



Welcome to Working with the RDKit in KNIME Analytics Platform

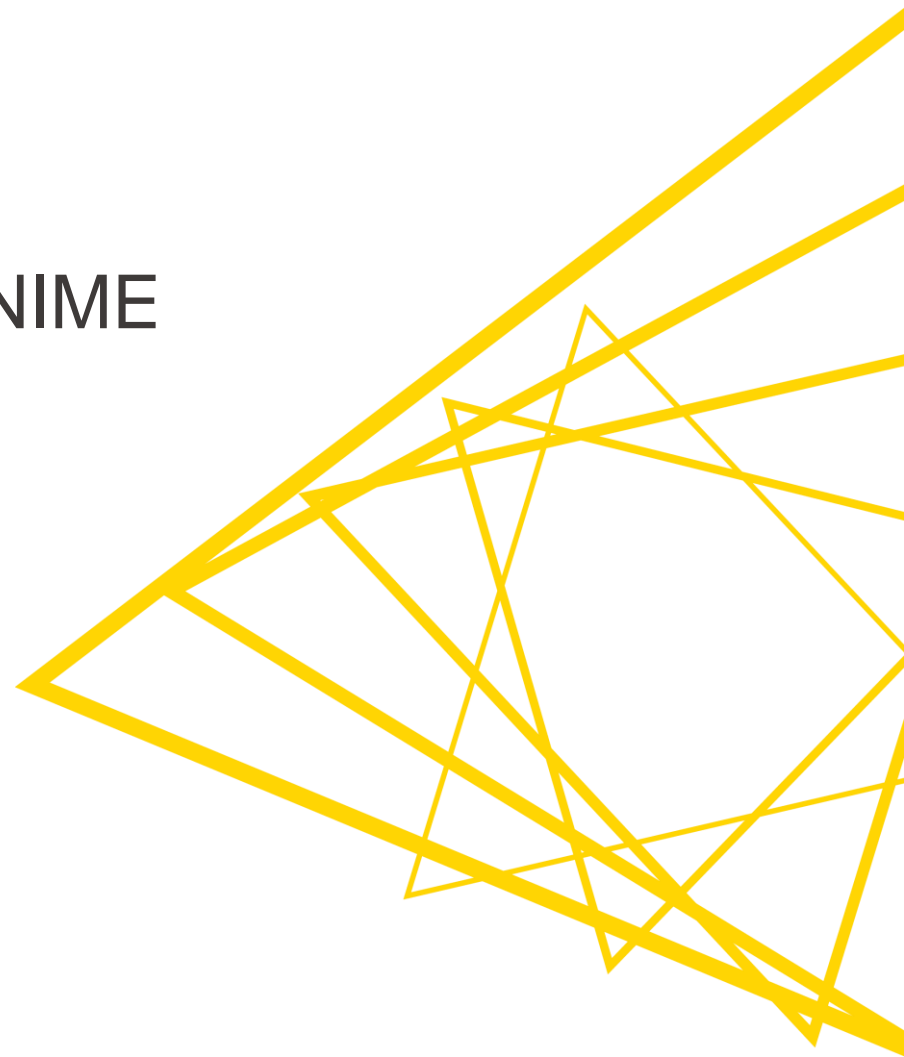
Going live at:

Chicago 11:00 am

San Francisco 9:00 am

New York 12:00 pm

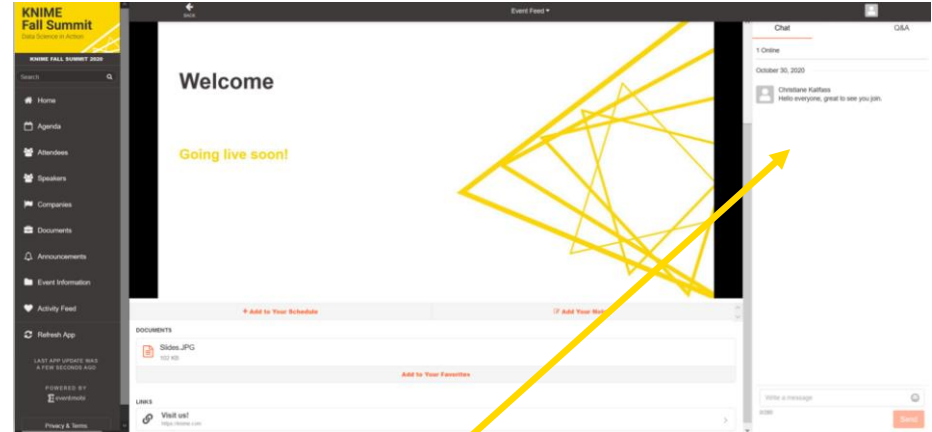
Berlin 6:00 pm



Housekeeping

- Post in the chat where you are dialing in from and discuss with other attendees
- Questions? Post them in the Q&A

Questions will be answered after the presentation.




The agenda


- Brief introduction to KNIME
- **Part I: Alice**
Reactions in RDKit
- **Part II: Greg**
3D visualization of molecules

Post your questions in the forum



<https://forum.knime.com/c/special-interest-groups/cheminformatics/36>



Open for Innovation
KNIME

Hub Blog Forum Events Use Cases Career Contact [Download](#) 

SOFTWARE / PRICING / COMMUNITY / LEARNING / PARTNERS / ABOUT






















[Log In](#)  

Community Groups ▾

Cheminformatics ▾

Latest

Top

| Topic | | Replies | Views | Activity |
|--|---|---------|-------|----------|
| Altering the structure of a chemical compound |   | 4 | 57 | 12m |
| Identifying fragments of a multi-fragment structure |    | 4 | 51 | 7d |
| Extracting data from MedChem patent? |    | 3 | 76 | 18d |
| Error using CCG node |   | 1 | 62 | 23d |
| Input data format for Model Acceptability Criteria node enalos |    | 7 | 318 | Oct 14 |
|   3D rmsd of substructure vs larger molecule |    | 13 | 160 | Oct 8 |
| Error in workflow execution with Padel |    | 3 | 151 | Sep 28 |

The workflows

https://hub.knime.com/knime/spaces/Life%20Sciences/latest/Events/2020_11_20_KNIME_Summit_RDKit_Workshop/

Public space

Life Sciences

Last update: 25 Jun 2020

Home > Events > 2020_11_20_KNIME_Summit_RDKit_Workshop



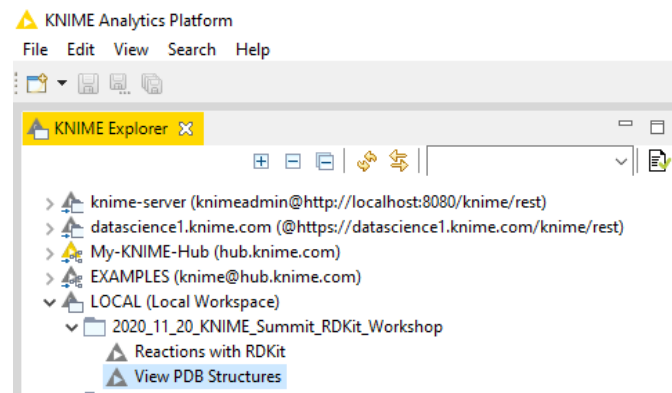
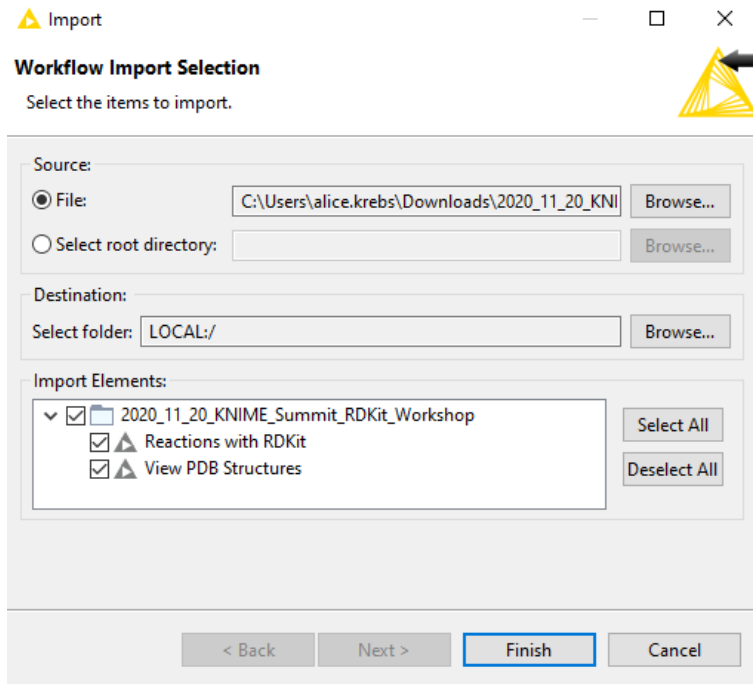
Reactions with RDKit



