QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.1

Tutorial on how to predict skin sensitisation potential by standardized workflow

Outlook

- **Background**
- Objectives
- Specific Aims
- Standardized workflow for Skin sensitization
- The exercise
- Standardized workflow execution

Background

- This is a step-by-step presentation designed to take the Toolbox user through the filling of skin sensitization data gaps using a standardized workflow.

Outlook

- Background
- **Objectives**
- Specific Aims
- Standardized workflow for Skin sensitization
- The exercise
- Standardized workflow execution

Objectives

This presentation demonstrates a number of functionalities of the Toolbox:

- Identify analogues for a target chemical;
- Retrieve experimental results available for those analogues;
- Color the profiling schemes according to their suitability for subcategorization;
- Fill data gaps by standardized workflow;

Outlook

- Background
- Objectives
- **Specific Aims**
- Standardized workflow for Skin sensitization
- The exercise
- Standard workflow execution

Specific Aims

- To introduce to the Toolbox user the standard workflow for predicting of skin sensitization potential;
- To familiarize the user with the new Toolbox interface;
- To familiarize the user with the new notification messages;
- To explain to the user the rationale behind each step of the exercise.

Outlook

- Background
- Objectives
- Specific Aims
- **Standardized workflow for Skin sensitization**
- The exercise
- Standardized workflow execution

Standardized workflow for Skin sensitization

Overview

- The standardized workflow (SW) is designed to apply data gap filling for discrete chemicals only
- The SWs has been developed to be applicable for the same endpoints used for application of the AWs (i.e. Skin sensitization, in vivo, LLNA and GPMT).
- Once started, the SW follows the implemented logic under the user control.
- As opposite to the automated workflow (AW), the domain of application is expanded in the SWs (including other species, durations, etc.) and SWs allow interactions by the user.
- In case more than one further application is possible, the workflow stops and waits for the decision of the user.
- SW can be executed for one chemical as well as for a batch of chemicals.

Outlook

- Background
- Objectives
- Specific Aims
- Standardized workflow for Skin sensitization
- **The exercise**
- Standardized workflow execution

The Exercise

- In this exercise we will predict the skin sensitization potential of Ethylparaben [CAS# 120-47-8], which will be the "target" chemical.
- This prediction will be accomplished by using of the developed standardized workflow for skin sensitization.

The Exercise
Sidebar On Sensitization

- Allergic contact dermatitis that results from skin sensitization is a significant health concern.
- Skin sensitization is a toxicological endpoint that is complex and conceptually difficult.
- However, there is growing agreement that most organic chemicals must react covalently with skin proteins in order to behave as skin sensitizers.
- Therefore, mechanisms by which organic chemicals bind with proteins are relevant to grouping chemicals that may be skin sensitizing agents.

Outlook

- Background
- Objectives
- Specific Aims
- Standardized workflow for Skin sensitization
- The exercise
- **Standardized workflow execution**

Standardized workflow execution

- Only three of the general Toolbox modules are used in a sequential workflow:
  - Input
  - Data Gap Filling
  - Report

The rest of the modules – *Profiling*, *Data* and *Category definition* are included as a part of the algorithm of the standardized workflow. The workflow stops at them and waits for the decision of the user.

Outlook

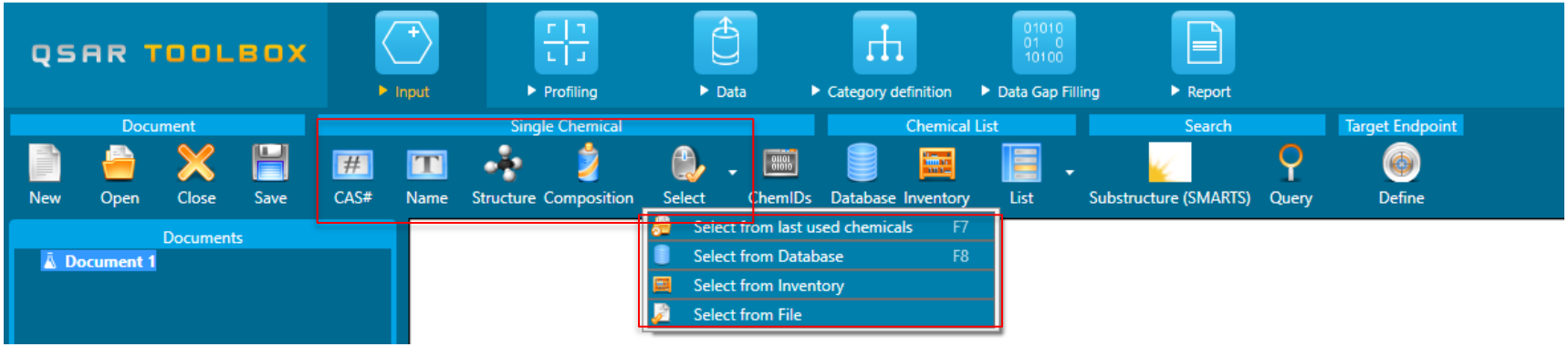
- Background
- Objectives
- Specific Aims
- Standardized workflow for Skin sensitization
- The exercise
- **Standardized workflow execution**
  - Input

Input
Overview

- This module provides the user with several means of entering the chemical of interest (i.e. the target chemical).
- Since all subsequent functions are based on chemical structure, the goal here is to make sure the molecular structure assigned to the target chemical is the correct one.

Input

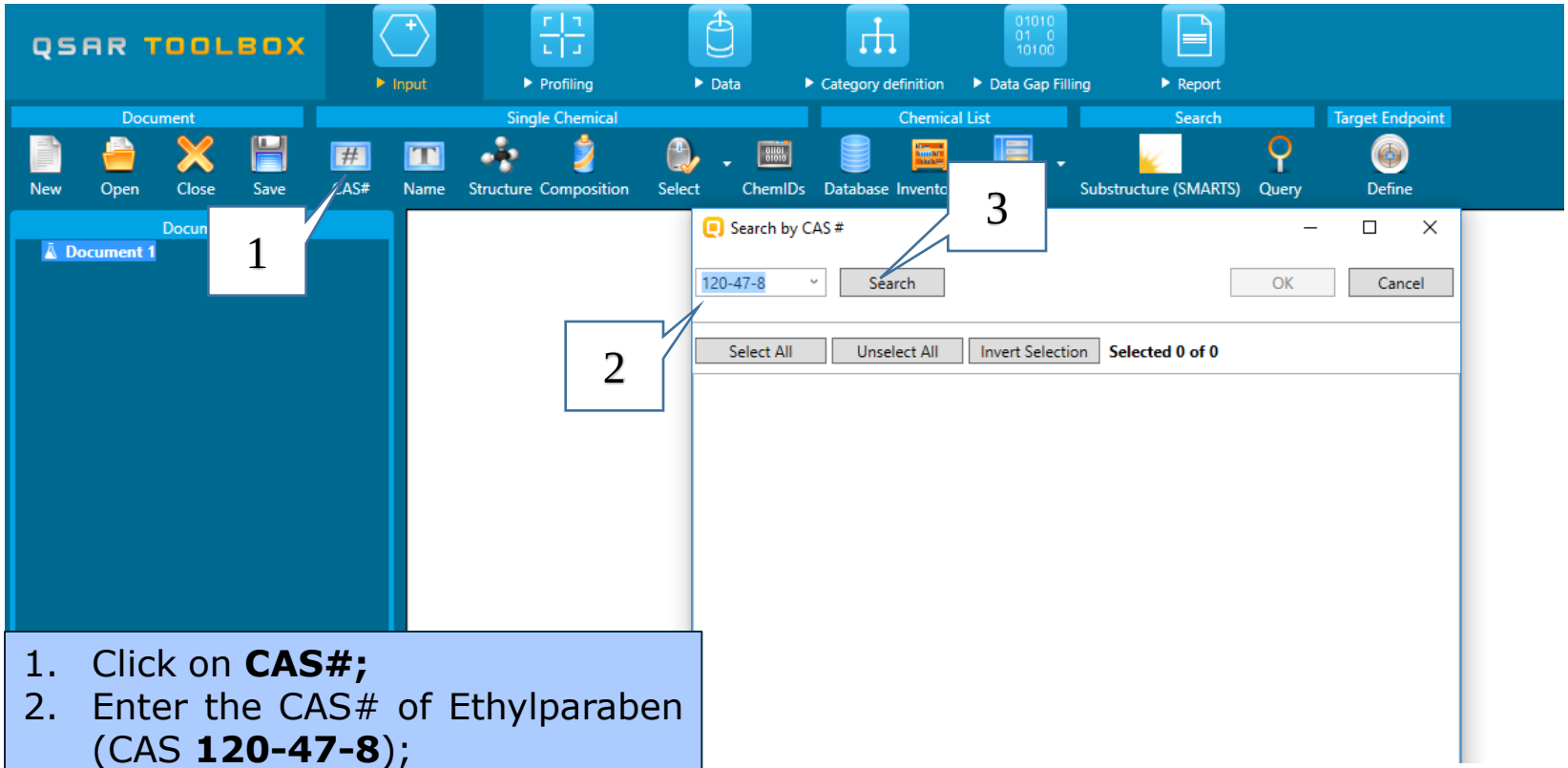
Ways of Entering a Chemical



- Chemical Name
- Chemical Abstract Services (CAS) number (#)
- Drawing chemical structure
- Select from User List/Inventory/Databases

