ECHA
EUROPEAN CHEMICALS AGENCY

Illustrative examples with the OECD QSAR Toolbox workflow

Part 2: Case studies

Illustrative examples with the OECD QSAR Toolbox workflow
Part 2: Case studies

Reference: ECHA-14-R-07-EN
Cat. number: ED-02-14-124-EN-N
ISBN: 978-92-9244-433-4
DOI: 10.2823/83105
Date: February 2014
Language: English

If you have questions or comments in relation to this document please send them (quoting the reference and issue date) using the information request form. The form can be accessed via the 'Contact ECHA' page at: http://echa.europa.eu/contact

European Chemicals Agency
Mailing address: P.O. Box 400, FI-00121 Helsinki, Finland
Visiting address: Annankatu 18, Helsinki, Finland

Table of Contents

1. Introduction

Part 2: "Case studies" contains four illustrative examples of in-silico prediction of chosen endpoints, generated by using the QSAR Toolbox version 3.2. Computational methods cannot reliably cover all possible data gaps for all the imaginable chemical structures and endpoints. Thus, the chemicals in this document have been chosen at random amongst typical industrial chemicals. The endpoints selected are relevant for the third REACH registration deadline and require animal testing (vertebrates in three examples out of four) for the experimental assessment.

The case studies cover three different endpoints:

- Skin sensitisation (Annex VII of the REACH Regulation)
- Short-term toxicity testing on invertebrates (preferred species Daphnia) (Annex VII of the REACH Regulation)
- Short-term toxicity testing on fish (Annex VIII of REACH Regulation)

For each chemical, a single endpoint result is predicted using experimental data for similar substances. The analogues with available data are called "source" substances, and the ones with data-gaps (to be filled) are called "target" substances.

For skin sensitisation, the two case studies have been selected to show examples of positive and negative predictions for the Murine Local Lymph Node Assay (LLNA), which is a first choice test under REACH to assess the skin sensitisation endpoint.

One example is provided for short-term toxicity on Daphnia and one for fish short-term toxicity.

The illustrative examples follow a common layout. First, a general introduction of the endpoint is provided. Second, the Toolbox-relevant contents for the endpoint are described. Third, the prediction workflows are illustrated for each target chemical. Finally, the reliability of the prediction and the robustness of the category are commented. It is important to note that no effort has been made to validate the predictions by experimental data. The exercise is not intended to provide examples for documentation of prediction but rather to illustrate the workflow for the selection of analogues, starting with structural similarity (common functional groups), followed by mechanistic similarity and other refinements (if necessary) and assessment of the robustness elements.

The use of high quality data is crucial to obtain reliable predictions. The resident databases of the Toolbox have been incorporated into the Toolbox as they have been donated. The OECD does not formally provide any quality assurance of data within the Toolbox. Therefore, whilst the OECD expects these databases to include reliable data it does not take any responsibility for their use. Nevertheless, as far as available, the Toolbox provides information on the quality assurance process performed by the donator, as well as references to experimental results published in the open literature. This should help the user to assess whether the data are adequate for their purposes.

The quality of the data reported in the "ECHA Chem" database is under the responsibility of the registrants and has not been curated by ECHA. The Toolbox user who intends to use the "ECHA Chem" database is recommended to carefully verify the quality of every data point used for the prediction. The database includes many endpoint data for registered substances under REACH and allows access to more detailed information for the quality check of data, such as (robust) summaries. There is a hyperlink in the Toolbox that connects data points to the relevant information in the registration dossiers. In these examples, "ECHA Chem" data have not been used.

2. Case studies

2.1 SKIN SENSITISATION

2.1.1 General description of the endpoint

A sensitiser is an agent that is able to cause an allergic response in susceptible individuals, i.e. the characteristic adverse health effects of allergic contact dermatitis or atopic dermatitis may be provoked by subsequent exposure via the skin.

A range of *in vivo* methods exists that has been proven to be very accurate in terms of the predictive identification of chemicals that possess skin sensitising properties. The necessity to reduce the number of animal tests and to refine them, together with the need for simpler protocol and a quantitative outcome, has led the Local Lymph Node Assay (LLNA) to become the method of first choice for assessing the skin sensitisation potential. It is based upon the characteristics of induced proliferative responses of lymphocytes in draining lymph nodes following topical exposure of chemicals to mice. The endpoint is the stimulation index (SI), which gives a ratio of thymidine incorporation in the proliferating lymphocytes in lymph nodes from dosed animals compared to the thymidine incorporation in the proliferating lymphocytes in lymph nodes from vehicle-treated control animals. The test is positive when the stimulation index (SI) is greater than 3 for any of the dose concentrations. The EC3 value, interpolated from the dose response curve, is the effective concentration of the test substance required to produce a threefold increase in the stimulation index compared to vehicle-treated controls.

Under REACH, sensitising potential needs to be assessed for chemicals above the one tonne threshold according to Annex VII.

2.1.2 QSAR Toolbox relevant contents for skin sensitisation

The Toolbox covers the skin sensitisation endpoint with specific profilers and dedicated databases.

There are three relevant profilers under the general mechanistic profiler branch:

- Protein binding by OASIS v 1.2 (101 categories)
- Protein binding by OECD (102 categories)
- Protein binding potency (90 categories)

Protein binding by OASIS is also present in the endpoint specific profiler group with the name "Protein binding alerts for skin sensitisation by OASIS v 1.2". It contains a total of 100 categories but these are defined slightly differently compared to the mechanistic profiler "Protein binding by OASIS v 1.2". The protein-binding is relevant for skin sensitisation since one unifying characteristic of chemical allergens is that they react with proteins for the effective induction of skin sensitisation. The majority of chemical allergens are electrophilic and react with nucleophilic amino acids. More details about these profilers are reported here:

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). Profilers for speciation/transformations could be applied in a separate step. Different

alert compilations were analysed in order to place the information from 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 (acylation, Michael addition (MA), Schiff base formers, nucleophilic substitution (where the central carbon atom is directly accessible to the nucleophile, SN2) and nucleopilic substitution in an aromatic ring (SNar)), 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 either at a level of a mechanistic alert (e.g. Michael addition) or at a structural alert level (e.g. Michael addition/quinones) using this profiler. The "Protein binding by OECD" profiler contains 16 mechanistic alerts covering 52 structural alerts. These data are supported by mechanistic chemistry and references to the scientific literature (the metadata).

Protein binding by OASIS version 1.2
These protein-binding alerts have been developed by industry consortia involving ExxonMobil, Procter&Gamble, Unilever, Research Institute for Fragrance Materials (RIFM), Dow and Danish National Food Institute with the Laboratory of Mathematical Chemistry Bourgas and the partnership of Dr Roberts, as a part of the TIMES model to predict skin sensitisation. Under the scope of a research agreement signed in 2007 with Professor Mekenyan (OASIS - LMC), L'Oreal contributed to the assessment and refinement of chemical categories provided in the QSAR Toolbox. The scope of this profiler is to investigate the presence of alerts within target molecules responsible for interaction with proteins. The list of 101 structural alerts has been separated into 11 mechanistic domains. Each of the mechanistic domains has been separated into more than two mechanistic alerts. The profiling result assigns a target to the corresponding structural alert, mechanistic alerts and domain.

Protein binding potency
This profiler is developed on the base of empirical data for thiol reactivity expressed by the *in chemico* RC50 value. Data are obtained by measuring target chemical covalent binding with the thiol group of glutathione (GSH). The structural alerts for protein-binding are extracted from about 400 chemicals comprised within GSH Experimental RC50. All the chemicals have two common electrophilic mechanisms of interaction with GSH – interaction via SN2 and interaction via MA mechanism. The other three mechanistic domains of the protein-binding profilers are currently not associated with protein-binding potency prediction (a "Not possible to classify…" result appears in the data matrix).

Protein binding for skin sensitisation by OASIS
These protein-binding alerts have been developed by industry consortia involving ExxonMobil, Procter&Gamble, Unilever, Research Institute for Fragrance Materials (RIFM), Dow and Danish National Food Institute with the Laboratory of Mathematical Chemistry Bourgas and the partnership of Dr Roberts, as a part of the TIMES model to predict skin sensitisation. The scope of this profiler is to investigate the presence of alerts within target molecules responsible for interaction with proteins and especially with skin proteins. This profiler accounts for incapability of some chemicals having an alert to interact with skin due to electronic and steric factors. This is explicitly defined by inhibition masks associated with the some alerts. The list of 100 structural alerts has been separated into 11 mechanistic domains. Each of the mechanistic domains has been separated into more than two mechanistic alerts. The profiling result assigns a target to the corresponding structural alert, mechanistic alerts and domain.

The Toolbox encloses two dedicated databases for skin sensitisation:

A. "Skin sensitisation", which includes 1 035 chemicals and 1 570 experimental data points (includes the OASIS skin sensitisation database and the Liverpool John Moores University skin sensitisation database).
B. "Skin sensitisation ECETOC", with 39 chemicals and 42 experimental data points.

A. Skin sensitisation database gathers two data sets:
"OASIS Skin Sensitisation" includes 876 chemicals tested either in LLNA, or in Guinea Pig Maximisation Test (GPMT), or chemicals from the BfR list (Kayser, D., E. Schlede. (eds.) Chemikalien und Kontagtalergie - eine bewertende Zusammenstellung. Medizin&Wissen Verlagges, Munchen. ISBN 3-86094-163- 1, 2001). For 102 chemicals more than one skin sensitisation data point are stored in the database. Based on the observed skin sensitisation effect, the chemicals are assigned to three classes:

- Class 2 (strong sensitisers). These are chemicals with EC3 < 10% in the LLNA test, or showing positive response in more than 30% of tested animals in the GPMT or classified as significant contact allergens by BfR (Category A).
- Class 1 (weak sensitisers). These are chemicals with EC3 within the range of 10% to 50% in the LLNA test, or showing a positive response in 1% to 30% of tested animals in the GPMT or possessing a solid-based indication for contact allergenic effects according to BfR (Category B).
- Class -1 (non sensitisers). These are chemicals without positive effects in the LLNA and GPMT or with insignificant/questionable contact allergenic effects according to BfR (Category C). Note that class 1 (weak sensitisers) and Class -1 (non sensitisers) are different categories and not to be confused.

The LJMU database complements (with partial overlapping) the OASIS database with skin sensitisation data: it contains 212 chemicals tested by more than one test method. The chemicals are assigned into four classes.

- 2 (strongly sensitising) - four chemicals;
- 1 (moderately sensitising) - 47 chemicals;
- -1 (non sensitising) 155 chemicals
- 0 (ambiguous) - six chemicals.

B. The "Skin sensitisation ECETOC" database includes experimental results on skin and respiratory sensitisation. The ECETOC database is a compilation of different test results and provides qualitative outcome. The data are stored in the fields Human health/Sensitisation/Skin/*In vivo*/Undefined Assay.

In the examples below, a qualitative assessment of skin sensitisation potential is provided according to the "Danish EPA scale", which converts the results into positive/negative/equivocal according to the following rules:

- for quantitative data, EC3 values >50% are converted to "negative", EC3 values <50% converted to "positive".
- for qualitative data recorded according to different scales, only the non-sensitisers are converted to "negative", while all the other possible categories (weak/moderate/strong sensitisers) are considered "positive". Ambiguous outcomes are converted into "equivocal".

For classification purposes, the thresholds in the CLP Regulation and the respective guidance should be checked.

The scale is selected during the data-gap filling, when a dialogue "Possible data inconsistency" window pops up before the generation of the first plot requiring, among others, the selection of the scale/unit to use for the prediction.

A proof-of-concept tool to predict an adverse outcome pathway (AOP) for skin sensitisation from *in vitro* data is also implemented in version 3.2 of the Toolbox but this is not discussed in this document.

2.2 EXAMPLE 1. 3-METHYL-2-BUTANOL

The first example illustrates the workflow for a qualitative assessment of the skin sensitisation potential for 3-methyl-2-butanol (CAS number 598-75-4, SMILES: C(C)(O)C(C)C).

2.2.1 Input

The structure can be introduced in the Toolbox using the drawing tool, the CAS number or the SMILES notation. A structure could also be selected by a database, inventory, or user list.

2.2.2 Profiling

The outcomes of the mechanistic/endpoint relevant profilers for skin sensitisation potential are:

- No alert found (Protein binding by OASIS)
- No alert found (Protein binding by OECD)
- Not possible to classify (Protein binding potency)
- No alert found (Protein binding alerts for skin sensitisation by OASIS)

It could be noted that the protein-binding potency profiler predicts the reactivity quantitatively only if the protein-binding mechanism is MA or SN2. Figure 1 shows the profiling result for 3-methyl-2-butanol (black font indicates no alert, red font indicates alerts; yellow indicates parameters, which have not been calculated in the Toolbox).

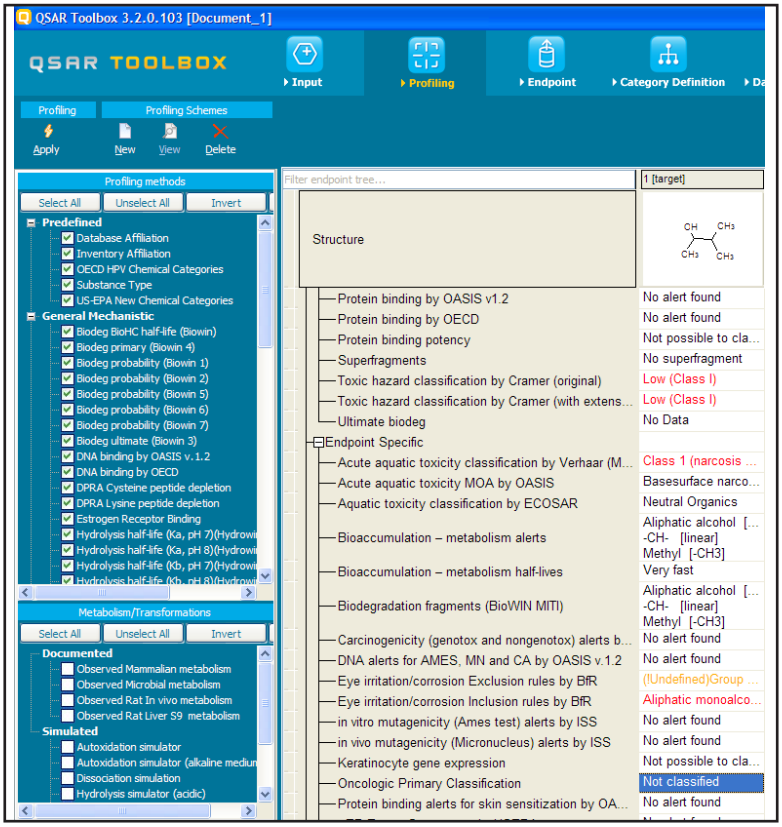


FIGURE 1. RESULTS OF PROFILING FOR 3-METHYL-2-BUTANOL

2.2.3 Endpoint

The two skin sensitisation databases are selected:

A. Skin sensitisation
B. Skin sensitisation ECETOC

No experimental data were found in the Toolbox databases for this substance.

2.2.4 Category definition

The starting category is based on a chemical profiler "US EPA new chemical categories", which classifies the molecule as "Neutral Organic" ("Strict" option checked), Figure 2. The sub-categorisation will be performed in the data-gap filling module in order to visualise and better follow the process. 119 chemicals are gathered in the category (118 molecules from the databases and the target) and data are extracted/gathered for all category members.