View PDB Structures

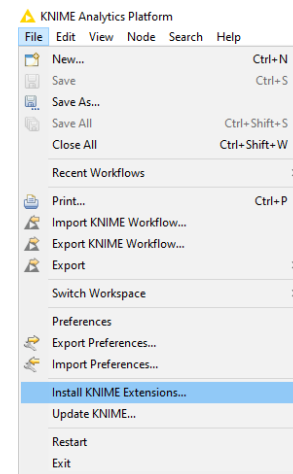
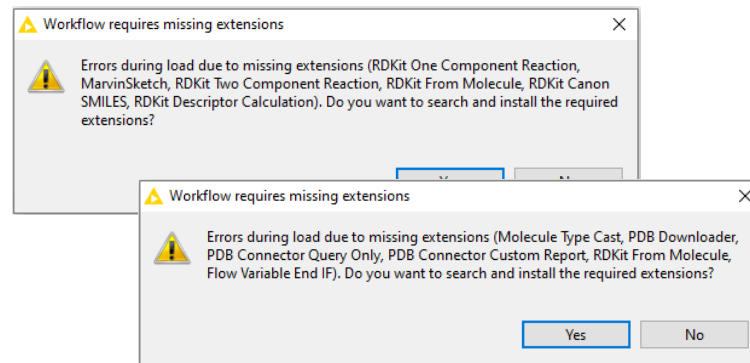


Download



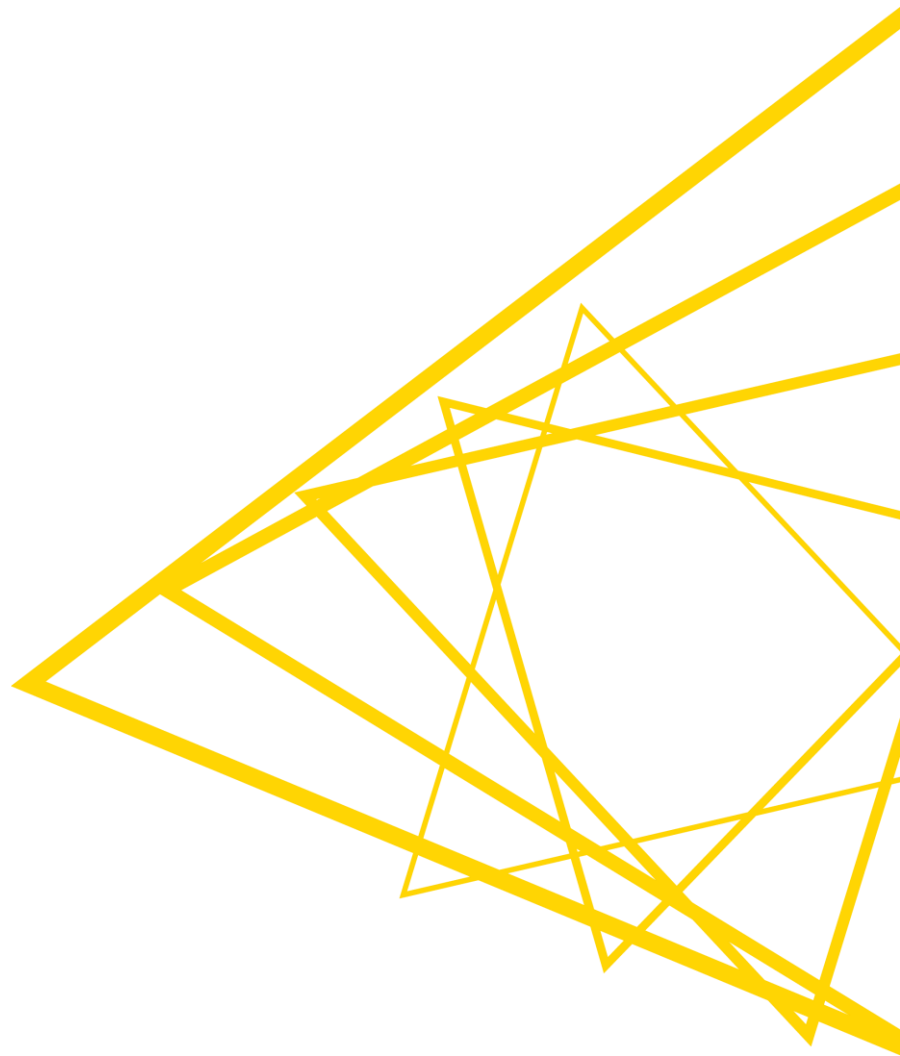
Install extensions

- Open the workflows...
- ... or via File > Install KNIME Extensions
 - ChemAxon/Infocom Marvin Extensions Feature
 - RDKit KNIME integration
 - KNIME Base Chemistry Types & Nodes
 - Vernalis KNIME Nodes





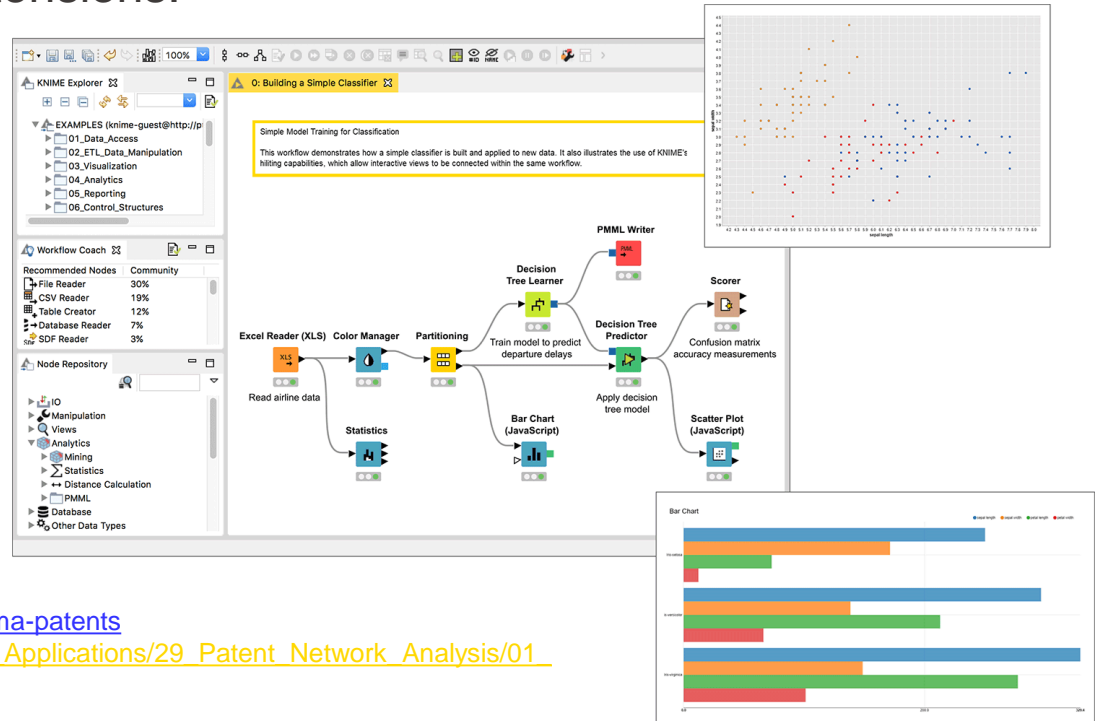
Introduction to KNIME



What is KNIME Analytics Platform?

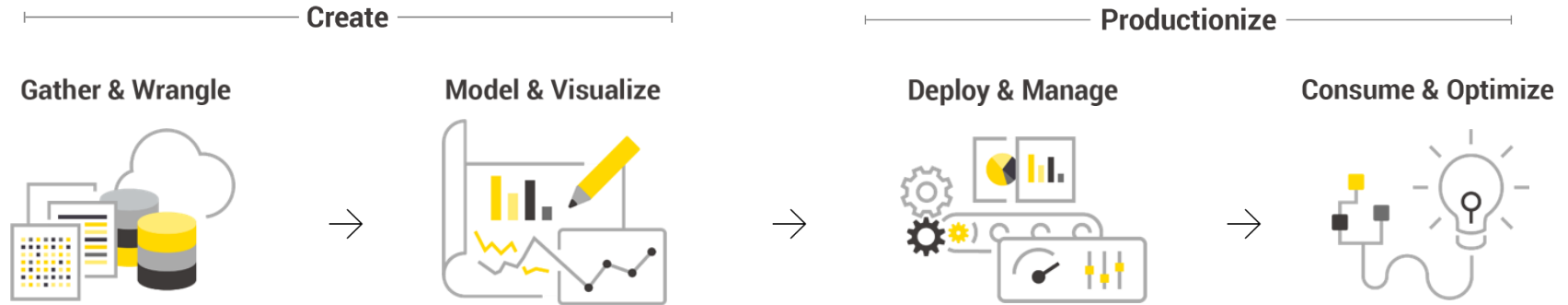
- A tool for data analysis, manipulation, visualization, and reporting
- Based on the graphical programming paradigm
- Provides a diverse array of extensions:

- Text Mining
- Network Mining
- Cheminformatics
- Many integrations, such as Java, R, Python, Weka, Keras, Plotly, H2O, etc.



