

The Schrödinger KNIME extensions

Computational Chemistry and Cheminformatics in
a workflow environment

Jean-Christophe Mozziconacci

Volker Eyrich

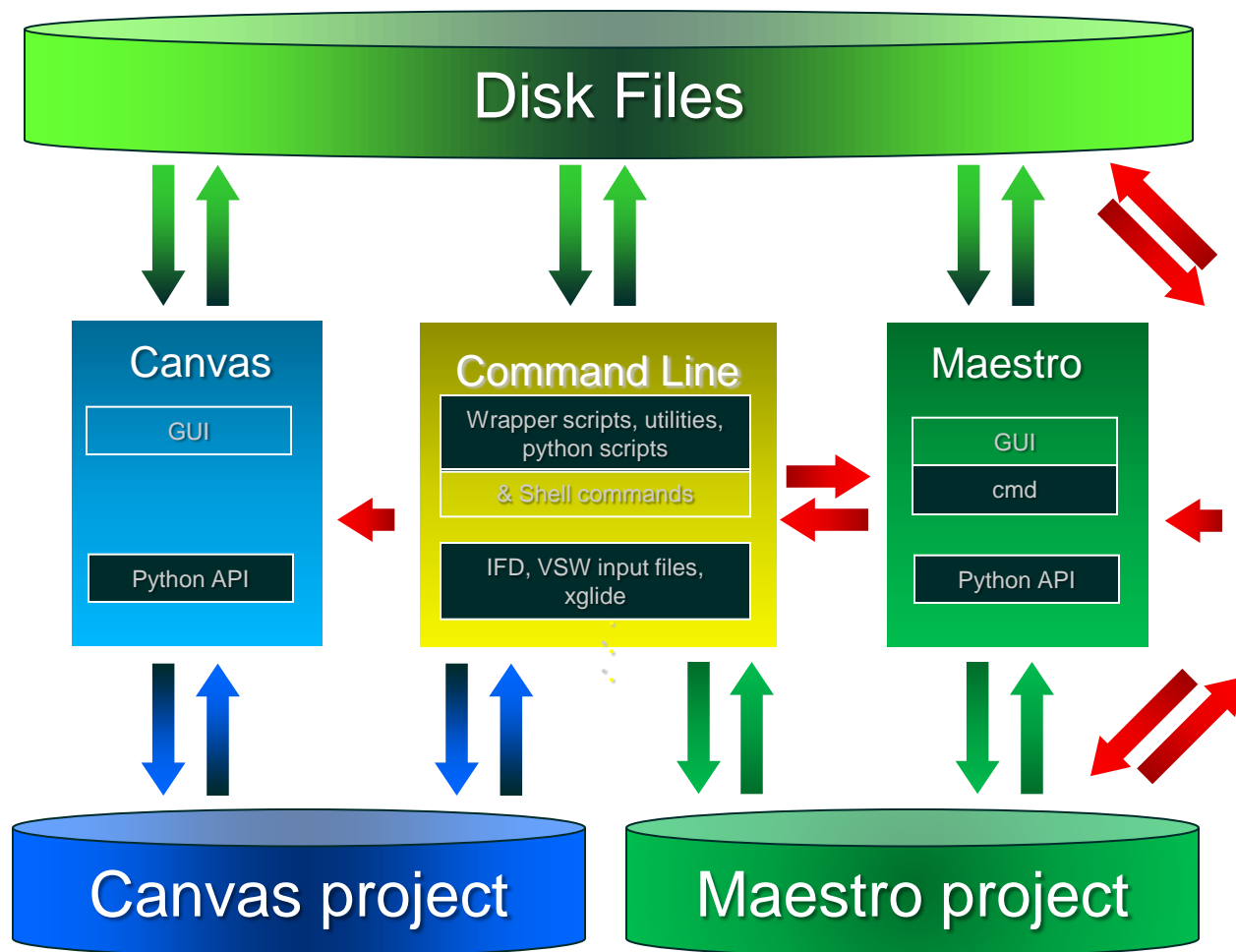


SCHRÖDINGER.

Topics

- Integration options
- What are the Schrödinger extensions?
- Workflow examples
- Roadmap

Integration options



Knime

- Various tools
- Third party codes (e.g. statistics, data-mining, visualization, scientific modules from other vendors)
- Databases (e.g. Oracle-based chemistry cartridges)
- Web services integration and web-based interfaces
- Prototyping
- Share, maintenance

What are the Schrödinger extensions?

