ELSEVIER

Reaxys® Pipeline Pilot components for Pipeline Pilot 9.5

Pipeline Pilot 18.1 supported

Reaxys Pipeline Pilot Components v1.3.10 Installation and User Guide

Version 1.2

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 Pipeline Pilot components 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 components 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 Pipeline Pilot components. The components are free-licensed software. To use the components productively, a valid subscription to Reaxys and/or Reaxys Medicinal Chemistry and to the relevant API is needed.

About BIOVIA Pipeline Pilot

Built on the BIOVIA Foundation, Pipeline Pilot enables scientists to rapidly create, test and publish scientific services that automate the process of accessing, analyzing and reporting scientific data, either for the scientist's personal use or for sharing across the scientific community. Using Pipeline Pilot, scientists, researchers, engineers and analysts with little or no software development experience can create scientific protocols that can be executed through a variety of interfaces including BIOVIA Web Port, other BIOVIA solutions such as BIOVIA Electronic Lab Notebook, Isentris and Chemical Registration, third-party applications such as Microsoft SharePoint, and customerdeveloped applications. These protocols aggregate and provide immediate access to volumes of disparate research data that is in separate silos. They automate the scientific analysis of the data and enable researchers to rapidly explore, visualize and report results.

Pipeline Pilot enables organizations to:

- Save time by automating tedious manual and error-prone tasks, enabling scientists to spend more time on innovation
- Lower costs by avoiding unnecessary experiments through improved access to scientific information locked in disparate silos
- Improve operating efficiency by automating common scientific workflows and by empowering scientists to address their constantly changing requirements without relying on IT

The Reaxys Pipeline Pilot Components

The Reaxys Pipeline Pilot components are free-licensed, closed-source software. To use the components 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 Pipeline Pilot components.
- 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 Pipeline Pilot components.
- 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 Pipeline Pilot components.

The Reaxys Substances Pipeline Pilot component 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 Pipeline Pilot component allows the retrieval of all reaction data in Reaxys. This includes reaction detail (variation) information and reaction structures.

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

The Reaxys Citations Pipeline Pilot component 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 Pipeline Pilot components requires a stable internet connection.

The Reaxys Pipeline Pilot components have been created with Pipeline Pilot 9.5 and support Pipeline Pilot Server versions 9.5 through 18.1.

Installation and updating

To install the latest release of the Reaxys Pipeline Pilot components, please follow this procedure.

Install the Reaxys Pipeline Pilot package using this link:

http://supportcontent.elsevier.com/Support Hub/Reaxys/pipelinepilot/9.5/ReaxysPipelinepilotUpdate.zip

The pkgutil tool is used to install the Reaxys Pipeline Pilot components package and is described in Chapter 6 of the *Application Packaging Guide* of Pipeline Pilot. Installation instructions are detailed below. An account with the appropriate privileges must be used.

Linux

1. Log on to the Pipeline Pilot Server.
2. Set up the environment.

    > cd <Pipeline Pilot install Directory>/linux_bin

    > source ./ppvars.sh

3. List the installed packages

    > pkgutil -l

4. If Reaxys Package (elsevier/reaxysapi) is already listed, remove it.

    > pkgutil -u elsevier/reaxysapi

    > rm -rf ../apps/elsevier/

5. Unzip the Reaxys Package and place it into the apps directory on the server.

    > unzip -d ../apps/ /path/to/ReaxysPipelinepilotUpdate.zip

6. Install the Reaxys Package.

    > pkgutil -i elsevier/reaxysapi

Windows

1. Log on to the Pipeline Pilot Server.
2. Set up the environment (note that the command prompt may require Administrator privileges depending on your Pipeline Pilot setup).

    > cd <Pipeline Pilot install Directory>\bin

3. List the installed packages.

    > pkgutil.exe -l

4. If Reaxys Package (elsevier/reaxysapi) is already listed, remove it.

    > pkgutil.exe -u elsevier/reaxysapi

    > rmdir /S ..\apps\elsevier

5. Unzip the Reaxys Package and place it into the <Pipeline Pilot install Directory>\apps directory on the server.

6. Install the Reaxys Package

    > pkgutil.exe -i elsevier/reaxysapi

When the clients are restarted the components will appear under Components > Elsevier.

Verify installation

1. Start the Pipeline Pilot Client.

2. Select the Components tab.

3. Verify that the Elsevier directory is present.

4 Verify that the directory contains the four Reaxys Pipeline Pilot components (Bioactivities, Citations, Reactions and Substances).

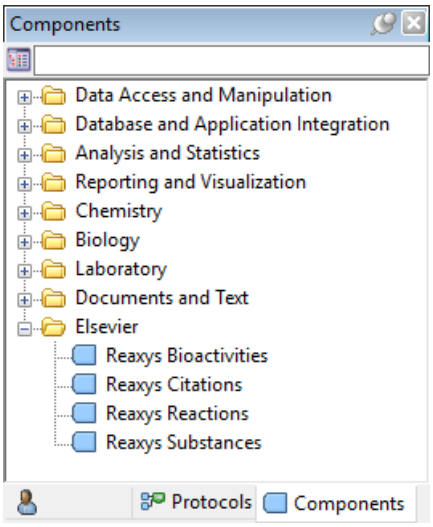


Figure 1. Reaxys Pipeline Pilot component installation

The Reaxys Pipeline Pilot components

General description

The package includes four components—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 components.

In general, the Reaxys Pipeline Pilot components 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 component to create a query.

The compnents access the Reaxys database server via the Reaxys API using the credentials given in the configuration dialog of the component. The components 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 component, 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 component, 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.

Input: Parameters of the Pipeline Pilot components

In the Reaxys Connection section, the user needs to provide the credentials for accessing the Reaxys database server. This section also allows the user to configure the connection using a proxy server as well as debug settings (Figure 2 – 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 (Input data Column) and by defining the type of the input data (Input data Type). The user can specify how the instances retrieved per query are sorted (Figure 2 – 2).

The user can further add a query restriction that is applied to each query. The query restriction can be created by typing into the query restriction free text field (Figure 2 – 2).

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. In the “Output Options”, the user also needs to specify which data should be retrieved for the substance, reaction, bioactivity or citation instances matching the query.

The user can specify the maximum number of retrieved instances per query. If this value is set to less than zero, no limit on the number of retrieved instances per query will be imposed (Figure 2 – 3).

