OECD QSAR Toolbox v.4.0

Tutorial on how to predict skin sensitisation potential by automated workflow
Outlook

• Background
• Objectives
• Specific Aims
• Automated workflow
• The exercise
• Execution of AW for Skin sensitization
Background

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

- Background
- Objectives
- Specific Aims
- Automated workflow
- The exercise
- Execution of AW for Skin sensitization
Objectives

This presentation demonstrates how to:

• Fill data gaps for a chemical by automated workflow for skin sensitization (SS);

• Fill data gaps for a batch of chemicals by automated workflow for SS.
Outlook

• Background
• Objectives
• **Specific Aims**
  • Automated workflow
  • The exercise
• Execution of AW for Skin sensitization
Specific Aims

• To introduce to the user the automated workflow for predicting of skin sensitization potential of a chemical;
• To introduce to the user the automated workflow for predicting of skin sensitization potential of a batch of chemicals;
• 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
• Automated workflow
• The exercise
• Execution of AW for Skin sensitization
Automated workflow
Overview

• The automated workflow (AW) is designed to apply data gap filling for discrete chemicals only;
• The AW follows the implemented logic and finished with prediction;
• The AWs do not allow interactions during the workflow process;
• Analogous sets of chemicals are often selected based on the hypothesis that the toxicological effects of each member of the category will show a common behaviour;
• AW can be executed for one chemical as well as for a batch of chemicals.
Outlook

• Background
• Objectives
• Specific Aims
• Automated workflow
• The exercise
• Execution of AW for Skin sensitization
The Exercise

• In this exercise we will predict the skin sensitization potential for:
  o One chemical – Chlorpyrifos [CAS# 2921-88-2], which will be the “target” chemical;
  o Batch of chemicals - the first five compounds of the Skin sensitization ECETOC database, which will be the “target” chemicals.

• This prediction will be accomplished by using of the developed automated 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
• Automated workflow
• The exercise
• **Execution of AW for Skin sensitization**
Execution of AW for Skin sensitization

• Only three out of six Toolbox modules are used in a sequential workflow:
  o Input
  o Data Gap Filling
  o Report

The rest of the modules – Profiling, Data and Category definition are included as a part of the algorithm of the automated workflow.
Outlook

• Background
• Objectives
• Specific Aims
• Automated workflow
• The exercise

• Execution of AW for Skin sensitization—Example 1
  o Input
Input Overview

• This module provides the user with several ways 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 Single Chemical

• Chemical Name
• Chemical Abstract Services (CAS) number (#)
• Drawing chemical structure with or without additional defining of the composition
• Select from User List/Inventory/Databases
1. Click on **CAS#**;
2. Enter the CAS# of Chlorpyrifos (CAS **2921-88-2**);
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-dimenensional 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 to be retained for the subsequent workflow.
Outlook

• Background
• Objectives
• Specific Aims
• Automated workflow
• The exercise
• **Execution of AW for Skin sensitization– Example 1**
  o Input
  o **Data Gap Filling**
Data Gap Filling
Overview

• “Data Gap Filling” (DGF) module give access to five different data gap filling tools:
  o Read-across
  o Trend analysis
  o (Q)SAR models
  o Standardized workflow
  o Automated workflow

• Depending on the situation, the most relevant data gap mechanism should be chosen, taking into account the following considerations:
  o 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.
  o 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.
  o “(Q)SAR models” can be used to fill a data gap if no adequate analogues are found for a target chemical.
  o 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 type of workflows are represented on the next slide.
In this example, we will use the Automated workflow approach.
Data Gap Filling
Apply Automated workflow

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

Once the workflow is started, the **Workflow controller** appears. It 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

Skin sensitization and Skin sensitization ECETOC databases are selected. 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)
During the process, the user can see so called “Helpers” (1). This is a new functionality intended to help the user with specific information associated with analogues used in gap filling approach.
Data Gap Filling
Information messages and functionalities

The helpers provide different type of information, e.g.:

• warning messages, alerting the user:
  - for qualifiers availability
  - for analogue(s) with composition

• notification messages:
  - if the read-across prediction is obtained by more analogues than the default 5. This could be due to chemicals with equal descriptor values (logKow values)
Data Gap Filling
Information messages and functionalities

Functionality used to follow the progress of the active action or to cancel an active action when an Automated workflow is run.

- To see the progress of the active action click on the clock (1)
- To cancel an active action click on Stop button (2)
Data Gap Filling
Apply Automated workflow

The automated workflows follow the implemented logic in them and finish with accepted prediction.

