

**ANR**  
**AGENCE NATIONALE DE LA RECHERCHE**  
**PROGRAMME IMPBIO**

**Rapport de fin de projet**

( à transmettre en 2 exemplaires )

**I - FICHE D'IDENTITE DU PROJET**

---

**DECISION D'AIDE N°03 2 633**

**Numéro de référence du projet :**

**Titre du Projet :** MathResoGen "Méthodes mathématiques pour l'identification d'acteurs critiques dans des processus biochimiques régulés par un réseau de gènes"

**Coordinateur du projet :** Ovidiu Radulescu

**Tél du coordinateur du projet :** 02 9984 7306

**Mél du coordinateur du projet :** ovidiu.radulescu@univ-rennes1.fr

**Laboratoire de rattachement du coordinateur:** IRMAR, Institut de Recherches Mathématiques de Rennes, UMR 6625

**Adresse postale du laboratoire :** Campus de Beaulieu, 35042 Rennes

**Numéro d'unité:** UMR 6625

**Montant global** (en KEurosTTC) : 137

**Durée :** 3 ans

<b>Partenaires du Projet</b>	<b>Laboratoire (pas de sigle)</b>
Ovidiu Radulescu Elisabeth Pécou Anne Siegel Frédéric Grognard Sandrine Lagarrigue Alain Lilienbaum Nathalie Theret	Institut de Recherches Mathématiques de Rennes Institut de Mathématiques de Bourgogne Projet Symbiose, IRISA (UMR CNRS-INRIA) - Rennes Projet Comore, INRIA-Sophia Antipolis UMR598 Génétique Animale, INRA, Agrocampus Rennes Laboratoire Cytosquelette et Développement, CNRS Paris Détoxication et Réparation Tissulaire, Inserm Rennes

N.B : S'il y a eu modifications des partenaires initialement prévus, en expliquer les raisons.

## II – BILAN DETAILLE DU PROJET

---

### 1 – Rappel des objectifs initiaux du projet :

MathResoGen a été défini en tant que projet interdisciplinaire destiné à produire des outils mathématiques pour l'analyse du fonctionnement des systèmes biologiques. Il a ciblé trois applications principales : le métabolisme des lipides du foie chez le poulet; la signalisation de TGF $\beta$  dans le cancer du foie; l'induction de NF $\kappa$ B, régulateur de la signalisation intercellulaire et du cycle cellulaire. Les réponses souhaitées à ces problématiques ont été inspirées par les systèmes dynamiques différentiels et probabilistes :

- a) Existence de plusieurs échelles de temps, abordée par des techniques de perturbations singulières.
- b) Complexité des systèmes, dont la maîtrise impose d'avoir une approche hiérarchique et modulaire ; identification des paramètres pertinents expliquant le phénomène biologique par affinement successif de l'analyse en suivant la hiérarchie du modèle.
- c) Stochasticité des processus biochimiques, conduisant d'une part à des observations bruitées (notamment celles provenant de la microscopie à fluorescence), et d'autre part à des fluctuations du phénotype et à des instabilités ; l'utilisation de l'analyse stochastique comme complément de validation des modèles.

Les méthodes mathématiques ont été formalisées dans des algorithmiques et intégrées dans des outils informatiques. Ensuite, elles ont été appliquées au traitement des données expérimentales biologiques et à la mise en place de nouvelles expérimentations en fonction des analyses obtenues.

L'objectif final de ce projet était d'aider à la découverte de propriétés dynamiques de réseaux biologiquement pertinentes, impossible à trouver sans l'aide de modèles. Dans cette optique, nous avons proposé d'enrichir par des méthodes mathématiques la panoplie de méthodes utiles au biologiste théoricien : quand et comment est-il légitime de simplifier les modèles discrets, différentiels voire stochastiques du métabolisme et des régulations géniques ? Peut-on les regrouper en grandes classes de modèles pour lesquels il est possible de caractériser certains comportements ? De même, nous avons proposé de déterminer, dans des réseaux complexes

mettant en relation métabolisme, gènes et voies de signalisation (dans des organismes pluricellulaires évolués), les variables génétiques ou métaboliques qui sont susceptibles d'influencer sensiblement le fonctionnement du réseau.

De par sa nature, MathResoGen représente sans aucun doute un projet en biologie systémique. Depuis le départ nous avons souhaité nous distinguer par rapport à d'autres projets en biologie systémique (on peut mentionner le projet de H. de Jong, J.-L. Gouzé et H. Geiselmann relatif à la régulation de la transcription chez E.Coli par des modèles discrets ou linéaires par morceaux) en intégrant une panoplie de méthodes mathématiques complémentaires et plus vastes. De même, nous avons voulu nous distinguer par rapport à la communauté sur le contrôle métabolique en cherchant des nouvelles méthodologies permettant les analyses qualitatives et la modélisation de la régulation génétique du métabolisme. Finalement, conformes aux problématiques naissantes de la biologie systémique, nous avons proposé l'étude des principes fondamentaux du fonctionnement du vivant : la stabilité, la robustesse, la hiérarchisation, la modularité, la stochasticité.

## **2 – Rapport final**

### **2.1 Programme des actions engagées**

- en cas de modification du programme des travaux préciser les raisons et les réorientations
- préciser les modalités concrètes de travail interdisciplinaire

Nous avons suivi les objectifs principaux du projet, en consacrant d'abord un effort important au développement de nouvelles méthodes mathématiques d'analyse de systèmes et données biologiques. Ces méthodes mathématiques ont été appliquées aux problèmes biologiques ciblés par le projet. La généralité des méthodes nous a permis de les appliquer aussi à d'autres problèmes, qui n'étaient pas initialement prévus dans le projet, tels que la régulation de la transcription chez E.Coli, ainsi que les interactions entre les gènes du développement chez la Drosophile.

L'étude de la stabilité de la morphogenèse chez la Drosophile s'ajoute aux objectifs de modélisation initiaux. En fait, nous ont été amené tout naturellement à considérer ce problème lors de nos réflexions sur la robustesse des réseaux biologiques. Les idées développées pour

expliquer la robustesse du réseau de signalisation de NFκB, ont pu être utilisées aussi pour étudier la canalisation du développement. Il faut mentionner que la robustesse en général et la stabilité du développement en particulier font partie des grands problèmes ouverts en biologie systémique. Les aborder, nous a permis de donner plus de poids à notre activité et a ouvert la porte à des collaborations internationales.

