

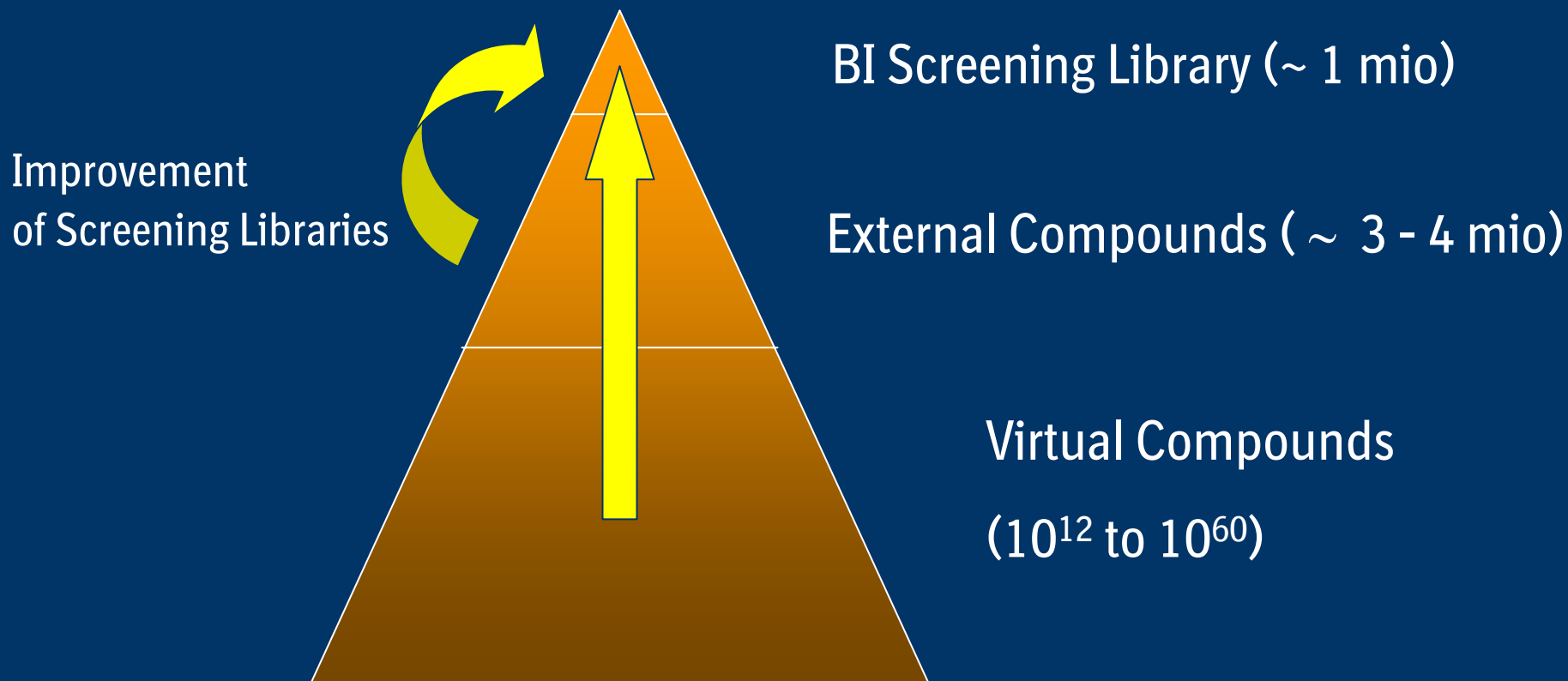
8th International Conference on Chemical Structures

Searching Fragment Spaces with Feature Trees

Uta Lessel

Boehringer Ingelheim Pharma GmbH & Co. KG

The Chemical Universe



How can we exploit this huge space?

