QSAR TOOLBOX

The OECD QSAR Toolbox for Grouping Chemicals into Categories

# OECD QSAR Toolbox

## Tutorial on SMARTS structures search

# Outlook

- **Background**
- Aims
- Overview of SMARTS language
- SMARTS Editor window
- The exercise
- Workflow process

# Background

- This is a step-by-step presentation designed to take the Toolbox user through the SMARTS Editor functionalities for substructure search;

- The SMARTS substructure search is useful when the user does not have a specific chemical in mind but rather wants to find a group of chemicals containing the same structural characteristics.

# Outlook

- Background
- **Aims**
- Overview of SMARTS language
- SMARTS Editor window
- The exercise
- Workflow process

# Aims

- To introduce to the Toolbox user the SMARTS Editor;
- To familiarize the user with the new query fragments for substructure search.

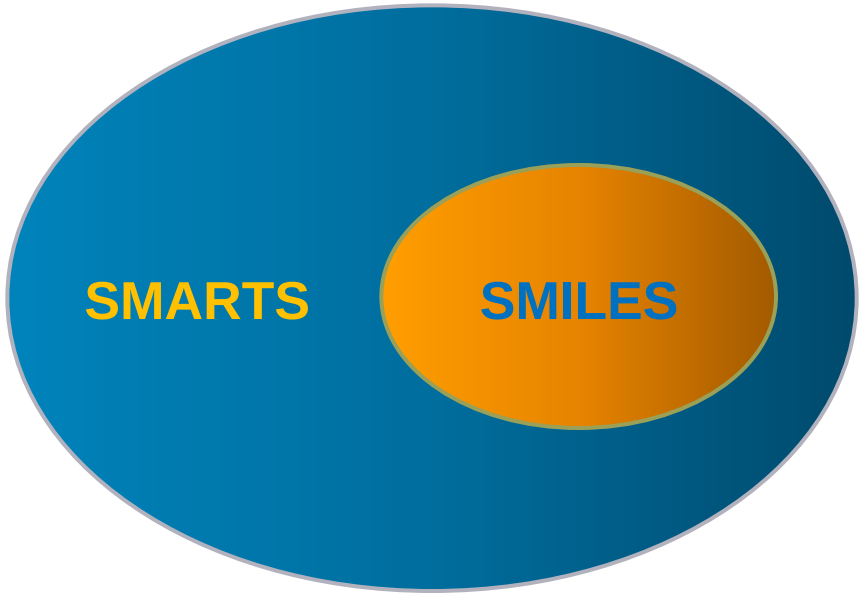
# Outlook

- Background
- Aims
- **Overview of SMARTS language**
- SMARTS Editor window
- The exercise
- Workflow process

# Overview of SMARTS language
## SMARTS vs. SMILES

- SMARTS (***SM****iles* ***AR****bitrary* ***T****arget* ***S****pecification*) - A language for describing molecular patterns used for fragment matching;

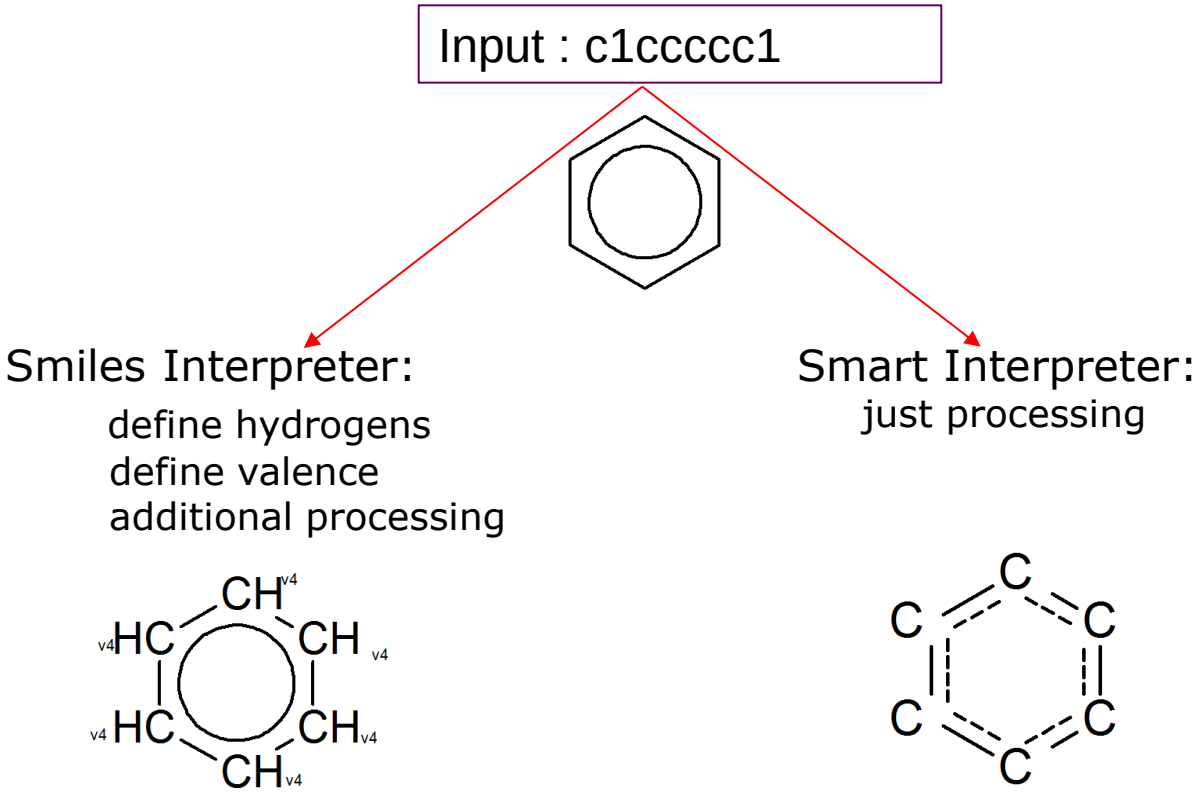
- SMARTS allows you to specify substructures using rules that are straightforward extensions of SMILES*




*http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html

# Overview of SMARTS language
## SMARTS vs. SMILES

Any valid SMILES is valid SMARTS.



# Outlook

- Background
- Aims
- Overview of SMARTS language
- **SMARTS Editor window**
- The exercise
- Workflow process

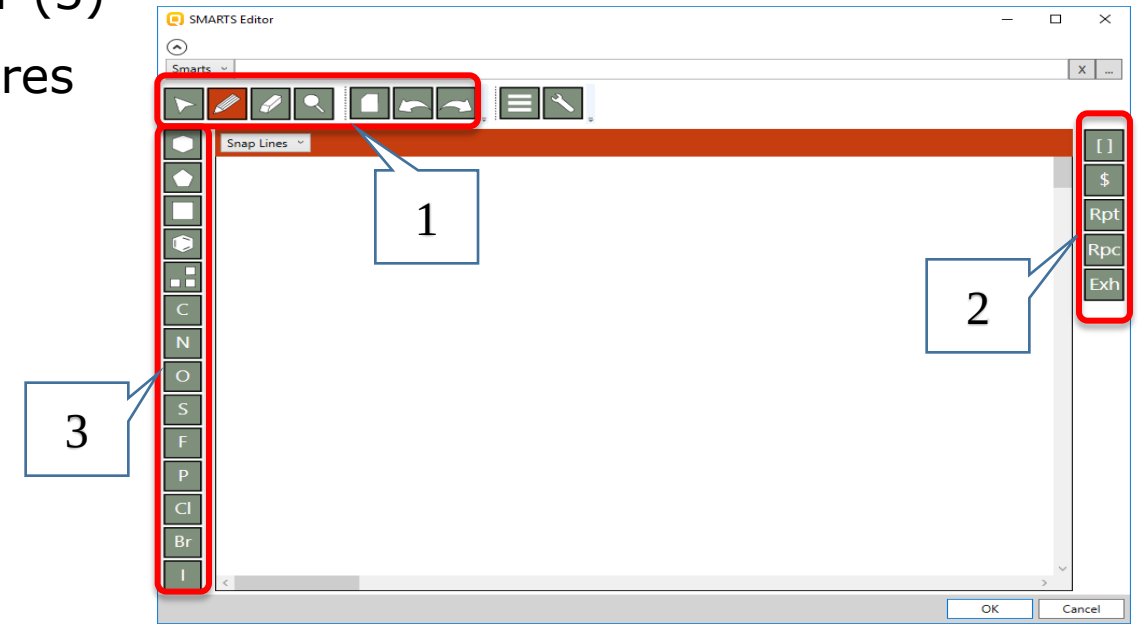
# SMARTS Editor
## Overview

- The new editor is developed to be compatible with the SMARTS language;
- Editor fragments of the previous Toolbox versions are updated and extended to facilitate the users in writing SMARTS;
- The SMARTS Editor can be used for substructure search as well as for creating of custom profiles.

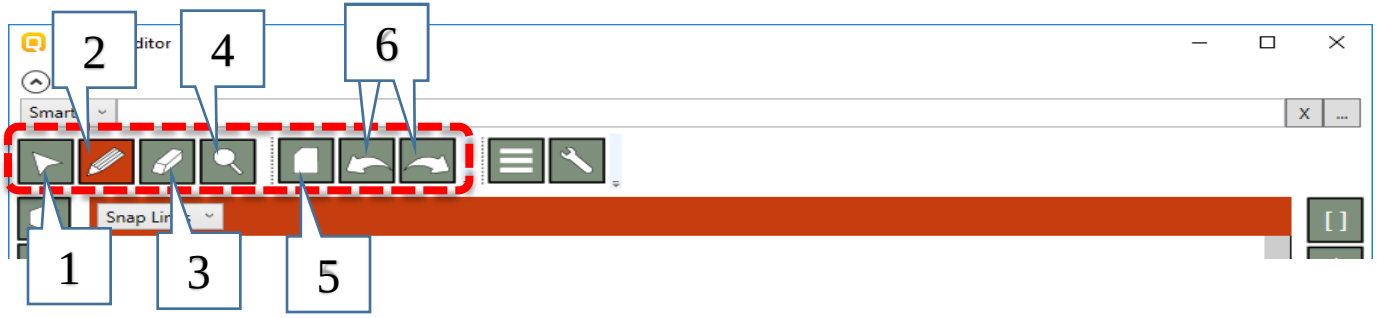
# SMARTS Editor window

The SMARTS Editor window includes:

- Action toolbar (1)
- Fragments toolbar (2)
- Auxiliary toolbar (3)
- Additional features



# SMARTS Editor window
## Action toolbar



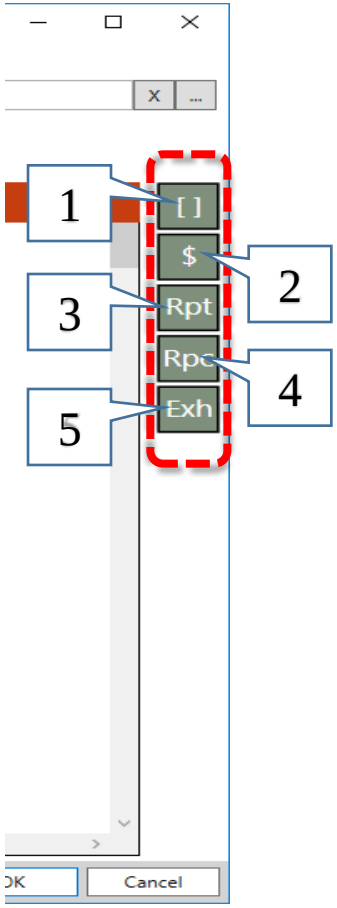

The **Action toolbar** includes:

- **Selection tool** (1) – allows to define atom characteristics, to mark or rotate a structure, to start new activity, etc;
- **Pencil** button (2) – allows drawing of hydrocarbon structures (scaffolds);
- **Eraser** button (3) – removes selected bonds or other parts of a drawn structure;
- **Magnifying glass** button (4) – allows to zoom in (zoom out) with the mouse scroller;
- **Blank page** (5) – clears all in the drawing pane;
- **Arrows** buttons (6) – allow to undo any action or redo right after undo.

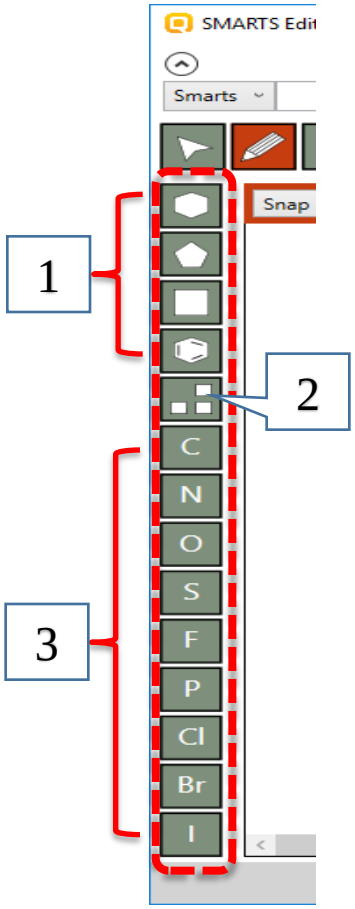
# SMARTS Editor window
## Fragments toolbar

The **Fragments toolbar** includes:

- "**[]**" button (enumeration) (1): allows enumeration of atoms which could be bonded to a given atom in the molecule;
- "**$**" button (recursive SMARTS expression) (2): allows to describe the surrounding of an atom. For example it could be used to describe N -atom, which is not part of nitro group;
- "**Rpt**" button (repeat fragment) (3): allows to specify how many times an atom or a fragment could be repeated in the molecule;
- "**Rpc**" button (replace fragment) (4): gives an option to replace a given atom in the molecule with different chemical elements;
- "**Exh**" buttons (exhaust fragment) (5): allows only those atoms or fragments listed in the exhaust fragment to be bonded to a given atom.

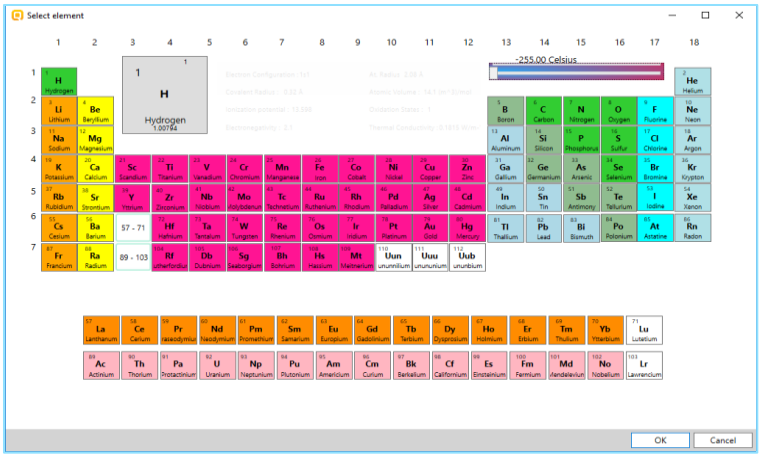


# SMARTS Editor window
## Auxiliary toolbar




The **Auxiliary toolbar** includes:

1. Some of the frequently used hydrocarbon structures, which can be used as scaffolds;
2. Periodic table – allows to select any chemical element. The new developed periodic table is more colorful and informative. It also allows the user to see the physical state of the elements at different temperature.




