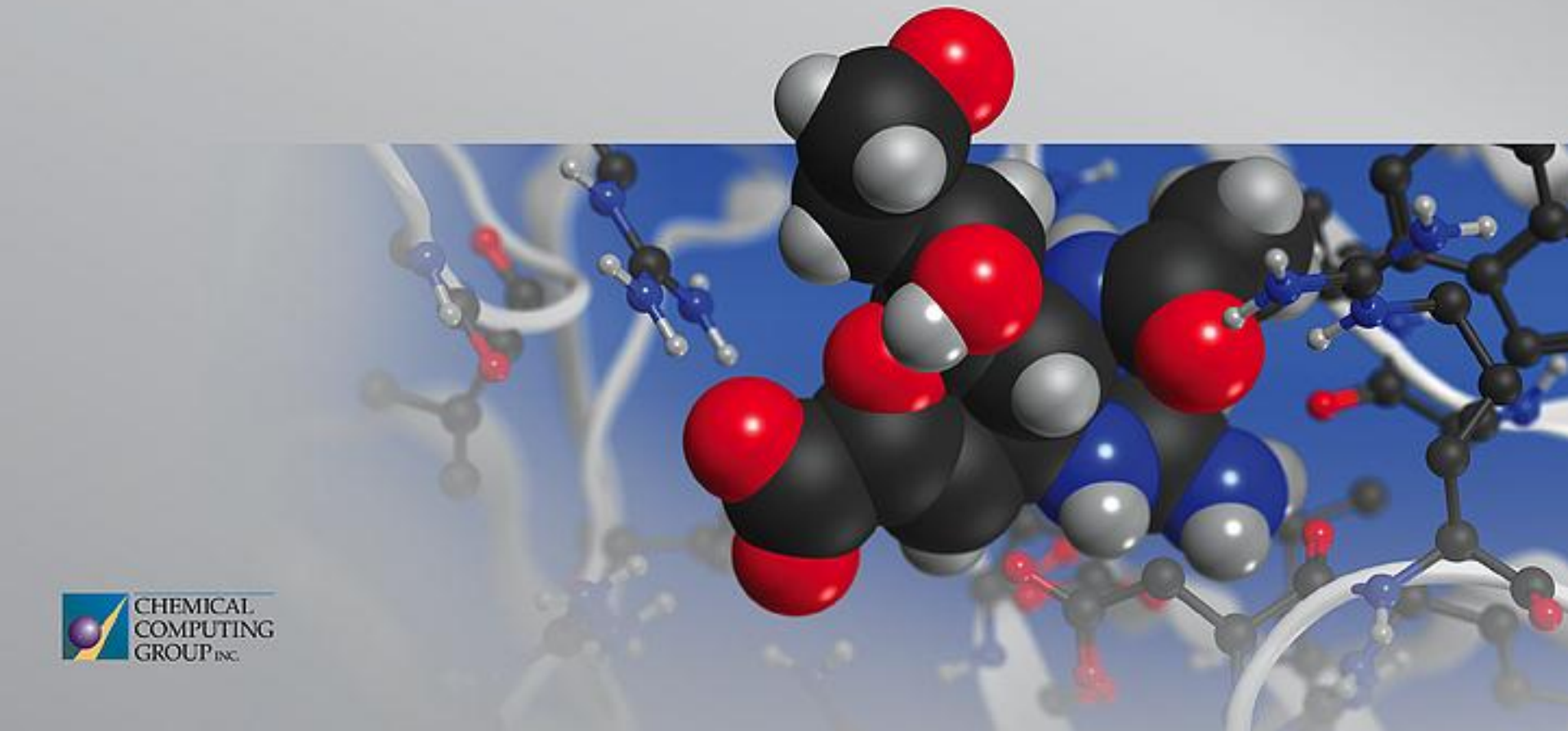
MOE Extensions for KNIME

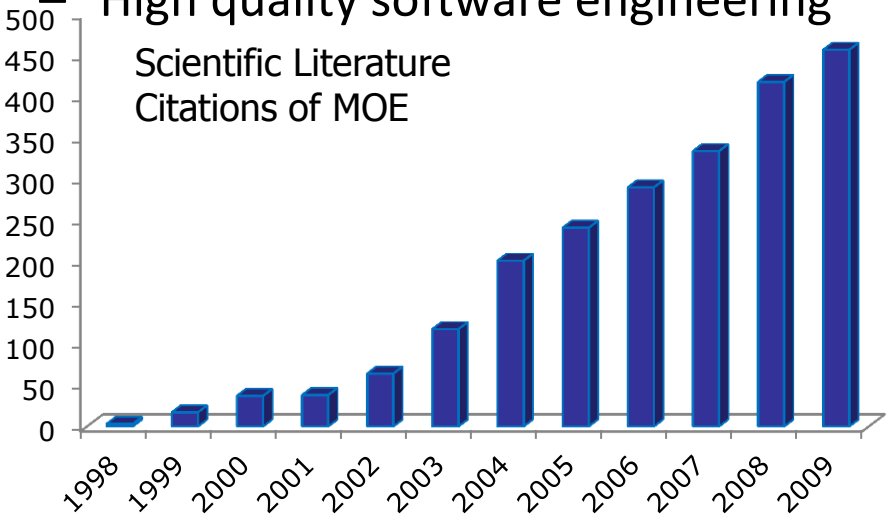
Version 2.0.0



CHEMICAL COMPUTING GROUP INC.

Chemical Computing Group

- **CCG's two main products**
  - Molecular Operating Environment™
  - PSILO® – Protein Silo
- **CCG's strengths**
  - Leading scientific R&D group in CAMD
  - Collaborative customer support
  - High quality software engineering



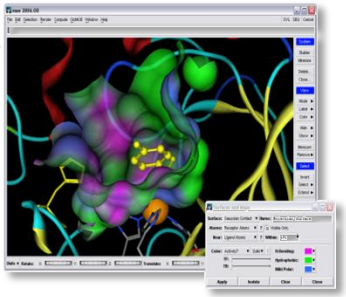
CCG founded in 1994
- Life sciences software company
- Headquarters in Montreal/Canada
- Offices: UK, Germany
- 800+ customers worldwide

Molecular Operating Environment™
- Pharmacophore Discovery
- Structure-Based Drug Design
- Protein and Antibody Modeling
- Molecular Modeling and Simulations
- Cheminformatics and (HTS) QSAR
- Medicinal Chemistry Applications
- Methods Development & Deployment

