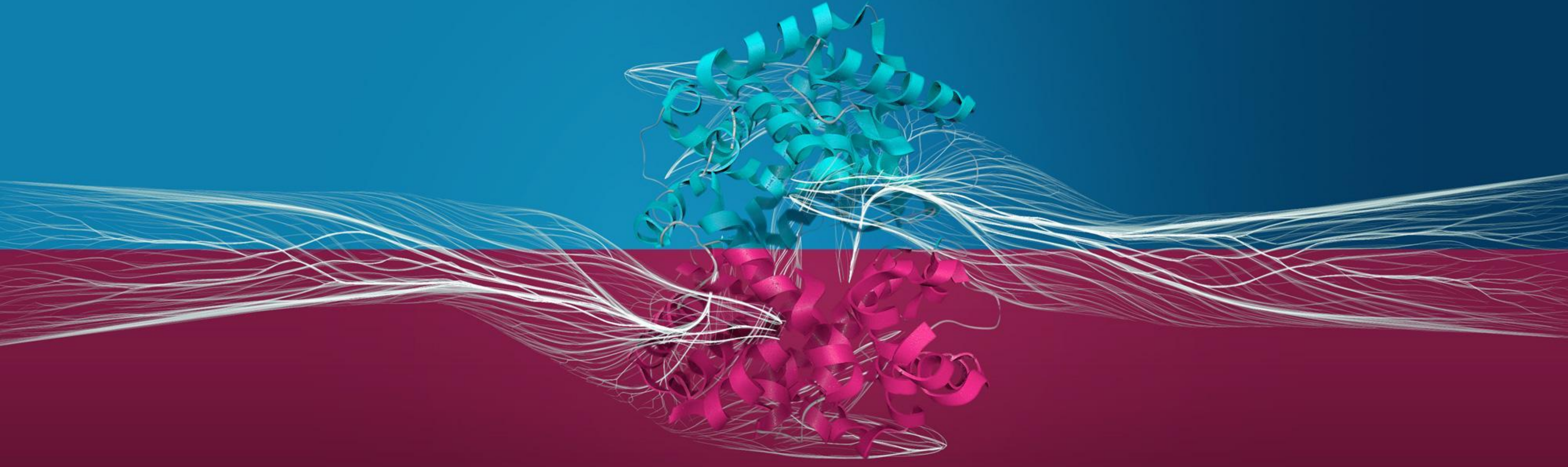


congruence TX

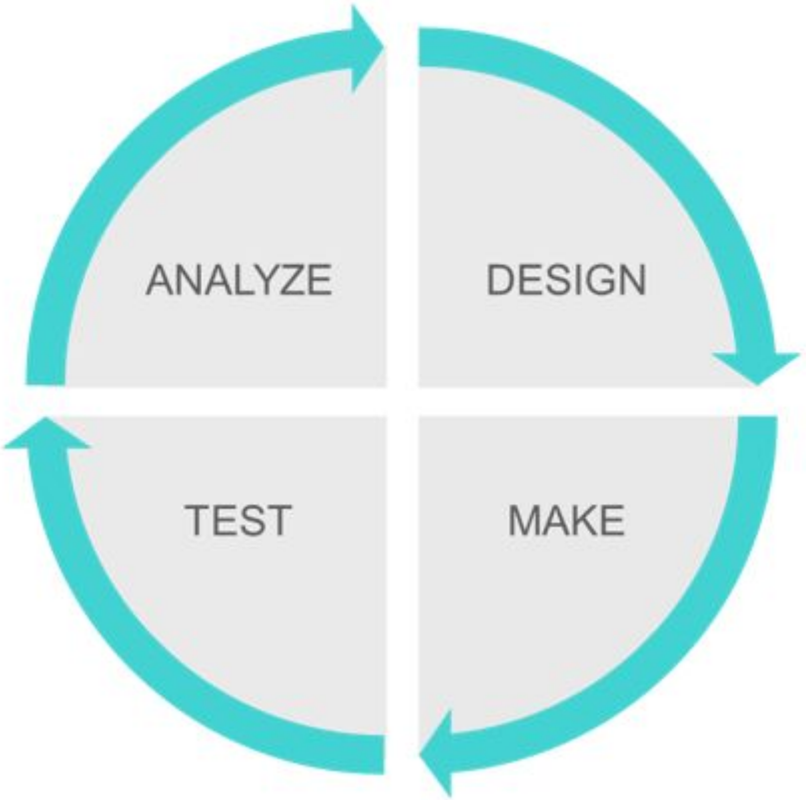
Accelerating Drug Discovery

KNIME DataHop U.S. - October 24, 2023, New York

The central paradigm of lead optimization in drug discovery

NEW HYPOTHESIS GENERATION

DESIGN MOLECULES TO TEST NEW HYPOTHESES



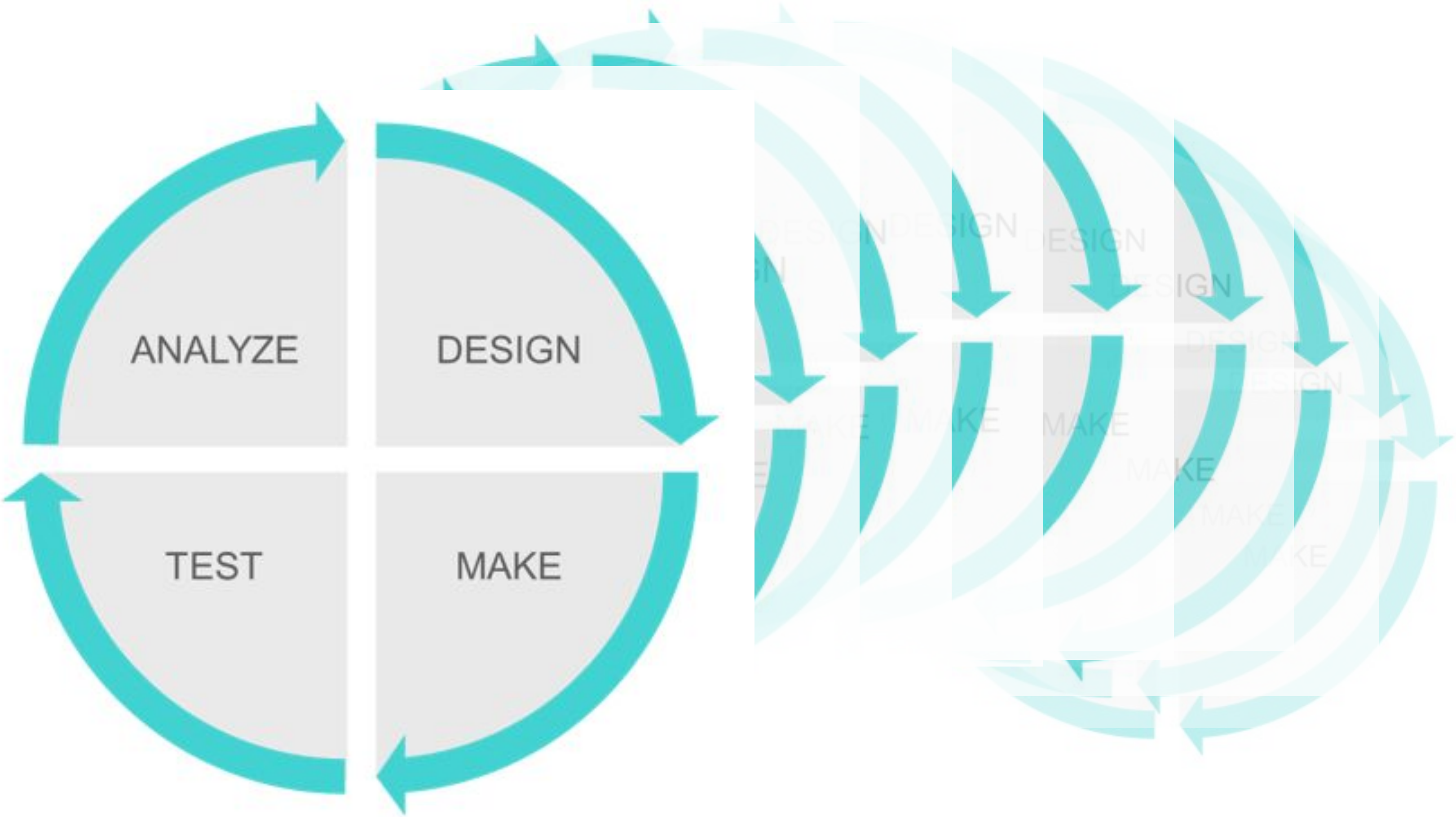
COLLECT DECISION DRIVING DATA EARLY

EFFECTIVELY DEPLOY CHEMISTRY RESOURCES

Iteration through design cycles with efficiency is critical

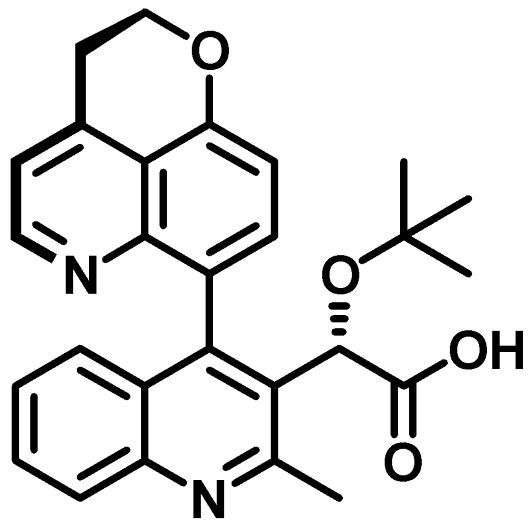
Optimization of design cycle time for efficient hypothesis testing

Minimize the number of iterations for rapid project progression



A key step in cheminformatics is the storage of chemical structures.

- A key challenge is representing chemical compounds in a way that **both humans and machines** can comprehend.
- Chemists use **2D and 3D visualizations** to represent chemical structures.
- Chemical structure representations may contain **explicit and implicit** information
