KNIME workflow for assisting analogue selection (KWAAS) for PBK modelling

Information from a PBK model for a chemical rich in data (source chemical) can be used to create a PBK model for a similar chemical lacking data (target chemical). This document explains how to use the KNIME workflow that was created to help determine similar analogues with available PBK models. A flow chart of the steps taken to use the tool and determine the most similar chemicals from the PBK dataset is shown in Figure 1. In the examples here, refinement of analogue selection was based on molecular weight (MW), logarithm of the octanol:water partition coefficient (log P), distribution coefficient (log D) and relative acidity or basicity (pKa or pKb). Note that the KNIME workflow is adaptable and flexible. A user can upload their own data on physico-chemical properties or absorption, distribution, metabolism or excretion (ADME) data and use these to refine analogue selection. The user can also choose to narrow or broaden the property ranges to decrease or increase the number of analogues identified. The models identified by the workflow can be used either as a template to develop a new PBK model for a target chemical, or provide information to adapt a generic PBK model, rapidly providing a more accurate chemical-specific model. Thus, this will assist more accurate safety assessment of chemicals by providing kinetic information for chemicals of interest.

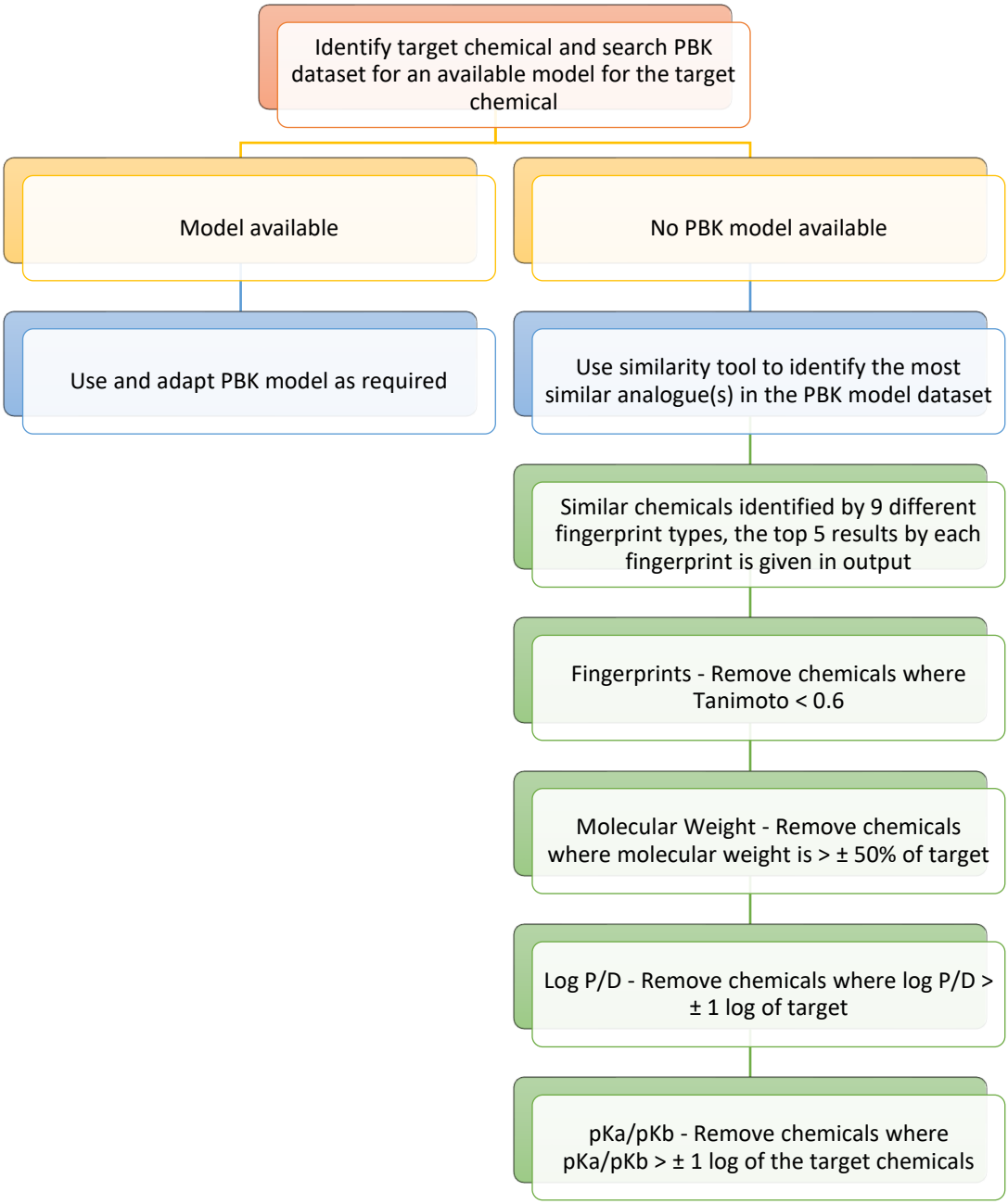


Figure 1 Flowchart of the steps taken to determine the most similar chemicals in the PBK dataset to a target chemical in the example used here, note that alternative properties may be used to refine analogue selection.

