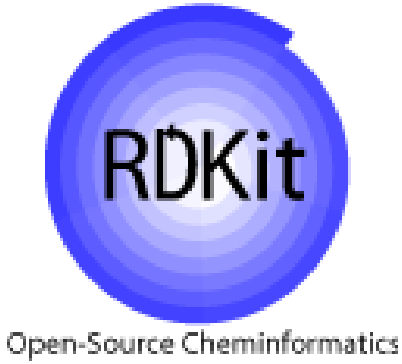


Open source cheminformatics in KNIME with the RDKit: update

Gregory Landrum

NIBR IT

Novartis Institutes for BioMedical Research, Basel

5th KNIME Users Group Meeting

Zurich, 2 February 2012

Acknowledgements

- Novartis:
  - Tom Digby (Legal)
  - John Davies (CPC)
  - Steve Litster (NIBR IT)
  - Richard Lewis (GDC)
  - Patrick Warren (NIBR IT)
  - Andy Palmer (NIBR IT)
  - Manuel Schwarze (NIBR IT)
  - Dillip Kumar Mohanty (NIBR IT)
  - Sayantan Sengupta (NIBR IT)
- Rational Discovery:
  - Santosh Putta (currently at Nodality)
  - Julie Penzotti
- RDKit open-source community
- KNIME forum members
- knime.com
  - Michael Berthold
  - Thorsten Meinl
  - Bernd Wiswedel

NOVARTIS

RDKit: What is it?

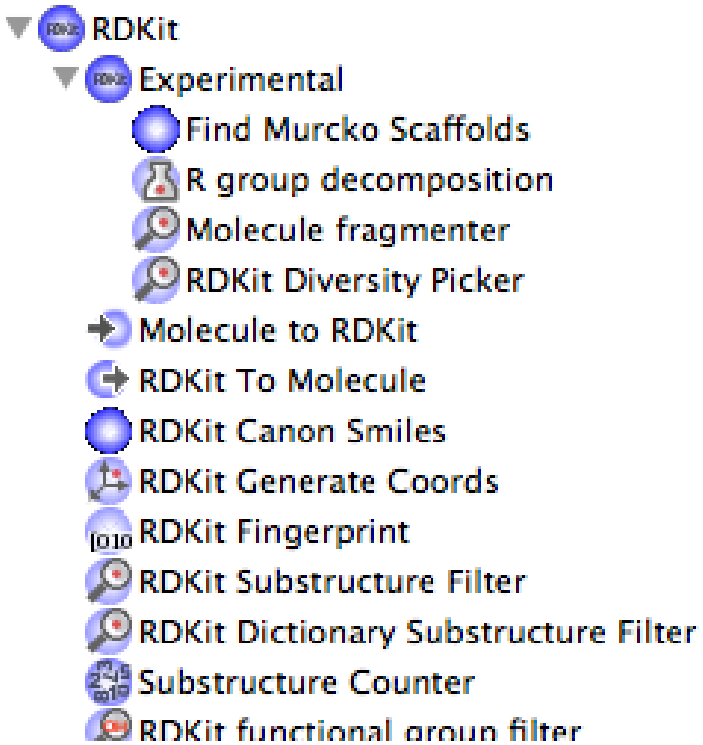
- Python (2.x), Java, C++ toolkit for cheminformatics
  - Core data structures and algorithms in C++
  - Heavy use of Boost libraries
  - Python wrapper generated using Boost.Python
- Functionality:
  - 2D and 3D molecular operations
  - Descriptor generation for machine learning
  - Database cartridge for substructure and similarity searching
  - Supports Mac/Windows/Linux
- History:
  - 2000-2006: Developed and used at Rational Discovery for building predictive models for ADME, Tox, biological activity
  - June 2006: Open-source (BSD license) release of software, Rational Discovery shuts down
  - to present: Open-source development continues, use within Novartis, contributions from Novartis back to open-source version

NOVARTIS

Knime integration[1]

- Out of the box Knime is strong on data processing and mining, weak on chemistry.
- Goal: develop a set of *open-source* RDKit-based nodes for Knime that provide basic cheminformatics functionality

