

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

Cheminformatics Operating in Chemical Space

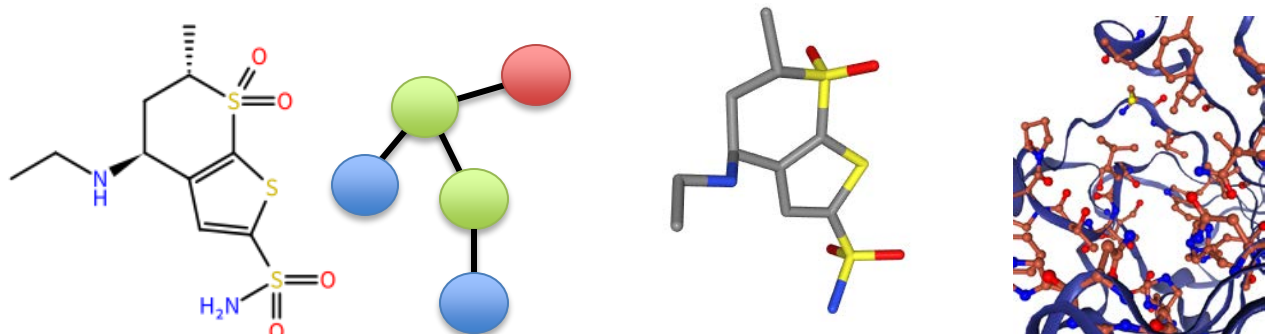
MATTHIAS RAREY

Research Group for Computational Molecular Design

Forschungsgruppe Algorithmisches Molekulares Design (AMD)

Early-Phase Drug Discovery

The Query



Top. Simil. MCS Red.-Graphs Shape Pharma-cophore Docking

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

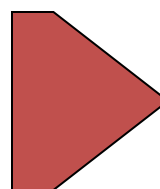
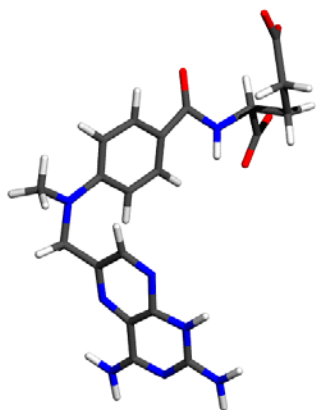
The Search Space

10-1000

1mio-100mio

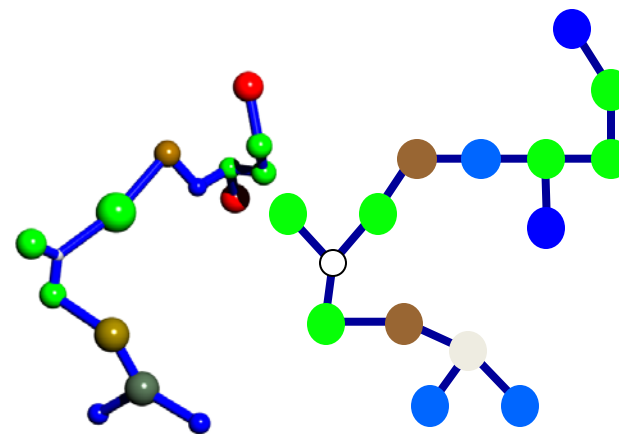
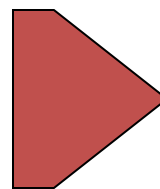
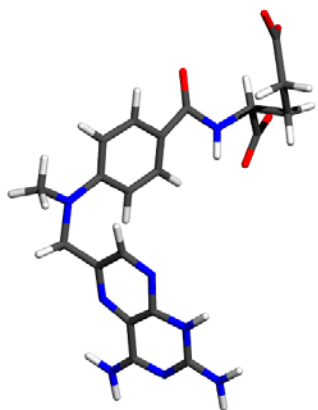
10^{10} - 10^{60}

Feature Trees (1998)

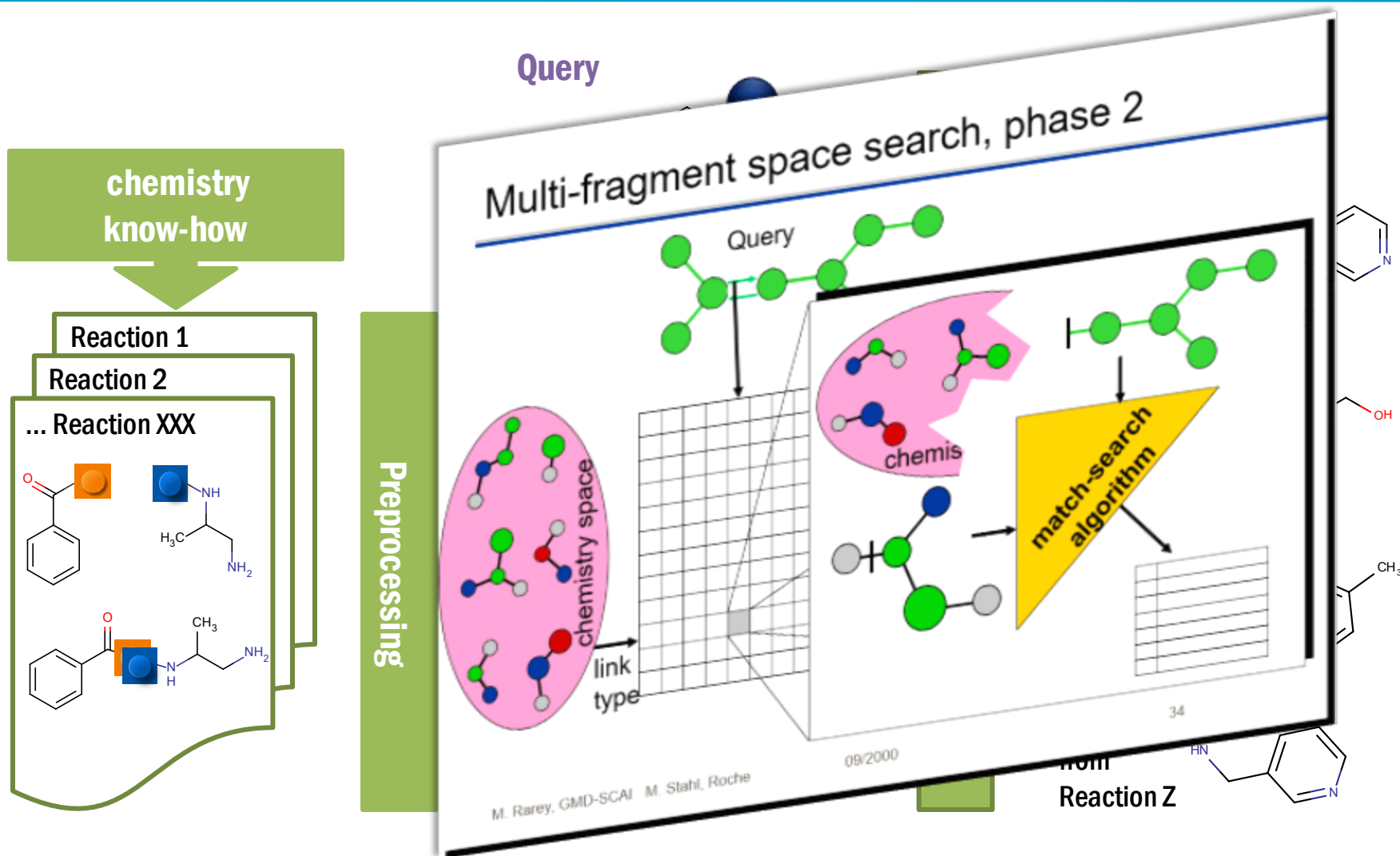


01101001011101001110101

■ Non-linear descriptors: Feature Trees



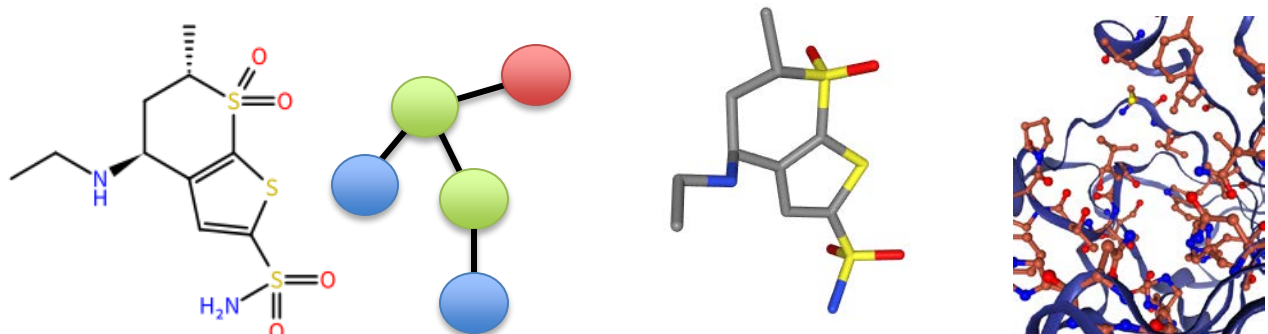
FTrees Fragment Spaces: Searching Instead of Screening (2001)



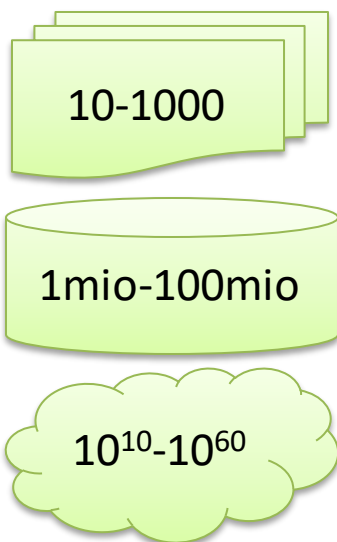
Courtesy of Franca Klingler, BioSolveIT GmbH

Early-Phase Drug Discovery

The Query

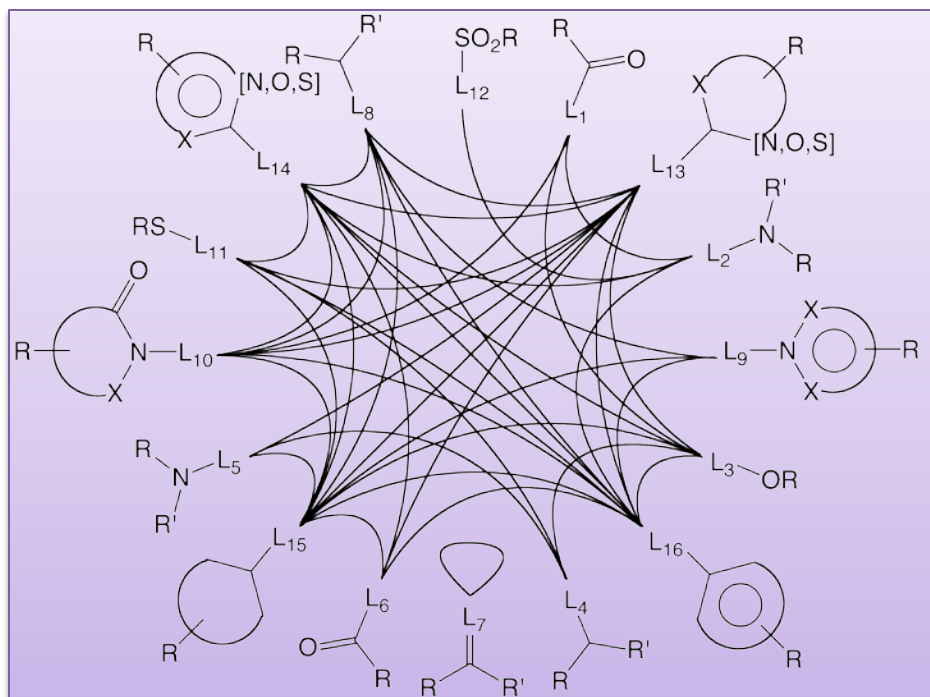


Top. Simil. MCS Red.-Graphs Shape Pharma-cophore Docking



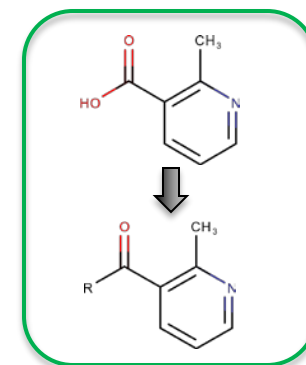
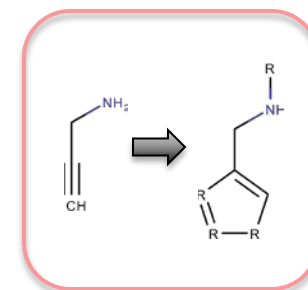
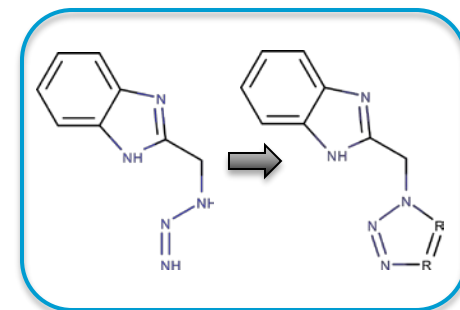
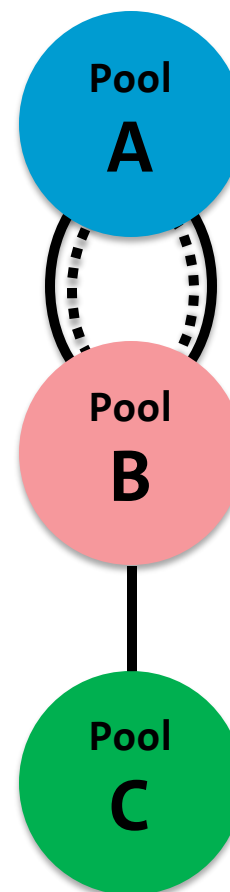
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