Input Screen

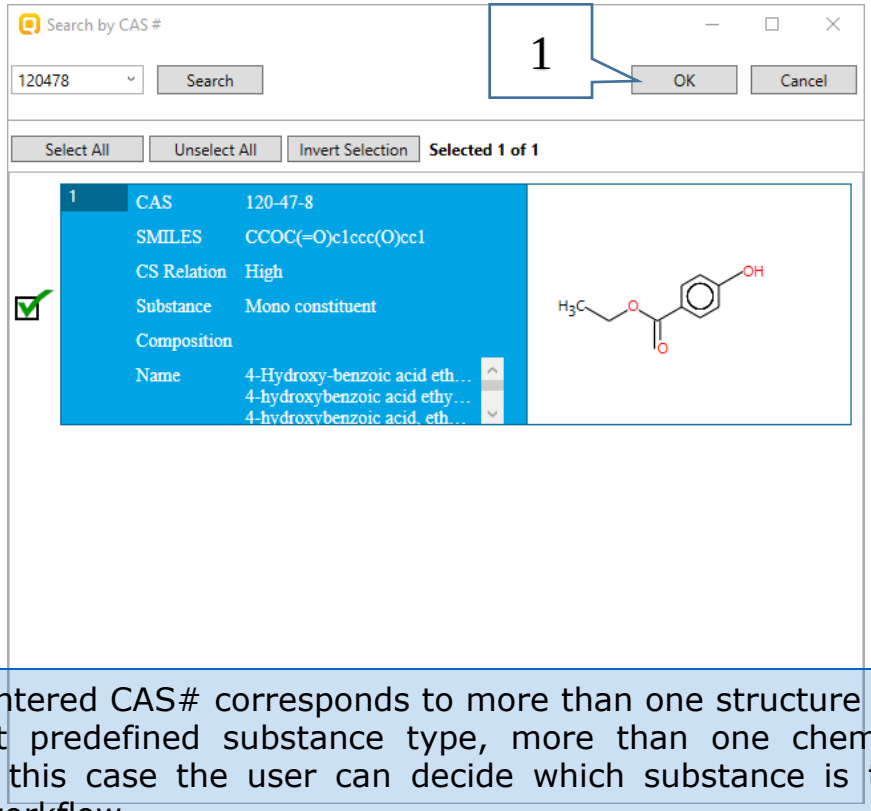
Input target chemical by CAS#



1. Click on **CAS#;**
2. Enter the CAS# of Ethylparaben (CAS **120-47-8**);
3. Click on **Search**.

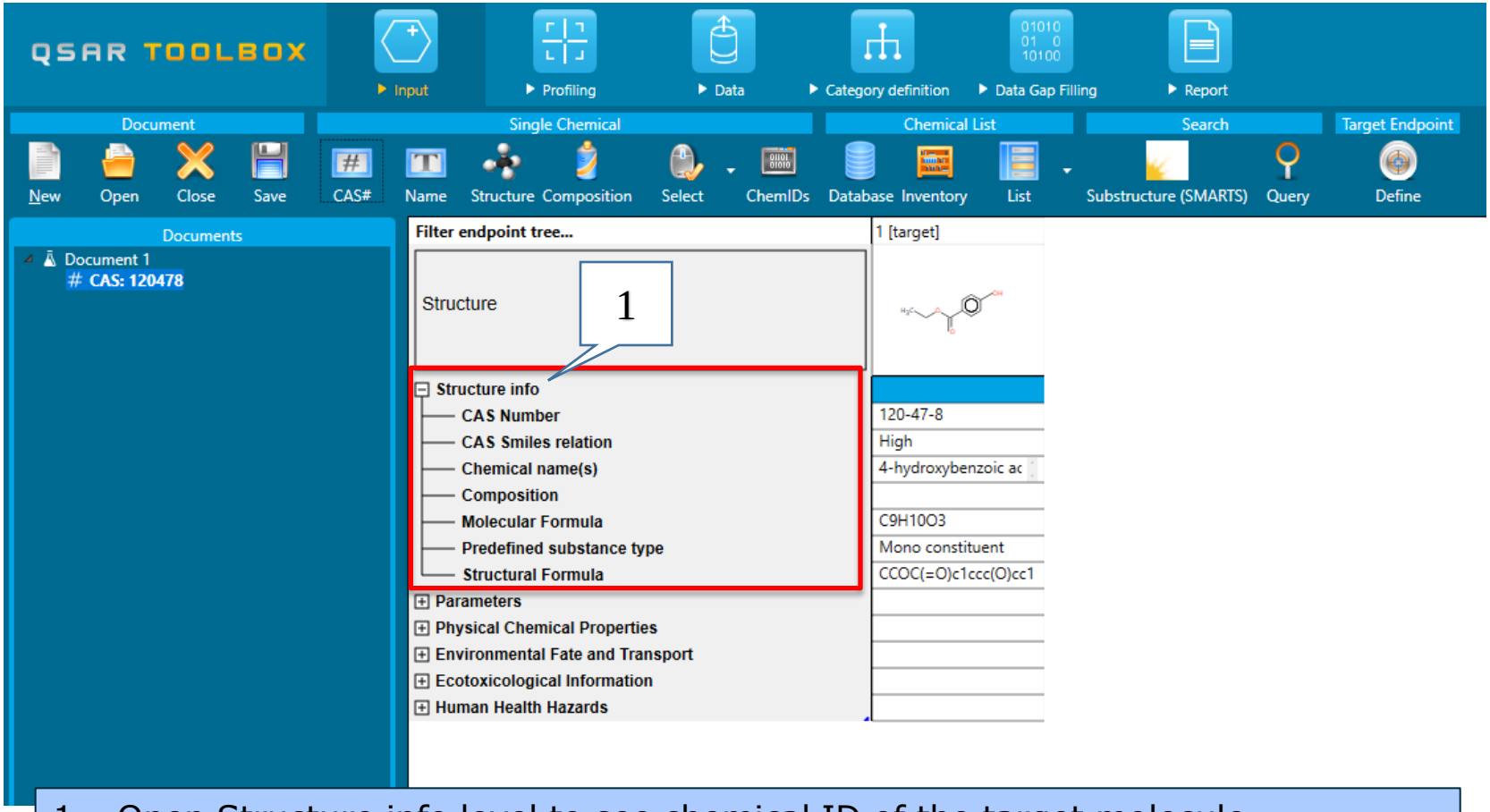
Input
Target chemical identity

The Toolbox now searches the databases to find out if the CAS# you entered is linked to a molecular structure stored in the Toolbox. It is displayed as a 2-demensional depiction. Click on **OK** (1).



In case the entered CAS# corresponds to more than one structure or to one structure but with different predefined substance type, more than one chemical identity could be retrieved. In this case the user can decide which substance is to be retained for the subsequent workflow.

Input
Target chemical identity



1. Open Structure info level to see chemical ID of the target molecule

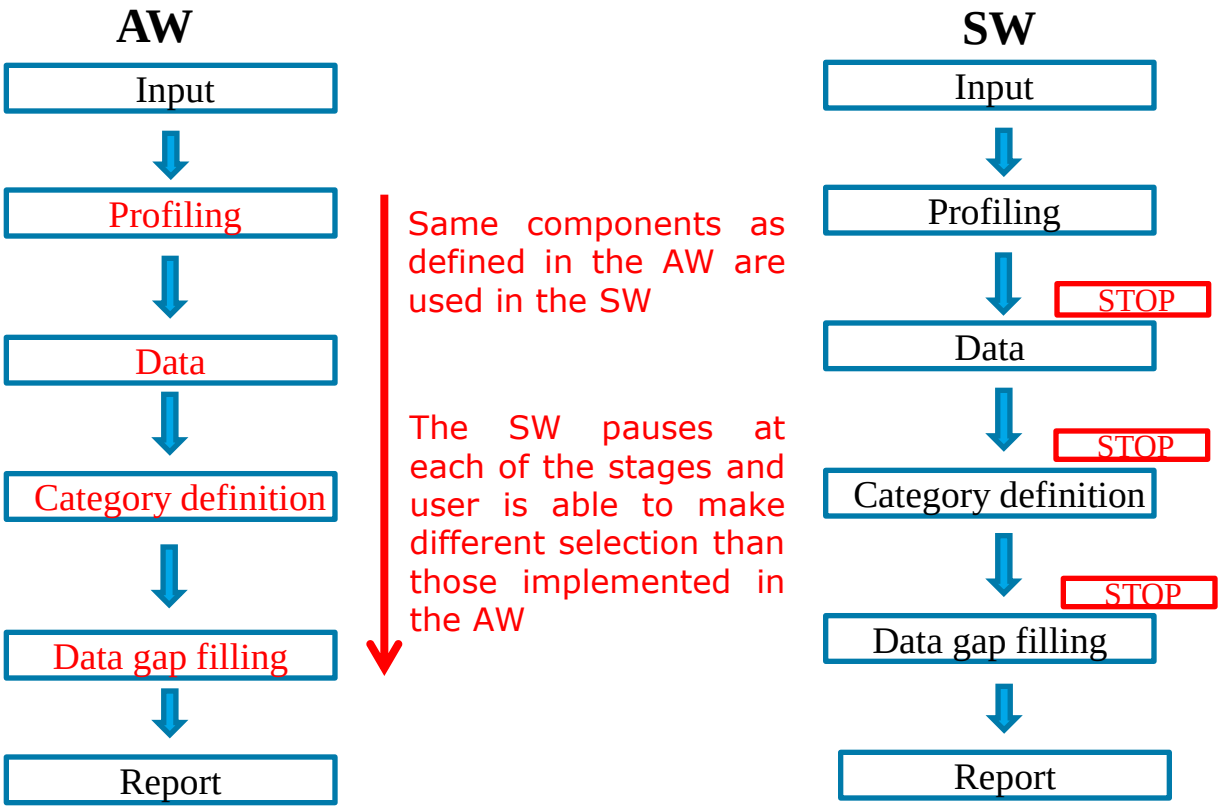
Outlook

- Background
- Objectives
- Specific Aims
- Standardized workflow for Skin sensitization
- The exercise
- **Standardized workflow execution**
  - Input
  - **Data Gap Filling**

Data Gap Filling
Overview

- "Data Gap Filling" (DGF) module give access to five different data gap filling tools:
  - Read-across
  - Trend analysis
  - (Q)SAR models
  - Standardized workflow
  - Automated workflow
- Depending on the situation, the most relevant data gap mechanism should be chosen, taking into account the following considerations:
  - Read-across is the appropriate data-gap filling method for "qualitative" endpoints like skin sensitisation or mutagenicity for which a limited number of results are possible (e.g. positive, negative, equivocal). Furthermore read-across is recommended for "quantitative endpoints" (e.g., 96h-LC50 for fish) if only a low number of analogues with experimental results are identified.
  - Trend analysis is the appropriate data-gap filling method for "quantitative endpoints" (e.g., 96h-LC50 for fish) if a high number of analogues with experimental results are identified.
  - "(Q)SAR models" can be used to fill a data gap if no adequate analogues are found for a target chemical.
  - Standardized and Automated workflows are developed to facilitate the users work. Once started, they follow the implemented logic and finish with prediction. The general differences between the two types of workflows are represented on the next slide.

Data Gap Filling
Overview



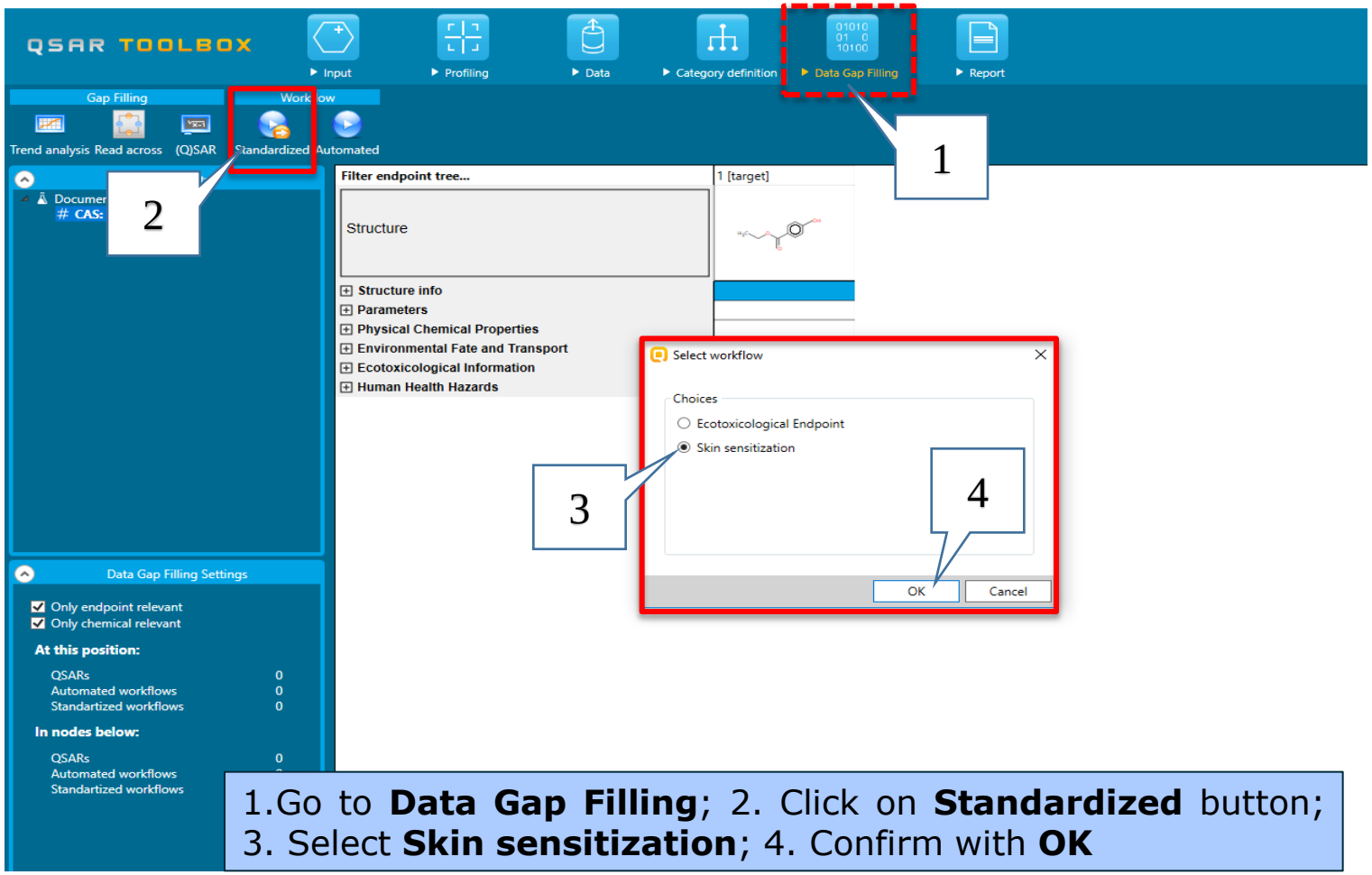
Databases with data for the target endpoint are listed and user select to use all of them or make specific selection.

