8th Advanced In Silico Drug Design workshop

27 - 31 January 2025
Olomouc, Czech Republic

Univerzita Palackého v Olomouci

CReM: features and applications

Pavel Polishchuk

Institute of Molecular and Translational Medicine
Faculty of Medicine and Dentistry
Palacky University

pavlo.polishchuk@upol.cz
qsar4u.com

Polishchuk *J Cheminform* (2020) 12:28
https://doi.org/10.1186/s13321-020-00431-w

Journal of Cheminformatics

SOFTWARE Open Access

CReM: chemically reasonable mutations framework for structure generation

Pavel Polishchuk*

Abstract

Structure generators are widely used in de novo design studies and their performance substantially influences an outcome. Approaches based on the deep learning models and conventional atom-based approaches may result in invalid structures and fail to address their synthetic feasibility issues. On the other hand, conventional reaction-based approaches result in synthetically feasible compounds but novelty and diversity of generated compounds may be limited. Fragment-based approaches can provide both better novelty and diversity of generated compounds but the issue of synthetic complexity of generated structure was not explicitly addressed before. Here we developed a new framework of fragment-based structure gene
provides flexible control over diversity, novelty, sy
framework was implemented as an open-source
exploration of chemical space.

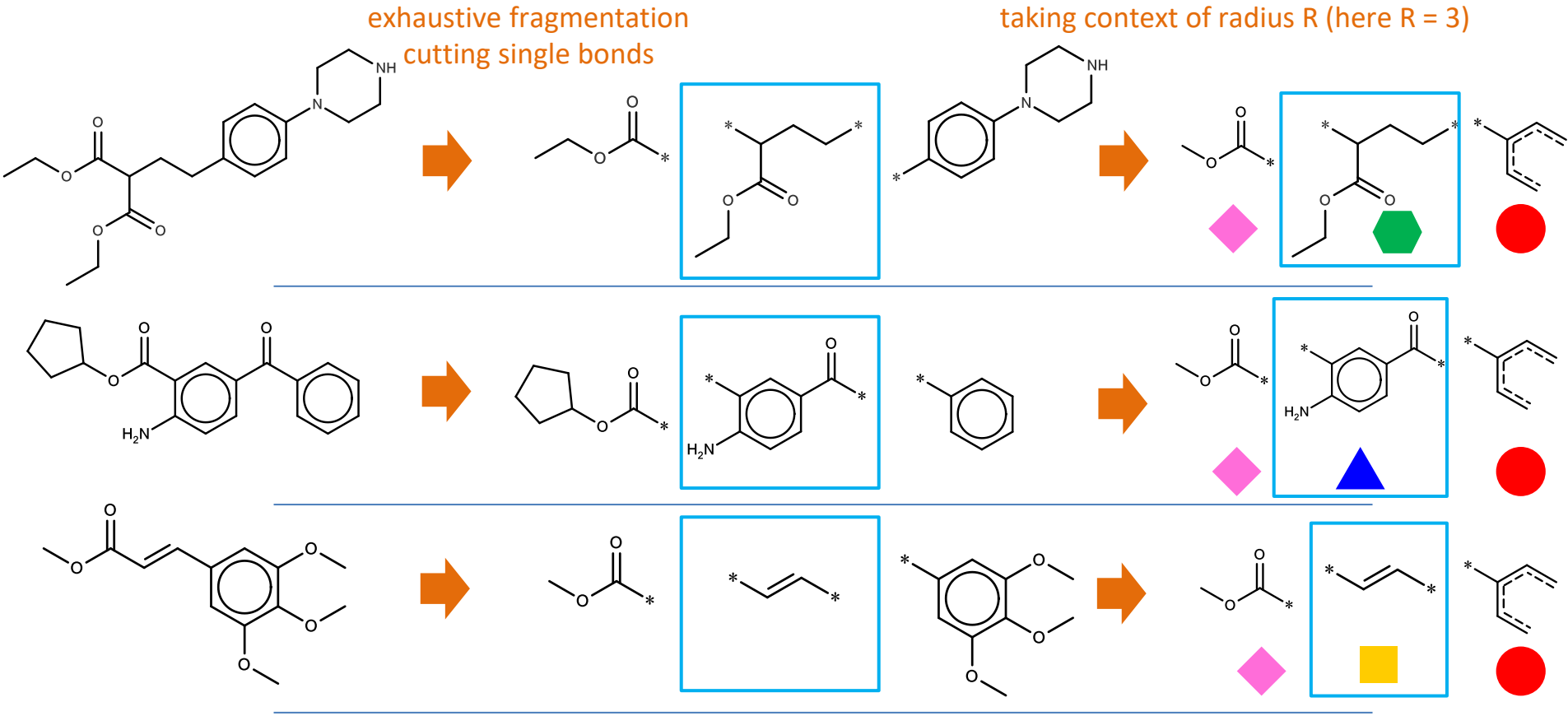
Keywords: De novo structure generation, De no

```
(rd2209) pavel@pavel-nb:~/python/streamd$  pypistats overall crem
```

category	percent	downloads
with_mirrors	100.00%	23,869
without_mirrors	82.25%	19,632
Total		23,869

```
Date range: 2024-07-31 - 2025-01-27
```

Chemically reasonable mutations (CReM)



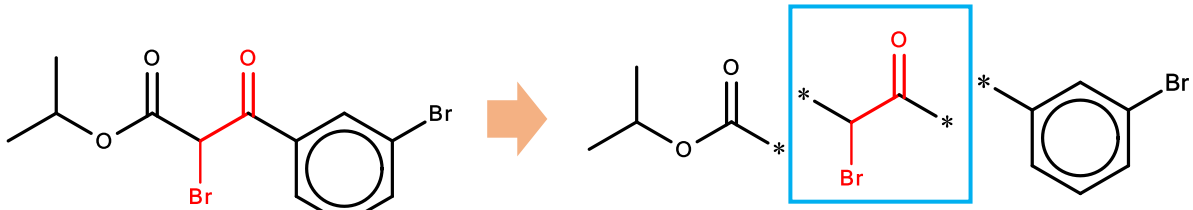
DB of replacements

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

interchangeable fragments

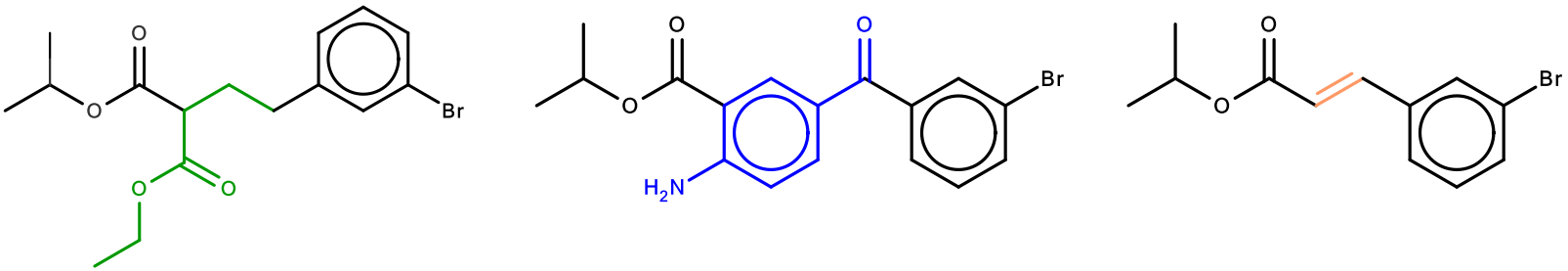
Polishchuk, P., CReM: chemically reasonable mutations framework for structure generation. *J. Cheminf.* **2020**, 12 (1), 28.

Chemically reasonable mutations (CReM)



DB of replacements

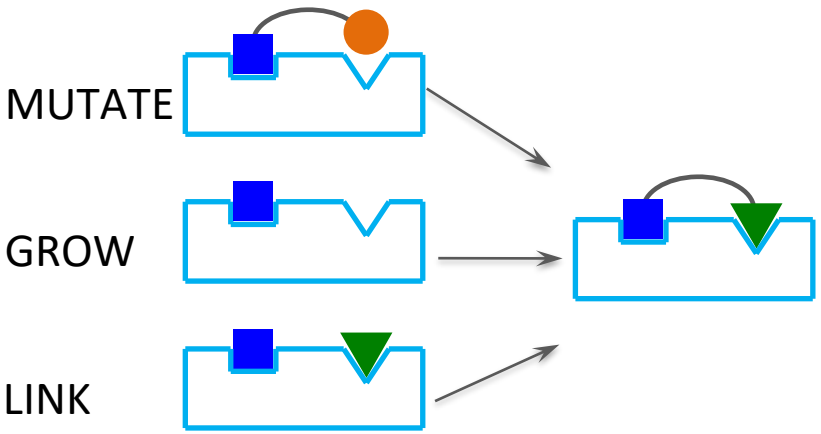
environment (radius = 3)	fragments

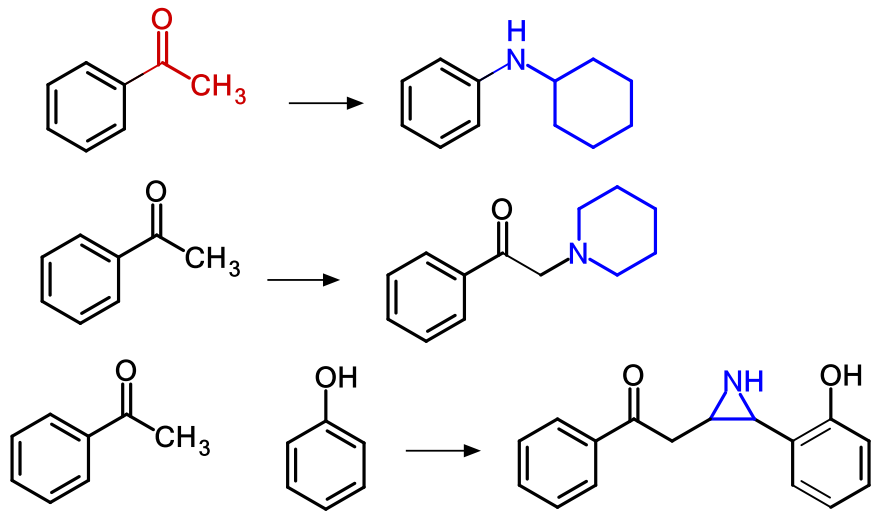


Generated structures are always chemically valid!

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

CReM features & applications

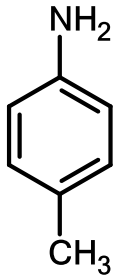




- use a custom (in-house) **fragment database** to generate more synthetically accessible compounds enriched with specific chemotypes
- choose larger **radiuses** to make replacements more conservative and resulting to more synthetically accessible compounds
- specify the **size of replaced and replacing fragments** to control granularity of steps in chemical space
- specify **atoms** to **protect** **or** **replace** to direct structural modifications
- specify the topological **distance** between attachment points in a linker

