

QSAR-ME Profiler 2025 Manual

Software and models development

QSAR Research Unit in Environmental Chemistry and Ecotoxicology
Department of Theoretical and Applied Sciences (DiSTA)
University of Insubria, Varese, Italy
https://dunant.dista.uninsubria.it/qsar/

Funding information

University of Insubria, Post Doc grant (2021-2022) "In silico solutions for the assessment of biotransformation related endpoints of organic chemicals in multiple organisms."

Contacts

Prof. Ester Papa - e-mail: ester.papa@uninsubria.it (project responsible)
Dr. Nicola Chirico, PhD - e-mail: nicola.chirico@uninsubria.it (software development)

How to cite

Nicola Chirico, Arianna Sgariboldi, Marco Evangelista and Ester Papa. QSAR-ME Profiler 2025: A New Software for QSA(P)R Predictions Supported by Structural Analysis. *Chemical Research in Toxicology*, 2026, 39, 5, 813-818. DOI: 10.1021/acs.chemrestox.5c00552

Limitations of liability and disclaimer of warranty

QSAR-ME Profiler 2025 and the accompanying materials and manuals are provided "as they are" without warranty of any kind. The authors do not warrant, guarantee, or make any representations, either expressed or implied, regarding the use, or the results from the use of QSAR-ME Profiler 2025, the accompanying materials and manuals, in terms of correctness, accuracy, reliability, currentness, or otherwise.

You assume the entire risk as to the result and performance of QSAR-ME Profiler 2025.

In no event shall the authors be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with QSAR-ME Profiler 2025 or the use or other dealings in QSAR-ME Profiler 2025, even if the authors have been advised of the possibility of such damages.

QSAR-ME Profiler 2025 software, the accompanying materials and manuals are protected by copyright: 2026, University of Insubria, https://www.uninsubria.it - Varese, Italy.

Contents

1 Introduction 1

2 QSAR-ME Profiler 2025 requirements, installation and running 1

3 Using QSAR-ME Profiler 2025 1
- 3.1 Main window . . . 1
- 3.2 Selection of a QSAR . . . 2
- 3.3 Selection of QSARs chemicals and neighbors detection . . . 2
- 3.4 Prediction of target chemicals endpoint value . . . 3
- 3.5 Selection of target chemicals and neighbors detection . . . 5
- 3.6 MLR QSAR predicted endpoint diagnostics . . . 5
- 3.7 LDA QSAR predicted endpoint diagnostics . . . 6
- 3.8 Target chemicals prediction tables . . . 7
- 3.9 Export report and similarity table . . . 9
- 3.10 Other options . . . 10
  - 3.10.1 Sessions . . . 11
  - 3.10.2 Charts . . . 11
  - 3.10.3 Similarity . . . 11
  - 3.10.4 Significant digits . . . 11

4 Running QSAR-ME Profiler 2025 as a command line application 12

5 QSAR-ME Profiler 2025 QSARs customization 12
- 5.1 Configure QSAR categories . . . 12
  - 5.1.1 Add a user-defined QSAR . . . 13
  - 5.1.2 QSAR files . . . 14
  - 5.1.3 Create MLR QSAR xml and csv files . . . 14
  - 5.1.4 Create MLR QSAR xml and csv files - Toxtree version . . . 18
  - 5.1.5 Create LDA QSAR xml and csv files . . . 19

6 QSARs shipped with QSAR-ME Profiler 2025 22

7 Acknowledgments 23

8 References 24

1 Introduction

QSAR-ME Profiler 2025 (Quantitave Structure-Activity Relationship Multiple Endpoint Profiler) is a multi-platform software developed for the application of different classes of QSA(P)Rs[1] for the prediction of the activity/property of chemicals.

QSAR-ME Profiler 2025 is shipped with a set of QSA(P)Rs[2] aiming to the assessment of the potential hazard and risk posed by heterogeneous organic chemicals. These QSA(P)Rs, developed by the QSAR Research Unit in Environmental Chemistry and Ecotoxicology of the University of Insubria, cover physical-chemical properties, global indexes, aquatic toxicity, *in vivo* and *in vitro* mammalian biotransformations. In addition, these QSA(P)Rs are compliant to the OECD principles for regulatory purposes of (Quantitative) Structure-Activity Relationship models and are accompanied by the corresponding QMRF (QSAR Model Reporting Format) documents[3].

QSAR-ME Profiler 2025 can apply QSA(P)Rs in batch and automates many steps, including the application of Toxtree [1] for the detection of metabolic reactions and the generation of metabolites, to be further processed by QSAR-ME Profiler 2025. Predictions, both for the target chemicals and detected metabolites, can be explored in full details in the context of the applied QSA(P)Rs, and further supported by the comparison of the activity/property of the most similar chemicals automatically detected[4] by the software.

From now on, to simplify reading, QSAR will also mean QSPR, therefore QSA(P)R will be replaced by QSAR.

2 QSAR-ME Profiler 2025 requirements, installation and running

QSAR-ME Profiler 2025 requires Java™ SE 21 runtime environment or a more recent version.

Unzip the downloaded file in a folder of choice then, to run QSAR-ME Profiler 2025, double click on the QSAR-ME-Profiler.jar icon or open a terminal pointing to the QSAR-ME Profiler 2025 main folder and type in

```
java -jar QSAR-ME-Profiler.jar
```

3 Using QSAR-ME Profiler 2025

3.1 Main window

QSAR-ME Profiler 2025 main window is splitted in four resizable panels, whose relevant sections are shown in Figure 1. The top left panel allows for QSARs selection, setup and predictions tables of target chemicals and metabolites. The top right panel graphically shows predictions in the corresponding QSAR context. The bottom left panel shows a detailed view of the selected QSAR and corresponding target/metabolites predictions. The bottom right panel depicts the chemicals of interest and their neighbors.

[1] QSAR-ME Profiler 2025 includes also Structure-Properties (P) models.

[2] See Section 6 for further details.

[3] QSA(P)Rs shipped with QSAR-ME Profiler 2025 are calibrated using full datasets, aiming for maximum structural applicability domain. QMRFs equations may be calibrated using a training set and an external set, for validation purposes (as required by the OECD principles), generated by splitting the full dataset.

[4] From the training set of the applied QSA(P)Rs.

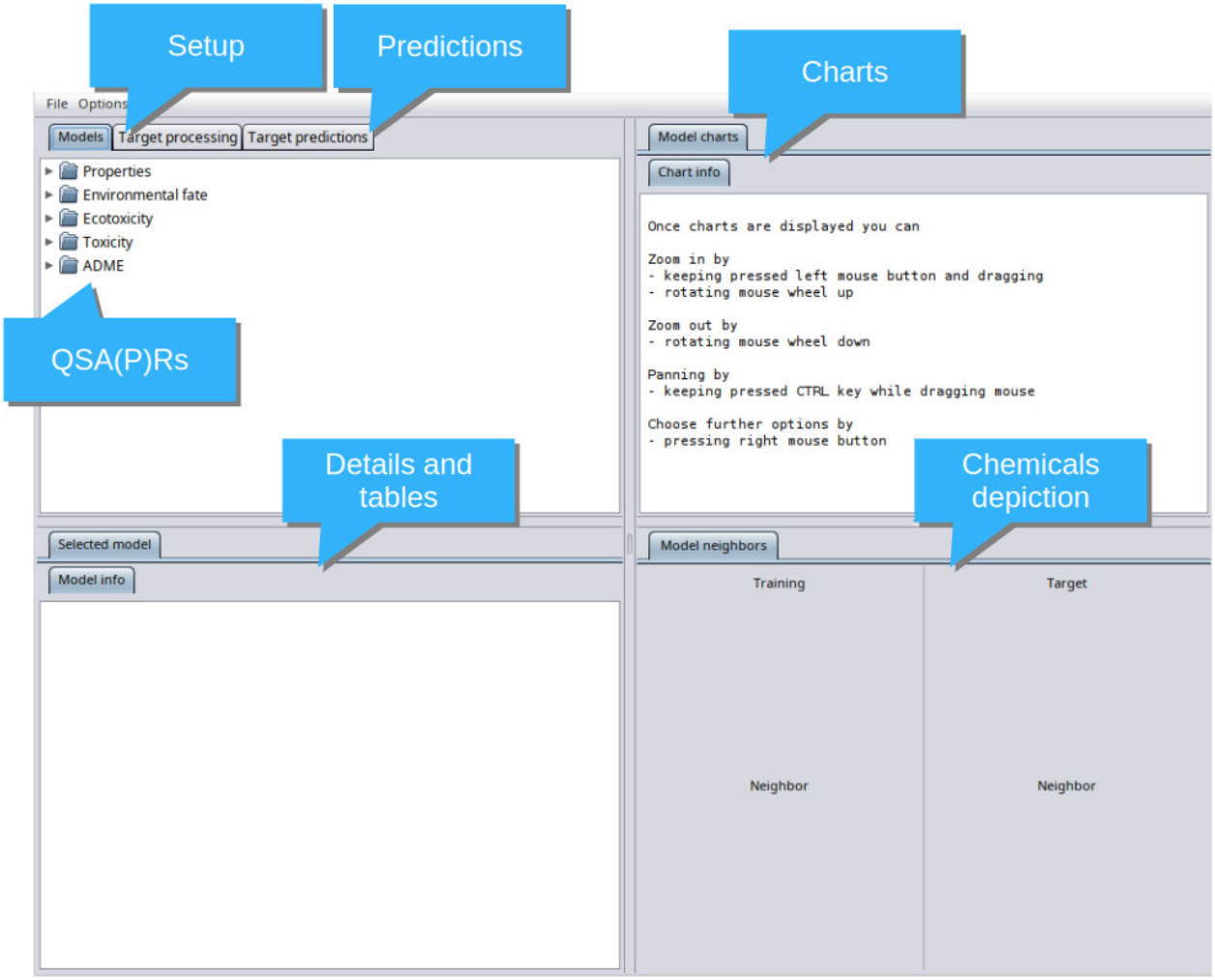

Figure 1: QSAR-ME Profiler 2025 main window.

3.2 Selection of a QSAR

QSARs, both shipped with QSAR-ME Profiler 2025 and user-defined, are listed in the Models tab, and are organized as groups (like e.g., Properties and Ecotoxicity) containing more specific subgroups (like e.g., Fish pLC50 and Algae pEC50). QSARs can be selected from the subgroups, as shown in Figure 2. Once selected, QSAR details and performance charts are displayed in the main window, as in Figure 3 example. By right clicking over the selected QSAR, the corresponding QMRF can be displayed(provided that a PDF viewer is installed[5]).

3.3 Selection of QSARs chemicals and neighbors detection

To explore QSARs in deeper details, QSARs training set chemicals can be selected both from the Model data table and by clicking on one point[6] on the performance charts[7], as shown in Figure 4. Once selected, the chemical is highlighted on the performance charts by two crossing lines, and the chemical structure is depicted in the Training pane of the Model neighbors tab. Chemical selection triggers neighbors detection, which are then collected as a table in the Training neighbors tab. Neighbors can

[5]To manually open a QMRF, go to the qsar folder located in the main folder of QSAR-ME Profiler 2025, then open the qsar_me_profiler folder and locate the QSARs subgroup folder, then select the PDF of interest.

[6]By hovering the mouse pointer over a chart point, the name of the corresponding chemical is shown.

