QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

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**

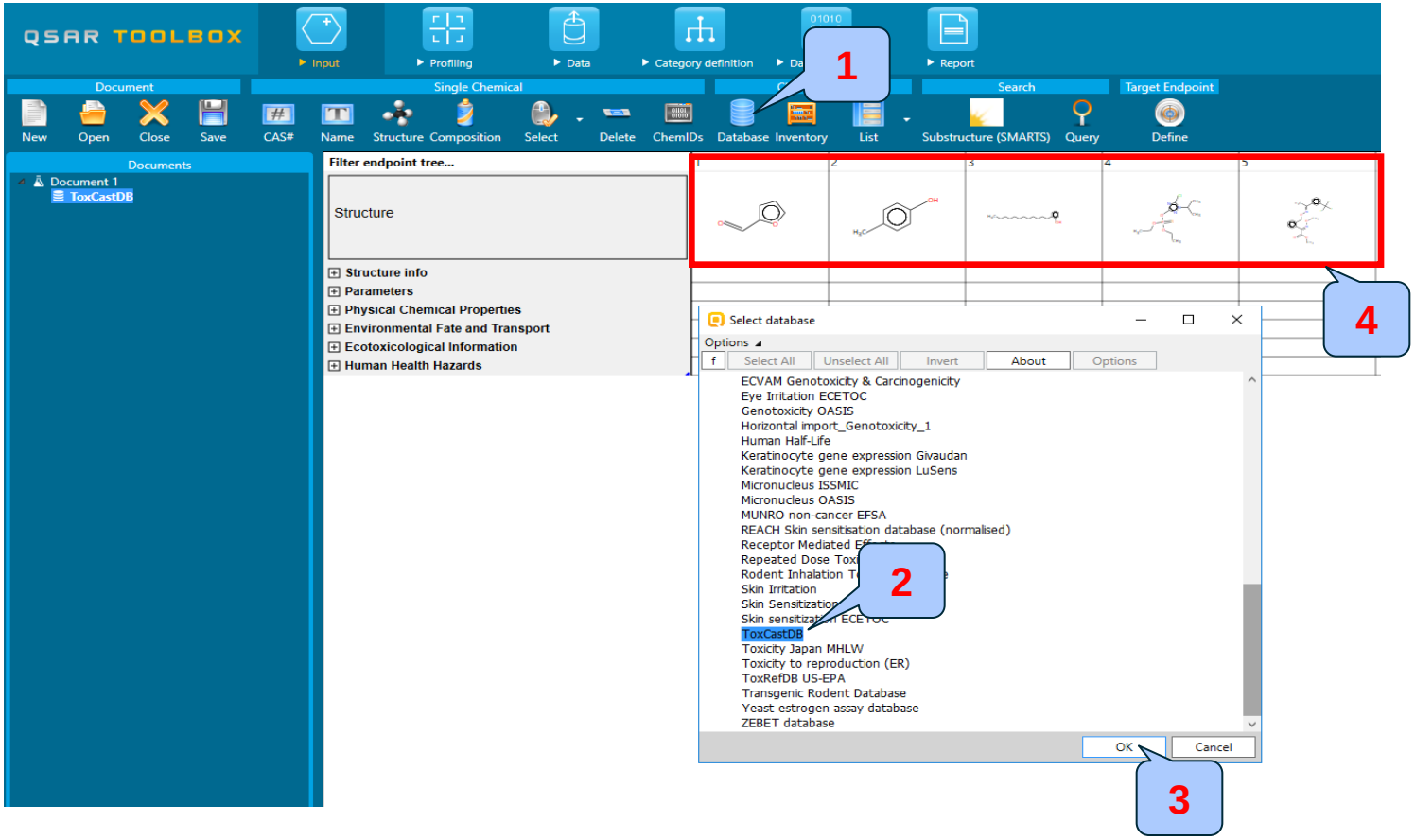
Workflow

- **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

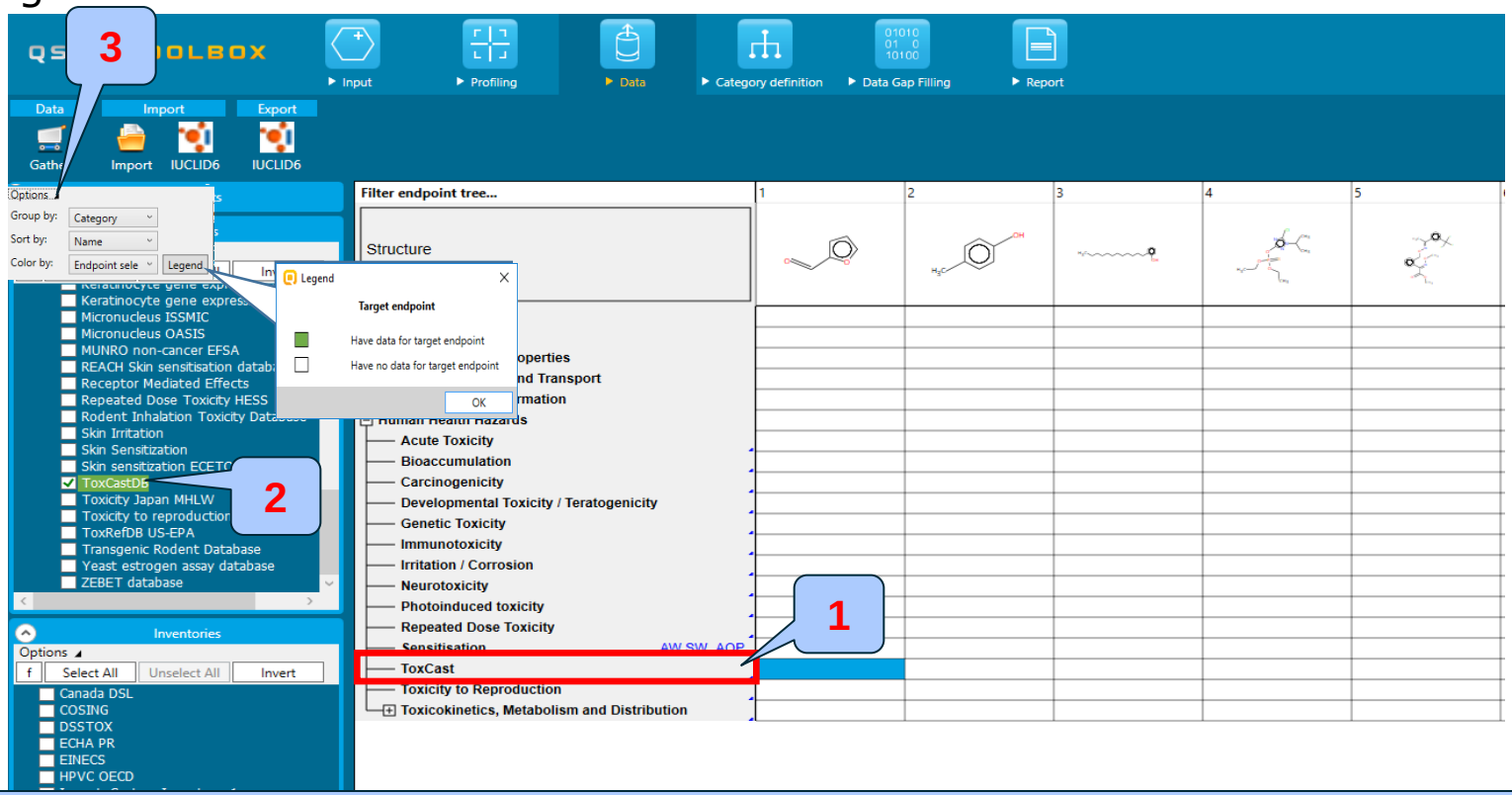


1. **Click** on "Database" button; 2. **Select** "ToxCast DB"; 3. **Click** "OK"; 4. Chemicals are loaded on data matrix

