QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

User Manual

Tips and Tricks

Document history

Version	Comment
Version 1.0	December 2010: Tips and Tricks for Toolbox version 2.0
Version 1.1	February 2011: Amended section 3

Issue date: February 2011

Language: English

If you have questions or comments that relate to this document, please send them to ehscont@oecd.org or visit the QSAR Toolbox discussion forum at https://community.oecd.org/community/toolbox_forum.

Table of Contents

Foreword

This document is one of a series of guidance documents accompanying the OECD QSAR Toolbox Version 2.0. The main objective of the Toolbox is to allow the user to use (Q)SAR methodologies to group chemicals into categories and to fill data gaps by read-across, trend analysis and (Q)SARs. For in-depth background information on the concept of chemical categories, the user is invited to consult the guidance document for grouping of chemicals published in the Series on Testing and Assessment of the OECD Environment, Health and Safety Publications [OECD (2007); ENV/JM/MONO(2007)28: http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)28].

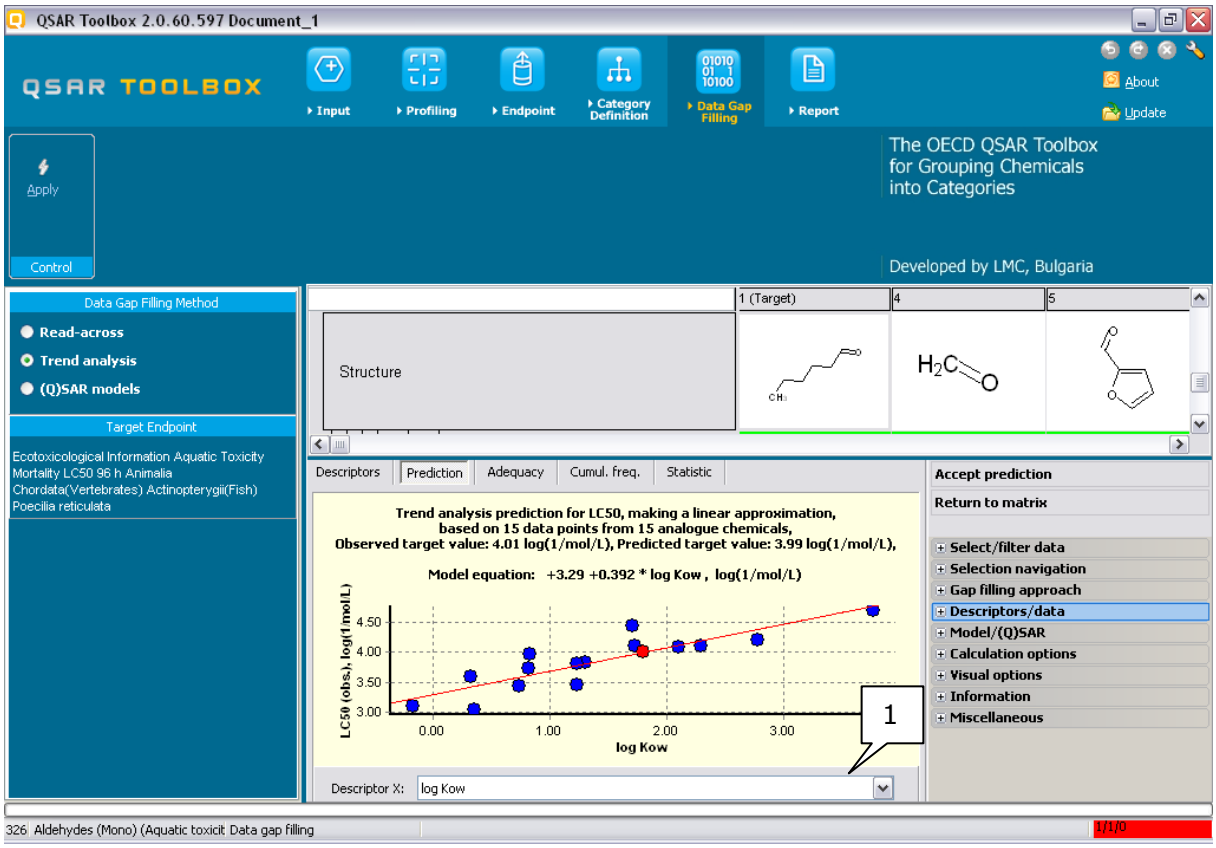
The current document is aimed at providing some guidance on some advanced features for using the Toolbox. It will be updated and expanded as further guidance becomes available.

Additional guidance and training material will be made available on the internet site for the OECD (Q)SAR Project [www.oecd.org/env/existingchemicals/qsar and http://www.qsartoolbox.org/] as well as the internet site of the Laboratory of Mathematical Chemistry [http://toolbox.oasis-lmc.org/]. There is also the OECD QSAR Toolbox Discussion Forum [https://community.oecd.org/community/toolbox_forum] where users can exchange their experiences with using the Toolbox. The user is invited to regularly consult these internet sites.

The OECD QSAR Toolbox has been developed with the financial assistance of the European Union.

