

Avis de Soutenance

Monsieur Pierre-Yves LIBOUBAN

Chimie

Soutiendra publiquement ses travaux de thèse intitulés

Prédiction de l'affinité de liaison des complexes protéine-ligand en combinant des simulations de dynamique moléculaire avec des algorithmes d'apprentissage profond

dirigés par Monsieur Pascal BONNET

Ecole doctorale : Santé, Sciences Biologiques et Chimie du Vivant - SSBCV

Unité de recherche : ICOA - Institut de Chimie Organique et Analytique

Soutenance prévue le **lundi 11 décembre 2023** à 14h00

Lieu : 3 avenue de la recherche scientifique, Délégation régionale (DR8) Campus CNRS, 45071 Orléans

Salle : amphithéâtre Charles Sadron

Composition du jury proposé

M. Pascal BONNET	Université d'Orléans	Directeur de thèse
M. Olivier SPERANDIO	Institut Pasteur	Rapporteur
M. Matthieu MONTES	Conservatoire national des arts et métiers (CNAM)	Rapporteur
Mme Jana SOPKOVA	Université de Caen	Examinatrice
M. Xavier BARRIL	Université de Barcelone	Examinateur
M. Gary TRESADERN	Janssen	Co-encadrant de thèse
Mme Samia ACI-SECHE	Université d'Orléans	Co-encadrante de thèse

Mots-clés : apprentissage profond, dynamique moléculaire, affinité de liaison, intelligence artificielle, complexe protéine-ligand, conception de médicaments

Résumé :

Les interactions des petites molécules avec leurs protéines cibles sont essentielles à la recherche pharmaceutique. L'affinité de liaison des complexes protéine-ligand peut être mesurée par des expériences in-vitro, mais ces tests sont coûteux en argent et en temps. Aujourd'hui, les réseaux de neurones profonds utilisant les structures en trois dimensions des complexes sont capables de prédire cette affinité de liaison. Cependant, des limitations persistent malgré l'implémentation de nouveaux réseaux de neurones. Ceci est principalement dû au manque de données structurales, qui nécessitent un travail expérimental conséquent pour être déterminées. Ce projet vise à améliorer notre capacité à prédire l'affinité des complexes protéine-ligand en combinant des approches d'apprentissage profond avec des simulations de dynamique moléculaire. Ainsi, il est possible d'augmenter la quantité de données utilisées lors de l'apprentissage des modèles statistiques, en extrayant des structures supplémentaires des simulations de dynamique moléculaire. En outre, celles-ci fournissent des informations temporelles sur les interactions protéine-ligand qui peuvent être utilisées pour améliorer les modèles. Nous avons créé un ensemble de données de 63 000 simulations, obtenues à partir de 6 300 complexes. Puis nous avons développé des réseaux de neurones, tel que le LSTM à convolutions, capables d'analyser à la fois les informations spatiales et temporelles issues des simulations. Ces réseaux combinent un réseau de neurones à convolutions capable d'extraire l'information spatiale des structures en trois dimensions à chaque pas de temps, tandis que le LSTM suit l'évolution de cette information sur l'ensemble de la simulation. En utilisant les simulations de dynamique moléculaire en tant qu'augmentation de données, nos modèles obtiennent des résultats prometteurs.