

A ChemAxon/KNIME based tool for designing chemical libraries

Tim Parrott
Dart NeuroScience
September 25, 2013

Background

Dart NeuroScience (DNS)

200+ Scientists

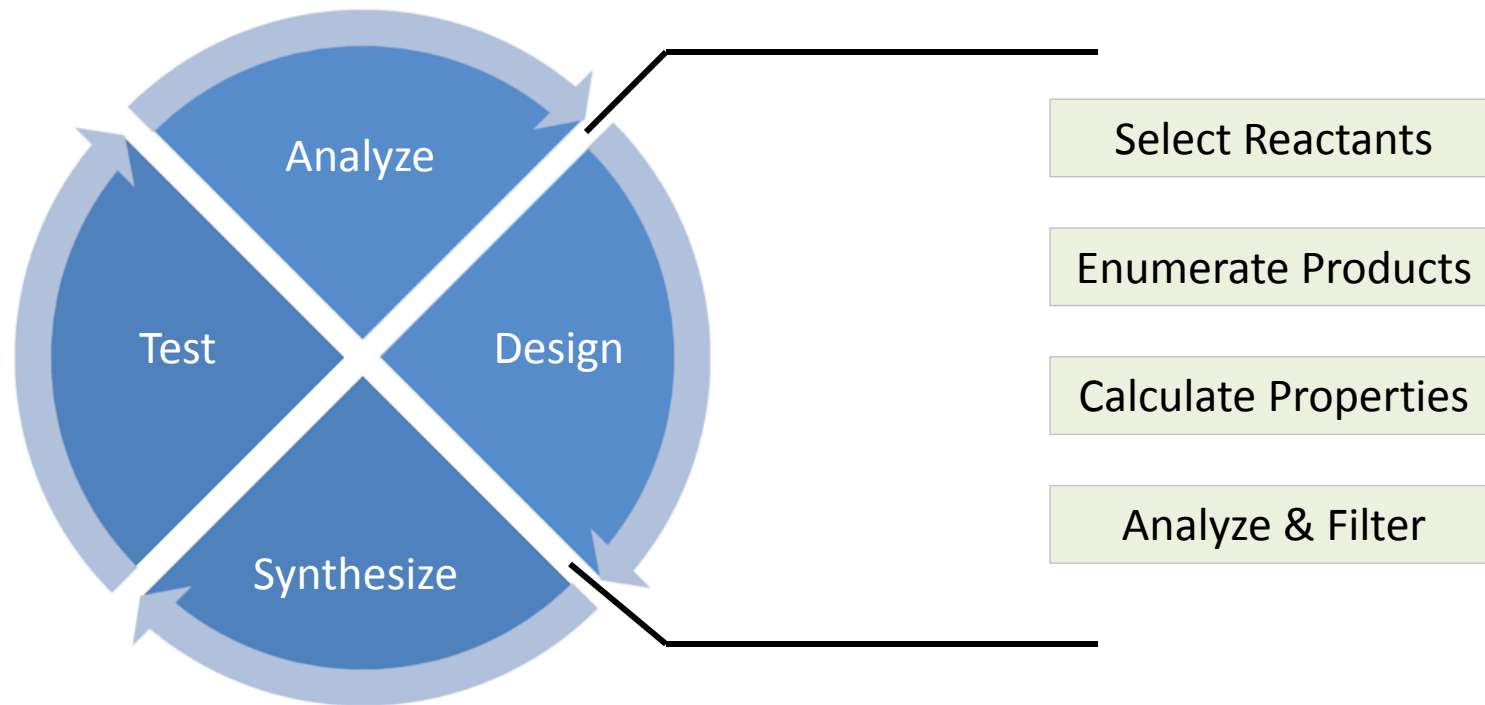
50+ Chemists

Parallel Synthesis Group

About 20 chemists involved in the design and creation of chemical libraries

We need a chemical library design tool !

