

Laboratoire: Laboratoire de Physique des Solides

Responsable du stage: Guillaume TRESSET

E-mail: guillaume.treset@cnrs.fr

Téléphone: +33 (0)1 69 15 53 60

Site web: <https://equipes2.lps.u-psud.fr/sobio/>

Adresse: Université Paris-Saclay, rue Nicolas Appert, Bâtiment 510, Orsay

Possibilité de poursuivre en thèse: oui (concours école doctorale PIF)



Décrypter l'assemblage du virus de l'hépatite B par l'apprentissage automatique

L'hépatite B est une maladie qui frappe chroniquement près de 300 millions de personnes à travers le monde et tue environ 1 million d'entre elles annuellement. Elle est causée par le virus de l'hépatite B (VHB), une nanocapsule de 35 nm de diamètre faite de protéines et de lipides protégeant le code génétique viral. Le virus se multiplie dans les cellules infectées via un processus d'auto-assemblage qui permet aux composants moléculaires de former de nouvelles particules virales infectieuses. Des molécules antivirales perturbant cette étape d'auto-assemblage sont en cours de développement dans les laboratoires, et certaines sont même en essai clinique. Néanmoins, un mystère persiste : comment ces molécules perturbent-elles la formation du virus ? Quelle est leur influence sur sa structure et sa cinétique d'assemblage ?

Au carrefour de la physique, de la virologie et de la chimie, notre consortium international (Institut de Biologie Intégrative de la Cellule@Gif-sur-Yvette, Université Emory, Université de Californie Riverside), est en mesure de reconstituer in vitro une partie du virus en présence d'antiviraux. Grâce à la diffusion des rayons X à haute résolution (synchrotron SOLEIL), nous capturons des données massives dans l'espace réciproque, clés pour percer les mécanismes d'assemblage du VHB à l'échelle nanométrique.

L'objet du stage sera de consolider un modèle d'apprentissage automatique et capable de décoder les paramètres structuraux de particules virales à partir de leurs signatures de diffusion. Pour y parvenir :

- Vous construirez une banque de données de particules en cours d'assemblage à l'aide d'un code Monte Carlo.
- Vous entraînerez un algorithme pour reconstruire des morphologies nanométriques à partir d'un modèle d'assemblage de type Fokker-Planck. Il s'agira soit d'une régression avec régularisation, soit d'un réseau de neurones récurrents selon les performances.
- Vous le testerez sur des séries de données récemment collectées au synchrotron et vous interpréterez les mécanismes sous-jacents régulés par des antiviraux et/ou des agents d'encombrement.

Le ou la stagiaire devra avoir une appétence pour la modélisation physique et l'analyse de données complexes. La maîtrise de Python est nécessaire mais l'apprentissage des bibliothèques spécialisées se fera éventuellement au fil de l'eau. Enfin, une curiosité pour les applications biomédicales de la physique et de l'IA est indispensable.

