# Accelrys Chemistry Extensions for KNIME

Keith Taylor, PhD

4th KNIME User Group Meeting
March 2-3, 2011

accelrys®

# Overview

- Goal of the Accelrys Chemistry Extensions
- Present capabilities
- Future enhancements

# Our Goal

- Expose Accelrys Chemistry capabilities in an easy-to-use framework
- Provide access to powerful KNIME data mining tools
- Support real chemistry
  - Reaction-based enumeration
  - Stereoenumeration
  - Rgroup decomposition/enumeration
  - Structure, Substructure, Similarity searches
- Be extensible
  - Open Source Nodes allow customization
  - Cheshire Script node

# Chemistry Extension Features

- Core Feature
  - Does not require any additional Accelrys products
- Cheshire Feature
  - Requires a Accelrys Cheshire installation on the machine running KNIME
- Database Feature
  - Requires access to Direct 6.x (does not have to be on the machine running KNIME)

# Core Nodes

- Molfile reader
- Rxnfile reader
- SDfile reader
- SDfile writer
- Deduplication
- Chime String Reader
- Chime String Writer

# Cheshire Nodes

- Reaction-based enumeration
- Transform definition (for Reaction-based enumeration)
- Reactant addition (for Reaction-based enumeration)
- Reaction To Query Converter (for Reaction-based enumeration)
- Rgroup decomposition
- Scaffold to Query Converter (for Rgroup Decomposition)
- Specific enumeration (generic -> specifics)

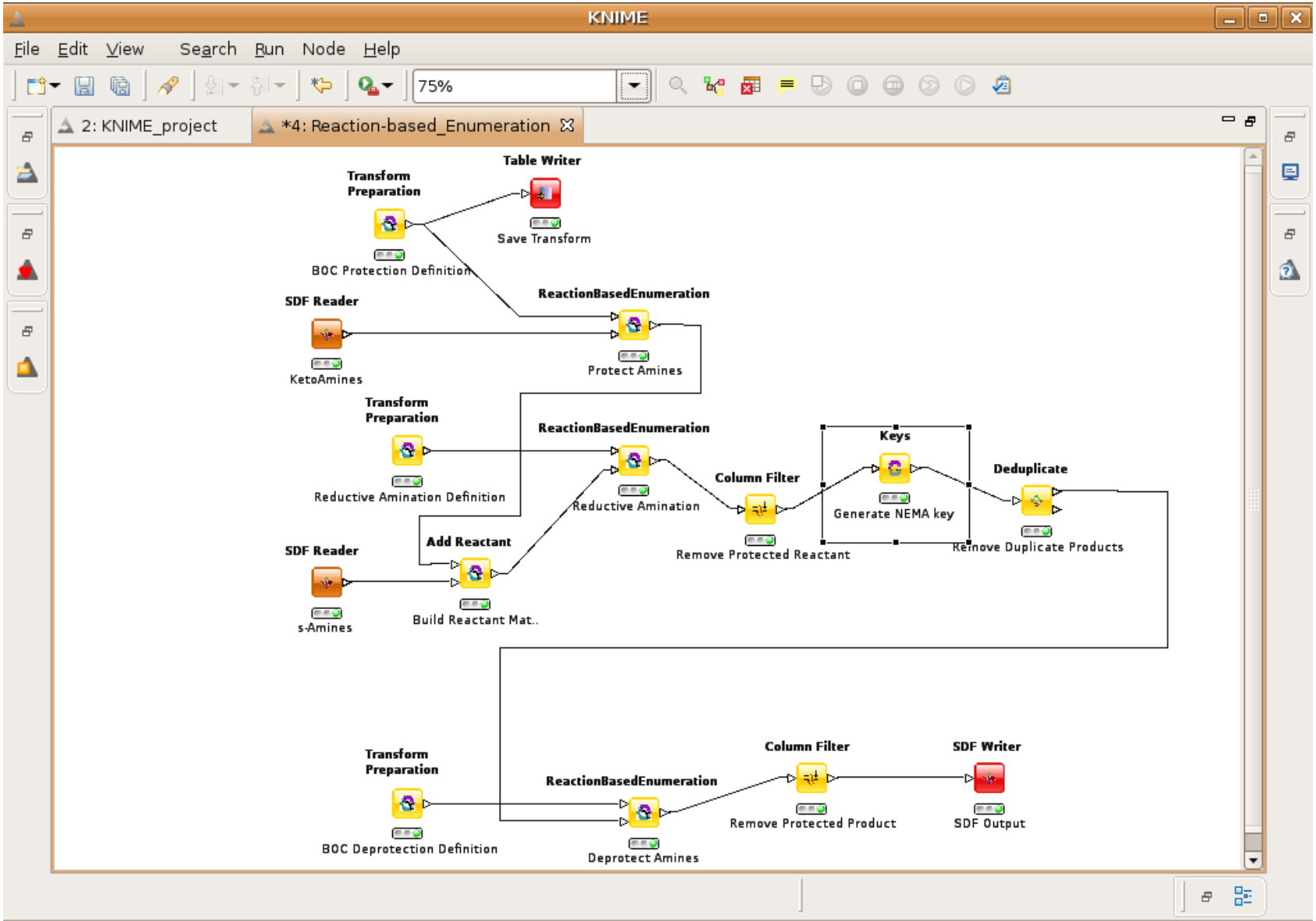
# Cheshire Nodes (cont.)

