A ChemAxon/KNIME based tool for designing chemical libraries

Tim Parrott
Dart NeuroScience
September 25, 2013

Background

Dart NeuroScience (DNS)

200+ Scientists

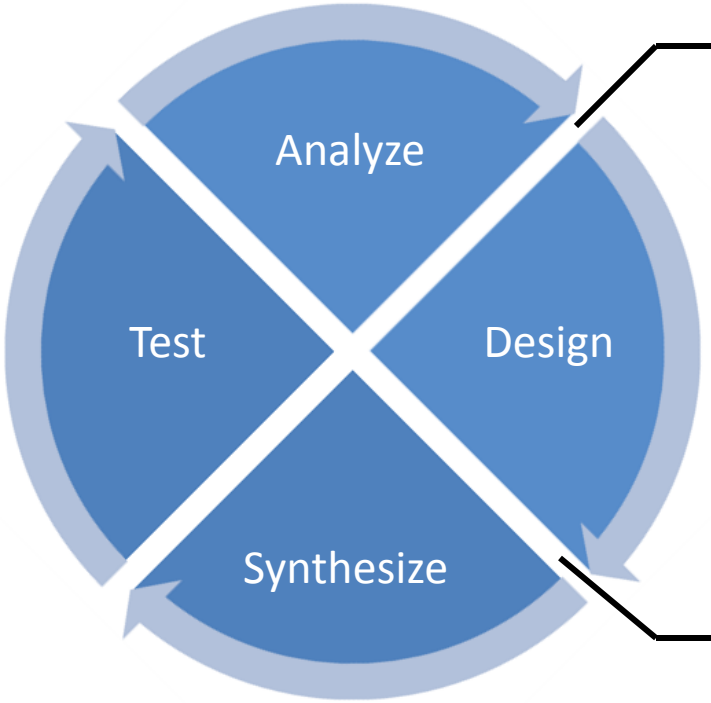
50+ Chemists

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



- Select Reactants
- Enumerate Products
- Calculate Properties
- Analyze & Filter

