8th Advanced In Silico Drug Design workshop

27 - 31 January 2025
Olomouc, Czech Republic

Univerzita Palackého v Olomouci

CACHE#1: searching for hit molecules in ultra-large chemical databases guided by de novo design

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

IMTM

CACHE challenge

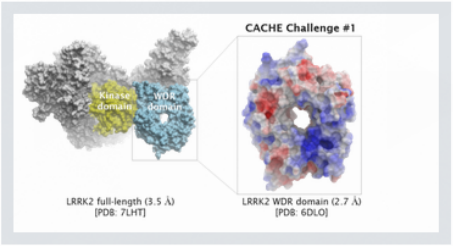
Competition among top chemoinformatics groups world-wide

Benefits supposed by organizers:

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

The first CACHE challenge

COMPETITION #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.

Why the WDR domain?	PD-associated LRRK2 mutations tend to promote LRRK2 filament formation and enhance LRRK2 interaction with microtubules. Recent structural data reveals that only compounds stabilizing the open form of LRRK2 antagonize the pathogenic formation of LRRK2 filaments in cells, but most kinase inhibitors stabilize the closed form of LRRK2. An alternative and so far overlooked strategy is to pharmacologically target the WDR domain of LRRK2, which is juxtaposed to the kinase domain. The WDR domain in LRRK2 may be important for recruiting LRRK2 signalling partners or for binding to tubulin. WDR domains are disease-associated and druggable. Identifying chemical starting points binding to the WDR domain of LRRK2 is a novel approach to target this protein.
Potential impact	The public release of chemical starting points for an understudied domain of LRRK2 will offer opportunities to target LRRK2 via an allosteric mechanism and make PROTACs to induce its degradation with ligands not directly interfering with the catalytic activity of the target.

https://cache-challenge.org/

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

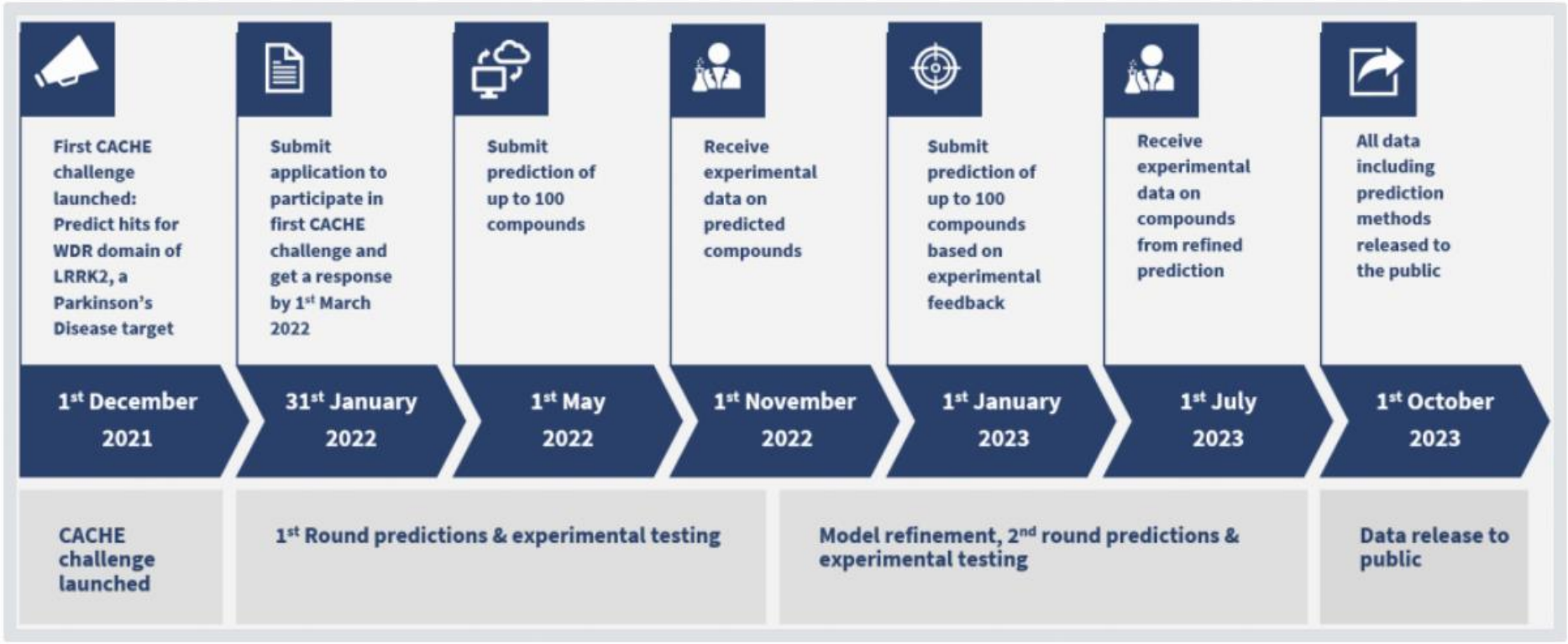
The challenge pipeline

Application opens
2021-12-01

Application closes
2022-01-31

Application form
Download

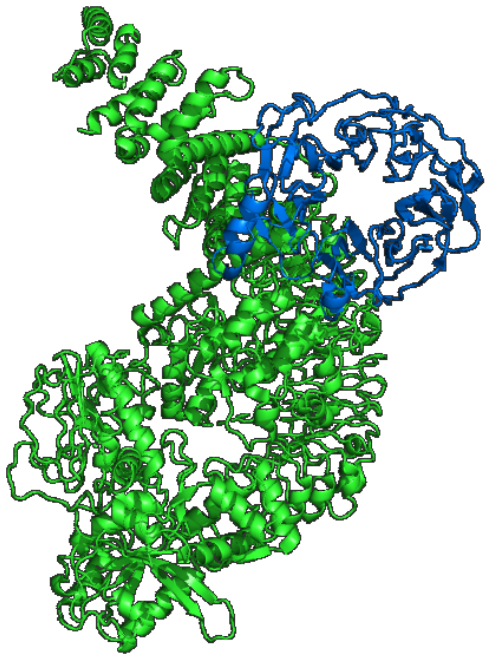
TIMELINE

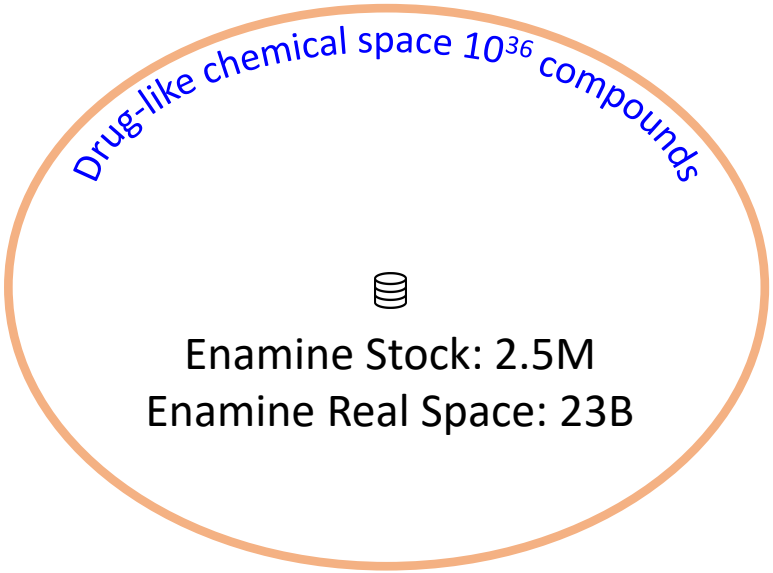


LRRK2 and WDR domain

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

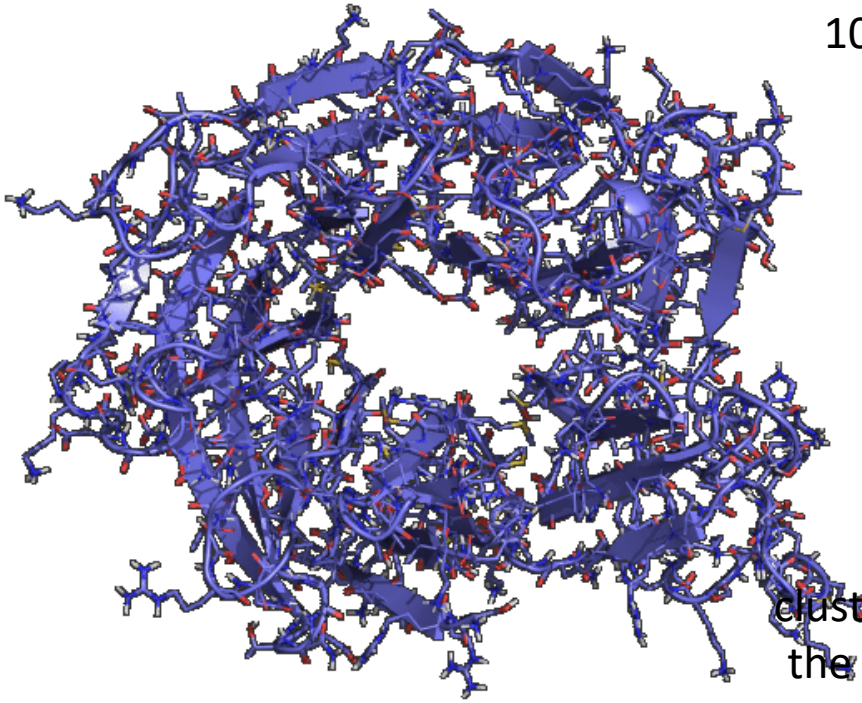
No known active molecules:
- large chemical space to explore