Different Approaches

Diverse methods for De novo Design

Recombination of fragments generated according to RECAP rules

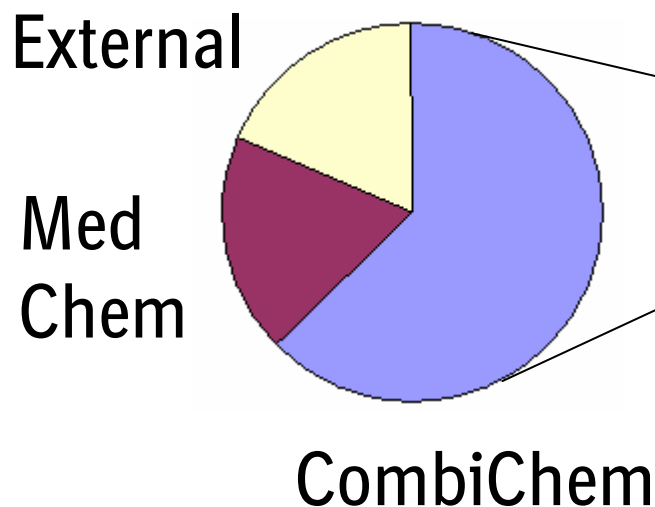
...

but

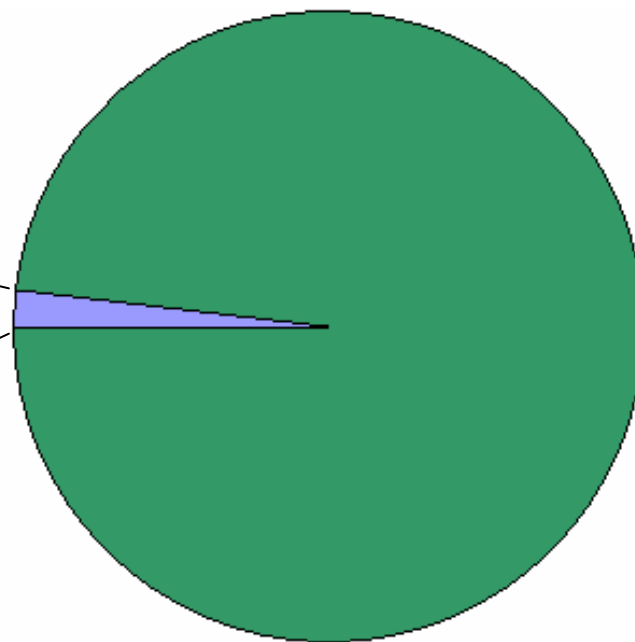
lacking chemical feasibility of most products is a severe disadvantage

BI CLAIM

BI Screening Pool



BI CLAIM



per
scaffold:

1,000 - 10,000
cpds. realized

millions of
virtual cpds.

Collection of Compounds

Considering all compounds which might be synthesized by combinatorial chemistry leads to chemistry spaces with about 10^{12} compounds or even more.

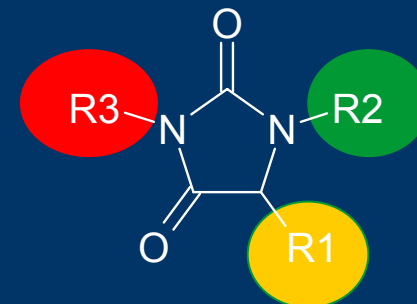
Consequences:

- **Storage:** special formats to store all the molecules
- **Access:** special virtual screening techniques which can cope with these special formats
- **Time:** intelligent screening techniques avoiding to screen the huge space completely

Fragment Spaces

General procedure

- split compounds into fragments
- define rules to combine them



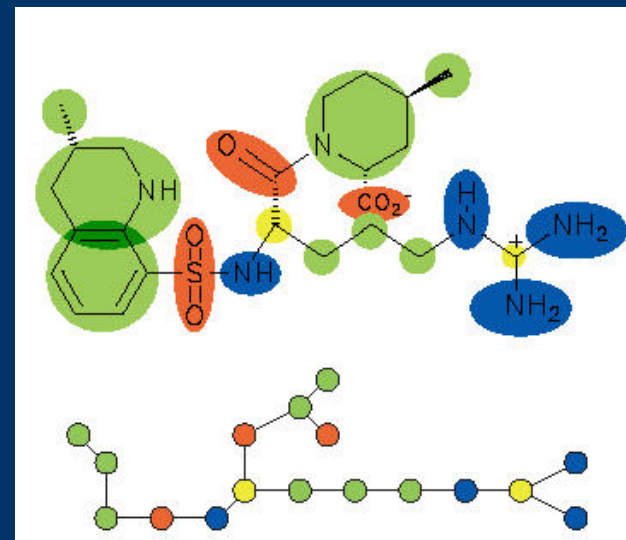
This way the amount of data to be stored is reduced to a manageable number of reagents and reactions.

