

# **KNIME Enalos+ nodes**



#### A Brief Tutorial

Novamechanics Ltd Contact: info@novamechanics.com Version 1, June 2017

#### **Table of Contents**

| Introdu | 1 netion                          |
|---------|-----------------------------------|
| Step 1- | Workbench overview                |
| Step 2- | Building a workflow2              |
| 1.      | Node status2                      |
| 2.      | Ports                             |
| Step 3- | Activate the Enalos+ nodes        |
| Step 4- | A Brief Introduction              |
| 1.      | Modelling                         |
| 2.      | Molecular Descriptors4            |
| 3.      | NCI4                              |
| 4.      | PubChem4                          |
| 5.      | UniChem                           |
| Step 5- | Adding Nodes                      |
| Step 6- | Connecting Nodes                  |
| Step 7- | Configuring nodes                 |
| Step 8- | Executing nodes                   |
| Step 9- | Inspecting the results            |
| Step 10 | 0-Extending the main Workflow11   |
| Embarl  | k your own voyage of discovery!16 |



## Introduction

Real world research seldom involves a single question being answered by means of a single operation, and the fields of molecular modelling, chemoinformatics and nanoinformatics are no exception. While researchers can create custom scripts to automate common procedures, this solution is less than ideal when projects demand rapid workflow prototyping, interactive data analysis, and robust, appropriately validated models. These are exactly the conditions for which Novamechanics Ltd Enalos+ nodes are best suited to open-source KNIME interface. Enalos+ nodes are designed to perform molecular modelling and help the user get straight access to multiple Chemical Databases for data mining and manipulation.

Enalos+ nodes built upon the existing KNIME infrastructure are divided in five main categories (Modelling, Molecular Descriptors, NCI, PubChem and UniChem) and significantly increase the number of the available nodes, the data handling tools and bridge different chemoinformatics and modelling tools upon the same interface.

The current tutorial is designed to help the user in going step-by-step through the process of building a KNIME workflow, using Enalos+ nodes of Novamechanics Ltd. This case study deals with the chemical compound abacavir, which is a medication used to prevent and treat HIV/AIDS.

## **Step 1-Workbench overview**

The KNIME workbench is organized as follows:

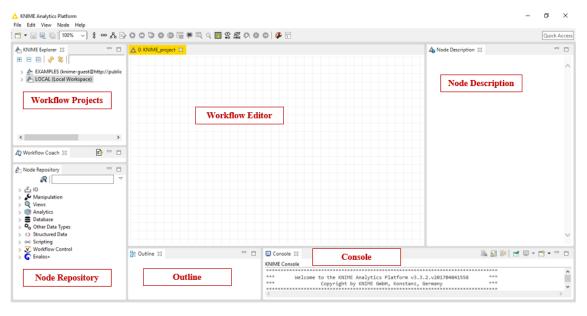


Fig. 1: KNIME workbench

It is composed of 6 main "windows": The Workflow Projects, the Workflow Editor, the Node Description, the Node Repository, the Outline and the Console. A short description of the KNIME's interface windows follows in Table 1:



| Workflow Projects   | Workflow Editor  | Node Description   |
|---|--|--|
| Each workflow refers to a<br>workflow project. All<br>projects are displayed here.<br>Import and export of<br>workflows is supported.<br>Status (closed, idle,<br>executing and executed) is<br>indicated by an icon. | Here the workflows are<br>assembled by dragging nodes<br>onto this editor, connecting,<br>configuring and executing<br>them. | Provides help about the<br>selected node, its dialog<br>options, views, expected<br>input data and resulting<br>output.      |
| Node Repository   | Outline  | Console  |
| Find all KNIME nodes here,<br>ordered by categories. Help<br>for selected nodes is<br>displayed in the Node<br>Description. Drag them onto<br>the editor in order to add<br>them to the workflow.                     | Overview over the workflow<br>and navigation help for large<br>workflows.  | Status information, warnings<br>and error messages are<br>logged here. This<br>information is also written to<br>a log file. |

#### Table 1: Description of KNIME interface

#### Step 2-Building a workflow

The nodes are the basic processing units of a KNIME workflow. A workflow is built by dragging nodes from the Node Repository onto the Workflow Editor and connecting them, creating pipelines: Each node has a number of input-and/or output ports. Data (or a model according to each particular case) is transferred over a connection from an out-port to the in-port of another node.

#### 1. Node status

When a node is dragged onto the workflow editor the status light shows red, which means that the node has to be configured in order to be able to be executed. A node is configured by right clicking it, choosing "Configure", and adjusting the necessary settings in the node's dialog. When the dialog is closed by pressing the "OK" button, the node is configured and the status light changes to yellow: the node is ready to be executed. Right-click on the node again shows an enabled "Execute" option; pressing it will execute the node and the result of this node will be available at the out-port (Fig. 2). After a successful execution the status light of the node is green. The result(s) can be inspected by exploring the out-port view(s): the last entries in the context menu open them. The above options "Configure", "Execute" and "View" are also available in the top ribbon of the KNIME interface window.

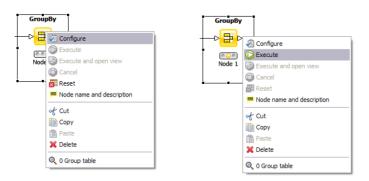


Fig. 2: Configuring and executing nodes



#### 2. Ports

Ports on the left are input ports, where the data from the out-port of the predecessor node are provided. Ports on the right side of the node are called out-ports. The result of the node's operation on the data is provided at the out-port to successor nodes.

### **Step 3-Activate the Enalos+ nodes**

In order to activate the Enalos+ nodes, the user has to copy the .jar file in the plugins folder and the .lic file in the license folder in the KNIME file location.

## **Step 4-A Brief Introduction**

The Enalos+ nodes are divided into 5 main categories: Modelling, Molecular Descriptors, NCI, PubChem and UniChem.

#### 1. Modelling

Modelling contains 11 nodes specified for data handling, preprocessing, testing modeling robustness and testing the accuracy of the predictions:

| Create New Molecules   | Domain APD   | Domain Leverage   |
|--|--|---|
| <i>Create New Molecules</i><br>enables the user to create a list<br>of molecules by combining a<br>series of substituents with a<br>core molecule.   | <i>Domain APD</i> enables the user<br>to define the domain of<br>applicability of the model<br>using a method based on the<br>Euclidean distances.                                   | <i>Domain Leverage</i> enables the<br>user to define the domain of<br>applicability of the model<br>using a method based on the<br>extent of extrapolation.   |
| Int 2 Double   | Kennard and Stone  | MLR   |
| <i>Int 2 Double</i> converts integer values of all columns to doubles.   | <i>Kennard-Stone</i> node allows<br>the selection of two<br>representative subsets (as<br>training and test sets) with a<br>uniform distribution over an<br>initial dataset.         | <i>MLR</i> node performs Multiple<br>Linear Regression in order to<br>model the relationships<br>between a scalar dependent<br>variable y and two or more<br>independent variables<br>denoted as X.   |
| Model Acceptability  | Remove Column  | Remove Duplicates   |
| Criteria   |  |   |
| Model Acceptability Criteria<br>gives information about the<br>Quality of Fit and Predictive<br>Ability of a continuous<br>QSAR Model.   | <i>Remove Column</i> node<br>removes the selected input<br>columns of the table that<br>contain the same values at a<br>percentage equal or higher<br>than a specified cutoff limit. | <i>Remove Duplicates</i> enables<br>the user to remove the rows of<br>the input table that contain the<br>same values in selected<br>columns. The filtered table<br>contains all rows that are<br>unique and the first one of<br>each repeated row. |
| Sphere Exclusion   | Y Randomization  |   |
| Sphere Exclusion node allows<br>the selection of two<br>representative subsets (such<br>as training and test sets). This<br>method attempts to specify<br>compounds which most<br>effectively cover the<br>available data space. | <i>Y Randomization</i> (or Y-scrambling) is a technique, applied to ensure a QSAR model's robustness.  |   |



## 2. Molecular Descriptors

Molecular Descriptors contains *EnalosMold2* node.

