Symyx Chemistry Extensions for KNIME

Keith Taylor, PhD

KNIME Life Sciences Day
February 25, 2010

Symyx®

Overview

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

Symyx

Our Goal

- Expose Symyx 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

Symyx

Chemistry Extension Features

- Core Feature
  - > Does not require any additional Symyx products
- Cheshire Feature
  - > Requires a Symyx 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)

Symyx

Core Nodes

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

Symyx

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.)

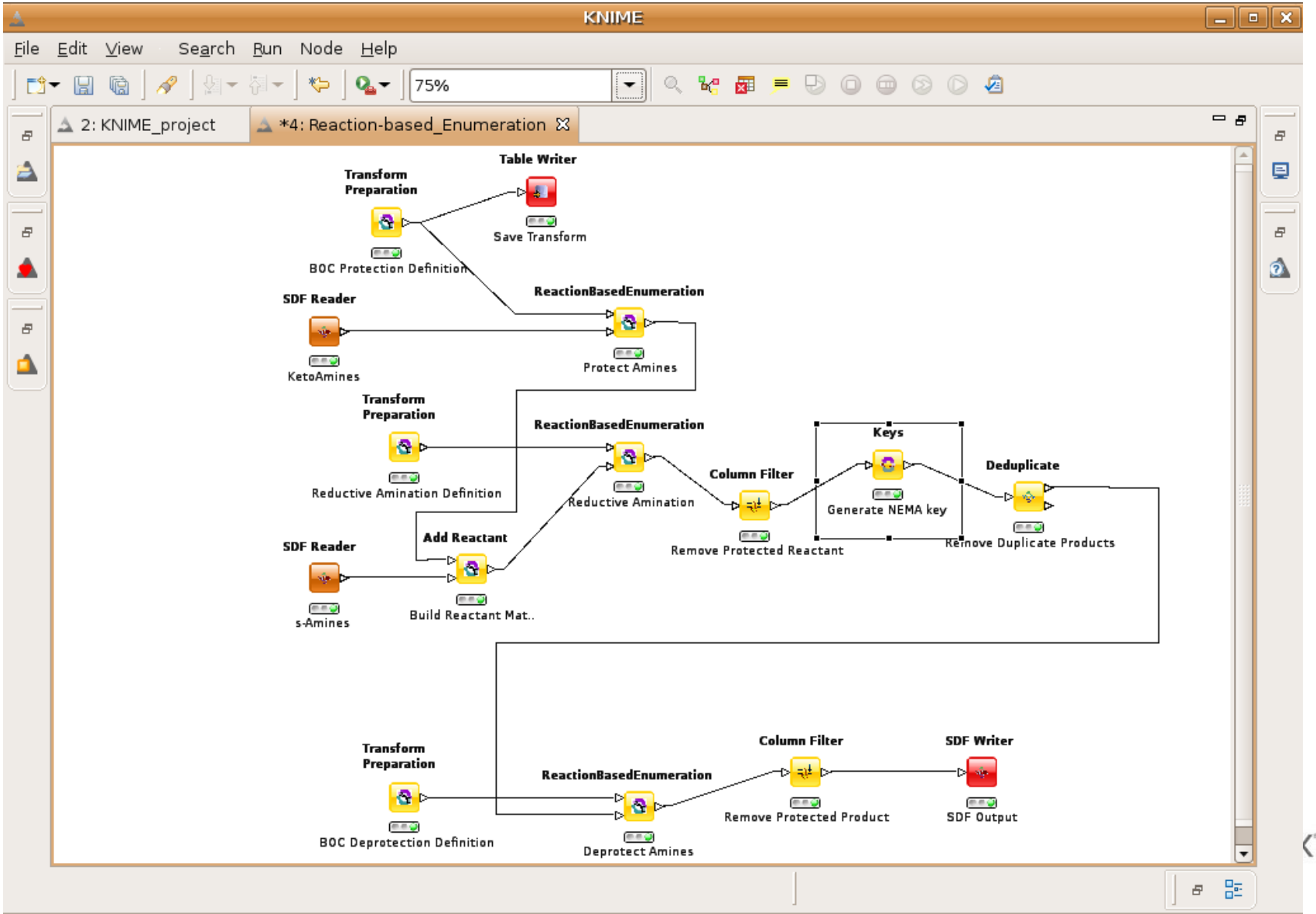
- Stereoisomer enumeration
- Keys (NEMA, SSKeys, SSKeys_subset)
- Cheshire script execution
- Remove V3000 features

