

Guide to ToxRead (0.17 BETA)

1. Introduction

ToxRead is a tool to assist user in making reproducible read-across evaluations. It processes the given target molecule by applying some sets of rules / structural alerts and extracting similar molecules with available experimental data from its database.

The results are given to the user as an interactive chart where it is possible to browse and examine the various rules and similar molecules found.

2. Installation

ToxRead is a Java application. It does not require installation, after the download the user has to unzip the downloaded package, and run the file ToxRead.jar with the standard Java Runtime Environment (on most systems, this can be done by simply double clicking the file). On Linux systems, the application can be also started by launching the ToxRead.sh script.

It is recommended to have installed the latest version of Java (check the website www.java.com for more details).

3. How to use

Once the application is started, the user must insert the target molecule, as a SMILES string. Optionally, the user can modify the number of similar compounds that will be shown (3 by default). The user must choose the endpoint of interest, and then the calculation will be started by clicking the “Run read-across” button.

After the calculation has been performed, the interactive chart will be shown. The right part of the chart reports a depiction of the target compound, its molecular weight and LogP (experimental value, if found in the database, otherwise a prediction performed with the VEGA model) and the list of rules that apply to it. The left part reports the graph itself.

In the graph, chemicals are shown as circles. The target chemical is in the center, while the similar chemicals are filled with a different color depending on their experimental activity. The size of the circle is proportional to the similarity with the target compound, which is calculated with the similarity index implemented within VEGA. Similarity goes from 1 (identity) to 0. If the similarity value is lower than 0.75, the two chemicals are quite different; thus the relevance of the similar compound is probably not high, and this requires further examination of other chemicals to support the overall evaluation.

For each chemical, its CAS number is shown. If the same chemical appears more than once in the graph, the circle is drawn with a dashed line. If the user moves the pointer over a molecule that appears more than once, all the occurrences of the same molecule in the graph will be highlighted in yellow. The user can see the structure of the chemical of interest double clicking on the circle. A form will be shown, with the structure, the similarity value, the experimental value for the selected endpoint and some additional experimental values for other properties, if found in the database.

The rules / structural alerts associated with the target chemical are represented with other shapes (depending on the endpoint) by triangles, and filled with colors depending on the meaning of the

rules. The user can visualize details on the rule by double clicking it. A form will be shown, with a depiction of the rule (for instance, the structure of the alert), the name and meaning of the rule, some statistical indices for the rule, and finally it is possible to visualize the most similar molecules (up to 100) that share the rule with the target compound.

It is common to have conflicting results: similar compounds which are both toxic and non toxic, for instance, or the presence of both toxicity and non toxicity structural alerts. In this case ToxRead helps the user in taking the decision. The user should evaluate the similarity of the related chemicals, look at the structures and evaluate the similarity index. More relevance should be given to the most similar compounds, and particular attention is necessary if the similarity is below 0.75. The evaluation should keep into account the occurrence of rules / structural alerts in common (or not) between the target compound and the similar chemicals, and this also is given by ToxRead.

4. Endpoint: Mutagenicity (Ames test)

For the Mutagenicity (Ames test) endpoint, four rule sets are available: Benigni/Bossa (as available in ToxTree v. 1), SARpy rules (extracted by Politecnico di Milano, with the automatic tool SARpy), IRFMN rules (extracted by human experts at Istituto di Ricerche Farmacologiche Mario Negri) and CRS4 rules (extracted by CRS4 Institute with automatic tools).

Molecules are depicted only with two colors, red and green, on the basis of their mutagenicity experimental activity (mutagenic/non-mutagenic). Rules are represented with red inverted triangles (with the tip pointing down) for mutagenic rules and green triangles for non-mutagenic rules. The intensity of the red (or green) is related to the percentage of toxic or non toxic chemicals (accuracy) for which the rule applies found in the ToxRead database. The size of the triangles is proportional to the Fisher Test P-Value of the rule, i.e. rules that have less statistical significance (thus reliability) have smaller triangles.

