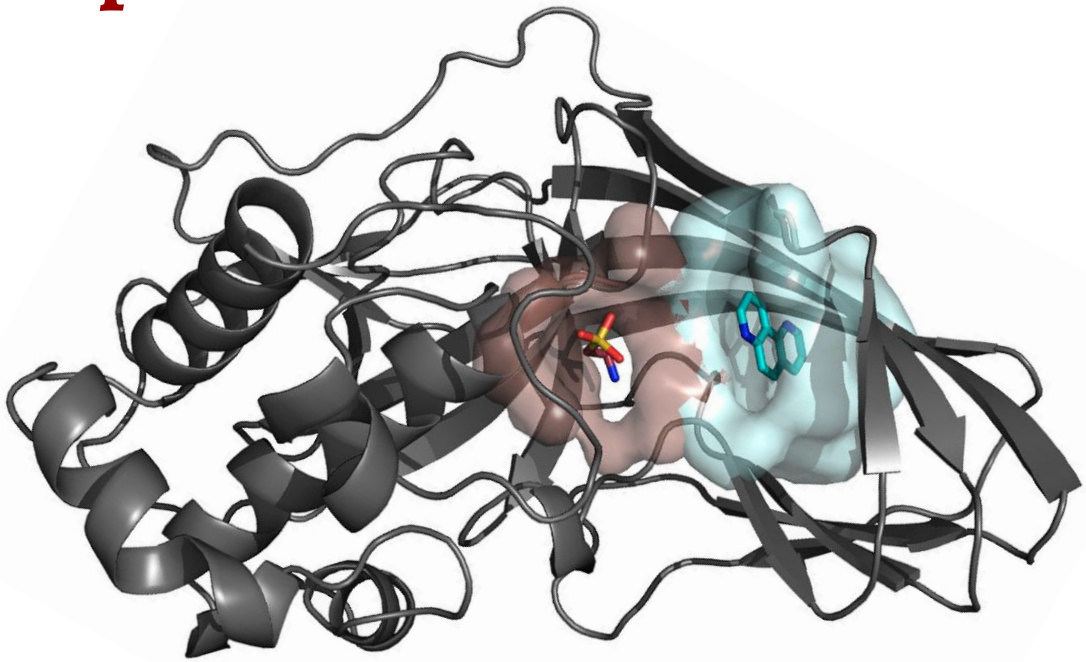
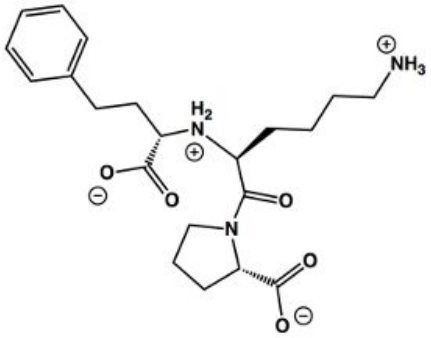
Classification of ligand-binding pockets in proteins



Nicolas Silva - Deep Learning Project - Octobre 2019

Context

- A lot of proteins ⇒ around 100 000 types
- All proteins bind with ligand:

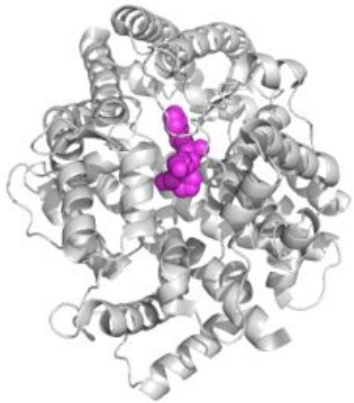


Ligand

\+

Protein

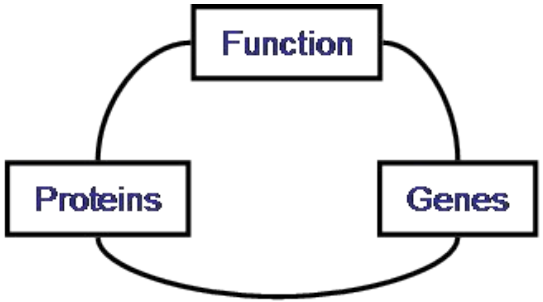
⇌



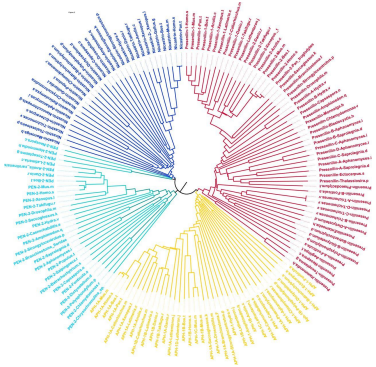
Ligand-protein complex

Context

Important to understand protein-ligand interactions



Molecular function

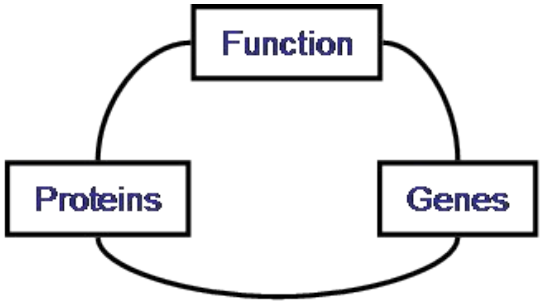


Evolutionary history

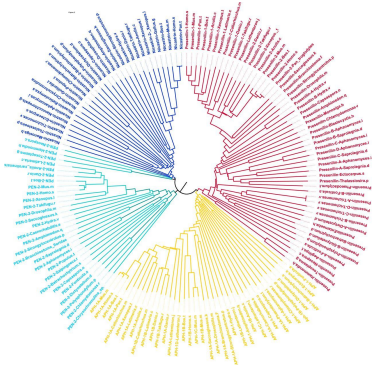
New medications

Context

Important to understand protein-ligand interactions



Molecular function

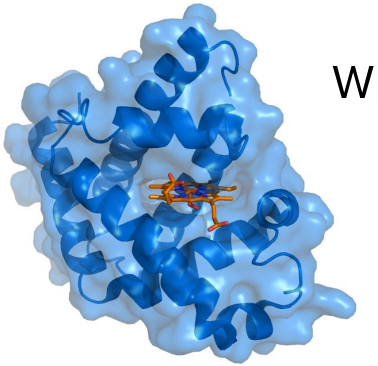


Evolutionary history

New medications

Problem ⇒ Great similarity between ligand-binding pockets in proteins

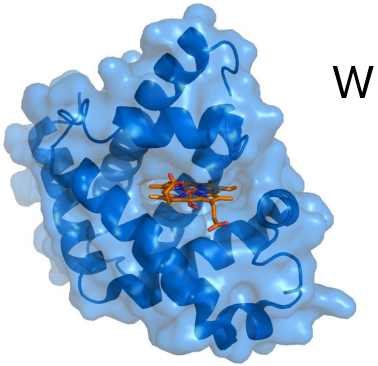
Objective



Which type ?

Be able to discriminate proteins in function of electrostatic interaction potential between the protein and its ligand

