IMTM

ENBIK 2025
National Bioinformatic Conference

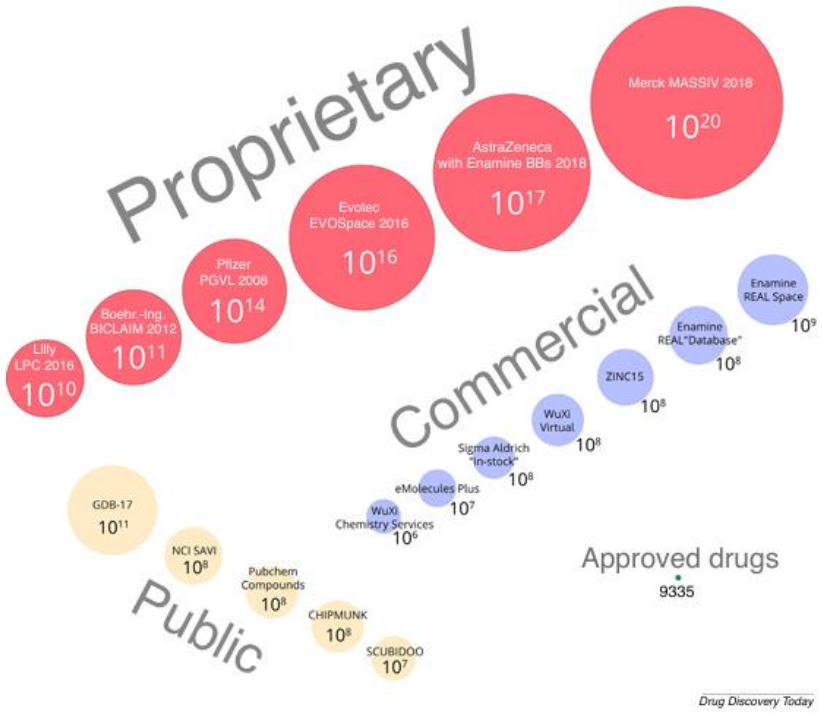
9-11 June 2025
Mala Moravka, Czech Republic

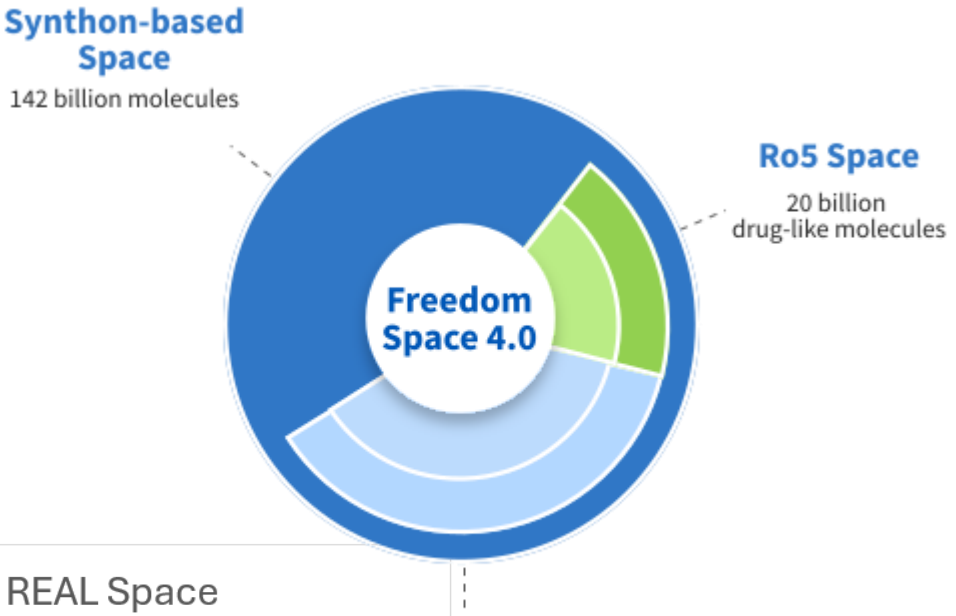
Fragment-based de novo design and searching for hit molecules in ultra-large chemical libraries

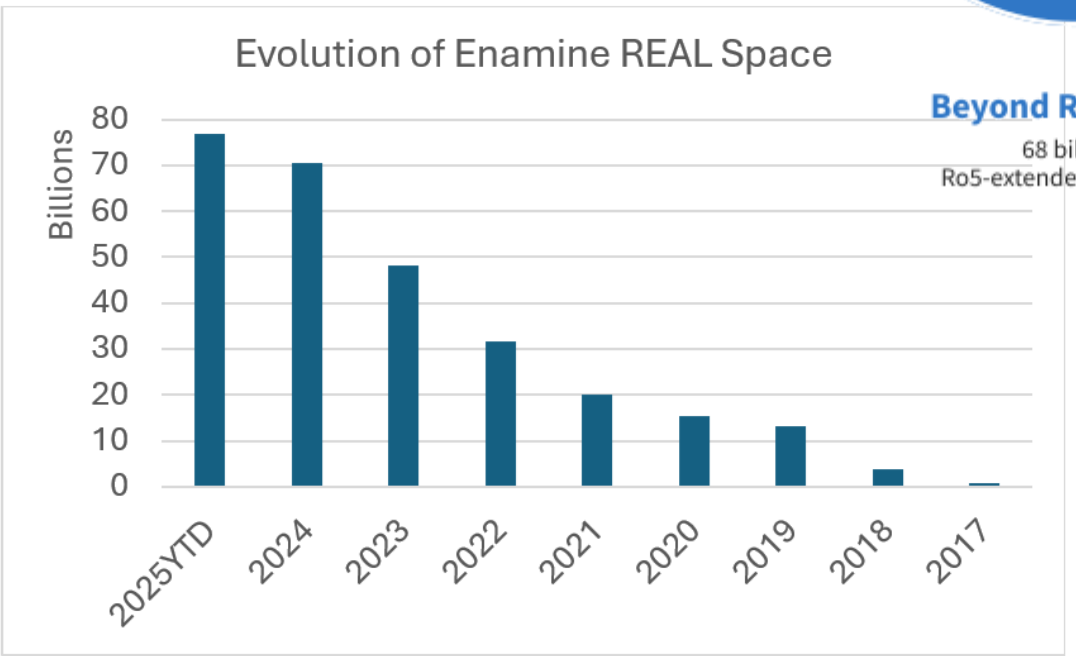
Pavel Polishchuk

Institute of Molecular and Translational Medicine
Faculty of Medicine and Dentistry
Palacky University
Czech Republic
pavlo.polishchuk@upol.cz
https://imtm.cz/laboratories/chemoinformatics-and-drug-design

Ultra-large combinatorial libraries







Hoffmann, T.; Gastreich, M., The next level in chemical space navigation: going far beyond enumerable compound libraries. *Drug Discovery Today* **2019**, 24, 1148-1156

CACHE challenges

Competition among chemoinformatics groups world-wide

Supposed benefits:

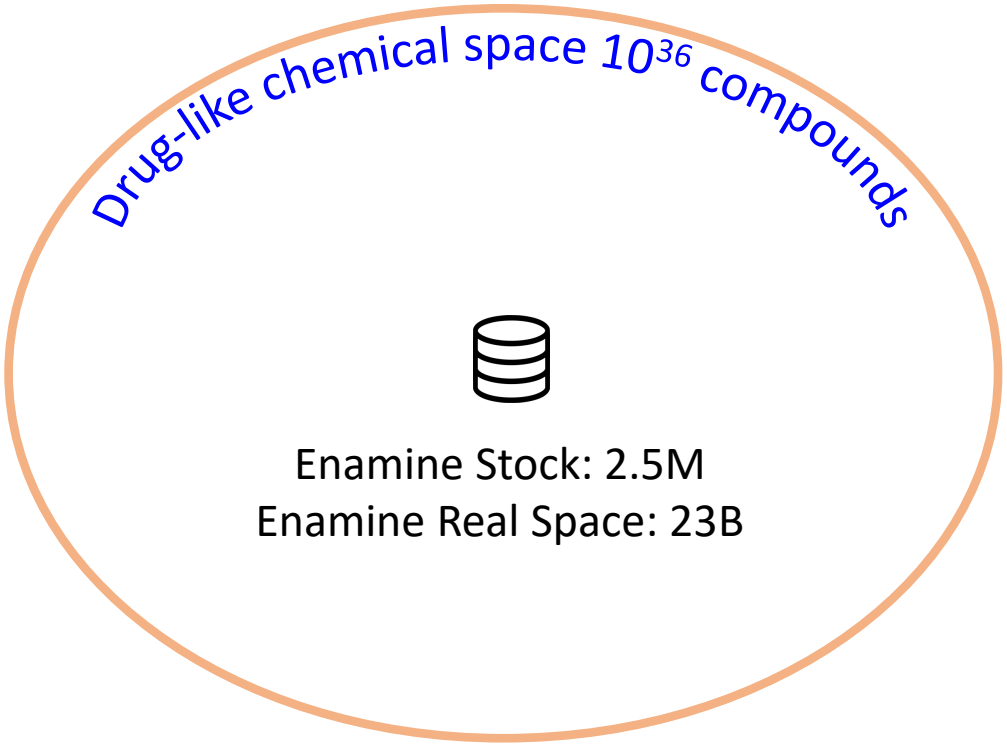
1. Encourage development and improvement of computational tools
2. Create a platform for prospective validation and comparison of different modeling tools and pipelines
3. Identify hit compounds for challenging or emerging targets/diseases
4. Contribute to open science to accelerate researches in a chosen direction

Ackloo, S. et al. *Nature Reviews Chemistry* **2022**, 6, 287-295.

