

The Schrödinger KNIME extensions

Computational Chemistry and Cheminformatics in
a workflow environment

Jean-Christophe Mozziconacci

Volker Eyrich



SCHRÖDINGER.

What are the Schrödinger extensions?

- **Modelling and computational chemistry**
- **150+ nodes**
Linux and Windows 32 and 64 bit

Molecular Mechanics – Macromodel

Molecular Dynamics – Desmond

Quantum Mechanics – Jaguar

Cheminformatics – Canvas

Pharmacophore modeling – Phase

Combinatorial Libraries – Combiglide

Docking – Glide

Protein Structure Prediction – Prime,
IFD

Protein preparation

Ligand preparation – LigPrep, Epik

Property generation – Qikprop ...

Filtering

Tools for data and structure manipulation

Scripting – shell, Python

Reporting

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

Cheminformatics

- **Substructure Search**
- Clustering, diversity selection, similarity search
- **Database analysis , MCS**

Docking and post-processing

- **Protein preparation and Glide grid generation**
- Docking and scoring, Virtual screening, Ensemble docking
- **Loop over docking parameters**
- **Validate docking parameters**

Pharmacophore modeling

- **Phase Shape screening**
- Phase hypothesis identification
- **Phase database screening** ◀

Molecular Mechanics

- **Compare conformational search methods**

Quantum mechanics

- **Conformational search and QM optimization**
(using the Report designer)

Library design

- Library enumeration

Homology modeling

- **Model building and refinement**
- Induced Fit Docking

Real World Examples

- Vendor database preparation

Metanodes

- **Run Maestro 1:1**
- **SiteMap**

General tools

- Python script node use-cases
including a Run PyMOL prototype
- Chemistry external tool node use-cases
- Run maestro command node use-cases
- Output column structure option philosophy

KNIME workbench

- **Workflows in the current workspace**
- GroupBy Use-cases

KNIME workbench- looper

- Group Looper, Unpivot

Simplest, **most exciting**, **new** and improved ◀ workflows

New features

- **Infrastructure**

- **Mac** port
- Ease of use: new **start-up script options** (including one to run stand-alone KNIME installation)
improved console log, faster start-up

- **KNIME menu in Maestro**

- **Connect to KNIME** mode to exchange structures with a KNIME session
- Build, import, edit and **run workflows from Maestro** on project table data. Dynamically generated GUI to alter some parameters.

Disconnected calculations and progress tracking in Jobcontrol.

Workflow tools and examples available on the Workflow web page:

