Chemical Reactions

Molecules

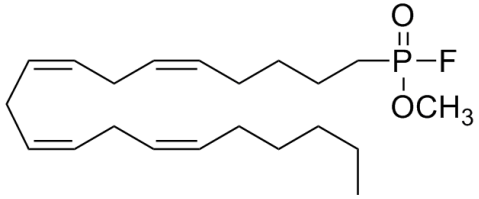
Chemical Energy

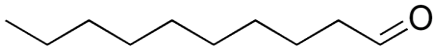
Reactivity

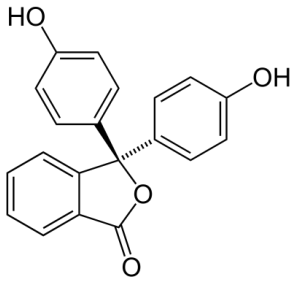
Synthesis Planning

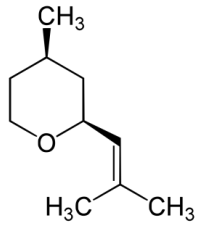
Big Data

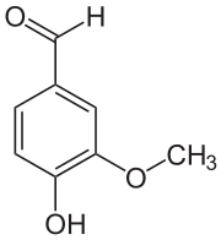
Molecules

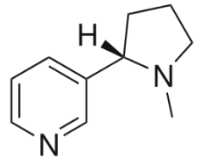




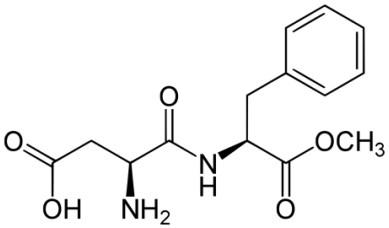


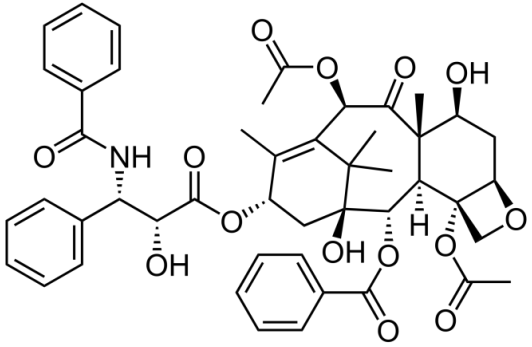


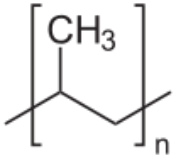


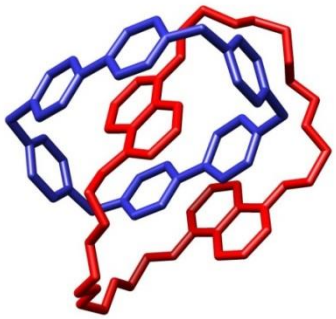


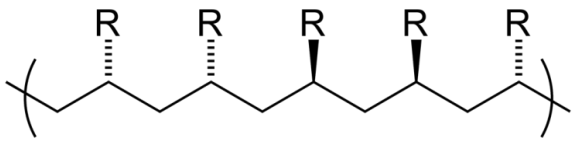
- Drugs (taxol)
- Natural products (nicotine)
- Fragrances (decanal, rose oxide, vanillin)
- Dyes (phenolphthalein)
- Probes (mafp)
- Peptides (aspartam)
- Antibody-drug conjugates
- DNA/RNA (DNA cage)
- Polymers (polypropylene)
- Others (catenane)





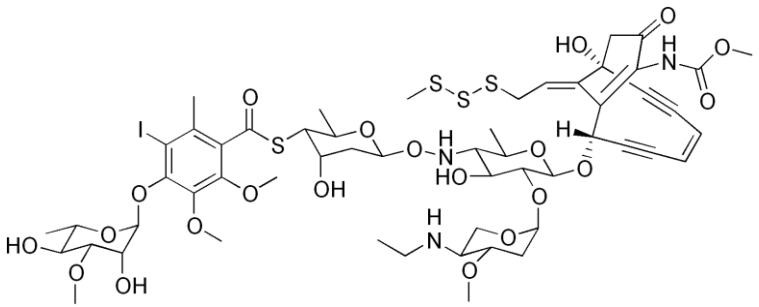




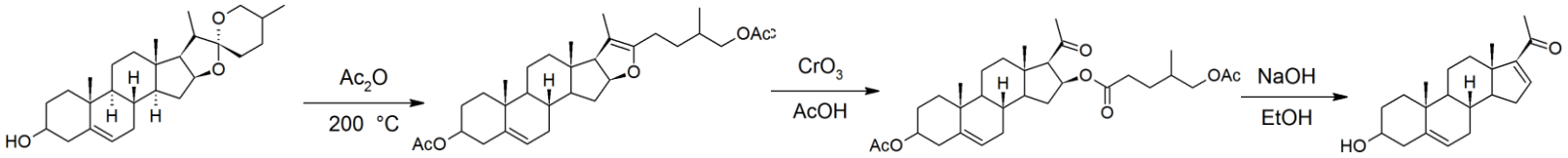


How to make molecules?

- Chemical synthesis (multistep)
- Fermentation (synthetic biology)
- Partial syntheses

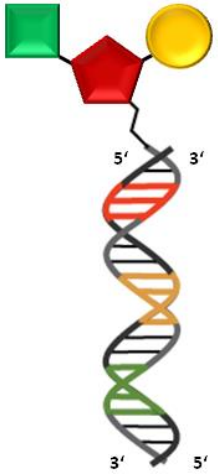


calicheamicin γ1



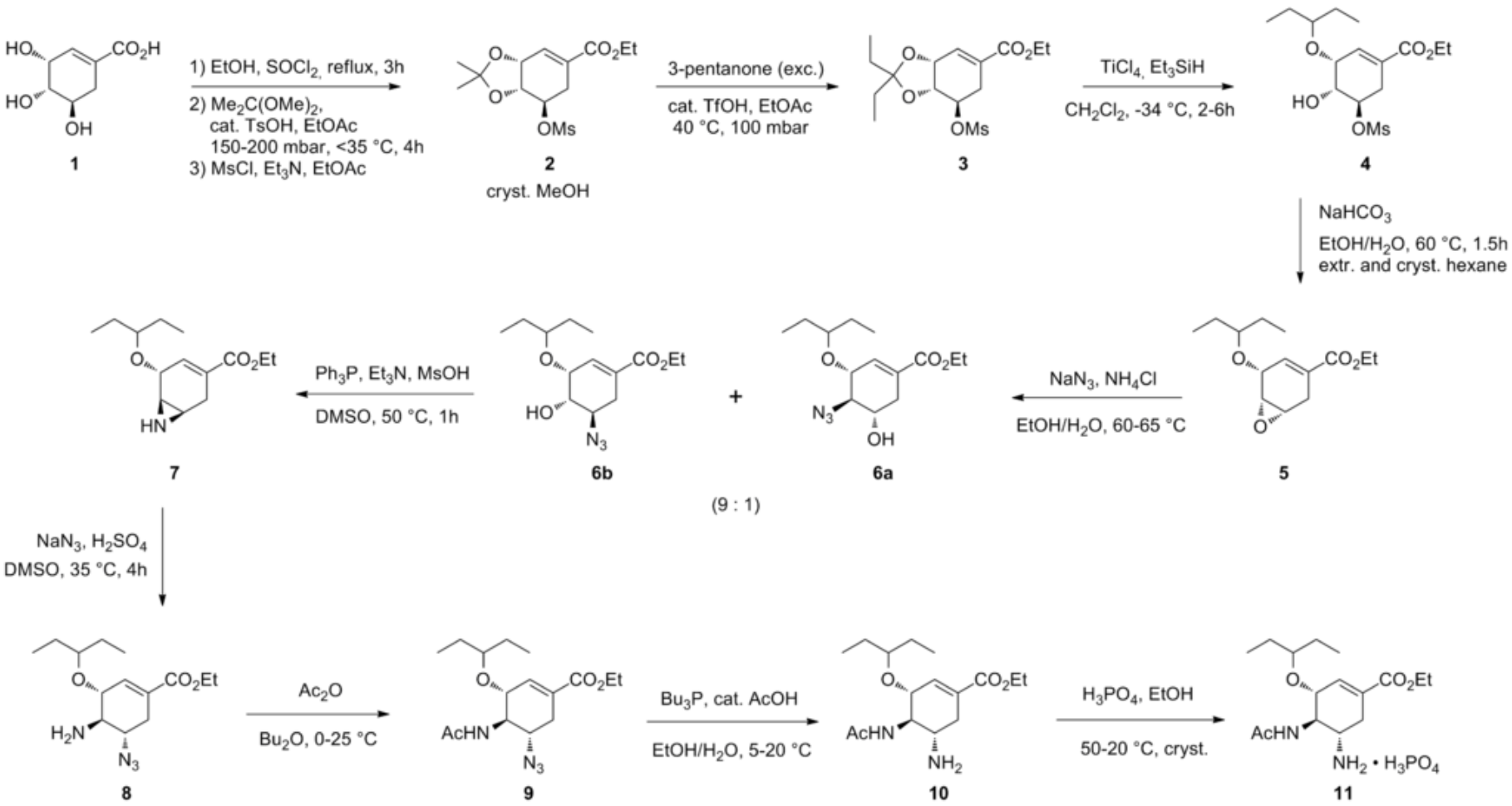
Marker degradation

- Solid-phase synthesis (peptides, DNA/RNA)

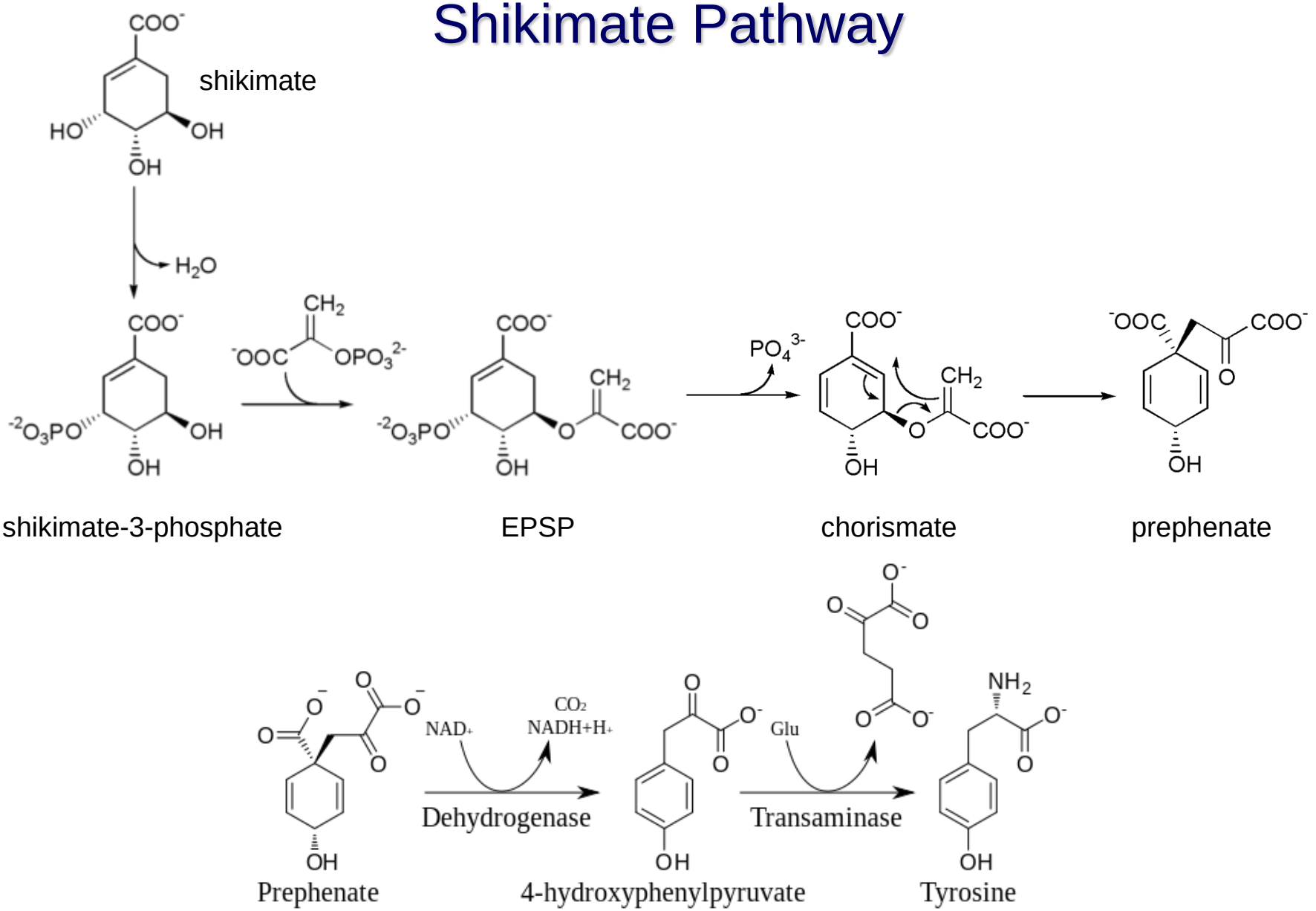


- Library synthesis
  - Combinatorial synthesis (beads, microarrays)
  - DNA-encoded chemistry
  - Parallel synthesis (automated)

