



Automated and Standardized workflows for predicting Acute aquatic toxicity and Skin sensitization

Laboratory of Mathematical Chemistry,
University “Prof. Assen Zlatarov” - Bourgas, Bulgaria

February 2020

Outlook

- Endpoints
- Specificities
- Components
- Executing module
- Algorithm of Ecotoxicological workflow
- Predicting Ecotoxicity by using the workflows - examples
- Algorithm of Skin sensitization workflow
- Predicting Skin sensitization by using the workflows - examples

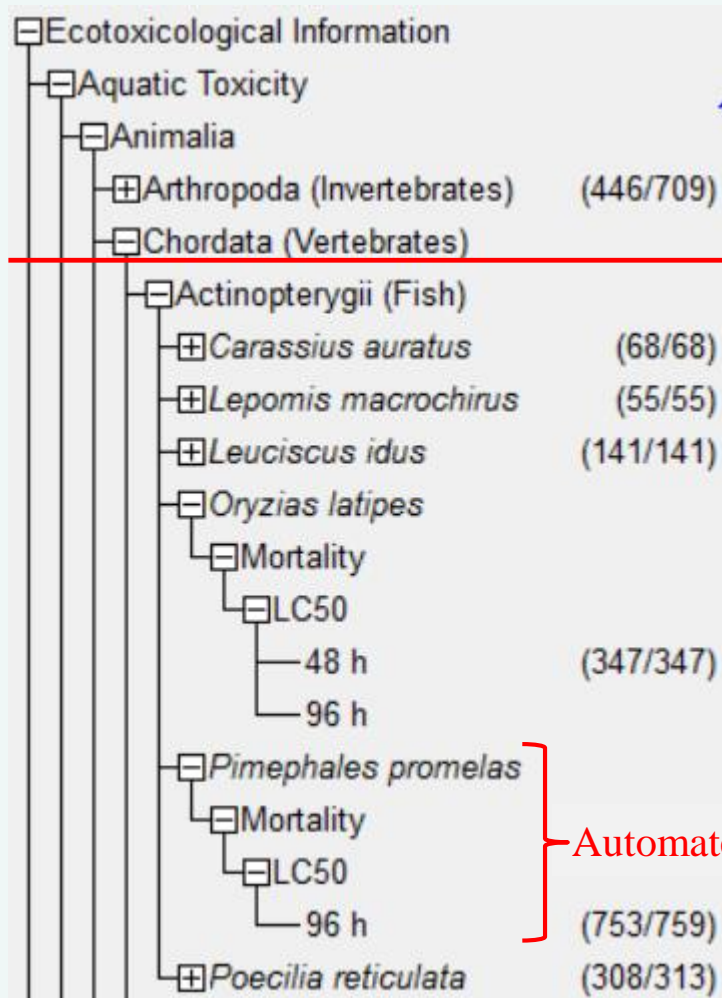
Outlook

- Endpoints
- Specificities
- Components
- Executing module
- Algorithm of Ecotoxicological workflow
- Predicting Ecotoxicity by using the workflows - examples
- Algorithm of Skin sensitization workflow
- Predicting Skin sensitization by using the workflows - examples

Automated and Standardized workflows in Toolbox

Endpoints

- Acute aquatic toxicity – Domain of application



- Sub-hazard – Aquatic toxicity
- Effect – Mortality, Intoxication, Immobilization, Growth, Population
- Endpoint – LC50;EC50; IC50
- Duration: 48 - 96 h
- Species:
 - all fish: *P. promelas*, *P. reticulata*, etc.
 - Crustaceans
 - Green algae

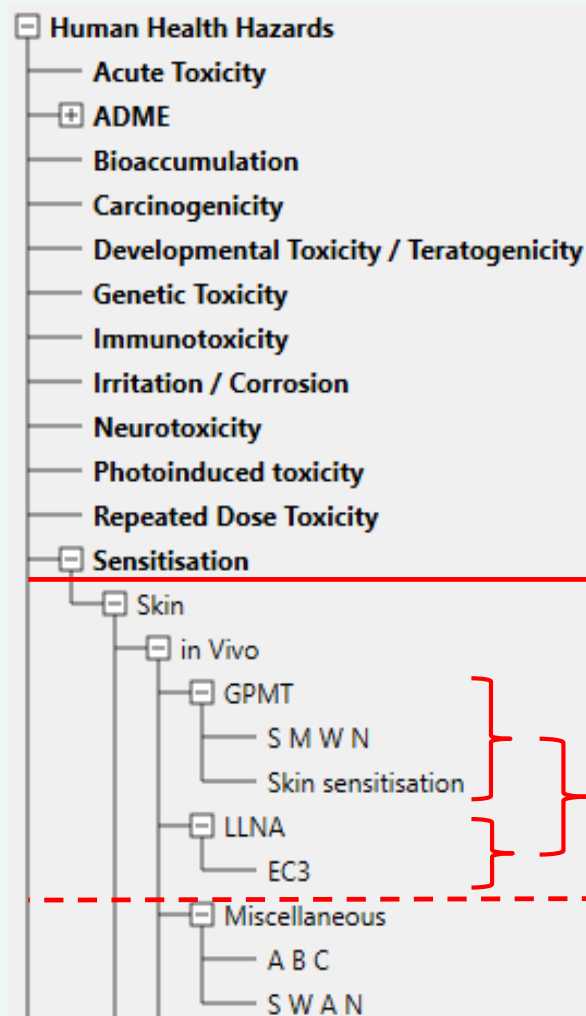
Standardized workflow

Automated workflow

Automated and Standardized workflows in Toolbox

Endpoints

- Skin sensitization – Domain of application



- Sub-hazard – Skin sensitization
- Method – *in vivo*
- Assays – LLNA + GPMT, LLNA only, GPMT only
- Endpoint – EC3 + SMWN converted in dichotomous scale

Outlook

- Endpoints
- Specificities
- Components
- Executing module
- Algorithm of Ecotoxicological workflow
- Predicting Ecotoxicity by using the workflows - examples
- Algorithm of Skin sensitization workflow
- Predicting Skin sensitization by using the workflows - examples

Automated and Standardized workflows in Toolbox

Specificities

Automated workflow (AW)

- On the fly data gap filling – the user is not able to control the workflow
- Input information – only target chemical ID
- Independency – application is not affected by the user activities (proceeding or subsequent)
- Batch mode execution

Standardized workflow (SW)

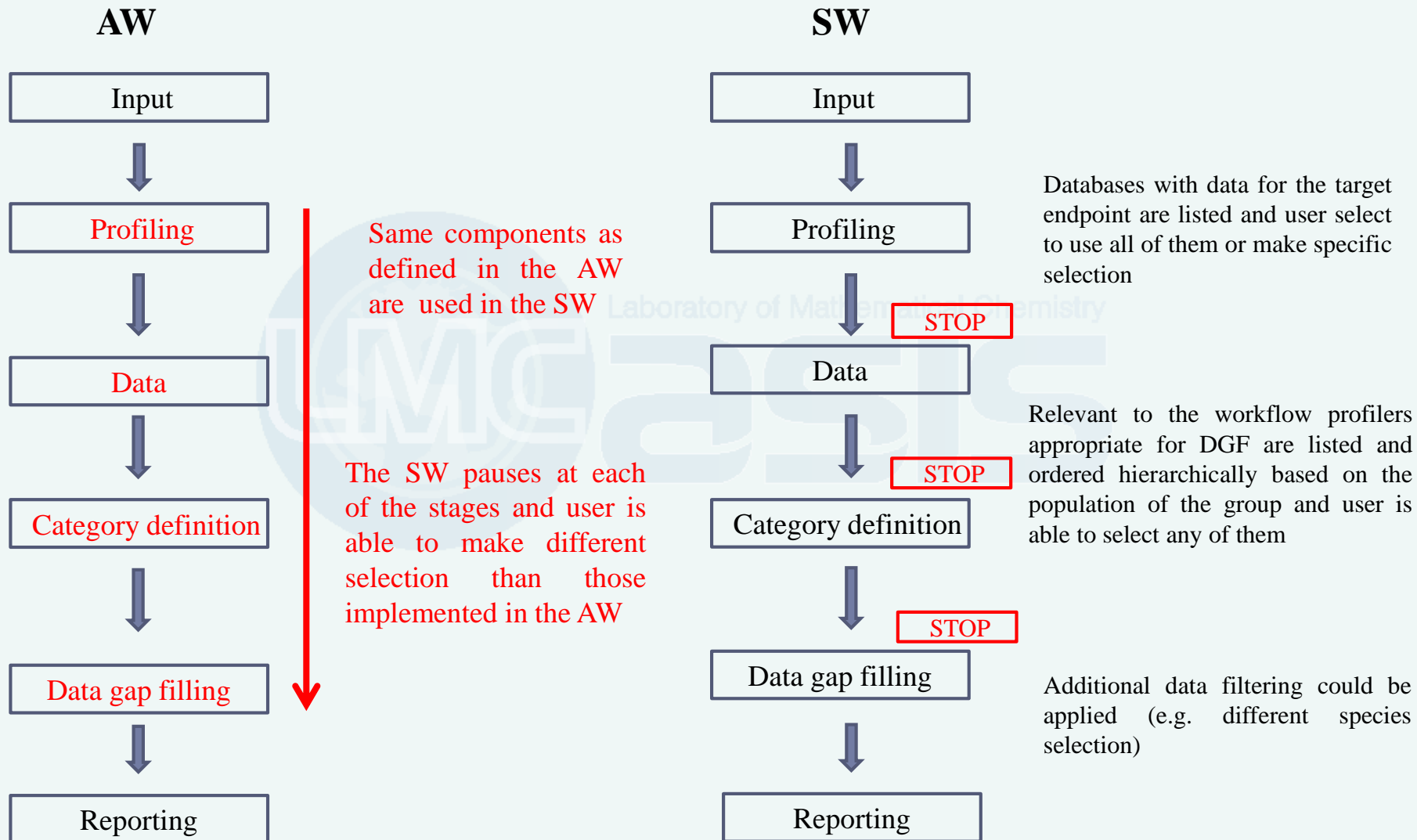
- Applicability domain of SW is expanded as compared to the AWs, including additional endpoints, effects, durations and species.
- The SWs follow same steps as in the AWs.
- The SWs stop at the each step of the workflows allowing the user to make different selection.