- **Line notations** benefit both computational processes and human chemists in cheminformatics tasks.



Example from the literature. See: dx.doi.org/10.1021/ml500002n | *ACS Med. Chem. Lett.* **2014**, *5*, 422−427

SMILES string: CC1=NC2=CC=CC=C2C(=C1[C@@H](C(=O)O)OC(C)(C)C)C3=C4C5=C(C=C3)OCCC5=CC=N4

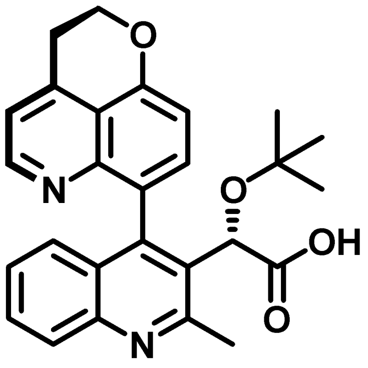
InChI:
1S/C27H26N2O4/c1-15-21(25(26(30)31)33-27(2,3)4)23(17-7-5-6-8-19(17)29-15)18-9-10-20-22-16(12-14-32-20)11
-13-28-24(18)22/h5-11,13,25H,12,14H2,1-4H3,(H,30,31)/t25-/m0/s1

InChIKey: MIXIIJCBELCMCZ-VWLOTQADSA-N

Creation of chemical fingerprints is a crucial step in cheminformatics

- Chemical fingerprints encode structure-based descriptors

Structure



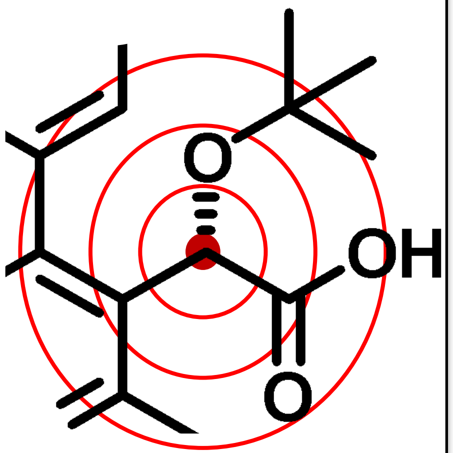
Descriptors

HEP_hCLint	HEP_mCLint	HEP_rCLint
8.75574502838689	19.2917832534951	12.2078510236194

Predictions

serum shift (50% HS), fold-change[b]	2.1
HLM/RLM ($t_{1/2}$), min	210/>300
Caco-2 (P_{app}), × 10^6 cm/s	14
CyP450 inh. (IC_{50}, 3A4/2D6), μM	23/>30
$logD_{7.4}$	0.44
solubility[c] (pH = 6.8), mg/mL	>0.85

