

7th Advanced In silico Drug Design workshop/challenge

29 January - 2 February 2024
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

SOFTWARE

Open Access



CReM: chemically reasonable mutations framework for structure generation

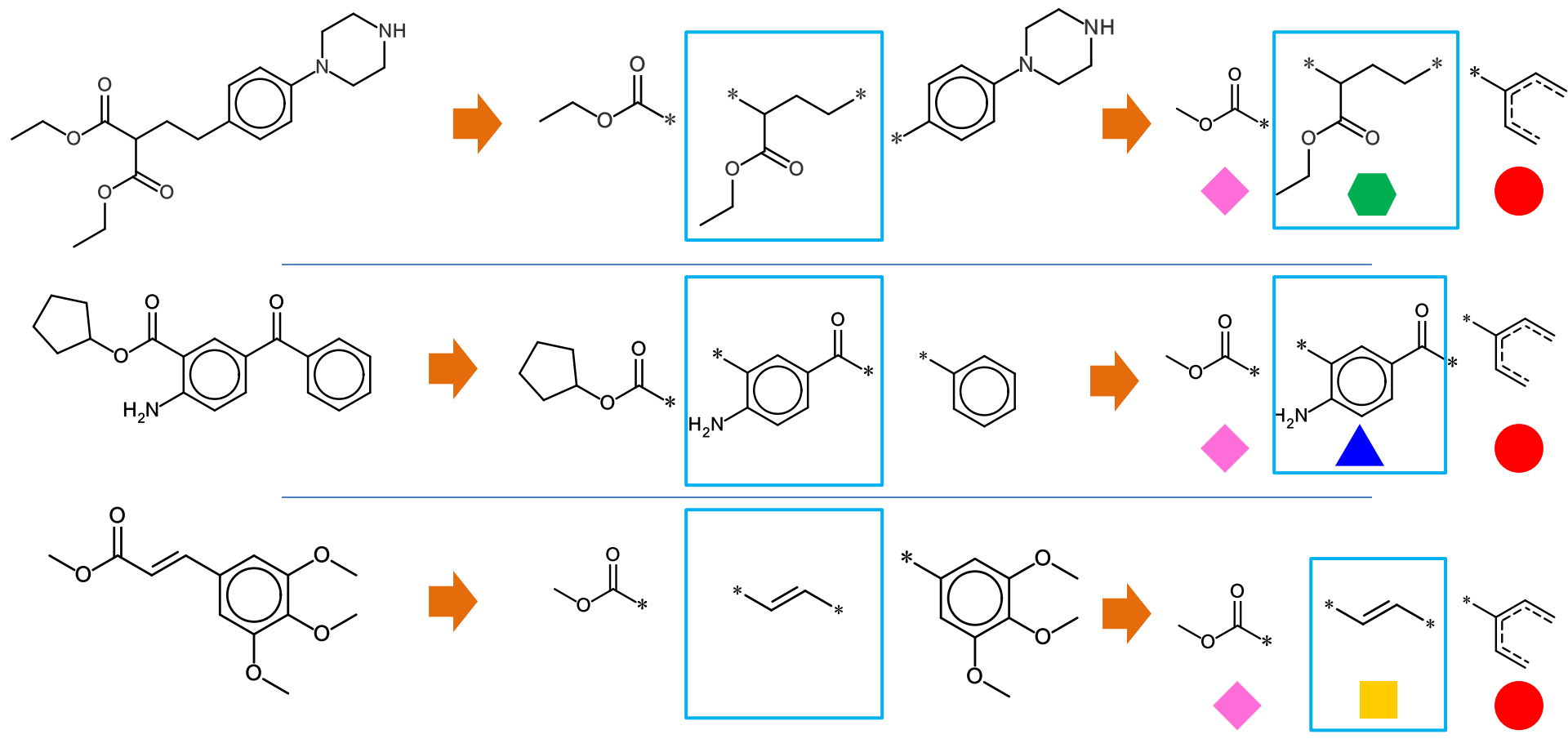
Pavel Polishchuk^{*} 

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 generation that, by design, results in the chemically valid structures and provides flexible control over diversity, novelty, synthetic complexity and chemotypes of generated compounds. The framework was implemented as an open-source Python module and can be used to create custom workflows for the exploration of chemical space.

Keywords: De novo structure generation, De novo design, Matched molecular pairs

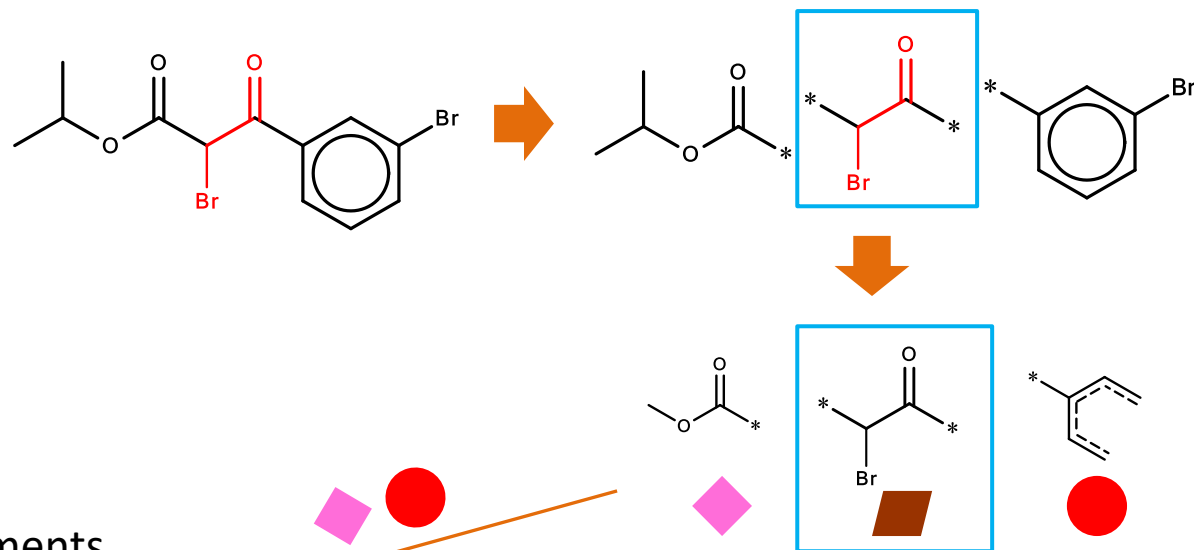
Chemically reasonable mutations (CReM)



DB of replacements

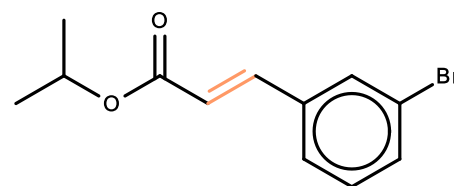
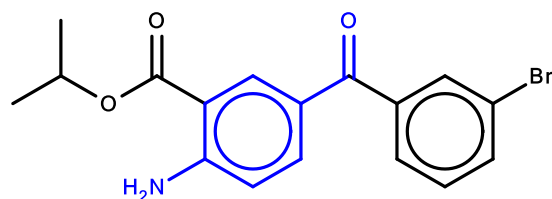
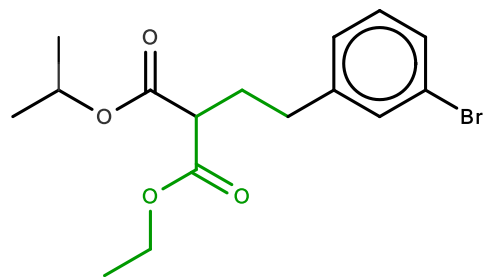
environment (radius = 3)	fragments
 	   ...
...	...

Chemically reasonable mutations (CReM)

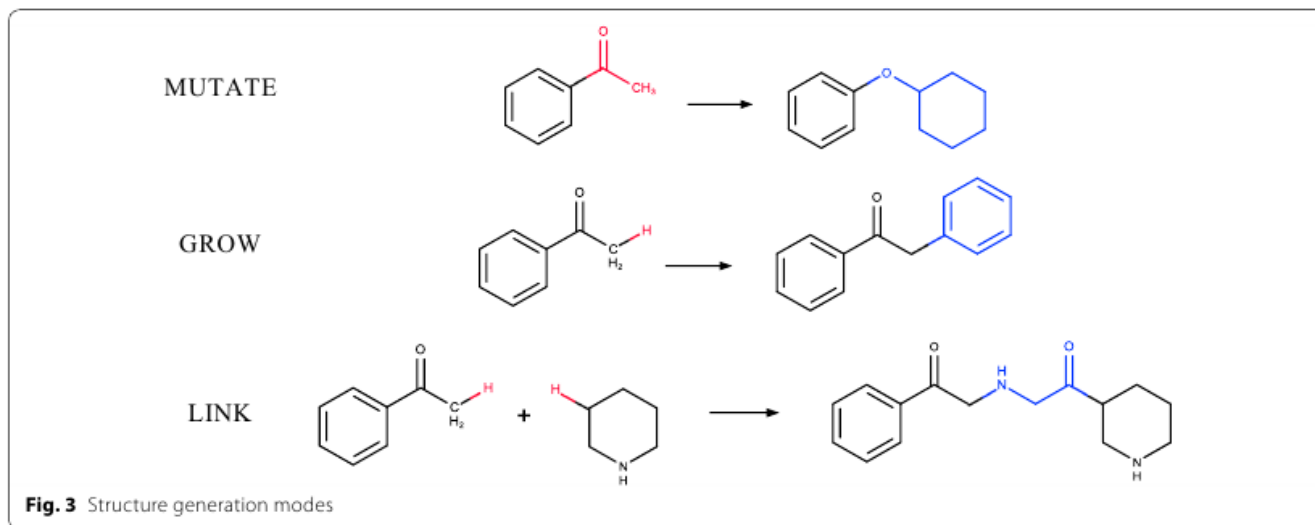
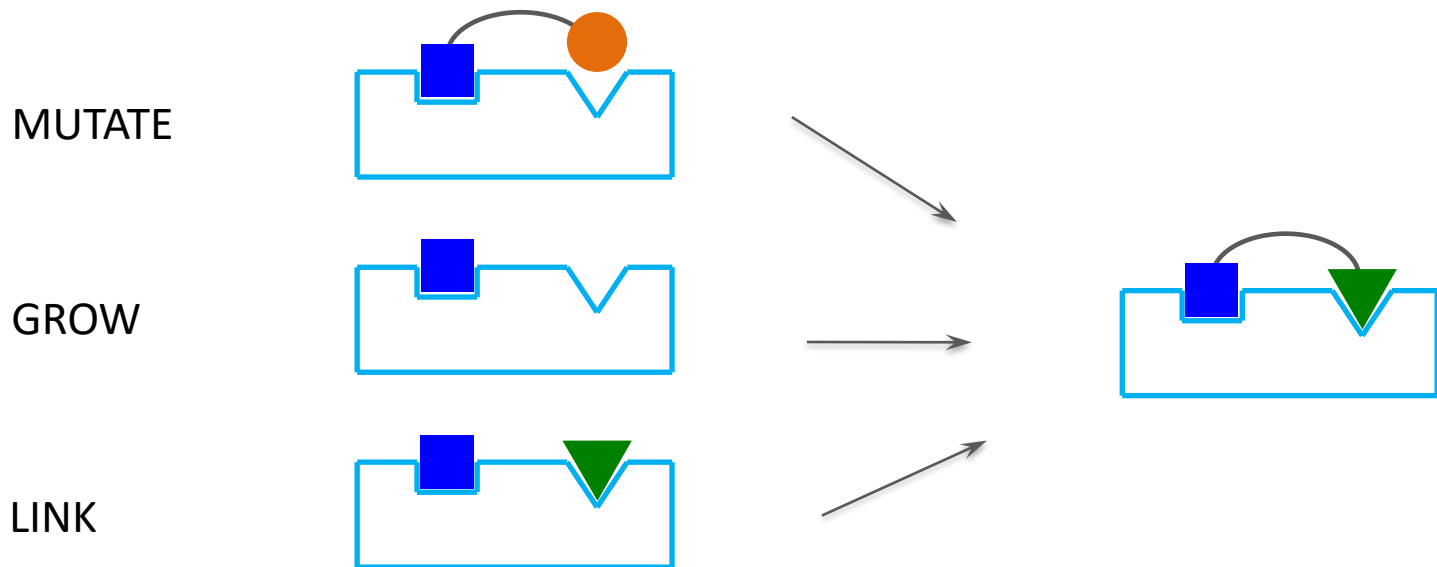