1 Changing the X-axis descriptor when performing trend analysis with the Toolbox

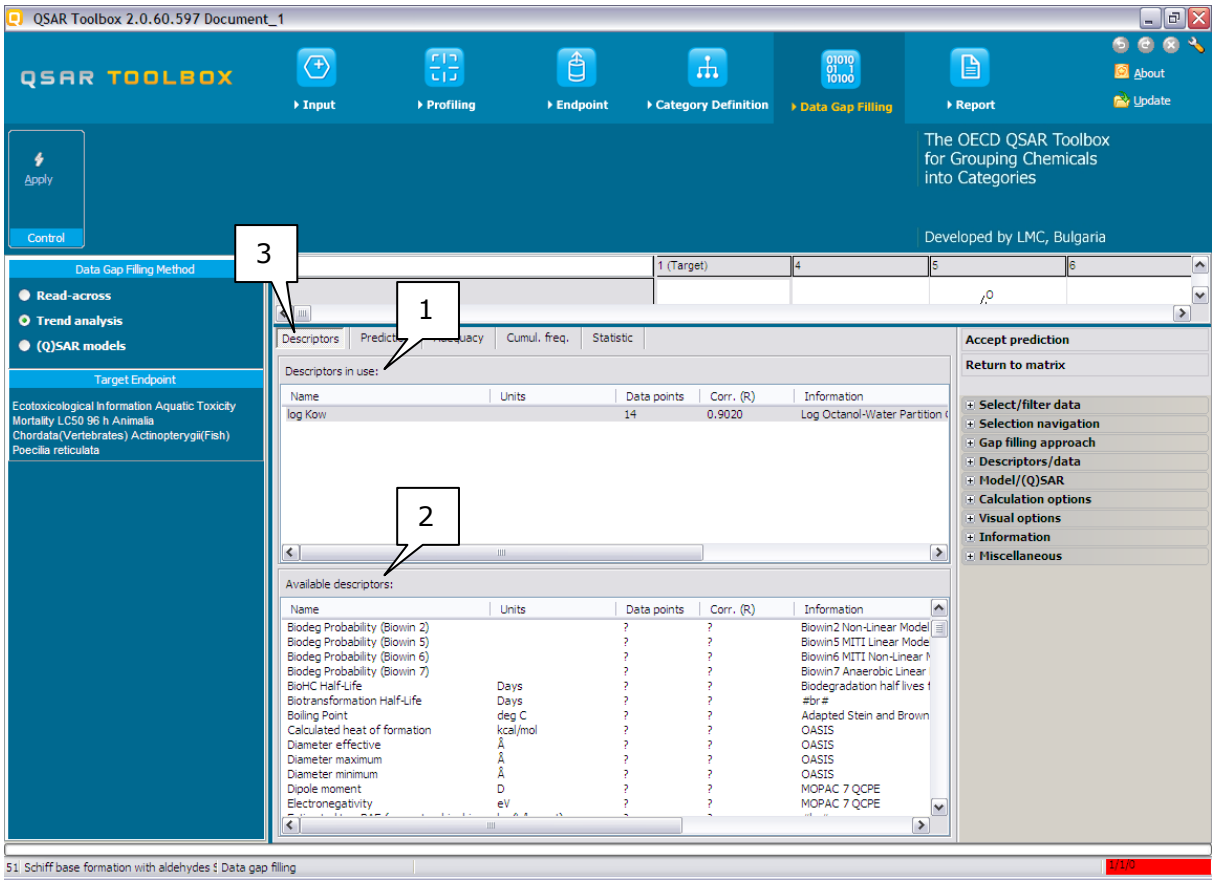
When performing a trend analysis within a category to estimate missing data with the Toolbox, the default X-axis descriptor is **log Kow**, which is estimated with the EPISUITE program (1) [see screenshot below]. While this descriptor may be pertinent for many endpoints, e.g. aquatic toxicity, alternative descriptors may be warranted for other endpoints (e.g. vapour pressure for acute inhalation toxicity).



There are two types of descriptors in the Toolbox – active descriptors (or **Descriptors in use** (1)) and inactive (or **Available descriptor** (2)). Both types of descriptors are listed separately in the in the tab **Descriptors** (3) [see screenshot below].

Active descriptors are the descriptors that were used in the calculations for predicting the endpoint.

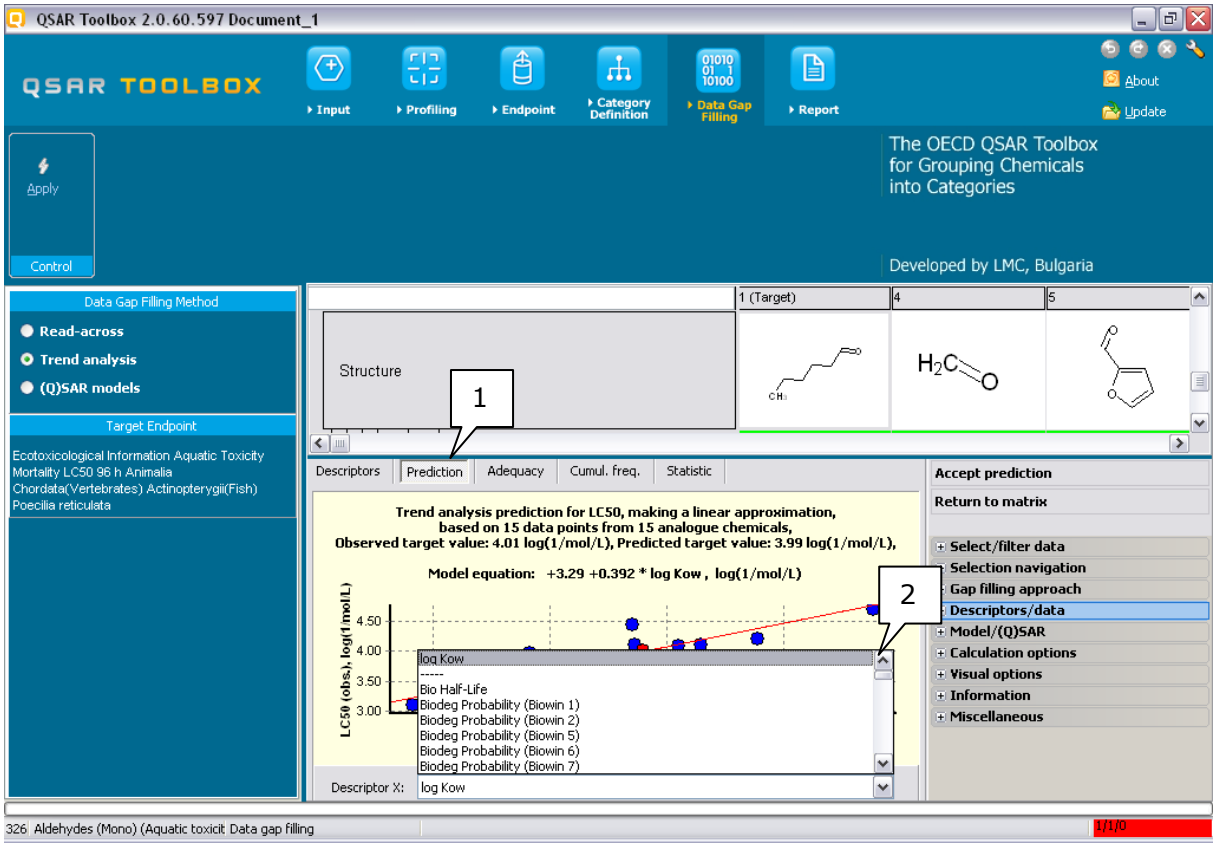
Inactive descriptors are all the other descriptors supported by the Toolbox but which were not used for predicting the endpoint.



