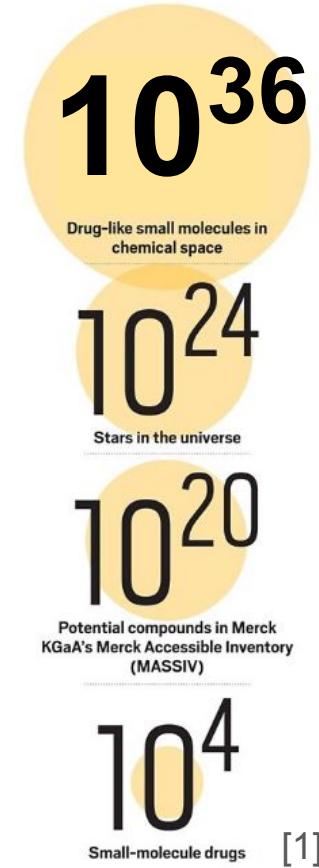
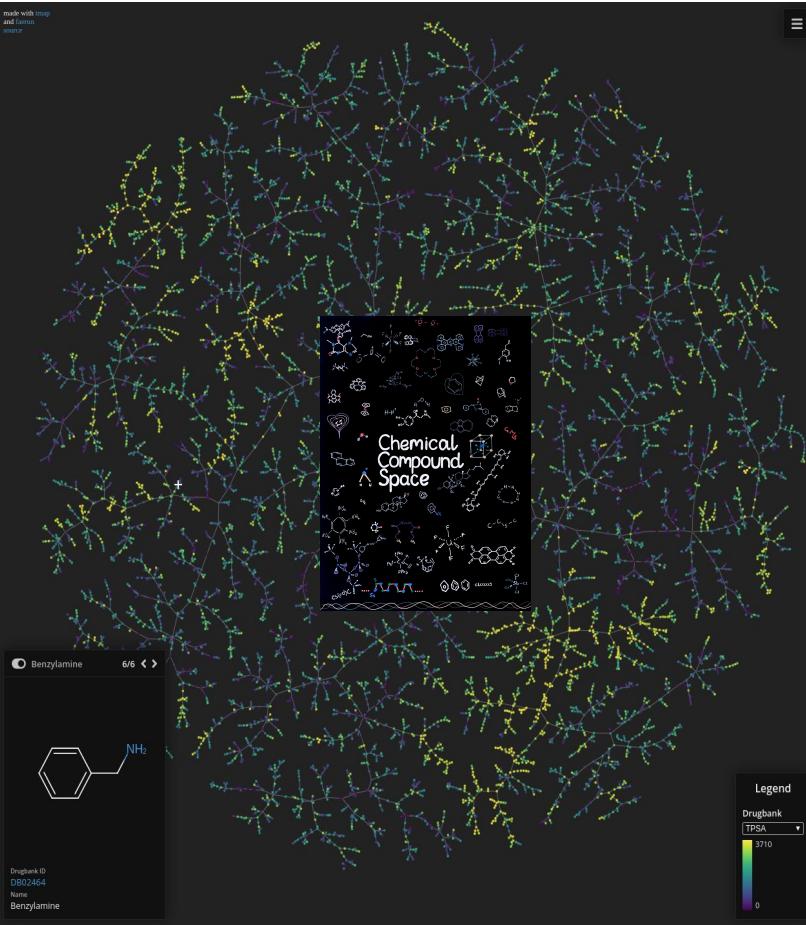


# GenCReM: de novo design guided by explainable docking

Aleksandra Ivanova<sup>1</sup>  
Pavel Polishchuk<sup>1</sup>

<sup>1</sup> Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University and University Hospital in Olomouc

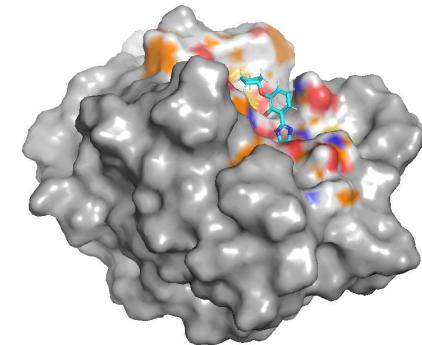


### \* Drug-like molecules:

- based on extrapolation on GDB-17<sup>[2]</sup> -  $10^{36}$
- based on stitching together up to 30 carbon, nitrogen, oxygen, and sulfur atoms in different arrangements -  $10^{63}$