Relevant to the workflow profilers appropriate for DGF are listed and ordered hierarchically based on the population of the group and user is able to select any of them.

Additional data filtering could be applied (e.g. different species selection)

In this example, we will use the Standardized workflow approach.

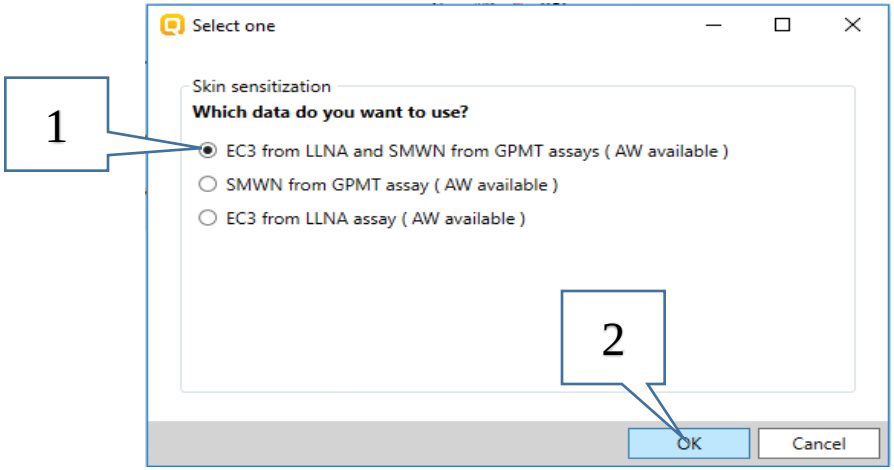
Data Gap Filling
Apply Standardized workflow



1.Go to **Data Gap Filling**; 2. Click on **Standardized** button; 3. Select **Skin sensitization**; 4. Confirm with **OK**

Data Gap Filling
Apply Standardized workflow

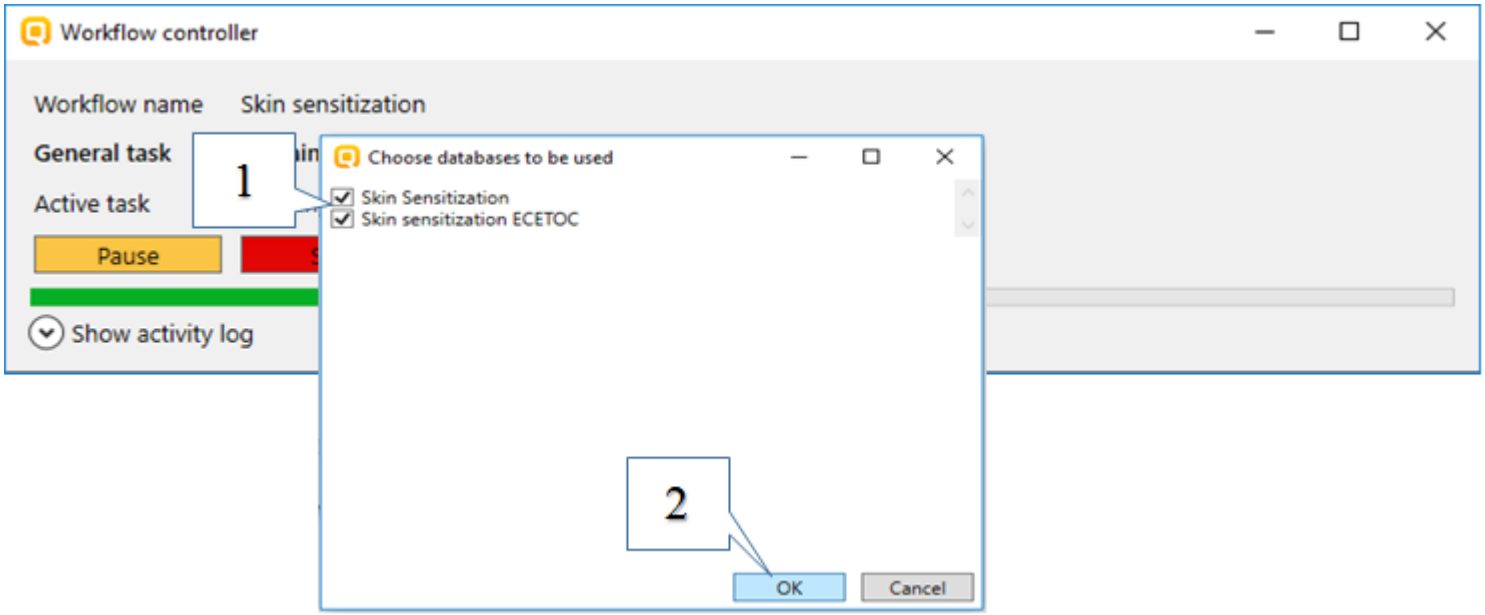
There are three options for selection of endpoint and user should select one of them. The first option allows more analogues to be found because of the using of two types of data – EC3 and SMWN. During the workflow process all skin data of these endpoints will be converted to the default scale for Skin Sensitisation - "Skin Sensitisation ECETOC". It converts all skin data into: Positive and Negative.



1. Select the first combined endpoint – EC3 from LLNA and SMWN from GPMT assays; 2. Confirm with **OK**

Data Gap Filling
Apply Standardized workflow

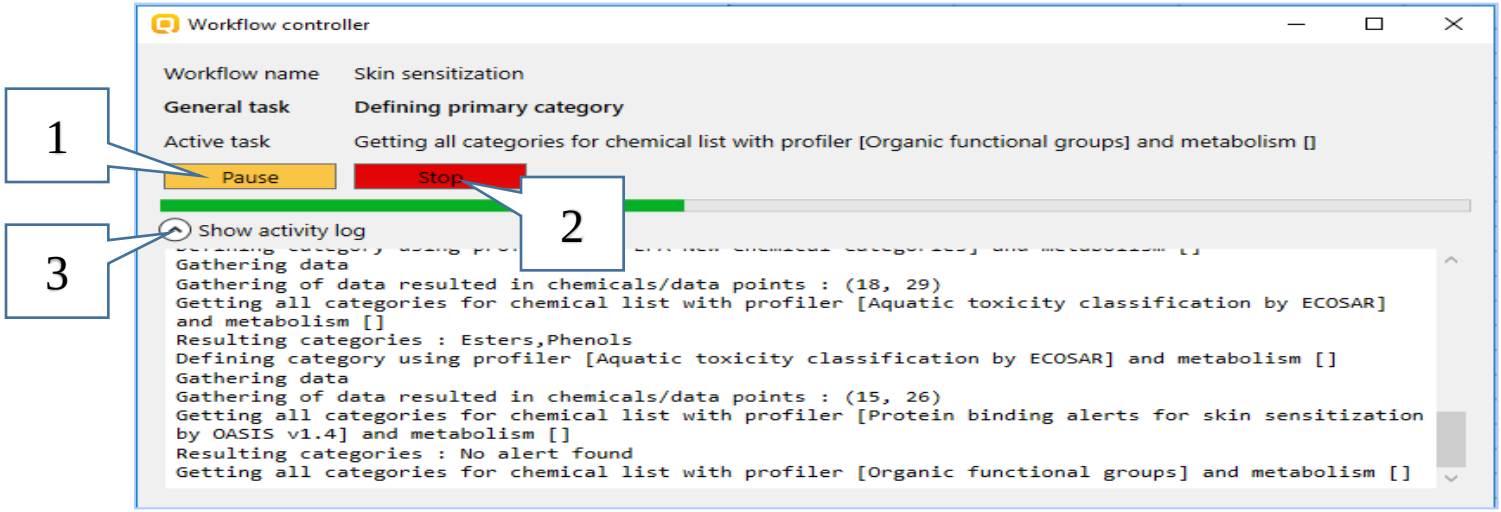
Once the workflow is started, *Workflow controller* appears. The first choice, which the user have to make is to select one or all skin sensitization databases, where the analogues will be searched.



1. Select the both databases – **Skin Sensitization** and **Skin sensitization ECETOC**;
2. Click on **OK**

Data Gap Filling
Apply Standardized workflow

The **Workflow controller** has two main buttons **Continue/Pause** (1) allowing to continue or pause and **Stop** (2) – which will stop the workflow. Additionally, all actions that will be done during the execution of the workflow are tracked down and could be seen from the **Show activity log** part (3) of the Workflow controller.



Data Gap Filling
Algorithm of Skin sensitization workflow

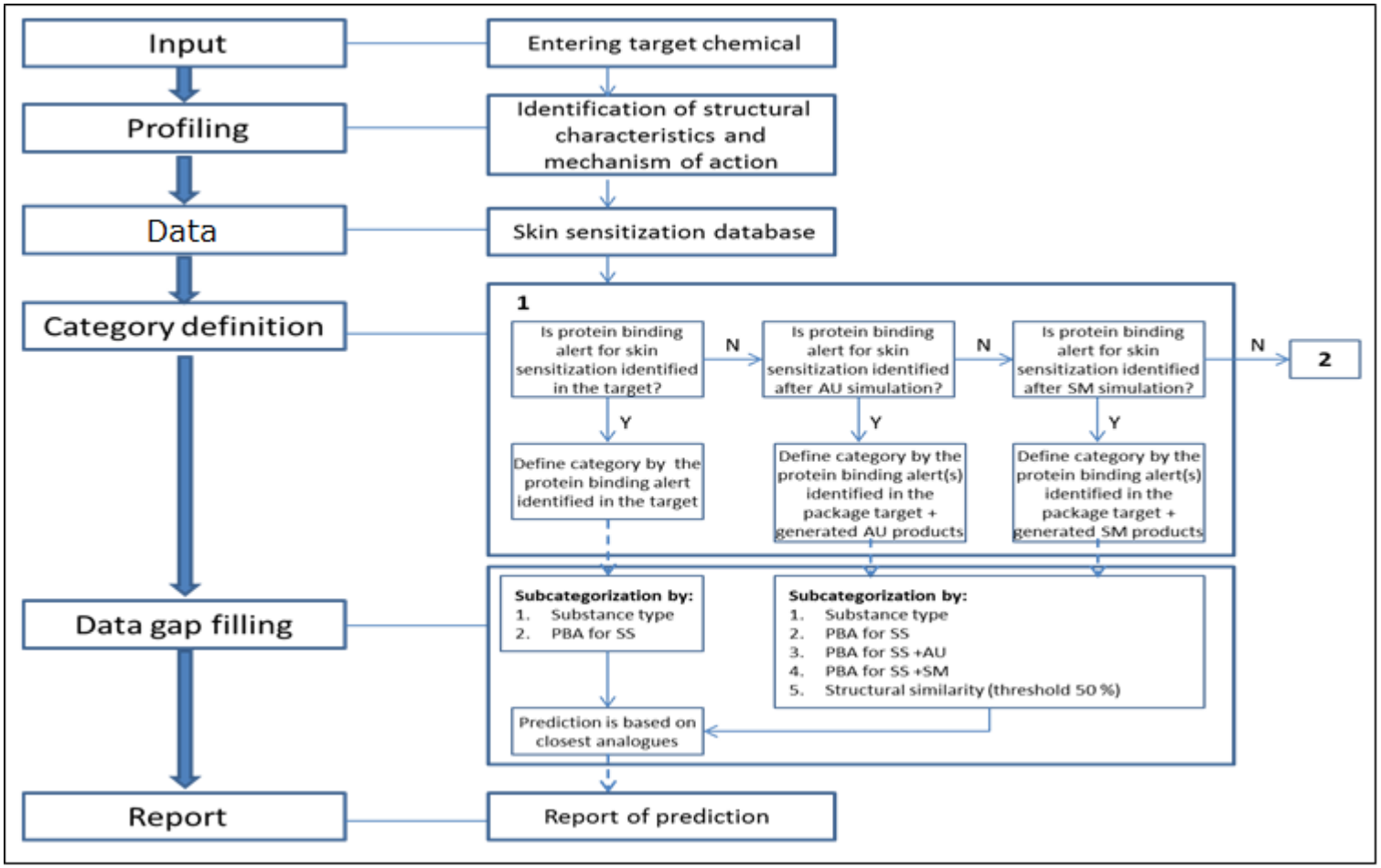
Once finished with the selection of databases, the workflow continues with application of the relevant profilers. There are three possible cases to form an analogues set:

1) if the target have an active alert as a parent;
2) if the target have an active alert as a result of autoxidation or skin metabolism activation;
3) if no alert is found in the target or its metabolites.

