



The
University
Of
Sheffield.

Knowledge Based *De Novo* Design using Reaction Vectors

Hina Patel
University of Sheffield

Talk Overview

- *De novo* design
- Reaction vectors
- Application of reaction vectors as transforms
- Reaction cleaning
- A desktop tool for generating structures using one-step transforms
- Conclusions & future work



De Novo Design

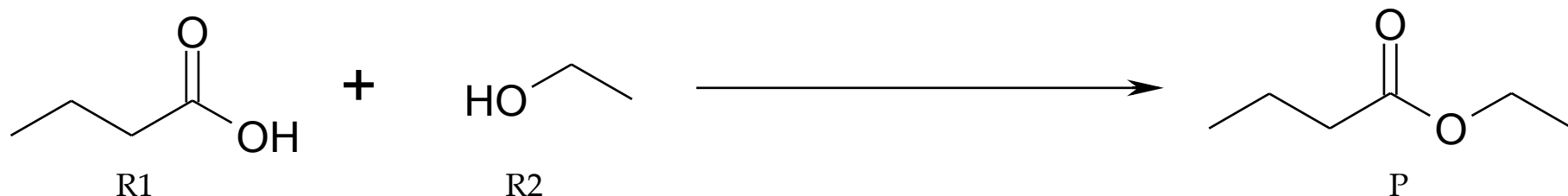
- Generation of virtual lead compounds
- Ideally should be soluble, non-toxic, synthetically feasible etc
- Limitations of current tools:
 - Restrictions in computing power
 - Lack of synthesisable structures



Aim

- Design a tool for generating synthetically feasible structures using reaction vectors

Reaction Vectors¹



I	1	2	3	4
Bond	C-C	C=O	C-OH	C-OR
#	4	1	2	0

REACTANT VECTOR
(R1 + R2)

I	1	2	3	4
Bond	C-C	C=O	C-OH	C-OR
#	4	1	0	2

PRODUCT VECTOR (P)

I	1	2	3	4
Bond	C-C	C=O	C-OH	C-OR
#	0	0	-2	2

REACTION VECTOR (D)

$$D = P - (R1 + R2)$$

¹ Broughton, H. B. et al. (2003) Methods for Classifying and Searching Chemical Reactions. United States Patent Application 367550. 25 Sept

Descriptor

Use of modified Atom Pairs descriptor

$$AP2 : X1(n, p, r) - 2(BO) - X2(n, p, r)$$

$$AP3 : X1(n, p, r) - 3 - X2(n, p, r)$$

Where:

$X1$ and $X2$ - chemical symbols

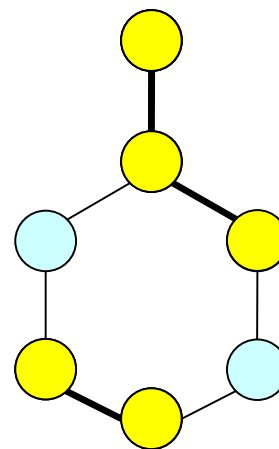
n - number of heavy atoms

p - number of π bonds

r - number of ring memberships

d - distance between atoms

BO - bond order





Reaction Vectors ($D = P - R$) as Reaction Transforms

The reaction vector D is given by:

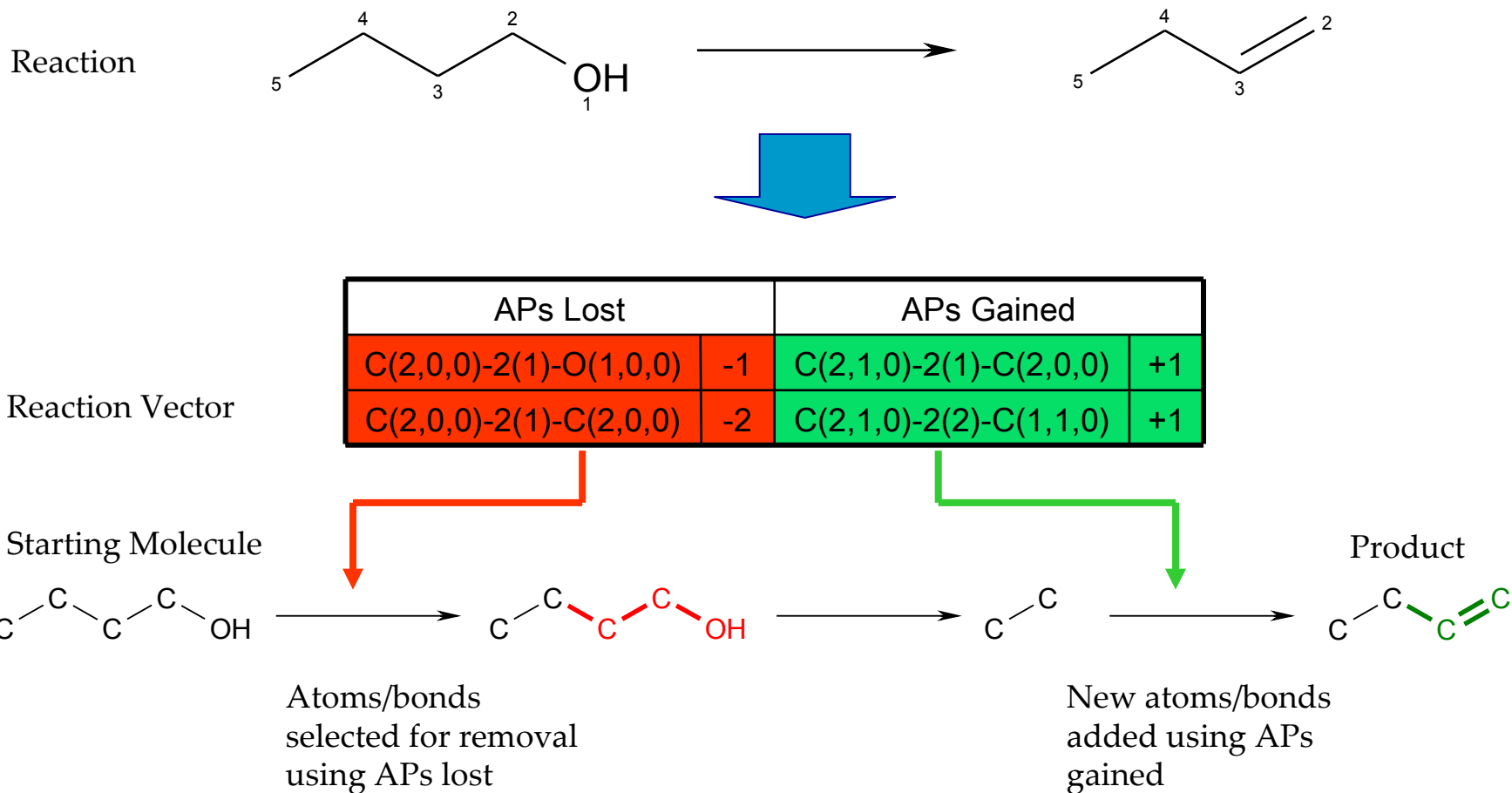
$$D = P - R$$

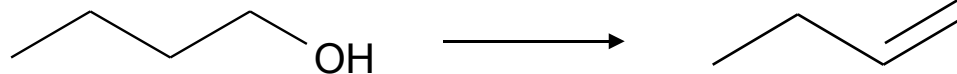
Where P is the product vector and R the reactant vector.

