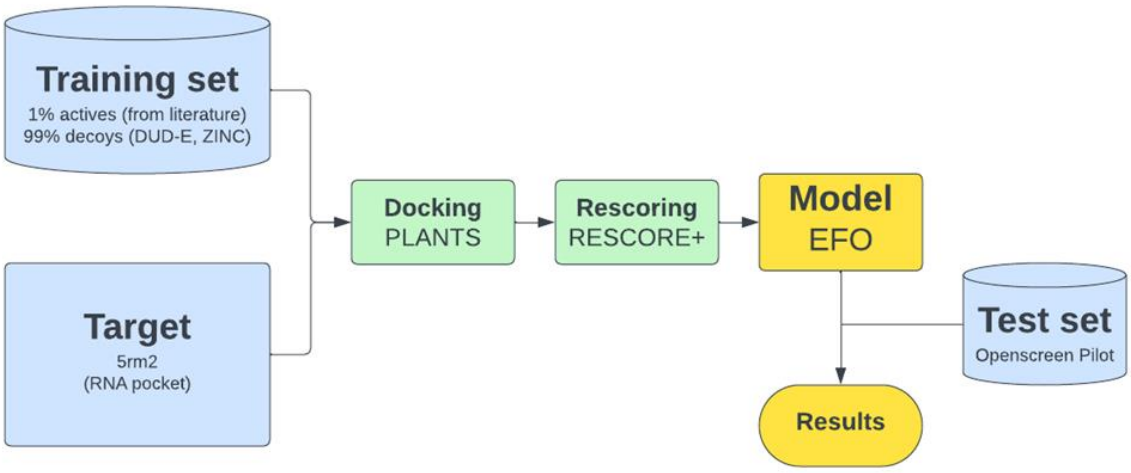
An *in silico* approach to find potential SARS-CoV-2 NSP13 inhibitors

Team MILANO 2: Ali Nassir, Gaia Gastoldi, Mahsa Babaie, Negar Esfahanizadeh Hosseinpour



Training set preparation: we decided to start inhibitors whose activity had been tested *in vitro* and we selected 9 inhibitors. Decoys were used as inactive molecules and they were retrieved from DUD-E and ZINC. All the molecules were converted to 3D, ionized and minimized using VEGA ZZ and then filtered according to physical-chemical properties of the 9 active molecules (applicability domain). We obtained an unbalanced dataset of 900 molecules (9 active and 891 inactives).

Pocket selection: we focused on the 5'-RNA-binding pocket of NSP13 because it is one of the inhibitor-binding site proposed in literature for its importance in the helicase functioning and according to previous studies. To perform molecular docking, we used the crystal structure with PDB code 5RM2 as already prepared in the Meet-EU 2023 website. The pocket's coordinates were defined using the co-crystallised ligand as the centre with a radius of 8 Å.

Docking: the redocking of the minimised co-crystallised ligand was performed to test the suitability of the decided algorithm (PLANTS as implemented in VEGA ZZ) for the docking. We used ChemPLP as scoring function and obtained an RMSD below 2 Å by using the shape constraints of the original co-crystallised ligand.

The docking of the training set was executed with PLANTS and 10 poses were generated for each molecule of the training set.

Rescore: the 9 000 poses were then rescored with RESCORE+ in VEGA ZZ to compute new scores using other scoring functions.

Model design: we tried to develop two different models, one working only with the best pose (according to ChemPLP scoring function) for each ligand (*BEST model*) and one working with the mean scores of the 10 poses for each ligand (*MEAN model*).

We used the Enrichment Factor Optimization algorithm (in VEGA ZZ) that combines one or more docking scores (as in our case) or descriptors to obtain a linear equation that maximises the Enrichment Factor (EF). The Enrichment Factor was defined as below.

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

$Actives_{\%}$= number of corrected identified active molecules in the 1% top compounds.
$N_{\%}$= total number of molecules in the 1% top compounds (9)
$Actives_{tot}$= number of active molecules in the whole training set (9)
N_{tot}= number of molecules in the whole training set (900)

Among the equations generated with EFO we discarded the ones that were redundant and then chose the ones with the higher EF. We obtained:

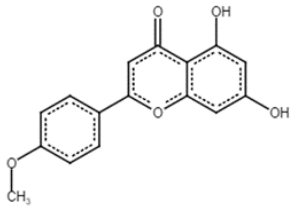
- PLANTS_CHEMPLP_NORM_HEVATMS + 0,09523810 MLPINS_2 + 0,09523810 ELECTDD (EF=44.44 for the *BEST model*)
- PLANTS_CHEMPLP_NORM_WEIGHT + 0,00948905 ELECTDD (EF=55.55 for the *MEAN model*)

The cut-off value to discriminate active and inactive molecules was the highest scores in the top 1% of molecules. The Enrichment Factor values were satisfying.

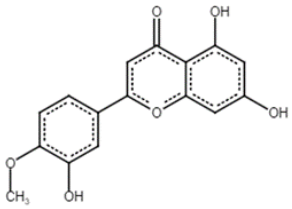
Virtual screening: we prepared Openscreen ECBL Pilot Library as we did for the training set and we filtered it to keep only the molecules in the applicability domain. We performed the docking and the rescore with the same parameters of the training set and we applied the two equations.

Results

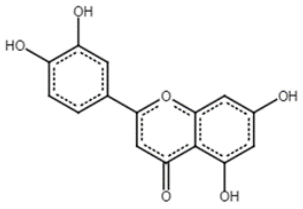
The BEST model and the MEAN model predicted respectively 11 and 15 molecules of the Openscreen database as active. Among them, 5 molecules were in common and their structure is shown in the figure below.



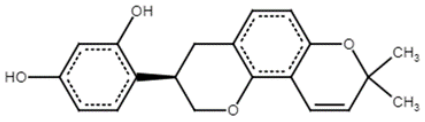
EOS100027



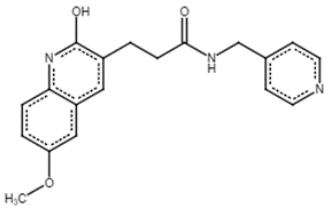
EOS100653



EOS98616



EOS100072



EOS1341

As it can be seen, 3 out of these 5 molecules are flavonoids. Flavonoids activity against SARS-CoV-2 have actually been proven also in vitro for different compounds. However, these results suggested that our method can be improved in future by adding more diverse compounds in the active set of molecules that we used to train the model.