Rules in each rule set are organized with hierarchies. A rule can be a parent for another rule, if it is a more generic alert that contains the latter. Only the most specific rules are initially shown; when they have some parent rules available, they will be connected to a plus sign, and clicking on it the tree of the parent rules will be shown (where all parent rules are connected, on the basis of their hierarchy).



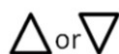
target compound



intensity of red related to the value of mutagenicity



intensity of green related to the value of non-mutagenicity



non-toxicity or toxicity rule, size related to the Fisher Test P-Value of the rule



circle size related to the similarity with the target compound



repeated compound

For the Mutagenicity endpoint, an integrated prediction is provided and on its basis the target molecule is depicted in green or red (mutagenic or non-mutagenic prediction), or in yellow if the prediction is conflicting. This prediction is based on two assessments: an automatic calculation of a read-across prediction based on the output of ToxRead, and a consensus between different QSAR models available in the VEGA software.

The read-across prediction is performed on the basis of the similar molecules found by ToxRead, where for each alert are not considered the molecules with the experimental value in disagreement with the toxicity reported by the alert. This means that each similar molecule can be counted more than one time (if it is reported as a similar compound in different alert), and this is considered the cardinality of the compound. Based on this list of compounds, two scores are calculated for mutagenic and non-mutagenic activities: each score is the sum of the cardinality multiplied by the similarity for each compound having the same toxicity activity. The final scores are then normalized (in the range between 0 and 1) and the score with the highest value represents the read-across prediction.

The QSAR prediction is a consensus between four available QSAR models from the open-source software tool VEGA (version 1.1.2). More details about the models and the integration approach can be found directly in the VEGA software, available from the website www.vega-qsar.eu

5. Endpoint: BCF

For the BCF endpoint a rule set developed by Istituto di Ricerche Farmacologiche Mario Negri is available, and in addition to these rules a logP/BCF interpolation chart is available for the target molecule and for each rule.

Molecules are depicted with different shades of color, according to a scale reported on the chart, where green is related to low BCF values, red is related to high BCF values, and yellow is related to intermediate values (around 3.0 log units). Rules are represented with rectangles, and their color corresponds to the mean experimental value calculated on all molecules in the ToxRead database for which the rule applies. In the details of the rule, also the standard deviation is reported, along with the number of molecules in the database for which the rule applies.

Some rules have a different depiction, an inverted triangles (with the tip pointing down). These rules represent an upper BCF threshold, found for all the compounds sharing a particular physico/chemical property (logP or molecular weight lower or greater than a fixed value).

The logP/BCF interpolation chart reports the most similar molecules with their logP value (experimental, if available, otherwise predicted with the VEGA model) versus their experimental BCF value. A dotted vertical line is drawn in correspondence with the logP value of the target compound. Such a chart can be useful to evaluate the relationship and the trend between the logP and BCF values for the most similar compounds.



target compound



red shades related to high BCF values, yellow to intermediate values (around 3.0 log units) green to low values



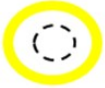
rules related with a mean experimental value for the compounds



