SEARCHING FOR HIT MOLECULES IN ULTRA-LARGE CHEMICAL LIBRARIES GUIDED BY DE NOVO DESIGN

Pavel Polishchuk[1], Guzel Minibaeva[1], Aleksandra Ivanova[1], Alina Kutlushina[1], Jan Beranek[2], Vojtech Spiwok[2]

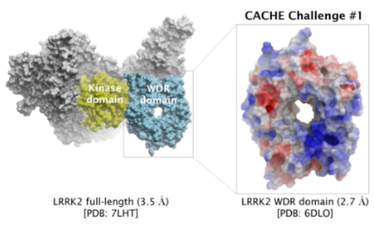
[1] Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University and University Hospital in Olomouc, Hnevotinska 5, 77900 Olomouc, Czech Republic

[2] Department of Biochemistry and Microbiology, University of Chemistry and Technology, Technická 5, 16628 Prague, Czech Republic

pavlo.polishchuk@upol.cz

CACHE CHALLENGE — CRITICAL ASSESSMENT OF COMPUTATIONAL HIT-FINDING EXPERIMENTS

CHALLENGE #1



PREDICT HITS FOR THE WDR DOMAIN OF LRRK2

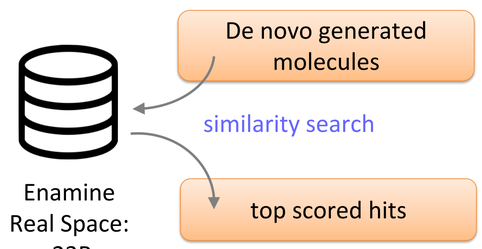
The first CACHE Challenge target is LRRK2, the most commonly mutated gene in familial Parkinson's Disease.

Participants are asked to find hits for the WD40 repeat (WDR) domain of LRRK2. Read more under Details below.

