Computer-Assisted Retrosynthetic Planning

Andrew Zahrt
Denmark Lab Group Meeting
01/30/18

Some Common Misconceptions

1. The machines are here to replace us all.

2. All computational chemists are intimately familiar with informatics and machine learning

3. There are inherent limitations in machine learning methods that will prevent machines from ever designing syntheses of complex molecules (automated synthetic planning is “mission impossible”)

What is Computer-Assisted Synthesis Planning?

Computer assisted synthetic planning generally applies methods from informatics and machine learning to solve chemical problems.

Chess

Rubik's cube

Chemical synthesis

Starting Position	*Predefined, with some moves not allowed*	*Random*	*Synthetic Target*
Relative Positions	*Discrete Configuration of Pieces*	*Configuration of Cube*	*Synthetic Intermediates*
End Position	*Checkmate or Draw*	*Successful Completion*	*Commercially available materials*
Movements	*Standard Chess Moves (small, predefined number)*	*Rotation of a layer*	*Chemical Transformations*

If synthetic positions and synthetic moves can be defined, all of the above problems become very similar.

Gryzbowski *et al., Angew. Chem. Int. Ed.* **2016** , *55*, 5904 –5937

Historical Overview

G.E. Vléduts and V. K. Finn, 1957: an Information Machine for Chemistry will:

(i) search for individual chemical compounds ✓

(ii) search from chemical compounds possessing a certain given combination of characteristics ✓

(iii) search for classes of reactions into which a definite individual compound can enter ✓

(iv) search for the class of reaction producing a particular chemical compound ✓

(v) search for the class of reactions which are of the same type chemically and are characterized by a transfer of given structural elements… from the initial molecules into other definite structural elements of final molecules ✓

(vi) search for the reaction that will take place between given compounds under given conditions ✓

(vii) search for ways of synthesizing a given compound from a definite number of permissible initial compounds ?

Fifty Years of Development

DENDRAL Project, 1965 – Edward Feigenbaum, Bruce Buchanan, Joshua Lederberg, and Carl Djerassi

Organic Chemical Simulation of Synthesis (OCSS), 1969 – Corey and Wipke

Logic and Heuristics Applied to Synthetic Analysis (LHASA) – Corey and Wipke

Simulation and Evaluation of Chemical Synthesis (SECS) – Wipke

SYNCHEM 1977-1998 – Stanford/Stony Brook

SYNLMA, 1989 – P. Y. Johnson

SYNGEN, 1977-1990 – J. B. Hendrickson

IGOR/IGOR2, 1974-1993 – Ugi

CHIRON, 1990-2005 – Stephen Hanessian

WODCA - Johann Gasteiger

WODCA
Computer Assisted Organic Synthesis

ARChem Route Designer – SymBioSys

ARChem
Automated Reasoning in Chemistry

IC$_{\text{SYNTH}}$ – ChemInfo

IC SYNTH

Chematica – Gryzbowski

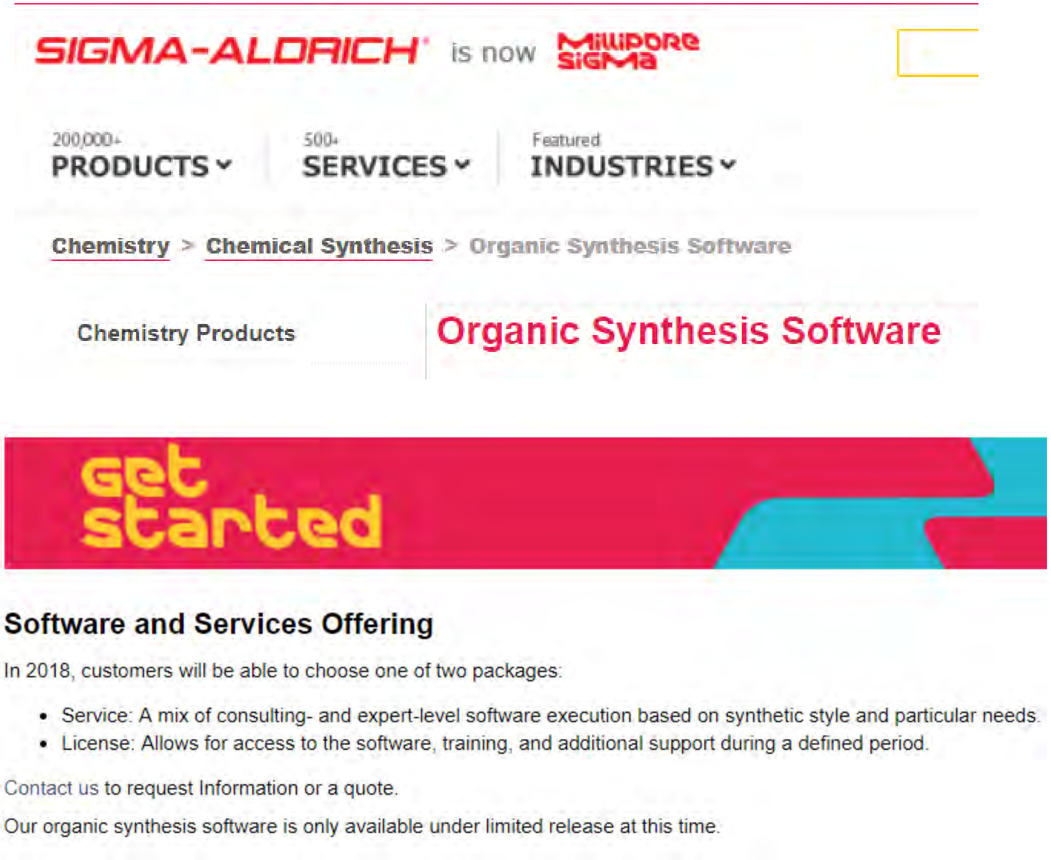
CHEMATICA

*1) Application of Artificial Intelligence for Organic Chemistry. The DENDRAL Project, McGraw-Hill, New York, 1980. 2) Science, **1969**, 166, 178-192 3) J. Am. Chem. Soc., **1972**, 94 , pp 421–430 4) Artificial Intelligence, **1978**, 11, 173-193 5) Science, **1977**, 197, 1041-1049 6) "Designing an Expert System for Organic Synthesis: The Need for Strategic Planning," Peter Y. Johnson, I. Burnstein, J. Crary, M.Evens, and T. Wang, Published in the **ACS Symposium Series 408 ";Expert System Applications in Chemistry**, p102-124, edited by Bruce Hohne and Thomas Pierce, 1989, Los Angeles, California*

Chematica Software

Bartosz Gryzbowski:
PhD. Harvard 2000 (Whitesides)
Post-doc Harvard 2000-2003
Northwestern (2003-2014)
UNIST (Distinguished Professor, Chemistry, 2014~Present)
ProChimia Surfaces (Chief Scientific Officer, 2002-Present)
GSI *L.L.C*. (President, 2009-Present)

CHEMATICA


SIGMA-ALDRICH is now MilliporeSigma
200,000+ PRODUCTS
500+ SERVICES
Featured INDUSTRIES
Chemistry > Chemical Synthesis > Organic Synthesis Software
Chemistry Products
get started

Organic Synthesis Software

Software and Services Offering

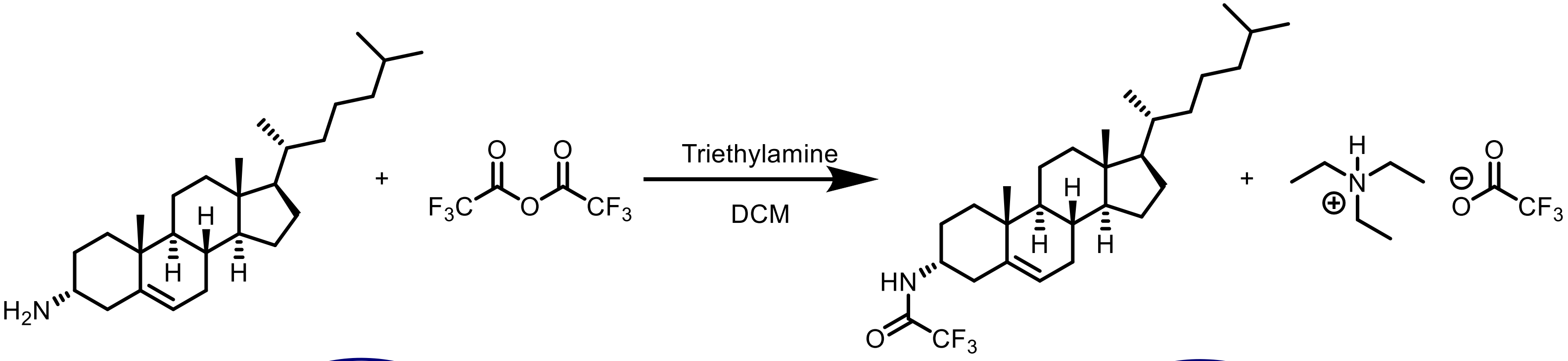
In 2018, customers will be able to choose one of two packages:

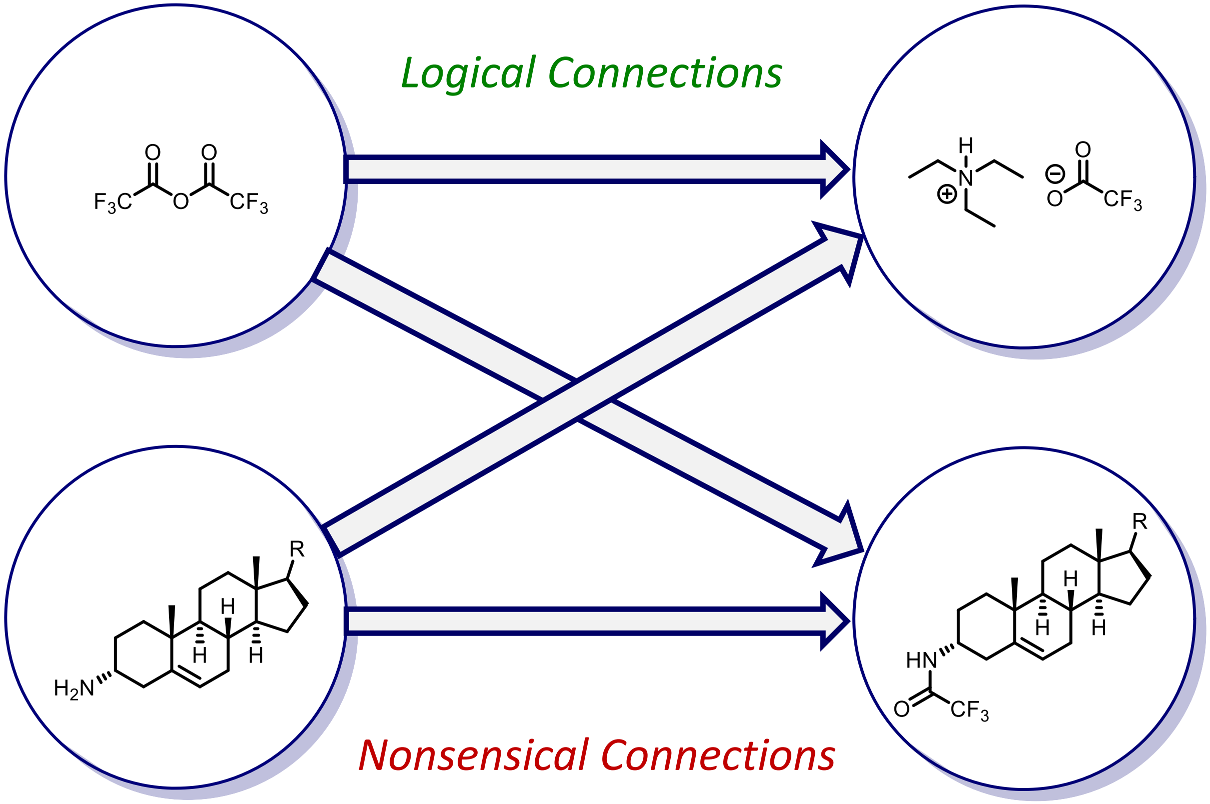
- Service: A mix of consulting- and expert-level software execution based on synthetic style and particular needs.
- License: Allows for access to the software, training, and additional support during a defined period.

Contact us to request Information or a quote.

Our organic synthesis software is only available under limited release at this time.

Position and Moves



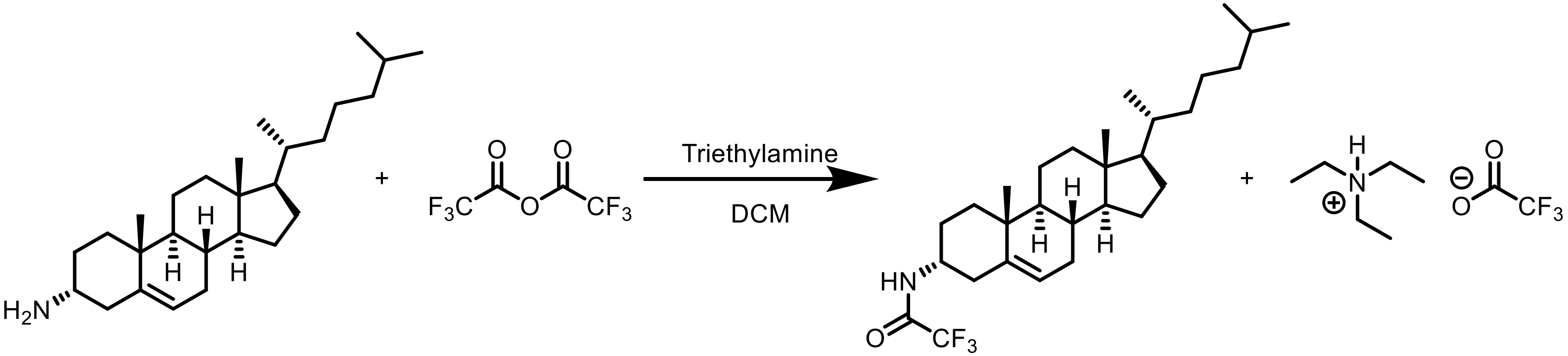


Logical Connections

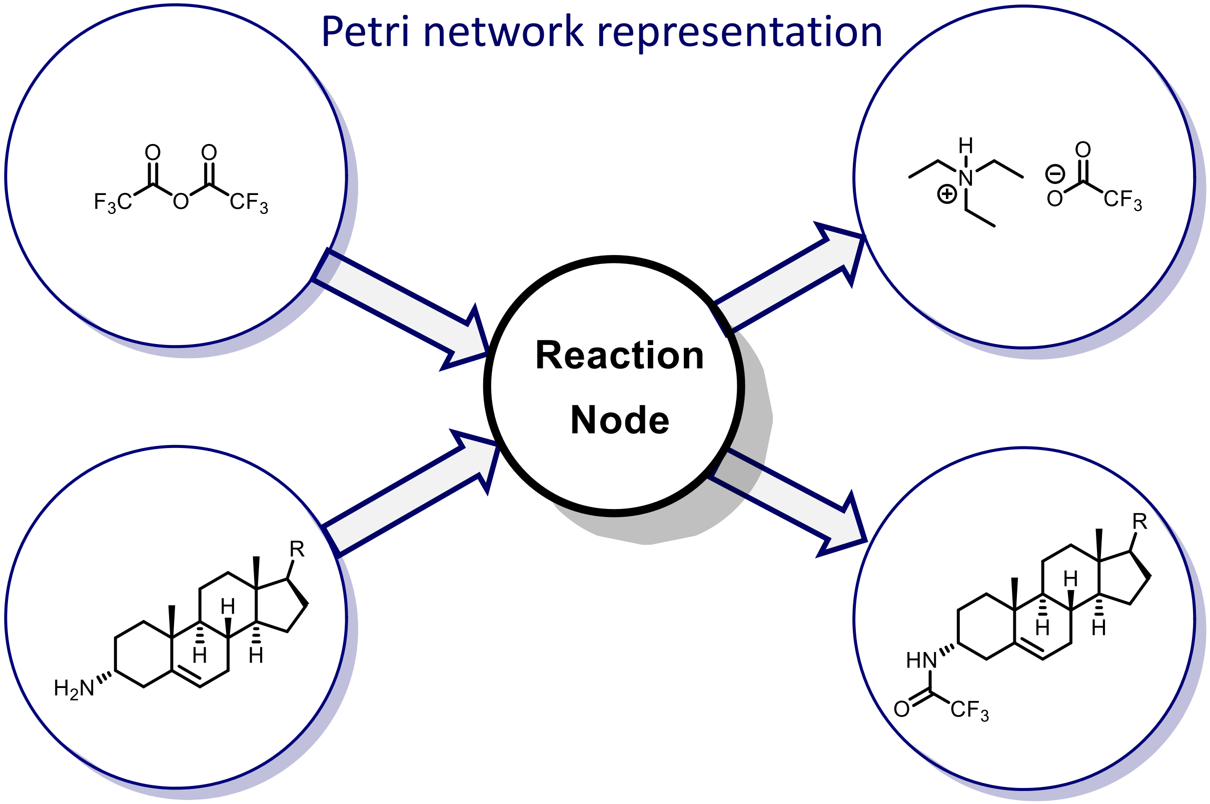
Nonsensical Connections

C. Chaouiya, Briefings Bioinf. 2007, 8,210 –219.
B.A. Grzybowski, K. J. M. Bishop,B.Kowalczyk, C. E. Wilmer, Nat. Chem. 2009, 1,31–36.