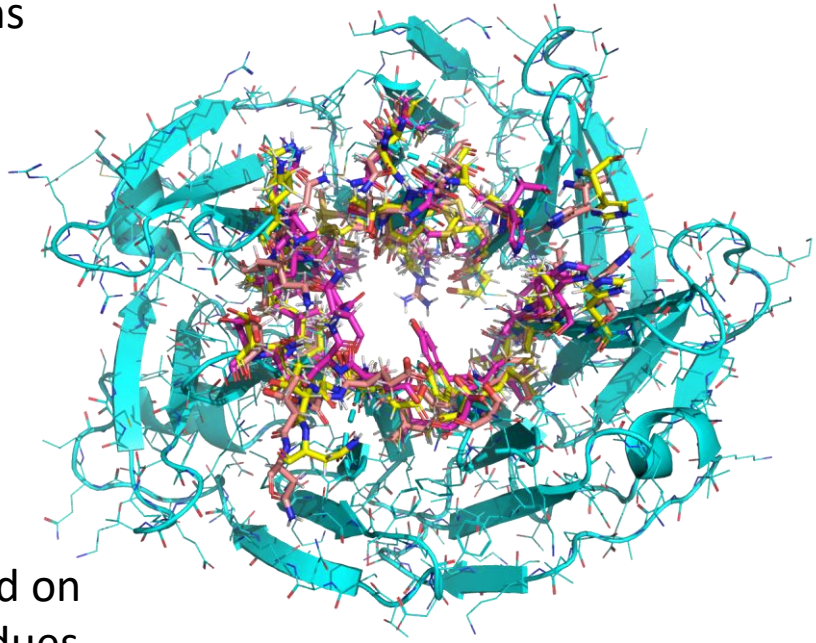
Round 1: protein structure challenge

WDR domain structure: 6DLO

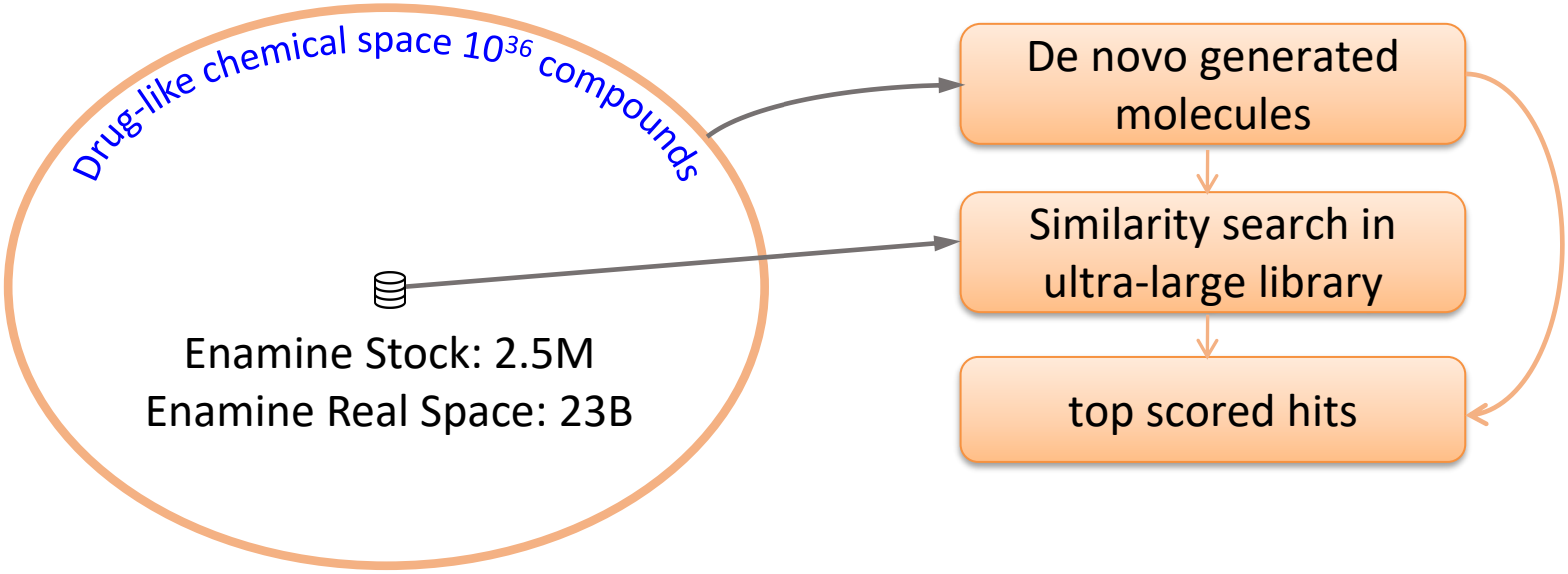


100 ns, 3 runs

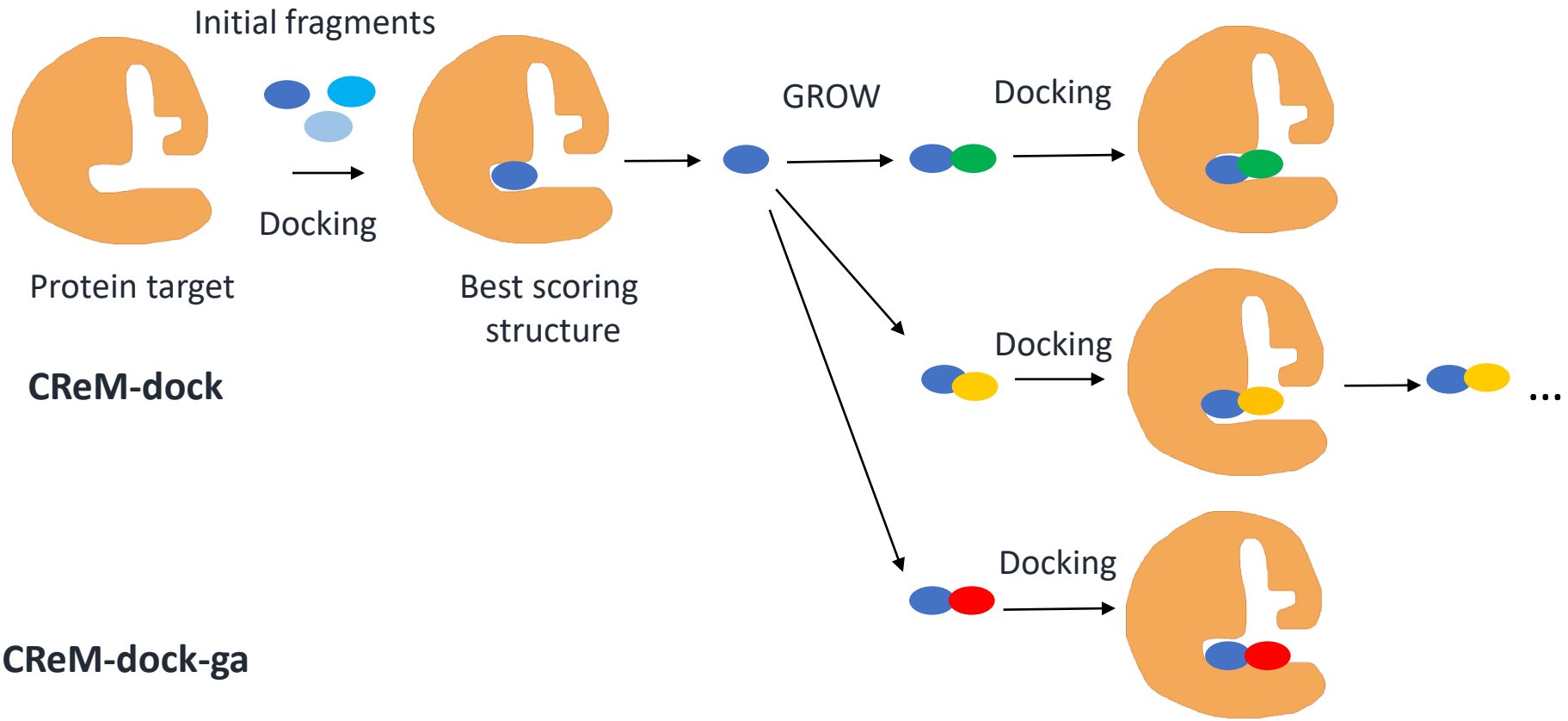
clustering based on
the cavity residues



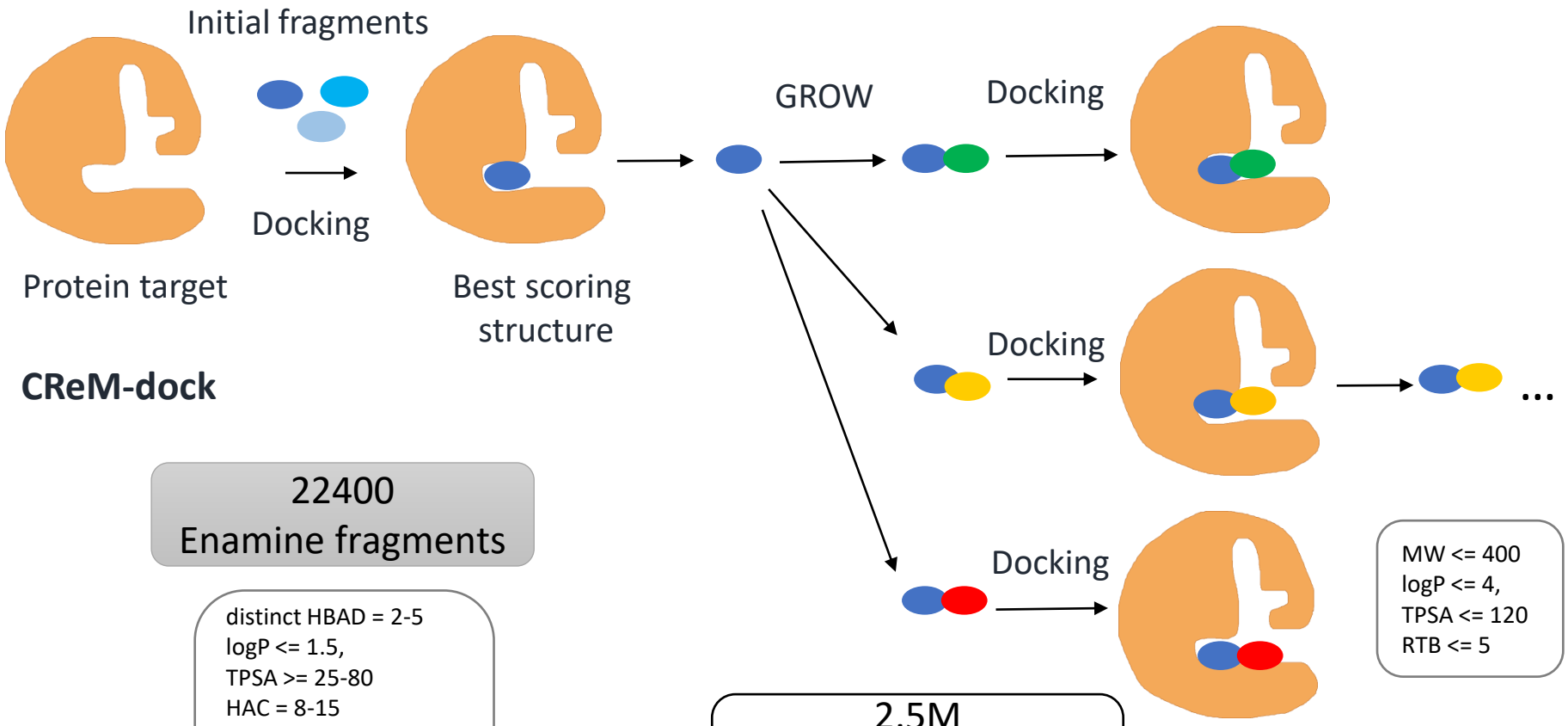
Round 1: chemical space exploration challenge



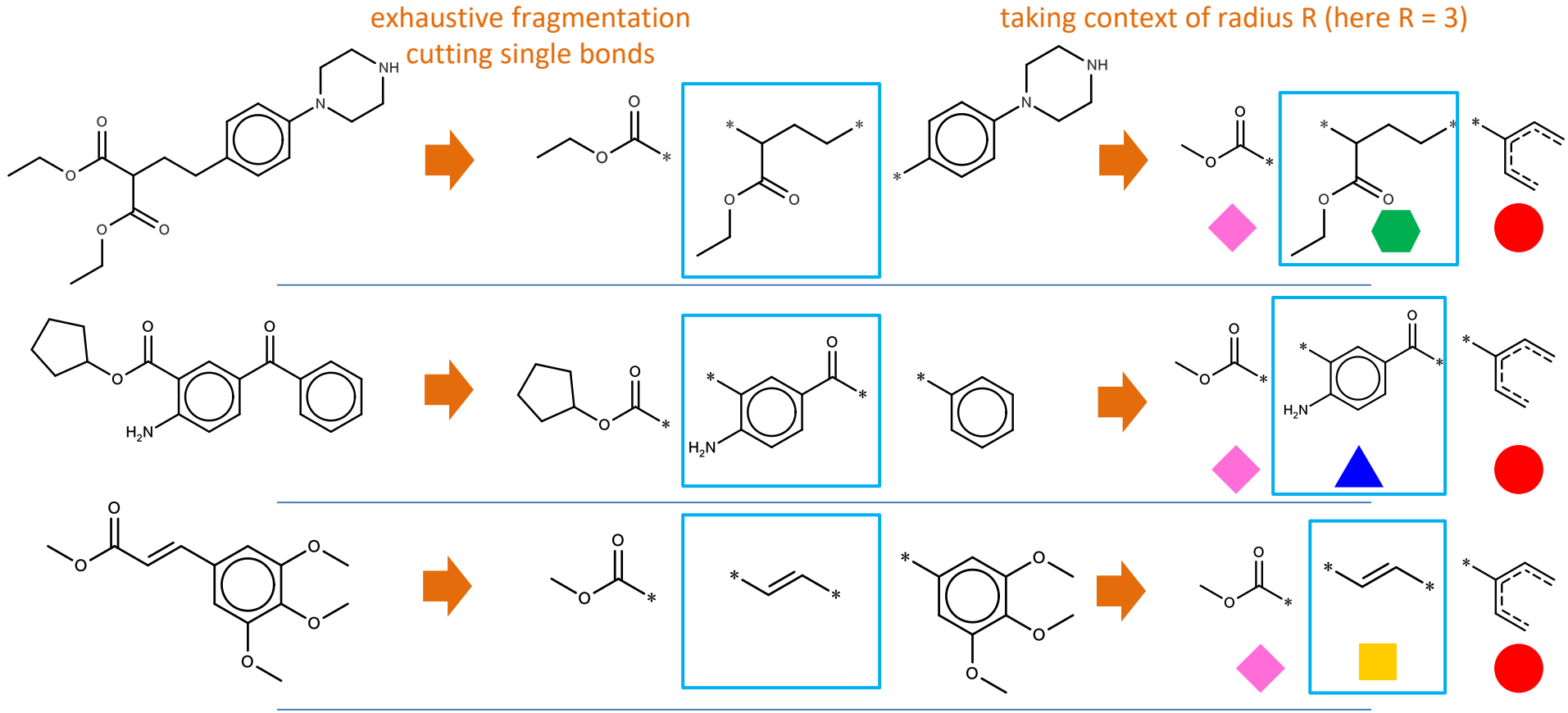
Round 1: strategy 1 (de novo design)



