

## RÉUNION SATELLITE des JOBIM

# Modélisation Dynamique de Réseaux de Régulation Biologique Des mesures aux modèles et des modèles aux mesures

samedi 9 juillet 2005

IBCP (Institut de Biologie et Chimie des Protéines)- Lyon



### Programme

- 8h30-9h00 **Accueil**
- 9h00 - 9h45 Bernard Vandebunder : *Structure, dynamique et robustesse des réseaux de régulation transcriptionnelle.*
- 9h45-10h30 Janine Guespin-Michel, Didier Filopon, Annabelle Mérieau, Gilles Bernot, Jean-Paul Comet, Benoit Polack : *Approche logique, bio-informatique et expérimentale de la dynamique d'un réseau de régulation chez Pseudomonas aeruginosa.*
- 10h30-11h00 **Pause**
- 11h00-11h45 Delphine Ropers, Danielle Bonaccio, Hidde de Jong, Dominique Schneider, Johannes Geiselmann : *Analyse de la réponse au stress nutritionnel chez Escherichia coli : utilisation des gènes rapporteurs gfp et lux.*
- 11h45-12h30 Gilles Curien : *Construction d'un modèle d'un système multienzymatique: hypothèses, simplifications et collecte de données biochimiques.*
- 12h30-14h00 **Repas**
- 14h00-14h45 Hubert Charles et Federica Calevro : *Mesure des niveaux d'expression des gènes chez les bactéries par la technique des biopuces : analyse du transcriptome de la bactérie Buchnera aphidicola en condition de stress trophique de son hôte, le puceron Acyrthosiphon pisum.*
- 14h45-15h30 Christine Brun : *Interactions protéine-protéine: du producteur au consommateur.*
- 15h30-16h00 **Pause**
- 16h00-16h45 Yves Vandebrouck : *Analyse de données de protéomique par spectrométrie de masse*
- 16h45-17h30 Discussion finale

### Organisation :

Claudine Chaouiya (LGPD/IBDM, Marseille)

Hidde de Jong (INRIA Rhône-Alpes)

### Remerciements :

Cette journée est organisée dans le cadre des activités de l'ACI Vicanne (Action Concertée Incitative IMPBio). Nous remercions également l'Institut de Biologie et Chimie des Protéines pour l'accueil ainsi que les organisateurs de JOBIM pour leur soutien financier et logistique.

---

## Structure, dynamique et robustesse des réseaux de régulation transcriptionnelle.

Bernard Vandebunder  
Institut de Recherche Interdisciplinaire, Lille

L'expression d'un gène est un processus séquentiel dont chaque étape est réalisée par des machines moléculaires constituées de plusieurs dizaines de sous unités. L'activité de ces machines est coordonnée et contrôlée par des réseaux complexes de régulation. L'étude de ces réseaux de régulation a été abordée par des approches globales à partir des ARNs extraits d'un tissu ou d'une culture cellulaire. Depuis une dizaine d'années, des nouvelles techniques d'imagerie permettent de visualiser chaque étape de la transcription dans des cellules vivantes individuelles. Grâce à ces approches cellulaires, on peut avoir accès plus directement aux propriétés dynamiques et spatiales des réseaux de régulation transcriptionnelle.

La connaissance des méthodes expérimentales, la quantification des concentrations des acteurs moléculaires impliqués et des constantes cinétiques des réactions élémentaires apportent des informations précieuses pour l'interprétation des données des DNA arrays, la construction de modèles représentant ces réseaux de régulation et la conception de nouvelles expériences qui permettront de tester la validité de ces modèles.

---

### Approche logique, bio-informatique et expérimentale de la dynamique d'un réseau de régulation chez *Pseudomonas aeruginosa*.

J.F. Guespin-Michel<sup>2</sup>.

D. Filopon<sup>1</sup>, A. Mérieau<sup>2</sup>, G. Bernot<sup>3</sup>, J.P. Comet<sup>3</sup>, B. Polack<sup>1</sup>,

<sup>1</sup> GREPI EA 2938 Université Joseph Fourier Grenoble 1, Laboratoire d'Hématologie, CHU de Grenoble, BP 217, 38043 Grenoble cedex 9

<sup>2</sup> Laboratoire de microbiologie du froid, EA 2123, Université de Rouen, F-76 821 Mt St Aignan

<sup>3</sup> LaMI, CNRS UMR 8042, Université d'Évry-Val-d'Essonne, Boulevard François

Cet exposé est destiné à illustrer une démarche : une hypothèse biologique est formulée, sa cohérence est testée par un modèle, qui est simulé grâce à un logiciel dédié, et une expérience, suggérée par ce modèle est réalisée qui valide l'hypothèse.

**L'hypothèse** est que l'acquisition d'un phénotype de cytotoxicité par la bactérie *Pseudomonas aeruginosa*, dans les poumons de malades atteints de mucoviscidose est due à un switch épigénétique. Il s'agit d'un processus dynamique (passage d'un état stationnaire à l'autre dans un système dynamique non linéaire à deux états stables), qu'il n'est pas possible d'étudier par les seules méthodes de la biologie moléculaire.