Position and Moves



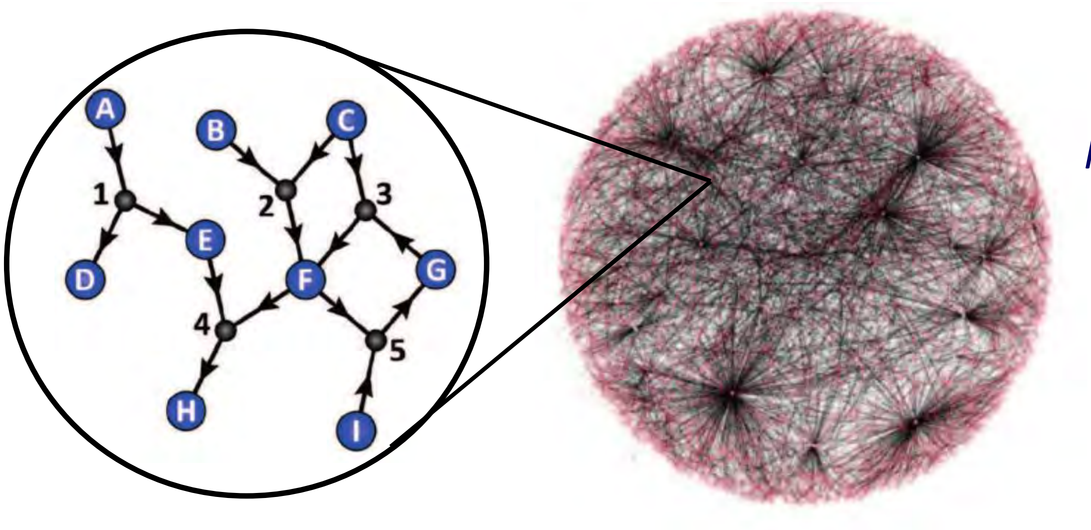
Petri network representation



C. Chaouiya, Briefings Bioinf. 2007, 8,210 –219.

B.A. Grzybowski, K. J. M. Bishop,B.Kowalczyk, C. E. Wilmer, Nat. Chem. 2009, 1,31–36.

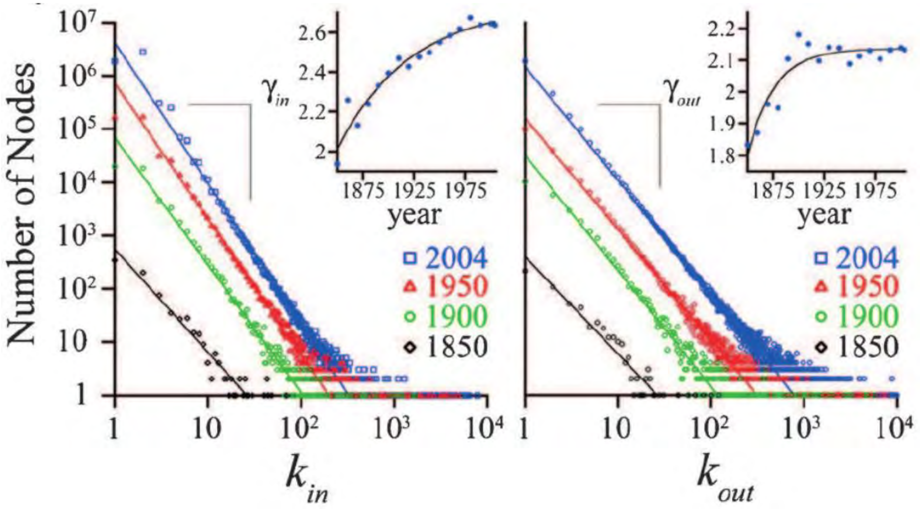
The Network of Organic Chemistry



NOC acquired from Beilstein Database

Over ten million unique structures as SMILES/SMARTS notation with ten million connections

Scale-Free Architecture



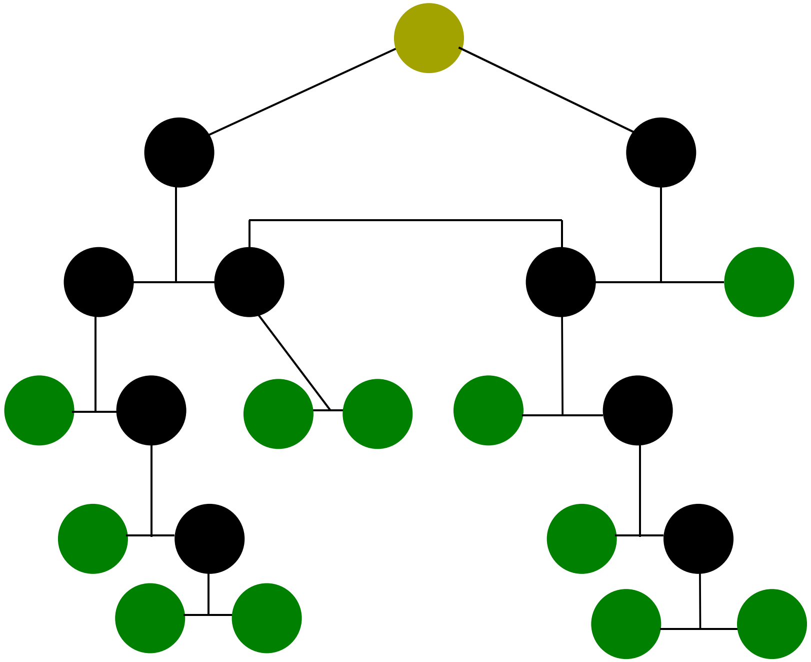
Is it possible to navigate the NOC rapidly generate synthetic pathways of known targets?

M. Fialkowski, K. J. M. Bishop, V. A. Chubukov, C. J. Campbell, B. A. Grzybowski, Angew. Chem. Int. Ed. 2005, 44,7263 –7269

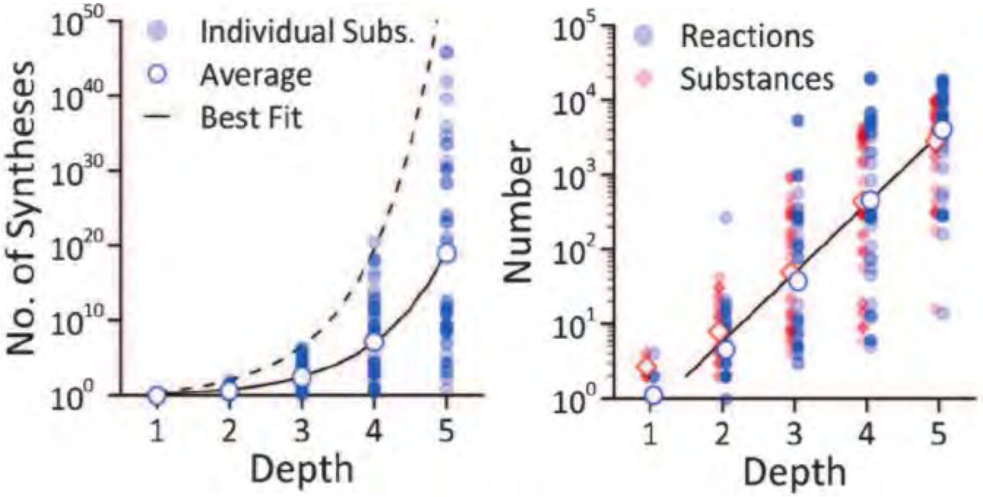
K. J. M. Bishop, R. Klajn, B. A. Grzybowski, Angew. Chem. Int. Ed. 2006, 45,5348 –5354

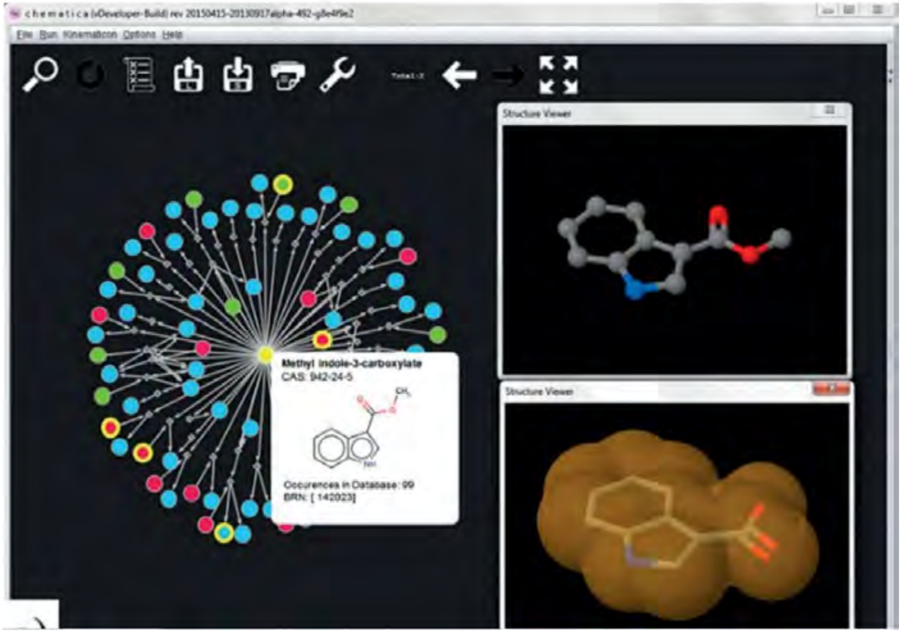
Transversing the NOC

Breadth First vs.
Depth First Searches



“The combinatorial explosion”
-E.J. Corey





chematica (vDeveloper-Build) rev 20150415-20130917alpha-492-g8e4f9e2
File Run Kinematicon Options Help
Total: 2
Structure Viewer
Structure Viewer
CH_3
O
O
NH

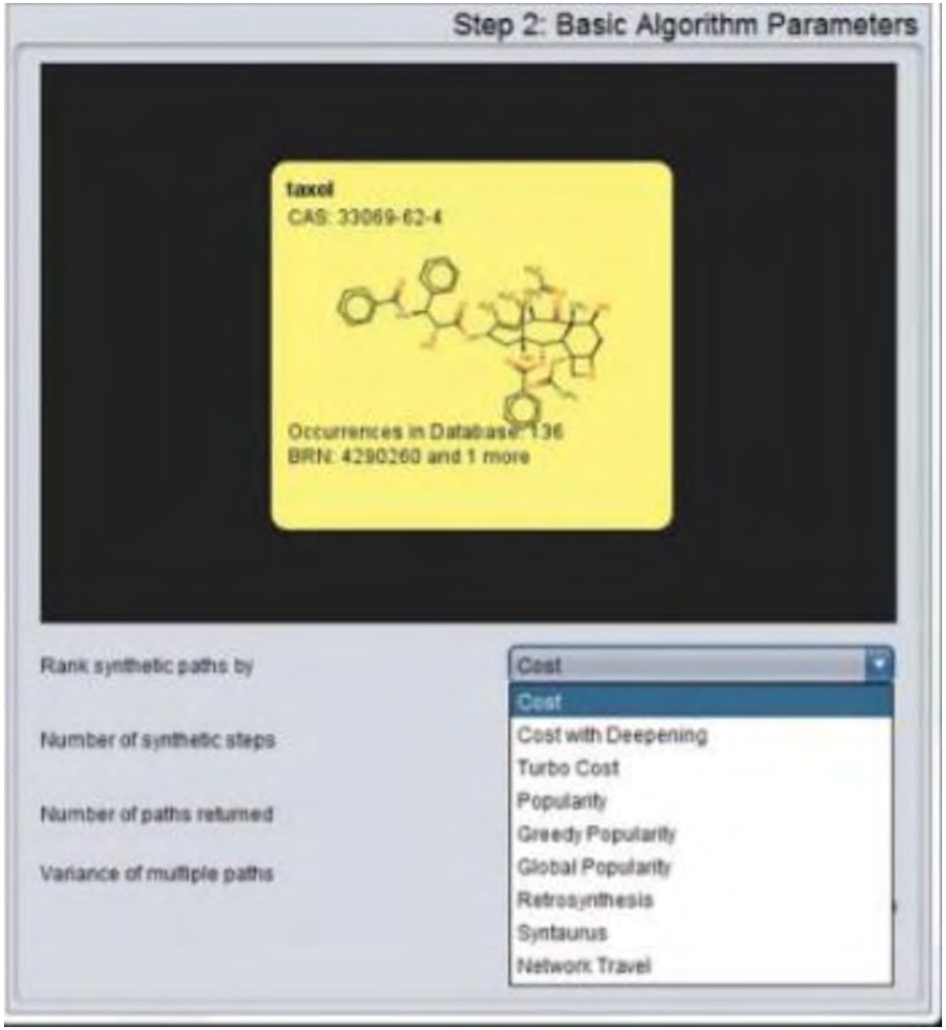
Methyl indole-3-carboxylate
CAS: 942-24-5

Occurences in Database: 99
BRN: [142023]

C. M. Gothard, S. Soh, N. A. Gothard, B. Kowalczyk, Y. H. Wei, B. Baytekin, B. A. Grzybowski, Angew. Chem. Int. Ed. 2012, 51,7922 –7927

Movie Break

Search Criteria and Constraints



Step 2: Basic Algorithm Parameters

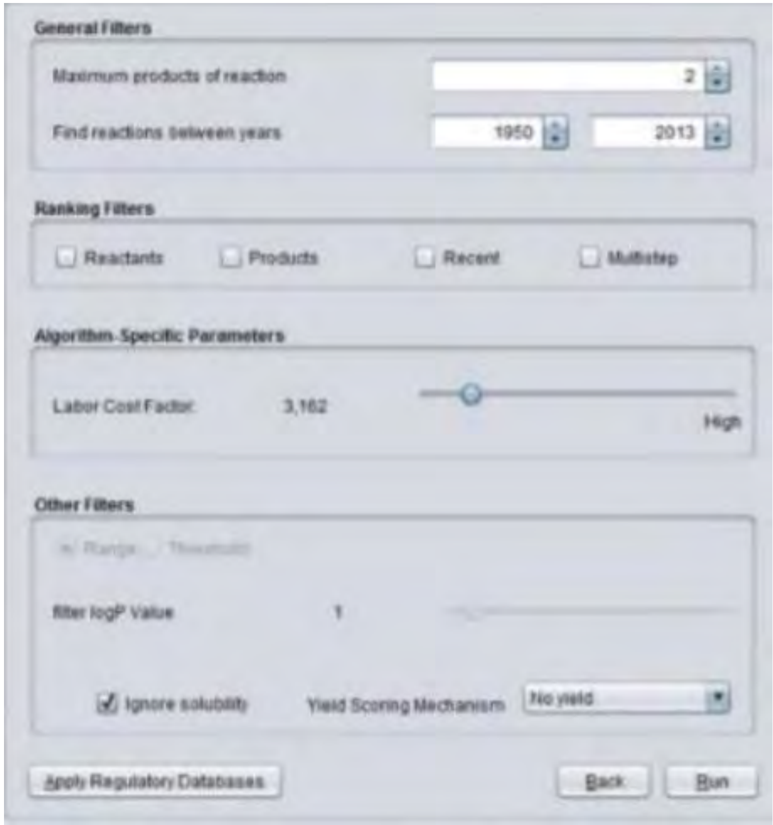
taxol
CAS: 33069-62-4
Occurrences in Database: 136
BRN: 4290260 and 1 more

Rank synthetic paths by: Cost

Number of synthetic steps

Number of paths returned

Variance of multiple paths



General Filters

Maximum products of reaction: 2

Find reactions between years: 1950 2013

Ranking Filters

☐ Reactants ☐ Products ☐ Recent ☐ Multistep

Algorithm-Specific Parameters

Labor Cost Factor: 3,162 High

Other Filters

● Range ○ Threshold

filter logP Value: 1

☑ Ignore solubility

Yield Scoring Mechanism: No yield

Cost:

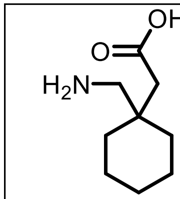
$$C_{tot} = C^{o}_{rxn} N_{rxn} + \sum_i C_{sub}(i)$$

Popularity:

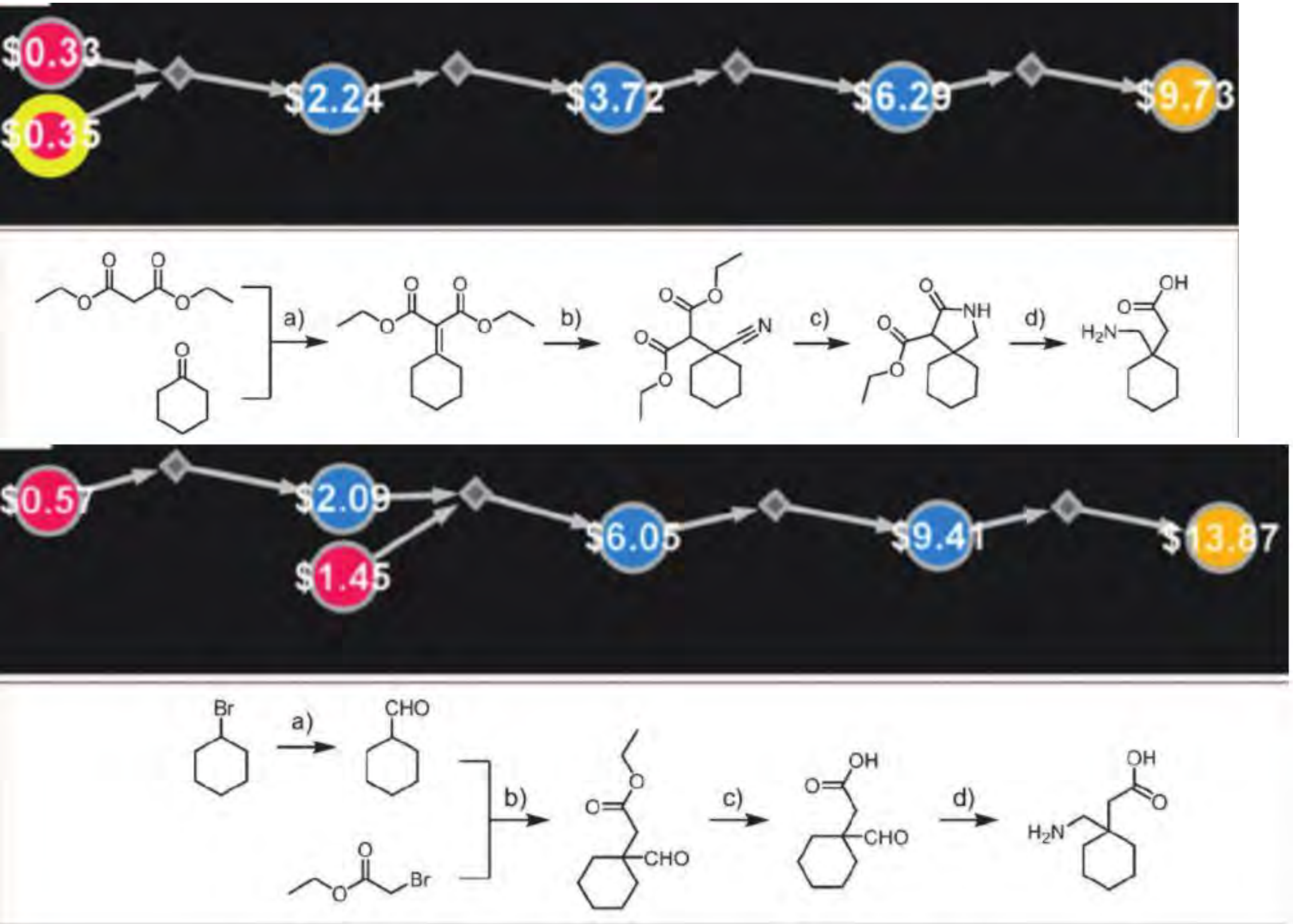
Function of k_{in} / k_{out}

Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

Example: Synthesis of Gabapentin



Gabapentin



Red Nodes Denote Commercially Available Materials

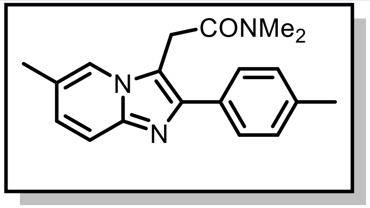
Blue Nodes Denote Synthetic Intermediates

Yellow Halos Denote Controlled Substances

Golden Node Denotes Target Compound

Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

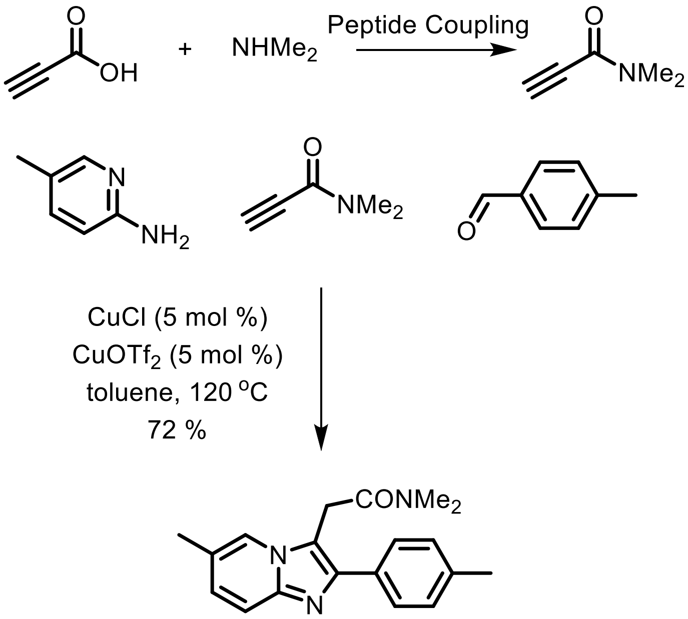
Example: Synthesis of Zolpidem



Zolpidem

C^{o}_{rxn}= 7.5

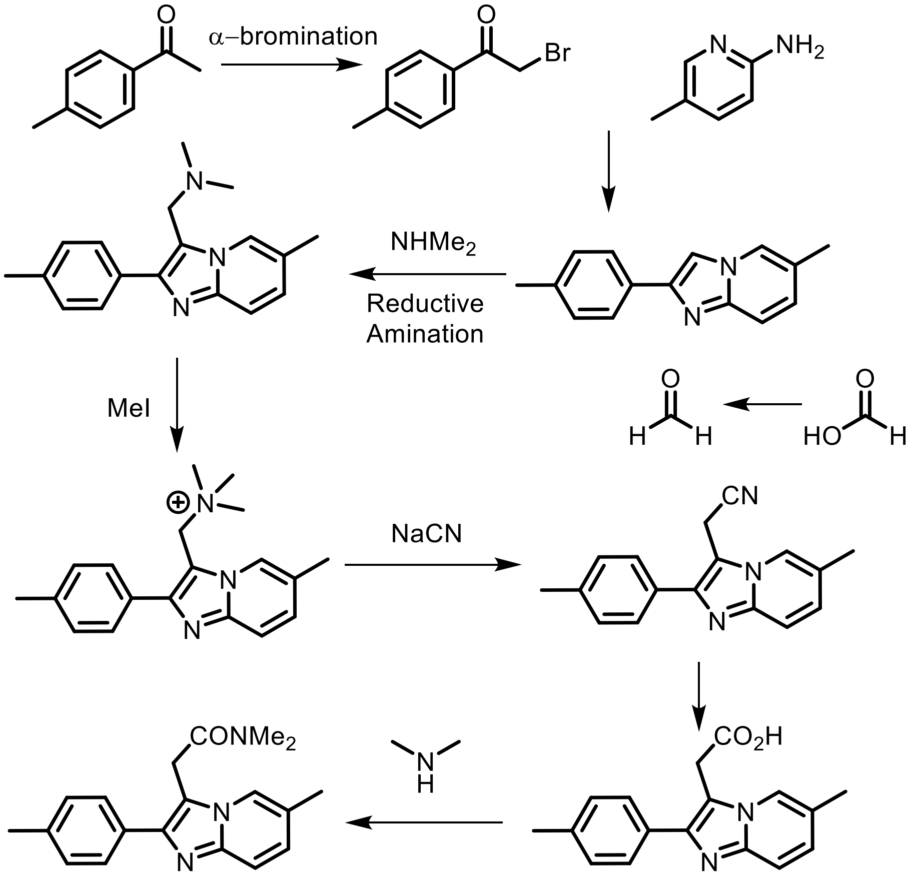
$ 2.76 / g
$ 196 / mol



Angew. Chem. Int. Ed. 2010, 49, 2743 –2746

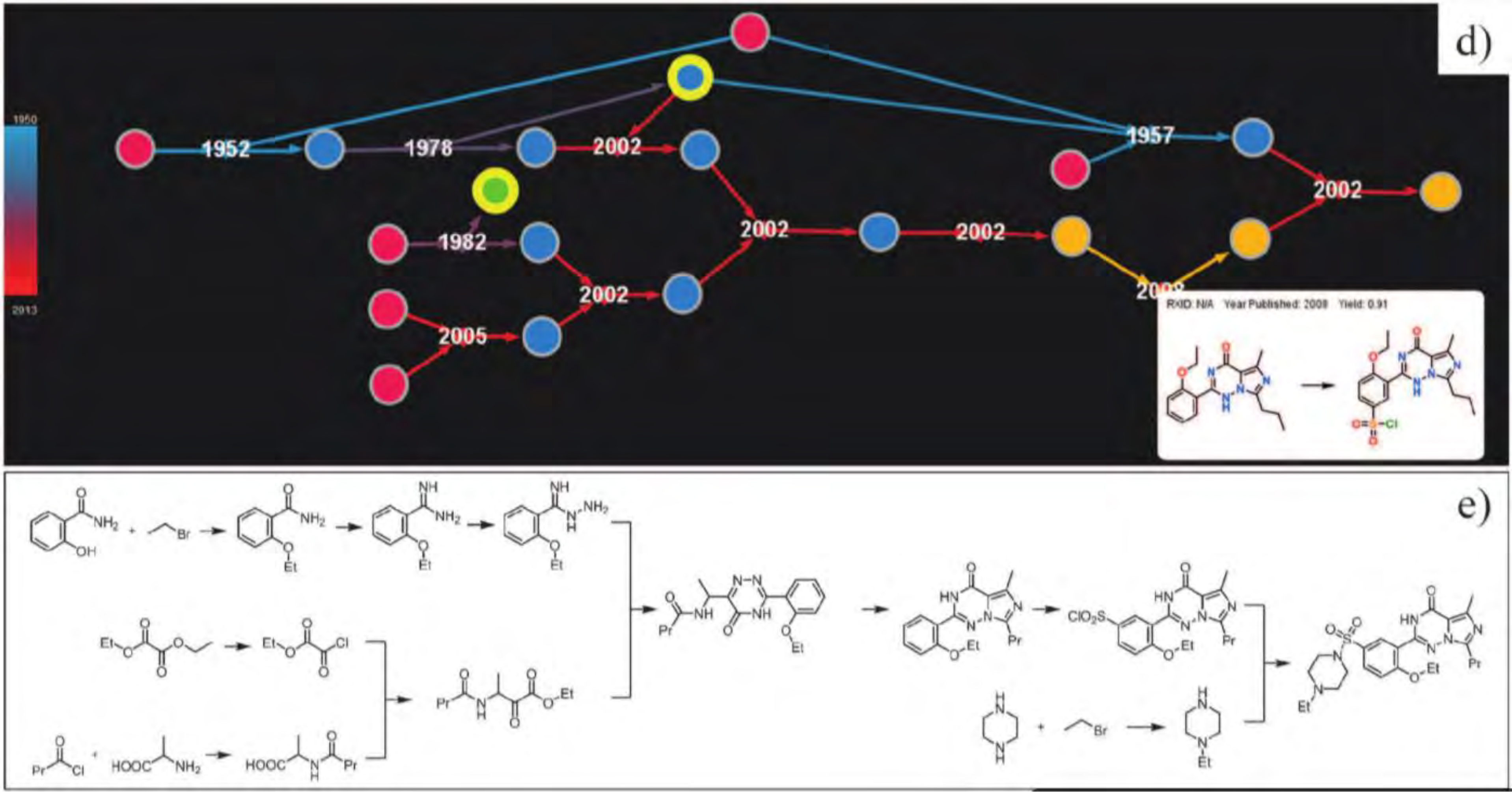
C^{o}_{rxn}= 0.0075

$ 1.61 / g
$175 / mol



Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

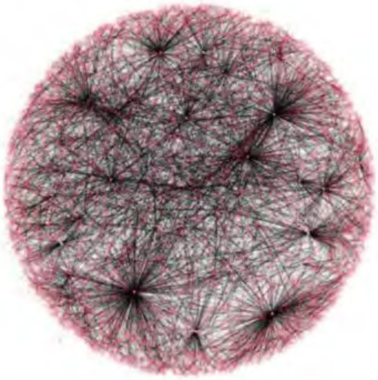
Example: Synthesis of Vardenafil



Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

Movie Break

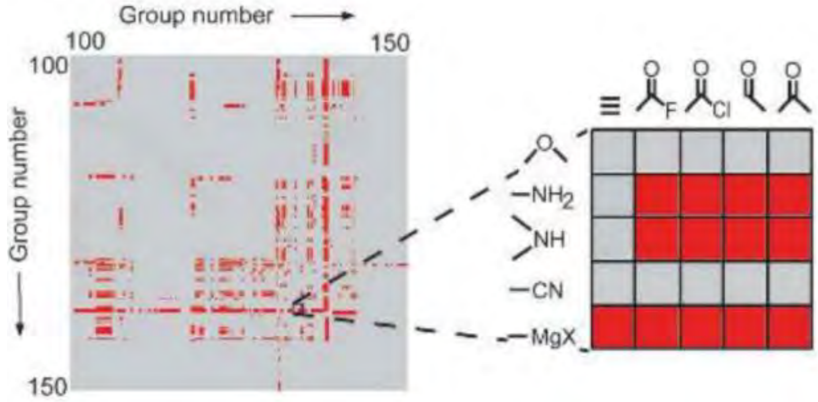
Development of One-Pot Reactions



Assign Molecules by Functional Groups

Compare Classification Against a 322 x 322 master grid of functional group compatibility

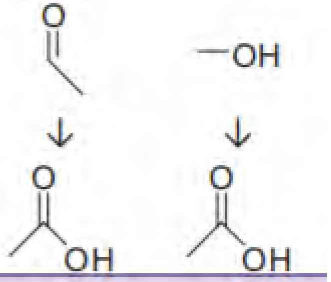
1



If uncompatible groups exists, either: 1) suggest compatible order of addition or 2) omit from candidate combinations

2

Assess if all functional groups are compatible with all reaction conditions



=	0	1
≡	0	1
—X	0	0

3 Solvent Miscibility

6 Acid / Base Compatibility

86,000 Chemical Criteria

4 Aqueous vs Nonaqueous

7 Hydride / Proton Incompatibility

Over 1 million 2-step sequences

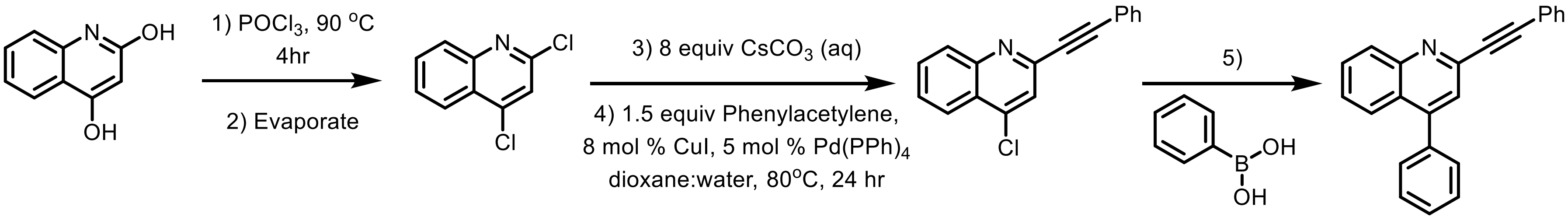
5 Oxidizing vs Reducing

8 Compatibility between Reagents and Functional Groups

14 two-step, 12 3-step, one 4-step sequences experimentally evaluated

Angew. Chem. Int. Ed. 2012, 51, 7922 –7927

Selected Examples of One-Pot Reactions



1) $POCl_3$, 90 °C 4hr

2) Evaporate

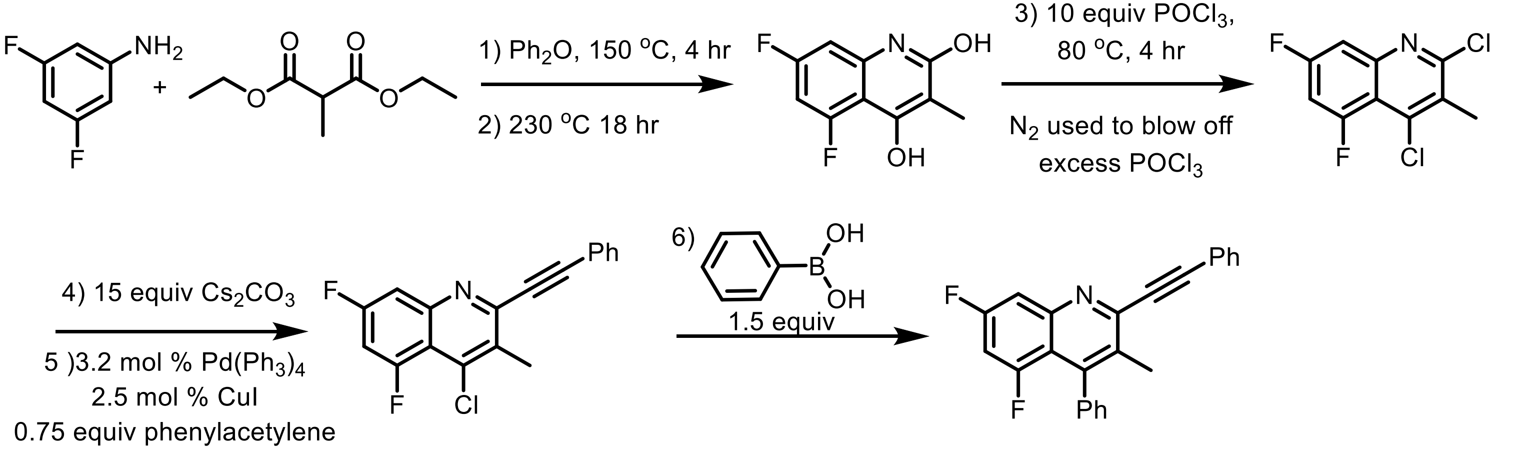
3) 8 equiv $CsCO_3$ (aq)

