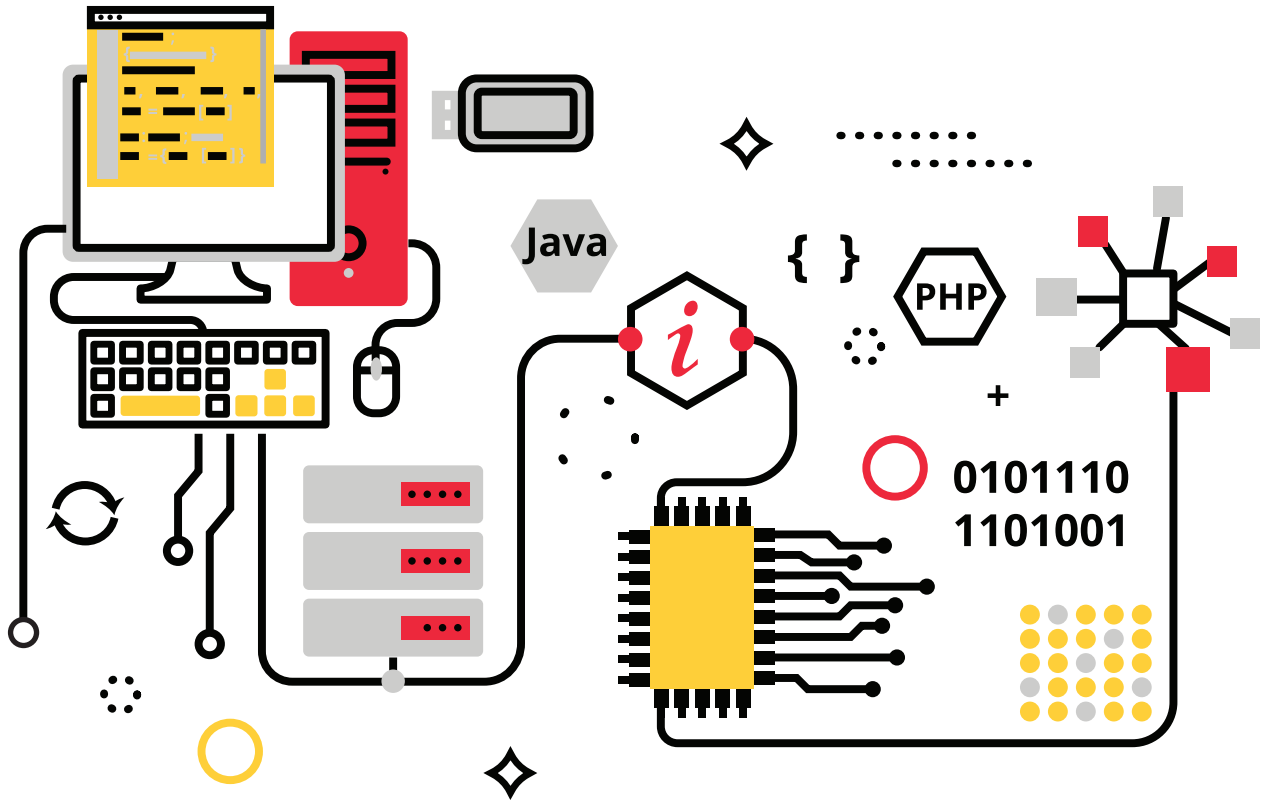
wellcome centre anti-infectives research



KNIME: Reaction Customisation

Wellcome Centre for Anti-Infectives Research (*WCAIR*)

TRAINING

Contents

1.0 About this exercise

From KNIME Reaction Enumeration exercise, you learned to virtually create an amide compound collection from a commercial catalogue. In this exercise, we will explore other coupling reactions using SMARTS to enumerate a compound collection.

Create a new folder in your "Documents" called "knime-workspace".
Ensure you download the following into this folder:

- A KNIME workflow (.knwf) file, **"WCAIR_01_Reaction_Customisation.knwf"**
- An SDF file, **"customization-example.sdf"** that contains a small table of alcohols and carboxylic acids to be used in this example

1.1 What you will need

1.1.1 SDF file

For this exercise, you have been provided an example SDF file "customization-example.sdf".
Once competent with the exercise, you may wish to use your own SDF customised table*.
Make sure your SDF file containing your molecule catalogue has at least the below three columns:

- "Molecule", containing the structure of your reactant
- "ID", containing a text identifier (e.g.: catalogue code) for this reactant
- "Catalog", containing the supplier (Sigma, Fisher, TCI, etc.)