En ce qui concerne les difficultés que nous avons rencontrées, nous devons signaler quelques retards dans la production de données et dans l'application de nos méthodes aux résultats de l'expérimentation. Ces difficultés sont tout à fait compréhensibles, vu l'ampleur du sujet, et n'affectent pas l'importance de nos résultats. La puce pangénomique destinée à l'étude du métabolisme des lipides a été obtenue seulement à la fin du projet, en grande partie à cause de problèmes lors de la conception de puces au centre de bioinformatique de l'Inra (sur lequel nous ne pouvions pas agir directement). Nous avons donc modifié nos objectifs initiaux en se concentrant sur l'étude d'un modèle simplifié du métabolisme et en développant une collaboration avec l'unité INRA (UR66) de Toulouse pour avoir accès à des données physiologiques de la dynamique à jeun-nourri. Nous avons aussi décidé de repousser l'expérimentation prévue sur l'étude des effets stochastiques dans la signalisation de NFκB, pour consolider notre compréhension théorique du système, dans le but de décider sur les cibles à observer et d'établir un plan optimal d'expérience.

Le travail interdisciplinaire a été rendu possible par plusieurs assemblées générales d'information et de travail (organisées à Rennes et à Paris) et aussi par la constitution de plusieurs groupes de travail sur les différentes thématiques qui se sont réunis périodiquement. Chaque groupe thématique a eu une composition interdisciplinaire. La modélisation du métabolisme a été étudiée par le groupe de S.Lagarrigue, E.Pécou, O.Radulescu et A.Siegel. A ce groupe se sont associées des membres d'autres unités INRA de Rennes et de Toulouse. Les méthodes qualitatives d'analyse de réseaux et leurs applications possibles ont été développées par le groupe C. Guziolowski (doctorante associée au projet depuis son stage de master en 2005), M. Le Borgne, O.Radulescu, A.Siegel, P.Veber (doctorant associé au projet depuis 2004). La signalisation TGFβ a été étudiée par N. Theret, J. Gruel (doctorant associé au projet depuis son stage de master en 2005), M. Le Borgne, P.Veber. La robustesse a été étudiée en relation avec la modélisation de NFκB et de la morphogenèse, par O.Radulescu, A.Zinovyev (de l'Institut Curie, associé au projet à partir de la deuxième année), A.Lilienbaum, D.Grigoryev, A.Gorban (collaborateur étranger) et S.Vakulenko

(collaborateur étranger). Les techniques d'analyse de réseaux par identification de modules ont été élaborées par O.Radulescu, A.Siegel, F.Grognard. Les phénomènes stochastiques ont été étudiés par O.Radulescu, A.Lilienbaum, A.Deussche et A.Crudu (doctorante associée au projet en 2005).

Nous avons ainsi obtenu :

- Une nouvelle méthode d'analyse de réseaux moléculaires, utilisant les équations qualitatives. Cette méthode permet de tester la cohérence entre les modèles d'interaction et les données transcriptomes, de proposer des corrections, de fournir de prédictions et des plans d'expérience. L'élaboration de cette méthode a été l'investissement le plus important en tant que quantité de travail, pendant toute la durée du projet. La méthode est implémentée dans un nouveau logiciel développé à Rennes : PYQUALI, qui sera disponible via la plateforme bioinformatique de Ouest-Genopole à la fin du premier trimestre 2007.
- Une approche mathématique nouvelle, proposant les processus déterministes par morceaux comme cadre naturel d'étude des phénomènes stochastiques en biologie moléculaire. Cette approche a été développée pendant la deuxième année du projet et a conduit au développement d'un algorithme accéléré de simulation stochastique.
- Un nouveau cadre conceptuel d'étude de la robustesse des systèmes biologiques, identifiant les phénomènes de concentration de mesure en grande dimension comme source de robustesse distribuée. Ces idées ont émergé pendant la dernière année du projet.
- Plusieurs techniques de réduction de modèle : une méthode utilisant la théorie des graphes, inspirée par Clarke et une méthode du type Carr-Pego (variété invariante) pour les modèles par EDP de systèmes étendus (en relation avec la morphogenèse de la Drosophile). La recherche des approches appropriées de réduction de modèle s'est déroulée pendant toute la durée du projet.
- Une méthode nouvelle d'identification de modules, basée sur un théorème d'univalence globale (Gale-Nikaido). Cette méthode a été inspirée par la modélisation du métabolisme des lipides, mais son domaine d'applicabilité est plus général. La méthode a été implémentée dans un logiciel qui utilise le calcul formel.

Nous avons également développé un certain nombre de modèles :

- Un modèle abstrait de métabolisme de lipides comportant une dizaine de variables. Ce modèle intègre plusieurs voies métaboliques (oxydation, synthèse, lipolyse) et leur régulation génétique et a pu être testé sur des données métaboliques et génétiques chez la souris.
- Un modèle étendu du type graphe d'interaction pour la régulation de la lipogenèse chez le poulet comportant plusieurs centaines de variables.
- Un modèle cinétique biochimique pour la signalisation de NFκB comportant plus d'une centaine de variables.
- Un modèle graphique bipartite (graphe de réactions biochimiques) pour la signalisation de TGFβ comportant plusieurs centaines de variables.

Le programme expérimental s'est concentré sur :

- L'étude de la dynamique à jeun-nourri du métabolisme des lipides.
- Les spécificités de la signalisation de TGFβ en relation avec la fibrogenèse.

## **2.2 Apports et résultats scientifiques, y compris les avancées permises par la collaboration interdisciplinaire.**

Les résultats sont présentés dans deux catégories : (I) méthodes, (II) avancées biologiques et modélisation.

### **I. Méthodes**

#### **Construction, correction et analyse qualitative de modèles du type graphe d'interaction**

Nous avons élaboré une stratégie de construction et d'étude de modèles qualitatifs du type graphe d'interaction [1-5]. Pour un problème étudié nous construisons d'abord une base de connaissances dédiée alimentée par des sources multiples (extraction manuelle de la littérature et bases de connaissances non-spécifiques telles que KEGG, INGENUITY, TRANSPATH). Nous avons construit une base sur la régulation du métabolisme des lipides et une autre sur la signalisation du TGFβ.