One possible way to search Fragment Spaces:
Feature Trees – Fragment Spaces (FTrees-FS)

M. Rarey, M. Stahl, J. Comp.-Aided Molecular Design, 15, 497-520, 2001

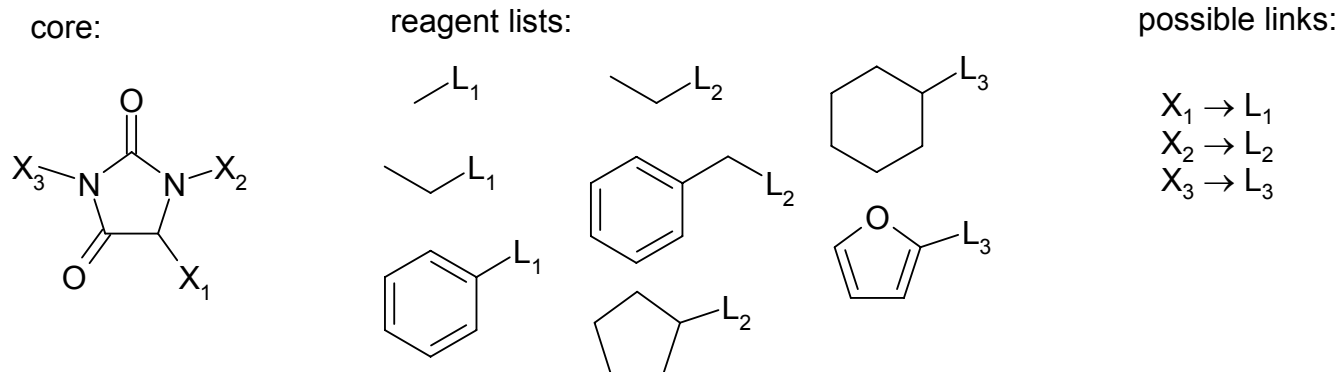
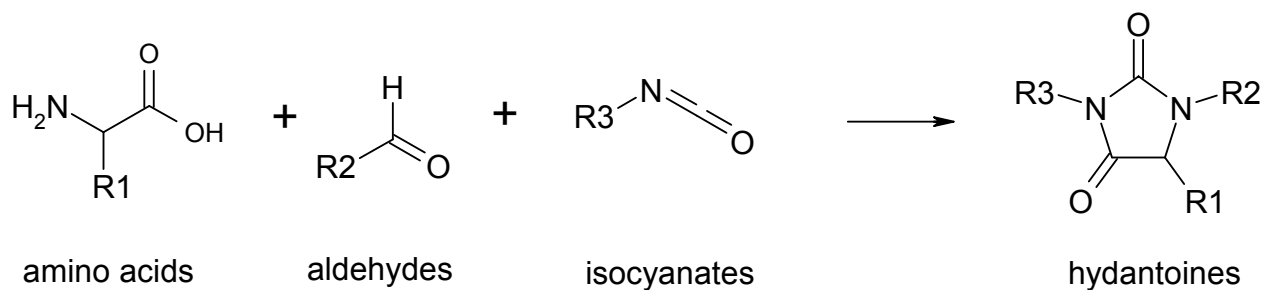
Introduction into FTrees

- First time described in 1998 by Rarey and Dixon
 (*Journal of Computer Aided Molecular Design* 12; 471-490, 1998)
- Method for fast similarity searches
- FTrees represent molecules as trees.
- The nodes contain the chemical and steric information of the corresponding fragments.
- The properties of the nodes are additive.
- The descriptor has already shown its potential in virtual screening.
 (*Journal of Medicinal Chemistry* 48; 5448-5465, 2005)
- The method allows extensive lead hopping