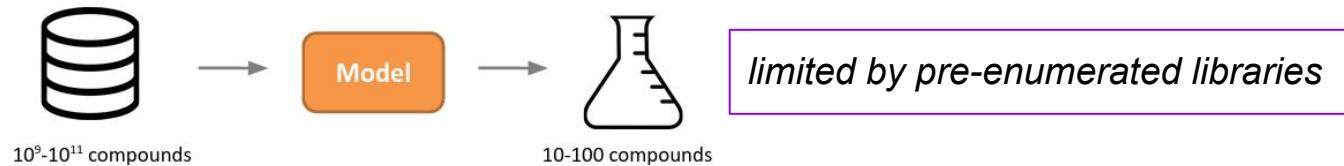


## Active compound?

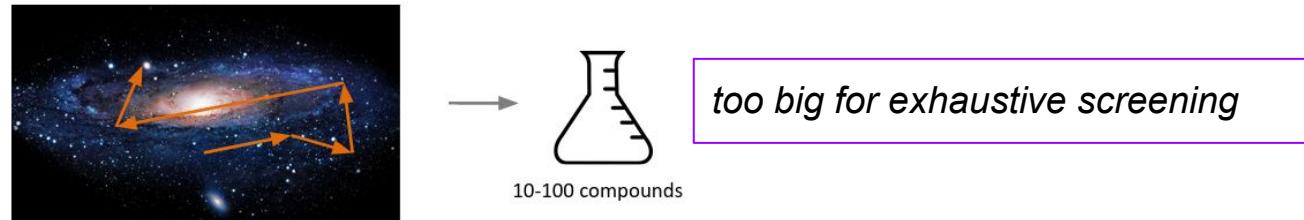


*looking for a needle in a haystack*

## Virtual screening

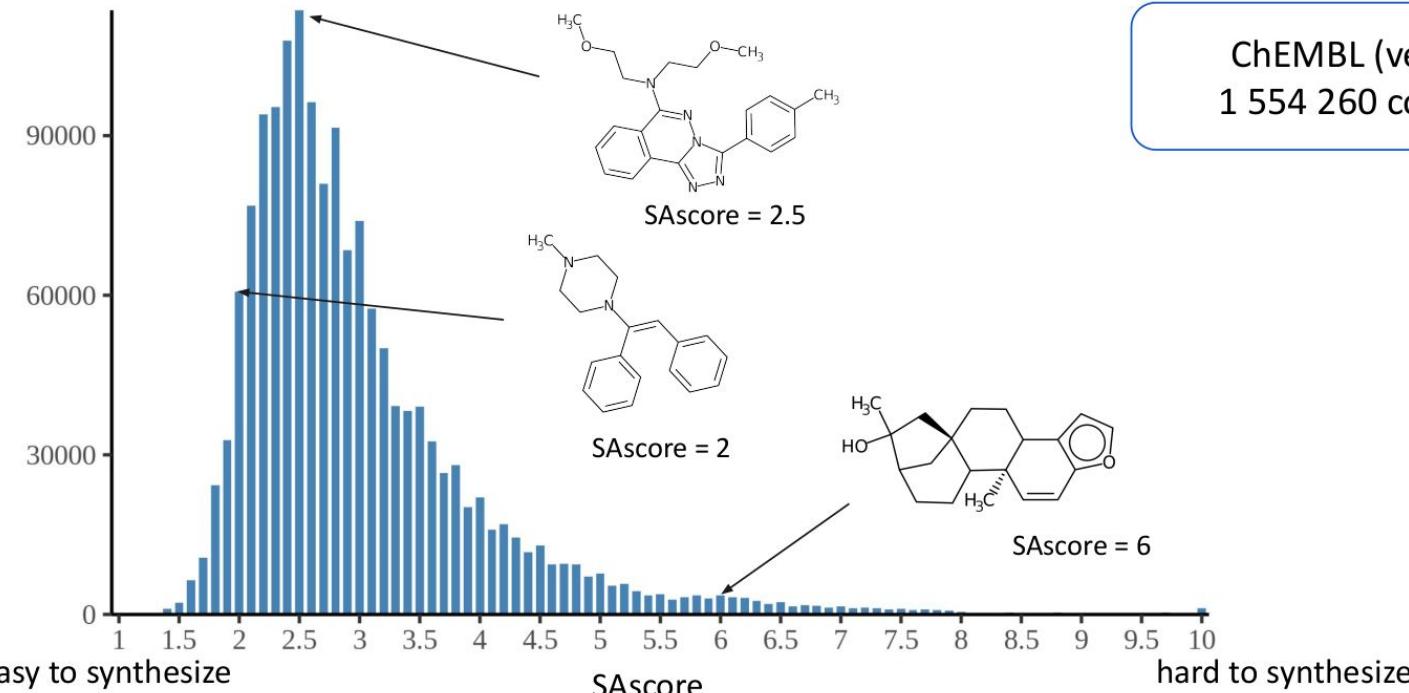


## De novo design



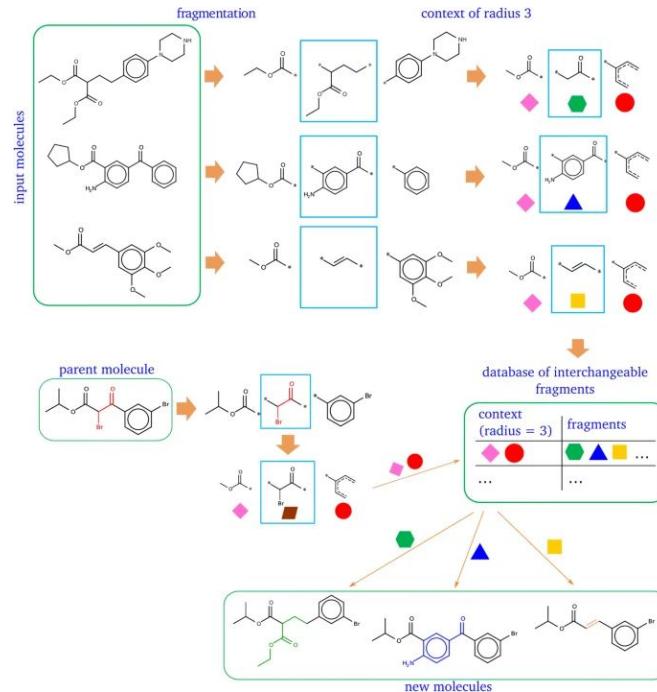
Need for **efficient, goal-driven exploration**

