

**Reaxys<sup>®</sup> KNIME nodes for KNIME 3.1**

# Reaxys KNIME Nodes Installation and User Guide

Version 1.0

## 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 high-quality, 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 3.1.

## 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”.

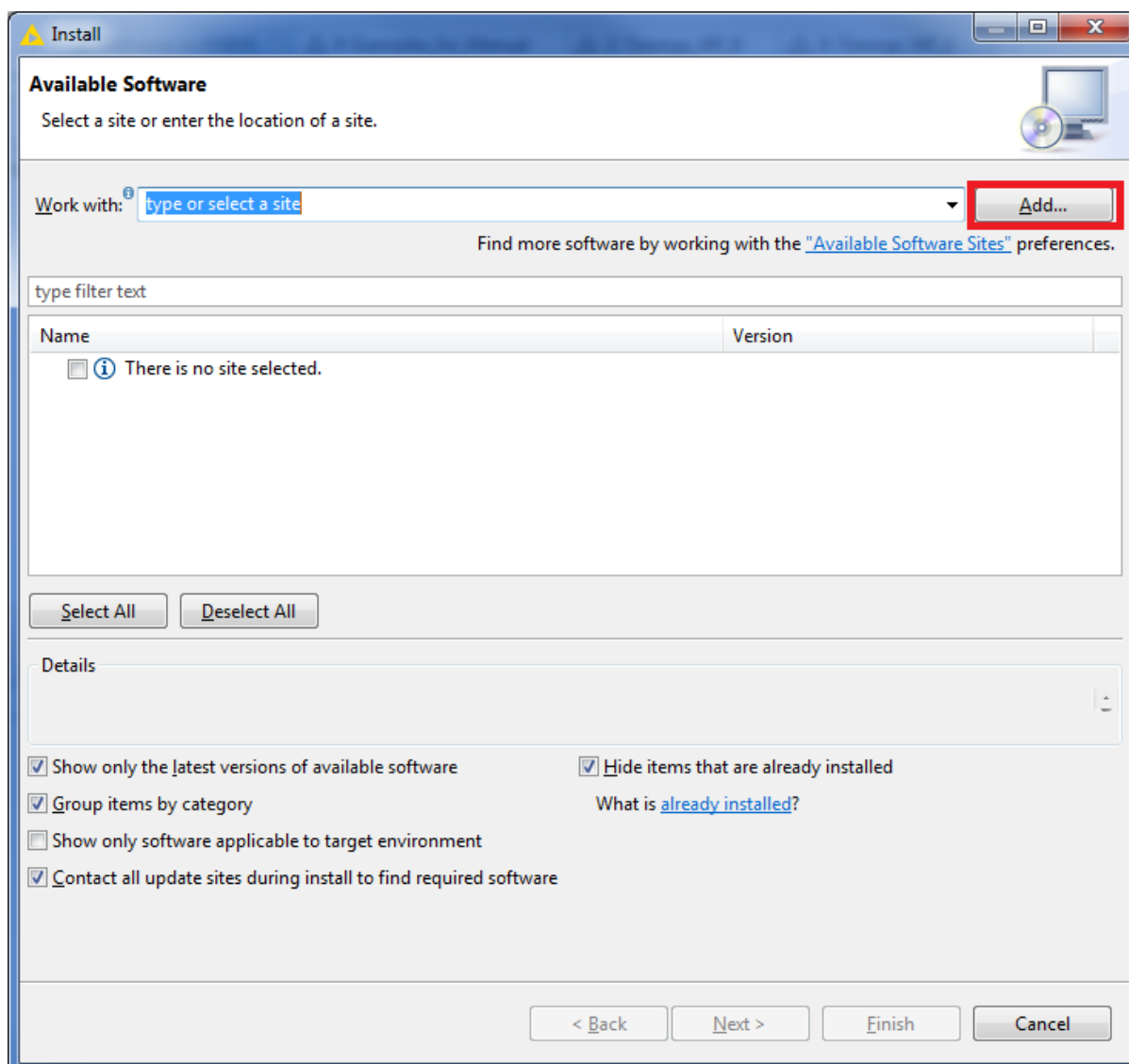


Figure 1. KNIME update site selection

2. Paste the URL of the Reaxys Knime node update site (<http://supportcontent.elsevier.com/Support Hub/Reaxys/knime/3.1/>) in the “Location” box, then click “OK”.

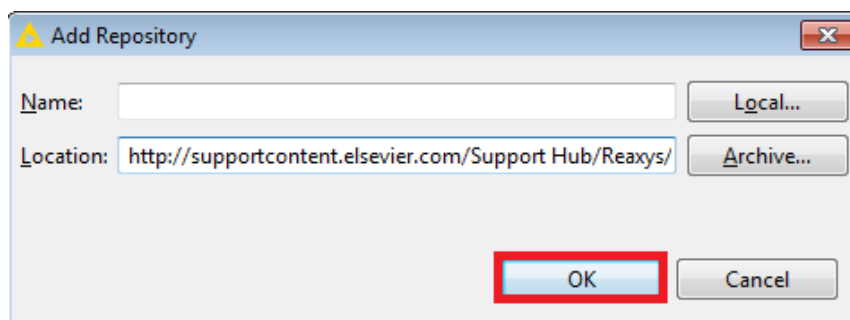


Figure 2. Reaxys KNIME node Location (URL)

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

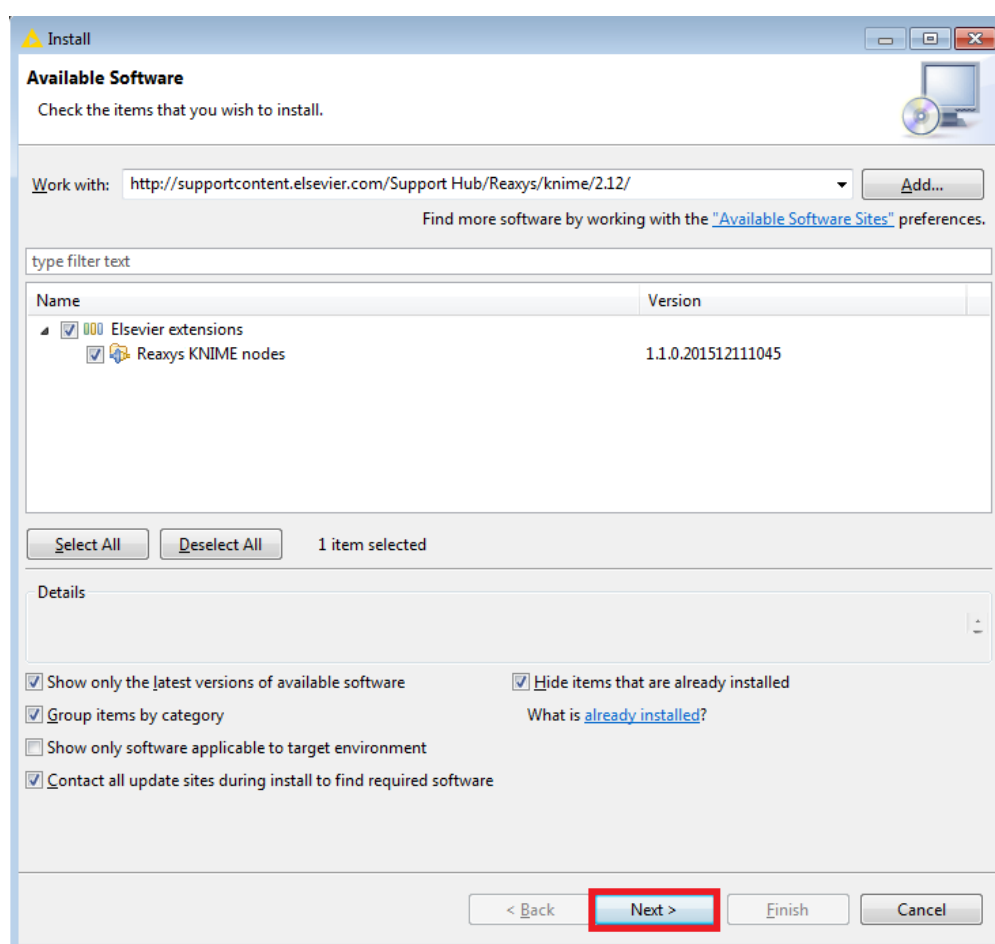


Figure 3. Reaxys KNIME nodes selection

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/3.1/ReaxysKnimeUpdate31.zip>.
2. Start KNIME, choose Help → Install new software, then click “Add”:

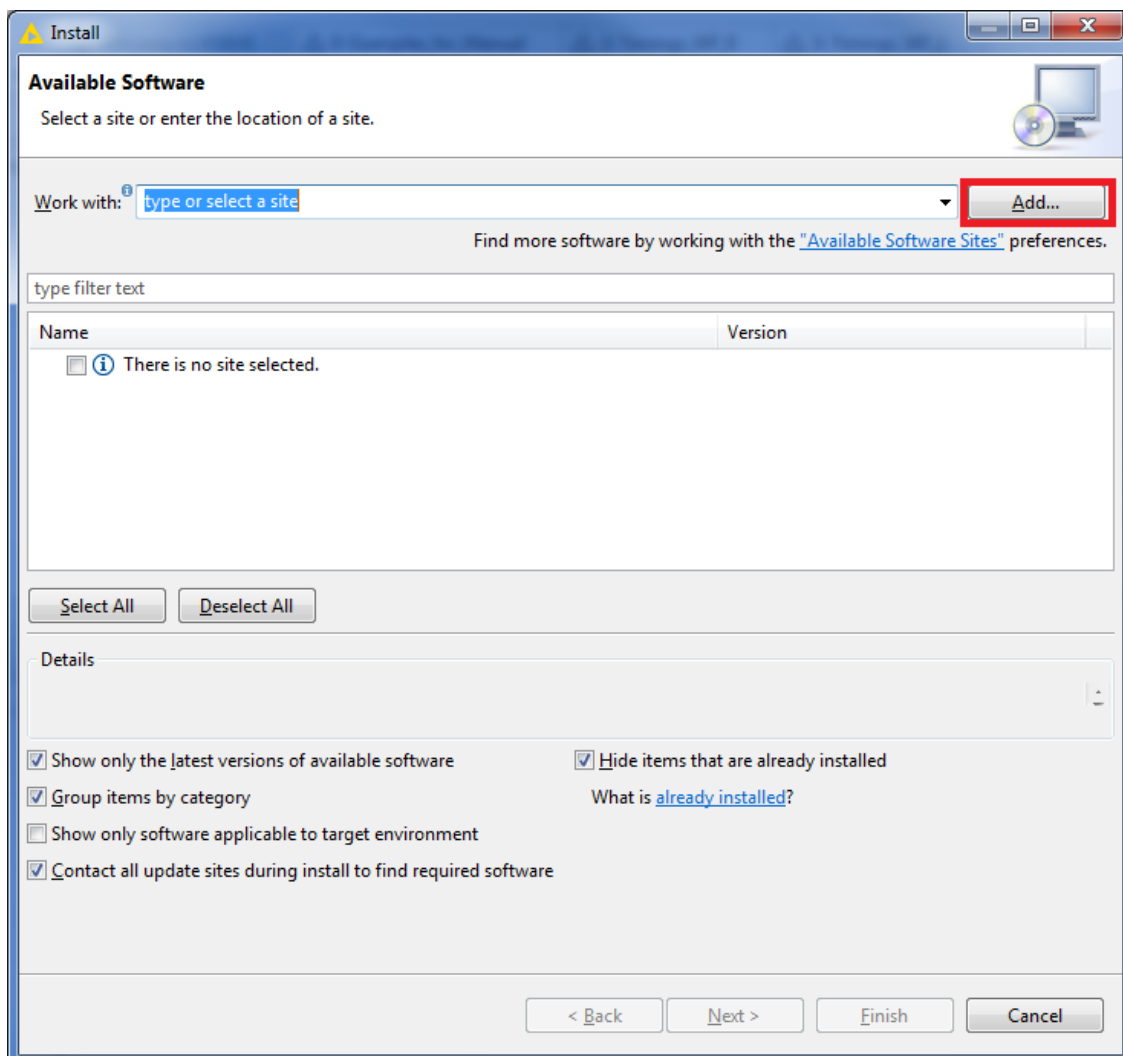


Figure 4. Reaxys KNIME archive file selection

3. Click the “Archive...” button:

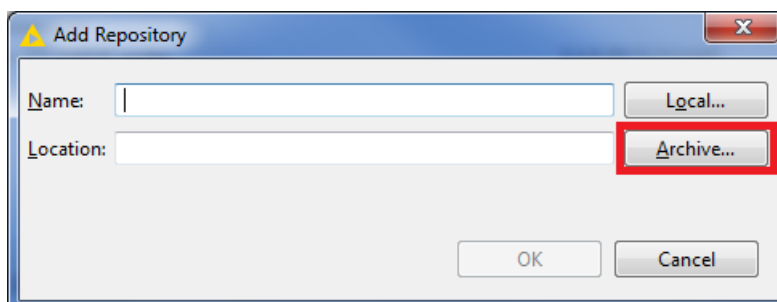


Figure 5. Reaxys KNIME archive file location

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.

### 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 6 – 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 6 – 2) and by defining the type of input data (Figure 6 – 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 6 – 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 6 – 5) or it can be created by typing into the query restriction free text field (Figure 6 – 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 6 – 7).

The user can further specify the maximum number of retrieved instances per query (Figure 6 – 8). If this value is set to less than zero, no limit on the number of retrieved instances per query will be imposed.



Dialog - 0:37 - Reaxys Substances

File

Options | Structure Search Options | Flow Variables | Memory Policy

Reaxys API URL and API Key

Reaxys API URL

API Key

User Credentials (optional)

Username

Password

Query

Input Data Column

Input Data Type

- CAS Registry Number
- Chemical Name
- InChI Key
- Patent Assignee
- Patent Number
- Reaxys Citation Number
- Reaxys Registry Number
- Structure
- Target Name

Order by

- Entry Date
- Molecular Weight
- Number of References
- Reaxys Registry Number
- Update Date

Order Direction

☐ Ascending ☒ Descending

Query Restriction Type  Query Restriction Operator  Query Restriction Value  Add

Query Restriction

Output

Output Data

- Abiotic Degradation, Hydrolysis
- Abiotic Degradation, Photolysis
- Adsorption (MCS)
- Association (MCS)
- Autoignition
- Azeotropes (MCS)
- Bioaccumulation, Biomagnification and Biomonitoring
- Biodegradation
- Boiling Point
- Boundary Surface Phenomena (MCS)

Maximum Number of Results per Query Value

Version: 1.1.8

OK Apply Cancel ?

Figure 6. 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 7 – 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 7 – 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 7 – 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 salts:** 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 7 – 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.