Step-by-step guide on using the KWAAS

Instructions on how and where to downloaded KNIME can be found at the below link: https://www.knime.com/downloads.

Key information for using KNIME:

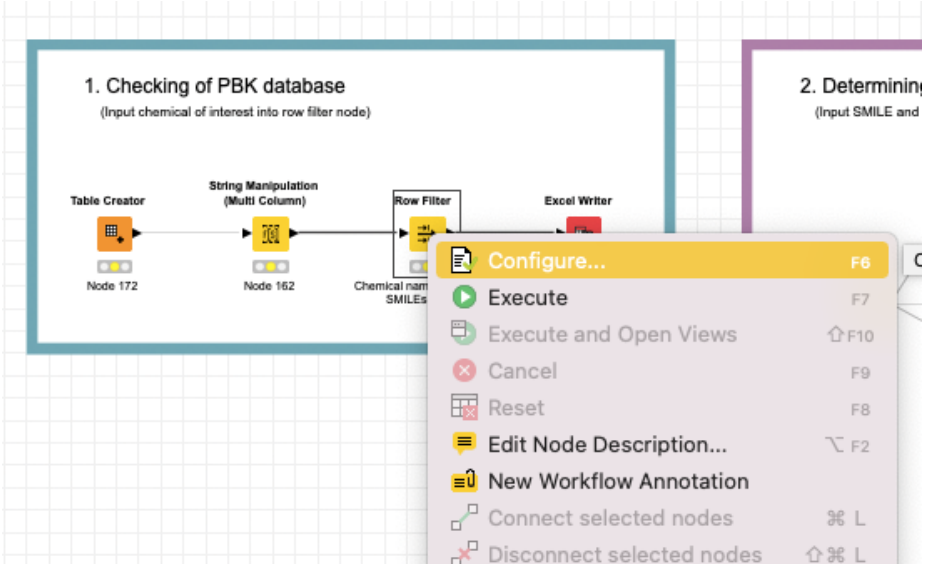
- Red light under node – node needs to be configured

- Yellow light under node – node is configured, and needs to be executed
- Green light under a node – node executed
- Right click any node to get a list of options of commands, including configure, execute and an option to view results.

Stage 1 – Checking the PBK model dataset for an existing model for a target

To check the PBK model dataset to determine if there are any PBK models currently available within the dataset for the target chemical of interest.

Right click and select configure on the 'row filter' node.

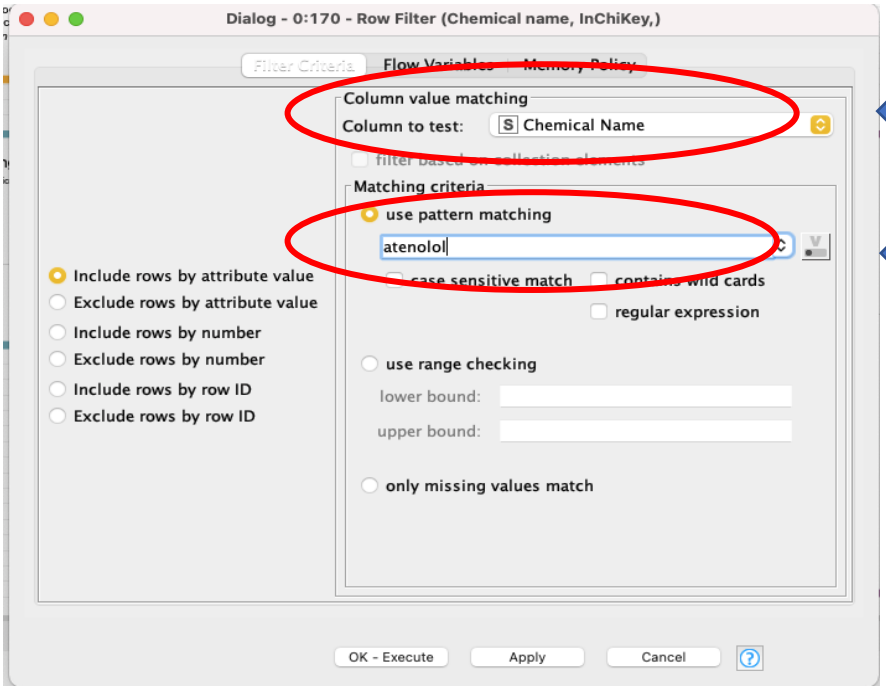


Choose a column option to search the dataset, either by Canonical SMILES, InChiKey or chemical name. Input the canonical SMILES, InChiKey or chemical name under 'use pattern matching' and click OK, then execute the node.