Database Nodes

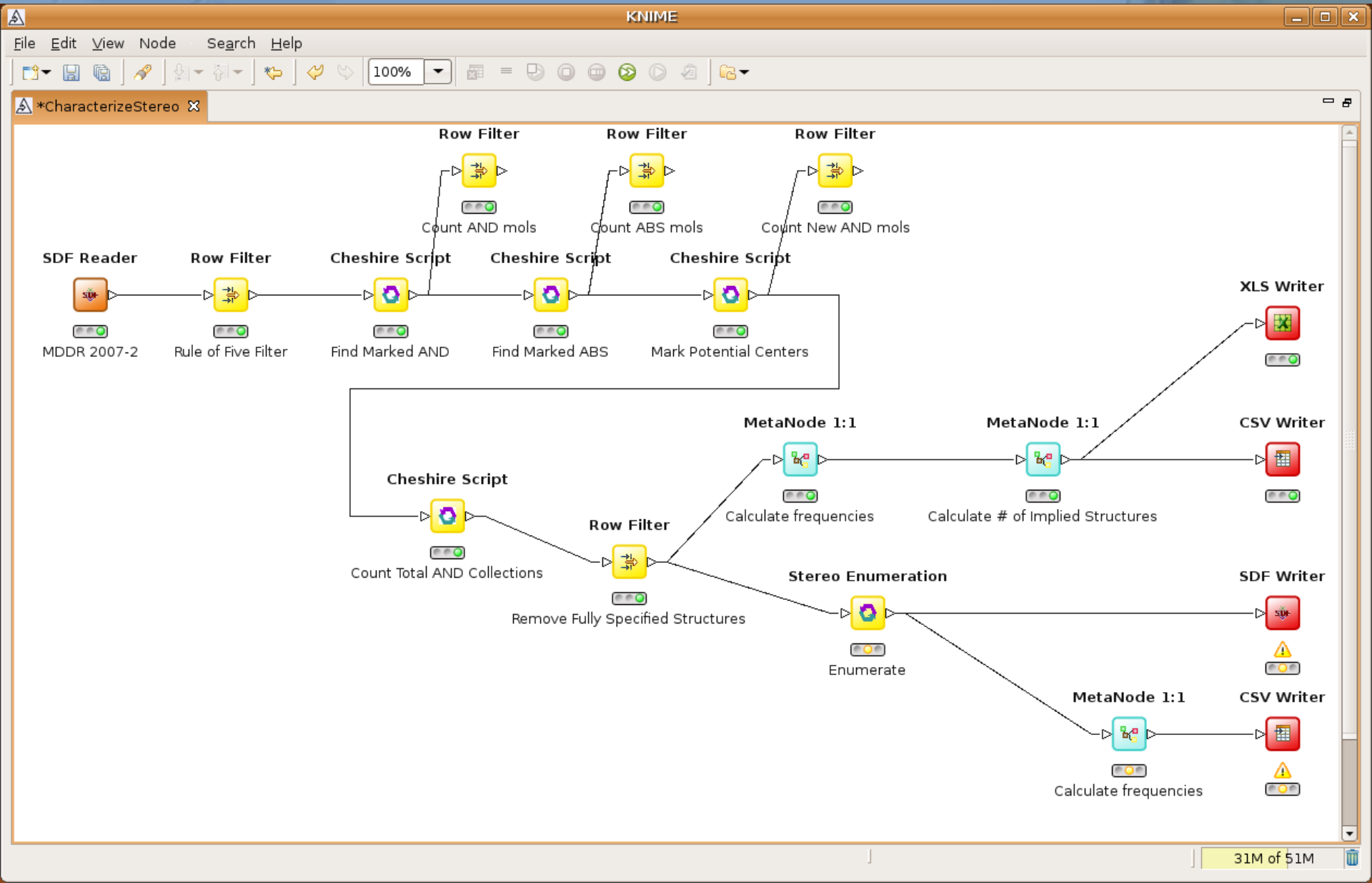
- Direct Database Molecule Search
- Direct Database Reaction Search