Given a reaction vector (D) and the reactant(s) (R) from which it was derived, can the product vector (P) be generated:

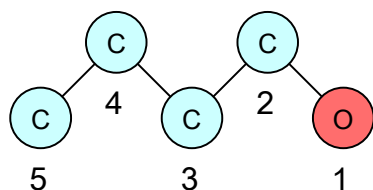
$$P = D + R \quad \text{or} \quad (P_1 + P_2) = D + (R_1 + R_2)$$

Applying Reaction Vectors





Starting molecule



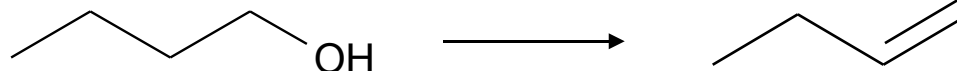
Atom Table

Atom No.	Type	N	P	R
1	O	1	0	0
2	C	2	0	0
3	C	2	0	0
4	C	2	0	0
5	C	1	0	0

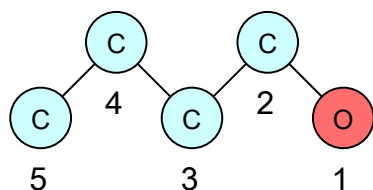
APs Lost	
C(2,0,0)-2(1)-O(1,0,0)	-1
C(2,0,0)-2(1)-C(2,0,0)	-2

Bond Table

AP	Atom 1	Atom 2	Bond Order
C(2,0,0)-2(1)-O(1,0,0)	2	1	1
C(2,0,0)-2(1)-C(2,0,0)	2	3	1
C(2,0,0)-2(1)-C(2,0,0)	3	4	1
C(2,0,0)-2(1)-C(1,0,0)	4	5	1



Starting molecule



Atom Table

Atom No.	Type	N	P	R
1	O	1	0	0
2	C	2	0	0
3	C	2	0	0
4	C	2	0	0
5	C	1	0	0

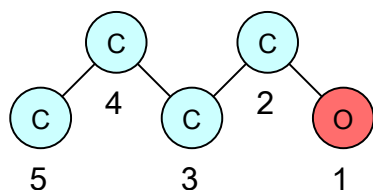
APs Lost	
C(2,0,0)-2(1)-O(1,0,0)	-1
C(2,0,0)-2(1)-C(2,0,0)	-2

Bond Table

AP	Atom 1	Atom 2	Bond Order
C(2,0,0)-2(1)-O(1,0,0)	2	1	1
C(2,0,0)-2(1)-C(2,0,0)	2	3	1
C(2,0,0)-2(1)-C(2,0,0)	3	4	1
C(2,0,0)-2(1)-C(1,0,0)	4	5	1



Intermediate molecule



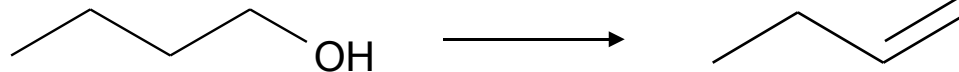
Atom Table

Atom No.	Type	N	P	R
1	O	1	0	0
2	C	2	0	0
3	C	2	0	0
4	C	2	0	0
5	C	1	0	0

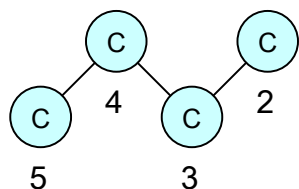
APs Lost	
C(2,0,0)-2(1)-O(1,0,0)	-1
C(2,0,0)-2(1)-C(2,0,0)	-2

Bond Table

AP	Atom 1	Atom 2	Bond Order
C(2,0,0)-2(1)-O(1,0,0)	2	1	1
C(2,0,0)-2(1)-C(2,0,0)	2	3	1
C(2,0,0)-2(1)-C(2,0,0)	3	4	1
C(2,0,0)-2(1)-C(1,0,0)	4	5	1



Intermediate molecule



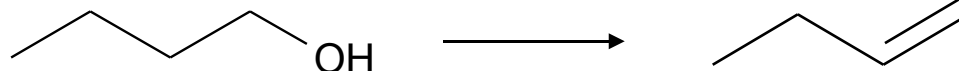
Atom Table

Atom No.	Type	N	P	R
1	O	1	0	0
2	C	2	0	0
3	C	2	0	0
4	C	2	0	0
5	C	1	0	0

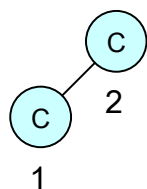
APs Lost	
C(2,0,0)-2(1)-O(1,0,0)	-1
C(2,0,0)-2(1)-C(2,0,0)	-2

Bond Table

AP	Atom 1	Atom 2	Bond Order
C(2,0,0)-2(1)-O(1,0,0)	2	1	1
C(2,0,0)-2(1)-C(2,0,0)	2	3	1
C(2,0,0)-2(1)-C(2,0,0)	3	4	1
C(2,0,0)-2(1)-C(1,0,0)	4	5	1



Intermediate molecule



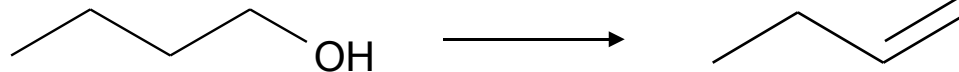
Atom Table

Atom No.	Type	N	P	R
1	C	1	0	0
2	C	2	0	0

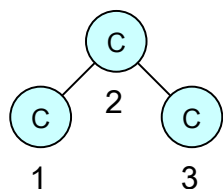
APs Gained	
C(2,1,0)-2(1)-C(2,0,0)	+1
C(2,1,0)-2(2)-C(1,1,0)	+1

Bond Table

AP	Atom 1	Atom 2	Bond Order
C(2,0,0)-2(1)-C(1,0,0)	1	2	1



Intermediate molecule



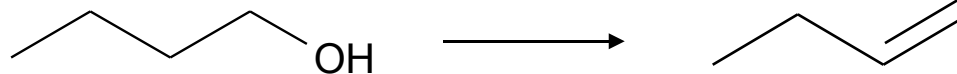
Atom Table

Atom No.	Type	N	P	R
1	C	1	0	0
2	C	2	0	0
3	C	2	1	0

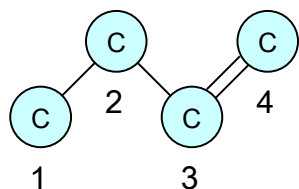
APs Gained	
C(2,1,0)-2(1)-C(2,0,0)	+1
C(2,1,0)-2(2)-C(1,1,0)	+1

Bond Table

AP	Atom 1	Atom 2	Bond Order
C(2,0,0)-2(1)-C(1,0,0)	1	2	1
C(2,1,0)-2(1)-C(2,0,0)	2	3	1



Intermediate molecule



Atom Table

Atom No.	Type	N	P	R
1	C	1	0	0
2	C	2	0	0
3	C	2	1	0
4	C	1	1	0

APs Gained	
C(2,1,0)-2(1)-C(2,0,0)	+1
C(2,1,0)-2(2)-C(1,1,0)	+1

Bond Table

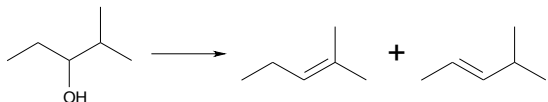
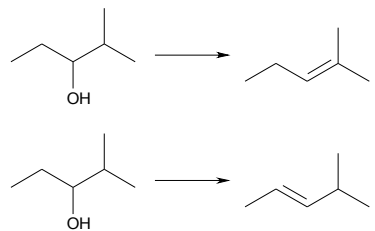
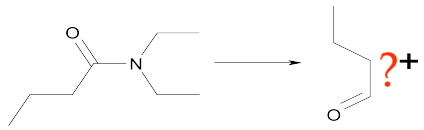
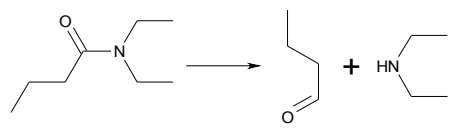
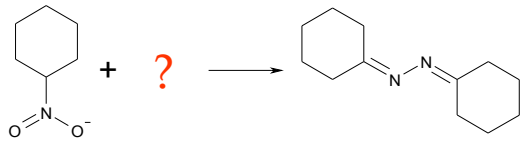
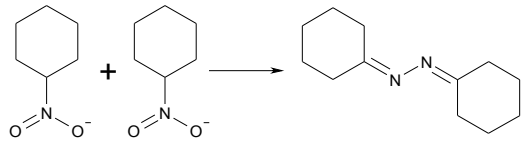
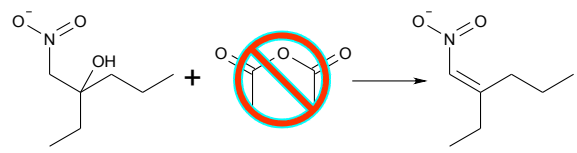
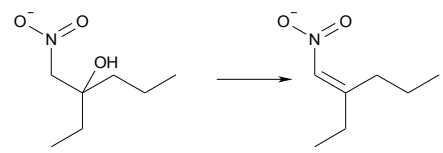
AP	Atom 1	Atom 2	Bond Order
C(2,0,0)-2(1)-C(1,0,0)	1	2	1
C(2,1,0)-2(1)-C(2,0,0)	2	3	1
C(2,1,0)-2(2)-C(1,1,0)	3	4	2

