ELSEVIER

Reaxys® KNIME nodes for KNIME 3.5

KNIME 3.5 and later supported

Reaxys KNIME Nodes v2.2.1 Installation and User Guide

Version 1.2

Elsevier R&D Solutions

Introduction

The Reaxys and Reaxys Medicinal Chemistry Application Programming Interfaces (APIs) programmatic access to the databases of these two Elsevier research solutions. This enables chemoinformaticians, researchers and information professionals to access and integrate the highquality, manually curated content into their existing informatics environment. This means that all data are readily available for further processing and analysis in the users chosen workflow solution.

The APIs allow the querying of substance property data, reaction information, bioactivity data and bibliographic data through factual and keyword searches. Furthermore, the user can perform exact structure, substructure and similarity searches.

Elsevier provides API wrappers to the most common workflow management systems KNIME and Pipeline Pilot. The Reaxys KNIME nodes and Pipeline Pilot components greatly facilitate the query and retrieval process, eliminating any need for implementation development at the customer side and enabling “drag and drop” programming.

The Reaxys nodes and components constitute everything that is needed for instantly accessing and seamlessly integrating Reaxys and Reaxys Medicinal Chemistry content within an existing environment of tools and processes. They enable the rapid and straightforward creation of integration mash-ups that include Elsevier content. New applications can also be rapidly prototyped.

This document describes how to configure the Reaxys KNIME nodes. The nodes are free-licensed software. To use the nodes productively, a valid subscription to Reaxys and/or Reaxys Medicinal Chemistry and to the relevant API is needed.

About KNIME

KNIME is a modern data analytics platform that allows users to perform sophisticated statistics and data mining to analyze trends and predict potential results. Its visual workbench combines data access, data transformation, initial investigation, powerful predictive analytics and visualization. KNIME also provides report generation functions and can automate the application of new insight into production systems.

The KNIME Analytics Platform is open source and available under GPL license. It can be extended with KNIME Commercial Software to include professional support, productivity and collaboration functionality.

The Reaxys KNIME Nodes

The Reaxys KNIME nodes are free-licensed, closed-source software. To use the nodes productively, a valid subscription to Reaxys and/or Reaxys Medicinal Chemistry and to the relevant API is needed.

- With a subscription to Reaxys and the Reaxys API, you have access to all substance data, reaction data and related bibliographic information from the Reaxys Substances, Reactions and Citations KNIME nodes.
- With a subscription to Reaxys Medicinal Chemistry and the Reaxys Medicinal Chemistry API, you have access to all bioactivity data and related target information, substance data and bibliographic information from the Reaxys Bioactivities, Substances and Citations KNIME nodes.
- With a combined license (Reaxys and Reaxys Medicinal Chemistry and the Reaxys API and Reaxys Medicinal Chemistry API) you have access to all data (substance data, reaction data, bioactivity data and related target and bibliographic information) from the Reaxys Substances, Reactions, Bioactivities and Citations KNIME nodes.

The Reaxys Substances KNIME node allows the retrieval of all substance data in Reaxys and/or Reaxys Medicinal Chemistry. This includes substance identification data, substance structures and substance property data.

The Reaxys Reactions KNIME node allows the retrieval of all reaction data in Reaxys. This includes reaction detail (variation) information and reaction structures.

The Reaxys Bioactivities KNIME node allows the retrieval of all bioactivity data in Reaxys Medicinal Chemistry. This includes target and bioassay information.

The Reaxys Citations KNIME node allows the retrieval of all bibliographic information in Reaxys and/or Reaxys Medicinal Chemistry. This includes basic citation information, keywords and abstracts and patent bibliographic data.

Querying and retrieval with the Reaxys KNIME nodes requires a stable internet connection.

The Reaxys KNIME nodes have been developed against KNIME 3.5 and support KNIME versions 3.5 through 3.6.

Installation and updating

To install the latest release of the Reaxys KNIME nodes, please follow one of the procedures described below.

Install the Reaxys KNIME nodes by adding the Reaxys KNIME download site to your software sources

1. Start KNIME, choose Help → Install new software, then click “Add”.

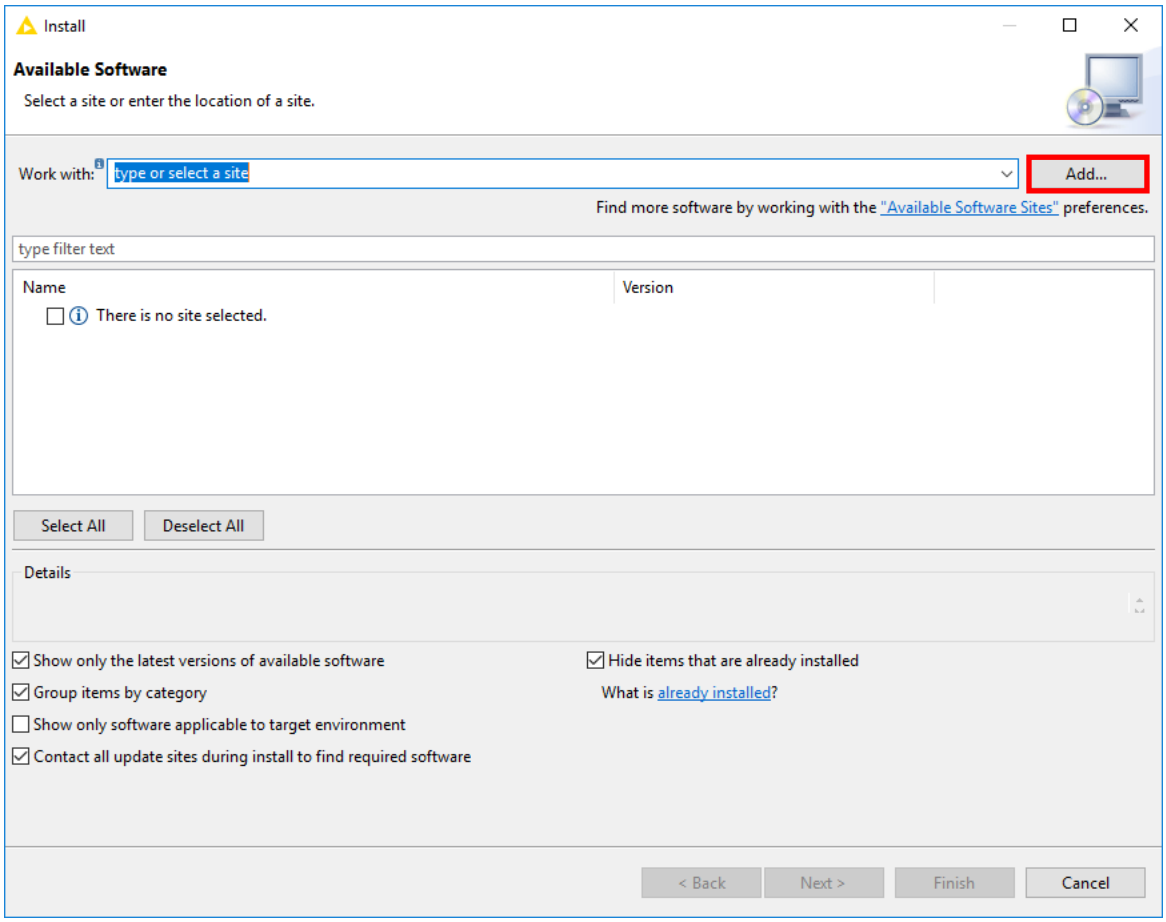


Figure 1. KNIME update site selection

2. Paste the URL of the Reaxys Knime node update site (http://supportcontent.elsevier.com/Support Hub/Reaxys/knime/3.5/) in the “Location” box, then click “OK”.

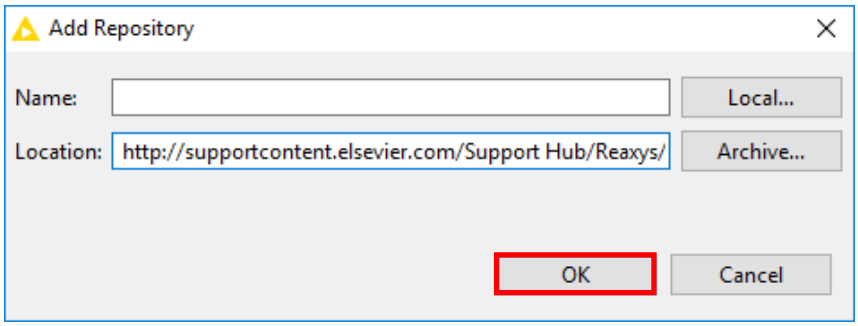


Figure 2. Reaxys KNIME node Location (URL)

After a few seconds, “Elsevier extensions” will appear on the screen. Select them.

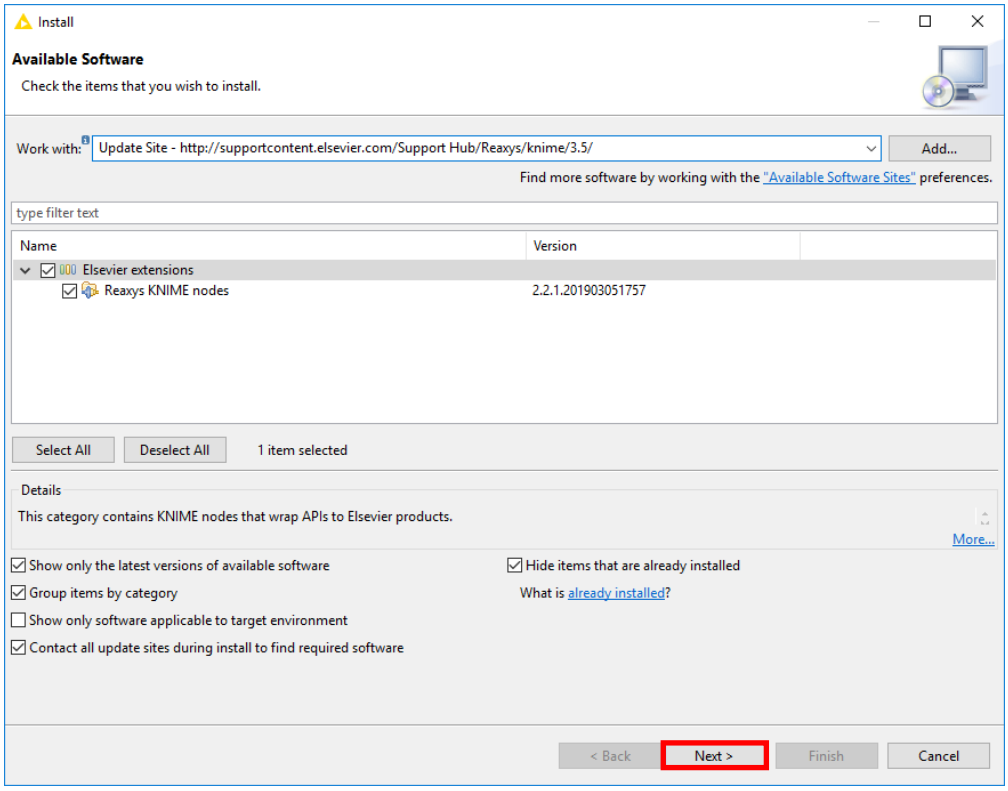


Figure 3. Reaxys KNIME nodes selection

3. Click “Next”, then “Next” again, accept the license agreement, click on “Finish” and restart KNIME to use the Reaxys nodes.