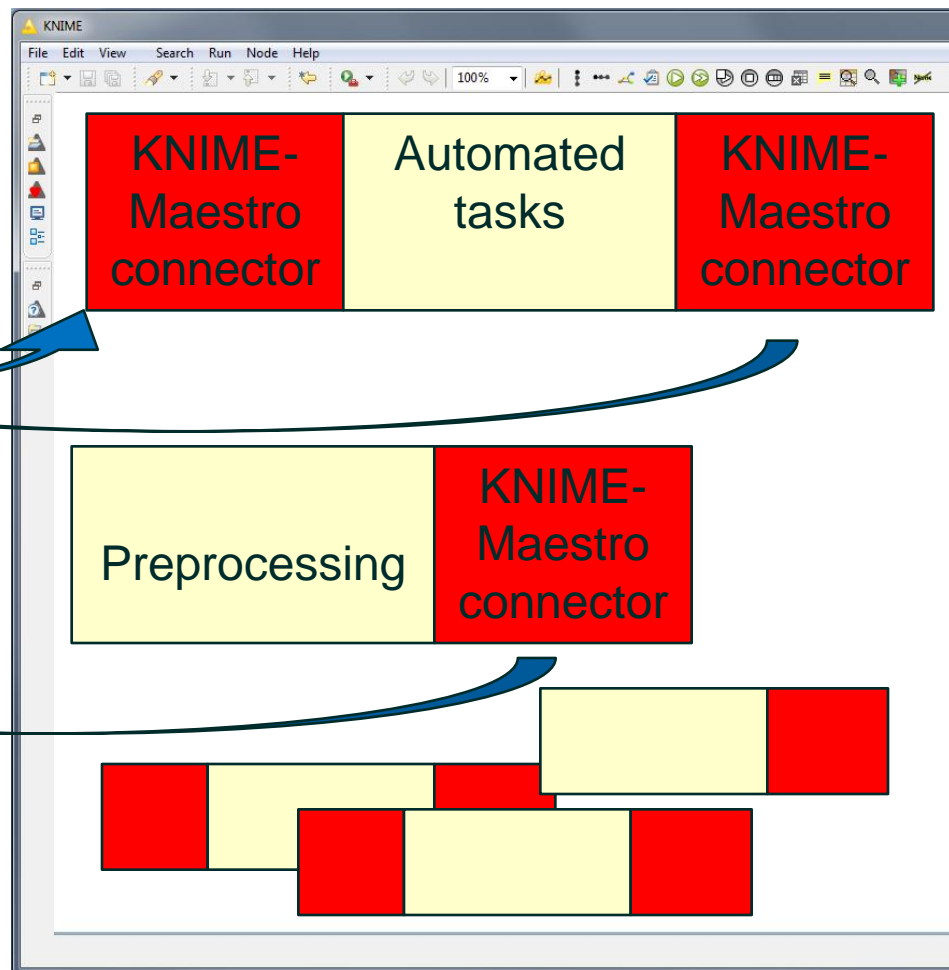
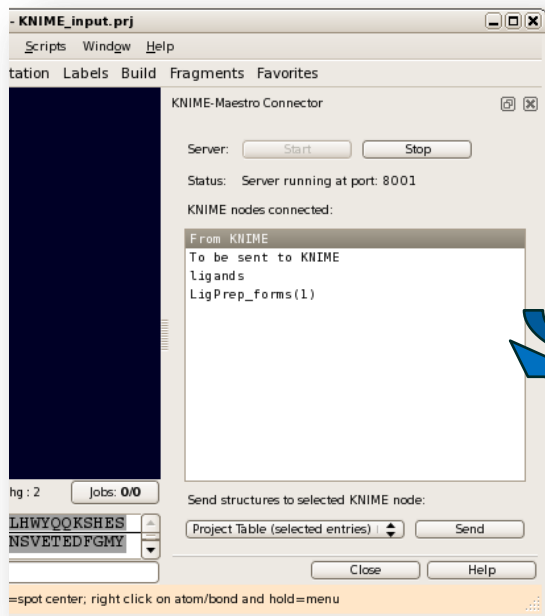
- Protein RMSD matrix, PDB list preparation, Align multimer chains
- Internal similarity, Vendor database filtering
- ...

KNIME – Maestro connection

Maestro
Visual inspection, selections ...

Server

KNIME
used interactively



Project table
Store the results

The screenshot shows the 'Project Table' window with a table of results. The table has columns for 'Row', 'Stars', 'In', 'Title', and 'docking score'. The data is as follows:

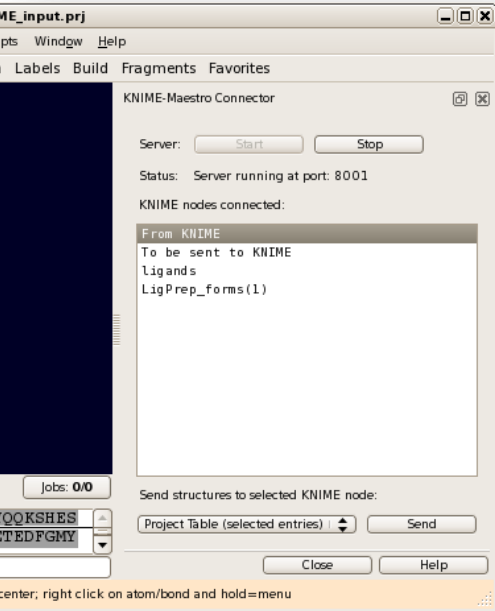
Row	Stars	In	Title	docking score
[3]		-	[1] - Crystal structures	
1	☆☆☆☆	☑	1AC8	
2	☆☆☆☆	☑	1AET	
3	☆☆☆☆	☑	1AC4	
[8]		+	[2] - Known active	
[17]		+	[3] - Inactive compounds	
[54]		+	[4] - Ligprep forms	
[55]		-	[5] - MaestroCmd_34402818517213_	

