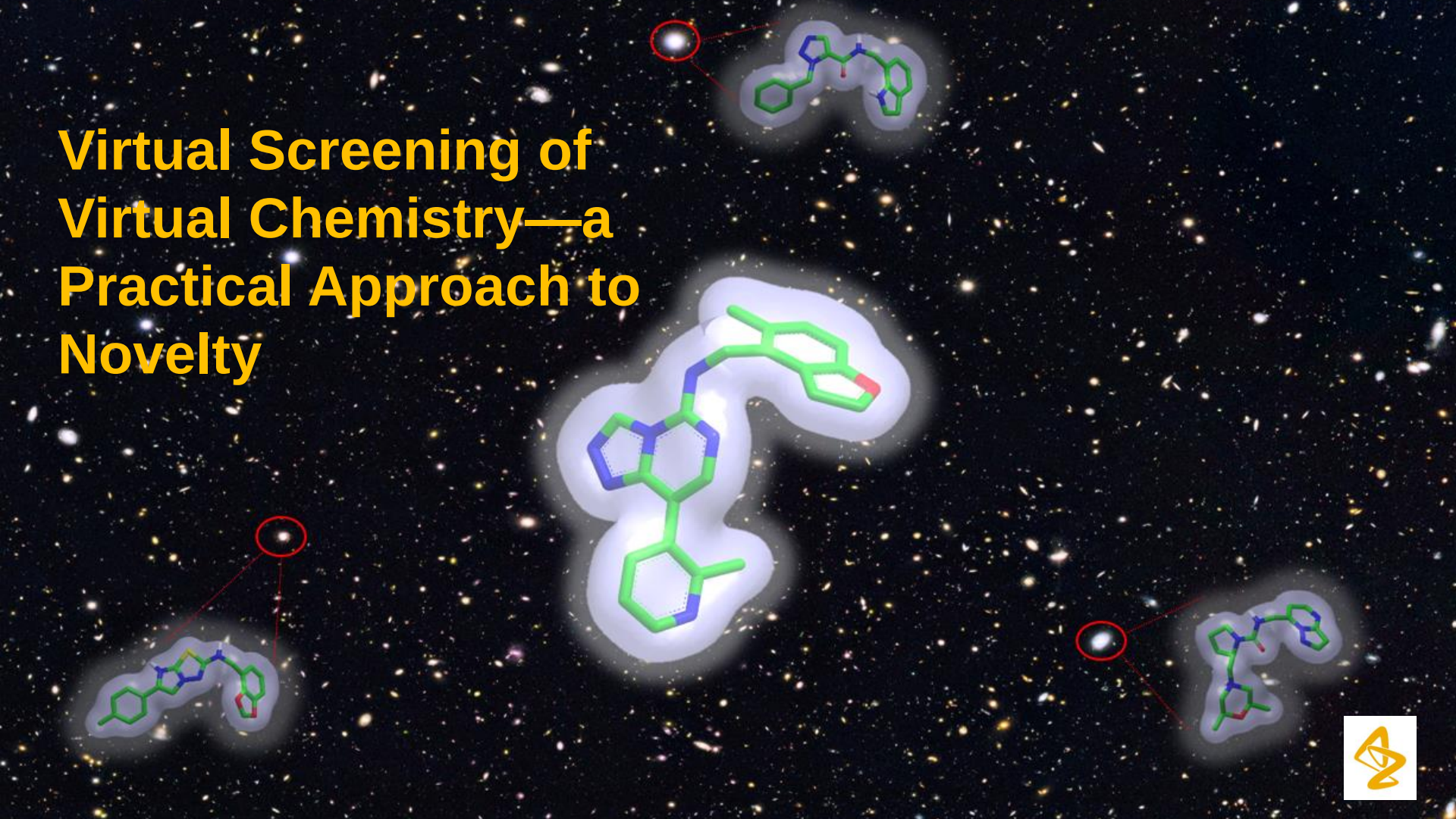
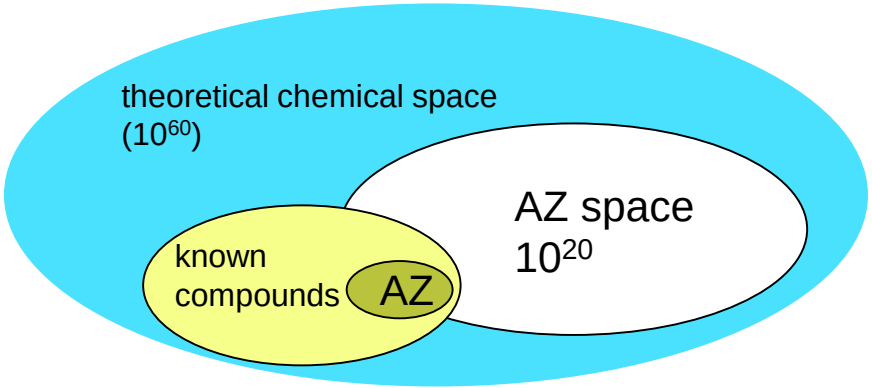
Virtual Screening of Virtual Chemistry—a Practical Approach to Novelty



