SORBONNE UNIVERSITÉ

MEET-EU

TEAM SORBONNE 6

Henri GUILHON

Mathys DELATTRE

Lorenzo BALLESTRA

QUENTIN DESPRETZ

CONTENT

INTRODUCTION

- Having a wide repertoire of drugs is important.
- Docking is very time consuming.
- Identify good candidates time efficiently.

OBJECTIVES

Objective n° 1

Develop a deep neural network capable of finding good ligand candidates.

Objective n° 2

Save time by avoiding to dock every molecule.

METHODS

A) Docking with Vina

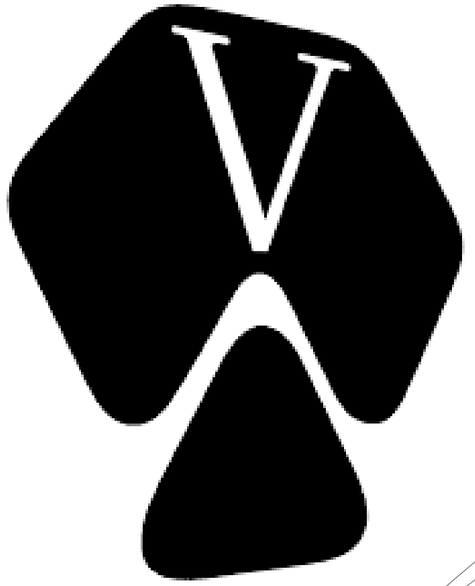
B) Docking Map

C) Creating Graphs

D) Model Training

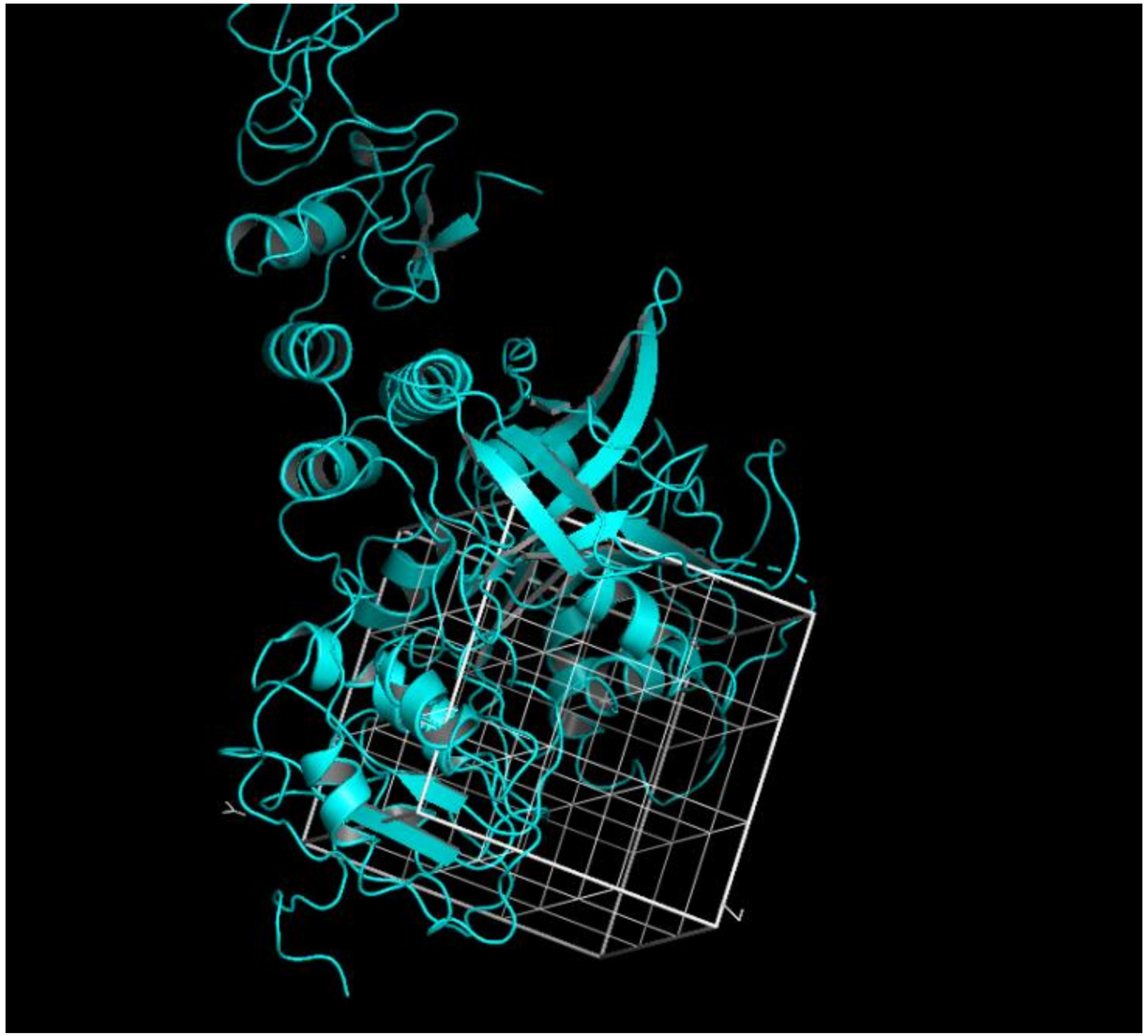
DOCKING WITH VINA

- Preparing ligand and receptor : ADFR Suite
  - PDB to PDBQT format
- Searching for a search universe
- Autodock Suite : Vina