How We Describe Fragment Spaces

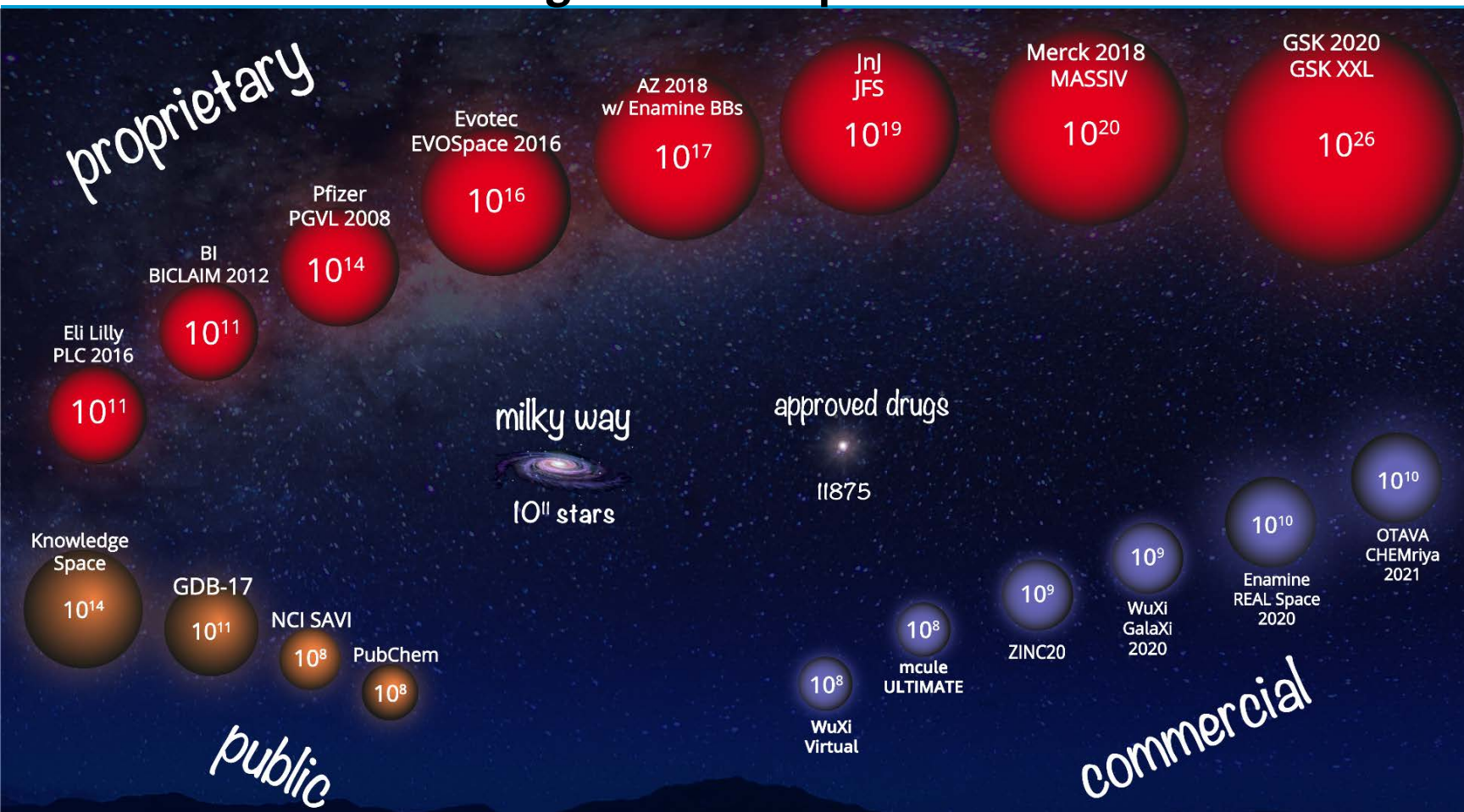


BRICS: Degen, Gerlach, Zaliani & Rarey, *ChemMedChem*, **3**, 1503 (2008)

Topology Graph

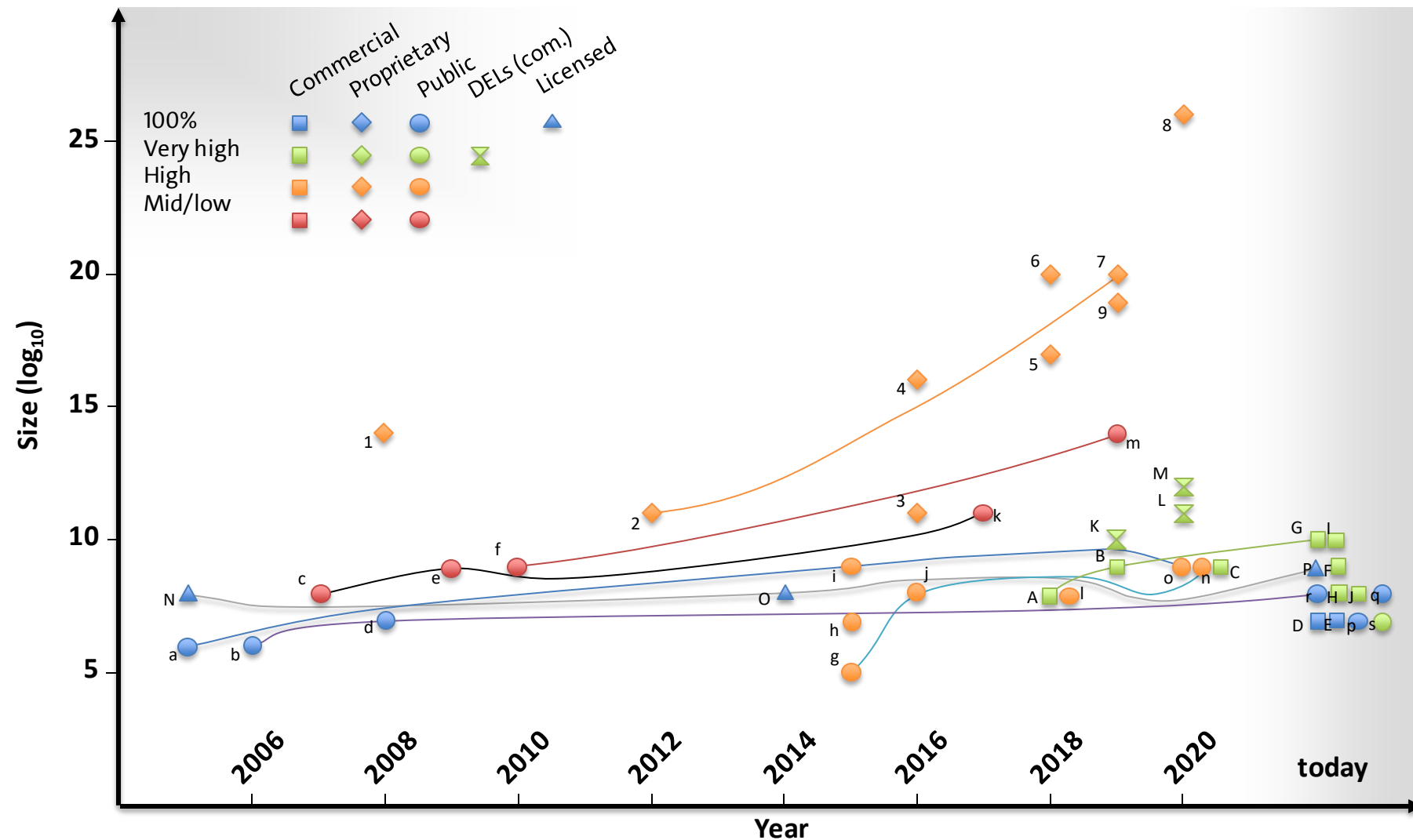


Advances in Modeling Chemical Space

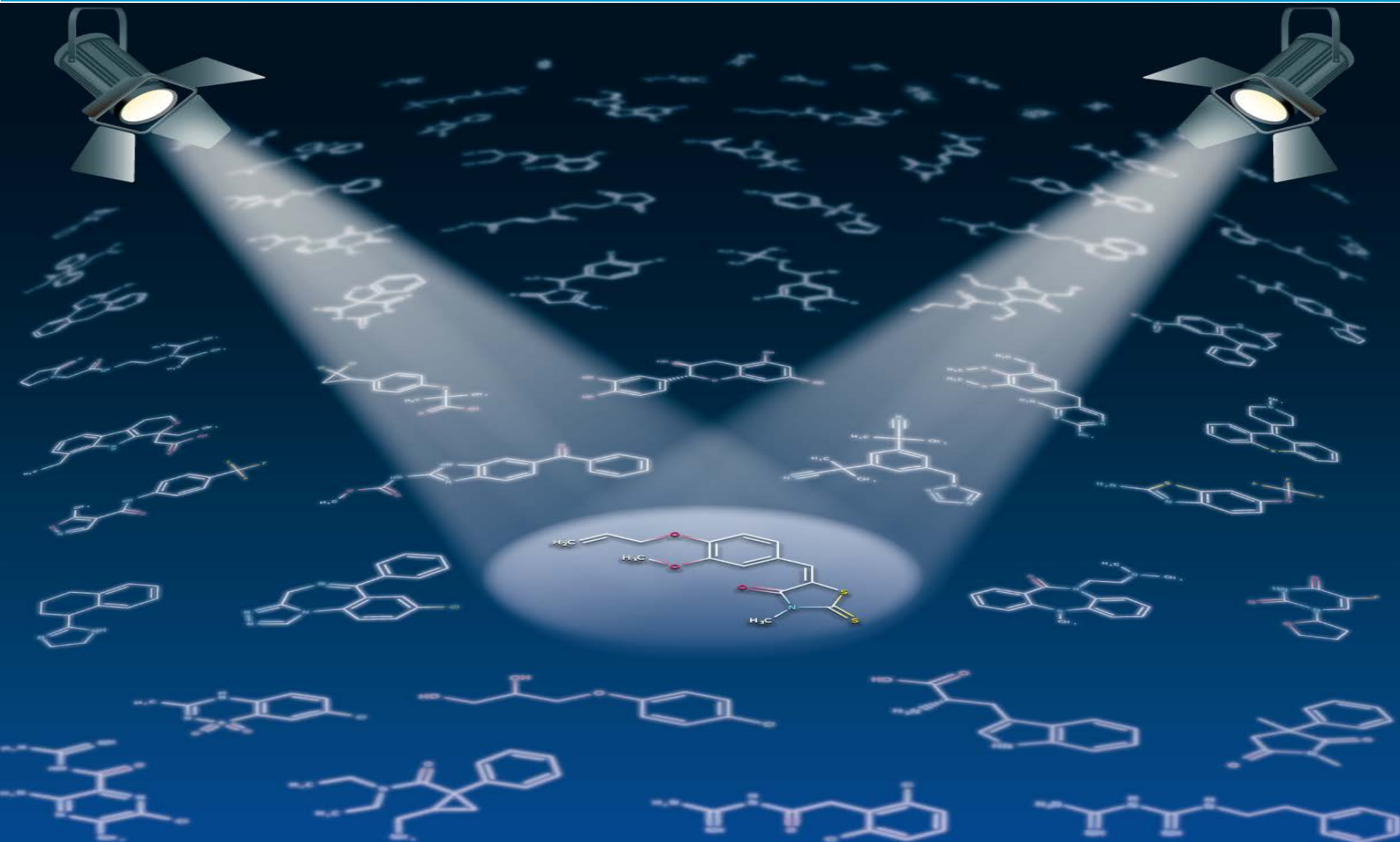


Adapted and updated from Hoffmann et al, DDT (2019), 24, pp 1148

Chemical Space over Time

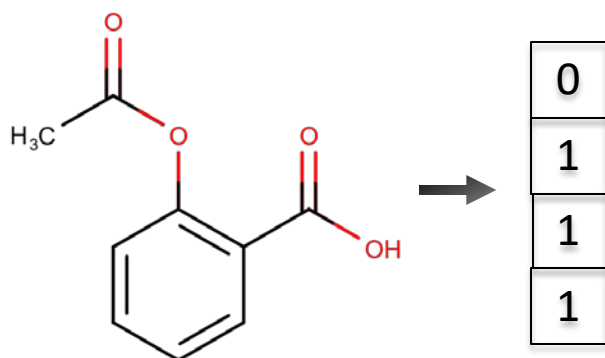


SpaceLight – Topological Searching in Chemical Space



Classic Molecular Fingerprints in a New Light

Employ classic fingerprints in a new search approach



Notation

ECFP_x: x max diameter for circular features

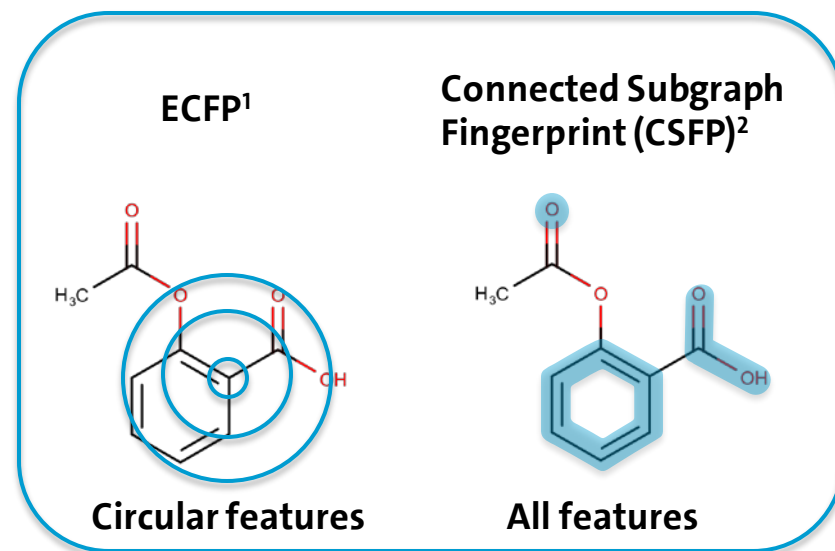
CSFP_{y.z}: y lower, z upper bound for #heavy atoms

