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

Laboratory of Mathematical Chemistry,
Bourgas University “Prof. Assen Zlatarov”
Bulgaria
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 4.0

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. reticulate*, etc.
  - Crustaceans
  - Green algae

Standardized workflow

Automated workflow
Automated and Standardized workflows in Toolbox 4.0

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 4.0

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 4.0

Components

**AW**

- Input
- Profiling
- Endpoint
- Category definition
- Data gap filling
- Reporting

**SW**

- Input
- Profiling
- Endpoint
- Category definition
- Data gap filling
- Reporting

Same components as defined in the AW are used in the SW

The SW pauses at each of the stages and user is able to make different selection than those implemented in the AW

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)
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 4.0

**Executing module**

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

![Workflow controller interface](image)

- The workflow is navigated by two main buttons:
  - Continue/Pause allowing to continue or pause the workflow and
  - Stop which cancel (deactivate) the workflow.
  - Additionally, all actions that are done during the execution of the workflow are tracked down and could be seen in the “Show activity 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 4.0

Algorithm of Ecotoxicological workflow
Automated and Standardized workflows in Toolbox 4.0

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 4.0

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
Automated and Standardized workflows in Toolbox 4.0

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 4.0

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 4.0

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*:
     - Substance type,
     - Protein binding (OASIS + OECD),
     - Chemical elements,
     - Str. Similarity

PGPs – primary grouping profilers
PS – profilers for subcategorization
Automated and Standardized workflows in Toolbox 4.0

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

---

PGPs – primary grouping profilers
PS – profilers for subcategorization
**Automated and Standardized workflows in Toolbox 4.0**

**Algorithm of Ecotoxicological workflow**

**Criteria for acceptance of prediction**

### Trend analysis

After sub-categorization by all PGPs and PS

<table>
<thead>
<tr>
<th>Condition</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2 \geq 0.7$ and $NA \geq 10$</td>
<td>THEN accept the prediction and generate report, ELSE switch to Read across</td>
</tr>
</tbody>
</table>

### Read-across

After sub-categorization by all PGPs and PS

<table>
<thead>
<tr>
<th>Condition</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpolation AND $LC50 \leq 2$ log units OR $log Kow \leq 2$ log units AND $NA \geq 5$</td>
<td>THEN accept prediction and proceed with Report</td>
</tr>
</tbody>
</table>

**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 units
- $log K_{ow} \leq 2$: for the 5 closest analogues the range of variation of $log K_{ow}$ is $\leq 2$ log 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 4.0

SW for Ecotoxicological endpoint - example

Target chemical: CAS 120-83-2

![Chemical Structure]

**Endpoint:** LC50 or EC50  
**Effect:** Mortality  
**Species:** Actinopergyii (all fish) – illustrated in SW  
**Species:** Pimephales promelas – illustrated in AW  
**Duration:** 48h
Automated and Standardized workflows in Toolbox 4.0

**SW for Ecotoxicological endpoint - example**

**Location:** The AWs and SWs are part of the DGF
Automated and Standardized workflows in Toolbox 4.0

SW for Ecotoxicological endpoint - example

**Activation:** The Standardized workflow is activated by clicking on the corresponding button.
Automated and Standardized workflows in Toolbox 4.0

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*
Automated and Standardized workflows in Toolbox 4.0

**SW for Ecotoxicological endpoint - example**

**Selection of databases:** Five databases are available. All of them are selected in this example.
Automated and Standardized workflows in Toolbox 4.0

SW for Ecotoxicological endpoint - example

**Category definition:**

- Categories are hierarchically ordered 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

![Selection of primary grouping: MOA](image)
Automated and Standardized workflows in Toolbox 4.0

**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
Automated and Standardized workflows in Toolbox 4.0

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
Subcategorization:

- The applied subcategorizations are as follows:
  - Substance type (mandatory)
  - Filter by WS (mandatory)
  - 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 4.02 mg/L (toxic) after application of SW for Ecotox.
Automated and Standardized workflows in Toolbox 4.0

**SW for Ecotoxicological endpoint - example**

**Report:** the report could be generated once the prediction is obtained
Automated and Standardized workflows in Toolbox 4.0

**AW for Ecotoxicological endpoint - example**

**Activation:** The Automated workflow is activated by clicking on the corresponding button.
Automated and Standardized workflows in Toolbox 4.0

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

Confirmation of endpoint for execution of AW
Automated and Standardized workflows in Toolbox 4.0