Ces connaissances issues de la littérature, qui définissent un premier modèle qualitatif (graphe d'interaction), sont presque toujours incomplètes et parfois incorrectes. Il est nécessaire de les

compléter et corriger. Pour cela, nous avons développé une théorie mathématique de la relation entre la topologie d'un réseau ou des parties d'un réseau et la réponse du réseau ou des parties à des perturbations. La situation formalisée est typique en génétique fonctionnelle, car le fonctionnement des grands réseaux de gènes, de protéines ou métabolites (dont la topologie est parfois incomplètement caractérisée) est observé à travers leur réponse aux changements des conditions extérieures ou à des actions sur des gènes, protéines, métabolites (mutation, délétion, amplification, etc.).

Nous avons montré que le problème étudié est en fait une extension du problème de Dirichlet pour l'équation de Laplace sur graphes orientés, d'une part pour des opérateurs et équations généralisant le laplacien, et d'autre part dans un contexte non-linéaire. La solution que nous proposons pour ce problème rappelle la théorie du feed-back utilisée en réseaux électriques (formules de Mason-Coates). Nous exprimons la relation entre le signe et la topologie des boucles dans le modèle graphique d'interaction et la variation des quantités définies sur les noeuds du graphe (concentrations de gènes, protéines, métabolites, etc.) lorsque le réseau est soumis à des perturbations. On peut dire que nous avons développé une *théorie de l'élasticité* des réseaux.

Très vite, nous nous sommes aperçus que notre formalisation mathématique de la relation entre la topologie et la réponse justifie l'écriture d'équations qualitatives à partir de graphes. A chaque noeud du réseau on associe une variable qualitative à trois niveaux représentant le comportement qualitatif (1="augmente", -1="décroit", 0="constant"). Les perturbations sont transmises par le réseau, mais le niveau qualitatif d'un noeud ne dépend que des niveaux qualitatifs de ses voisins. La relation entre les niveaux qualitatifs est l'addition dans l'algèbre de signes. On peut ainsi associer à tout couple modèle (graphe orienté) /données (observations de certains noeuds) un système d'équations qualitatives dans l'algèbre des signes. Le modèle et les données sont compatibles lorsqu'il existe au moins une solution de ce système. En absence de solution, on dispose d'algorithmes permettant de trouver des modifications du modèle (correction de modèle) ou des données (correction de données) qui sont minimales par rapport à un critère imposé (par exemple nombre de noeuds ou d'arcs modifiés).

Ces méthodes s'apparentent ainsi à des méthodes de diagnostic introduites dans les années 80 pour la détection et la réparation des erreurs dans des applications diverses en circuits électroniques, diagnostic médical, etc.. Un modèle cohérent avec les données produit un

certain nombre de prédictions que nous associons aux composantes dures des équations qualitatives (variables dont le signe ne change pas d'une solution à une autre).

Dans le domaine des réseaux de gènes, ceci ne représente pas une nouveauté absolue car ces techniques ont déjà porté leurs fruits sous un autre nom, celui des réseaux booléens introduits par Stuart Kauffman dans les années 70. Un réseau booléen associe seulement deux niveaux (0 pour inactif, 1 pour actif) à tout noeud d'un graphe orienté. La dynamique de ce réseau est un automate (synchrone ou asynchrone) qui réactualise les valeurs d'un noeud, par le résultat d'une fonction booléenne définie sur ses voisins. Ces modèles ont été (relativement) récemment utilisés par Ideker pour la reconstruction de réseaux de gènes à partir de perturbations. En comparaison avec ces travaux, nos résultats ont l'avantage d'une justification rigoureuse des opérations sur graphe à partir d'un modèle différentiel. Les modèles booléens souffrent d'une incertitude sur le choix du type de fonction logique (conjonction, disjonction, etc). Dans nos réseaux à trois valeurs, nous avons travaillé sur la justification du fait que la loi de composition doit être tout simplement l'addition dans l'algèbre de signes.

Un autre avantage de nos méthodes réside dans la résolution automatique des équations qualitatives. L'algorithme utilise le codage du système d'équations qualitatives par une seule équation polynomiale sur  $Z/3Z$ . Bien entendu, trouver les solutions du système d'équations qualitatifs est NP-complet. Néanmoins, le codage polynomial est optimisé par réduction sous forme de diagrammes de décision. Ceci permet de traiter des réseaux de tailles importantes (quelques centaines de gènes). La méthode des équations qualitatives permet de prendre en compte des données hétérogènes (données d'expression différentielle, microARN, CGH). Nous travaillons en ce moment sur l'extension de ces méthodes aux séries temporelles et aux données ChIP on chip. Cette stratégie (testée pour l'instant sur des données d'expression lors du stress nutritionnel chez E.coli) est automatisée et permettra l'amélioration des connaissances et du modèle qualitatif [4]. Le logiciel correspondant (un module Python) PYQUALI est en cours d'enregistrement à l'agence de protection de logiciels.

Une autre application de ces résultats est l'étude des compétitions dans un réseau [2]. Dans un réseau, les perturbations sont transportées sur les différents chemins reliant la frontière et l'intérieur du graphe. Les contributions des chemins sur un noeud du réseau peuvent avoir des signes différents. On parle alors de compétitions. L'étude de compétitions est un premier

outil d'analyse de la flexibilité d'un réseau biologique (caractérisée par la multitude de comportements possibles) et d'identification de cibles potentielles pour des thérapies.

### **Modèles différentiels et leur réduction**

En parallèle avec les méthodes qualitatives, nous avons développé des méthodes d'analyse de modèles dynamiques quantitatifs, notamment des modèles différentiels. La difficulté principale dans l'étude de tels modèles est posée par le nombre grand de paramètres et de variables dynamiques. Notre stratégie a été de développer des méthodes de réduction de modèle [13,14].

Actuellement, la plupart des modèles en biologie systémique sont des systèmes d'équations différentielles résultant de la cinétique d'un réseau de réactions bio-chimiques. En particulier, la loi d'action de masse pour la cinétique chimique produit des systèmes polynomiaux. La complexité (quantifiée par le nombre de variables et de paramètres) de ces modèles est très grande. Nous avons réduit la complexité en réduisant le nombre de variables et le nombre de paramètres [14]. Nous avons aussi étudié le concept de dimension dynamique minimale définie comme le nombre d'équations qui rend compte (à un difféomorphisme près) de la dynamique du système [10,14]. Ce concept est important étant donné que beaucoup de systèmes biologiques de structure complexe ont une dynamique simple. La complexité permet d'allier flexibilité et robustesse pour réaliser un nombre grand de tâches simples, de manière fiable [10,11].

Nous avons utilisé plusieurs méthodes de réduction.