Data from the wet lab

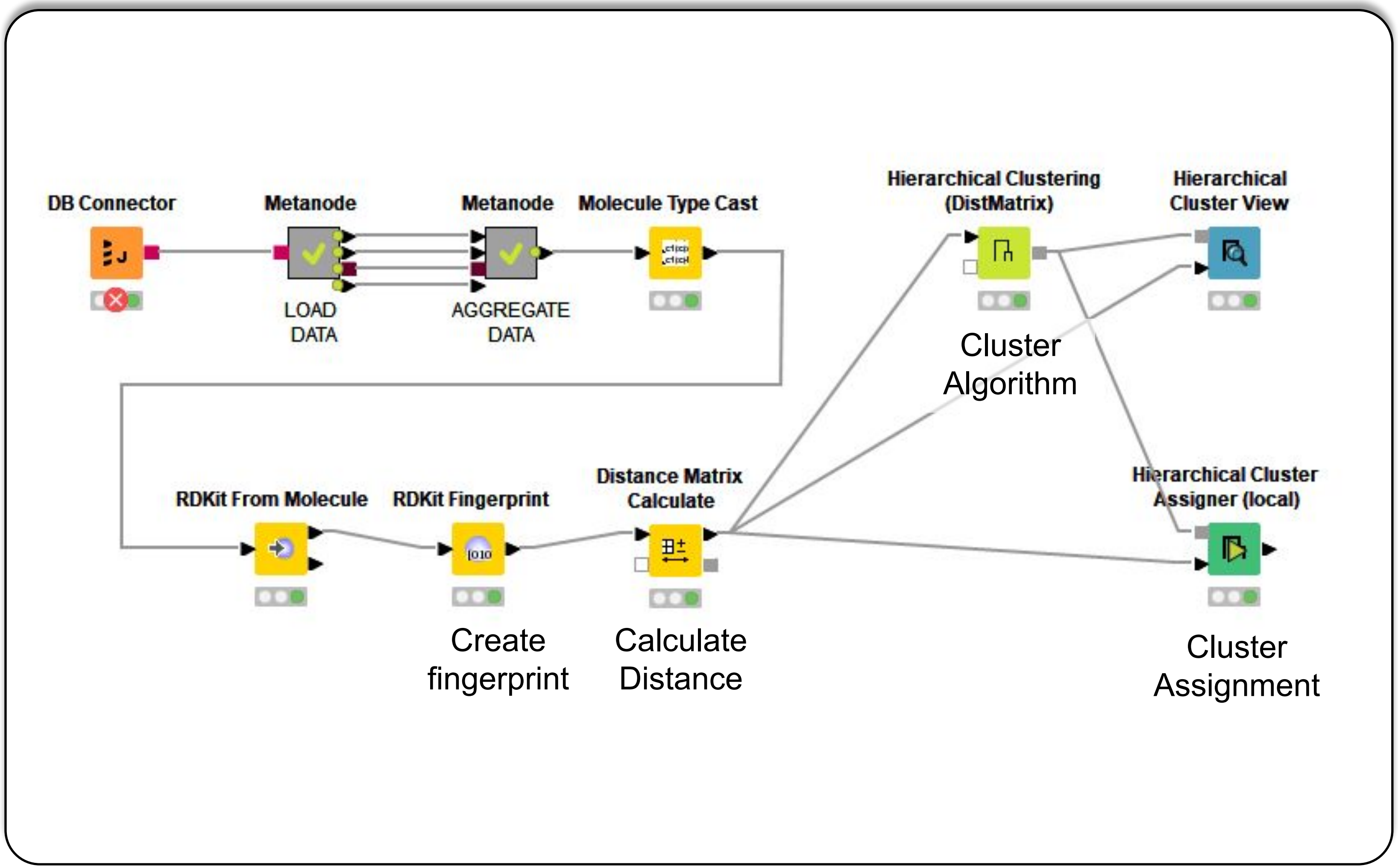


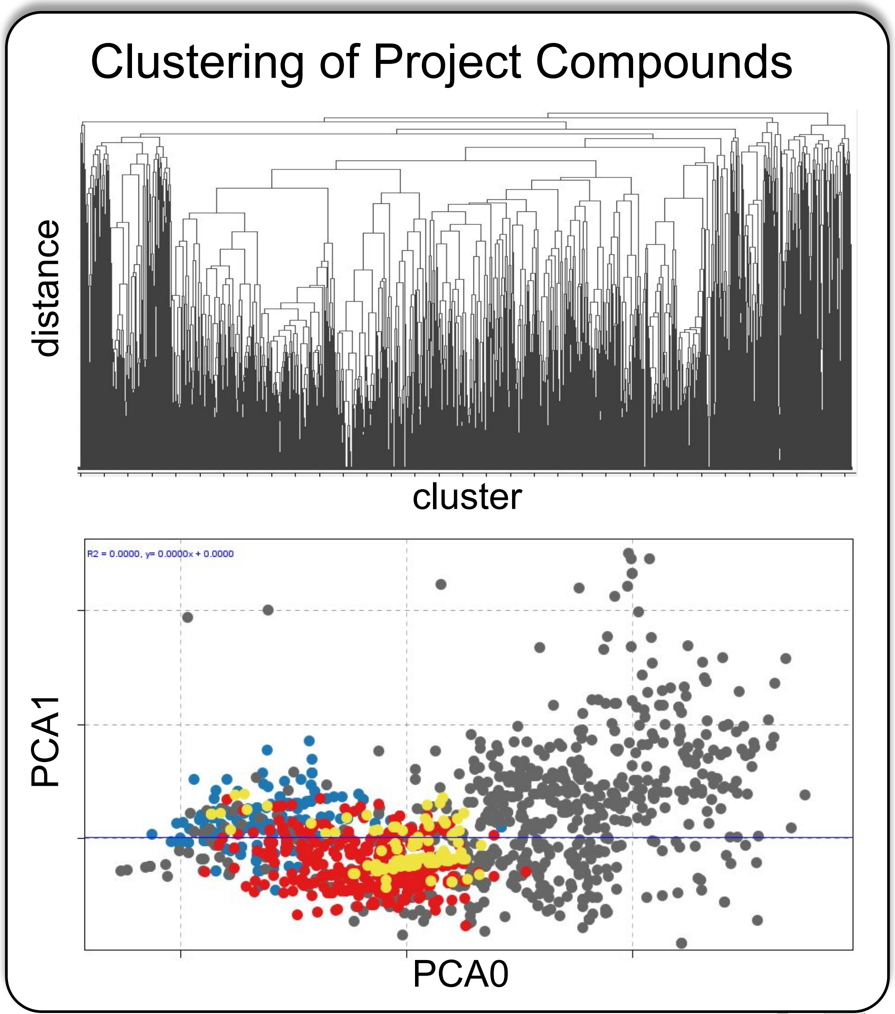
fingerprints

Bit vector representation

000000000000000000000000000001000000000000
010000000000000000000001000000000100000000
000000000000000100000000001000000000000000
000000010000000000010000010000000000000000
000000000010000001000000100010000000010010
001000000000100000000000000000001000000000
000000000000100000000001000000000000000000
000000000010010000000010000000000000010100
100100000000000100000000000000000000000010
000010000000000000000000000000000000000000
001000000000100000000000000000000000000000
000000000000000000000001000000000000000000
000000000000000000000000001000000000101000
000000000000000000000000000010000000000000
000000101000000000000010000000000000000001
001000000000000100100000000000100000000000
000000000001000100000000000000000001000000
000001000000000000000000000000000000001000
000000000000000000000000000000000000000001
000000000000000000000000000000000100000000
000000000000000000000000000010000000000000
000000000000000000001000000000000000000000
000000010000000000000000000001000000000000
000000010000000010000000000000000000000000
000000100000000000000000000001000000000010

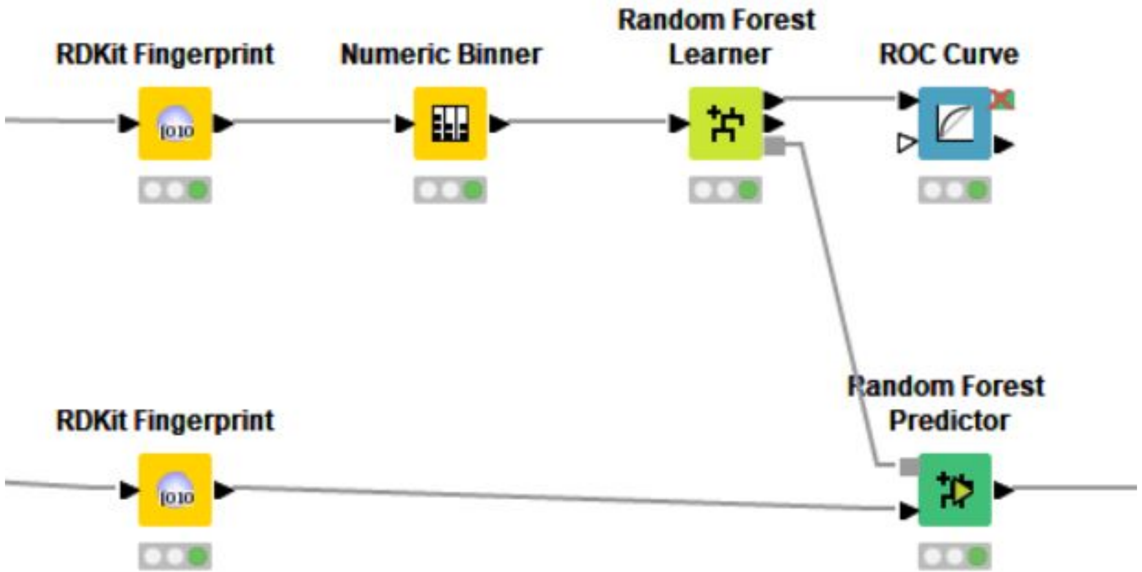
Clustering of compounds through similarity scoring is a common operation involving chemical fingerprints