Virtual screening of virtual chemistry

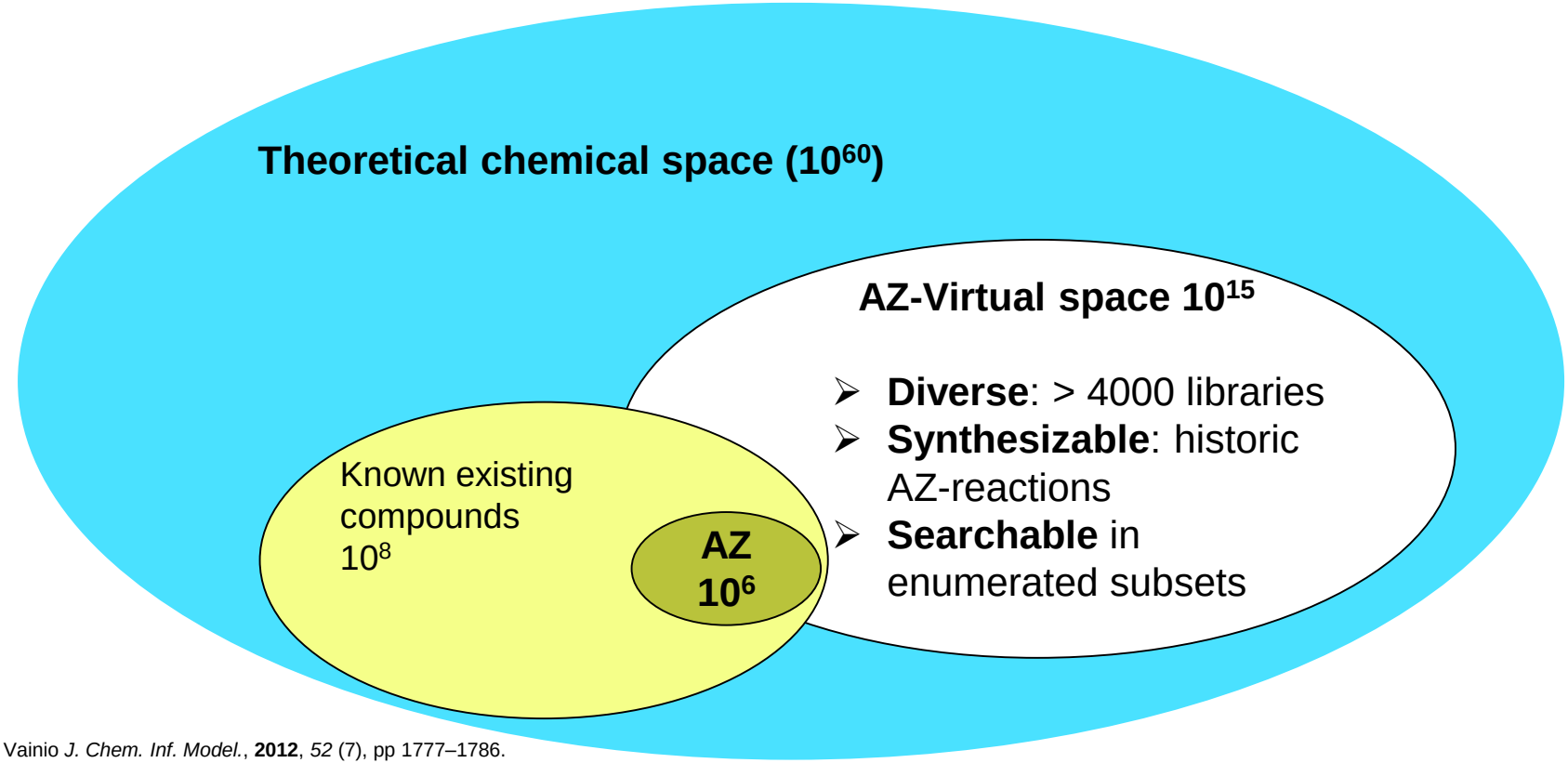


What if we had…
- fast search strategies (minutes-hours)
- 3D methods
- synthetically accessible libraries

