FAQ – RetroPath2.0 & rp2paths

Authors	Setareh Tasdighian Thomas Duigou
Last update	2021/01/28

Can I use an empty filed for compounds without INCHI?

No, you must provide 2 fields per line in the `sink.csv` file. If a compound has no InChI, the second field should be present, but empty: `""`. In general, using the compounds with no InChI is not advised, because it suggests that the compound is not known. Moreover, RetroPath eliminates such compounds from the sink.

Should I focus on reactions that have EC classifications?

This is not advised because there are well-annotated reactions without ECs, also in the biological domain.

Are the “double quotation” marks necessary in the sink file?

Not necessary, however, the string for some names and InChIs may include a comma, which is the delimiter character in the .csv file. As such, it is best to embrace every field with double quotations.

If I set the pathway length to 10, does the tool find only pathways of this length?

No, the tool tries to find any pathway with length 1 to 10.

Does the workflow stop if the source compound is found in the sink?

Yes, it does, and the output is printed in the sourceinsink.csv file in the output folder. You should remove any compound with the same InChI as the source from the sink in advance. Alternatively, you can run the workflow and use the `source-in-sink.csv` file to remove the compounds with similar InChI. The second solution will take longer if you have large sink.

If I leave the sink empty, I get an error and the workflow does not start?

Several users report this issue. Best to provide a file with only a header for sink.csv, rather than leaving the field empty.

Why do I not see a conventional text-book pathway among my enumerated pathways for a known source?

One possibility is that you did not enumerate enough pathways. rp2paths enumerated only the top 150 pathways by default. If you want to have a look at other pathways, you should increase it by default.

How were the rules set designed for the tutorial examples?

To reproduce the known pathways of the example compounds, the reaction rules with the desired ECs were selected to be included in the rule set.

What does the co-factor list provide in the GUI?

Nothing as far as reaction rules provided by RetroRules (https://retrorules.org) are concerned (i.e. mono-component reaction rules). Notice that cofactors (such as ATP, NADP, …) are not eliminated from compounds outputted by RetroPath.

If I use the pre-parsed rules which were generated for diameter 2 to 16, does it make any difference to set the diameter value to higher value in the GUI?

No. The max-diameter would be 16. However, if you set the max to a lower value, for example to 10, it only applies the rules up to diameter 10.

Does the red cross on the *Core* node always mean that the workflow has failed?

No, it is possible that the `.svg` graphs are not produced due to incompatibility with user IDs, for example if the compounds are not using MNX-IDs, the tool cannot visualize them. However, you should always look inside the output folder for the `scope.json` and `scope.csv` files, which are the main output.

Where do I get those MNX IDs and is there a mapper included in the tool?

You get the MNX IDs from MetaNetX website and you should write your own parser to map your IDs to those of MNX if you wish to see the structures.

Does using the tool make me eligible to use the data?

No. For instance, if you are using MetaNetX data (https://www.metanetx.org), you should comply with the license agreements stated from it (see https://www.metanetx.org/mnxdoc/mnxref.html). RetroRules

come with a separate license as well. As a rule of thumb, the tool licenses are often dissociated to the licenses of the data used as input.

What are the CMPD IDs in the output of rp2paths?

These are “dummy” compounds IDs generated by rp2paths. Any compounds that does not match a sink compound will have such CMPD prefix. The reason compounds do not have a real ID is that rp2paths does use at this stage to retrieve it.

Are the enumerated pathways the result?

No, there is still a “completion step” needed. RetroPath 2.0 applies the reaction rules to one compound at a time for bi-substrate reactions. As such, some of the compounds are not present in the reactions of the pathways (output of rp2paths). As such, these compounds must be “put back”. This operation is now implemented in the rptools package (https://github.com/brsynth/rptools), and can be performed using the SynBioCAD Galaxy portal (https://galaxy-synbiocad.org) : use the “Complete Reactions” tool (from “SynBioCAD RetroSynthesis” category) and provides RetroPath2.0 and rp2paths outputs.

Is it possible to run this workflow on a paid KNIME-server? If so, how should the input and output nodes be re-configured?

Although not been tested, this is very likely to be possible.

When would the completion step be added to the tool?

The completion step has been developed as a separated tool, included into the rptools package (https://github.com/brsynth/rptools).

How to run the workflow for multiple compounds in parallel, while the workflow reads the same input files from the same location on one instance of KNIME running on a Linux machine?

Although feasible, it is preferable to execute the workflow in a sequential way because of the heavy memory consumption in RetroPath2.0. To automatize executions, one could use the batch mode. Here below an example:

```
# Helpful paths
KPATH="/Applications/KNIME.app/Contents/MacOS/knime"  # Mac OS example
# KPATH="/opt/knime/knime"                          # Linux example
RP_PATH="/Users/jdoe/Downloads/RetroPath2.0”
```

```
# Make output folder if needed
mkdir -p "${RP_PATH}/res”

# RetroPath2.0 call
$KPATH \
-nosplash \
-nosave \
-reset \
--launcher.suppressErrors \
-application org.knime.product.KNIME_BATCH_APPLICATION \
-workflowFile="${RP_PATH}/RetroPath2.0.knwf" \
-workflow.variable=input.dmin,0,int \
-workflow.variable=input.dmax,1000,int \
-workflow.variable=input.max-steps,5,int \
-workflow.variable=input.sourcefile,"${RP_PATH}/tutorial_data/carotene/source.csv",String \
-workflow.variable=input.sinkfile,"${RP_PATH}/tutorial_data/carotene/sink.csv",String \
-workflow.variable=input.rulesfile,"${RP_PATH}/tutorial_data/carotene/rules.csv",String \
-workflow.variable=output.dir,"${RP_PATH}/res",String \
-workflow.variable=output.solutionfile,"result.csv",String \
-workflow.variable=output.sourceinsinkfile,"source-in-sink.csv”,String
```

The code available at https://github.com/brsynth/retropath2-wrapper proposes a wrapper around KNIME and RetroPath2.0 to facilitate the use through a terminal.

Is it possible to optimize the run-time of the workflow using parallel computations inside the workflow nodes?

This is not available in RetroPath2.0.

The visualizations in Scope Viewer for larger scopes is not helpful. Are there any plans to improve the tool?

Yes, this is now implemented in “rpviz”, a tool to visualize individual pathways that lies in the scope. See: https://github.com/brsynth/rpVisualiser .

Would it be possible to also visualize the enumerated pathways on top of the scope graph within the new version of Scope Viewer?

Indeed, this is what rpviz does :)