#### 3. NCI

NCI contains CIR node.

| Table 3: | Molecular | Descriptors | and NCI |
|----------|-----------|-------------|---------|
| rubic 5. | moncomm   | Descriptors | and not |

| EnalosMold2                                     | CIR   |
|---|---|
| Molecular Descriptors by EnalosMold2            | Enalos+ CIR node enables the user to get        |
| calculates a large and diverse set of molecular | direct access to CIR (Chemical Identifier       |
| descriptors (777) encoding two-dimensional      | Resolver) through KNIME. The user has the       |
| chemical structure information.                 | option to select several output formats through |
|   | a GUI menu.                                     |

#### 4. PubChem

PubChem contains 8 nodes that give direct access to PubChem database through KNIME in order to extract useful information:

| Assay Class                                     |
|---|
| Assay Class node searches in PubChem            |
| database according to one or more given AIDs    |
| (BioAssay identification numbers) and           |
| displays only the active or inactive            |
| compounds.                                      |
|   |
| Patent  |
| Patent node gives the user straight access to   |
| the PubChem database in order to obtain         |
| information about the patent coverage           |
| information for thousands of compounds with     |
| one request.                                    |
| •   |
|   |
|   |
| Sid   |
| Sid node exports the CIDs (Compound IDs) of     |
| a given list of SIDs (Substance IDs), searching |
| the PubChem database. The user can search       |
| the PubChem database and obtain information     |
| about the CIDs for thousands of compounds       |
| with one request.                               |
| Vendor  |
| Vendor node enables the user to search the      |
| PubChem database and obtain information         |
| about the commercial availability for           |
| thousands of compounds with one request.        |
|   |
|   |
|   |
|   |
|   |

Table 4: PubChem nodes



#### 5. UniChem

UniChem contains 2 nodes for accessing UniChem databases:

| Table  | 5:         | UniChem  | nodes |
|--------|------------|----------|-------|
| 1 0000 | <i>v</i> . | onconcin | nouco |

| UniChem   | UniChem Connectivity  |
|---|---|
| Enalos <i>UniChem</i> gives the user direct access<br>to UniChem databases through KNIME.<br>UniChem is a superset of all 27 available<br>databases, separated in 5 friendly and easily<br>recognizable categories. | UniChem Connectivity is an expanded version<br>of the standard UniChem tool that allows you<br>to find related molecules. Connectivity Search<br>allows molecules to be first matched on the<br>basis of complete identity between the<br>connectivity layer of their corresponding<br>Standard InChIs, and the remaining layers then<br>compared to highlight stereo-chemical and<br>isotopic differences. |

#### **Step 5-Adding Nodes**

In the Node Repository, expand the *IO* and the contained *Other* category and choose *Table Creator* node (Fig. 3). Then, drag & drop the *Table Creator* icon into the Workflow Editor window.



Fig. 3: Node Depository interface

Subsequently, expand the *Enalos*+ category followed by the *PubChem* category and drag into the Workflow Editor *Main PubChem* and *Similarity*. The final nodes are *Assay*, *Patent* and *Vendor* and are positioned to the right of the *Similarity* node as shown below (Fig. 4).



| A NOME Analytics Platform       -< |             |             |                |              |           |                    |           |              |
|--|-------------|-------------|----------------|--------------|-----------|--------------------|-----------|--------------|
| 👆 KNIME Explorer 🛛   |             | 🔥 *4: abaca | wir_tutorial 🔀 |              |           |                    | A Node    | Descri 🛛 🗖 🛙 |
| 🗄 🖻 🖻 🗇 🍣 🗍  | ~           | [           |                |              |           |                    |           |              |
| <  | >           |             |                |              |           |                    |           |              |
|  | _           |             |                |              |           | • <mark>6</mark> • |           |              |
| Workflow Coach 23  | B           | I           |                |              |           |                    |           |              |
| Recommended Nodes  | Community 1 | <b>`</b>    |                |              |           | Node 4             |           |              |
|  |             |             |                |              |           |                    |           |              |
| CSV Reader   |             |             |                | Main PubChem |           | Patent             |           |              |
|  |             |             | <b>.</b>       | × 6 ×        | × C +     | + <mark>6</mark> + |           |              |
|  |             | ,           |                |              |           |                    |           |              |
|  |             |             |                |              |           |                    |           |              |
| Node Repository  |             | 1           |                |              |           |                    |           |              |
| <u>a</u>   | 7           | 7           |                |              |           | Vendor             |           |              |
|  | -           | ×           |                |              |           |                    |           |              |
| > 👿 Workflow Control   |             |             |                |              |           |                    |           |              |
|  |             |             |                |              |           |                    |           |              |
|  |             |             |                |              |           |                    |           |              |
|  |             |             |                |              |           |                    |           |              |
|  |             |             |                |              |           |                    |           |              |
|  |             |             |                |              |           |                    |           |              |
| G Main PubChem   |             |             |                |              |           |                    |           |              |
|  |             |             |                |              |           |                    |           |              |
|  |             |             |                |              |           |                    |           |              |
| Similarity   |             | Se Outline  | 22             |              | Console X |                    | 🖹 🖬 🕬 🖬 🚽 | 🗉 🗸 📸 🖛 🖻 f  |
|  |             |             |                |              |           |                    |           | - La -       |

Fig. 4: Workflow editor

## **Step 6-Connecting Nodes**

Now, you need to connect the nodes, in order to get the data flowing. Click an output port and drag the connection to an appropriate input port. Complete the flow as pictured below (Fig. 5). The nodes will not show a green status as long as they are not yet configured and executed.

| 🕈 🕶 🔛 🖳 🐚 🚺 100% 🔍                      | \$ ው ዱ 🗈   | 0000         |                     |                    | ۶            |                     |         | Quick Acc    |
|---|------------|--------------|---------------------|--------------------|--------------|---------------------|---------|--------------|
| KNIME Explorer 🛛                        | -          | 🗆 🔥 *4: a    | abacavir_tutorial 🔀 |                    |              |                     | A Node  | Descri 🛛 🙂 E |
| E 🗉 🖗 🍫 🗐                               |            | ~ [          |                     |                    |              |                     |         |              |
| <                                       | >          |              |                     |                    |              | Assay               |         | ,            |
| Workflow Coach 🛛                        | B -        | -            |                     |                    |              | /* <mark>©</mark> * |         |              |
|   |            |              |                     |                    |              |                     |         |              |
| ecommended Nodes                        | Community  | <u> </u>     |                     |                    | /            | Node 4              |         |              |
| File Reader<br>CSV Reader               | 34%<br>19% |              |                     |                    | a            |                     |         |              |
| Table Creator                           | 13%        |              | Table Creator       | Main PubChem       | Similarity   | Patent              |         |              |
| →Database Reader                        | 7%         |              |                     | • <mark>6</mark> • |              |                     |         |              |
| Excel Reader (XLS)                      | 7%         | ~            |                     |                    |              |                     |         |              |
|   | 2          | •            | Node 1              | Node 2             | Node 3       | Node 6              |         |              |
| Node Repository                         | -          |              |                     |                    |              |                     |         |              |
| R                                       |            |              |                     |                    |              | Vendor              |         |              |
| (++> Scripting                          |            | ^            |                     |                    |              | · <b>· · · ·</b>    |         |              |
| Vorkflow Control                        |            |              |                     |                    |              |                     |         |              |
| C Enalos+                               |            |              |                     |                    |              | Node 5              |         |              |
| > G Modelling > G Molecular Descriptors |            |              |                     |                    |              |                     |         |              |
| > C NCI                                 |            | 11 M M       |                     |                    |              |                     |         |              |
| V C PubChem                             |            |              |                     |                    |              |                     |         |              |
| 🕥 Assay                                 |            |              |                     |                    |              |                     |         |              |
| S Assay Class                           |            |              |                     |                    |              |                     |         |              |
| G Main PubChem                          |            |              |                     |                    |              |                     |         |              |
| Patent to Sid                           |            |              |                     |                    |              |                     |         |              |
| G Sid                                   |            |              |                     |                    |              |                     |         |              |
| Similarity                              |            | \Xi Out      | line S?             |                    | Console 🔀    |                     | 🖹 🖬 🕅 🖃 | E • 📑 • 🗆    |
| C Vendor                                |            | - <u>~</u> ~ | - 203               |                    | NIME Console |                     |         | - 10         |

