QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

OECD QSAR Toolbox v.4.1

Tutorial illustrating new options of the structure similarity

Outlook

- **Background**
- Aims
- PubChem features
- The exercise
- Workflow

Background

- This presentation is designed to familiarize the Toolbox user with the new structure similarity features;

- Structure similarity options in TB 4.0 have been expanded including PubChem substructure features.

Outlook

- Background
- **Aims**
- PubChem features
- The exercise
- Workflow

Aims

- To show to the Toolbox user how to compare two chemicals with respect to PubChem substructure similarity features;
- To show to the Toolbox user how to compare a chemical with list of chemicals with respect to PubChem substructure similarity features.

Outlook

- Background
- Aims
- **PubChem features**
- The exercise
- Workflow

PubChem features
Overview

The *PubChem* generates a binary substructure fingerprint for each chemical structure. These fingerprints are used by PubChem for similarity neighboring and similarity searching.

A substructure is a fragment of chemical structure. A fingerprint is an ordered list of binary (1/0) bits. Each bit represent a Boolean determination of specific atom or test features used further for similarity neighboring and similarity searching. Seven groups of PubChem features are defined and used:

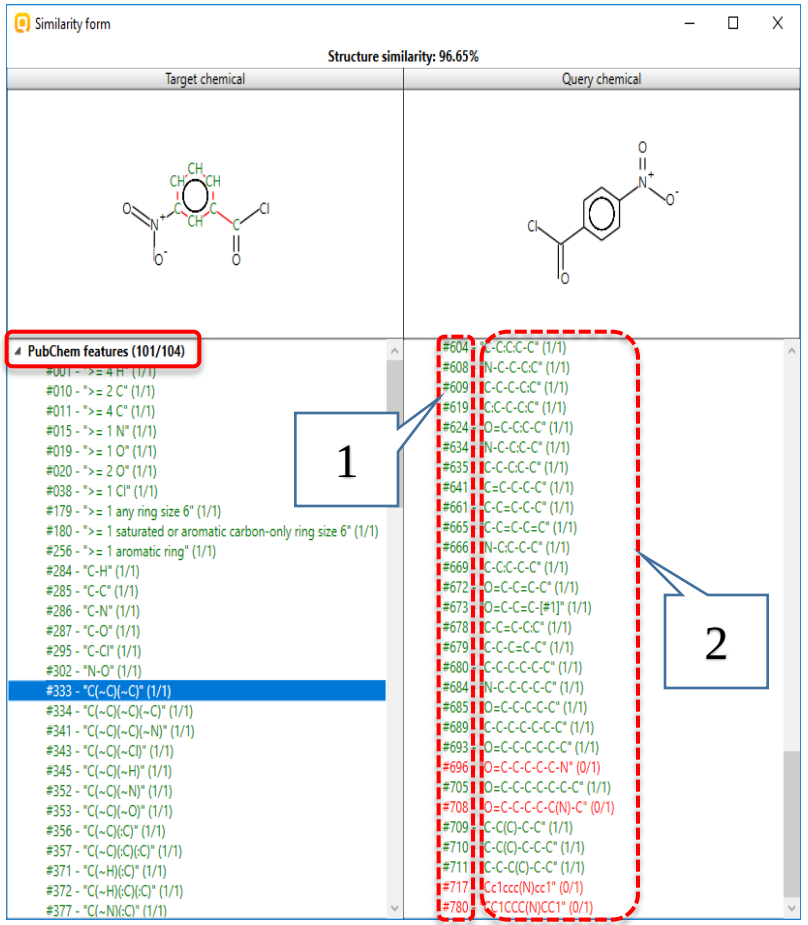
- Hierarchical element counts;
- Rings;
- Simple atom pairs;
- Simple atom nearest neighbors;
- Detailed atom neighbors;
- Simple SMARTS patterns;
- Complex SMARTS patterns.

PubChem features
Overview

Numbers in brackets show how many are the common substructure features between the two compared structures out of all features found in a single chemical.

- Green colored features are common for both structures;
- Red colored features are unique.

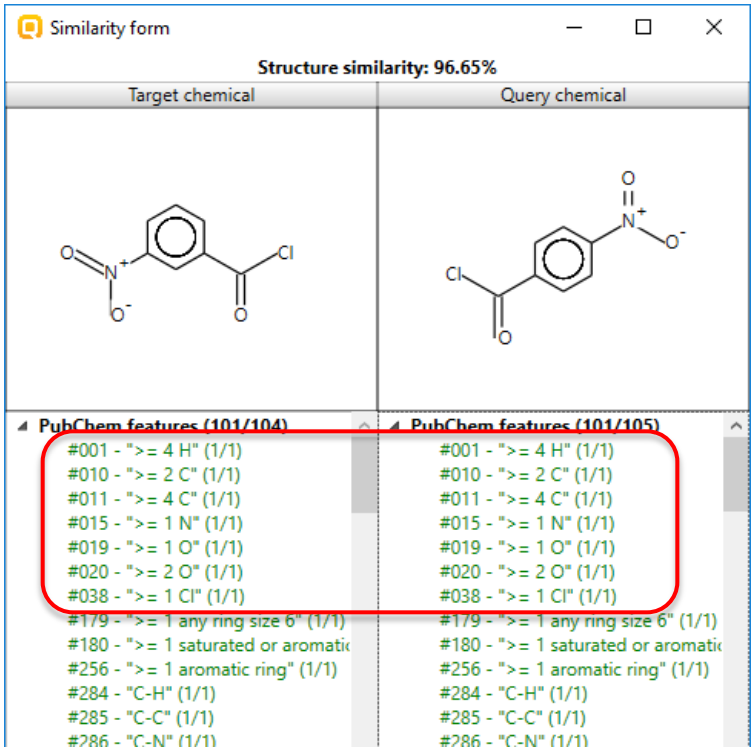
Each of the *PubChem* features has **bit position** (1) which correspond to a **bit substructure** (2).



PubChem features
Hierarchical element counts

- **Hierarchical element counts** - These bits test for the presence or count of individual chemical atoms represented by their atomic symbol.

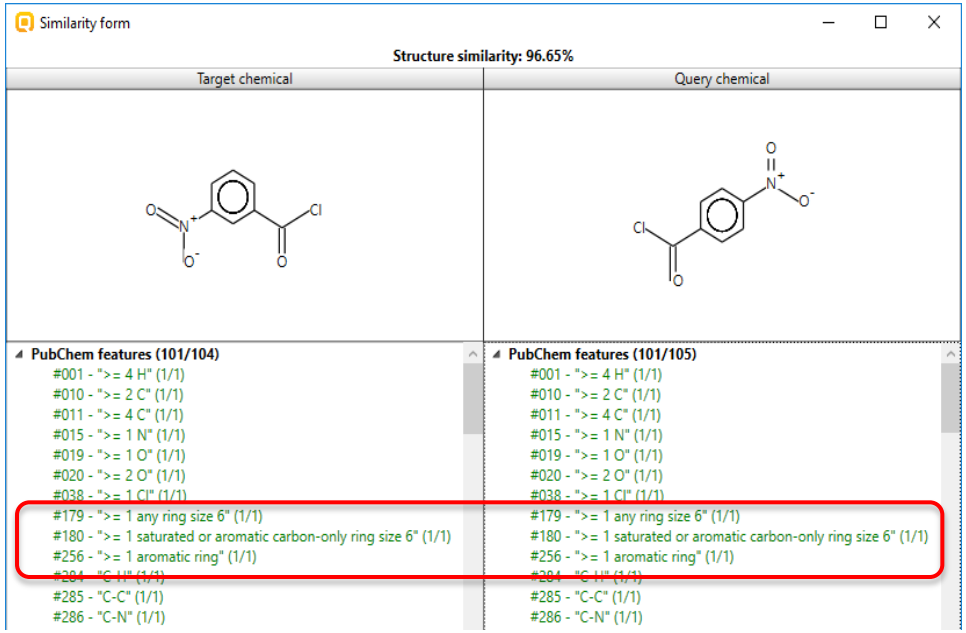
They include bit positions from 001 to 115.