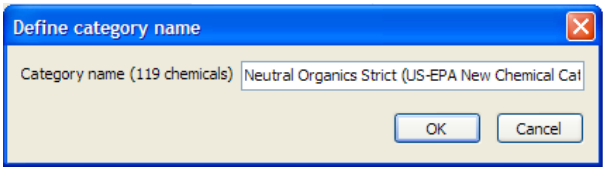


FIGURE 2. POP-UP WINDOW FOR NEUTRAL ORGANICS CATEGORY

2.2.5 Data-gap filling

In the data-gap filling module the user chooses:

- the endpoint to predict (skin sensitisation potential according to LLNA)
- the type of approach for filling the data-gap (read-across)
- the sub-categorisation strategy (described below)

At the end of the process the user can decide if the result is satisfactory. If so, the prediction can be accepted for reporting and reported as a standalone file or exported to IUCLID.

In this example, to predict the skin sensitisation potential of the target chemical the LLNA EC3 is selected as the endpoint to be predicted as shown in Figure 3.

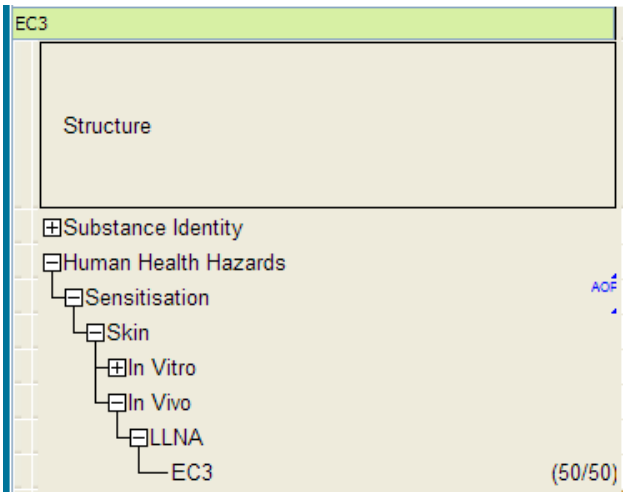


FIGURE 3. SELECTION OF LLNA EC3 IN THE "ENDPOINT TREE". 50 DATA POINTS FOR 50 SUBSTANCES ARE AVAILABLE

50 data points for as many molecules are available (one data point per molecule). The qualitative prediction is performed selecting "Read-across" to fill the data-gap and selecting the scale/unit "Skin sensitisation (Danish EPA)" from the "Possible data inconsistency window".

The first plot is then automatically generated (Figure 4). It shows the aggregated test data for the LLNA as the dependent variable (on Y-axis), and the calculated from EpiWIN (KOWIN ver. 1.67) octanol-water partition coefficient (log K_{ow}) as the dependent variable (on the X-axis) for the Neutral Organics category. The red dot is the target substance and the brown dots are the nearest neighbours in the descriptor space of the plot (log K_{ow} in this case, could be also water solubility, vapour pressure, etc.). Notably, the nearest neighbours may change if a different descriptor has been used.

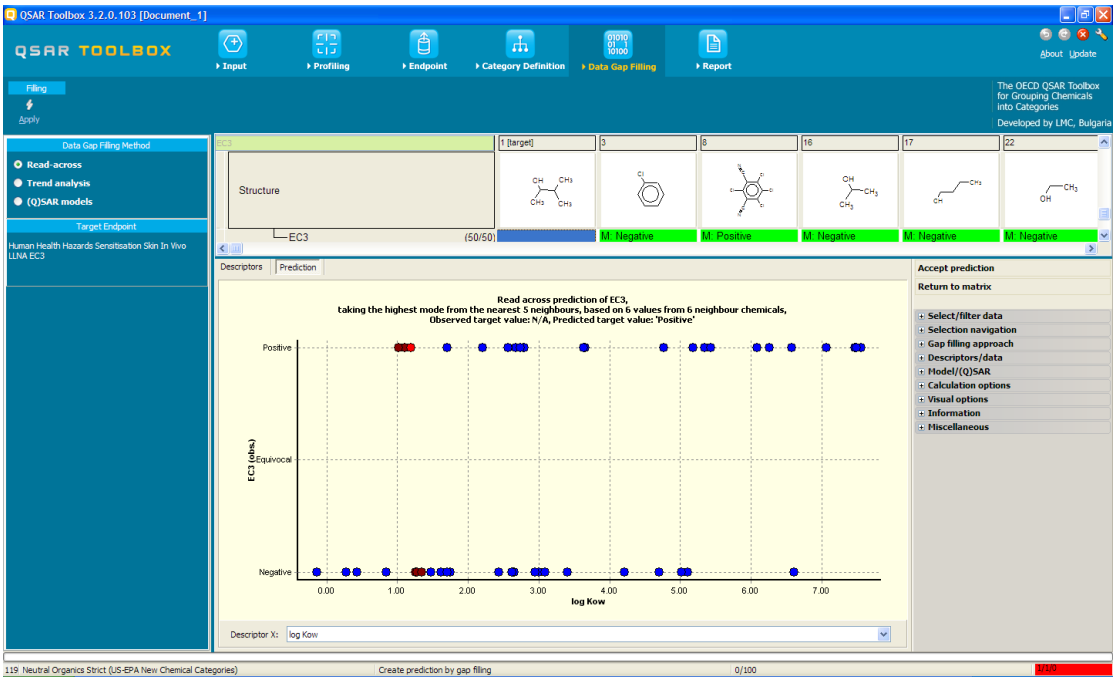


FIGURE 4. AUTOMATICALLY GENERATED PLOT FOR THE STARTING CATEGORY.

The default "read-across" technique in the Toolbox prediction based on five nearest neighbours, produces a "negative" result for the target chemical, however it is obvious that the category is still wide and the results too disperse for a meaningful prediction.

A sub-categorisation according to "organic functional group (nested)" profiler is then performed (Figure 5). Only alcohols are selected (blue dots on the plot below) and they remain in the analysis, while chemicals with more or different functional groups are excluded from the set of analogues (the green dots on the plot below). A special case is represented by 2-propanol, which is a simple alcohol that should be comprised in the group but would be eliminated because the profiler recognises an "overlapping groups" functionality (iso-propyl and alcohol groups on the same carbon). In order to keep this molecule, which potentially brings useful additional information for the prediction, the "Overlapping groups" category has been manually deselected from the list of categories whose members have to be removed.

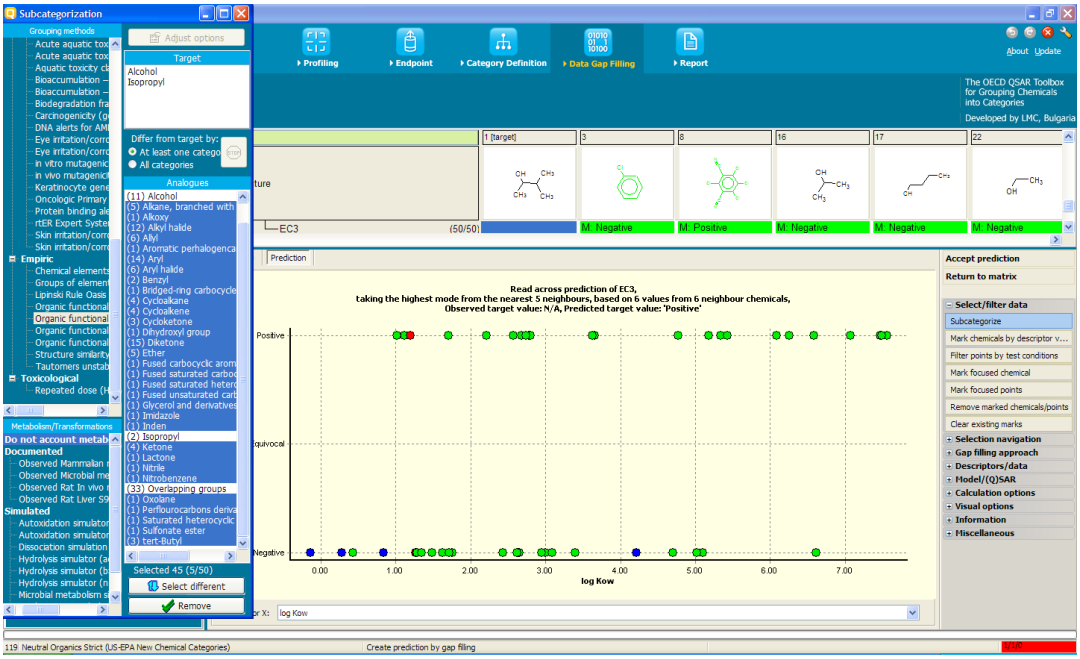


FIGURE 5. ALCOHOL SUB-CATEGORISATION ACCORDING TO ORGANIC FUNCTIONAL GROUPS (NESTED).

Analogues with other functionalities, such as aryl, ketone, ether, and so on are eliminated. Five chemicals (Figure 6) are left in the group of analogues, which have a hydroxyl group attached to the primary or secondary carbon in an alkyl saturated chain. The analysis of the available data for these five analogues (source substances) shows that all of them have an experimental negative outcome from LLNA, so the target is predicted to be negative as well. The prediction is made using the "highest mode" option, based on data from five neighbours by default. The number of neighbours could be changed by the user at this point. The approximation type could also be changed to "minimal", "maximal", "median", and "mode" (in addition to "highest mode").

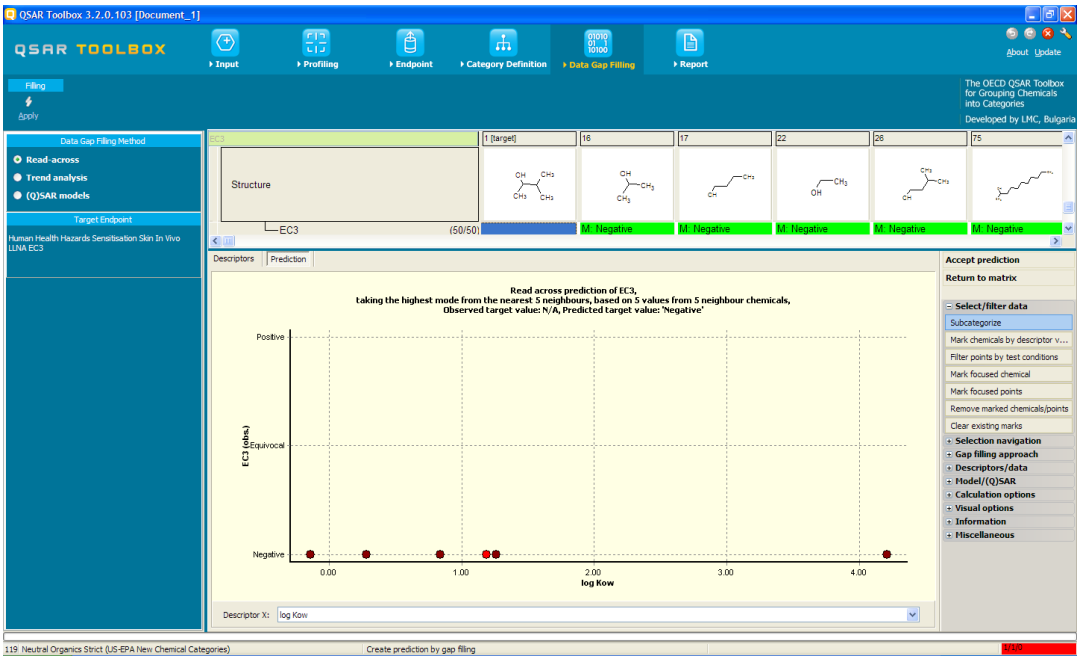


FIGURE 6. STRUCTURES OF THE TARGET AND OF THE SOURCE SUBSTANCES. THE EXPERIMENTAL DATA FOR SKIN SENSITISATION ARE HIGHLIGHTED IN GREEN.

Up to this point of the analysis, the mechanistic or endpoint specific profilers have not been employed. The protein-binding profiles have to be checked to confirm the mechanistic similarity between the analogues and the target. Sub-categorising according to the three skin sensitisation specific profilers does not exclude any other chemical, since they all have the same profile (no alert found, Figure 7).

2.2.6 Category robustness and prediction reliability

The selection of analogues started from a large pool of generally non-reactive chemicals (neutral organics according to the US EPA new chemical categories), narrowed down the selection to aliphatic alcohols using "organic functional group (nested)" profiler, and then further limited the selection to the substances that have experimental data for skin sensitisation from the LLNA assay. In limiting the prediction dataset, we excluded substances with functional groups other than methyl, methylene and the hydroxyl group connected to a primary or secondary carbon, thus satisfying the requirement for structural and functional similarity.

The common protein-binding profiles of all category members indicate homogeneity of the selected analogues from a mechanistic point of view for skin sensitisation. None of the category members triggers alerts for protein-binding, and no selected analogue was experimentally found to be a skin sensitiser. This evidence supports the likelihood of the target being a non sensitiser as well. Notably, in alert-based systems a lack of alert may not always mean a lack of hazard. This is a common characteristic of the alert-based systems – they do not show an alert if there is no knowledge for it. In this case, we consider the class of the aliphatic alcohols to be well enough characterised for skin sensitisation. Consequently, the lack of alert could be interpreted as a lack of protein-binding, and the negative outcome of the LLNA for the analogues as a confirmation of the absence of skin sensitisation potential. Products of skin metabolism and auto-oxidation for the target have to be checked too. They can be predicted with the Toolbox in the profiling module. The software predicts two skin metabolites for the target, while no auto-oxidation products are predicted. The two metabolites do not fire protein-binding alerts and thus are not considered of concern for the skin sensitisation endpoint.

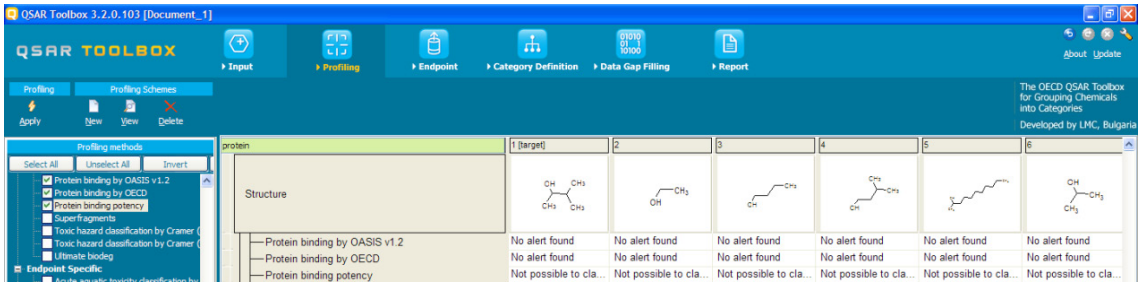


FIGURE 7. RESULTS OF PROTEIN-BINDING PROFILING FOR THE CATEGORY MEMBERS.

