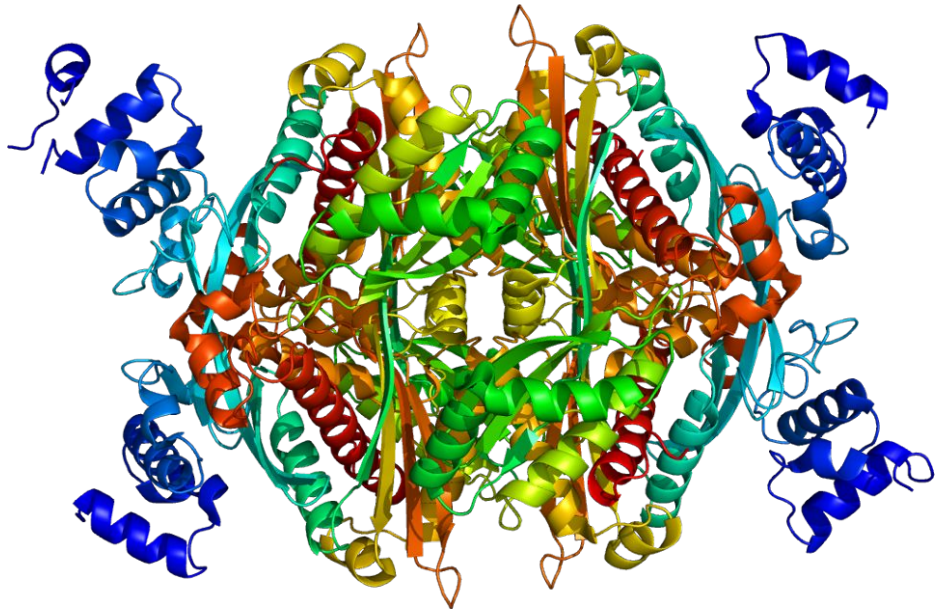
HMG-CoA reductase

Maarten Koffijberg, Youry van Uum | Leiden

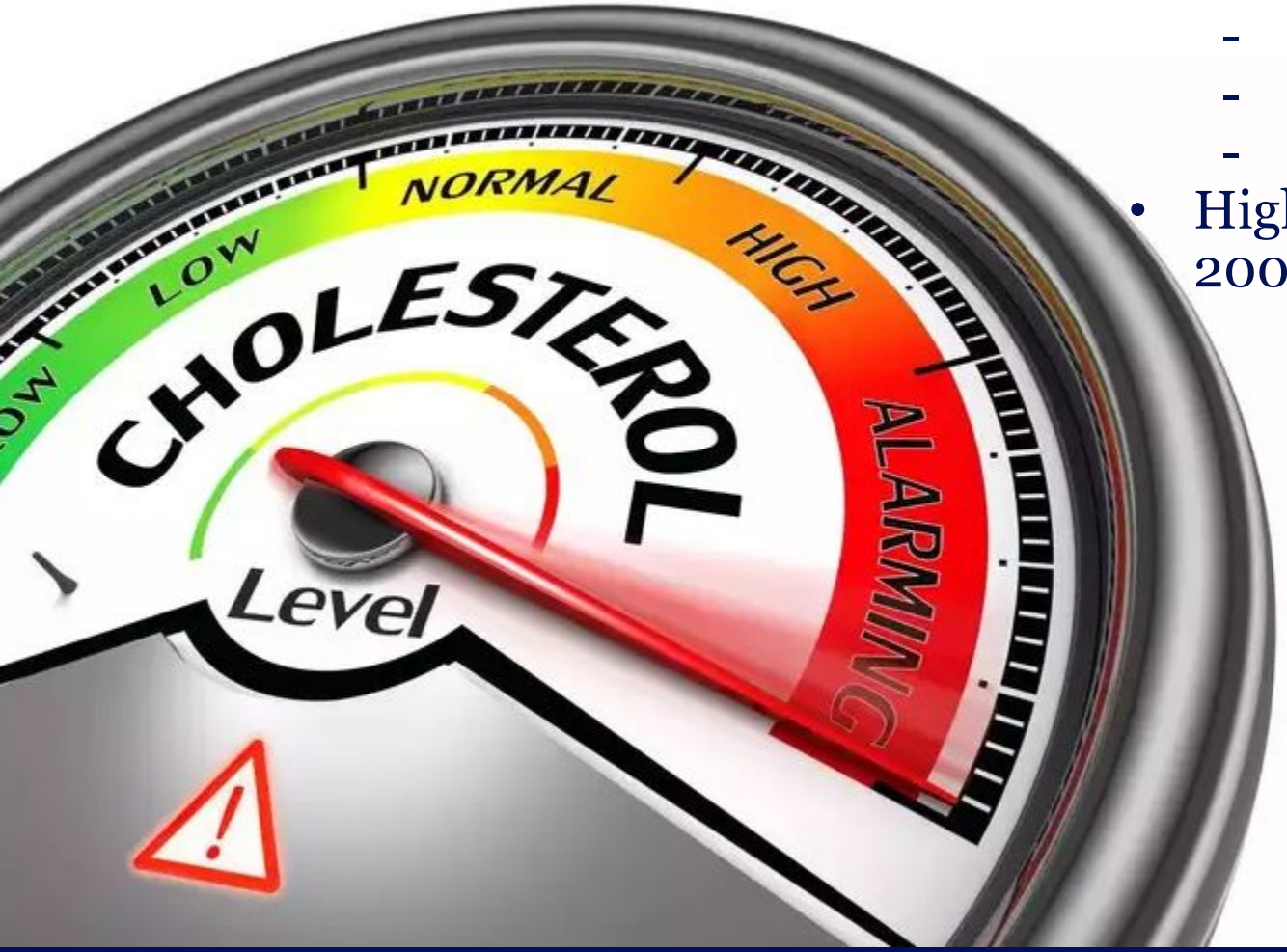
Universiteit
Leiden
The Netherlands

Index

1. Target
2. Data
3. Clustering
4. Machine learning models
   - Classifier