If a protein binding alert is identified in the target or is produced as a result of autoxidation or skin metabolism (case 1 or 2) then primary grouping is based on this alert.
In the last case (case 3) the primary group is defined using structurally based profiling schemes.

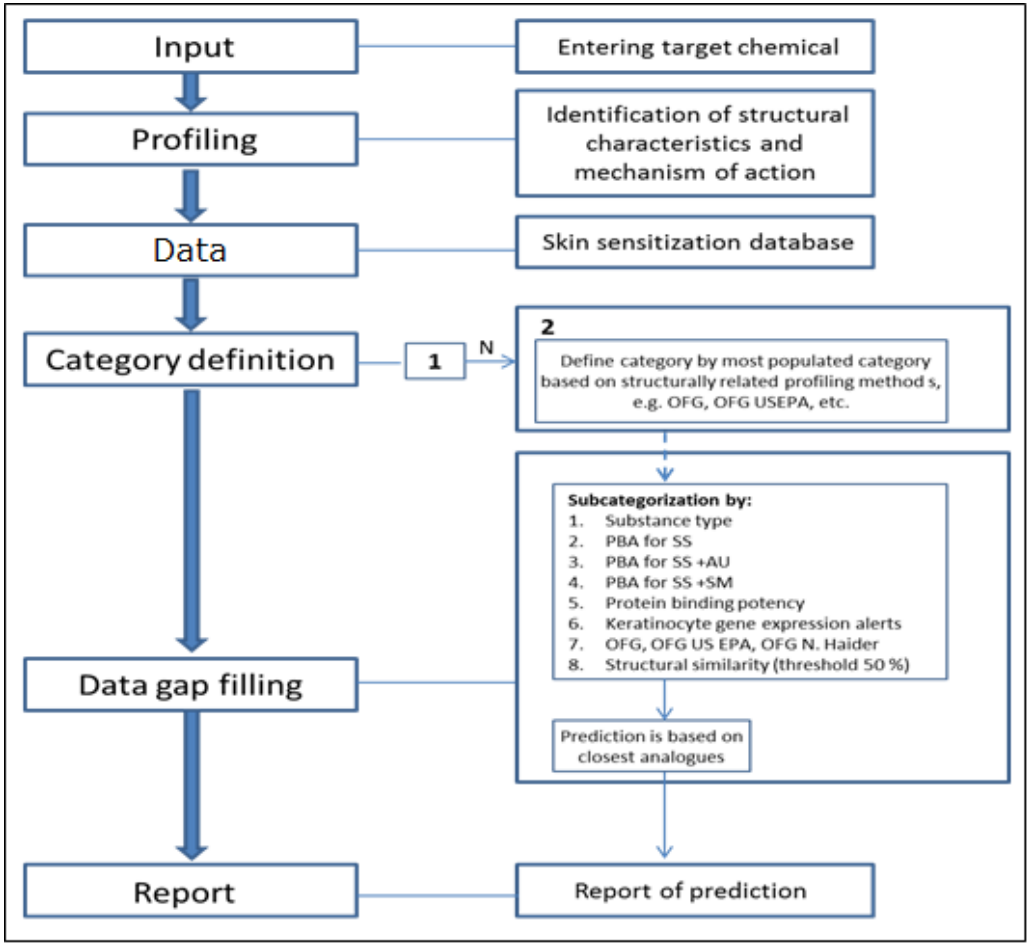
Data Gap Filling

Scenario 1 – an alert is identified in the parent or produced as a result of autoxidation or skin metabolism activation



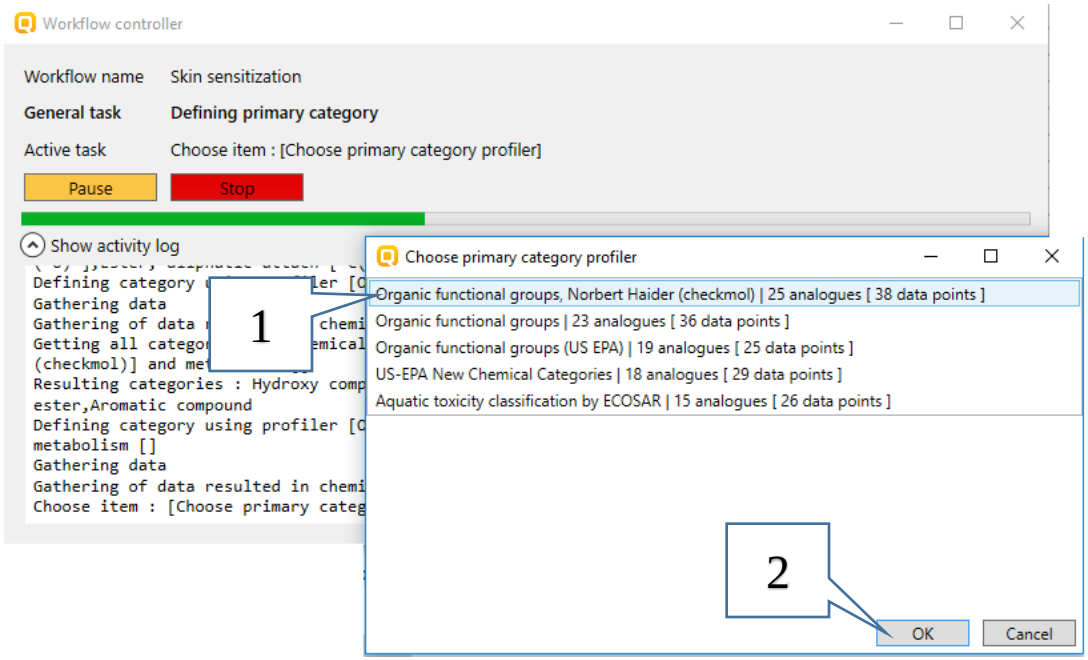
Data Gap Filling

Scenario 2 – no alert is identified in the parent neither in generated metabolites (autoxidation products and skin metabolites)



Data Gap Filling
Apply Standardized workflow

When a group cannot be defined by the workflow (more than one alert available in the target chemical or different primary groupings based on structure based profilers) then the standardized workflow wait for action from the user.



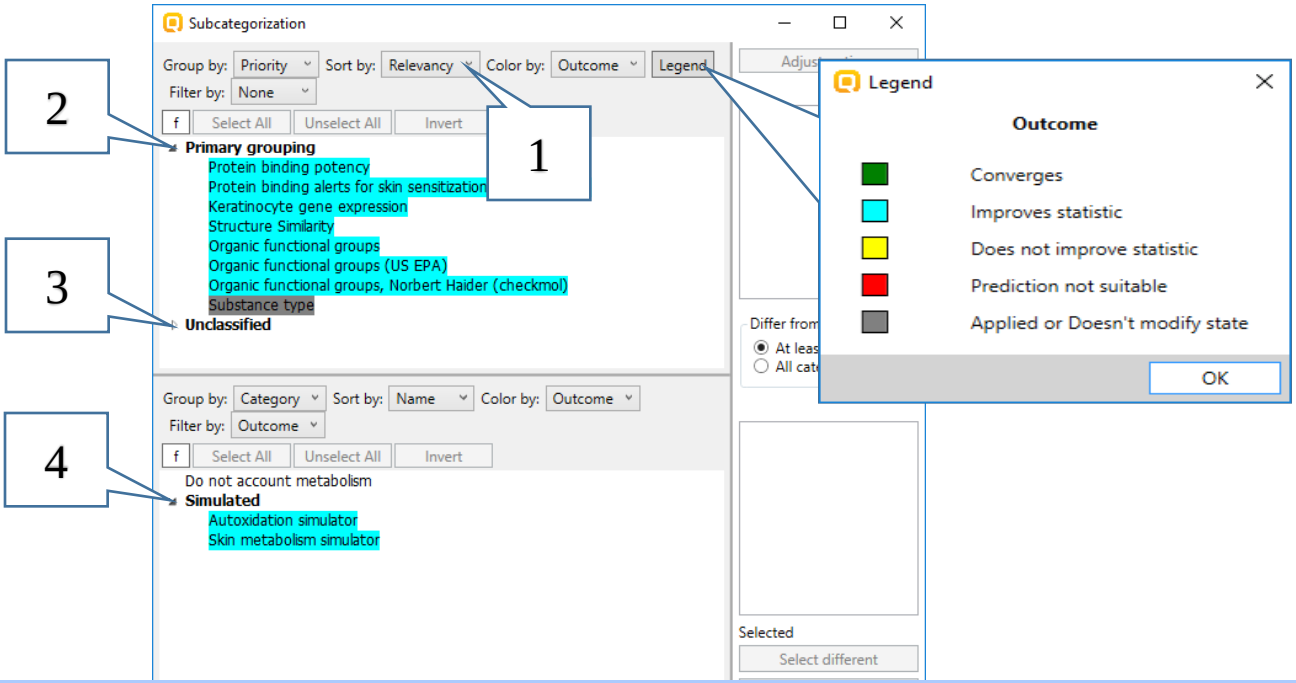
1. Select the most populated category; 2. Confirm with **OK**

Data Gap Filling
Apply Standardized workflow

When the profiling scheme for primary categorization is selected, the workflow makes a category and enters in Data Gap Filling, where the next step is subcategorization. Depending on the outcome obtained by the applied subcategorization, the profilers are colored as follows:

- **Green** – application of the profiler will satisfy the criteria for acceptance of the prediction
- **Blue** – application of the profiler increase the confidence of the prediction only
- **Yellow** – application of the profiler does not change the current state
- **Red** – criteria for acceptance the subcategorization will be not reached
- **Grey** – already applied profiler