Example: Flumioxazin

Select InChiKey and input the InChiKEy for flumioxzin (i.e. FOUWCSDKDDHKQP-UHFFFAOYSA-N)

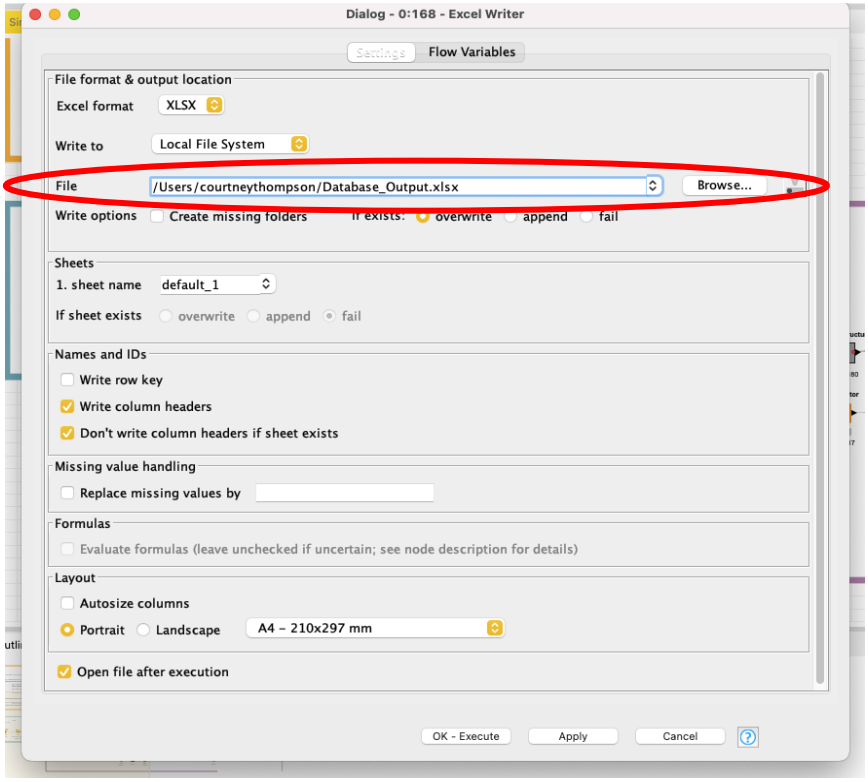
Output should include 2 models for atenolol (1 model for a human and 1 model for a rat).



Enter the term “Chemical Name”, “Canonical SMILES” or “InChiKey” here to select which column is to be searched

Enter the name of the chemical, its SMILES string or InChiKey, according to the column type selected above

To obtain the results in an Excel spreadsheet, right click and select configure on the ‘Excel writer’ node

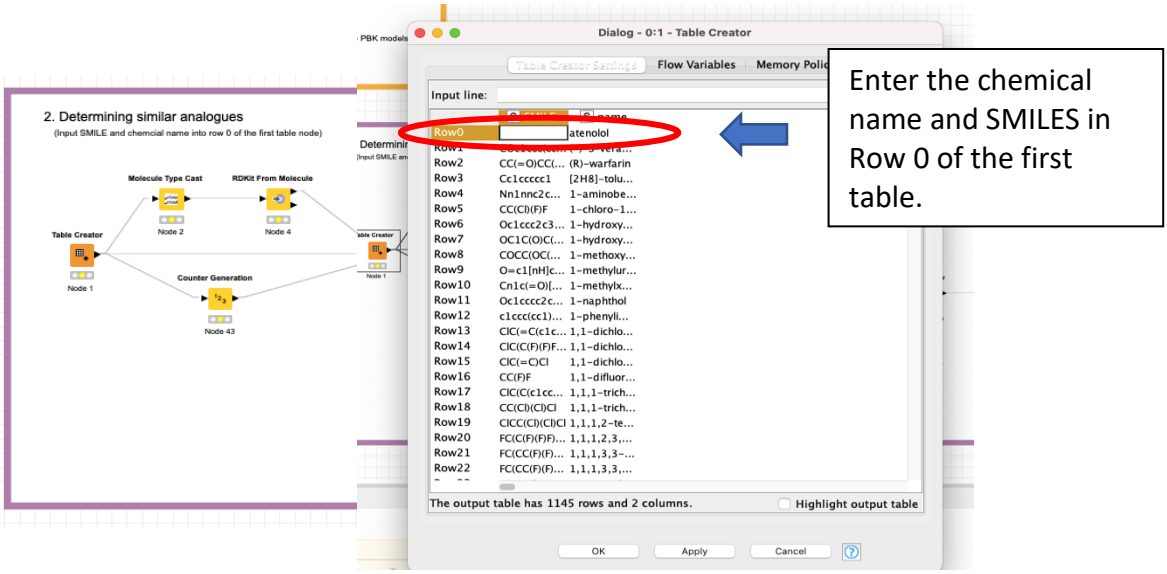


Under 'file' select the location where you want to save your excel file before executing the node.

Stage 2 – Finding similar analogues to a target chemical

Similar chemicals to a target chemical may be identified using chemical fingerprints and/or physico-chemical properties. Nine chemical fingerprints, Morgan, FeatMorgan, AtomPair, Torsion, RDKit, Avalon, Layered, MACCS, and Pattern are all used to calculate similarity between chemicals in which a score is given. Excel output can be obtained at every step of the workflow (follow steps in stage 1 on how to configure the Excel node).

Configure the first table creator by inputting the Canonical SMILES and chemical name in the first row for the target chemical.