Reaction Cleaning

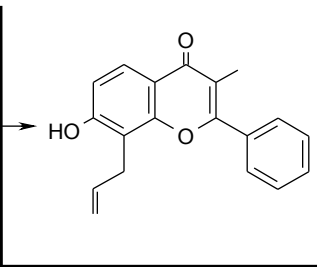
44% of reactions extracted were incomplete

95% of these were successfully "cleaned" using an algorithm

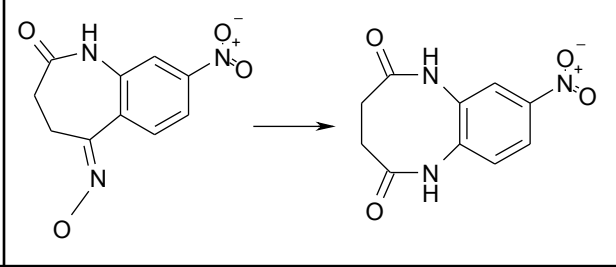
The dataset consisted of 6016 reactions made up of 29 different reaction types

Incomplete reaction	After "cleaning"
 <p>Reaction of 3-methylbutan-2-ol to two different alkenes, representing an incomplete reaction.</p>	 <p>Reaction of 3-methylbutan-2-ol to two different alkenes, representing a cleaned reaction.</p>
 <p>Reaction of N-ethylpropylacetamide to propylaldehyde and a question mark, representing an incomplete reaction.</p>	 <p>Reaction of N-ethylpropylacetamide to propylaldehyde and diethylamine, representing a cleaned reaction.</p>
 <p>Reaction of cyclohexyl nitrite to cyclohexylidenehydrazine with a question mark, representing an incomplete reaction.</p>	 <p>Reaction of two cyclohexyl nitrite molecules to cyclohexylidenehydrazine, representing a cleaned reaction.</p>
 <p>Reaction of a nitro-substituted alcohol to a nitro-substituted alkene with a crossed-out byproduct, representing an incomplete reaction.</p>	 <p>Reaction of a nitro-substituted alcohol to a nitro-substituted alkene, representing a cleaned reaction.</p>

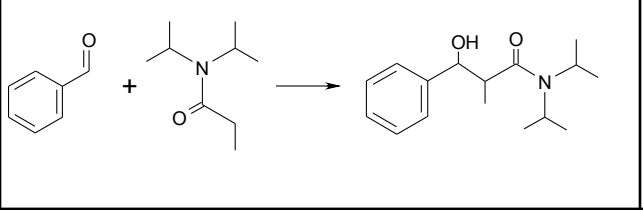
Result (5304 reactions)	% Reactions
Correct Product	90.3
Incorrect Product	4.7
No Product Generated	5.0