Fig. 5: Connecting nodes

## **Step 7-Configuring nodes**

Fully connected nodes showing a red status icon need to be configured. Start with the *Table Creator*, right click it and select "Configure" from the menu. You can rename "Column1" by "Name" or etc. You can also right click a random row and select "Row ID properties". In this menu you can set the prefix and the suffix of row IDs, as well as the number you want to start counting (usually: 0 (default) or 1). Then type in cell (1,1) the name of the chemical compound "abacavir". Press "Apply" and "OK" to close the dialog of the *Table Creator* node (Fig. 6). Once the node has been configured correctly, it switches to yellow (meaning ready for execution).



| Table Crea  | ator Settings Flow | v Variables | temory Policy |  |   |              |    |
|-------------|--------------------|-------------|---------------|--|---|--------------|----|
| Input line: |                    |             |               |  |   |              |    |
|             | S Name             |             |               |  |   |              |    |
| Row1        | abacavir           |             |               |  |   |              | 1  |
| Row2        |                    |             |               |  |   |              |    |
| Row3        |                    |             |               |  |   |              |    |
| Row4        |                    |             |               |  |   |              |    |
| Row5        |                    |             |               |  |   |              |    |
| Row6        |                    |             |               |  |   |              |    |
| Row7        |                    |             |               |  |   |              |    |
| Row8        |                    |             |               |  |   |              |    |
| Row9        |                    |             |               |  |   |              |    |
| Row 10      |                    |             |               |  |   |              |    |
| Row11       |                    |             |               |  |   |              |    |
| Row12       |                    |             |               |  |   |              |    |
| Row13       |                    |             |               |  |   |              |    |
| Row14       |                    |             |               |  |   |              |    |
| Row15       |                    |             |               |  |   |              |    |
| Row16       |                    |             |               |  |   |              |    |
| Row17       |                    |             |               |  |   |              |    |
| Row 18      |                    |             |               |  |   |              |    |
| Row 19      |                    |             |               |  |   |              |    |
| Row20       |                    |             |               |  |   |              |    |
| Row21       |                    |             |               |  |   |              |    |
| Row22       |                    |             |               |  |   |              |    |
| Row23       |                    |             |               |  |   |              | ١, |
|             | <                  |             |               |  |   | >            | 1  |
|             | table has 1 rows a |             |               |  | - | nt output ta | -  |

Fig. 6: Configuring Table Creator node

Then, configure *PubChem* node by choosing "Name" in the "Structure Identifier" and the "Input type" menus (Fig. 7). In this case, you have to choose "Name" because you have already given the name of the chemical compound (abacavir), while configuring the *Table Creator* node. Press "Apply" and "OK" to exit the dialog of the *Main PubChem*.

| 🛕 Dialog - 2:2 - Mair | PubChem     | ו            |        | - |        | $\times$ |
|-----------------------|-------------|--------------|--------|---|--------|----------|
| File                  |             |              |        |   |        |          |
| Enalos PubChem Flor   | w Variables | Memory Polic | y      |   |        |          |
| Structure Identifier  | Name        |              |        |   | $\sim$ |          |
| Input Type            | name        |              |        |   | $\sim$ |          |
|                       |             |              |        |   |        |          |
|                       |             |              |        |   |        |          |
|                       |             |              |        |   |        |          |
|                       |             |              |        |   |        |          |
|                       |             |              |        |   |        |          |
|                       |             |              |        |   |        |          |
|                       |             |              |        |   |        |          |
| ОК                    |             | Apply        | Cancel | ( | ?      |          |

Fig. 7: Configuring Main PubChem node

Subsequently, configure *Similarity* node. Choose a "Structure Identifier", an "Input Type" (CID for example) and a "Tanimoto Similarity" value as depicted in Fig. 8.



| 🛕 Dialog - 2:3 - Simili<br>ile | arity | 1              |               | _  | I    | ×    |
|--------------------------------|-------|----------------|---------------|----|------|------|
|                                |       |                |               |    |      |      |
| Enalos PubChem Similar         | rity  | Flow Variables | Memory Policy |    |      | <br> |
| Structure Identifier           | CID   | I              |               |    | ~    |      |
| Input Type                     | CID   |                |               |    | ~    |      |
| Tanimoto Similarity            |       |                |               |    | 99 🜲 |      |
|                                |       |                |               |    |      |      |
|                                |       |                |               |    |      |      |
|                                |       |                |               |    |      |      |
|                                |       |                |               |    |      |      |
|                                |       |                |               |    |      |      |
|                                |       |                |               |    |      |      |
| ОК                             |       | Apply          | Canc          | al | 0    |      |
| UK                             |       | Apply          | Canc          | ei | O    |      |

Fig. 8: Configuring Similarity node

Assay, Patent and Vendor nodes are configured by giving in the dialog window a "Structure Identifier" and an "Input Type". For more consistent results, when using PubChem Data Base, we usually choose CID for "Structure Identifier" and "Input Type" (Fig. 9).

| 📐 Dialog - 2:6 - Paten<br>ile | t              |               | _   |        | × |
|-------------------------------|----------------|---------------|-----|--------|---|
| Enalos PubChem Patent         | Flow Variables | Memory Policy |     |        |   |
| Structure Identifier          | CID            |               |     | ~      |   |
| Input Type                    | CID            |               |     | $\sim$ |   |
|                               |                |               |     |        |   |
|                               |                |               |     |        |   |
|                               |                |               |     |        |   |
|                               |                |               |     |        |   |
|                               |                |               |     |        |   |
|                               |                |               |     |        |   |
| ОК                            | Apply          | Can           | cel | ?      |   |

Fig. 9: Configuring Patent node

#### **Step 8-Executing nodes**

Now, right click on the *Assay* node and execute it. The workbench will execute all predecessor nodes for you. Then execute the *Patent* and the *Vendor* nodes. In a larger, more complex flow, you could select multiple nodes and trigger execution for all of them. The workflow manager will execute the nodes as needed, if possible in parallel. To execute all executable nodes press (Shift+F7).

