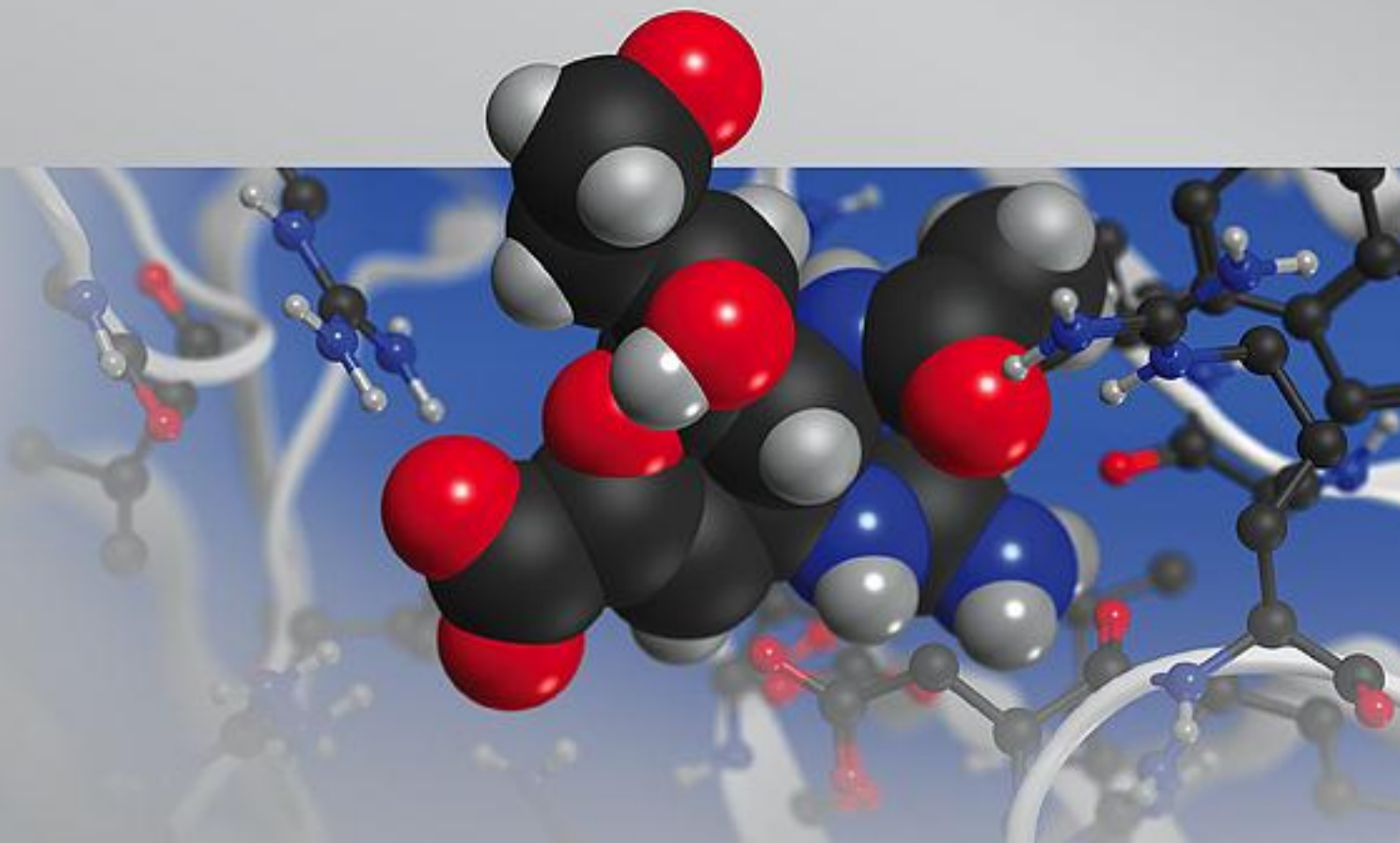


MOE Extensions for KNIME

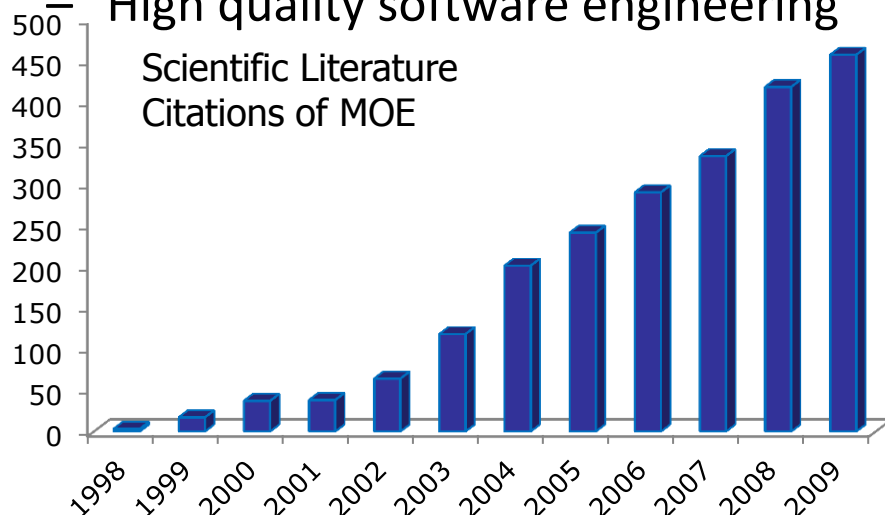
Version 2.0.0



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



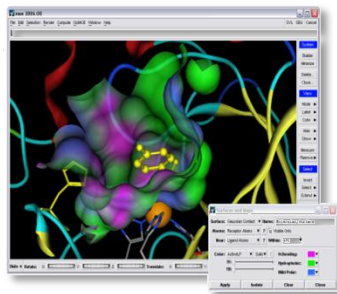
Molecular Operating Environment™

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

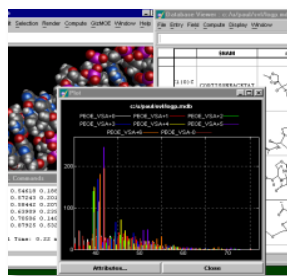
- **14+ years of consistent innovation and product releases**