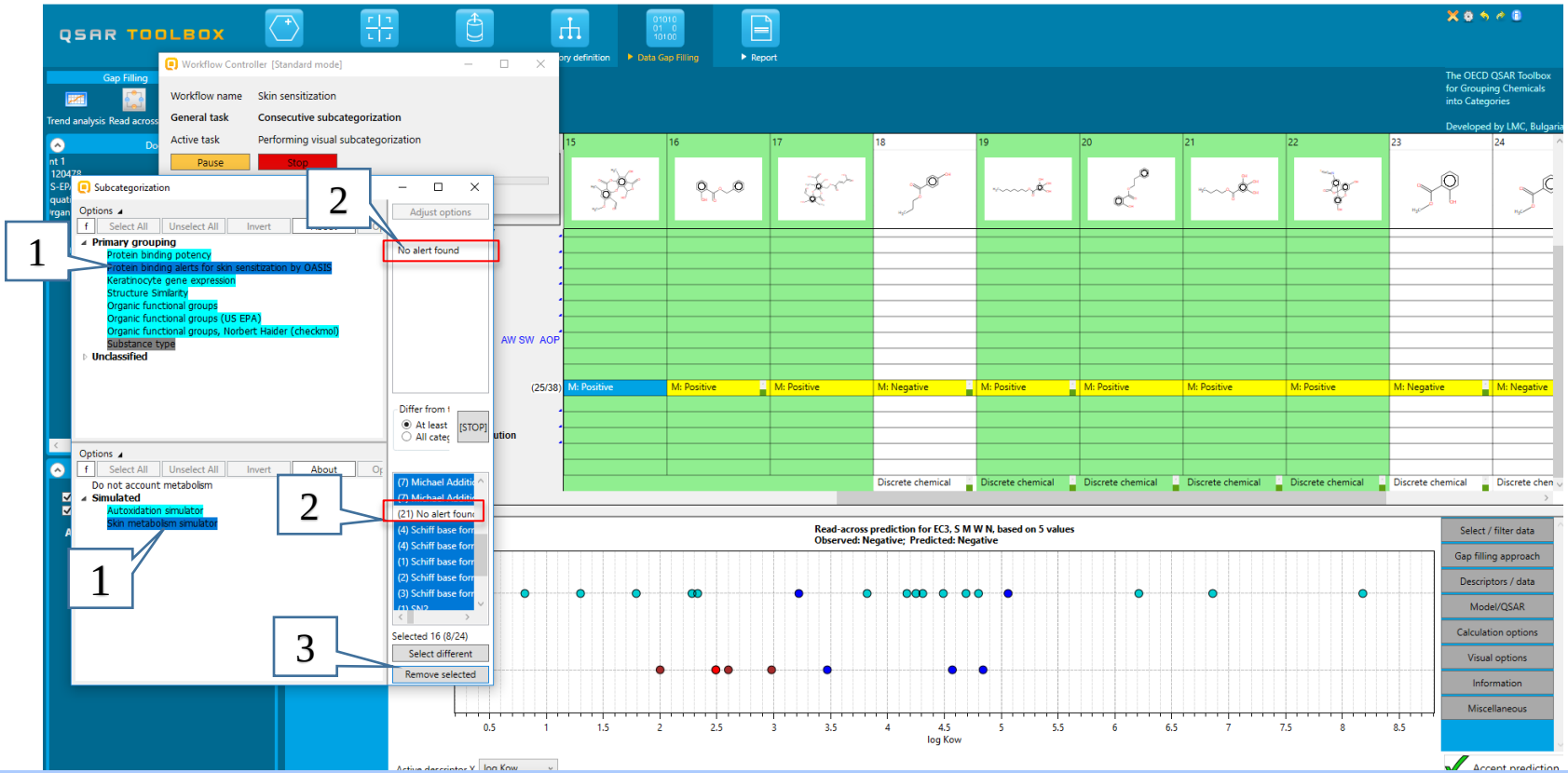
Data Gap Filling
Apply Standardized workflow



The profilers are sorted by ***Relevancy*** (1) and the highlighted profilers appear in group ***Primary grouping*** (2). All the rest profilers are under group ***Unclassified*** (3). Although the highlighted profiles are relevant to the endpoint, recommended and facilitating the user in their further choice, the rest of the profilers are not removed and user can select each of them for subcategorization.
In case the target have no alert as a parent, the related metabolic simulator will be also highlighted (4).

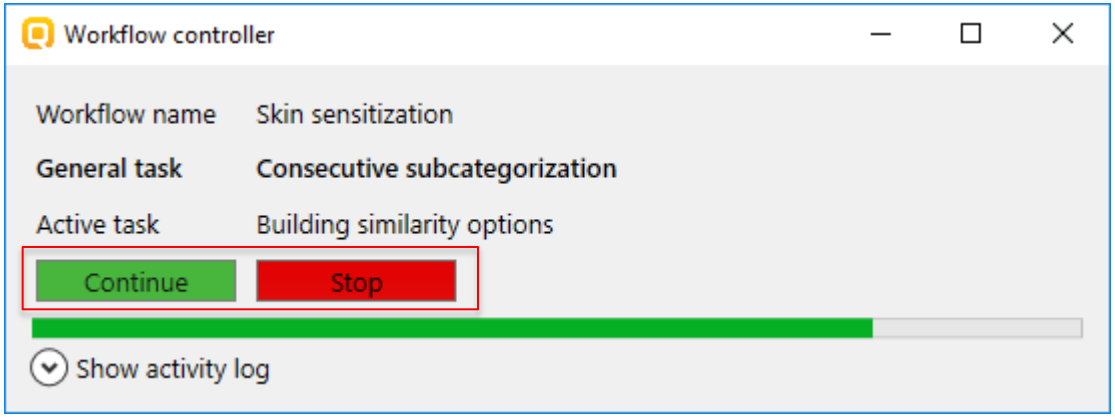
Data Gap Filling

Apply Standardized workflow



Select ***Protein binding alerts for skin sensitization by OASIS.*** in combination with ***Skin metabolism simulator*** (1). No alerts for SS are found in the target, neither in its metabolites (2). Click on **Remove selected** (3).

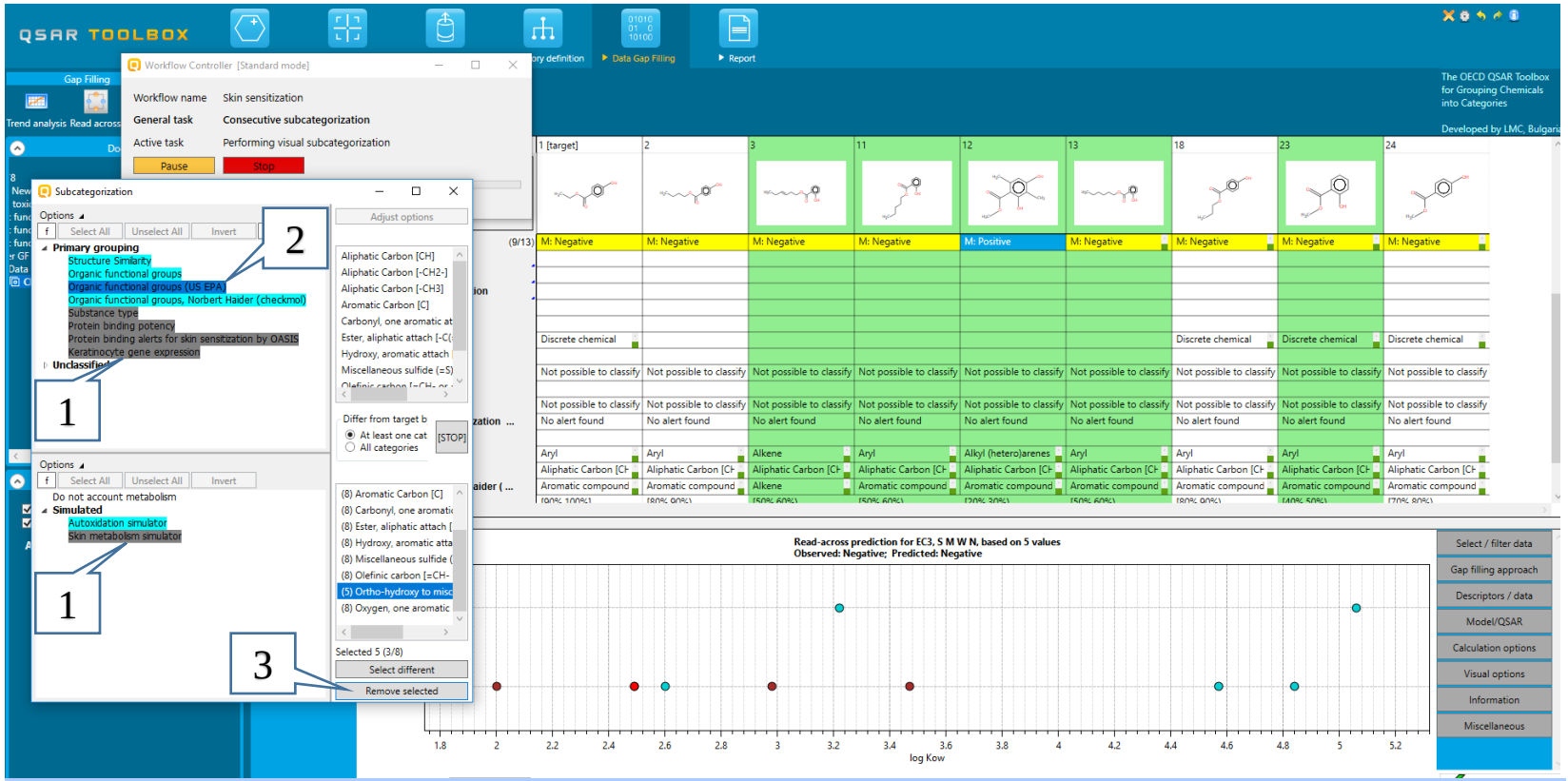
Data Gap Filling
Apply Standardized workflow



When the chemicals, which fulfill the requirements are removed, the *Workflow controller* stops before the next subcategorization. The user can continue to subcategorize or to stop, accepting the current state of the prediction.

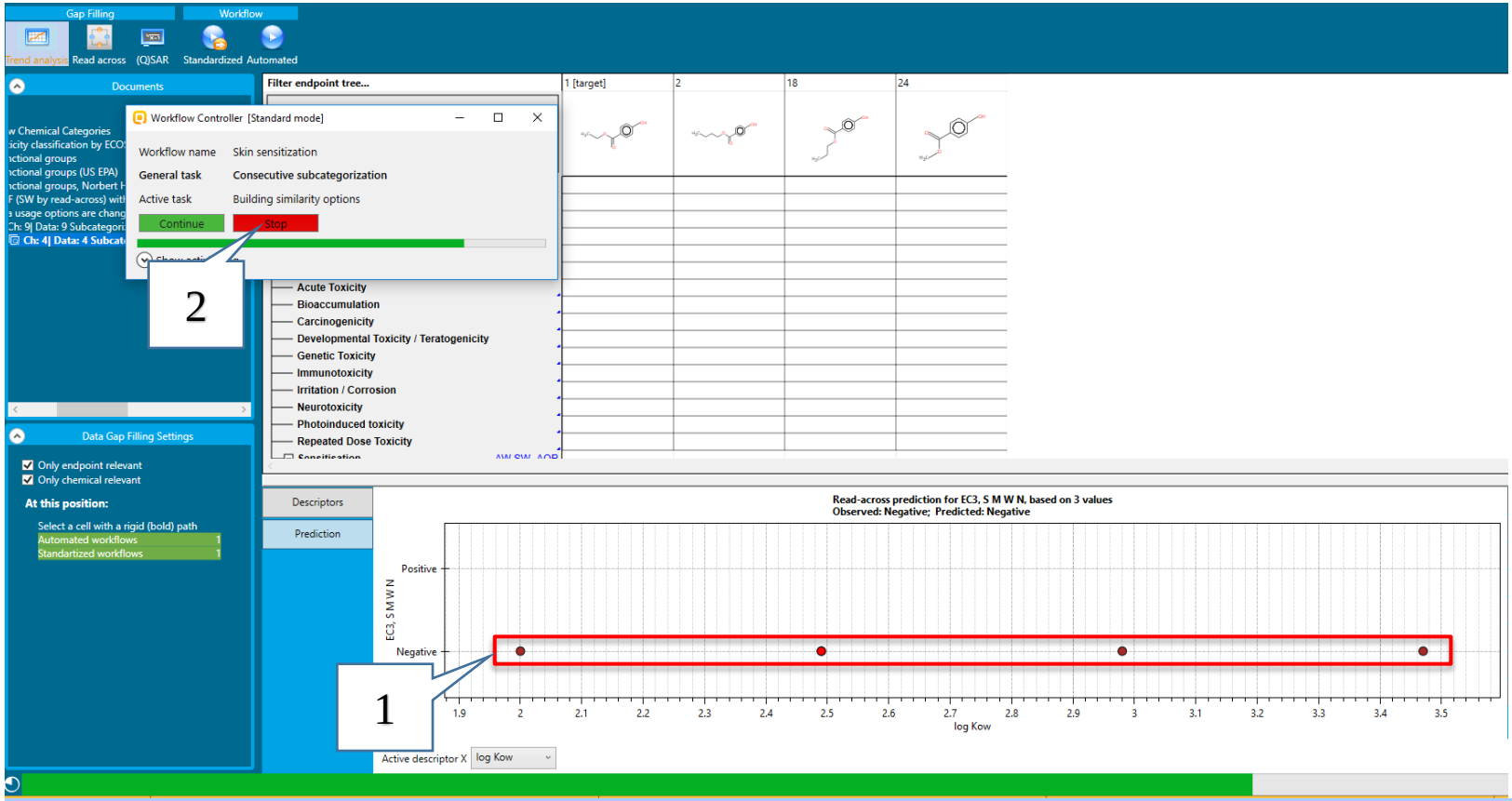
In this example we will continue with consecutive subcategorization. Click on Continue button