3. Some of the frequently used atoms.

# SMARTS Editor window
## Additional features

Some additional features are:

- **Buttons**

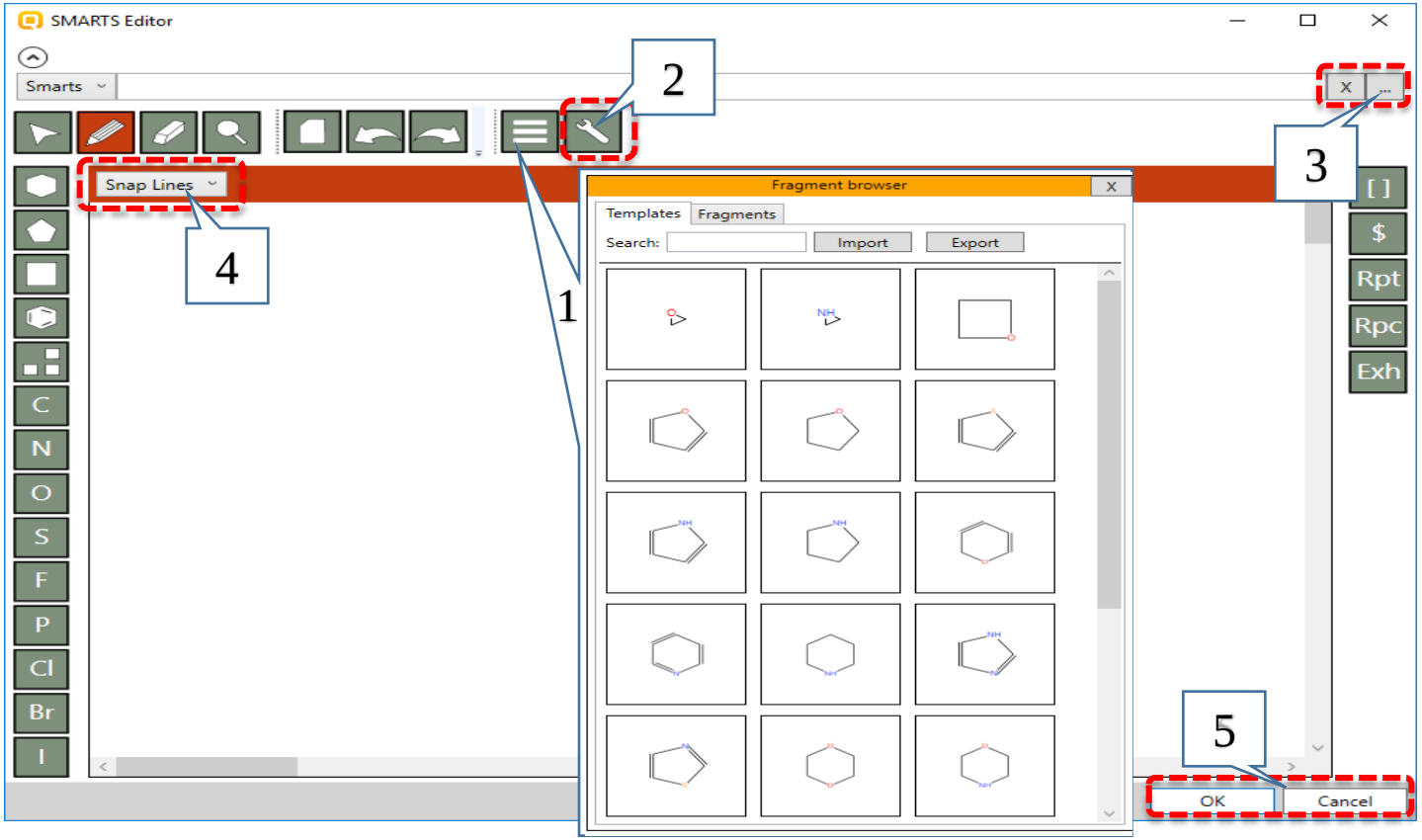
  1. Templates button - opens fragment browser, which contains list of templates of cyclic structures and list of fragments. The user can also add their structures and to make own fragment library;
  2. Wrench button – corrects all angles in the structure;
  3. Buttons allowing to delete or insert SMARTS;
  4. Red bar – includes some different additional secondary options;
  5. Buttons allowing to execute or cancel the search;

- **Object explorer**

***See on the next two slides.***

# SMARTS Editor window
## Additional features

- **Buttons**

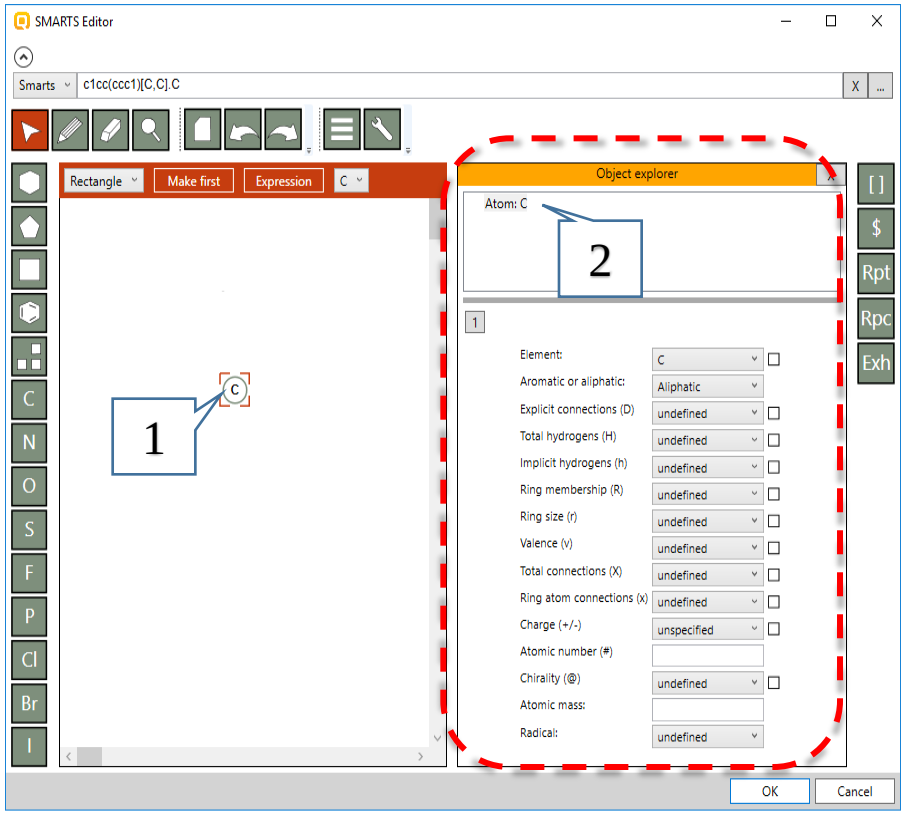


# SMARTS Editor window
## Additional features

- **Object explorer**

1. The *Object explorer* shows what information stays behind an atom or a fragment by click on the corresponding object in the drawing pane;

2. In the *Object explorer* the user can specify details for the selected object.



# Outlook

- Background
- Aims
- Overview of SMARTS language
- SMARTS Editor window
- **The exercise**
- Workflow process

# The exercise

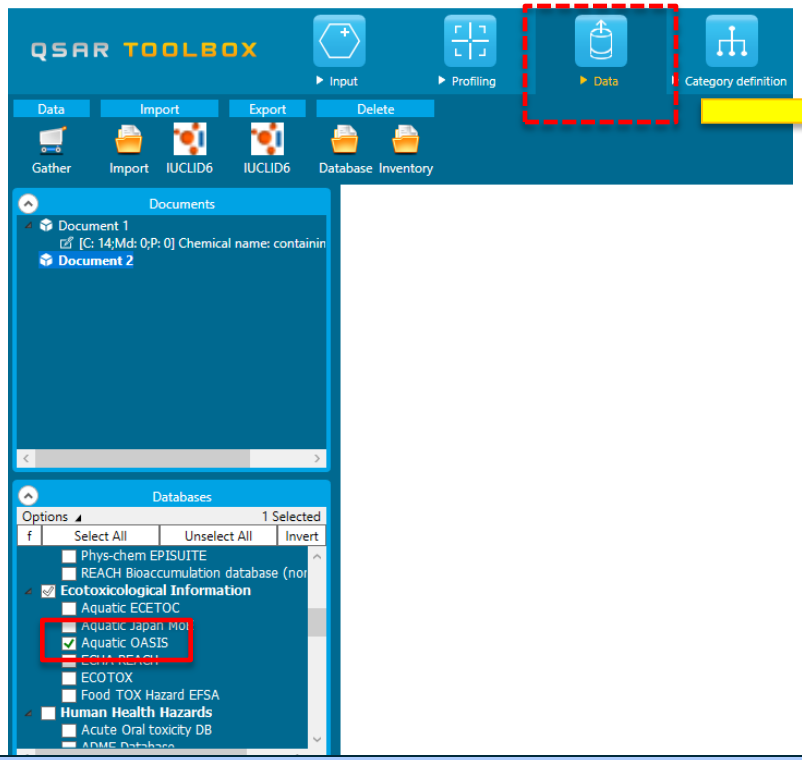
- In this exercise we will demonstrate how to construct different queries for structure searching;
- We will use each of the new fragments for the current purposes;
- *Aquatic OASIS* database will be used.

# Outlook

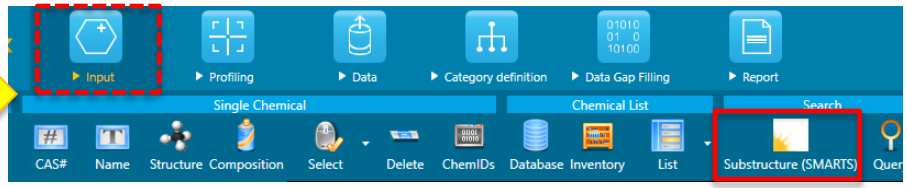
- Background
- Aims
- Overview of SMARTS language
- SMARTS Editor window
- The exercise
- **Workflow process**

# Workflow process
## Overview

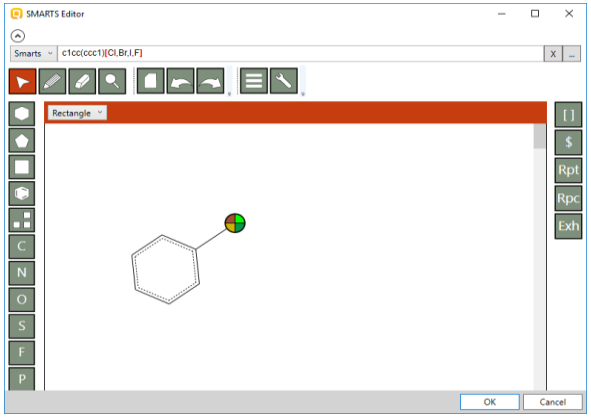
**Step 1**: Select database(s)/inventory(ies)




**Step 2**: Click on ***Substructure (SMARTS)***



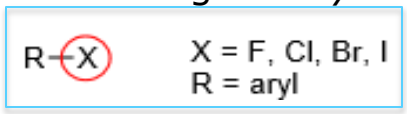

**Step 3**: Specify criteria for searching structures



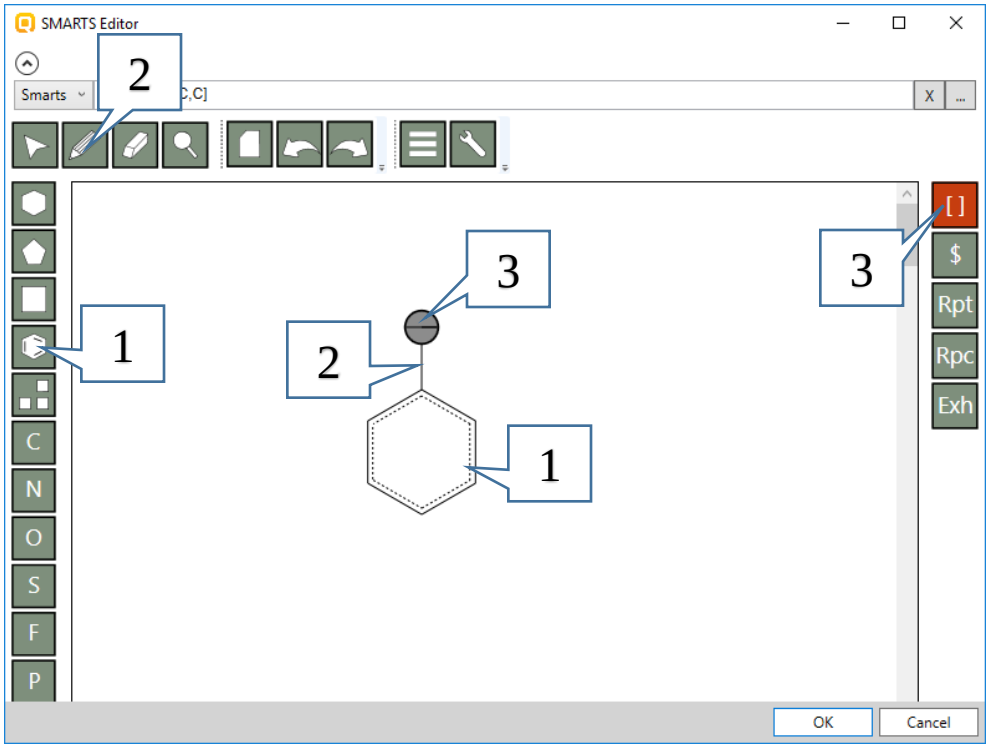

We will use **Aquatic OASIS** database for all examples. Searching criteria will be different only.

# Workflow process
## Structures search

**Example 1:** Searching for *Aryl halides*

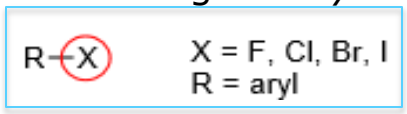



1. Click on the **benzene** template and then click in the drawing pane;
2. Click on the **pencil** button and then draw a bond;
3. Click on the "**[ ]**" (enumeration) button and then click over the new carbon atom.