Install the Reaxys KNIME nodes by downloading the zipped Reaxys KNIME download site

1. Download the zipped version of the Reaxys KNIME nodes from http://supportcontent.elsevier.com/Support Hub/Reaxys/knime/3.5/ReaxysKnimeUpdate35.zip

2. Start KNIME, choose Help → Install new software, then click “Add”:

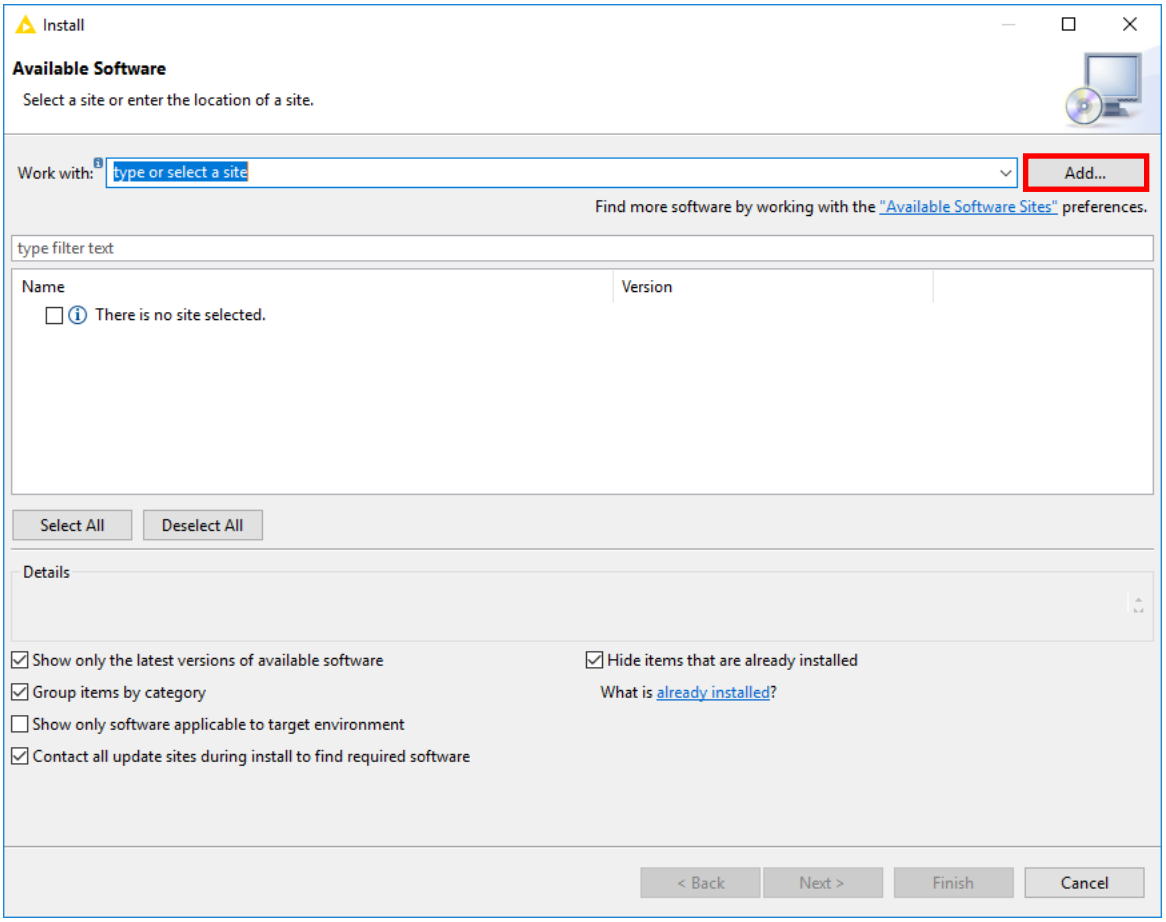


Figure 4. Reaxys KNIME archive file selection

3. Click the “Archive…” button:

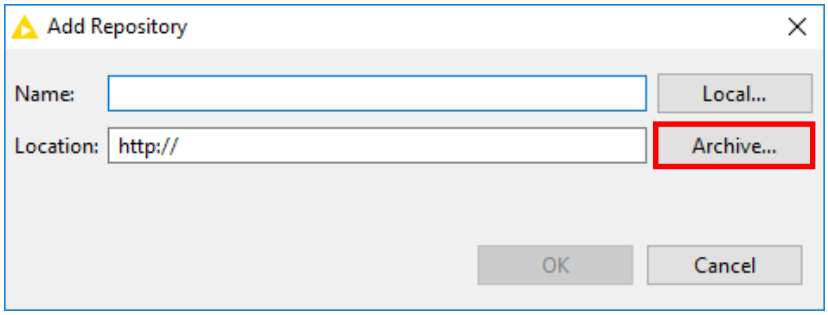


Figure 5. Reaxys KNIME archive file location

4. Browse for the “ReaxysKnimeUpdate.zip” file and click “OK”.

5. Select the “Elsevier extensions”.

6. Click “Next”, then “Next” again, accept the license agreement, click on “Finish” and restart KNIME to use the Reaxys nodes.

Proxy server configuration

In order to connect to the Reaxys API through a Proxy server configure the HTTPS Proxy entry in the KNIME Preferences. The configuration is accessible in KNIME from File → Preferences → General → Network Connections.

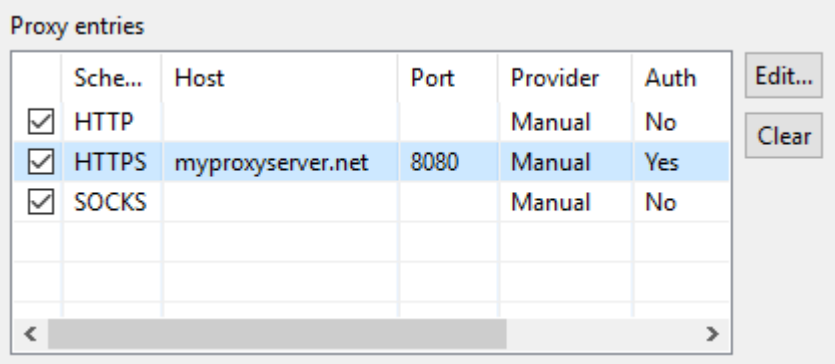


Figure 6. KNIME Proxy server configuration. Configure the HTTPS entry to connect to the Reaxys API though a Proxy server

If the Proxy server requires basic authentication, then KNIME will require the Proxy username and password to be entered when running a Reaxys workflow.

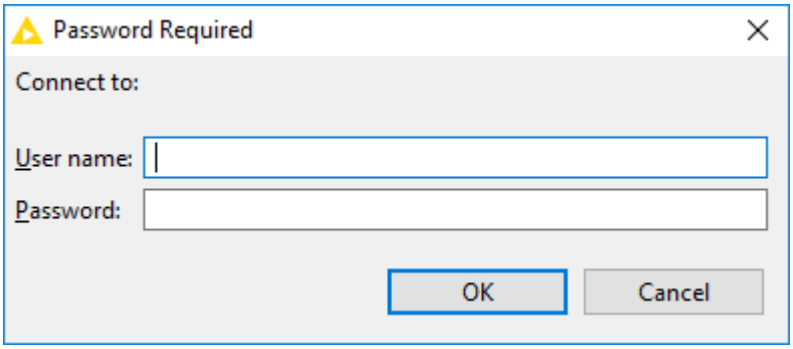


Figure 7. When using a Proxys server with basic authentication, KNIME will open a Dialog when running a Reaxys workflow to enter the Proxy server username and password.

Other suggested chemistry nodes for Knime

To get the most benefit from the chemical structure search using the Reaxys KNIME nodes, we suggest downloading these chemistry KNIME nodes to help manipulate chemical structures. These open-source nodes are available from the “Help” menu under “Install New Software”.

- **KNIME Base Chemistry Types & Nodes**
  - org.knime.features.chem.types.feature.group
    Produced by KNIME GmbH, Konstanz, Germany
- **KNIME Chemistry Add-Ons**
  - org.knime.features.ext.chem.tools.feature.group
    Produced by KNIME GmbH, Konstanz, Germany
- **ChemAxon/Infocom Marvin Extensions Feature**
  - jp.co.infocom.cheminfo.marvin.feature.feature.group
    Produced by Infocom Corporation
- **KNIME-CDK**
  - org.openscience.cdk.knime.feature.feature.group
    Produced by University Konstanz and EBI
- **RDKit KNIME integration**
  - org.rdkit.knime.feature.feature.group
    Produced by NIBR
- **Erlwood Knime Open Source Core**
  - org.erlwood.features.core.base.feature.group
    Produced by Erlwood
- **KNIME Interactive R Statistics Integration**
  - org.knime.features.r.feature.group
    Produced by KNIME GmbH, Konstanz, Germany
- **Vernalis KNIME Nodes**
  - com.vernalis.knime.feature.feature.group
    Produced by Vernalis

The Reaxys KNIME nodes

General description

The package includes four nodes—one each for query and retrieval of substance data, reaction data, bioactivity data points and bibliographic information (citations) from Reaxys and/or Reaxys Medicinal Chemistry. As described earlier, the available product and API license determines which data can be queried and retrieved with each of the nodes.

In general, the Reaxys KNIME nodes expect to receive input data from a table. Each row from the input data table represents a query value and is used together with the options given in the configuration dialog of the node to create a query.

The nodes access the Reaxys database server via the Reaxys API using the credentials given in the configuration dialog of the node. The nodes search all substance, reaction, bioactivity or citation instances matching the query.

Data retrieval occurs according to the specifications given by the user. With the Reaxys Substances node, the user can for example specify that they want to have all boiling point data retrieved for the substance instances matching the query. Similarly, the user can specify that they want to have structures as V2000 or V3000 mol files retrieved. With the Reaxys Reactions node, the user can for example specify that they want to have all reaction detail information retrieved for the reaction instances matching the query.

The requested data retrieved for all instances matching the queries is concatenated and returned as an output table.

The “Options” tab in the configuration dialog

In the “Options” tab, the user needs to provide the credentials for accessing the Reaxys database server (Figure 8 –1).

Furthermore, the user needs to specify how the query is constructed from the input table by indicating in which column of the input table the query values can be found (Figure 8 – 2) and by defining the type of input data (Figure 8 – 3).

Each query is separately sent to the server and evaluated. The result of each query is a set of instances of substances, reactions, bioactivities or citations. The user can specify how the instances retrieved per query are sorted (Figure 8 – 4).

The user can further add a query restriction that is applied to each query. The query restriction can be constructed using the query restriction builder (Figure 8 – 5) or it can be created by typing into the query restriction free text field (Figure 8 – 6).

In the “Options” tab, the user also needs to specify which data should be retrieved for the substance, reaction, bioactivity or citation instances matching the query (Figure 8 – 7).

The user can further specify the maximum number of retrieved instances per query (Figure 8 – 8). If this value is set to less than zero, no limit on the number of retrieved instances per query will be imposed.

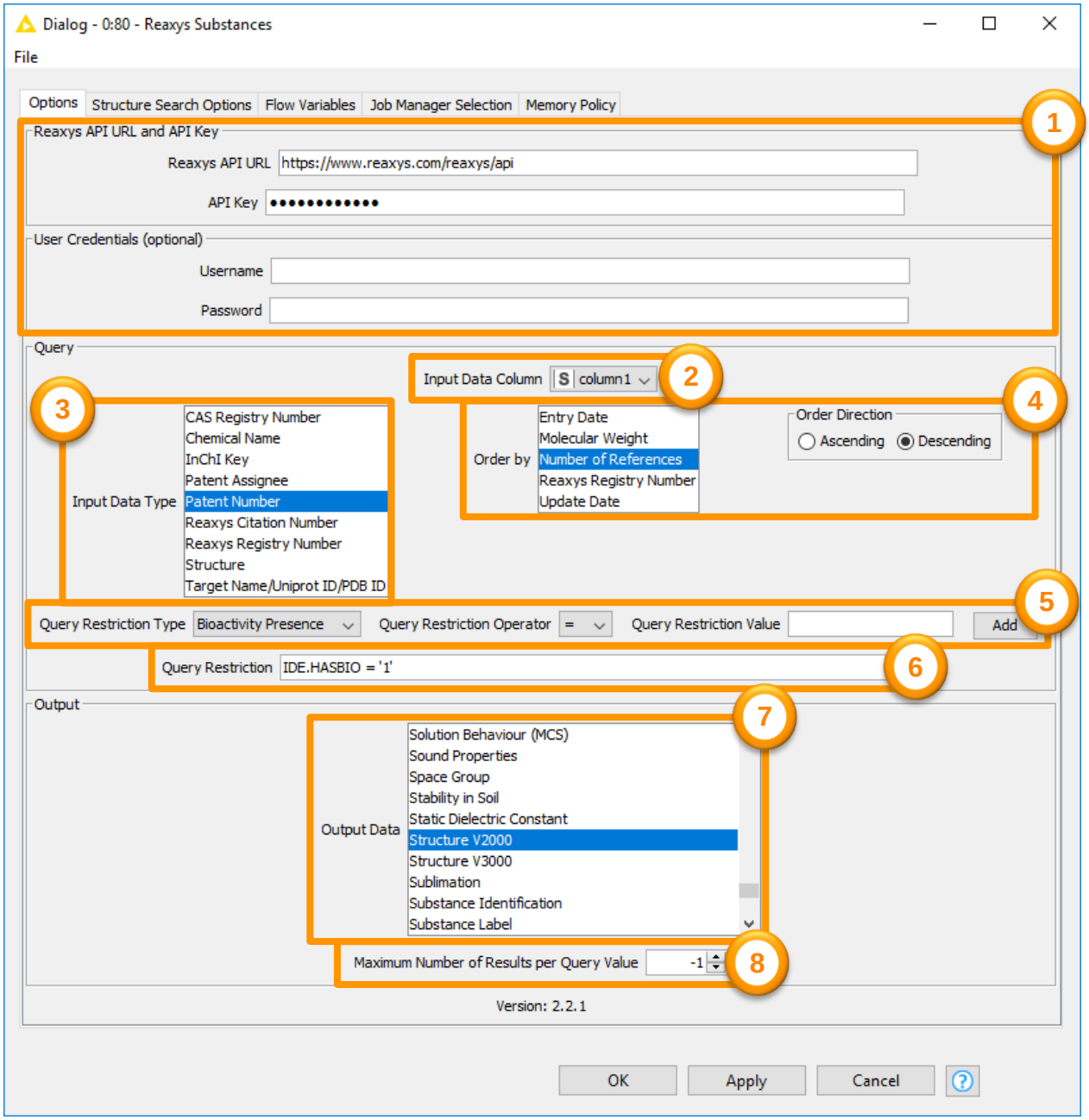


Figure 8. The "Options" tab in the configuration dialog. 1) The field for user credentials. 2) The field for specifying the input data column. 3) The field for defining the type of input data. 4) The field for specifying how the instances retrieved per query are sorted. 5 and 6) The fields for constructing query restrictions. The restriction can be built (5) or typed in (6). 7) The field for specifying which data should be retrieved. 8) the field for specifying the maximum number of instances retrieved per query.