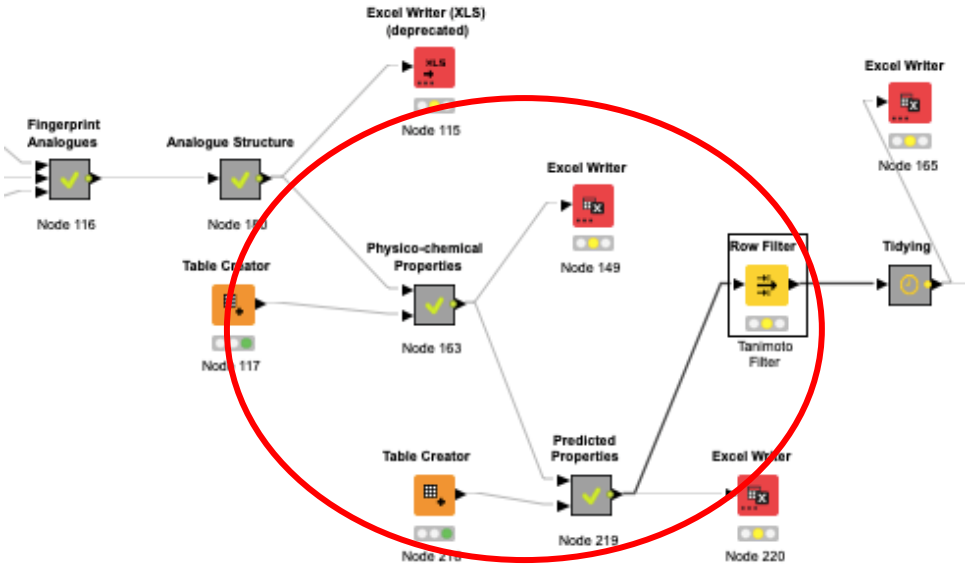


Example: Flumioxazin

Input under SMILES column: C#CCN1C(=O)COC2=CC(=C(C=C21)N3C(=O)C4=C(C3=O)CCCC4)F

Input under chemical name column: flumioxazin

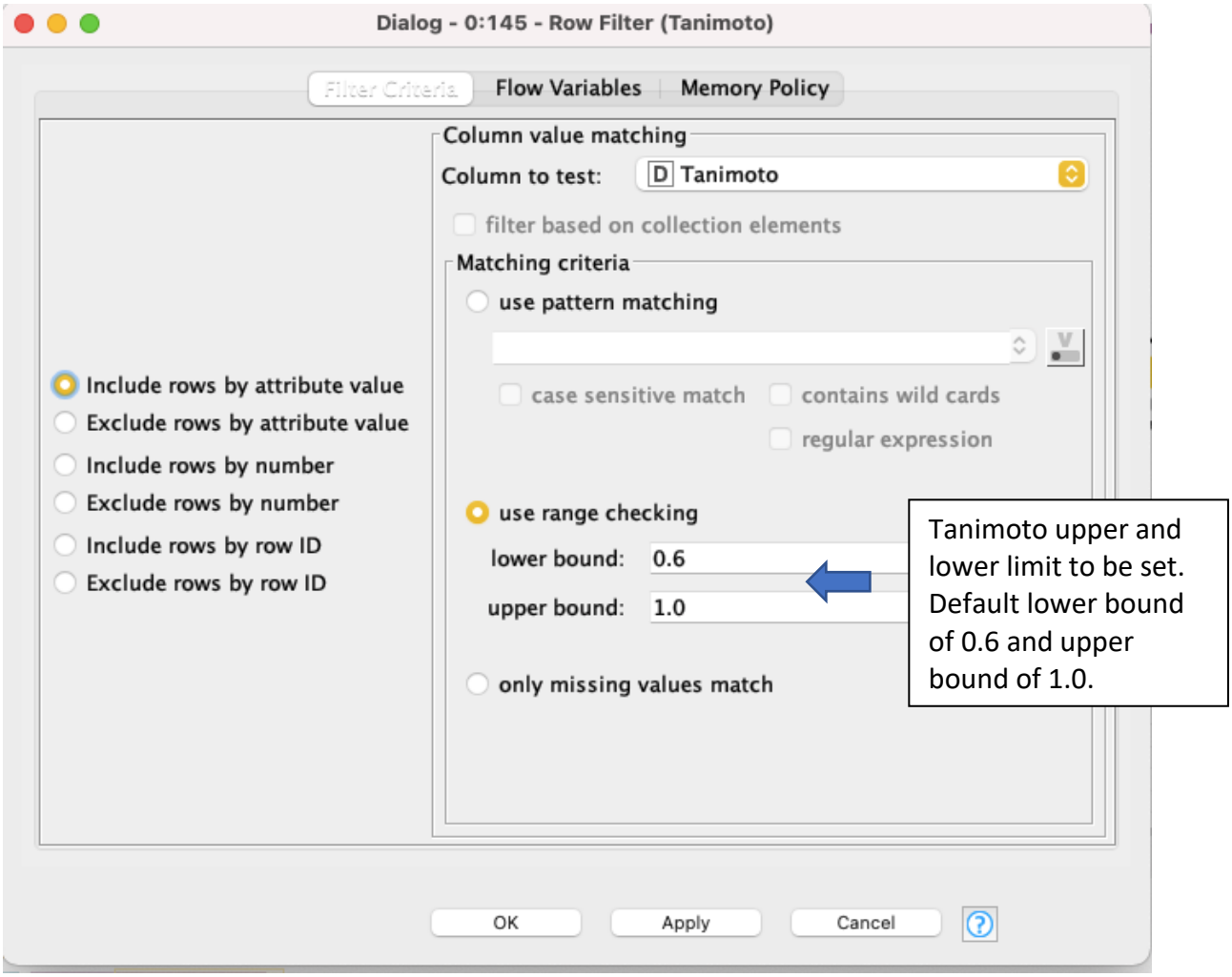
Execute the metanode ‘fingerprint analogues’ to obtain the analogues determined as being similar according to chemical fingerprints. The next three metanodes in the sequence, will provide the structures, physicochemical properties (molecular weight, log P, hydrogen bond donors and acceptors, TPSA (topological polar surface area), and the number of Lipinski rule violations) and predicted properties (pKa, pKb, log D at pH 5.5, log D at pH 7.4, and CL_{int}) for each suggested analogue when executed.