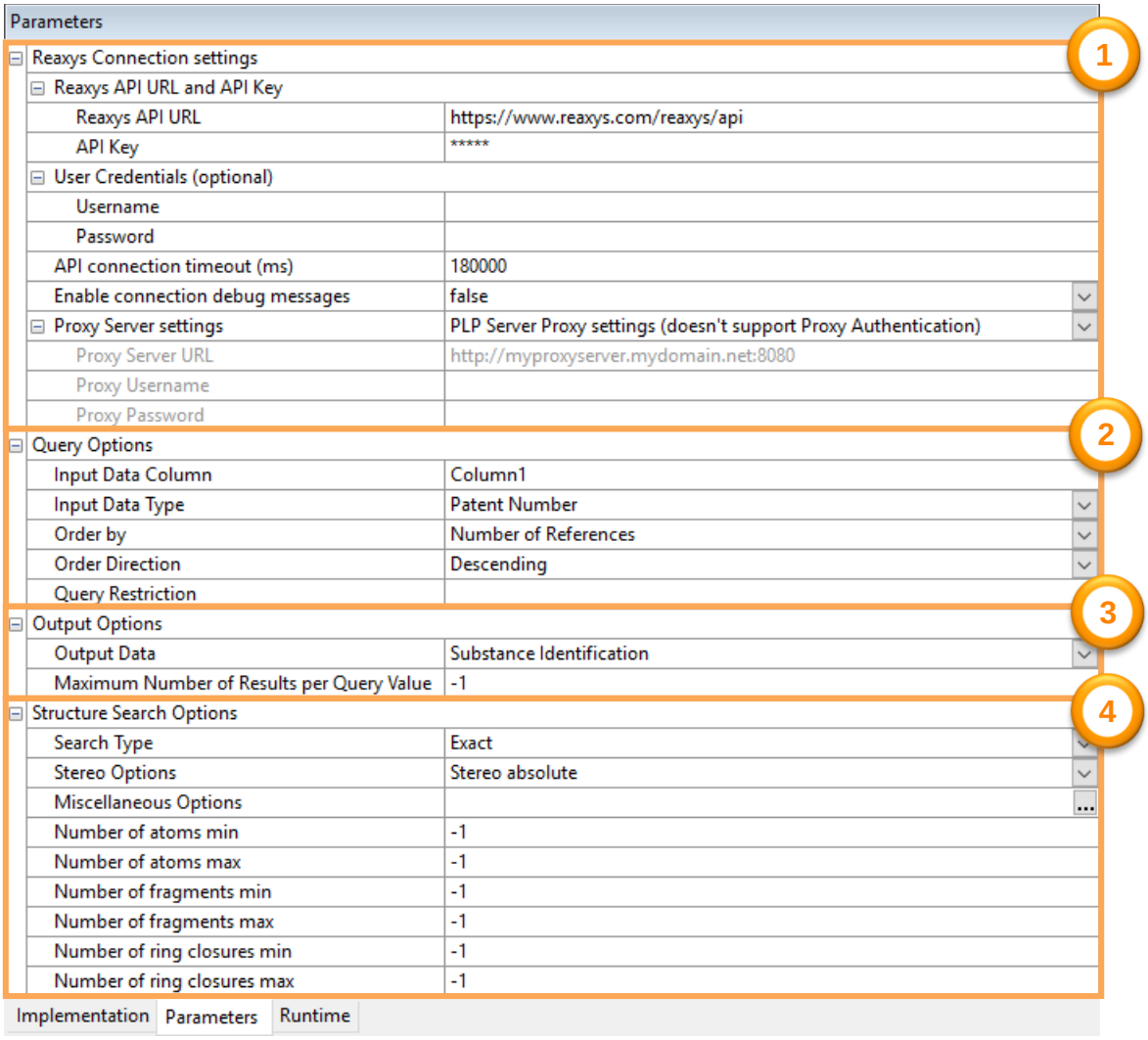
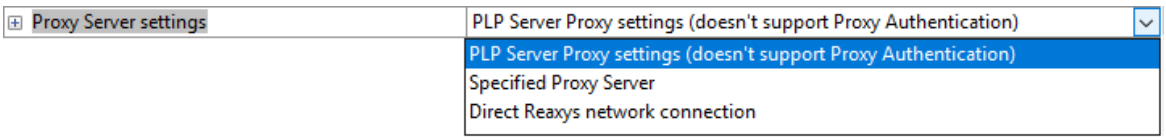
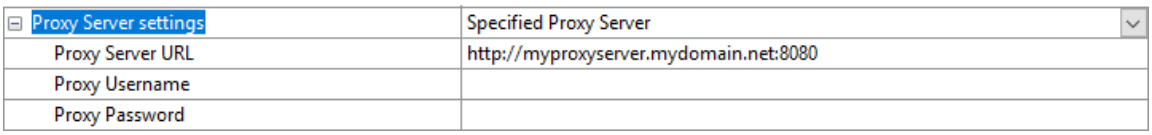


Figure 2. The "Parameters" in Pipeline Pilot. The Reaxys Substances Parameter fields are shown as an example. 1) The field for user credentials. 2) The field for specifying the input data column and the type of input data. 3) The output options field. 4) The field for structure search options.

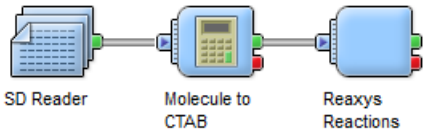
The "Reaxys Connection settings" field (Figure 2 – 1)





The "Structure Search Options" field (Figure 2 – 4)

The Reaxys Pipeline Pilot components can take substance structures (mol files) or reactions (rxn files) as query values in a CTAB format. Before using the Reaxys Pipeline Pilot component, the user needs to convert the mol files or rxn files into a CTAB format (Molecule to CTAB component):



In that case, the user needs to specify “Structure” as the input data type. The components will then perform a structure search. Exact match structure searching, substructure searching and searching by similarity are supported.

Note that the Reaxys Citations Pipeline Pilot component does not allow searching bibliographic information by structure.

The user needs to specify the structure search type (Figure 3).

- **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.

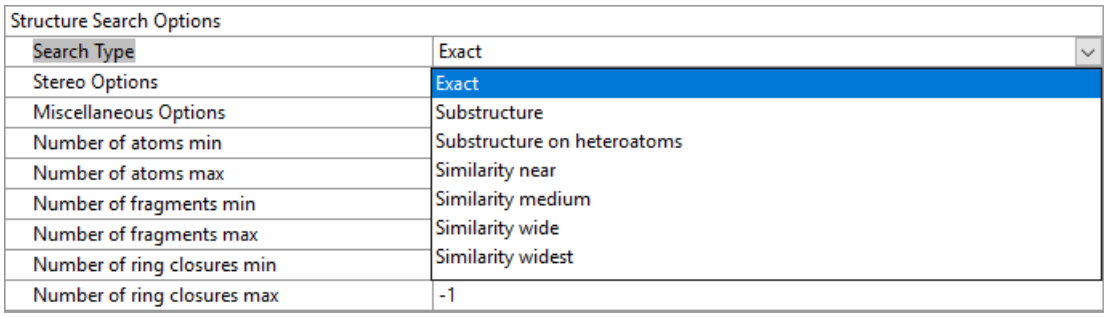


Figure 3. The field for defining the structure search type

The user also needs to specify how stereochemistry is treated (Figure 5):

- **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.

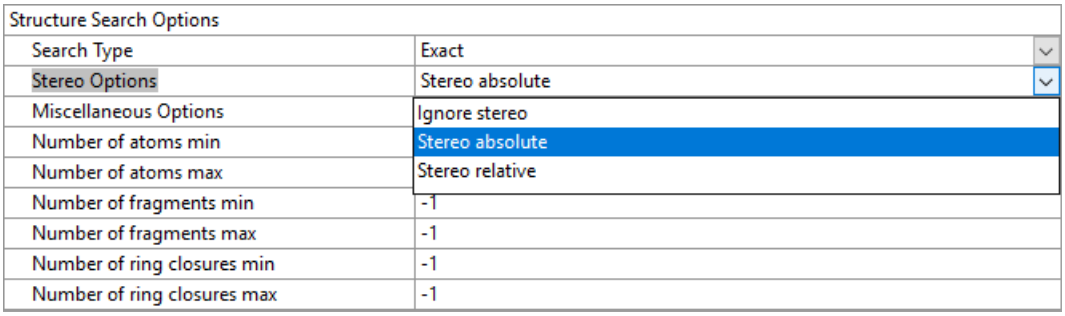


Figure 4. The field for defining the approach to stereochemistry

The user can further specify miscellaneous options (Figure 5).