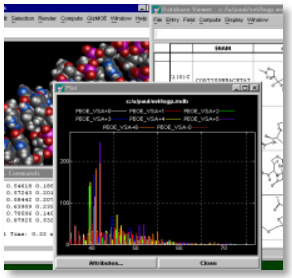
- **14+ years of consistent innovation and product releases**

MOE: Common Platform for Different Research Groups

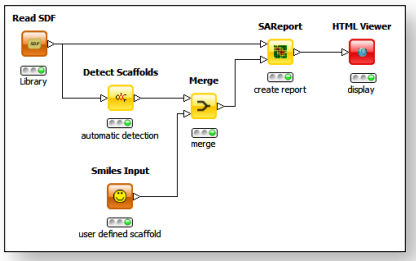
Computational Experts



Medicinal Chemists



Cheminformaticians



Occasional Users

Flat File
Clipboard
OEB

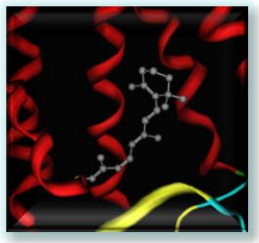
Intranet
Oracle
Socket

MOE

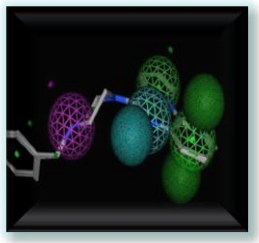
MOE/batch • MOE/smp
MOE/web • MOE/java

Local Files • Cut & Paste
MOE/java System for Connectivity

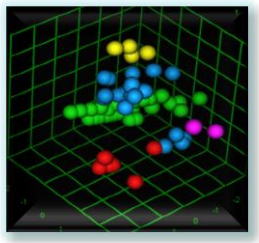
Desktop • Cluster • Pipeline Workflow
Windows • Linux • Unix • Browser



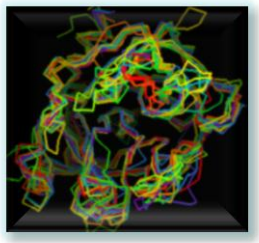
Structure Based Drug Design



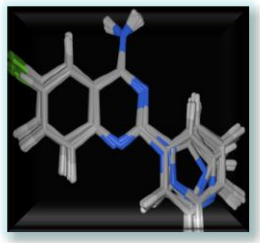
Pharmacophore Discovery



QSAR / HTS Cheminformatics



Protein Modeling



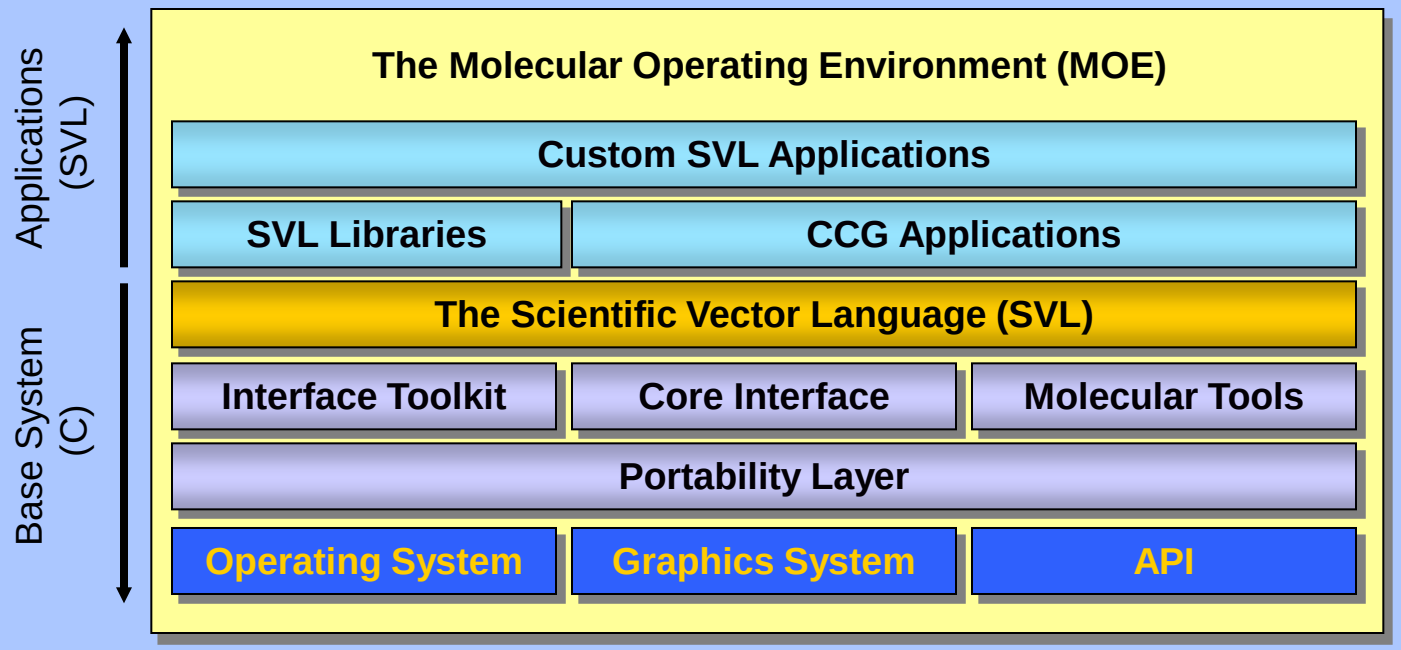
Modeling and Simulations

Methodology Development

SVL: Built-In Programming Language of MOE

MOE is built with a chemistry-aware embedded language

SVL is a portable high-level language
Applications/scripts are written in SVL – source code is shipped