To search 10^{20} molecules
- would take $3 \cdot 10^4$ years with current capability

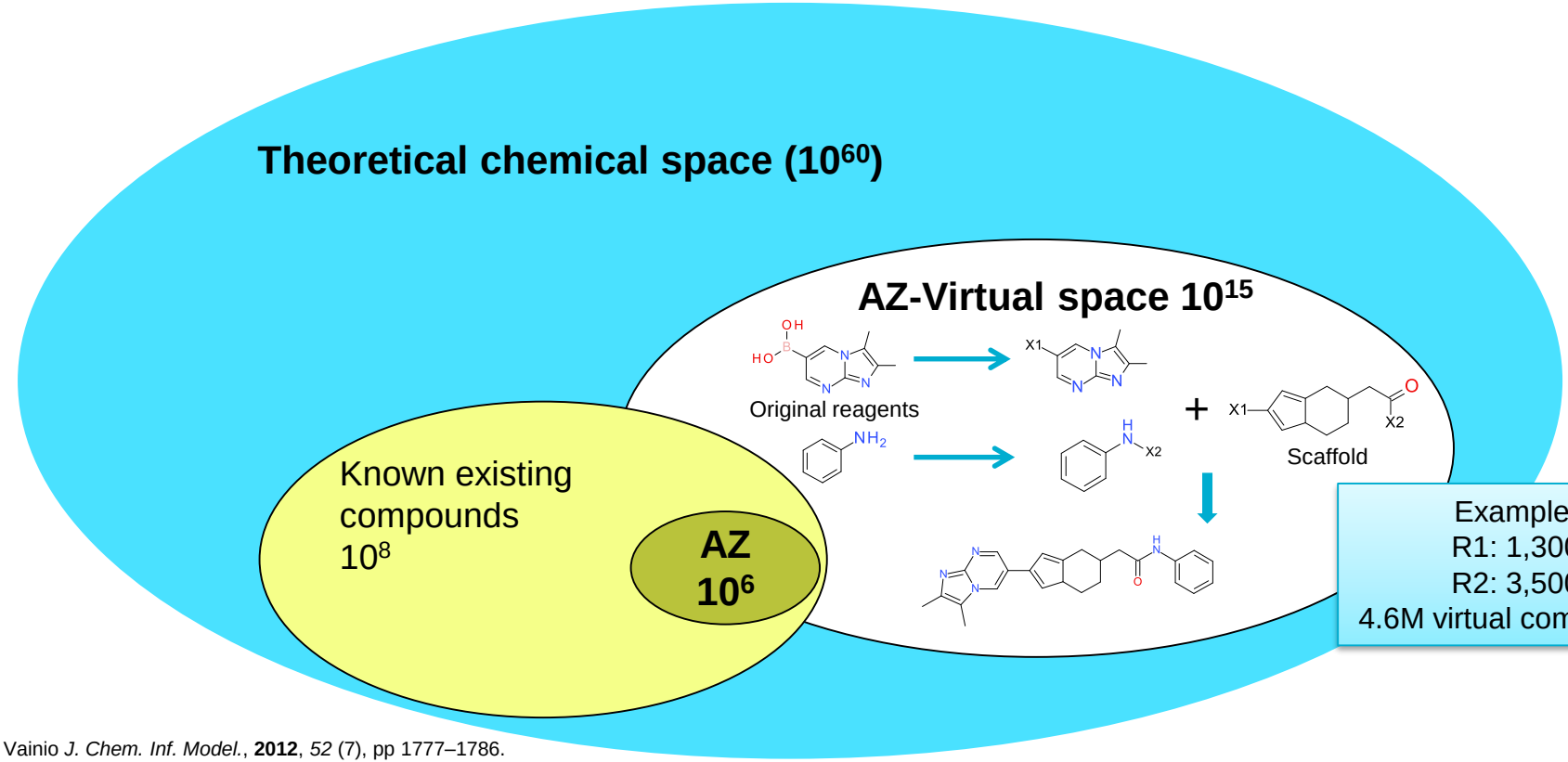
To store 10^{20} molecules
- would need 10^5 ExaByte of storage
- this is 100 times the 2015 world storage market (10^3 ExaByte)

