Conditions of Use:

With opening this file you agree to the following:

- Please feel free to read and share this document within your company or organization.
- When sharing, please make sure that this slide remains part of the presentation.
- When citing, we prefer if you cite the original scientific journal articles and not this slide deck.

- You are not allowed to share this presentation or parts of it in any public places or with people outside your organisation. Please feel free to share the download link.
- You are not allowed to take parts of this presentation into your presentation in case it is publically presented (i.e. outside your organisation). Please ask for written consent in this case.

- In case of questions please get in touch: rarey@zbh.uni-hamburg.de

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

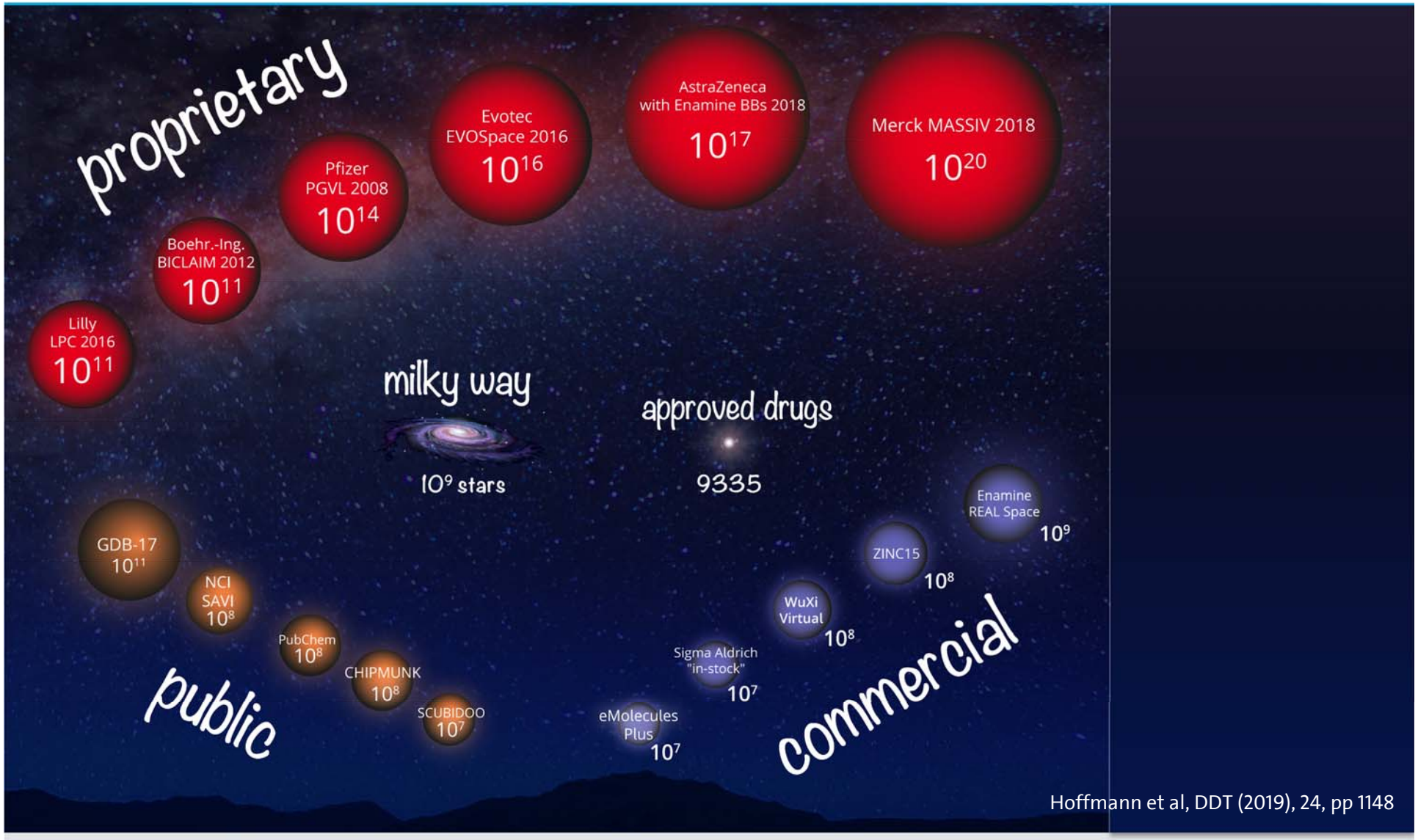
Some New Algorithmic Approaches for Cheminformatics and Modeling

MATTHIAS REAREY

Research Group for Computational Molecular Design

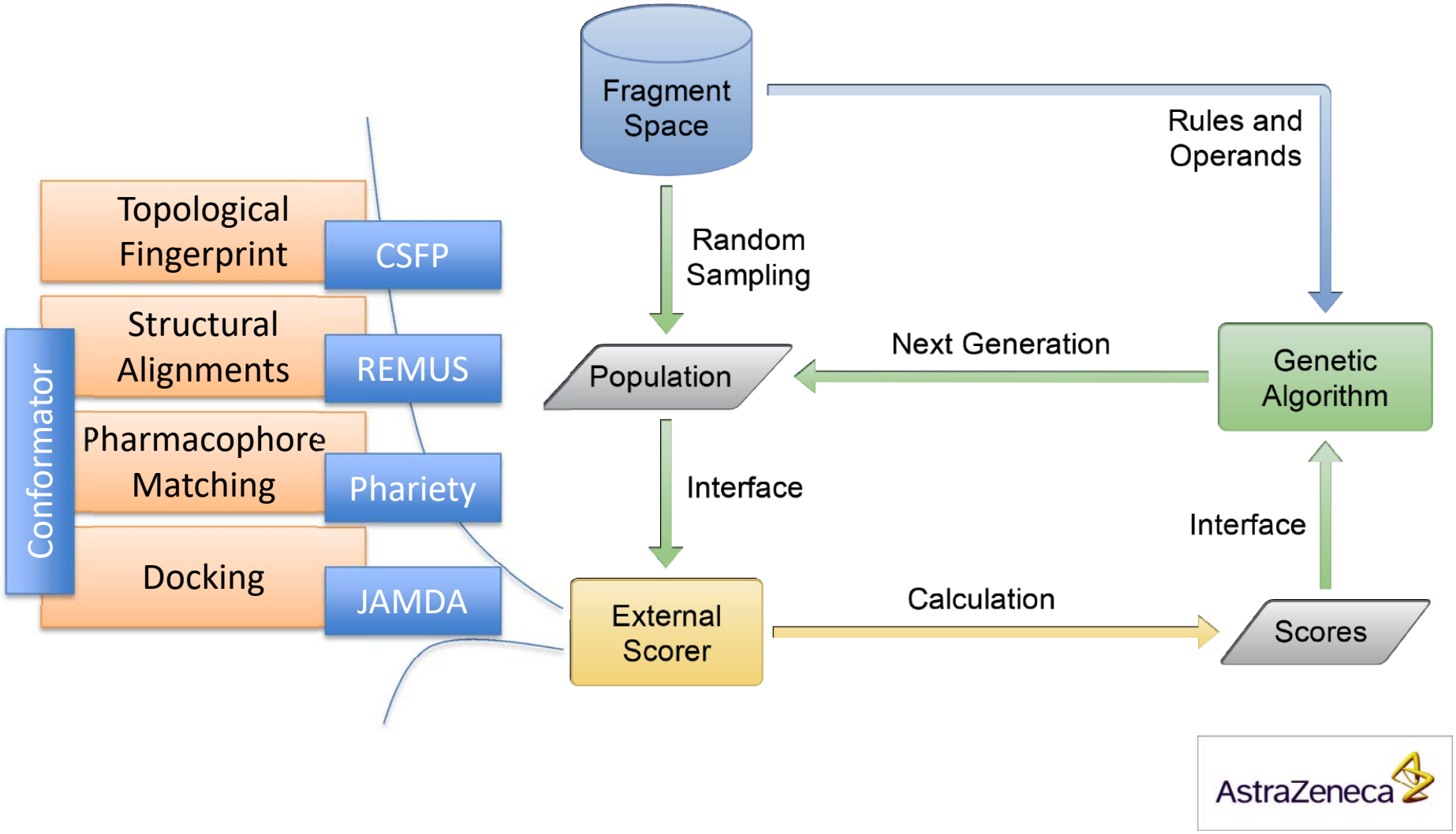
Forschungsgruppe Algorithmisches Molekulares Design (AMD)