KNIME – Maestro connection

Maestro



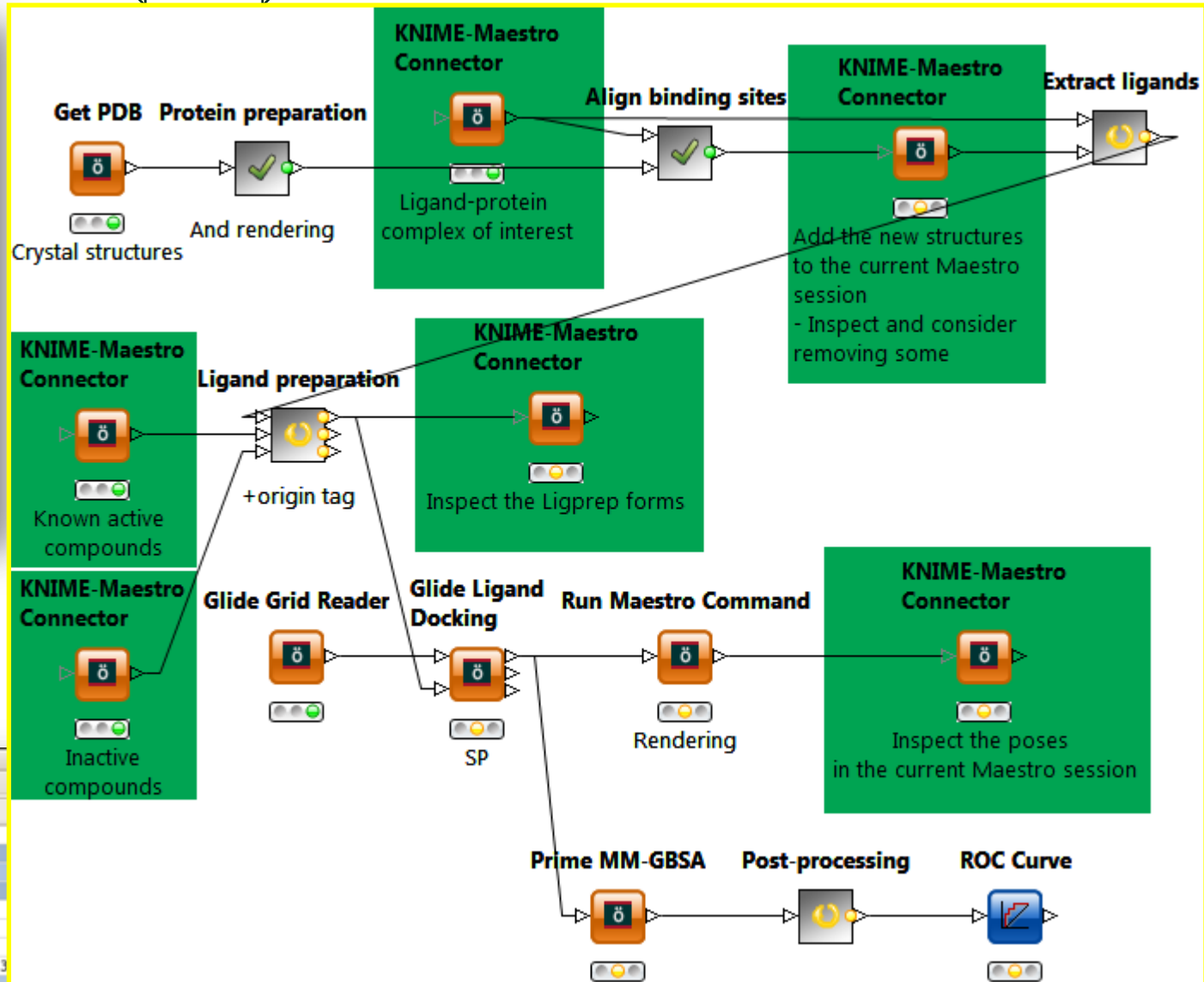
KNIME



Project table

Project Table — Scratch Project

Row	Stars	In	Title
[3]		-	[1] - Crystal structures
1	☆☆☆☆	<input type="checkbox"/>	1AC8
2	☆☆☆☆	<input type="checkbox"/>	1AET
3	☆☆☆☆	<input type="checkbox"/>	1AC4
[8]		+	[2] - Known active
[17]		+	[3] - Inactive compounds
[54]		+	[4] - Ligprep forms
[55]		-	[5] - MaestroCmd_34402818517213



