7th Advanced In silico Drug Design workshop/challenge

29 January - 2 February 2024
Olomouc, Czech Republic

Univerzita Palackého v Olomouci

The first CACHE challenge: searching for hit molecules in ultra-large chemical databases

Pavel Polishchuk

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

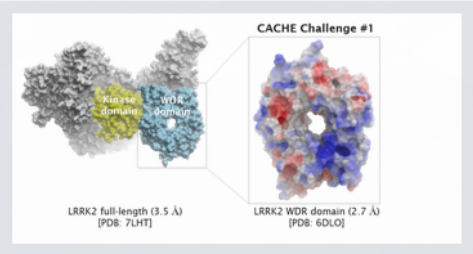
IMTM

Our motivation

1. Validate and improve our developed modeling tools in a competitive environment
2. Establish robust and reliable computational pipelines which can be further easily applied in other projects

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.

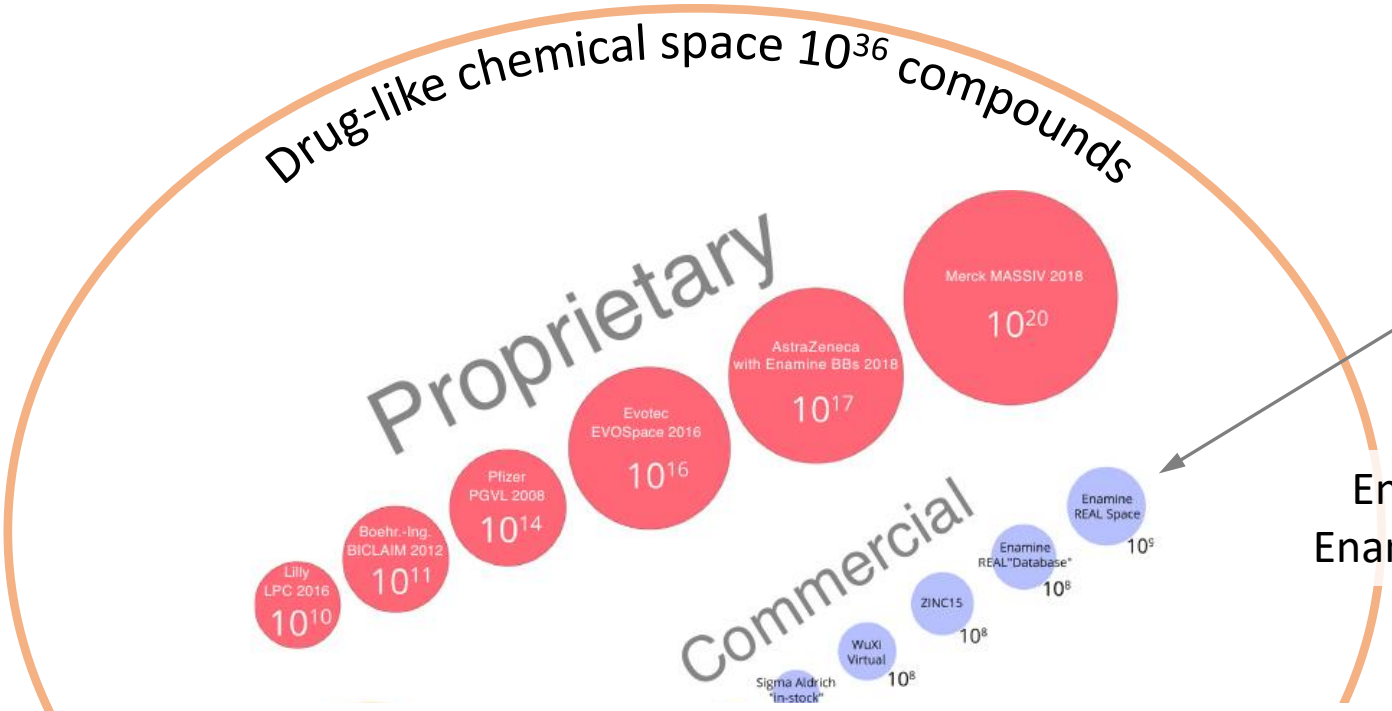
IMTM

LRRK2 and WDR domain

No known active molecules

No X-ray of protein-ligand complexes

Chemical search space



Enamine Stock: 2.5M
Enamine Real Space: 16B

Large space to explore

Traffic light score	Score	Binding affinity[a] (μM)	Solubility in water[a] (mg l^{-1})	logD (pH 7.5)[a]	MWcorr	PSA (Å^2)	Number of rotatable bonds	Fsp^3	Novelty[b]
	2	>10	<10	>4	>500	>140	≥11	<0.2	>0.6
	1	1–10	10–50	3–4	400–500	120–140	8–10	0.2–0.3	0.4–0.6
	0	<1	≥50	<3	≤400	≤120	≤7	>0.3	<0.4

Fsp^3, fraction of sp^3 hybridized carbon atoms, calculated based on Murcko scaffolds. [a]Measured experimentally. [b]Tanimoto distance relative to most similar structures binding that target, as calculated from RDKit. PSA, polar surface area.

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

Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A., Estimation of the size of drug-like chemical space based on GDB-17 data. *Journal of Computer-Aided Molecular Design* **2013**, 27, 675-679