CSFP versions

fCSFP: fine-grained similarity measurement

iCSFP: MCS-like descriptor

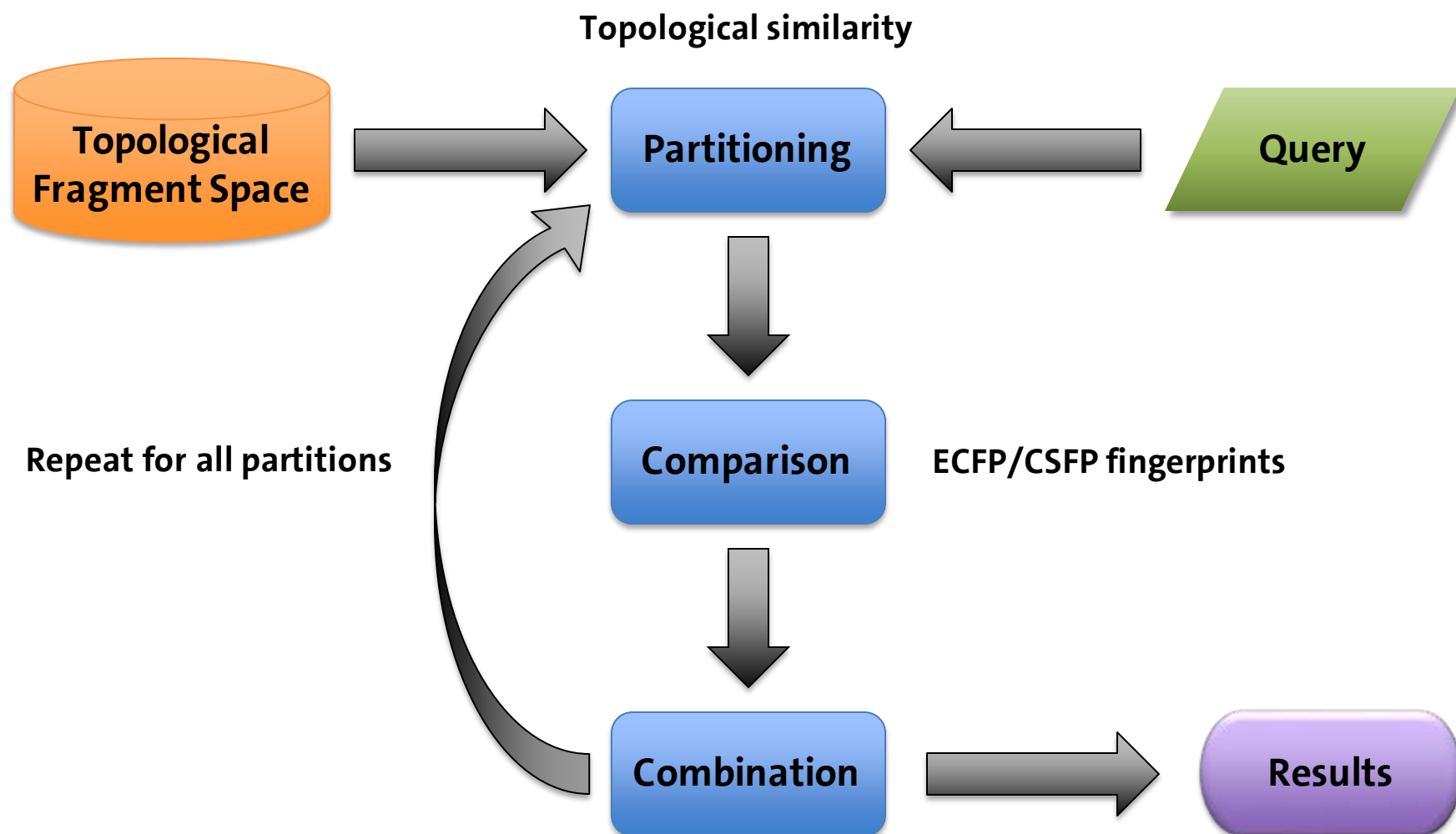
tCSFP: scaffold-hopping potential



1: Rogers and Hahn, J. Chem. Inf. Model. (2010), 50

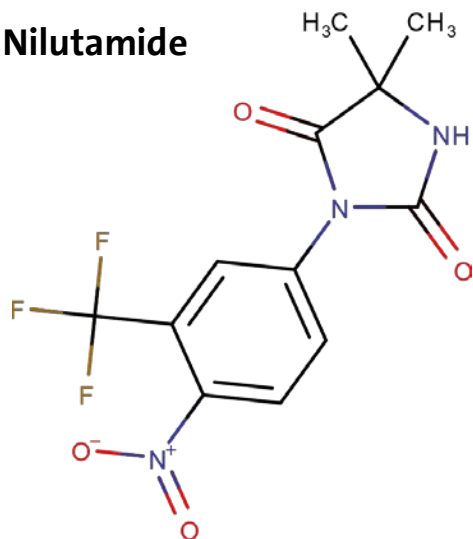
2: Bellmann et al., J. Chem. Inf. Model. (2019), 59, 11

SpaceLight Algorithm Overview



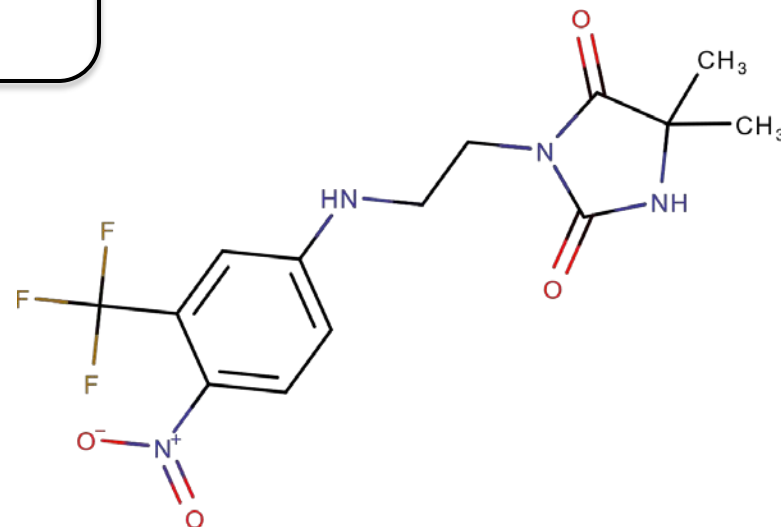
SpaceLight Search for Drugs in Enamine REAL Space

Nilutamide



Found with score 1.0

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

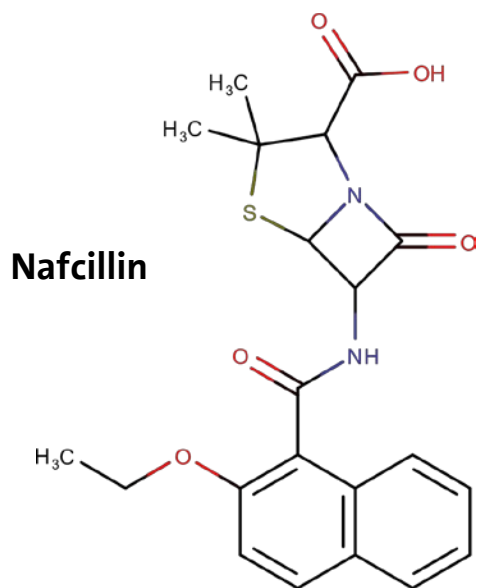


Second best with score 0.87

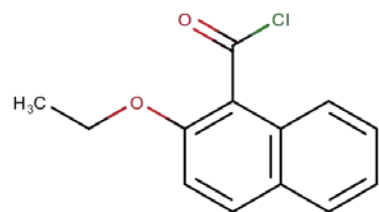
SpaceLight Gives You Ideas for Synthesis Routes

SpaceLight retrieves synthetic pathways of products

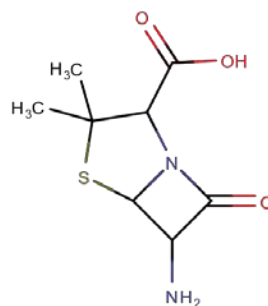
Knowledge Space
 120 synthesis protocols
 >142000 building blocks
 >10¹⁵ products



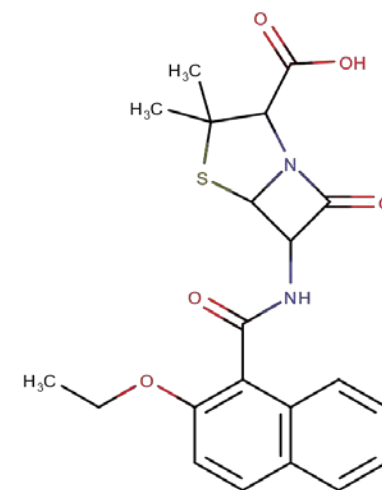
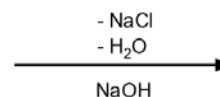
Found with score 1.0



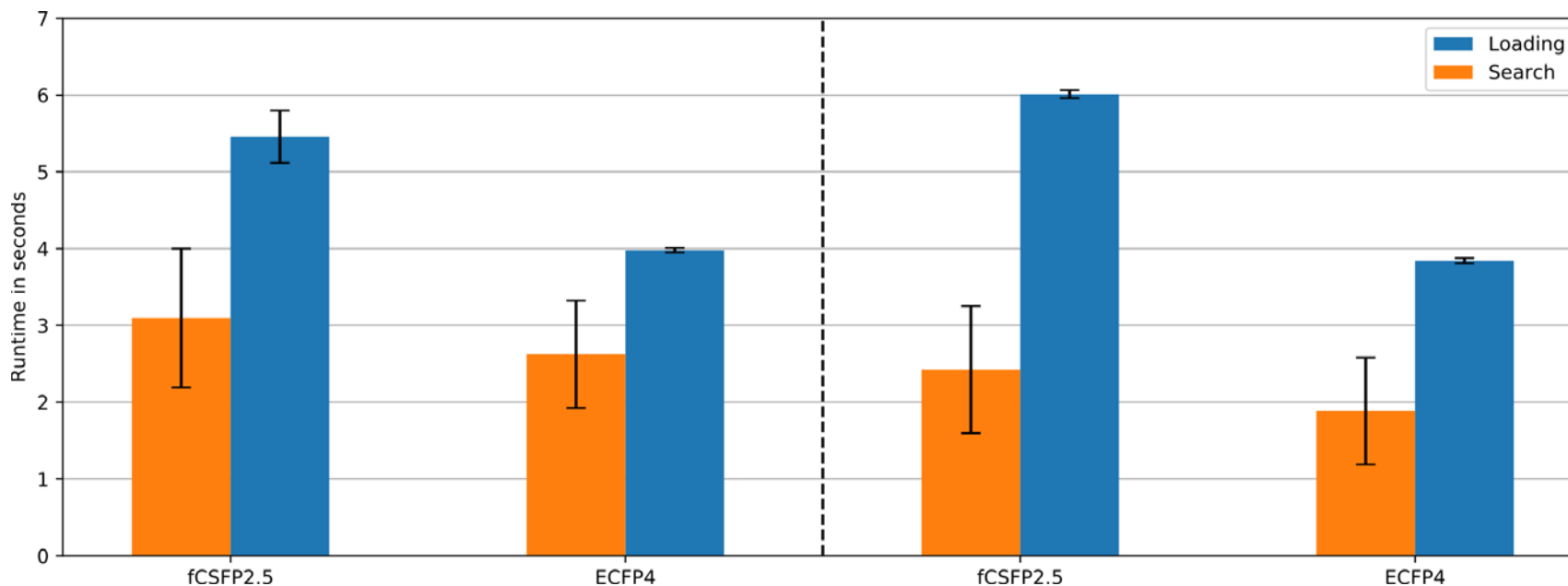
+



Schotten-Baumann



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



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

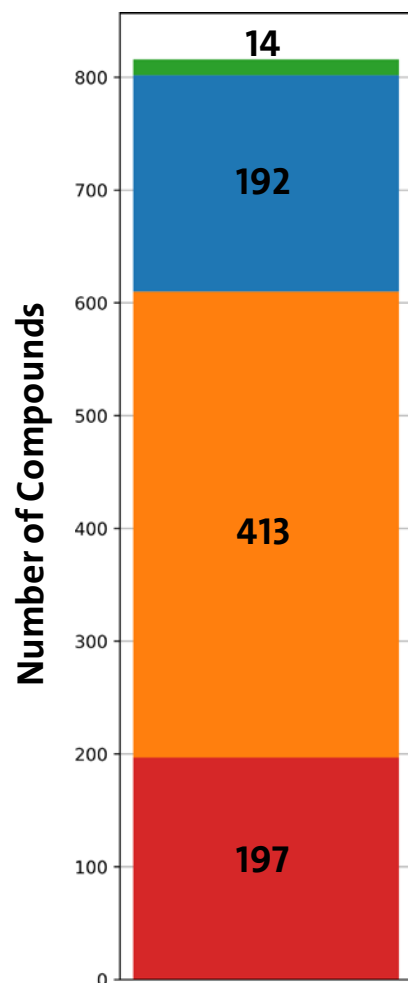
Intel Core i5-6500 @3.2GHZ
 16 GB memory
 Search with 3 threads

500 random queries
 from ZINC lead-like

Knowledge Space¹
 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 Space



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

Queries from DrugBank¹
1. Approved drugs
2. Molecular weight
between 300 and 500
=> 816 compounds