- <https://www.knime.com/competitive-intelligence-for-pharma-patents>
- https://hub.knime.com/knime/spaces/Examples/latest/50_Applications/29_Patent_Network_Analysis/01_Tarceva_neighbor_network_from_SureChEMBL

KNIME Software - One Ecosystem



KNIME Analytics Platform

KNIME Extensions

KNIME Integrations

Community Extensions

Partner Extensions

KNIME Server

KNIME WebPortal

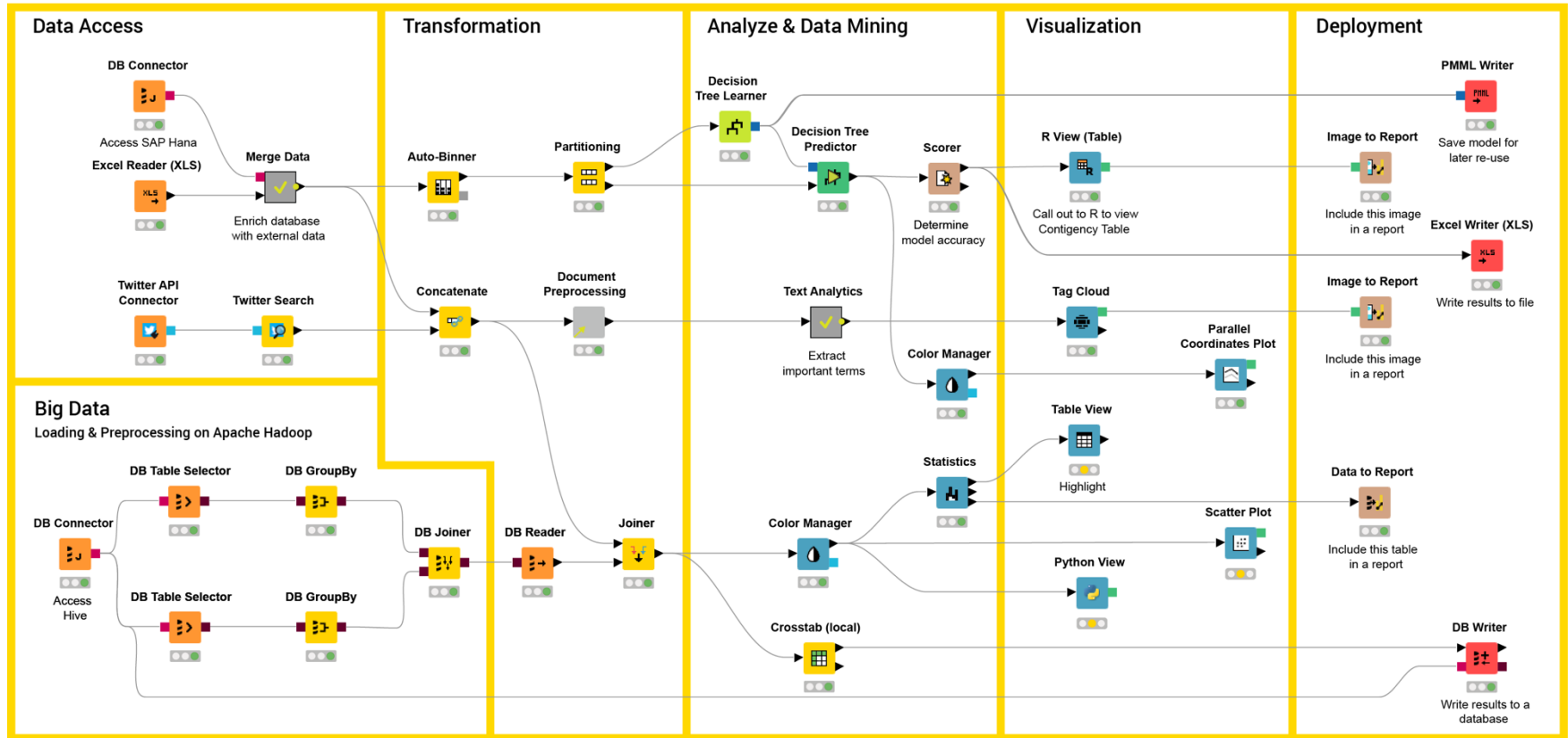
Data Science as a Service

The KNIME Workbench


The screenshot displays the KNIME Workbench interface with several key components highlighted by yellow callouts:

- Servers and Workflows:** Points to the **KNIME Explorer** on the left, which shows a project tree with folders like `data`, `exercises`, and `solutions`, and specific files like `01_Chemistry`.
- Node Recommendations:** Points to the **Workflow Coach** panel, which lists recommended nodes such as `File Reader` (28%), `CSV Reader` (20%), and `Table Creator` (13%).
- Node Repository:** Points to the **Node Repository** panel, which provides a categorized list of nodes including `IO`, `Manipulation`, `Views`, `DB`, `Analytics`, `Other Data Types`, `Structured Data`, `Scripting`, `Tools & Services`, `Community Nodes`, `KNIME Labs`, `Workflow Control`, `Workflow Abstraction`, `Reporting`, `Chemistry`, and `ChemAxon / Infocom`.
- Workflow Editor:** The central workspace showing a workflow with two steps: **Step 1: Read data from different sources** (containing `File Reader` and `SDF Reader`) and **Step 2: Remove duplicates** (containing `Concatenate`, `Molecule Type Cast`, `RDKit Canon SMILES`, and `Duplicate Row Filter`).
- Node Description:** Points to the **Description** panel on the right, which shows the title `01_Chemistry_basics` and a detailed description of the workflow's purpose and data sources.
- KNIME Hub:** Points to the **KNIME Hub Search** panel, which allows searching for workflows, nodes, and more.
- Console:** Points to the **Console** panel at the bottom right, which displays log messages and warnings, such as `Log file is located at: /Users/daria_knime/knime-workspace_rdkit_20200421/metadata/knime/` and various `WARN` messages about column specifications.
- Outline:** Points to the **Outline** panel at the bottom left, which provides a visual overview of the workflow structure.

3000+ Nodes for all Steps of End-To-End Data Science



KNIME Hub: Searching, Sharing, and Collaborating

Hub

Welcome to the

KNIME Hub

Solutions for data science: find workflows, nodes and components, and collaborate in spaces.