- Stereoisomer enumeration
- Keys (NEMA, SSKeys, SSKeys_subset)
- Cheshire script execution
- Remove V3000 features

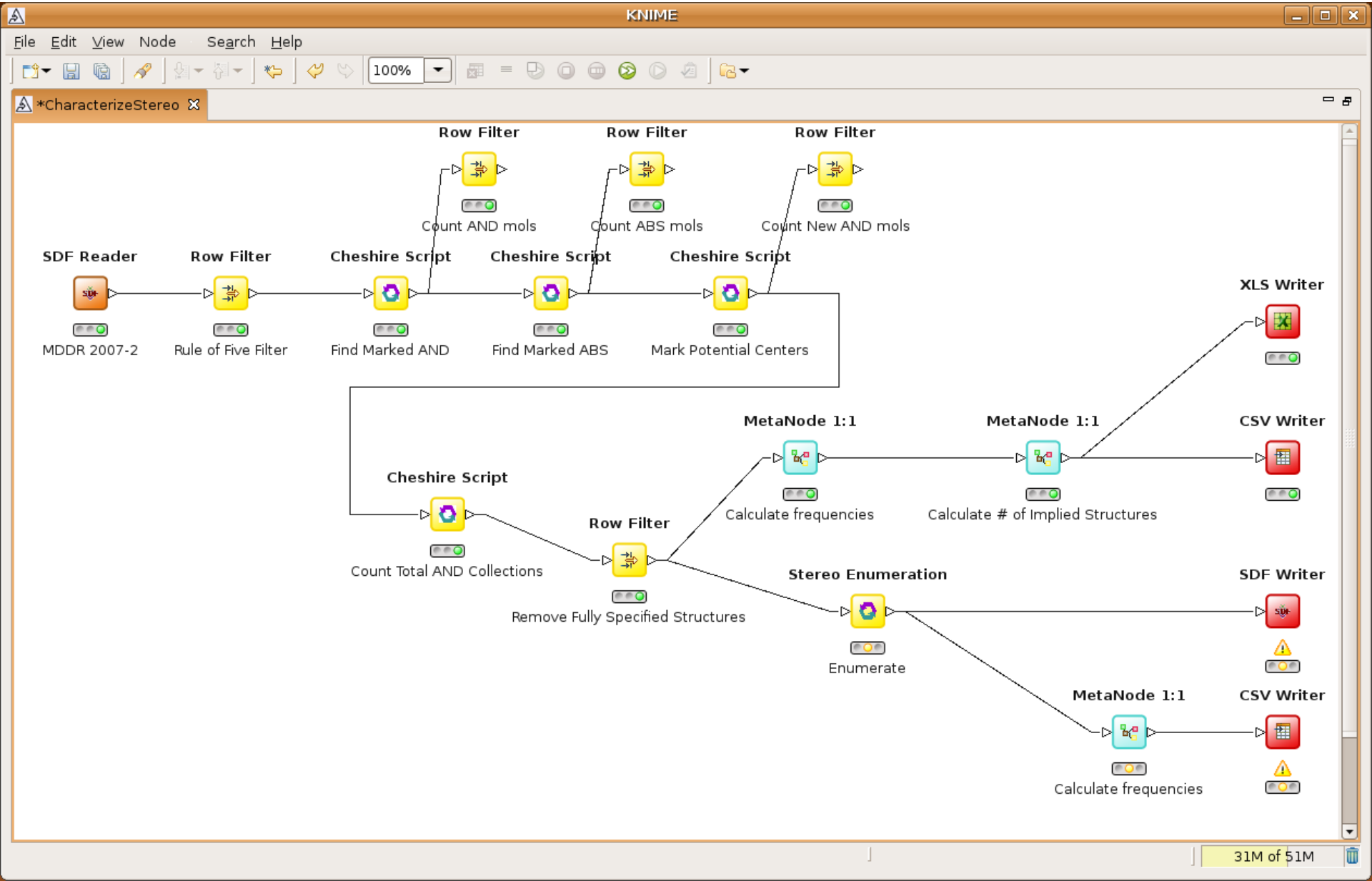
# Database Nodes

- Direct Database Molecule Search
- Direct Database Reaction Search

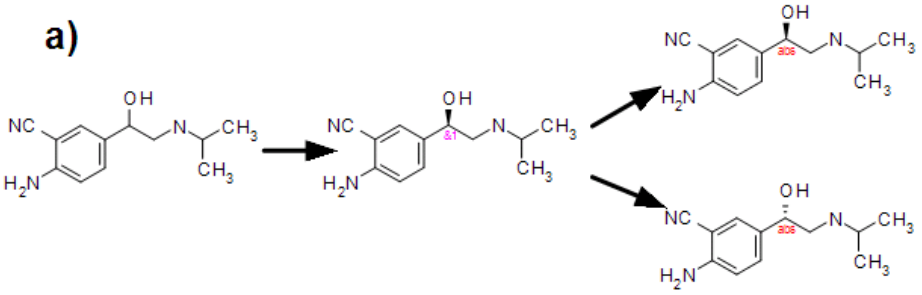
# Reaction-based Enumeration

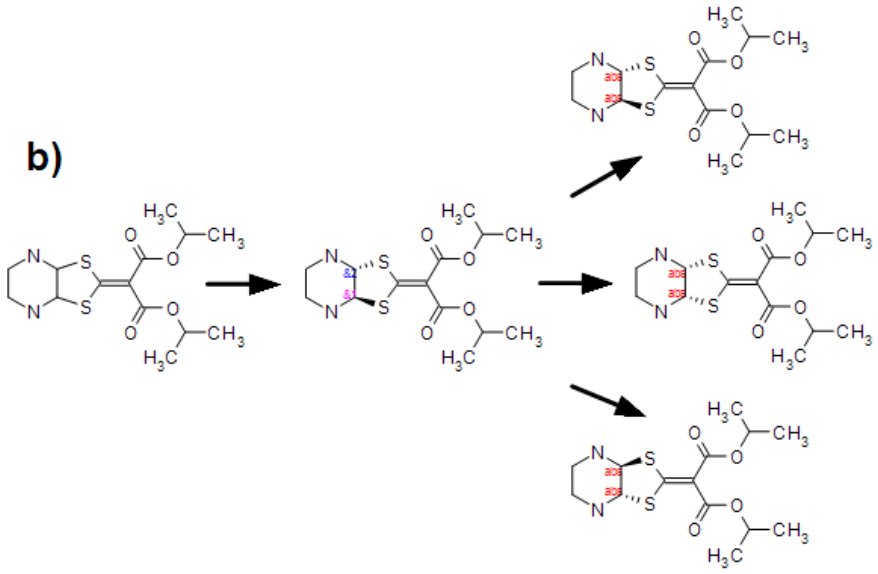


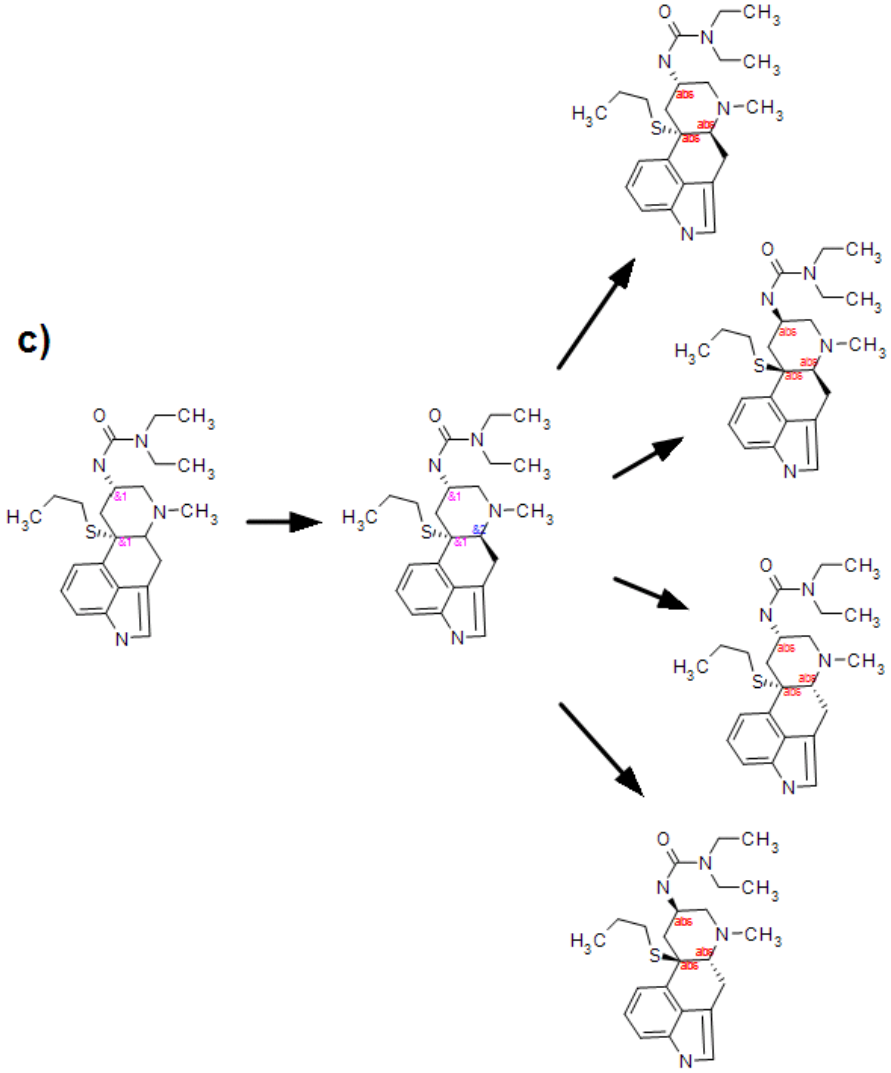
# Stereoenumeration Workflow



# Example Enumerations



# Possible New Nodes

- Structure Sketcher node
  - > Use the new Accelrys JDraw
- Support for Rgroup Decomposition/Analysis workflow
  - > Structure searching a Data Table
  - > Alternate Rgroup Decomposition

# Possible New Nodes (cont.)

- Support for Generics
  - Library enumeration
    - Support reporting of members used