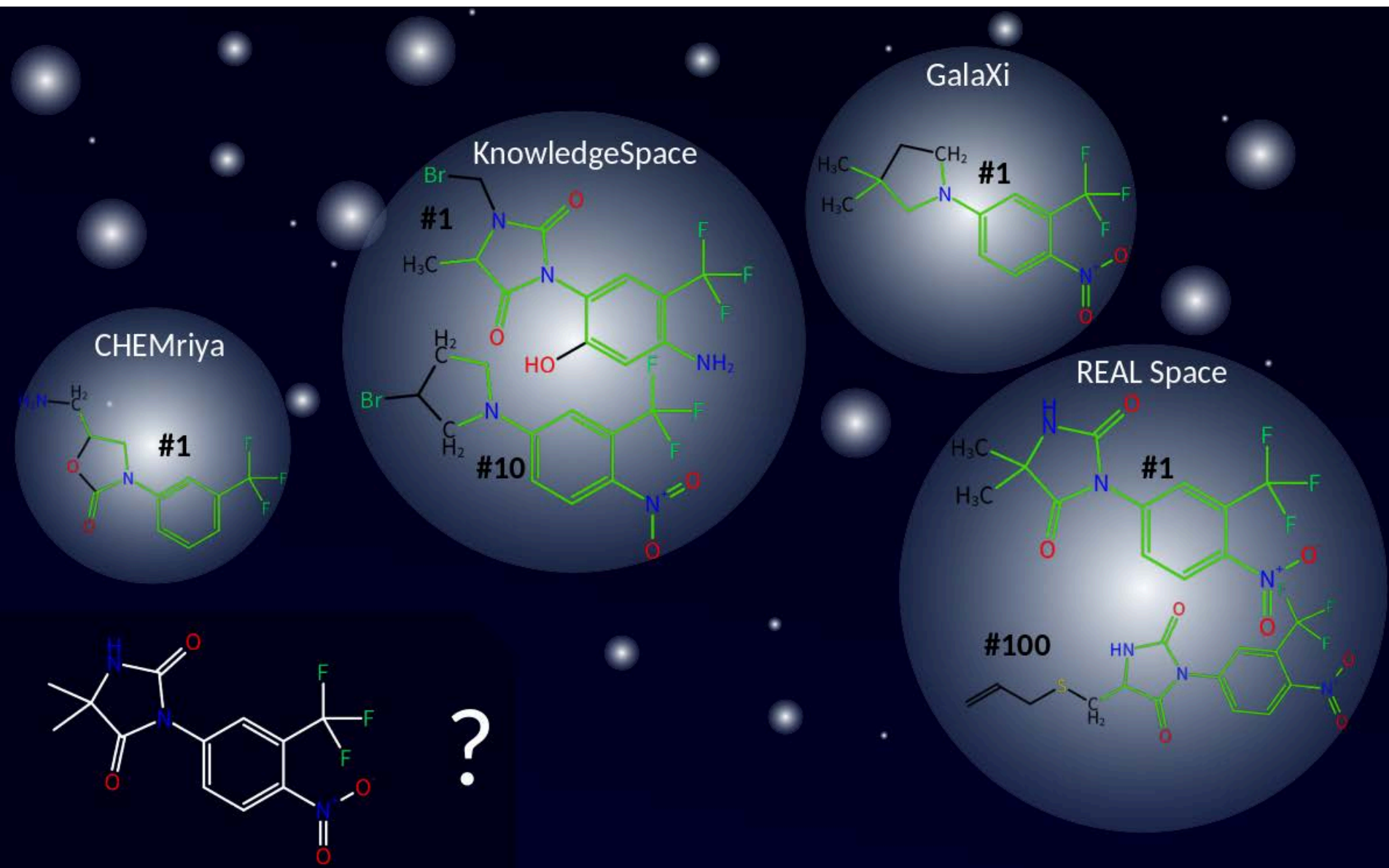
SpaceLight_{fCSFP2.5}

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

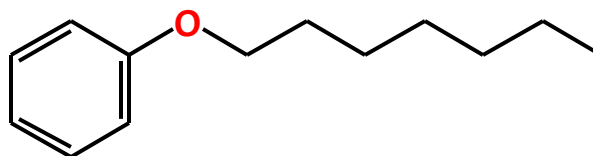
SpaceMACS – (Maximum Common) Substructure Searching in Chemical Space



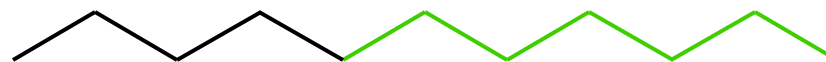
Maximum Common Substructure Search in Chemical Space



- connected MCS
- acyclic/acyclic bond matching



Qsize: 14 atoms



Tsize: 11 atoms

Msize: 7 atoms

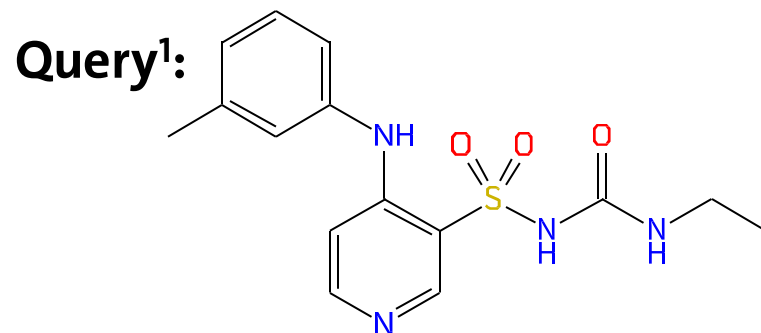
Hit list	

Sort criteria:

1. Msize | Tsize
2. MCS-Similarity:

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

A First Run: Why MCS-Similarity?

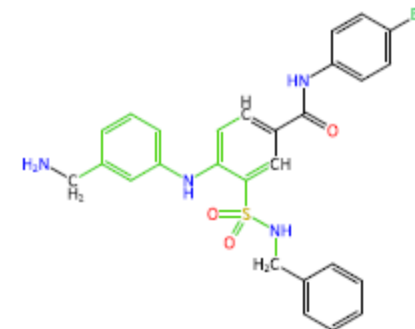


Space²: GalaXi-Space 2020-11
2.1 bn compounds

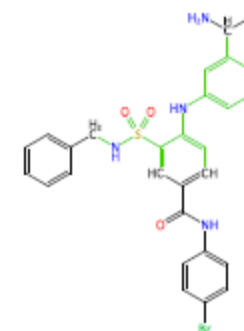
Hit list:

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

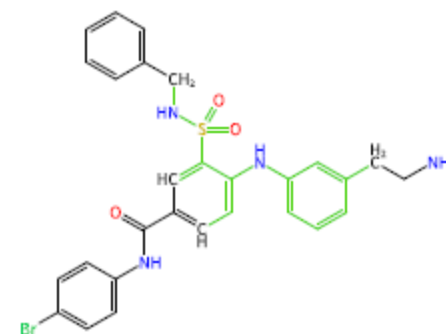
1.



2.



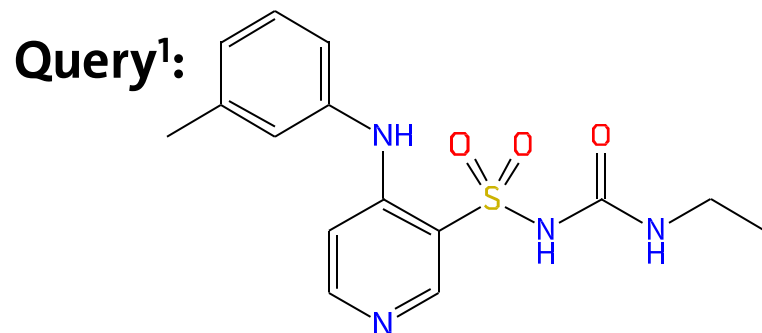
3.



¹Lessel & Lemmen, ACS Med.Chem.Lett. (2019), 10(10), 1504ff

²https://www.biosolveit.de/download/?file=2.1bn-GalaXi_2020-11.space

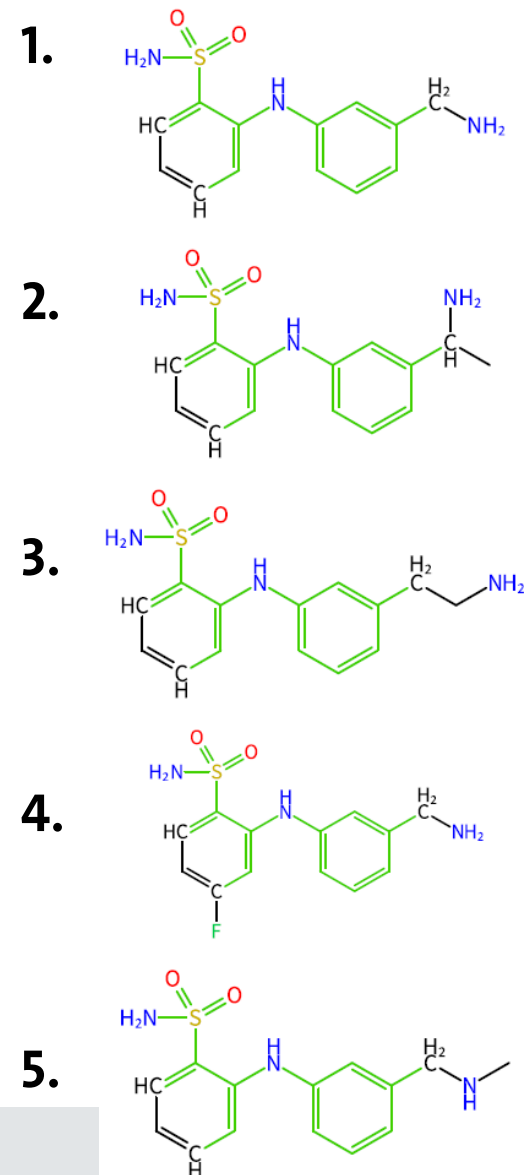
A First Run: Why MCS-Similarity?



Space²: GalaXi-Space 2020-11
2.1 bn compounds

Hit list:

(by MCS-Similarity (#M=17))



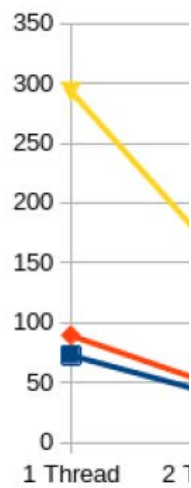
¹Lessel & Lemmen, ACS Med.Chem.Lett. (2019), 10(10), 1504ff

²https://www.biosolveit.de/download/?file=2.1bn-GalaXi_2020-11.space

Computing Times and Speed Up

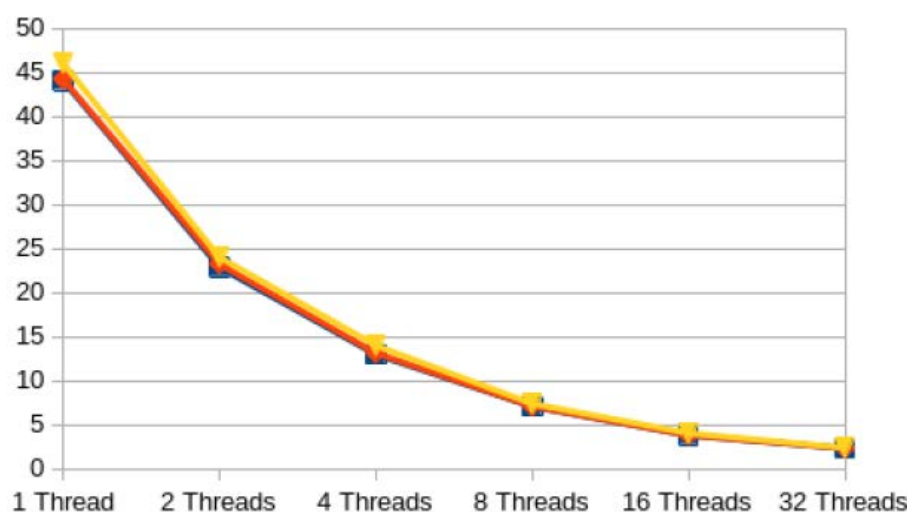
Mean REALspace Runtimes in s

Similarity



Mean REALspace Runtimes in s

Subset



■ 100 Results
◆ 1000 Results
▲ 10000 Results

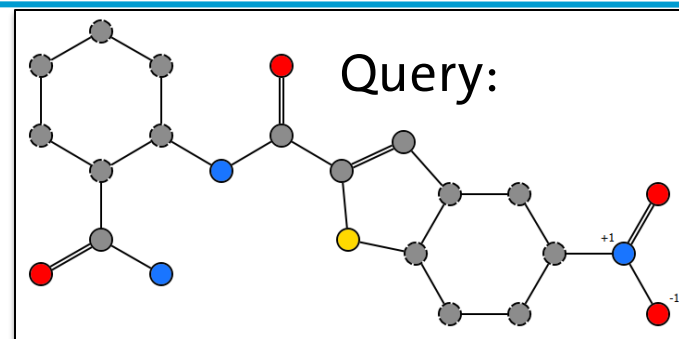
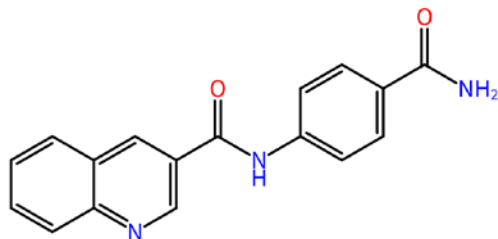
100 queries taken from [1],
 searching in Enamine REALspace² 2020-11 (15 bn compounds)

