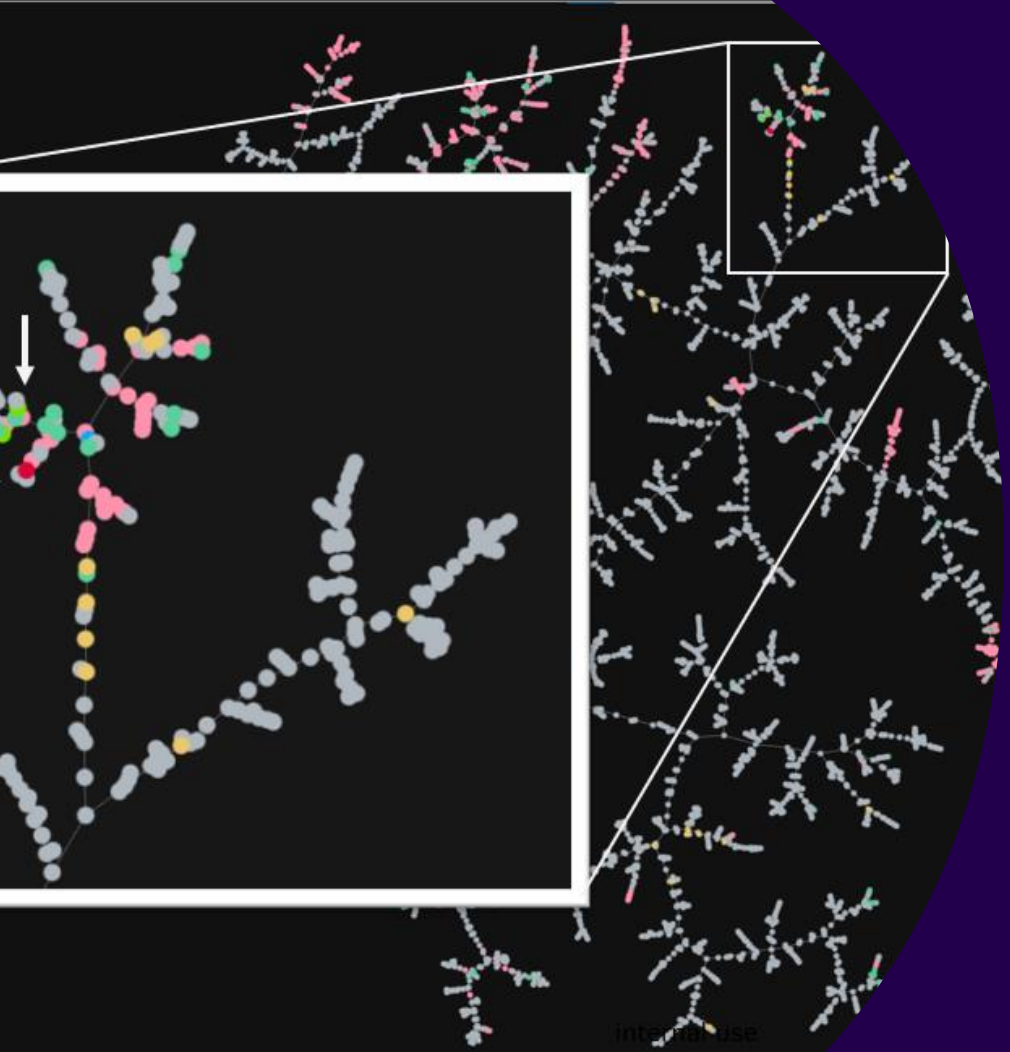

sanofi



sanofi

Practice of Generative AI in lead optimization

Marc Bianciotto

March 14th 2024

AiChemist Spring School - Berlin

Dr. Maximilien Levesque

(now Aqemia)

Dr. Maxime Langevin

Pr. Rodolphe Vuilleumier

And in Sanofi

CADD: B. Filoche, C. Terrier, C. Grebner, Y. Li, S. Guessreguen, H. Matter, S. Sauer

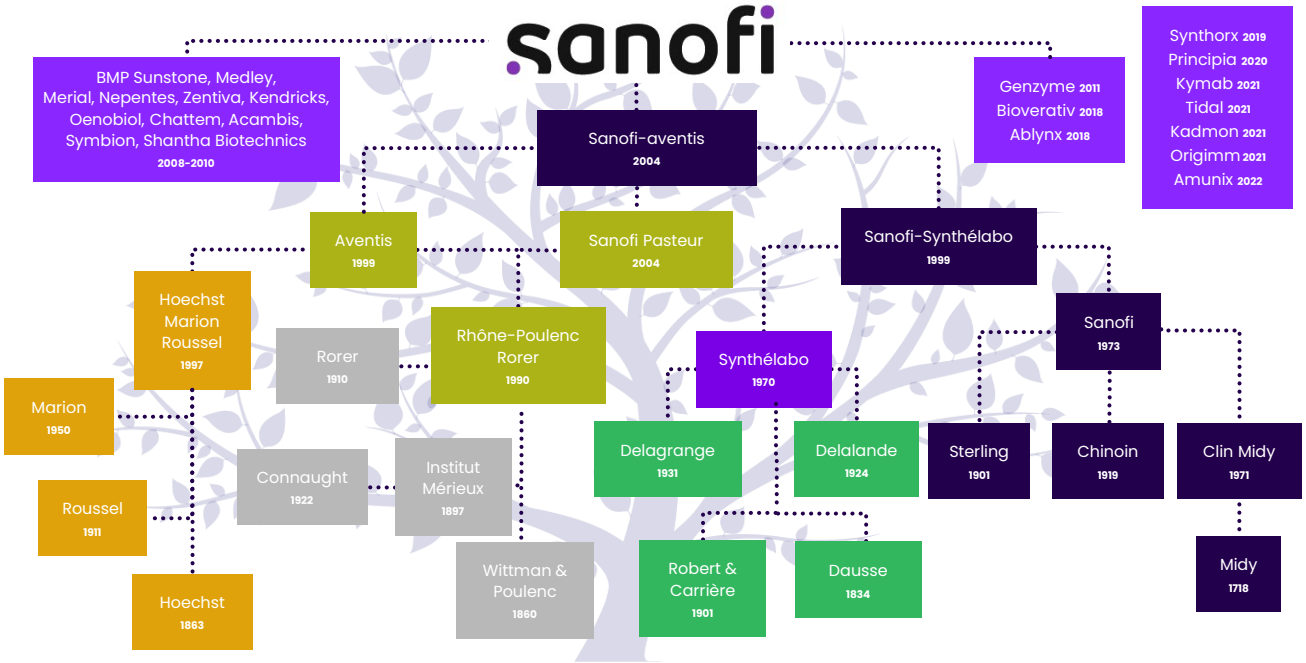
Chemistry: S. Desprets and the medicinal chemists involved in the Molecular Turing Test

"Marc Bianciotto is a Sanofi employee and may hold shares and/or stock options in the company"