# Current Limitations in De Novo Design

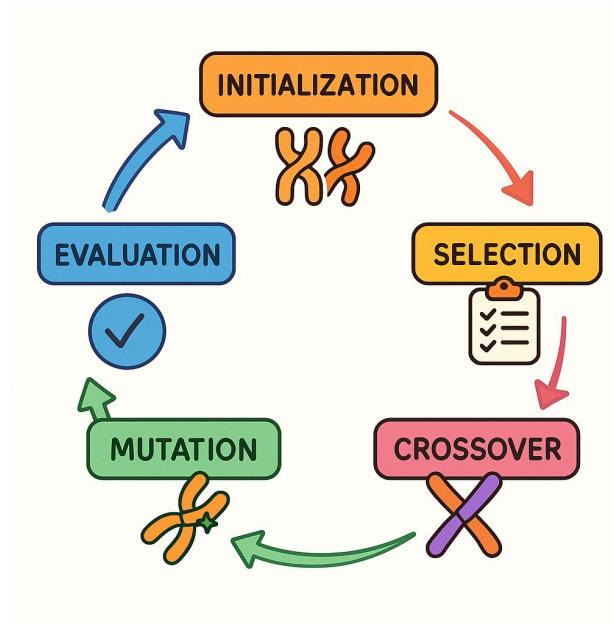


ChEMBL (version 22)  
1 554 260 compounds

# GenCReM = Genetic Algorithm & CReM

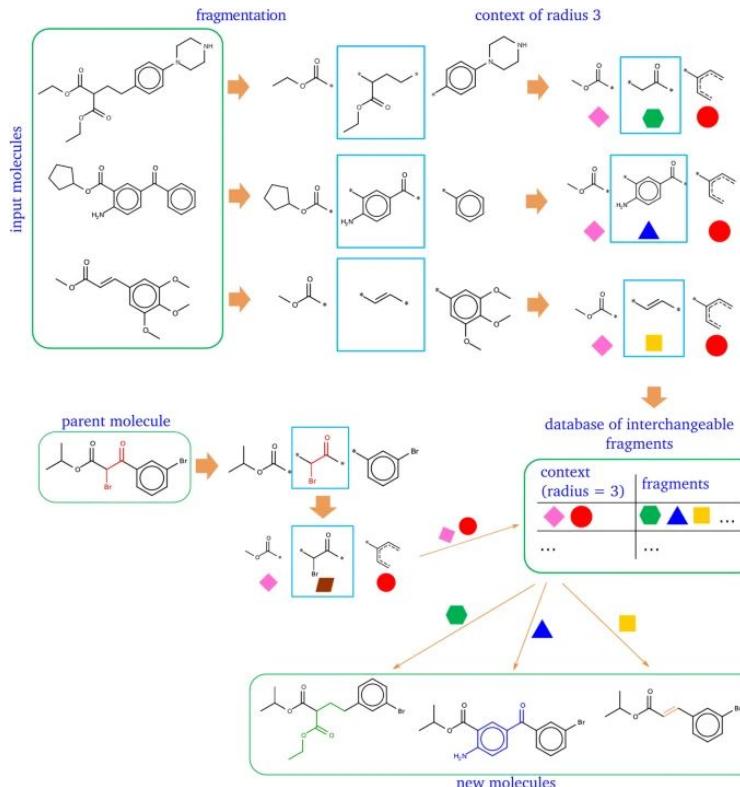


**Ensuring Synthetic Feasibility**



**Navigating Chemical Space Efficiently**

# GenCReM = Genetic Algorithm & CReM

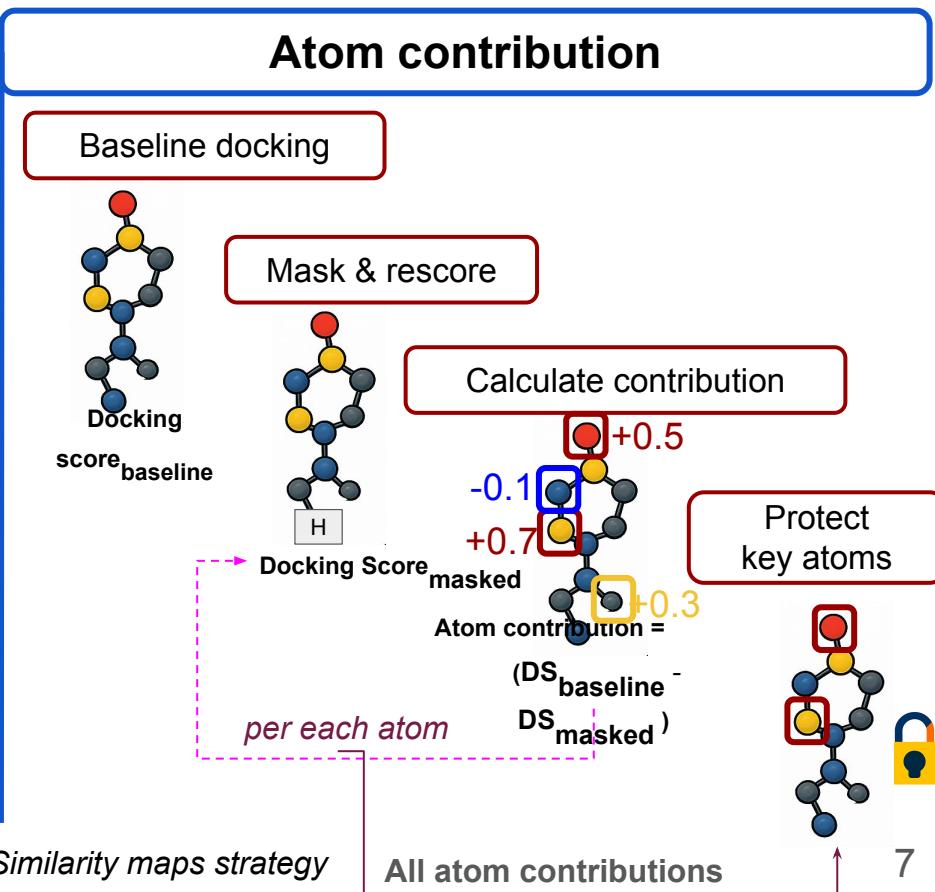
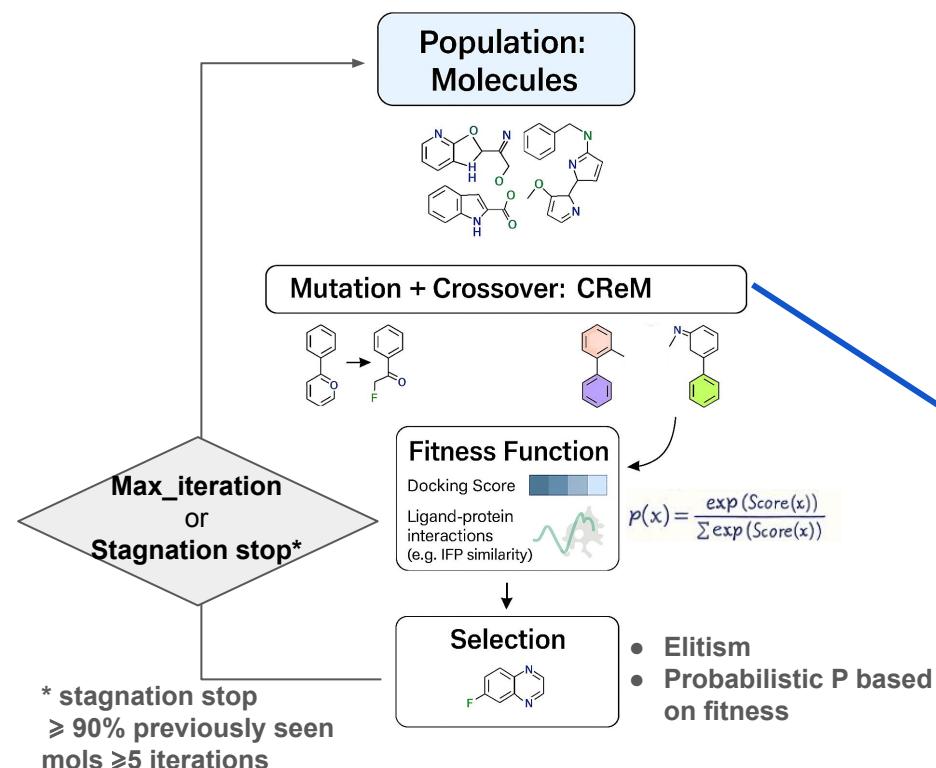


## CReM: chemically reasonable mutations

- Database-derived fragment replacements
- Chemical validity by design
- Applies context-aware replacements of fragments
- Indirect control of synthetic synthetic accessibility
- Custom fragment databases derived from in-house libraries

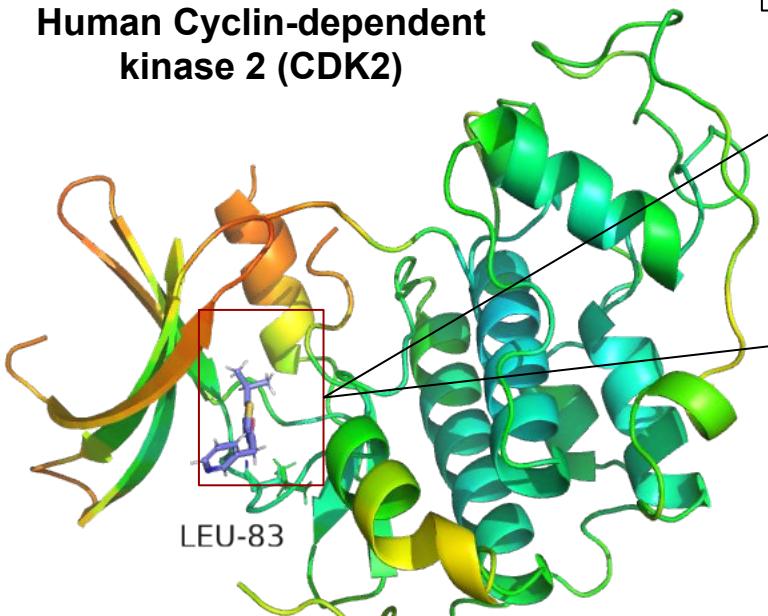
## Ensuring Synthetic Feasibility

# GenCReM = Genetic Algorithm & CReM



# Test study

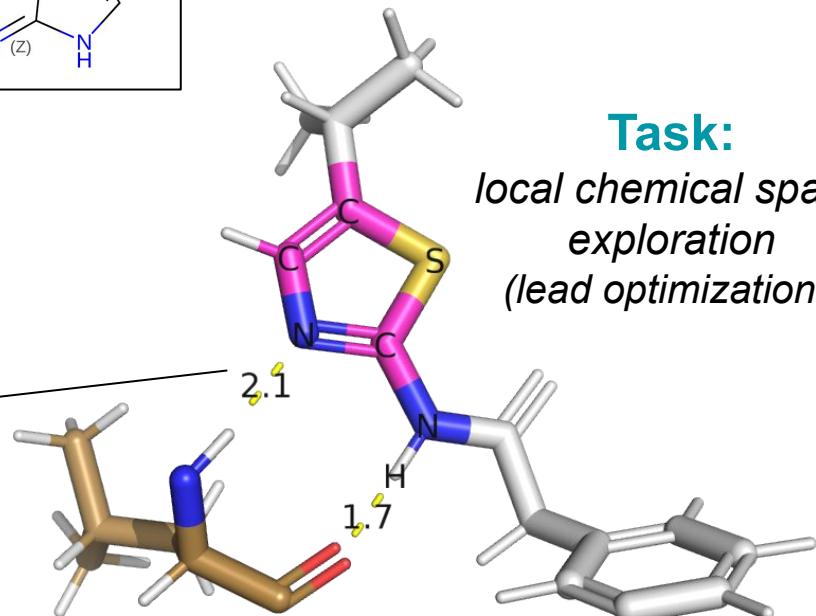
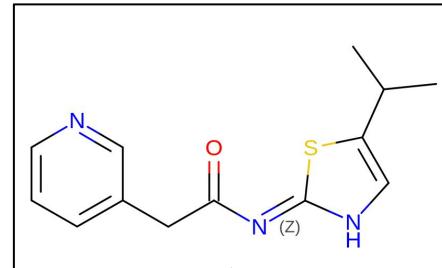
Human Cyclin-dependent kinase 2 (CDK2)



