



Marcus Gastreich

Spotting Binders in 10^{12} Compounds



Estimated no. of stars in universe: 10^{23}
Compounds searched traditionally: 10^7

“Interesting” compounds: 10^{60}

Given the “Similarity Principle”...

1. Which of the 10^{60} are accessible?
2. How to search that many compounds?
3. Four eyes may see more than two... - Right?

1. Which Ones are the Appealing Ones?

- “Lead-like”, good ADME/tox properties
- Synthetically accessible
- Unpatented
(but given the high number of molecules,
this should be a smaller problem...)

The Synthesis Dilemma

- Many of the compounds will not be accessible
- It is un-doable to test every molecule:

Assumption:

1 millisecond compute time per compound

Example:

$$10^{12} \times 0.001s = 10^9s$$

~ 16,666,666 mins

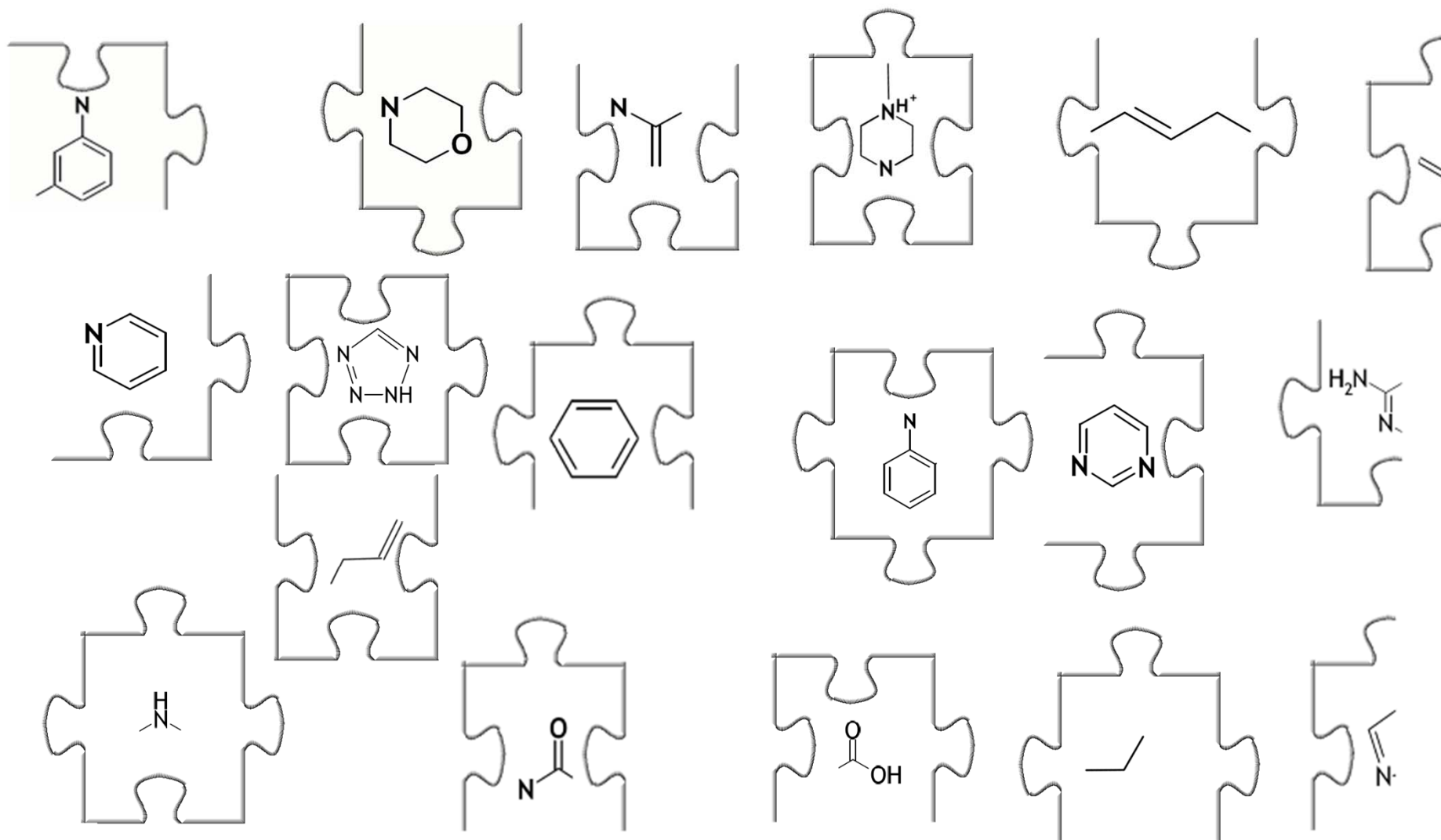
~ 277,777 hrs

~ 11,500 days

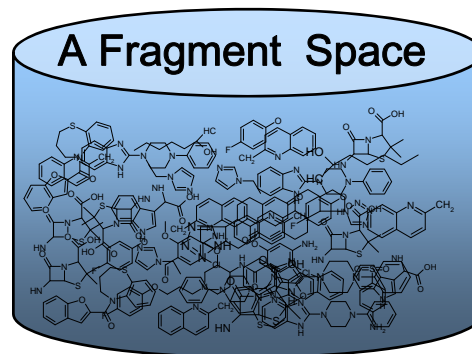
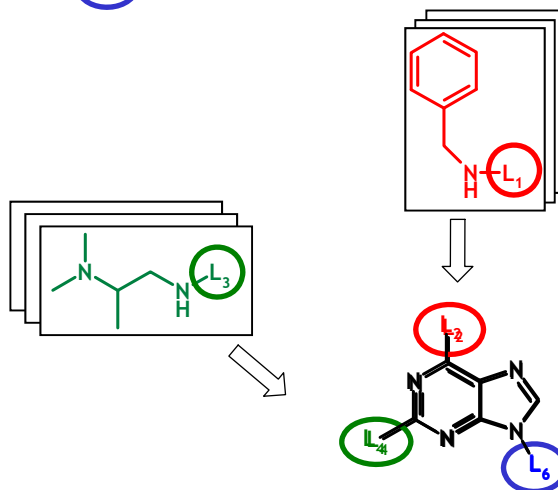
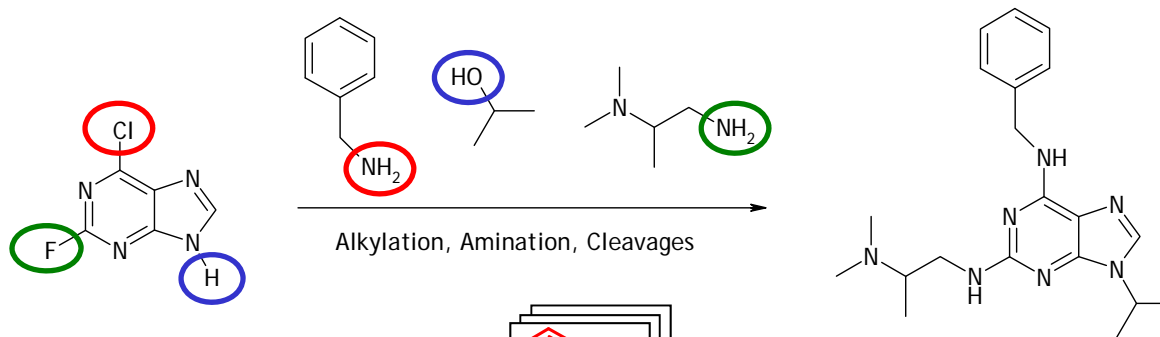
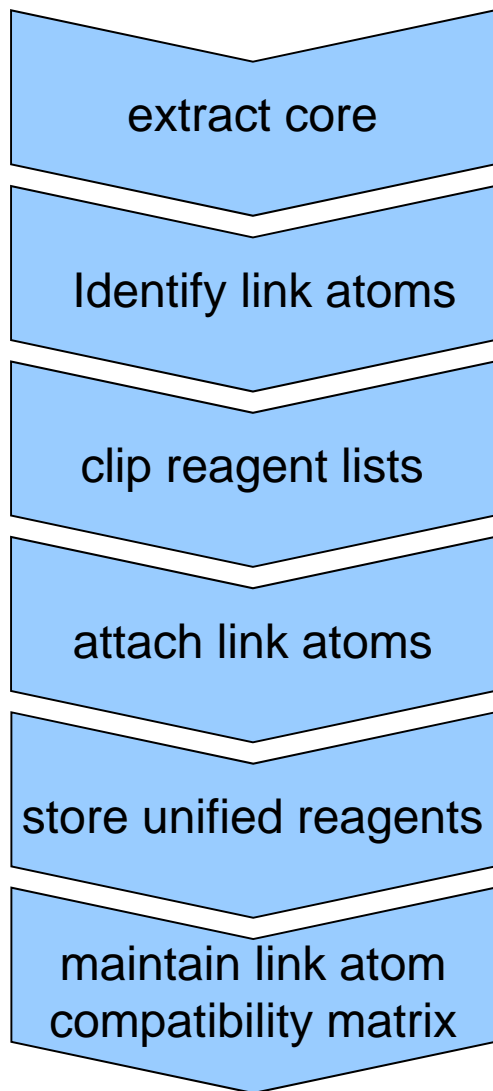
Solution: Do not Enumerate!