- **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 components.
- **Include related Markush**: Related Markush structures are retrieved. This option is not available from the Reactions component.
- **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 component.
- **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 component.

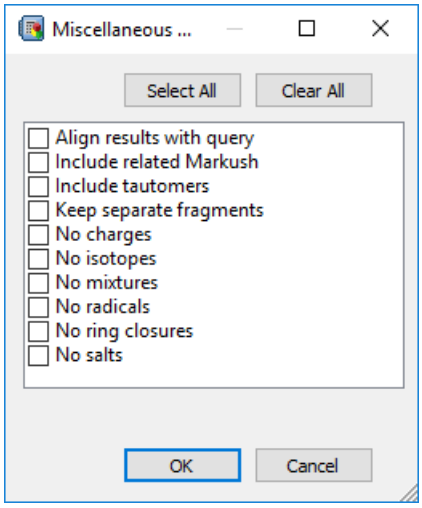


Figure 5. Miscellaneous options

The user can further specify the number of atoms, fragments and ring closures to be found in the results (Figure 6). These options are not available from the Reaction component. 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).

Structure Search Options	
Search Type	Exact
Stereo Options	Stereo absolute
Miscellaneous Options	
Number of atoms min	-1
Number of atoms max	-1
Number of fragments min	-1
Number of fragments max	-1
Number of ring closures min	-1
Number of ring closures max	-1

Figure 6. The field for defining the number of atoms, fragments and ring closures

In the “Structure Search Options” tab in the configuration dialog of the Reaxys Reactions Pipeline Pilot component, additional role options need to be specified (Figure 7). The role options define the role that the query structures play in the reaction results. Role options are only available from the Reactions component. 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.

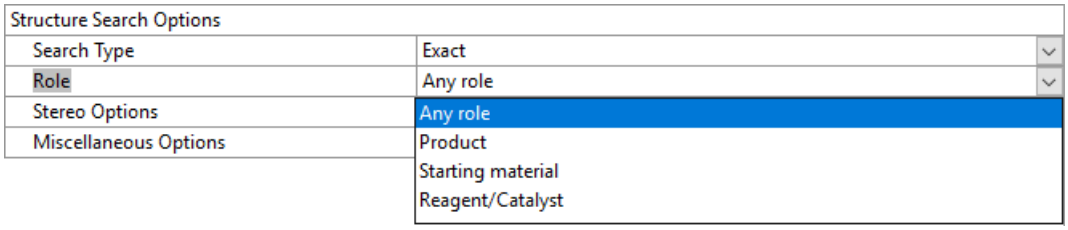


Figure 2. The field for defining the specific structure search option

Output

The output from the Reaxys Pipeline Pilot components is a data record where the requested data retrieved for all substance, reaction, bioactivity or citation instances matching the queries is concatenated. To transform this data record into a table, use the “Unmerge data” component from Pipeline Pilot library.

Each row in the data record contains 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 data record 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 Pipeline Pilot workflow.

Performance

This section lists examples of searches performed using Reaxys Pipeline Pilot components 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	8.3	4
2	Search substances by chemical name and retrieve substance property (boiling point) data	398 boiling point records	5.7	2.9
3	Search bioactive substances by patent number and retrieve the structures	174 chemical structures	7.6	1.3
4	Search preparations for a set of structures and retrieve the reaction structures	125 reaction structures	45.4	4.9
5	Search bioactivities by patent number	300 bioactivities	7.4	3.2
6	Search bioactivities and substances by target name and retrieve bioactivity datapoints and structures for the most active compounds	1000 datapoints and structures	199.5	7.5

Examples

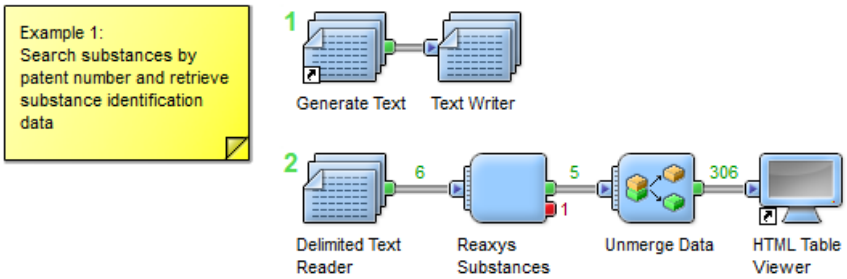
Reaxys Pipeline Pilot component examples as well as the required files can be download here:

http://supportcontent.elsevier.com/Support Hub/Reaxys/pipelinepilot/9.5/Reaxys pipelinepilot workflow Examples/Reaxys pipelinepilot workflow Examples 9_5.zip

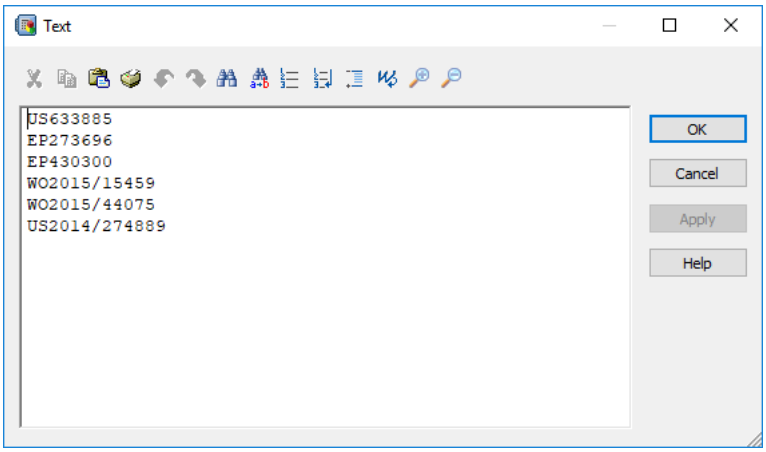
1. Search substances by patent number and retrieve substance identification data

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 “Generate Text” and “Text Writer” are used to generate a delimited text file store in the Pipeline Pilot server and further used by the “delimited text reader” component to define the query values. In this example, we used six patent numbers as query values.