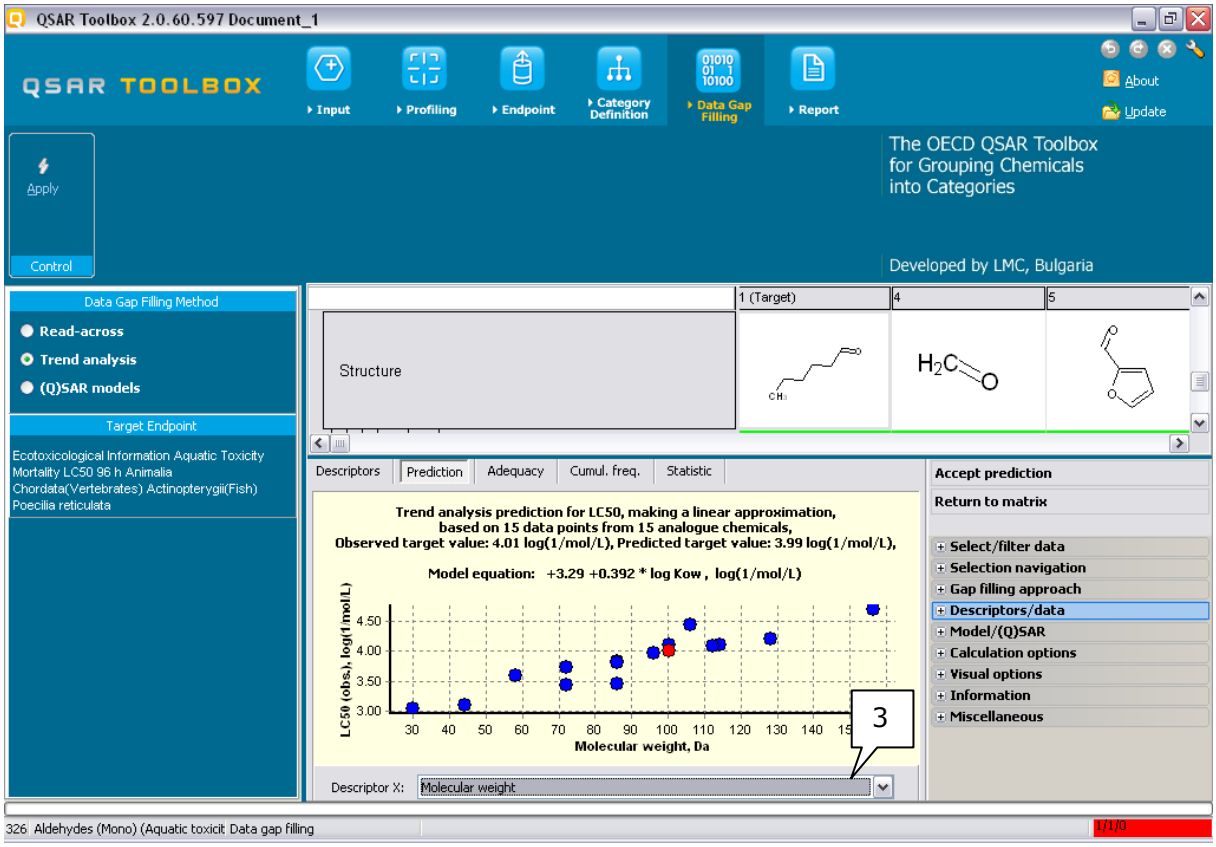
In the **Prediction** tab (1) the distribution of the observed data for the endpoint is plotted against any descriptor. By default, the Toolbox selects the first active descriptor to plot the available data. In case of trend analysis the equation line of the model is also visible.

If the user changes the descriptor, he/she will see how the observed data for the endpoint is distributed against the newly selected descriptor.

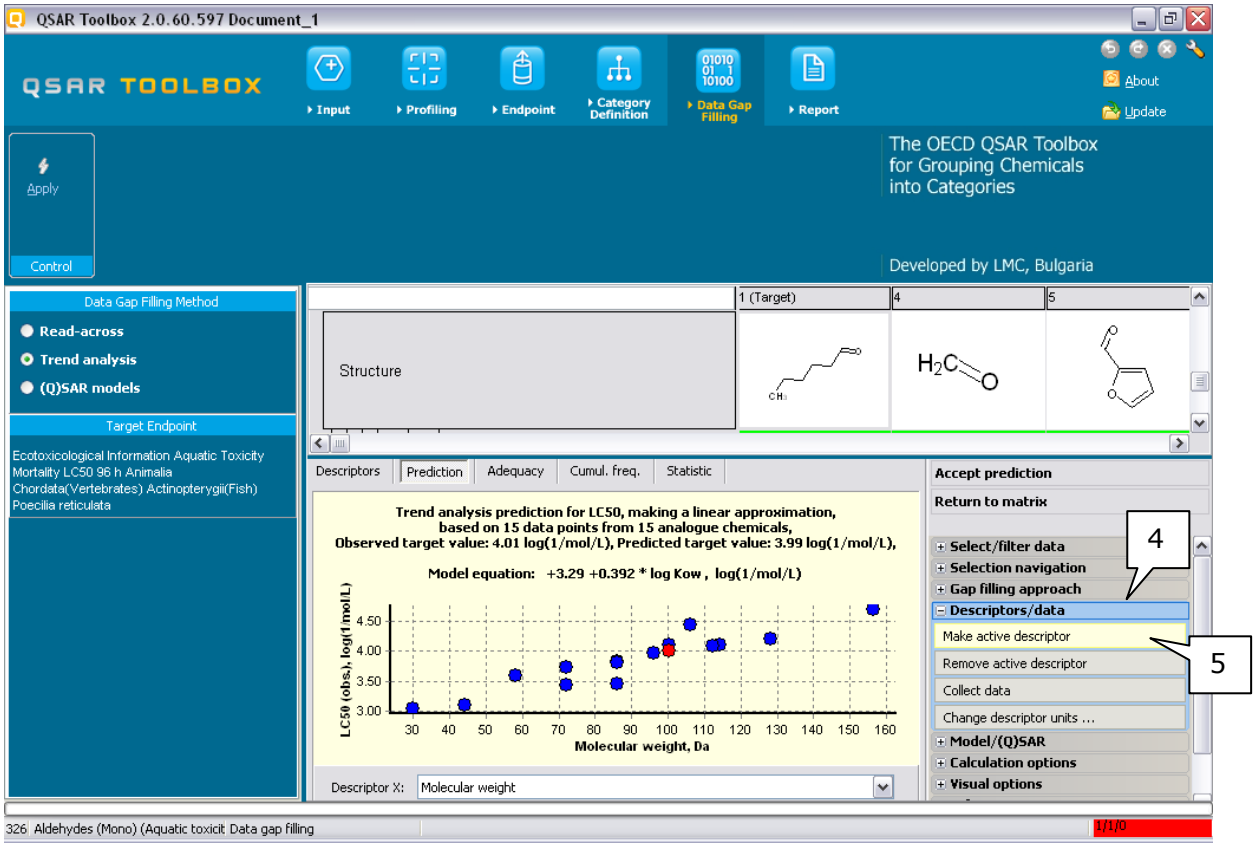
It is possible to change the descriptor and re-estimate the property. For that the user has to click the box of **Descriptor X** under the trend analysis graph (2). A list of available descriptors appears. The user can select a new descriptor from the list.