Access to Schrödinger tools via KNIME

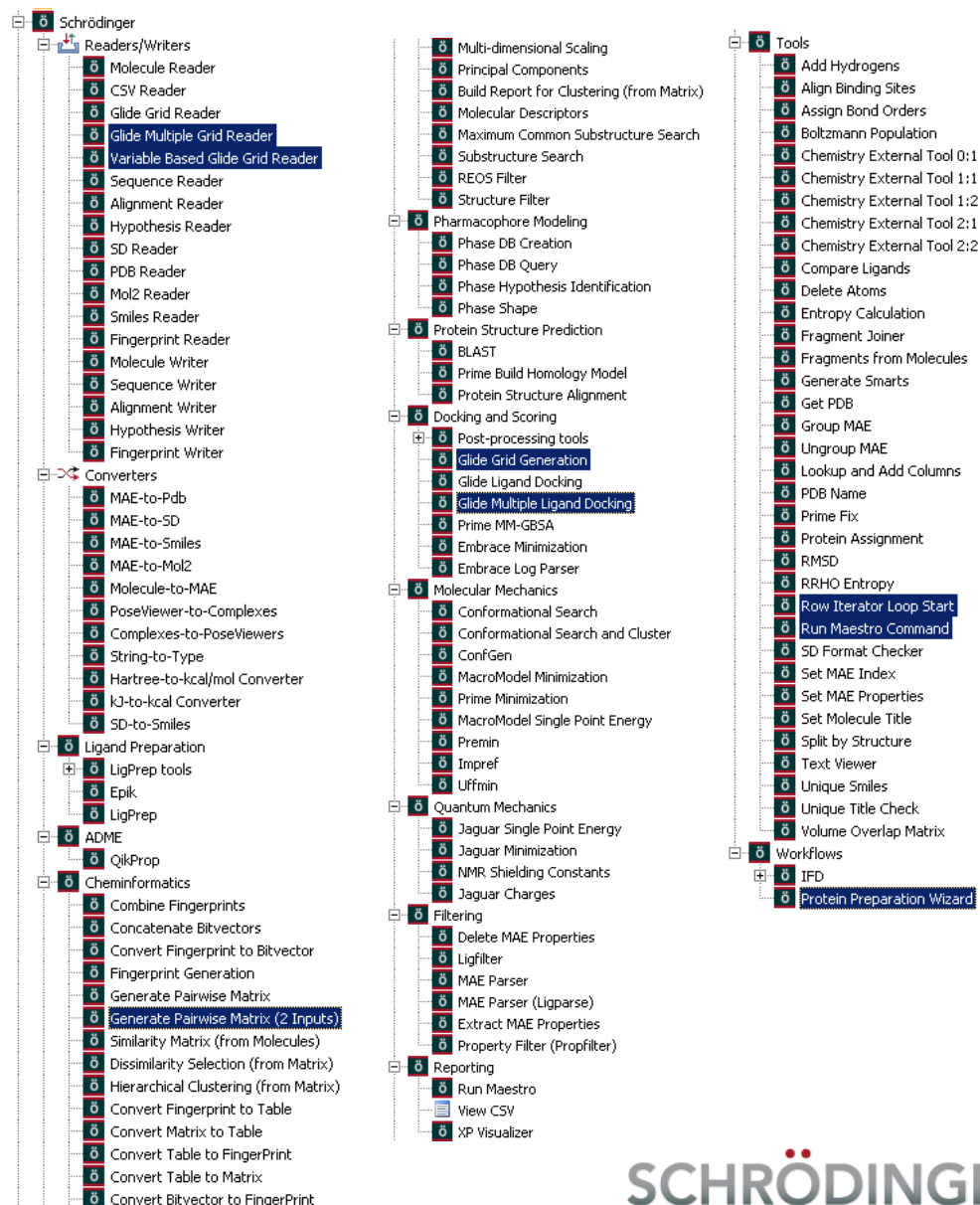
- Modeling and computational chemistry in a workflow environment
- Graphical assembly of workflows
- Generate, manipulate, analyze and visualize chemical data
- Interactive and automated analysis of datasets
- Presentation and communication of results

Access to Schrödinger tools via KNIME

Modeling and computational chemistry in a workflow environment

~140 nodes

- Ligand/protein preparation
- LBDD (cheminformatics, pharmacophore modeling)
- SBDD (Protein structure prediction, docking and scoring)
- Molecular and quantum mechanics



