Mid-term meeting

Progress report

Josep Arús-Pous
Prof. Jean-Louis Reymond
Dr. Hongming Chen
Dr. Thierry Kogej

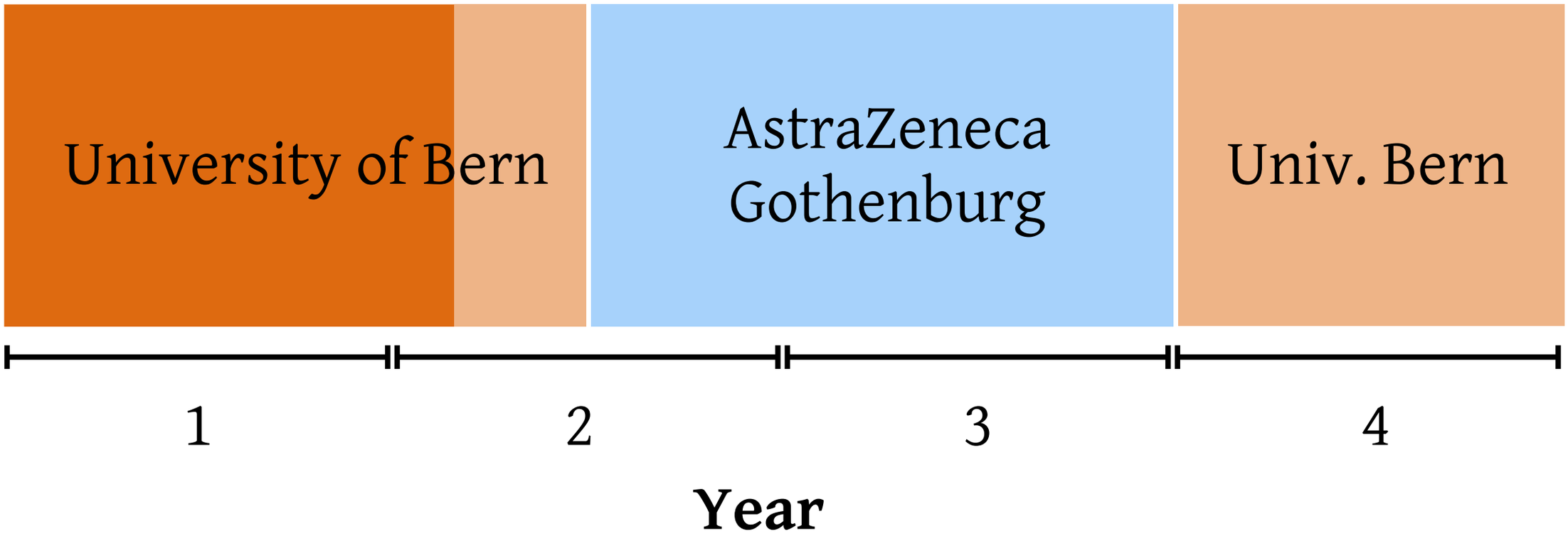
UNIVERSITÄT BERN

BigChem

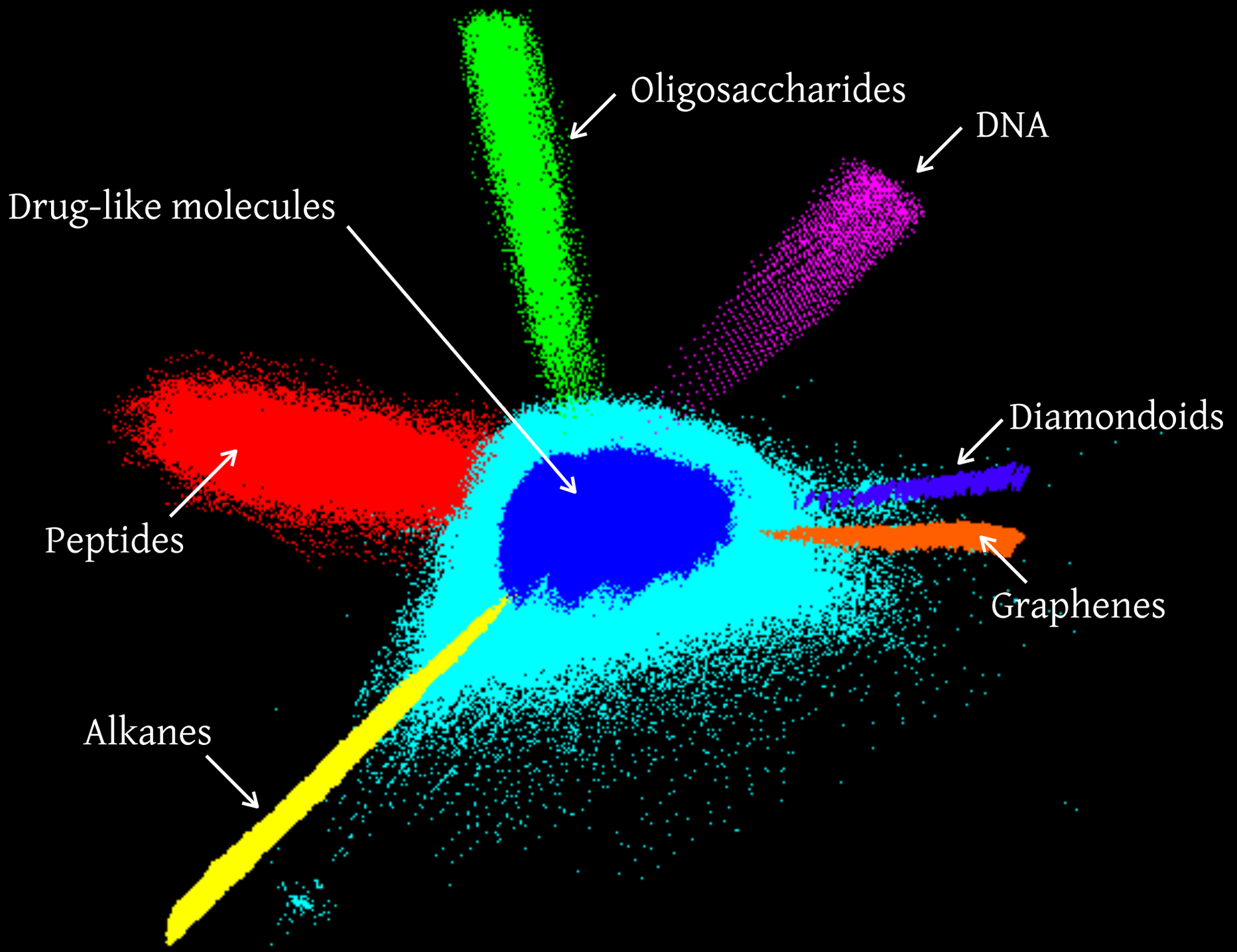
AstraZeneca

Who am I?

- I come from **Barcelona.**
- I'm a **Computer Scientist.**
- I have MSc in **Computer Eng.** and in **Bioinformatics.**



The Chemical Space



Reymond, J.-L.; Awale, M. ACS Chem. Neurosci. 2012, 3 (9), 649–657.

