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

Combinatorial Approaches for Searching Synthetically Accessible Chemical Space

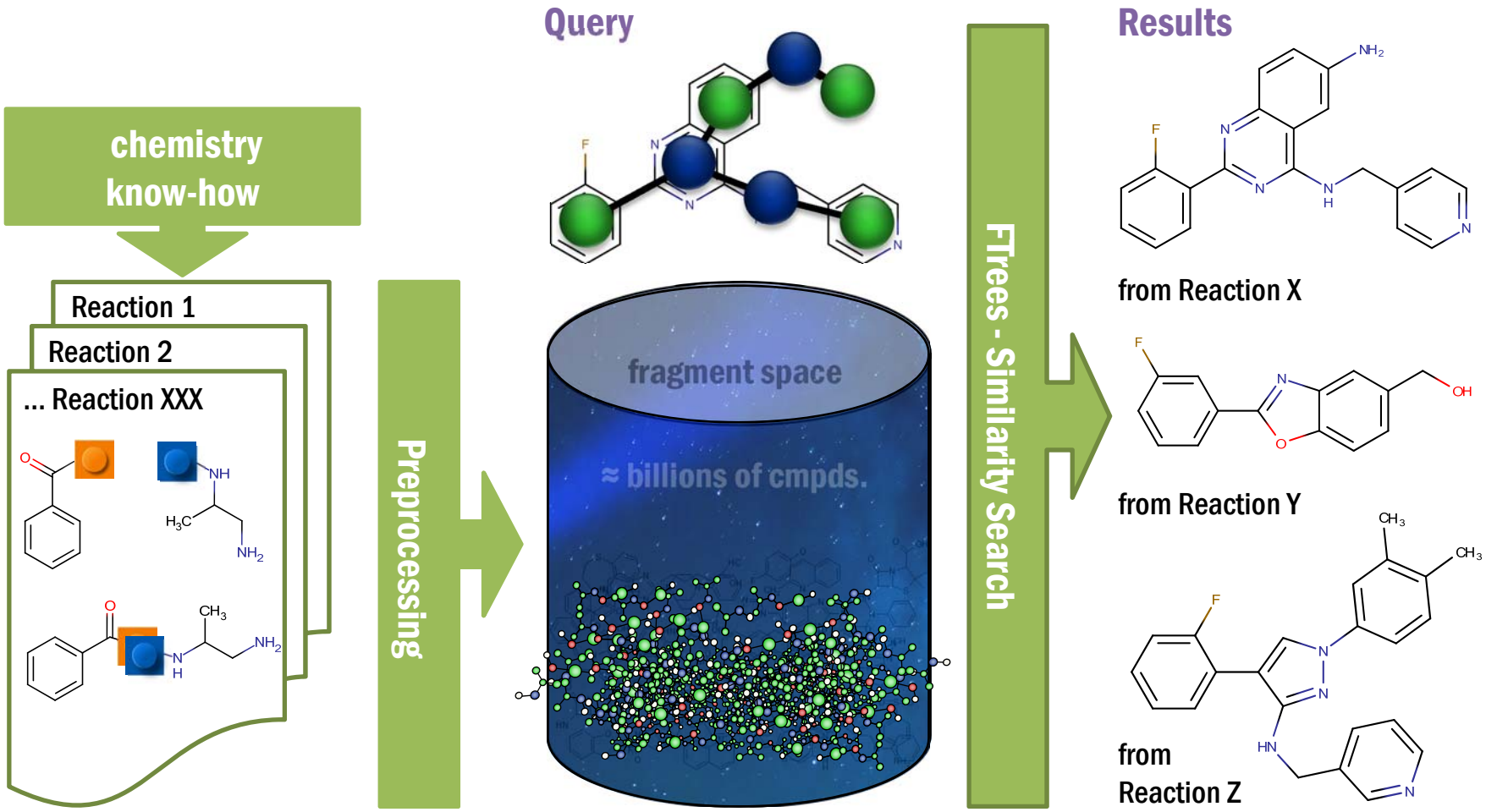
MATTHIAS RAREY

Research Group for Computational Molecular Design

Forschungsgruppe Algorithmisches Molekulares Design (AMD)

Ftrees-FS: Rarey & Stahl, (2001), J.Comput.-Aided Mol.Des., **15,** 497ff

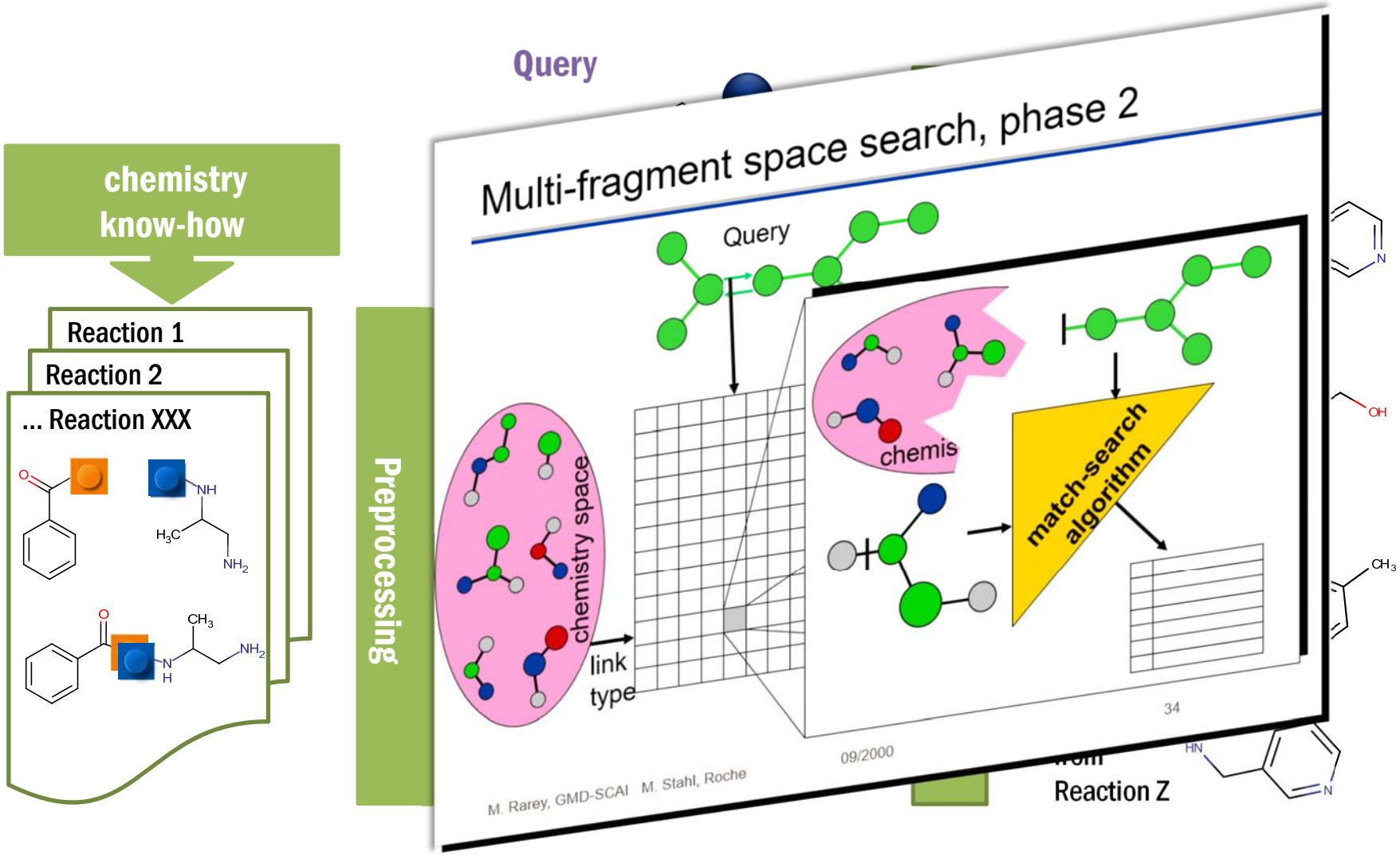
FTrees Fragment Spaces: Searching Instead of Screening



Courtesy of Franca Klingler, BioSolveIT GmbH

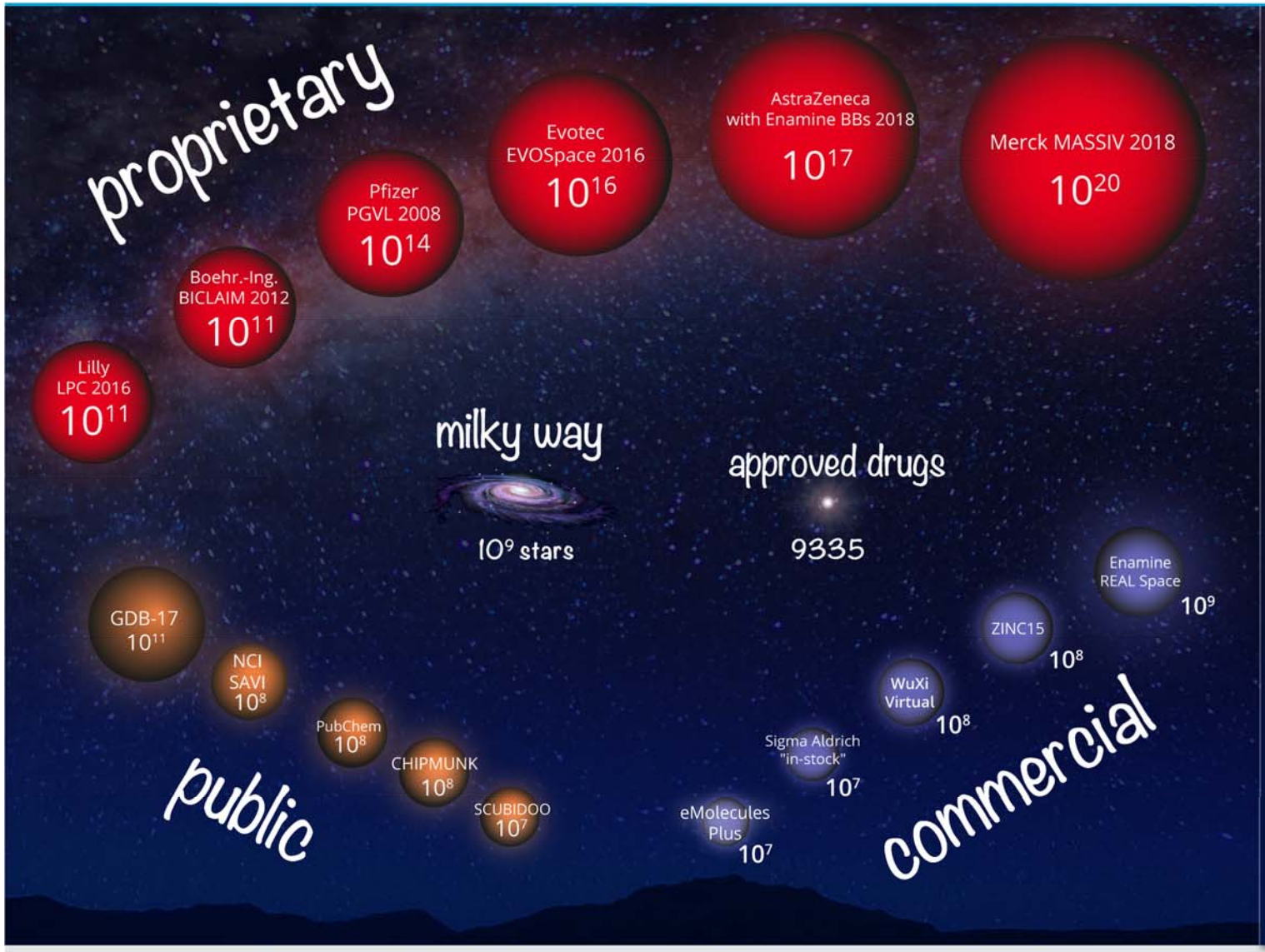
Ftrees-FS: Rarey & Stahl, (2001), J.Comput.-Aided Mol.Des., **15,** 497ff