rules that represent an upper threshold for the property



circle size related to the similarity with the target compound



repeated compound



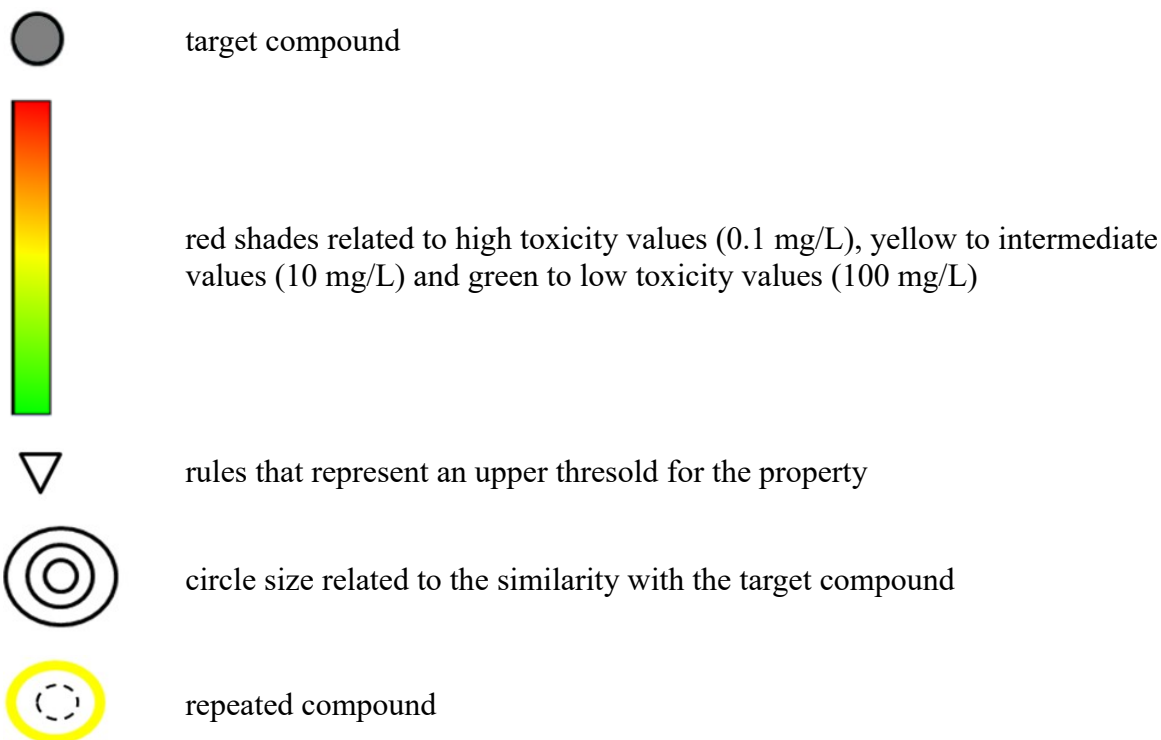
graph

6. Endpoint: Algae acute toxicity (biocides)

For the Algae acute toxicity endpoint, a rule set developed by IRFMN - IRCCS is available. The rule set was extracted using the tool SARpy, applied to biocides.

Molecules are depicted with different shades of color, according to a scale reported on the chart, where green is related to low toxicity, red is related to high toxicity, and yellow is related to medium toxicity. Rules are represented with triangles, and their color corresponds to the mean experimental value related to all molecules in the ToxRead database for which the rule applies. In the details of the rule (clicking on the rule), the experimental accuracy is also reported, along with the number of molecules in the database for which the rule applies.

Triangles are represented with the tip pointing down or up. In the first case, tip pointing down, the rule represents an upper threshold, found for all the compounds sharing a particular physico/chemical property. If the tip is pointing up, the rule represents a lower threshold, found for all the compounds sharing a particular chemical/physical property.