CACHE challenge #1: LRRK2 and WDR domain

No X-ray of protein-ligand complexes:
- unknown binding site
- unknown conformation of a protein in a bound state

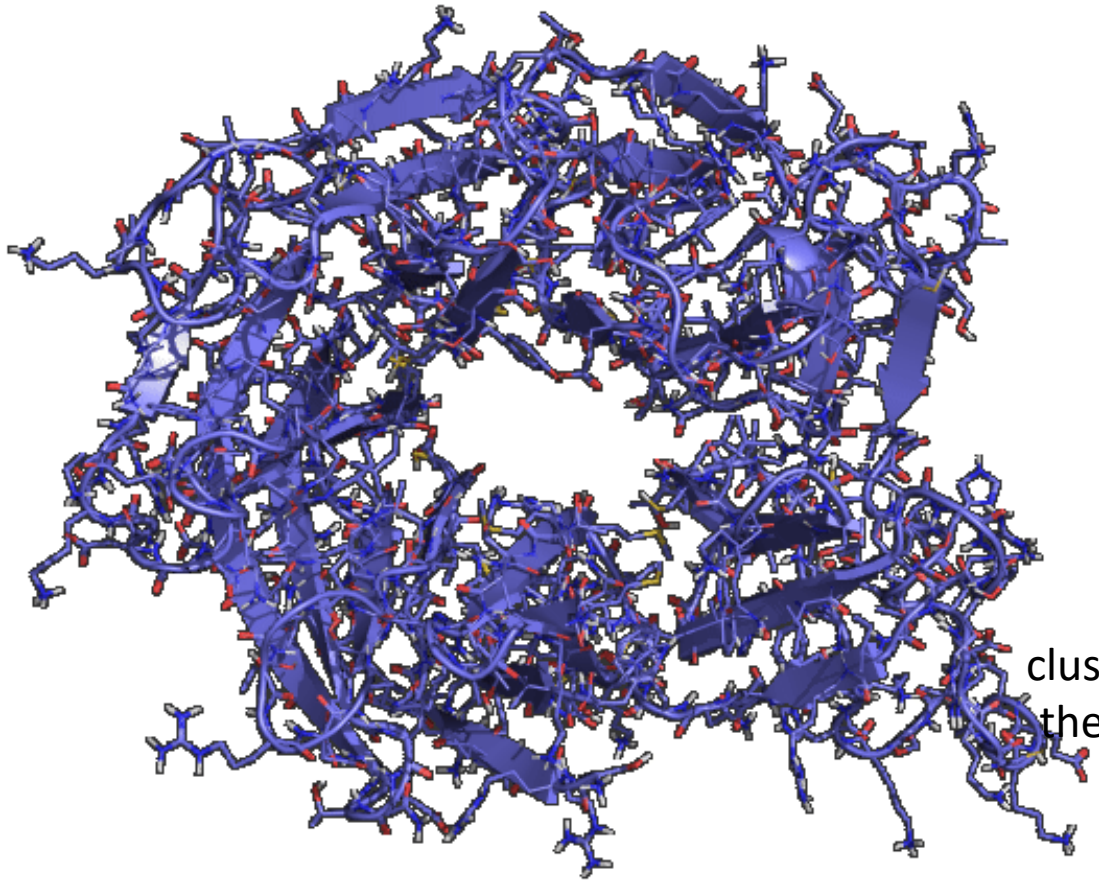
No known active molecules:
- large chemical space to explore



Ackloo, S. et al. *Nature Reviews Chemistry* **2022**, 6, 287-295.