FTrees Fragment Spaces: Searching Instead of Screening



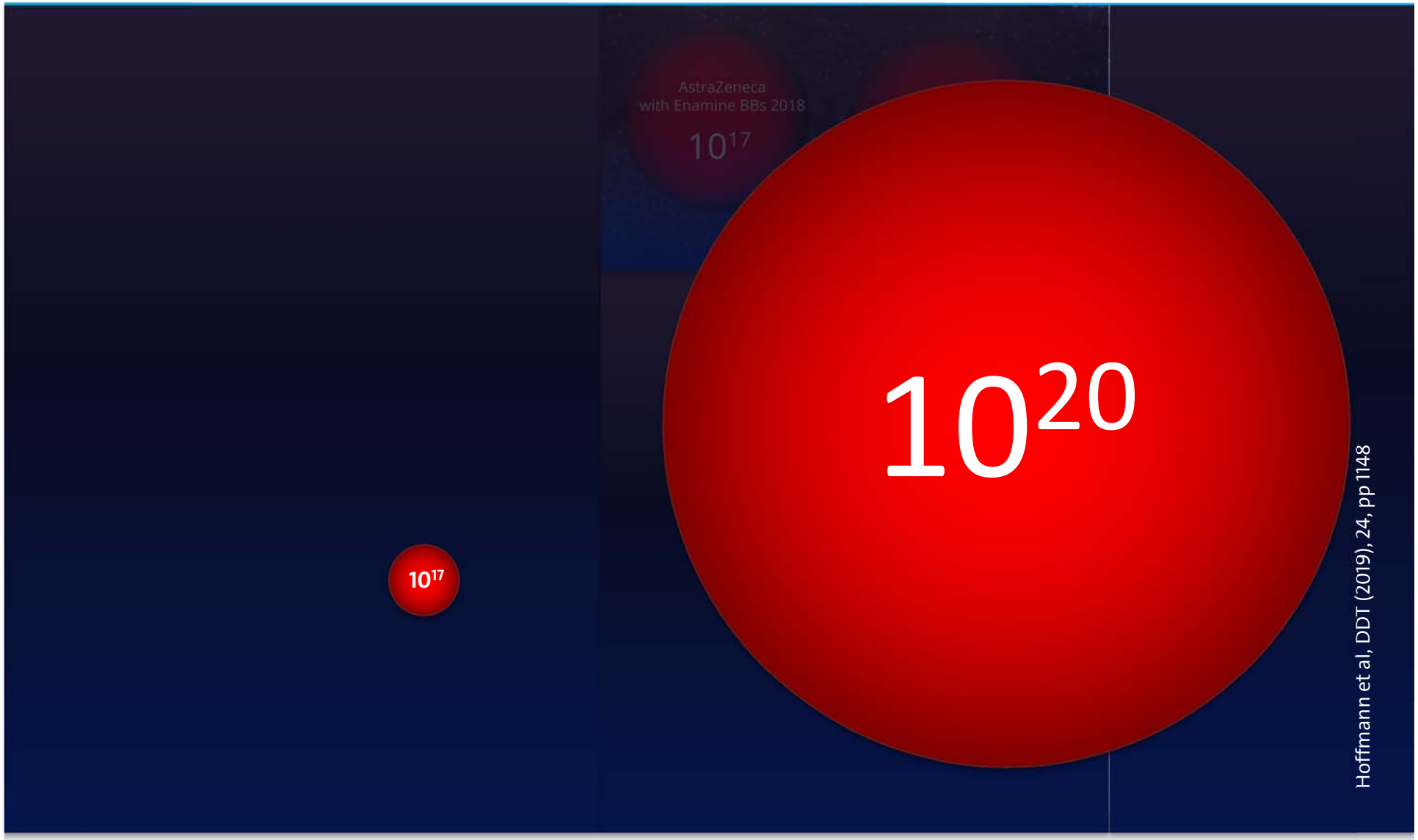
Courtesy of Franca Klingler, BioSolveIT GmbH