1. Scaffold decoration
2. Enumeration of analog series
3. Hit expansion
4. Lead optimization
5. De novo design

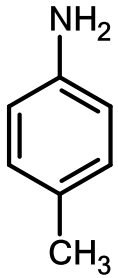
Radius of chemical context



GROW

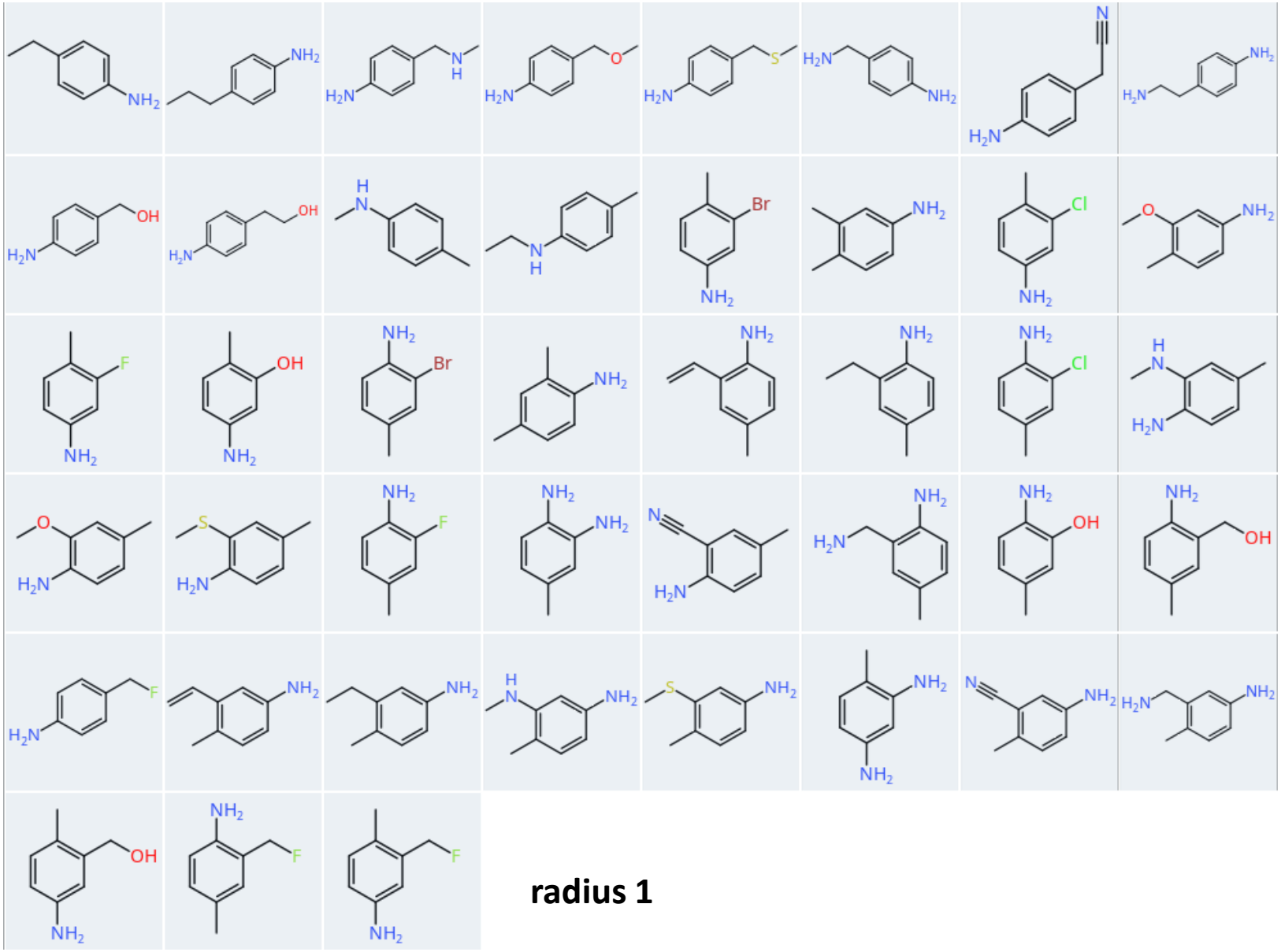
max_atoms=2

Radius of chemical context



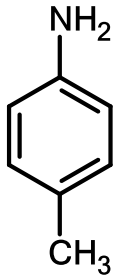
GROW

max_atoms=2



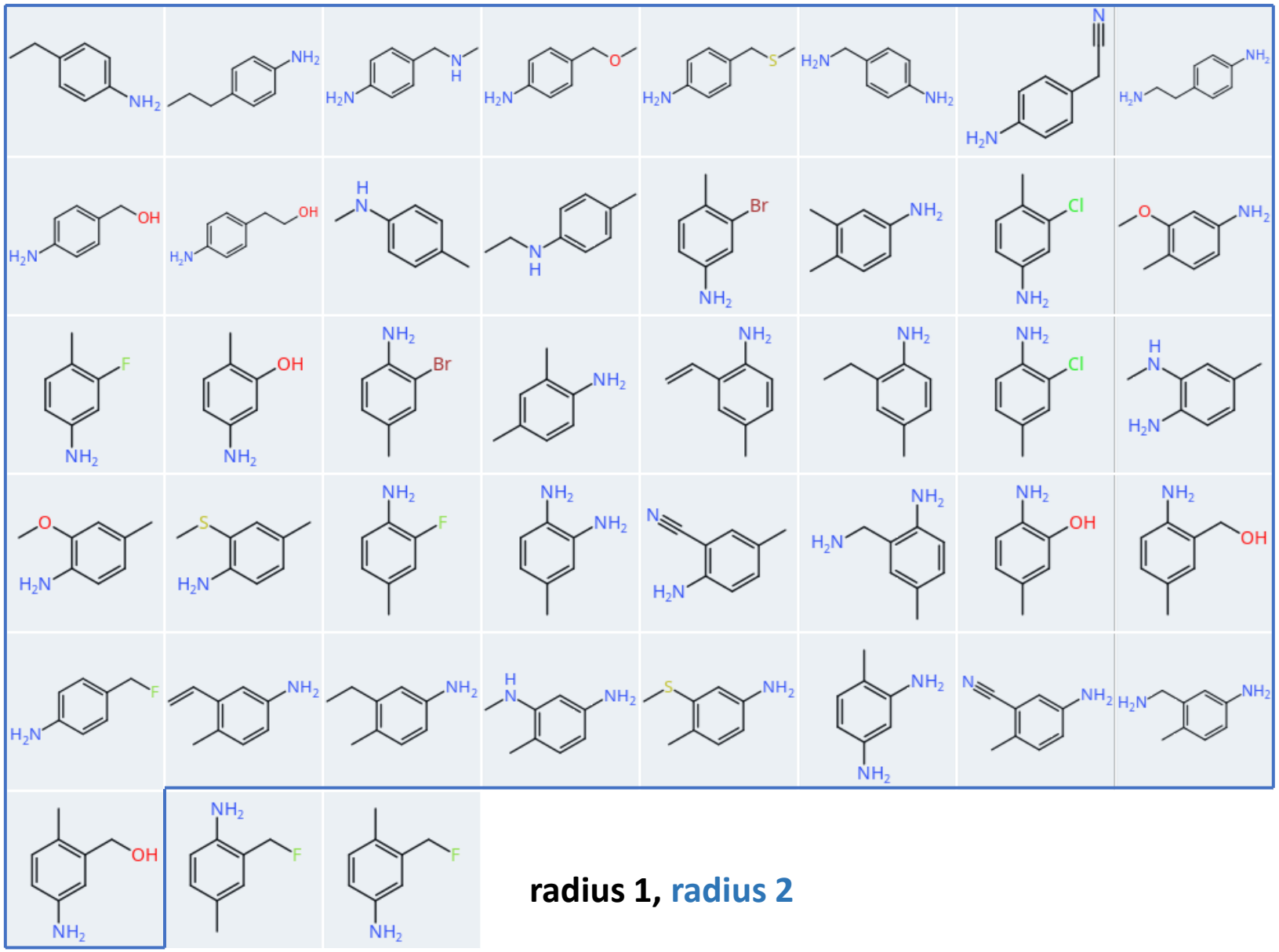
radius 1

Radius of chemical context



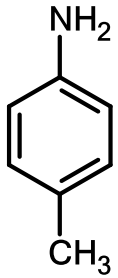
GROW

max_atoms=2



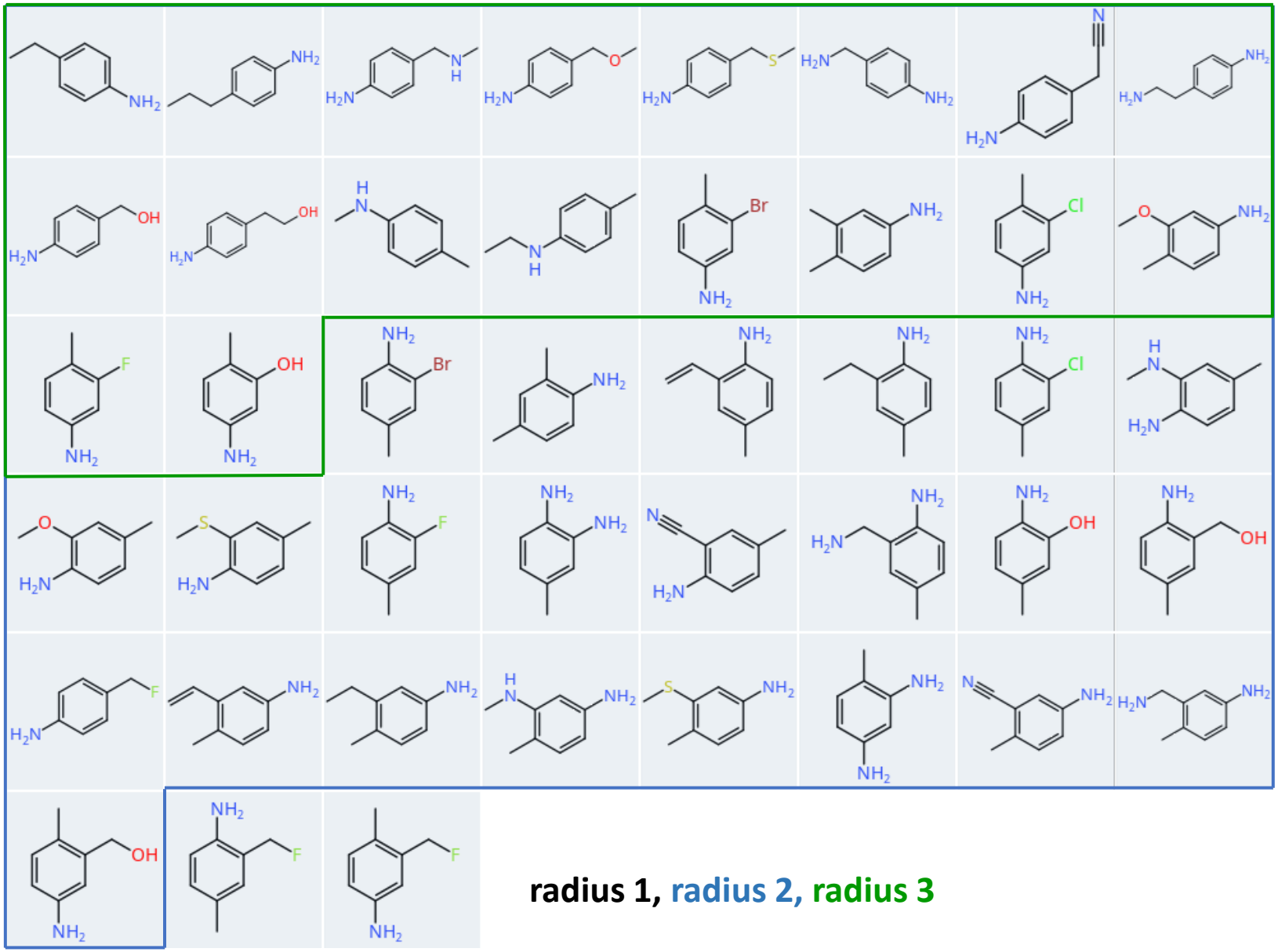
radius 1, radius 2

Radius of chemical context



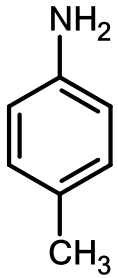
GROW

max_atoms=2



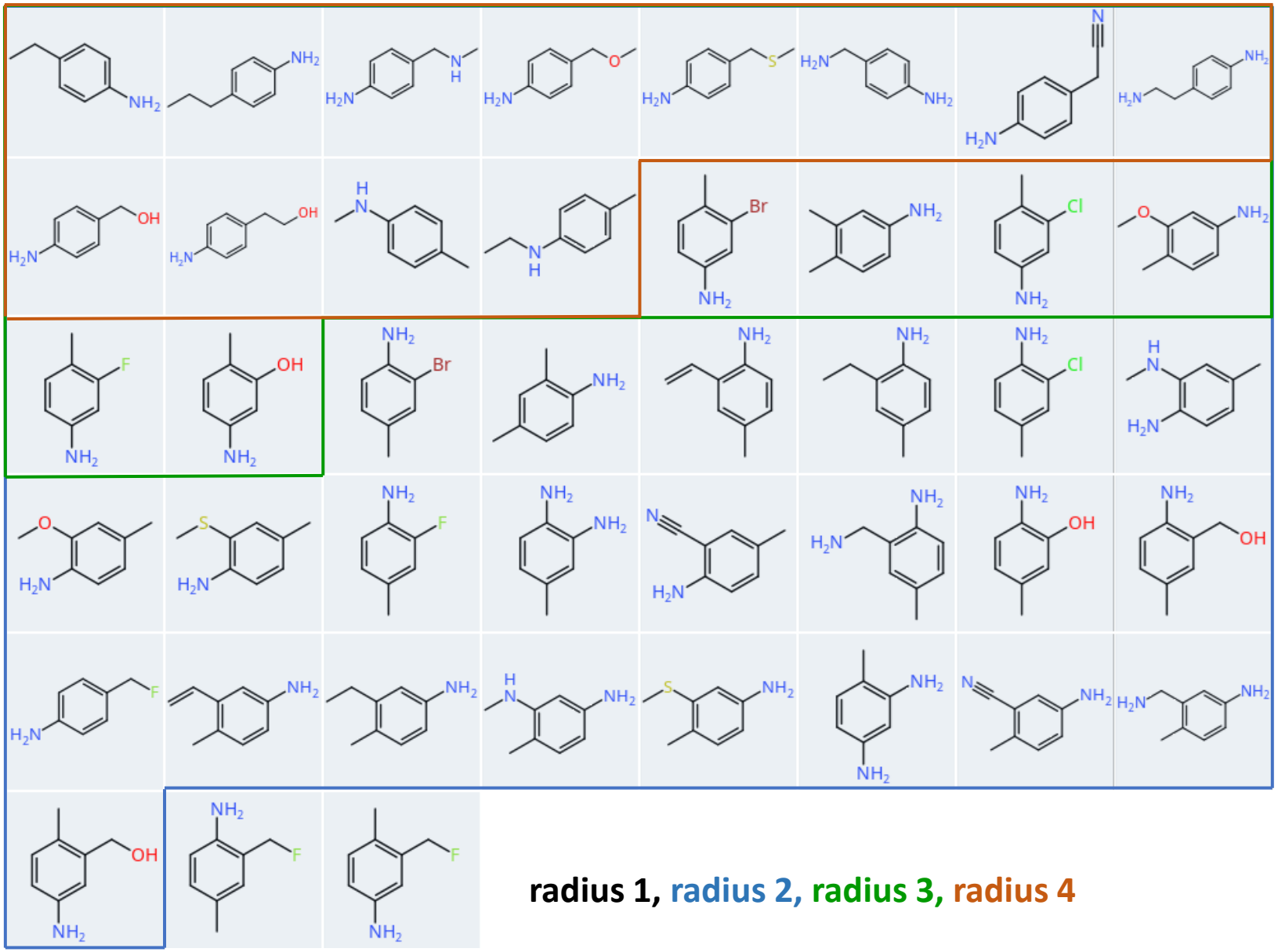
radius 1, radius 2, radius 3

Radius of chemical context



GROW

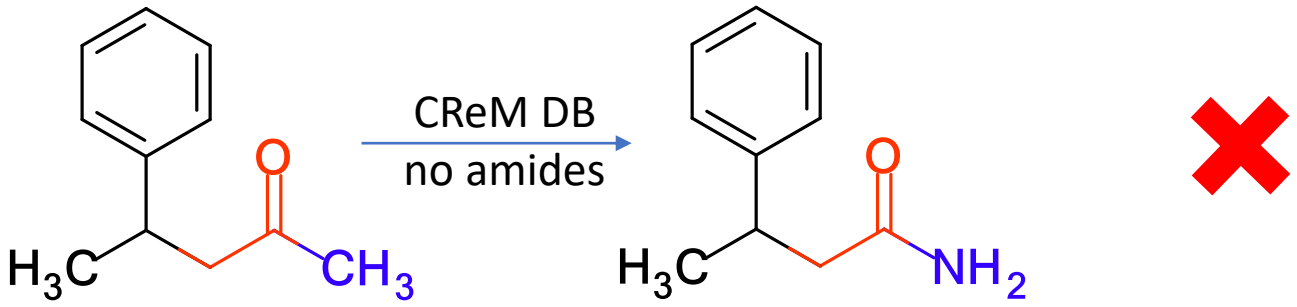
max_atoms=2



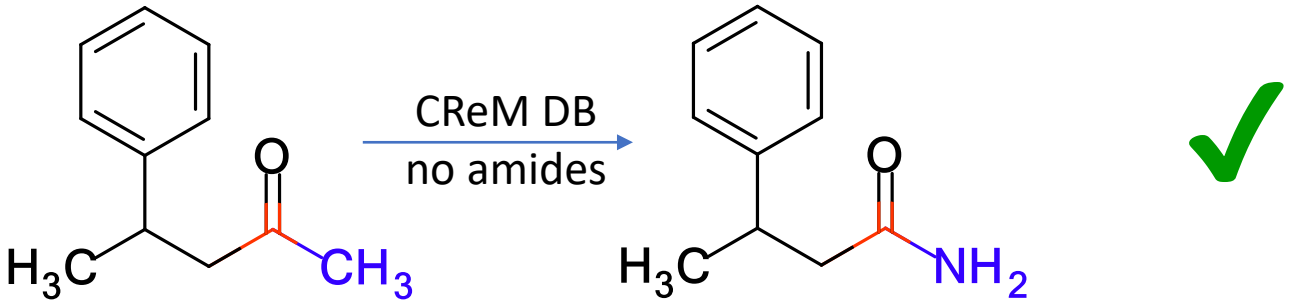
radius 1, radius 2, radius 3, radius 4

Radius of chemical context

Generated new chemotypes will have a size greater than a selected radius

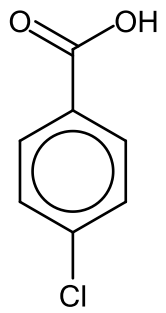


context radius 2

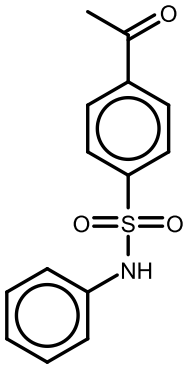


context radius 1

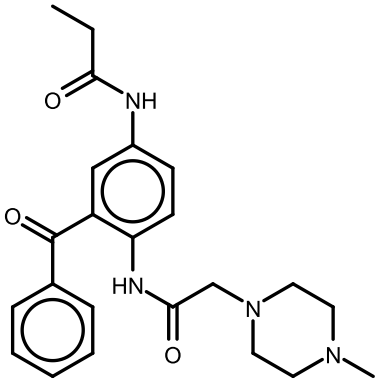
Synthetic accessibility of compounds



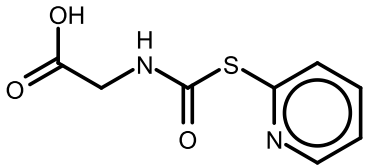
1.2
CHEMBL618



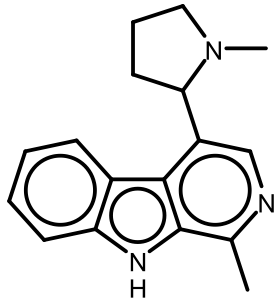
1.5
CHEMBL3310985



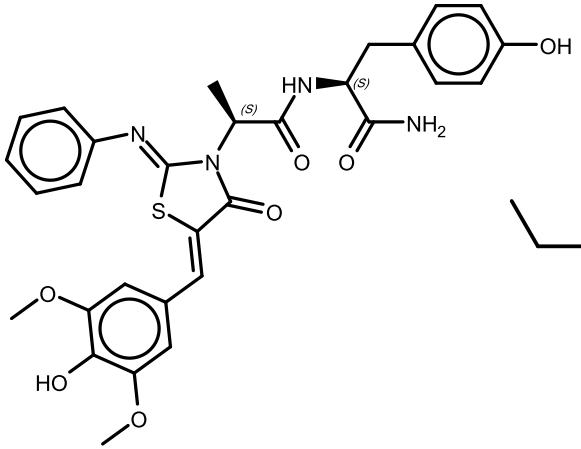
2.0
CHEMBL595820



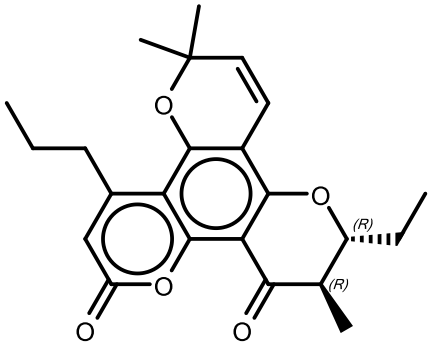
2.5
CHEMBL503660



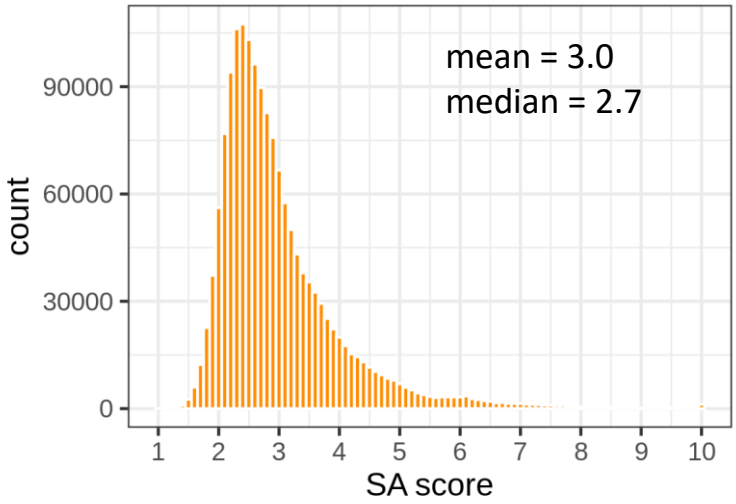
3.0
CHEMBL500286



3.5
CHEMBL582554



4.0
CHEMBL7633



Ertl, P.; Schuffenhauer, A., Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics* **2009**, 1, 8. (10.1186/1758-2946-1-8)

