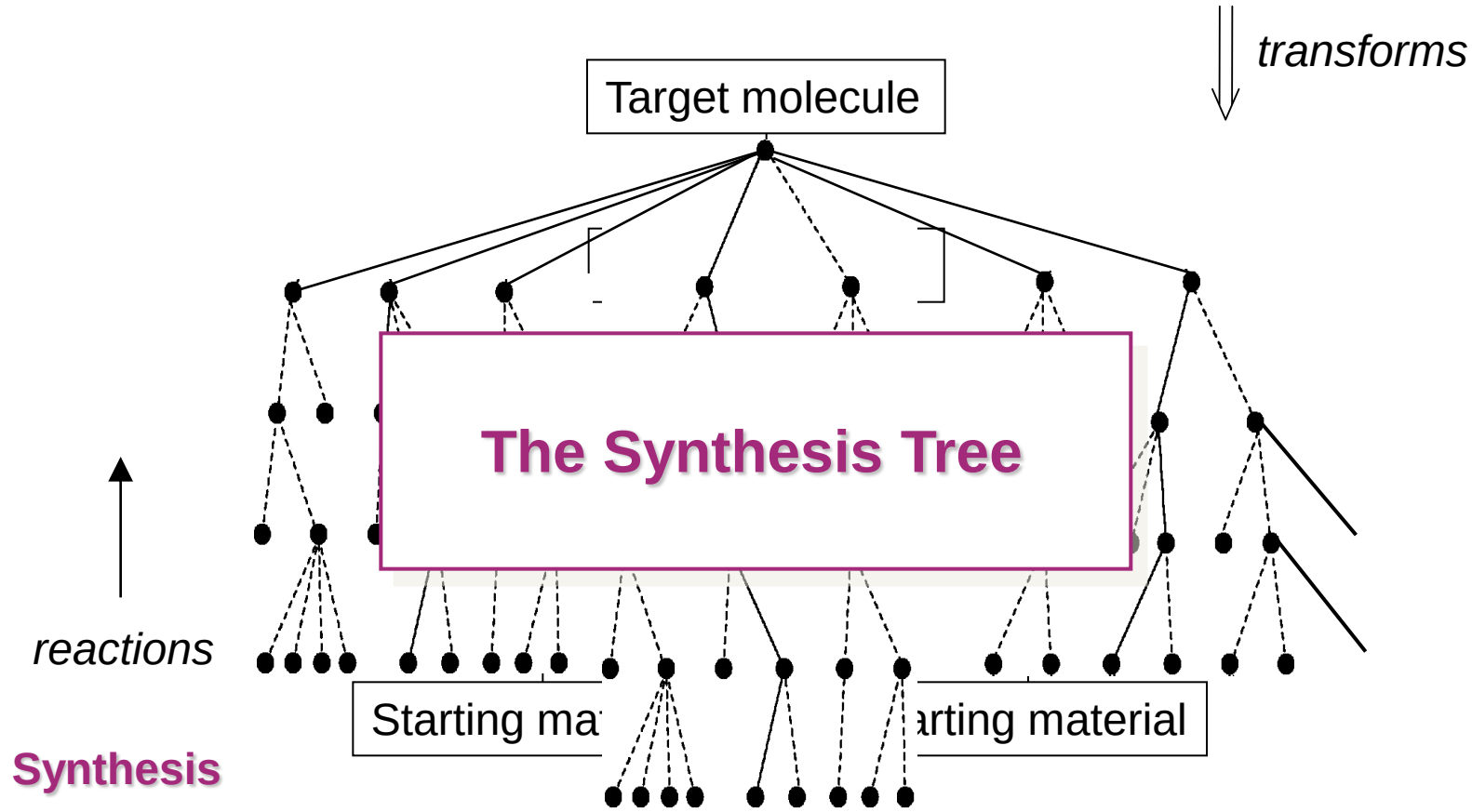
A View of Recent Computer Aided Synthesis Developments

- The past …
- … more recent attitudes
- Three modern CASD systems

Mike Hutchings
Heinz Saller
Hans Kraut
Peter Löw
InfoChem GmbH, Munich

E.J.Corey 1960s The Formalisation of Retrosynthesis



Retrosynthesis computerised

(Corey, then others, late 1960s ⇨)

CASD introduced widely into industry

Within industry CASD went into hibernation from the late 80s

- Heavy in-house maintenance & support
- Ease of use & results unsatisfactory
- Even those centrally involved developed other interests & priorities
- Management stopped support

- *The same applied across other European chemical companies*

- But chemists were NOT : "sceptical, even hostile" (*Ihlenfeld & Gasteiger*)

ALSO :
- *Mid-80s computer searchable reaction d/b started to appear*
- *More intuitive; easier to use*
- *Addressed some synthesis questions more readily*

On-line services became tools of choice

But in the last 5 years –

- A new generation of chemists
- *For them* - available tools inadequate
- Industrial chemists asked potential suppliers to re-visit CASD

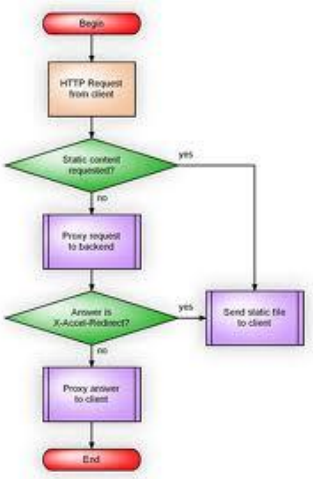
New Motivations

Extensive reaction databases

e.g. SPRESI (InfoChem): 4 M reactions

Algorithms / software

- Atom-atom-mapping
- Reaction centre i/d
- Reaction classification
- Fast search engine

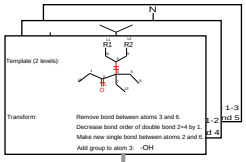


Computer hardware

- Powerful
- Inexpensive
- Networked
- Secure

Transforms From Reaction Databases

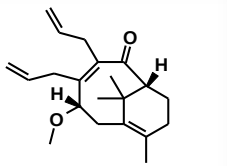
1. Pre-processing



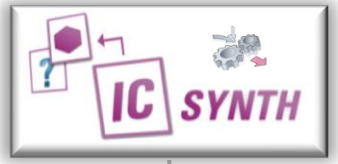
Automatic transform extraction
(IC*MAP*/CLASSIFY)