Data Gap Filling
Apply Standardized workflow



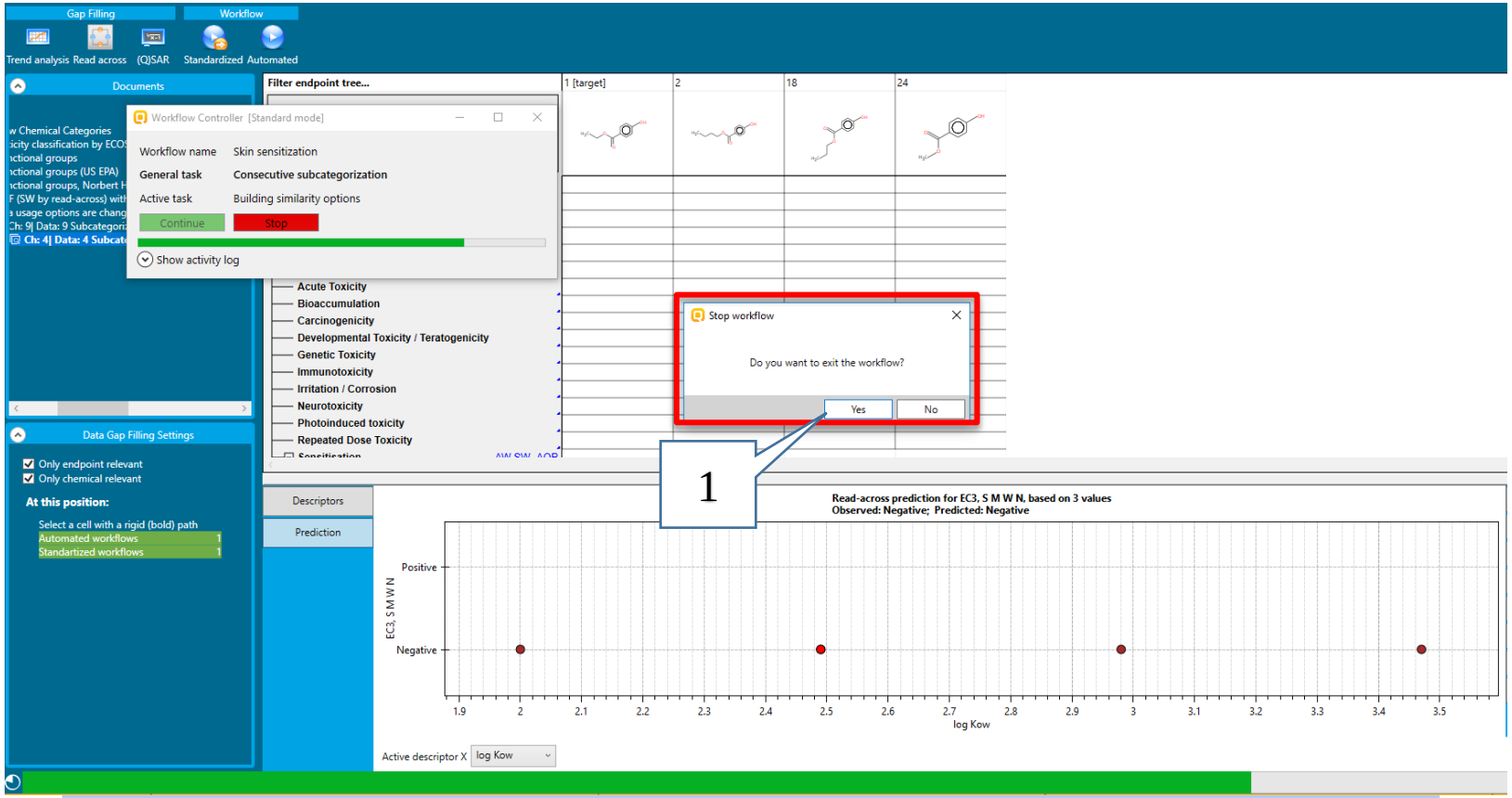
The used profiling schemes and metabolism simulators are colored in grey (1).
Now we select ***Organic functional groups (US-EPA)*** (2) and remove the different chemicals (3).

Data Gap Filling
Apply Standardized workflow



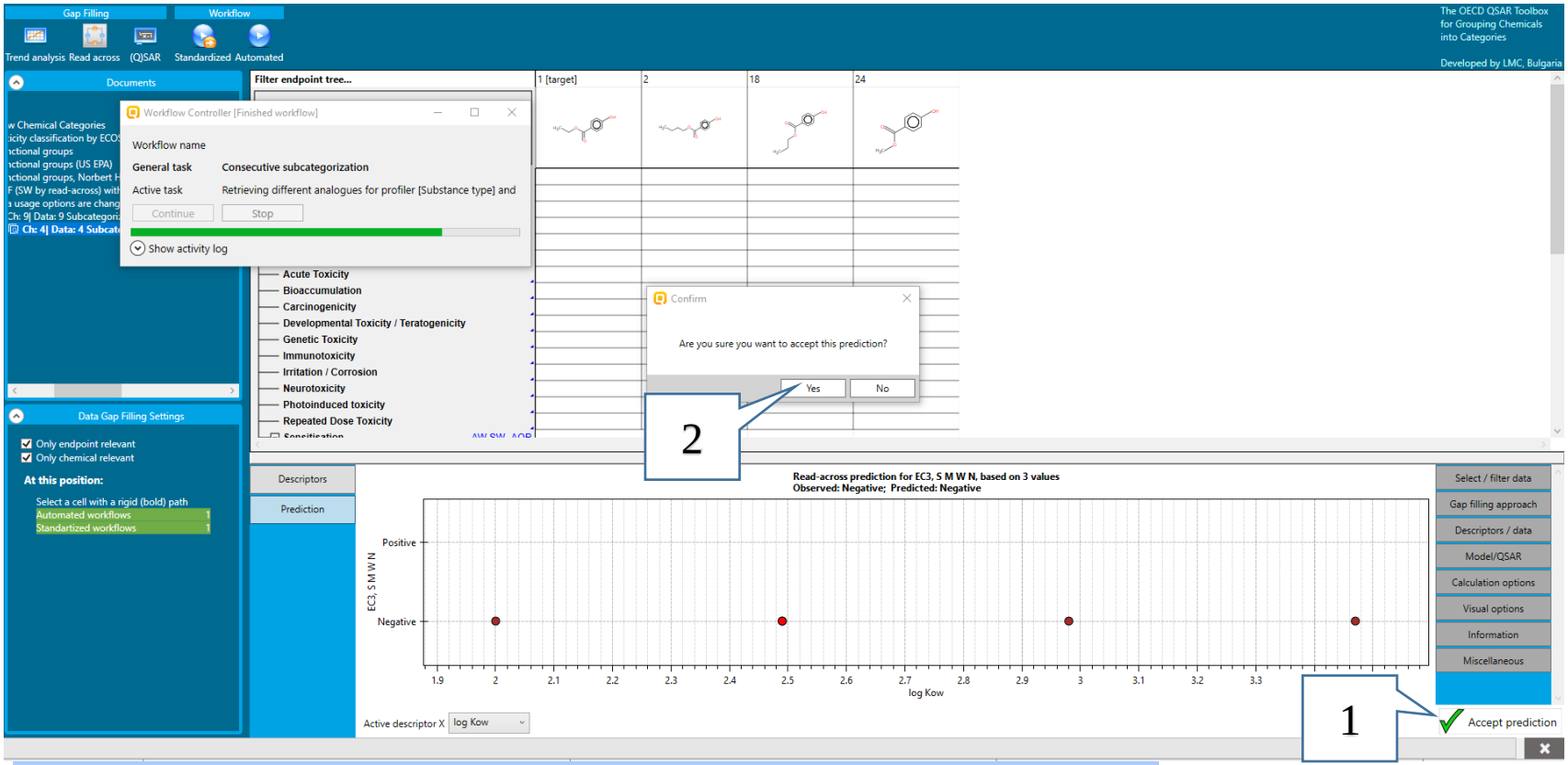
We see that all remained chemicals after the second subcategorization posses negative data (1). Therefore, here we click on the **Stop** button (2) of the workflow controller.

Data Gap Filling
Apply Standardized workflow



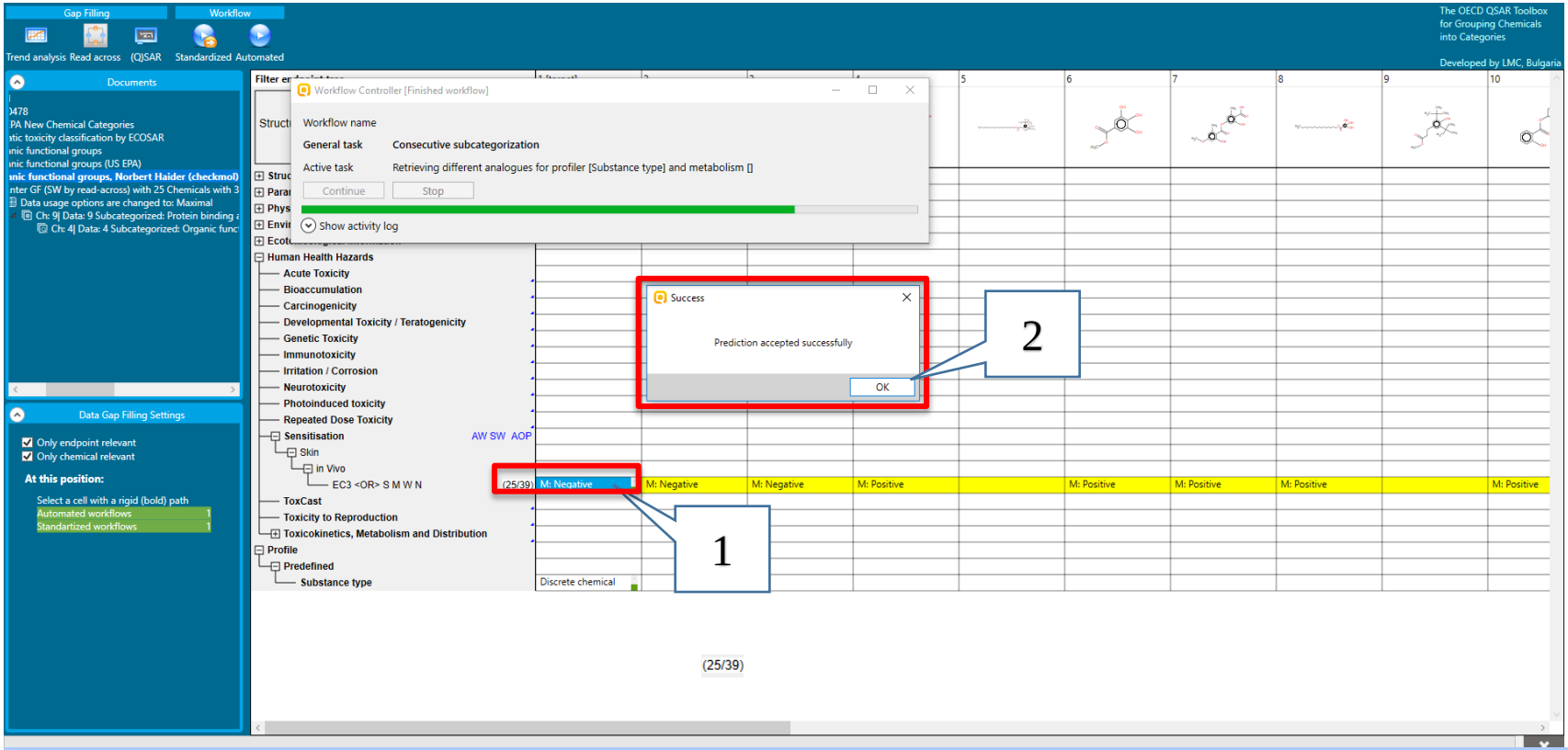
"Stop workflow" dialogue appears to ask the user if they are sure to exit the workflow. Confirm with "**Yes**" (1)