Oseltamivir



Shikimate Pathway



Chemical Reactions

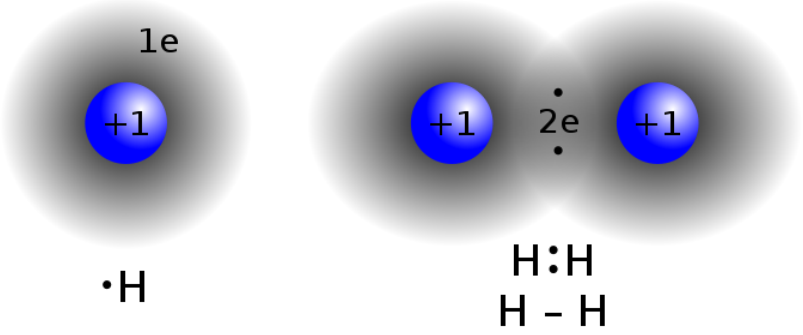
Molecules

Chemical Energy

Reactivity

Synthesis Planning

Big Data

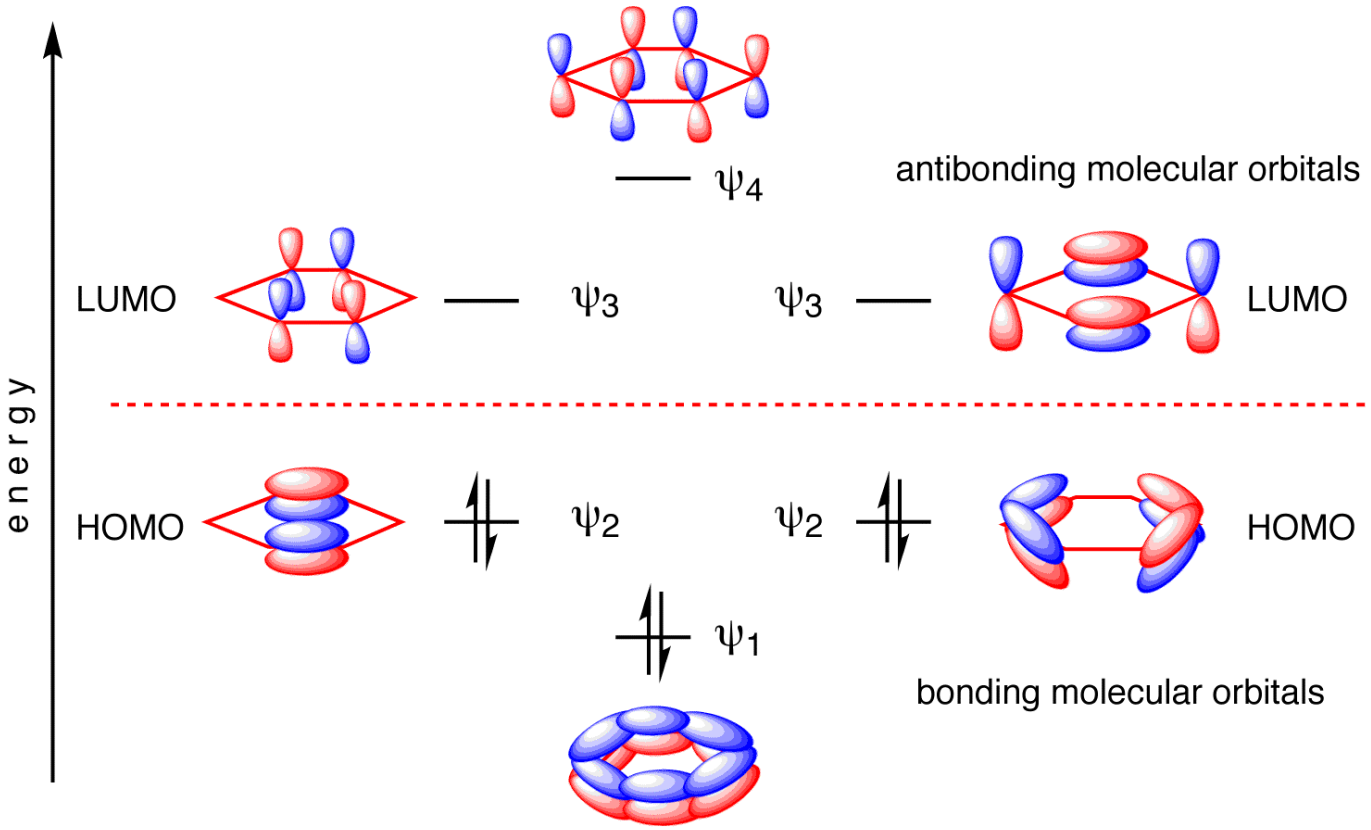


Bond Dissociation Energies

Average Bond Energies (kJ/mol)

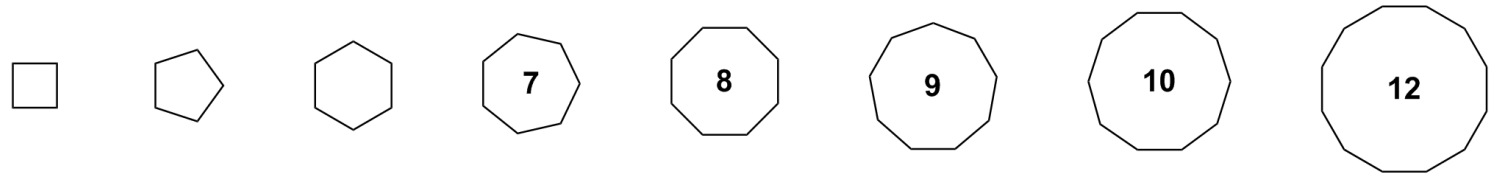
Bond	Energy	Bond	Energy	Bond	Energy	Bond	Energy
Single Bonds							
H—H	432	N—H	391	Si—H	323	S—H	347
H—F	565	N—N	160	Si—Si	226	S—S	266
H—Cl	427	N—P	209	Si—O	368	S—F	327
H—Br	363	N—O	201	Si—S	226	S—Cl	271
H—I	295	N—F	272	Si—F	565	S—Br	218
		N—Cl	200	Si—Cl	381	S—I	~170
C—H	413	N—Br	243	Si—Br	310		
C—C	347	N—I	159	Si—I	234	F—F	159
C—Si	301					F—Cl	193
C—N	305	O—H	467	P—H	320	F—Br	212
C—O	358	O—P	351	P—Si	213	F—I	263
C—P	264	O—O	204	P—P	200	Cl—Cl	243
C—S	259	O—S	265	P—F	490	Cl—Br	215
C—F	453	O—F	190	P—Cl	331	Cl—I	208
C—Cl	339	O—Cl	203	P—Br	272	Br—Br	193
C—Br	276	O—Br	234	P—I	184	Br—I	175
C—I	216	O—I	234			I—I	151
Multiple Bonds							
C=C	614	N=N	418	C≡C	839	N≡N	945
C=N	615	N=O	607	C≡N	891		
C=O	745 (799 in CO_2)	O_2	498	C≡O	1070		

Aromaticity

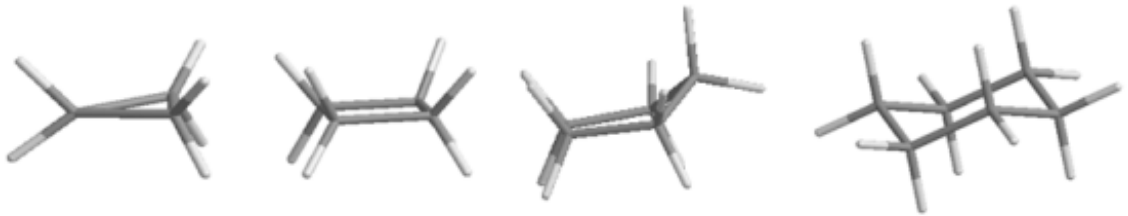


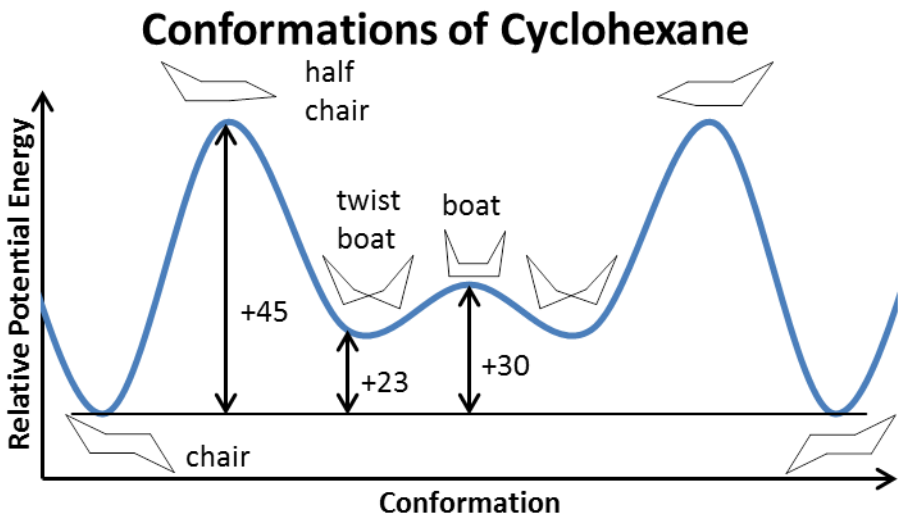
the π molecular orbitals for benzene. The dashed line represents the energy of an isolated p orbital – all orbitals below this line are bonding, all above it are antibonding. Benzene has six electrons in its π system so all the bonding MOs are fully occupied

Ring Strain and Conformation



Strain energy in kcal/mol	26.3	6.2	0.1	6.2	9.7	12.6	12.4	4.1





Chemical Reactions

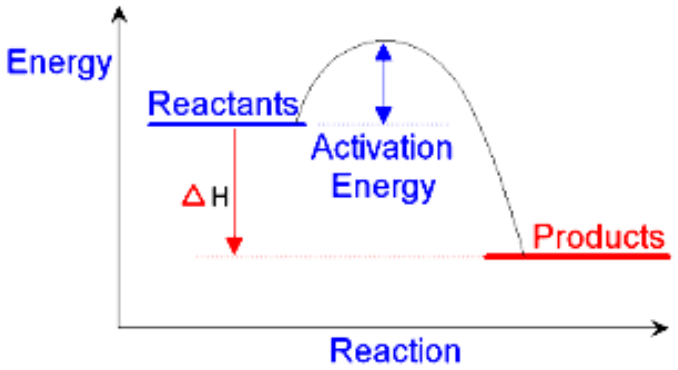
Molecules

Chemical Energy

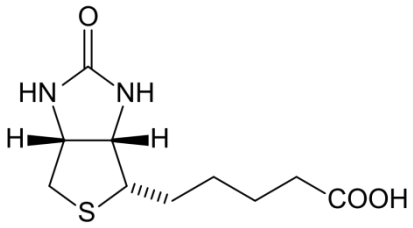
Reactivity

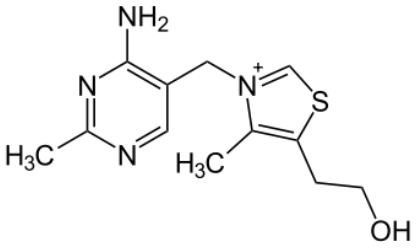
Synthesis Planning

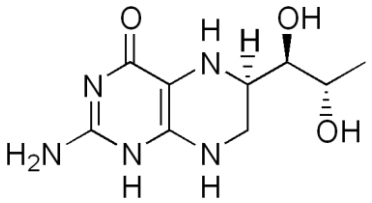
Big Data



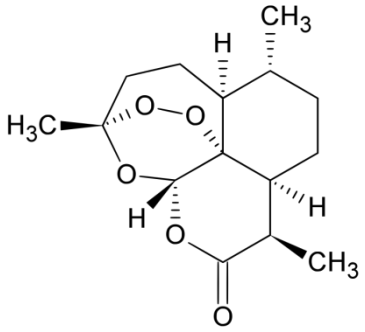
Reactive Biomolecules

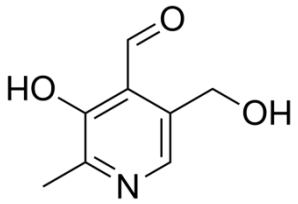


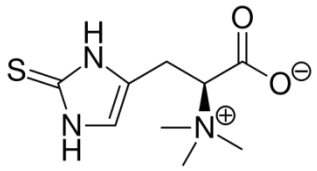


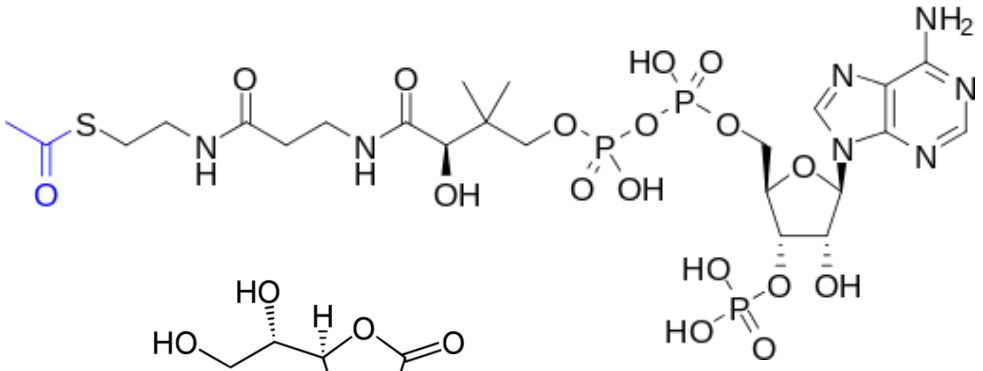


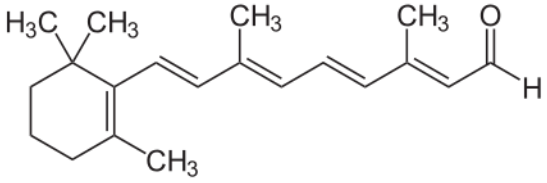
- > Acid/Base
- > Electrophile/nucleophile (Thiamin)
- > Acyl/CO_2 transfer (AcCoA, biotin)
- > Imine/enamine (pyridoxal, retinal)
- > Redox active (ascorbate, THBP)
- > Photochemistry (retinal)
- > Radicals (ergothioneine, artemisinin)