   - Regression
5. Docking

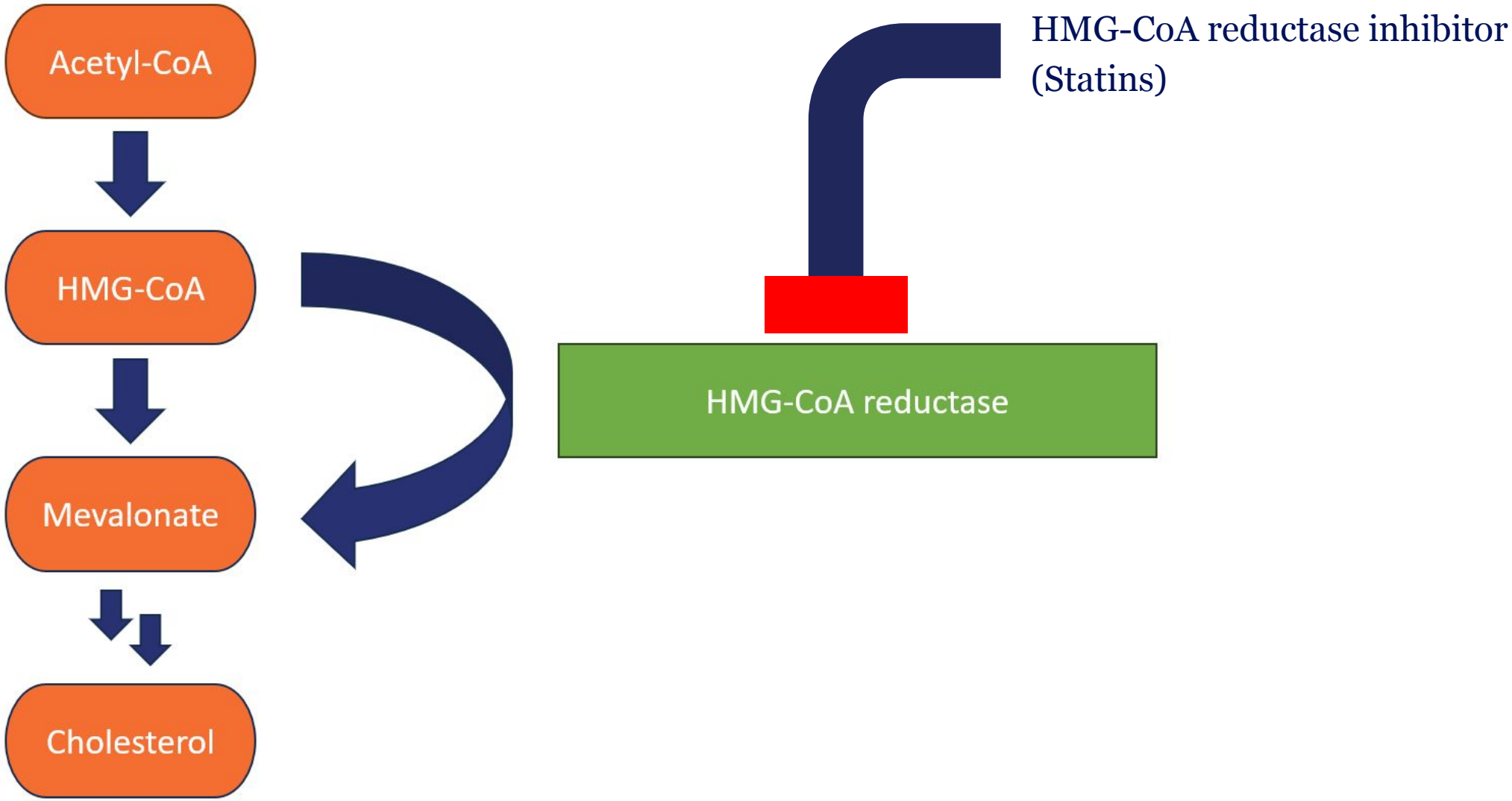


Target



- High cholesterol levels leads to atherosclerosis
- Increased risk of:
  - Heart attack
  - Stroke
  - Peripheral artery disease
- High incidence of cholesterol levels above 200mg/dL[2]