Virtual chemistry space



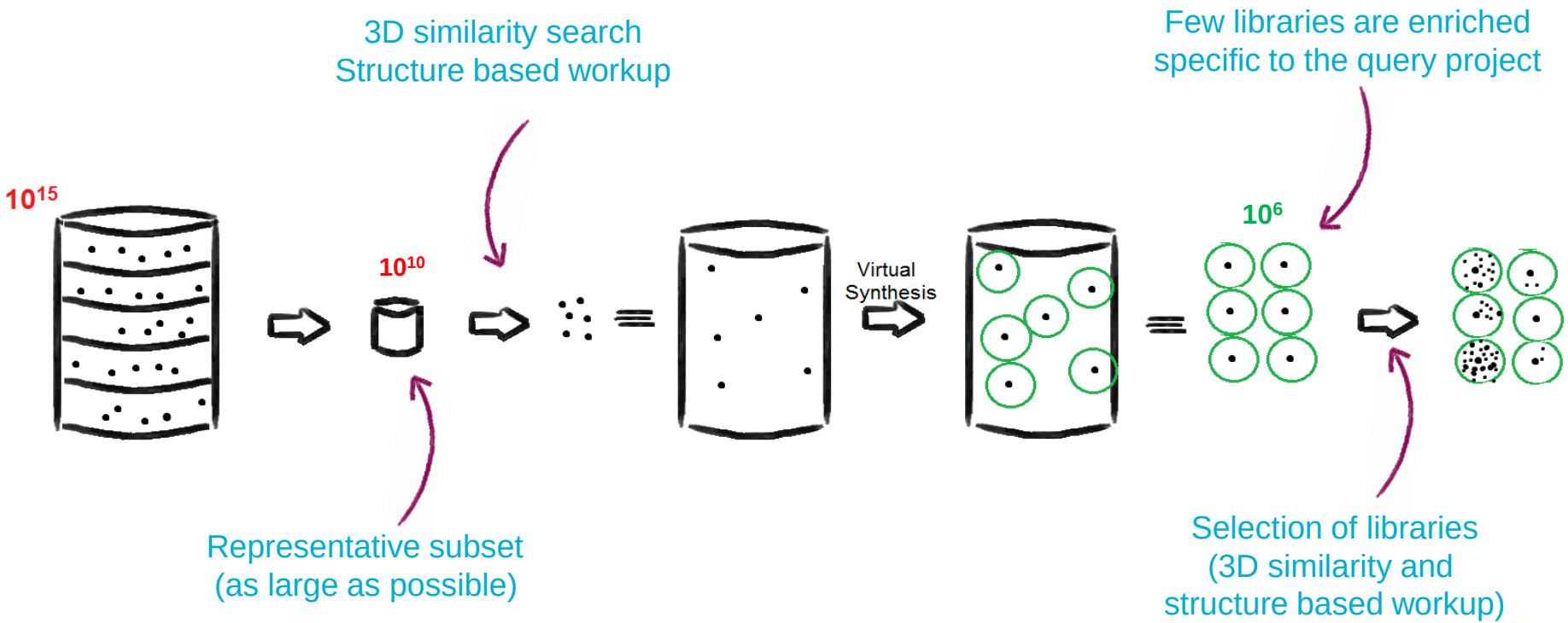
MJ Vainio *J. Chem. Inf. Model.*, **2012**, *52* (7), pp 1777–1786.

Virtual chemistry space



MJ Vainio *J. Chem. Inf. Model.*, **2012**, *52* (7), pp 1777–1786.

Computational strategy: brute force exploration of subsets

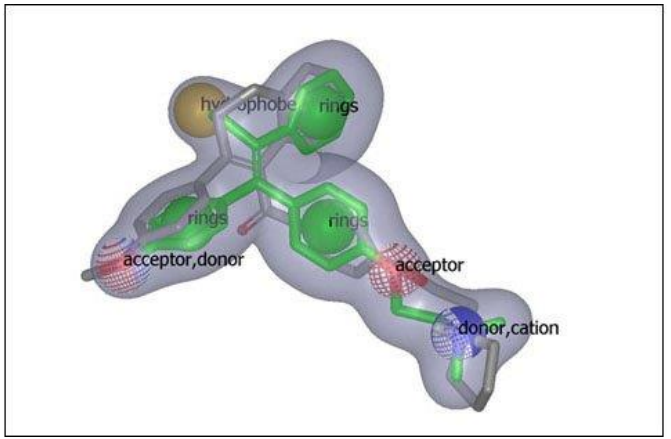


FastROCS: 3D similarity search

ROCS (rapid overlay of chemical structures) is a **shape**-based superposition method: aligning molecules based on their volume defined by Gaussian functions

The more similar the molecular shape and electrostatics, the more similar are the molecules.