The "Structure Search Options" tab in the configuration dialog

The Reaxys KNIME nodes can take substance structures (mol files) or reactions (rxn files) as query values. In that case, the user needs to specify "Structure" as input data type. The nodes will then perform a structure search. Exact match structure searching, substructure searching and searching by similarity are supported.

Note that the Reaxys Citations KNIME node does not allow users to search for bibliographic information by structure.

The user needs to specify the structure search type (Figure 9 – 1):

- **Exact:** For each query structure, an exact match structure search will be performed. Available valence on all atoms will be filled with hydrogen atoms. This is the default setting.
- **Substructure**: For each query structure, a substructure search will be performed. Free substitution is allowed on all atoms, where the valence is not filled.
- **Substructure on heteroatoms**: For each query structure, a substructure search will be performed. Free substitution is allowed on heteroatoms, where the valence is not filled.
- **Similarity near**: For each query structure, the search will include structures containing the same ring and chain systems (possibly multiple) with the original relative positions of substituents and extended by further simple substituents such as hydrocarbons. This is typically equivalent to 80% similarity.
- **Similarity medium**: For each query structure, the search will include structures with a wider range of rings and substituents: the degree of unsaturation, form and substitution pattern of rings is extended. This is typically equivalent to 60% similarity.
- **Similarity wide**: For each query structure, the search will include a still wider range of substituents but retaining, to some extent, the influence of the relative positions of substituents. This is typically equivalent to 40% similarity.
- **Similarity widest**: For each query structure, the search will be performed without any restrictions on the relative positions of substituents. This is typically equivalent to 20% similarity.

The user also needs to specify how stereochemistry is treated (Figure 9 – 2):

- **Ignore stereo**: For each query structure, ignore stereochemistry.
- **Stereo absolute**: For each query structure, all stereo centers match the mapped centers in the search result. This is the default setting.
- **Stereo relative**: For each query structure, all stereo centers match the mapped centers in the search result or its mirror image, where all centers are synchronously inverted.

The user can further specify miscellaneous options (Figure 9 – 3). Multiple options can be selected by pressing the CTRL key while clicking on an option.

- **Align results with query**: Checking this option ensures that the structures in the results will have the same general orientation as the structure in the query.
- **Ignore atom mapping**: If atom–atom mappings are defined in a query, this option will ignore those mappings in the search. This option is not available from the Substances or Bioactivities nodes.
- **Include related Markush**: Related Markush structures are retrieved. This option is not available from the Reactions node.
- **Include tautomers**: Tautomers of the query structures will also be found.
- **Keep separate fragments**: When the query contains more than one isolated component, checking this option ensures that they are to be retrieved as separate components.
- **No charges**: Charged compounds will be excluded from the results.
- **No isotopes**: Isotopes will not appear in the results.
- **No mixtures**: Mixtures (and polymers) containing the query are excluded from the results. This option is not available from the Reactions node.
- **No radicals**: Compounds containing radicals will not appear in the results.
- **No ring closures**: Ring closures between atoms or groups with free sites will be excluded from the results.
- **No salt**s: Multi-fragment substances such as salts or charge-transfer complexes are excluded from the results. This option is not available from the Reactions node.

The user can further specify the number of atoms, fragments and ring closures to be found in the results (Figure 9 – 4). These options are not available from the Reaction node. For each option, a lower and upper bound must be given:

- **Number of atoms**: The total number of atoms to be found in the retrieved structure.
- **Number of fragments**: The total number of fragments to be found in the retrieved structure (e.g., salts and addition compounds).
- **Number of ring closures**: The total number of rings to be found in the retrieved structure (defined as the smallest number of ring bonds which must be broken in order to convert the structure into an acyclic structure).

In the “Structure Search Options” tab in the configuration dialog of the Reaxys Reactions KNIME node, additional role options need to be specified. The role options define the role that the query structures play in the reaction results. Role options are only available from the Reactions node. If a reaction structure is searched, this option must be set to "Any role".

- **Any role**: The query appears in any role in the results. This is the default setting.
- **Product**: The query appears as a product in the results.
- **Starting material**: The query appears as a reactant in the results.
- **Reagent/Catalyst**: The query appears as a reagent or catalyst in the results.

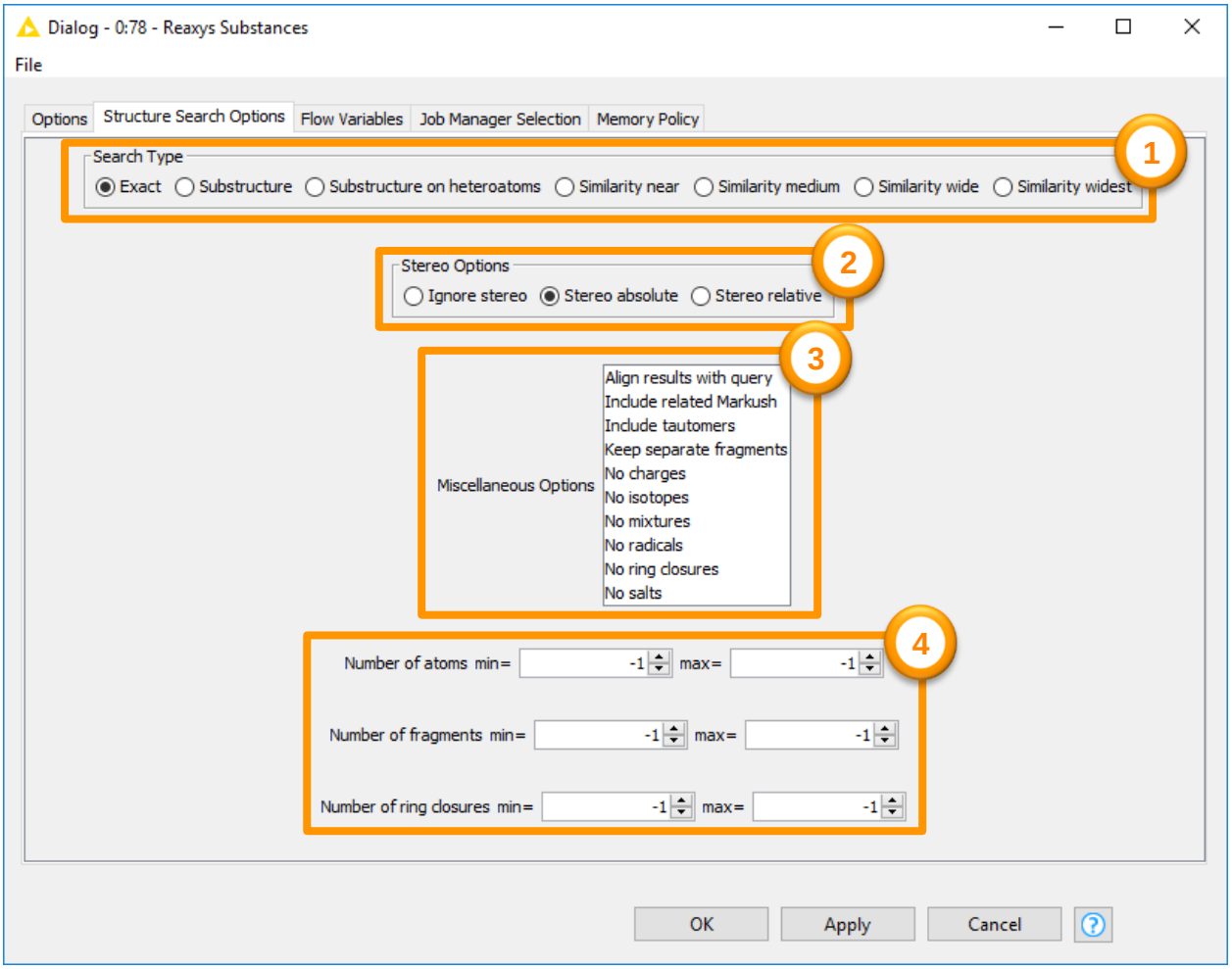


Figure 9. Structure Search Options” tab in the configuration dialog of the Substances node. 1) The field for specifying the search type. 2) The field for specifying how stereochemistry is treated. 3) The field for defining other options (see text for details). 4) The field for specifying structural details.

Output

The output from the Reaxys KNIME nodes is a table where the requested data retrieved for all substance, reaction, bioactivity or citation instances matching the queries is concatenated.

Each row in the output table contains the original query value in the first column and a specific object identifier for the substance, reaction, bioactivity or citation instances matching the query. The specific object identifier is the Reaxys Registry Number for substance instances, the Reaction ID for reaction instances, the Bioactivity ID for bioactivity instances and the Citation Number for citation instances.

All columns in the output table are of string type. The user might have to cast the value to an integer or floating-point number for numeric calculations. Note that Reaxys sometimes returns ranges as well as specific values. The user should consider filtering out or calculate the median for ranges in the KNIME workflow.

Strings representing mol or rxn files can be casted to molecules using for example the Molecule Type Cast KNIME node from the KNIME Chemistry Base nodes provided by KNIME GmbH, Konstanz, Germany.

The hierarchical organization of Reaxys data, such as Reaction stages and conditions, is represented by concatenating the values using level specific delimiters. The | character is used as delimiter to group values on the same level, such as solvents used in a given stage. The % character is used as delimiter for the parent level such as the stage level. The null indicates the non-existence of a value. Figure 10 illustrates an example for the hierarchical organization of reaction conditions for the Reaxys Reaction ID 23590888.The reaction has two stages, with each stage having different reaction temperatures and solvents. Figure 11 illustrates how these relationships are represented in the KNIME output table.

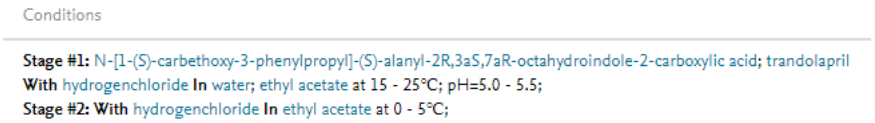


Figure 10. Reaxys Reaction ID 23590888 describes two stages with stage #1 having water and ethyl acetate as solvent and a temperature of 15-25 °C, stage #2 having ethyl acetate as solvent and a temperature of 0-5 °C

Row ID	S Solvent Reaxys Registry Number	S Solvent	S Temperature	S pH-Value
Row 1	3587155|506104%506104	water|ethyl acetate%ethyl acetate	15 - 25%0 - 5	5.0 - 5.5%null

Figure 11. Representation of the hierarchical and correlated reaction conditions in the KNIME output table. Multiple solvents of stage #1 are separated by the | character. The solvents for stage #1 and stage #2 are separated using the % character.

Performance

This section lists examples of searches performed using Reaxys KNIME nodes and the corresponding duration of the search in seconds. Average and standard deviation (SD) were calculated based on ten runs of the example workflows that are displayed in the next section.