Round 1: strategy 1 (de novo design)



Chemically reasonable mutations (CReM)



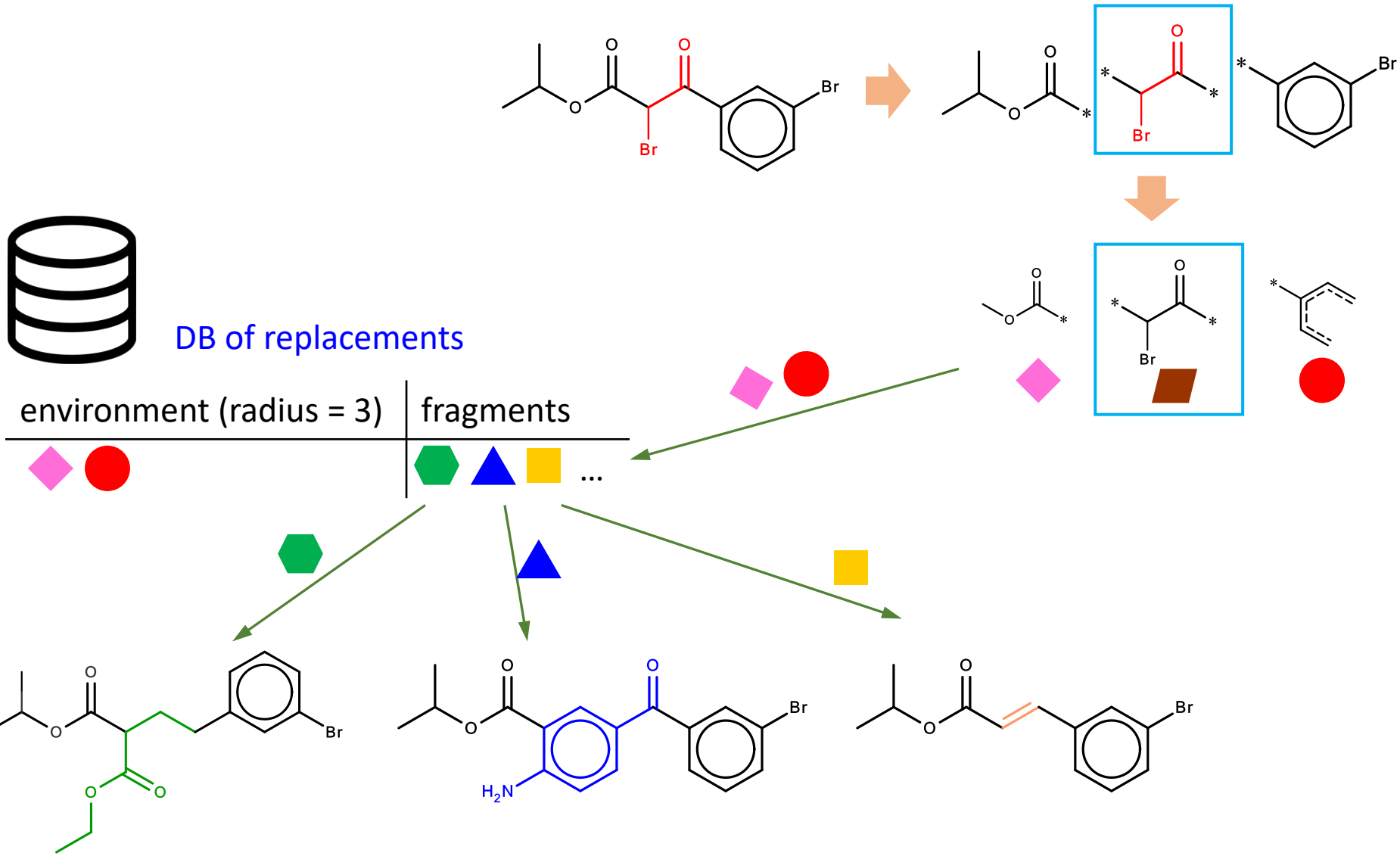
DB of replacements

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

interchangeable fragments

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

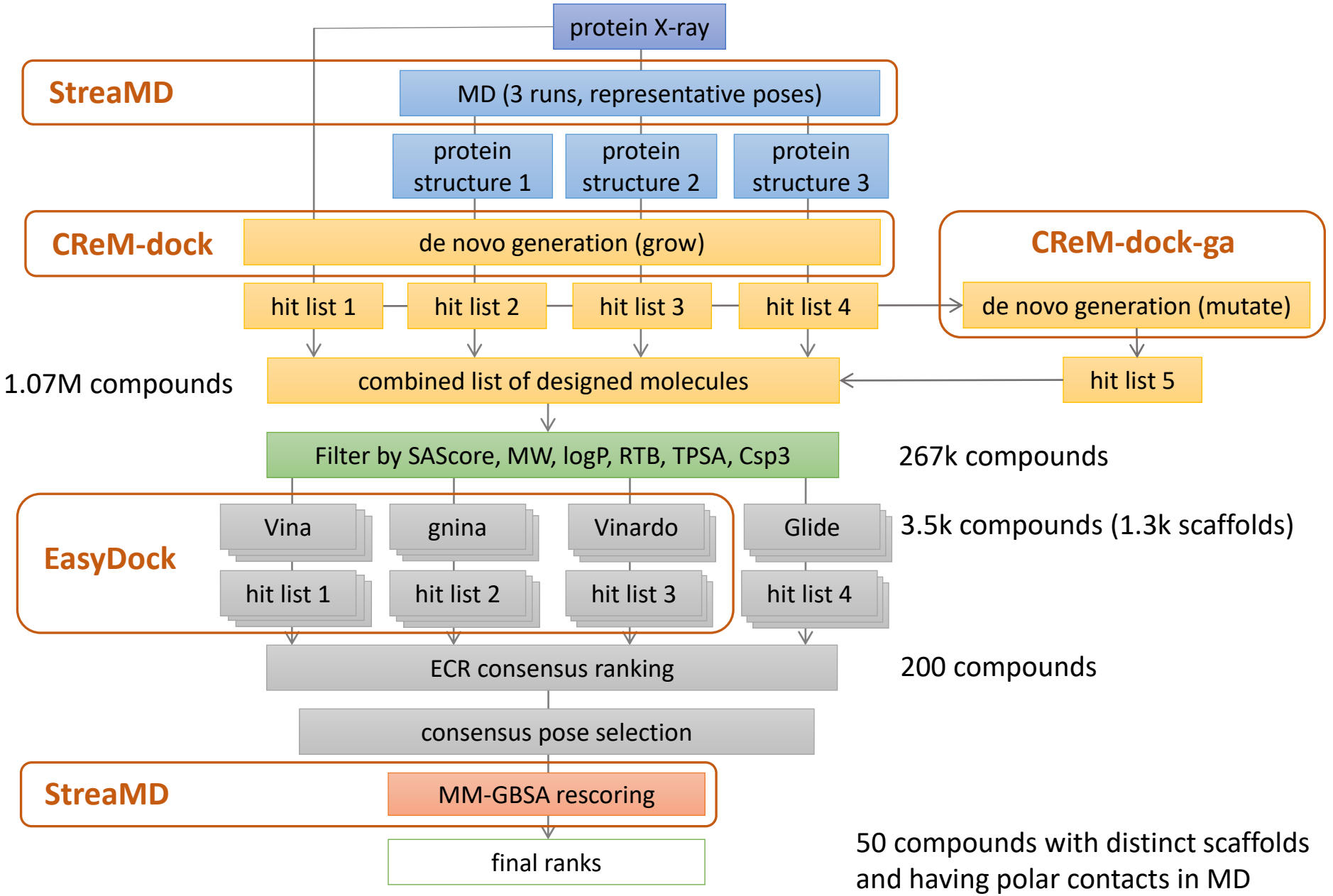
Chemically reasonable mutations (CReM)



Generated structures are always chemically valid!

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

Round 1: strategy 1 (de novo design)

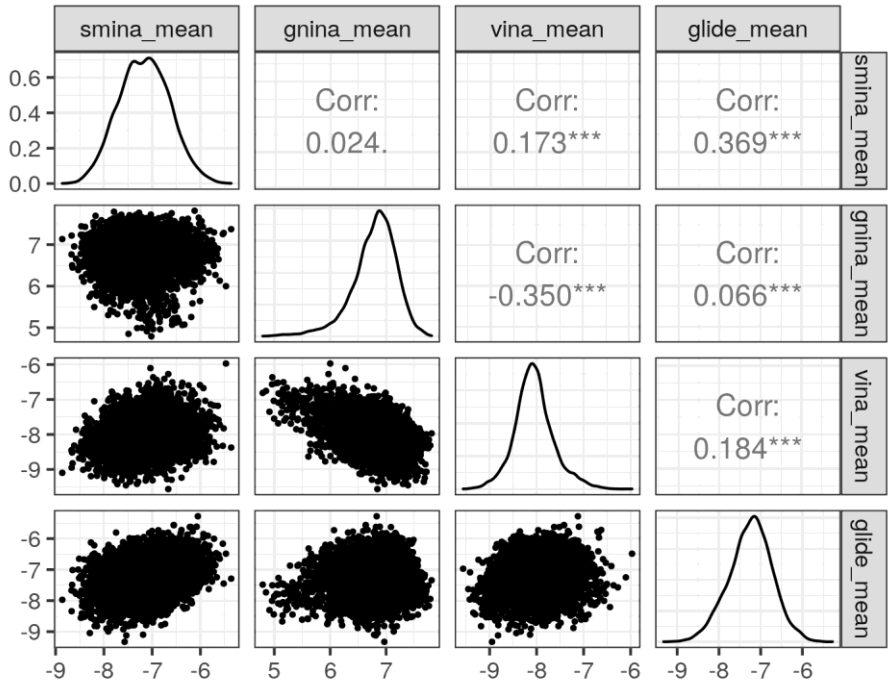


Round 1: strategy 1 (de novo design)

protein X-ray

StreaMD

MD (3 runs, representative poses)



CReM-dock

hit lis

1.07M compounds

CReM-dock-ga

de novo generation (mutate)

hit list 5

compounds

compounds (1.3k scaffolds)