A Basic Chemical Library Design Tool



Goals

Constraints

Limited IT/IM support

Time

Chemists already on software overload

Approach

Standardize calculations & reactions (services)

Simplify: wrap processes and minimize import/export operations

Enhance capabilities and speed by doing calculations remotely



Ease of Use



Support

=



Productivity



Platforms



Tool Overview

The image shows a screenshot of the KNIME software interface. On the left, a 'Node Repository' panel lists various nodes, including 'DNS Nodes' and 'Reactor Nodes'. A callout box labeled 'Custom Nodes' points to the 'Export For Spotfire' node in the repository. In the main workspace, a 'Spotfire Export' callout points to the 'Export For Spotfire' node in the 'Personal favorite nodes' list. On the right, a 'Selection & Configuration Panel' callout points to a blue button with the DNS logo and the text 'Select a DNS Node or drop a new one on the canvas'.

Node Repository

- DNS Nodes
 - Reactor Nodes
 - BiReactor
 - TriReactor
 - UniReactor
 - Batch Details
 - Cluster
 - Deduplication
 - Diversity Elements
 - Export For Spotfire
 - Library Plan Publisher
 - RN Lookup
 - ROCS Calc
 - Remote CalcP
 - Remove Additive
 - Stereochemistry Code Assignment
 - Structure Lookup
 - VIDA Launcher

Custom Nodes

Spotfire Export

Selection & Configuration Panel

Select a DNS Node or drop a new one on the canvas

Reactant Selection

*3: DEMO

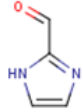
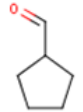
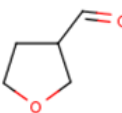
Node Description DNS Node Inspector

Reactant Classes: Aldehydes

0 of 226 selected

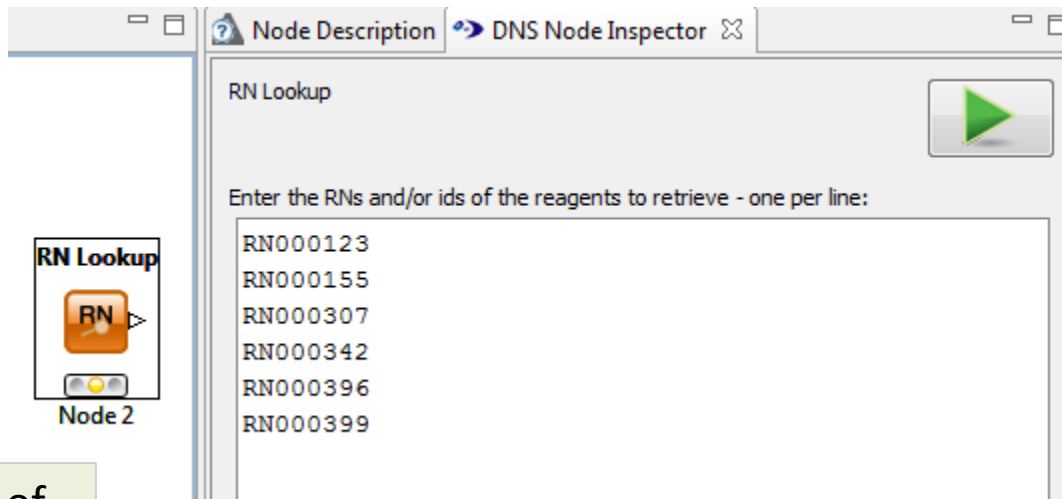
Diversity Elements

Aldehydes

Aldehydes Reactant Id	Aldehydes Structure	Aldehydes Stereo Info
<input type="checkbox"/> RN004001		achiral
<input type="checkbox"/> RN004010		achiral
<input type="checkbox"/> RN004011		racemic

Import curated
classes of
reactants

Reactant Selection



Import list of
Reagent
Numbers

Reactant Deduplication

Diversity Elements



Amines

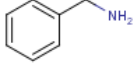
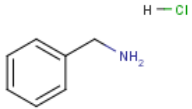
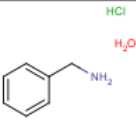
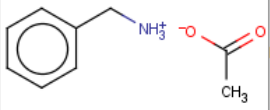
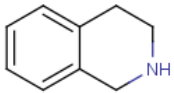
Deduplication



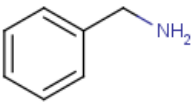
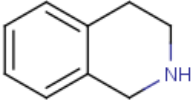
Deduplication
(Amines Structure)