GDB17
(2012, 165 B, TiB)

GDB13
(2009, 1 B, GiB)

GDB11
(2005, 26 M, MiB)

Reymond, J. L. Acc. Chem. Res. 2015, 48 (3), 722–730.

What next?

GDB21 would be **at least 10^{15} molecules** (PiB).

↓

Change the way we explore the chemical space.

Working with **GDB17** is **extremely difficult.**

↓

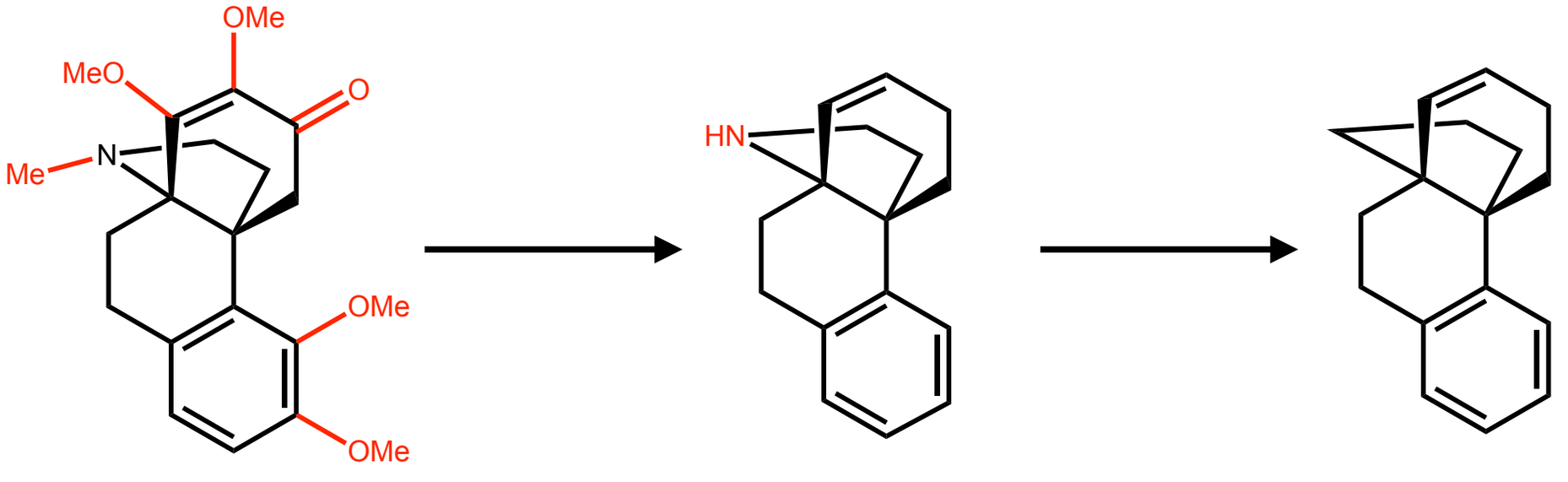
Create new tools to work with huge chemical databases.

