

OECD QSAR Toolbox v.4.1

Example illustrating endpoint vs. endpoint correlation using ToxCast data

Outlook

- **Background**
- Objectives
- The exercise
- Workflow

Background

This presentation is designed to introduce the user with:

- ToxCast database as part of the Toolbox database
- Illustration of endpoint vs. endpoint correlations using:
 - ToxCast data
 - ToxCast and Estrogen receptor data

Outlook

- Background
- **Objectives**
- The exercise
- Workflow

Objectives

- This presentation demonstrates endpoint vs. endpoint correlations using ToxCast and Estrogen receptor data

Outlook

- Background
- Objectives
- **The exercise**
- Workflow

The exercise

- Illustration of endpoint data correlations using the ToxCast and estrogen binding data between two type data:
 - AC50 vs. AC50 endpoints associated with different test type
 - AC50 vs. Estrogen receptor binding data

Outlook

- Background
- Objectives
- The exercise
- **Workflow**

Workflow

- **The Toolbox has six modules which are typically used in a workflow:**
 - Chemical Input
 - Profiling
 - Endpoints
 - Category Definition
 - Filling Data Gaps
 - Report
- **In this example we will use the modules in a different order, tailored to the aims of the example.**

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - **Load ToxCast database**

ToxCast database Loading database

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Document', 'Single Chemical', 'Search', and 'Target Endpoint'. The toolbar contains icons for 'New', 'Open', 'Close', 'Save', 'CAS#', 'Name', 'Structure', 'Composition', 'Select', 'Delete', 'ChemIDs', 'Database Inventory', 'List', 'Substructure (SMARTS)', 'Query', and 'Define'. The 'Database Inventory' dialog box is open, showing a list of databases. The 'ToxCastDB' entry is highlighted. The 'OK' button is also visible. Red callout boxes with numbers 1, 2, 3, and 4 indicate the steps: 1. Click on 'Database' button; 2. Select 'ToxCast DB'; 3. Click 'OK'; 4. Chemicals are loaded on data matrix.

1. **Click** on "Database" button;
2. **Select** "ToxCast DB";
3. **Click** "OK";
4. Chemicals are loaded on data matrix

ToxCast database

Sidebar of database relevancy

Once the endpoint is selected, the relevant databases are getting green highlighted.

The screenshot shows the QSAR Toolbox interface. At the top, there is a navigation bar with icons for Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this is a menu bar with 'Data', 'Import', and 'Export' options. The main workspace is divided into several panels:

- Left Panel (Inventories):** A list of databases. 'ToxCastDB' is highlighted in green. A callout box labeled '2' points to this entry. Above this list is an 'Options' menu with 'Group by: Category', 'Sort by: Name', and 'Color by: Endpoint selection'. A callout box labeled '3' points to the 'Options' menu.
- Middle Panel (Filter endpoint tree...):** A tree view of endpoints. 'ToxCast' is selected and highlighted in green. A callout box labeled '1' points to this entry. A 'Legend' dialog box is open, showing 'Target endpoint' with a green square for 'Have data for target endpoint' and a white square for 'Have no data for target endpoint'.
- Right Panel (Table):** A table with 5 columns and multiple rows. The first row contains chemical structures. The 'ToxCast' row is highlighted in green.

1. **Click** on the level ToxCast endpoint tree;
2. The database is getting green highlighted;
3. **Click** "Options" and ask for Legend;

ToxCast database Data gathering

1 Go to **Data**; **2** Check **ToxCast database**; **3** Click "Gather"; **4** The data appears on datamatrix on the level "ToxCast"

Structure	1	2	3	4	5	6	7	8	9
<chem>O=Cc1ccccc1</chem>									
<chem>Oc1ccc(O)cc1</chem>									
<chem>CCCCCCCCCO</chem>									
<chem>CC1(C)C(C)C(C)C(C)C1</chem>									
<chem>CC1(C)C(C)C(C)C(C)C1</chem>									
<chem>CCCCCCCCCO</chem>									
<chem>CC1(C)C(C)C(C)C(C)C1</chem>									
<chem>CC1(C)C(C)C(C)C(C)C1</chem>									
ToxCast			M: 0.0601 mg/L			M: 7.06 mg/L		M: 6.8 mg/L	M: 2.84 mg/L
ACEA (600/660)						M: 5.84 mg/L			
Apredica (425/2653)									
Attagene (1374/11710)		M: 4.33 mg/L	M: 0.756 mg/L	M: 0.627 mg/L	M: 17.9 mg/L	M: 16.2 mg/L	M: 0.67 mg/L	M: 9.86 mg/L	M: 1.14 mg/L
BioSeek (971/21906)	M: 0.127 mg/L	M: 3.41 mg/L		M: 4.17 mg/L	M: 0.674 mg/L	M: 2.69 mg/L			
NCGG (1475/6890)	M: 2.72 mg/L		M: 6.32 mg/L	M: 1.61 mg/L	M: 1.53 mg/L		M: 12.4 mg/L		
Novoscreen (975/8054)		M: 2.43 mg/L		M: 0.295 mg/L	M: 0.209 mg/L	M: 0.0122 mg/L	M: 8.61 mg/L	M: 0.415 mg/L	
Odyssey Thera (969/2794)		M: 6.89 mg/L	M: 0.299 mg/L	M: 9.54 mg/L	M: 2.57 mg/L		M: 17.1 mg/L	M: 14.1 mg/L	
Undefined Assay provider (2/2)									
Toxicity to Reproduction									
Toxicokinetics, Metabolism and Distribution									