Advances in Modeling Chemical Space

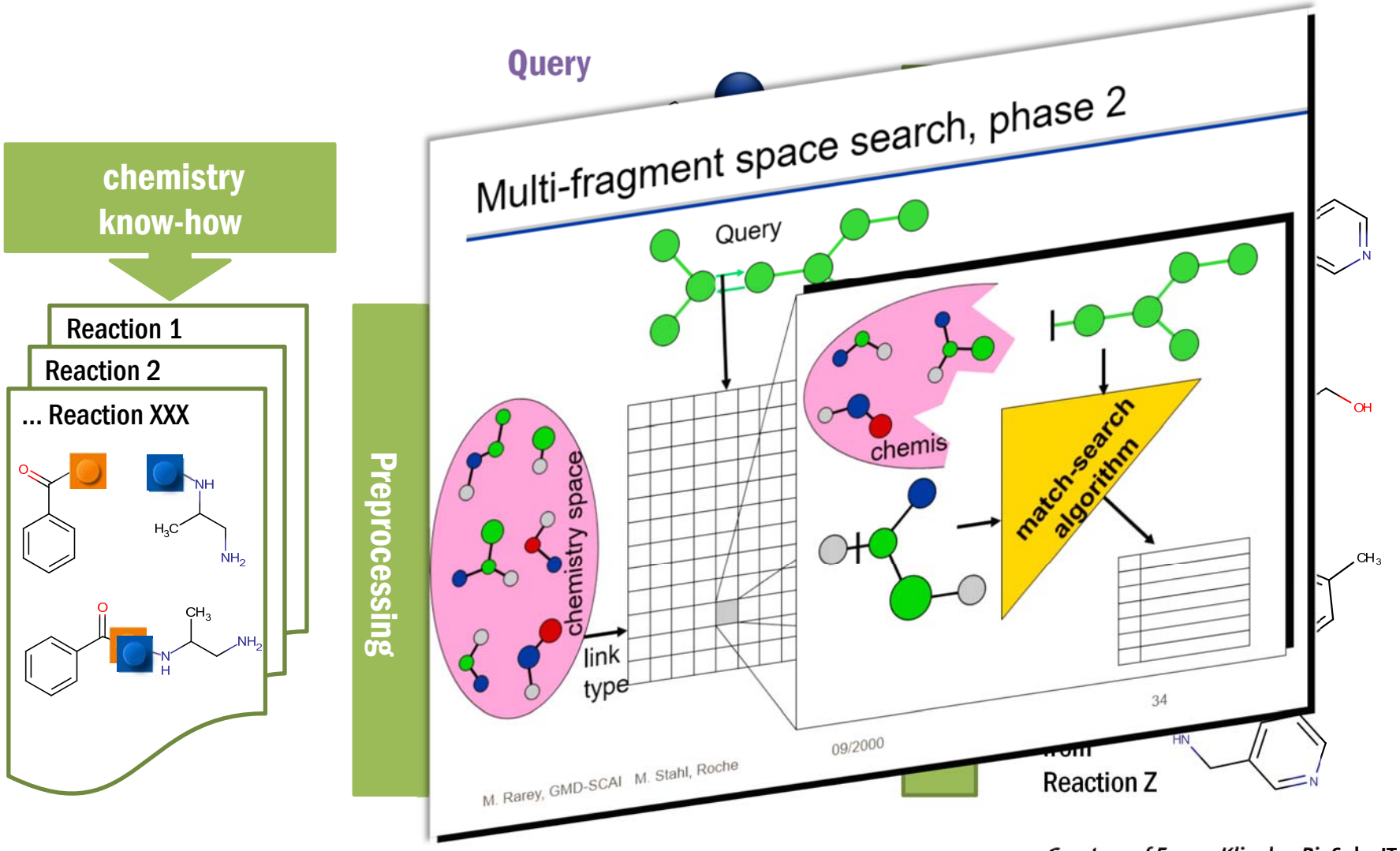


Hoffmann et al, DDT (2019), 24, pp 1148

Galileo: 3D Genetic Operators for Searching in Chemical Space

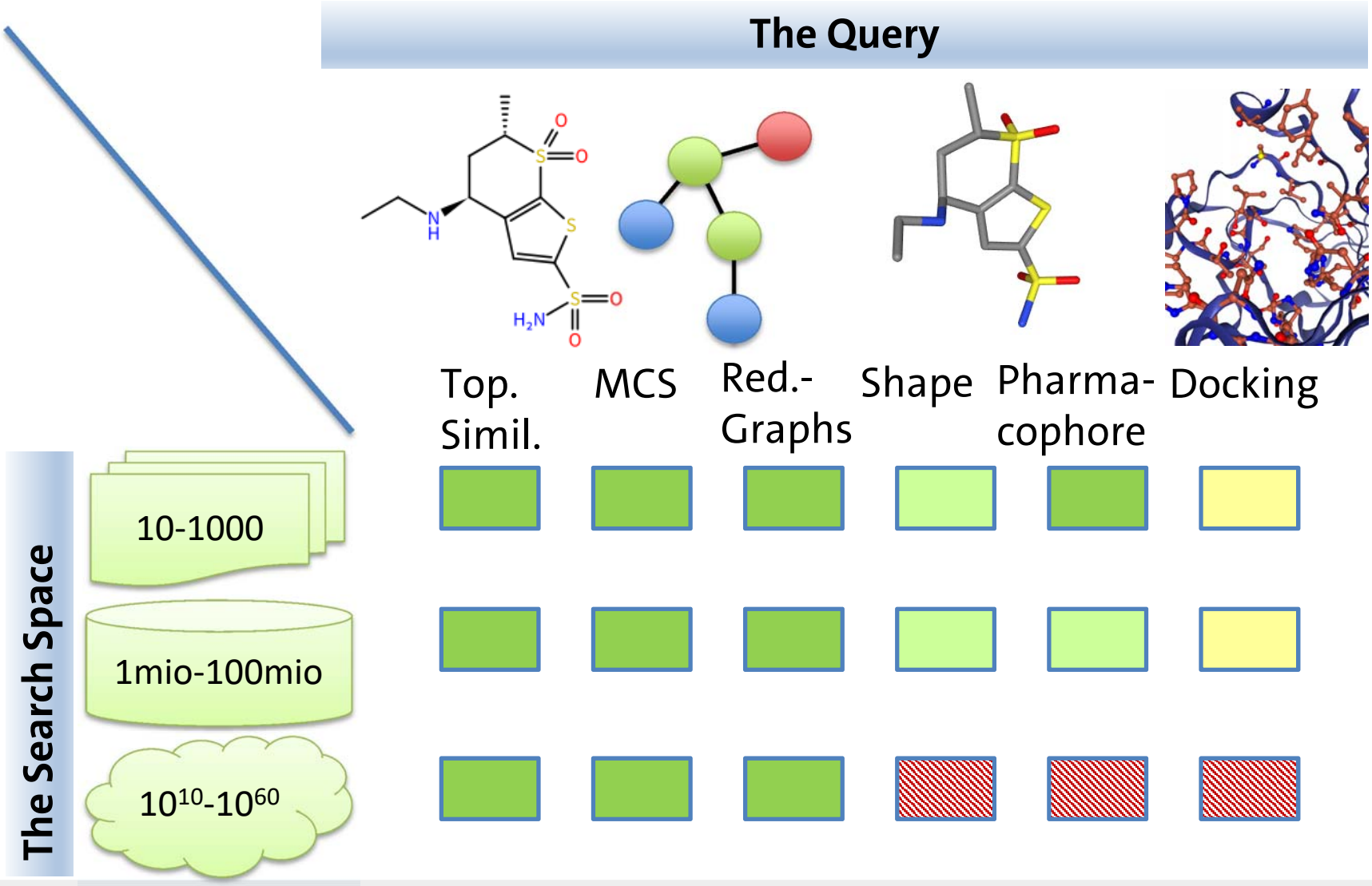


FTrees Fragment Spaces: Searching Instead of Screening



Courtesy of Franca Klingler, BioSolveIT GmbH

Molecular Design in a Nutshell: The First Phase



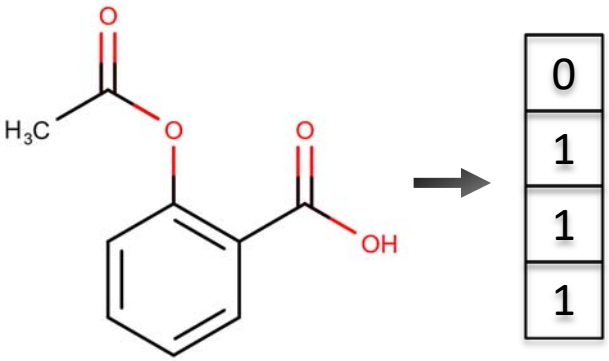
SpaceLight

SpaceLight: Bellmann et al, (2021), J.Chem.Inf.Model., **61**(1), **238ff**

SpaceLight – Topological Searching in Chemical Space

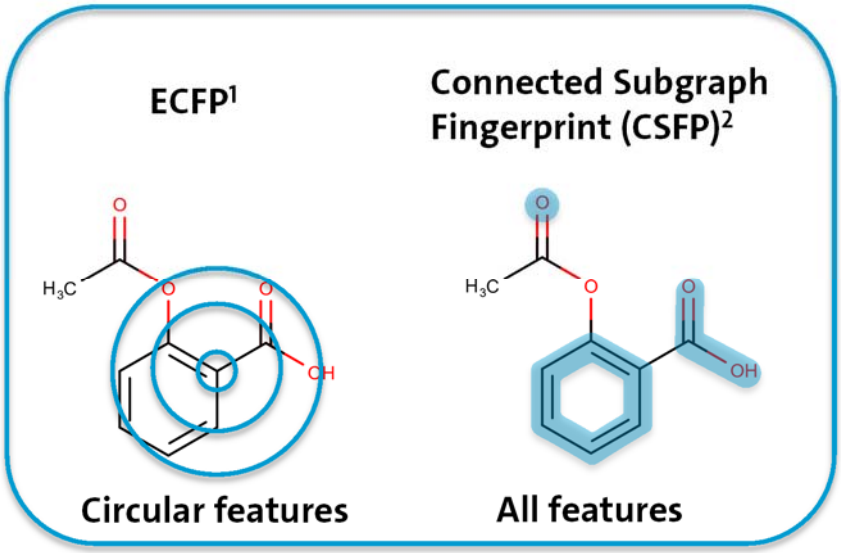
Classic Molecular Fingerprints in a New Light

Employ classic fingerprints in a new search approach



CSFP versions

fCSFP: fine-grained similarity measurement
iCSFP: MCS-like descriptor
tCSFP: scaffold-hopping potential

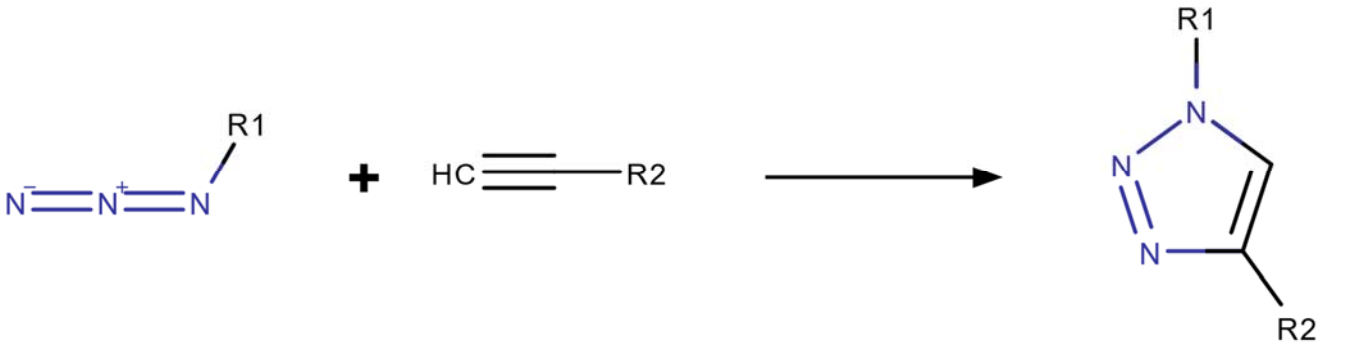


Notation

ECFPx: x max diameter for circular features
CSFPy.z: y lower, z upper bound for #heavy atoms

1: Rogers and Hahn, J. Chem. Inf. Model. (2010), 50
2: Bellmann et al., J. Chem. Inf. Model. (2019), 59, 11

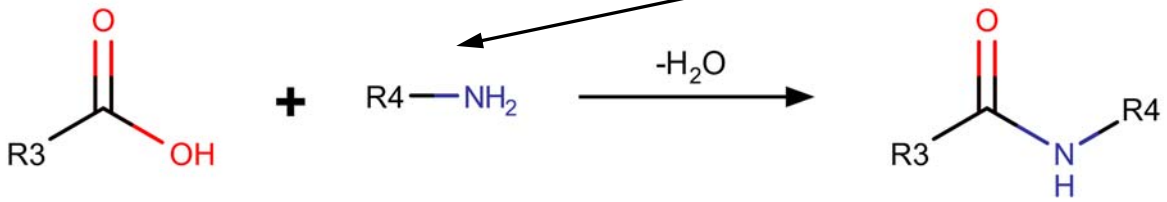
A Way Out: Combinatorial Nature of Chemistry



Pool A
1000 fragments

Pool B
1000 fragments

Intermediate
1000^2 = 1 million

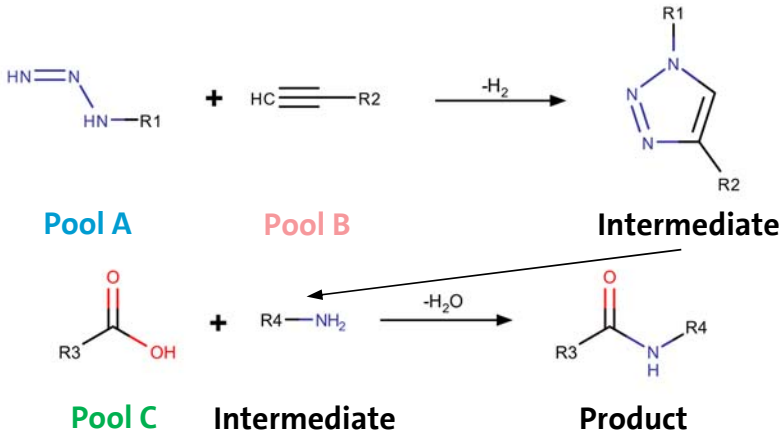


Pool C
1000 fragments

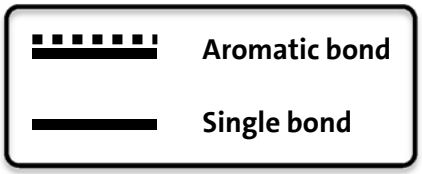
Intermediate
1000^2 = 1 million

Product
1000^3 = 1 billion

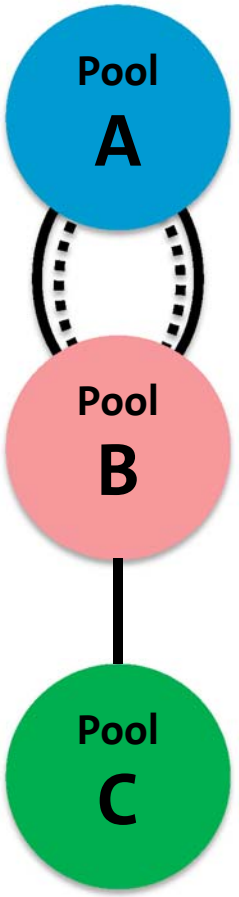
How We Describe Topological Fragment Spaces