Structure and data manipulation

- **Generate, manipulate, analyze and visualize chemical data**
- **Interactive and automated analysis of datasets**
- **Presentation and communication of results**

KNIME nodes

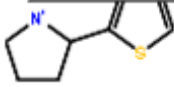
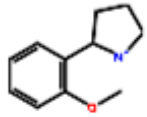
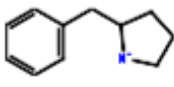
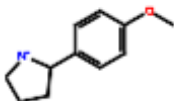

- Readers/Writers, converters, SD format checker,
- Filtering (Ligfilter, MAE parser, Compare ligands)
- Reporting (Text viewer, CSV viewer: OpenOffice/Excel, Run Maestro)
- Grouped structures (Group and Ungroup, Row iterator loop start)
- Structure and data manipulation (RMSD, Look up and add columns, Extract and Set MAE properties, Set MAE index, Split by structure, Delete atoms, Unique smiles, Unique title check, Align binding sites, Volume overlap matrix...)
- Scripting (Run Maestro commands, Python node, Chemistry external tool)

Canvas 2D view and use in the Report designer

Molecules in Maestro format - 5:41 - Mole...

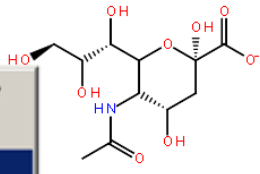
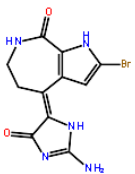
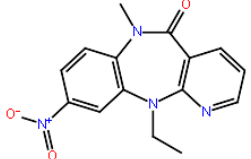
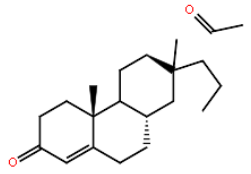
File

Table "default" - Rows: 1347 | Spec - Column: 1 | Properties

Row ID	CT
Row1	
Row2	
Row3	
Row4	
Row5	

Available Renderers ▶

- Maestro Molecule Summary
- Maestro File
- Canvas 2D
- String

BIRT Report Viewer				
Ligand	Complex	Ligand origin	Binding site	Mutations
	1a4q	1nsc	ligand A:DPC2	No
	1buh	1dm2	residues A:18,A:80,A:314	No
	1c1c	1rth	ligand Z:UNK999	A:102,A:227,A:234,B:102,B:234
	1dba	1dbb	ligand _:STR1	H:104

KNIME, Maestro and Canvas

- Launch Maestro or Canvas from KNIME
 - Interactive visualization
 - Run Maestro command scripts
- Launch KNIME from Maestro
 - Through a simple python script
 - Improvements under development
- ‘Visual coherence’ for application GUIs/configuration dialogs

Configuration panel layout – Maestro vs. KNIME

Read in
Ligands

Smiles Reader



Read 2D Ligands

LigPrep

Use structures from: **File**

File name: **Browse...**

Filter criteria file: **Create...** **Browse...**

Force field: **OPLS_2005**

Ionization:

☐ Retain original state

☐ Neutralize (best for QikProp)

☒ Generate possible states at target pH: +/-

Using:

☒ Ionizer

☐ Epik

☐ Add metal binding states

☒ Desalt ☒ Generate tautomers

Stereoisomers

Computation:

☒ Retain specified chiralities (vary other chiral centers)

☐ Determine chiralities from 3D structure

☐ Generate all combinations

Generate at most: per ligand

Generate low energy ring conformations: per ligand

Output format: ☒ Maestro ☐ SDF

Start... **Close** **Help**

Dialog - 2:3 - LigPrep (2D -> 3D)

File

Ligprep **Job control** **Flow Variables** **General Node Settings**

Column containing input: **SMI Smiles**

Output Column Structure

☐ Input plus Output

☒ Output replaces Input

☐ Output only

Match Input ☒

Force field: **OPLS_2005**

Ionization:

☐ Retain original state

☐ Neutralize (best for QikProp)

☒ Generate possible states at target pH: +/-

Using:

☒ Ionizer

☐ Epik

☐ Add metal binding states

Desalt ☒ Generate tautomers ☒

Stereoisomers

Computation:

☒ Retain specified chiralities (vary other chiral centers)

☐ Determine chiralities from 3D structure

☐ Generate all combinations

Generate at most: per ligand

Generate low energy ring conformations: per ligand

Output Type: **Maestro**

OK **Apply** **Cancel**

KNIME workflow page <http://www.schrodinger.com/knimeworkflows>

Select Workflows

Listed below are example KNIME Workflows that utilize many of the Schrödinger KNIME Extensions (Nodes) as well as many other built-in tools.