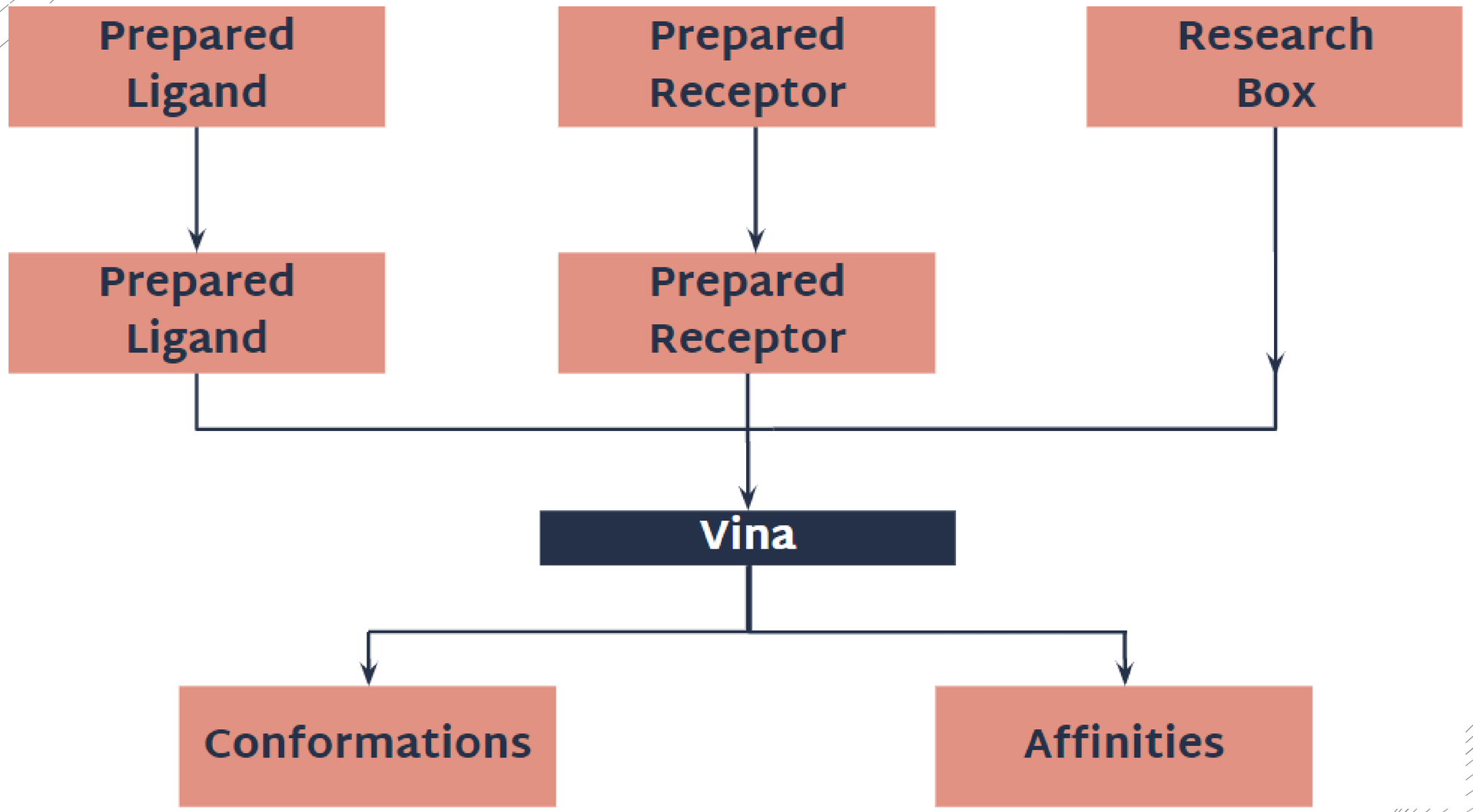
Vina : https://onlinelibrary.wiley.com/doi/10.1002/jcc.21334
https://pubs.acs.org/doi/10.1021/acs.jcim.1c00203

DOCKING MAP



https://doi.org/10.1038/s41467-021-25166-6

DOCKING MAP


Prepared Ligand
Prepared Receptor
Research Box
Prepared Ligand
Prepared Receptor
Vina
Conformations
Affinities

```
MODEL 1
REMARK VINA RESULT:    -6.869      0.000      0.000
REMARK INTER + INTRA:          -9.938
REMARK INTER:                  -8.877
REMARK INTRA:                  -1.061
REMARK UNBOUND:                -1.061
REMARK  5 active torsions:
REMARK  status: ('A' for Active; 'I' for Inactive)
REMARK    1  A    between atoms: C2_2  and  C3_3
REMARK    2  A    between atoms: C3_3  and  C4_4
REMARK    3  A    between atoms: C3_3  and  C6_6
REMARK       I    between atoms: C6_6  and  N1_8
REMARK    4  A    between atoms: N1_8  and  C7_9
REMARK    5  A    between atoms: N3_14  and  C10_15
ROOT
HETATM    1  C7  UNL     1     410.168 -23.901  76.160  1.00  0.00     0.139 C
HETATM    2  C8  UNL     1     410.515 -23.513  77.436  1.00  0.00     0.027 A
HETATM    3  C9  UNL     1     411.627 -22.677  77.286  1.00  0.00     0.051 A
HETATM    4  N2  UNL     1     411.986 -22.502  76.022  1.00  0.00    -0.295 N
ATOM      5  H2  UNL     1     412.806 -22.013  75.661  1.00  0.00     0.187 HD
HETATM    6  N3  UNL     1     410.999 -23.132  75.313  1.00  0.00    -0.239 N
HETATM    7  C16 UNL     1     412.300 -22.150  78.489  1.00  0.00     0.199 A
HETATM    8  S1  UNL     1     411.318 -22.869  79.835  1.00  0.00     0.154 S
HETATM    9  O2  UNL     1     410.620 -21.794  80.518  1.00  0.00    -0.227 OA
HETATM   10  O3  UNL     1     412.154 -23.800  80.570  1.00  0.00    -0.227 OA
HETATM   11  C17 UNL     1     410.118 -23.813  78.831  1.00  0.00     0.185 A
ENDROOT
BRANCH   1  12
HETATM   12  N1  UNL     1     409.279 -24.859  75.685  1.00  0.00    -0.311 N
```

```
Scoring function : vina
Rigid receptor: Target.pdbqt
Ligand: EOS1033/EOS1033.pdbqt
Grid center: X 405.56 Y -25.45 Z 72.5
Grid size  : X 30 Y 30 Z 30
Grid space : 0.375
Exhaustiveness: 64
CPU: 0
Verbosity: 1

Computing Vina grid ... done.
Performing docking (random seed: -841425399) ...
0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
**************************************************

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1       -6.869          0          0
   2       -6.736      1.294      1.397
   3       -6.668      4.063      6.534
   4       -6.583      2.277      4.849
   5       -6.492       2.08      4.823
   6       -6.421       3.79      6.628
   7       -6.414      3.778      5.853
   8       -6.196      3.556      6.027
```