Run workflows from Maestro

The image shows the Maestro software interface with the 'Workflows' menu open. The menu options include 'Connect to KNIME...', 'New...', 'Manage...', 'Install...', 'Update...', 'Install examples', 'Vendor database filtering...', 'Cheminformatics', 'Filtering', 'General tools', and 'PDB list preparation...'. The 'PDB list preparation...' option is highlighted, and a sub-menu is visible with the following items: 'KNIME', 'Protein Preparation Wizard...', 'Virtual Screening Workflow...', 'Induced Fit Docking...', 'QM-Polarized Ligand Docking...', 'Ligand and Structure-Based Descriptors...', and 'P450 Site of Metabolism'. The main window displays a 3D molecular model of a protein-ligand complex. The status bar at the bottom shows 'Jobs: 0/0', 'Atoms: 0/3511/5720', 'Entries: 1/1900', 'Res: 363', 'Chn: 1', 'Mol: 12', and 'Chg: -9'. The 'Commands' field contains the text: '9:1W84 A ERPTFYRQELNKTIWEVPERYQNLSPVGSAGYGSVCAAFDFTKTGLRVAVKKLSR' and '11723:Ligand: L12'. The 'PDB list preparation' workflow configuration panel is open on the right, showing the following details:

- Input:** None
- Output:**
 - New group
 - Maestro message Warning if the structures are prepared automatically
- Parameter:**
 - PDB code list - PDB codes: 1W84, 1W83, 1W7H, 10UY
 - PDB code list - Mode: Preprocess

There is an unchecked checkbox for 'Save after execution'. The 'Run', 'Cancel', and 'Help' buttons are visible at the bottom of the configuration panel.

Run workflows from Maestro

The image displays two software windows: Maestro and KNIME.

Maestro - KNIME_input.prj

- Menu: **Workflows** > **KNIME** > **PDB list preparation**
- Buttons: **Open in KNIME...** (circled in red), **Create a Copy**
- Text: **PDB list preparation**
Load a list of PDB codes and preprocess or prepare them automatically.
- Input:** None
- Output:**
 - New group
 - Maestro message Warning if the structures are prepared automatically
- Parameter:** (circled in red)
 - PDB code list - PDB codes: 1W84, 1W83, 1W7H, 10UY
 - PDB code list - Mode: Preprocess
- Checkbox: Save after execution
- Buttons: **Run**, **Cancel**, **Help**

KNIME

- Workflow: ***0: PDB_list_preparation** > ***0:25 - PDB code list**
- Nodes:
 - String Radio Buttons** (Mode GUIsettingCOMBO - circled in red)
 - Java Edit Variable** (Switch)
 - String Input** (PDB codes GUIsetting)
 - Get PDB**