3 960

Nodes

368


Components

3 188

Workflows

212

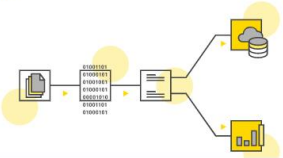
Extensions



Data Science in Action
KNIME Fall Summit 2020
Nov 16-20, Online
THE KNIME SHOW


How to
Getting started

From downloading through building your first workflow



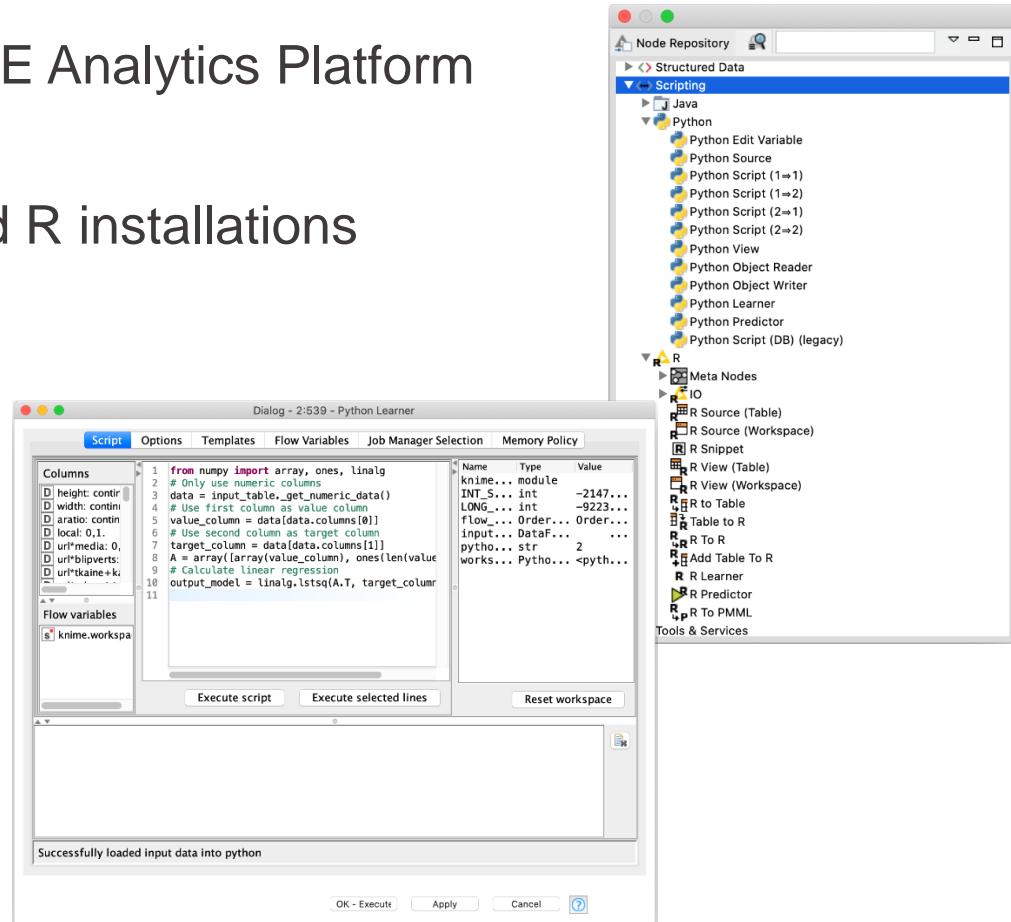
Verified Components

Easily reuse bundled functionalities, verified by KNIME Experts



Scripting Integrations: R and Python

- Run R or Python code in KNIME Analytics Platform
- Works with existing Python and R installations
- Syntax highlighting support
- Different nodes for many tasks



Selected Open Source Extensions for Cheminformatics

- RDKit
 - Converters
 - Modifiers
 - Calculators
 - RDKit Descriptor Calculation
 - RDKit Calculate Charges
 - Geometry
 - RDKit Generate Coords
 - RDKit Optimize Geometry
 - RDKit Add Conformers
 - RDKit Open 3D Alignment
 - RDKit RMSD Filter
 - Fingerprints
 - RDKit Fingerprint
 - RDKit Count-Based Fingerprint
 - RDKit Fingerprint Reader
 - RDKit Fingerprint Writer
 - RDKit Diversity Picker
 - Fragments
 - Searching
 - Reactions
 - Viewing
 - RDKit Interactive Table
 - RDKit SMILES Headers
 - RDKit Molecule Highlighting
 - Experimental
- CDK
 - 3D
 - 3D Coordinates
 - 3D D-Moments
 - 3D D-Similarity
 - 3D RMSD
 - 3D Viewer
 - 3D WHIM
 - AMBIT
 - I/O
 - 2D Coordinates
 - Atom Signatures
 - ChemSpider
 - Connectivity
 - Depiction
 - Element Filter
 - Fingerprint Similarity
 - Fingerprints
 - Hydrogen Manipulator
 - Lipinski's Rule-of-Five
 - Mass Calculator
 - Molecular Properties
 - OPSIN
 - SMARTS Query
 - Structure Sketcher
 - Substructure Search
 - Sugar Remover
 - Sum Formula
 - Symmetry
 - XLogP
- Erlwood Nodes
 - IO
 - Structure Data Format Converters
 - Structure Similarity
 - Fingerprint Similarity
 - Structure Properties
 - Plane of Best Fit Calculator
 - Virtual Screening
 - Virtual Screening Metrics
 - Evaluation and Ranking
 - Desirability Ranking
 - Pareto Ranking
 - SAR Analysis
 - Automated Matched Pairs
 - Free-Wilson Matched Pairs
 - Viewers
 - 2D/3D Scatterplot
 - Activity Cliffs Viewer
 - Similarity Viewer
 - Testing
- Vernalis
 - Databases
 - European PubMed Central
 - European PubMed Central Advanced Search
 - Fingerprints
 - Flow Control
 - IO
 - Matched Molecular Pairs (MMPs)
 - Filtering
 - Fragmentation
 - Pair Generation
 - Rendering
 - Transforms
 - MMP Calculate Maximum Cuts (RDKit)
 - MMP Fragmentation Type Loop Start
 - Uniquify IDs
 - Principal Moments of Inertia (PMIs)
 - RCSB PDB Tools
 - PDB Connector
 - PDB Connector (XML Query)
 - PDB Connector Combine XML Queries
 - PDB Connector Custom Report
 - PDB Connector Query Only
 - PDB Connector Query Only (XML Query)
 - PDB Connector XML Query Builder
 - PDB Describe Heterogens
 - PDB Downloader (Source)
 - PDB SMILES Query
 - PDB Downloader
 - Local PDB Tools
 - Sequence Tools
 - Speedy SMILES
 - Testing
 - Miscellaneous

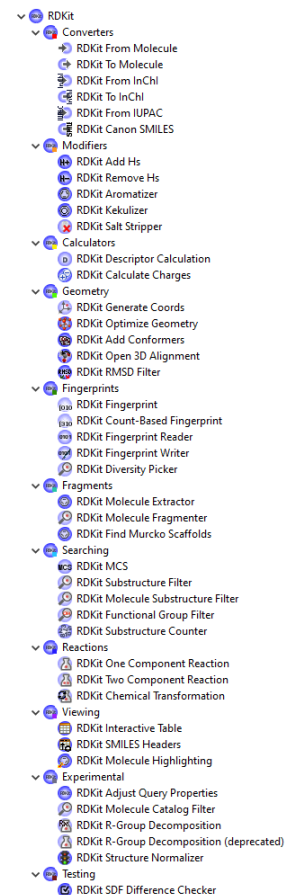
Selected Commercial Extensions for Cheminformatics

- ▼ BioSolveIT Nodes
 - ▶ CoLibri (Chemistry Spaces)
 - ▶ IO
 - Assess Affinity with Hyde in SeeSAR
 - Compute FTrees Similarity
 - Compute FlexS Alignments
 - Compute LeadIT Docking
 - Convert Molecules with Naomi
 - FTrees Query Generator
 - Filter Molecules with Naomi
 - FlexX Docking
 - Generate 3D Coordinates
 - Generate Protomers / Tautomers with Naomi
 - Interactive BioSolveIT Table
 - Interactive SeeSAR Viewer
 - Prepare Receptor with LeadIT
 - Run ReCore Interactively
 - Search FTrees Fragment Space
 - SeeSAR Project Generator
- ▼ ChemAxon / Infocom
 - ▼ JChem
 - ▶ IO
 - ▶ Converter
 - ▶ Marvin
 - ▶ Calculator Plugins
 - ▶ JChem Base
 - ▶ JChem Cartridge
 - ▶ Standardizer
 - ▶ Structure Checker
 - ▶ Name to Structure
 - ▶ Screen
 - ▶ JKlustor
 - ▶ Reactor
 - ▶ Markush Viewer
 - ▶ Metabolizer
 - ▶ Fragmenter
 - ▶ Marvin
- ▼ Cresset
 - ▼ Forge
 - ▶ Models
 - ▶ Project
 - ▶ Forge Align
 - ▶ Activity Miner
 - ▶ FieldTemplater
 - ▼ Spark
 - ▶ Spark Fragment Selector
 - ▶ Generate Spark Database
 - ▶ Spark Database Search
 - ▼ XedTools
 - ▶ XedMin
 - ▶ XedeX
 - ▶ Torch/Forge Molecule Viewer
- ▼ MOE
 - ▶ Input
 - ▶ Output
 - ▶ Convert
 - ▶ Transform
 - ▶ Process
 - ▶ Calculate
 - ▶ QuaSAR
 - ▶ Fingerprints
 - ▶ Simulations
 - ▶ Bioinformatics
 - ▶ Fragment Based Design
 - ▶ CombiChem
 - ▶ Miscellaneous
 - ▶ Pharmacophore
 - ▶ Materials
- ▼ Schrödinger
 - ▶ Readers/Writers
 - ▶ Converters
 - ▶ Ligand Preparation
 - ▶ Property Generation
 - ▶ Cheminformatics
 - ▶ Pharmacophore Modeling
 - ▶ Protein Structure Prediction
 - ▶ Docking and Scoring
 - ▶ Molecular Mechanics
 - ▶ Molecular Dynamics
 - ▶ Quantum Mechanics
 - ▶ Workflows
 - ▶ Filtering
 - ▶ Reporting
 - ▶ Scripting
 - ▶ Tools

What is the RDKit?