Need to identify and remove functionally equivalent reactants

Input

Molecule	S ID
	RN000123
	RN000456
	RN000789
	RN000911
	RN001000

Output

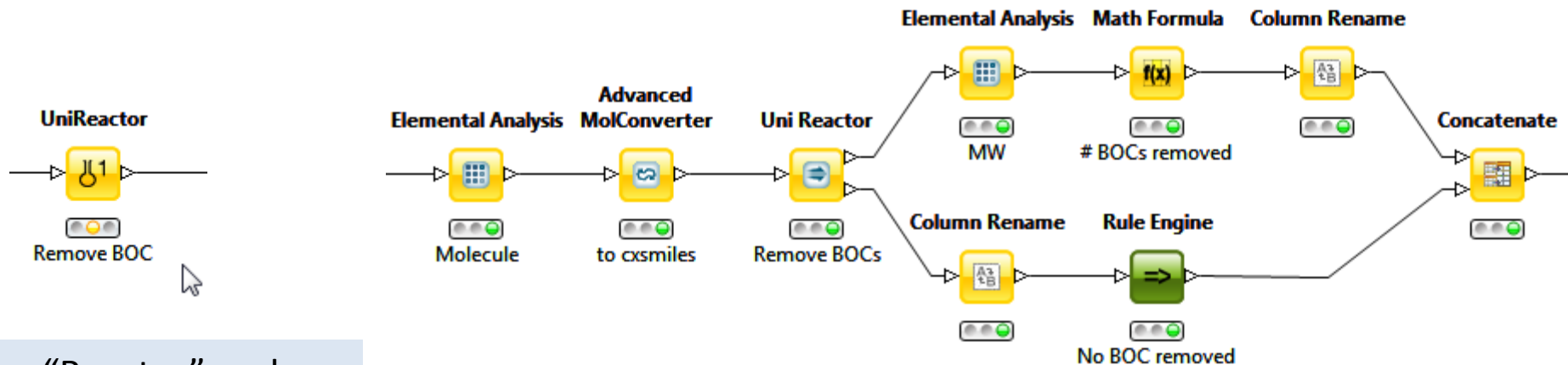
Product Molecule - Deduplication ...	S ID	S Removed IDs
	RN000123	RN000456,RN000789,RN000911
	RN001000	Unique

Reaction Selection

The screenshot displays a software interface for reaction selection. On the left, a workflow diagram shows 'RN Lookup' and 'Diversity Elements' (Aldehydes) feeding into a 'BiReactor' node. The 'BiReactor' node is labeled 'Node 4' and 'Reductive Amination'. On the right, the 'Available Protocols' section lists three reaction types:

- Reductive Amination**: Shown with a chemical reaction between benzaldehyde and a secondary amine to form a secondary amine derivative. Expected inputs: Amines, Aldehydes.
- Sonogashira Coupling**: Shown with a chemical reaction between an aryl iodide and an alkyne to form an aryl-alkyne. Expected inputs: Aryl and vinyl halides, Acetylenes.
- Stille Coupling**: Partially visible at the bottom.

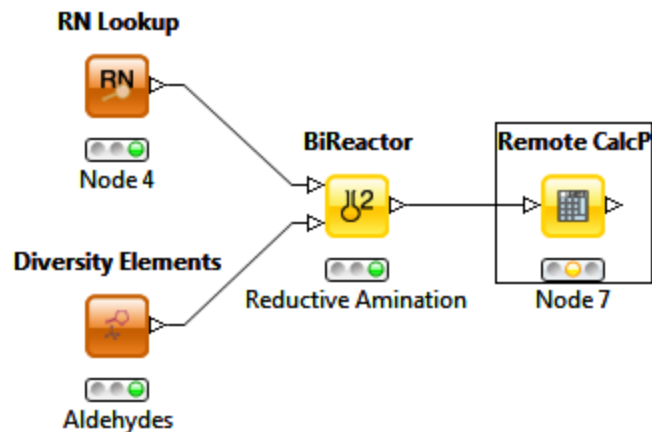
Reactions: A Look under the Hood



“Reactor” nodes can contain multi-step workflows.