AW for Ecotoxicological endpoint - example

All the consecutive steps are applied automatically:

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

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
Algorithm of Skin sensitization workflow

Part 1

Input → Entering target chemical → Identification of structural characteristics and mechanism of action → Skin sensitization data → Category definition → Data gap filling

Category definition:

1. Is protein binding alert for skin sensitization identified in the target?
   - Y: Define category by the protein binding alert identified in the target
   - N: Continue

2. Is protein binding alert for skin sensitization identified after AU simulation?
   - Y: Define category by the protein binding alert(s) identified in the package target + generated AU products
   - N: Continue

3. Is protein binding alert for skin sensitization identified after SM simulation?
   - Y: Define category by the protein binding alert(s) identified in the package target + generated SM products
   - N: Continue

Data gap filling:

Subcategorization:
1. Substance type
2. PBA for SS
3. PBA for SS + AU
4. PBA for SS + SM
5. Structural similarity (threshold 50 %)

Prediction is based on closest analogues

Report of prediction

Subcategorization by:
1. Substance type
2. PBA for SS
3. PBA for SS + AU
4. PBA for SS + SM
5. Structural similarity (threshold 50 %)
Automated and Standardized workflows in Toolbox 4.0

Algorithm of Skin sensitization workflow

Part 2

Input

Profiling

Endpoint

Category definition

Data gap filling

Reporting

Entering target chemical

Identification of structural characteristics and mechanism of action

Skin sensitization data

PBA for SS is found

PBA for SS is not found

1

Define category by most populated category based on structurally related profiling methods, e.g., OFG, OFG USEPA, etc.

2

Subcategorization by:
1. Substance type
2. PBA for SS
3. PBA for SS +AU
4. PBA for SS +SM
5. Protein binding potency
6. Keratinocyte gene expression alerts
7. OFG, OFG USEPA, OFG N. Haider
8. Structural similarity (threshold 50%)

Prediction is based on closest analogues

Report of prediction
Automated and Standardized workflows in Toolbox 4.0

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 N. Haider

- **Supporting profilers for further improvement of the category**
  - Substance type
  - Protein binding potency
  - Keratinocyte gene expression
  - Structure Similarity

- Abiotic and biotic activation of chemical is accounted by application of respective Autoxidation (AU) and Skin metabolism (SM) simulators
Automated and Standardized workflows in Toolbox 4.0

Algorithm of Skin sensitization workflow

- **Databases**
  - Skin sensitization
  - Skin sensitization ECETOC

- SS predictions are based on LLNA and GPMT exp. data.
  - LLNA – EC3, %
  - GPMT - Strong, Moderate, Weak and Non sensitizer
  - The dichotomous scale converting LLNA and GPMT potency categories into **Positive** and **Negative** SS is preferred when select to use LLNA and GPMT data
Automated and Standardized workflows in Toolbox 4.0

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**), **OFG USEPA**, **OFG N. Haider**
    - **Acute aquatic classification by ECOSAR**
    - **US-EPA New Chemical categories**

In this case, the most appropriate category is the collection of a **broader group** of analogues
Automated and Standardized workflows in Toolbox 4.0

Algorithm of Skin sensitization workflow

- **Reliability** of PBA for SS is evaluated based on:
  - Alert performance (AP) – ratio between correct vs. total number of applications of the alert
  - Number of chemicals in local training set (n)

- **Thresholds** for reliability of alerts could be assigned as follows:
  - High reliability – AP ≥ 60% and n ≥ 5
  - Low reliability - AP ≤ 60% and n ≥ 5
  - Undetermined reliability - 1 < n < 5
  - Undetermined theoretical reliability - no support by exp. data.

