Prediction of the binding affinity of protein-ligand complexes combining MD simulations with DL algorithms

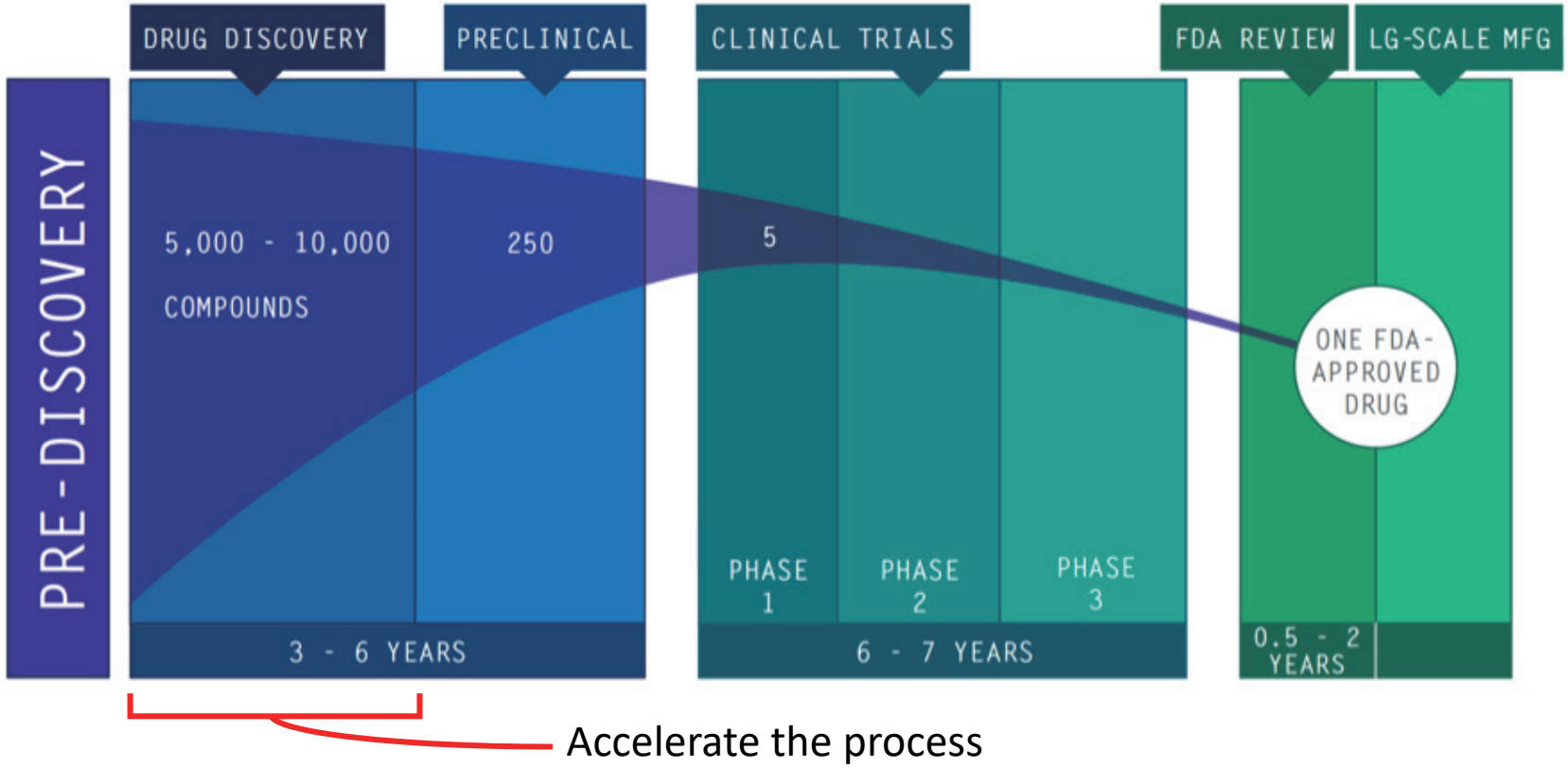
ICOA Supervisors: Pr. Pascal Bonnet / Dr. Samia Aci-Sèche

