The Schrödinger KNIME extensions

Computational Chemistry and Cheminformatics in a workflow environment

Jean-Christophe Mozziconacci
Volker Eyrich
Topics

- What are the Schrödinger extensions?
- Workflow application examples
- 2012 development plan: KNIME, Maestro, Canvas, PyMOL and Seurat integration
What are the Schrödinger extensions?

Access to Schrödinger tools via KNIME

- Modelling 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

- **Run on Linux and Windows**
  - ~150 nodes
- **Ligand/protein preparation**
- **LBDD (cheminformatics, pharmacophore modeling)**
- **SBDD (Protein structure prediction, docking and scoring)**
- **Molecular and quantum mechanics**

## Molecular Mechanics
- MacroModel Single Point Energy
- MacroModel Minimization
- MacroModel Coordinate Scan
- ConfGen Standard
- ConfGen
- Conformational Search
- Conformational Search and Cluster
- Premin
- Impref
- Uffmin

## Quantum Mechanics
- Jaguar Single Point Energy
- Jaguar Minimization
- NMR Shielding Constants
- Jaguar Charges

## Cheminformatics
### Fingerprint Based Tools
- Fingerprint Generation
- Generate Pairwise Matrix
- Generate Pairwise Matrix (2 Inputs)
- Similarity Matrix (from Molecules)
- Dissimilarity Selection (from Matrix)
- Build Report for Clustering (from Matrix)
- Hierarchical Clustering (from Matrix)

## Pharmacophore Modeling
- Phase Shape
- Phase DB Query
- Phase File Query
- Phase DB Creation
- Phase Hypothesis Identification

## Docking and Scoring
- Glide Grid Generation
- Glide Ligand Docking
- Glide Multiple Ligand Docking
- XP Visualizer

## Post-processing
- Prime MM-GBSA

## Protein Structure Prediction
- BLAST
- Prime Build Homology Model
- Prime Side Chain Sampling
- Prime Minimization

## Molecular and Quantum Mechanics
- Jaguar Single Point Energy
- Jaguar Minimization
- NMR Shielding Constants
- Jaguar Charges
Schrödinger nodes

Workflows
- Protein Preparation
  - Protein Preparation Wizard
  - Protein Assignment
  - Induced-fit docking
  - IFD and individual steps
- Ligand Preparation
  - LigPrep
  - Ligprep individual tools (Ionizer, Desalter, Neutralizer…)
  - Epik
- Property Generation
  - QikProp
  - Molecular Descriptors
  - Calculate properties
- Filtering
  - Ligfilter
  - Ligparse
  - Property Filter (Propfilter)

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

Report
- Run Maestro
- Run Canvas
- View CSV (open xls/ooffice)
- Text Viewer

Tools
- Combinatorial Libraries
  - CombiGlide Library Enumeration
  - CombiGlide Reagent Preparation
- Fragments
  - Fragment Joiner
  - Fragments from Molecules

Structure Manipulation
- Add Hydrogens
- Delete Atoms
- Split by Structure
- MAE Parser
  - Extract MAE Properties
  - Delete MAE Properties
  - Set MAE Properties
  - Set Molecule Title
  - Set MAE Index

Utilities
- Get PDB
- Align Binding Sites
- Protein Structure Alignment
- Prime Fix
- RMSD
  - Assign Bond Orders
  - Unique Title Check
  - PDB Name
  - SD Format Checker
  - Generate Smarts
  - Unique Smiles
  - Entropy Calculation
  - RRHO Entropy
  - Boltzmann Population
  - Volume Overlap Matrix

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

Scripting
- Run Maestro Command
- Chemistry External Tool
  - Maestro Command
  - Chemistry External Tool
  - Python Script
  - Maestro Command
  - Chemistry External Tool
  - Python Script

Readers/ Writers
- CSV Reader
- Molecule Reader
- SD, PDB, Mol2 Reader nodes
- Sequence Reader
- Alignment Reader
- Fingerprint Reader
- Hypothesis Reader
- Glide Grid Reader
- Variable Based Glide Grid Reader
- Molecule Writer
- Sequence Writer
- Alignment Writer
- Hypothesis Writer
- Fingerprint Writer

Converters
- Molecule-to-MAE
- MAE-to-Pdb, to-SD, to-Smiles and to-Mol2
- SD-to-Smiles
- PoseViewer-to-Complexes
- Complexes-to-PoseViewers
- String-to-Type
- Hartree-to-kcal/mol Converter
- kJ-to-kcal Converter
Schroedinger Suite 2011- some new features

Infrastructure
- Canvas **2D renderer on 64 bit** KNIME workbench, **Windows** installers
- New **start-up script options** to pass user/machine/OS-specific parameters (memory limit, temporary directory…)

New nodes
- **Desmond** nodes (System builder and Molecular dynamics, Trajectory manipulation, extract frames, Desmond trajectory and CMS Readers)
- **Phase query on files** (already possible on a Phase database)
- **Canvas model** building and prediction nodes (PLS and Bayes classification)
- **Run Canvas**
  - Filter structures based on an ASL expression, Calculate properties