When the prediction is accepted successfully, it appears on the matrix (1). In this case the predicted result (R:) corresponds to the experimental data (M:). Click on “OK” (2)
Data Gap Filling
Apply Automated 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 Automated workflow execution. After that the controller have to be closed by click on the close button (2).
The new organization of the Toolbox allows the Documents panel (1) to be consistent in each of the toolbar sections. This gives the possibility to the user to go back/forward through the subcategorization steps by moving up/down to the document tree. In this way the user can follow each step of the executed automated workflow.
Double click on the cell with the prediction (1) opens the *Data points* window. Here the user can see more information for their prediction. It is seen that only EC3 data is used for the prediction (2) and the prediction approach is read across executed by AW “Skin sensitization”.

---

**Data Gap Filling**

**Apply Automated workflow**
Data Gap Filling
Recap

The application of the automated workflow aims to facilitate the Toolbox user`s work limiting their activities. The AW for SS requires only user activities such as:

- specification of the endpoint
Outlook

• Background
• Objectives
• Specific Aims
• Automated workflow
• The exercise
• **Execution of AW for Skin sensitization – Example 1**
  o Input
  o Data Gap Filling
  o **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.
1. Go to the **Report** module; 2. Click on the cell with the prediction; 3. Click on **Create**
The user can customize the report content (1) and appearance (2). Generation of the reports happens by click on the Create report button (3).
• **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.

![Generated report files window](image-url)
Using of an automated workflow for predicting of skin sensitization potential is noted in the *Prediction report*. 

**Report**

Generated report files
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.
Outlook

- Background
- Objectives
- Specific Aims
- Automated workflow
- The exercise

- **Execution of AW for Skin sensitization— Example 2**
  - Input
Input
Ways of Entering a Chemical List

• Database
• Inventory
• List: Last used files/ From examples folder
Input Screen
Input database

1. Click on **Database**;
2. Select **Skin sensitization ECETOC**;
3. Click on **OK**.
All chemicals of the selected database are loaded on the data matrix. **Skin sensitization ECETOC** database consists of 39 compounds.
Outlook

• Background
• Objectives
• Specific Aims
• Automated workflow for
• The exercise

• **Execution of AW for Skin sensitization– Example 2**
  o Input
  o **Data Gap Filling**
1. Go to **Data Gap Filling**; 2. Click on **Automated** button; 3. Select **Skin sensitization**; 4. Confirm with **OK**
Data Gap Filling
Apply Automated workflow

1. Select the first combined endpoint – EC3 from LLNA and SMWN from GPMT assays; 2. Confirm with OK
Data Gap Filling
Apply Automated workflow

When the preferred endpoint is selected, a new dialog related to the application range of the workflow appears.

The user can define the exact range of chemicals, on which the automated workflow will be executed. The numbering of chemicals corresponds to their number on the data matrix.

The whole range is set as default.
Data Gap Filling
Apply Automated workflow
Data Gap Filling
Apply Automated workflow

1. Select to execute the automated workflow for the first five chemicals;
2. Confirm with OK
All chemicals in the selected range are predicted automatically in a row. The user can see the number of the running chemical out of all chemicals for prediction (1).
Data Gap Filling
Apply Automated workflow

• The workflow in a batch mode finishes when automated workflow is executed for each of the chemicals in the previously selected range.

• No messages for the individual predictions appear during the workflow.

• In some cases following the implemented logic in the workflow, no analogues are found or they are removed during the subcategorization process. Then the prediction result lacks (see on the next slide).
Data Gap Filling
Apply Automated workflow

There is no prediction
Data Gap Filling
Apply Automated workflow

If automated workflow in single mode is executed for such kind of chemicals, message proposing to try the standardized workflow appears.
Outlook

• Background
• Objectives
• Specific Aims
• Automated workflow
• The exercise
• **Execution of AW for Skin sensitization – Example 2**
  o Input
  o Data Gap Filling
  o **Report**
Prediction report and Data matrix report can be generated for each of the predictions.

It is necessary just to be positioned on the cell with the desired prediction and then click on the Create button.
1. Click on the prediction for 3rd chemical
2. Click on **Create**
Report
Generated report files

Prediction report
Data matrix report
Congratulations

• You have completed the tutorial on the automated workflow for skin sensitization data gap filling.
• You have been introduced to the automated workflow in a single and in a batch mode.
• You have now been introduced to the consecutive steps of the AW for SS and the rationale behind each step.
• Note proficiency comes with practice.