4) 1.5 equiv Phenylacetylene, 8 mol % CuI, 5 mol % $Pd(PPh)_4$ dioxane:water, 80°C, 24 hr

5)

1.5 equiv

one-pot: 95 %
stepwise: 70 %



1) Ph_2O, 150 °C, 4 hr

2) 230 °C 18 hr

3) 10 equiv $POCl_3$, 80 °C, 4 hr

N_2 used to blow off excess $POCl_3$

4) 15 equiv Cs_2CO_3

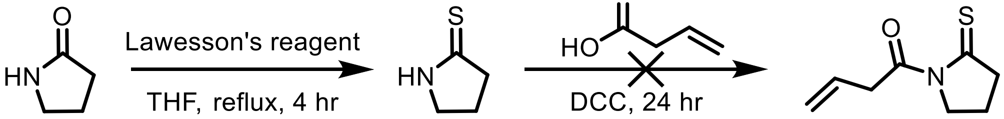
5)3.2 mol % $Pd(Ph_3)_4$
2.5 mol % CuI
0.75 equiv phenylacetylene

6)

1.5 equiv

45 % one-pot yield

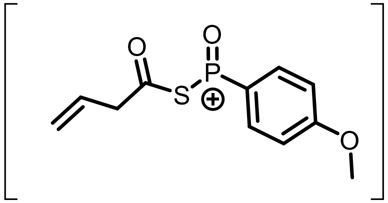
Nothing's Perfect…



Lawesson's reagent

THF, reflux, 4 hr

DCC, 24 hr



Angew. Chem. Int. Ed. 2012, 51, 7922 –7927

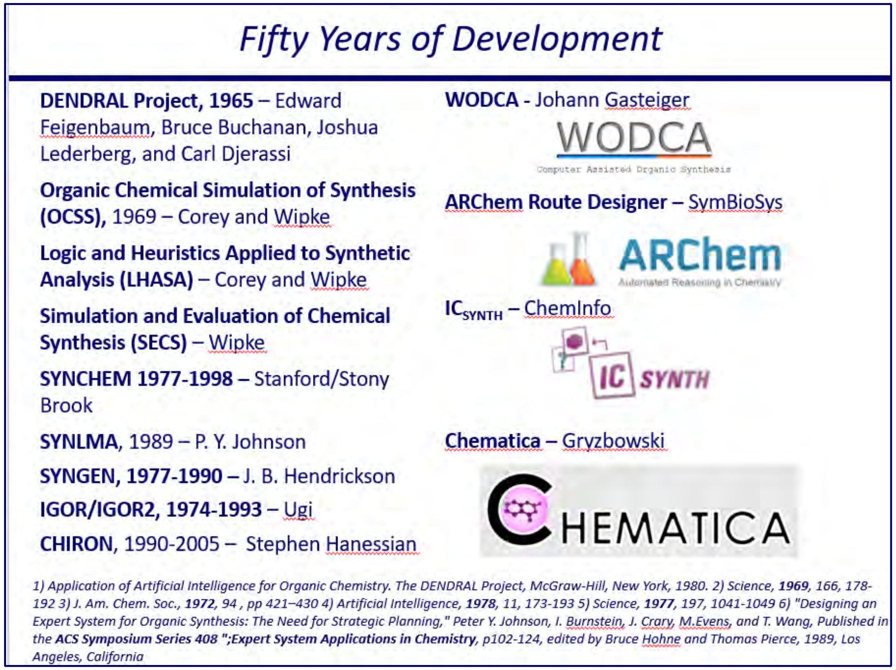
"Intelligent" Retrosynthetic Analysis

Navigating the NOC might make the computer seem smart… but is it really?

1) In NOC, the only positions and moves available are taken directly from the literature – novel transformations or novel compounds are unattainable

2) In NOC, all synthetic positions are static – Expert organic chemists use a dynamic network

How do we make a computer "smart enough" to solve synthetic problems?



Fifty Years of Development

DENDRAL Project, 1965 – Edward Feigenbaum, Bruce Buchanan, Joshua Lederberg, and Carl Djerassi

Organic Chemical Simulation of Synthesis (OCSS), 1969 – Corey and Wipke

Logic and Heuristics Applied to Synthetic Analysis (LHASA) – Corey and Wipke

Simulation and Evaluation of Chemical Synthesis (SECS) – Wipke

SYNCHEM 1977-1998 – Stanford/Stony Brook

SYNLMA, 1989 – P. Y. Johnson

SYNGEN, 1977-1990 – J. B. Hendrickson

IGOR/IGOR2, 1974-1993 – Ugi

CHIRON, 1990-2005 – Stephen Hanessian

WODCA - Johann Gasteiger

ARChem Route Designer – SymBioSys

IC$_{\text{SYNTH}}$ – ChemInfo

Chematica – Gryzbowski

WODCA
Computer Assisted Organic Synthesis
ARChem
Automated Reasoning in Chemistry
IC SYNTH
CHEMATICA

*1) Application of Artificial Intelligence for Organic Chemistry. The DENDRAL Project, McGraw-Hill, New York, 1980. 2) Science, **1969**, 166, 178-192 3) J. Am. Chem. Soc., **1972**, 94 , pp 421–430 4) Artificial Intelligence, **1978**, 11, 173-193 5) Science, **1977**, 197, 1041-1049 6) "Designing an Expert System for Organic Synthesis: The Need for Strategic Planning," Peter Y. Johnson, I. Burnstein, J. Crary, M.Evens, and T. Wang, Published in the **ACS Symposium Series 408 ";Expert System Applications in Chemistry**, p102-124, edited by Bruce Hohne and Thomas Pierce, 1989, Los Angeles, California*

Possible Reasons for failure:

- Needed Better Computers / Algorithms
- Oversimplification of the Problem

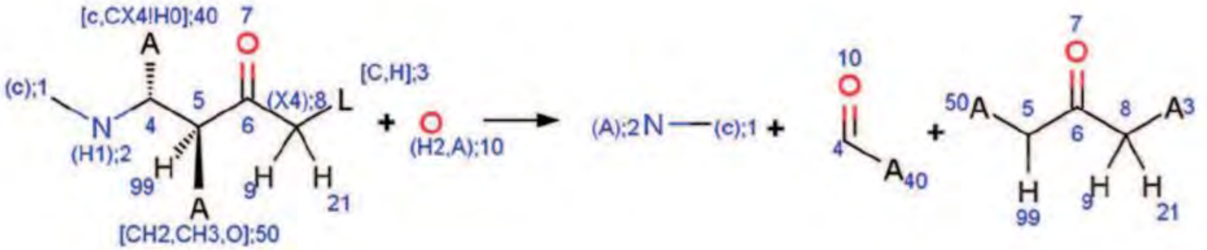
Exhaustive Searches vs. Heuristics

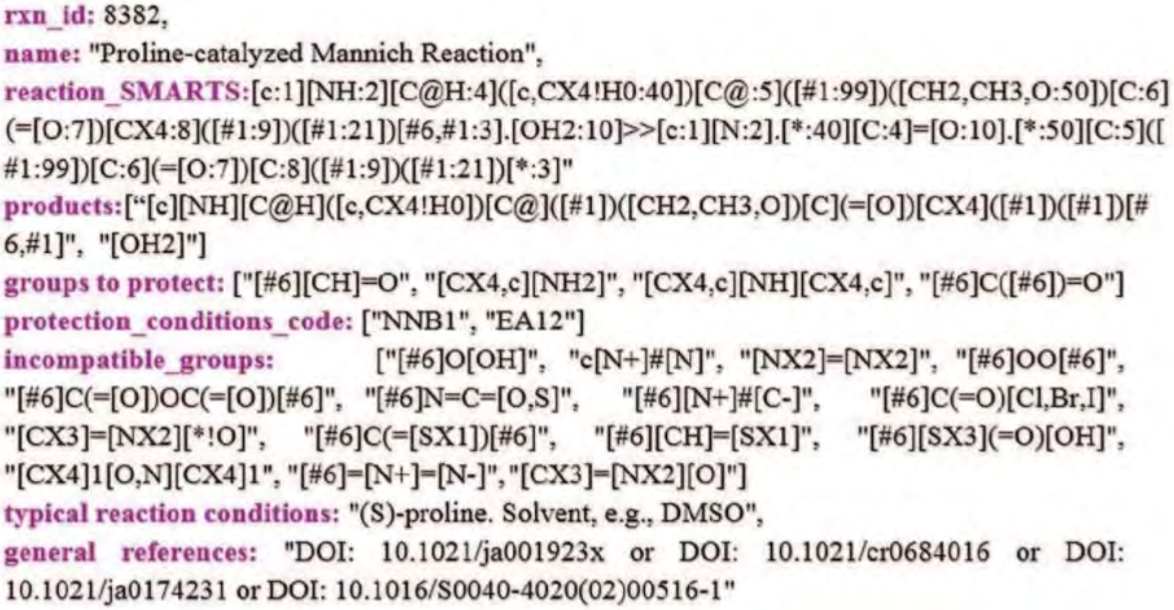
How do we do exhaustive searches on dynamic networks? How does this influence synthetic position? How do we account for sparse events / stereochemistry / molecular context?

Angew. Chem. Int. Ed. 2012, 51, 7922 –7927

Syntaurus

Molecular Representations (SMARTS/SMILES) combined with a set of expert rules




rxn_id: 8382,
name: "Proline-catalyzed Mannich Reaction",
reaction_SMARTS:[c:1][NH:2][C@H:4]([c,CX4!H0:40])[C@:5]([#1:99])([CH2,CH3,O:50])[C:6](=[O:7])[CX4:8]([#1:9])([#1:21])[#6,#1:3].[OH2:10]>>[c:1][N:2].[*:40][C:4]=[O:10].[*:50][C:5]([#1:99])[C:6](=[O:7])[C:8]([#1:9])([#1:21])[*:3]"
products:["[c][NH][C@H]([c,CX4!H0])[C@]([#1])([CH2,CH3,O])[C](=[O])[CX4]([#1])([#1])[#6,#1]", "[OH2]"]
groups to protect: ["[#6][CH]=O", "[CX4,c][NH2]", "[CX4,c][NH][CX4,c]", "[#6]C([#6])=O"]
protection_conditions_code: ["NNB1", "EA12"]
incompatible_groups: ["[#6]O[OH]", "c[N+]#[N]", "[NX2]=[NX2]", "[#6]OO[#6]", "[#6]C(=[O])OC(=[O])[#6]", "[#6]N=C=[O,S]", "[#6][N+]#[C-]", "[#6]C(=O)[Cl,Br,I]", "[CX3]=[NX2][*!O]", "[#6]C(=[SX1])[#6]", "[#6][CH]=[SX1]", "[#6][SX3](=O)[OH]", "[CX4]1[O,N][CX4]1", "[#6]=[N+]=[N-]", "[CX3]=[NX2][O]"]
typical reaction conditions: "(S)-proline. Solvent, e.g., DMSO",
general references: "DOI: 10.1021/ja001923x or DOI: 10.1021/cr0684016 or DOI: 10.1021/ja0174231 or DOI: 10.1016/S0040-4020(02)00516-1"

Over 20,000 rules (in 2016, Aldrich website says 50,000), 200,000 specialized reactions in addition to "conditional rules of chemistry"

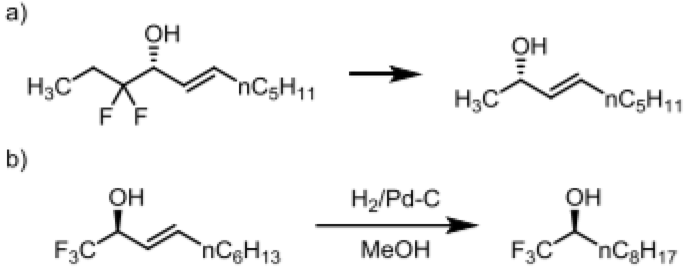
Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

Early Failures

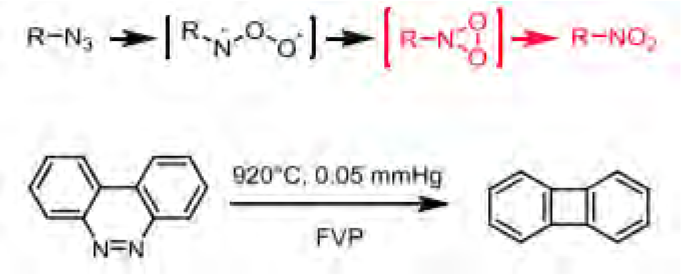
Attempt 1: Use machine-extracted transforms as rules