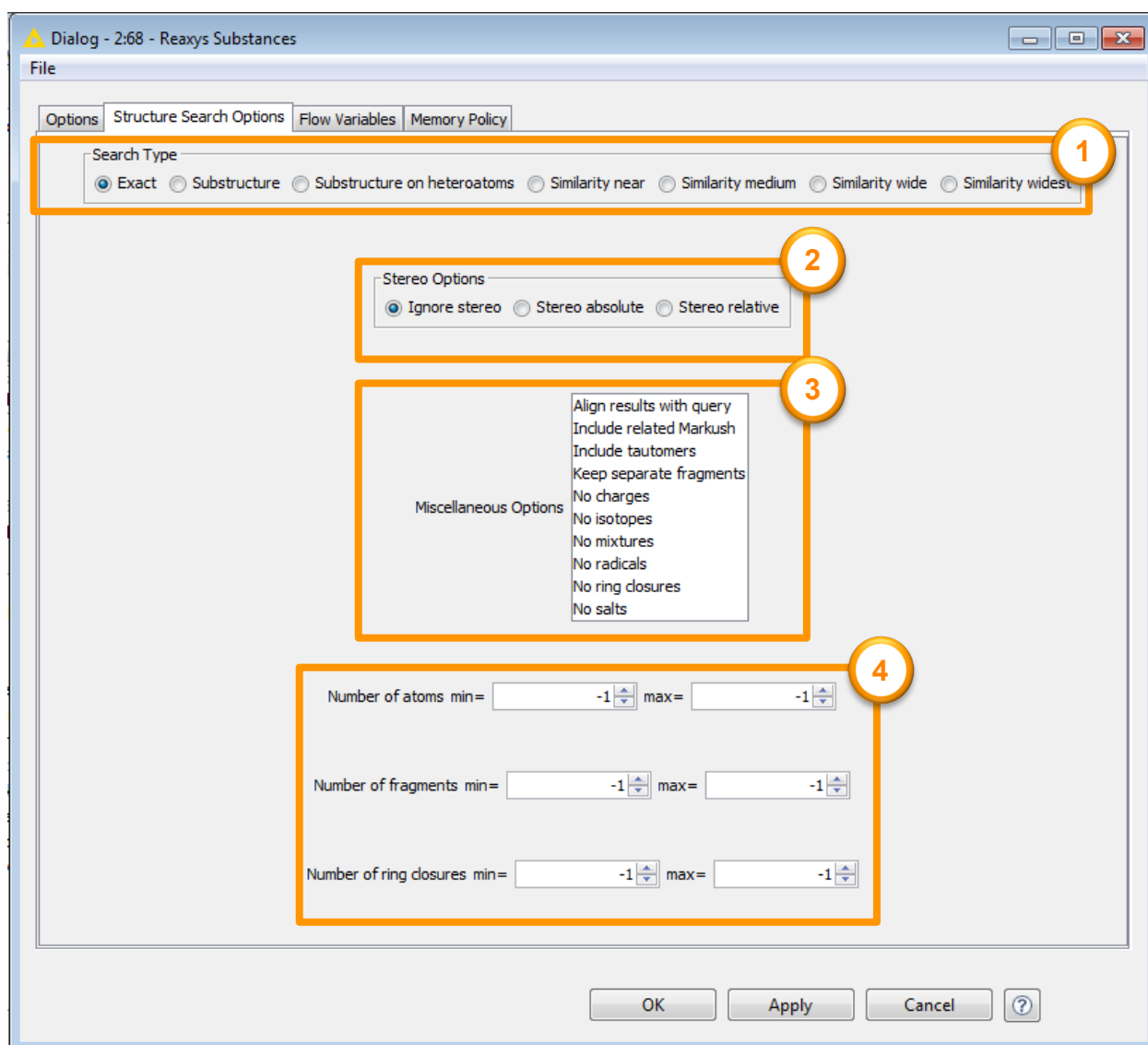


Figure 7. “Structure Search Options” tab in the configuration dialog of the Substances node. 1) The field for specifying the search type. 2) The field for specifying how stereochemistry is treated. 3) The field for defining other options (see text for details). 4) The field for specifying structural 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.

## 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/3.1/Reaxys KNIME workflow Examples/Reaxys KNIME workflow Examples 3\\_1.zip](http://supportcontent.elsevier.com/Support Hub/Reaxys/knime/3.1/Reaxys KNIME workflow Examples/Reaxys KNIME workflow Examples 3_1.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.

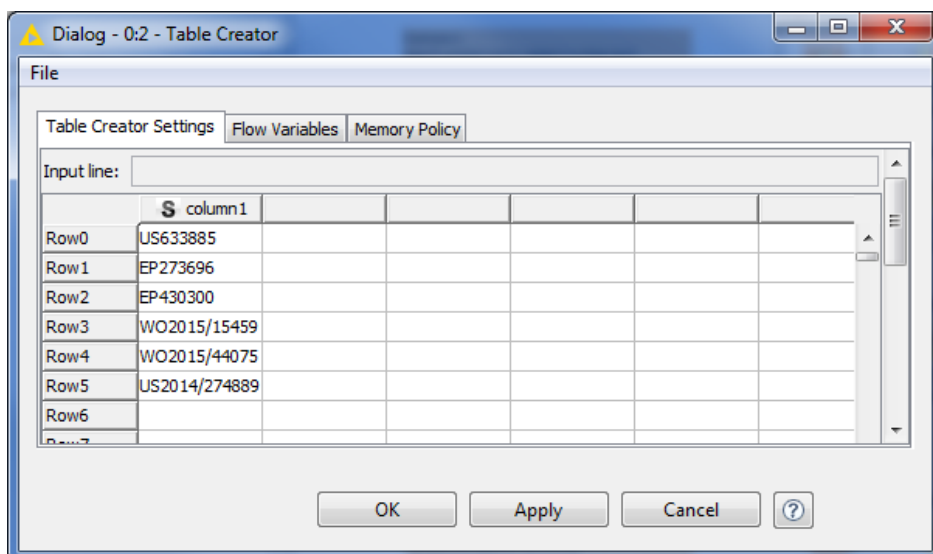
The overall workflow appears as follows:

Example 1:  
Search substances by patent number and  
retrieve substance identification data