ToxCast database

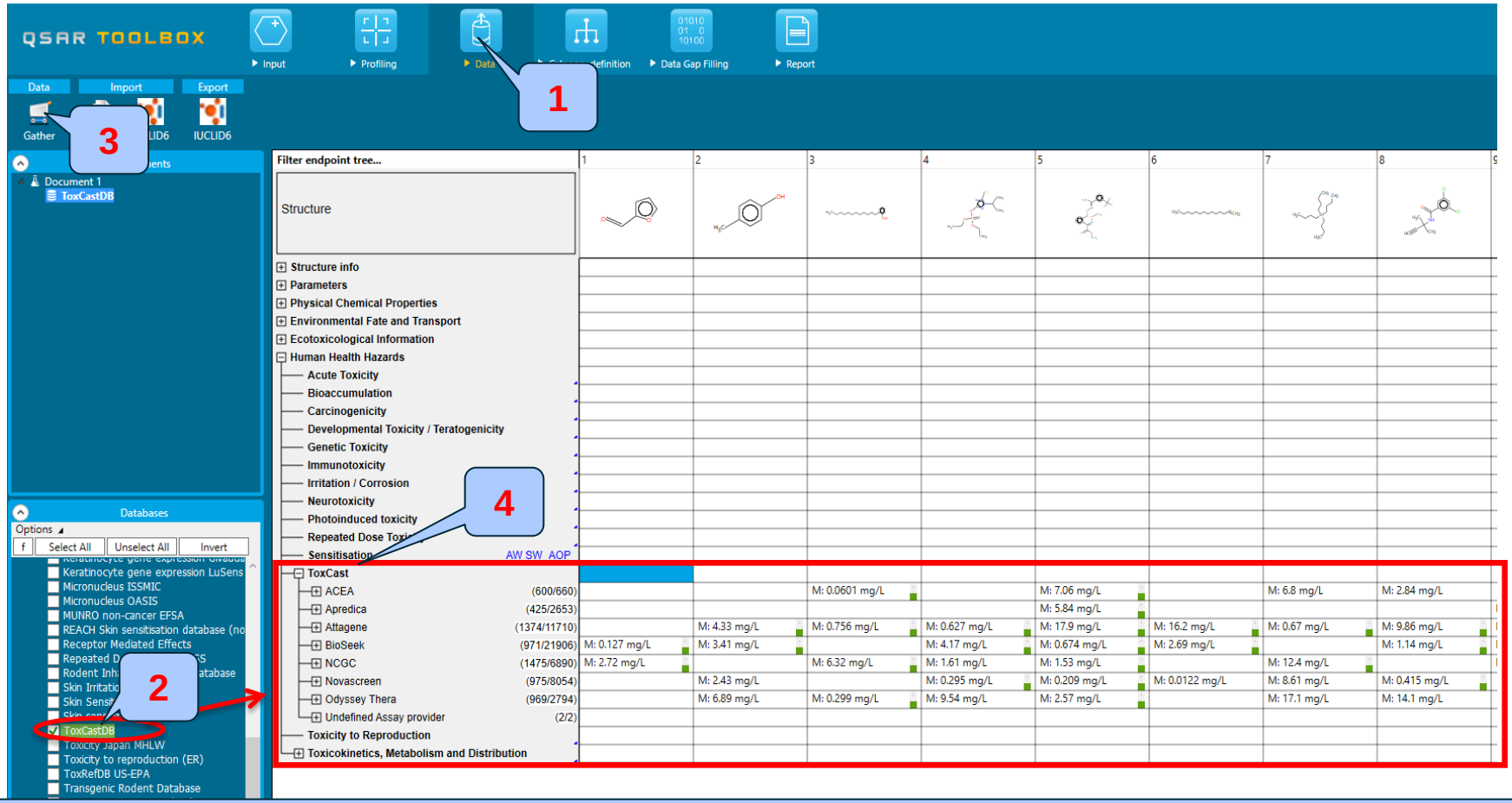
Sidebar of database relevancy

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



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"

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 introduce 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 continues type endpoint data or so called ratio data correlates each other (e.g.LC50 vs. EC50 data)

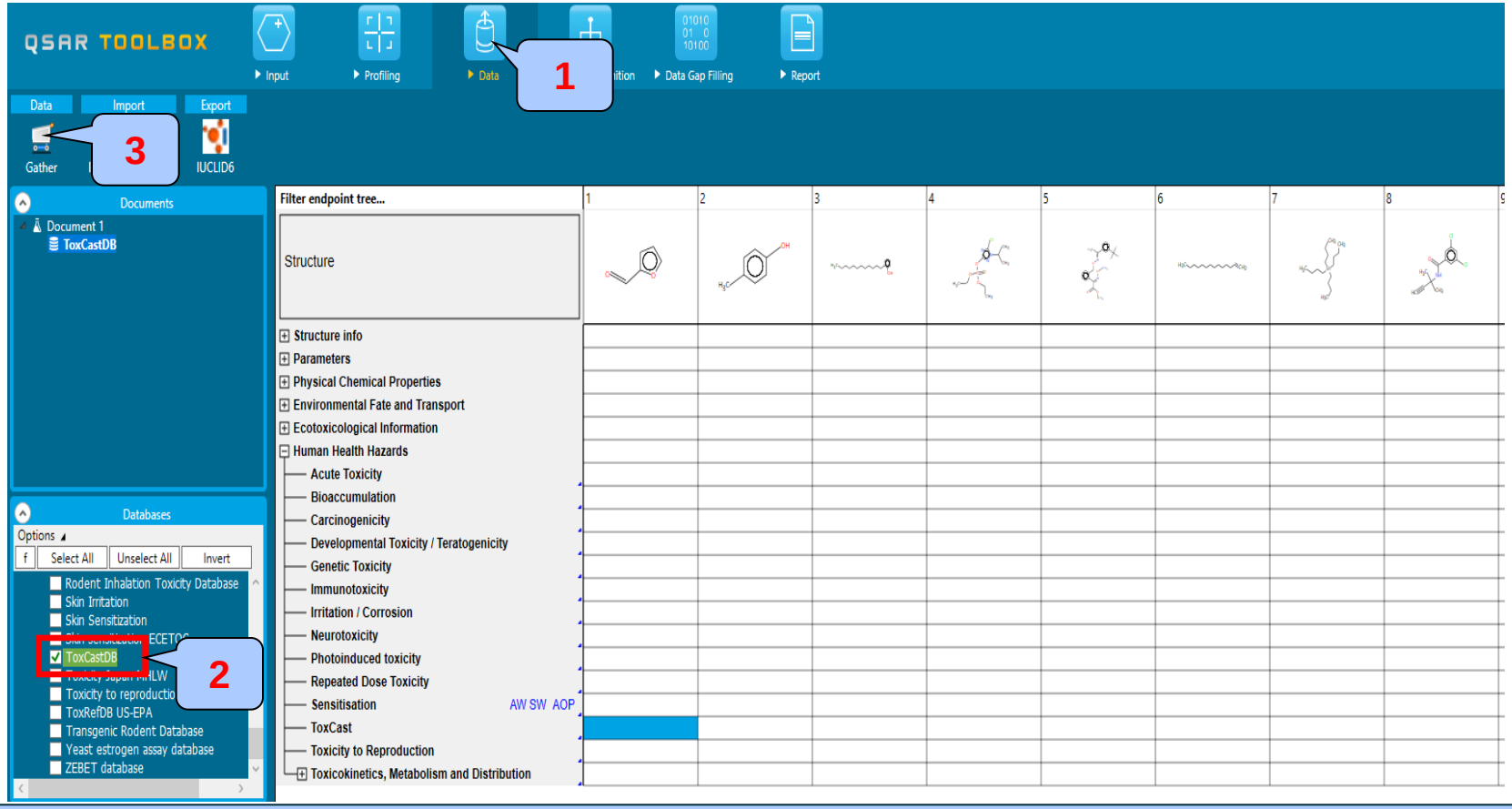
- In this example we will illustrated how AC50 data associated with two different test assays extracted from ToxCast DB correlates each other:
  - NCGC Reporter Gene Assay ERa Agonist, Estrogen receptor 1 (assay 1)
  - Tox21_Erα_BLA_Agonist_ch2 (assay 2)

- 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



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

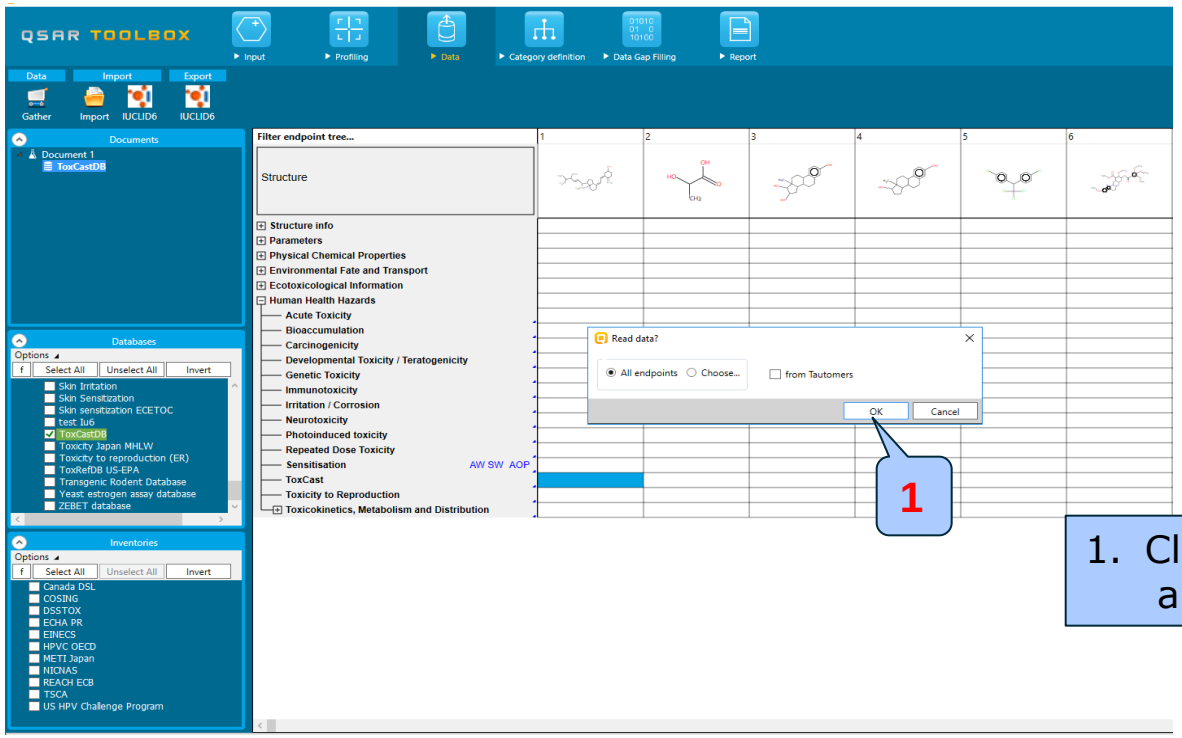
Types endpoint correlations

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.