S Aldehydes...	S Primary Amine...	S Isocyanates...	S Number_of_BOC_groups_removed
RN001298	RN001235	RN000934	1.0
RN001298	RN002207	RN000934	1.0
RN002071	RN000663	RN000752	0

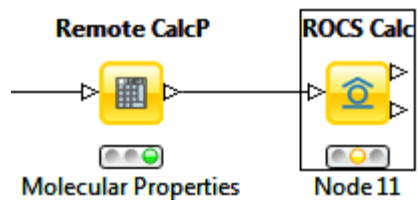
Calculations



The screenshot shows the 'DNS Node Inspector' window. The 'Input structure' dropdown is set to 'Product Molecule - Reductive A...'. A green play button is visible in the top right corner. Below the input field, there is a section titled 'All Properties' with a checkbox. A list of properties is displayed, each with a checkbox:

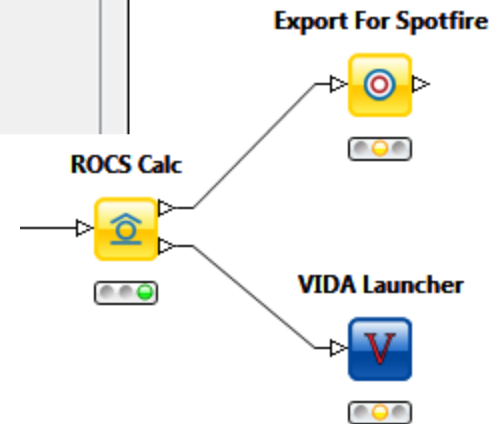
- Molecular Weight
- Monoisotopic Mass
- ClogP
- TPSA (7.4)
- Strongest Acidic pKa
- Strongest Basic pKa
- HBA atoms (parent)
- HBDs (parent)
- Molecular Formula
- Elemental Composition
- logD (7.4)
- Charge (7.4)
- Second Strongest Acidic pKa
- Second Strongest Basic pKa
- HBA atoms (7.4)
- HBDs (7.4)

Calculations --- OpenEye ROCS



Protocols Target and Binding Mode	
Target:	Binding Mode:
ALL JNXS	ROFL1b: 12 2D Queries (Jan 2013) Open wiki
JNX1	
JNX2	
JNX5	
LOL2	
LOL3	
ROFL1b	ROFL1b: 14 3D Queries (Jan 2013) Open wiki

ROCS output includes the Shape/Pose that scored best and the Tanimoto Score against that query.



Clustering

Cluster
Node 12

Node Description DNS Node Inspector

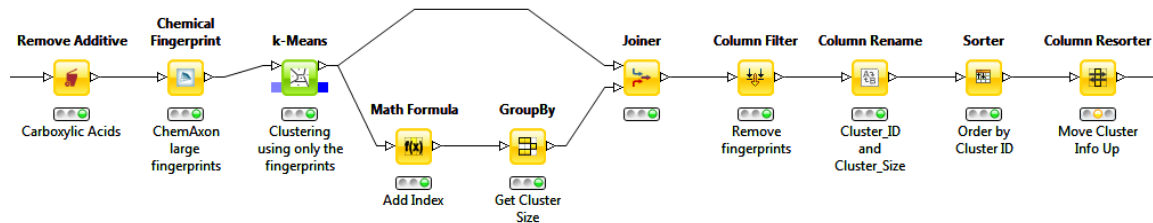
Input structure: Carboxylic Acids Structure

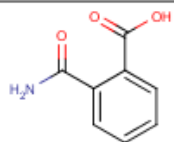
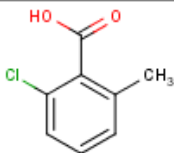
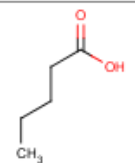
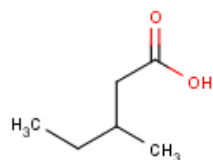
Clustering Methods and Parameters

Method: K-Means Clustering

Clusters: 50

Max Iterations: 99



Carboxylic Acids Structure	Cluster_ID	Cluster_Size
	cluster_15	9
	cluster_15	9
	cluster_16	12
	cluster_16	12

Server-side Execution

- Use existing services (property calculations)
- Parallelized across multiple processors
- Queuing system