Data Gap Filling
Apply Standardized workflow



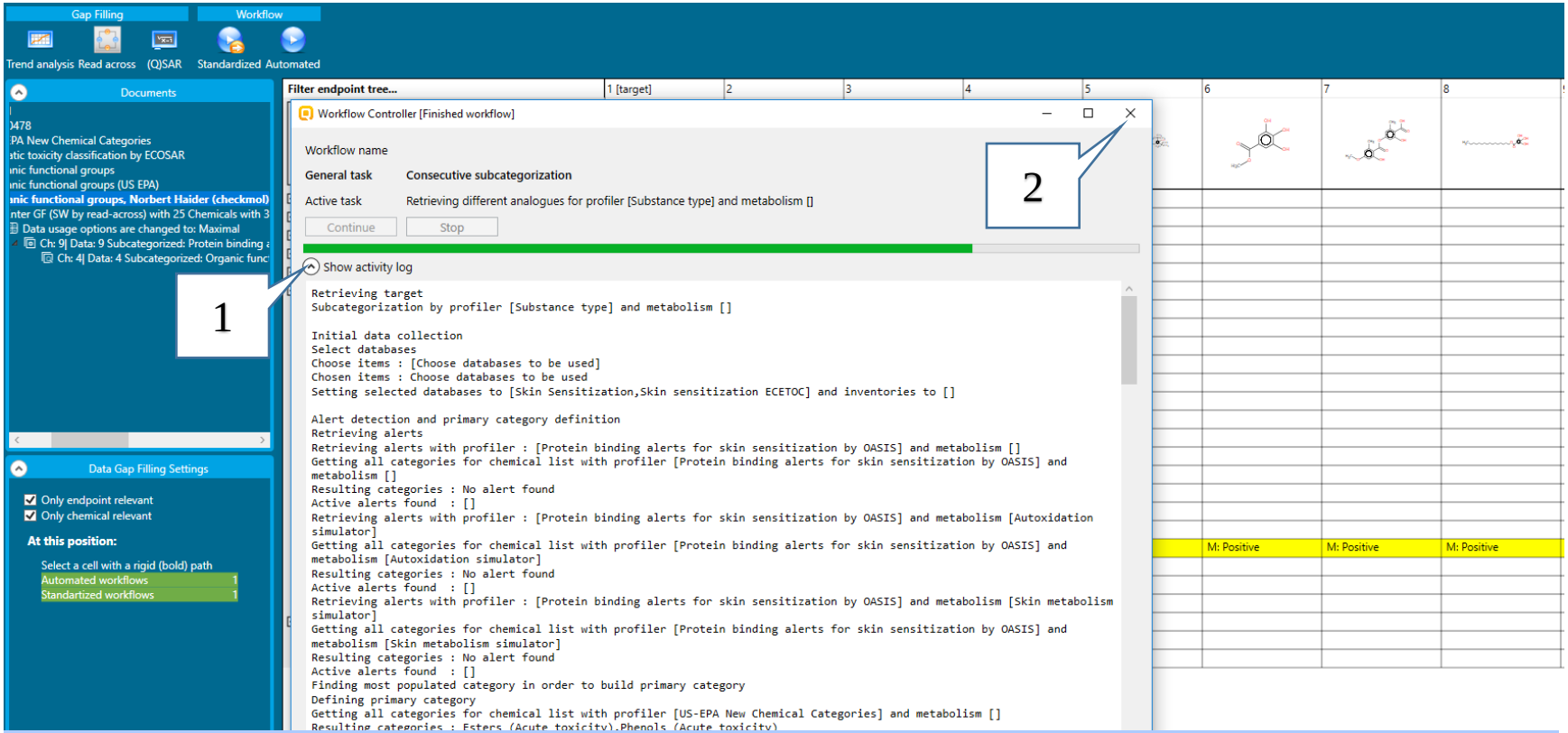
Now we can accept the prediction by click on **Accept prediction** (1) and to confirm this activity by the "**Yes**" button (2).

Data Gap Filling
Apply Standardized workflow



The next message informs that the prediction is accepted successfully and it appears on the matrix (1). Click on “**OK**” (2)

Data Gap Filling
Apply Standardized workflow



The *Workflow Controller* **does not** close itself automatically. The user can expand the *activity log* (1) and to examine all performed steps during the Standardized workflow execution. After that the controller have to be closed by click on the close button (2).

Data Gap Filling
Recap

The application of the SW for skin sensitization requires user activities such as:

- specification of the endpoint;
- selection of databases;
- selection of primary grouping in case no alert is identified in the target as parent and after autoxidation and skin metabolism activation;
- selection of highlighted profiling schemes for subcategorization in order to increase the confidence of the prediction;
- accepting the prediction.

Outlook

- Background
- Objectives
- Specific Aims
- Read across and analogue approach
- The exercise
- **Workflow**
  - Input
  - Data Gap Filling
  - **Report**

Report
Overview

- Report module could generate report on any of predictions performed with the Toolbox.
- Report module contains predefined report template which users can customized.
- Two type of report files are generated:

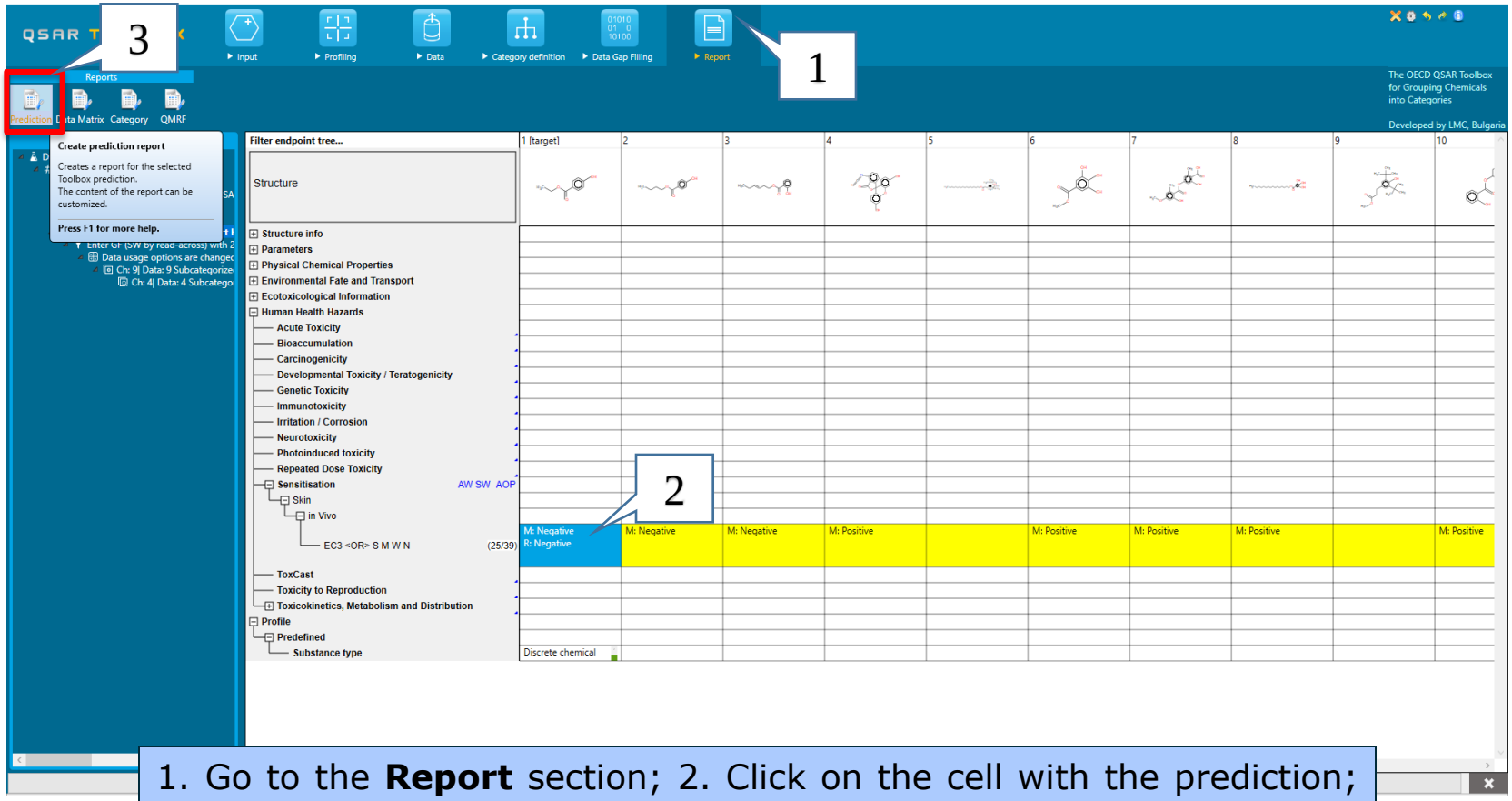
- *Prediction report* – containing information for the target

- *Data matrix* – containing information for the analogues used for the prediction