CREATING GRAPHS

Atomic Number, Degree, Formal charge, Is Aromatic

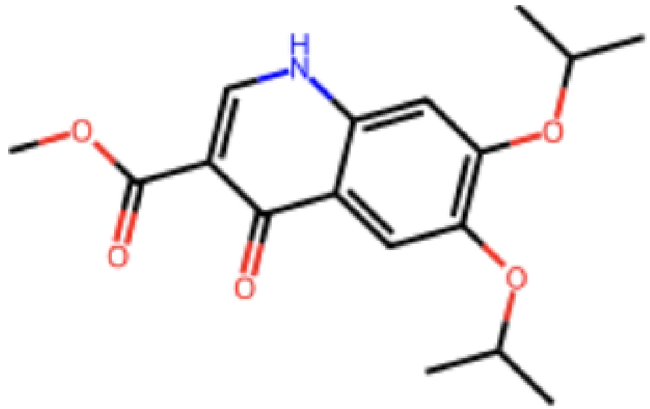
Atom 1 [6,4,0,0]

Atom 2 [6,3,0,1]

[...]

Atom n [8,1,0,1]

Rdkit molecule :



Graph.edge_index:

Edge 1 Edge n

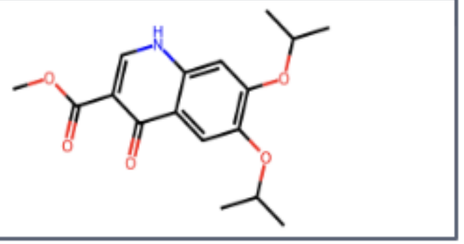
[0,1,...,18,22]

[1,0,...,22,18]

Rdkit Chem Package : https://www.rdkit.org/docs/source/rdkit.Chem.html

MODEL TRAINING

Input



Layer 1

Returns 16 channels

Layer 2

Returns 32 channels

Layer 3

Returns 8 channels

Pooling

Returns a one dimensional vector

Linear layer

Returns the predicted affinity

Evaluation :
Mean Squared Error Loss

Training :
Adam Optimizer with a learning rate = 0.001

Training data :
373 molecules

Testing data :
161 molecules

Network inspiration : github.com/vaiteaopuu/gnn_mol_example
Thomas N. Kipf, Max Welling, 2016, Semi-Supervised Classification with Graph Convolutional Networks, https://arxiv.org/abs/1609.02907