We operate in
90
countries
There are
91,000
Sanofians
worldwide
We work in
59
manufacturing
sites
We have
20
R&D sites

sanofi

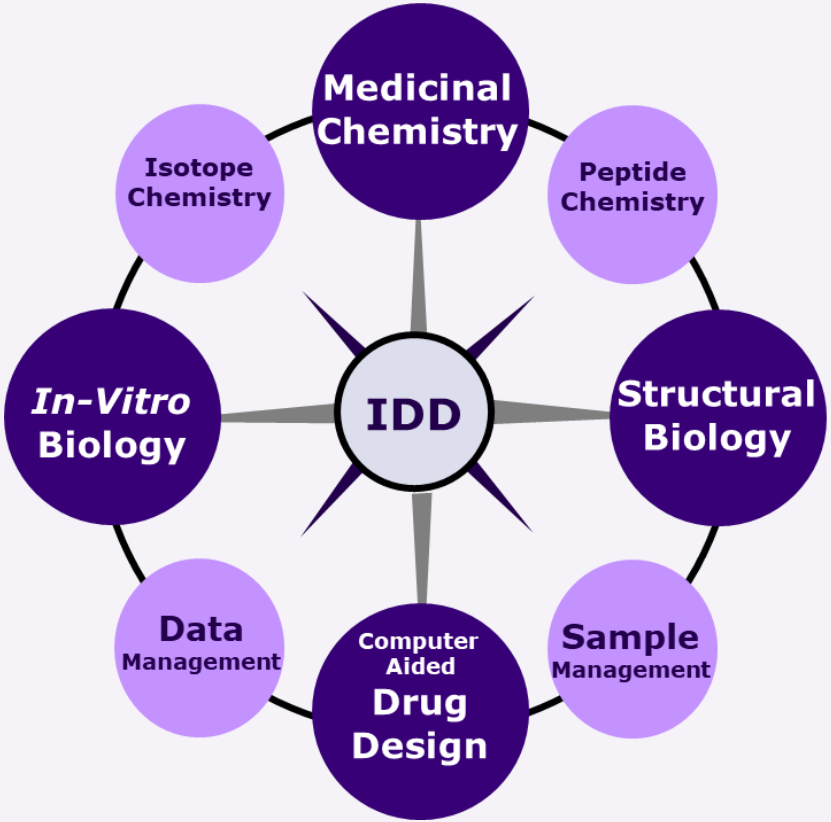
The history of Sanofi



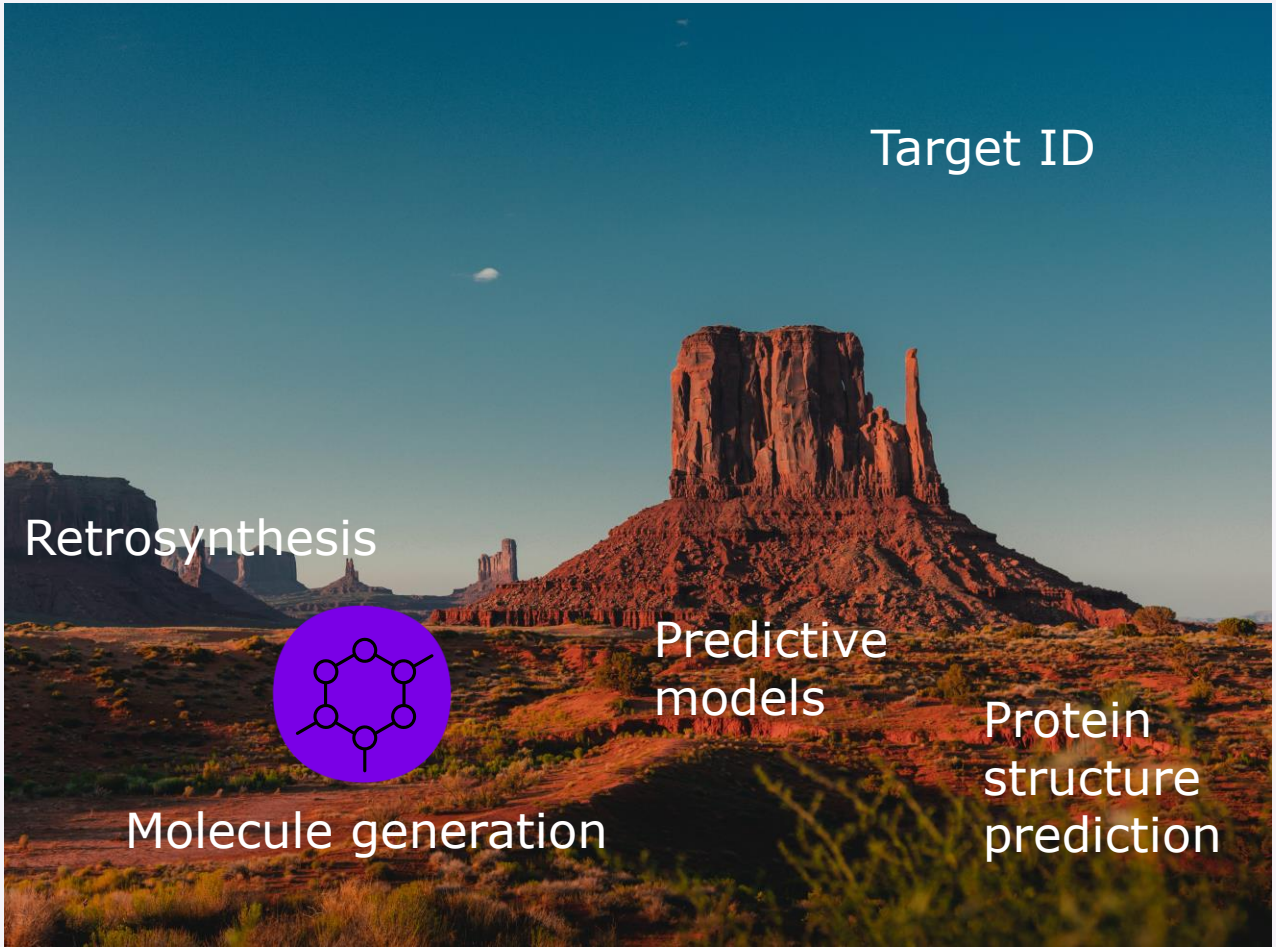
Integrated Drug Discovery