Outlook

- Endpoints
- Specificities
- **Components**
- Executing module
- Algorithm of Ecotoxicological workflow
- Predicting Ecotoxicity by using the workflows - examples
- Algorithm of Skin sensitization workflow
- Predicting Skin sensitization by using the workflows - examples

Automated and Standardized workflows in Toolbox

Components



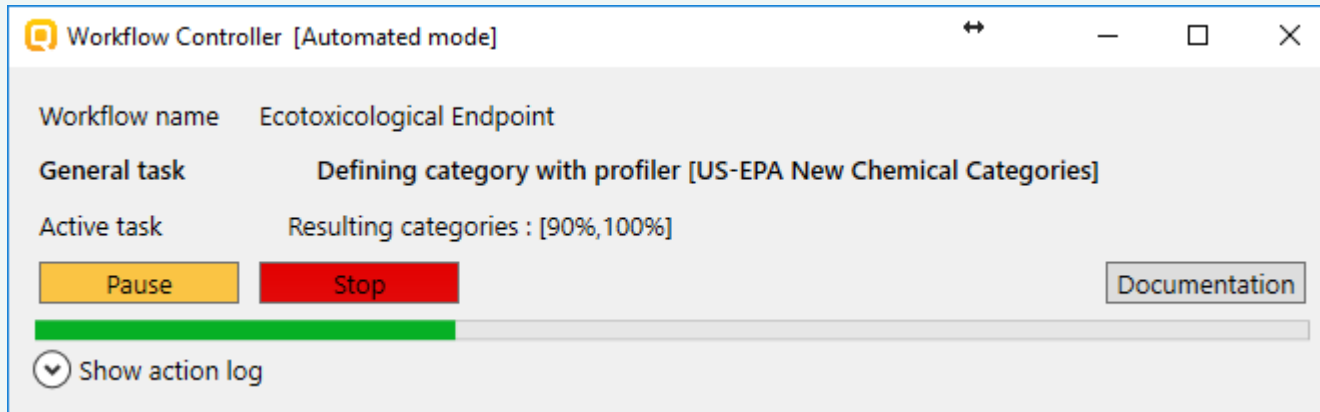
Outlook

- Endpoints
- Specificities
- Components
- Executing module
- Algorithm of Ecotoxicological workflow
- Predicting Ecotoxicity by using the workflows - examples
- Algorithm of Skin sensitization workflow
- Predicting Skin sensitization by using the workflows - examples

Automated and Standardized workflows in Toolbox

Executing module

- The AWs and SWs are controlled by the “*Workflow controller*”



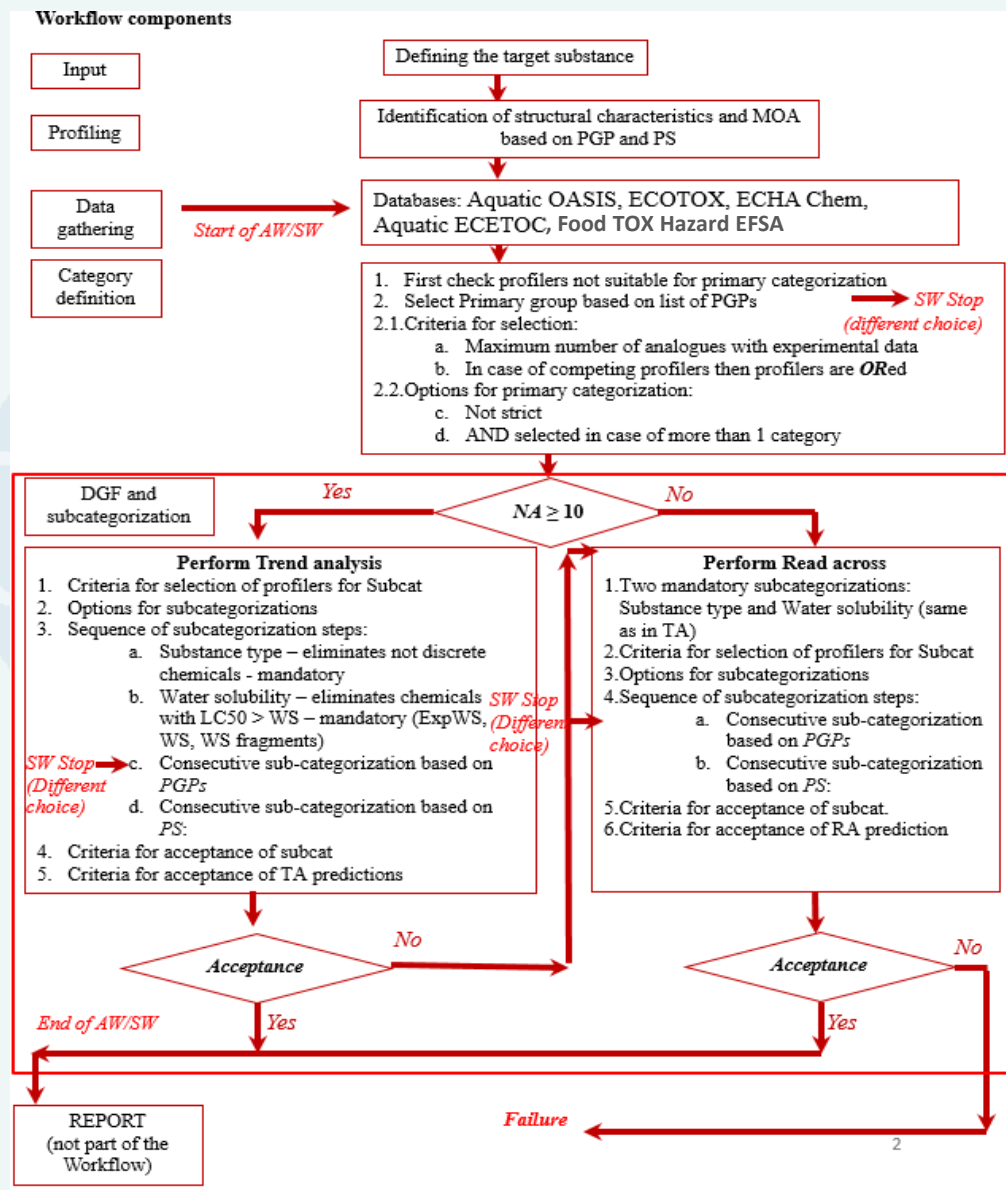
- The workflow is navigated by two main buttons:
 - *Continue/Pause* allowing to continue or pause the workflow
 - *Stop* which cancel (deactivate) the workflow.
- The algorithm of the current workflow could be seen by the *Documentation* button.
- Additionally, all actions that are done during the execution of the workflow are tracked down and could be seen in the “*Show action log*”.

Outlook

- Endpoints
- Specificities
- Components
- Executing module
- **Algorithm of Ecotoxicological workflow**
- Predicting Ecotoxicity by using the workflows - examples
- Algorithm of Skin sensitization workflow
- Predicting Skin sensitization by using the workflows - examples

Automated and Standardized workflows in Toolbox

Algorithm of Ecotoxicological workflow



Automated and Standardized workflows in Toolbox

Algorithm of Ecotoxicological workflow

▪ Databases

- Aquatic OASIS (used in AW)
- Aquatic ECETOX (used in AW)
- ECOTOX (used in AW)
- ECHA CHEM (used in AW)
- Aquatic Japan*

Laboratory of Mathematical Chemistry
OASIS

*Aquatic japan MoE is not used in the AW because there are no data for L(E)C50, 96h, Pimephales promelas, Mortality

Automated and Standardized workflows in Toolbox

Algorithm of Ecotoxicological workflow

▪ Profilers used for primary categorization:

- US EPA New Chemical Categories
- Acute aquatic toxicity classification by Verhaar (Modified)
- Acute aquatic toxicity MOA by OASIS
- Aquatic toxicity classification by ECOSAR
- Organic functional groups (OFG)
- Organic functional groups US-EPA,
- Organic functional groups, Norbert Haider

▪ Options for primary categorization:

- Maximum number of analogues with experimental data
- In case of competing profilers then profilers are **ORed**

Automated and Standardized workflows in Toolbox

Algorithm of Ecotoxicological workflow

▪ Data Gap Filling

- Trend analysis is the default approach
- Read across is applied if:
 - Prediction by Trend analysis is not acceptable, or
 - The number of analogues is < 10
- Gap filling and subcategorizations are sequence of logical operations (if, then), combined with criteria for acceptance.