[7]Concerning LDA models, only applicability domain charts allow the selection of chemicals by mouse clicking.

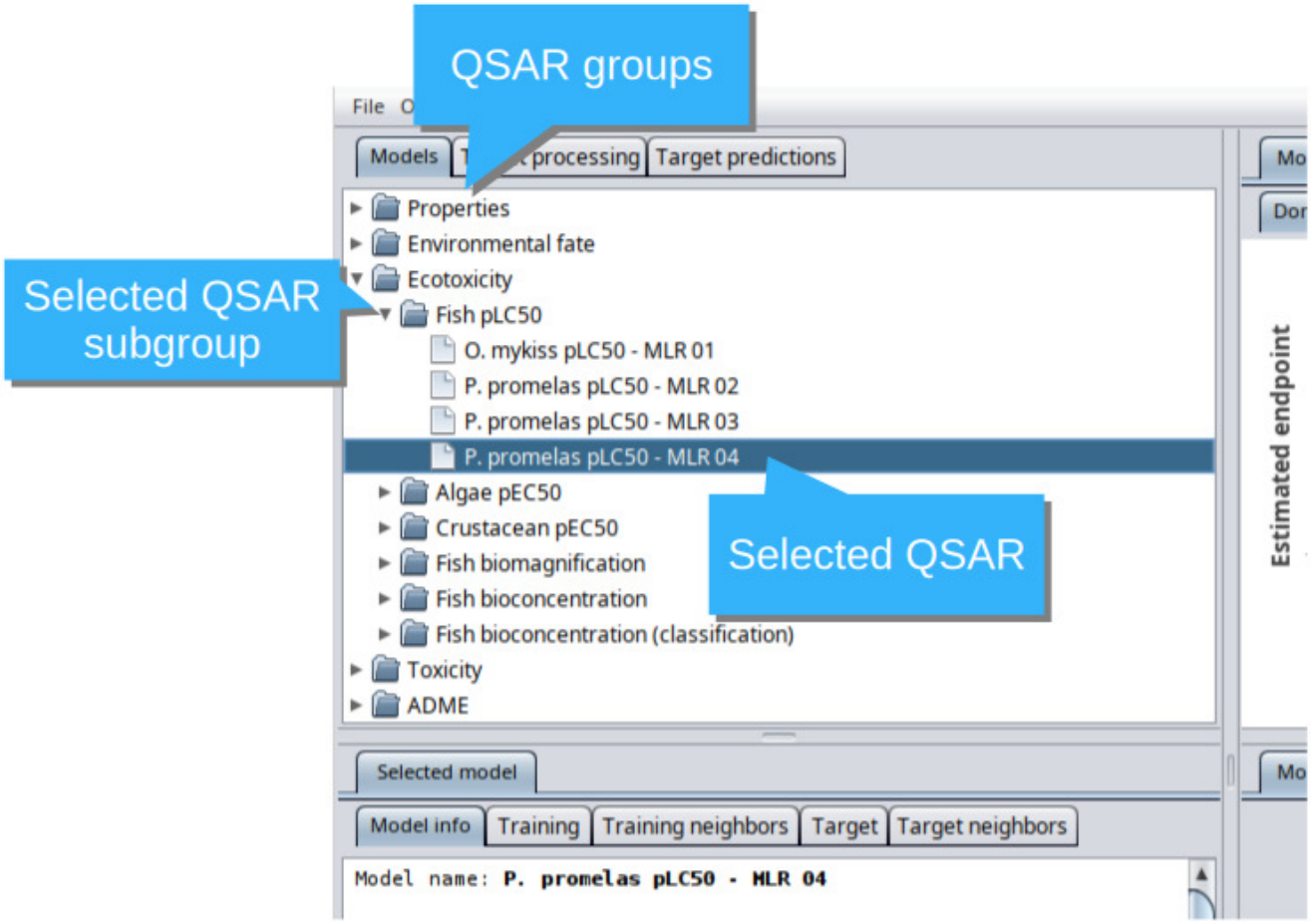

Figure 2: QSAR selection.

also be selected, and are both highlighted by crossing lines in the performance charts and depicted in the Neighbor pane, still in the Model neighbors tab, as shown in Figure 5. Neighbors may change according to the fingerprint[8] and distance measures[9], see also Section 3.10.3.

3.4 Prediction of target chemicals endpoint value

Target chemicals[10] endpoint value can by predicted from the target processing pane, by selecting the Target processing tab, as shown Figure 6. Chemicals structure must be coded as SMILES, and must be entered in the TARGET SMILES input area[11] in the smi format i.e., as SMILES followed by space or tabulation, and then a label (e.g. CAS number or chemical name).

QSAR-ME profiler 2025 is shipped with some categories of QSARs developed according to CYP P-450 reactions. To select pertinent QSARs, QSAR-ME Profiler 2025 needs to know the target chemicals most probable metabolic reactions. Available reactions can be selected from the USER SELECTED REACTION list. If no reactions are selected, QSAR-ME Profiler 2025 is set to run Toxtree for automatic detection.

[8]Available fingerprints are Pubchem, E-State, Klekota and Roth, Klekota and Roth count, substructure and substructure count.

[9]Available distances are Tanimoto, cosine and dice.

[10]In QSAR-ME Profiler 2025, target chemicals are intended as the user-entered chemicals whose endpoint value should be predicted.

[11]By clicking the right mouse button on the input area, SMILES can be copy/pasted or imported from an external source.

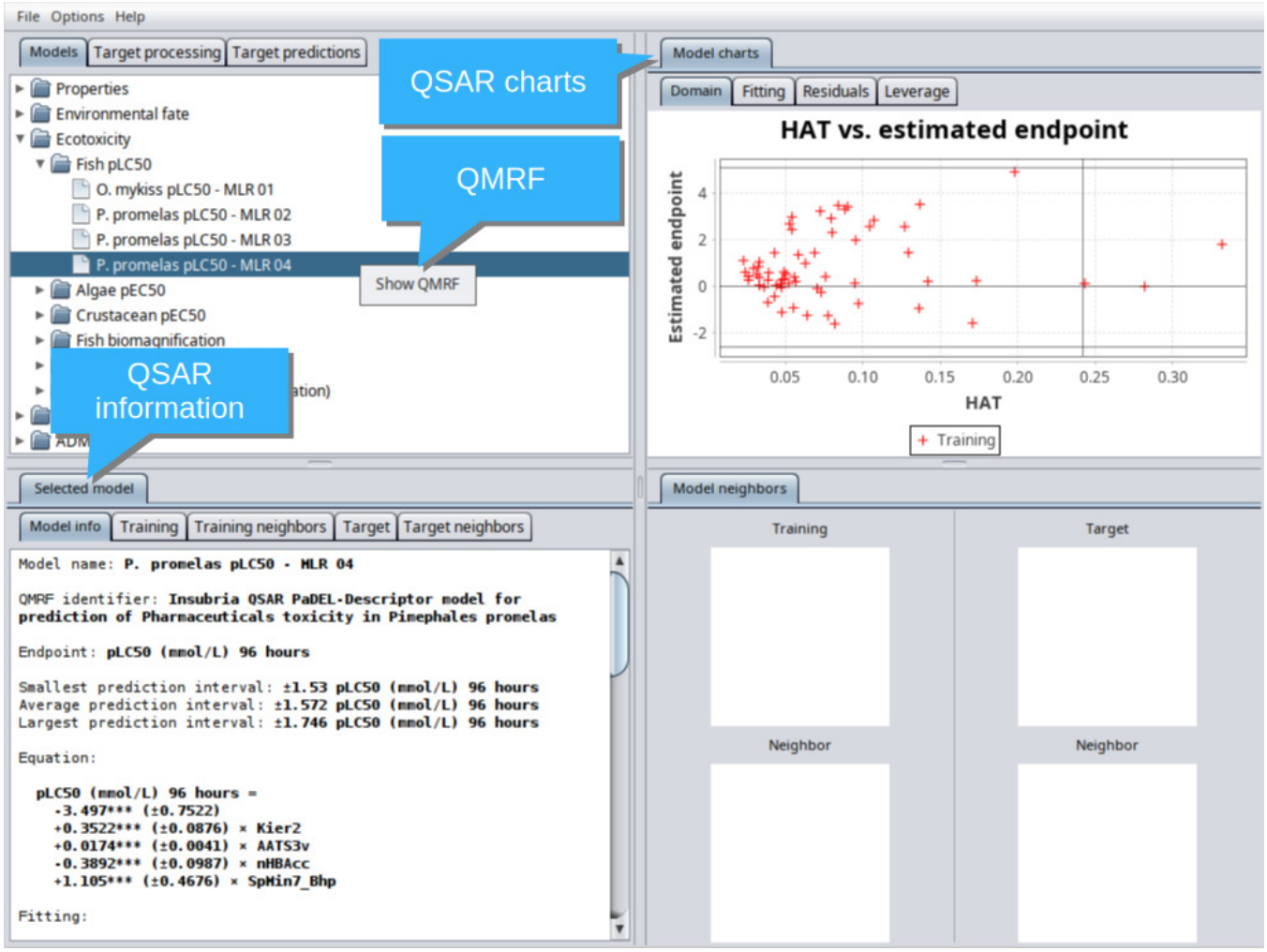

Figure 3: Selected QSAR details and performances.

If checked, the Add detected metabolites to predictions option sets QSAR-ME Profiler 2025 to add target chemicals metabolites, which will be automatically detected by running Toxtree, to the target predictions.

Once the setup is complete, by pressing the Process button, processing of the chemicals begins. Concerning QSARs shipped with QSAR-ME Profiler 2025, target (and metabolites) descriptors are calculated by QSAR-ME Profiler 2025 itself. Concerning user-defined QSARs, QSAR-ME Profiler 2025 asks for a CSV file containing the descriptors values, formatted in the following way. The first column must be filled with the chemicals names while the following columns must be filled by the descriptors values. The first row must be filled with the first column name (e.g. Name) followed by the descriptors name. Below follows a descriptors CSV file example[12].

```
Name,    GGI1,              GGI2,              GGI3,              GATS1p,            MW
MOL_01, 9.32170283982964, 21.0376216363349, 6.04764452295474, 0.148619042403139, 1.285074622688880
MOL_02, 9.80350145766020, 50.3339498186955, 22.1746965841674, 0.140821514317611, 0.833539090654583
MOL_03, 10.6320742681355, 18.3235081995985, 14.1111705535611, 0.287791197362692, 1.145218503007730
```

If the Add detected metabolites to predictions option has been checked, QSAR-ME Profiler 2025 also needs a CSV descriptors file calculated for the metabolites. Therefore, QSAR-ME Profiler 2025 first generate a .smi files, which can be used as input for the software used to calculate the descriptors. Then, the resulting CSV file, formatted as explained above, can be loaded by QSAR-ME Profiler 2025. All

[12]In production CSV files, items should be separated only by the separator char (usually comma). Items are here aligned to simplify reading.