RESULTS

A) Best model Parameters

B) Testing model Accuracy

C) Ranking of Molecules

D) Predicting molecule affinities

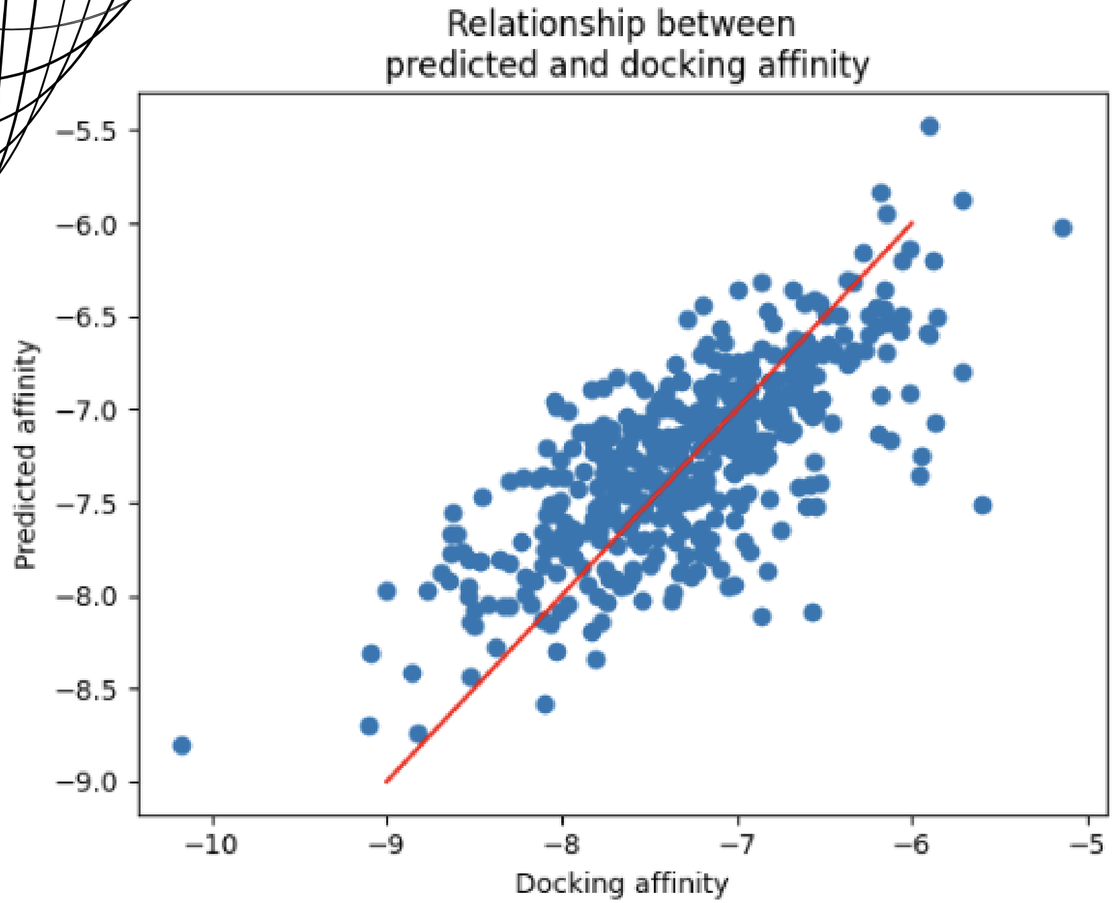
BEST MODEL PARAMETERS

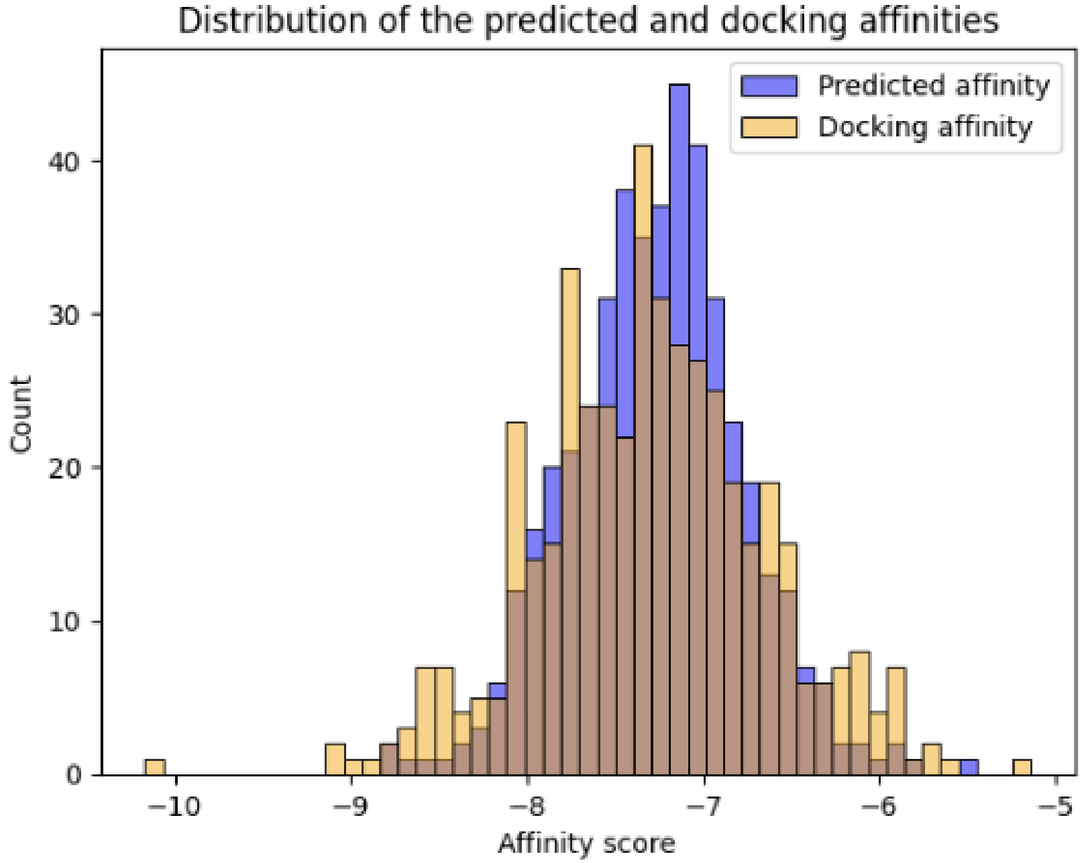
1. Batches extend the training time
2. Batches lowers the accuracy
3. Adding Hydrogen extends the training time
4. Adding Hydrogen improves the accuracy

Full batch + Hydrogen

	Hydrogen	No Hydrogen
Full Batch	0.170	0.223
Batch of 32	0.202	0.220
Batch of 16	0.224	0.226