Example		Output	Duration in seconds	
Number	Details		Average	SD
1	Search substances by patent numbers and retrieve substance identification data (Reaxys Ids, InChI Key, chemical, name, molecular weight, etc…) – structures are not included	306 substances	7.0	2.8
2	Search substances by chemical name and retrieve substance property (boiling point) data	398 boiling point records	5.6	0.4
3	Search bioactive substances by patent number and retrieve the structures	174 chemical structures	7.7	0.9
4	Search preparations for a set of structures and retrieve the reaction structures	125 reaction structures	98.0	37.9
5	Search bioactivities by patent number	300 bioactivities	7.1	2.7
6	Search bioactivities and substances by target name and retrieve bioactivity datapoints and structures for the most active compounds	1000 datapoints and structures	215	25.3

Examples

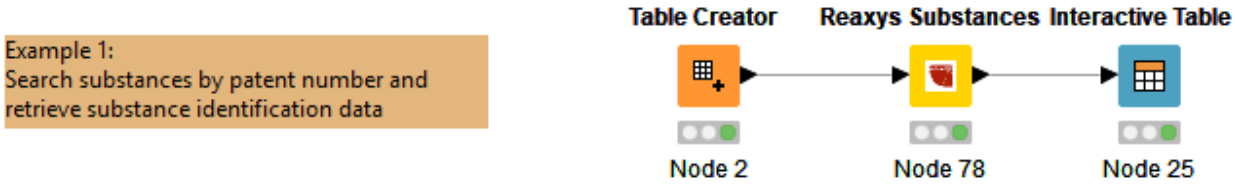
Reaxys KNIME node examples as well as the required files can be download here:

http://supportcontent.elsevier.com/Support Hub/Reaxys/knime/3.5/Reaxys KNIME workflow Examples/Reaxys KNIME workflow Examples 3_5.zip

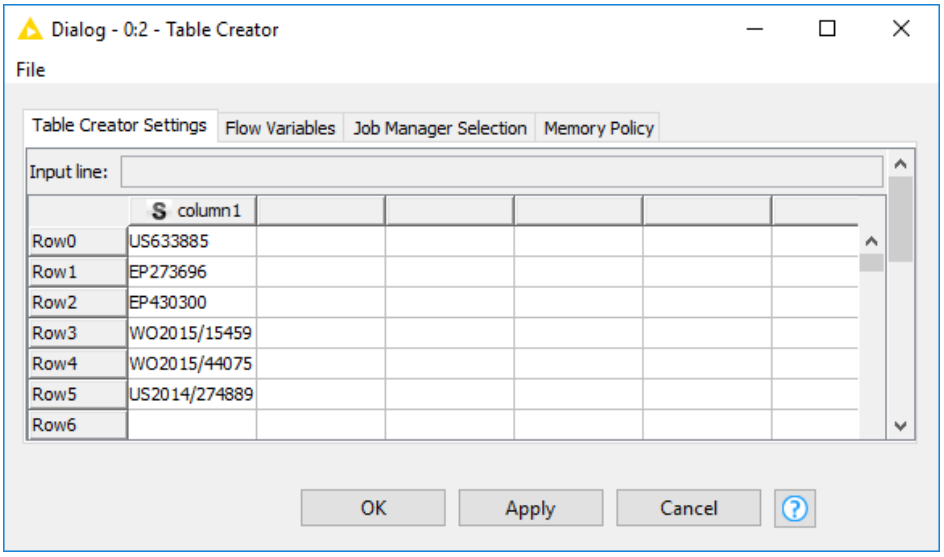
1. Search substances by patent number and retrieve substance identification data

In this example, we search substances by patent number and retrieve substance identification data for all the identified substances.

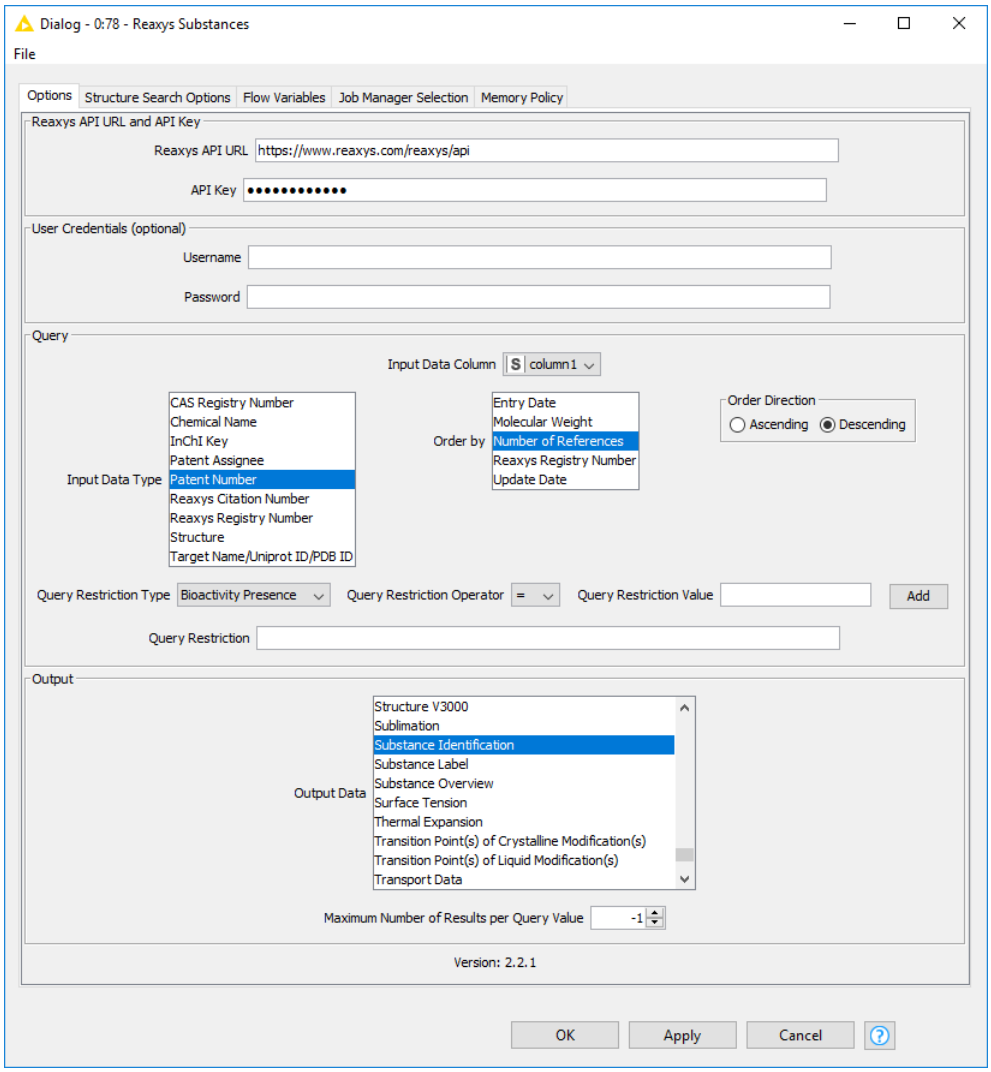
The overall workflow appears as follows:



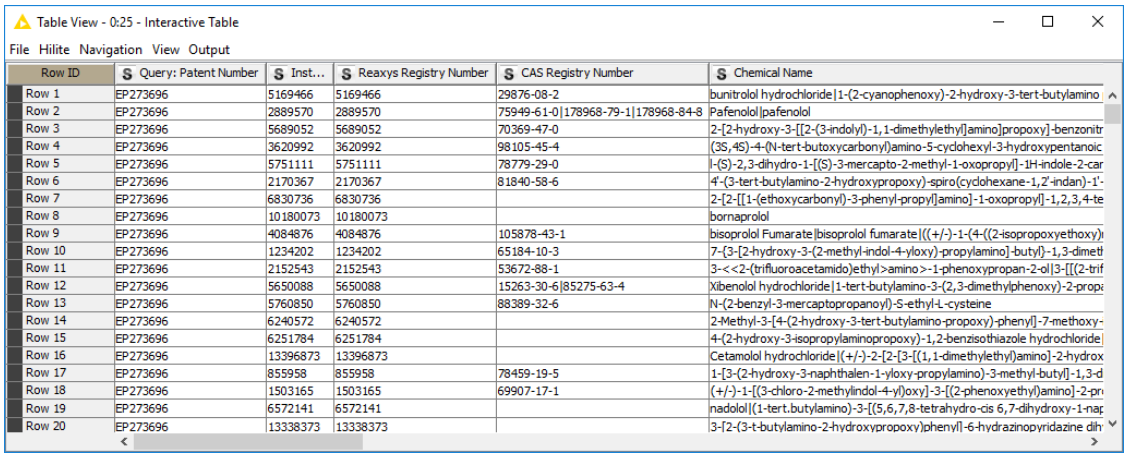
The “Table Creator” node is used to define the input table with the query values. In this example, we use six query values.



In the “Reaxys Substances” node, we need to define the type of the input data and the output data.



The output is a table with this format:



Row ID	Query: Patent Number	Inst...	Reaxys Registry Number	CAS Registry Number	Chemical Name
Row 1	EP273696	5169466	5169466	29876-08-2	bunitrolol hydrochloride|1-(2-cyanophenoxy)-2-hydroxy-3-tert-butylamino
Row 2	EP273696	2889570	2889570	75949-61-0|178968-79-1|178968-84-8	Pafenolol|pafenolol
Row 3	EP273696	5689052	5689052	70369-47-0	2-[2-hydroxy-3-[[2-(3-indolyl)-1,1-dimethylethyl]amino]propoxy]-benzonitr
Row 4	EP273696	3620992	3620992	98105-45-4	(3S,4S)-4-(N-tert-butoxycarbonyl)amino-5-cyclohexyl-3-hydroxypentanoic
Row 5	EP273696	5751111	5751111	78779-29-0	l-(S)-2,3-dihydro-1-[(S)-3-mercapto-2-methyl-1-oxopropyl]-1H-indole-2-car
Row 6	EP273696	2170367	2170367	81840-58-6	4'-(3-tert-butylamino-2-hydroxypropoxy)-spiro(cyclohexane-1,2'-indan)-1'-
Row 7	EP273696	6830736	6830736		2-[2-[[1-(ethoxycarbonyl)-3-phenyl-propyl]amino]-1-oxopropyl]-1,2,3,4-te
Row 8	EP273696	10180073	10180073		bornaprolol
Row 9	EP273696	4084876	4084876	105878-43-1	bisoprolol Fumarate|bisoprolol fumarate|((+/-)-1-(4-((2-isopropoxyethoxy)(
Row 10	EP273696	1234202	1234202	65184-10-3	7-[3-[2-hydroxy-3-(2-methyl-indol-4-yloxy)-propylamino]-butyl]-1,3-dimeth
Row 11	EP273696	2152543	2152543	53672-88-1	3-<<2-(trifluoroacetamido)ethyl>amino>-1-phenoxypropan-2-ol|3-[[(2-trif
Row 12	EP273696	5650088	5650088	15263-30-6|85275-63-4	Xibenolol hydrochloride|1-tert-butylamino-3-(2,3-dimethylphenoxy)-2-propa
Row 13	EP273696	5760850	5760850	88389-32-6	N-(2-benzyl-3-mercaptopropanoyl)-S-ethyl-L-cysteine
Row 14	EP273696	6240572	6240572		2-Methyl-3-[4-(2-hydroxy-3-tert-butylamino-propoxy)-phenyl]-7-methoxy-
Row 15	EP273696	6251784	6251784		4-(2-hydroxy-3-isopropylaminopropoxy)-1,2-benzisothiazole hydrochloride|
Row 16	EP273696	13396873	13396873		Cetamolol hydrochloride|(+/-)-2-[2-[3-[(1,1-dimethylethyl)amino]-2-hydrox
Row 17	EP273696	855958	855958	78459-19-5	1-[3-(2-hydroxy-3-naphthalen-1-yloxy-propylamino)-3-methyl-butyl]-1,3-d
Row 18	EP273696	1503165	1503165	69907-17-1	(+/-)-1-[(3-chloro-2-methylindol-4-yl)oxy]-3-[(2-phenoxyethyl)amino]-2-pr
Row 19	EP273696	6572141	6572141		nadolol|(1-tert.butylamino)-3-[(5,6,7,8-tetrahydro-cis 6,7-dihydroxy-1-nap
Row 20	EP273696	13338373	13338373		3-[2-(3-t-butylamino-2-hydroxypropoxy)phenyl]-6-hydrazinopyridazine dih

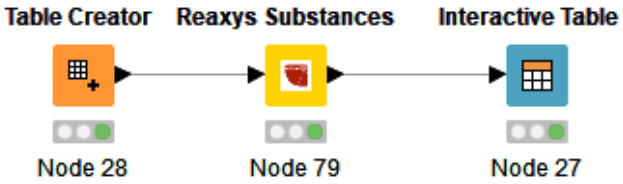
Result: 306 records of substance identification data are retrieved for the compounds matching the query.