MOE: Common Platform for Different Research Groups

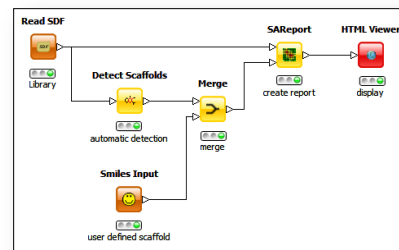
Computational Experts



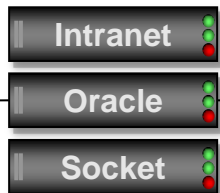
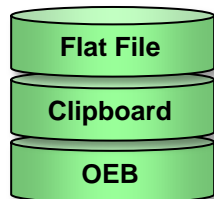
Medicinal Chemists



Cheminformaticians

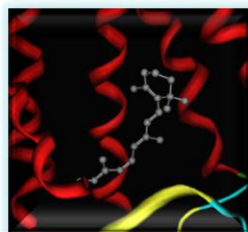


Occasional Users

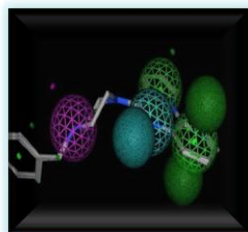


Local Files • Cut & Paste
MOE/java System for Connectivity

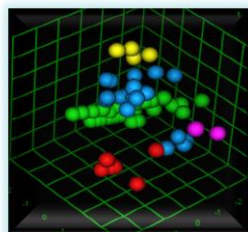
Desktop • Cluster • Pipeline Workflow
Windows • Linux • Unix • Browser