D'abord les méthodes de type variété invariante attractive (également applicables à des systèmes d'EDP). Elles peuvent s'appliquer lorsque, après une étape transitoire, le système dynamique évolue sur une variété de dimension petite. Il y a plusieurs situations quand cette réduction est possible: a) les systèmes multi-échelle (singulières pour lesquels les modes rapides sont soumises (slaved) à la dynamique des modes lentes [13], ou modulés pour lesquels un mode lent module la solution d'une équation plus simple [14]; on peut mentionner, selon le cas, au moins trois méthodes d'approximation: les perturbations singulières, les méthodes à deux ou plusieurs échelles, l'homogénéisation) b) les systèmes distribués avec brisure de symétrie continue et apparition d'un nombre petit de modes lentes (par exemple les

EDP de réaction-diffusion nonlinéaires admettant des solutions du type interface mobile) c) les systèmes proches de bifurcations (par exemple bifurcation de Hopf). En biologie on retrouve les trois types cités: les systèmes biologiques sont multiéchelle conduisant au type a) (modes rapides dissipatifs comme souvent dans les systèmes métaboliques régulés génétiquement, conduisant à une dynamique maître-esclave, modes rapides oscillants ou systèmes avec loi de conservation brisée en signalisation, conduisant à des modulations), le type b) est rencontré en morphogenèse le type c) s'applique aux oscillateurs biologiques. Malheureusement, même quand l'existence d'une variété invariante est prouvée, il est difficile en général de caractériser la variété invariante et encore plus difficile d'étudier sa dépendance dans les paramètres. En collaboration avec Alexander Gorban et Andrei Zinovyev, nous avons développé des méthodes numériques de construction et de caractérisation de la variété invariante pour des systèmes de réactions biochimiques [10]. En collaboration avec Sergei Vakulenko, nous avons étudié l'application de la décomposition du type variété invariante aux EDP modélisant la morphogenèse [15,16].

Dans notre analyse hiérarchique de la signalisation de NFkB [14] nous avons utilisé la projection Karhunen-Loève, qui utilise l'analyse statistique des trajectoires du système pour trouver la plus petite variété linéaire contenant la dynamique. Notons que le choix de la métrique peut être important et la non-linéarité peut limiter l'applicabilité de ces méthodes. Par exemple, lorsque l'amplitude des modes linéaires décroît lentement il faut prendre en compte beaucoup de modes pour obtenir une bonne approximation de la dynamique (par exemple pour l'équation Kuramoto-Sivashinsky). Dans ce cas il faut utiliser des méthodes de projection non-linéaires et un choix adapté de la métrique.

Nous avons également développé un troisième type de méthode, utilisant la théorie des graphes et permettant de préserver la structure du modèle en tant que réseau de réactions biochimiques [14]. Cette méthode est utile pour comparer des modèles de complexités différentes, produire des hiérarchies de modèles et pourra être associée à des bases de modèles pour la biologie systémique (collaboration avec Upi Bhalla de Bangalore).

### **Modélisation stochastique**

Les modèles mathématiques qui décrivent des processus biologiques peuvent être soit déterministes, soit stochastiques. Ayant comme base de départ la cinétique chimique, les

modèles aux équations différentielles ordinaires sont utilisés pour décrire l'évolution temporelle des concentrations d'espèces moléculaires du système. D'autre part, les modèles stochastiques sont étudiés pour décrire le bruit interne dû aux espèces en faible nombre, ainsi que pour illustrer le bruit externe dû aux fluctuations de l'environnement.

Les processus de Markov à sauts permettent la modélisation des phénomènes stochastiques en biologie moléculaire. Néanmoins, il y a peu de résultats mathématiques sur la dynamique de ces processus. Aussi, leur simulation sur ordinateur rencontre des difficultés reliées au temps d'exécution. Nos résultats permettent de réduire la complexité de la dynamique stochastique. Ces méthodes utilisent des théorèmes limites probabilistes [9].

Les trajectoires des processus de Markov à sauts convergent vers des trajectoires déterministes, solutions d'équations différentielles, lorsque le nombre de chaque type de réaction par unité de temps est important (Kurtz, 1970,1971). Dans cette limite, appelée limite fluide par analogie avec le même concept en recherche opérationnelle, les phénomènes stochastiques disparaissent. Il s'agit de la loi des grands nombres pour les processus de Markov. Dans les mêmes conditions, il existe aussi un théorème central limite qui permet l'approximation des processus de Markov par des diffusions (Kurtz, 1971, Allain, 1976, Gillespie, 2000). Nous avons montré que, pour les systèmes à plusieurs échelles (certaines espèces moléculaires sont en faible nombre et d'autres en très grande nombre, ou certaines réactions sont lentes et d'autres sont rapides) les limites singulières des processus de Markov à sauts conduisent à des modèles réduits [9]. Une classe importante de modèles est constituée par les processus déterministes par morceaux. Ces processus sont obtenus à partir des processus de Markov à sauts, lorsque la limite fluide est partielle, c'est à dire qu'elle s'applique uniquement à ces espèces moléculaires qui existent en grand nombre. La dynamique de ces espèces peut être décrite, avec une bonne approximation, par des équations différentielles déterministes, paramétrées par les nombres de molécules rares. Les nombres de molécules rares suivent de processus de Markov à sauts. A chaque saut (variation du nombre de molécules rares), les paramètres du système dynamique déterministe changent de façon discrète et aléatoire. Nous avons classifié les différentes situations possibles selon la dimension  $n$  de la dynamique déterministe, selon le nombre  $m$  d'attracteurs et selon le nombre d'états discrets  $p$  du système. Ainsi, on peut imaginer que pour un état discret fixé on peut avoir un ou plusieurs attracteurs stables. Ces attracteurs changent lorsqu'on passe d'un état discret à un autre. Nous avons étudié l'ergodicité ainsi que certaines propriétés qualitatives

des trajectoires (loi invariante, temps de retour à niveau, oscillations, intermittence, pics stochastiques) pour un modèle en dimension un et avec deux états discrets ( $n=1, m=1, p=2$ ). En outre, nos résultats ont conduit à un algorithme de simulation accélérée de systèmes moléculaires stochastiques (plus performante que la méthode classique de Gillespie) [9].