- [Cheminformatics](#)
- [Docking / Docking Post-processing](#)
- [Molecular Mechanics](#)
- [Homology Modeling](#)
- [Pharmacophore Modelling](#)
- [KNIME Workbench](#)
- [General Tools](#)

Check the boxes next to the workflows you wish to download, read the Important Notes, then [proceed below](#) to check the Important Notes checkbox and click the Download button.

Note: If you are not already logged in to the site, your selection will be saved and you will be prompted to log in. If you do not already have a logon account, please [click here](#) to request one.

Select All

Select Workflows Updated in the Last

30

Days

Reset

Cheminformatics

Select	Description	Author	Updated
<input type="checkbox"/>	Cluster by Fingerprint: Cluster structures via binary fingerprints and inspect the clustering statistics to choose a good number of clusters. Generate the clusters, inspect the corresponding list of structures and the clusters' representative molecules. [List of nodes: File Reader, Column Filter, Line Plot, Generate Pairwise Matrix, Row Filter, Build Report for Clustering _from Matrix_, RowID, Fingerprint Generation]	Schrödinger	09/08/2009
<input type="checkbox"/>	Maximum Common Substructure Search: Create all possible MCS groups and list the groups identified. Inspect the compounds in the first group and list the compounds that aren't included in this group. List the MCS groups that contain a compound of interest. Create all the groups with the option limiting each compound to at most one		09/08/2009

General Tools

Select	Description	Author	Updated
<input type="checkbox"/>	Ensure molecule title uniqueness: Ensure unique molecule titles (using the KNIME RowID) in the context of Canvas. [List of nodes: Set Molecule Title, RowID, Similarity Matrix _from Molecules_, Molecule Reader, Extract MAE Properties]	Schrödinger	09/08/2009

 [Back to top](#)

Important Notes

NOTICE: THE WORKFLOWS ON THIS WEB SITE HAVE BEEN GENERATED BY SCHRÖDINGER OR BY USERS OF SCHRÖDINGER SOFTWARE. THE WORKFLOWS ARE PROVIDED FREE OF CHARGE. THE WORKFLOWS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, AND MAY NOT BE ERROR FREE. SCHRÖDINGER AND THE AUTHORS OF THESE WORKFLOWS DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL SCHRÖDINGER OR THE AUTHORS OF THESE WORKFLOWS BE LIABLE FOR SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES OF ANY KIND.

☒ Include optional example data in the download.

☐ I have read the [Important Notes](#) above. You must check this box before downloading.