Transform library

2. Retrosynthesis planning



Target molecule



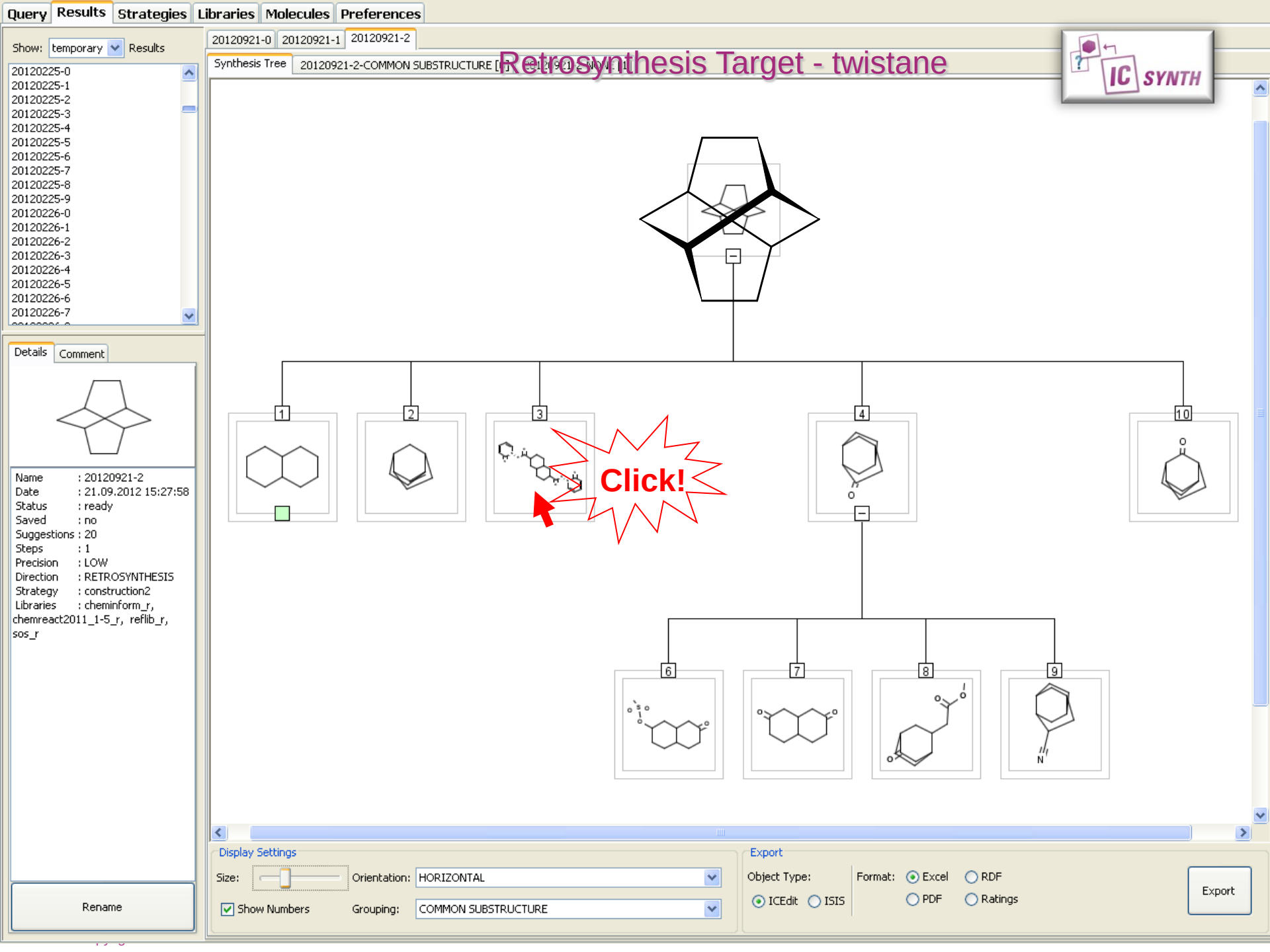
Reaction database
e.g.
SPRESI
proprietary databases
commercial databases

Lookup
stored
example
reactions

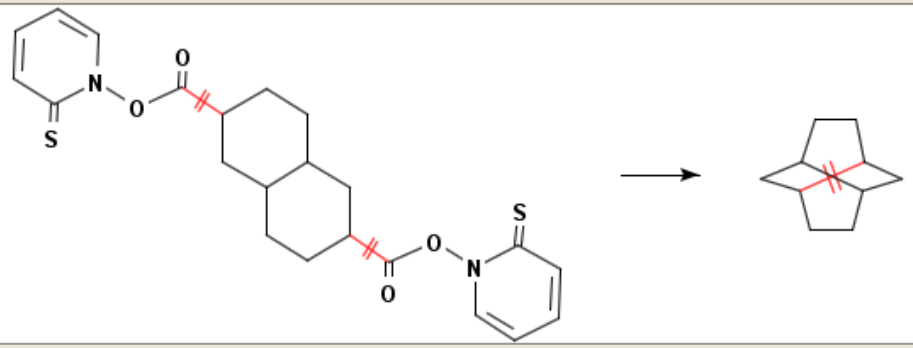
Strategies for evaluation & prioritisation – chemical intelligence

Transform x → Precursor(s) 1
Transform y → Precursor(s) 2
...
Transform z → Precursor(s) n

Retrosynthesis Target - twistane

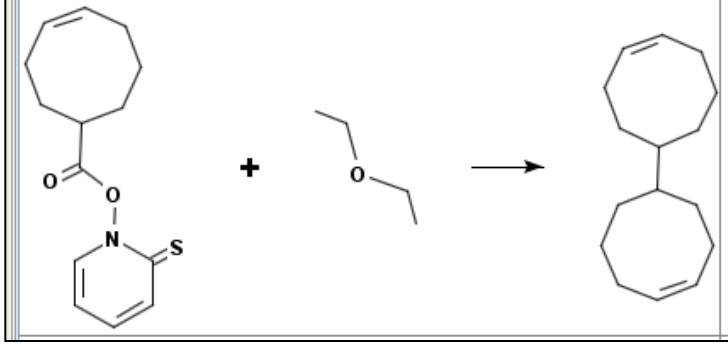


Model Reaction:



Rating: **141500**

References:

Structure	Similarity	Source	Citation	Yield	Conditions	Name	Report Data ...
	69	SPRESI	J. CHEM. SOC. CHEM. COMMUN.,(1987) N 10, 786-788	27	diethyl ether, irradiation	--	Report Error

Click!