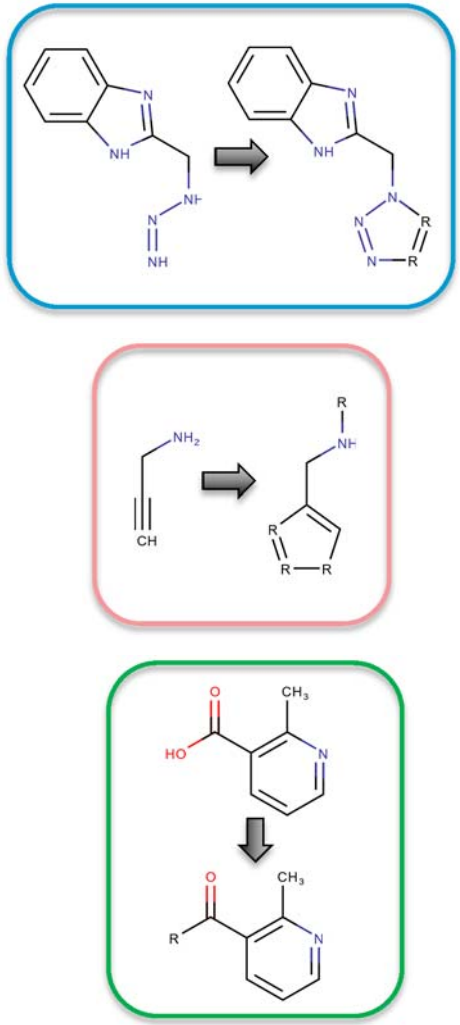


Capture newly formed bonds and structural changes in educts

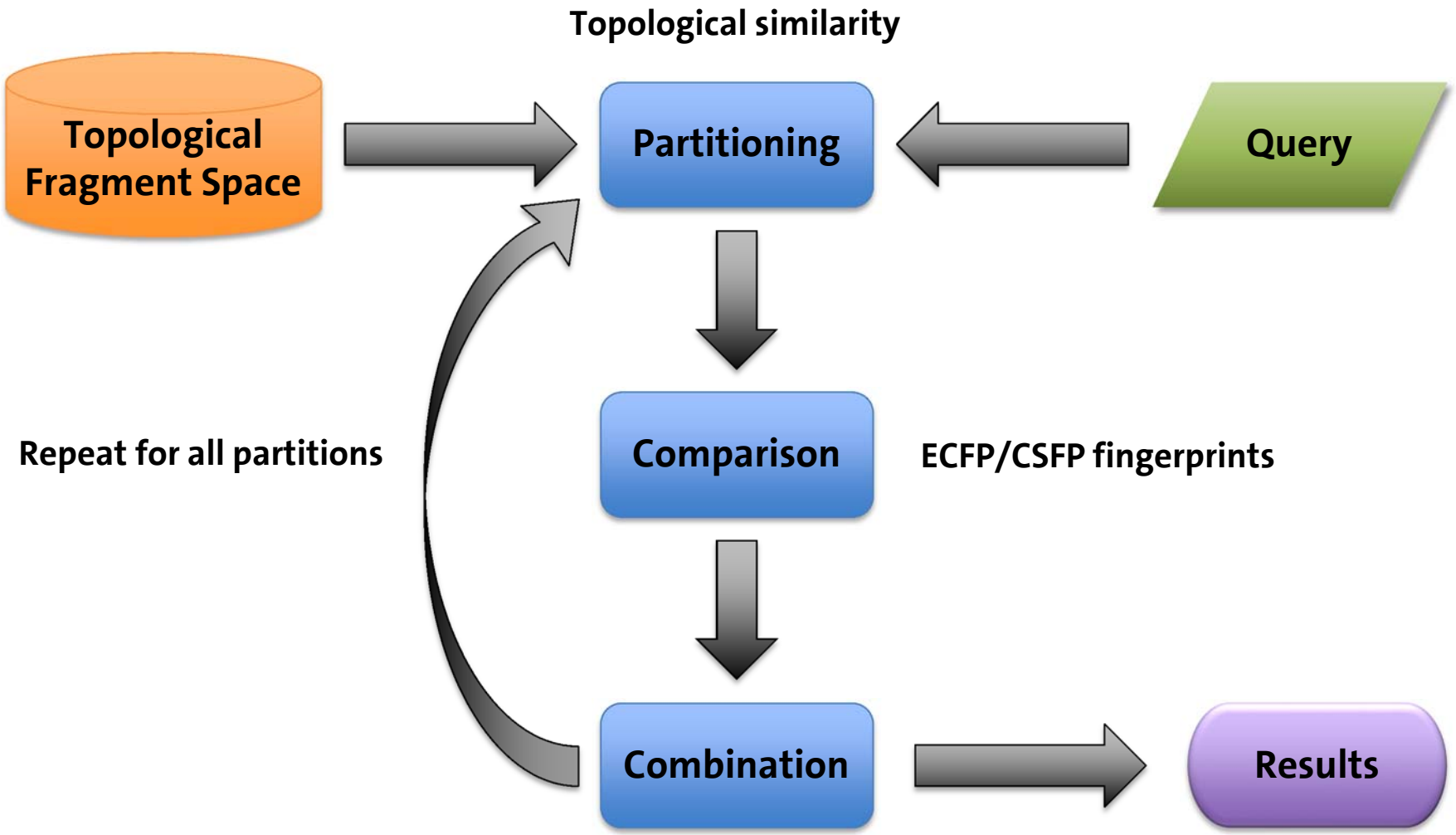


Topology Graph



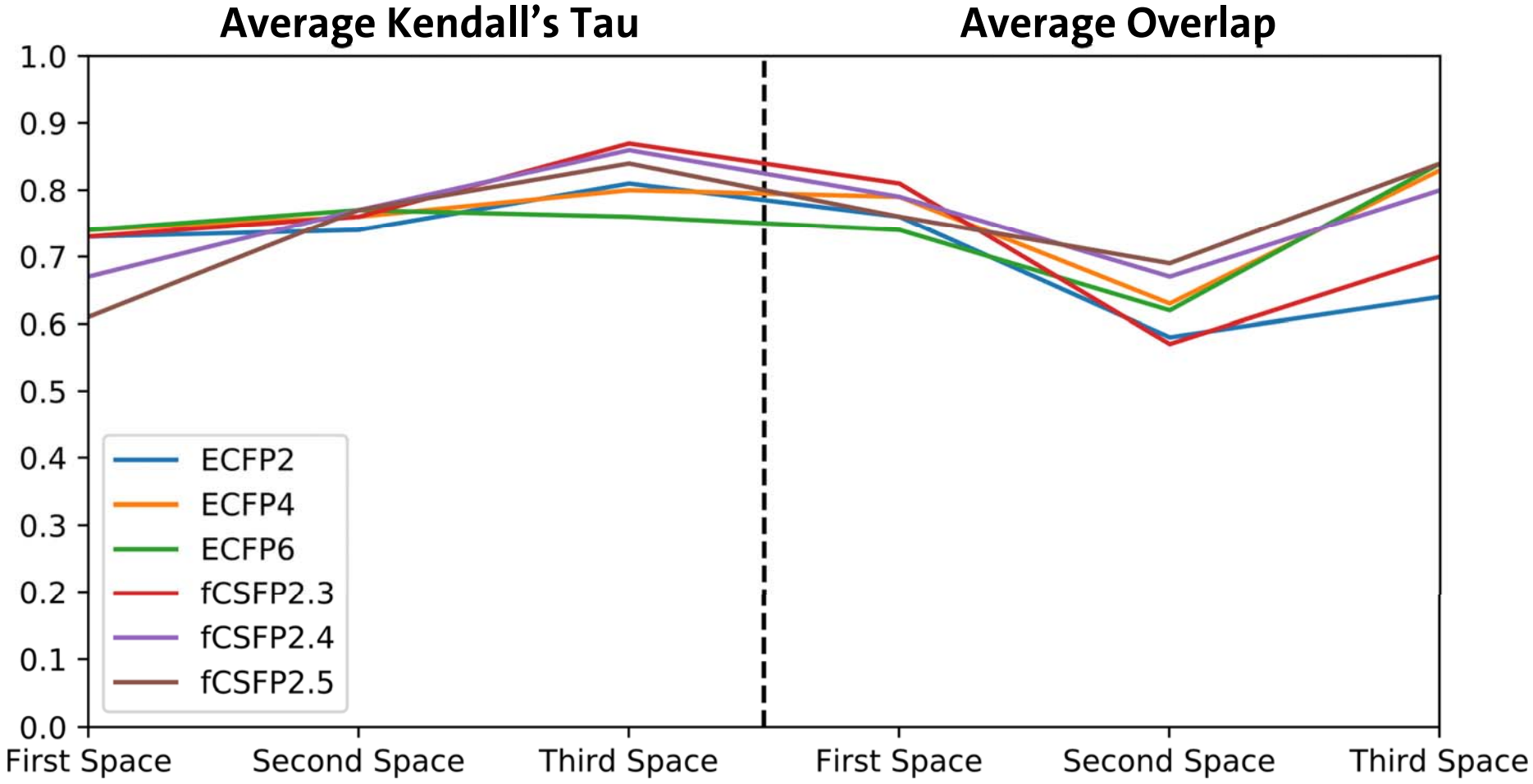


SpaceLight Algorithm Overview



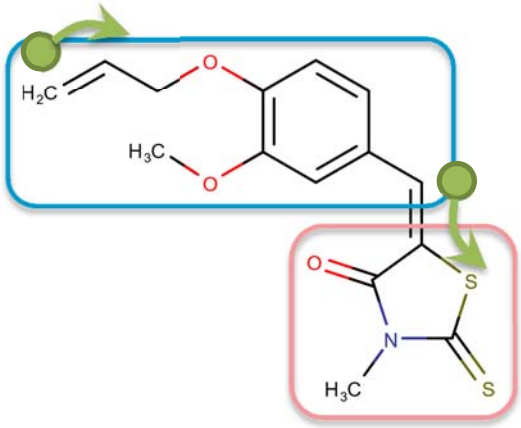
SpaceLight: Bellmann et al, (2021), J.Chem.Inf.Model., **61**(1), **238ff**

SpaceLight Correlation with Sequential Search

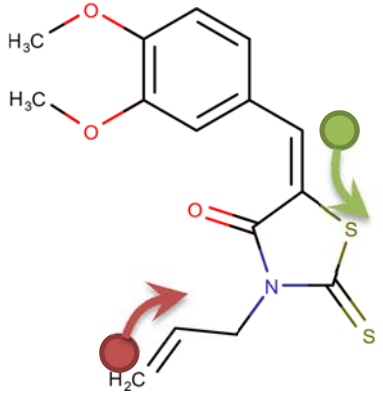


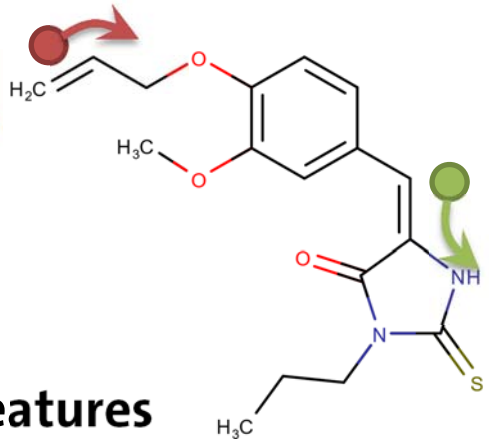
SpaceLight generates results similar to classic fingerprints

SpaceLight vs. Sequential ECFP4 Search



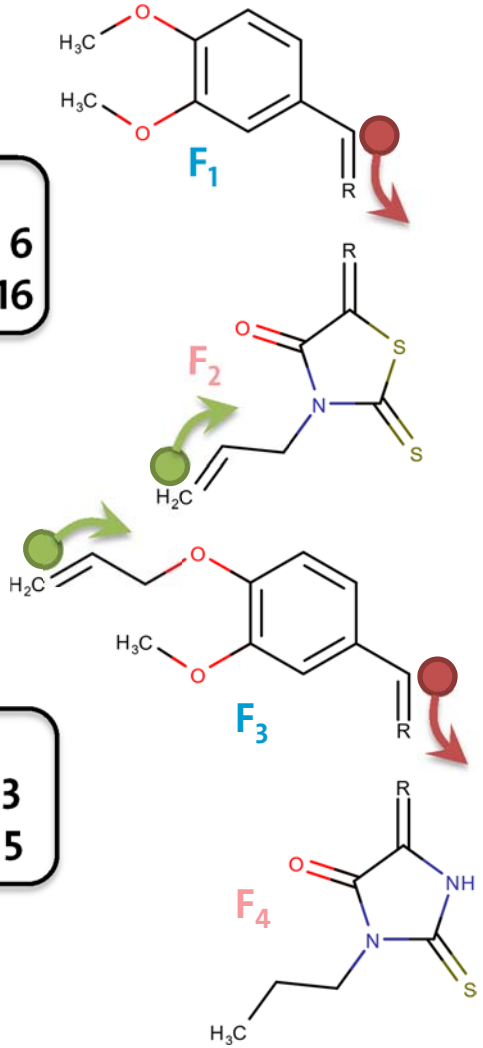
ZINC029932278





Product Ranks
Classic ECFP4: 6
SpaceLight$_{ECFP4}$: 16

Product Ranks
Classic ECFP4: 53
SpaceLight$_{ECFP4}$: 5

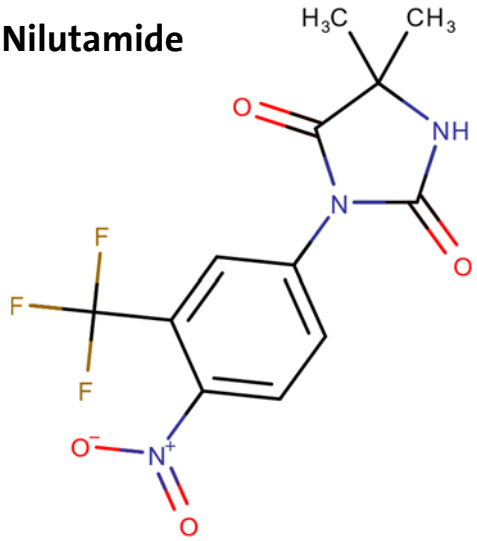


SpaceLight considers locality of chemical features

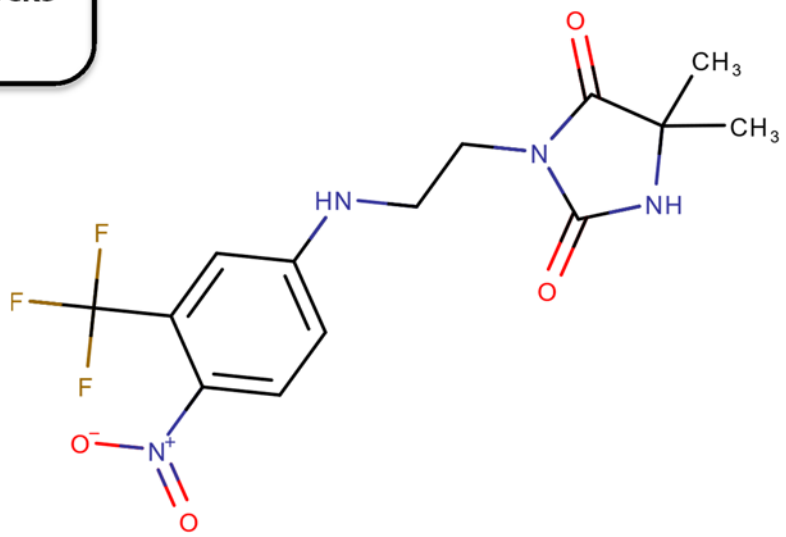
SpaceLight: Bellmann et al, (2021), J.Chem.Inf.Model., **61**(1), **238ff**

