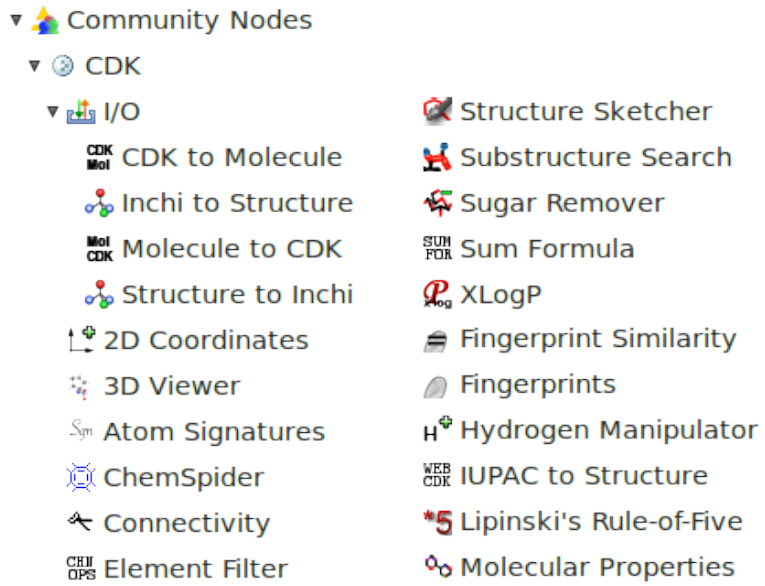
CDK Community Feature

Chemistry Development Kit (CDK)

An Open Source Java™ Library for Structural Chem- and Bioinformatics.

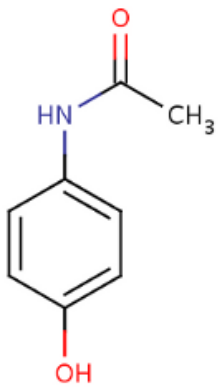
- Input/Output
- Visualisation
- Modeling
- Chemical Graphs
- Structure Generation
- Molecular Properties

→ http://tech.knime.org/community/cdk ←



Mass-based Sum Formula Prediction and Selection

1. Read in accurate masses.

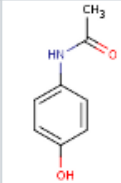
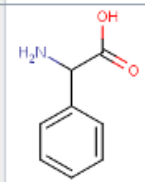
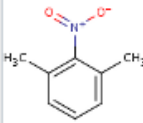
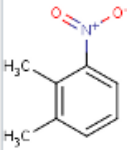
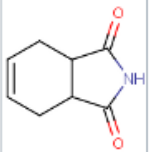


C8H9NO2
=
151.063

2. Calculate likely sum formulas.

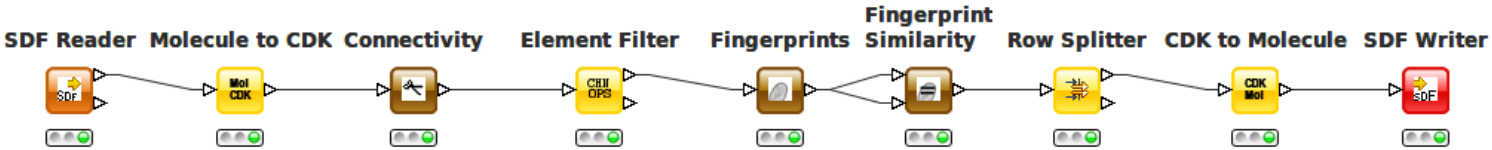
Split Value 1	C8H9NO2
Split Value 2	C4H5N7
Split Value 3	C3H9N3O4
Split Value 4	C2H9N5O3
Split Value 5	C7H9N3O
Split Value 6	C5H5N5O
Split Value 7	C4H9NO5
Split Value 8	H5N7O3
Split Value 9	CH9N7O2
Split Value 10	C11H5N
Split Value 11	C5H13NO4
Split Value 12	C6H9N5
Split Value 13	C6H5N3O2
Split Value 14	C2HN9
Split Value 15	CH5N5O4
Split Value 16	H9N9O
Split Value 17	C4H13N3O3
Split Value 18	C7H5NO3
Split Value 19	C9H13NO
Split Value 20	C3HN7O
Split Value 21	C2H5N3O5
Split Value 22	C3H13N5O2
Split Value 23	C9HN3
Split Value 24	C8H13N3
Split Value 25	C4HN5O2