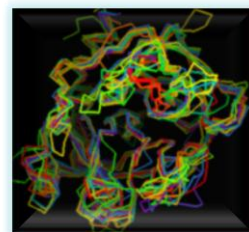
Structure Based Drug Design



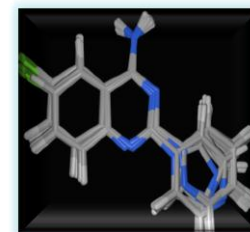
Pharmacophore Discovery



QSAR / HTS Cheminformatics



Protein Modeling



Modeling and Simulations



Methodology Development

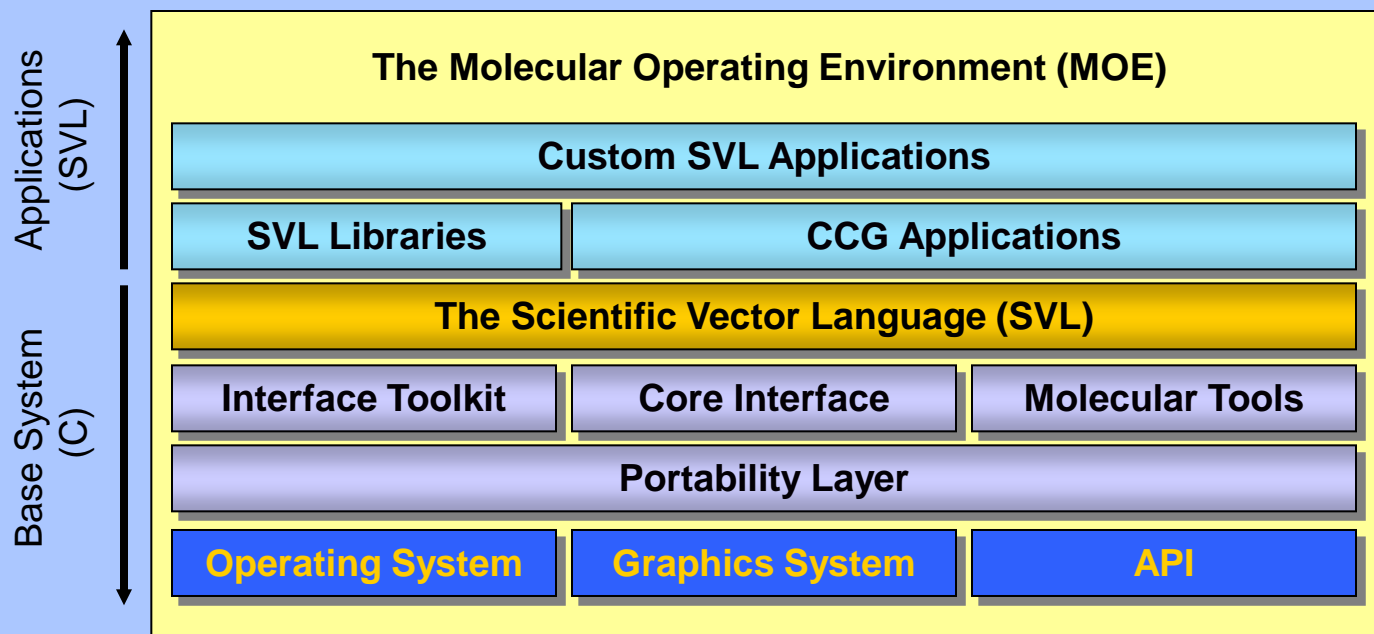
MOE Molecular Operating Environment

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

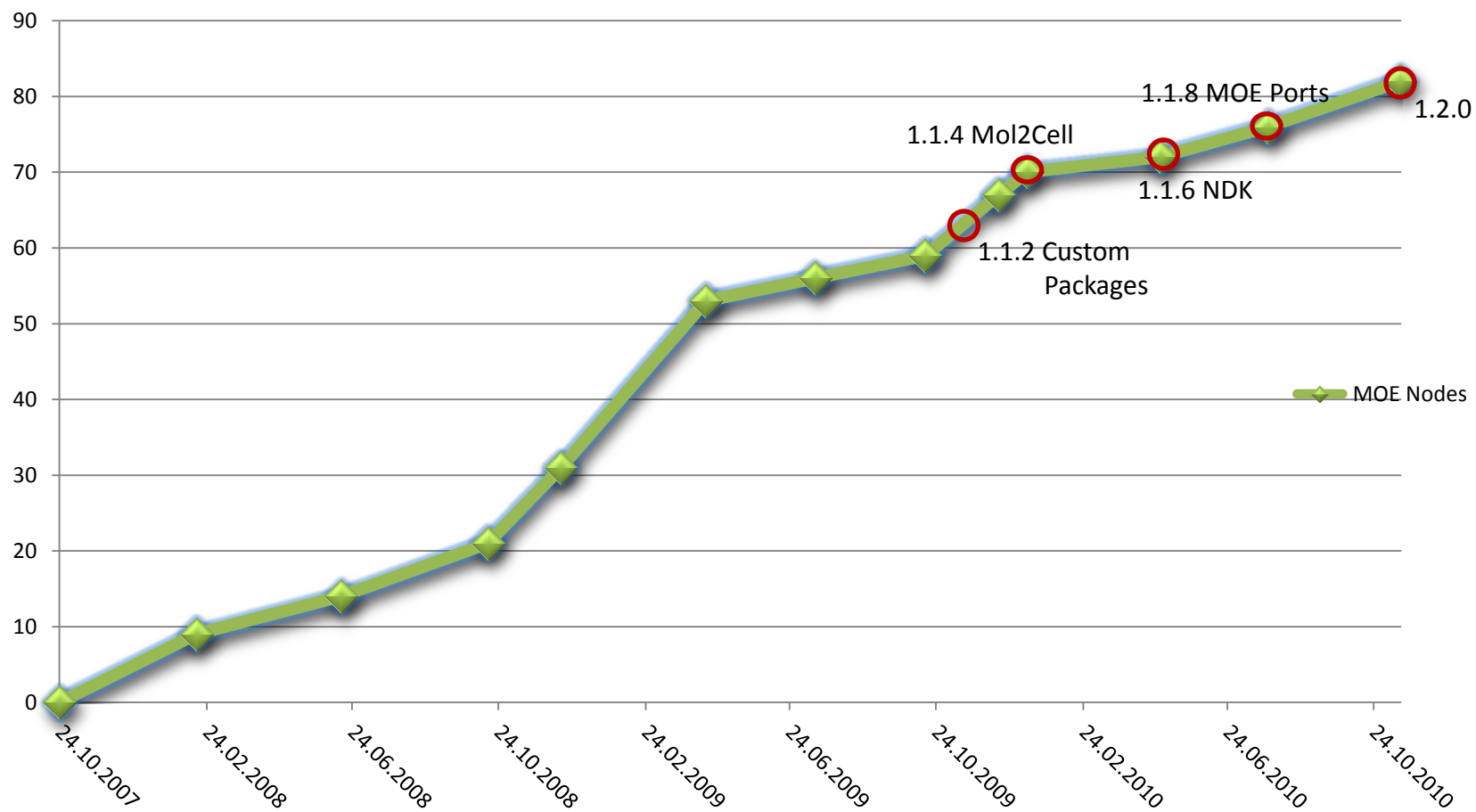
SVL and the Portability Layer eliminate dependencies on hardware and operating system.

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

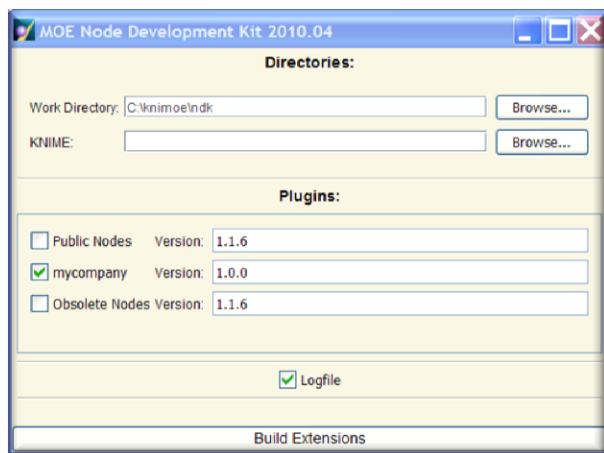
▲ MOE

- ▷  Input
- ▷  Output
- ▷  Convert
- ▷  Process
- ▷  Transform
- ▷  Calculate
- ▷  QuaSAR
- ▷  CombiChem
- ▷  Simulations
- ▷  Homology
- ▷  View