Ring systems

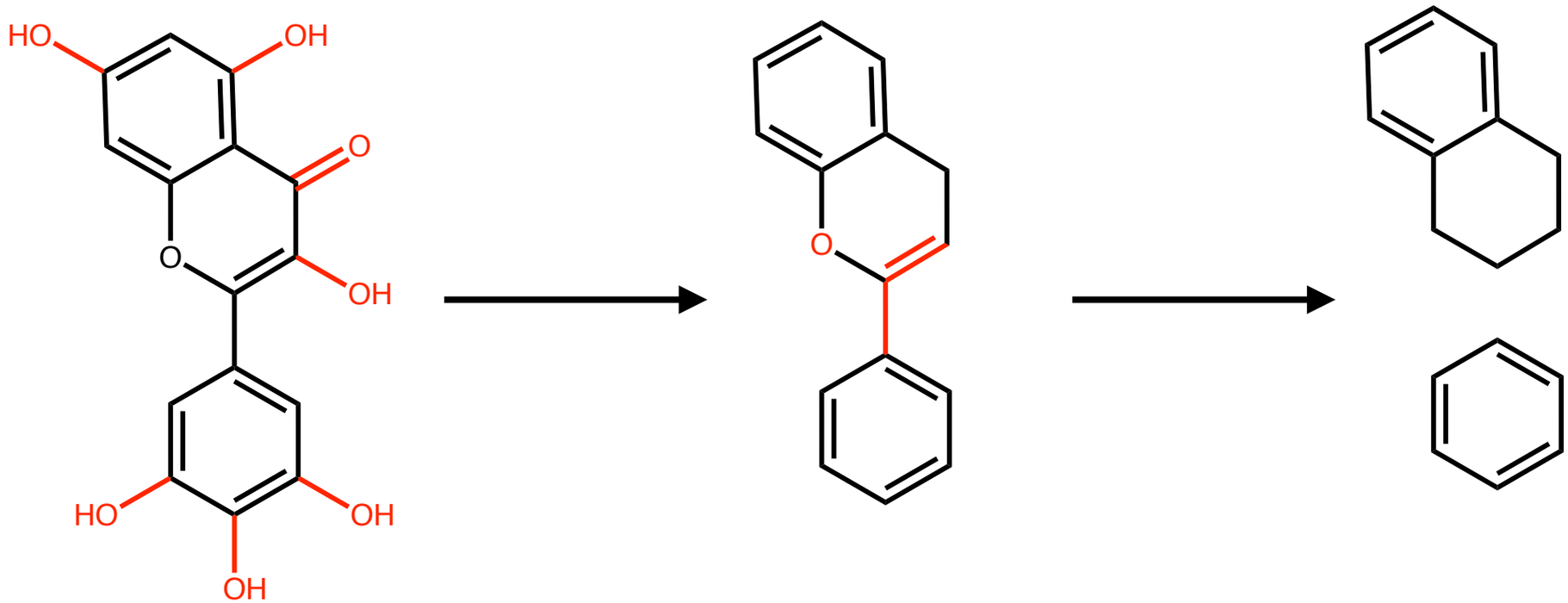
Molecule

Scaffold

Ring systems



Hasubanonine



Myricetin

GDB4c

Rules

- ✓ Max. **4 rings.**
- ✓ **First** ring **size ≤ 14.**
- ✓ **Second** ring **size ≤ 7.**
- ✓ **Other rings size ≤ 6.**
- ✓ **Aromatic** rings size 5 or 6 are allowed.
- ✓ **No small rings** next to **aromatic rings.**

Database	# RS
GDB4c	916,130
GDB4c3D	6,555,929
Reference DB	12,536
Reference DB3D	95,309

Just accepted

Visini, R.; Arus-Pous, J.; Awale, M.; Reymond, J.-L. J. Chem. Inf. Model. 2017, acs.jcim.7b00457.

