

250611

KNIME in Chemistry for Beginners

xFOREST Therapeutics Co., Ltd. (Kyoto, Japan)

Hiroki YOSHIDA

Online meeting

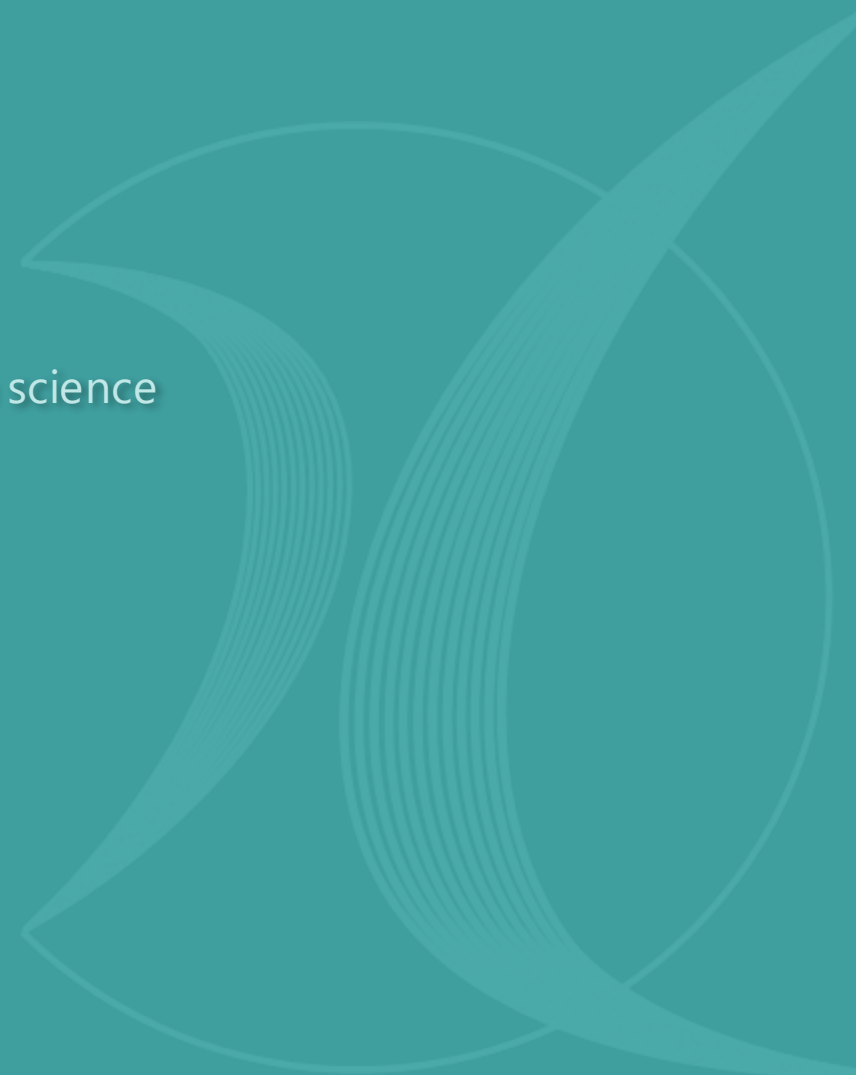


Agenda

- 00 Current situation
- 01 **Why** KNIME for chemistry?
- 02 **What** can KNIME do? Case studies
- 03 **How** to learn KNIME easily?
- 99 Appendix

00

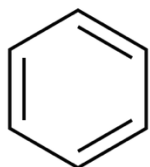
Current situation: Chemistry – data science



Importance of AI Knowledge in Research



- AI technologies (such as ChatGPT) are rapidly gaining attention across various fields
- Knowledge and skills in IT and AI are becoming essential beyond traditional disciplines



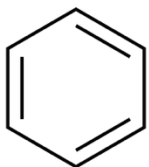
- In drug discovery as well as other fields, the integration of IT skills is increasingly necessary
- AI-driven drug discovery using AlphaFold, MD simulation, predictive ADMET and many IT tools have been focused recently



- Even wet-lab researchers are expected to have a fundamental understanding of programming and related technologies

KNIME as a starting point for chemo informatics

Traditional way



Compound as input



```

from rdkit import Chem
from rdkit.Chem import SaltRemover, rdmolops, Draw

def preprocess_smiles(smiles):
    mol = Chem.MolFromSmiles(smiles)
    if mol is None:
        return None

    remover = SaltRemover.SaltRemover()
    mol = remover.StripMol(mol, dontRemoveEverything=True)

    mol = Chem.AddHs(mol)
    Chem.SanitizeMol(mol)

    rdmolops.Kekulize(mol, clearAromaticFlags=True)

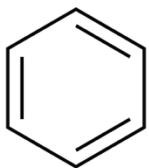
    return mol
    
```