PubChem features
Rings in a canonic ESSSR ring set

- **Rings in a canonic Extended Smallest Set of Smallest Rings (ESSSR) ring set** - These bits test for the presence or count of the described chemical ring system. An ESSSR ring is any ring which does not share three consecutive atoms with any other ring in the chemical structure. For example, naphthalene has three ESSSR rings (two phenyl fragments and the 10-membered envelope), while biphenyl will yield a count of only two ESSSR rings.

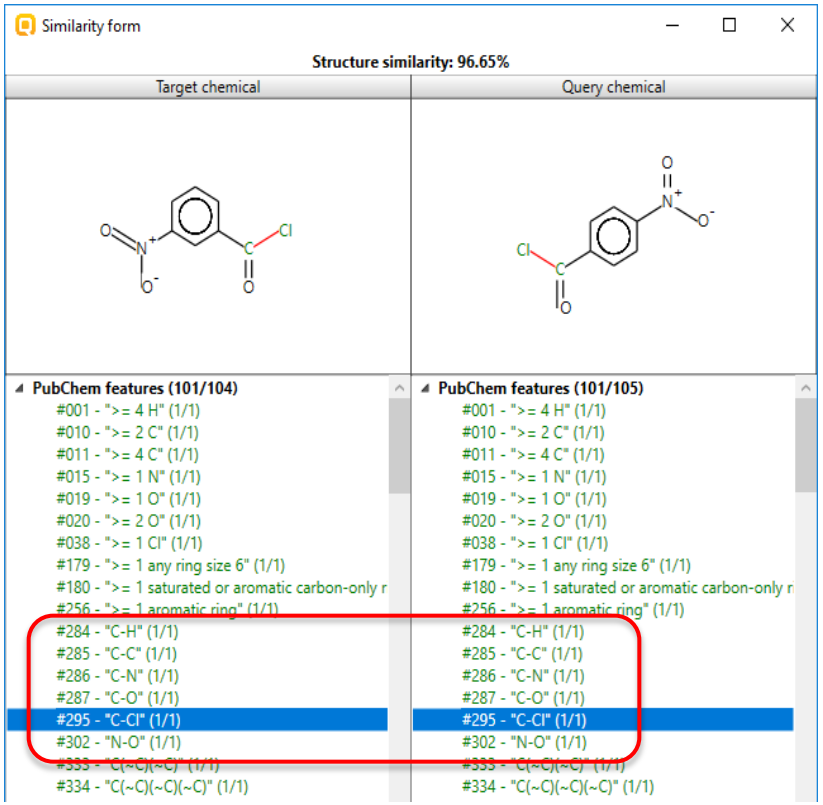
They include bit positions from 116 to 263.



PubChem features
Simple atom pairs

- **Simple atom pairs** – These bits test for the presence of patterns of bonded atom pairs, regardless of bond order or count.

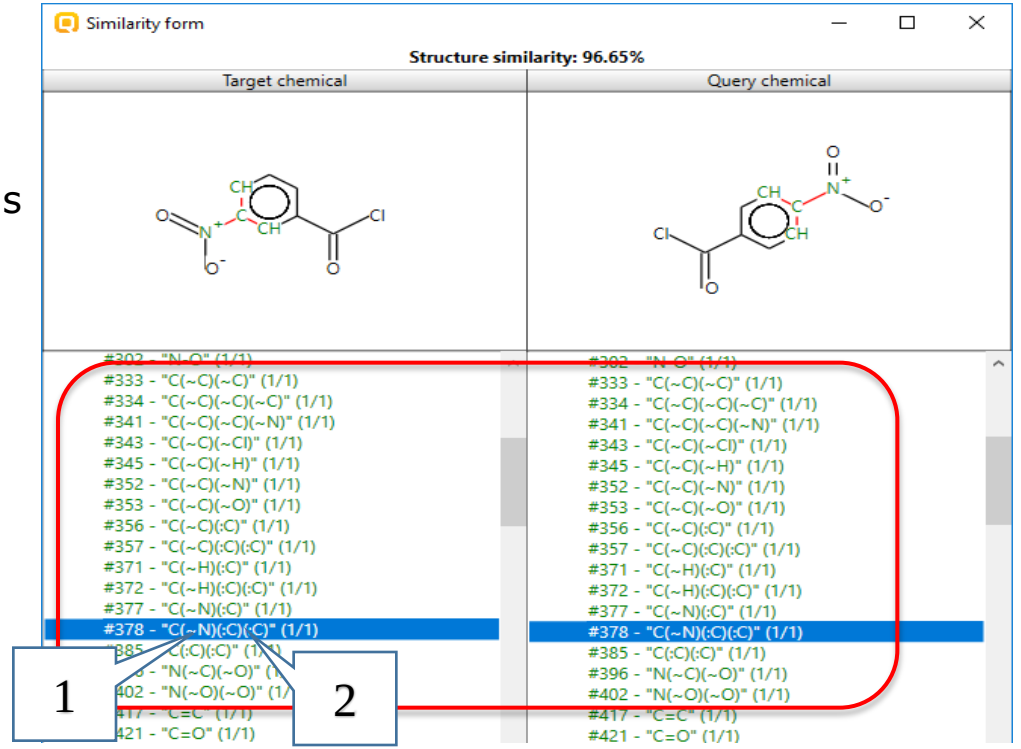
They include bit positions from 264 to 327.



PubChem features
Simple atom nearest neighbors

- **Simple atom nearest neighbors** – These bits test for the presence of atom nearest neighbor patterns, regardless of bond order (denoted by "~") (1) or count, but where bond aromaticity (denoted by ":") (2) is significant.

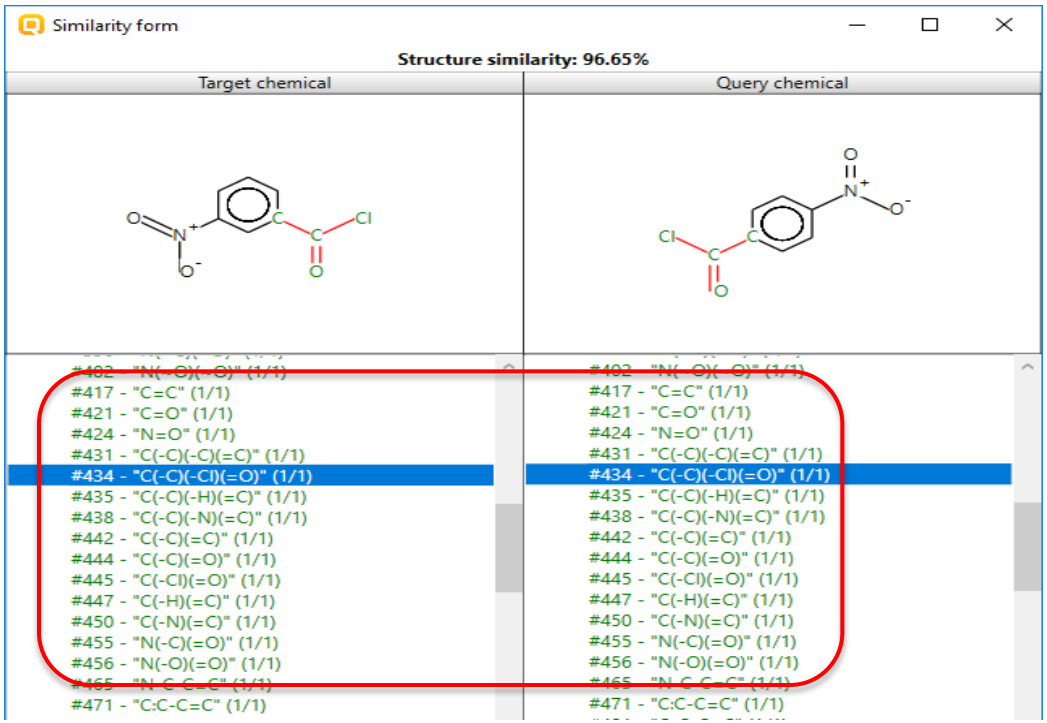
They include bit positions from 328 to 416.