### **Modules pour l'étude d'équilibres.**

Les modèles biologiques sont des systèmes d'équations différentielles, polynomiales ou rationnelles, contenant un certain nombre de paramètres inconnus (mais sur lesquels on peut éventuellement agir). Nous avons souhaité obtenir un certain nombre de conditions formelles sur les paramètres, pour que des propriétés biologiquement intéressantes soient satisfaites. Par exemple nous nous sommes concentrés sur l'identification de conditions d'unicité de l'état stationnaire, ou de conditions pour que certaines quantités faisant intervenir les interactions entre les molécules soient toujours positives ou toujours négatives [6].

Pour l'étude des états stationnaires nous utilisons une méthode d'élimination. Notre principale innovation est de décomposer le système en modules, chacun satisfaisant les conditions d'un théorème d'univalence globale (Gale-Nikaido) et possédant un unique équilibre pour toutes valeurs des variables d'entrée [6,10]. Les variables internes à un module sont éliminées simultanément. Cette méthode est analogue à la technique de décomposition en modules monotones, utilisée récemment par Sontag et testée aussi au sein de notre projet [12]. La condition de monotonie imposée aux modules par Sontag est plus forte que la nôtre; ceci constitue l'avantage principal de notre méthode. En contrepartie, nous ne pouvons rien dire sur la stabilité des équilibres. La décomposition en modules monotones est une méthode complémentaire, elle est conçue pour des études de stabilité [12].

Nous avons développé une suite de programme MAPE qui permet de manipuler les conditions d'univalence globale du type Gale-Nikaido (stage de licence, Clément Chatelain, 2006). En outre, ceci permet d'identifier (par des vérifications des signes des mineurs principaux de la matrice jacobienne du système) le plus grand module possédant un unique état stationnaire. La matrice jacobienne intervient aussi dans la définition des élasticités du graphe [1]. Nous utilisons le même type de manipulations formelles pour tester les signes de ces élasticités ainsi que les signes des variations des élasticités suite à un changement du

système (par exemple mutation). Ces techniques ont été appliquées à l'étude qualitative d'un modèle de métabolisme [6].

## **Robustesse et complexité**

Un des objectifs de ce projet est l'identification d'acteurs critiques. Notre étude de la robustesse et de la complexité précise les limites de cette recherche. Nous montrons que le contrôle des propriétés des systèmes biologiques peut être distribué, par opposition au contrôle localisé. Dans le premier cas, il n'y a pas un seul acteur critique. Nous avons ainsi introduit et justifié mathématiquement les notions de robustesse distribuée et de r-robustesse [10,11]. La robustesse distribuée a été associée à des phénomènes de concentration de mesure dans des espaces de grande dimension de paramètres. Nous distinguons deux types de phénomènes de concentration : la concentration du cube correspondant à la loi des grands nombres et la concentration du simplexe correspondant aux propriétés des statistiques d'ordre.

O.Radulescu a montré (avec A.Gorban) qu'il y a des relations entre la complexité topologique du graphe de réactions biochimiques (définie en termes de boucles et ponts dans le réseau) et la robustesse [11]. Ce résultat complète des résultats similaires obtenus par David Rand à Warwick sur la relation entre la complexité et la flexibilité (Rand, Shulgin, Salazar et Millar, JTB 2006).

Dima Grigoryev (avec S.Vakulenko) a obtenu des résultats sur la complexité des motifs spatio-temporels (en morphogenèse) qui peuvent être produits par certaines classes de modèles. Le modèle de type Hopfield peut ainsi engendrer toute dynamique structurellement stable [17-19]. De même, en utilisant la méthode de Khovanski ils ont trouvé des bornes inférieures pour le nombre de gènes de développement nécessaire pour produire une dynamique donnée [17,18]. Puis, ils ont étudié la stabilité de ces modèles par rapport à l'évolution du graphe d'interaction selon quelques schémas d'évolution inspirés des réseaux aléatoires. Ainsi, ils ont montré que l'évolution à complexité constante (telle que dans le schéma de Erdős-Renyi) ne peut pas être stable [20].

## **II. Avancés biologiques et modélisation.**

### **Métabolisme des lipides.**

La motivation première de la modélisation du métabolisme des lipides a été de développer un programme interdisciplinaire qui pourrait contribuer à une meilleure interprétation des données transcriptomiques générées dans le cadre de la thématique de l'unité UMR de Génétique animale. On s'est proposé ainsi d'identifier les quelques gènes directement responsables (de par une variation allélique) de la variabilité d'état d'engraissement entre deux lignées de poulets divergentes pour l'engraissement (de tels gènes sont communément appelé QTL pour Quantitative Trait Loci). Nous avons localisé 5 régions QTL dans des intervalles de 5 à 20cM [8]. La puce développée en 2002/2003 dédiée « foie » de 300 gènes d'expression hépatique, dont 50% étaient identifiés chez les mammifères comme intervenant dans le métabolisme hépatique des lipides, a été hybridées avec 10 mélanges d'ARN de foies de poulets des lignées grasse et maigre. Moins de 10 gènes présentaient des niveaux d'expression différents entre lignées, il s'agit notamment de gènes de la lipogénèse (ACC, FAS, SCD1) ou impliqués dans sa régulation (SREBP-1) [7]. De tels résultats suggèrent que le facteur de transcription LXRA présent dans une région QTL et connu pour trans-activer SREBP-1 soit un des 5 gènes QTL recherchés comme responsable de la variabilité de l'engraissement [7]. La recherche d'une mutation causale par séquençage au niveau de ce gène LXRA est en cours. Ces résultats montrent un nombre très réduit de gènes différentiellement exprimés entre lignées grasse et maigre.

Néanmoins, on s'attend avec l'utilisation en 2007 de puces pangénomiques de 20 000 gènes, à identifier des sous-groupes de gènes plus conséquents dont l'expression est contrôlée par des régions précises du génome et qui sont ainsi la signature de l'effet d'un gène régulateur plus amont présent dans ces régions, régulateur qu'il nous faudra identifier. Aussi, avons nous fixé comme objectif, en concertation avec la modélisation, d'accéder à une compréhension intégrée du métabolisme des lipides et de sa régulation. Les premiers protocoles expérimentaux réalisés dans l'unité UMRGA sont d'explorer la transition du métabolisme des lipides d'un état de lipogénèse (état nourri) à un état de lipolyse (à jeun). L'étude de cette transition nourri-à jeun apparaît particulièrement complexe car elle fait intervenir un facteur temps (mobilisation séquentielle des réserves énergétiques de l'organisme, régulations biochimiques et génétiques ne s'opérant pas aux mêmes échelles de temps); elle fait intervenir

de nombreux acteurs moléculaires (métabolites, cofacteurs, hormones, cytokines, ATP, etc.) et implique des régulations coordonnées au niveau de divers organes (principalement foie, tissu adipeux et muscles). Pour cela, une cinquantaine de poulets mâles de 4 semaines ont été mis dans 6 états de jeûne et de renutrition (6 à 9 poulets par état). Ces états nutritionnels sont les suivants : nourriture ad libitum, jeûne de 16h ou de 48h, jeûne de 16h suivi d'une réalimentation pendant 5h ou 16h, jeûne de 48h suivi d'une réalimentation pendant 24h. Les acides gras du plasma et du foie de ces 50 individus ont été dosés. Par ailleurs, les ARN de foies ont été extraits et hybridés fin 2006 sur puces pangénomiques de 20 000 gènes. Les données transcriptomiques ainsi générées sont en cours d'analyse, en association avec les données d'acides gras.