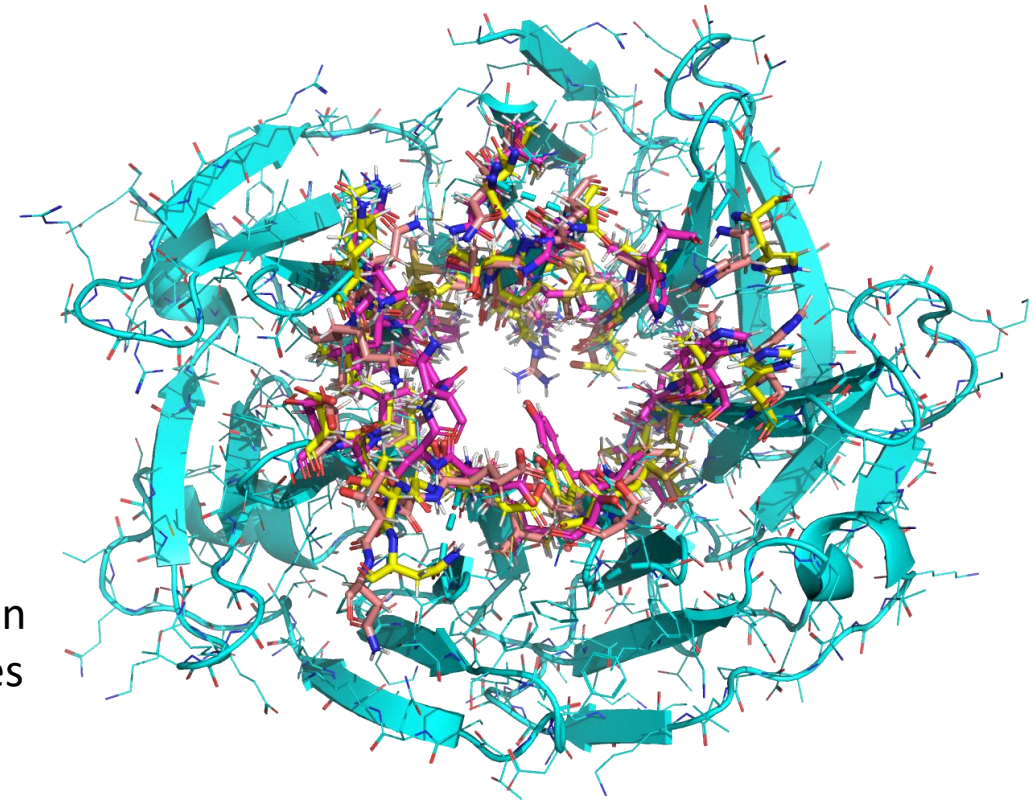
Protein structure challenge

WDR domain structure: 6DLO

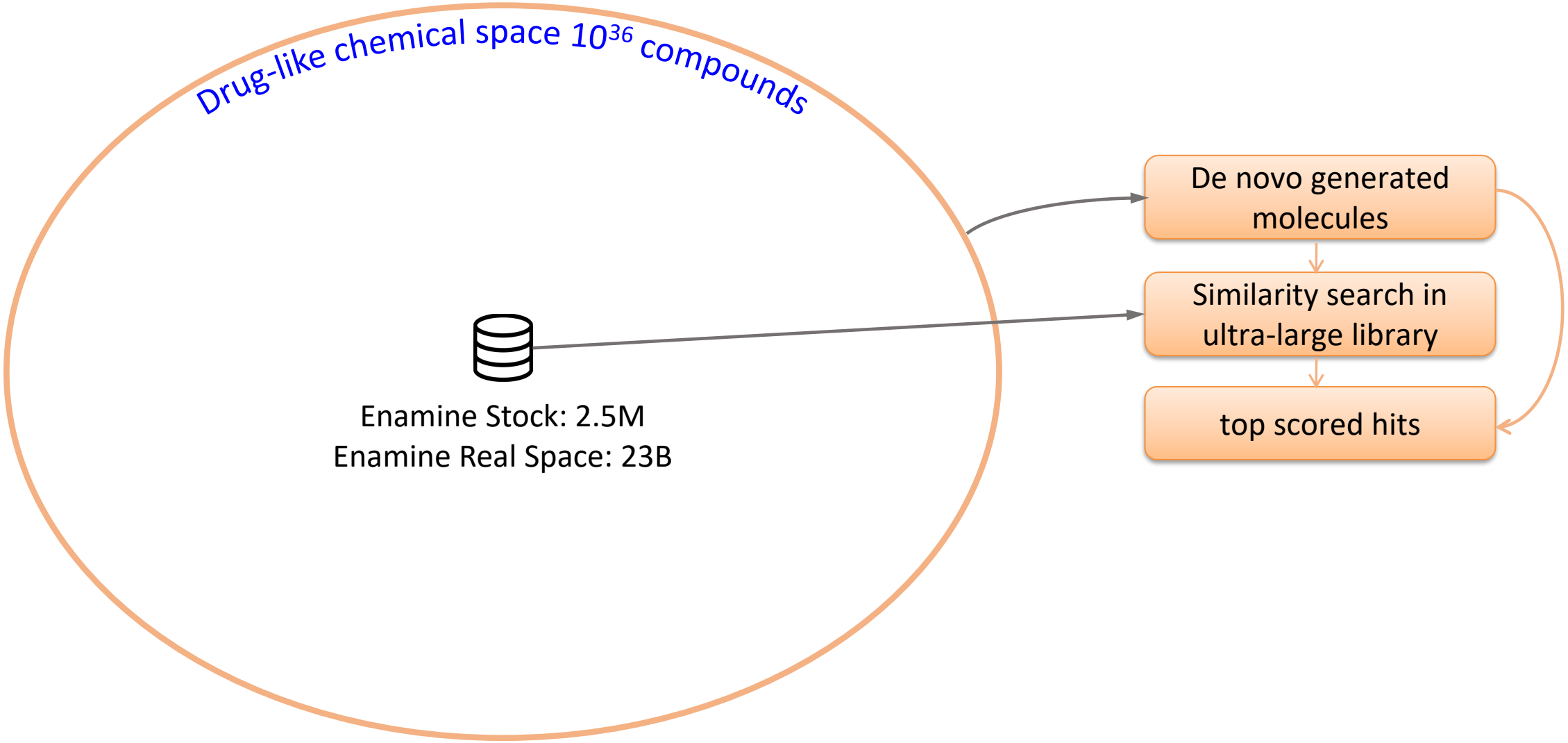


100 ns, 3 runs

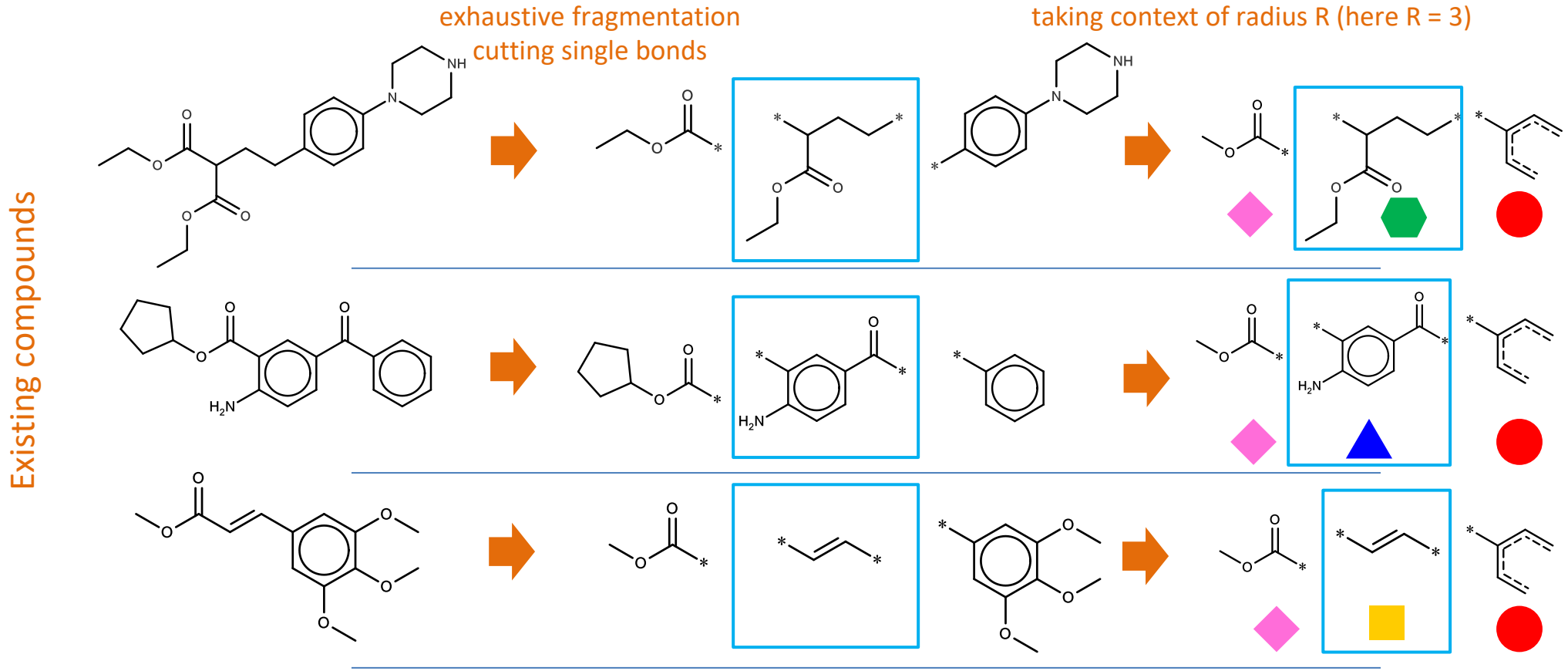
clustering based on the cavity residues



Chemical space exploration challenge



Chemically reasonable mutations (CReM)



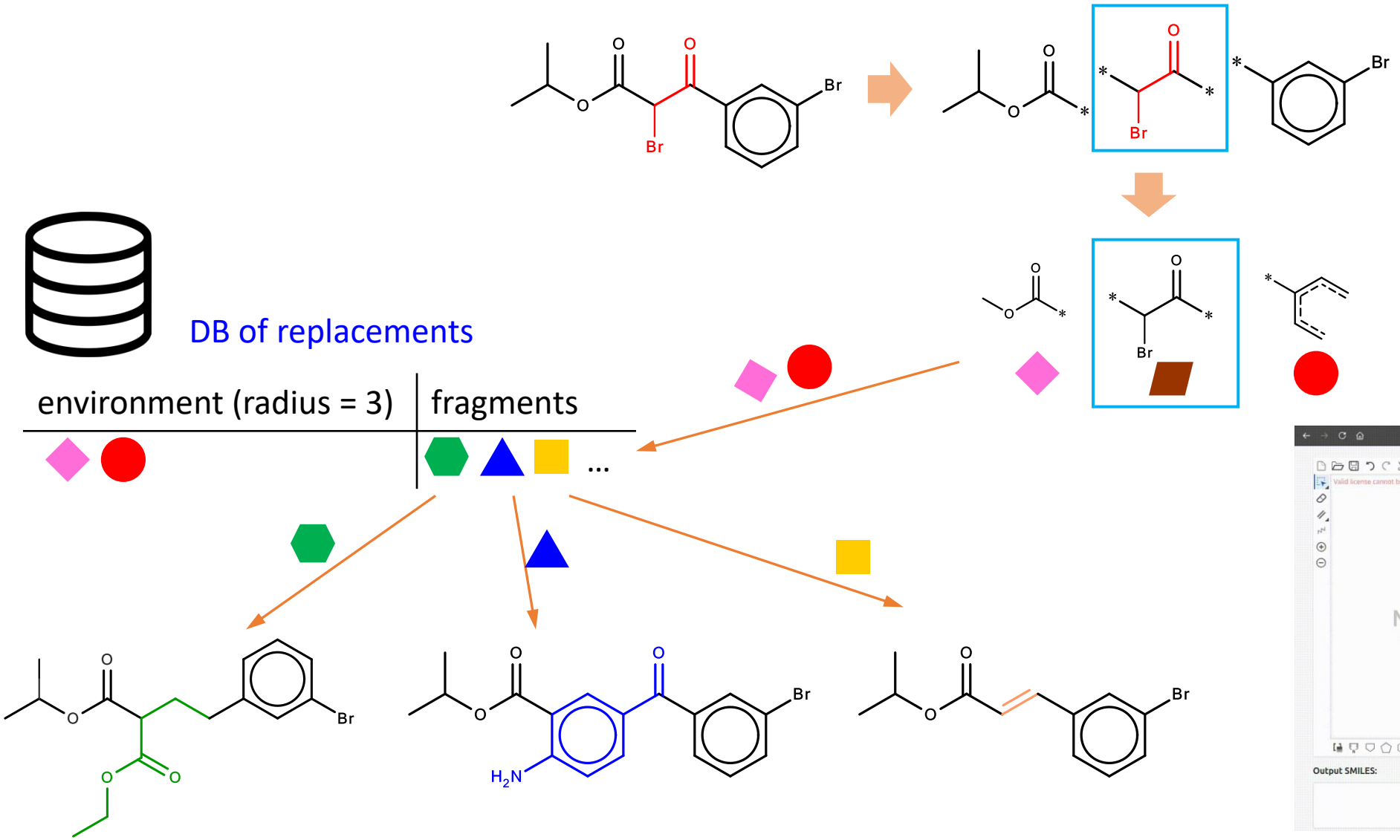
DB of replacements

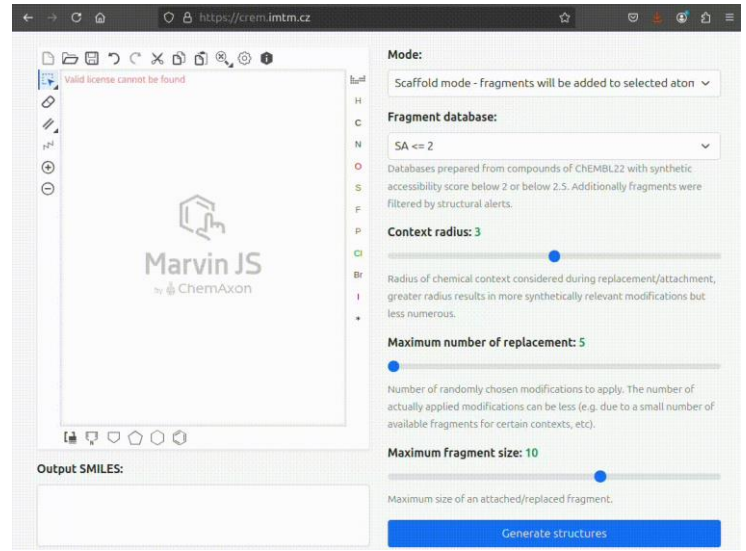
environment (radius = 3)	fragments
◆ ●	⬢ ▲ ■ ...
...	...

interchangeable fragments

Polishchuk, P., CReM: chemically reasonable mutations framework for structure generation. *J. Cheminf.* **2020**, 12 (1), 28.