JANSSEN Supervisor: Dr. Gary Tresadern

JCAD – December 13th, 2021

Pierre-Yves Libouban

Introduction

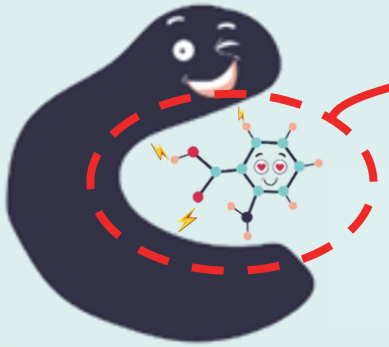


- The time to develop a drug last from 10 to 15 years and it cost of a drug is up to 1 billion €
- Goal: Accelerate the process and reduce the cost

Source: Drug Discovery and Development, Understanding de R&D Process" Pharmaceutical Research and Manufacturers of America, 2007, innovation.org

#MT180

#MT180

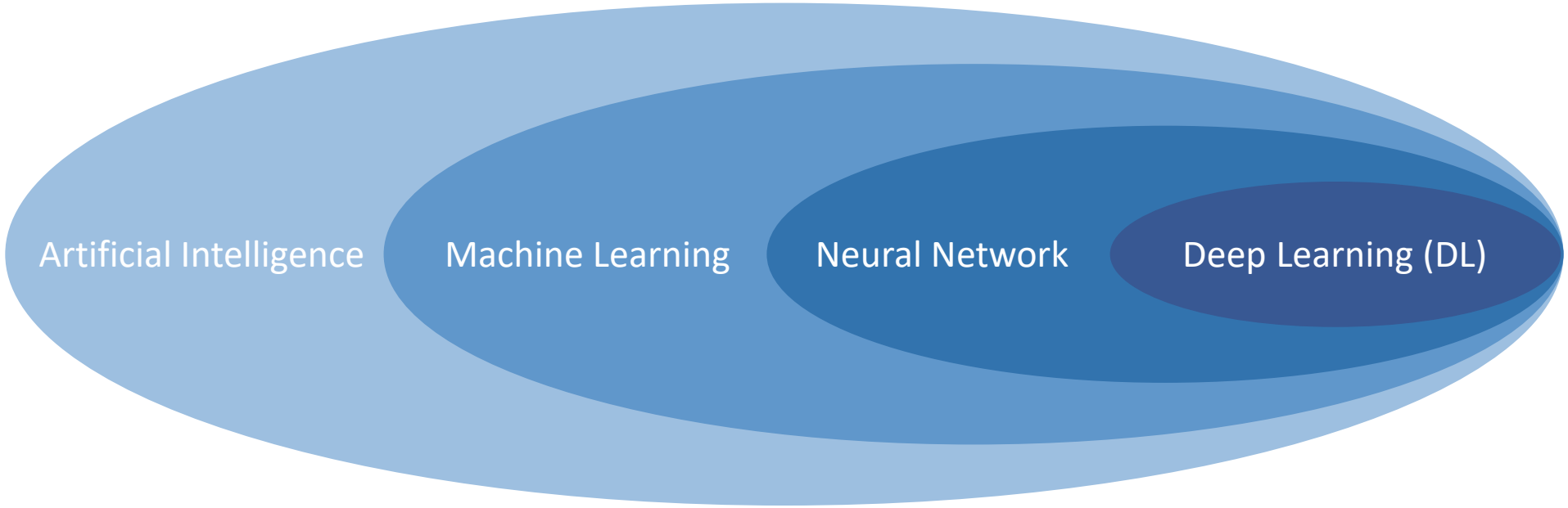


Lead compound

#MT180

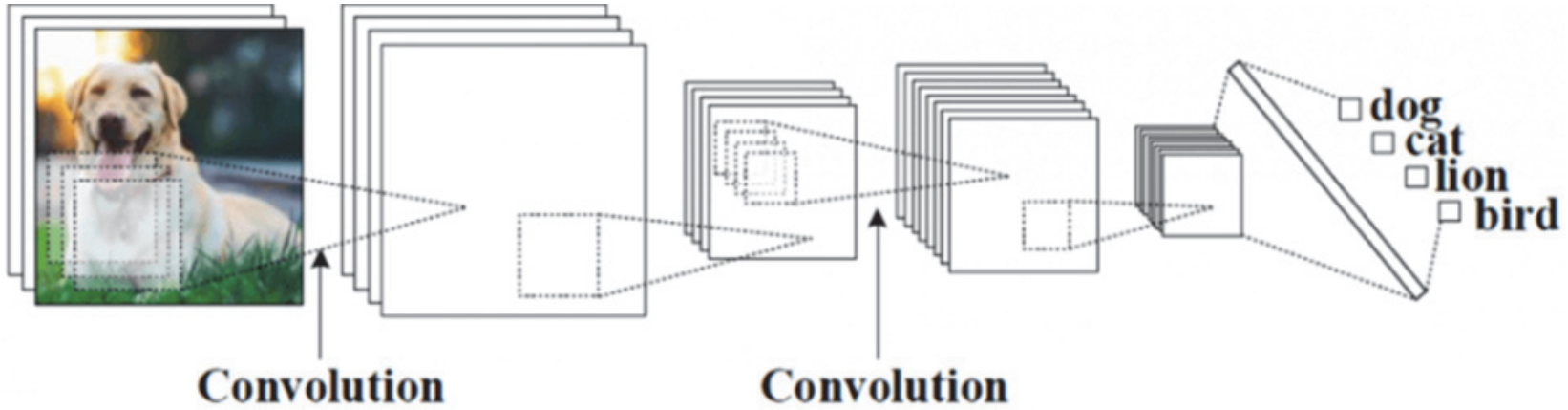
Methods: Deep learning

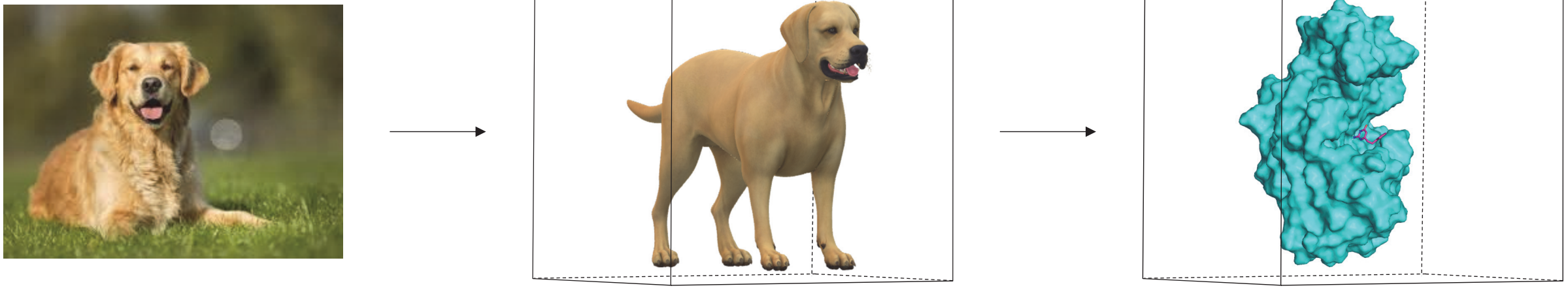
Artificial intelligence, machine learning and deep learning



Methods: Deep learning

Spatial data: Convolutional neural networks (**CNN**)