These three columns are **mandatory**.

During the workflow execution, KNIME will look for these columns and raise an error if they are not present.

* *You can use the free and open-source tool DataWarrior to create SDF files (**Figure 1**). After creating your own catalogue table, you can use the menu File > Save Special > SD-File to export the table as an SDF file.*

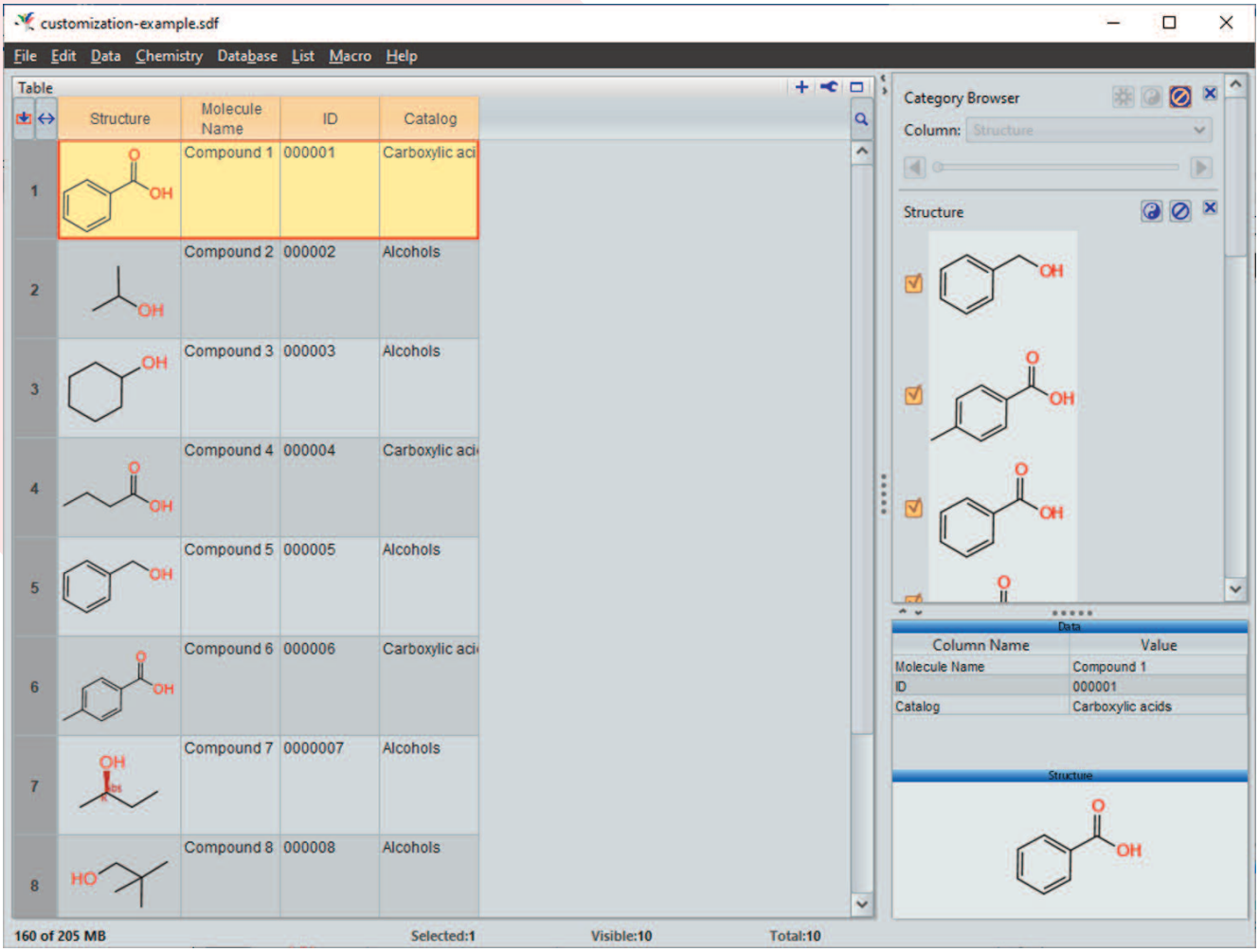


Figure 1 *DataWarrior window with an example of a catalogue SDF file.*

1.1.2 SMARTS reaction strings

You will also need your own SMARTS string for your reactions if you're planning to use it for something other than amide couplings (see KNIME Reaction Enumeration).