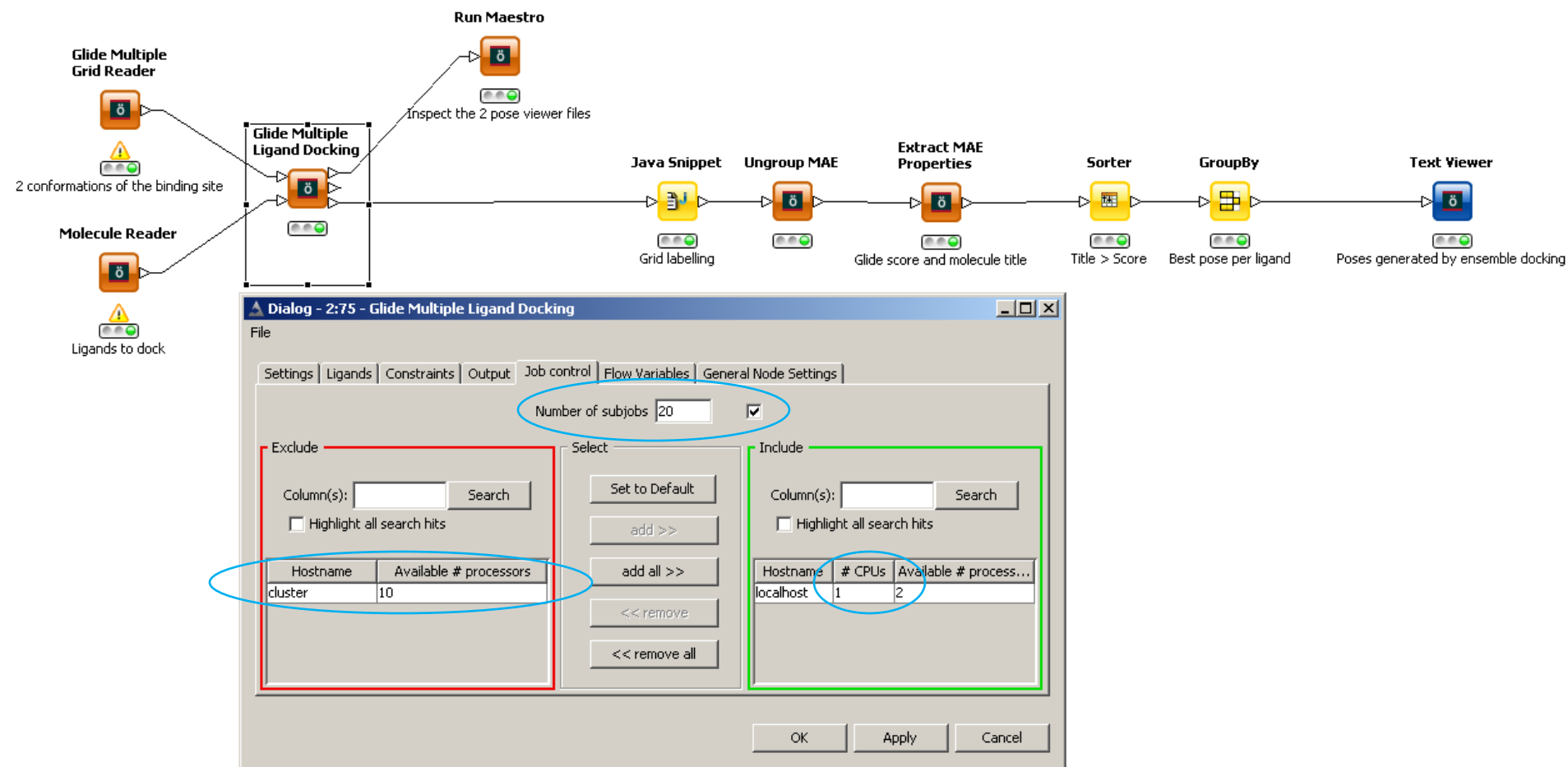
Download

Reset

Select All

SCHRÖDINGER.

