

Stage de Master 2 en bioinformatique : Analyse des données métabolomiques pour la détection de contamination

### 1) Contexte du projet

La métabolomique par spectrométrie de masse est une technologie récente qui permet de décrire de manière exhaustive et non ciblée le contenu d'un échantillon en biomolécules de faible poids moléculaire (e.g. médicaments, hormones, pesticides). Elle est utilisée à la fois dans le domaine biomédical (recherche de biomarqueurs) et dans l'environnement (détection de contaminants). Le processus bioinformatique depuis le signal brut en sortie d'appareil jusqu'à l'identification des métabolites représente un défi (Katajamaa and Oresic, 2007; Trygg *et al.*, 2007) en raison de la taille (de l'ordre du Go) et de la complexité des données (grandes dimensions). Notre laboratoire de recherche technologique, au sein du CEA à Saclay, est spécialisé dans le traitement du signal spectral (correction de la ligne de base, extraction des pics) et l'analyse des données métabolomiques (régression PLS, classification, modèles à effets mixtes), en étroite collaboration avec les partenaires expérimentaux du laboratoire d'Etude du Métabolisme des Médicaments (Lenuzza *et al.*, 2011 ; Werner *et al.*, 2008).

### 2) Sujet proposé

Le stage de Master 2 consistera à développer les outils bioinformatiques pour contrôler la qualité d'échantillons et détecter la présence de contaminants (connus ou non). Ces derniers se trouvant à l'état de trace, le stage nécessitera la mise en œuvre de méthodes fines de séparation signal/bruit, puis de détection robuste d'outliers (Charlton *et al.*, 2008), et enfin d'aide à la caractérisation automatique des métabolites inconnus. Les logiciels utilisés seront R et C/C++.

### 3) Profil recherché

Nous recherchons un candidat avec de solides compétences en mathématiques, statistiques et informatique appliquées à la biologie, ainsi qu'une forte motivation pour le travail interdisciplinaire académique et industriel. Le stage pourra déboucher sur une thèse.

### 4) Candidature

Merci d'envoyer un mail décrivant vos motivations et incluant votre CV à :

Etienne Thévenot  
Laboratoire d'Outils pour l'Analyse de Données  
CEA, Centre de Saclay, DRT/LIST/DCSI, 91191 Gif-sur-Yvette  
Courriel : [etienne.thevenot@cea.fr](mailto:etienne.thevenot@cea.fr)  
Internet : <http://www-centre-saclay.cea.fr>

### 5) Références

Charlton *et al.* (2008). Non-targeted detection of chemical contamination in carbonated soft drinks using NMR spectroscopy, variable selection and chemometrics. *Analytica Chimica Acta*, **618**:196-203.

Katajamaa and Oresic (2007). Data processing for mass spectrometry-based metabolomics. *Journal of Chromatography A*, **1158**:318-328.

Lenuzza *et al.* (2011). Development of a PK library of parent drug/metabolite integrated models (2011). *Conference of Population Approach Group in Europe*, Athens.

Trygg *et al.* (2007). Chemometrics in Metabonomics. *Journal of Proteome Research*, **6**:469-479.

Werner *et al.* (2008). Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends. *Journal of Chromatography B*, **871**:143-163.