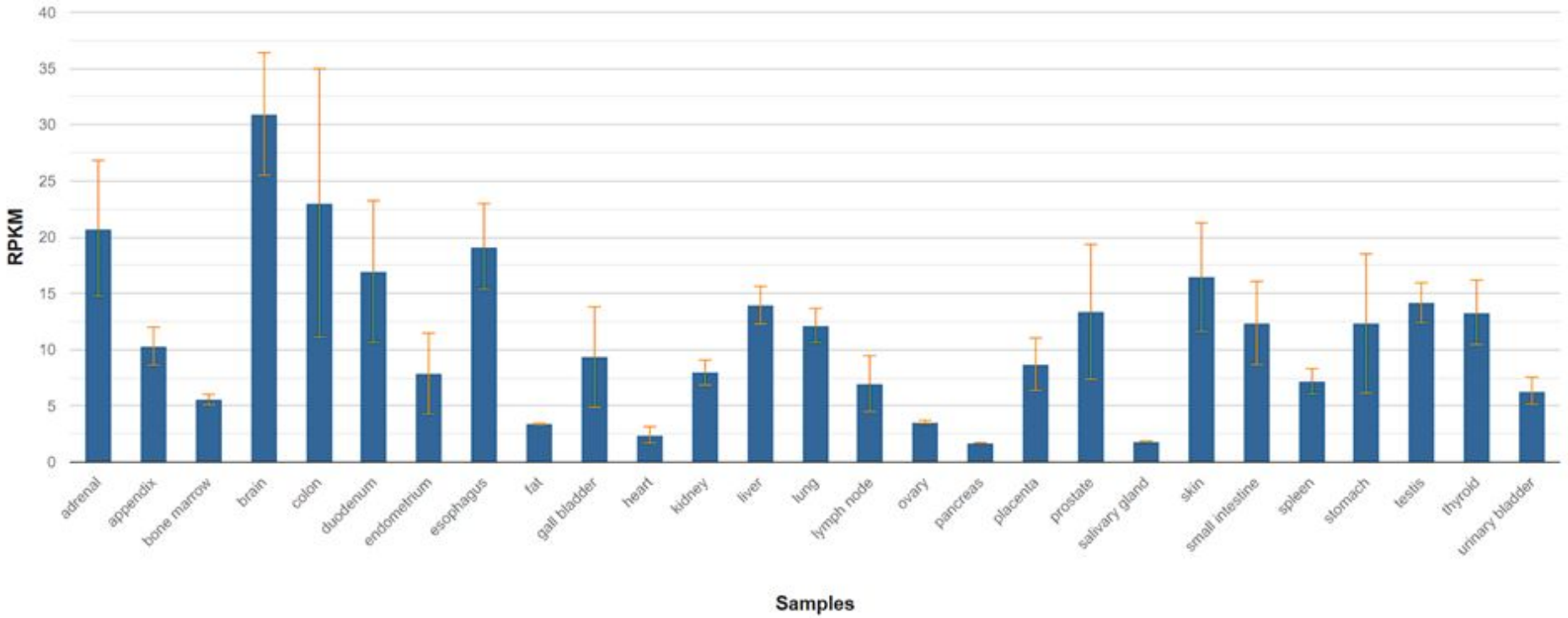
Target - HMG-CoA reductase



Target - HMG-CoA reductase

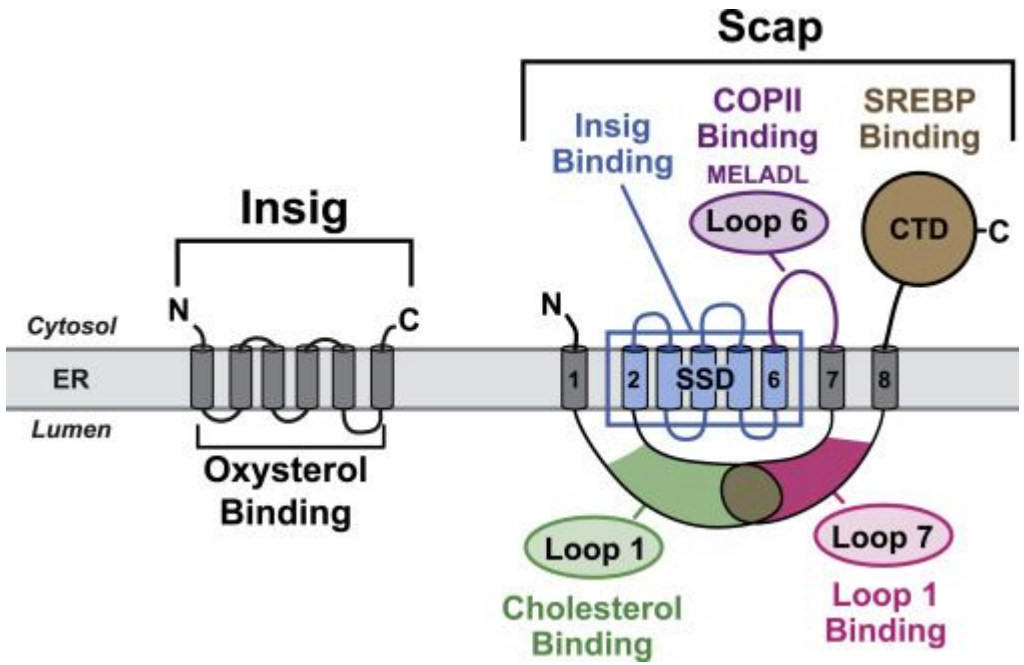
Our target (Uniprot code: P04035):

- Part of the HMG-CoA reductase family
- This family is found in humans, mice and rice.
- Can be found all over the body



Target - HMG-CoA reductase

- Most similar human protein (based on protein sequence): Scap (26%)
- Similarity in Sterol Sensing Domain (SSD)
- Both Scap and our target are cholesterol regulated



Target - HMG-CoA reductase

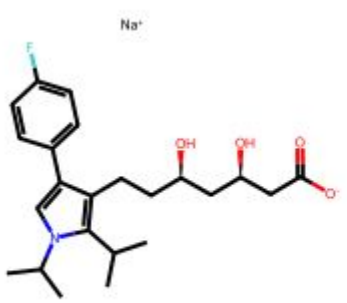
- Most similar human protein (based on protein sequence): Scap (26%)
- Similarity in Sterol Sensing Domain (SSD)
- Both Scap and our target are cholesterol regulated

- Most similar protein in other species:
- Pan Paniscus or Pygmy Chimpanzee (99%)
- High levels of conservation between species

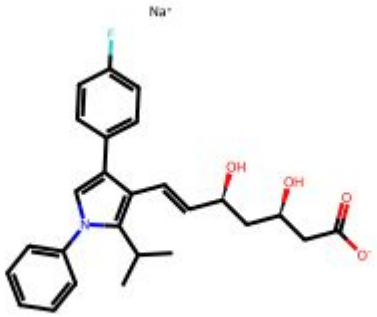
Data aquisition & clustering

- Scanned Chembl only for available datapoints
- total of ~2400 entries
- only ~300 with usable IC50 values
- ADME optimisation yielded 159 drug-like molecules.
- clustering: largest cluster of 35 molecules

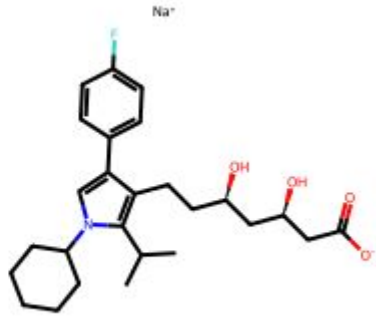
found three most potent structures:



CHEMBL348798

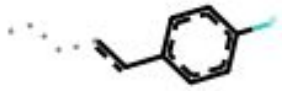


CHEMBL162642

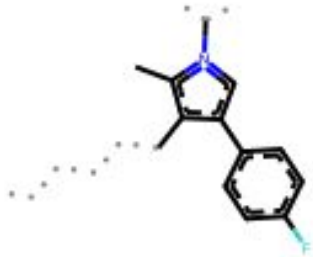


CHEMBL159951

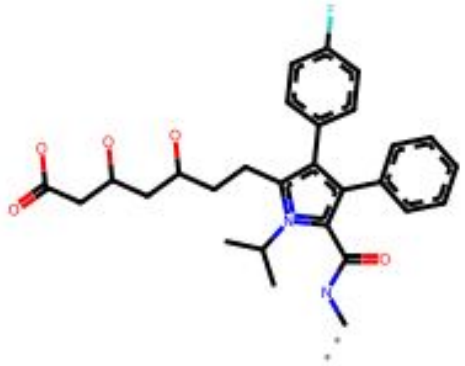
- Did a MCS (maximum common substructures):