FastROCS: GPU-implementation of ROCS; capable of doing more than **1M overlays/sec/GPU** vs 1k overlays/sec/CPU with ROCS



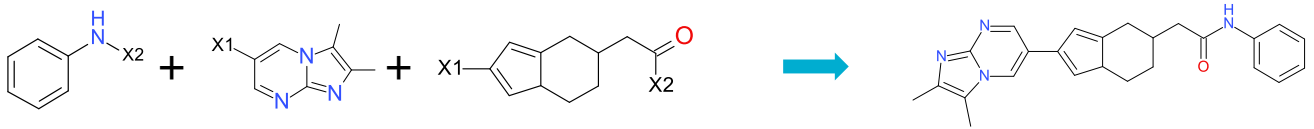
Grant, J.A., Gallardo, M.A., Pickup, B., *J. Comp. Chem.*, **1996**, *17*, 1653.
https://www.eyesopen.com/rocs

FastROCS: 3D similarity search

What do we need:

- Databases of pre-enumerated compounds
- Generation of 3D conformers.
- The size of the database defines the search time.
- Enumeration of a representative set defines search quality and diversity.

→ Enumerate as much as possible and searchable in reasonable time!

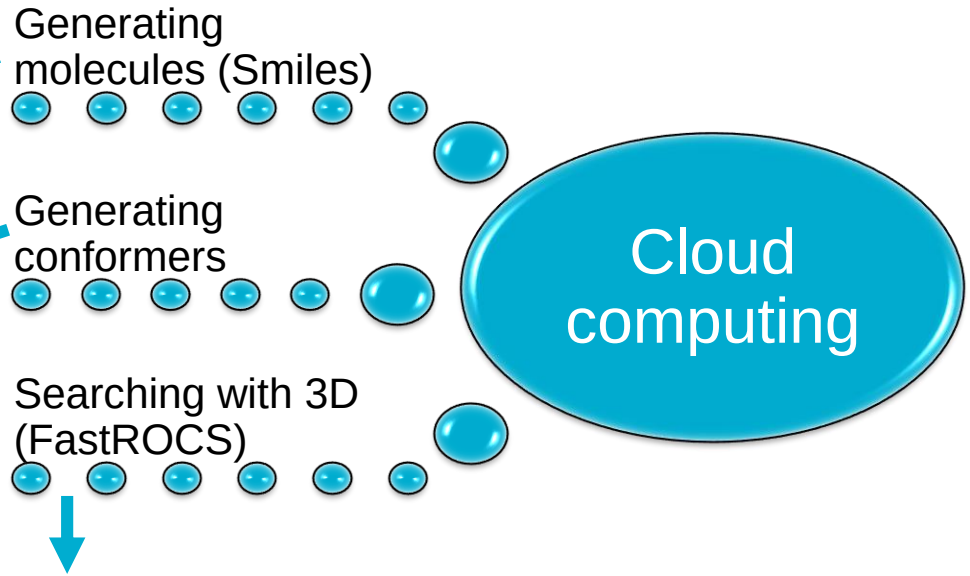


R1 + R2 + Scaffold → Product

FastROCS: 3D similarity search

- 10^9 in a few hours
- **10^{10} ~1.5 days on 400 CPUs**

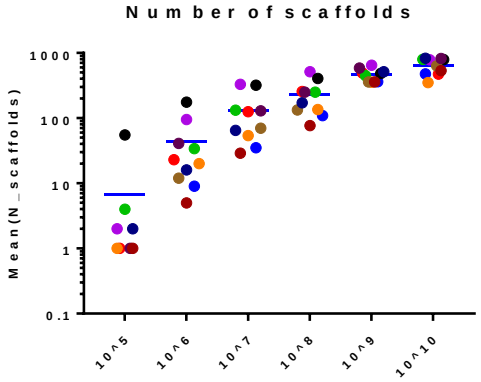
- 10^9 molecules in 24h on 16000 CPUs
- **10^{10} molecules in 55h on 29000 CPUs**
- TBs of data to store: 10^9 ~ 2TB; **10^{10} ~20TB**