SPRESIweb 2.11

SEARCH ...
- Quick Search
- Molecules
- Reactions
- References

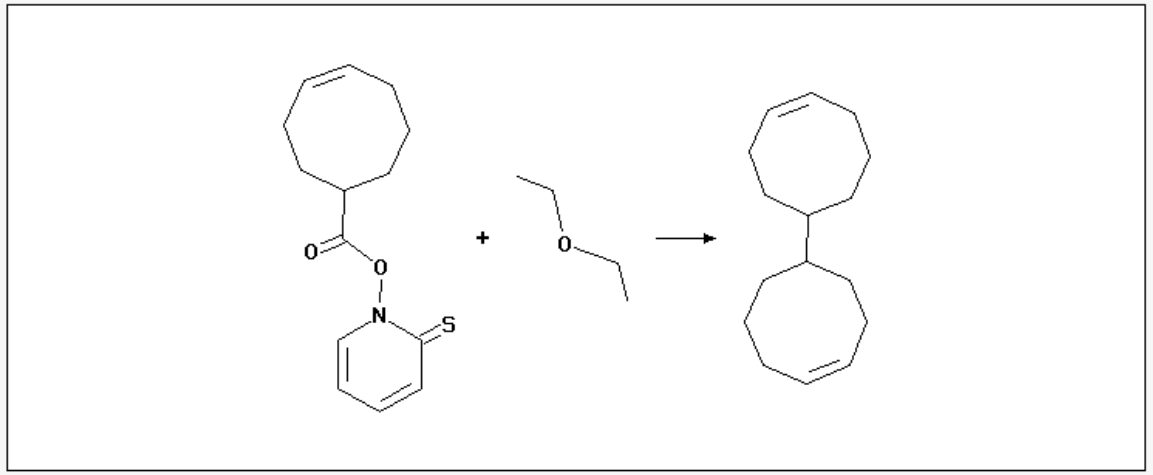
QUERY ...
- New
- Save
- Open
- Delete

HIT LIST ...
- Save
- Open
- Delete
- Import
- Export
- Download statistic

HELP ...
- Help
- Get Plugins
- About...
- Your opinion...

OTHERS ...
- Logout
- Home
- InfoChem
- Versions

2 Reaction Regno: 1333298 »



DETAILS SIMILAR ZOOM ERROR

Reaction Data

Yield: 27%

Conditions: diethyl ether, irradiation

Reference

BLOODWORTH, A. J.;CRICH, DAVID;MELVIN, I

»THE CIS-CYCLO-OCT-4-ENYL RADICAL IS ...NGE

J. CHEM. SOC. CHEM. COMMUN.,(1987) N 10

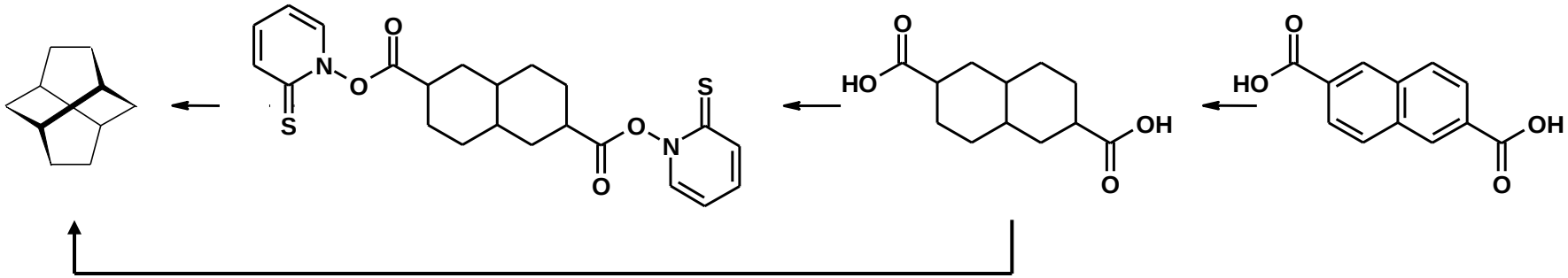
View abstract/full paper on publisher's page

Click!

Publisher websites ⇨ literature

The best literature route to twistane is **4 steps**

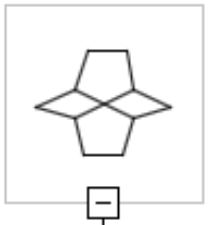
IC*SYNTH* idea : twistane from a bulk monomer – **3 steps**

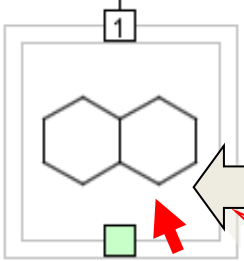


Further consideration of this idea suggests **2 steps** via Kolbe electrolysis

Telescope into a **1-pot** process ???? !!!

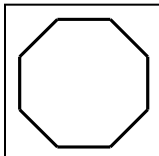
NOISE *versus* NOVELTY



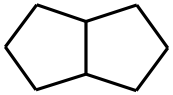


The highest priority suggestion
Is this reasonable?
NOISE or NOVELTY?

Click!