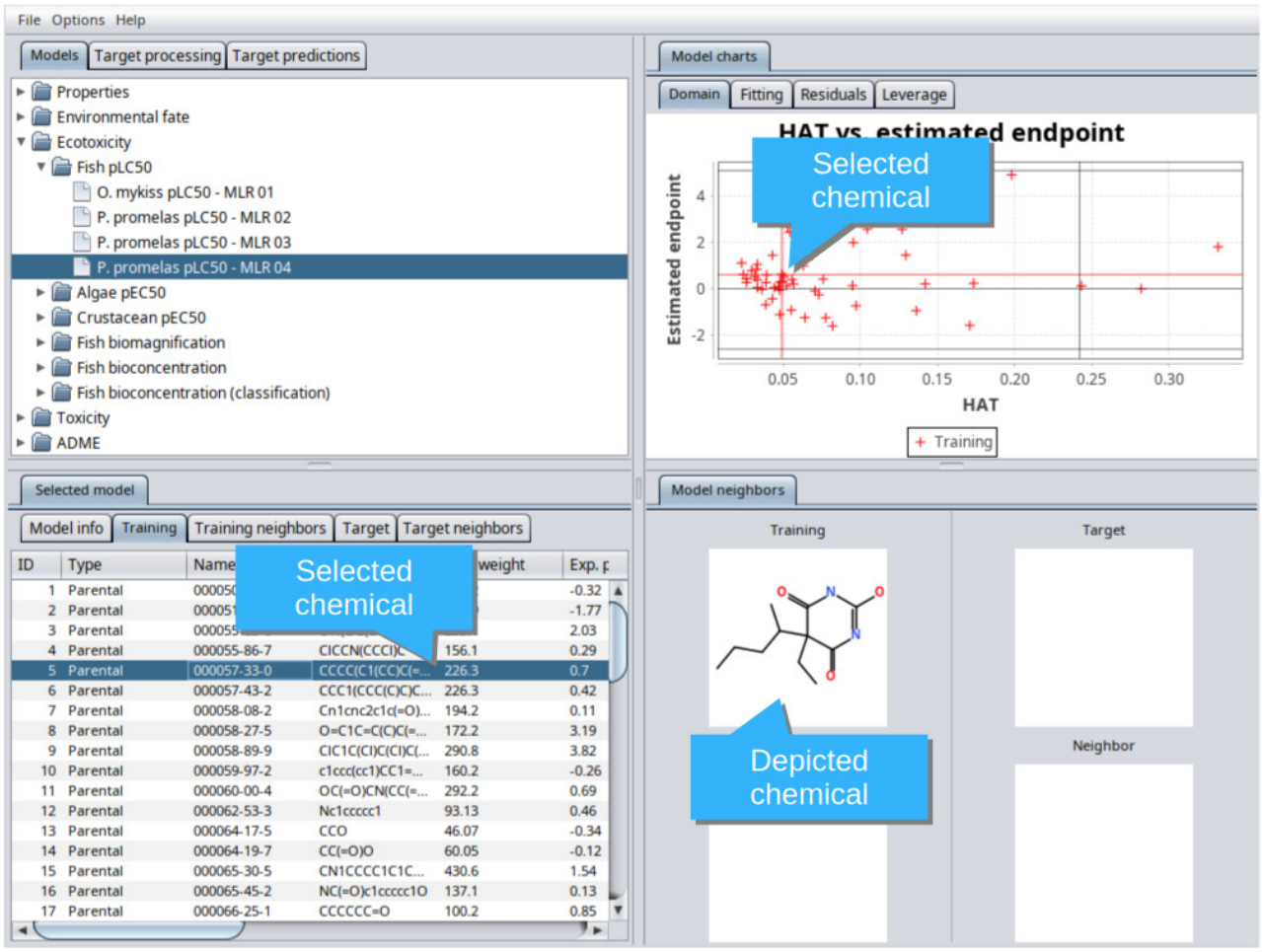

Figure 4: Chemicals selection.

these steps are supported by explicative dialogs, during the process, by QSAR-ME Profiler 2025.

3.5 Selection of target chemicals and neighbors detection

Selection and depiction of target chemicals, and corresponding neighbors, follows the same logic of Section 3.3, except that the focus is now on Target instead of Training, see also Figure 7.

3.6 MLR QSAR predicted endpoint diagnostics

QSARs diagnostic charts help to identify the context of the endpoint predicted values. Concerning MLR QSARs, the HAT vs. estimated endpoint chart (Model charts → Domain tab) allows plotting the predicted target endpoint value against the QSAR training set values. This chart, see also the upper left chart in Figure 8, can be used to check the compliance of the target chemicals with the QSAR applicability domain, both in terms of chemical structure and estimated endpoint value. The vertical dark grey line is the threshold HAT value,[13] where the chemicals on the right of the threshold line are considered structural outliers. Chemicals above or below the horizontal dark grey lines i.e., the range of experimental endpoint, are considered endpoint extrapolations.

[13]The HAT threshold is calculated as $3p'/n$, where p' is the number of the descriptors + 1 and n the number of compounds.

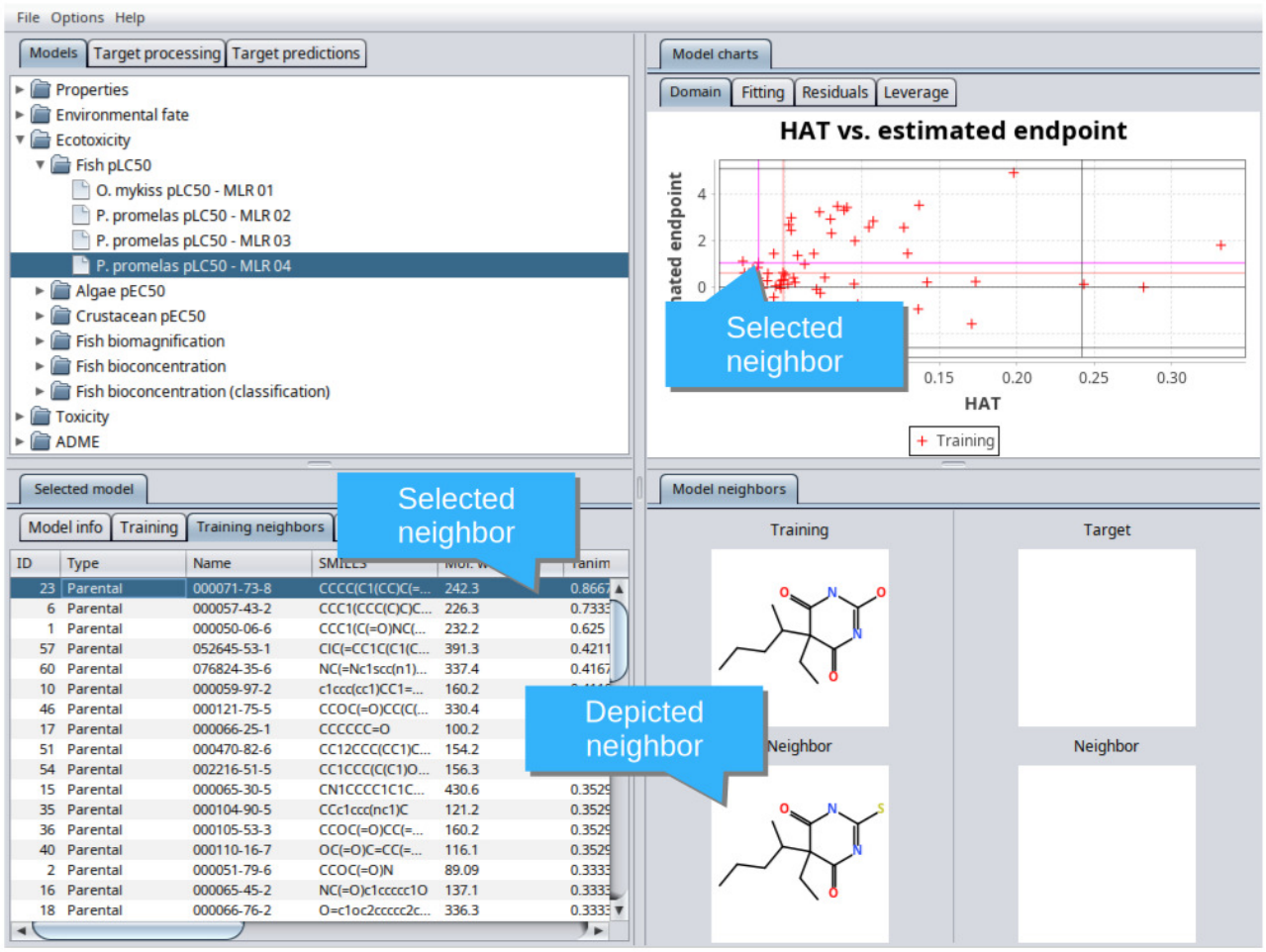

Figure 5: Neighbors selection.

Concerning QSARs training set only performances, the experimental vs. predicted endpoint (training set) chart (Model charts → Fitting tab), see also Figure 8 lower left chart), plots the experimental vs the predicted endpoint values, while the Endpoint residual (training set) chart (Model charts → Residuals tab), see also Figure 8 upper right chart, shows their difference vs the predicted value on the x-axis. The HAT vs. standardized residuals (training set) chart (Model charts → Leverage tab), see also Figure 8 lower right chart, is similar to the Domain chart, but plots standardized residuals on the y-axis instead of the predicted endpoint, whose thresholds are arbitrarily set to 2.5 standardized deviation unit (horizontal black lines). Chemicals falling outside these thresholds are considered outliers.

3.7 LDA QSAR predicted endpoint diagnostics

LDA Applicability Domain (event) chart (Model charts → Domain event tab) charts are defined in terms of chemical structure and probability of a classification event. The endpoint domain [2] is here arbitrarily defined by post probability thresholds (horizontal grey lines of left chart of Figure 9) smaller than 0.25 of not being the event, and bigger than 0.75 of being the event, therefore a classification is considered uncertain if the post probability is between 0.25 and 0.75. A chemical structure is considered an outlier if its distance, measured as the average of the 3 nearest $cos\ \alpha$ neighbors [2, 3], is smaller than the 0.95 quantile of the k-nearest neighbors within the distribution of all training set distances [4]. This threshold is depicted by a vertical grey line in the left chart example of Figure 9.