#### **Step 9-Inspecting the results**

In order to examine the data and the results, open the nodes' views. From *PubChem* output port information is extracted about PubChem CID, IUPAC Name, InChI, InChI-Key Molecular Formula, Molecular Weight, Canonical SMILES and PubChem URL (Fig. 10).



| 🔥 Compound          | l information - | - 2:2 - Main PubChem — 🗆 🗙                                  |    |
|---------------------|-----------------|---|----|
| File                |                 |   |    |
| Table "default" - I | Rows: 1 Spec    | - Columns: 9 Properties Flow Variables                      |    |
| Row ID              | T CID           | S IUPAC Name  |    |
| Row 0               | 441300          | [(1S4R)-4-[2-amino-6-(cyclopropylamino)-9-purinyl]-1-cyclop | er |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     |                 |   |    |
|                     | <               |   | >  |
|                     |                 |   |    |

Fig. 10: Main PubChem results

*Similarity* node extracts the CIDs of the similar compounds, along with their molecular formula, their molecular weight and the number of rotatable bonds (Fig. 11).

| ble "default" - | Rows: 27 Spec - | Columns: 14 | Properties Flor | w Variables |        |   |
|-----------------|-----------------|-------------|-----------------|-------------|--------|---|
| Row ID          | S Initial r     | I CID       | S Molecul       | D Molecul   | Rotata |   |
| Row 0           | Row 0           | 441300      | "C14H18N6O"     | 286.339     | 4      |   |
| Row 1           | Row 0           | 6328608     | "C14H18N6O"     | 286.339     | 4      | - |
| Row 2           | Row 0           | 1971        | "C14H18N6O"     | 286.339     | 4      |   |
| Row 3           | Row 0           | 65140       | "C14H18N6O"     | 286.339     | 4      |   |
| Row 4           | Row 0           | 5742631     | "C14H18N6O"     | 286.339     | 4      |   |
| Row 5           | Row 0           | 469584      | "C14H18N6O"     | 286.339     | 4      |   |
| Row 6           | Row 0           | 15953951    | "C14H17N5O"     | 271.324     | 4      |   |
| Row 7           | Row 0           | 18646149    | "C15H20N6O"     | 300.366     | 5      |   |
| Row 8           | Row 0           | 18671944    | "C15H20N6O"     | 300.366     | 5      |   |
| Row 9           | Row 0           | 20598906    | "C14H18N6O"     | 286.339     | 4      |   |
| Row 10          | Row 0           | 57174705    | "C15H20N6O"     | 300.366     | 5      |   |
| Row 11          | Row 0           | 66939068    | "C14H18N6O"     | 286.339     | 4      |   |
| Row 12          | Row 0           | 73205335    | "C14H17N5O"     | 271.324     | 4      |   |
| Row 13          | Row 0           | 464151      | "C14H17N5O"     | 271.324     | 4      |   |
| Row 14          | Row 0           | 10660866    | "C14H18N6O"     | 286.339     | 4      |   |
| Row 15          | Row 0           | 18000486    | "C14H18N6O"     | 286.339     | 4      |   |
| Row 16          | Row 0           | 19767005    | "C14H19CIN      | 322.797     | 4      |   |

Fig. 11: Similarity results

Assay node exports information from PubChem about the Assay name, AID (BioAssay identification number), Bioassay Type, Bioactivity Outcome, Target IG, Target GeneID, Activity Value in  $\mu$ m (whenever available) and the Activity Name (Fig. 12).



| ble "default" - | Rows: 328 Sp | ec - Columns: 29 | Properties Flo | w Variables |           |    |
|-----------------|--------------|------------------|----------------|-------------|-----------|----|
| Row ID          | S AID        | S AID Ver        | S AID Re       | S Panel M   | S SID     |    |
| Row 0           | 880          | 2                | 1              |             | 26757979  | ۰, |
| Row 1           | 880          | 2                | 1              |             | 26757979  |    |
| Row 2           | 1469         | 1                | 1              |             | 26757979  |    |
| Row 3           | 1471         | 2                | 1              |             | 26757979  | -  |
| Row 4           | 1479         | 1                | 2              |             | 26757979  | -  |
| Row 5           | 1487         | 1                | 1              |             | 26757979  |    |
| Row 6           | 1490         | 2                | 1              |             | 26757979  |    |
| Row 7           | 1688         | 1                | 1              |             | 26757979  | -  |
| Row 8           | 1766         | 1                | 1              |             | 26757979  |    |
| Row 9           | 1766         | 1                | 1              |             | 26757979  |    |
| Row 10          | 1768         | 1                | 1              |             | 26757979  |    |
| Row 11          | 1768         | 1                | 1              |             | 26757979  |    |
| Row 12          | 2528         | 1                | 2              |             | 26757979  |    |
| Row 13          | 2546         | 1                | 1              |             | 26757979  |    |
| Row 14          | 2551         | 1                | 1              |             | 26757979  |    |
| Row 15          | 2662         | 2                | 1              |             | 26757979  |    |
| Row 16          | 200001       | 9                | 2              |             | 103463308 | •  |

#### Fig. 12: Assay results

*Patent* node exports the Patent IDs, where every single compound is found. For the compounds that are not found in any patent of the PubChem, the node will export an appropriate message (Fig. 13).

| 📐 Patent IDs ·<br>Ie | - 2:6 - Patent |                 |              | _              |          | ×          |
|----------------------|----------------|-----------------|--------------|----------------|----------|------------|
| able "default" -     | Rows: 23243 Sp | ec - Columns: 1 | 6 Properties | Flow Variables |          |            |
| Row ID               | S Initial r    | S Patent ID     | S Initial r  | CID            | S Mole   | cul.       |
| Row 0                | Row 0          | EP0628044A1     | Row 0        | 441300         | "C14H18M | 160        |
| Row 1                | Row 0          | EP0777669A1     | Row 0        | 441300         | "C14H18M | 16C        |
| Row 2                | Row 0          | EP0777669B1     | Row 0        | 441300         | "C14H18N | 160        |
| Row 3                | Row 0          | EP0817637A1     | Row 0        | 441300         | "C14H18N | 160        |
| Row 4                | Row 0          | EP0846110B1     | Row 0        | 441300         | "C14H18M | 160        |
| Row 5                | Row 0          | EP0910386A1     | Row 0        | 441300         | "C14H18N | 160        |
| Row 6                | Row 0          | EP0970100B1     | Row 0        | 441300         | "C14H18M | 160        |
| Row 7                | Row 0          | EP0994890A2     | Row 0        | 441300         | "C14H18M | <b>16C</b> |
| Row 8                | Row 0          | EP1003743A4     | Row 0        | 441300         | "C14H18M | <b>16C</b> |
| Row 9                | Row 0          | EP1009405A4     | Row 0        | 441300         | "C14H18M | 160        |
| Row 10               | Row 0          | EP1015642A1     | Row 0        | 441300         | "C14H18M | 16C        |
| Row 11               | Row 0          | EP1015642A4     | Row 0        | 441300         | "C14H18M | <b>16C</b> |
| Row 12               | Row 0          | EP1037633A1     | Row 0        | 441300         | "C14H18M | 160        |
| Row 13               | Row 0          | EP1054867A1     | Row 0        | 441300         | "C14H18M | 160        |
| Row 14               | Row 0          | EP1068568A2     | Row 0        | 441300         | "C14H18M | 160        |
| Row 15               | Row 0          | EP1079831A1     | Row 0        | 441300         | "C14H18M | 160        |
| Row 16               | Row 0          | EP1079846A4     | Row 0        | 441300         | "C14H18M | 160        |
|                      | <              |                 |              |                |          | >          |

Fig. 13: Patent results

*Vendor* node extracts information from PubChem about: PubChem SID, Compound URL, Supplier and Category of the Vendor (Fig. 14).