Advances in Modeling Chemical Space



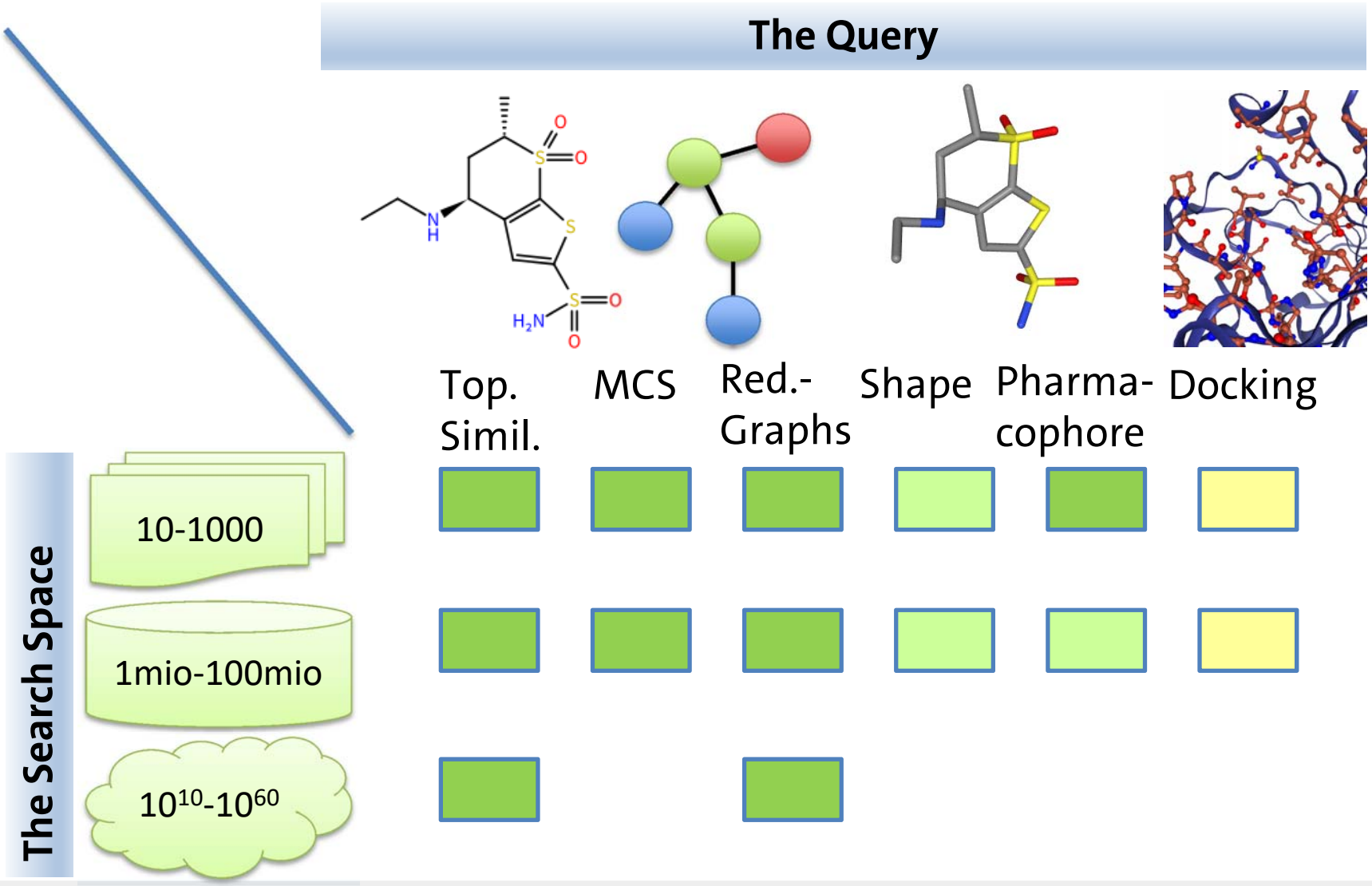
Hoffmann et al, DDT (2019), 24, pp 1148

Advances in Modeling Chemical Space



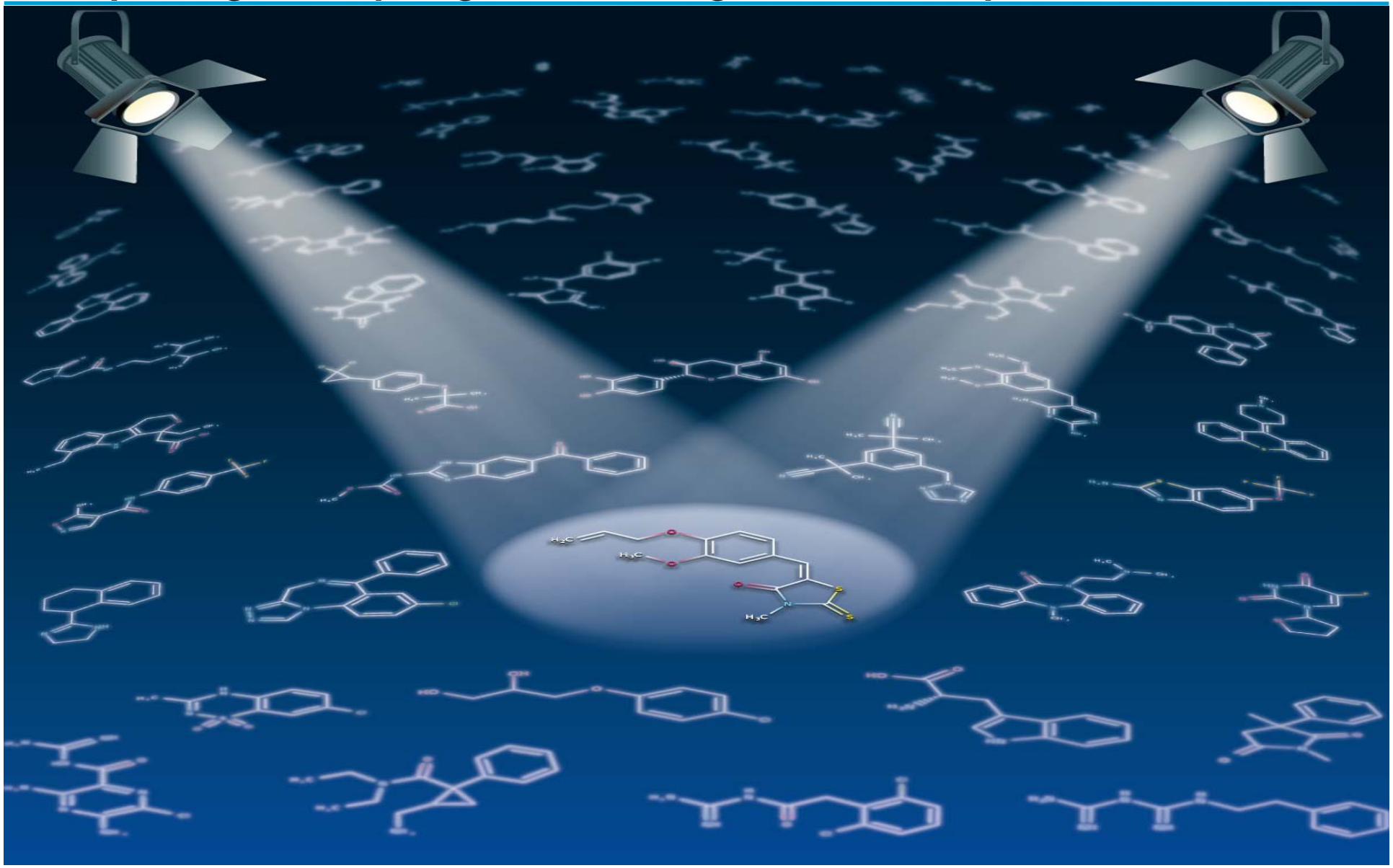
Hoffmann et al, DDT (2019), 24, pp 1148

Hit Identification Methods and Chemical Space



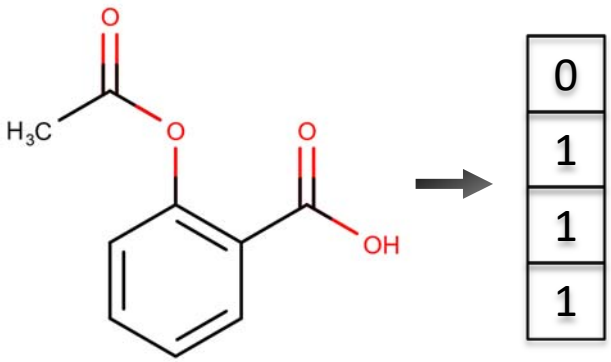
SpaceLight: Bellmann et al, (2020), J.Chem.Inf.Model., **ASAP**

SpaceLight – Topological Searching in Chemical Space



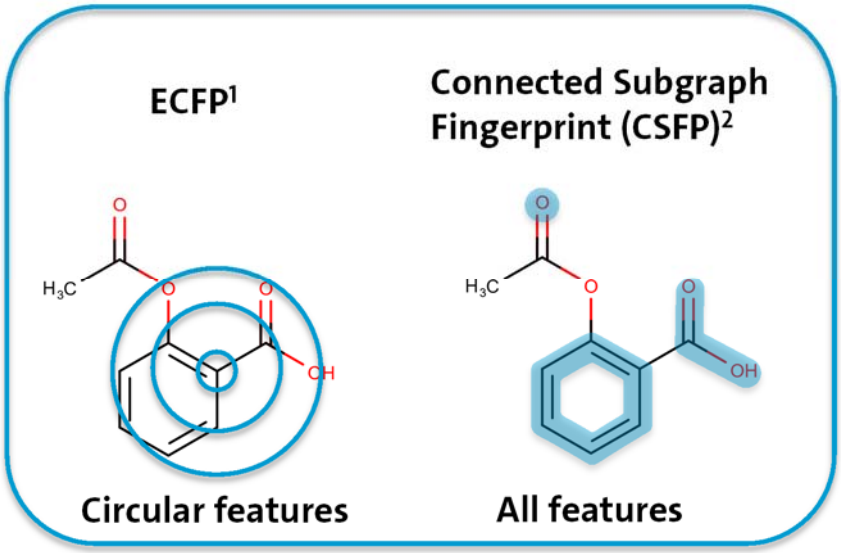
Classic Molecular Fingerprints in a New Light

Employ classic fingerprints in a new search approach



CSFP versions

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



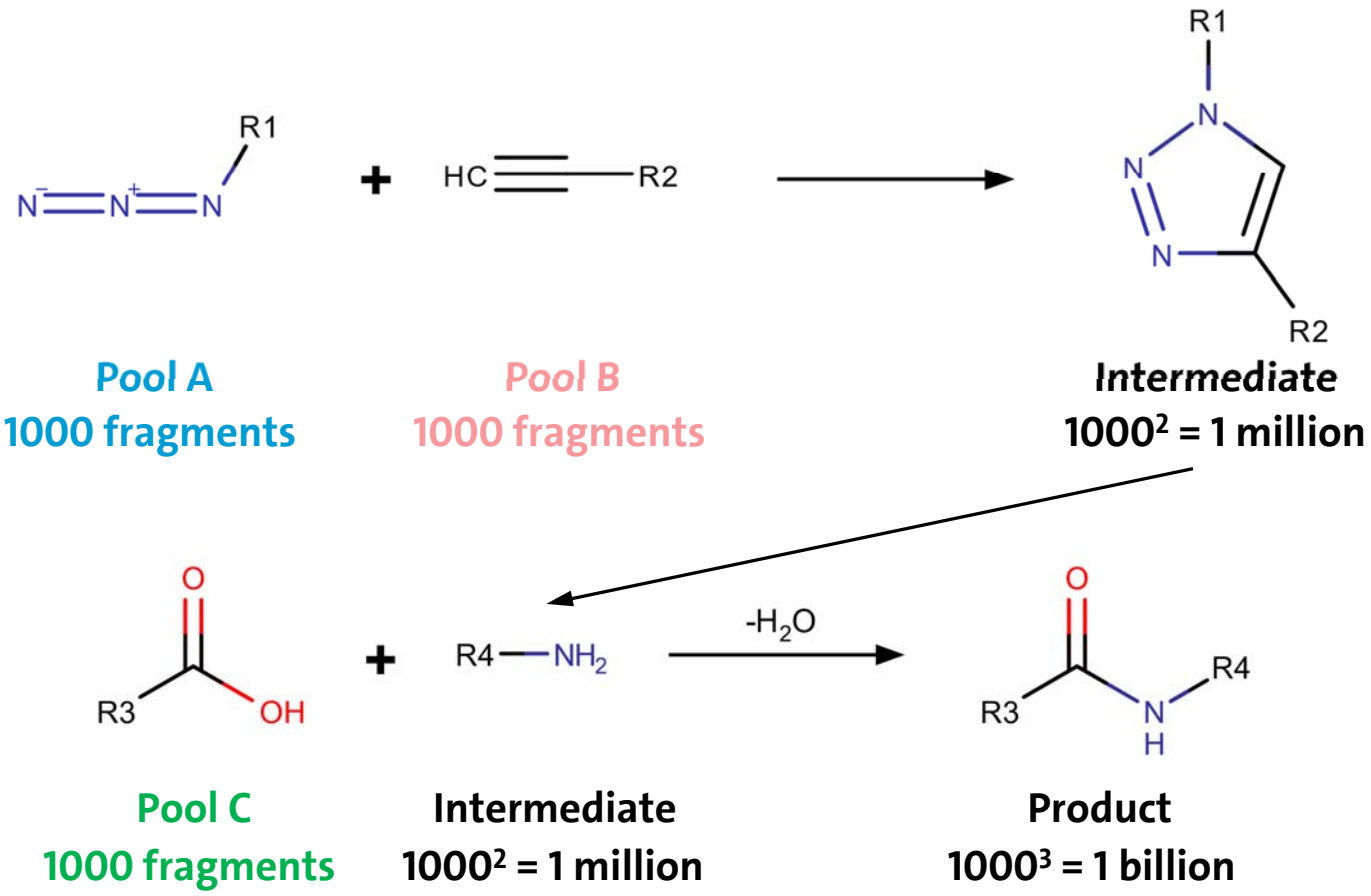
Notation

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

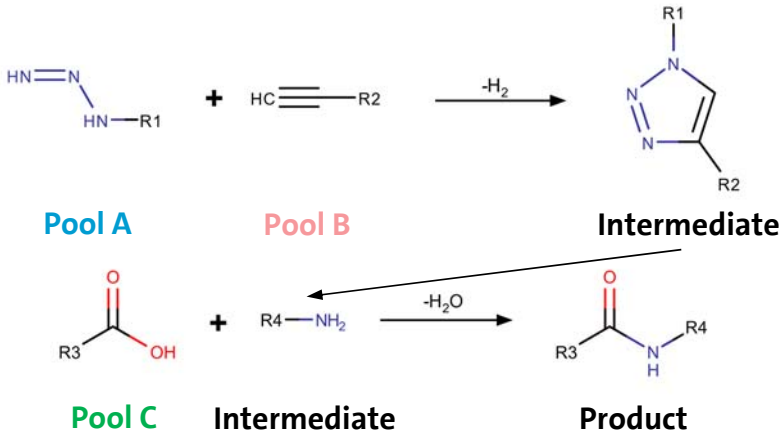
Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