¹Lessel & Lemmen, ACS Med.Chem.Lett. (2019), 10(10), 1504ff

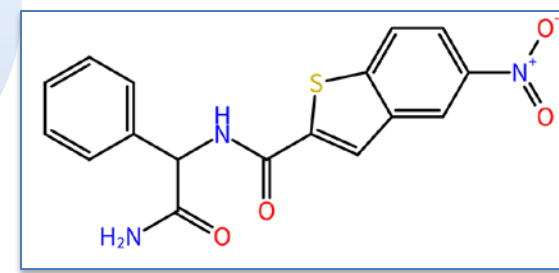
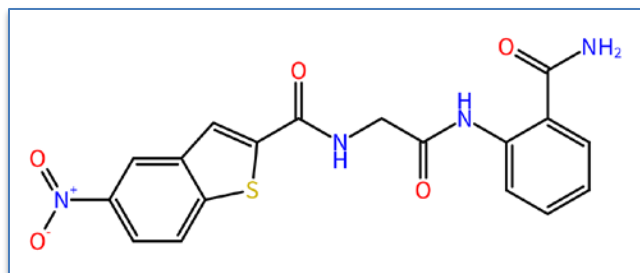
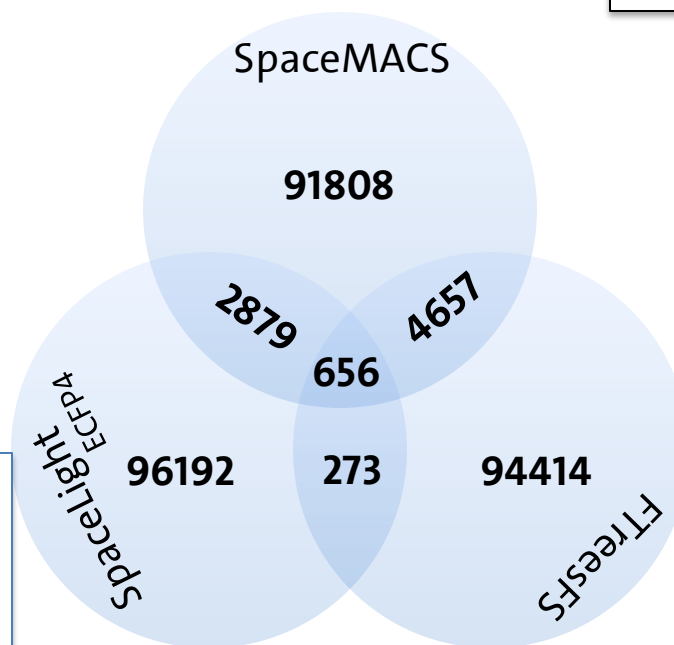
²<https://www.biosolveit.de/infiniSee/#realspace>

Overlap: SpaceMACS – SpaceLight – FTreesFS

- 4 Molecules only found by FTreesFS:

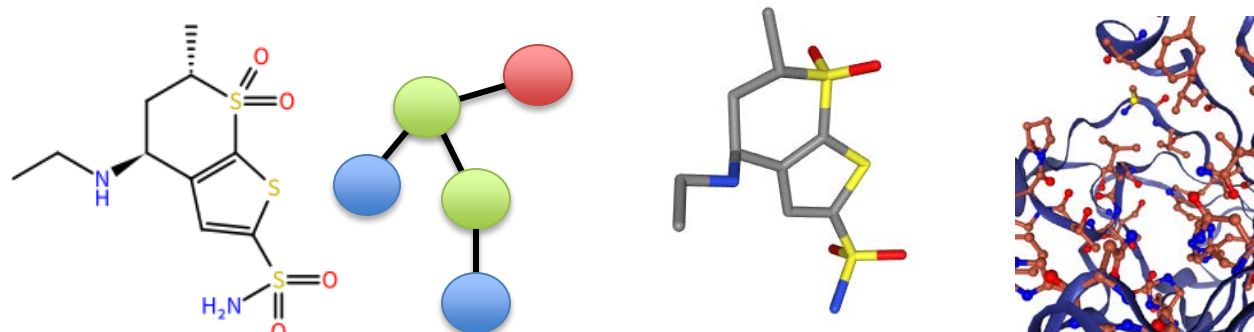


- Hit list overlap:



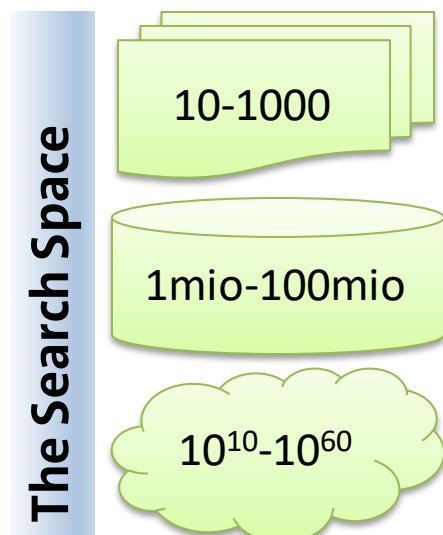
Early-Phase Drug Discovery

The Query

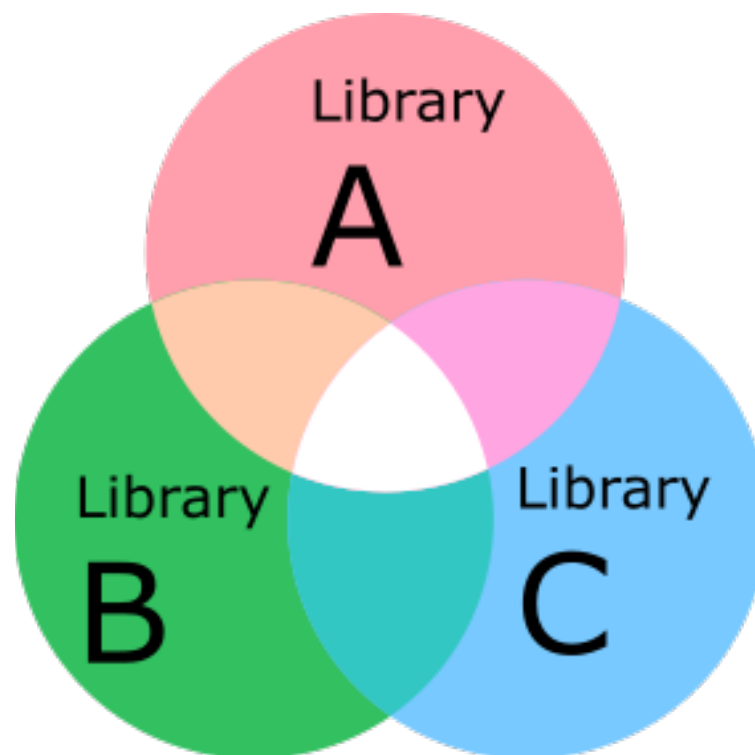


Top. Simil. MCS Red.-Graphs Shape Pharma-cophore Docking

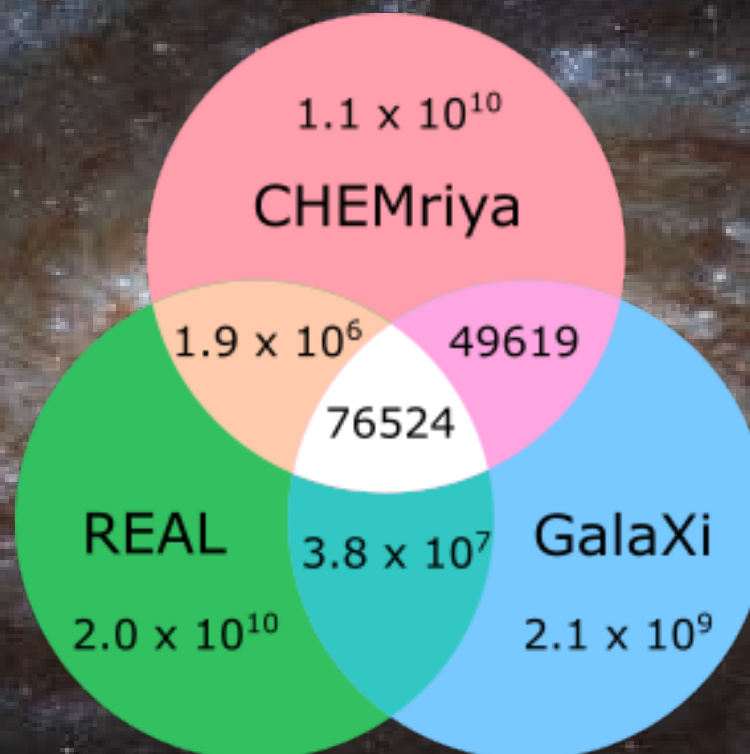
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



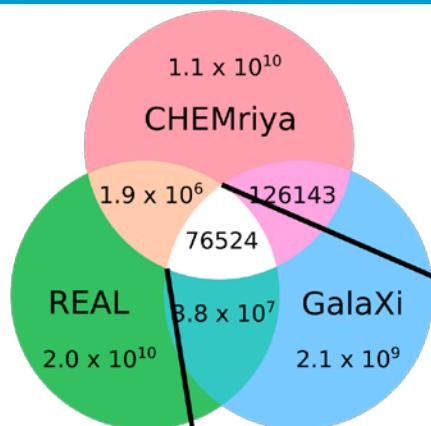
Tools to Analyze and Design Chemical Fragment Spaces



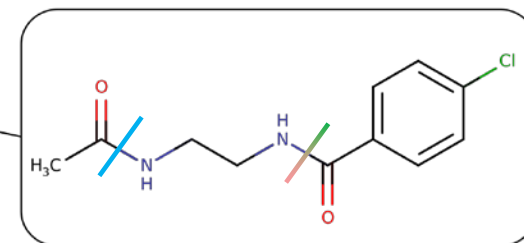
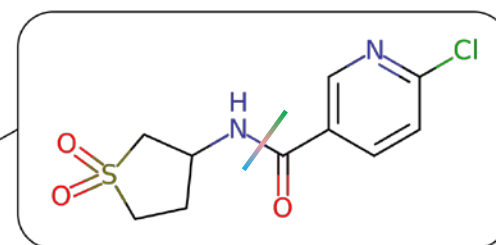
Tools to Analyze and Design Chemical Fragment Spaces



Vendors Mostly Agree on Synthesis in Overlap

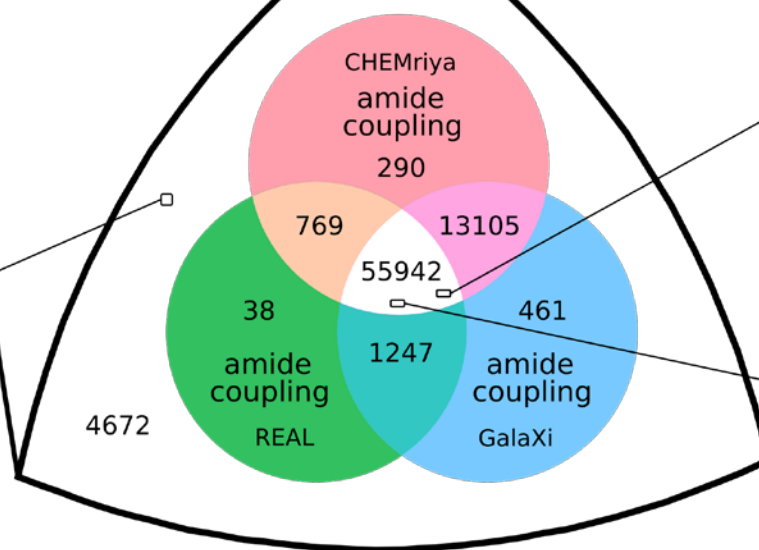
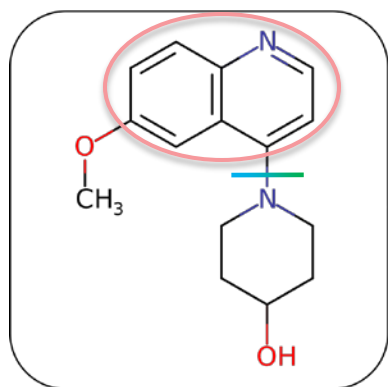