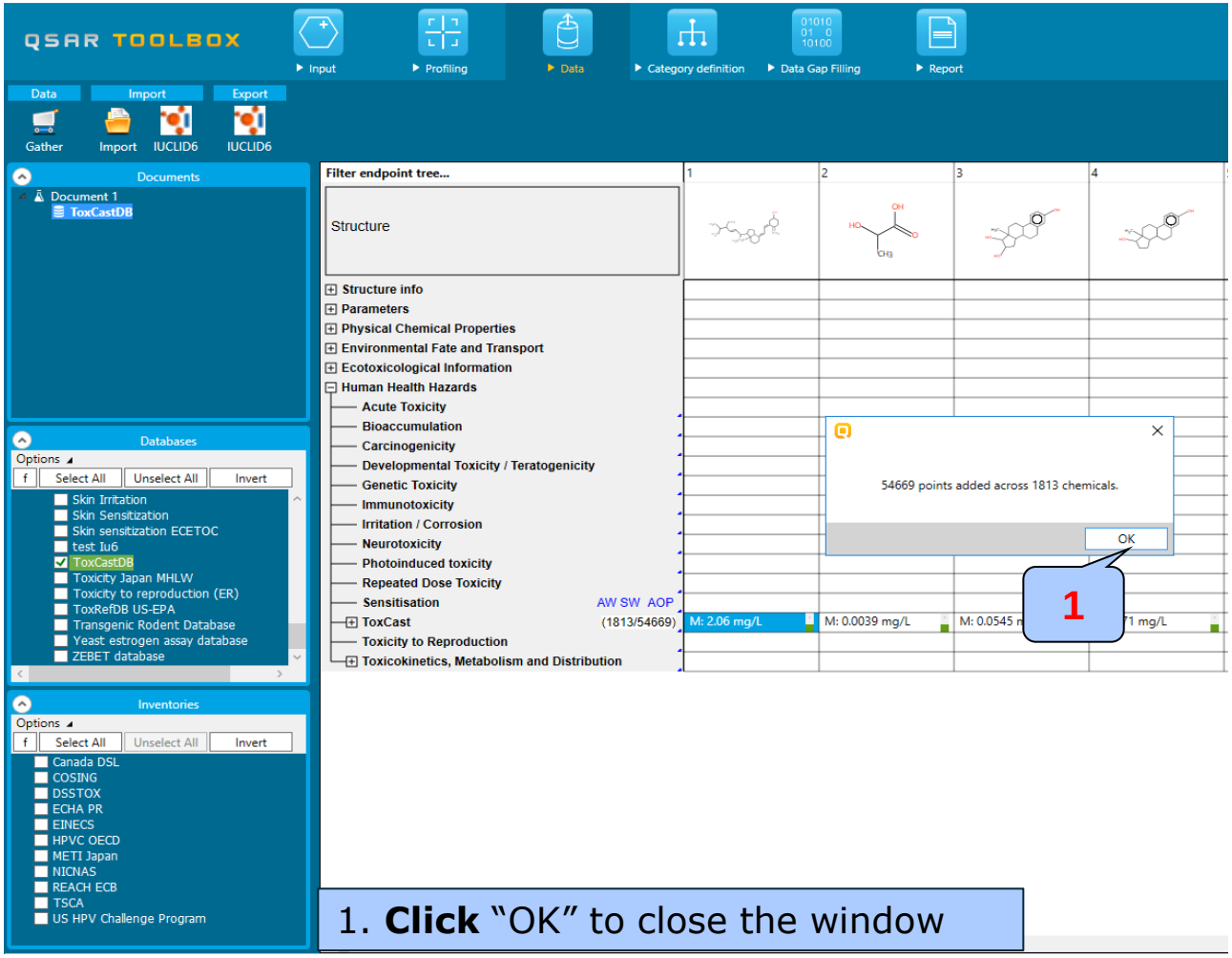


1. Click **OK** to read all available data

Types endpoint correlations

Continuous vs. continuous

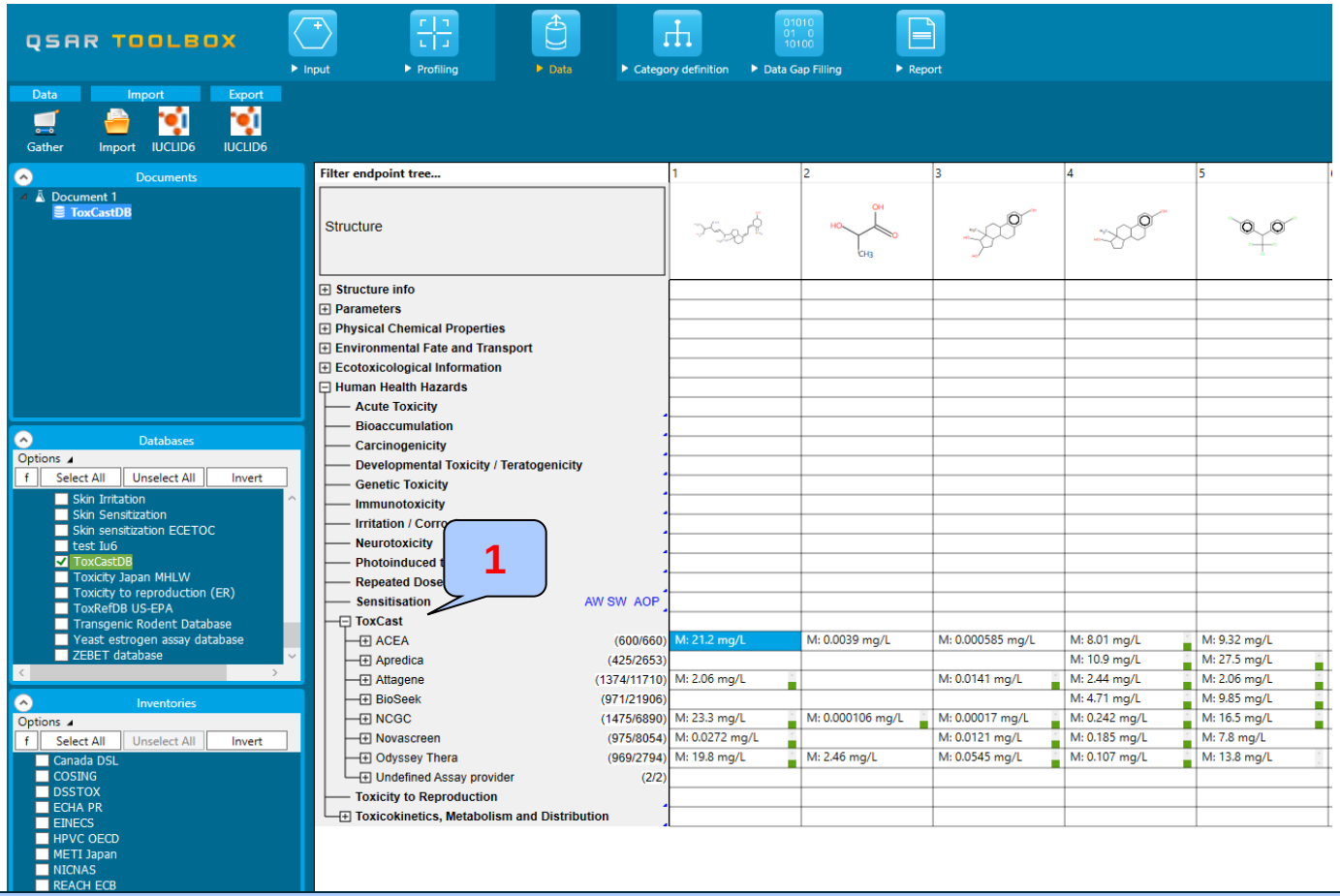
Gather experimental data – step 1



Types endpoint correlations

Continuous vs. continuous

Gather experimental data – step 1

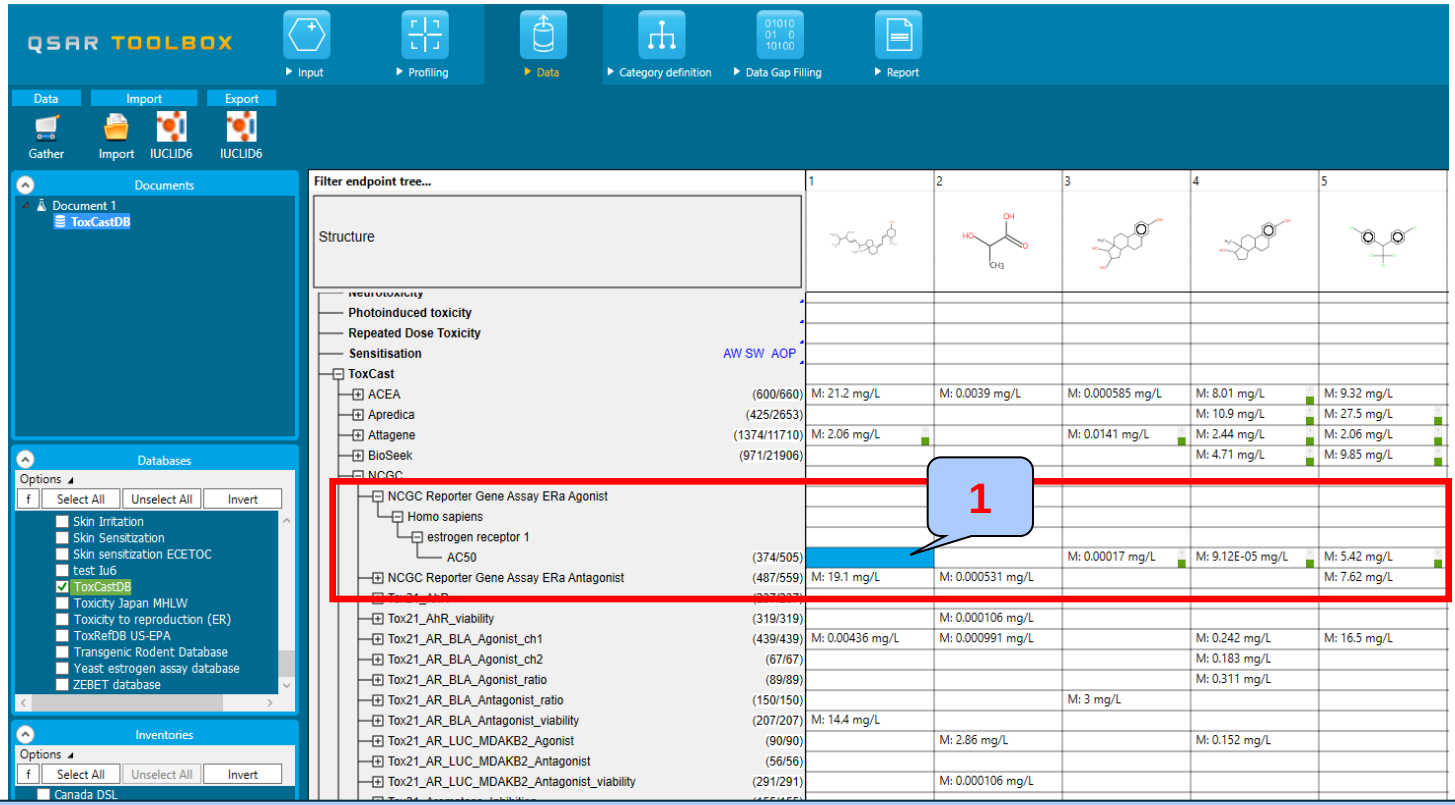


1. **ToxCast data** has been loaded on datamatrix in a separate node of "Endpoint tree" called "ToxCast"