DB of replacements

environment (radius = 3) | fragments



CReM modes



CReM applications

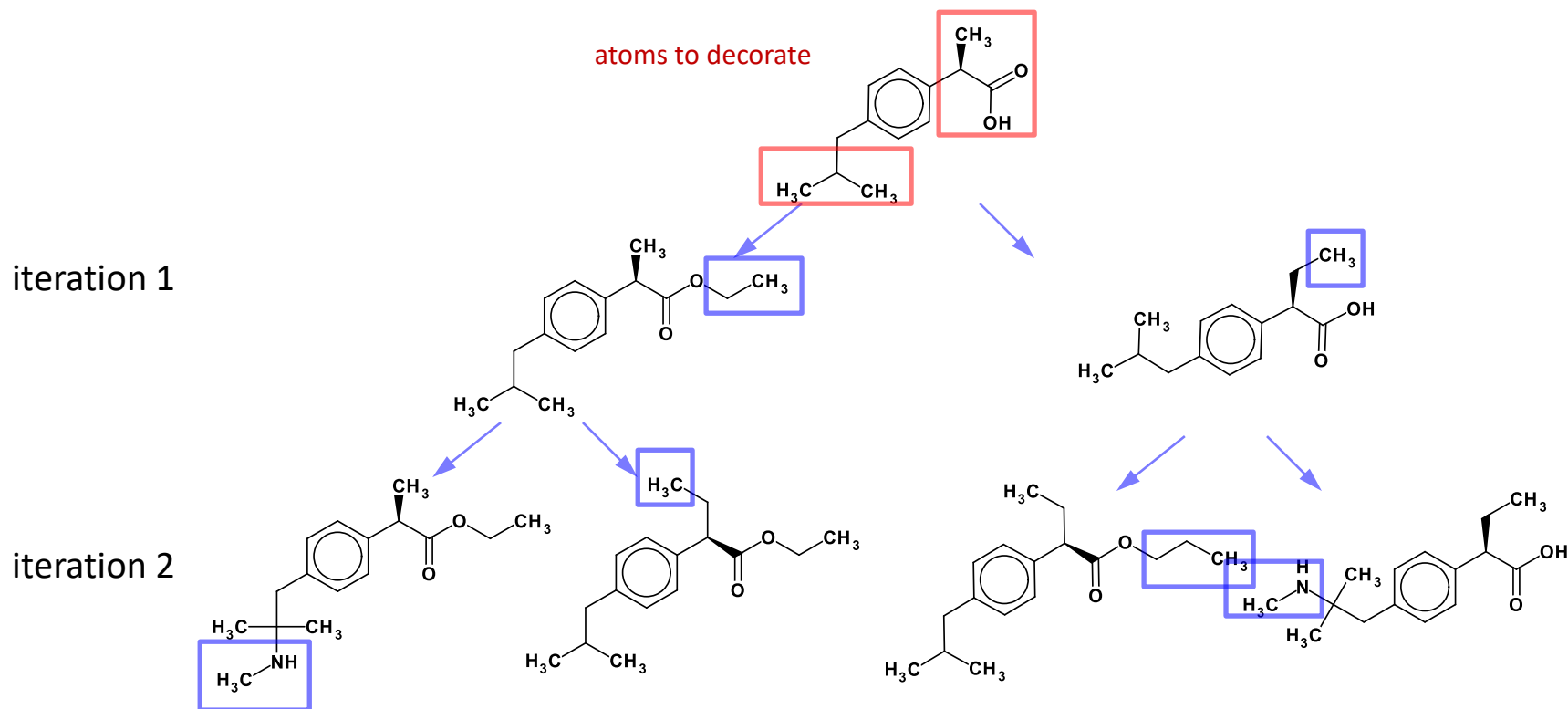
1. Scaffold decoration
2. Enumeration of analog series
3. Hit expansion
4. Lead optimization
5. De novo design

Scaffold decoration

Sequential apply of GROW mode to the seed molecule.

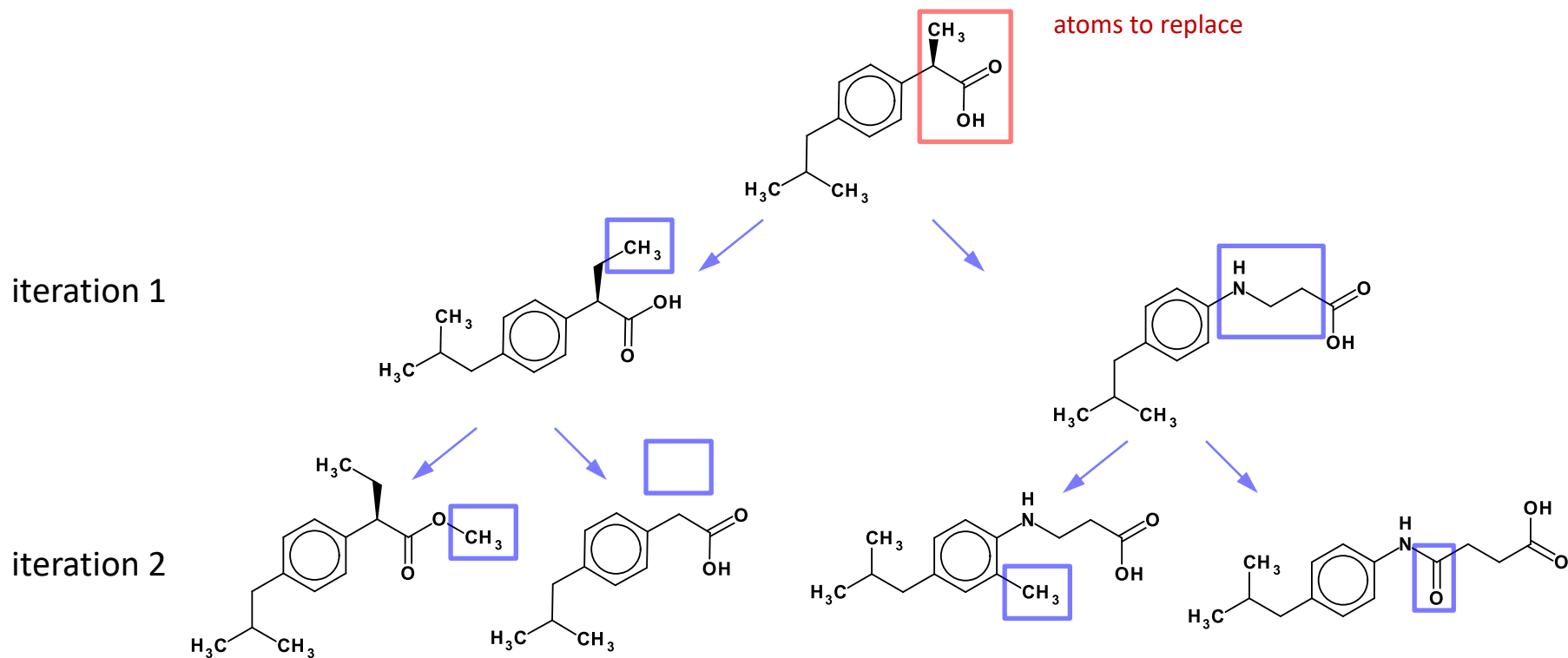
Particular positions of the scaffold can be protected from substitution.

Size of added fragments can be adjusted.



Enumeration of analog series

Sequential apply of MUTATE mode to the seed molecule.
Particular atoms (scaffold) can be protected from replacements.
Size of replaced and replacing fragments can be adjusted.



<http://crem.imtm.cz>

CReM web app

<http://crem.imtm.cz>

Valid license cannot be found

Mode:
Scaffold mode - fragments will be added to selected atoms

Fragment database:
SA <= 2

Databases prepared from compounds of ChEMBL22 with synthetic accessibility score below 2 or below 2.5. Additionally fragments were filtered by structural alerts.

Context radius: 3

Radius of chemical context considered during replacement/attachment, greater radius results in more synthetically relevant modifications but less numerous.

Maximum number of replacement: 5

Number of randomly chosen modifications to apply. The number of actually applied modifications can be less (e.g. due to a small number of available fragments for certain contexts, etc).

Maximum fragment size: 10

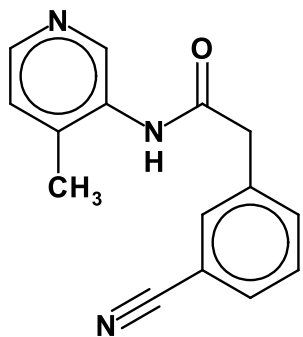
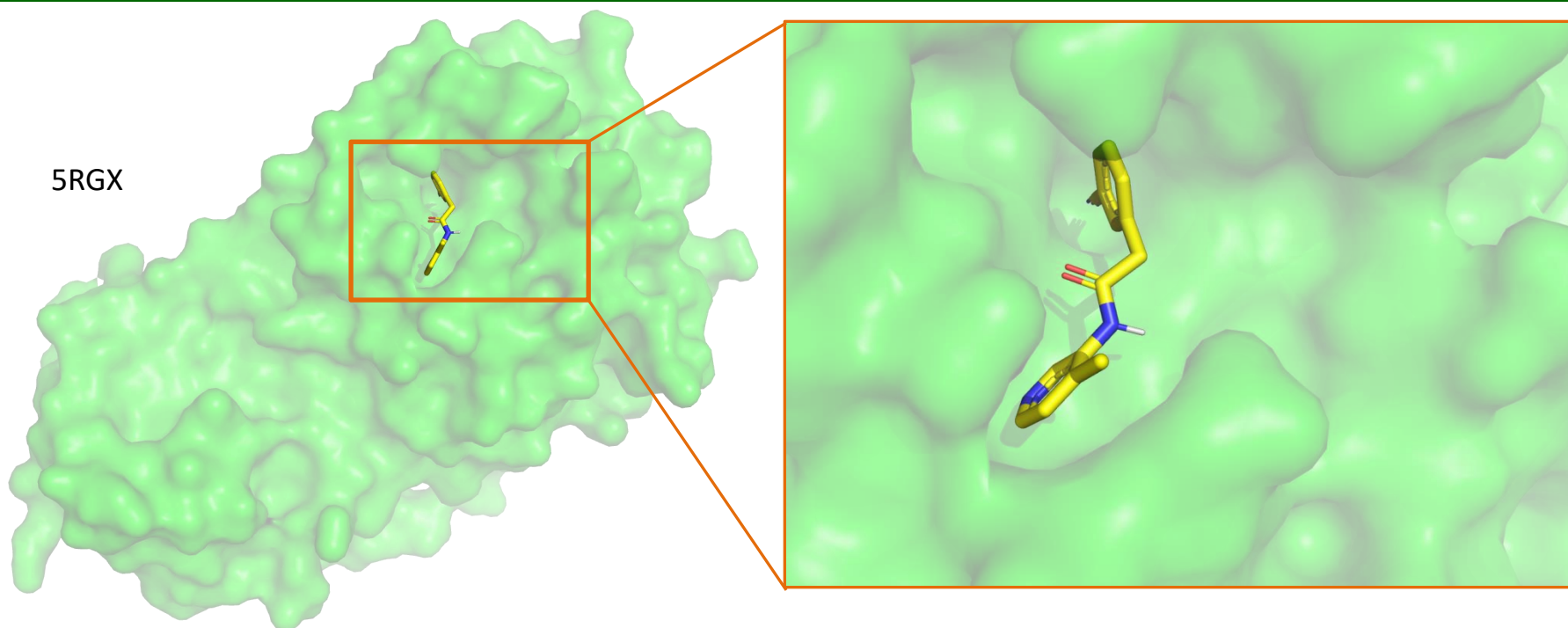
Maximum size of an attached/replaced fragment.