1. Go to **Data**; 2. Check **ToxCast database**; 3. Click "Gather"; 4. The data appears on datamatrix on the level "ToxCast"

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Load ToxCast database
 - **ToxCast database - overview**

ToxCast database

Background

- A major part of EPA's CompTox research is the ToxCast™ project. ToxCast is a multi-year project launched in 2007 that uses automated chemical screening technologies (called "high-throughput screening assays") to expose living cells or isolated proteins to chemicals. The cells or proteins are then screened for changes in biological activity that may suggest potential toxic effects. These innovative methods have the potential to limit the number of required laboratory animal-based toxicity tests while quickly and efficiently screening large numbers of chemicals.
- ToxCast has evaluated over 2,000 chemicals from a broad range of sources including: industrial and consumer products, food additives, and potentially "green" chemicals that could be safer alternatives to existing chemicals. Chemicals were evaluated in over 700 high-throughput assays that cover a range of high-level cell responses and approximately 300 signaling pathways.
- ToxCast results are contributed to the federal agency collaboration called Toxicity Testing in the 21st Century (Tox21). Tox21 pools chemical research, data and screening tools from multiple federal agencies including the National Toxicology Program. So far, Tox21 has compiled high-throughput screening data on nearly ten thousand chemicals.

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Load ToxCast database
 - ToxCast database – overview
 - **Correlation of data - background**

Correlation of endpoint data

Background

- This functionality introduces the user with opportunity to analyze correlations between selected gap filling endpoint (endpoint used for prediction) and other endpoint data.
- It is applicable for correlation analysis of data presented in ordinary, interval or ratio scale.
- If correlated data are measured in interval or ratio scale they are transformed in ordinary scale and the strength of the correlation is estimated by Spearman correlation coefficient.
- Basically, this functionality provides a correlation between target endpoint (this is the initial endpoint selected by the user) displayed on ordinate axis (Y-axis) and other endpoint data displayed on abscissa (X-axis).

Correlation of endpoint data

Spearman coefficient factor

- Spearman's rank correlation coefficient is a nonparametric rank statistic proposed by Charles Spearman as a measure of the strength of an association between two variables. It assesses how well the relationship between two variables can be described using a monotonic function.
- Spearman correlation coefficient could be used for exploring the covary between:
 - two ranked variables
 - one measurement variable and one ranked variable (in this case, the measurement variable need to be to converted to ranks)
- Spearman correlation varies from -1 to +1 and the interpretation of the coefficient factor is provided below:
 - 0.00 – 0.19 – very weak correlation
 - 0.20 – 0.39 – weak correlation
 - 0.40 – 0.59 – moderate correlation
 - 0.60 – 0.79 – strong correlation
 - 0.80 – 1.0 – very strong

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Load ToxCast database
 - ToxCast database – overview
 - Correlation of data – background
 - **Types endpoint correlations**

Types endpoint correlations

Types endpoint correlations are as follows:

- Continuous vs. continuous
- Categorical vs. categorical*:
 - ✓ Categorical vs. categorical
 - ✓ Categorized continuous vs. categorical
 - ✓ Categorized continuous vs. categorized continuous

*All type categorical vs. categorical correlations are not illustrated in this presentations. These type correlations are shown in presentation "Tutorial 13 TB4.1. Example illustrating endpoint vs. endpoint correlation for apical endpoints"

Outlook

- Background
- Objectives
- The exercise
- **Workflow**
 - Load ToxCast database
 - ToxCast database – overview
 - Correlation of data – background
 - **Types endpoint correlations**
 - Continuous vs. continuous

Types endpoint correlations

Continuous vs. continuous

- The aim of this type correlation is to illustrate how continuous type endpoint data or so called ratio data correlates each other (e.g.LC50 vs. EC50 data)
- In this example we will illustrate how AC50 data associated with two different test assays extracted from ToxCast DB correlates each other:
 - NCGC Reporter Gene Assay ERa Agonist, Estrogen receptor 1 (assay 1)
 - Tox21_Era_BLA_Agonist_ch2 (assay 2)
- Step by step workflow is presented on next few slides. Summary of the workflow steps are provided below:
 - *Gather experimental data (step 1)*
 - *Define target endpoint (step 2)*
 - *Enter Gap filling (step 3)*
 - *Change default X-descriptor (logKow) with AC50 data (step 5)*