All nodes up until the first ‘row filter’ node can be executed without configuration and excel output obtained at each stage. A Tanimoto index for similarity is generated giving a score between 0 and 1, where 0 is not similar and 1 indicates an identical chemical. The row filter node named ‘Tanimoto filter’ is set at default to include all analogues of a with a Tanimoto index of 0.6 or higher. However, this Tanimoto value can be set by the user.

Example: Flumioxazin

After filtering by Tanimoto > 0.6, there should be 25 rows of results consisting of 17 chemicals, for this example.



Refining analogue selection by Molecular Weight

Once a set of suitable analogues has been identified these can be further refined. The 'MW filter' row filter node sets upper and lower boundaries to include analogues that fall within those limits.

It is recommended to set a lower bound of 50% less than the target chemical's molecular weight and an upper bound of 50% greater than the target chemical's molecular weight (although the range can be set to any value of the user's choice).

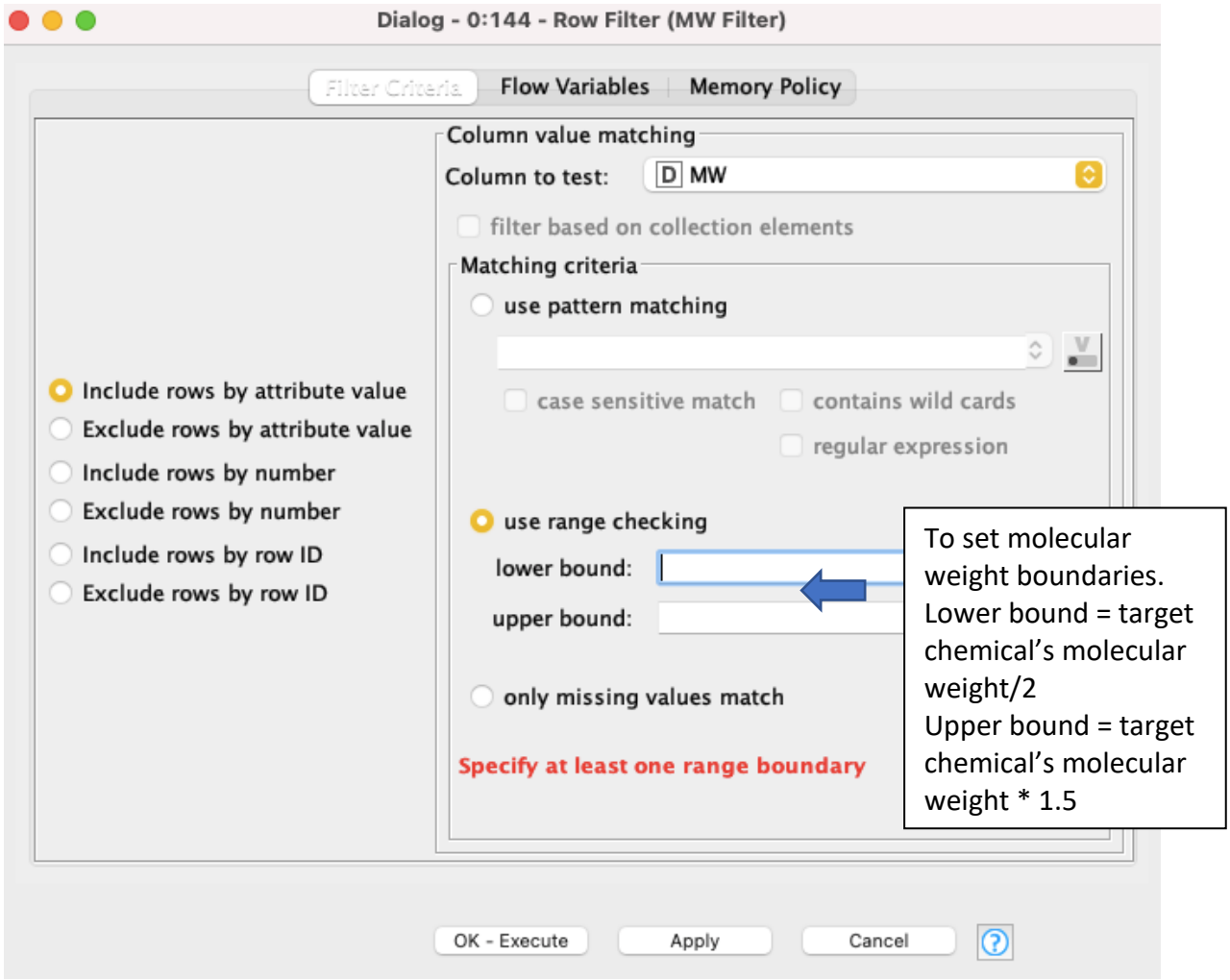
Example: Flumioxazin

Molecular weight for flumioxazin: 354.3 Da.