PubChem features
Detailed atom neighborhoods

- **Detailed atom neighborhoods** – These bits test for the presence of detailed atom neighborhood patterns, regardless of count, but where bond orders are specific, bond aromaticity matches both single and double bonds, and where "-", "=", and "#" matches a single bond, double bond, and triple bond order, respectively.

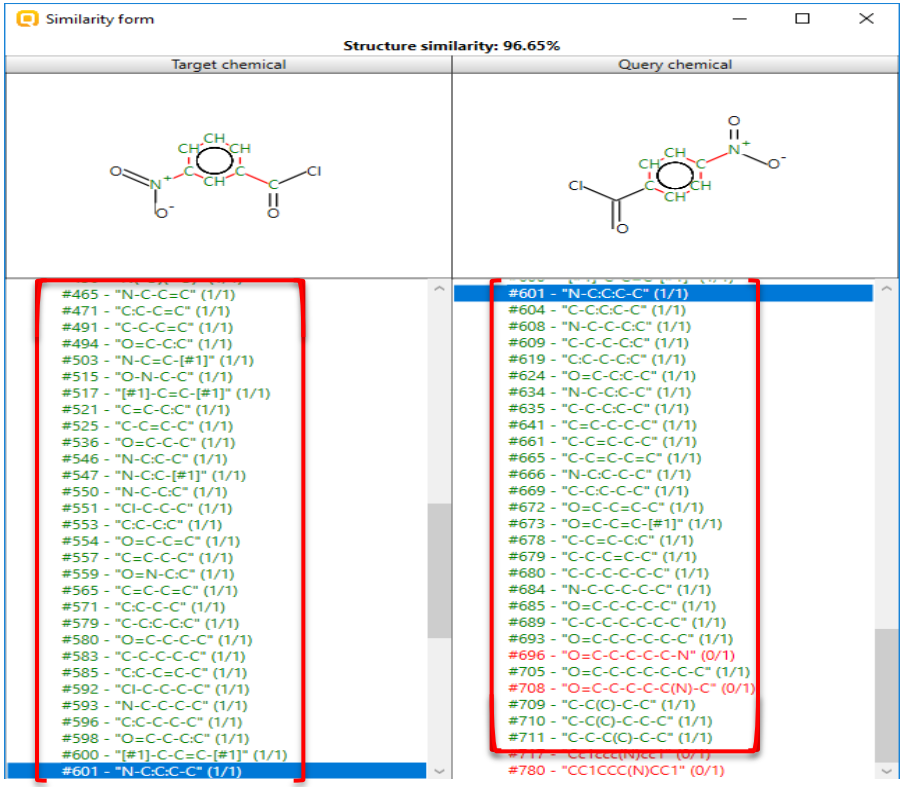
They include bit positions from 417 to 460.



PubChem features
Simple SMARTS patterns

- **Simple SMARTS patterns** – These bits test for the presence of simple SMARTS patterns, regardless of count, but where bond orders are specific and bond aromaticity matches both single and double bonds.

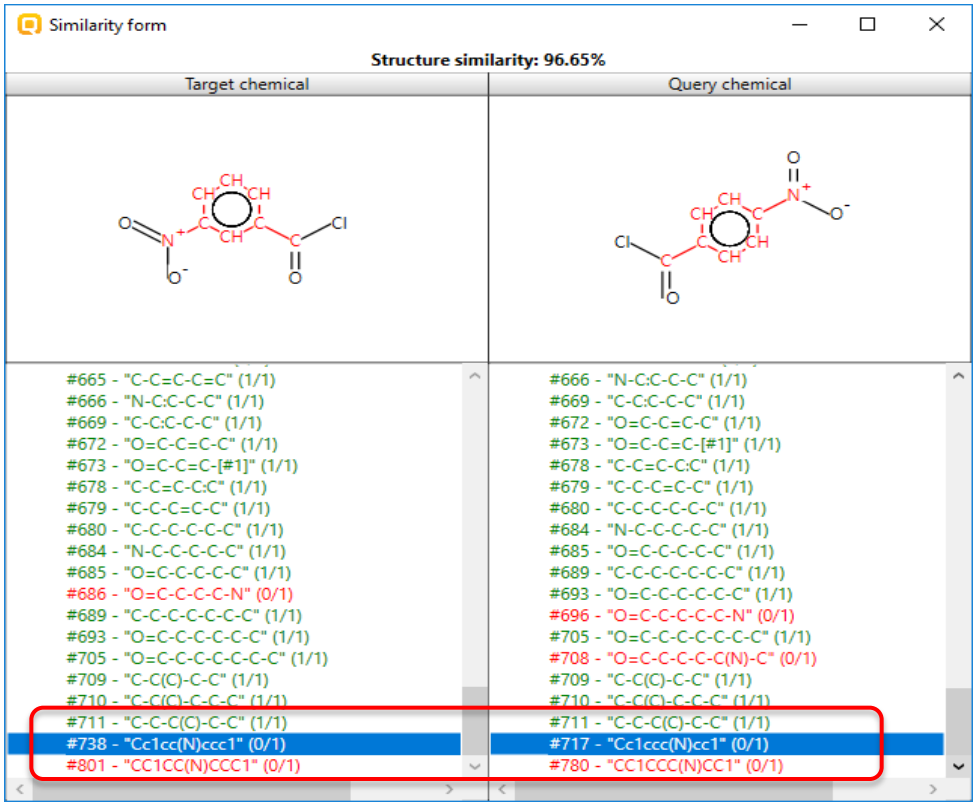
They include bit positions from 461 to 713.



PubChem features
Complex SMARTS patterns

- **Complex SMARTS patterns** – These bits test for the presence of complex SMARTS patterns, regardless of count, but where bond orders and bond aromaticity are specific.

They include bit positions from 714 to 881.