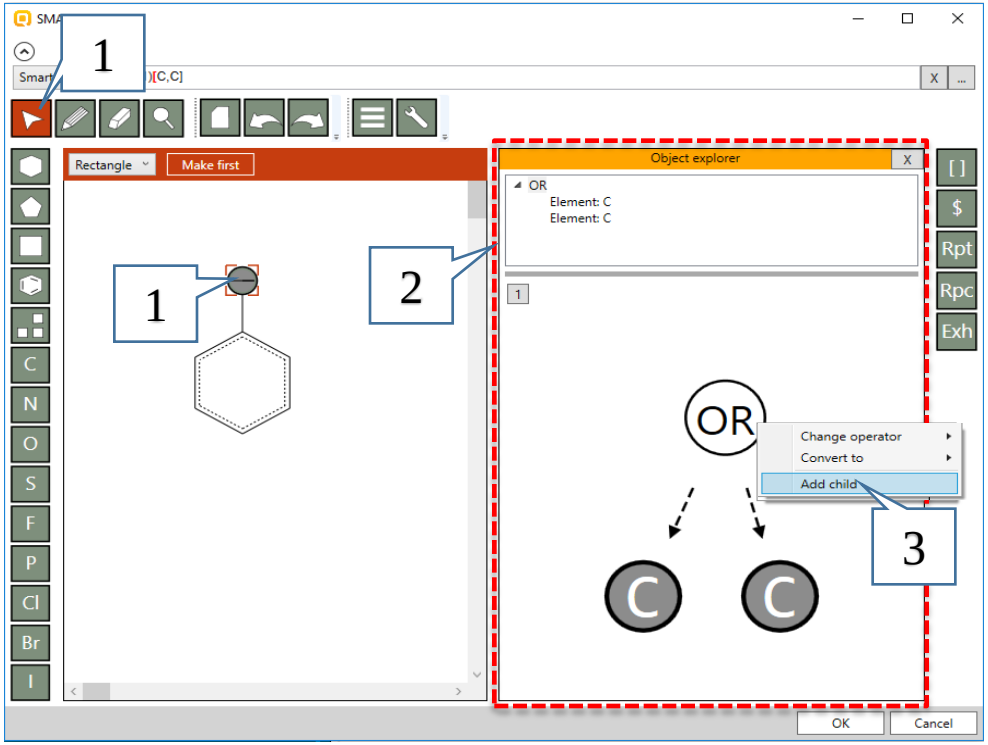


# Workflow process
## Structures search

**Example 1:** Searching for *Aryl halides*

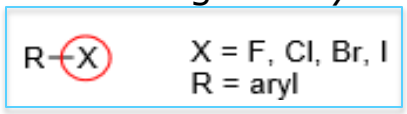



1. Click on the **selection tool** and then click on the new object;
2. *Object explorer* appears;
3. Right click over the "***OR***" expression node and select **Add child**. Repeat this step again.

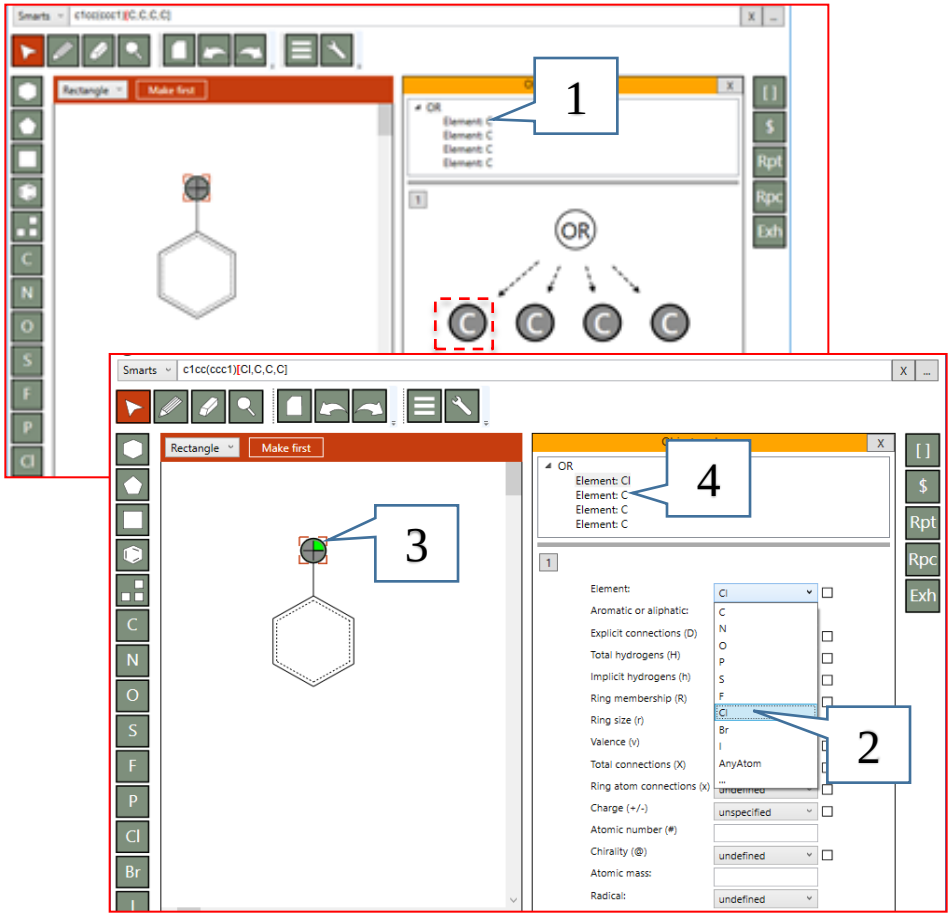


# Workflow process
## Structures search

**Example 1:** Searching for *Aryl halides*

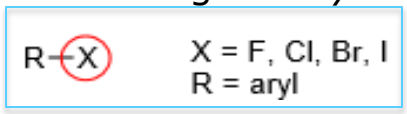



1. Click on the first element which corresponds to the first “child” query. The panel with the atom characteristics appears;
2. Select ***Cl*** from the *Element* drop-down menu;
3. The part of the SMARTS fragment corresponding to the first element have colored.
4. Click on the next elements selecting ***Br***, ***I***, ***F*** successively.

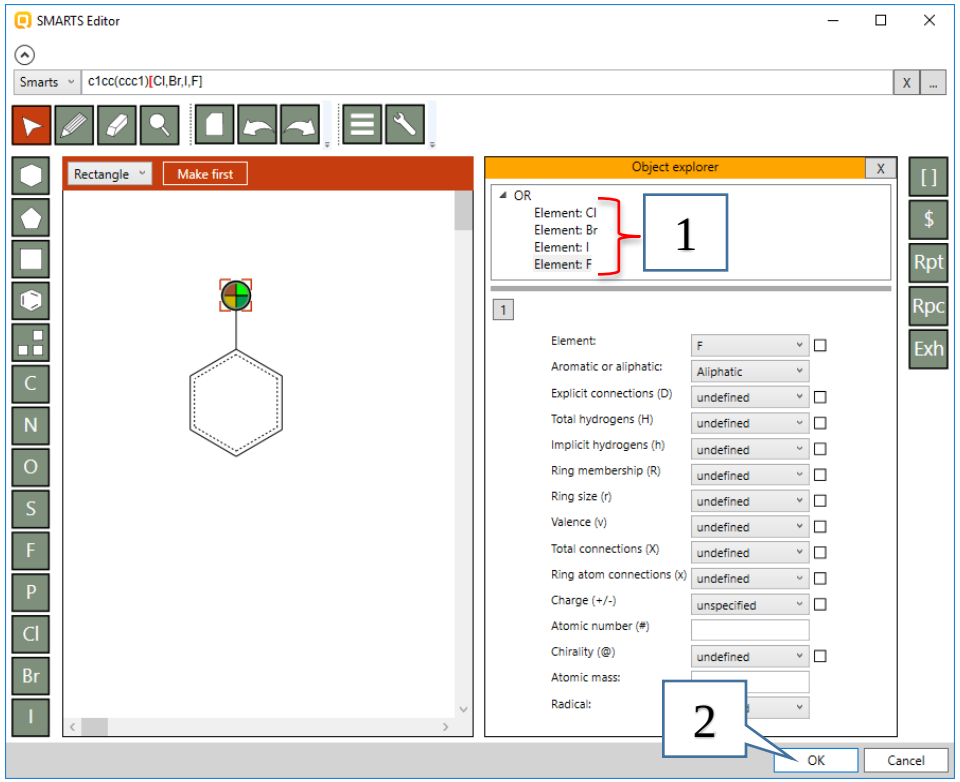


# Workflow process
## Structures search

**Example 1:** Searching for *Aryl halides*



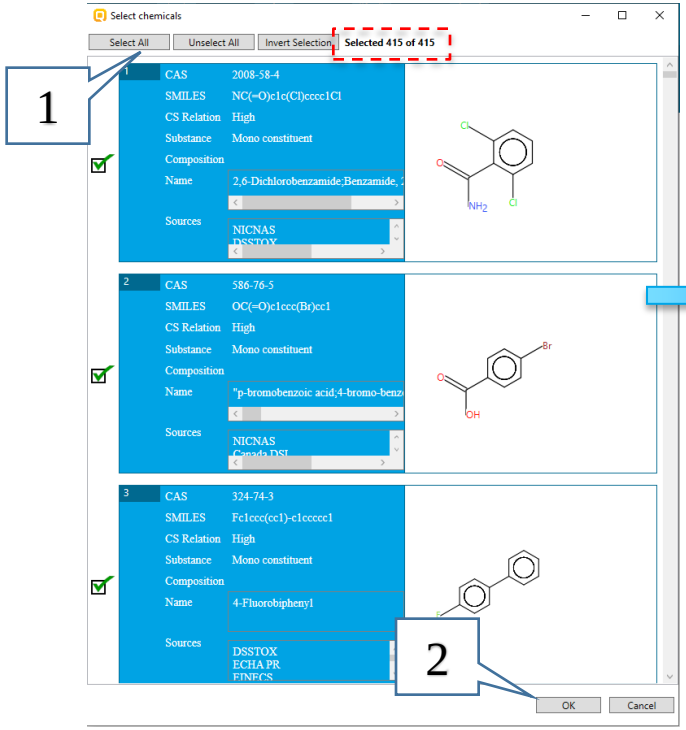

When all elements are selected (1) click on **OK** (2) to execute the search.

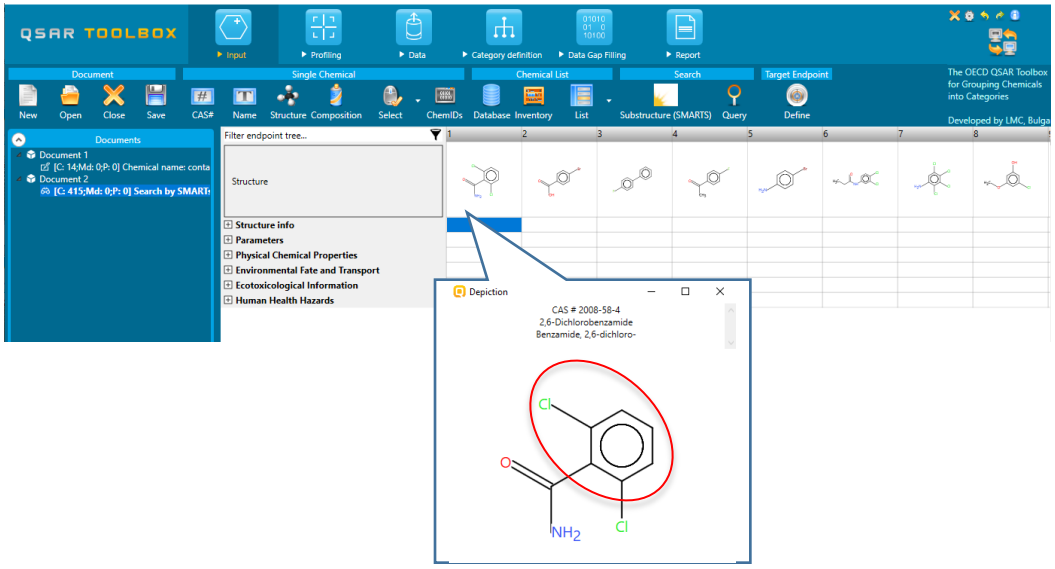


# Workflow process
## Structures search

**Example 1:** Searching for *Aryl halides*

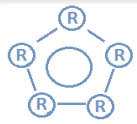
415 chemicals are found in the selected database. Click on Select all button *(1)* and then **OK** (2) to load all of them on the data matrix. You can see that all of the chemicals fulfill the criteria.



# Workflow process
## Structures search

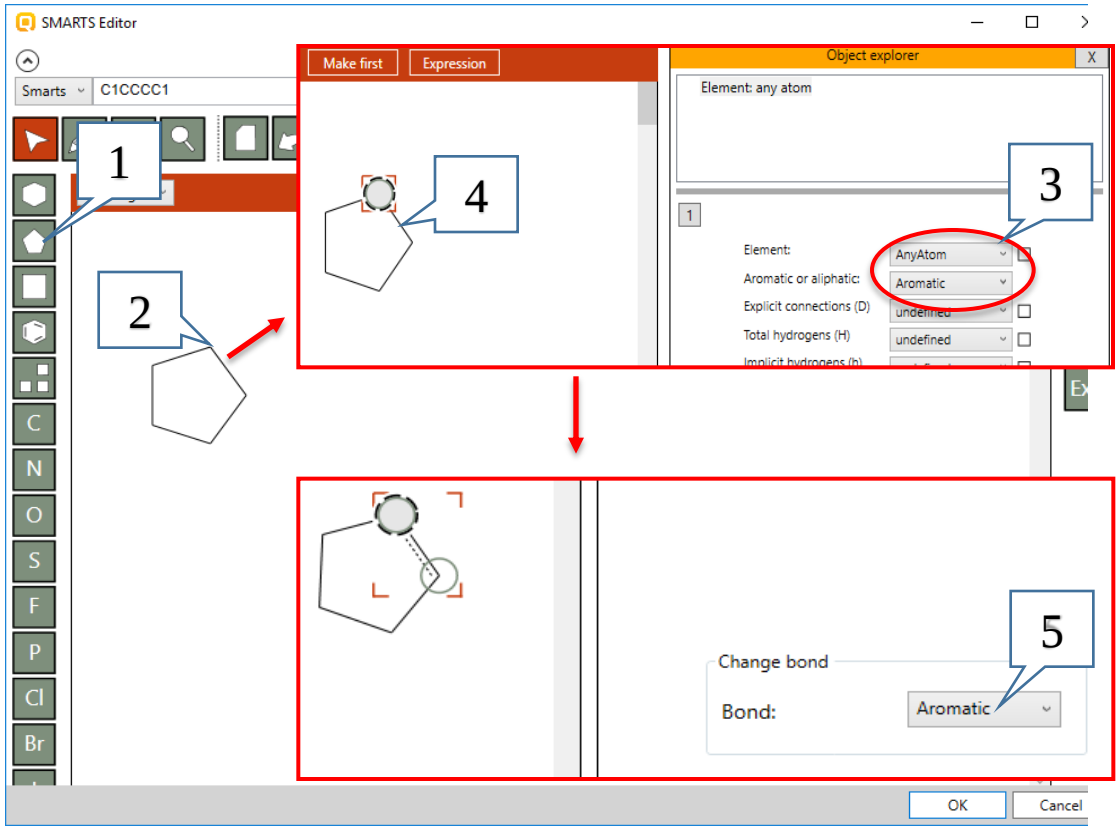
**Example 2:** Searching for *Five-membered heteroaromatic compounds*




R = C (1 to 4), S, N, O

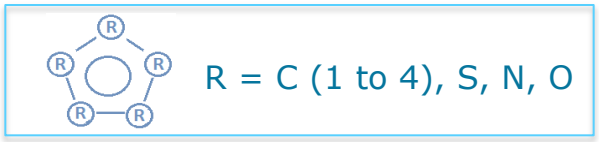
1. Use the pentane template;
2. Click on one of the carbons;
3. Specify the atom to be ***AnyAtom*** and ***Aromatic*** from the drop-down menus;
4. Click on a bond;
5. Change the bond to ***Aromatic***.

Repeat these steps for each of the atoms and bonds in the structure.