Pausing Local Execution



Ready to Execute



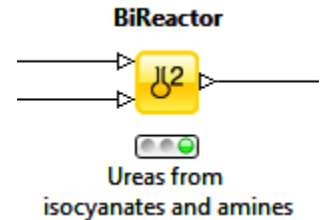
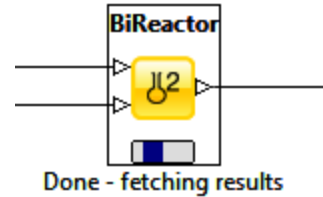
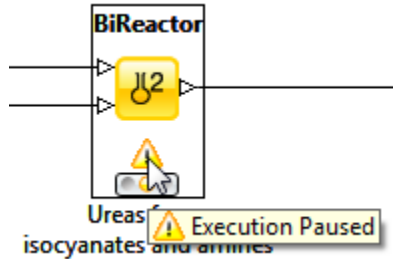
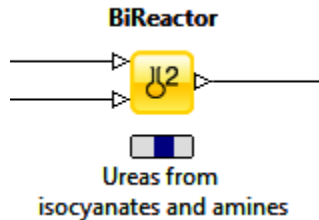
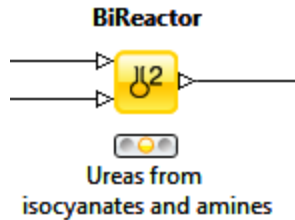
Executing – can't be paused yet



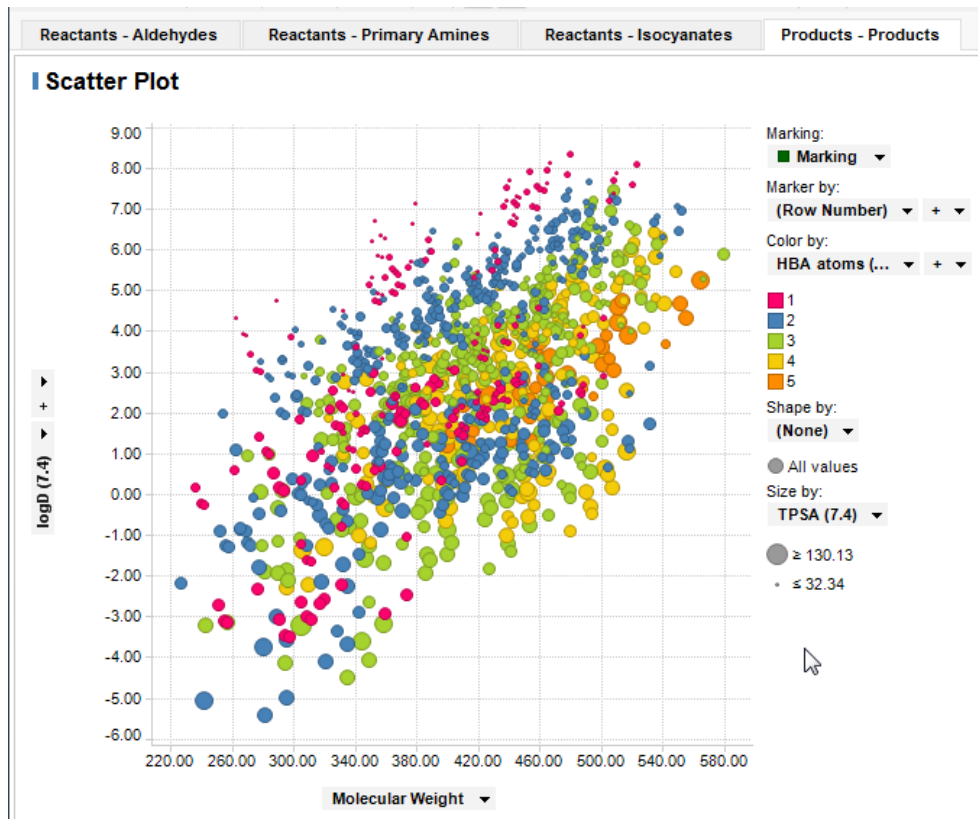
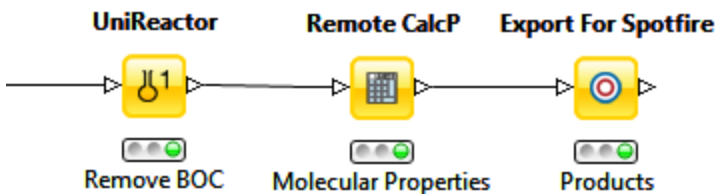
Executing – can be paused