5% Pt on Al_2O_3 / Ti(0)



***JACS* 1986, 1716**

IC SYNTH

- Released commercially
- AZ getting valuable o/p - “Paid for itself”
- Process chemistry brainstorms
- Real, and surprising, applications

- A new type of *forward* search engine

- Continuing development - fundamental re-design
  - complexity driven strategies
  - new type of transform
  - improved user interface & display …

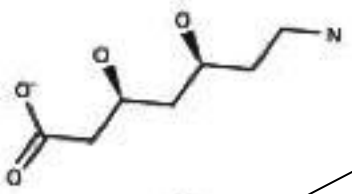
Both designed for retrosynthetic analysis
Both based on transforms (rules) derived from literature chemistry

Differences

- Approaches to “strategies”
- Output display formats

- Transforms algorithmic; shell-based ± chemistry
- Transforms *optionally* derived from as few as 1 lit example: unexpected solutions
- No stereochemistry yet

- Rules capture key chemistry
- Rules based on ≥10 lit examples, tried-and-tested chemistry only
- Enantioselective chemistry

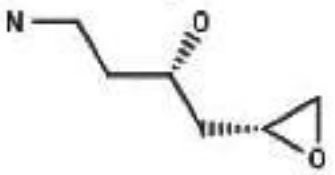


Priority of this suggestion

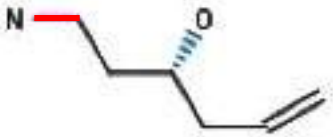
1180 examples, 90% | 25 of 40

Number of lit examples of this rxn

Number of suggested routes to this target



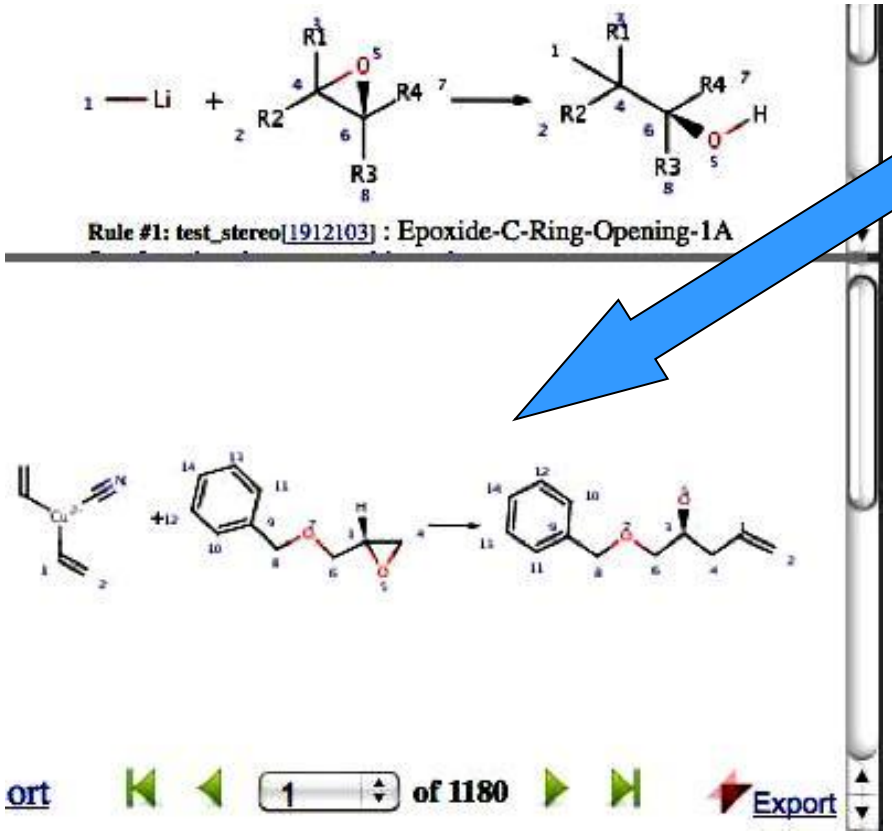
3108 examples, 77% | 4 of 24

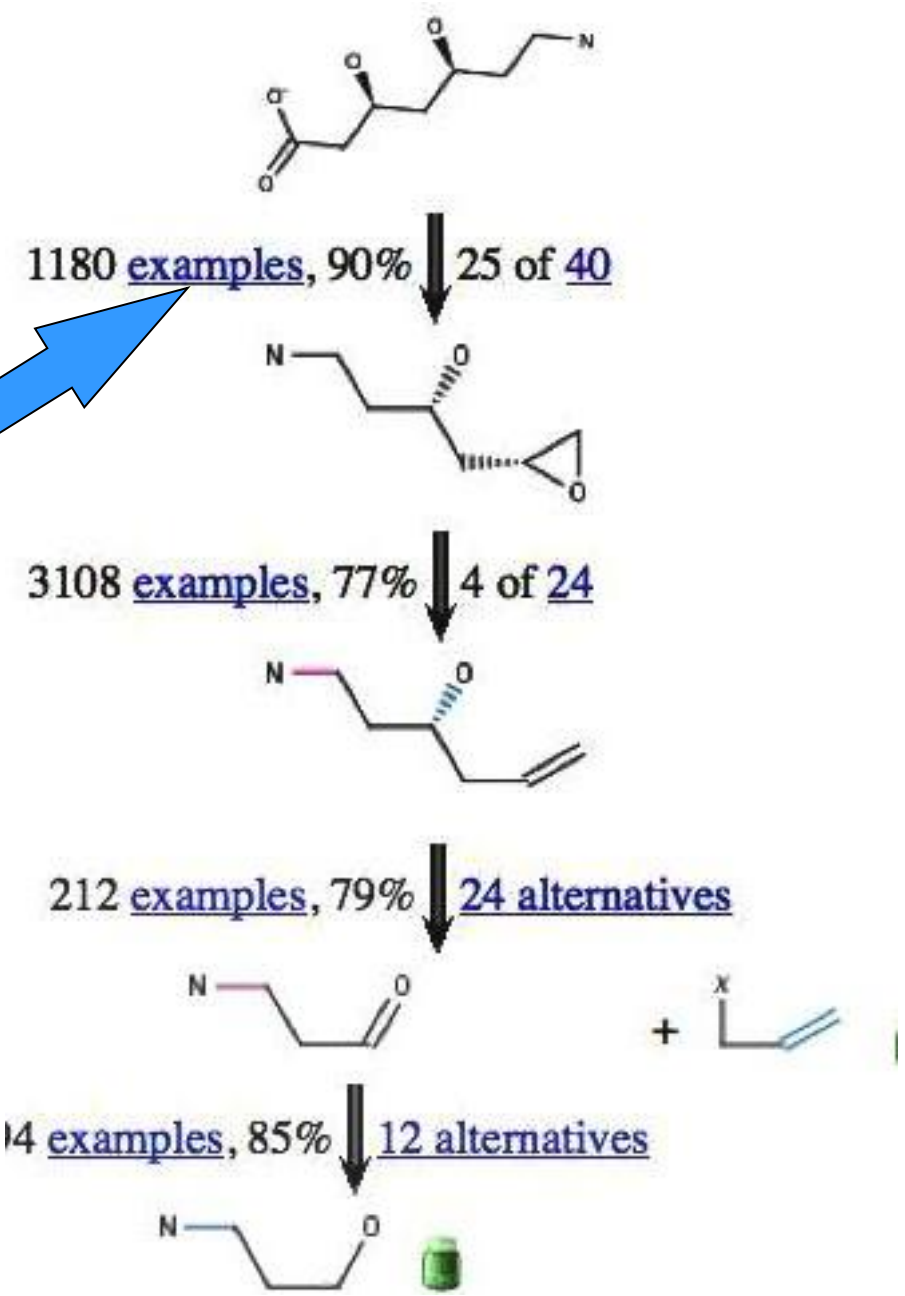


Commercially available

212 examples, 79% | 24 alternatives

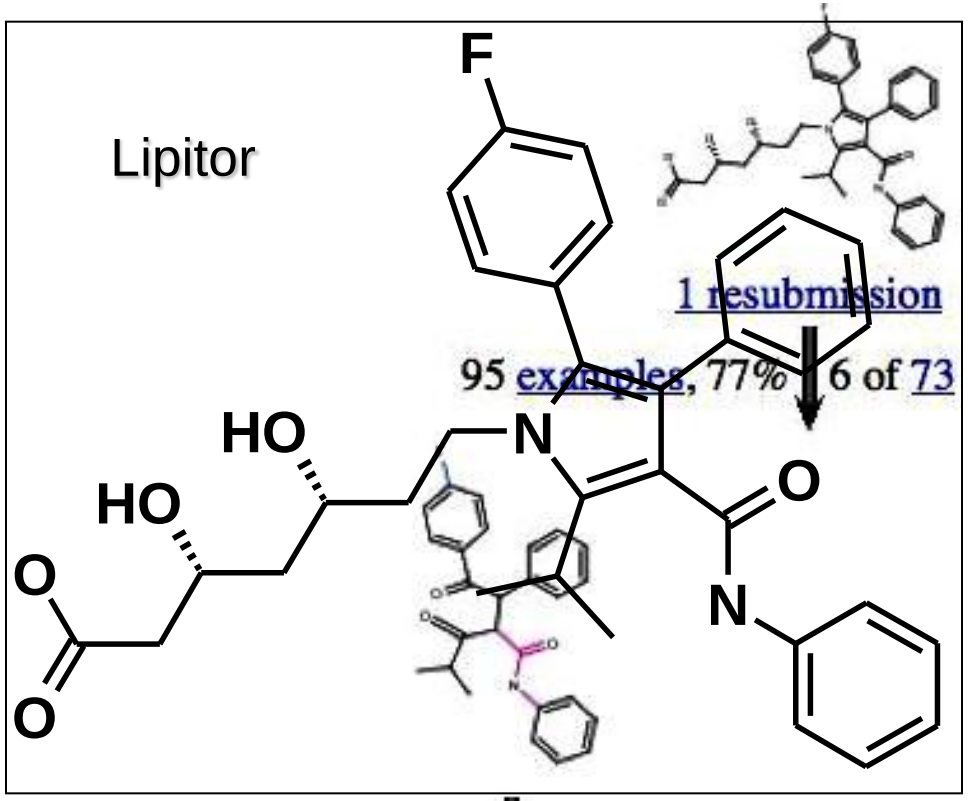
1194 examples, 85% | 12 alternatives


Rule #1: test_stereo[1912103] : Epoxide-C-Ring-Opening-1A
ort
1
of 1180
Export


1180 examples, 90%
25 of 40
3108 examples, 77%
4 of 24
212 examples, 79%
24 alternatives
4 examples, 85%
12 alternatives