Types endpoint correlations

Continuous vs. continuous

Gather experimental data – step 1

The screenshot shows the QSAR Toolbox software interface. The top menu bar includes 'Data', 'Import', and 'Export'. The 'Data' menu is open, showing options like 'Gather', 'IUCLID6', and 'Data Gap Filling'. A callout '1' points to the 'Data' menu. A callout '3' points to the 'Gather' button. On the left, the 'Databases' panel is open, showing a list of databases with 'ToxCastDB' selected. A callout '2' points to 'ToxCastDB'. The main window displays a 'Filter endpoint tree...' on the left and a data matrix on the right. The data matrix has 8 columns and 15 rows. The first row shows chemical structures for columns 1 through 8. The 'ToxCast' endpoint is highlighted in blue in the matrix.

Follow the steps if you already load Toxcast data on data matrix. 1. **Go** to "Data" 2. **Select** "ToxCast" DB 3. **Click** "Gather"

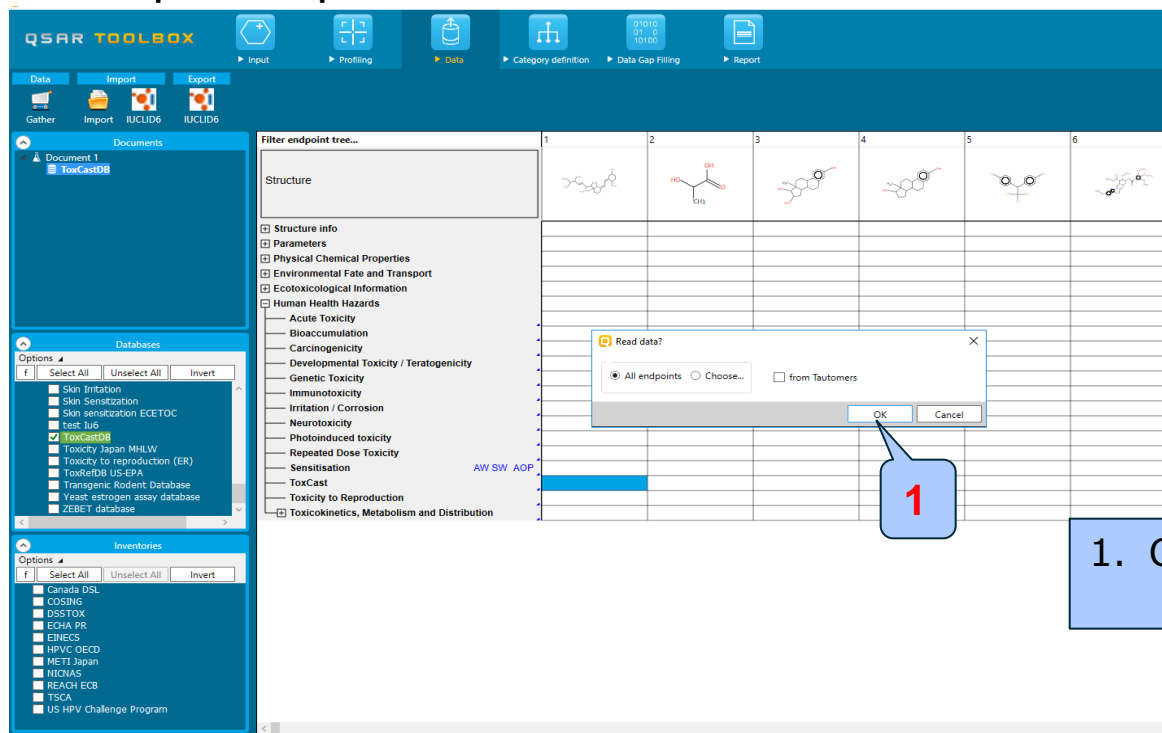
Types endpoint correlations

Continuous vs. continuous

Gather experimental data – step 1

Toxicity information on the target chemical is electronically collected from the selected datasets.

A window with “Read data?” appears. Now the user could choose to collect “all” or “endpoint specific” data.



1. Click **OK** to read all available data

Types endpoint correlations

Continuous vs. continuous
Gather experimental data – step 1

The screenshot shows the QSAR Toolbox interface with the 'Filter endpoint tree...' window open. The 'Databases' panel on the left has 'ToxCastDB' selected. The 'Inventories' panel also has several options checked. The main window displays a table with chemical structures and their corresponding data points. A dialog box is overlaid on the table, displaying the message '54669 points added across 1813 chemicals.' and an 'OK' button. A red '1' in a blue box points to the 'OK' button.