2.0 Using your own catalogue table

2.1 Uploading your .sdf file

Open "WCAIR_01_Reaction_Customisation.knwf" > double click "Choose SDF to open" node.

Locate example file "customization-example.sdf" by clicking the "Browse" button (**Figure 2**). Once selected, click "OK" to close this window.

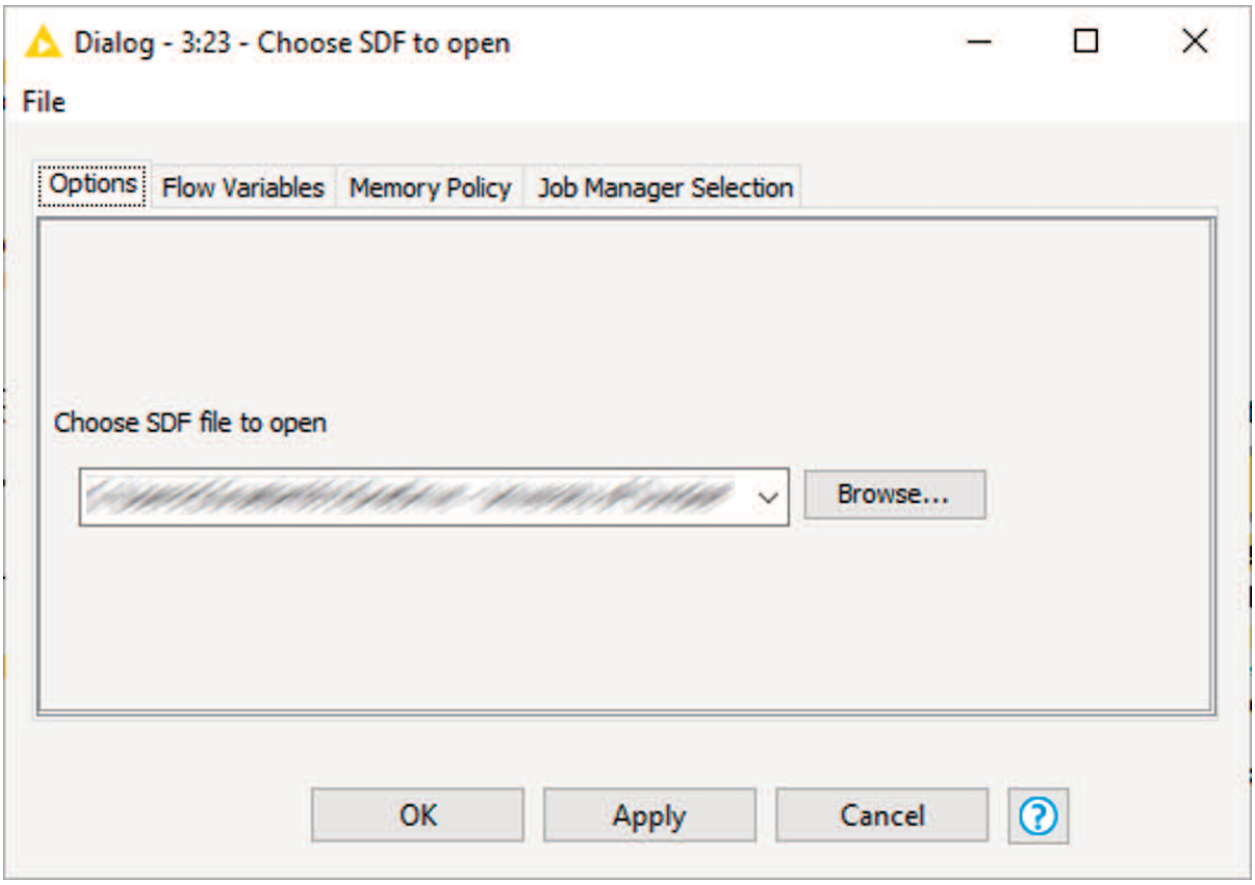


Figure 2 *Choose SDF to open" configuration window*

A warning message will appear and due to the change in data file, the node will have to be reset.

Click "OK".

Right click "Choose SDF to open" node > "Execute".

2.2 Changing the Split Compound Catalogue node

From the Reaction Enumeration exercise, the Split node is configured to split the Sigma-Aldrich catalogue into two tables: amines and carboxylic acids. If you're intending to perform other reactions, you will need to change the internals of this node.

