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?

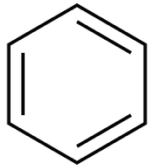
99 Appendix

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

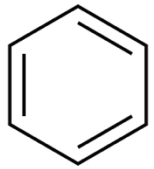


- In drug discovery as well as other fields, the integration of IT skills is increasingly necessary
- AI-driven drug discovery using AlphaFold, MD simulation, predictive ADMET and many IT tools have been focused recently

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

KNIME as a starting point for chemo informatics

Traditional way



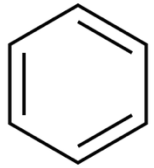
Compound as input

```python
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

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?

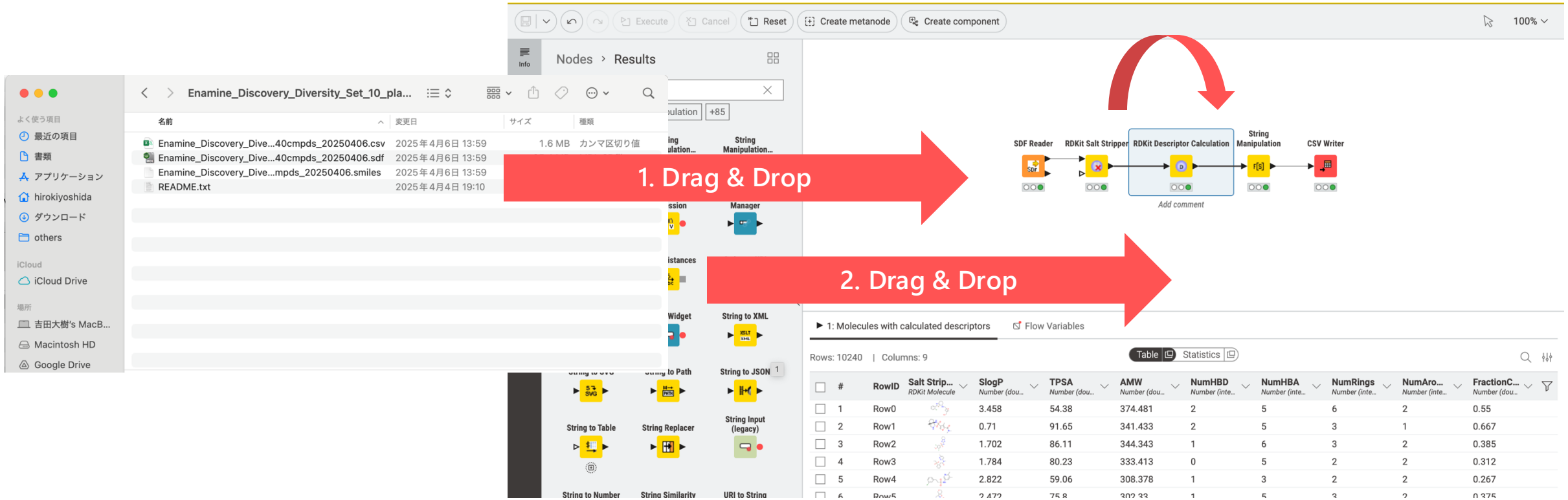
Why KNIME for chemistry?