- Instead of checking compounds afterwards, we can encode “reactions” and assemble molecules on-the-fly.

Our Approach: Fragments to Leads



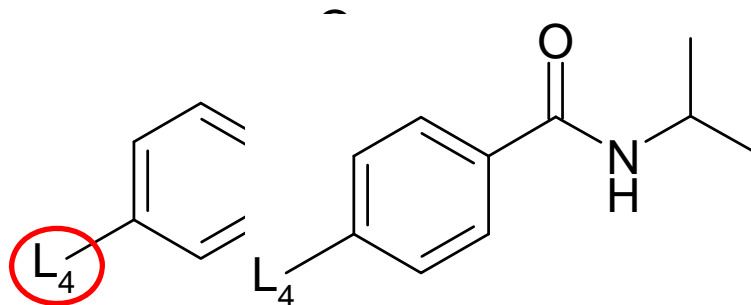
Encoding Reactions in a Nutshell



	L ₁	L ₂	L ₃	L ₄	L ₅	L ₆	
L ₁		X					
L ₂	X						
L ₃				X			...
L ₄			X				
L ₅						X	
L ₆					X		

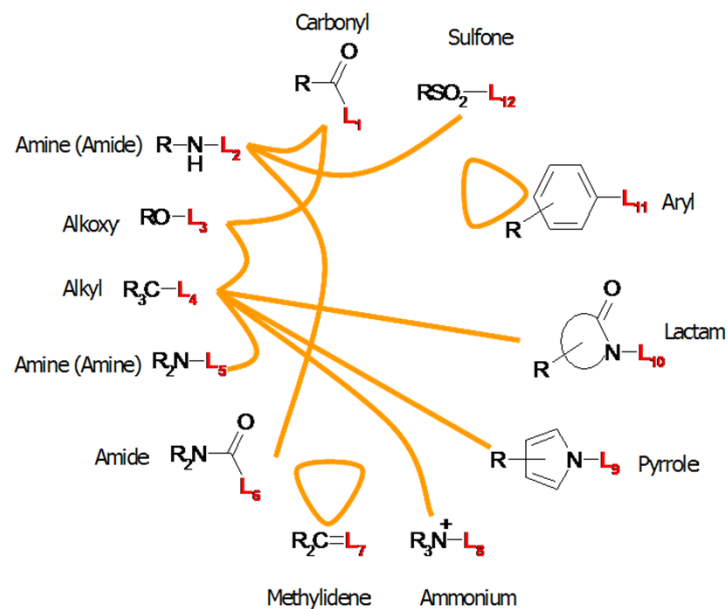
What makes a Fragment Space ?

- A fragment to us is a virtual molecule with link atoms



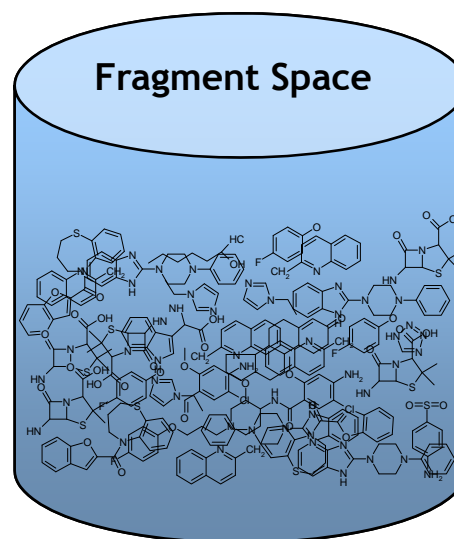
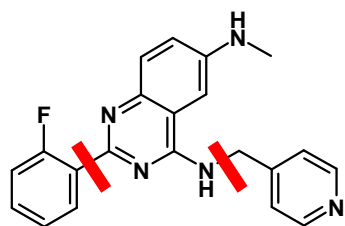
- Fragments can be connected

- fragments
+ link compatibility rules
= Fragment Space



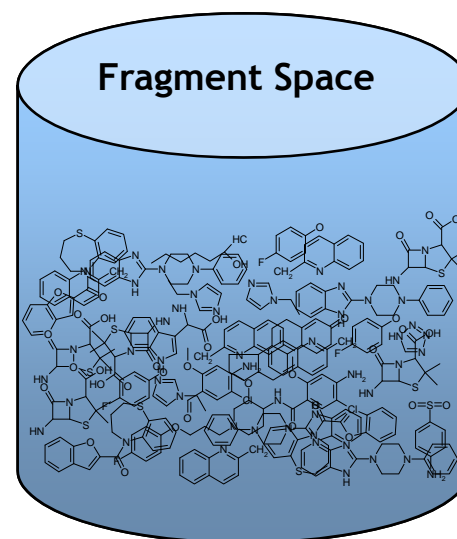
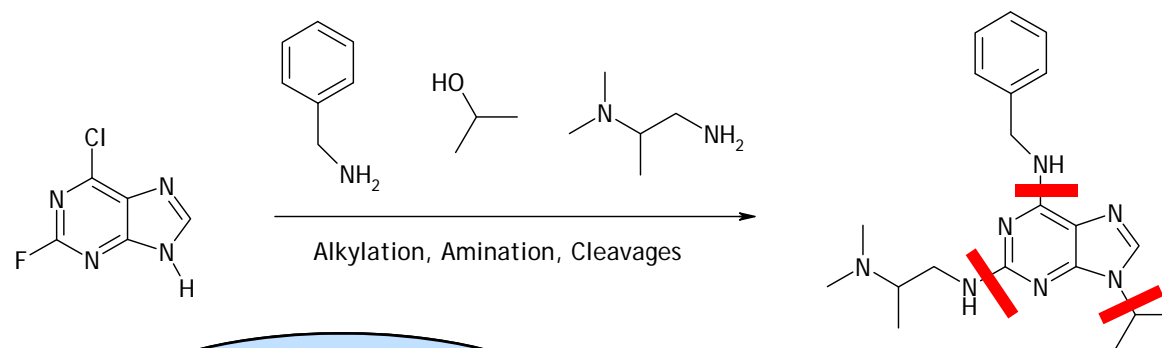
The Classification of Spaces

Shred Molecules:



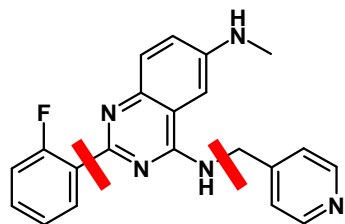
The Classification of Spaces

Exploit combinatorial chemistry:

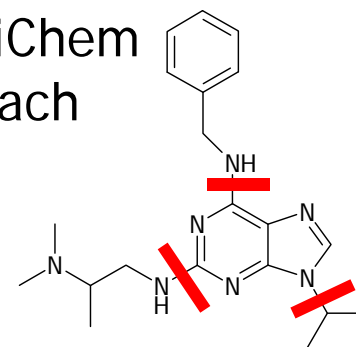


The Classification of Spaces

Shredding Approach:

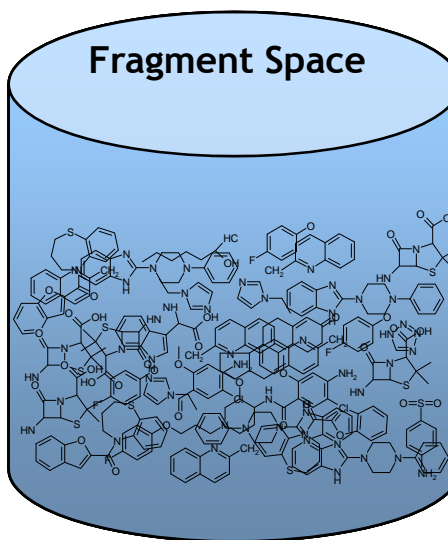


CombiChem Approach



Shredding

Fragment Space



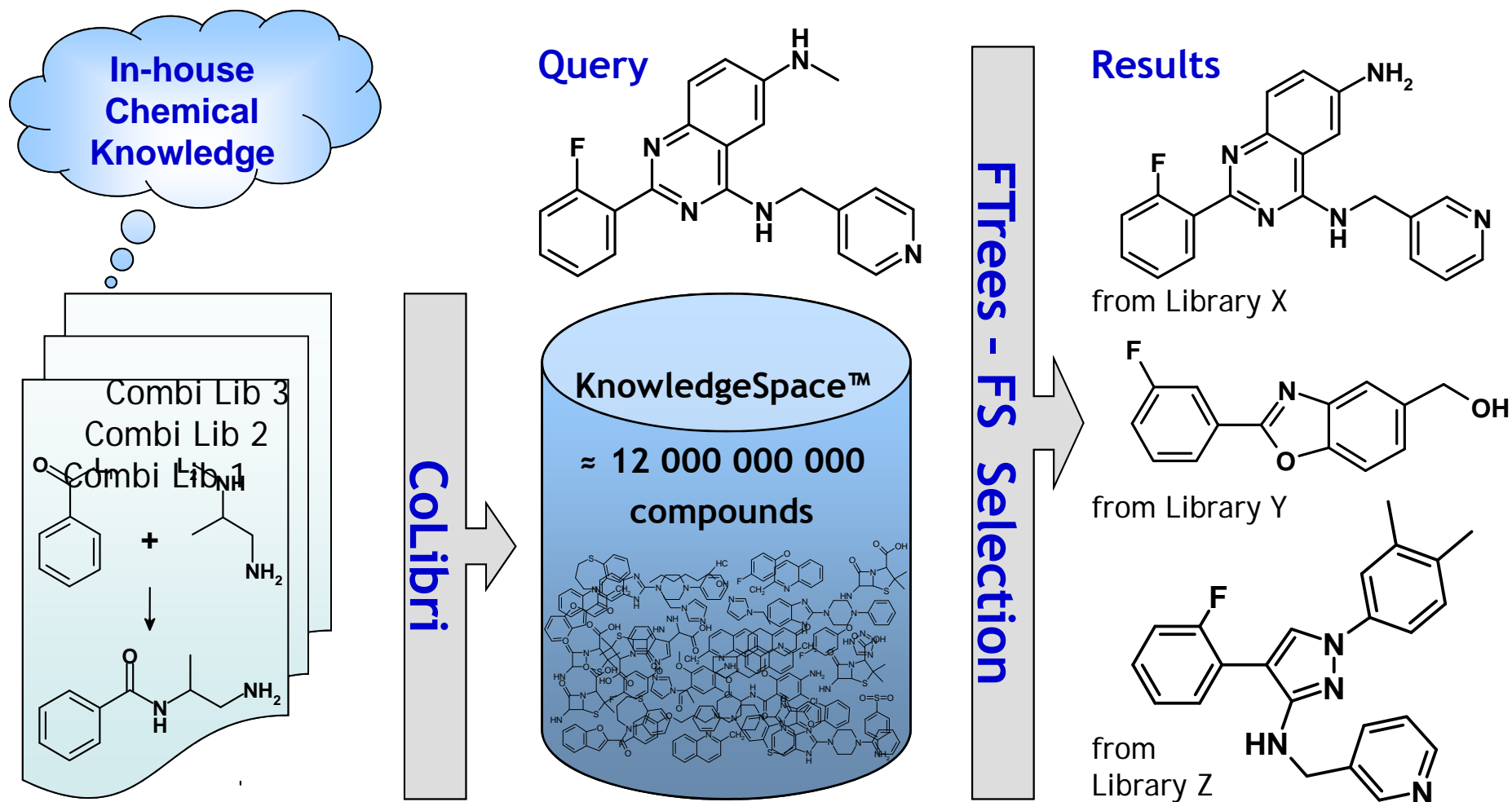
Clipping

Complete Compounds vs. Combinatorics

- Searching n **complete** ('enumerated') compounds leads to n possible compounds searched.
 - IP Gain = 0
 - Synthetic access: usually no problem.
 - Computation times are! (must touch every cpd!)

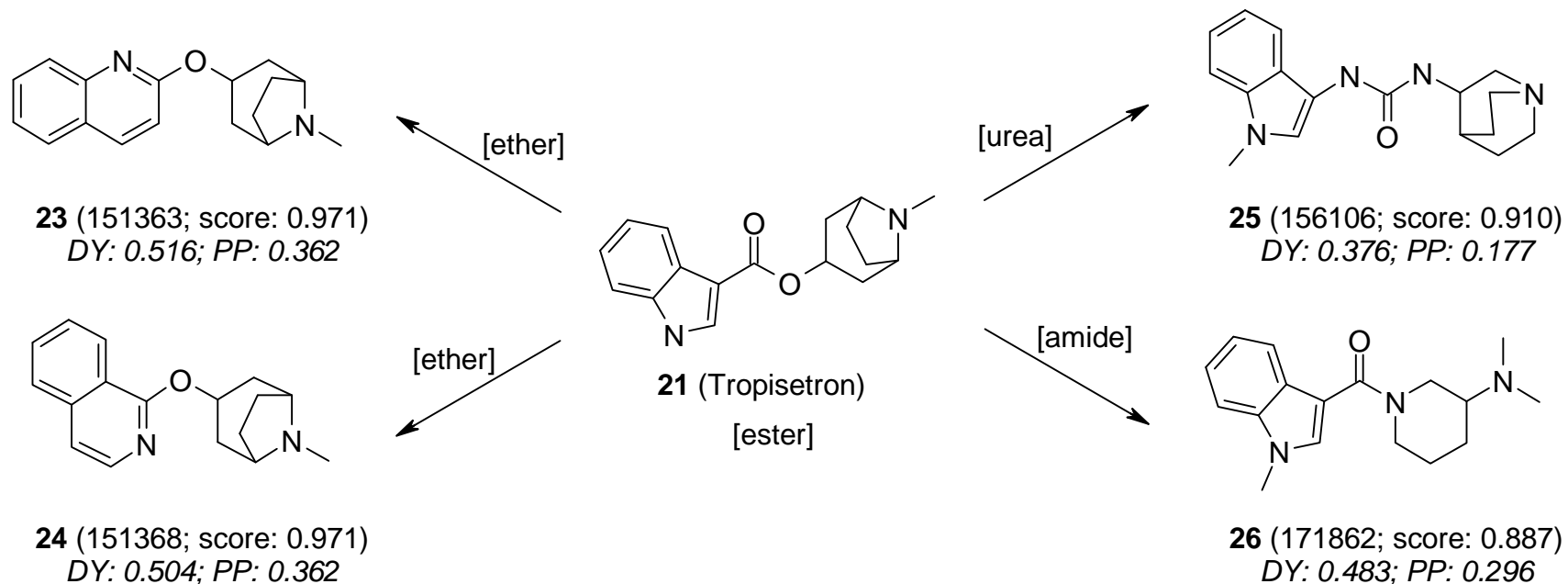
- Exploiting combinatorial chemistry surely can lead to many more compounds to search.
 - IP Gain much higher.
 - Synthetic access: must be taken care of!
 - Incomputable - unless clever approach used.

How is This Done In Real Life?



It Works: Pfizer's PGVL Space: 10^{12} Cpd

- Search for 5HT3 antagonists



Böhm et al., J. Med. Chem. **51**, 2468–2480, 2008

2. Search Technology

The *Similarity Principle* - umm... *Assumption*:

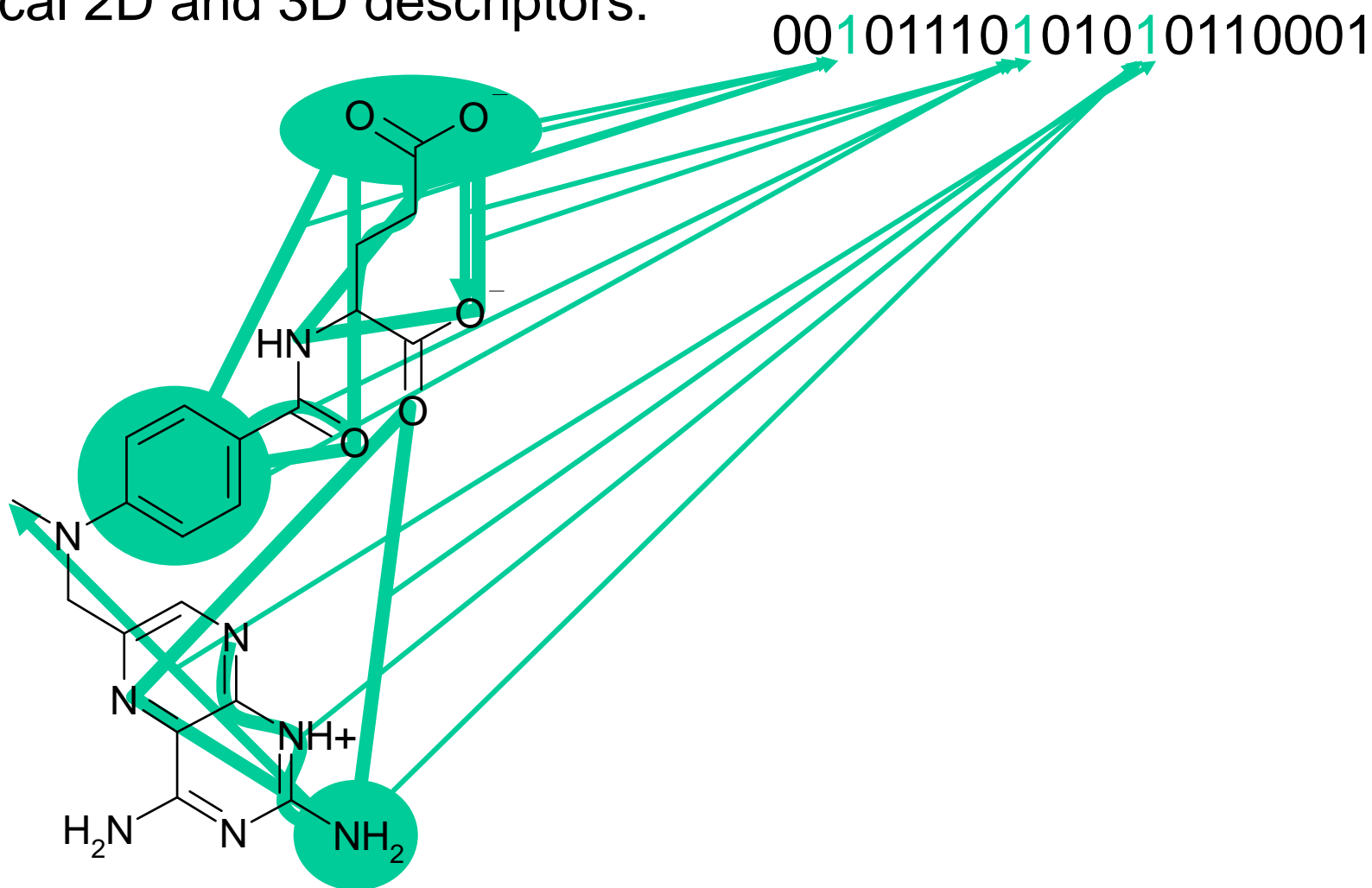
“What looks similar, behaves similarly.”