Coding based on many words by typing



- Python is very useful for cheminformatics
- **However**, it seems tough for a person from wet lab to code from the beginning

KNIME



Compound as input



#	RowID	Salt Stripped Molecule
3	Row2	
4	Row3	

Workflow always with structure (by drag & drop)



- KNIME always shows structural information, making a chemist feel good
- Visualizing makes it easy to understand what you're doing

01

Why KNIME for chemistry?



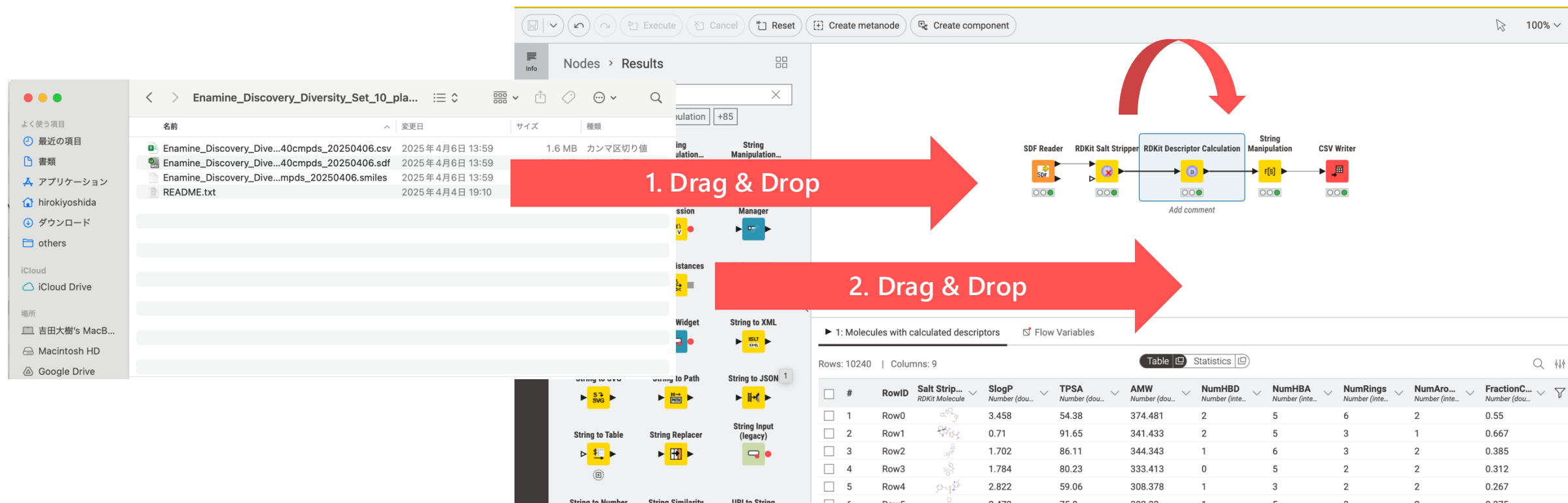
Why KNIME for chemistry?

1. **Quick** to build up programming
2. **Interactive** to visualize chemical structures
3. **Easy** to develop useful application

Why KNIME for chemistry?

Generating chemical descriptors just in 3 steps (QUICK)

It needs 1 min to make chemical descriptors



Protocol

1. Drag & Drop SDF file (input)
2. Drag & Drop Nodes and connect them
3. Run, and export as csv file

Why KNIME for chemistry?

Generating chemical descriptors just in 3 steps (QUICK)