1. Quick to build up programing
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



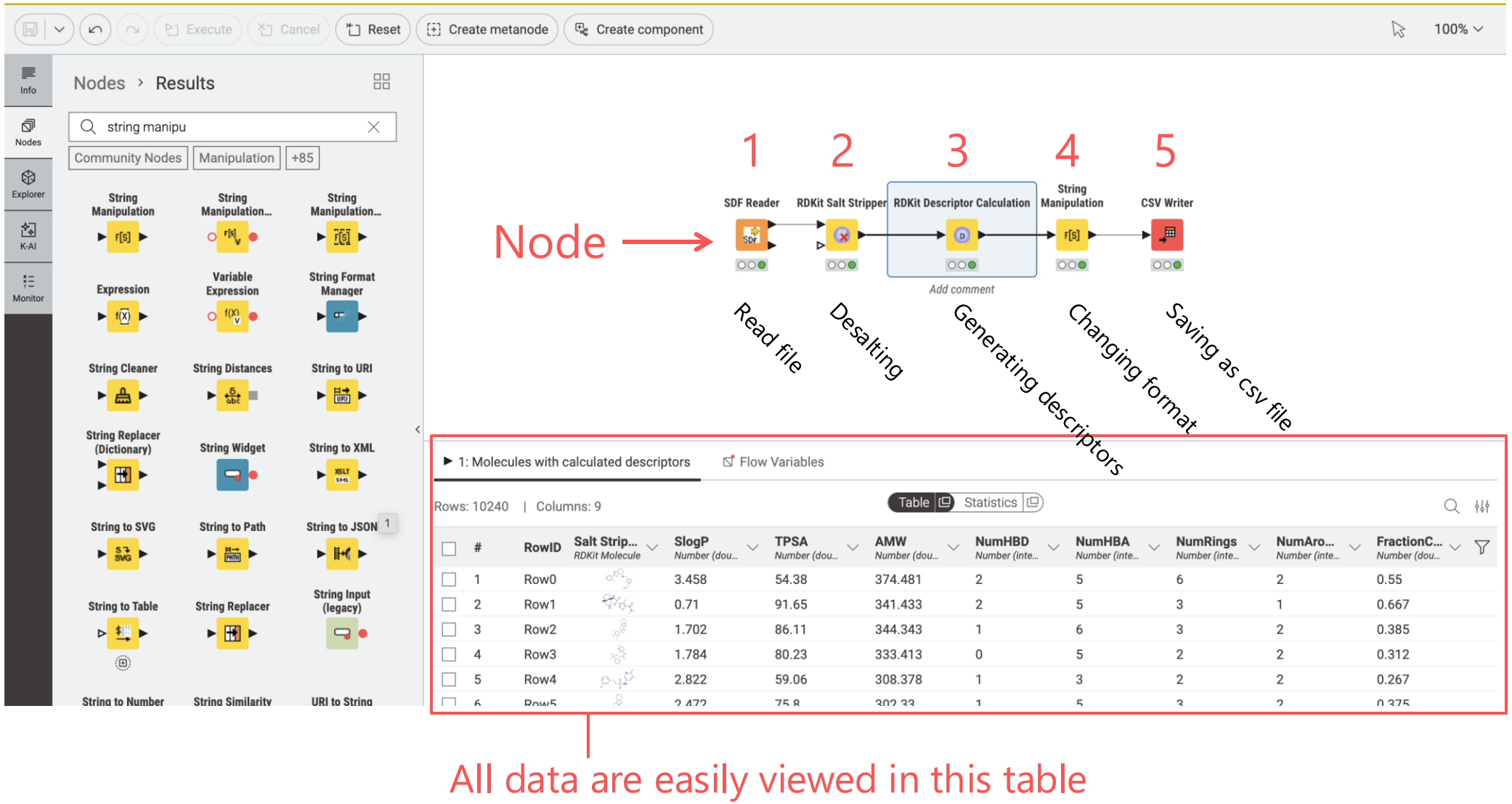
Protocol

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)

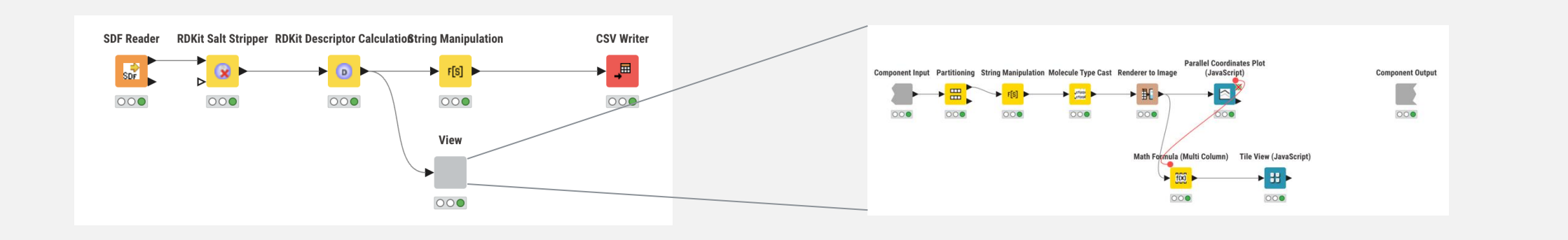
It needs 1 min to make chemical descriptors