TESTING MODEL ACCURACY

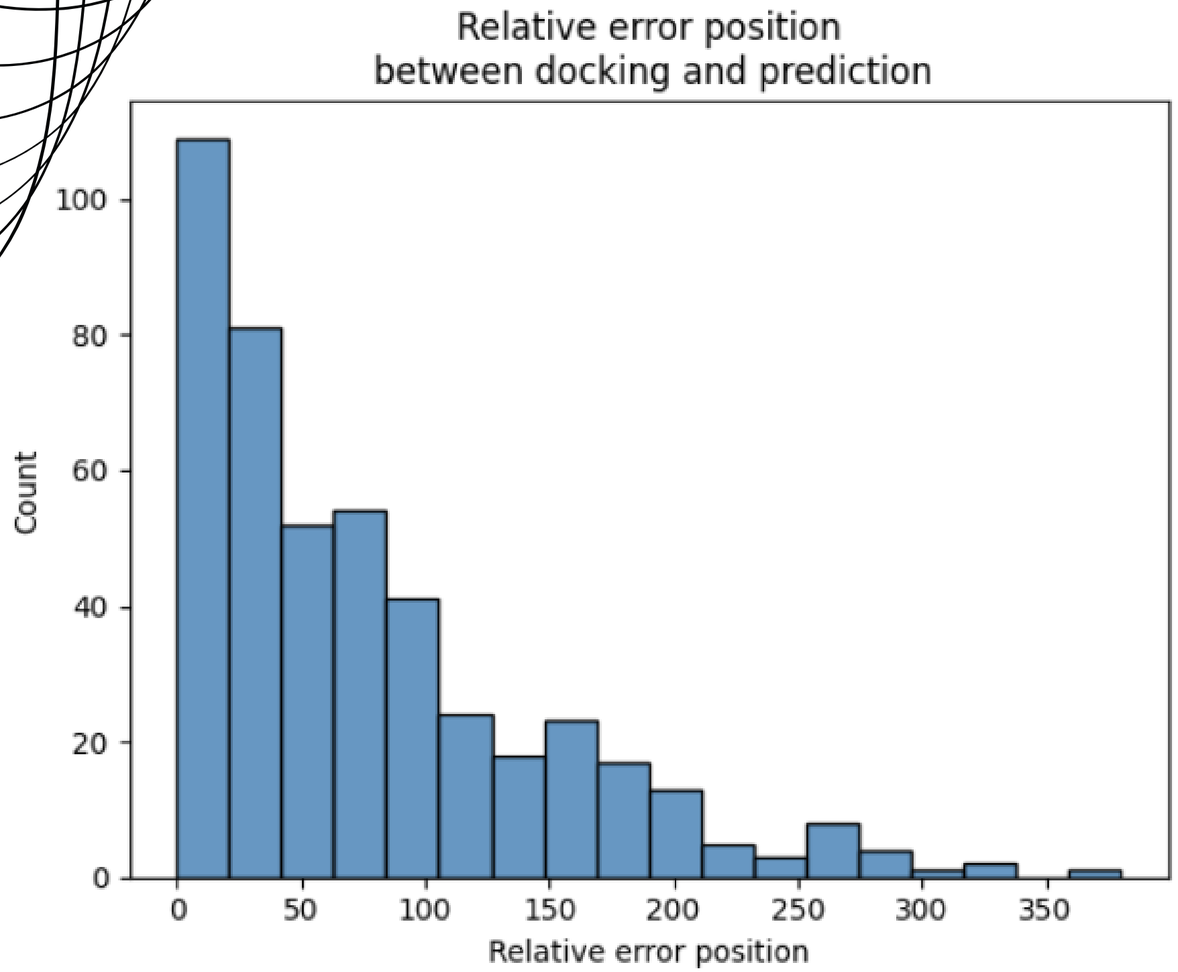




- *The mean of the prediction is -7.27*
- *The standard deviation is 0.48*

- *The mean of the validation set is -7.30.*
- *The standard deviation is 0.65*

RANKING THE MOLECULES



456 sorted molecules

Compare position with the validation set

Mostly well predicted

Some really badly ranked

PREDICTING MOLECULE AFFINTIES

EOS	Docking affinity
101357	-9.631
102307	-7.507
101186	-6.559
100233	-8.048
101072	-8.513
101472	-8.083
93	-9.247
1195	-10.18

EOS	Docking affinity
100814	-7.56
101803	-8.026
264	-8.78
100811	-7.993
1824	-7.304
100687	-7.708
101096	-8.189
98622	-8.45

PREDICTING MOLECULE AFFINITIES

EOS	Docking affinity
101357	-9.631
102307	-7.507
101186	-6.559
100233	-8.048
101072	-8.513
101472	-8.083
93	-9.247
1195	-10.18

EOS	Docking affinity
100814	-7.56
101803	-8.026
264	-8.78
100811	-7.993
1824	-7.304
100687	-7.708
101096	-8.189
98622	-8.45

- Mean affinity : -8.26
- Great to extract a promising sample
- Not good at extracting absolute best molecules