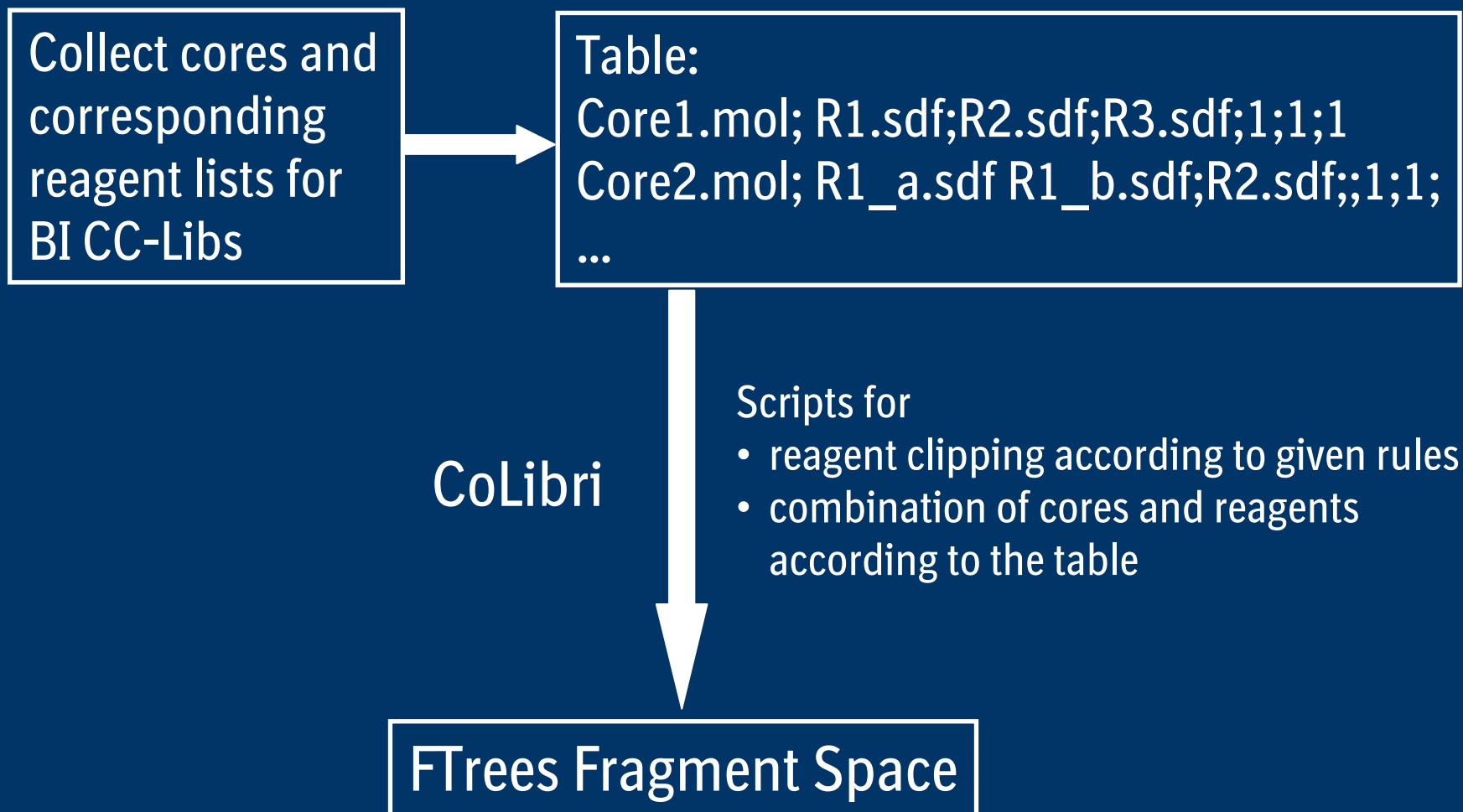


Example of a Feature Tree

Generation of Fragment Spaces



BI CLAIM and CoLibri



Current BI CLAIM Fragment Space

Today BI CLAIM contains:

- ~ 1,600 scaffolds
- ~ 30,000 reagents

encoding about 500,000,000,000 compounds

Creation of this fragment space with CoLibri
takes about 8 h (single Intel Xeon CPU 3.2 GHz)

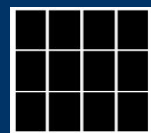
FTrees-FS searches can be performed in a few minutes
depending on the size of the query molecules.