Output SMILES:

```
CC(=O)OCCc1ccccc1
N#Cc1ccc(NCc2ccccc2)cc1
CCC(C)NC(=O)CCCc1ccccc1
C(=NOCc1ccccc1)c1ccccc1
```

Generate structures

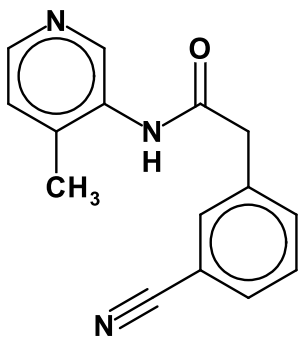
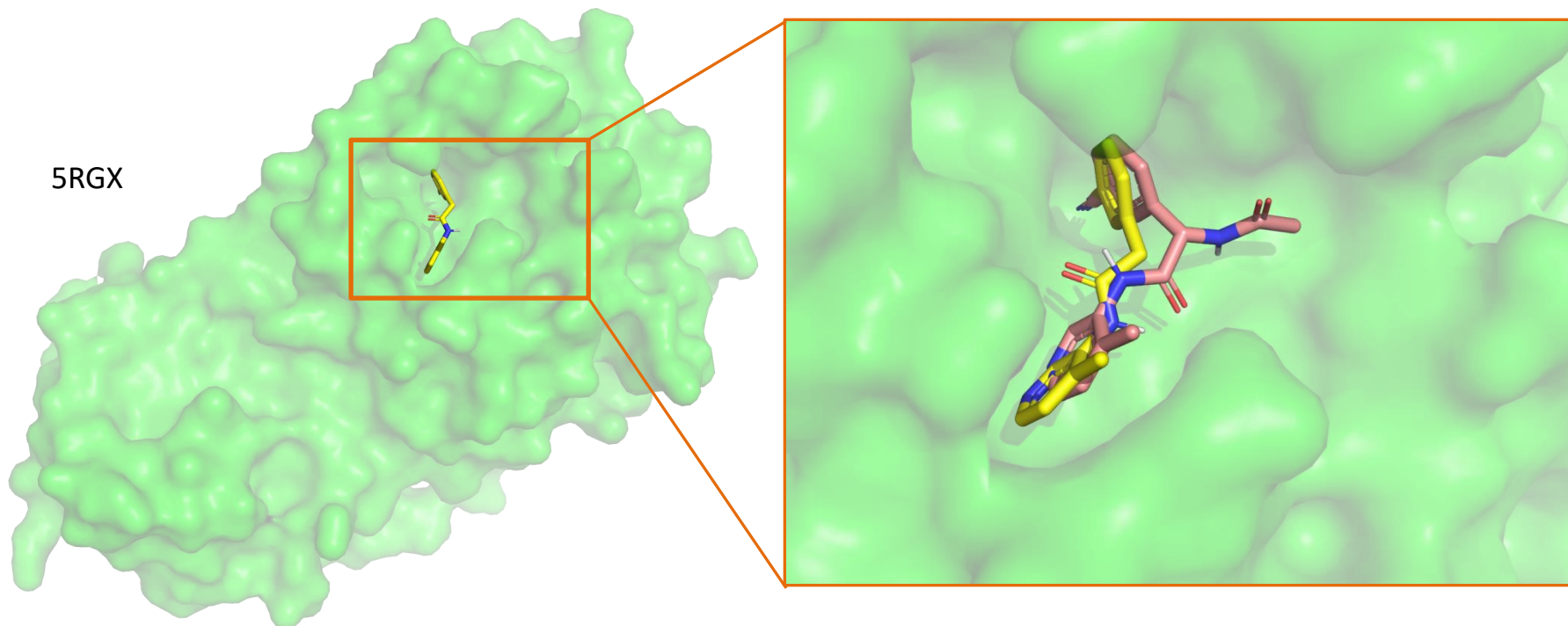
Hit expansion: inhibitors of main protease SARS-CoV2



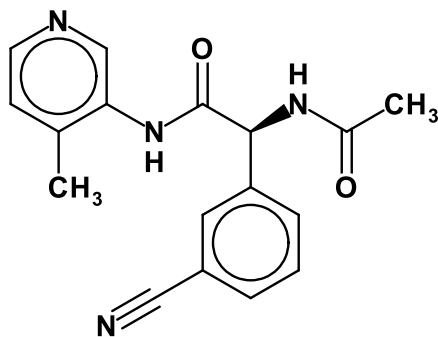
-6.2

docking score (Autodock Vina)

Hit expansion: inhibitors of main protease SARS-CoV2



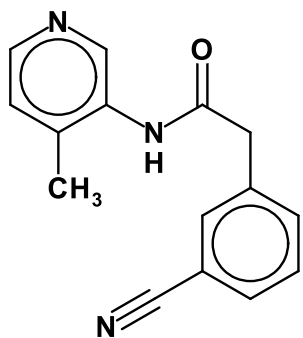
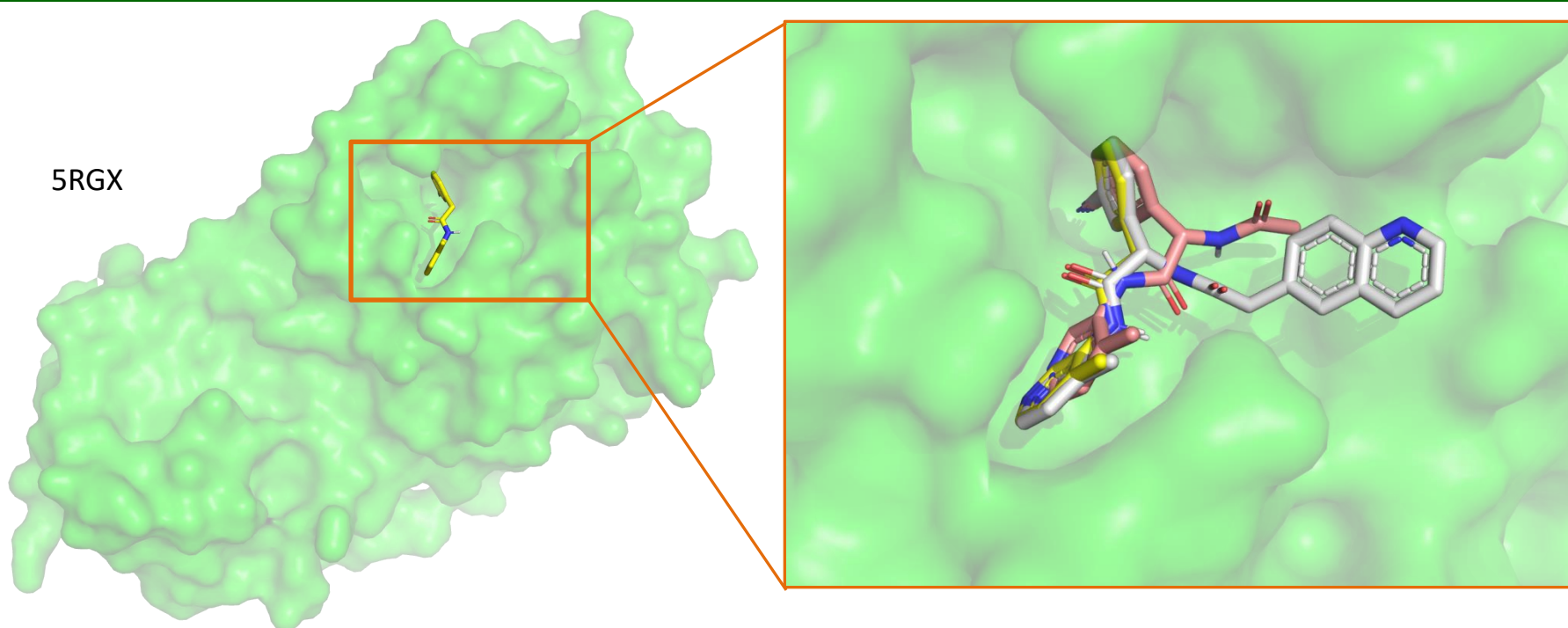
-6.2



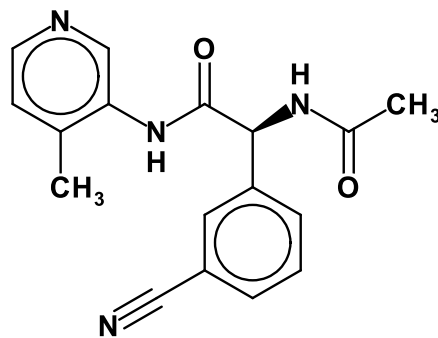
-7.0

docking score (Autodock Vina)

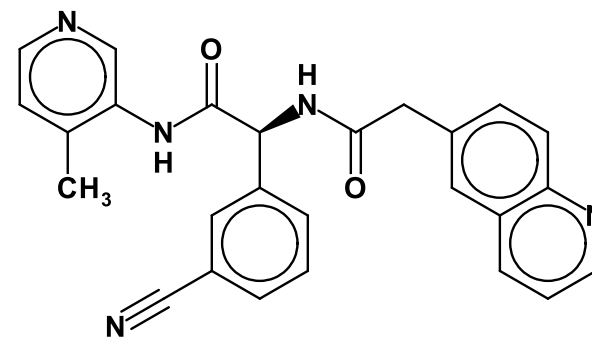
Hit expansion: inhibitors of main protease SARS-CoV2



-6.2



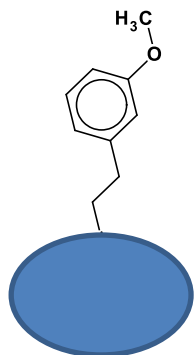
-7.0



-9.4

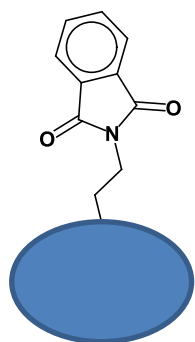
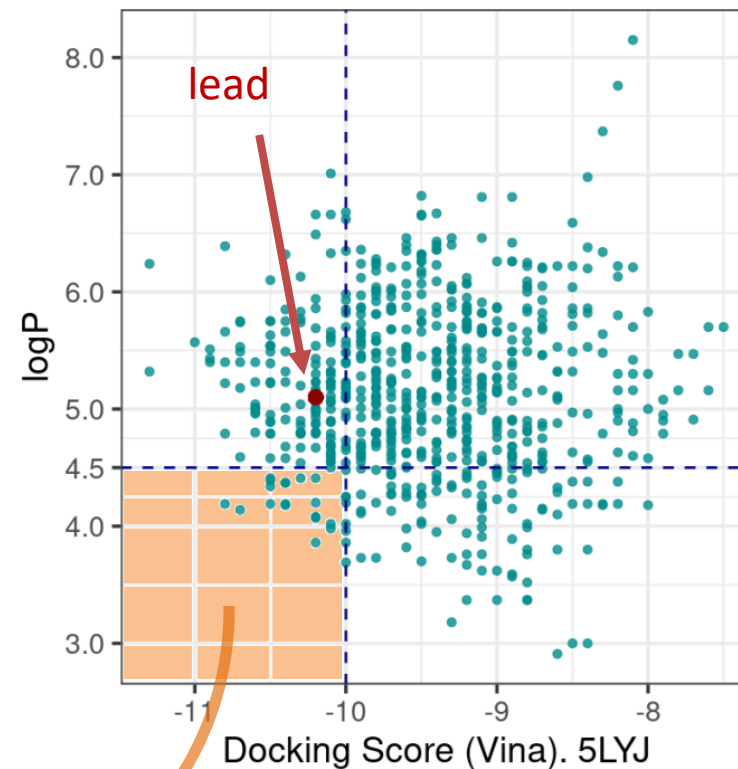
docking score (Autodock Vina)