Types endpoint correlations

Continuous vs. continuous

Define target endpoint – step 2



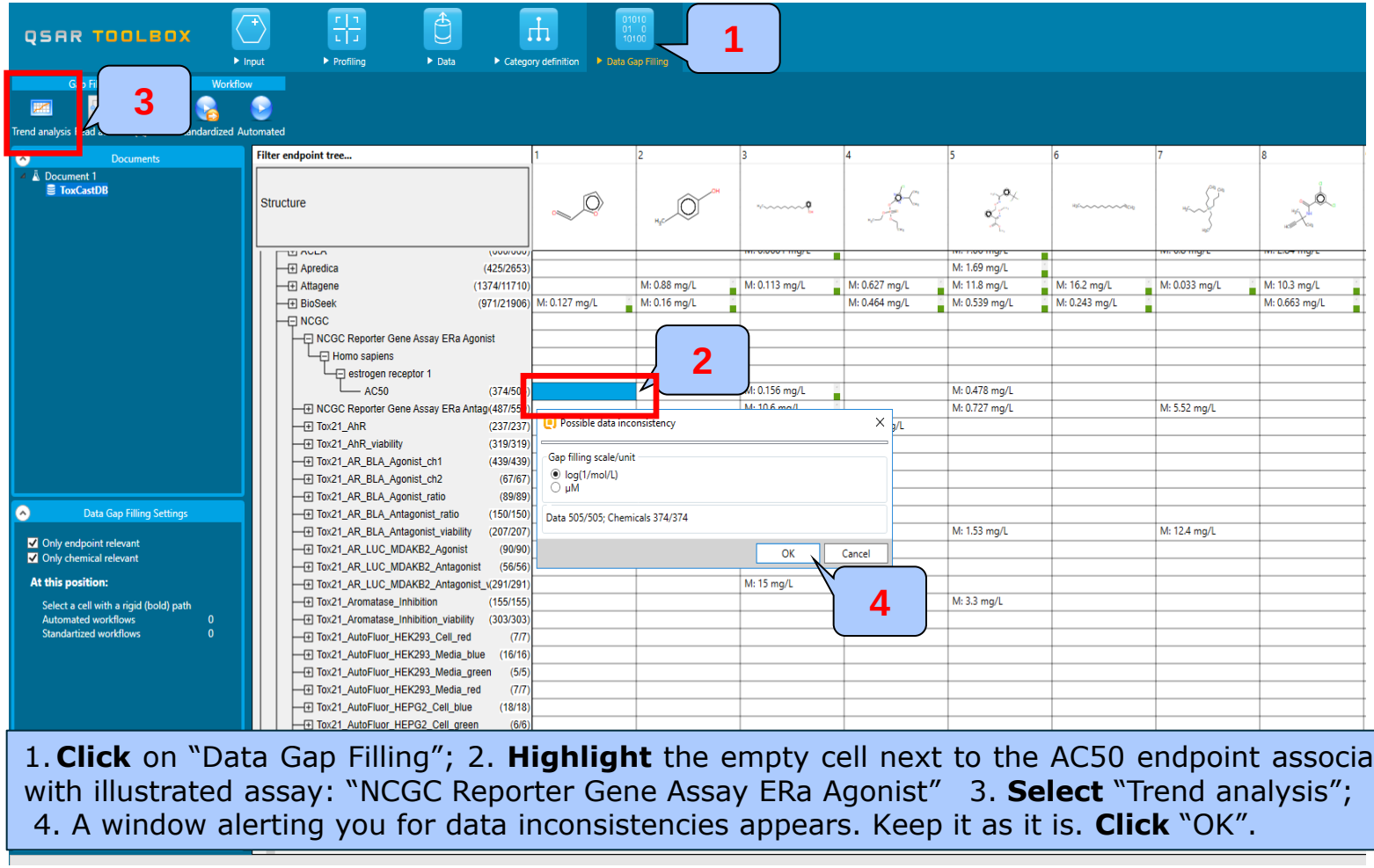
The target endpoint is AC50 associated with assay "NCGC Reporter Gene Assay ERa Agonist"

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

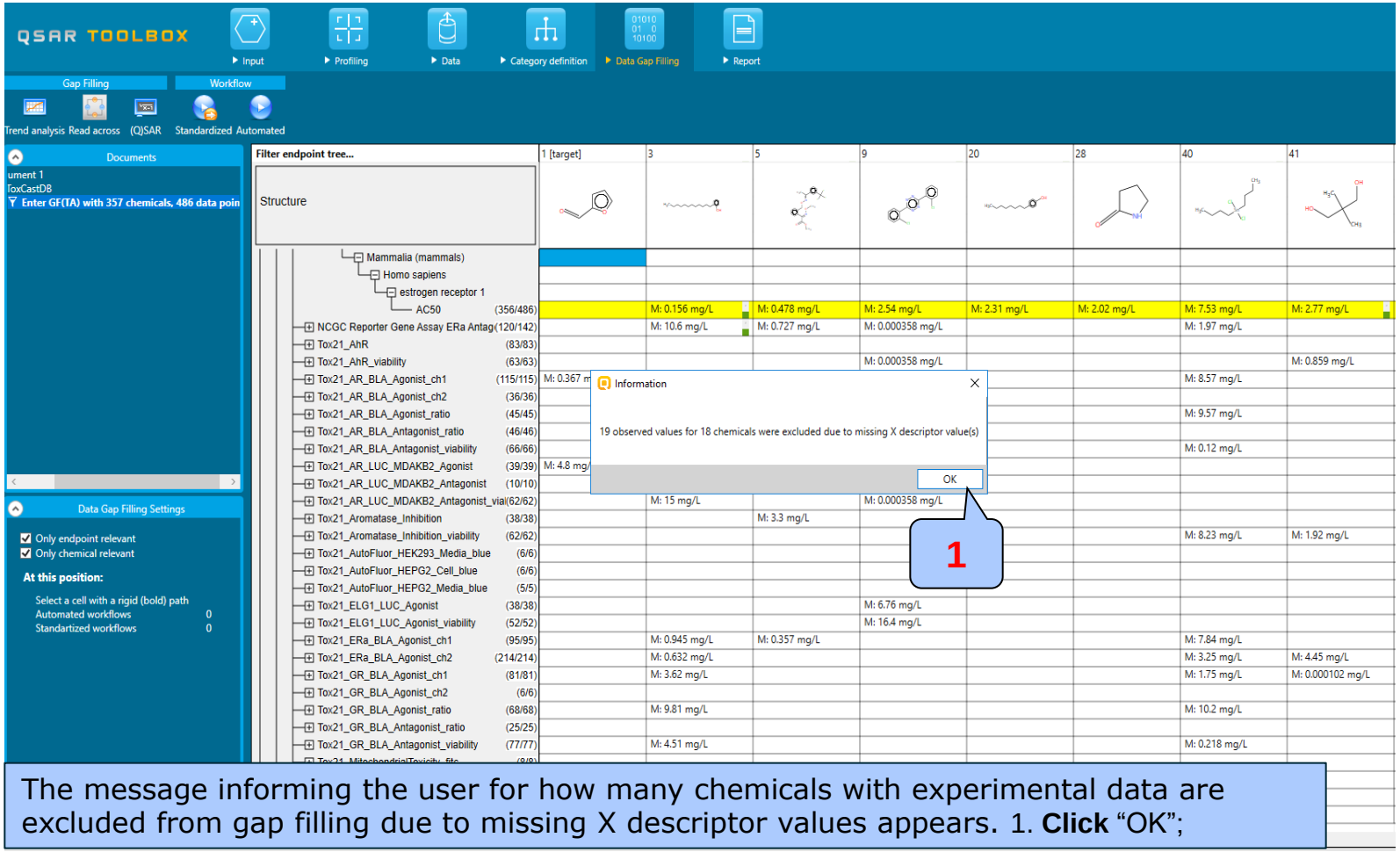


1. **Click** on "Data Gap Filling"; 2. **Highlight** the empty cell next to the AC50 endpoint associated with illustrated assay: "NCGC Reporter Gene Assay ERa Agonist" 3. **Select** "Trend analysis"; 4. A window alerting you for data inconsistencies appears. Keep it as it is. **Click** "OK".