Ensemble docking



Similarity search

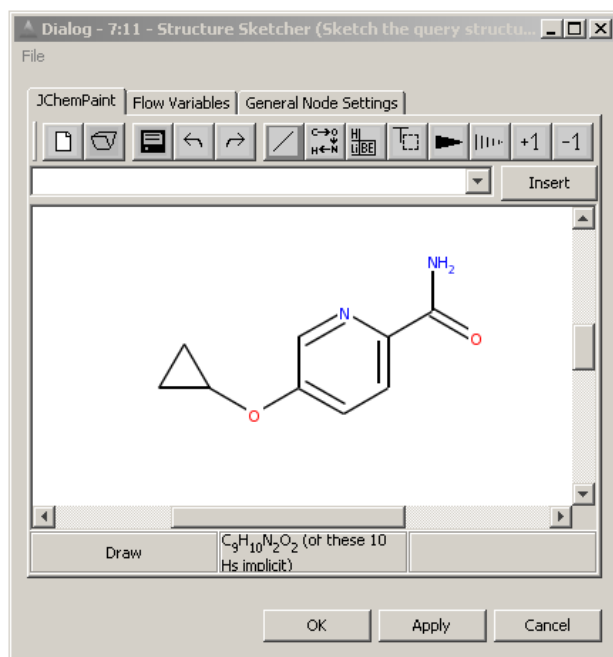
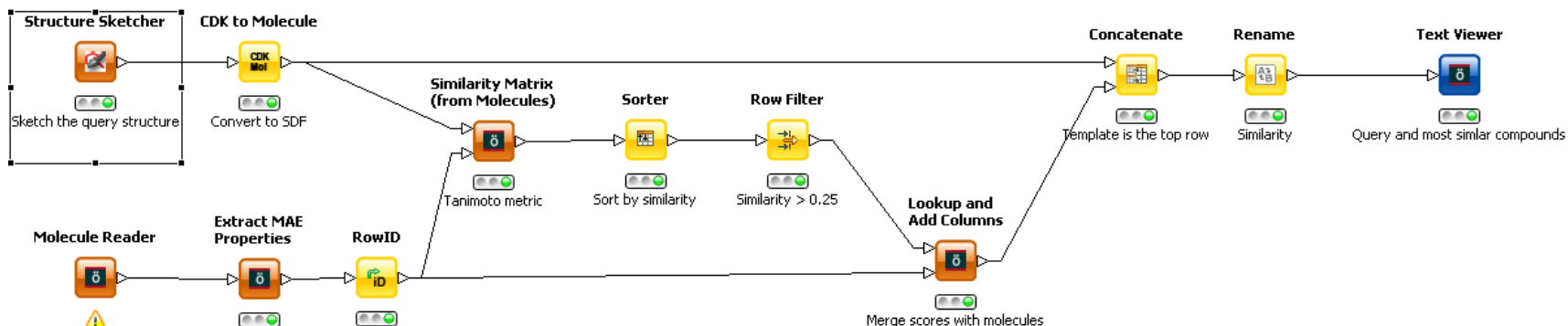
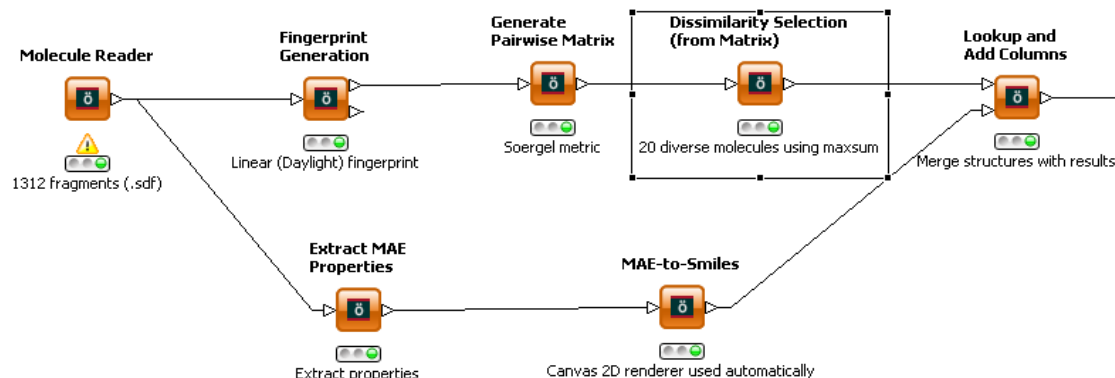


Table View - 7:38 - Text Viewer (Query and most similar compounds) (19 x 4)

File Hilite Navigation View Output

Row ID	SDF Query	S name	D similarity	CT
Structure_0		?	?	?
Row1	?	500015737	0.466	
Row2	?	500019047	0.39	
Row3	?	500024183	0.365	
Row4	?	500016211	0.353	
Row5	?	500031759	0.327	
Row6	?	500006085	0.303	

Diversity selection



Dialog - 7:76 - Dissimilarity Selection (fr...)

File

CanvasDBCS | Flow Variables | General Node Settings

Column containing input :

Number of Subsets:

Distance Threshold:

Selection Method Type:

Metric Type:

Ignore any scaled Finger Print values ☐

Initialization Method:

Seed for random Number generation

OK Apply Cancel

Dialog - 7:59 - Lookup and Add Columns (Merge structur...)

File

LookupAndAdd | Flow Variables | General Node Settings

Key column :

Lookup column :

Include column :

Include One Column ☒

Include All Columns ☐

Add String to column names in Input 2 (if duplicate)

Use Hash (Keep all necessary data in memory) ☒

OK Apply Cancel

Out-Port name - 7:59 - Lookup and Add Columns (...)

File

Table "default" - Rows: 19 | Spec - Columns: 3 | Properties

Row ID		S_BestRe...	SMI Smiles
Row1	2	500006741	<chem>CC(C)C1CCC(CC1)CO</chem>
Row2	3	500032051	<chem>Clc1ccc2nc3cc(Br)ccc3nc21</chem>
Row3	4	500034435	<chem>CN1CCc2c(c1)nc(C#N)s2</chem>
Row4	5	500016575	<chem>CC1(C)OC(=O)C1C(=O)O</chem>
Row5	6	500020093	<chem>C1CC2(C1)CCNCC2</chem>
Row6	7	500032675	<chem>Nc1ccccc1-n1ccncc1</chem>
Row7	8	500021417	<chem>ClCc1nc2ccccc2o1</chem>
Row8	9	500022357	<chem>CC1(C)OC(=O)C1C(=O)O</chem>