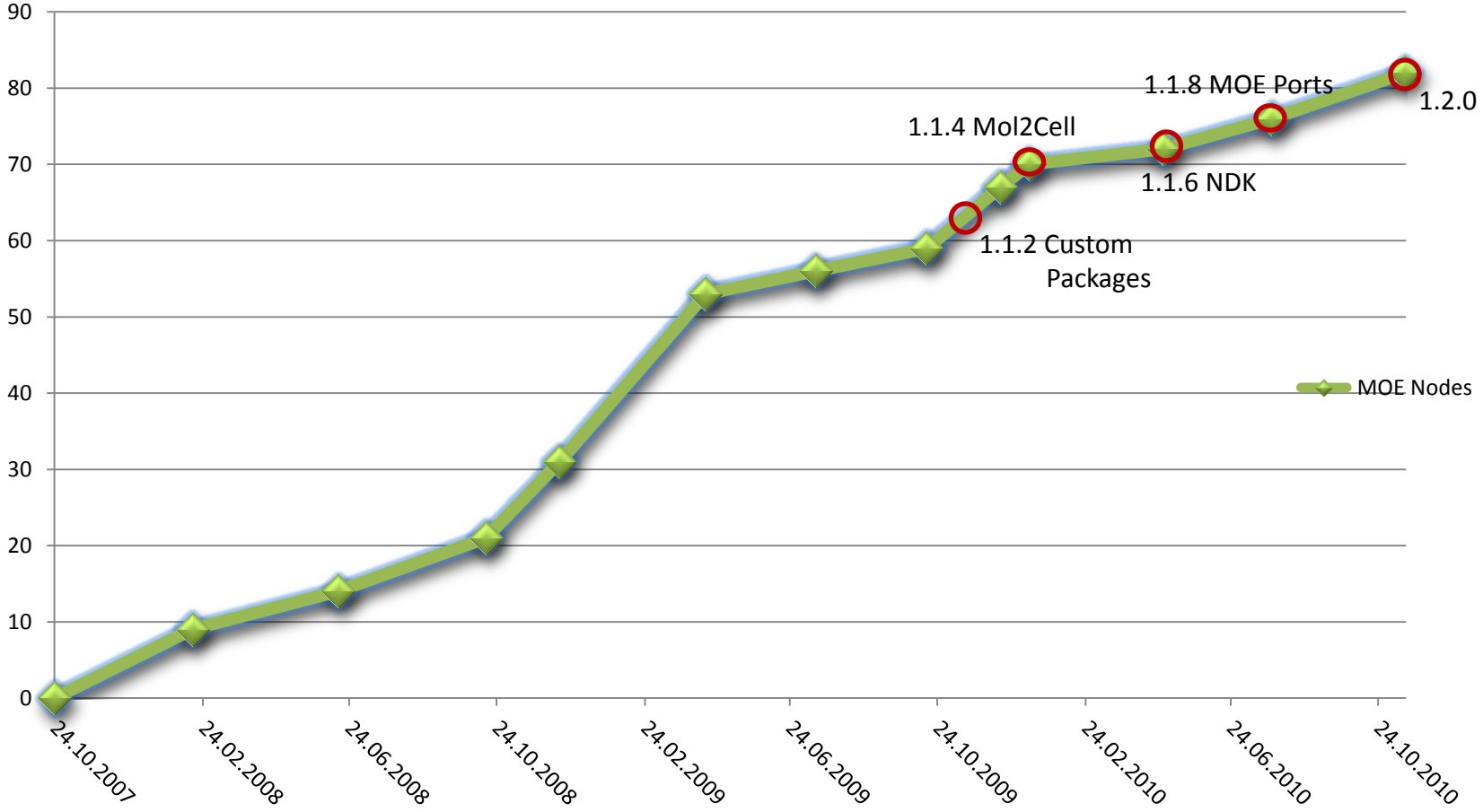


Platform Independence
SVL and the Portability Layer eliminate dependencies on hardware and operating system.

Application Program Interface
In-house or external code libraries can be incorporated to extend the built-in functionality.

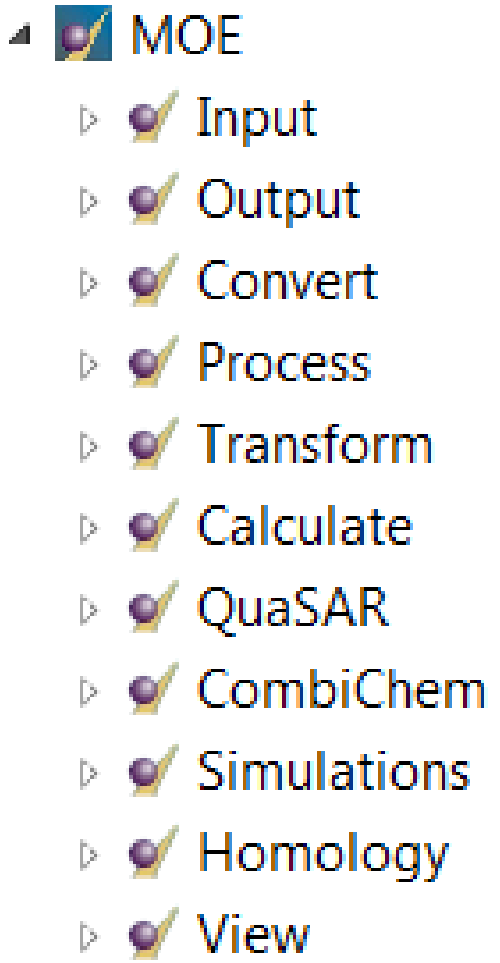
MOE Extensions: A Success Story

- Active development since 2007
- Customer driven node development
- KNIME Partners since 2011



MOE Extensions: Facts and Benefits

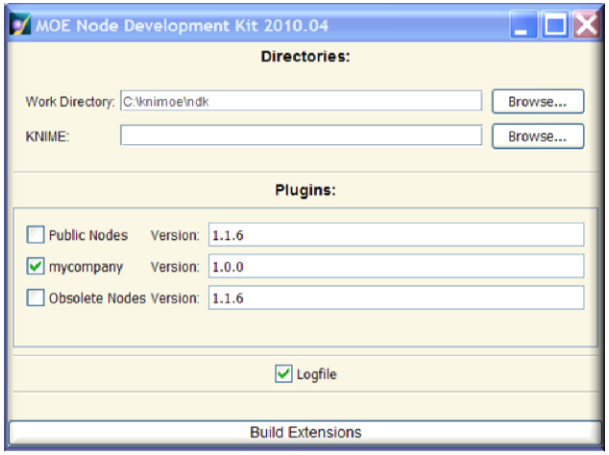
- **CCG offerings for KNIME**
  - 80+ pre-built MOE nodes included
  - Open SVL and Java source code
  - Add your own extensions
  - Tested with MOE 2010.10
  - Tested with KNIME 2.3.1
- **Supported by CCG**
  - Instructions available on the SVL Exchange
  - Mailing list for updates
  - Comprehensive documentation
- **Available free of charge to MOE Users**