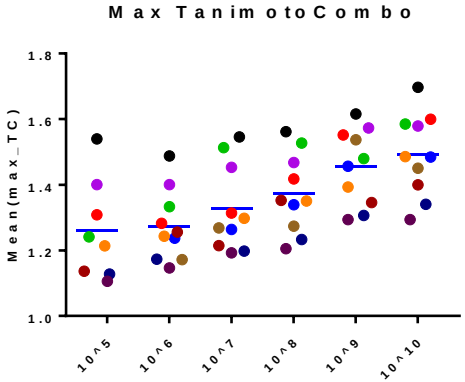


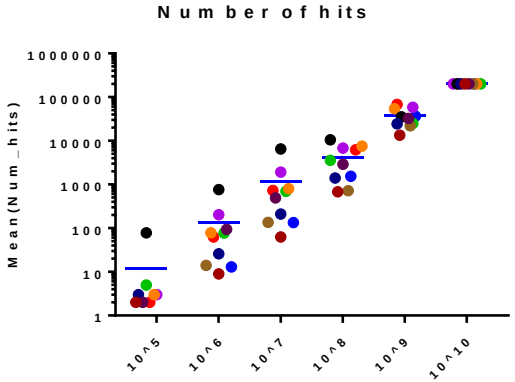
#molecules [million]	Loading [min]	Querying [min]	#GPUs
1330	23	18	50
12700	60+60	29+29	(85+85)

Is it worth it (in terms of chemical novelty)?

- Evaluate the chemical space (10^{15} molecules) using subsets and FastROCS
- Results relative to MinTanimotoCombo @200k hits in 10^{10}