PERSPECTIVES

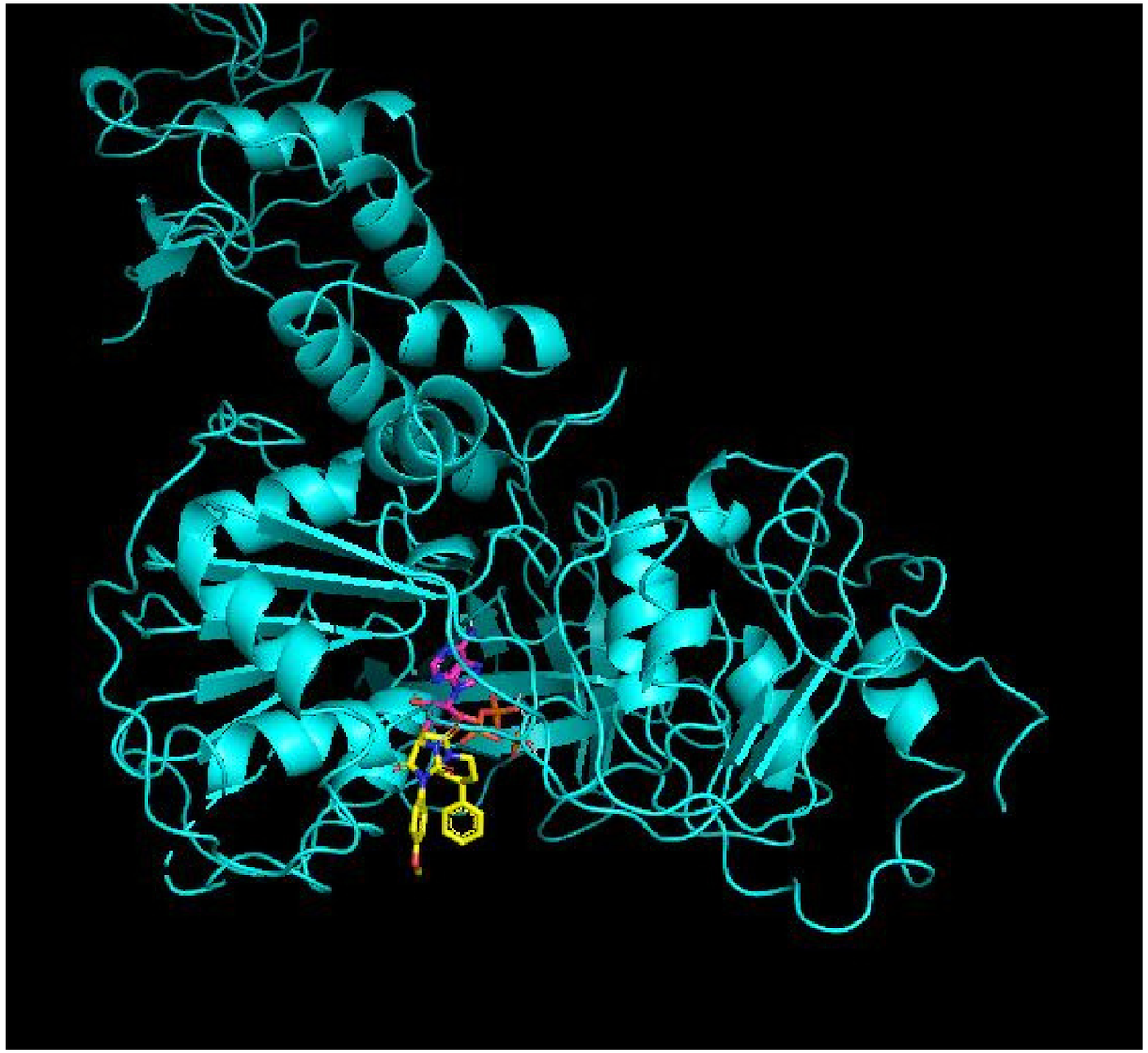
A) Introducing Pocket Occupation

B) Calculating PO rate

C) Selecting a conformation

D) Was the selection needed ?

POCKET OCCUPATION



POCKET OCCUPATION

- Simplify amino acid by their alpha carbon (AC)
- Register distances of closest AC to all ATP atoms
    - Mean distance : 4.7Å
- Register all AC close to an ATP atom < 5Å
- For all 9 conformations : taking only AC found more or equal to 10 times
    - Pool of reference AC

POCKET OCCUPATION RATE

- Register all atoms of the molecule at least 5Å of a reference AC
    - Found : 1
    - Not Found : 0