It needs 1 min to make chemical descriptors

Node →

1: Molecules with calculated descriptors

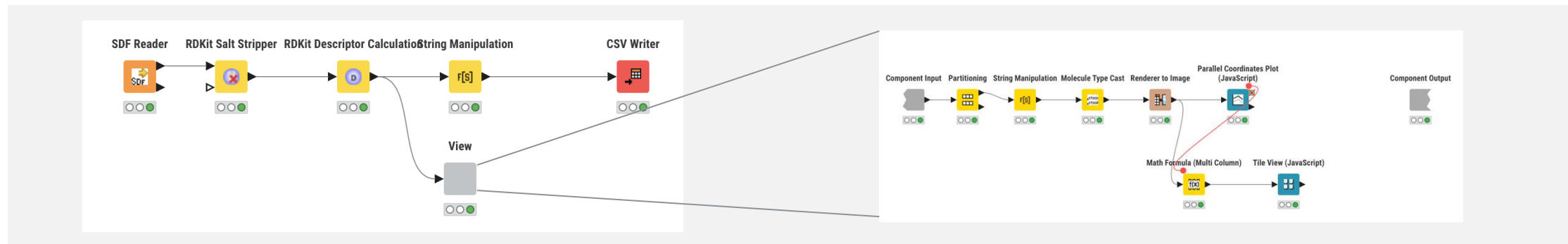
Rows: 10240 | Columns: 9

#	RowID	Salt Strip... RDKit Molecule	SlogP Number (dou...)	TPSA Number (dou...)	AMW Number (dou...)	NumHBD Number (inte...)	NumHBA Number (inte...)	NumRings Number (inte...)	NumAro... Number (inte...)	FractionC... Number (dou...)
1	Row0		3.458	54.38	374.481	2	5	6	2	0.55
2	Row1		0.71	91.65	341.433	2	5	3	1	0.667
3	Row2		1.702	86.11	344.343	1	6	3	2	0.385
4	Row3		1.784	80.23	333.413	0	5	2	2	0.312
5	Row4		2.822	59.06	308.378	1	3	2	2	0.267
6	Row5		2.472	75.8	302.33	1	5	3	2	0.375

All data are easily viewed in this table

Developing **INTERACTIVE** visualization to identify compounds with certain properties

It needs a few minutes to develop a function to search compounds with a certain properties