New features

- **Node functionalities**

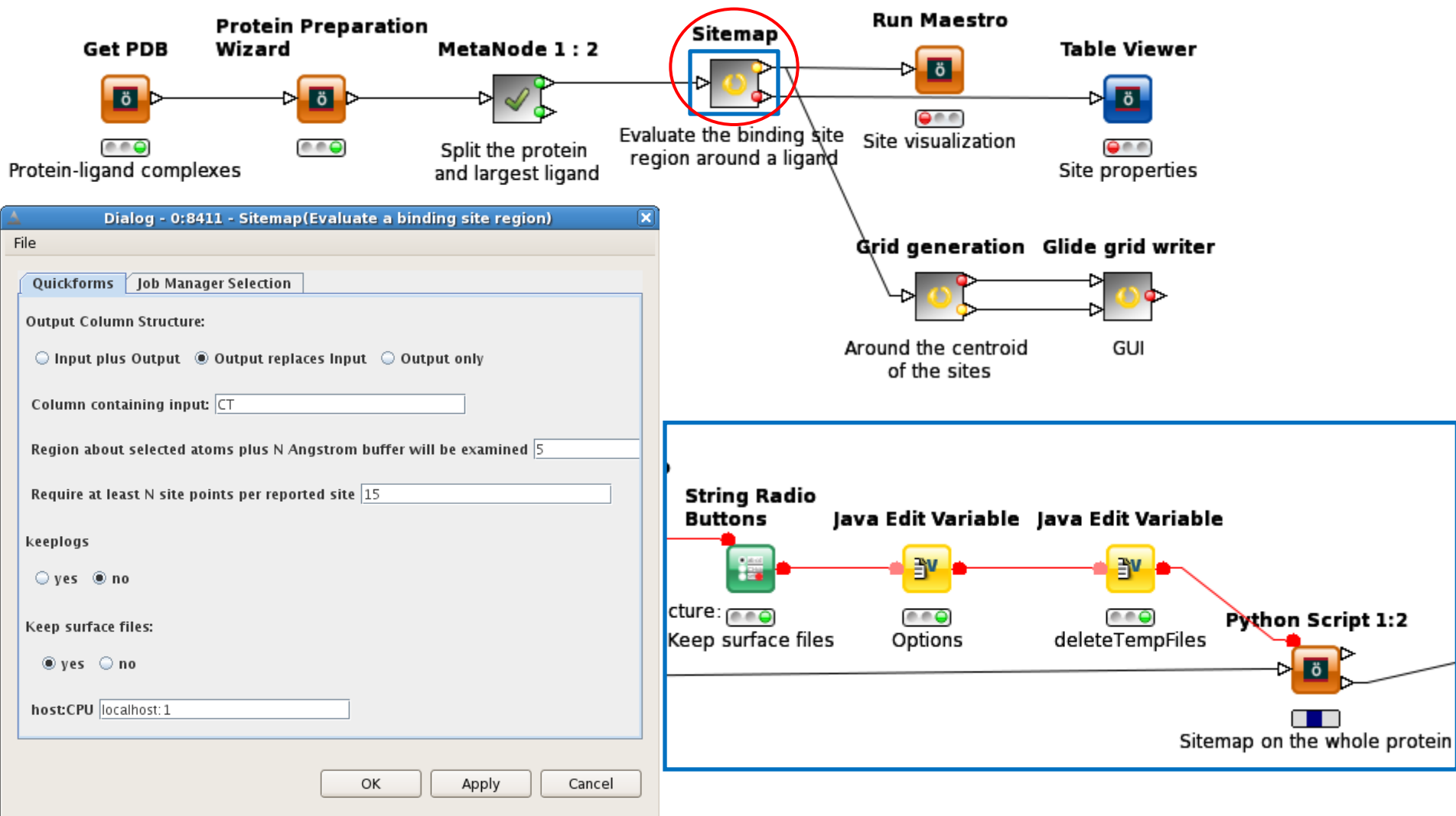
- Many fixes and minor improvements
- Column types: new **Surface type**. Phase hypothesis and Glide grid type improvement
- **Python script node** new functions (easier metanode development), access to Flow variables

- **Nodes/metanodes**

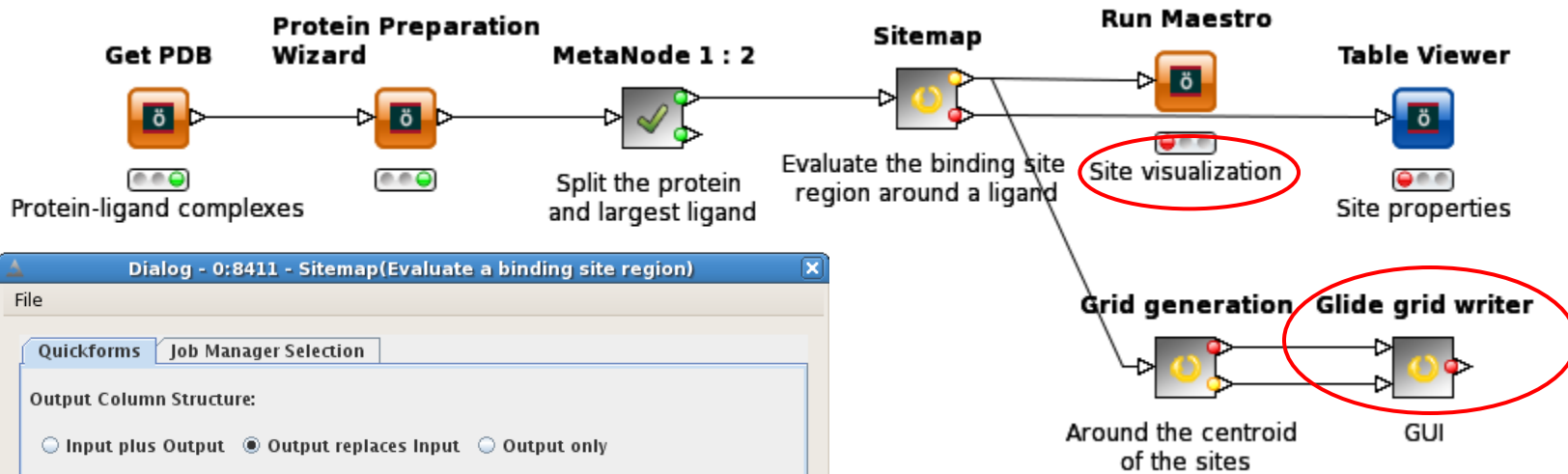
- Run Maestro 1:1,* Run PyMOL *
- Jaguar pKa, Jaguar ESP surface
- Sitemap *
- Glide grid writer
- Prime loop sampling, Prime energy