|                      | ails - 2:5 - Vendo | r             | - 0  | × |
|----------------------|--------------------|---------------|--|---|
| e<br>ble "default" - | Rows: 109 Spec     | - Columns: 19 | Properties Flow Variables  |   |
| Row ID               | S Initial r        | I SID         | S URL  |   |
| Row 6                | Row 0              | 131298145     | "http://www.amadischem.com/en-US/ProductDetail.aspx?catalog=A807079" |   |
| Row 7                | Row 0              | 136345668     | "http://www.synchem.de/chemical_Abacavir.html"                       |   |
| Row 8                | Row 0              | 160964383     | "http://www.chembase.cn/molecule-920.html"                           |   |
| Row 9                | Row 0              | 162011475     |  |   |
| Row 10               | Row 0              | 162176802     | "http://www.aksci.com/item_detail.php?cat=71906"                     |   |
|                      | <                  |               |  | > |

Fig. 14: Vendor results

The executed workflow is depicted in Fig. 15:

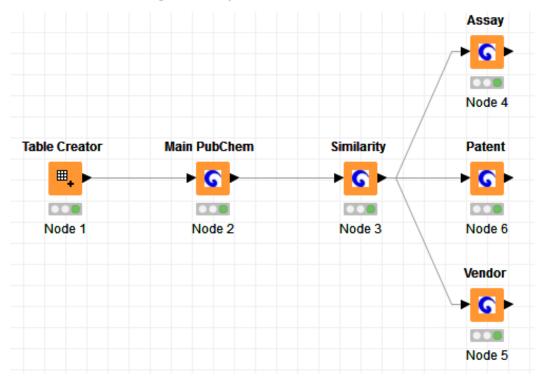


Fig. 15: Workflow using Enalos+ nodes

#### Step 10-Extending the main Workflow

Now you can extend the previous workflow by adding other NCI, PubChem and UniChem nodes. You can connect the *Main PubChem's* output with *CIR*, *UniChem* and *UniChem Connectivity*. To configure *CIR* node, you will need to specify the "Target Column" and the "Selected Representations" as shown in Fig. 16. In *UniChem* node you need to select InChI-Key as Input and consequently choose some of the available databases (Fig. 17). *UniChem Connectivity* node is configured in the same way as *UniChem* (Fig. 18).



| ▲ Dialog - 2:14 - CIR   |   | -  |  | Dialog - 2:12 - UniChem   |               |                    |   | - | × |
|---|---|--|--|---|---------------|--------------------|---|---|---|
| Enalos CIR Flow Variables Memory Policy   |   |  | Fi   |   |               |                    |   |   |   |
| Target column: InChIKey   |   |  |  | Enalos Unichem Flow Variables   | Memory Policy |                    |   |   | _ |
|   |   |  |  | InChiKey column InChIKey  |               | ~                  |   |   |   |
| Available representations:  |   | Selected representations:<br>Number of Hydrogen Bond Acceptors and Do  | opore  | Patents   | ~             | Category databases | Search in databases   |   |   |
| A Standard InfoRey<br>InchThey Simplified<br>Standard InfoRey<br>NCL(2AD FICTS Identifier<br>NCL(2AD FICTS Identifier<br>NCL(2AD FICTS Identifier<br>NCL(2AD FICTS Identifier<br>NCL(2AD FICTS Identifier<br>NCL(2AD FICTS Identifier<br>NCL(2AD FICTS Identifier<br>Chemispider ID<br>Notecular Weight<br>Chemispider IT<br>Number of Hydrogen Bond Acceptors<br>Number of Rule of 5 Violations<br>Number of Rule of 5 Violations<br>Number of Fictively Rotatable Bonds |   | value or nya oper ovatale Fonds<br>Number of Ferry Rotatale Fonds<br>Number of Rings<br>Number of Ring Systems |  | Add all Des   |               | Add DB             | SureCheNEL (surechembl )<br>IBM Patent System (bm)<br>PubChem Compounds (pubchem)<br>CheNeL (chembl)<br>BindraDB (bindingdb)<br>CheBI (cheb)<br>ZINC (chem)<br>ACToR (Chemical Toxisity) (actor)<br>Nikáz (nikázi)<br>PubChem: Thomson Pharma (pubche<br>evlolecules (molecules)<br>Remove DB | ^ |   |
|   |   |  |  |   |               |                    |   |   |   |
|   | OK                                      | Apply Cancel   | 0  |   |               | ОК                 | Apply Cancel  | ( |   |
| Fig. 16:  | Configuring                             | CIR node   |  |   | Fig. 17       | : Configuring U    | niChem node   |   |   |
| 0   | ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,, |  |  |   | _             | 0000               |   |   |   |
|   |   | △ Dialog - 2:13 - UniChem Co<br>File   | onnectivity  |   | >             | ζ                  |   |   |   |
|   |   | Enalos Unichem Connectivity Fi   | flow Variables Memory Policy                             |   |               |                    |   |   |   |
|   |   | InChilKey column InChilKey   | ~  |   |               |                    |   |   |   |
|   |   | Patents<br>Add al Dis  | Category data     Add DB                                 | ases Sauch in databases<br>DurchOBEs (unchonesis)<br>IMM Facture Composition (subdref)<br>Criefies (subdref)<br>Criefies (subdref)<br>Criefies (subdref)<br>Criefies (subdref)<br>ALCISE (Chemical Braues) (setter<br>National Streft)<br>ALCISE (Chemical Braues)<br>ALCISE (Chemical Braues)<br>ALCISE (Chemical Braues)<br>Criefies (subdref)<br>Remove DB | 0             |                    |   |   |   |
|   |   |  | match on FIXHB<br>The Query InChI matches the InChI assi | ned to the src compared id  | ~             |                    |   |   |   |
|   |   | Frequency block (0-500)  |  | an a set a Charlenn Ca  | 0             |                    |   |   |   |
|   |   | InChI length block (0-2000)<br>Labels A  | Add labels   |   | ••            |                    |   |   |   |
|   |   |  | only current   |   | ~             |                    |   |   |   |
|   |   |  |  |   |               |                    |   |   |   |
|   |   |  |  | OK Annhy Cance  |               | 1                  |   |   |   |

Fig. 18: Configuring UniChem Connectivity node

Then, try to add the following PubChem nodes as shown in Fig. 19.

- *Remove Duplicates* node removes the rows of the input table that contain the same values in selected columns. In this case you can select the CID column in the configuring menu (Fig. 23).
- Consequently, you can add 2 *Assay Class* nodes, in order to display both active and inactive compounds. In Fig. 24 you can see the "active" *Assay Class* configuring menu.
- In *Patent to Sid* node, give the Patent's ID as input, in order to obtain compounds' SIDs that are covered by the patents in request (Fig. 25).
- Finally, the *Sid* node, takes the compounds' SIDs as input and outputs the PubChem CIDs and synonyms (Fig. 26).



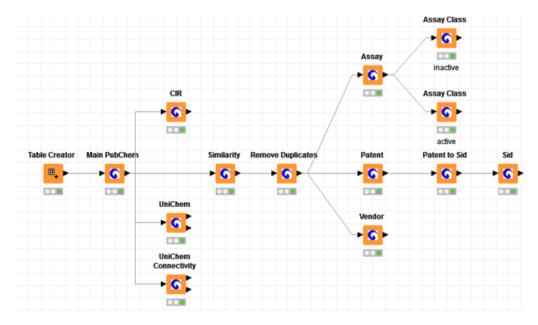


Fig. 19: Extended version of the main Workflow

- In order to control easily the data flow, you can use *Column Splitter* and *Row Splitter* nodes. Connect the *Column Splitter* to the output of *Remove Duplicates, Assay, Patent* and *Patent to Sid* nodes. You can select only one column for the Top partition and all the other columns for the Bottom partition (Fig. 21).<sup>1</sup> The Top partition Column will be used as input to the following nodes. For example, C.S.4 Top Partition ("SID") is used as *SID* node input. You can also connect *Row Splitter* to the output of *Assay* and *Patent* nodes, in order to export a data-table with less rows meeting the specified criteria (Fig. 22).
- The converted workflow, using the *Column Splitter* and *Row Splitter* nodes is depicted in Fig. 20.