Automated and Standardized workflows in Toolbox

Algorithm of Ecotoxicological workflow

▪ Subcategorization

- The aim is to increase the similarity of analogues with the target
- It is consecutive process of application of primary grouping profilers (PGPs) and profilers for subcategorization (PS)
- Hierarchy of application of PGPs and PS depends on the number of analogues they have collected
- Sub-categorization process is based on:
 - Sequence of subcategorization steps
 - Criteria for acceptance of subcategorization steps

PGPs – primary grouping profilers

PS – profilers for subcategorization

Automated and Standardized workflows in Toolbox

Algorithm of Ecotoxicological workflow

▪ Subcategorization

Sequence of subcategorization steps

1. *Substance type* – eliminates not discrete chemicals
2. *Water solubility* (WSKOWWIN + WATERNT)
– eliminates chemicals with $LC50 > WS$
3. Consecutive sub-categorization based on *PGPs*:
 - US EPA,
 - Verhaar,
 - MOA,
 - ECOSAR,
 - OFG (without nested)
4. Consecutive sub-categorization based on *PS*:
 - Protein binding (OASIS + OECD),
 - Chemical elements,
 - Str. Similarity* (remove all constituents with similarity <50%)

*Default options: Dice, Atom centered fragments

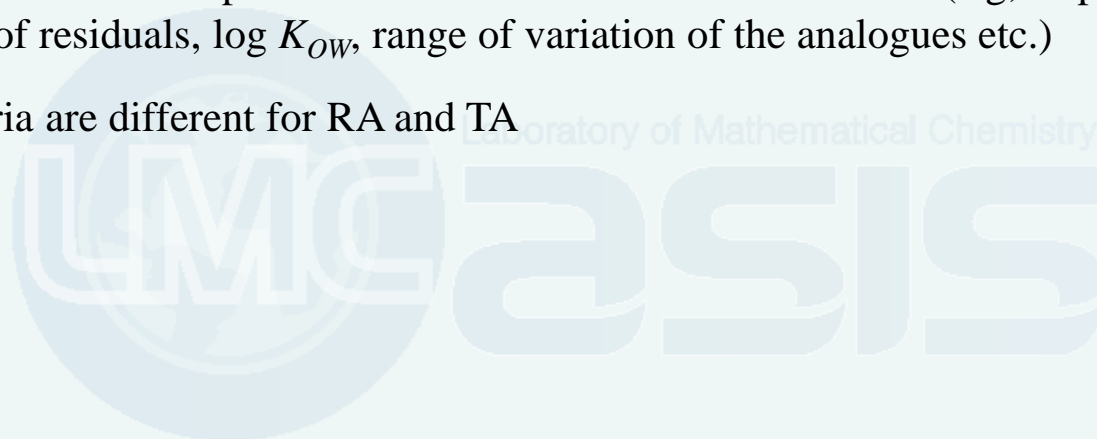
Automated and Standardized workflows in Toolbox

Algorithm of Ecotoxicological workflow

▪ Subcategorization

Criteria for acceptance of subcategorization step:

- Depends on the specific statistical and structural criteria (e.g, experimental error, 95% of residuals, $\log K_{ow}$, range of variation of the analogues etc.)
- Criteria are different for RA and TA



Automated and Standardized workflows in Toolbox

Algorithm of Ecotoxicological workflow

Criteria for acceptance of prediction

Trend analysis

After sub-categorization by all PGPs and PS

IF ($R^2 \geq 0.7$ and $NA \geq 10$)

THEN accept the prediction and generate report,

ELSE switch to Read across

Read-across

After sub-categorization by all PGPs and PS

IF *Interpolation* **AND** $LC50 \leq 2 \log \text{units}$ **OR** $\log K_{ow} \leq 2 \log \text{units}$ **AND** $NA \geq 5$

THEN accept prediction and proceed with Report

In case, criteria are not met, then no prediction is obtained

Definitions

R^2 – Correlation coefficient

Interpolation: $\log K_{ow}$ of the target should be within the range of $\log K_{ow}$ of analogues

$LC50 \leq 2$: for the 5 closest analogues the range of variation of $LC50$ is $\leq 2 \log \text{units}$

$\log K_{ow} \leq 2$: for the 5 closest analogues the range of variation of $\log K_{ow}$ is $\leq 2 \log \text{units}$

NA – Number of analogues

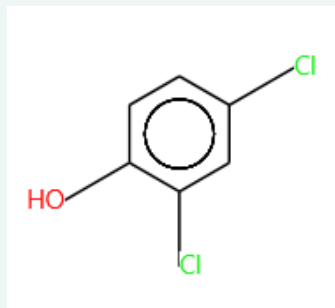
Outlook

- Endpoints
- Specificities
- Components
- Executing module
- Algorithm of Ecotoxicological workflow
- Predicting Ecotoxicity by using the workflows - examples
- Algorithm of Skin sensitization workflow
- Predicting Skin sensitization by using the workflows - examples

Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

Target chemical: CAS 120-83-2



Endpoint: LC50 or EC50

Effect: Mortality

Species: Actinopergyii (all fish) – illustrated in SW

Species: Pimephales promelas – illustrated in AW

Duration: 96h

Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

Location: The AWs and SWs are part of the DGF

The screenshot displays the QSAR Toolbox interface. The top menu bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The 'Data Gap Filling' menu is open, showing options for 'Gap Filling', 'Workflow', 'Trend analysis', 'Read across', '(Q)SAR', 'Standardized', and 'Automated'. The 'Workflow' option is highlighted with a red arrow and a label 'Automated workflow'. The 'Standardized' option is also highlighted with a red arrow and a label 'Standardized workflow'. The 'Data Gap Filling Settings' panel on the left shows the 'Only endpoint relevant' checkbox checked. The 'At this position:' section shows counts for QSARs (0), Automated workflows (0), and Standardized workflows (0). The 'In nodes below:' section also shows counts for QSARs (0), Automated workflows (0), and Standardized workflows (0). The main workspace displays a chemical structure of 2,4-dichlorophenol and a table with the following data:

Structure info	Parameters	Physical Chemical Properties	Environmental Fate and Transport	Ecotoxicological Information	Human Health Hazards

Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

Activation: The Standardized workflow is activated by clicking on the corresponding button

The screenshot displays the QSAR Toolbox software interface. The top toolbar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The 'Data Gap Filling' button is highlighted with a red dashed box. Below the toolbar, the 'Workflow' section contains 'Standardized' and 'Automated' buttons, with the 'Standardized' button highlighted by a red arrow and a red box. A red text label 'Standardized workflow' points to this button. On the left, the 'Documents' panel shows 'Document 1' with its CAS number. The 'Data Gap Filling Settings' panel on the bottom left shows a table of workflow counts:

At this position:	
QSARs	0
Automated workflows	0
Standardized workflows	0
In nodes below:	
QSARs	0
Automated workflows	0
Standardized workflows	0

A 'Select workflow' dialog box is open in the foreground, titled 'List with available standardized workflows'. It contains a 'Choices' section with two radio button options: 'Ecotoxicological Endpoint' (selected) and 'Skin sensitization'. The 'OK' button at the bottom right of the dialog is also highlighted with a red box.

Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

Selection of endpoint: The SW for Ecotoxicological endpoint needs the endpoint to be confirmed. There are 5 options for endpoint selection. In our case study we apply SW for:

Endpoint: EC50 or LC50 for Fish, duration: 96h

The screenshot displays the QSAR Toolbox software interface. The top toolbar includes icons for Input, Profiling, Data, Category definition, Data Gap Filling (highlighted with a red dashed box), and Report. Below the toolbar, the 'Gap Filling' and 'Workflow' tabs are visible. The 'Documents' panel on the left shows 'Document 1' with chemical identifiers: # [C: 1;Md: 0;P: 0] CAS: 120832. The 'Data Gap Filling Settings' panel on the bottom left shows 'Only endpoint relevant' checked, and 'At this position:' and 'In nodes below:' both set to 0 for QSARs, Automated workflows, and Standardized workflows. The main workspace shows a chemical structure and a 'Filter endpoint tree...' panel. A dialog box titled 'Selection of endpoint for execution of SW' is open, asking 'Which data do you want to use?'. The dialog lists five options under 'Aquatic toxicity':

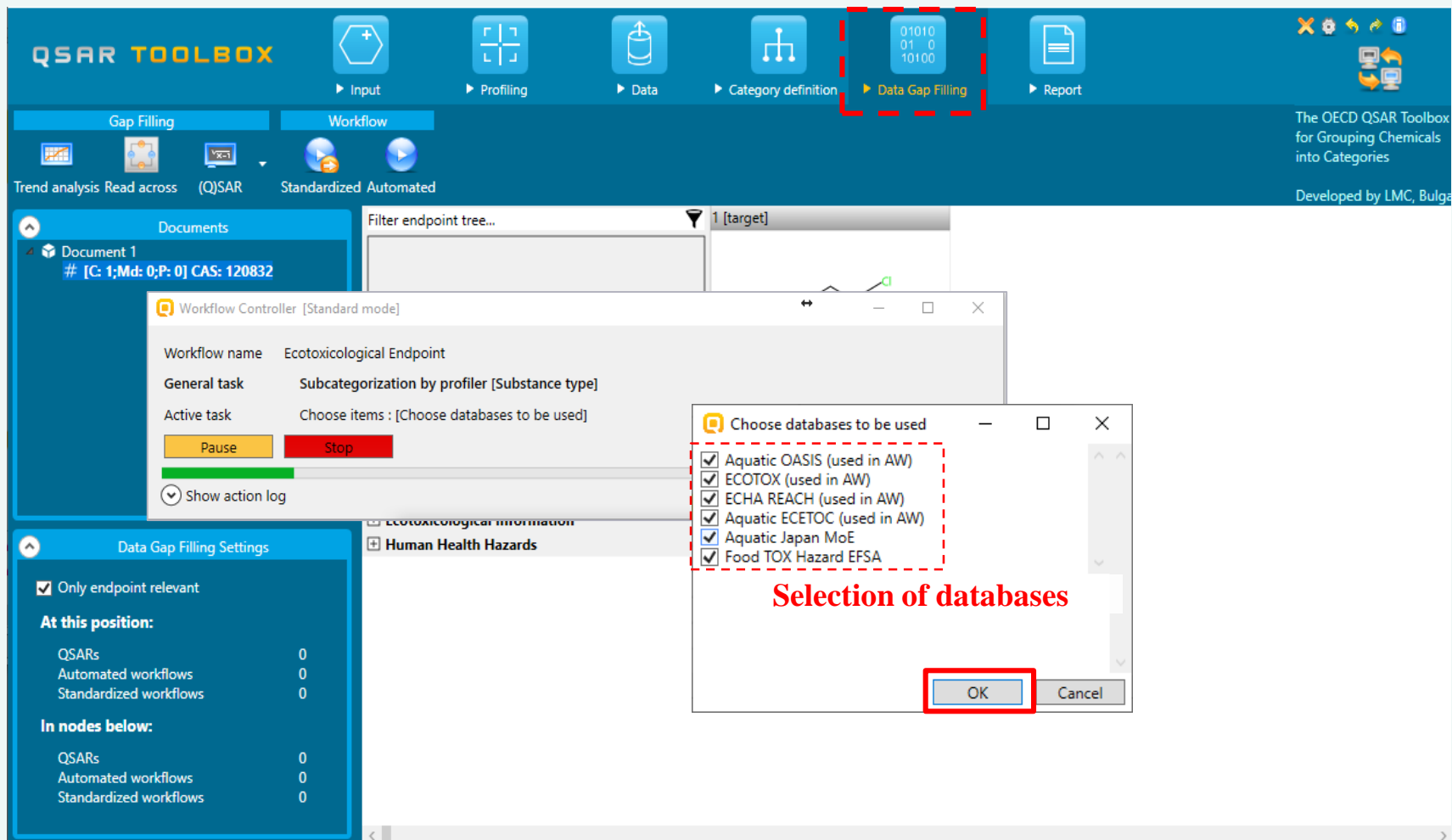
- ☐ Fish, LC50(EC50) at 96h for Pimephales promelas (mortality) (AW available)
- ☒ Fish, LC50(EC50) at 96h for Actinopterygii(mortality)
- ☐ Invertebrates, EC50 (LC50) at 48h for Branchiopoda(mortality,intoxication,immobilization)
- ☐ Algae, EC50(IC50,LC50) at 72-96h for Chlorophyceae (population)
- ☐ Algae, EC50(IC50,LC50) at 72-96h for Chlorophyceae (growth)

The 'OK' button in the dialog box is highlighted with a red box. The 'Documentation' button is also visible at the bottom left of the dialog.

Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

Selection of databases: Six databases are available. All of them are selected in this example

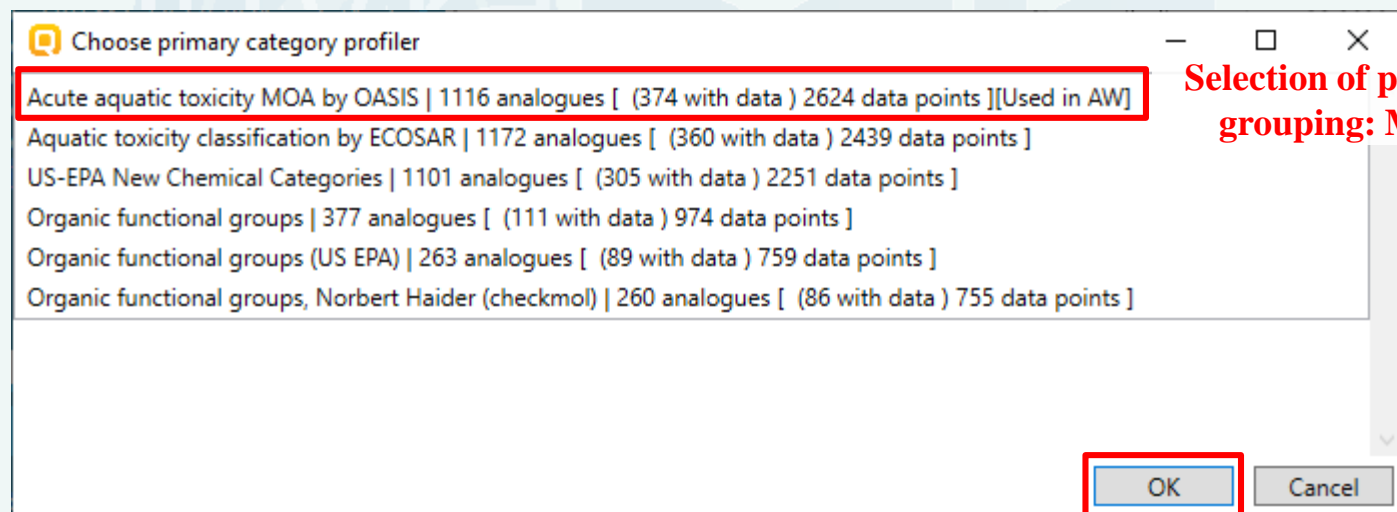


Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

Category definition:

- Categories are hierarchically order depending on the number of analogues with data
- The most populated category is used in the AW
- In SW the user is able to make different selection
- In this example we apply Acute aquatic toxicity MOA



Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

Data Gap Filling: Before entering to DGF, the workflow allows filtering of duration; endpoint and unit

Filtering of data: keep EC50 and LC50 data

Possible data inconsistency

Metadata

- Duration
 - ☒ 96 h (374 chemicals; 2624 data)
- Effect
 - ☒ Mortality (374 chemicals; 2624 data)
- Endpoint
 - ☒ EC50 (52 chemicals; 217 data)
 - ☒ LC50 (370 chemicals; 2407 data)
- Kingdom
 - ☒ Animalia (animals) (374 chemicals; 2624 data)
- Native scale/unit
 - ☒ (1 chemicals; 1 data)
 - ☒ % saturation concentration (2 chemicals; 2 data)
 - ☒ % v/v (1 chemicals; 3 data)
 - ☒ M (2 chemicals; 2 data)
 - ☒ mM (1 chemicals; 2 data)
 - ☒ mg/L (252 chemicals; 853 data)
 - ☒ mmol/L (4 chemicals; 4 data)

Required unit/scale

Select scale/unit to use

- ☐ M [2 native data and 2395 converted]
- ☐ g/m³ [0 native data and 2397 converted]
- ☐ log(1/mol/L) [0 native data and 2397 converted]
- ☐ mM [2 native data and 2395 converted]
- ☐ mg/L [853 native data and 1544 converted]

Converted data

- 2 from scale/unit M
- 2 from scale/unit mM
- 853 from scale/unit mg/L
- 4 from scale/unit mmol/L

Chemicals 363/374; Data 2397/2624

OK Cancel

Workflow Controller [Standardized mode]

Workflow name: Ecotoxicological Endpoint

General task: Defining category with pre...

Active task: Performing gap filling, appro...

Pause Stop

Show action log

Only endpoint relevant

At this position:

Category	Count
QSARs	0
Automated workflows	0
Standardized workflows	0

In nodes below:

Category	Count
QSARs	0
Automated workflows	0
Standardized workflows	0

Documents

- Document 1
 - # [C: 1;Md: 0;P: 0] CAS: 120832
 - [C: 1101;Md: 2251;P: 0] Phenols (Acute
 - [C: 1172;Md: 2439;P: 0] Phenols (Aquat
 - [C: 377;Md: 974;P: 0] Aryl<AND>Aryl h
 - [C: 263;Md: ...
 - [C: 260;Md: ...
 - [C: 1116;Md: ...

Filter endpoint tree...

Structure

Sediment Toxicity

Terrestrial Toxicity

Human Health Hazard

Profiling

Predefined

Substance type

1116

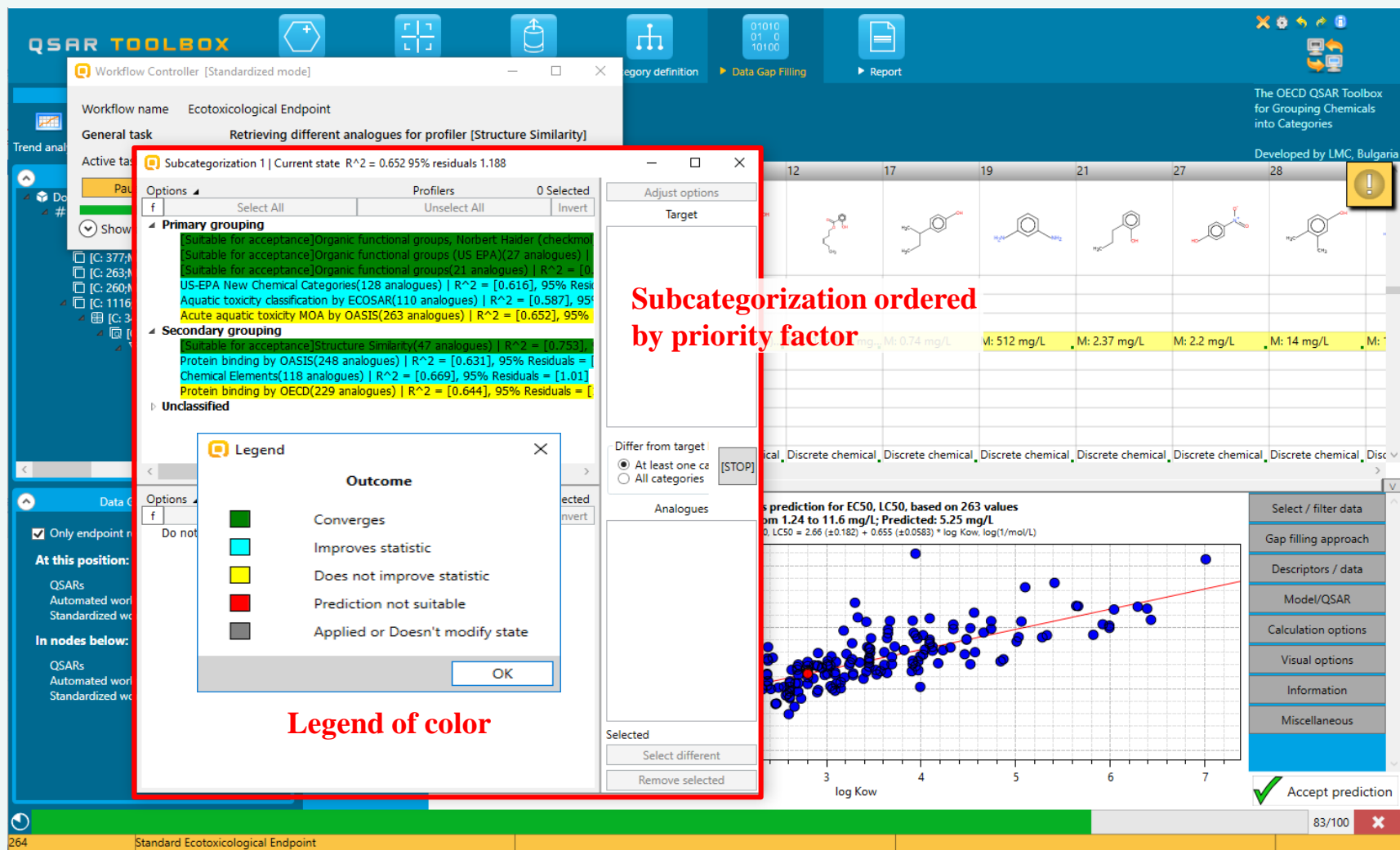
Standard Ecotoxicological Endpoint

50/100

Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

Data Gap Filling: Subcategorization dialogue provides the list with relevant profilers highlighted in different color and order by priority factor