On the "Split Compound Catalogue" node, right click on it, choose "Component" and "Open" (**Figure 3**).

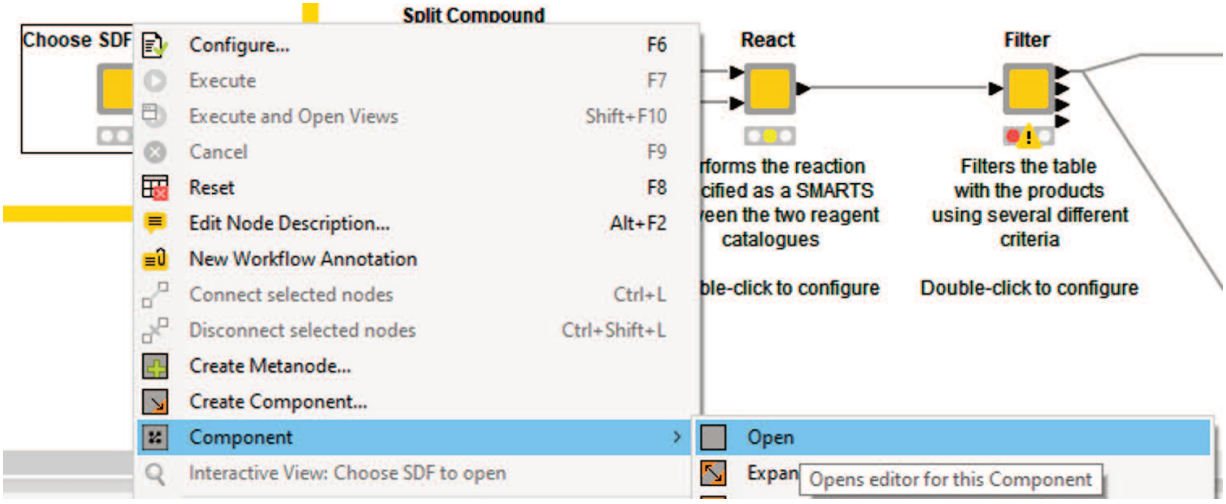


Figure 3 *Component context menu highlighting the 'Component' options*

A new tab will open on the KNIME Main pane (**Figure 4**).