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

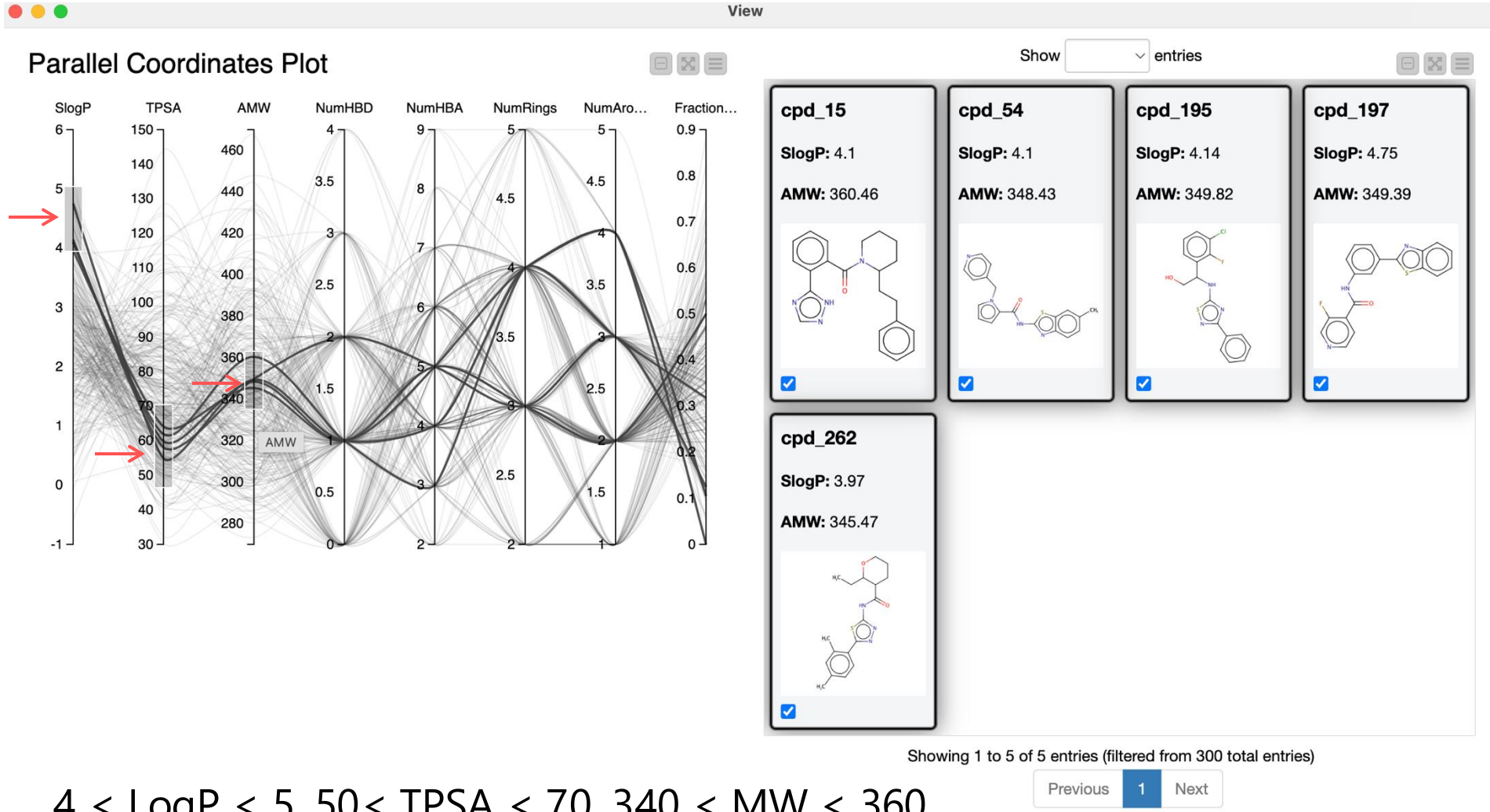


Let's say you want to search compounds having properties below:

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

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

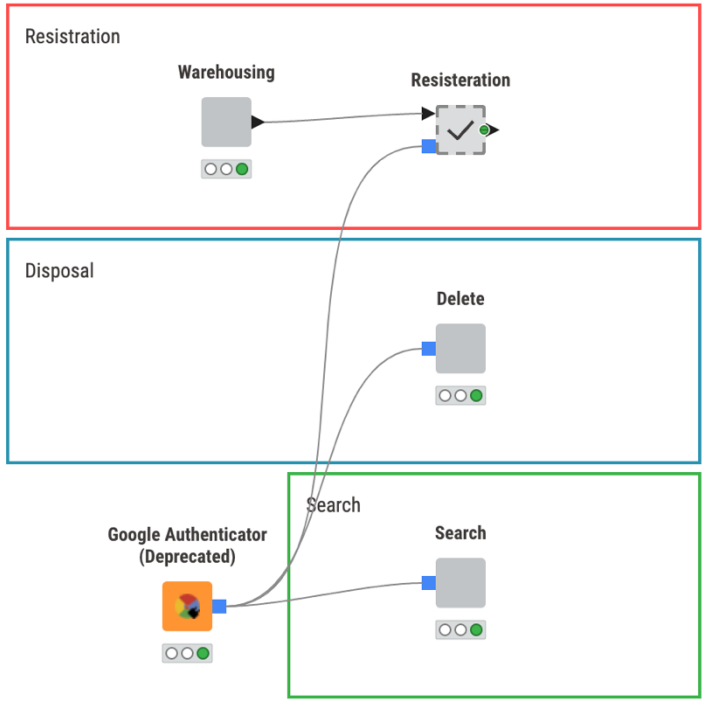