Prediction uncertainty can be seen by selecting one table from the Selected model tab, and is esti-

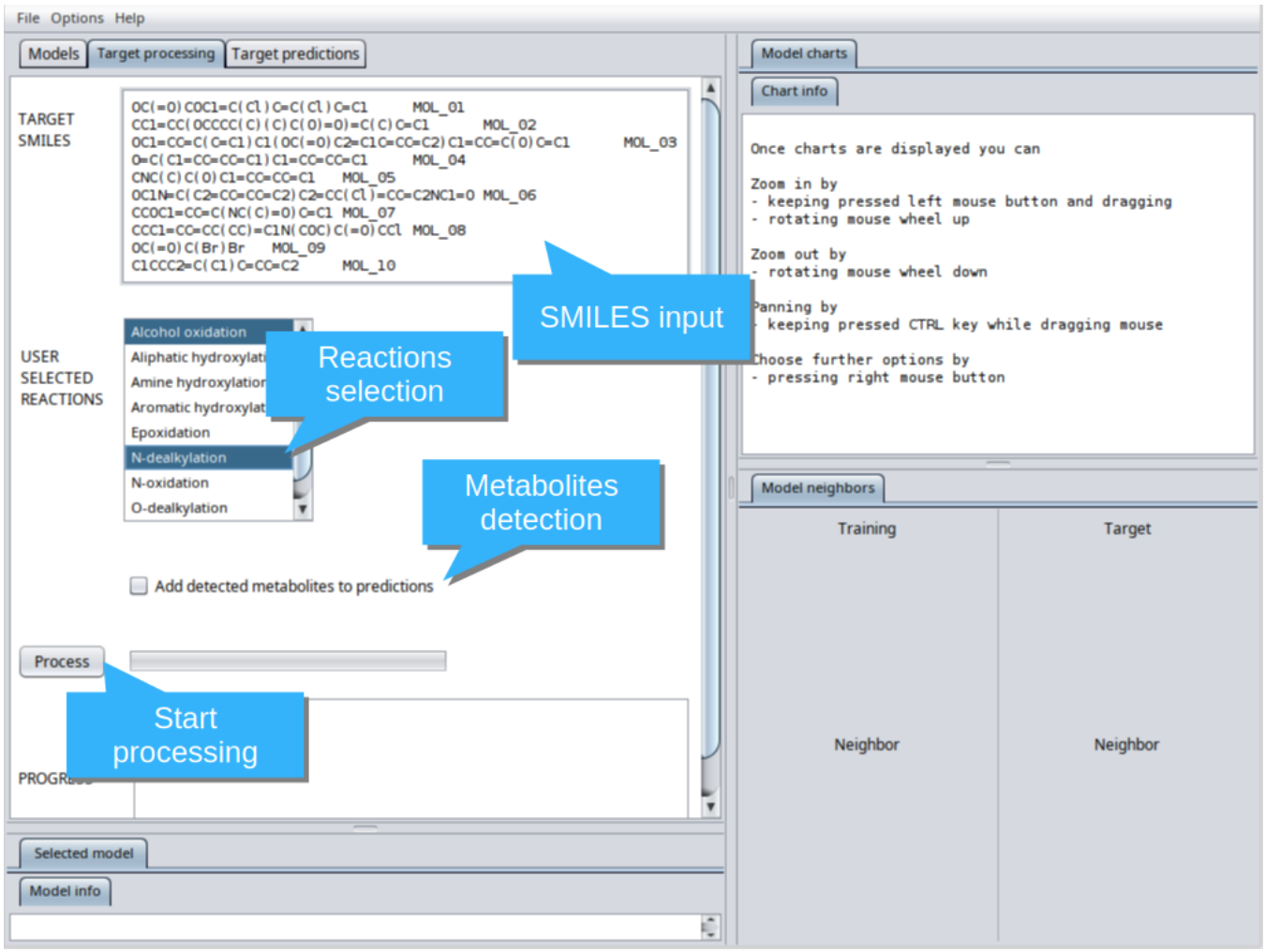

Figure 6: Target chemicals processing pane.

mated by the Shannon entropy, calculated as $-\sum_n p(x_n)\log p(x_n)$, where p is the event post probability.

ROC (event) (Receiver Operating Characteristic) charts (Model charts → ROC event tab) are also available as a reference for the classification QSARs (see Figure 9 right chart, as an example) performances. When classification involves more than two classes, ROC charts are depicted using the one versus the rest approach, where the class under consideration is the event and the remaining are the non-event. By hovering the mouse cursor over the chart points, hints about probability, sensitivity and specificity are shown.

3.8 Target chemicals prediction tables

Once target processing has been completed (see Section 3.4), prediction tables are generated in the Target predictions tab, as shown in the example Figure 10. Predictions are organized according to the QSAR-ME Profiler 2025 QSARs groups and subgroups as explained in Section 3.2. For each target chemical, the ID is reported followed by Type, which can be either Parental for the target chemicals or Metabolite if the Add detected metabolites to predictions option, from the Target processing tab has been selected (see section 3.4). Endpoint reports a brief endpoint description while Rank, Reaction and SOM score concern metabolites[14] only. SOM score [5] is related to the probability[15] of an atom of being a site of metabolism (SOM). Rank spans from 1 to>=4 and is inversely related to the SOM score (the

[14]Metabolites are detected according to the most likely reactions mediated by cytochrome P-450.
[15]The higher the score the higher the probability.

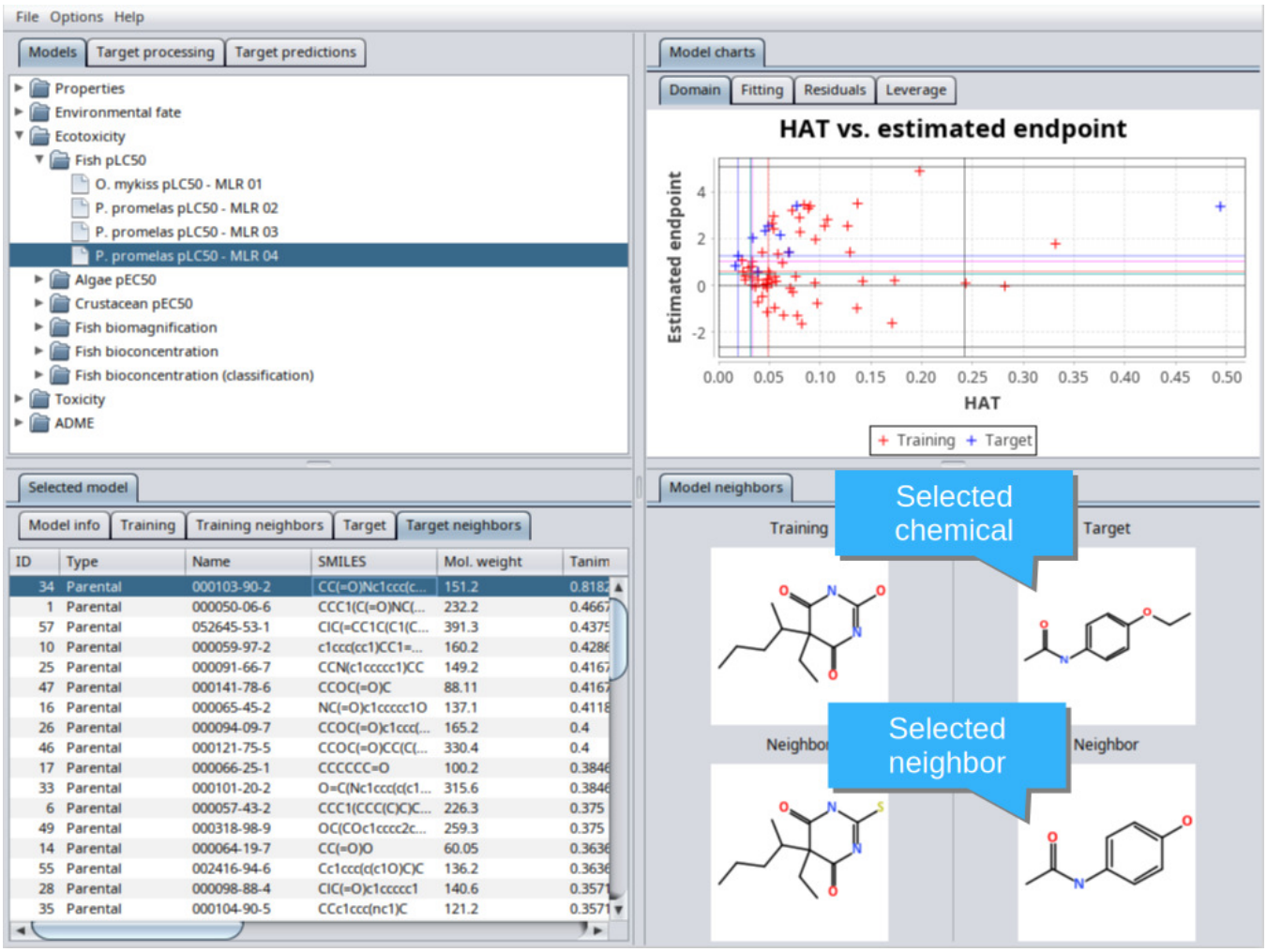

Figure 7: Depiction of target chemicals and corresponding neighbors.

lower the rank the higher the score)[16]. Reaction is the detected metabolic reaction.

Consensus weighted ALL and Consensus weighted AD contain combined endpoint predictions, computed as weighted averages[17] [7] of the predictions of the selected QSARs, by checking the corresponding boxes (see Figure 10). Consensus weighted ALL averages all predictions while Consensus weighted AD averages only predictions within the applicability domain. Missing compliance with the applicability domain is detailed in the QSAR columns after the predicted values, and is reported as * if the prediction is outside the experimental endpoint domain, # if the chemical is outside the chemical structure domain and $ if the prediction interval is bigger than the maximum calculated for the training set of the QSAR. Concerning QSARs developed according to the most likely reactions mediated by cytochrome P-450, prediction tables add information concerning the ranks and reactions used to develop the QSARs (see Figure 11).

LDA QSARs tables are similar to the MLR ones, except that endpoints are discrete classes. Consensus

[16] If only one metabolite is detected it is associated to rank 1.

[17] Weight is calculated as $w = 1/\sigma^2$, where σ^2 is the square of its uncertainty. For N predictions, weighted average is $\frac{\sum w_i \hat{y}_i}{\sum w_i}$ and its uncertainty is $\frac{1}{\sqrt{\sum w_i}}$, where $i = 1, ..., N$. Prediction interval of individual uncertainty is calculated as [6] $\sigma = \pm t_{stud\ \alpha/2} \cdot s \times \sqrt{1 + h_{ii}}$, where $t_{stud\ \alpha/2}$ is t-student for $\alpha/2 = 0.025$ and s is $\sqrt{\frac{\sum_n (y_n - \hat{y}_n)^2}{n-p-1}}$, where n is the number of the training set chemicals and p the number of descriptors.h_{ii} is the leverage value calculated as $h_{ii} = x_i \cdot (X \cdot X^T)^{-1} \cdot x_i^T$, where X is the matrix of the training descriptors and i the ith chemical descriptors values vector.