Structure	1	2	3	4
Structure info				
Parameters				
Physical Chemical Properties				
Environmental Fate and Transport				
Ecotoxicological Information				
Human Health Hazards				
Acute Toxicity				
Bioaccumulation				
Carcinogenicity				
Developmental Toxicity / Teratogenicity				
Genetic Toxicity				
Immunotoxicity				
Irritation / Corrosion				
Neurotoxicity				
Photoinduced toxicity				
Repeated Dose Toxicity				
Sensitisation				
ToxCast				
Toxicity to Reproduction				
Toxicokinetics, Metabolism and Distribution				
AW SW AOP (1813/54669)	M: 2.06 mg/L	M: 0.0039 mg/L	M: 0.0545 mg/L	M: 0.0545 mg/L

1. Click "OK" to close the window

Types endpoint correlations

Continuous vs. continuous

Gather experimental data – step 1

The screenshot displays the QSAR TOOLBOX interface. At the top, a workflow menu includes: Input, Profiling, Data, Category definition, Data Gap Filling, and Report. Below this, there are buttons for Data (Gather), Import (Import, IUCLID6), and Export (IUCLID6). The main workspace is divided into several panels:

- Documents:** Shows 'Document 1' containing 'ToxCastDB'.
- Databases:** A list of databases with checkboxes. 'ToxCastDB' is selected and highlighted in green. Other databases include Skin Irritation, Skin Sensitization, ECETOC, test Iu6, Toxicity Japan MHLW, etc.
- Inventories:** A list of assay inventories with checkboxes. 'Canada DSL', 'COSING', 'DSSTOX', etc., are listed.
- Filter endpoint tree...:** A tree view of toxicity endpoints. A red callout box with the number '1' points to the 'ToxCast' node under 'Human Health Hazards'.
- Datamatrix:** A table with 5 columns (labeled 1-5) and multiple rows. The first row shows chemical structures for five different compounds. Subsequent rows show numerical data for various endpoints, with the 'ToxCast' row highlighted in blue. The data includes values like 'M: 21.2 mg/L', 'M: 0.0039 mg/L', etc.

1. **ToxCast data** has been loaded on datamatrix in a separate node of "Endpoint tree" called "ToxCast"

Types endpoint correlations

Continuous vs. continuous

Define target endpoint – step 2

The screenshot shows the QSAR Toolbox interface with a datamatrix table. The table has 5 columns representing different chemical structures. The first column is labeled 'Structure' and contains chemical structures. The subsequent columns contain numerical values representing endpoint correlations. A red box highlights a cell in the table, and a callout bubble with the number '1' points to it.

Structure	1	2	3	4	5
Chemical Structure 1					
Chemical Structure 2					
Chemical Structure 3					
Chemical Structure 4					
Chemical Structure 5					
ACEA (600/660)	M: 21.2 mg/L	M: 0.0039 mg/L	M: 0.000585 mg/L	M: 8.01 mg/L	M: 9.32 mg/L
Apradina (425/2653)				M: 10.9 mg/L	M: 27.5 mg/L
Attagene (1374/11710)	M: 2.06 mg/L		M: 0.0141 mg/L	M: 2.44 mg/L	M: 2.06 mg/L
BioSeek (971/21906)				M: 4.71 mg/L	M: 9.85 mg/L
NCGC Reporter Gene Assay ERa Agonist (374/505)					
Homo sapiens					
estrogen receptor 1					
AC50	M: 19.1 mg/L	M: 0.000531 mg/L	M: 0.00017 mg/L	M: 9.12E-05 mg/L	M: 5.42 mg/L
NCGC Reporter Gene Assay ERa Antagonist (487/559)					M: 7.62 mg/L
Tox21_AhR (319/319)		M: 0.000106 mg/L			
Tox21_AR_BLA_Agonist_ch1 (439/439)	M: 0.00436 mg/L	M: 0.000991 mg/L		M: 0.242 mg/L	M: 16.5 mg/L
Tox21_AR_BLA_Agonist_ch2 (67/67)				M: 0.183 mg/L	
Tox21_AR_BLA_Agonist_ratio (89/89)				M: 0.311 mg/L	
Tox21_AR_BLA_Antagonist_ratio (150/150)			M: 3 mg/L		
Tox21_AR_BLA_Antagonist_viability (207/207)	M: 14.4 mg/L				
Tox21_AR_LUC_MDAKB2_Agonist (90/90)		M: 2.86 mg/L		M: 0.152 mg/L	
Tox21_AR_LUC_MDAKB2_Antagonist (56/56)					
Tox21_AR_LUC_MDAKB2_Antagonist_viability (291/291)		M: 0.000106 mg/L			

The target endpoint is AC50 associated with assay "NCGC Reporter Gene Assay ERa Agonist"

1. **Click** on the cell related to the investigated endpoint, below the first chemical of datamatrix

Types endpoint correlations

Continuous vs. continuous

Define target endpoint – step 2

The screenshot displays the QSAR Toolbox software interface during a 'Data Gap Filling' workflow. The top toolbar includes icons for 'Input', 'Profiling', 'Data', 'Category definition', and 'Data Gap Filling'. A 'Documents' panel on the left shows 'ToxCastDB'. A 'Filter endpoint tree...' panel is open, displaying a hierarchical list of endpoints. A table of chemical structures and their associated data values is shown. A 'Possible data inconsistency' dialog box is open, with 'OK' and 'Cancel' buttons. Red boxes and numbers 1, 2, 3, and 4 highlight specific actions: 1. Clicking 'Data Gap Filling' in the toolbar; 2. Highlighting an empty cell in the table; 3. Selecting 'Trend analysis' in the 'Data Gap Filling' dropdown; 4. Clicking 'OK' in the dialog box.