Outlook

- Background
- Aims
- PubChem features
- **The exercise**
- Workflow

The Exercise

- In this exercise we will compare:

  1. two chemicals with respect to *PubChem* substructure similarity features (we will use *m-Chloroaniline* and *benzoic acid*);

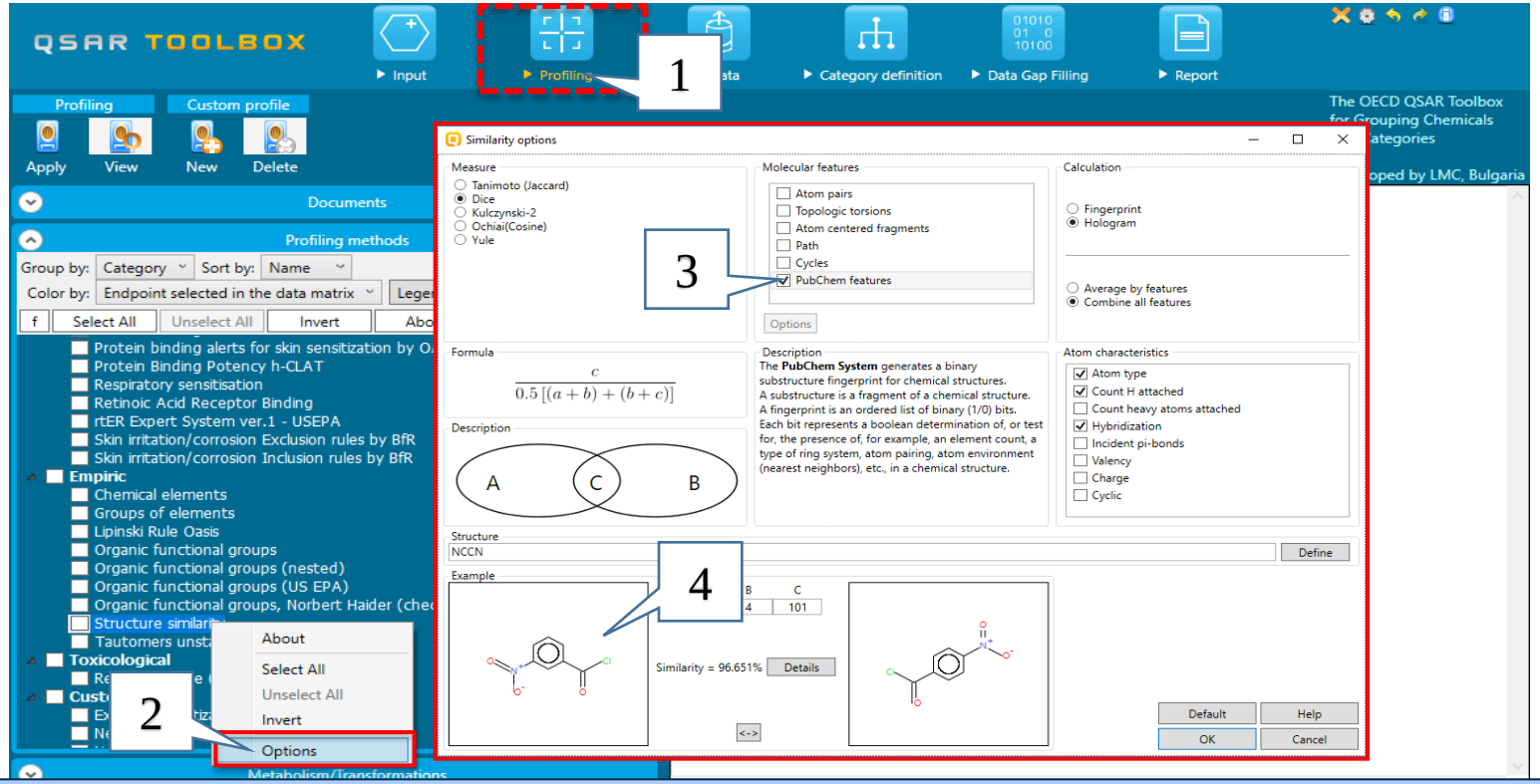
  2. One chemical with a list of chemicals with respect to *PubChem* substructure similarity features (we will use *m-Chloroaniline* and Skin sensitization ECETOC database).

Outlook

- Background
- Aims
- PubChem features
- The exercise
- **Workflow**
  - **Substructure similarity between two chemicals**

Workflow

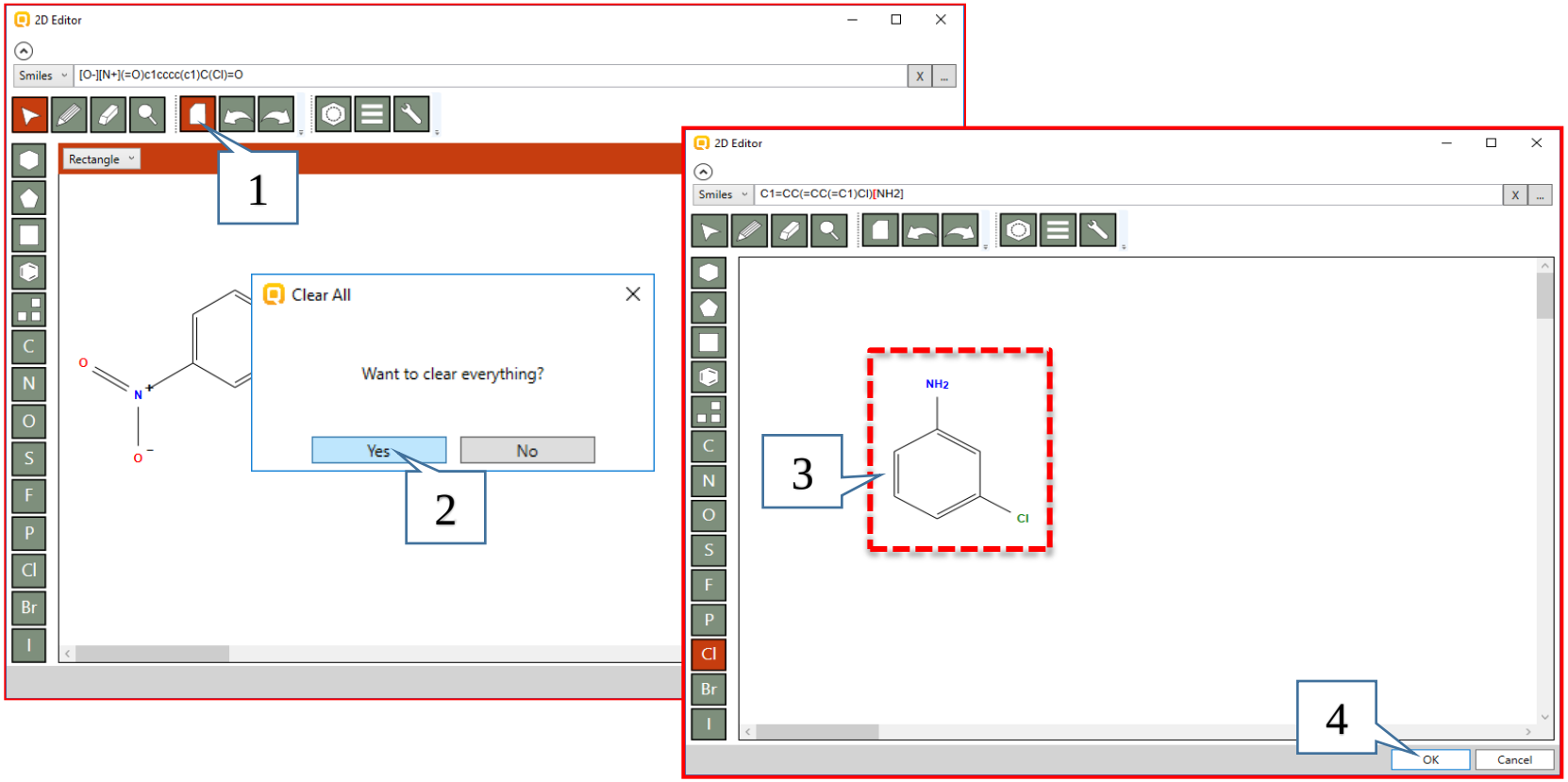
Substructure similarity between two chemicals



1.Go to *Profiling* module; 2. Right click over the *Structure similarity* profiler and select **Options**; 3.Uncheck all molecular features and select only *PubChem* features. The additional similarity options (e.g. *Calculation* and *Atom characteristics*) do not have influence to the *PubChem* features.; 4. Double click on the structure in left which will be our target. *2D editor* window appears.