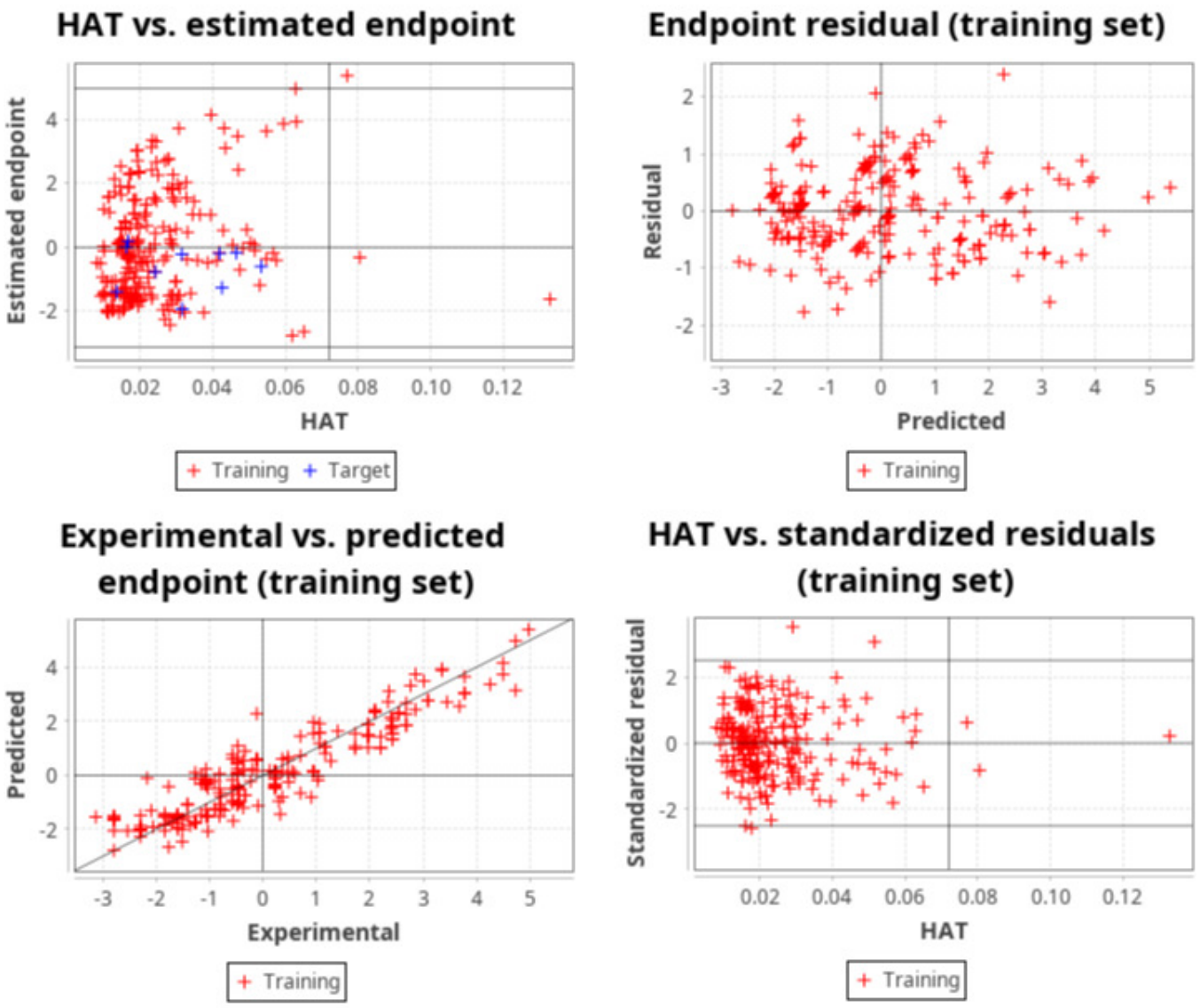

Figure 8: MLR diagnostic charts.

weighted ALL is calculated by counting the most represented event from the selected QSARs while Consensus weighted AD does the same but using only predicted events compliant with the applicability domain. Missing compliance with the applicability domain is detailed in the QSAR columns after the predicted values, and is reported as * if the prediction is outside the experimental classification domain, # if the chemical is outside the structural domain, and $ if the Shannon entropy is larger than the maximum calculated from the training set of the QSAR.

3.9 Export report and similarity table

Target chemicals and metabolites reports, containing the applied QSAR information, predicted values and detected neighbors, can be generated by right-clicking on the chemical of interest from the Selected model → Target table, and then by selecting the Build report option from the drop-down menu.

Concerning neighbors detection, choosing the best combinations different similarity measures and fingerprints may be tricky. Indeed it could require many trials for different chemicals, therefore, to help this task, similarity tables calculated by combining all distance measures and fingerprints, against all training chemicals, can be generated by selecting the Build similarity table menu item.

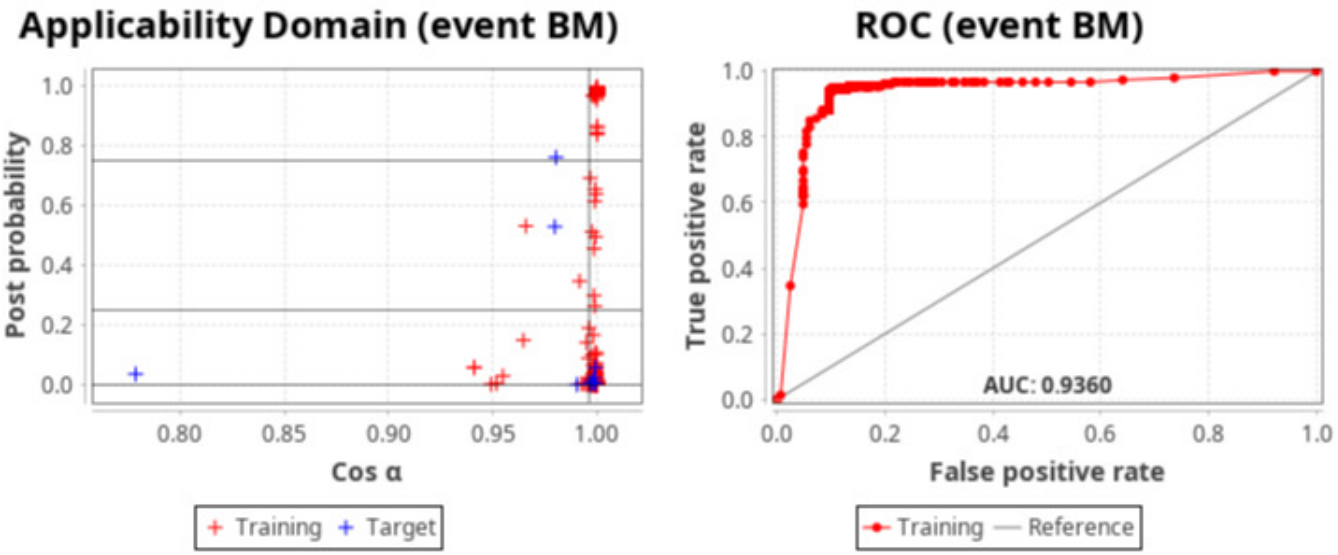

Figure 9: LDA diagnostic charts.

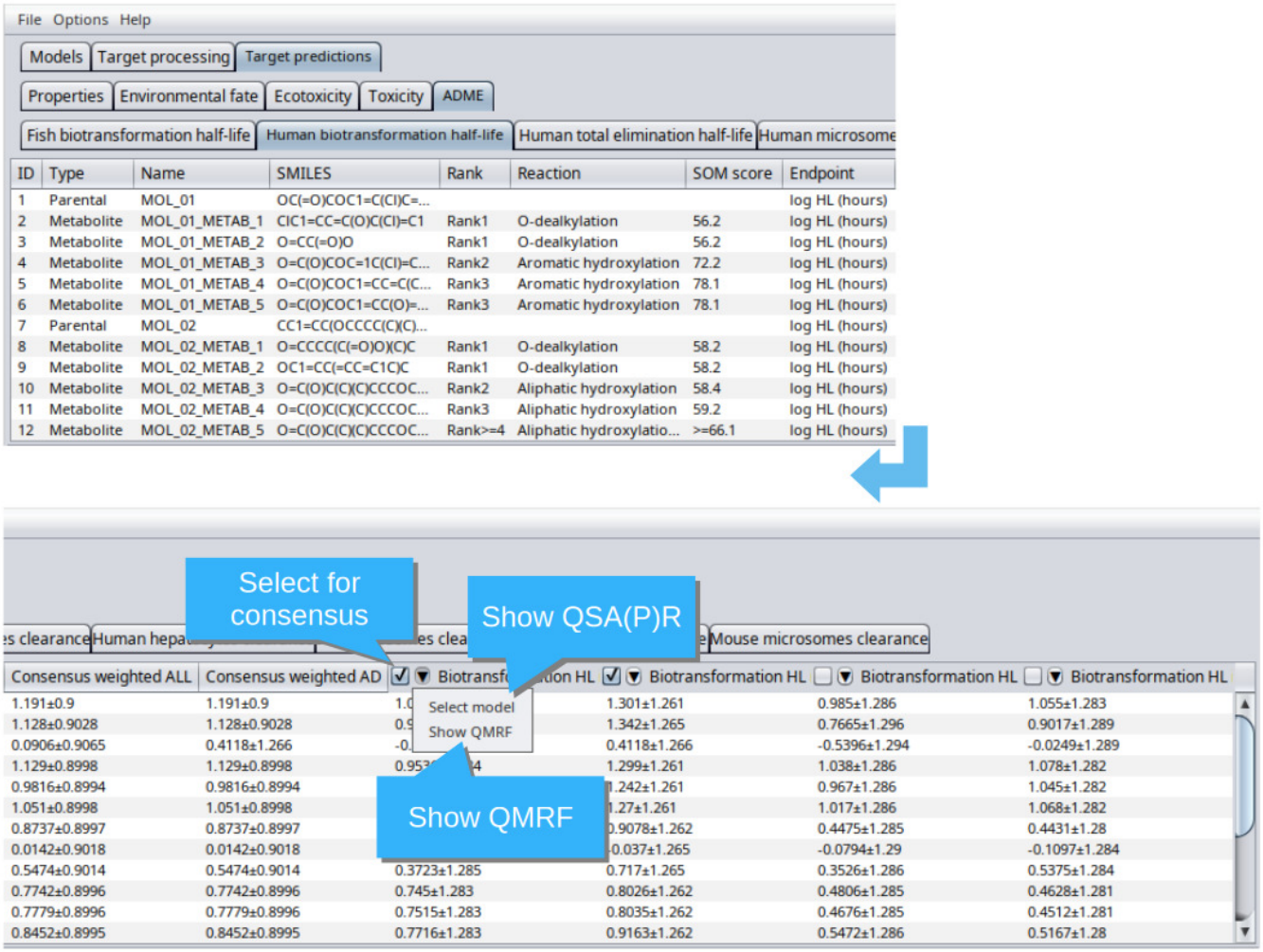

Figure 10: MLR QSARs prediction tables.

3.10 Other options

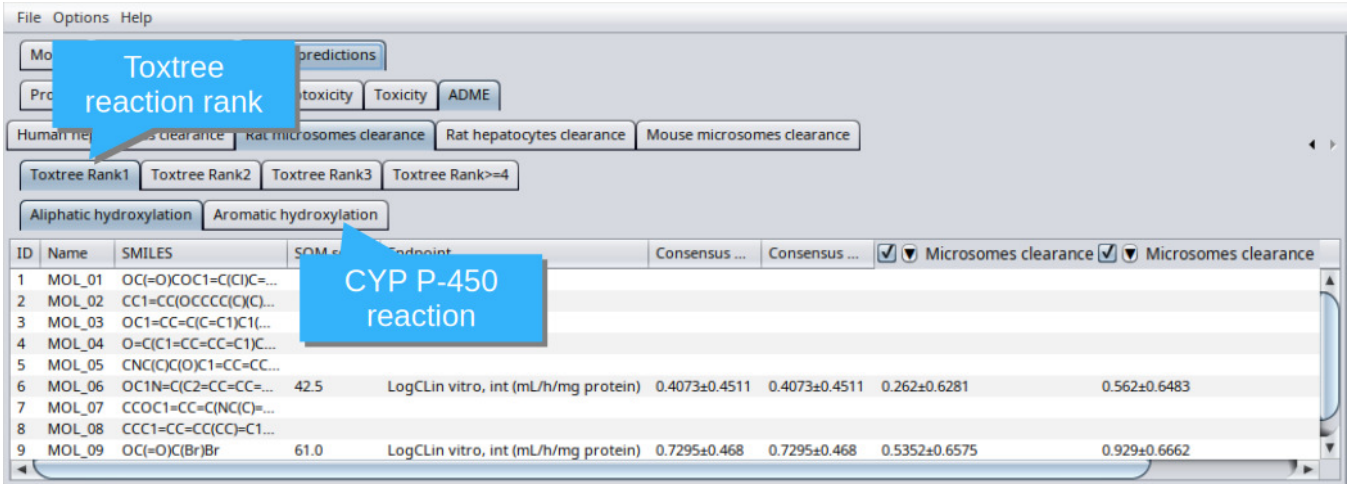

Figure 11: Prediction tables of cytochrome P-450 reaction based MLR QSARs.