Automated and Standardized workflows in Toolbox

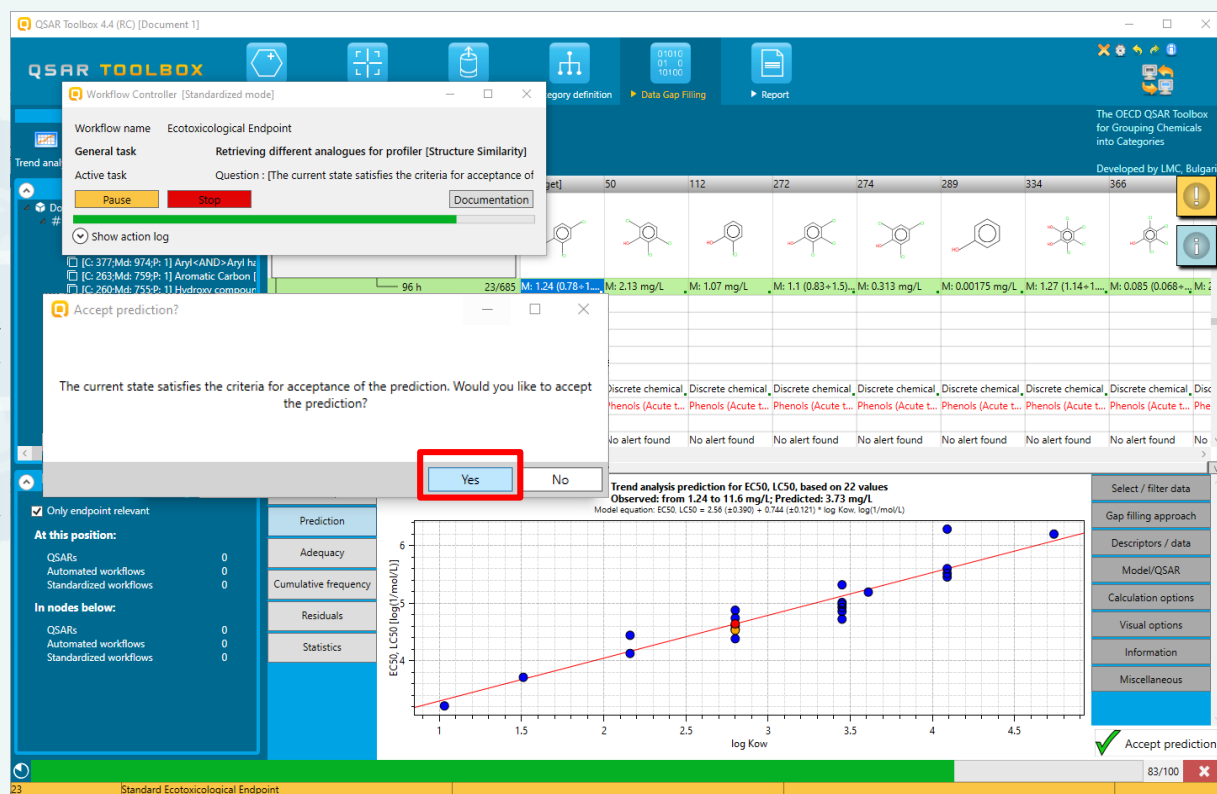
SW for Ecotoxicological endpoint - example

Subcategorization: The applied subcategorizations are as follows:

- ✓ Substance type (mandatory – automatically applied)
- ✓ Filter by WS (mandatory – automatically applied)
- ✓ OFG (USEPA)
- ✓ US-EPA Chemical categories

Note: After each subcategorization the workflow controller wait for activity by the user – “Continue” button needs to be clicked.

Once the prediction satisfy the acceptance criteria a message appears.



Predicted LC50(EC50) for all fish of 2,4-dichlorophenol is 3.73 mg/L (toxic) after application of SW for Ecotox

Automated and Standardized workflows in Toolbox

SW for Ecotoxicological endpoint - example

- Report:**
- The report section is not part of the workflows
 - Report could be generated once the prediction is accepted

The screenshot displays the QSAR Toolbox software interface. The main window is titled "Prediction of EC50, LC50 for Dichlorophenol" and shows the "Prediction summary" section. A red box highlights the "Prediction summary" section, which includes the following information:

- Predicted endpoint:** EC50, LC50; Mortality; No species specified; 96 h; No guideline specified
- Predicted value:** 3.73 (from 1.13 to 12.3)
- Unit/scale:** mg/L
- Data gap filling method:** Trend analysis, Standardized workflow for Ecotoxicological Endpoint
- Summary:** manually editable field
- Not provided by the user

A chemical structure of Dichlorophenol is shown. The structure is a benzene ring with two chlorine atoms (Cl) and one hydroxyl group (OH). The SMILES string is Oc1ccc(Cl)cc1Cl.

The "Generated report files" dialog box is open, showing the following files:

- Prediction report
- Category report
- Data matrix
- PDF file containing the prediction report

The "Create report" button is highlighted with a red box. A red arrow points to the "Prediction report" file in the list.

Automated and Standardized workflows in Toolbox

AW for Ecotoxicological endpoint - example

Activation: The Automated workflow is activated by clicking on the corresponding button

The screenshot displays the QSAR Toolbox software interface. The top toolbar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The 'Data Gap Filling' button is highlighted with a red dashed box. Below the toolbar, the 'Workflow' section shows 'Standardized' and 'Automated' options, with 'Automated' highlighted by a red box and an arrow pointing to it. A red text box labeled 'Automated workflow' is positioned over the 'Automated' button. The left sidebar shows a 'Documents' list with 'Document 1' selected, displaying its CAS number: 120832. The bottom left panel shows 'Data Gap Filling Settings' with a table of workflow counts:

At this position:	
QSARs	0
Automated workflows	0
Standardized workflows	0
In nodes below:	
QSARs	0
Automated workflows	0
Standardized workflows	0

A 'Select workflow' dialog box is open in the center, titled 'List with available workflows'. It contains a 'Choices' section with three radio button options: 'Ecotoxicological Endpoint' (selected), 'Skin sensitization', and 'Skin sensitization for defined approaches'. The 'OK' button is highlighted with a red box.

Automated and Standardized workflows in Toolbox

AW for Ecotoxicological endpoint - example

Selection of endpoint: The AW for Ecotoxicological endpoint needs the endpoint to be confirmed.

Endpoint: EC50 or LC50 for Pimephales promelas, duration: 96h

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, there are tabs for Gap Filling and Workflow. The left sidebar shows a list of documents, with 'Document 1' selected, displaying its CAS number (120832). The main workspace shows a chemical structure and a list of endpoints. A dialog box titled 'Select one' is open, asking 'Which data do you want to use?' for 'Aquatic toxicity'. The selected option is 'Fish, LC50(EC50) at 96h for Pimephales promelas (mortality)'. The 'OK' button is highlighted with a red rectangle.

Confirmation of endpoint for execution of AW

Automated and Standardized workflows in Toolbox

AW for Ecotoxicological endpoint - example

The following steps are applied automatically:

- Selection of databases
- Selection of primary group
- Sequence of subcategorizations
- Acceptance of the prediction

The screenshot displays the QSAR Toolbox software interface. The top toolbar contains icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below the toolbar, there are tabs for Gap Filling and Workflow. The main window shows a 'Filter endpoint tree...' on the left, a 'Structure' panel in the center, and a 'Workflow Controller' dialog box on the right. The 'Workflow Controller' dialog box shows a 'Finished workflow' and a 'Success' message: 'Prediction accepted successfully'. A red box highlights the 'Success' message. The background shows a chemical structure and a list of chemical properties.