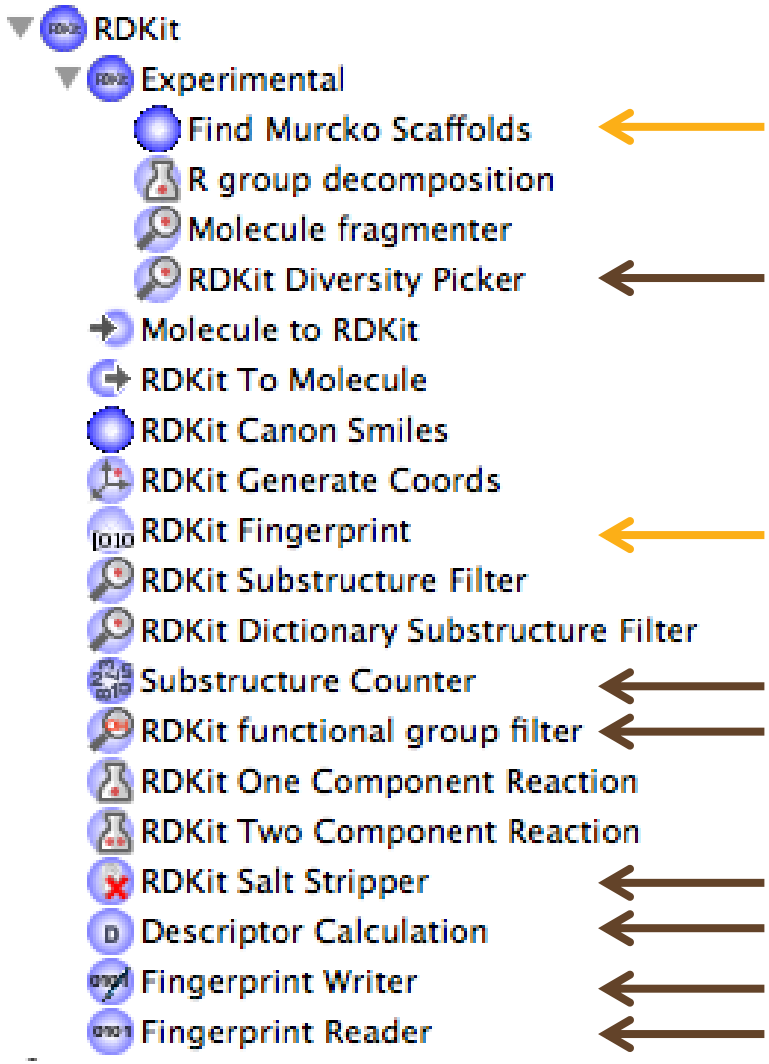
- Distributed from knime community site
- Binaries available as an update site (no RDKit build/installation required)
- Work in progress: more nodes being added (new wizard makes it *easy*)



[1] Work done together with knime.com

NOVARTIS

What's there?



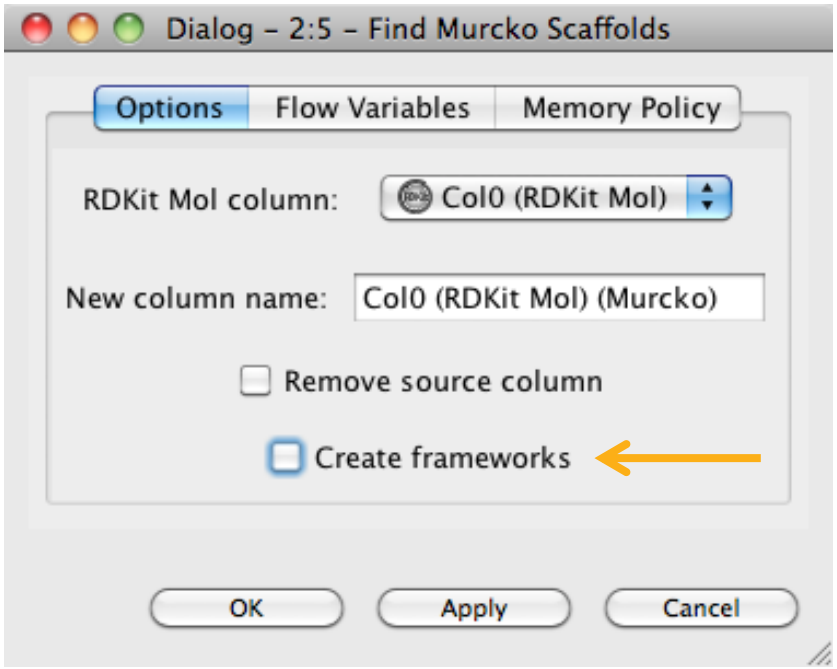
Another new piece: 64bit Windows support

