

## Construction de Modèles d'États Markoviens pour l'analyse de simulations de dynamique moléculaire

Responsables : Florian Blanc, Olivier Walker\*  
Équipe Biophysique des Systèmes Complexes, ISA

e-mail: [olivier.walker@univ-lyon1.fr](mailto:olivier.walker@univ-lyon1.fr); [fblanc@biophys.mpg.de](mailto:fblanc@biophys.mpg.de)

**Mots-clés** : Python, Jupyter, machine learning, dynamique moléculaire, bio-informatique structurale, apprentissage profond

Les simulations de dynamique moléculaire tout-atome sont un outil puissant pour révéler les changements conformationnels des protéines, et donc comprendre leur fonction. Cependant, l'interprétation des trajectoires de simulations est difficile du fait de leur complexité structurale et dynamique. Une méthode interprétative féconde est la construction de Modèles d'États de Markov (Markov States Models, MSM). Cette approche repose sur l'identification, à partir de trajectoires de simulations, d'états conformationnels dominants à l'aide d'apprentissage machine non-supervisé (réduction de dimensionnalité et *clustering*) ; puis la construction d'une matrice de transition reflétant les probabilités de transition entre ces états. Une fois obtenue, cette matrice encapsule toute l'information thermodynamiques et cinétiques sur le système, et peut être construite efficacement à l'aide de modules Python comme PyEMMA. Le but du projet est de mettre en place, sous la forme d'un notebook Jupyter bien documenté, un *pipeline* de construction d'un modèle d'états de Markov à partir de trajectoires de dynamique moléculaire de protéines obtenues au laboratoire. Les étudiants utiliseront pour cela le module PyEMMA (url) et pourront notamment s'appuyer sur sa documentation. En fonction des progrès, le projet pourra être poursuivi avec la mise en place d'un *pipeline* pour le module *deeptime*, qui implémente des approches récentes pour construire le modèle markovien à l'aide d'apprentissage profond.

**Profil recherché** : connaissance de l'écosystème Python scientifique (Jupyter, Numpy, Scipy, pandas, scikit-learn) ; familiarité avec les protéines (biologie/bioinformatique structurale) ; des connaissances préalables en dynamique moléculaire et/ou apprentissage profond seraient un plus.