

**Domaine de recherche :**

Bio-informatique, simulation bio-moléculaire / Sciences du vivant  
Mathématiques - Analyse numérique - Simulation / Sciences pour  
l'ingénieur

**Intitulé du sujet :**

Prédiction "in silico" de la biodégradabilité de composés  
xénobiotiques dans des environnements spécifiques

**Résumé du sujet :**

Because of increasing environmental concerns, it becomes important to be able to predict whether a particular compound is biodegradable, and if alternate routes can be engineered for compounds already known to be biodegradable.

Biodegradation pathways rely on microbial organisms, the vast majority of which are presently unculturable. However, metagenomics, empowered by recent technological developments in high-throughput sequencing, provides an alternative route to access the uncultivable genomes, hence the biochemical potential of the corresponding unculturable organisms.

On the other hand, the use of computational tools to predict the biodegradation of xenobiotics can aid in identifying the reactions needed to degrade these compounds, thus providing insight into the fate of xenobiotic compounds in the environment.

The aim of this thesis is to make use of large scale metagenomic sequence data derived from several waste water treatment plant bioreactors to predict the biodegradability, or in a more restricted sense the biotransformability, of undesirable compounds. This will be achieved by combining the computational identification of enzymes in the sequence data with chemical knowledge bases describing metabolite biotransformation rules and the reactivity of chemical functional groups.

Most of current biodegradation prediction methods are rule based and rely on extensive databases, e.g. the University of Minnesota Biocatalysis and Biodegradation Database (UM-BDD), describing the transformation of chemical functional groups. One such method makes use of the Kegg RPAIR database containing structure transformation patterns (so-called RDM patterns) to predict bacterial biodegradation pathways. This method uses a large database of RDM Patterns and limits the possible compounds produced from the query compound based on substrate and reaction specificity.

Thesis work will also be inspired by related and more recent approaches, like the BNICE method based on a framework developed to generate every possible biochemical reaction based on a set of enzyme reaction rules (e.g. based on the enzyme commission (EC) classification system) and starting compounds.

Besides metagenomic sequences and chemical knowledge bases, further experimental data consisting in meta-metabolic measurements (i.e. metabolite fingerprints) will be generated during the time frame of the thesis, and will be available both for the analysis of biodegradation networks and for the verification of computational predictions.

---

**Formation recommandée :** M2 chemoinformatique ou bioinformatique

---

**Informations pratiques :** Institut de Génomique  
Genoscope - Centre national de séquençage  
Genoscope - Centre national de séquençage  
Date souhaitée pour le début de la thèse : 01/10/2010  
Centre : Fontenay-aux-roses

---

**Personne à contacter :** Thomas Brüls  
CEA / DSV/IG/Genoscope/LABIS  
Genoscope / Institut de Génomique du CEA  
  
2 rue Gaston Crémieux  
  
F-91000 Evry France

Courriel : [thomas.bruls@cea.fr](mailto:thomas.bruls@cea.fr)  
Téléphone : 01 60 87 25 00

---

**En savoir plus :** <http://www.genoscope.cns.fr>

---

**Université / Ecole Doctorale :** Evry Val d'Essonne  
Des Génomes aux Organismes - Evry -

---

**Directeur de Thèse :** Jean Weissenbach  
CEA / DSV/IG//Genoscope  
IG/GENOSCOPE  
  
2, rue Gaston Cremieux  
  
91000 Evry

---