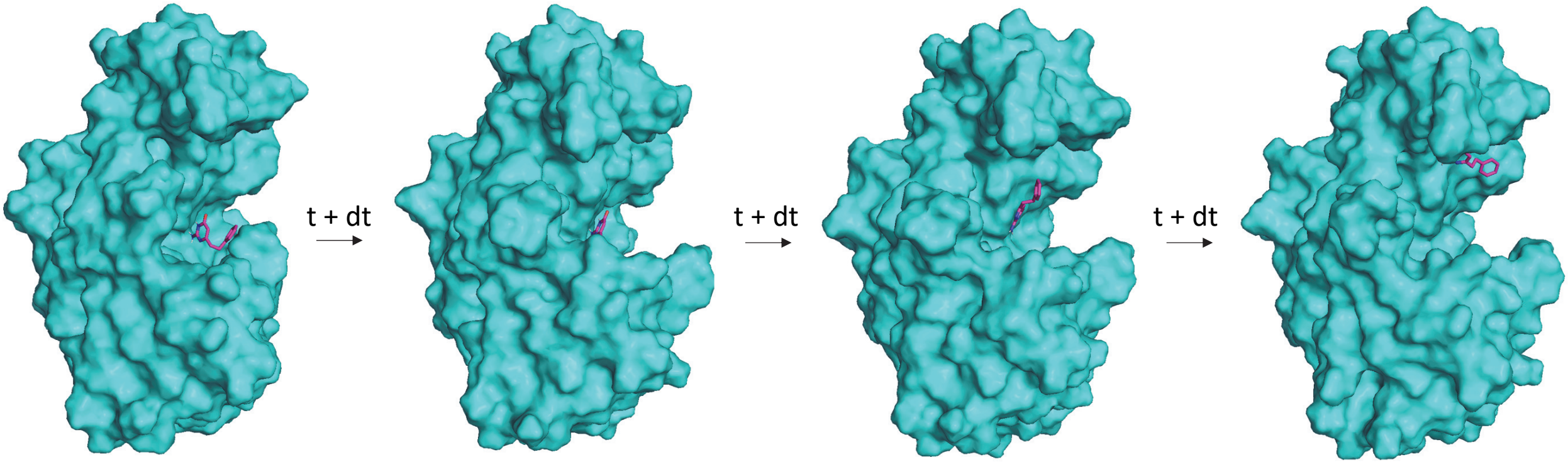




2D 3D 3D

Source: https://www.aspexit.com/en/neural-network-lets-try-to-demystify-all-this-a-little-bit-3-application-to-images/

Methods: Molecular dynamics simulations



Methods: Molecular dynamics simulations

Dataset: a subset of the PDBbind
~9 000 complexes

Simulation adapted to run on GPU.

Performing molecular dynamics on several clusters of GPU:

- Jean ZAY (IDRIS): V100
- Myria (CRIANN): P100
- Our laboratory cluster: RTX 2080

Carried out simulations for 4500 complexes.