Structure	1	2	3	4	5	6	7	8
Apredica (425/2653)			M: 0.0001 mg/L		M: 1.69 mg/L			
Attagene (1374/11710)		M: 0.88 mg/L	M: 0.113 mg/L	M: 0.627 mg/L	M: 11.8 mg/L	M: 16.2 mg/L	M: 0.033 mg/L	M: 10.3 mg/L
BioSeek (971/21906)	M: 0.127 mg/L	M: 0.16 mg/L		M: 0.464 mg/L	M: 0.539 mg/L	M: 0.243 mg/L		M: 0.663 mg/L
NCGC								
NCGC Reporter Gene Assay ERA Agonist								
Homo sapiens								
estrogen receptor 1								
AC50 (374/505)			M: 0.156 mg/L		M: 0.478 mg/L			
NCGC Reporter Gene Assay ERA Antag (487/555)			M: 10.6 mg/L		M: 0.727 mg/L		M: 5.52 mg/L	
Tox21_AhR (237/237)								
Tox21_AhR_viability (319/319)								
Tox21_AR_BLA_Agonist_ch1 (439/439)								
Tox21_AR_BLA_Agonist_ch2 (67/67)								
Tox21_AR_BLA_Agonist_ratio (89/89)								
Tox21_AR_BLA_Antagonist_ratio (150/150)								
Tox21_AR_BLA_Antagonist_viability (207/207)								
Tox21_AR_LUC_MDAKB2_Agonist (90/90)								
Tox21_AR_LUC_MDAKB2_Antagonist (56/56)								
Tox21_AR_LUC_MDAKB2_Antagonist_v (291/291)								
Tox21_Aromatase_Inhibition (155/155)			M: 15 mg/L					
Tox21_Aromatase_Inhibition_viability (303/303)								
Tox21_AutoFluor_HEK293_Cell_red (7/7)								
Tox21_AutoFluor_HEK293_Media_blue (16/16)								
Tox21_AutoFluor_HEK293_Media_green (5/5)								
Tox21_AutoFluor_HEK293_Media_red (7/7)								
Tox21_AutoFluor_HEPG2_Cell_blue (18/18)								
Tox21_AutoFluor_HEPG2_Cell_green (6/6)								

1. Click on "Data Gap Filling"; 2. Highlight the empty cell next to the AC50 endpoint associated with illustrated assay: "NCGC Reporter Gene Assay ERA Agonist" 3. Select "Trend analysis"; 4. A window alerting you for data inconsistencies appears. Keep it as is. Click "OK".

Types endpoint correlations

Continuous vs. continuous

Define target endpoint – step 2

The screenshot shows the QSAR Toolbox software interface during a data gap filling workflow. The main window displays a 'Filter endpoint tree...' on the left and a data table on the right. The table has columns for target endpoints (1, 3, 5, 9, 20, 28, 40, 41) and rows for various chemical categories. An information dialog box is overlaid on the table, stating '19 observed values for 18 chemicals were excluded due to missing X descriptor value(s)'. A blue callout bubble with the number '1' points to the 'OK' button in the dialog box.

The message informing the user for how many chemicals with experimental data are excluded from gap filling due to missing X descriptor values appears. 1. Click “OK”;

Types endpoint correlations

Continuous vs. continuous

Enter Gap filling – step 3

1 Data Gap Filling stage

2 Trend analysis

3 AC50 (log(1/(1-0.5)))

4 log Kow

Filter endpoint tree...	1 [target]	3	4	5	6	10	14	18
Structure								
AC50 (356/496)	M: 19.1 mg/L	M: 0.00017 mg/L	M: 0.00016 mg/L	M: 1.53 mg/L	M: 9.43 mg/L	M: 4.53 mg/L	M: 2.32 mg/L	M: 13.3 mg/L
Tox21_AHR (83/83)				M: 7.62 mg/L		M: 9.37 mg/L		
Tox21_AHR_viability (83/83)								
Tox21_AR_BLA_Agonist_ch1 (115/115)	M: 0.00436 mg/L		M: 0.242 mg/L	M: 16.5 mg/L				M: 0.435 mg/L
Tox21_AR_BLA_Agonist_ch2 (63/63)			M: 0.183 mg/L		M: 23.3 mg/L			
Tox21_AR_BLA_Agonist_ratio (45/45)			M: 0.311 mg/L					
Tox21_AR_BLA_Antagonist_ratio (46/46)		M: 3 mg/L			M: 4.62 mg/L			
Tox21_AR_BLA_Antagonist_viability (67/67)	M: 14.4 mg/L				M: 12.7 mg/L			
Tox21_AR_LUC_MDAKB2_Agonist (38/38)			M: 0.152 mg/L					
Tox21_AR_LUC_MDAKB2_Antagonist (10/10)								
Tox21_AR_LUC_MDAKB2_Antagonist_via (62/62)						M: 0.000104 mg/L	M: 6.52 mg/L	
Tox21_Aromatase_inhibition (38/38)								
Tox21_viability (62/62)								
Tox21_via_blue (6/6)							M: 4.19 mg/L	
Tox21_via_blue (6/6)								