**2BTR**

**pIC50<sub>exp</sub> = 7.02**

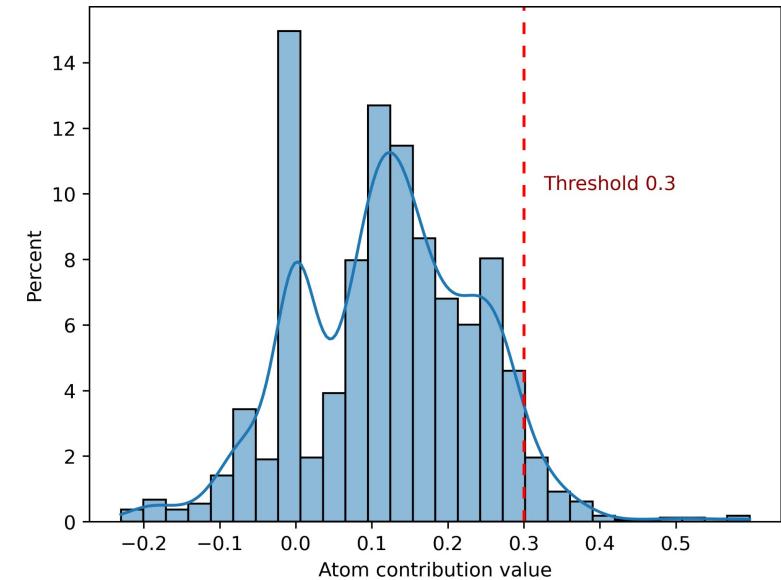
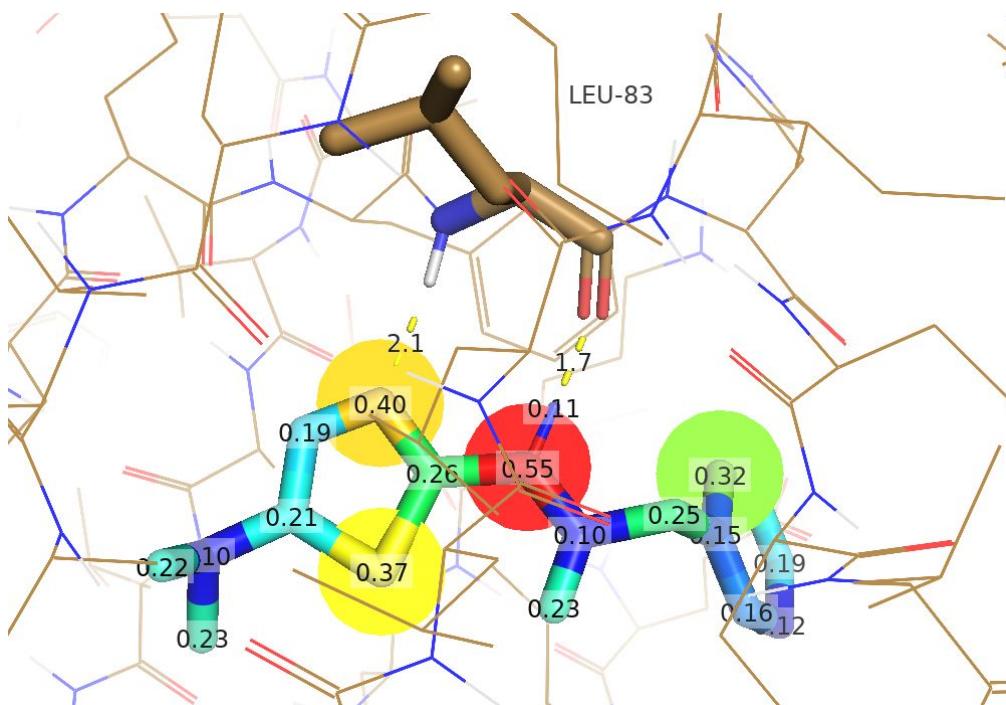
**Gnina score = 6.9**



*hinge region*

**Task:**  
local chemical space  
exploration  
(lead optimization )

# Atom contributions



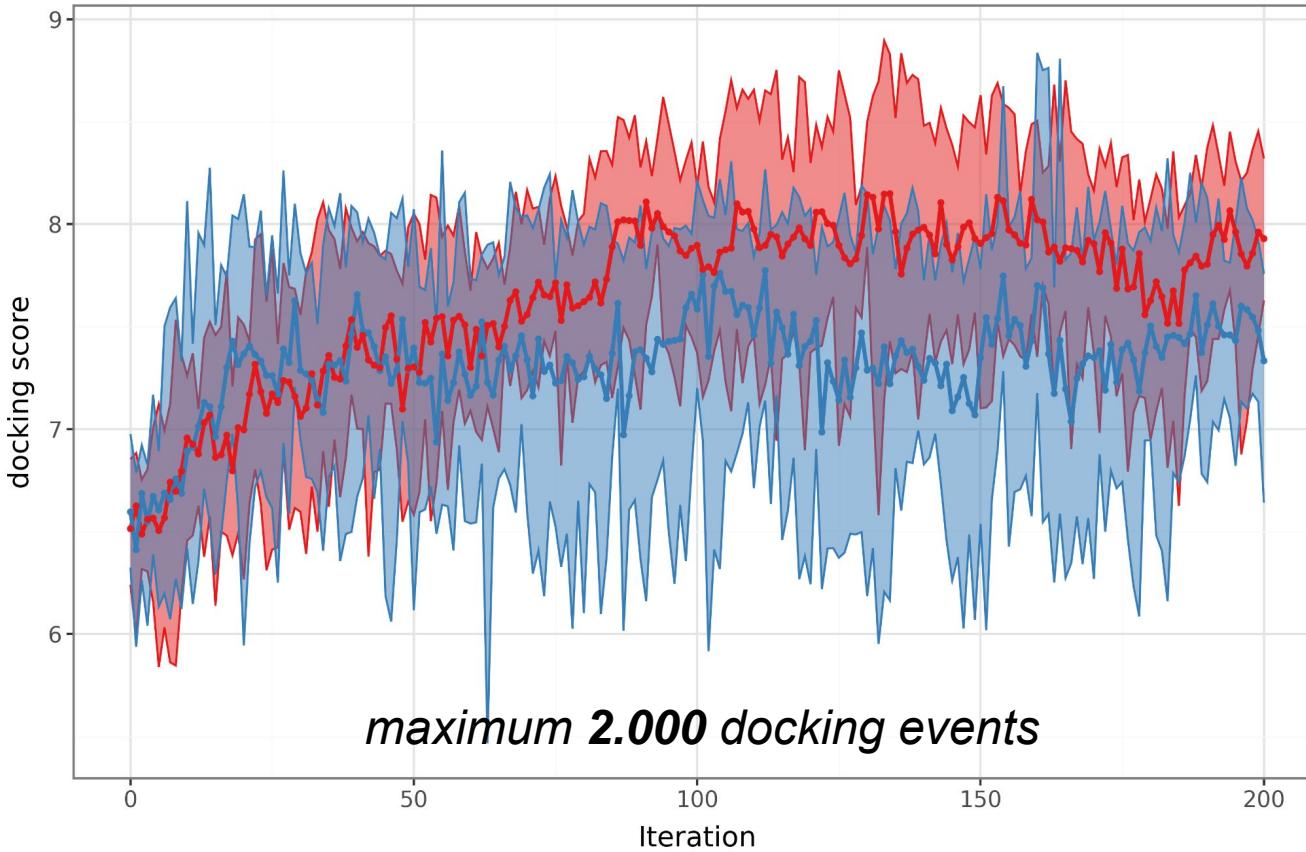
# What has been done

- CReM radius 3 / 5
- Low ( 10 molecules in the population, *max 2.000 dockings events*) - medium (100 molecules in the population, *max 20.000 dockings events*) - high budget (500 molecules in the population, *max 100.000 dockings events*) structure generations
- Gnina