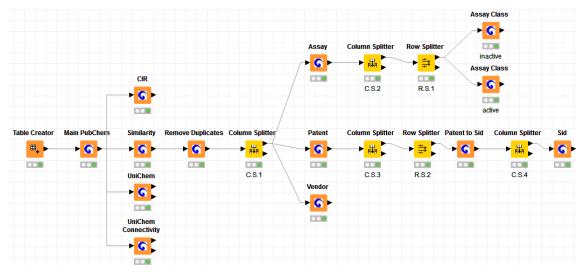


Fig. 20: Final version of the Workflow

<sup>&</sup>lt;sup>1</sup> In the 1<sup>st</sup> Column Splitter select only CID for the Top Partition

In the  $2^{nd}$  Column Splitter select only AID for the Top Partition

In the 3<sup>rd</sup> Column Splitter select only Patent ID for the Top Partition

In the 4<sup>th</sup> Column Splitter select only SID for the Top Partition



| 🛕 Dialog - 2:18 - Column Splitter (C.S.1)  | - 0   | × | ▲ Dialog - 2:16 - Row Splitter (R.S.)<br>File  | 2)  | - 🗆 | × |
|--|---|---|--|---|-----|---|
| File   |   |   | Filter Criteria Flow Variables Memo  | ary Policy  |     |   |
| Settings Flow Variables Memory Policy  |   |   | Providence Menu  | Set the filter criteria for the upper port  |     |   |
| Manual Selection      Wildcard/Regex Selection      Type   | pe Selection  |   |  | Row number range  |     |   |
| Top     Select     Bottom       Column(s):     Search     >  |   |   |  | First row number     1 •/•       to the end of the table       Last row number       10 •/• |     |   |
| CID     S Initial roo     S Molecula     Molecula     Constant     Constant | rFormula<br>Weight<br>eBondCount<br>ame<br>/<br>r Formula<br>r Weight<br>i SMILFS | • | Indude rows by attribute value     Exclude rows by attribute value     Indude rows by number     Exclude rows by number     Indude rows by row ID     Exclude rows by row ID |   |     |   |
| OK Apply   | Cancel  |   |  |   | 0   |   |
| Fig. 21: Configuring Column Sp   | litter  |   | Fig. 2   | 22: Configuring Row Splitter  |     |   |
| A Dislog 27 Parroya Duplicator   | – 🗆 ×   |   |  |   | _   |   |
| Dialog - 2:7 - Remove Duplicates File  | - U X   |   | 🛕 Dialog - 2:9 - Assay Cla<br>File   | ıss (active) —  |     | × |
| Enalos Remove Duplicates Flow Variables Memory Policy  |   |   |  |   |     |   |
| Columns' Selection Selected Colum  |   |   | Enalos PubChem Assay Cla   | SS Flow Variables Memory Policy   |     |   |
| Initial row id   |   |   | Assay ID (AID) AID   | ~   |     |   |
| MolecularFormula   |   |   |  |   |     |   |
| MolecularWeight<br>RotatableBondCount  |   |   | Output Type sids   | ~   | 1   |   |
| CID #1<br>IUPAC Name   |   |   | En. Type active  | ~   | /   |   |
| InChI  |   |   |  |   | _   |   |
| InChIKey<br>Molecular Formula  |   |   |  |   |     |   |
| Molecular Weight   |   |   |  |   |     |   |
| Canonical SMILES<br>PubChem URL  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
| Add Column Remove Colu   | an in   |   |  |   |     |   |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
| OK Apply Ca  | ncel  |   | ОК   | Apply Cancel  | ?   |   |
|  |   |   |  |   |     |   |
| Fig. 23: Configuring Remove Duplic   | ates node   |   | Fig. 24:   | Configuring Assay Class nod   | le  |   |
| 🛕 Dialog - 2:8 - Patent to Sid   | – 🗆 X   |   | 📐 Dialog - 2:15 - Sid  | _   |     | × |
| File   |   |   | ile  |   |     | ~ |
| riie   |   |   | lie  |   |     |   |
| Enalos PubChem Patent to SID Flow Variables Memory Police  | 4   |   | Enalos PubChem SID Flow  | Variables Memory Policy   |     |   |
| Structure Identifier Patent ID   | ~   |   | Structure Identifier   | SID   | ~   |   |
|  |   |   |  |   |     | · |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
| L  |   |   |  |   |     |   |
|  |   |   |  |   |     |   |
| OK Apply Cancel  |   |   | ОК   | Apply Cancel  | ?   |   |
|  | 1 1   |   |  | 26. Configuring Sidnada   |     |   |

Fig. 25: Configuring Patent to Sid node

Fig. 26: Configuring Sid node



Now, inspect the results of the added nodes:

- *CIR* node exports a table with all pre-requested representations (Fig. 27).
- UniChem and UniChem Connectivity nodes output the InChiKeys that were found in the preselected UniChem databases and the InChiKeys that were not found in these databases (Fig. 28).
- Assay Class node exports a table with information from PubChem about PubChem SIDs (Substance IDs) of compounds that are active or inactive (Fig. 29, Fig. 30).
- *Patent to Sid* extracts a table with information from PubChem about PubChem SIDs of compounds that are covered by the patents in request. In case that no compounds are found under a patent, the number -1 is assigned as SID of this patent (Fig. 31).
- Sid outputs information from PubChem about CIDs and synonyms (Fig. 32).