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

Platforms


Spotfire

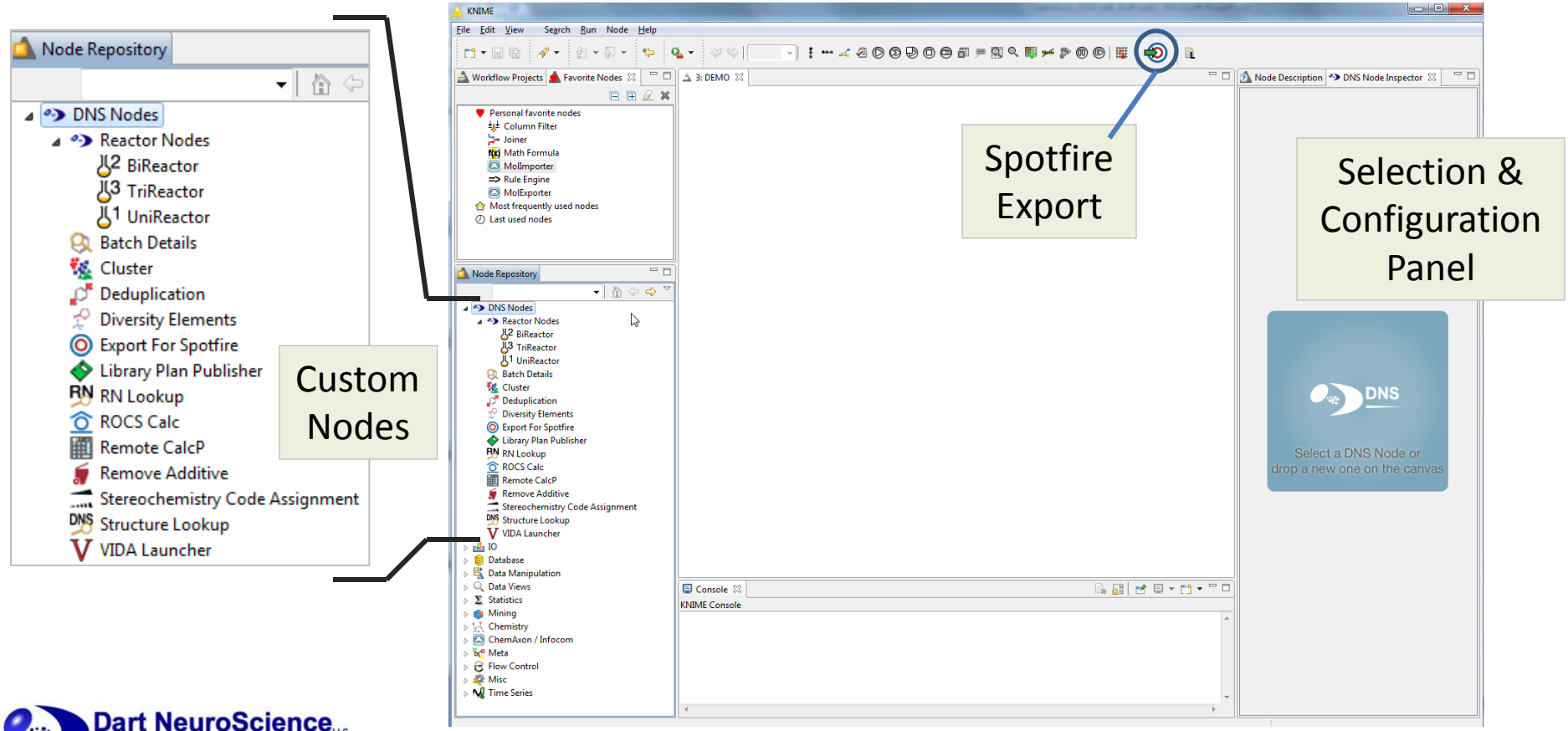

OpenEye
Scientific Software


KNIME


infocom

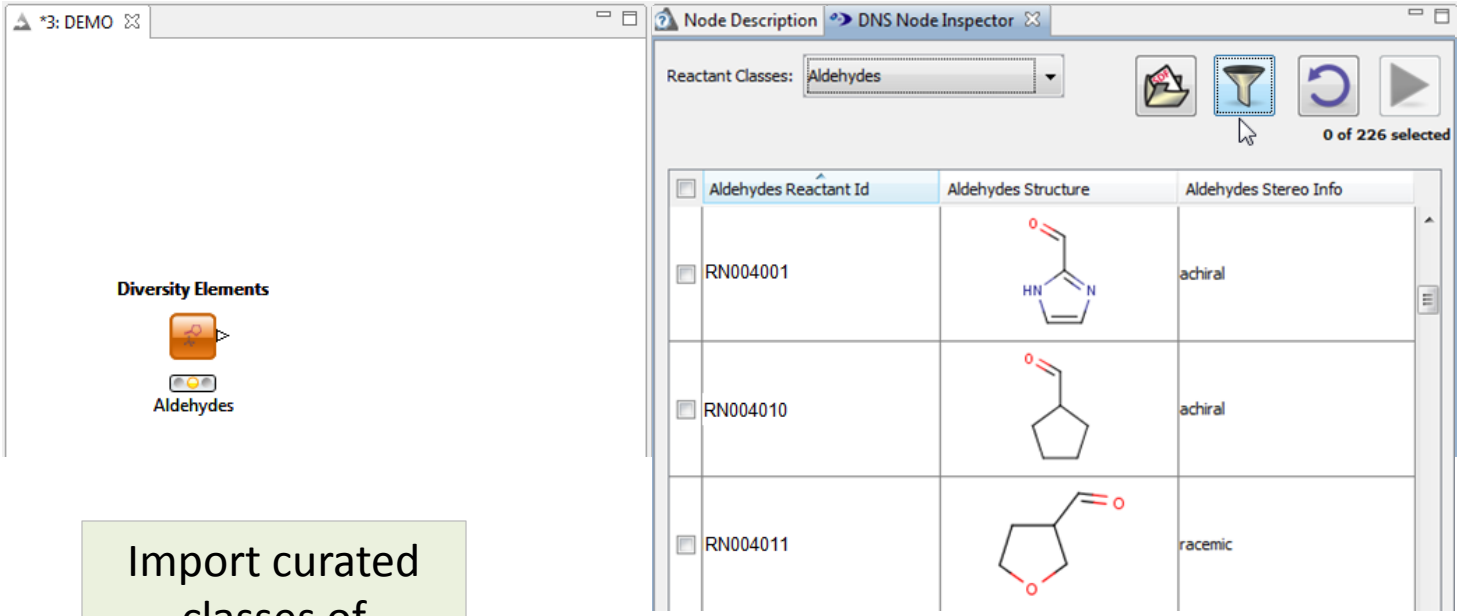

ChemAxon

Tool Overview



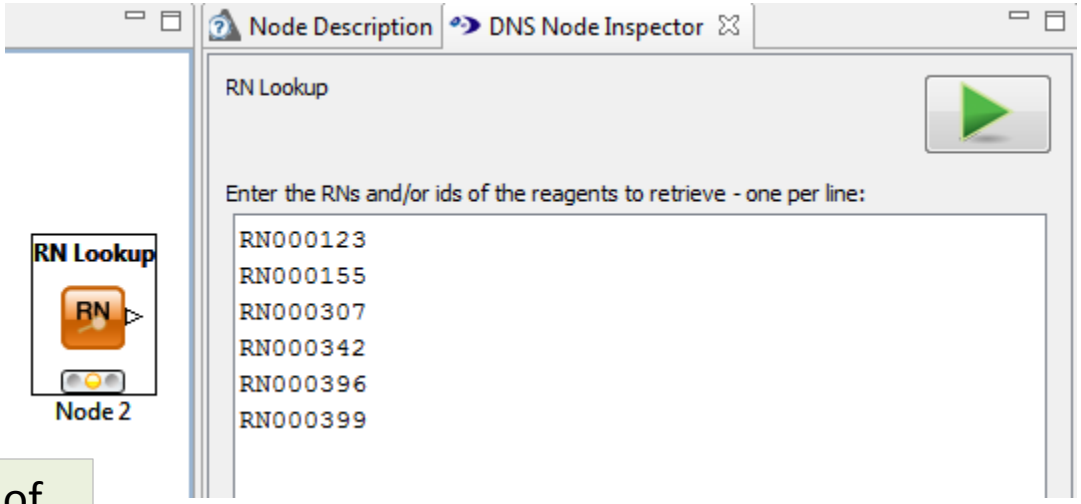
Dart NeuroScience LLC
quality of mind

Reactant Selection



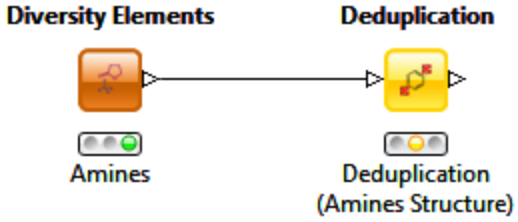
Import curated classes of reactants

Reactant Selection



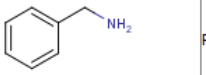
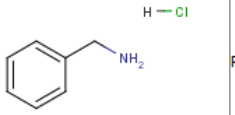
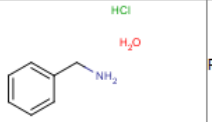
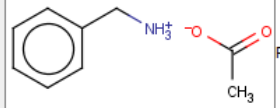
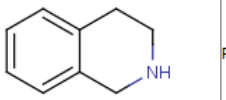
Import list of Reagent Numbers

Reactant Deduplication

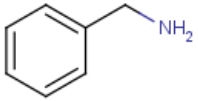
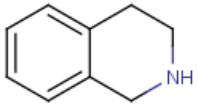