SpaceLight Search for Drugs in Enamine REAL

Nilutamide



Enamine REAL Space
>180 synthesis protocols
>115000 building blocks
>13 billion products



Found with score 1.0

Second best with score 0.87

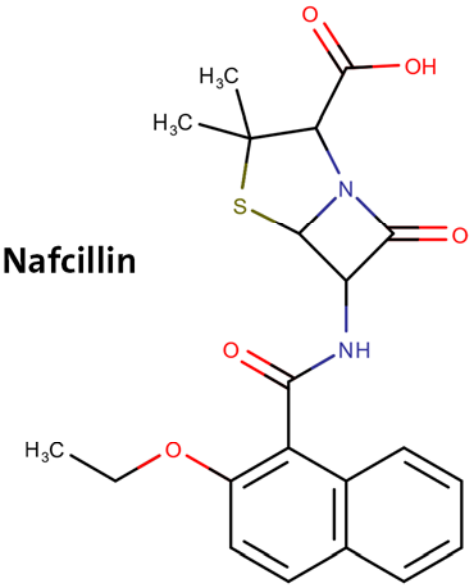
SpaceLight: Bellmann et al, (2021), J.Chem.Inf.Model., **61**(1), **238ff**

SpaceLight Gives You Ideas for Synthesis Routes

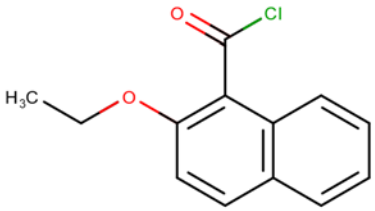
SpaceLight retrieves synthetic pathways of products

Knowledge Space
120 synthesis protocols
>142000 building blocks
>10^{15} products

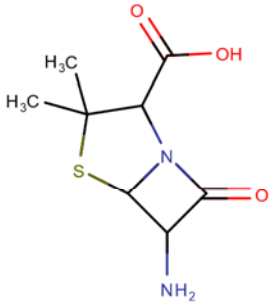
Nafcillin



Found with score 1.0



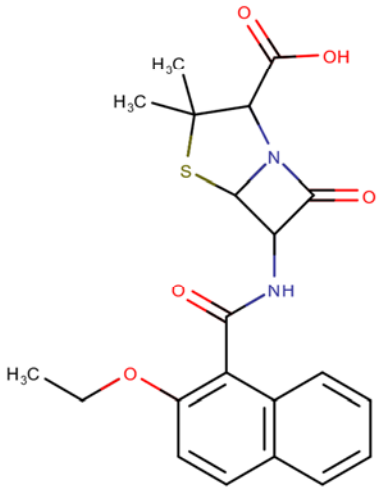
+



Schotten-Baumann

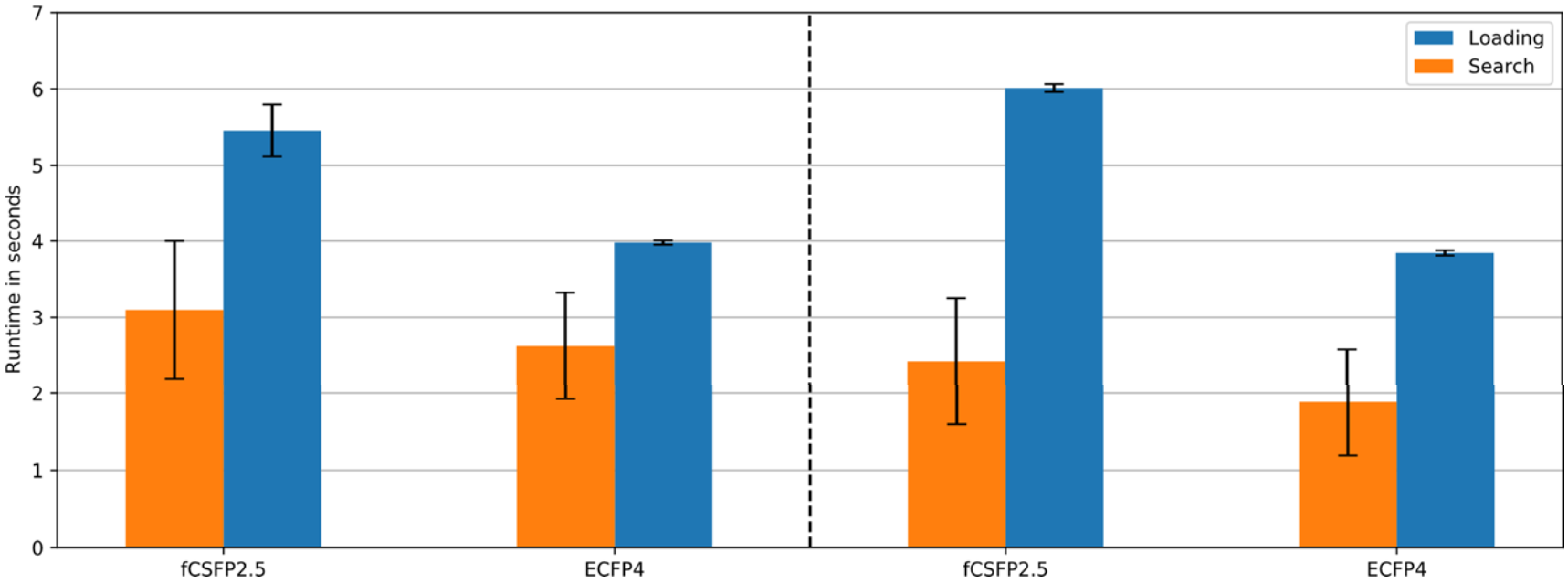
- NaCl
- H_2O

NaOH



SpaceLight: Bellmann et al, (2021), J.Chem.Inf.Model., **61**(1), **238ff**

SpaceLight Searches 10^{15} Products In a Few Seconds



Enamine REAL Space
>180 synthesis protocols
>115000 building blocks
>13 billion products

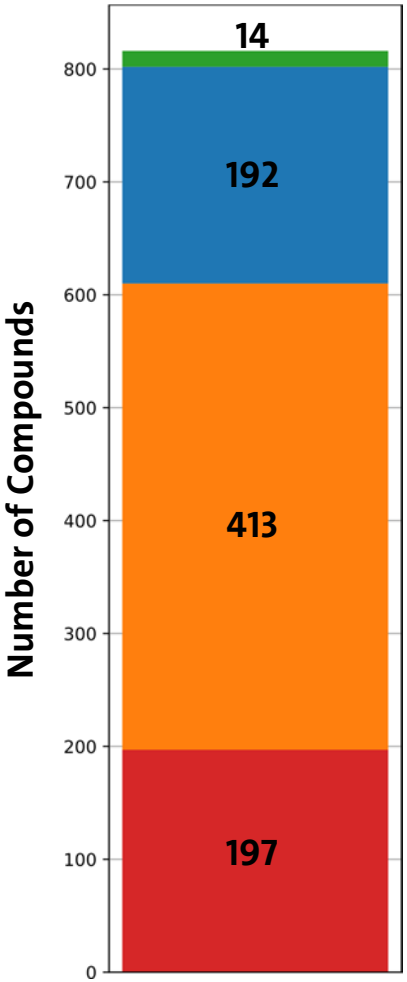
Intel Core i5-6500 @3.2GHZ
16 GB memory
Search with 3 threads

500 random queries from ZINC lead-like

Knowledge Space[1]
120 synthesis protocols
>142000 building blocks
>10^{15} products

1: https://www.biosolveit.de/CoLibri/spaces.html#knowledgespace

SpaceLight Search for Drugs in Enamine REAL



Enamine REAL Space
>180 synthesis protocols
>115000 building blocks
>13 billion products

Queries from DrugBank[1]
1. Approved drugs
2. Molecular weight between 300 and 500
=> 816 compounds

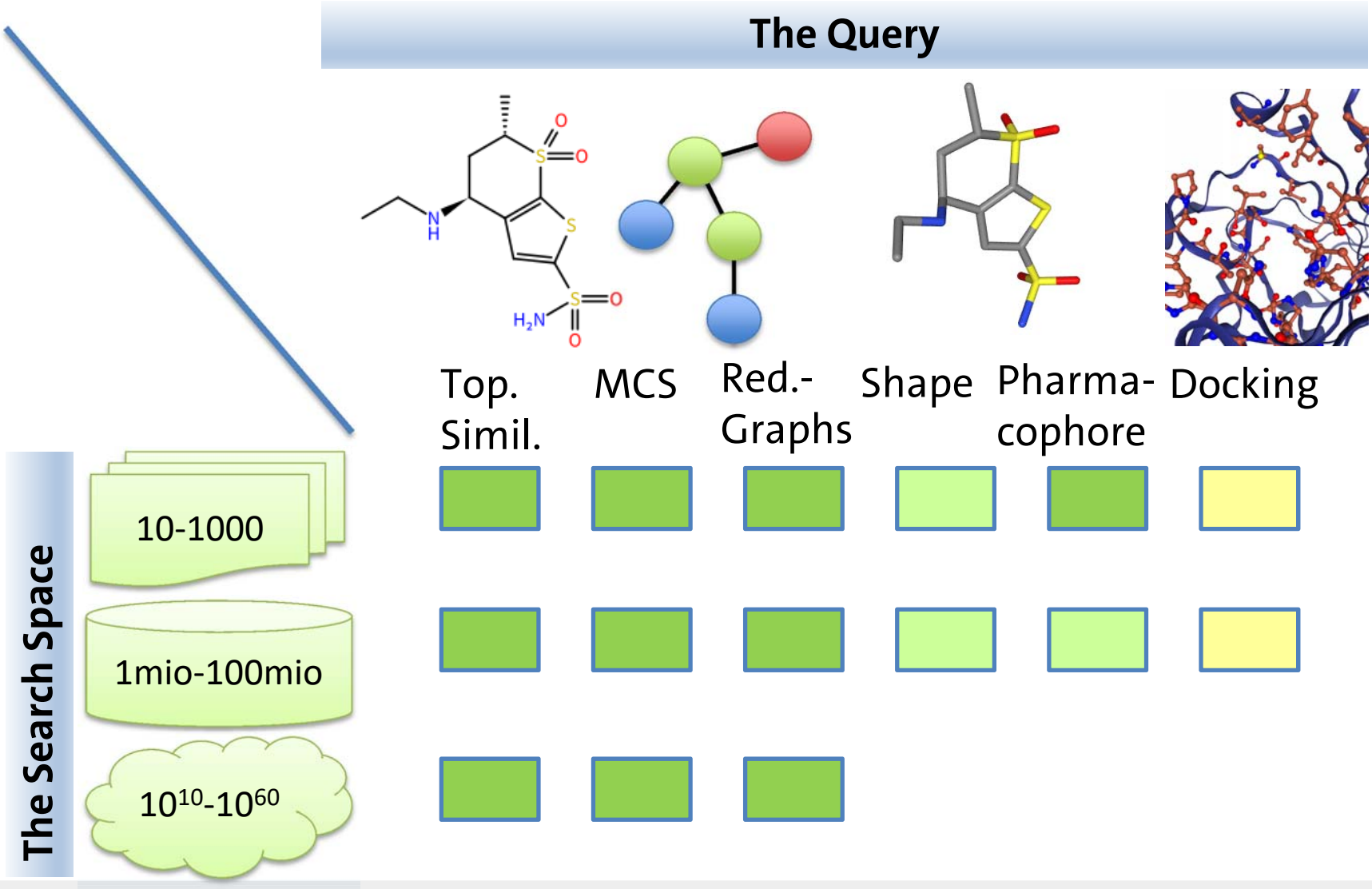
SpaceLight$_{fCSFP2.5}$

Search run took 22 min (1.6 s per compound)