Synthetic accessibility of compounds

ChEMBL22
(1.55 M)

ChEMBL

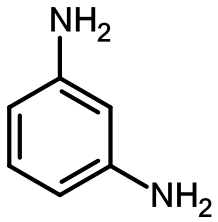
BMS
Dundee
Glaxo
Inpharmatica
PAINS

SA ≤ 2.5

SA ≤ 2

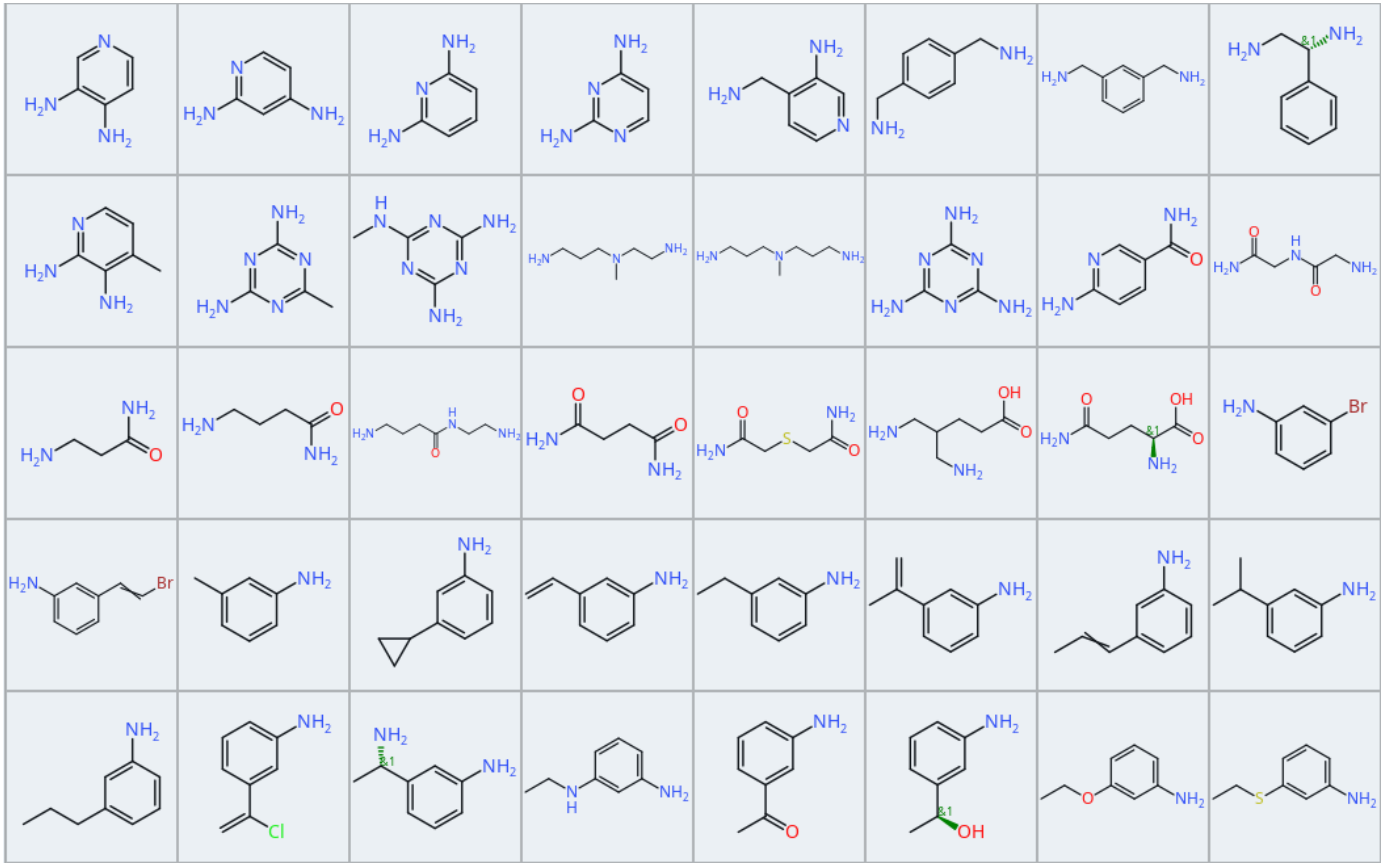
CReM DB	n (molecules)	n (distinct fragments, 12 atoms)	number of distinct fragment/context pairs for each radius				
			radius 1	radius 2	radius 3	radius 4	radius 5
all	818 174	988 585	2 263 436	4 051 790	7 133 534	11 007 247	15 271 543
SA2.5 (SA ≤ 2.5)	338 422	272 988	671 140	1 263 268	2 319 377	3 752 375	5 419 544
SA2 (SA ≤ 2)	67 970	55 498	143 434	267 156	472 126	754 905	1 087 492

Synthetic accessibility of compounds

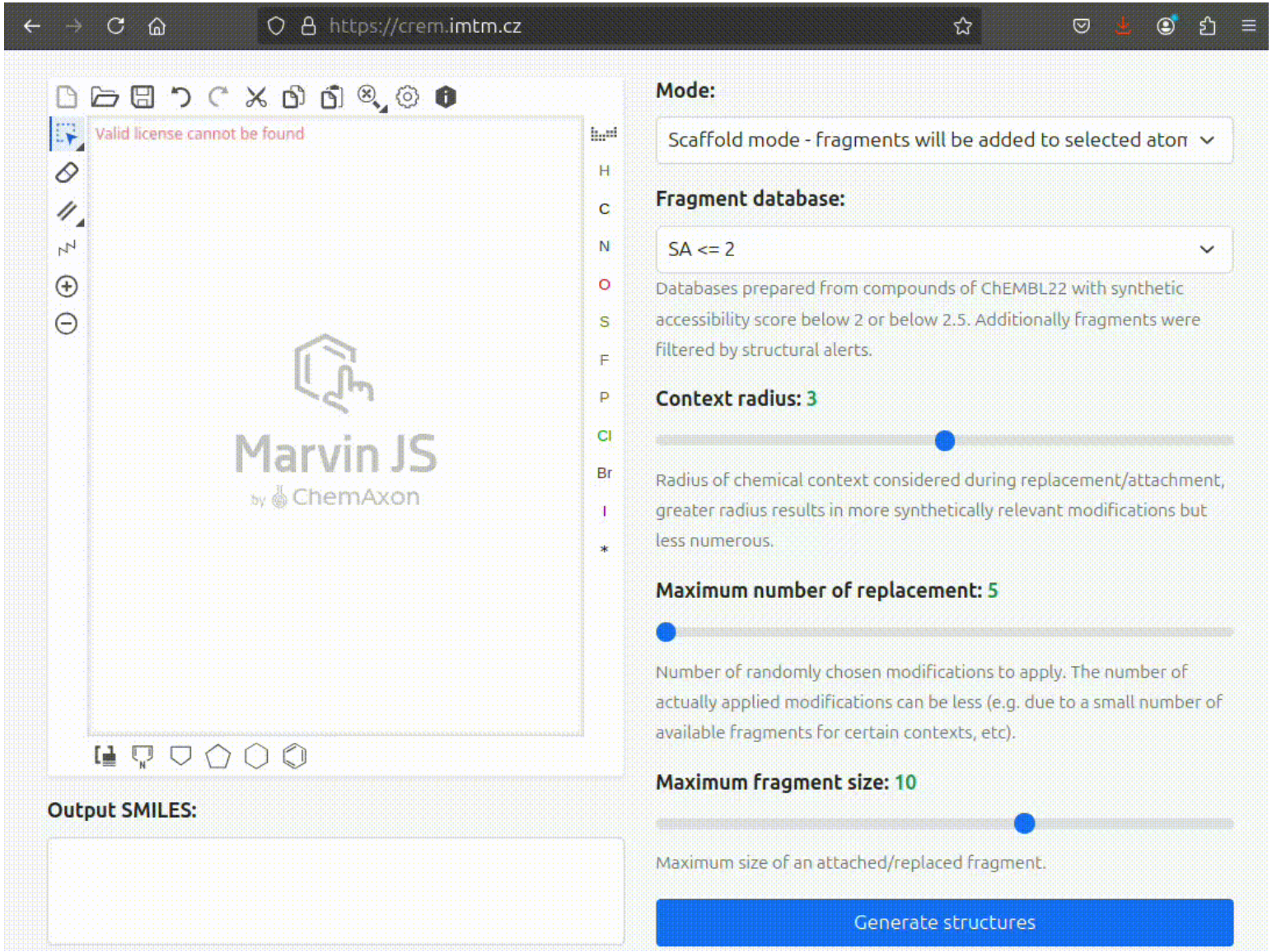


MUTATE

	radius 1	radius 2	radius 3	radius 4	radius 5
ChEMBL SA2.5	329	327	323	288	288
ChEMBL SA2	161	158	154	123	123