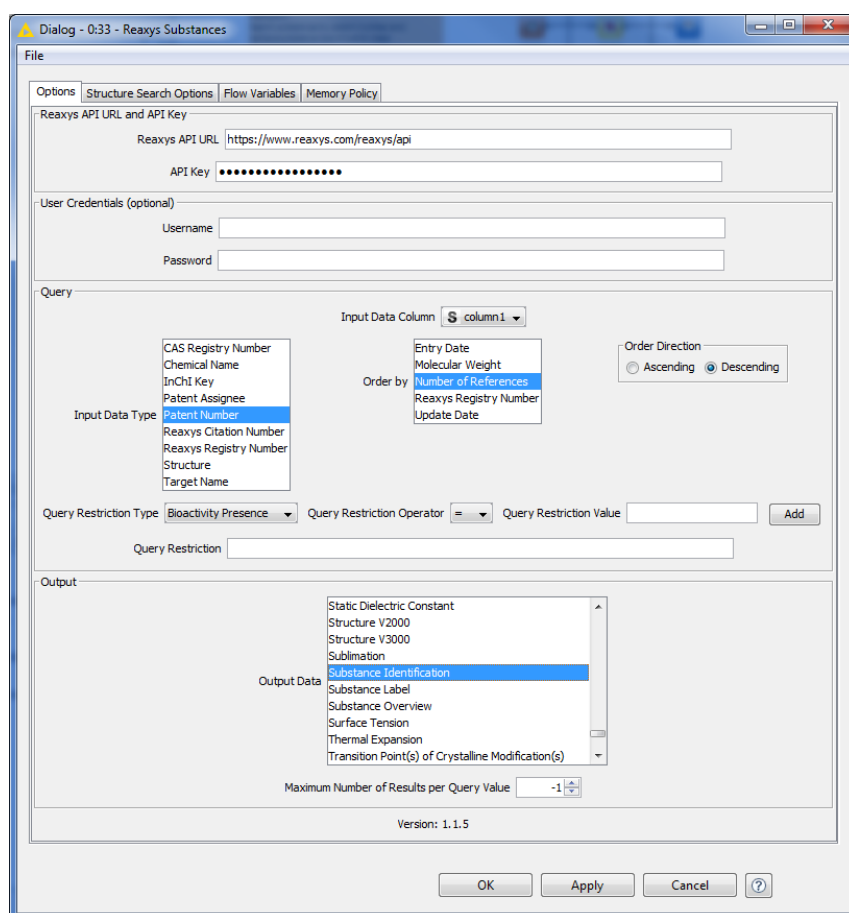
Table Creator Reaxys Substances Interactive Table



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	Instance Identifier: Reaxys Registry Number	CAS Registry Number	Chemical Name
Row 137	EP430300	26096785	136420-90-1	
Row 138	EP430300	26096788	136420-91-2	
Row 139	EP430300	26096789	136420-92-3	
Row 140	EP430300	26096790		
Row 141	EP430300	26096791	136420-94-5	
Row 142	EP430300	26096792	1349243-06-6	
Row 143	EP430300	26096794	1348229-65-1	
Row 144	EP430300	26096795	136420-98-9	
Row 145	WO2015/15459	561195	69975-86-6	2-(7-theophyllinemethyl)-1,3-dioxolane[Doxofylline]
Row 146	WO2015/44075	969135	74-88-4	methyl iodide [iodomethane] [methyl iodide]
Row 147	WO2015/44075	635680	67-64-1	acetone [dimethyl ketone] [propanone]
Row 148	WO2015/44075	605461	536-74-3	phenylacetylene [ethynylbenzene] [PhC≡CH] [phenylethyne] [phenyl acetylene]
Row 149	WO2015/44075	1787583	32315-10-9	bis(trichloromethyl) carbonate [triphosgene] [triphosgene]
Row 150	WO2015/44075	606965	134-20-3	2-carbomethoxyaniline [methyl Anthranilate]
Row 151	WO2015/44075	1211945	64269-79-0	Carbonyldimidazole [carbonyldimidazole]
Row 152	WO2015/44075	2081106	42016-93-3	2-chloro-4-iodo-aniline [4-iodo-2-chloro-aniline] [2-chloro-4-iodoaniline] [4-iodo-2-chloro-aniline]
Row 153	WO2015/44075	383687	16135-36-7	methyl 4-aminopyridine-3-carboxylate [4-amino-nicotinic acid methyl ester] [4-aminor]
Row 154	WO2015/44075	14217835	637336-53-9	4-amino-1-methyl-1H-pyrazole-3-carboxylic acid methyl ester [methyl 4-amino-1-me
Row 155	WO2015/44075	14529252	877603-44-6	2-chloro-4-(2-phenylethynyl)aniline [2-chloro-4-(2-phenyl-1-ethynyl)aniline]
Row 156	WO2015/44075	28037395		2,6-difluoro-4-(2-phenylethynyl)phenylamine [2,6-difluoro-4-phenylethynylphenyla
Row 157	WO2015/44075	28037408		2-chloro-6-fluoro-4-[2-(3-pyridyl)ethynyl]aniline
Row 158	WO2015/44075	21315164	17318-50-2	methyl 2-(isopropylamino)benzoate
Row 159	WO2015/44075	28037399		2,6-difluoro-4-[2-(3-pyridyl)ethynyl]aniline [2,6-difluoro-4-pyridin-3-ylethynylpheny
Row 160	WO2015/44075	28037377		3-[2-chloro-4-(2-phenylethynyl)phenyl]-1H-pyrido[3,2-d]pyrimidine-2,4-dione
Row 161	WO2015/44075	28037378		3-(2-chloro-4-iodophenyl)-1H-pyrido[4,3-d]pyrimidine-2,4-dione

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

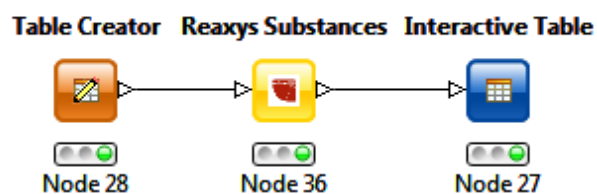


## 2. Search substances by chemical name and retrieve substance property data

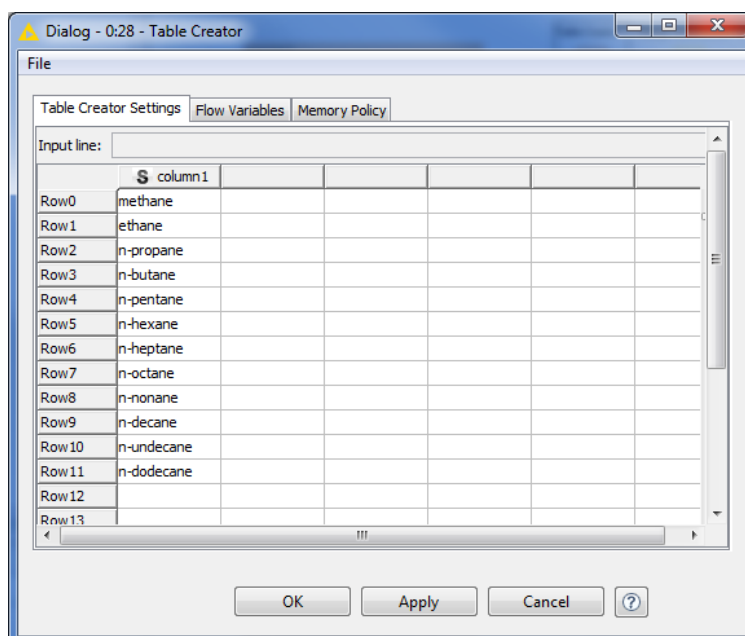
In this example, we search substances by chemical name and retrieve boiling point data for all the identified substances.

The overall workflow appears as follows:

**Example 2:**  
Search substances by chemical name and retrieve substance property (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:

Dialog - 0:36 - Reaxys Substances

File

Options | Structure Search Options | Flow Variables | Memory Policy

Reaxys API URL and API Key

Reaxys API URL:

API Key:

User Credentials (optional)

Username:

Password:

Query

Input Data Column:

Input Data Type: 

- CAS Registry Number
- Chemical Name**
- InChi Key
- Patent Assignee
- Patent Number
- Reaxys Citation Number
- Reaxys Registry Number
- Structure
- Target Name

Order by: 

- Entry Date
- Molecular Weight
- Number of References**
- Reaxys Registry Number
- Update Date

Order Direction: ☐ Ascending ☒ Descending

Query Restriction Type:  Query Restriction Operator:  Query Restriction Value:

Add

Query Restriction:

Output

Output Data: 

- Abiotic Degradation, Hydrolysis
- Abiotic Degradation, Photolysis
- Adsorption (MCS)
- Association (MCS)
- Autoignition
- Azeotropes (MCS)
- Bioaccumulation, Biomagnification and Biomonitoring
- Biodegradation
- Boiling Point**
- Boundary Surface Phenomena (MCS)

Maximum Number of Results per Query Value:

Version: 1.1.5

OK Apply Cancel ?