- Identical: Score 1.0
- Highly similar: Score above 0.7
- Less similar: Score above 0.4
- Dissimilar: Score below 0.4

1: Wishart, Nucleic Acids Res. (2008), 46

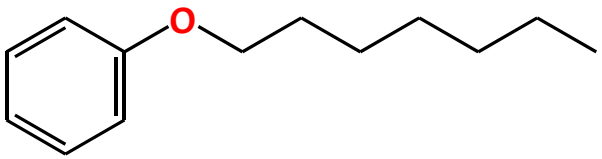
Molecular Design in a Nutshell: The First Phase

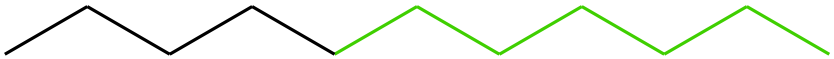


SpaceMACS

Schmidt et al., in preperation

Maximum Common Substructure Search in Chemical Space





Qsize: 14 atoms

- connected MCS
- acyclic/acyclic bond matching

Tsize: 11 atoms

Msize: 7 atoms

Hit list

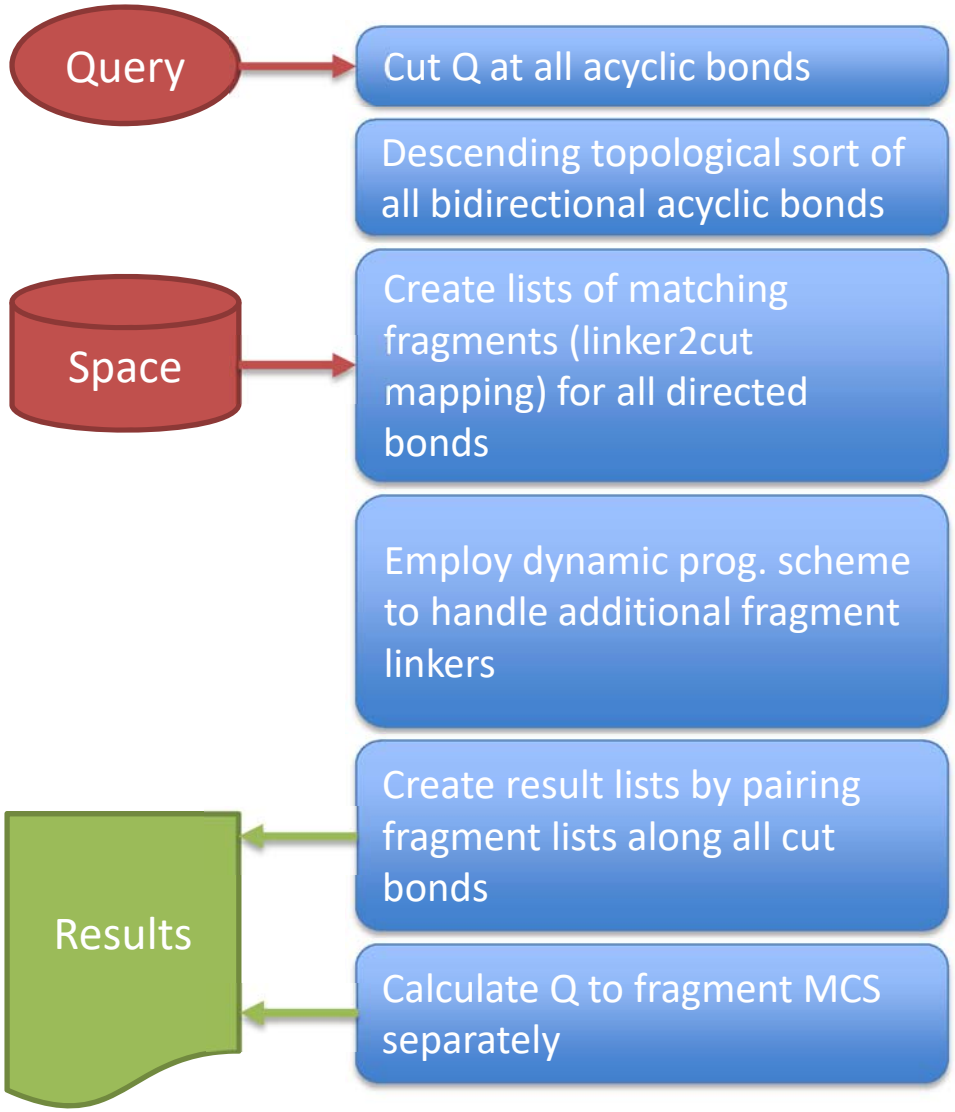
Sort criteria:

1. Msize | Tsize
2. MCS-Similarity:

$$\frac{\text{Msize}}{\text{Qsize} + \text{Tsize} - \text{Msize}}$$

SpaceMACS Algorithm

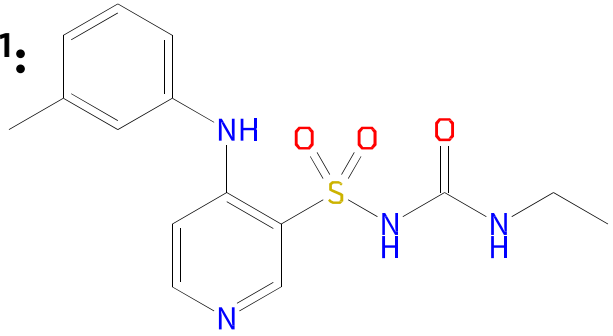
- RIMACS library[1] for elementary connected MCS calculation
- FTreesFS[2] dynamic programming scheme to handle matches over fragment borders
- Precise for Msize sorting, heuristic for MCS-similarity sorting on lower ranks
- SMARTS features in compatibility check



1: Schmidt et al, JCIM (2021), **61**(1), 167ff
2: Rarey et al, JCAMD (2001), **15**, 497ff

A First Run: Why MCS-Similarity?

Query[1]:

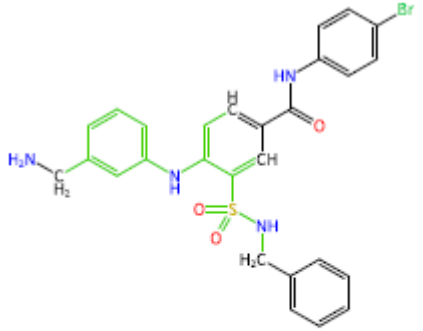


Space[2]: GalaXi-Space 2020-11
2.1 bn compounds

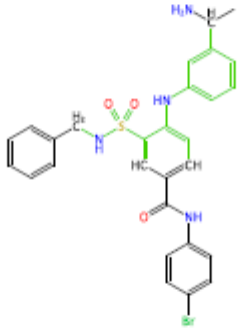
Hit list:

(by MCS-Size (#M=18))

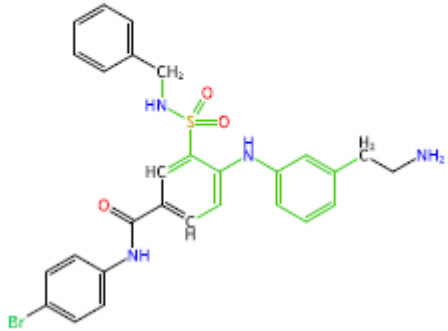
1.



2.



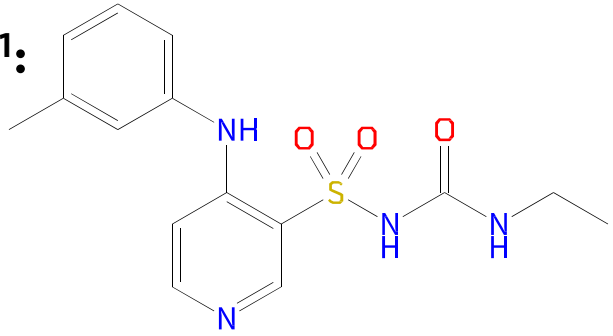
3.



[1]Lessel & Lemmen, ACS Med.Chem.Lett. (2019), 10(10), 1504ff
[2]https://www.biosolveit.de/download/?file=2.1bn-GalaXi_2020-11.space

A First Run: Why MCS-Similarity?

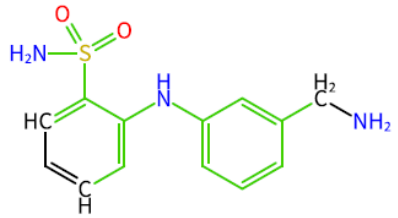
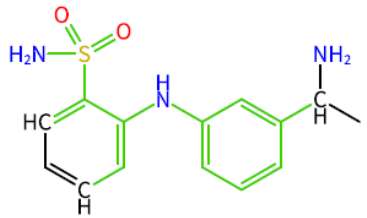
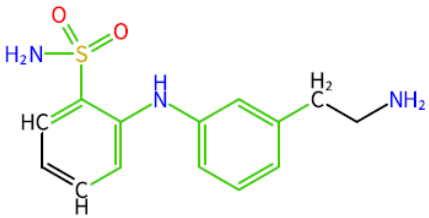
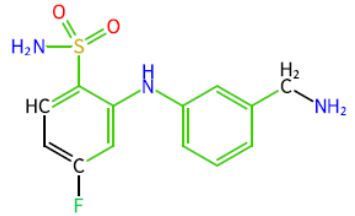
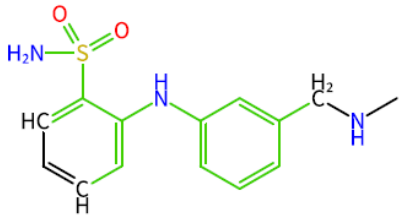
Query[1]:



Space[2]: GalaXi-Space 2020-11
2.1 bn compounds

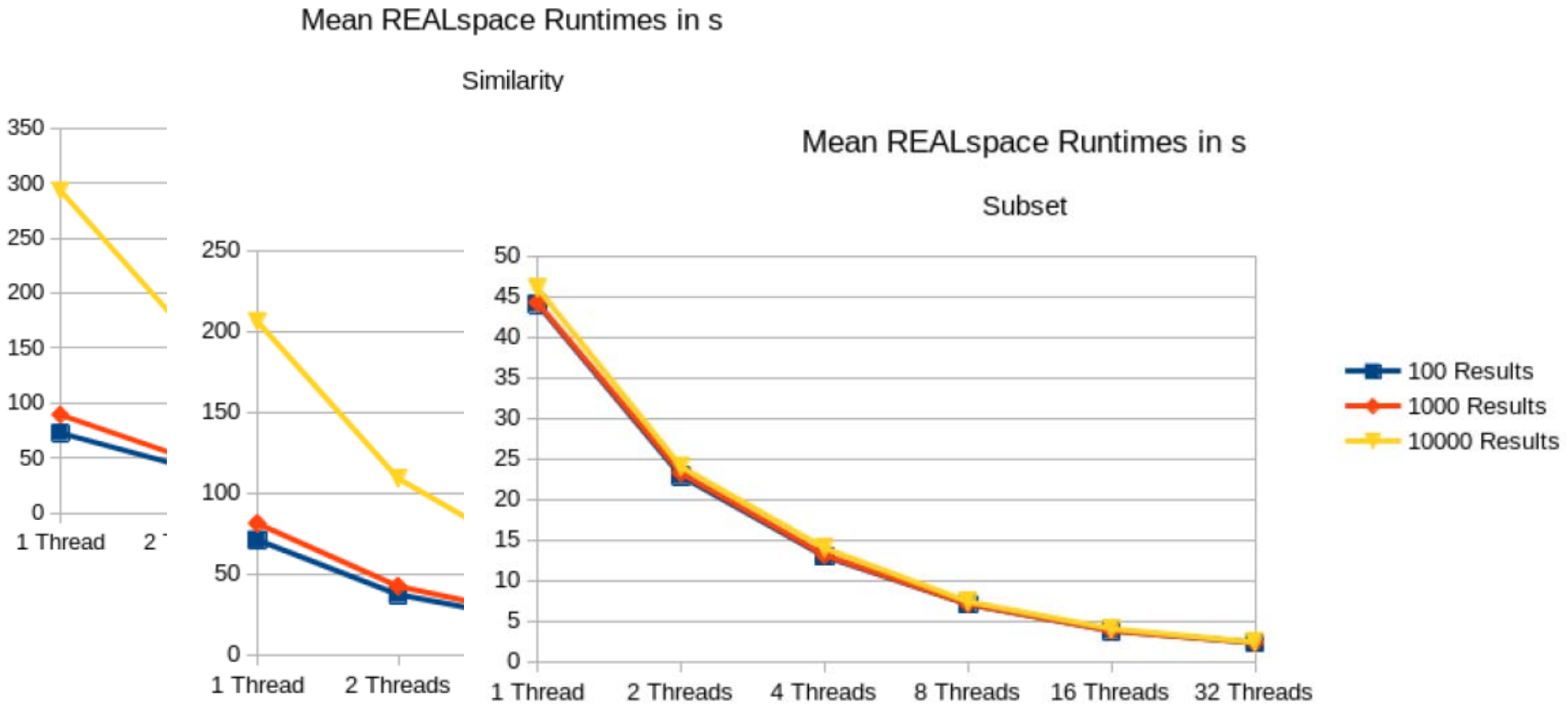
Hit list:

(by MCS-Similarity (#M=17)

1. 
2. 
3. 
4. 
5. 