Need to identify and remove functionally equivalent reactants

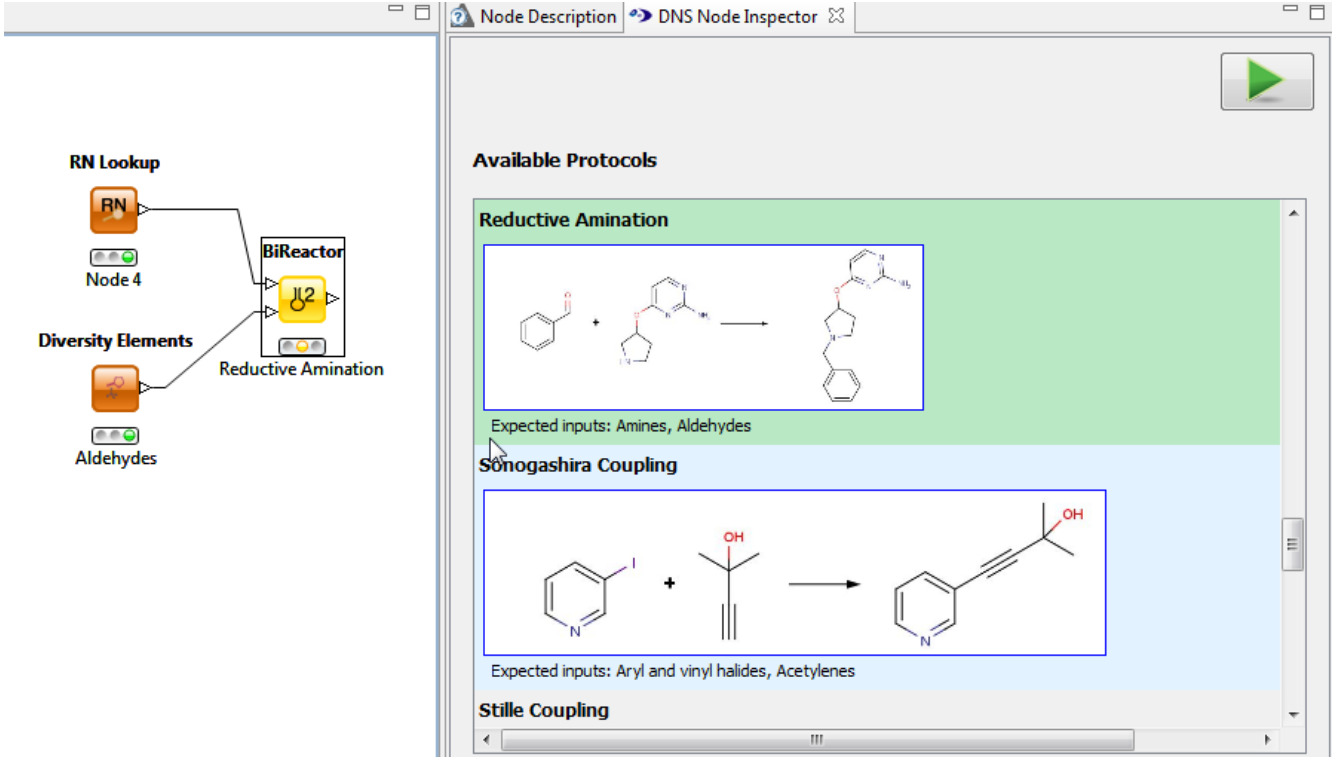
Input

Molecule	ID
	RN000123
	RN000456
	RN000789
	RN000911
	RN001000

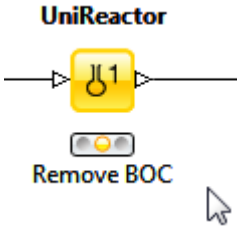
Output

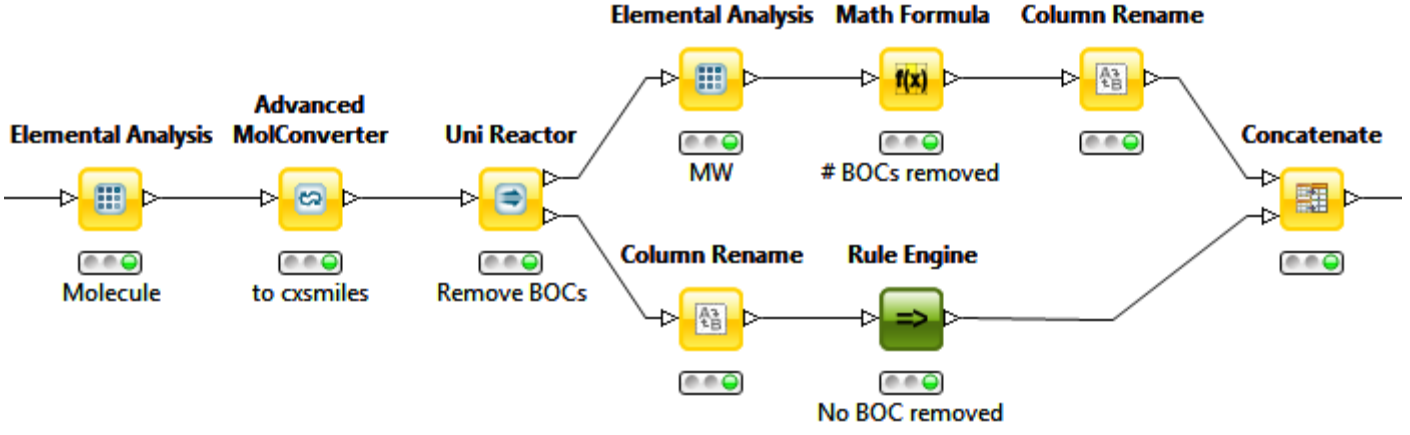
Product Molecule - Deduplication ...	ID	Removed IDs
	RN000123	RN000456,RN000789,RN000911
	RN001000	Unique

Reaction Selection



Reactions: A Look under the Hood

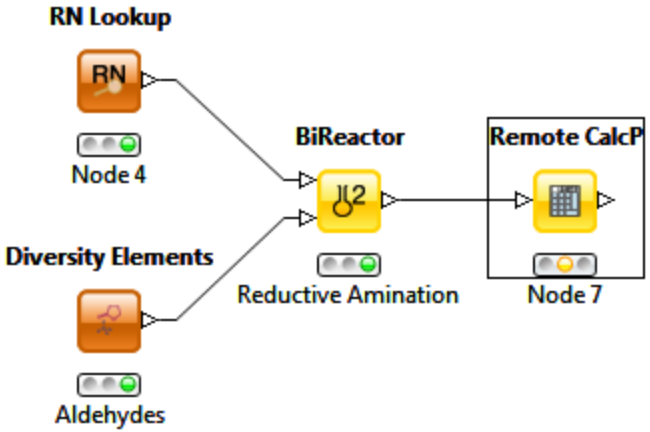




"Reactor" nodes can contain multi-step workflows.