The output is a table with this format:

Table View - 0:27 - Interactive Table

File | Highlight | Navigation | View | Output

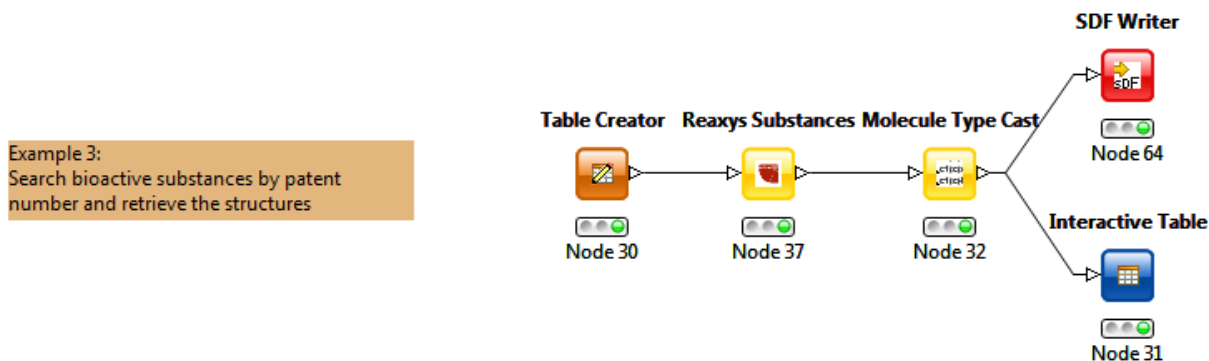
Row ID	Query: Chemical Name	Instance identifier: Reaxys Registry Number	Reaxys Citation Number	Boiling Point	Pressure	Entry Date	Location	Comment
Row 73	ethane	1730716	1304793	-107.9	288	2007/10/05		
Row 74	ethane	1730716	1304793	-100.4	385	2007/10/05		
Row 75	ethane	1730716	1304793	-97.1	454	2007/10/05		
Row 76	ethane	1730716	1304793	-94	546	2007/10/05		
Row 77	ethane	1730716	1304793	-91.5	630	2007/10/05		
Row 78	ethane	1730716	1304793	-89.3	713	2007/10/05		
Row 79	ethane	1730716	1304793	-88.5	760	2007/10/05		
Row 80	ethane	1730716	1294447	-84.1	749	2007/10/05		
Row 81	ethane	1730716	1290628	-93	760	2007/10/05		
Row 82	ethane	1730716	1290628	0	18088	2007/10/05		
Row 83	ethane	1730716	1290628	23.5	30704	2007/10/05		
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 37...	-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		Druckbaengigke
Row 89	n-butane	969129	1314973 1327560 1334638 13...	-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		

Result: 399 records of boiling point data are retrieved.

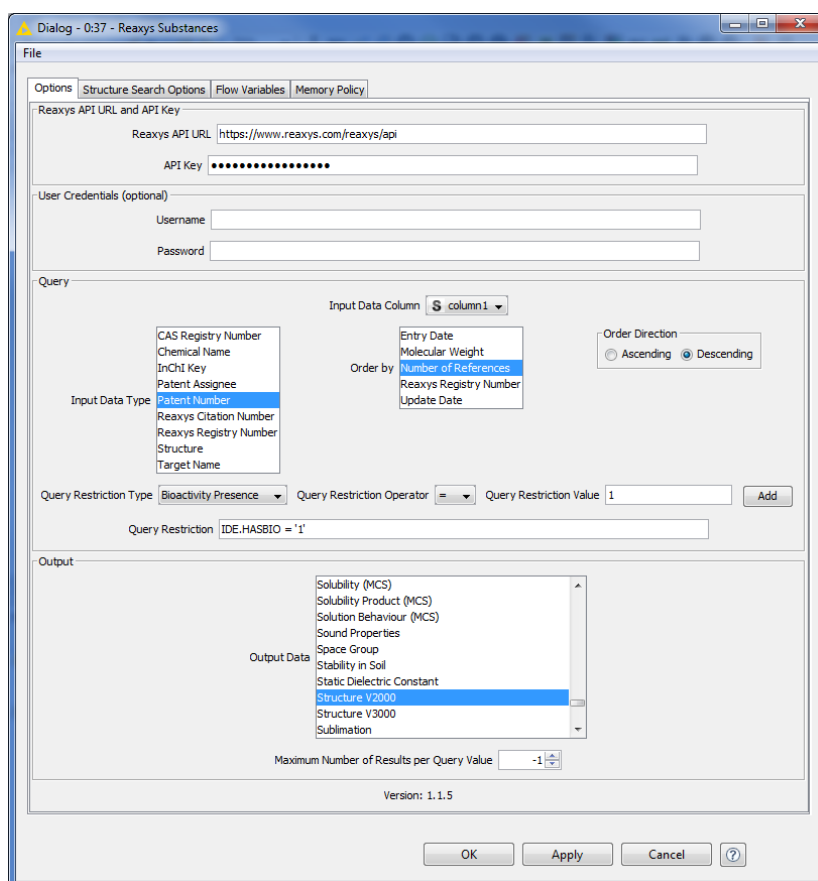
### 3. Search bioactive substances by patent number and retrieve the structures

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.

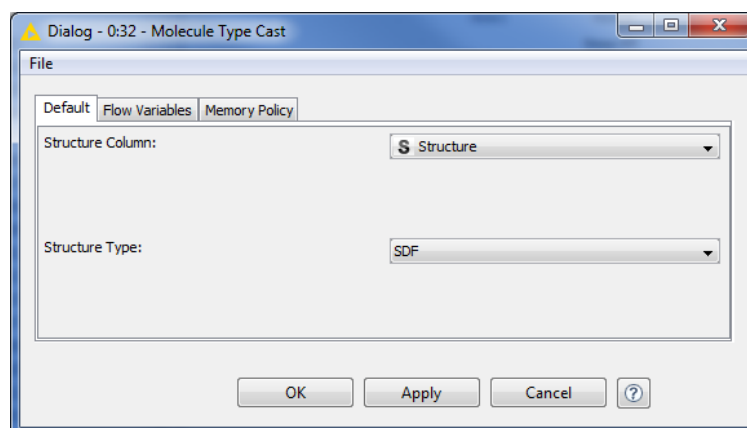
The overall workflow appears as follows:



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.