- Open source cheminformatics library in C++
- Wrappers for KNIME maintained by the open source community
- Useful for:
 - Descriptor calculation
 - Cleaning structures
 - InChI conversion
 - Canonical SMILES
 - Fingerprints
 - Scaffolds/substructures
 - Reaction simulation
 - and more...

<https://www.rdkit.org>

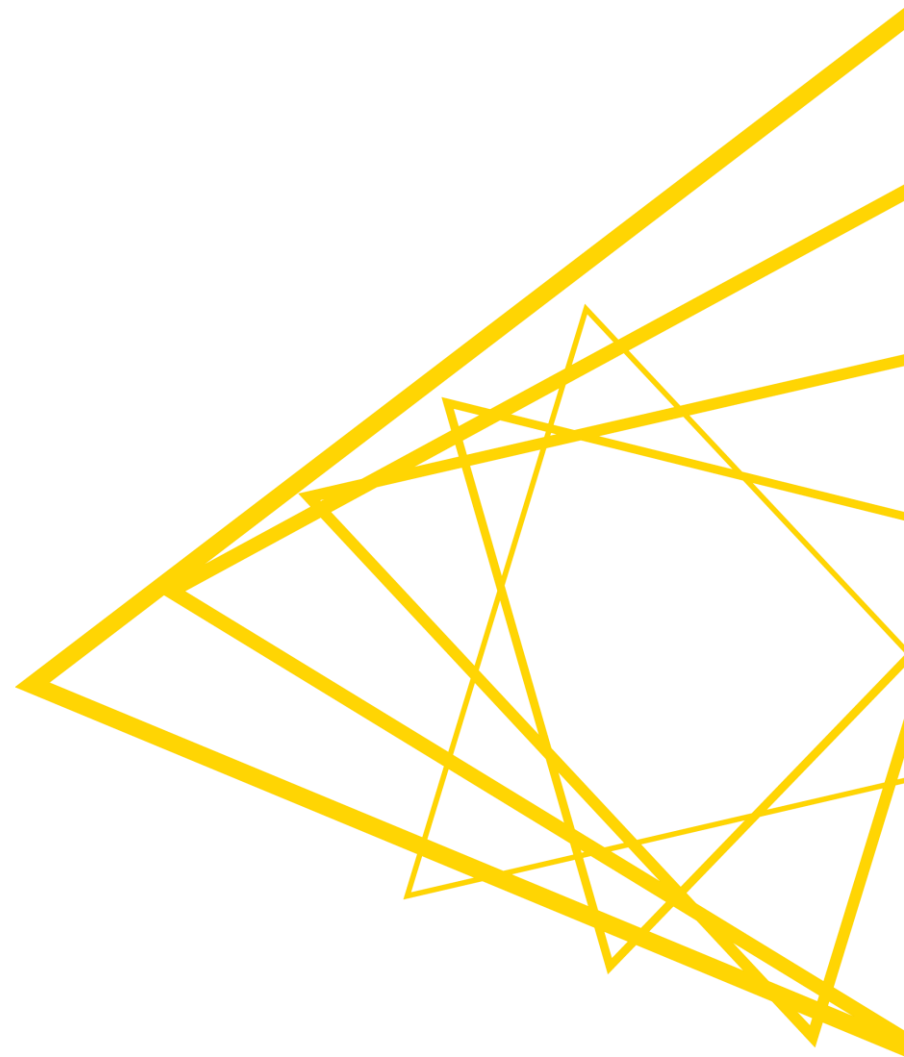


Reactions with RDKit

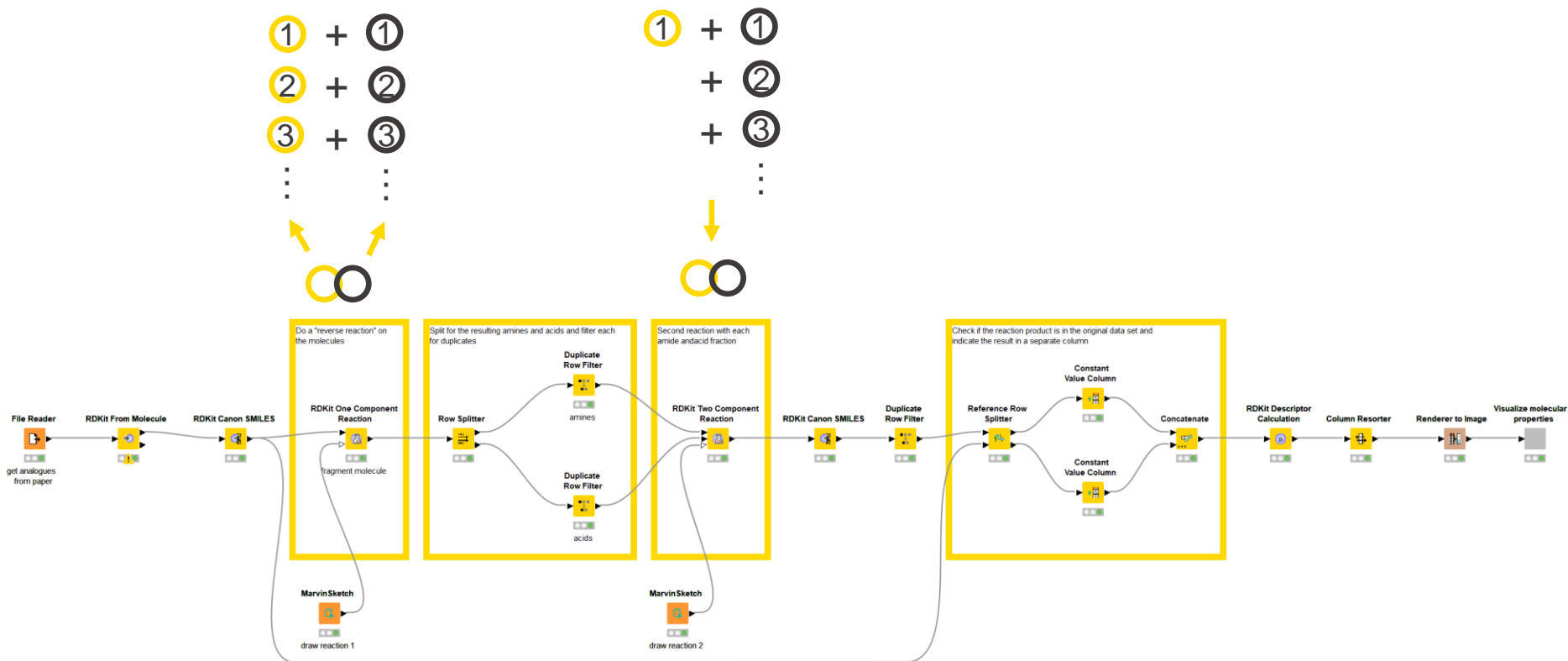
Alice Krebs

KNIME Fall Summit Webinar

20 November 2020



The workflow



https://hub.knime.com/knime/spaces/Life%20Sciences/latest/Events/2020_11_20_KNIME_Summit_RDKit_Workshop/Reactions%20with%20RDKit

The data

Journal of
**Medicinal
Chemistry**

pubs.acs.org/jmc

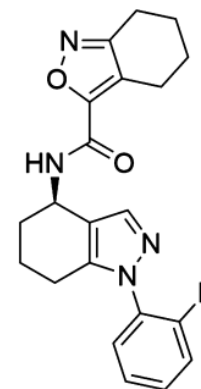
This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.