- *Category report* – containing information for the analogues in the category

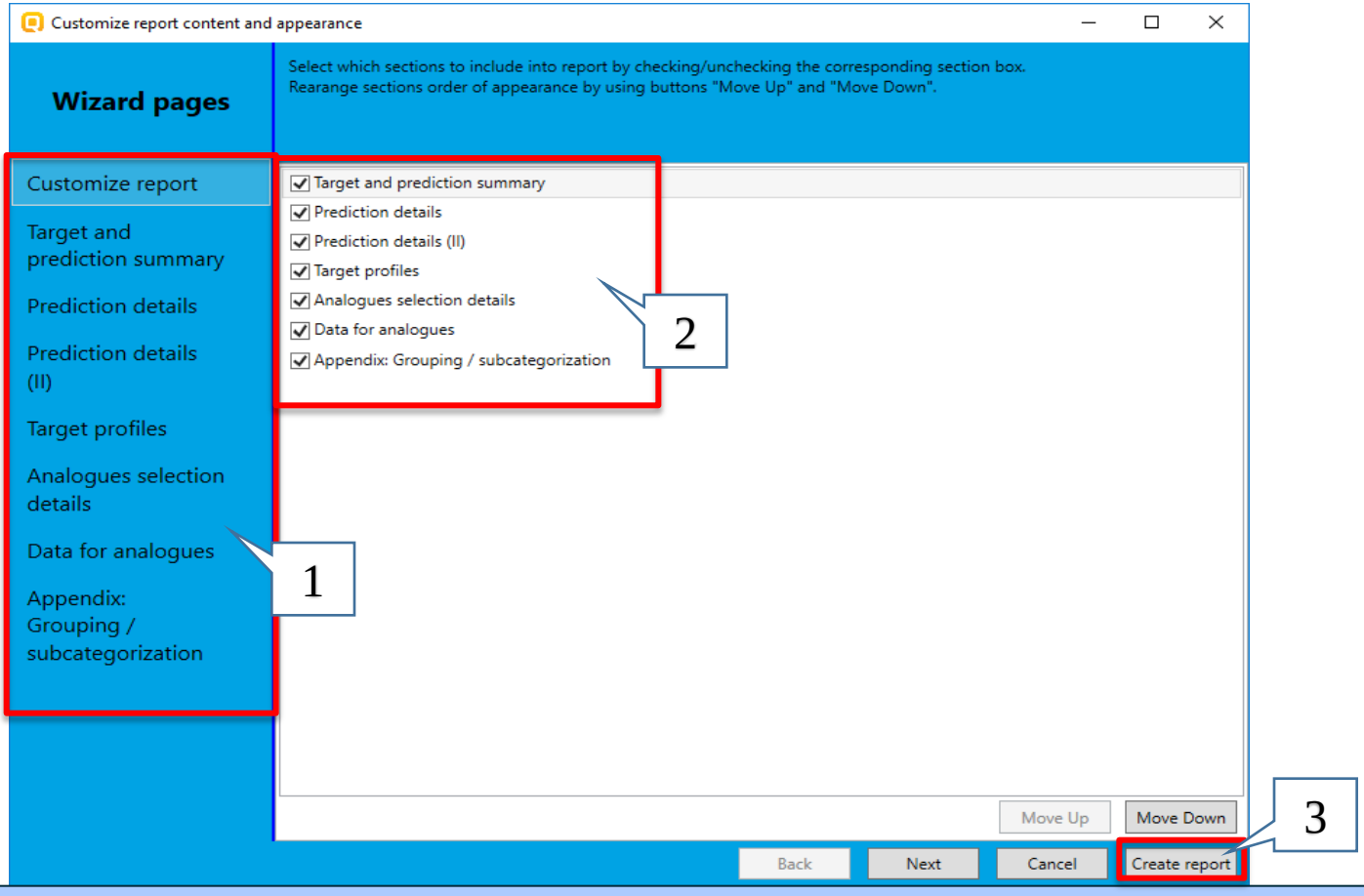
- *QMRF report* – containing information for the applied QSAR model.

Report
Generation report



1. Go to the **Report** section; 2. Click on the cell with the prediction;
3. Click on **Prediction**

Report
Generation report



The user can customize the report content (1) and appearance (2). Generation of the reports happens by click on the **Create report** button (3).

Report
Wizard pages

- **Customized report** - the user is able to include or exclude the sections in the report.

- **Target and prediction summary** – This section includes substance ID of the target chemical and the prediction outcome. Fields which are automatically populated by the system are indicated. Here the user could add information for the author, contact details and summary information

- **Prediction details** and **Prediction details (II)** – section prediction details provides details about the prediction and its reliability. Prediction details (II) is optional it provides specific information about the prediction depending on the gap filling approach.

- **Target profiles** – this section summarize profiles used for the prediction. Additional profiles could be also included by the user.

Report
Wizard pages

- **Analogues selection details** – This section illustrates how analogues were selected. It displays selected databases, category boundaries and applicability domain.

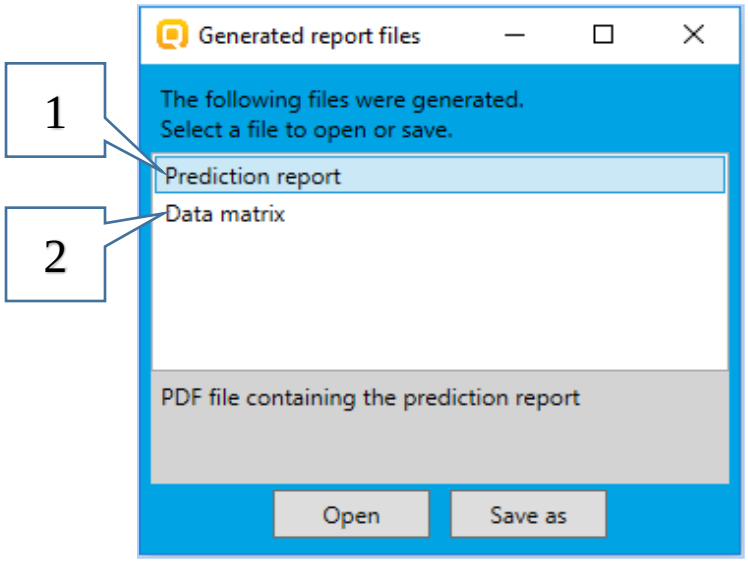
- **Data for analogues** – This section provides details information about the analogues used for obtaining the prediction including parameters, profilers and experimental data. Additional data, available for the used analogues into Toolbox, could be also included by the user.

- **Appendix: Grouping / subcategorization** – This section includes the categories used for primary categorization as well as the categories and metadata removed during the subcategorization automatically populated by the system.

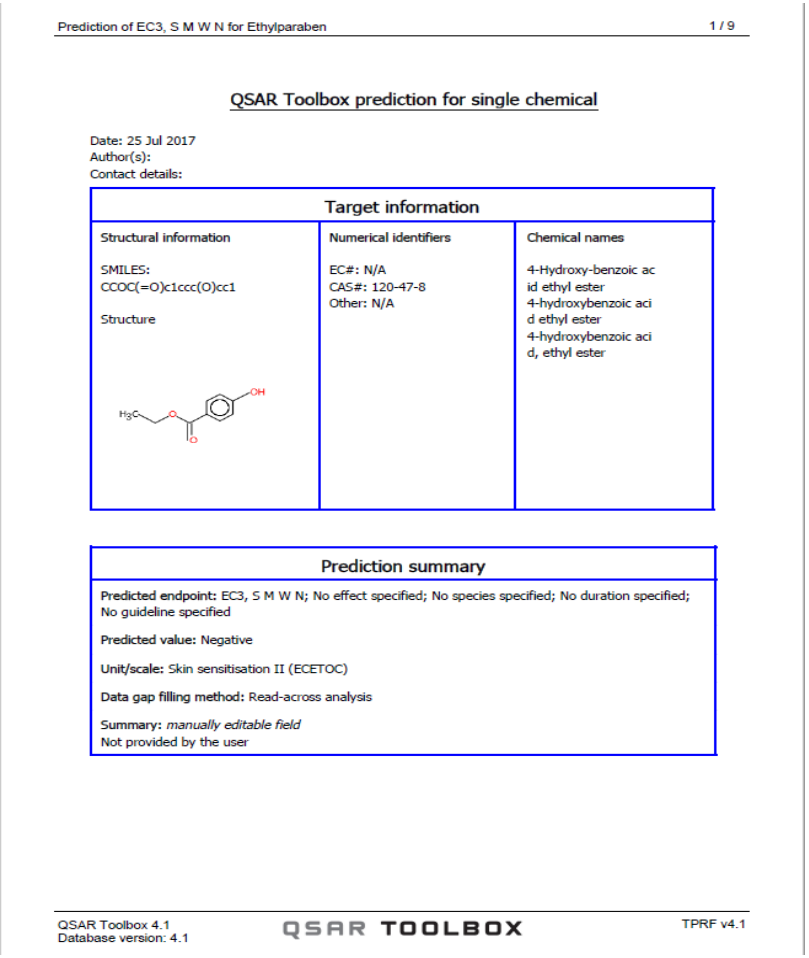
Report
Generation report

After the click on the Create report button, *Generated report files* window appears. It contains two type of files:

1) **Prediction report** - a PDF file containing the prediction information related to the target.
2) **Data matrix** - a MS Excel file containing chemicals used for prediction along with their data for selected parameters, profiles and endpoint tree positions.



Report
Generated report files



Prediction of EC3, S M W N for Ethylparaben 1 / 9

QSAR Toolbox prediction for single chemical

Date: 25 Jul 2017
Author(s):
Contact details:

Target information		
Structural information SMILES: CCOC(=O)c1ccc(O)cc1 Structure	**Numerical identifiers** EC#: N/A CAS#: 120-47-8 Other: N/A	**Chemical names** 4-Hydroxy-benzoic ac id ethyl ester 4-hydroxybenzoic aci d ethyl ester 4-hydroxybenzoic aci d, ethyl ester

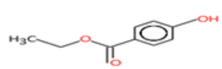
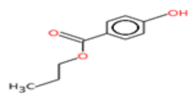
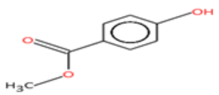
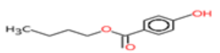
Prediction summary
Predicted endpoint: EC3, S M W N; No effect specified; No species specified; No duration specified; No guideline specified
Predicted value: Negative
Unit/scale: Skin sensitisation II (ECETOC)
Data gap filling method: Read-across analysis
Summary: *manually editable field* Not provided by the user

QSAR Toolbox 4.1
Database version: 4.1
QSAR TOOLBOX
TPRF v4.1

Using of a standardized workflow for predicting of skin sensitization potential is noted in the *Prediction report*.

Report

Generated report files

#		unit	Target chemical	Neighbour #1	Neighbour #2	Neighbour #3
2	**Substance identity**					
3	Structure					
4	CAS number		120-47-8	94-13-3	99-76-3	94-26-8
5	Chemical name		Ethylparaben	Propylparaben	Methylparaben	Butylparaben
6	Other identifier					
7	SMILES		CCOC(=O)c1ccc(O)cc1	CCCOC(=O)c1ccc(O)cc1	COC(=O)c1ccc(O)cc1	CCCCOC(=O)c1ccc(O)cc1
8						
9	**Parameters**	**unit**				
10						
11	**Profilers**					
12	***Profiles used for grouping/subcategorization***					
13	Organic functional groups, Norbert Haider (checkmol) (primary grouping)		Hydroxy compound; Phenol; Carboxylic acid derivative; Carboxylic acid ester; Aromatic compound	Hydroxy compound; Phenol; Carboxylic acid derivative; Carboxylic acid ester; Aromatic compound	Hydroxy compound; Phenol; Carboxylic acid derivative; Carboxylic acid ester; Aromatic compound	Hydroxy compound; Phenol; Carboxylic acid derivative; Carboxylic acid ester; Aromatic compound
14	Protein binding alerts for skin sensitization by		No alert found	No alert found	No alert found	No alert found
15	Organic functional groups (US EPA) (subcategorization)		Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Aromatic Carbon [C]; Carbonyl, one aromatic attach [-C(=O)-]; Ester, aliphatic attach [-C(=O)O]; Hydroxy, aromatic attach [-OH]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]; Oxygen, one aromatic attach [-O-]	Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Aromatic Carbon [C]; Carbonyl, one aromatic attach [-C(=O)-]; Ester, aliphatic attach [-C(=O)O]; Hydroxy, aromatic attach [-OH]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]; Oxygen, one aromatic attach [-O-]	Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Aromatic Carbon [C]; Carbonyl, one aromatic attach [-C(=O)-]; Ester, aliphatic attach [-C(=O)O]; Hydroxy, aromatic attach [-OH]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]; Oxygen, one aromatic attach [-O-]	Aliphatic Carbon [CH]; Aliphatic Carbon [-CH2-]; Aliphatic Carbon [-CH3]; Aromatic Carbon [C]; Carbonyl, one aromatic attach [-C(=O)-]; Ester, aliphatic attach [-C(=O)O]; Hydroxy, aromatic attach [-OH]; Miscellaneous sulfide (=S) or oxide (=O); Olefinic carbon [=CH- or =C<]; Oxygen, one aromatic attach [-O-]

Analogues used for the target prediction can be seen the **Data matrix** report. Their selected profiling results, experimental data and/or parameters are also shown.

Congratulation

- You have completed the tutorial on the standardized workflow for skin sensitization data gap filling.
- You have now been introduced to the consecutive steps of the standardized workflow of the QSAR Toolbox and the rationale behind each step.
- Note proficiency comes with practice.