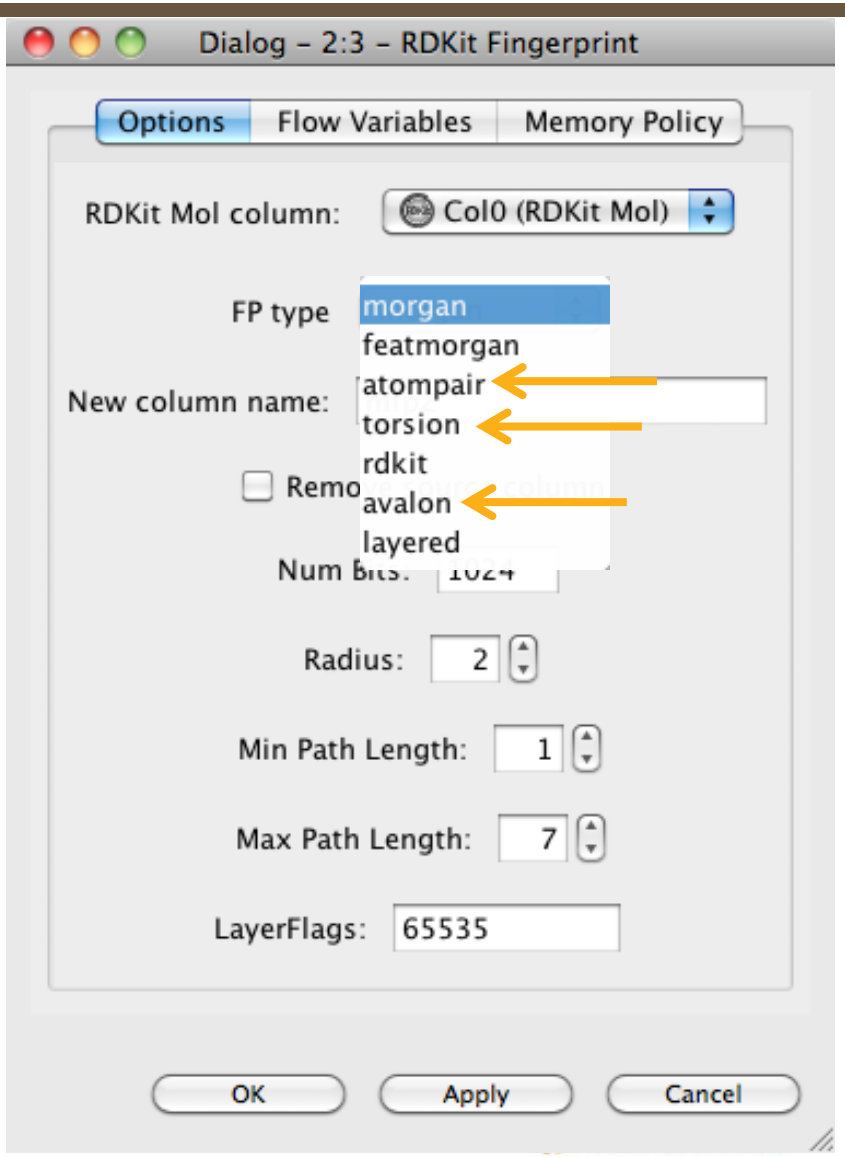
New

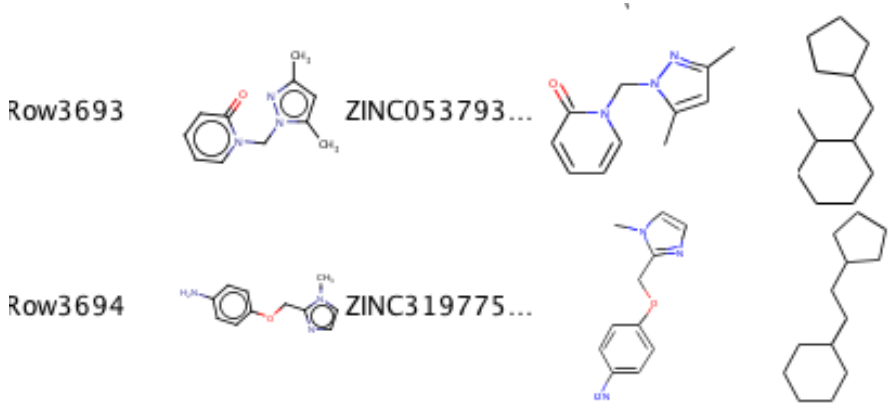
Updated

NOVARTIS

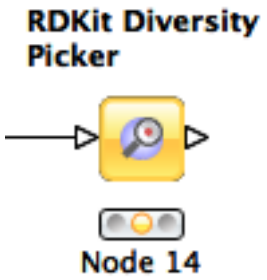
Modified nodes

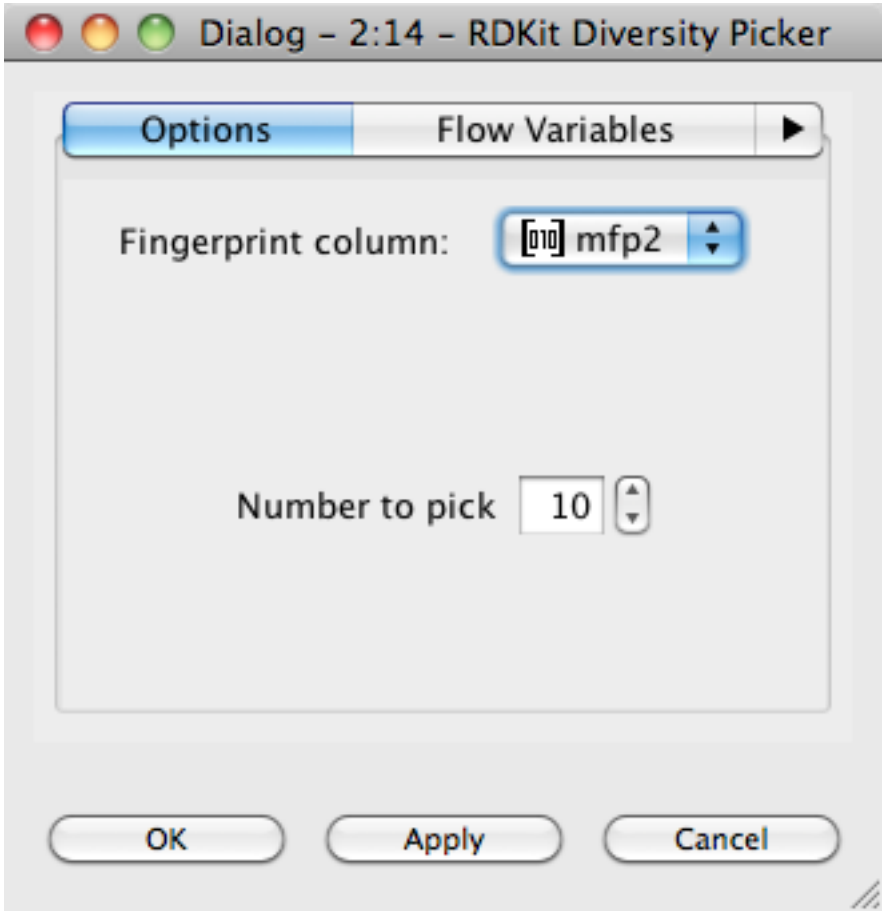






Diversity Picker





NOVARTIS

Functional group filter

Dialog – 2:6 – RDKit functional group filter

Settings | Flow Variables | Memory Policy

Select molecule column

Column name : Col0 (RDKit Mol)