Lead optimization: tubulin inhibitors

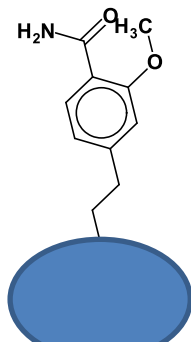


docking score: -10.2

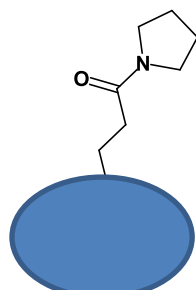
	IC ₅₀
A549	0.033
CCRF-CEM	0.058
CEM-DNR	0.097
HCT116	0.029
HCT116p53-	0.029
K562	0.029
K562-TAX	0.087
U2OS	0.038



-10.7



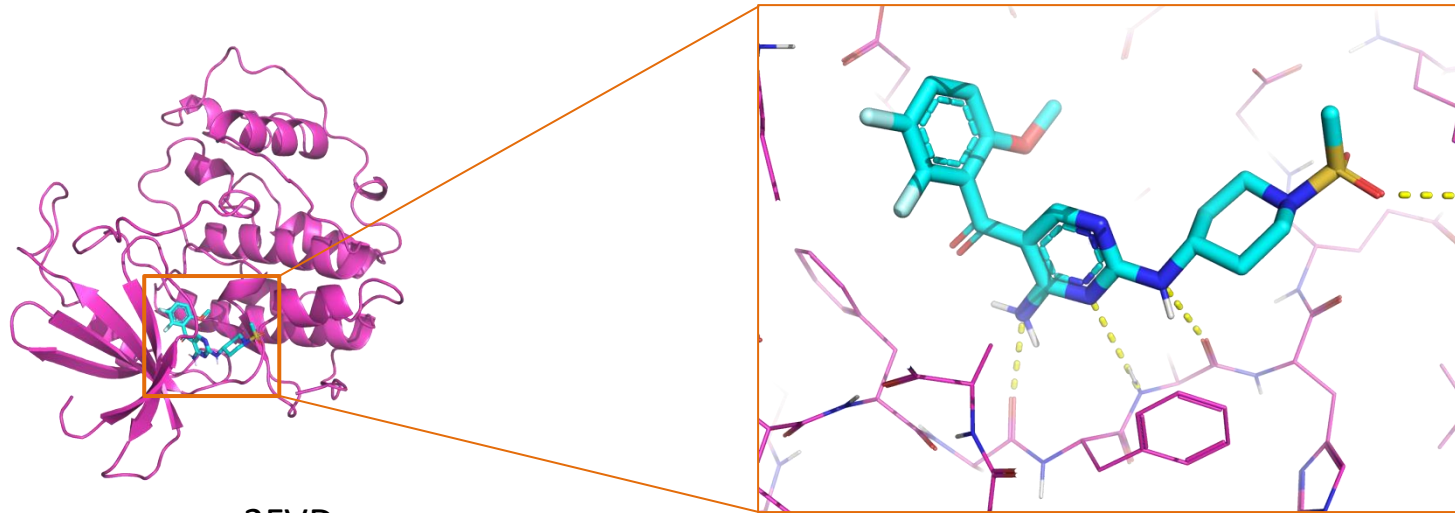
-10.4



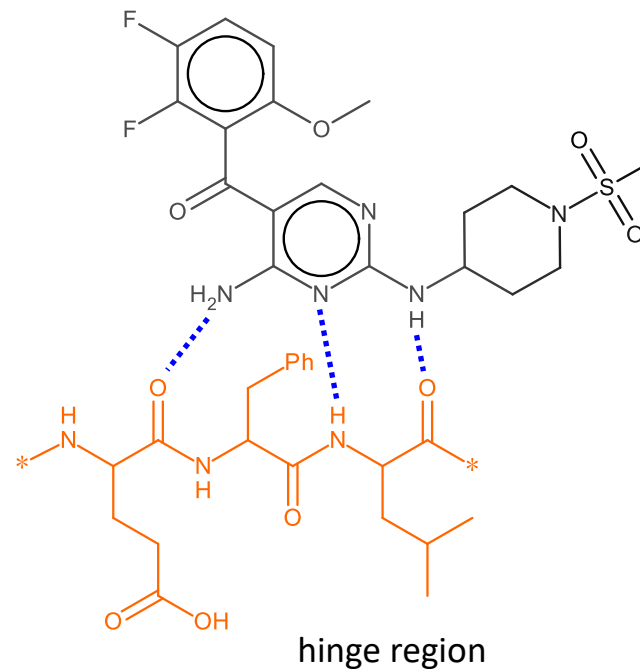
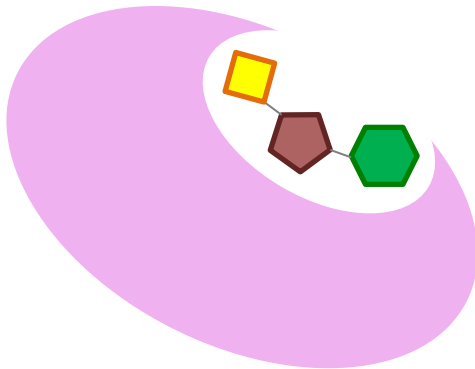
-10.0

docking score

De novo design using docking (CDK2 inhibitors)

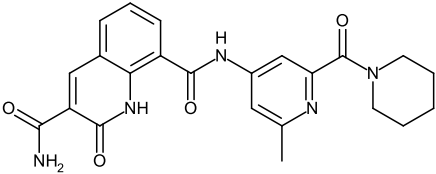
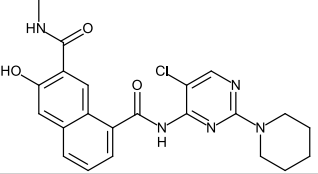
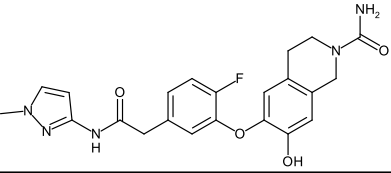
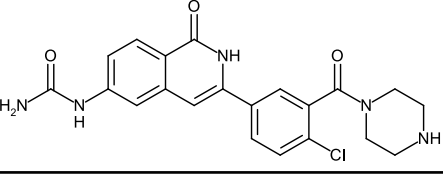
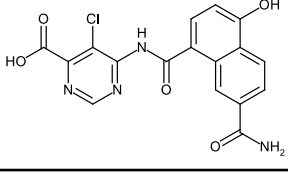


2FVD
 $K_i = 3 \text{ nM}$
docking score = -8.7



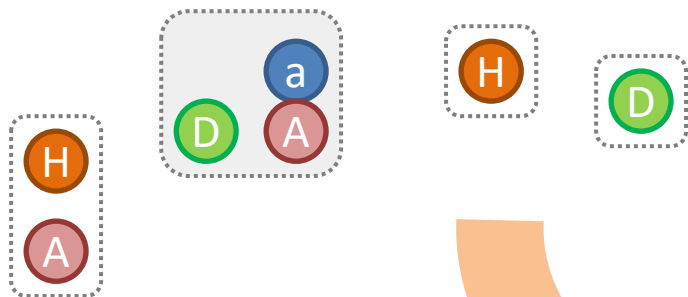
De novo design using docking (CDK2 inhibitors)

Examples of top structures with SA < 3, logP < 4, QED > 0.5

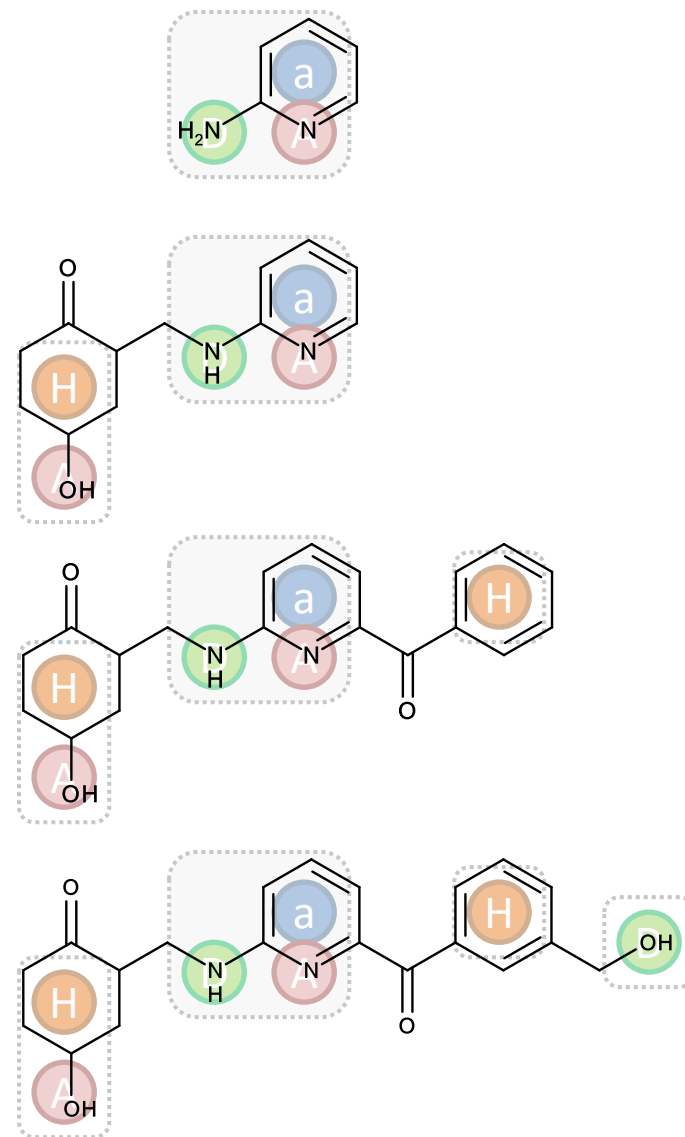
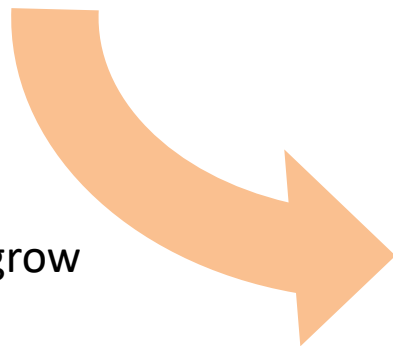
smi	id	Docking score	logP	MW	QED	SA
	002-009344-00	-11.6	2.21	433	0.579	2.48
	002-016242-00	-11.4	3.59	440	0.573	2.49
	002-022716-00	-11.2	2.68	439	0.564	2.7
	002-006301-00	-10.9	2.38	425	0.514	2.48
	002-012455-00	-10.8	2.04	386	0.534	2.49