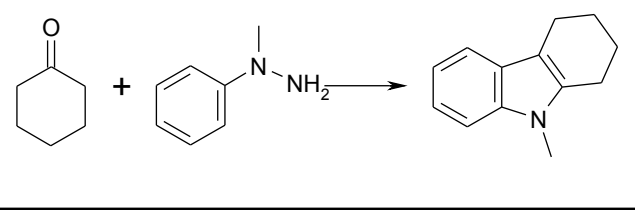
Rearrangement



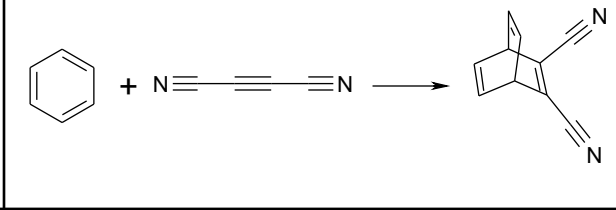
Beckmann rearrangement



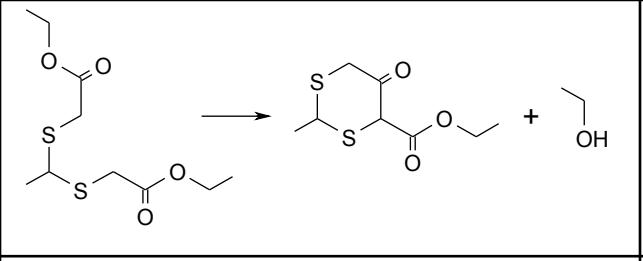
Aldol condensation



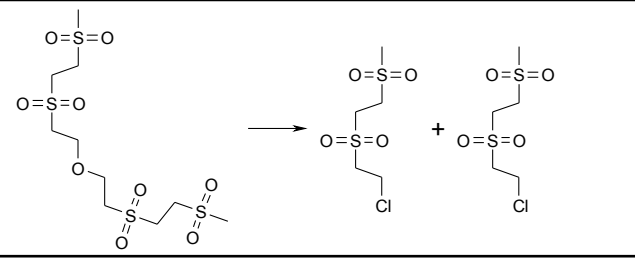
Fischer indole



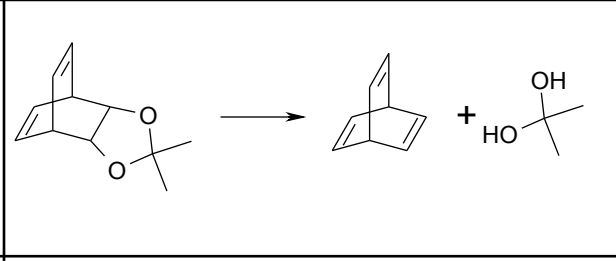
Diels-Alder cycloaddition



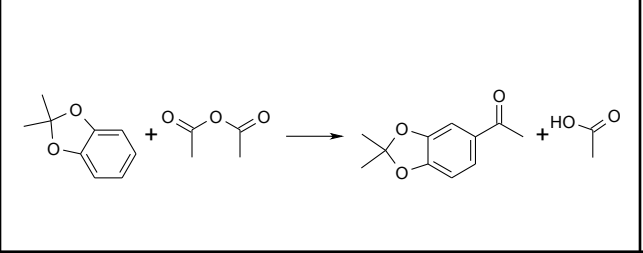
Dieckmann condensation



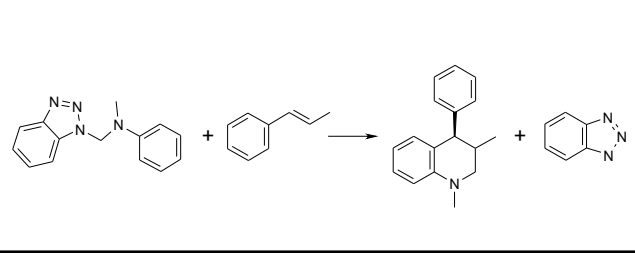
Ether halogenation



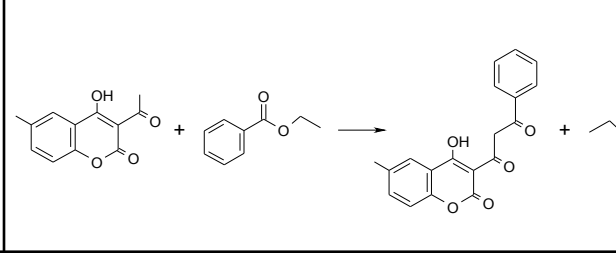
Olefination



Friedel-Crafts acylation



Hetero Diels-Alder

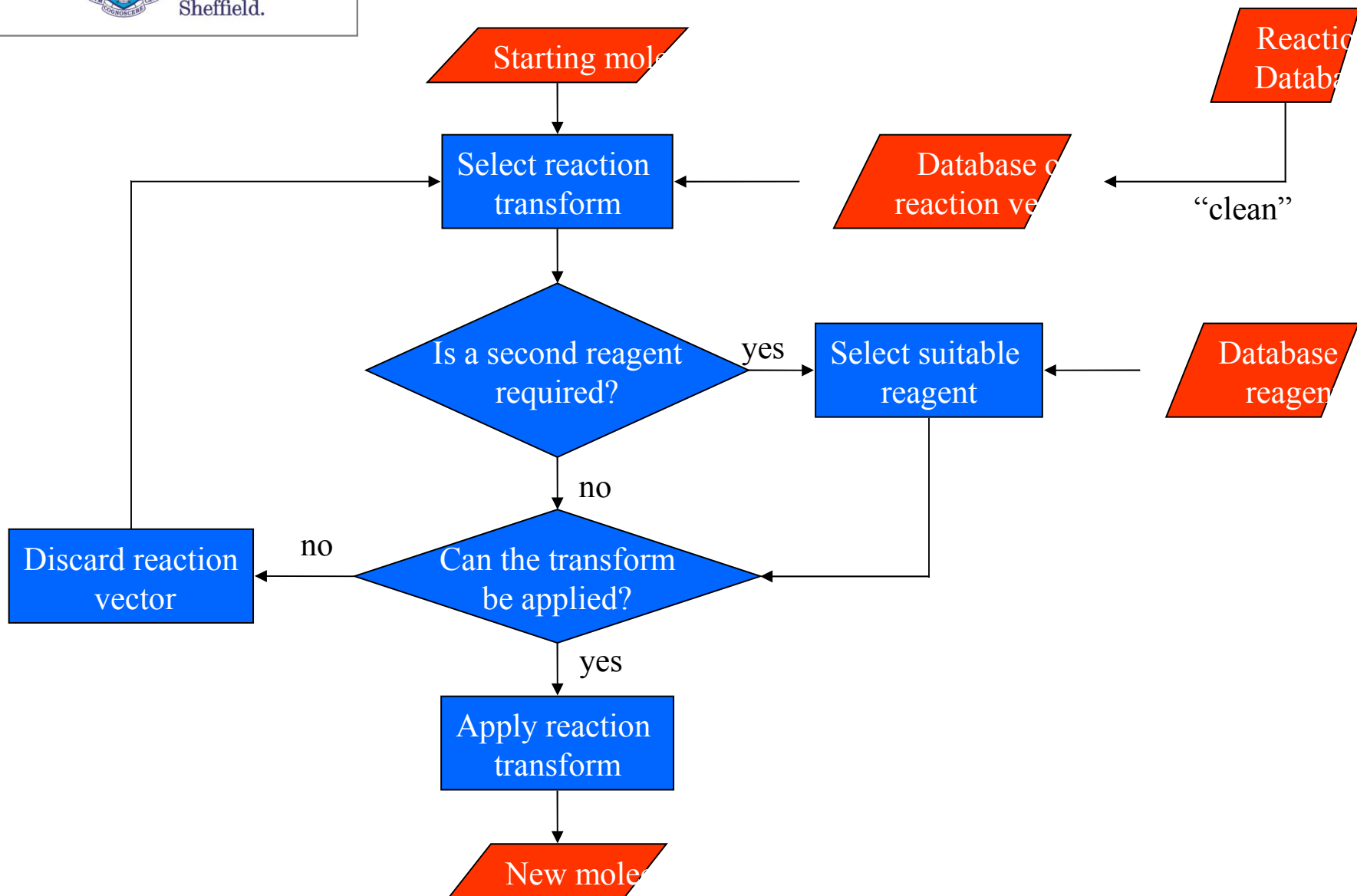


Claisen condensation



De Novo Design

Given a starting molecule and a set of reaction vectors can we identify suitable reagents and build novel molecules of potential therapeutic use?



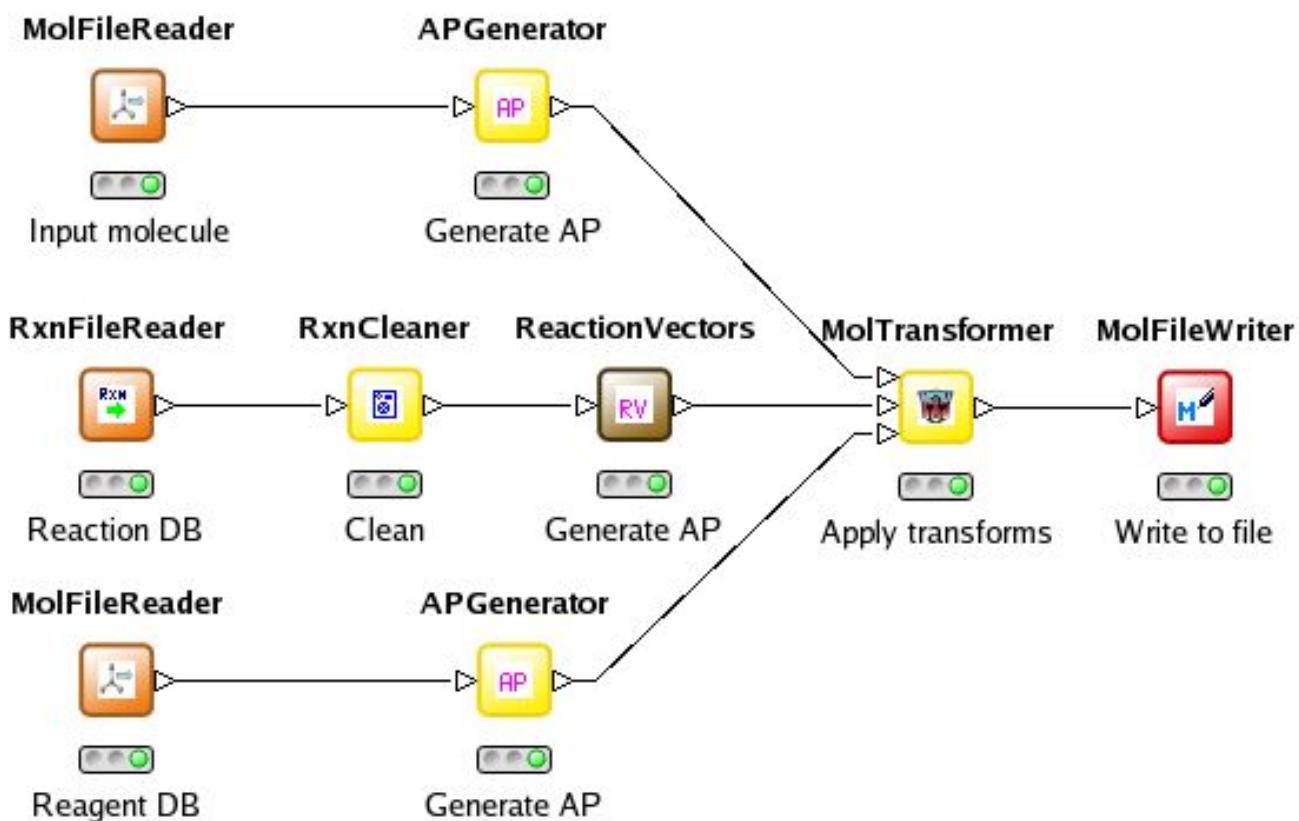
A Desktop Tool Using KNIME¹

- KNIME is a java based platform allowing creation of visual data flows
- Nodes can be built and modified rapidly to generate the required workflow
- Allows a plug and play approach
- Chemists can explore the next synthetic step