| 2  |  |  |  |                  |           | File   |  |   |   |                  |  |
|--|--|--|--|------------------|-----------|--|--|---|---|------------------|--|
| ble "New repr  | resentations" - Rov  | ws: 1 Spec - C   | olumns: 13 Prop  | erties Flow      | Variables | Table "InChiKey:   | found" - Rows: 18  | Spec - Col  | umns: 15 Properties   | Flow Variables   |  |
| Row ID   | S Input  |  | S Numb   | oer of Hyd       | S Number  | Row ID   | S Initial r  | S InChi K   | ey  | S Database       | Info                                   |
| Row 0  | MCGSCOLBFJ   | QGHM-SCZZXKL   | OSA-N 7  |                  | 4 5       | Row 0  | Row 0  | MCGSCOLBF   | JQGHM-SCZZXKLOSA-   | N A database of  | bioacti                                |
|  |  |  | I  |                  | 11        | Row 1  | Row 0  |   | JOGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 2  | Row 0  |   | JQGHM-SCZZXKLOSA-   |                  | resour                                 |
|  |  |  |  |                  |           | Row 3  | Row 0  |   | JQGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 4  | Row 0  |   | JOGHM-SCZZXKLOSA-   |                  | ly ava                                 |
|  |  |  |  |                  |           | Row 5  | Row 0  | MCGSCOLBF   | JQGHM-SCZZXKLOSA-   | N A free databas | se of c                                |
|  |  |  |  |                  |           | Row 6  | Row 0  | -   | JOGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 7  | Row 0  | MCGSCOLBF   | JQGHM-SCZZXKLOSA-   | N The data are p | orovide                                |
|  |  |  |  |                  |           | Row 8  | Row 0  |   | JQGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 9  | Row 0  | MCGSCOLBF   | JQGHM-SCZZXKLOSA-   | N SureChEMBL au  | utoma                                  |
|  |  |  |  |                  |           | Row 10   | Row 0  |   | JQGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 11   | Row 0  |   | JQGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 12   | Row 0  |   | JQGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 13   | Row 0  |   | JOGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 14   | Row 0  | MCGSCOLBF   | JQGHM-SCZZXKLOSA-   | N The LINCS DCI  | IC faci                                |
|  |  |  |  |                  |           | Row 15   | Row 0  |   | JQGHM-SCZZXKLOSA-   |                  |  |
|  |  |  |  |                  |           | Row 16   | Row 0  | -   | JQGHM-SCZZXKLOSA-   |                  |  |
| Active/Ina   | <<br>ctive IDs - 2:9 - A   | Fig. 27: C   |  | _                | ,<br>     |  | · · · · · ·  | 3. 28: Un   | iChem results   | - 0              | 2                                      |
| 2  | ctive IDs - 2:9 - A  | issay Class (act   | tive)  | _                |           | Active/Inac  | Fiz  | g. 28: Un<br>Assay Class (  | iChem results   | - 0              | >                                      |
| le "default" -   | ctive IDs - 2:9 - A  | ec - Columns: 3  | Properties Flor  | <br>w Variables  |           | Active/Inac<br>File<br>Table "default" -   | Fiz  | g. 28: Un<br>Assay Class (<br>pec - Column  | iChem results<br>inactive)<br>s: 3 Properties Flow  | - 0              | >                                      |
| le "default" -<br>Row ID   | ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r   | ssay Class (act<br>c - Columns: 3  | Properties Flo   | —<br>w Variables |           | Active/Inac<br>File<br>Table "default" -<br>Row ID   | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 S<br>S Initial r  | g. 28: Un<br>Assay Class (<br>pec - Column<br><b>S</b> SID  | iChem results<br>inactive)<br>s: 3 Properties Flow  | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 0  | ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r<br>Row 0  | ec - Columns: 3  | Properties Flow  | —<br>w Variables |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0  | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0   | g. 28: Un.<br>Assay Class (<br>pec - Column<br>§ SID<br>2433099   | iChem results<br>inactive)<br>s: 3 Properties Flow<br>S AID<br>880  | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 0<br>Row 1   | ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r<br>Row 0<br>Row 0   | ec - Columns: 3<br>S SID<br>204<br>710   | Properties Flow<br>S AID<br>880<br>880   | —<br>w Variables |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0<br>Row 1   | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 S<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130   | iChem results<br>inactive)<br>s: 3 Properties Flow<br>S AID<br>880<br>880   | - 0              | ;                                      |
| Row ID<br>Row 0<br>Row 1<br>Row 2  | ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r<br>Row 0<br>Row 0<br>Row 0  | ec - Columns: 3<br>S SID<br>204<br>710<br>1072   | Properties Flow<br>S AID<br>880<br>880<br>880  | —<br>w Variables |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2  | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0<br>Row 0<br>Row 0   | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726   | iChem results<br>inactive)<br>s: 3 Properties Flow<br>S AID<br>880<br>880<br>880  | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3   | ctive IDs - 2:9 - A<br>·Rows: 1007 Spr<br>S Initial r<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0   | assay Class (act<br>ac - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1302   | Properties Flow<br>8 AID<br>880<br>880<br>880<br>880<br>880                            |                  |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3   | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0   | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856  | iChem results<br>inactive)<br>s: 3 Properties Flow<br>880<br>880<br>880<br>880<br>880   | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4  | ctive IDs - 2:9 - A<br>•Rows: 1007 Spr<br><b>S</b> Initial r<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0  | assay Class (act<br>ac - Columns: 3<br>5 SID<br>204<br>710<br>1072<br>1302<br>1309   | Properties Flow<br>880<br>880<br>880<br>880<br>880                                     | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4  | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 S<br>S Initial r<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0  | g. 28: Un<br>Assay Class (<br>pec - Column<br>S SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892   | iChem results<br>inactive)<br>s: 3 Properties Flow<br>S AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880        | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5   | ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0  | ssay Class (act<br>ac - Columns: 3<br>5 SID<br>204<br>710<br>1072<br>1302<br>1309<br>1720  | tive) Properties Flov S AID 880 880 880 880 880 880 880 880 880 88                     |                  |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5  | <i>Fiz</i><br>tive IDs - 2:11 - <i>A</i><br>Rows: 228929 S<br><b>S</b> Initial r<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0<br>Row 0   | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587   | iChem results<br>inactive)<br>s: 3 Properties Flow<br>\$ AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88 | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 1<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6  | ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r<br>Row 0<br>Row 0   | ec - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1309<br>1720<br>1730   | Properties Flor<br><b>S</b> AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88 | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 1<br>Row 2<br>Row 4<br>Row 5<br>Row 6  | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br>S SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185  | iChem results<br>inactive)<br>s: 3 Properties Flow<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88           | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7  | Ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r<br>Row 0<br>Row 0  | Assay Class (act<br>ac - Columns: 3<br><b>S</b> SID<br>204<br>710<br>1072<br>1309<br>1720<br>1730<br>1830  | tive) Properties Flow 880 880 880 880 880 880 880 880 880 88                           | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7  | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 S<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247  | iChem results<br>inactive)<br>s: 3 Properties Flow<br><b>S</b> AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88         | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 6<br>Row 7<br>Row 8   | ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r<br>Row 0<br>Row 0   | ec - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1302<br>1309<br>1720<br>1730<br>1830<br>2051   | tive)  Properties Flow  S AID  880  880  880  880  880  880  880  8                    | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8  | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247<br>280907  | iChem results<br>inactive)<br>s: 3 Properties Flow<br><b>S</b> AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88         | - 0              | ;                                      |
| Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9  | ctive IDs - 2:9 - A<br>Rows: 1007 Spe<br>S Initial r<br>Row 0<br>Row 0      | assay Class (act<br>ec - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1302<br>1309<br>1720<br>1730<br>1830<br>2051<br>2259   | tive)  Properties Flow  S AID  880  880  880  880  880  880  880  8                    | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 8<br>Row 9  | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247<br>280907<br>1314338   | iChem results<br>inactive)<br>s: 3 Properties Flow<br>S AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88                | - 0              | >                                      |
| Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10   | ctive IDs - 2:9 - A<br>-Rows: 1007 Spe<br><b>S</b> Initial r<br>Row 0<br>Row 0<br>Ro | assay Class (act<br>ac - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1309<br>1720<br>1730<br>1830<br>2051<br>2259<br>2794   | tive)  Properties Flow  880 880 880 880 880 880 880 880 880 8                          | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 1<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10   | Fig<br>trive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0<br>Row 0 | g. 28: Un.<br>ssay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247<br>280907<br>1314338<br>2301266   | iChem results<br>inactive)<br>s: 3 Properties Flow<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88           | - 0              | ;                                      |
| Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10<br>Row 11   | ctive IDs - 2:9 - A<br>•Rows: 1007 Spr<br><b>S</b> Initial r<br>Row 0<br>Row 0<br>Ro | assay Class (act<br>c - Columns: 3<br>5 SID<br>204<br>710<br>1072<br>1309<br>1720<br>1730<br>1830<br>2051<br>2259<br>2794<br>3194                                  | tive)  Properties Flov  S AID  880  880  880  880  880  880  880  8                    | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 1<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10<br>Row 10 | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un<br>Assay Class (<br>pec - Column<br>S SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247<br>280907<br>1314338<br>2301266<br>3243963   | iChem results<br>inactive)<br>s: 3 Properties Flow<br><b>S</b> AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88         | - 0              | `````````````````````````````````````` |
| le "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 9<br>Row 10<br>Row 11<br>Row 12                      | Ctive IDs - 2:9 - A  | Assay Class (act<br>ec - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1309<br>1720<br>1730<br>1830<br>2051<br>2259<br>2294<br>3194<br>3383                         | tive)  Properties Flov  S AID  880  880  880  880  880  880  880  8                    | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 1<br>Row 2<br>Row 4<br>Row 5<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 8<br>Row 9<br>Row 10<br>Row 11<br>Row 12                  | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 s<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br>S SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247<br>280907<br>1314338<br>2301266<br>3243963<br>567049                              | iChem results<br>inactive)<br>s: 3 Properties Flow<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88           | - 0              | ;                                      |
| le "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10<br>Row 11<br>Row 12<br>Row 13                     | Ctive IDs - 2:9 - A  | Assay Class (act<br>ec - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1302<br>1720<br>1730<br>1830<br>2051<br>2259<br>2794<br>3194<br>3383<br>3450                 | tive)  Properties Flow  880  880  880  880  880  880  880  8                           | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10<br>Row 11<br>Row 12<br>Row 13                 | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 S<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247<br>280907<br>1314338<br>2301266<br>3243963<br>567049<br>6023095            | iChem results<br>inactive)<br>s: 3 Properties Flow<br><b>S</b> AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88         | - 0              |  |
| le "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10<br>Row 11<br>Row 11<br>Row 12<br>Row 13<br>Row 14 | Ctive IDs - 2:9 - A<br>Rows: 1007 Spanner<br>Row 0<br>Row | Assay Class (act<br>ac - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1302<br>1730<br>1730<br>1730<br>1830<br>2051<br>2259<br>2794<br>3194<br>3383<br>3450<br>3589 | tive)  Properties Flow 880 880 880 880 880 880 880 880 880 88                          | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10<br>Row 10<br>Row 11<br>Row 12<br>Row 13<br>Row 14      | Fig<br>tive IDs - 2:11 - A<br>Rows: 228929 S<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247<br>280907<br>1314338<br>2301266<br>3243963<br>567049<br>6023095<br>5331183 | iChem results<br>inactive)<br>s: 3 Properties Flow<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88           | - 0              | >                                      |
| le "default" -   | Ctive IDs - 2:9 - A  | Assay Class (act<br>ec - Columns: 3<br>S SID<br>204<br>710<br>1072<br>1302<br>1720<br>1730<br>1830<br>2051<br>2259<br>2794<br>3194<br>3383<br>3450                 | tive)  Properties Flow  880  880  880  880  880  880  880  8                           | w Variables      |           | Active/Inac<br>File<br>Table "default" -<br>Row ID<br>Row 0<br>Row 1<br>Row 2<br>Row 3<br>Row 4<br>Row 5<br>Row 6<br>Row 7<br>Row 8<br>Row 9<br>Row 10<br>Row 11<br>Row 12<br>Row 13                 | Fiz<br>tive IDs - 2:11 - A<br>Rows: 228929 S<br>S Initial r<br>Row 0<br>Row 0  | g. 28: Un.<br>Assay Class (<br>pec - Column<br><b>S</b> SID<br>2433099<br>1259130<br>714726<br>6602856<br>2598892<br>707587<br>16190185<br>1188247<br>280907<br>1314338<br>2301266<br>3243963<br>567049<br>6023095            | iChem results<br>inactive)<br>s: 3 Properties Flow<br><b>S</b> AID<br>880<br>880<br>880<br>880<br>880<br>880<br>880<br>88         | - 0              | >                                      |