Amide coupling with same building blocks

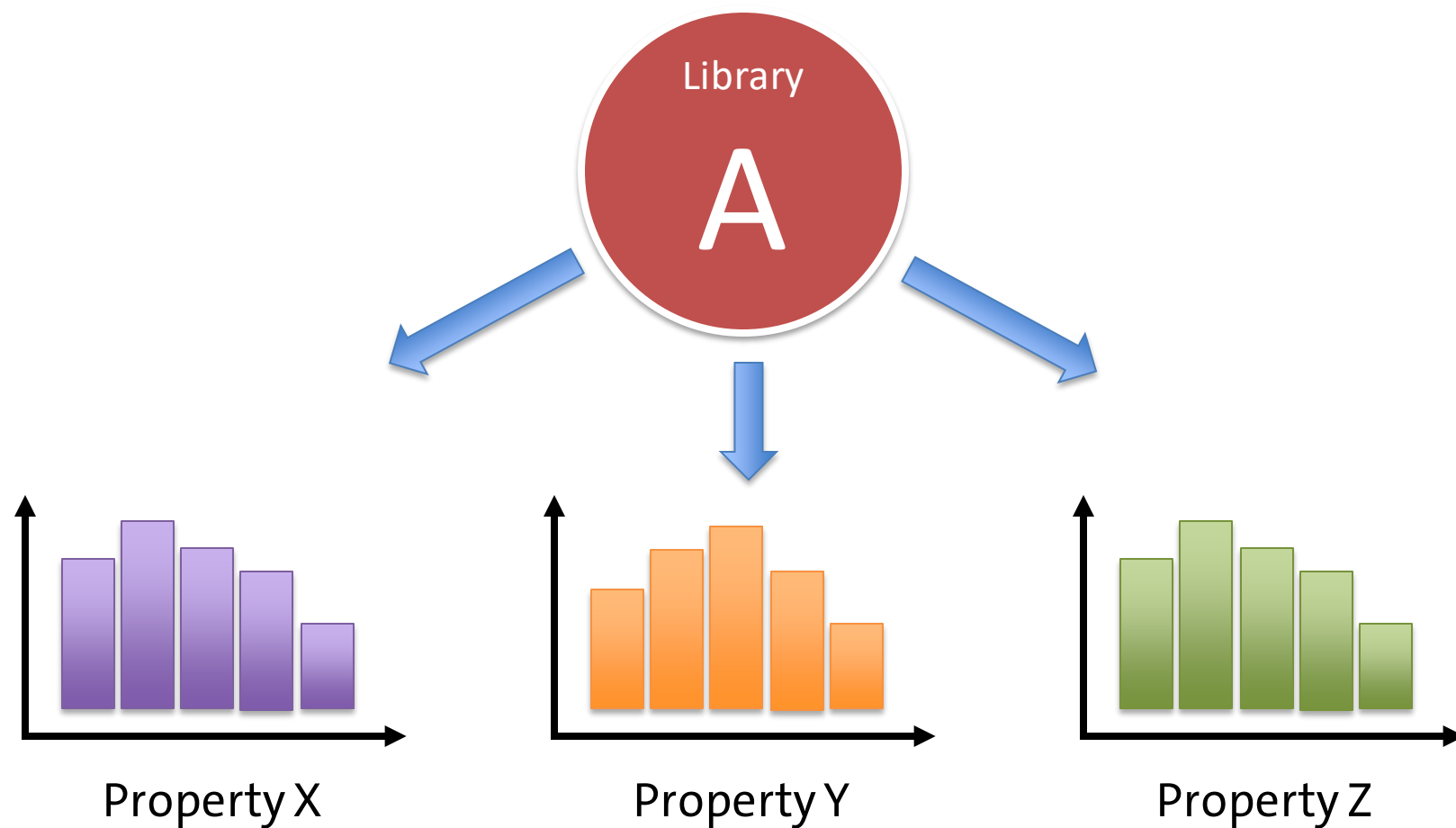


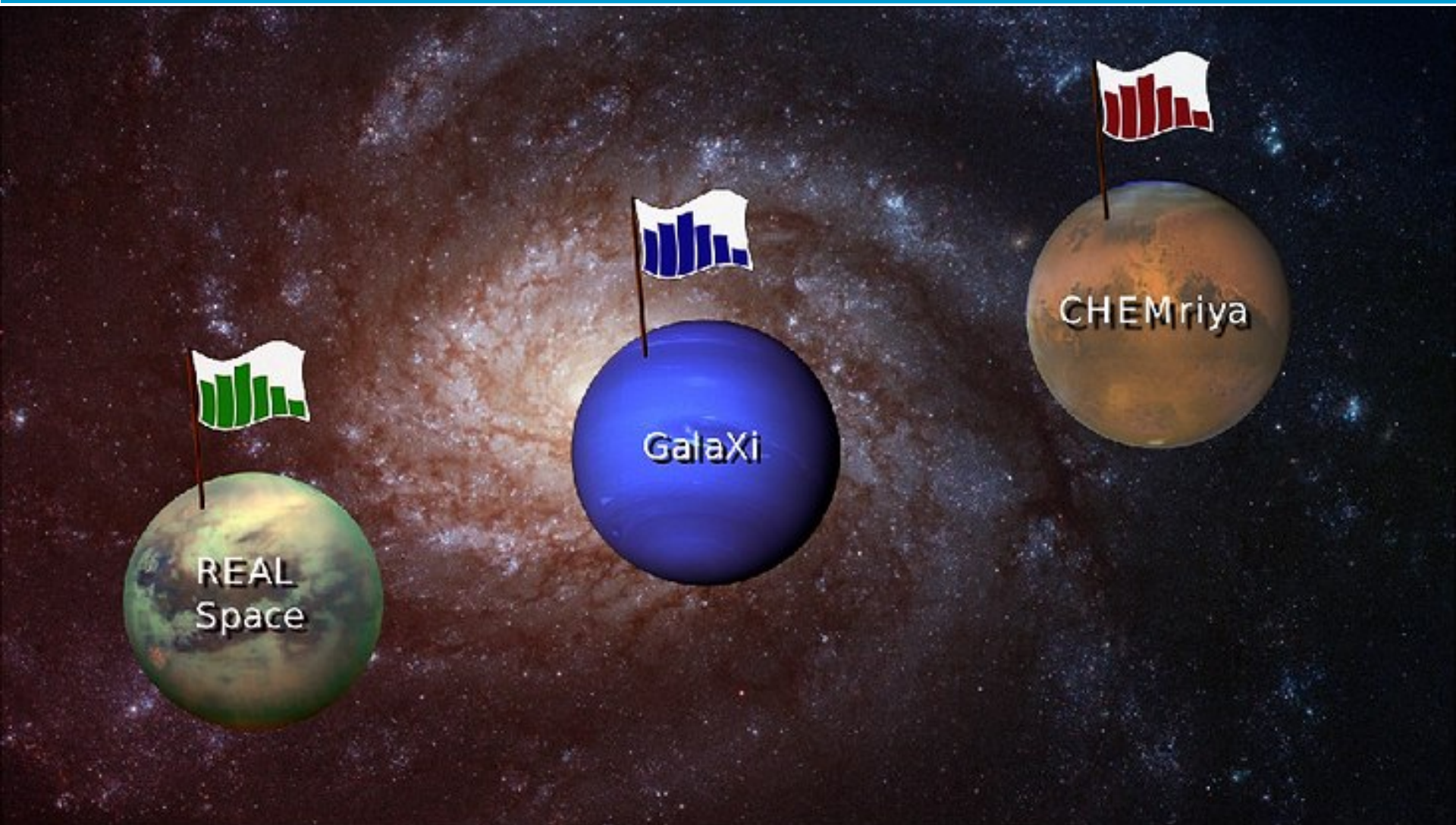
Amide coupling with different building blocks

Different reactions for synthesis



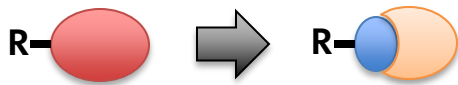
Tools to Analyze and Design Chemical Fragment Spaces



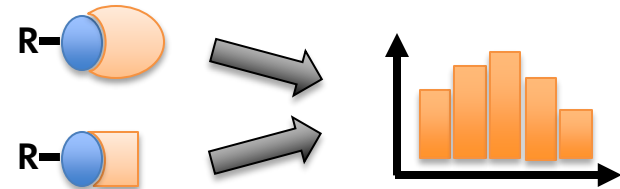


SpaceProp Algorithm Outline

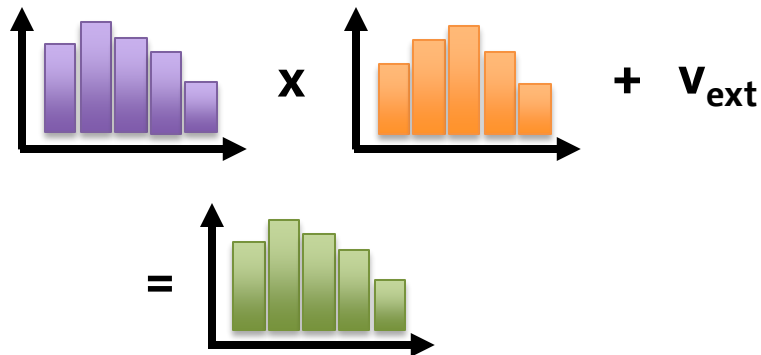
1: Calculate Internal/External Components



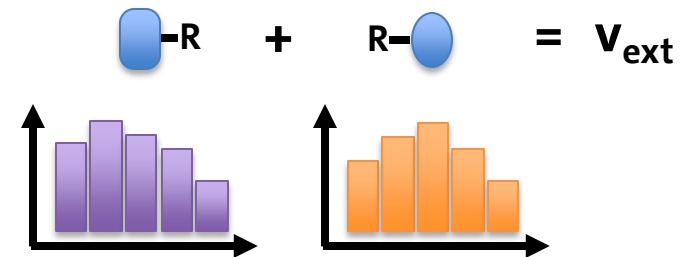
2: Pool Fragments with Same External Component



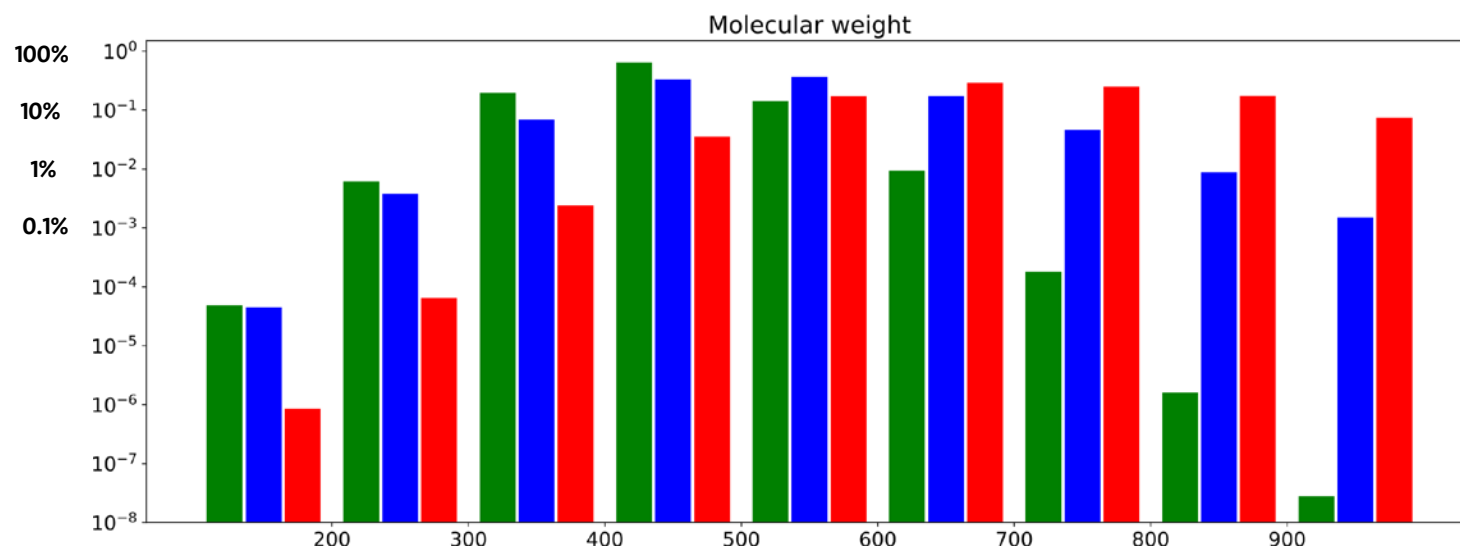
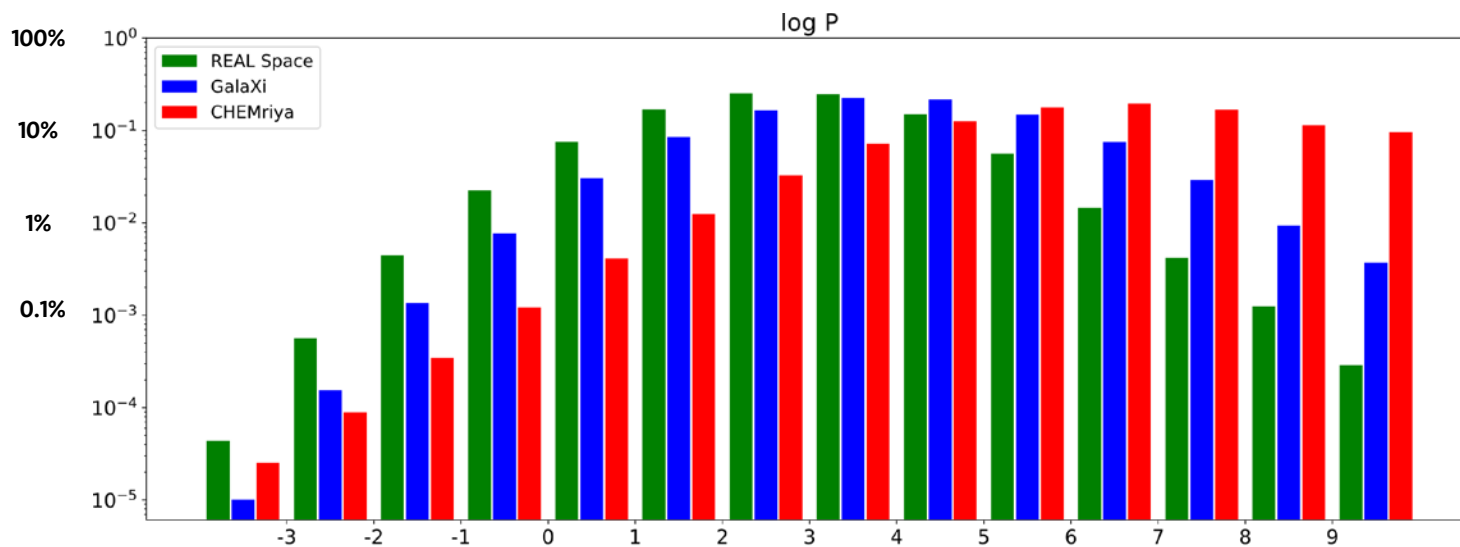
4: Combine Distributions