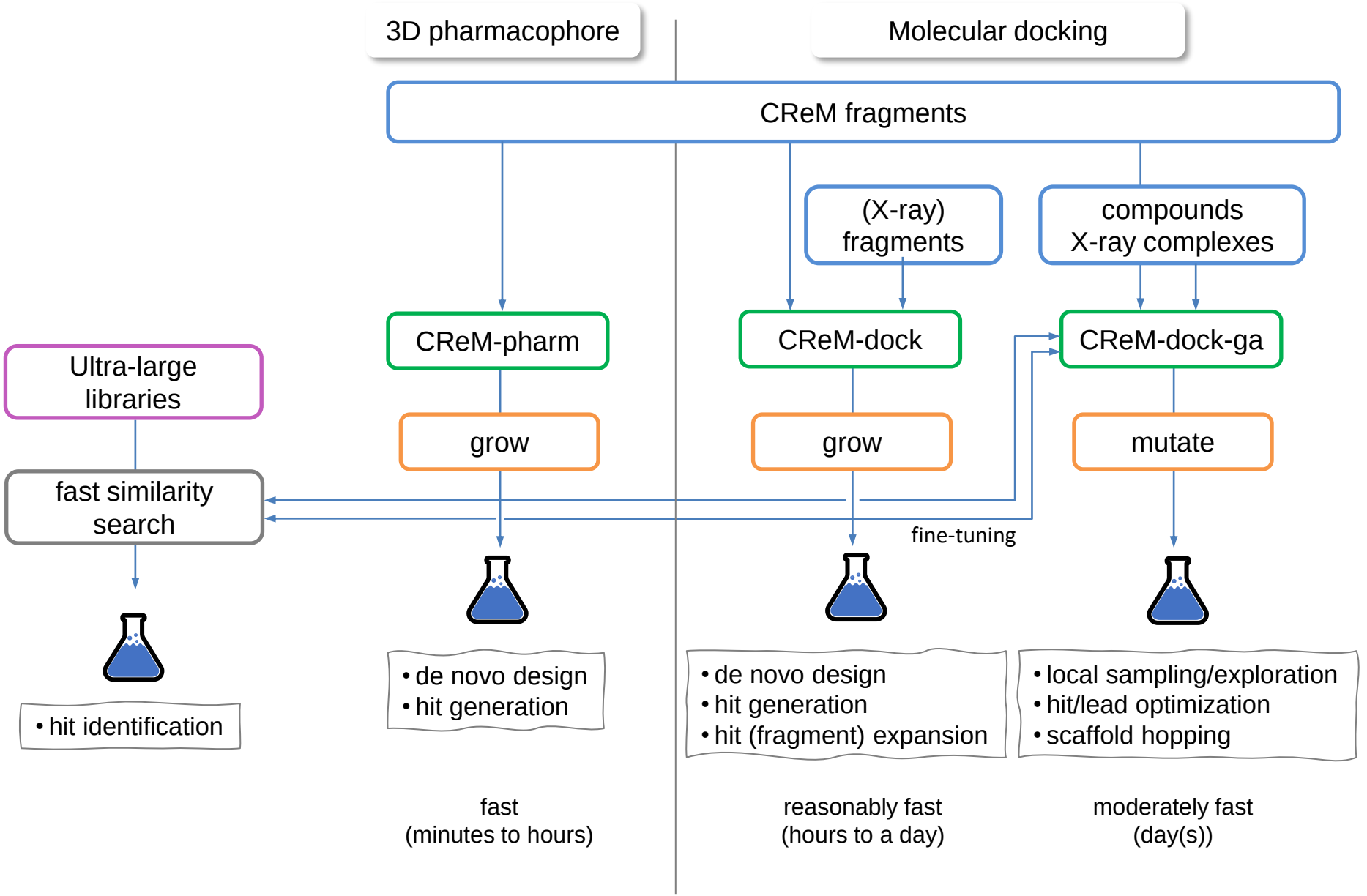


CReM online

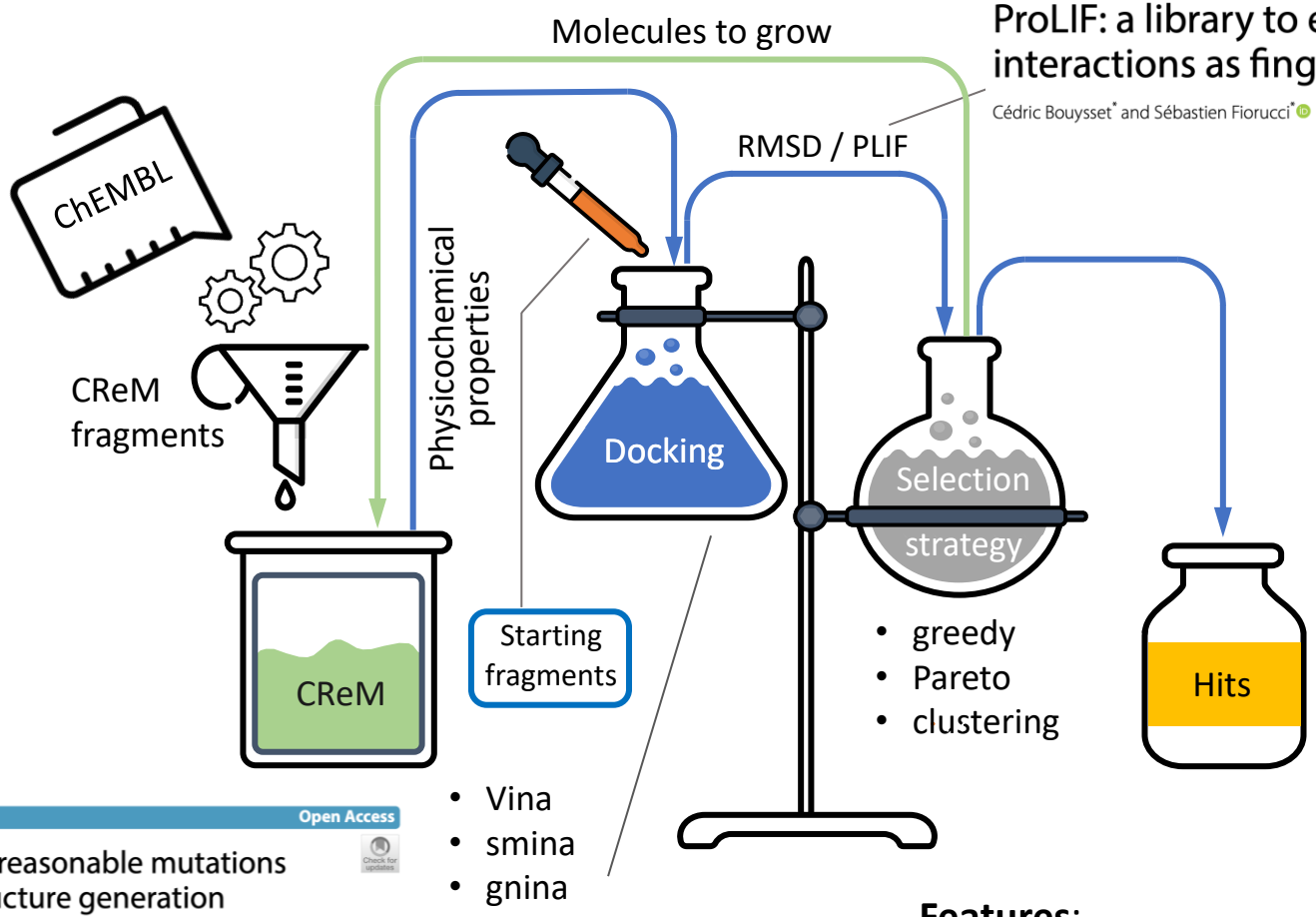


crem.imtm.cz

CReM-based applications



CReM-dock



SOFTWARE

ProLIF: a library to encode molecular interactions as fingerprints

Cédric Bouysset* and Sébastien Fiorucci*

SOFTWARE Open Access

CReM: chemically reasonable mutations framework for structure generation

Pavel Polishchuk*

SOFTWARE Open Access

EasyDock: customizable and scalable docking tool

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

Features:

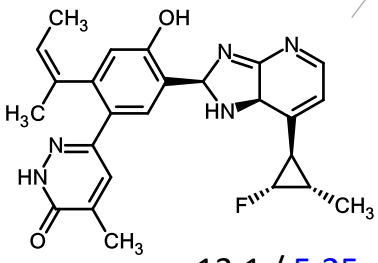
- control physicochemical properties
- control protein-ligand interactions
- keep the initial pose
- support different docking tools via EasyDock

Minibaeva and Polishchuk, CReM-dock: de novo design of synthetically feasible compounds guided by molecular docking, https://doi.org/10.26434/chemrxiv-2024-fpzqb-v2

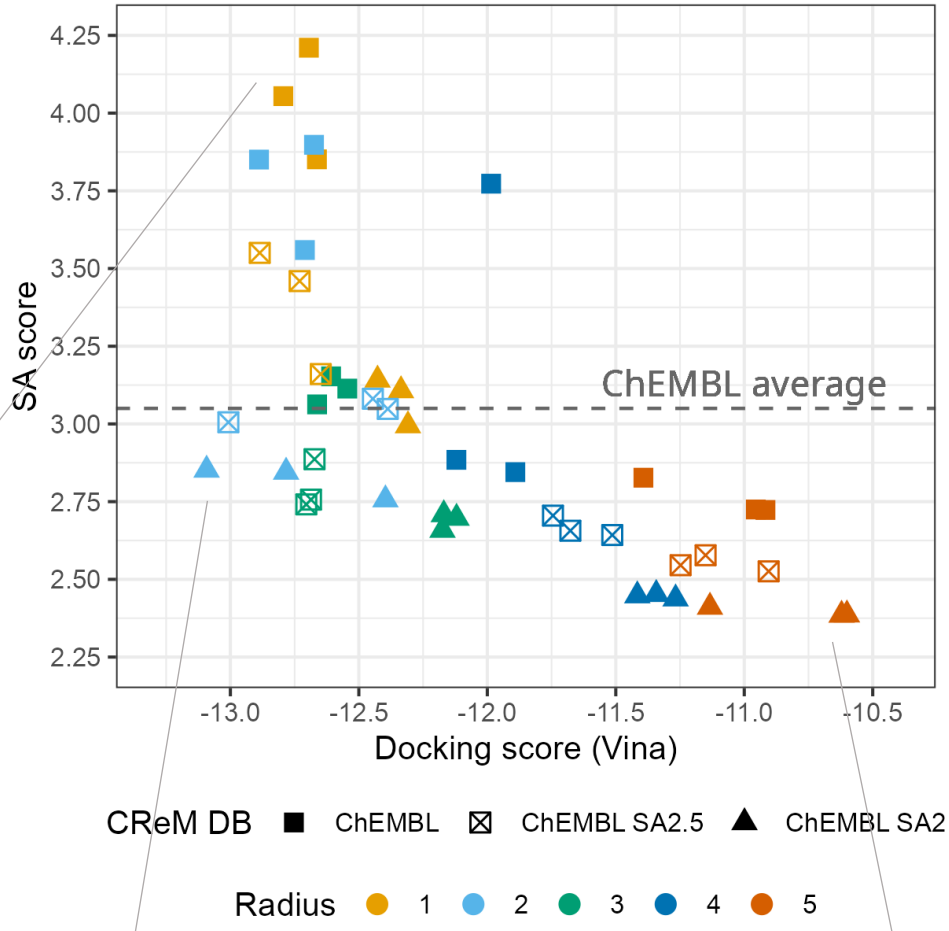
CReM-dock: CDK2 example

Settings:

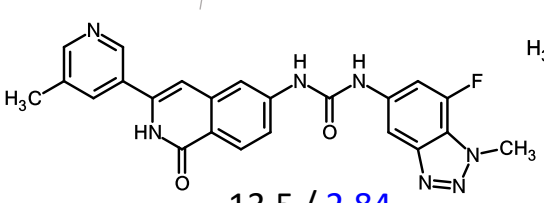
- CDK2 (2BTR)
- MW ≤ 450, logP ≤ 4, TPSA ≤ 120, RTB ≤ 7
- PLIF – hinge region interaction
- maximum number of replacements: 2000
- selection strategy: clustering (25 clusters, top 2 mols)
- 3 independent runs
- top 100 compounds by docking score



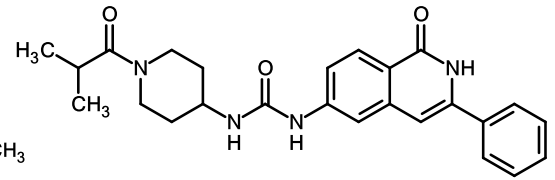
-13.1 / 5.25



- a clear trade-off between SA and docking scores
- SA scores are predictably changed with changing of a radius and a fragment database



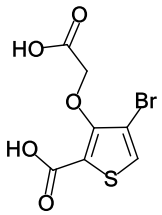
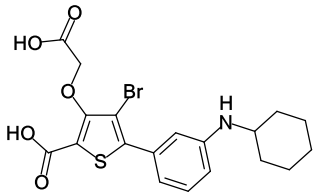
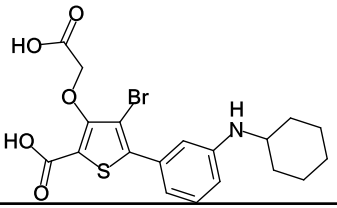
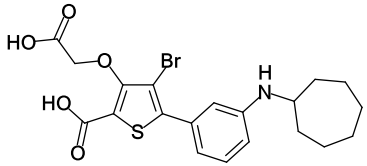
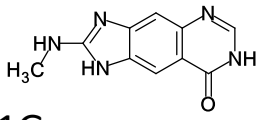
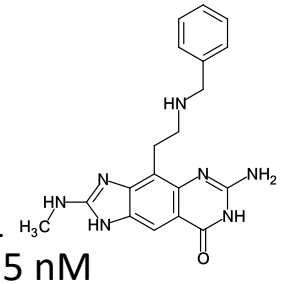
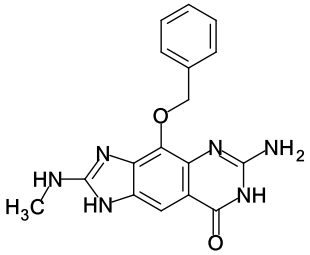
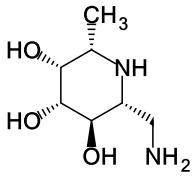
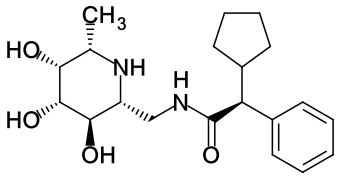
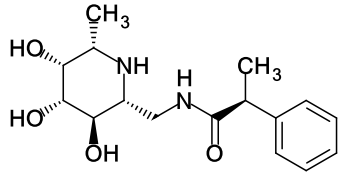
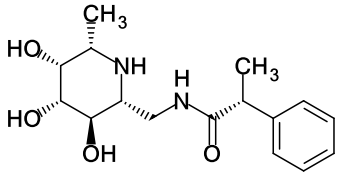
-13.5 / 2.84



-11.8 / 2.34

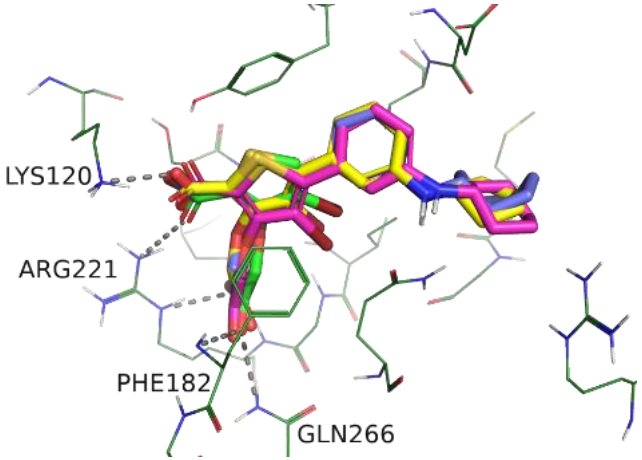
docking score / SA score

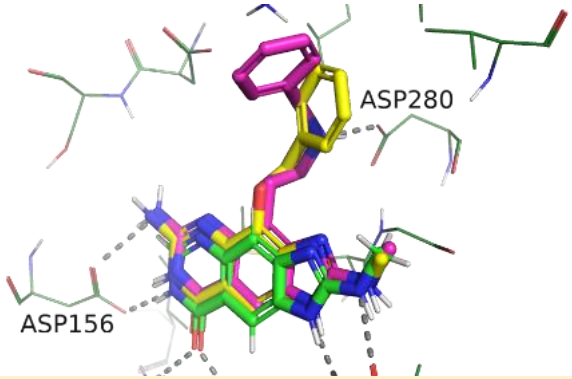
CReM-dock: fragment expansion

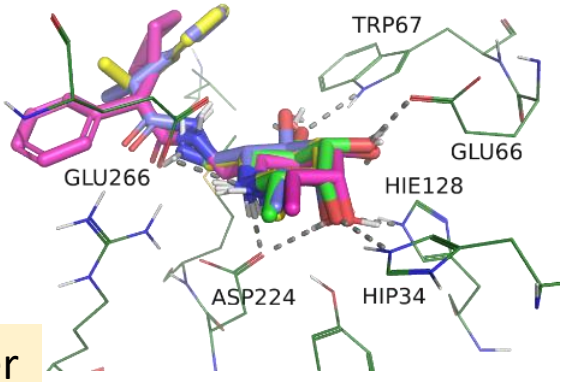
Starting ligand	Target ligand	Similarity of starting and target	Generated molecules most similar to the target one	Similarity to the target ligand	RMSD to the starting ligand, Å
 2HB1 K_i = 160 µM	 2QBS K_i = 210 nM	0.36		1	1.25
				1	1.52
 3S1G K_i = 6500 nM	 3GC4 K_i = 25 nM	0.32		0.63	0.06
 2ZWZ K_i = 16.3 nM	 2ZX9 K_i = 0.054 nM	0.32		0.69	0.86
				0.69	1.03