Maximum Common Substructure Search

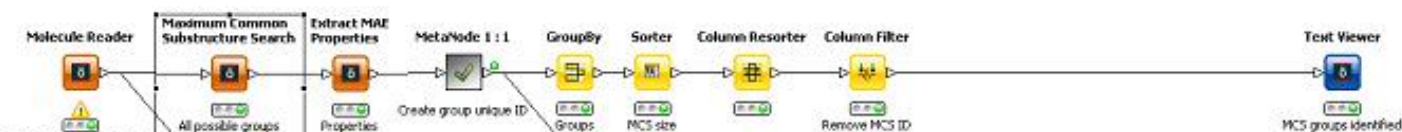


Table View - 10:46 - Text View...

Row ID	sm s_canvas_MCS_SMA...	Count
Row5		26
Row5		23
Row3		21
Row4		21
Row2		19
Row1		18
Row0		14
Row0		9

Two columns, group name and count

Row ID	MCS_ID	CT
Row1	26_1	#CTs: 9
Row2	23_1	#CTs: 10
Row3	21_1	#CTs: 11
Row4	21_2	#CTs: 11
Row5	19_1	#CTs: 12
Row6	18_1	#CTs: 16
Row7	14_1	#CTs: 19
Row8	9_1	#CTs: 20
Row9	7_1	#CTs: 21

Table View - 10:52:21 - Text Viewer (Inspect the first group members) (9 x 4)

Row ID	MCS_ID	sm Smiles	Title	sm s_canvas_M
Row1	26_1		4785	
Row2	26_1		4810	
Row3	26_1		16782	
Row4	26_1		16796	

Dialog - 10:01:53 - Maximum Common Substructure Search (All possible groups)

CanvasMCS | Flow Variables | General Node Settings

Column containing molecules:

Column containing title for smiles:

Minimum number of molecules that must match MCS:

Maximum number of molecules that must match MCS:

Stop searching when atom and bond count go below:

Maximum number of molecules to process:

Marking atoms in output SMARTS: ☐

Restrict compound membership to at most a single MCS group: ☐

Atom/bond typing

7: Atomic number and bond order

8: Atoms distinguished by ring size, aromaticity, HB acceptor/donor, ionization potential, whether benzy