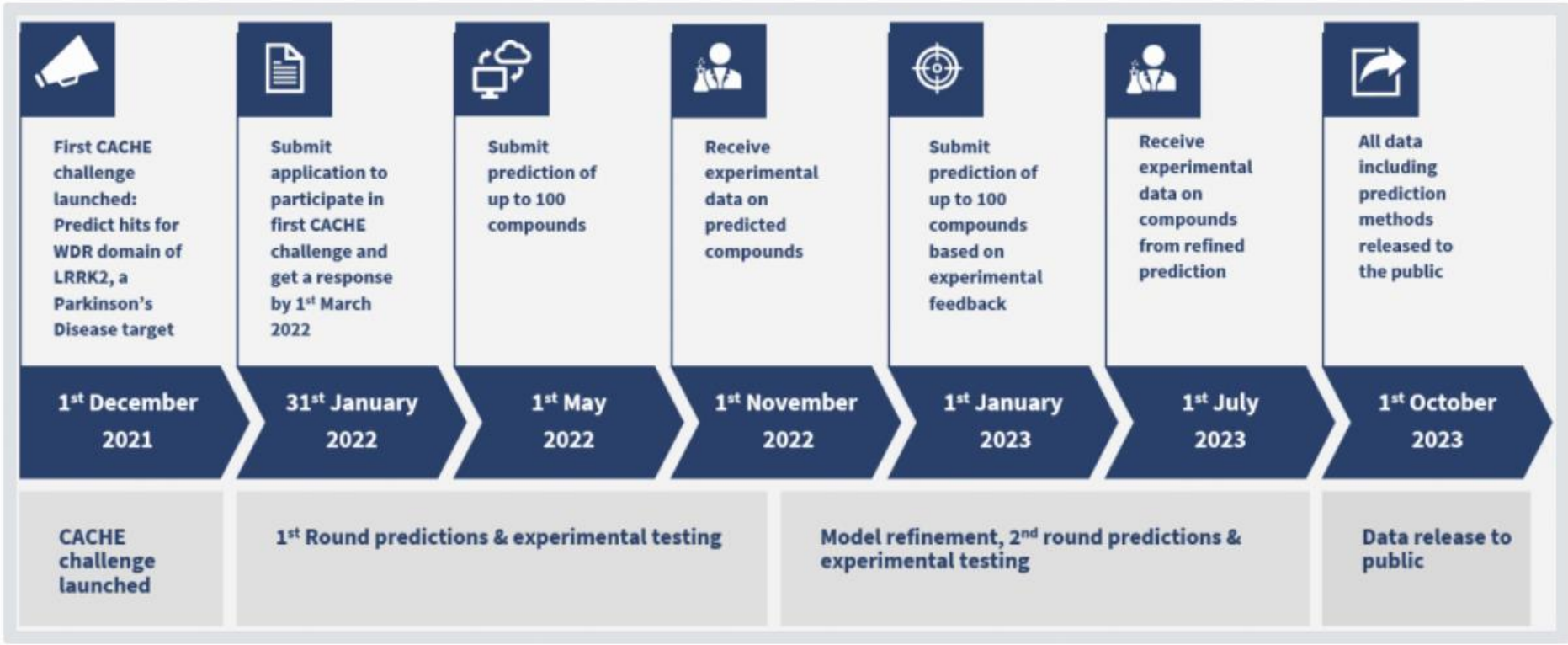
CACHE challenge pipeline

Application opens
2021-12-01

Application closes
2022-01-31

Application form
Download

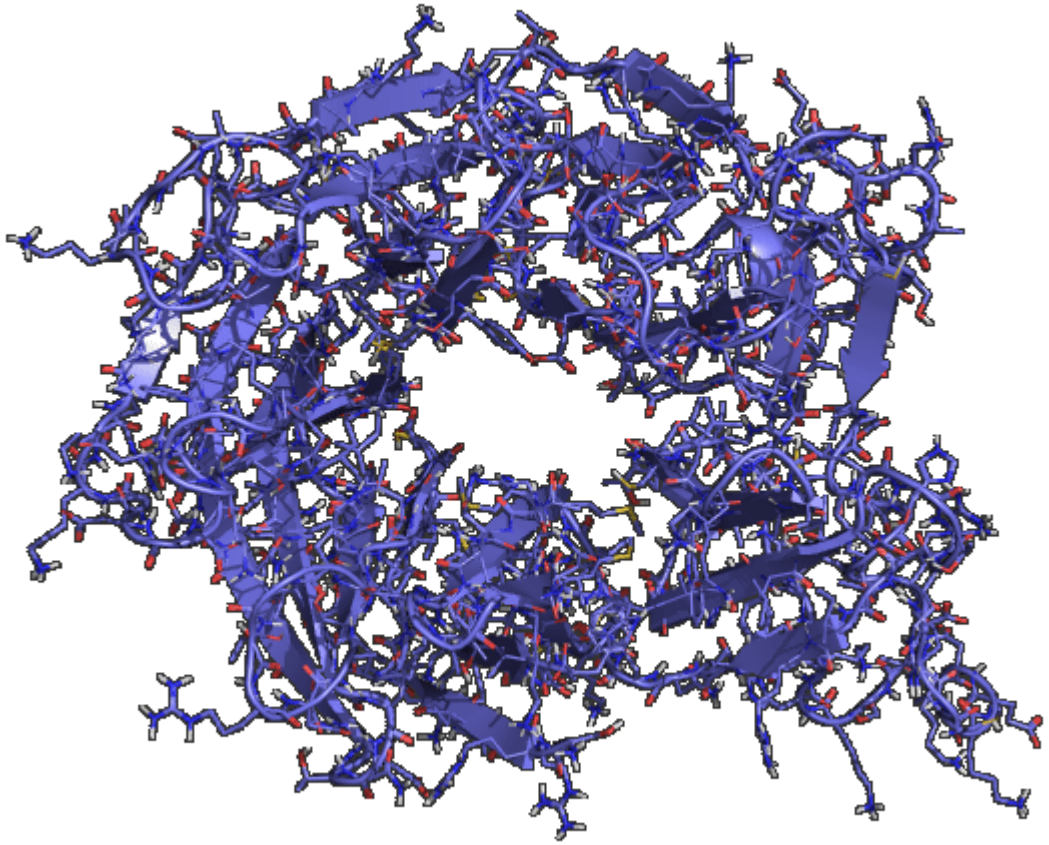
TIMELINE



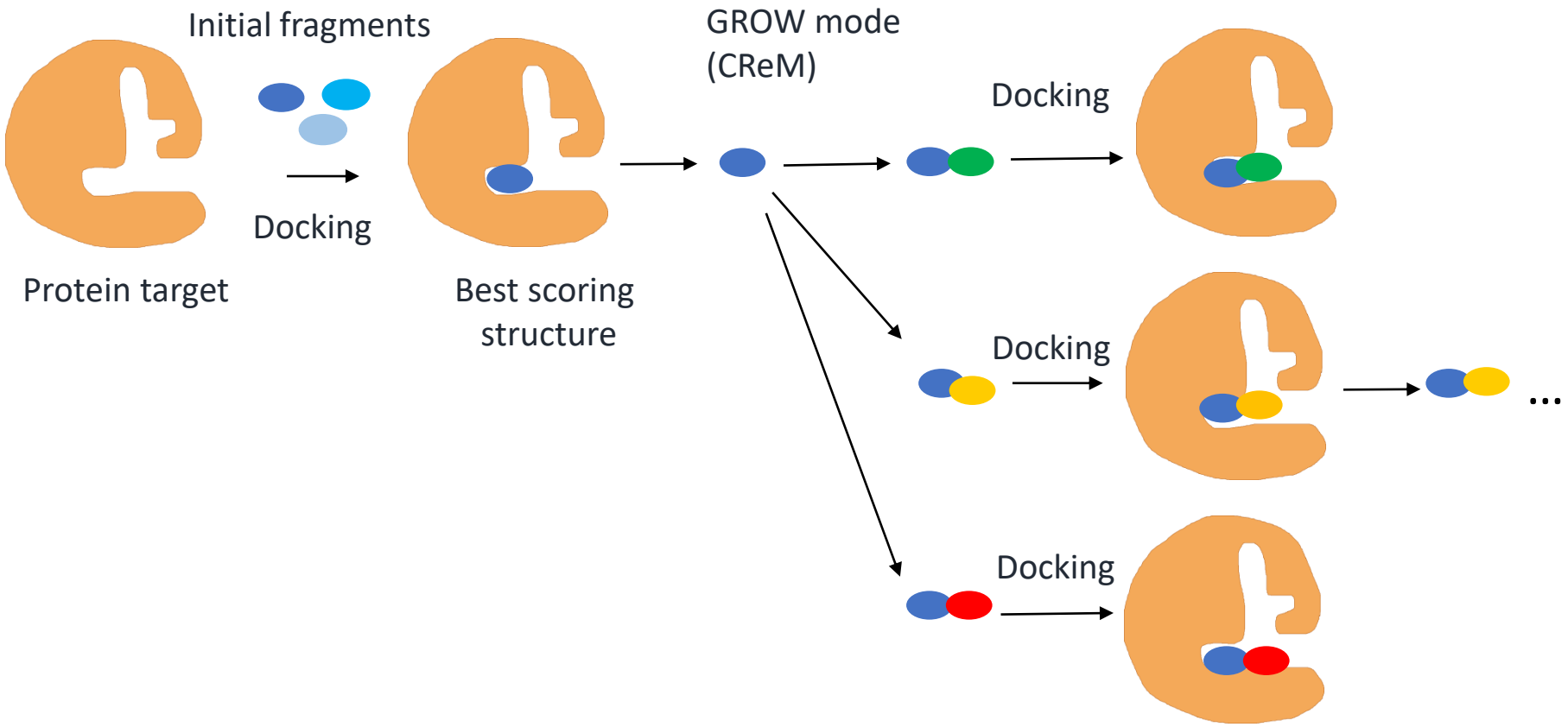
Round 1

WDR domain structure is **available**: 6DLO
Known ligand are **not available**

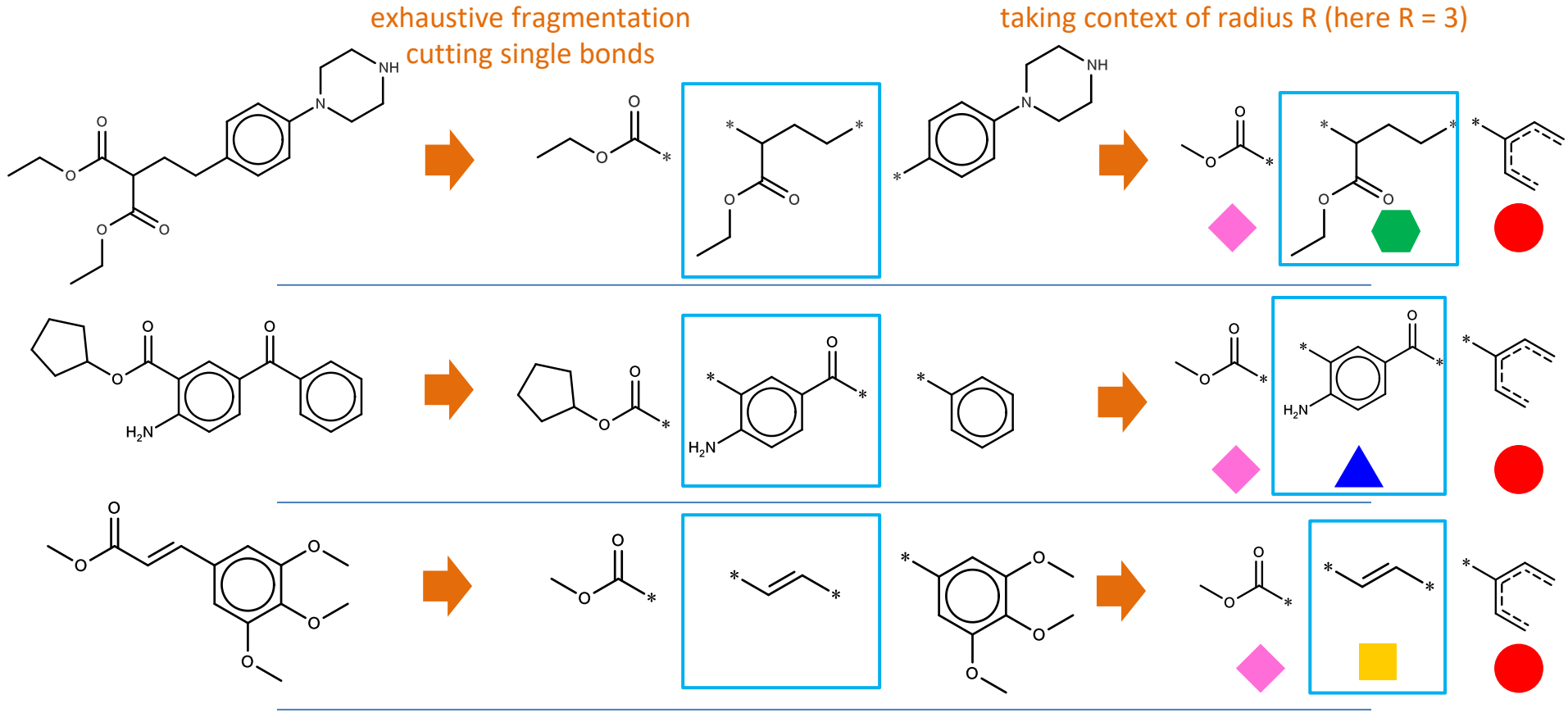
Only structure-based approaches are applicable: **molecular docking** and **dynamics**



Round 1: strategy 1 (de novo design)



Chemically reasonable mutations (CReM)



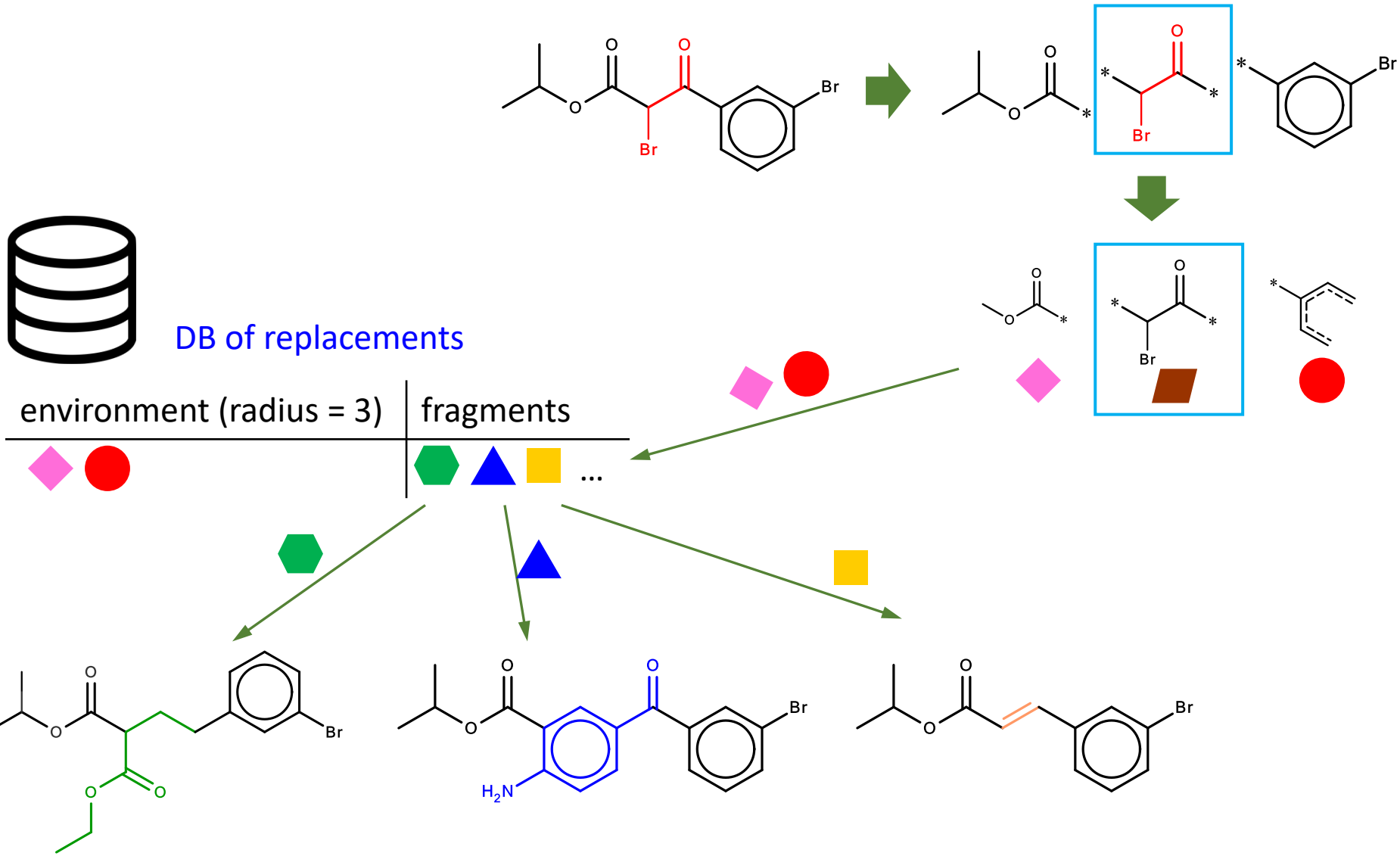
DB of replacements

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

interchangeable fragments

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

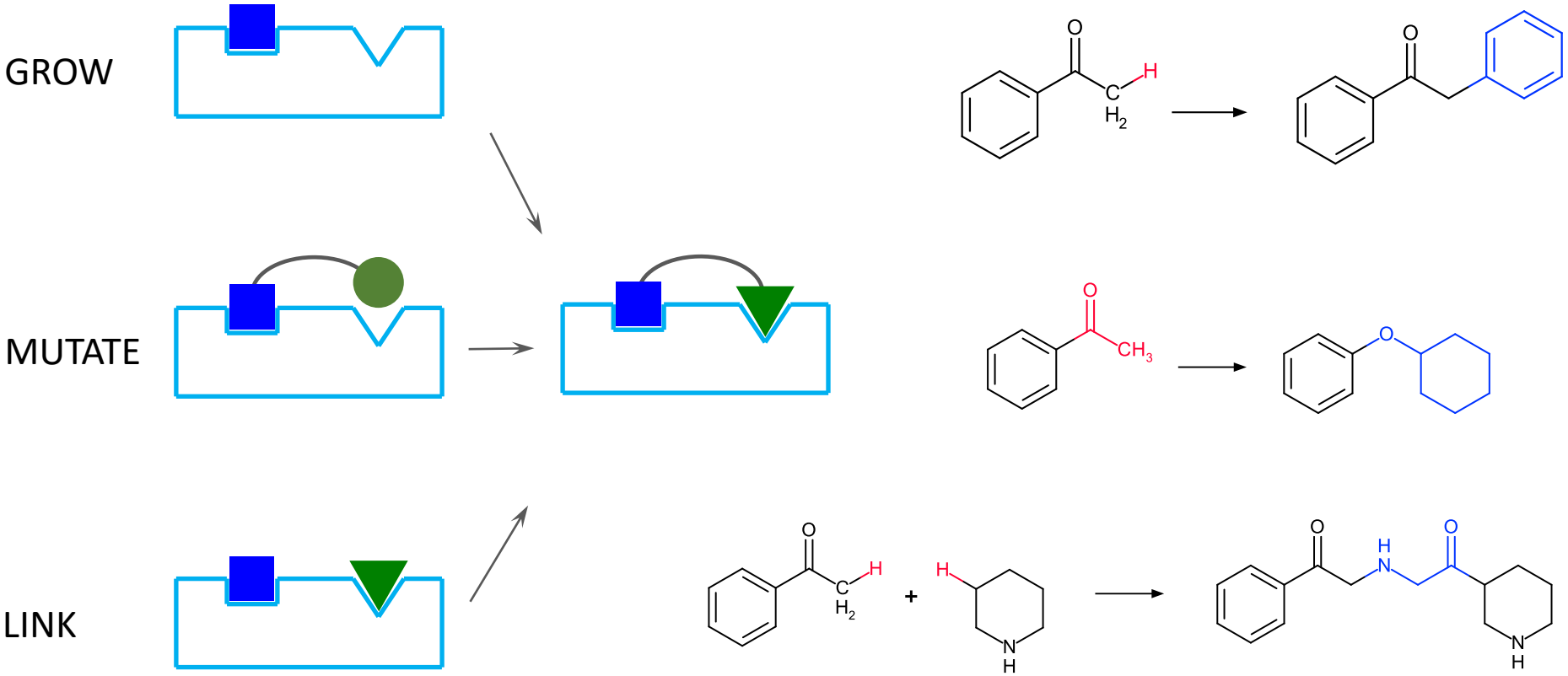
Chemically reasonable mutations (CReM)



Generated structures are always chemically valid!