EasyDock

Vina

hit list 1

ompounds

consensus pose selection

StreaMD

MM-GBSA rescoring

final ranks

50 compounds with distinct scaffolds and having polar contacts in MD

Round 1: strategy 1 (de novo design)

protein X-ray

StreaMD

MD (3 runs, representative poses)

protein structure 1 | protein structure 2 | protein structure 3

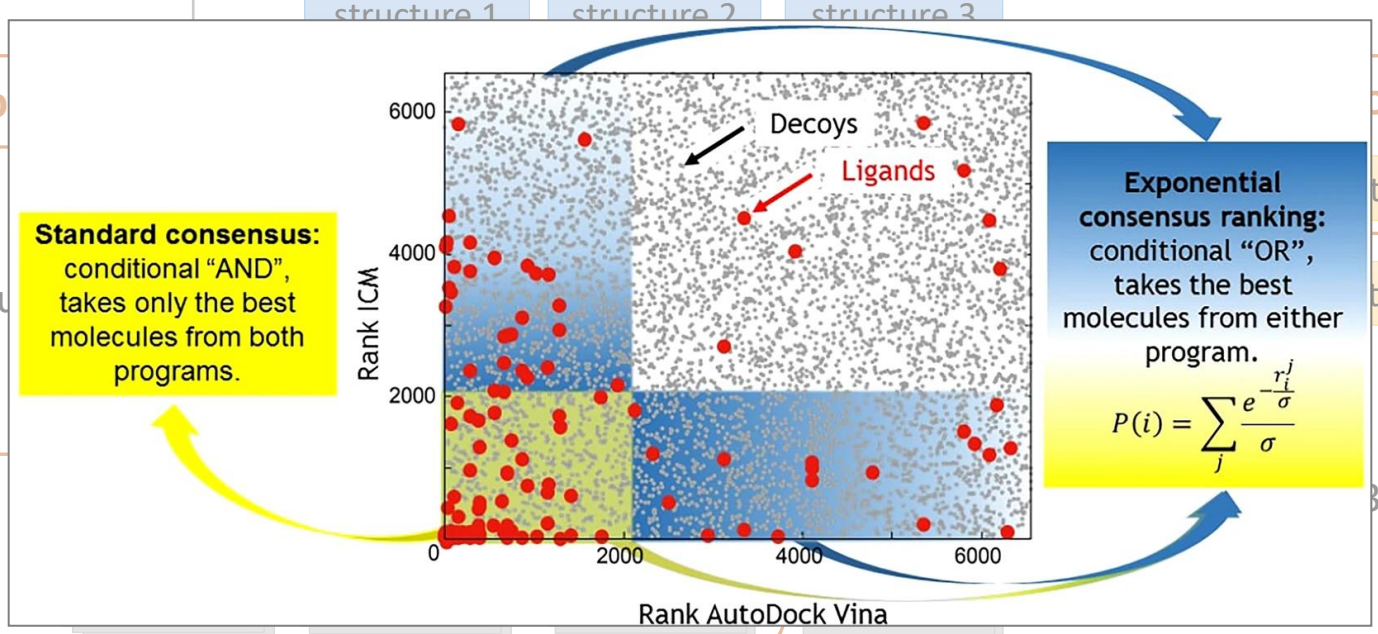
CReM-do

ock-ga

tion (mutate)

1.07M compou

t 5



EasyDock

3k scaffolds)

ECR consensus ranking

200 compounds

consensus pose selection

StreaMD

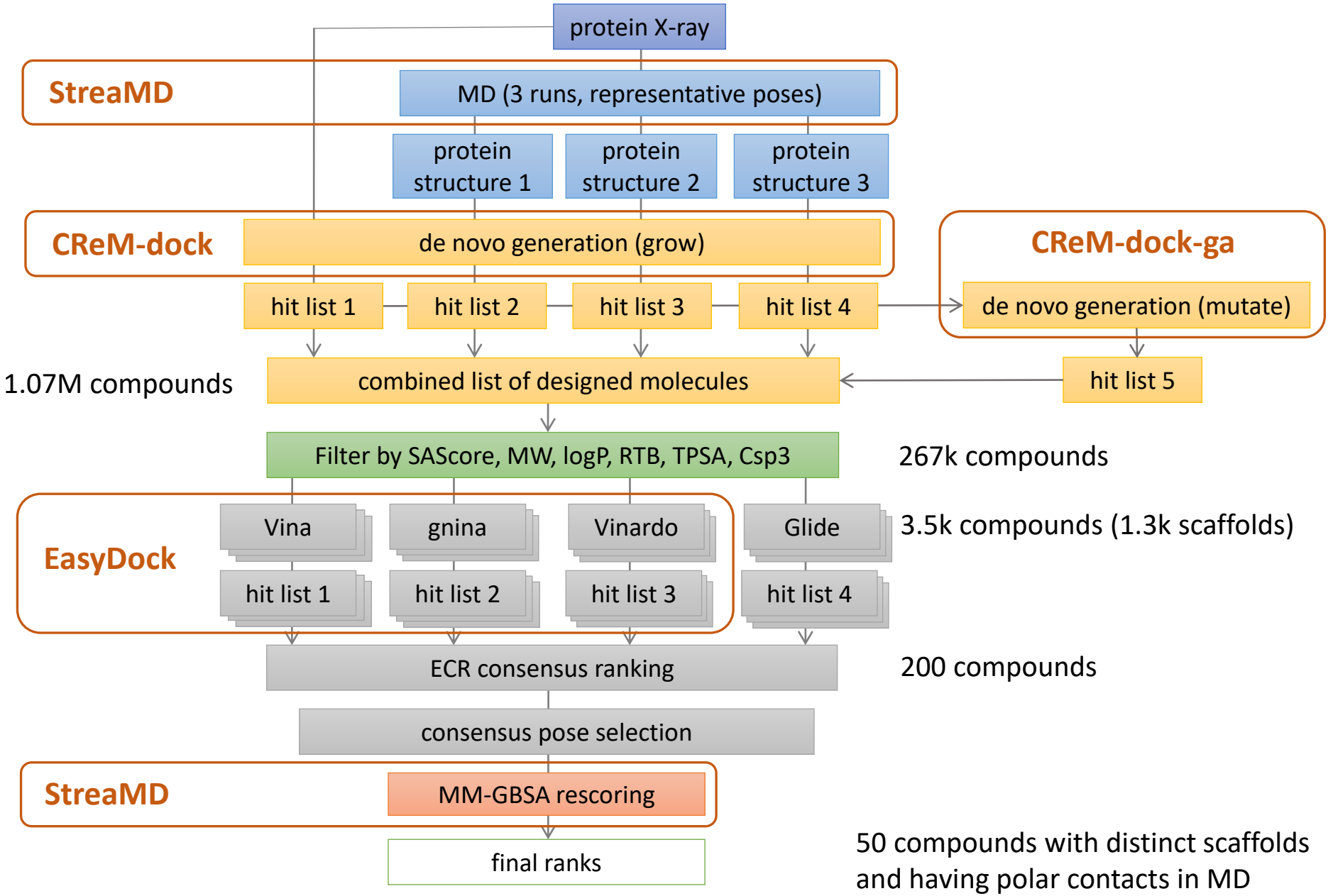
MM-GBSA rescoring

final ranks

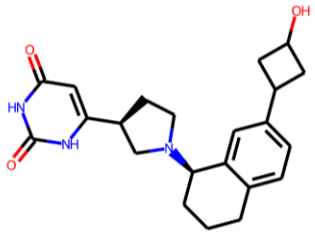
50 compounds with distinct scaffolds and having polar contacts in MD

Palacio-Rodríguez, K.; Lans, I.; Cavasotto, C. N.; Cossio, P. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. *Scientific Reports* **2019**, 9 (1), 5142.

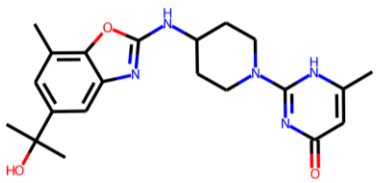
Round 1: strategy 1 (de novo design)



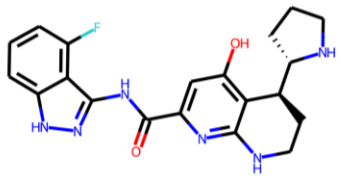
Round 1: strategy 1 (de novo design)