ARChem Route Designer

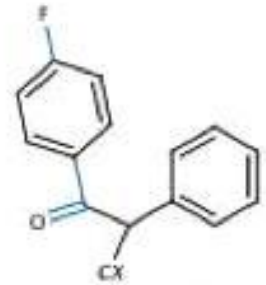
Lipitor



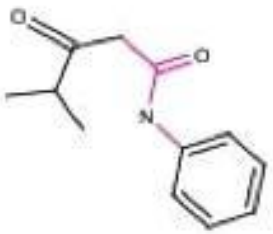
1 resubmission

95 examples, 77% 6 of 73

501 examples, 65% 58 alternatives

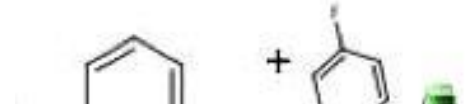


+

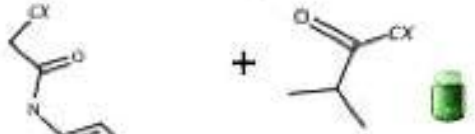


umples, 86% 10 alternatives

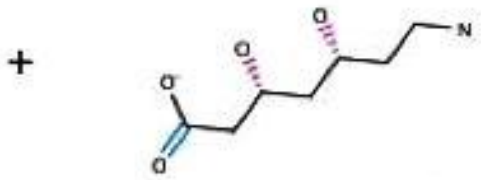
130 examples, 76% 26 alternatives



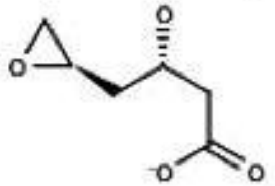
+



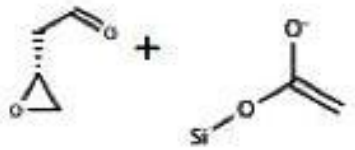
+



1180 examples, 90% 22 of 40



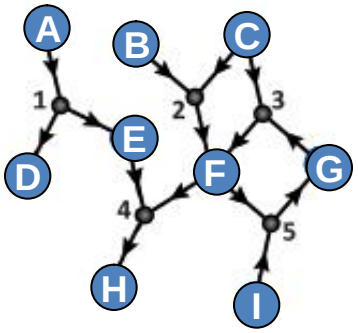
273 examples, 90% 3 of 3



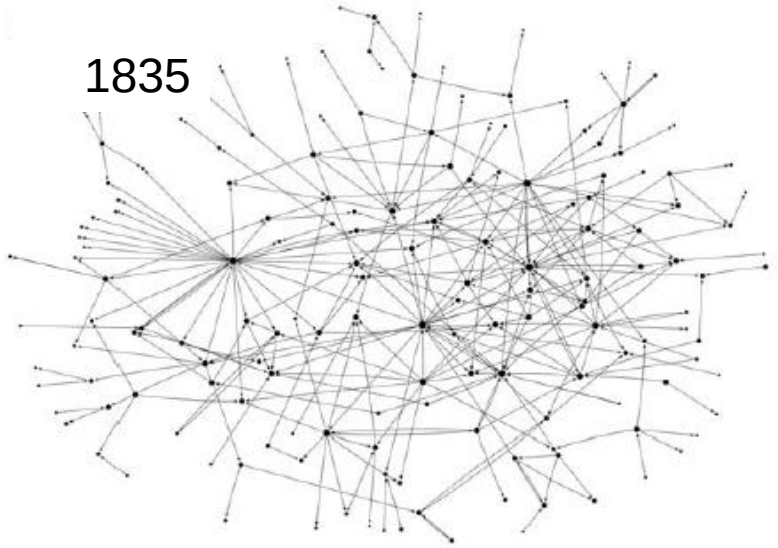
A new approach : **Network of Organic Chemistry**

Reactions

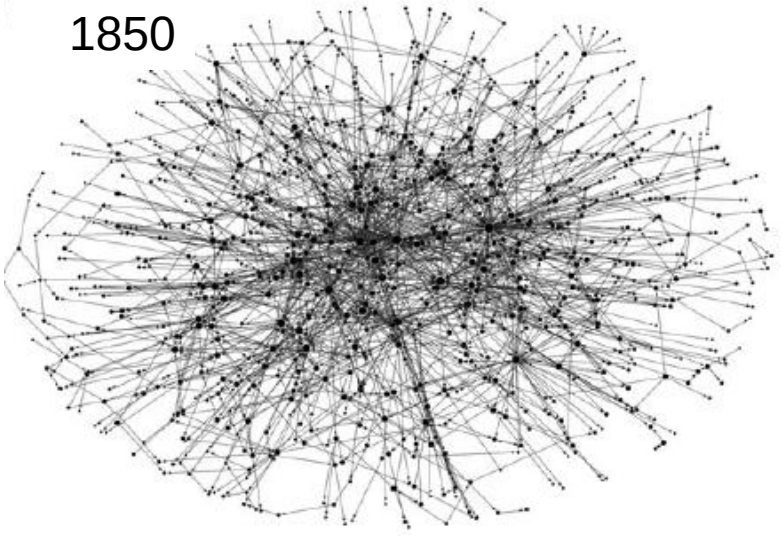
1) A → D + E
2) B + C → F
3) C + G → F
4) E + F → H
5) F + I → G