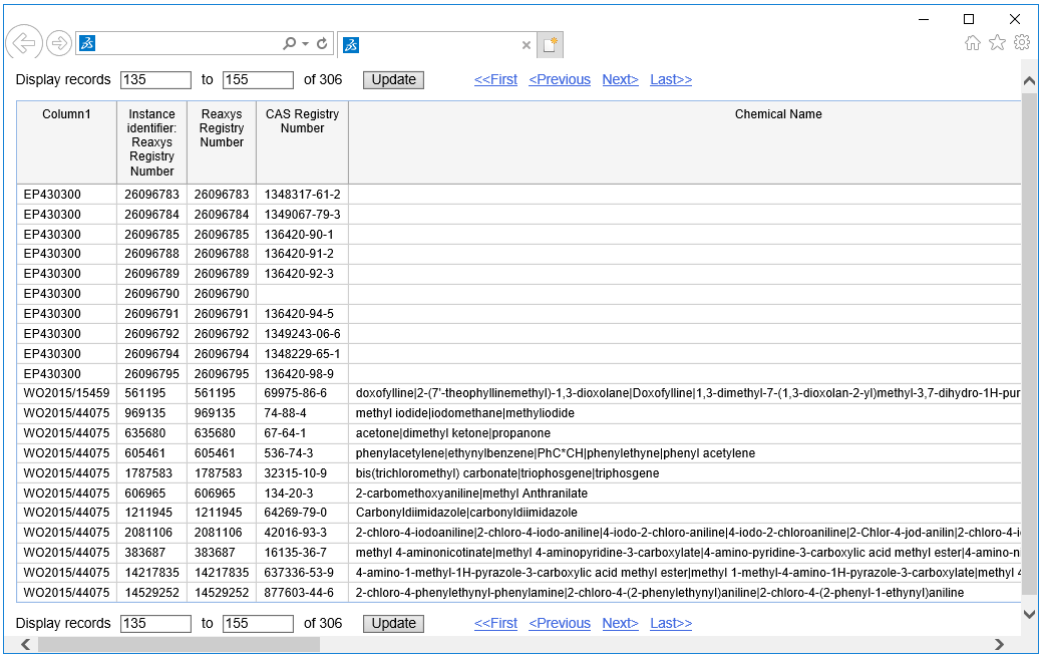


In the “Reaxys Pipeline Pilot Substances” component, we need to define the type of the input data and the output data.

Parameters	
Reaxys Connection settings	
Reaxys API URL and API Key	
Reaxys API URL	https://www.reaxys.com/reaxys/api
API Key	*****
User Credentials (optional)	
Username	
Password	
API connection timeout (ms)	180000
Enable connection debug messages	false
Proxy Server settings	PLP Server Proxy settings (doesn't support Proxy Authentication)
Query Options	
Input Data Column	Column1
Input Data Type	Patent Number
Order by	Number of References
Order Direction	Descending
Query Restriction	
Output Options	
Output Data	Substance Identification
Maximum Number of Results per Query Value	-1
Structure Search Options	
Search Type	Exact
Stereo Options	Stereo absolute
Miscellaneous Options	
Number of atoms min	-1
Number of atoms max	-1
Number of fragments min	-1
Number of fragments max	-1
Number of ring closures min	-1
Number of ring closures max	-1

Implementation Parameters Runtime

The output is a table with this format:



Display records 135 to 155 of 306 Update <<First <Previous Next> Last>>

Column1	Instance identifier: Reaxys Registry Number	Reaxys Registry Number	CAS Registry Number	Chemical Name
EP430300	26096783	26096783	1348317-61-2	
EP430300	26096784	26096784	1349067-79-3	
EP430300	26096785	26096785	136420-90-1	
EP430300	26096788	26096788	136420-91-2	
EP430300	26096789	26096789	136420-92-3	
EP430300	26096790	26096790		
EP430300	26096791	26096791	136420-94-5	
EP430300	26096792	26096792	1349243-06-6	
EP430300	26096794	26096794	1348229-65-1	
EP430300	26096795	26096795	136420-98-9	
WO2015/15459	561195	561195	69975-86-6	doxofylline|2-(7'-theophyllinemethyl)-1,3-dioxolane|Doxofylline|1,3-dimethyl-7-(1,3-dioxolan-2-yl)methyl-3,7-dihydro-1H-pur
WO2015/44075	969135	969135	74-88-4	methyl iodide|iodomethane|methyliodide
WO2015/44075	635680	635680	67-64-1	acetone|dimethyl ketone|propanone
WO2015/44075	605461	605461	536-74-3	phenylacetylene|ethynylbenzene|PhC*CH|phenylethyne|phenyl acetylene
WO2015/44075	1787583	1787583	32315-10-9	bis(trichloromethyl) carbonate|triophosgene|triphosgene
WO2015/44075	606965	606965	134-20-3	2-carbomethoxyaniline|methyl Anthranilate
WO2015/44075	1211945	1211945	64269-79-0	Carbonyldiimidazole|carbonyldiimidazole
WO2015/44075	2081106	2081106	42016-93-3	2-chloro-4-iodoaniline|2-chloro-4-iodo-aniline|4-iodo-2-chloro-aniline|4-iodo-2-chloroaniline|2-Chlor-4-jod-anilin|2-chloro-4-i
WO2015/44075	383687	383687	16135-36-7	methyl 4-aminonicotinate|methyl 4-aminopyridine-3-carboxylate|4-amino-pyridine-3-carboxylic acid methyl ester|4-amino-n
WO2015/44075	14217835	14217835	637336-53-9	4-amino-1-methyl-1H-pyrazole-3-carboxylic acid methyl ester|methyl 1-methyl-4-amino-1H-pyrazole-3-carboxylate|methyl 4
WO2015/44075	14529252	14529252	877603-44-6	2-chloro-4-phenylethynyl-phenylamine|2-chloro-4-(2-phenylethynyl)aniline|2-chloro-4-(2-phenyl-1-ethynyl)aniline

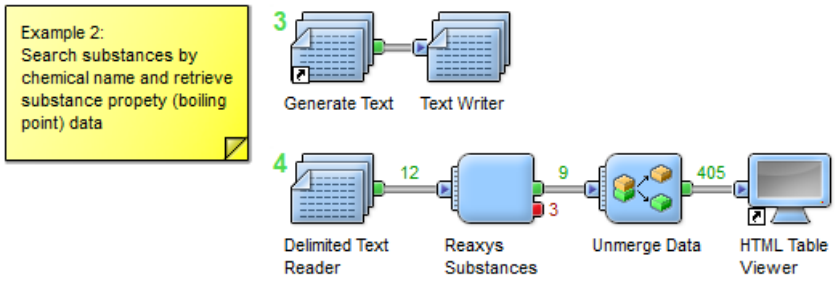
Display records 135 to 155 of 306 Update <<First <Previous Next> Last>>

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

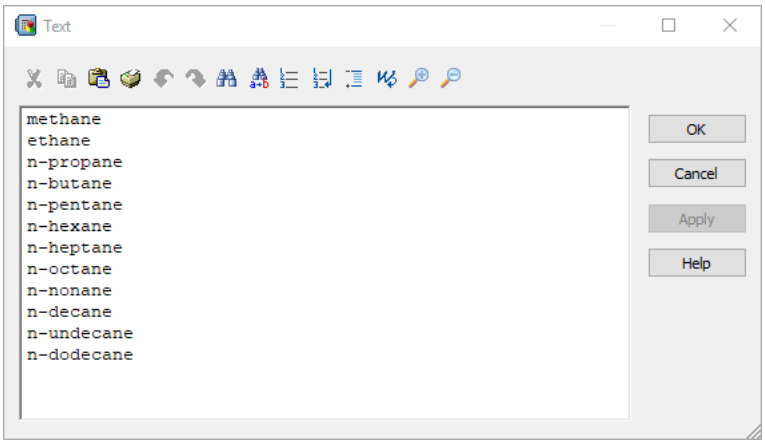
2. Search substances by chemical name and retrieve substance property data

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:

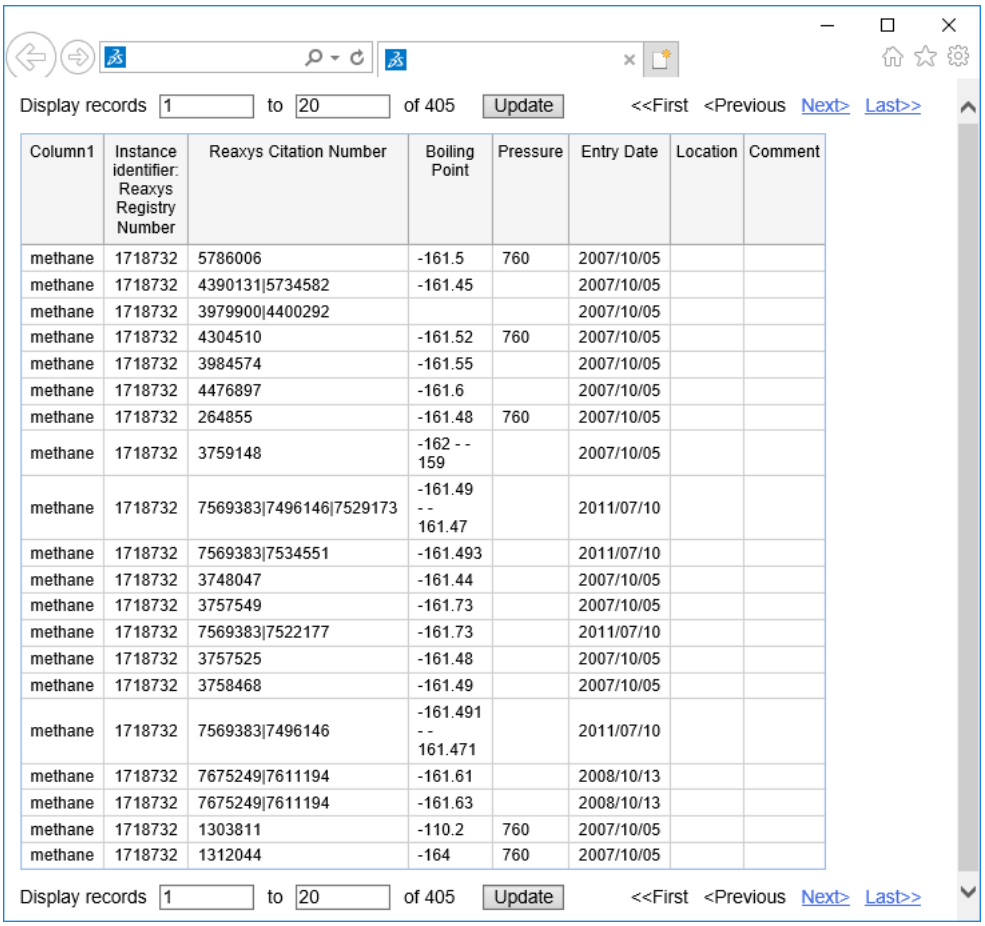


The “Generate Text ” and “Text Writer” are used to generate a delimited text file store in the Pipeline Pilot server and further used by the “delimited text reader” component to define the query values. In this example, we use 12 Substance Chemical Names as query values.



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

The output is a table with this format:



Display records 1 to 20 of 405 Update <<First <Previous Next> Last>>

Column1	Instance identifier: Reaxys Registry Number	Reaxys Citation Number	Boiling Point	Pressure	Entry Date	Location	Comment
methane	1718732	5786006	-161.5	760	2007/10/05		
methane	1718732	4390131|5734582	-161.45		2007/10/05		
methane	1718732	3979900|4400292			2007/10/05		
methane	1718732	4304510	-161.52	760	2007/10/05		
methane	1718732	3984574	-161.55		2007/10/05		
methane	1718732	4476897	-161.6		2007/10/05		
methane	1718732	264855	-161.48	760	2007/10/05		
methane	1718732	3759148	-162 - - 159		2007/10/05		
methane	1718732	7569383|7496146|7529173	-161.49 - - 161.47		2011/07/10		
methane	1718732	7569383|7534551	-161.493		2011/07/10		
methane	1718732	3748047	-161.44		2007/10/05		
methane	1718732	3757549	-161.73		2007/10/05		
methane	1718732	7569383|7522177	-161.73		2011/07/10		
methane	1718732	3757525	-161.48		2007/10/05		
methane	1718732	3758468	-161.49		2007/10/05		
methane	1718732	7569383|7496146	-161.491 - - 161.471		2011/07/10		
methane	1718732	7675249|7611194	-161.61		2008/10/13		
methane	1718732	7675249|7611194	-161.63		2008/10/13		
methane	1718732	1303811	-110.2	760	2007/10/05		
methane	1718732	1312044	-164	760	2007/10/05		

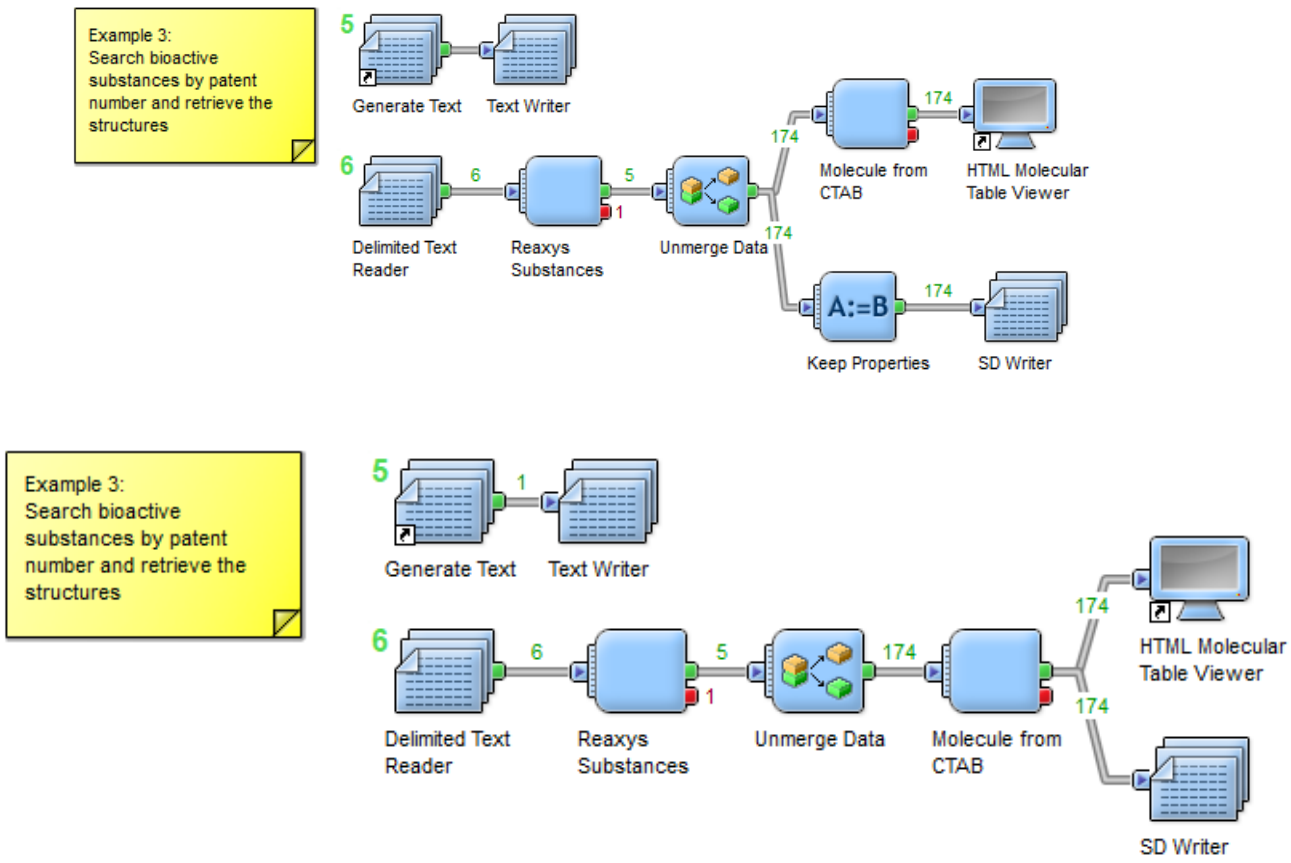
Display records 1 to 20 of 405 Update <<First <Previous Next> Last>>