Chemically reasonable mutations (CReM)



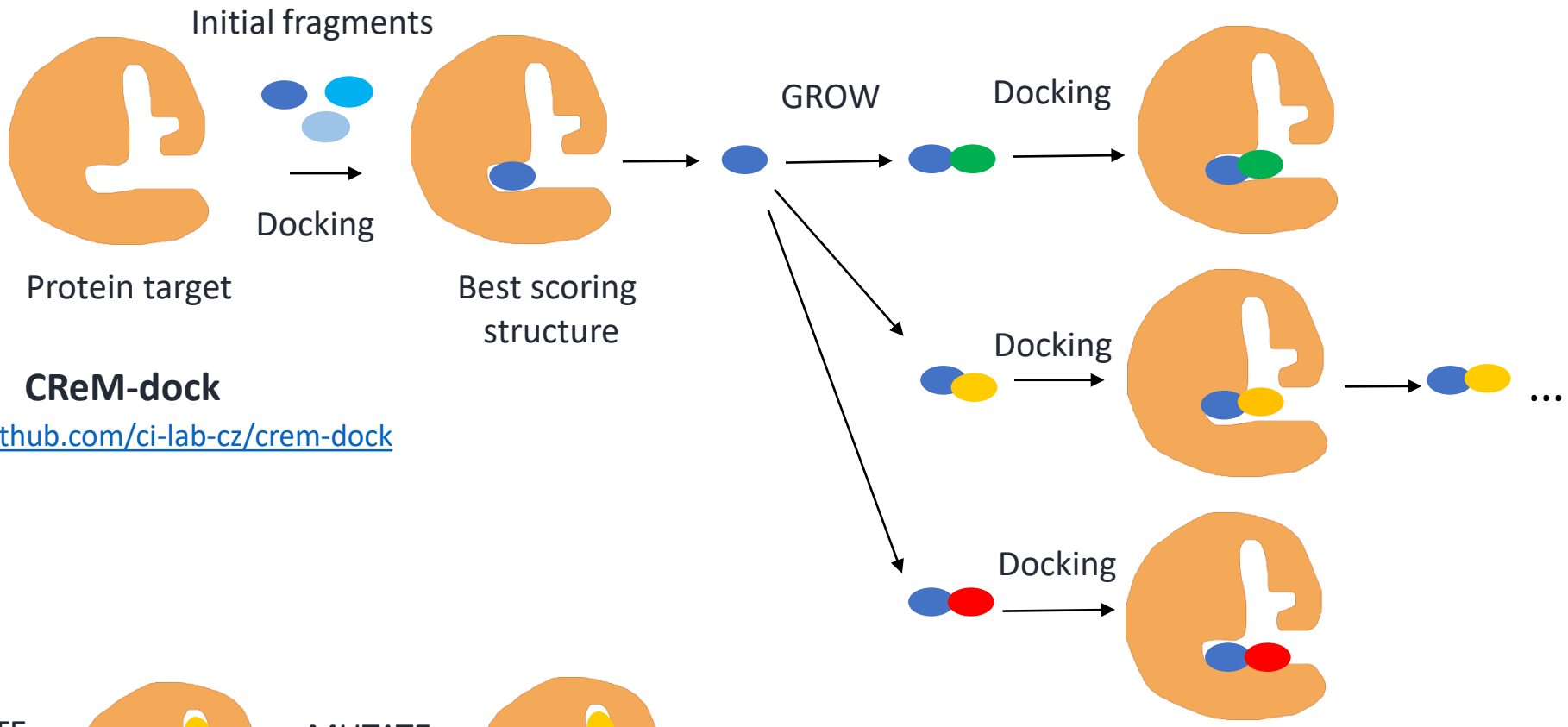


Generated structures are always chemically valid!

crem.imtm.cz

Polishchuk, P., CReM: chemically reasonable mutations framework for structure generation. *J. Cheminf.* **2020**, 12 (1), 28.

Strategy 1 (de novo design)