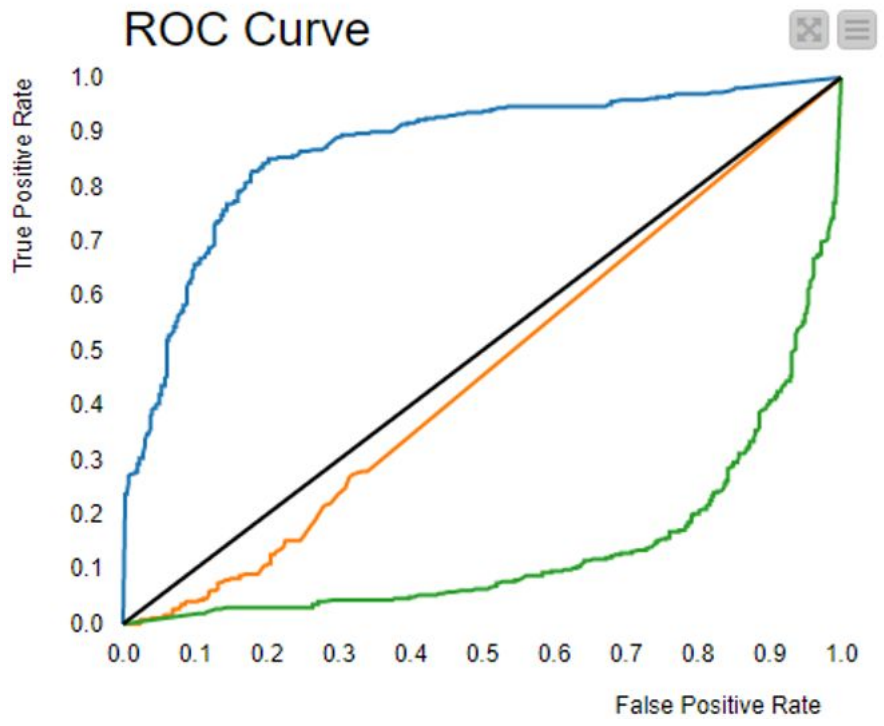
Quantitative Structure Activity Relationships (QSAR) is a contemporary approach to property prediction in Med Chem

Easily Trained and deployed in KNIME

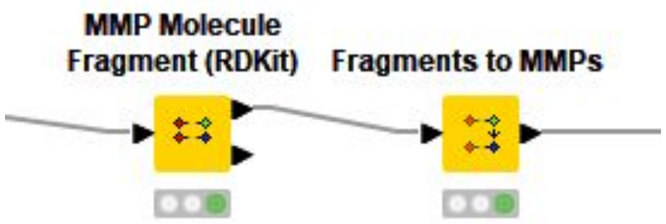


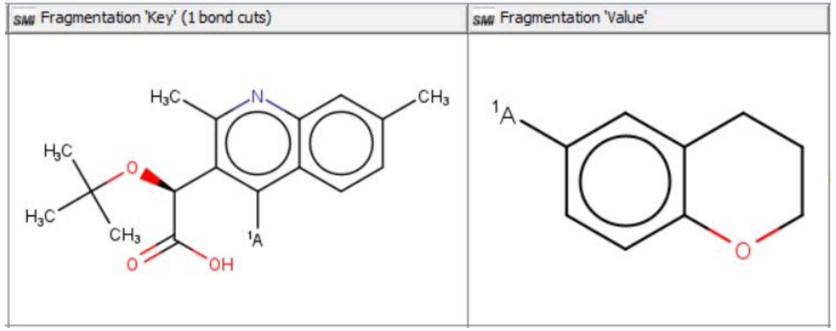
Consensus Approach with complimentary fingerprints often gives best results

Performance is suitable for guiding design



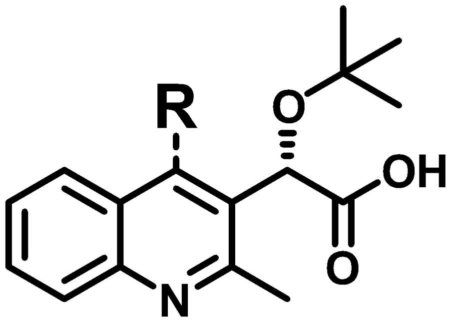
Matched Molecular Pair (MMP) analysis is an effective approach to SAR interpretation

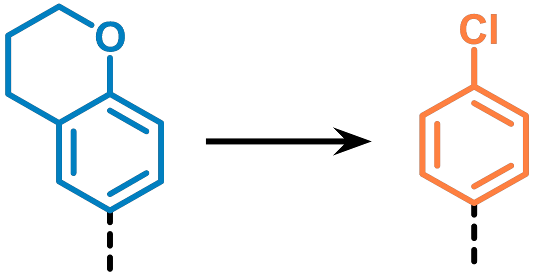




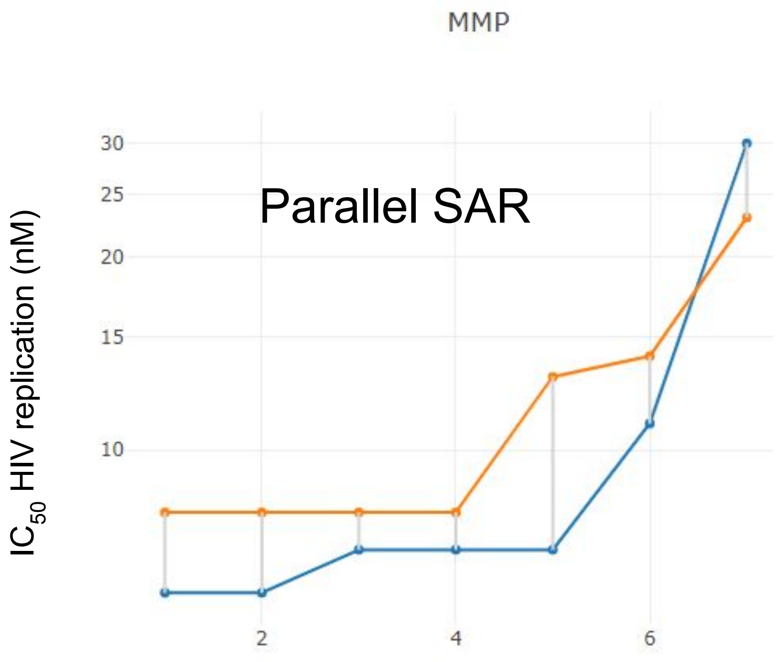
Fragment Key
(constant in a matched pair)

Fragment Value
(variable in a matched pair)