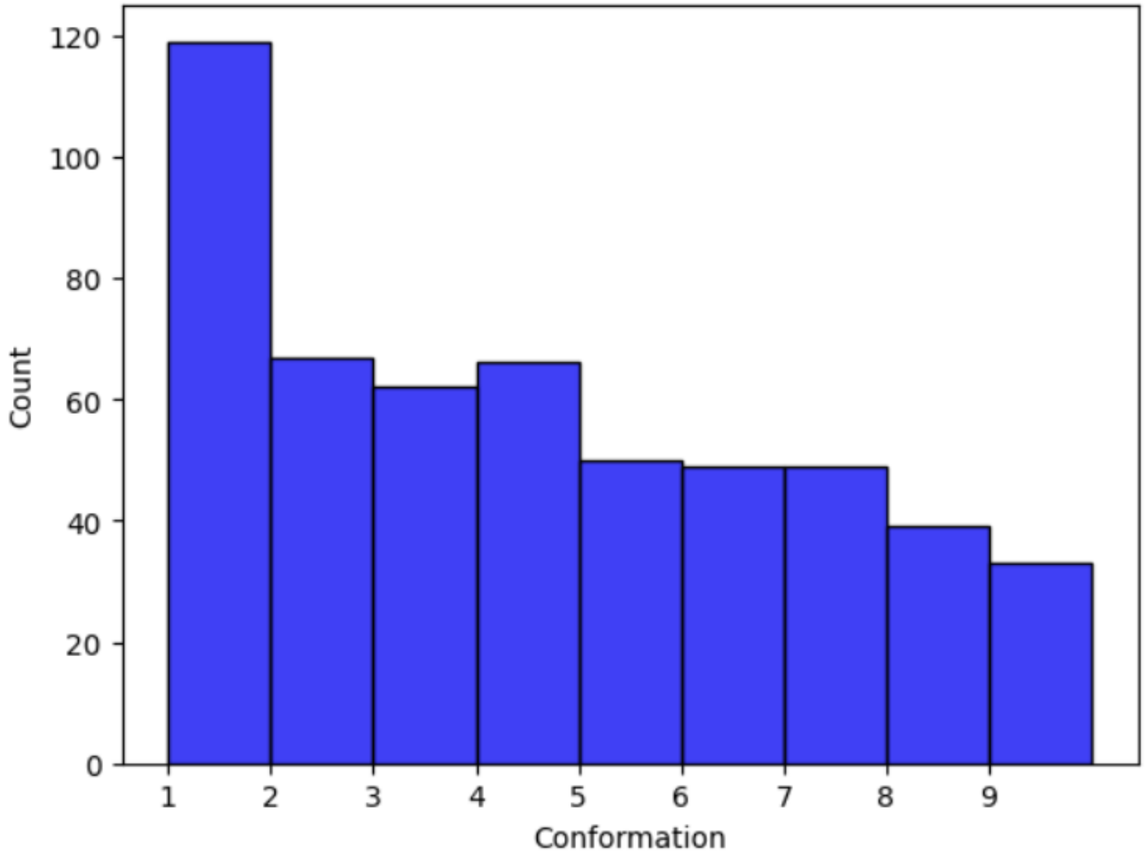
- Calculating fraction of found atoms

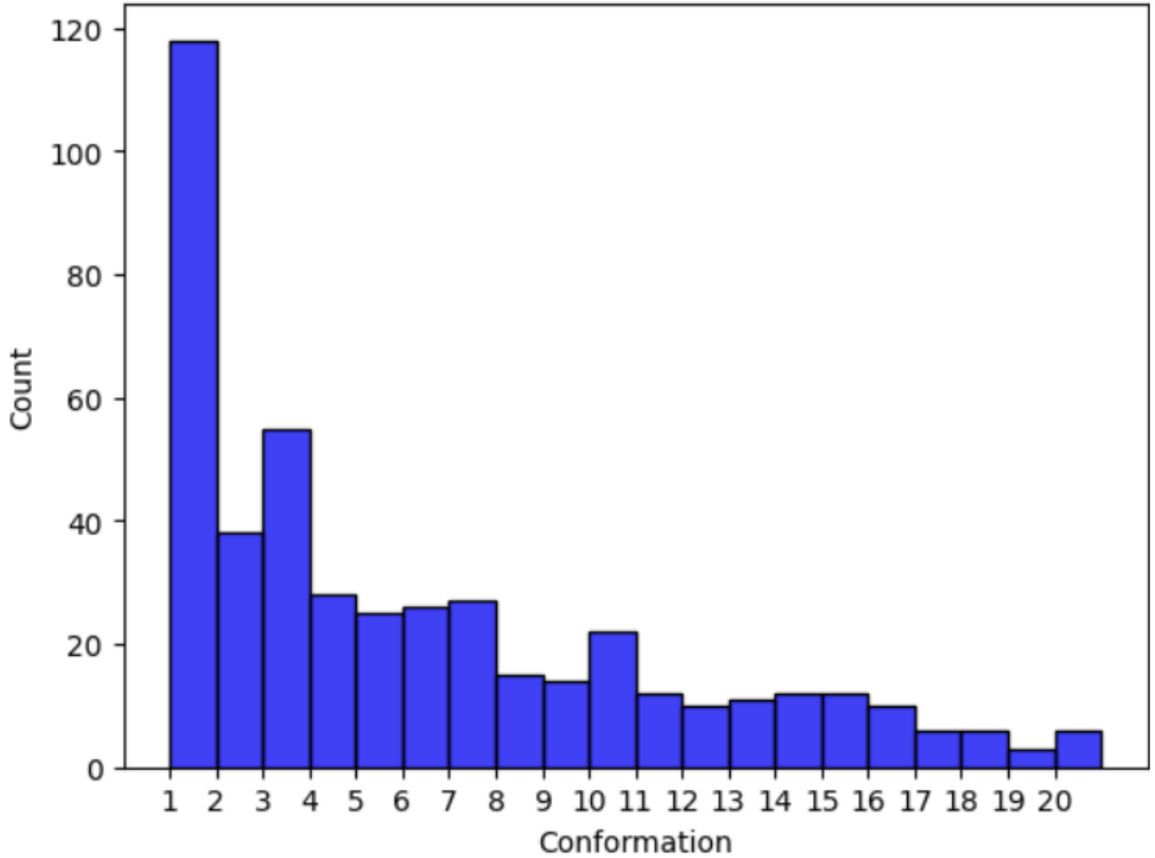
SELECTING A CONFORMATION

```
9
EOS1023
Conformation 1:  0.818% < 5 A with affinity -6.774
Conformation 2:  0.773% < 5 A with affinity -6.718
Conformation 3:  0.0% < 5 A with affinity -6.37
Conformation 4:  0.636% < 5 A with affinity -6.288
Conformation 5:  0.0% < 5 A with affinity -6.261
Conformation 6:  0.682% < 5 A with affinity -6.19
Conformation 7:  0.864% < 5 A with affinity -6.175
Conformation 8:  0.773% < 5 A with affinity -6.107
Conformation 9:  0.636% < 5 A with affinity -6.071
Best conformation: 7 with affinity : 0.636
```