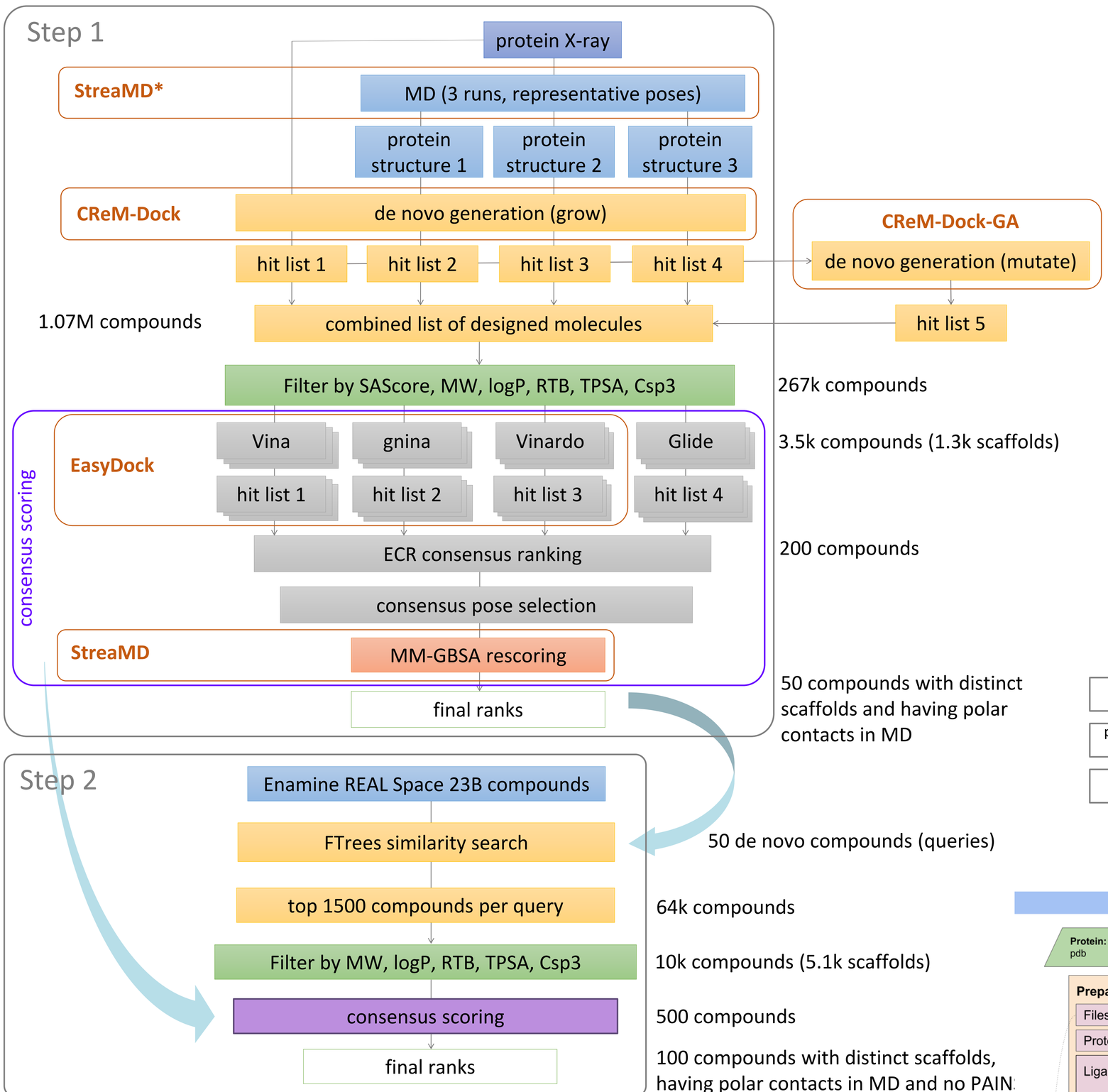
PARTICULAR CHALLANGES

1. X-ray structure of a protein in the apo-form was only available
2. No known binders
3. The search chemical space should be accessible to Enamine (stock, ~3M compounds, or REAL space, ~23B compounds)