Article

Optimization of Tetrahydroindazoles as Inhibitors of Human Dihydroorotate Dehydrogenase and Evaluation of Their Activity and In Vitro Metabolic Stability

Gergana Popova,* Marcus J. G. W. Ladds, Lars Johansson, Aljona Saleh, Johanna Larsson, Lars Sandberg, Sara Häggblad Sahlberg, Weixing Qian, Hjalmar Gullberg, Neeraj Garg, Anna-Lena Gustavsson, Martin Haraldsson, David Lane, Ulrika Yngve, and Sonia Lain



(R)-HZ05

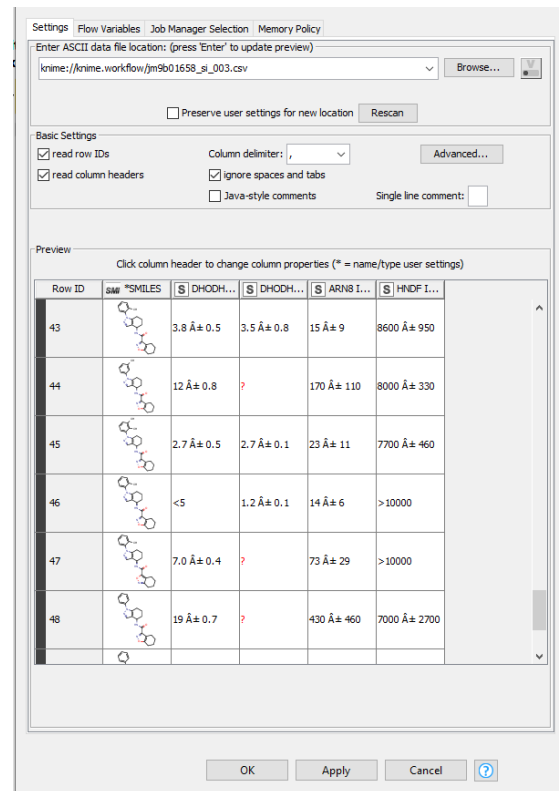
<https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.9b01658>

The nodes – File Reader

- tip: workflow-relative paths to the data

Local: "C:\Users\alice.krebs\knime-workspace\2020_11_20_KNIME_Summit_RDKit_Workshop\Reactions with RDKit\jm9b01658_si_003.csv"

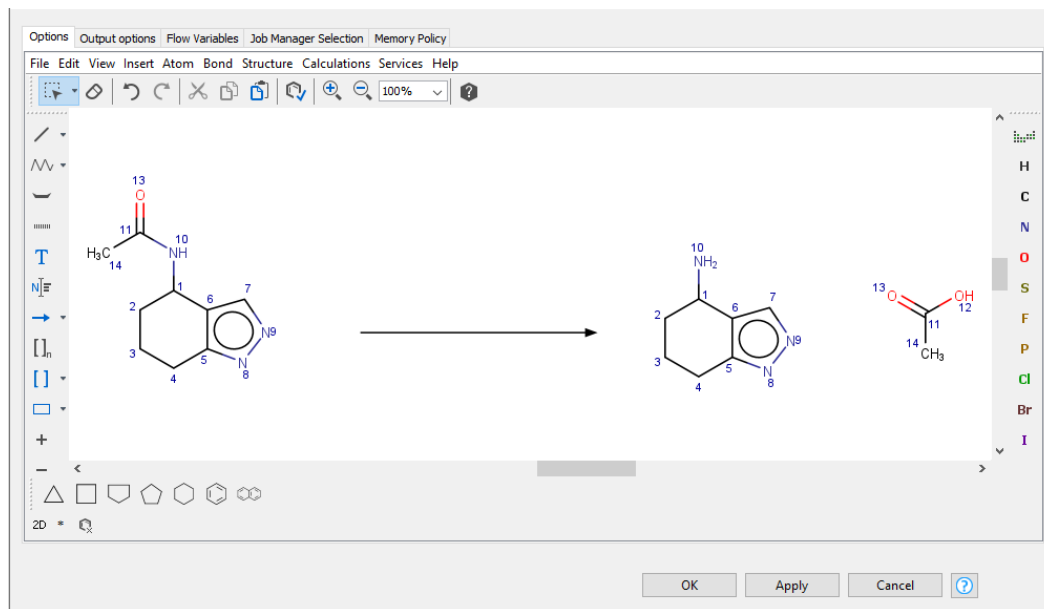
File Reader



<https://www.youtube.com/watch?v=U9sP4g4yGwY&t=124s>

The nodes – Marvin Sketch

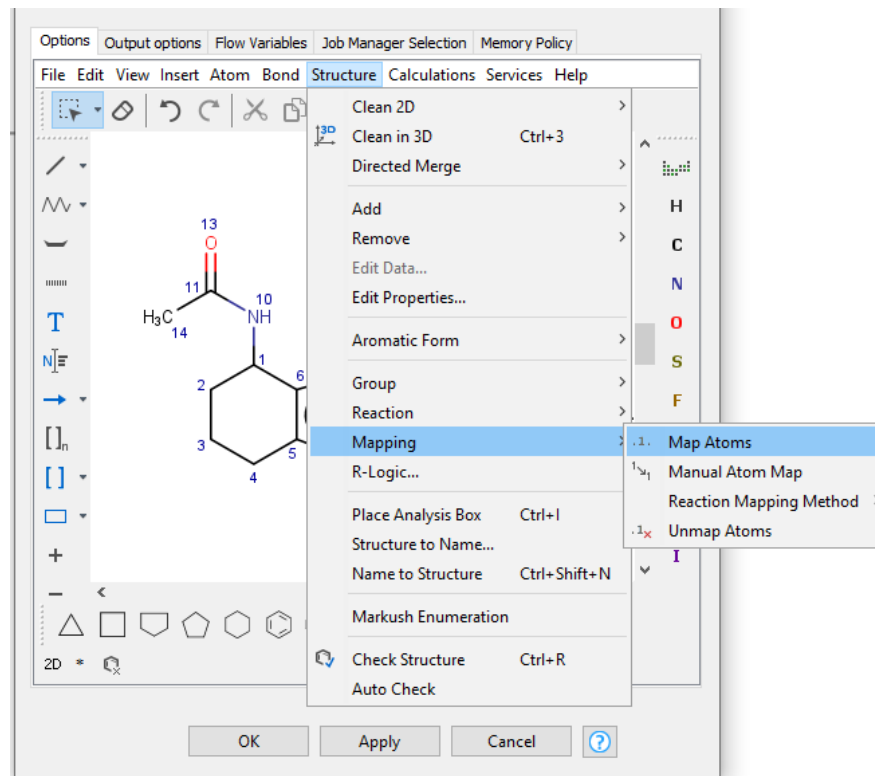
MarvinSketch



The nodes – Marvin Sketch

- Mapping the atoms

MarvinSketch



The nodes – RDKit One Component Reaction

RDKit One Component Reaction



Options | Flow Variables | Job Manager Selection | Memory Policy

Reaction

Reactant RDKit Mol column:

RDKit Rxn column:

Randomization

☐ Randomize reactants

Maximum number of random reactions:

Random seed (or -1 to be ignored):

Other Options

☐ Uniquify products

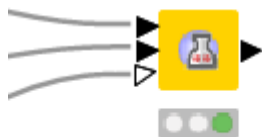
OK Apply Cancel

Table "Output" - Rows: 54 Spec - Columns: 4 Properties Flow Variables

| Row ID | Product | Product Index | Reactant 1 sequence index | Reactant 1 |
|--------|---------|---------------|---------------------------|------------|
| 24_0_0 | | 0 | 24 | |
| 24_0_1 | | 1 | 24 | |
| 25_0_0 | | 0 | 25 | |
| 25_0_1 | | 1 | 25 | |


The nodes – RDKit Two Component Reaction


RDKit Two Component Reaction




Options | Flow Variables | Job Manager Selection | Memory Policy

Reaction

Reactants 1 RDKit Mol column:  Product

Reactants 2 RDKit Mol column:  Product

RDKit Rxn column:  Rxn Molecule

Randomization

☐ Randomize reactants

Maximum number of random reactions:

Random seed (or -1 to be ignored):

Other Options






















☐ Unify products

☒ Do matrix expansion

OK Apply Cancel ?