Of course, this is not always true,
but sometimes we cannot do better.

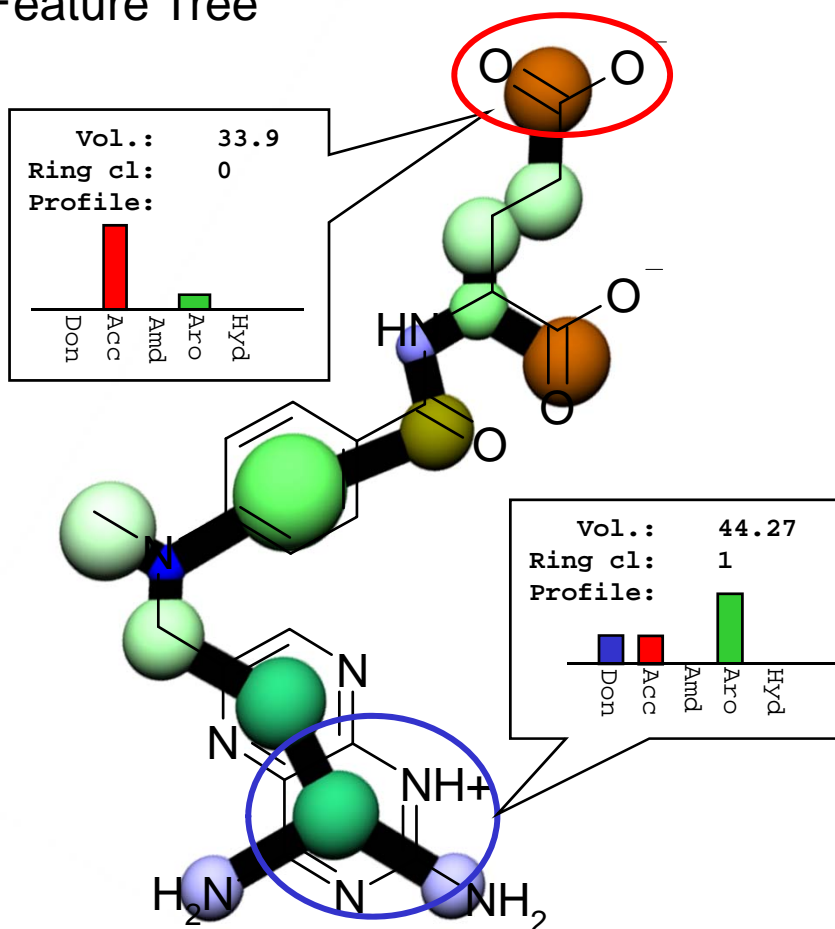
Similarity Descriptors: Bitstrings

typical 2D and 3D descriptors:



Similarity Descriptors: FTrees

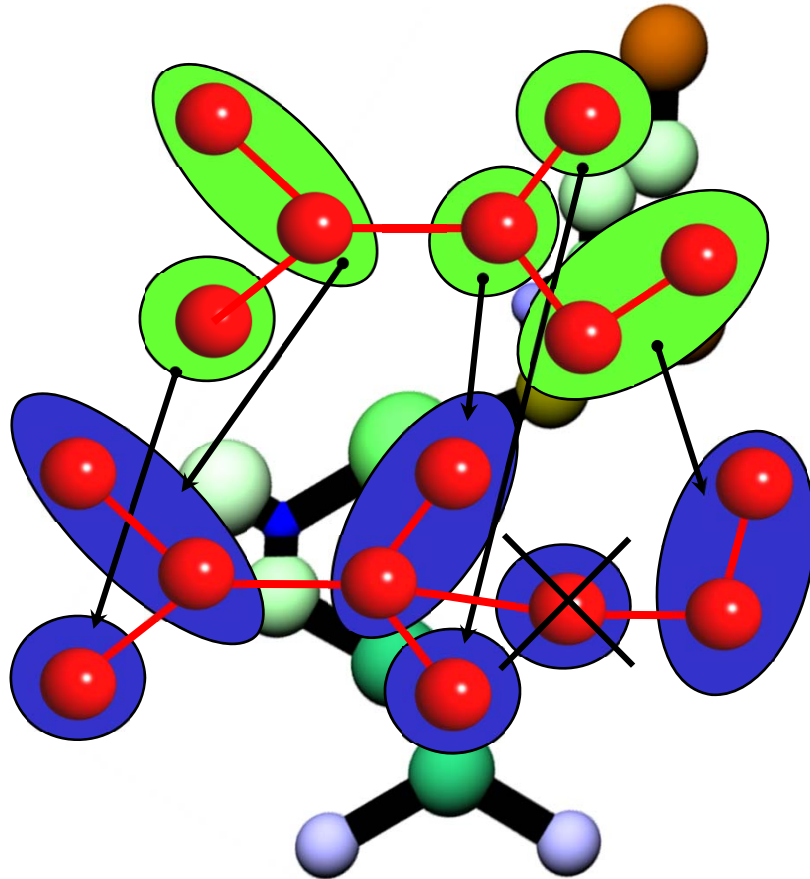
2½D FTrees descriptor:
Feature Tree



- Molecule represented as a tree
 - no bit string
 - no 3D conformers
- Nodes: chemistry/physics
- Topology preserved

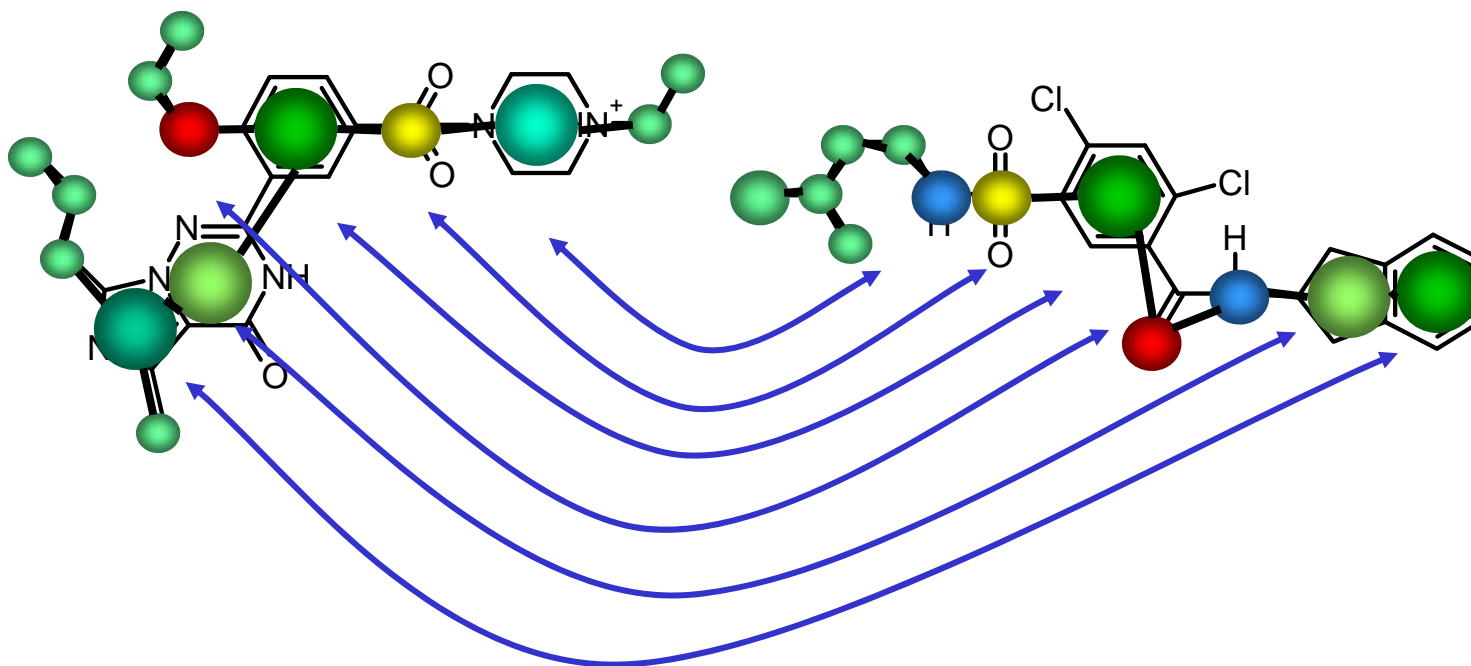
Original Idea, first implementation:
M.Rarey and S.Dixon, *JCAMD*, **12** (1998)