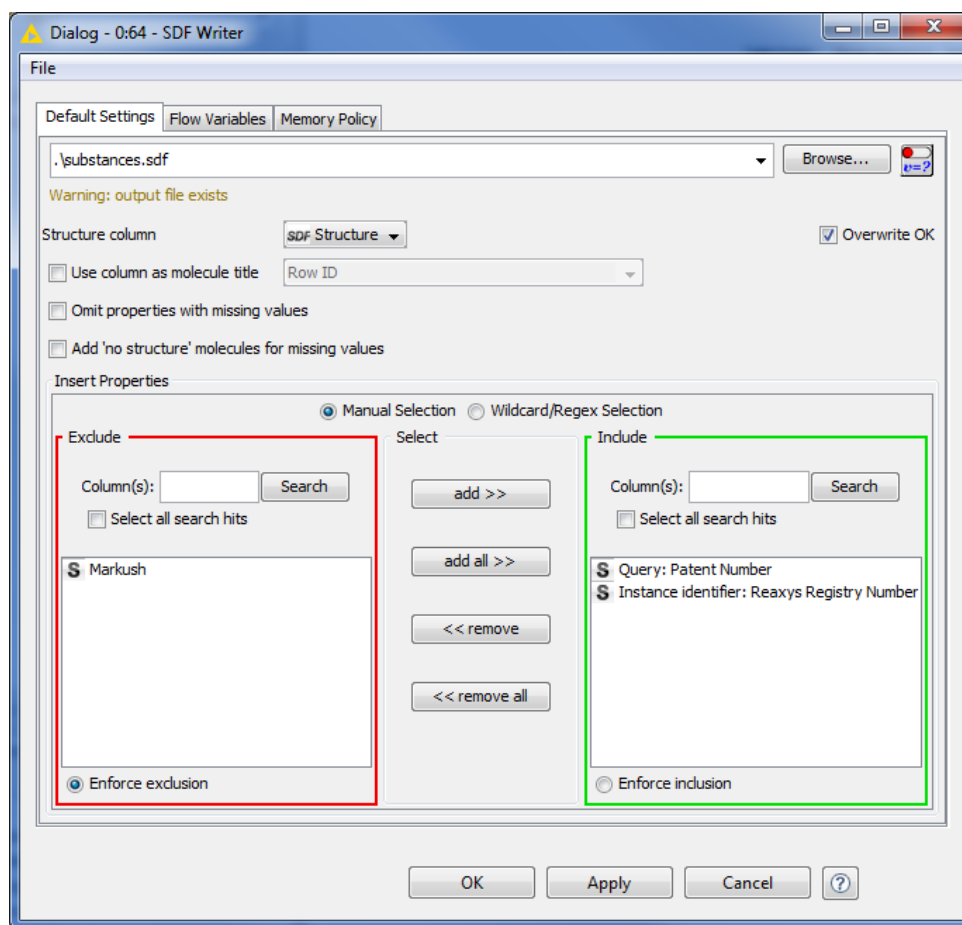


The output is a table with this format:

The screenshot shows a window titled "Table View - 0:31 - Interactive Table". It has a menu bar with "File", "Hilite", "Navigation", "View", and "Output". The table below has five columns: "Row ID", "S Query: Patent Number", "S Instance identifier: Reaxys Registry Number", "sdf Structure", and "S Markush". The table contains six rows of data, with the first three rows highlighted in grey. Each row in the "sdf Structure" column contains a chemical structure image.

Row ID	S Query: Patent Number	S Instance identifier: Reaxys Registry Number	sdf Structure	S Markush
Row 58	EP273696	25545612		
Row 59	EP273696	25545613		
Row 60	EP273696	25545616		
Row 61	EP430300	261704		
Row 62	EP430300	235783		
Row 63	EP430300	26096762		

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.



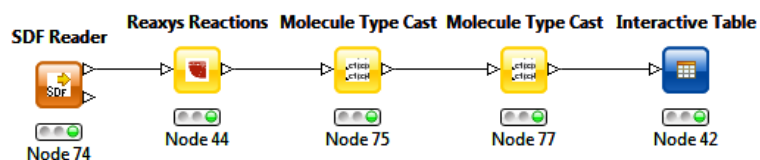
**Result: 174 structures are retrieved and exported.**

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

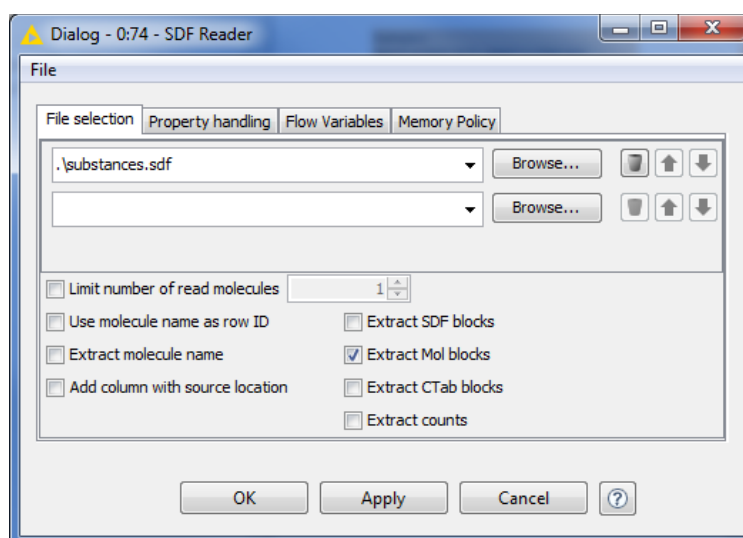
In this example, we use the 174 substance structures that have been exported from the previous workflow as a structure query. We search preparations and retrieve reaction structures.

The overall workflow appears as follows:

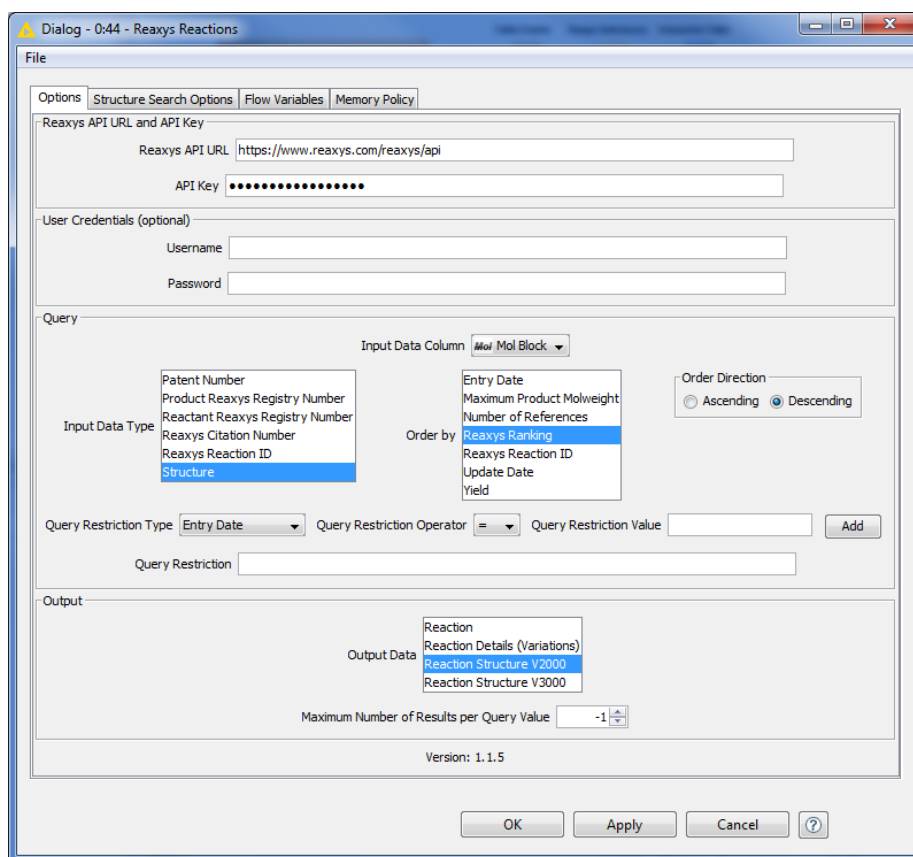
Example 4:  
Search preparations for a set of structures and  
retrieve the reaction structures