* first version already available on the Workflow page

Some new metanodes and surface cell type



Application example



Dialog - 0:8411 - Sitemap(Evaluate a binding site region)

File

Quickforms Job Manager Selection

Output Column Structure:

Input plus Output Output replaces Input Output only

Column containing input:

Region about selected atoms plus N Angstrom buffer will be examined

Require at least N site points per reported site

keeplogs

yes no

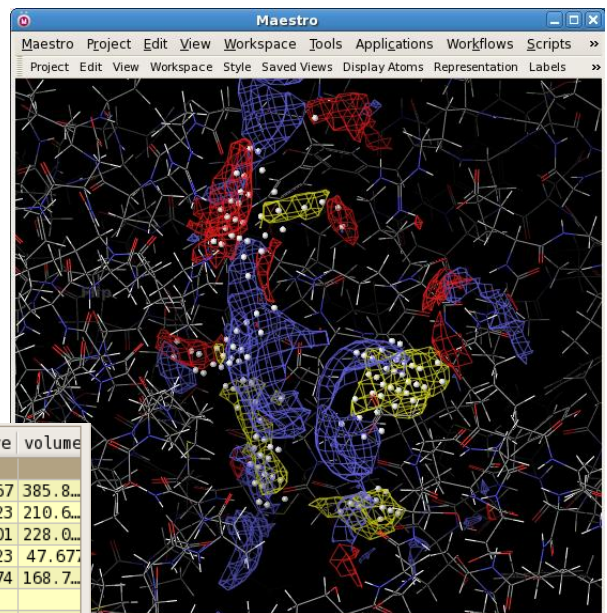
Keep surface files:

yes no

host:CPU

OK

Row	Stars	In	Title	Entry ID	SiteScore	size	Dscore	volume
[12]			[1] - internal M...					
1	☆☆☆☆	■	1YTK_site_1	S 1	1.027	125	1.067	385.8...
2	☆☆☆☆	□	1YTK_site_2	S 2	0.966	81	0.723	210.6...
3	☆☆☆☆	□	1YTK_site_3	S 3	0.956	73	1.001	228.0...
4	☆☆☆☆	□	1YTK_site_4	S 4	0.806	21	0.723	47.67...
5	☆☆☆☆	□	1YTK_site_5	S 5	0.772	56	0.774	168.7...
6	☆☆☆☆	■	1YTK	6				
7	☆☆☆☆	□	3DT1_site_1	S 7	1.169	125	1.238	289.8...
8	☆☆☆☆	□	3DT1_site_2	S 8	1.020	150	0.951	468.1...
9	☆☆☆☆	□	3DT1_site_3	S 9	0.958	230	0.989	694.5...
10	☆☆☆☆	□	3DT1_site_4	S 10	0.773	56	0.718	110.7...
11	☆☆☆☆	□	3DT1_site_5	S 11	0.683	34	0.637	117.3...
12	☆☆☆☆	□	3DT1	12				



Schrodinger KNIME extensions

Jean-Christophe

KNIME Associate Product Manager

jcmozzic@schrodinger.com

Volker Eyrich

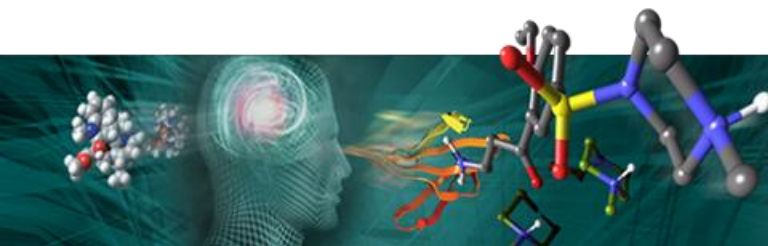
VP of Technology

volker@schrodinger.com

Ravikiran Kuppuraj

Main developer

Schrödinger Developers



What are the Schrödinger extensions?