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

Chemically reasonable mutations (CReM)



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

Tweak synthetic accessibility within CReM

Content of fragmented library

all ChEMBL compounds (1 554 160)

compounds with SA score ≤ 2.5 (572 527)

compounds with SA score ≤ 2 (107 806)

Context radius

1 less conservative replacements

2

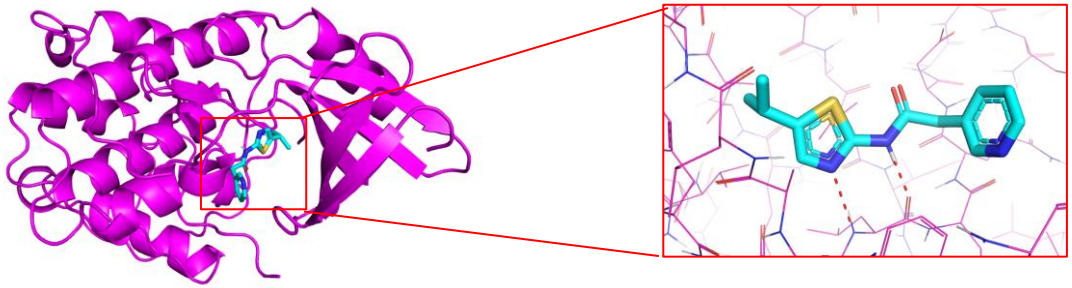
3

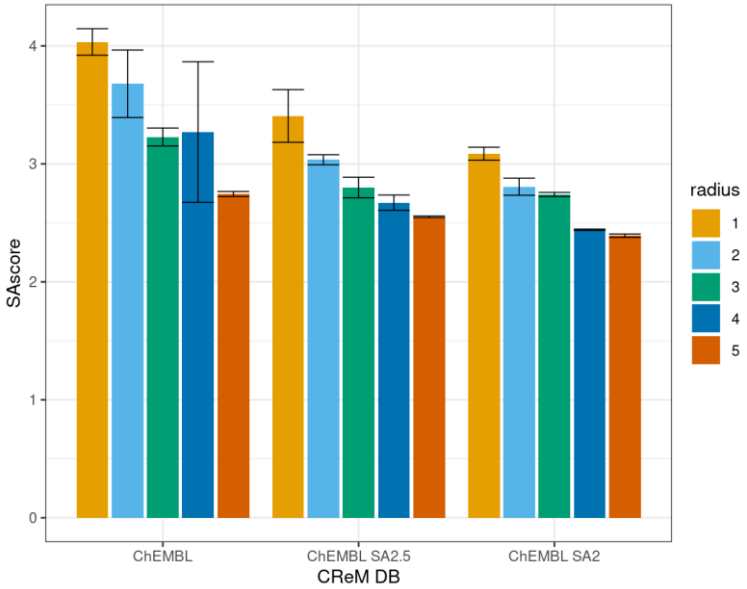
4 more conservative replacements

5

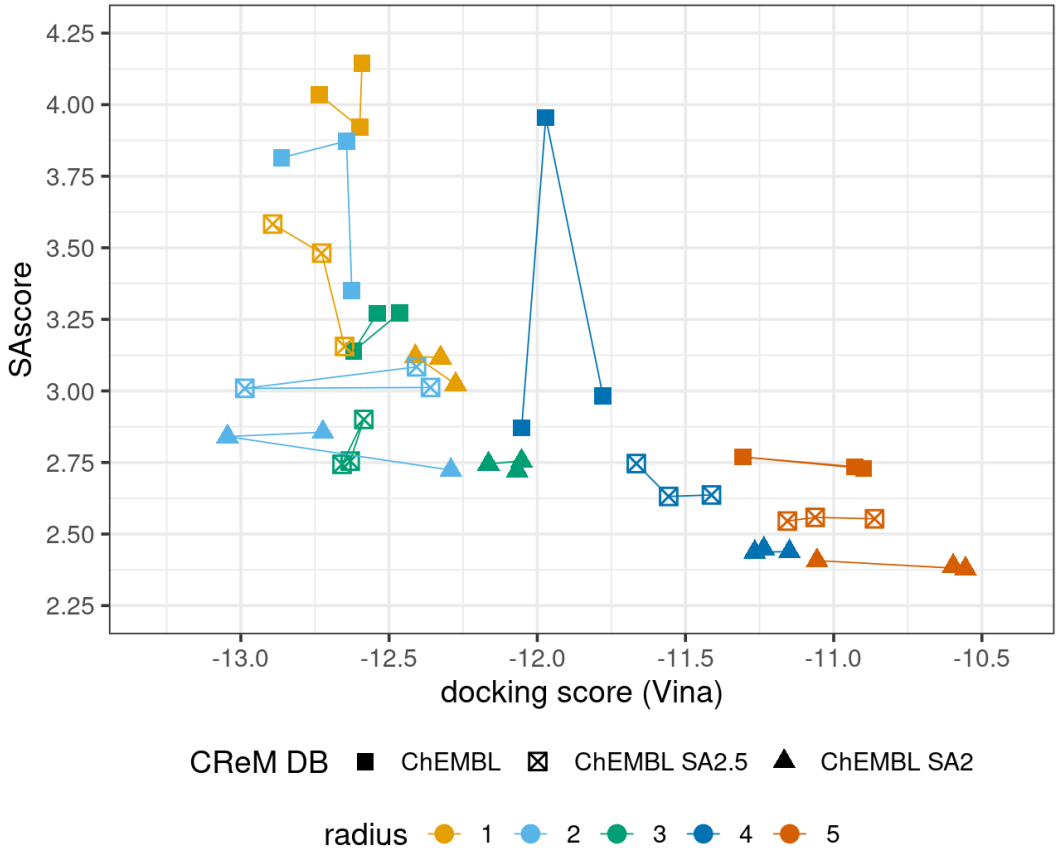
De novo design using docking (example)

2BTR
IC_{50} = 95 nM
docking score = -7.86





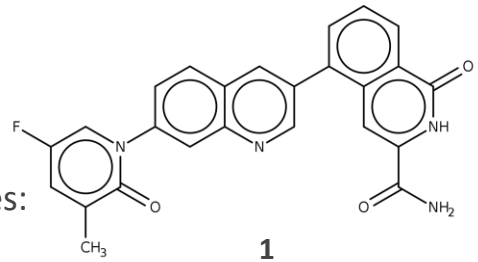
Average docking and SA scores for top 100 molecules from each run



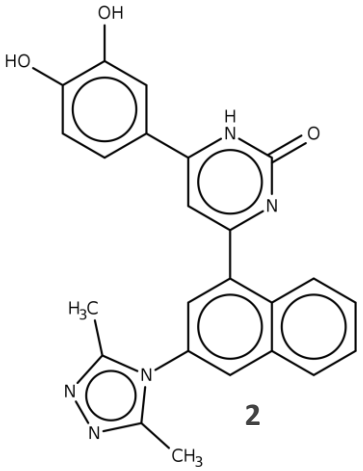
De novo design using docking (example)

Constant conditions:
- hinge region binding
- ChEMBL SA2
- radius 2

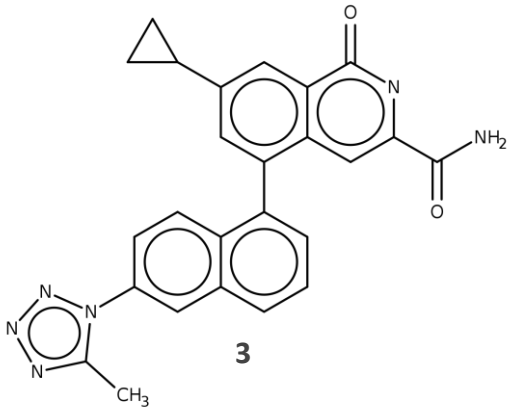
Variable conditions:
different CDK2 complexes:
- 2BTR
- 2FVD
- 3RAL
- 6GUH