Parallèlement à la production de données, nous avons élaboré une méthodologie de modélisation. L'originalité de cette méthodologie consiste à proposer deux niveaux de complexité distincts dans la modélisation : 1) un modèle de connaissance (modèle étendu) qui détaille les interactions qualitatives entre tous les acteurs moléculaires connus et représente ces connaissances sous la forme de graphes d'interactions signées et 2) un modèle réduit (modèle abstrait) où des variables du modèle étendu sont regroupées en sous-groupes fonctionnels et qui permet l'application de méthodes mathématiques d'analyse de la dynamique du système (états d'équilibre, réponses aux perturbations, prédictions du comportement de mutants, mise en évidence de composantes chefs d'orchestre, etc.). Ce schéma s'applique tout naturellement aux données disponibles, mais il y a aussi une raison plus fondamentale de l'appliquer à l'analyse des systèmes biologiques en général, notamment elle fournit un cadre d'étude de la robustesse (voir plus loin).

Nous avons appliqué cette méthodologie à la modélisation de la régulation du métabolisme des lipides dans les hépatocytes. Nous avons proposé un modèle abstrait intégratif [6], car il rend compte des processus les plus importants intervenant dans le métabolisme des lipides (glycolyse, synthèse, oxydation, lipolyse, production de corps cétoniques). Le modèle est une abstraction de faible complexité car beaucoup de voies métaboliques ont été modélisées par une seule réaction biochimique. Le modèle est multi-échelle car il rend compte de régulations à vitesses très différentes (métabolique rapide, voies génétiques des récepteurs nucléaires PPAR, LXR plus lentes). Ce modèle a été validé sur un jeu de données transcriptomiques et métaboliques lors de jeun prolongé pendant 72 h chez la souris normale et PPAR mutante (présentation à ICSB Yokohama 2006). Les données souris ont été obtenues par une

collaboration avec l'unité INRA de Toulouse (Pascal Martin). Le même modèle sera appliqué aux données chez le poulet. Les résultats de ce travail de modélisation n'ont pas fourni seulement une identification des paramètres du modèle abstrait. Ils ont aussi répondu à des questions sur l'importance relative des différentes voies métaboliques et phénomènes de transport. Initialement le modèle abstrait considérait uniquement deux types d'acides gras (AG) : les AG monosaturés et saturés qui sont synthétisés de novo par le foie et les AG polyunsaturés qui proviennent de l'alimentation et dont certains agissent sur les récepteurs nucléaires. A jeun, les deux types d'AG sont mobilisés et entrent dans le foie où ils sont oxydés. Dans les souris PPAR mutantes, l'oxydation est réduite et les deux types d'AG doivent s'accumuler dans le foie (produisant entre autres la stéatose du foie). De façon surprenante, nous avons remarqué une différence importante de comportement entre les deux types d'AG (accumulation limitée pour les AG polyunsaturés). Ce phénomène ne peut s'expliquer que par une modulation simultanée de l'oxydation et de la voie de synthèse par la mutation. Ceci nous a amenés à considérer un modèle à trois types d'AG avec une voie synthétique supplémentaire : la voie de synthèse de novo des AG polyunsaturés à longue chaîne, assurée par des désaturases et élongases. L'importance d'une étude plus approfondie du comportement des désaturases a été soulignée aussi lors d'une étude qualitative d'un modèle étendu de la voie de synthèse des AG [1]. De même, l'étude du modèle abstrait a montré la nécessité de l'acquisition de dosages d'AG dans le plasma, pour une meilleure compréhension des phénomènes de mobilisation et de transport. Ces dosages sont en train d'être produits.

### **Signalisation de NFκB**

Notre stratégie sur ce sujet a été d'utiliser ce système en tant que système modèle pour l'application de nos différentes techniques. En effet, la signalisation de NFκB a déjà été modélisée par quelques équipes (groupe de D. Baltimore à Pasadena, groupe de M.R.H. White à Liverpool, etc.) en relation avec les oscillations des concentrations de NFκB et de son inhibiteur. Les modèles utilisés ont une complexité moyenne. Nous avons voulu comprendre ce qu'on peut retirer d'une modélisation plus réaliste des mécanismes de transcription. De même, nous avons voulu identifier les observables et les paramètres importants pour un plan d'expérience. Finalement, nous avons voulu comprendre l'impact des phénomènes stochastiques sur la dynamique.

Nous avons produit un modèle complexe avec un traitement réaliste des mécanismes moléculaires intervenant lors de la transcription, la traduction et la translocation des différents acteurs [14]. Ce modèle a été ensuite réduit successivement en appliquant nos méthodes de réduction de modèle [14]. La réduction a produit une hiérarchie de modèle qui nous a permis a) une comparaison avec les modèles existants b) le choix de plusieurs niveaux de complexité selon les propriétés que nous souhaitons étudier (ce choix est important pour optimiser la confrontation entre modèle et données) c) le choix de paramètres les plus influents sur l'amortissement et la période des oscillations [11]. Nous avons aussi effectué des études sur la robustesse de ce système [11], en relation avec des phénomènes de concentration de mesure dans l'espace de paramètres. Nous avons aussi distingué entre le comportement différent de la période et du temps d'amortissement en fonction de l'accumulation de perturbations dans un modèle numérique, en identifiant une concentration du type cube pour la période et du type simplex pour le temps d'amortissement. Cette différence prédite pourrait avoir des conséquences biologiques notamment sur la variabilité épigénétique. Ainsi, selon nos résultats il y a deux sources de variabilités dans le modèle : la variabilité paramétrique (dont les effets sont limités par les phénomènes de concentration du type cube ou simplex) et le bruit moléculaire (que nous avons modélisé en simulant le processus de Markov à sauts par la méthode de Gillespie).