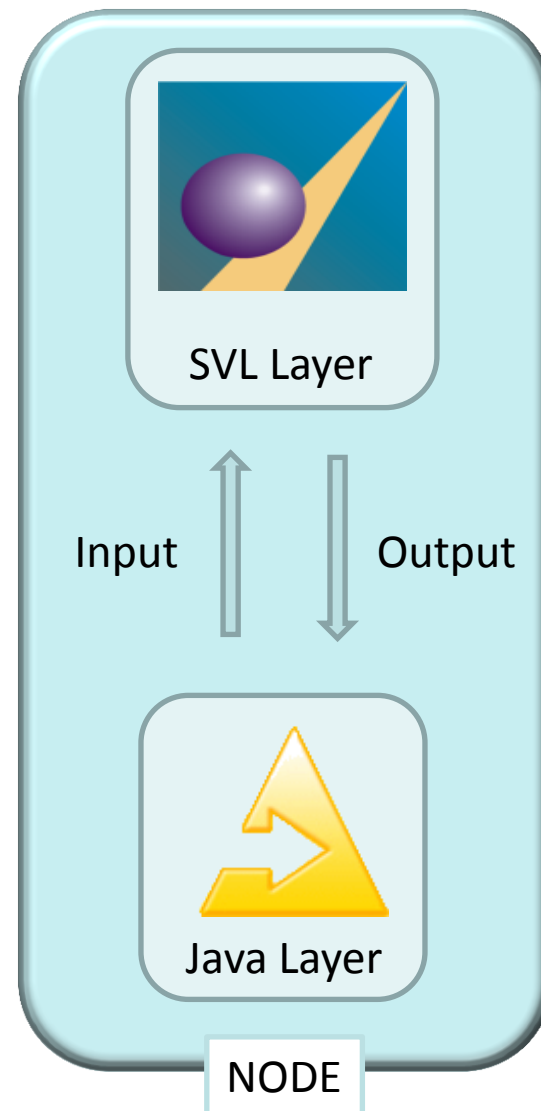
*Subcategories in MOE Node
Repository*

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

Protonate3D



Retrosynthetic
Synthesizability



▪ Nodes for SVL Exchange / Support Applications

- Read Isis Draw Sketch Files
- Shannon Entropy Model

Shannon
Entropy Model



Read SKC



▪ Nodes for Applications requested by Users

- Murcko Framework
- Write Pharmacophore Model
- CATS

Murcko Framework

Write Pharmacophore



CATS Descriptor



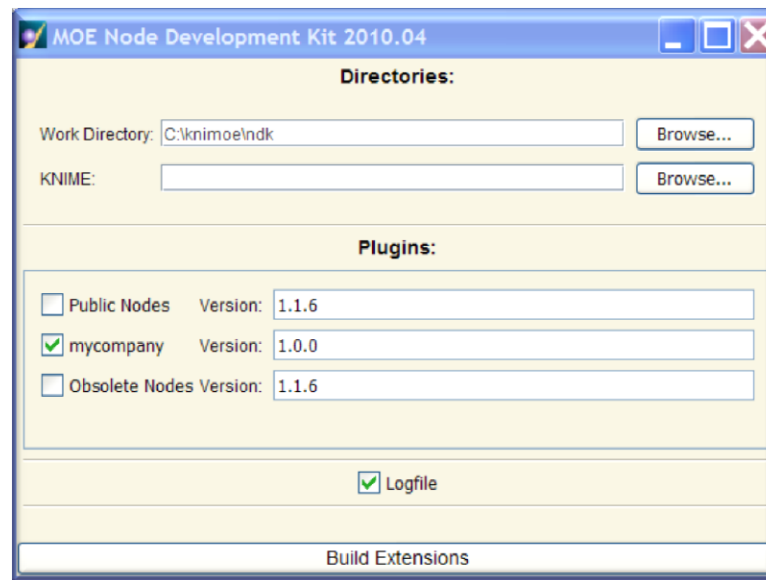
Available MOE Nodes

- QuaSAR
 - Binary Model
 - DMC
 - DynaMAD
 - E-State Descriptors
 - Fingerprint
 - Fingerprint model
 - MAD
 - Model Evaluate
 - PCR Model
 - PLS model
 - QSAR Descriptors
 - RDF Descriptors
 - Retrosynthetic Synthesizability
 - Shannon Entropy Model
 - Sterimol Descriptors
 - Toxicophore Descriptor
 - x3D Descriptors
- Simulations
 - Complex Scoring
 - Conformations
 - Docking Placement
 - Flexible Alignment
 - Pharmacophore Search
 - Pose Refinement
 - Transform
 - BREED
 - Complex Splitter
 - Depict
 - Detect Scaffolds
 - Energy Minimizer
 - Enumerate
 - Fragmentation
 - Protonate3D
 - Stereo Enumeration
 - Topological Alignment
 - Wash
- Homology
 - Homology Model
 - Homology Search
 - Protein Alignment
 - Sequence Extraction
 - View
 - HTML Viewer
- Calculate
 - Database Report
 - Free-Wilson Matrix
 - InChI
 - Ligand Interaction Data
 - Ligand Interaction Report
 - Molecule Name Extraction
 - Murcko Framework
 - PDP
 - PLIF Scores
 - Pharmacophore Efficiency
 - RMSD
 - SARReport
 - SHED
 - Scaffold Classification Approach
 - Set Molecule Name
 - CombiChem
 - Combi Reaction
 - R-Group Clip
- Output
 - URL 2 Zip
 - Write ASCII
 - Write MDB
 - Write Model
 - Write SDF
 - Convert
 - Mol2 to Sdf
 - Molecule to SMILES
 - SMILES to Molecule
 - Sdf to Mol2
- Process
 - Diverse Subset
 - Filter
 - Merge
 - Pool
 - Remove Duplicates
 - SMARTS Query
 - Sort
 - SortUnique
 - Input
 - Read ASCII
 - Read CIF
 - Read MDB
 - Read MOE
 - Read Model
 - Read PDB
 - Read SDF
 - Read Sequence
 - Smiles Input

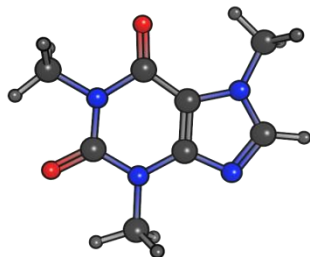
The current set of nodes covers the whole field of life-science research.

