

## **Des approches in silico**

Le très grand nombre de données stockées dans les bases de données informatisées représente une source d'information précieuse pour mieux comprendre la régulation du métabolisme hépatique des lipides. Des approches informatiques ont donc été réalisées et comprennent trois étapes :

- . la recherche de séquences régulatrices,
- . l'extraction de connaissances de bases bibliographiques,
- . la modélisation de réseau de régulation.

### **Recherche de séquences régulatrices**

Les séquences régulatrices de gènes dont les facteurs contrôlent plusieurs gènes du métabolisme hépatique des lipides pourraient présenter des structures analogues dans certaines régions. Pour rechercher de telles structures, une collaboration avec l'équipe Symbiose de l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA) à Rennes a été engagée au sein du programme Mathresogen (<http://name.math.univ-rennes1.fr/ovidiu.radulescu/ACI.html>).

### **Extraction de connaissances de bases bibliographiques**

Le projet Caderige (<http://caderige.imag.fr/>), auquel participent les laboratoires Symbiose (IRISA, Rennes), l'Institut d'Informatique et Mathématiques Appliquées (IMAG, CNRS, Grenoble), le Laboratoire d'Informatique de Paris Nord (LIPN, Université de Paris 13), le Laboratoire de Recherche en Informatique (LRI, Université de Paris 11), le laboratoire de Mathématique, Informatique et Génome (MIG, INRA, Versailles), Le laboratoire de Génétique Microbienne (GM, INRA, Jouy en Josas) et le Laboratoire de Recherches sur la Viande (LRV, INRA, Jouy en Josas) et le laboratoire, comprend 3 étapes :

- . la « recherche d'information » permet de sélectionner dans une base bibliographique, un ensemble de résumés pertinents quant au métabolisme des lipides. 22022 résumés ont été isolés de la base Medline.
- . la « sélection de fragments » permet d'identifier la portion de texte du résumé qui concerne précisément le métabolisme des lipides. D'après la liste des 220 gènes intervenant dans ce métabolisme, 1813 phrases ont été retenues.
- . l'« extraction d'information » permet d'extraire et de donner une représentation formelle de l'information pertinente de manière automatique. Cette dernière étape nécessite au préalable qu'une petite partie des phrases soit annotée manuellement par le biologiste afin d'aider l'apprentissage. C'est la phase qui est en route actuellement.

### **Modélisation de réseau de régulation**

Avec l'équipe Symbiose de l'IRISA, un réseau d'interactions géniques sous forme de graphes est formalisé afin de mieux en analyser la complexité. De nouvelles hypothèses de régulation pourront alors émerger.