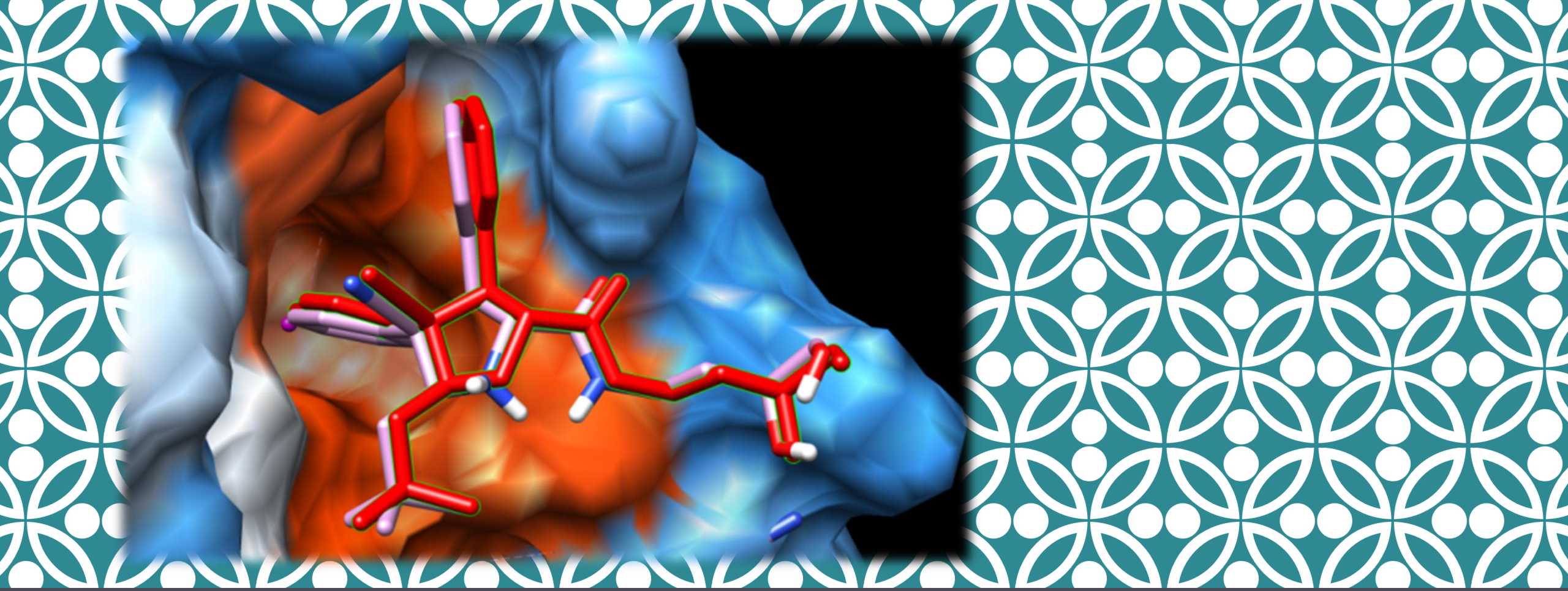
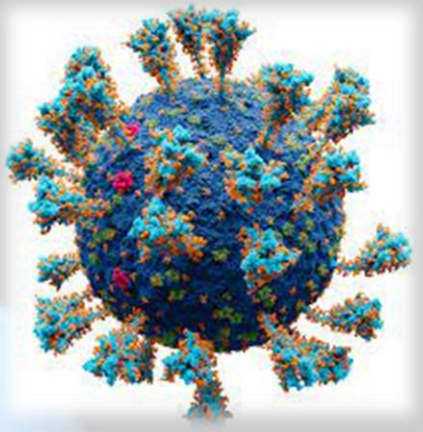


AN *IN SILICO* APPROACH TO FIND POTENTIAL SARS-COV-2 NSP13 INHIBITORS

Group Milano2:
Gaia Gastoldi
Ali Nassiri
Negar Esfahanizadeh
Mahsa Babaie (Presenter)
Prague, Feb. 2024

INTRODUCTION:

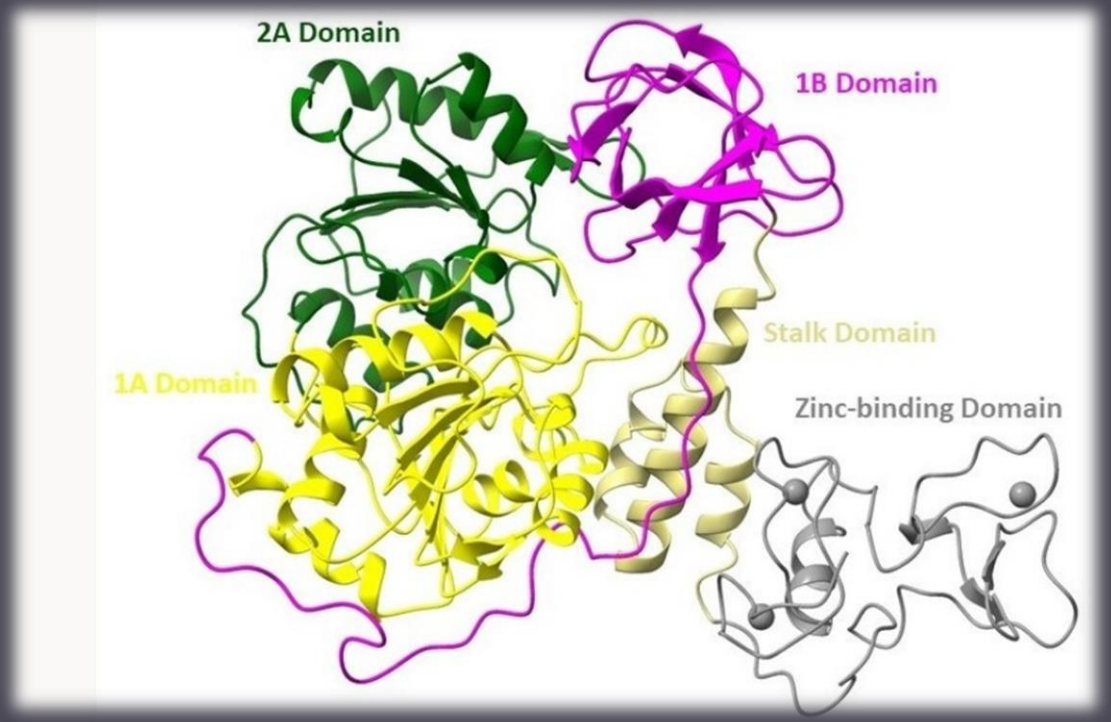


SARS-CoV-2 is the etiological agent of a respiratory disease whose first outbreak was reported in December 2019 in Wuhan, China and then it spread all over the world as a pandemic.