3. Query Pubchem for matching structures.

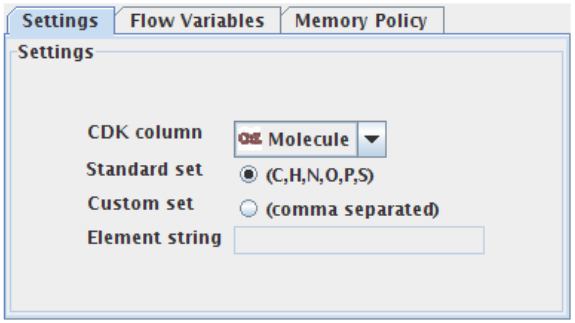
Row0		C8H9NO2
Row1		C8H9NO2
Row2		C8H9NO2
Row3		C8H9NO2
Row4		C8H9NO2

EMBL-EBI

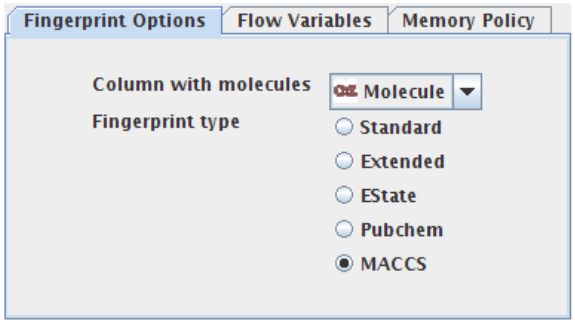
Similarity-driven Molecule Library Creation



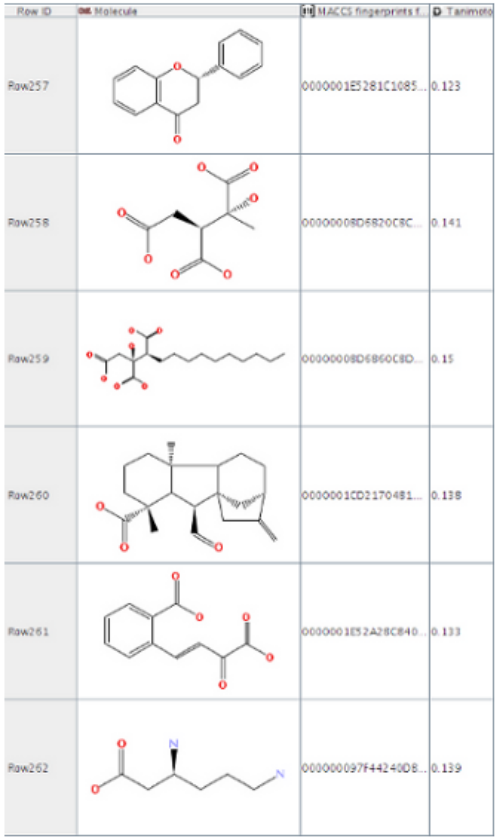
1. Filter molecules by elements.



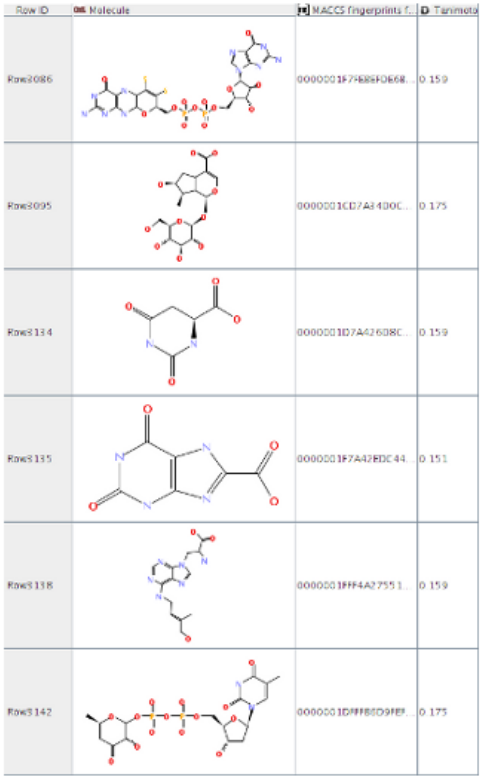
2. Calculate MACCS Fingerprint.