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

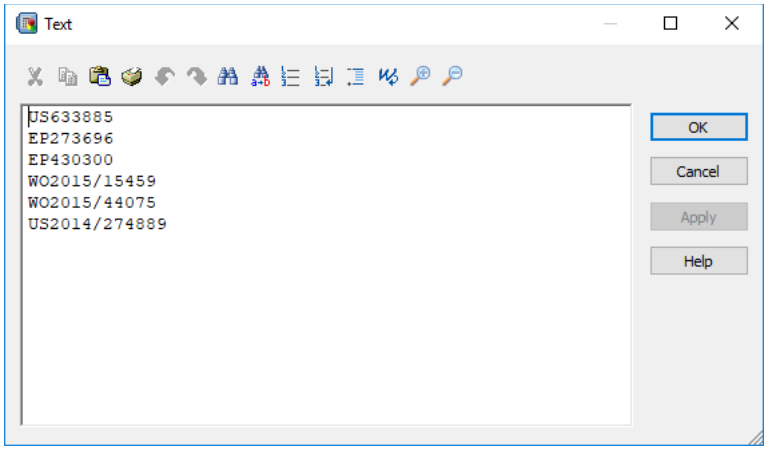
3. Search bioactive substances by patent number and retrieve the structures

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.

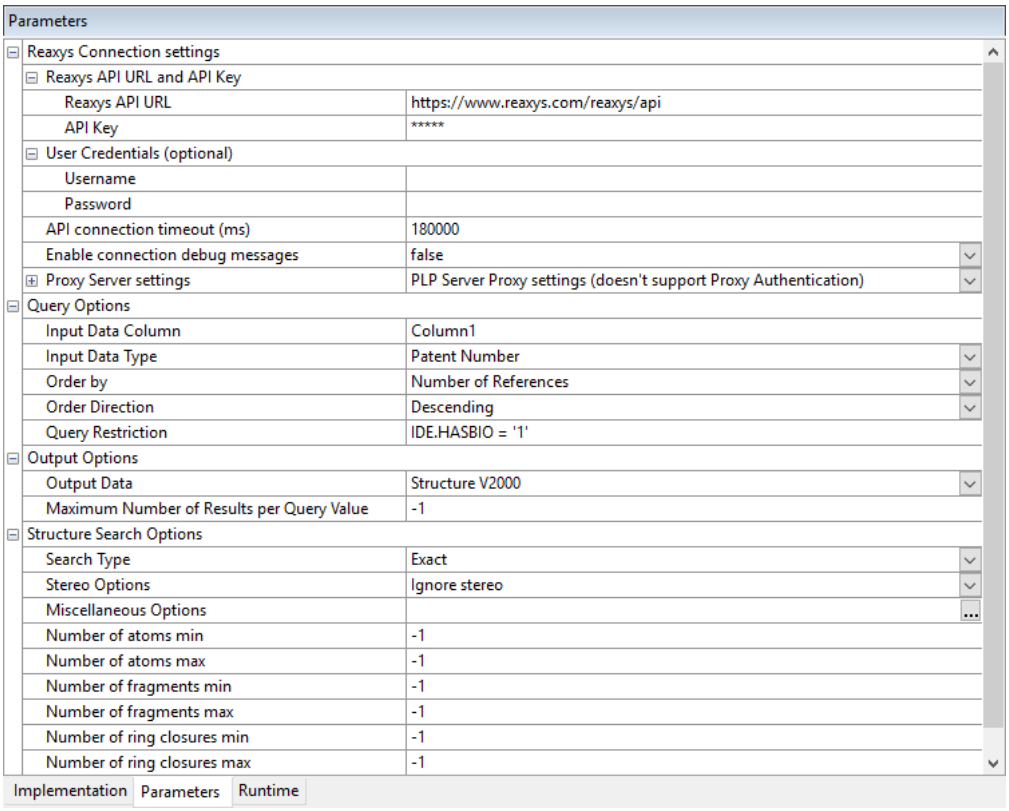
The overall workflow appears as follows:



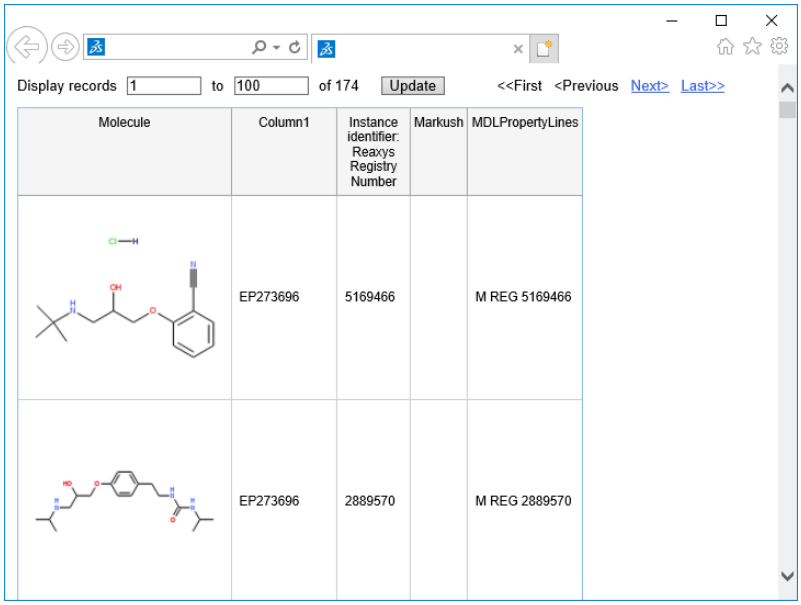
The “Generate Text” and “Text Writer” are used to generate a delimited text file store in the Pipeline Pilot server and further used by the “delimited text reader” component to define the query values. In this example, we used six patent numbers as query values.


Text
US633885
EP273696
EP430300
WO2015/15459
WO2015/44075
US2014/274889
OK
Cancel
Apply
Help

In “Reaxys Pipeline Pilot Substances” component, 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.



The output is a table with this format:



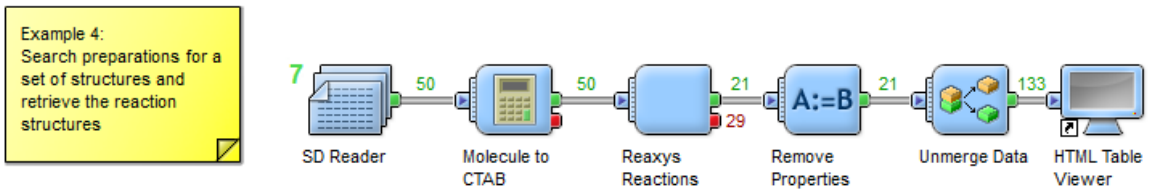
In addition to displaying the structures on screen, we want to export them into an SDF file. We use the “SDF Writer” component for that.

