



**The Efficient Enumeration
of
Reaction Networks.**



Reaction Based Enumeration.

- Developed jointly with Evotec OAI.
- ‘Reaction Based’ - reactions are predefined.
- Choice of Reactant Classes, aids specificity.
- Flexible reagent selection.
- ‘Reaction Network’ GUI developed through discussions with Harvard Med. School.
- Integrated with Modgraph Registry and CSC (Chemical Supplier Catalogue).



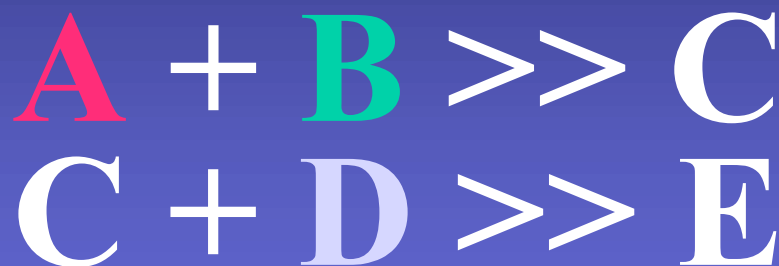
Library Builder

- Exploits systematic nature of Daylight SMIRKS and SMARTS.
- SMIRKS generated on the fly.
- Many thousands of different SMIRKS. (Only one bug found).
- Selection from Inventory and selection of reaction centre are the same 'thing'.
- Design of SMARTS is a compromise between trying to select only hits which are 'useful', (minimising list size) and excluding structures which are inappropriate.