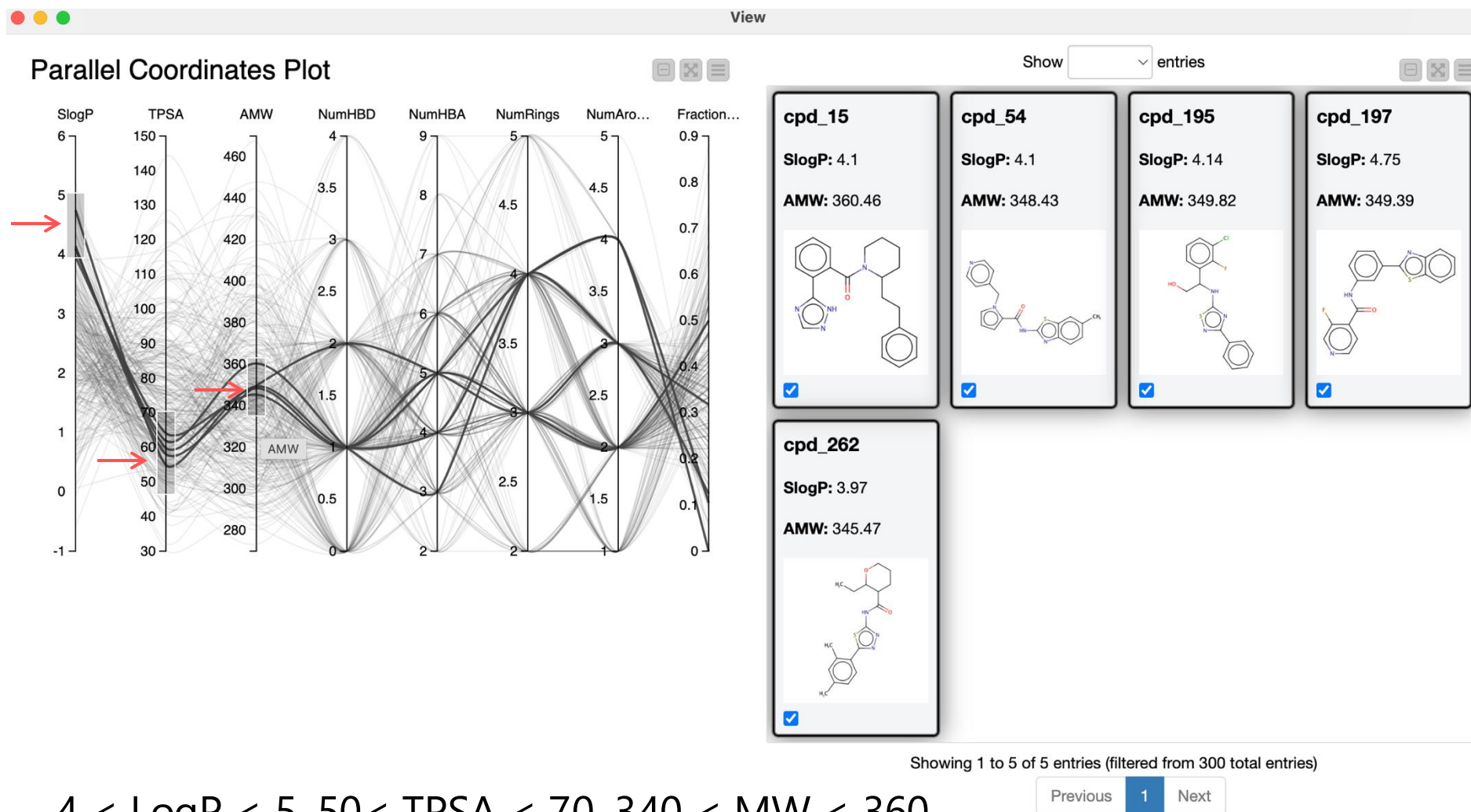


Let's say you want to search compounds having properties below:

$$\begin{aligned} 4 < \text{LogP} < 5 \\ 50 < \text{TPSA} < 70 \\ 340 < \text{MW} < 360 \end{aligned}$$

Developing **INTERACTIVE** visualization to identify compounds with certain properties

It needs a few minutes to develop a function to search compounds with a certain properties

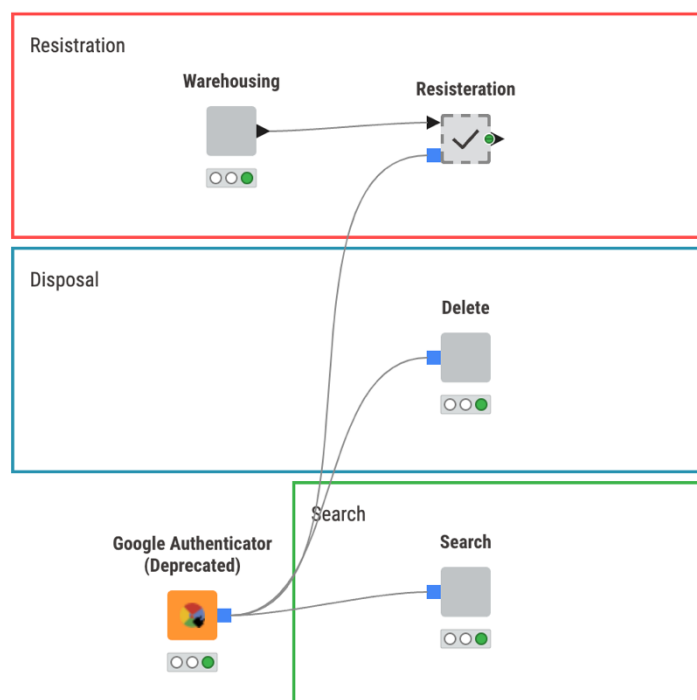


Why KNIME for chemistry?