A Way Out: Combinatorial Nature of Chemistry

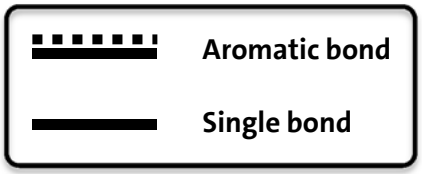


Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

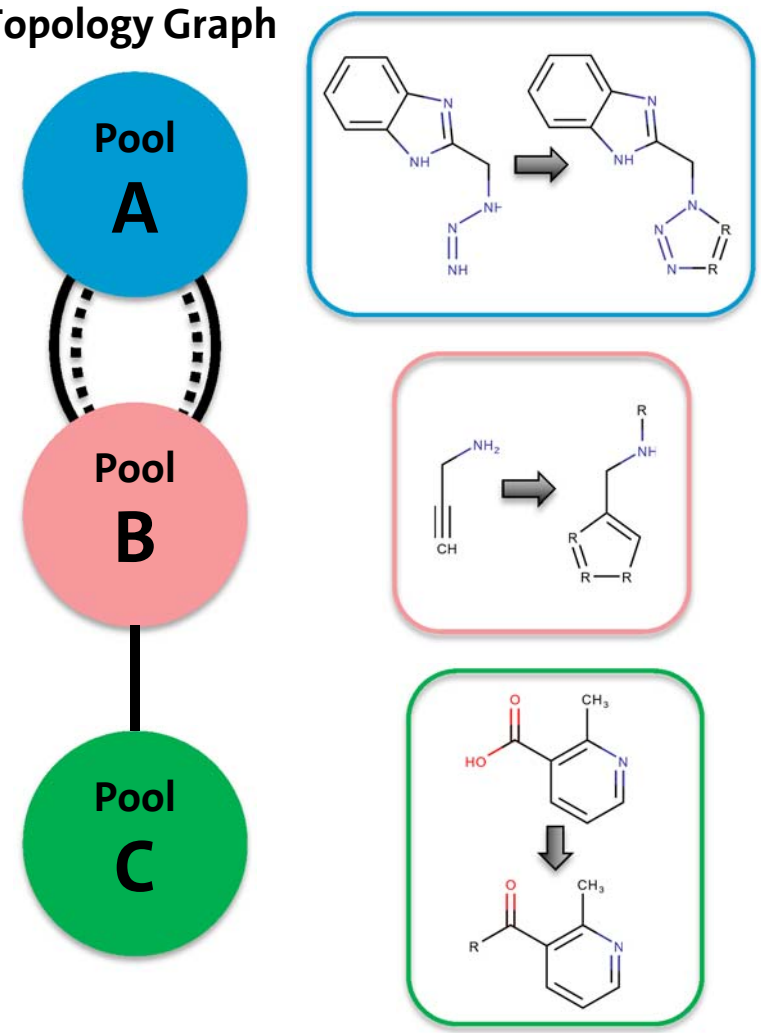
How We Desribe Topological Fragment Spaces



Capture newly formed bonds and structural changes in educts

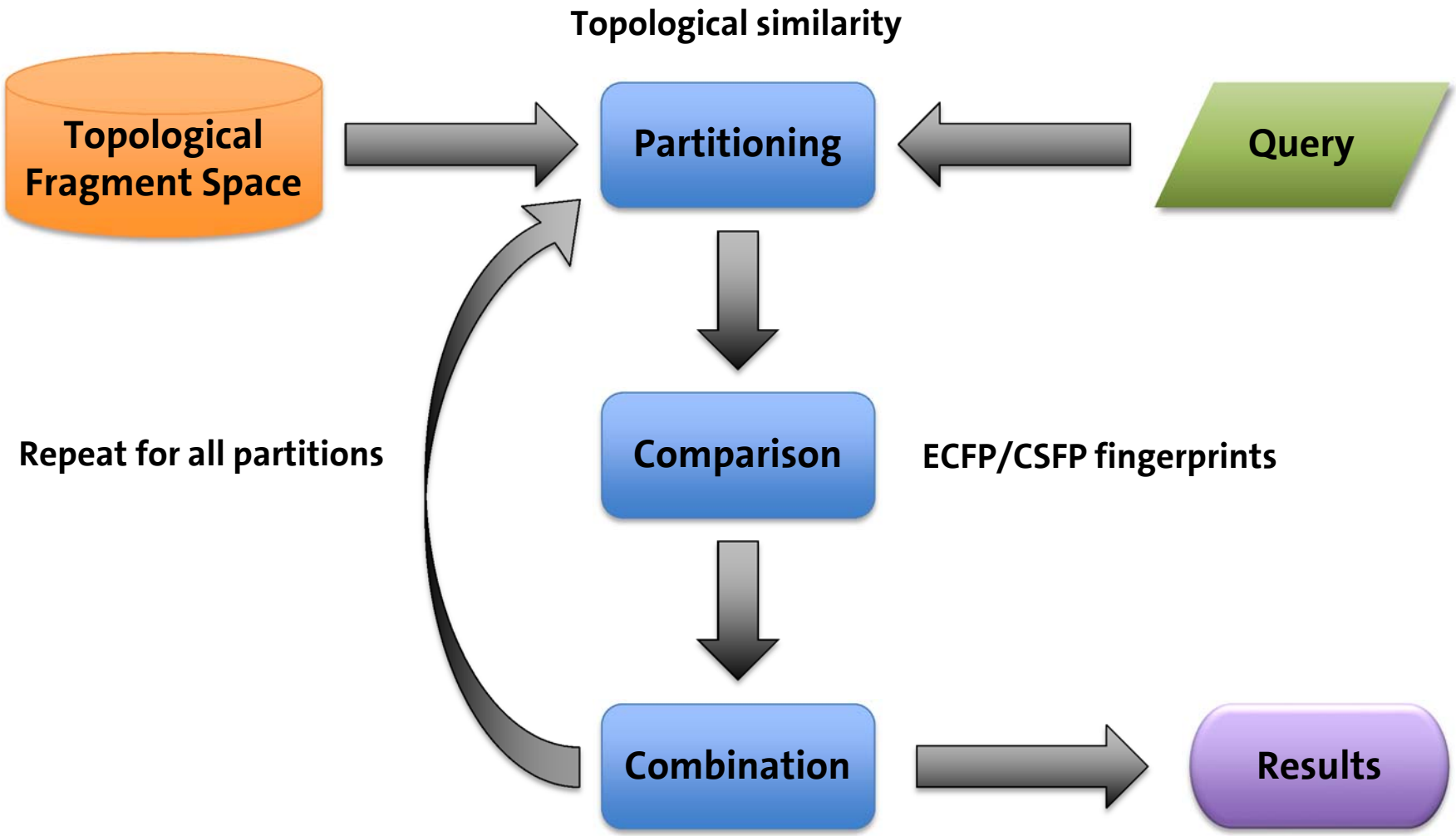


Topology Graph



Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

SpaceLight Algorithm Overview

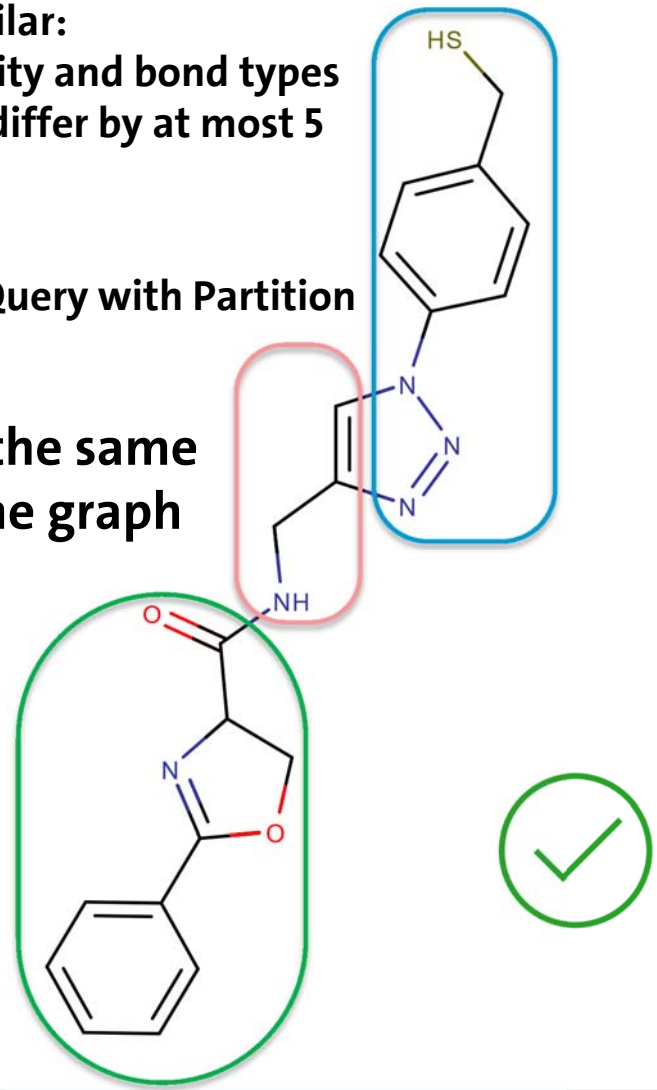


Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

Step 1: Partition Query Compound

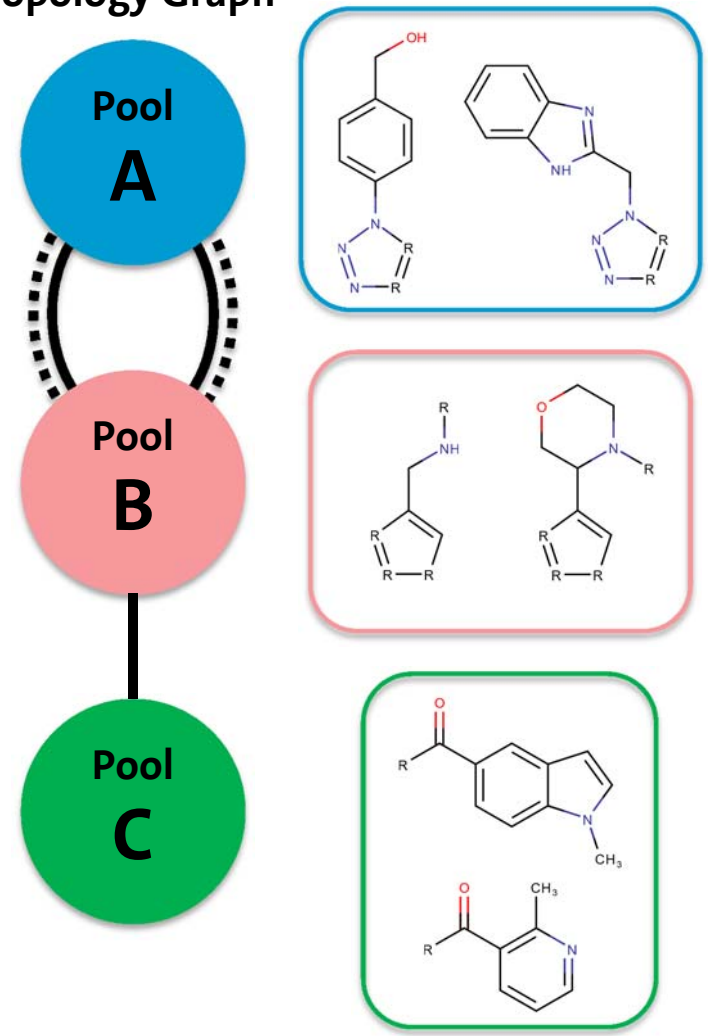
Topologically similar:
1. same connectivity and bond types
2. #heavy atoms differ by at most 5