Types endpoint correlations

Continuous vs. continuous

Define target endpoint – step 2

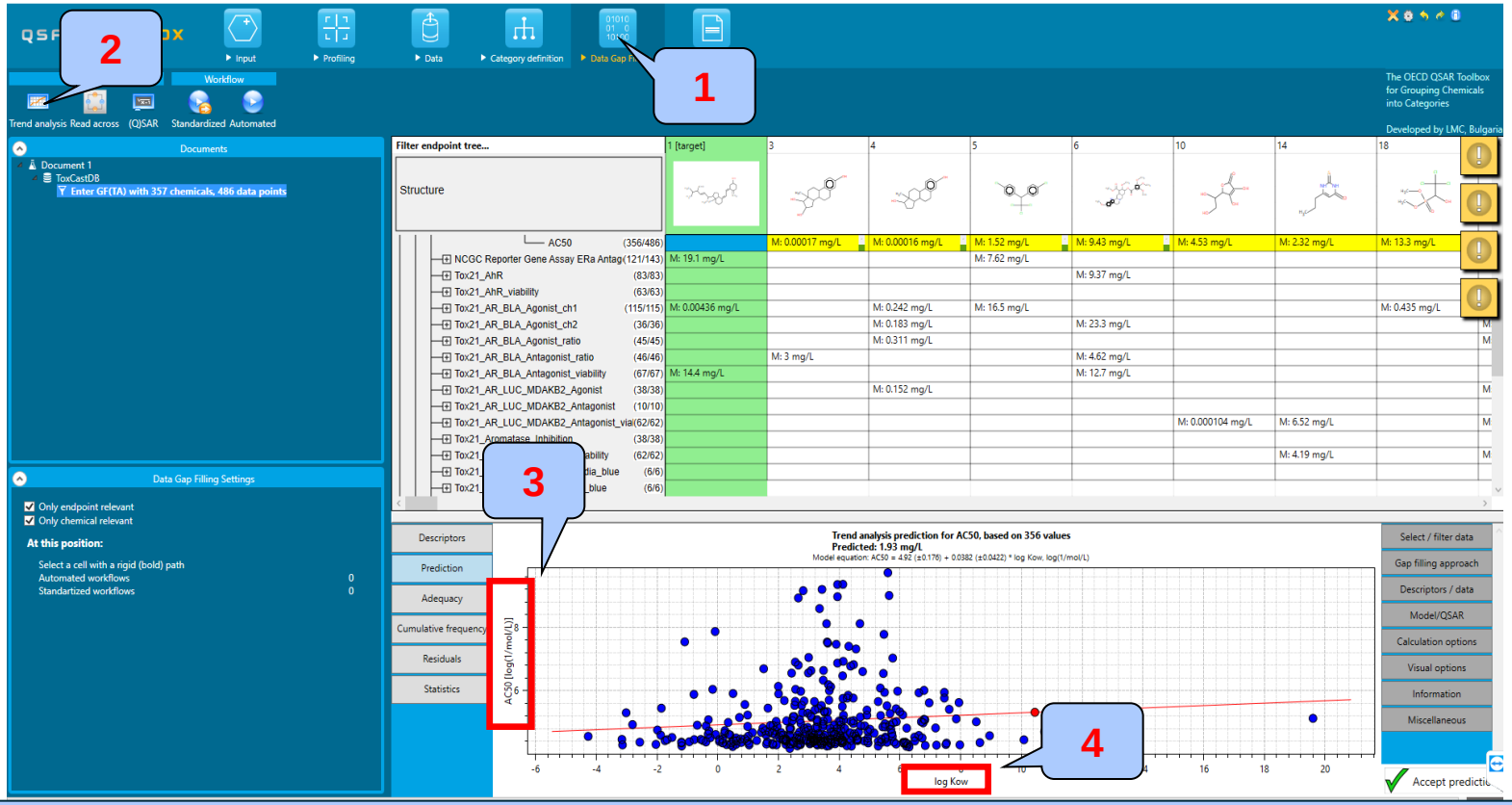


The message informing the user for how many chemicals with experimental data are excluded from gap filling due to missing X descriptor values appears. 1. **Click** "OK";

Types endpoint correlations

Continuous vs. continuous

Enter Gap filling – step 3

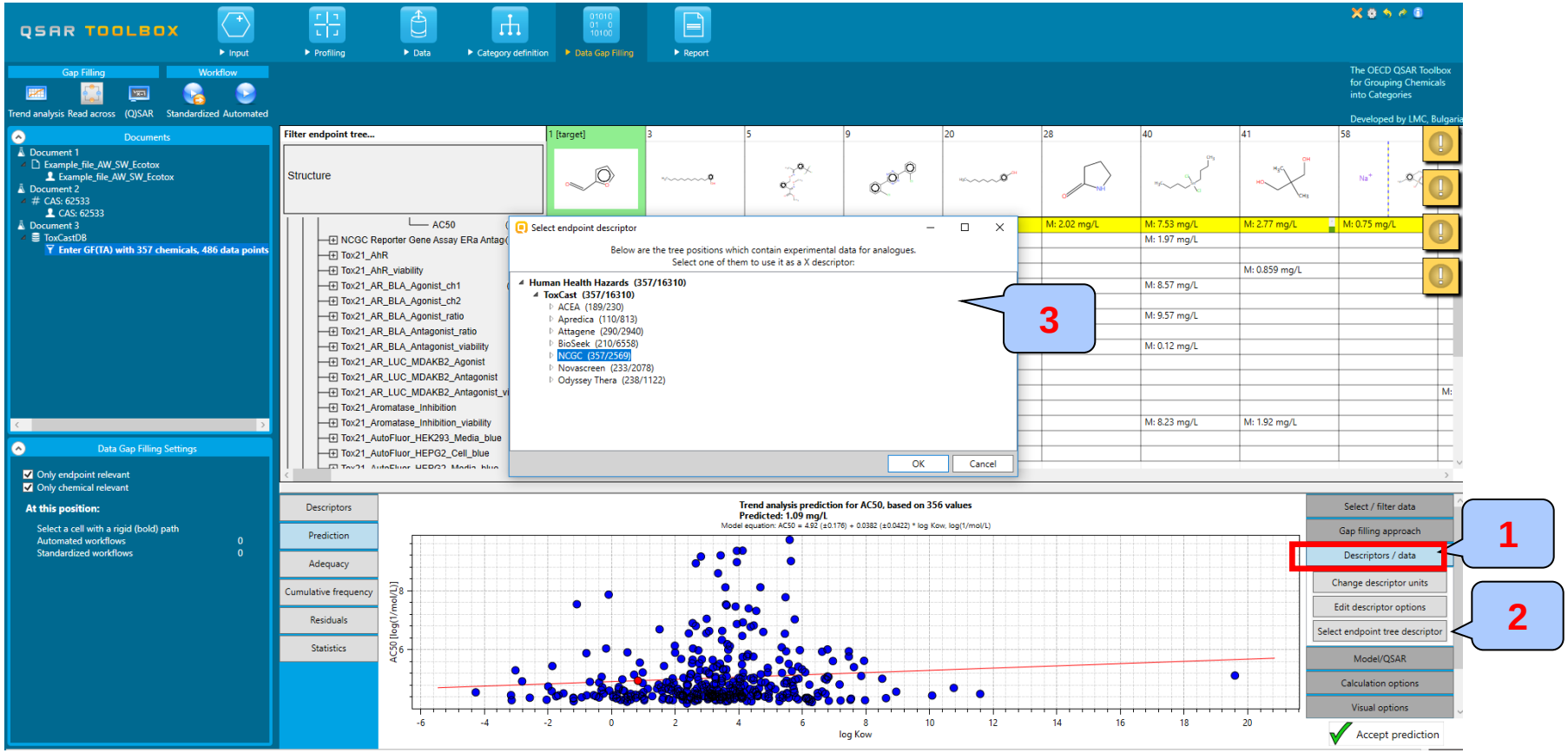


Enter Gap filling applying trend analysis. Trend analysis is applied because the target endpoint is in continues range of data and there is enough data to build a linear regression.
1. Data Gap filling stage 2. Trend analysis approach is applied 3. AC50 endpoint related to ER enzyme assay ia plotted 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

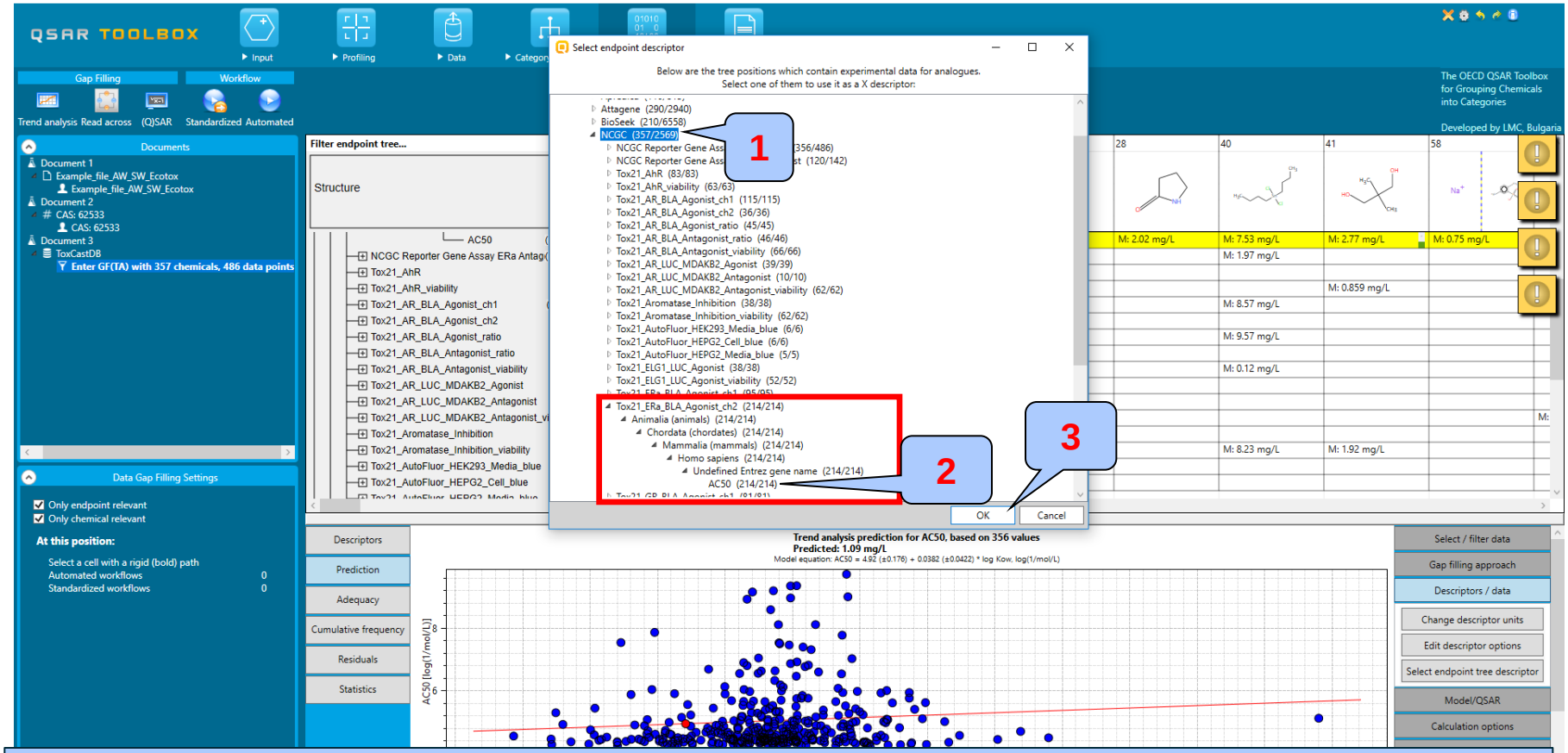


1. **Click** on "Descriptors /data"; 2. **Go** on "Select endpoint tree descriptor"; 3. A window with arranged "Endpoint data tree" appears. Expand the endpoint tree;