Subcategories in MOE Node Repository

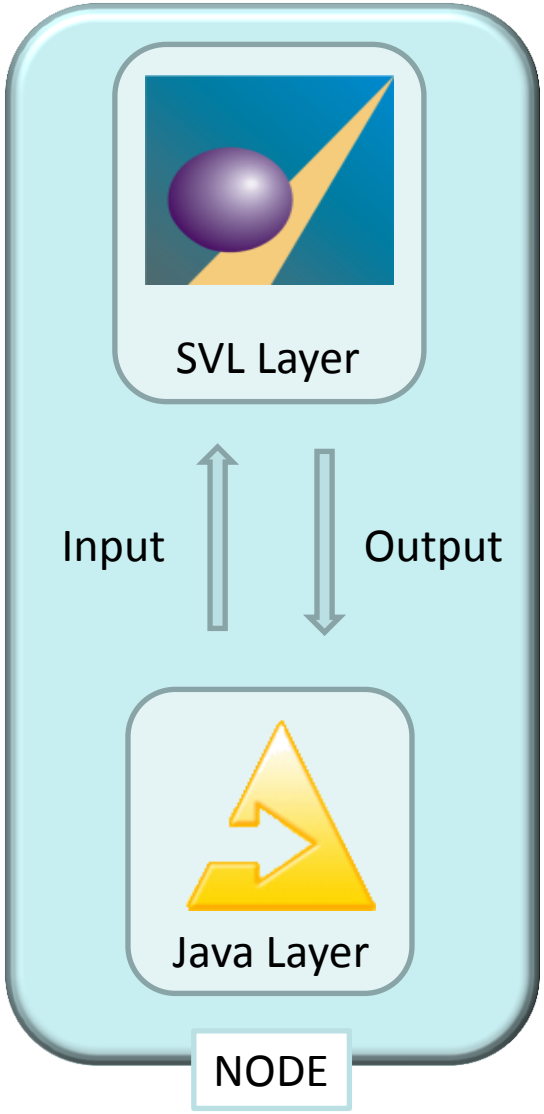
KNIME Nodes Powered by MOE/batch

- **Two-Layer Architecture**
  - Ensures identical results as in MOE/batch
  - SVL files can be edited independently
- **Java Layer controls In- and Output**
  - MOE used for chemistry calculations only
  - Efficient license token usage
  - Communication revised for speed



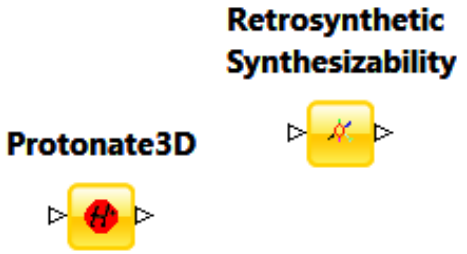
Generates Java Layer

from SVL

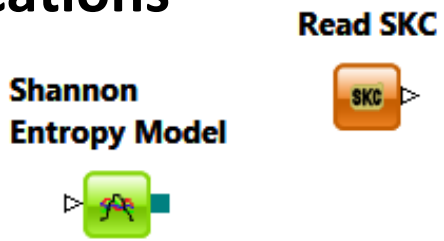


CCG Node Development

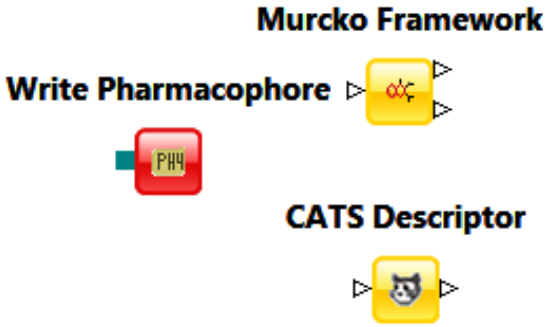
- **Nodes for MOE Applications**
  - Synthetic 'feasibility' (rsynth)
  - Protonate3D for protein preparation



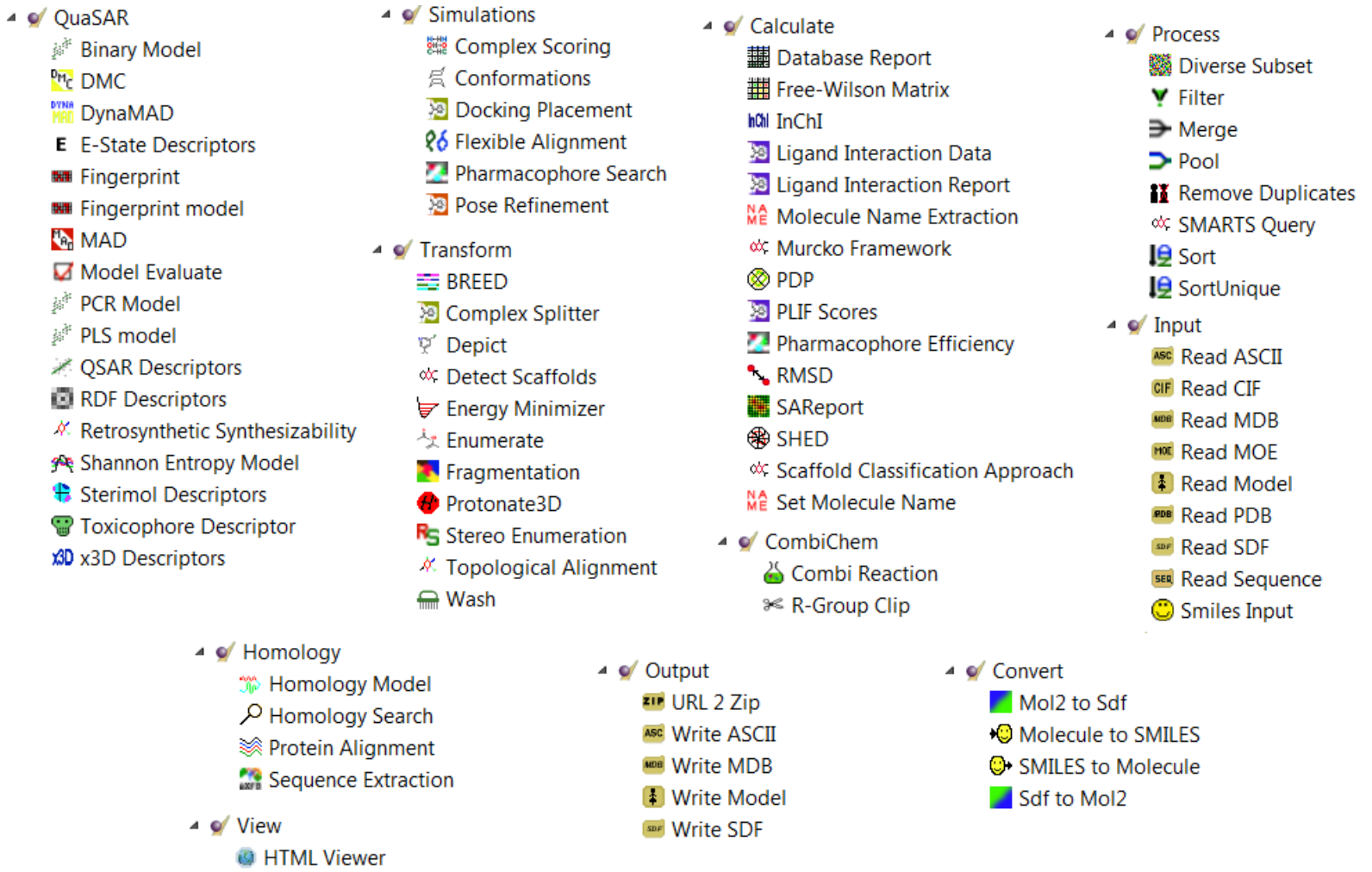
- **Nodes for SVL Exchange / Support Applications**
  - Read Isis Draw Sketch Files
  - Shannon Entropy Model