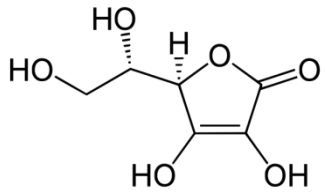












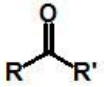
Organic Functional Groups

Alkane

Alkene

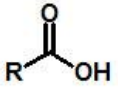
Conjugated Alkene

Alkyne

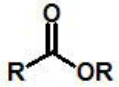


Ketone

Aldehyde

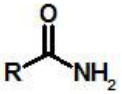


Carboxylic Acid

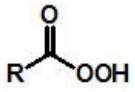


Ester

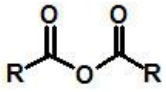
Acid Halide



Amide



Peroxy Acid

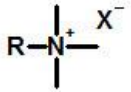


Anhydride

Primary Amine

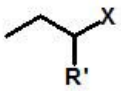
Secondary Amine

Tertiary Amine



Quaternary Ammonium Salt

Primary Alkyl Halide



Secondary Alkyl Halide

Tertiary Alkyl Halide

R—OH

Alcohol

R—O—R

Ether

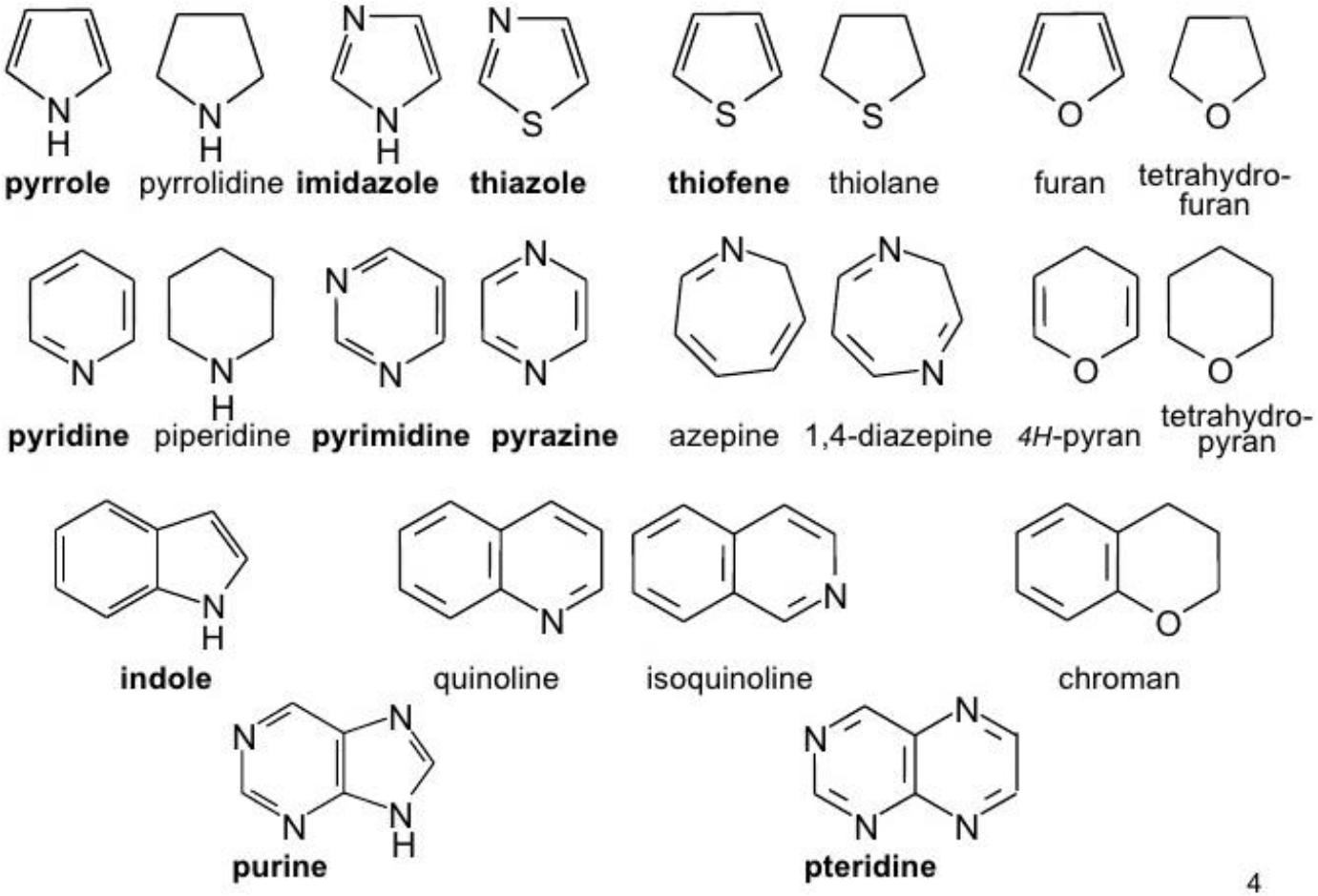
R—SH

Thiol

R—S—R

Thioether

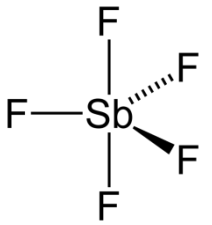
Heterocycles

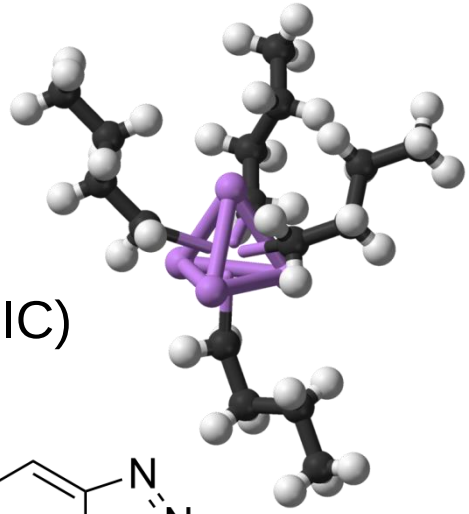


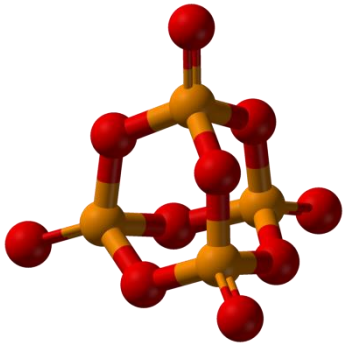
4

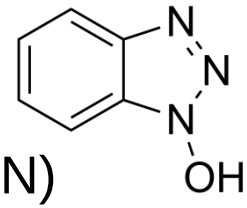
Reagents and Catalysts

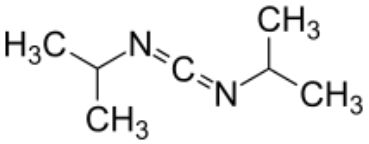
- Acids (SbF_5)
- Bases (*n*-BuLi)
- Condensing agents (P_2O_5, DIC)
- Nucleophiles (HOBt)
- Reducing agents ($NaBH_4$)
- Oxidizing agents (IBX)
- Radical sources (NBS, AIBN)
- Enzymes (L-Pro)
- Transition metals ($Pd(PPh_3)_4$)

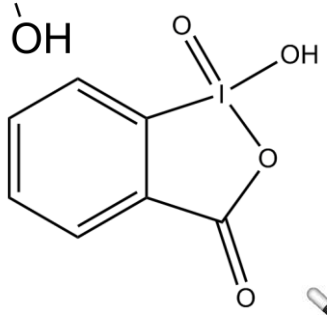


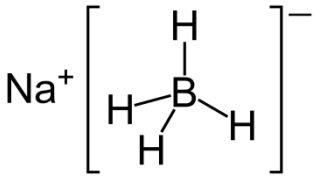


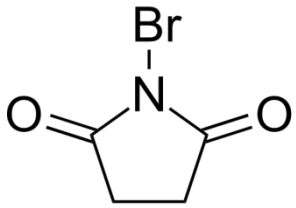


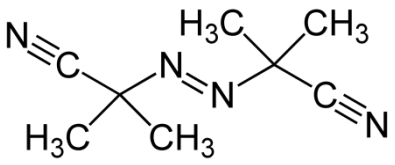


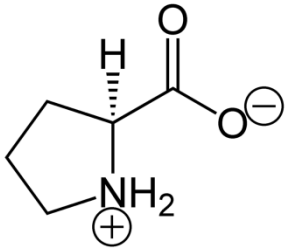


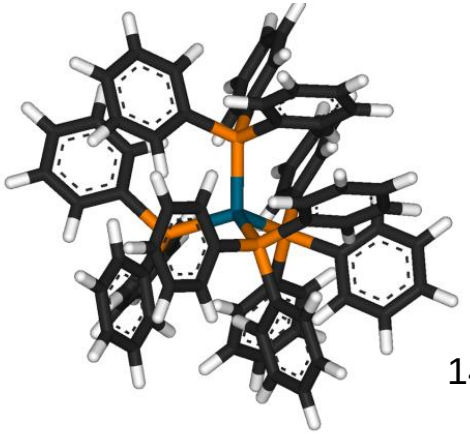












A Reaction in the Lab

- Materials
  - Toxicity/safety
  - Scale, equipment
  - Purity of starting material and reagents
- Conditions
  - Solubility, stirring
  - Temperature control (dry ice, reflux, mw)
  - Molar ratios
  - Exclusion of water, air, light
- Product isolation
  - Reaction analytics, timing
  - Quenching
  - Purification, yield
  - Structure determination

Chemical Reactions

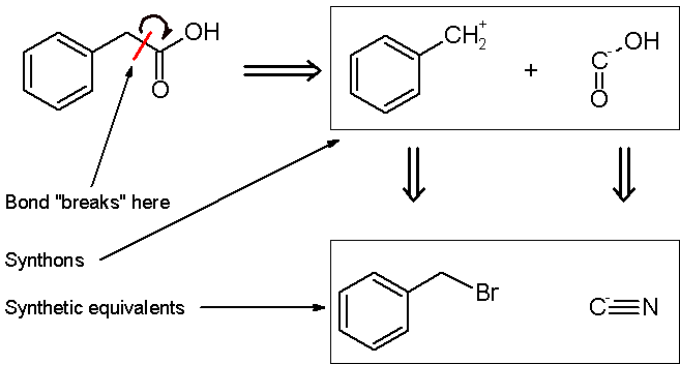
Molecules

Chemical Energy

Reactivity

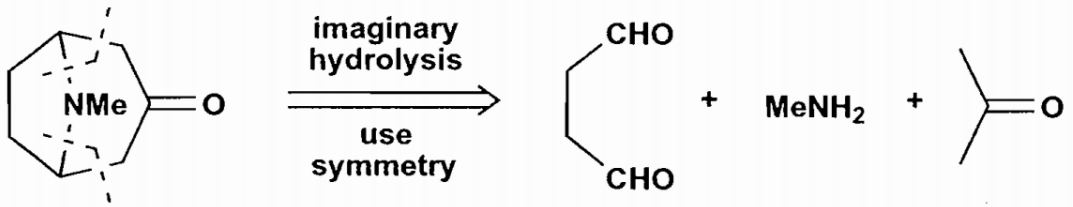
Synthesis Planning

Big Data



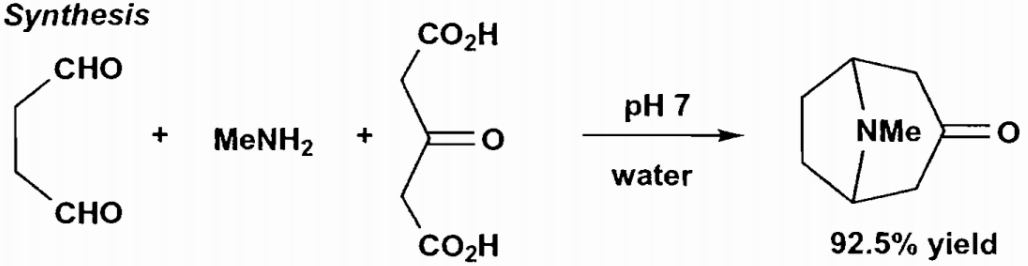
The Disconnection Approach

Tropinone: Robinson's Analysis



This was a famous synthesis because it is so short and simple and also because it makes a natural product in a way that imitates nature. The reaction is carried out at pH 7 in water. In fact Robinson didn't use acetone, as suggested by his 'imaginary hydrolysis', but acetone dicarboxylic acid. This procedure is an improved one invented by Schöpf[2] in 1935.

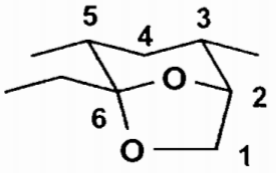
Tropinone: Synthesis



Amazingly, nobody picked up the idea until the 1960s when E. J. Corey at Harvard was considering how to write a computer program to plan organic syntheses.[3] He needed a systematic logic and he chose the disconnection approach, also called retrosynthetic analysis. All that is in this book owes its origin to his work. The computer program is called LHASA and the logic survives as a way of planning syntheses used by almost all organic chemists. It is more useful to humans than to machines.

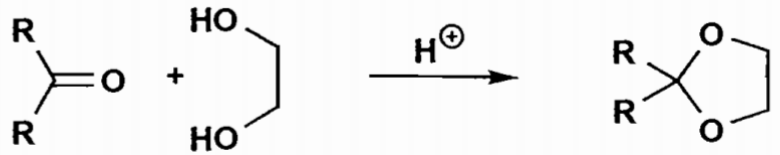
The Synthesis of Multistriatin

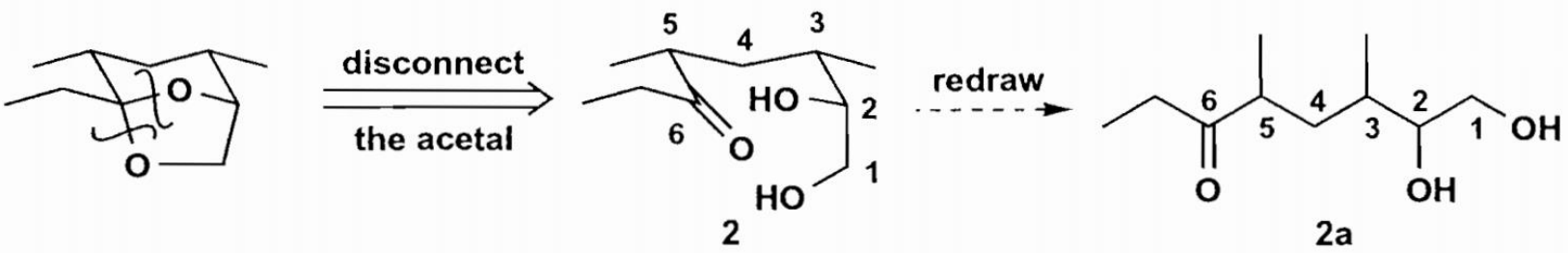
Multistriatin **1** is a pheromone of the elm bark beetle. This beetle distributes the fungus responsible for Dutch elm disease and it was hoped that synthetic multistriatin might trap the beetle and prevent the spread of the disease. It is a cyclic compound with two oxygen atoms both joined to the same carbon atom (C-6 in **1**) and we call such ethers *acetals*.