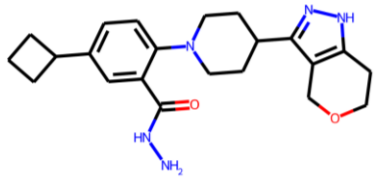
CREM1777121



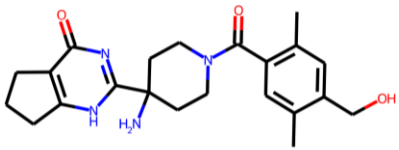
CREM0329741



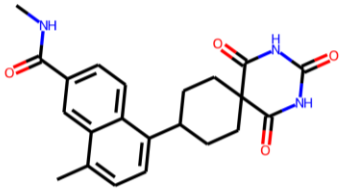
CREM1661038



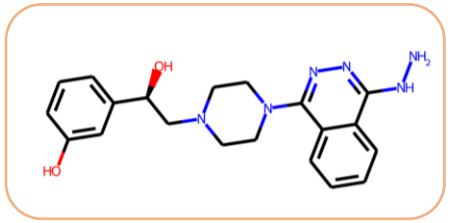
CREM1506273



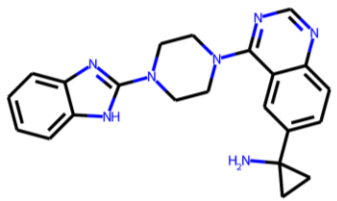
CREM0340409



CREM1089720

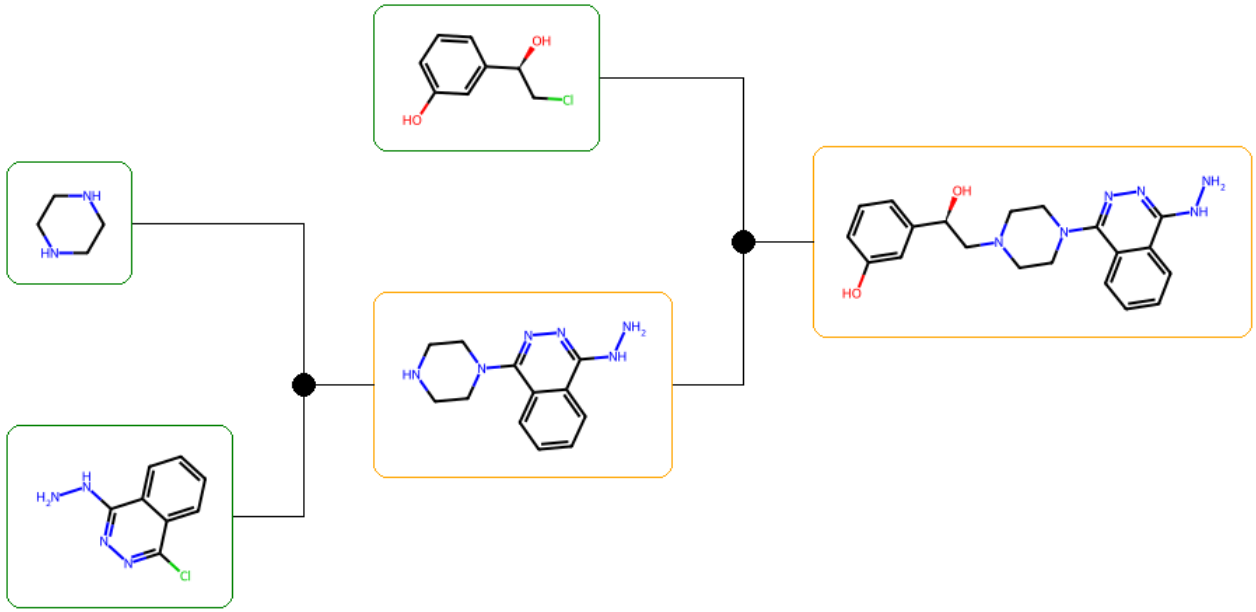


CREM1507777

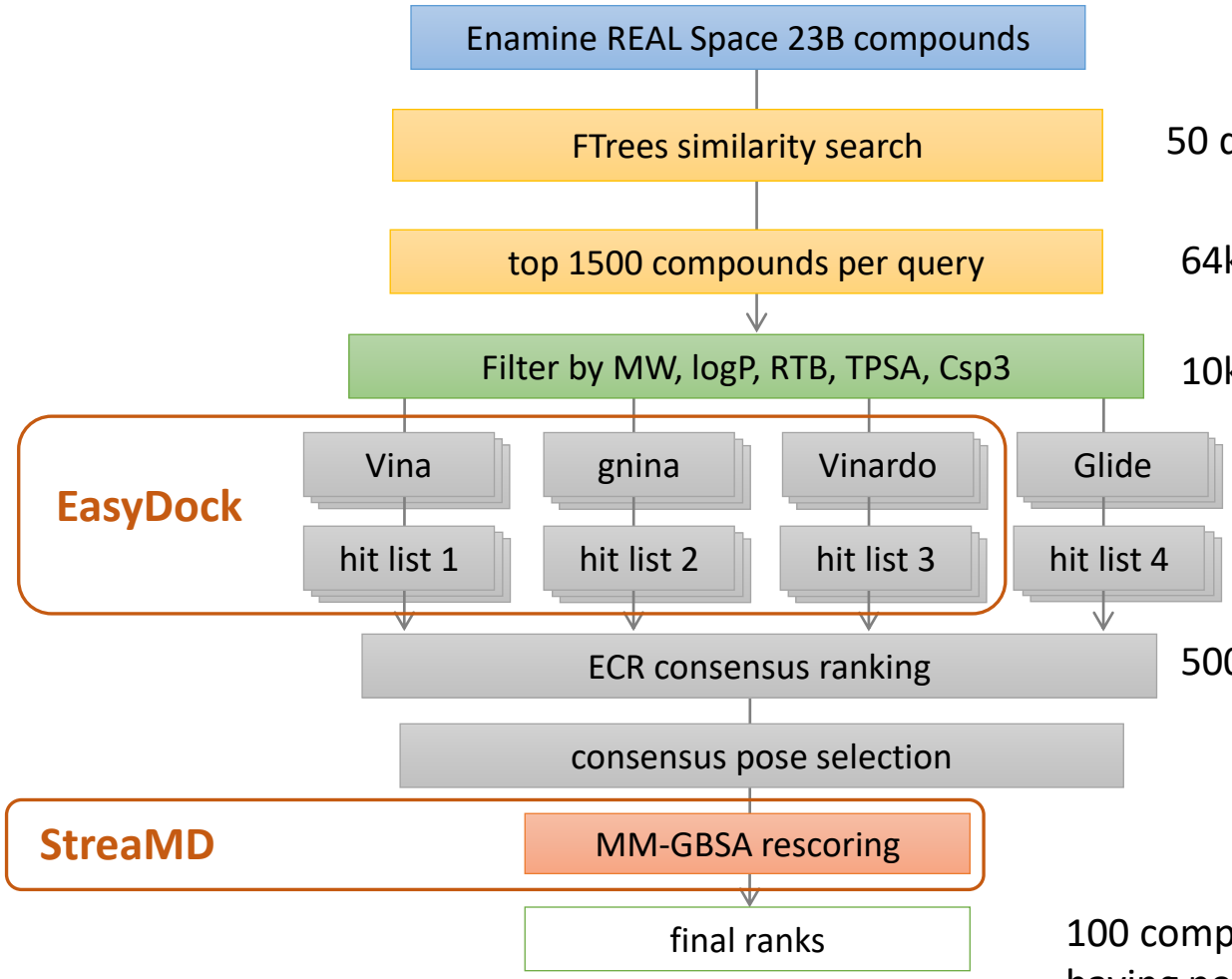


CREM1468894

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



Round 1: strategy 2 (similarity search)



50 de novo compounds (queries)

64k compounds

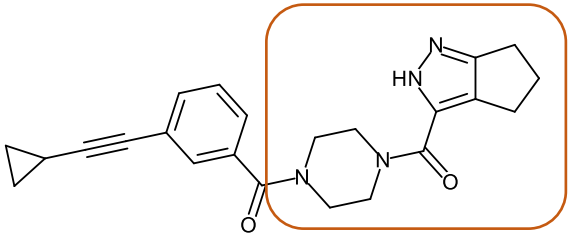
10k compounds (5.1k scaffolds)

500 compounds

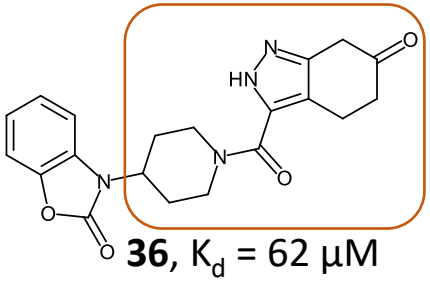
100 compounds with distinct scaffolds, having polar contacts in MD and no PAINS

Round 1: experimental results

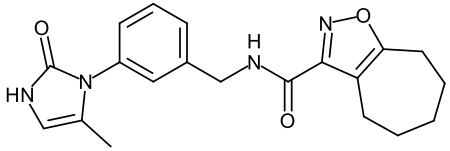
50 de novo + 100 similar compounds
91 compounds were selected (within the budget 9000$)
82 compounds were synthesized
8 compounds demonstrated activity (K_d = 25-117 µM by SPR)



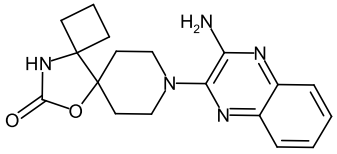
1, K_d = 61 µM



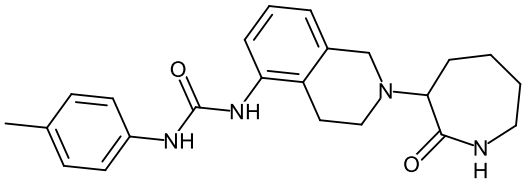
36, K_d = 62 µM



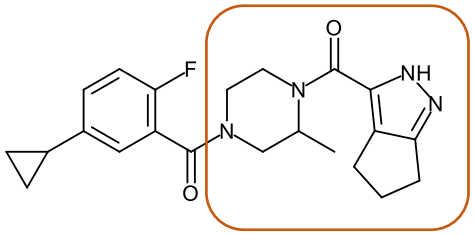
59, K_d = 32 µM



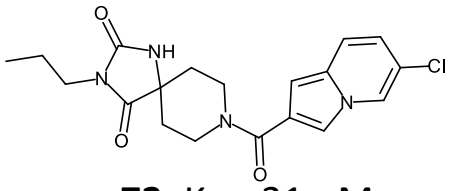
62, K_d = 25 µM



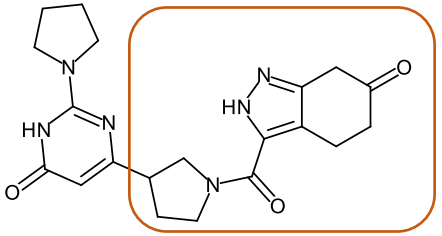
65, K_d = 56 µM



69, K_d = 117 µM

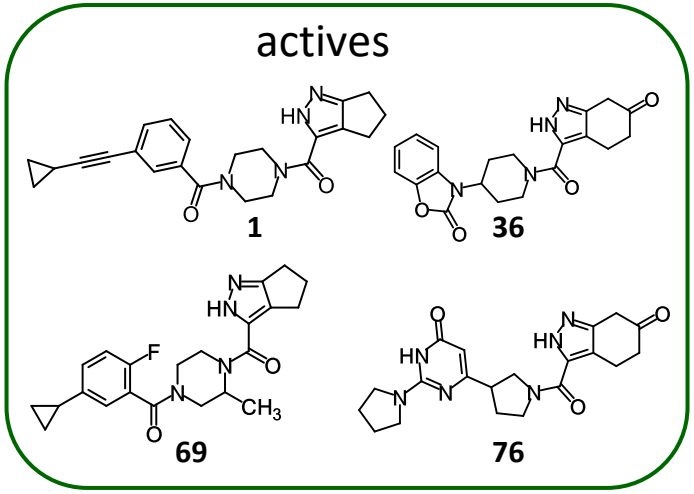


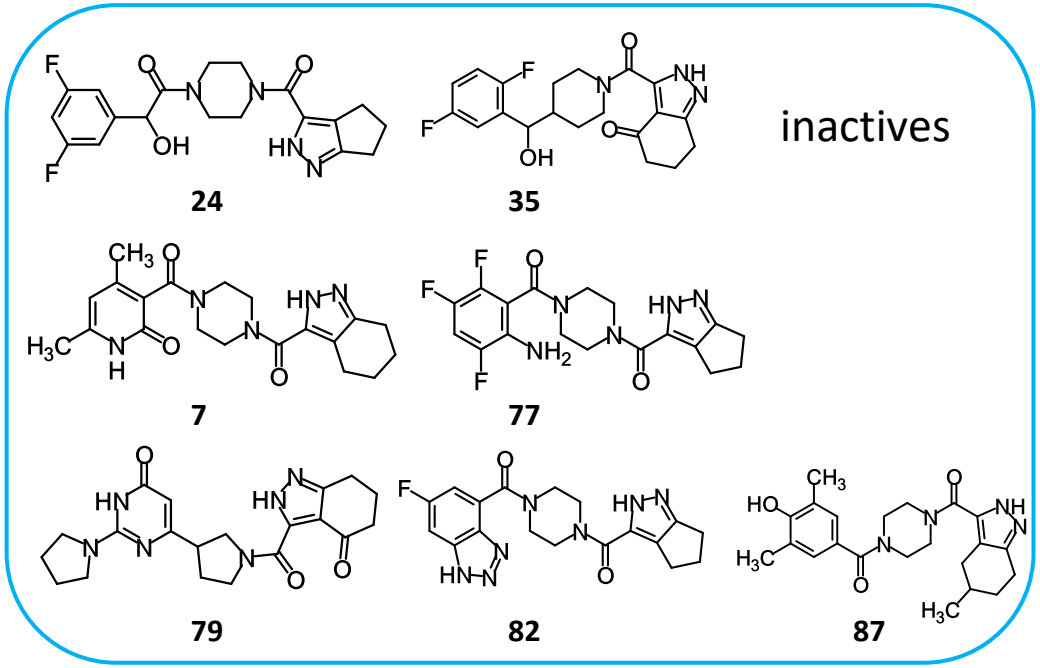
73, K_d = 31 µM



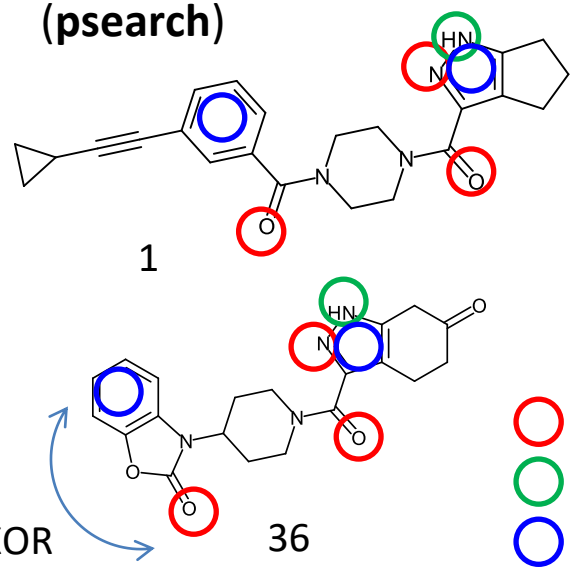
76, K_d = 74 µM

Round 2: hit optimization (compound pool 1)





3D ligand-based pharmacophores (**psearch**)



precision: 0.43-0.5
recall: 0.75
EF: 7.2-8.4

H-bond acceptor
H-bond donor
aromatic/hydrophobic

2.5M
Enamine Stock

the most restrictive
pharmacophore model

155 compounds

Round 2: hit optimization (compound pool 2)