2. Search substances by chemical name and retrieve substance property data

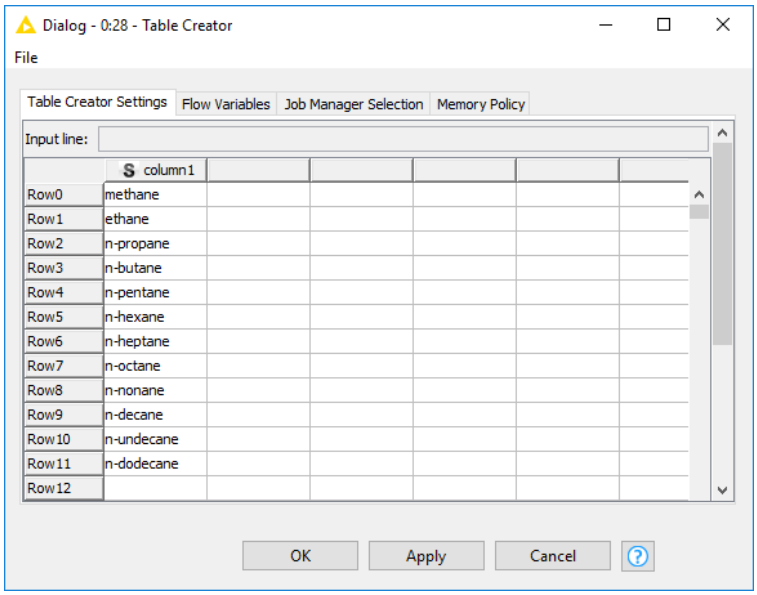
In this example, we search substances by chemical name and retrieve boiling point data for all the identified substances.

The overall workflow appears as follows:

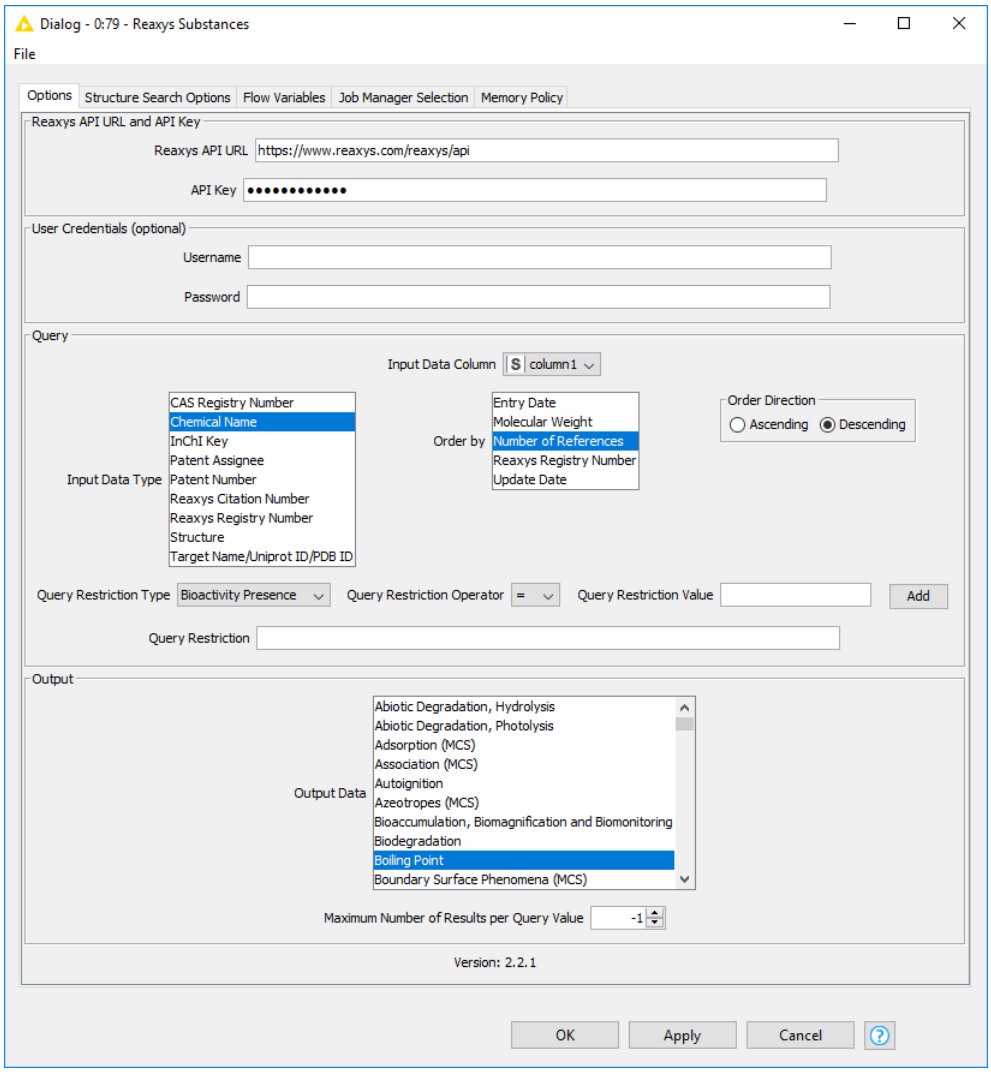
Example 2:
Search substances by chemical name and retrieve substance propety (boiling point) data



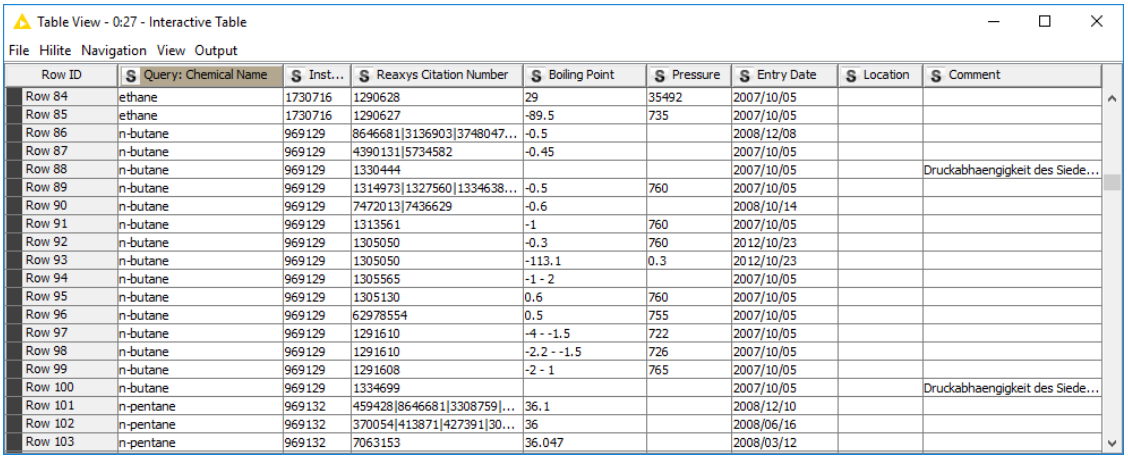
The “Table Creator” node is used for defining the input table with the query values. In this example, we use six query values.



In the “Reaxys Substances” node, we need to define the type of the input data and the output data:



The output is a table with this format:



Row ID	Query: Chemical Name	Inst...	Reaxys Citation Number	Boiling Point	Pressure	Entry Date	Location	Comment
Row 84	ethane	1730716	1290628	29	35492	2007/10/05		
Row 85	ethane	1730716	1290627	-89.5	735	2007/10/05		
Row 86	n-butane	969129	8646681|3136903|3748047...	-0.5		2008/12/08		
Row 87	n-butane	969129	4390131|5734582	-0.45		2007/10/05		
Row 88	n-butane	969129	1330444			2007/10/05		Druckabhaengigkeit des Siede...
Row 89	n-butane	969129	1314973|1327560|1334638...	-0.5	760	2007/10/05		
Row 90	n-butane	969129	7472013|7436629	-0.6		2008/10/14		
Row 91	n-butane	969129	1313561	-1	760	2007/10/05		
Row 92	n-butane	969129	1305050	-0.3	760	2012/10/23		
Row 93	n-butane	969129	1305050	-113.1	0.3	2012/10/23		
Row 94	n-butane	969129	1305565	-1 - 2		2007/10/05		
Row 95	n-butane	969129	1305130	0.6	760	2007/10/05		
Row 96	n-butane	969129	62978554	0.5	755	2007/10/05		
Row 97	n-butane	969129	1291610	-4 - -1.5	722	2007/10/05		
Row 98	n-butane	969129	1291610	-2.2 - -1.5	726	2007/10/05		
Row 99	n-butane	969129	1291608	-2 - 1	765	2007/10/05		
Row 100	n-butane	969129	1334699			2007/10/05		Druckabhaengigkeit des Siede...
Row 101	n-pentane	969132	459428|8646681|3308759|...	36.1		2008/12/10		
Row 102	n-pentane	969132	370054|413871|427391|30...	36		2008/06/16		
Row 103	n-pentane	969132	7063153	36.047		2008/03/12		

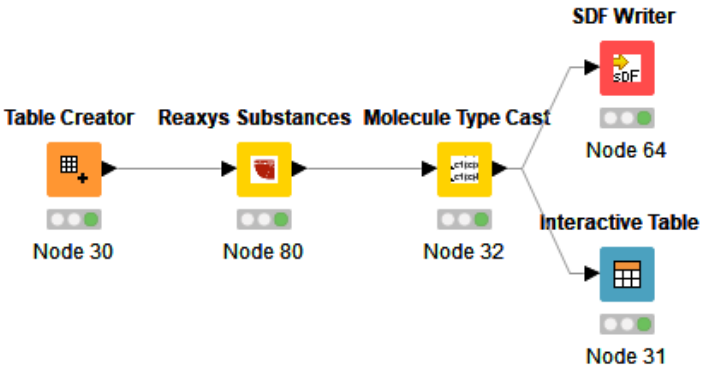
Result: 405 records of boiling point data are retrieved (as of this writing).

3. Search bioactive substances by patent number and retrieve the structures

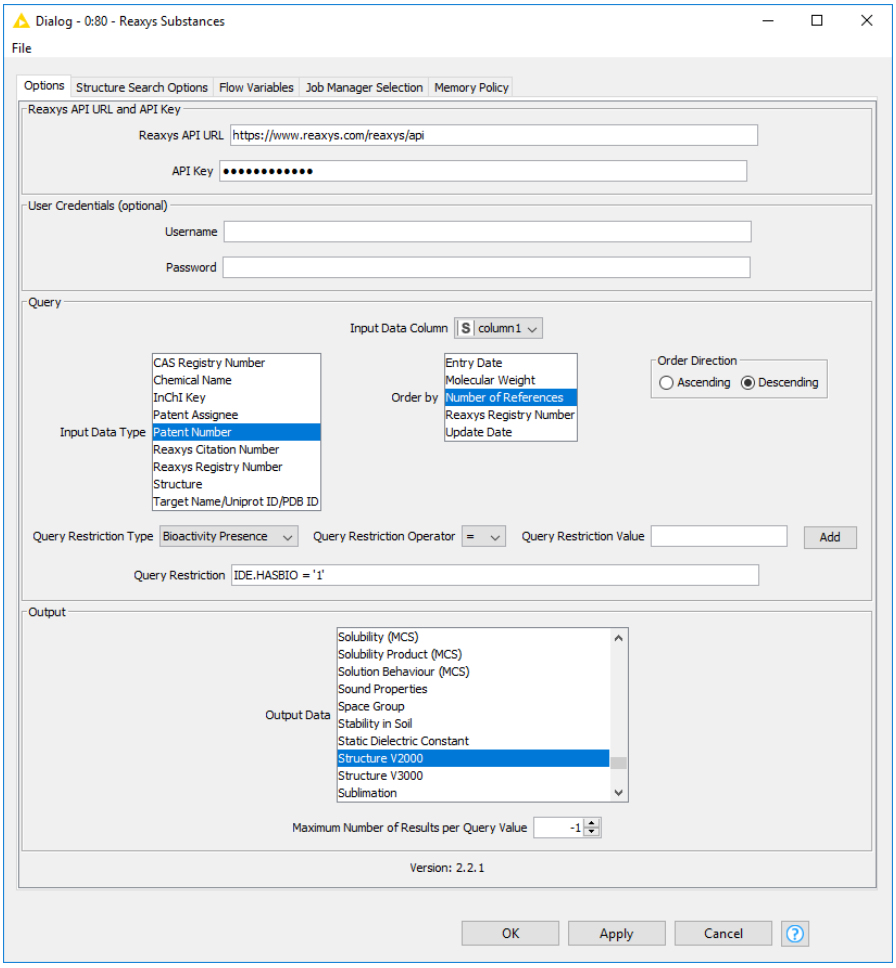
In this example, we search substances by patent number and we retrieve the structures (V2000 mol files) for all the identified substances. We will display the structures on screen and write an SDF file.