The data quality also needs to be addressed. The Toolbox does not contain reliability scores for experimental toxicity data. However, the user can consult and extract metadata regarding the experimental conditions for each data point. In Table 1 we report the reference source for each data point (although more information can be extracted from the software). Unfortunately, two of the references in that table are not traceable. It could be possible to retrieve references from sources other than the Toolbox. In the present example, it is still possible to exclude these two data points from the analysis and refer only to the remaining three analogues for read-across. There are additional considerations that support the exclusion of ethanol and 2-undecanol – the ethanol is used often as a solvent for other substances and could be considered by some leaving the test system due to its volatility. 2-Undecanol was commented already by others and seems to be lacking sufficient solubility in water for skin penetration. Thus, the consistency in the experimental conditions should also be checked in the source or in the metadata provided by the Toolbox or in the literature.

CAS	NAME	Reference source
64-17-5*	Ethanol	Givaudan
71-36-3	1-Butanol	Gerberick GF et al, Dermatitis, 16 (4): 1-46, 2005
1653-30-1*	2-Undecanol	Givaudan
67-63-0	2-Propanol	Gerberick GF et al, Dermatitis, 16 (4): 1-46, 2005
123-51-3	3-Methyl-1-butanol	Patlewicz G et al, Regulatory Toxicology and Pharmacology, 48: 225-239, 2007

TABLE 1. EXPERIMENTAL VALUES REFERENCES. * INDICATES ANALOGUES LACKING TRACEABLE SOURCE.

As a conclusion, the negative prediction for 3-methyl-2-butanol based on three nearest neighbours (structurally similar and mechanistically homogenous) can be considered coherent with the absence of alerts and supported by experimental data for similar substances. Thus, the prediction can be accepted for reporting. Please note that here the term “accept” describes a Toolbox button indicating that the prediction is ready for reporting, and is not related to acceptance in terms of regulatory application.

2.2.7 Report and category summary

Once the prediction is accepted for reporting, the report module of the Toolbox gives the possibility to semi-automatically generate a standalone file to record the result with some pre-defined information, coming from the modelling procedure (applicability domain, information about the group members, etc.). The report form also contains manually editable fields that the user should fill to justify the procedure that was

followed. Here, we summarise the considerations that need to be given for the assessment of the category and the prediction. They are based on good practices and expert judgement and are not hard-coded in the Toolbox as such.

a. Category definition
A category of four analogues (three analogues with experimental data and the target substance) is defined for the purpose of data-gap filling for skin sensitisation (as a qualitative value for EC3 from LLNA assay) for 3-methyl-2-butanol. The analogues are shown in Table 1 (the table also shows the analogues that were subjectively excluded because of a lack of traceability of their toxicological information). In the example, considerations on impurities are omitted. For REACH substances, impurities need to be analysed and discussed.

b. Hypothesis for grouping chemicals
The working hypothesis is that the primary and secondary aliphatic alcohols with the C3-C5 aliphatic chain can be grouped together for the prediction of skin sensitisation. The hypothesis is based on the structural, physicochemical and mechanistic similarity of the group members and on the assumption that transformation products as a result of auto-oxidation and skin metabolism do not break this hypothesis.

c. Category description
- The category members are aliphatic alcohols that do not have any other functional group but methyl and methylene as described above, with a chain length between three and five carbon atoms. For each analogue, an experimental value for skin sensitisation, obtained in the LLNA assay, is available. All analogues are found to be negative in the LLNA assay. Original quantitative EC values are converted into a qualitative scale using the 50% threshold (encoded in the "Danish EPA" scale).

- The covered hydrophobicity range in terms of experimental log K_{ow} for the group of three analogues with traceable data is between 0.05 and 1.25. Accounting the experimental variability and uncertainty, the log K_{ow} of the target substance (1.28) can reasonably be considered in this range. Experimental water solubility covered is from $2.67*10^4$ mg/L to $1.00*10^6$ mg/L. The experimental solubility of the target compound is $5.60*10^4$ mg/L. The hydrophobicity and solubility are relevant parameters for the endpoint of skin sensitisation because they determine the potential for dermal bioavailability of the substances through passive diffusion. Molecular weight and vapour pressure of the target are also in the range of these properties, determined by the group members (from 60.1 to 88.1 Da for molecular weight and from 2.37 to 45.5 mmHg for experimental vapour pressure). Based on these parameters, it is assumed that all group members, including the target substance, are absorbed well into intact skin and the prediction could be done as interpolation in the descriptor space, defined by the hydrophobicity and solubility. The target falls also in the structural domain defined by the group members. In fact, all structural variations are covered by experimental data.

- The mechanistic similarity is evaluated using the protein-binding profilers available in the Toolbox. All profiling results show no protein-binding for all group members. The assumption is that for the concerned chemical class a lack of alerts for protein-binding indicates a lack of skin sensitising potential. The assumption is supported by the fact that all the source substances have negative experimental outcome from the LLNA assay. For other chemical classes and combinations of functional groups, the uncertainty might be too high to make such an assumption from the profiling results.

- Metabolites and auto-oxidation products were considered. There are no simulated auto-oxidation products and the predicted metabolites do not fire any alert for protein-binding and skin

sensitisation. The target molecule does not ionise/hydrolyse, and does not produce tautomers. For other chemistry, molecular speciation and especially the tautomerism might also need to be checked.

d. Strategy used (data-gap filling method)
Qualitative read-across. The prediction is made using the "highest mode" option, based on data from three (selected) neighbours (five by default). If "trend analysis" has been used, the details should be provided here. However, "trend analysis" is not considered very suitable for predicting qualitative (yes/no) endpoints.

2.3 EXAMPLE 2. P-NITROBENZOYL CHLORIDE

The second example illustrates the workflow for a qualitative assessment of the skin sensitisation potential for p-nitrobenzoyl chloride (CAS number 122-04-3, SMILES: C(=O)(Cl)c1ccc(N(=O)=O)cc1).

2.3.1 Input

The structure can be introduced in the Toolbox using the drawing tool, the CAS number or the SMILES notation. A structure could also be selected by a database, inventory, or user list.

2.3.2 Profiling

The outcomes of the mechanistic/endpoint relevant profilers for skin sensitisation potential are:

- (Thio)Acyl and (thio)carbamoyl halides and cyanides (Protein binding by OASIS)
- Acyl halides (Protein binding by OECD)
- Not possible to classify (Protein binding potency)
- (Thio)Acyl and (thio)carbamoyl halides and cyanides (Protein binding for skin sensitisation by OASIS)

It could be noted that the protein-binding potency profiler only predicts the reactivity quantitatively if the protein-binding mechanism is of Michael-type or SN2. In this step, Figure 8 shows the profiling results for the target chemical (black font indicates no alert; red font indicates alerts; yellow indicates parameters, which have not been calculated in the Toolbox).

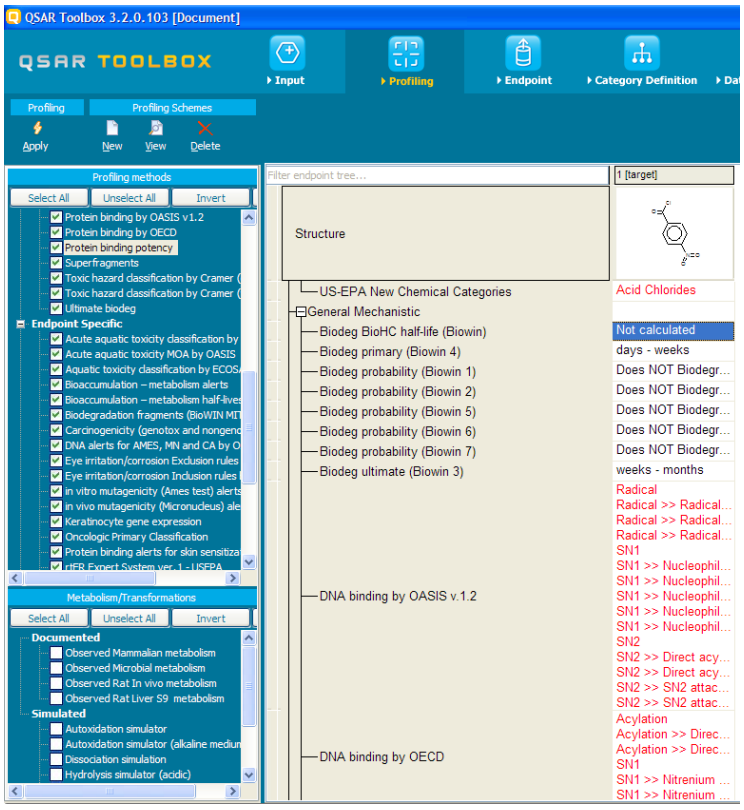


FIGURE 8. RESULTS OF PROFILING FOR THE TARGET SUBSTANCE.

2.3.3 Endpoint

The two skin sensitisation databases are selected:

A. "Skin sensitisation"
B. "Skin sensitisation ECETOC"

No experimental data were found in the Toolbox databases for this substance.

2.3.4 Category definition

The starting category is based on the chemical profiler "US EPA new chemical categories", which classifies the molecule as "Acid chloride" (strict), Figure 9. The sub-categorisation will be performed in the data-gap filling module in order to visualise and better follow the process. Ten chemicals are gathered in the group (the target and nine molecules from the databases) and data are extracted/gathered for all category members.

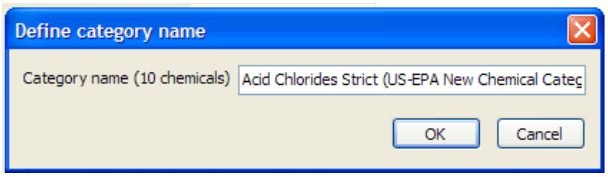


FIGURE 9. POP-UP WINDOW FOR ACID CHLORIDES CATEGORY.

2.3.5 Data-gap filling

In the data-gap filling module the user chooses:

- the endpoint to predict (skin sensitisation potential according to LLNA)
- the type of approach for filling the data-gap (read-across)
- the sub-categorisation strategy (described below)

At the end of the process, the user can decide if the result is satisfactory. If so, the prediction can be accepted and reported as a standalone file or exported to IUCLID.

In this example, to predict the skin sensitisation potential of the target chemical the LLNA EC3 (Figure 10) is selected as the endpoint to be predicted. Eight of these analogues had EC3 data for the LLNA assay.

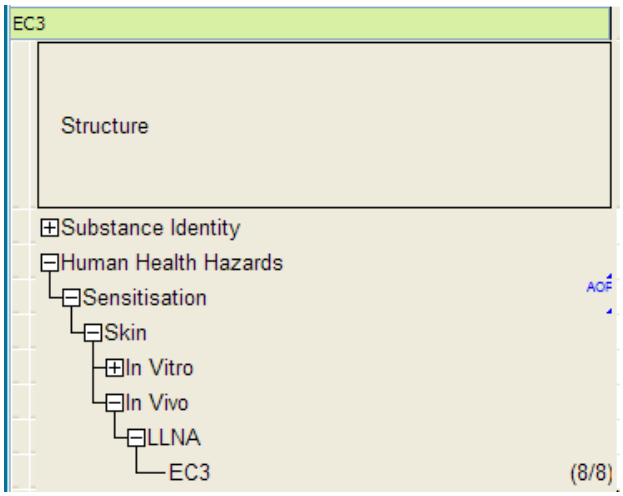


FIGURE 10. SELECTION OF LLNA EC3 IN THE "ENDPOINT TREE". EIGHT EC3 DATA POINTS FOR EIGHT SUBSTANCES ARE AVAILABLE.

"Read-across" is selected for filling the data-gap and the "Skin sensitisation (Danish EPA)" is selected as the scale/unit. Then, the first plot is generated (Figure 11). It shows the aggregated test data for the LLNA as the dependent variable (on Y-axis), and the calculated from EpiWIN (KOWIN ver. 1.67) octanol-water partition coefficient (log K_{ow}) as the dependent variable (on the X-axis). The red dot is the target substance and the brown ones by default indicate the nearest neighbours in the descriptor (log K_{ow}) space.

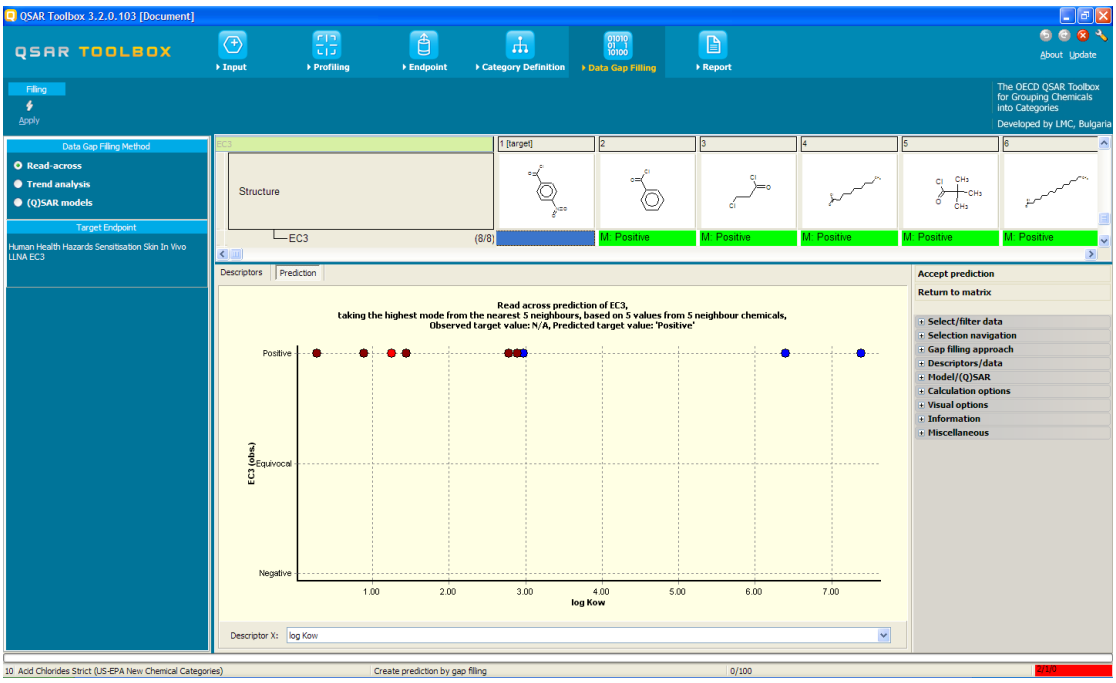


FIGURE 11. AUTOMATICALLY GENERATED PLOT FOR THE STARTING CATEGORY.

Eight analogues are found with the structural profiler (US EPA new chemical categories). The mechanistic profilers (both OASIS and OECD protein-binding profilers) identify one different substance (propanoyl chloride, 3-chloro-), which was removed because it was recognised as an alpha-haloalkane (in addition to acyl

halide). The remaining seven compounds show the same mode of action, determined by the presence of the only acyl halide functional group. Thus, the selection resulted in seven source substances with experimental LLNA data (Figure 12).

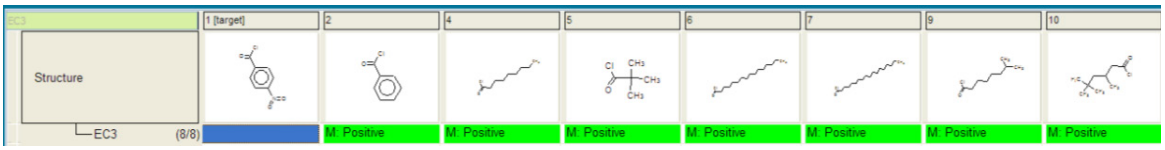


FIGURE 12. STRUCTURES OF THE TARGET AND SOURCE SUBSTANCES. THE EXPERIMENTAL EC3 VALUE FOR LLNA IS HIGHLIGHTED IN GREEN.