3: Resolve External Components

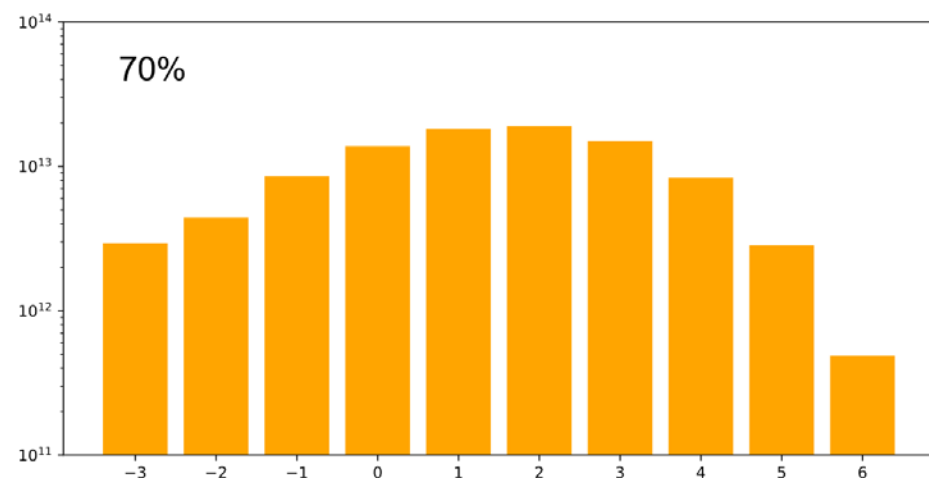
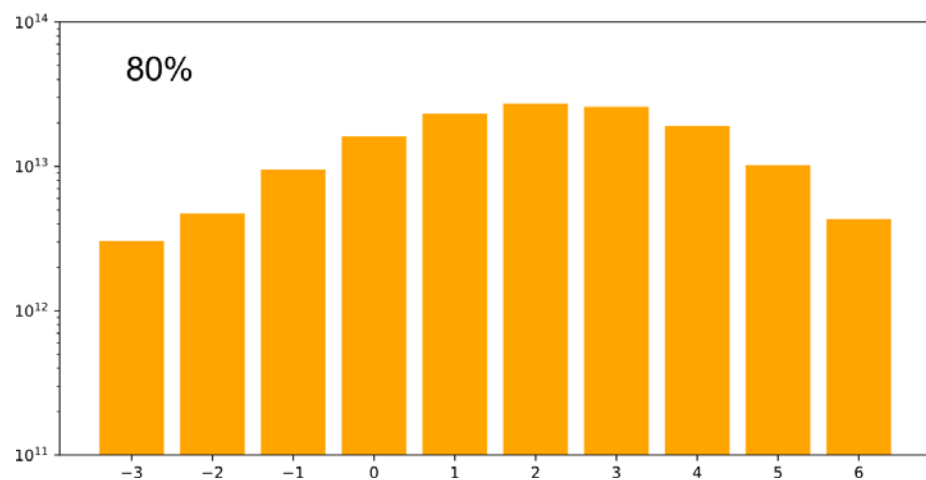
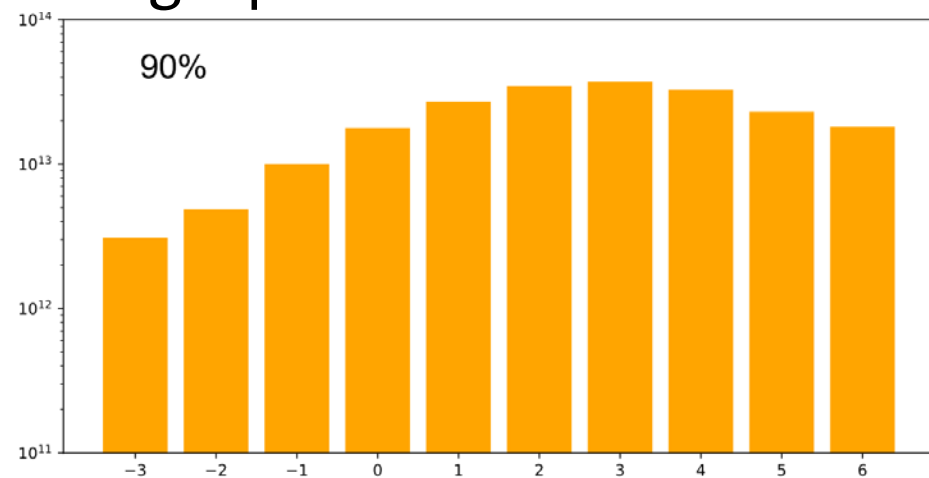
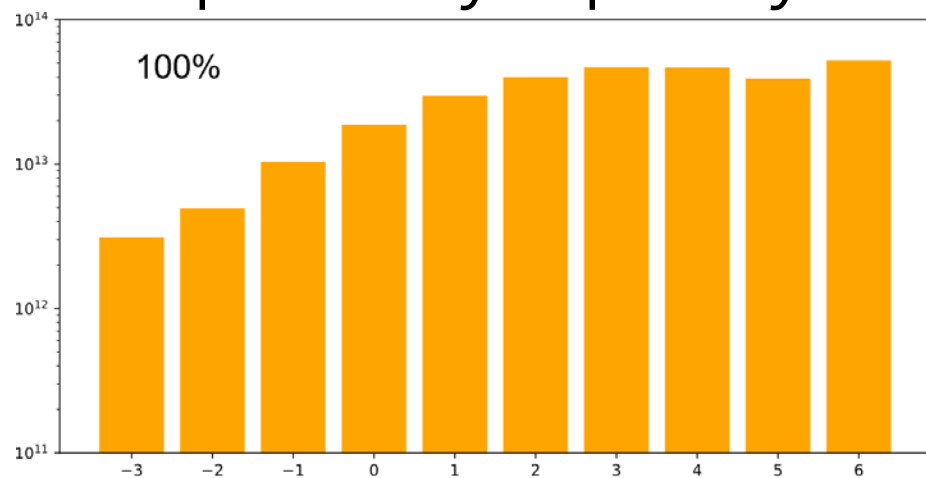


Property Distributions of Ultra-Large Vendor Catalogs



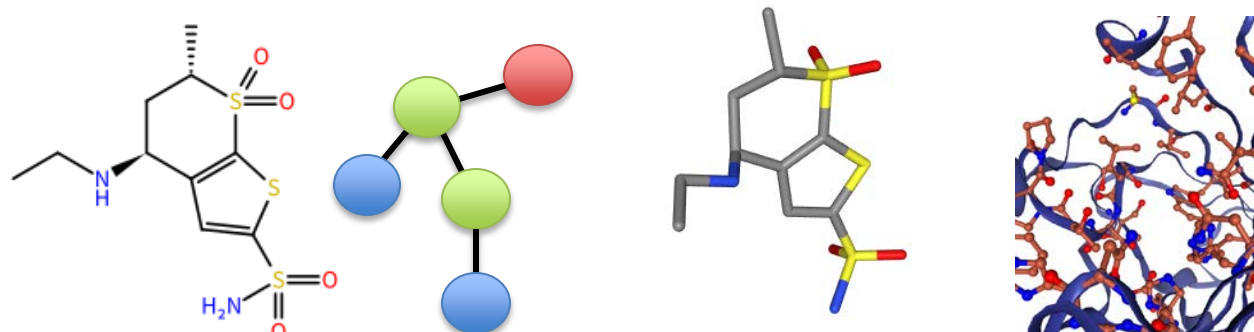
SpaceProp: Property Distribution in Fragment Spaces