Types endpoint correlations

Continuous vs. continuous

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

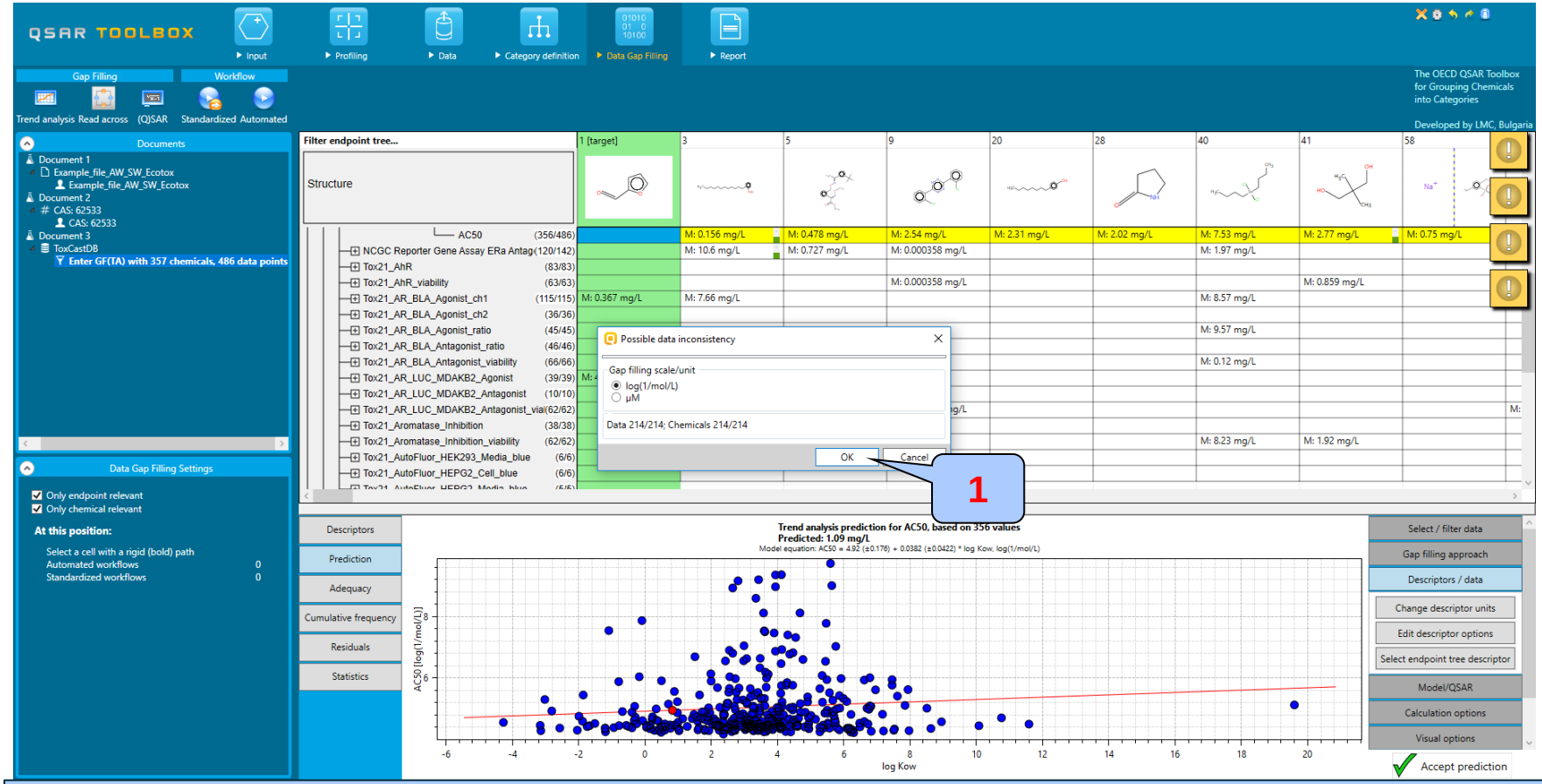


1. **Click** on "NCGC" node to open the sub-nodes; 2. **Select** endpoint, which will be placed on X-axis circled in red box; point the mouse on the level of AC50 (214/214); 3. **Click** "OK" button

Types endpoint correlations

Continuous vs. continuous

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

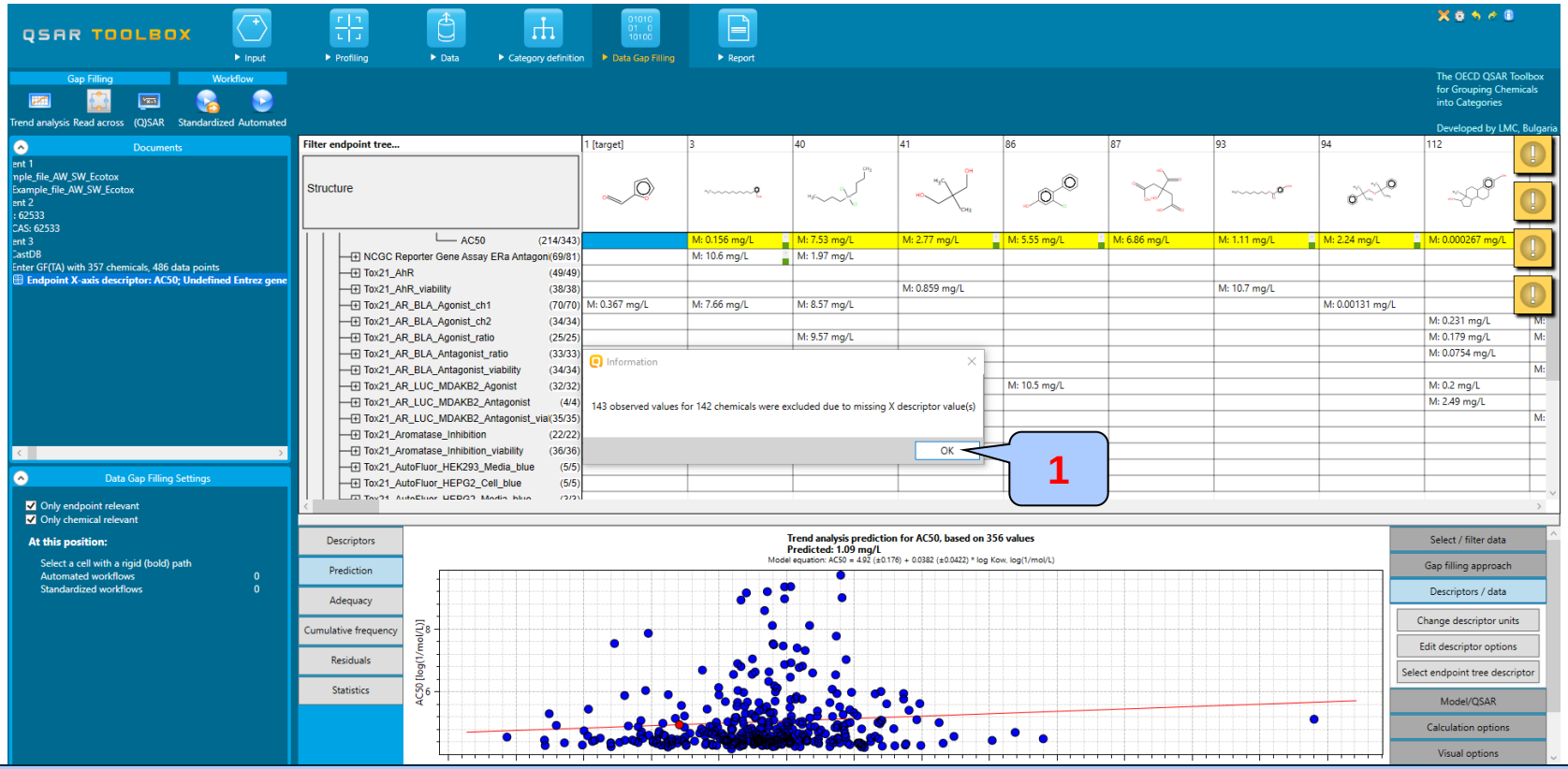


1. **Click** "OK" on the message alerting you for data inconsistency;
The aim of this example is to see how the data correlates.