Workflow

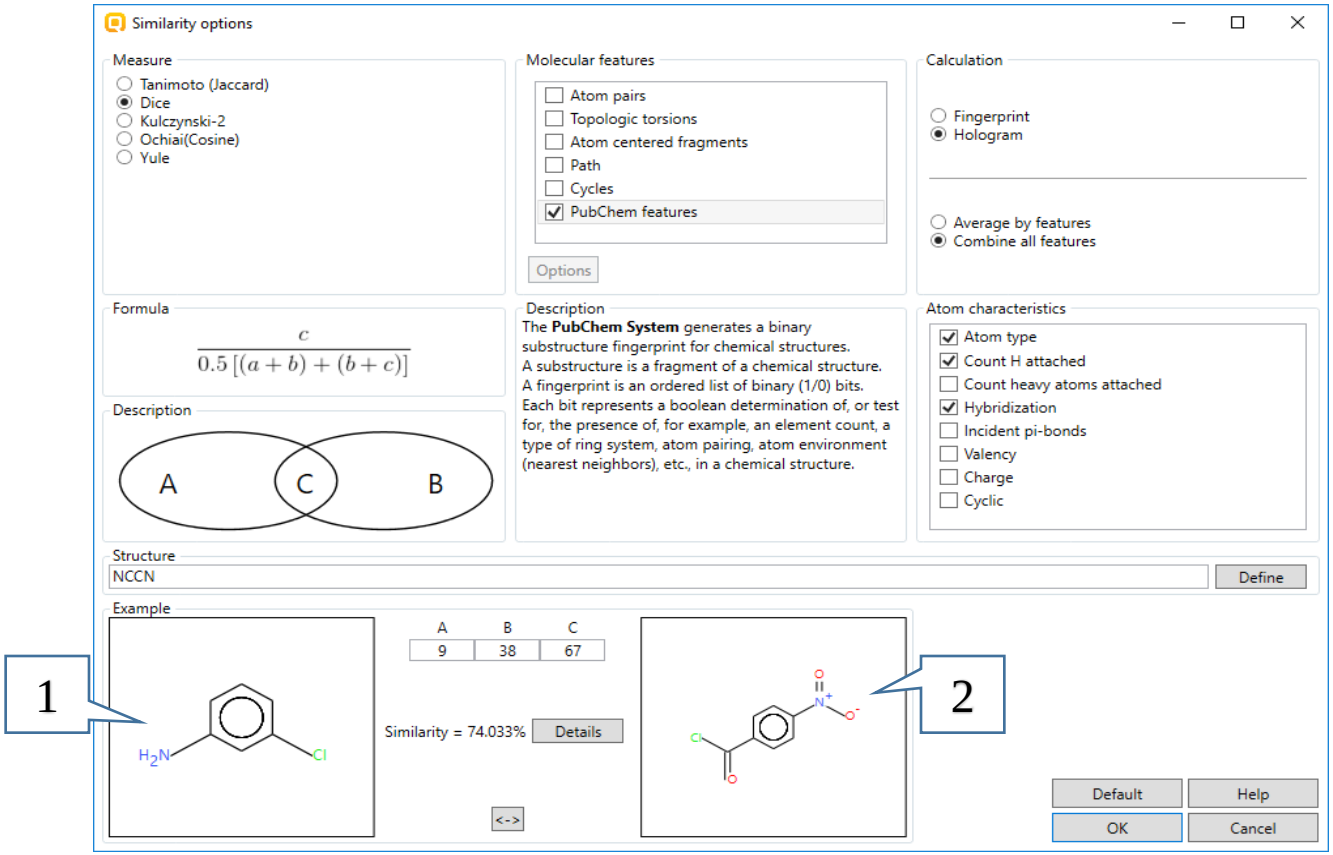
Substructure similarity between two chemicals



1.Click on the ***Blank page*** button; 2. Confirm that you want to clear everything by click on **Yes**; 3. Draw *m-chloroaniline* structure; 4. Click on **OK**

Workflow

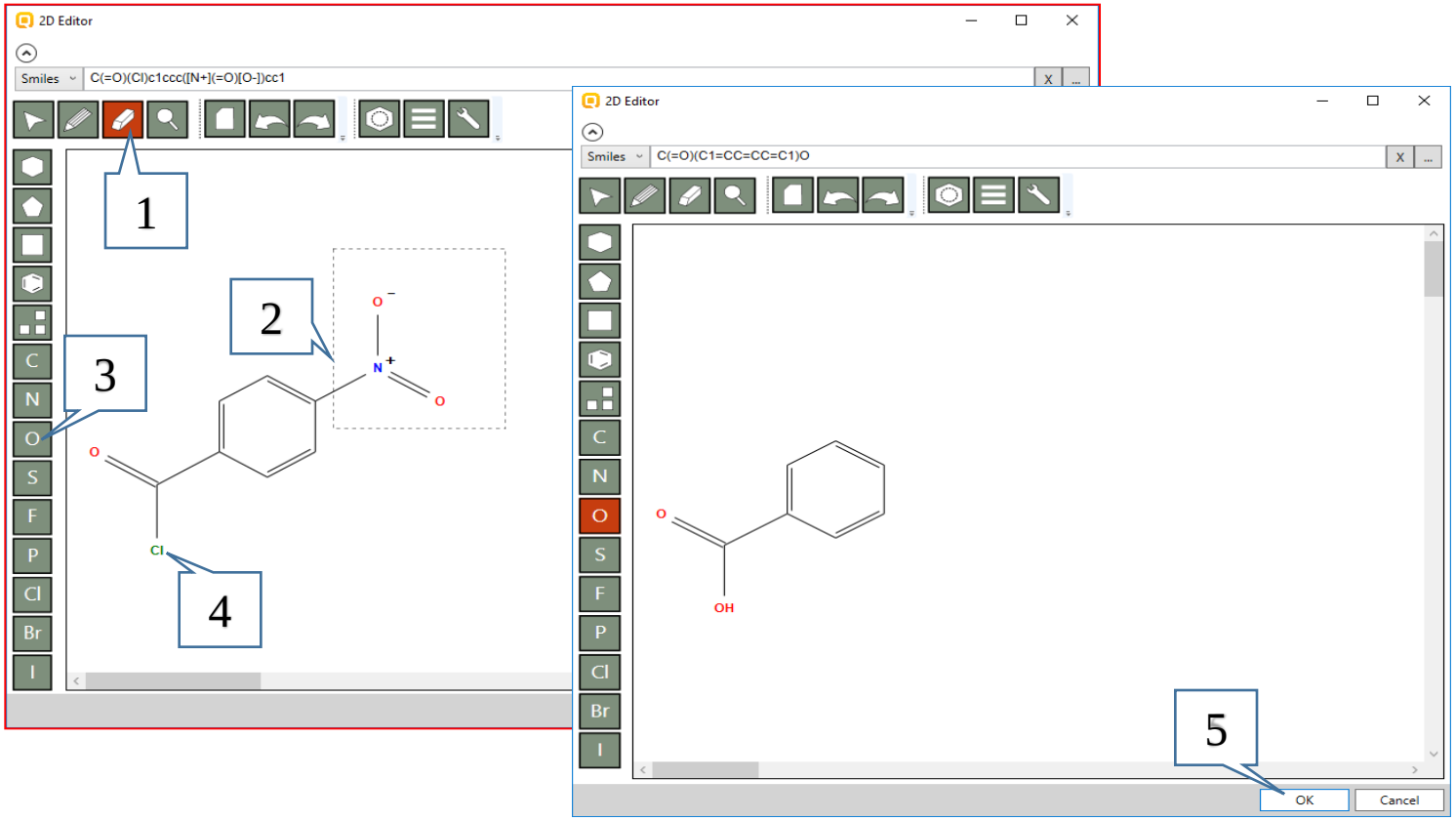
Substructure similarity between two chemicals