The overall workflow appears as follows:

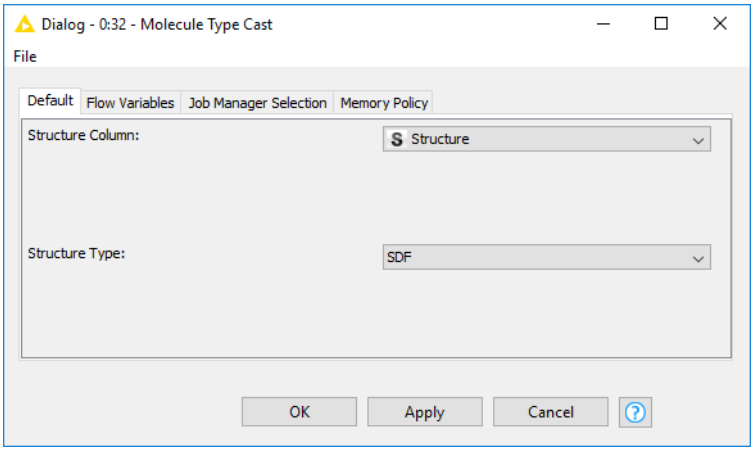
Example 3:
Search bioactive substances by patent number and retrieve the structures



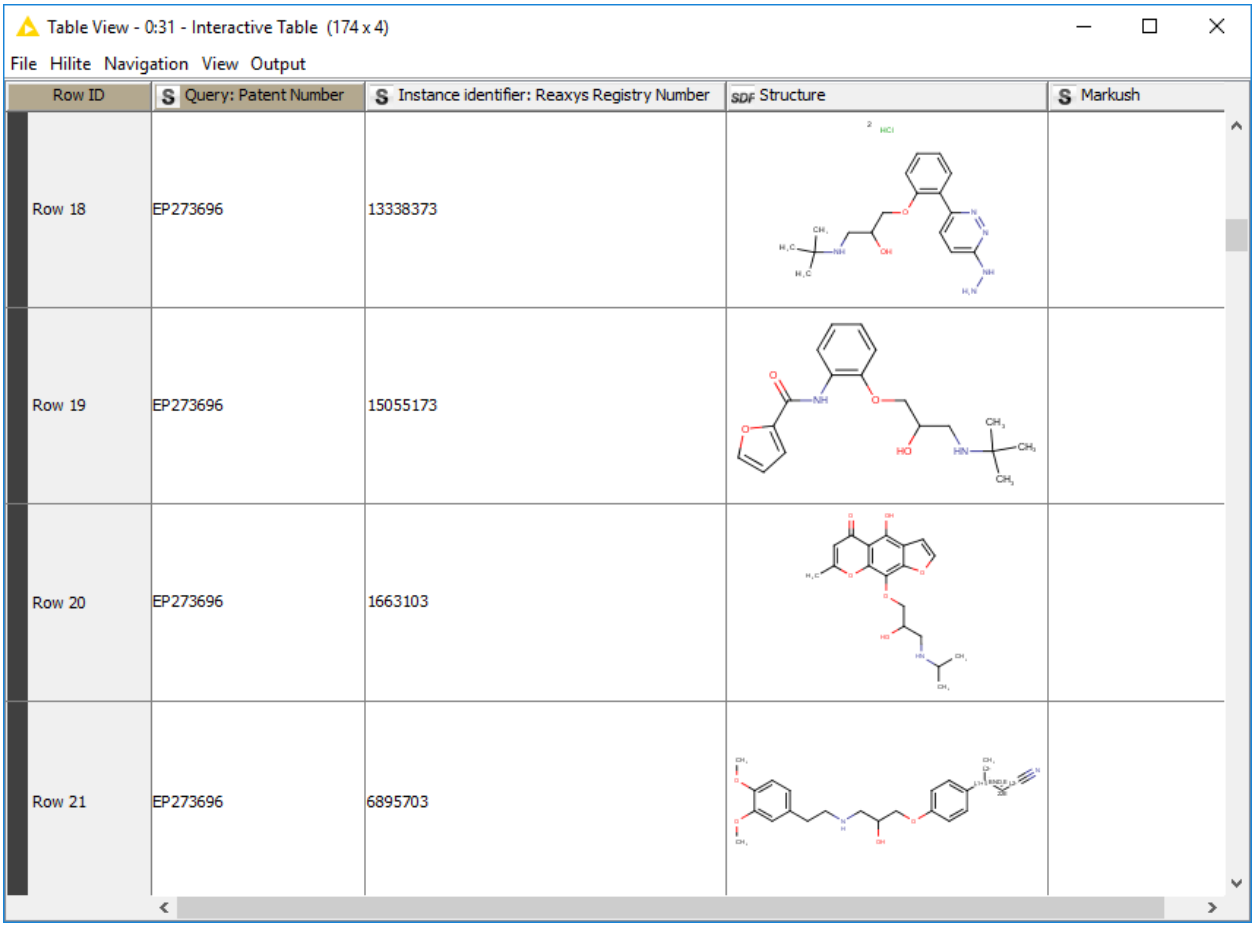
In the “Reaxys Substances” node, we need to define the type of the input data and the output data. Note that a query restriction is used for searching only bioactive substances, i.e., substances that have bioactivity data from Reaxys Medicinal Chemistry associated with them.



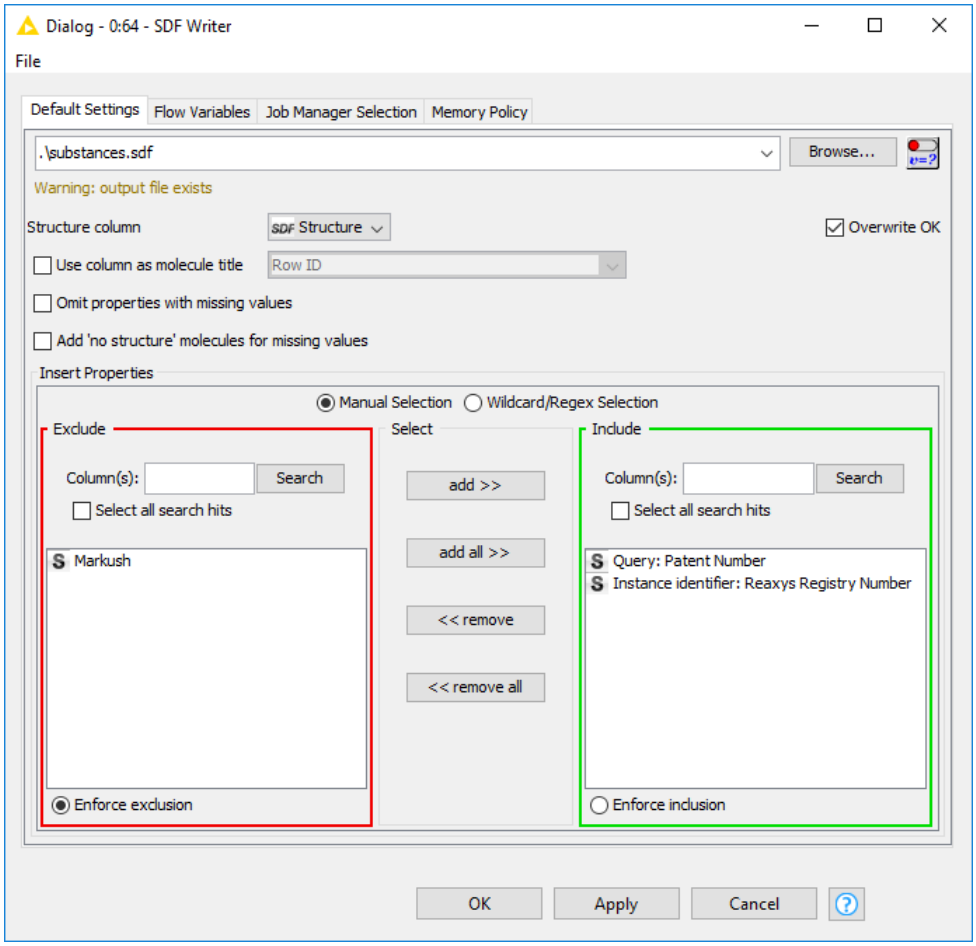
Structures are retrieved as mol file strings. Prior to displaying the structures on screen, we need to cast the mol file strings to structure types. We use the “Molecule Type Cast” node.



The output is a table with this format:



In addition to displaying the structures on screen, we want to export them into an SDF file. We use the “SDF Writer” node for that.

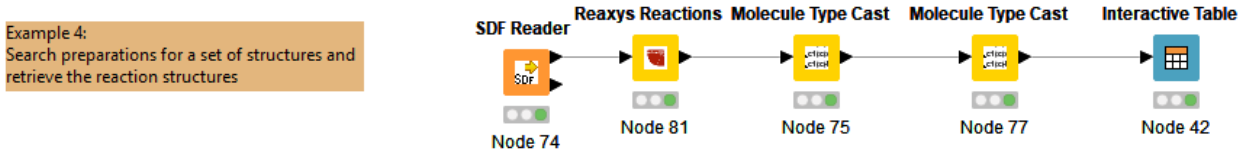


Result: 174 structures are retrieved and exported.

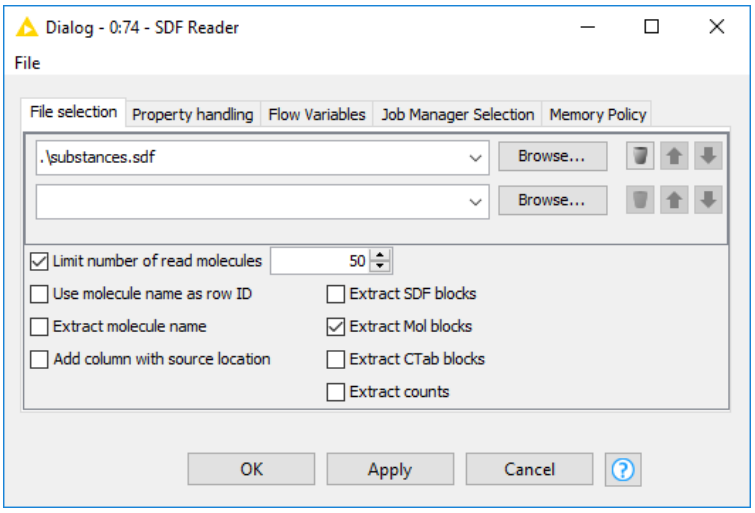
4. Search preparations for a set of structures and retrieve the reaction structures

In this example, we use the 174 substance structures that have been exported from the previous workflow as a structure query. We search preparations and retrieve reaction structures.