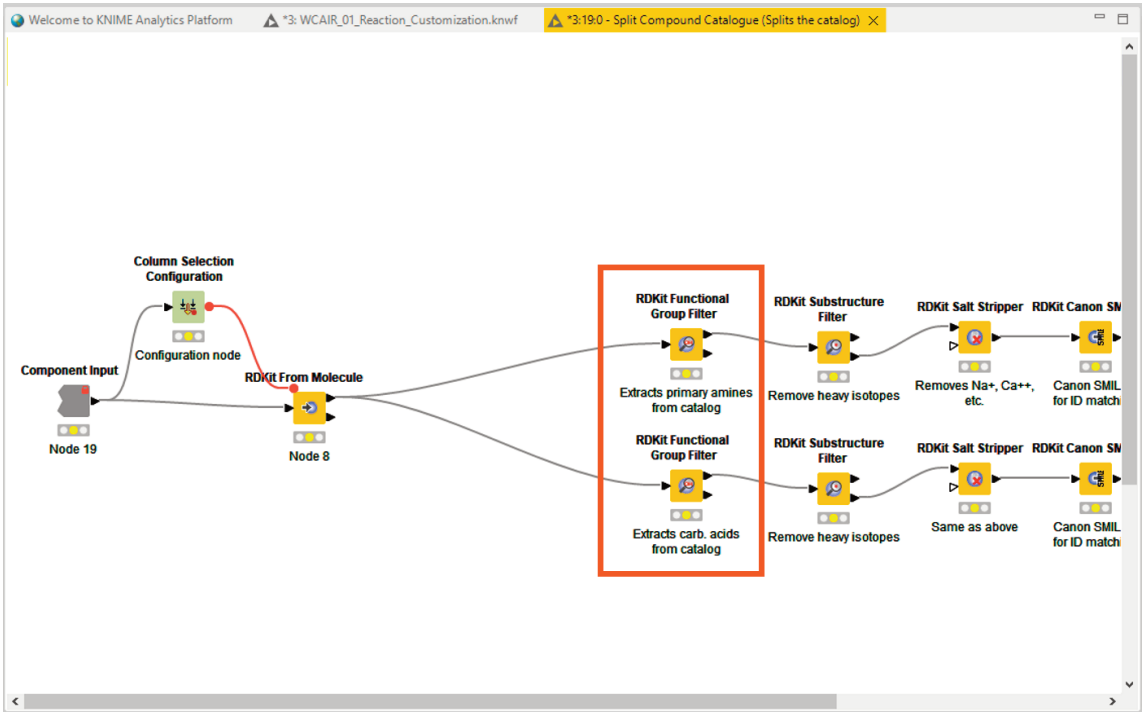


Figure 4 *Split Compound component tab. RDKit Function Group Filters have been highlighted in orange.*

You are now seeing the internals of the “Split Compound Catalogue” node.

You will need to execute the first three nodes as previously described: “Component Input”, “Column Selection Configuration” and “RDKit from Molecule”.

We will now focus on only “RDKit Functional Group Filter” node. You will notice in **Figure 4** that two Group Filter nodes are highlighted. One node is performing the filtering of the amines and the other node is filtering the carboxylic acids.

The provided customized table (“customization-example.sdf”) contains carboxylic acids and alcohols. One of the filtering nodes (**Figure 4**) is already filtering for carboxylic acids. The other one – currently filtering for amines – needs to be reconfigured.

Double click on the top node that filters for primary amines. A dialog box will appear (**Figure 5**).

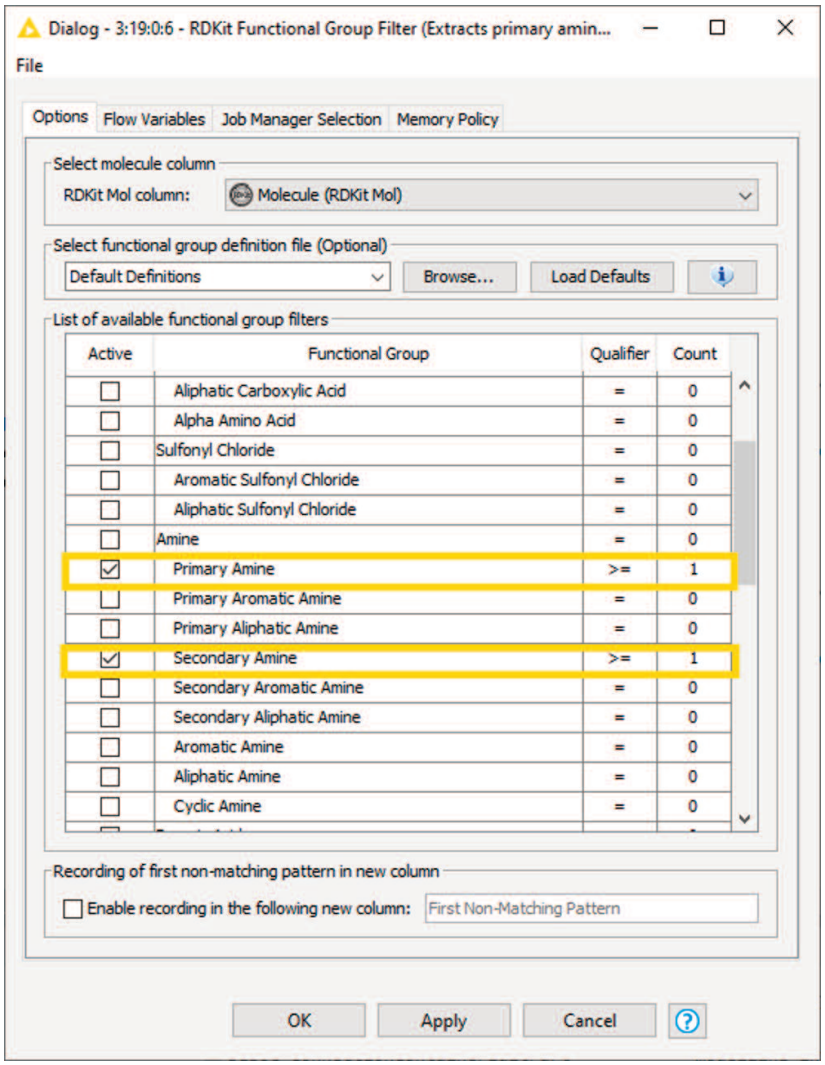


Figure 5 *RDKit Functional Group Filter configuration window configured to filter for primary and secondary amines*

Here you will change only the list of functional group filters. The figure shows the current configuration options for primary and secondary amines. We will now be filtering for alcohols.