EASILY developing storage application for reagents

Substructure search by drawing structure is possible

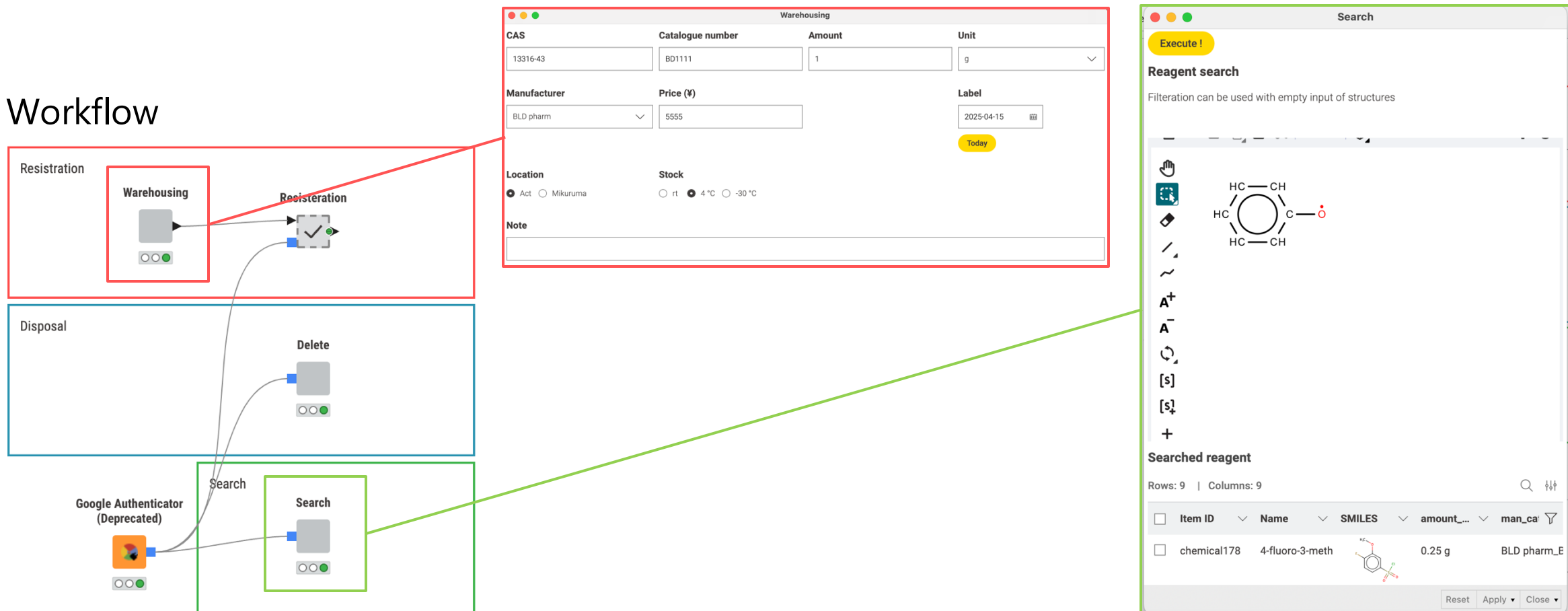
Workflow



EASILY developing storage application for reagents

Substructure search by drawing structure is possible

Workflow

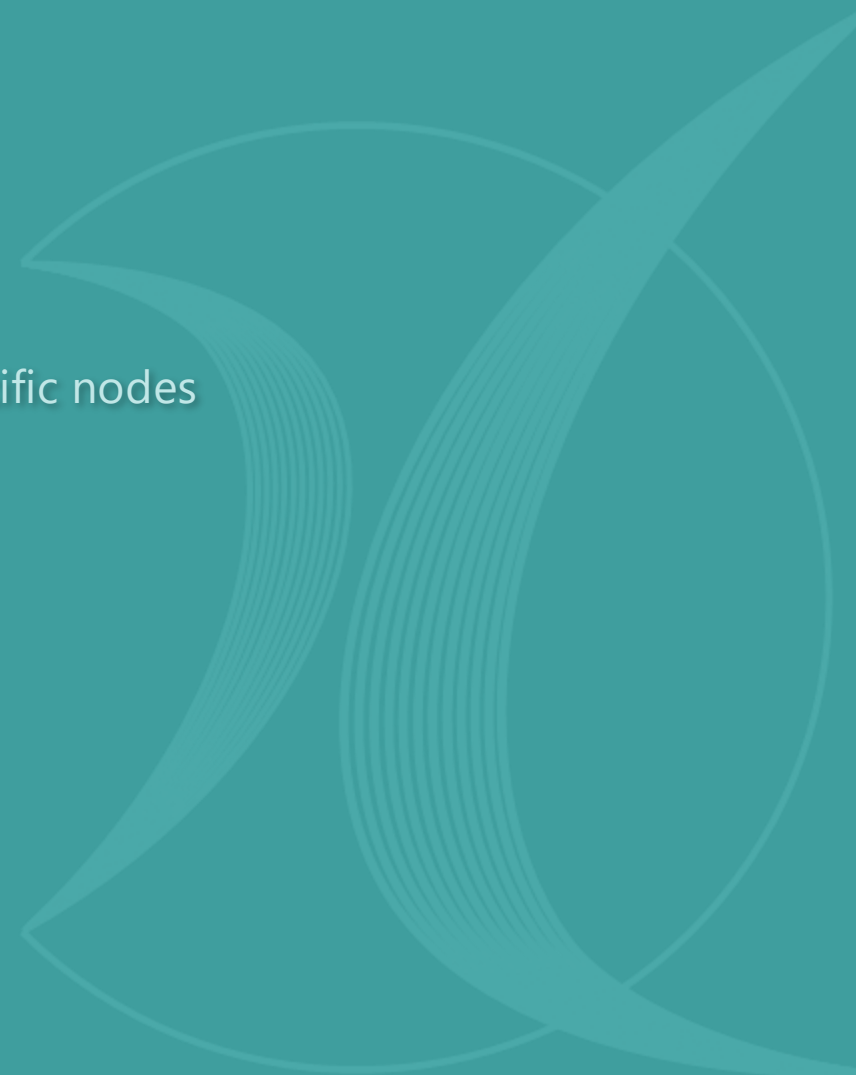


(Deprecated)