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

Calculations

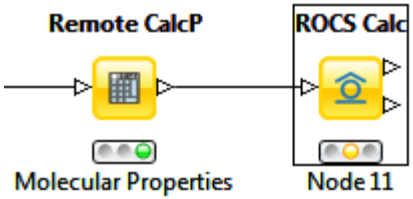


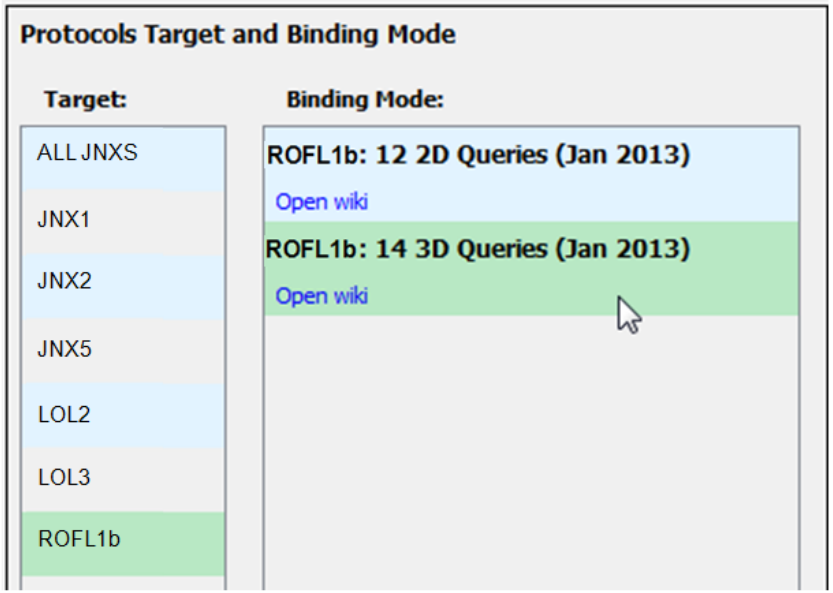
Node Description | DNS Node Inspector

Input structure: Product Molecule - Reductive A...

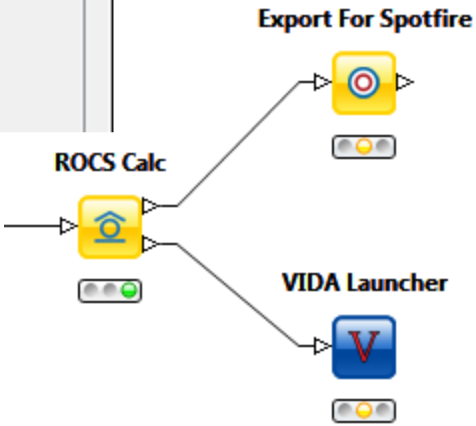
All Properties	
☐ Molecular Weight	☐ Molecular Formula
☐ Monoisotopic Mass	☐ Elemental Composition
☐ ClogP	☑ logD (7.4)
☑ TPSA (7.4)	☐ Charge (7.4)
☐ Strongest Acidic pKa	☐ Second Strongest Acidic pKa
☐ Strongest Basic pKa	☐ Second Strongest Basic pKa
☐ HBA atoms (parent)	☐ HBA atoms (7.4)
☐ HBDs (parent)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