MCS1



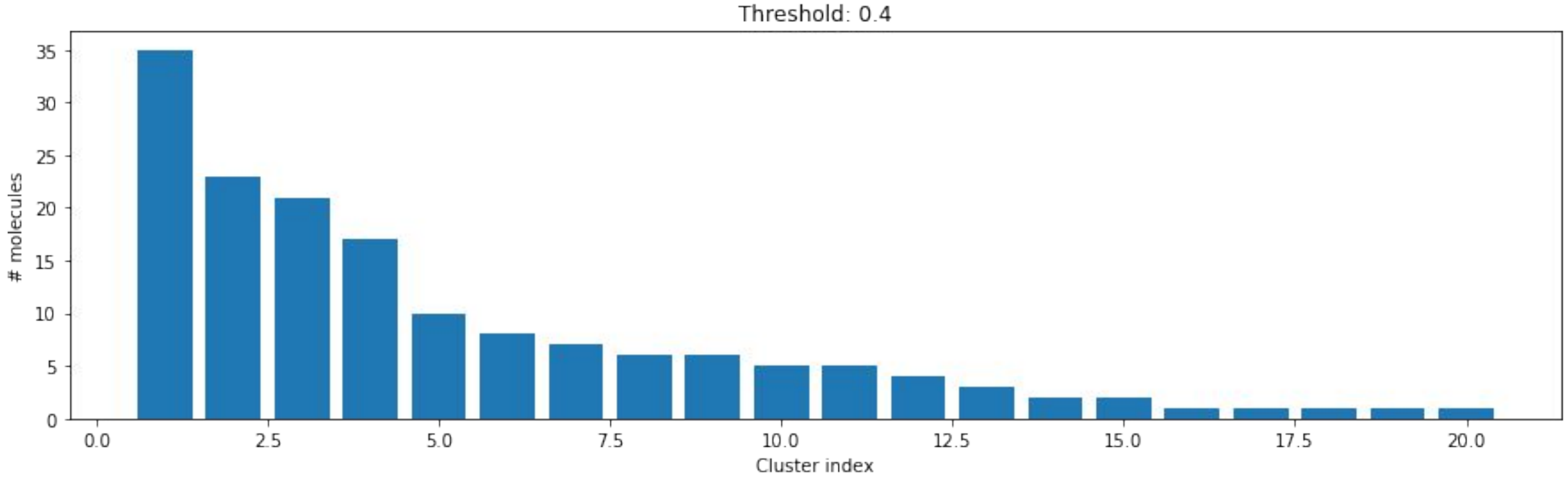
MCS2: +treshold=0.8



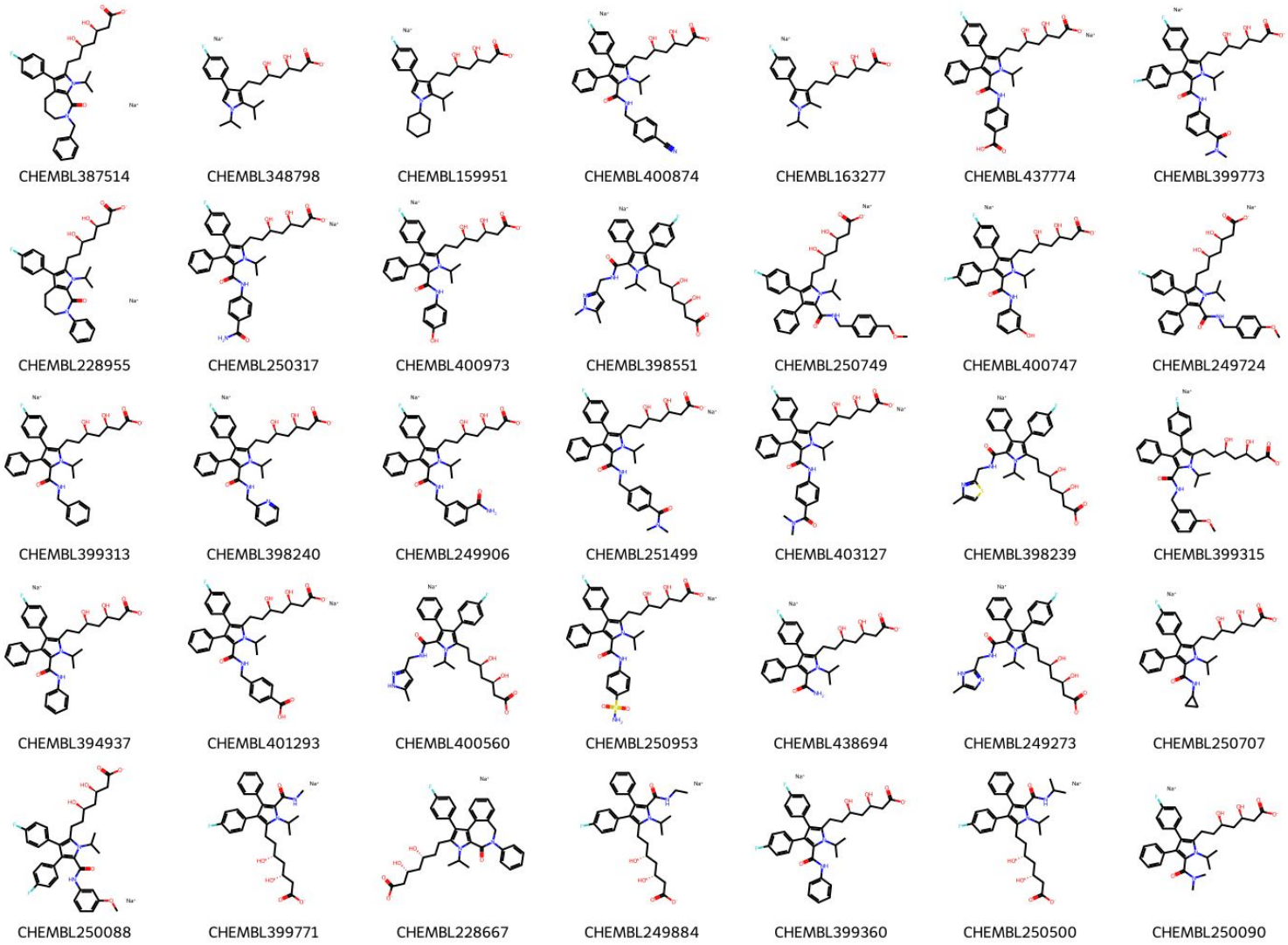
mcs3: +ringmatch

Data aquisition & clustering

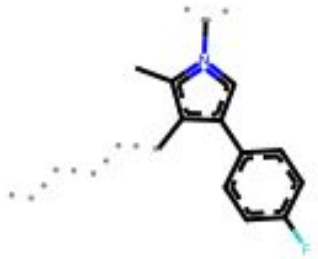
```
Number of clusters: 20 from 159 molecules at distance cut-off 0.40
Number of molecules in largest cluster: 35
Similarity between two random points in same cluster: 0.64
Similarity between two random points in different cluster: 0.40
```