First, all options will need to be reset.

Unselect “Primary Amine” and “Secondary Amine”.

For these two options, change the “Qualifier” to “=” and “Count” to “0”.

To select alcohols: Select “Alcohol”, change the “Qualifier” to “>=” and the “Count” to “1” (**Figure 6**).

Click “OK”.

You have now instructed the filter node to look up all the structure in the stable and keep only the ones that have alcohols.

Repeat the above procedure if you want to filter for other functional groups.

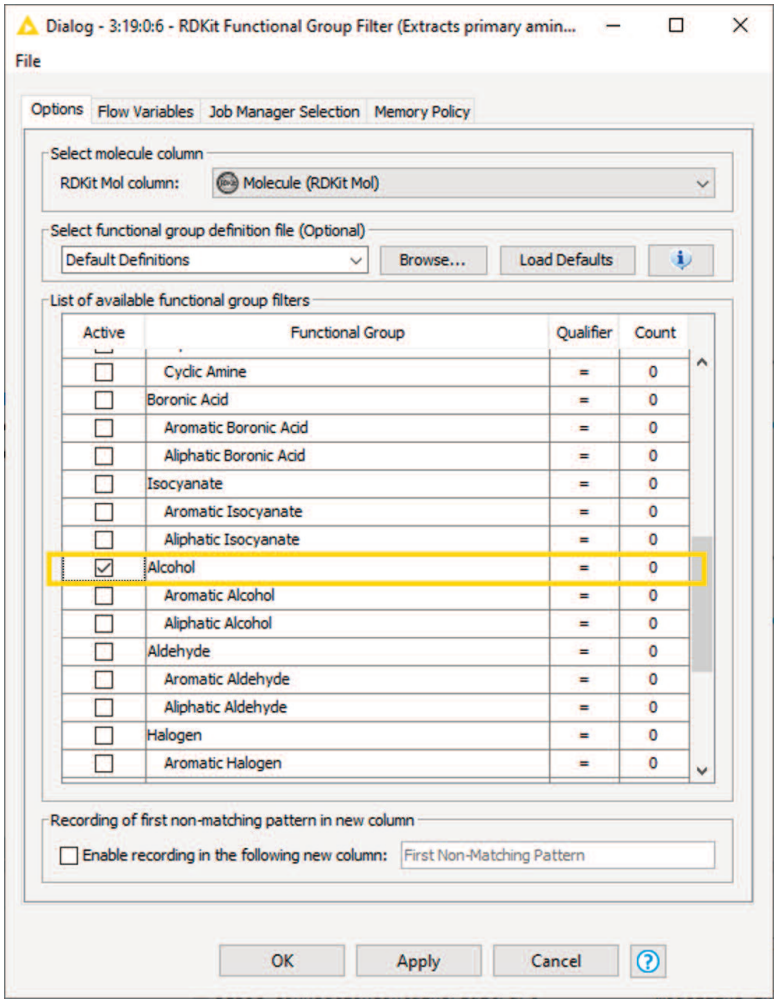


Figure 6 *RDKit Functional Group Filter configuration window*

Close the component tab to return to the main workflow tab.

Save the file ("File" > "Save").

Execute the "Split Component Catalogue" node to perform the filtering operations.

2.3 React using SMARTS?

The next node, "React", will require a customised SMARTS string to perform your customised reaction.

SMARTS – or Smiles Arbitrary Target Specification – is a language used for describing molecular patterns and properties, that specifies which kind of molecules it will try to match to a given reaction or structural search. You can learn more about the SMARTS language here. You can create your own SMARTS strings using the Marvin JS demo, saving your reaction scheme as SMARTS, and copying the result code. You can also use the smarts.plus webservice to check if your SMARTS string is valid.

You have been provided (**Table 1**) four SMARTS strings to four common two-component reactions for your own use.