Six of the analogues have a quantitative result for EC3 from LLNA (expressed as % value, all below 10%), and for one substance (CAS 328230-2) there is only a semi-quantitative description (strongly positive). Thus, all selected analogues have a strongly positive result to LLNA. The selected analogues may look structurally different but they all contain an acid chloride group. The assumption is that the mechanism of action and consequent skin sensitisation potential is due to the presence of the alert, which in this case is more relevant than the structural similarity in the rest of the molecule. Figure 13 shows the results from the mechanistic/ endpoint specific profilers, which are consistent within the group members.

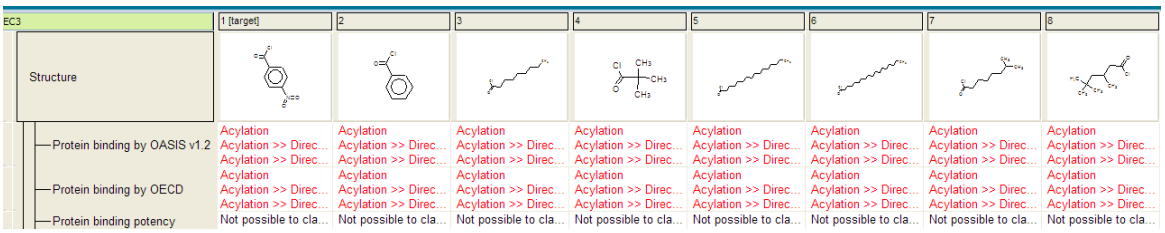


FIGURE 13. RESULTS OF THE PROTEIN-BINDING PROFILING FOR THE CATEGORY MEMBERS.

2.3.6 Category robustness and prediction reliability

The selection of the analogues started from the chemical category of Acid halides, according to US EPA new chemical categories profiler. The group was then restricted to chemicals having experimental LLNA data.

The three mechanistic protein-binding profilers and the protein-binding profiler for skin sensitisation provide the same outcome for all the category members, confirming the homogeneity of the category from a mechanistic point of view. All protein-binding profilers (without the potency profiler, where prediction was not possible) refer to the acylation mechanism. Figure 14 shows the mechanistic explanation of the protein-binding provided by the Toolbox (the user can see the profilers' explanation by right-clicking on the profiler outcome in the matrix, and then selecting "explain" and "details"). The mechanism refers to the possible nucleophilic reaction of the target substance with proteins, which is one of the established modes of action for the skin sensitisers.

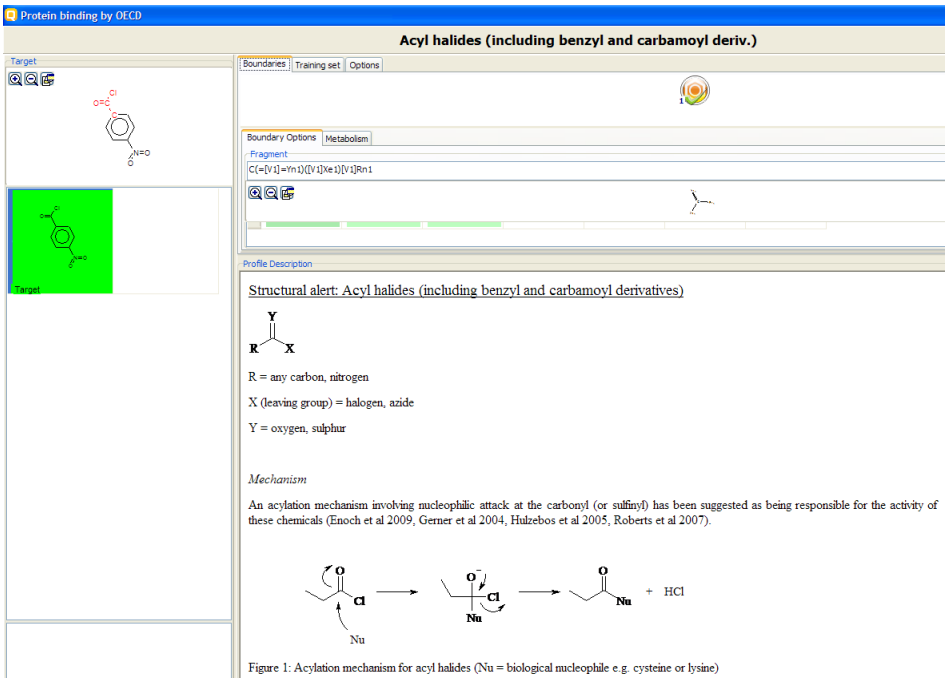


FIGURE 14. EXPLANATION OF THE MECHANISM OF ACTION FOR THE CATEGORY OF ACYL HALIDES ACCORDING TO PROTEIN-BINDING BY OECD PROFILER.

Data quality also needs to be addressed. The Toolbox does not contain reliability scores for experimental toxicity data. However, the user can consult and extract metadata regarding the experimental conditions for each data point and can check the references. In Table 2, we report the reference source for each data point (although more information can be extracted from the software). Unfortunately, two of the experimental values are not traceable from the Toolbox. An additional literature search indicated that the skin sensitisation potential of benzoyl chloride is published by Basketter DA et al, Food and Chemical toxicology, 37 (12): 1167-1174, 1999 and that of octadecanoyl chloride was found in Ashby J et al, Toxicology 103: 177-194, 1995.

CAS	NAME	Reference
98-88-4	Benzoyl chloride	Unilever
764-85-2	Nonanoyl chloride	Gerberick GF et al, Dermatitis, 16 (4): 1-46, 2005
3282-30-2	Propanoyl chloride	Patlewicz, G et al, Regulatory Toxicology and Pharmacology, 48: 225-239, 2007
112-67-4	Hexadecanoyl chloride	Gerberick GF et al, Dermatitis, 16 (4): 1-46, 2005
112-76-5	Octadecanoyl chloride	Unilever
57077-36-8	Isononanoyl chloride	Gerberick GF et al, Dermatitis, 16 (4): 1-46, 2005
36727-29-4	3,5,5-Trimethylhexanoyl chloride	Gerberick GF et al, Dermatitis, 16 (4): 1-46, 2005

TABLE 2. REFERENCES FOR THE EXPERIMENTAL VALUES RETRIEVED FOR THE GROUP MEMBERS.

All the evidence supports the likelihood of the chemical to be positive to LLNA. Moreover, quantitative data for the identified analogues suggest that the substance could be also qualified as a "strong sensitiser" (EC3 <10%). The prediction can be considered reliable and the result accepted for reporting. Please note that here the term "accept" describes a Toolbox button indicating that the prediction is ready for reporting, and is not related to acceptance in terms of regulatory application.

2.3.7 Report and category summary

Once the prediction is accepted for reporting, the report module of the Toolbox gives the possibility to semi-automatically generate a standalone file to record the result with lots of pre-defined information, coming from the modelling procedure (applicability domain, information about the group members, etc.). The report also contains manually editable fields that the user should fill to justify the procedure that was followed. Here, we summarise the considerations that need to be given for the assessment of the category and the prediction. They are based on good practice and expert judgement and are not hard-coded in the Toolbox as such.

a. Category definition
   A category of eight analogues (seven source substances with experimental data and a target) is defined for the purpose of data-gap filling for skin sensitisation (as a qualitative value for EC3 from LLNA assay) for the p-nitro benzoyl chloride. The seven analogues are shown in Table 2. In the example, considerations on impurities are omitted. For REACH substances, impurities need to be analysed and discussed.

b. Hypothesis for grouping chemicals
   The working hypothesis is that acid chlorides (both aliphatic and aromatic) can be grouped together (within physico-chemical boundaries determined by the group members) for the prediction of skin sensitisation because of their structural and mechanistic similarity. The transformation products are not checked in this case because the parents themselves are found to be sensitisers.

c. Category description
   - The category members are aliphatic and aromatic acyl halides. The aliphatic analogues could be linear and branched, without influence if the acyl halide group is connected to primary, secondary, tertiary, or aromatic carbon. The nitro group in the target substance is not supposed to reduce the activity – on the contrary, it may increase the electrophilicity of the molecule due to its electro-withdrawing effect. For each analogue a strongly positive experimental value for skin sensitisation, obtained in the LLNA assay, is available. Original quantitative EC values are converted into a qualitative scale using the 50% rule (encoded in the "Danish EPA" scale). However, the original EC values could be analysed as well.

   - The covered hydrophobicity range in terms of log K_{ow} for this group is between approximately 0.273 and 7.39. The target substance has a log K_{ow} of 1.25, which is well in the range defined by the postulated group members. The solubility covered is from $3.86*10^{-3}$ mg/L to $3.62*10^{4}$ mg/L. The solubility of the target compound is $4.13*10^{3}$ mg/L (please note that for this group of chemicals the Toolbox does not contain experimental information for log K_{ow} and water solubility). Here we refer to values calculated by the Toolbox. In the case of water solubility, "Water solubility (fragments)" was used for calculating water solubility. The hydrophobicity and solubility are relevant parameters for the endpoint of skin sensitisation because they determine the potential for dermal bioavailability of the substances through passive diffusion. The molecular weight of the target is also in the range of these properties, determined by the group members (from 123.0 to 303.0 Da). The vapour pressure of the target is just below the range defined by the analogues. Nevertheless,

a lower vapour pressure could not reduce the skin sensitisation potential (as it could be for higher values of vapour pressure, due to the volatility issues during the experiment). Based on these parameters, it is assumed that all group members, including the target substance, are absorbed well into the skin and the prediction could be done as interpolation in the descriptor space, defined by the hydrophobicity and solubility.

- The mechanistic similarity is evaluated using the protein-binding profilers available in the Toolbox. A nucleophilic reaction is identified as the mechanism underlying the protein-binding and consequent skin sensitisation effect. This assumption is confirmed by the strong positive outcome in the LLNA assay for all the source substances.

- Metabolites, auto-oxidation products, molecular speciation, and tautomerism were not checked in this example because the parent compounds are sensitisers themselves and drive a conservative prediction.

d. Strategy used (data-gap filling method)
Qualitative read-across. The predicted result is "positive" since all seven analogues had "positive" experimental results according to the "Danish EPA scale". The "Highest rank" option was used to fill the data-gap in the Toolbox.

There are indications that the category members might be corrosive to skin. This may affect the interpretation of the result. Under REACH column 2, a waiver could be used i.e. if the substance is corrosive to the skin no *in vivo* study is needed for skin sensitisation. We assume that in this case the test data for the analogues were generated according to the OECD Test Guideline for LLNA (OECD Guideline 429. Skin sensitisation: the local lymph node assay). Thus, the predicted data point can assumingly compare with these results. In fact, the experimental data points are taken from the compilation and a check of the original data source is always recommended.

2.4 SHORT TERM AQUATIC TOXICITY

(Due to the similar background of the endpoints, daphnid and fish short-term aquatic toxicity will be introduced together.)

2.4.1 General description of the endpoint

Information on aquatic toxicity is required to assess the hazard and risk of chemical substances to marine and freshwater organisms living in the water column. Information on aquatic toxicity can also be used as a basis for screening approaches (the equilibrium partitioning method) to predict toxicity to terrestrial and sediment organisms. A detrimental effect of a chemical can be measured after short-term and/or long-term exposure. The short-term (acute) toxicity is most often measured as a concentration, which is lethal to 50% of the organisms (lethal concentration, LC50, typically used for toxicity to fish) or causes a measurable effect to 50% of the test organisms (effective concentration, EC50, typically used for toxicity to Daphnia).

With respect to aquatic (pelagic) toxicity, Annex VII of the REACH Regulation requires information on short-term toxicity testing on invertebrates (preferred species Daphnia) and growth inhibition studies on aquatic plants (algae preferred); Annex VIII additionally requires short-term toxicity data on fish but the registrant can consider long-term toxicity data instead of short-term. Annexes VII and VIII also specify when the studies do not need to be conducted. Information on aquatic toxicity may be acquired from studies performed according to existing national and international guidelines, such as the EU testing methods and OECD test guidelines (TGs) which refer to internationally agreed testing methods.

For Daphnia acute toxicity (Annex VII), the reference TG is the OECD 202 (Daphnia sp., Acute Immobilisation Test), where the EC50 is determined at 48 hours.

For fish acute toxicity (Annex VIII), several tests fulfil REACH requirements, including OECD TG 203 (Fish, Acute Toxicity Test) which has a 96 hours duration.

Some so-called "difficult" substances may require special attention in terms of generation of new data or interpretation of existing data. Such substances might be associated with problems of solubility (not or very poorly dissolving in the water phase), bioavailability (e.g. resulting in non-constant exposure concentrations during the test), and/or concentration measurement (suitability of the analytical method). Examples of properties that can be challenging include low water solubility, ionisation, ability to form complexes, surface activity, colour, volatility, adsorption, abiotic or biotic degradation and photodegradation.

2.4.2 QSAR Toolbox relevant contents for aquatic toxicity

The Toolbox covers the aquatic toxicity with specific profilers and dedicated databases.

There are five relevant profilers for aquatic toxicity (not taking into account stability and degradation aspects), three endpoint-specific profilers and two mechanistic profilers. The endpoint-specific profilers contain structural alerts that have been identified as being associated with trends in toxicity from an analysis of aquatic toxicological data. The endpoint-specific profilers include:

Aquatic toxicity classification by ECOSAR
The Aquatic Toxicity Classification by ECOSAR profiler consists of molecular definitions developed by LMC and OECD to mimic the structural definitions of chemical classes within the US Environmental Protection Agency's Ecological Structure-Activity Program (ECOSAR). ECOSAR contains a library of

class-based SARs for predicting aquatic toxicity, overlaid with an expert decision tree based on expert rules for selecting the appropriate chemical class for evaluation of the compound. ECOSAR is currently programmed to identify 118 chemical classes. The profiler is introduced for chemical categorisation purposes using the class definitions from ECOSAR.

Acute aquatic toxicity MOA by OASIS
This profiler divides chemicals in different categories according to their acute toxic mode of action (MOA). 2D structural information is used to identify the MOA of chemicals. Based on theoretical and empirical knowledge, the following seven hierarchically ordered MOAs are distinguished: Aldehydes; alpha, beta-Unsaturated alcohols; Phenols and Anilines; Esters; Narcotic Amines; Basesurface narcotics.

Acute aquatic toxicity classification by Verhaar
The classification system separates a large number of small to intermediate organic chemicals into four distinct classes that can either be assigned an MOA, or that can otherwise be assigned quantitative relationships between the structure of the classified chemicals and their acute aquatic toxicity. These four classes are: (1) inert chemicals (baseline toxicity); (2) less inert chemicals, (3) reactive chemicals; and (4) specifically acting chemicals. In the Toolbox, class 5 is introduced to indicate chemicals, which cannot be classified according to the rules of this system.

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. These profilers currently comprise:

Protein binding by OECD
The protein binding by OECD profiler was developed by analysis of direct acting structural alerts based on theoretical organic chemistry (the profiler does not contain metabolically/abiotically activated structural alerts). 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 16 mechanistic alerts covering 52 structural alerts. These data are supported by mechanistic chemistry and references to the scientific literature (the metadata).