3. Calculate average Tanimoto distance over all molecules.

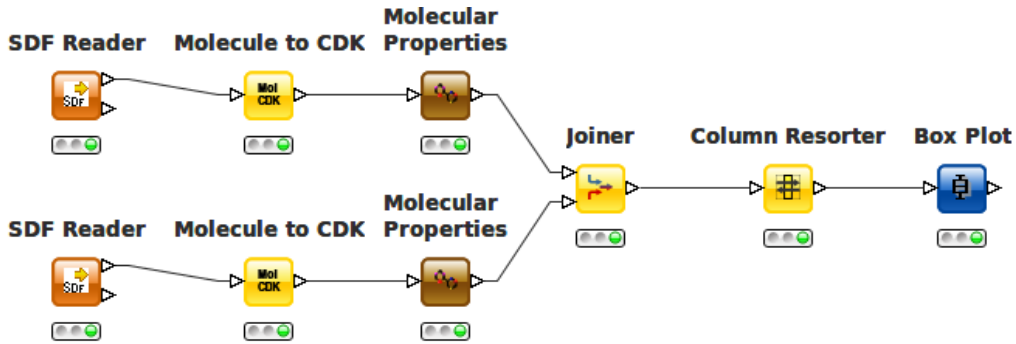


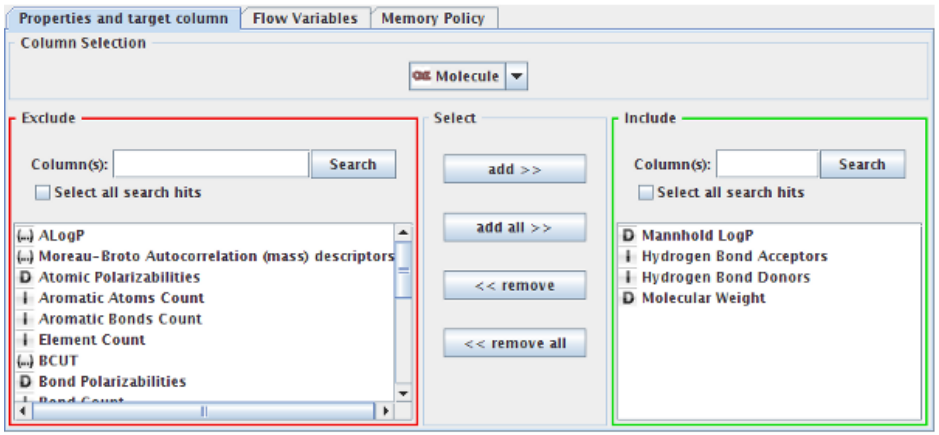
4. Keep only molecules with common structural features.

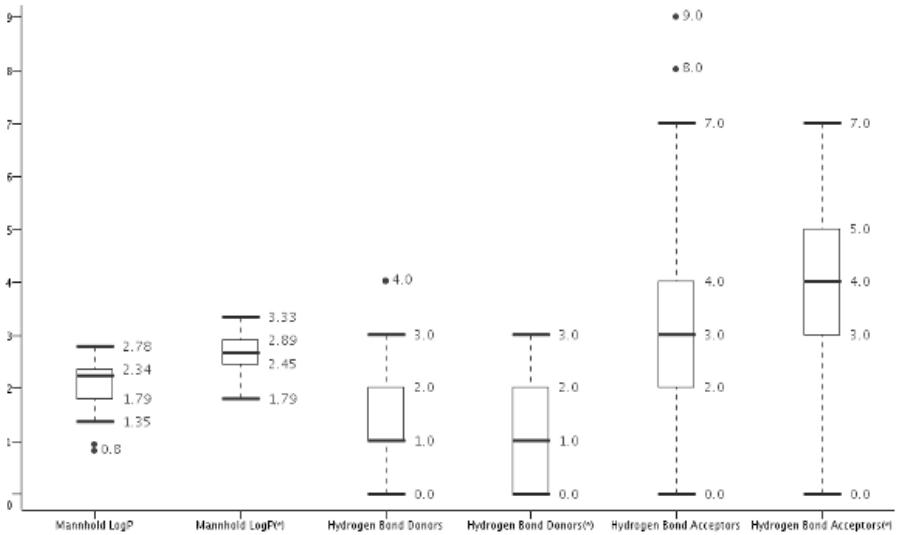


EMBL-EBI

Molecular Properties Calculation and Comparison







EMBL-EBI

Acknowledgements

The Chemoinformatics and Metabolism Team

- Christoph Steinbeck

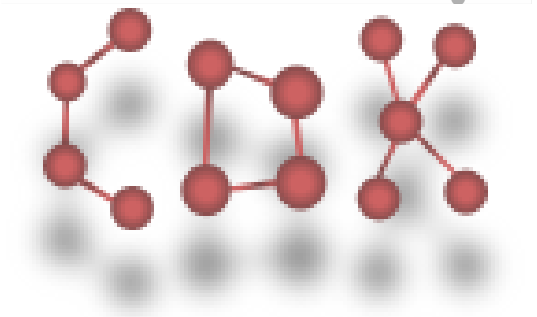
The CDK Project Admins

- Egon Willighagen
- Miguel Rojas
- Christoph Steinbeck

The KNIME Team

- Thorsten Meinl
- Bernd Wiswedel

All CDK Developers & Contributors,
Syngenta AG, The University of Cambridge



- Steinbeck, C.; Hoppe, C.; Kuhn, S.; Guha, R.; Willighagen, E. L. Current Pharmaceutical Design 2006, 12, 2111-2120.
- Steinbeck, C.; Han, Y. Q.; Kuhn, S.; Horlacher, O.; Luttmann, E., Willighagen, E. Journal of Chemical Information and Computer Sciences 2003, 43, 493-500.