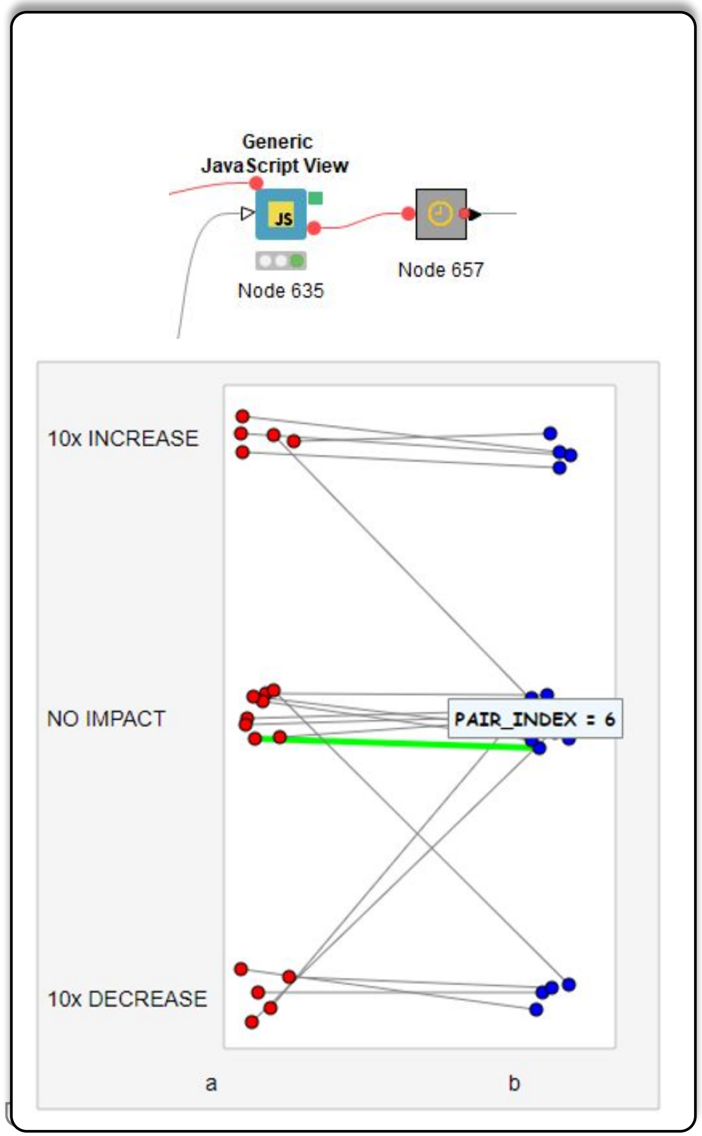


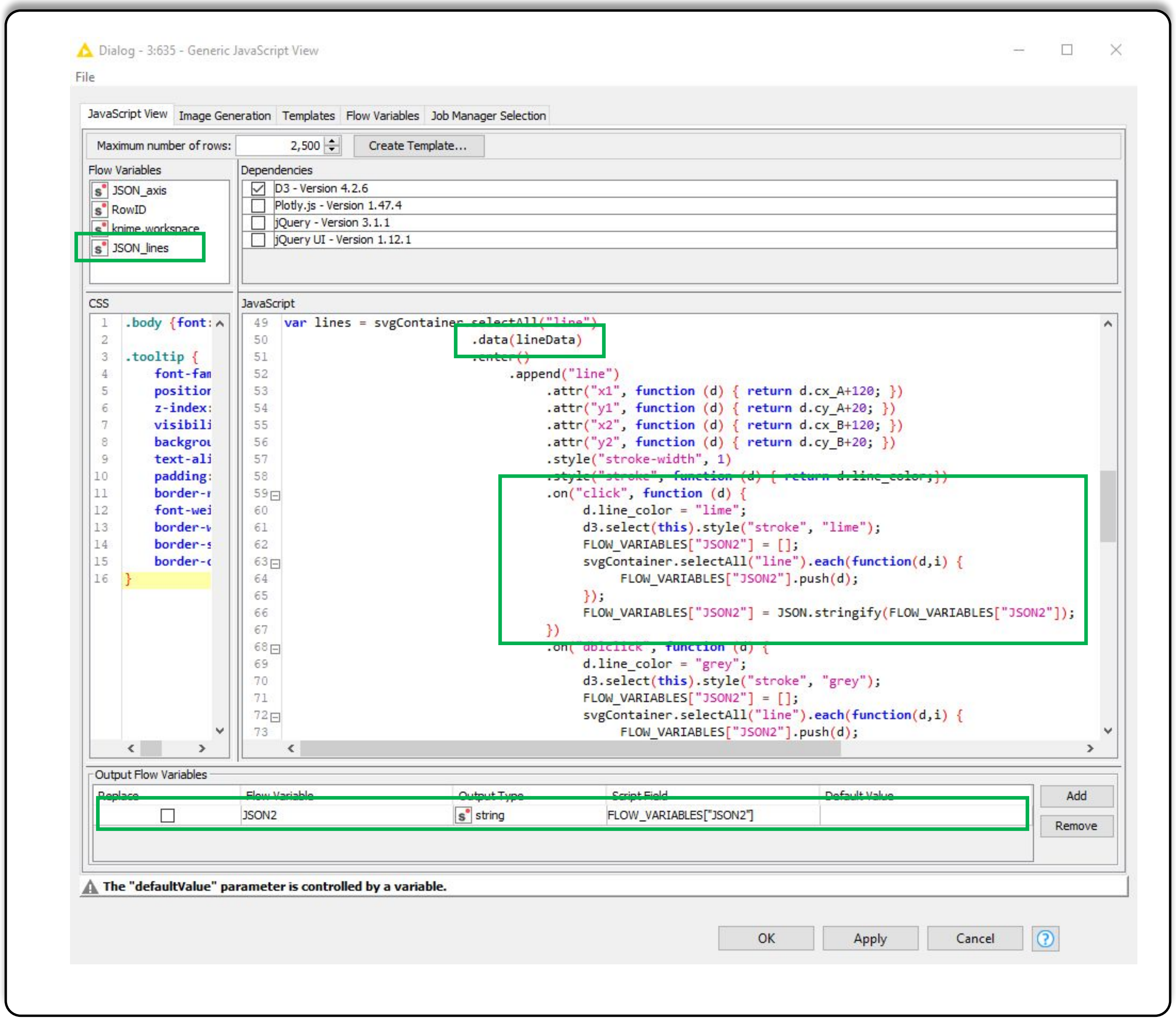


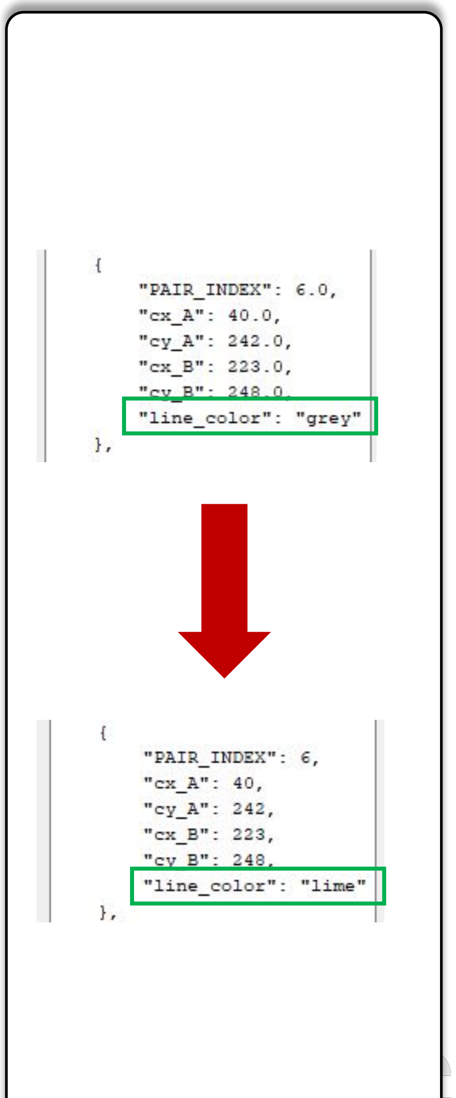
MMP Transformation



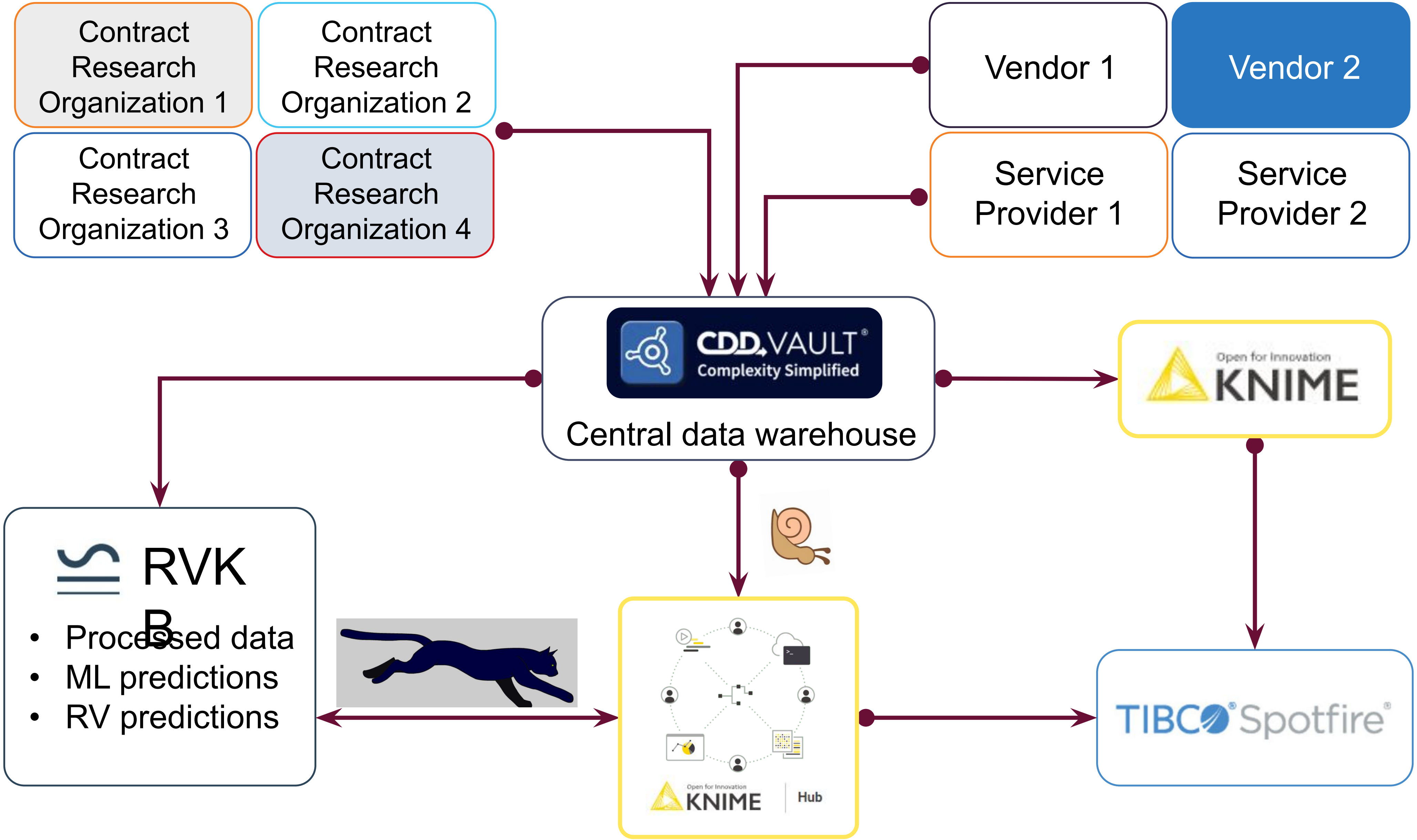