Advantages of FTrees



- more than just a similarity score
- fast like 2D, but retains topology
- fuzzy

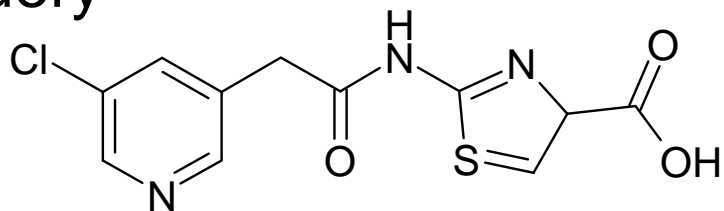
Principle of FTrees



- FTrees Similarity: 0.85
 - global for whole molecule
 - local for each match
- Mapping of substructures

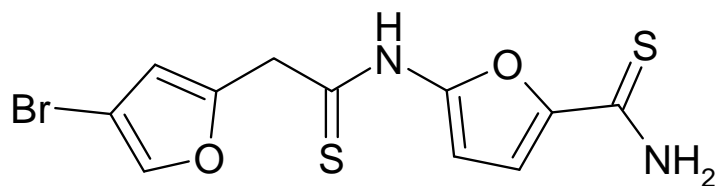
FTrees “Intuition”

query



D.R. Flower, *JCICS*, **38**, 379-386 (1998)

“On the properties of bit string-based measures of chemical similarity”



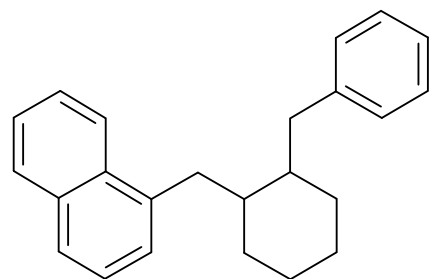
2D Fingerprints

(Unity)

0.203

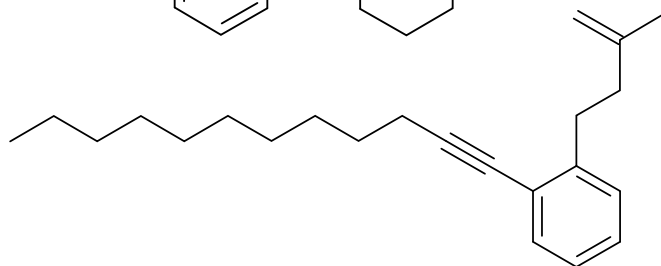
FTrees

0.764



0.203

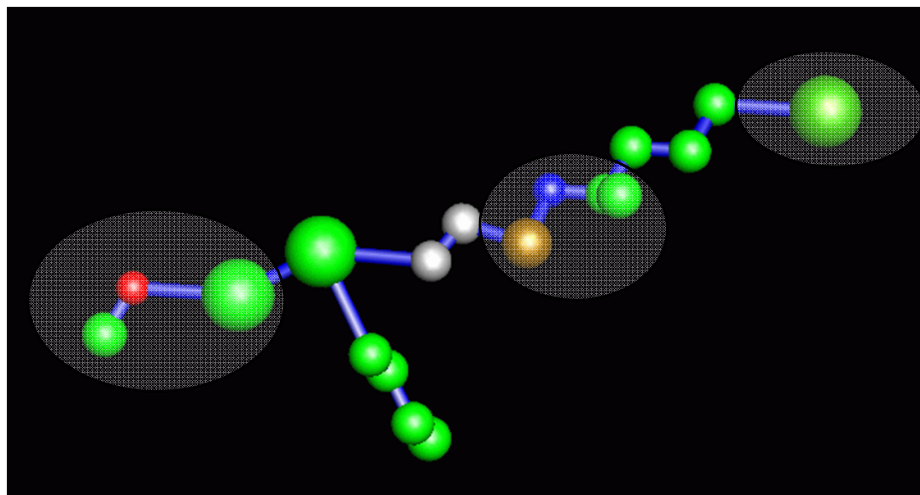
~~0.655~~



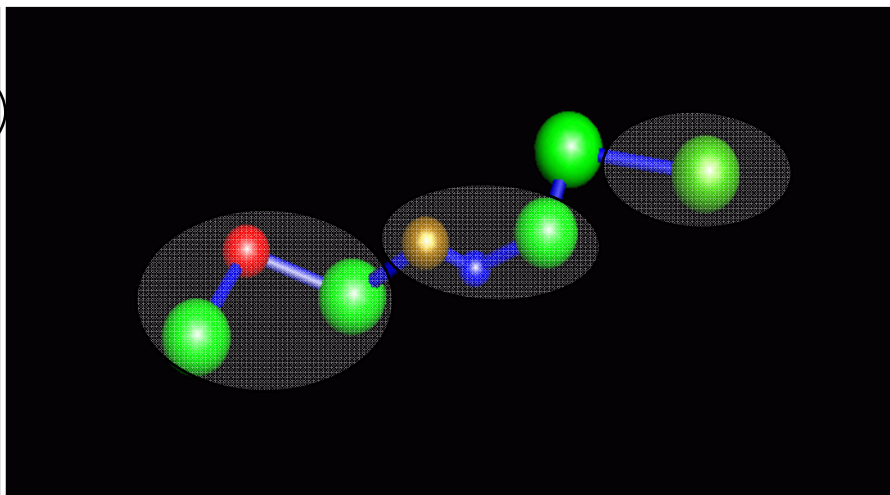
0.203

~~0.550~~

PAF Application by Boehringer-Ingelheim



query
PAF antagonist



hit
PAF antagonist

2D: rank 729 / 957
FTrees: rank 5 / 957

H.Briem, U.Lessel, *PD3*, 20, 231 (2000)

“In vitro and in silico affinity fingerprints: Finding similarities beyond structural classes”

GPCR Application by Sanofi-Aventis

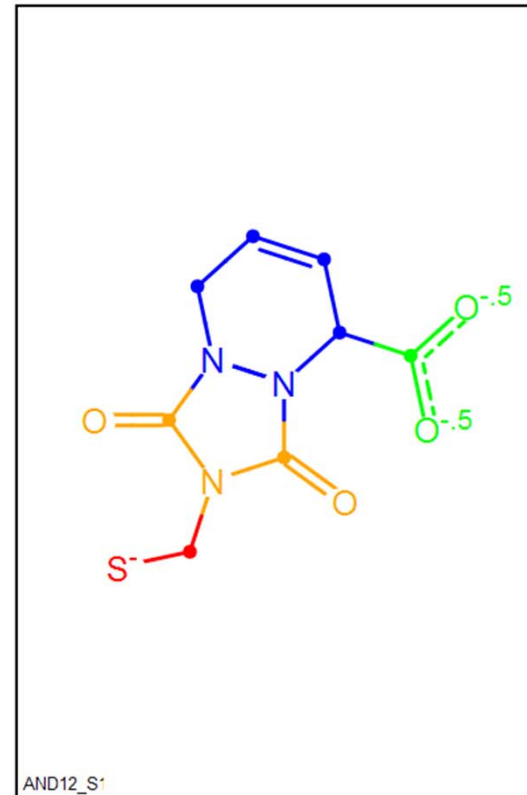
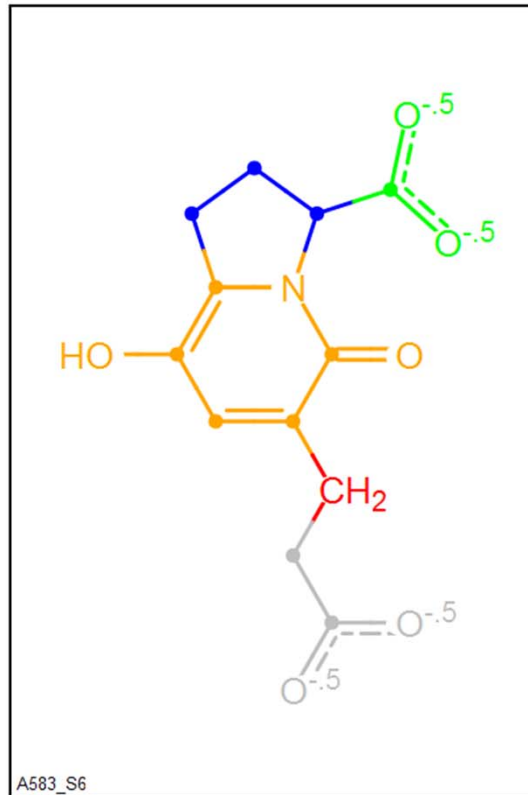
Table 2. Enrichment Factors and Hit Rates from Screening the Virtual Screening Library Using the Alpha1A, 5HT2A, D2, and M1 Feature Tree and Catalyst Models^a