    - Support filtering
  - Building Block Generation

- Support for biologics
  - UniProt to molfile
  - Generate Structure Activity tables for chemically modified biologics

- Access DiscoveryGate web service
  - Public data
    - Synonyms
    - MDL number
  - Private data with license

# Conclusion

- Accelrys Chemistry Extensions for KNIME expose Accelrys'considerable chemistry capabilities
- The Accelrys Chemistry Extensions enable real chemistry workflows
- Future releases of the Accelrys Chemistry Extensions will:
  - enable additional workflows
  - expose new Cheshire capabilities
  - expose new Direct capabilities

# Status

- Accelrys Chemistry Extensions for KNIME 2.0 released July 2009
  - Compatible with KNIME 2.0 (not with 1.x)
  - SDFReader — allows extraction of the molecule name and comments lines from the molfile
  - SDFWriter — supports writing gzipped SDfiles.
  - Rgroup Decomposition Node/Rgroup Decomposition Query Node — choose which column in the input contains the scaffold used for Rgroup decomposition
  - Rendering — supports rendering in reports using the optional BIRT nodes.

# Sample Workflows

- Reaction-based Enumeration
  - Support for sophisticated reaction definitions
    - inclusion/exclusion sites
    - “run to completion”
    - stereoenumeration
    - deduplication
  - Scalable – designed for hundreds of products, runs with millions
- Characterizing Stereochemistry
  - How many structures have what stereoinformation?
  - Enumeration of unspecified stereochemistry