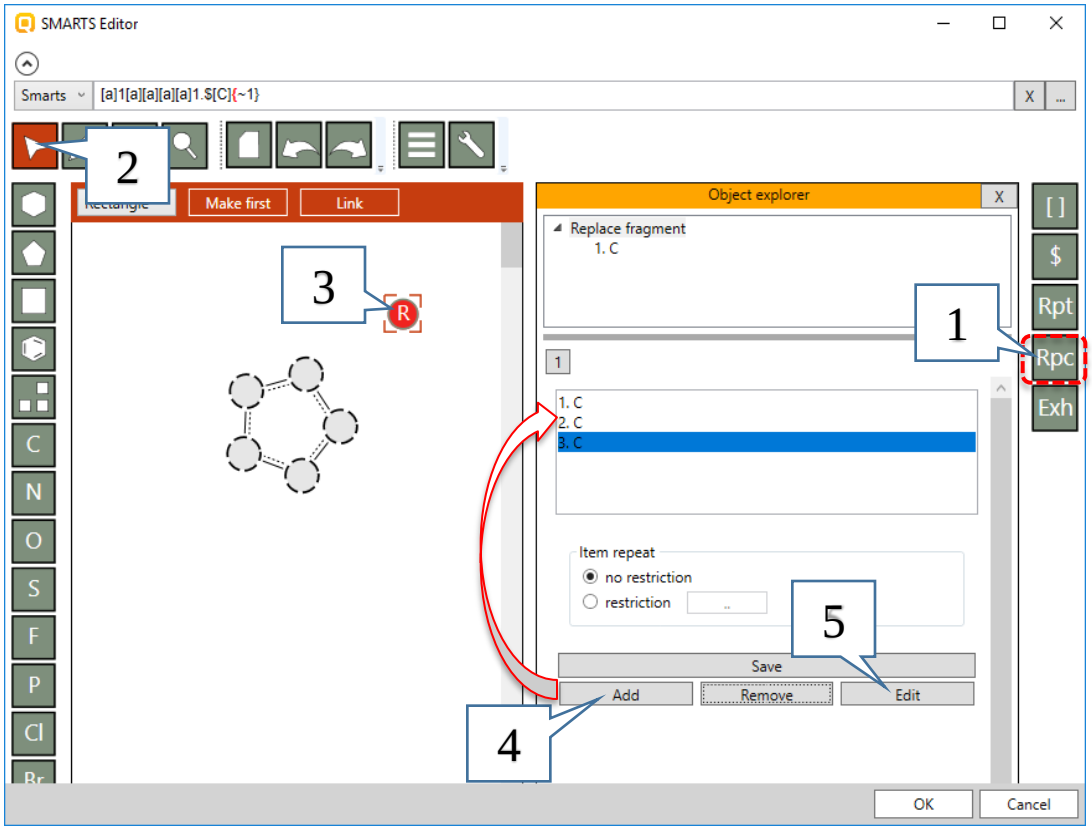


# Workflow process
## Structures search

**Example 2:** Searching for *Five-membered heteroaromatic compounds*

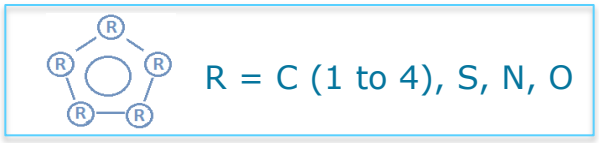



1. Click on the **Rpc** (replace) fragment and paste it in the drawing pane;
2. Click on the **selection tool**;
3. Click on the **red R** (Rpc) fragment;
4. Click on **Add** button three times to include three elements. The elements are aliphatic carbons by default.
5. Click on **Edit** button.

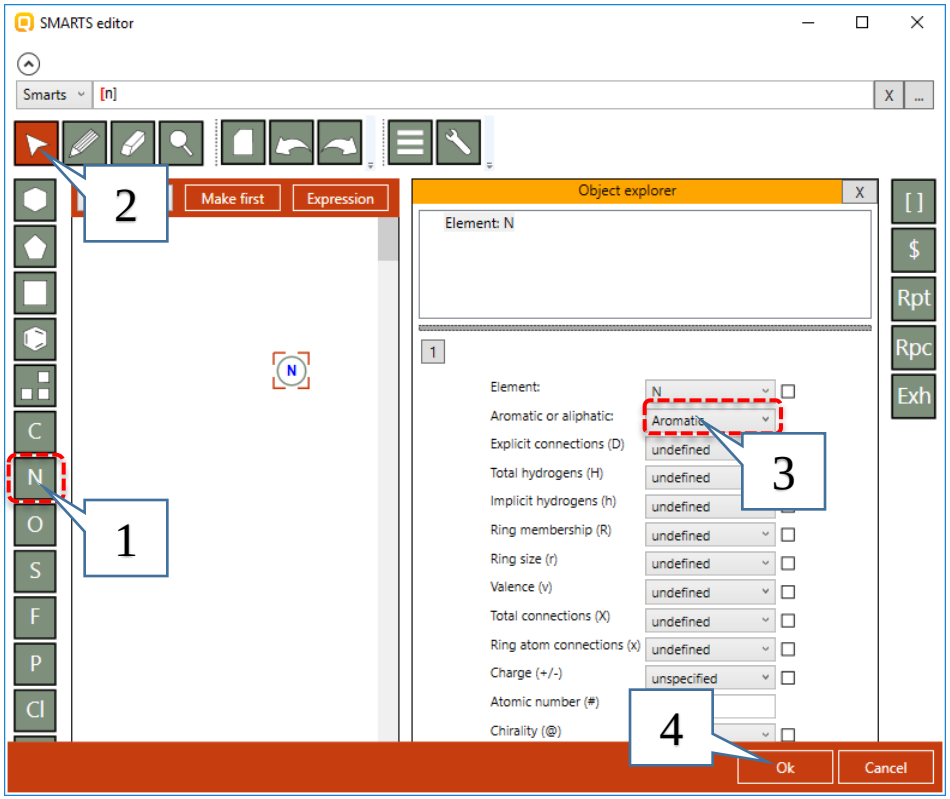


# Workflow process
## Structures search

**Example 2:** Searching for *Five-membered heteroaromatic compounds*

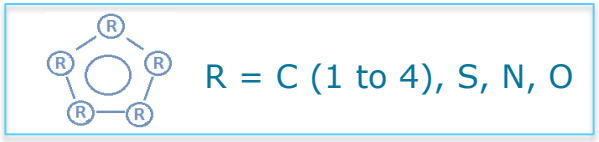



1. Select **N** symbol and put it over the C atom;
2. Click on **selection tool**;
3. Define the nitrogen atom to be aromatic;
4. Click **OK** to confirm



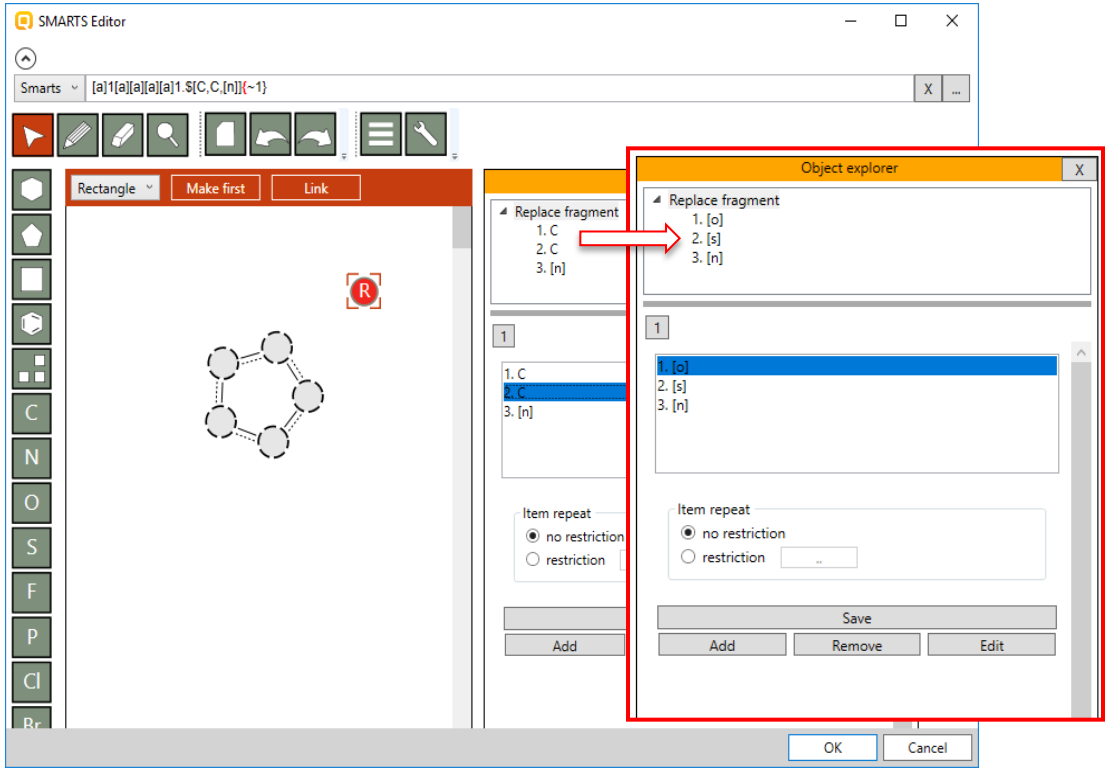
# Workflow process
## Structures search

**Example 2:** Searching for *Five-membered heteroaromatic compounds*



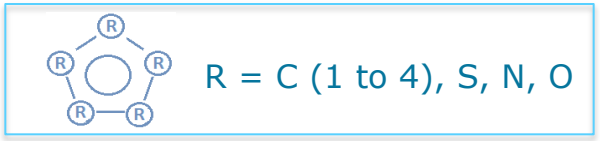

R = C (1 to 4), S, N, O

Change the rest aliphatic carbons to aromatic sulphur and oxygen in the same way;



# Workflow process
## Structures search

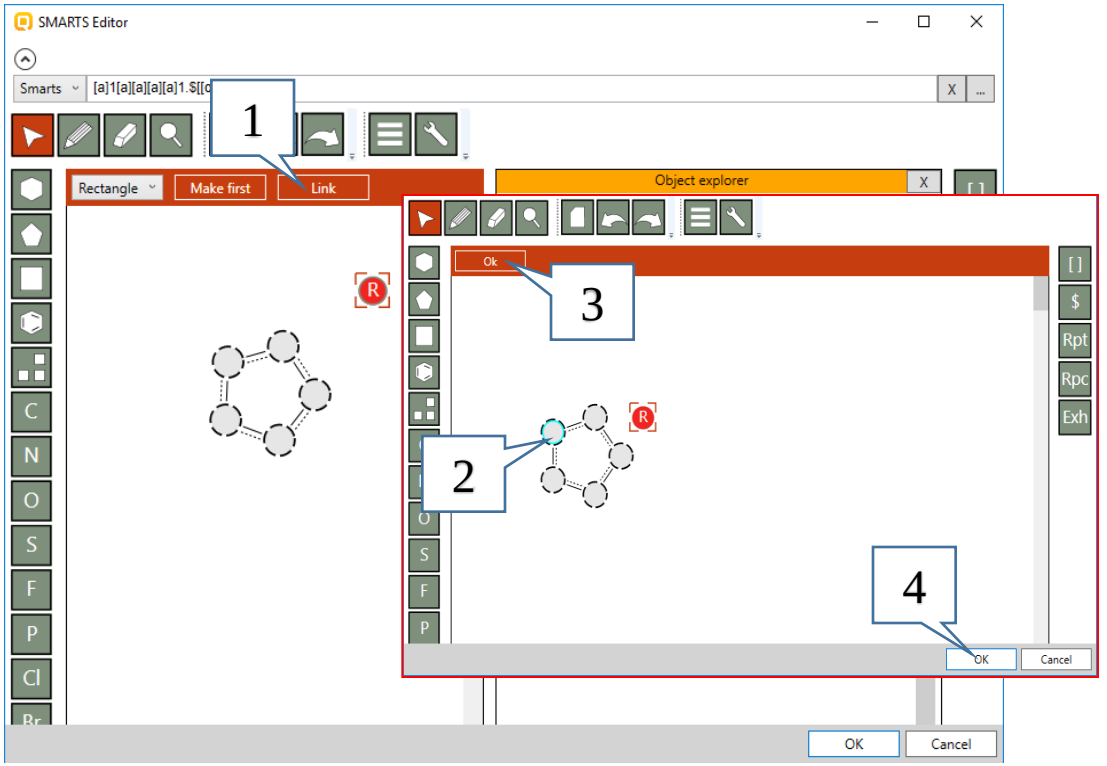
**Example 2:** Searching for *Five-membered heteroaromatic compounds*




1. Click on **Link** button to show the atoms, which will be replaced;
2. Click over one of the atoms;
3. Press ***OK*** to confirm the criteria;

In this way we say that we want the found structures to have at least one heteroaromatic atom;

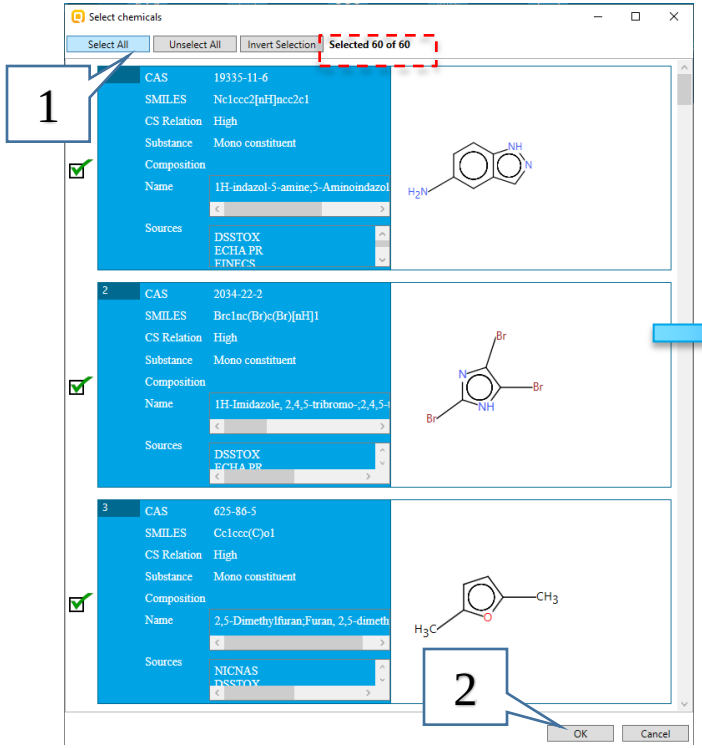
4. Execute the search by the main **OK** button

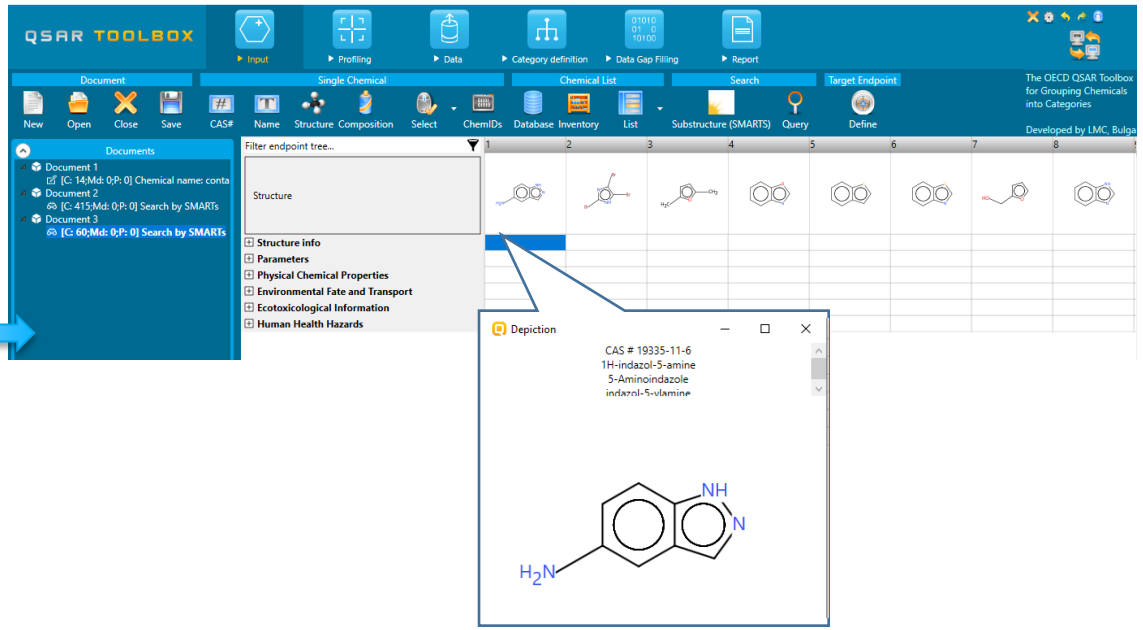


# Workflow process
## Structures search

**Example 2:** Searching for *Five-membered heteroaromatic compounds*

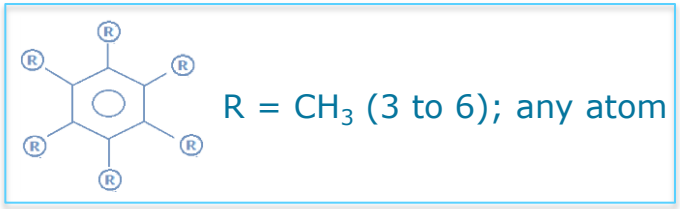
60 chemicals are found in the selected database. Click on **Select All** button (1) then **OK** (2) to load all of them on the data matrix. You can see that all of the chemicals fulfill the criteria.