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

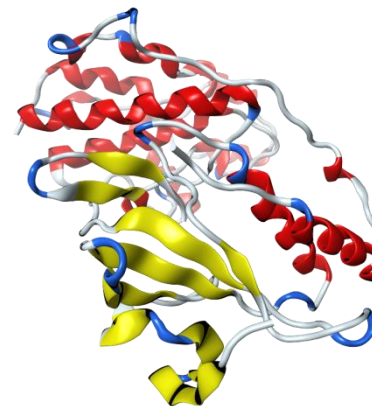


SdfCell, MolCell, CtabCell: Drugsize Molecules

-  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



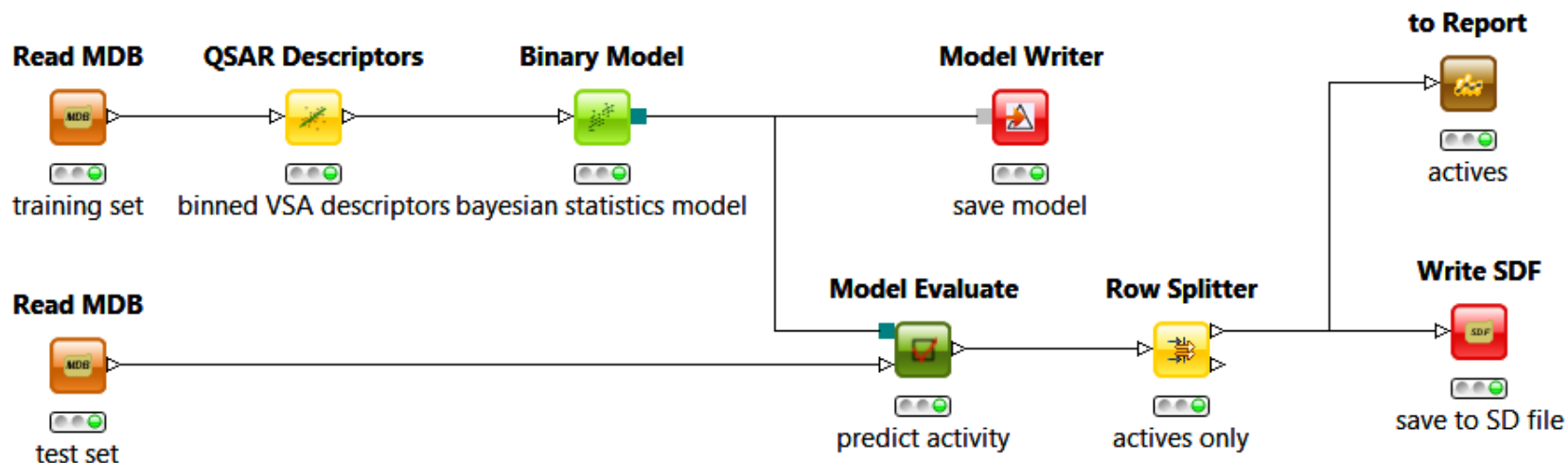
Design should allow to avoid unnecessary converter nodes
All types of molecules can be handled