CReM-dock

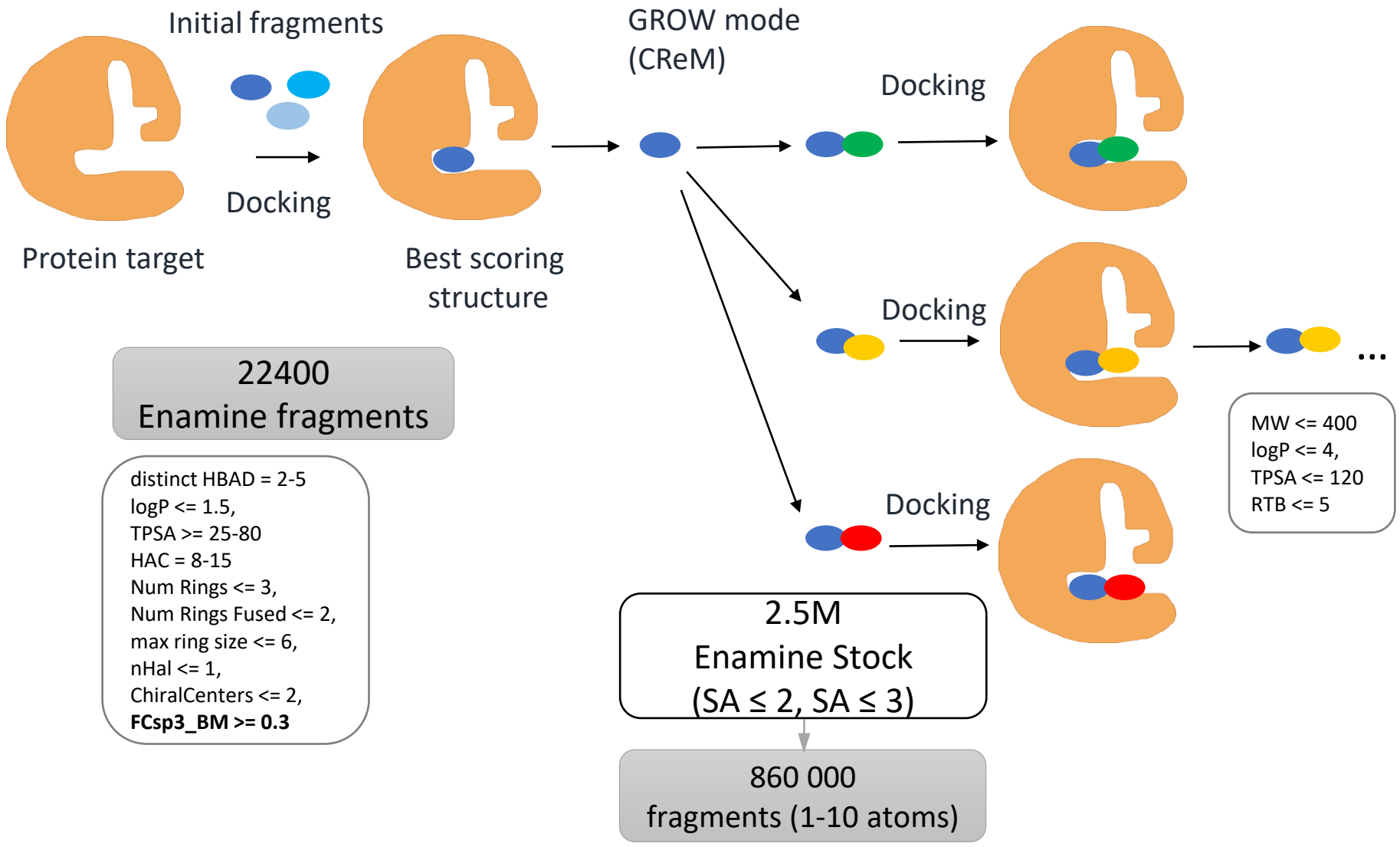
https://github.com/ci-lab-cz/crem-dock

CReM-dock-ga

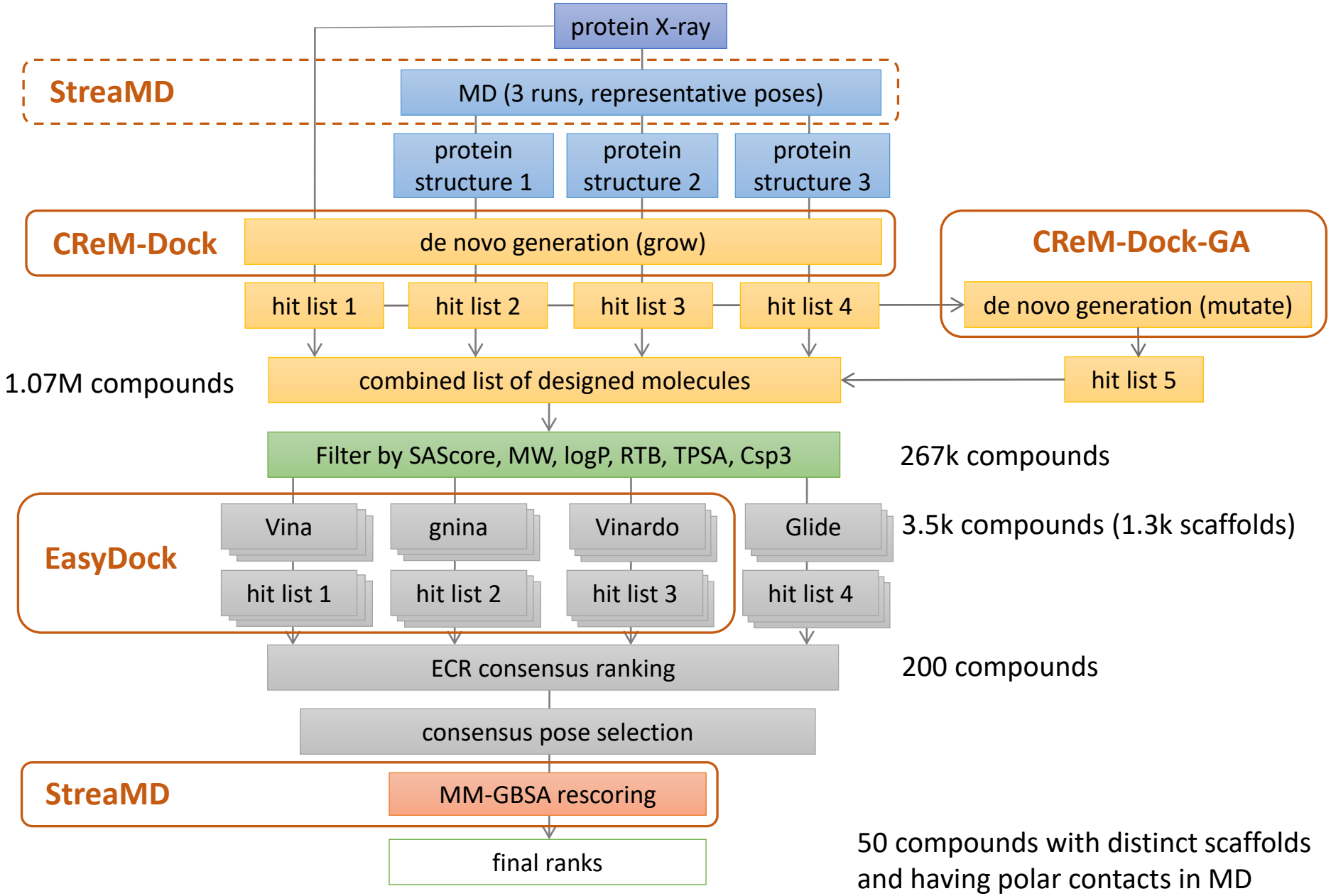
genetic algorithm (GA)

Minibaeva, G.; Polishchuk, P. CReM-dock: de novo design of synthetically feasible compounds guided by molecular docking. *ChemRxiv* ***2024***

Round 1: strategy 1 (de novo design be CReM-dock)



Strategy 1 (de novo design)



50 compounds with distinct scaffolds and having polar contacts in MD

Strategy 1 (de novo design)

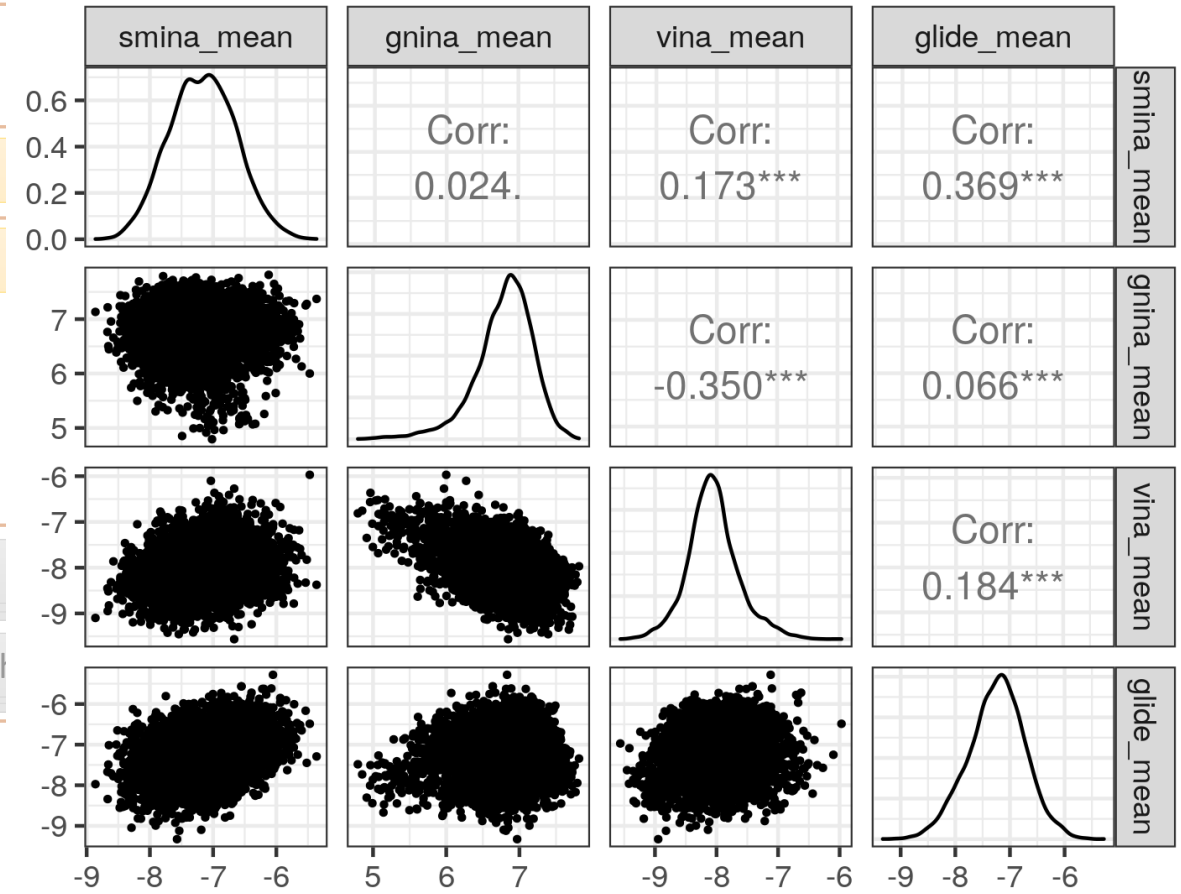
protein X-ray

StreaMD

CReM-Dock

1.07M compounds

EasyDock



M-Dock-GA

eneration (mutate)

nit list 5

s

s (1.3k scaffolds)

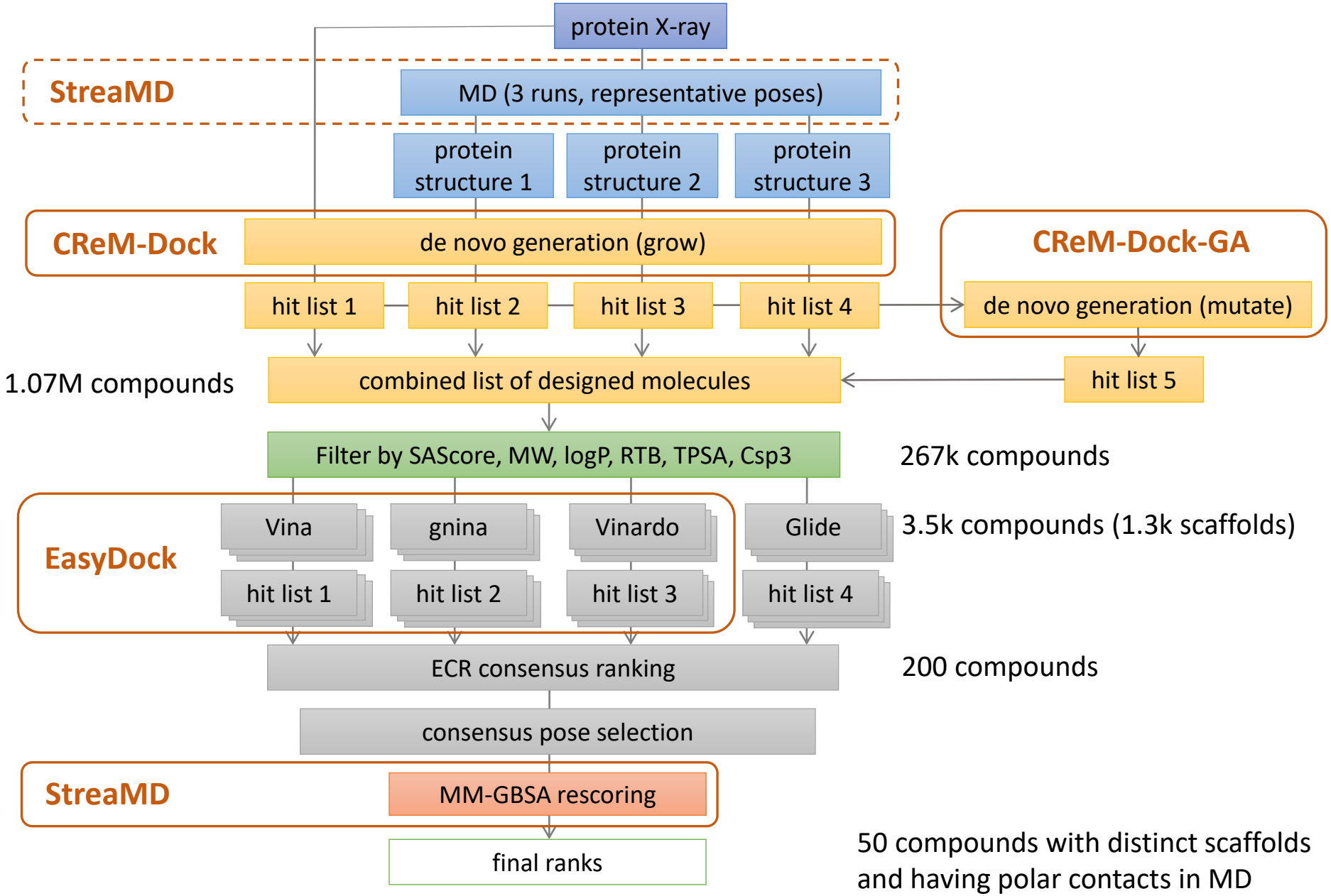
StreaMD

MM-GBSA rescoring

final ranks

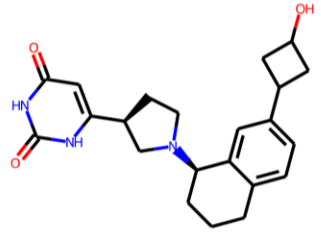
50 compounds with distinct scaffolds and having polar contacts in MD