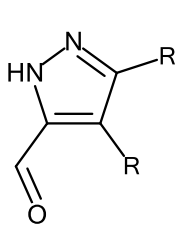
2.5M
Enamine Stock

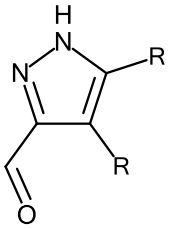
↓

substructure search

↓

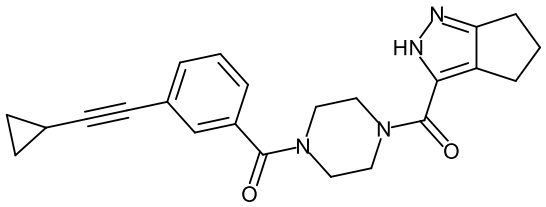
18 411 compounds





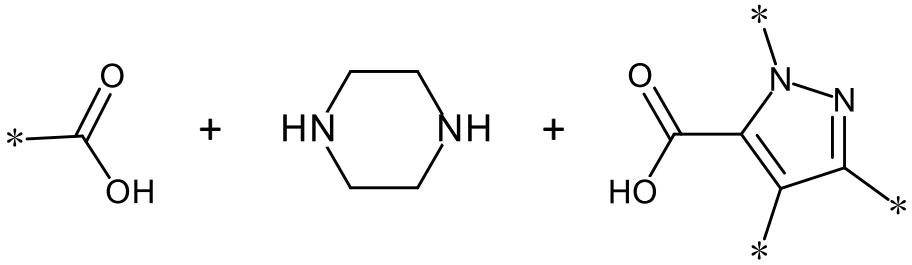
Round 2: hit optimization (compound pool 3)

Enamine fragments → substructure search → 18 845 building blocks



1, IC_{50} = 61 µM

Enamine fragments → substructure search → 474 building blocks

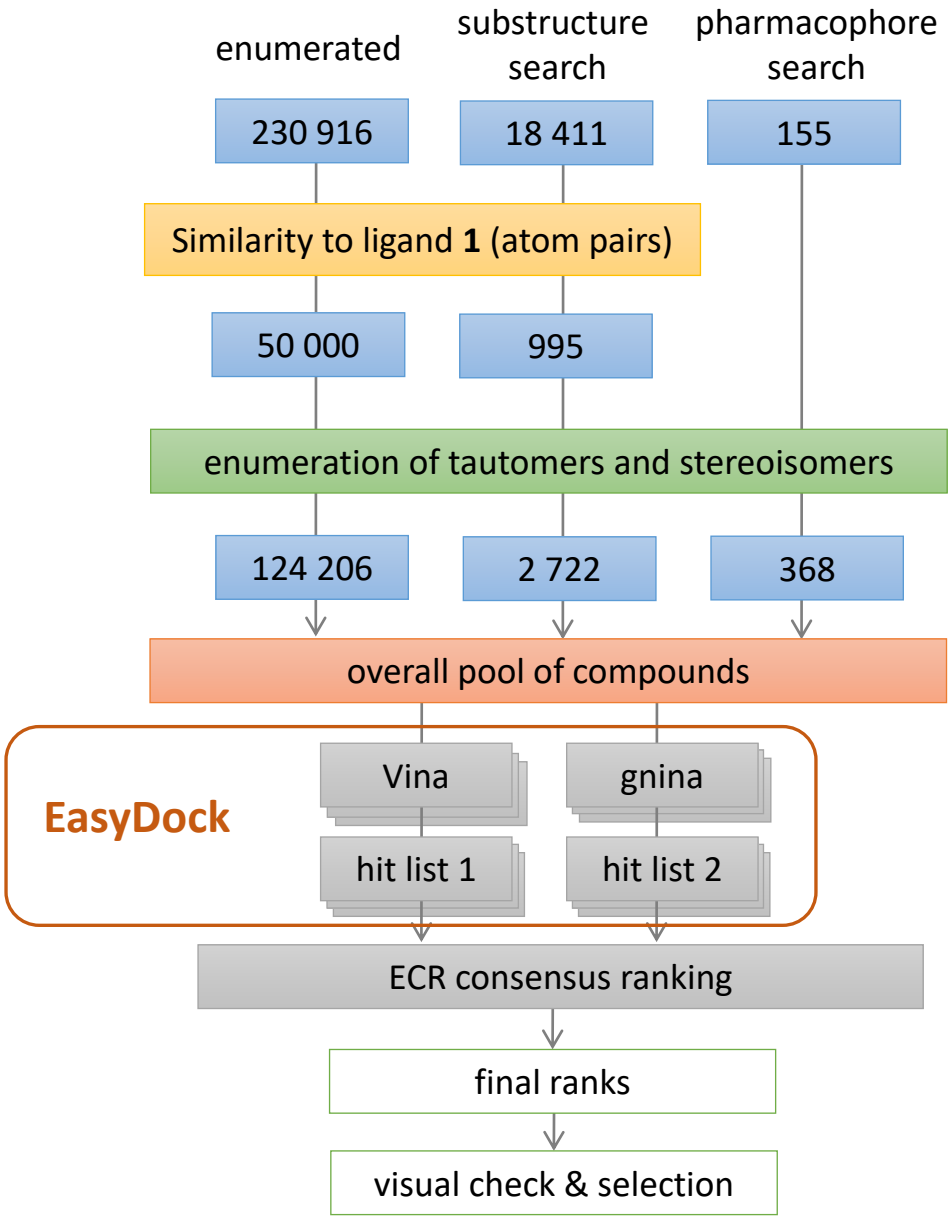


2 943 486 enumerated molecules

Filter by MW, logP, TPSA, RTB, Csp3

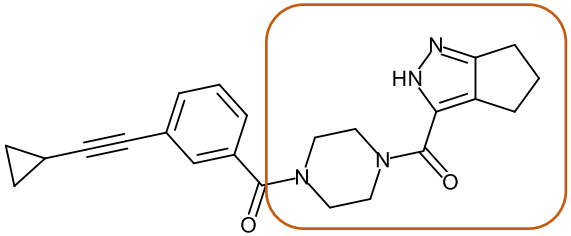
230 916 compounds

Round 2: hit optimization (screening pipeline)

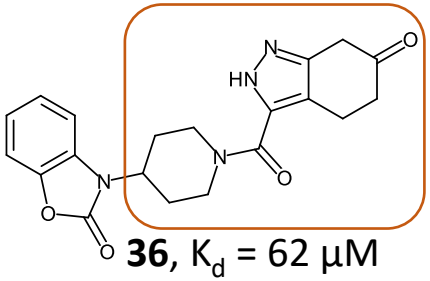


Round 2: hit optimization (experimental results)

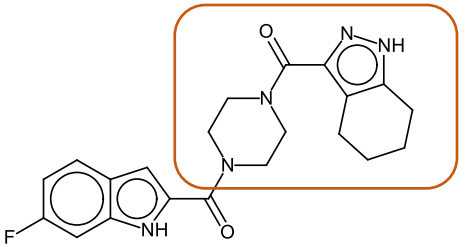
38 compounds were selected (within the budget 4500$)
35 compounds were synthesized
4 compounds demonstrated some effect in SPR
1 scaffold had confirmed selectivity



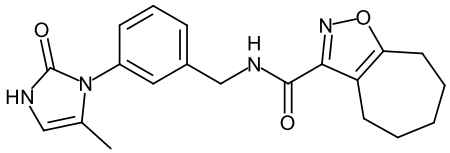
1, K_d = 61 µM



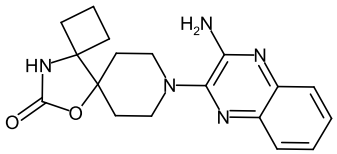
36, K_d = 62 µM



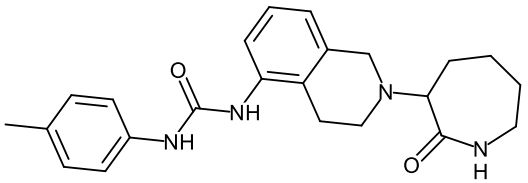
HO-15, K_d = 71 µM



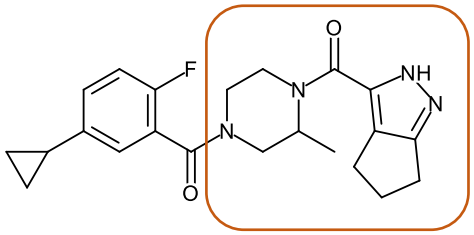
59, K_d = 32 µM



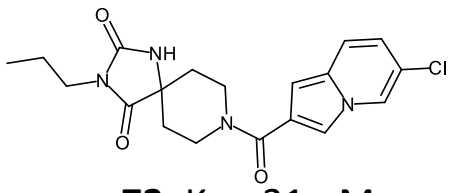
62, K_d = 25 µM



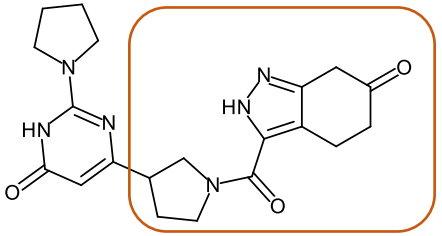
65, K_d = 56 µM



69, K_d = 117 µM



73, K_d = 31 µM

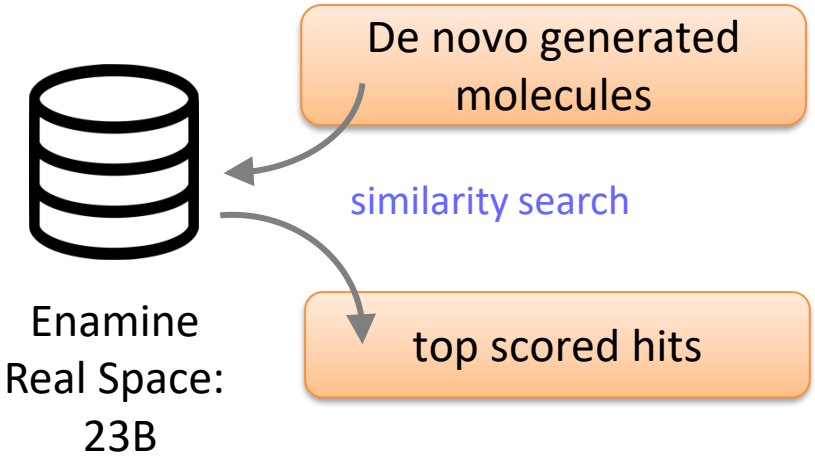


76, K_d = 74 µM

Summary of the CReM-based pipeline

Round 1

- 1.27M docking events and 700 MM-GBSA were enough to discover 8 primary hits among 82 compounds retrieved from Enamine REAL Space
- no human selection



Round 2

while 4 compounds demonstrated some effect on WDR among 35 tested ones, only one had confirmed selectivity. The observed SAR is inconclusive.