sanofi

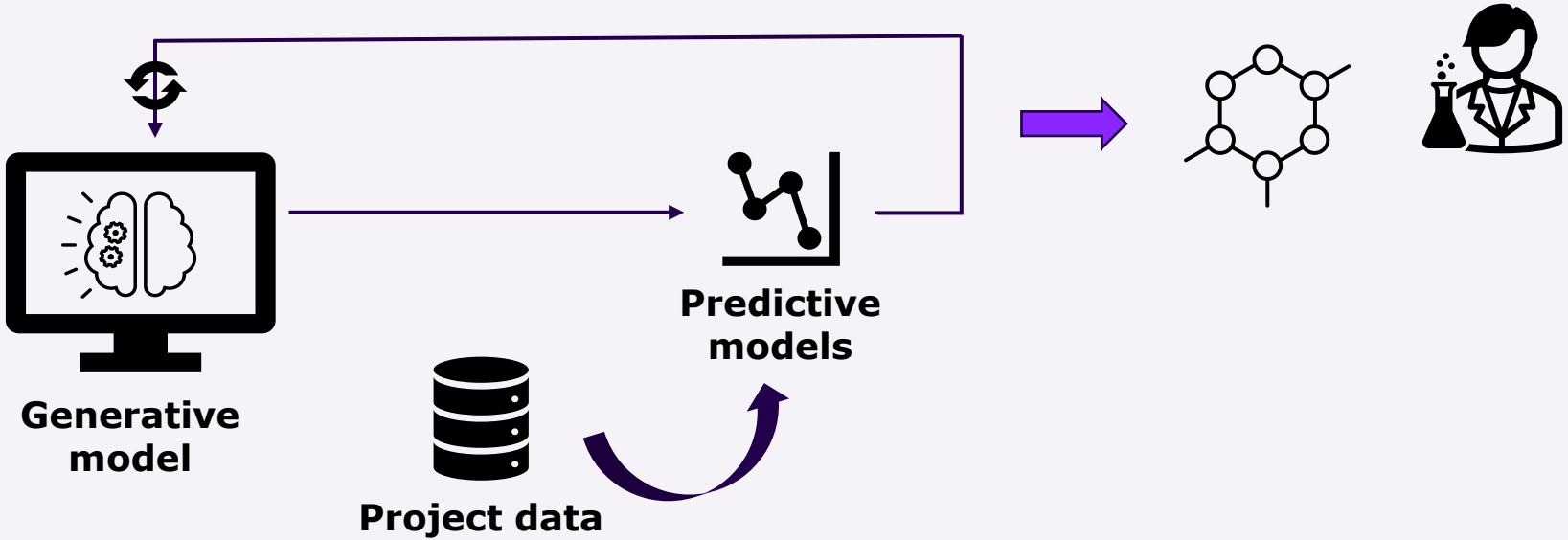
Integrated Drug Discovery France



AI for Drug Discovery: the Gold Rush



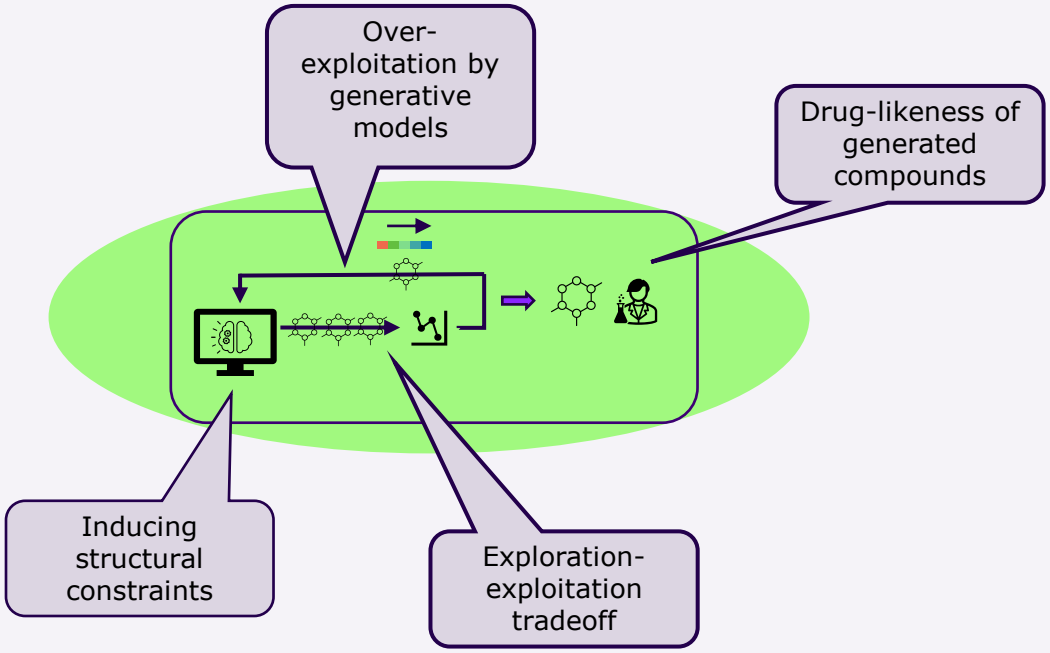
Generate molecules with AI?

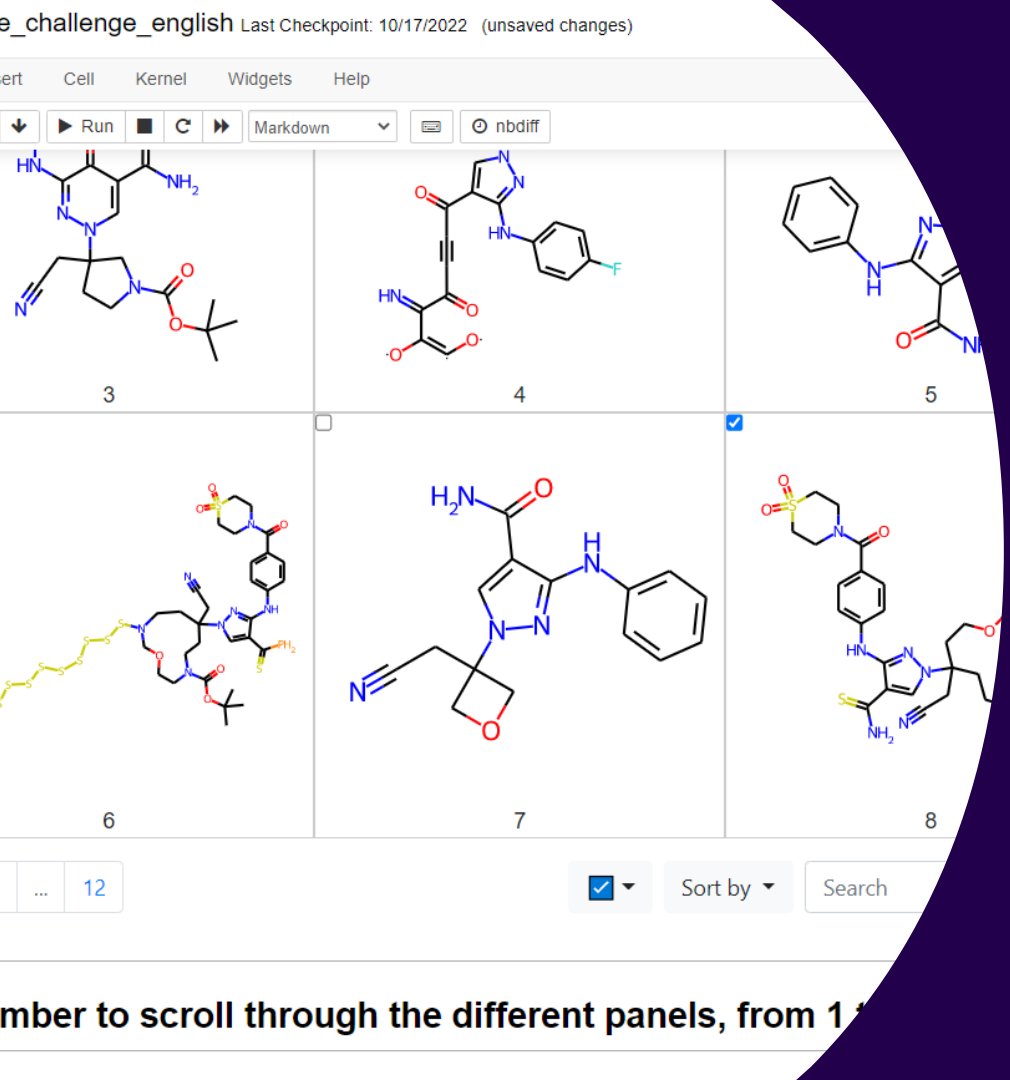


Generate molecules with AI?