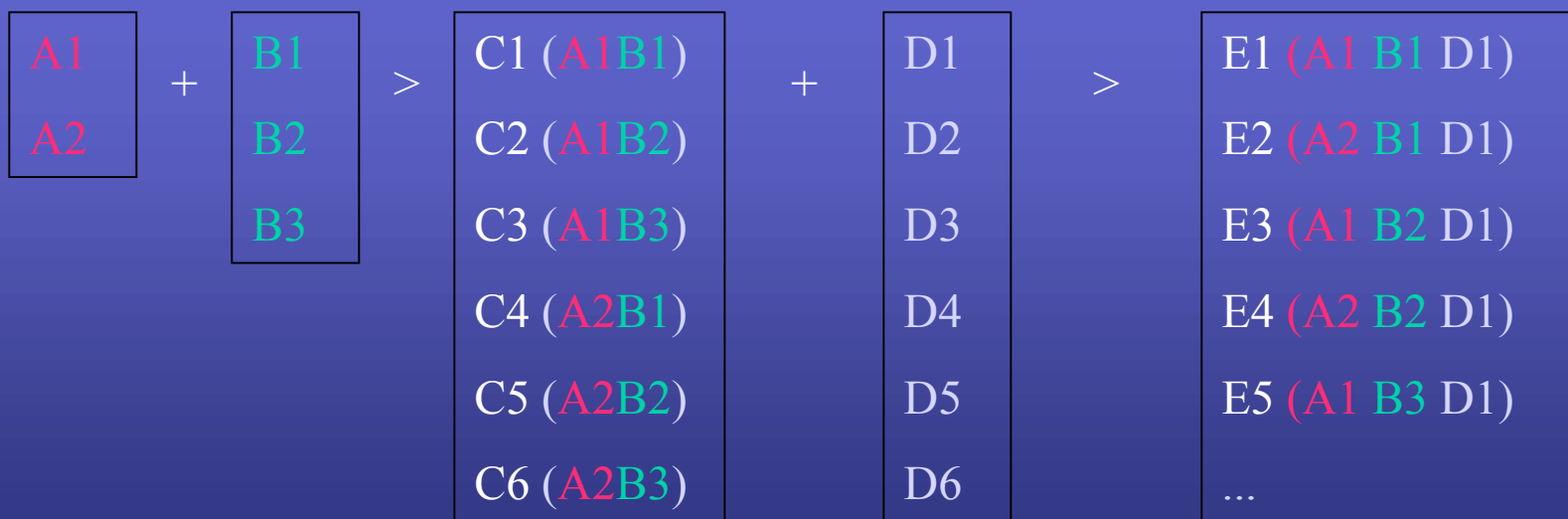
Enumeration Example

2 Step Reaction



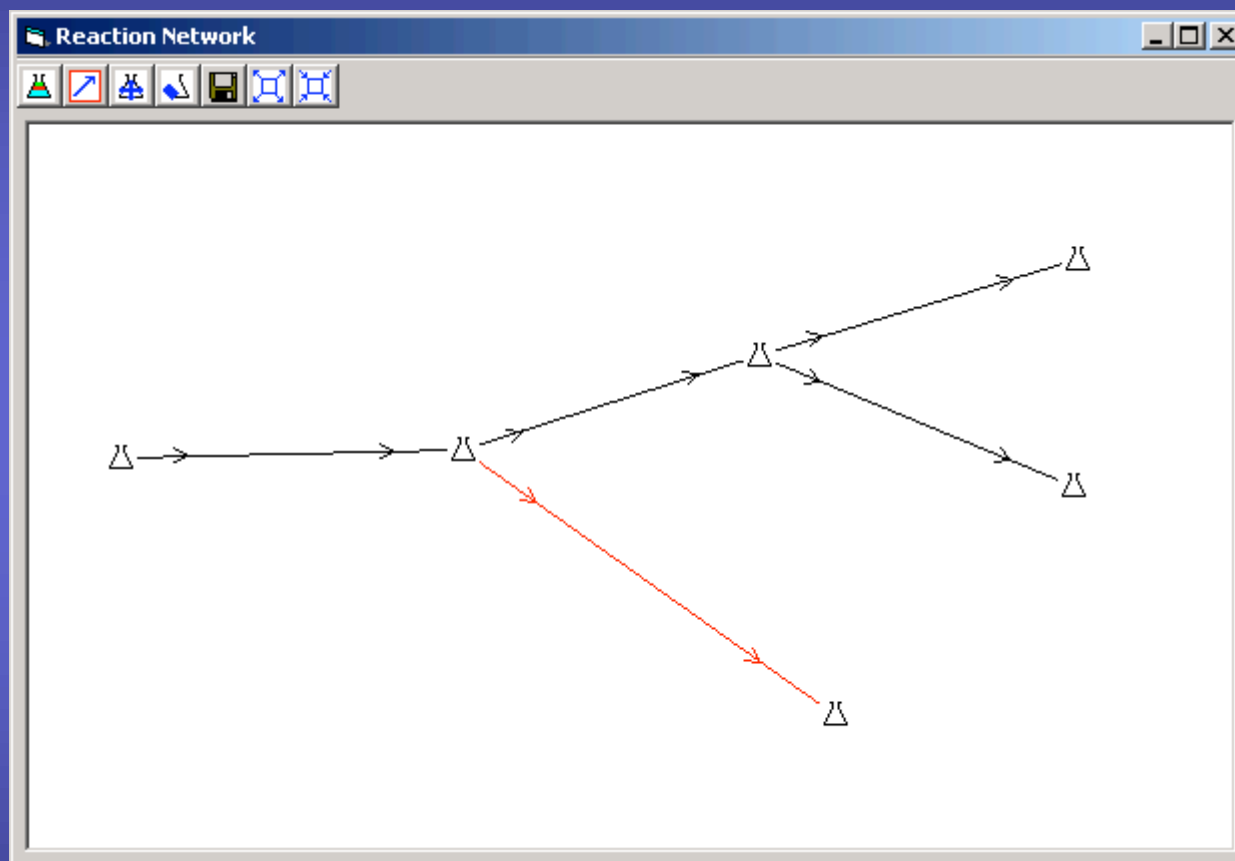
$$2 \text{ A } \times 3 \text{ B } = 6 \text{ C}$$

$$\times 6 \text{ D } = 36 \text{ E}$$





Network Design



First design the reaction network.

Nodes can be added, deleted and moved, zoom-in, zoom-out. Rules.