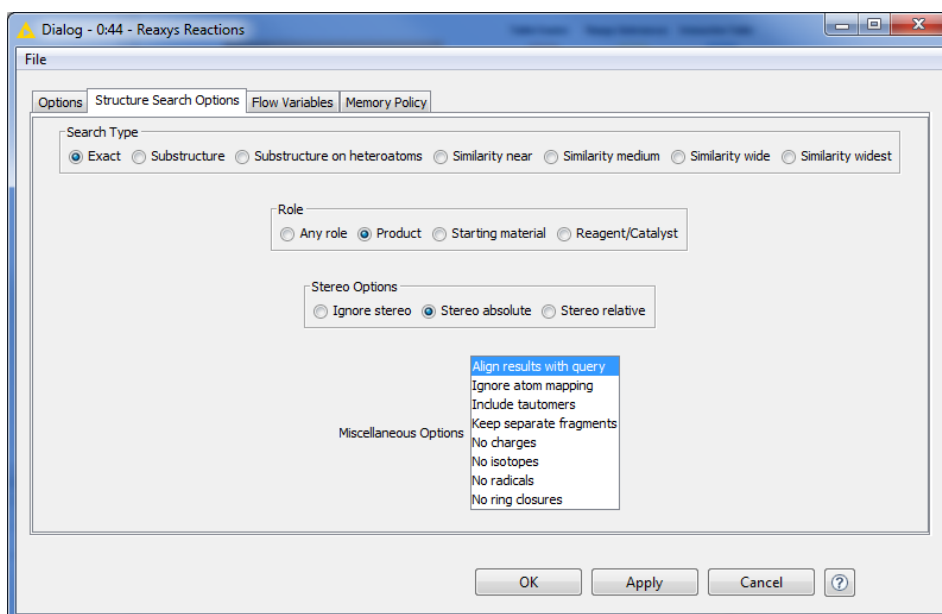
Reading the SDF file, we see:



The settings for the Reaxys node are:



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



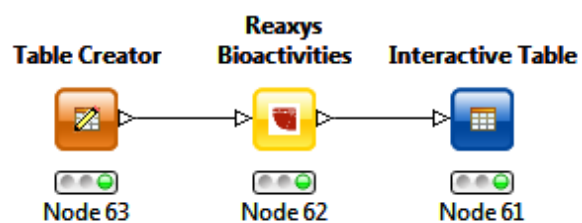
**Result: 125 reaction structures are retrieved.**

### 5. Search for bioactivities by patent number

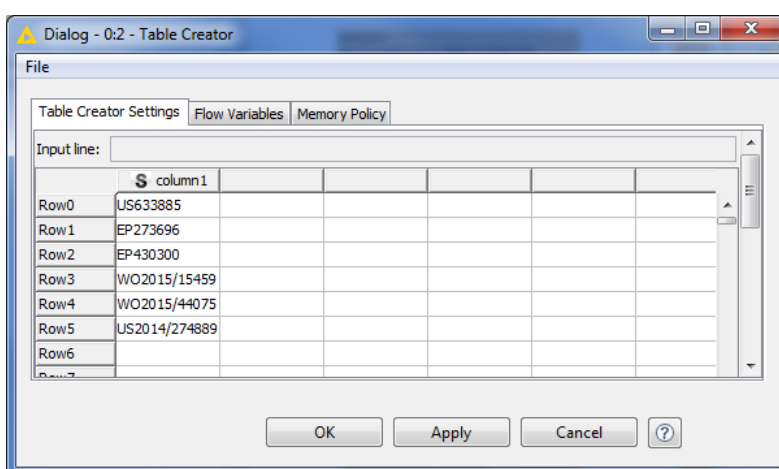
In this example, we search for bioactivities associated with a given patent number.

The overall workflow appears as follows:

Example 5:  
Search bioactivities by patent number



We have an input table with 6 patent numbers.





We use the Bioactivities node.

Dialog - 2:62 - Reaxys Bioactivities

File

Options | Structure Search Options | Flow Variables | Memory Policy

Reaxys API URL and API Key

Reaxys API URL:

API Key:

User Credentials (optional)

Username:

Password:

Query

Input Data Column:

Input Data Type:

Order by:

Order Direction: ☐ Ascending ☒ Descending

Query Restriction Type:  Query Restriction Operator:  Query Restriction Value:

Query Restriction:

Output

Output Data:

Maximum Number of Results per Query Value:

Version: 1.1.8

OK Apply Cancel ?

The output is a table with this format:

Table View - 2:61 - Interactive Table

File | Hilit | Navigation | View | Output

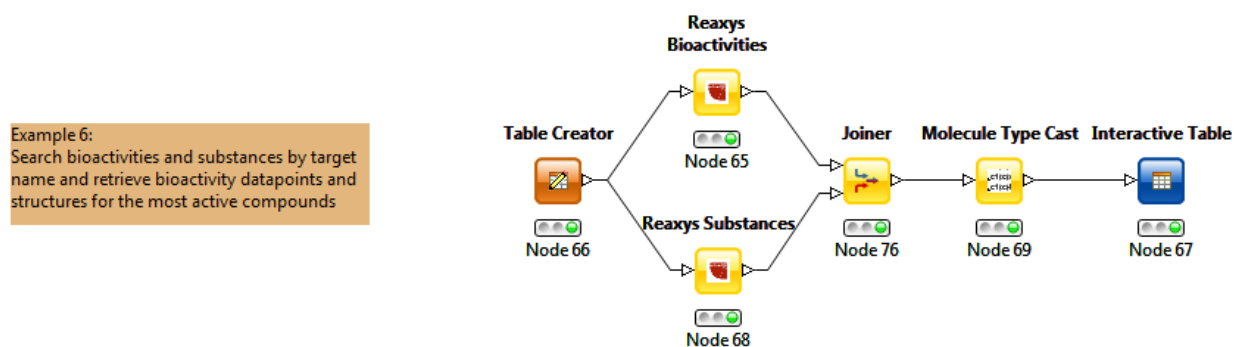
Row ID	Highest...	Substra...	Substra...	Measur...	Value T...	Bioassa...	Bioassa...	Bioassay ...	Animal ...	Bi
Row 276				% Decrease				In Vivo (Animal...		
Row 277	Marketed			% Decrease				In Vivo (Animal...		
Row 278				% increase				In Vivo (Animal...		
Row 279	Marketed			% increase				In Vivo (Animal...		
Row 280	Marketed			Qualitative				In Vivo (Animal...		
Row 281	Marketed			Qualitative				In Vivo (Animal...		
Row 282	Marketed			Qualitative				In Vivo (Animal...		
Row 283	Marketed			% increase				In Vivo (Animal...		
Row 284				Tested				In Vivo (Animal...		
Row 285	Marketed			Tested				In Vivo (Animal...		
Row 286				Tested				In Vivo (Animal...		
Row 287	Marketed			Tested				In Vivo (Animal...		
Row 288				Tested				In Vivo (Animal...		
Row 289	Marketed			Tested				In Vivo (Animal...		
Row 290	Marketed			% change				In Vivo (Animal...		
Row 291	Marketed			% change				In Vivo (Animal...		
Row 292	Marketed			% change				In Vivo (Animal...		
Row 293	Marketed			% change				In Vivo (Animal...		
Row 294	Marketed			% change				In Vivo (Animal...		
Row 295	Marketed			% change				In Vivo (Animal...		
Row 296	Marketed			% Decrease				In Vivo (Animal...		
Row 297	Marketed			Amount				In Vivo (Animal...		
Row 298	Marketed			Amount				In Vivo (Animal...		
Row 299	Marketed			% Decrease				In Vivo (Animal...		
Row 300	Marketed			% Decrease				In Vivo (Animal...		