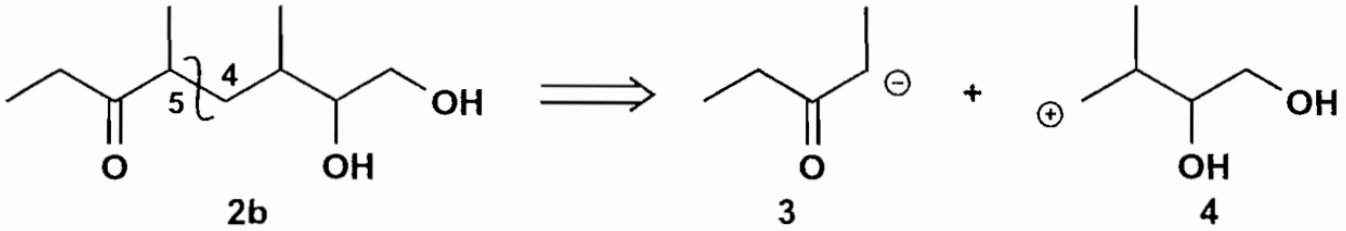


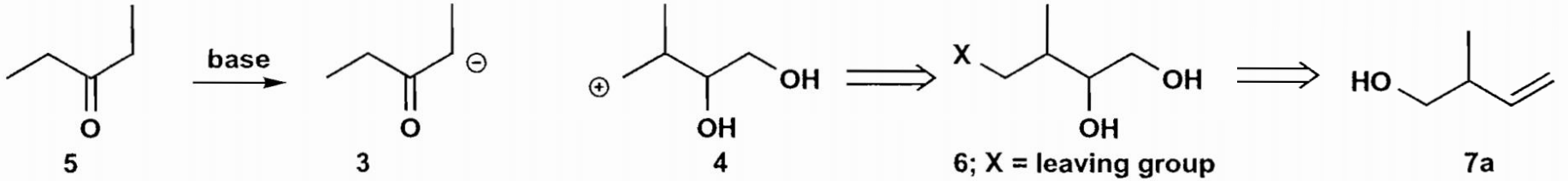
1; multistriatin

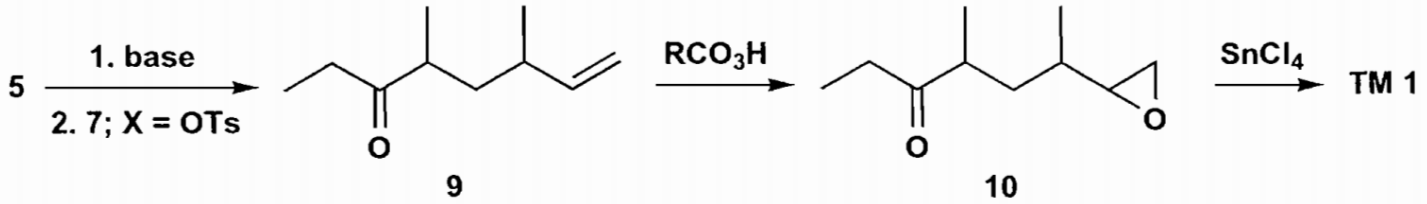
We know one good way to make acetals: the reliable acid-catalysed reaction between two alcohols or one diol and an aldehyde or ketone.



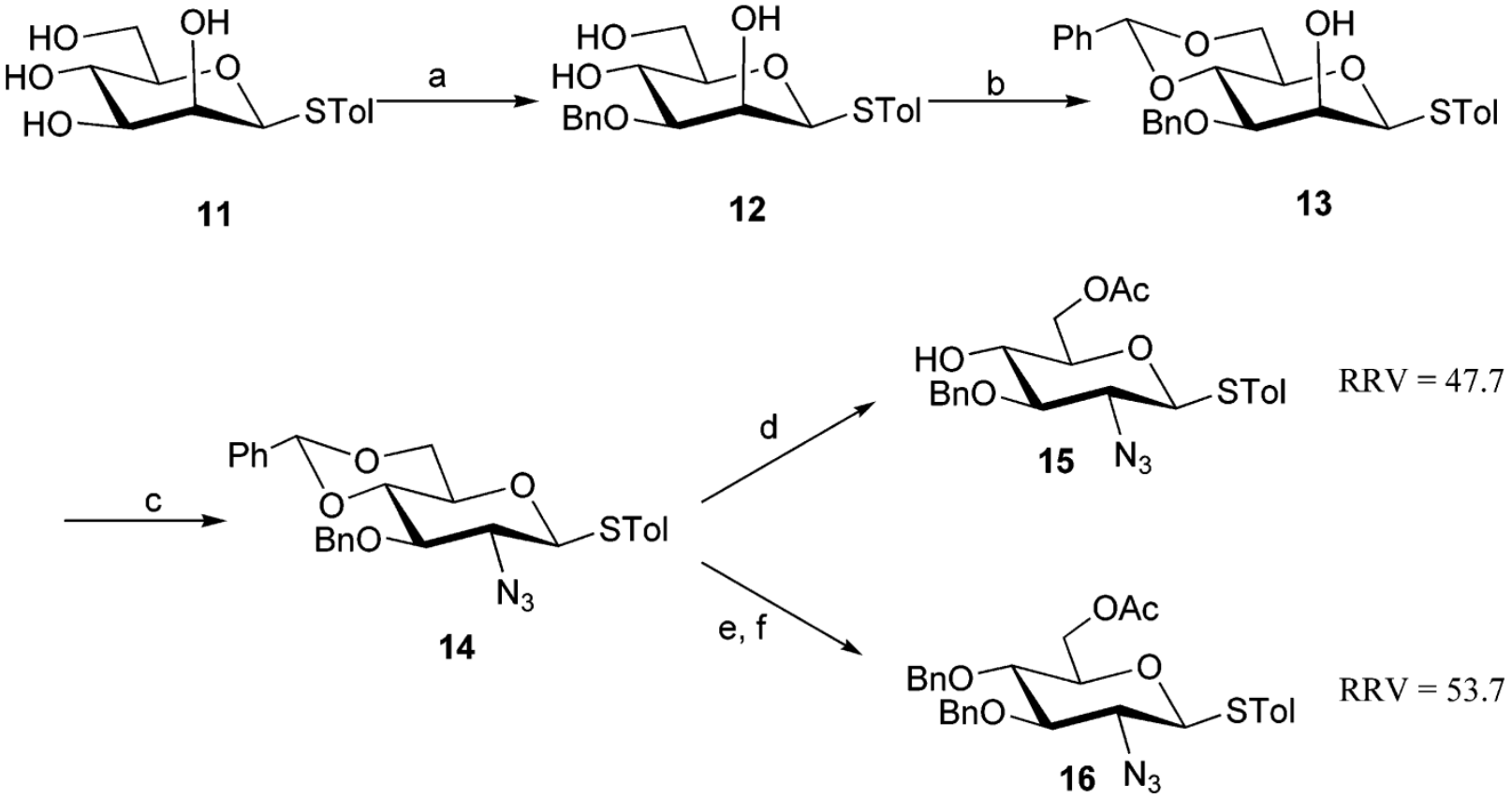

disconnect
the acetal
redraw
2
2a


2b
3
4

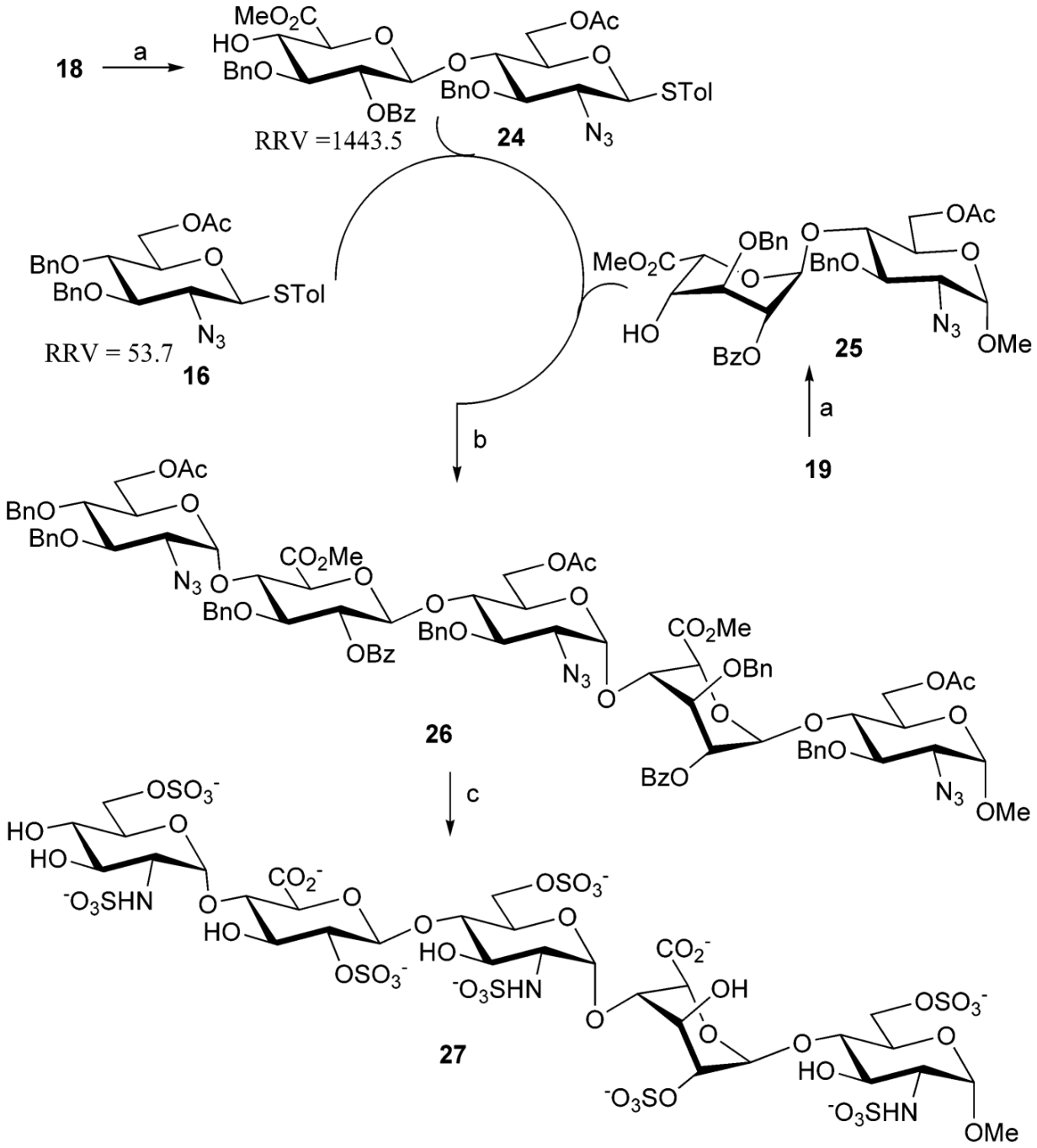

base
5
3
4
6; X = leaving group
7a


1. base
2. 7; X = OTs
RCO3H
SnCl4
TM 1
9
10

Protecting Groups



J. AM. CHEM. SOC. 2007, *129*, 12795−12800


18
a
RRV =1443.5
24
RRV = 53.7
16
25
19
b
a
26
c
27

Key Steps

Natural Product Total Synthesis

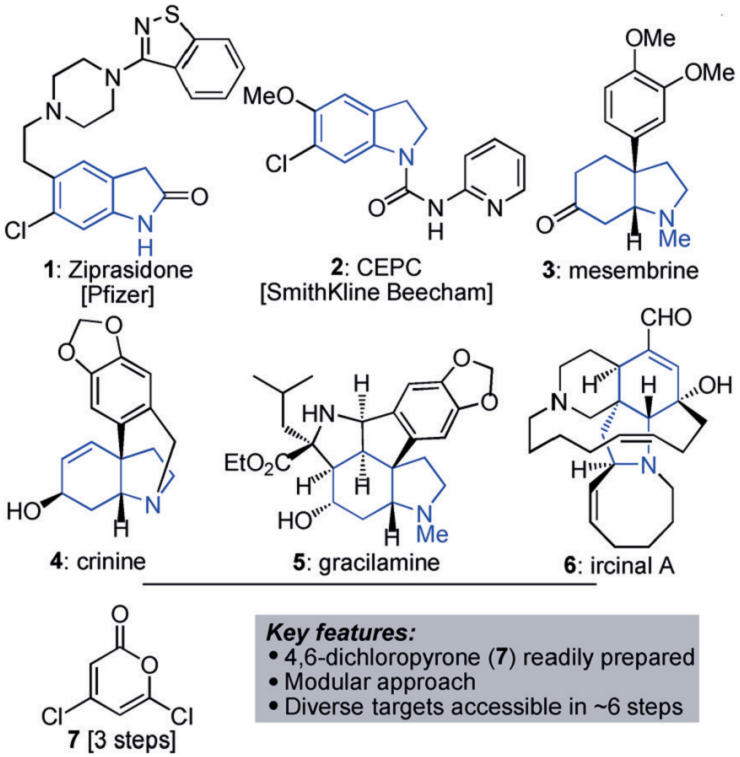
International Edition: DOI: 10.1002/anie.201510520
German Edition: DOI: 10.1002/ange.201510520