Types endpoint correlations

Continuous vs. continuous

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

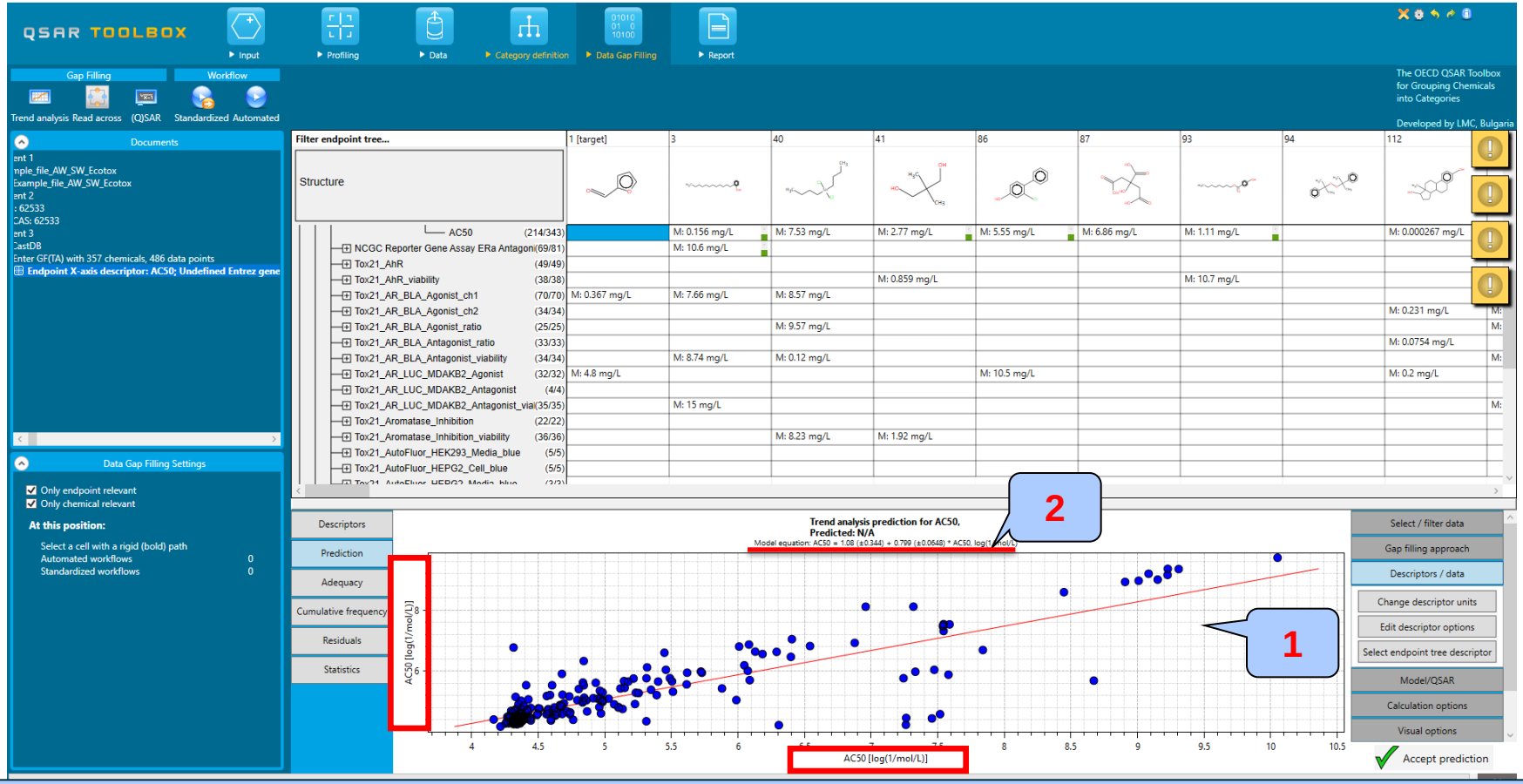


1. **Click** "OK" on the message informing you excluded number of chemicals due to missing X-descriptor data. They are analogues with no such type AC50 data. This will not affect the value of correlation coefficient;

Types endpoint correlations

Continuous vs. continuous

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



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)

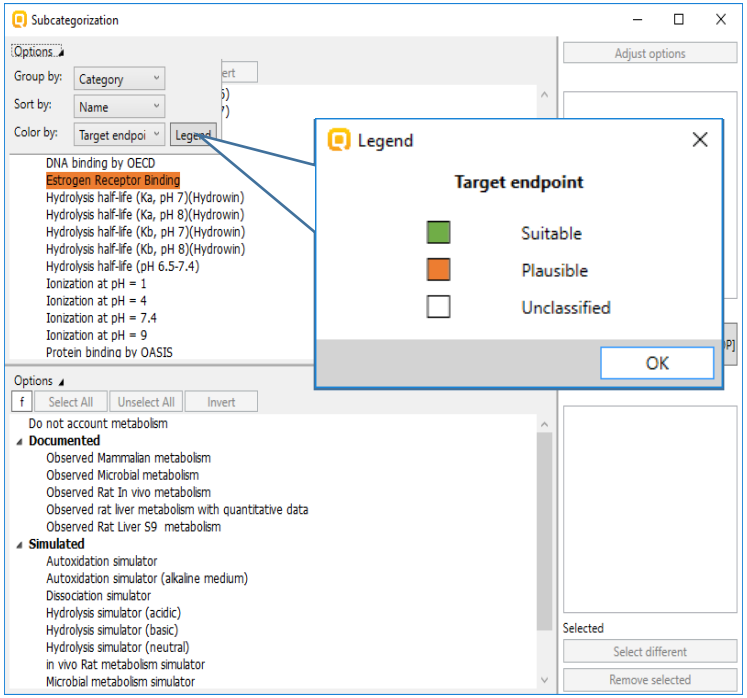
Types endpoint correlations

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.

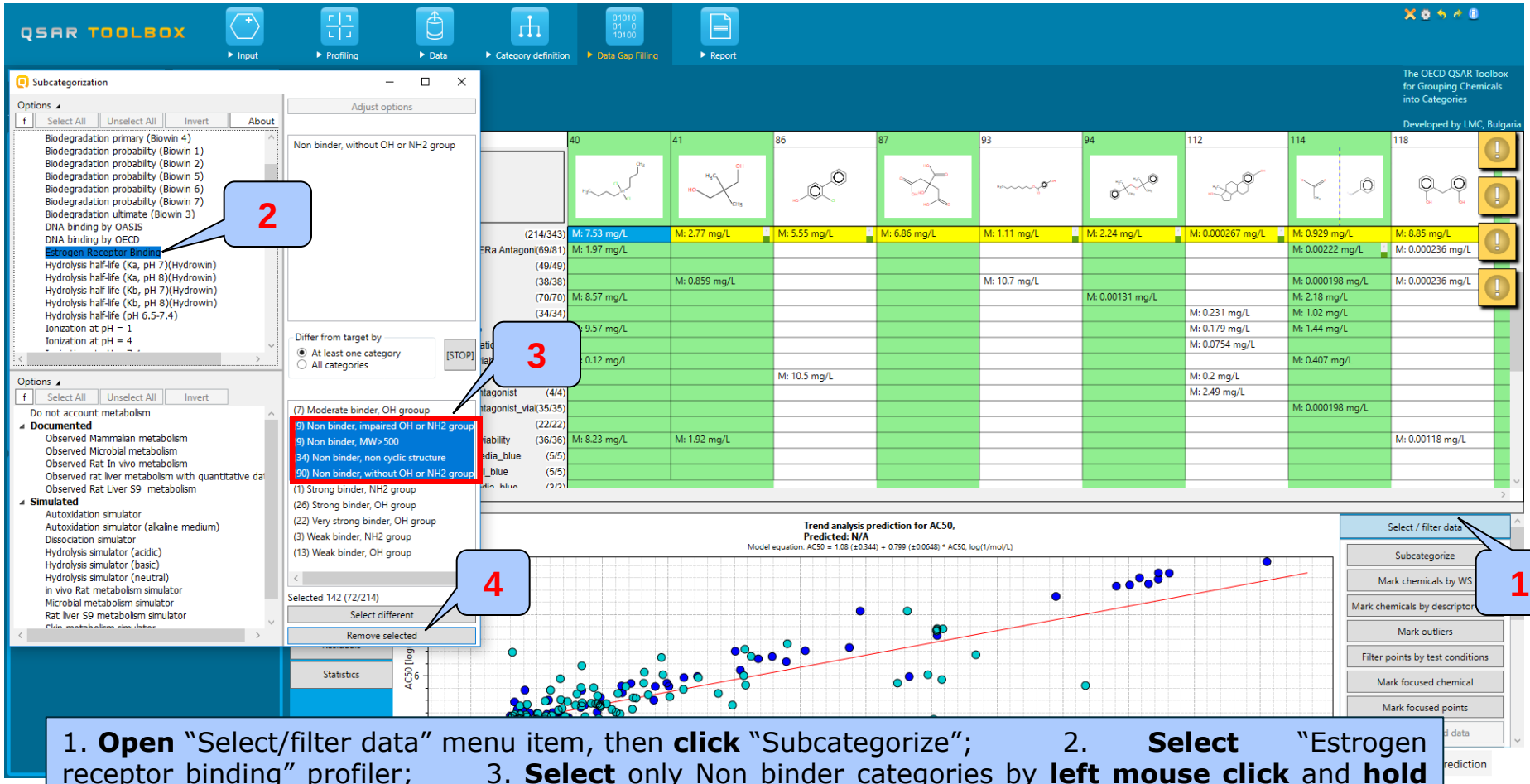


- **Suitable** - developed using data/knowledge for the target endpoint;
- **Plausible** – structure-based; form broader group of analogues;
- **Unclassified** – all profilers, which are not classified in any of the categories above.