4 < LogP < 5, 50< TPSA < 70, 340 < MW < 360

EASILY developing storage application for reagents

Substructure search by drawing structure is possible

Workflow

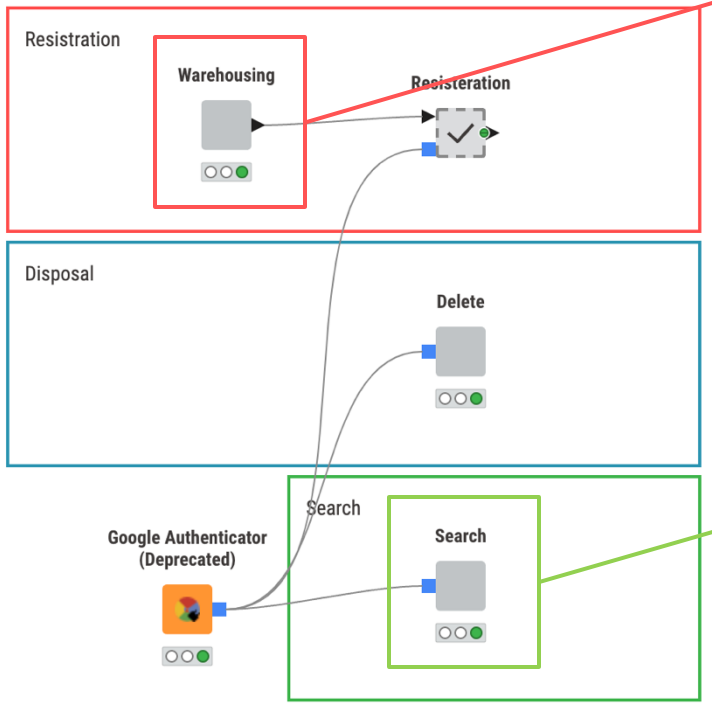


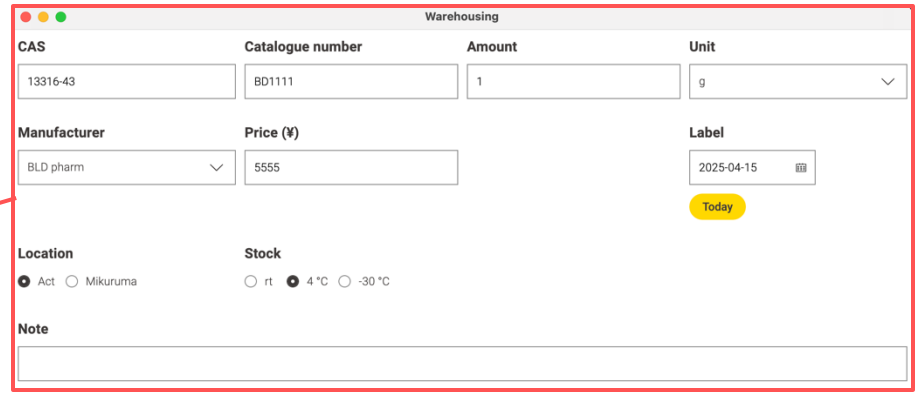
Why KNIME for chemistry?