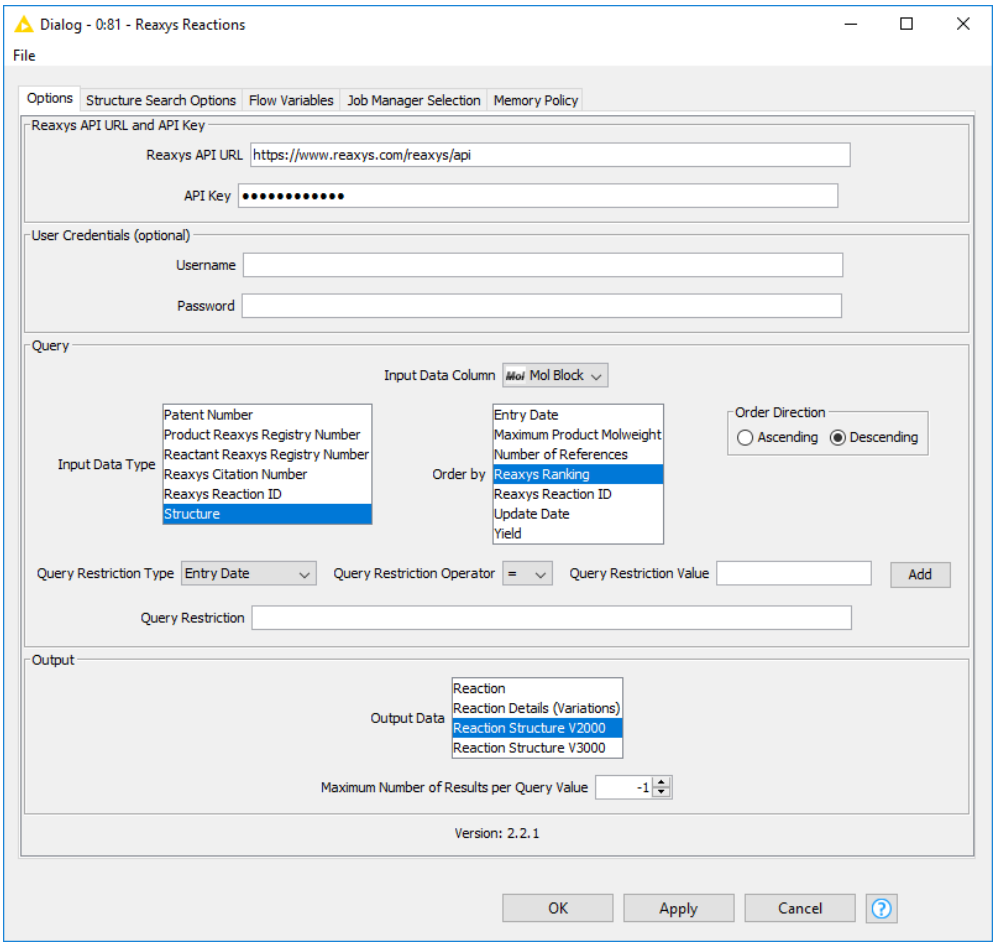
The overall workflow appears as follows:



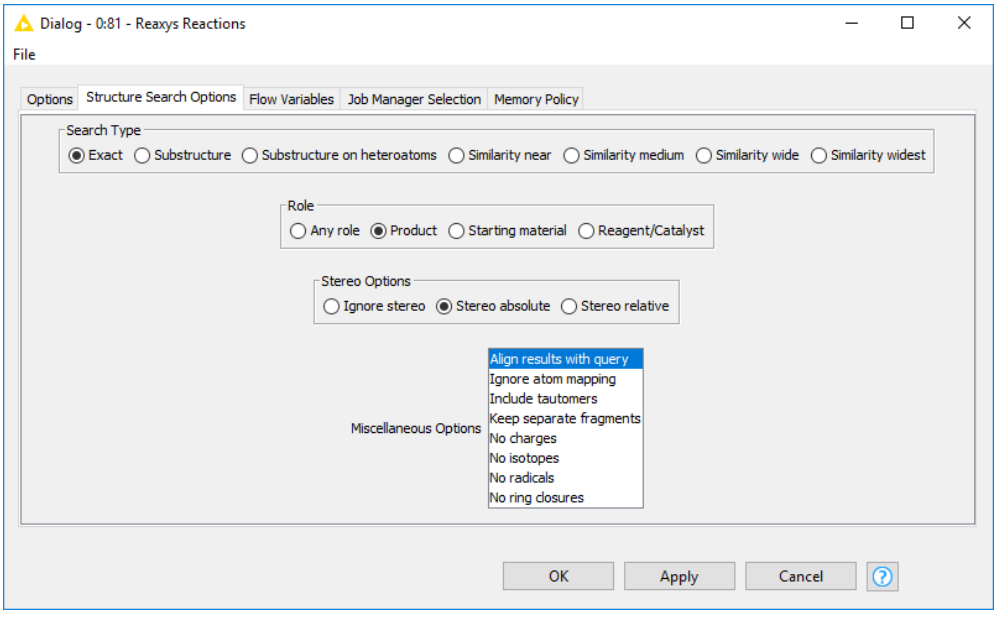
Reading the SDF file, we see:



The settings for the Reaxys node are:



It is a structure query, so specific structure search options must be set.

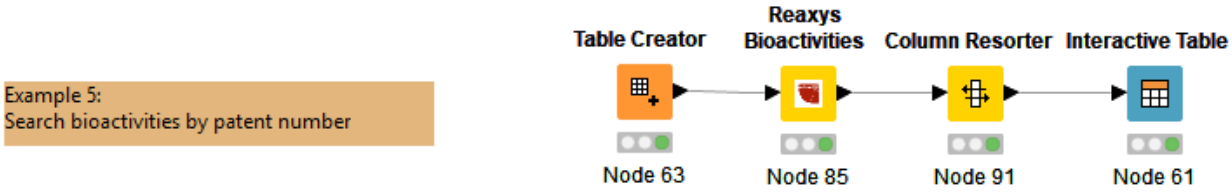


Result: 133 reaction structures are retrieved (as of this writing).

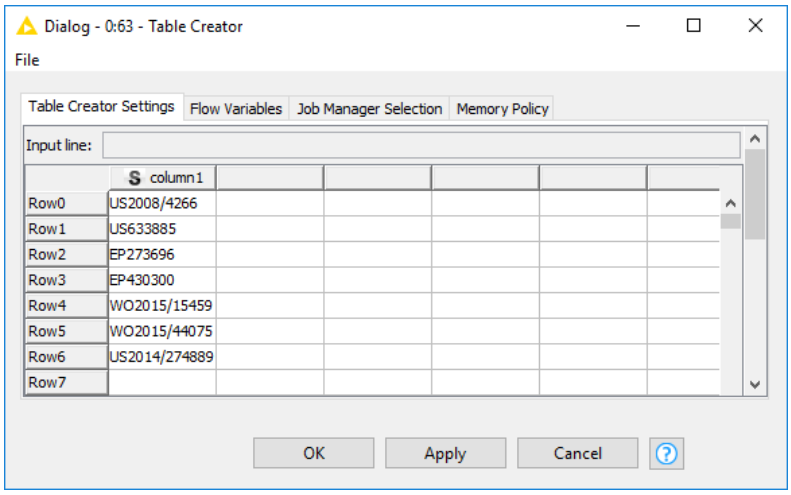
5. Search for bioactivities by patent number

In this example, we search for bioactivities associated with a given patent number.

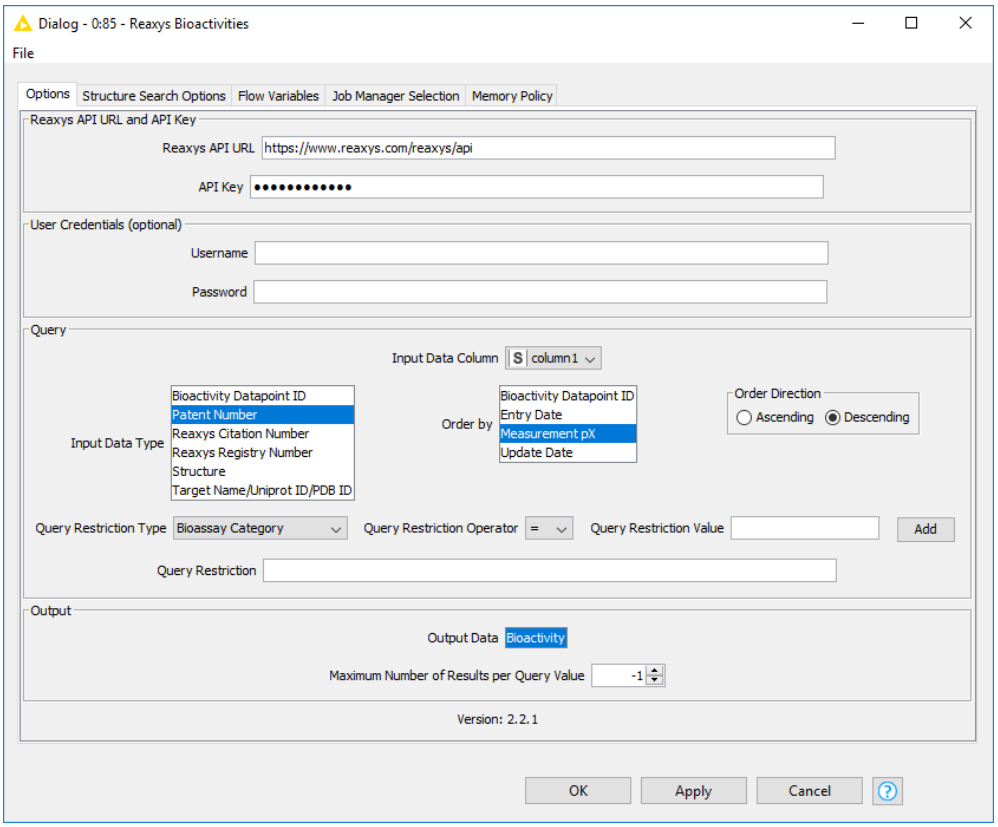
The overall workflow appears as follows:



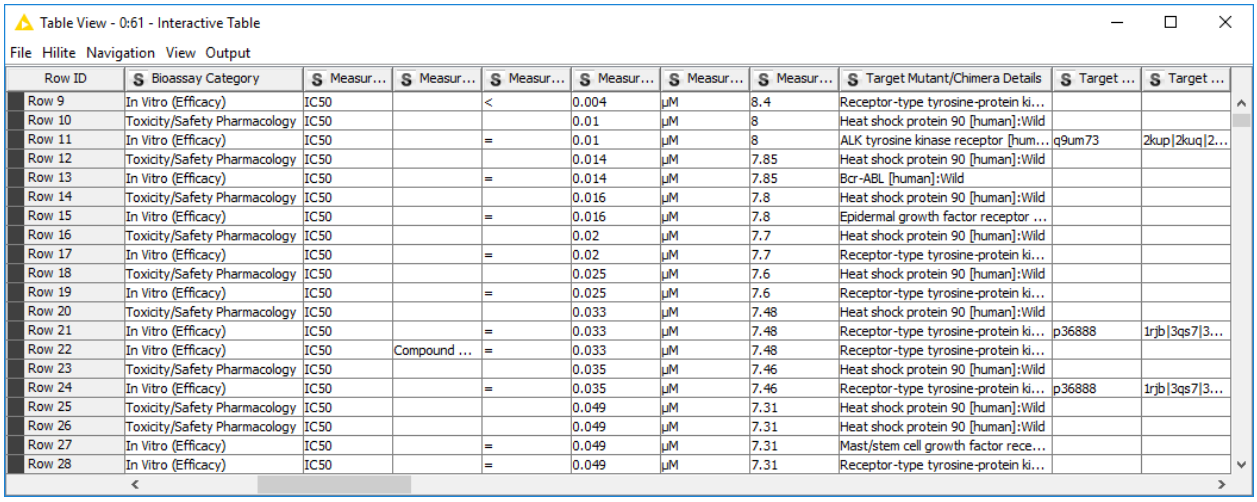
We have an input table with 6 patent numbers.



We use the Bioactivities node.