The structure which you have drawn appears (1). Double click on the structure in right (2) to draw the second chemical.

Workflow

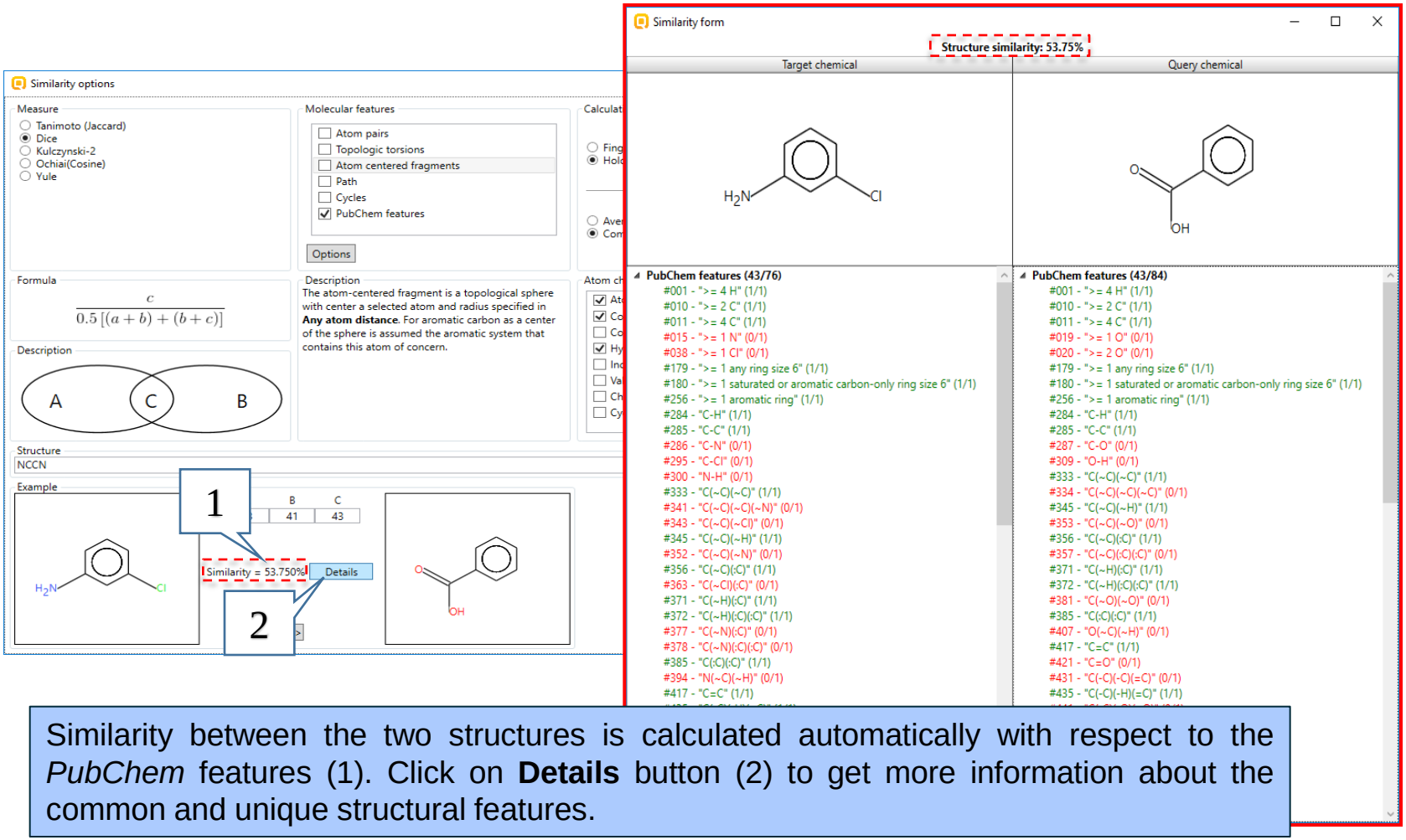
Substructure similarity between two chemicals



Click on the **Eraser** button (1) and remove the nitro group (2). Select the oxygen symbol (3) and click over the chorine atom (4). Now you are ready with drawing of benzoic acid and have to click on **OK** (5)

Workflow

Substructure similarity between two chemicals



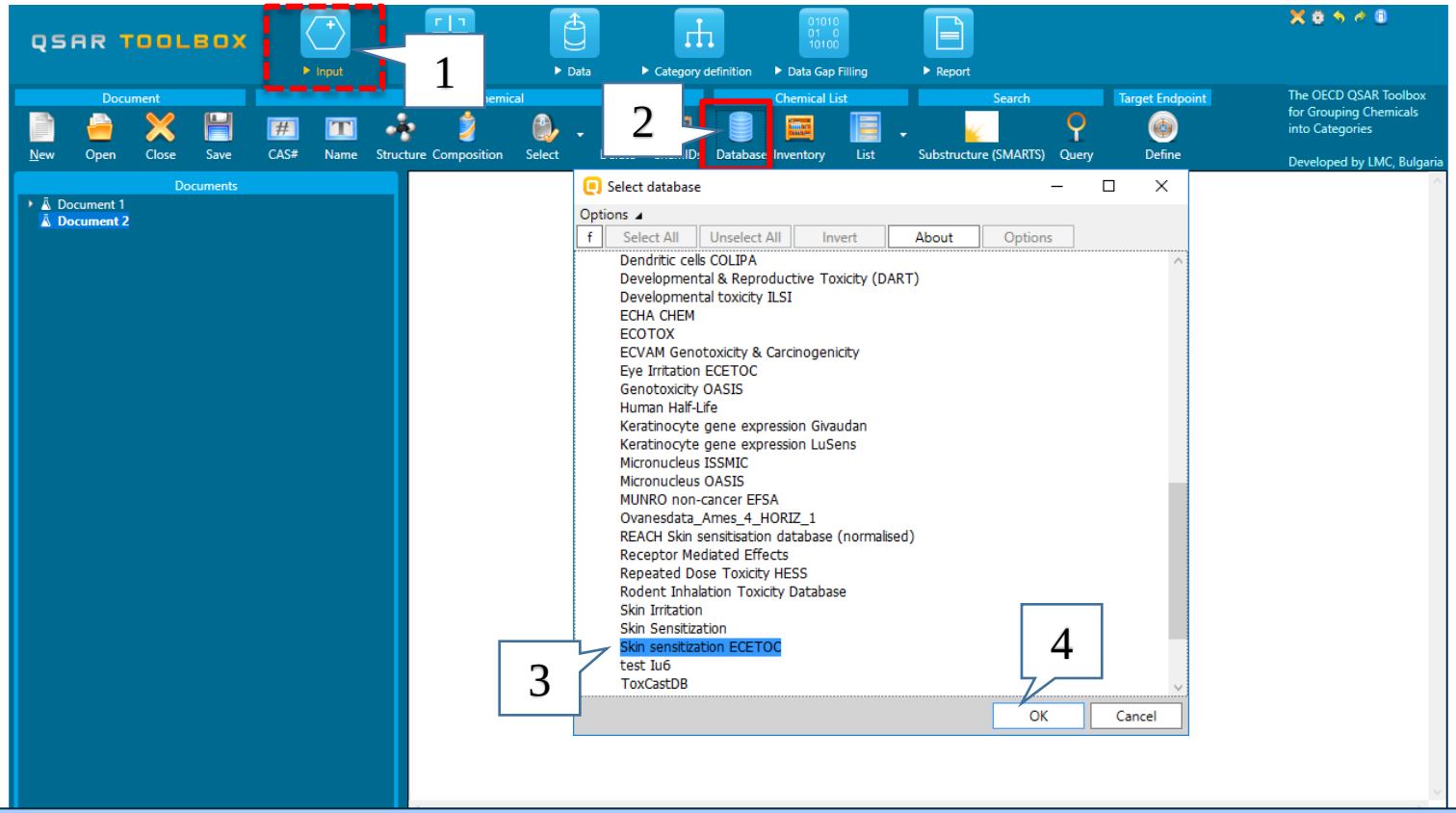
Similarity between the two structures is calculated automatically with respect to the *PubChem* features (1). Click on **Details** button (2) to get more information about the common and unique structural features.