Strategy 1 (de novo design)

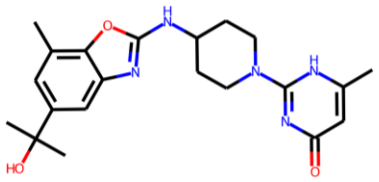


50 compounds with distinct scaffolds and having polar contacts in MD

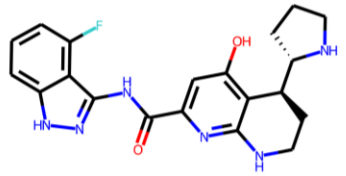
Strategy 1 (de novo design)



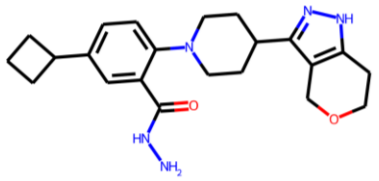
CREM1777121



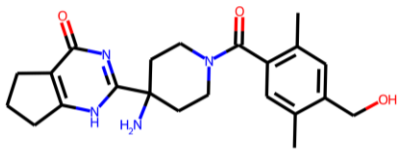
CREM0329741



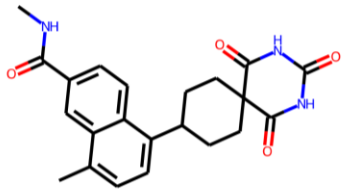
CREM1661038



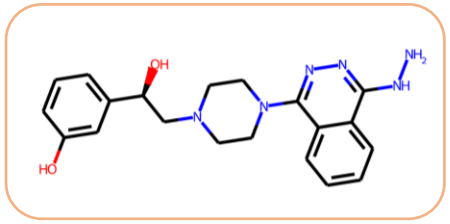
CREM1506273



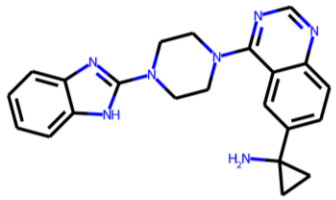
CREM0340409



CREM1089720

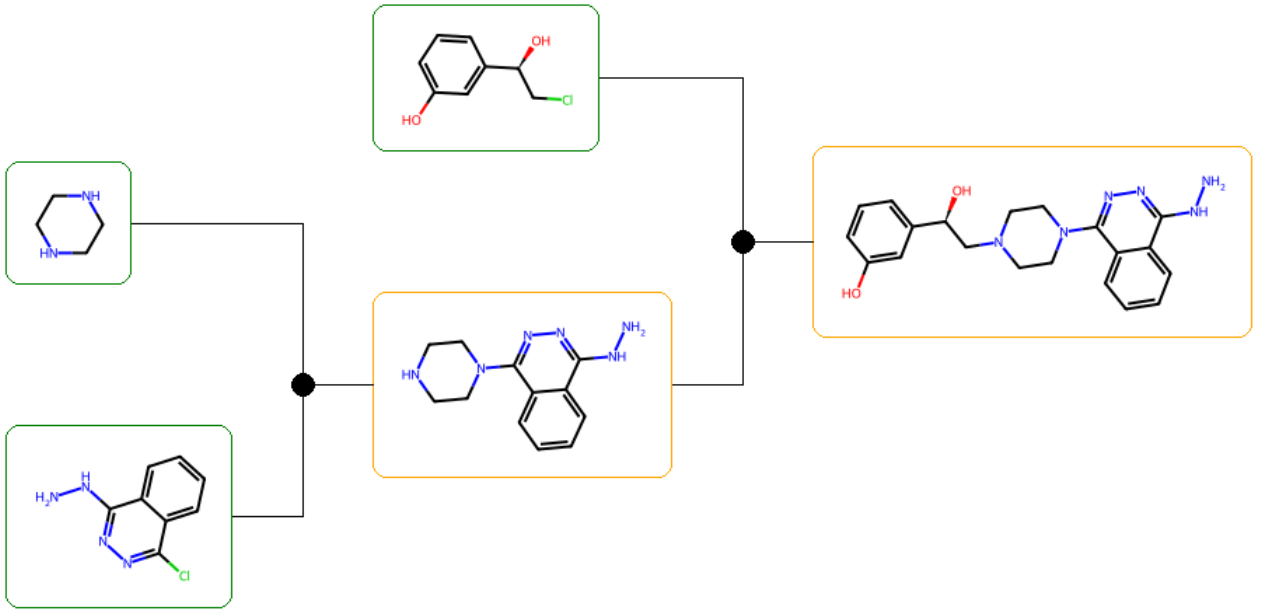


CREM1507777

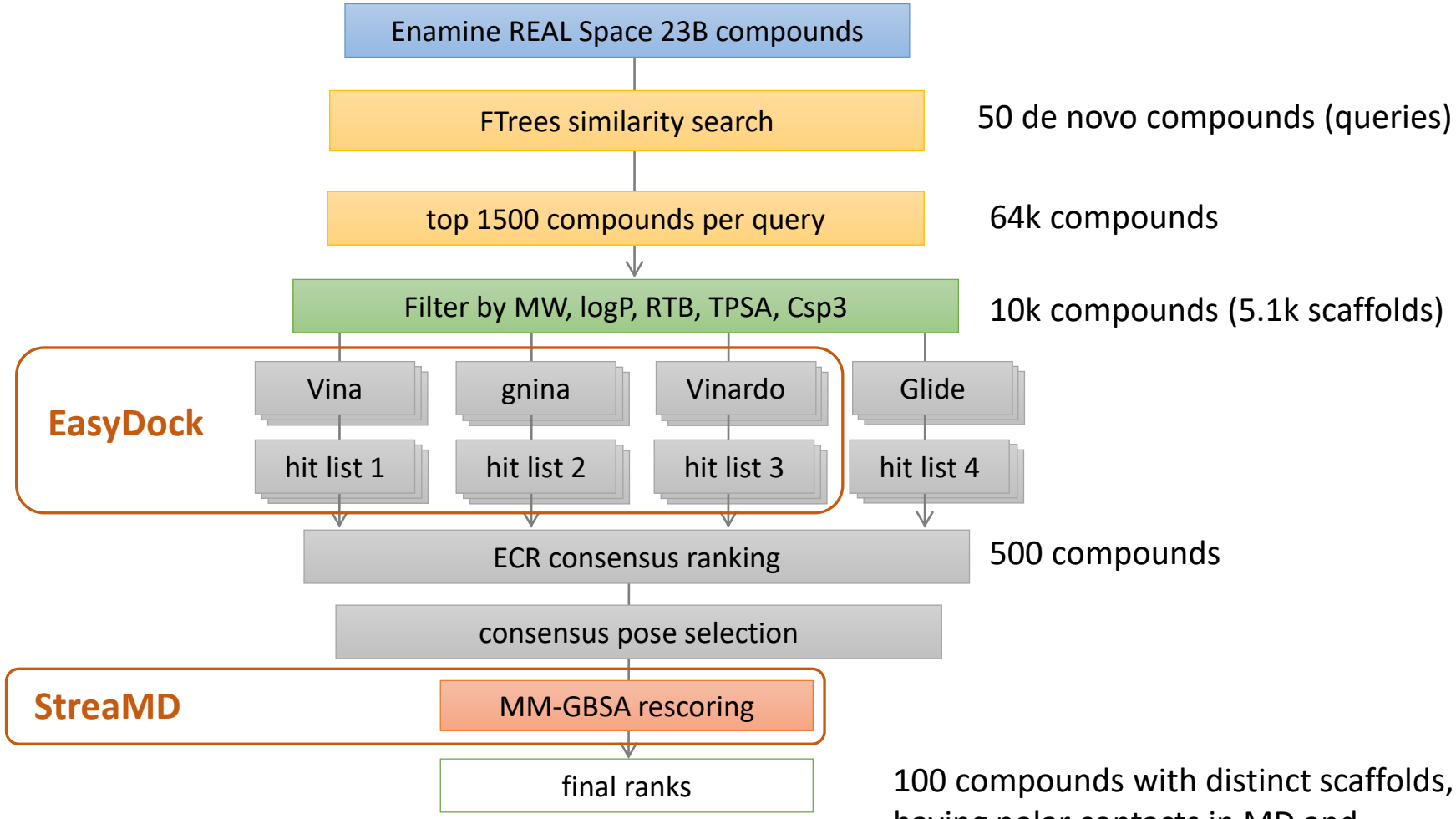


CREM1468894

- 50 de novo compounds
- SA score < 3
- 11 reconstructed retrosynthetic pathways with AiZynthFinder (2-5 steps)