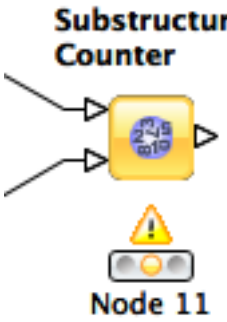
Select functional group definition file (Optional)

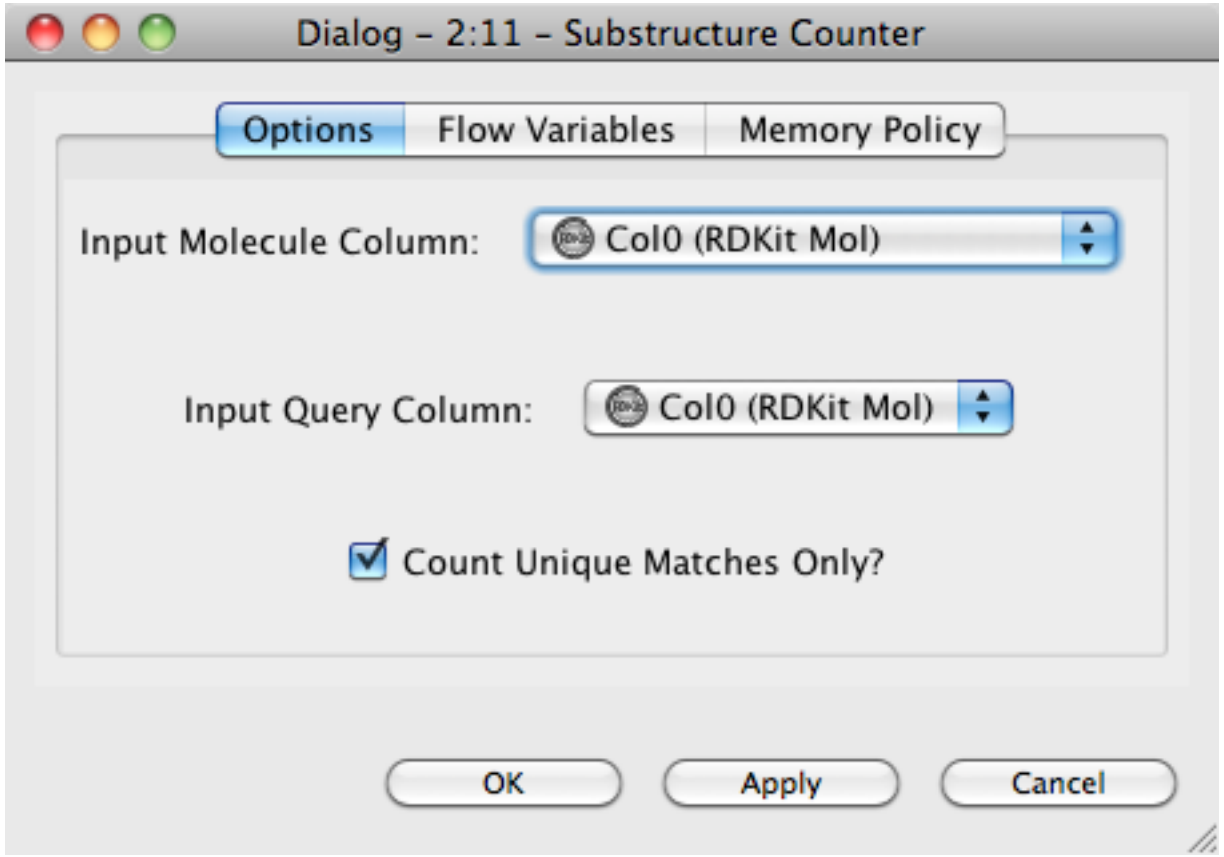
File path : Browse... Load default

List of available functional group filters

Select	Functional Group Name	Qualifier	Count
☐	AcidChloride	Exactly(=)	0
☐	Aromatic AcidChloride	Exactly(=)	0
☐	Aliphatic AcidChloride	Exactly(=)	0
☐	CarboxylicAcid	Exactly(=)	0
☐	Aromatic CarboxylicAcid	Exactly(=)	0
☐	Aliphatic CarboxylicAcid	Exactly(=)	0
☐	AlphaAmino CarboxylicAcid	Exactly(=)	0
☐	SulfonylChloride	Exactly(=)	0
☐	Aromatic SulfonylChloride	Exactly(=)	0
☐	Aliphatic SulfonylChloride	Exactly(=)	0
☑	Amine	Exactly(=)	1
☐	Primary Amine	Exactly(=)	0