The output is a table with this format:



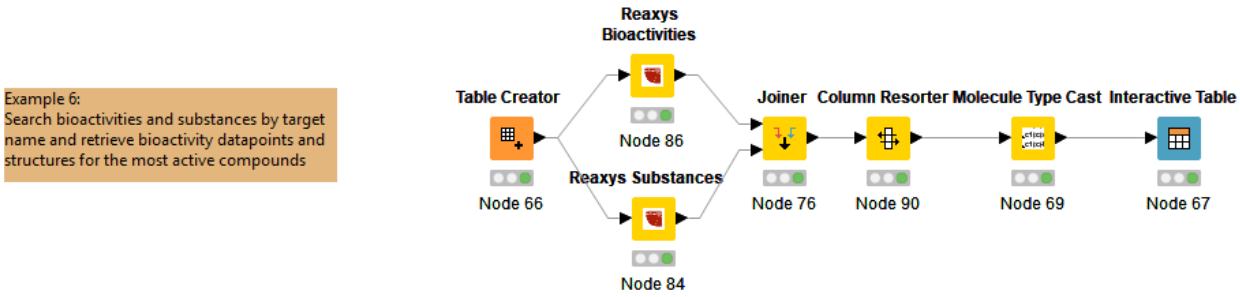
Row ID	Bioassay Category	Measur...	Measur...	Measur...	Measur...	Measur...	Measur...	Target Mutant/Chimera Details	Target ...	Target ...
Row 9	In Vitro (Efficacy)	IC50		<	0.004	µM	8.4	Receptor-type tyrosine-protein ki...		
Row 10	Toxicity/Safety Pharmacology	IC50			0.01	µM	8	Heat shock protein 90 [human]:Wild		
Row 11	In Vitro (Efficacy)	IC50		=	0.01	µM	8	ALK tyrosine kinase receptor [hum...	q9um73	2kup|2kuq|2...
Row 12	Toxicity/Safety Pharmacology	IC50			0.014	µM	7.85	Heat shock protein 90 [human]:Wild		
Row 13	In Vitro (Efficacy)	IC50		=	0.014	µM	7.85	Bcr-ABL [human]:Wild		
Row 14	Toxicity/Safety Pharmacology	IC50			0.016	µM	7.8	Heat shock protein 90 [human]:Wild		
Row 15	In Vitro (Efficacy)	IC50		=	0.016	µM	7.8	Epidermal growth factor receptor ...		
Row 16	Toxicity/Safety Pharmacology	IC50			0.02	µM	7.7	Heat shock protein 90 [human]:Wild		
Row 17	In Vitro (Efficacy)	IC50		=	0.02	µM	7.7	Receptor-type tyrosine-protein ki...		
Row 18	Toxicity/Safety Pharmacology	IC50			0.025	µM	7.6	Heat shock protein 90 [human]:Wild		
Row 19	In Vitro (Efficacy)	IC50		=	0.025	µM	7.6	Receptor-type tyrosine-protein ki...		
Row 20	Toxicity/Safety Pharmacology	IC50			0.033	µM	7.48	Heat shock protein 90 [human]:Wild		
Row 21	In Vitro (Efficacy)	IC50		=	0.033	µM	7.48	Receptor-type tyrosine-protein ki...	p36888	1rjb|3qs7|3...
Row 22	In Vitro (Efficacy)	IC50	Compound ...	=	0.033	µM	7.48	Receptor-type tyrosine-protein ki...		
Row 23	Toxicity/Safety Pharmacology	IC50			0.035	µM	7.46	Heat shock protein 90 [human]:Wild		
Row 24	In Vitro (Efficacy)	IC50		=	0.035	µM	7.46	Receptor-type tyrosine-protein ki...	p36888	1rjb|3qs7|3...
Row 25	Toxicity/Safety Pharmacology	IC50			0.049	µM	7.31	Heat shock protein 90 [human]:Wild		
Row 26	Toxicity/Safety Pharmacology	IC50			0.049	µM	7.31	Heat shock protein 90 [human]:Wild		
Row 27	In Vitro (Efficacy)	IC50		=	0.049	µM	7.31	Mast/stem cell growth factor rece...		
Row 28	In Vitro (Efficacy)	IC50		=	0.049	µM	7.31	Receptor-type tyrosine-protein ki...		

Result: 300 bioactivity datapoints are retrieved.

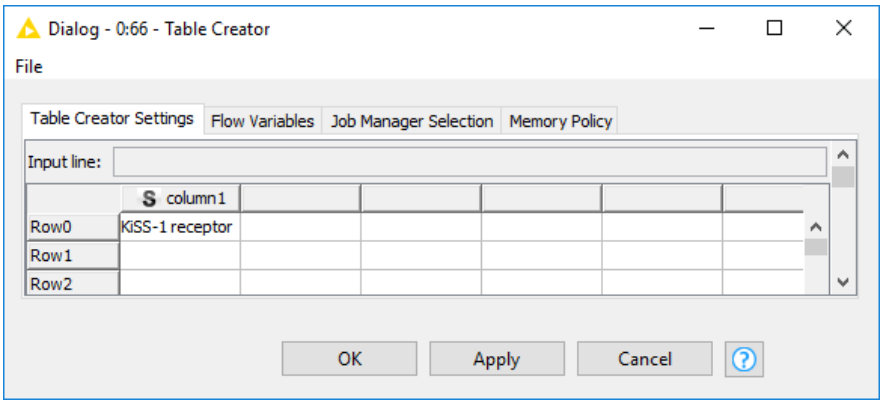
6. Search bioactivities and substances by target name and retrieve bioactivity datapoints and structures for the most active compounds

In this example, we search bioactivities and substances associated with a given target and then retrieve the bioactivity datapoints and structures for the most active compounds.

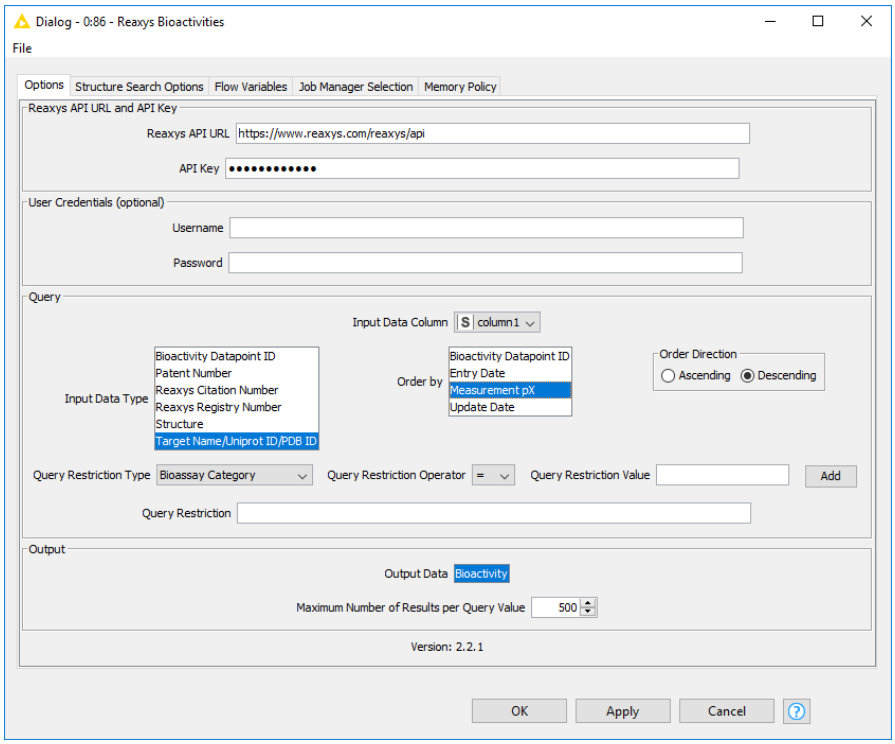
The overall workflow appears as follows:

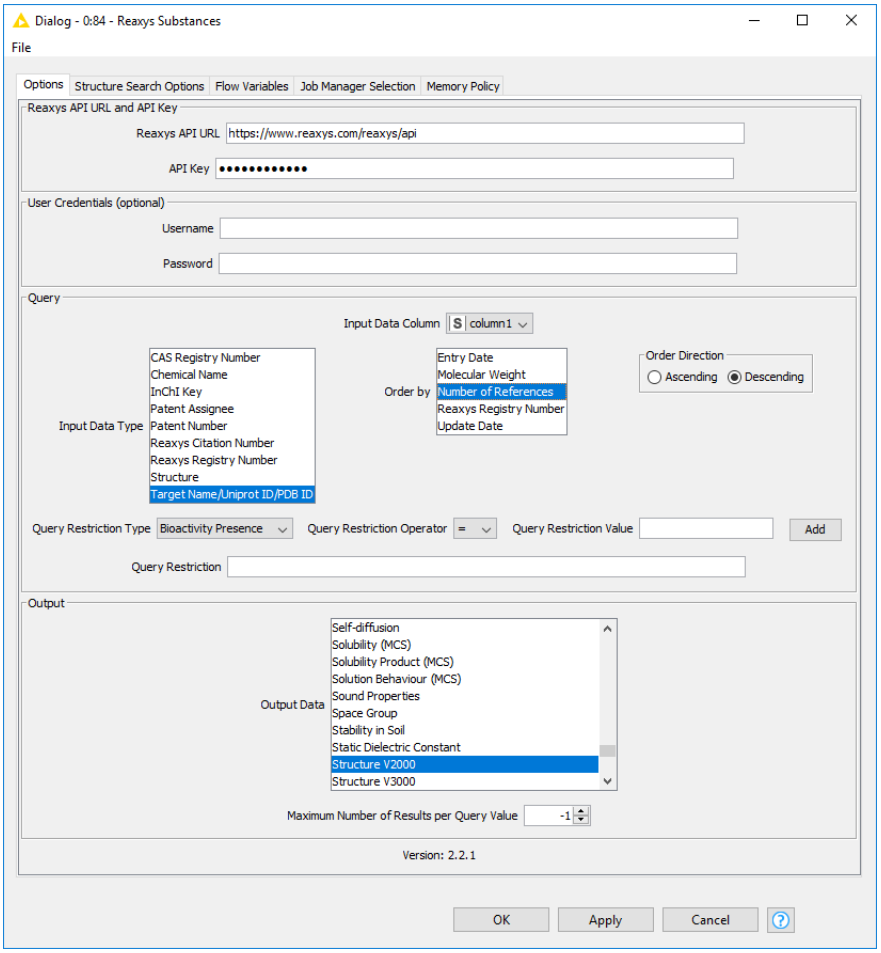


We have an input table with 1 target name.



We use the Bioactivities node and the Substances node.





The output is a table with this format:

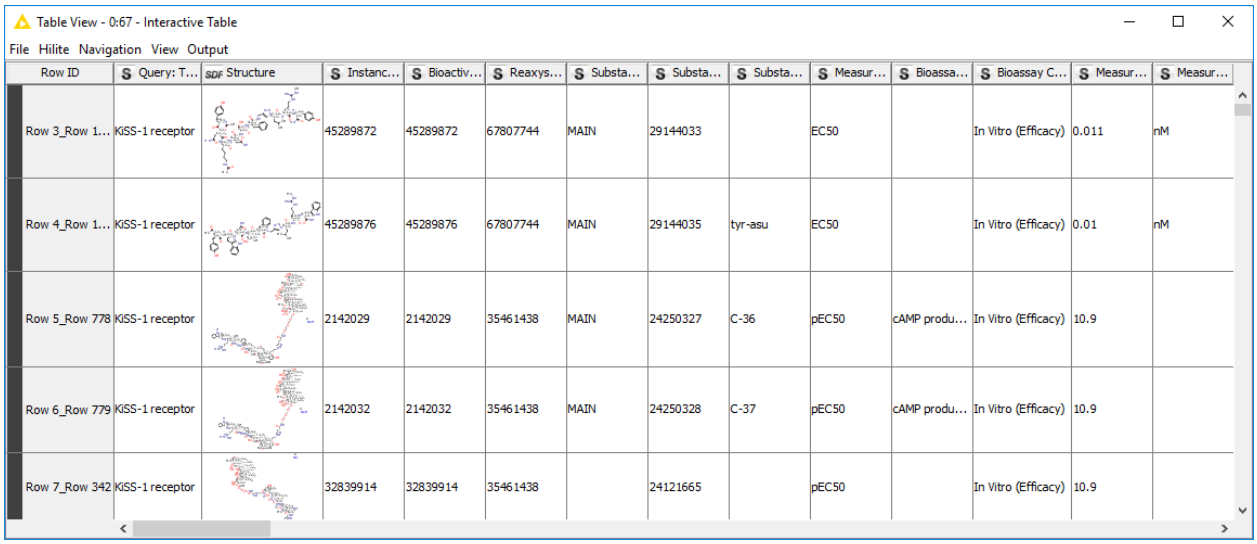


Table View - 0:67 - Interactive Table

File Hilite Navigation View Output

Row ID	Query: T...	Structure	Instanc...	Bioactiv...	Reaxys...	Substa...	Substa...	Substa...	Measur...	Bioassa...	Bioassay C...	Measur...	Measur...
Row 3_Row 1...	KiSS-1 receptor		45289872	45289872	67807744	MAIN	29144033		EC50		In Vitro (Efficacy)	0.011	nM
Row 4_Row 1...	KiSS-1 receptor		45289876	45289876	67807744	MAIN	29144035	tyr-asu	EC50		In Vitro (Efficacy)	0.01	nM
Row 5_Row 778	KiSS-1 receptor		2142029	2142029	35461438	MAIN	24250327	C-36	pEC50	cAMP produ...	In Vitro (Efficacy)	10.9	
Row 6_Row 779	KiSS-1 receptor		2142032	2142032	35461438	MAIN	24250328	C-37	pEC50	cAMP produ...	In Vitro (Efficacy)	10.9	
Row 7_Row 342	KiSS-1 receptor		32839914	32839914	35461438		24121665		pEC50		In Vitro (Efficacy)	10.9	

Result: 500 bioactivity datapoints and structures are retrieved.