The prediction is accepted if answer the criteria of gap filling approach

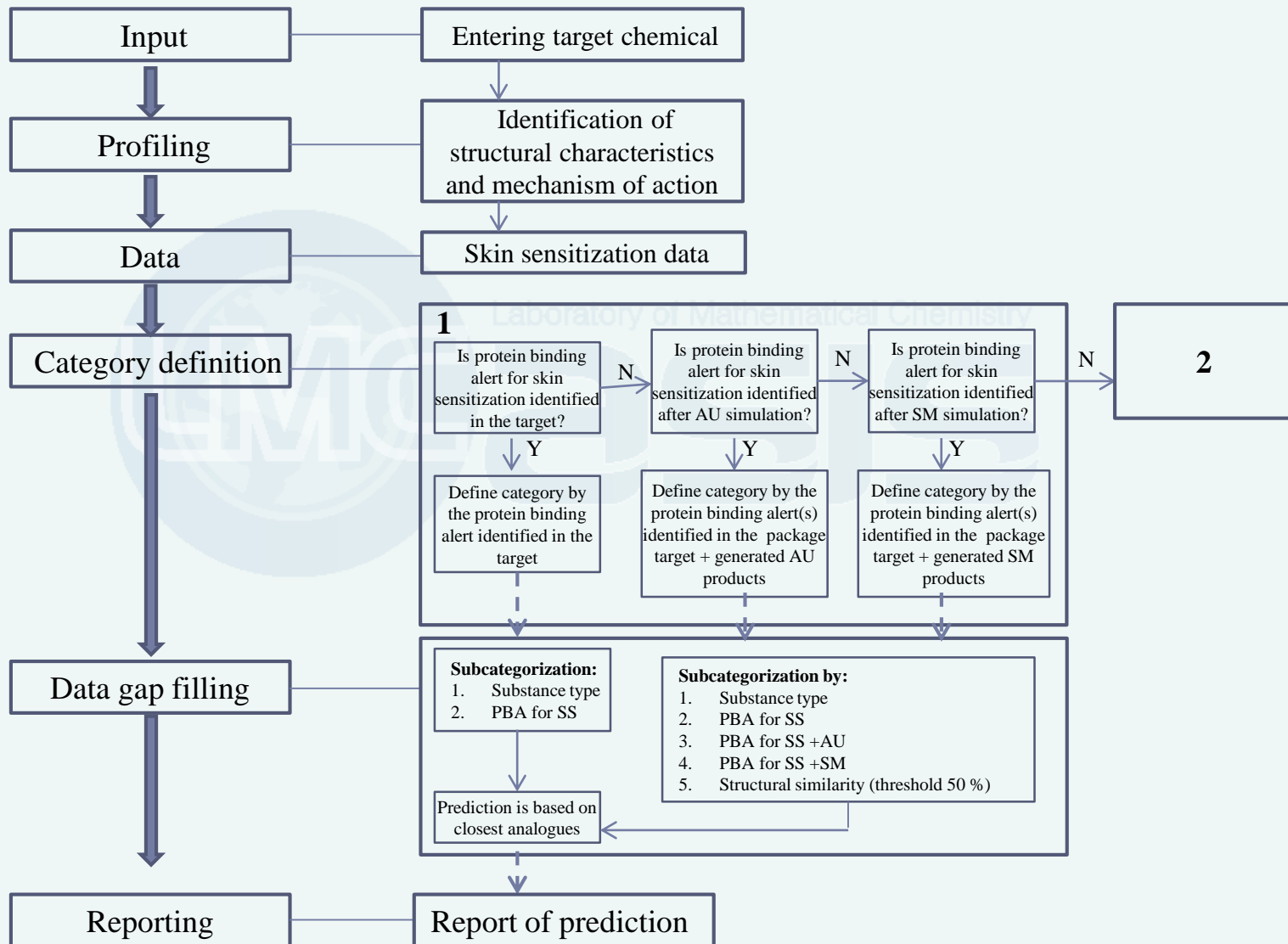
Outlook

- Endpoints
- Specificities
- Components
- Executing module
- Algorithm of Ecotoxicological workflow
- Predicting Ecotoxicity by using the workflows - examples
- **Algorithm of Skin sensitization workflow**
- Predicting Skin sensitization by using the workflows - examples

Automated and Standardized workflows in Toolbox

Algorithm of Skin sensitization workflow

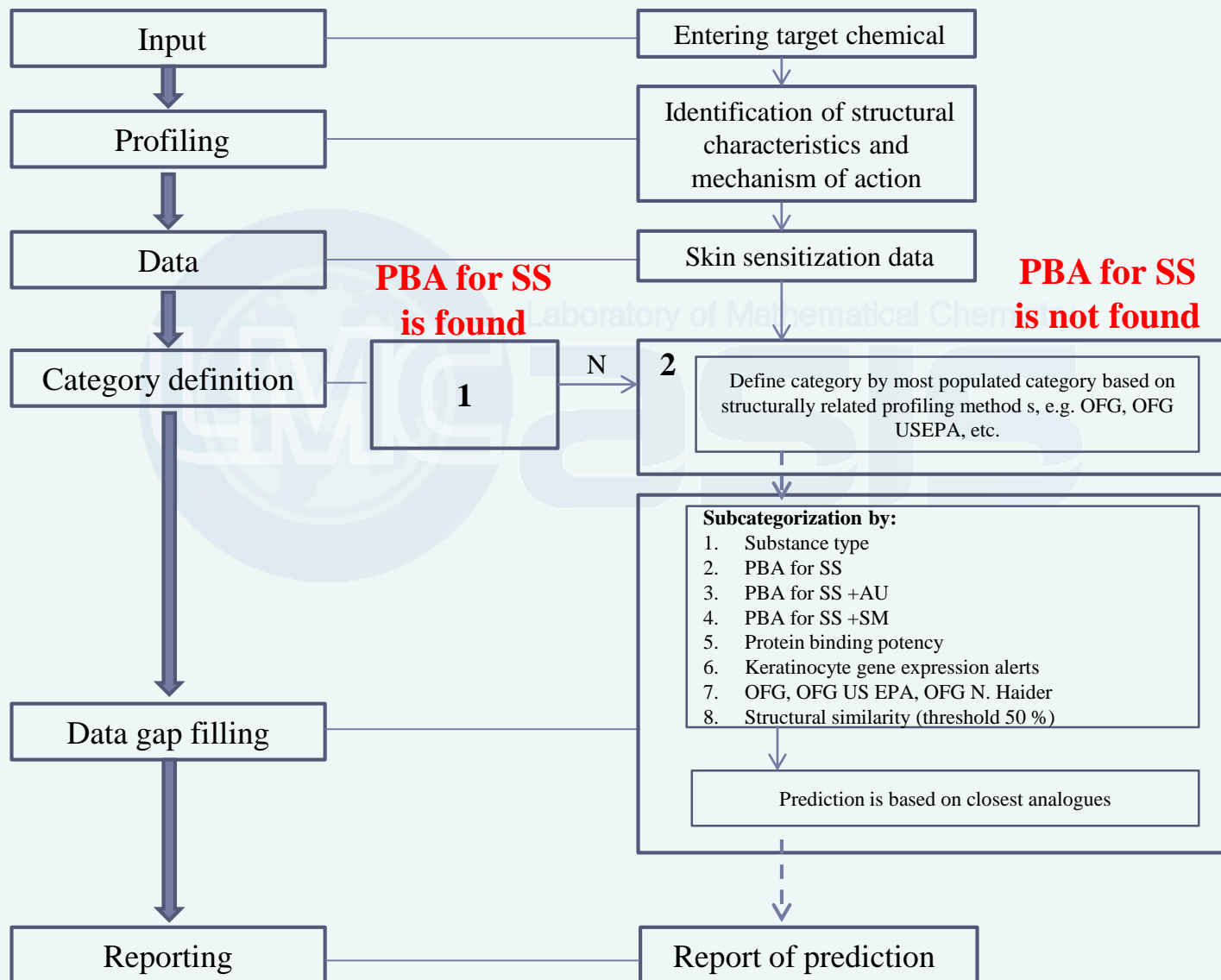
Part 1



Automated and Standardized workflows in Toolbox

Algorithm of Skin sensitization workflow

Part 2



Automated and Standardized workflows in Toolbox

Algorithm of Skin sensitization workflow

▪ Databases

- Skin sensitization
- Skin sensitization ECETOC
- REACH Skin sensitisation database (normalized)

▪ SS predictions are based on LLNA and GPMT exp. data.

- ✓ LLNA – EC3, %
- ✓ GPMT – Positive; Negative
- ✓ LLNA and GPMT – Positive; Negative

Laboratory of Mathematical Chemistry

Automated and Standardized workflows in Toolbox

Algorithm of Skin sensitization workflow

- **Profilers for primary grouping**
 - US EPA New Chemical Categories
 - Aquatic toxicity classification by ECOSAR
 - Protein binding alerts for Skin sensitization effect
 - Organic Functional Groups
 - Organic Functional Groups by US EPA
 - Organic Functional Groups by Norbert Haider
- **Abiotic and biotic activation of chemical is accounted for by application of respective Autoxidation (AU) and Skin metabolism (SM) simulators**

Automated and Standardized workflows in Toolbox

Algorithm of Skin sensitization workflow

▪ Category definition

- If protein binding alert for skin sensitization (**PBA for SS**) is identified in the target structure then the primary category is based on this alert
- If **PBA for SS** is identified after AU or SM simulation then the primary category is defined accounting the metabolic simulation
- If **more than one PBA for SS** are identified in the parent structure or in the generated metabolites, then:
 - the category is defined based on all available PBA as presented in the target structure
 - In case no analogues found, the selection of alert is based on the criteria for **reliability** of alerts, i.e. most reliable alert is selected (see next slide)
- If **No PBA for SS** is identified in the parent structure and in the generated metabolites, then the primary category is defined on global molecular features by using:
 - *Organic Functional Groups* (OFG) **OR**
 - *OFG USEPA* **OR**
 - *OFG N. Haider* **OR**
 - *Acute aquatic classification by ECOSAR* **OR**
 - *US-EPA New Chemical categories* **OR**

In this case, the primary category will be formed based on the profiler leading to **the largest group** of analogues.

Automated and Standardized workflows in Toolbox

Algorithm of Skin sensitization workflow

▪ Data gap filling

- Prediction is based on up to five closest analogues with respect to logKow
- Read across is applied as default gap filling approach
- Specific subcategorizations are applied depending on the profiling result and subsequent primary group formation (see next slide for more information)

If no analogues with data are found, then no prediction is obtained after application of the workflows

Outlook

- Endpoints
- Specificities
- Components
- Executing module
- Algorithm of Ecotoxicological workflow
- Predicting Ecotoxicity by using the workflows - examples
- Algorithm of Skin sensitization workflow
- Predicting Skin sensitization by using the workflows - examples

Automated and Standardized workflows in Toolbox

SW for Skin sensitization endpoint - example

Example for the Standardized workflow:

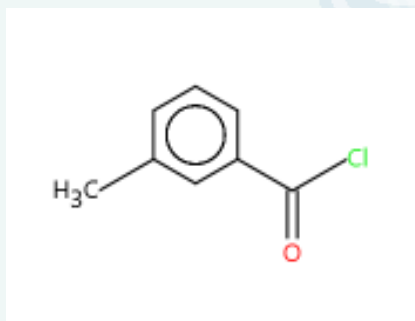
Endpoint: EC3

Type of method: in vivo

Assay: LLNA

Organ: Skin

Target chemical: CAS 1711-06-4



Example for the Automated workflow:

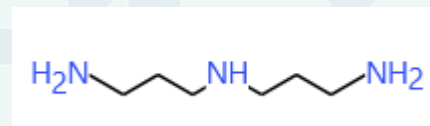
Endpoint: EC3 and Skin sensitization

Type of method: in vivo

Assay: LLNA and GPMT

Organ: Skin

Target chemical: CAS 56-18-8



Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

Location: The AWs and SWs are part of the DGF

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes 'Gap Filling', 'Workflow', 'Trend analysis', 'Read across', '(Q)SAR', 'Standardized', and 'Automated'. The 'Workflow' menu is open, showing 'Standardized' and 'Automated' options. A red arrow points from the 'Automated workflow' label to the 'Automated' option in the 'Workflow' menu. Another red arrow points from the 'Standardized workflow' label to the 'Standardized' option in the 'Workflow' menu. The 'Data Gap Filling' button is highlighted with a red box. The main workspace shows a chemical structure of 4-methylbenzoyl chloride (CC(=O)c1ccc(C)cc1) and a list of endpoints: Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards. The 'Data Gap Filling Settings' panel on the left shows 'Only endpoint relevant' checked, and counts for 'At this position' and 'In nodes below' for QSARs, Automated workflows, and Standardized workflows.