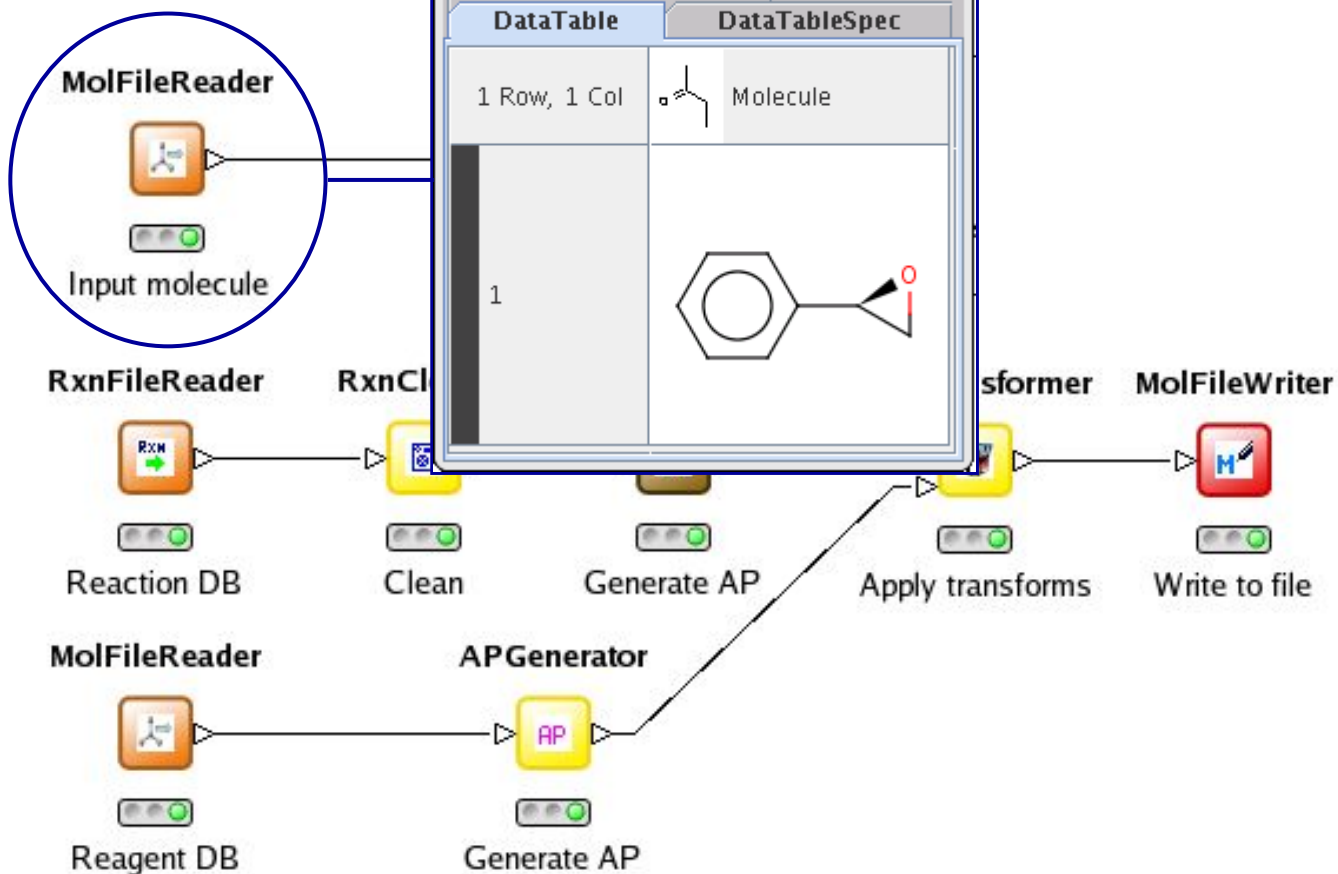
¹ Konstanz Information Miner - www.knime.org



KNIME Nodes



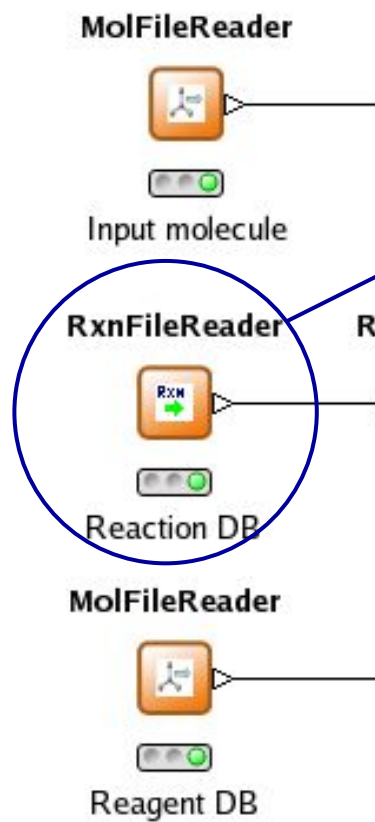
KNIME NO





The University Of Sheffield.