3.10.1 Sessions

In QSAR-ME Profiler 2025 a session consists of the user-entered chemicals and the calculated predictions tables. Sessions can be saved by selecting File → Save session... or reloaded[18] by selecting File → Open session... from the File menu of the main window. Do drop an active session by starting a new one, select File → New session....

3.10.2 Charts

Charts can be zoomed in by dragging the mouse while keeping pressed the left mouse button, or by rotating the mouse wheel up, while charts can be zoomed out by rotating the mouse wheel down. Charts can be panned by pressing the CTRL (control) key while dragging the mouse. By right clicking the mouse over a chart, a popup menu will allow to copy (Copy item) and save (Save as item) the chart, to auto range (Auto range current chart option, which resizes the current chart to data point ranges), and reset all charts (Reset all charts option, which resets all charts to their original state). By accessing sub-menus from Options → Charts chart points color, shape and size can be customized.

3.10.3 Similarity

By accessing sub-menus from Options → Similarity, fingerprints and distances can be chosen to modify structural similarity detection in QSAR-ME Profiler 2025, see Section 3.3 for further details.

3.10.4 Significant digits

The number of significant digits, displayed on tables and QSARs information, can be changed by the Display → Significant digits... option.

[18] If the QSARs scheme is modified by the user (see Section 5) after saving a session, schemes must be consistent with the old ones otherwise reloading will trigger an error.

4 Running QSAR-ME Profiler 2025 as a command line application

QSAR-ME Profiler 2025 can be run as a command line application. When running in this mode no GUI is opened, while QSAR-ME Profiler automatically apply embedded and user-defined QSARs developed using QSAR-WB Descriptor descriptors (manual selection of reactions and user-defined QSARs developed using descriptors from external tools are not allowed). Results are saved to files, which are suitable for calculation pipelines. To get help from the command line type

```
java -jar QSAR-ME-Profiler.jar --help
```

the following output will be generated

```
Usage -i [SMILES FILE] -o [OUTPUT FILES NAME HEADER] ...
Example: 'java -jar QSAR-ME-Profiler.jar -i smiles.smi -o output

File options:
-i <file>    SMILES file (.smi format)
-o <file>    Output files name header
-a           Add detected metabolites (optional)
```

Option -i must be fulfilled and is used for loading chemical structures from a text file, as SMILES written in the .smi format [19]. Option -o must also be fulfilled, and is used for choosing the output files header. For example, if chosen header is output the following files are generated output_mlr.csv (MLR QSARs), output_mlr_toxtree.csv (MLR QSARs developed according to reactions detected by Toxtree) and output_lda.csv (LDA QSARs). Option -a is discretionary and is used for also applying QSARs to metabolites automatically detected by Toxtree.

5 QSAR-ME Profiler 2025 QSARs customization

QSAR-ME Profiler 2025 QSARs are encoded as XML (Extensible Markup Language) and CSV (Comma-Separated Values) files which can be modified using a text editor. Concerning XML files, it is recommended avoiding special characters, like &, < etc. which may conflict with the QSAR-ME Profiler 2025 XML reader. Therefore it is here suggested using simple alphanumeric characters and _ as words spacing. XML tags that are not discussed in this manual are intended to be left untouched[20]. ↪ symbols in the code listings means continuation of the previous line. Code syntax must be strictly applied and thoroughly checked, since even a misplaced or missing character could prevent QSAR-ME Profiler 2025 from running[21].

5.1 Configure QSAR categories

QSAR categories in QSAR-ME Profiler 2025 are collections of QSARs sharing the same endpoint in a certain group. Categories can be modified by editing the qsar_layout.xml file, located in the config folder, which is in the QSAR-ME Profiler 2025 main folder. MLR models developed using Toxtree or not, and LDA models, are the allowed categories type in QSAR-ME Profiler 2025, as shown in the following conceptual configuration file:

```
<?xml version="1.0" encoding="UTF-8"?>
<model_tree>
```

[19] SMILES strings are followed by one or more spaces/tabulations followed by the corresponding chemical name.

[20] Code to be left untouched is grayed in the following sections.

[21] When booting, QSAR-ME Profiler 2025 checks the configuration files and stops by warning the user in case of inconsistencies. However, there is no guarantee that all problems are detected.

```
  <qsar source="qsarme" descriptors="qsarme" group="Group 1" sub_group="Sub group 1" folder="qsars_1"
      ↪ type="mlr" toxtree="no" />
  <qsar source="qsarme" descriptors="qsarme" group="Group 1" sub_group="Sub group 2" folder="qsars_2"
      ↪ type="mlr" toxtree="yes" />
  <qsar source="qsarme" descriptors="qsarme" group="Group 1" sub_group="Sub group 3" folder="qsars_3"
      ↪ type="lda" toxtree="no" />
</model_tree>
```

A QSAR category begins with <qsar and ends with />, while code between these marks specifies the QSAR category. source= specifies whether QSARs are those embedded ("qsarme") or are user-defined ("userde"). descriptors= specifies how descriptors are calculated using one of the following options: "qsarme" (used only by QSARs shipped with QSAR-ME Profiler 2025), "qsarwb" for automatic calculation of QSARs developed using QSAR-WB Descriptor descriptors (this option is available both for QSARs shipped with QSAR-ME Profiler 2025 and for user-defined QSARs) and "extern" if descriptors values must be manually loaded (this option is available only for user-defined QSARs). group= creates a QSAR group which name is specified between quotation marks. sub_group= works similarly to group=, and creates a QSAR subgroup within a group. folder= specifies the folder containing the subgroup QSARs. Accoding to the source= option, the subgroup folder must be either located in qsar_me_profiler for embedded QSARs or in user_defined for user-defined QSARs (both folders are located in the qsar folder in the main QSAR-ME Profiler 2025 folder). type= specifies whether QSARs within a subgroup are multiple linear regressions ("mlr") or linear discriminant analysis ("lda"). toxtree= specifies whether the QSARs were developed according to the reactions and ranks calculated by Toxtree ("yes") or not ("no"). Within each group, subgroups must be different, and so the name of the corresponding folders. Below follows as example of a fictional configuration file.

```
<?xml version="1.0" encoding="UTF-8"?>
<model_tree>
  <qsar source="qsarme" descriptors="qsarme" group="Ecotoxicity" sub_group="Fish pLC50" folder="
      ↪ fish_acute_tox_plc50" type="mlr" toxtree="no" />
  <qsar source="qsarme" descriptors="qsarme" group="Ecotoxicity" sub_group="Algae pEC50" folder="
      ↪ algae_acute_tox_pec50" type="mlr" toxtree="no" />
  <qsar source="qsarme" descriptors="qsarme" group="Ecotoxicity" sub_group="Crustacean pEC50" folder=
      ↪ "crustacean_acute_tox_pec50" type="mlr" toxtree="no" />
  <qsar source="qsarme" descriptors="qsarme" group="Ecotoxicity" sub_group="Fish biomagnification (
      ↪ classification)" folder="fish_diet_bmf_lda" type="lda" toxtree="no" />
  <qsar source="qsarme" descriptors="qsarme" group="Metabolic transformations" sub_group="Human
      ↪ microsomes clearance" folder="invtr_human_mic" type="mlr" toxtree="yes"/>
  <qsar source="qsarme" descriptors="qsarme" group="Metabolic transformations" sub_group="Human
      ↪ hepatocytes clearance" folder="invtr_human_hep" type="mlr" toxtree="yes"/>
</model_tree>
```

The first group is Ecotoxicity which includes Fish pLC50, Algae pEC50, Crustacean pEC50 and Fish biomagnification (classification) subgroups. The first three subgroups contain multiple linear regressions QSARs while the last subgroup contains linear discriminant analysis QSARs. All QSARs do not require Toxtree. The second group is Metabolic transformations, which includes Human microsomes clearance and Human hepatocytes clearance subgroups. All QSARs are multiple linear regressions requiring Toxtree.

5.1.1 Add a user-defined QSAR

The following steps should be followed to add a user-defined a QSAR.

1. Locate and go to in the qsar folder within the main folder of QSAR-ME Profiler 2025.
2. Go to in the user_defined subfolder.
3. According on whether descriptors are calculated by a tool of choice or by QSAR-WB Descriptors, go to in the external_descr subfolder or qsar_wb_descr.

4. Create a new folder which will contain a QSARs subgroup.
5. Create the XML (extension .xml) and CSV (extension .csv) QSAR/s files.
6. Create the QMRF PDF file/s (extension .pdf) or use empty placeholder/s.
7. Copy the XML, CSV and PDF files in the folder of point 3.
8. Edit qsar_layout.xml accordingly (see Section 5.1).

Examples of already compiled files can be found in the examples folder within the main folder of QSAR-ME Profiler 2025.

5.1.2 QSAR files

QSARs are coded as XML files containing the model description and CSV files containing the corresponding dataset. Correspondence between XML and CSV files of a QSAR is dictated by the file name, which must be the same e.g., new_qsar.xml and new_qsar.csv. To help writing the XML files, the following templates are provided in the templates folder located in the main QSAR-ME Profiler 2025 folder.

- mlr_template.xml - multiple linear regression QSAR.
- mlr_toxtree_template.xml - multiple linear regression QSAR developed using Toxtree.
- lda_template.xml - linear discriminant analysis QSAR.

5.1.3 Create MLR QSAR xml and csv files

Let us assume we need to write files for an MLR QSAR called "New QSAR", to be placed in a QSAR subgroup called "New subgroup" within a QSAR group called "New group MLR". The following steps should be followed.

- Copy the mlr_template.xml file and rename as new_mlr_qsar.xml, then open this file in a text editor [22] .
- Locate the model section, then locate the item followed by type="name" and write New QSAR between the quotation marks of value=. See also example below.

```
<model>
    <item type="name" value="New QSAR"/>
</model>
```

- Locate the qmrf section, then locate the item followed by type="name" and write New QSAR QMRF between the quotation marks of value=. Locate now the item containing type="file" and write qmrf_placeholder.pdf[23] between the quotation marks of value=. See also example below.

```
<qmrf>
    <item type="name" value="New QSAR QMRF"/>
    <item type="file" value="qmrf_placeholder.pdf"/>
</qmrf>
```

[22] For text editing, software like Microsoft Word or LibreOffice are not the right tools. Use text editors like gedit (Linux), Notepad (Windows), TextEdit (macOS) or more advanced software for editing code like e.g., Kate, Notepad++ or Visual Studio Code.

[23] This is an empty placeholder PDF used instead of a compiled QMRF PDF file.