Query with Partition



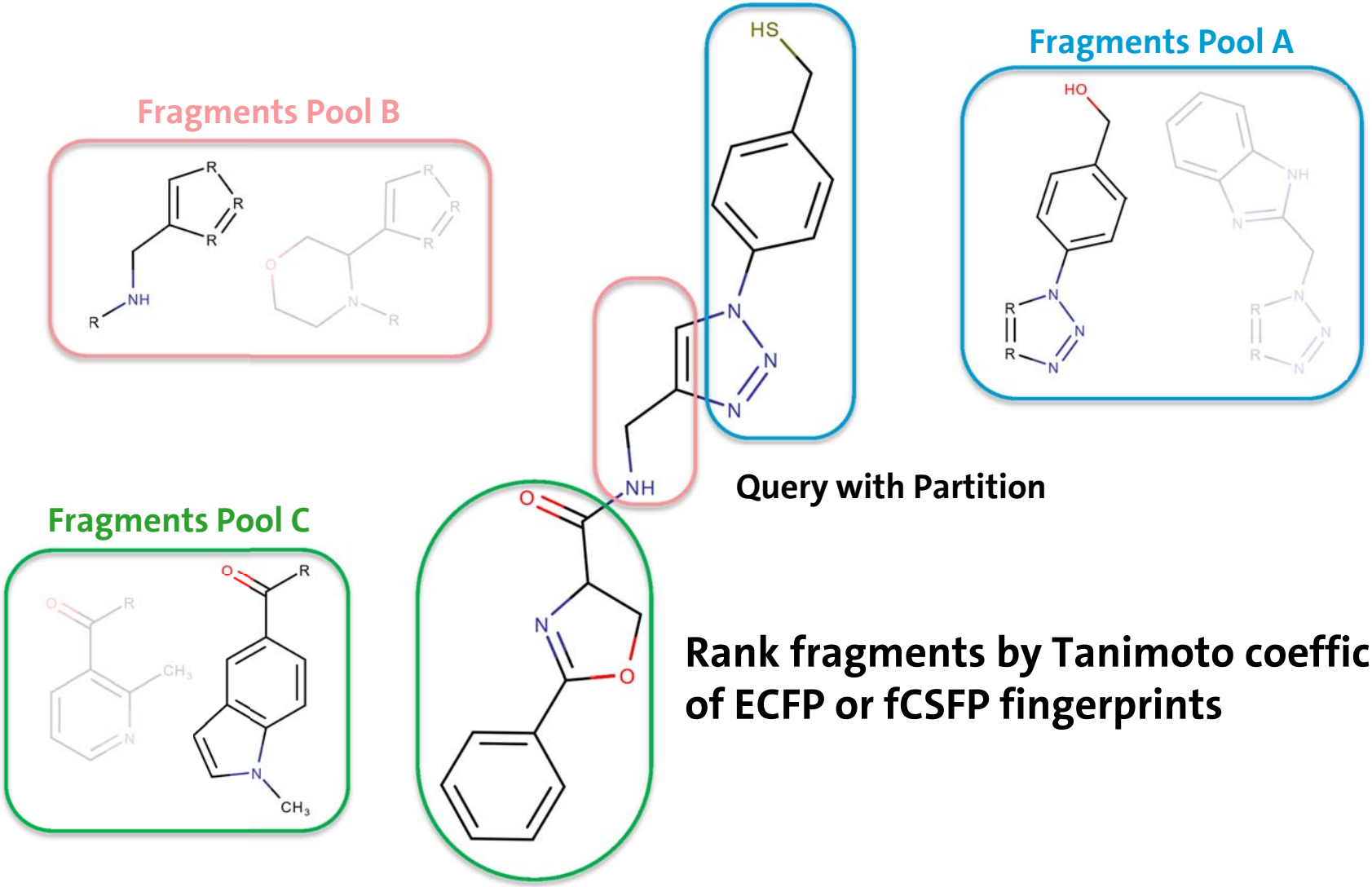
Partition has the same topology as the graph

Topology Graph



Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

Step 2: Find Similar Fragments

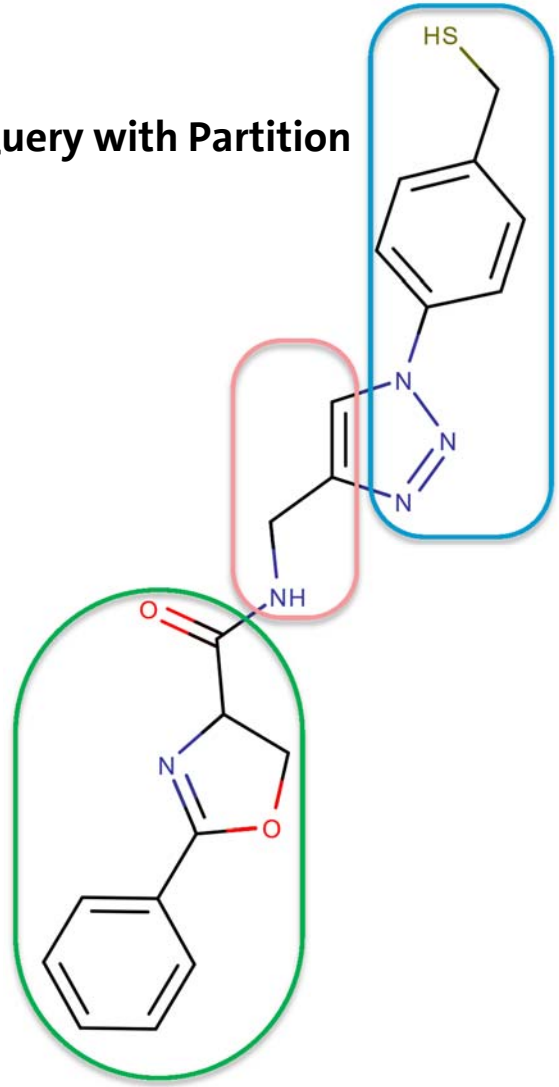


Rank fragments by Tanimoto coefficient of ECFP or fCSFP fingerprints

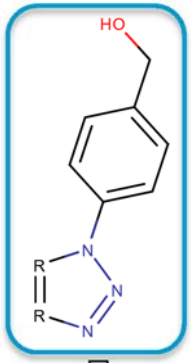
Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

Step 3: Combine Fragments

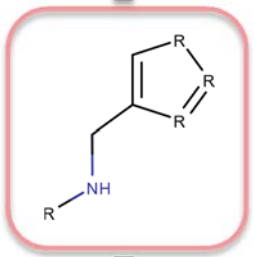
Query with Partition



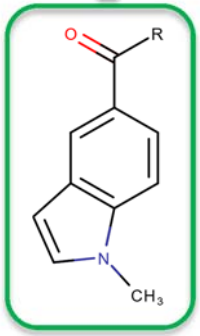
similar



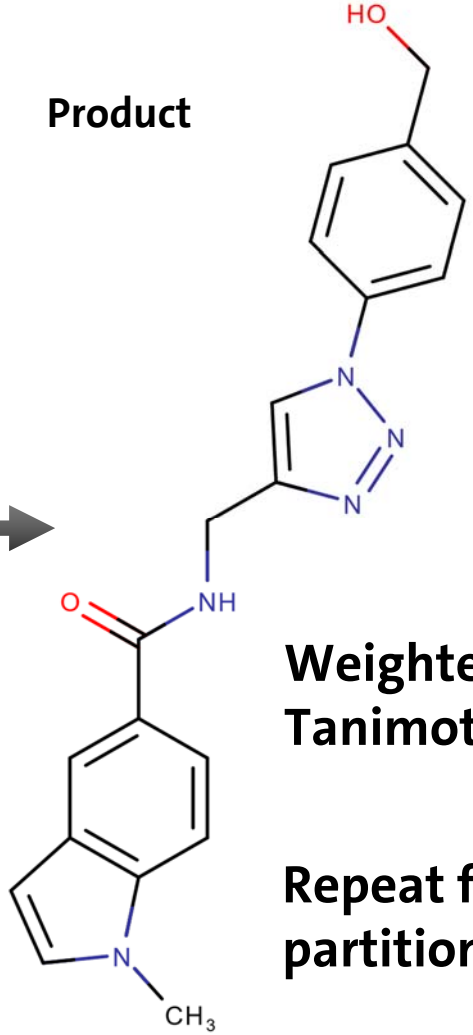
“identical”



