

**Sujet de stage M2 – Bioinformatique structurale  
année scolaire 2018-19**

Contact et encadrant: Thomas Simonson (directeur de recherches CNRS et professeur Ecole Polytechnique)

Coordonnées: Laboratoire de Biochimie (UMR CNRS 7654), Ecole Polytechnique, Palaiseau;  
[thomas.simonson@polytechnique.fr](mailto:thomas.simonson@polytechnique.fr), 0169334860

Titre du stage: **Le dessin computationnel de protéines: mesures de performance et nouveaux développements**

Mots-clés: Monte Carlo, ingénierie des protéines, logiciel Proteus, programmation perl/python, analyse de séquences et structures moléculaires

Description du sujet: **voir page suivante**

Date de début et durée estimée: début entre janvier et mars; durée 5 ou 6 mois

Rémunération académique règlementaire, soit 554 Euros bruts/mois

Date limite de candidature: fin janvier 2019 (mais clôture avant si un candidat est choisi)

Compétences essentielles recherchées: notions de biologie structurale et/ou bioinformatique structurale et/ou de biophysique moléculaire et/ou de chimie physique.

Plus généralement, le stage peut convenir à une personne motivée avec une formation en bioinformatique, biochimie, biophysique, biologie moléculaire, chimie moléculaire, chimie théorique, ou équivalent, le sujet pouvant être infléchi en fonction du profil.

## **Le dessin computationnel de protéines: mesures de performance et nouveaux développements**

Equipe de bioinformatique structurale, Laboratoire de Biochimie, Ecole Polytechnique

Contact: Thomas Simonson, [thomas.simonson@polytechnique.fr](mailto:thomas.simonson@polytechnique.fr), 0169334860

Notre équipe développe et applique des méthodes de dessin computationnel, ainsi qu'un des rares logiciels européens du domaine, le logiciel Proteus. *Le travail de stage proposé* consistera à déterminer les performances de notre dernier modèle sur plusieurs famille de protéines. Proteus permet des simulations Monte Carlo dans l'espace des séquences et conformations, avec un squelette fixe ou mobile, en présence ou non d'un ou plusieurs ligands, en permettant ou non les changements d'état redox ou acid/base de groupes particuliers, et en exploitant plusieurs méthodes d'échantillonnage dont le "Replica Exchange Monte Carlo", la métadynamique, et un panel de méthodes de "umbrella sampling". Il utilise un champ de force récent de mécanique moléculaire (Amber) et un solvant implicite récent de type Born Généralisé. Ces éléments lui ont permis de produire, pour plusieurs systèmes, des séquences de qualité comparable au meilleur logiciel américain, Rosetta. Il obtient d'excellentes performances pour le placement de chaînes latérales (comparables à Scwrl4) et pour les calculs acide/base (supérieures à Rosetta). Il est appliqué à l'ingénierie de complexes protéine-ligand, en collaboration avec plusieurs groupes expérimentaux et théoriques. Nous étudions notamment les aminoacyl-ARNt synthetases (avec une perspective d'ingénierie du code génétique) et les domaines PDZ (avec une perspective d'inhibition d'interactions protéine-protéine). Ainsi, nous avons obtenu une tyrosyl-ARNt synthétase avec une stéréospécificité inversée, préférant la D-tyrosine comme substrat.

Nous travaillons activement pour améliorer notre modèle de solvant, et notamment le traitement des interactions non-polaires ou hydrophobes. Le travail de stage pourra consister à étudier les performances en dessin computationnel d'un ou plusieurs variants de notre tout dernier modèle. En effet, celui-ci a été paramétré et testé jusque-là pour des problèmes plus simples: prédictions de structure et de changements de stabilité suite à des mutations ponctuelles. Nous prévoyons un gain de performance et de vitesse grâce à ce nouveau modèle hydrophobe. Les techniques employées comprendront ainsi: la simulation Monte Carlo, la programmation perl et/ou python et shell, l'analyse de séquences et de structures, la simulation de dynamique moléculaire (pour tester certaines séquences), la modélisation implicite du solvant.

Nous sommes une équipe de modélisation moléculaire expérimentée (3 chercheurs permanents; 8 thèses et 60 articles depuis 10 ans), appartenant au Laboratoire de Biochimie de l'Ecole Polytechnique. Nous sommes situés dans le centre de recherche interdisciplinaire de l'Ecole et, au-delà, de l'Université Paris-Saclay. La poursuite en thèse est envisageable (à travers l'Ecole Doctorale de l'Ecole Polytechnique). Quelques articles récents sont référencés ci-dessous; cf aussi [biology.polytechnique.fr/biocomputing/biblio.html](http://biology.polytechnique.fr/biocomputing/biblio.html).

### **Références choisies (Noms en gras: stagiaires M2 au moment du travail)**

- Polydorides, Michel, Druart, Mignon, Archontis, Simonson (2016) In "*Methods in Molecular Biology: Design and Creation of Protein Ligand Binding Proteins*," Springer Verlag, New York. Proteus and the design of ligand binding sites.
- Simonson, Ye-Lehmann, Palmal, Amara, Wydau-Demattis, Bigan, **Druart**, Moch, Plateau (2016) *Proteins* 84:240-53. Redesigning the stereospecificity of tyrosyl-tRNA synthetase.
- Mignon, Panel, **Chen**, Fuentes, Simonson (2017) *J Chem Theory Comp* 13:2271-89. doi.org/10.1038/s41598-017-16221-8 Computational design of the Tiam1 PDZ domain and its ligand binding.
- **Opuu**, **Silvert**, Simonson (2017) *Scientific Reports* 7, article 15873. doi.org/10.1038/s41598-017-16221-8 Computational design of fully overlapping coding schemes for protein pairs and triplets.
- Villa, Panel, **Chen**, Simonson (2018) *J Chemical Physics*, 149, article 072302. doi.org/10.1063/1.5022249 Adaptive landscape flattening in amino acid sequence space for the computational design of protein:peptide binding.