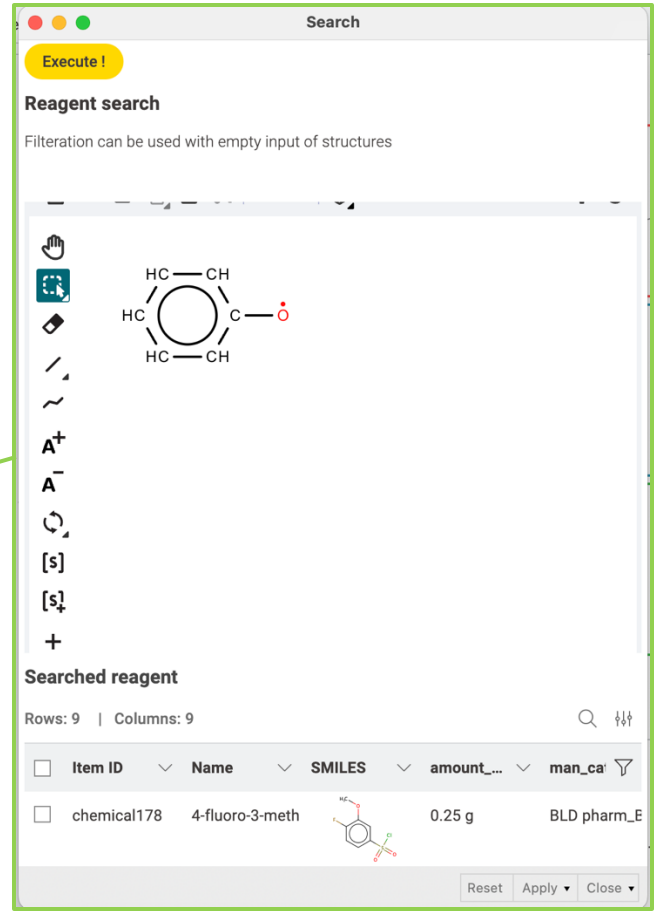
EASILY developing storage application for reagents

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

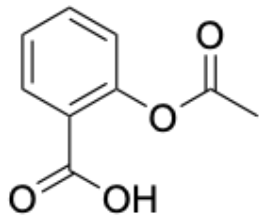
Case study

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

Converting structure into machine-readable formats

Many types of data are used for compounds

Chemical structure



2D structure

SMILES

CC(=O)Oc1ccccc1C(=O)O

InChIKey

BSYNRYMUTXBXSQ-UHFFFAOYSA-N

SMARTS

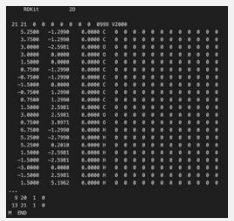
[#6]-[#6](=O)-[#8]-[#6]-1=[#6]-[#6]=[#6]-[#6]=[#6]-1-[#6](-[#8])=O

Fingerprint

10000000001000000000000000000101101010100000010110……..

2D or 3D structure

Mol file (SDF file)



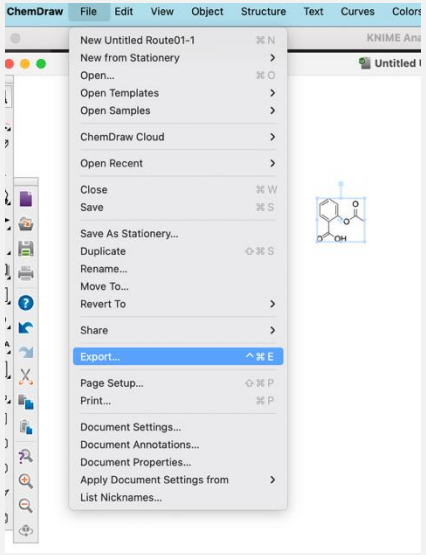
- ✓ Each property has distinct characteristics tailored for specific applications.