9: Carhart atom types (atom-pairs approach), all bonds equivalent

10: Daylight invariant atom types; bonds distinguished by bond order

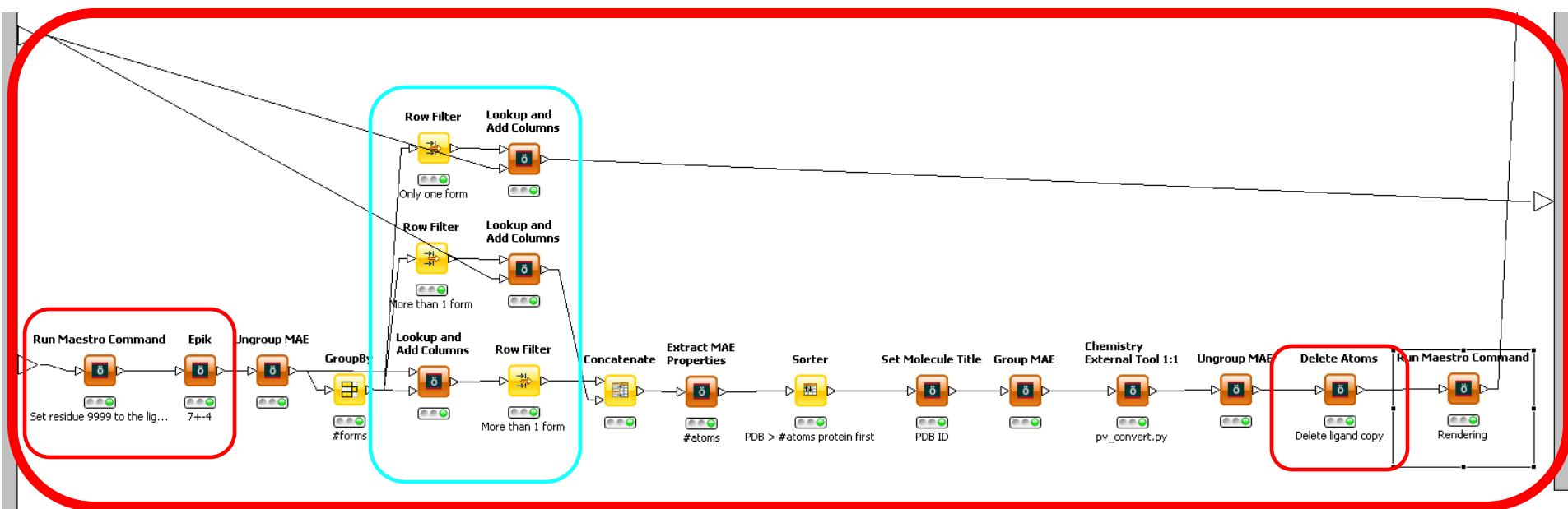
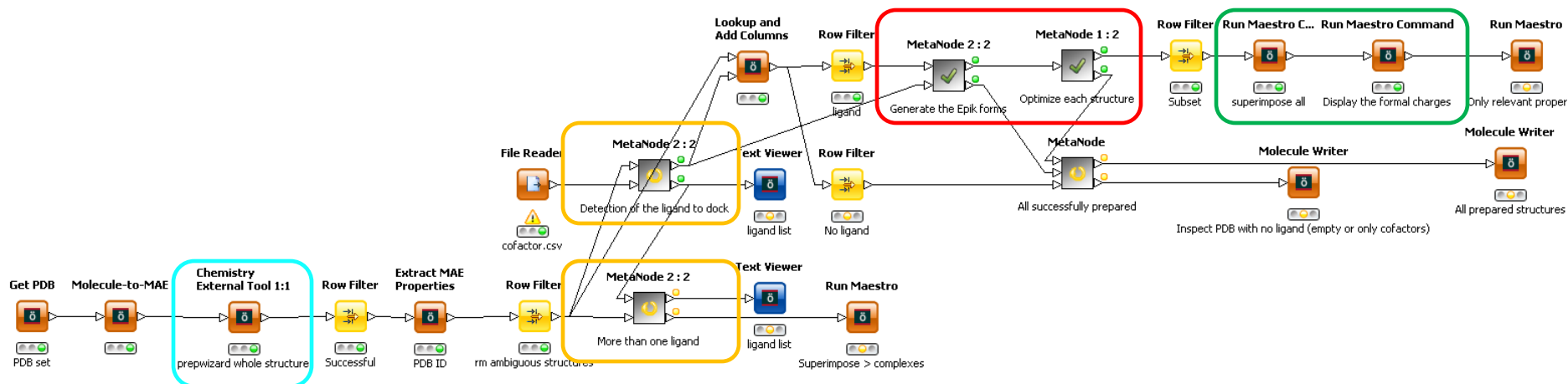
11: Atomic number and aromaticity; bond order

Atom Type Definition File (only with Custom Atom Type): Browse...

Output Type:

OK Apply Cancel

Automatic protein preparation



Roadmap

- Expose more back-end functionality through KNIME
Canvas model building, Macromodel coordinate scan, CombiGlide, Desmond ...
- Usability
- Stability and robustness
- Provide better integration with Maestro/Canvas
- Better integration with KNIME Enterprise products

Schrodinger KNIME extensions

Jean-Christophe

KNIME Associate Product Manager

jcmozzic@schrodinger.com

Volker Eyrich

VP of Technology

volker@schrodinger.com

Blaine Bell

Developer

Hina Patel

Developer

Schrödinger Developers

Woody Sherman

VP of Applications Science

Applications scientists



SCHRÖDINGER.



Why KNIME?

- Leading open-source ‘pipelining/workflow’ tool
- Extensible
- Cross-platform
- Vendor neutral
- Computational/Workflow model
- Freely available to academic and industrial researchers
- Commercial development and support
- Enterprise functionality available through [KNIME.com](https://www.knime.com)

Timeline & Resources

- Timeline
 - Q3 2006: Started evaluation of core platform
 - Q3 2007: First release of Schrödinger extensions (version 1.0)
 - Q2 2008: Schrödinger Release 1.1 and workflow examples
 - Q2 2009: Schrödinger Release 1.2 and more examples
 - Q2 2010: New release
- Resources
 - Two 2 FTEs for development since the beginning
 - QA and documentation resources
 - Internal applications

Access to Schrödinger tools via KNIME

Modeling and computational chemistry in a workflow environment

~140 nodes

- Ligand preparation
- ADME
- Cheminformatics
- Pharmacophore modeling
- Protein structure prediction
- Docking and scoring
- Molecular mechanics
- Quantum mechanics
- Workflows