## Settings:

- The maximum size of the replacement fragment can not be > 10 atoms
- The strict physico-chemical filters were used during generations:  
**MW <= 450, RTB <= 5, logP <= 4, tpsa <= 120**
- All runs were replicated 3 times (**The best replica (by docking score) is used** for the further analysis)
- Crossover procedure was not used yet

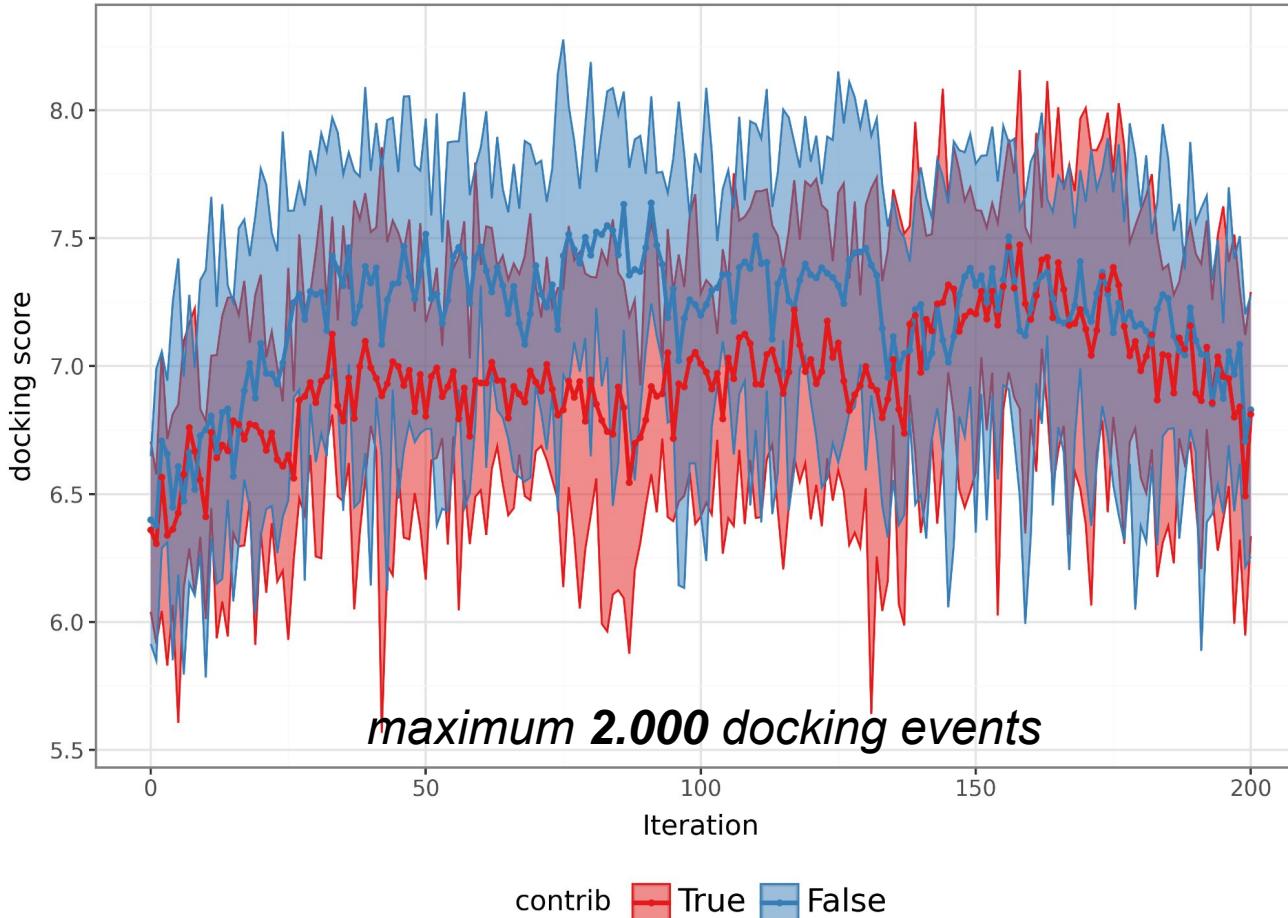
# Population size 10. Radius 3



Low-budget  
structure  
generations

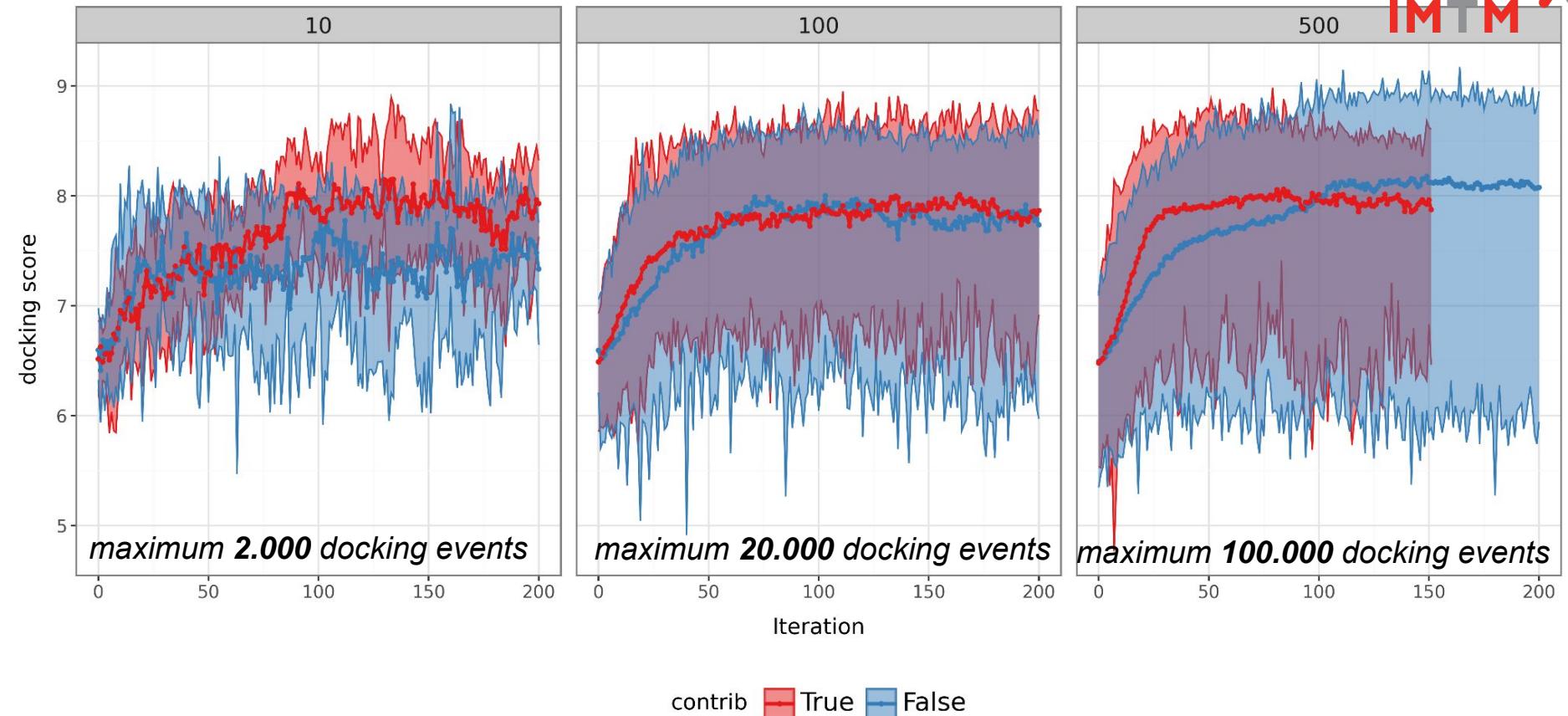
contrib  True  False

Population size 10. Radius 5



At radius of 5 the  
impact of the atom  
contributions  
became negligible