Result: 174 structures are retrieved and exported.

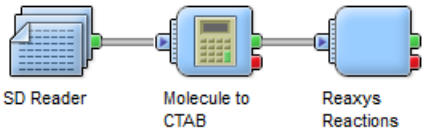
4. Search preparations for a set of structures and retrieve the reaction structures

In this example, we use the top 50 of 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:



Before using the Reaxys Pipeline Pilot component, the user needs to convert the mol files, the rxn files or Sd files into a CTAB format (Molecule to CTAB component).



These are the settings:

Parameters	
Reaxys Connection settings	
Reaxys API URL and API Key	
Reaxys API URL	https://www.reaxys.com/reaxys/api
API Key	*****
User Credentials (optional)	
Username	
Password	
API connection timeout (ms)	180000
Enable connection debug messages	false
Proxy Server settings	PLP Server Proxy settings (doesn't support Proxy Authentication)
Query Options	
Input Data Column	CTAB
Input Data Type	Structure
Order by	Reaxys Ranking
Order Direction	Descending
Query Restriction	
Output Options	
Output Data	Reaction Structure V2000
Maximum Number of Results per Query Value	-1
Structure Search Options	
Search Type	Exact
Role	Product
Stereo Options	Stereo absolute
Miscellaneous Options	Align results with query

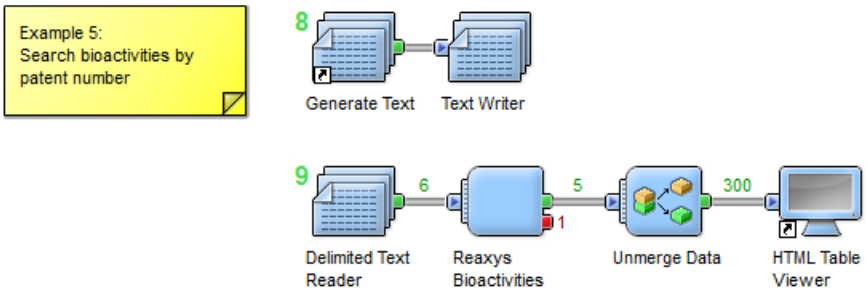
Implementation | Parameters | Runtime

Result: 155 reaction structures are retrieved.

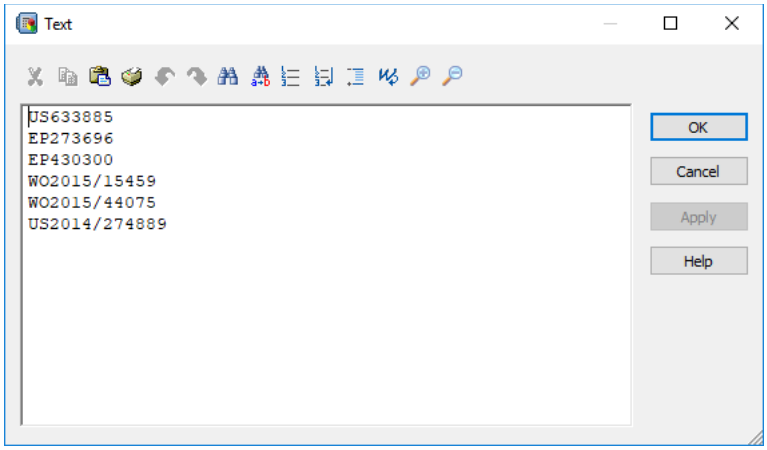
5. Search for bioactivities by patent number

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

The overall workflow appears as follows:



The “Generate Text ” and “Text Writer” are used to generate a delimited text file store in the Pipeline Pilot server and further used by the “delimited text reader” component to define the query values. In this example, we used six patent numbers as query values.

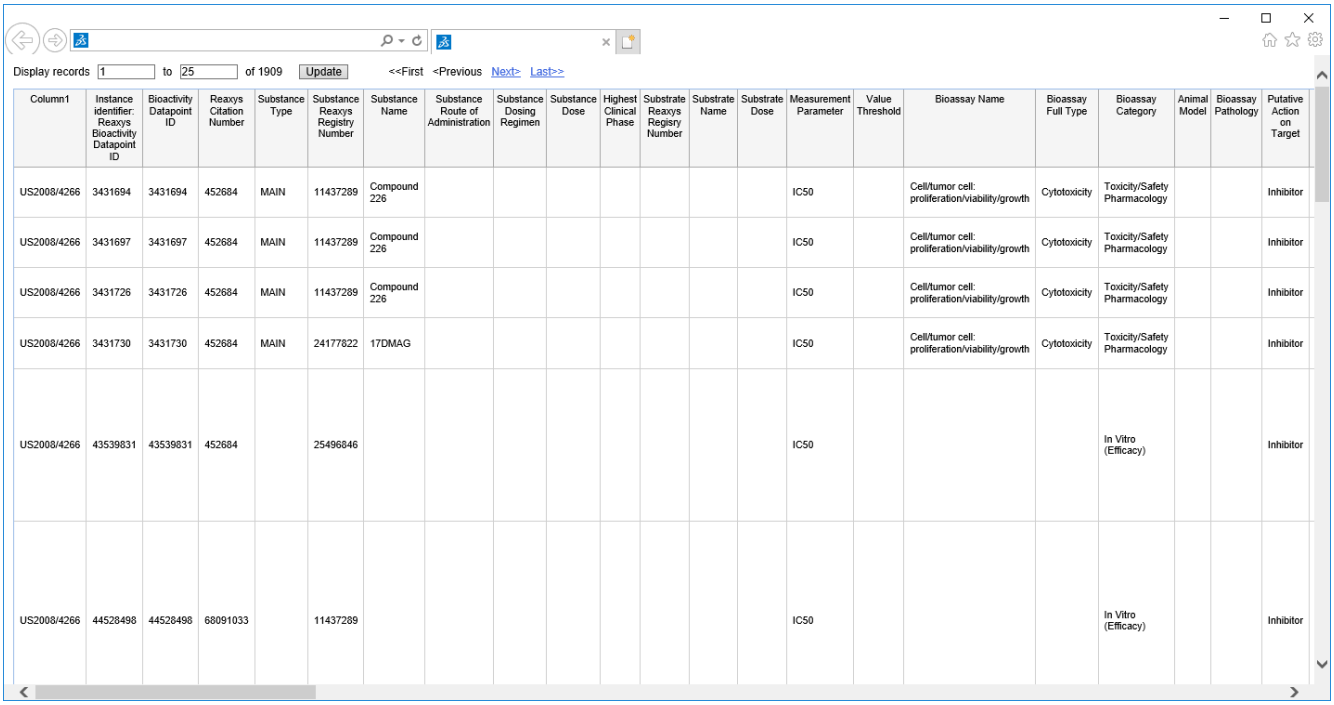


These are the Bioactivities parameters.