- ✓ Efficient reagent storage management is valuable in a university laboratory, especially since we don't need to purchase an external system.
- ✓ Substructure search from stock must be very useful in traditional organic chemistry lab

02

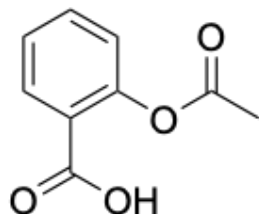
Case studies with chemical specific nodes



Case study

1. **RDKit** in KNIME
2. **Vernalis** in KNIME
3. **Other application** with KNIME

Chemical structure



SMILES

CC(=O)Oc1ccccc1C(=O)O

InChIKey

BSYNRYMUTXBXSQ-UHFFFAOYSA-N

SMARTS

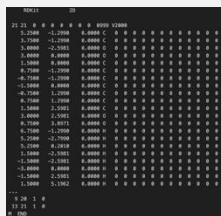
$$[\#6]-[\#6](=O)-[\#8]-[\#6]-1=[\#6]-[\#6]=[\#6]-[\#6]=[\#6]-1-[\#6](-[\#8])=O$$

Fingerprint

100000000010000000000000000000101101010100000010110.....

2D or 3D structure

Mol file (SDF file)

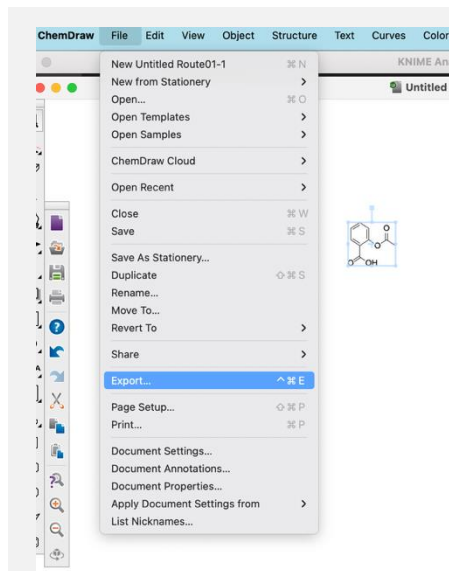


- ✓ Each property has distinct characteristics tailored for specific applications.
- ✓ SMILES and SDF are good starting points for beginner
- ✓ SMILES and SDF file can be easily made by ChemDraw

1st step to generate chemical descriptor from ChemDraw data

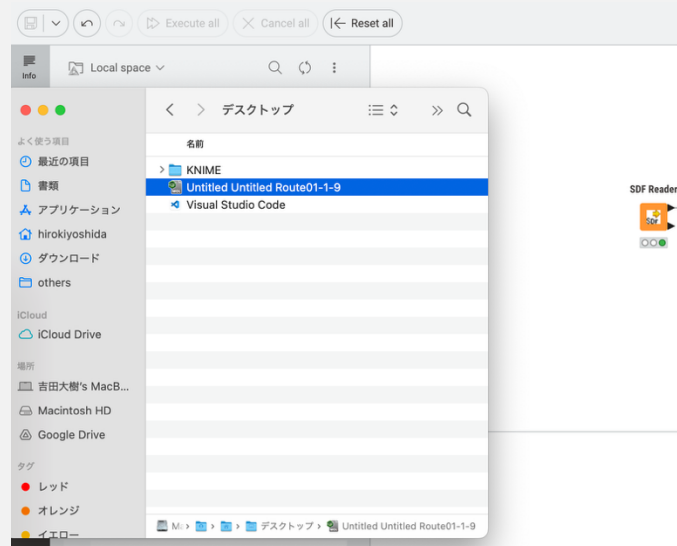
The simplest way to make chemical descriptors