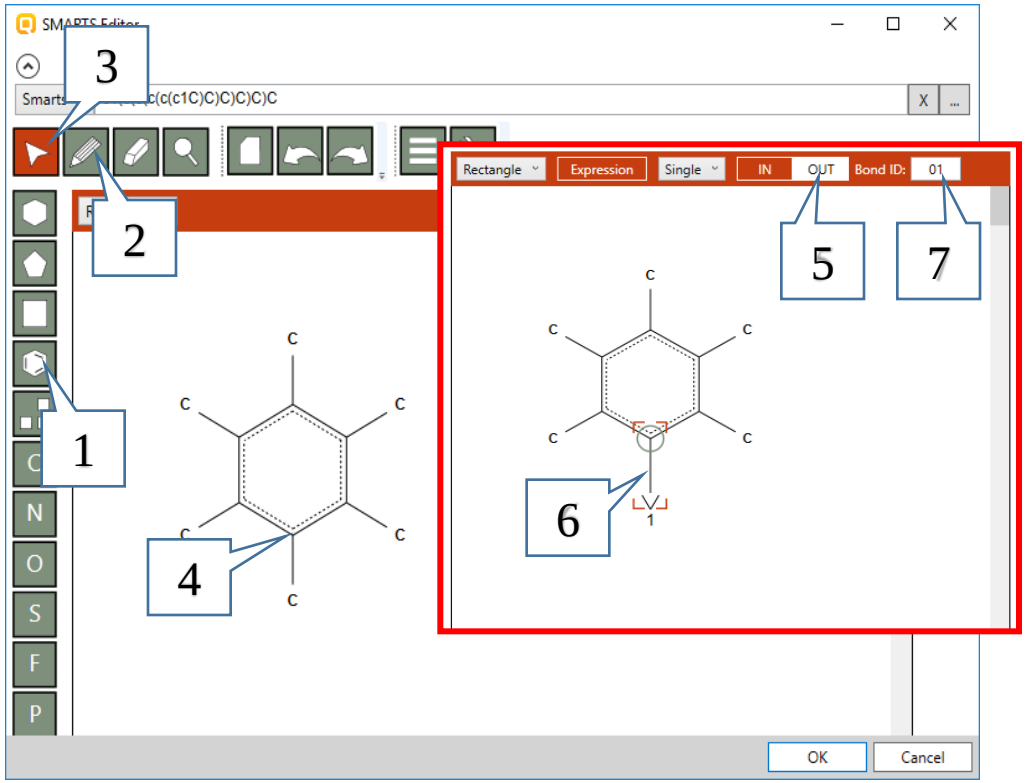
# Workflow process
## Structures search

**Example 3:** Searching for *aryl compounds with three or more methyl groups*



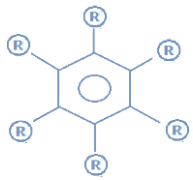

1. Select **benzene** template and put it in the drawing pane;
2. Select the **pencil** and draw bonds to each of the benzene atoms;
3. Click on **selection tool**;
4. Click over a bond;
5. Select "**Out**" of the new appeared buttons in the red bar;
6. Click over the same bond again;
7. Write "**1**" in the *Bond ID* field.

Repeat the steps above (from 4 to 7) for each of the bonds.



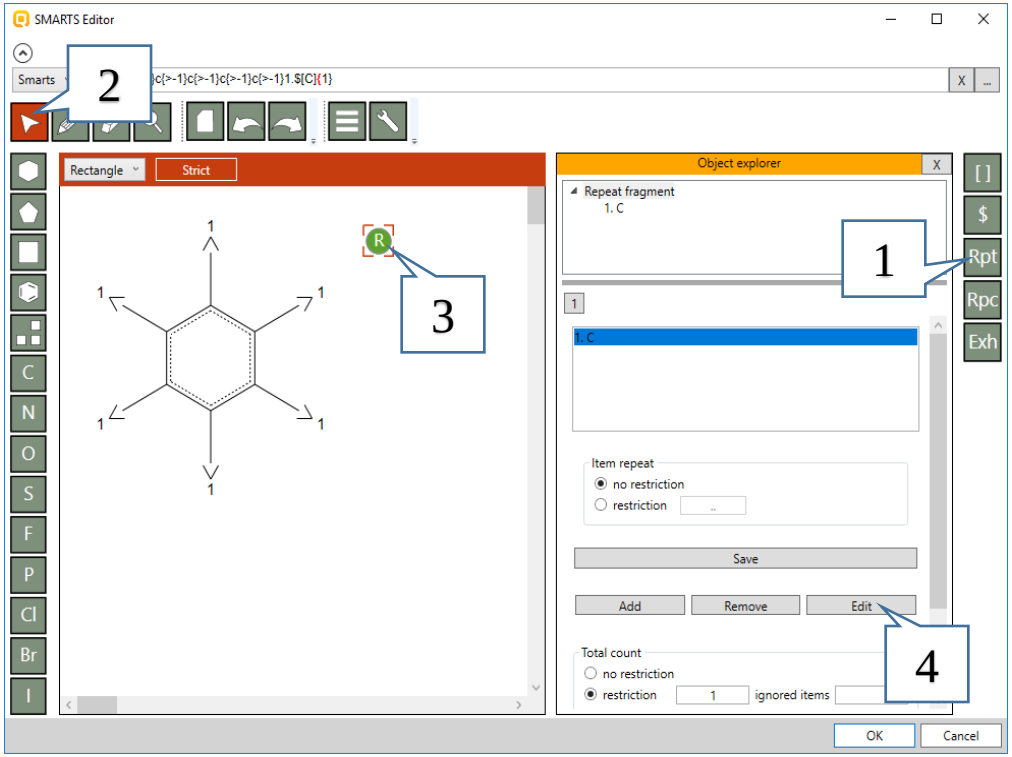
# Workflow process
## Structures search

**Example 3:** Searching for *aryl compounds with three or more methyl groups*



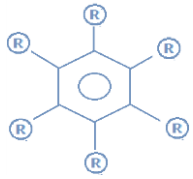
R = $CH_3$ (3 to 6); any atom

1. Select **Rpt** (repeat) fragment and put it in the drawing pane;
2. Click on the **selection tool**;
3. Click over the **green R** (Rpt);
4. Click on the **Edit** button.



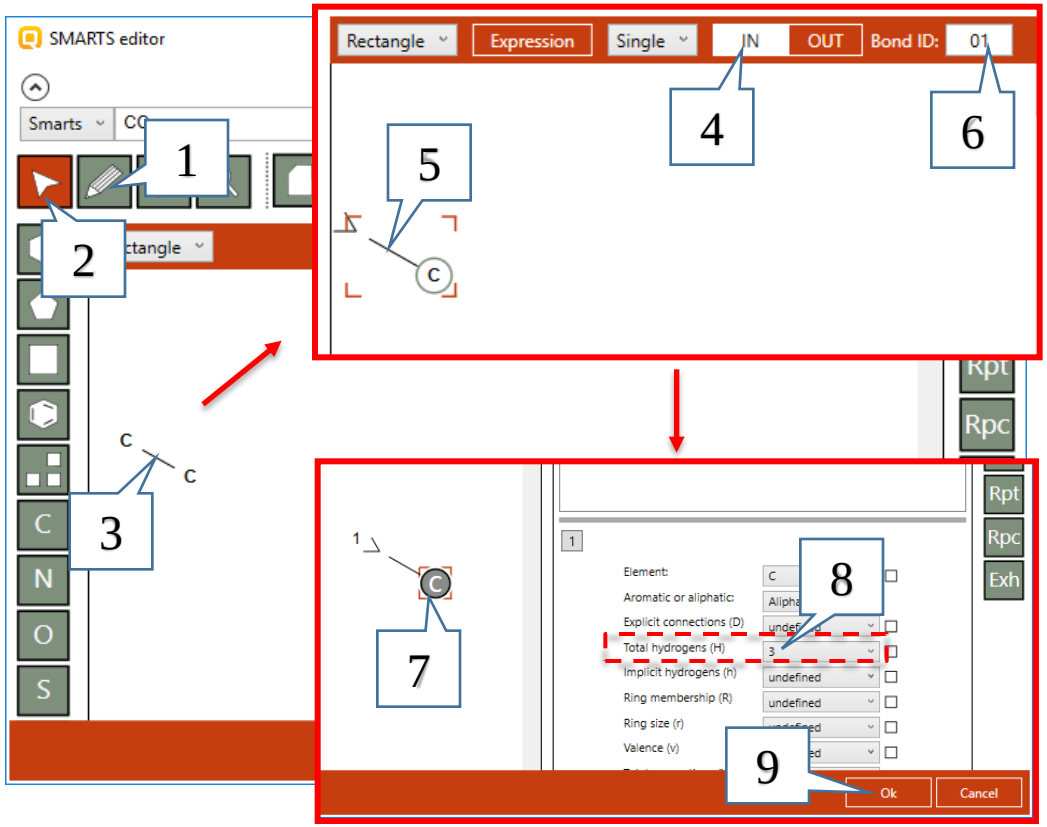
# Workflow process
## Structures search

**Example 3:** Searching for *aryl compounds with three or more methyl groups*



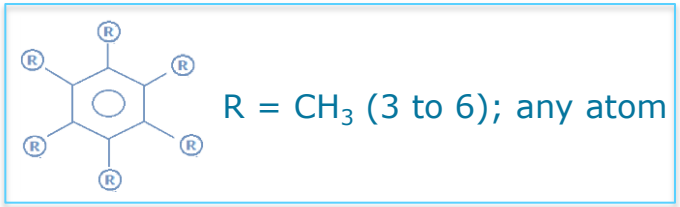
R = $CH_3$ (3 to 6); any atom

1. Select the **pencil** and draw a bond;
2. Click on the **selection tool**;
3. Click over the bond;
4. Select "**In**" of the new appeared buttons in the red bar;
5. Click over the same bond again;
6. Write "**1**" in the *Bond ID* field.
7. Click over the carbon atom;
8. Define total hydrogens to be three;
9. Confirm with **OK**.