Pour résumer, ces études nous ont conduits à des avancées théoriques importantes sur la stabilité et la variabilité des systèmes biologiques, à une meilleure compréhension du système NFκB et à un plan d'expérience pour identifier les différentes sources de bruit dans le fonctionnement de la signalisation de NFκB.

### **Signalisation de TGFb**

Notre démarche dans ce domaine a été interdisciplinaire. Elle a été basée sur des échanges continus entre les biologistes (Nathalie Théret de INSERM U620) et les modélisateurs (équipe Symbiose de Rennes).

En modélisation, nous avons produit une carte d'interactions fonctionnelles entre les molécules intervenant dans la signalisation de TGFb en partant de la littérature et des bases de données existantes (TRANSPATH). Cette carte comporte 361 variables biologiques 250 interactions et plus d'une centaine de références bibliographiques. Néanmoins, les variables

concernent uniquement le trafic moléculaire dans le cytoplasme allant jusqu'à quelques facteurs de transcription. La régulation génétique, reliant la signalisation à la physiologie et au phénotype est trop peu documenté pour alimenter le réseau. Nous avons essayé de construire cette partie en recherchant les sites de régulations dans les promoteurs des gènes cibles du TGF- $\beta$ , mais cette approche s'est avérée insuffisante.

Parallèlement à la construction de modèles nous avons cherché à identifier les gènes responsables des pathologies étudiés, notamment le cancer et la fibrose. Nous sommes arrivés à la conclusion que la situation du cancer du foie est très complexe, elle résulte dans la majorité des cas de la complication de pathologie chronique induisant une fibrose et une cirrhose hépatique [22,24]. Dans les tissus, les données transcriptionnelles, issu des puces à ADN, résultent de la contribution de différents types cellulaires avec une forte signature donnée par les cellules étoilées, qui sont activées dans cette pathologie. Ceci a justifié la nécessité d'une étude expérimentale rigoureuse du comportement des cellules étoilées et la mise en place d'un protocole d'une puce destinée à cette étude.

Pour utiliser notre modèle de signalisation, nous avons voulu d'abord définir les spécificités du trafic dans le cas de la fibrose. Nathalie Theret a identifié la molécule ADAM12, une Disintégrine et Métalloprotéase comme protéase régulée positivement par la voie du TGF $\beta$  [23]. Cette protéase intervient dans la signalisation du trafic du TGF $\beta$  (voie des SMAD) en facilitant le trafic des récepteurs vers les endosomes dans leurs phases précoces. Du point de vue de la théorie générale de la réponse des graphes cette boucle de régulation positive pourrait augmenter la sensibilité de la voie [1]. Les données suggèrent que la présence de ADAM12 dans les endosomes est spécifique à la fibrose et pourrait ainsi être une des causes du phénotype fibrotique. Nous sommes en train d'intégrer ce résultat dans un modèle de trafic résultant de notre carte d'interaction et de la modélisation des processus physico-chimiques liés à l'internalisation et les échanges entre les récepteurs.

Ainsi, le résultat principal de la collaboration entre biologistes et modélisateurs a été de cibler la stratégie expérimentale, au départ trop générale, vers l'étude d'une pathologie (la fibrose) et d'un type cellulaire (cellules étoilées). Concernant le travail de modélisation en cours, nous avons pu identifier des acteurs importants pour la modélisation, notamment ADAM12. De même, la réflexion sur le rôle de cette molécule nous a conduit à la considération des effets physico-chimiques, ce qui représente une direction prometteuse de la modélisation du trafic.

## **Morphogenèse et développement**

On s'est proposé d'analyser un problème central de la biologie du développement qui est celui de la stabilité. Les mécanismes moléculaires, soumises à du bruit et à la variabilité, comment engendrent-ils une morphogenèse canalisée et reproductible? Même pour des organismes très bien étudiés tels que la *Drosophile*, on ne sait pas encore répondre à cette question. Cette étude a été rendue possible par nos avancées sur la robustesse des systèmes biologiques et par une collaboration avec le groupe de John Reinitz à Stony Brook.

Pour les premiers stades de développement de la *Drosophile* (blastoderme, avant la gastrulation) nous disposons (grâce à la collaboration avec John Reinitz de Stony Brook) d'un ensemble de données important qui consiste essentiellement en profils d'expression de gènes du développement en fonction du temps et de la position dans l'embryon. Des profils pour des organismes mutants sont également disponibles ou en cours de production.

Un modèle du type Hopfield avec un terme de diffusion [15] peut expliquer l'évolution des profils. Un premier problème : pour rendre compte des observations il faut identifier les paramètres du modèle. Ces paramètres décrivent les interactions entre les gènes du développement. L'identification des paramètres est difficile (les minima de la fonction objective sont en grand nombre et peuvent être dégénérés; actuellement le groupe de Reinitz utilise le recuit simulé et les algorithmes génétiques). On a souhaité d'un part développer des méthodes plus rapides, d'autre part limiter le nombre de solutions en imposant des contraintes.

Un deuxième problème: le modèle est déterministe (il ne prend pas en compte des phénomènes stochastiques) et phénoménologique. Or, nous nous proposons de comprendre les mécanismes moléculaires et stochastiques de la morphogenèse. Une description des aspects moléculaires de la régulation génétique (même grossière en absence de connaissances détaillée) est nécessaire.

Un troisième problème: pour ce modèle phénoménologique nous souhaitons comprendre la stabilité et la canalisation de la morphogenèse. Les études numériques demandent beaucoup de temps et sont difficiles à interpréter.

Pour résoudre le premier problème, nous avons proposé une méthode de simplification de modèle, du type variété centrale pour les EDP. Dans cette approche, la dynamique (en dimension infinie) des profils est approchée par la dynamique (en dimension finie) d'un système d'interfaces en interaction. Les bases d'une telle approche sont décrites dans [15]. Une première application a été présentée dans [16].

Le deuxième problème demande un effort de modélisation important. Ceci a été commencé dans le groupe de Reinitz, par des descriptions assez détaillées des mécanismes de transcription. Notre contribution a porté sur la modélisation des aspects stochastiques de la morphogenèse.