– ca. 115,000 unique reaction classes

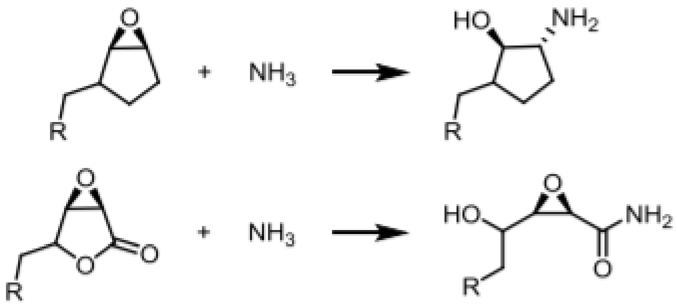
Errors in Databases



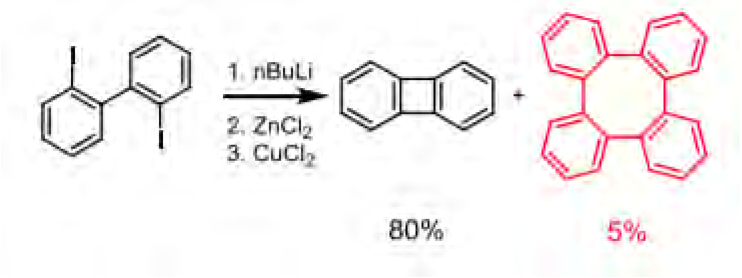
Non-synthetically Useful Conditions



"Context dependent" Cases

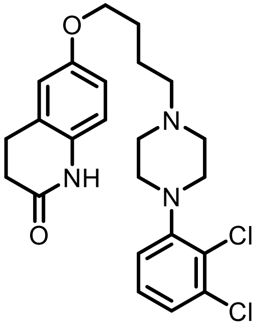


Relative Abundance of Products



Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

Group Problem 1



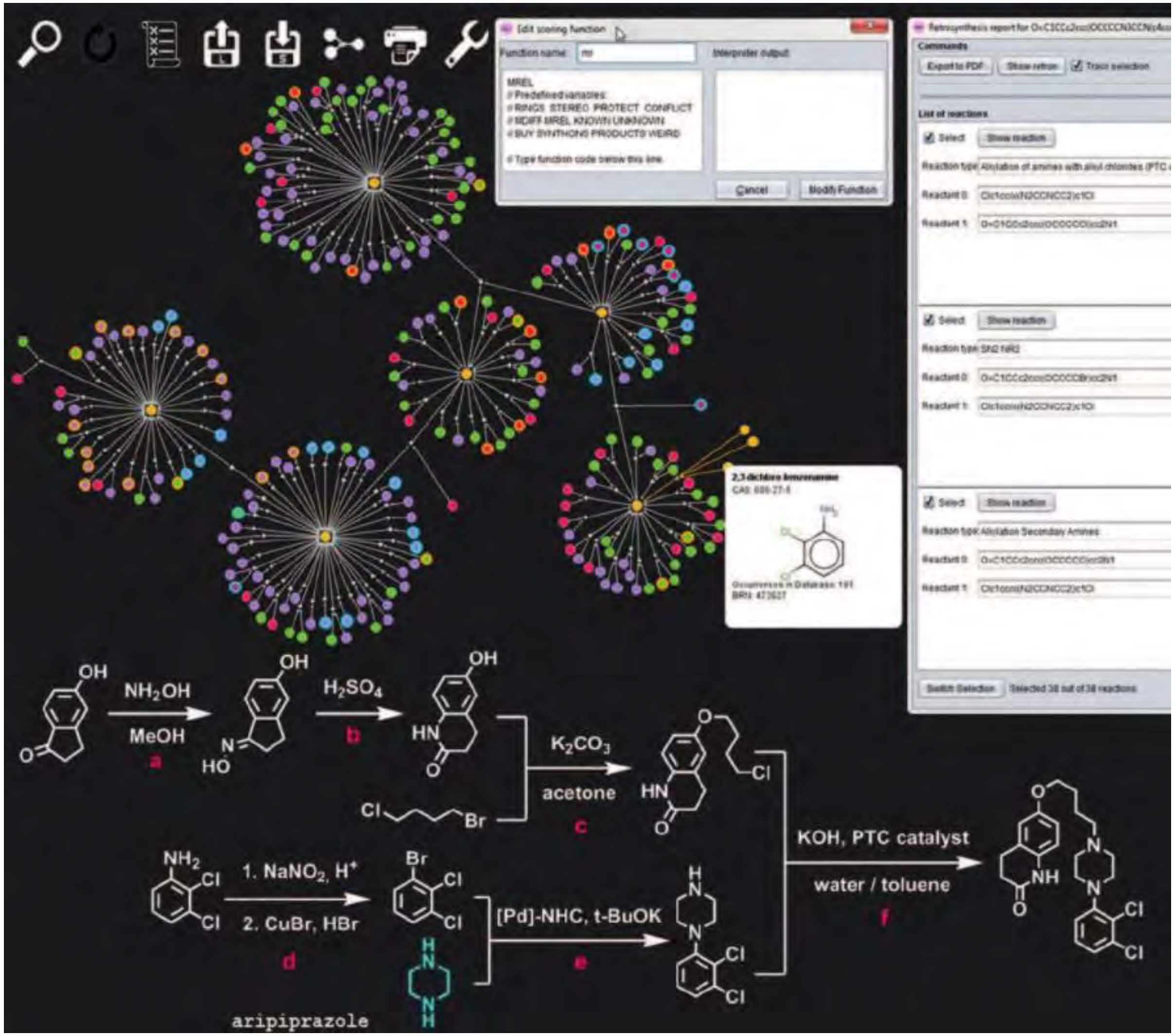
Aripiprazole

Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

Group Problem 1

Built-in script language for custom scoring:

- MREL: favorably scores substrates of comparable molecular weights
- STEREO: favorably scores creating enriched stereocenters
- RINGS: favorably scores creating rings
- BUY: promotes substrates that are commercially available
- CONFLICT: Penalizes reactions with functional group incompatibility
- PROTECT: Penalizes the necessity for protecting groups



Edit scoring function

Function name: rel

```
MREL
// Predefined variables:
// RINGS STEREO PROTECT CONFLICT
// SDIFF MREL KNOWN UNKNOWN
// BUY SYNTHONS PRODUCTS WEIRD

// Type function code below this line
```

Interpreter output

Retrosynthesis report for O=C1CCc2ccc(OCCCCN3CCN(c4c

Commands

☑ Trace selection

List of reactions

☑ Select

Reaction type: Alkylation of amines with alkyl chlorides (PTC

Reactant 0: Clc1cccc(N2CCNCC2)c1Cl

Reactant 1: O=C1CCc2ccc(OCCCCCl)cc2N1

☑ Select

Reaction type: SN2 NR2

Reactant 0: O=C1CCc2ccc(OCCCCBr)cc2N1

Reactant 1: Clc1cccc(N2CCNCC2)c1Cl

☑ Select

Reaction type: Alkylation Secondary Amines

Reactant 0: O=C1CCc2ccc(OCCCCCl)cc2N1

Reactant 1: Clc1cccc(N2CCNCC2)c1Cl

Selected 38 out of 38 reactions

2,3-dichloro benzenamine

CAS: 608-27-5

Occurrences in Database: 181

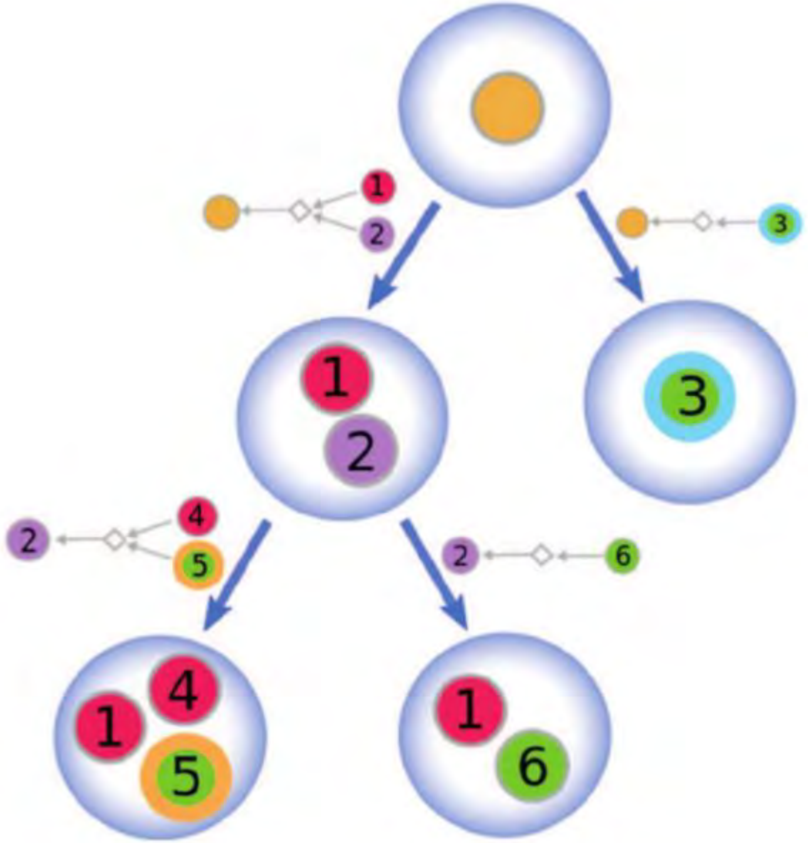
BRN: 473627

Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

Automation

Individual searches are inefficient – manual selection of each step makes finding the optimal route highly improbable and very time consuming.

Searching on a Dynamic Network is more difficult – no information is available about subsequent layers and evaluation of step *and* position is necessary



Chemical Scoring Function

RINGS, STEREO, MASS, SMILES_LEN, KNOWN, WEIRD, BUY, KNOWN

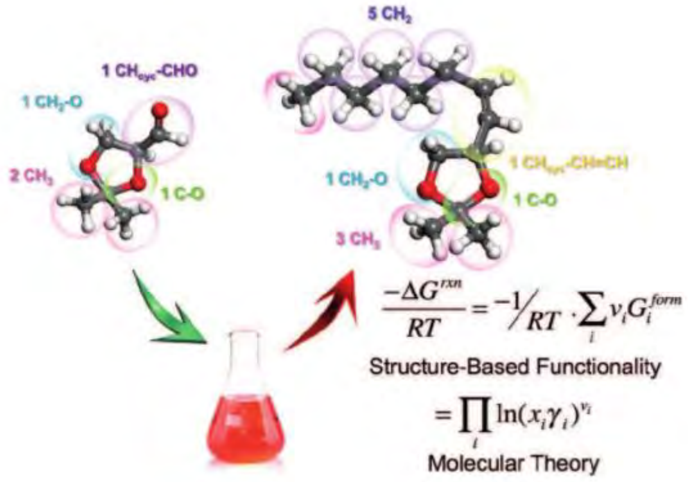
Reaction Scoring Function

PROTECT, CONFLICT, YIELD

If CSF = 0 and RSF = 1:
Each reaction step costs +1
– minimizes number of steps

Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

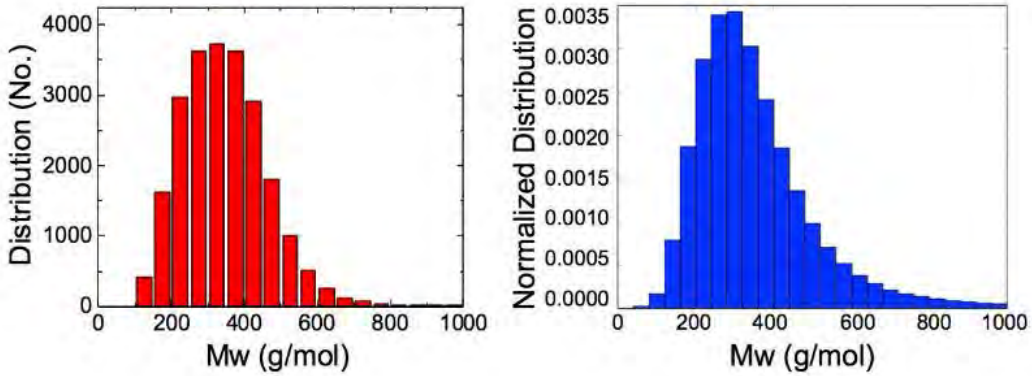
Estimation of Reaction Yields


5 CH₂
1 CH$_{cyc}$-CHO
1 CH₂-O
2 CH₃
1 C-O
1 CH₂-O
1 CH$_{cyc}$-CH=CH
1 C-O
3 CH₃
$\frac{-\Delta G^{rxn}}{RT} = -1/RT \cdot \sum_i \nu_i G_i^{form}$
Structure-Based Functionality
$= \prod_i \ln(x_i \gamma_i)^{\nu_i}$
Molecular Theory

Concept: Calculate accurate G^{form} and use them to calculate ΔG^{rxn}, which should then correlate with yield.

Assumptions: Most reactions are under thermodynamic control, the training set is representative of most chemical reactions

1) Obtain Training Set of 23,000 reported reactions (MW of reactant 100-1000 g / mol)



2) Decompose training molecules into 296 distinct functional groups and assign guess ΔG^{form} values

⇩

3) $\Sigma\Delta G^{form} = \Delta G^{calc}$

ΔG^{calc} does not account for non-ideality.

Must use experimental yields with perturbed-chain statistical associating fluid theory to attain more accurate values.

Angew. Chem. Int. Ed. 2015, 54,10797 –10801

Estimation of Reaction Yields

4a) Solve for mole fraction

$$\xi = (n_i^o - x_i n^o) \,/\, (x_i \nu - \nu_i)$$

ξ = Experimental Yield

n_i^o = Initial mols of *i*

x_i = mol fraction *i*

n^o = total number of initial mols

ν = total stoichiometry coefficient

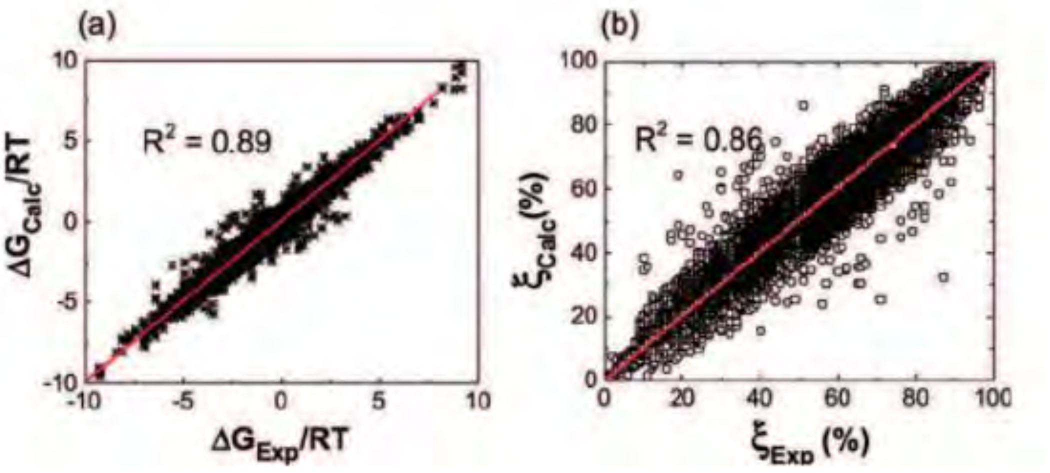
ν_i = stoichiometry coefficient for *i*

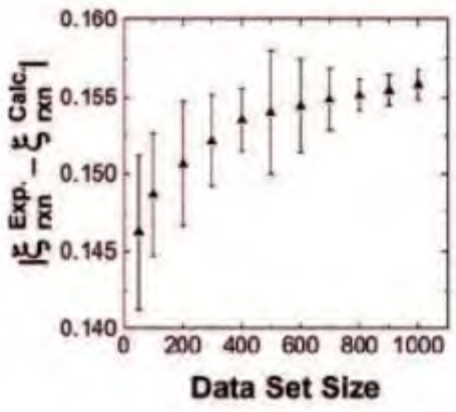
4b) Use PC-SAFT to calculate γ

$$\Delta G^{exp} = -\mathrm{RT}\ln \prod (x_i \gamma_i)^{\nu_i}$$

5) Optimize ΔG^{form} to fit experimental data

$$OBJ = \sqrt{(\Delta G^{exp} - \Delta G^{calc})^2/2}$$





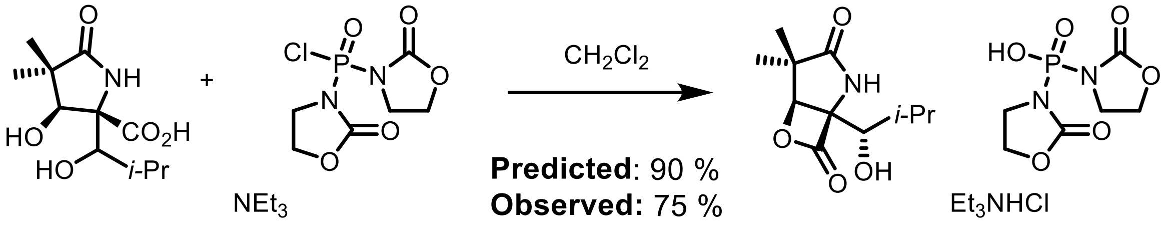
Approx. 15 % error – still good enough to provide qualitative assessment.

Seminal publication of a this approach to predict reaction yield.