not similar



Product

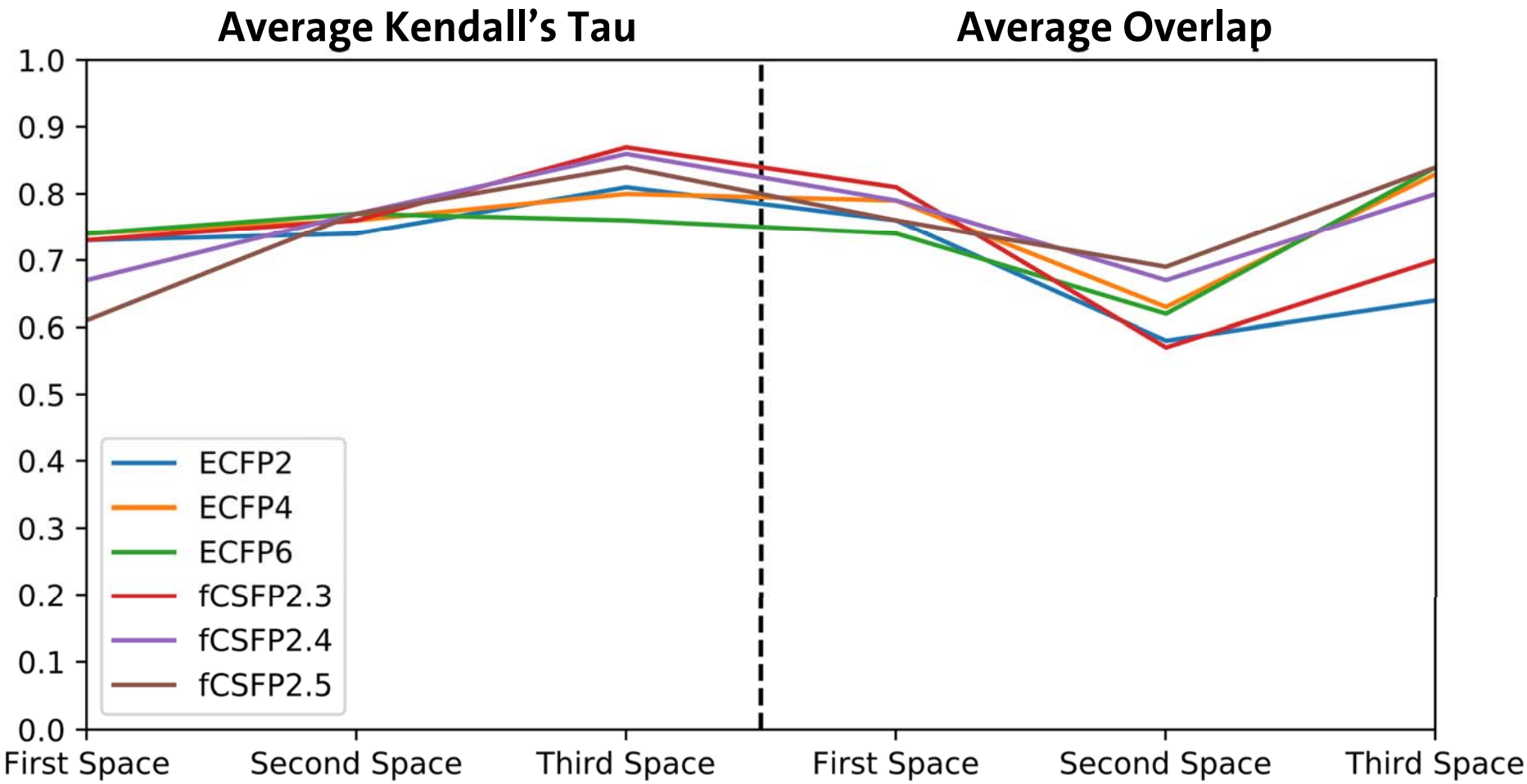


Weighted sum of Tanimoto scores

Repeat for all partitions

Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

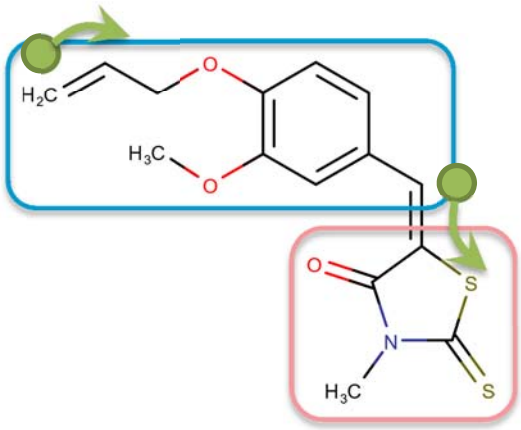
SpaceLight Correlation with Sequential Search



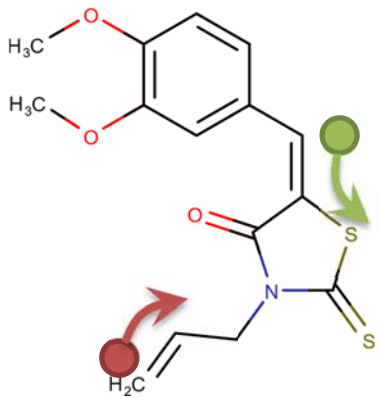
SpaceLight generates results similar to classic fingerprints

Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

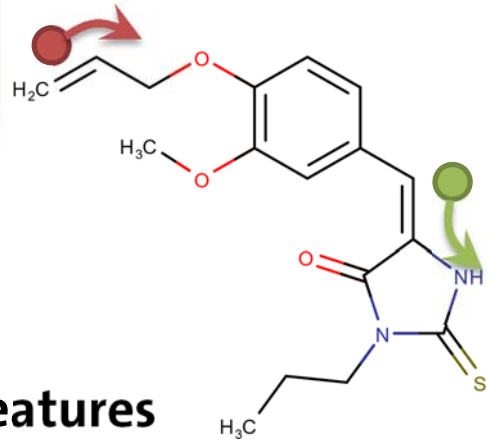
SpaceLight vs. Sequential ECFP4 Search



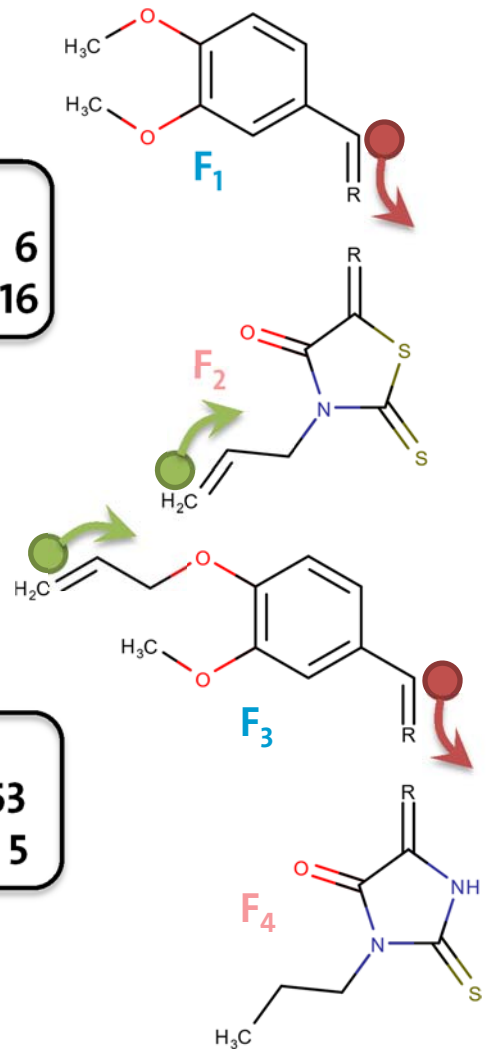
ZINC029932278



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



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



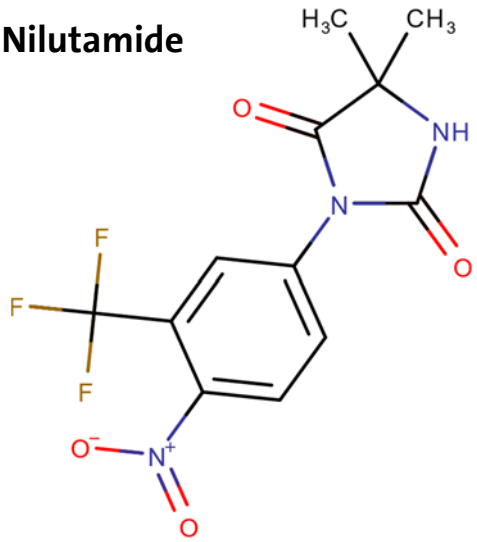
SpaceLight considers locality of chemical features

Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

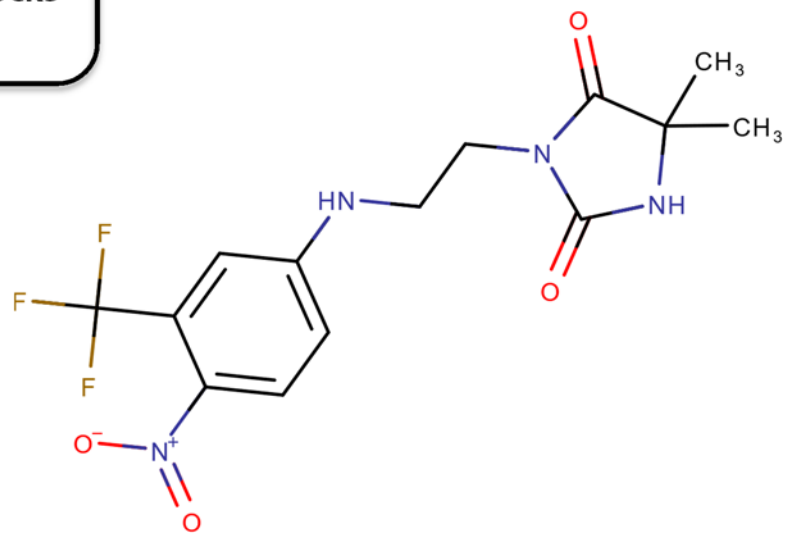
SpaceLight Search for Drugs in Enamine REAL