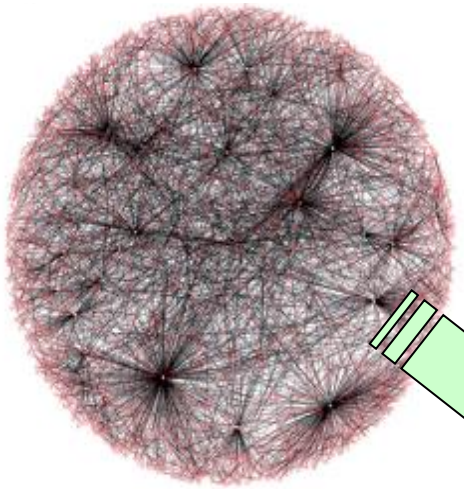
1835



1850



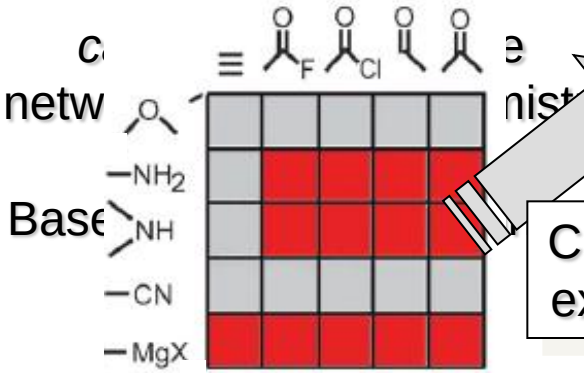
The Network of Organic Chemistry



NOC

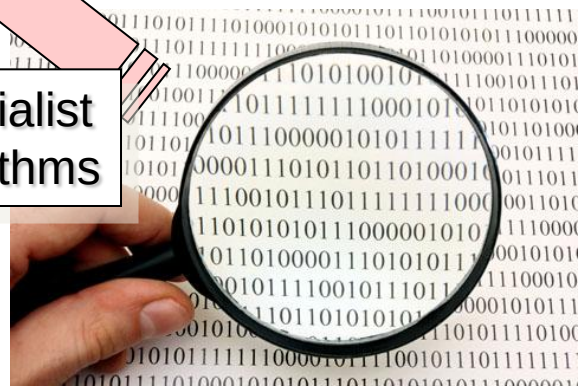
Modern computers

New Aspects of Synthesis Planning



Chemical expertise

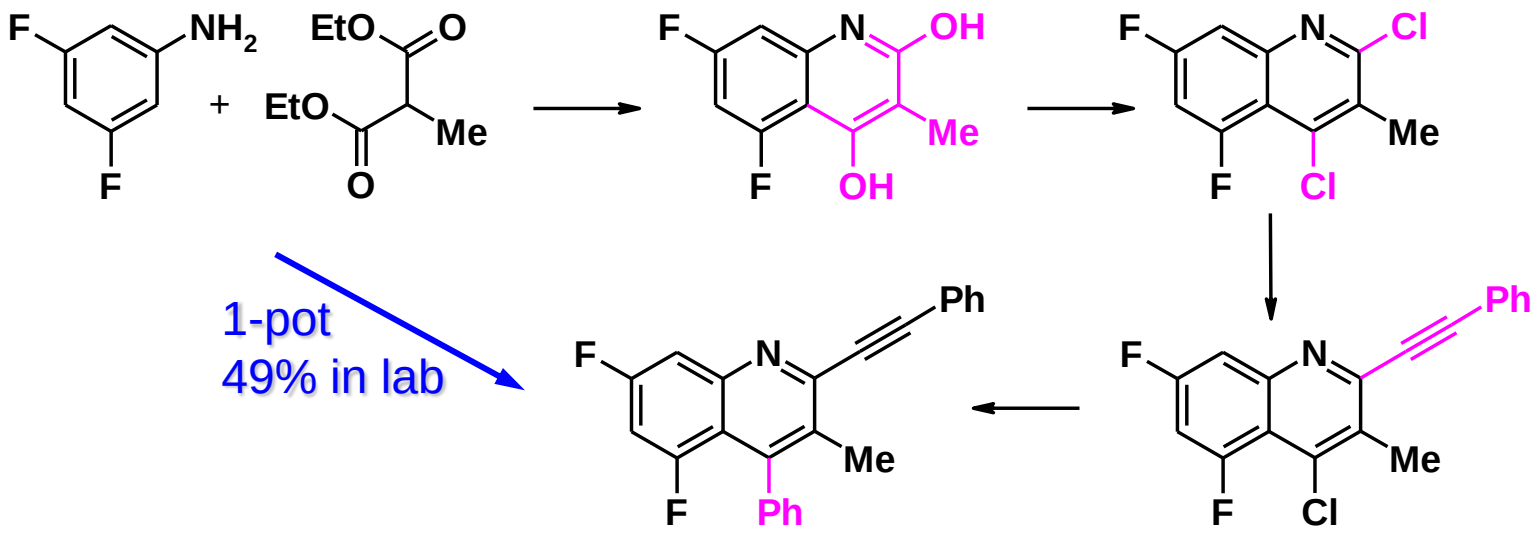
Specialist algorithms



Rewiring Chemistry

Algorithmic Discovery of One-pot Reactions in the NOC

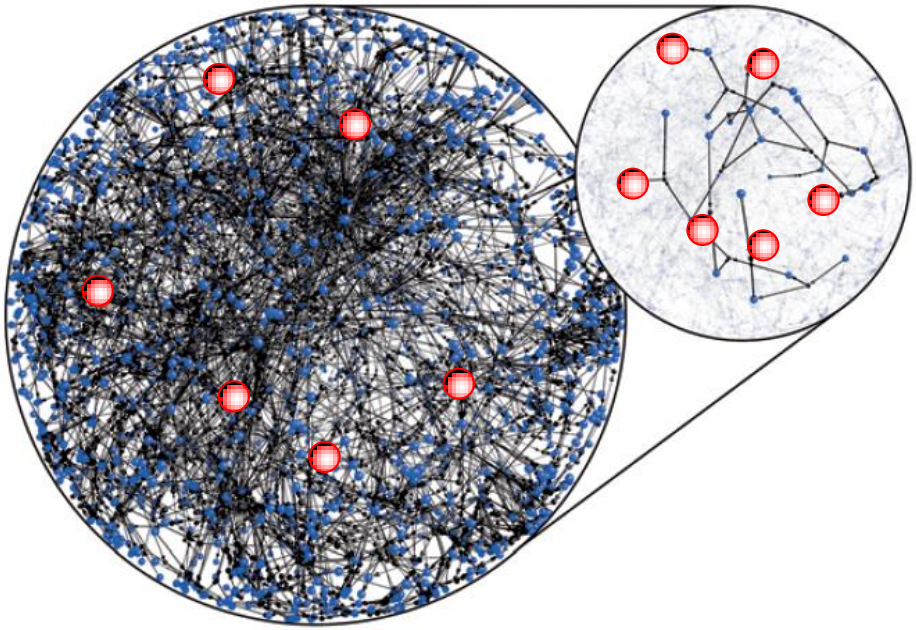
Angew. Chem. Int. Ed. 2012, *51, 7922-27*



"…so far generated >1,000,000 structurally diverse one-pot sequences…"

Parallel Optimisation of Synthetic Pathways
within the Network of Organic Chemistry

Angew. Chem. Int. Ed. **2012**, ***51, 7928-32***



Given:
- Several related products
- Some common starting materials
- Various chemistry

The problem:
- Optimise profitability
- by selecting SM
- & chemistry used

The solution:
- NOC + new algorithms

Cost reduction of 45%
in one case of 51 related products

The Grzybowski Group

SELF ASSEMBLY AND ADAPTIVE SYSTEMS

NORTHWESTERN UNIVERSITY

Chematica

" A software product for NOC searches…

…will be commercially available later in 2012 "

Acknowledgments

InfoChem (*Munich*)
- Fanny Irlinger
- Christoph Oppawsky

SimBioSys (*Toronto*)
- Orr Ravitz
- Peter Johnson (*Leeds University*)

AstraZeneca (*Södertälje*)
- Adrian Clark

ChemNotia (*Stockholm*)
- Anders Bogevig
- Fernando Huerta
- Tobias Rein

Thank you

Click!