% database screened		alpha1A		5HT2A		D2		M1	
		EF	hit rate [%]	EF	hit rate [%]	EF	hit rate [%]	EF	hit rate [%]
Feature Tree models									
1	class1 ^b	12.0	60	18.0	90	20.0	100	18.0	90
	class2 ^c	16.0	80	16.0	80	16.0	80	10.0	50
	winner ^d	12.0	60	20.0	100	20.0	100	20.0	100
5	class1	6.8	34	9.6	48	8.8	44	8.8	44
	class2	9.2	46	6.0	30	5.2	26	4.4	22
	winner	7.6	38	11.2	56	9.6	48	10.4	52
10	class1	4.8	24	6.8	34	5.0	25	5.0	25
	class2	5.4	27	4.2	21	4.2	21	3.2	16
	winner	5.8	29	8.0	40	6.2	31	5.8	29
Catalyst models									
1	class1	14.0	70	10.0	50	6.0	30	18.0	90
	class2	12.0	60	8.0	40	2.0	10	16.0	80
	winner	14.0	70	10.0	50	0.0	0	16.0	80
5	class1	6.4	32	8.4	42	4.8	24	9.2	46
	class2	10.0	50	8.4	42	4.0	20	7.6	38
	winner	10.4	52	8.8	44	4.0	20	9.2	46
10	class1	3.6	18	4.4	22	2.8	14	5.6	28
	class2	5.4	27	6.4	32	4.4	22	5.0	25
	winner	5.8	29	7.4	37	4.6	23	5.6	28

Evers et al., J.Med.Chem., 48, 5448-65 (2005)

Alignments Explain Similarities

Overall similarity: 0.797



Local similarities

Match 1: 0.895

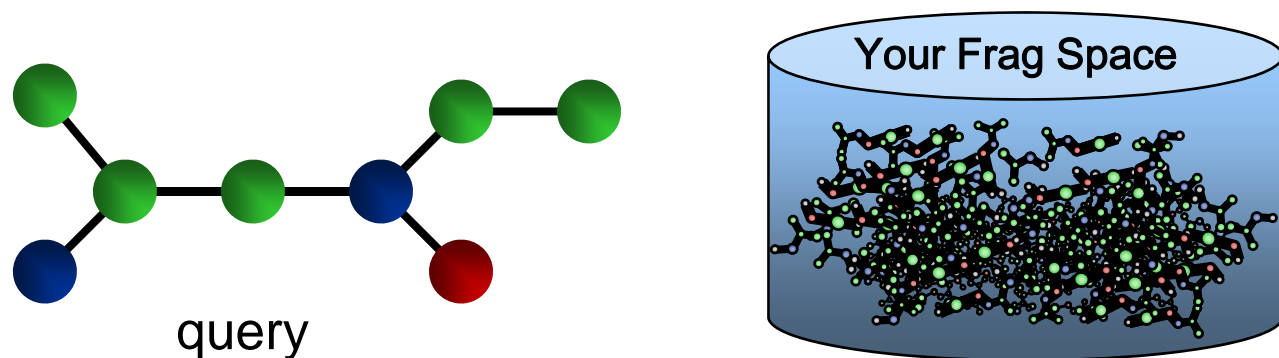
Match 2: 1.000

Match 3: 0.802

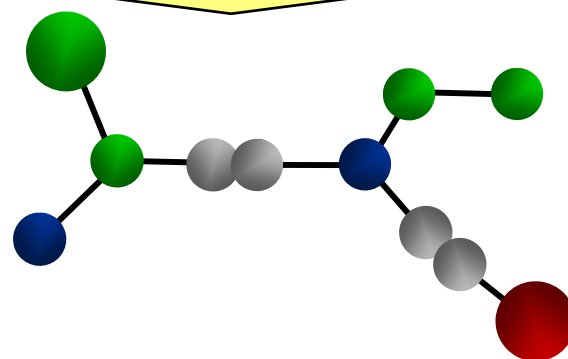
Match 4: 0.750

Not matched

The Actual Searching



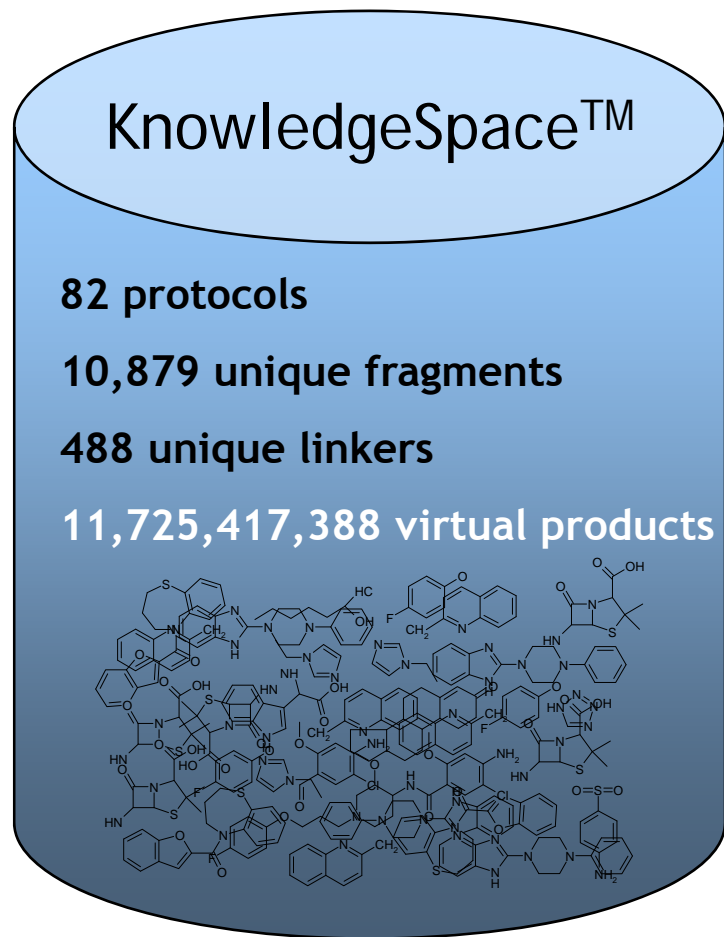
dynamic programming *



* deterministic:
- same result for every run
- always identify best mol

M.Rarey, M. Stahl, *JCAMD*, **15**, 479–520 (2001)

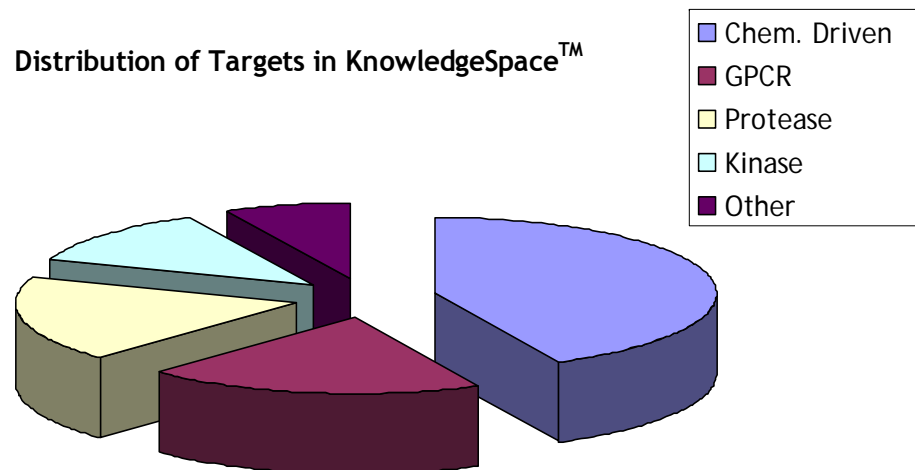
KnowledgeSpace™: a Public cLib-Space (10^{10})



<http://www.biosolveit.de/datasets>

J. Ambler et al., Bioorganic & Medicinal Chemistry Letters 9 (1999) 73
P. Tempest et al., Tetrahedron Letters 42 (2001) 4959--4962
B. Beck, Dissertation, TU München, Fakultät für Chemie, 2003
C. Kalinski, Dissertation, TU München, Fakultät für Chemie, 2006
A.M. Boldi et al., Tetrahedron Letters 40 (1999) 619-622

Distribution of Targets in KnowledgeSpace™



I.R Hardcastle et al., Bioorganic & Medicinal Chemistry Letters 15(2000) 1173
E. Farrant et al., Tetrahedron Letters 41 (2000) 5383-5386
D. Fancelli et al., J. Med. Chem. 2005, 48, 3080-3084
O. de Frutos et al., J. Comb. Chem. 2000, 2, 639-649
C.Gil et al., Chem. Eur. J. 2005, 11, 2680 -- 2688
S.Gray et al., SCIENCE VOL 281 24 JULY 1998
J.P. Killburn et al., Tetrahedron Letters 41 (2000) 5419-5421
R. Kuang et al., J. Am. Chem. Soc. 1999, 121, 8128-8129
S.H. Lee et al., Tetrahedron Letters 39 (1998) 9469-9472
C.E. Lee et al., J. Am. Chem. Soc. 1998, 120, 9735
S.V. Ley et al., J. Comb. Chem. 2000, 2, 104
Lu et al., J. Comb. Chem. 2000.

KnowledgeSpace™ Is Publicly Available:

- www.biosolveit.com/KnowledgeSpace

The screenshot shows the BioSolveIT website interface. The header includes the BioSolveIT logo and the tagline "The Premier Scientific Solution Provider". A navigation menu on the left lists: HOME, SOFTWARE, SUPPORT, SERVICES (newsletter, workshops, conferences, publications, datasets), DOWNLOAD, and COMPANY. The main content area is titled "Datasets" and features a pie chart comparing an "x million in-house library" (commercially available) to "x.000.000.000.000 synthetically accessible compounds". Below the chart, text states: "No matter how big your in-house library and no matter how many compounds you acquire to add to it, it will only be a tiny fraction of what your chemists are capable of synthesizing." A section titled "KnowledgeSpace" highlights "Fragments + Reactions = 23,000,000,000 Synthesizable(!) Compounds" and provides details about the search capabilities and the composition of the dataset.