Angew. Chem. Int. Ed. 2015, 54,10797 –10801

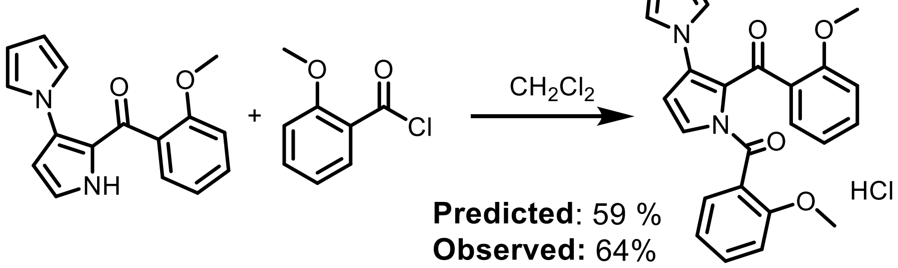
Selected Examples of Yield Prediction

(-)-7-Methylmuralide (Corey and Shenvi)

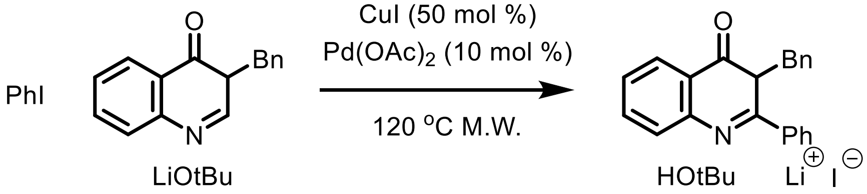


Predicted: 90 %
Observed: 75 %

Marinopyrrole A (Nicolaou)



Predicted: 59 %
Observed: 64%

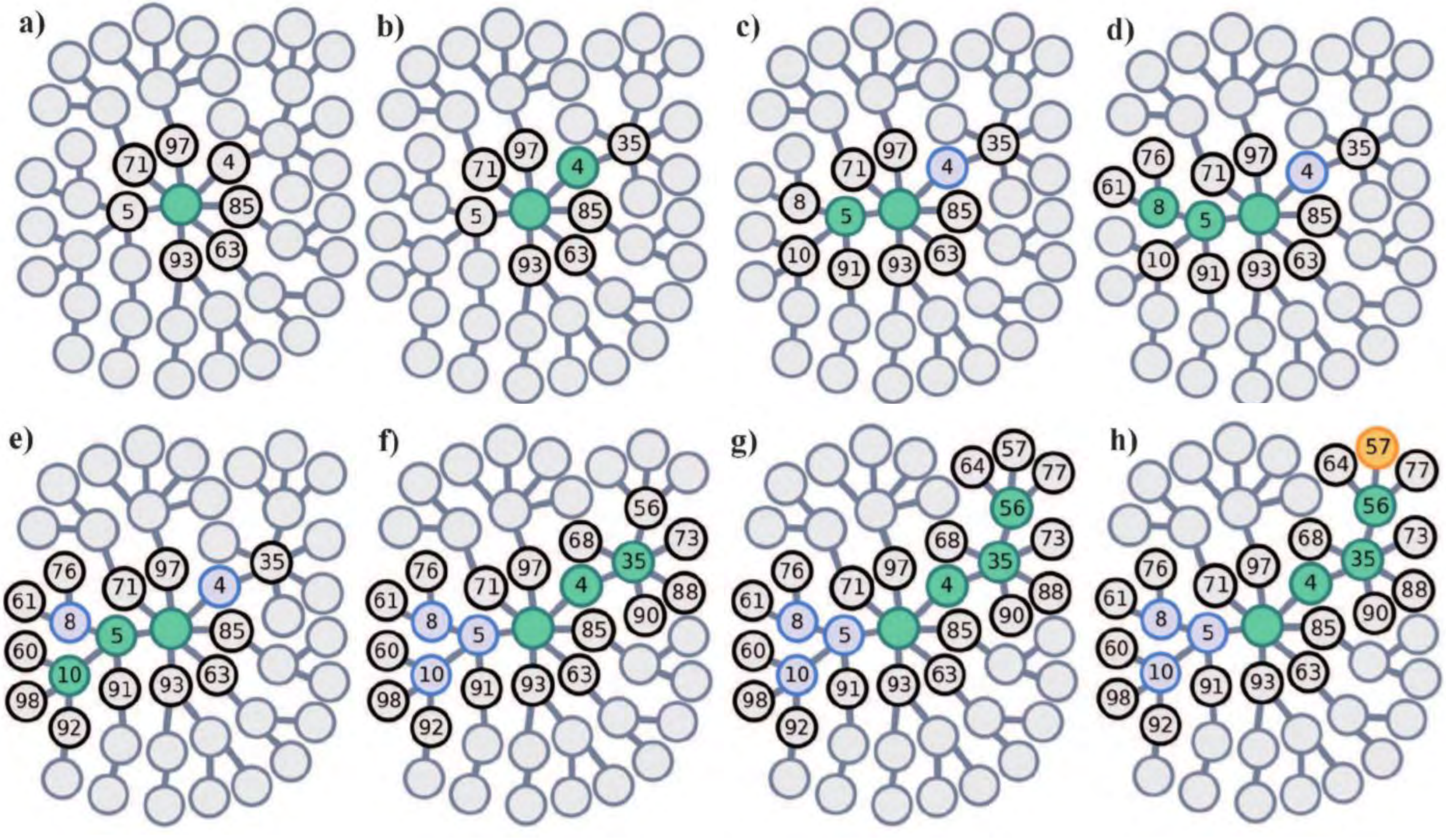


Solvent	Yield (P/O)
DMF	91% / 82 %
DMA	43% / 59 %
DMPU	0 % / 12 %

Angew. Chem. Int. Ed. 2015, 54,10797 –10801

Searching in Syntaurus

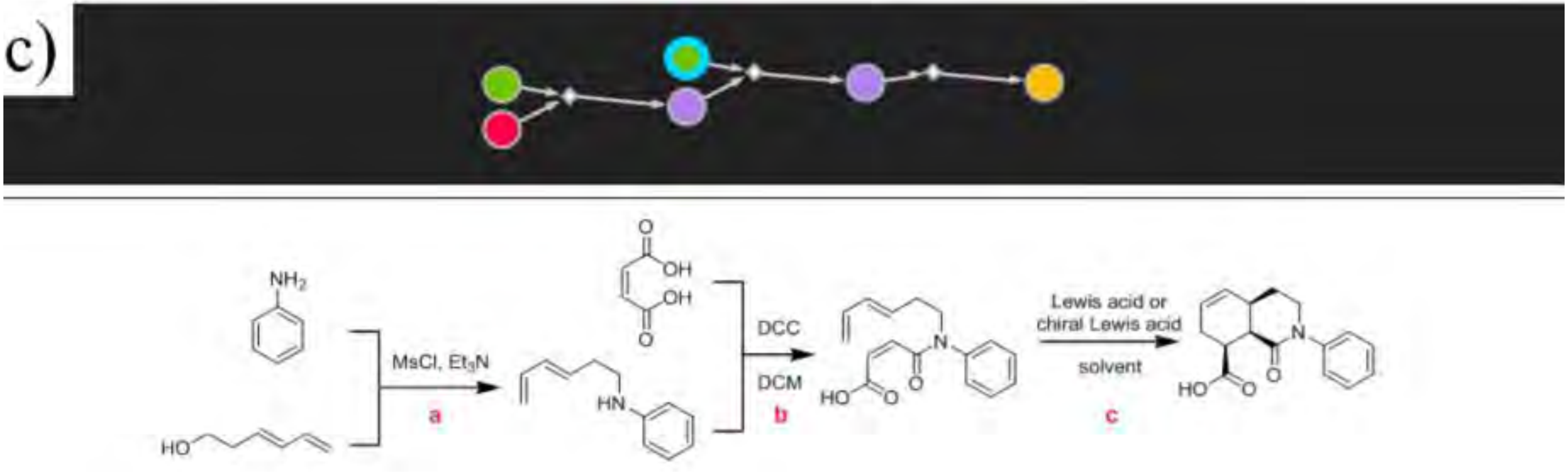
Search algorithm should be 1) non-local 2) strategizing and 3) self-correcting

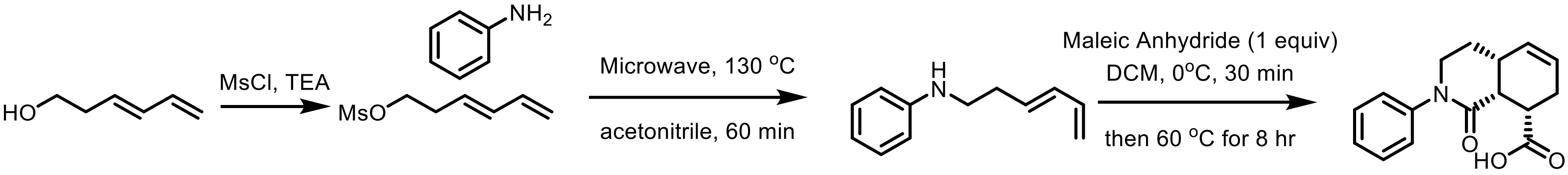


Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

Movie Break

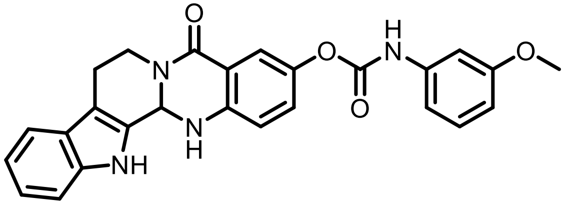
“Rediscovery” of Published Synthesis

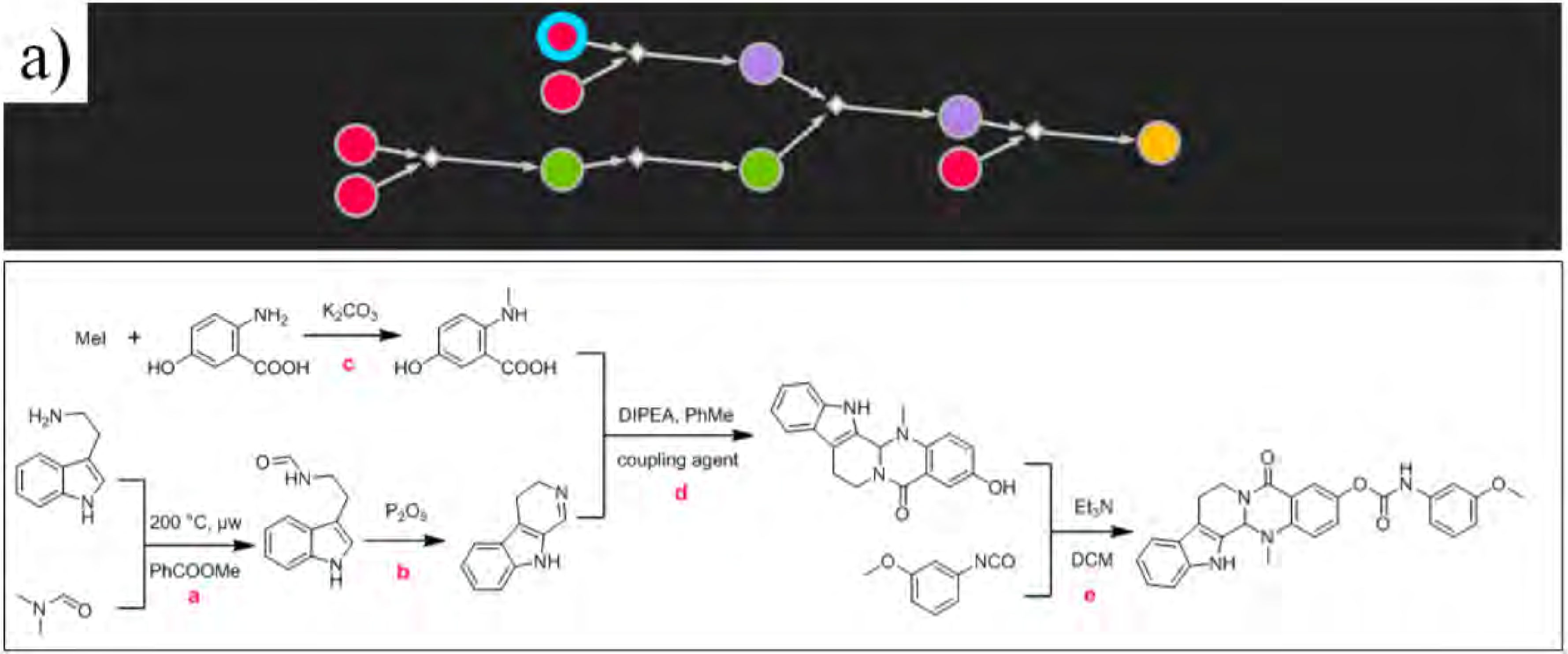




Tantillo, Aubé, *et al.* J. Org. Chem. 2015, 80, 5260−5271

Rediscovery of Published Synthesis

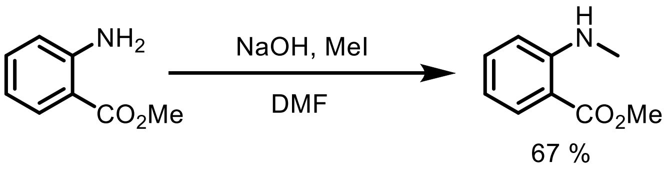




Angew. Chem. Int. Ed. 2016, 55, 5904 – 5937

Rediscovery of Published Synthesis

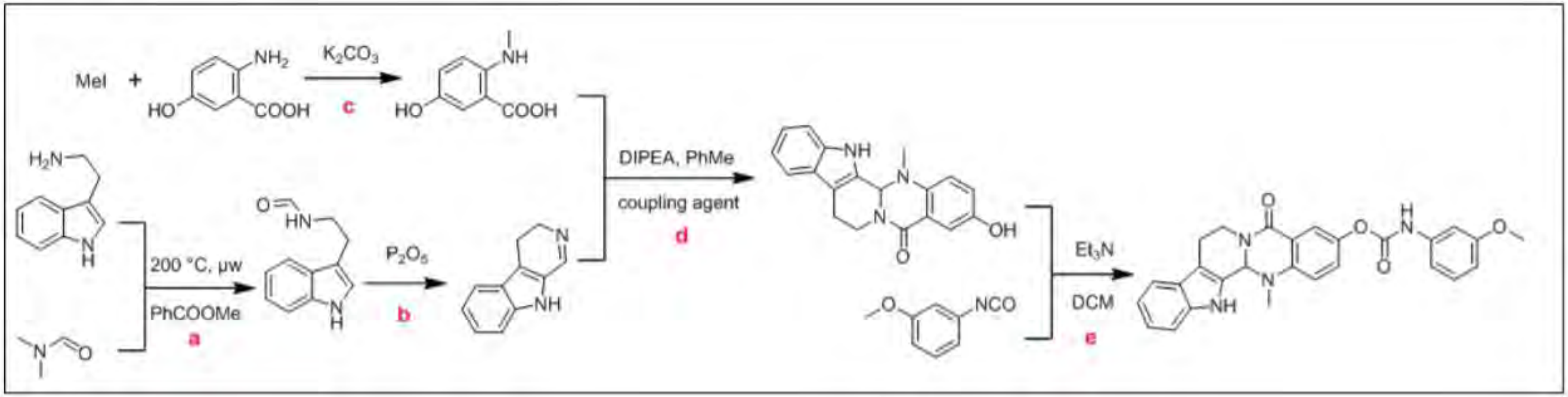
C



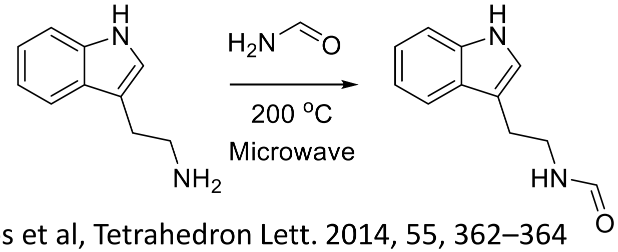
Yuan et al, Med. Chem. Res. 2014, 23, 2169- 2177

E: Exact Procedure

Decker et al, Eur. J. Med. Chem. 2014, 81, 15–21



A

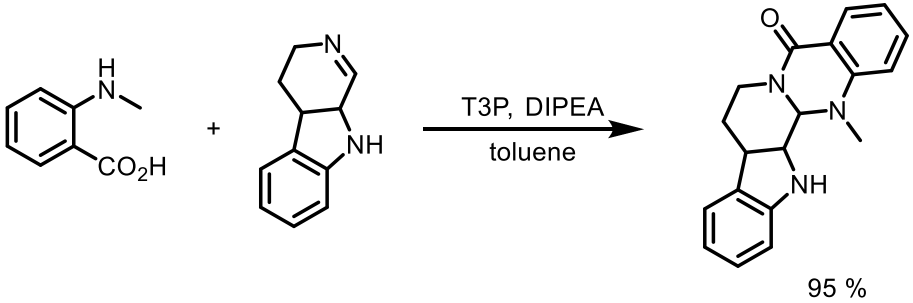


Legros et al, Tetrahedron Lett. 2014, 55, 362–364

B: Exact Procedure

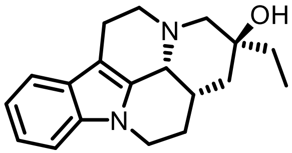
Dodd et al, Bioorg. Med. Chem. 2001, 9, 2155–2164