Lower bound input: 177.15

Upper bound input: 531.45

After filtering there should be 20 rows of results consisting of 12 chemicals.



Refining analogue selection by log P/log D

Results can be refined further to find the most similar chemicals. The 'log P filter' and 'log D filter' row filter node sets upper and lower boundaries to include analogues that fall within those limits.

A log D at pH 5.5 or pH 7.4 needs to be chosen before inputting boundaries to obtain results within the specified range. It is recommended to set a lower bound of 1 log less than the target chemical's log P/log D and an upper bound of 1 log greater than the target chemical's log P/log D (although the range can be set to any value of the user's choice).

Example: Flumioxazin

If choosing to filter by log P. Log P for flumioxazin: 1.9281

Lower bound input: 0.9281

Upper bound input: 2.9281

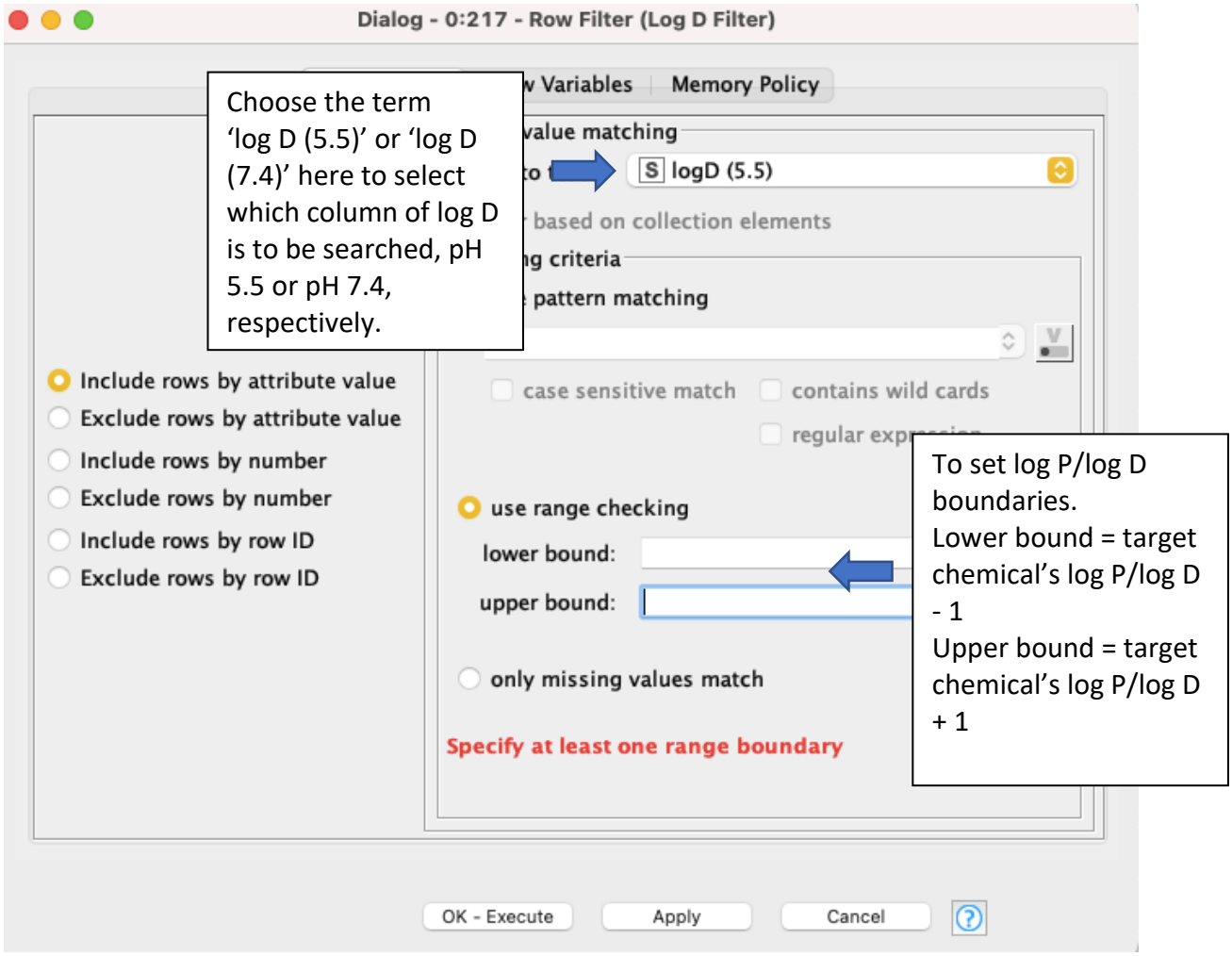
After filtering there should be 15 rows of results consisting of 7 chemicals, for this example.