- ✓ SMILES and SDF are good starting points for beginner
- ✓ SMILES and SDF file can be easily made by ChemDraw

Chemical descriptor (RDKit)

1st step to generate chemical descriptor from ChemDraw data

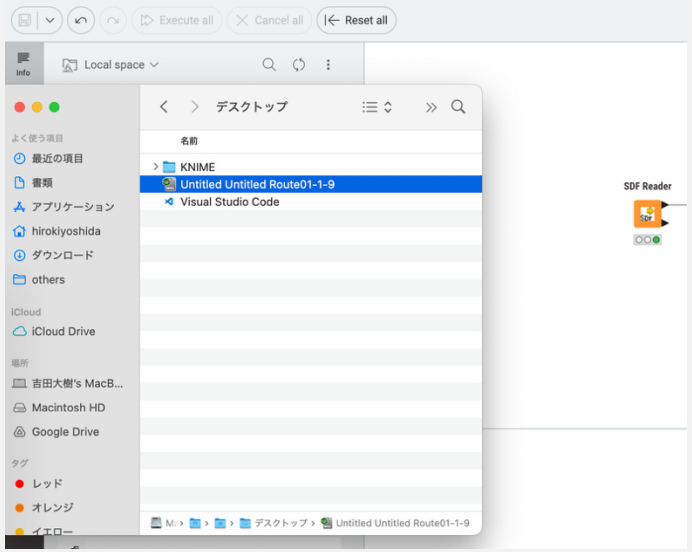
The simplest way to make chemical descriptors

1. Chem draw



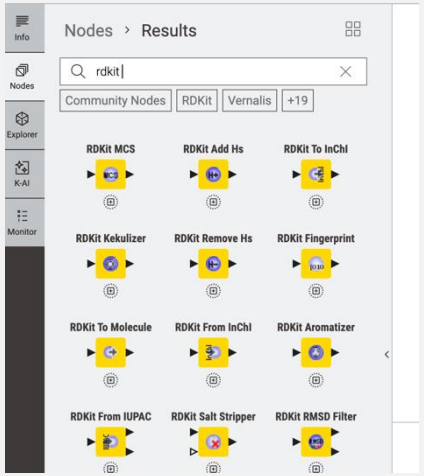
Draw cpd & export as SDF file

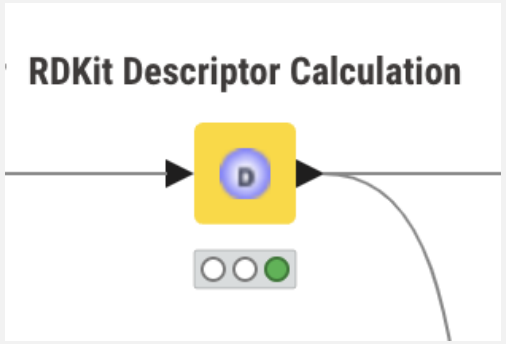
2. KNIME-data import



Drag the SDF file to KNIME

3. KNIME-descriptor generation