Options and functionalities
- Prime Build homology model- include ligand and cofactors
- Run Maestro- import as groups, read Hypotheses, specify a project to which structures are added
- **Cheminformatics**
  - Cluster by fingerprint
  - Database analysis
  - Maximum Common Substructure
  - Select diverse molecules
  - Similarity search
  - Substructure Search

- **Docking and post-processing**
  - Protein preparation and Glide grid generation
  - Docking and scoring
  - Ensemble docking
  - Induced Fit Docking protocol
  - Loop over docking parameters
  - Validate docking parameters
  - Virtual screening

- **Pharmacophore modeling**
  - Phase Shape screening
  - Phase database screening
  - Phase hypothesis identification

- **Molecular Mechanics**
  - Compare conformational search methods

- **Quantum mechanics**
  - Conformational search and QM optimization

  Using the Report designer

- **Homology modeling**
  - Model building and refinement

- **KNIME workbench**
  - Group by use-cases

- **General tools**
  - Python script node use-cases
  - Chemistry external tool node use-cases
  - Ensure molecule title uniqueness
  - Group Looper
  - Output column structure option philosophy
  - Run maestro command node use-cases
  - Protein structure alignment

Simplest, most exciting, new and improved workflows
Vendor database preparation
import schrodinger.structureutil as structureutil
iterator = inData[0].iterator()
while iterator.hasNext():
    row = iterator.next()
    mae_ct = row.getCell(0)
    for st in mae_ct.getStructureReader():
        RingNumber = len(structureutil.find_rings(st))
        newIntCell1 = IntCell(RingNumber)
        newRow = AppendedColumnRow(row, [newIntCell1])
        newRow.colNames = ["ring_number"]
        outContainer[0].addRowToTable(newRow)
Vendor database preparation - filtering on various criteria
Database shape screening and ensemble docking

Prefiltering of the database to screen

Molecule Reader

Query structure (1FDS co-crystallized ligand)

66,200 structures to screen

Property filtering

Filtering at 0.7

MetaNode 1:1

Phase Shape

Inspect the alignment of the query and the compounds similar in shape

MetaNode 2:1

Sort by similarity

Property filtering

Query structure (1FDS co-crystallized ligand)

Molecule Reader

MetaNode 1:1

Phase Shape

Inspect the alignment of the query and the compounds similar in shape

MetaNode 2:1

Sort by similarity

Property filtering

Query structure (1FDS co-crystallized ligand)

Molecule Reader

Prefiltering of the database to screen

Molecule Reader

Extract MACS properties

pdb id

PDB ID

50 Factor Xa PDB structures

Glide Grid Generation

LigPrep

Glide Multiple Ligand Docking

Run Maestro

MetaNode 1:1

Best pose per ligand

Fosses generated by ensemble docking

Molecule Writer

MAE to Smiles

Data to Report

SCHRODINGER.
Conformational search and QM optimization

Isomethone conformational analysis

Conformer populations

<table>
<thead>
<tr>
<th>Conformer</th>
<th>Relative Energy</th>
<th>Ring conformation</th>
<th>Boltzmann population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conformer 0</td>
<td>6.0</td>
<td>-104.2</td>
<td>69.9</td>
</tr>
<tr>
<td>Conformer 1</td>
<td>11.0</td>
<td>-80.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Conformer 2</td>
<td>10.29</td>
<td>-68.9</td>
<td>1.1</td>
</tr>
<tr>
<td>Conformer 3</td>
<td>15.96</td>
<td>-18.4</td>
<td>0.1</td>
</tr>
<tr>
<td>Conformer 4</td>
<td>10.55</td>
<td>-18.3</td>
<td>1.0</td>
</tr>
<tr>
<td>Conformer 5</td>
<td>7.17</td>
<td>-11.1</td>
<td>12.1</td>
</tr>
<tr>
<td>Conformer 6</td>
<td>6.09</td>
<td>-1.3</td>
<td>2.5</td>
</tr>
<tr>
<td>Conformer 7</td>
<td>4.52</td>
<td>6.5</td>
<td>11.4</td>
</tr>
</tbody>
</table>

Jaguar

MacroModel

Reporting preparation

Populations

Aligned conformations

Rendering

Run Maestro Command

Run Maestro

Joiner

Concatenate

Best conformers and Boltzmann population

Best conformer and Boltzman population

MetaNode 1:2

Joiner

Test Viewer

Populations

Report preparation

Conformers and Chart Analysis
KNIME, Maestro, PyMOL, Canvas and Seurat integration

- Launch Maestro, Canvas and PyMOL from KNIME
  - Interactive visualization
  - Run Maestro command scripts
- Launch KNIME from Maestro
  - Through a simple python script
  - Dedicated menu in 2012
- Tighter integration in 2012 including Seurat
Schroedinger KNIME extensions

Jean-Christophe  KNIME Associate Product Manager
jcmozzic@schrodinger.com

Volker Eyrich  VP of Technology
volker@schrodinger.com

Schrödinger Developers

Woody Sherman  VP of Applications Science

Applications scientists
Conformational search - compare methods

30 diverse ligands from J. Chem. Inf. Model. 2010, 534