Calculations --- OpenEye ROCS



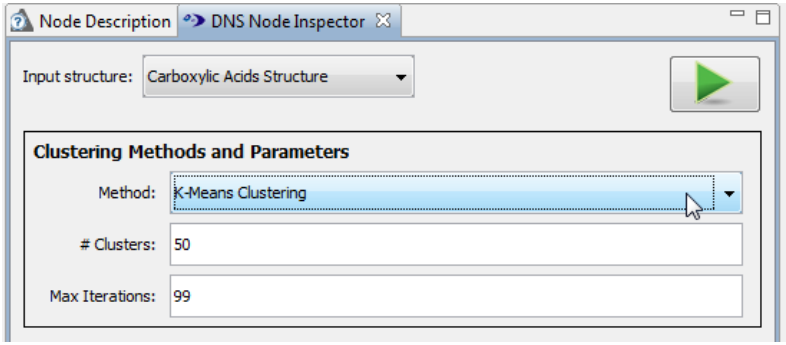


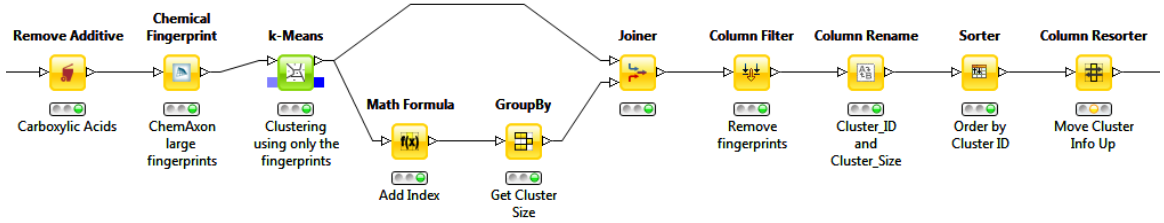
ROCS output includes the Shape/Pose that scored best and the Tanimoto Score against that query.

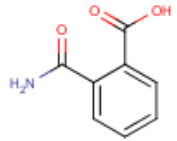
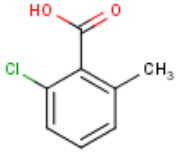
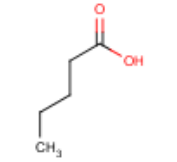
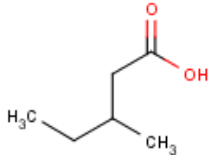