BioSolveIT
The Premier Scientific Solution Provider

contact impressum legal notice search print this page

HOME
SOFTWARE
SUPPORT
SERVICES
newsletter
workshops
conferences
publications
datasets
DOWNLOAD
COMPANY

Workshop 2009
Noordwijkerhout
The Netherlands
May 8, 2009

10²⁰ virtual compounds here

Datasets

x million in-house library
commercially available

x.000.000.000.000 synthetically accessible compounds

No matter how big your in-house library and no matter how many compounds you acquire to add to it, it will only be a tiny fraction of what your chemists are capable of synthesizing.

KnowledgeSpace

Fragments + Reactions = 23,000,000,000 Synthesizable(!) Compounds

Now you have the possibility of searching incredibly large numbers of molecules in a sensible way: **Compounds you find can actually be made!** Search our new, free *KnowledgeSpace* – within just minutes!

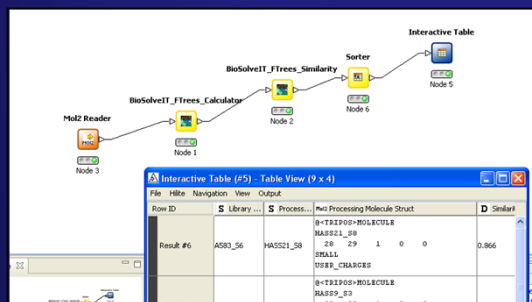
The *KnowledgeSpace* (Version 1.0) has been compiled based on 82 synthesis protocols from the literature. Some are target-specific (GPCRs, proteases, kinases, ...), and some are purely chemistry-driven. Altogether they cover **17,794** unique fragments and comprise about 23,000,000,000 virtual products.

The *KnowledgeSpace* has first been tested to retrieve known drugs. See the details in a recent [presentation](#).

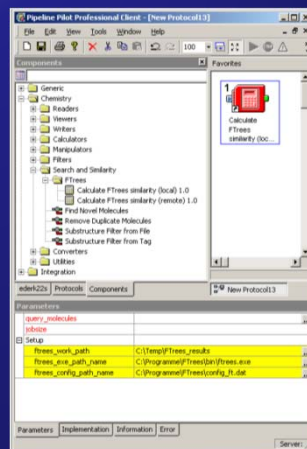
The *KnowledgeSpace* is ready to use for *FTrees-FS* and *FlexNovo* searches. It can be easily uploaded into related Pipeline Pilot components.

FTrees: Interfaces and GUIs

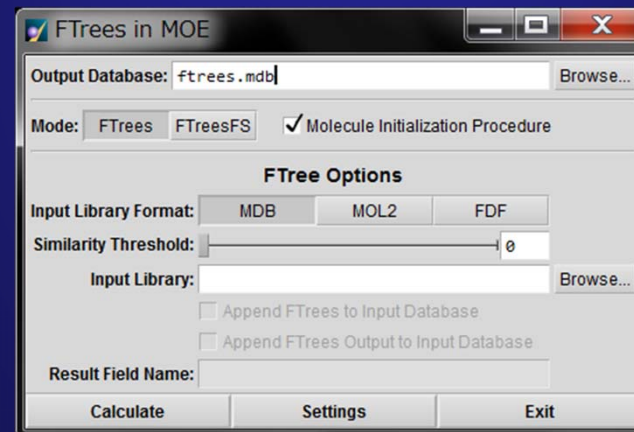
Pipelining Platforms:



KNIME®



PipelinePilot®

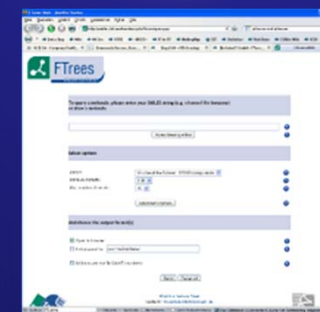


All possibilities: MOE®

The screenshot shows the mcule.com website with the following content:

- Navigation: Home, Pricing, Docs, Sign in, Sign up
- Payment period: Month, Academic pricing
- Cart: (0) 0 USD
- Product: **FTrees Visual Similarities** (BioSolveIT)
- Description: FTrees Visual Similarities is a highly efficient tool for scaffold hopping and ligand-based virtual screening. Its underlying topological descriptor (the Feature Tree) captures connectivity and physico-chemical properties of functional groups. The reduced graph representation that preserves the pharmacophore characteristics of the ligands in a fuzzy way enables the identification of novel scaffolds. Moreover, the topological Feature Tree descriptor makes similarity assessment extremely fast compared to 3D approaches and is also exempt from the uncertainties of 3D coordinate calculation.
- Link: <http://www.biosolveit.de/>

Pay-per-Use including visualization:
mcule.com



Free & simple web interface:
biosolveit.de/FTreesWeb

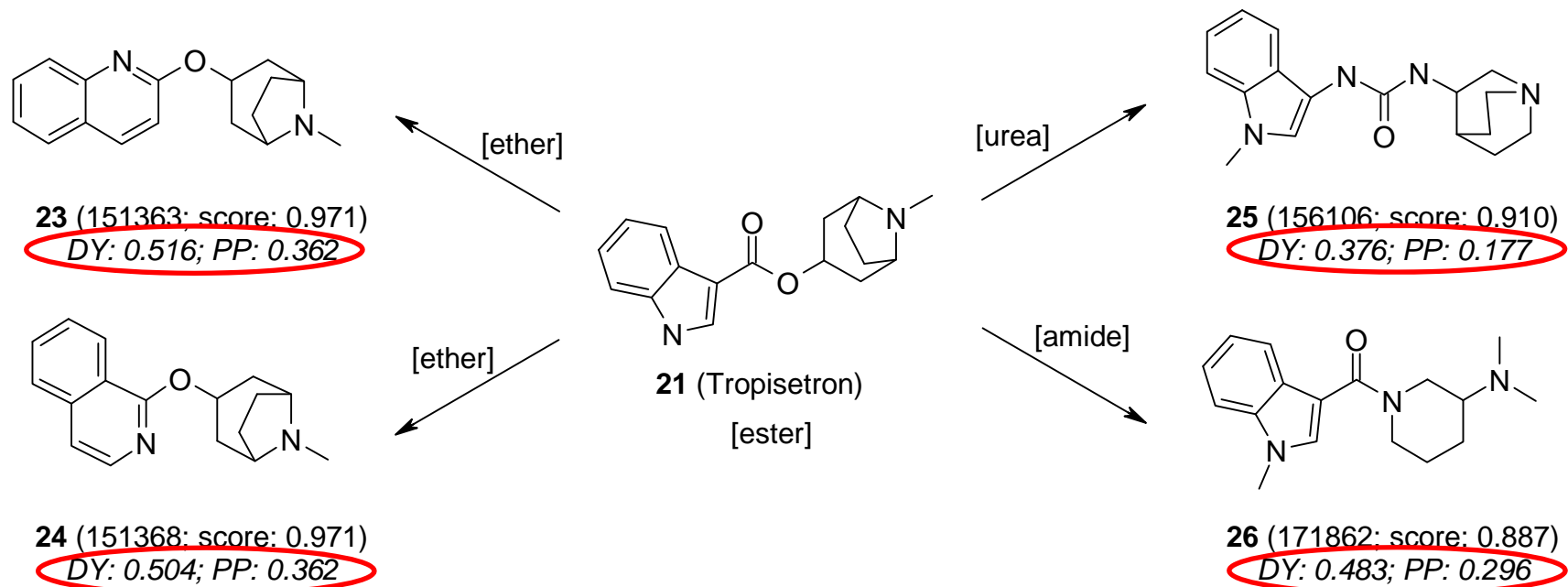
FTrees Successes - A First Summary

- Searches synthetically accessible molecules
 - reports the reaction alongside
 - covers 10^{12} possibilities in 5 minutes
- Highly efficient
 - 250.000 molecule:molecule comparisons per minute
 - high enrichment ratios
- Helps with various drug design tasks
 - scaffold hopping
 - virtual screening
 - open chain vs. ring structures are found

3. Four Eyes May See More Than Two...

It Works: Pfizer's PGVL Space: 10^{12} Cpd

- Search for 5HT3 antagonists



Böhm et al., J. Med. Chem. **51**, 2468–2480, 2008

Highly Effective: An Orthogonal View!



What Complements the FTrees-“View”

- FTrees do not “see” stereochemistry.
- FTrees cannot distinguish ortho-, meta-, para.
- FTrees make one “blob” per cycle.
- FTrees are conformation independent.

=> Let us use 3D then!