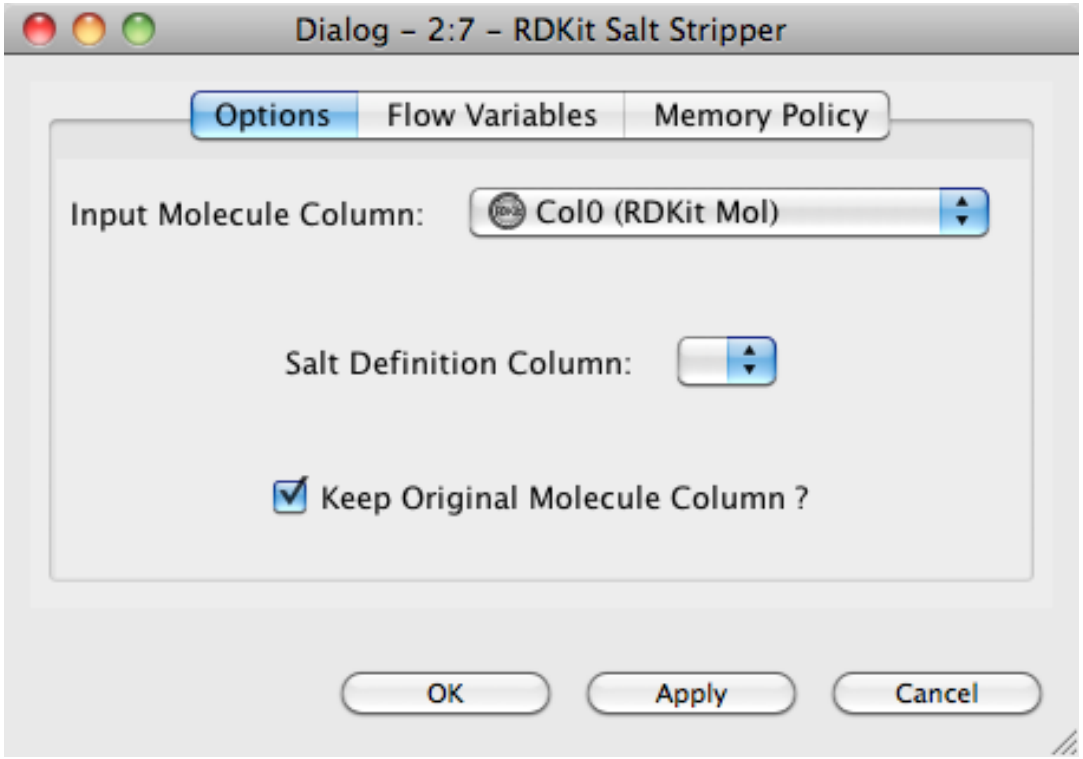
Substructure Counter





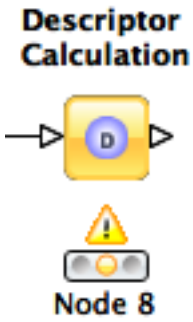
NOVARTIS