	Lead Gen.	Lead Opt.
Project data	Less	More
Predictive models performance	Worse	Better
Chemical space to explore	Larger	More focused
Synthesis likelihood	Lower	Higher
Integration in the legacy workflow	More challenging	Less challenging

Topics of concern in small molecule GenAI for lead optimization





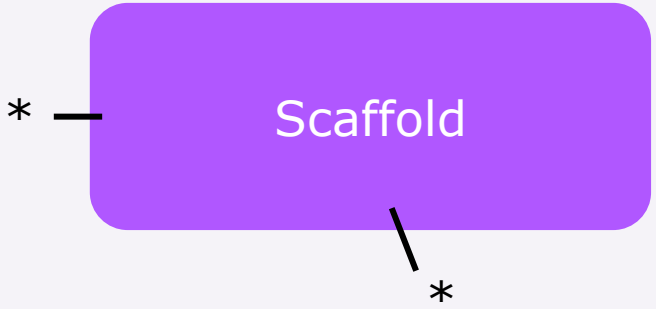
sanofi

Translation to practice

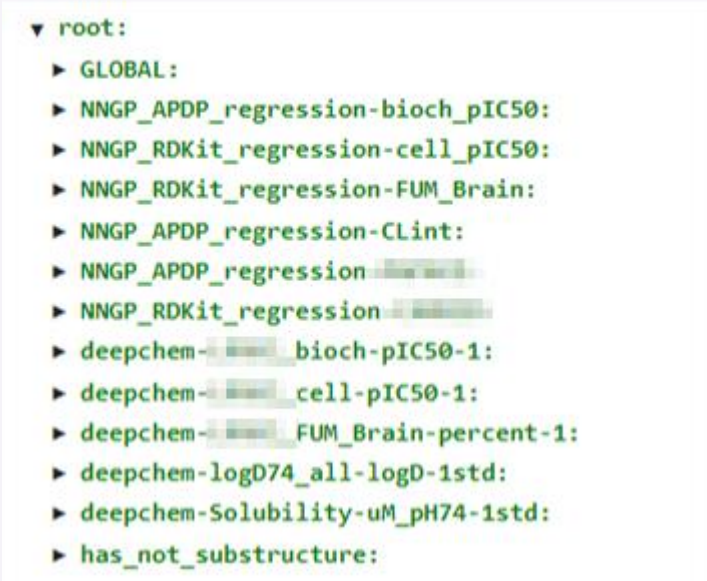
Generative AI for protein kinase X: last effort

Objectives:

- Evaluate Med Chemists proposals
- Generate a last round of compounds
  - Close to the pre-candidate (same scaffold, 2 diversification points)
  - **Optimized on 8 properties**
  - Synthetically readily accessible

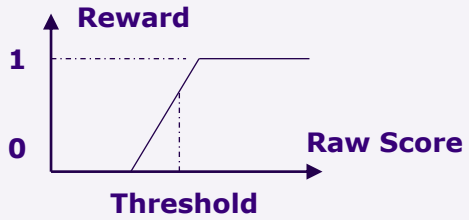


Protein kinase X: reward function



- Composite scoring function
  - Biochemical, cell pIC50s, counter-targets, unbound fraction in Brain, solubility, logD
  - Several models per readout if the models are correct
- AD for generation

- Score thresholds from the Med. Chemists' propositions



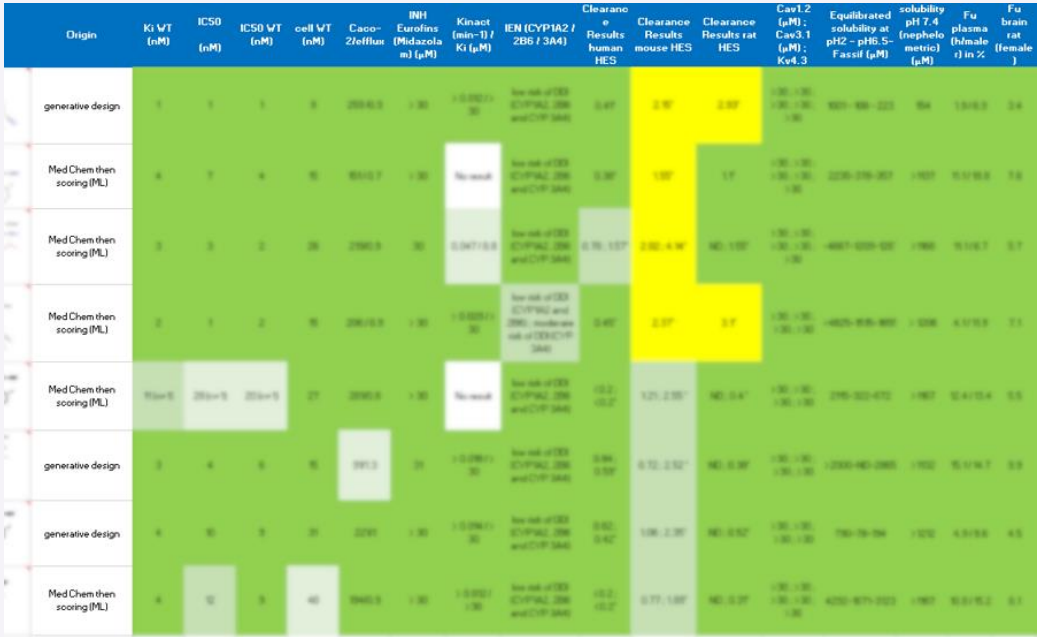
Generation outcome

- Human intelligence + ML score:
  - **73 close analogs of known cpds**
  - 12 « flat cpds » selected

  => 23 synthesized with stereo

- AI design:
  - 857 cpds generated in-criteria
    - **108 selected for Med Chemists**
  - 10 « flat » cpds selected for synthesis