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

Nilutamide



Found with score 1.0

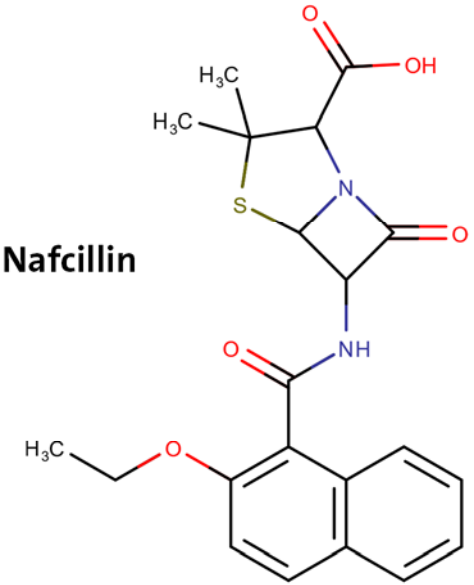


Second best with score 0.87

Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

SpaceLight Gives You Ideas for Synthesis Routes

Nafcillin

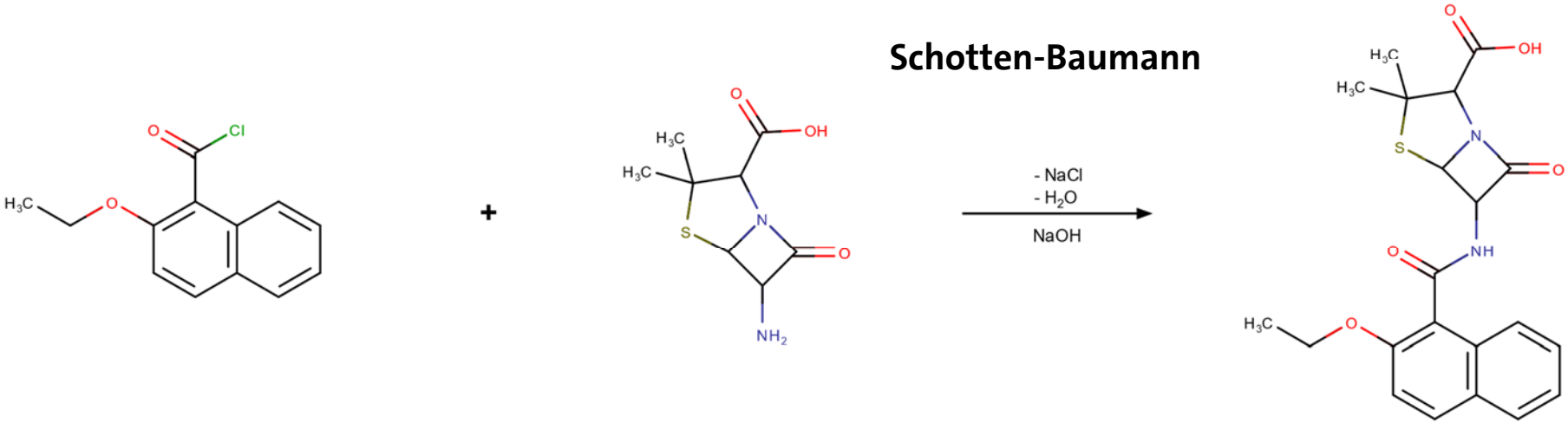


SpaceLight retrieves synthetic pathways of products

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

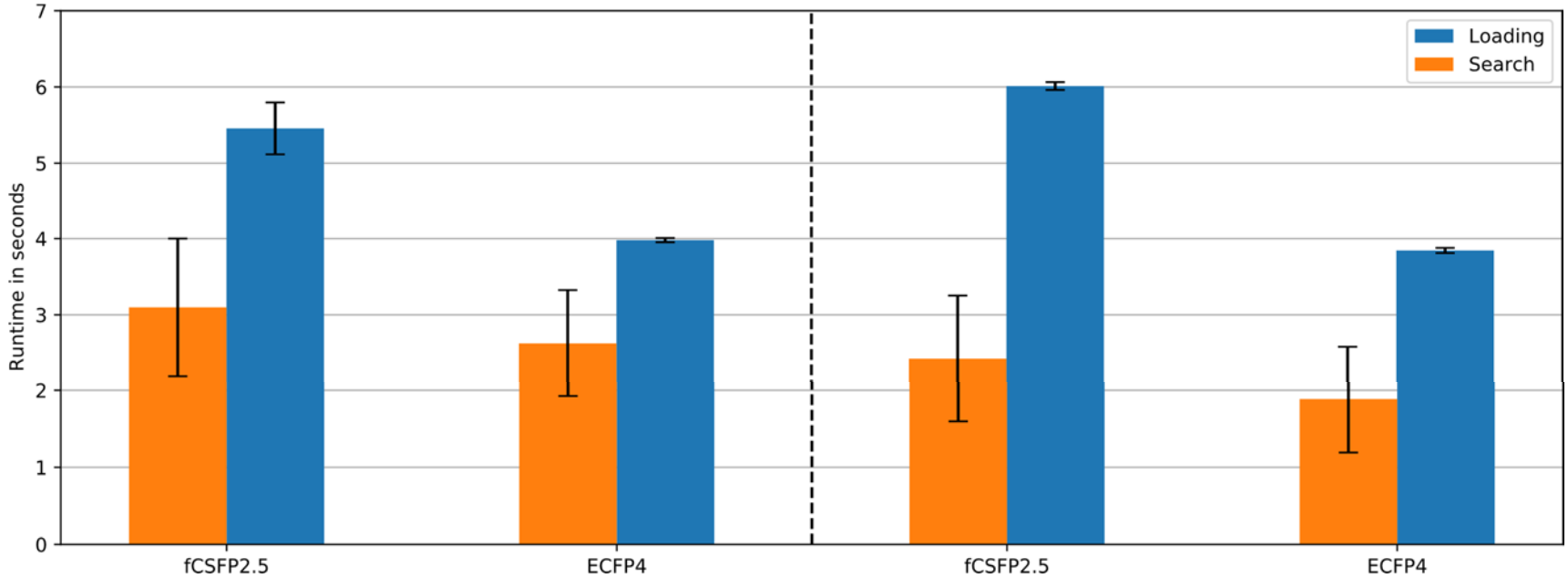
Found with score 1.0

Schotten-Baumann



Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

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

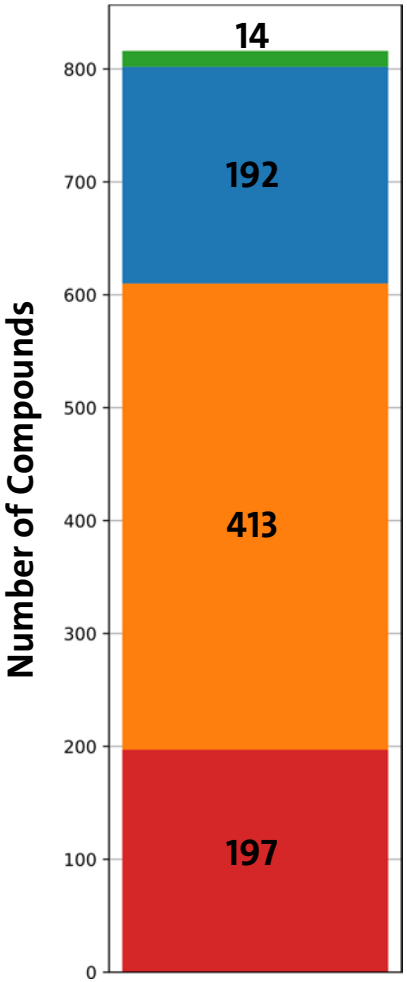
500 random queries from ZINC lead-like

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

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

Bellmann et al., J. Chem. Inf. Model. (2020), ASAP

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

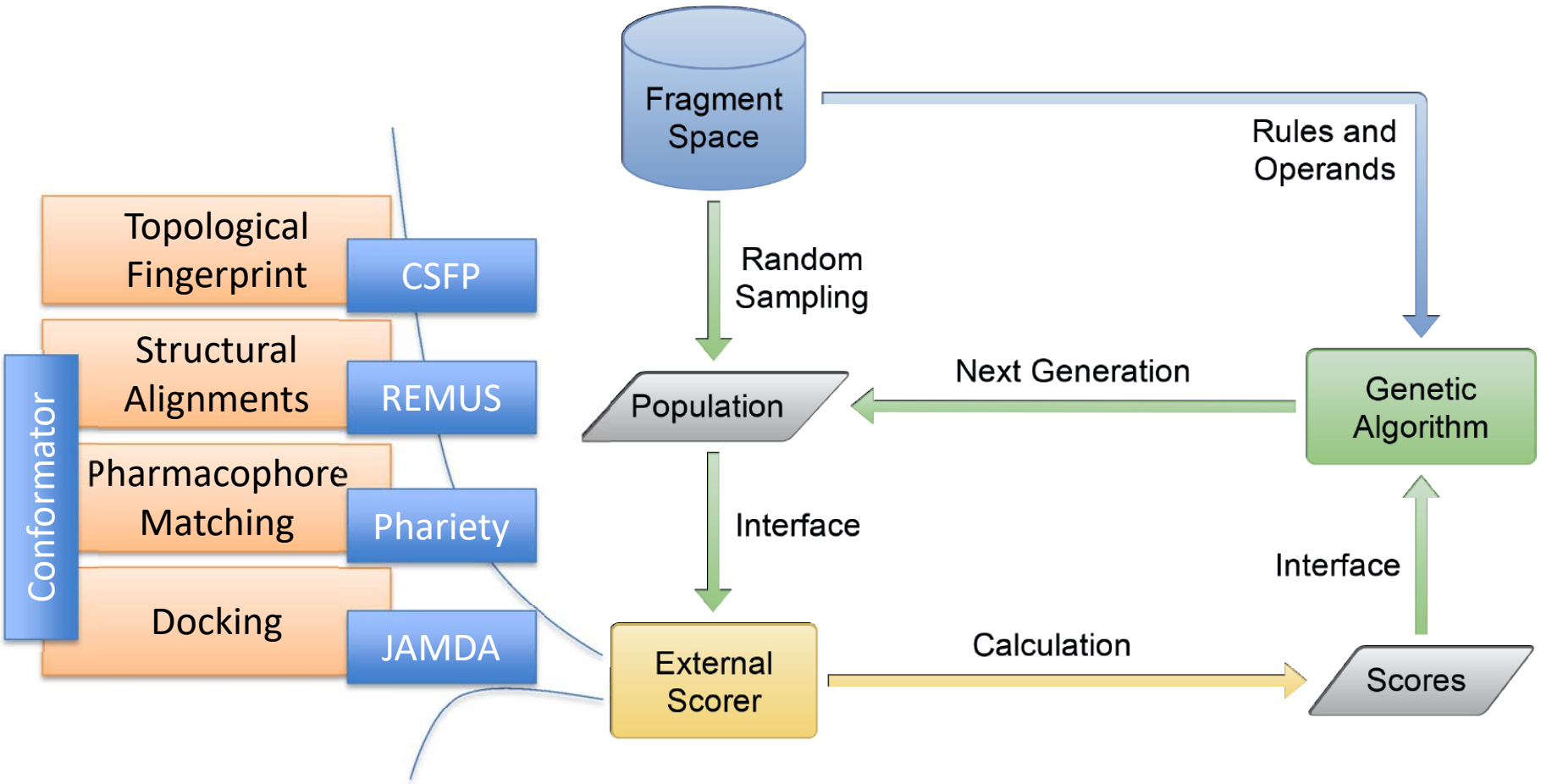
1: Wishart, Nucleic Acids Res. (2008), 46

towards

3D

Meyenburg & Dolfus et al., in preperation

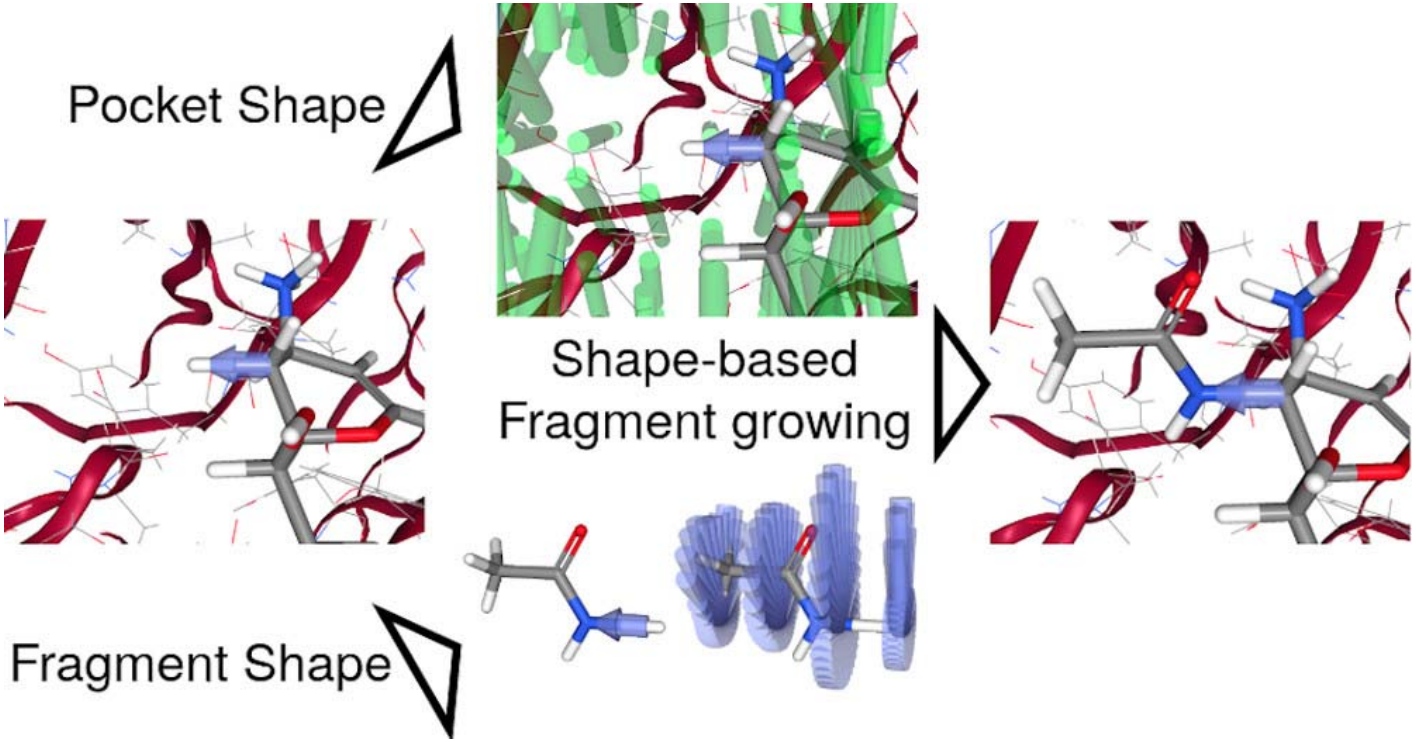
Galileo: 3D Genetic Operators for Searching in Chemical Space