De novo design using pharmacophores

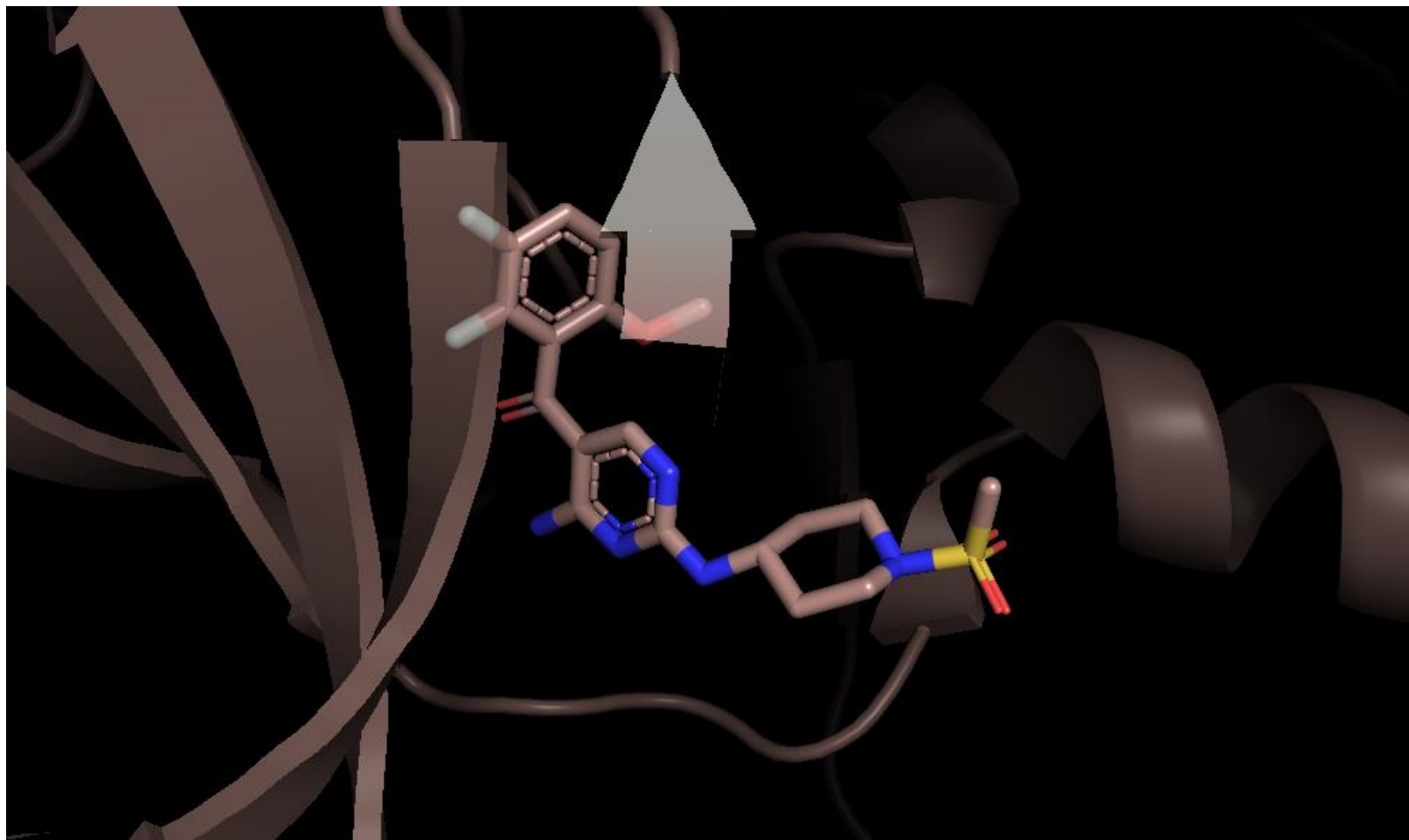
Pharmacophore model



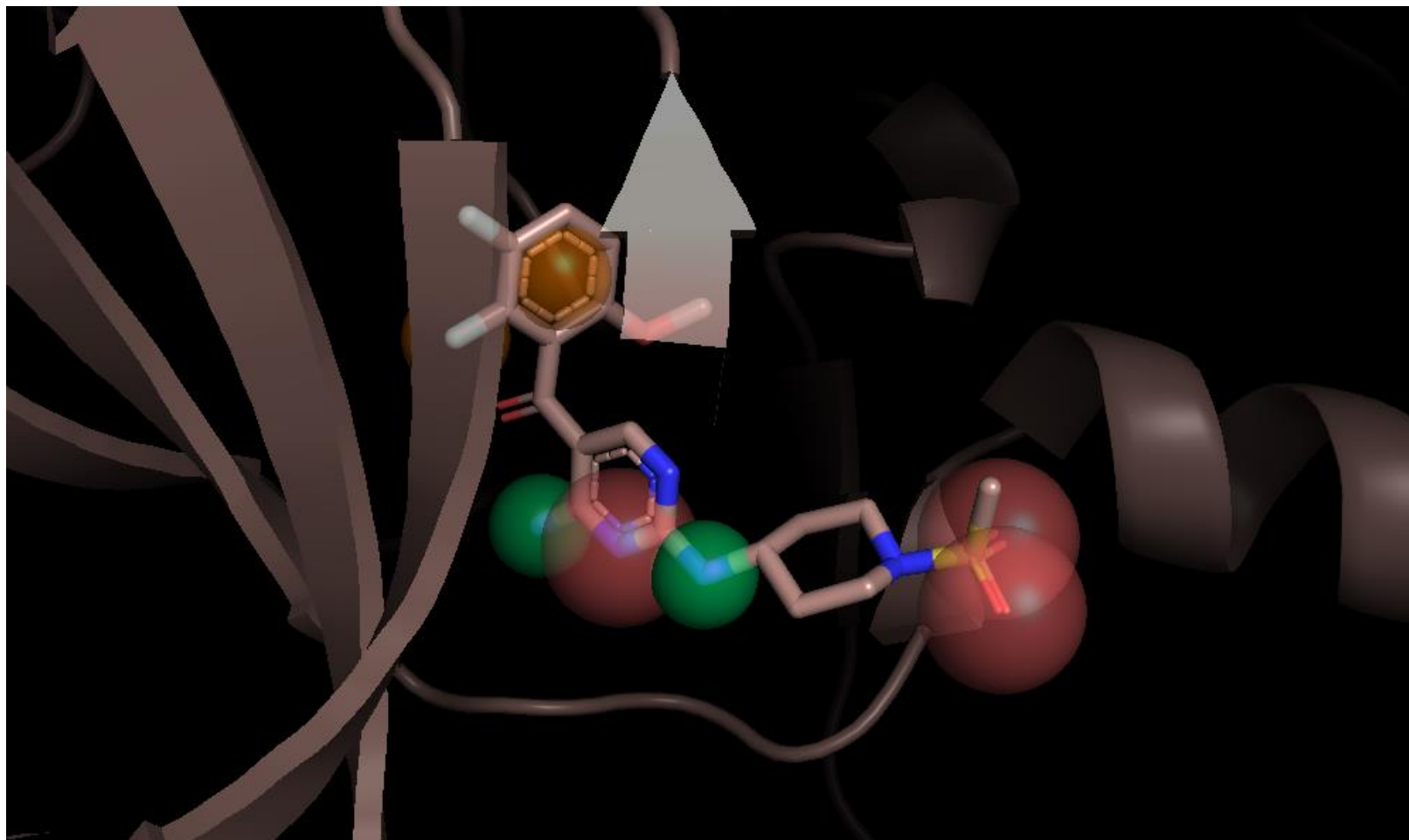
Iterative grow



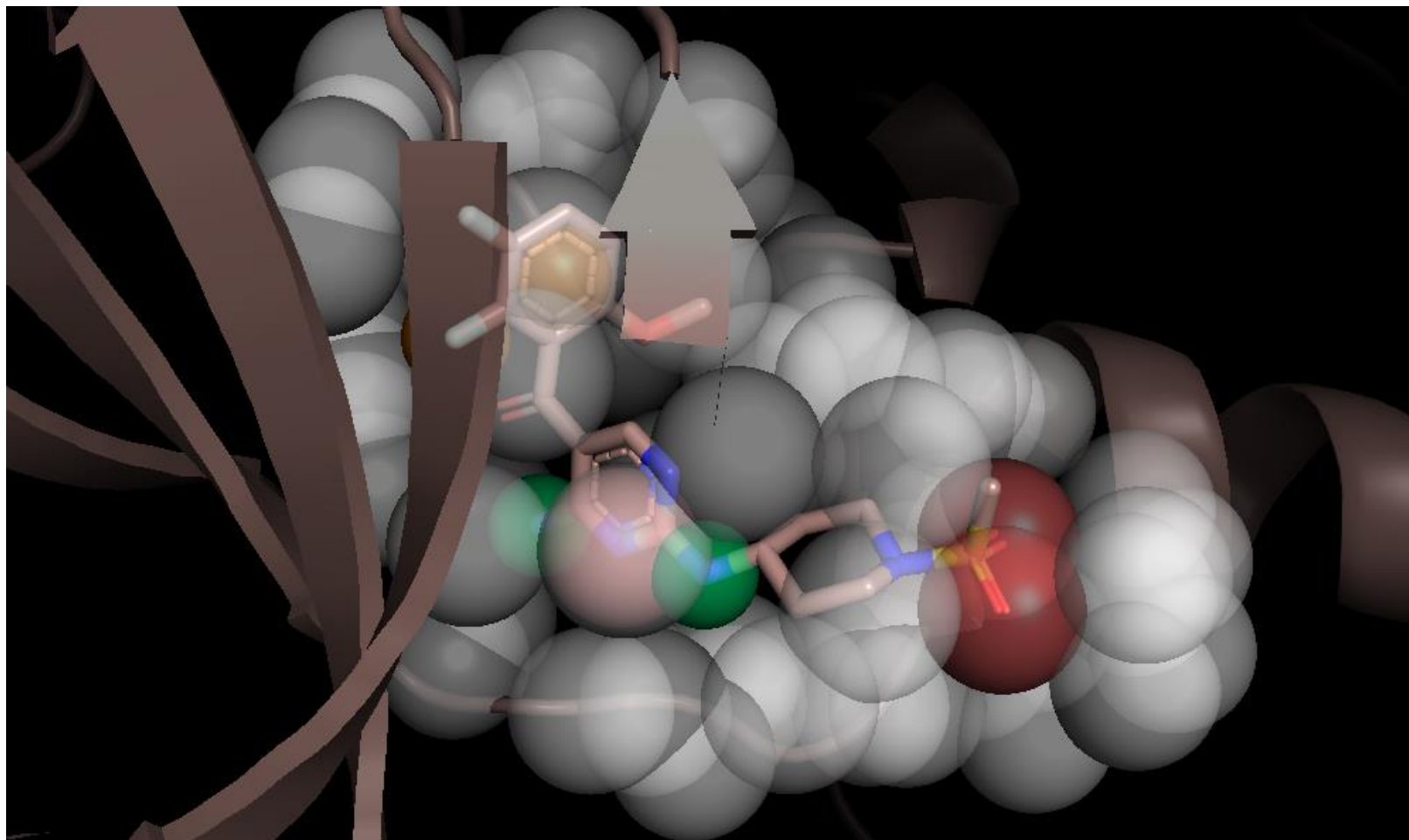
De novo design using pharmacophores (CDK2 example)



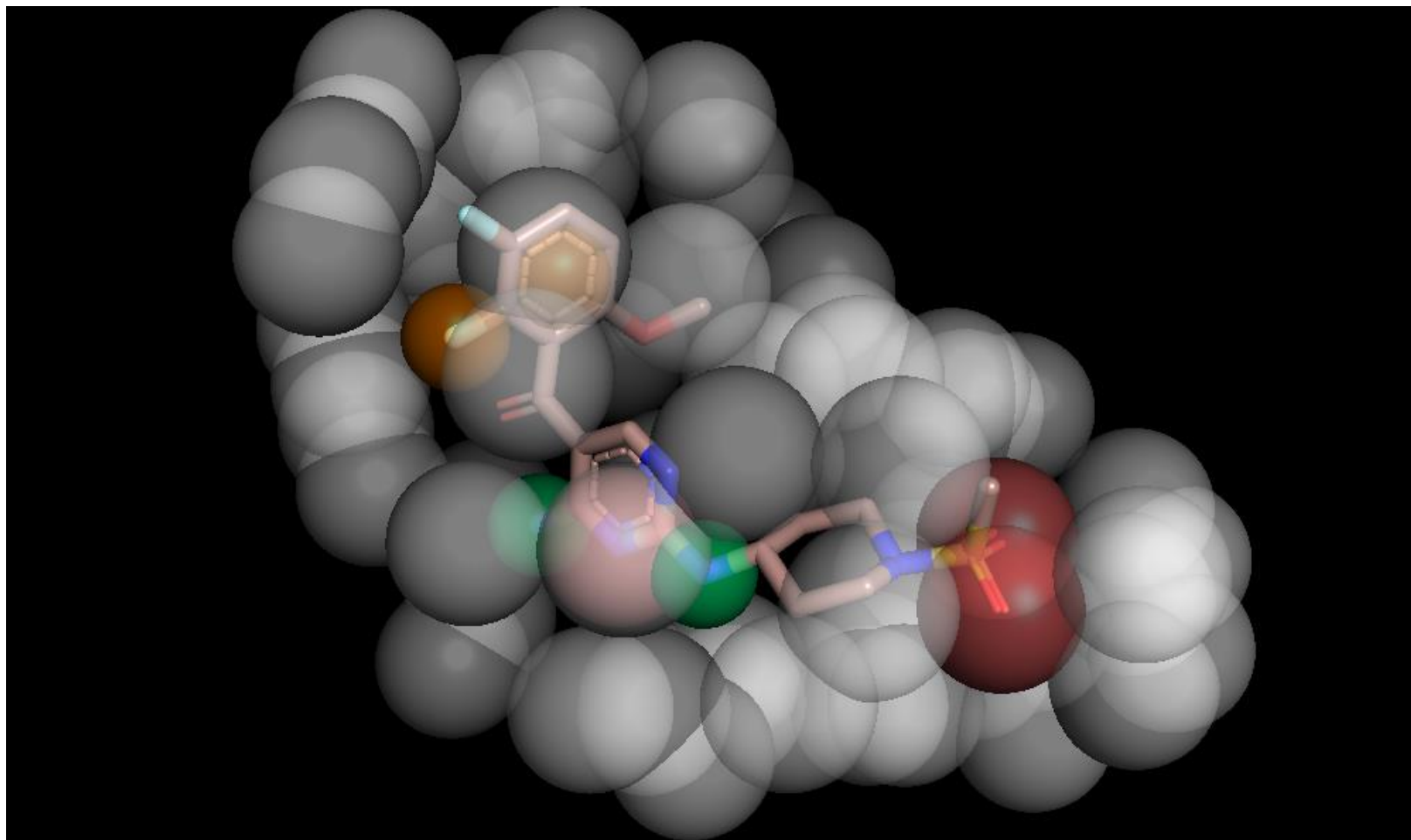
De novo design using pharmacophores (CDK2 example)