  => 17 synthesized with stereo.
  - **Non-trivial R-groups**



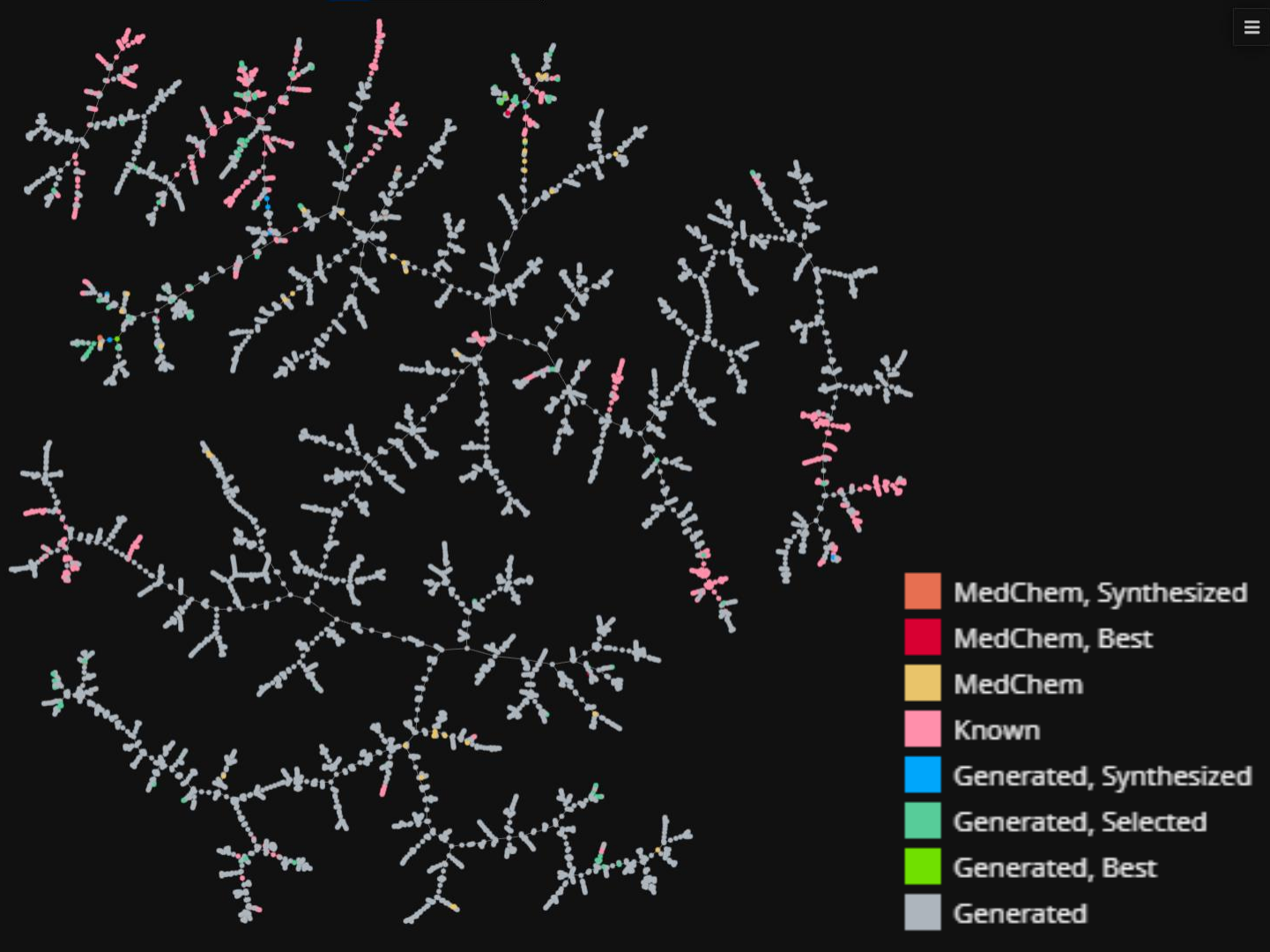
A shortlist of 5+3 cpds that fulfill the vast majority of 15 criteria