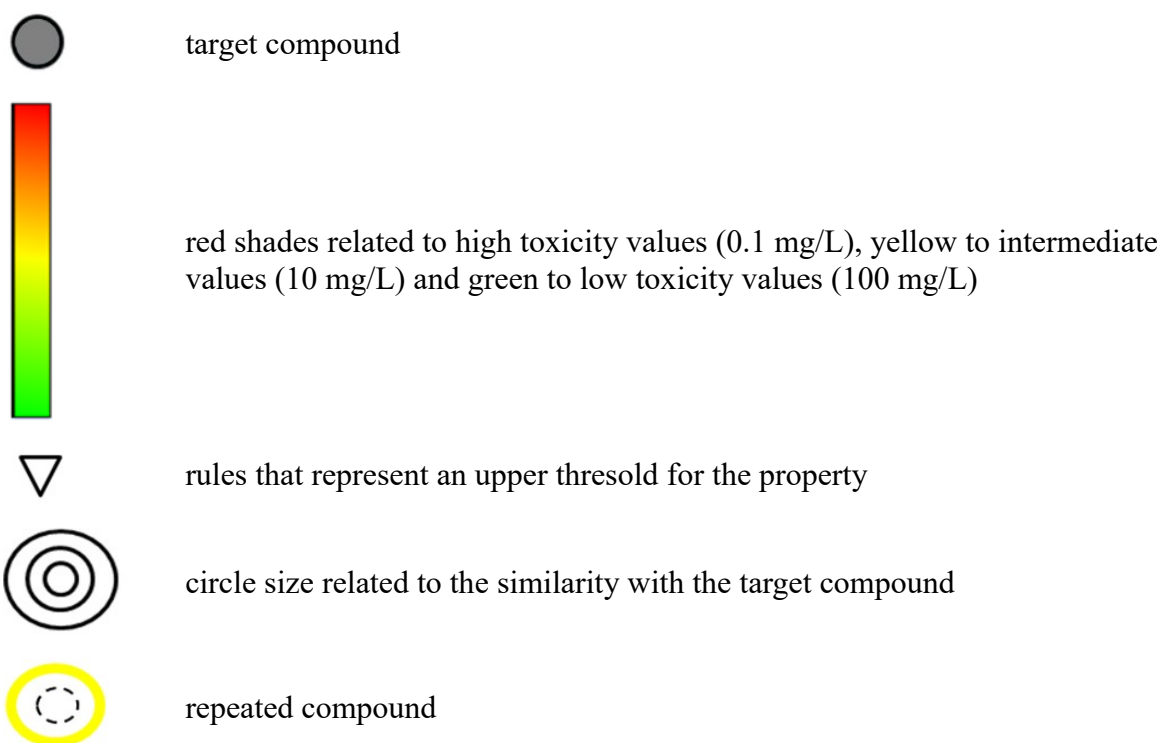


7. Endpoint: Microbial toxicity (biocides)

For Microbial toxicity endpoint, a rule set developed by IRFMN - IRCCS is available. The rule set was extracted using the tool SARpy, applied to biocides.

Molecules are depicted with different shades of color, according to a scale reported on the chart, where green is related to low toxicity, red is related to high toxicity, and yellow is related to medium toxicity. Rules are represented with triangles, and their color corresponds to the mean experimental value related to all molecules in the ToxRead database for which the rule applies. In the details of the rule (clicking on the rule), the experimental accuracy is also reported, along with the number of molecules in the database for which the rule applies.

Triangles are represented with the tip pointing down or up. In the first case, tip pointing down, the rule represents an upper threshold, found for all the compounds sharing a particular physico/chemical property. If the tip is pointing up, the rule represents a lower threshold, found for all the compounds sharing a particular chemical/physical property.