1

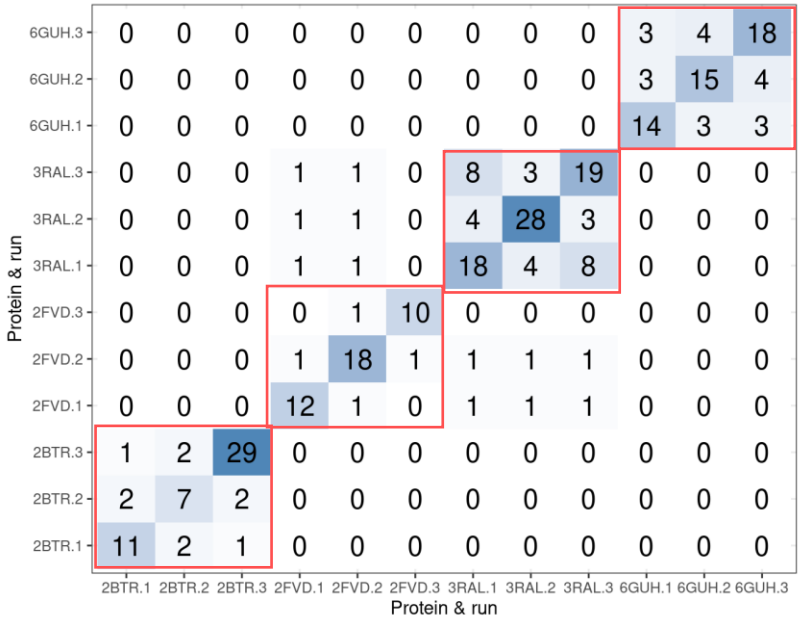


2

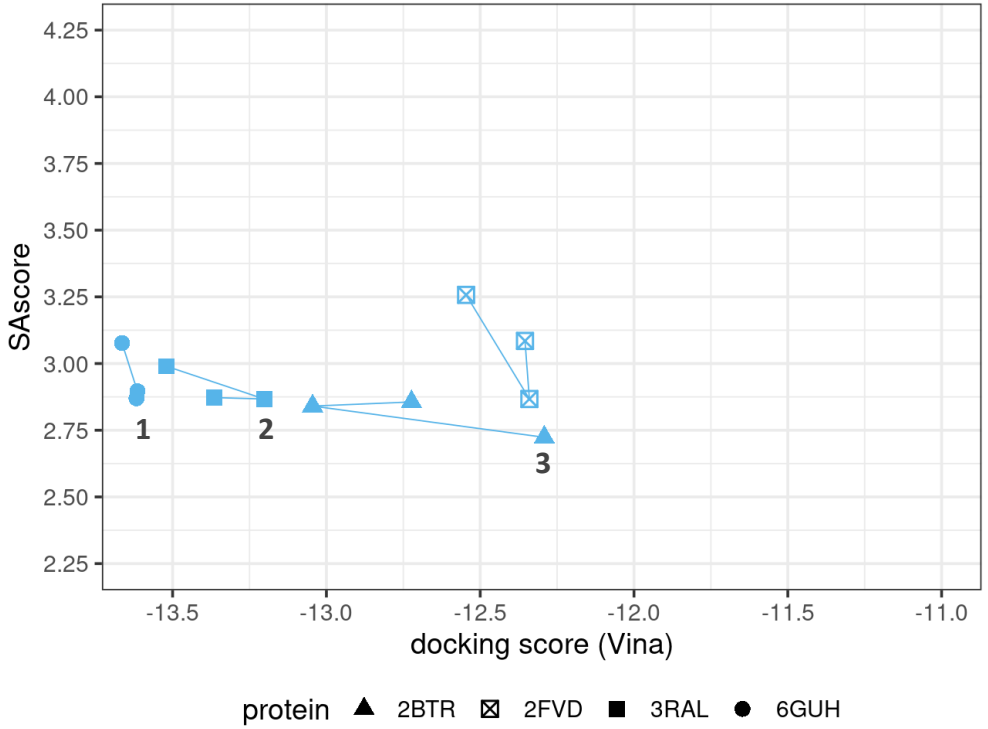


3

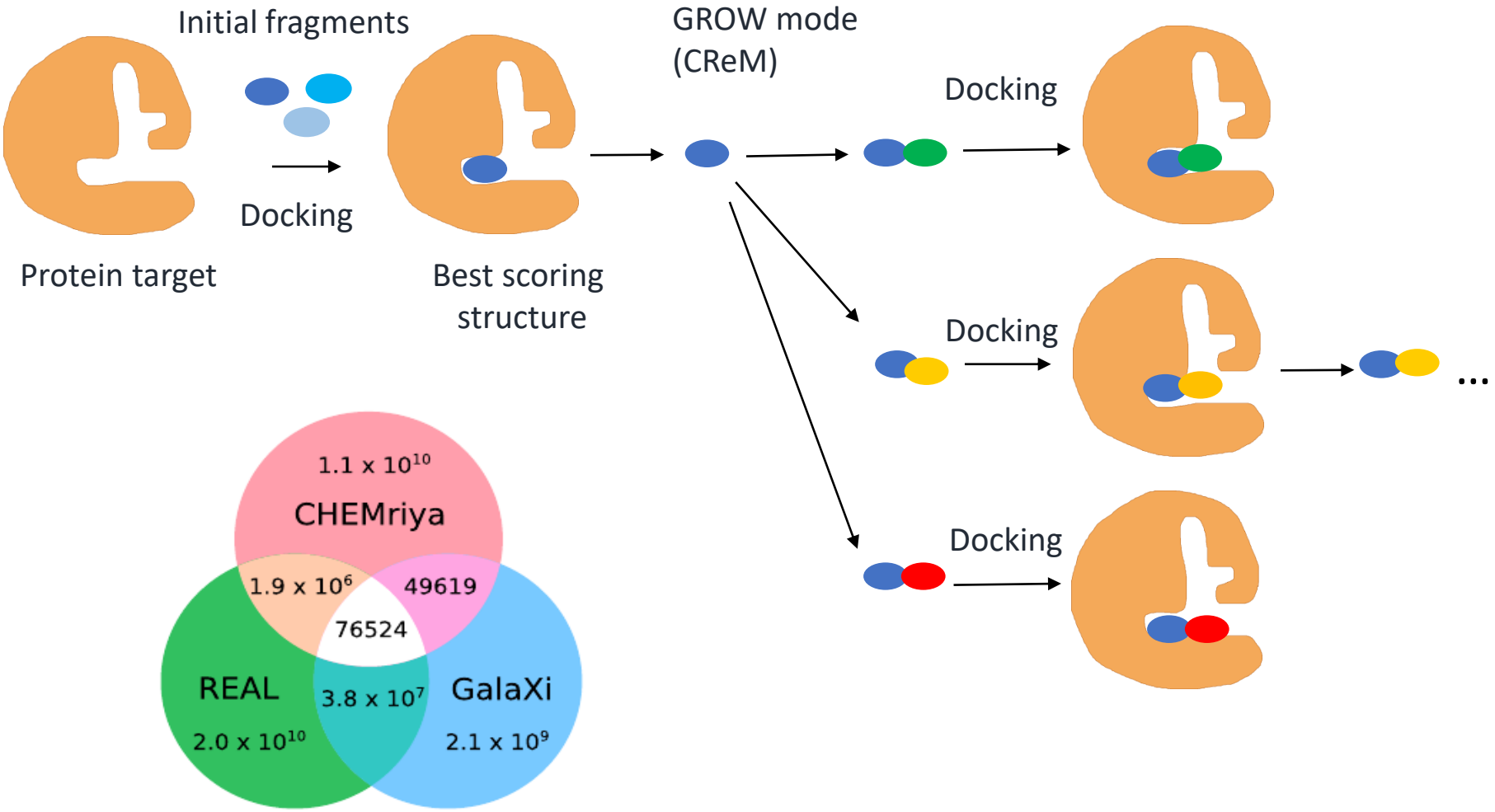
The number of distinct Murcko scaffolds in top 100 scored compounds in different runs and their intersection across runs



Average docking and SA scores for top 100 molecules from each run

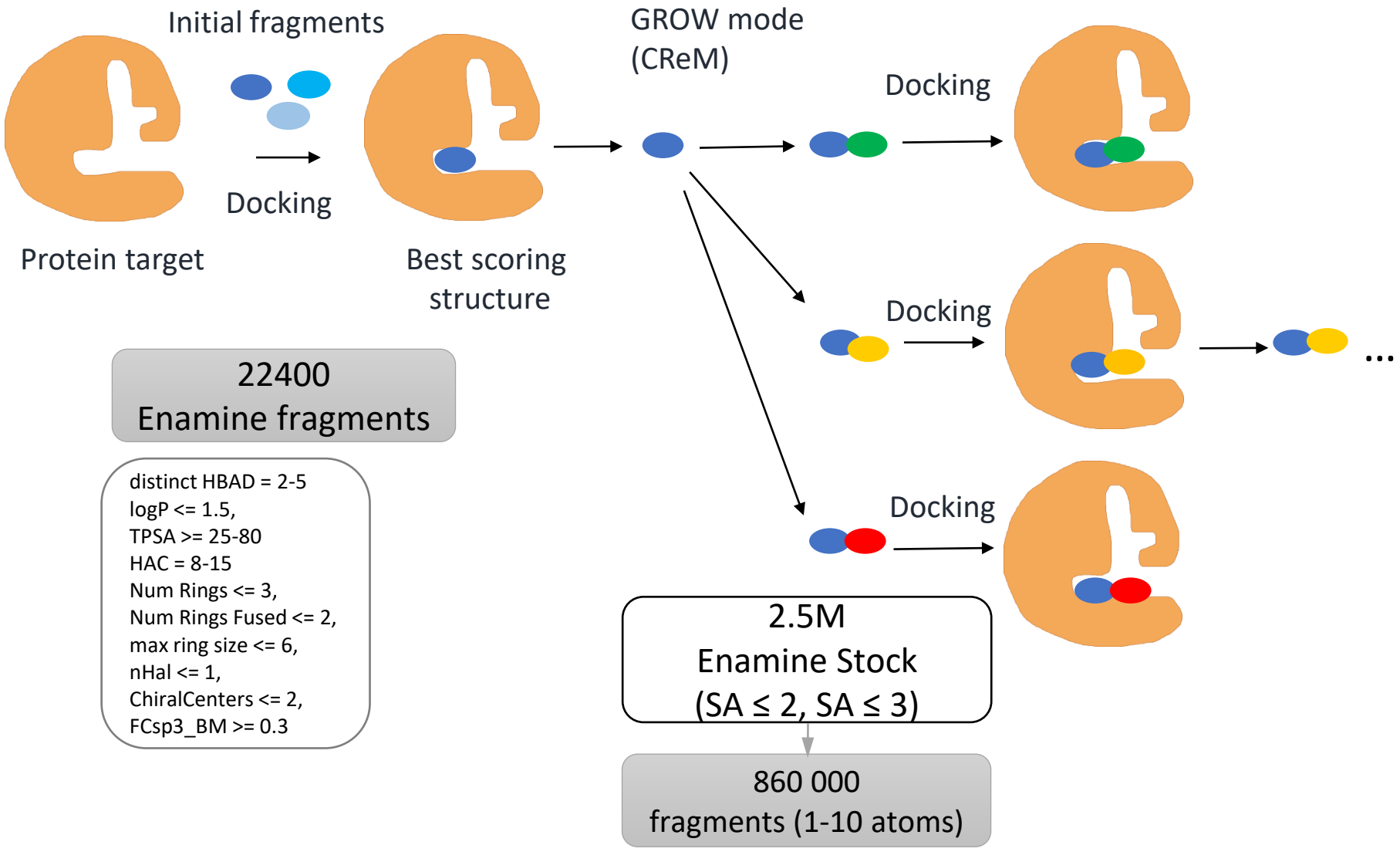


Round 1: strategy 1 (de novo design)

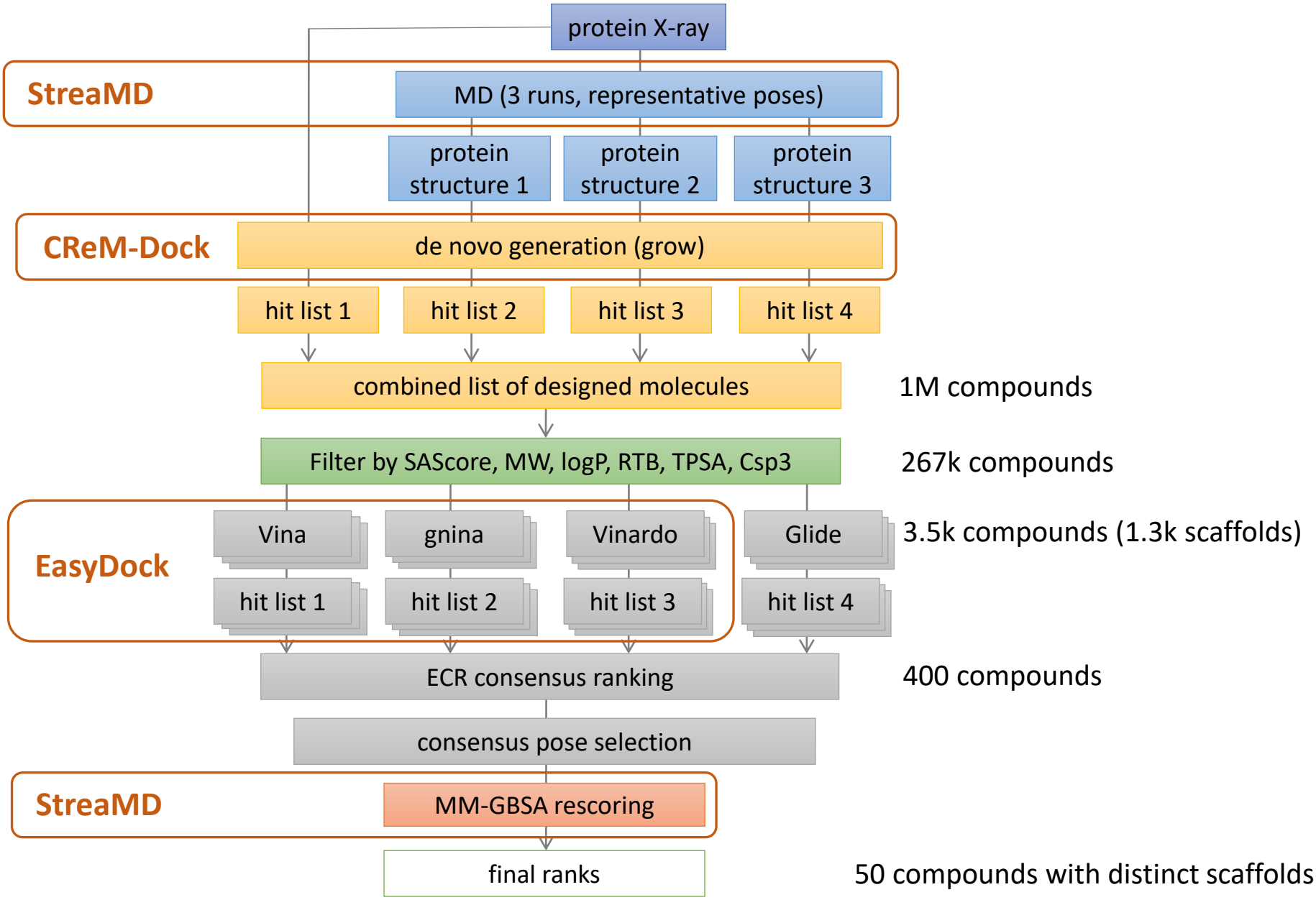


Bellmann, L.; Penner, P.; Gastreich, M.; Rarey, M., Comparison of Combinatorial Fragment Spaces and Its Application to Ultralarge Make-on-Demand Compound Catalogs. *J. Chem. Inf. Model.* **2022,** 62 (3), 553-566.

Round 1: strategy 1 (de novo design)



Round 1: strategy 1 (de novo design)



Round 1: strategy 1 (de novo design)

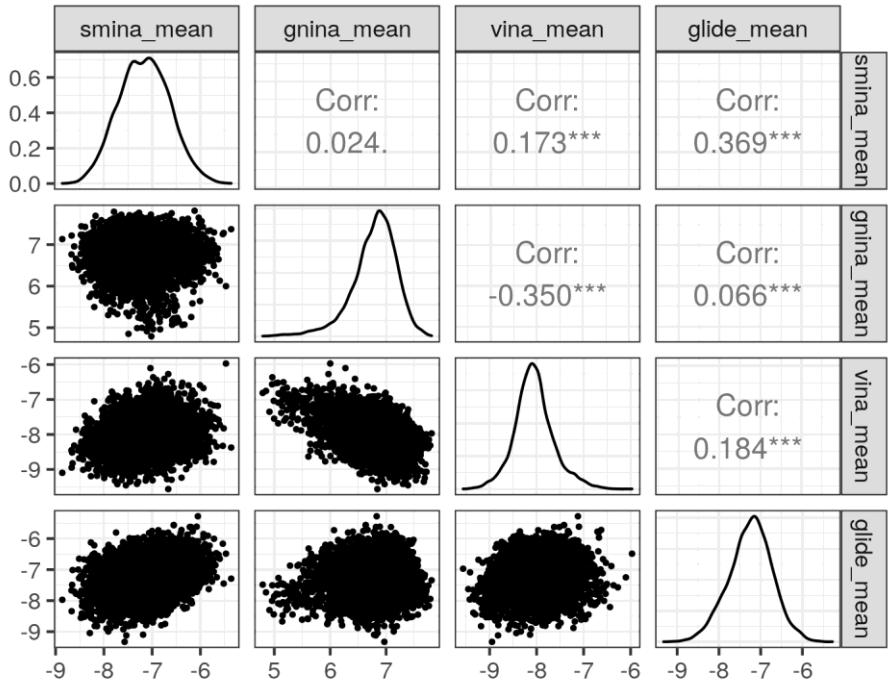
protein X-ray

StreaMD

MD (3 runs, representative poses)