Objective

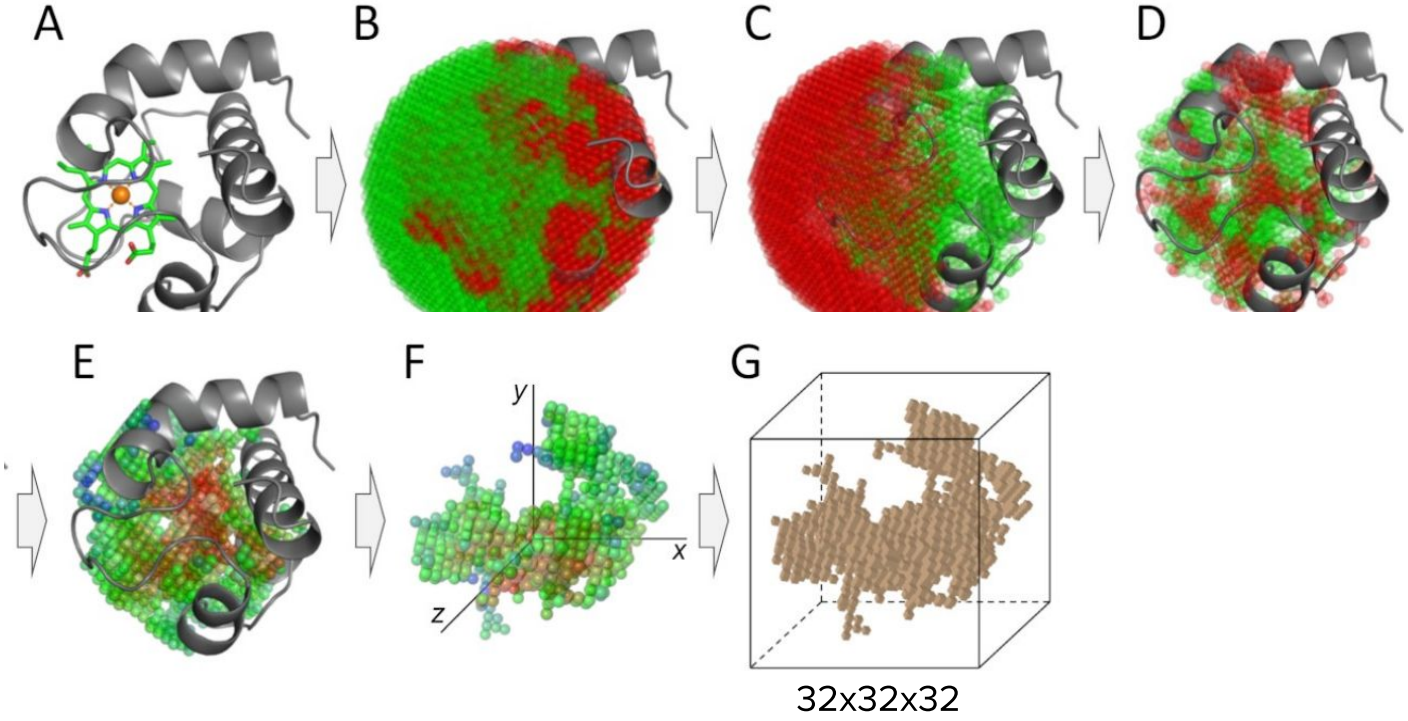


Which type ?

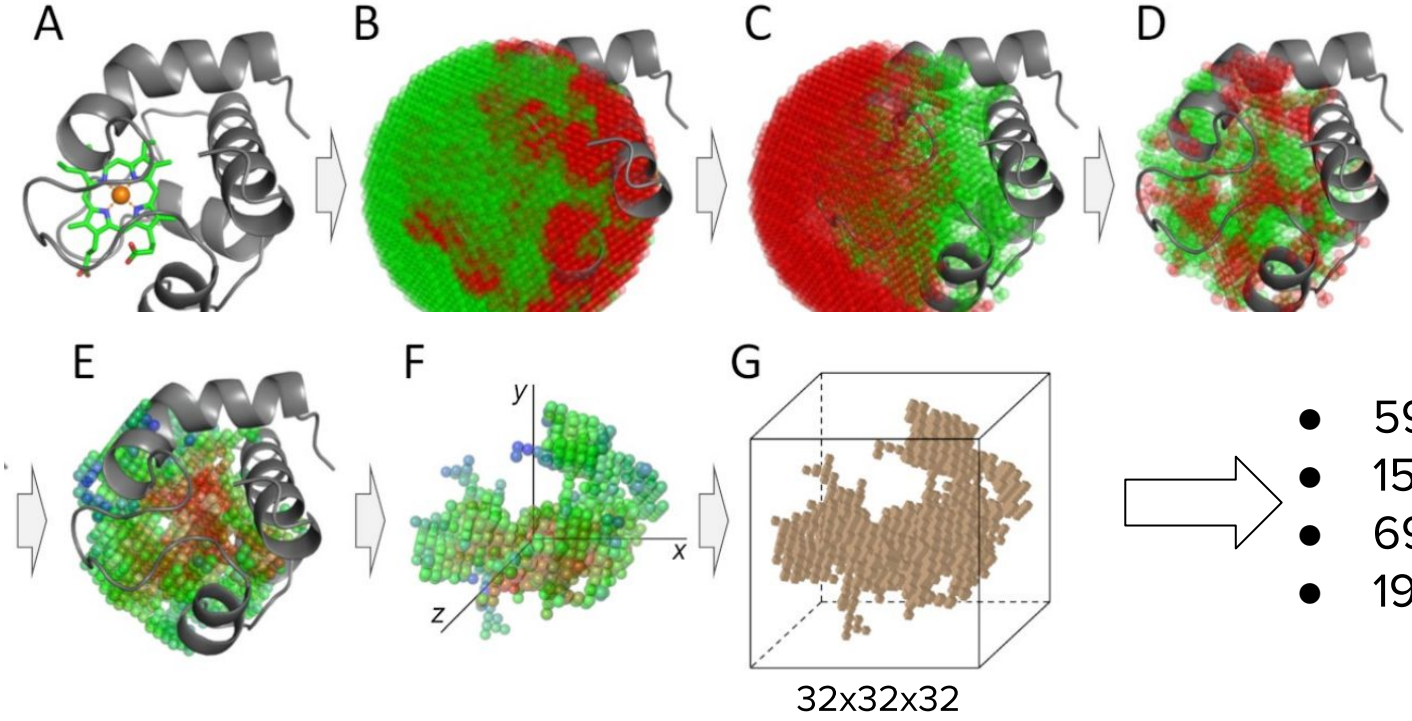
Be able to discriminate proteins in function of electrostatic interaction potential between the protein and its ligand