Radius 3

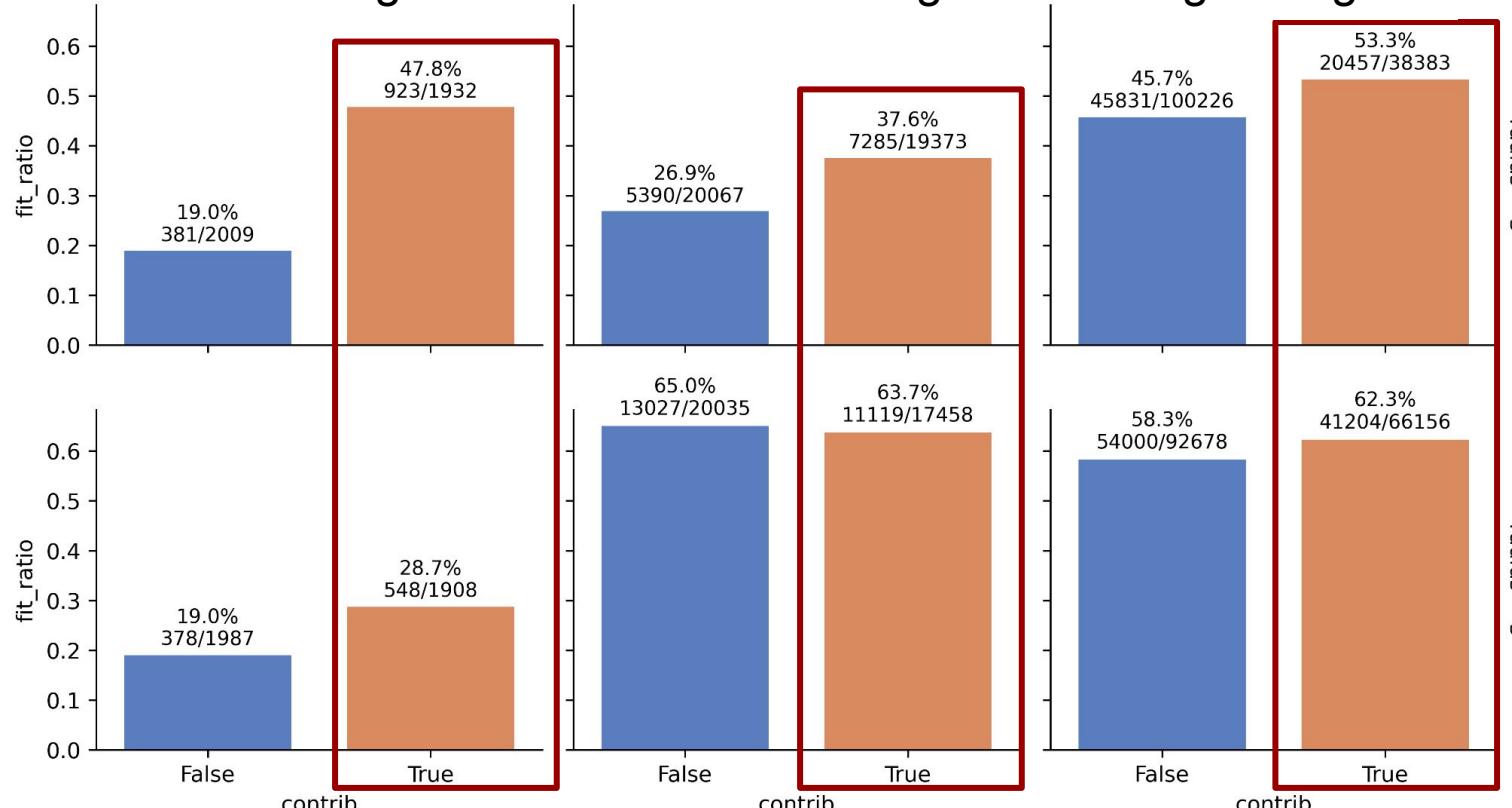


Atom contributions usage leads to early convergence in medium- and high-budget scenarios

## Low-budget

## Medium-budget

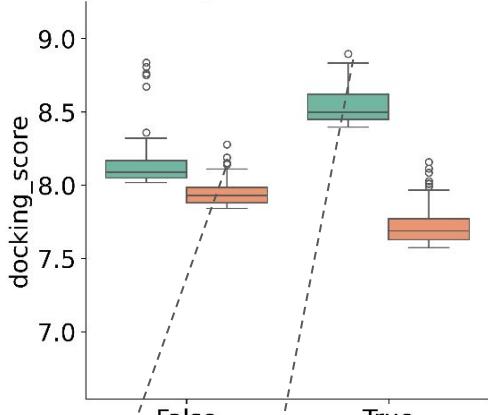
## High-budget



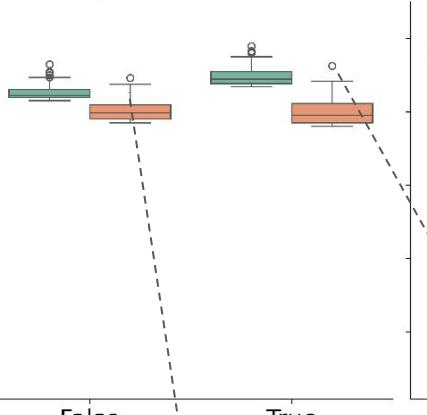
**Atom contributions help to partially preserve important for binding Prot-Lig contacts**

Docking scores for the top100 molecules

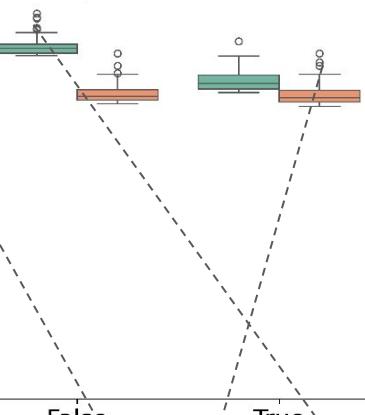
Low-budget



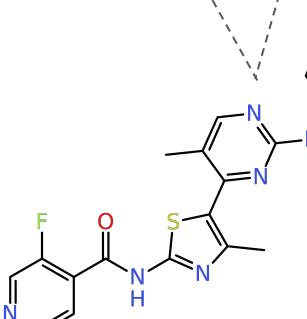
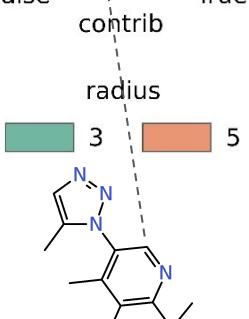
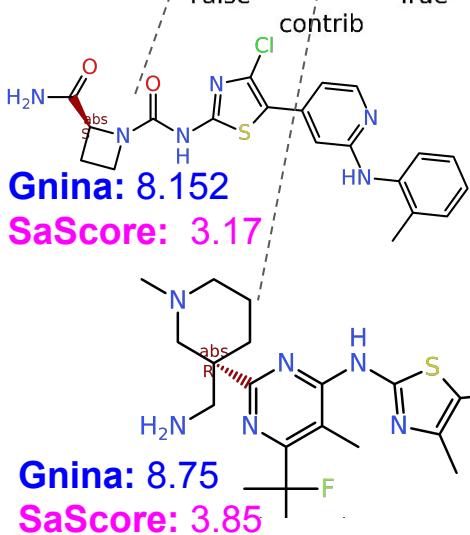
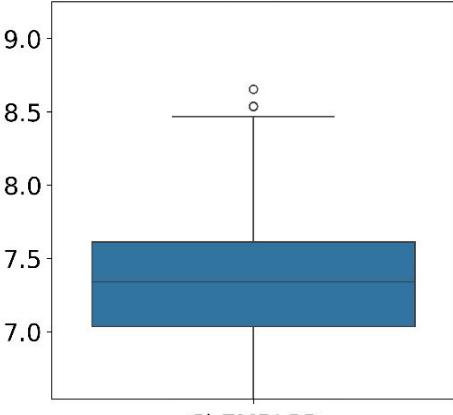
Medium-budget