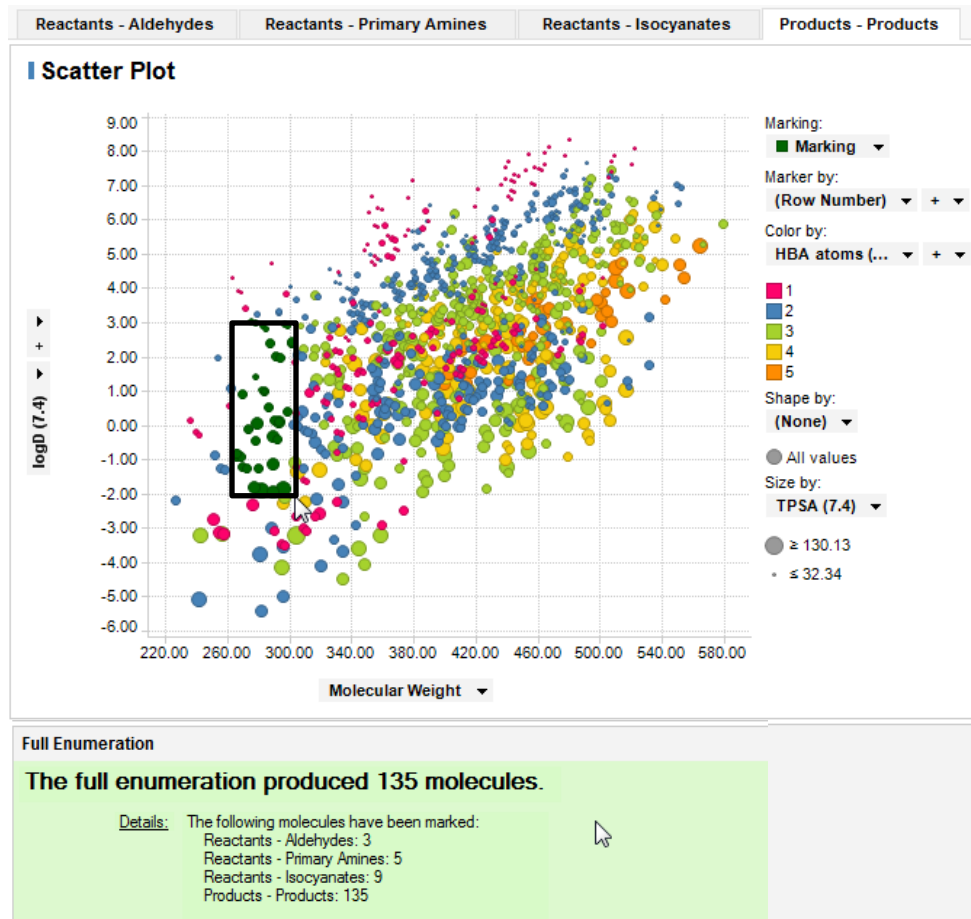
Local Execution has been paused



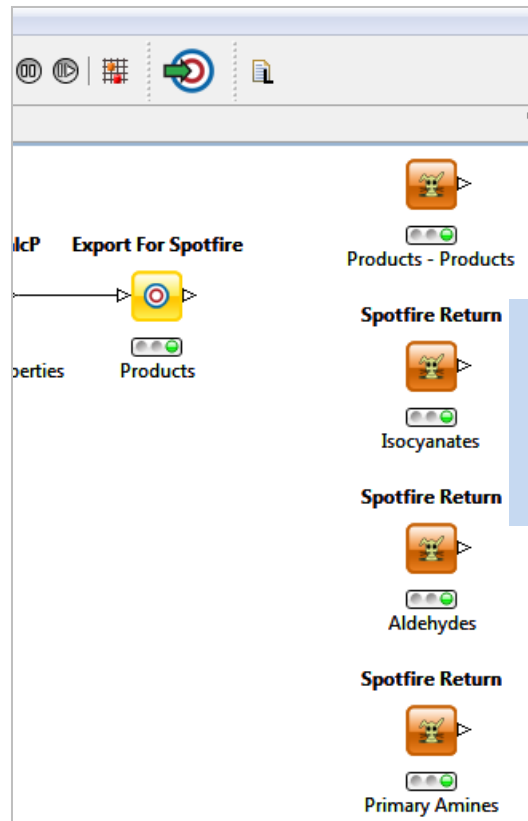
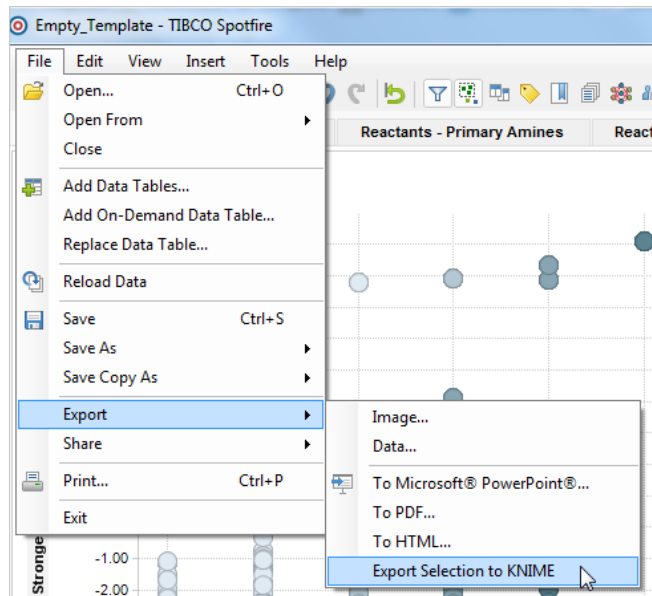
Export to Spotfire