Protein-binding by OASIS
The protein-binding alerts have been developed by industry consortia involving ExxonMobil, Procter&Gamble, Unilever, Research Institute for Fragrance Materials (RIFM), Dow and Danish National Food Institute with the Laboratory of Mathematical Chemistry Bourgas and the partnership of Dr D.Roberts, as a part of the TIMES model to predict skin sensitisation. Under the scope of a research agreement signed in 2007 with Professor Mekenyan (OASIS - LMC), L'Oreal contributed to the assessment and refinement of chemical categories provided in the QSAR Toolbox. Categories relating to protein-binding have been the focus of this project.

The Toolbox also includes four databases with experimental data on aquatic toxicity (ECHA Chem also contains experimental data on aquatic toxicity but was not used for these examples):

- Aquatic ECETOC (734 chemicals, 9 487 data, 33 endpoints)
- Aquatic Japan Ministry of Environment (464 chemicals, 2 900 data, four endpoints)
- Aquatic OASIS (2 390 chemicals, 4 826 data, eight endpoints)
- ECOTOX (7 188 chemicals, 491 536 data, 191 endpoints)

2.5 EXAMPLE 3. 2-BUTANAMINE

The third example illustrates the workflow for a quantitative assessment of Daphnia short-term toxicity for 2-butanamine (CAS number 13952-84-6, SMILES: C(C)(N)CC).

2.5.1 Input

The structure can be introduced in the Toolbox using the CAS number. Alternatively, the use of the drawing tool or the SMILES notation returns four chemicals with the same structure and the user can select the chemical structure according to its CAS number. A structure could also be selected from a database, inventory, or user list.

2.5.2 Profiling

The outcomes of the endpoint specific profilers for aquatic toxicity are:

- Aliphatic Amines (Aquatic toxicity classification by ECOSAR)
- Narcotic Amine (Acute aquatic toxicity MOA by OASIS)
- Class 2 (Acute aquatic toxicity classification by Verhaar)

The outcomes of the mechanistic profilers for aquatic toxicity are:

- Protein-binding by OECD: no alert found
- Protein-binding by OASIS: no alert found

According to the endpoint specific profilers, the chemical will exhibit the ecotoxicity typical of amines rather than the baseline toxicity. The structure does not trigger any protein-binding alert. The full profiling results are shown in Figure 15.

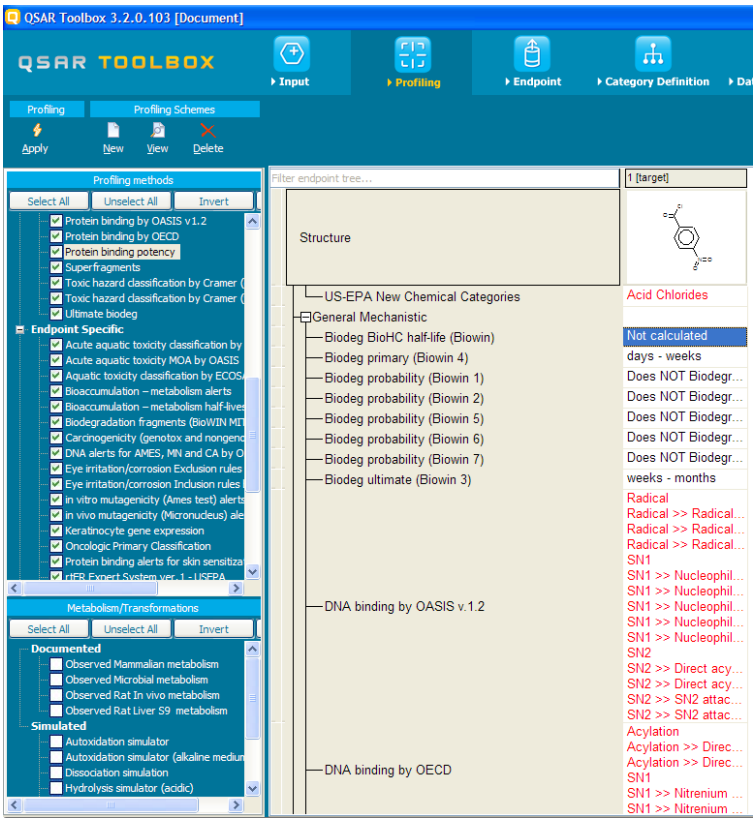


FIGURE 15. PROFILING RESULTS FOR THE TARGET CHEMICAL.

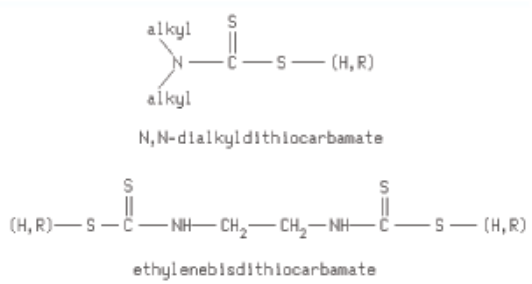
The Toolbox provides further information to explain alerts. The information can easily be accessed by right-clicking the mouse button and selecting 'Explain'. The underlying documentation informs on which part of the chemical triggers the alert and provides links to the original source for further information.

In addition to these endpoint-specific and mechanistic profilers, the following predefined profilers are giving alerts for this chemical:

Alert: OECD HPV Chemical Categories – Primary amines
Under the OECD Cooperative Chemicals Assessment Programme, primary amines including the substance of interest were assessed as the C1-C13 Primary Amines category and concluded at the OECD Assessment Meeting (SIAM) in 2011; http://webnet.oecd.org/Hpv/UI/handler.axd?id=9e86965a-715b-4cb8-99a4-f7113a364ea9).

Alert: US-EPA New chemical categories – Aliphatic Amines
Aliphatic Amines constitute a US-EPA New chemical category and information on environmental toxicity including a proposed testing strategy is provided:

> *Category: Aliphatic Amines Environmental Toxicity*
>
> *This category includes primary amines, secondary amines and tertiary amines; or monoalkyl amines, dialkly amines and trialkyl amines, respectively. This group includes alkanes, alkenes and alkynes; substitutions on carbon (alkyl) chains may include but not be limited to halogens and hydroxyls; insertions in alkyl chains may include but not be limited to ethoxys, propoxys, ethers, sulfides, disulfides and polysulfides; amine oxides are also included in this category; fatty polyamines (e.g. diamines, triamines, tetraamines, pentamines, etc) are also included; amines may either be un-ionised (free) or ionised; and strong ion pairs may also be included.*



Boundaries: There are no lower boundaries and the upper boundary is unknown at this time. It is known that a C13-NH3 Cl is still toxic to fish at less than 1 mg/L. An upper boundary for carbon chain length will probably be about 20 carbons but more information is needed at this time. Generally, members of this category will have molecular weights less than 1 000.

2.5.3 Endpoint

To retrieve any existing data that may be in one of the Toolbox databases, go to Endpoint, select the relevant databases and click "Gather". The retrieval can be further limited to endpoints. For example, in this case the user can click on "choose…" in the pop up window "Read data?" and select only "aquatic toxicity").

Four relevant databases are selected:

- Aquatic ECETOC
- Aquatic Japan MoE Aquatic OASIS
- Aquatic OASIS
- ECOTOX

2.5.4 Category definition

The starting group is formed by use of the chemical profiler "Organic functional groups (nested)" (strict option checked), which determines the molecule as "Aliphatic amine, primary", Figure 16. The sub-categorisation will be performed in the data-gap filling module in order to visualise and better follow the process. There are 35 chemicals gathered in the category (the target and 34 molecules from the databases) and data is extracted/gathered for all category members. None of the data found in the Toolbox for the target chemical refers to 48 hour toxicity to Daphnia.

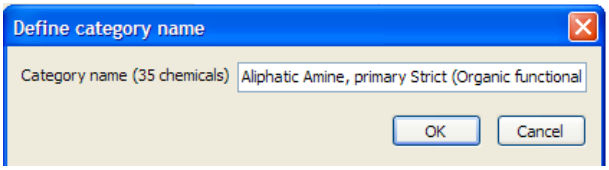


FIGURE 16. POP-UP WINDOW FOR ALIPHATIC AMINE, PRIMARY "STRICT" CATEGORY.

2.5.5 Data-gap filling

In the data-gap filling module the user chooses:

- the endpoint to predict (48h EC50 for Daphnia magna)
- the type of approach for filling the data-gap (trend analysis)
- the sub-categorisation strategy (described below)

At the end of the process, the user can decide if the result is satisfactory. If so, the prediction can be accepted for reporting. Standalone files or IUCLID files can be produced.

The most appropriate endpoint for the prediction of Daphnia short-term toxicity for REACH purposes is the 48h EC50 (Figure 17), in agreement with the OECD TG 202. It should be noted that this guideline refers to the immobilisation of Daphnia. This immobilisation is due to intoxication. This explains why in the Toolbox most of the tests are reported under the branch "intoxication" instead of "immobilisation". It is possible to overcome the terminology inconsistency by rearranging the hierarchy tree. In this example, we used the default tree as shown in Figure 17 (i.e. Effect, Endpoint, duration, Test organisms (species). If a different hierarchy tree is used (e.g. Test organisms/species, Duration, Endpoint, Effect), the data for Immobilisation, Intoxication or Mortality can be combined. The tree hierarchy can be changed by hovering above "Aquatic toxicity" with the mouse and clicking the right mouse button. A pop up window appears and "Set tree hierarchy…" can be selected.

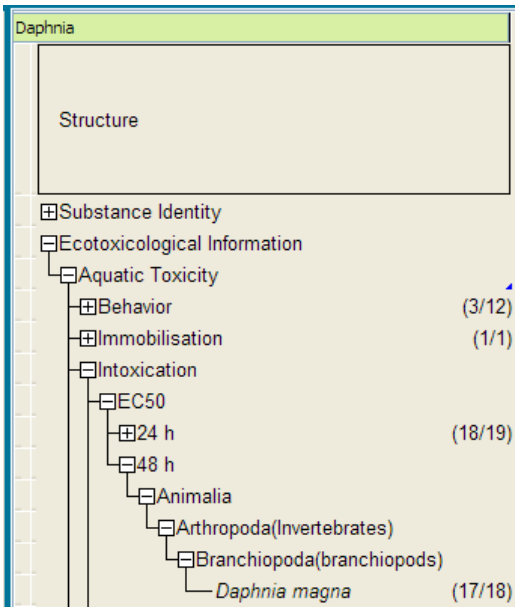


FIGURE 17. SELECTION OF 48H EC50 FOR DAPHNIA MAGNA IN THE "ENDPOINT TREE". 18 DATA POINTS FOR 17 SUBSTANCES ARE AVAILABLE.

The databases contain 18 data points for the category of aliphatic amines. For chemicals with multiple values, by default the Toolbox uses the average value for the calculation (in this case only one chemical presents two values; in example 4, where several chemicals present multiple values, the issue is addressed in details). To predict a quantitative endpoint, the approach of first choice for the data-gap filling is "Trend analysis" because it returns a quantitative outcome based on relationship with a physico-chemical descriptor. Thus, the first plot is generated (Figure 18). It shows the aggregated test data for 48h EC50 as the dependent variable (on Y-axis, expressed as log 1/EC50 in mol/L), and the octanol-water partition coefficient (log K_{ow}) calculated from EpiWIN (KOWIN ver. 1.67) as the dependent variable (on the X-axis). The

red dot indicates the target substance and the blue dots are the selected analogues.

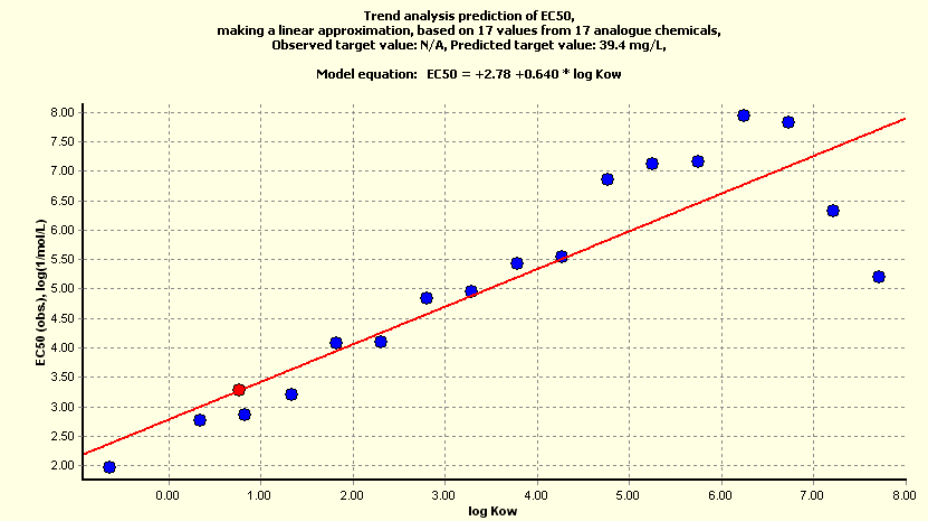


FIGURE 18. AUTOMATICALLY GENERATED PLOT FOR THE INITIAL CATEGORY.

The plot shows strong linear relationship between the EC50 values and the log K_{ow} of up to about six.

For each chemical, it should be checked that the EC50 value does not exceed the water solubility value. Normally, the data points with LC50/EC50 values above the water solubility should be excluded from the analysis. This operation can be performed automatically with the Toolbox in the section "Select/filter data" > "Mark chemical by WS" > Water solubility (fragments). This step excludes two chemicals (the two chemicals with the highest log K_{ow}), leaving a group of 15 analogues as shown in Figure 19.

2.5.6 Category robustness and prediction reliability

The molecules are profiled according to aquatic toxicity profilers. No analogues are removed when selecting "Acute aquatic toxicity MOA by OASIS" and "Acute toxicity classification by ECOSAR" profilers. On the other hand, a sub-categorisation according to the profiler "Acute aquatic toxicity classification by Verhaar" leads to the identification of the analogues whose log K_{ow} is not in the range of 0 to 6, as required for Class 2 category members – less inert compounds. Four analogues were excluded from further analysis because it was considered that the existing knowledge does not cover a wider hydrophobicity range. This exclusion did not affect the prediction value significantly but contributed to increased confidence in making the prediction. The outcome of the aquatic toxicity profilers and the visual inspection of the structures of the analogues indicate that the remaining 12 chemicals form a group robust enough to obtain a reliable prediction. The predicted EC50 value is 93.1 mg/L. The high R2 and Q2 values support the reliability of the prediction.