■ Optimize Hydrophilicity of KnowledgeSpace



Molecular Design in a Nutshell: The First Phase

The Query



Top. Simil. MCS Red.-Graphs Shape Pharma-cophore Docking

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

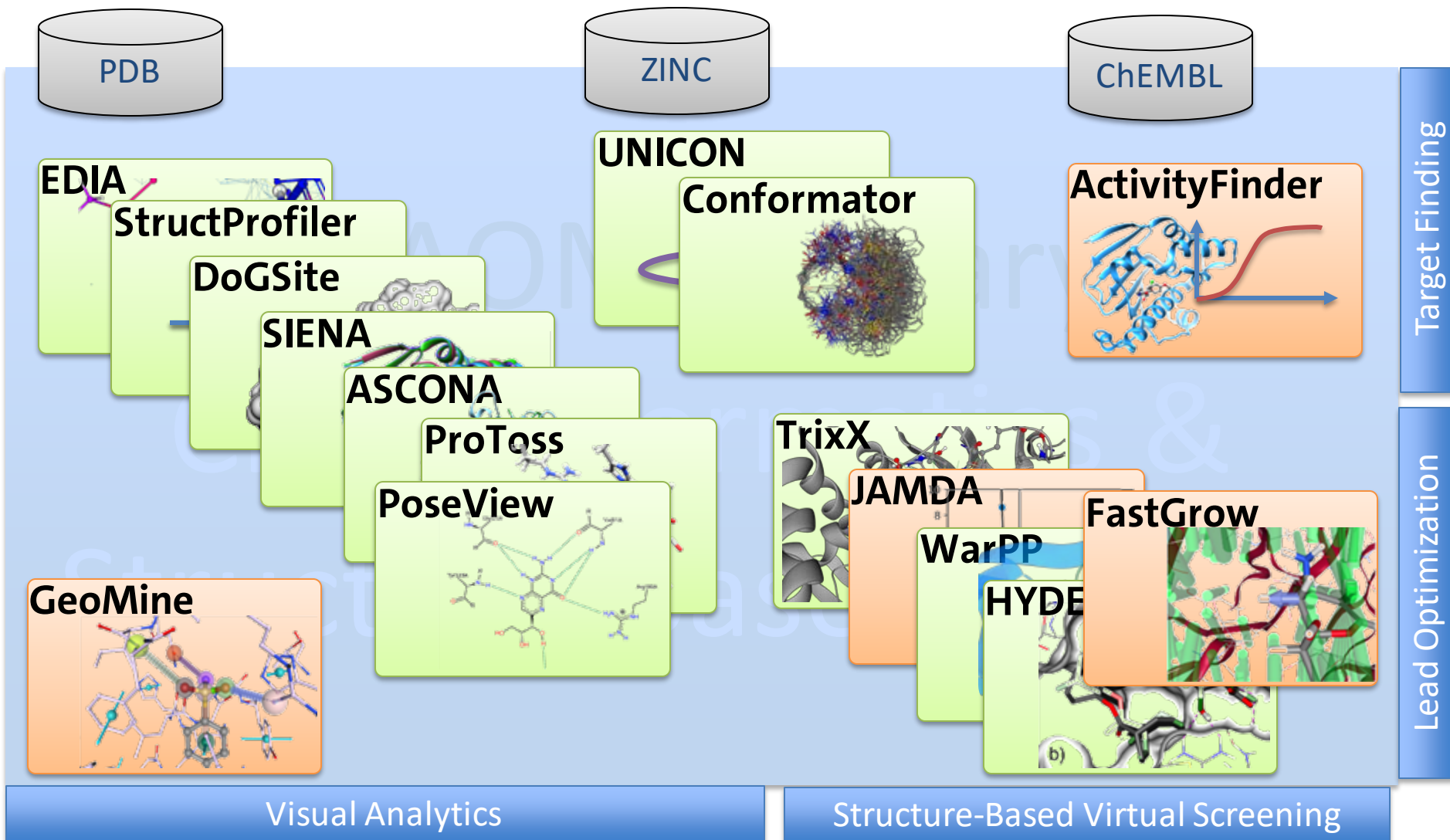
The Search Space

10-1000

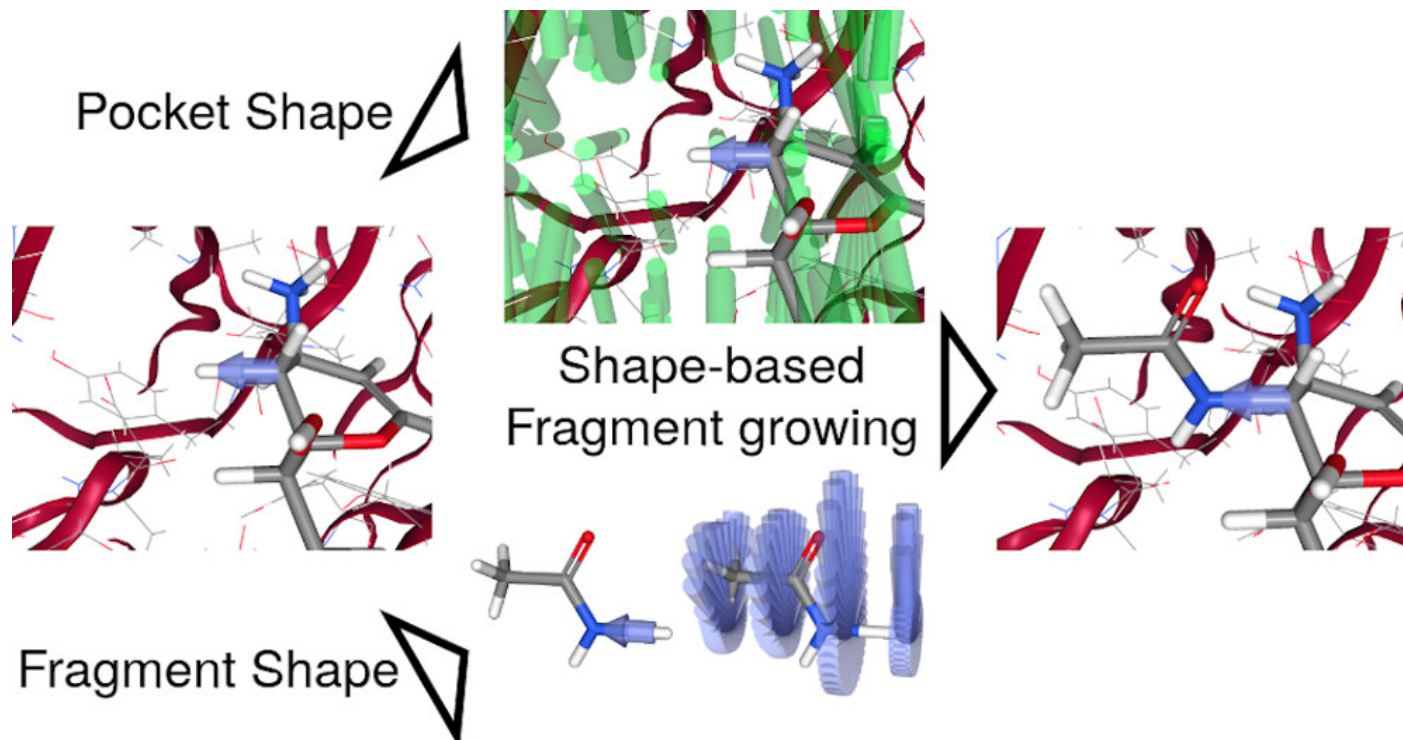
1mio-100mio

10^{10} - 10^{60}

Protein Structure Processing and Exploitation



Efficient Fragment Growing via RVM Descriptors



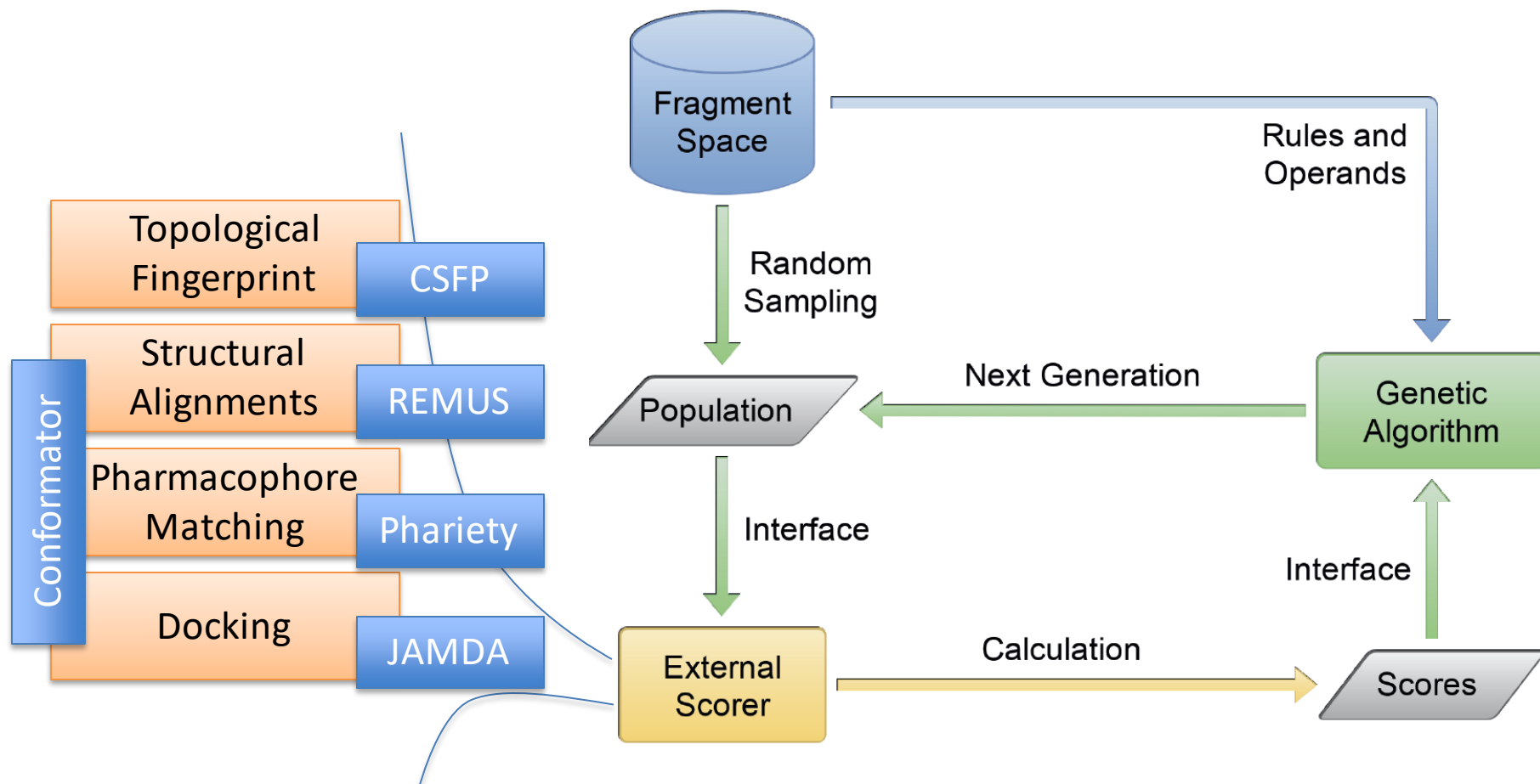
Accuracy (pose < 2Å RMSD)
cross-growing: 67.5% (425 cases)
anchored docking (DOCK): 54,3%

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



BioSolveIT
expect actives!

Galileo: 3D Genetic Operators for Searching in Chemical Space

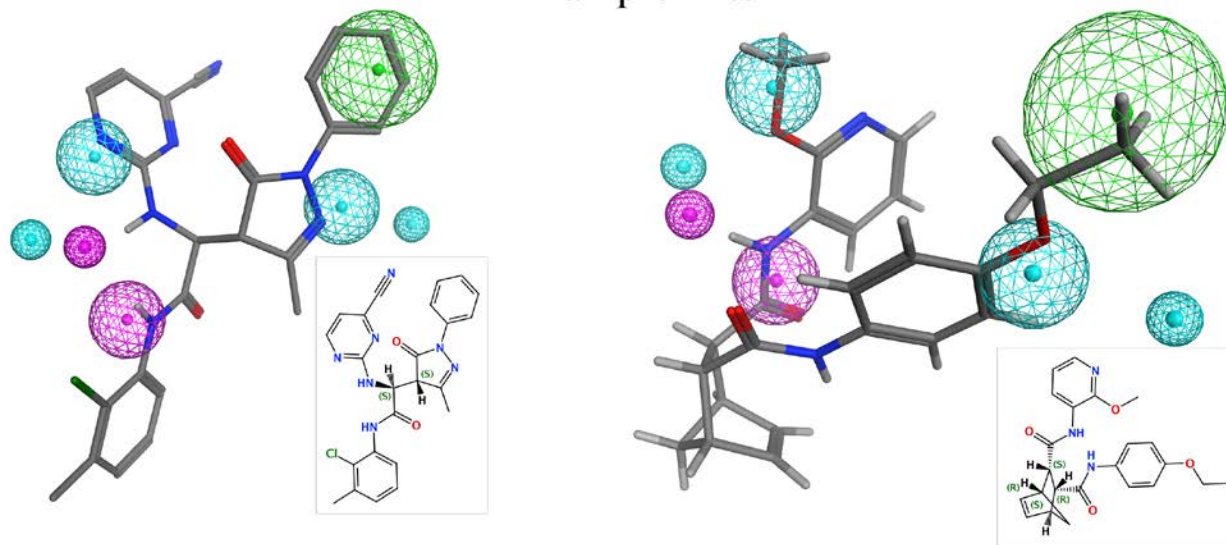
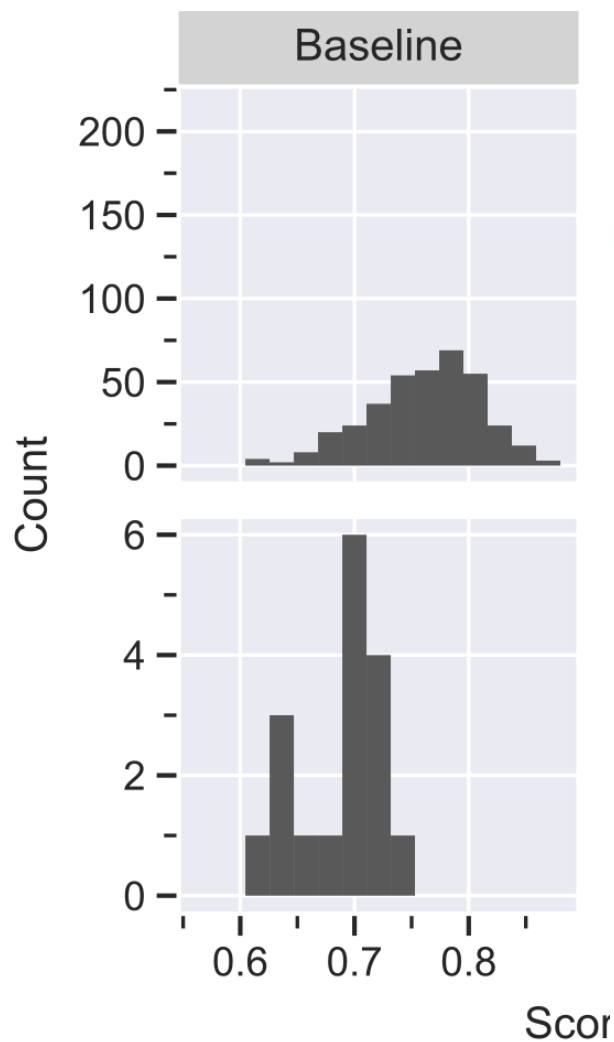


joint work with

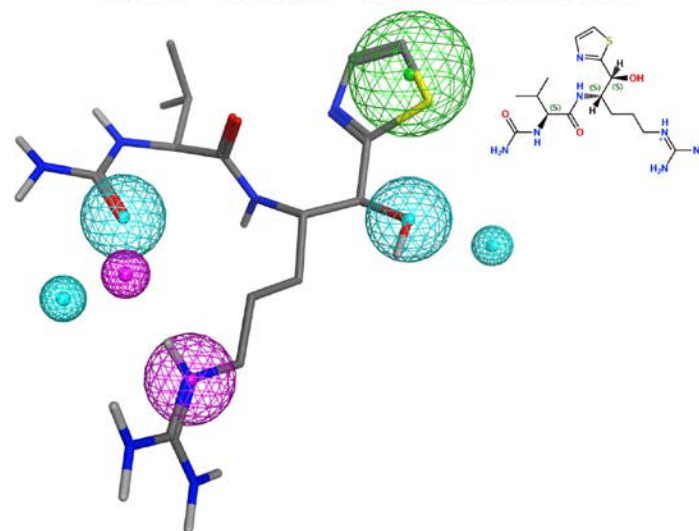


A Pharmacophore Search in Enamine REAL Space

Example Hits



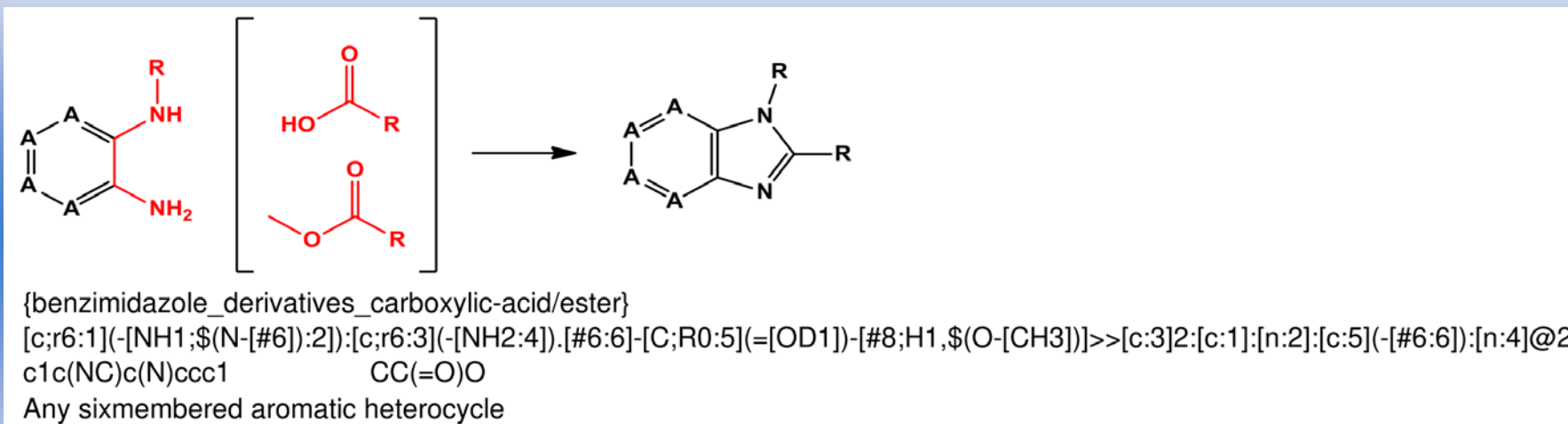
FX1a: model with directions



Some Concluding Remarks

- Make-on-demand catalogues and custom-made chemical spaces
 - are orders of magnitude larger than screening collections
 - substantially impacts early-phase drug discovery
- Standard cheminformatics (fingerprint similarity (max. common substructure search, reduced graphs, overlap, histograms) can be solved efficiently with combinatorial algorithms
- 3D searching (pharmacophores, shape similarity, docking) is possible, but remain heuristic in nature.

Coming Soon: ReactionSMARTSviewer



Contributions / Acknowledgements




current (alphabetic order): Louis Bellmann, Ole Berg, Theresa Cavasin, Konrad Diedrich, Uschi Dolfus, Emanuel Ehmki, Christiane Ehrt, Gerd Embruch, Joel Graef, Melanie Geringhoff, Michael Grössler, Torben Gutermuth, Tobias Harren, Sophia Hönig, Malte Korn, Christian Meyenburg, Patrick Penner, Jonathan Pletzer-Zelgert, Thorben Poburski, Martin Poppinga, Jochen Sieg, Annett Ungethüm

former: Mathias von Behren, Stefan Bietz, Wolf-Guido Bolick, Jörg Degen, Hans-Christian Ehrlich, Janna Eich, Rainer Fährrolfes, Robert Fischer, Florian Flachsenberg, 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, Robert Schmidt, 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]



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>