Scaffold database

Rules

- ✓ Max. **size** = 30.
- ✓ Max. 5 **rings**.
- ✓ **Ring size ≤** 7.
- ✓ Max. **2 rings per ring system**
- ✓ **1 small ring** allowed.
- ✓ No aromatics
- ✓ No heteroatoms.

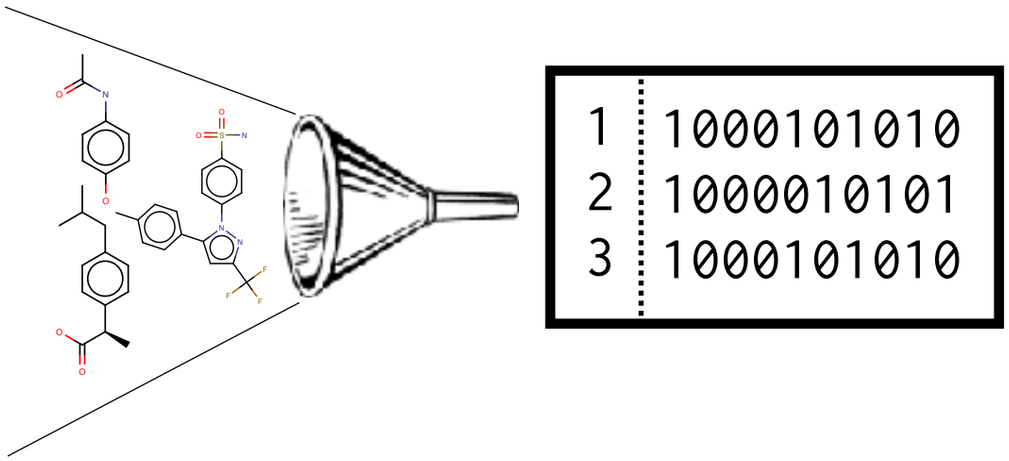
- Contains carbon skeletons (CSKs).
- Contains **~46 M** scaffolds.
- **97%** of novel scaffolds.
- **84,7%** of all known scaffolds in absolute numbers.

Work in progress

Tools - Codex

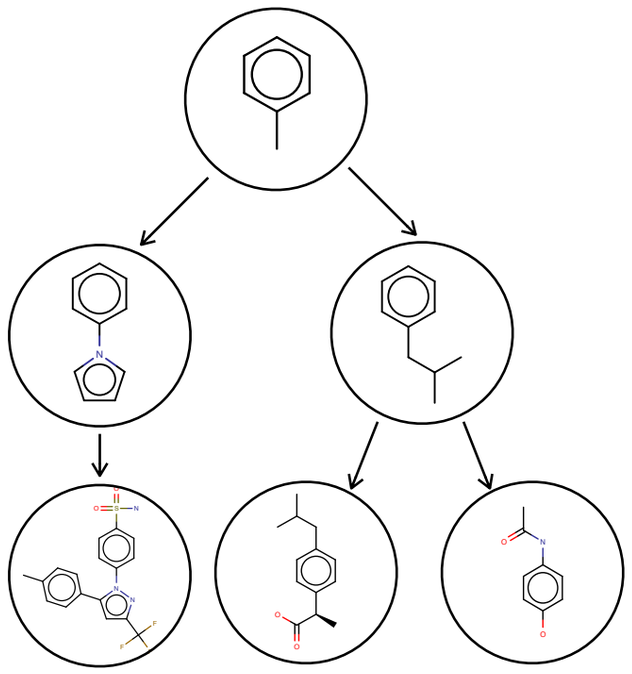
- A specific database for chemical entities with the following features:

1) Writable once ➔ Read-only



1	1000101010
2	1000010101
3	1000101010

2) Data is fuzzily pre-ordered for fast search



3) Heavily compressed

Snappy **ZSTD**

Currently under development

Achievements

Publications:

Visini, R.; Arus-Pous, J.; Awale, M.; Reymond, J.-L. *J. Chem. Inf. Model.* 2017, acs.jcim.7b00457.