Mike Hutchings
Heinz Saller
Hans Kraut
Peter Löw

InfoChem GmbH, Munich

InfoChem's expertise : Customers & Business Partners

BASF The Chemical Company
Pfizer
LANXESS Energizing Chemistry
CAS A division of the American Chemical Society
Leibniz Gemeinschaft
Umwelt Bundes Amt Für Mensch und Umwelt
Bayer
AstraZeneca life inspiring ideas
Lonzagroup
IPB
SCAI
Institut Algorithmen und Wissenschaftliches Rechnen
MERZ
Mercedes-Benz
ASC
Boehringer Ingelheim
DyStar
THOMSON
Symyx
NOVARTIS
SIGMA-ALDRICH
MERCK
Waters
VOGEL VERLAG HEINRICH VOGEL
ELSEVIER
Thieme
Springer the language of science
WILEY Publishers Since 1807
VWR INTERNATIONAL
ACROS ORGANICS

AstraZeneca requested tools
beyond SciFinder, Reaxys, SPRESI[web] …
&
beyond professional chemist experience & expertise

Chemists need NEW IDEAS *e.g.* for synthetic pathways

BUT: augment chemists, not replace them

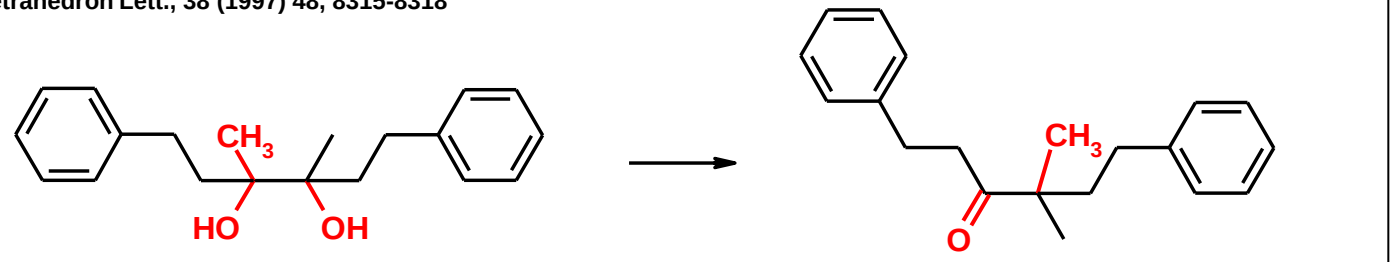
Led to InfoChem's New CASD Tool :

Reactions ↔ Transforms

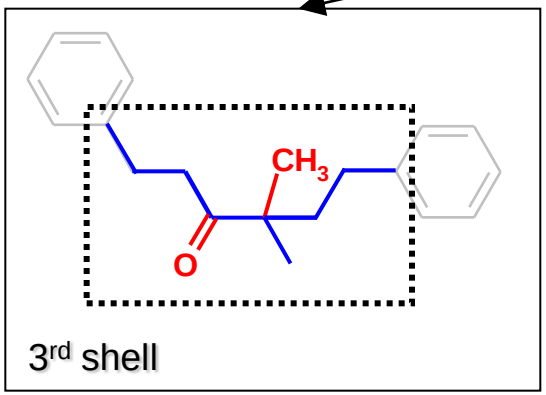
Reaction database

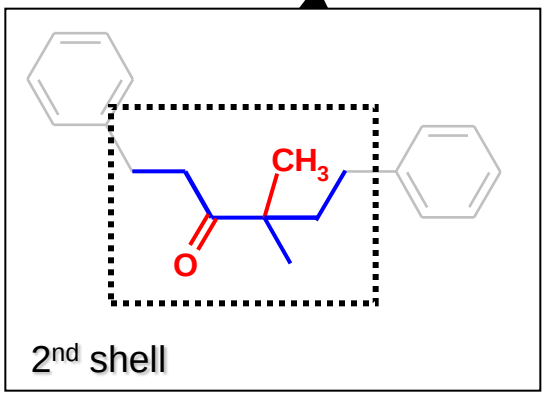
SPRESI Reaction Reg-No.: 3179546

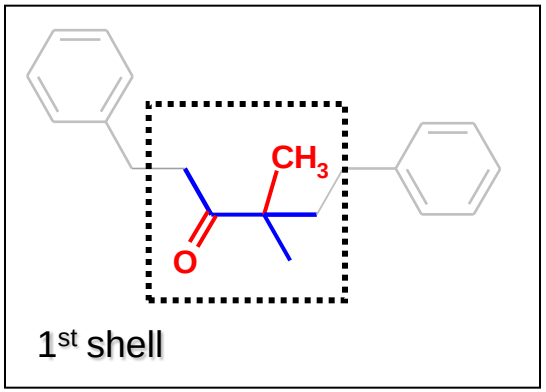
Kita Yasuyuki, Yoshida Yutaka, Mihara Sachiko, Fang Dai-Fei, Higuchi Kazuhiro, Furukawa Akihiro, Fujioka Hiromichi
Efficient pinacol rearrangement mediated by trimethyl orthoformate
Tetrahedron Lett., 38 (1997) 48, 8315-8318



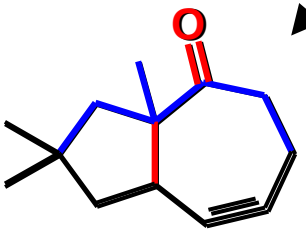
Template





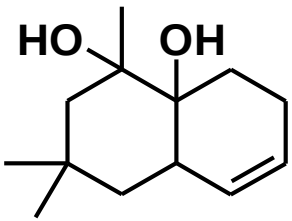


Transform Application

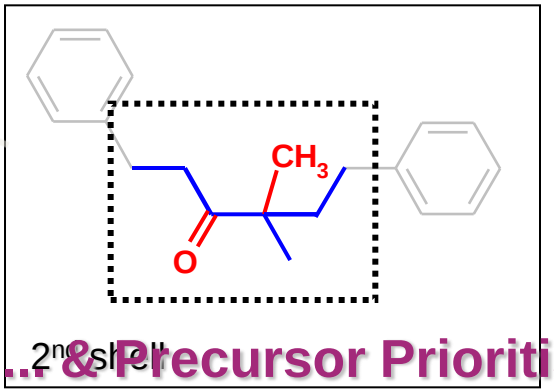


Target

Transform



Precursor



2nd shell ...

... & Precursor Prioritisation

Depends on strategy parameters

- bond type (C-C, C-X ...)
- ring (type)
- branching site
- bond symmetry
- number of new bonds
- number of precedent rxns
- best lit yield
- exact literature product?
- commercial availability
- number & size of precursor molecules
- complexity reduction
- etc ...

Spectrum of Result Values
NOISE *versus* NOVELTY

Bad

D/b error – literature
– abstracting
– machine translation

Inappropriate – strain
– selectivity
– other over-extrapolation…

Known – literature

New – expected – "routine"
– unexpected – "idea"

Good

Appears as unwanted o/p if not removed - NOISE

But if incorrectly penalised may be missing potential NOVELTY