D

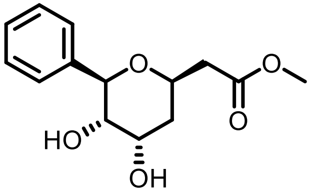


Taylor et al, Org. Lett. 2013, 15, 258–261

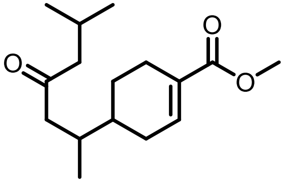
Group Problem 2



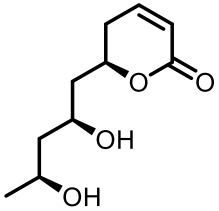
tacamonidine



goniothalesdiol A

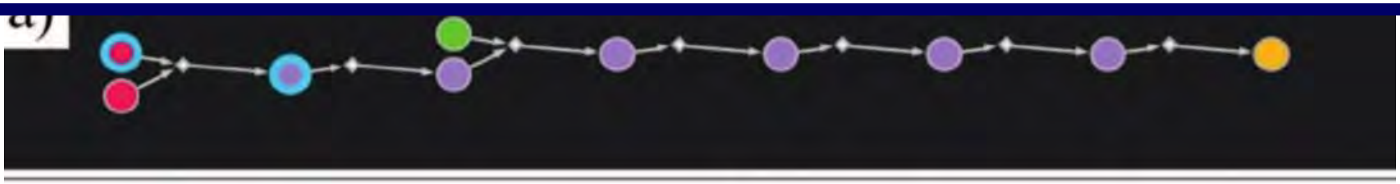


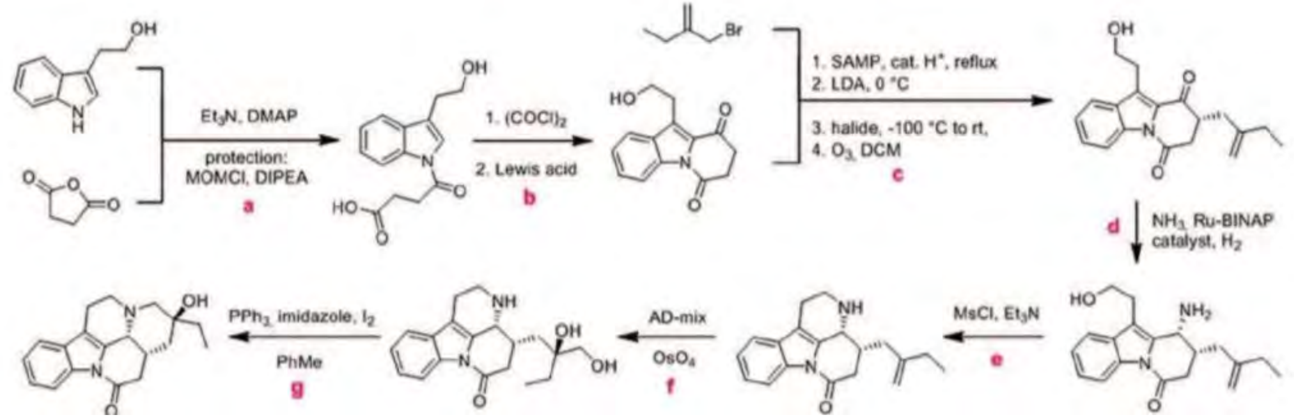
juvabione



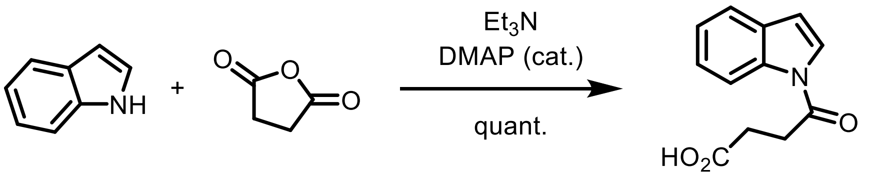
(*R*)-6-((2*S*,4*S*)-2,4-dihydroxypentyl)-5,6-dihydro-2*H*-pyran-2-one

Group Problem 2

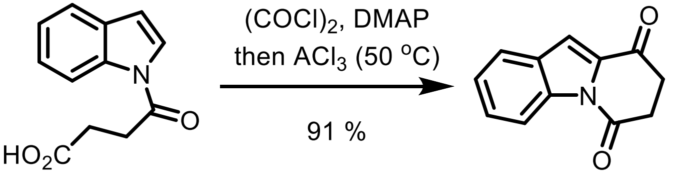




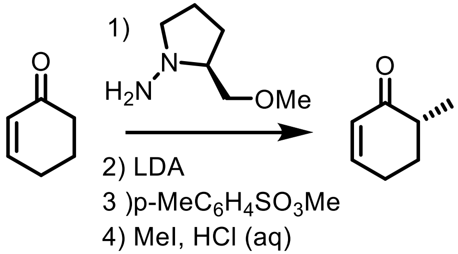
A



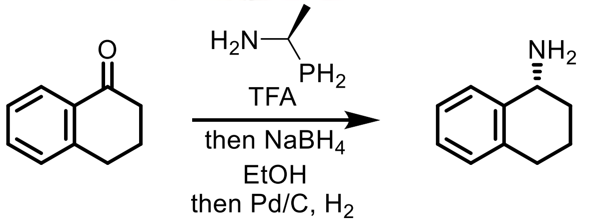
B



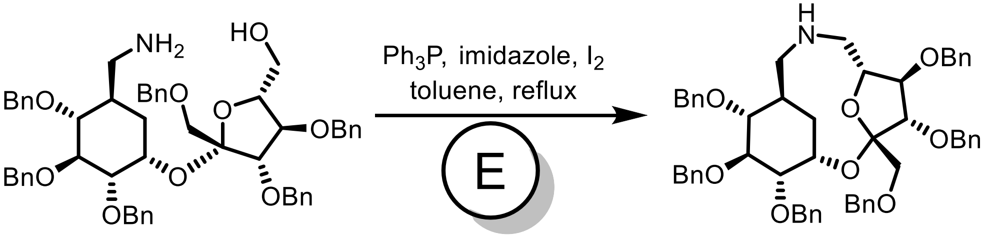
C



D

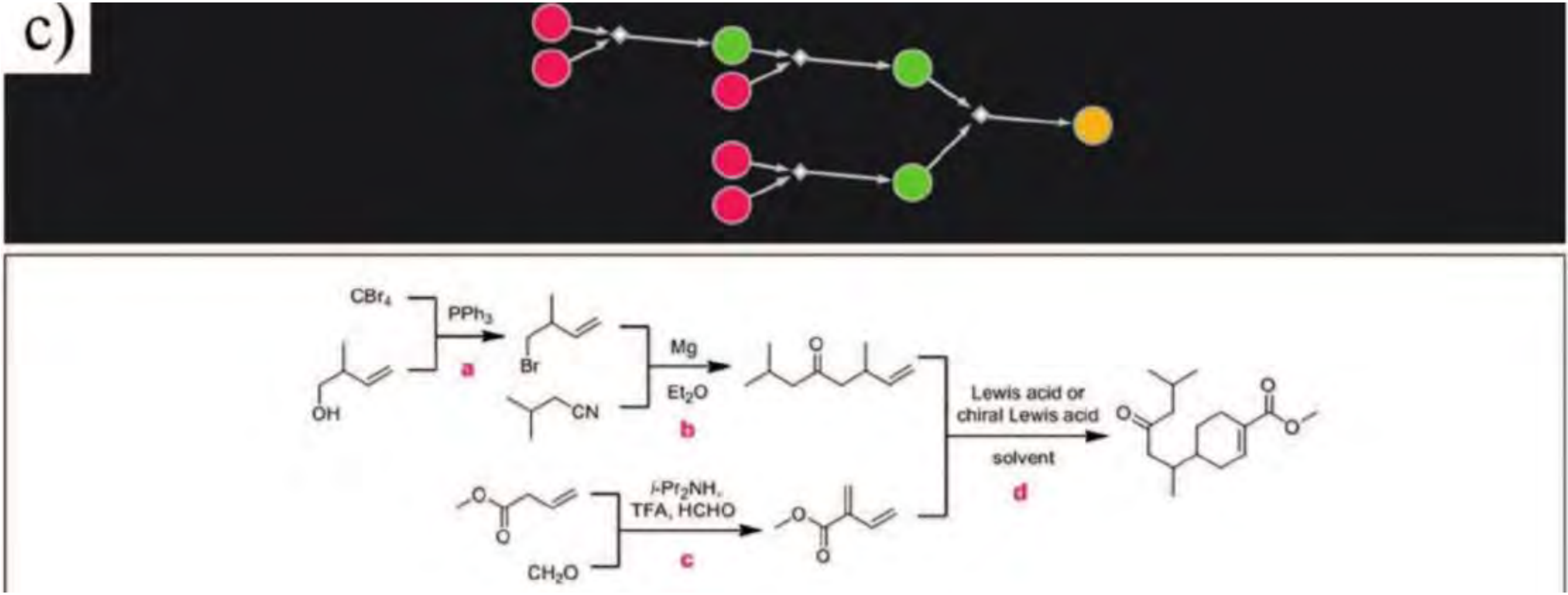


E



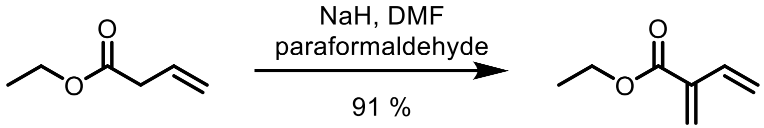
. Kam, J. Nat. Prod. 2015, 78, 1129 – 1138, Han, Chem. Eur. J. 2012, 18, 9784 – 9788, Tokoroyama, J. Chem. Soc. Chem. Commun. 1987, 358 – 359, Lazny, Synlett 1998, 721 – 722, F. Polyak, Tetrahedron: Asymmetry 1998, 9, 4369 – 4379, Jarosz, Tetrahedron: Asymmetry 2012, 23, 1474 – 1479

Group Problem 2

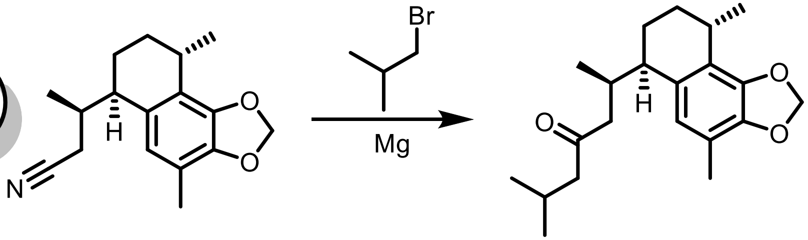


A **Straight from Literature**

C

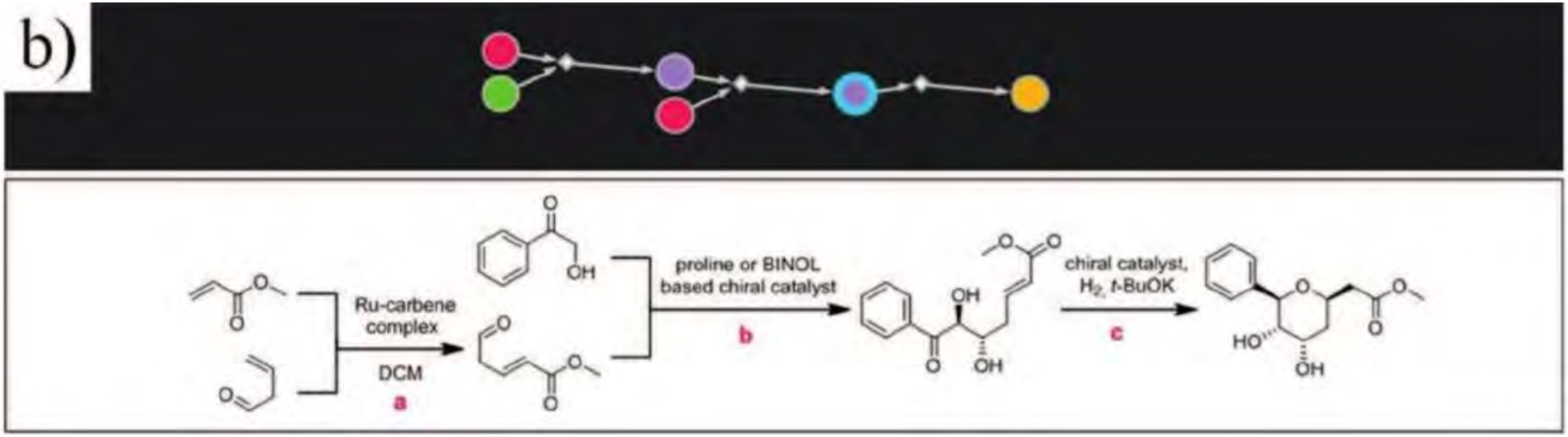


B

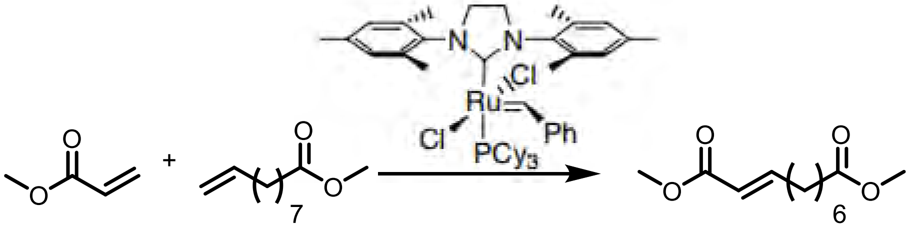


Kaplan, J. Med. Chem. 1982, 25, 1292 – 1299, Schmalz, Org. Lett. 2001, 3, 3579 – 3582, . Suh, Bioorg. Med. Chem. Lett. 2012, 22, 6750 – 6755.

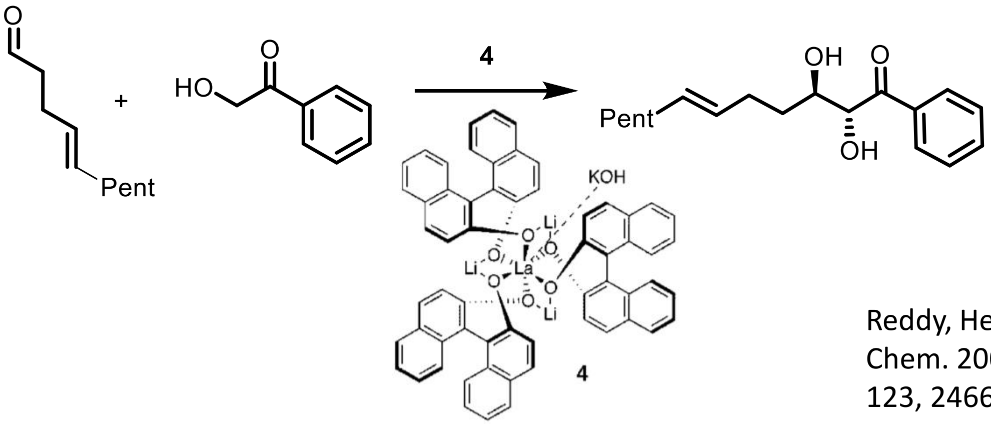
Group Problem 2


b)
Ru-carbene complex
DCM
a
OH
proline or BINOL based chiral catalyst
b
OH
OH
chiral catalyst, H_2, t-BuOK
c
O
HO
OH