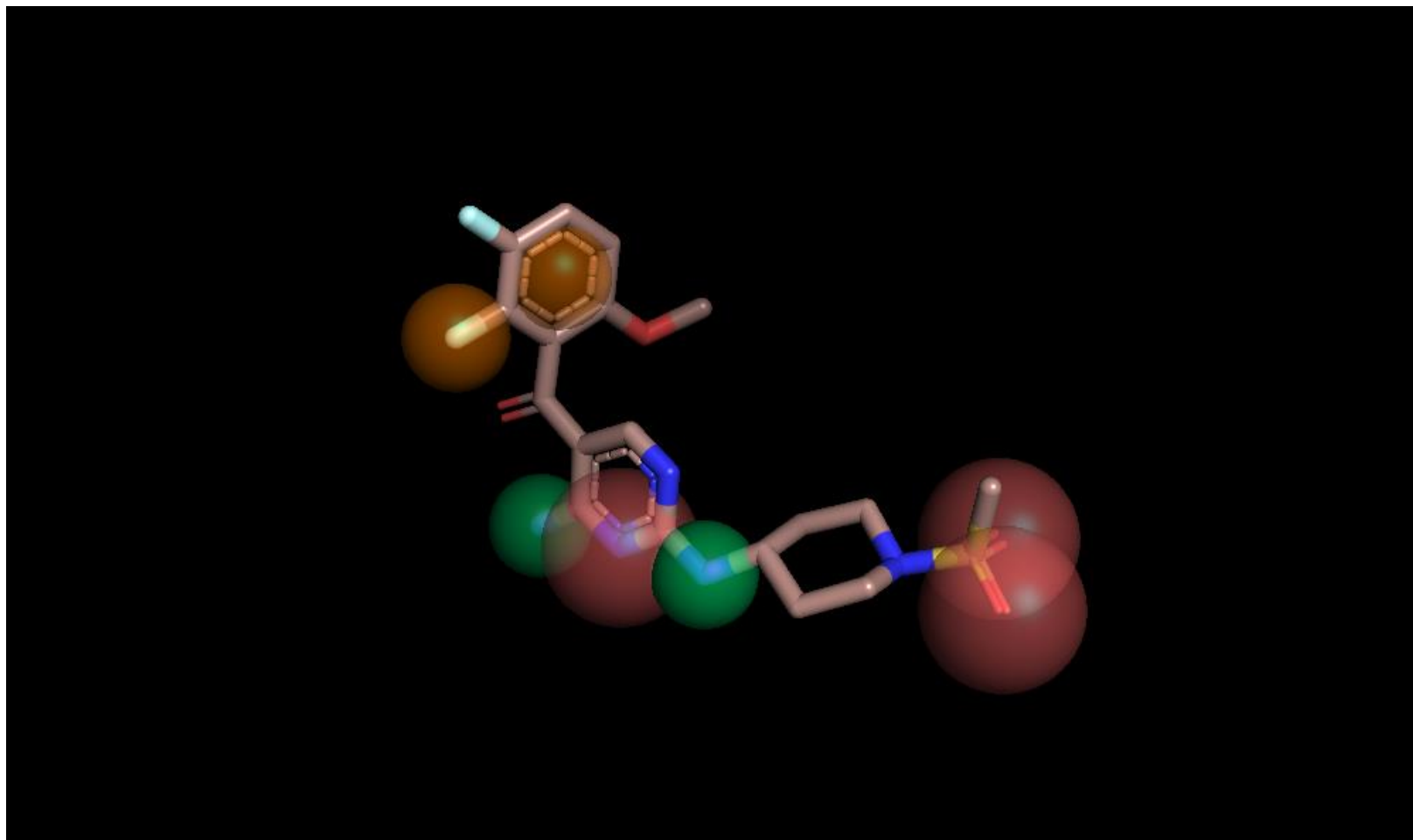
De novo design using pharmacophores (CDK2 example)



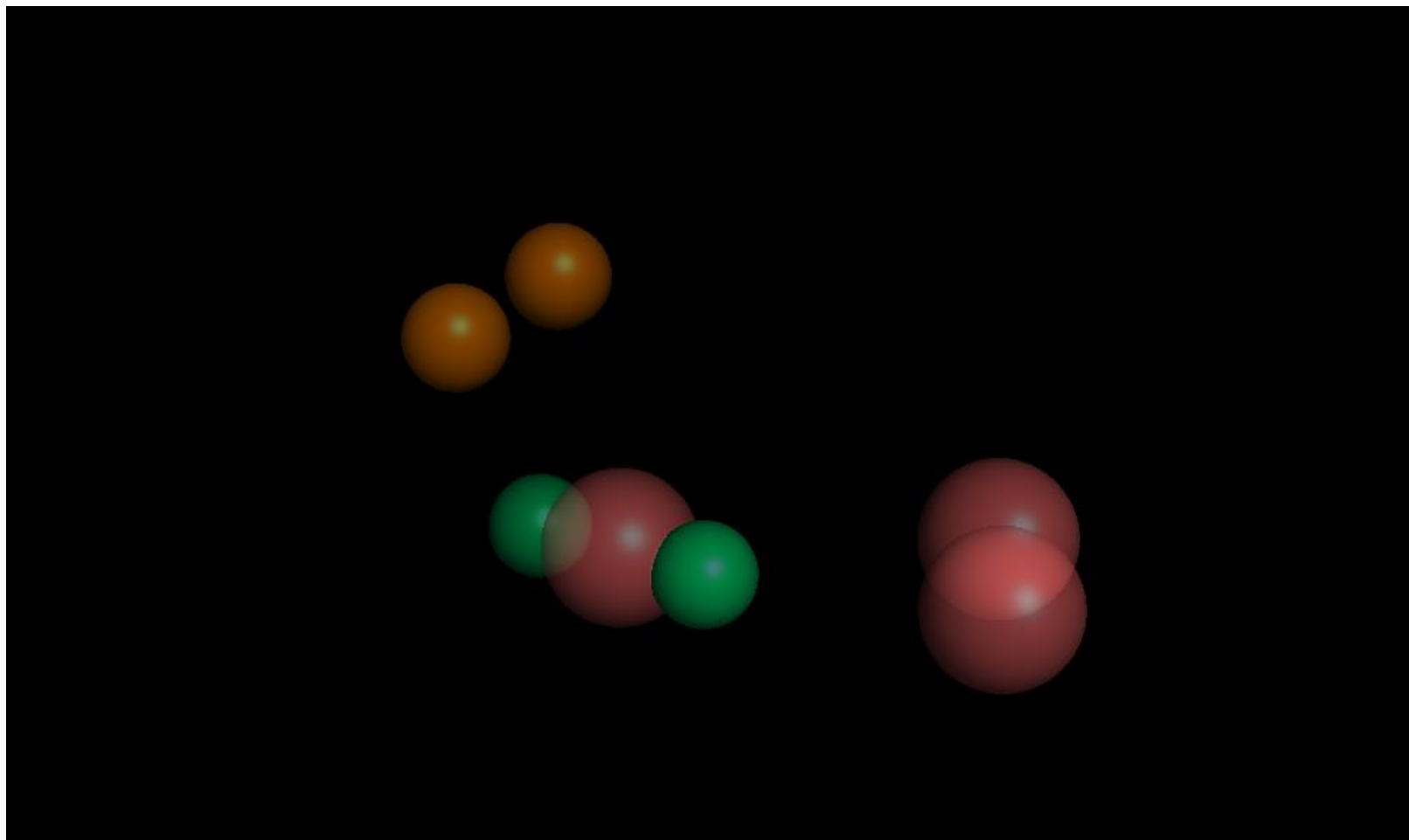
De novo design using pharmacophores (CDK2 example)



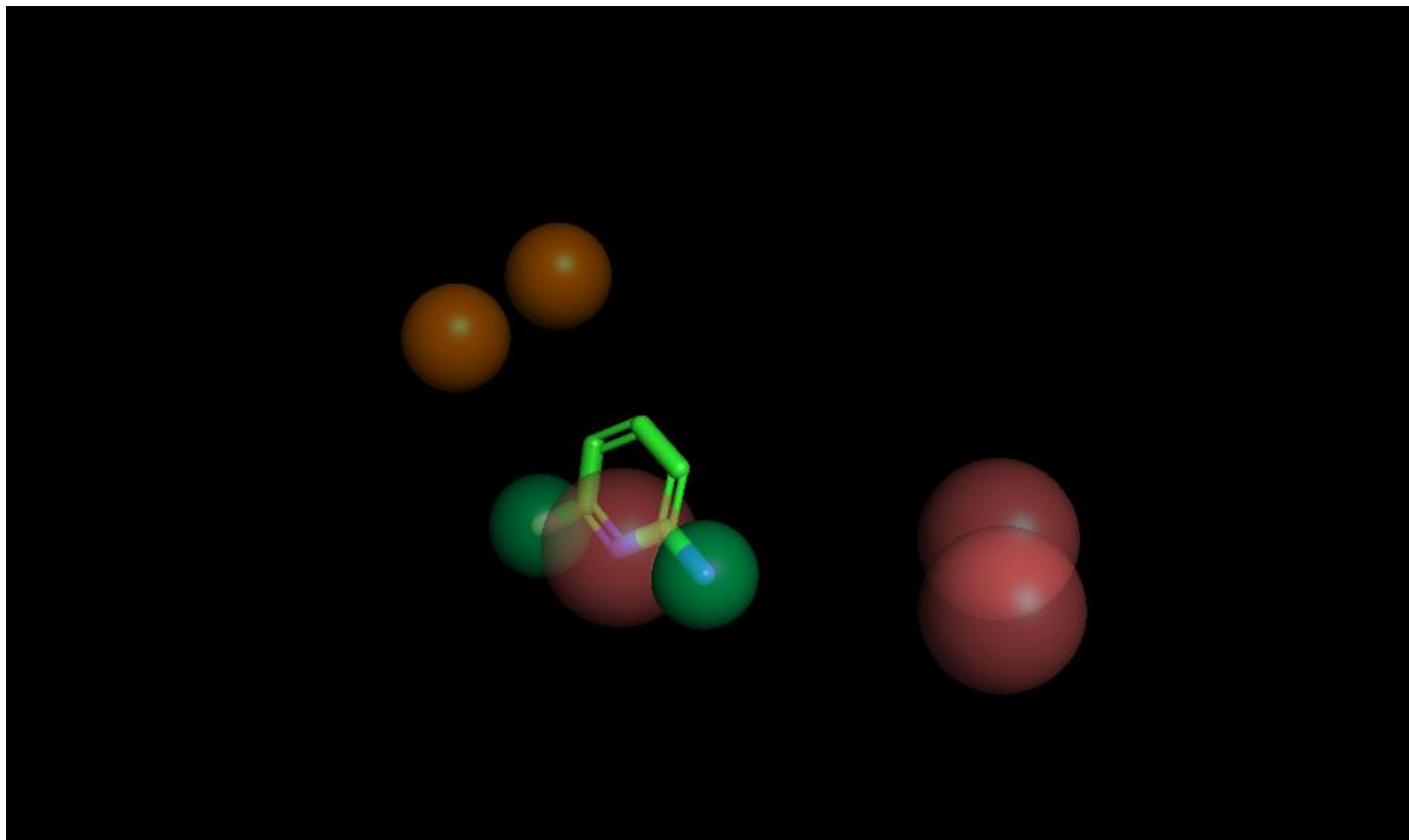
De novo design using pharmacophores (CDK2 example)