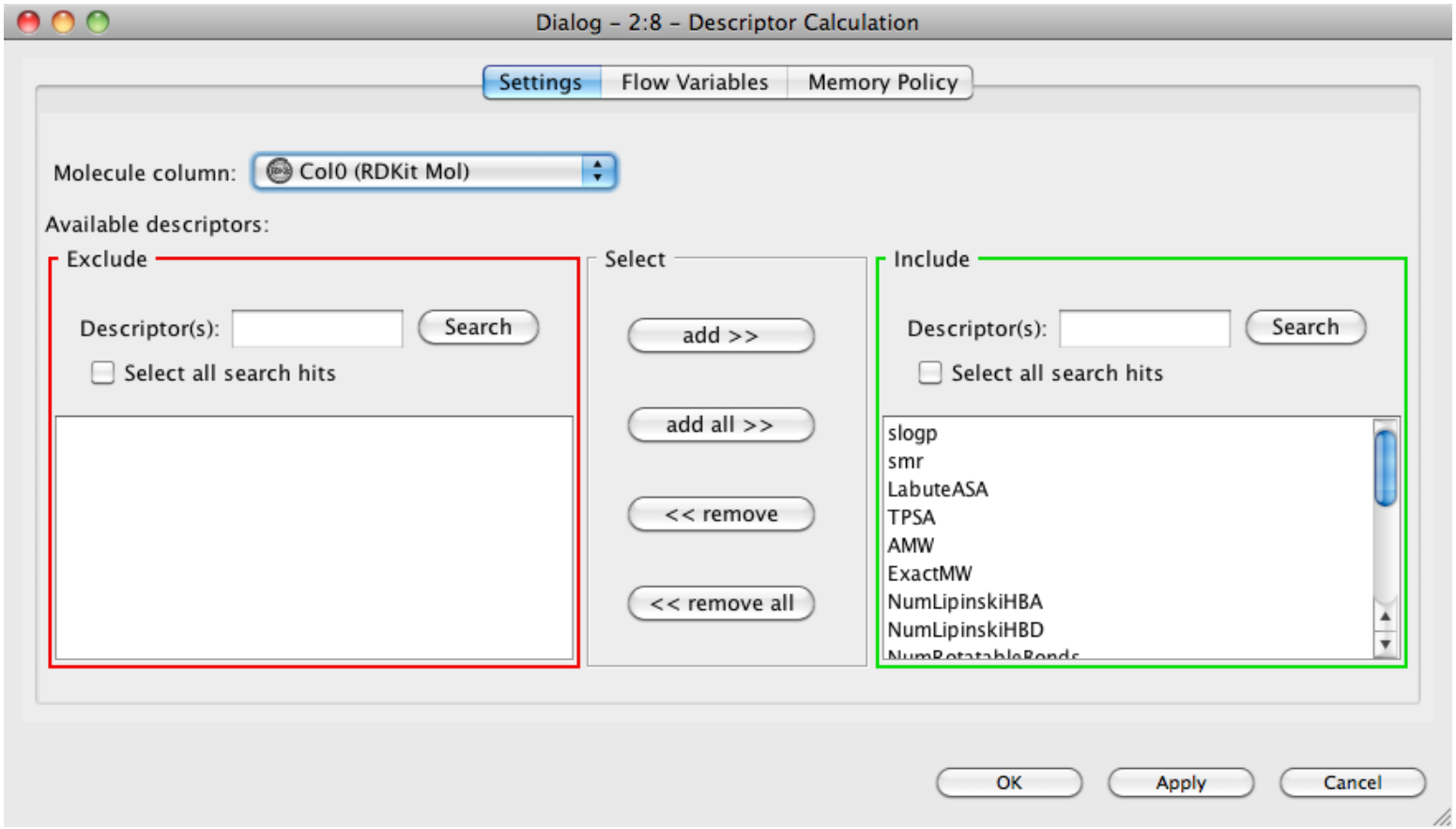
Salt stripper



NOVARTIS

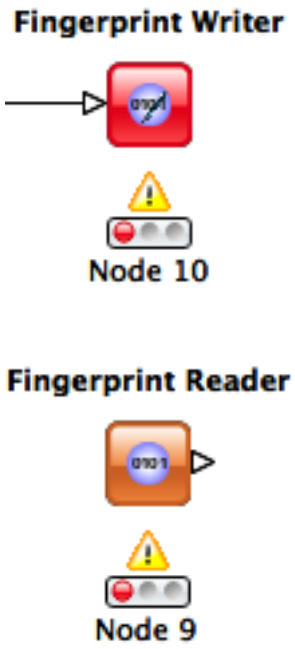
Descriptor calculator