| e               |                |                |                 |           |   | File              |                |                |              |                |        |   |
|-----------------|----------------|----------------|-----------------|-----------|---|-------------------|----------------|----------------|--------------|----------------|--------|---|
| ble "default" - | Rows: 2035 Spe | c - Columns: 3 | Properties Flow | Variables |   | Table "default" - | Rows: 2035 Spe | c - Columns: 6 | Properties I | Flow Variables |        |   |
| Row ID          | S Initial r    | I SID          | S Patent ID     |           |   | Row ID            | S Initial r    | I SID          | I CID        | S synonyms     | S name | e |
| Row 0           | Row 0          | 127415408      | EP0628044B1     |           | ~ | Row 51            | Row 51         | 135764025      | 22721337     |                | 816    | ī |
| Row 1           | Row 0          | 127427561      | EP0628044B1     |           |   | Row 52            | Row 52         | 135855649      | 11108834     |                | 816    |   |
| Row 2           | Row 0          | 127564905      | EP0628044B1     |           |   | Row 53            | Row 53         | 135855990      | 21310051     |                | 816    |   |
| Row 3           | Row 0          | 127602502      | EP0628044B1     |           |   | Row 54            | Row 54         | 135933200      | 11459511     |                | 816    |   |
| Row 4           | Row 0          | 127720939      | EP0628044B1     |           |   | Row 55            | Row 55         | 135996759      | 57133849     |                | 816    |   |
| Row 5           | Row 0          | 127804557      | EP0628044B1     |           |   | Row 56            | Row 56         | 136042524      | 57155707     |                | 816    | Ĩ |
| Row 6           | Row 0          | 127807700      | EP0628044B1     |           |   | Row 57            | Row 57         | 136045151      | 57157246     |                | 816    | Ĩ |
| Row 7           | Row 0          | 127965093      | EP0628044B1     |           |   | Row 58            | Row 58         | 136049641      | 57159919     |                | 816    |   |
| Row 8           | Row 0          | 128037226      | EP0628044B1     |           |   | Row 59            | Row 59         | 136178896      | 57236764     |                | 816    |   |
| Row 9           | Row 0          | 128090495      | EP0628044B1     |           |   | Row 60            | Row 60         | 136209496      | 57255035     |                | 816    | Ĩ |
| Row 10          | Row 0          | 128110507      | EP0628044B1     |           |   | Row 61            | Row 61         | 226393245      | 6101         | SCHEMBL34      | 4488   | Ĩ |
| Row 11          | Row 0          | 128165640      | EP0628044B1     |           |   | Row 62            | Row 62         | 226393282      | 6547         | SCHEMBL122     | 4488   |   |
| Row 12          | Row 0          | 128167461      | EP0628044B1     |           |   | Row 63            | Row 63         | 226393293      | 244          | SCHEMBL147     | 4488   | Ĩ |
| Row 13          | Row 0          | 128211172      | EP0628044B1     |           |   | Row 64            | Row 64         | 226393308      | 21226206     | SCHEMBL172     | 4488   |   |
| Row 14          | Row 0          | 128246082      | EP0628044B1     |           |   | Row 65            | Row 65         | 226393327      | 7509         | SCHEMBL205     | 4488   |   |
| Row 15          | Row 0          | 128310878      | EP0628044B1     |           |   | Row 66            | Row 66         | 226393487      | 7799         | SCHEMBL454     | 4488   |   |
| Row 16          | Row 0          | 128413036      | EP0628044B1     |           |   | Row 67            | Row 67         | 226393712      | 107428       | SCHEMBL762     | 4488   |   |
| Row 17          | Row 0          | 128428285      | EP0628044B1     |           | ~ |                   | <              |                |              |                |        |   |

## Embark your own voyage of discovery!

Now, this was just a simple example to get you started. There is a lot more to discover. Try to explore it! We tried to keep it simple and intuitive. We would love to receive your feedback and find out what you liked and what you did not like; things you find not functional or things that did not seem to work.