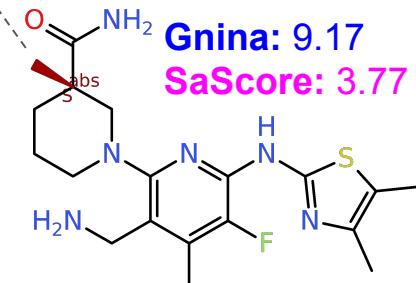
High-budget



Known actives IC50 >= 6  
Gnina, top1000



**Gnina: 8.81**  
**SaScore: 2.87**

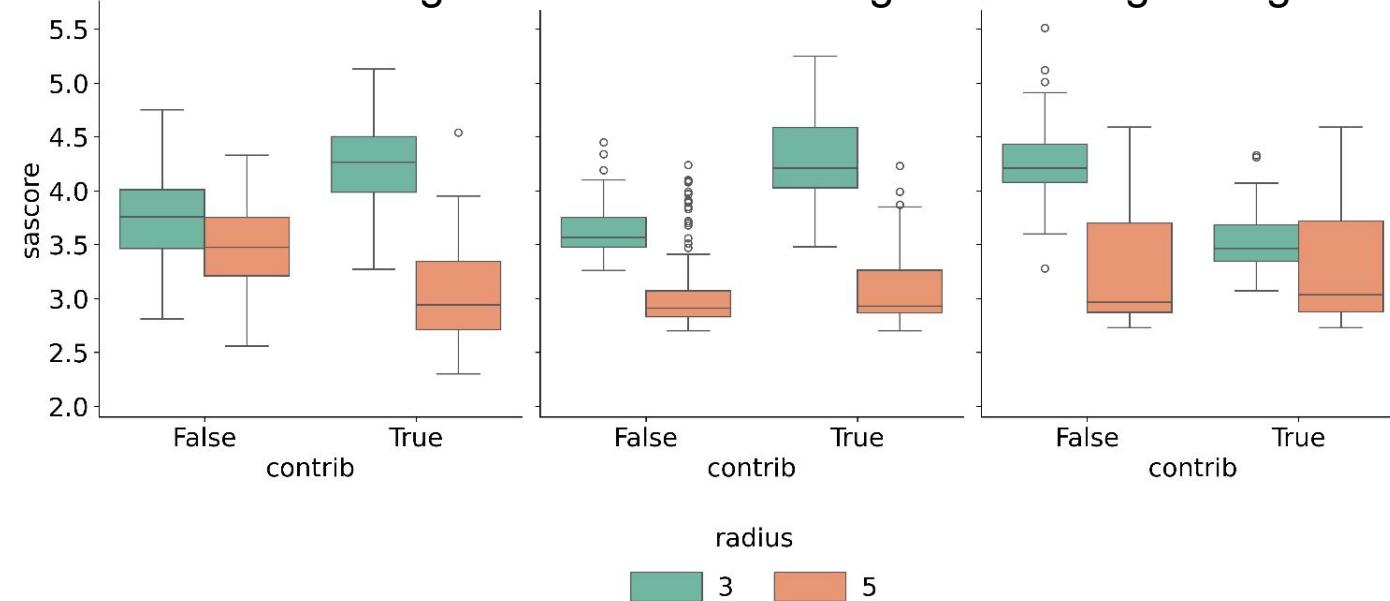


SA scores for the top100 molecules

Low-budget

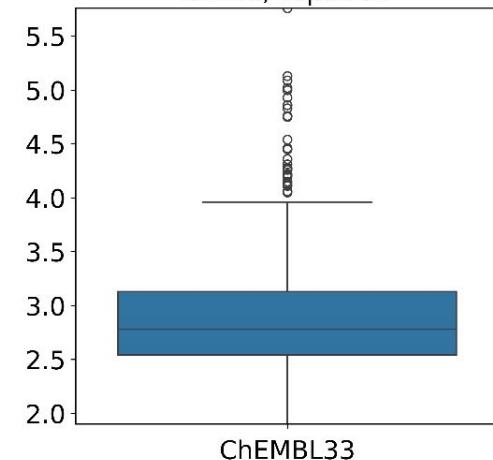
Medium-budget

High-budget



**Radius has a greater impact on the SA scores of the generated molecules, than the atom contributions usage or population size**

Known actives IC<sub>50</sub> >= 6  
Gnina, top1000



# Low-budget, Radius 3

starts from Gnina Score 6.9

**Known CDK2 inhibitors**



atom contrib = False

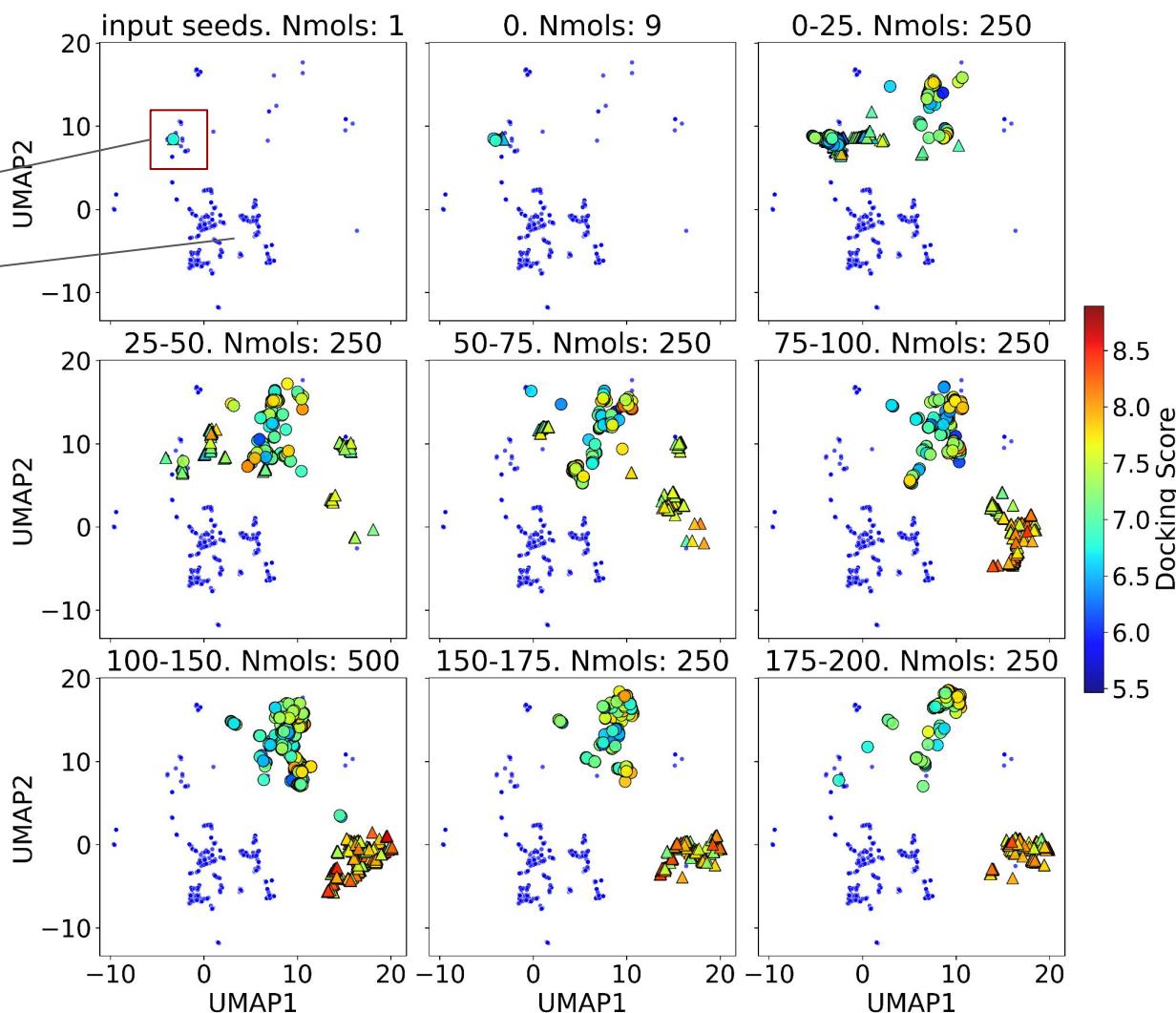
Max Gnina Score: 8.8  
Iteration (achieved the best score): 160