Table. 1 Calculated APs for selected PBA for SS accounting AU and SM activation

<table>
<thead>
<tr>
<th>Category</th>
<th>Parents only</th>
<th>AU activated</th>
<th>SM activated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of chemicals</td>
<td>AP</td>
<td>Number of chemicals</td>
</tr>
<tr>
<td></td>
<td>Pos  Neg</td>
<td>Pos  Neg</td>
<td>Pos  Neg</td>
</tr>
<tr>
<td>Activated aryl esters</td>
<td>42    1</td>
<td>42    1</td>
<td>43    1</td>
</tr>
<tr>
<td>Quinone type compounds</td>
<td>2     1</td>
<td>68    5</td>
<td>83    7</td>
</tr>
<tr>
<td>Anhydrides (sulphur analogues of anhydrides)</td>
<td>3     1</td>
<td>3     1</td>
<td>3     1</td>
</tr>
<tr>
<td>Epoxides, Aziridines and Sulfuranes</td>
<td>5     1</td>
<td>25    1</td>
<td>18    1</td>
</tr>
<tr>
<td>Aldehydes</td>
<td>28    2</td>
<td>37    9</td>
<td>93    15</td>
</tr>
<tr>
<td>alpha,beta-Carbonyl compounds</td>
<td>32    4</td>
<td>43    5</td>
<td>33    4</td>
</tr>
<tr>
<td>alpha,beta-Aldehydes</td>
<td>11    1</td>
<td>15    1</td>
<td>14    1</td>
</tr>
</tbody>
</table>
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 4.0

SW for Skin sensitization - examples

Location: The AWs and SWs are part of the DGF
Automated and Standardized workflows in Toolbox 4.0

SW for Skin sensitization - examples

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

**Example 1:** Target chemical has protein binding alert

![Standardized workflow](image-url)
Automated and Standardized workflows in Toolbox 4.0

**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
Automated and Standardized workflows in Toolbox 4.0

SW for Skin sensitization - examples

**Selection of databases:** Two databases are available

**Example 1:** Target chemical has protein binding alert

![List with databases]

**Workflow controller is activated**

CAS 1711-06-4
Automated and Standardized workflows in Toolbox 4.0

**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

![Example chemical structure](image)
Automated and Standardized workflows in Toolbox 4.0

**SW for Skin sensitization - examples**

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

**Example 1:** Target chemical has protein binding alert

![Q SAR Toolbox](image)

**9 chemicals are found based on the primary group**

CAS 1711-06-4
Automated and Standardized workflows in Toolbox 4.0

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

Relevant to the workflow profilers are listed here and ordered in a way they are executed in AW

Legend of different color meaning

CAS 1711-06-4
Automated and Standardized workflows in Toolbox 4.0

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 rest of profilers are also here and can be used to subcategorize
Automated and Standardized workflows in Toolbox 4.0

**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
Automated and Standardized workflows in Toolbox 4.0

**SW for Skin sensitization - examples**

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

**Example 2**: Target chemical is activated metabolically
Automated and Standardized workflows in Toolbox 4.0

SW for Skin sensitization - examples

Illustration of the active metabolite obtained after skin metabolism simulation

**Example 2:** Target chemical is activated metabolically
Automated and Standardized workflows in Toolbox 4.0

**SW for Skin sensitization - examples**

**Category definition:** Primary group is defined accounting skin metabolism activation

**Example 2:** Target chemical is activated metabolically

![Status bar showing that current grouping is applied with metabolism](image)

**CAS 56188**
Automated and Standardized workflows in Toolbox 4.0

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

Example 2: Target chemical is activated metabolically
Automated and Standardized workflows in Toolbox 4.0

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

**Example 2:** Target chemical is activated metabolically

Apply the following subcategorizations:
- Substance type;
- Protein binding alerts for skin sensitization by OASIS;
- Protein binding alerts for skin sensitization by OASIS + Autoxidation simulator;
- Protein binding alerts for skin sensitization by OASIS + Skin metabolism simulator;
- Structure similarity – select all having at least 50% similarity with respect to the target

After each subcategorization the Subcategorization dialogue is closed (see next slide)
Automated and Standardized workflows in Toolbox 4.0

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

**Example 2:** Target chemical is activated metabolically

Continue button should be pressed in order to open the subcategorization
Automated and Standardized workflows in Toolbox 4.0

**SW for Skin sensitization - examples**

**Report**

- Report is not part of the workflows
- Information specifying that the prediction is produced by AW or SWs is available in the report

![Image showing report content](image_url)

**Reporting of additional information (from data matrix) is provided in excel**
Automated and Standardized workflows in Toolbox 4.0

**AW for Skin sensitization endpoint - example**

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

CAS 56188
Automated and Standardized workflows in Toolbox 4.0

AW for Skin sensitization endpoint - example

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

- List with available workflows
- List with endpoints and experimental data for Skin sensitization: Select EC3 from LLNA
Automated and Standardized workflows in Toolbox 4.0

AW for Skin sensitization endpoint - example

All the consecutive steps are applied automatically:

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

The prediction is accepted

CAS 56188