CReM-dock: fragment expansion

starting ligand
target ligand

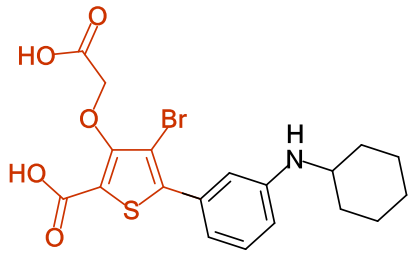


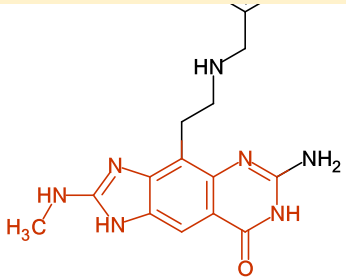


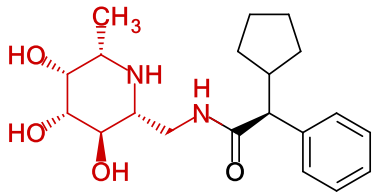


Fragments may grow in a proper direction which was previously explored as active

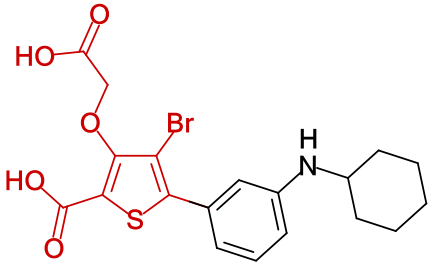
target ligand

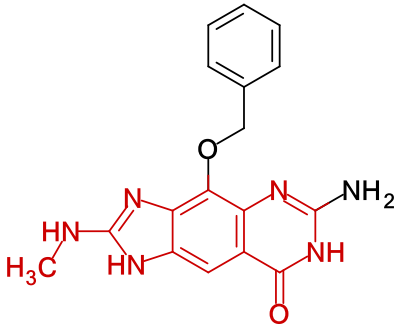


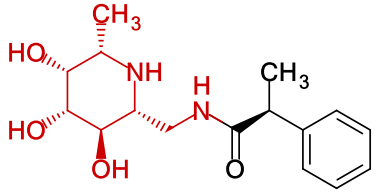




most similar ligand

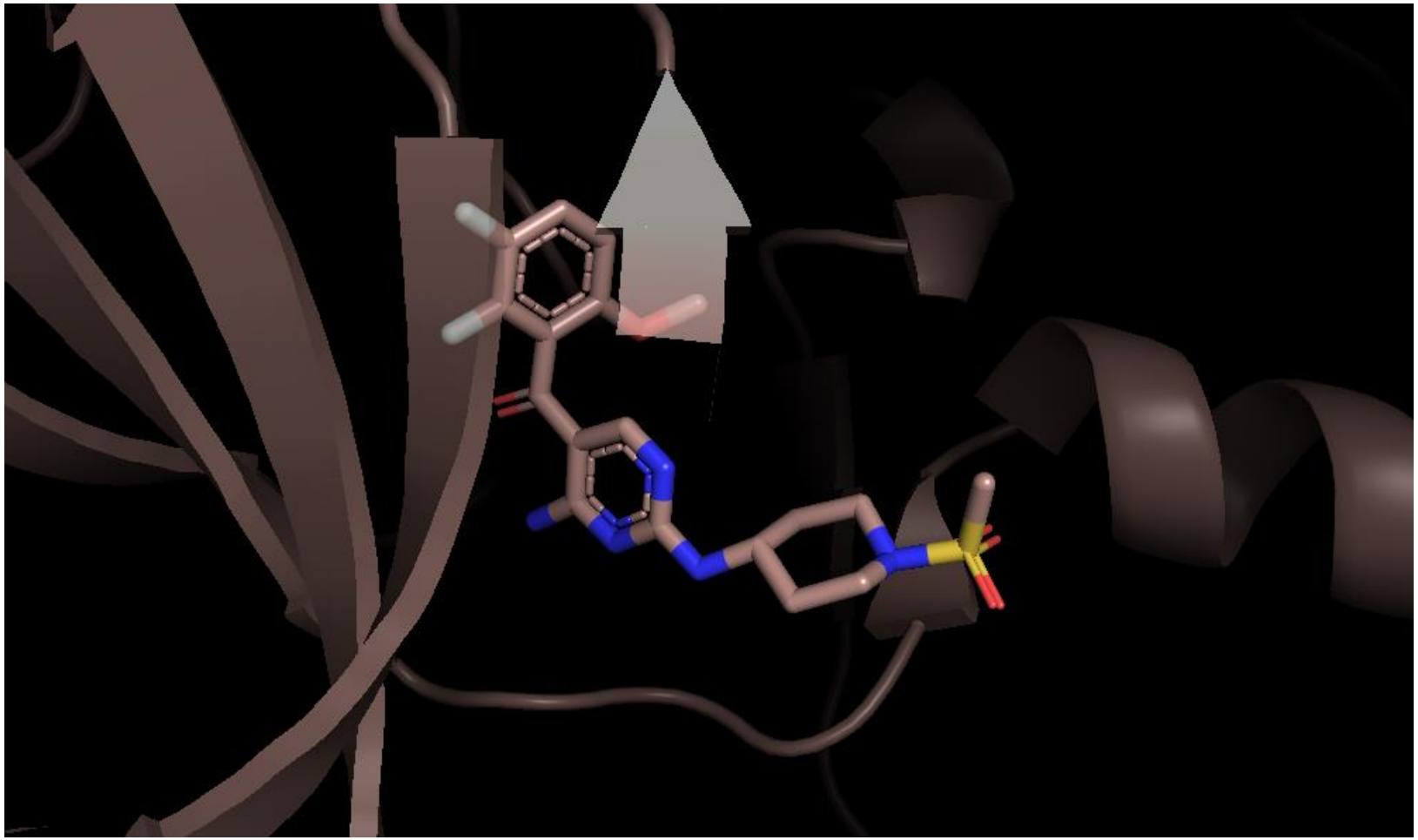






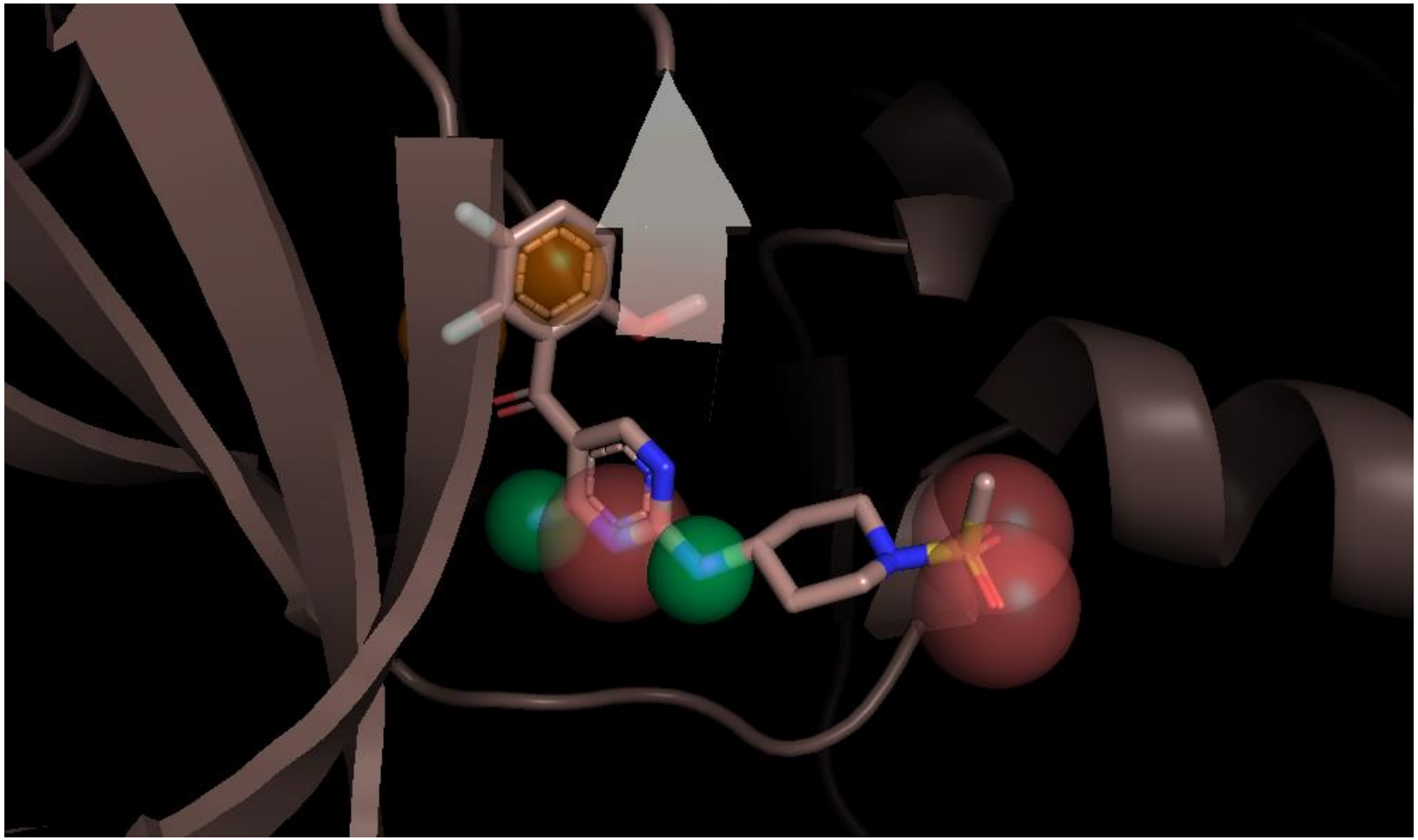
IMTM

CReM-pharm: generation example



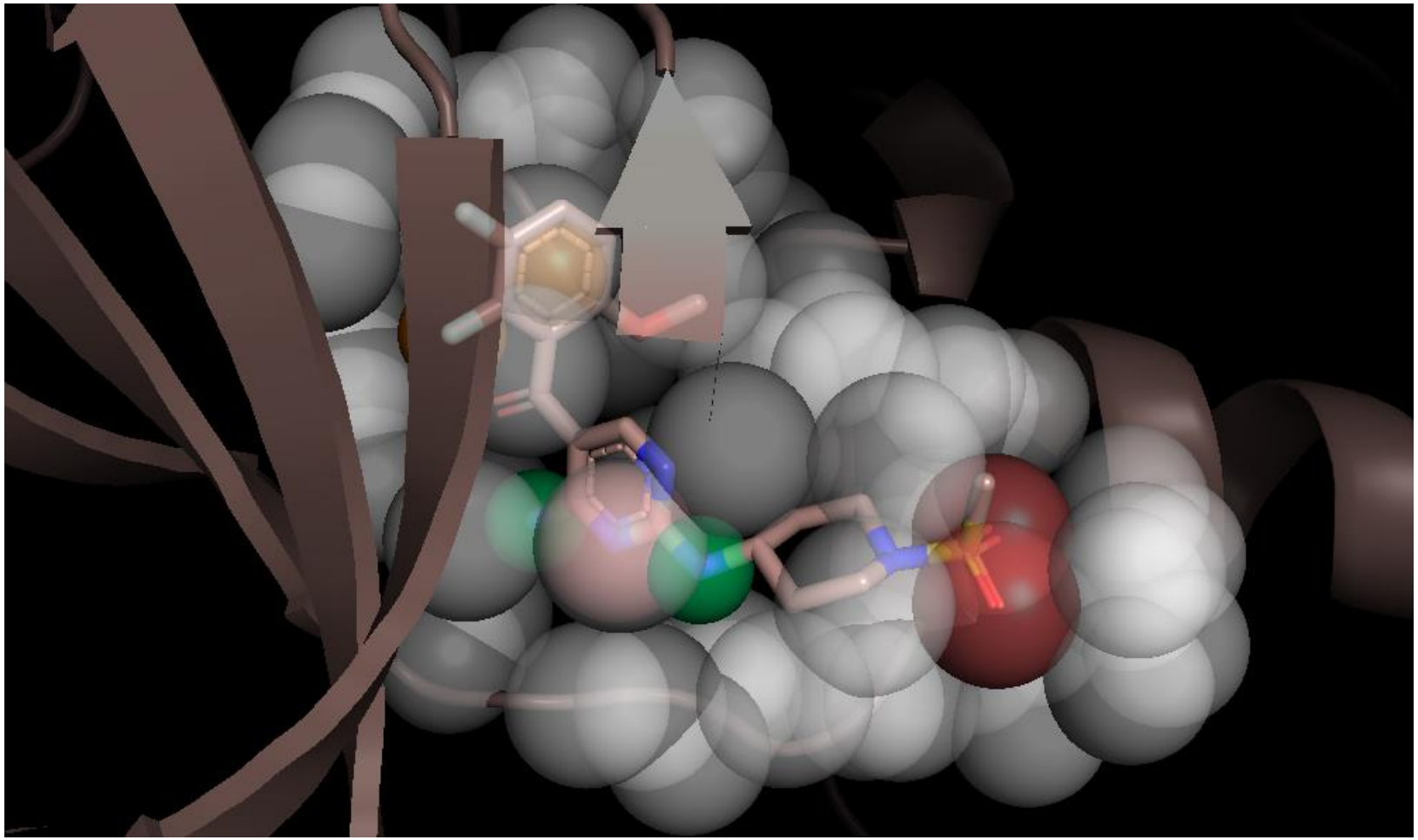
IMTM

CReM-pharm: generation example



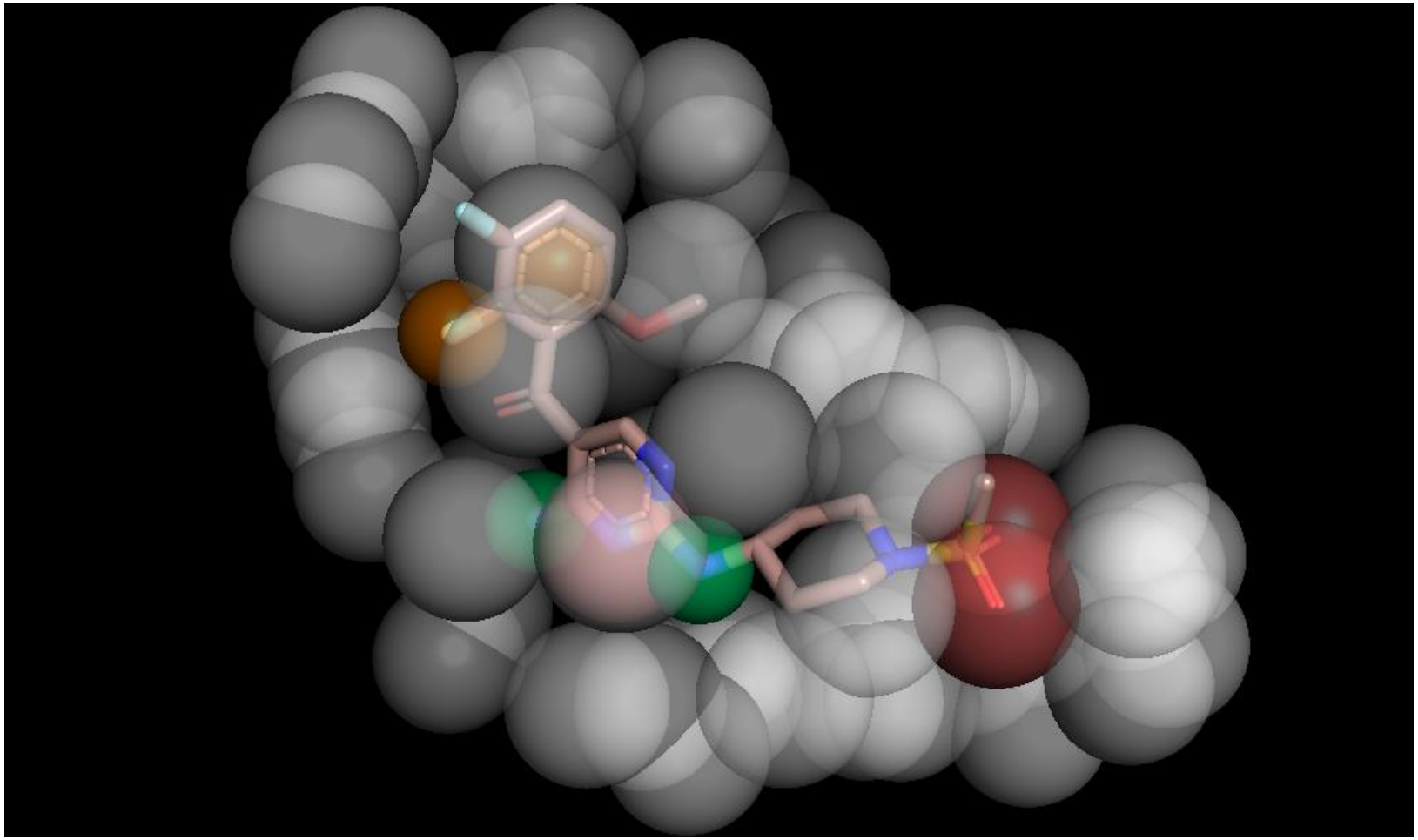
IMTM

CReM-pharm: generation example