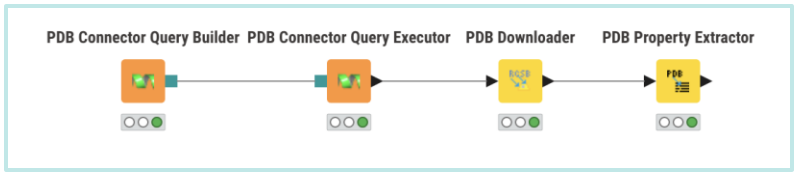


Drag “RDKit Descriptor Calculation” from Node section, connect this to SDF reader double-click to make some setting and then run

- 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

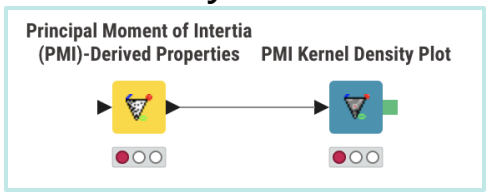
Wide range of functions in KNIME

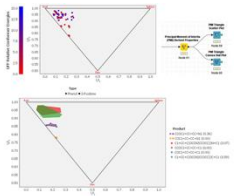
1. PDB analysis



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

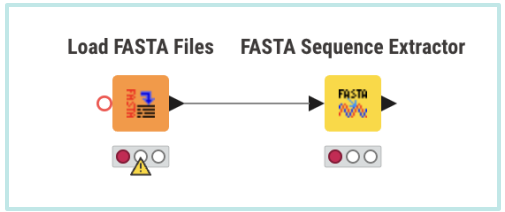
2. PMI analysis