Parameters	
Reaxys Connection settings	
Reaxys API URL and API Key	
Reaxys API URL	https://www.reaxys.com/reaxys/api
API Key	*****
User Credentials (optional)	
Username	
Password	
API connection timeout (ms)	180000
Enable connection debug messages	false
Proxy Server settings	PLP Server Proxy settings (doesn't support Proxy Authentication)
Query Options	
Input Data Column	Column1
Input Data Type	Patent Number
Order by	Measurement pX
Order Direction	Descending
Query Restriction	
Output Options	
Output Data	Bioactivity
Maximum Number of Results per Query Value	-1
Structure Search Options	
Search Type	Exact
Stereo Options	Stereo absolute
Miscellaneous Options	
Number of atoms min	-1
Number of atoms max	-1
Number of fragments min	-1
Number of fragments max	-1
Number of ring closures min	-1
Number of ring closures max	-1

Implementation | Parameters | Runtime

The output is a table with this format:



Display records 1 to 25 of 1909 Update <<First <Previous Next> Last>>

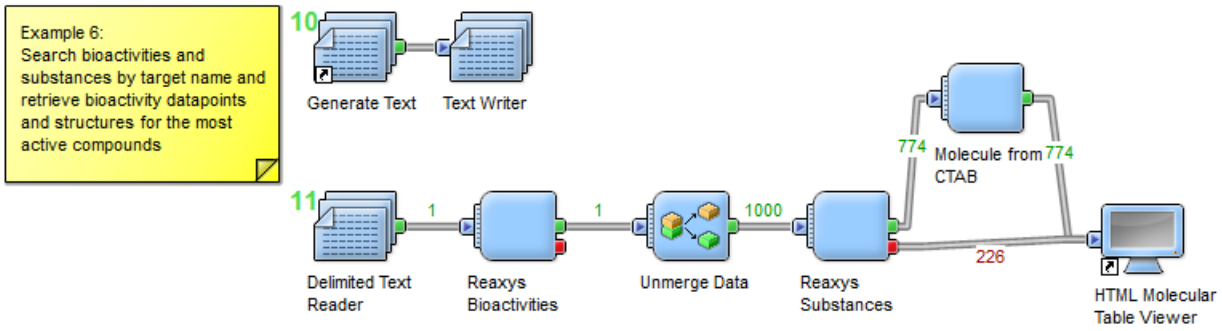
Column1	Instance identifier: Reaxys Bioactivity Datapoint ID	Bioactivity Datapoint ID	Reaxys Citation Number	Substance Type	Substance Reaxys Registry Number	Substance Name	Substance Route of Administration	Substance Dosing Regimen	Substance Dose	Highest Clinical Phase	Substrate Reaxys Regisry Number	Substrate Name	Substrate Dose	Measurement Parameter	Value Threshold	Bioassay Name	Bioassay Full Type	Bioassay Category	Animal Model	Bioassay Pathology	Putative Action on Target
US2008/4266	3431694	3431694	452684	MAIN	11437289	Compound 226								IC50		Cell/tumor cell: proliferation/viability/growth	Cytotoxicity	Toxicity/Safety Pharmacology			Inhibitor
US2008/4266	3431697	3431697	452684	MAIN	11437289	Compound 226								IC50		Cell/tumor cell: proliferation/viability/growth	Cytotoxicity	Toxicity/Safety Pharmacology			Inhibitor
US2008/4266	3431726	3431726	452684	MAIN	11437289	Compound 226								IC50		Cell/tumor cell: proliferation/viability/growth	Cytotoxicity	Toxicity/Safety Pharmacology			Inhibitor
US2008/4266	3431730	3431730	452684	MAIN	24177822	17DMAG								IC50		Cell/tumor cell: proliferation/viability/growth	Cytotoxicity	Toxicity/Safety Pharmacology			Inhibitor
US2008/4266	43539831	43539831	452684		25496846									IC50				In Vitro (Efficacy)			Inhibitor
US2008/4266	44528498	44528498	68091033		11437289									IC50				In Vitro (Efficacy)			Inhibitor

Result: 300 bioactivity datapoints are retrieved.

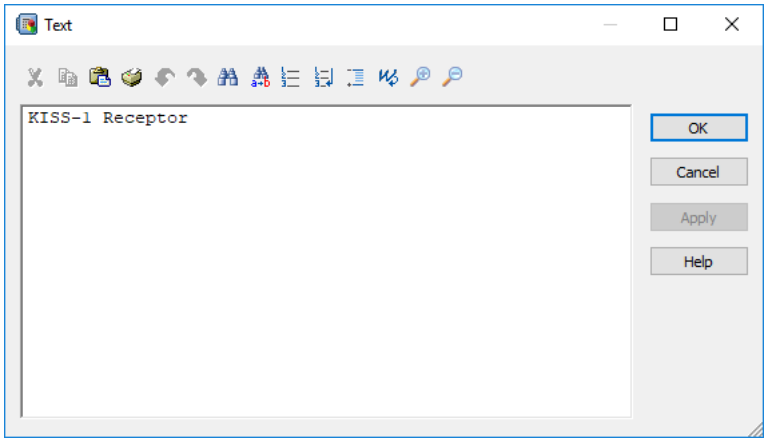
6. Search bioactivities and substances by target name and retrieve bioactivity datapoints and structures for the most active compounds

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:



The “Generate Text ” and “Text Writer” are used to generate a delimited text file store in the Pipeline Pilot server and further used by the “delimited text reader” component to define the query values. In this example, we used 1 target name: KiSS-1 receptor.



We use the Bioactivities and the Substances parameters.

Parameters	
Reaxys Connection settings	
Reaxys API URL and API Key	
Reaxys API URL	https://www.reaxys.com/reaxys/api
API Key	*****
User Credentials (optional)	
Username	
Password	
API connection timeout (ms)	180000
Enable connection debug messages	false
Proxy Server settings	PLP Server Proxy settings (doesn't support Proxy Authentication)
Query Options	
Input Data Column	Column1
Input Data Type	Target Name/Uniprot ID/PDB ID
Order by	Measurement pX
Order Direction	Descending
Query Restriction	
Output Options	
Output Data	Bioactivity
Maximum Number of Results per Query Value	1000
Structure Search Options	
Search Type	Exact
Stereo Options	Stereo absolute
Miscellaneous Options	
Number of atoms min	-1
Number of atoms max	-1
Number of fragments min	-1
Number of fragments max	-1
Number of ring closures min	-1
Number of ring closures max	-1

Implementation Parameters Runtime

Parameters	
Reaxys Connection settings	
Reaxys API URL and API Key	
Reaxys API URL	https://www.reaxys.com/reaxys/api
API Key	*****
User Credentials (optional)	
Username	
Password	
API connection timeout (ms)	180000
Enable connection debug messages	false
Proxy Server settings	PLP Server Proxy settings (doesn't support Proxy Authentication)
Query Options	
Input Data Column	Substance Reaxys Registry Number
Input Data Type	Reaxys Registry Number
Order by	Number of References
Order Direction	Descending
Query Restriction	
Output Options	
Output Data	Structure V2000
Maximum Number of Results per Query Value	-1
Structure Search Options	
Search Type	Exact
Stereo Options	Ignore stereo
Miscellaneous Options	Align results with query
Number of atoms min	-1
Number of atoms max	-1
Number of fragments min	-1
Number of fragments max	-1
Number of ring closures min	-1
Number of ring closures max	-1

Implementation Parameters Runtime

Result: 1000 bioactivity datapoints and 721 structures are retrieved. 279 compounds have no available structure.