Neural Network inspired from Pu *et al.* (2018), **DeepDrug3D**: *Classification of ligand-binding pockets in proteins with a convolutional neural network.*

Materials and methods: Data generation



Materials and methods: Data generation



- 596 Hèmes
- 1553 Nucleotides
- 69 Steroïds
- 1946 Controls

Each voxels can have **14** different electrostatic interaction potentials

Materials and methods: Data train

Pocket Data (shuffled

Hemes x 300

Nucleotides x 300

Steroïds x 69

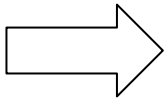
Controls x 300

Materials and methods: Data train

	Pocket Data (shuffled	Class Data (shuffled
Hemes	x 300	[1, 0, 0, 0]
Nucleotides	x 300	[0, 1, 0, 0]
Steroïds	x 69	[0, 0, 1, 0]
Controls	x 300	[0, 0, 0, 1]

Materials and methods: Data train

	Pocket Data (shuffled	Class Data (shuffled
Hemes	x 300	[1, 0, 0, 0]
Nucleotides	x 300	[0, 1, 0, 0]
Steroïds	x 69	[0, 0, 1, 0]
Controls	x 300	[0, 0, 0, 1]



Running the training of the model

Materials and methods: Neural Network

3 different models:

Model A:

1 x 3D convolution
with **16** neurons

Model B:

2 x 3D convolution
with **4** and **16** neurons

Model C:

2 x 3D convolution
with **16** and **16**
neurons

Materials and methods: Neural Network

3 different models:

Model A:

1 x 3D convolution
with **16** neurons

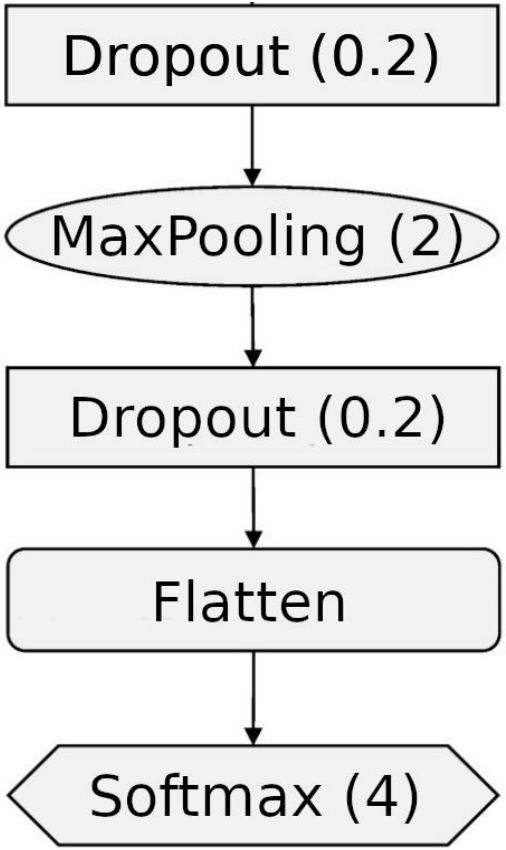
Model B:

2 x 3D convolution
with **4** and **16** neurons

Model C:

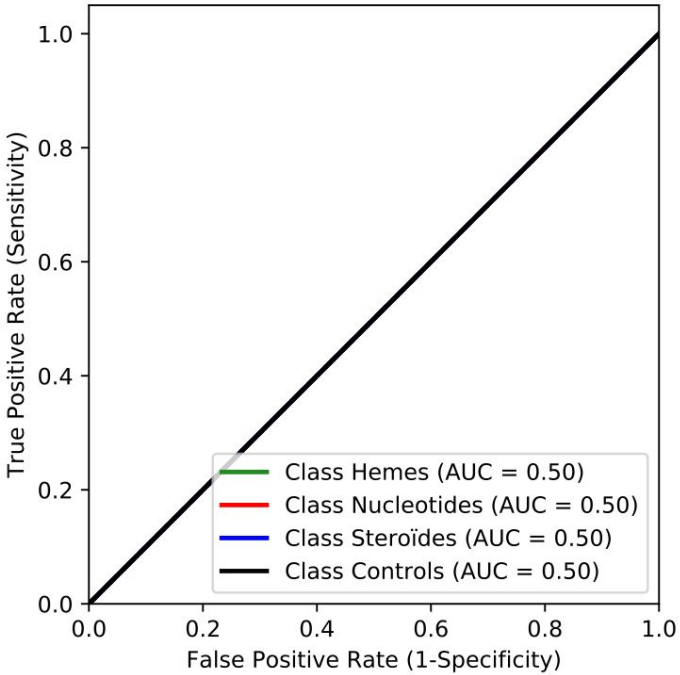
2 x 3D convolution
with **16** and **16**
neurons

+



Results

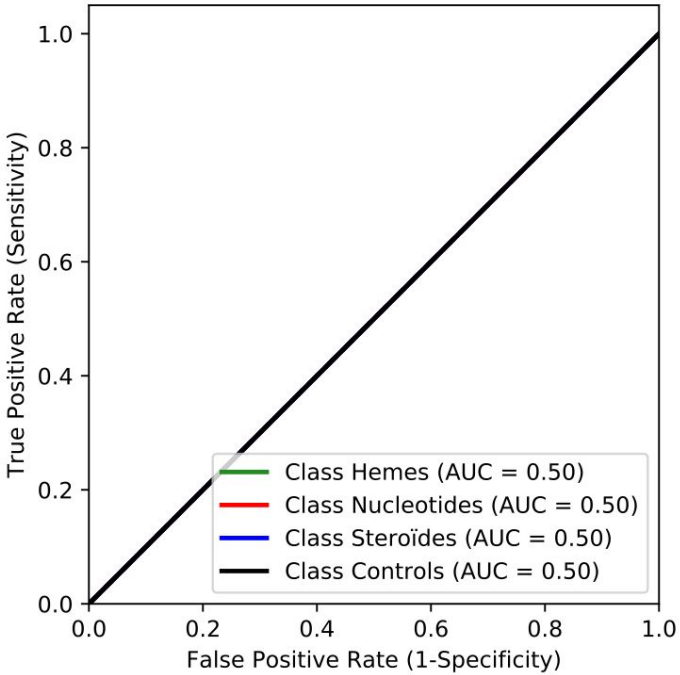
Model with one convolution layer 16 neurons