[1]Lessel & Lemmen, ACS Med.Chem.Lett. (2019), 10(10), 1504ff
[2]https://www.biosolveit.de/download/?file=2.1bn-GalaXi_2020-11.space

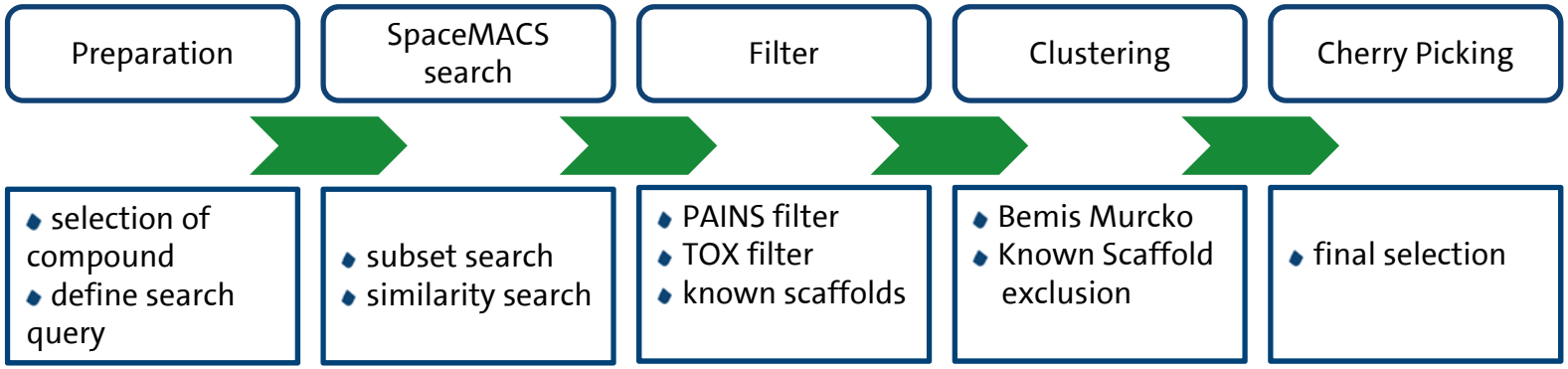
Computing Times and Speed Up



100 queries taken from [1],
searching in Enamine REALspace[2] 2020-11 (15 bn compounds)

[1]Lessel & Lemmen, ACS Med.Chem.Lett. (2019), 10(10), 1504ff
[2]https://www.biosolveit.de/infiniSee/#realspace

SAR-by-SPACE Application Scenario (by Raphael Klein, BSI)



ACS Medicinal Chemistry Letters

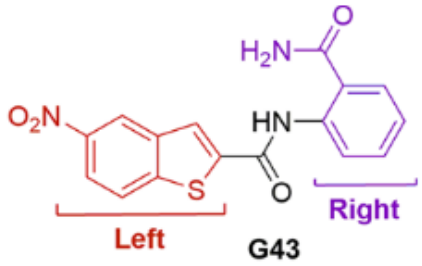
Discovery of Potent Inhibitors of *Streptococcus mutans* Biofilm with Antivirulence Activity

Bhavitavya Nijampatnam, Parmanand Ahirwar, Piyasuda Pukkanasut, Holly Womack, Luke Casals, Hua Zhang, Xia Cai, Suzanne M. Michalek, Hui Wu*, and Sadanandan E. Velu*

https://doi.org/10.1021/acsmedchemlett.0c00373

Starting Point

- lead compound G43:



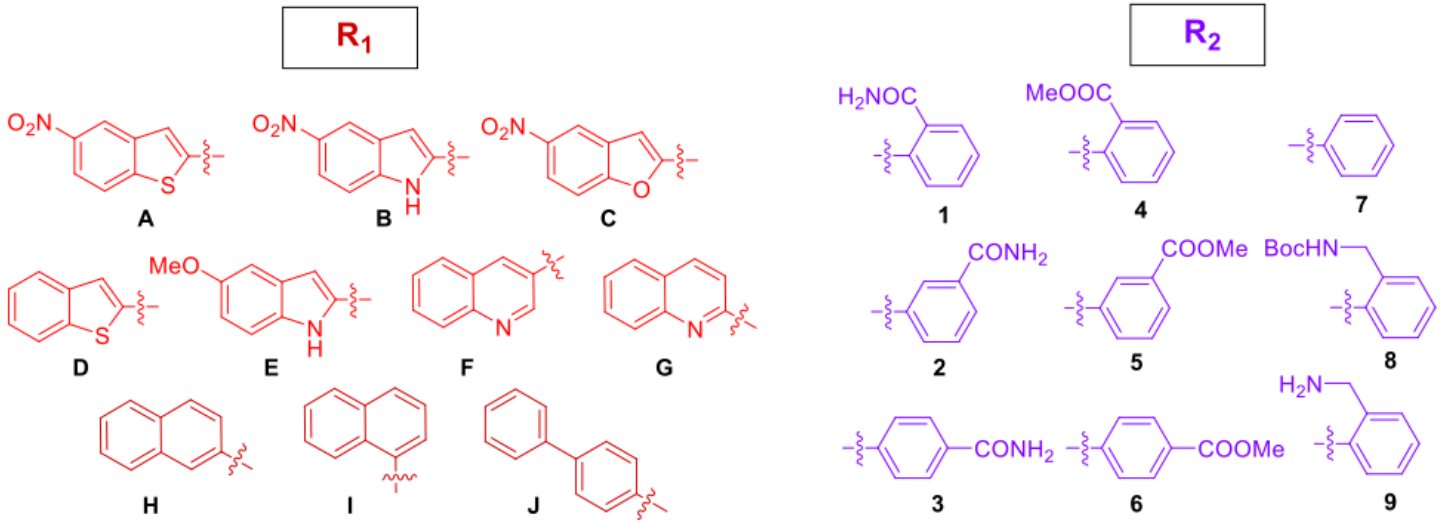
G43

Biofilm IC_{50} = 16.7 µM

Gtf IC_{50} = 10 µM

Growth: No inhibition up to 200 µM

- authors of publication generated 90 analogs by replacing left and right part

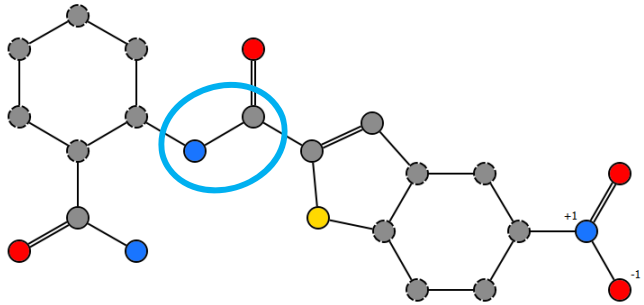


Preparation and SpaceMACS Call

- Determine query SMILES:

S1c2c(cc([N+]([O-])=O)cc2)C=C1C(=O)Nc3c(cccc3)C(=O)N

=> MCS-similarity search (=> SAR-by-Space)



- `SpaceMACS –f 17.6bn-REALSpace_2021-03.space -i smiles-input.smi –n 50000 -m similarity -o search50.csv >search50-log.txt`

  Search50-log.txt => Total program runtime: 144 seconds

- `SpaceMACS –f 2.1bn-GalaXi_2020-11.space –I smiles-input.smi -n 50000 -m similarity -o search49.csv >search49-log.txt`

  Search49-log.txt => Total program runtime: 41.4 seconds

Hardware used: Intel(R) Xeon(R) W-3223 CPU @ 3.50GHz, 8 cores (16 threads), openSUSE 15.1