Basic Application Concept

Large space of potentially
interesting molecules

FTrees-FS search
Filter e.g. FT_sim ≥ 0.85

Space reduced
to a manageable size



Use the resulting cpds.
for conventional
Virtual Screening

Add matrix format
Design focused
Combi-Libs

Idea generator

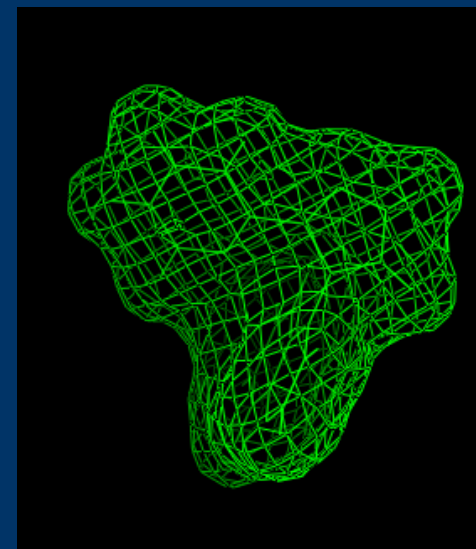
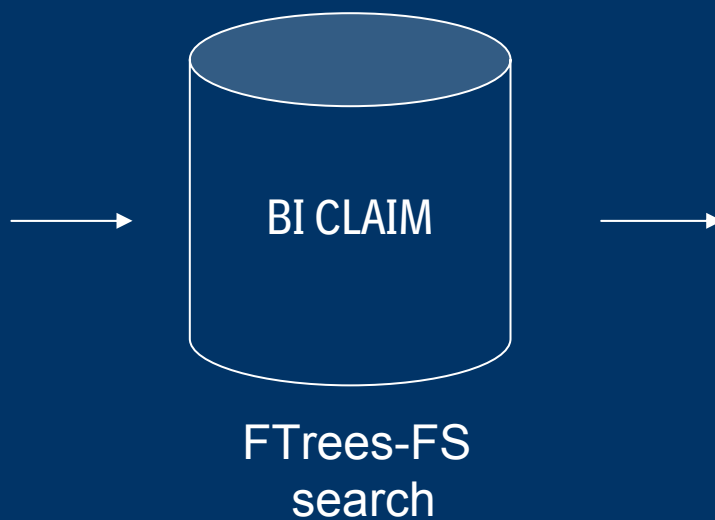
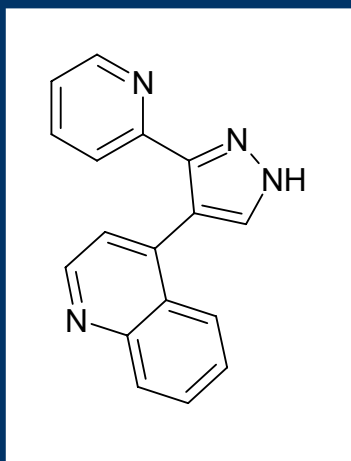
10^{12}

10^5

10^2

Typical Workflow @BI

Query, e.g.:

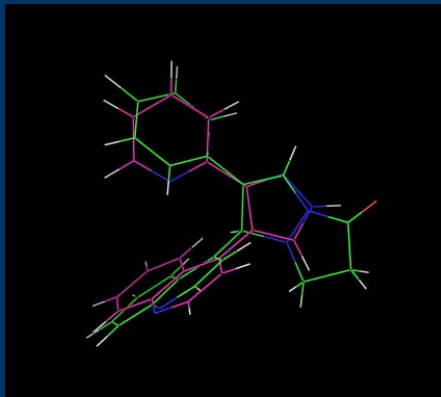


Post-processing
(e.g. 3D SHAPE Filter)

Visual inspection
of the results

Discussion with Combinatorial Chemists

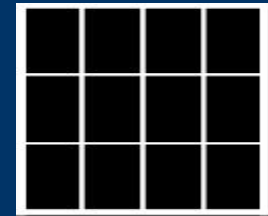
Selected scaffolds:



show 3D
alignment

collect results from
similar existing
compounds

selection of
interesting
scaffolds:



design and synthesis
of focused
combinatorial library

selection and
synthesis or purchase
of prototypes

Case Study TGFB β

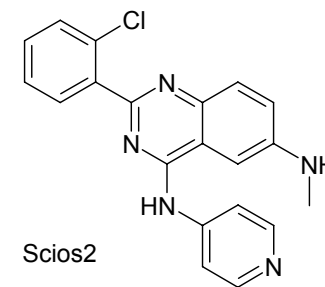
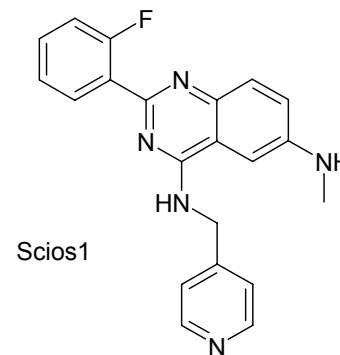
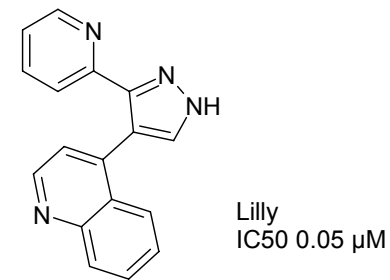
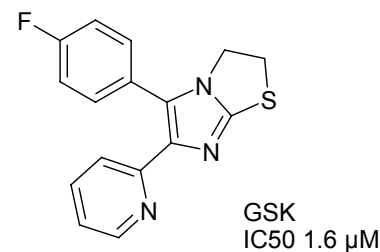
Knowledge:

739,605 compounds
tested in HTS

6,343 compounds with
 $IC_{50} < 100 \mu M$

Hit rate: 0.9%

Queries for searches in BI CLAIM:

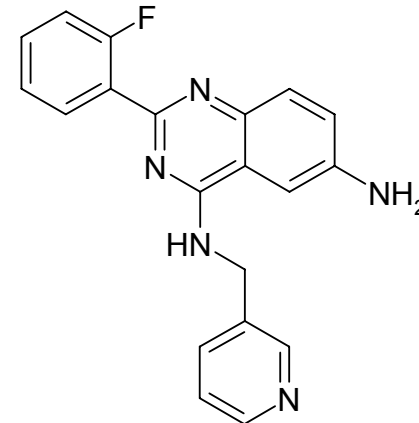
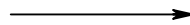
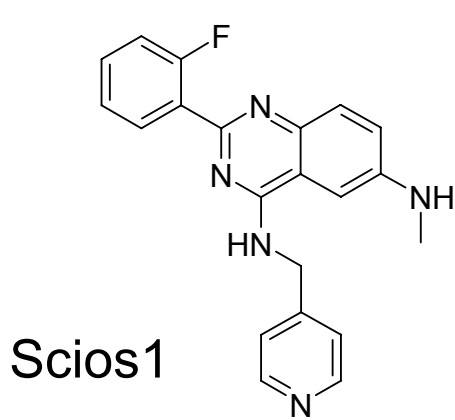


Case Study TGFB – Numeric Results

Query	# hits	# known activity	# actives	# unknown activity
GSK	7,653	15	none	7,638
Lilly	8,538	15	1 (34.4 μ M)	8,523
Scios1	25,005	18	1 (5.8 μ M)	24,987
Scios2	20,733	58	none	20,675

Enrichment of actives can not be assessed!