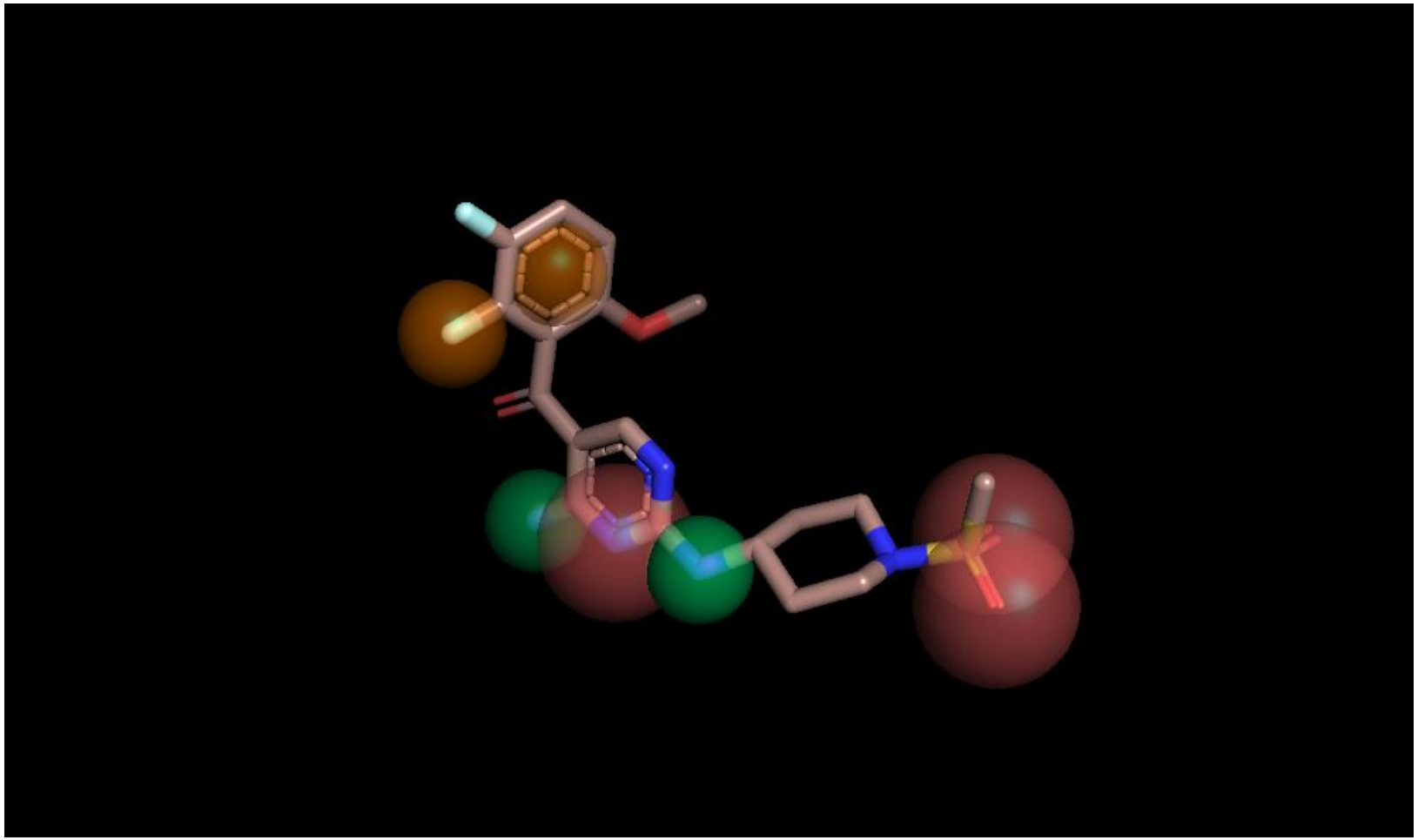
IMTM

CReM-pharm: generation example



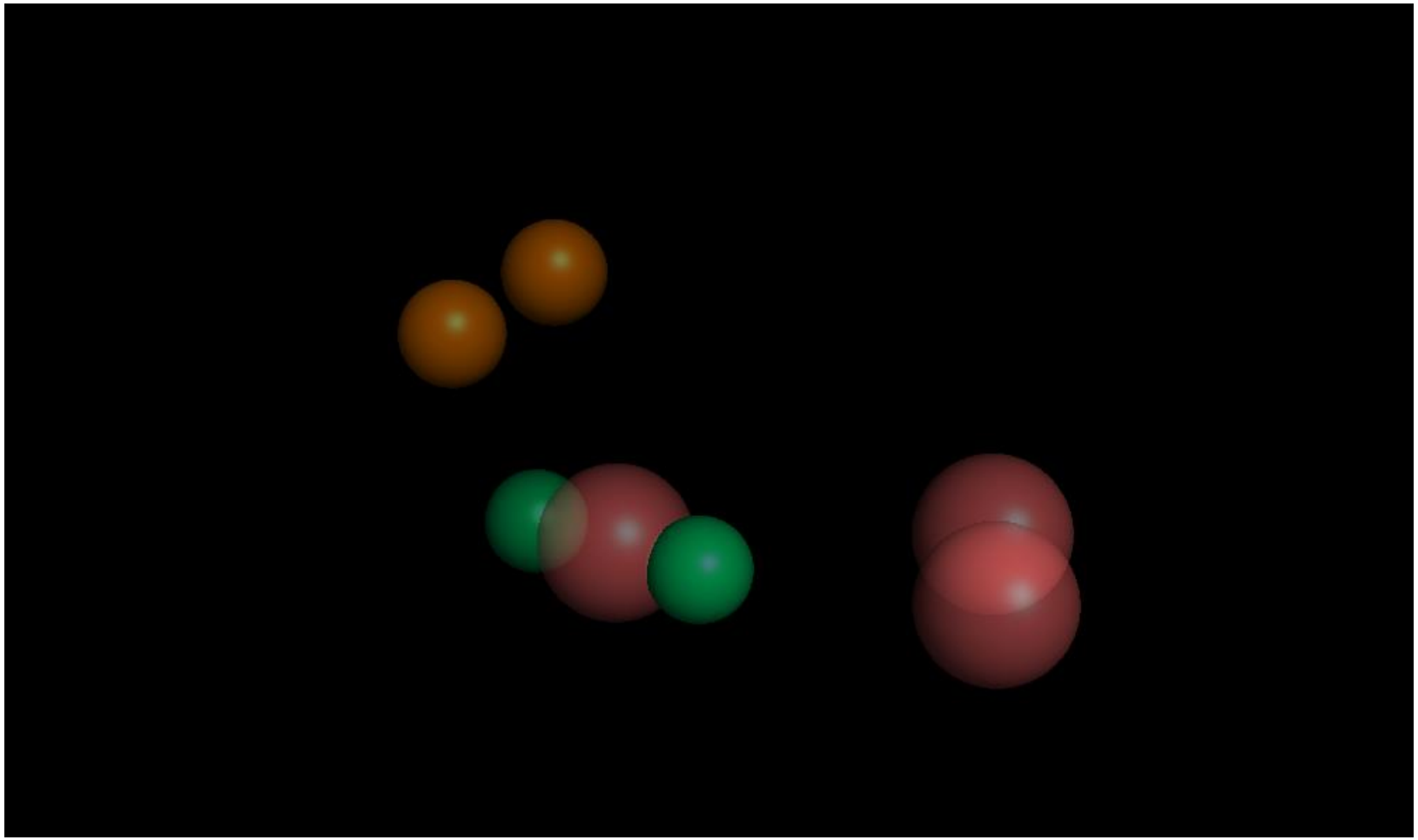
IMTM

CReM-pharm: generation example



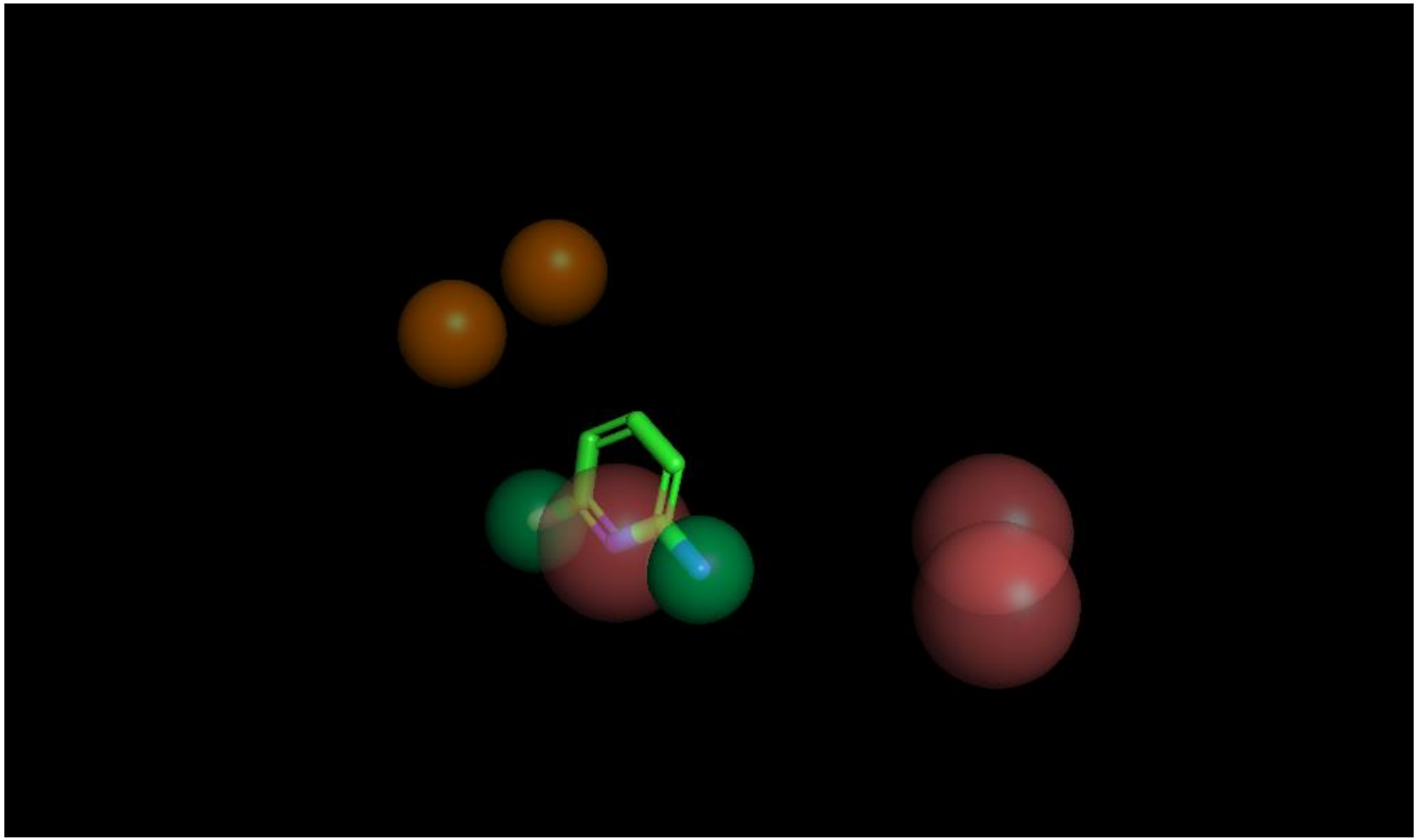
IMTM

CReM-pharm: generation example



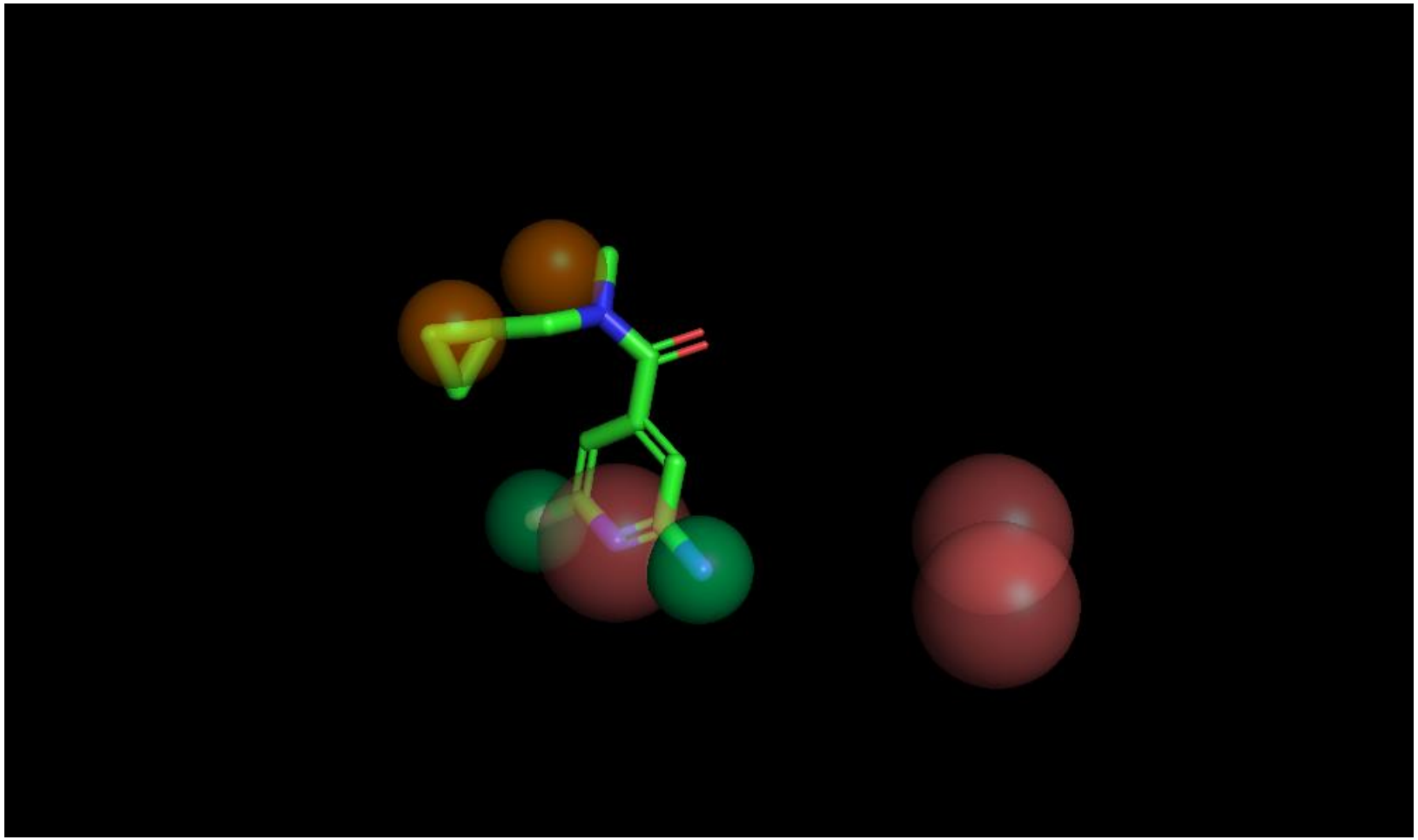
IMTM

CReM-pharm: generation example



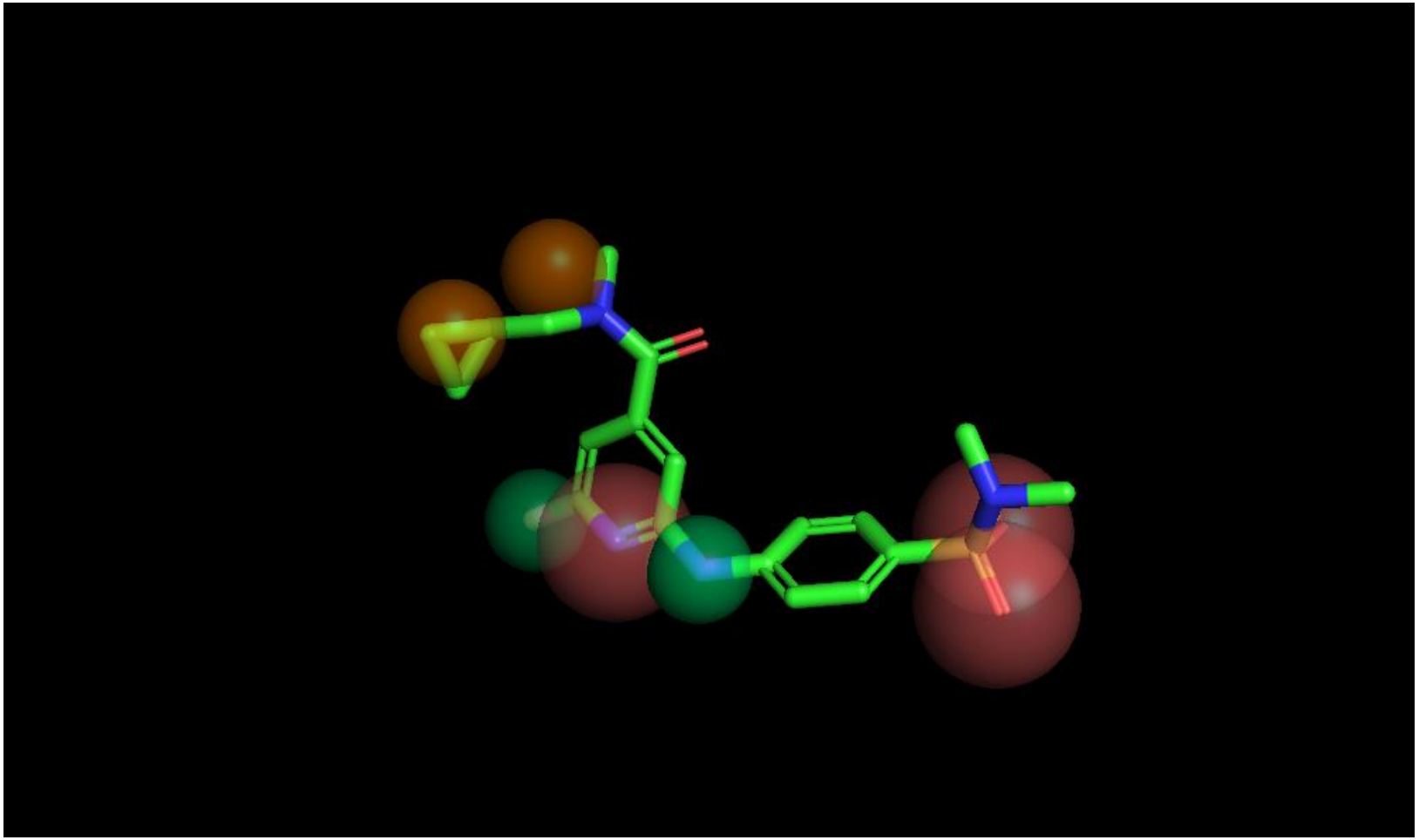
IMTM

CReM-pharm: generation example

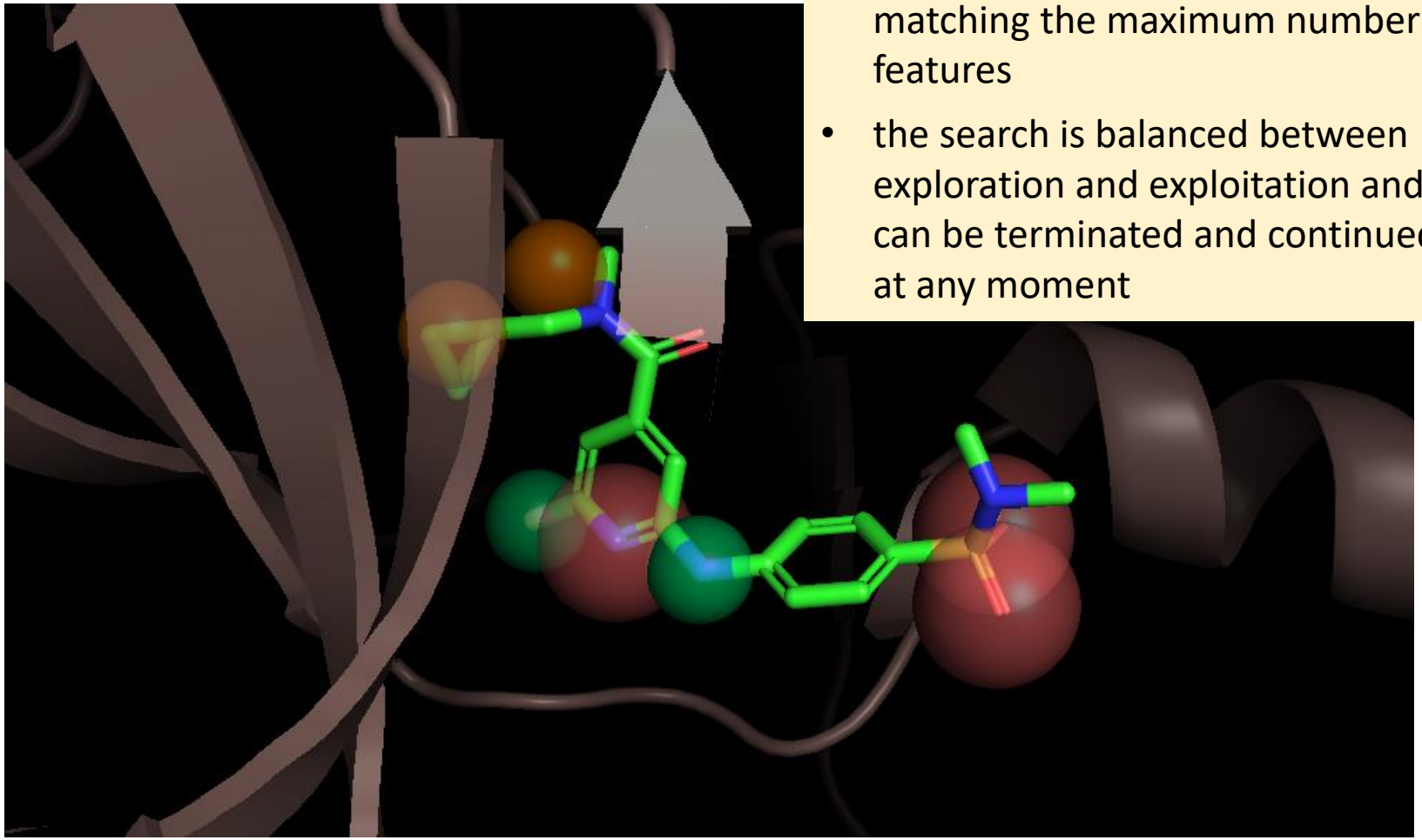


IMTM

CReM-pharm: generation example

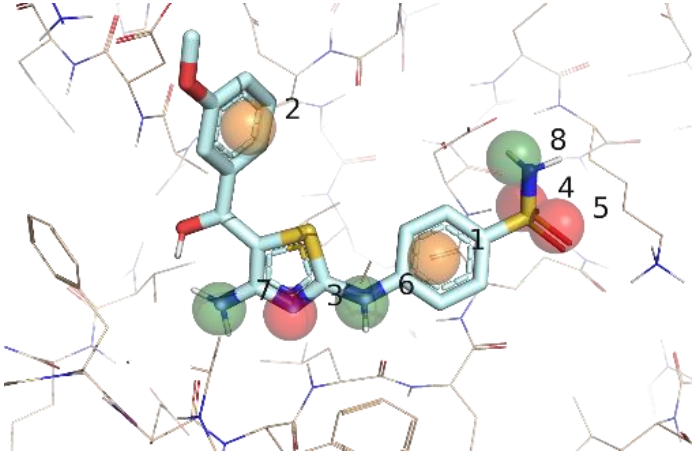


CReM-pharm: generation example



- design minimum-sized molecules matching the maximum number of features
- the search is balanced between exploration and exploitation and can be terminated and continued at any moment