If choosing to filter by log D. Log D (at pH 5.5) for flumioxazin: 2.55

Lower bound input: 1.55

Upper bound input: 3.55

After filtering there should be 13 rows of results consisting of 5 chemicals, for this example.



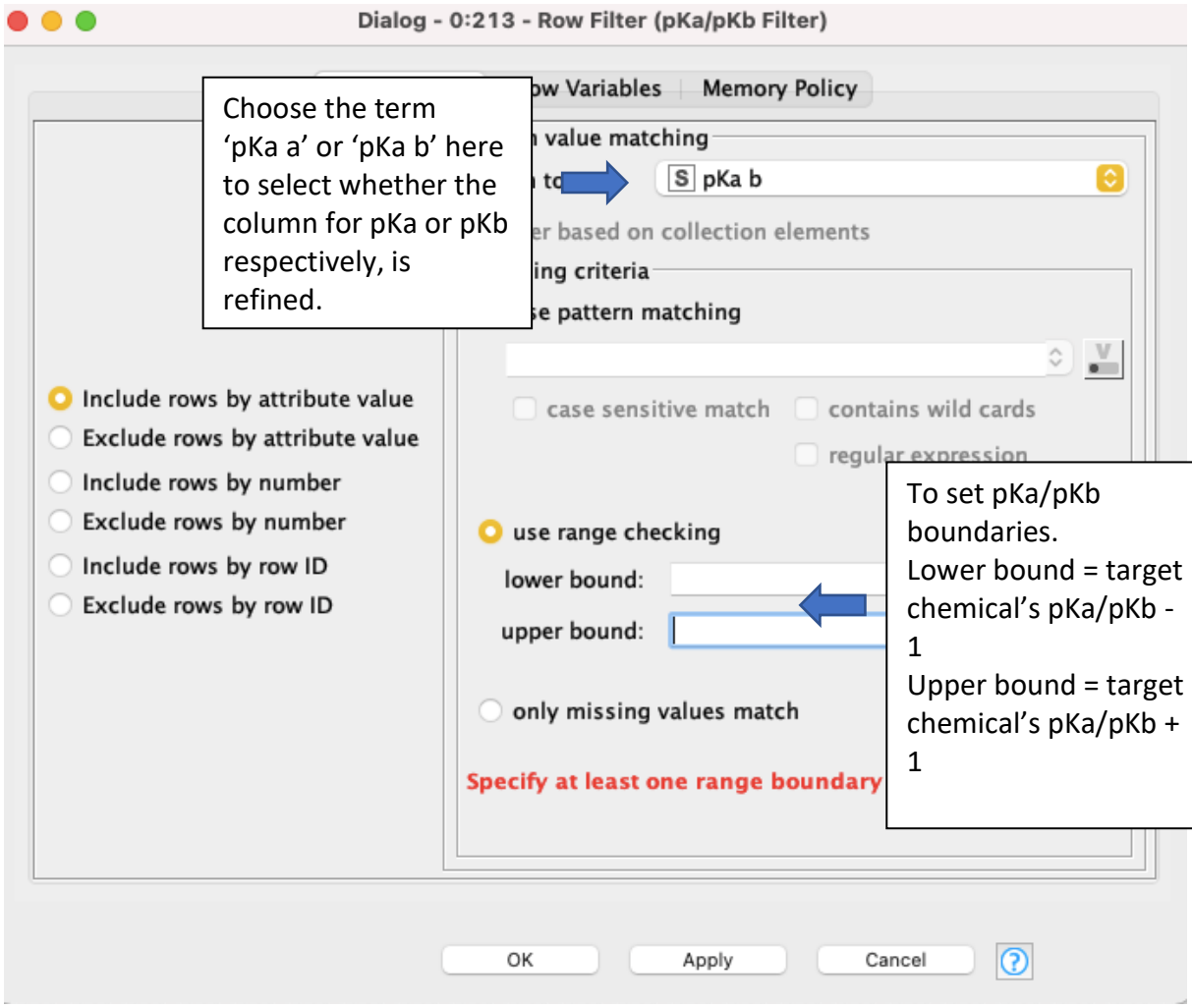
Refining analogue selection by pKa/pKb

Results can be refined further to find the most similar chemicals. The 'pKa/pKb filter' row filter node sets upper and lower boundaries to include analogues that fall within them limits.

It is recommended to set a lower bound of 1 log less than the target chemical's pKa/pKb and an upper bound of 1 log greater than the target chemical's pKa/pKb (although the range can be set to any value of the user's choice).

Example: Flumioxazin
pKb for flumioxazin: 3.31
Lower bound input: 2.31
Upper bound input: 4.31

After filtering there should be 10 rows of results consisting of 2 chemicals, for this example.