Case Study TGFB - Virtual Hits

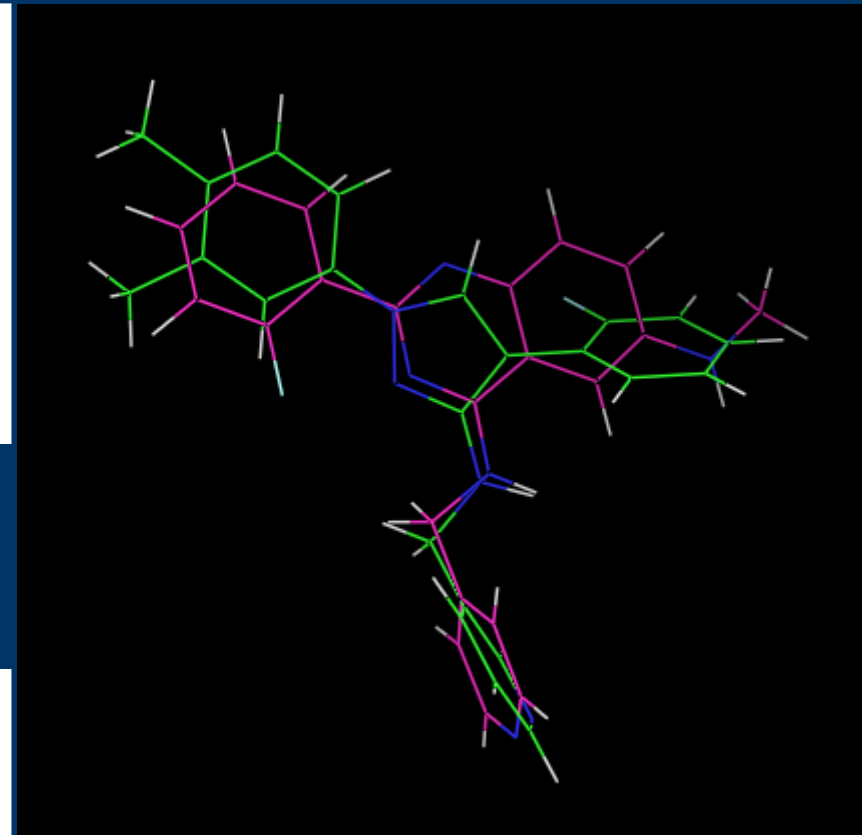
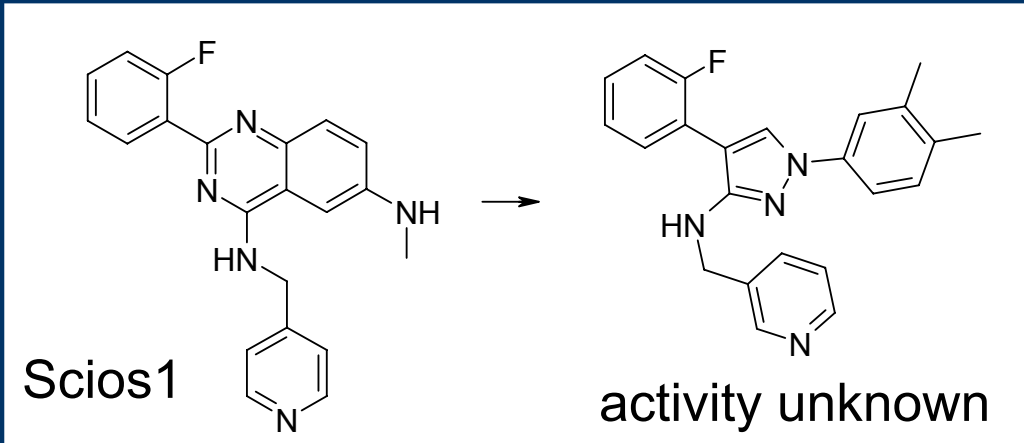


query

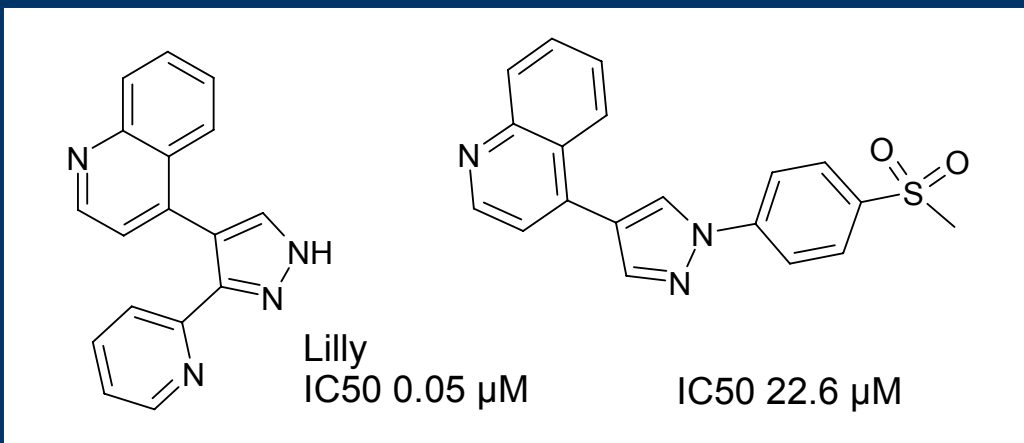
virtual hit

Trivial result, any program should find it!