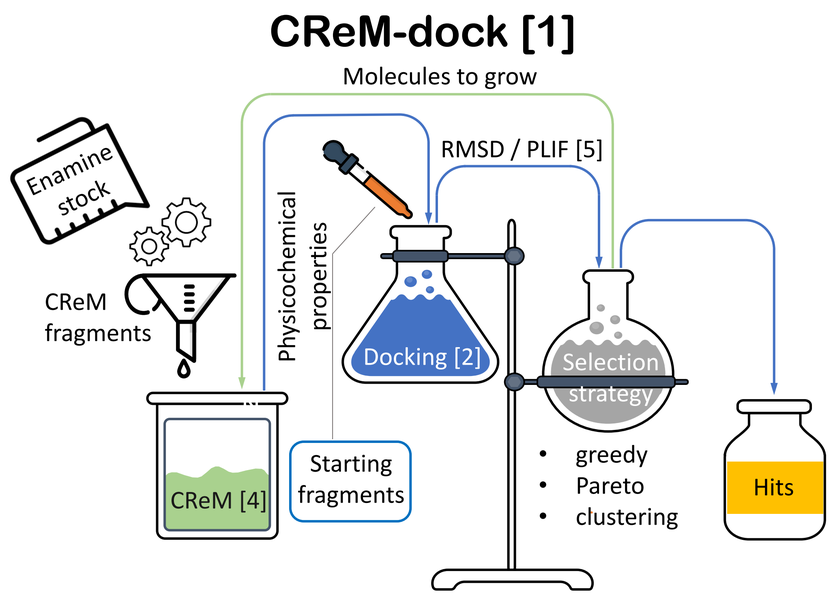
SOLUTION CONCEPT



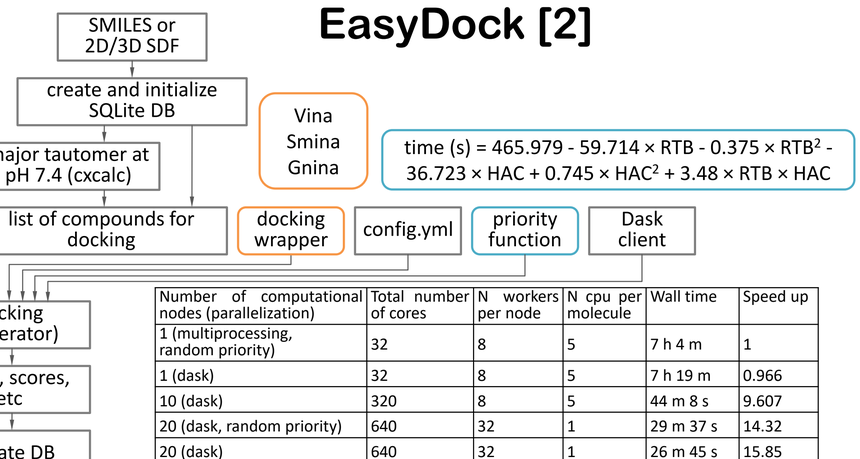
SOLUTION PIPELINE



CReM-dock [1]

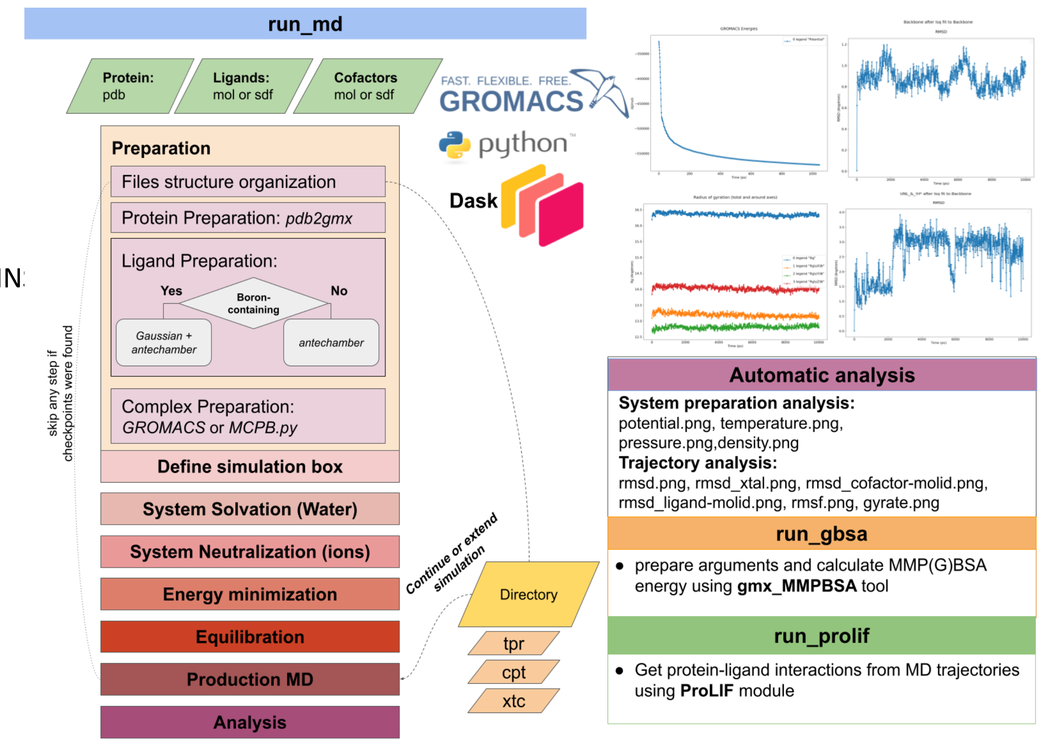


EasyDock [2]



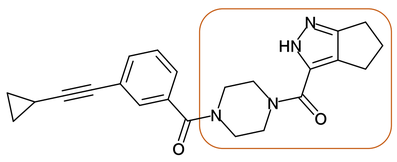
Number of computational nodes (parallelization)	Total number of cores	N workers per node	N cpu per molecule	Wall time	Speed up
1 (multiprocessing, random priority)	32	8	5	7 h 4 m	1
1 (dask)	32	8	5	7 h 19 m	0.966
10 (dask)	320	8	5	44 m 8 s	9.607
20 (dask, random priority)	640	32	1	29 m 37 s	14.32
20 (dask)	640	32	1	26 m 45 s	15.85
20 (dask)	640	8	5	22 m 14 s	19.07
20 (dask, random priority)	640	8	5	22 m 35 s	18.77

