

## Acquisition, stockage, analyse et représentation des réseaux d'interactions entre macromolécules biologiques

G.Launay<sup>1</sup>, N. Thierry-Mieg<sup>2</sup>, C. Blanchet<sup>3</sup>, S Ricard-Blum<sup>1</sup>

UMR 5086 CNRS - Université Lyon 1<sup>1</sup>, Institut de Biologie et Chimie des Protéines<sup>3</sup>, 7  
passage du Vercors, 69367 Lyon Cedex 07

UMR 5525 CNRS - Université Joseph Fourier, Laboratoire TIMC-IMAG, Domaine de la Merci,  
38706 La Tronche Cedex

### **Contexte**

Les interactions entre macromolécules sont des événements polymorphes pour lesquels une description biologique pertinente requiert un niveau important de détails. Une approche bio-informatique à même de permettre l'exploitation de ces données "interactomiques" nécessite donc la création et l'emploi d'objets et d'algorithmes parfois complexes. Sous l'impulsion de l'EBI une communauté de bases de données s'est organisée en consortium (the International Molecular Exchange consortium, [www.imexconsortium.org/](http://www.imexconsortium.org/)) pour promouvoir un standard de curation de données d'interactions riche et pour implémenter un format d'échanges de données. Cela permet actuellement aux 11 bases de données du consortium de maintenir un ensemble de 1225771 interactions entre macromolécules (essentiellement protéine-protéine), toutes accessibles de manière programmatique. Notre équipe, qui a créé la base de données d'interactions extracellulaires MatrixDB ([matrixdb.ibcp.fr/](http://matrixdb.ibcp.fr/)), est membre du consortium et est à la fois producteur de contenu (curation de la littérature) et fournisseur de service pour la visualisation et l'exploitation bio-informatique de ces données.

### **Partie Client**

Le serveur MatrixDB subit actuellement une évolution majeure afin d'accroître les possibilités d'exploitation des données d'interactions par l'utilisateur. Pour mener à bien cette évolution, l'approche suivie par notre équipe sépare les aspects représentation et manipulation de l'information des aspects stockage de données et inférence d'information. Les premiers sont confiés au client web de l'utilisateur et les seconds au serveur MatrixDB. Le client web fournit une représentation des interactions en proposant la visualisation de données extraites par les curateurs à partir des publications qui ont mis en évidence les interactions. Ces données correspondent notamment à la méthode de détection des interactions, aux paramètres cinétiques gouvernant les interactions, à l'affinité des interactions et à la localisation des partenaires sur les partenaires.

Aujourd'hui, de nombreux services bio-informatiques, comme Expasy ou Mobylye, permettent l'analyse des macromolécules biologiques. Notre équipe consacre un effort important à l'interopérabilité entre le client web et ces services pour la base de données MatrixDB. Le résultat est une navigation transparente entre ces

contenus biologiques complémentaires qui augmente la richesse des informations accessibles à l'utilisateur. Le client web MatrixDB permet, également, les manipulations simultanées de nombreuses interactions sous la forme d'un réseau. L'implémentation du visualisateur est conçue afin de préserver l'interopérabilité entre la représentation des macromolécules comme noeuds du graphe et les ressources web qui leur sont afférentes. Une des fonctionnalités permises par cette interopérabilité est, par exemple, l'utilisation des données d'expression (telles que celles d'ArrayExpress, [www.ebi.ac.uk/arrayexpress/](http://www.ebi.ac.uk/arrayexpress/)) pour inférer dans un contexte biologique donné les concentrations des protéines du réseau. Le graphe peut donc être contextualisé selon les tissus et/ou les concentrations d'effecteurs par exemple et modulé par les informations expérimentales disponibles en mettant par exemple la cinétique de formation, la stabilité ou l'affinité des interactions comme poids sur les arêtes, données qui sont obtenues expérimentalement dans l'équipe par résonance plasmonique de surface (SPR). Selon les propriétés des graphes, différentes méthodes sont automatiquement éligibles pour mettre en évidence des structures sous-jacentes aux réseaux.

### **Partie Serveur**

L'extraction des informations dans des grands graphes biologiques multivalués est un champ scientifique en développement. La mise en œuvre des algorithmes issus de ces recherches est coûteuse en calcul et en stockage, et doit être réalisée par l'infrastructure coté serveur. Pour cela nous envisageons une approche basée sur des technologies de "cloud computing". Nous nous appuyerons sur une plateforme expérimentale de "cloud" pour la bio-informatiques au sein de la plateforme "Infrastructure Distribuée pour la Biologie". L'extraction et le traitement des données se traduisent le plus souvent à travers la mise en œuvre de pipelines complexes impliquant plusieurs outils et sources de données hétérogènes. Ces traitements peuvent demander le déploiement d'infrastructures spécifiques et complexes. Les technologies de "cloud" peuvent aider à simplifier la mise en œuvre de ces combinaisons d'algorithmes à travers le déploiement à la demande d'infrastructures virtuelles prédéfinies avec les outils bio-informatiques nécessaires comme nous le faisons déjà pour des logiciels standards comme par exemple FastA/SSearch, ClustalW/omega, OMSSA et X!Tandem, et des interfaces scientifiques comme Galaxy ou MobyLe.

### **L'équipe souhaiterait participer à l'atelier PROSPECTOM autour de 3 axes**

- Apport d'expertise pour aider à la création d'un modèle de représentation des données "interactomiques".
- Une participation au déploiement et à l'évaluation des algorithmes permettant l'extraction d'informations dans des réseaux riches et de grande taille comme les réseaux d'interactions biomoléculaires.
- Travailler à la conception ou au déploiement des méthodes pour l'intégration de données hétérogènes (cinétique et affinité) et la visualisation de réseaux biologiques modulables en fonction des contextes étudiés par les utilisateurs.