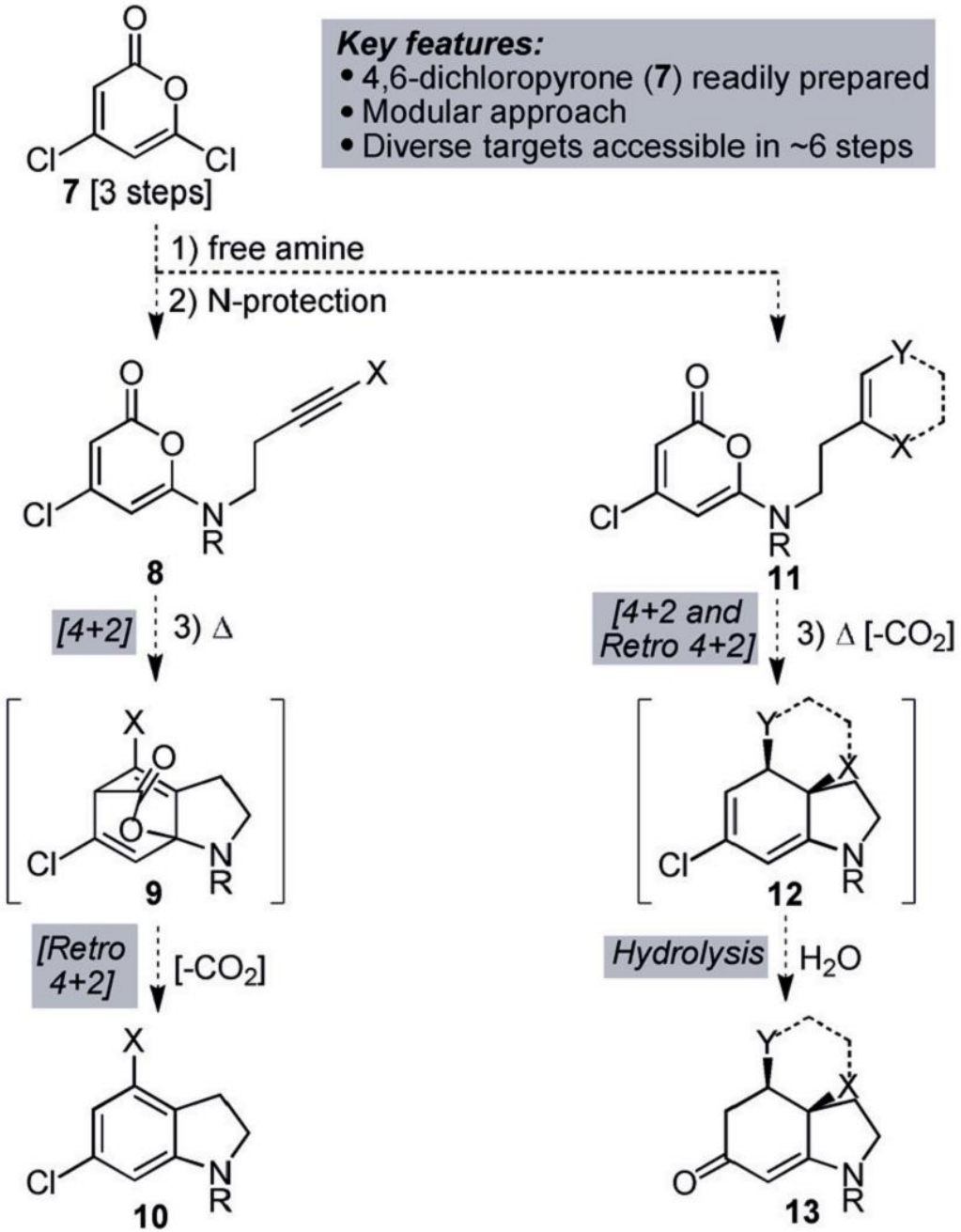
Pyrone Diels–Alder Routes to Indolines and Hydroindolines: Syntheses of Gracilamine, Mesembrine, and Δ^7-Mesembrenone

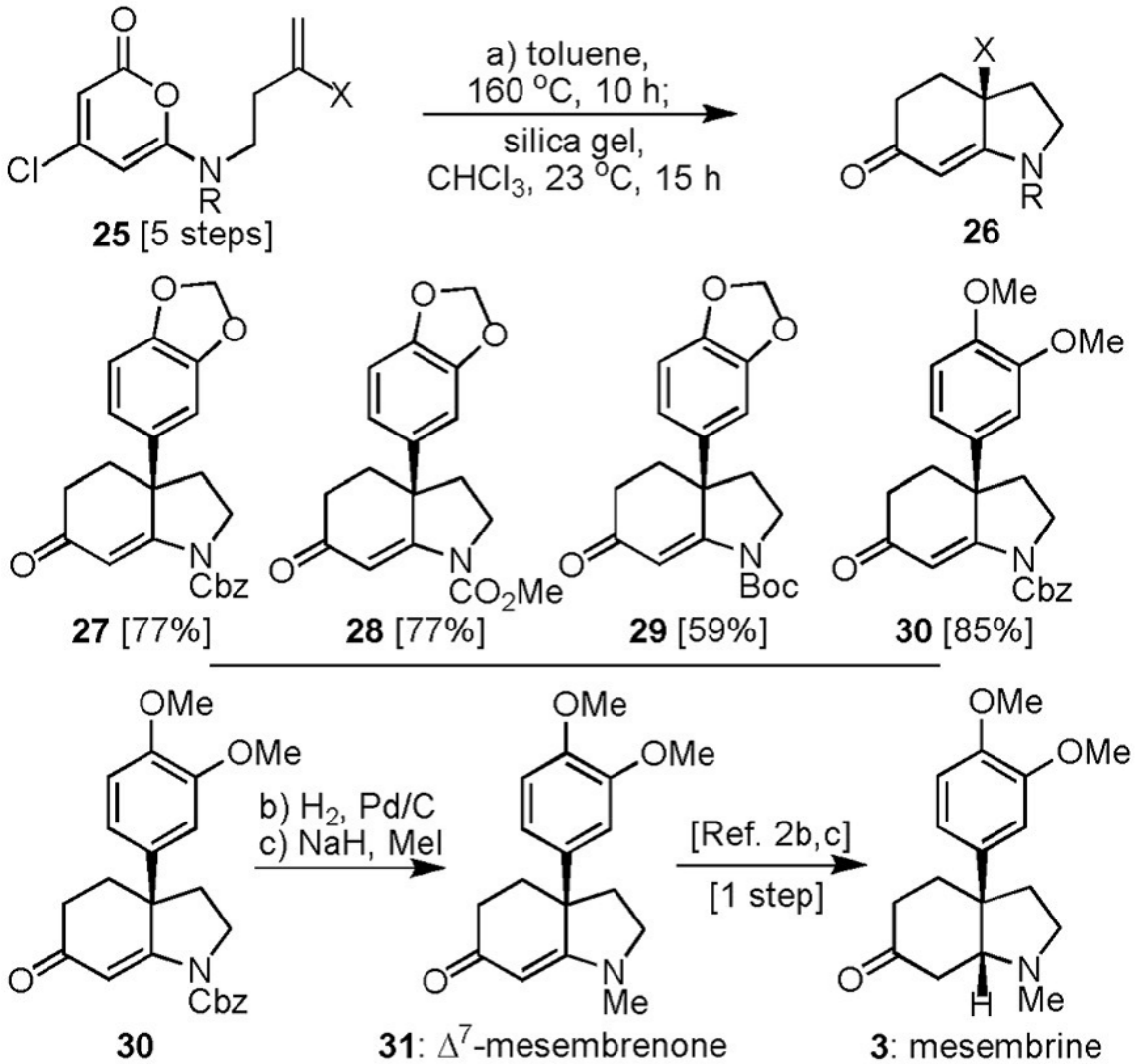
*Pei Gan, Myles W. Smith, Nathaniel R. Braffman, and Scott A. Snyder**

Abstract: *Although the Diels–Alder reaction has long been utilized for the preparation of numerous heterocycles, opportunities to extend its power remain. Herein, we detail a simple, modular, and robust approach that combines various amines regioselectively with 4,6-dichloropyrone to create substrates which, under appropriate conditions, can directly deliver varied indolines and hydroindolines through [4+2] cycloadditions with substitution patterns difficult to access otherwise. As an initial demonstration of the power of the strategy, several different natural products have been obtained either formally or by direct total synthesis, with efforts toward one of these—the complex amaryllidaceae alkaloid gracilamine—affording the shortest route to date in terms of linear step count.*

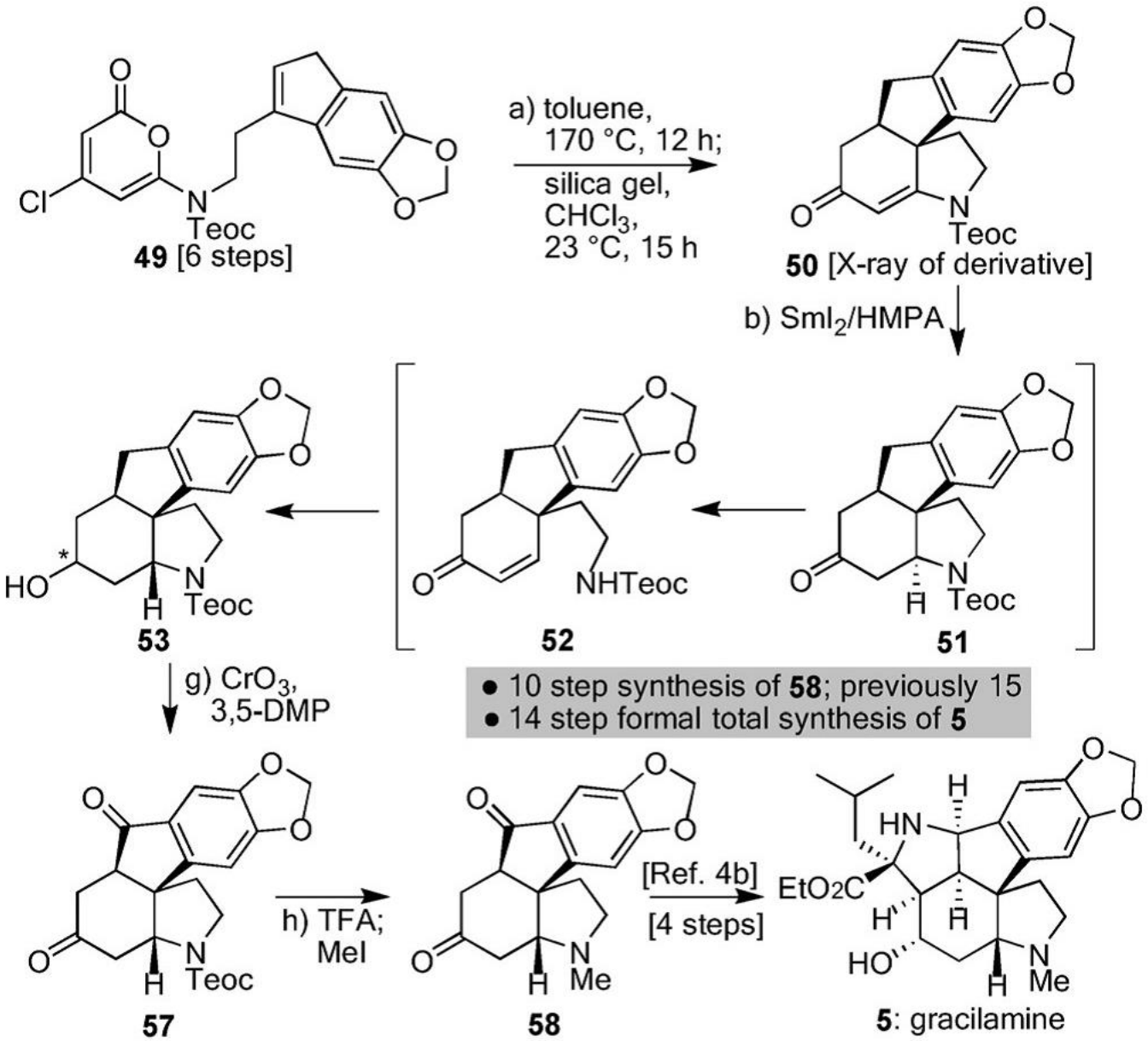
Indolines, oxindoles, and other variants of differential oxidation state are found in a plethora of pharmaceutical agents and natural products, a small selection of which is shown in the top part of Scheme 1.[1] Such ubiquity has induced the development of myriad synthetic approaches for



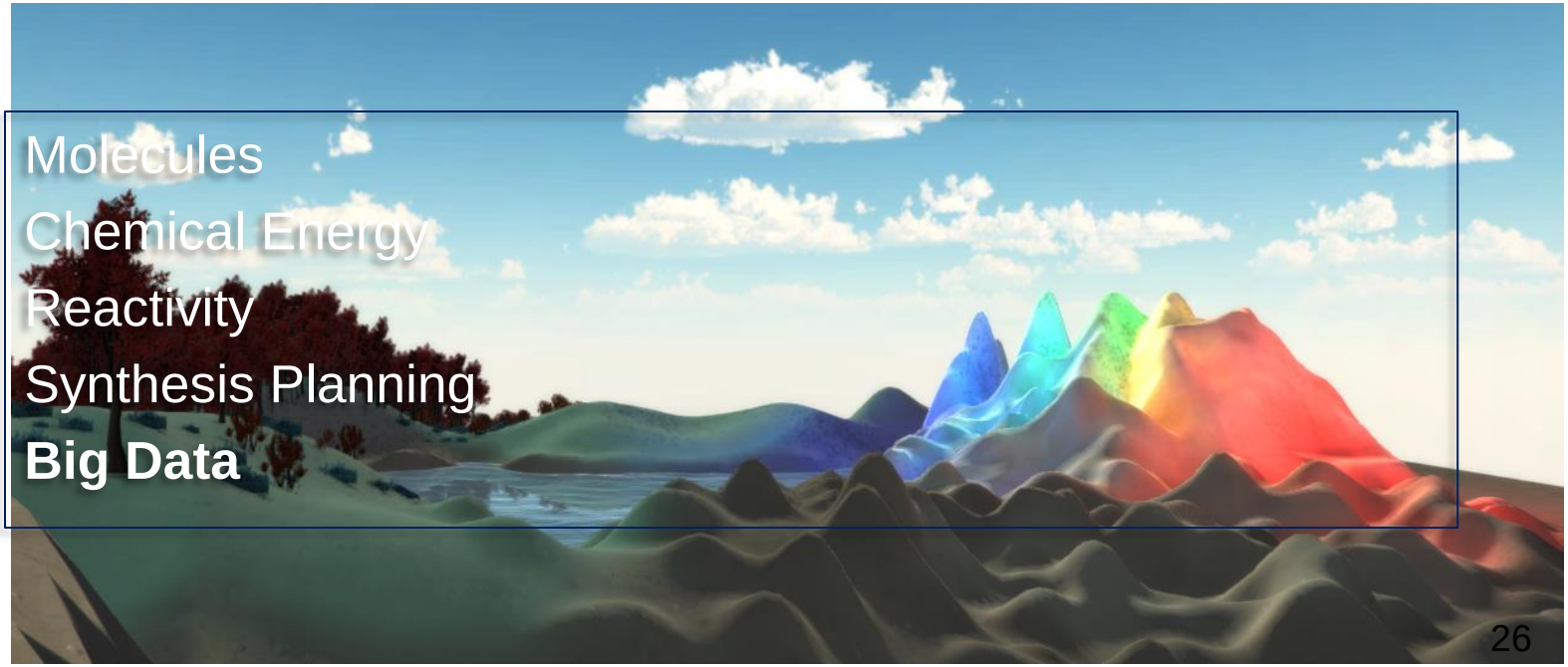

Key features:
• 4,6-dichloropyrone (7) readily prepared
• Modular approach
• Diverse targets accessible in ~6 steps
7 [3 steps]
1) free amine
2) N-protection
8
11
[4+2]
3) Δ
[4+2 and Retro 4+2]
3) Δ [-CO2]
9
12
[Retro 4+2]
[-CO2]
Hydrolysis
H2O
10
13



- 8 step synthesis of Δ^7-mesembrenone; 9 step synthesis of mesembrine (**3**)


a) toluene, 170 °C, 12 h; silica gel, CHCl3, 23 °C, 15 h
49 [6 steps]
50 [X-ray of derivative]
b) SmI2/HMPA
51
52
53
g) CrO3, 3,5-DMP
• 10 step synthesis of 58; previously 15
• 14 step formal total synthesis of 5
57
h) TFA; MeI
58
[Ref. 4b]
[4 steps]
5: gracilamine

Chemical Reactions



What to do with reaction information?