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**

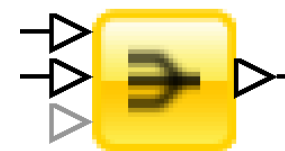


MOE Model Ports are depicted as greenish-blue squares

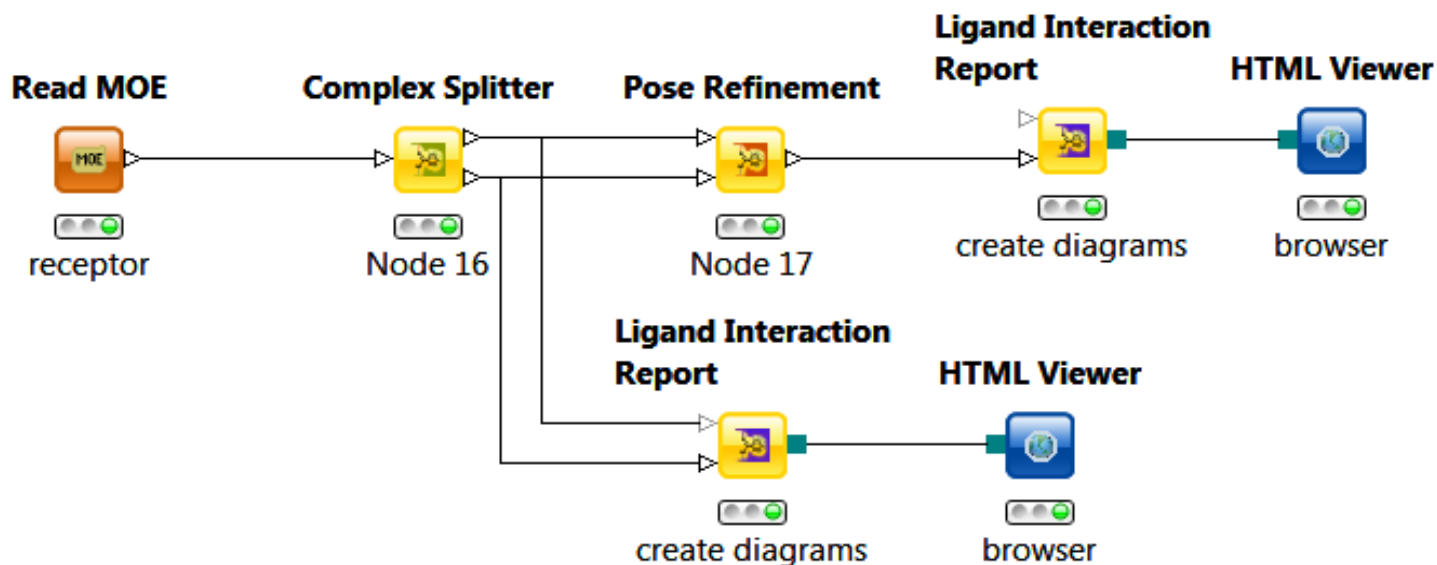


Optional Ports Supported by MOE Extensions

- **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, ...



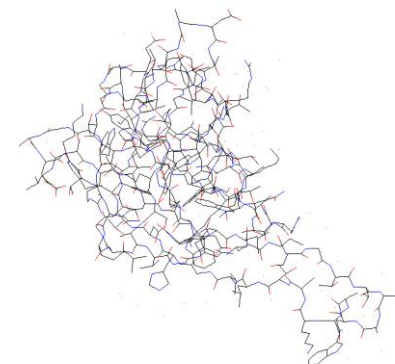
Optional Ports are depicted as grey triangles



Molecule Rendering in Data Tables

- **Customize rendering styles**

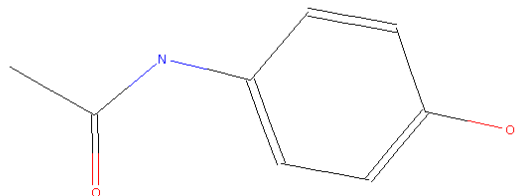
- Show or hide hydrogens
- Display large molecules as sigma skeleton
- Color heteroatoms



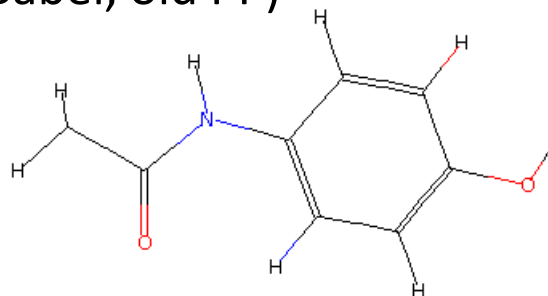
Mol2Cell – Sigma Skeleton

- **Available for large and small molecules**

- Mol2 supported
- SDF, Mol, Ctab (V2000 + V3000) supported
- Robust against format errors (OpenBabel, old PP)



SdfCell – Hydrogens hidden

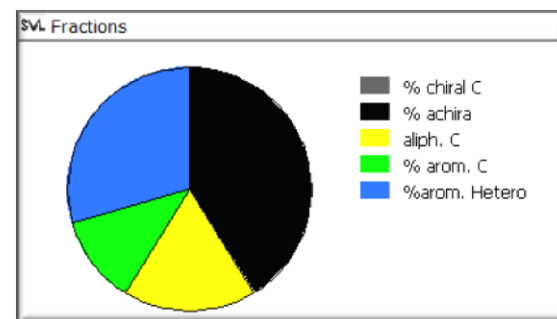
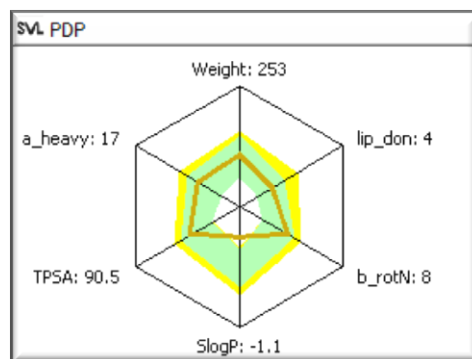
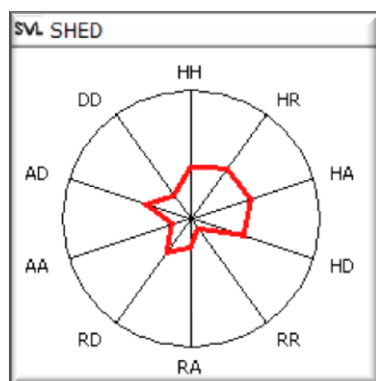
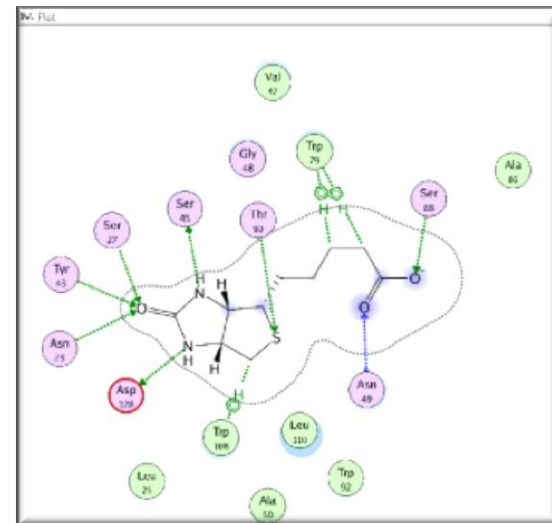