The Generic JavaScript View: A workhorse for data visualization-driven decision making



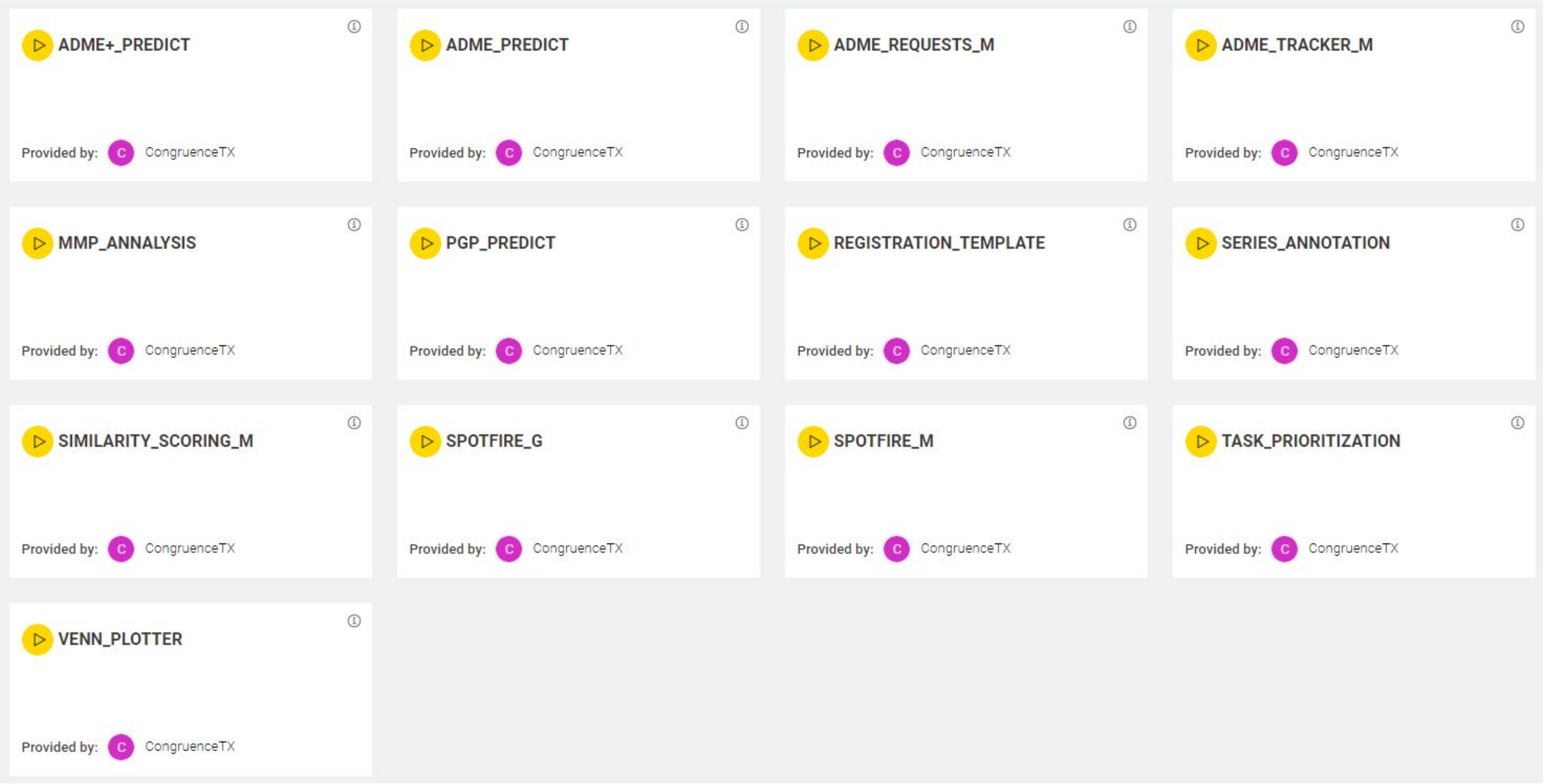




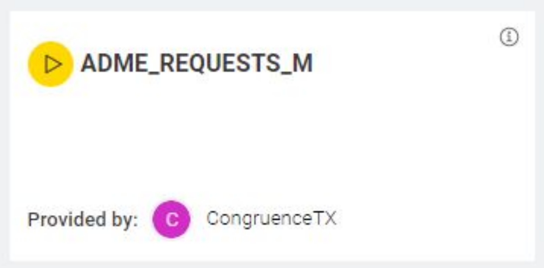
How does compound data flow at Congruence