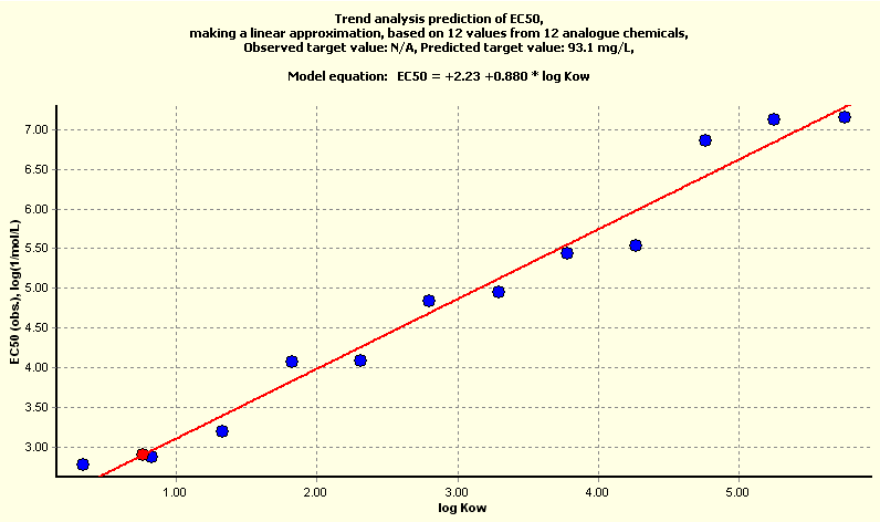


FIGURE 19. SUB-CATEGORY PLOT WITH THE 12 ANALOGUES. IT SHOWS AN EVIDENT TREND BETWEEN THE EC50 VALUE AND LOG K_{OW} OF THE CHEMICALS.

Statistical characteristics and the performance of the identified trend are automatically calculated by the Toolbox and can be seen by clicking on "Statistics" above the plot (Figure 20).

Statistical characteristics	TA model
Number of data points, (N)	12
Coefficient of determination, (R2)	0.973
Adjusted coefficient of determination, (R2adj)	0.970
Coefficient of determination - leave one out, (Q2)	0.960
Coefficient of correlation for external set, (r2)	-
Sum of squared residuals, (SSR)	0.750
Standard deviation of residuals, (sN)	-
Sample standard deviation of residuals, (s)	0.274
Fisher function, (F)	356
Fisher threshold for statistical significance, (Fa)	7.43
b0	
- model descriptor	Intercept
- coeff. value	2.23
- coeff. range	± 0.36
- significance	Yes
- max. covariation	0.297 (vs b1)
b1	
- model descriptor	log Kow
- coeff. value	0.880
- coeff. range	± 0.104
- significance	Yes
- max. covariation	0.297 (vs b0)

FIGURE 20. STATISTICS OF THE MODEL DERIVED WITH THE 12 ANALOGUES.

A visual inspection of the chemical structures of the analogues (Figure 21), shows that all members of the series of homologues are primary amines whose structures are very similar to that of the target, with the exception of the chain length. In this case, chain length is expected to modulate the absorption and not the mechanism of action, within the selected range of log K_{ow}.

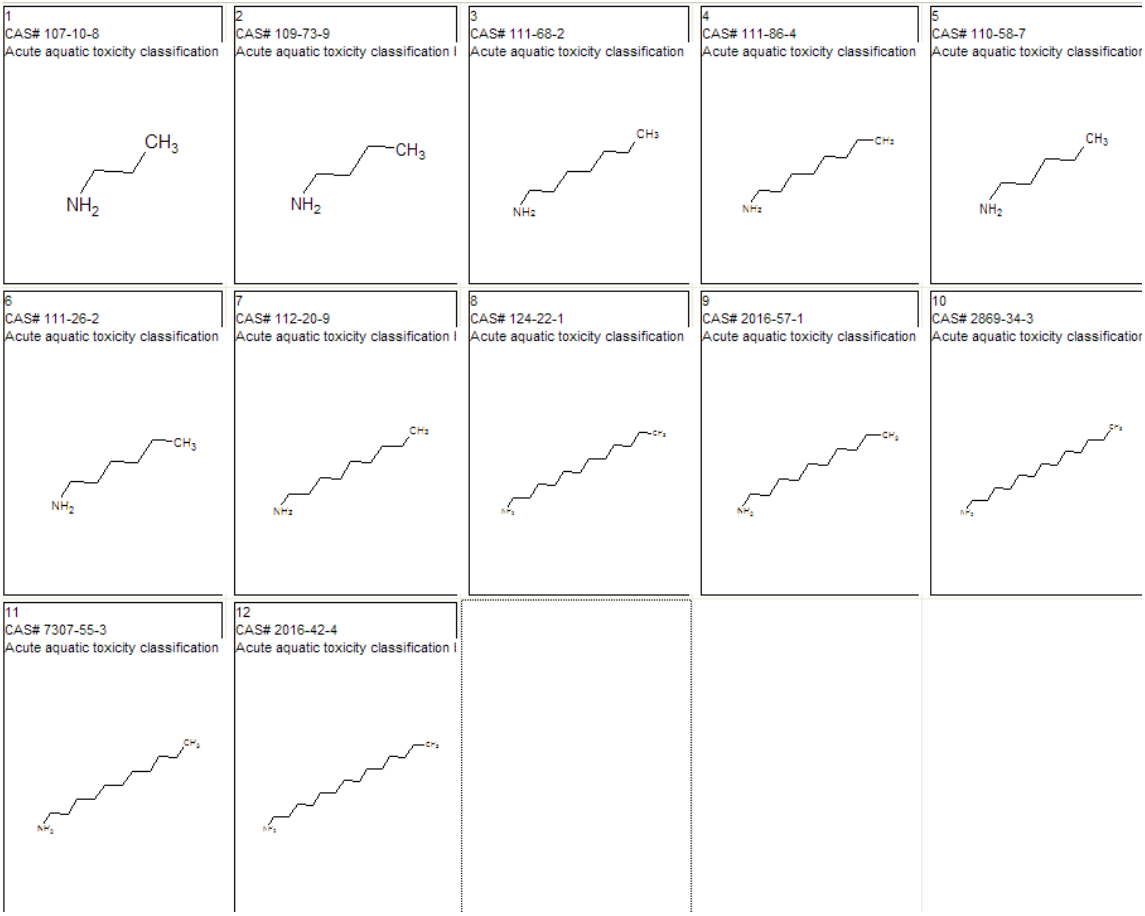


FIGURE 21. CHEMICAL STRUCTURE AND CAS NUMBER OF THE SOURCE SUBSTANCES.

Data quality also needs to be addressed before accepting the predicted EC50 value (93.1 mg/L). The Toolbox does not contain reliability scores for experimental toxicity data. However, the user can consult and extract metadata regarding the experimental conditions for each data point and can check the references. For illustrative purposes, in Table 3 we report the reference source for data points shown in Figure 19 (although more information can be extracted from the software).

CAS	Name	Reference	Exp. Value(s) (mg/L)
107-10-8	Propylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	> 100
109-73-9	Butylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	> 100
110-58-7	Amylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	56 (27;170)
111-68-2	Heptylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	9.4 (6.9;12)
111-26-2	Hexylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	8.6 (7;11)
111-86-4	Octylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	1.9 (1.5;2.4)

CAS	Name	Reference	Exp. Value(s) (mg/L)
112-20-9	Nonylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	1.6 (1.2;2)
2016-57-1	Decylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	0.58 (0.39;0.79)
7307-55-3	Undecylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	0.5 (0.4;0.61)
124-22-1	Dodecylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	0.026 (0.018;0.036)
2869-34-3	Tridecylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	0.015 (0.011;0.019)
2016-42-4	Tetradecylamine	Proj.No.303587, Report to Danish EPA, Copenhagen, Denmark, 1998	0.015 (0.011;0.019)

TABLE 3. REFERENCE TABLE

It is also possible to use experimental log K_{ow} values instead of calculated, when available. The experimental log K_{ow} values should be checked for reliability too. To use experimental hydrophobicity, during the data-gap filling step "Exp log P" has to be selected as the X descriptor (in picklist below the plot) and has to be made the "active descriptor" using the "Descriptors/data" menu. Experimental values are taken from EPISUITE/KOWWIN database. In this example, the experimental values are available for seven analogues (Figure 22). The predicted LC50 for the 2-butanamine is 119 mg/L, very close to that obtained with the predicted log K_{ow} (93.1 mg/L). Also with the experimental log K_{ow}, the correlation with EC50 values is satisfactory. Table 4 lists the experimental and predicted log K_{ow} values. The user has the possibility to also extract other physico-chemical properties and find additional trends that might support the prediction or the robustness of the category.

CAS	NAME	Experimental log K_{ow}	Calculated log K_{ow}
107-10-8	Propylamine	0.48	0.34
109-73-9	Butylamine	0.97	0.84
110-58-7	Amylamine	1.49	1.33
111-68-2	Heptylamine	2.57	2.31
111-26-2	Hexylamine	2.06	1.82
111-86-4	Octylamine	2.9	2.8
112-20-9	Nonylamine	NA	3.29
2016-57-1	Decylamine	NA	3.78
7307-55-3	Undecylamine	NA	4.27
124-22-1	Dodecylamine	NA	4.76
2869-34-3	Tridecylamine	NA	5.25
2016-42-4	Tetradecylamine	NA	5.75

TABLE 4. EXP LOG K_{OW} VALUE FOR SOME OF THE GROUP MEMBERS

The user can decide to accept and report the prediction obtained with experimental or calculated log K_{ow} values. In this example, the predicted values with calculated or experimental log K_{ow} values are in the same order of magnitude, although, the EC50 obtained using the calculated log K_{ow} (93.1 mg/L) indicates higher toxicity than that obtained using experimental log K_{ow} (119 mg/L). The first prediction also relies on more data points (12 against 6). In order to follow the precautionary principle, the EC50 of 93.1 mg/L is accepted. Please note that here the term "accept" describes a Toolbox button indicating that the prediction is ready for reporting, and is not related to acceptance in terms of regulatory application.

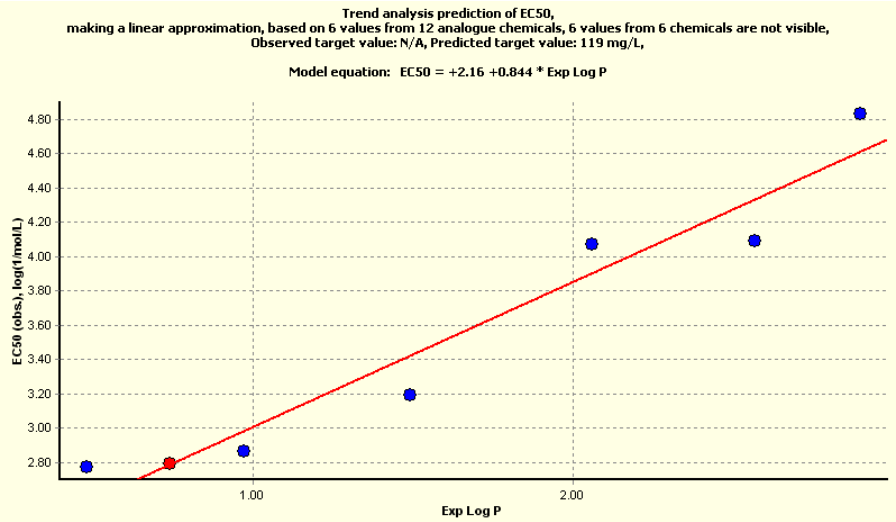


FIGURE 22. EC50 VS EXP LOG K_{OW} PLOT

2.5.7 Report and category summary

Once the prediction is accepted for reporting, the report module of the Toolbox gives the possibility to semi-automatically generate a standalone file to record the result with lots of pre-defined information, coming from the modelling procedure (applicability domain, information about the group members, etc.). The report also contains manually editable fields that the user should fill to justify the procedure that was followed. Here, we summarise the considerations that need to be given for the assessment of the category and the prediction. They are based on good practices and expert judgement and are not hard-coded in the Toolbox as such.

a. Category definition
   A category of 13 analogues (a target and 12 source chemicals) is defined for the purpose of data-gap filling of Daphnia short-term toxicity (as 48h EC50). The analogues are shown in Table 4. In the example, considerations on impurities are omitted. For REACH substances, impurities need to be analysed and discussed.

b. Hypothesis for grouping chemicals
   The working hypothesis is that primary aliphatic amines with a C3-C14 aliphatic chain can be grouped together for the prediction of *Daphnia* short-term toxicity. The hypothesis is based on the observation (reported by the US Environmental Protection Agency) that aliphatic amines can be highly toxic to all groups of freshwater organisms (i.e. fish, aquatic invertebrates and green algae). Observed corrosive properties seem to overwhelm the systemic toxicity of the primary amines in most cases, including acute toxicity to mammals and aquatic species. This means that the compounds can destroy

the proteins and the lipids in the tissues of the living organisms and thus irreversibly disrupt their function.

Toxicity is related to the length of the hydrophobic carbon chains: the longer (or greater the number of carbons) the chain, the more toxic to aquatic organisms when the number of amines is constant (up to the limit of solubility). In general, the hydrophobicity cut-offs can be dictated either by the absorption/availability being too small or the water solubility being too low. All primary non-branched amines are likely to be metabolised in the same way to products, which are not expected to have an excess toxicity (do not trigger an alert for protein-binding in the Toolbox). The only metabolic exception in this series is expected to be the methylamine, which has been removed from the group. This is noted also in a recent OECD report on assessment of primary amines (OECD, 2011). In the category herein, a clear trend between log K_{ow} and Daphnia 48h EC50 is observed.

c. Category description
   - The category members are primary aliphatic non-branched amines with a single amine group and carbon chain length ranging from three to 14. The category members do not present any additional functional groups. For each analogue a 48h EC50 value for Daphnia short-term toxicity is available. The EC50 values are expressed in mg/L. For modelling, these were converted on a molar basis, and converted again in mg/L for reporting.

   - The covered hydrophobicity range in terms of log K_{ow} is between 0.34 and 5.75 (Toolbox calculated value, the experimental value is not available for all the analogues). The log K_{ow} of the target chemical is 0.74 (experimental) or 0.76 (calculated). The two values are similar to each other and both fit in the range defined by the analogues. Solubility covered by the analogues is from 1.74 to $3.9*10^5$ mg/L (Toolbox calculated value from "Water solubility (fragments)" descriptor, the experimental value is not available for all the analogues). The solubility for the target is $1.12*10^5$ (experimental) or $2.35*10^5$ (calculated). The two values are similar and both fit in the range defined by the analogues. The molecular weight of the target is 73.1 Da, in the range defined by the analogues (from 59.1 to 269 Da). The pKa of all the group members has similar values (between 10 and 11).

d. Strategy used (data-gap filling method)
   Quantitative trend analysis. The regression equation is:

   EC50 = 2.23(±0.36) + 0.883(±0.104) * log K_{ow} (as shown in Figure 19).
   The EC50 is presented in log (1/EC50 mol/L).
   Number of data points (N) = 12
   Statistical characteristics of the model (exhaustive list in Figure 20):
   Coefficient of determination (R2) = 0.973
   Coefficient of determination, leave-one-out (Q2) = 0.960
   The sample standard deviation of residuals (s) = 0.274
   Mechanistic explanation and applicability domain: provided at points b and c, respectively.

2.6 EXAMPLE 4. PROPIOPHENONE

The fourth example illustrates the workflow for a quantitative assessment of fish short-term toxicity for propiophenone (CAS number 93-55-0, SMILES: C(=O)(c1ccccc1)CC).

2.6.1 Input

The structure can be introduced in the Toolbox using the drawing tool, the CAS number or the SMILES notation. A structure could also be selected by a database, inventory, or user list.

2.6.2 Profiling

The outcomes of the endpoint specific profilers for aquatic toxicity are:

- Class 5 (Not possible to classify) (Acute aquatic toxicity classification by Verhaar)
- Basesurface narcotics (Acute aquatic toxicity MOA by OASIS)
- Neutral organics (Aquatic toxicity classification by ECOSAR)

Figure 23 shows the results of the profiling for the target chemical (black font indicates no alert; while red font highlights the presence of alerts; yellow indicates parameters which have not been calculated in the Toolbox).