- Link substrates and products into networks to facilitate synthesis planning
  - Scifinder
  - Reaxys
  - Chematica
- Annotate reaction with conditions and experimental outcome to accelerate optimization
  - eNovalys
- Combine known substrates and known reactions to enumerate possible products
  - Pfizer
  - Enamine
  - Academic groups (P. Johnsson, G. Schneider, P. Kolb)
- Use theoretical transformations («mutations») to generate new molecules
  - Chemical space travel

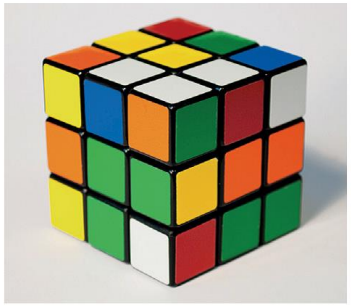
Computer-Aided Synthetic Planning

International Edition: DOI: 10.1002/anie.201506101
German Edition: DOI: 10.1002/ange.201506101

Computer-Assisted Synthetic Planning: The End of the Beginning

*Sara Szymkuć, Ewa P. Gajewska, Tomasz Klucznik, Karol Molga, Piotr Dittwald, Michał Startek, Michał Bajczyk, and Bartosz A. Grzybowski**

Table 1: Comparison of chess, the Rubik's cube and chemical synthesis.[a]

	Chess	Rubik's cube	Chemical synthesis
			
Number of players	Two	One	One
Movements	Small set of moves defined for each piece, some moves may not be allowed for some positions	Rotation of cube's single layer; always the same number of moves allowed	Very large (>10000) number of possible moves (i.e., reaction rules); applicable moves depend on the structure of the molecule; database of moves can grow as chemistry advances
Start position	Always the same initial arrangement of pieces on the board; "white" player starts	(Random) rearrangement of the cube	Target that needs to be synthesized
Position	Current configuration of the pieces on the board	Configuration of the cube	Set of substrates/synthons at each step

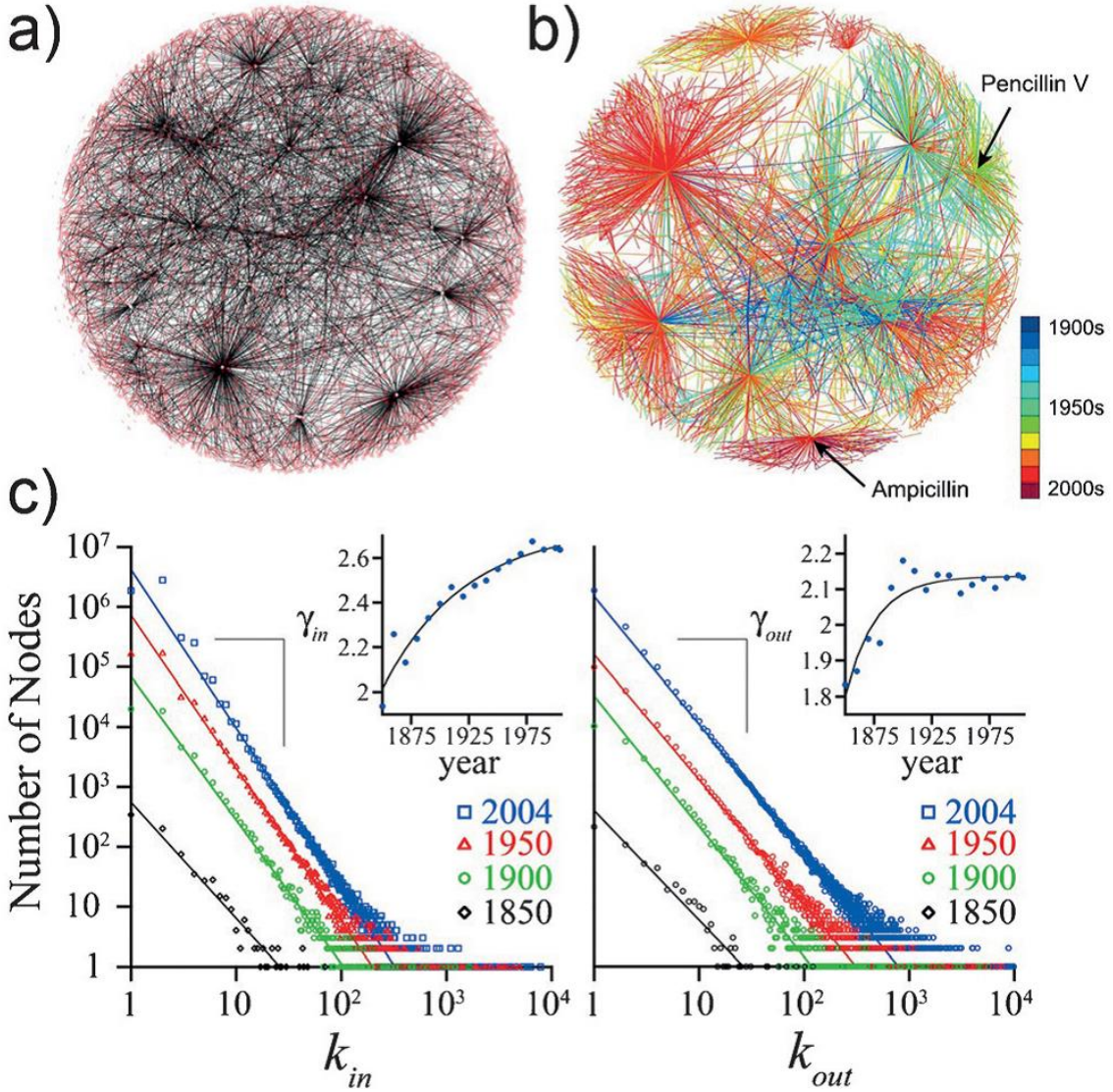
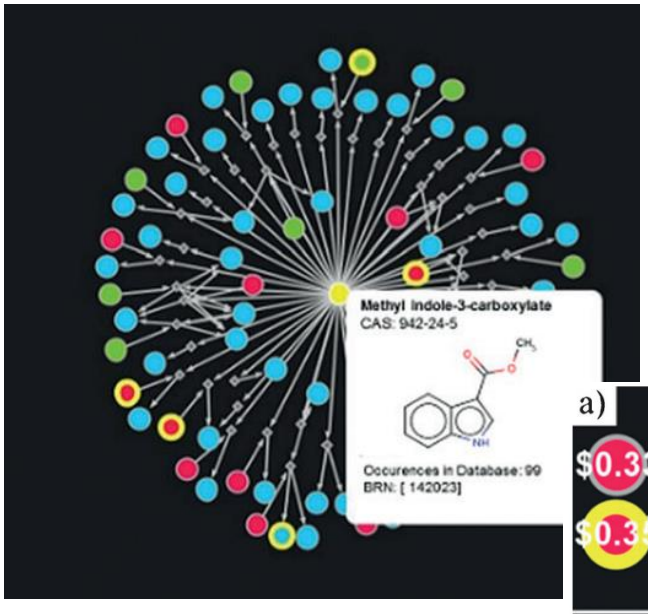
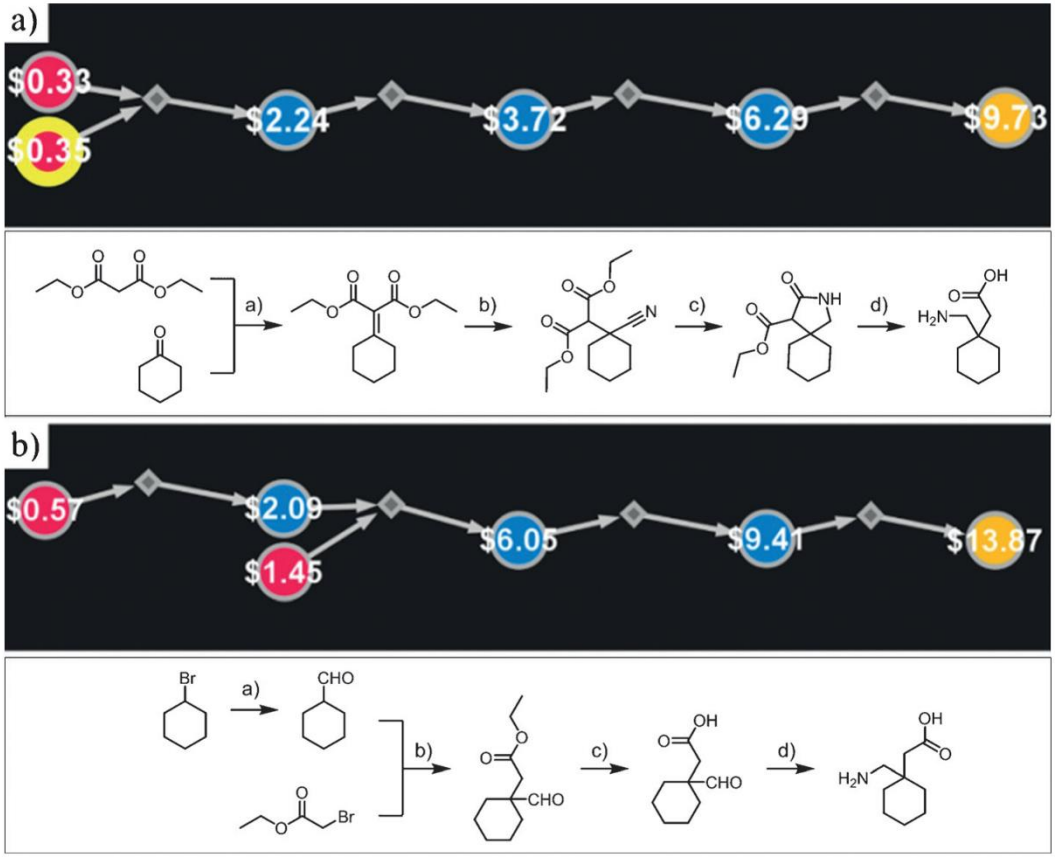


Figure 2. The structure and dynamics of the Network of Chemistry (NOC). a) A small (≈5500 nodes) fragment of NOC where individual nodes represent the molecules and arrows represent reactions. The

Chematica


Methyl indole-3-carboxylate
CAS: 942-24-5
Occurences in Database: 99
BRN: [142023]


a)
$0.33
$0.35
$2.24
$3.72
$6.29
$9.73
a)
b)
c)
d)
b)
$0.57
$2.09
$1.45
$6.05
$9.41
$13.87
Br
CHO
a)
b)
c)
d)

ACS Combinatorial Science

Review

pubs.acs.org/acscombsci

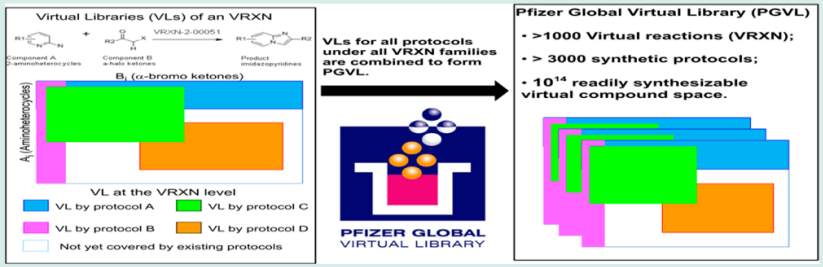
Pfizer Global Virtual Library (PGVL): A Chemistry Design Tool Powered by Experimentally Validated Parallel Synthesis Information

Qiyue Hu,*,† Zhengwei Peng, Scott C. Sutton, Jim Na, Jaroslav Kostrowicki, Bo Yang, Thomas Thacher, Xianjun Kong, Sarathy Mattaparti, Joe Zhongxiang Zhou, Javier Gonzalez, Michele Ramirez-Weinhouse, and Atsuo Kuki

Pfizer Global Research and Development, La Jolla Laboratories, 10770 Science Center Drive, San Diego, California 92121, United States

Supporting Information

ABSTRACT: An unprecedented amount of parallel synthesis information was accumulated within Pfizer over the past 12 years. This information was captured by an informatics tool known as PGVL (Pfizer Global Virtual Library). PGVL was used for many aspects of drug discovery including automated reactant mining and reaction product formation to build a synthetically feasible virtual compound collection. In this report, PGVL is discussed in detail. The chemistry information within PGVL has been used to extract synthesis and design information using an intuitive desktop Graphic User Interface, PGVL Hub. Several real-case examples of PGVL are also presented.



KEYWORDS: *drug discovery, cheminformatics, molecular design, parallel synthesis, combinatorial library, synthesis protocol, knowledge system, Pfizer Global Virtual Library (PGVL), reactant, product, enumeration, canonical parallel synthetic protocols (CPSP), VL reaction registrars, reaction/protocol developers, automated reactant mining objects (ARM Object), virtual reaction objects (VRXN Object), , virtual library (VL)*

Review of previous VL

component to defining a synthetically feasible virtual library. In 2005, Nikitin and co-workers also constructed a large and diverse space of virtual compounds with potential applications in drug discovery.[7] This collection was built based on reaction schemes from approximately two hundred literature papers and four hundred individual combinatorial libraries. The building blocks were either from the same literature papers or chosen from commercially available reactants using drug-like and reaction suitability filters. The Nikitin work represented a significantly broader coverage of diversity in chemistry and the virtual compounds were more likely to be synthetically accessible since they were based on experimental precedence. One of the best examples of the association of synthetic feasible virtual space with tight integration of synthesis protocols and design is found in the AIDD system published by Manly.[8] Because of the limited number of synthetic protocols captured in AIDD, the virtual compound space of 150 million compounds was, understandably, smaller than those from Andrews et al[6] and Nikitin at al.[7] Lessel and co-workers reported on similar software called BI-CLAIM[9] in 2009, which used 300 000 reactants leading to about 5×10^{11} products.