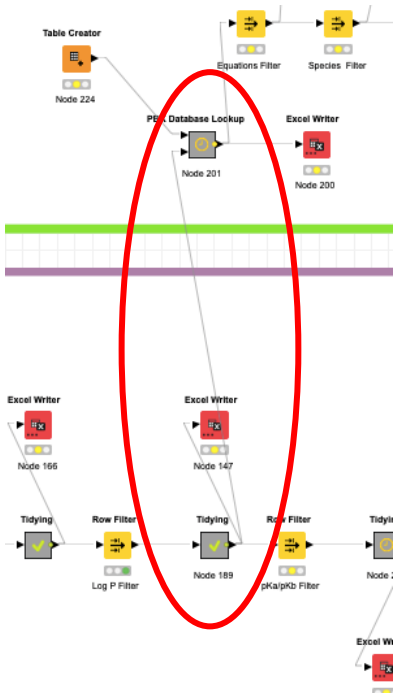


Generating the Excel output after employing these filters will result in an Excel file containing the analogues in the PBK model dataset that are considered most similar to the target.

Similarity after all these steps has been determined in terms of the Tanimoto score for nine chemical fingerprints and molecular weight, log P, pKa and log D falling within set boundaries.

Stage 3 – Refining selection of models from the dataset for chemicals determined to be similar

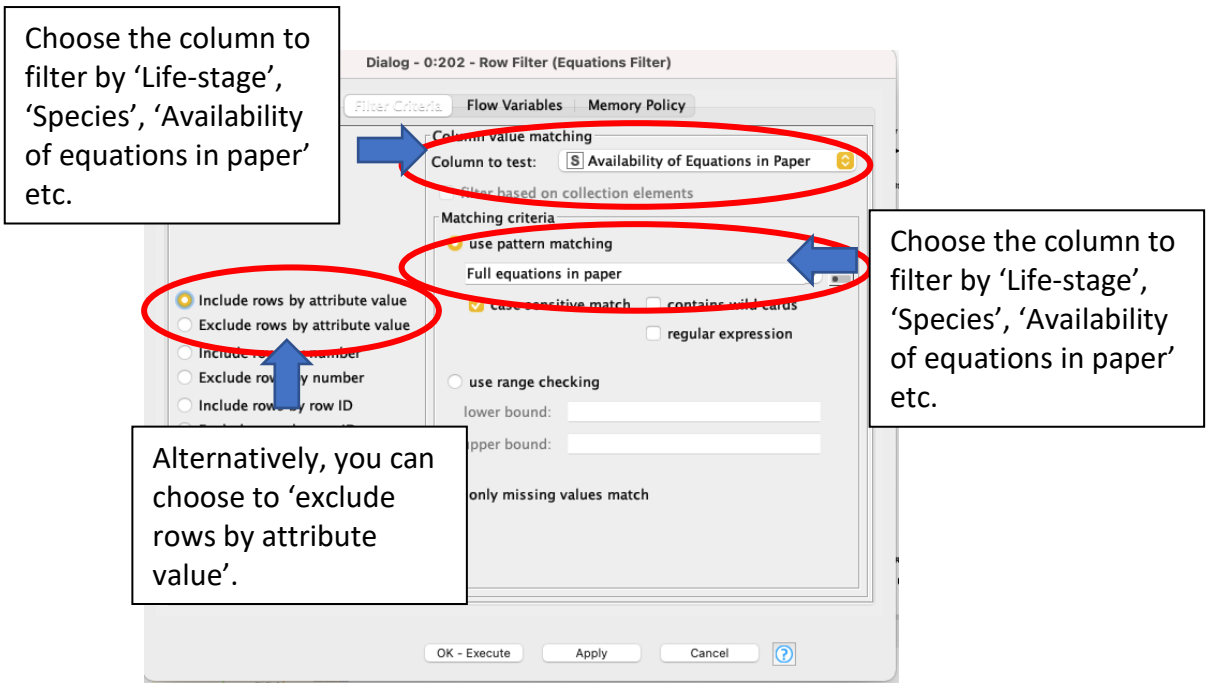
The PBK model dataset can be searched again, this time searching for all analogues that have been identified as being similar to the target and refining to show the analogue models most specific to the user's requirements. Excel output can be obtained at each step of filtering (again follow the steps in stage 1 as to how to configure the Excel node). Output can be exported at any stage to investigate the models available for the identified chemicals to see if these models meet the user's requirements.



Connect the last row filter node used within determining similarity (stage 2) to the 'PBK model dataset lookup' metanode (in stage 3).
This can be done by clicking the arrow coming out of the last node used and dragging it to the bottom input arrow of the PBK model dataset lookup' metanode.
Executing this metanode will result in all available models associated with the analogues determined as being similar at the last filter stage.

Each row filter connected to the 'PBK model dataset lookup' metanode can be set and executed as desired. Currently it is set to filter in the order of availability of equations, species, software, life-stage, then route of administration. Filtering the PBK dataset results allows the user to refine the available models in respect of the purpose of the search, e.g. refine the species by human, and life-stage by pregnant to obtain only pregnancy human PBK models.

When configuring a row filter node, first select a 'column to test'. Then under 'use pattern matching' select the statement you would like to only include from the drop-down option.



Alternatively, the user can choose to exclude models for a specific species, sex, life-stage, software or availability of equations. This can be achieved through choosing to ‘exclude rows by attribute value’ and following the same steps as performed for including a specific row.