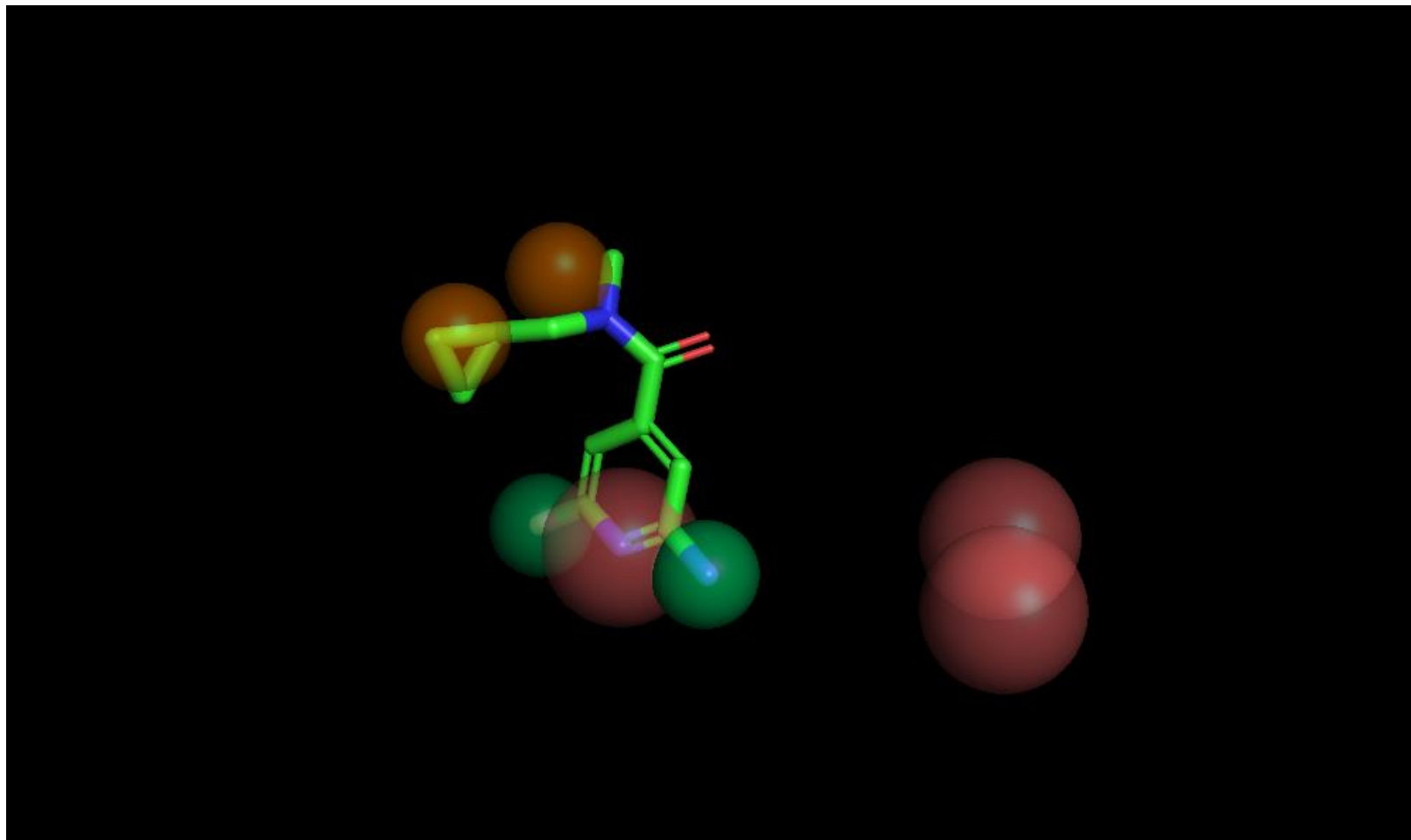
De novo design using pharmacophores (CDK2 example)



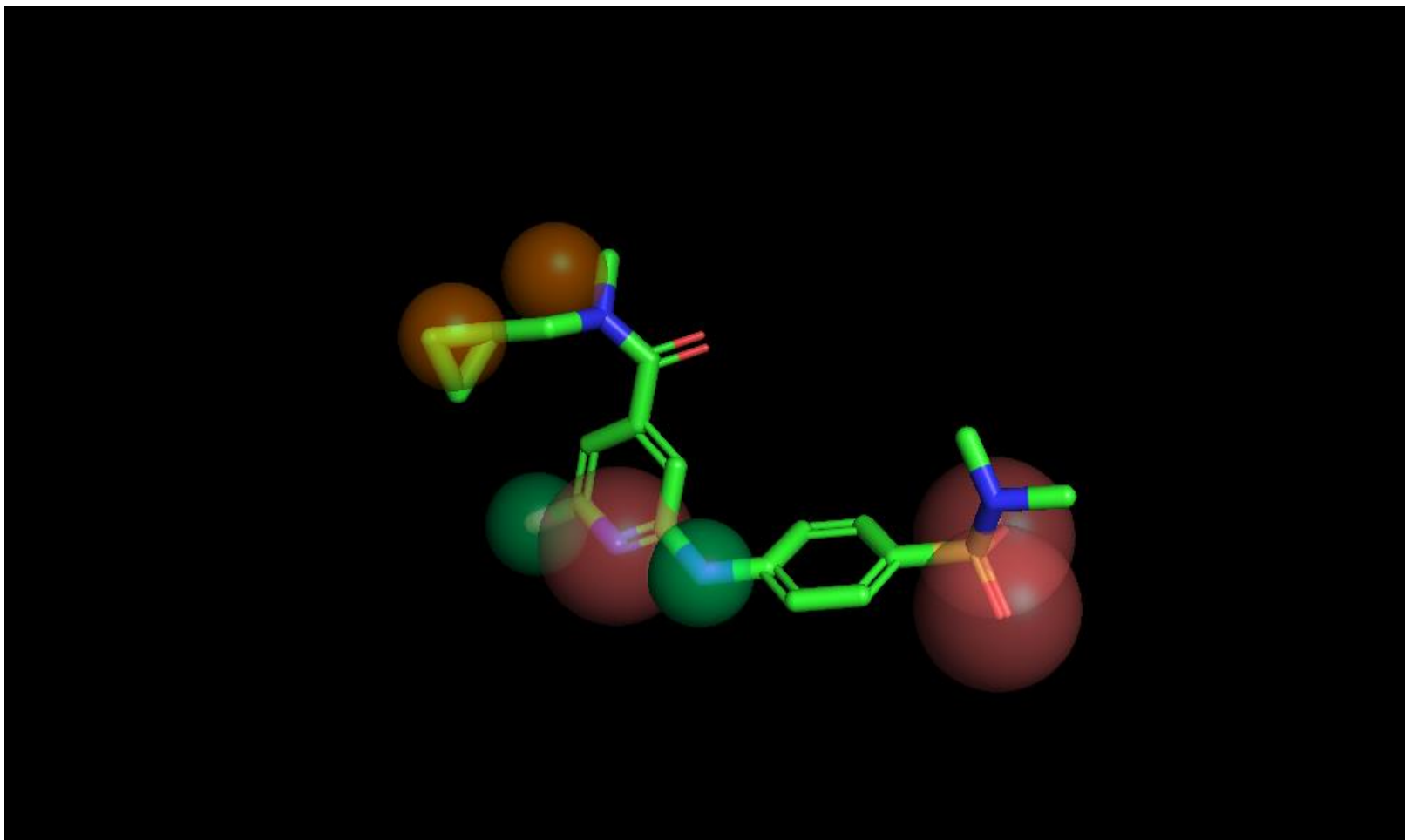
De novo design using pharmacophores (CDK2 example)



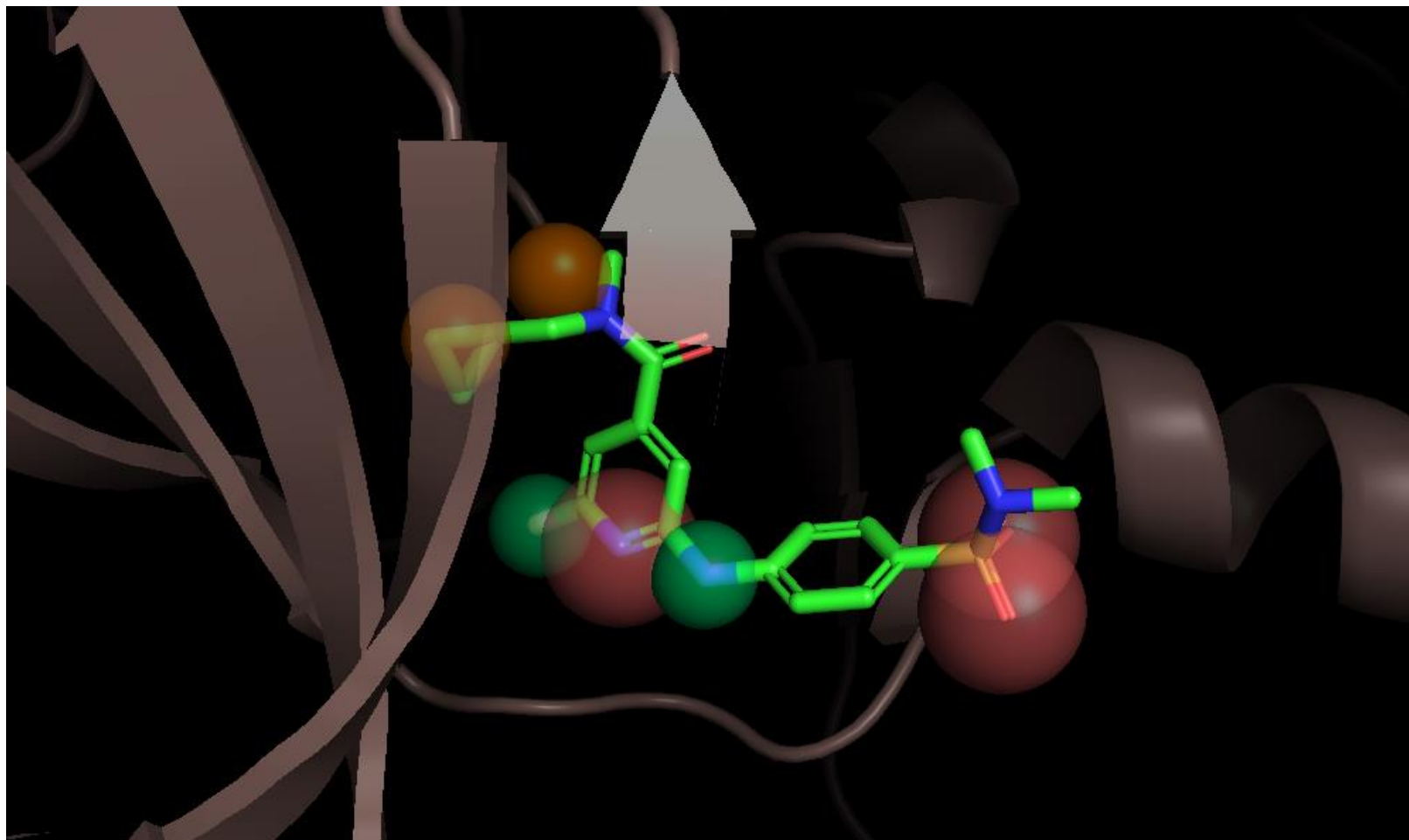
De novo design using pharmacophores (CDK2 example)