SdfCell – Hydrogens shown

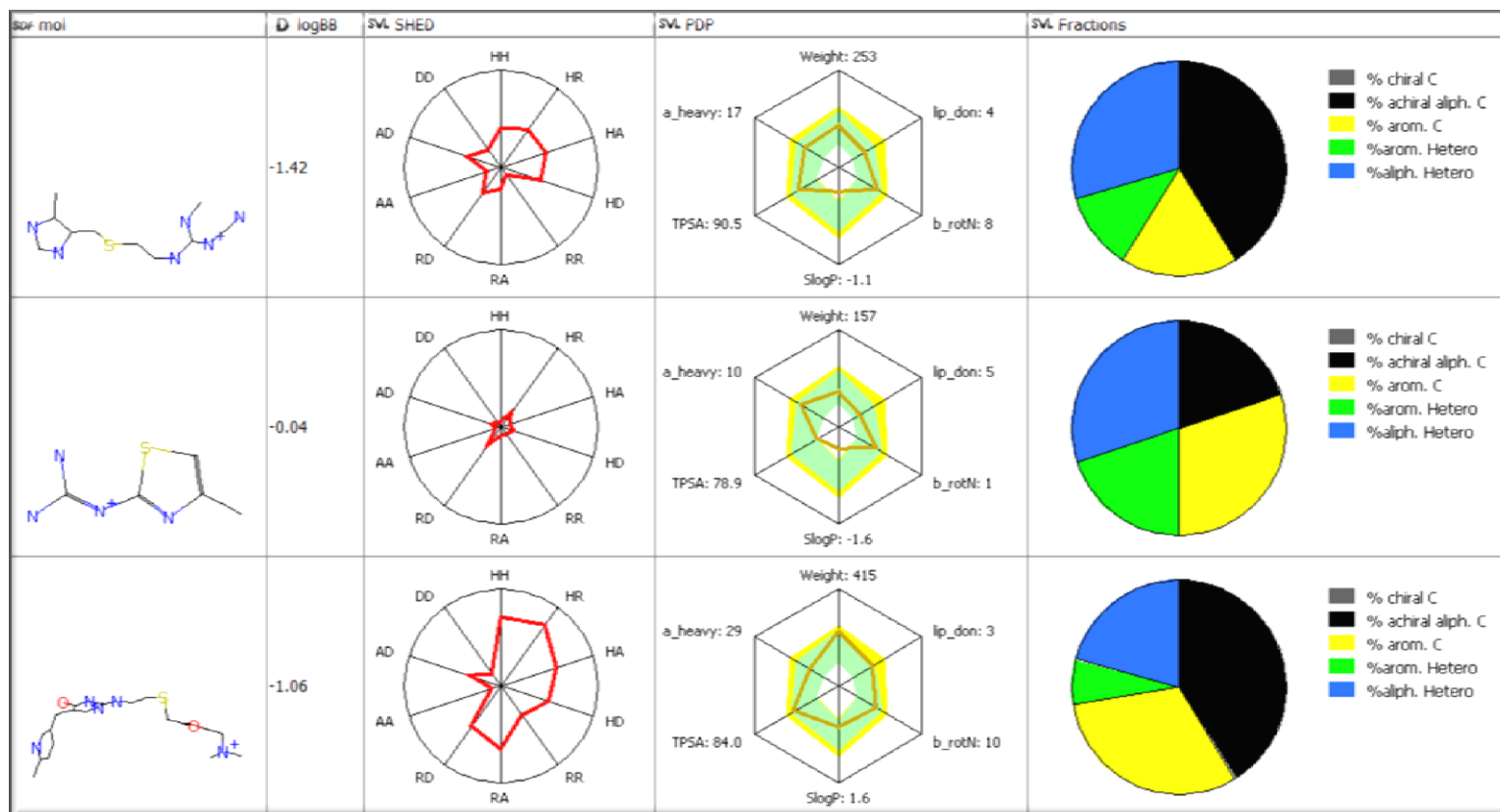
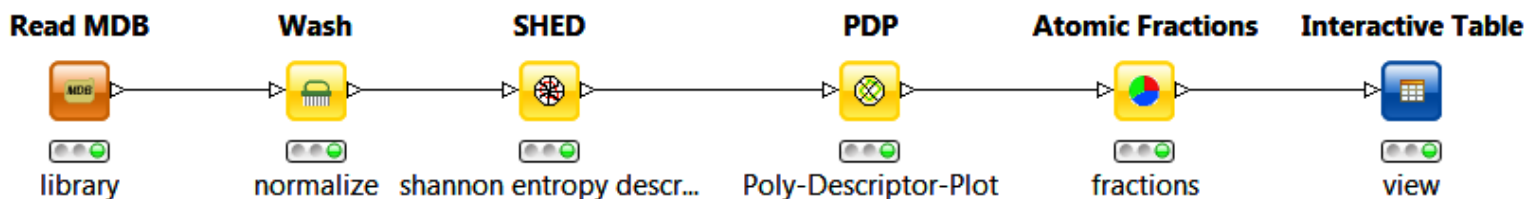
Future Developments: Improve quality, support wedges, annotate aromatic rings, ...