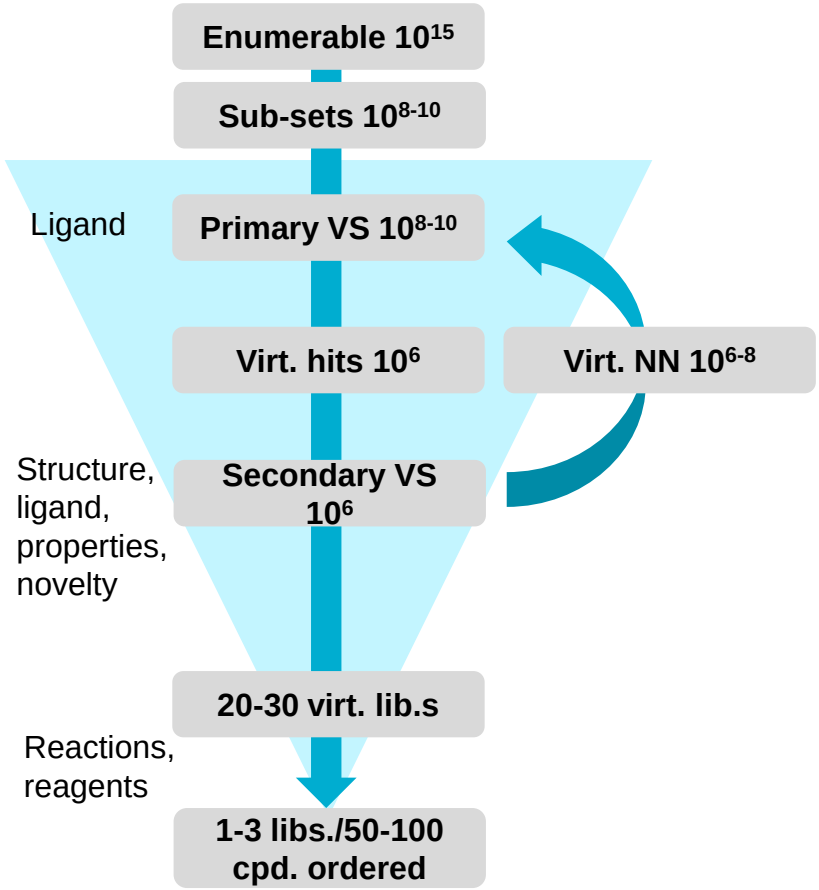


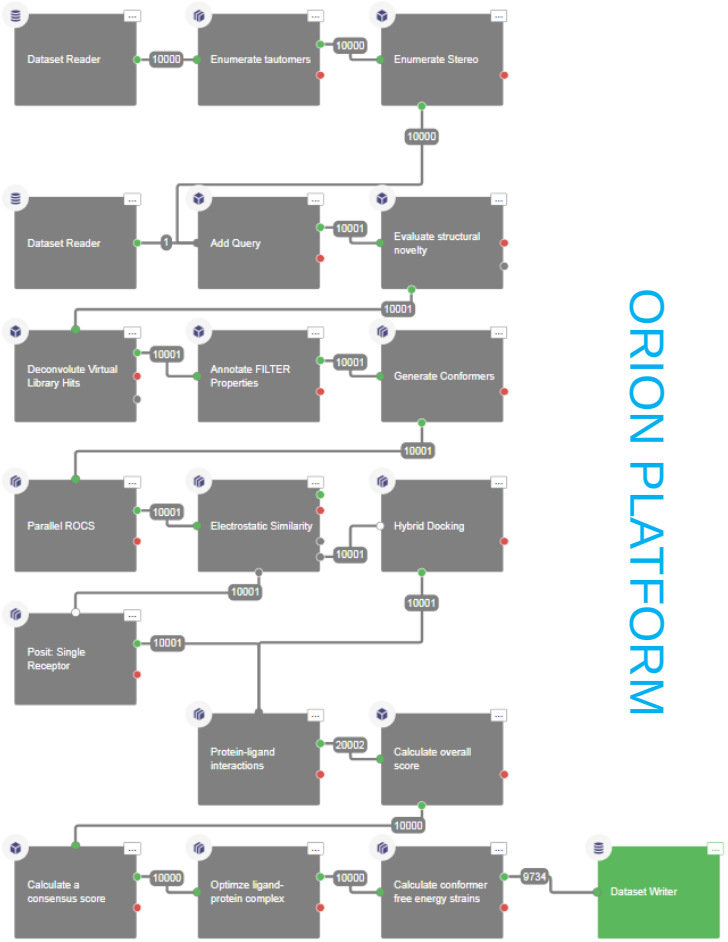




- ➢ Linear increase of hit numbers and scores
- ➢ Diversity (number of scaffolds) seems to converge
- ➢ More similar compounds are found
- ➢ Still „low“ sample rate: 10^{10} molecules = 0.001% of the whole space