Table 1 *SMARTS strings for common organic reactions*

Reaction	SMARTS strings
Esterification using a carboxylic acid and alcohol	[#6:1][OH:2].[OH][#6:3]=[O:4]>>[#6:1][O:2][#6:3]=[O:4]
Suzuki coupling using a boronic acid and an aryl halide	[#6;a:2][#17,#35,#53].[#6:1]-[#5](-[#8])-[#8]>>[#6:1]-[#6;a:2]
(aliphatic) Nucleophilic substitution between an amine and an alkyl halide	[N;!H0:1]-[#6:2].[#6;A:3]-[#17,#35,#53]>>[*:2]-[N:1]-[*:3]
Aromatic Nucleophilic substitution between an amine and an aryl halide	[#7;!H0:1]-[*:2].[#6;a:3]-[#17,#35,#53]>>[#6;a:3]-[#7;!H0:1]-[*:2]
An amide or sulphonamide coupling using primary amines and either carboxylic acids, acyl chlorides or sulphonyl chlorides (this is the original example used in the workflow walkthrough)	[#7;!H0!H1!$([#7]-C=[O,N,S])!$([#7]~[!#6]):1].[#8H,#17;A][#6,S:2]=[O:3]>>[#7:1]-[#6,S:2]=[O:3]

In this example we are performing an esterification.

Copy the SMARTS string from row 1 from **Table 1.**

In KNIME, double click the "React" node.

Paste the new SMARTS string into the "Enter reaction SMARTS" text box (**Figure 7**).

Deselect "Diversity Picker" as in this example, the reagents table does not contain enough molecules to react.

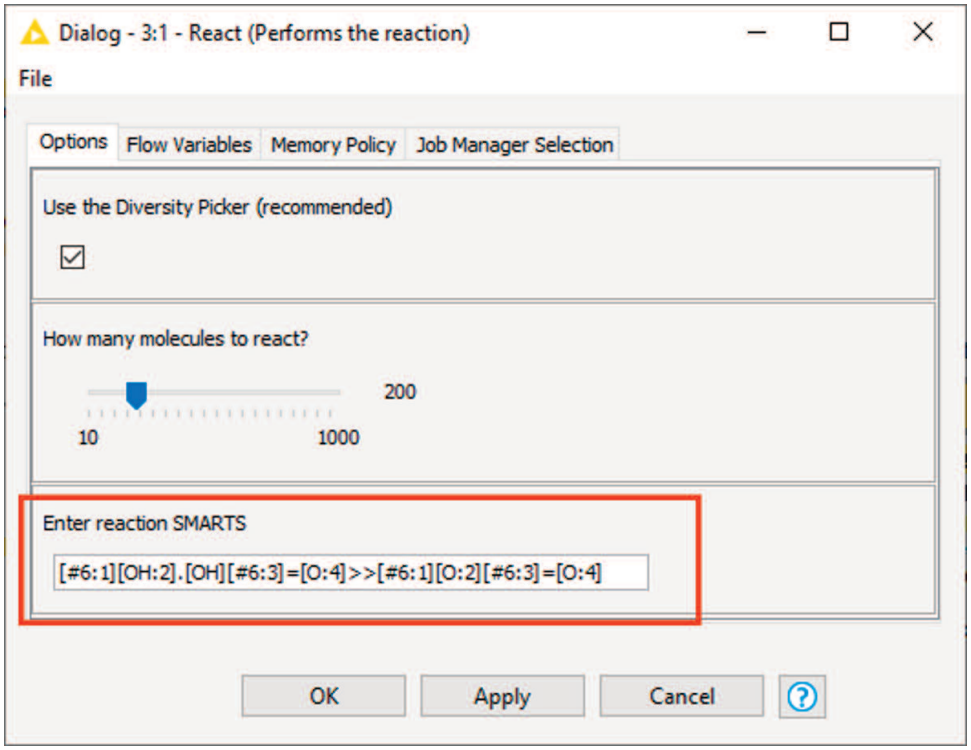


Figure 7 *React node configuration window.*

Press OK to close this window and execute the node.

2.4 Filtering and producing compounds

The "React" node has reacted our list of alcohols and acids together to form esters. To view this, right click "React" node > "Products". A table will now appear with all the reacted compounds.

How many esters have been calculated?

After the execution of the reaction node, you can configure and execute the Filter node to filter the reaction results using the pre-defined rules.

Continue executing all nodes and view the results as previous described (*WCAIR Reaction Enumeration*).

2.5 Using your own data tables

Repeat the above procedure if you want to filter for other functional groups.

Please note that the order of reagents in your SMART strings must match your order of tables in the react node. The example given was alcohols followed by acids.