AstraZeneca

Penner et al., J. Chem. Inf. Model. (2020), ASAP

Efficient Fragment Growing via RVM Descriptors

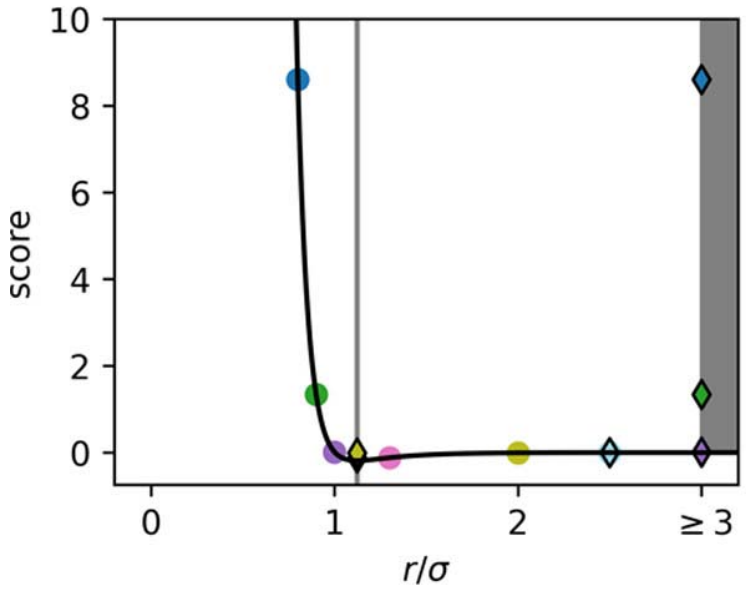


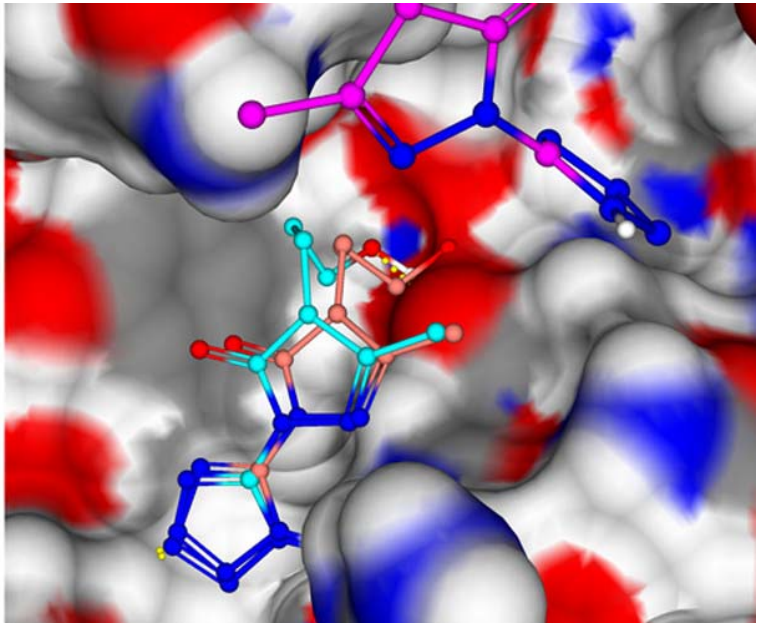
Accuracy (pose < 2Å RMSD)
self-growing: 84% (3299 cases)
cross-growing: 66% (326 cases)

Performance (i5 @ 3.4GHz, 16GB):
32,000 – 77,000 conformers/sec

Flachsenberg et al., J. Chem. Inf. Model. (2020), ASAP

Fast 3D Searching Requires Precise Pose Optimization



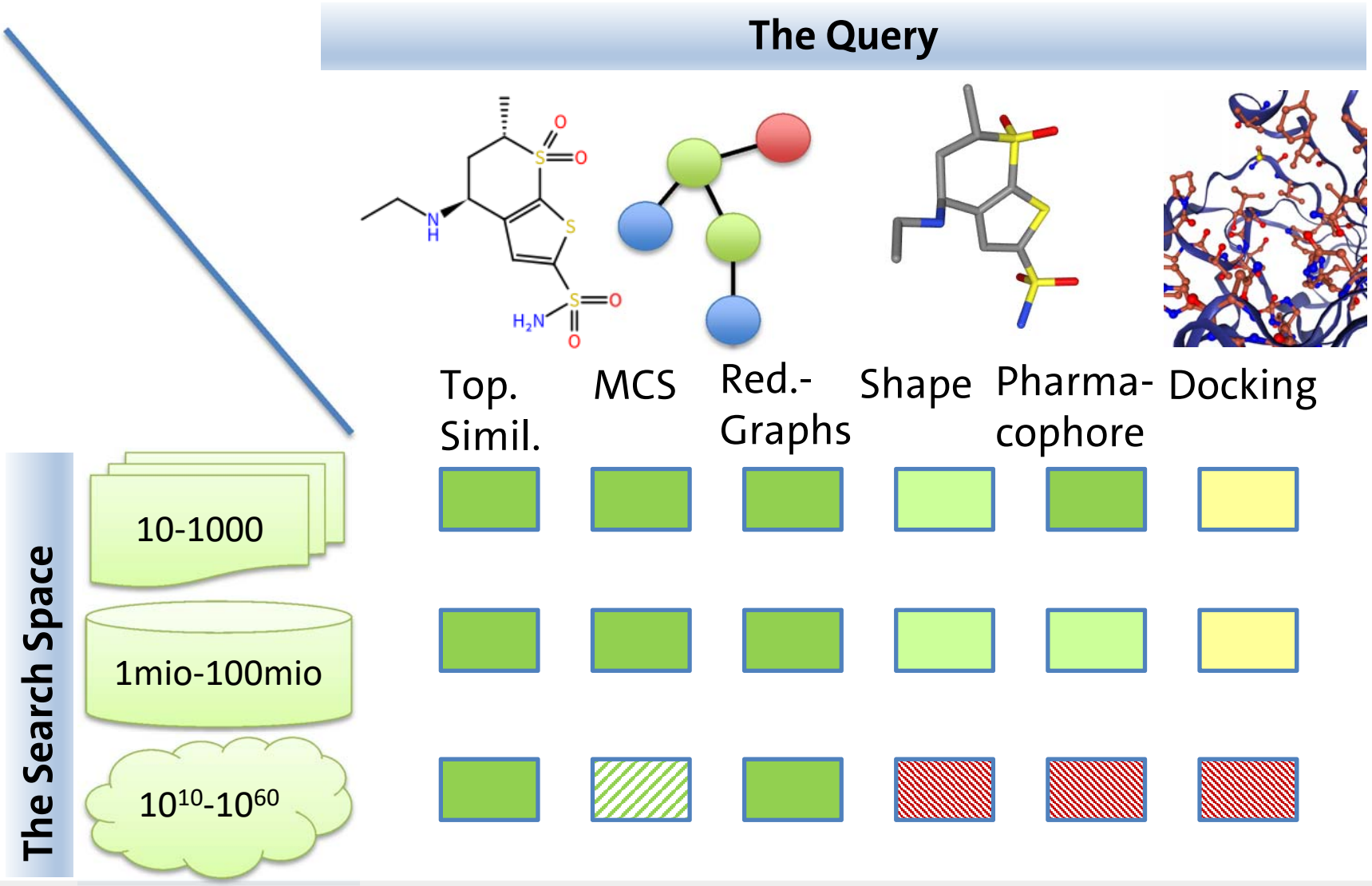


JAMDAscorer:

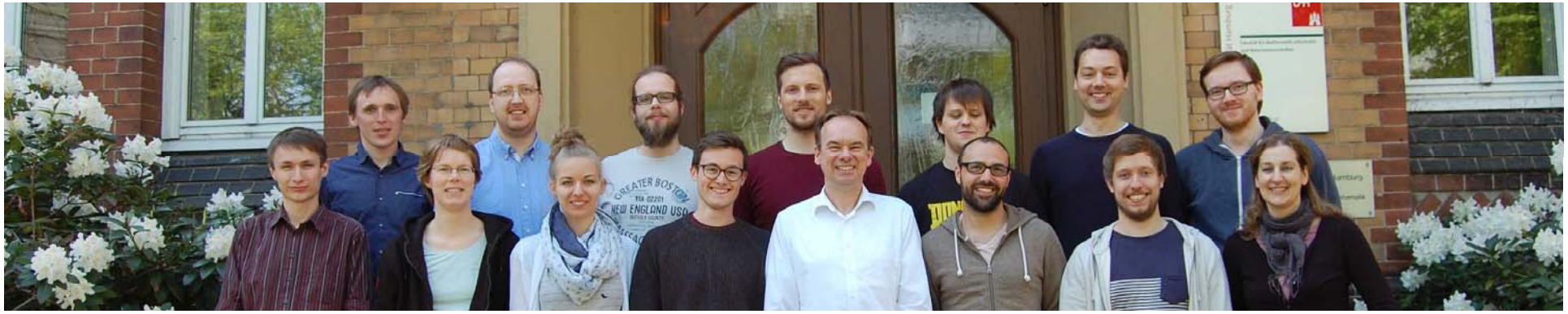
twice differentiable score, new LSL-BFGS gradient-based optimizer

→ high docking power, high optimization locality, low number of score evaluations

Concluding Remarks on Chemical Space Search



Contributions / Acknowledgements



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

former: Mathias von Behren, Stefan Bietz, Jörg Degen, Hans-Christian Ehrlich, Janna Eich, 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 & BioSolveIT 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