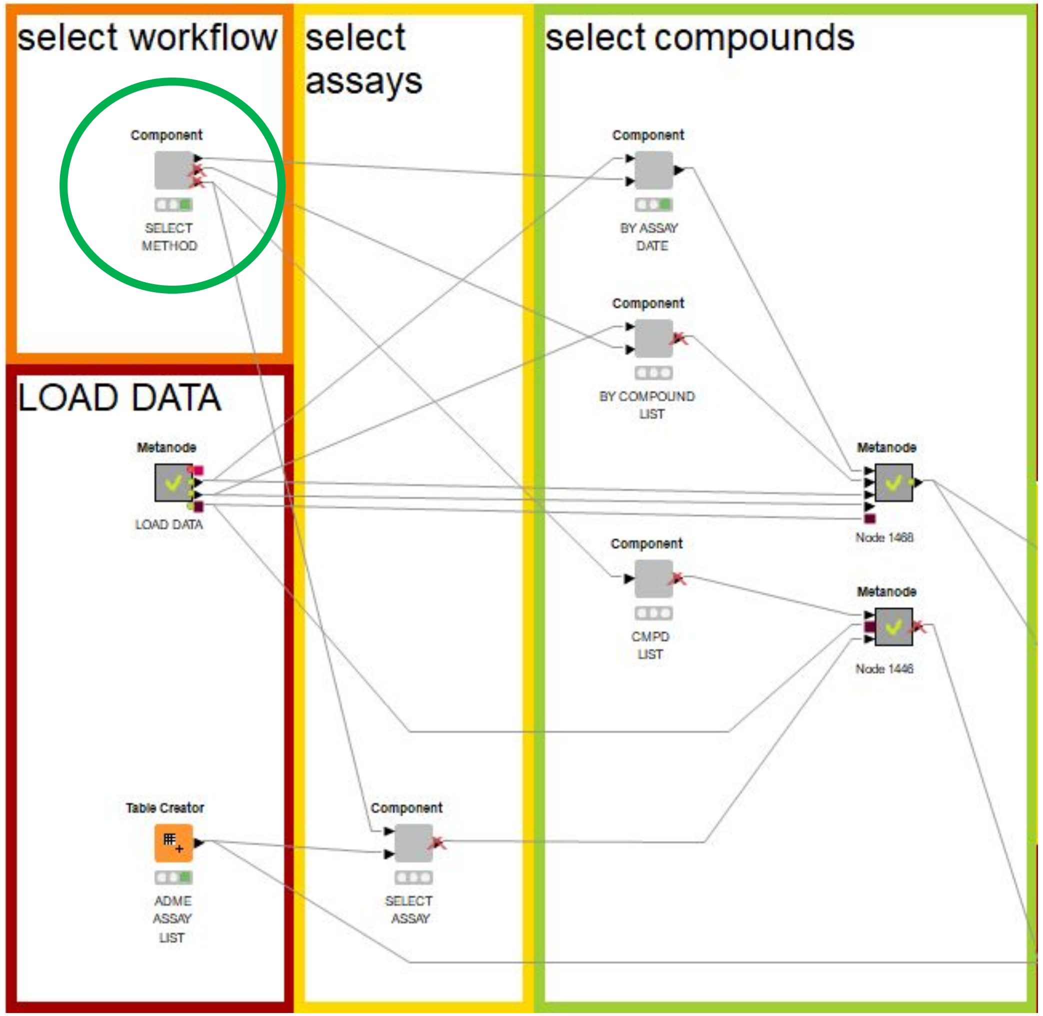


The KNIME Hub allows for rapid deployment of cheminformatic tools to guide design and decision making

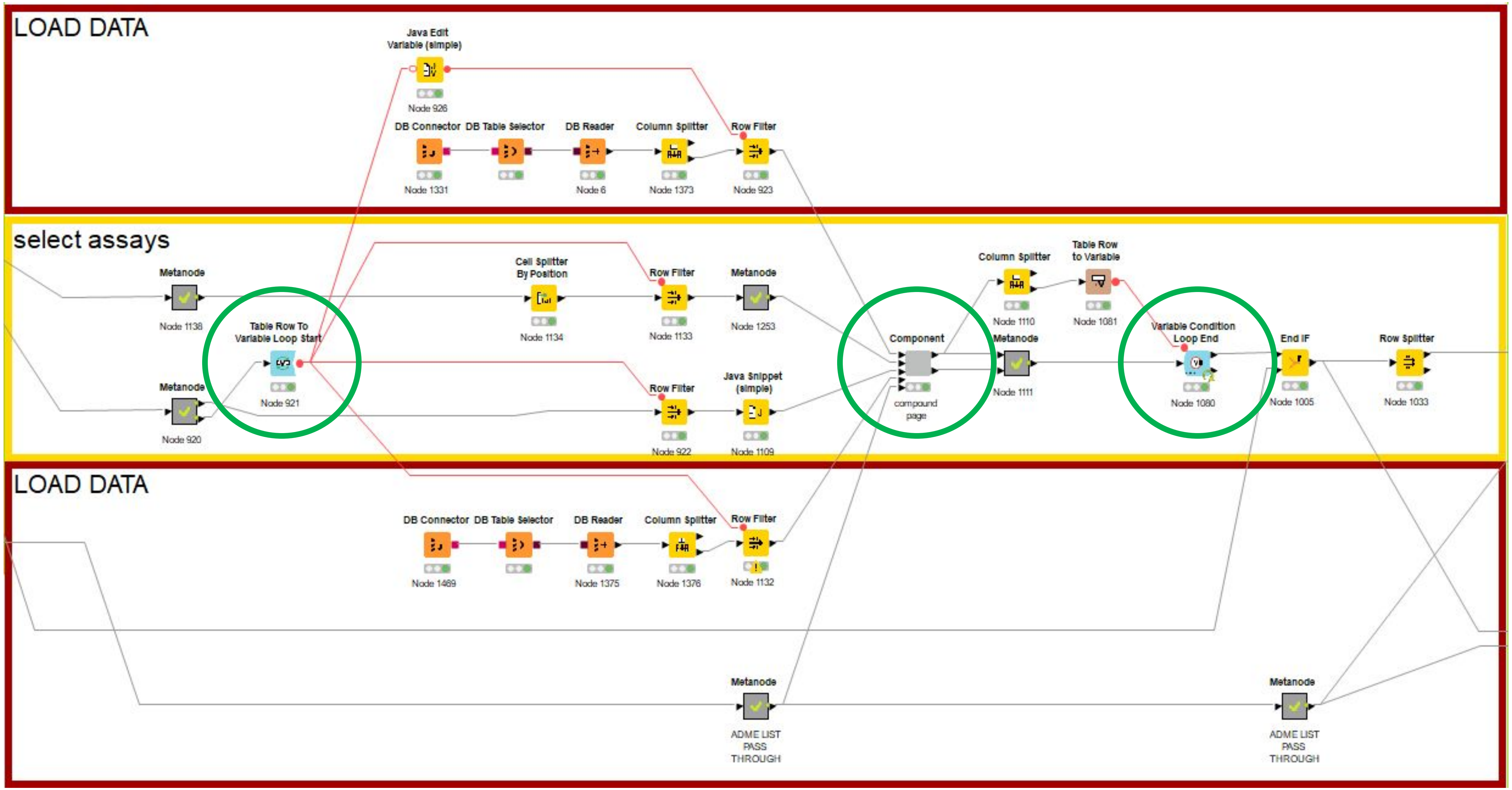


Bringing it all together:
Using KNIME Business Hub to accelerate drug discovery

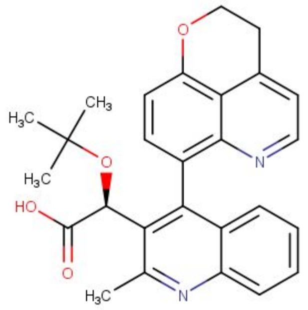


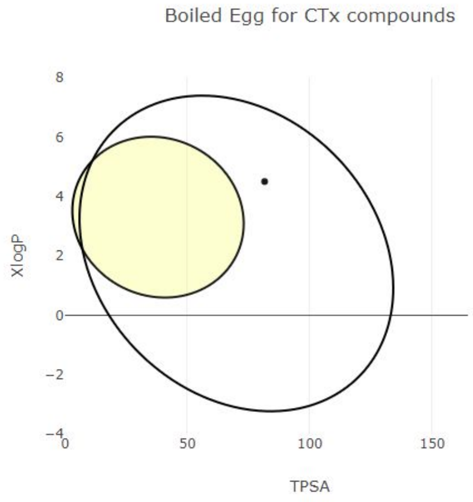


Data Apps with Components within Loops: A versatile approach to decision making



With KNIME and KNIME Hub, decision making tools designed to retrieve and analyze data are easily assembled and deployed