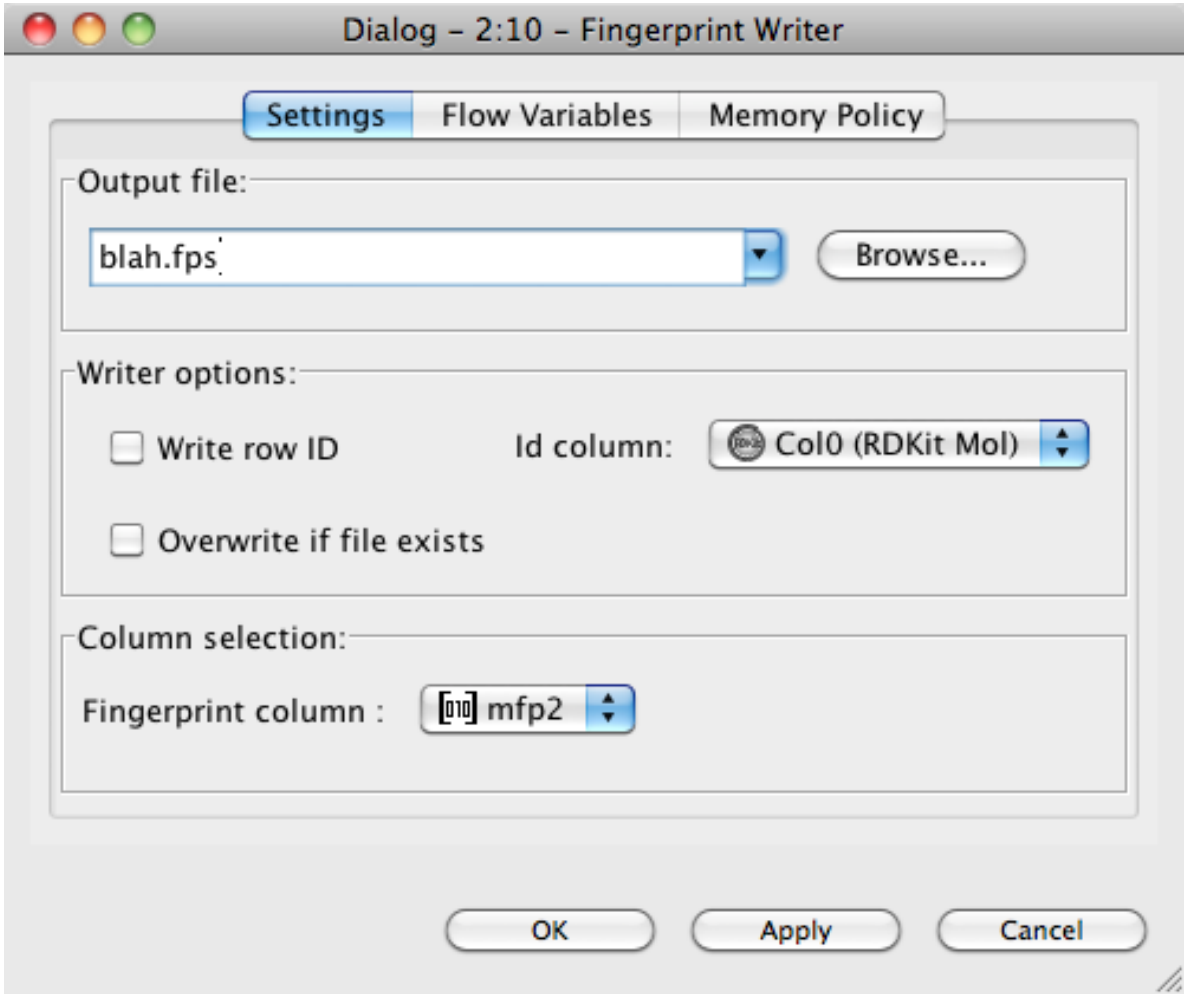




NOVARTIS

Fingerprint reader/writer





http://code.google.com/p/chem-fingerprints/

NOVARTIS