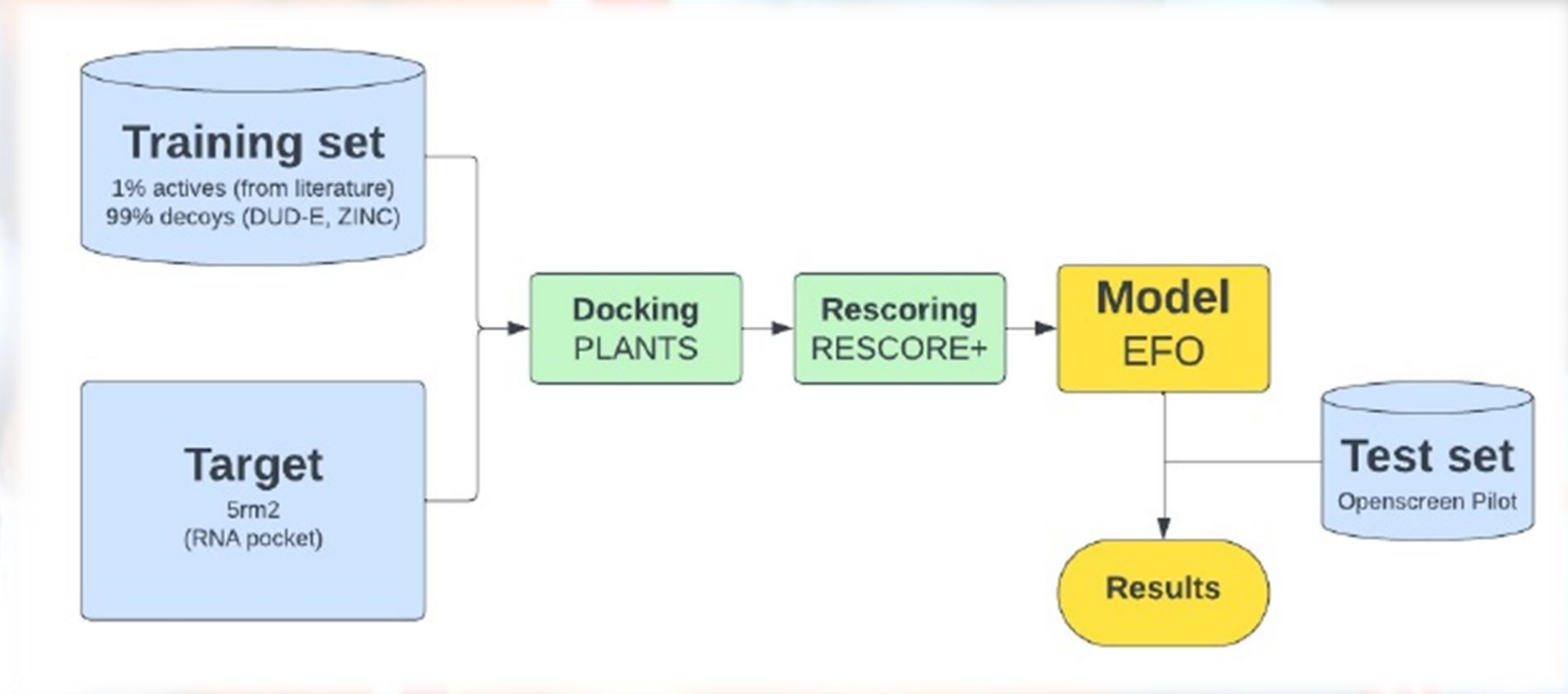
Among SARS-CoV-2 proteins, the structural proteins are obvious target to prevent viral entry into the host cells and viral assembly. However, also the NSPs (non-structural proteins) implied in viral replication are very interesting target to act against SARS-CoV-2 infection. NSP13 is the helicase of SARS-CoV-2 and it works in coordination with other NSPs.

NSP13 Structure:

NSP13 belongs to the helicase superfamily 1B and it is highly conserved among coronaviruses. It is composed by five domains: two ReacA-like domains (1A and 2A domains), the zinc-binding domain, the stalk domain and the 1B domain.



What did we do?!



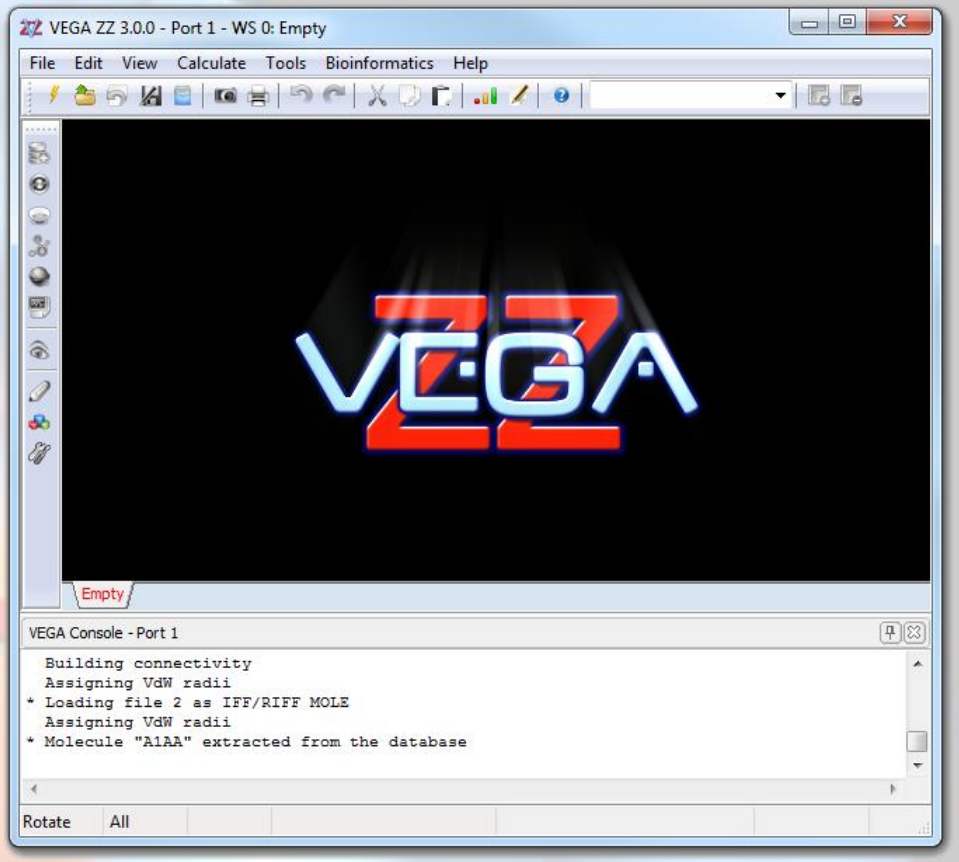
Virtual screening workflow followed in this project.