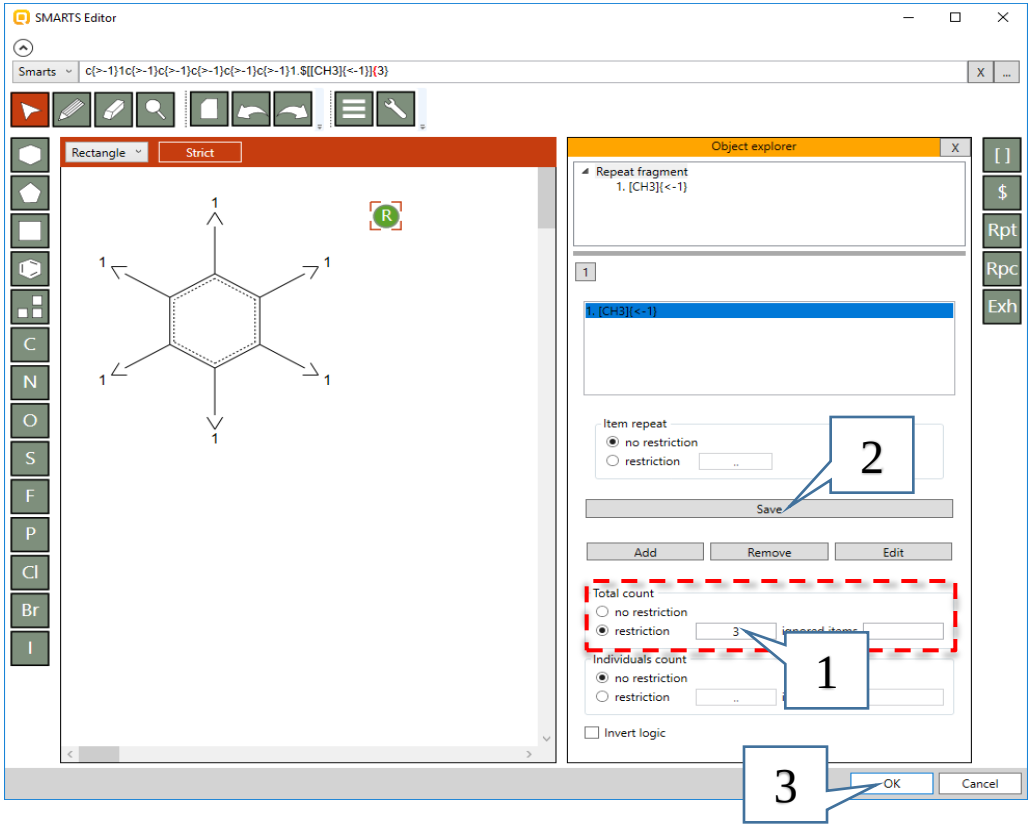


# Workflow process
## Structures search

**Example 3:** Searching for *aryl compounds with three or more methyl groups*



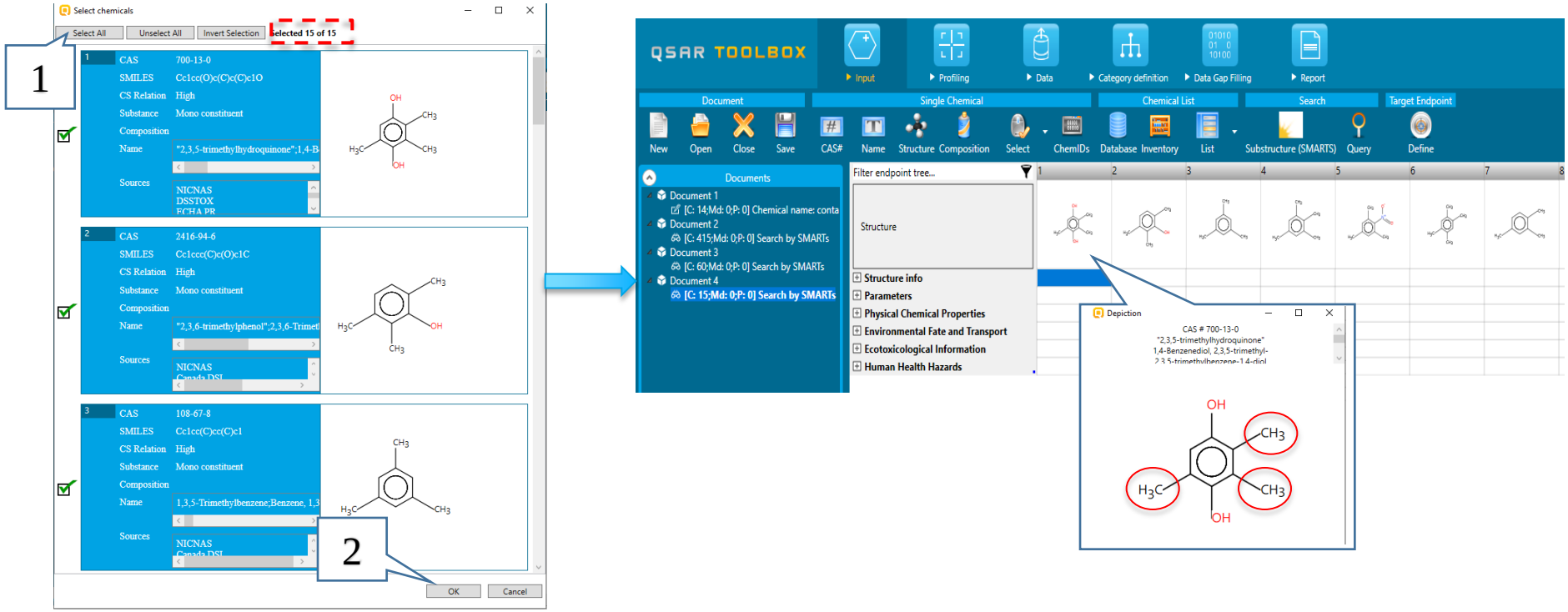

1. Type **3** in the restriction field for the ***Total count***. Structures with at least three methyl groups will be searched;
2. Click on **Save**;
3. Execute the search by click on **OK**



# Workflow process
## Structures search

**Example 3:** Searching for *aryl compounds with three or more methyl groups*

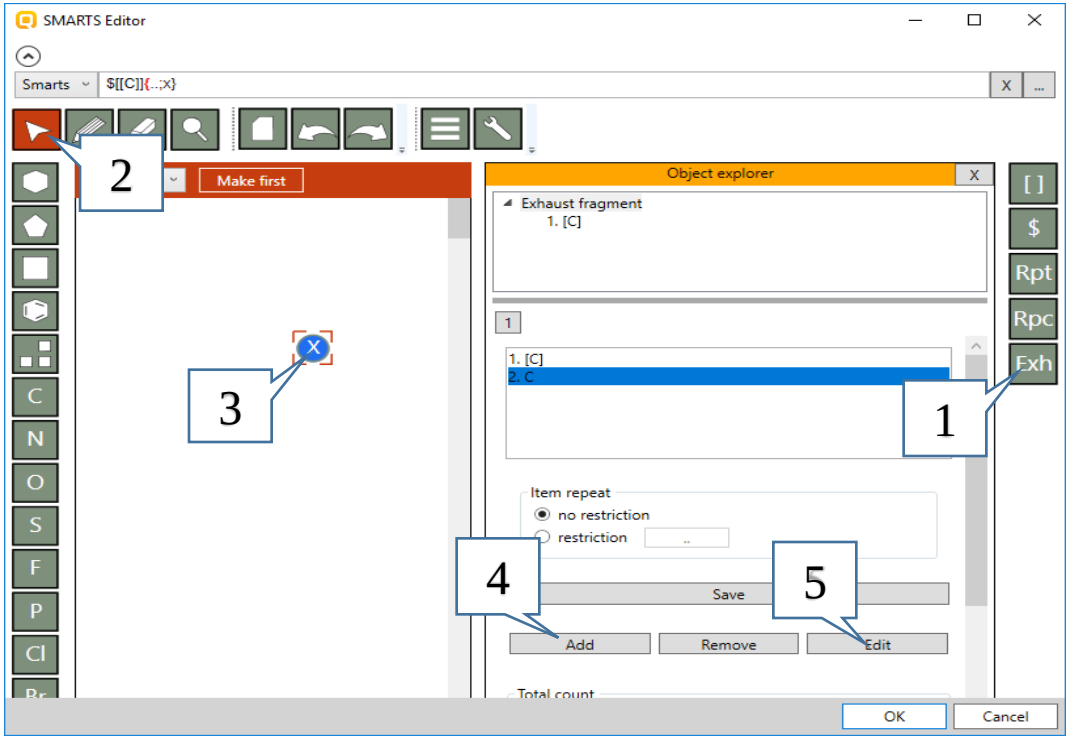
15 chemicals (1) are found in the selected database. Click on **OK** (2) to load all of them on the data matrix. You can see that all of the chemicals fulfill the criteria.



# Workflow process
## Structures search

**Example 4:** Searching for hydrocarbon structures with 3 to 6 atoms and at least one hydroxyl group

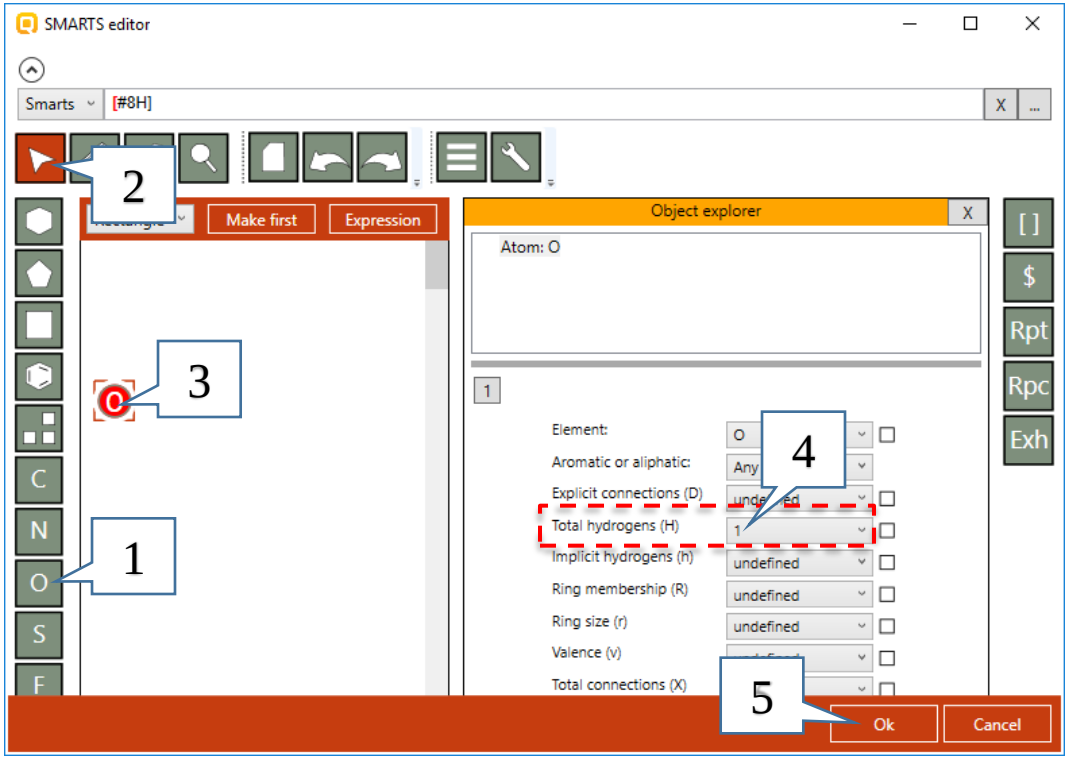
1. Select **Exh** (exhaust) fragment and put it in the drawing pane;
2. Click on **selection tool**;
3. Click over the **X** (Exh);
4. Click on **Add** button;
5. Click on **Edit** button.



# Workflow process
## Structures search

**Example 4:** Searching for hydrocarbon structures with 3 to 6 atoms and at least one hydroxyl group

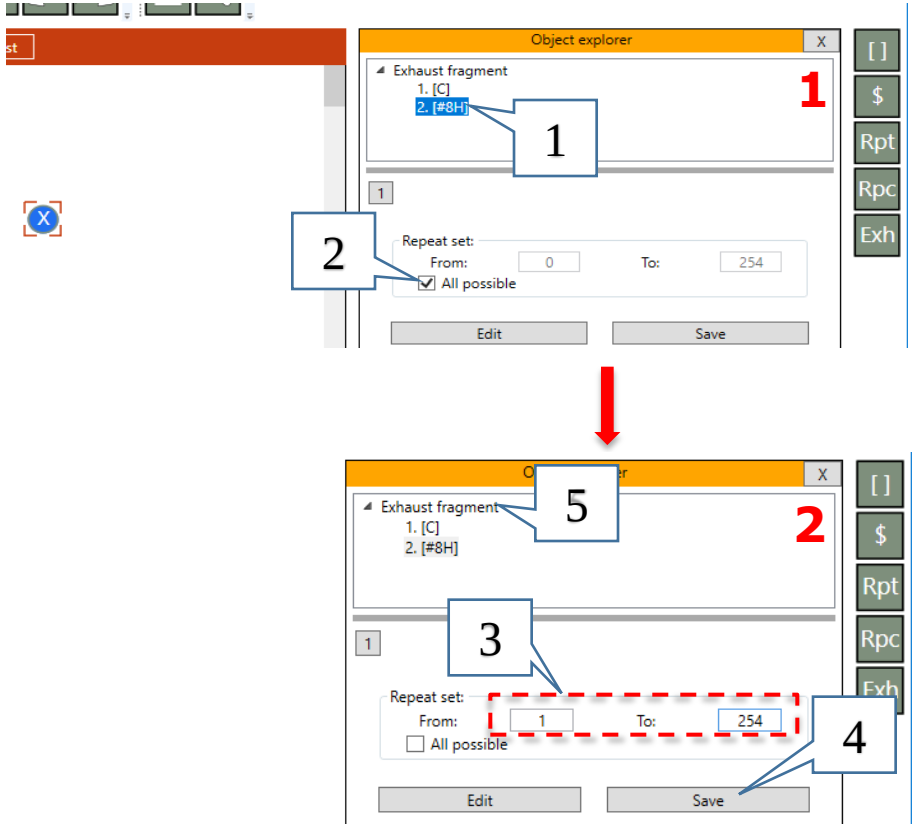
1. Select oxygen symbol (**O**) and put it over the carbon atom;
2. Click on **selection tool**;
3. Click over the **oxygen**;
4. Define number of *Total hydrogens (H)* to be one;
5. Confirm with **OK**.



# Workflow process
## Structures search

**Example 4:** Searching for hydrocarbon structures with 3 to 6 atoms and at least one hydroxyl group

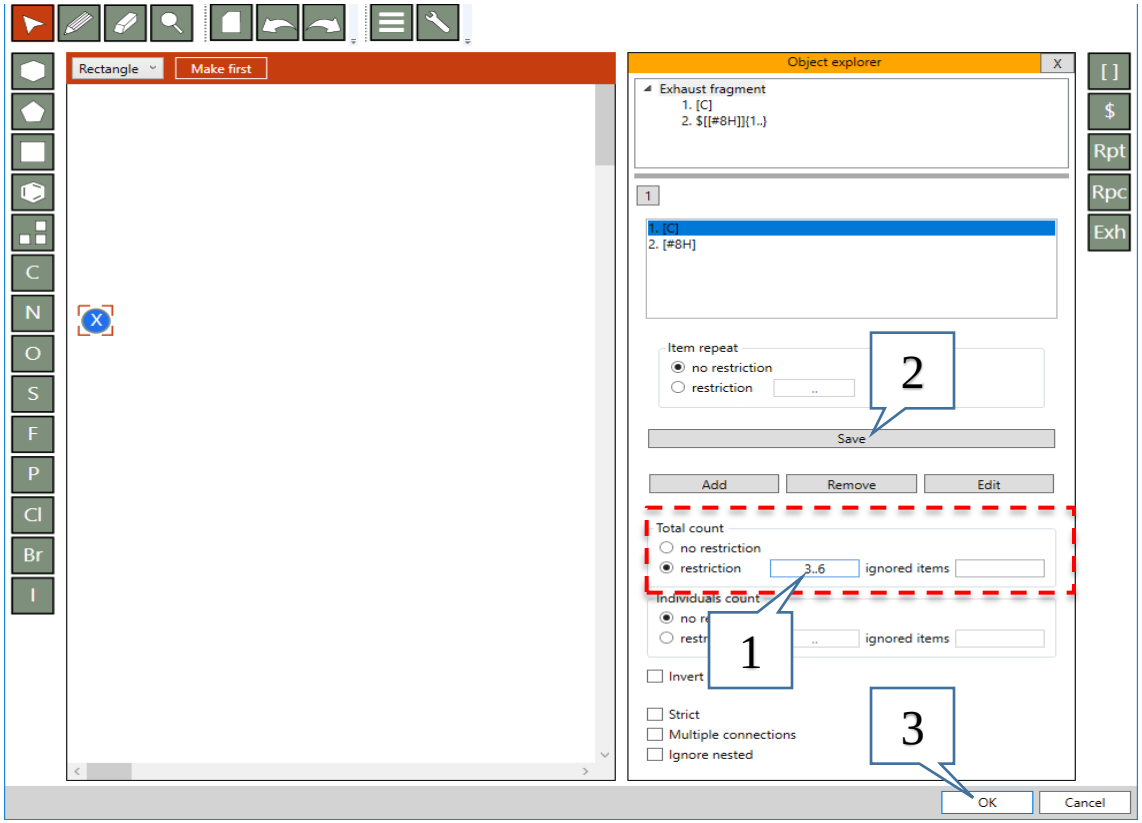
1. Click on the second item (hydroxyl group) in the exhaust fragment;
2. Uncheck *All possible;*
3. Define range from 1 to 254 (or to present at least one time in the structure);
4. Click on **Save**;
5. Go back to the ***Exhaust fragment***.



