

# Optimisation du positionnement de chaînes latérales de protéines sous contraintes d'observations Cryo-EM.



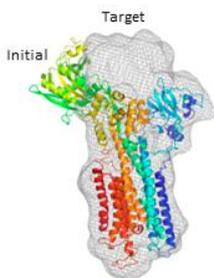
**Lieu:** L'Unité de Mathématiques et Informatique Appliquées de Toulouse (centre INRA Castanet Tolosan) <http://mia.toulouse.inra.fr>

**Contact :** [david.allouche@inra.fr](mailto:david.allouche@inra.fr) / (☎ 05 61 28 52 77)

Le stage sera réalisé en collaboration avec l'équipe du F. Tama et O. Miyashita (U. Nagoya et RIKEN). Sur la base de méthodes initiées dans les équipes françaises et japonaises.

**Durée :** 3 à 6 mois selon convention

**Keywords:** modèle graphique, optimisation combinatoire, branch and bound, Combinatorial computational geometry, mécanique moléculaire.



**Contexte :** Mise au point récemment, la Cryo-microscopie électronique moléculaire est l'une des plus récentes méthodes de détermination de structure moléculaire. Cette méthode de spectroscopie permet de voir des objets moléculaires uniques ("Single molecule Approach"). Couronnée d'un prix Nobel en 2017<sup>1</sup>, elle a permis, notamment, pour la première fois de construire des modèles 3D de virus ou de gros assemblage moléculaire à l'échelle de l'ordre du micron.

La construction du modèle 3D fait appel à une approche de reconstruction hybride faisant intervenir, le traitement d'image pour définir une enveloppe 3D (grille). Puis dans un Second temps, une étape de modélisation "moléculaire" vise à positionner le squelette de la molécule ou du complexe de molécules dans l'enveloppe 3D issue de l'observation de microscopie.

**Objectif :** Le but du stage est d'accroître la résolution du modèle 3D via la prise en compte des chaînes latérales du squelette protéique.

Dans un premier temps, il conviendra d'incorporer les données Cryo-EM (-i.e. l'enveloppe 3D-) dans le modèle et le cas échéant de modifier les algorithmes d'optimisation pour en améliorer la résolution. Du point de vue mathématique, il s'agira :

D'abord de modifier le calcul d'une matrice d'énergie de complexité ( $O(N^2)$ ) où  $N$  est le nombre d'acides aminés du système. Puis déterminer, la solution de probabilité maximum du modèle graphique probabiliste associé via sa formulation en problème d'optimisation combinatoire. Le problème est NP-difficile, mais peut en pratique, être modélisé via un réseau de fonction de coût, résolu par méthode de recherche branch and bound.

Après implémentation du modèle, l'algorithme de recherche pourra, selon les résultats et l'avancement, être adapté pour en accélérer la résolution dans le contexte des données Cryo-EM.

## Bibliographie :

### ➤ **Optimisation combinatoire:**

1. Unified Variable neighborhood search with cost function network (UAI Sydney 2017) :<http://auai.org/uai2017/proceedings/papers/197.pdf>
2. Variable Neighborhood Search with Cost Function Networks To Solve Large Computational Protein (JCIM 2018)
3. Anytime Hybrid Best-First Search with Tree Decomposition for Weighted CSP : David Allouche &al cp 2015

### ➤ **Modèle bio physique: / computational geometry :**

4. Flexible Multi-scale Fitting of Atomic Structures into Low-resolution Electron Density Maps with Elastic Network Normal Mode Analysis J. Mol. Biol.(2004) 337, 985–999
5. Hierarchical mesh decomposition using fuzzy clustering and cuts [S Katz, A Tal](#) - ACM Transactions on Graphics (TOG), 2003 - dl.acm.org

## Connaissances/notions scientifiques utiles:

- Théorie des graphes
- Optimisation
- Computationnel géométrie
- Réseau gaussien
- Mécanique classique.

**Profil** : Le candidat doit avoir un goût prononcé pour la modélisation et le développement de méthode dans un contexte interdisciplinaire. Une maîtrise du langage C++ est indispensable, ainsi que des facilités de communication en anglais à l'écrit et à oral.



Le stage donne droit à une indemnité forfaitaire d'environ 560 €/mois avec accès à restauration sur place au tarif étudiant.