Types endpoint correlations

Continuous vs. continuous

Subcategorization by Estrogen receptor binding profiler

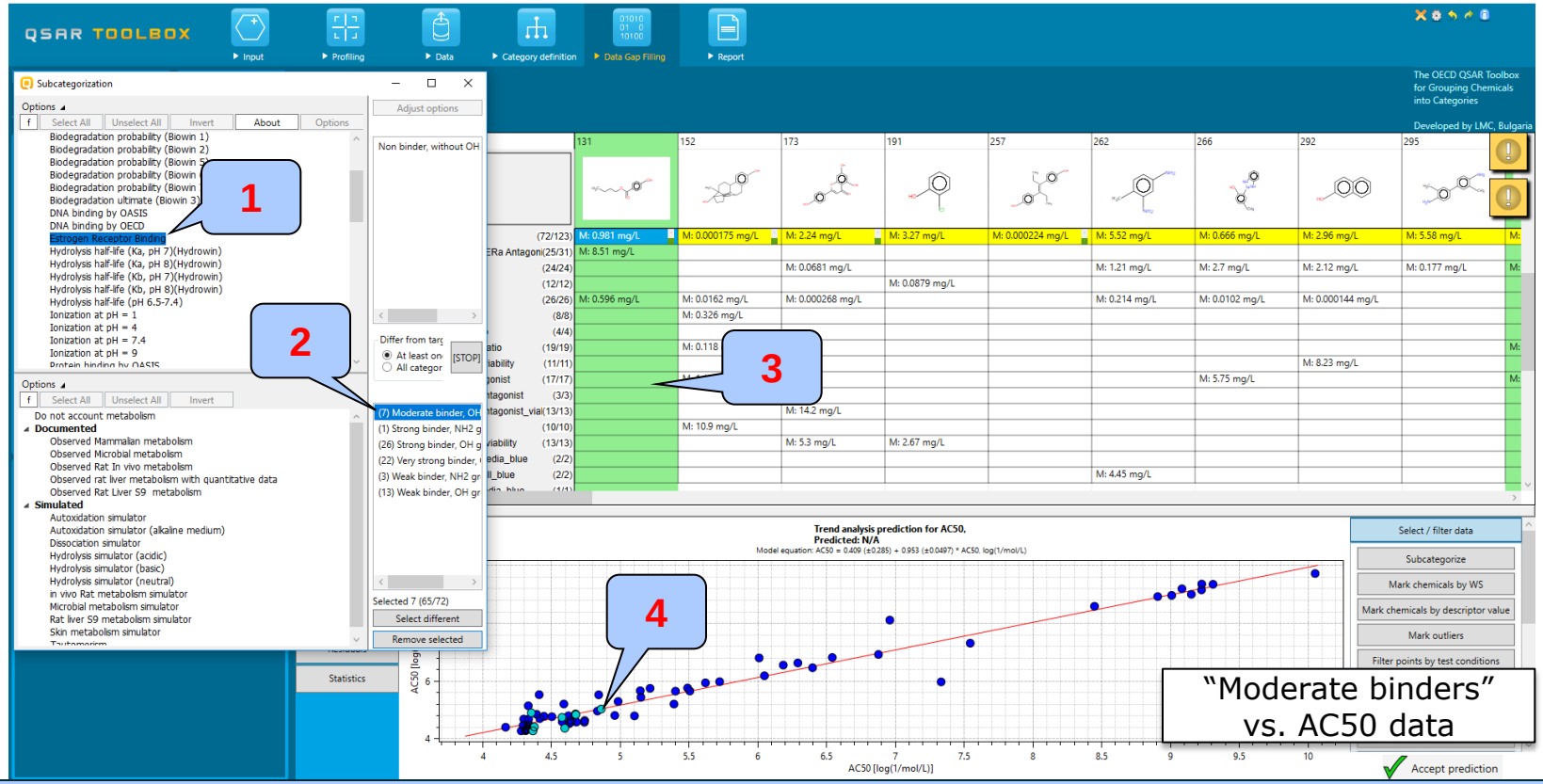


1. **Open** "Select/filter data" menu item, then **click** "Subcategorize"; 2. **Select** "Estrogen receptor binding" profiler; 3. **Select** only Non binder categories by **left mouse click** and **hold** "Ctrl" button; 4. **Click** "Remove" button;

Types endpoint correlations

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint

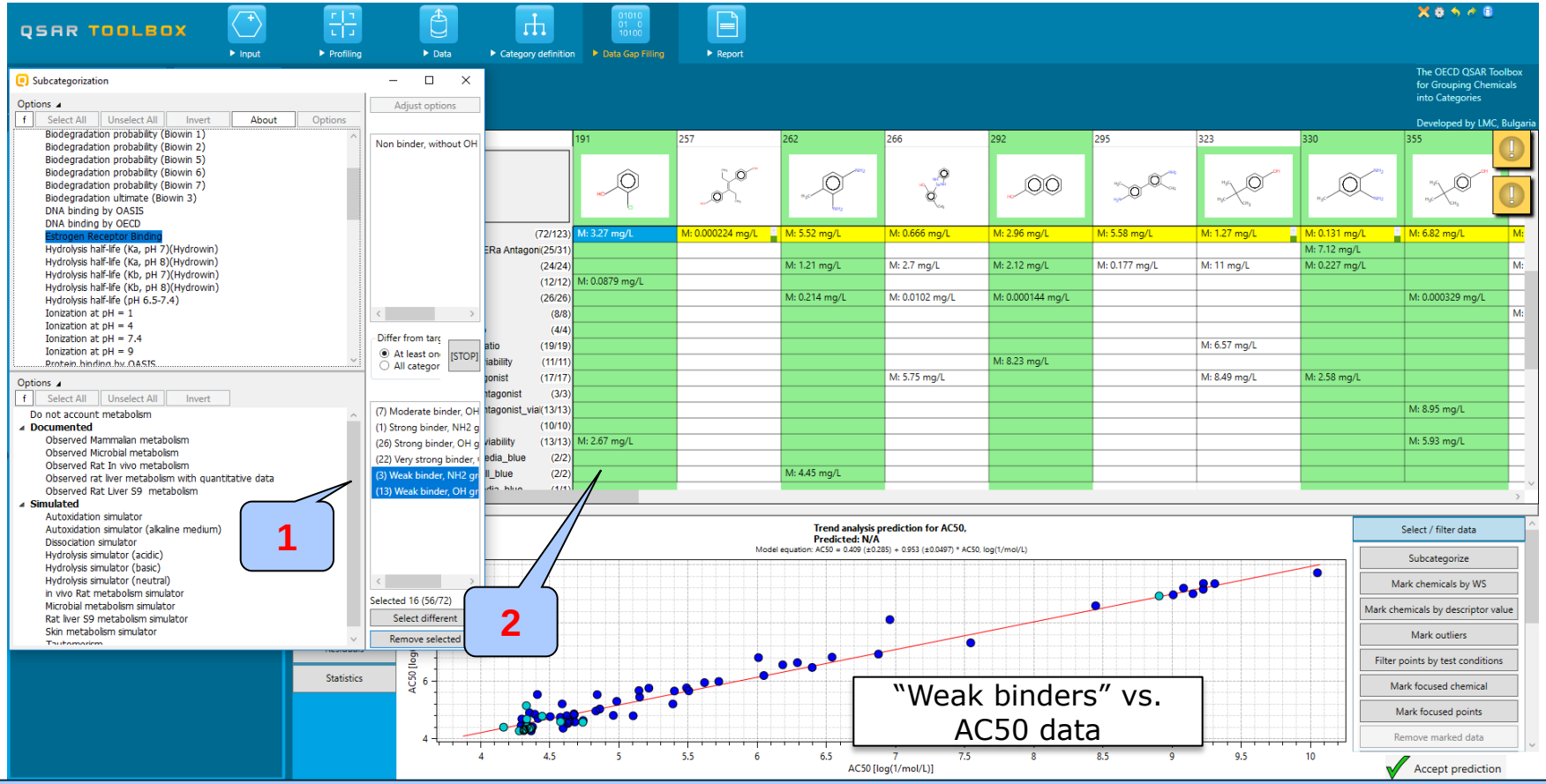


1. Click again on Estrogen receptor binding profiler
2. **Select** "Moderate binder" categories
3. The chemicals corresponding to the selected categories are highlighted in green; 4 and in light blue on the graph

Types endpoint correlations

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint

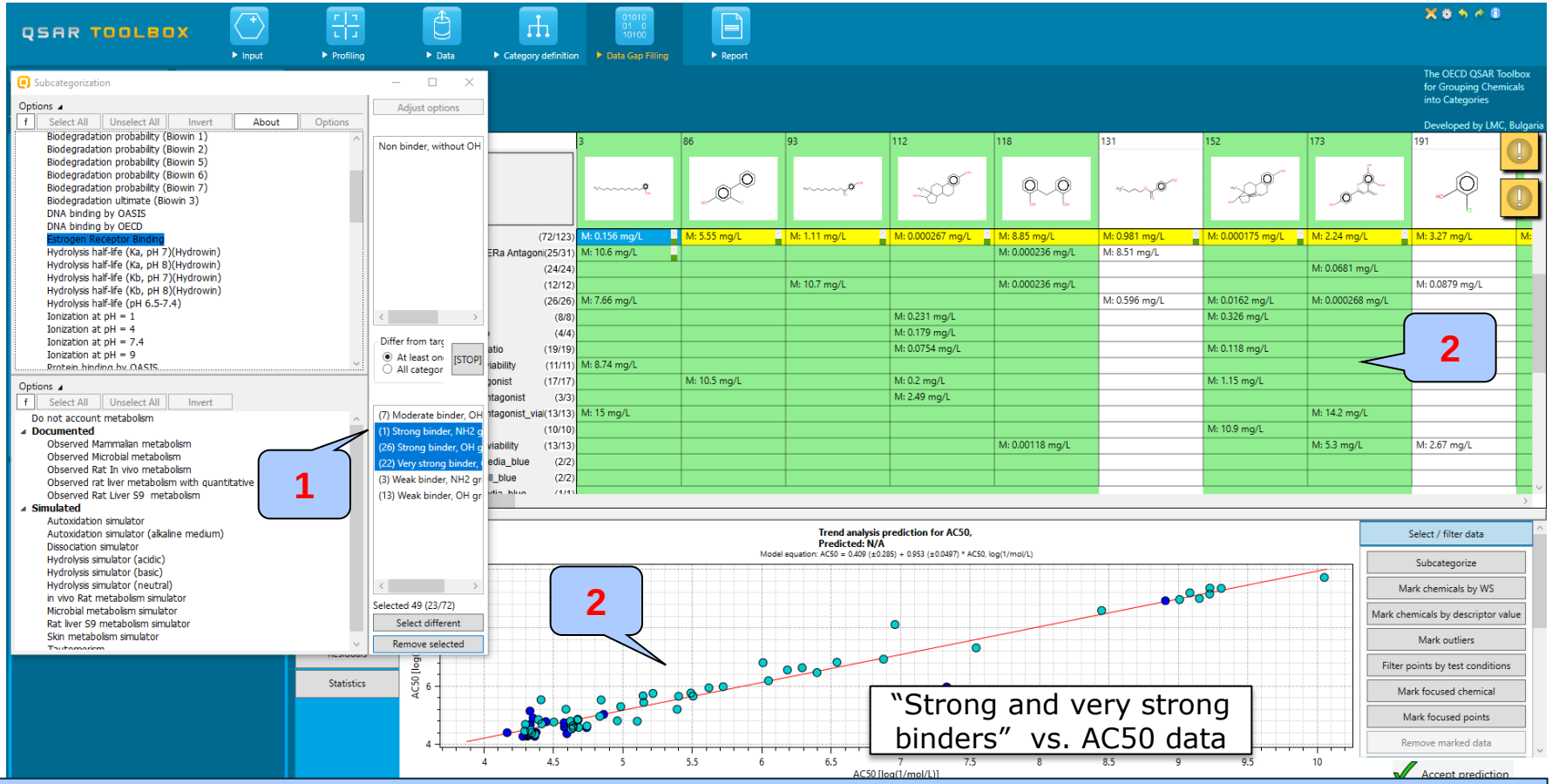


1. **Select** "Weak binder" categories (left mouse click and hold "Ctrl" button);
2. The chemicals corresponding to the selected categories are highlighted in green;

Types endpoint correlations

Continuous vs. continuous

Correlation of active Estrogen receptor categories vs.AC50 endpoint



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