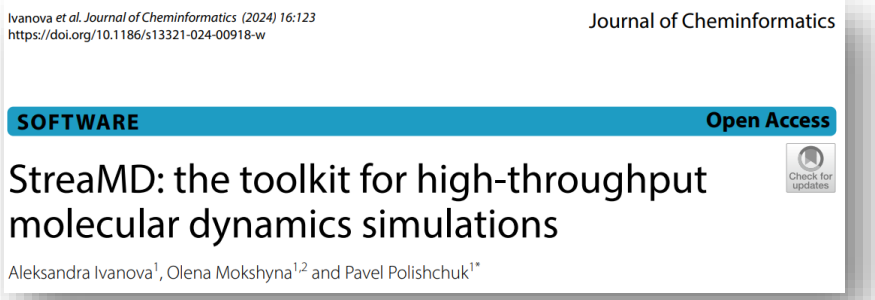

Minibaeva *et al. Journal of Cheminformatics* (2023) 15:102
https://doi.org/10.1186/s13321-023-00772-2

Journal of Cheminformatics

SOFTWARE Open Access

EasyDock: customizable and scalable docking tool

Guzel Minibaeva[1], Aleksandra Ivanova[1] and Pavel Polishchuk[1*]


Ivanova *et al. Journal of Cheminformatics* (2024) 16:123
https://doi.org/10.1186/s13321-024-00918-w

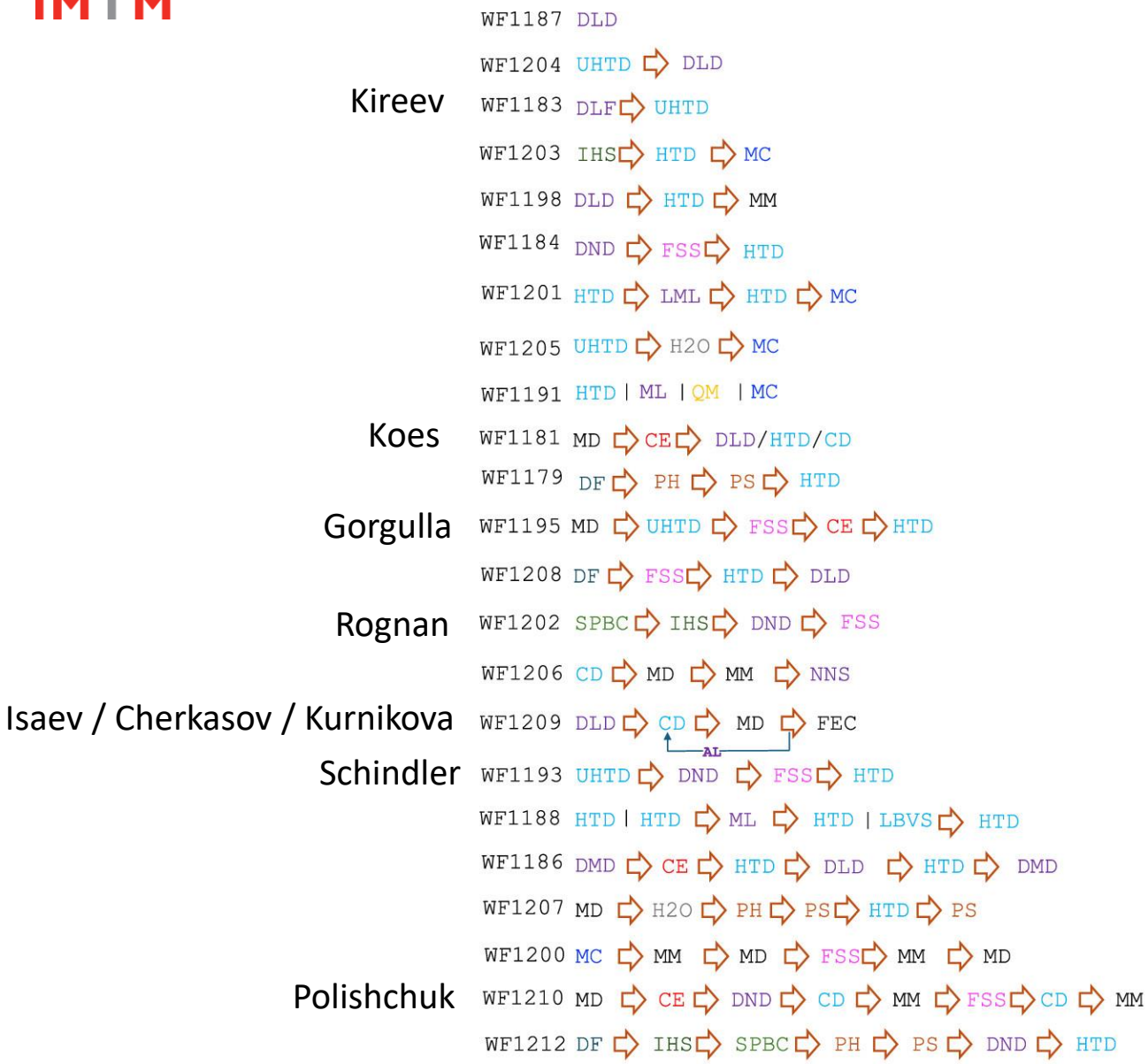
Journal of Cheminformatics

SOFTWARE Open Access

StreaMD: the toolkit for high-throughput molecular dynamics simulations

Aleksandra Ivanova[1], Olena Mokshyna[1,2] and Pavel Polishchuk[1*]

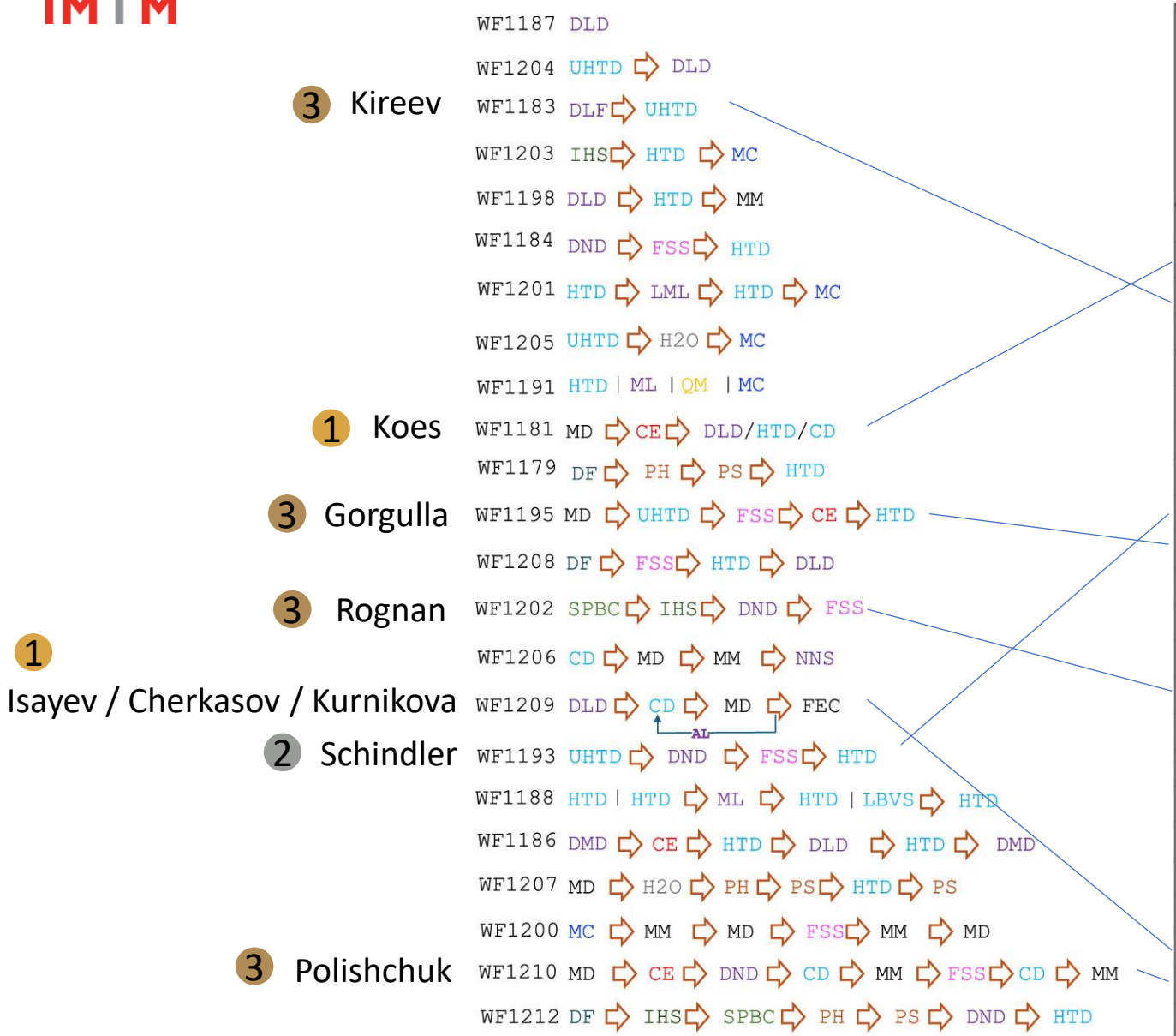
Pipelines of all participants



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.

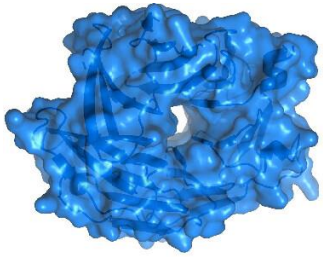
Pipelines of all participants



Participant	Compounds tested experimentally	Compounds advancing to Round 2	Hit rate
1179	100	2	2
1181	84	2	2
1183	37	2	5
1184	65	2	3
1186	99	11	11
1187	72	4	6
1188	113	3	3
1191	95	0	0
1193	92	4	4
1195	84	10	12
1198	79	0	0
1200	91	2	2
1201	101	1	1
1202	94	5	5
1203	105	2	2
1204	71	0	0
1205	98	3	3
1206	83	0	0
1207	101	1	1
1208	100	4	4
1209	59	7	12
1210	82	8	10
1212	50	0	0

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.

David Koes pipeline 1



Crystal Structure

Docking with GNINA

molport

~7 million molecules

Top 1k Molecules by GNINA score

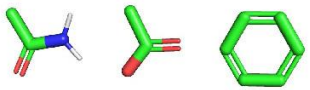
Docking with GNINA

1000 molecules, + GNINA Scores

Molecular Weight < 550 Da

Cluster by FP2 fingerprint

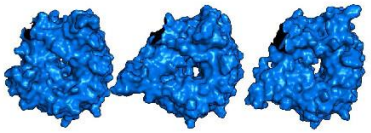
Vendor availability



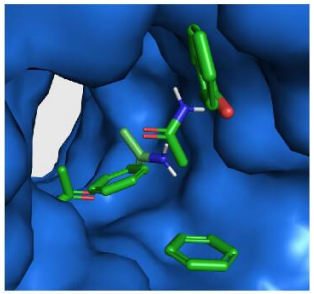
Molecule Fragments

- ZINC20: 20 mil molecules
- MCULE: 45 mil molecules
- MCULE-ULTIMATE: 126 mil molecules