# Workflow process
## Structures search

**Example 4:** Searching for hydrocarbon structures with 3 to 6 atoms and at least one hydroxyl group

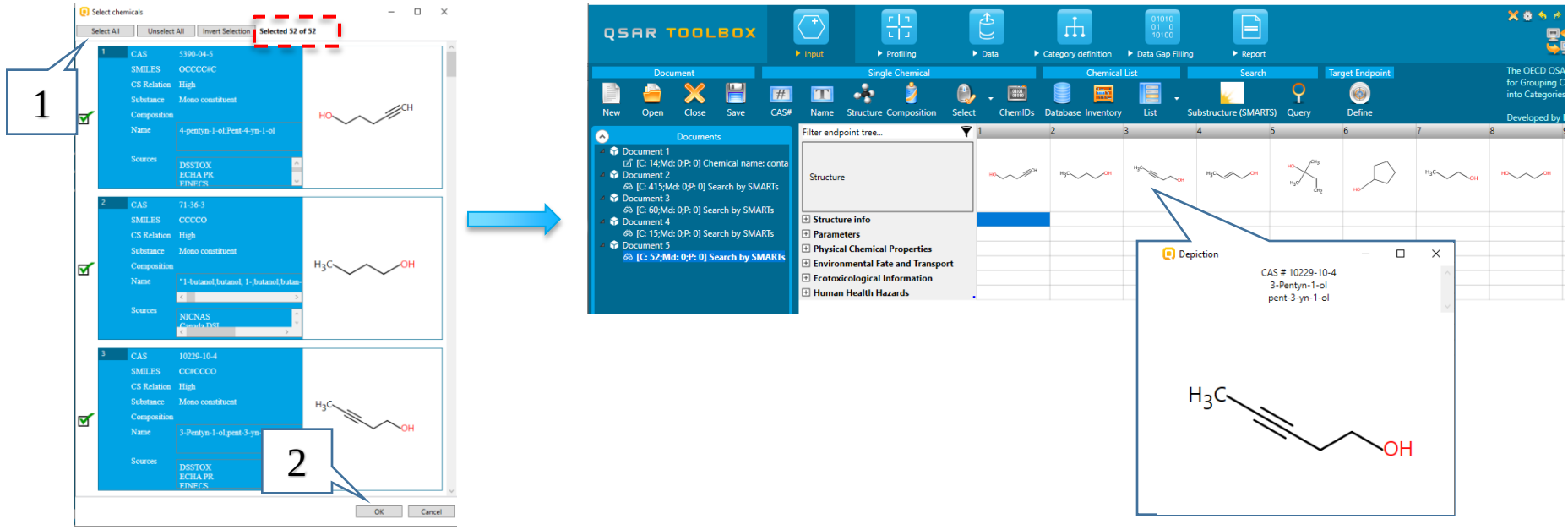
1. Set total restriction of the atoms to be from **3** to **6**.;
2. Click on ***Save***;
3. Execute the search by **OK**



# Workflow process
## Structures search

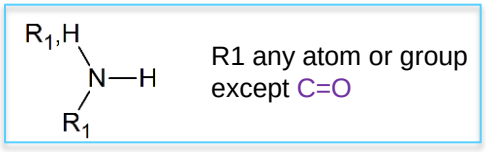
**Example 4:** Searching for hydrocarbon structures with 3 to 6 atoms and at least one hydroxyl group

52 chemicals (1) are found in the selected database. Click on **Select All** (1) button and then **OK** (2) to load all of them on the data matrix. You can see that all of the chemicals fulfill the criteria.

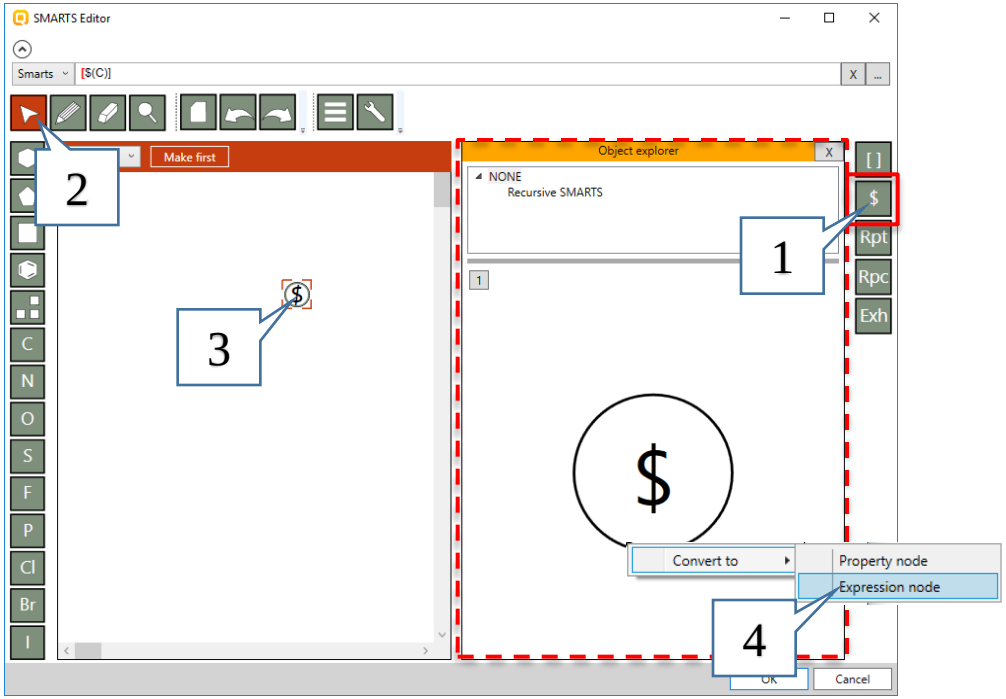


# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide

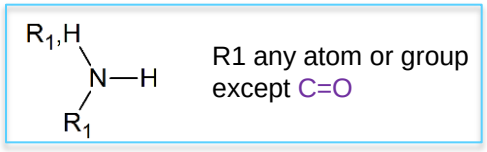



1. Select **$** (recursive SMARTS) fragment and put it in the drawing pane;
2. Click on the **selection tool**;
3. Click over the **$** fragment;
4. Right click over the $ in the *Object explorer* panel and select ***Convert to > Expression node***



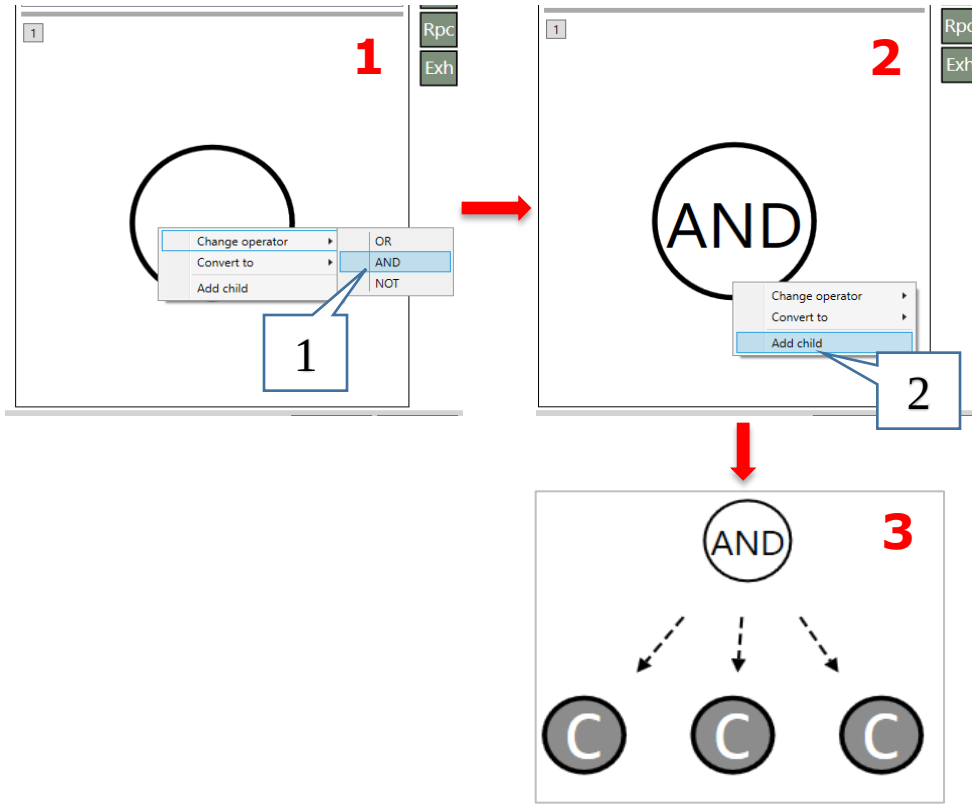
# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide



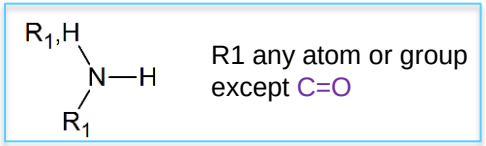

R1 any atom or group except C=O

1. Right click over the ball and select ***Change operator > AND***
2. Right click over the AND operator and select **Add child**.
3. Repeat the 2nd step three times to obtain three “child” queries.



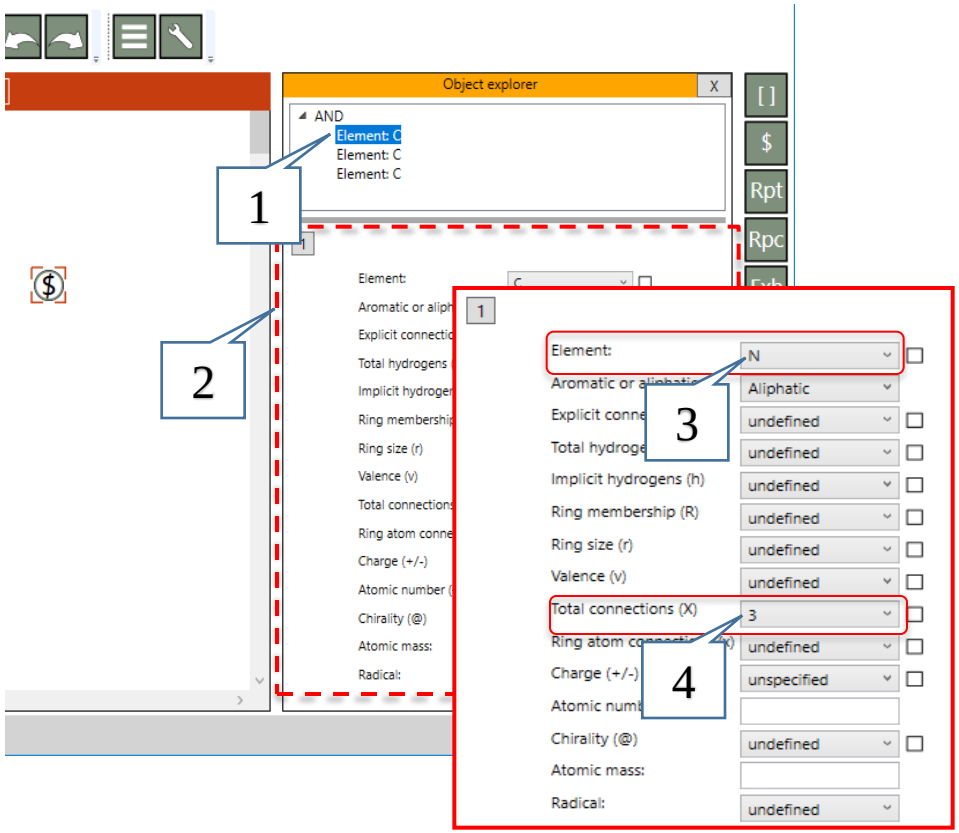
# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide



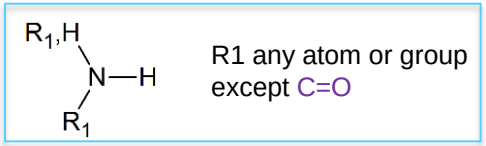

R1 any atom or group except C=O

1. Click on the first item of "**AND**";
2. Atom characteristics window appears;
3. Select *Element* to be nitrogen from the drop-down menus;
4. Select *Total connections* to be three from the drop-down menu.

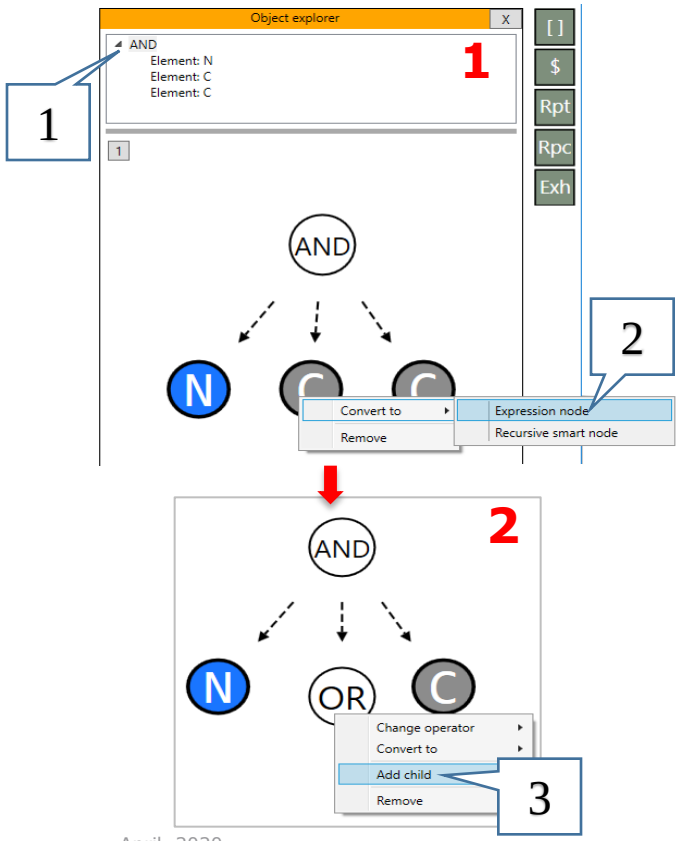


# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide

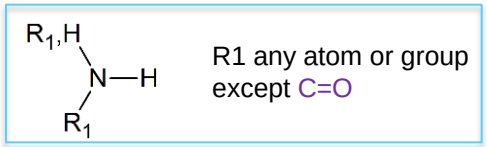



1. Go back to the "**AND**";
2. Right click over the second query and select ***Convert to > Expression node***;
3. Right click over the **OR** and select ***Add child***. Add two child queries.



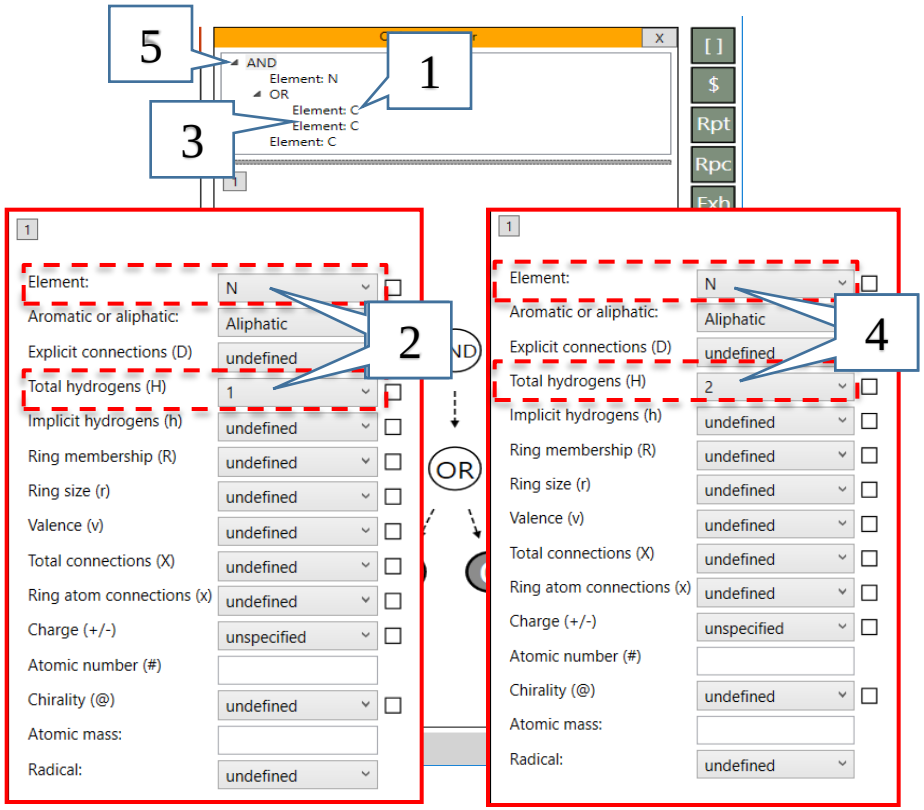
# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide



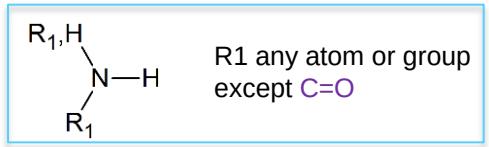

R1 any atom or group except C=O

1. Click on the first item of "***OR***"
2. Define primary amine group by select the *Element* to be **N** and *Total hydrogens* to be **1**;
3. Click on the second item of "***OR***"
4. Define secondary amine group by select the *Element* to be **N** and *Total hydrogens* to be **2**;
5. Go back to the **AND** root.