Product Enumeration

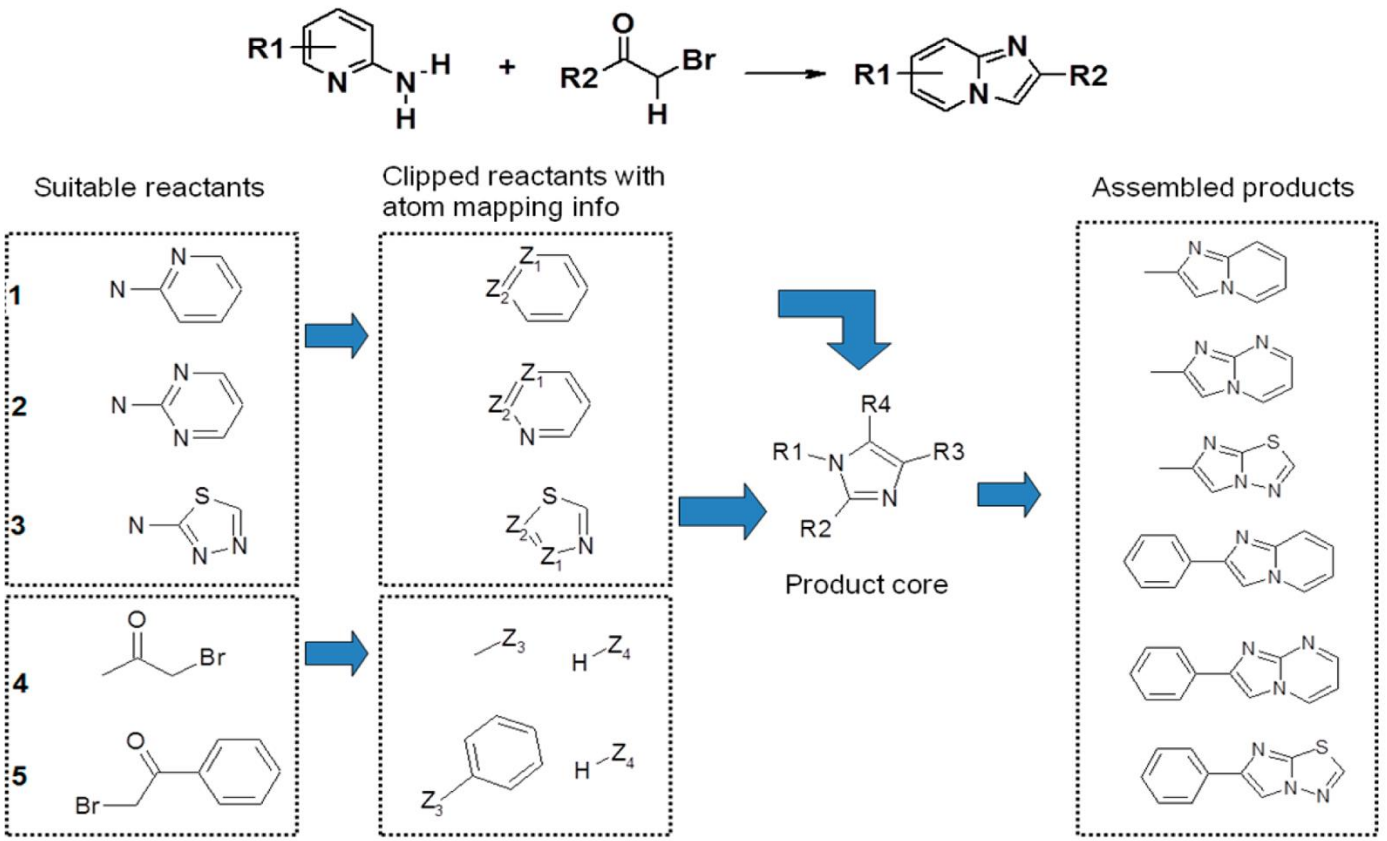


Figure 5. Product enumeration using reactant clipping and product assembly. The imidazopyridine-forming reaction using aminoheterocycles and alpha-bromo ketones is shown as an example. The first step involves clipping of reactant molecules at their respective reactive centers and label the clipping site with numbered *z*-labels (z_1, z_2, etc.). The second step involves assembly of products based on a product core structure with numbered R-group attachments (R_1, R_2, etc.) and forming new bonds at the appropriate z-labeled sites.

Calculated Library Size

The virtual library sizes within PGVL are summarized in Table 1. A total of 1244 unique transformations, called VRXNs, are

Table 1. Snapshot of PGVL's Scale (as of January 2011)

starting material	no. VRXNs	no. of basis products	no. of virtual products
reagents from in-house inventory			
2-component RXN	436	2 453 094	4.034×10^{9}
3-component RXN	725	5 907 543	1.102×10^{13}
4-component RXN	83	786 037	2.993×10^{14}
total	1244	9 146 674	3.103×10^{14}
reagents from ACD of Accelrys			
2-component RXN	436	23 308 473	4.905×10^{11}
3-component RXN	725	46 975 889	9.216×10^{15}
4-component RXN	83	5 504 761	1.269×10^{18}
total	1244	75 789 123	1.278×10^{18}

captured. The enumeration of all possible chemistry space when using the Pfizer available inventory encompasses 10^{14} virtual compounds. When using the commercial (ACD) inventory, the possible chemistry space encompasses 10^{18} virtual compounds.

Heterocycles

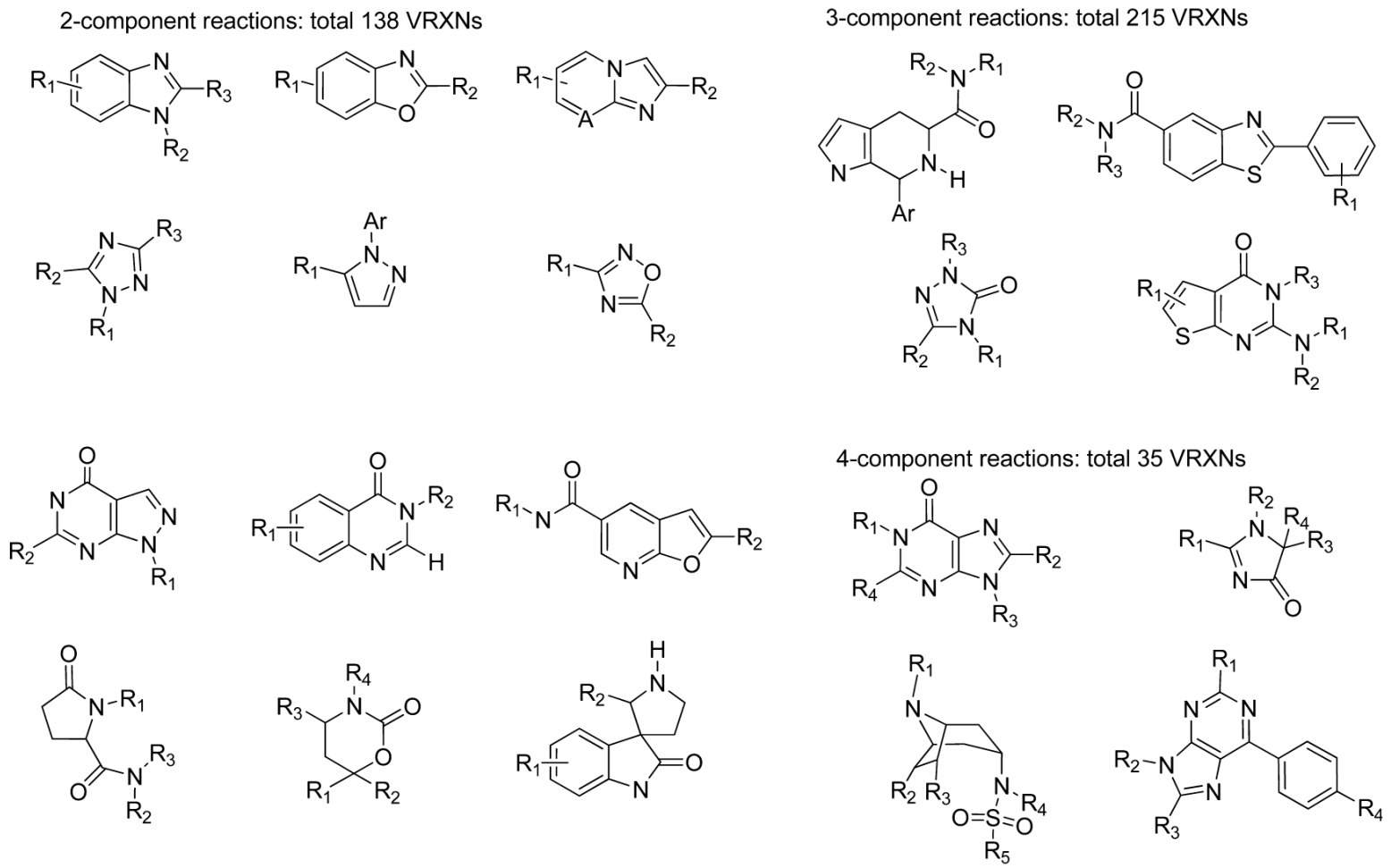


Figure 7. Summary and some examples of VRXNs that form heterocyclic rings.

Hit Optimization

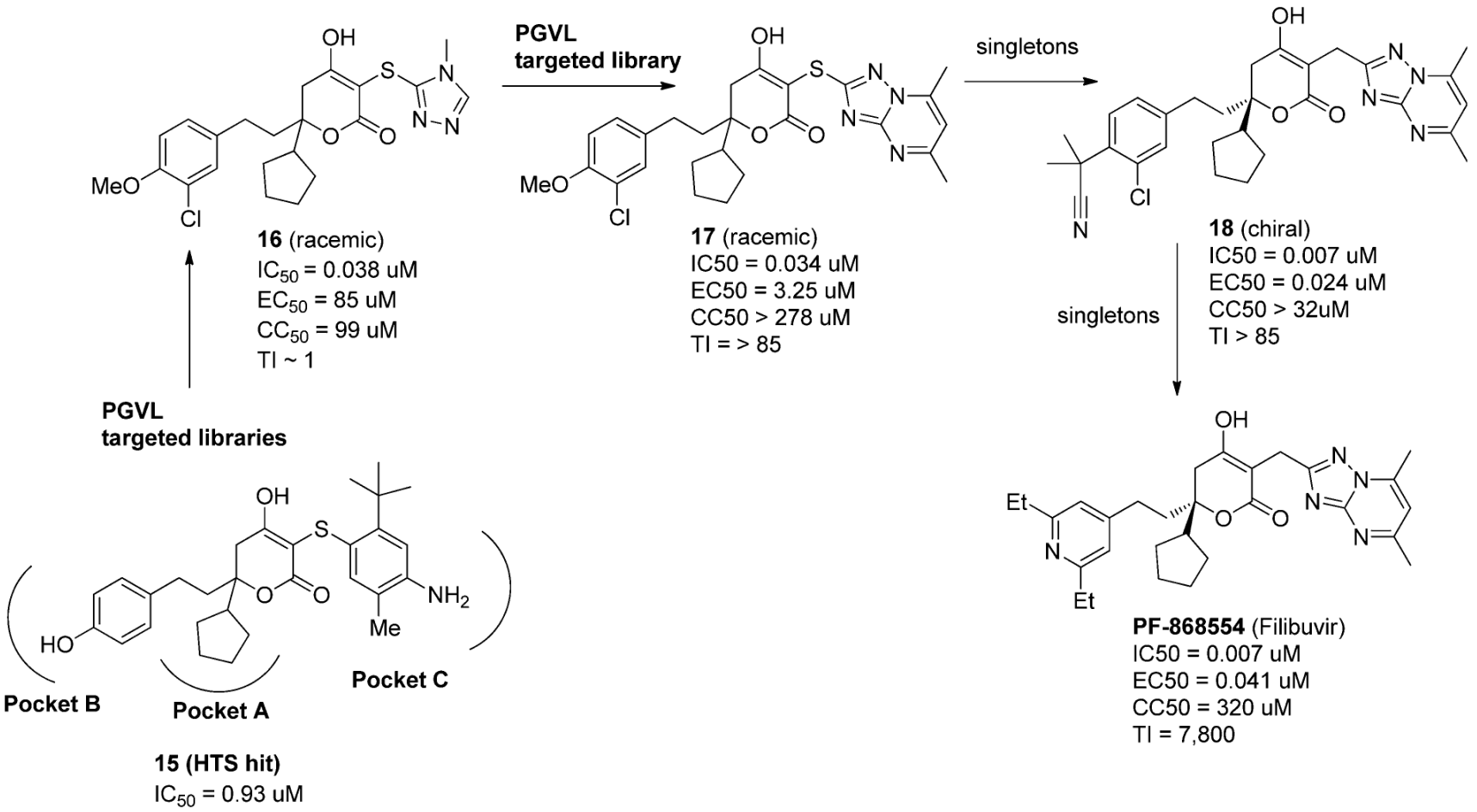
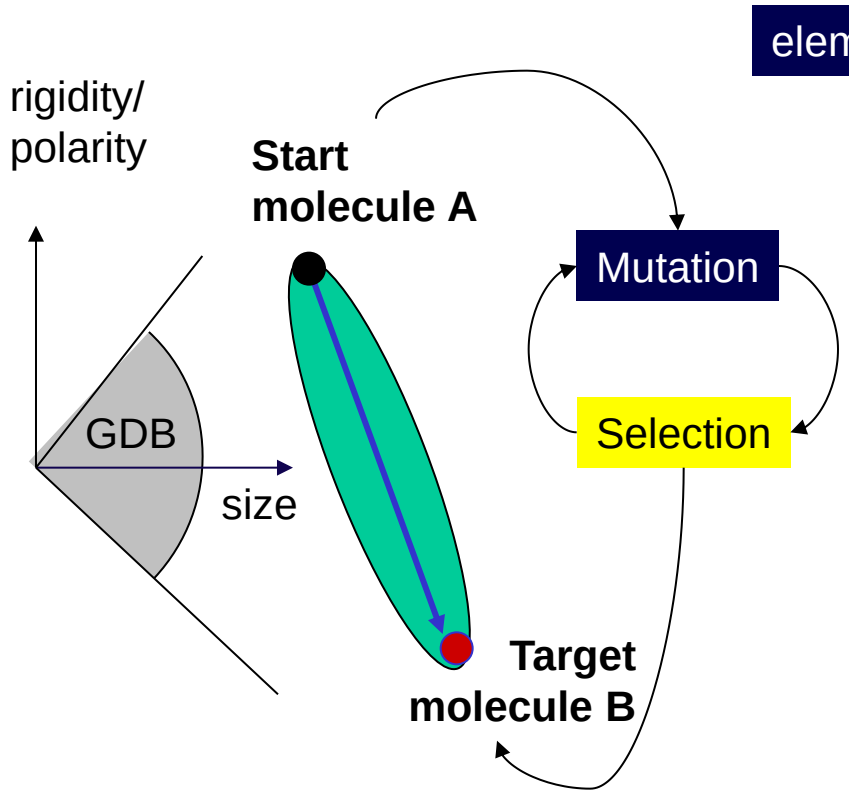


Figure 9. Discovery of Filibuvir from HTS hit **15**.