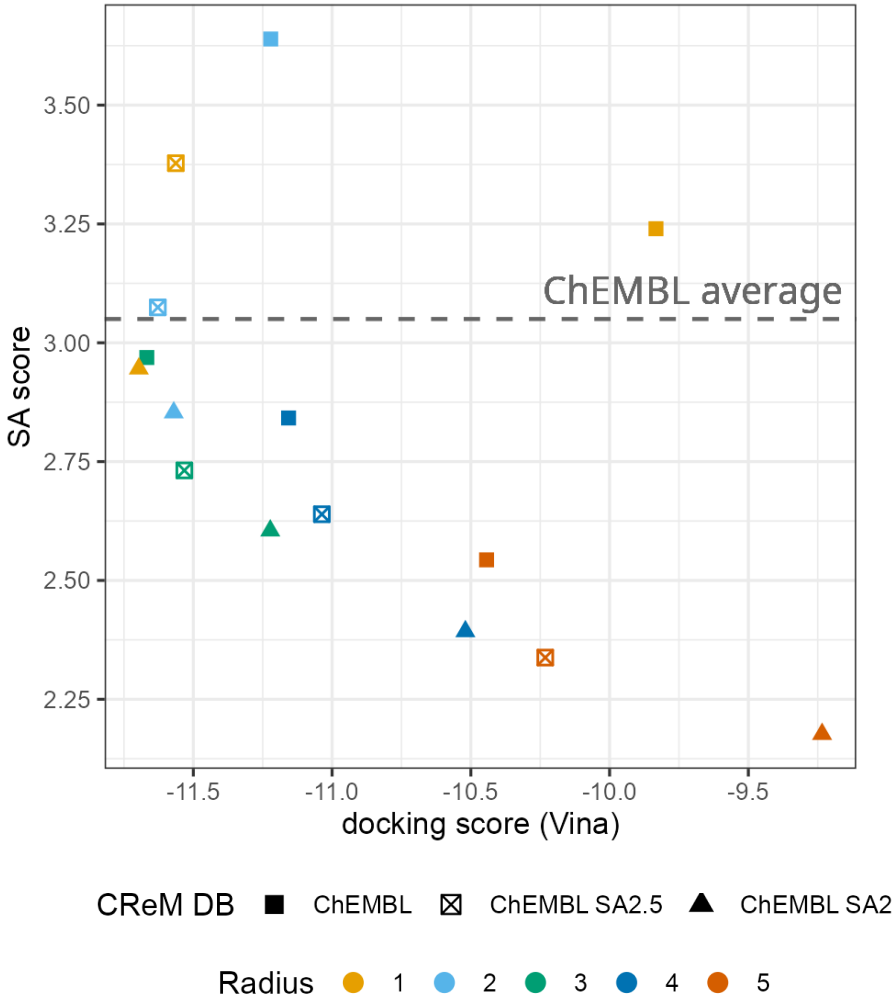
CReM-pharm: CDK2 example



3RAL

Settings:

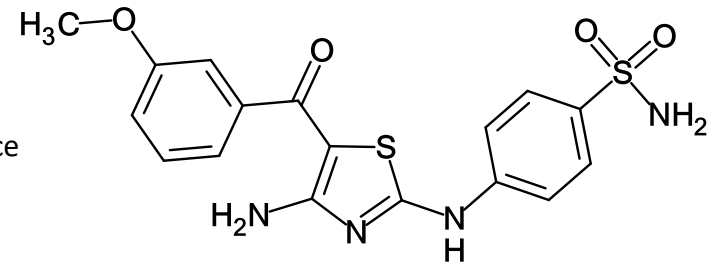
- MW ≤ 450, logP ≤ 4, TPSA ≤ 120, RTB ≤ 7
- maximum number of replacements: all
- top 100 compounds by docking score

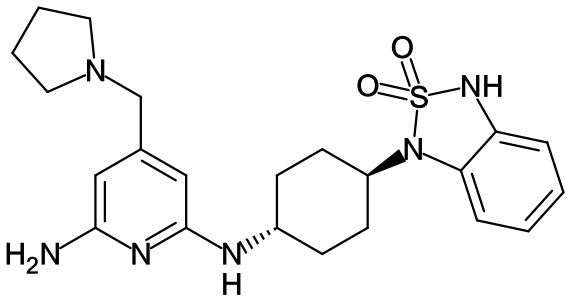


CReM-pharm example

CDK2 (3RAL)

reference ligand

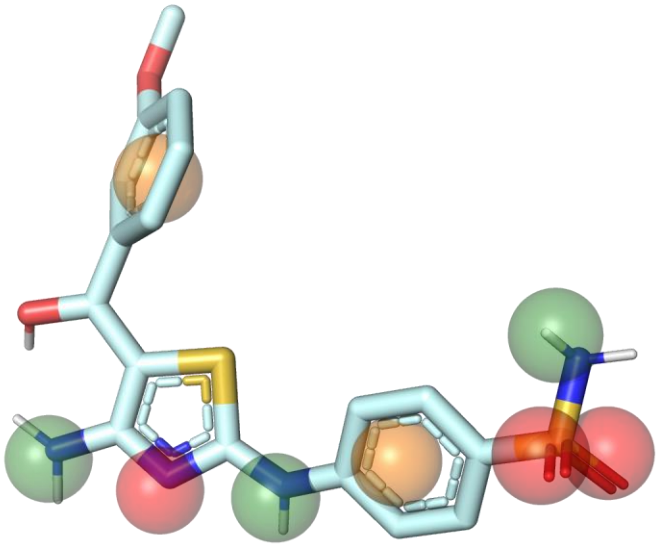


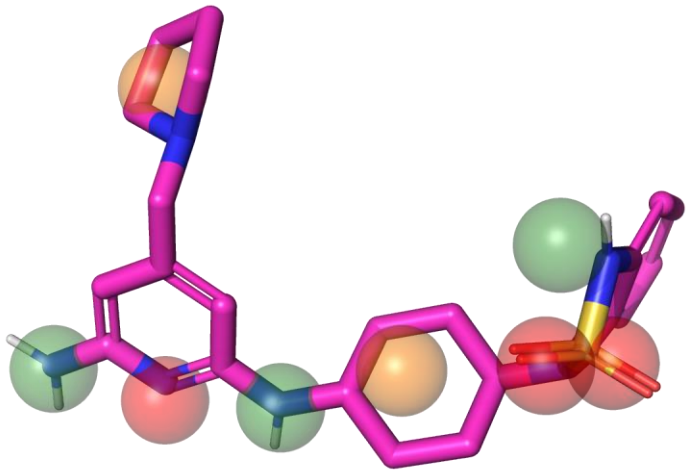


designed ligand

-11.4 / 3.0

docking score / SA score

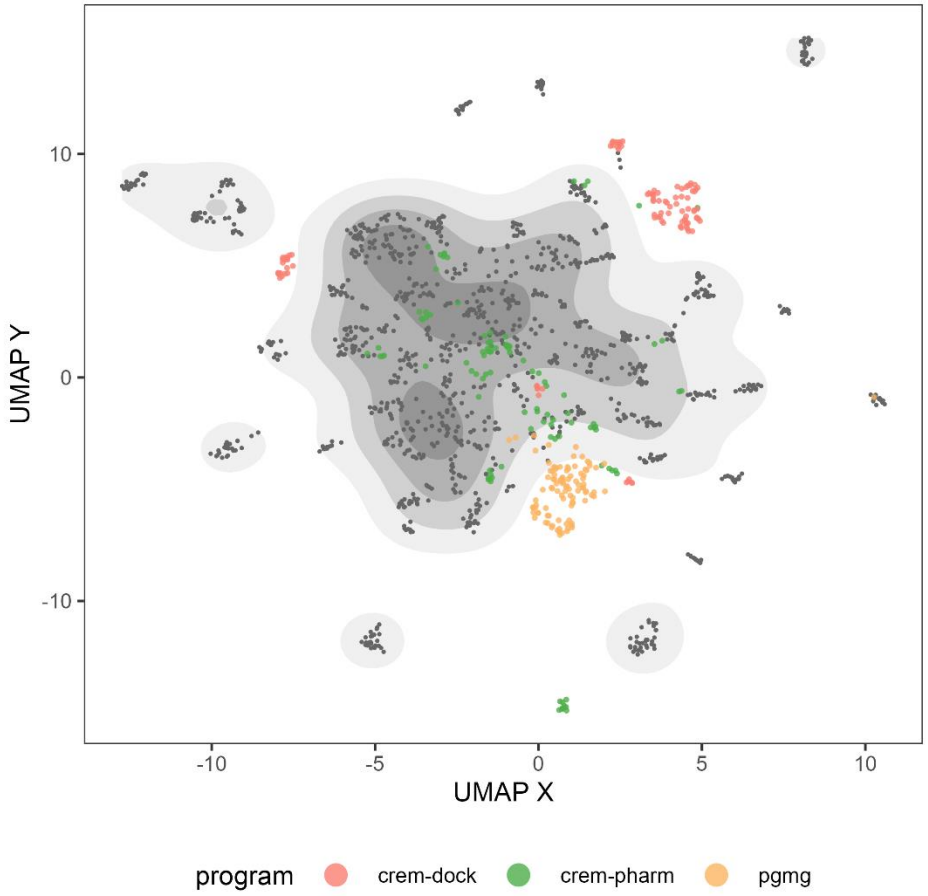




- designed compounds have high docking scores and fit to protein pockets
- SA scores are not very sensitive to complexity of pharmacophore models