- **Modelling and computational chemistry in a workflow environment**
- **150+ nodes covering the whole Schrodinger Suite**
- **Run on Linux and Windows 32 and 64 bit versions**

Molecular Mechanics

- MacroModel Single Point Energy, Minimization, Coordinate Scan
- ConfGen
- Conformational Search
- Premin, Impref, Uffmin

Quantum Mechanics

- Jaguar Single Point Energy, Minimization
- NMR Shielding Constants
- Jaguar Charges

Molecular Dynamics

- Desmond System builder
- Desmond Molecular Dynamics
- Trajectory extract frames and manipulation
- Trajectory reader, CMS reader

Cheminformatics

Fingerprint Based Tools

- Fingerprint Generation
- Generate Pairwise Matrix, and 2 Inputs
- Similarity Matrix, Dissimilarity Selection
- Build Report and Hierarchical Clustering

Filters and Mining Tools

- MCS
- Substructure Search
- Structure, REOS Filters

Utilities and Converters

- PCA, MDS
- Combine Fingerprints, Concatenate Bitvectors, Convert Fingerprint to Bitvector, etc

Modeling

- Bayes Model Building, prediction
- PLS Model Building, Prediction

Pharmacophore Modeling

- Phase Shape
- Phase DB Query, File Query
- Phase DB Creation
- Phase Hypothesis Identification

Combinatorial Libraries

- CombiGlide Reagent Preparation and Library Enumeration
- Fragments from Molecules and joiner

Docking and Scoring

- Glide Grid Generation
- Glide Ligand Docking
- Glide Multiple Ligand Docking
- XP Visualizer

Post-processing

- Prime MM-GBSA
- Embrace Minimization
- Strain Rescore, Pose Entropy
- Pose Filter, Glide Merge, Sort Results
- Glide Ensemble Merge

Protein Structure Prediction

- BLAST
- Prime Build Homology Model
- Prime Side Chain Sampling, Minimization
- IFD and individual steps

Schrödinger nodes

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

Protein Preparation

- Protein Preparation Wizard
- Protein Assignment

Ligand Preparation

- LigPrep and the individual steps
- Epik

Property Generation

- QikProp, Molecular Descriptors
- Canvas descriptors

Filtering

- Ligfilter, Ligparse, Propfilter

Scripting

- Run Maestro Command
- Chemistry External nodes
- Python Script nodes

Reporting

- Run Maestro, Run Canvas
- Spreadsheet viewer
- Table viewer

Tools

Data Manipulation

- Compare Ligands
- Lookup and Add Columns
- Group and Ungroup MAE

Structure Manipulation

- Add Hydrogens
- Delete Atoms
- Split by Structure
- MAE atom/bond property Parser
- Extract, Set and Delete MAE Properties
- Set Molecule Title, MAE Index

Utilities

- Get PDB
- Align Binding Sites
- Protein Structure Alignment
- RMSD

- Assign Bond Orders
- Unique Title Check
- Check PDB Name
- SD Format Checker
- Generate Smarts, Unique Smiles
- RRHO Entropy
- Boltzmann Population
- Volume Overlap Matrix

Readers/Writers

- CSV Reader (several inputs)
- Molecule Reader and Writer
- SD, PDB, Mol2 Reader nodes
- Sequence, Alignment Readers and Writers
- Fingerprint Reader and Writer
- Hypothesis Reader and Writer
- Glide Grid and Multiple Grid Reader
- Variable Based Glide Grid Reader

Converters

- String-to-Type
- Molecule-to-MAE, MAE-to-Pdb, to-SD, to-Smiles and to-Mol2, SD-to-Smiles
- PoseViewer-to-Complexes and Complexes-to-PoseViewers
- Hartree-to-kcal/mol and kJ-to-kcal