Display Images and Charts in Data Tables

- **Private SvIVarCell 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

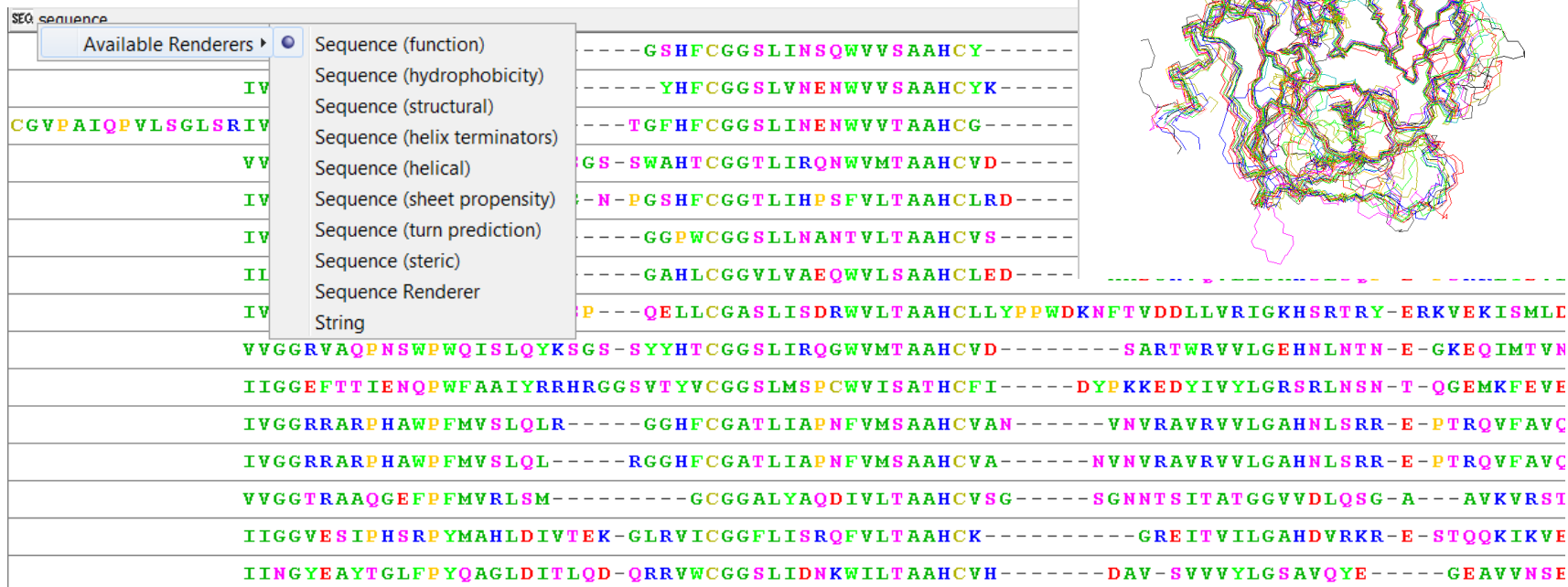


Wheel, Star and Pie Chart Types Supported



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



The screenshot displays the MOE software interface. On the left, a menu titled 'SEQ. sequence' shows 'Available Renderers' with a dropdown list of nine options: Sequence (function), Sequence (hydrophobicity), Sequence (structural), Sequence (helix terminators), Sequence (helical), Sequence (sheet propensity), Sequence (turn prediction), Sequence (steric), Sequence Renderer, and String. The 'String' option is selected. The main window shows a multi-sequence alignment of amino acid sequences, with each sequence rendered in a different color. To the right of the alignment, a 3D ribbon diagram of a protein structure is shown, colored to match the sequences in the alignment.

SEQ. sequence	Available Renderers	Sequence
	Sequence (function)	-----GSHFCGGS LINSQWVVSAAHCY-----
	Sequence (hydrophobicity)	-----YHFCGGS LVNENWVVSAAHCYK-----
	Sequence (structural)	-----TGFHFCGGS LINENWVVTAAHCG-----
	Sequence (helix terminators)	-----GS-SWAHTCGGTLIRQNWVMTAAHCVD-----
	Sequence (helical)	-----N-PGSHFCGGTLIHPSFVLTAAHCLRD-----
	Sequence (sheet propensity)	-----GGPWCGGSLLNANTVLTAAHCVS-----
	Sequence (turn prediction)	-----GAHLCGGVLVAEQWVLSAAHCLED-----
	Sequence (steric)	-----P-----
	Sequence Renderer	-----P-----
	String	-----P-----
		VVGGRRVAQPNSWPWQISLQYKSGS-SYYHTCGGSLIRQGWVMTAAHCVD-----SARTWRVVLGEHNLNTN-E-GKEQIMTVN
		IIGGEFTTIEHQPFWFAAIYRRHRGGSVTYVCGGSLMSPCWVISATHCFI-----DYPKKEDIYVYLGSRSLNSN-T-QGEMKFEVE
		IVGRRRAPHAWPFMVSLQLR-----GGHFCGATLIAPNFVMSAAHCVAN-----VNVRAVRVVLGAHNLSRR-E-PTRQVFAVC
		IVGRRRAPHAWPFMVSLQL-----RGGHFCGATLIAPNFVMSAAHCVA-----NVNVRAVRVVLGAHNLSRR-E-PTRQVFAVC
		VVGGTRAAQGEFPFMRVLSM-----CGGALYAQDIVLTAAHCVSG-----SGNNTSITATGGVVDLQSG-A---AVKVRST
		IIGGVESIPHSRPMYMAHLDIVTEK-GLRVICGGFLISRQFVLTAAHCK-----GREITVILGAHDVRKR-E-STQQKIKVE
		IINGYEAYTGLFPYQAGLDITLQD-QRRVWCGGSLIDNKWILTAAHCVH-----DAV-SVVVYLGSAVQYE-----GEAVVNSE

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 (SARReport, 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