- **Nodes for Applications requested by Users**
  - Murcko Framework
  - Write Pharmacophore Model
  - CATS



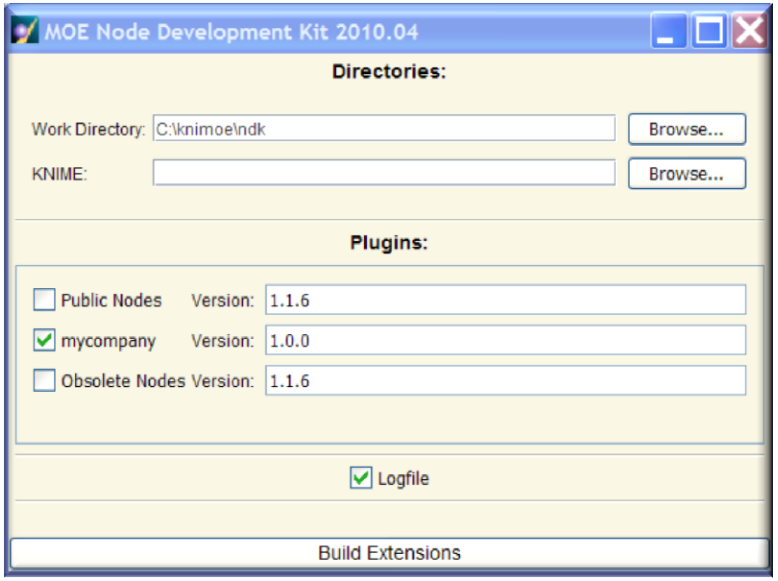
Available MOE Nodes



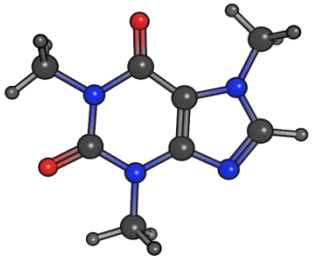
The current set of nodes covers the whole field of life-science research.

Compile MOE/batch SVL into KNIME Nodes

- **User-friendly graphical interface for cross-compilation**
  - Java based (Sun JDK required)
  - Windows, Linux and Mac OS X supported
  - Eclipse installation not required
- **Create custom MOE-based extensions**
  - Compile new nodes from SVL sources
  - Share nodes with others
  - SVL templates provided
  - Function documentation available
- **Manage multiple extensions**
  - Modular architecture
  - Folders are interpreted as separate packages



Chemistry Types for Small and Large Molecules

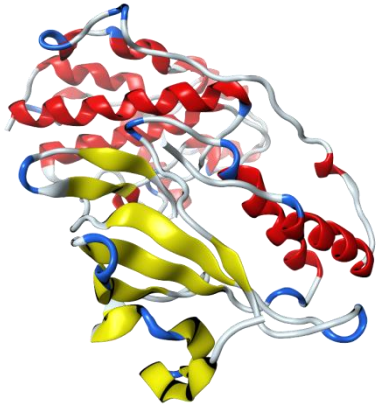


SdfCell, MolCell, CtabCell: Drugsize Molecules

- V2000 is a widely supported standard
- 999 atoms limit
- Lacks partial charges
- V3000 is supported by MOE, but not a standard elsewhere

Mol2Cell: Large Molecules (Tripos MOL2 Format)

- No hard coded atom limit
- Partial charges are supported
- Non-standard

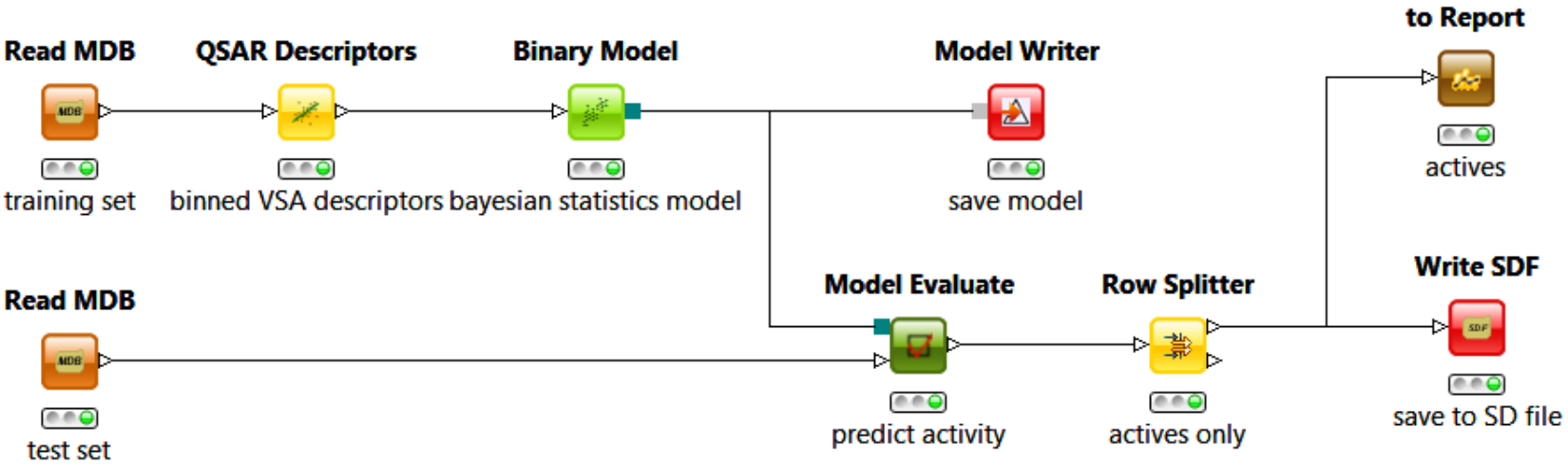


Design should allow to avoid unnecessary converter nodes
All types of molecules can be handled