Selections made in Spotfire

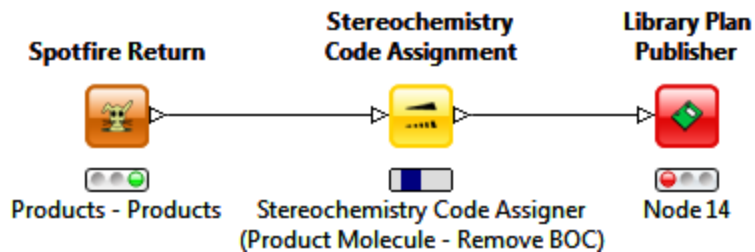


Spotfire Selections returned to KNIME



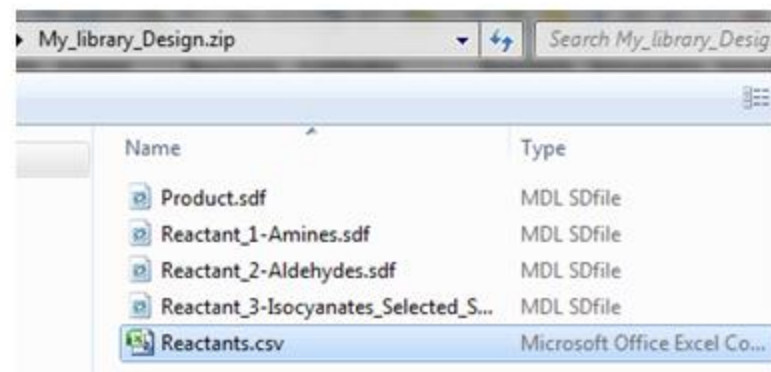
New nodes with selected products & reactants appear in KNIME

Final Steps



stereo_category_code
achiral
achiral
oneenant
racemic

Stereochemical codes needed for registration are assigned based on structure.



The library design plan contains separate sdf files for the products and each reactant, along with a .csv file listing how many times each reactant is used. The zipped file is parsed on import into a chemist's electronic laboratory notebook.

Deployment

Local install of KNIME for each user

DNS Node update site to push out updates

Turn off KNIME and InfoCom updates to keep consistent versions across all users

Spotfire updates pushed out to users

Service updates are largely invisible to users

Summary

- June 2011 Parallel Synthesis Group formed
- June 2012 First release of Library Design Tool (LDT)
- Sept 2012 Additional KNIME training
- November 2012 Second release (Clustering, ROCS)
- April 2013 Pausable Nodes, Deduplication
- August 2013 RN Lookup, Stereo Code Assigner
40 Total Reactions

Acknowledgments

Node Development

loki der quaeler

Services & Deployment

Ron Blanford

Karen Do

Kenny Leung

Zach Young

Daniel Garden

Testing and troubleshooting

Eileen Tompkins

Andrew Burritt

The SGC Team

Management & PM

Melanie Nelson

Heather Jones

Brock Luty