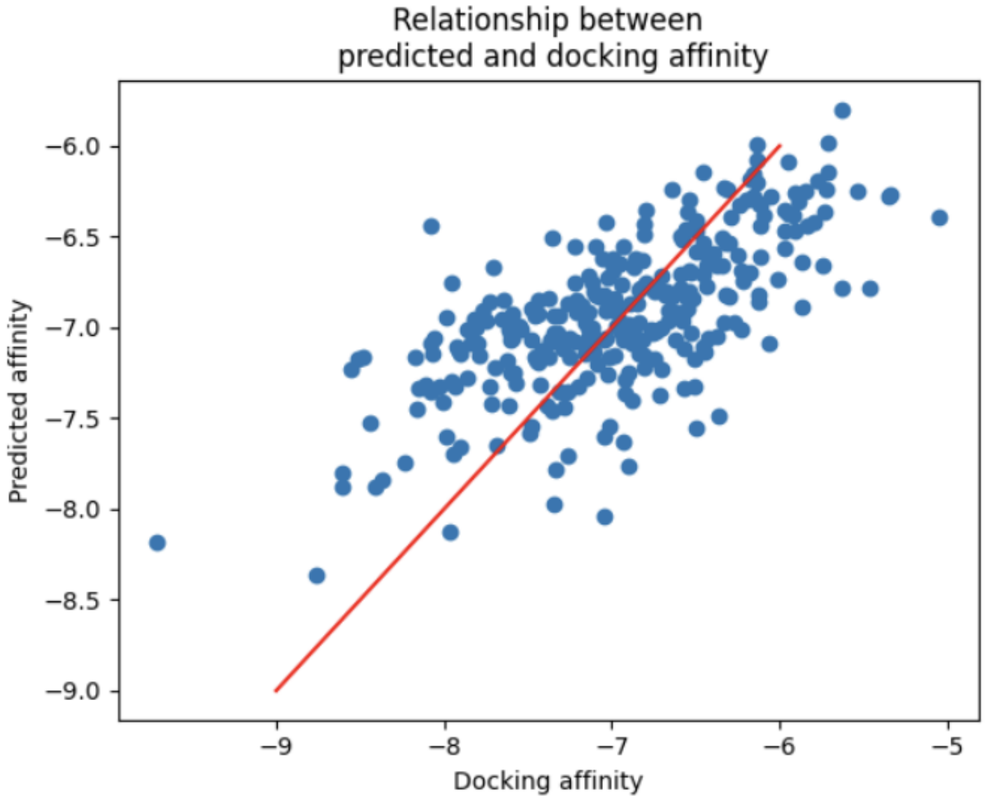
```
8
EOS1033
Conformation 1:  0.654% < 5 A with affinity -6.869
Conformation 2:  0.692% < 5 A with affinity -6.736
Conformation 3:  0.731% < 5 A with affinity -6.668
Conformation 4:  0.731% < 5 A with affinity -6.583
Conformation 5:  0.615% < 5 A with affinity -6.492
Conformation 6:  0.769% < 5 A with affinity -6.421
Conformation 7:  0.692% < 5 A with affinity -6.414
Conformation 8:  0.731% < 5 A with affinity -6.196
Best conformation: 6 with affinity : 0.731
```

WAS THE SELECTION NEEDED ?

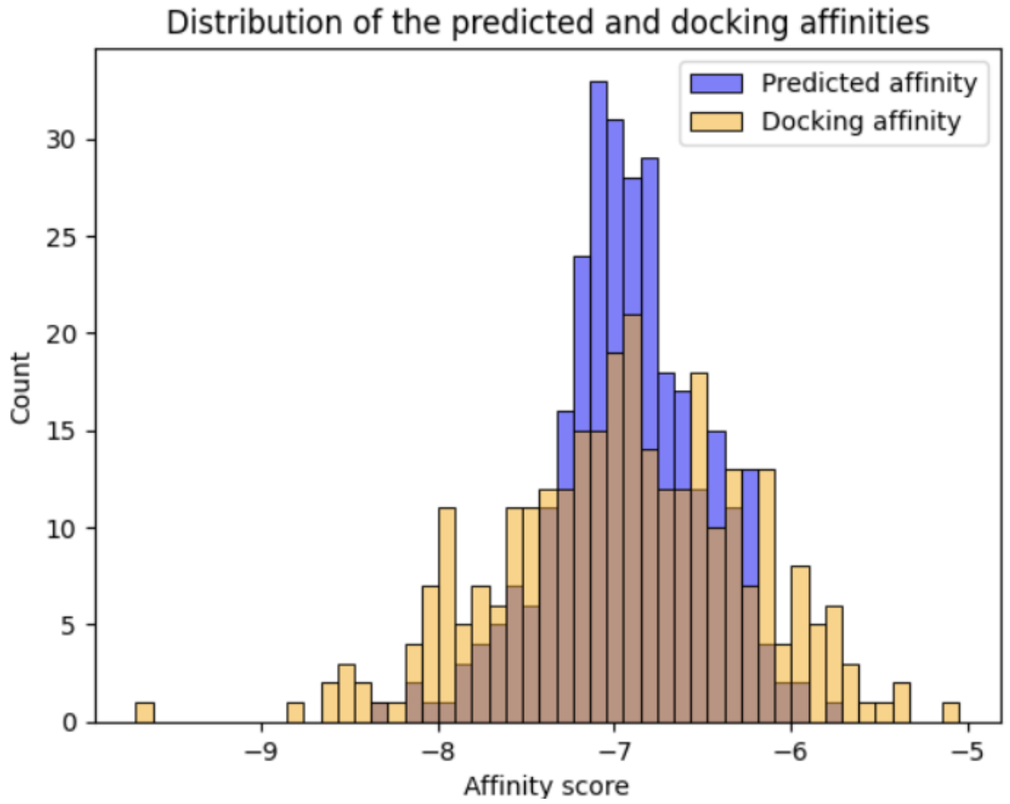




WAS THE SELECTION NEEDED ?



- *The mean of the prediction is -6.925*
- *The standard deviation is 0.42*



- *The mean of the validation set is -6.957*
- *The standard deviation is 0.71*

CONCLUSION

Neural network's evaluation :

Adding hydrogens gives best performances. → Longer training time.

Full batch → Better results and faster training.
More sensitive to local minimums.

After comparison with docking results → Contains mostly high scoring ligands.
But also some bad candidates

CONCLUSION

Relevance of the approach :

Model capable of sorting the best candidates within seconds.

Predictions are not exact → But returns best candidates

Relying solely on affinity is not sufficient → Introducing “pocket occupation score”

DISCUSSION

About the network :

Need to explore other features of the ligand.

Some features may have been poorly selected.

Need to try different architectures and batch sizes.

About the metrics :

Affinity alone is not sufficient → Pocket occupation is an interesting metric

ACKNOWLEDGMENTS

- Vaitea OPUU for his help on building the neural network and the orientation of our project

- Elodie LAINE for introducing us to other criteria to assess the quality of the ligands

- Juliana SILVA BERNARDES for supervising our project

- Charles University for welcoming us