Trend analysis prediction for AC50, based on 356 values
 Predicted: 1.93 mg/L
 Model equation: AC50 = 4.92 (±0.176) + 0.0382 (±0.0422) * log Kow, log(1/mol/L)

Enter Gap filling applying trend analysis. Trend analysis is applied because the target endpoint is in continues range of data and there is enough data to build a linear regression.

1. Data Gap filling stage 2. Trend analysis approach is applied 3. AC50 endpoint related to ER enzyme assay ia plotted on Y-axis 4. Pay attention that default descriptor displayed on X-axis is log Kow.

Types endpoint correlations

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with AC50 data – step 4

The screenshot shows the QSAR Toolbox interface during the 'Data Gap Filling' step. The main window is titled 'Filter endpoint tree...' and displays a tree structure on the left and a table of chemical structures and their corresponding AC50 values on the right. A dialog box titled 'Select endpoint descriptor' is open, showing a tree of descriptors with 'Human Health Hazards (357/16310)' expanded. A callout bubble with the number '3' points to this dialog. On the right side of the interface, a vertical menu is visible with callouts '1' and '2'. Callout '1' points to the 'Descriptors /data' option, and callout '2' points to the 'Select endpoint tree descriptor' option. At the bottom, a scatter plot shows 'Trend analysis prediction for AC50, based on 356 values' with a regression line. The predicted value is 1.09 mg/L. The model equation is $AC50 = 432 (\pm 0.176) + 0.0382 (\pm 0.0422) * \log Kow. \log(1/mol/L)$.

1. Click on "Descriptors /data";
2. Go on "Select endpoint tree descriptor";
3. A window with arranged "Endpoint data tree" appears. Expand the endpoint tree;

Types endpoint correlations

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with other AC50 data – step 4

The screenshot displays the QSAR TOOLBOX interface. A 'Select endpoint descriptor' dialog box is open, showing a tree of endpoints. Callout 1 points to the 'NCGC (857/2569)' node. Callout 2 points to the 'AC50 (214/214)' node, which is highlighted with a red box. Callout 3 points to the 'OK' button. The background shows a 'Filter endpoint tree' with 'AC50' selected, and a 'Trend analysis prediction for AC50' graph with a scatter plot and a regression line. The graph shows 'AC50 [log(1/mol/L)]' on the y-axis and an unlabeled x-axis. The regression equation is $AC50 = 452 (\pm 0.176) + 0.0382 (\pm 0.0422) * \log Kow, \log(1/mol/L)$. The predicted value is 1.09 mg/L. The background also shows a table of chemical structures and their corresponding AC50 values.

28	40	41	58
<chem>C1=CC=NC=C1</chem>	<chem>CC1=CC=CC=C1</chem>	<chem>CC(O)C</chem>	<chem>[Na+]</chem>
M: 2.02 mg/L	M: 7.53 mg/L M: 1.97 mg/L	M: 2.77 mg/L	M: 0.75 mg/L
	M: 8.57 mg/L	M: 0.859 mg/L	
	M: 9.57 mg/L		
	M: 0.12 mg/L		
	M: 8.23 mg/L	M: 1.92 mg/L	

1. **Click** on “NCGC” node to open the sub-nodes; 2. **Select** endpoint, which will be placed on X-axis circled in red box; point the mouse on the level of AC50 (214/214); 3. **Click** “OK” button

Types endpoint correlations

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with other AC50 data – step 4

The screenshot displays the QSAR Toolbox interface during a data gap filling process. A dialog box titled "Possible data inconsistency" is open, prompting the user to select a gap filling scale/unit for data 214/214. The options are $\log(1/\text{mol/L})$ (selected) and μM . A red "1" in a blue box highlights the "OK" button. Below the dialog, a scatter plot titled "Trend analysis prediction for AC50, based on 356 values" shows a predicted value of 1.09 mg/L. The plot's y-axis is AC50 (log(1/mg/L)) and the x-axis is log Kow. A red regression line is visible. On the right side of the interface, there are several control buttons, including "Accept prediction" at the bottom right.

1. **Click** "OK" on the message alerting you for data inconsistency;
The aim of this example is to see how the data correlates.