Template Approval

Library Builder

New Open Save Search Overlap LNB Data Info Cluster Network

Library Name

Lists

Step/Reactant	nCpds
Route 1, Step 1 Reactant 1	0 + 0
Subtotal Route 1	0
Subtotal Route 2	0
Subtotal Route 3	0
Total All Routes	0

Oxidation of Olefin: Aliphatic, Alkene

Transform Query Query+ Enum Only Sample Enumerate View Library Failures Export Inv No Export SDF Overview

No Compounds = 0 of 0 (0 Failures)

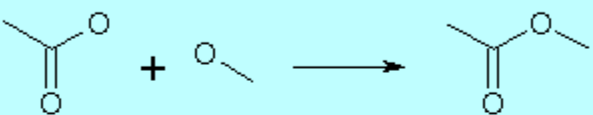
Clicking on step
(edge) displays
selected reaction



Reaction Selection

Transform: Mode = 'View'

Template ID Alcohol/Acid Esterification



Close OK

Transform

Intramolecular

No. Reactants = 2

Alcohol Acid

Reactant Class

- Aliphatic
- Aromatic
- Acyclic Anhydride
- Cyclic Anhydride
- Previous Step
- Amino Acid
- Excess
- Unsaturated
- Generic

Product-Based

- Acid Amide
- Epoxide
- Carbonate
- C-C double bond
- Alcohol
- Amine
- Ammonium salt
- Carbamoyl chloride
- Sulfonic acid derivatives
- Urea/thiourea
- Ketone
- Carbodiimide
- Thioamide
- Ester
 - Alcohol/Acid Esterification
 - Oxime/Acid Substitution
 - Ester Alkylation
- Carboxylic Acid
- Aldehyde
- Alkane
- Imine

Reaction
selection

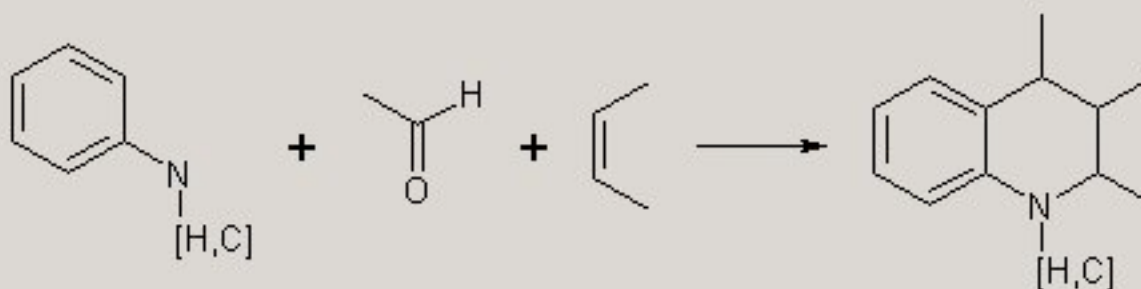
'Reactant Class'
selection



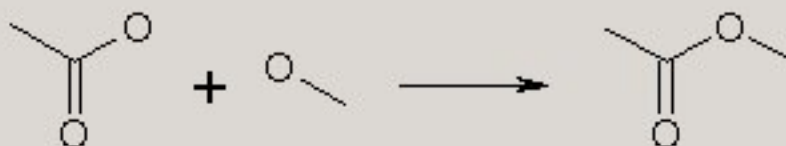
Reaction Overview

Overview: template = WX24V

Quinoline: Aliphatic, Aldehyde + Primary, Aromatic Amine + Aliphatic, Alkene



Alcohol/Acid Esterification: Primary, Alcohol + Aromatic, Previous Step, Acid



Overview can
be printed



Reactant Selection

Search for reactants is based on either:

- Inventory number
- Exact structure
- Any Combination of
 1. Default SMARTS - the SMARTS based on reactant class selection.
 2. Sub-structure
 3. Filter - applies additional SMARTS based on reactant classes
 4. Molecular weight range.
- Two Stage Selection. Select list and select from list.



Reagent Selection - Stage 1.