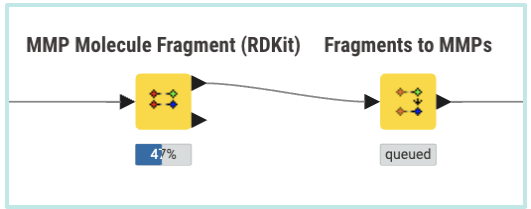
- ✓ Easily visualize chemical properties based on Rod-Sphere-Disc

3. FASTA analysis



- ✓ DNA-RNA information could be dealt in KNIME

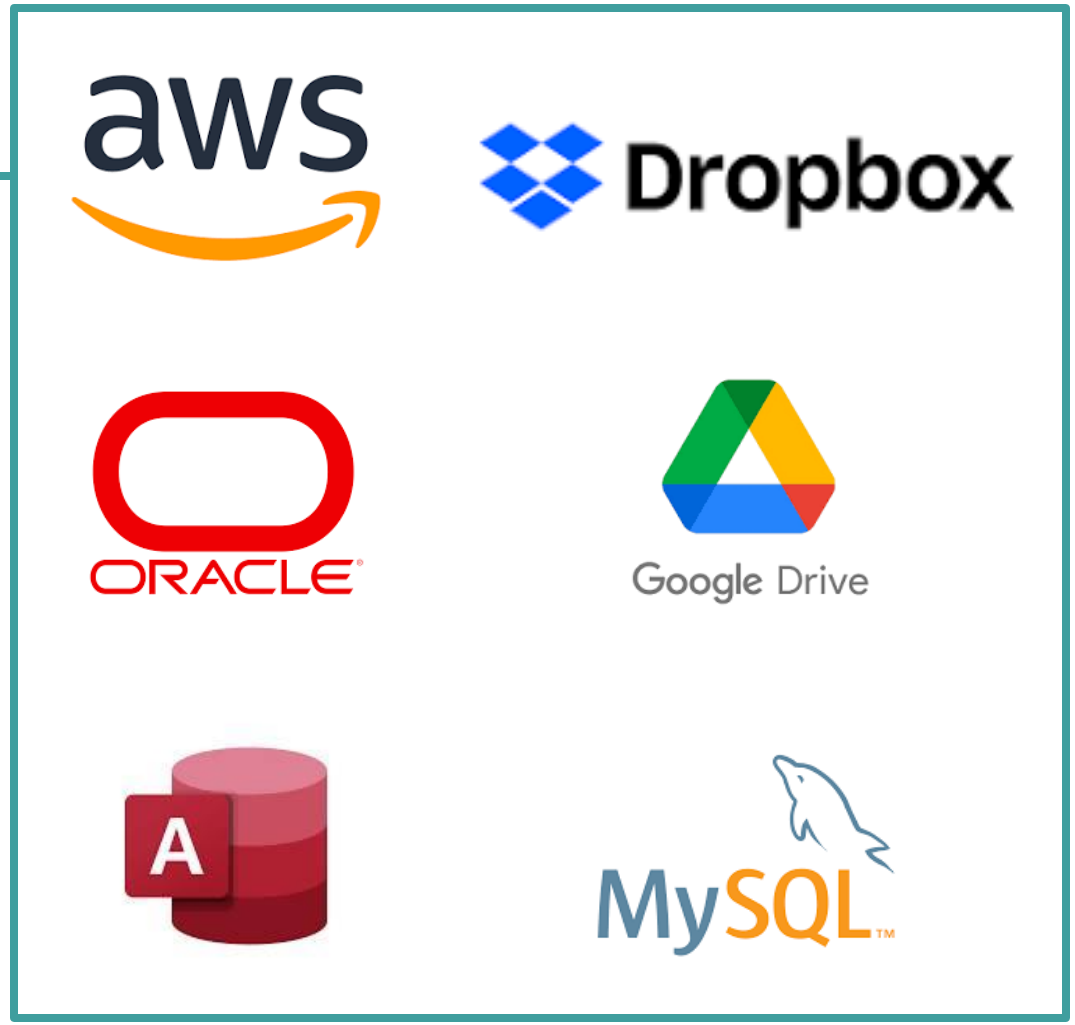
4. Matched Molecule Pair analysis



- ✓ 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

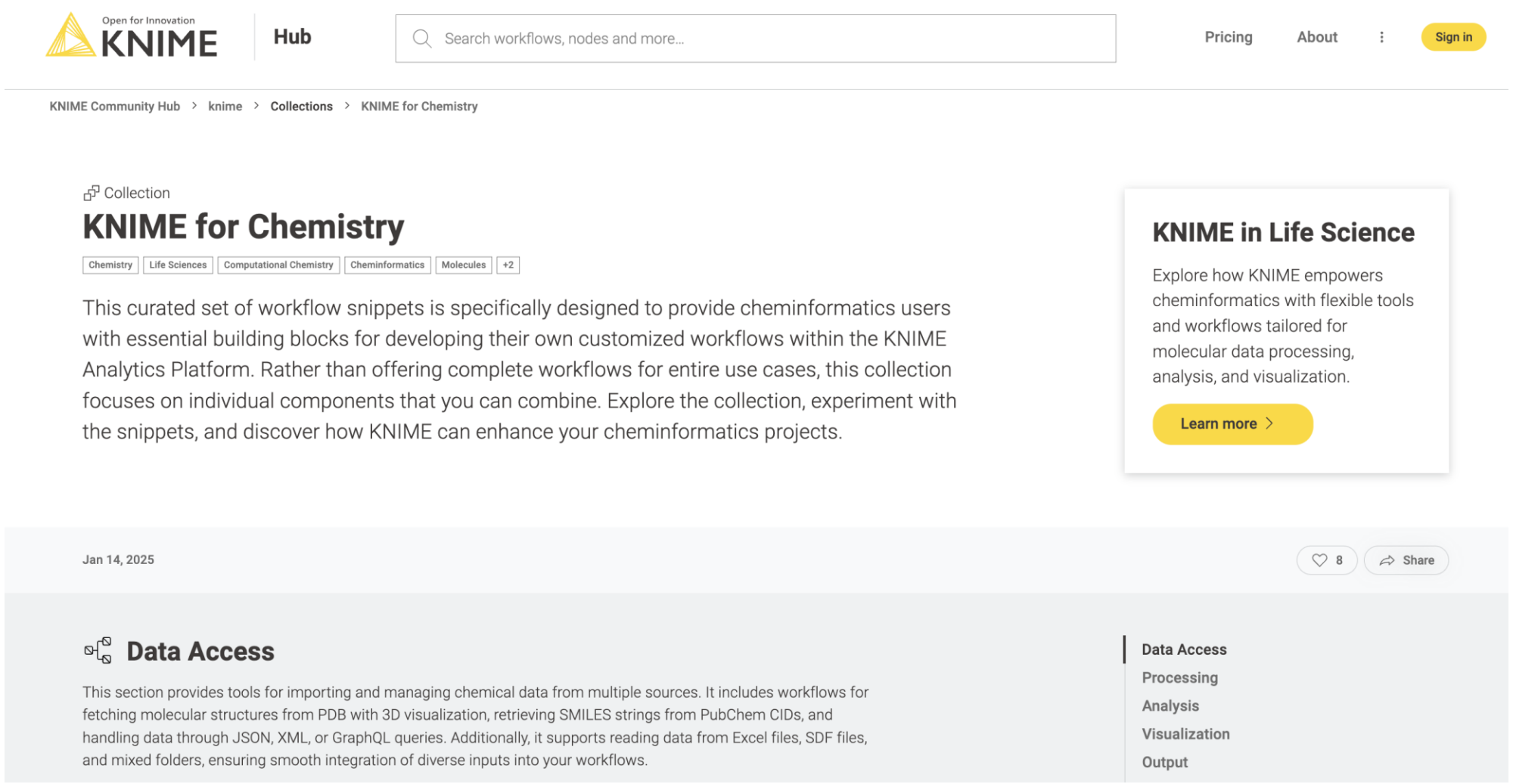
03

KNIME community hub

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



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