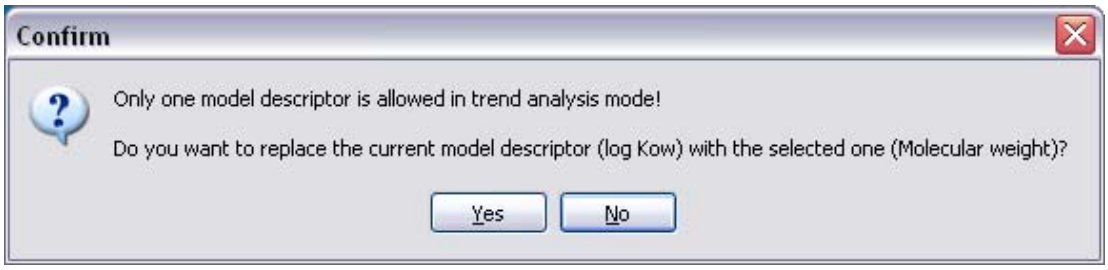
By clicking on the new descriptor (in this example "Molecular weight"), the selected descriptor is shown in the **Descriptor X** box (3) and the new descriptor appears also in the trend analysis graph.



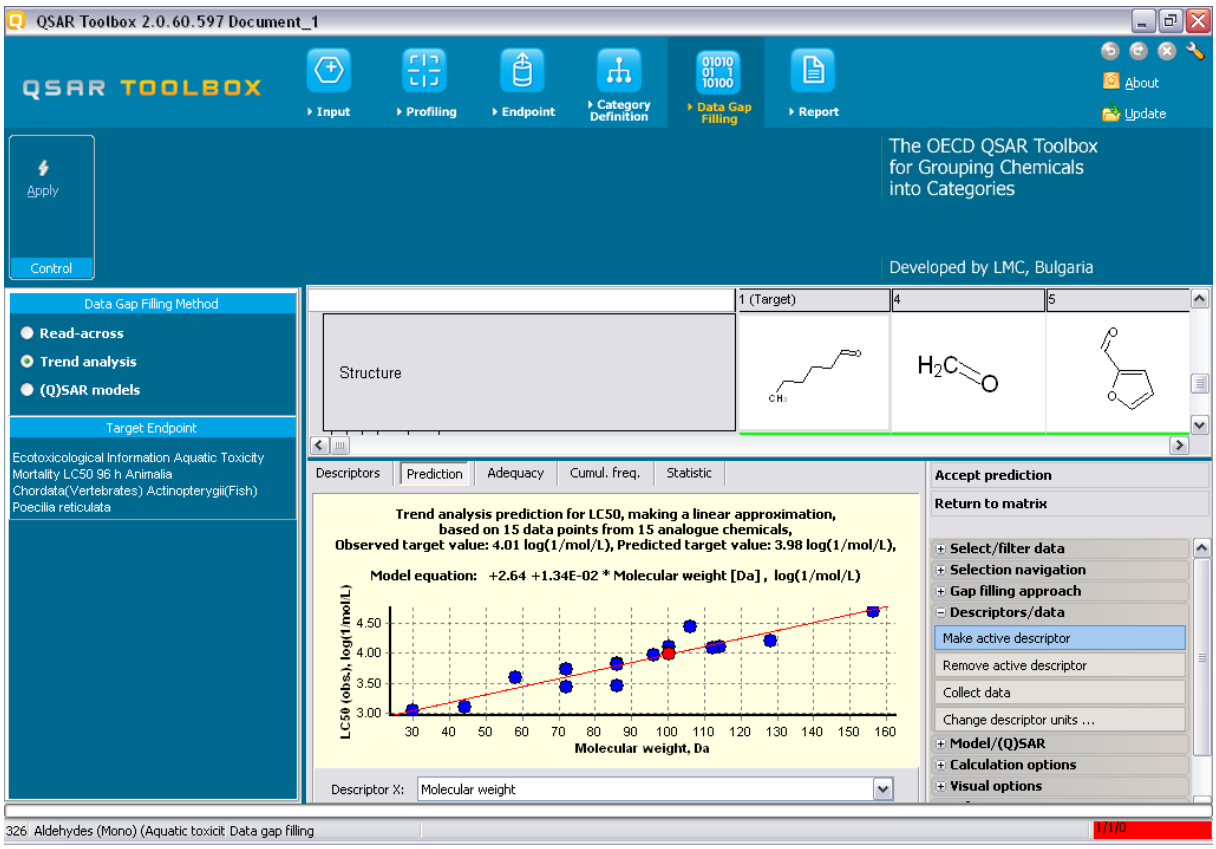
At this stage, however, the equation is **NOT** yet re-estimated. To make a prediction using the newly selected descriptor the user must activate it. This is done by clicking **Descriptors/data** (4) on the right-hand menu and then selecting **Make active descriptor** (5) [or by dragging the new descriptor from the **Available descriptors** list to the **Descriptor**s tab].



A dialogue box asking for confirmation of replacing the descriptor pops up.



By clicking the **Yes** button, the trend analysis for the new X-axis descriptor is completed.


QSAR Toolbox 2.0.60.597 Document_1
QSAR TOOLBOX
Input
Profiling
Endpoint
Category Definition
Data Gap Filling
Report
About
Update
Apply
Control
The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria
Data Gap Filling Method
Read-across
Trend analysis
(Q)SAR models
Target Endpoint
Ecotoxicological Information Aquatic Toxicity Mortality LC50 96 h Animalia Chordata(Vertebrates) Actinopterygii(Fish) Poecilia reticulata
Structure
1 (Target)
4
5
Descriptors
Prediction
Adequacy
Cumul. freq.
Statistic
Trend analysis prediction for LC50, making a linear approximation, based on 15 data points from 15 analogue chemicals,
Observed target value: 4.01 log(1/mol/L), Predicted target value: 3.98 log(1/mol/L),
Model equation: +2.64 +1.34E-02 * Molecular weight [Da], log(1/mol/L)
LC50 (obs.), log(1/mol/L)
Molecular weight, Da
Descriptor X: Molecular weight
Accept prediction
Return to matrix
Select/filter data
Selection navigation
Gap filling approach
Descriptors/data
Make active descriptor
Remove active descriptor
Collect data
Change descriptor units ...
Model/(Q)SAR
Calculation options
Visual options
326 Aldehydes (Mono) (Aquatic toxicit Data gap filling
1/1/0

2 Using experimental results for the X-axis descriptor

The user also has the possibility to load datasets with experimental data to be used instead of the estimated descriptors. This is illustrated with an example below of loading a database with experimental log Kow results.