1. Chem draw



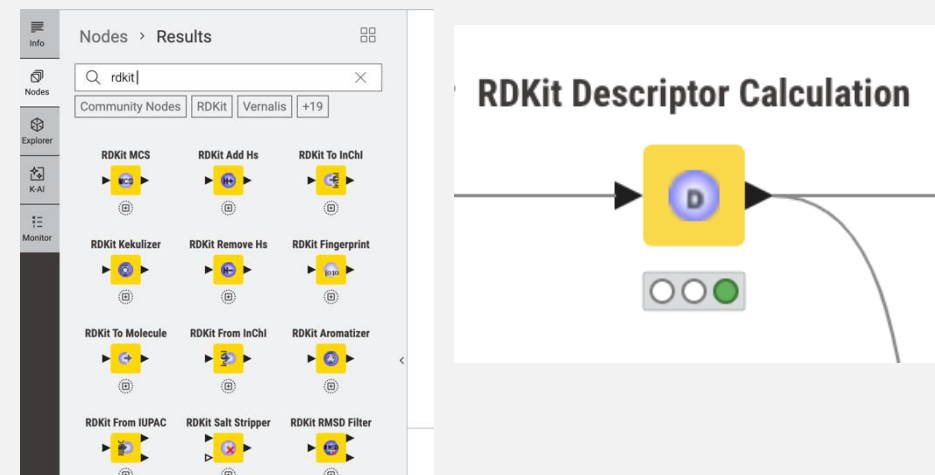
Draw cpd & export as SDF file

2. KNIME-data import



Drag the SDF file to KNIME

3. KNIME-descriptor generation

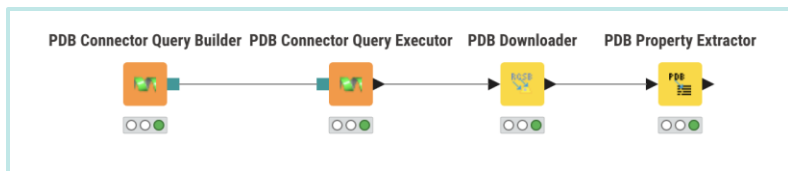


Drag "RDKit Descriptor Calculation" from Node section, connect this to SDF reader double-click to make some setting and then run

- After this 1st trial , please start to learn "desalting", "aromatizing", "Kekulizing", "deionizing" and related-normalizing method to analyze compounds as data (not explained in today's talk)
- Please download any compound library from any company and try to analyze compounds

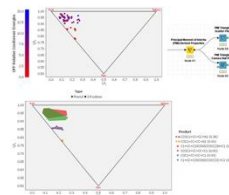
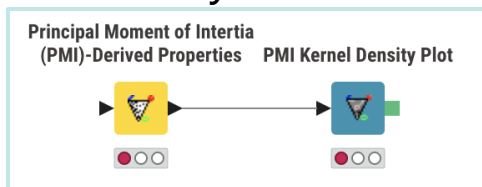
Wide range of functions in KNIME

1. PDB analysis



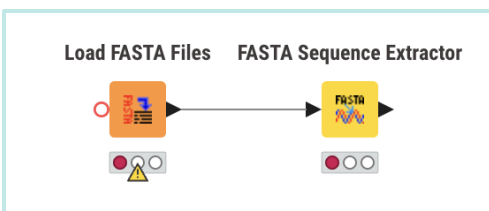
- ✓ Directly connect PDB and extract data
- ✓ 3D conformation can be also viewed

2. PMI analysis



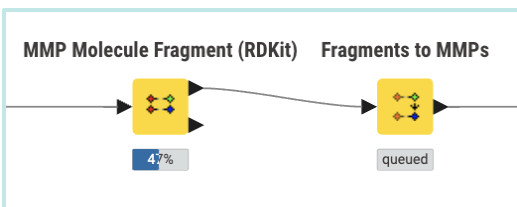
- ✓ Easily visualize chemical properties based on Rod-Sphere-Disc