PROGRESS

compound 10 out of 11

☐ Check this box to stop after this entry.

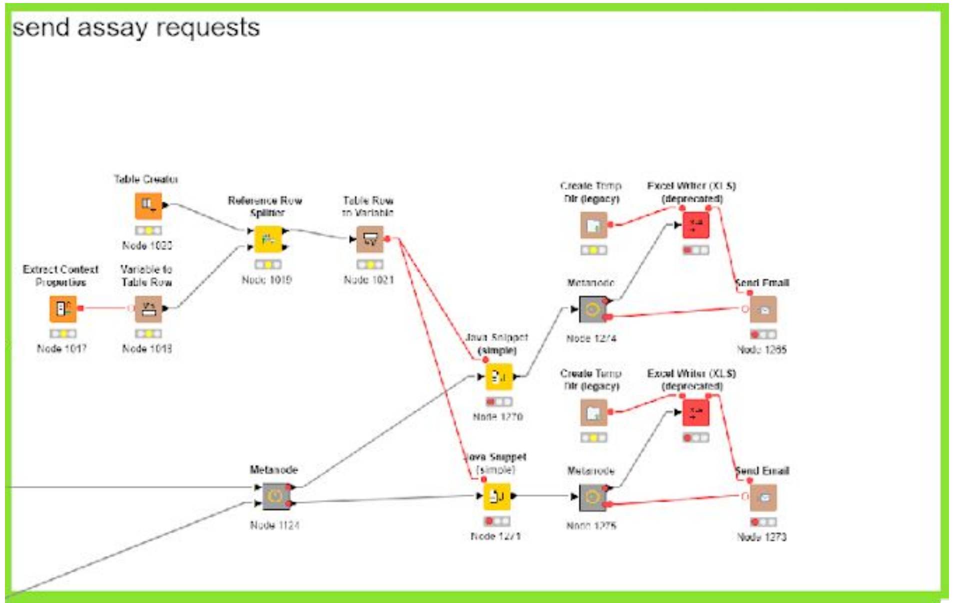
RowID	Row0
Topological Polar Surface Area	81.54
Molecular Weight	442.18925731200073
XLogP	4.499
CYP_HLM_CLint	10.424613703671
HEP_hCLint	8.75574502838692
HEP_mCLint	19.291783253495
HEP_rCLint	12.2078510236195
UGT1A1	Yes (85%)
UGT1A10	No (83%)
UGT1A3	Yes (78%)
UGT1A4	No (64%)
UGT1A6	No (97%)

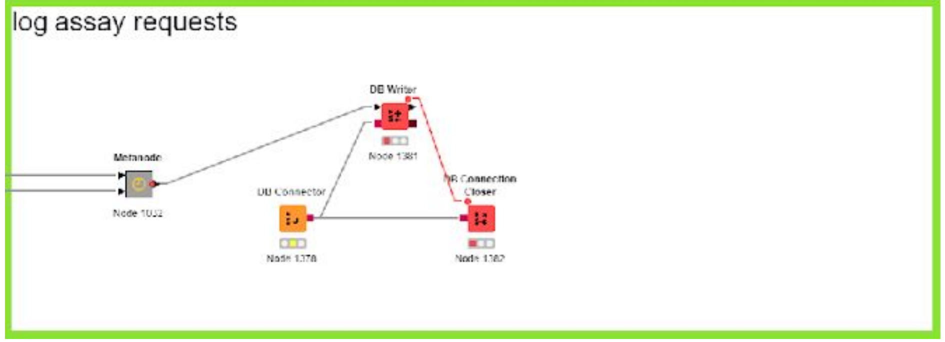
Assay	Days since request
Stability in liver microsomes (5 time points, only NADPH)_mouse	3
Stability in liver microsomes (5 time points, only NADPH)_mouse	3
Stability in liver microsomes (5 time points, only NADPH)_human	3
Stability in liver microsomes (5 time points, only NADPH)_human	3
hERG single point automatedl patch-clamp system (% inh at 10 uM, minimun 10 compounds)	3

Showing 1 to 5 of 5 entries

ASSAY REQUEST LIST

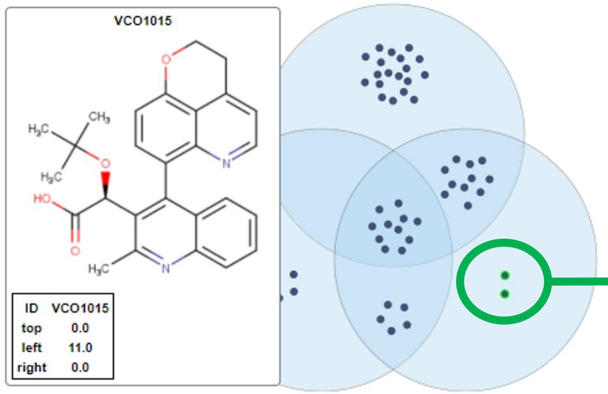
☐	ASSAY	Compound Number
☐	Caco-2 (A to B)	not requested
☐	Stability in liver microsomes (5 time points, only NADPH)_mouse	pending
☐	Stability in liver microsomes (5 time points, only NADPH)_human	pending





Example: the Venn Diagram-Scatter Plot hybrid
A custom visualization for guided analytics

Refresh



Filter based on selection

```
.on("click", function (d) {
    d.line_color = "lime";
    d3.select(this).style("stroke", "lime");
    FLOW_VARIABLES["JSON2"] = [];
    svgContainer.selectAll("line").each(function(d,i) {
        FLOW_VARIABLES["JSON2"].push(d);
    });
    FLOW_VARIABLES["JSON2"] = JSON.stringify(FLOW_VARIABLES["JSON2"]);
```

Number of Compounds with all three data points:

56

left: HLM 2pt (%QH)	top: CYP Inhibition HTL: % Inh 3A4-T	right: hERG Inhibition APC Single Point: % Inhibition at 10 uM
max	max	max
75	50	50
min	min	min
0	0	0

Key Takeaways

- In medicinal chemistry, the central paradigm of compound optimization follows a Design-Make-Test-Analyze cycle that is iterated over as efficiently as possible.
- To do this effectively, different representations of chemical structures that are interpretable by machines and can be translated to human interpretable representations are needed.
- Chemical descriptors can be derived from chemical structures and assembled into fingerprints, which are used for myriad analytical tasks, modelling and machine learning.
- Aggregating predictions, biological data and other cheminformatic outputs enables informed decision making in drug discovery.
- The KNIME Business Hub allows for flexible construction of workflows, services and data apps that facilitate decision making and accelerate drug discovery.