CReM-Dock

hit lis



ompounds

compounds

compounds (1.3k scaffolds)

EasyDock

Vina

hit list

ompounds

consensus pose selection

StreaMD

MM-GBSA rescoring

final ranks

50 compounds with distinct scaffolds

Round 1: strategy 1 (de novo design)

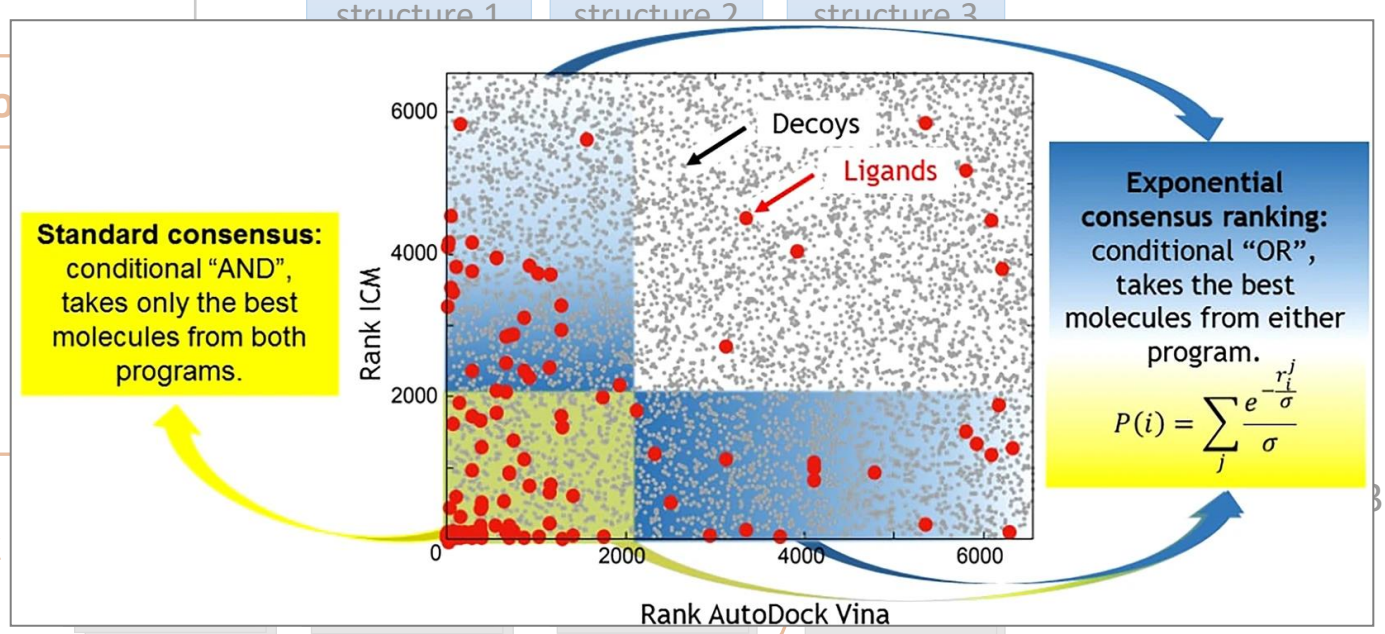
protein X-ray

StreaMD

MD (3 runs, representative poses)

protein structure 1 | protein structure 2 | protein structure 3

CReM-Do



EasyDock

3k scaffolds)

ECR consensus ranking

400 compounds

consensus pose selection

StreaMD

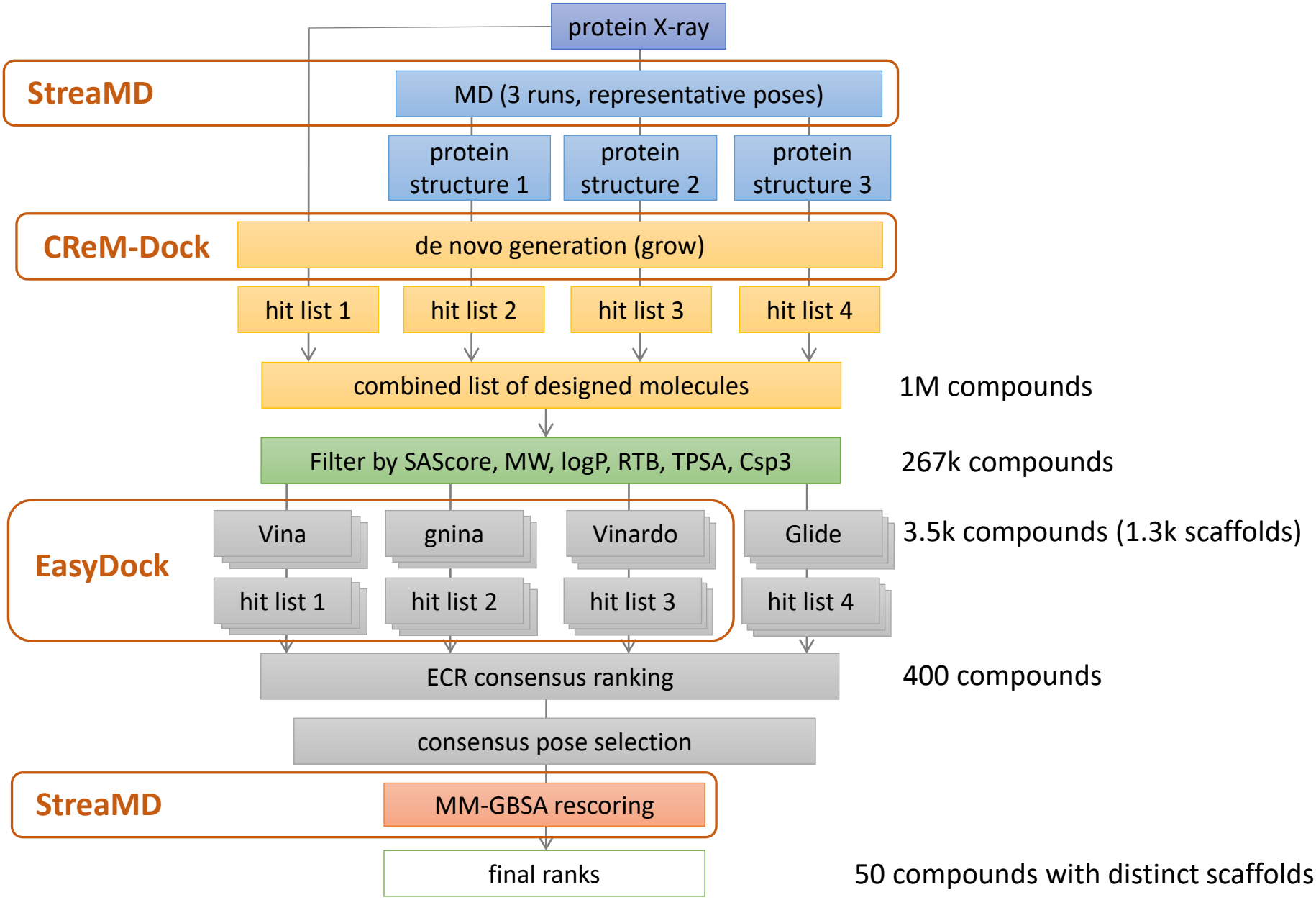
MM-GBSA rescoring

final hits

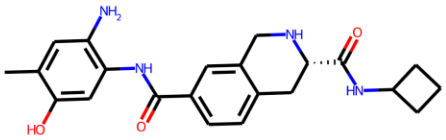
50 compounds with distinct scaffolds

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.

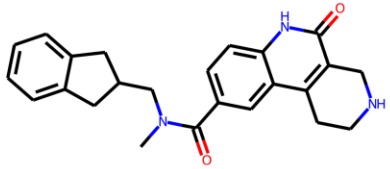
Round 1: strategy 1 (de novo design)



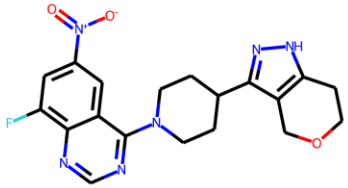
Round 1: strategy 1 (de novo design)



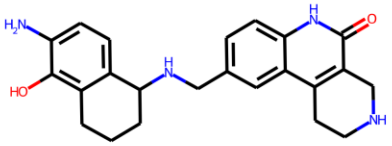
CREM0402551



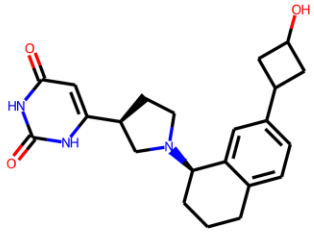
CREM0978670



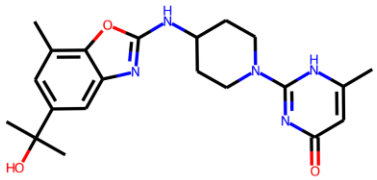
CREM1515848



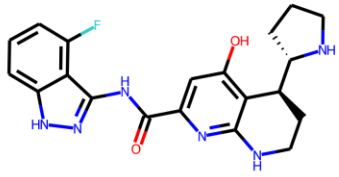
CREM1480106



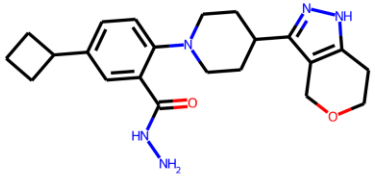
CREM1777121



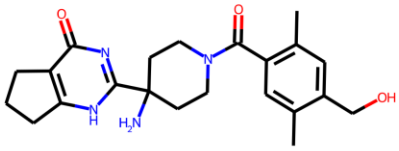
CREM0329741



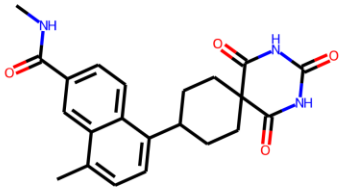
CREM1661038



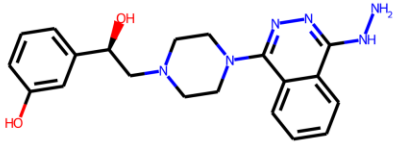
CREM1506273



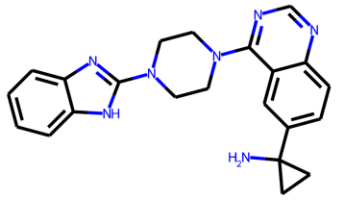
CREM0340409



CREM1089720



CREM1507777



CREM1468894

50 de novo compounds

SA score < 3

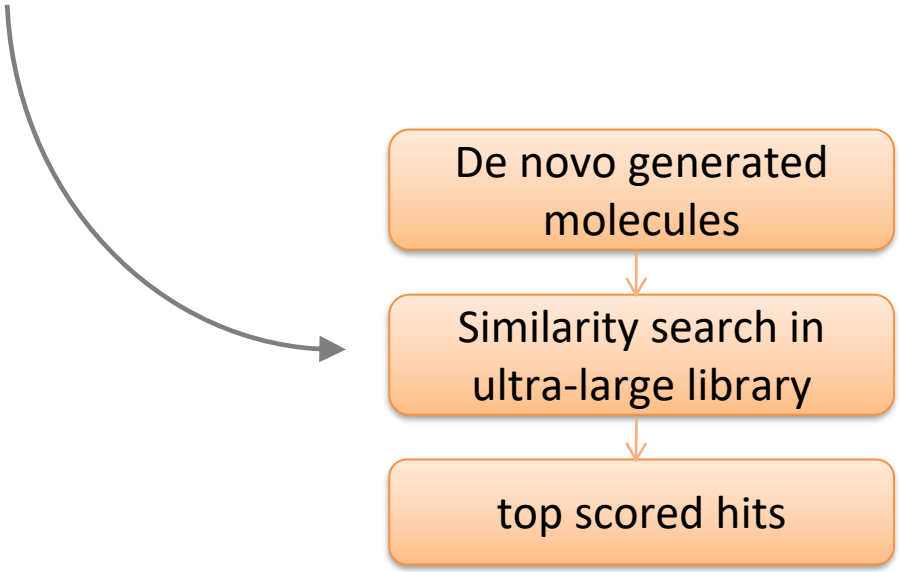
11 reconstructed retrosynthetic pathways with AiZynthFinder (2-5 steps)

Round 1: strategy 2 (similarity search)