Symyx

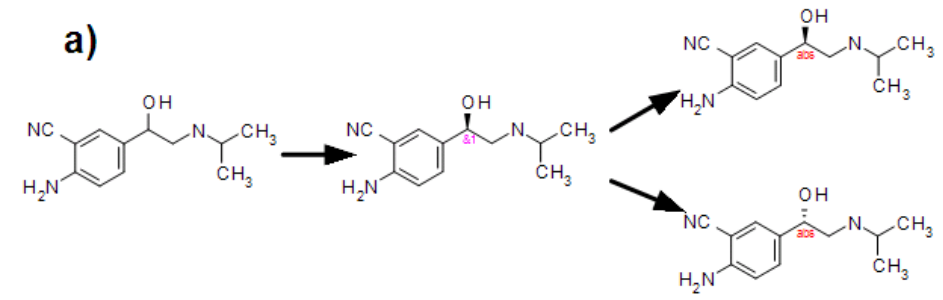
Reaction-based Enumeration

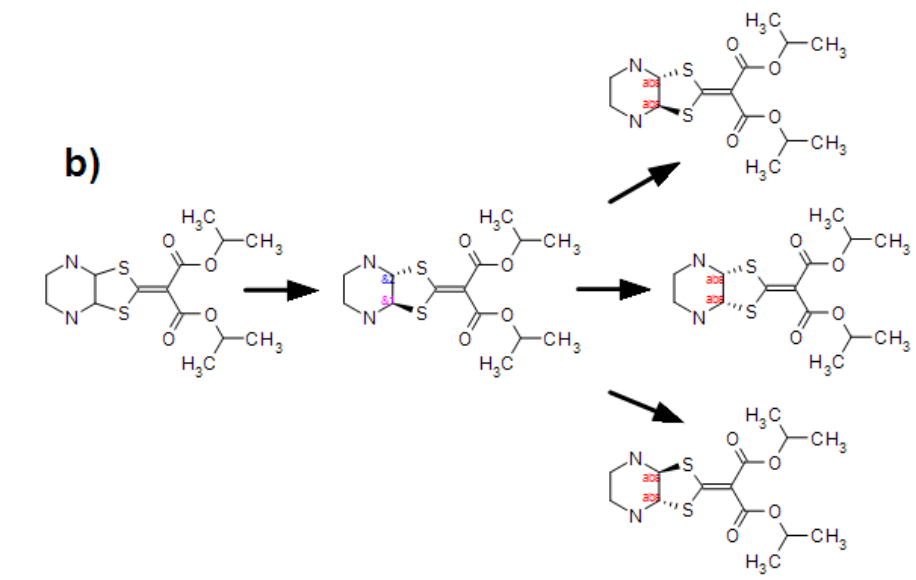


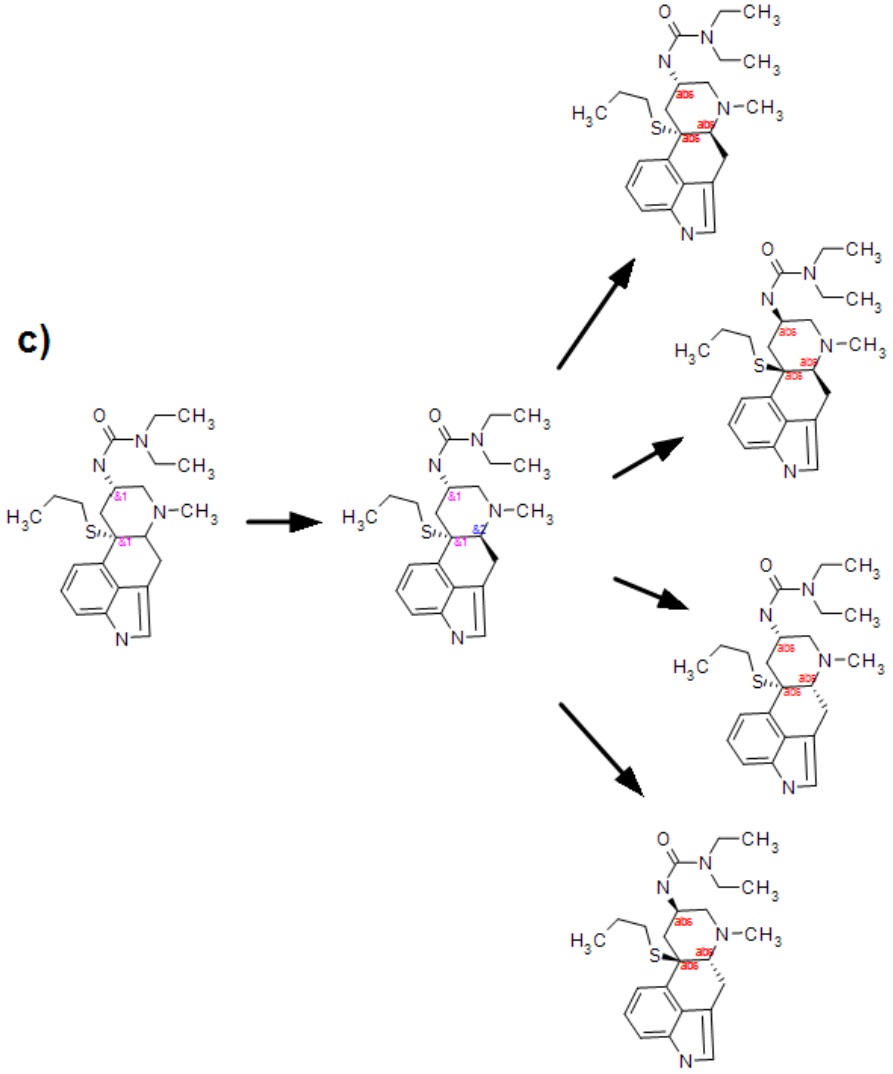
Stereoenumeration Workflow



Example Enumerations







Possible New Nodes

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

Symyx

Possible New Nodes (cont.)

- Support for new Cheshire features
  - Structure-to-Name, Name-to-Structure
  - SMILES
  - InChI name, InChI Key
- Support for Generics (Direct 6.3)
  - Library enumeration
    - Support reporting of members used
    - Support filtering
  - Building Block Generation
- Support for biologics
  - UniProt to molfile
  - Molfile to UniProt

Symyx

Conclusion

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

Symyx

Status

- Symyx 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.

Symyx

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

Symyx