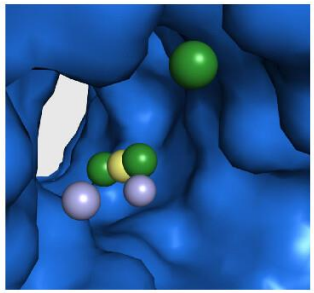
Molecule Libraries



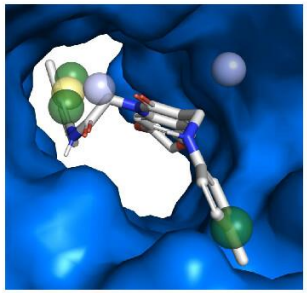
Molecular Dynamics Ensemble Structures



Docking with GNINA



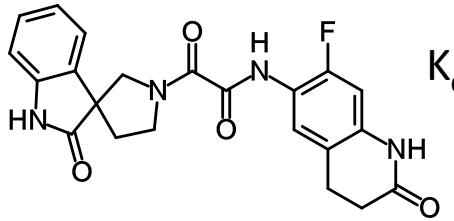
Pharmacophore Generation



Pharmacophore Screening with PHARMIT

Docking with GNINA

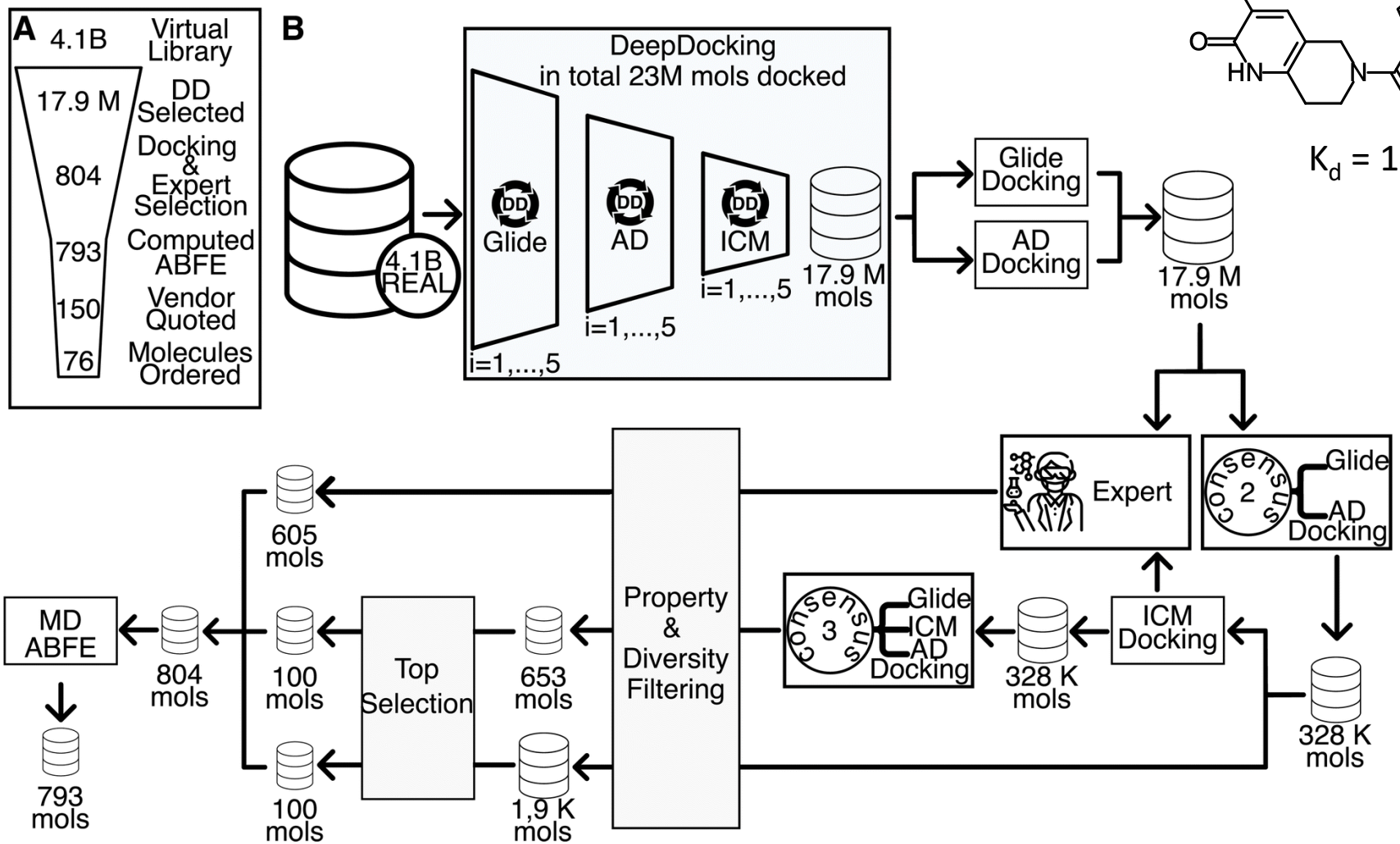
3572 molecules + GNINA scores

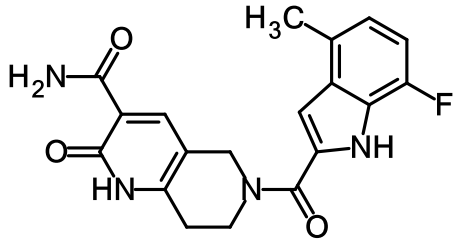


K_d = 56 µM

Dunn, I.; Pirhadi, S.; Wang, Y.; Ravindran, S.; Concepcion, C.; Koes, D. R. CACHE Challenge #1: Docking with GNINA Is All You Need. *J. Chem. Inf. Model.* **2024**, 64 (24), 9388-9396.

Isayev/Cherkasov/Kurnikova pipeline 1

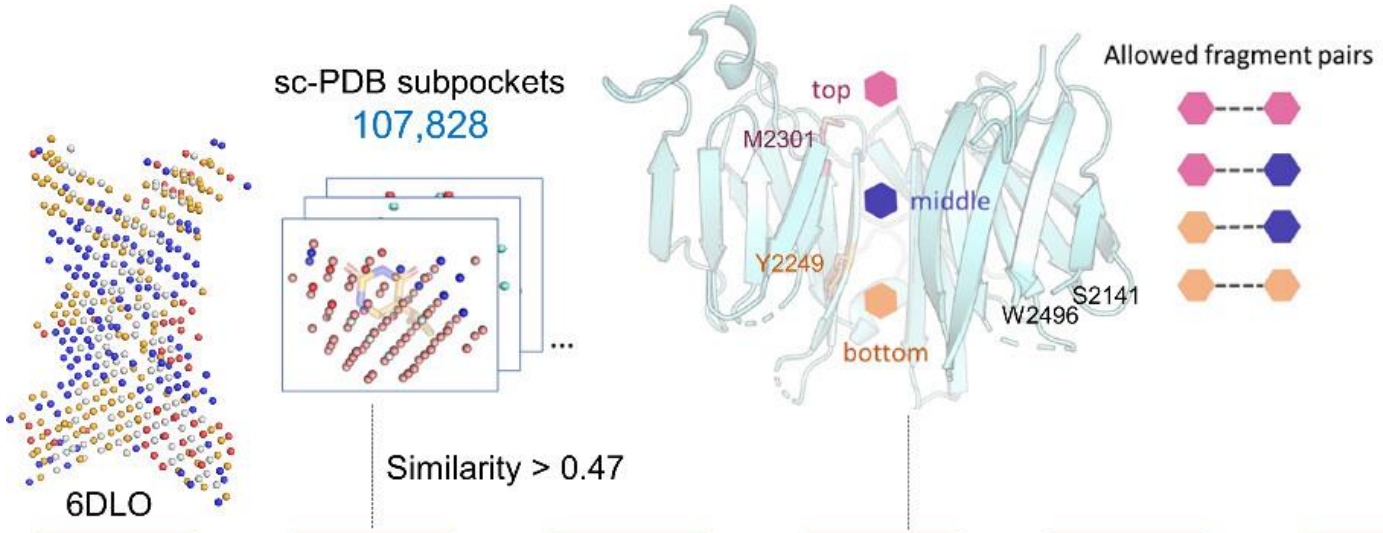


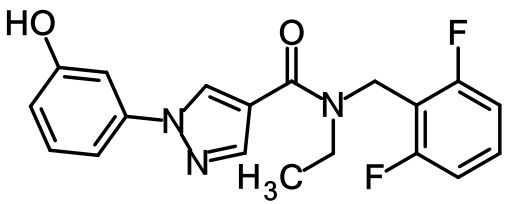


K_d = 142 μM

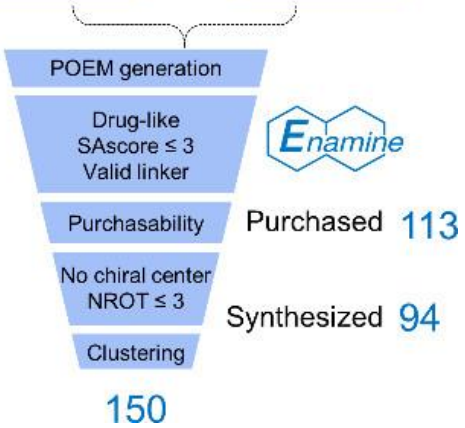
Gutkin, E.; Gusev, F.; Gentile, F.; Ban, F.; Koby, S. B.; Narangoda, C.; Isayev, O.; Cherkasov, A.; Kurnikova, M. G. In silico screening of LRRK2 WDR domain inhibitors using deep docking and free energy simulations. *Chem. Sci.* **2024**, 15 (23), 8800-8812.

Didier Rognan pipeline 3



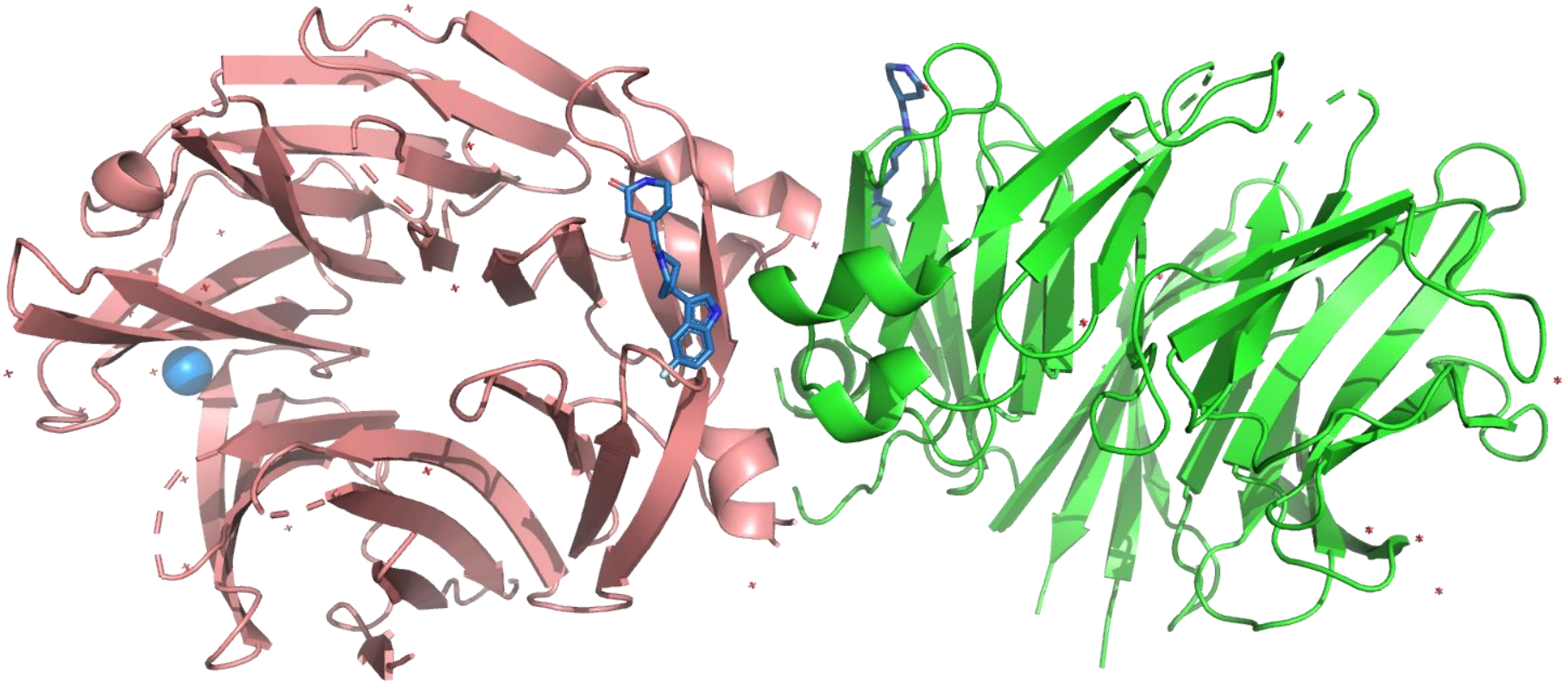


K_d = 111 µM



Eguida, M.; Bret, G.; Sindt, F.; Li, F.; Chau, I.; Ackloo, S.; Arrowsmith, C.; Bolotokova, A.; Ghiabi, P.; Gibson, E.; et al. Subpocket Similarity-Based Hit Identification for Challenging Targets: Application to the WDR Domain of LRRK2. *J. Chem. Inf. Model.* **2024**, 64 (13), 5344-5355.

Attempt to solve X-ray structure



https://cache-challenge.org/results-cache-challenge-1

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

IMTM

Palacky University (Olomouc):

Guzel Minibaeva

Aleksandra Ivanova

Alina Kutlushina

University of Chemistry and Technology (Prague):

Dr. Vojtech Spiwok

Jan Beranek

Thank you for your attention

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). We also acknowledge the contributions from the project ENOCH (CZ.02.1.01/0.0/0.0/16_019/0000868).