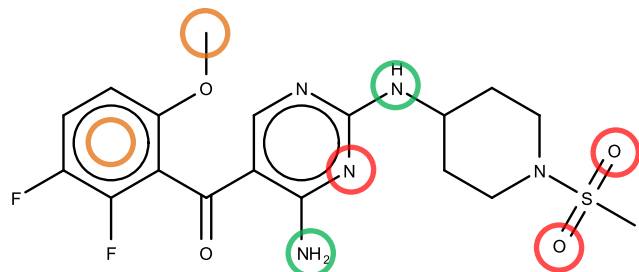
De novo design using pharmacophores (CDK2 example)



De novo design using pharmacophores (CDK2 example)

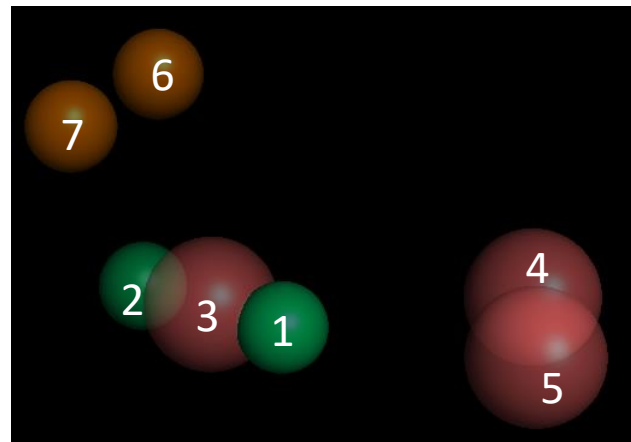


De novo design using pharmacophores (CDK2 example)

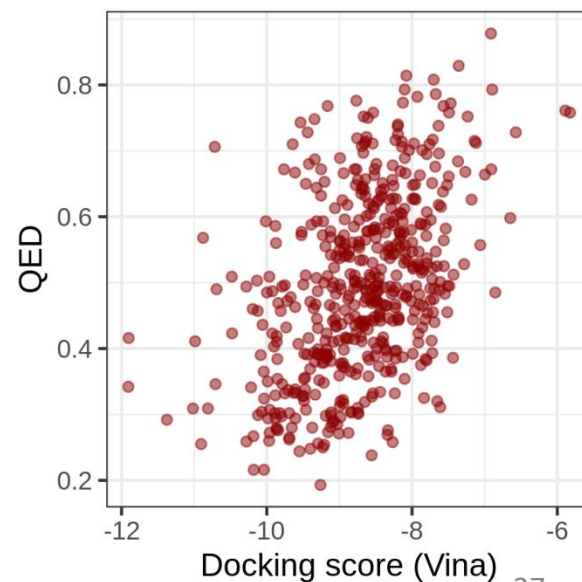
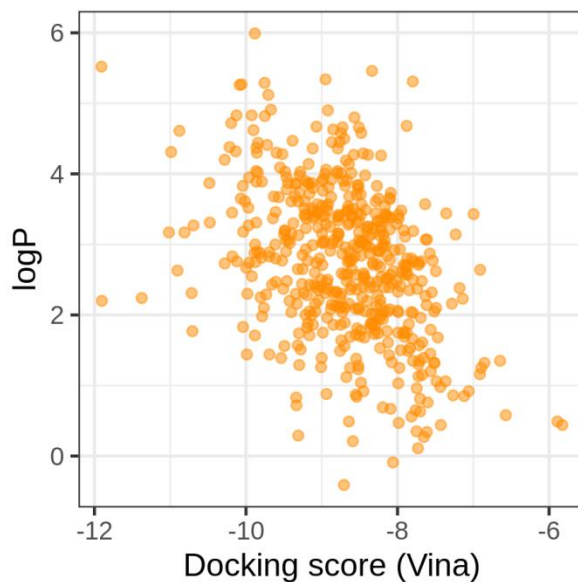
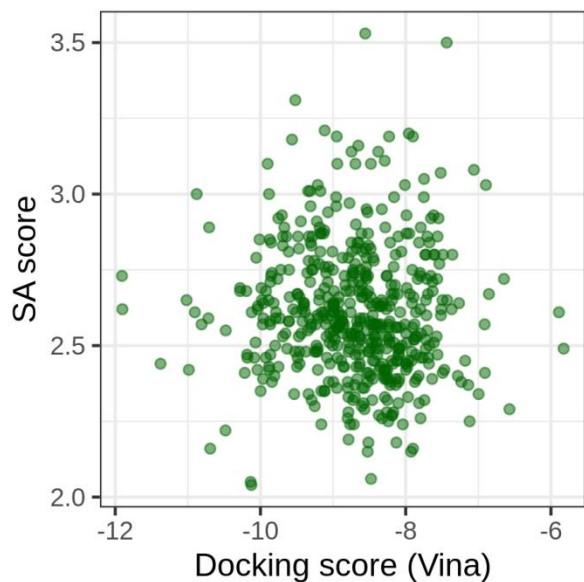


2FVD
 $K_i = 3 \text{ nM}$
docking score = -8.7

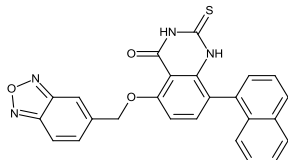
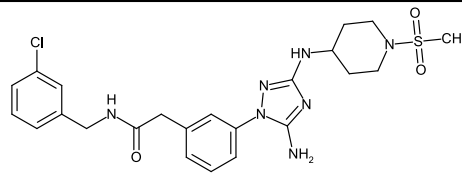
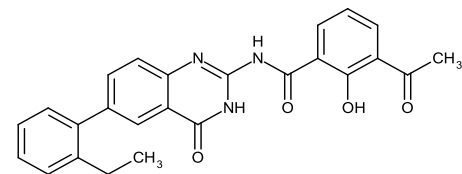
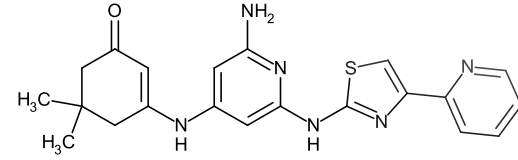
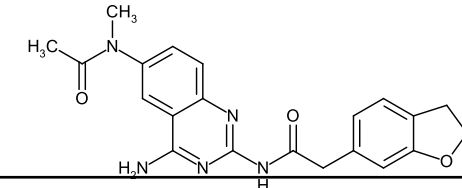
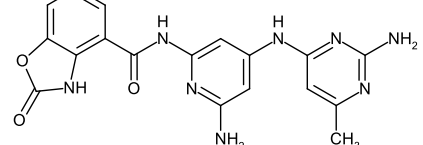
structure-based



233 635 structures were generated
521 matched at least 5 features

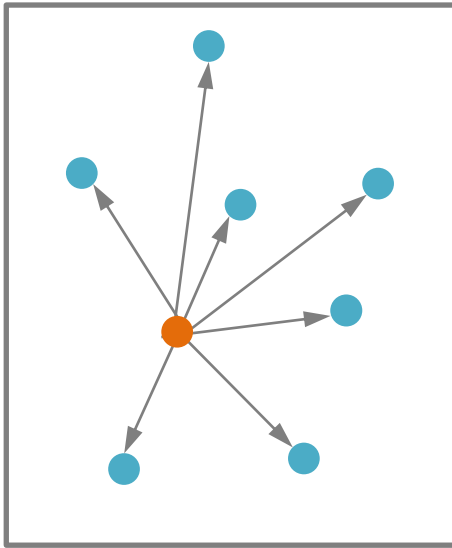


Results from top 10

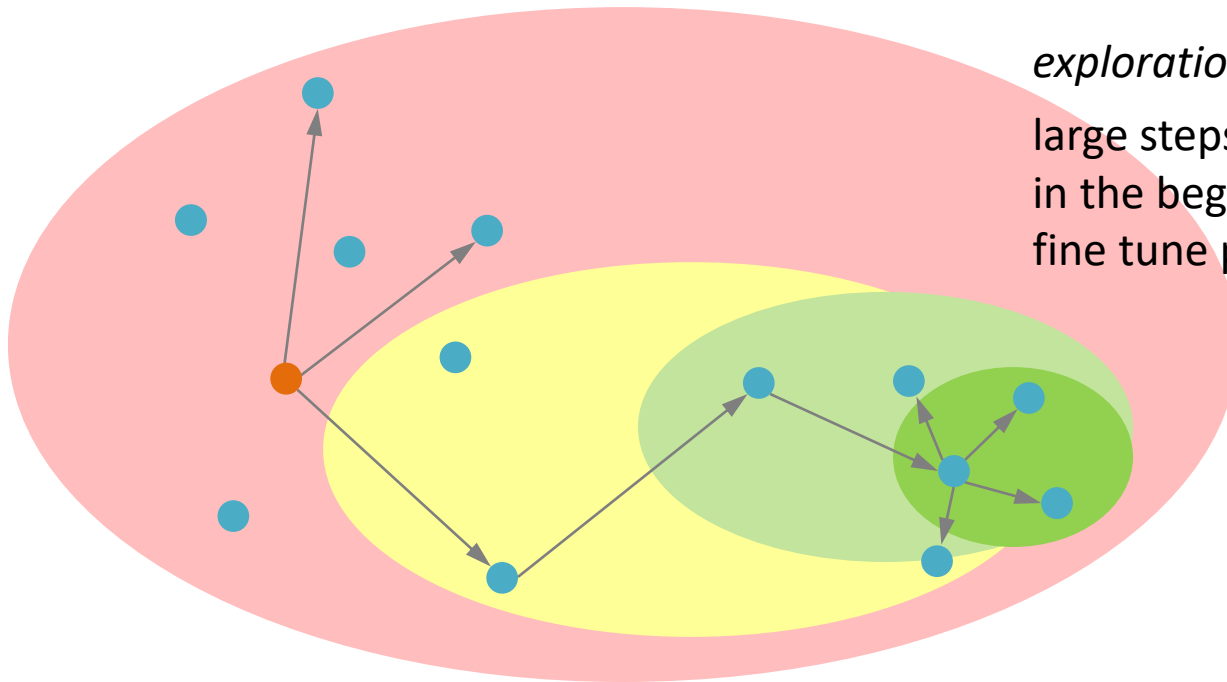
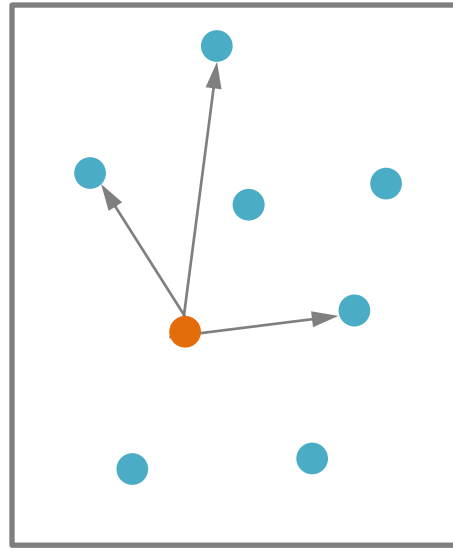
smi	id	Docking score	logP	MW	QED	SA
	225	-11.9	5.52	452	0.342	2.73
	732	-11.9	2.2	518	0.416	2.62
	397	-10.9	4.31	427	0.411	2.42
	487	-10.8	4.61	406	0.568	3
	377	-10.7	2.31	391	0.706	2.59
	517	-10.7	1.77	392	0.346	2.89

Tips for own workflow implementation

radius = 1



radius = 3



exploration vs. exploitation

large steps (large structural changes)
in the beginning and small steps to
fine tune promising candidates