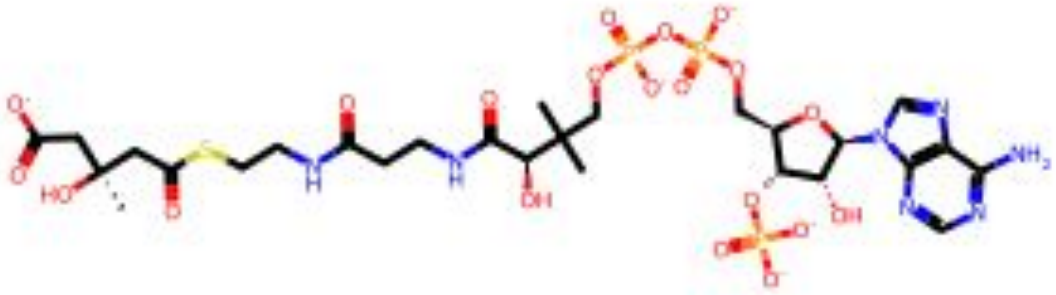
Data aquisition & clustering



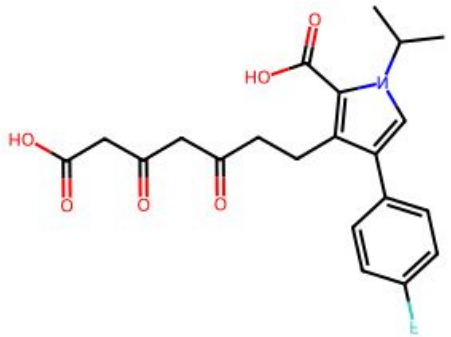
Scaffold - Ligand Design



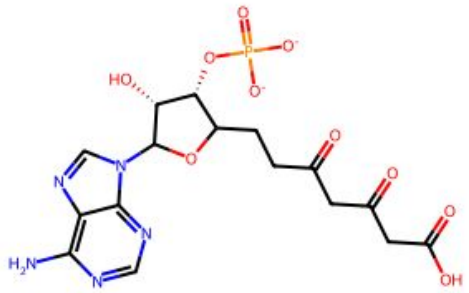
MCS2: +treshold=0.8



HMG-CoA



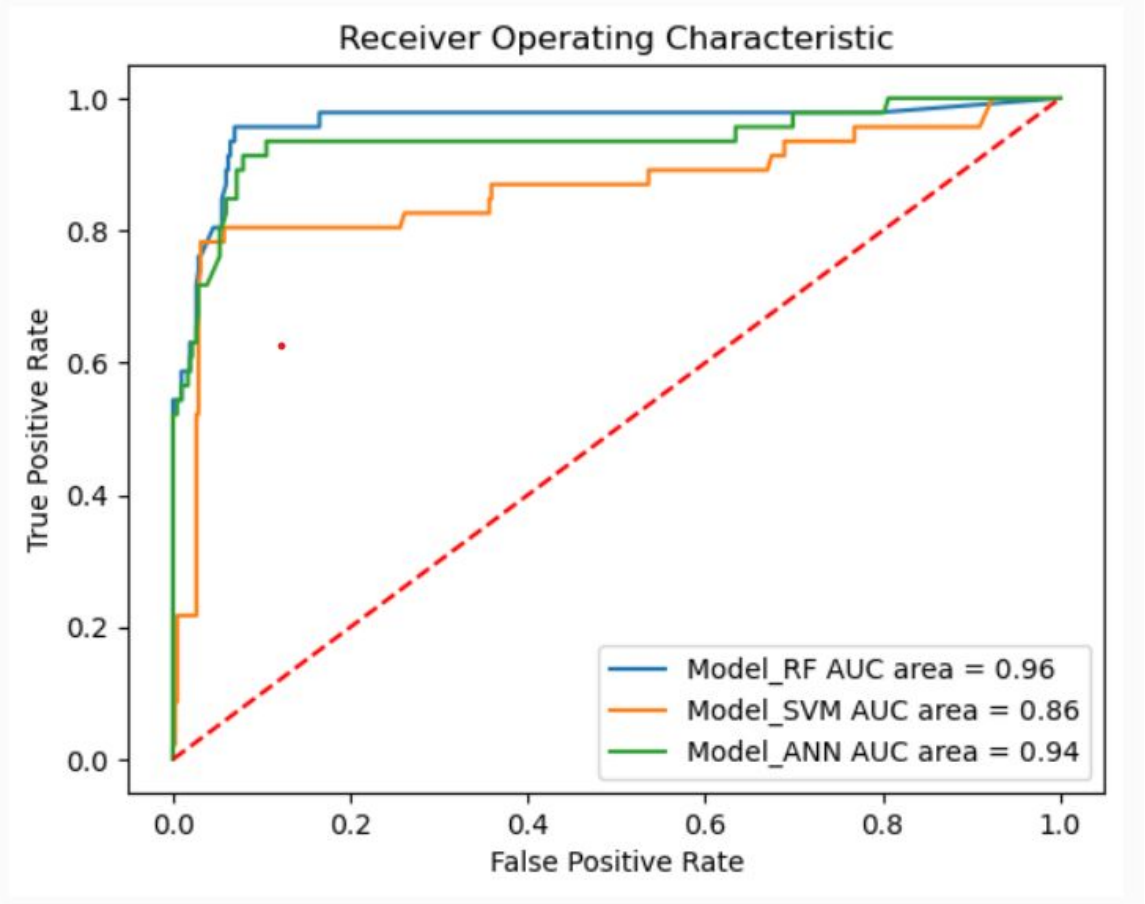
Self-Designed LIGAND001



Self-Designed LIGAND002

Machine learning: classification

- RF and ANN performed the best: similar AUC values
- Trained on all available data
  - Not filtered on Lipinski's rule of 5
  - Trained on pChEMBL instead of IC50




```
=======
Model_RF
Mean accuracy: 0.94     and std : 0.01
Mean sensitivity: 0.68  and std : 0.10
Mean specificity: 0.97  and std : 0.00
Mean AUC: 0.96  and std : 0.01
Time taken : 34.62s


=======
Model_SVM
Mean accuracy: 0.90     and std : 0.02
Mean sensitivity: 0.18  and std : 0.06
Mean specificity: 0.98  and std : 0.00
Mean AUC: 0.85  and std : 0.04
Time taken : 172.07s


=======
Model_ANN
Mean accuracy: 0.94     and std : 0.01
Mean sensitivity: 0.62  and std : 0.09
Mean specificity: 0.98  and std : 0.01
Mean AUC: 0.96  and std : 0.02
Time taken : 150.69s
```

Machine learning regression

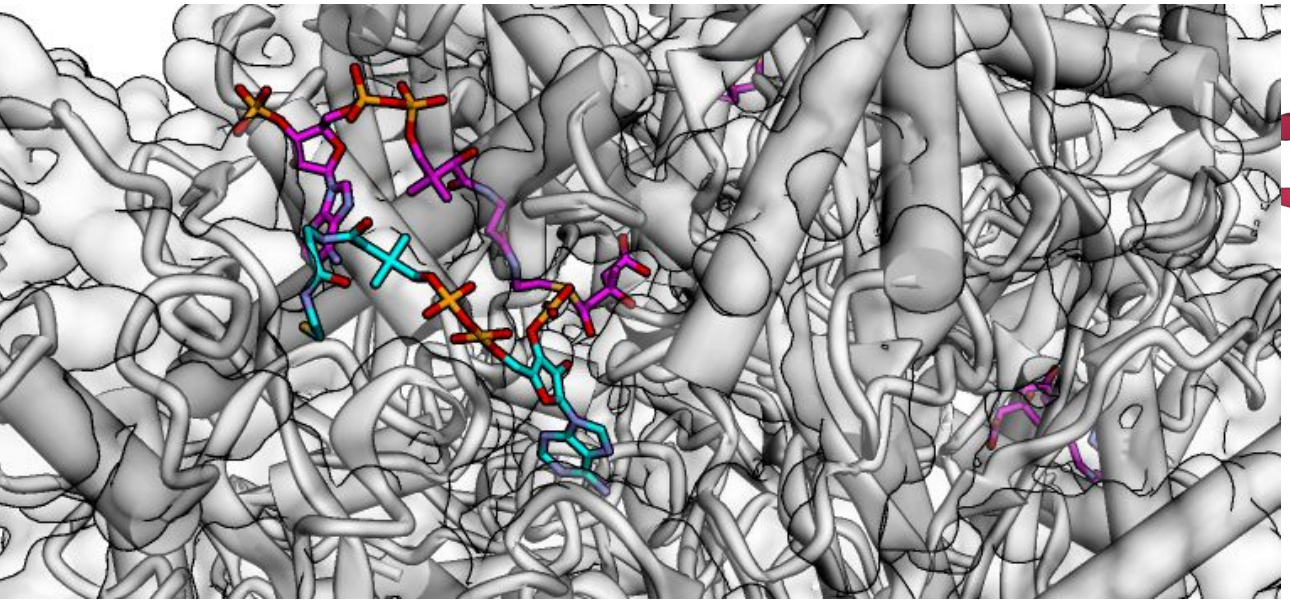
- Few datapoints → high MAE and RMSE values
  - Values around 0.6 are considered decent
- Despite that, some pChEMBL values came close:

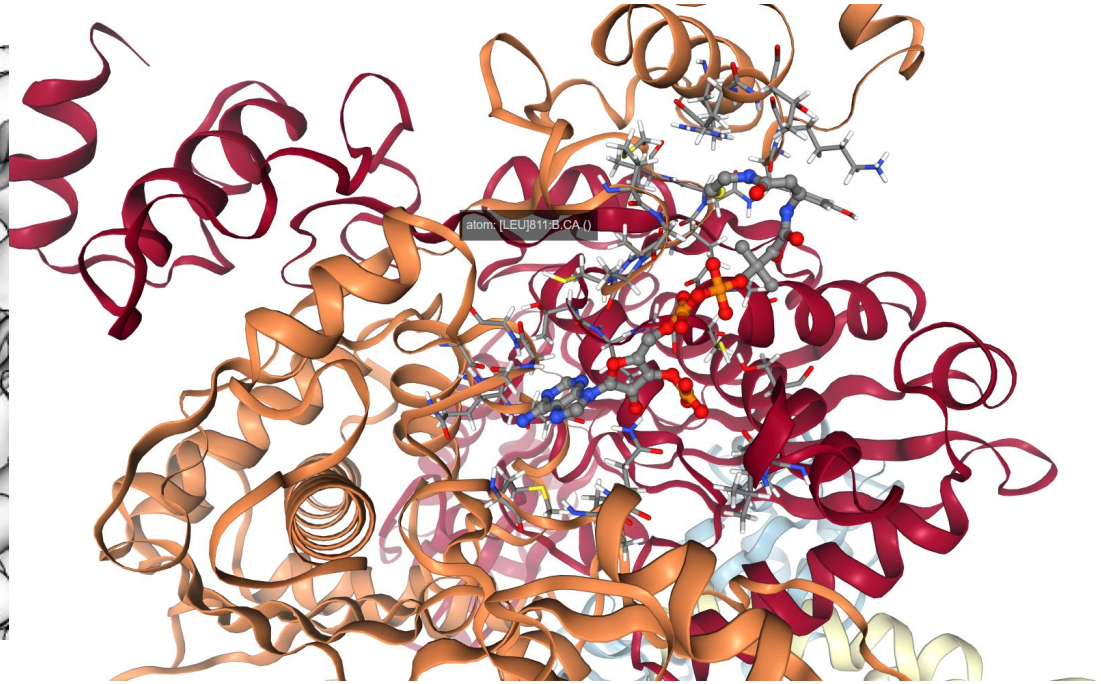
	Value	Std
MAE	0.73	0.11
RMSE	1.03	0.17

Name	Function	Predicted pChEMBL value (QSAR)	Predichted pChEMBL value (Docking)	Actual pChEMBL value
3HI / CHEMBL1565	HMG-CoA Inhibitor	7.92	5.08	7.56
LIGAND001	Self-designed ligand	8.40	5.30	-
LIGAND002	Self-designed ligand	5.87	5.28	-
HMG-CoA	Substrate	5.62	4.84	-
Atorvastatin	Approved drug (HMG-CoA inhibitor)	8.63	5.42	8.21

docking

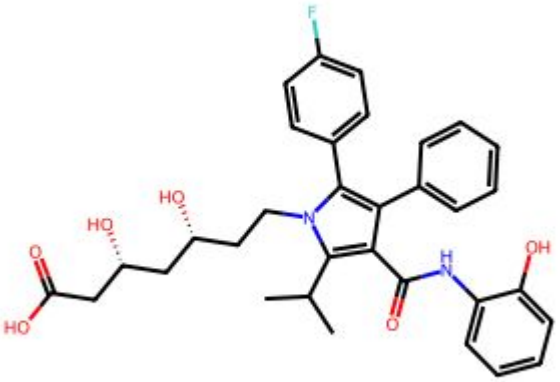
HMG-COA and HMG-COA
The affinity predicted by Vina = -6.603 (kcal/mol)
Which (approximately) corresponds to a pChEMBL of 4.84