MOE Model Ports

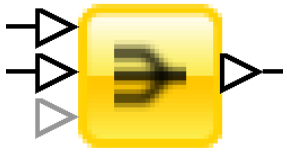
- **Special MOE Model Ports guide visual node connections**
  - Connect to other MOE Model Ports
  - Connect to generic KNIME Model Ports
- **Used for variety of MOE based models**
  - QSAR
  - Fingerprint
  - Composite models
  - …
- **Remove ambiguity when connecting nodes**

MOE Model Ports are depicted as greenish-blue squares

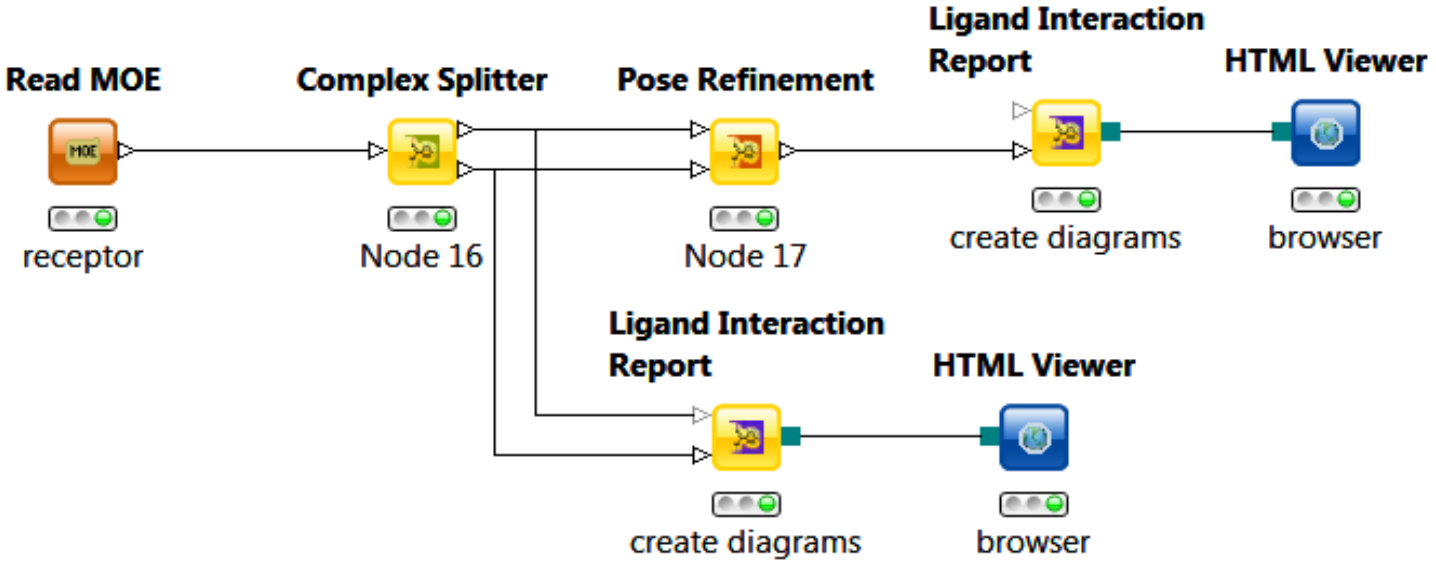


Optional Ports Supported by MOE Extensions

- **Problem: KNIME nodes required fixed number of instreams**
  - Duplicate nodes were necessary for similar tasks
- **Add flexibility with Optional Ports**
  - Unified node with arbitrary number of input streams
  - Available since KNIME 2.2
- **Implemented MOE nodes using Optional Ports**
  - Merge, SAReport, Complex Scoring, …

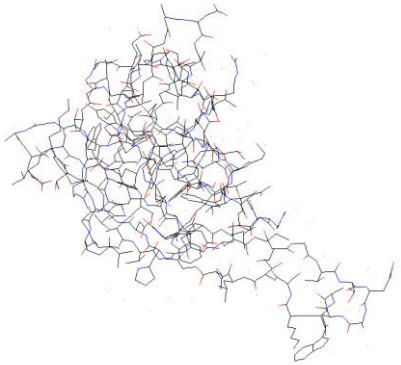


Optional Ports are depicted as grey triangles

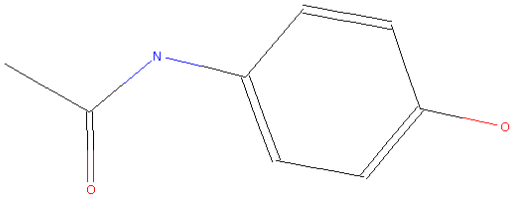


Molecule Rendering in Data Tables

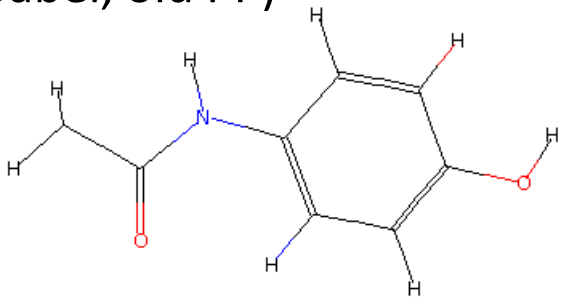
- **Customize rendering styles**
  - Show or hide hydrogens
  - Display large molecules as sigma skeleton
  - Color heteroatoms
- **Available for large and small molecules**
  - Mol2 supported
  - SDF, Mol, Ctab (V2000 + V3000) supported
  - Robust against format errors (OpenBabel, old PP)