Automated workflow

Standardized workflow

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (Q)SAR Standardized Automated

Documents

Document 1
[C: 1;Md: 0;P: 0] CAS: 1741064

Filter endpoint tree... 1 [target]

Structure info
Parameters
Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information
Human Health Hazards

Data Gap Filling Settings

☒ Only endpoint relevant

At this position:

QSARs	0
Automated workflows	0
Standardized workflows	0

In nodes below:

QSARs	0
Automated workflows	0
Standardized workflows	0

Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

Activation: The SW is activated by clicking on the corresponding button

Example 1: Target chemical has protein binding alert

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (Q)SAR Standardized Automated

Document 1
[C: 1;Md: 0;P: 0] CAS: 1741064

Filter endpoint tree... 1 [target]

Structure info
Parameters
Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information
Human Health Hazards

Data Gap Filling Settings

☒ Only endpoint relevant

At this position:

QSARs	0
Automated workflows	0
Standardized workflows	0

In nodes below:

QSARs	0
Automated workflows	0
Standardized workflows	0

Standardized workflow

Chemical structure: CC1=CC=C(C=C1)C(=O)Cl

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulg.

Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

Selection of endpoint: Three options for data usage are provided: EC3, GPMT and EC3/GPMT

Example 1: Target chemical has protein binding alert

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options: Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, a sub-menu shows Trend analysis, Read across, (Q)SAR, and Standardized Automated. The main workspace shows a 'Filter endpoint tree...' window with a search filter '1 [target]' and a chemical structure of a benzene ring with a chlorine atom and a carbonyl group. A 'Select workflow' dialog box is open, showing a list of choices: 'Ecotoxicological Endpoint', 'Skin sensitization' (selected), and 'Skin sensitization for defined approaches'. A red box highlights the 'Skin sensitization' option, and a red arrow points to it. A second dialog box, 'Select one', is also open, showing the question 'Which data do you want to use?' with three radio button options: 'EC3 from LLNA or Skin sensitisation from GPMT assays (AW available)', 'Skin sensitisation from GPMT assay (AW available)', and 'EC3 from LLNA assay (AW available)' (selected). A red text overlay on the right side of the image reads: 'List with endpoints and experimental data for Skin sensitization: Select EC3 from LLNA'. The bottom left corner shows 'Data Gap Filling Settings' with a checkbox for 'Only endpoint relevant' and a table of workflow counts.

List with implemented workflows: Select “Skin sensitization”

List with endpoints and experimental data for Skin sensitization: Select EC3 from LLNA

Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

Selection of databases: Three databases are available

Example 1: Target chemical has protein binding alert

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options: Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, a secondary bar shows Trend analysis, Read across, (Q)SAR, and Standardized Automated. The left sidebar contains a 'Documents' panel with 'Document 1' (CAS: 1711064) and a 'Data Gap Filling Settings' panel with checkboxes for 'Only endpoint relevant' and 'At this position: QSARs, Automated workflows, Standardized workflows'. The main workspace shows a 'Filter endpoint tree...' with a chemical structure of 4-chlorobenzaldehyde (O=Cc1ccc(Cl)cc1). A red text overlay 'Workflow controller is activated' is positioned over the main workspace. A 'Workflow Controller [Standardized mode]' dialog box is open, showing 'Workflow name: Skin sensitization', 'General task: Subcategorization by profiler [Substance type]', and 'Active task: Choose items : [Choose databases to be used]'. It includes 'Pause' and 'Stop' buttons and a 'Show action log' checkbox. A 'Choose databases to be used' dialog box is also open, listing three checked databases: 'Skin Sensitization (used in AW)', 'Skin sensitization ECETOC (used in AW)', and 'REACH Skin sensitisation database (normalised) (used in AW)'. A red text overlay 'List with databases' is placed over this dialog, and the 'OK' button is highlighted with a red rectangle.

Workflow controller is activated

List with databases

Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

Category definition: Primary group is defined based on the alert found in the target

Example 1: Target chemical has protein binding alert

Explanation for: Protein binding alerts for skin sensitization by OASIS -> Acylation -> Direct acylation involving a leaving group -> (Thio)Acyl and (thio)carbamoyl halides and cyanides

Categories

Filter:

- Protein binding alerts for skin sensitization by OASIS
 - Acylation
 - (Thio)carbamoylation of protein nucleophiles
 - Isocyanates, Isothiocyanates
 - Acyl transfer via nucleophilic addition reaction
 - Carbodiimides
 - Direct acylation involving a leaving group
 - (Thio)Acetates
 - (Thio)Acyl and (thio)carbamoyl halides and cyanides**
 - Anhydrides (sulphur analogues of anhydrides)
 - Azlacones and unsaturated lactone derivatives
 - Carbamates
 - Diacyl peroxides, anhydrides (sulphur analogues of diacyl)
 - N-Acylloxysuccinimides
 - N-Carbonyl heteroaryl amines
 - N-Carbonylsulfonamides
 - N-Haloacylamides
 - Phosphonyl halides or cyanides
 - Sulphonyl halides or cyanides
 - Thiosulfonates

[84] (Thio)Acyl and (thio)carbamoyl halides and cyanides

Definition Properties Training Set Literature MetaInfo Table Scheme

Mechanistic Domain: Acylation

Mechanistic Alert: Direct acylation involving a leaving group

Structural Alert: (Thio)Acyl and (thio)carbamoyl halides, cyanides, azides, etc.

The chemical is a strong sensitizer as a result of Protein acylation by (thio)acyl halides and (thio)carbamoyl derivatives:

$$\text{R}-\text{C}(=\text{O})-\text{Hal} \xrightarrow{\text{Pr}-\text{NH}_2} \text{R}-\text{C}(=\text{O})-\text{NH}-\text{Pr} + \text{H}-\text{Hal}$$

Hal = F, Cl, Br, I
R = any C

SMARTS

[#6][#6](=[S])

View mode:

R¹ = C (sp³), N (sp³)
R² = F, Cl, Br, I, CN

Map 1

[0] Structure Query Metabolism

Contents

- Queries

Complex search options

- ☒ Exact connectivity
- ☐ Ignore stereo information
- ☐ Exact match

Queries execution mode All

Mapping

- ☒ Unique mappings

Max maps 1000

Left click on any marked atom to explore

Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

Category definition: Primary group is defined based on the alert found

Example 1: Target chemical has protein binding alert

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Trend analysis Read across (Q)SAR Standardized Automated

Documents

Document 1

[C: 1-Md: 0:P: 0] CAS: 1711064

[C: 17;Md: 0:P: 0] Acylation >> Direc

Filter endpoint tree...

1 [target] 2 3 4 5 6 7 8

Structure

Ecotoxicological Information

Workflow Controller [Standardized mode]

Workflow name Skin sensitization

General task Defining category with profiler [Protein binding alerts for skin sensitization by OASIS]

Active task Active alerts found

Pause Stop

Show action log

Grouping results

17 chemical(s) found.

Photoinduced toxicity

Repeated Dose Toxicity

Sensitisation

Skin

in Vivo

LLNA

EC3

ToxCast

Toxicity to Reproduction

Toxicokinetics, Metabolism and Distribution

Profiling

Predefined

Substance type

Discrete chemical

At this position:

QSARs 0

Automated workflows 0

Standardized workflows 0

In nodes below:

QSARs 0

Automated workflows 0

Standardized workflows 0

Standard Skin sensitization

SW for Skin sensitization - examples

Example 1: Target chemical has protein binding alert



Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

Data Gap Filling: Relevant to the workflow profilers are provided and colored

Example 1: Target chemical has protein binding alert

The screenshot displays the QSAR Toolbox interface. The top navigation bar includes buttons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The 'Data Gap Filling' tab is active. On the left, the 'Subcategorization' window is open, showing a list of profilers. A red box highlights the 'Protein binding alerts for skin sensitization by OASIS' profiler under the 'Unclassified' category. A red arrow points to this profiler with the text 'The rest of profilers are also here and can be used to subcategorize'. The background shows a grid of chemical structures and a scatter plot of log Kow values. The bottom status bar indicates 'Standard Skin sensitization' and '80/100'.

QSAR TOOLBOX

Input Profiling Data Category definition Data Gap Filling Report

Gap Filling Workflow

Subcategorization

Options Select All Unselect All Invert

Primary grouping

Substance type

Protein binding alerts for skin sensitization by OASIS

Unclassified

Inventory Affiliation

OECD HPV Chemical Categories

Database Affiliation

US-EPA New Chemical Categories

(AOT)Protein binding by OASIS v1

Estrogen Receptor Binding

Biodegradation probability (Biowin 2)

Ionization at pH = 9

Protein binding potency GSH

Biodegradation probability (Biowin 5)

Hydrolysis half-life (Kb, pH 8)(Hydrowin)

Adjust options

Target 1 [target] 5 6 9 13 14 16 17

The rest of profilers are also here and can be used to subcategorize

Different analogues for profiler [Protein binding alerts for skin sensitization by OASIS]

Visual subcategorization

Differ from

At least [STOP]

Documentation

Options Select All Unselect All Invert

Metabolisms

Do not account metabolism

Selected

Select different

Remove selected

log Kow

log Kow

Accept prediction

80/100

Standard Skin sensitization

Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

Data Gap Filling: Relevant to the workflow profilers are provided and colored

Example 1: Target chemical has protein binding alert

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options: Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows a document tree with a selected chemical structure. The main workspace displays a workflow tree with various endpoints, including Sensitisation, Skin, in Vivo, and LLNA. A 'Success' dialog box is overlaid on the workflow, indicating 'Prediction accepted successfully'. The bottom of the screen shows a table of results for the selected chemical structure.