Global accuracy = 0.30

Results

Model with one convolution layer 16 neurons

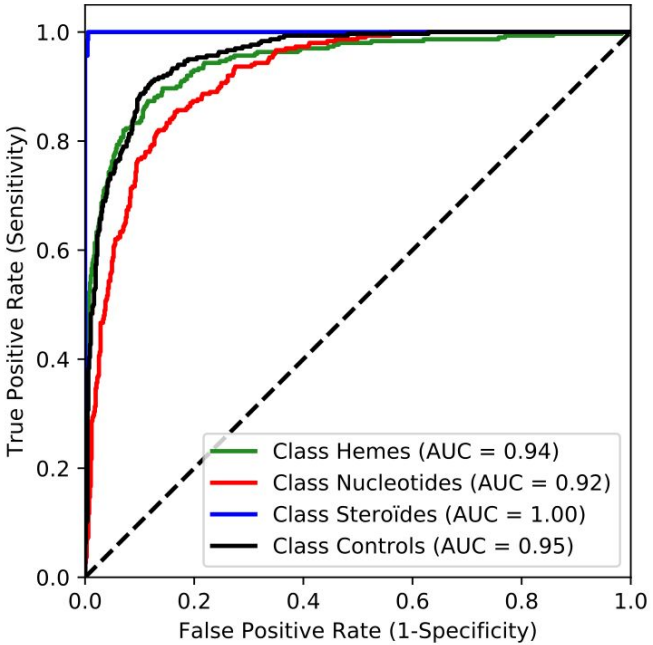


Global accuracy = 0.30

This model is not learning !

Results

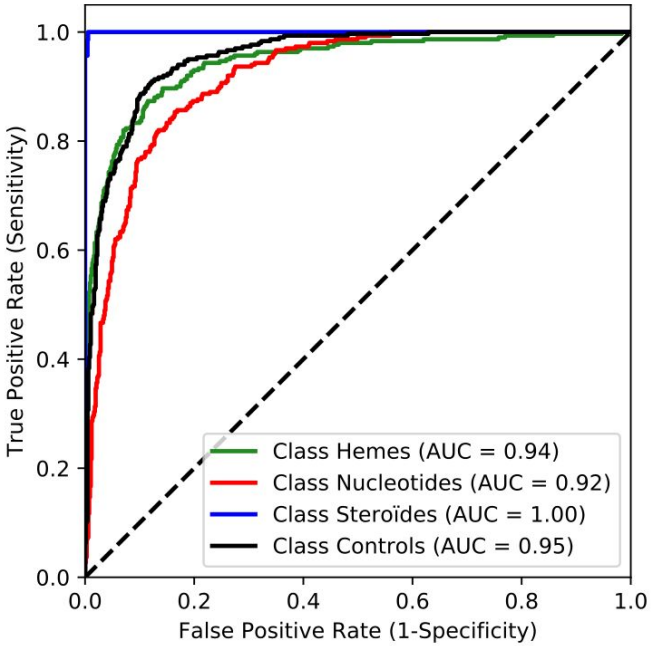
Model with double convolution layer of 4 and 16 neurons



Global accuracy = 0.80

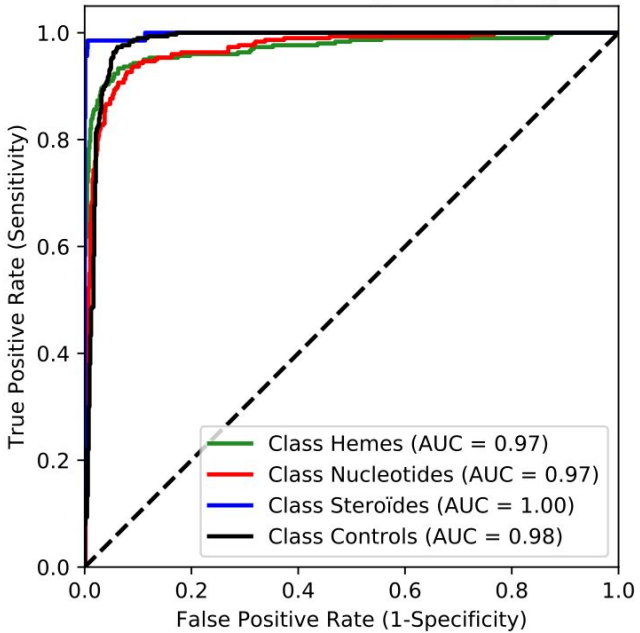
Results

Model with double convolution layer of 4 and 16 neurons



Global accuracy = 0.80

Model with double convolution layer of 16 and 16 neurons



Global accuracy = 0.90

Discussion

- Implemented model C pretty good
- Not enough data on steroïds (discrimination accuracy biased)
- Hemes and Nucleotides discrimination close from Pu and his team
- Necessity to implement double convolution layer and learning data set shuffling