Query List: No. of Databases selected = 1

Open Save Clear Search Export List Import List

Step: 1 Aliphatic Aldehyde Primary Aromatic Amine Aliphatic Alkene

Structure
 Customer = EOAI Database = EOAI_INVENTORY
 MWT = 139.114487 MF = C5H5N3O2
 Name = 4-AMINOPYRIMIDINE-5-CARBOXYLIC ACID
 Clipboard = Nc1ncnc1C(=O)O

11 Hits

Inv No
10168
10197
10381
10661
12264
15702
24644
24645
24656
24657
26345

3 Picks

Inv No	Status
10168	E
10197	E
10369	E

Remove

Query

Inv No. Exact Structure

Default Query Sub Structure Filter

MWT >= 100 < 150

Input Selection Run Query Reset Query Filter

Filter
IS Aromatic Acid

QUERY IS Primary Aromatic Amine AND IS Aromatic Acid AND MWT >= 100 AND MWT < 150

Multiple, customer specific databases.

Reactant Structure and details.

Default Query

+ Filter

+ MW range

Filter selected by name/class.

English description of Query.



Reagent Selection - Stage 2

Hits List: 500 compounds. Display: 0 - 8 Selected: 3

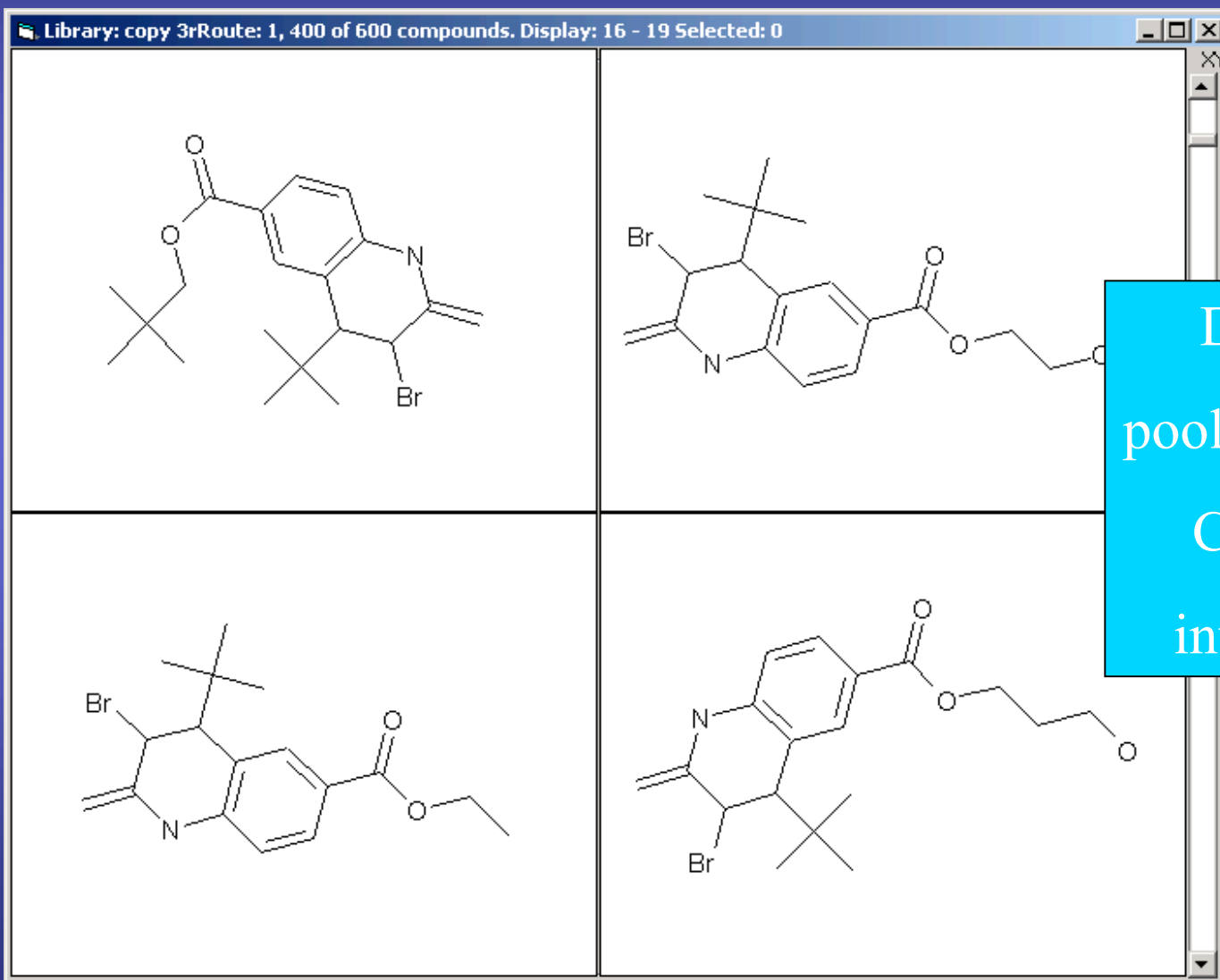
The screenshot displays a grid of chemical structures. The top-right cell is highlighted in grey. The structures are:

- Top-left: A benzene ring with an amino group (-NH₂), a hydroxyl group (-OH), and a bromine atom (-Br).
- Top-middle: A benzene ring with an amino group (-NH₂), a bromine atom (-Br), and a fluorine atom (-F).
- Top-right (highlighted): A benzene ring with an amino group (-NH₂) and two hydroxyl groups (-OH).
- Middle-left: A benzene ring with an amino group (-NH₂), a chlorine atom (-Cl), and a hydroxyl group (-OH).
- Middle-middle: A benzene ring with an amino group (-NH₂), a bromine atom (-Br), and a hydroxyl group (-OH).
- Middle-right: A benzene ring with a hydroxyl group (-OH) and a hydroxyl group (-OH).
- Bottom-left: A benzene ring with two amino groups (-NH₂).
- Bottom-middle: A benzene ring with a hydroxyl group (-OH) and a hydroxyl group (-OH).
- Bottom-right: A benzene ring with a hydroxyl group (-OH) and a hydroxyl group (-OH).

Structures are displayed
with Inventory No.



Library View





Product Intermediates

Product Intermediates

Clipboard=
LNB =

CCD No
10143
10144
10168
10459

Inventory Nos
of reactants.



Enumeration Algorithm

- Fast and efficient enumeration of very large libraries.
- The library is split into a number of ‘experiments’, such that the number of compounds in one experiment is minimized, subject to no intermediate being synthesized more than once.
Minimum memory usage; maximum speed.
- Minimising memory usage is more important, since writing to a db is likely to be RDS.
- Library Sample Algorithm - minimum library size consistent with every reactant being used at least once.

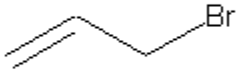
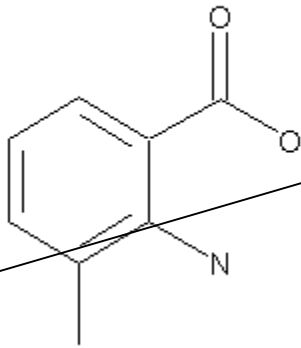
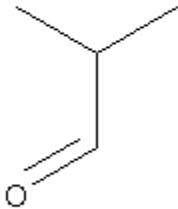


Failures List

Enumeration Failures

Inv No
10581
10367
10459

#	Step
1	Step 1
2	Step 1
3	Step 1
4	Step 1
5	Step 1
6	Step 1
7	Step 1
8	Step 1
9	Step 1
10	Step 1
11	Step 1
12	Step 1
13	Step 1
14	Step 1
15	Step 1
16	Step 1
17	Step 1
18	Step 1



Shows the step no. which failed and reagents up to that step.

Here the failure is due to absence of aromatic H ortho to NH₂



Routes can be Disjoint.