Endpoint	1 [target]	2	3	4	5	6	7	8	9
Structure	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>	<chem>O=C1C=CC(=O)C=C1</chem>
Ecotoxicological Information									
Workflow Controller [Finished workflow]									
Workflow name									
General task									
Active task									
Pause									
Stop									
Show action log									
Photoinduced toxicity									
Repeated Dose Toxicity									
Sensitisation									
Skin									
in Vivo									
LLNA									
EC3	13/14	R: 3.25 (-6.68+13...	M: Negative	M: Strongly posit...	M: Positive	M: 2.9 %	M: 0.23 %	M: Negative	M:
ToxCast									
Toxicity to Reproduction									
Toxicokinetics, Metabolism and Distribution									
Profiling									
Predefined									
Substance type	Discrete chemical								

Automated and Standardized workflows in Toolbox

SW for Skin sensitization - examples

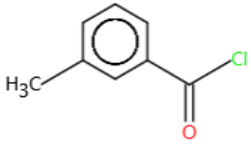
Report

- The report section is not part of the workflows
- Report could be generated once the prediction is accepted

Prediction of EC3 for m-toluyl chloride 1 / 6

QSAR Toolbox prediction for single chemical

Date: 28 Jan 2020
Author(s):
Contact details:

Target information		
Structural Information	Numerical Identifiers	Chemical names
SMILES: <chem>Cc1cccc(c1)C(Cl)=O</chem>	CAS#: 1711-06-4 Other: EC Number: 2169768	3-methylbenzoyl chloride Benzoyl chloride, 3-methyl- m-toluyl chloride
Structure 		

Prediction summary

Predicted endpoint: EC3; No effect specified; No species specified; No duration specified; No guideline specified

Predicted value: 3.25 (from -6.68 to 13.2)

Unit/scale: %

Data gap filling method: Read-across analysis, Standardized workflow for Skin sensitization

Summary: manually editable field
Not provided by the user

Generated report files

The following files were generated.
Select a file to open or save.

- Prediction report
- Category report
- Data matrix

PDF file containing the prediction report

Open Save as

Move Up Move Down

the PDF files.
be specified in the first page of the report

Back Next Cancel Create report

Discrete chemical

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulg

7 8 9

M: Negative M:

3

Automated and Standardized workflows in Toolbox

AW for Skin sensitization - examples

Activation: The Automated workflow is activated by clicking on the corresponding button

Example 2: Target chemical is activated metabolically

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options: Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, a sub-menu for 'Workflow' is open, showing 'Automated' and 'Standardized' options. The 'Automated' option is highlighted with a red box, and a red arrow points to it from a text box labeled 'Automated workflow'. The main workspace shows a chemical structure of a diamine (NCCCCNCCCCN) and a list of endpoints: Structure info, Parameters, Physical Chemical Properties, Environmental Fate and Transport, Ecotoxicological Information, and Human Health Hazards. The bottom panel shows 'Data Gap Filling Settings' with checkboxes for 'Only endpoint relevant' and 'At this position:' (all set to 0) and 'In nodes below:' (all set to 0).

Automated workflow

Automated and Standardized workflows in Toolbox

AW for Skin sensitization - examples

In this case the parent chemical is not active but is activated as a result of skin metabolism simulation

Example 2: Target chemical is activated metabolically. The primary group will be formed based on the alert(s) identified in the metabolite(s)

QSAR TOOLBOX

Filter endpoint tree... 1 [target]

Structure

Photoinduced toxicity

Repeated Dose Toxicity

Sensitisation

EC3 <OR> S M W N <OR> Skin...

ToxCast

Toxicity to Reproduction

Toxicokinetics, Metabolism and Distribution

Profilage

Predefined

Substance type

Endpoint Specific

Protein binding alerts for skin sensitiz...

Metabolism/Transformation

Autoxidation simulator

Endpoint Specific

Protein binding alerts for skin s...

Skin metabolism simulator

Endpoint Specific

Protein binding alerts for skin sensitization by OASIS

Explanation for: Protein binding alerts for skin sensitization by OASIS -> Schiff base formation with carbonyl compounds >> Aldehydes

Categories

Filter:

C-Nitroso compounds

Radical reactions

Free radical formation

Generated free radicals

Hydroperoxides

Schiff base formation

Benzoyl Schiff base formation

Benzoyl phosphine oxides

Direct acting Schiff base formers

1,2-Dicarbonyls and 1,3-Dicarbonyls

Di-substituted alpha,beta-unsaturated aldehydes

Pyrazolones and Pyrazolidinones derivatives

Pyrazolones and Pyrazolidinones

Schiff base formation with carbonyl compounds

Activated Carbonyl compounds

Aldehydes

alpha-Ketoesters

Aromatic carbonyl compounds

Bis aldehydes

SN1

Mechanistic Domain: Schiff base formation

Mechanistic Alert: Schiff base formation with carbonyl compounds

Structural Alert: Aldehydes

All aliphatic aldehydes can potentially undergo **Schiff base formation** with a primary amine, which is a reversible reaction (optimal at pH 3-4) and proceeds in two stages via a tetrahedral intermediate.

$$\text{R}-\text{C}(=\text{O})-\text{H} + \text{H}_2\text{N}-\text{R}' \rightleftharpoons \text{R}-\text{C}(\text{OH})(\text{H})-\text{N}(\text{R}')-\text{R}'$$

R = any carbon atom

According to Roberts *et al.*, 2015 simple mono-aldehydes are not highly reactive. However, the nature of the substituents will affect the reactivity of the carbonyl groups. For example, in TIMES SS model chemicals having the following structural boundaries will be predicted as strong skin sensitizers:

$$\text{R}-\text{C}(=\text{O})-\text{H}$$

R = H, -Csp3Ar, -Csp2C=C, -C(=O)OH, -OC

In the Local Lymph Node Assay, aromatic aldehydes are much less potent than expected from their reactivity as Schiff base electrophiles, and this applies to a smaller extent to human sensitization. In TIMES SS model aryl aldehydes without ortho-OH in their structure are predicted as Weak sensitizers when logKow is between 2 and 5.

3 x Schiff base formation

3 x Schiff base formation >> Schiff base formation with carbonyl...

2 x Schiff base formation >> Schiff base formation with carbonyl...

1 x Schiff base formation >> Schiff base formation with carbonyl...

2 x No alert found

Automated and Standardized workflows in Toolbox

AW for Skin sensitization - examples

Activation: The Automated workflow is activated by clicking on the corresponding button

Example 2: Target chemical is activated metabolically

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The left sidebar shows a list of documents, with 'Document 1' selected, showing its CAS number (56188). The main workspace displays a chemical structure of a target chemical (H₂N-CH₂-CH₂-NH-CH₂-CH₂-NH₂). A 'Filter endpoint tree...' dialog is open, showing a list of endpoints. A red box highlights the 'Skin sensitization' option in the 'Choices' list. A red arrow points from this box to a 'Select one' dialog box. This dialog box asks 'Which data do you want to use?' and lists three options: 'EC3 from LLNA or Skin sensitisation from GPMT assays' (selected), 'Skin sensitisation from GPMT assay', and 'EC3 from LLNA assay'. A red text box on the right side of the image provides instructions: 'List with endpoints and experimental data for Skin sensitization: Select EC3 from LLNA or Skin sensitization from GPMT assays'.

List with implemented workflows: Select "Skin sensitization"

List with endpoints and experimental data for Skin sensitization: Select EC3 from LLNA or Skin sensitization from GPMT assays

Automated and Standardized workflows in Toolbox

AW for Skin sensitization - examples

The following steps are applied automatically:

- Selection of databases
- Selection of primary group
- Sequence of subcategorizations
- Acceptance of the prediction

The screenshot displays the QSAR Toolbox software interface. The top menu bar includes options like Input, Profiling, Data, Category definition, Data Gap Filling, and Report. The main workspace shows a workflow tree on the left with nodes for Document 1, [C: 27;Md: 59;P: 1] Grouping with me, [C: 25;Md: 59;P: 1] Enter GF (AW by), [C: 25;Md: 59;P: 1] Data usage of, [C: 15;Md: 35;P: 1] Subcategorization, and Workflow Controller [Finished workflow]. The central panel shows a filter endpoint tree with a selected endpoint '1 [target]'. The bottom panel displays a table of results for various endpoints, including Neurotoxicity, Photoinduced toxicity, Repeated Dose Toxicity, Sensitisation, Skin, in Vivo, GPMT <OR> LLNA, EC3 <OR> S M W N <O... 25/60, R: Positive, M: Positive, M: 1.68 %, M: Negative, M: Positive, M: Category 1A, M: Negative, M: 8.4 %, and M: . The table also includes columns for ToxCast, Toxicity to Reproduction, Toxicokinetics, Metabolism and Distribution, Profiling, Predefined, and Substance type. A red box highlights a 'Success' dialog box with the message 'Prediction accepted successfully' and an 'OK' button. The text 'The prediction is accepted' is written in red above the dialog box.

The prediction is accepted

Success

Prediction accepted successfully

OK

More information about the automated and standardized workflows could be found in:

[Yordanova, D., Schultz, T.W., Kuseva, C., Tankova, K., Ivanova, H., Dermen, I., Pavlov, T., Temelkov, S., Chapkanov, A., Georgiev, M., Gissi, A., Sobanski, T., Mekenyan, O.G. 2019. Automated and standardized workflows in the OECD QSAR Toolbox. *Comput. Toxicol.* 10, pp. 89-104.](#)

Additional materials related to Automated and standardized workflows are available at Toolbox website:

<https://qsartoolbox.org>

List with tutorials:

- [Tutorial on how to predict skin sensitization potential by automated workflow](#)
- [Tutorial on how to predict skin sensitization potential by standardized workflow](#)
- [Tutorial of how to use Automated workflow for ecotoxicological prediction](#)
- [Tutorial of how to use Standardized workflow for ecotoxicological prediction](#)