Types endpoint correlations

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with other AC50 data – step 4

The screenshot displays the QSAR Toolbox interface. The 'Filter endpoint tree' dialog is open, showing a list of endpoints on the left and a table of predicted values on the right. The table includes columns for chemical IDs (1 [target], 3, 40, 41, 86, 87, 93, 94, 112) and their corresponding AC50 values. A message box is overlaid on the table, indicating that 143 observed values for 142 chemicals were excluded due to missing X descriptor values. A red '1' in a blue callout box points to the 'OK' button of this message box. Below the dialog, a scatter plot titled 'Trend analysis prediction for AC50, based on 356 values' shows the relationship between log Kow and AC50. The plot includes a regression line and the following statistics: Predicted: 1.09 mg/L, Model equation: $AC50 = 4.92 (\pm 0.176) + 0.0382 (\pm 0.0422) * \log Kow$.

1. **Click** "OK" on the message informing you excluded number of chemicals due to missing X-descriptor data. They are analogues with no such type AC50 data. This will not affect the value of correlation coefficient;

Types endpoint correlations

Continuous vs. continuous

Replacement of default X-descriptor (logKow) with other AC50 data – step 4

The screenshot displays the QSAR Toolbox software interface during a workflow step. The main window shows a data table with columns for chemical structures and their corresponding AC50 values. The 'Filter endpoint tree...' panel on the left lists various endpoints, with 'AC50' selected. The scatter plot at the bottom visualizes the relationship between the X-descriptor (log Kow) and the Y-descriptor (AC50). The plot shows a positive correlation, with a red regression line and a model equation: $AC50 = 1.08 (\pm 0.344) + 0.799 (\pm 0.0648) \cdot AC50 \cdot \log(1/mol/L)$. The predicted value is N/A. The right-hand panel contains settings for the analysis, including 'Select / filter data', 'Gap filling approach', and 'Accept prediction'.

1. The graph obtained after replacing log Kow with Toxcast endpoint is visualized;
2. The equation including endpoint data is rebuild;

Types endpoint correlations

Continuous vs. continuous

Interpretation of correlation results

- In this example, we have correlated two AC50 endpoints associated with different type assay
- As seen from the graph, a linear relationship between two endpoints has been observed
- In order to assess only the chemicals having positive estrogen activity we remove the “Non-binders” chemicals based on subcategorization by “Estrogen receptor binding by OASIS” profiler (illustrated on next slide)

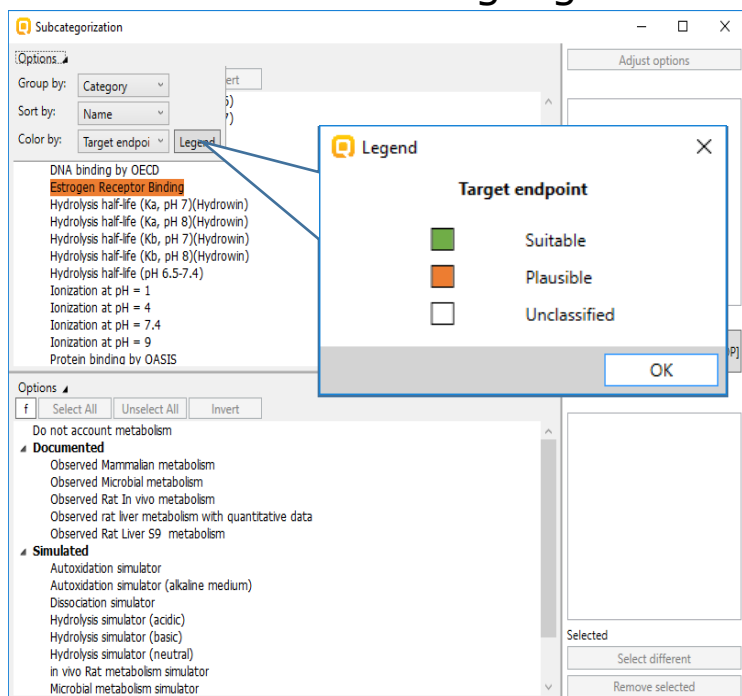
Types endpoint correlations

Continuous vs. continuous

Subcategorization by Estrogen receptor binding profiler

Sidebar of profiles relevancy

Once the endpoint is selected, the relevant profiles and metabolic transformations are highlighted.



- **Suitable** - developed using data/knowledge for the target endpoint;
- **Plausible** – structure-based; form broader group of analogues;
- **Unclassified** – all profilers, which are not classified in any of the categories above.