Library Approval found 15 hits

New Open Save Search Overlap LNB Data Info Cluster Network

Library 0 Name MUG 2 Routes

Lists

Step/Reactant	nCpds
Route 1, Step 1 Reactant 1	3 + 0
Route 1, Step 1 Reactant 2	3 + 0
Subtotal Route 1	9
Route 1, Step 2 Reactant 1	3 + 0
Route 1, Step 2 Reactant 2	2 + 0
Subtotal Route 1	6
Total All Routes	15

Amine/Acid Condensation: Primary, Amine + Aromatic, Acid

No Compounds = 0 of 15 (0 Failures)

Transform Query Query+ Enum Only Sample Enumerate View Library Failures Export Inv No Export SDF Overview



Query +

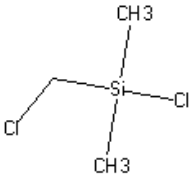
Query List: No. of Databases selected = 1

Open Append Clear Search Export List Import List

Step: 2 Secondary DR Previous Step Amide Aliphatic Haloalkane/Triflate

Structure

Customer = Database = INV2
 MWT = 143.089218 MF = C3H8Cl2Si
 Name = (CHLOROMETHYL)DIMETHYLCHLOROSILANE



500 Hits

Inv No
46485
46487
46494
46516
46547
46548
46562
46576
46623
46635
46674
46723
46736
46741
46742
46751
46752
46753
46755
46765
46769
46779
46794
46796
46802
46803
46804

7 Picks

Inv No	Status
46547	E
46741	E
46765	E
46804	E
46855	N
46976	N
47149	N

Remove

Filter

Query

Inv No. Exact Structure

Default Query Sub Structure Filter

MWT >= <

Input Selection Run Query Reset Query Filter

QUERY IS Aliphatic Haloalkane/Triflate

Three extra compounds added.

Avoid re-enumerating whole library. (Time, Data).



Query +

Library Approval found 45 hits

Library: 0 Name: MUG Net 2 Plus

N-Alkylation: Secondary, Previous Step, Amide + Aliphatic, Haloalkane/Triflate

CC(=O)N[C@@H](C)OS(=O)(=O)(F)F.[F-]C(Br)(F)F
 \rightarrow
CC(=O)N(C)C(F)(F)F.[H+].[O-]

Step/Reactant	nCpds
Route 1, Step 1 Reactant 1	3 + 0
Route 1, Step 1 Reactant 2	3 + 0
Route 1, Step 2 Reactant 1 (previous)	9
Route 1, Step 2 Reactant 2	7 + 3
Subtotal Route 1	90
Route 2, Step 3 Reactant 1 (previous)	9
Route 2, Step 3 Reactant 2	2 + 0
Subtotal Route 2	18
Total All Routes	108

No Compounds = 45 of 108 (0 Failures)

Transform Query Query+ **Enum Only** Sample Enumerate View Library Failures Export Inv No Export SDF Overview

$3 \times 3 \times 3 = 27$
 $9 \times 2 = 18$
 $27 + 18 = 45$



Sampling

Library Approval found 24 hits

Library: 0 Name: MUG Net 2 Plus

N-Alkylation: Secondary, Previous Step, Amide + Aliphatic, Haloalkane/Triflate

Lists

Step/Reactant	nCpds
Route 1, Step 1 Reactant 1	3 + 0
Route 1, Step 1 Reactant 2	3 + 0
Route 1, Step 2 Reactant 1	0
Route 1, Step 2 Reactant 2	10 + 0
Subtotal Route 1	16
Route 2, Step 3 Reactant 1	0
Route 2, Step 3 Reactant 2	2 + 0
Subtotal Route 2	8
Total All Routes	24

No Compounds = 24 of 24 (0 Failures)

Transform Query Query+ Enum Only **Sample** Enumerate View Library Failures Export Inv No Export SDF Overview

Arithmetic
rather than
Geometric.



Futures and Options

- Customisation. Integration with other applications/business process. Plate layout, library filtering, LNB assignment etc.
- Variable Sample size.
- Improvements to Reagent selection.
- Specification of Reaction Centre per reactant.



Acknowledgements

- John Bradshaw
- Jack Delaney
- Caroline McGee - formerly Evotec OAI