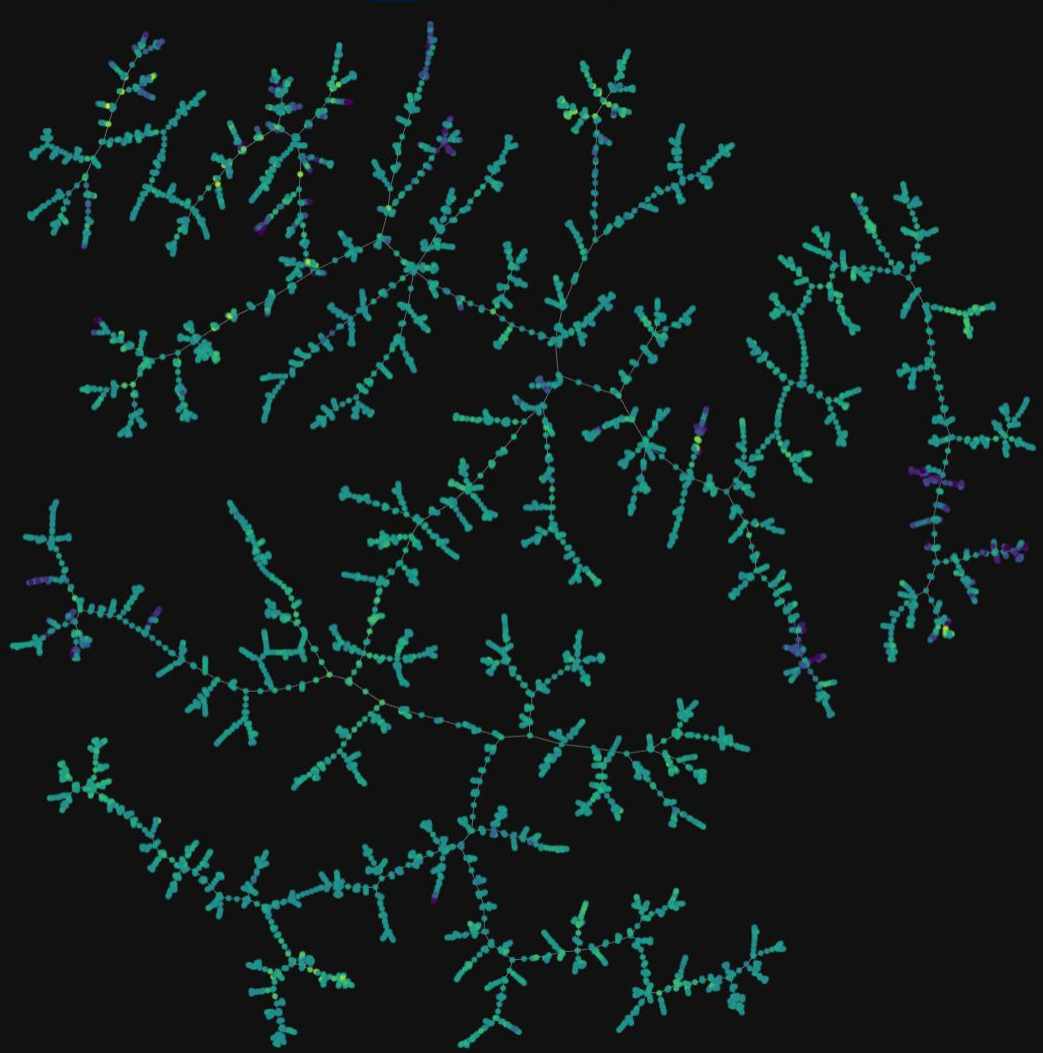
Chemical space

Colored by

molecule origin and fate

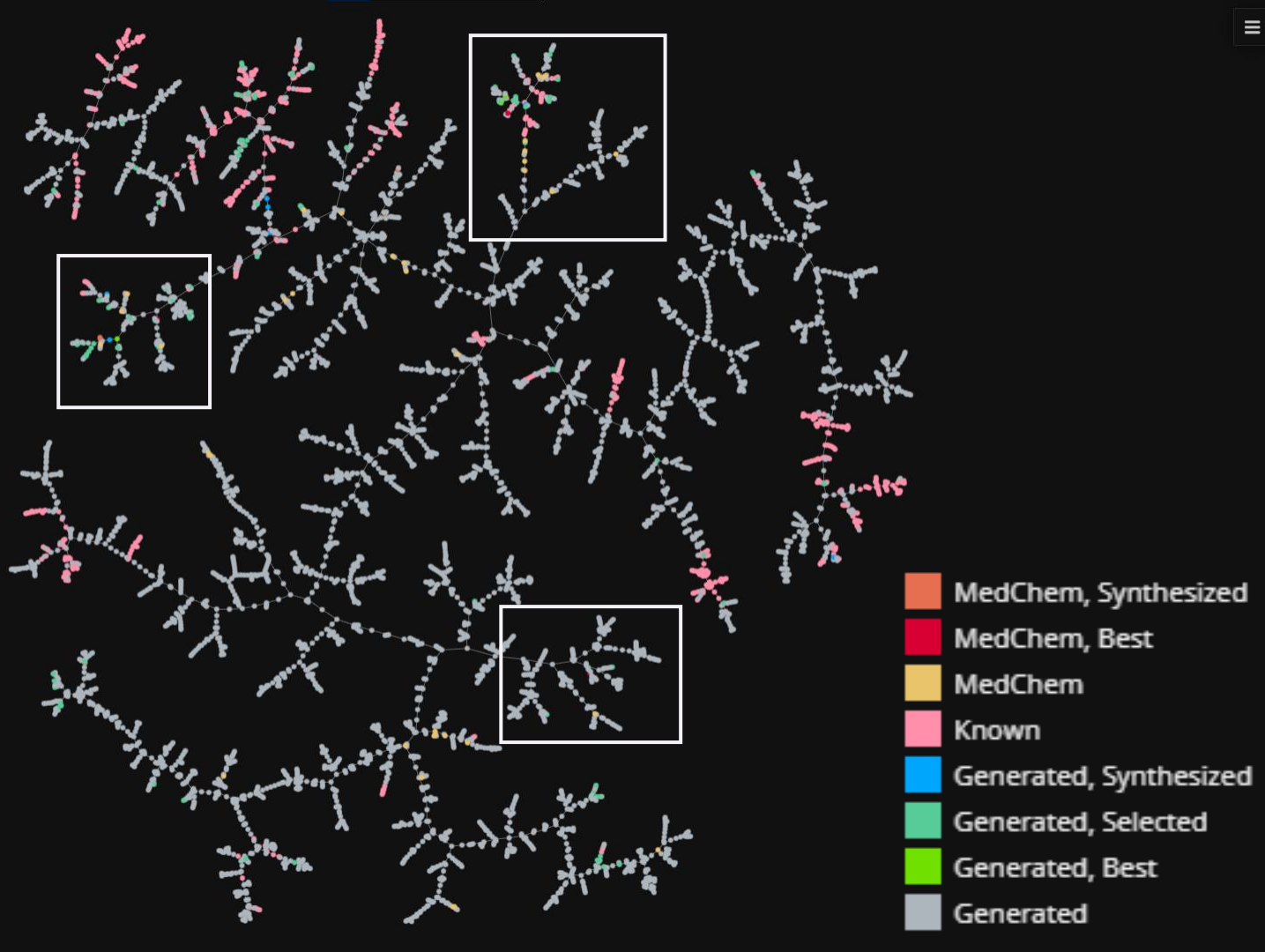
sanofi

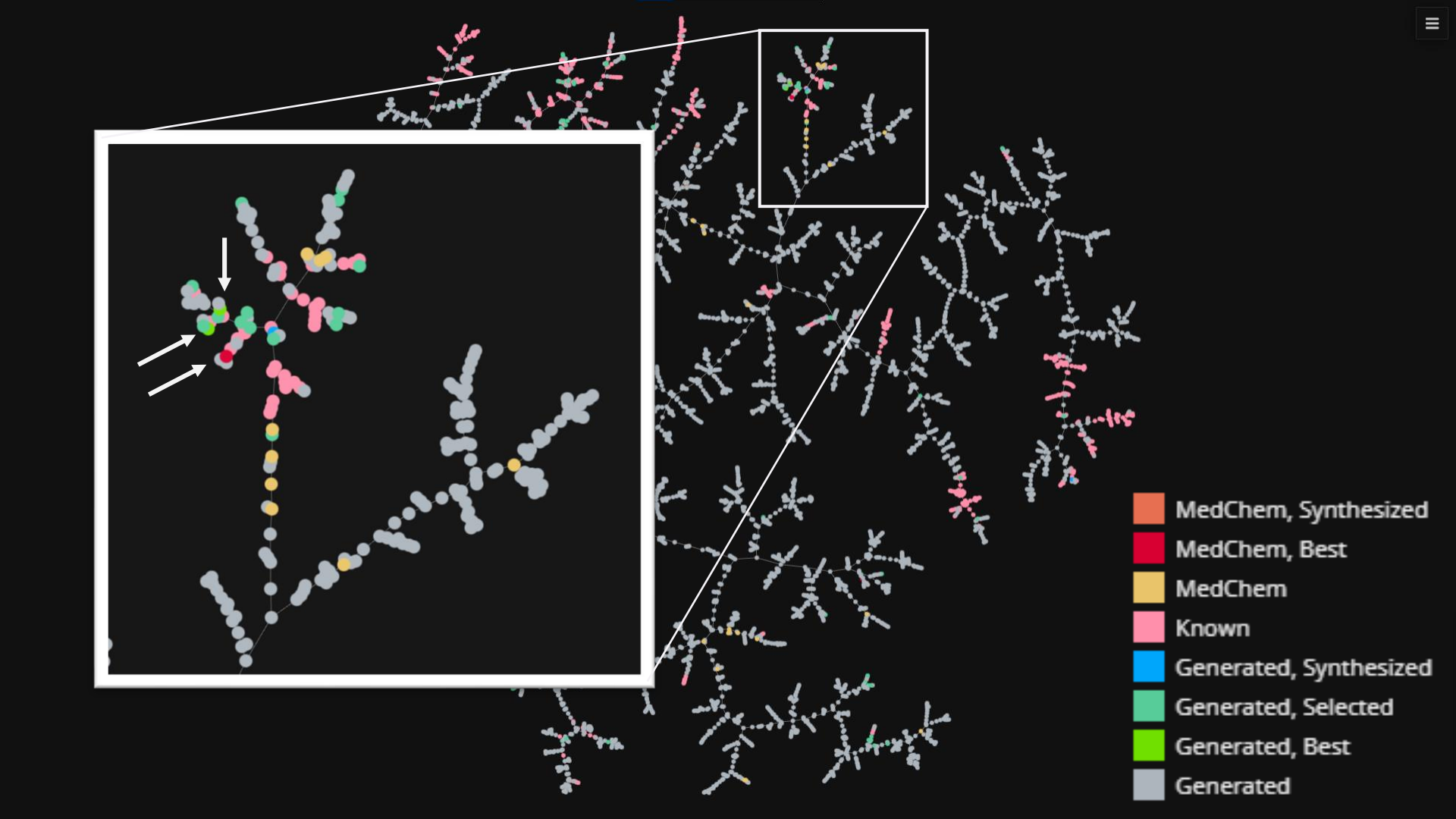

MedChem, Synthesized
MedChem, Best
MedChem
Known
Generated, Synthesized
Generated, Selected
Generated, Best
Generated