Mol2Cell – Sigma Skeleton



SdfCell – Hydrogens hidden

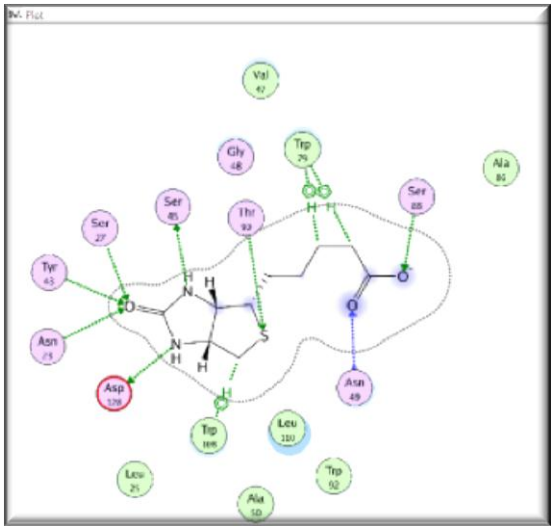


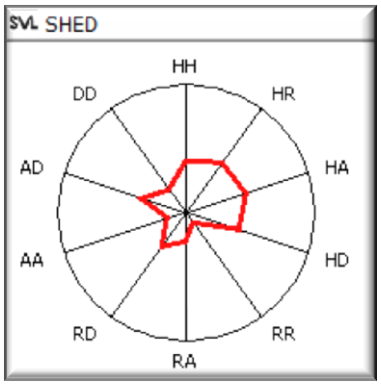
SdfCell – Hydrogens shown

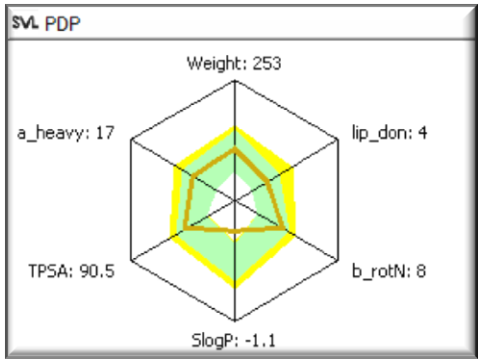
Future Developments: Improve quality, support wedges, annotate aromatic rings, …

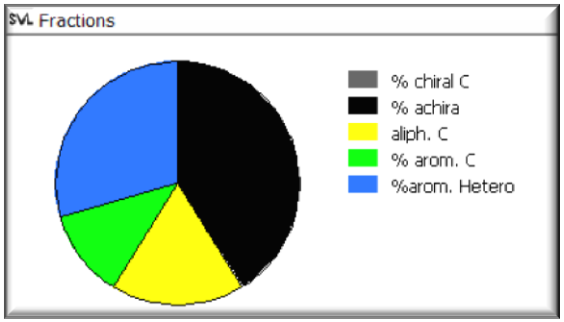
Display Images and Charts in Data Tables

- **Private SvlVarCell Type allows registration of CCG renderer**
  - Information is stored as an SVL vector
  - Render MOE images in tables
  - Three different chart types already supported
- **Not limited to MOE nodes**
  - No MOE tokens used during display
  - Coded in Java