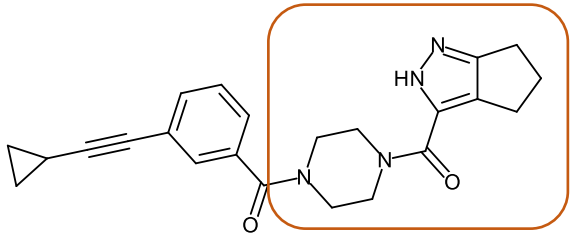


Strategy 2 (similarity search)

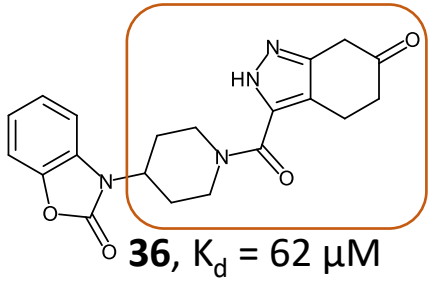


Rounds 1 & 2: results

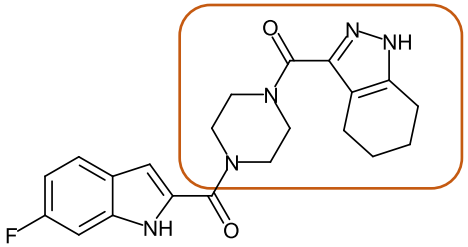
- 50 de novo + 100 similar compounds
- 91 compounds were selected (within the budget 9000$)
- 82 compounds were synthesized
- 8 compounds demonstrated activity (K_d = 25-117 µM by SPR)



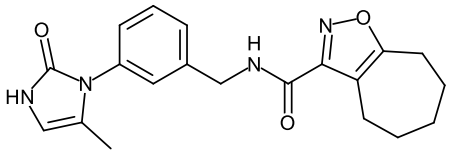
1, K_d = 61 µM



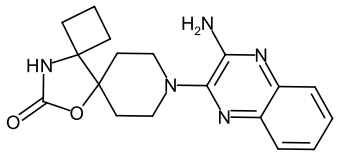
36, K_d = 62 µM



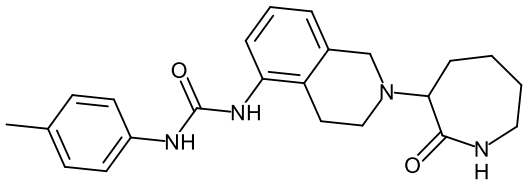
HO-15, K_d = 71 µM



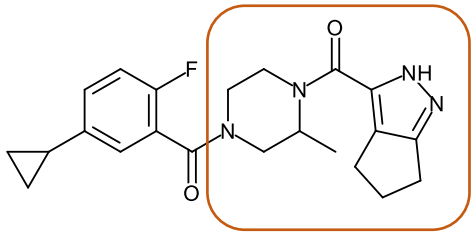
59, K_d = 32 µM



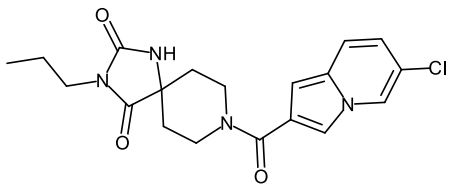
62, K_d = 25 µM



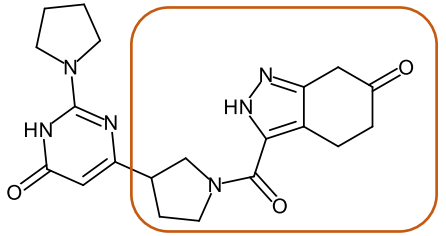
65, K_d = 56 µM



69, K_d = 117 µM



73, K_d = 31 µM



76, K_d = 74 µM

- 1.27 million docking events and 700 short MD simulations were made
- no human decision and compound selection across the whole pipeline

IMTM

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

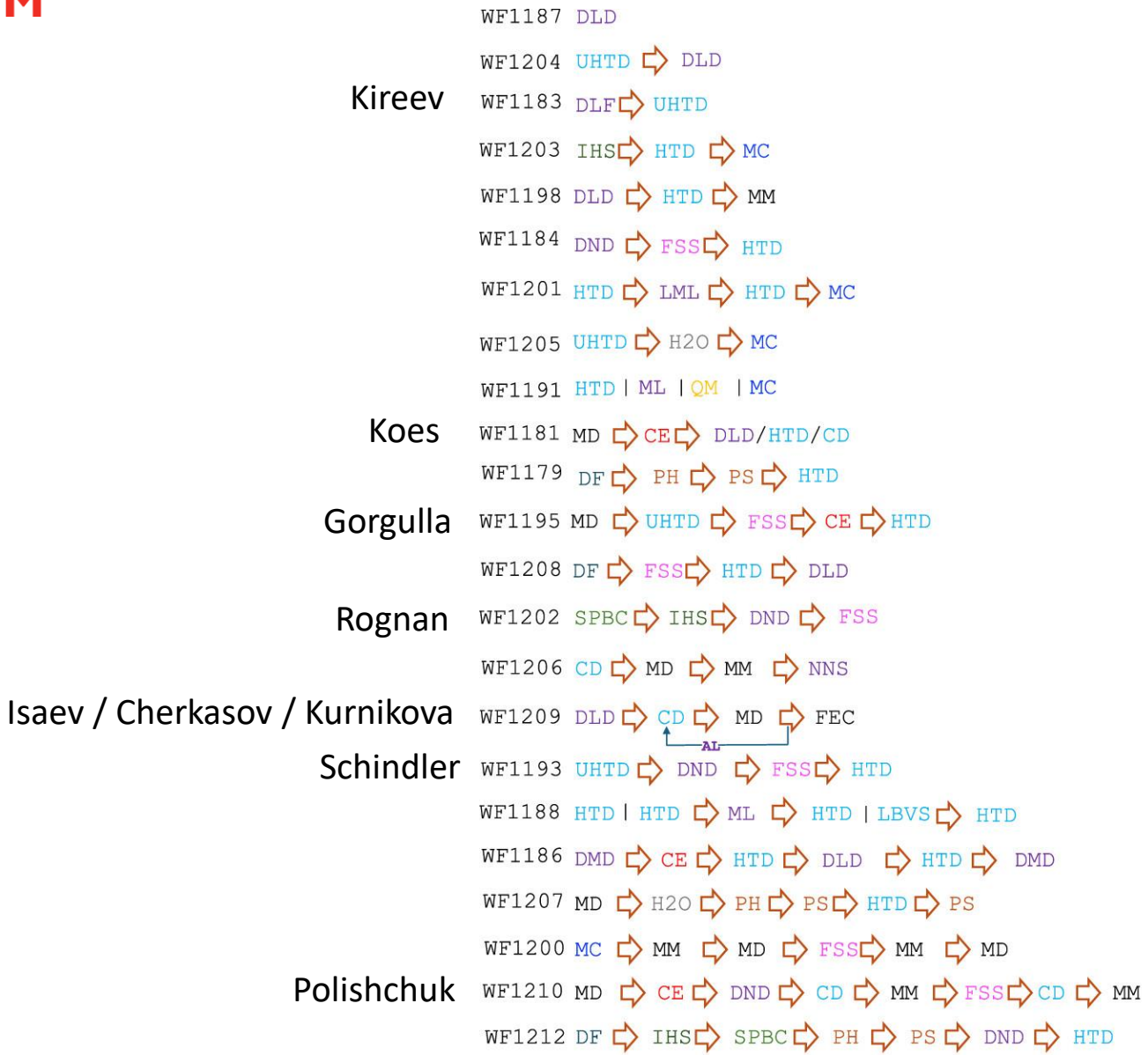
pubs.acs.org/jcim Article

CACHE Challenge #1: Targeting the WDR Domain of LRRK2, A Parkinson's Disease Associated Protein

Fengling Li, Suzanne Ackloo, Cheryl H. Arrowsmith, Fuqiang Ban, Christopher J. Barden, Hartmut Beck, Jan Beránek, Francois Berenger, Albina Bolotokova, Guillaume Bret, Marko Breznik, Emanuele Carosati, Irene Chau, Yu Chen, Artem Cherkasov, Dennis Della Corte, Katrin Denzinger, Aiping Dong, Sorin Draga, Ian Dunn, Kristina Edfeldt, Aled Edwards, Merveille Eguida, Paul Eisenhuth, Lukas Friedrich, Alexander Fuerll, Spencer S Gardiner, Francesco Gentile, Pegah Ghiabi, Elisa Gibson, Marta Glavatskikh, Christoph Gorgulla, Judith Guenther, Anders Gunnarsson, Filipp Gusev, Evgeny Gutkin, Levon Halabelian, Rachel J. Harding, Alexander Hillisch, Laurent Hoffer, Anders Hogner, Scott Houliston, John J Irwin, Olexandr Isayev, Aleksandra Ivanova, Celien Jacquemard, Austin J Jarrett, Jan H. Jensen, Dmitri Kireev, Julian Kleber, S. Benjamin Koby, David Koes, Ashutosh Kumar, Maria G. Kurnikova, Alina Kutlushina, Uta Lessel, Fabian Liessmann, Sijie Liu, Wei Lu, Jens Meiler, Akhila Mettu, Guzel Minibaeva, Rocco Moretti, Connor J Morris, Chamali Narangoda, Theresa Noonan, Leon Obendorf, Szymon Pach, Amit Pandit, Sumera Perveen, Gennady Poda, Pavel Polishchuk, Kristina Puls, Vera Pütter, Didier Rognan, Dylan Roskams-Edris, Christina Schindler, François Sindt, Vojtěch Spiwok, Casper Steinmann, Rick L. Stevens, Valerij Talagayev, Damon Tingey, Oanh Vu, W. Patrick Walters, Xiaowen Wang, Zhenyu Wang, Gerhard Wolber, Clemens Alexander Wolf, Lars Wortmann, Hong Zeng, Carlos A. Zepeda, Kam Y. J. Zhang, Jixian Zhang, Shuangjia Zheng, and Matthieu Schapira*