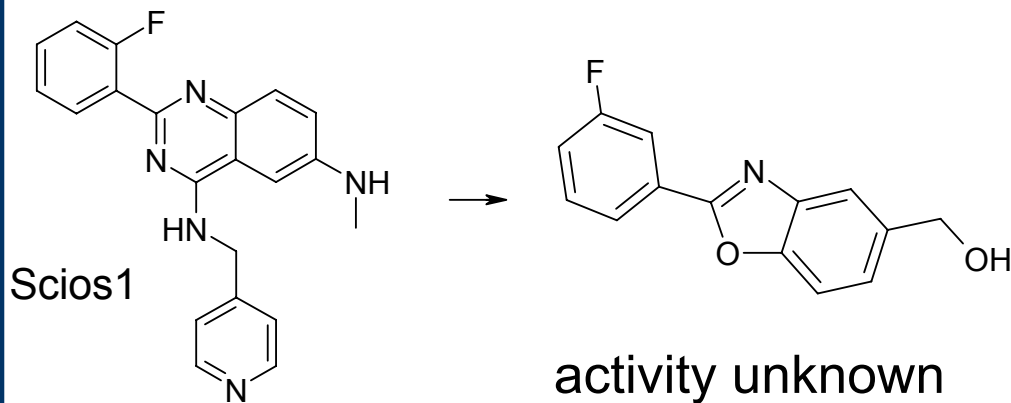
Case Study TGFβ - Non-Trivial Virtual Hits (I)



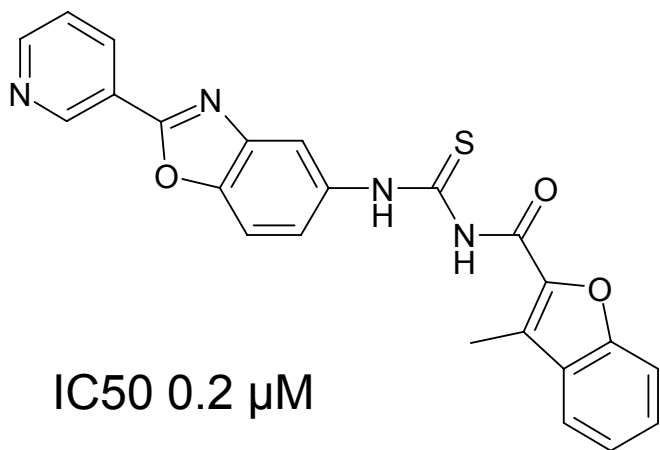
Known Pyrazoles:



Case Study TGFβ - Non-Trivial Virtual Hits (II)



Known Benzoxazole:



Two application examples of BI CLAIM (I)

GPCR-Project:

- An FTrees-FS search based on a literature compound yielded some 1,000 hits.
- These hits were analyzed with ROCS in a post-filtering step regarding their shape similarity to the query structures.
- In a manual selection step (“chemist’s eye”) two compound classes were selected for libraries with a few hundred compounds each.
- One of these libraries provided hits in the 100 nM range.

Two application examples of BI CLAIM (II)

Proteinase project:

- Literature known actives served as query structures for an FTrees-FS search.
- The initial BI CLAIM pre-screening library of 1,200 cpds. from 10 scaffolds yielded hits with $IC_{50} > 10 \mu M$.
- A first iteration (synthesis of ~ 30 compounds) yielded two active scaffolds with IC_{50} values of 5 – 10 μM .
- A cocrystallized compound gave a hint for the optimization of a sidechain orientation.
- A second and third iteration yielded compounds in the 10 nM range.

Summary and Conclusions

- Fragment space searches with FTrees allow to expand the chemical space to be screened by VS enormously.
- A large pool of reagents allows both diverse and focused decoration of scaffolds detected by the search.
- Knowledge of combinatorial chemistry and dedicated chemistry capacity for synthesis of VS hits are needed.
- Due to the validated chemistry protocols the time lines for synthesis fit to the projects' needs.

Acknowledgements

Holger Claußen*
Markus Lilienthal*
Christian Lemmen*

Christoph Hoenke
Dieter Wiedenmayer

Jürgen Mack
Ingo Mügge

Alexander Weber
Bernd Wellenzohn

Herbert Köppen
Ulrike Küfner-Mühl

*: BiosolveIT, St. Augustin