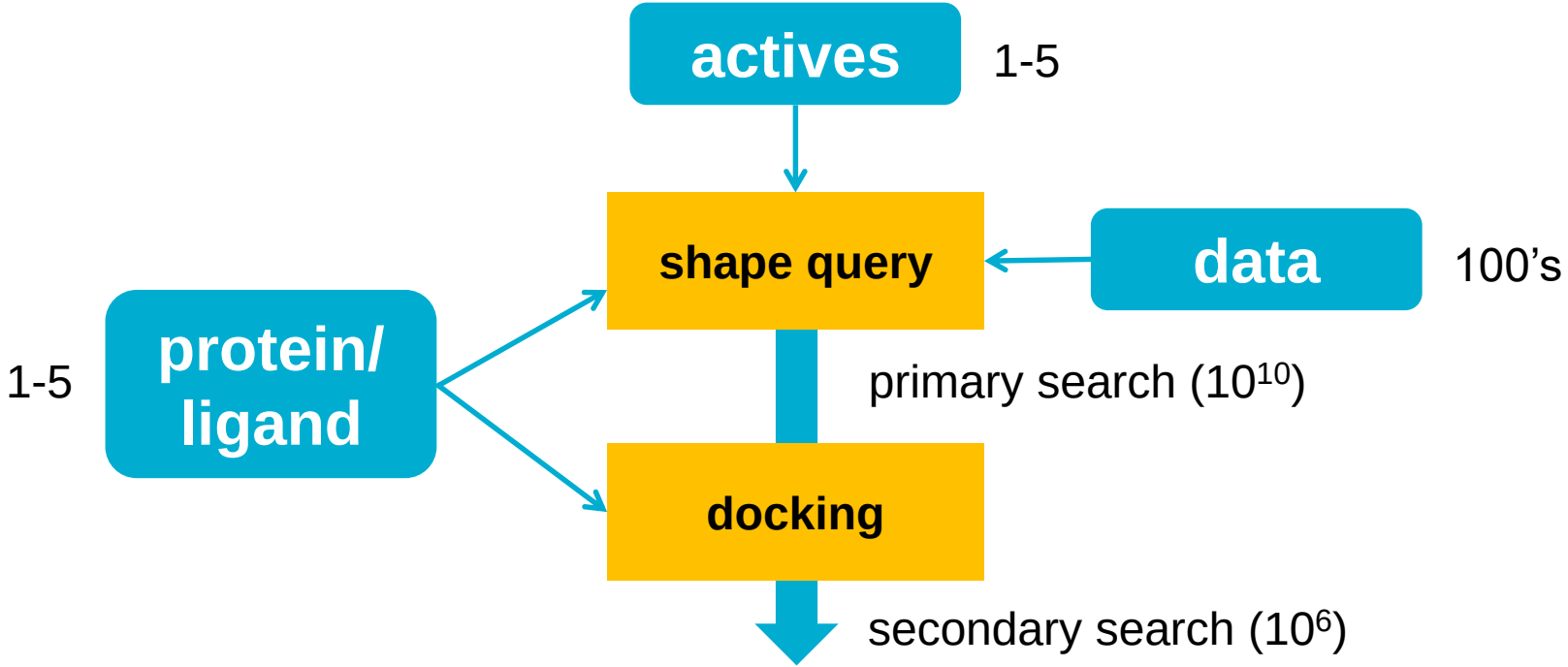
Iterative screening cascades





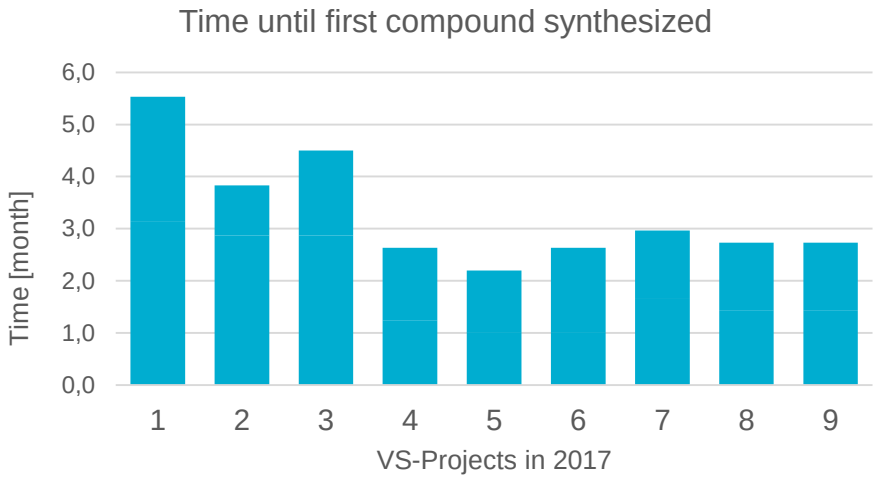
ORION PLATFORM

Prerequisites—what is the ideal project?



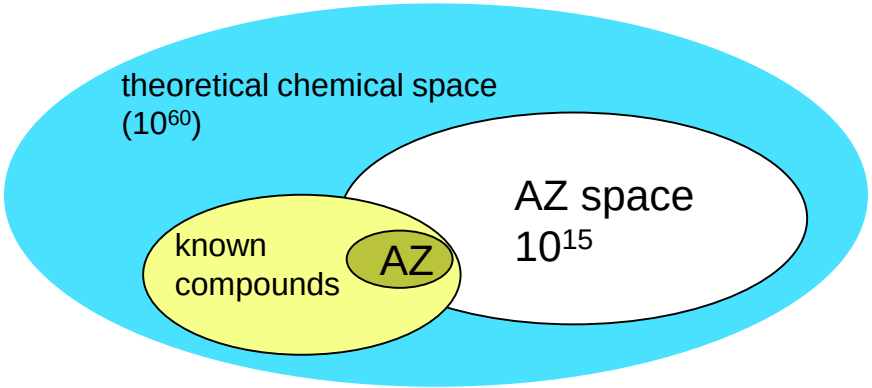
- 1 month from start of VS to compound order
- synthesis funded
- compounds ready before HTS re-test

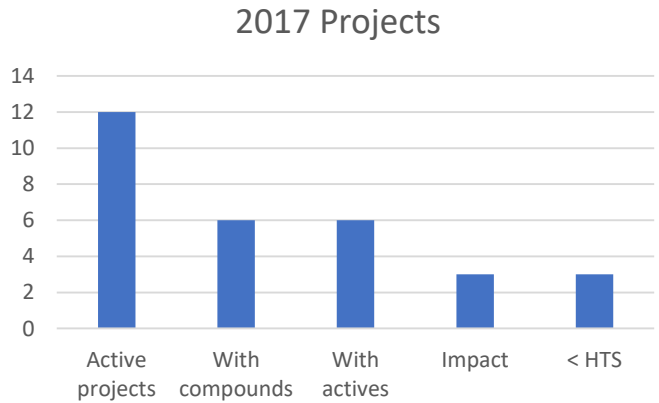
Can we make the compounds?



Average success rate in making compounds: ~70%

Summary





We use Orion for virtual screening

- Frequent use of virtual screening workflows
- Flexible way of working up millions of hits

We can search large sub-sets

- Able to enumerate 10 billion compounds and to generate 3D conformations
- Searching the space using FastROCS

We propose novel molecules, we make them, and we find hits

- Application of our workflows in projects
- Several novel active molecules
- Short delivery time: We know how to make these compounds

People

- Comp. Chem.: **Erik Malmerberg, Christoph Grebner, Thierry Kogej**
- Chemistry: **Michael Kossenjans, Nidhal Selmi**

- Global Comp. Chem.: **Jonas Boström, Anders Hogner, Martin Packer, Samantha Hughes, Christian Tyrchan**

- Hit ID Futures: **Steve Rees, Garry Pairaudeau**

Confidentiality Notice