StreaMD [3]

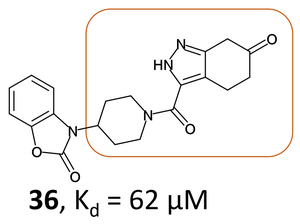


EXPERIMENTAL EVALUATION OF HITS

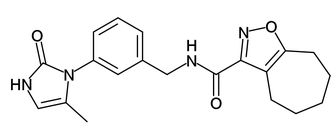
- 50 de novo + 100 similar compounds
- 91 compounds were selected (within the budget 9000$)
- 82 compounds were synthesized
- 8 compounds demonstrated activity (K_d = 25-117 μM by SPR)



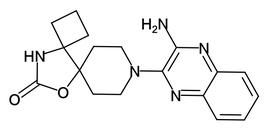
1, K_d = 61 μM



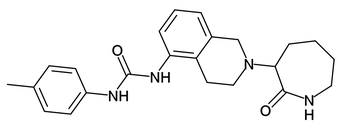
36, K_d = 62 μM



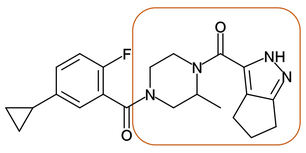
59, K_d = 32 μM



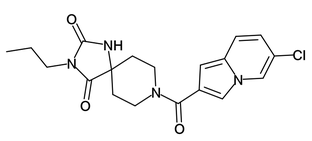
62, K_d = 25 μM



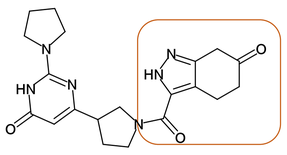
65, K_d = 56 μM



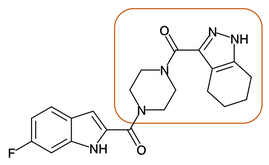
69, K_d = 117 μM



73, K_d = 31 μM



76, K_d = 74 μM



HO-15, K_d = 71 μM

CONCLUSIONS

- Fully automated pipeline (no human intervention on any step)
- 1.27 million docking events and 700 short MD simulations were made
- 8 compounds were active among 82 tested ones (hit rate ~10%) from 23B REAL Space
- Several tools were developed and validated (EasyDock, StreaMD, CReM-dock, CReM-dock-ga)

[1] Minibaeva, G.; Polishchuk, P. CReM-dock: de novo design of synthetically feasible compounds guided by molecular docking. *ChemRxiv* **2024** - https://doi.org/10.26434/chemrxiv-2024-fpzqb-v2

[2] Minibaeva, G.; Ivanova, A.; Polishchuk, P. EasyDock: customizable and scalable docking tool. *J. Cheminf.* **2023**, 15 (1), 102. - https://github.com/ci-lab-cz/easydock

[3] Ivanova A., Mokshyna O., Polishchuk P. StreaMD: the toolkit for high-throughput molecular dynamics simulations. *ChemRxiv* **2024** - https://doi.org/10.26434/chemrxiv-2024-2rjqz, https://github.com/ci-lab-cz/streamd

[4] Polishchuk, P. CReM: chemically reasonable mutations framework for structure generation. *J. Cheminf.* **2020**, 12 (1), 28 - https://github.com/DrrDom/crem, https://crem.imtm.cz (scaffold decoration & analogs enumeration)

[5] Bouysset, C.; Fiorucci, S. ProLIF: a library to encode molecular interactions as fingerprints. J. Cheminf. 2021, 13 (1), 72.- https://github.com/chemosim-lab/ProLIF

The work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through INTER-EXCELLENCE II LUAUS23262, the e-INFRA CZ (ID:90140), ELIXIR-CZ (LM2018131, LM2023055), CZ-OPENSCREEN (LM2018130, LM2023052) grants and by European and Regional Fund project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868).