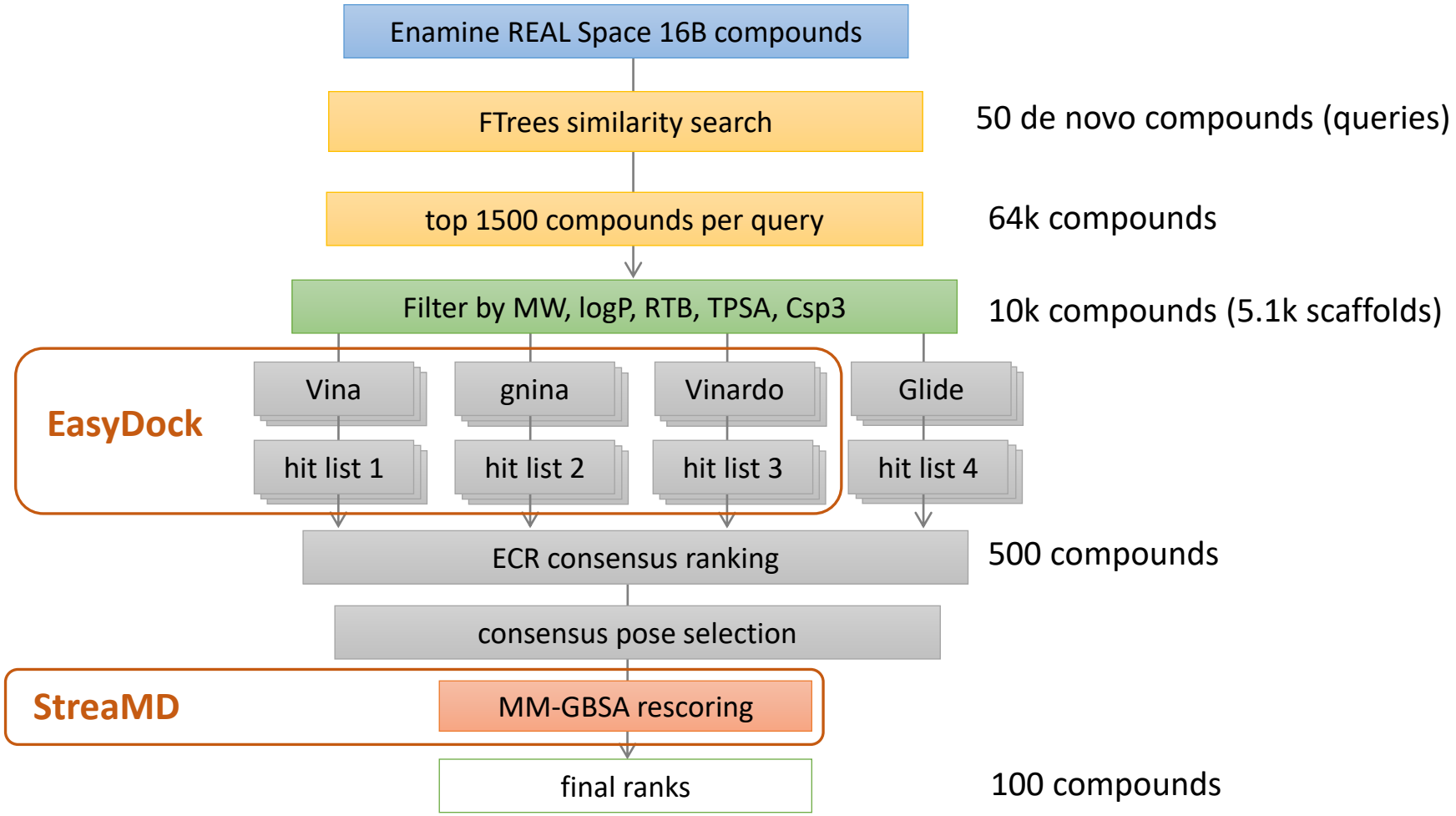
Enamine Real Space: 16B

Docking of a whole ultra-large library (>10 B compounds) is extremely expensive

(if one docking takes 1 sec, it will take 317 years on a single core)

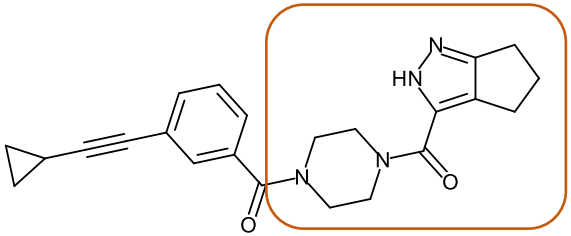


Round 1: strategy 2 (similarity search)

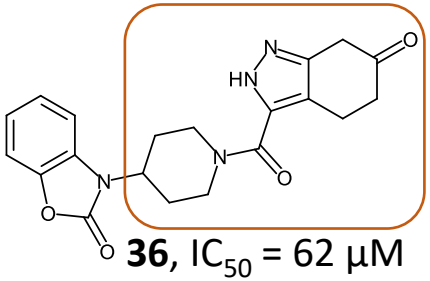


Round 1: experimental results

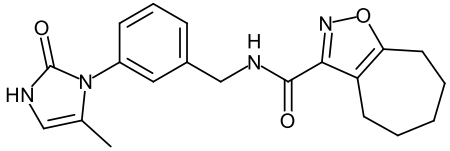
50 de novo + 100 similar compounds
91 compounds were selected (within the budget 9000$)
82 compounds were synthesized
8 compounds demonstrated activity (K_d = 25-117 µM by SPR)



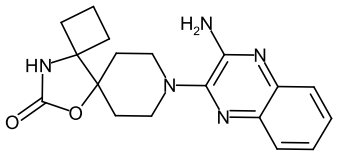
1, IC_{50} = 61 µM



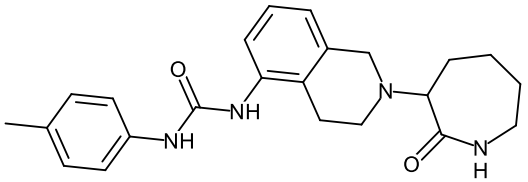
36, IC_{50} = 62 µM



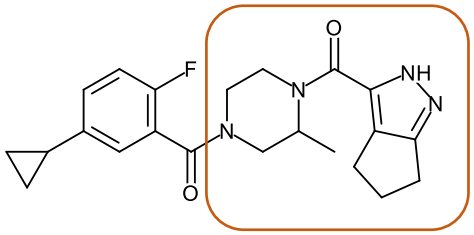
59, IC_{50} = 32 µM



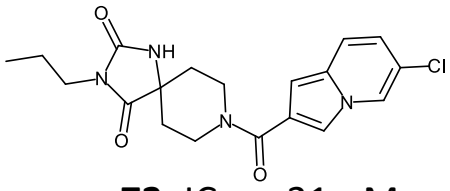
62, IC_{50} = 25 µM



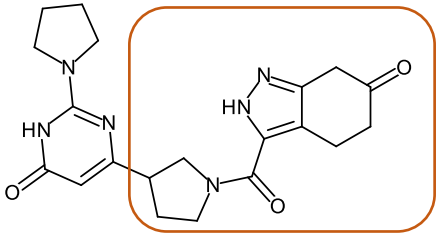
65, IC_{50} = 56 µM



69, IC_{50} = 117 µM

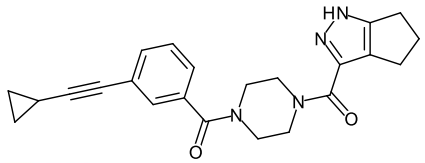


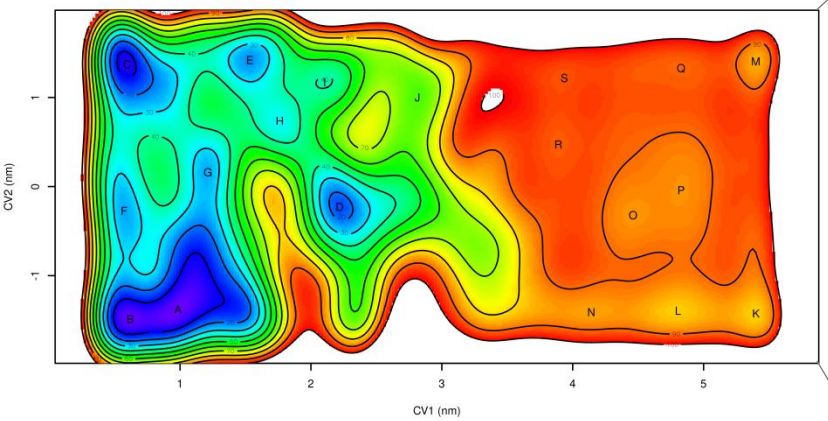
73, IC_{50} = 31 µM

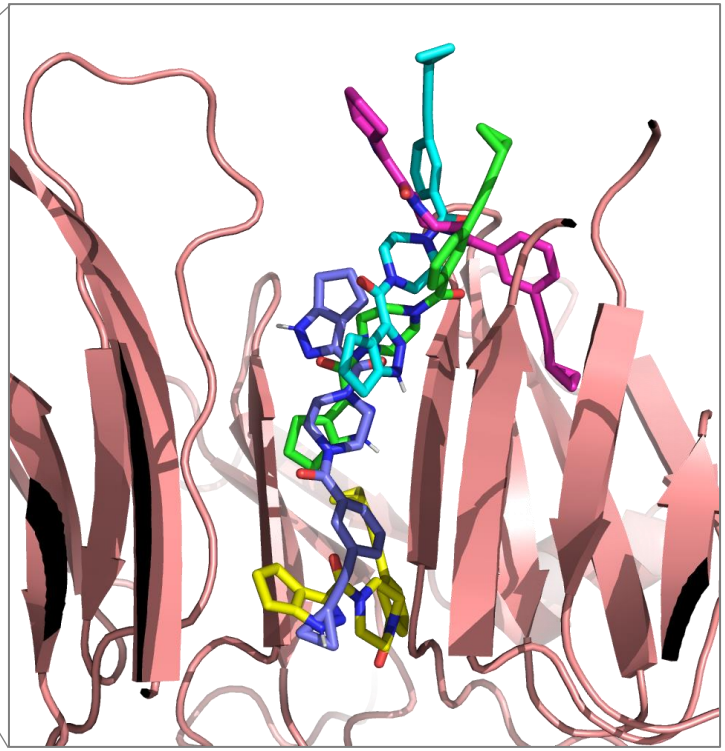


76, IC_{50} = 74 µM

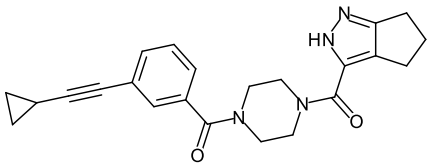
Round 2: hit optimization (metadynamics)

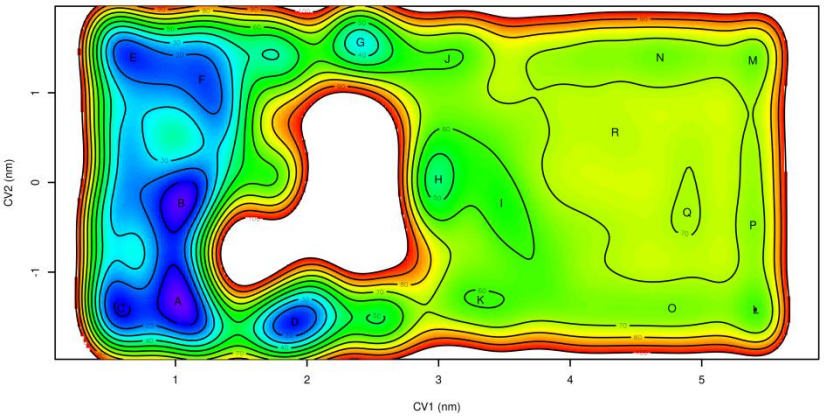


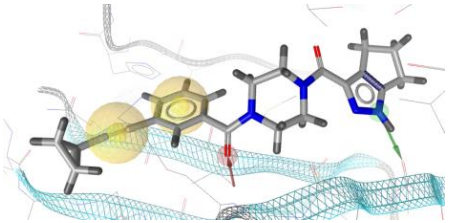


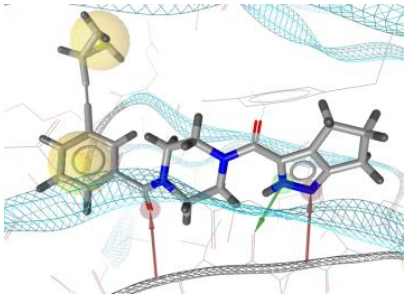


A
B
C
D
E



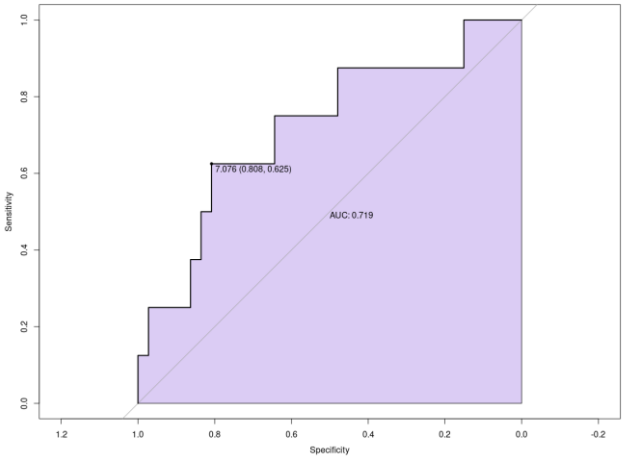




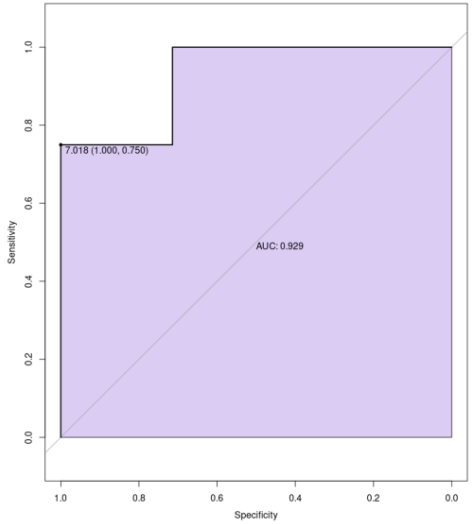


Round 2: hit optimization (compound pool 1)