1. Prepare the external file with parameter (calculator) values

The external file should be a text format with experimental data with the following structure:

SMILES<Tab>Value<TAB>Units

Please note that in case the descriptor is dimensionless, each line should nevertheless contain a <TAB> after the Value. Example of a text file is presented in Box 1.

Box 1. Format of a text file with experimental results

C(C)S	<TAB>	1.4	<TAB>
C(S)CCCCC	<TAB>	3.3	<TAB>
C(S)CCCCCCC	<TAB>	4.3	<TAB>

Notes:

- There should be no title line and no empty line at the bottom of the file (i.e. no “Carriage Return” at the end of the last line).

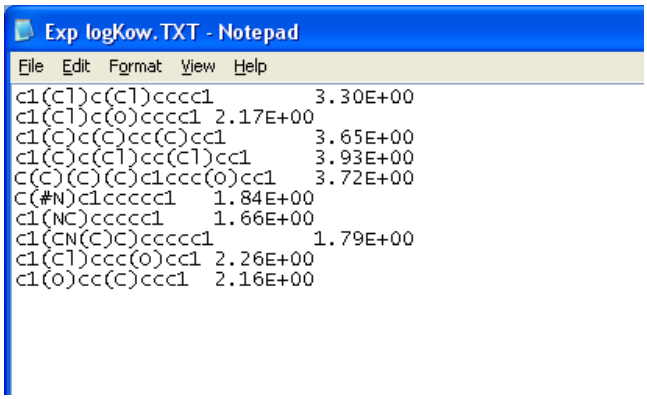
- The file has to be saved with a TXT extension (e.g. Log_KOW_internal.txt).

Example of external file with experimental parameter (calculator) values:

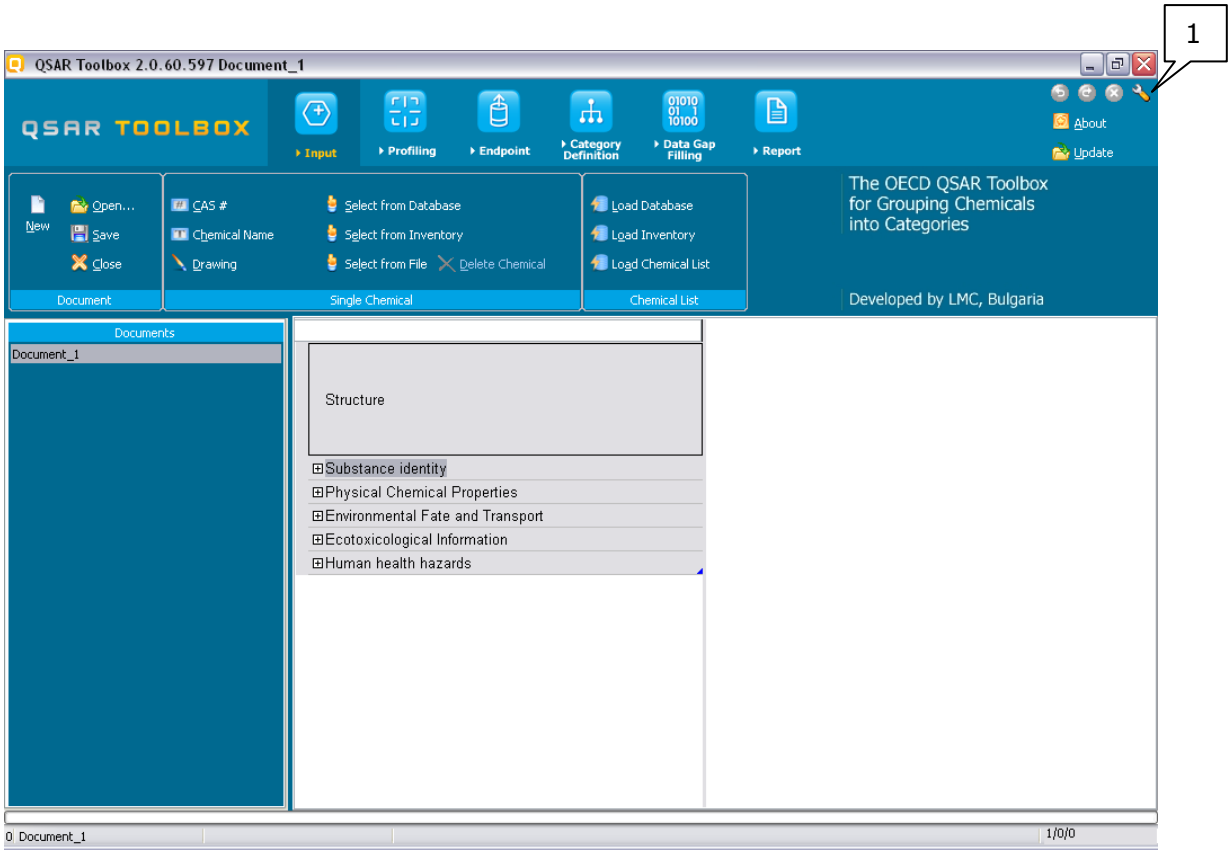
SMILES	Value	Units
c1(Cl)c(Cl)cccc1	3.30E+00	
c1(Cl)c(O)cccc1	2.17E+00	
c1(C)c(C)cc(C)cc1	3.65E+00	
c1(C)c(Cl)cc(Cl)cc1	3.93E+00	
C(C)(C)(C)c1ccc(O)cc1	3.72E+00	
C(#N)c1ccccc1	1.84E+00	
c1(NC)ccccc1	1.66E+00	
c1(CN(C)C)ccccc1	1.79E+00	
c1(Cl)ccc(O)cc1	2.26E+00	
c1(O)cc(C)ccc1	2.16E+00	

Copy these columns into the Notepad and save it as a file with ".txt" extension.