atom contrib = True

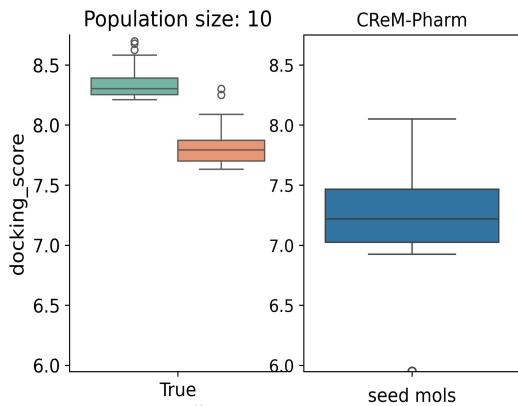
Max Gnina Score: 8.9  
Iteration (achieved the best score): 133

**Optimization procedure**

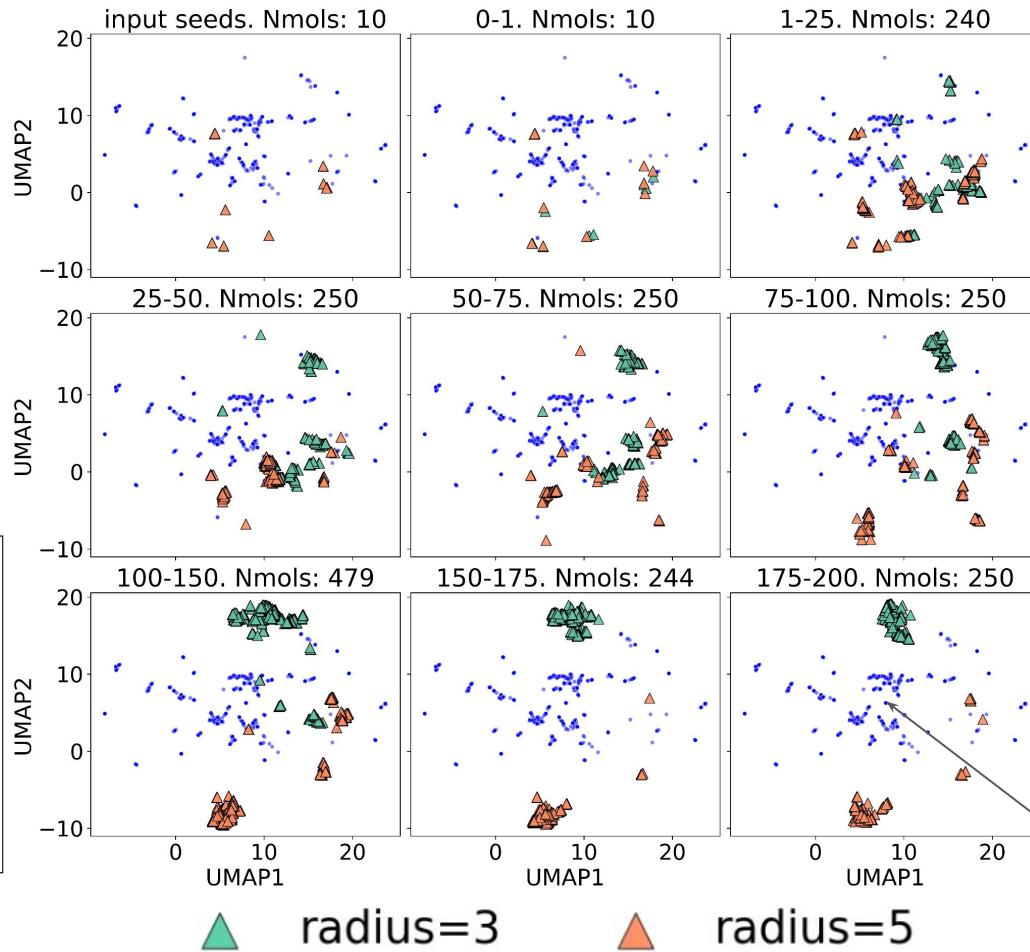
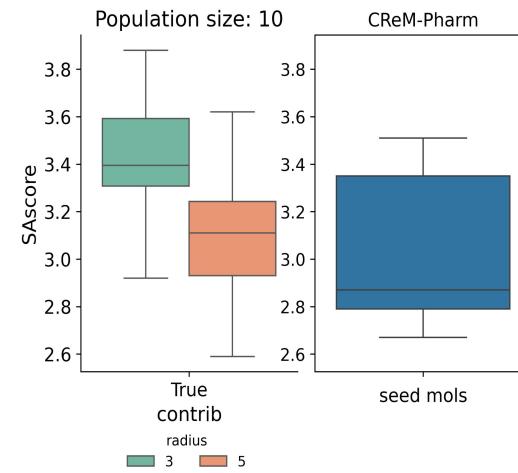


# Low-budget

Docking scores for the top100 molecules



SA scores for the top100 molecules



CReM-Pharm

Automatic grow procedure combined with 3D pharmacophores

Known CDK2 inhibitors

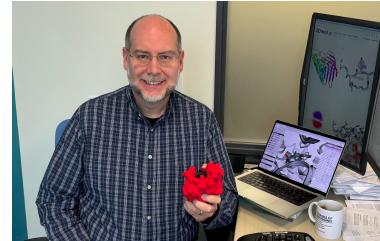
# Conclusions

- Combination of the **CReM** and **Genetic algorithm** allows to merge **synthetically feasible structure generation** with **efficient, goal-driven chemical space exploration**
- Generations guided by **atoms contributions** improve **low budget exploration in virtual chemical spaces of larger sizes** relatively to unguided mutations due to faster convergence.
- **Synthetic accessibility** of generated structures **depends only on the chosen context radius** within a particular database and **is not sensitive to other settings**.

# DRUG DISCOVERY:

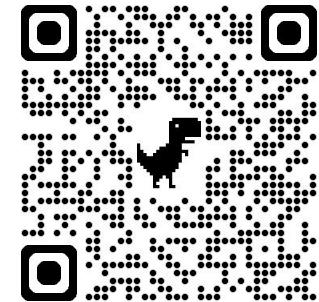


# Thank you!



## Acknowledgements

Special thanks to Dr. David Ryan Koes for his guidance and support with *Gnina*.



<https://github.com/ci-lab-cz>

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) and projects ELIXIR-CZ (LM2023055) and CZ-OPENSCREEN (LM2023052). 20