Potential Follow-Up: Scaffold filter and BM-Clustering

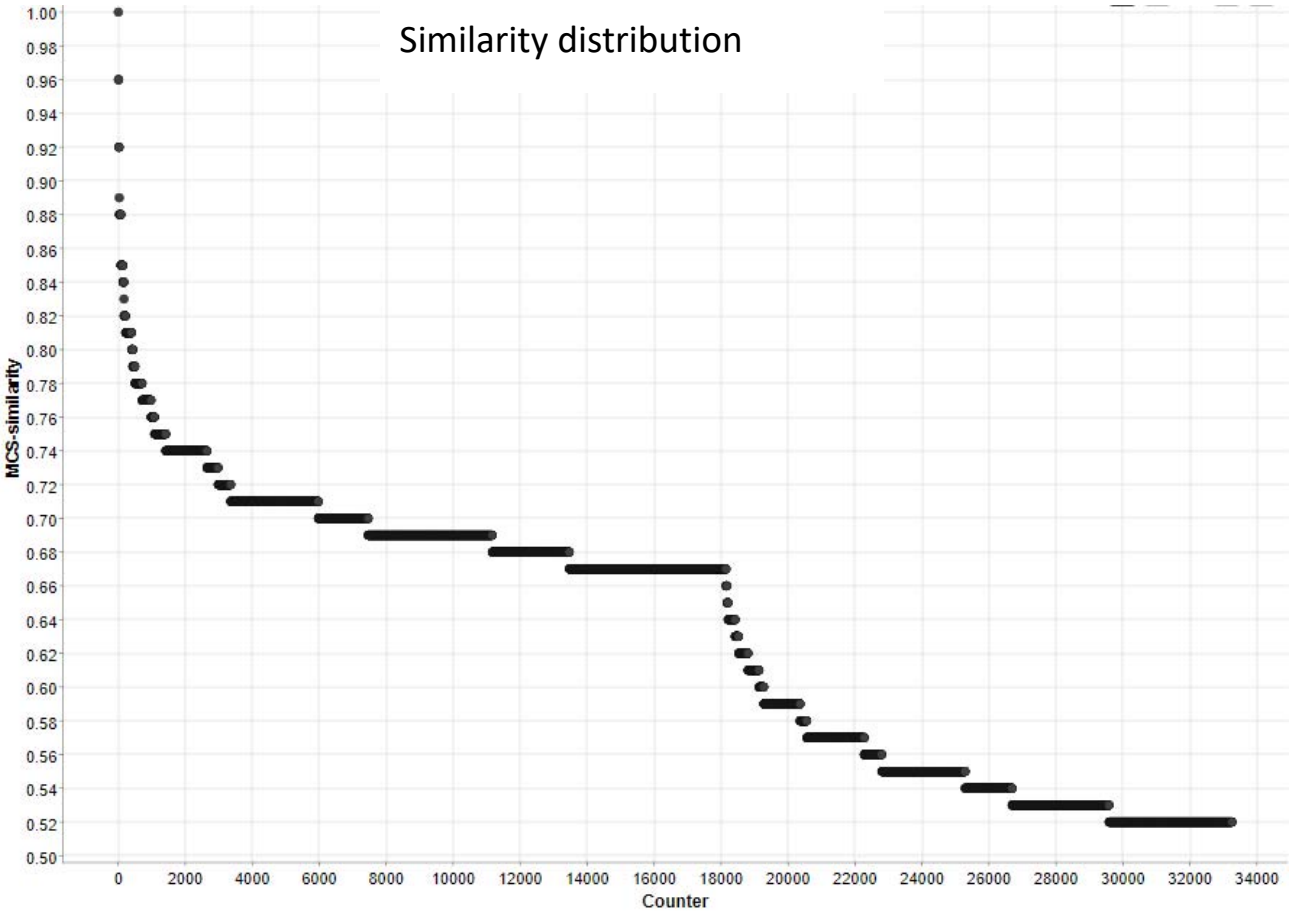
SpaceMACS
=> 100k

Lead-like filter
=> 33k

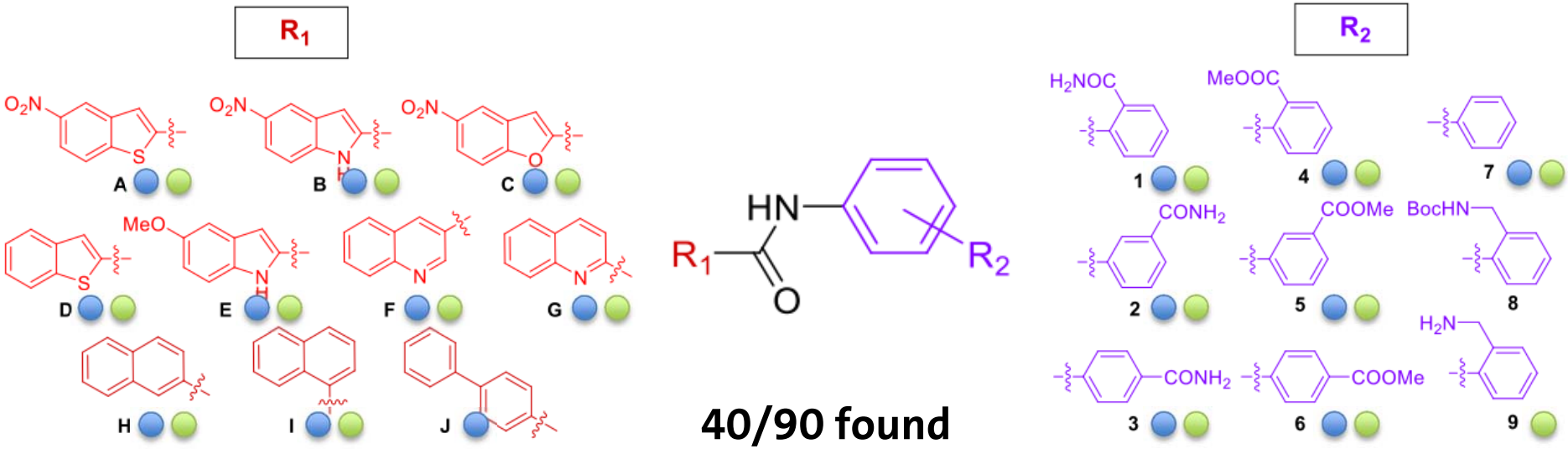
Similarity >0.67
=>18k

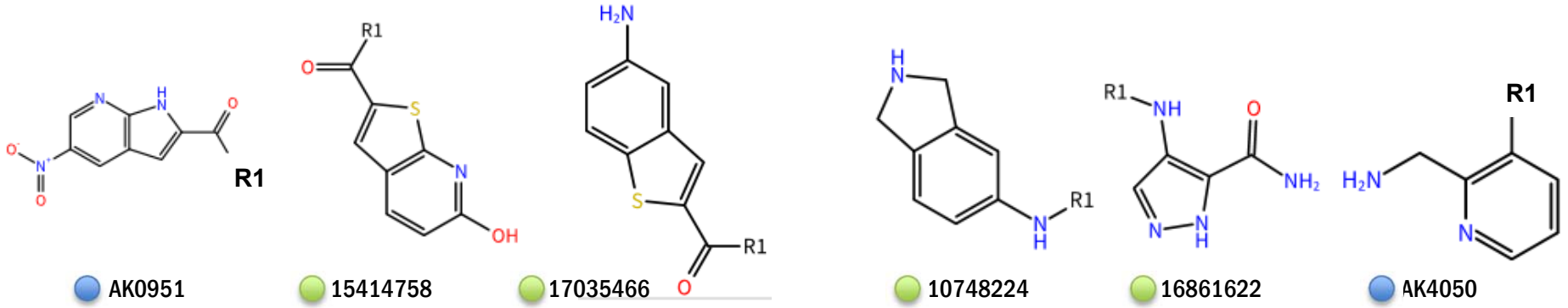
known scaffolds
(BM, atom-typed)
=> 10k

886 BM scaffolds



Recall of Compounds from the Publication

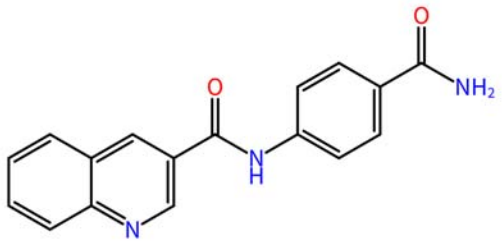




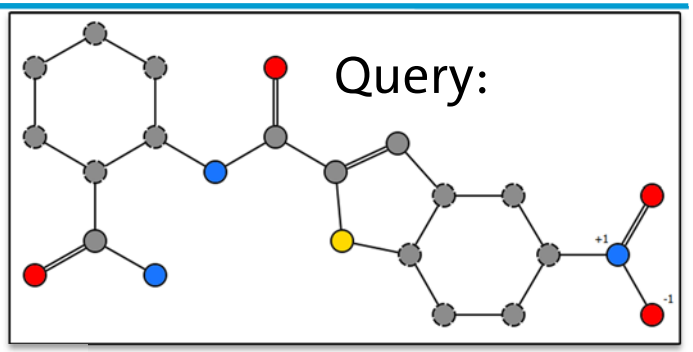
2.1bnGalaXi_2020-11 17.6bnRealSpace_2021-03

Overlap: SpaceMACS – SpaceLight – FTreesFS

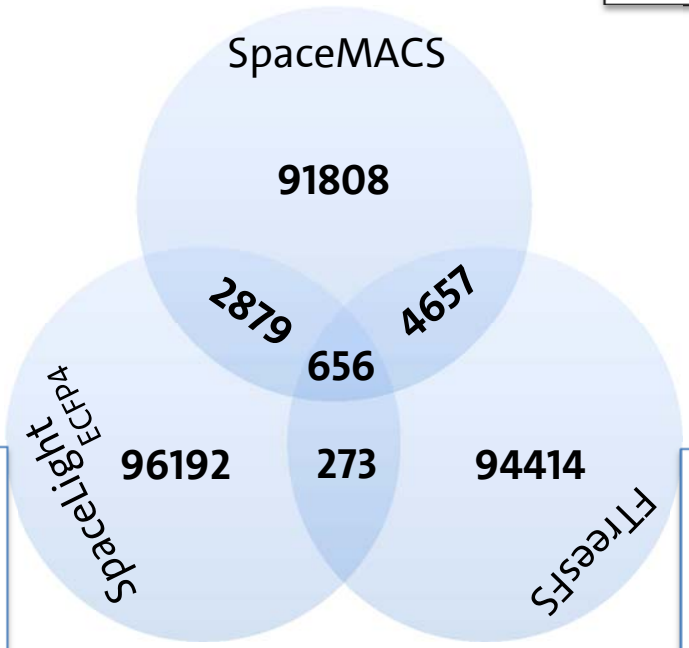
- 4 Molecules only found by FTreesFS:

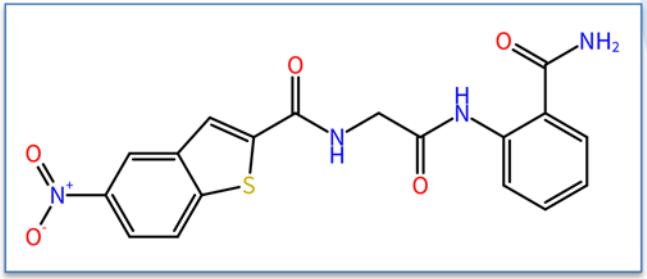


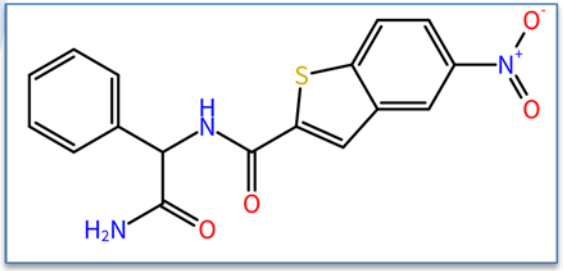
Query:



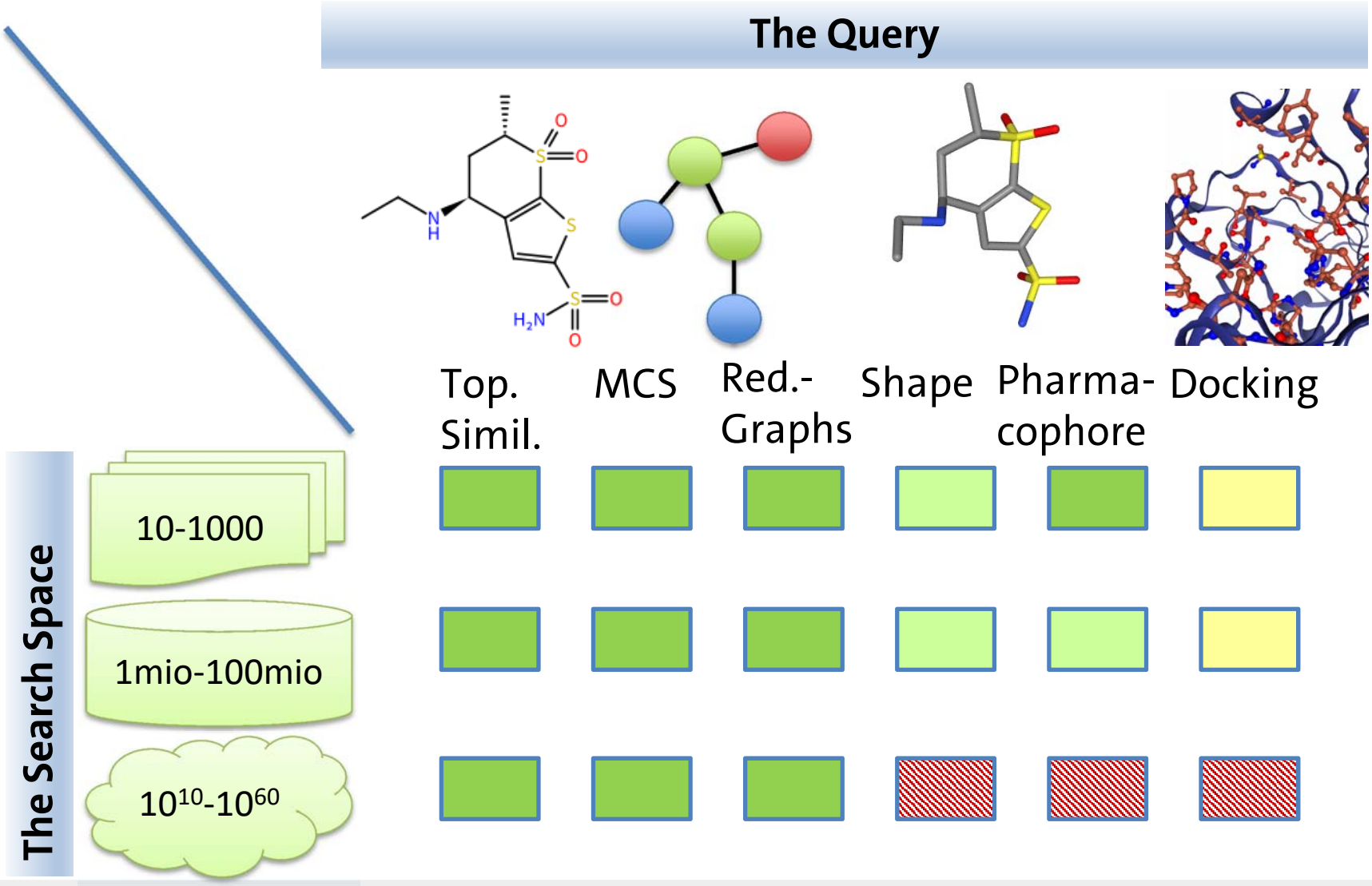
- Hit list overlap:



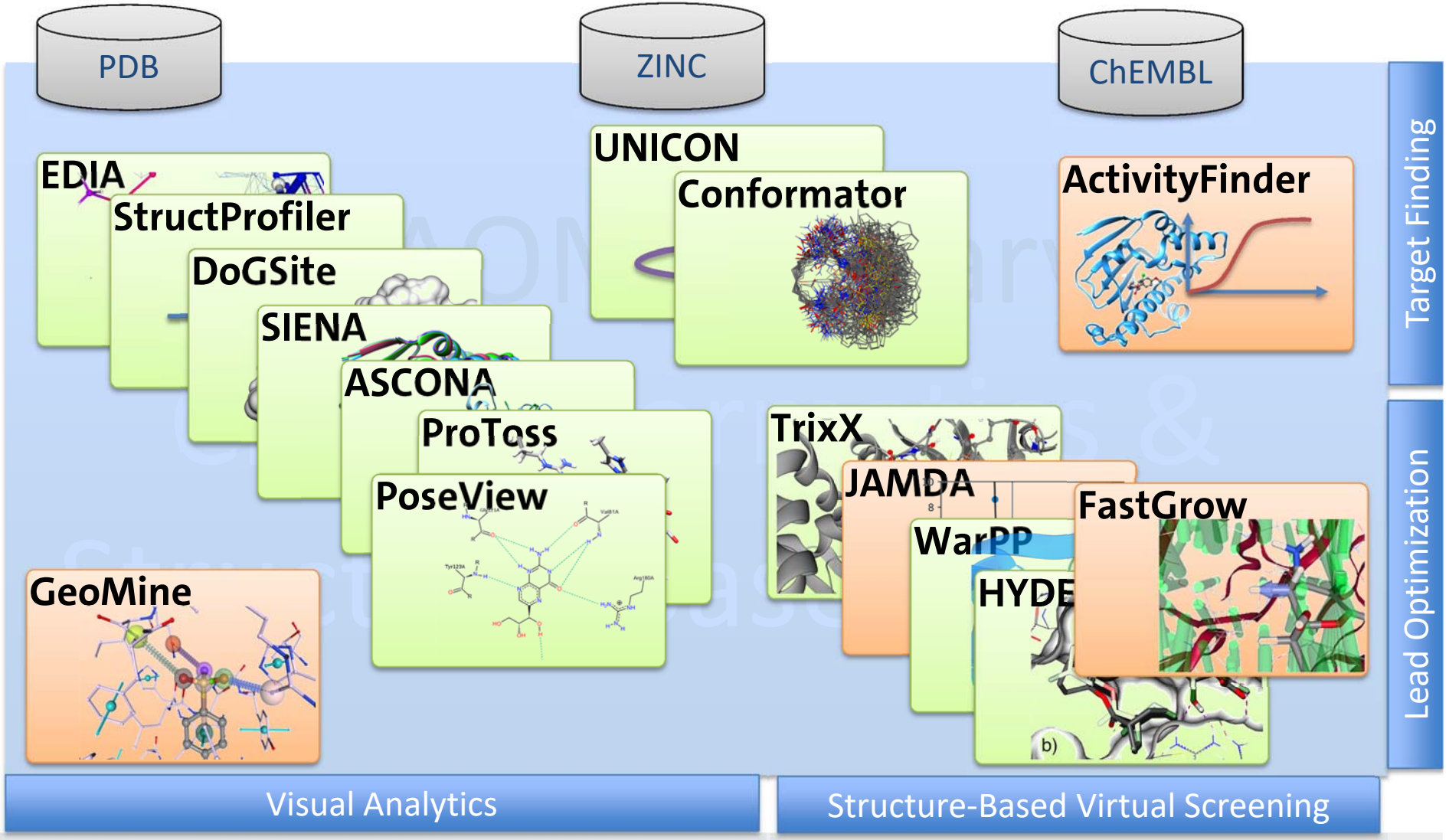




Molecular Design in a Nutshell: The First Phase

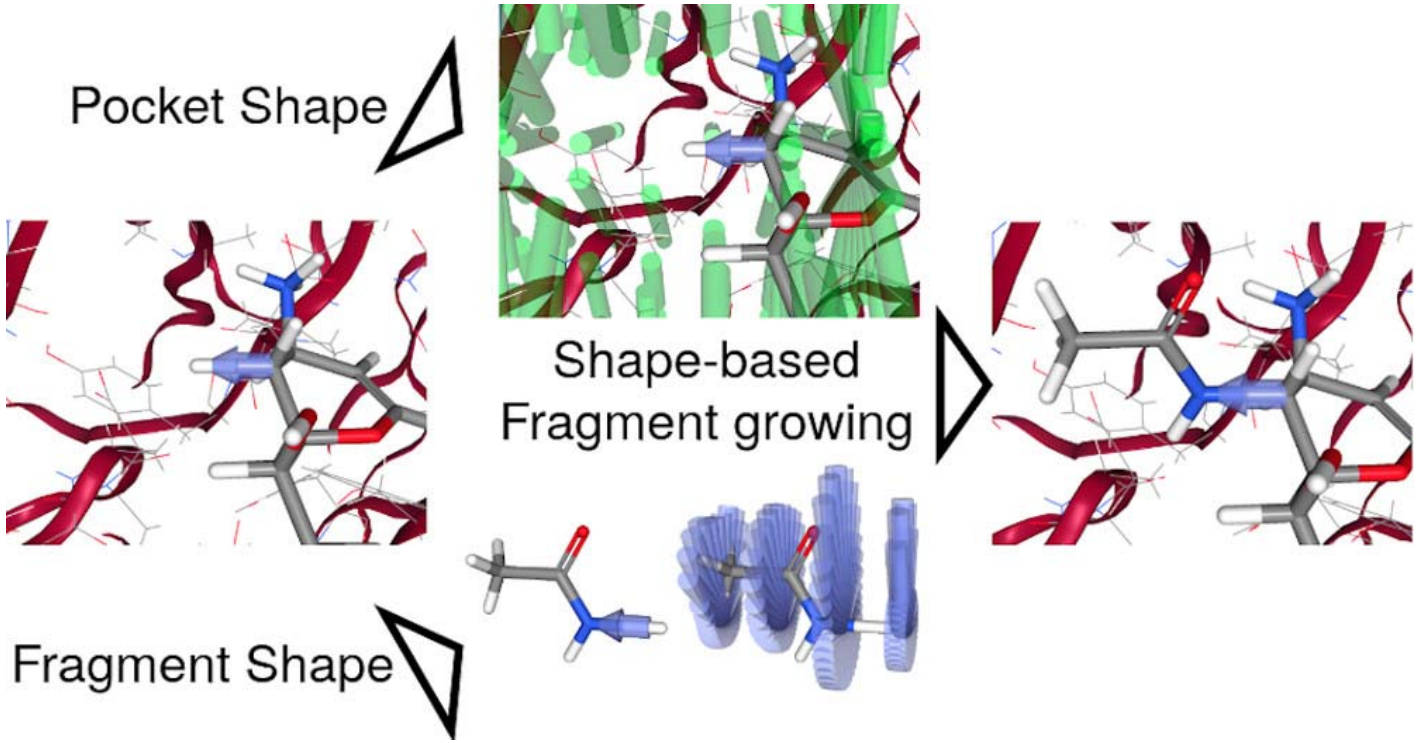


Protein Structure Processing and Exploitation



Penner et al., J. Chem. Inf. Model. (2020), **60**(12), 6269ff

Efficient Fragment Growing via RVM Descriptors



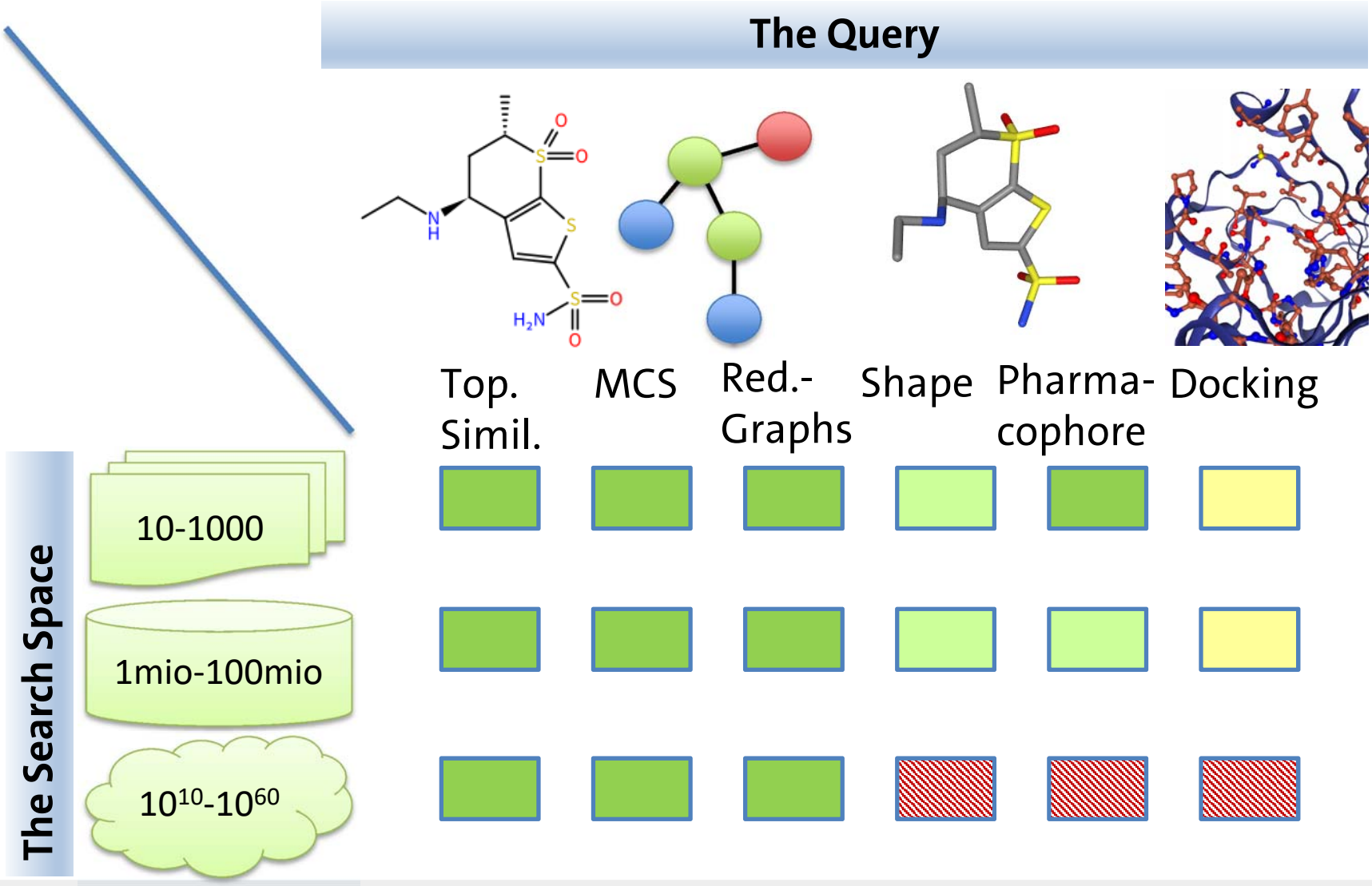
Accuracy (pose < 2Å RMSD)
self-growing: 84% (3299 cases)
cross-growing: 66% (326 cases)

Performance (i5 @ 3.4GHz, 16GB):
~10,000 conformers/sec

SERVIER

BioSolveIT
expect actives!

Molecular Design in a Nutshell: The First Phase



Contributions / Acknowledgements

current (alphabetic order): ***Louis Bellmann***, Ole Berg, Wolf-Guido Bolick, Theresa Cavasin, Konrad Diedrich, Uschi Dolfus, Emanuel Ehmki, Christiane Ehrt, Gerd Embruch, ***Florian Flachsenberg***, Joel Graef, Melanie Geringhoff, Michael Grössler, Torben Gutermuth, Tobias Harren, Sophia Hönig, ***Christian Meyenburg***, ***Patrick Penner***, Jonathan Pletzer-Zelgert, Thorben Puburski, Martin Poppinga, ***Robert Schmidt***, Katrin Schöning-Stierand, Jochen Sieg

former: Mathias von Behren, Stefan Bietz, Jörg Degen, Hans-Christian Ehrlich, Janna Eich, Rainer Fährrolfes, Robert Fischer, Nils-Ole Friedrich, Axel Griewel, Tim Harder, Lennart Heinzerling, Angela Henzler, Viola Hingst, Matthias Hilbig, Sarah Hofmann, Therese Inhester, Adrian Kolodzik, Tobias Lang, Florian Lauck, Tobias Lippert, Patrick Maaß, Agnes Meyder, Eva Nittinger, Juri Pärn, Thomas Otto, Ingo Reulecke, Christian Rhein, Christin Schärfer, Ingo Schellhammer, Jochen Schlosser, Nadine Schneider, Karen Schomburg, Birte Seebeck, Kai Sommer, Katrin Stierand, Sascha Urbaczek, Andrea Volkamer, Björn Windshügel

Software Development Partner:
BioSolveIT [Christian Lemmen, Marcus Gastreich, ***Raphael Klein***, Alexander Neumann & the whole team]

PROTEINS PLUS

Funded by bmbf as part of de.NBI
(Grant ID: 031L0105)

Thanks for Funding + Cooperation, currently:
3x BMBF, Helmholtz Association, Merck, Servier, Bayer

Additional Reading and Software Availability

NAOMI ChemBio Suite:
http://uhh.de/naomi

Modeling Support Server:
http://proteins.plus

SMARTS Pattern Analyzer Server
http://smarts.plus

Professional tool collection:
http://www.biosolveit.com