- Locate the description section and write a QSAR description between CDATA[and], like Description of the new QSAR in the example below.

```
<description>
    <![CDATA[Description of the new QSAR]]>
</description>
```

- Locate the <endpoint type="dataset"> section, then locate the item followed by type="name" and write pEC50 mol/L as the endpoint description between the quotation marks of value=. See also the example below.

```
<endpoint type="dataset">
    <item type="name" value="pEC50 mol/L"/>
</endpoint>
```

- Transformed endpoints can be configured in the optional section <endpoint type="transformation">. The transformation formula must be written as reverse polish notation (RPN). Accepted tokens in QSAR-ME Profiler 2025 are:

  N → number to be transformed

  W → weight to be applied

  + - * / → algebraic operations

  log, ln, exp, e, sqrt, tan, atan, sin, asin, cos, acos → functions

  Let us assume we need to transform pEC50 mol/L as EC50 mg/L. The conversion formula is $10^{-N} \cdot W \cdot 1000$, which is coded as RPN as N - exp W * 1000 *.

  Like the previous point, locate the item containing type="name" and write EC50 mg/L as the endpoint description between the quotation marks of value=, then locate the item containing type="rpn_formula" and write N - exp W * 1000 * between the quotation marks of value=, finally locate the item containing type="weight_type" and write molecular weight between the quotation marks of value=[24]. See also example below.

```
<endpoint type="transformation">
    <item type="name" value="EC50 mg/L"/>
    <item type="rpn_formula" value="N - exp W * 1000 *"/>
    <item type="weight_type" value="molecular weight"/>
</endpoint>
```

- Let now assume we developed, or took from the literature, an MLR QSAR with the following equation:

  pEC50 mol/L = 0.5473 - 0.1606 × C1SP2 + 0.1434 × minHBint4

  Locate the equation section type="normal", then locate the type="intercept" item and write the intercept value[25] between the value= quotation marks. Locate the <item type="coefficient" line and copy it below since we need two coefficients. Locate the type="coefficient" item, then write the first descriptor name between the name= quotation marks, and the corresponding value of the coefficient between the value= quotation marks. Go to the next line and do the same for the second descriptor. Repeat the same steps for the standardized coefficients by locating the equation section type="standardized". See also example below.

[24] Currently only molecular weight has been implemented in QSAR-ME Profiler 2025.

[25] Numerical values should conform to 64 bits double precision format IEEE 754, therefore strings should contain up to 15 significant digits, with a range of values approximately from $4.9 \cdot 10^{-324}$ to $1.8 \cdot 10^{308}$.

```
<equation type="normal">
    <item type="intercept" value="0.5473"/>
    <item type="coefficient" name="C1SP2" value="-0.1606"/>
    <item type="coefficient" name="minHBint4" value="0.1434"/>
</equation>

<equation type="standardized">
    <item type="coefficient" name="C1SP2" value="-0.7053"/>
    <item type="coefficient" name="minHBint4" value="0.6089"/>
</equation>
```

- Intercept and coefficients/ statistics are located in the confidence, significance (p-value) and standard_error statistic sections types. These sections should be compiled similarly to the equation sections, as in the example below.

```
<statistic type="confidence">
    <item type="intercept" value="0.2280"/>
    <item type="coefficient" name="C1SP2" value="0.06261"/>
    <item type="coefficient" name="minHBint4" value="0.06472"/>
</statistic>

<statistic type="significance">
    <item type="intercept" value="0.000009098"/>
    <item type="coefficient" name="C1SP2" value="0.00004420"/>
    <item type="coefficient" name="minHBint4" value="0.0002081"/>
</statistic>

<statistic type="standard_error">
    <item type="intercept" value="0.1075"/>
    <item type="coefficient" name="C1SP2" value="0.02953"/>
    <item type="coefficient" name="minHBint4" value="0.03053"/>
</statistic>
```

- Fitting performances are located in the performance section, type="fitting". <item type="value" name ="s" and <item type="value" name ="h*" must be filled since they are needed by some QSAR-ME Profiler 2025 calculations. "s" is the standard error of the estimate (i.e., the square root of the squared residuals mean) and "h*" is the leverage cut-off value (see Section 3.6). Further items are optional and can be any, since they are only displayed in QSAR-ME Profiler 2025 and do not involve any calculation. Here it follow an example with obligatory and optional performance measures[26].

```
<performance type="fitting">
    <item type="value" name="R2" value="0.7372"/>
    <item type="value" name="R2adj" value="0.7044"/>
    <item type="value" name="RMSE" value="0.1858"/>
    <item type="value" name="MAE" value="0.1470"/>
    <item type="value" name="CCC" value="0.8487"/>
    <item type="value" name="F" value="22.44"/>
    <item type="value" name="s" value="0.2025"/>
    <item type="value" name="h*" value="0.4737"/>
</performance>
```

- Cross validation performances are located in the performance section, type="cross_validation". This section, which is optional, works similarly to the fitting performances section, see the example below.

[26] Measures are shown in QSAR-ME Profiler 2025 in the same order of the performance section.

```
<performance type="cross_validation">
    <item type="value" name="Q2LOO" value="0.6580"/>
    <item type="value" name="RMSE" value="0.1736"/>
    <item type="value" name="CCC" value="0.7973"/>
</performance>
```

- The dataset section links the XML definition of a QSAR to its training dataset file, which must be written in the CSV format (items must be separated by comma). Below follows the <dataset> section of the QSAR example.

```
<dataset>
    <item type="file" value="new_mlr_qsar.csv"/>
    <item type="object" value="Name"/>
    <item type="smiles" value="SMILES"/>
    <item type="exp_endpoint" value="Exp. endpoint"/>
    <item type="pred_endpoint" value="Pred. endpoint"/>
    <item type="residual" value="Residual"/>
    <item type="std_residual" value="Std. Residual"/>
    <item type="hat" value="Hat"/>
    <item type="descriptor" value="C1SP2"/>
    <item type="descriptor" value="minHBint4"/>
</dataset>
```

The first item, <item type="file", is used to link QSAR-ME Profiler 2025 to the CSV file, whose file name, in this example, could be new_mlr_qsar.csv[27]. The file name must be between the quotation marks of the corresponding value= item, as shown above. Below follows part of a fictional example of the CSV dataset file content (SMILES are shortened using three dots to make the example more compact.) Note: if a chemical name contains comma/s, like e.g. 2,4-Diaminotoluene, the field must be embraced by quotation marks like "2,4-Diaminotoluene", otherwise QSAR-ME Profiler 2025 will consider commas within chemical names as field separators [28]. Below follows a CSV file [29] example.

```
Name,        SMILES,      Exp. pEC50, Pred. pEC50, Resid.,  Std. Resid., Hat,    Mp,     GGI8
000078-93-3, CCC(=O)C,    2.1587,     2.3373,      0.1786,  0.4659,      0.2075, 0.5903, 0.0000
000100-41-4, CCc1ccccc1,  4.4693,     3.9986,     -0.4708, -1.1493,      0.0949, 0.6663, 0.0000
000100-42-5, C=Cc1ccccc1, 5.1600,     4.7288,     -0.4311, -1.0702,      0.1244, 0.6996, 0.0000
000124-18-5, CCCCCCCCCC,  3.2034,     2.5863,     -0.6171, -1.5898,      0.1871, 0.5870, 0.0247
```

Concerning the XML file, the text between the quotation marks of value= of the items from <item type="object" to the last <item type="descriptor" are links to the columns of the CSV file. The value= text between the quotation marks can be any, so far it matches the name of the corresponding column in the CSV file i,e.

<item type="object" links to the names of the chemicals.

<item type="smiles" links to the SMILES of the chemicals.

<item type="exp_endpoint" links to the experimental endpoint values.

<item type="pred_endpoint" links to the predicted endpoint values.

<item type="residual" links to the residuals of the endpoint.

[27] For CSV files it is recommended to use the same file name (excluding the extension) of the XML file, which is, in this example, new_qsar.xml).

[28] Note also that, concerning labels and names, quotation marks and spaces count therefore, for example, "000078-93-3" is different from 000078-93-3 and □000078-93-3 is different from 000078-93-3.

[29] Items are here aligned to simplify reading. In production CSV files, items should be separated only by the separator char (usually comma).

<item type="std_residual" links to the standardized residuals of the endpoint.

<item type="hat" links hat (leverage) values of the chemicals.

<item type="descriptor" links to a descriptor of the chemicals.

Additional columns in the CSV file containing data not referenced by the dataset section are ignored by QSAR-ME Profiler 2025, therefore can be left in place if preferred.

- Create a new folder called user_qsar in the user_defined folder, within the qsar folder located in the main folder of QSAR-ME Profiler 2025. Drop the XML, CSV and PDF files in this folder, then edit the configuration file as described in Section 5.1.1 by adding the following configuration line <qsar source="userde" group="User group" sub_group="User subgroup" folder="user_qsar" type="mlr" toxtree="no"/>
- Run QSAR-ME Profiler 2025, which will automatically build supporting data for internal use.

5.1.4 Create MLR QSAR xml and csv files - Toxtree version

QSAR-ME Profiler 2025 can run Toxtree (using the SMARTCyp module Cytochrome P450-Mediated Drug Metabolism and metabolites prediction plugin) to automatically select QSARs developed according to their potential reactivity, based on putative cytochrome P450 (CYP) mediated reactions.

- The configuration of the XML file is very similar to Section 5.1.3. After specifying the name item, additional model items must be filled i.e., organism, assay, rank and reaction. See also example below.

```
<model>
   <item type="name" value="Microsomes clearance in rat (alcohol oxidation) - MLR 01"/>
   <item type="organism" value="Rat"/>
   <item type="assay" value="Microsomes"/>
   <item type="rank" value="Rank1"/>
   <item type="reaction" value="Alcohol oxidation"/>
</model>
```

- Organisms and assays are user-defined and can be any, while ranks and reactions are ruled by the Toxtree SMARTCyp SDF output, as shown in the SDF portion below.

```
> <SMARTCyp.Rank1.Reaction>
O-dealkylation
```

In the SDF files, the rank name is reported between the two fullstops file i.e., Rank1 in this example. The reaction item is reported in the line below in the SDF file i.e., O-dealkylation in this example. To configure this section, write the organism name between the quotation marks of value= of the <item type="organism" item, the assay name between the quotation marks of value= of the <item type="assay" item, the Toxtree rank between the quotation marks of {<item type="rank" and the Toxtree reaction name between the quotation marks of value= of the <item type="reaction" item.

- Create the CSV file, as explained in section 5.1.3
- Create a new folder called user_qsar in the user_defined folder, within the qsar folder located in the main folder of QSAR-ME Profiler 2025. Drop the XML, CSV and PDF files in this folder, then edit the configuration file as described in Section 5.1.1 by adding the following configuration line, then edit the configuration file as explained in Section 5.1.1, by adding the following line <qsar source="userde" group="User group" sub_group="User subgroup" folder="user_qsar_tox" type="mlr" toxtree="yes"/>.
- Run QSAR-ME Profiler 2025, which will automatically build supporting data for internal use.

5.1.5 Create LDA QSAR xml and csv files

The first part of the configuration of an LDA QSAR XML file is similar to the MLR XML configuration explained in Section 5.1.3.