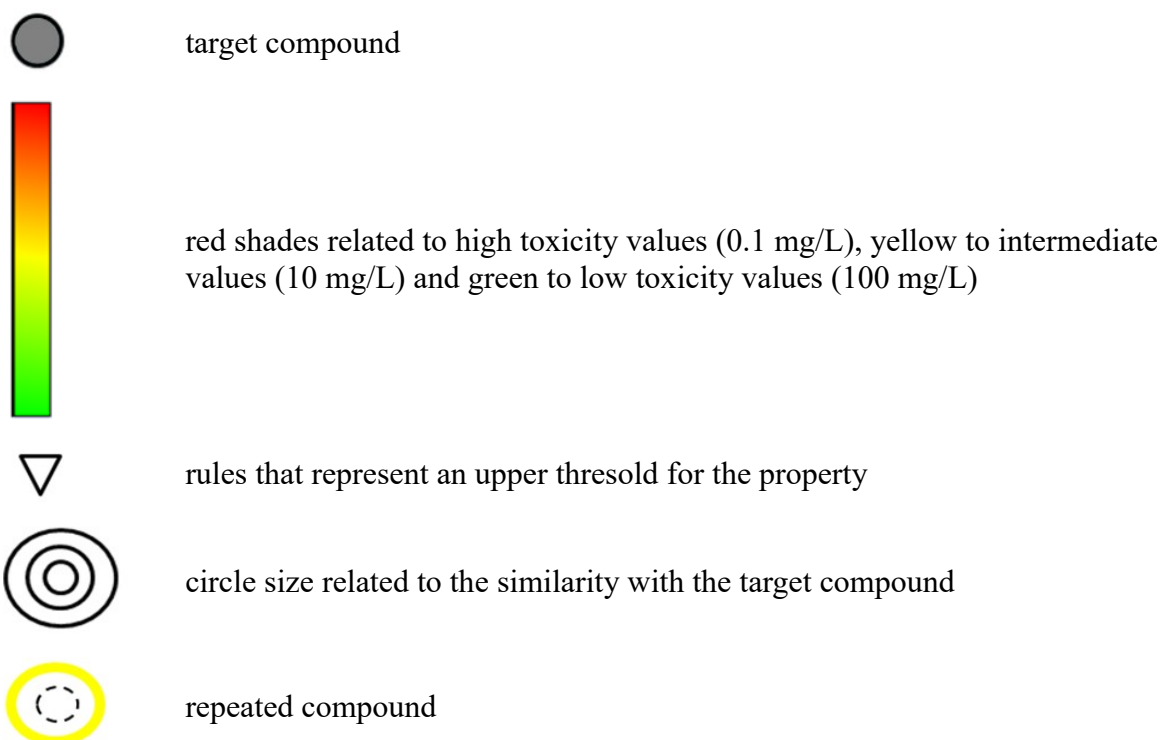


8. Endpoint: Fish acute toxicity (biocides)

For Fish toxicity endpoint, a rule set developed by IRFMN - IRCCS is available. The rule set was extracted using the tool SARpy, applied to biocides.

Molecules are depicted with different shades of color, according to a scale reported on the chart, where green is related to low toxicity, red is related to high toxicity, and yellow is related to medium toxicity. Rules are represented with triangles, and their color corresponds to the mean experimental value related to all molecules in the ToxRead database for which the rule applies. In the details of the rule (clicking on the rule), the experimental accuracy is also reported, along with the number of molecules in the database for which the rule applies.

Triangles are represented with the tip pointing down or up. In the first case, tip pointing down, the rule represents an upper threshold, found for all the compounds sharing a particular physico/chemical property. If the tip is pointing up, the rule represents a lower threshold, found for all the compounds sharing a particular chemical/physical property.



9. Endpoint: *Daphnia magna* acute toxicity (biocides)

For *Daphnia magna* acute toxicity endpoint, a rule set developed by IRFMN - IRCCS is available. KNIME Analytical Platform was used by finding the relevant fragments from data. In particular, RDKit and MoSS methods were used. Biocides have been used to obtain the rules.

Molecules are depicted with different shades of color, according to a scale reported on the chart, where green is related to low toxicity, red is related to high toxicity, and yellow is related to medium toxicity. Rules are represented with triangles, and their color corresponds to the mean experimental value related to all molecules in the ToxRead database for which the rule applies. In the details of the rule (clicking on the rule), the experimental accuracy is also reported, along with the number of molecules in the database for which the rule applies.

Triangles are represented with the tip pointing down or up. In the first case, tip pointing down, the rule represents an upper threshold, found for all the compounds sharing a particular physico/chemical property. If the tip is pointing up, the rule represents a lower threshold, found for all the compounds sharing a particular chemical/physical property.