Training Set Preparation:

Training set preparation: 9 inhibitors were selected from literature:

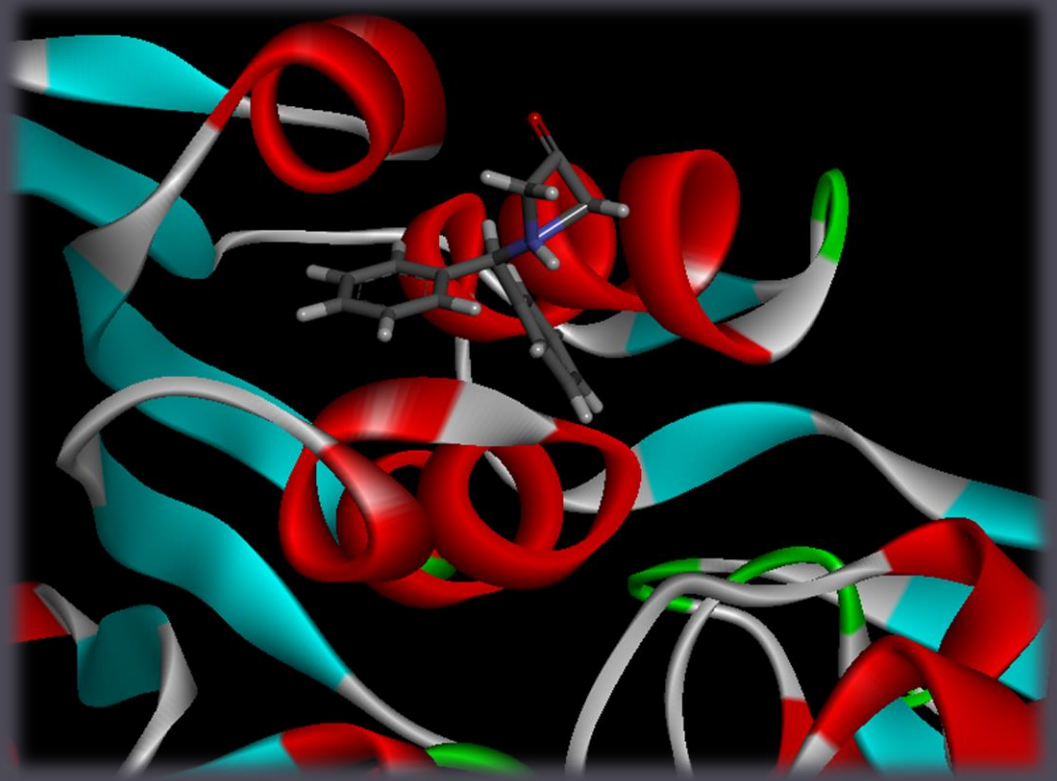
Baicalein
Flavanone
Flavanone-7-O-glucoside
Kaempferol
Licoflavone
Myricetin
Quercetin
SSYA10-001
Zafirlukast.

Their SMILES were retrieved from PubChem and used to build the 2D molecules in VEGA ZZ.