**Un modèle** a été construit par la méthode de logique généralisée.

**Un logiciel dédié de model checking** permet de trouver les conditions pour lesquelles ce modèle est conforme à l'hypothèse épigénétique. Ceci montre la cohérence entre l'hypothèse et le modèle.

Cette méthode permet aussi de **proposer une expérience** validant l'hypothèse.

**Cette expérience a été réalisée en laboratoire.** La cytotoxicité est acquise par un switch épigénétique, chez *Pseudomonas aeruginosa*.

Cette démarche peut être généralisée dans plusieurs directions.

---

## **Analyse de la réponse au manque de carbone chez *Escherichia coli* : utilisation des gènes rapporteurs *gfp* et *lux***

Delphine Ropers, Danielle Bonaccio, Hidde de Jong, Dominique Schneider, Johannes Geiselmann  
Institut National de Recherche en Informatique et en Automatique (INRIA) Unité de recherche  
Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex

Chez la bactérie *Escherichia coli*, un réseau de régulation génique composé de gènes, protéines et métabolites contrôle l'adaptation de la croissance bactérienne à la disponibilité en nutriments. Les composants clés de ce réseau sont des régulateurs globaux de la transcription qui régulent l'expression d'ensembles de gènes suivant les conditions environnementales [1]. Si l'identité de ces régulateurs est bien connue, ceux-ci sont nombreux et reliés par des boucles de rétroaction complexes qui rendent difficile la compréhension intuitive du fonctionnement du réseau. Afin de comprendre comment l'adaptation de la croissance d'*E. coli* émerge des interactions entre les régulateurs globaux, un modèle mathématique du réseau a été construit sur la base des données de la littérature et de banques de données publiques [2]. Ce modèle est basé sur une classe d'équations différentielles linéaires par morceaux qui possède des propriétés favorables à l'analyse qualitative du système [3,4].

Le modèle a été utilisé pour simuler la réponse de la bactérie à la présence ou l'absence de source de carbone, ce qui a permis de mieux comprendre le rôle des régulateurs globaux et leurs interactions dans l'adaptation de la croissance bactérienne [2]. Ce modèle a également motivé la réalisation d'expériences qui sont actuellement en cours. Il s'agit en particulier de vérifier des prédictions surprenantes, pour lesquelles il n'existe pas à l'heure actuelle de données expérimentales permettant de les valider ou non. La vérification de ces prédictions nécessite d'obtenir des données dynamiques d'expression des régulateurs globaux du réseau lors de l'adaptation de la croissance bactérienne. Dans cet objectif, les gènes rapporteurs codant la protéine GFP ("Green Fluorescent Protein") et la luciférase constituent des outils précieux pour mesurer l'activité des gènes à la fois au niveau transcriptionnel et traductionnel. Leur application à la compréhension de la réponse au manque de carbone chez *E. coli* sera présentée.

[1] Hengge-Aronis, 2000, Bacterial Stress Responses, G. Storz and R. Hengge-Aronis (eds), ASM Press, Washington D.C, pp. 161-178

[2] Ropers et al., 2005, BioSystems, accepté

[3] de Jong et al., 2004, Bull. Math. Biol., 66(2):301-340

[4] Glass et Kauffman, 1973, J. Theor. Biol., 39(1):103-129

---

## **Construction d'un modèle d'un système multienzymatique: hypothèses, simplifications et collecte de données biochimiques.**

Gilles Curien

Laboratoire de Physiologie Cellulaire Végétale  
DRDC-PCV - CEA/CNRS/INRA  
17 avenue des Martyrs – 38054 Grenoble

Dans les systèmes métaboliques contenant de nombreuses interactions allostériques agissant à plusieurs niveaux la fonction d'une interaction donnée est difficilement explicable en l'absence de modèles mathématiques. Pour être utiles au biochimiste, de tels modèles doivent permettre de faire un aller-retour entre le comportement du système dans lequel les enzymes officient et les propriétés des enzymes isolées. Ces modèles doivent permettre d'identifier les variables pertinentes du

système et ainsi de concevoir des expériences permettant de mesurer celles-ci *in vivo*. Nous décrivons une procédure que nous suivons actuellement pour obtenir un modèle mathématique du système de 25 enzymes assurant la répartition du squelette carboné de l'aspartate vers les produits "terminaux" lysine, thréonine, méthionine/*S*-adénosylméthionine et acides aminés branchés chez *Arabidopsis*. Nous exposerons les hypothèses nécessaires et justifierons la délimitation du système modélisé (dissection) ainsi que les simplifications du système et des équations de vitesses. Nous montrerons l'importance de la prise en compte du contexte métabolique dans cette étape de construction du modèle. La modélisation et les simulations d'une sous partie "autonome" du système nous permettront de montrer le caractère itératif de la modélisation et les intérêts de cette approche pour le biochimiste et le physiologiste.

---

**Mesure des niveaux d'expression des gènes chez les bactéries par la technique des biopuces : analyse du transcriptome de la bactérie *Buchnera aphidicola* en condition de stress trophique de son hôte, le puceron *Acyrtosiphon pisum*.**

Charles H. et Calevro F.