File Edit Hilitte Navigation View

Table "Output" - Rows: 340 Spec - Columns: 6 Properties Flow Variables

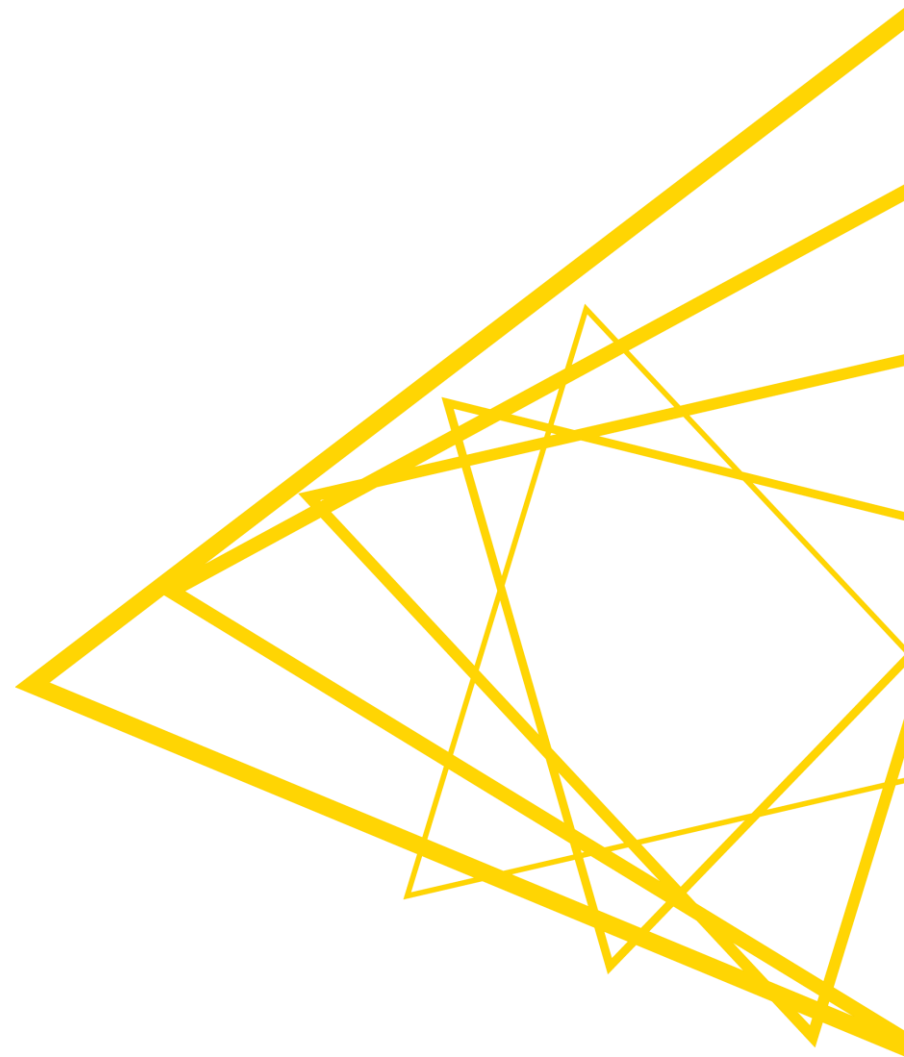
| Row ID |  Product | Product Index | Reactant 1 sequence index |  Reactant 1 | Reactant 2 sequence index |  Reactant 2 |
|---------|---|---------------|---------------------------|--|---------------------------|--|
| 0_0_0_0 |  | 0 | 0 |  | 0 |  |
| 0_1_0_0 |  | 0 | 0 |  | 1 |  |
| 0_2_0_0 |  | 0 | 0 |  | 2 |  |
| 0_3_0_0 |  | 0 | 0 |  | 3 |  |
| 0_4_0_0 |  | 0 | 0 |  | 4 |  |
| 0_5_0_0 |  | 0 | 0 |  | 5 |  |

3D Visualization of Molecules

Greg Landrum

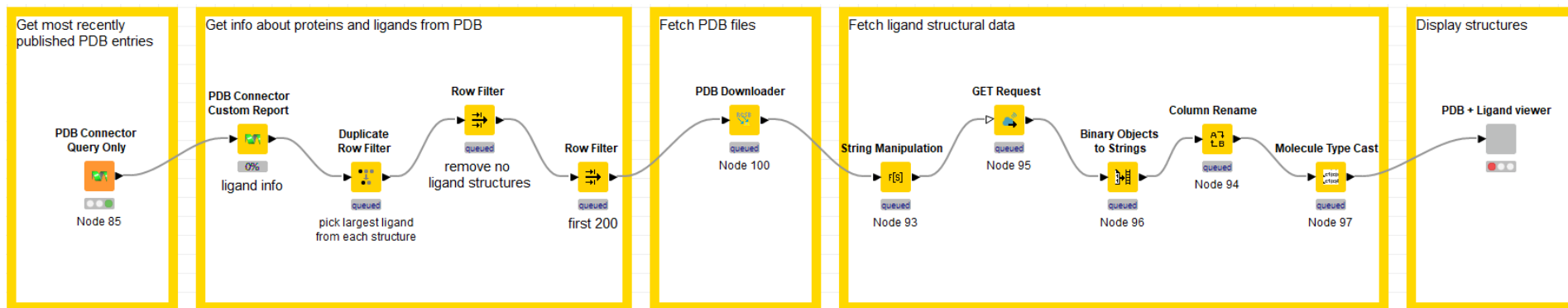
KNIME Fall Summit Webinar

20 November 2020



The use case

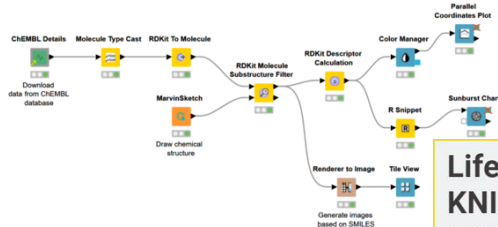
- Build a browser for the most recent protein–ligand crystal structures from the PDB
- The workflow:
 - Download most recent x-ray structures from the PDB that contain a co-crystallized “ligand”
 - Download the 3D structure for the ligand
 - Provide an interactive view allowing the ligands and proteins to be viewed in 3D



Why KNIME for Life Sciences?

<https://www.knime.com/why-knime-for-life-science>

Versatile and Open Analytics for the Life Sciences



Find out more about accessing, transforming, and interacting with large amounts of life science data.

Contact us

Why KNIME for Life Sciences

Work with your specific life science data types easily and in one single environment.

Manage large amounts of data all in one place.

Take advantage of machine learning capabilities in KNIME.

Use life science community nodes such as [RDKit](#), [Vernalis](#), [SeqAn](#), and more.

Draw on expertise from the KNIME Team as well as the community via the [KNIME Forum](#).

Analytical data
Cheminformatics Clinical data
AI/Machine learning QA/QC Metabolomics
Bioinformatics Genomics
Digital pathology **Life Sciences** Text mining
Interactive data exploration High-content screening
Computational chemistry Multi-objective optimization
Data integration High-throughput screening
Competitive intelligence

Life Sciences on the KNIME Blog

Scale and orchestrate the modeling process to train and evaluate 300,000 models or bioactivity prediction.

Explore, analyze, visualize: create interactive views using sunburst charts, tag clouds, and networks based on the example of investigating disease-related genes.

Learn how to use Python code found in Jupyter notebooks in KNIME as well as how to execute KNIME workflows directly from within Python.

Answer questions from the area of pharmaceutical research by linking and querying different datasets stored in BigQuery.

Visit KNIME Blog



KNIME in Action

Examples of KNIME in Action from our community of Life Science users:

Deep Learning: From Mastering the Game of Go to Revolutionizing Microscopy - by Florian Jug (deNBI). [Open slides](#)

Building a Clinically Significant Rare Disease Data Master: Approach and Workflows - by Sebastian Lefebvre (Alexion Pharmaceuticals). [Open slides](#)

Using KNIME to Build a Data-Driven Culture (and Workflows) in a Biopharma Setting - by Kenneth Longo (WAVE Life Sciences). [Open slides](#)

Additional Resources

KNIME pages (<https://www.knime.com>)

- RESOURCES **LEARNING HUB** <https://www.knime.com/learning-hub>
- RESOURCES **HUB** <https://hub.knime.com/>
- BOOK **WILL THEY BLEND** <https://www.knime.com/knimepress/will-they-blend>

KNIME Tech pages

FORUM for questions and answers <https://forum.knime.com>

- **DOCUMENTATION** for docs, FAQ, changelogs, ... <https://docs.knime.com/>
- **COMMUNITY CONTRIBUTIONS** for dev instructions and third party nodes
<https://www.knime.com/community>

KNIME TV on **YouTube** <https://www.youtube.com/user/KNIMETV>

KNIME courses



[Home](#) > [Learning](#) > [Events and Courses](#)

KNIME Events

Explore our events all over the world. Find upcoming events near you or browse past events!