KNIME M



RxnFileReader (#52), Out-Port, Table: default, Cols: 1

File
DataTable | DataTableSpec | DataColumnProperties

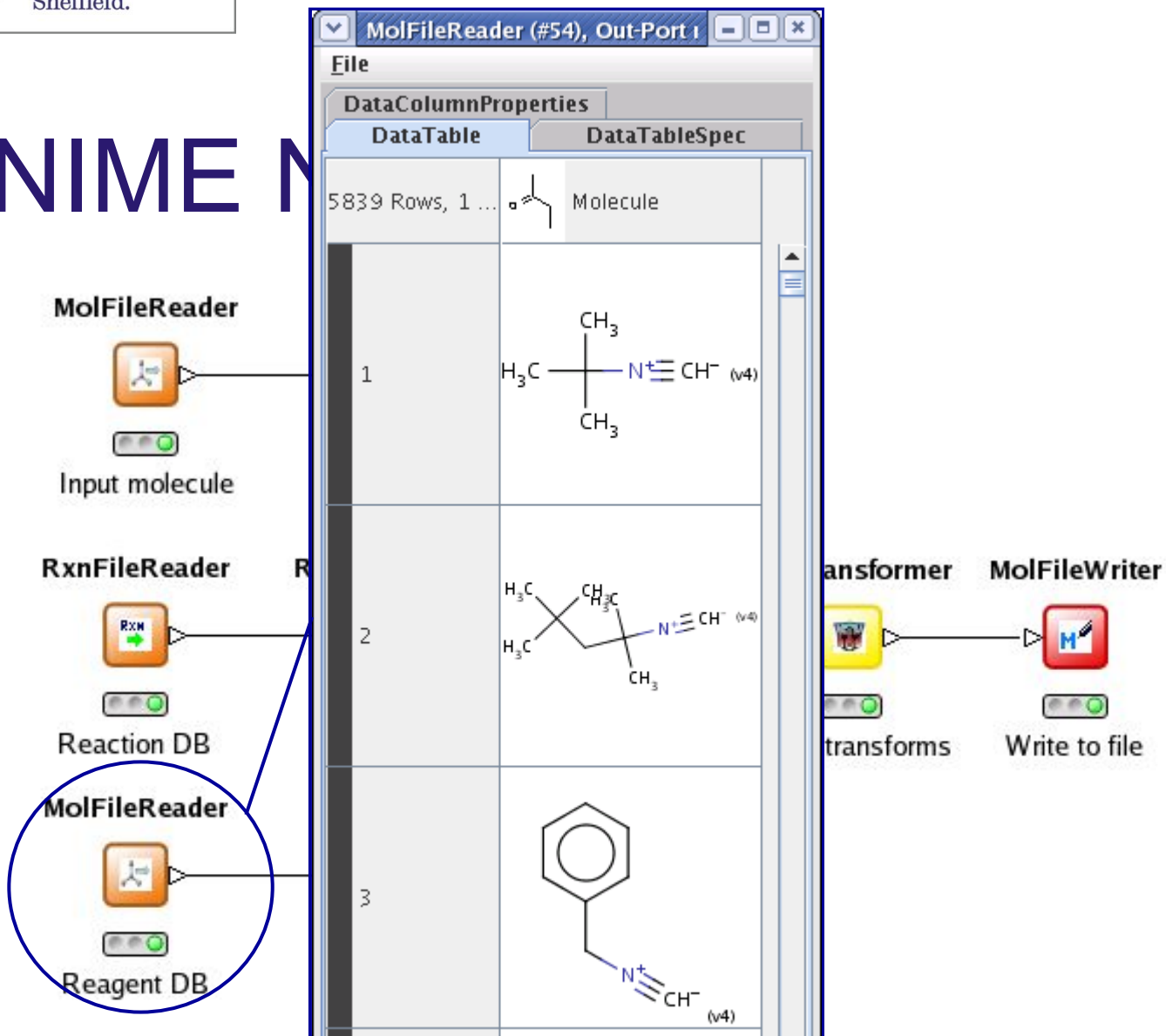
26676 Ro... A+B Reactions

1	
2	
3	

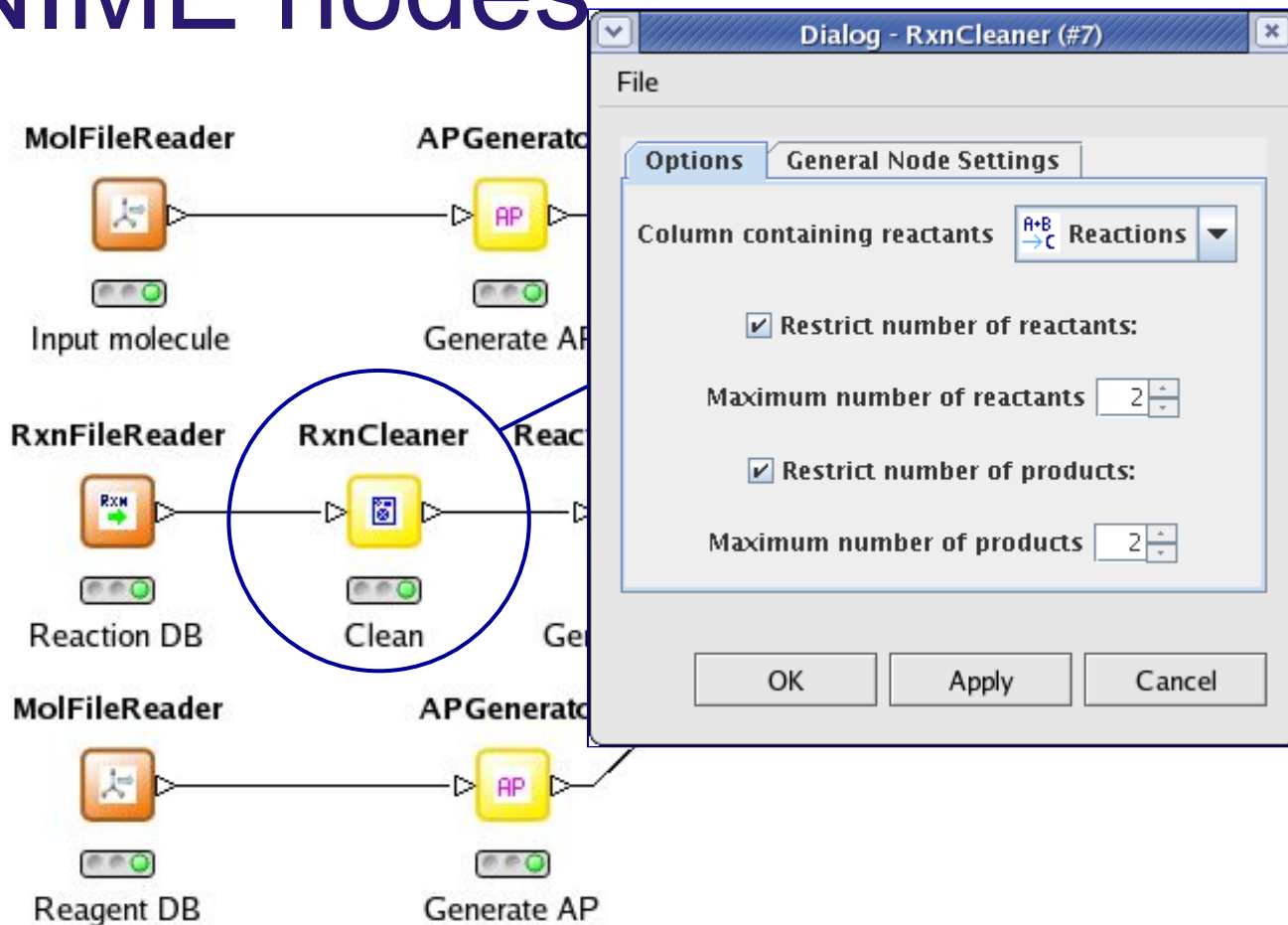
Writer
to file



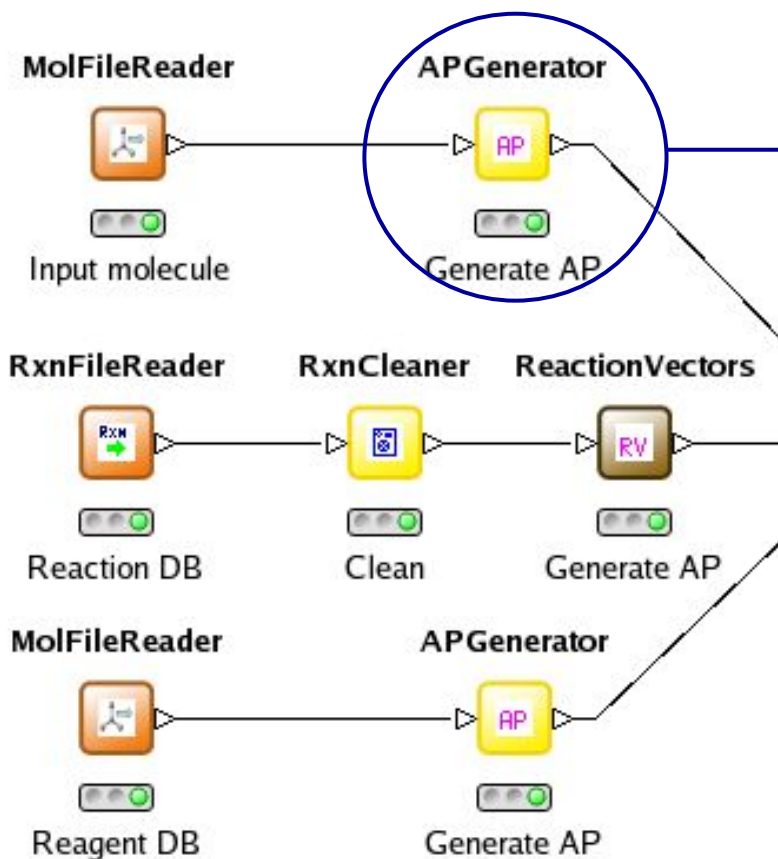
KNIME M



KNIME nodes



KNIME nodes



Dialog - APGenerator (#71)

File

Options General Node Settings

Data information:

Column containing molecules CC(=O)C Molecule

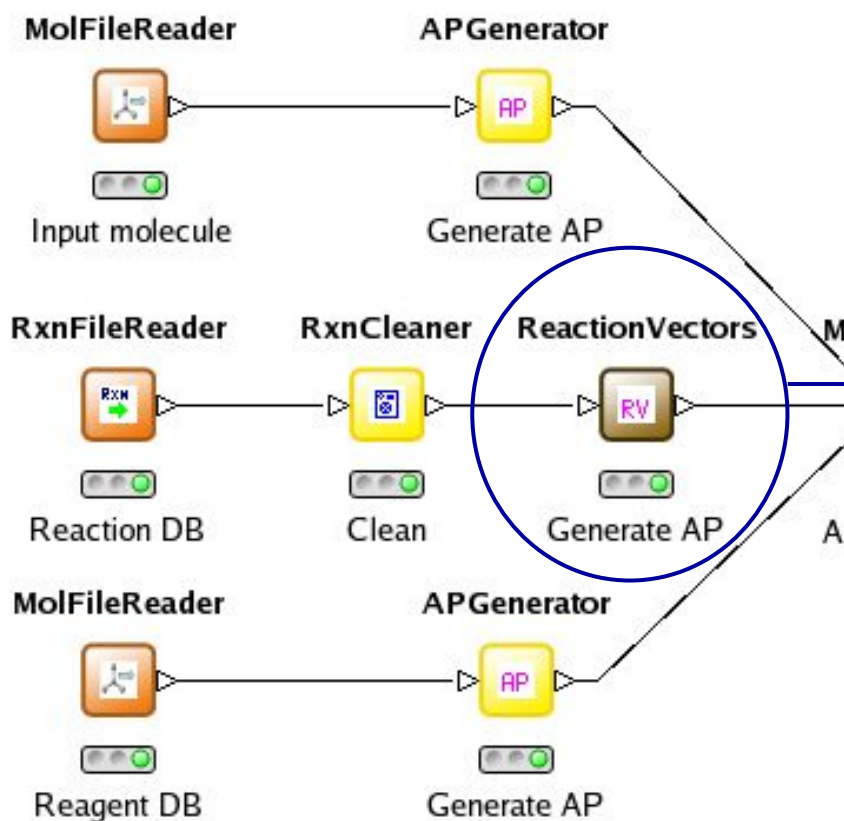
Atom Pairs settings:

Maximum atom distance to calculate Atom Pairs 3

- Use pi bonds:
- Use ring memberships:
- Use bond order for AP2:
- Store atom/bond info:

OK Apply Cancel

KNIME nodes



Dialog - ReactionVectors (#2)

File

Options General Node Settings

Data information:

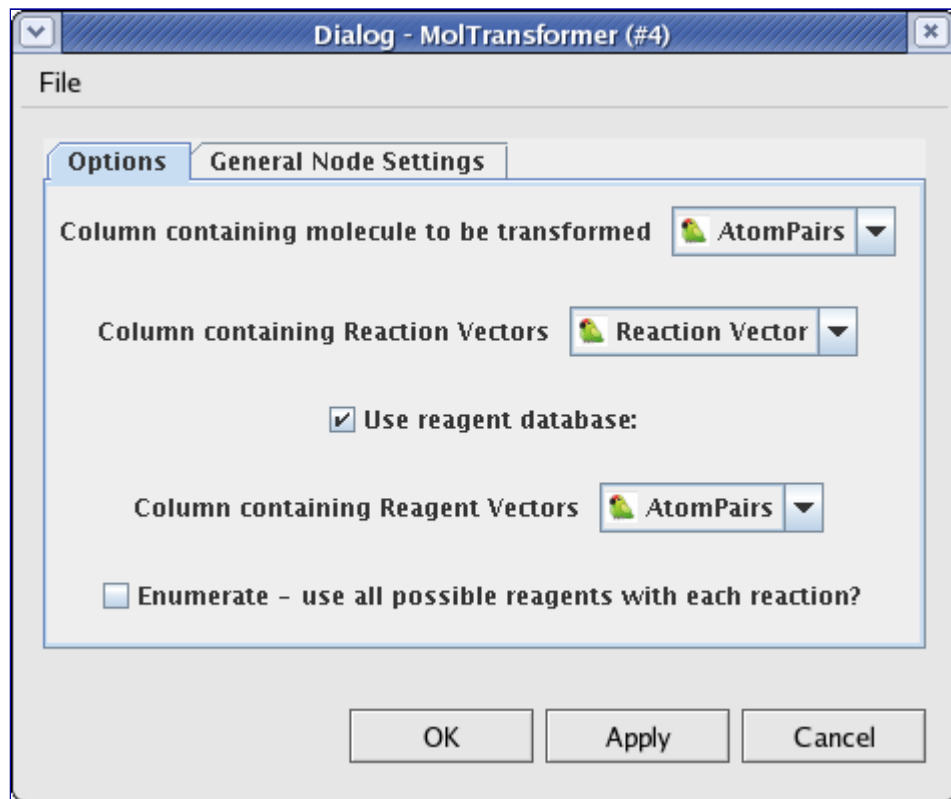
Column containing reactants $A \cdot B \rightarrow C$ Reaction

Atom Pairs settings:

Maximum atom distance to calculate Atom Pairs 3

- Use pi bonds
- Use ring memberships
- Use bond order for AP2
- Remove duplicates