3. FASTA analysis



- ✓ DNA-RNA information could be dealt in KNIME

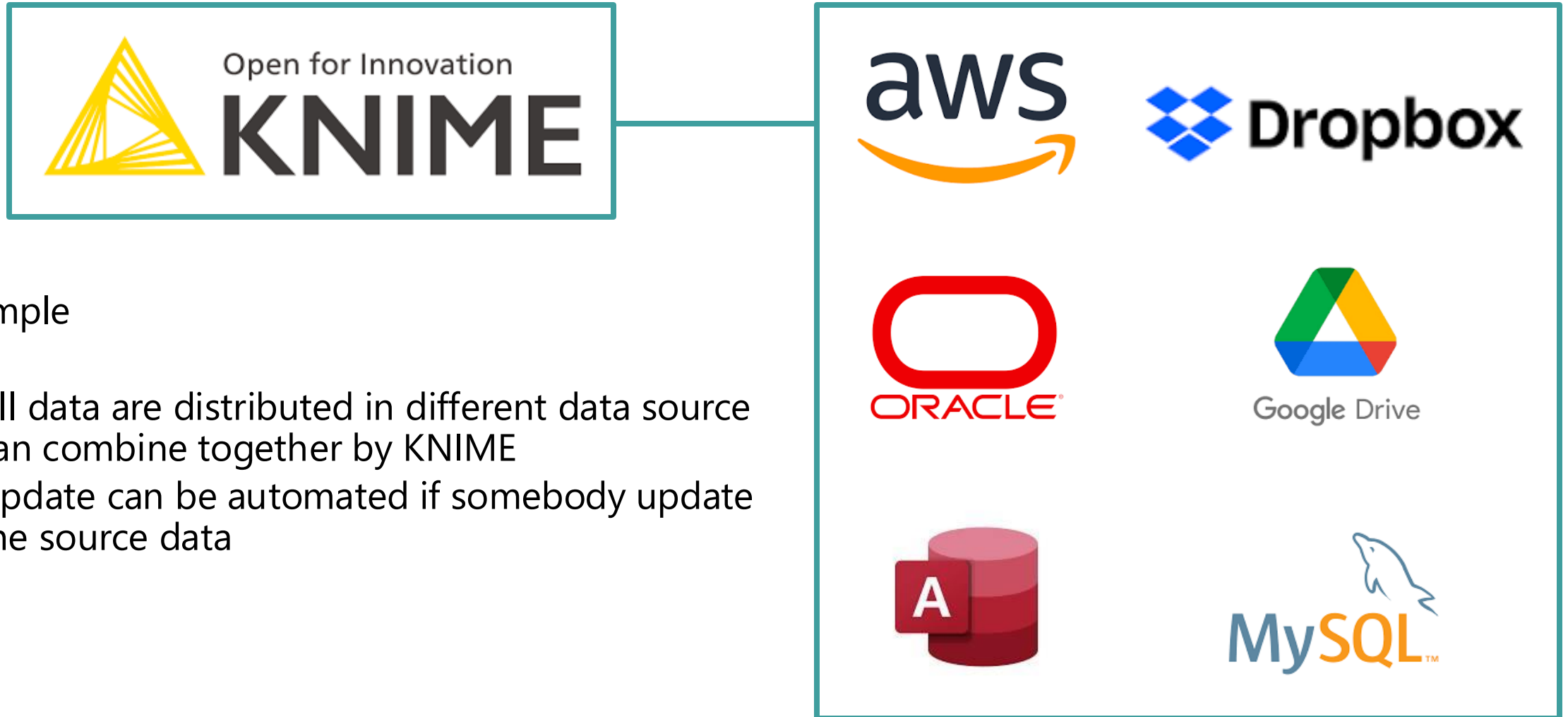
4. Matched Molecule Pair analysis



- ✓ MMP analysis can be done in KNIME
- ✓ Any modification is possible

<https://www.knime.com/blog/matched-molecular-pair-analysis-with-knime>

Making automation flow to combine all data from different data source



Example

- ✓ All data are distributed in different data source can combine together by KNIME
- ✓ Update can be automated if somebody update the source data


03

KNIME community hub



A lot of knowledges are shared freely in KNIME community hub/forum

Ready to use workflow also exist. It is easy to copy someone's workflow at single step



Hub

Search workflows, nodes and more...

PricingAboutSign in

KNIME Community Hub > knime > Collections > KNIME for Chemistry

Collection

KNIME for Chemistry

ChemistryLife SciencesComputational ChemistryCheminformaticsMolecules+2

This curated set of workflow snippets is specifically designed to provide cheminformatics users with essential building blocks for developing their own customized workflows within the KNIME Analytics Platform. Rather than offering complete workflows for entire use cases, this collection focuses on individual components that you can combine. Explore the collection, experiment with the snippets, and discover how KNIME can enhance your cheminformatics projects.


KNIME in Life Science

Explore how KNIME empowers cheminformatics with flexible tools and workflows tailored for molecular data processing, analysis, and visualization.

[Learn more >](#)

Jan 14, 2025

8Share



Data Access

This section provides tools for importing and managing chemical data from multiple sources. It includes workflows for fetching molecular structures from PDB with 3D visualization, retrieving SMILES strings from PubChem CIDs, and handling data through JSON, XML, or GraphQL queries. Additionally, it supports reading data from Excel files, SDF files, and mixed folders, ensuring smooth integration of diverse inputs into your workflows.

Data Access

- Processing
- Analysis
- Visualization
- Output

<https://hub.knime.com/knime/collections/KNIME%20for%20Chemistry~UEFCrTRANXd09rbz>