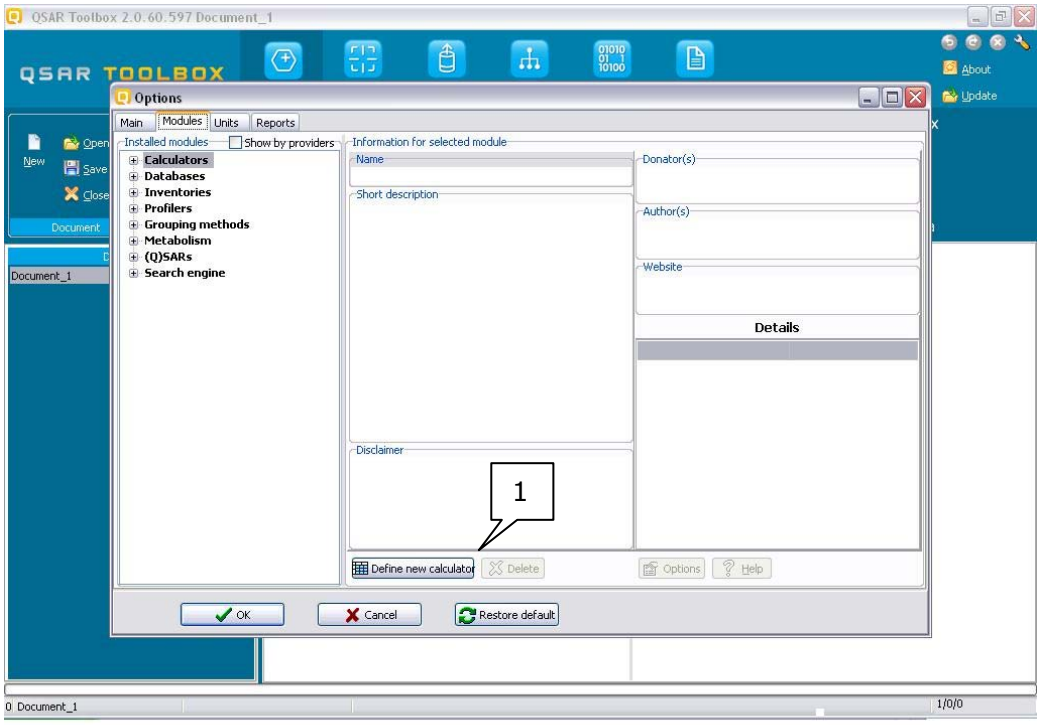
When copying columns above, please do not forget to include blanks in the last column so that there is a TAB between each Value and Unit.



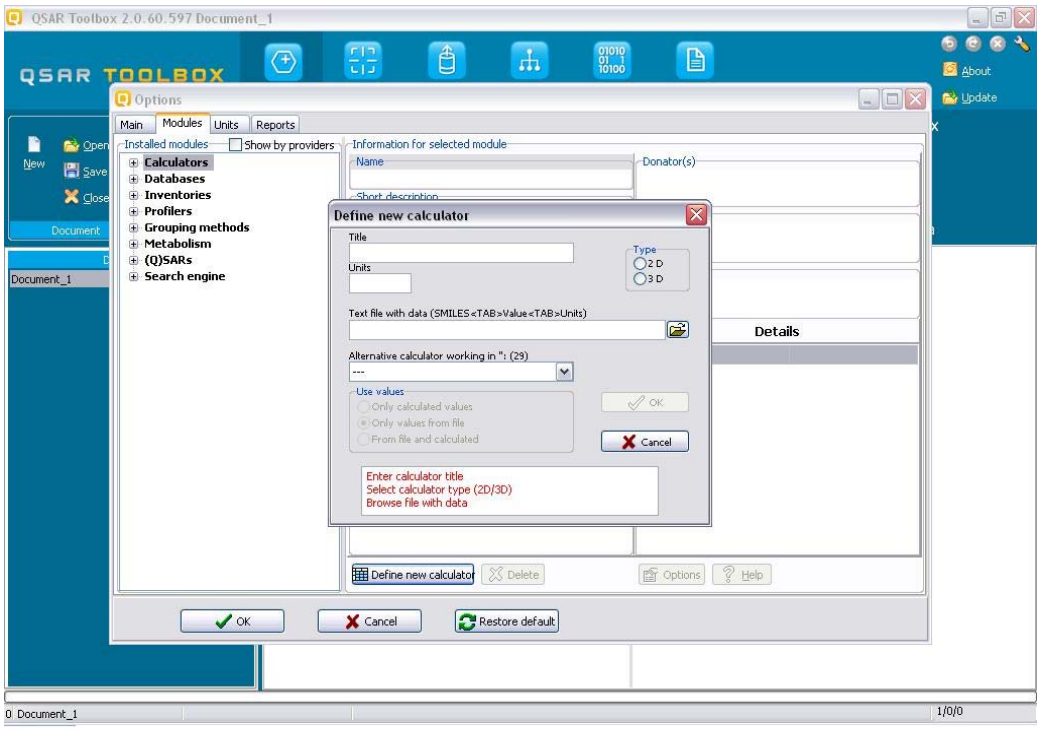
2. Go to the Options by clicking on the Options icon (in the shape of a wrench) in the upper right-hand corner (1).



Click on Define new calculator (1).



Then a dialogue box **Define new calculator** pops up.



3. The steps of introducing the external file with parameter (calculator) values in the Toolbox are (see also screenshot below):

(1) Add a title. In this example "Experimental log Kow"

(2) Add a dimension or unit. In this example the descriptor is dimensionless and the field is left blank.