Chemical Space Travel



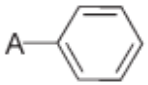
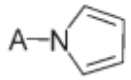
elementary changes in chemical structure

Similarity to Target B:
T_{SF} x T_{Phfp}

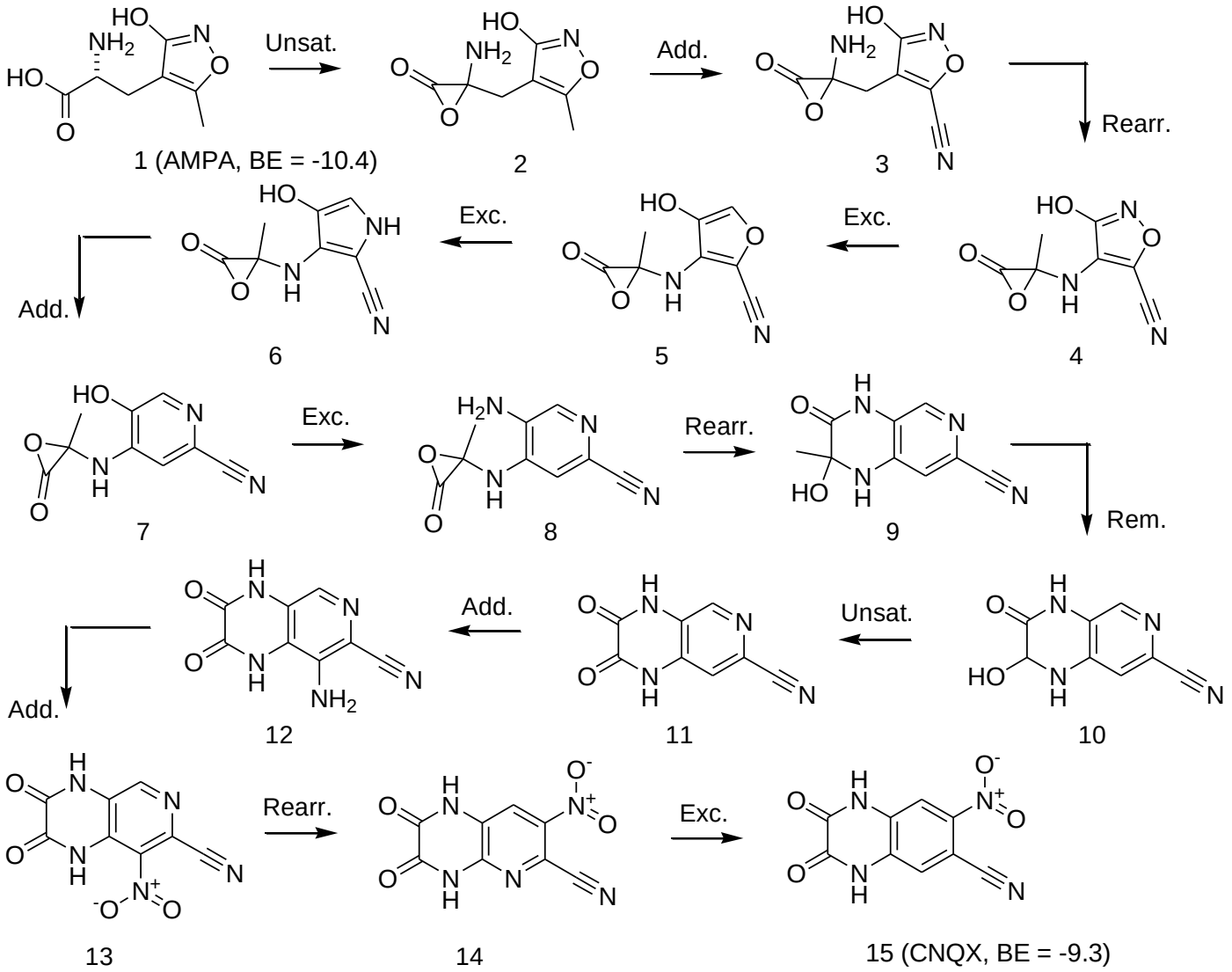
R. Van Deursen, J.-L. Reymond *ChemMedChem* **2007**, *2*, 636-640

Mutations

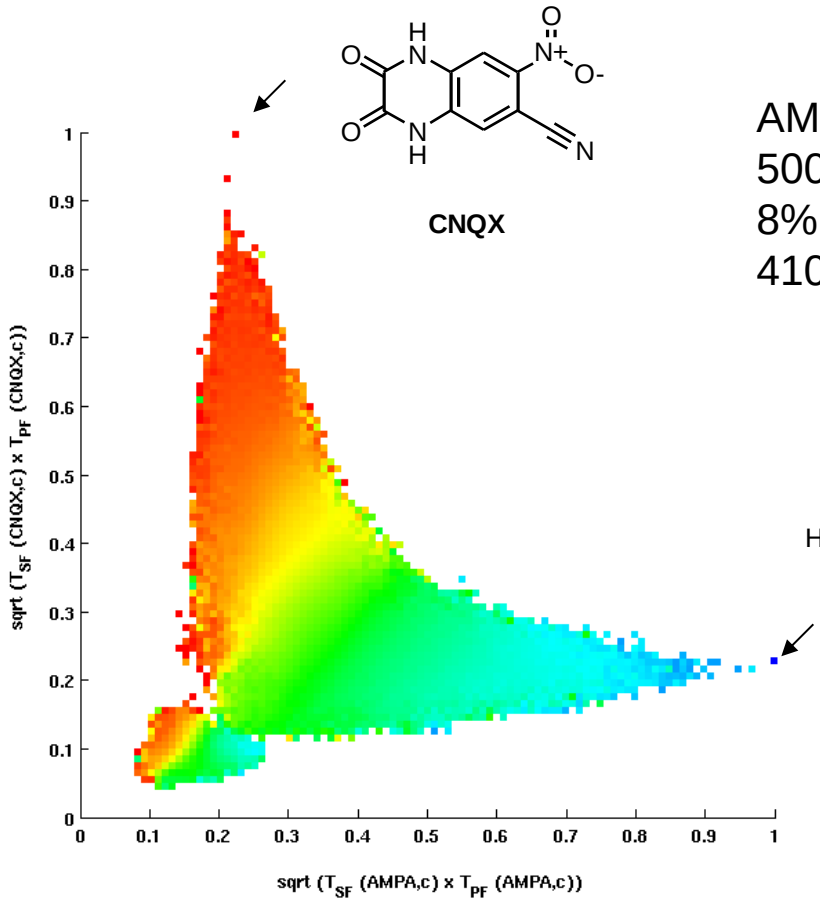
Table 1. Structural mutations used with the spaceship program.

Nearest neighbour mutations[a]	
Atom type exchange[b,c]	Replaces any atom by another atom type
Atom inversion[c]	Inverts two neighbouring atoms
Atom removal[c]	Primary: A–X→A
	Secondary: A–X-A→A–A
	Tertiary: XA_3→A–A–A
	(max. 6 combinations if 3 different A's)
	$A_2CH–CHA_2$ or $A_2C{=}CA_2$→CA_4
	Quaternary: XA_4→A–A–A–A or $A(A)_3$
	(max. 16 combinations if 4 different A's)
Atom addition[b,c]	On terminal atoms: A→A–X
	In any bond: A–A→A–X–A
	In chains: A–A–A→XA_3; A–A–A–A→XA_4
	Quaternary centres:
	CA_4→$A_2CH–CHA_2$ and $A_2C{=}CA_2$
	(max. 6 combinations if 4 different A's)
Bond saturation[c]	Breaks a cyclic σ- or any π-bond
Bond unsaturation	Makes a cyclic σ- or π-bond
Bond rearrangement[c]	Breaks a σ- or π-bond and inserts it anywhere else in the molecule
Non-nearest neighbour mutations	
	A-CH_3→ 
Aromatic ring addition[c,d]	A-NH_2→ 
	H_2O→ 

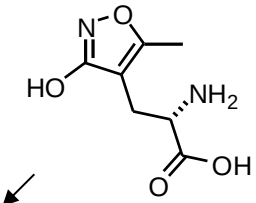
Trajectory Example



AMPA ↔ CNQX

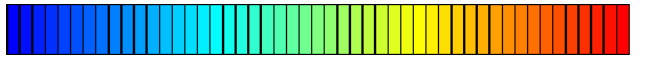


CNQX

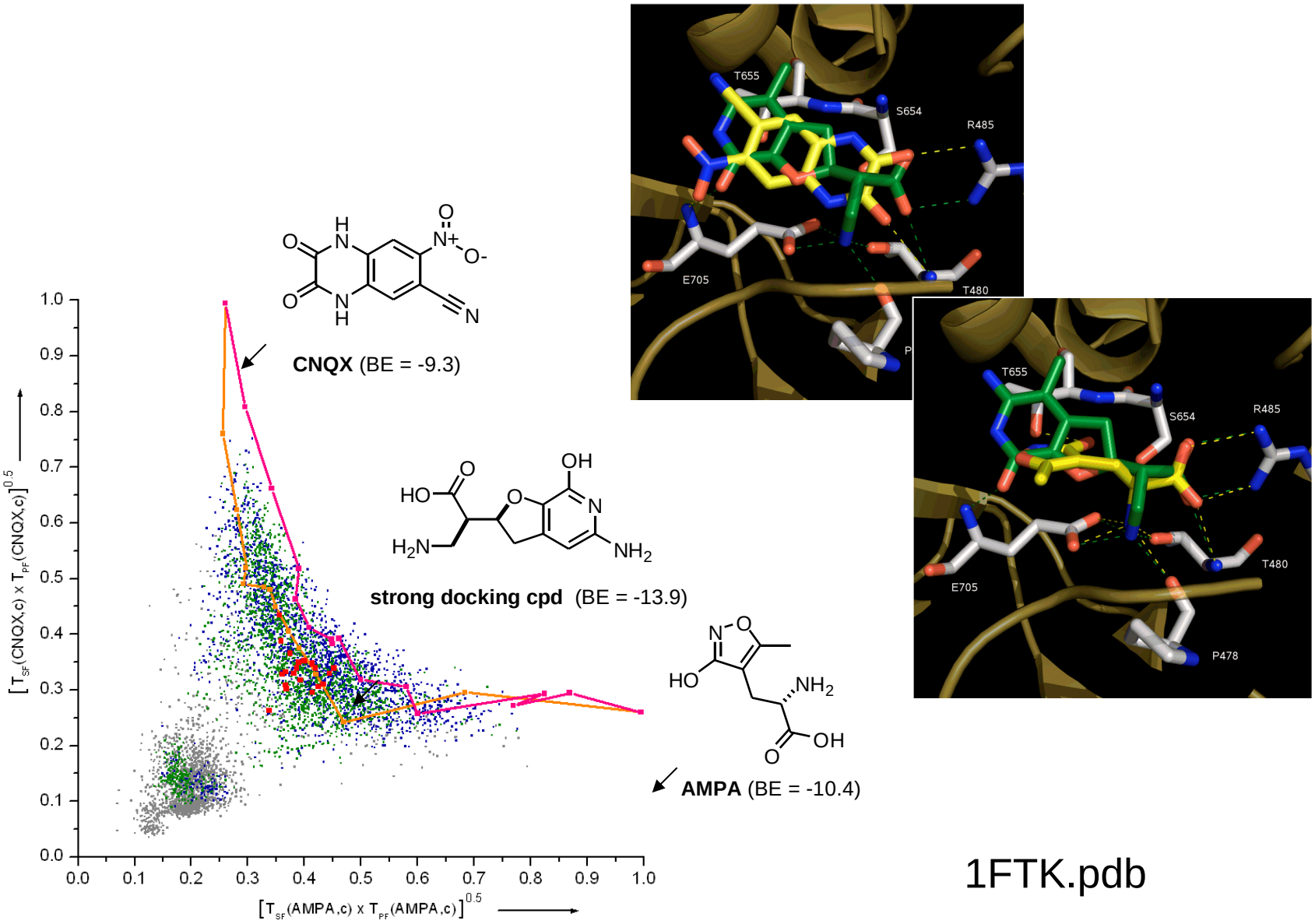


AMPA

AMPA ↔ CNQX
500 runs, 559,658 cpds
8% from 6.9 million generated
410 cpds (0.1%) also in ZINC



0 **0.50** **1.0**


T655
S654
R485
E705
T480
P
T655
S654
R485
T480
E705
P478
CNQX (BE = -9.3)
strong docking cpd (BE = -13.9)
AMPA (BE = -10.4)
[T_SF(CNQX,c) x T_PF(CNQX,c)]^0.5
[T_SF(AMPA,c) x T_PF(AMPA,c)]^0.5
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

1FTK.pdb

Recommended Reading

- «Organic Synthesis: The Disconnection Approach, 2nd Edition» Stuart Warren, Paul Wyatt

- «Pfizer Global Virtual Library (PGVL): A Chemistry Design Tool Powered by Experimentally Validated Parallel Synthesis Information», Q. Hu *et al.*, *ACS Comb. Sci.*, **2012**, 14, 579–589

- «Computer-Assisted Synthetic Planning: The End of the Beginning», S. Szymkuć *et al.*, *Angew. Chem. Int. Ed.* **2016**, *55*, 5904-5937

- «Chemical Space Travel», R. van Deursen *et al.* *ChemMedChem* **2007**, *2*, 636-640.