Awale M.; Visini R.; Probst D.; Arús-Pous J.; Reymond J.-L. *Chimia*. 2017

Conferences:

12-14 February 2017 **Frontiers in Medicinal Chemistry 2017**; University of Bern

21-22 August 2017 **SCS Fall Meeting 2017**; University of Bern POSTER

5-7 November 2017 **German Conference on Chemoinformatics**; Mainz POSTER

Awards:

4-5 September 2017 **First year PhD symposium**; University of Bern/Fribourg/ Neuchâtel; Best presentation

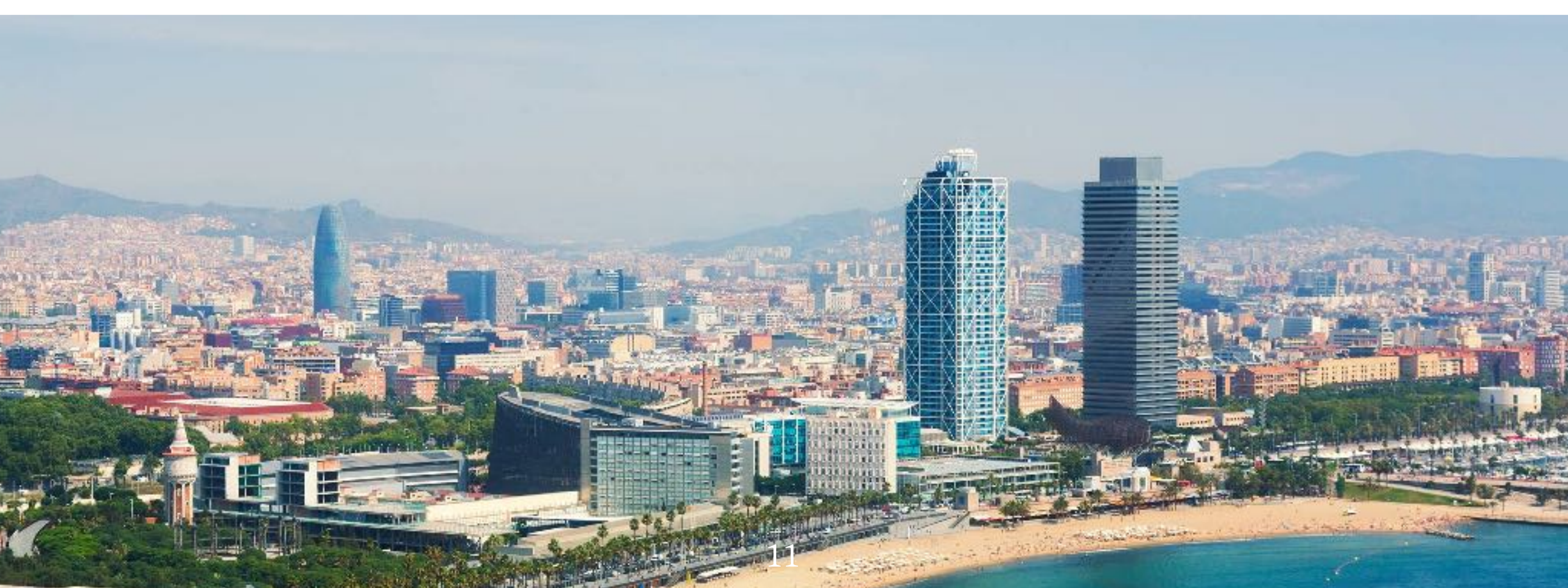
Outreach events:

5 November 2016 **Tag der öffenen Tür**; University of Bern

16 September 2017 **Nacht der Forschung**; University of Bern

BigChem Schools

- I participated in the **München Autumn School 2016.**
- I was part of the organization of the **BigChem BCN 2017.**
  - Located at the **PRBB**, one of the most important biomedical research centers in Europe.
  - **More than 50** participants, including ~30 external.



Future

- **Finish** the two projects:
  - Have a working prototype of the **database technology**.
  - Finish to characterize the **scaffold database.**
- Secondment with Dr. Jordi Mestres in **Chemotargets.**
- In March 2018 I start the **second half** of the PhD in Gothenburg.

Thank you for your attention

Thanks to:
Prof. Jean-Louis Reymond
Dr. Hongming Chen
Dr. Thierry Kogej
Ricardo Visini
Mahendra Awale
Daniel Probst
Sacha Javor

The whole Reymond Group