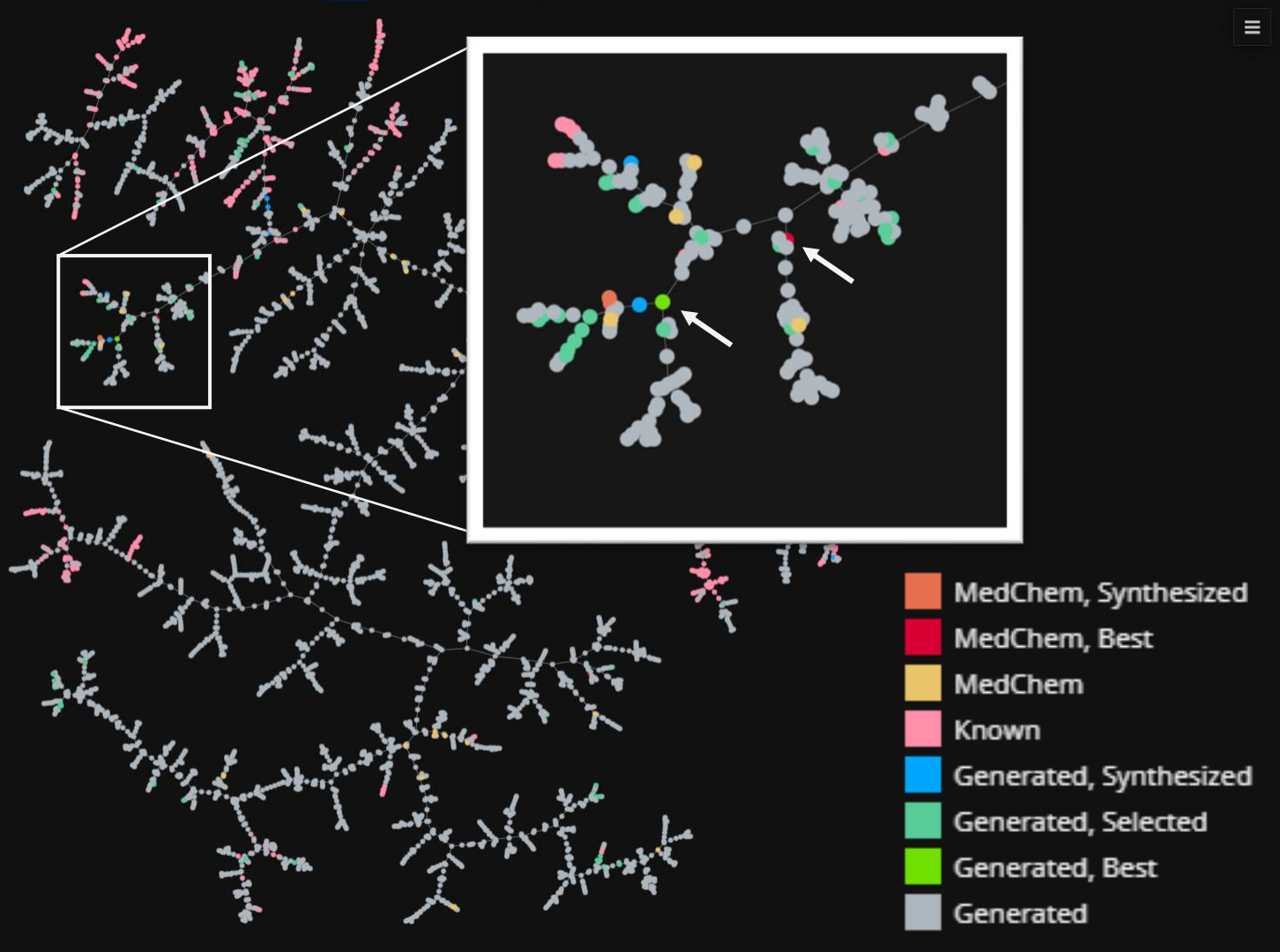


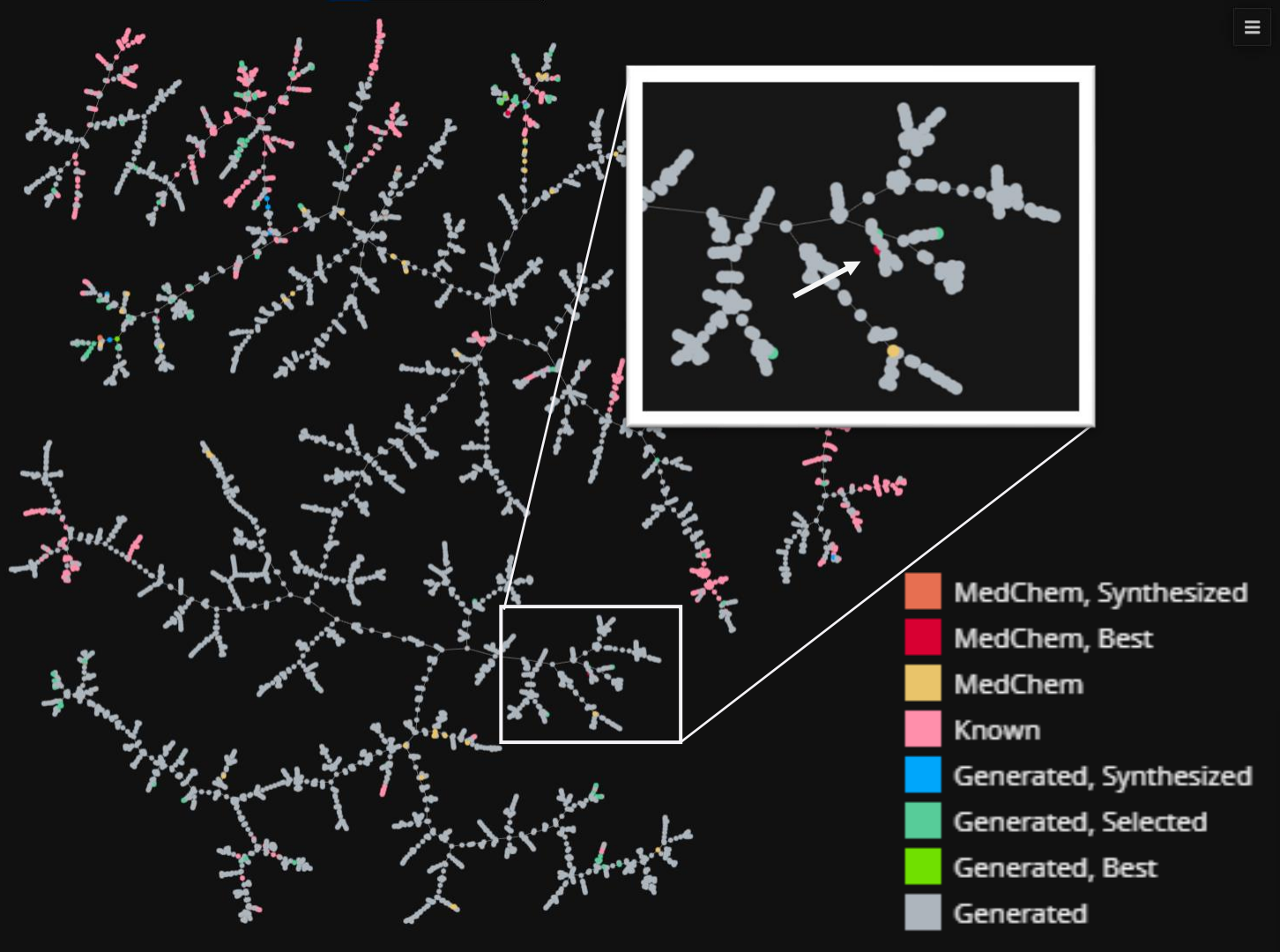
Colored by

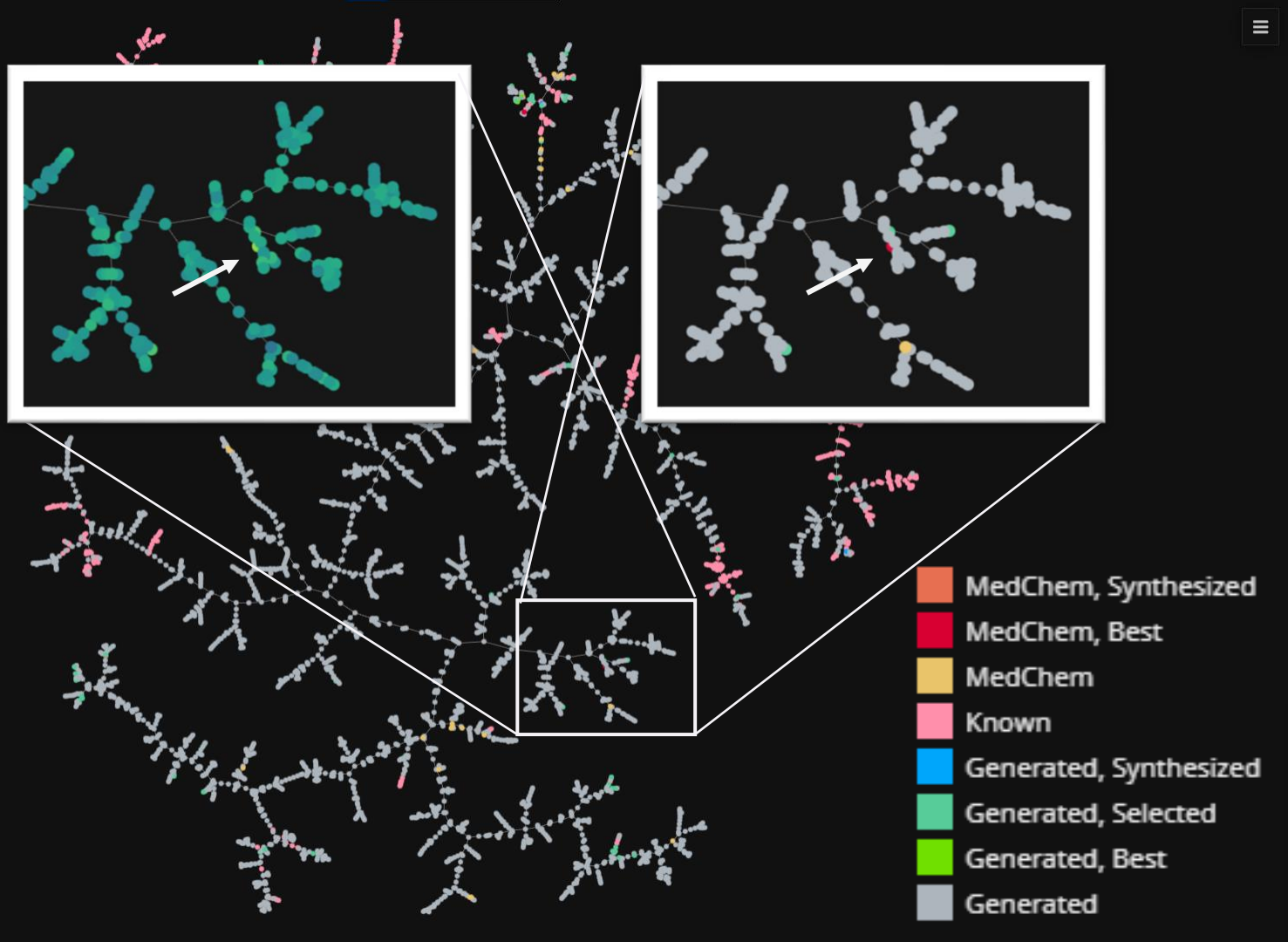
Multi-parametric score


MedChem, Synthesized
MedChem, Best
MedChem
Known
Generated, Synthesized
Generated, Selected
Generated, Best
Generated


MedChem, Synthesized
MedChem, Best
MedChem
Known
Generated, Synthesized
Generated, Selected
Generated, Best
Generated

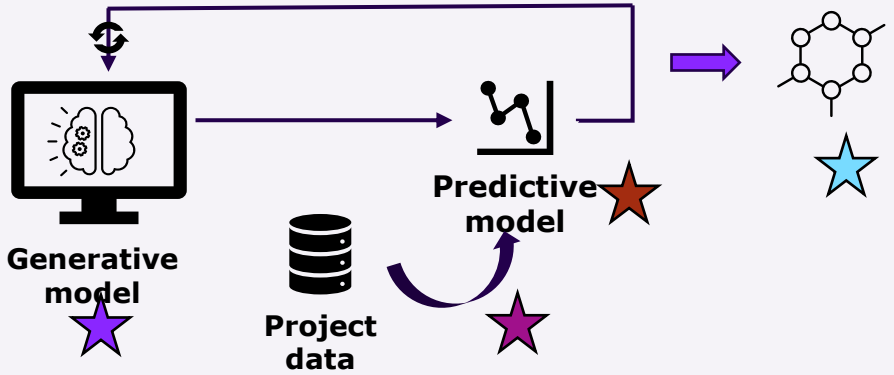

MedChem, Synthesized
MedChem, Best
MedChem
Known
Generated, Synthesized
Generated, Selected
Generated, Best
Generated


MedChem, Synthesized
MedChem, Best
MedChem
Known
Generated, Synthesized
Generated, Selected
Generated, Best
Generated


MedChem, Synthesized
MedChem, Best
MedChem
Known
Generated, Synthesized
Generated, Selected
Generated, Best
Generated