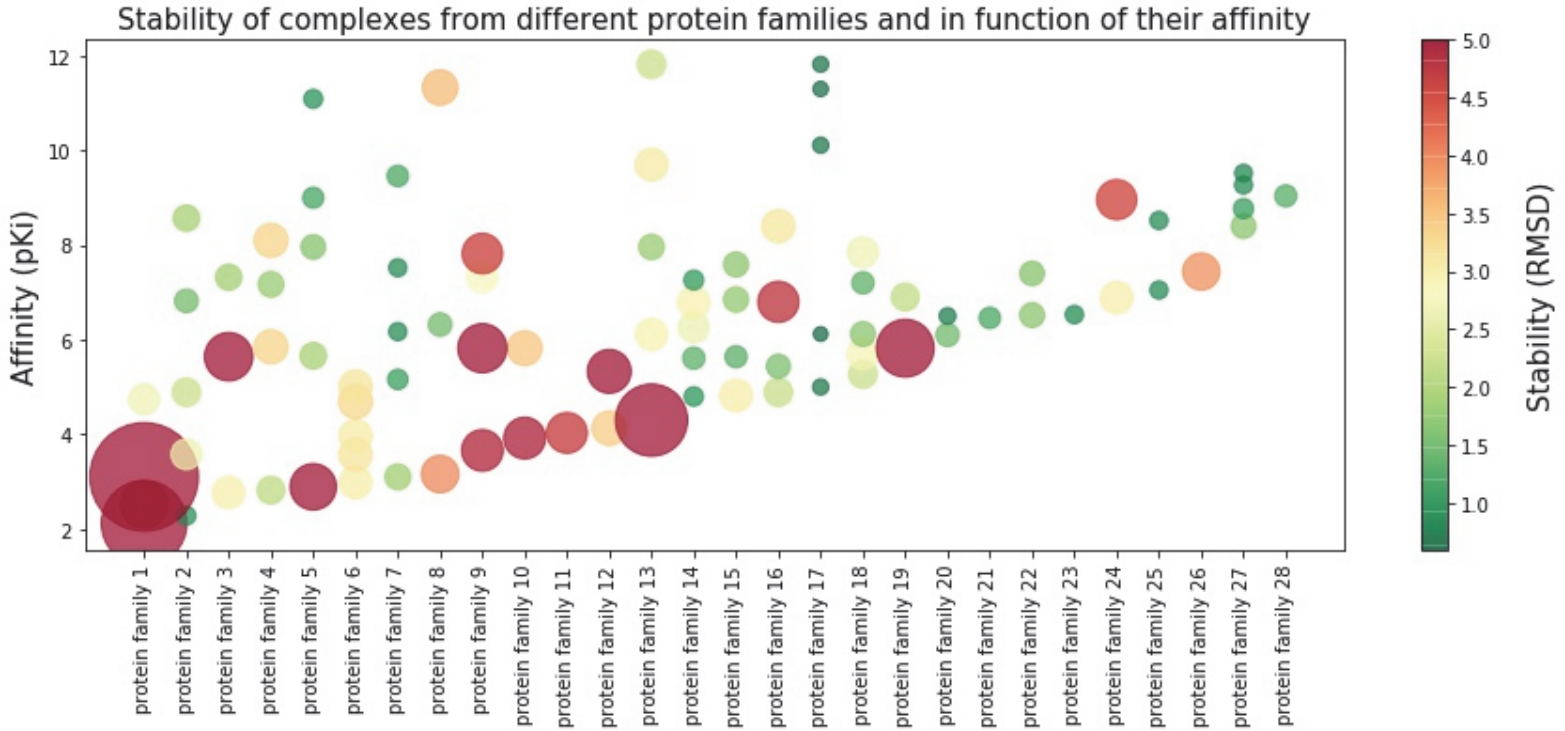
1 simulation can require up to a day of computation, depending on the GPU used.

Molecular dynamics protocol:

- Amber force field
- All atoms
- 10 simulations of 10 ns for each complexes

Methods: Molecular dynamics simulations

Evaluation of the stability of ligands during MD in function of their affinity

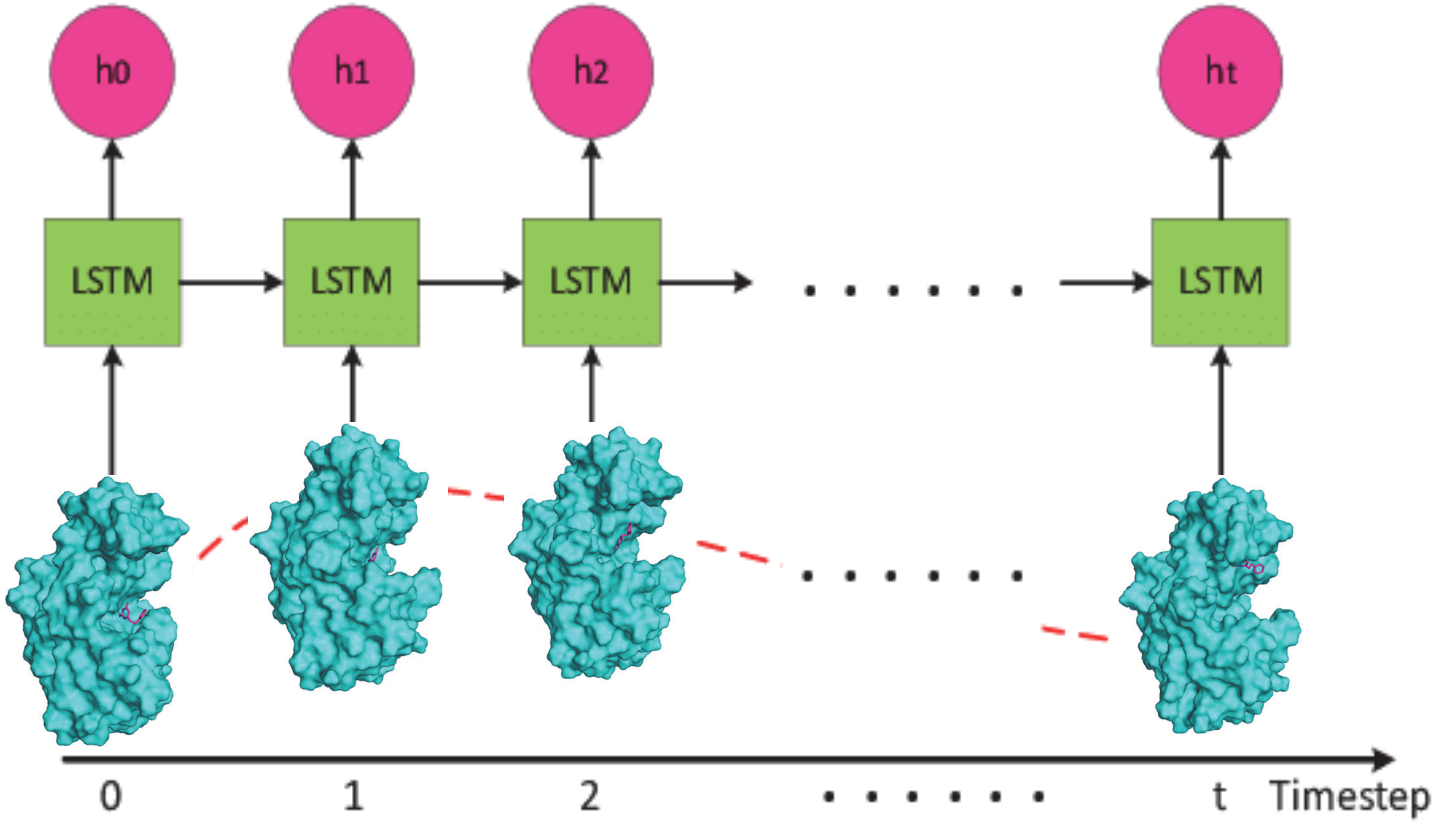


We display the stability of the ligand during the simulation. Each protein family contain several complexes with different affinities. The goal is to see if **complexes with lower affinity** (low pKi) within the same protein family, **also have a lower stability** (high RMSD) of the ligands.

Ligand RMSD: distance between the position of the ligand at the beginning of the simulation and the its position at a given time.

Methods: Molecular dynamics simulations + deep learning

Temporal data: Long short time memory (**LSTM**)



Zhou H, Zhang Y, Yang L, Liu Q, Yan K, Du Y. Short-Term Photovoltaic Power Forecasting Based on Long Short Term Memory Neural Network and Attention Mechanism. IEEE Access. 2019;PP:1-.

Conclusion and perspectives

- Conclusion
  - Hypothesis: MD simulations contain **temporal information** about **protein-ligand interaction** useful for **DL prediction**
  - Carried out MD simulations for **4500 complexes** to set up a dataset for DL algorithms to learn from

- Perspectives
  - Finish to carry out MD **simulations on the whole dataset**
  - **Combine CNN** and **LSTM** to be able to learn on **MD simulations** and create models from it

Acknowledgements

- SB&C team
  - Pascal Bonnet
  - Samia Aci-Sèche
  - Stéphane Bourg
  - Gautier Peyrat
  - Pascal Krezel
  - Adnane Aouidate
  - Lydia Benkaidali
  - Xiaojun Mao
- Janssen

Pierre-Yves.Libouban@univ-orleans.fr

Thank you for your attention !