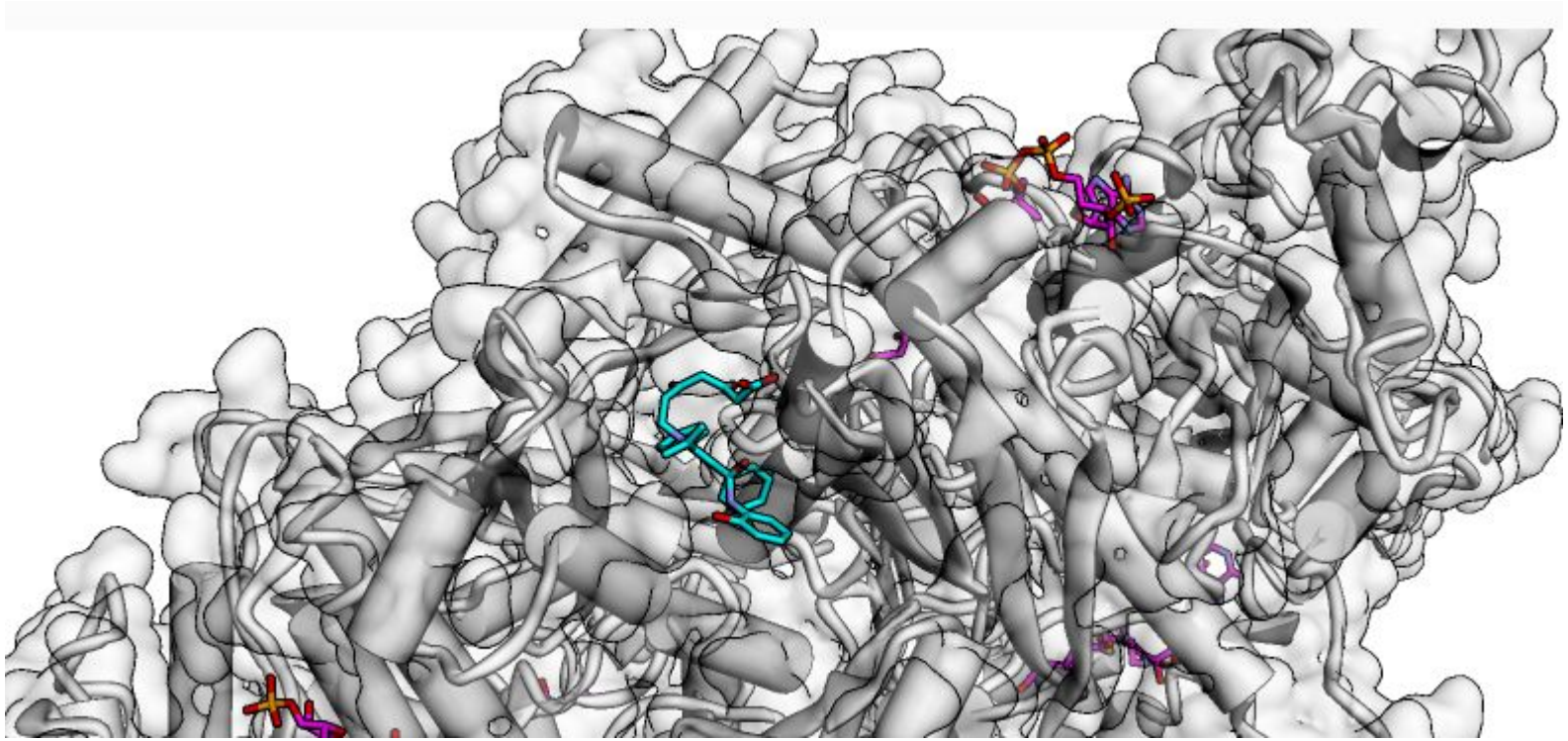




docking

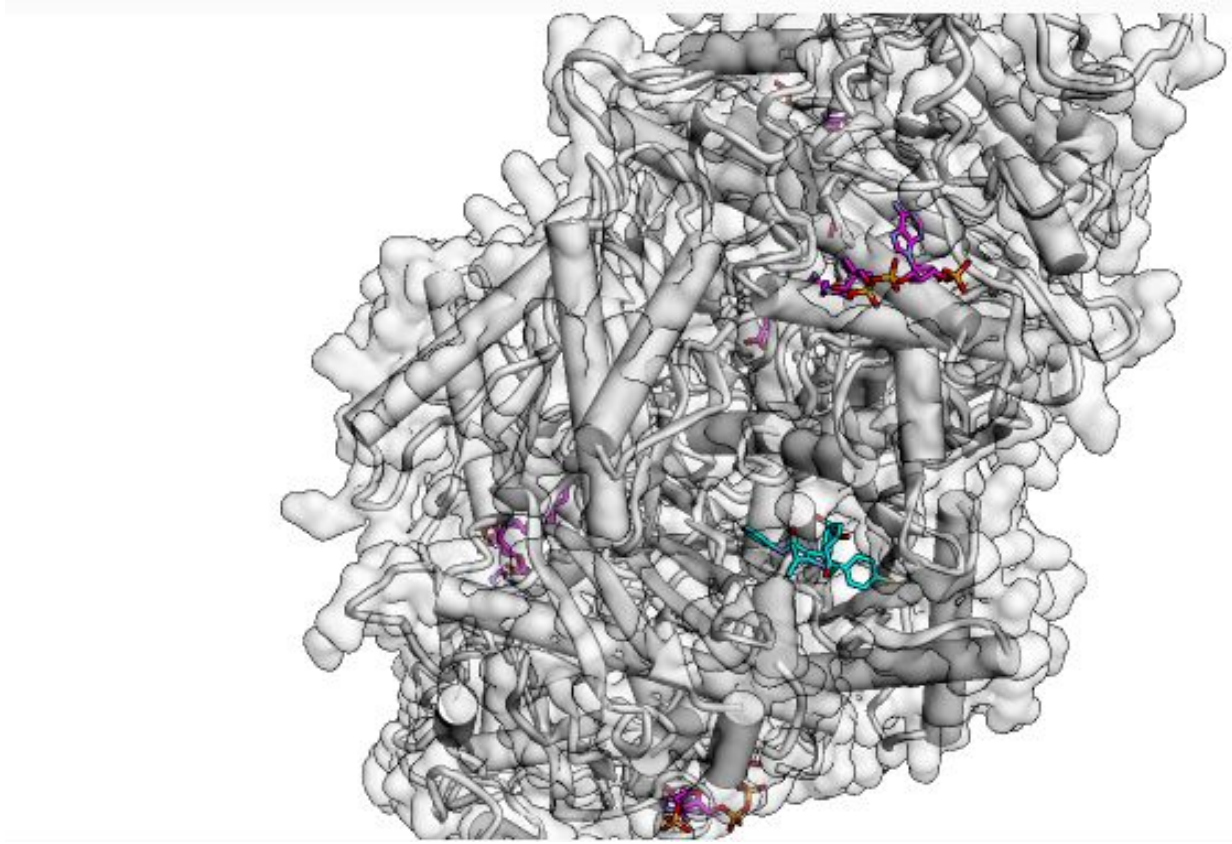
HMG-COA and PDB code 3HI
docking shows difficulties
The affinity predicted by Vina = -6.926 (kcal/mol)
Which (approximately) corresponds to a pChEMBL of 5.08

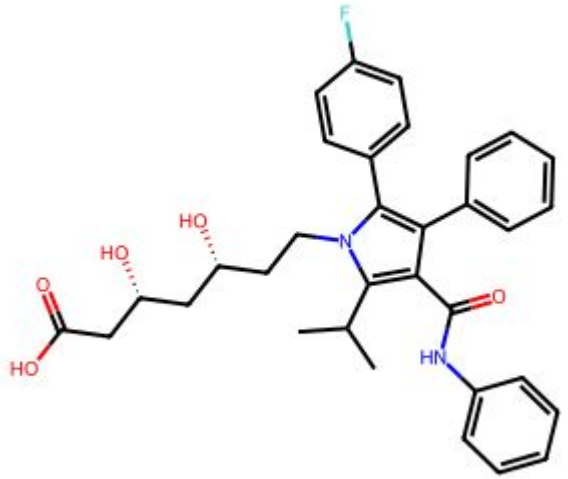




docking

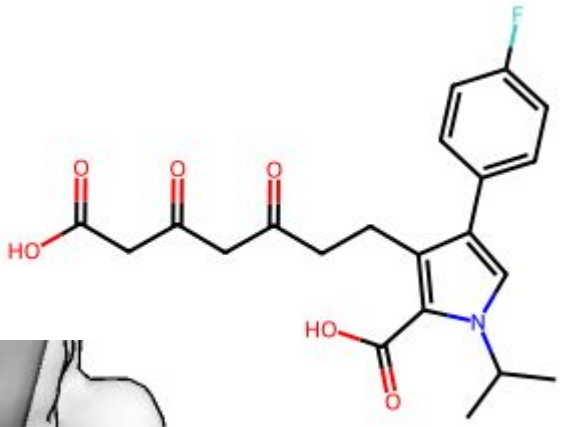
HMG-COA and Avorstatin
The affinity predicted by Vina = -7.383 (kcal/mol)
Which (approximately) corresponds to a pChEMBL of 5.42