Outlook

- Background
- Aims
- PubChem features
- The exercise
- **Workflow**
  - Substructure similarity between of two chemicals
  - **Substructure similarity between single chemical and chemical list**

Workflow

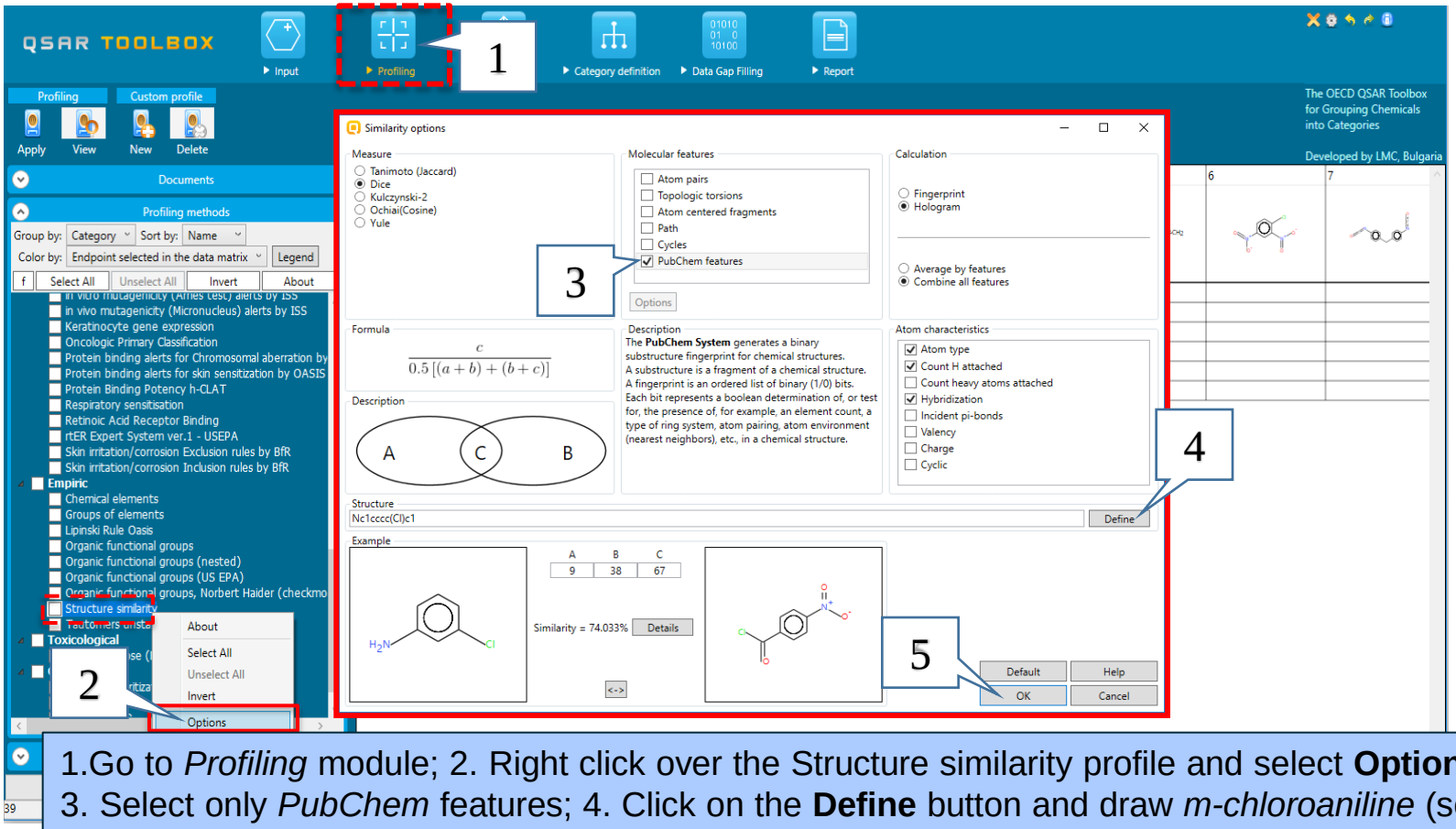
Substructure similarity between single chemical and chemical list



1.Go to *Input* module; 2. Click on *Chemical List* > **Database**; 3. Select **Skin sensitization ECETOC** database; 4. Click on **OK**.