- Below follows an example of model, qmrf and description sections, for a fictional QSAR named New QSAR LDA.

```
<model>
    <item type="name" value="New QSAR LDA"/>
</model>

<qmrf>
    <item type="name" value="New QSAR QMRF"/>
    <item type="file" value="qmrf_placeholder.pdf"/>
</qmrf>

<description>
    <![CDATA[Description of the new QSAR]]>
</description>
```

- The endpoint section is textual and must be written similarly to the description section between the CDATA square brackets, as in the following example which assumes three classes, so the endpoint description could be Three classes endpoint, for the fictional QSAR.

```
<endpoint>
    <![CDATA[Three classes endpoint]]>
</endpoint>
```

- Locate the equation section, type="discrim_function". For each class, the intercept and the coefficient/s of the discriminant functions must be written. First locate the intercept item (<item type="intercept"), then type the class name between the quotation marks of the class= item and type the intercept value between the quotation marks of value=. Do the same for the coefficients (three in this example) by locating the coefficient items (<item type="coefficient"), and then write the descriptor name between the quotation marks of the name= item followed by the coefficient value between the quotation marks of value= item, as in the example below.

```
<equation type ="discrim_function">
    <item type="intercept" class="1" value="-11.52"/>
    <item type="coefficient" class="1" name="MW" value="15.09"/>
    <item type="coefficient" class="1" name="Mp" value="29.61"/>
    <item type="coefficient" class="1" name="MAXDP" value="-6.680"/>
    <item type="intercept" class="2" value="-1.860"/>
    <item type="coefficient" class="2" name="MW" value="3.266"/>
    <item type="coefficient" class="2" name="Mp" value="10.74"/>
    <item type="coefficient" class="2" name="MAXDP" value="2.097"/>
    <item type="intercept" class="3" value="-2.670"/>
    <item type="coefficient" class="3" name="MW" value="-6.085"/>
    <item type="coefficient" class="3" name="Mp" value="8.024"/>
    <item type="coefficient" class="3" name="MAXDP" value="10.96"/>
</equation>
```

- Locate the in <item type="value" item within the probability section, then locate and write the prior probability of each class between the quotation marks of the class= item, and the corresponding probability value in the subsequent value=, as in the example below.

```
<probability type="prior">
    <item type="value" class="1" value="0.2638"/>
```

```
    <item type="value" class="2" value="0.4584"/>
    <item type="value" class="3" value="0.2778"/>
</probability>
```

- Fitting performances are optional, since they are shown but are not used for calculations in QSAR-ME Profiler 2025. Locate the <item type="value" item within the fitting section, then write the name of the fitting measure between the quotation marks of the name= item and the corresponding value between the quotation marks of the value= item, as in the example below.

```
<performance type="fitting">
    <item type="value" name="Accuracy" value="0.8056"/>
    <item type="value" name="No Information Rate" value="0.4583"/>
    <item type="value" name="P-Value [Acc greater than NIR]" value="1.553e-09"/>
    <item type="value" name="Sensitivity class 1" value="0.8947"/>
    <item type="value" name="Specificity class 1" value="0.9623"/>
    <item type="value" name="Sensitivity class 2" value="0.7273"/>
    <item type="value" name="Specificity class 2" value="0.8718"/>
    <item type="value" name="Sensitivity class 3" value="0.8500"/>
    <item type="value" name="Specificity class 3" value="0.8654"/>
</performance>
```

- The dataset section is similar to the one explained in Section 5.1.3. The first item, <item type="file", is used to link QSAR-ME Profiler 2025 to the CSV file, whose file name (in this fictional example new_lda_qsar.csv) must be typed between the quotation marks of the corresponding value=. The text between the quotation marks of value= of the items from <item type="object" to the last <item type="descriptor" are links to the CSV file columns. The value= text between the quotation marks can be any, so far it matches the name of the corresponding column in the CSV file.

  <item type="object" links to the names of the chemicals.

  <item type="smiles" links to the SMILES of the chemicals.

  <item type="exp_endpoint" links to the experimental endpoint values.

  <item type="post_prob" links to the post probability values.

  <item type="descriptor" links to a descriptor if the chemicals.

  Here it follows an example of a dataset section.

```
<dataset>
    <item type="file" value="new_lda_qsar.csv"/>
    <item type="object" value="CAS"/>
    <item type="smiles" value="SMILES"/>
    <item type="exp_endpoint" value="Exp. class"/>
    <item type="pred_endpoint" value="Pred. class"/>
    <item type="post_prob" class="1" value="Post. prob. 1"/>
    <item type="post_prob" class="2" value="Post. prob. 2"/>
    <item type="post_prob" class="3" value="Post. prob. 3"/>
    <item type="descriptor" value="MW"/>
    <item type="descriptor" value="Mp"/>
    <item type="descriptor" value="MAXDP"/>
</dataset>
```

- Create the CSV file, as explained in section 5.1.3

- Create a new folder called user_qsar in the user_defined folder, within the qsar folder located in the main folder of QSAR-ME Profiler 2025. Drop the XML, CSV and PDF files in this folder, then edit the configuration file as described in Section 5.1.1 by adding the following configuration line

<qsar source="userde" group="User group" sub_group="User subgroup" folder="user_qsar_lda" type="lda" toxtree="no"/>

- Run QSAR-ME Profiler 2025, which will automatically build supporting data for internal use.

6 QSARs shipped with QSAR-ME Profiler 2025

Table 1: QSARs shipped with QSAR-ME Profiler 2025

Group	Sub-group	Available models
Properties	Partition coefficients	1 x Soil organic carbon-water partition [8,9]
	Global Warming Potential (GWP)	1 x Global Warming Potential (GWP)
Environmental fate	Half-life	1 x Global Half-Life Index (GHLI) [8, 10]
	PBT	1 x Insubria PBT index [11, 12]
Ecotoxicity	Fish pLC50	1 x O. mykiss [13, 14]
		3 x P. promelas [13, 14]
	Algae pEC50	2 x P. subcapitata [13, 14]
	Crustacean pEC50	1 x D. Magna [13, 14]
	Fish biomagnification	1 x Dietary biomagnification factor (BMF) [15]
	Fish bioconcentration	1 x Dietary bioconcentration factor (BCF) [15]
	Fish bioconcentration (classification)	1 x Dietary bioconcentration factor (BCF) [16]
Toxicity	Human transthyretin disruption	1 x ANSA binding affinity [17]
		1 x FITC-T4 binding affinity [17]
		1 x T4-hTTR competing potency [17]
		1 x PFAS httr disruption [18]
	Human transthyretin disruption (classification)	1 x PFAS httr disruption [18]
ADME	Fish biotransformation half-life	2 x metabolic biotransformation [19]
	Human biotransformation half-life	4 x Whole-body [19]
	Human total elimination half-life	1 x Whole body [19]
	Human microsomes clearance	32 x in vitro clearance [20]
	Human hepatocytes clearance	8 x in vitro clearance [20]
	Rat microsomes clearance	13 x in vitro clearance [20]
	Rat hepatocytes clearance	6 x in vitro clearance [20]
	Mouse microsomes clearance	14 x in vitro clearance [20]

7 Acknowledgments

We acknowledge the University of Insubria for funding the post doc grant "In silico solutions for the assessment of biotransformation related endpoints of organic chemicals in multiple organisms" (2021-2022), to Dr. Nicola Chirico.

8 References

[1] Ideaconsult Ltd. Toxtree version 3.1.0. https://toxtree.sourceforge.net.

[2] Waldemar Klingspohn, Miriam Mathea, Antonius Ter Laak, Nikolaus Heinrich, and Knut Baumann. Efficiency of different measures for defining the applicability domain of classification models. *Journal of cheminformatics*, 9:1–17, 2017.

[3] Iurii Sushko, Sergii Novotarskyi, Robert Körner, Anil Kumar Pandey, Artem Cherkasov, Jiazhong Li, Paola Gramatica, Katja Hansen, Timon Schroeter, Klaus-Robert Müller, et al. Applicability domains for classification problems: benchmarking of distance to models for ames mutagenicity set. *Journal of chemical information and modeling*, 50(12):2094–2111, 2010.

[4] Miriam Mathea, Waldemar Klingspohn, and Knut Baumann. Chemoinformatic classification methods and their applicability domain. *Molecular Informatics*, 35(5):160–180, 2016.

[5] Patrik Rydberg, David E Gloriam, Jed Zaretzki, Curt Breneman, and Lars Olsen. Smartcyp: a 2d method for prediction of cytochrome p450-mediated drug metabolism. *ACS medicinal chemistry letters*, 1(3):96–100, 2010.

[6] Julian James Faraway. *Practical regression and ANOVA using R.*, volume 168. University of Bath, 2002.

[7] J.R. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, 2022.

[8] Paola Gramatica, Stefano Cassani, and Nicola Chirico. Qsarins-chem: Insubria datasets and new qsar/qspr models for environmental pollutants in qsarins. *Journal of computational chemistry*, 35(13):1036–1044, 2014.

[9] Paola Gramatica, Elisa Giani, and Ester Papa. Statistical external validation and consensus modeling: a qspr case study for koc prediction. *Journal of Molecular Graphics and Modelling*, 25(6):755–766, 2007.

[10] Paola Gramatica and Ester Papa. Screening and ranking of pops for global half-life: Qsar approaches for prioritization based on molecular structure. *Environmental science & technology*, 41(8):2833–2839, 2007.

[11] Ester Papa and Paola Gramatica. Qspr as a support for the eu reach regulation and rational design of environmentally safer chemicals: Pbt identification from molecular structure. *Green Chemistry*, 12(5):836–843, 2010.

[12] Ester Papa, Fulvio Villa, and Paola Gramatica. Statistically validated qsars, based on theoretical descriptors, for modeling aquatic toxicity of organic chemicals in pimephales p romelas (fathead minnow). *Journal of chemical information and modeling*, 45(5):1256–1266, 2005.

[13] Paola Gramatica, Stefano Cassani, and Alessandro Sangion. Aquatic ecotoxicity of personal care products: Qsar models and ranking for prioritization and safer alternatives' design. *Green Chemistry*, 18(16):4393–4406, 2016.

[14] Alessandro Sangion and Paola Gramatica. Hazard of pharmaceuticals for aquatic environment: prioritization by structural approaches and prediction of ecotoxicity. *Environment International*, 95:131–143, 2016.

[15] Linda Bertato, Nicola Chirico, and Ester Papa. Qsar models for the prediction of dietary biomagnification factor in fish. *Toxics*, 11(3):209, 2023.

[16] Linda Bertato, Olivier Taboureau, Nicola Chirico, and Ester Papa. Classification-based qsars for predicting dietary biomagnification in fish. *SAR and QSAR in Environmental Research*, 33(4):259–271, 2022.

[17] Marco Evangelista, Nicola Chirico, and Ester Papa. In silico models for the screening of human transthyretin disruptors. *Journal of Hazardous Materials*, 480:136188, 2024.

[18] Marco Evangelista, Nicola Chirico, and Ester Papa. New qsar models to predict human transthyretin disruption by per-and polyfluoroalkyl substances (pfas): Development and application. *Toxics*, 13(7):590, 2025.

[19] Ester Papa, Alessandro Sangion, Jon A Arnot, and Paola Gramatica. Development of human biotransformation qsars and application for pbt assessment refinement. *Food and chemical toxicology*, 112:535–543, 2018.

[20] IVBP-Suite: In vitro biotransformation prediction suite. https://dunant.dista.uninsubria.it/qsar.