Using Chemical Language Models: Practical considerations on execution

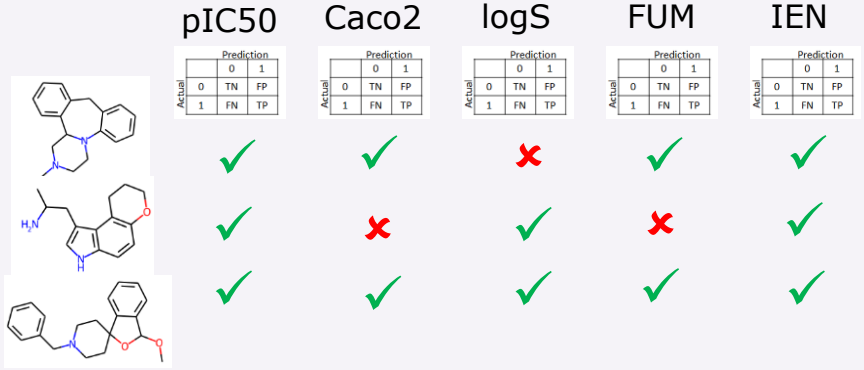
- Timing
  - The lead optimization GenAI window
    - Enough data for good models
    - Still useful to explore (vs exploit)
- "Data Science is teamwork"
  - Even with chemical lead opt data
  - Curate curate curate: Speed and quality



Using Chemical Language Models: Practical considerations on integration

- Define clear and shared objectives
- From screen to compound:
  - Quality & synthetisability
  - Or BB availability?
- From screen to data: cycle time
- Acculturate on statistics

- Explicit, transparent, shared design choices
- Generative approaches as part of a strategy
  - Explainable decision making



5 binary models,
success rate 0.8

$$0{,}8^5 = 0.33$$

Thank *you!*

sanofi