OK Apply Cancel



MolFileReader

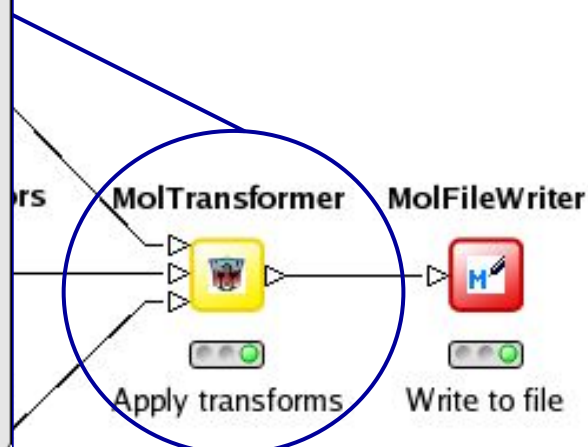


Reagent DB

APGenerator



Generate AP



File

DataTable

DataTableSpec

DataColumnProperties

24 Rows, 3 ...

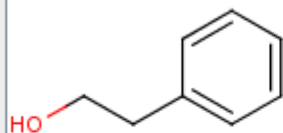


Solutions

S Reaction ID

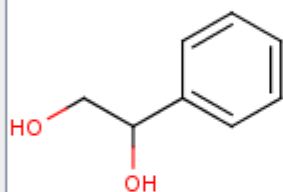
S Reagent ID

Mol 1: 21



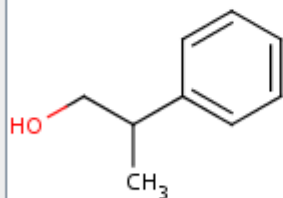
reaction 6329

Mol 1: 22



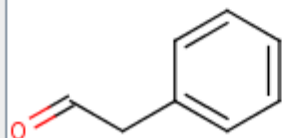
reaction 11178

Mol 1: 23

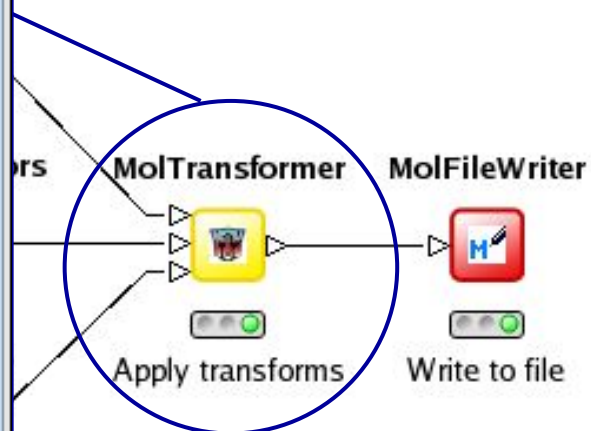


reaction 13801

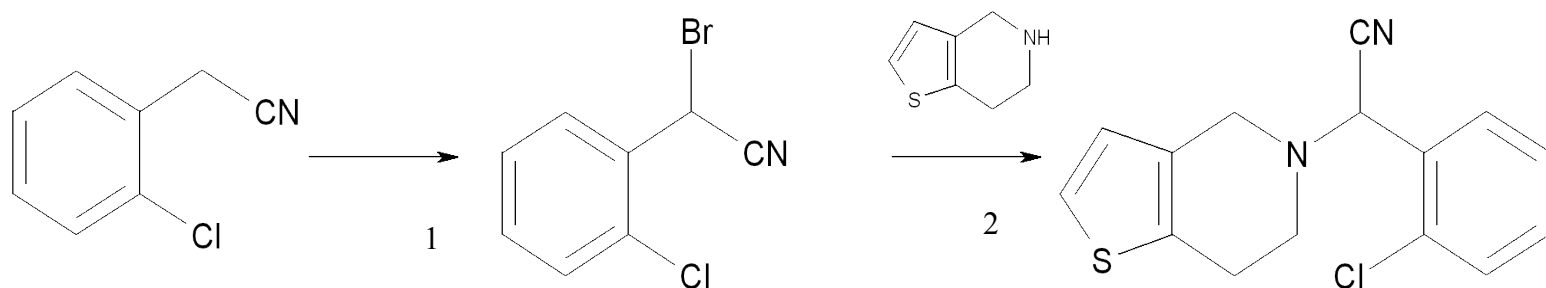
Mol 1: 24



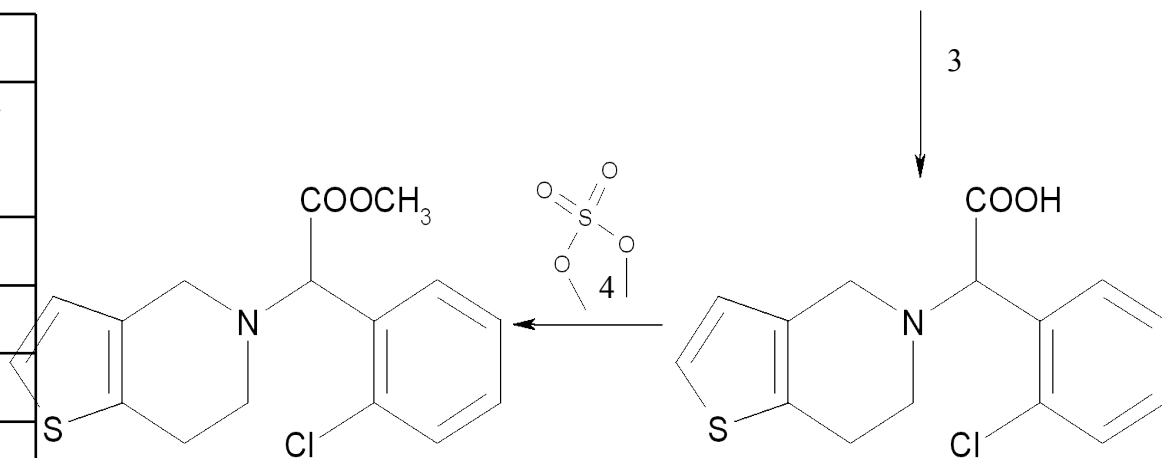
reaction 20182



Replicating the Synthetic Route of a Known Drug – Plavix™¹

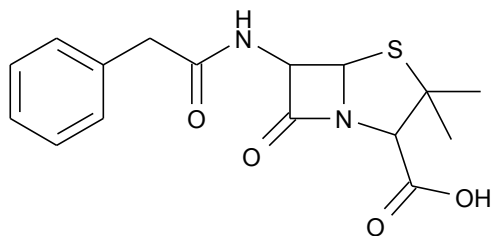


Plavix™		
Step	No. applicable reaction vectors	No. reagents for correct reaction
1	20	1
2	21	3
3	15	1
4	23	1

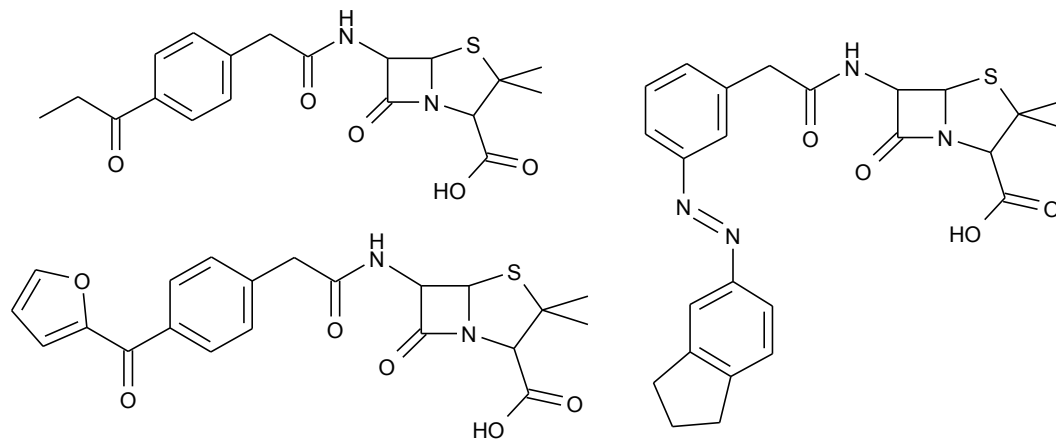


¹ Wang, L. et al. (2007), *Org. Process Res. Dev.*, 11 (3), 487-489

Lead Optimisation



Penicillin G

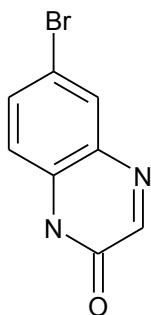


Examples of novel molecules generated

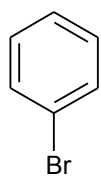
- 24 analogues generated - 11 found in literature
- Average cosine similarity to real products of the reactions used was 0.55

Enumeration

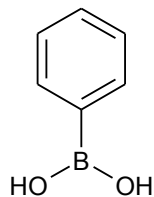
Core structure



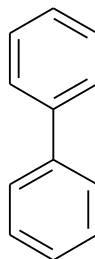
+



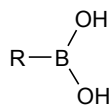
+



→

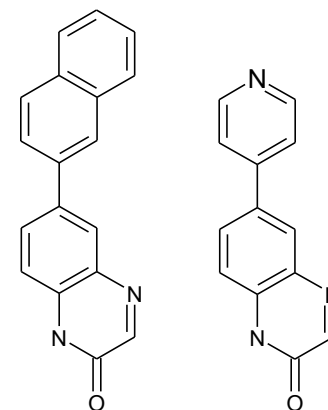
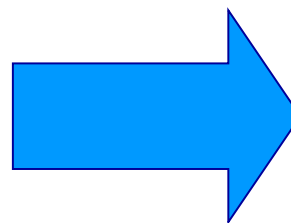


+

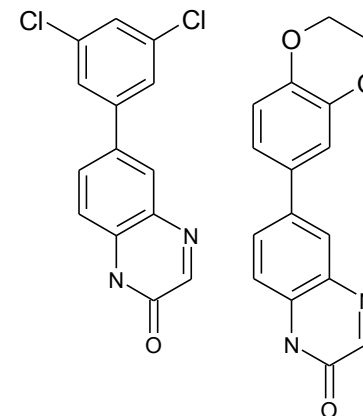


Suzuki
reaction

628 boronic acids



292 output
structures



Conclusions & Future Work

- Reaction vectors can be applied successfully as reaction transformations
- Reaction databases are continuously being updated and this is one way to utilise the knowledge base
- A desktop tool was created using KNIME to generate new molecules from a starting molecule, using one-step transformations
- Next step is to incorporate this method into a multi-objective *de novo* design tool



The
University
Of
Sheffield.

Acknowledgments

Val Gillet (CISRG group University of Sheffield)

Beining Chen (Chen group University of Sheffield)

Michael Bodkin

CADD group at Lilly UK

Funding:



Answers That Matter.

Software and technical support:

JOELIB (University of Tübingen)

KNIME (www.knime.org)

Marvin Beans (ChemAxon)