Dart NeuroScience LLC
quality of mind

Clustering





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

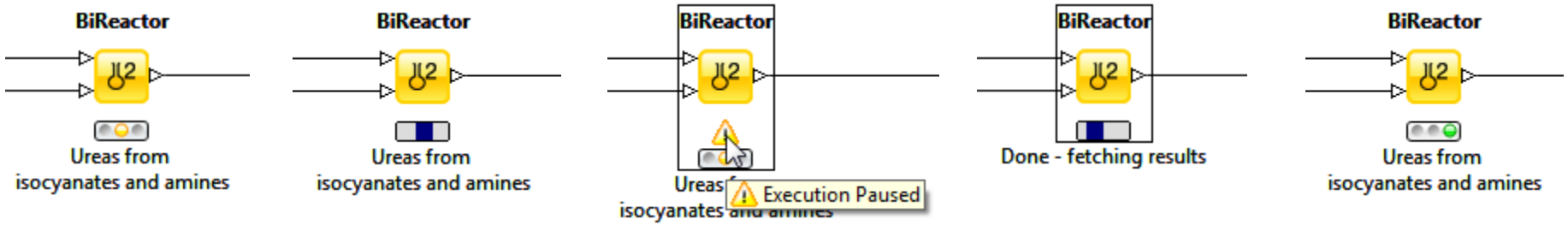
Pausing Local Execution

Ready to Execute

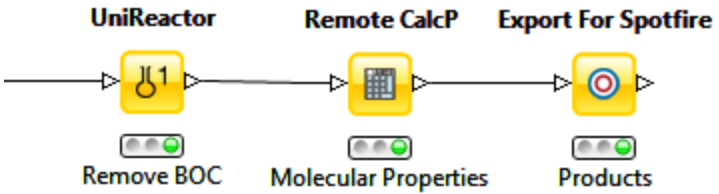
Executing – can't be paused yet

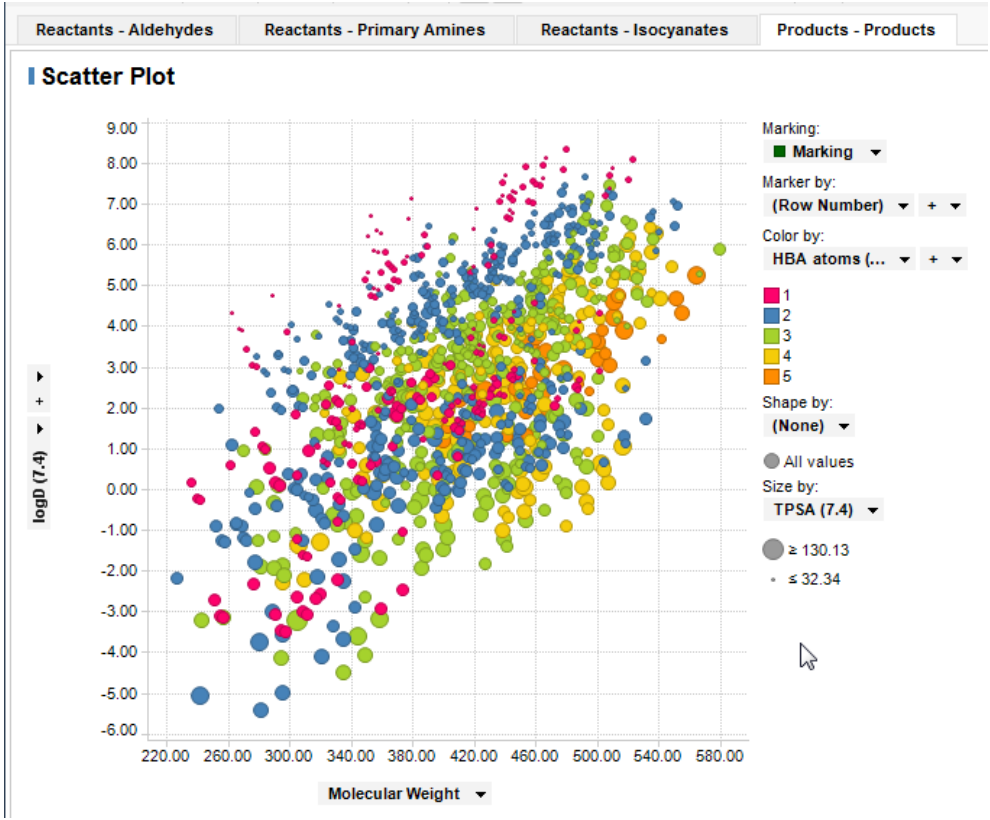
Executing – can be paused

Local Execution has been paused



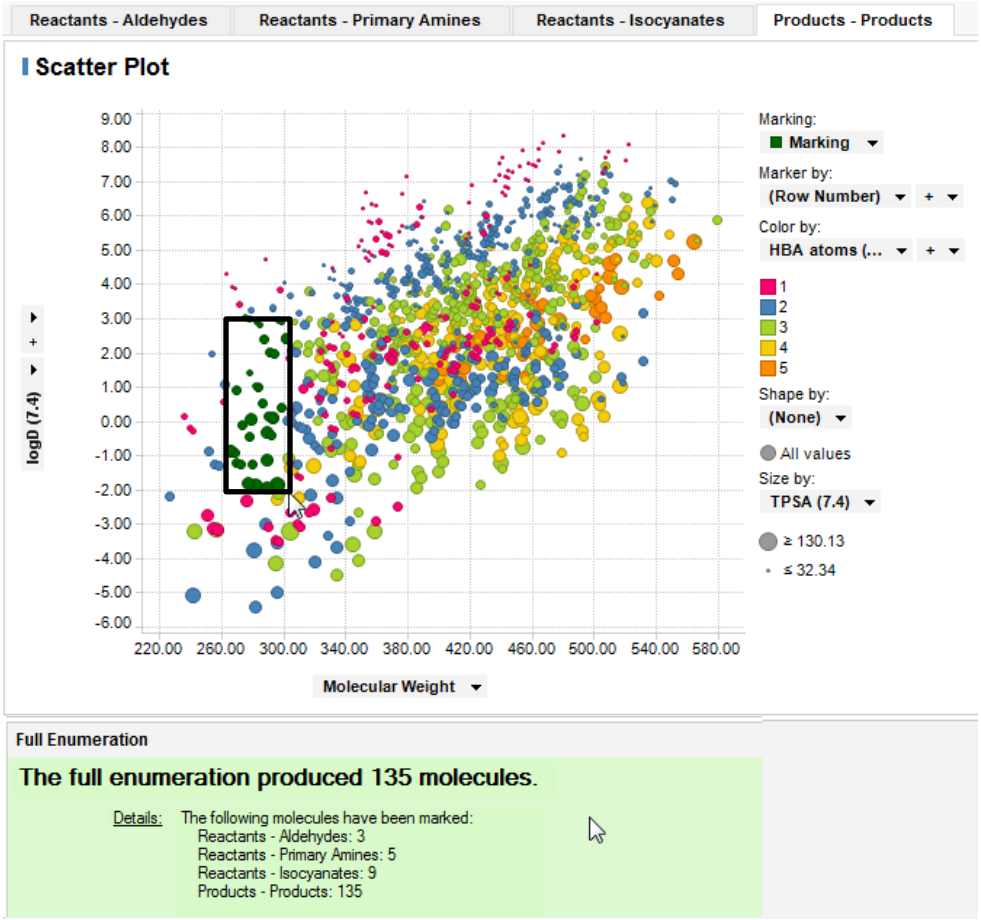
Export to Spotfire