Pocket definition

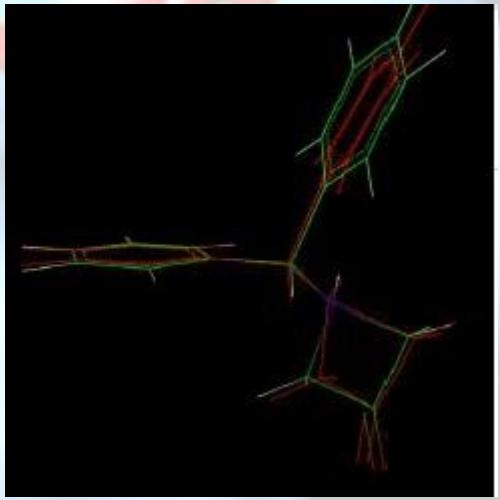
Once the training set was completed, we considered the protein structure. Starting from literature, we chose to work on the 5'-RNA pocket. Among the structures available on MEET-EU 2023 website, we chose **5RM2** because it was the **only structure with a small molecule docked at the RNA-binding** site. Using the co-crystallized small molecule as the center of the binding pocket in VEGA ZZ, we defined the coordinates of the binding pocket (-24.6800; 17.7100; -10.6000) using an 8 Å radius around the ligand.



REDOCKING:

- We used **PLANTS** to perform the docking as it is already implemented in VEGA ZZ and the **scoring function** we implemented was **ChemPLP**.
- We decided to use the original pose of the ligand as a shape constraint in the docking procedure: for an optimal atom-overlap of a ligand and a shape atom –3.0 value is added to the score. In this way we obtained an **RMSD of 0.3783 Å**

Overlapping structures of the original co-crystallised ligand and the minimized-and-docked ligand after the applying of the shape constraints



DOCKING AND RESCORING:

* The docking of the **training set** was performed with PLANTS in the VEGA ZZ environment. 10 poses were generated for each molecule (9000 poses were obtained in total).

* The rescoring of the generated poses was performed with **RESCORE+** as implemented in the VEGA ZZ suite. Rescoring is a useful tool to calculate new scores after molecular docking, to refine the already-obtained scores and generate new others to have more features to be used in the model generation. The selected scoring functions were: CHARMM, Elect, ElectDD, MLPInS, MLPIns_2, MLPInS_3, MLPInS_F, Contacts, ChemPlp, Plp, Plp95, X-Score.

OUR MODEL DESIGN:

The model was developed using the **Enrichment Factor Optimization algorithm** as implemented in VEGA ZZ. EFO is a classification algorithm that allows to calculate a new scoring function by linearly combining different descriptors (in this case, the docking scores). The result of the calculation is a first-degree equation whose coefficients are calculated by maximizing the enrichment factor (EF), calculated as below. The EF quantifies the capacity of the model to rank the correct compounds in the top-scoring molecules and the bigger the EF is, the more robust the model is.

$$EF_{\%} = \frac{\frac{Actives_{\%}}{N_{\%}}}{\frac{Actives_{tot}}{N_{tot}}}$$

We have used two approaches:

1- Considering the **best pose score** for each molecule

2-Considering the **mean score of 10 poses** for each molecule.

The model with the higher EF and without redundancy in the equation was chosen for each pathway.