Result: 300 bioactivity datapoints are retrieved.

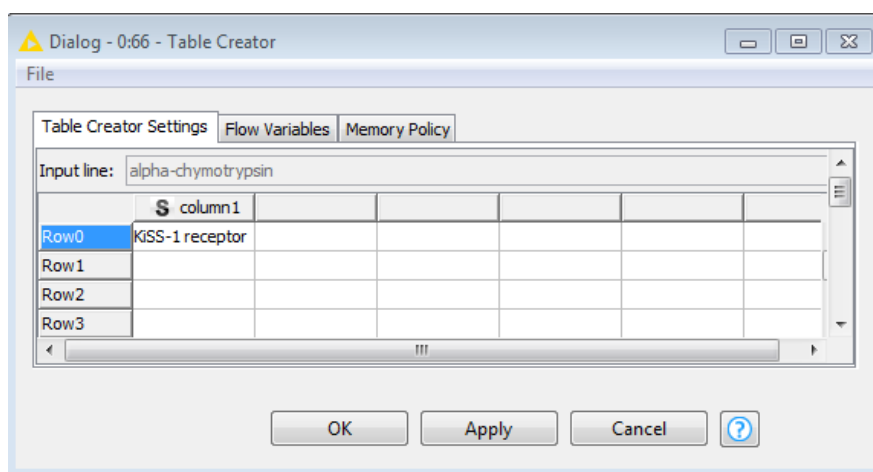
## 6. Search bioactivities and substances by target name and retrieve bioactivity datapoints and structures for the most active compounds

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.

Dialog - 2:65 - Reaxys Bioactivities

File

Options | Structure Search Options | Flow Variables | Memory Policy

Reaxys API URL and API Key

Reaxys API URL:

API Key:

User Credentials (optional)

Username:

Password:

Query

Input Data Column:

Input Data Type: 

- Bioactivity Datapoint ID
- Patent Number
- Reaxys Citation Number
- Reaxys Registry Number
- Structure
- Target Name

Order by: 

- Bioactivity Datapoint ID
- Entry Date
- Measurement pX
- Update Date

Order Direction: ☐ Ascending ☒ Descending

Query Restriction Type:  Query Restriction Operator:  Query Restriction Value:

Add

Query Restriction:

Output

Output Data:

Maximum Number of Results per Query Value:

Version: 1.1.8

OK Apply Cancel ?

Dialog - 2:68 - Reaxys Substances

File

Options | Structure Search Options | Flow Variables | Memory Policy

Reaxys API URL:

API Key:

User Credentials (optional)

Username:

Password:

Query

Input Data Column:

Input Data Type: 

- CAS Registry Number
- Chemical Name
- InChI Key
- Patent Assignee
- Patent Number
- Reaxys Citation Number
- Reaxys Registry Number
- Structure
- Target Name

Order by: 

- Entry Date
- Molecular Weight
- Number of References
- Reaxys Registry Number
- Update Date

Order Direction: ☐ Ascending ☒ Descending

Query Restriction Type:  Query Restriction Operator:  Query Restriction Value:

Add

Query Restriction:

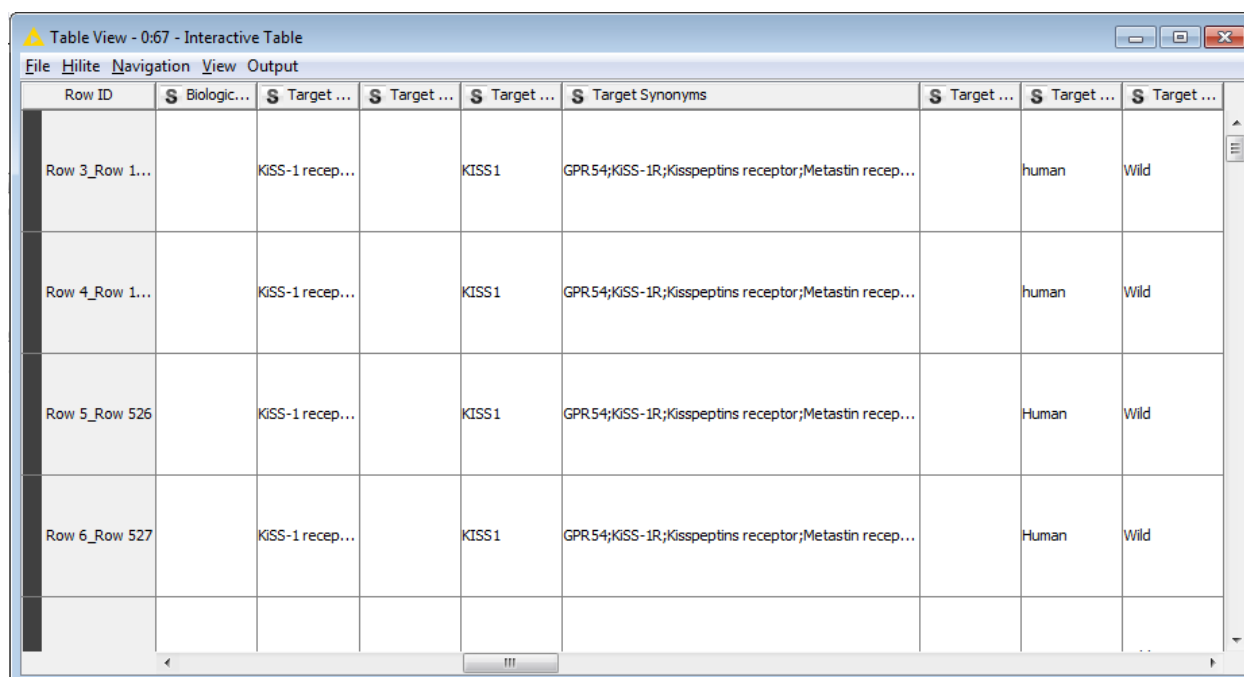
Output

Output Data: 

- Stability in Soil
- Static Dielectric Constant
- Structure V2000
- Structure V3000
- Sublimation
- Substance Identification
- Substance Label
- Substance Overview

OK Apply Cancel ?

The output is a table with this format:



Row ID	S Biologic...	S Target ...	S Target ...	S Target ...	S Target Synonyms	S Target ...	S Target ...	S Target ...
Row 3_Row 1...		KISS-1 recep...		KISS1	GPR54;KISS-1R;Kisspeptins receptor;Metastin recep...		human	Wild
Row 4_Row 1...		KISS-1 recep...		KISS1	GPR54;KISS-1R;Kisspeptins receptor;Metastin recep...		human	Wild
Row 5_Row 526		KISS-1 recep...		KISS1	GPR54;KISS-1R;Kisspeptins receptor;Metastin recep...		Human	Wild
Row 6_Row 527		KISS-1 recep...		KISS1	GPR54;KISS-1R;Kisspeptins receptor;Metastin recep...		Human	Wild

**Result: 1000 bioactivity datapoints and structures are retrieved.**

REAXYS is a trademark of RELX Intellectual Properties SA, used under license. Copyright © 2016, Elsevier Information Systems GmbH. All rights reserved.