Each reagent in the SMART string is delimited by a dot (“.”). The reaction arrow is represented as two ‘greater than’ signals together (“>>”). If you’re performing an esterification, the order in the SMART string (alcohols first, then carboxylic acids) should be the order of the two input tables in the “React” node (i.e.: first input must be the alcohol table). If the order of your tables is not matching, edit your SMART string as in the box below:

Using the esterification reaction example, let’s see its SMARTS code:

`[#6:1][OH:2].[OH][#6:3]=[O:4]>>[#6:1][O:2][#6:3]=[O:4]`

This could be translated as:

Alcohol + Carboxylic acid → Ester

What if, in your workflow, the table with carboxylic acids is the first input port and the alcohol table is the second?

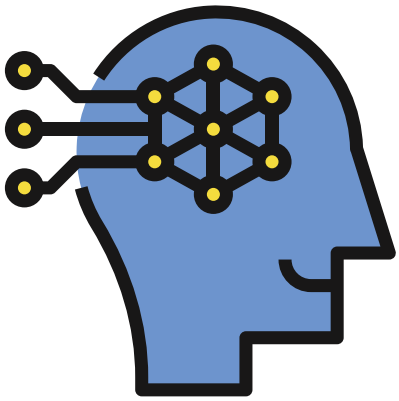
A quick fix to this would be to edit the SMARTS code as the following:

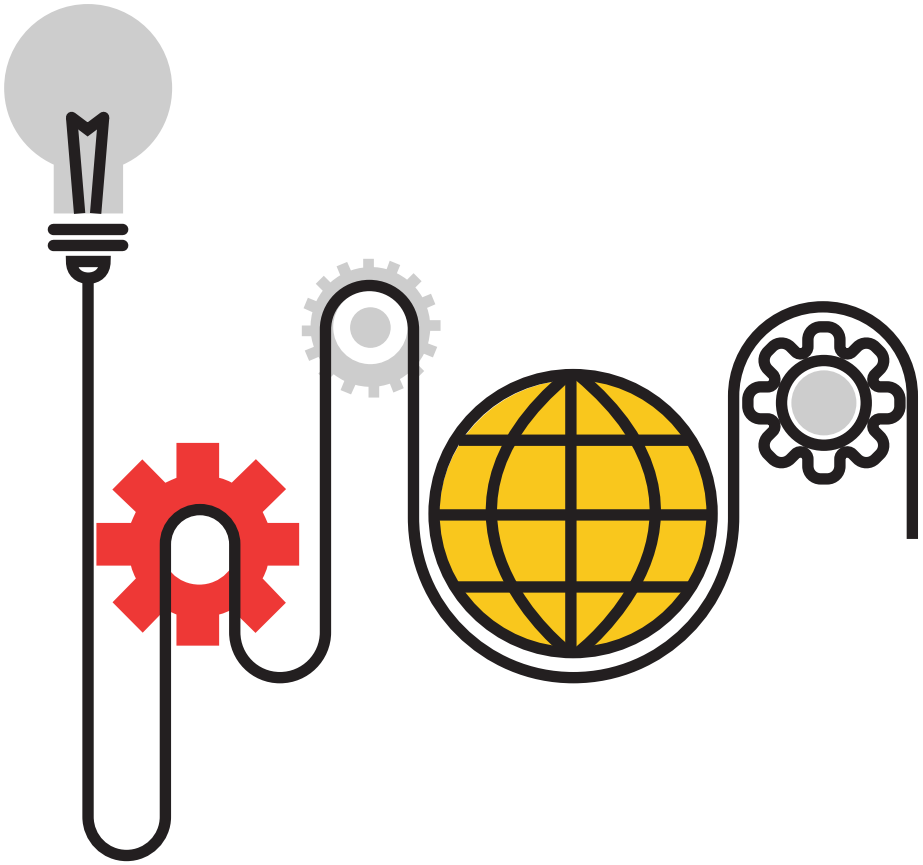
`[OH][#6:3]=[O:4].[#6:1][OH:2]>>[#6:1][O:2][#6:3]=[O:4]`

So now the order of the reagents matches the order in the SMARTS code.

And now...

you’ve learned how to modify and configure this workflow to run other two-component reactions of your choice, using your own molecule catalogue!





Wellcome Centre for Anti-Infectives Research
School of Life Sciences
University of Dundee
Dow Street
Dundee DD1 5EH

WCAIR@dundee.ac.uk

www.wcair.dundee.ac.uk

@WCAIRDundee

University of Dundee

wellcome centre anti-infectives research