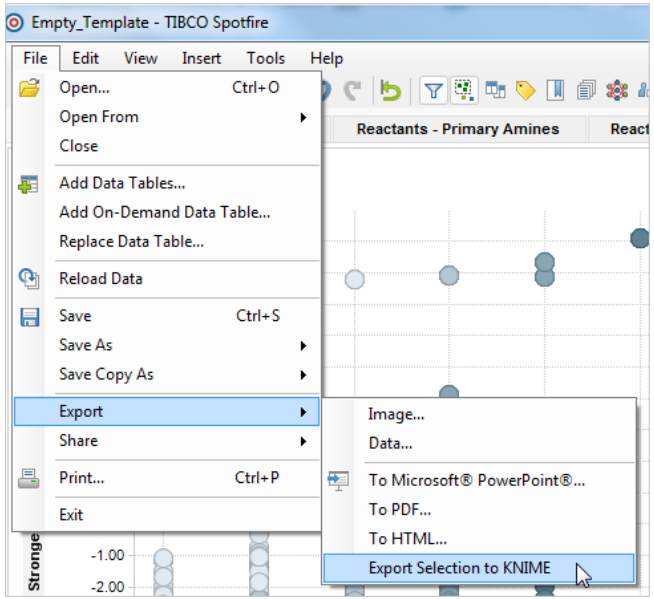
Dart NeuroScience LLC
quality of mind

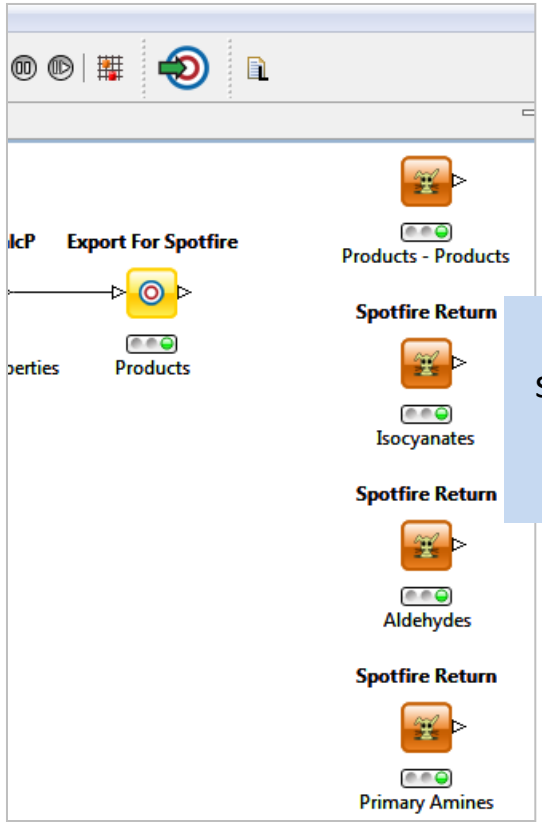
Selections made in Spotfire



Dart NeuroScience LLC
quality of mind

Spotfire Selections returned to KNIME

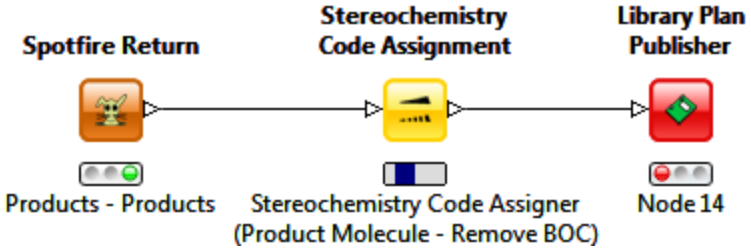




New nodes with selected products & reactants appear in KNIME

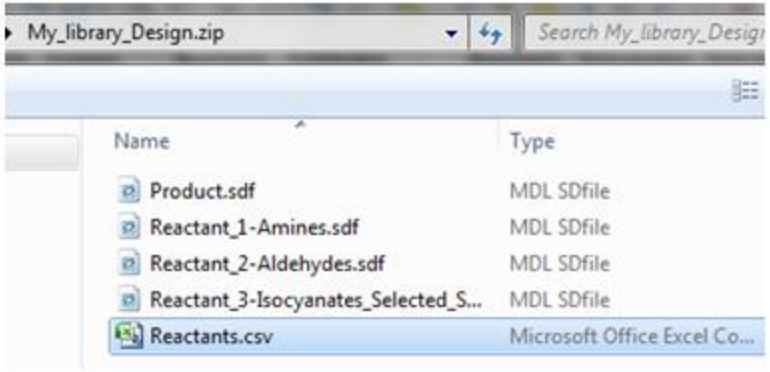
Dart NeuroScience LLC
quality of mind

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