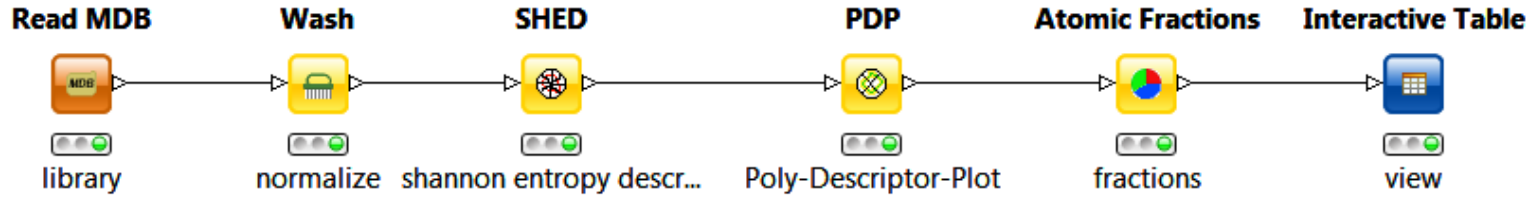


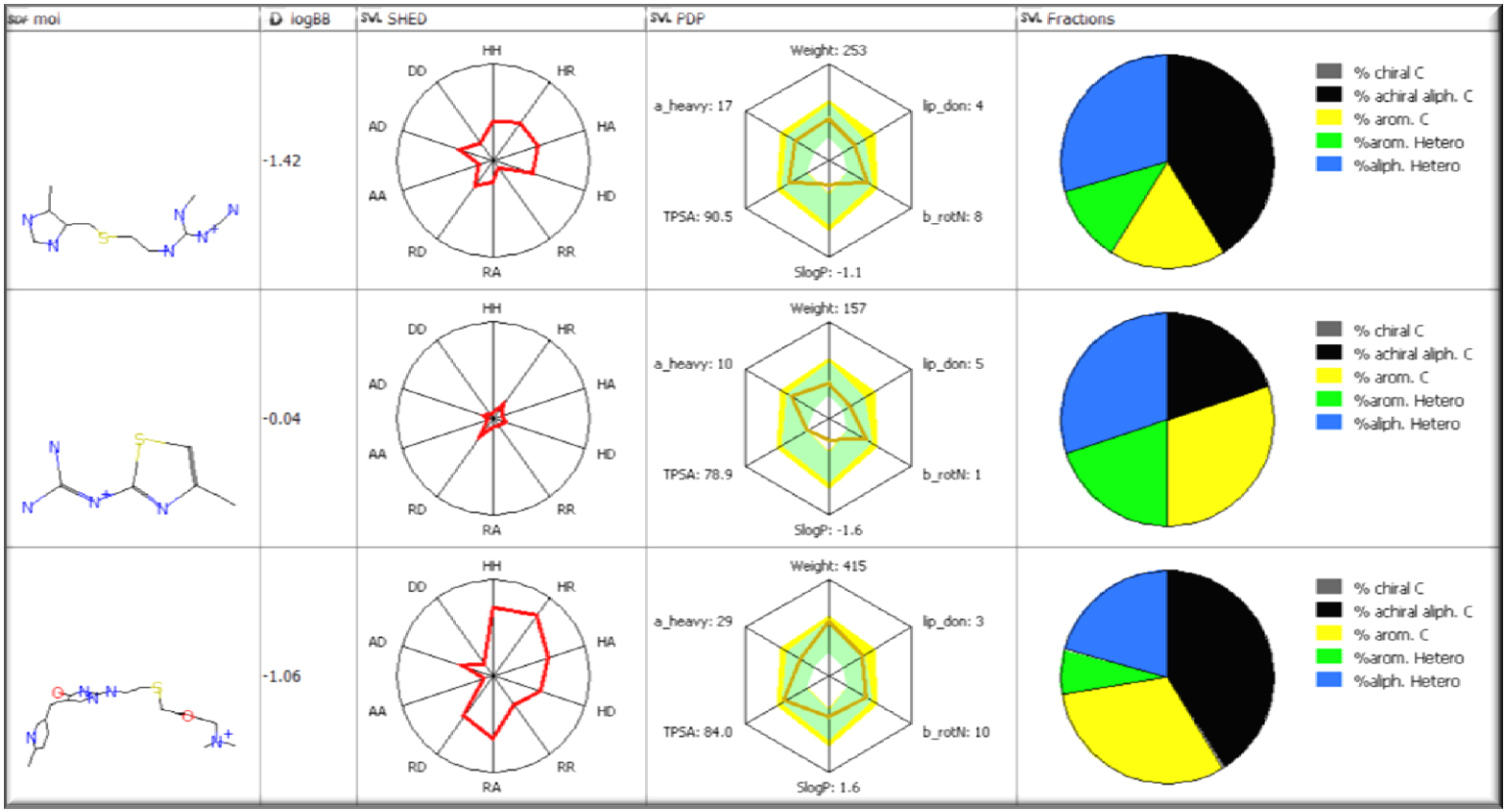






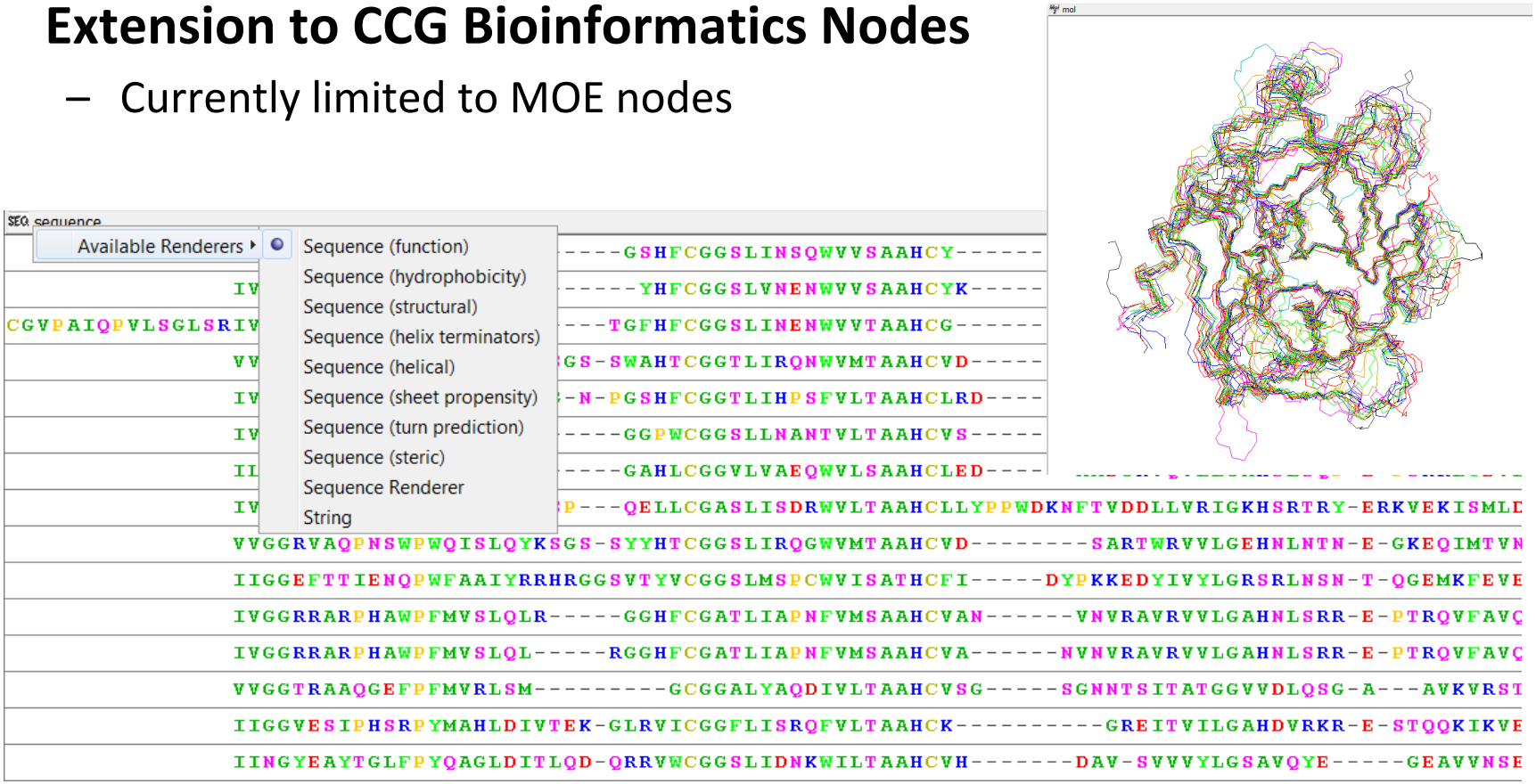
Wheel, Star and Pie Chart Types Supported





Rendering Sequences and Alignments

- **Renderer for Amino Acid Sequences and Alignments**
  - Nine different coloring schemes implemented
  - Sequences stored in special cell type (to be replaced by official type)
- **Extension to CCG Bioinformatics Nodes**
  - Currently limited to MOE nodes



Outline of New and Future Developments

- **I/O Layer rewritten for Speed and Compatibility**
  - Generic support for custom Cell Types
  - Optimized binary I/O protocol
- **Java Classes revised**
  - Dedicated support for more cell types (Sdf, Mol, Ctab, Mol2, Smiles)
  - Enhanced support for flow variables
  - Deprecated methods replaced (ongoing)
- **New SVL functions to connect to enhanced Java Classes**
  - SVL functions as frontends to new java methods available
- **Future Framework Improvements**
  - Support for MOE SOAP server
  - Batch Execution of KNIME workflows in MOE
  - New Nodes ☺

Summary

- **80+ nodes currently available in MOE Extensions**
  - Cheminformatics and data processing (SAReport, toxicophore descriptors, Murcko Framework, etc.)
  - Simulations (docking, conformational search pharmacophore, etc.)
  - Bioinformatics (alignment, Ramachandran plots, PLIF, etc.)
- **Compilation wizard for custom node development**
  - Reuse of validated SVL Sources
  - Easy development of new node by preconfigured templates
- **Robust Renderers**
  - Molecules in Sdf, Mol, Ctab and Mol2Cells
  - Sequences and Alignments
  - Three different Chart Types
- **Free (included in MOE support) and open source**

More Information

Contact:

support@chemcomp.com