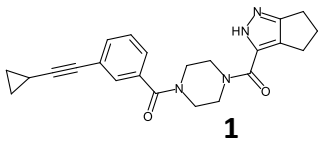
8 actives / 74 inactives (82 compounds)

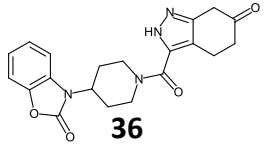


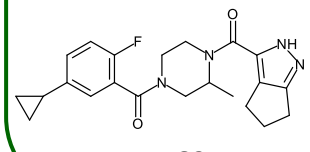
4 actives / 8 inactives (12 pyrazol derivatives)

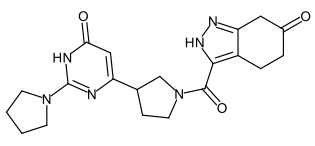


actives

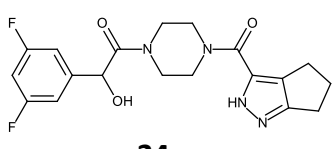


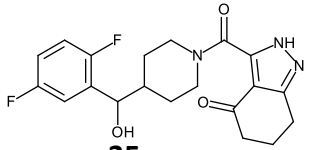


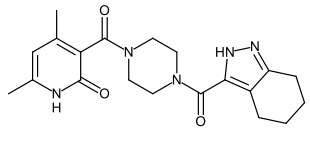


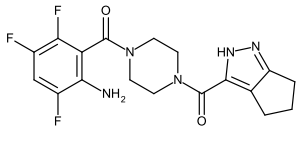


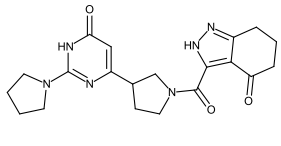
inactives

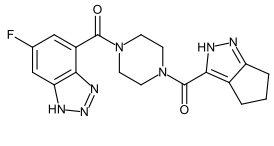


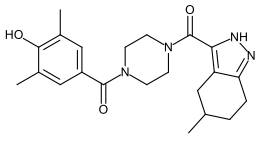






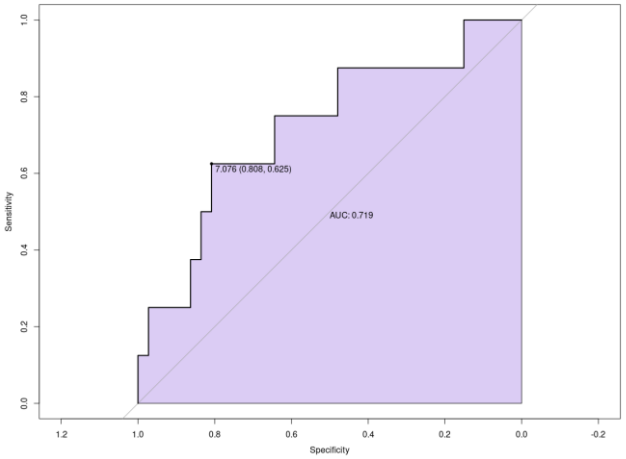




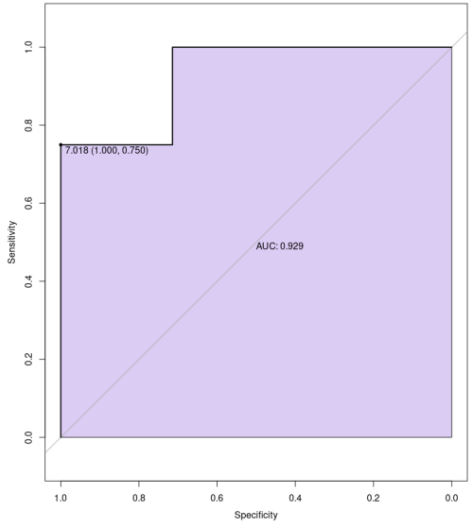


Round 2: hit optimization (compound pool 1)

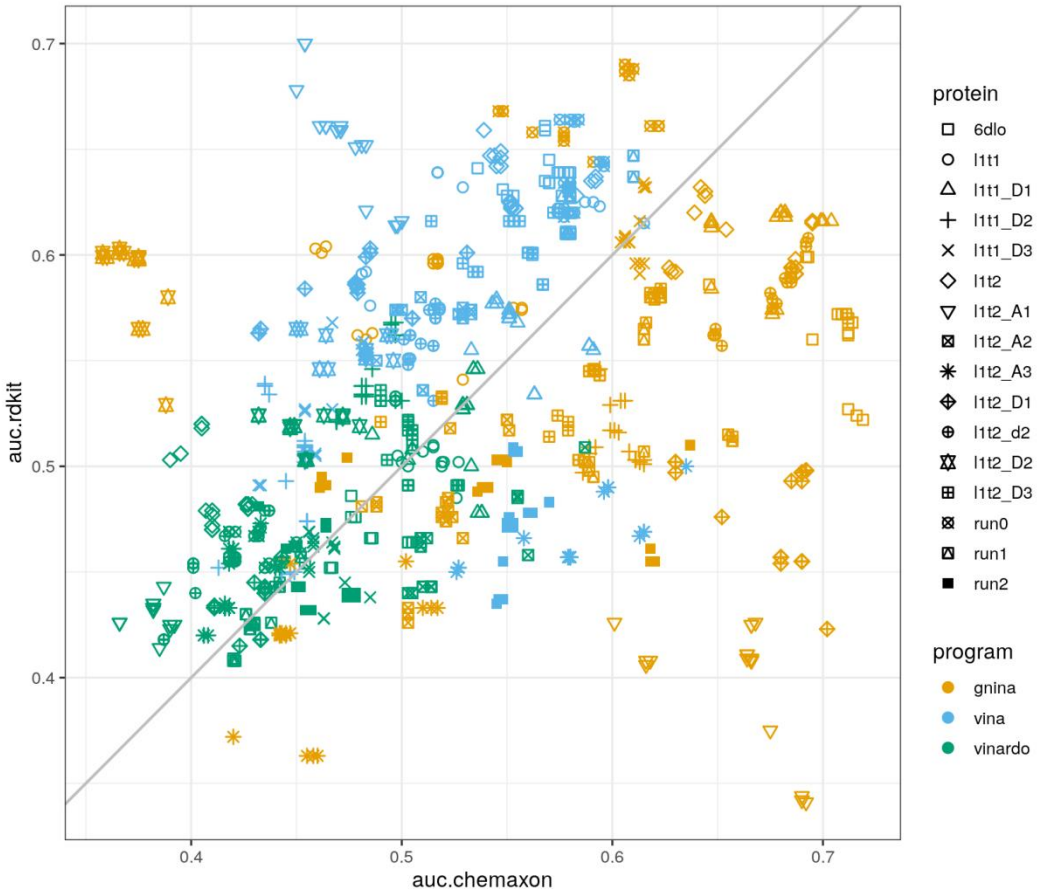
8 actives / 74 inactives (82 compounds)



4 actives / 8 inactives (12 pyrazol derivatives)

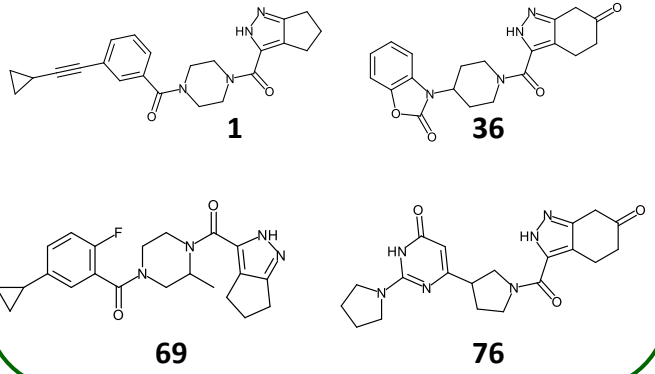


Initial conformer generation – RDKit vs. Chemaxon

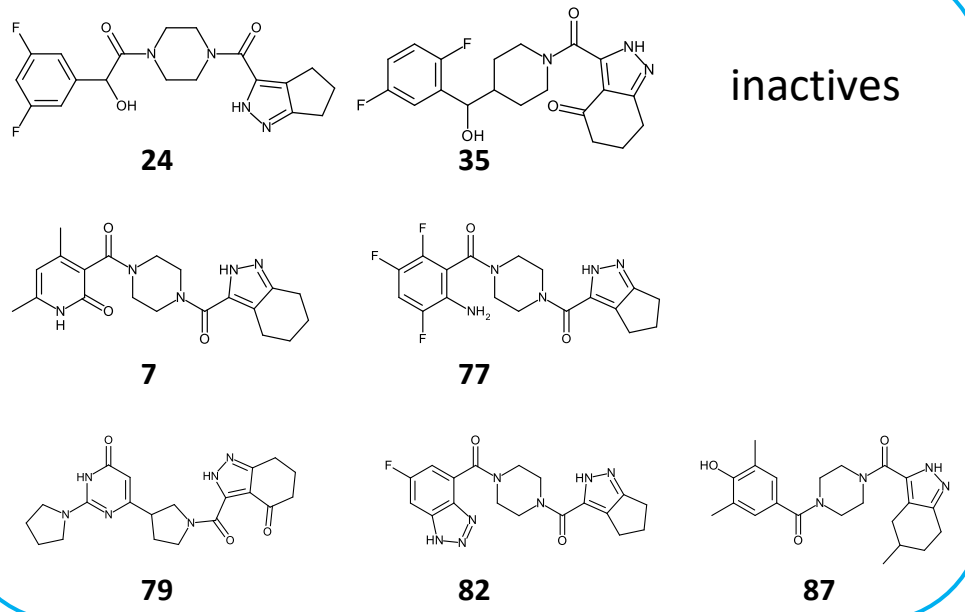


Round 2: hit optimization (compound pool 1)

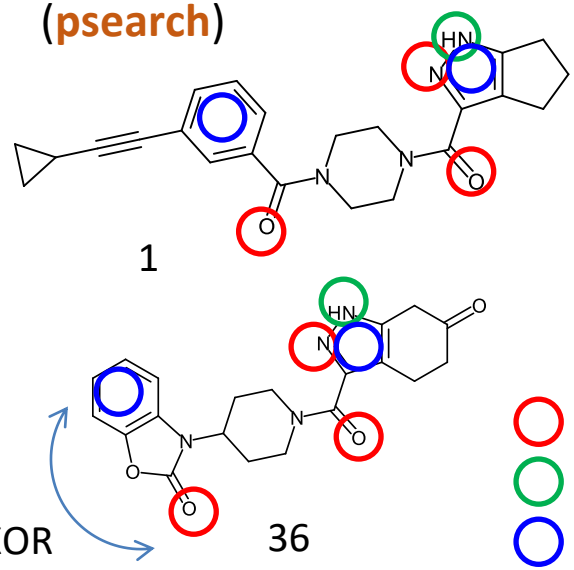
actives



inactives



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)

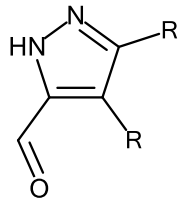
2.5M
Enamine Stock

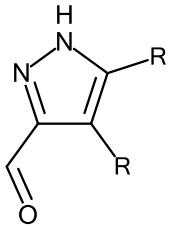
↓

substructure search

↓

18 411 compounds

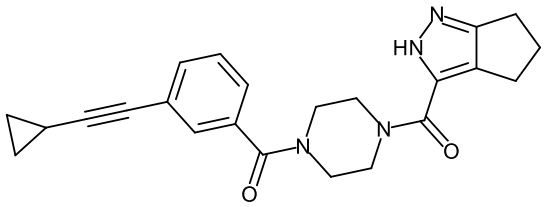




chemicalite-scripts

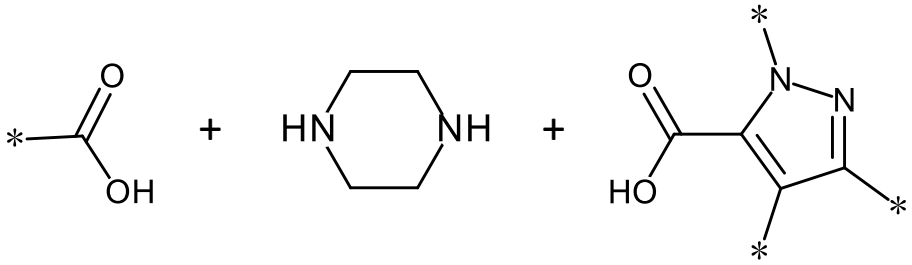
Round 2: hit optimization (compound pool 3)