UMR 203 INRA / INSA de Lyon, Biologie Fonctionnelle, Insectes et Interactions, INSA, Bât. Louis Pasteur, 69621 Villeurbanne Cedex, Biologie des Systèmes et Modélisation Cellulaire  
(<http://bsmc.insa-lyon.fr/>)

Les techniques d'analyse dites « à haut débit » (séquençage, SAGE et biopuces) ont révolutionné la biologie de ces dernières années. Alors que jusqu'à présent la communauté des biologistes moléculaires était confrontée à l'analyse de données essentiellement qualitatives et restreintes à un gène unique ou à un petit groupe de gènes, elle doit maintenant faire face à des volumes de données très importants qui demandent le développement d'outils nouveaux pour le stockage, la représentation, l'analyse et la modélisation des résultats.

Dans l'UMR BF2I, nous avons initié en 2001 une thématique de recherches sur l'analyse du transcriptome par la technique des biopuces appliquée à la bactérie symbiotique intracellulaire *Buchnera aphidicola*, associée au puceron *Acyrtosiphon pisum*. Le contexte de notre étude est la compréhension, à un niveau moléculaire et physiologique, des interactions entre la bactérie et son hôte, notamment pour ce qui concerne la complémentarité nutritionnelle en acides aminés que *Buchnera* fournit au puceron. De nombreux développements ont du être entrepris pour aller de la conception du plan expérimental à l'interprétation fonctionnelle des résultats.

La conception de la biopuce est la première étape de ce processus. Le choix et l'optimisation des séquences sondes, le choix du nombre de répétitions et la détermination de témoins négatifs et positifs sont des contraintes importantes dans ce type de projet. Un logiciel de sélection de sondes oligonucléotidiques a été développé en conséquence par N. Reymond dans l'unité (<http://pbil.univ-lyon1.fr/rosa/>).

La procédure d'obtention des données physiques inclut un très grand nombre d'étapes successives dont la préparation des échantillons biologiques, la conception du plan expérimental, le marquage des ARN cibles, le spottage des lames, la réaction d'hybridation, le lavage et le scanning. Chacune de ces étapes peut apporter un certain nombre de biais qu'il convient d'intégrer dans l'analyse des données par la mise en œuvre d'une méthode de normalisation adéquate. Ainsi, il est très important de développer une démarche qualité tout au long du processus. Un système d'information pour le stockage, la visualisation et l'échange des données de biopuce a été développé dans ce but (<http://sitrans.insa-lyon.fr/>).

Ce n'est qu'au terme de ce processus que le modélisateur pourra intégrer les données d'expression dans des analyses statistiques (détection de gènes différentiels ou classification) ou comme référence dans les cas de simulation par des modèles mathématiques ou informatiques.

---

## **Interactions protéine-protéine: du producteur au consommateur.**

Christine Brun

LGPD/IBDM Institut de Biologie du Développement de Marseille

Dans une première partie, les principales méthodes d'identification d'interactions protéine-protéine à grande échelle seront présentées, leurs limitations et la fiabilité de leur résultats discutées.

Les données produites par ces méthodes expérimentales chez des organismes divers (levure, ver, mouche) ont permis la représentation d'une partie du fonctionnement de la cellule sous forme de grands réseaux d'interactions protéine-protéine. Je présenterai donc nos méthodes d'analyse de la structure de ces réseaux qui conduisent à des classifications de protéines d'après les interactions protéine-protéine chez la levure et la mouche. J'illustrerai particulièrement l'importance de la robustesse de ces méthodes à la qualité des données.

J'évoquerai ensuite comment l'utilisation d'informations biologiques différentes des données d'interactions, (contenues par exemple dans les ontologies de fonction, de localisation, d'anatomie, de phénotype, etc) en aval de l'analyse de la structure de réseaux d'interactions protéine-protéine, peut contribuer à y introduire des notions de dynamique.

---

## **Analyse de données de protéomique par spectrométrie de masse.**

Yves Vandembrouck

Laboratoire de Biologie, Informatique et Mathématiques Département de Réponse et Dynamique  
Cellulaires (DRDC)

CEA/DSV/DRDC/BIM, Bât D3, 17 Rue des martyrs  
38054 Grenoble Cedex 9, France

Les défis méthodologiques et technologiques posés par la protéomique à haut-débit tiennent à la fois au très grand nombre de protéines à analyser, à l'hétérogénéité de leurs propriétés physicochimiques, et au fait que de très nombreuses protéines sont faiblement exprimées. Parallèlement au changement d'échelle des techniques expérimentales, la bioinformatique est amenée à jouer un rôle majeur : il ne s'agit plus seulement d'identifier une liste de protéines, mais aussi d'être capable de proposer d'autres types de résultats, comme les mesures d'expression et l'annotation de génomes. Dans ce contexte, l'analyse des données de spectrométrie de masse peut se révéler une tâche ardue face à la quantité de données produites, la complexité des informations utilisables, la précision des analyses quantitatives et la mise en oeuvre des ressources bioinformatiques appropriées. Ces différents aspects seront abordés à la lumière des approches expérimentales et informatiques actuellement développées.