Workflow

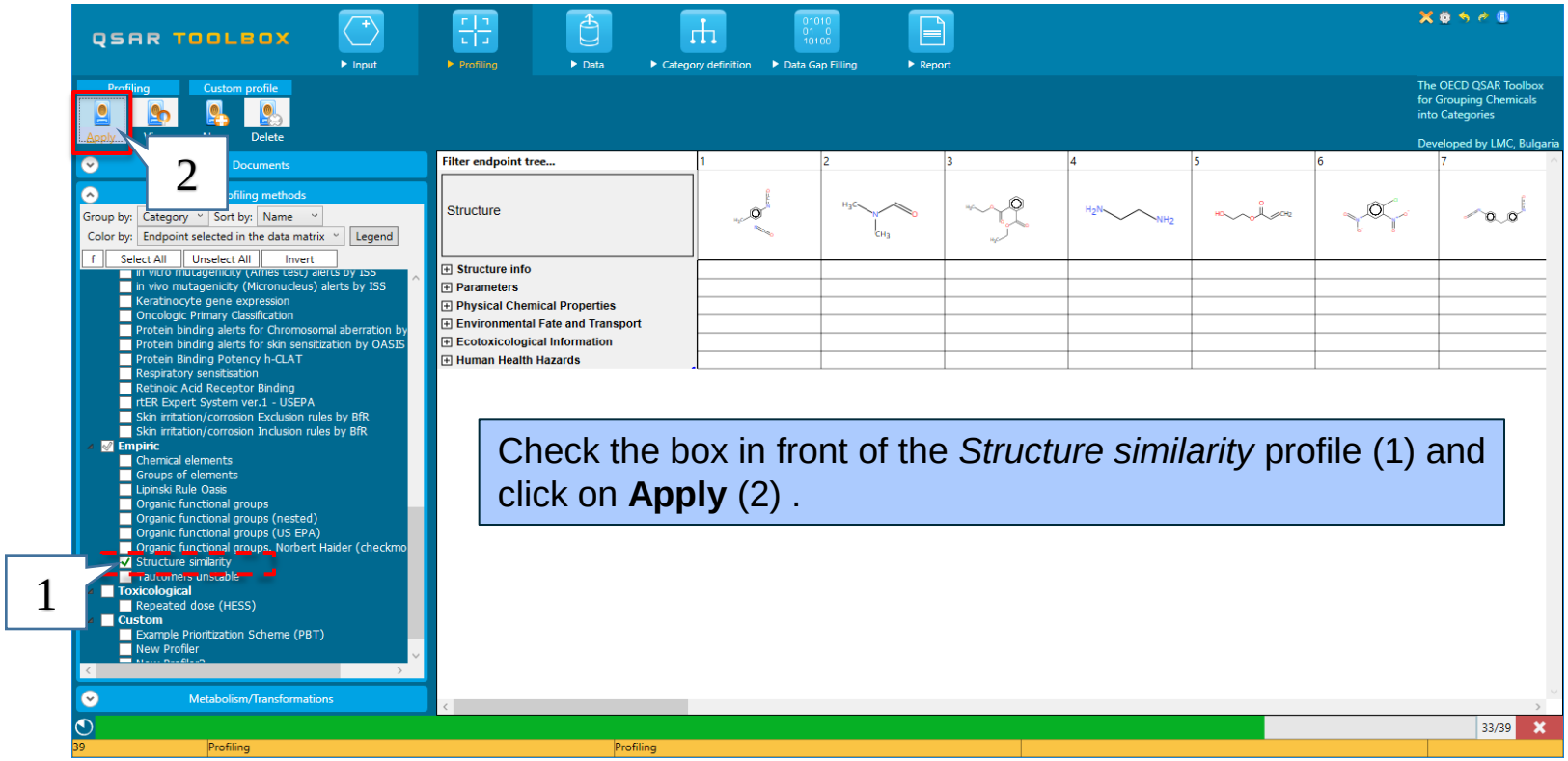
Substructure similarity between single chemical and chemical list



1.Go to *Profiling* module; 2. Right click over the Structure similarity profile and select **Options**; 3. Select only *PubChem* features; 4. Click on the **Define** button and draw *m-chloroaniline* (see slide 20); 5. Click on **OK**.

Workflow

Substructure similarity between single chemical and chemical list

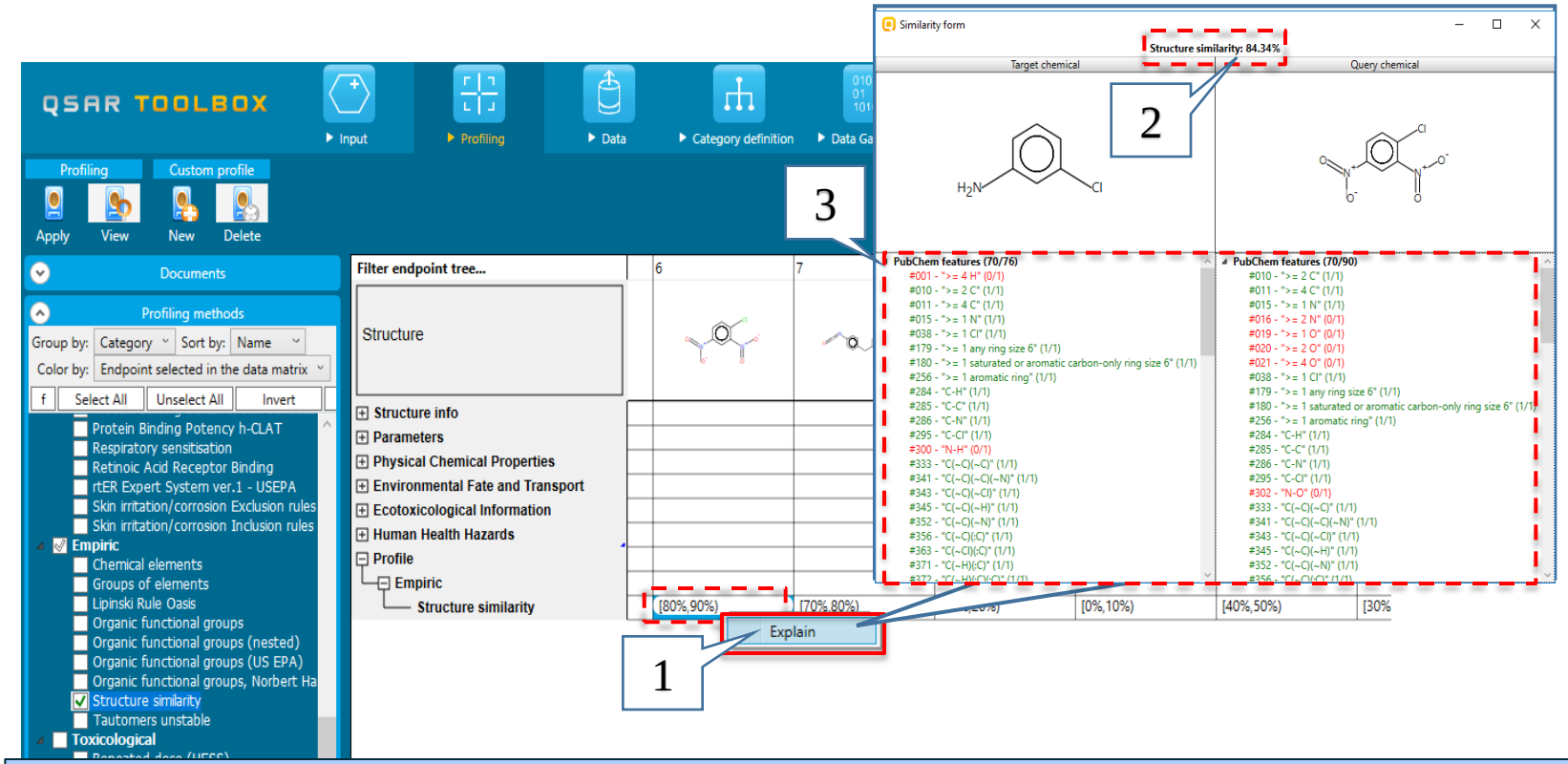


Check the box in front of the *Structure similarity* profile (1) and click on **Apply** (2) .

Keep in mind that the chemical list which you are insert have no a target structure. Otherwise the structure similarity will be calculated based on the target structure in the data matrix.

Workflow

Substructure similarity between single chemical and chemical list



You can see more details about the structure similarity by right click over the result and selection of **Explain** (1). In the *Similarity form* you can see the exact percentage of structure similarity (2) as well as all similar and different features (3) between the two molecules.

Congratulation

- You have now been familiarized with *PubChem* substructure similarity features;
- You have compared: 1) two chemicals and 2) single chemical and chemical list with respect to PubChem substructure similarity;
- Note proficiency comes with practice.