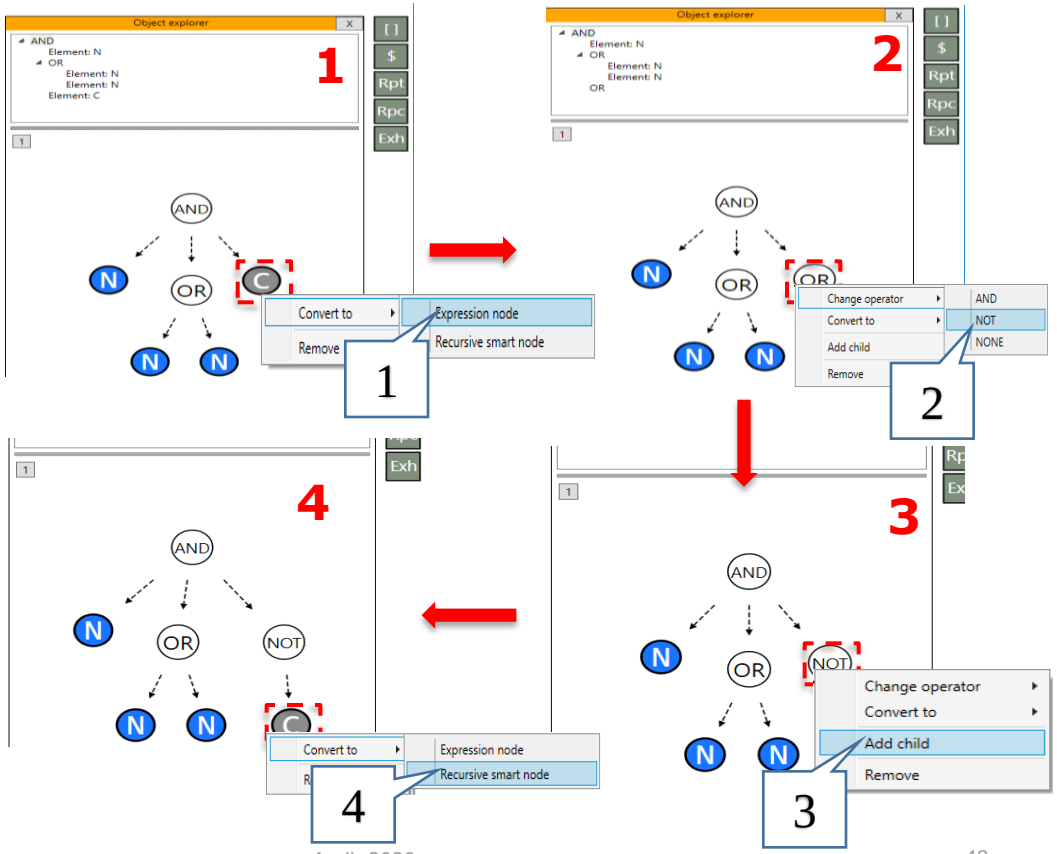
# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide



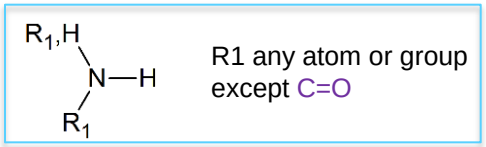

R1 any atom or group except C=O

1. Right click on the last query and select ***Convert to > Expression node***;
2. Right click on the new **OR** expression node and select ***Change operator > NOT***;
3. Right click on the **NOT** query and select ***Add child***;
4. Right click on the new child query and select ***Convert to > Recursive smart node***.



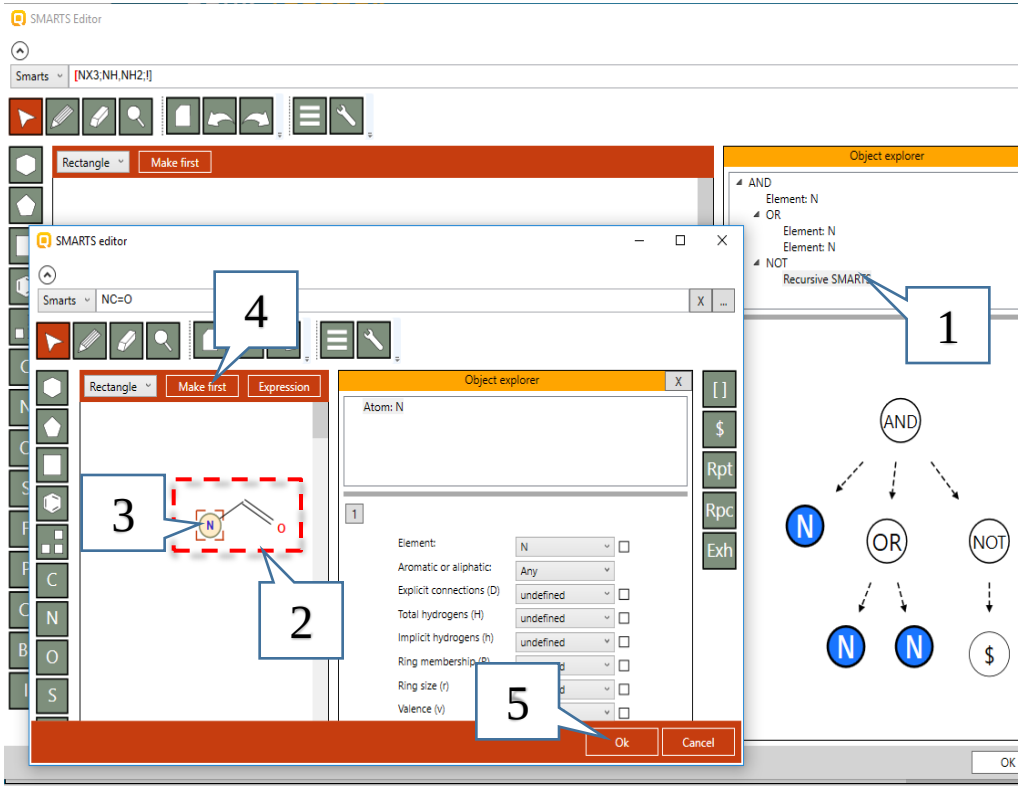
# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide



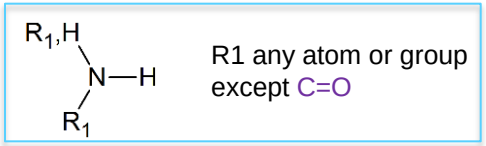

R1 any atom or group except C=O

1. Click on the ***Recursive SMARTS*** label. New SMARTS Editor window appears;
2. Draw amide group (NC=O);
3. Click over the nitrogen (N) of the amide group;
4. Select ***Make first***;
5. Confirm with **OK**.



# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide

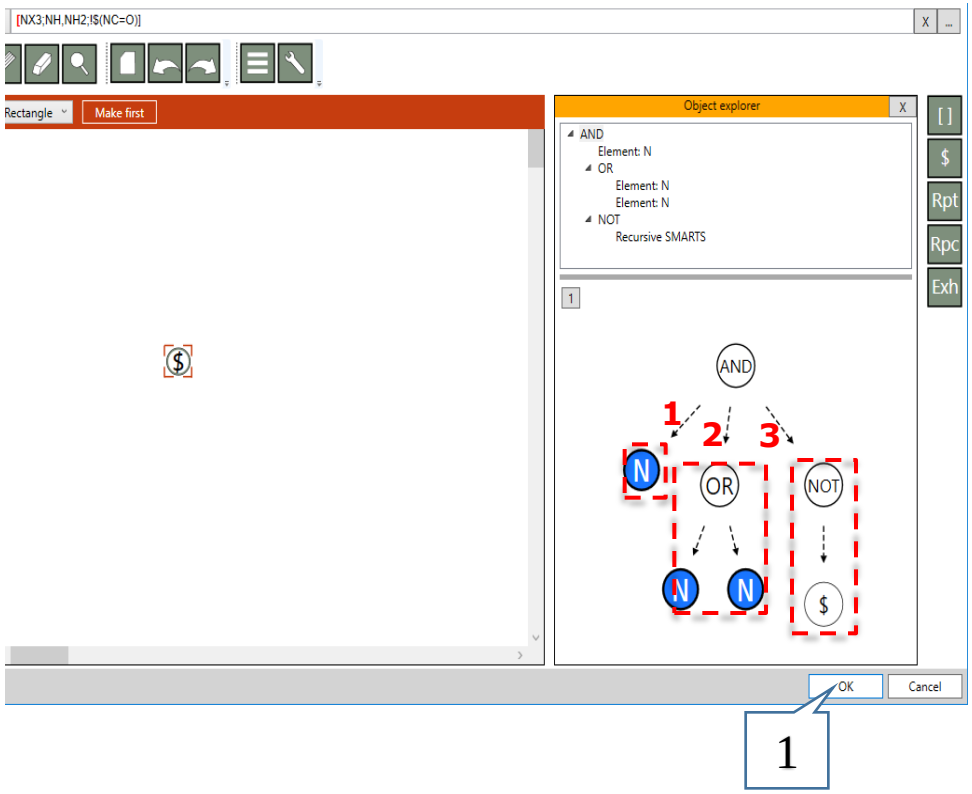



***Now we are ready to execute the search***.

Recap:

1. We defined that we will search for chemicals with nitrogen atom with three connections to other atoms;
2. We defined that the nitrogen to be part of primary or secondary amine functional group;
3. We defined that the nitrogen to be not part of amide functional group.

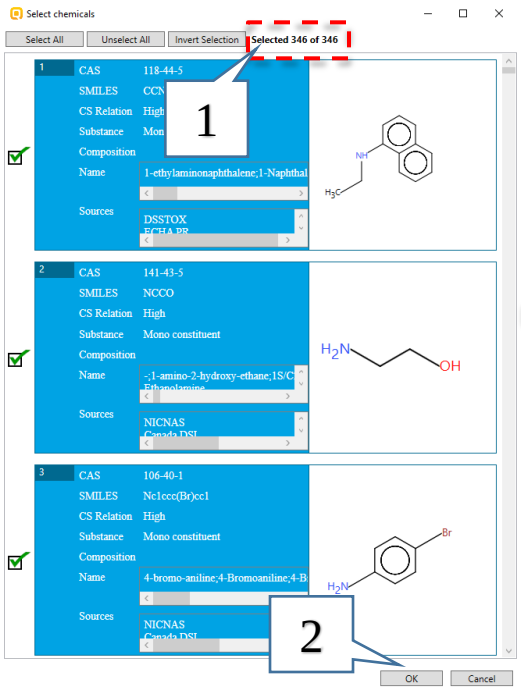
Click on **OK** (1) to execute the search.

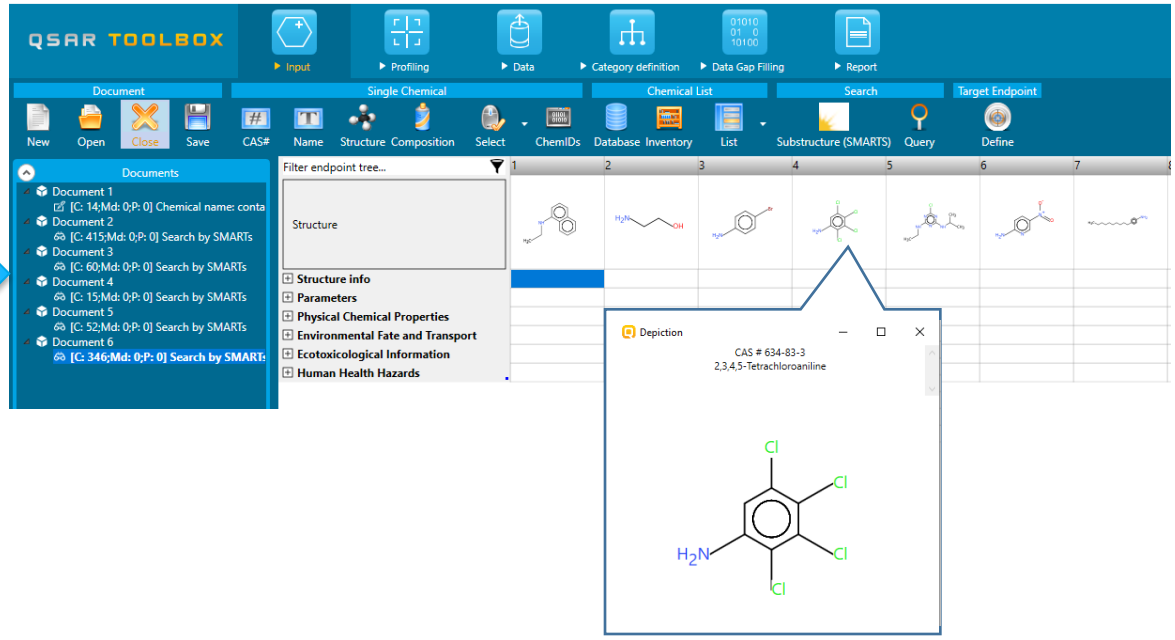


# Workflow process
## Structures search

**Example 5:** Searching for chemicals having primary or secondary amine group, but not amide

346 chemicals (1) are found in the selected database. Click on **Select All** button and then click on **OK** (2) to load all of them on the data matrix. You can see that all of the chemicals fulfill the criteria.



# Workflow process
## Overview

- Five examples illustrating searching for structures by SMARTS editor fragments were represented:
  - **[ ]** – Searching for Aryl halides;
  - **Rpc** – Searching for Five-membered heteroaromatic compounds;
  - **Rpt** – Searching for aryl compounds with three or more methyl groups;
  - **Exh** – Searching for hydrocarbon structures with 3 to 6 atoms and at least one hydroxyl group;
  - **$** - Searching for chemicals having primary or secondary amine group, but not amide

# Congratulation

- You have completed the tutorial on the substructure search by SMARTS editor;
- You have now been introduced to each of the general TB SMARTS fragments;
- Note proficiency comes with practice.