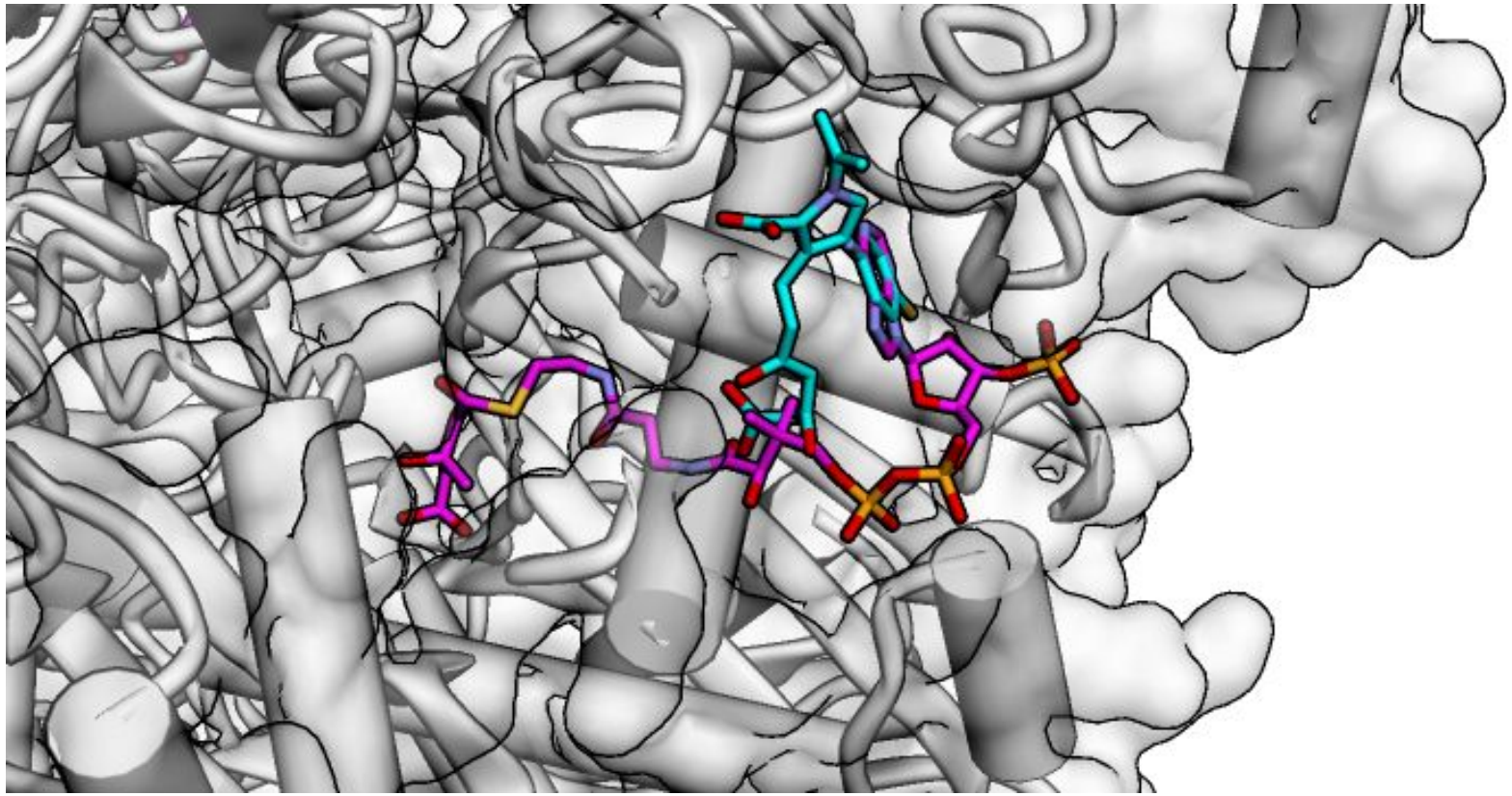




docking

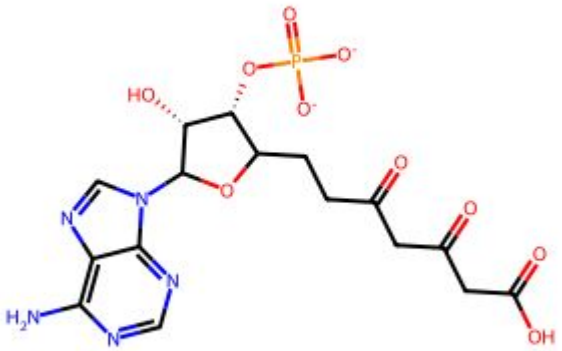
HMG-COA and LIGAND_001
The affinity predicted by Vina = -7.229 (kcal/mol)
Which (approximately) corresponds to a pChEMBL of 5.30

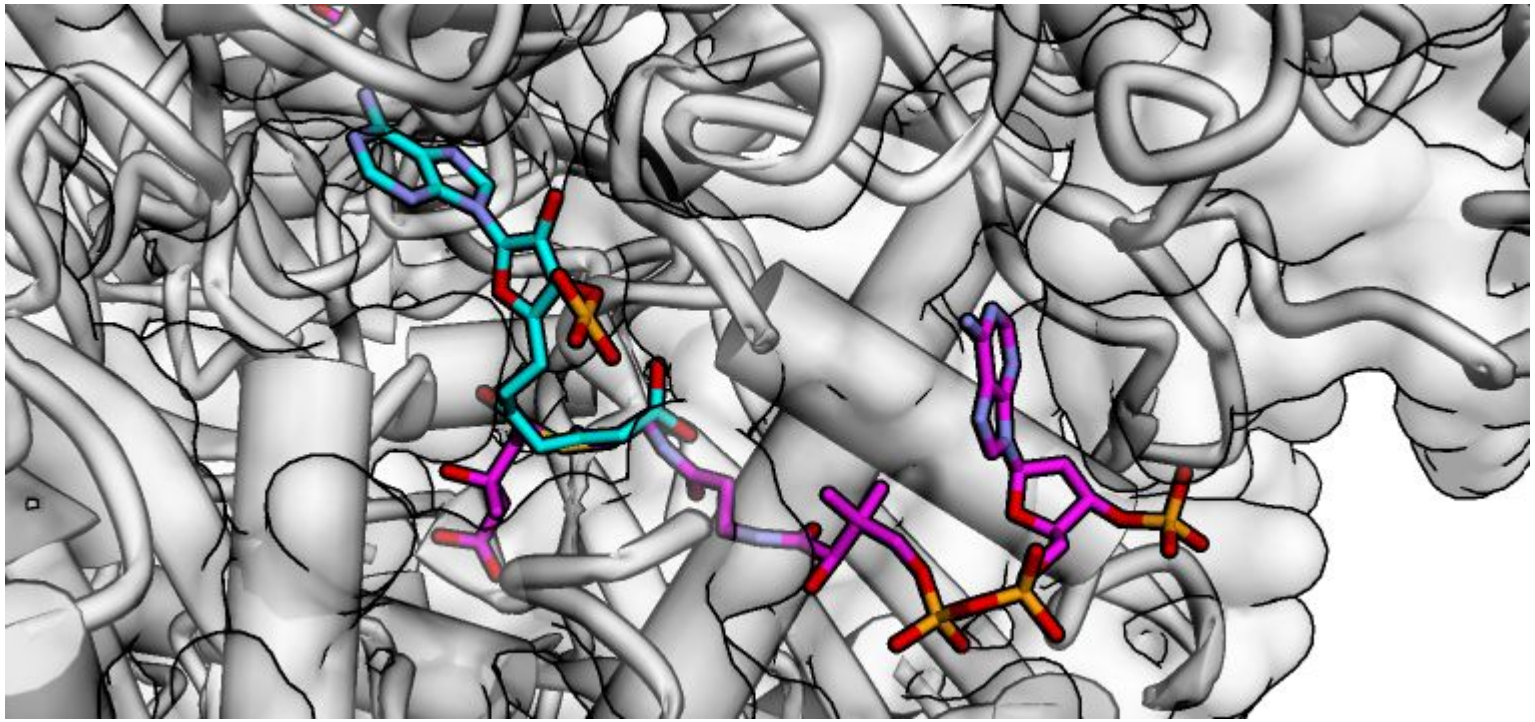




docking

HMG-COA and Ligand002
The affinity predicted by Vina = -7.197 (kcal/mol)
Which (approximately) corresponds to a pChEMBL of 5.28





Overview

Name	Function	Predicted pChEMBL value (QSAR)	Predichted pChEMBL value (Docking)	Actual pChEMBL value
3HI / CHEMBL1565	HMG-CoA Inhibitor	7.92	5.08	7.56
LIGAND001	Self-designed ligand	8.40	5.30	-
LIGAND002	Self-designed ligand	5.87	5.28	-
HMG-CoA	Substrate	5.62	4.84	-
Atorvastatin	Approved drug (HMG-CoA inhibitor)	8.63	5.42	8.21

Thanks for listening

Universiteit
Leiden
The Netherlands

Discover the world at Leiden University