Summit

KNIME Fall Summit 2020

November 16 - 20, 2020 - Online



Online Course

[L4-BD] Introduction to Big Data - Online

November 16 - 17, 2020 - Online



Certification

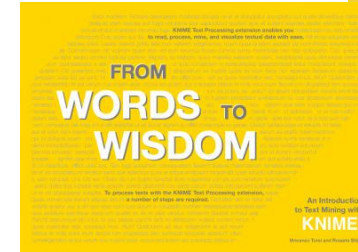
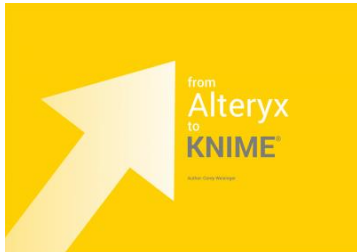
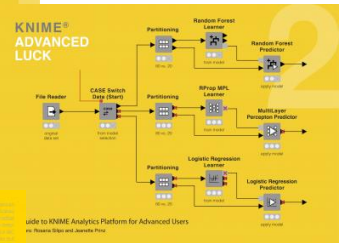
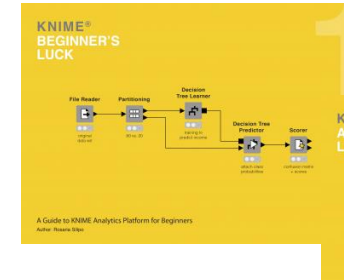
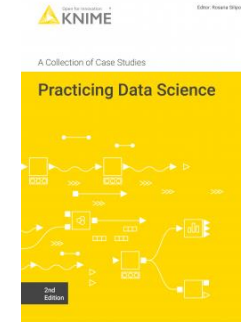
KNIME Certification - Online

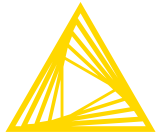
November 18 - 19, 2020 - Online

<https://www.knime.com/learning/events>

KNIME Books

- Course books downloadable from **KNIME Press**
- <https://www.knime.com/knimepress>
- Code: **FALL-SUMMIT-WORKSHOP**
Valid for: All Books
Expires: Jan 31, 2021

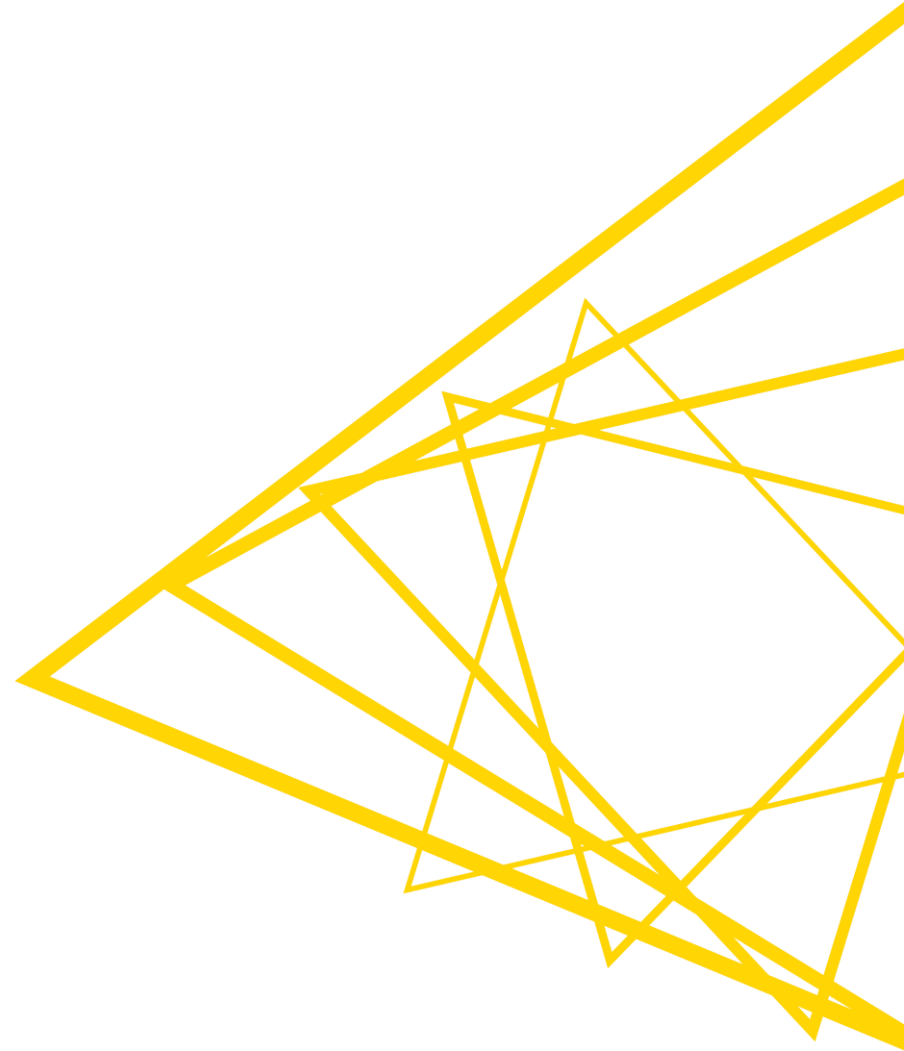




Open for Innovation

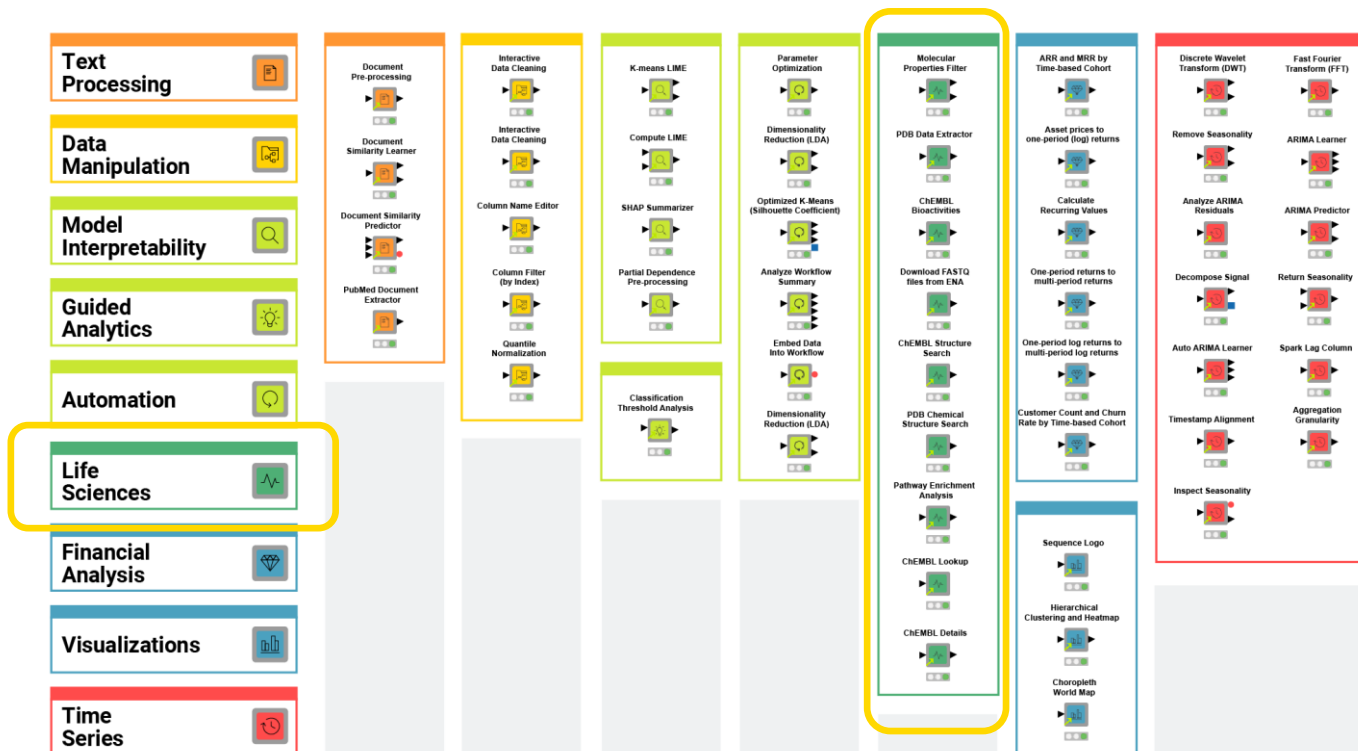
KNIME

Thank you for joining!

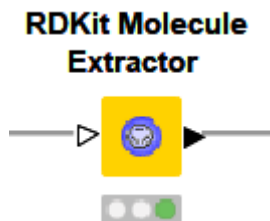


KNIME Verified Components

- [Verified Components](#) reuse bundled functionalities, verified by KNIME experts
- Released and updated on the [KNIME Hub](#)



The nodes – RDKit Molecule Extractor



Options Advanced Flow Variables Job Manager Selection Memory Policy

Table Input

RDKit Mol column: Product

Reference column (e.g. an ID): ? <RowID>

Output

Column name for extracted molecules: Molecules

Column name for copied reference data: Reference

OK Apply Cancel

Options Advanced Flow Variables Job Manager Selection Memory Policy

☐ Sanitize fragments

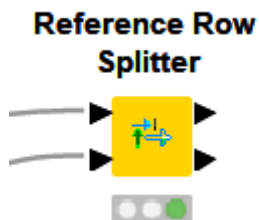
How to react on conversion errors: Create a missing cell, generate warning

How to react on empty (missing) cells: Create a missing cell, no warning

How to react on empty (zero atom) molecules: Skip result, no warning

OK Apply Cancel ?

The nodes – Reference Row Splitter



Options | Flow Variables | Job Manager Selection | Memory Policy

Reference columns

Data table column:

Reference table column:

OK Apply Cancel ?