- **For the best poses:**

PLANTS_CHEMPLP_NORM_HEVATMS + 0,09523810 MLPINS_2 + 0,09523810 ELECTDD

Enrichment factor (at 1%) = 44.44

- **For the mean of 10 poses:**

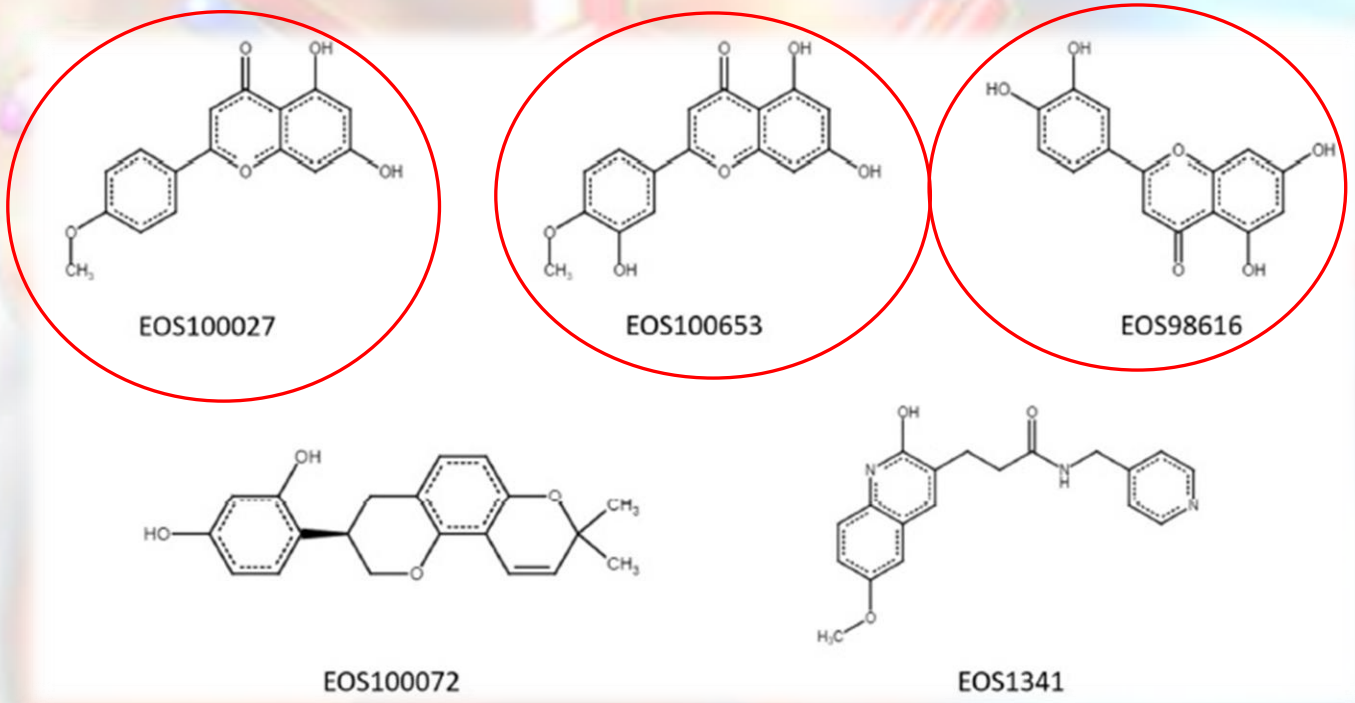
PLANTS_CHEMPLP_NORM_WEIGHT + 0,00948905 ELECTDD

Enrichment factor (at 1%) = 55.55

RESULTS:

BEST	MEAN
1,00000000 PLANTS_CHEMPLP_NORM_HEVATMS + 0,09523810 MLPINS_2 + 0,09523810 ELECTDD	1,00000000 PLANTS_CHEMPLP_NORM_WEIGHT + 0,00948905 ELECTDD
ModID_1	ModID_2
11 molecules predicted as active	15 molecules predicted as active
cut-off: -3,7314	cut-off: -0,22935
EOS100027	EOS98616
EOS100104	EOS100653
EOS100653	EOS100677
EOS100947	EOS593
EOS98616	EOS100072
EOS100072	EOS102169
EOS1251	EOS1341
EOS102267	EOS100027
EOS1601	EOS1763
EOS2290	EOS100804
EOS1341	EOS102409
	EOS101604
	EOS1240
	EOS100470
	EOS102410

As can be seen, 5 molecules are recognized as active by both models: EOS100027, EOS100653, EOS98616, EOS100072, EOS1341. 3 of the top-scoring hits are flavonoids. Among flavonoids, the *in vitro* activity as NS13 inhibitors has been proven for several molecules and in fact, among the inhibitors selected for the training set, seven were flavonoids. In conclusion, our model allowed us to search for inhibitors of NSP13, using docking, rescoring and the EFO algorithm. For the construction of the training set, we relied on the available experimental data about inhibitors whose activity has been confirmed in vitro.



COMPARISON WITH THE PAIRED GROUPS

Warsaw 3 team → peptides

Sorbonne 4 team → ZINC20 and AutoDock Vina

CONCLUSION

During our project we learned how to perform a virtual screening starting from an available crystal structure of the target protein and some experimentally-active molecules. In the Openscreen ECBL Pilot Library the molecules identified as active by both of our models were mainly flavonoids and this is probably because the inhibitors that we selected from literature were mainly flavonoids as well. To improve the predictive power of our models it would be useful to expand the training set with more diverse compounds.

In conclusion, our approach allowed us to exploit the potential of different tools for drug-design but always relying on the available experimental data about NSP13 to find new potential inhibitors.

The end…