Types endpoint correlations

Continuous vs. continuous

Subcategorization by Estrogen receptor binding profiler

The screenshot displays the QSAR Toolbox interface for subcategorization. The top navigation bar includes 'Input', 'Profiling', 'Data', 'Category definition', 'Data Gap Filling', and 'Report'. The left sidebar shows various tool options, with 'Estrogen Receptor Binding' selected under the 'Documented' section. The central data table shows a grid of chemical structures and their predicted values for various endpoints. The bottom plot shows a trend analysis prediction for AC50, with a model equation: $AC50 = 1.08 (\pm 0.344) + 0.799 (\pm 0.0648) * AC50_{log(1/mol/L)}$. The predicted value is N/A. The 'Adjust options' dialog is open, showing the selected category 'Non binder, without OH or NH2 group' and the 'Remove selected' button.

1. **Open** "Select/filter data" menu item, then **click** "Subcategorize";
2. **Select** "Estrogen receptor binding" profiler;
3. **Select** only Non binder categories by **left mouse click** and **hold** "Ctrl" button;
4. **Click** "Remove" button;

Types endpoint correlations

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint

The screenshot displays the QSAR Toolbox interface. On the left, the 'Subcategorization' window is open, showing a list of simulation options. Callout 1 points to 'Estrogen receptor binder' under the 'Documented' section. Callout 2 points to '(7) Moderate binder, OH' under the 'Options' section. The central table shows a grid of chemical structures and their corresponding AC50 values. Callout 3 points to the highlighted rows in this table. At the bottom, a scatter plot shows the correlation between AC50 [log] (x-axis) and AC50 [log] (y-axis). Callout 4 points to the highlighted data points in the plot. A text box at the bottom right of the plot reads "Moderate binders" vs. AC50 data. The plot also includes a trend line and a legend for 'Accept prediction'.

1. Click again on Estrogen receptor binding profiler
2. **Select** "Moderate binder" categories
3. The chemicals corresponding to the selected categories are highlighted in green; 4 and in light blue on the graph

Types endpoint correlations

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint

The screenshot displays the QSAR Toolbox interface. On the left, the 'Subcategorization' panel is open, showing a list of categories. The 'Estrogen Receptor Binding' category is selected, and its sub-categories are listed. A red callout box with the number '1' points to the 'Weak binder, OH group' category. The main table shows a grid of chemical structures and their AC50 values. A red callout box with the number '2' points to the highlighted cells in the table. Below the table, a 'Trend analysis prediction for AC50' plot is shown, with a red regression line and a text box stating 'Weak binders vs. AC50 data'. The plot shows a positive correlation between AC50 [log] and AC50 [log(1/mol/L)].

Category	191	257	262	266	292	295	323	330	355
Non binder, without OH									
ERa Antagoni(25/31)	M: 3.27 mg/L	M: 0.000224 mg/L	M: 5.52 mg/L	M: 0.666 mg/L	M: 2.96 mg/L	M: 5.58 mg/L	M: 1.27 mg/L	M: 0.131 mg/L	M: 6.82 mg/L
(72/123)									
(24/24)			M: 1.21 mg/L	M: 2.7 mg/L	M: 2.12 mg/L	M: 0.177 mg/L	M: 11 mg/L	M: 0.227 mg/L	
(12/12)	M: 0.0879 mg/L								
(26/26)			M: 0.214 mg/L	M: 0.0102 mg/L	M: 0.000144 mg/L				M: 0.000329 mg/L
(8/8)									
(4/4)									
atio									
liability								M: 6.57 mg/L	
onist				M: 5.75 mg/L			M: 8.49 mg/L	M: 2.58 mg/L	
tagonist									
tagonist_via(13/13)									M: 8.95 mg/L
(10/10)									
(1) Strong binder, NH2 g									
(26) Strong binder, OH g									M: 5.93 mg/L
(2) Very strong binder, dia_blue									
(2/2)									
(3) Weak binder, NH2 g			M: 4.45 mg/L						
(2/2)									
(13) Weak binder, OH g									
(1/1)									

1. **Select** "Weak binder" categories (left mouse click and hold "Ctrl" button);
2. The chemicals corresponding to the selected categories are highlighted in green;

Types endpoint correlations

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint

The screenshot displays the QSAR Toolbox interface for subcategorization. The left sidebar shows a list of categories, with 'Strong binder, NH2 group' and 'Very strong binder, OH group' highlighted. The central data table shows chemical structures and their corresponding AC50 values for various categories. The bottom plot shows a trend analysis prediction for AC50, with a model equation: $AC50 = 0.409 (\pm 0.285) + 0.953 (\pm 0.0497) * AC50_{log(1/mg/L)}$. A callout box '1' points to the selected categories in the sidebar, and a callout box '2' points to the highlighted green cells in the data table and the corresponding data points in the plot.

1. **Select** "Strong and very strong binder" categories (left mouse click and hold "Ctrl" button)
2. The chemicals corresponding to the selected categories are highlighted in green;

Types endpoint correlations

Continuous vs. continuous

Correlation results

- The two AC50 endpoints associated with different type assay have been correlated each other
- Non binders according to Estrogen receptor binding profiler have been eliminated from the correlation
- User can analyse the distribution of remaining ER binders (Very strong, Strong, Moderate and Weak) across selected AC50 endpoint