Cite This: J. Chem. Inf. Model. 2024, 64, 8521−8536 Read Online

Pipelines of all participants

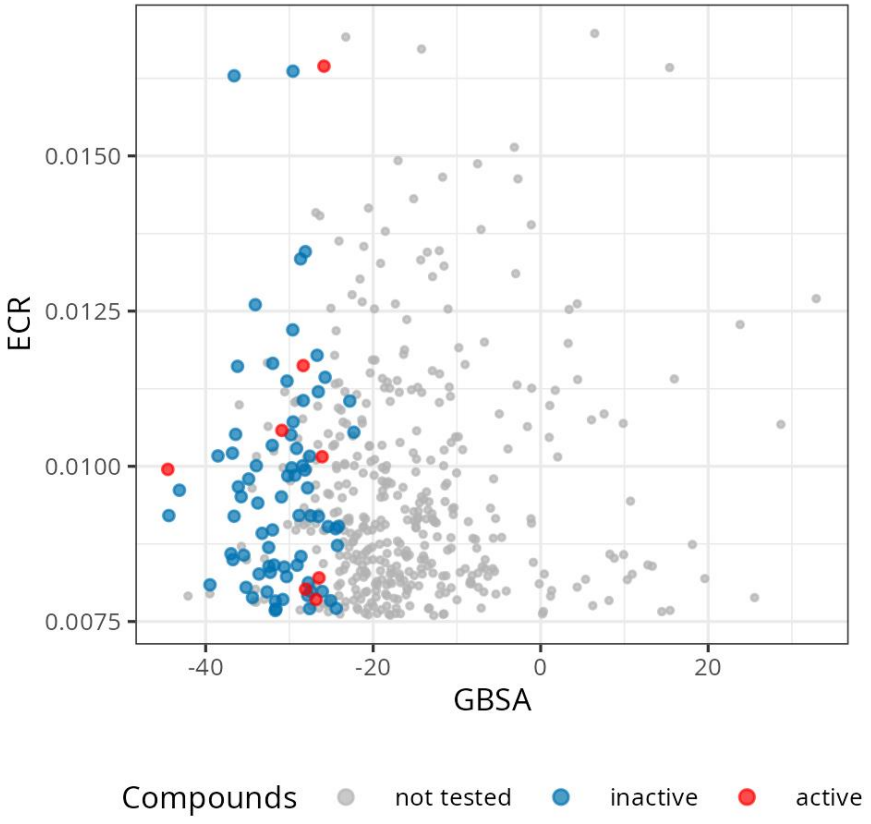


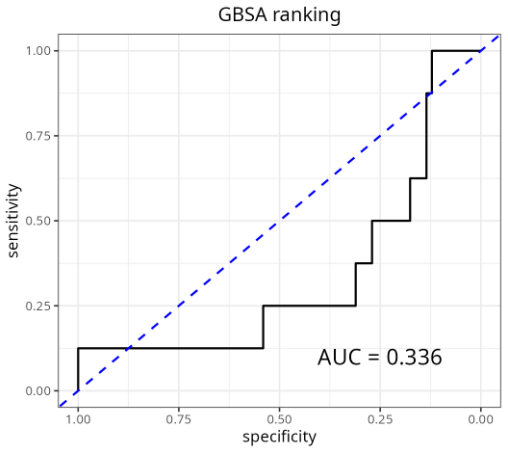
MC : medicinal chemist
CE: conformational ensemble
H2O: map stable water molec.
IHS: interaction hot spots
SPBC: similar pocket in PDB with bound compound
FSS: fingerprint similarity search
PS: pharmacophore search
PH: pharmacophore hypothesis
DND: de novo design
DLD: deep learning docking
DMD: deep molecular dynamics
NNS: NN scoring
DF: dock fragments
HTD : high-throughput docking
UHTD: ultra HTD
CD: consensus docking
MM: molecular mechanics
MD: molecular dynamics
FEC: free energy calculation

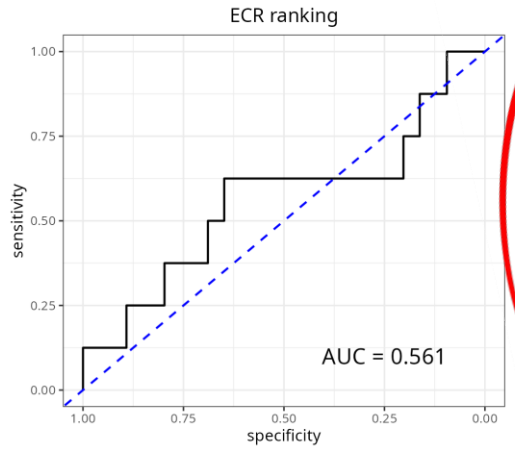
Li, F. et al. CACHE Challenge #1: Targeting the WDR Domain of LRRK2, A Parkinson's Disease Associated Protein. *J. Chem. Inf. Model.* **2024**, 64 (22), 8521-8536.

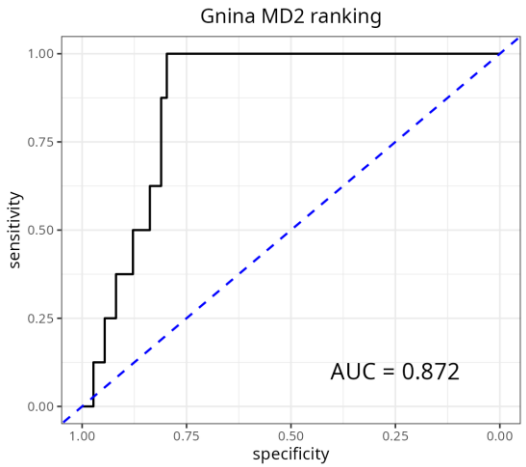
Overall pipeline of hit finding guided by de novo design

protein target → protein MD → de novo design → physicochemical filters → consensus docking → MM/GBSA → **De novo hits** → similarity search in REAL Space → physicochemical filters → consensus docking → MM/GBSA → **Similarity hits**





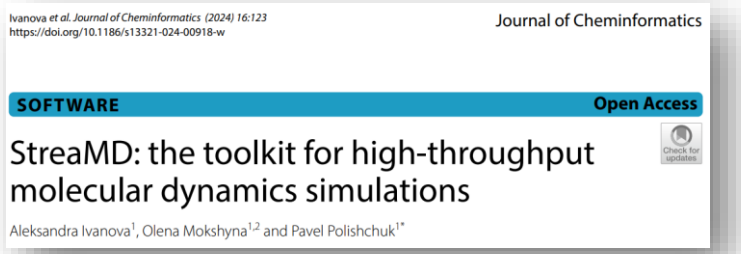




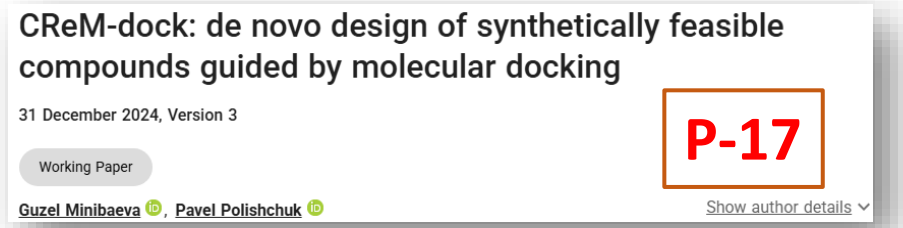
Summary

- Similarity search based on de novo designed templates looks promising as it allows to save resources. However, further studies are required.
- CACHE challenge was great and motivated us to develop several open-source tools: CReM-dock, Easydock and StreaMD.

https://github.com/ci-lab-cz/easydock



https://github.com/ci-lab-cz/streamd



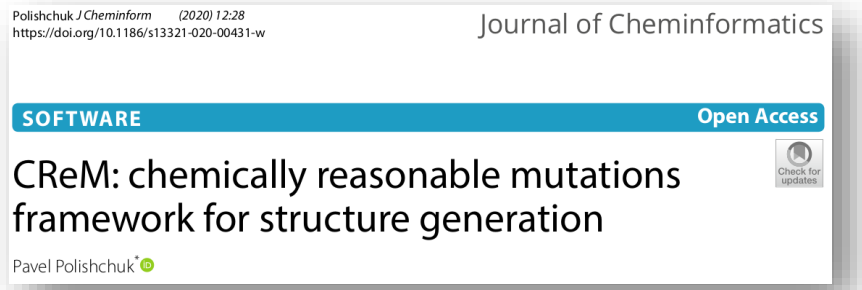
https://github.com/ci-lab-cz/crem-dock

Guzel Minibaeva

Aleksandra Ivanova

Alina Kutlushina

UCT:
Dr. Vojtech Spiwok
Jan Beranek



https://github.com/DrrDom/crem

The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through INTER_EXCELLENCE II grant LUAUS23262, the e-INFRA CZ (ID:90254), projects ELIXIR-CZ (LM2023055) and CZ-OPENSCREEN (LM2023052)