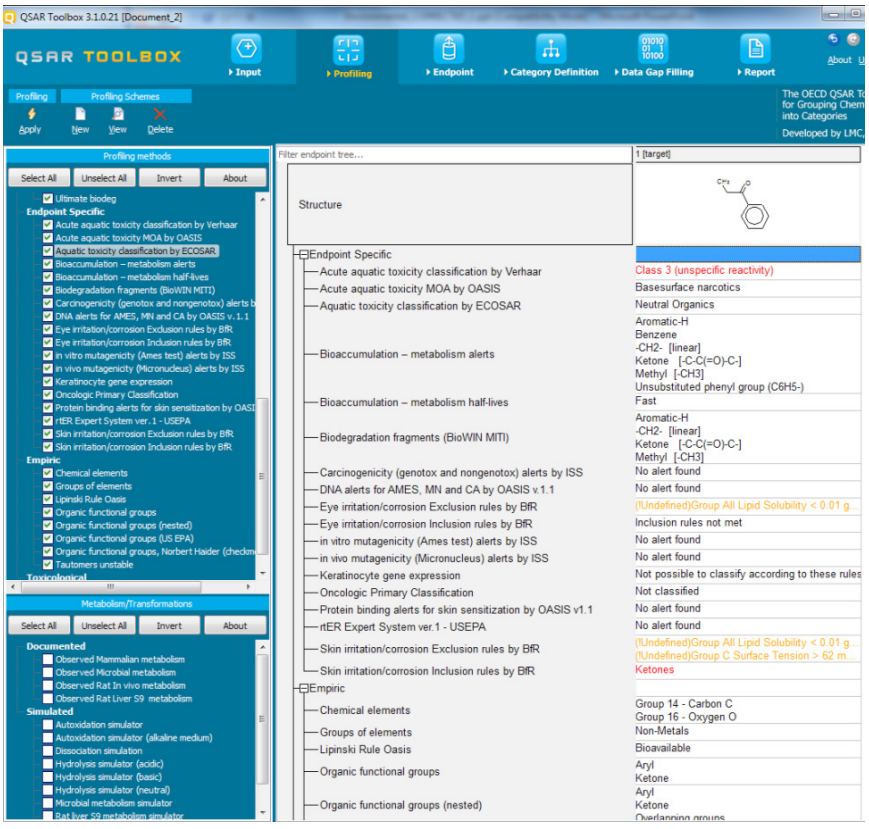


FIGURE 23. PROFILING RESULTS FOR THE TARGET CHEMICAL.

2.6.3 Endpoint

These four relevant databases are selected:

- Aquatic ECETOC
- Aquatic Japan MoE
- Aquatic OASIS
- ECOTOX

2.6.4 Category definition

The starting category is based on a chemical profiler, the US EPA new chemical categories, which classifies the molecule as "Neutral Organics", Figure 24. The sub-categorisation will be performed in the data-gap filling module in order to visualise and better follow the process. There are 937 chemicals gathered in the category (936 molecules from the databases and the target) and data are extracted/gathered for all category members.

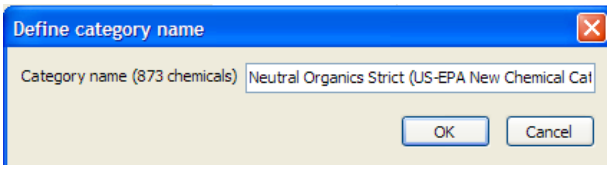


FIGURE 24. POP-UP WINDOW FOR NEUTRAL ORGANICS CATEGORY.

2.6.5 Data-gap filling

In the data-gap filling module the user chooses:

- the endpoint to predict (LC50 at 96h to *Pimephales promelas*)
- the type of approach for filling the data-gap (trend analysis)
- the sub-categorisation strategy (explained below)

At the end of the process, the user can decide if the result is satisfactory. If so, the prediction can be accepted for reporting and exported as a standalone report and to IUCLID.

Fish short-term toxicity laboratory tests have a different duration and involve different species. Several validated procedures (e.g. OECD TG 203) usually require LC50 at 96 hours. In the example, we use *Pimephales promelas* experimental data since it is one of the recommended fish species in the guidelines and in this example presents the highest number of data points (500 measures for 176 chemicals, as shown in Figure 25). It is also possible to use many different fish species in the same analysis, but the inter-species variability can lead to worse correlation and misleading results. For this reason, it is suggested to identify a trend for a single species (if enough data points are available) and then investigate if a more sensitive species exists using advanced functions available in the Toolbox, described later in this example. Here, after predicting the LC50 for *Pimephales promelas*, the toxicity to other species is also analysed.

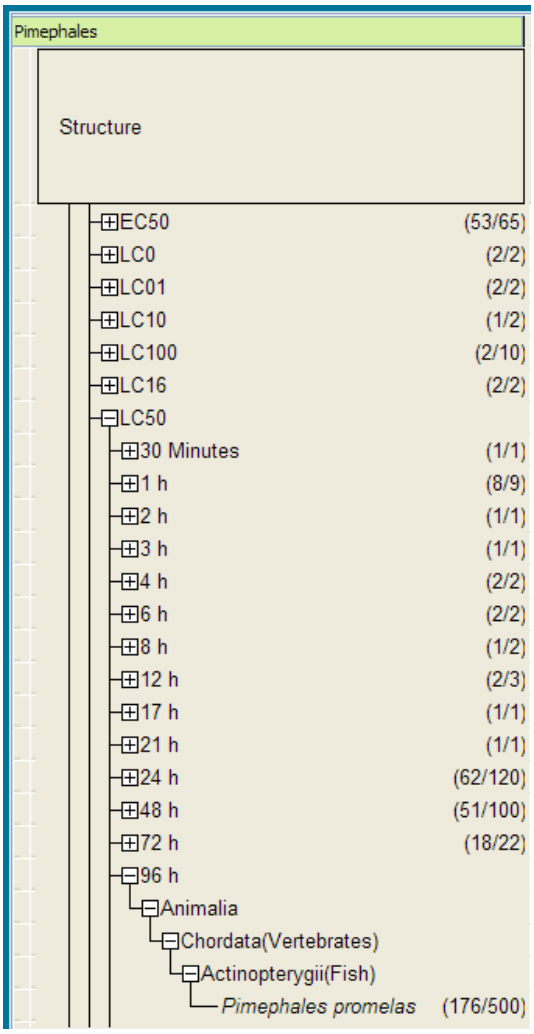


FIGURE 25. SELECTION OF 96H LC50 FOR PIMEPHALES PROMELAS IN THE "ENDPOINT TREE". 500 DATA POINTS FOR 176 SUBSTANCES ARE AVAILABLE. TIP: USING THE FILTER FUNCTIONALITY ON TOP HELPS TO FIND THE SPECIES.

The approach for filling the data-gap for the quantitative endpoint is "trend analysis". The plot for the preliminary category of Neutral organics (Figure 26) shows some correlation between LC50 (Y-axis, expressed as log 1/LC50 in mol/L) and log K_{ow} (X-axis) but with many outliers.

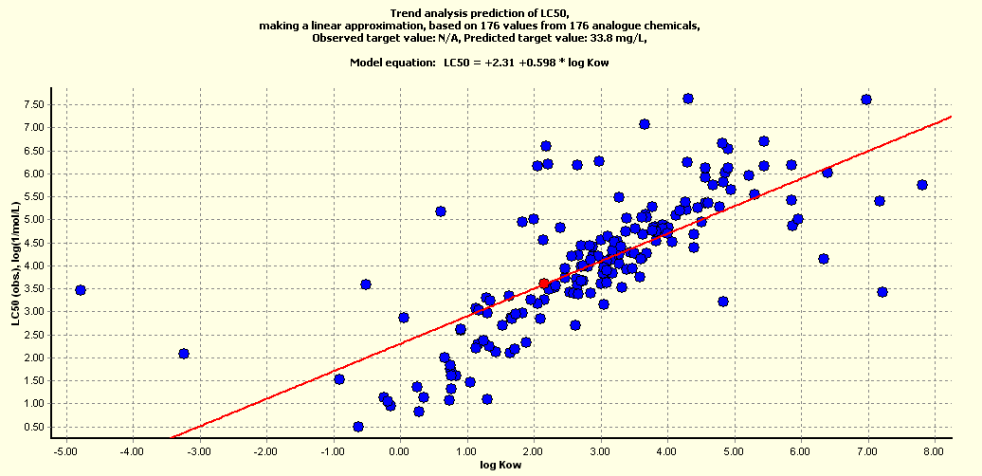


FIGURE 26. AUTOMATICALLY GENERATED PLOT FOR THE STARTING CATEGORY.

The category can now be narrowed down according to the "organic functional group (nested)" profiler, which identifies "aryl" and "ketone" functionalities in the target molecule. The molecules with these functionalities are kept and all the others are removed from the analysis. The plot of observed acute toxicity to fish and log K_{ow} for 22 chemicals is shown in Figure 27.

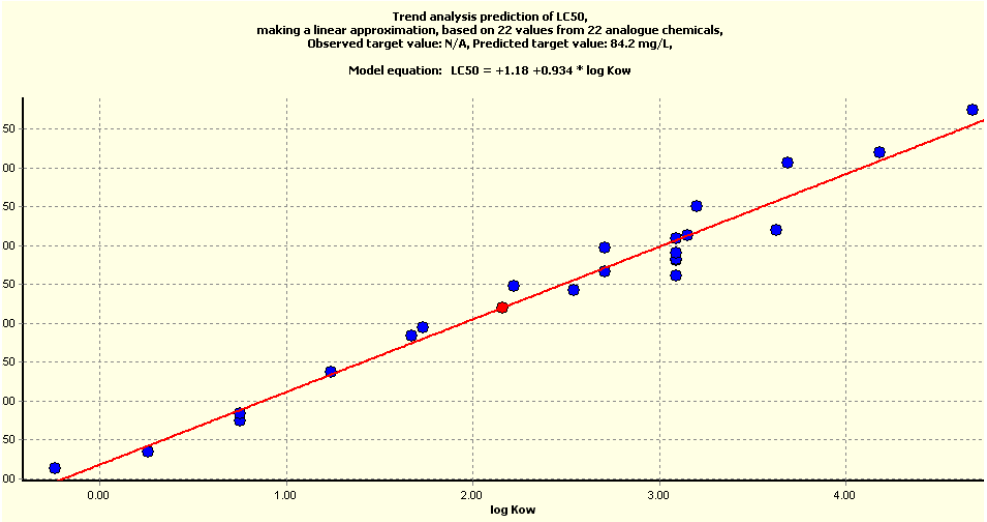


FIGURE 27. SUB-CATEGORY PLOT WITH THE 22 ANALOGUES. IT SHOWS A CLEAR TREND BETWEEN THE LC50 VALUE AND LOG K_{OW} OF THE CHEMICALS.

The trend is now statistically satisfactory and does not show any outliers.

2.6.6 Category robustness and prediction reliability

The statistics (shown in Figure 28) related to the identified trend are automatically calculated by the Toolbox, together with plots that reflect the statistical performance of the trend.

Statistical characteristics	TA model
Number of data points, (N)	22
Coefficient of determination, (R2)	0.966
Adjusted coefficient of determination, (R2adj)	0.964
Coefficient of determination - leave one out, (Q2)	0.959
Coefficient of correlation for external set, (r2)	-
Sum of squared residuals, (SSR)	1.07
Standard deviation of residuals, (sN)	-
Sample standard deviation of residuals, (s)	0.231
Fisher function, (F)	565
Fisher threshold for statistical significance, (Fa)	6.27
b0	
- model descriptor	Intercept
- coeff. value	1.18
- coeff. range	± 0.23
- significance	Yes
- max. covariation	0.267 (vs b1)
b1	
- model descriptor	log Kow
- coeff. value	0.934
- coeff. range	± 0.082
- significance	Yes
- max. covariation	0.267 (vs b0)

FIGURE 28. STATISTICS OF THE MODEL DERIVED WITH THE 22 ANALOGUES.

A visual inspection of the chemical structures of the analogues shows relative diversity in chemical structures selected to be the source substances for the selected target (Figure 29).

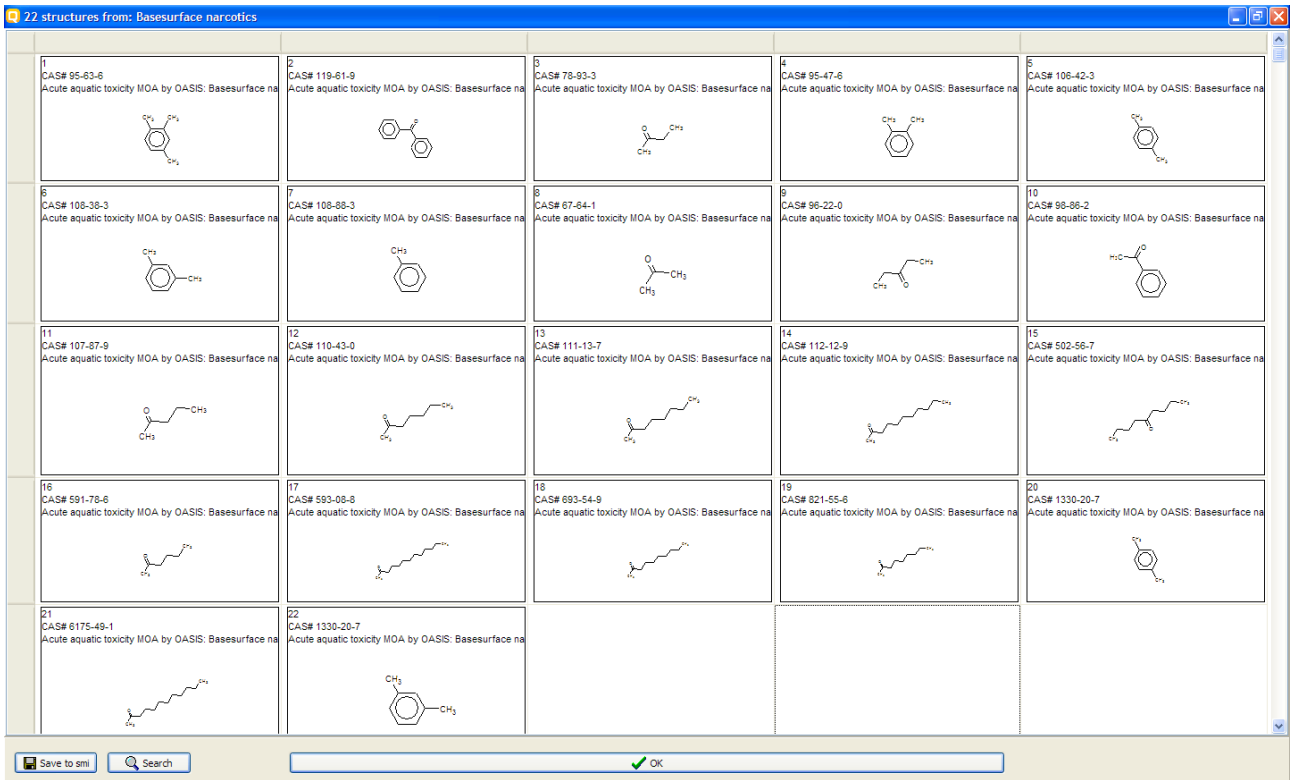


FIGURE 29. CHEMICAL STRUCTURE AND CAS NUMBER OF SOURCE SUBSTANCES.