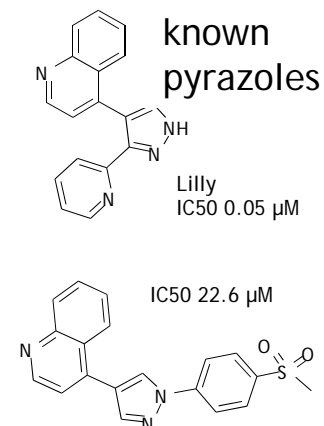
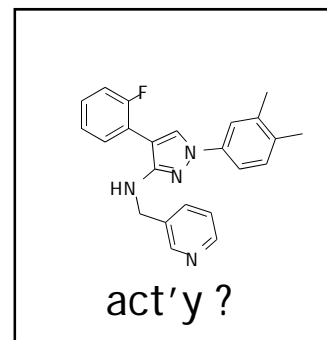
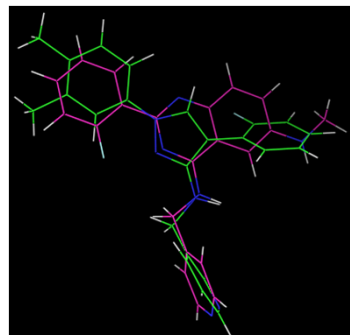
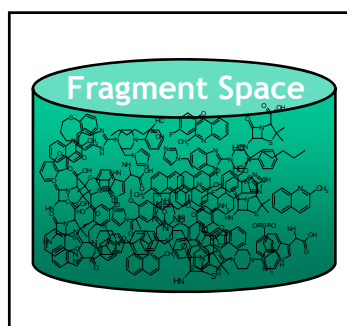
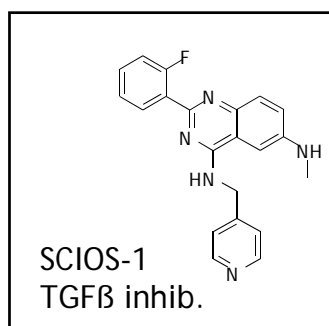
Prospectively: Boehringer's BI-CLAIM (10¹¹)

Search Space (BI-CLAIM):

1,600 scaffolds + 30,000 reagents -> 500,000,000,000 virtual products

A typical workflow (part 1):

lit query ➡ 1000s hits ➡ shape filter ➡ vis. inspect.



The typical workflow (part 2)

select scaffolds ➡ design focused lib / purchase prototypes ➡ refinement

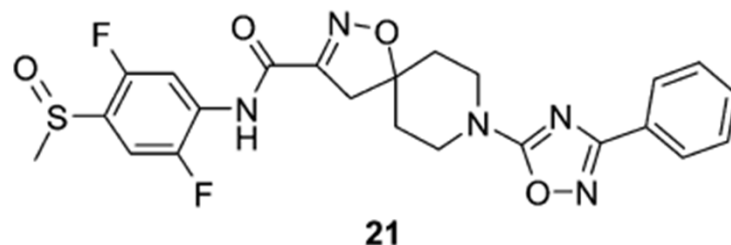
Rapid success stories

- GPCR: 1000 virtual hits, 2 focused libs, 100 nM inhibitors identified
- Proteinase: 1200 compd screened, 2 active scaffolds, refined to 10 nM

Lessel et al., ICCS'08 & J Chem Inf Model. 2009 Feb;49(2):270-9.

Boehringer Ingelheim J.Med.Chem. ASAP

This just appeared online:



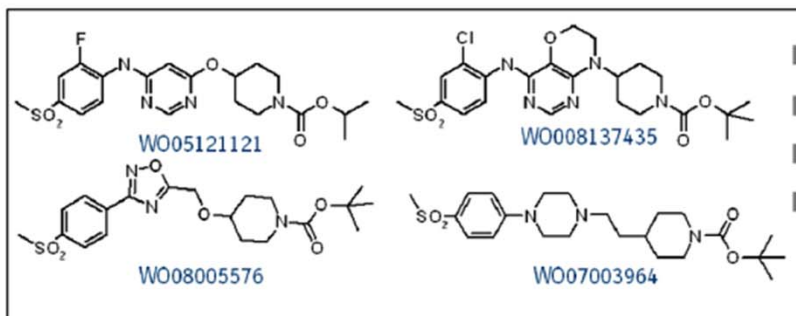
21 +/- 15nM activity
against GPR119 (a GPCR relevant for diabetes)

Wellenzohn et al., J.Med.Chem. 2012 asap:

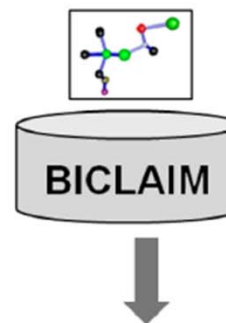
"Identification of new Potent GPR119 Agonists by Combining Virtual Screening and Combinatorial Chemistry

The Search Cascade Used (taken from JMC)

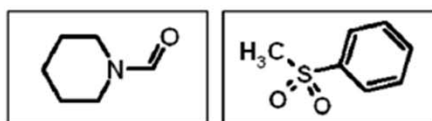
1.) Definition of FTrees queries



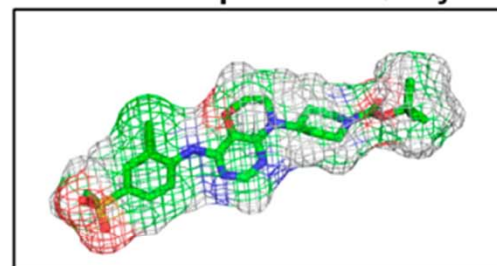
2.) Individual FTrees search in the BICLAIM fragment space



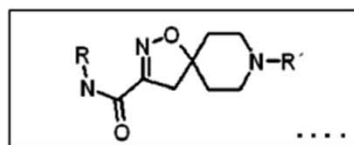
4.) Filtering by substructures and definition of activity anchor(s)



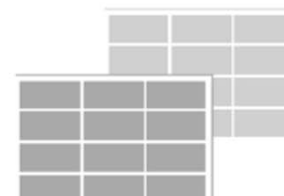
3.) Merging of FTrees results and generation of ROCS overlap with 3D Query



5.) Visual inspection of the remaining ROCS overlaps and selection of the cores



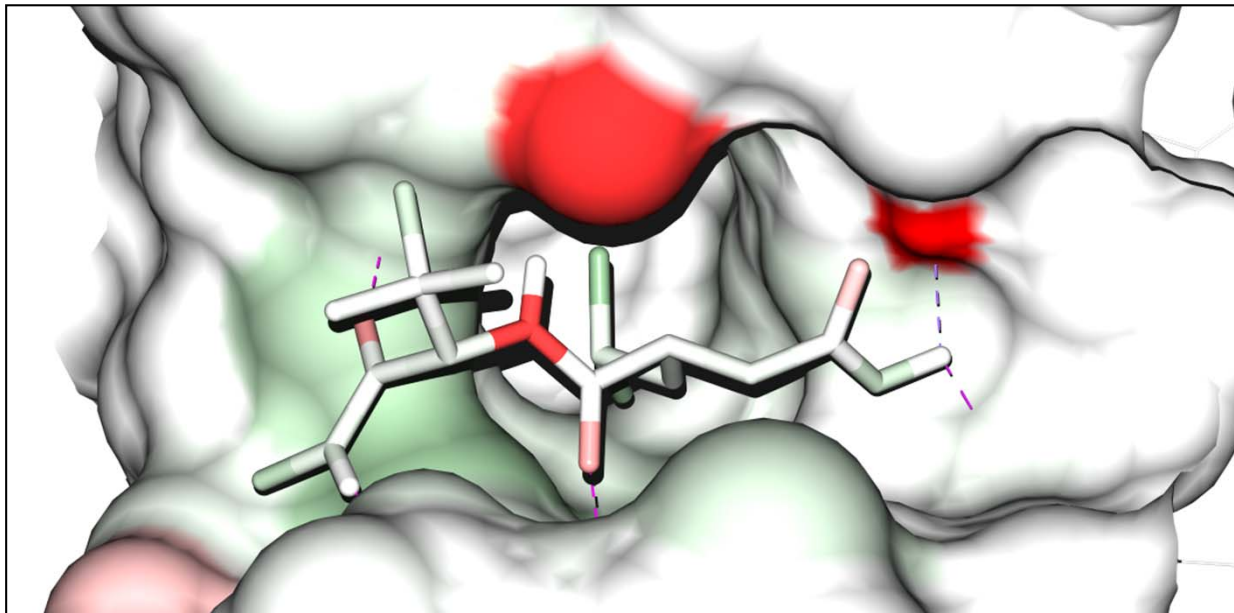
6.) Design and synthesis of combinatorial libraries of the selected core(s)



Another, New Orthogonal Method: HYDE

If you have a protein structure...

- Visual affinity computation using HYDE
- Hydrophobic effect and H-bonds (etc.) balanced intrinsically



PDB: 1GKC

Methods described in a paper JUST out:

doi: 10.1007/s10822-012-9626-2 Schneider et al., JCAMD Dec2012

So... Orthogonal Views Help a Great Lot



Summary

1. Trillions of compounds can be searched
 2. High likelihood of synthetic access / proven success in Pharma
 3. Combine 2D and 3D
-
- CAVEAT:
Searching your own chemistry involves yourself!

Happy FTrees Users:

Pfizer, Novartis, AstraZeneca, Merck USA/NL, Merck-Serono, Hoffmann-LaRoche, Johnson&Johnson, Sanofi-Aventis, Bayer, Boehringer-Ingelheim, Dupont, and many more