Enamine fragments → substructure search → 18 845 building blocks



1, IC_{50} = 61 µM

Enamine fragments → substructure search → 474 building blocks

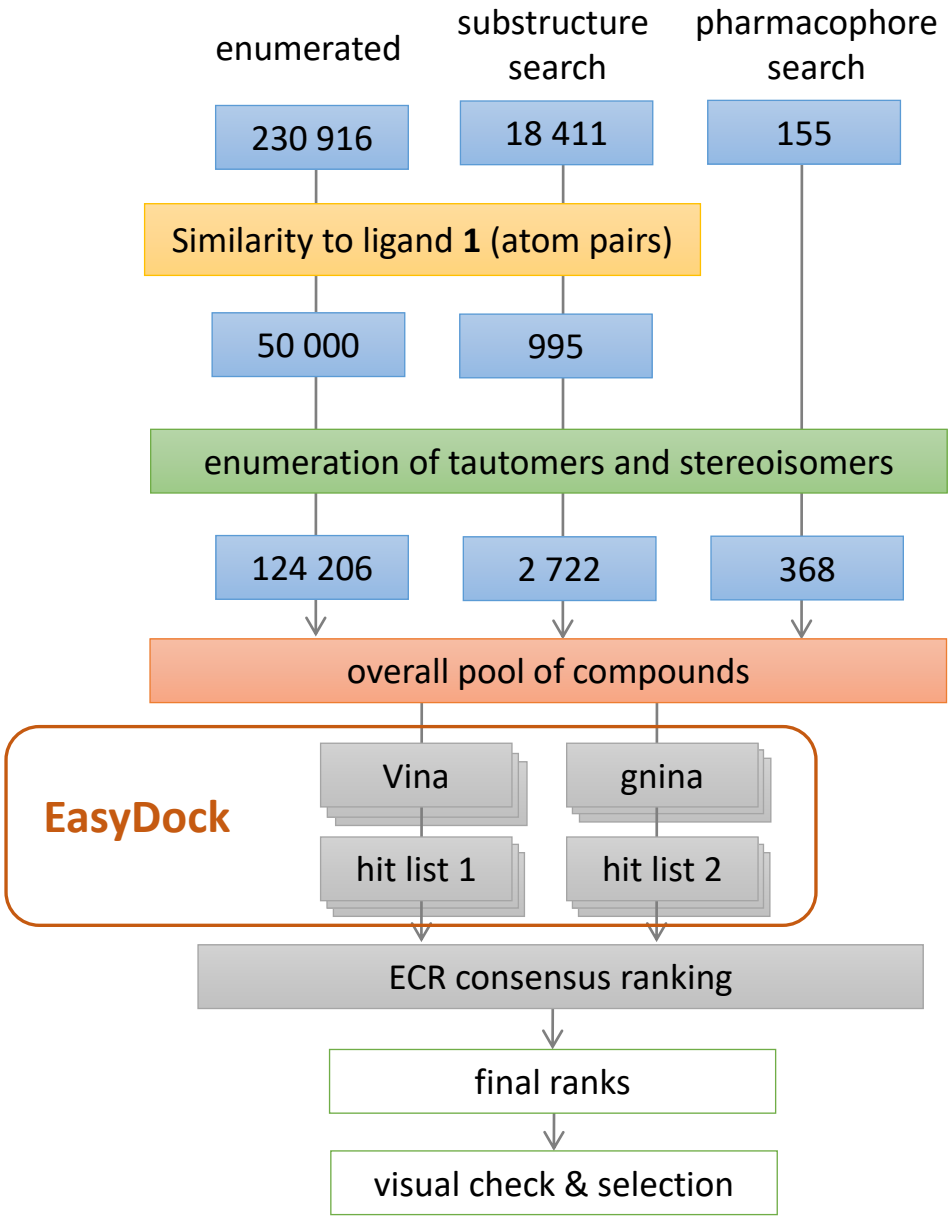


2 943 486 enumerated molecules

Filter by MW, logP, TPSA, RTB, Csp3

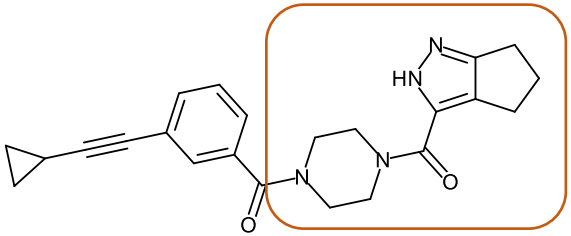
230 916 compounds

Round 2: hit optimization (screening pipeline)

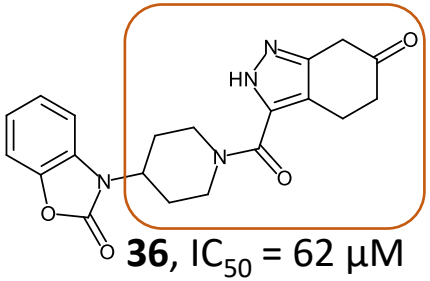


Round 2: hit optimization (experimental results)

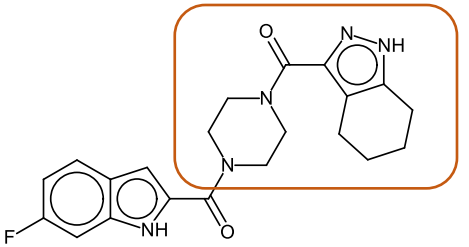
38 compounds were selected (within the budget 4500$)
35 compounds were synthesized
4 compounds demonstrated dose-response effect in SPR
1 scaffold had confirmed selectivity



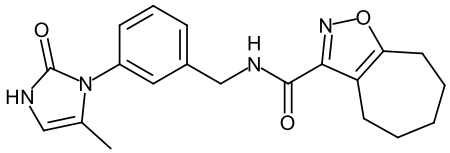
1, IC_{50} = 61 µM



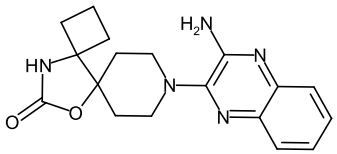
36, IC_{50} = 62 µM



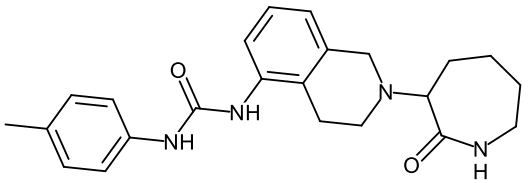
HO-15, IC_{50} = 71 µM



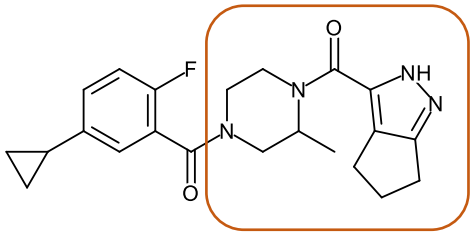
59, IC_{50} = 32 µM



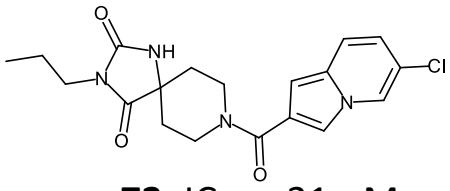
62, IC_{50} = 25 µM



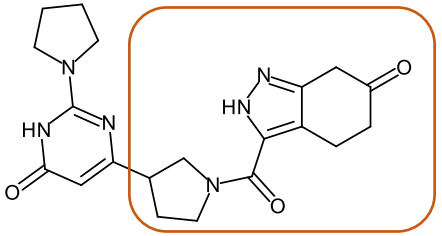
65, IC_{50} = 56 µM



69, IC_{50} = 117 µM



73, IC_{50} = 31 µM



76, IC_{50} = 74 µM

IMTM

Overall statistics of all groups

Round1 compounds	Round1 hits	Round2 compounds	Round2 SPR hits	Selective scaffolds confirmed in orthogonal methods	Promising chemical series
72	4	23	3	2	1
84	2	33	10	2	1
84	10	44	9	1	1
82	8	35	4	1	1
59	7	37	11	1	1
94	5	32	8	1	
92	4	39	6	1	
113	3	49	6	1	1
37	2	47	7	1	1
101	1	38	5	1	
98	3	46	4	0-2	
99	11	47	3	0	
100	4	49	3	0	
100	2	41	8	0	
105	2	25	1	0	
65	2	44	4	0	
91	2	36	4	0	
101	1	49	4	0	
79	0	0	0	0	
95	0	0	0	0	
71	0	0	0	0	
83	0	0	0	0	
50	0	0	0	0	

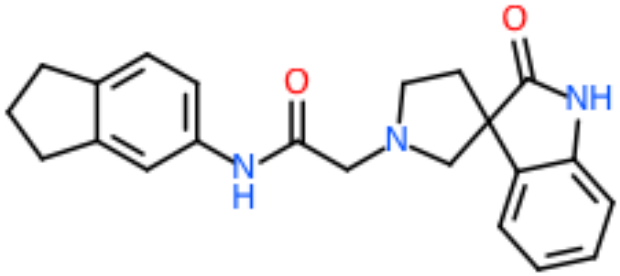
IMTM

Final ranking

Participant	Participant ID	Aggregated score	Computational Method
David Koes, University of Pittsburgh	1181	18	Link
Olexandr Isayev, Carnegie Mellon University & Artem Cherkasov, University of British Columbia	1209	18	Link
Christina Schindler, Merck KGaA	1193	17	Link
Dmitri Kireev, University of Missouri	1183	16	Link
Christoph Gorgulla, Harvard University	1195	16	Link
Didier Rognan, Université Strasbourg	1202	16	Link
Pavel Polishchuk, Palacky University	1210	16	Link

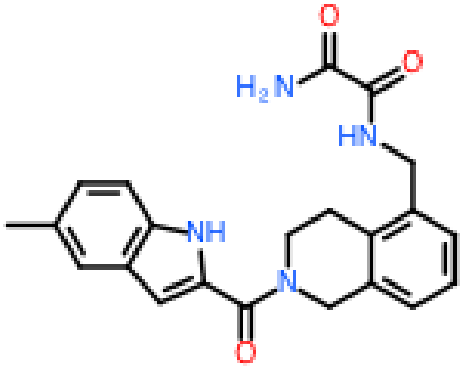
Examples of compounds selected by medchem experts

D,Koes



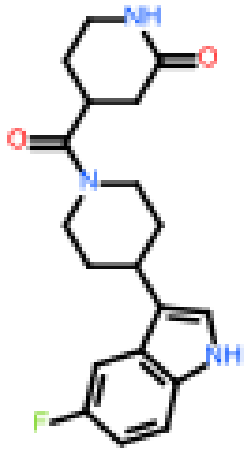
K_d = 123 µM

Isaev & Cherkasov



K_d = 11 µM

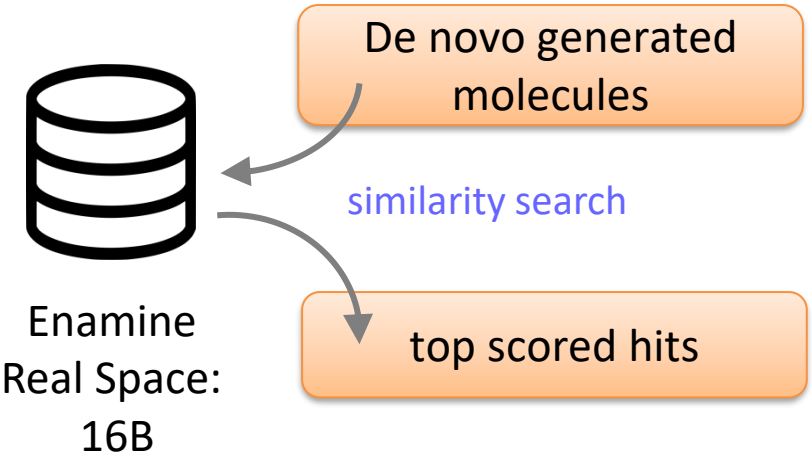
C. Schindler



K_d = 48 µM

Conclusions

- You should always have plan B, C, D...
- Unbiased *in silico* hit selection works (hit rate at Round 1 was almost 10%)
- The proposed strategy to search for hits in ultra-large libraries using similarity search guided by de novo designed compounds works



- The designed multi-step virtual screening pipeline which includes docking to multiple apo-protein structures, consensus scoring and re-scoring using MM-GBSA approach also works
- This project accelerated the development of new tools for automated docking (EasyDock) and molecular dynamics (StreaMD) which run on supercomputers. It allowed validate our de novo generation approach (CReM-Dock) and 3D ligand-based pharmacophore modeling tool (psearch) and FTrees tool for similarity search in large databases provided by BioSolvIT company.