For each chemical, its LC50 value has to be checked so that it does not exceed its water solubility value. This operation can be performed automatically with the Toolbox in the section "Select/filter data" > "Mark chemical by WS". No chemicals are excluded in this example.

The next step is the sub-categorisation according to endpoint specific profilers. The category is robust for the Acute aquatic toxicity MOA by OASIS (all substances are defined as basesurface narcotics) and Acute toxicity classification by ECOSAR (all substances are defined as neutral organics). Acute aquatic toxicity classification by Verhaar identifies the target chemical, benzophenone (CAS 119-61-9), acetone (CAS 67-64-1), and acetophenone (CAS 98-86-2) as Class 5 chemicals (not possible to classify). Acetophenone and benzophenone could be seen as alpha-beta unsaturated compounds, which are not expected to be activated by a polarisable substituent (carbonyl, nitrile, amide, nitro, sulphone, etc.) in the aromatic ring (the same as the target substance). Acetone is found to be different because of its log K_{ow} being lower than (but close to) 0. Moreover, acetone, acetophenone and benzophenone lie perfectly on the trend line defined by the analogues, so it is likely that they act with the same mechanism as the other selected compounds. Sub-categorisation according to the Verhaar profiler is not performed in the example. Another consideration is that some structures are reasonably similar to propiophenone, in particular, acetophenone (CAS# 98-86-2) is closely related to the target chemical. In one-to-one read-across, the prediction would be estimated as 178 mg/L, while the predicted LC50 from the trend is 84.2 mg/L. Thus, the trend prediction is both more robust and more conservative than the prediction from the read-across alone. Eventually, such comparison could be added in the justification for registration submission to ECHA.

The user can decide to accept the prediction obtained for reporting with experimental or calculated log K_{ow} values. The experimental log K_{ow} values should be checked for reliability too. In this example, the predicted log K_{ow} value for the target is quite similar to the experimental one, as shown in Table 5. The LC50 obtained using the calculated log K_{ow} is 84.2 mg/L. The LC50 obtained using experimental log K_{ow} is 110.0 mg/L. The two calculated LC50 values are obviously close and we prefer to use the lower one further. The first prediction using calculated log K_{ow} relies on 22 data points, while the prediction using the experimental log K_{ow} relies on 20 data points (figure 30). In order to follow the precautionary principle, the LC50 of 83.3 mg/L is accepted for reporting. Please note that here the term "accept" describes a Toolbox button indicating that the prediction is ready for reporting, and is not related to acceptance in terms of regulatory application.

CAS	NAME	Experimental log K_{ow}	Calculated log K_{ow}
93-55-0	Propiophenone	2.19	2.1647
67-64-1	Acetone	-0.24	-0.235
78-93-3	2-Butanone	0.29	0.2561
96-22-0	3-Pentanone	0.99	0.7472
107-87-9	2-Pentanone	0.91	0.7472
591-78-6	2-Hexanone	1.38	1.2383
98-86-2	Acetophenone	1.58	1.6736
110-43-0	2-Heptanone	1.98	1.7294
111-13-7	2-Octanone	2.37	2.2205

CAS	NAME	Experimental log K_{ow}	Calculated log K_{ow}
108-88-3	Toluene	2.73	2.5403
502-56-7	5-Nonanone	2.88	2.7116
1330-20-7	Xylenes (mixed)	3.2	3.0876
95-47-6	o-Xylene	3.12	3.0876
108-38-3	m-Xylene	3.2	3.0876
821-55-6	2-Nonanone	3.14	2.7116
119-61-9	Diphenyl methanone	3.18	3.1471
1330-20-7	Xylenes	3.15	3.0876
106-42-3	p-Xylene	3.15	3.0876
95-63-6	1,2,4-Trimethyl-benzene	3.63	3.6349
693-54-9	2-Decanone	3.73	3.2027
112-12-9	2-Undecanone	4.09	3.6938
6175-49-1	2-Dodecanone	NA	4.1849
593-08-8	2-Tridecanone	NA	4.676

TABLE 5. EXPERIMENTAL AND CALCULATED LOG K_{OW} VALUES FOR THE GROUP MEMBERS.

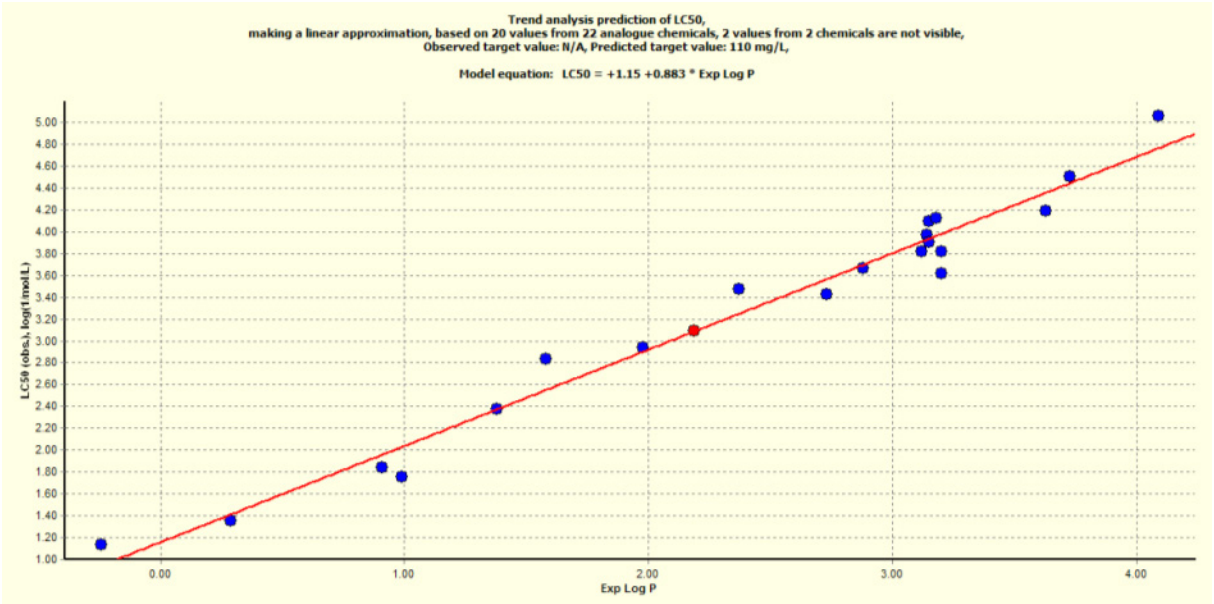


FIGURE 30. LC50 VS EXP LOG K_{OW} PLOT

Data quality also needs to be addressed. The Toolbox does not contain reliability scores for experimental toxicity data nor for log K_{ow}. However, the user can consult and extract metadata regarding the experimental conditions for each data point and can check the references (not shown for this example).

Inter-species variability considerations

The prediction is derived from experimental data for *Pimephales promelas*. However, the interspecies variability for this set of chemicals and the existence of a more sensitive species can be addressed. This possibility exists because the Toolbox can support different results for the same substance. Going back to the data-gap filling, it is possible to select the endpoint without defining the species to be considered but by gathering all the available 96h LC50 values. Then, setting the "data usage" as "all" under the "calculation options", it is possible to obtain the plot shown in Figure 31. A point-to-point analysis (not shown here) demonstrates that the highest values related to each chemical come from different species and that a species-specific trend does not exist.

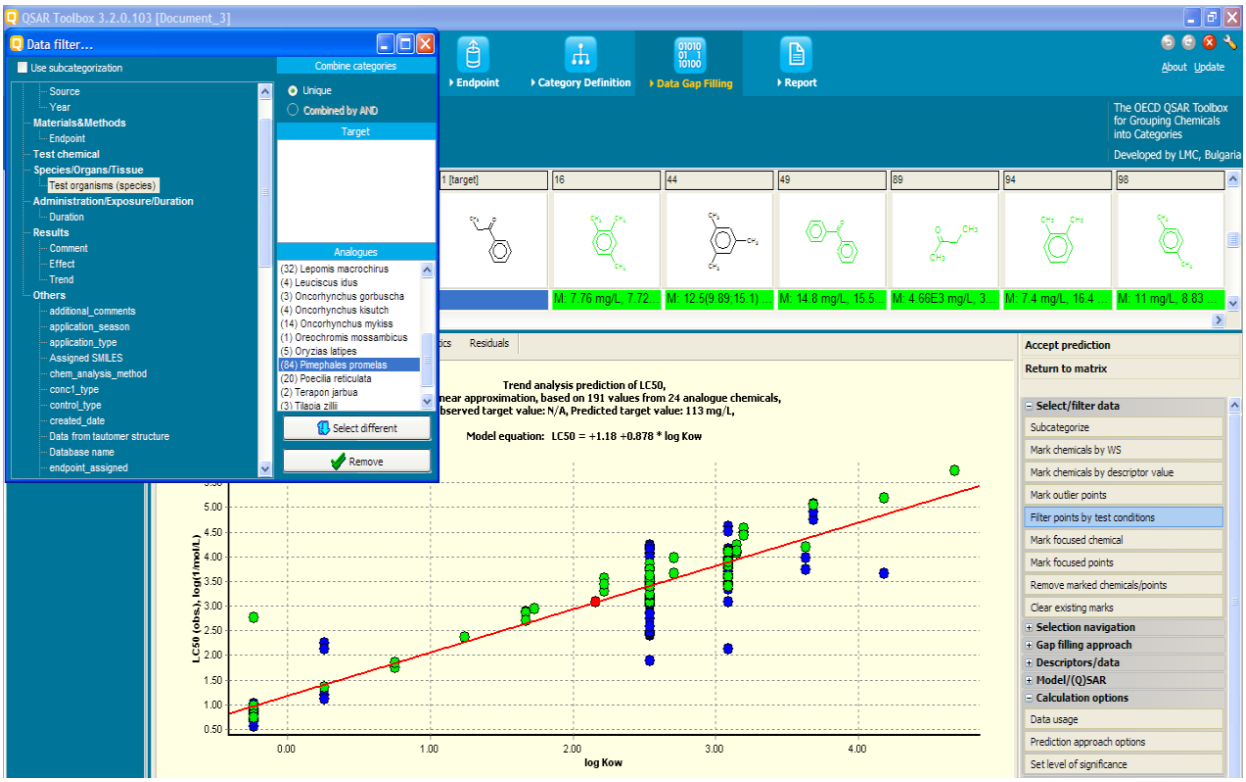


FIGURE 31. PLOT WITH DATA POINTS RELATED TO EXPERIMENT PERFORMED WITH DIFFERENT FISH SPECIES.

For some data rich substances, the acute aquatic toxicity to fish could vary more than two log units in mol/L scale and thus a prediction is not recommendable. There are many different possibilities of selecting the data (guideline species, most ecologically relevant species, most sensitive species, all data pooled, etc). The important issue is to clearly report and justify why a certain approach was taken.

2.6.7 Report and category summary

Once the prediction is accepted, the report module of the Toolbox gives the possibility to semi-automatically generate a standalone file to record the result with lots of pre-defined information, coming from the modelling procedure (applicability domain, information about the group members, etc.). The

report also contains manually editable fields that the user should fill to justify the procedure that was followed. Here, we summarise the considerations that need to be given for the assessment of the category and the prediction. They are based on good practices and expert judgement and are not hard-coded in the Toolbox as such.

a. Category definition
A category of 23 analogues (a target and 22 analogues) is defined for the purpose of data-gap filling for fish short-term toxicity (as 96h LC50 to Pimephales promelas). The chemical structure of the analogues is shown in Figure 29. In the example, considerations on impurities are omitted. For REACH substances, impurities need to be analysed and discussed.

b. Hypothesis for grouping chemicals
The working hypothesis is that chemicals with aryl or ketone functionalities can be grouped together for the prediction of fish short-term toxicity. The hypothesis is based on the assumption that the analogues share the same mode of action for aquatic toxicity – the non-polar narcotic effect. This assumption was supported by the robust trend found between LC50 and log K_{ow} for all analogues, selected neutral organics with aryl or ketone functionality only. On a comparison between the equation obtained in that example and different narcosis models (JRC EUR 21749, 2005), the slope and intercept of the obtained model are most similar to the one of the non-polar narcosis. The non-polar narcosis model for acute fish toxicity is a robust relationship that has proven many times over the years, whether a substance is neutral organics. A small variation in the coefficients does not matter so much as the belonging of a chemical to that group (the applicability of this concept to the chemical). In fact, the prediction by the JRC non-polar narcosis model LC50 is 81.3 mg/L, which is very close to our prediction (83.3 mg/L). However, it is the strength of the relationship, and the slope and the intercept of the line that matter because a single value could be approximated by many different trends. The narcosis effect is based on the assumption that the substances enter the body of the fish by absorption through passive diffusion. This mechanism is non-structure dependent and is driven by hydrophobicity. Thus, it is possible to observe a trend between LC50 and the log K_{ow} of the narcotic chemicals, regardless of structural similarity. The simulated metabolites of the analogues (including observed metabolites for some of them) are different. However, the metabolites do not trigger alerts for protein-binding and are not expected to provoke higher toxicity than the parents. Thus, the metabolism does not raise reason for concern within this group and endpoint (concern in addition to the one driven by the hydrophobicity). Potential metabolism should always be considered as a possible reason for trend-breaking, especially if structural and mechanistic diversity could be found.

c. Category description
The category members are chemicals exclusively containing aryl and/or ketone functional groups (non-activated carbonyl groups). The category does not contain analogues with ionisable and hydrolysable groups. For each analogue, a 96h LC50 value for Pimephales promelas in mg/L is available in the Toolbox. For modelling, these were converted on a molar basis, and converted again in mg/L for reporting.

The covered hydrophobicity range in term of log K_{ow} is between -0.23 and 4.68 (Toolbox calculated value, the experimental value is not available for all the analogues). The log K_{ow} of the target chemical is 2.19 (experimental) or 2.16 (calculated). The two values are similar and both fit well in the range defined by the analogues. Experimental solubility covered by the analogues is from 57 to $1*10^6$ mg/L. The experimental solubility for the target is $2*10^3$ mg/L, which is well in the range defined by the analogues. The molecular weight of the target is 134 Da, inside the range covered by the analogues (from 58.1 to 198 Da).

d. Strategy used (data-gap filling method)
Quantitative trend analysis. The regression equation is:

LC50 = 1.18 (±0.23) + 0.934 (±0.082) * log K_{ow} (as shown in Figure 27)
Statistical characteristics of the model (exhaustive list in Figure 28)
N = 22
R2 = 0.966
Q2 = 0.959
s = 0.231
Mechanistic explanation and applicability domain: provided at points b and c, respectively.

EUROPEAN CHEMICALS AGENCY
ANNANKATU 18, P.O. BOX 400,
FI-00121 HELSINKI, FINLAND
ECHA.EUROPA.EU

ECHA-14-R-07-EN - ED-02-14-124-EN-N - ISBN: 978-92-9244-433-4 - DOI: 10.2823/83105