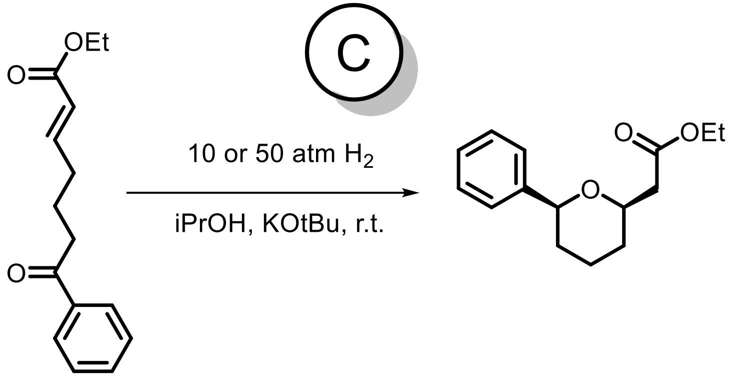
A



B

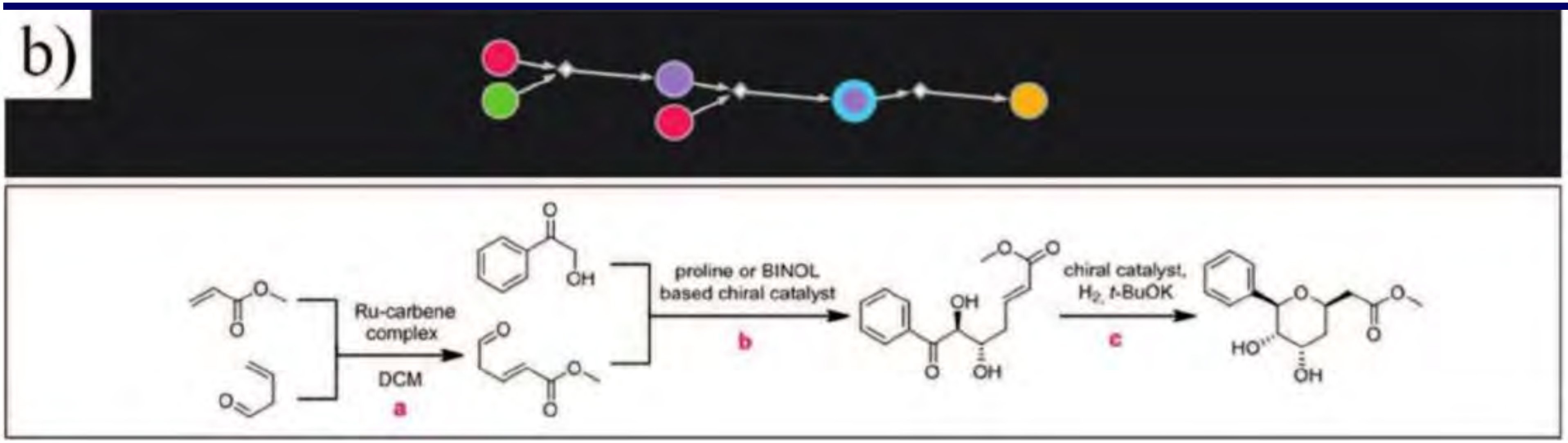


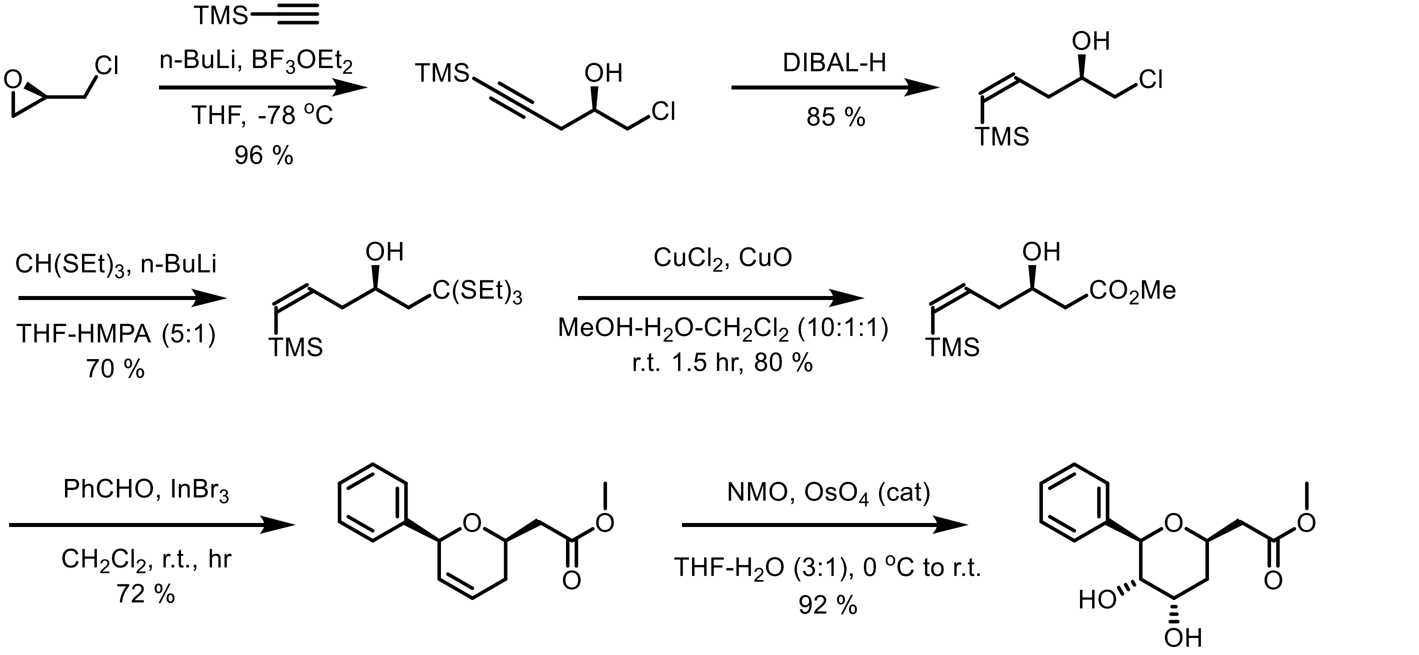
C



Reddy, Helv. Chim. Acta 2010, 93, 1362 – 1368, Tooze, J. Organomet. Chem. 2005, 690, 5863 – 5866, . Shibasaki, J. Am. Chem. Soc. 2001, 123, 2466 – 2467, Zhou, Org. Lett. 2012, 14, 4758 – 4761

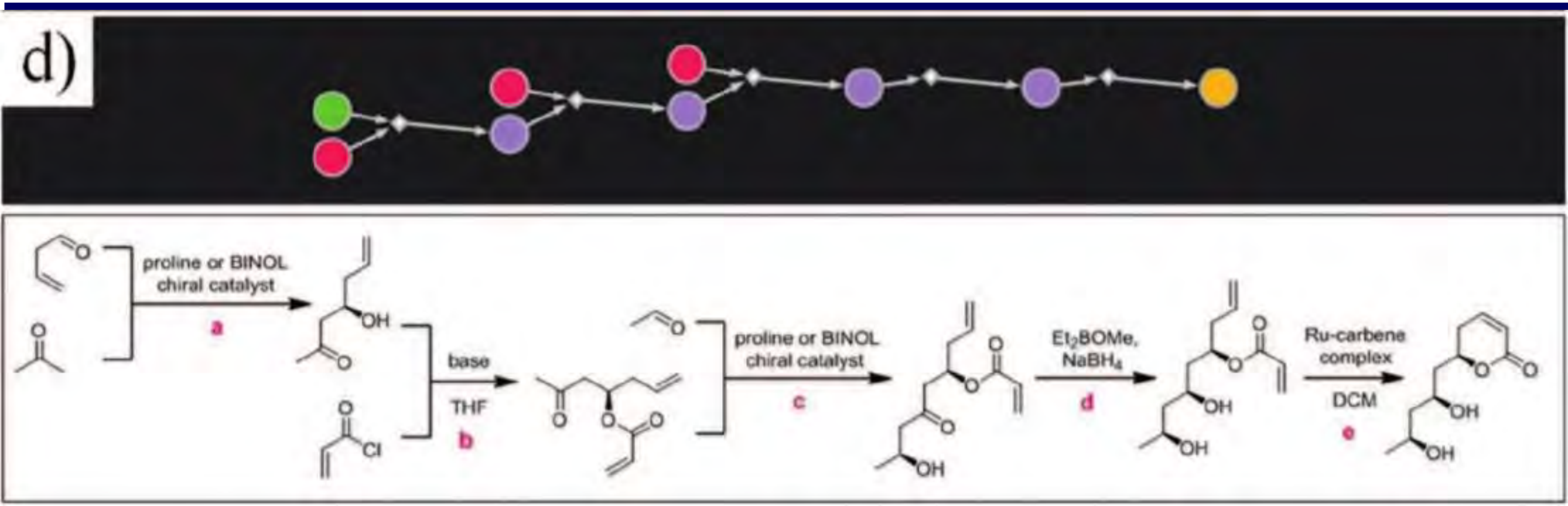
Group Problem 2

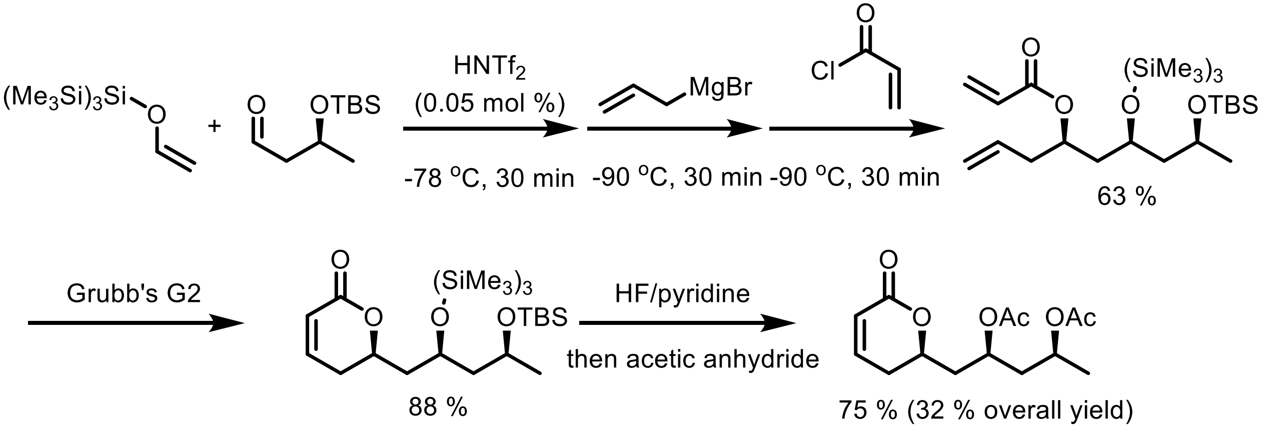

b)
Ru-carbene complex
DCM
a
proline or BINOL based chiral catalyst
b
chiral catalyst, H₂, t-BuOK
c
OH
HO
O



She, Synlett 2010, 15, 2283 – 2284

Group Problem 2


d)
proline or BINOL chiral catalyst
a
OH
O
base
THF
b
Cl
proline or BINOL chiral catalyst
c
Et_2BOMe, $NaBH_4$
d
Ru-carbene complex
DCM
e
OH

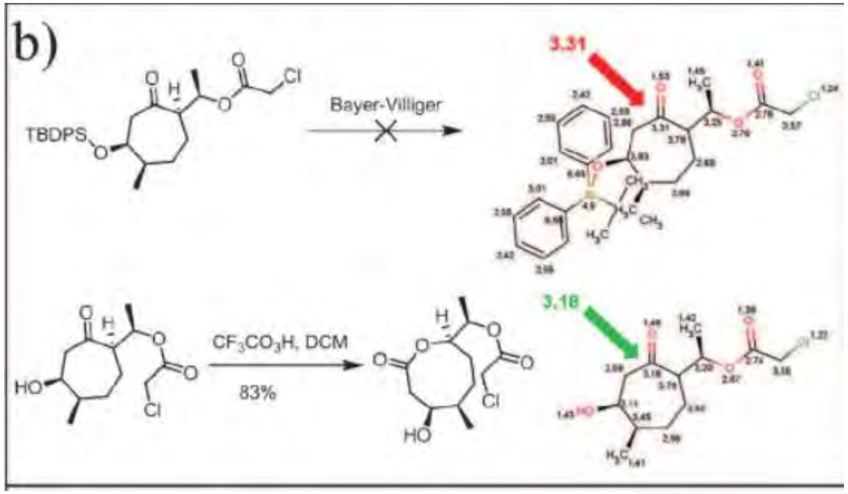


Yamamoto, JACS. 2007, 129, 2762-2763

Points of Improvement

More efficient searches: intrinsic (molecular topology) vs extrinsic (number of stereocenters created) metrics, synthetic accessibility, outcomes dictated by non-local contributions, the combinatorial explosion

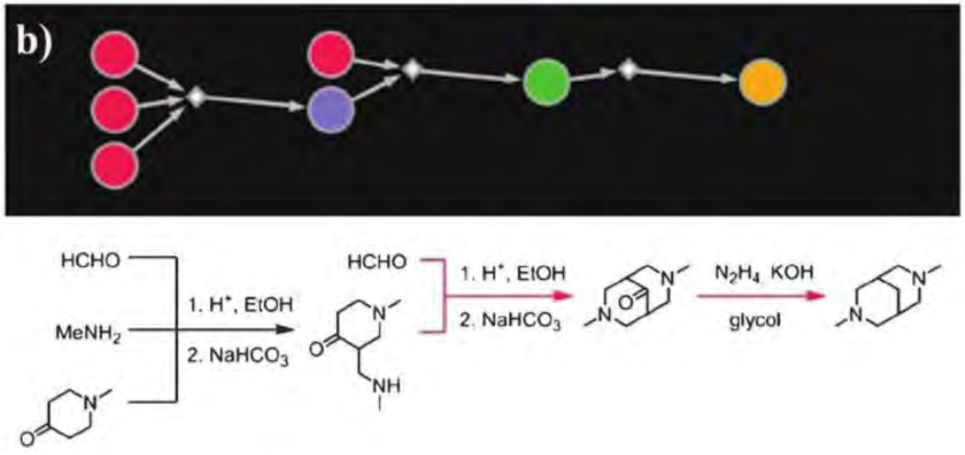
Reaction rules: Back to machine-learned rules?

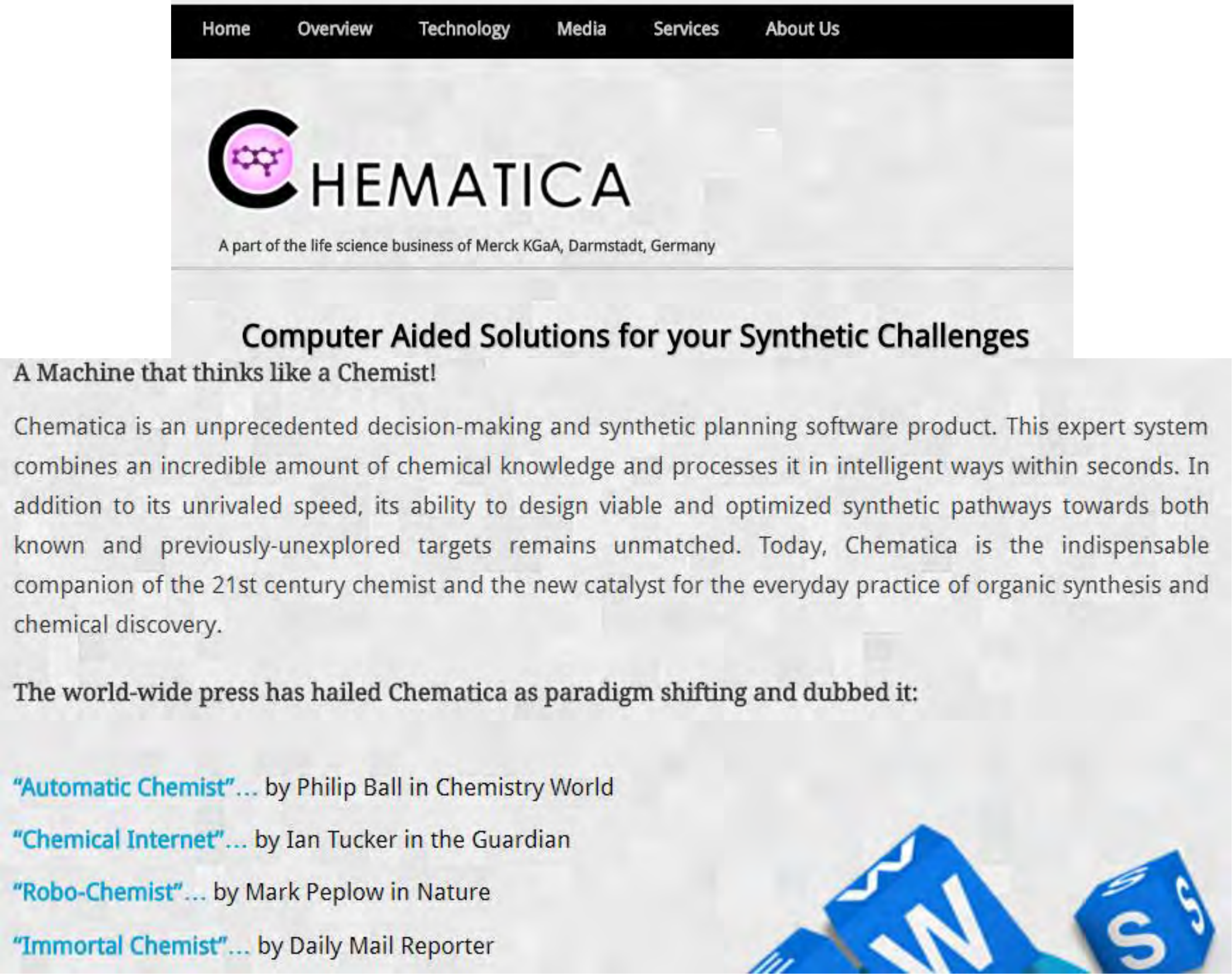


Cao and Liu's Topological Steric Effect Index

Conformer Distributions?

Quantum Chemical Calculations?





A part of the life science business of Merck KGaA, Darmstadt, Germany

Computer Aided Solutions for your Synthetic Challenges

A Machine that thinks like a Chemist!

Chematica is an unprecedented decision-making and synthetic planning software product. This expert system combines an incredible amount of chemical knowledge and processes it in intelligent ways within seconds. In addition to its unrivaled speed, its ability to design viable and optimized synthetic pathways towards both known and previously-unexplored targets remains unmatched. Today, Chematica is the indispensable companion of the 21st century chemist and the new catalyst for the everyday practice of organic synthesis and chemical discovery.

The world-wide press has hailed Chematica as paradigm shifting and dubbed it:

"Automatic Chemist"... by Philip Ball in Chemistry World

"Chemical Internet"... by Ian Tucker in the Guardian

"Robo-Chemist"... by Mark Peplow in Nature

"Immortal Chemist"... by Daily Mail Reporter