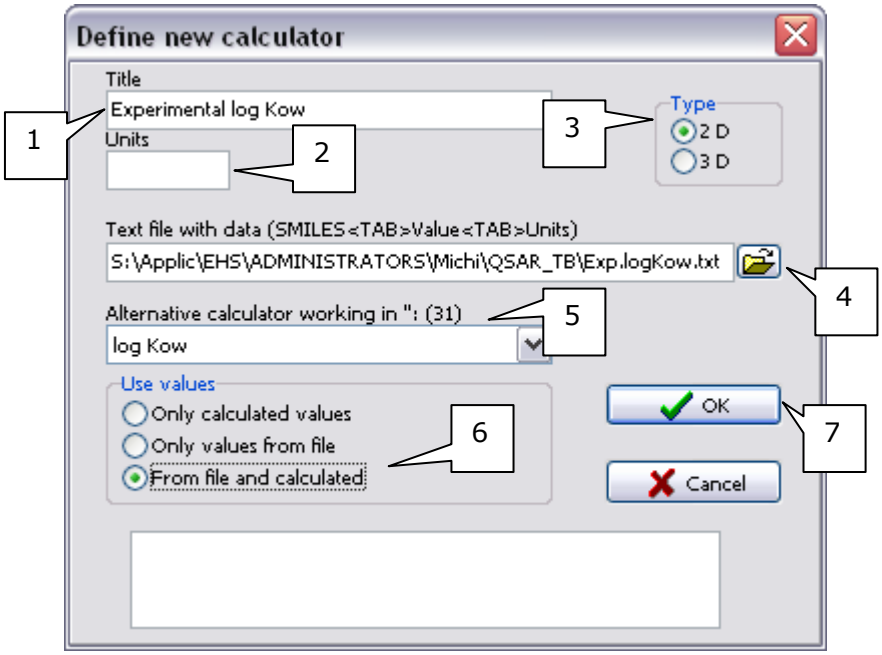
(3) Select a type: 2D or 3D. The distinction is made depending on whether the descriptor could be calculated based on a 2D or 3D configuration of the structure. Kow being usually derived from a 2D structure, the 2D button can be selected.

(4) Indicate the path of the file containing external experimental data.

(5) The user can choose an existing calculator to estimate the descriptor for chemicals not listed in the file with experimental data. In this example, "log Kow" (calculated by the EPISUITE model) could be chosen to fill data gaps in the experimental file.

(6) Please note that if an alternative calculator is chosen, the radio button **From file and calculated** has to be checked.

(7) By clicking the **OK** button, the file is imported as a new descriptor. And the user can return to a normal work session.



Please Note: For the newly defined parameter (calculator) to be active, the system must be restarted.

4. How to use the new parameter (calculator):

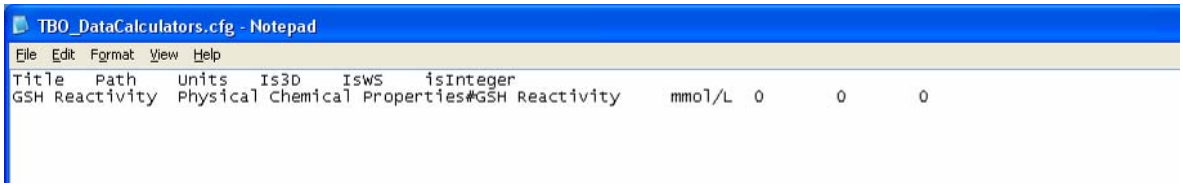
Under Data gap filling the user can change the X-axis descriptor with a parameter using external experimental data by following the steps explained in Chapter 1.

3 Using experimental results from the Toolbox databases for the X-axis descriptor

No interface is provided for this approach. For this purpose one needs to:

Edit the TBO_DataCalculators.cfg file
The default location of this file is C:\Program Files\Common Files\QSAR Toolbox\Ver 2.0\config

It can be edited with a simple text editor.



The file contains a tab delimited table in the following format:

Title<TAB>Path<TAB>Units<TAB>Is3D<TAB>IsWS<TAB>IsInteger

Each row from this file defines a new calculator with:
Title – this is the name of the calculator

Path – Position of the parameter in the endpoint tree – where the data is associated with when read from databases

Units – defined units of the parameter

Is3D- is the type specification of parameter. It is zero (0) for 2D; OR one (1) if the parameter is 3D

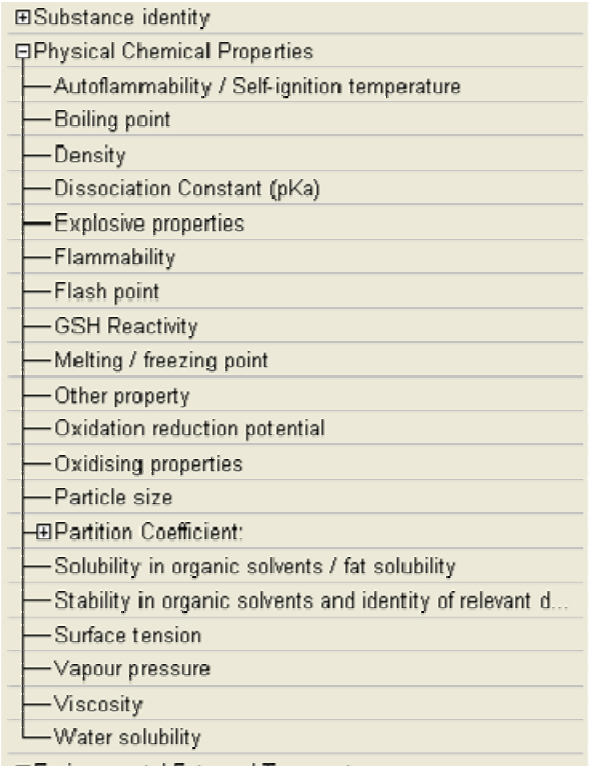
IsWS- It is zero (0) if the parameter is not water solubility or one (1) if the parameter is WS

IsInteger – Parameter get integer values only (for example "Number of aromatic bonds") if it is one (1), otherwise it is zero (0).

For instance, in order to add the experimental data in the QSAR Toolbox for boiling point as an X-axis descriptor, the user should follow these steps:

1. Open the "TBO_DataCalculators.cfg" file

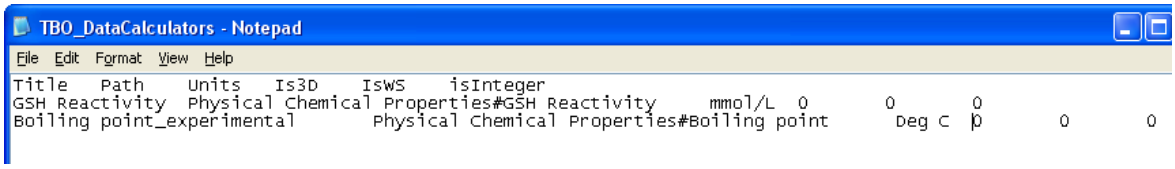
2. Find where the endpoint you wish to add is located in the endpoint tree of the Toolbox in order to define the path:



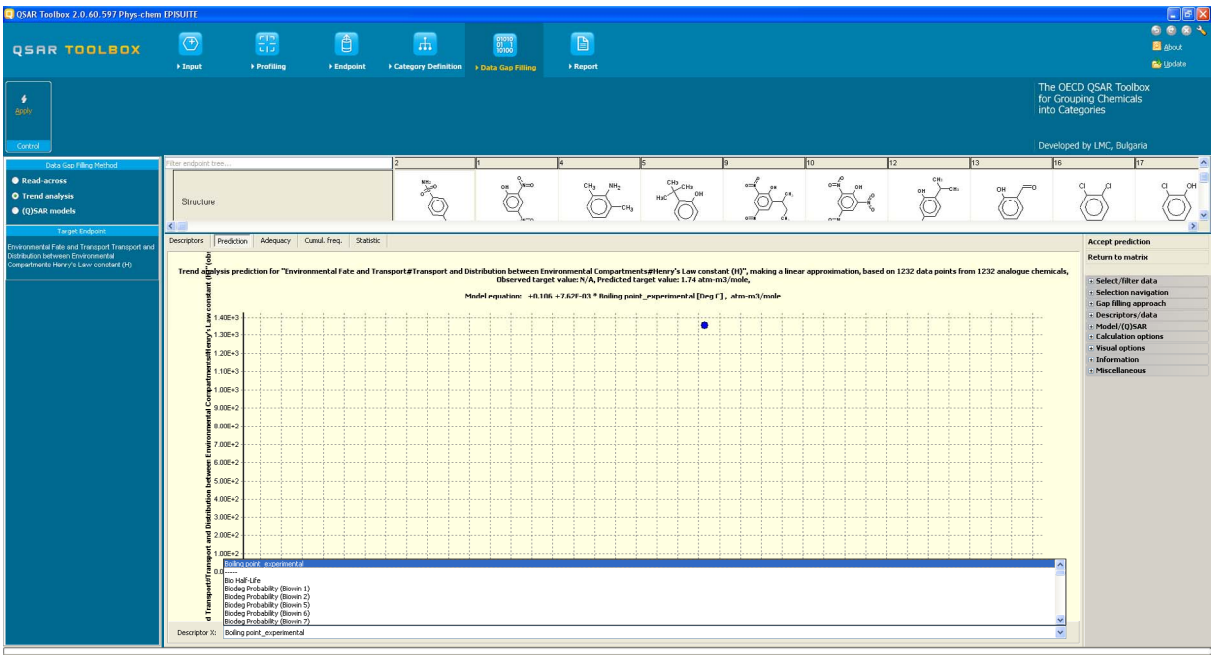
3. The path is indicated by separating each branching of the tree with "#". In this case, that means that the path for boiling point is "Physical Chemical Properties#Boiling point"

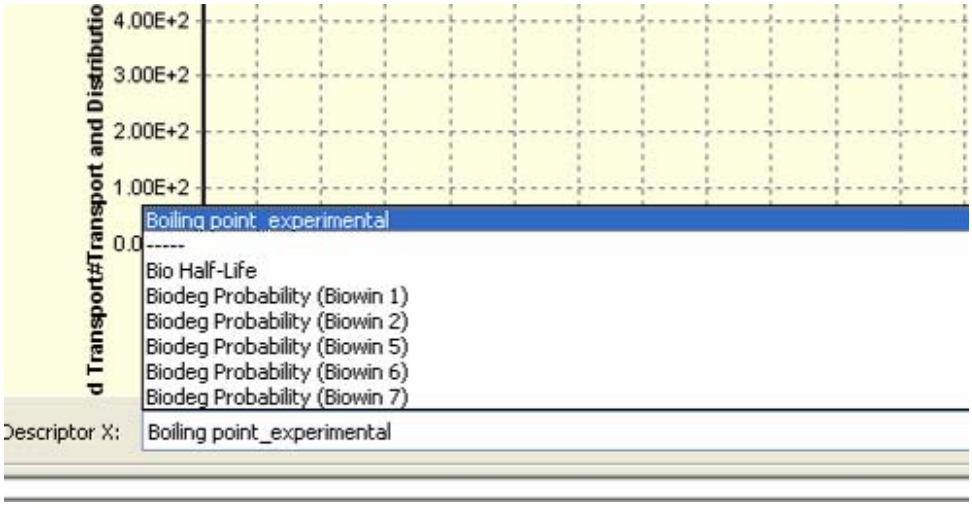
4. This parameter is expressed in Deg C, it is 2D, it is not water solubility and it is not limited to integer values. You can give any name you want to the descriptor (e.g.:Boiling point_experimental) Hence, you would need to add to the file "TBO_DataCalculators.cfg" the following line:

Boiling point_experimental <TAB> Physical Chemical Properties#Boiling point<TAB> Deg C<TAB>0<TAB>0<TAB>0.



5. Save and close the "TBO_DataCalculators.cfg" file. The next time the QSAR Toolbox is opened this new descriptor will be availabel in the drop down menu of the Data Gap filling graph.


QSAR Toolbox 2.0.60.597 Phys-chem EPISUITE
QSAR TOOLBOX
Input
Profiling
Endpoint
Category Definition
Data Gap Filling
Report
The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria
Data Gap Filling Method
Read-across
Trend analysis
(Q)SAR models
Target Endpoint
Environmental Fate and Transport Transport and Distribution between Environmental Compartments Henry's Law constant (H)
Structure
Descriptors
Prediction
Adequacy
Cumul. freq.
Statistic
Trend analysis prediction for "Environmental Fate and Transport#Transport and Distribution between Environmental Compartments#Henry's Law constant (H)", making a linear approximation, based on 1232 data points from 1232 analogue chemicals,
Observed target value: N/A, Predicted target value: 1.74 atm-m3/mole,
Accept prediction
Return to matrix
Select/filter data
Selection navigation
Gap filling approach
Descriptors/data
Model/(Q)SAR
Calculation options
Visual options
Information
Miscellaneous
Boiling point_experimental
Bio Half-Life
Biodeg Probability (Biowin 1)
Biodeg Probability (Biowin 2)
Biodeg Probability (Biowin 5)
Biodeg Probability (Biowin 6)
Biodeg Probability (Biowin 7)
Descriptor X:
Boiling point_experimental


4.00E+2
3.00E+2
2.00E+2
1.00E+2
Boiling point_experimental
Bio Half-Life
Biodeg Probability (Biowin 1)
Biodeg Probability (Biowin 2)
Biodeg Probability (Biowin 5)
Biodeg Probability (Biowin 6)
Biodeg Probability (Biowin 7)
Descriptor X:
Boiling point_experimental

OECD
2, rue André Pascal
75775 Paris Cedex 16
France

Tel.: +33 1 45 24 82 00
Fax: +33 1 45 24 85 00

ehscont@oecd.org