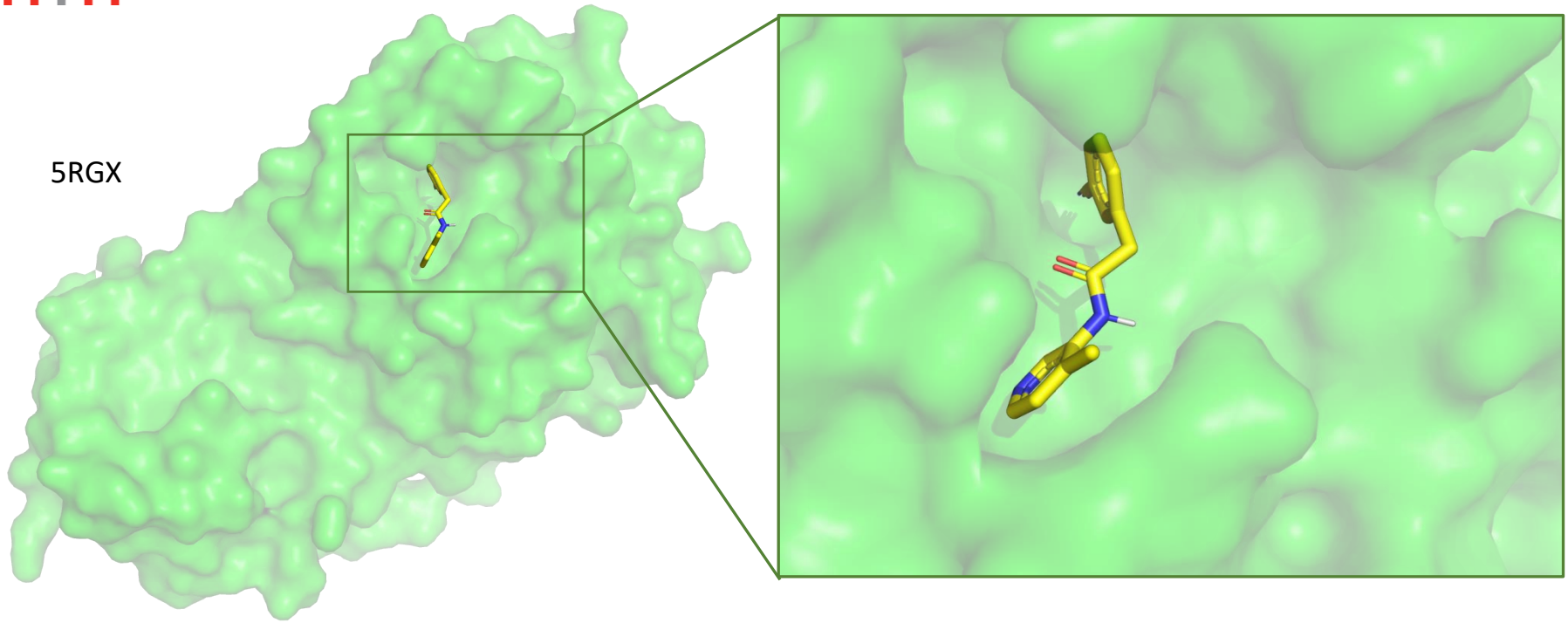
Chemical space overlap CReM-dock and CReM-pharm

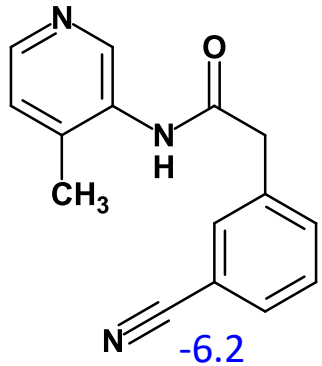
CDK2 (2BTR)



- CReM-dock and CReM-pharm structures generated for the same protein structure do not overlap much. Therefore, it can be suggested to use both approaches to get a greater number of diverse solutions

Hit expansion: inhibitors of main protease SARS-CoV2



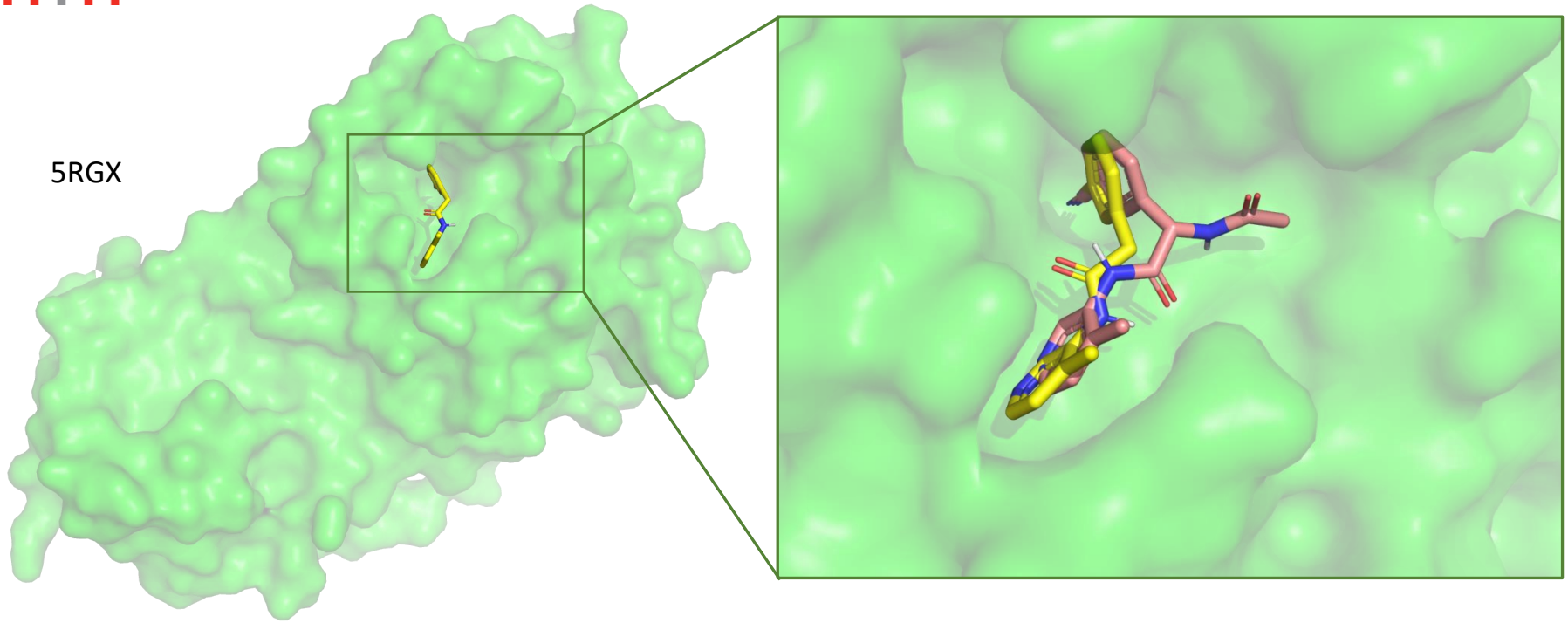


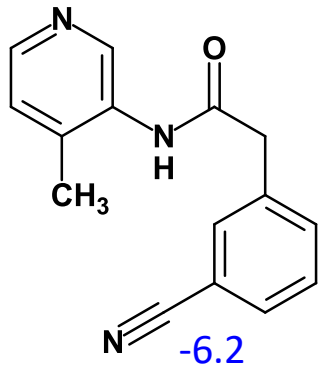
-6.2

docking score (Autodock Vina)

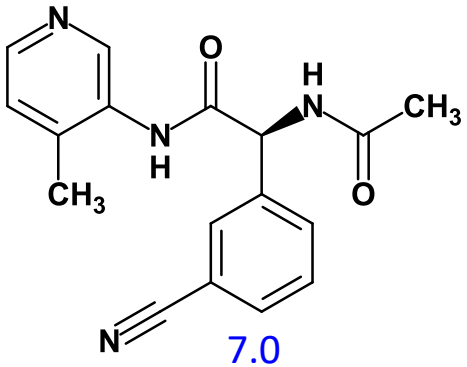
* in collaboration LifeChemicals company (Ukraine) and IOCB (Prague)

Hit expansion: inhibitors of main protease SARS-CoV2





-6.2

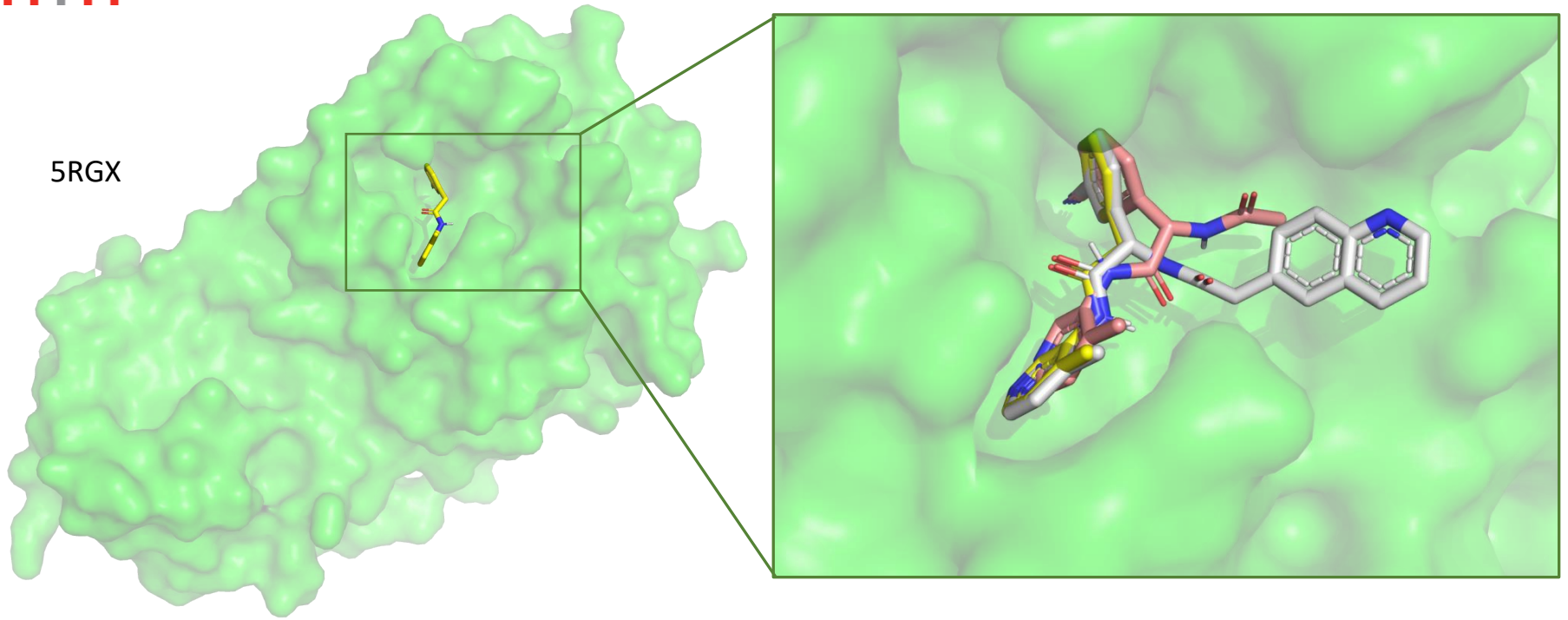


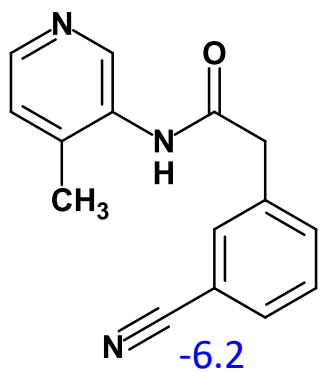
7.0

docking score (Autodock Vina)

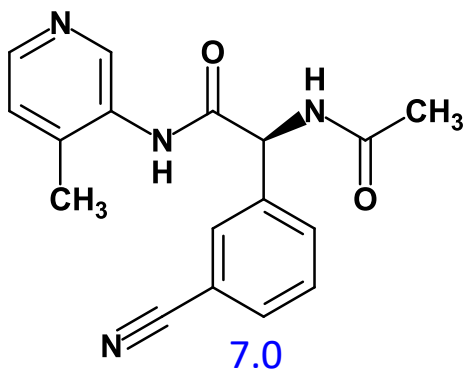
* in collaboration LifeChemicals company (Ukraine) and IOCB (Prague)

Hit expansion: inhibitors of main protease SARS-CoV2

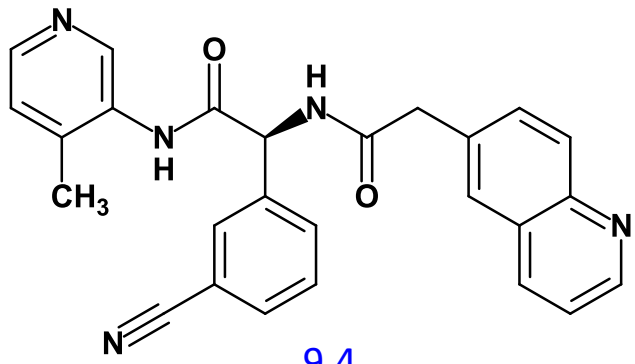




-6.2



7.0

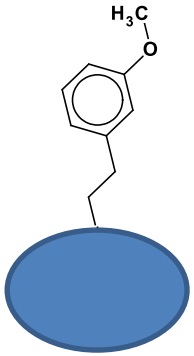


9.4

docking score (Autodock Vina)

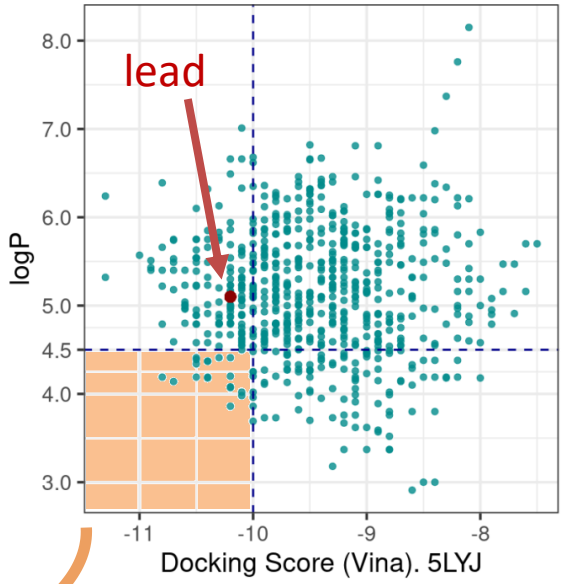
* in collaboration LifeChemicals company (Ukraine) and IOCB (Prague)

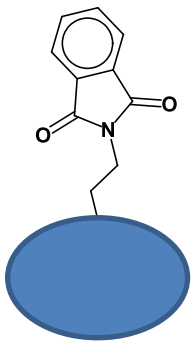
Optimization of tubulin inhibitors



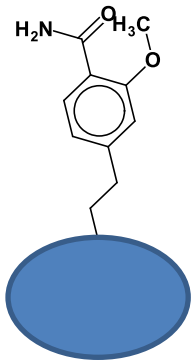
docking score: -10.2

Cell line	IC_{50}, µM
A549	0.033
CCRF-CEM	0.058
CEM-DNR	0.097
HCT116	0.029
HCT116p53-	0.029
K562	0.029
K562-TAX	0.087
U2OS	0.038
BJ	>50

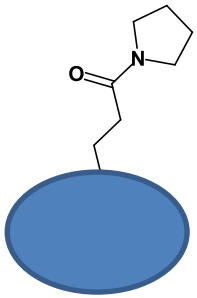




-10.7



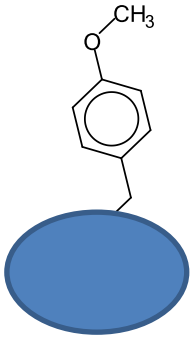
-10.4



-10.0

docking score

Cell line	IC_{50}, µM
A549	8.84
CCRF-CEM	6.46
CEM-DNR	-
HCT116	9.18
HCT116p53-	9.29
K562	2.65
K562-TAX	-
U2OS	6.44
BJ	> 50



Cell line	IC_{50}, µM
A549	0.034
CCRF-CEM	0.018
CEM-DNR	0.029
HCT116	0.017
HCT116p53-	0.021
K562	0.013
K562-TAX	0.030
U2OS	0.018
BJ	>50

* in collaboration with Miroslav Soural group (UPOL)