L'obtention d'un modèle moléculaire en dimension très grande et la simplification de modèle représente une stratégie d'étude de la stabilité (robustesse) que nous avons développée dans [10],[11]. Selon un vocabulaire que nous avons proposé dans [10],[11] il y a deux types de robustesse : la robustesse distribuée et la  $r$ -robustesse. La  $r$ -robustesse signifie la stabilité par rapport à un nombre limité de modifications, inférieur à un entier  $r$  (par exemple moins de  $r$  mutations). On parle de robustesse distribuée lorsque le nombre de modifications n'est pas borné, mais les modifications sont aléatoires et indépendantes. La  $r$ -robustesse est en général la conséquence de la redondance du contrôle de la morphogenèse. Des expériences récentes (Manu, groupe de Reinitz) confortées par nos calculs suggèrent que ceci pourrait être le cas du contrôle du domaine antérieur d'expression du gène *hunchback* chez la Drosophile (double redondance du contrôle par *krüppel* et *knirps*, donc 1-robustesse). La robustesse distribuée résulte de l'intégration d'un nombre grand d'effets. Ca pourrait être le cas de la deuxième bande du gène *eve* chez la Drosophile, dont le module MSE2 de régulation transcriptionnelle contient plus d'une vingtaine de sites de fixation pour *bicoid*, *hunchback*, *giant* et *krüppel*. Ces sites ne sont pas conservés d'une espèce à l'autre, pourtant ils sont responsables pour l'expression très précise de *eve*. Notre étude de la robustesse distribuée relie la structure des hiérarchies de modèles obtenus un de l'autre par projections (réductions de modèle) et les phénomènes de concentration de mesure dans des espaces de grande dimension [10], [11]. Deux niveaux de complexité sont privilégiés dans la hiérarchie : le modèle moléculaire le plus complexe et le modèle réduit le plus simple qui rend encore compte des caractéristiques étudiées de la morphogenèse (qui pourrait être le modèle du type Hopfield, ou encore un modèle d'interfaces en interaction obtenu à partir de celui-ci). La morphogenèse est d'autant

plus robuste que la différence de complexité entre le modèle moléculaire et le modèle réduit est grande.

### **2.3 Collaborations internationales**

L'activité de MathResoGen a bénéficié de plusieurs collaborations internationales. Certaines de ces collaborations sont bien établies et ont déjà donné lieu à des publications communes.

La collaboration avec Alexander Gorban, professeur à l'Université de Leicester en Angleterre sur le développement des notions de robustesse pour les réseaux biologique, ainsi que sur les méthodes de réduction de modèle, vient d'être renforcée par l'acceptation d'un projet dans le programme d'actions intégrées Alliance/Egide pour l'année 2007 (prolongeable jusqu'à deux fois) entre O.Radulescu et A.Gorban. L'objectif de ce projet est le développement de méthodes de réduction de modèle. Le doctorant INRA (voir 2.5) Pierre Blavy participera à ce programme d'échanges.

Une autre collaboration bien établie est celle avec Sergei Vakulenko, directeur de recherche à St.Petersburg, sur le sujet de la robustesse des réseaux génétiques, notamment en relation avec la morphogenèse de la Drosophile. Cette collaboration implique deux participants de MathResoGen, O.Radulescu et D.Grigoryev.

Parmi les collaborations bien établies il faut aussi mentionner celle avec le Laboratoire de Bioinformatique et Mathématiques du Génome (L.B.M.G.) (Centro de Modelamiento Matematico- Université du Chili à Santiago): E. Pécou est chercheur associé du LBMG :

- dans un projet pluridisciplinaire, en partenariat avec l'entreprise de biotechnologie BIOSIGMA S.A., pour l'étude, la modélisation et le contrôle de la régulation génétique du métabolisme des bactéries acidophiles intervenant dans la dégradation naturelle des minerais impliquant l'extraction de cuivre (procédé de biolixiviation). Pendant l'année de délégation au Chili (2005) de E. Pécou, A. Siegel et O. Radulescu ont tour à tour visité le LBMG (1 mois pour A. Siegel, 1 semaine pour O. Radulescu), pour échanger expériences et compétences.
- dans un projet académique (projet FONDEF) de développement d'une compétence en biologie systémique au Chili (organisation d'écoles, formation doctorale, etc).

La thèse d'un étudiant chilien, Guillermo Espinoza, commencera en janvier 2007. La direction sera en co-tutelle par Elisabeth Pécou et Alexandro Maass (Chili).

A Stony Brook, nous avons maintenu des multiples contacts avec John Reinitz sur le sujet de la robustesse et de la canalisation du développement. Cette collaboration s'est concrétisée dans un projet commun (soumis par O.Radulescu à HFSP) sur la robustesse, incluant aussi le groupe de Maria Samsonova de l'Université Polytechnique d'Etat de St.Petersburg comme partenaire.

O.Radulescu collabore aussi avec Francisco Lopez de Stony Brook et avec David Holloway de British Columbia sur des modèles stochastiques de la morphogénese.

Une autre collaboration avec Upi Bhalla de NCBS, Bangalore, concerne la réduction de modèles. Dans le cadre de cette collaboration, un stagiaire indien, Debasis Panda, commencera un internship de six mois à Rennes (sous l'encadrement de O. Radulescu). Un projet joint sera proposé avant fin mars 2007 à CEFIPRA pour une collaboration entre Rennes et Bangalore sur ce thème. Le stage de Debasis Panda devrait se poursuivre par une thèse en co-tutelle (O.Radulescu et U.Bhalla).

## **2.4 Publications par chercheurs impliqués**

Vu les collaborations multiples entre les chercheurs impliqués dans le projet, nous optons pour une liste de publications globale, **des publications liées au projet**, classées par sujets.

### **Equations qualitatives pour les réseaux moléculaires :**

[1] O. Radulescu, S. Lagarrigue, A. Siegel, , M. Le Borgne, P. Veber, Topology and static response of interaction networks in molecular biology, Royal Society Interfaces, 3 (2006) 185-196.

[2] A. Siegel, O. Radulescu, M. Le Borgne, P. Veber, J. Ouy, S. Lagarrigue, Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks, Biosystems 84 (2006) 153-174.

[3] Ph.Veber, M. Le Borgne, A. Siegel, O. Radulescu, Complex Qualitative Models in Biology: a new approach. *Complexus* 2004-05, 2, 140-151.

[4] C Guziolowski, P Veber, M Le Borgne, O Radulescu, and A Siegel (2006). Checking Consistency Between Expression Data and Large Scale Regulatory Networks: A Case Study, présenté à la conférence Réseaux d'interaction : analyse, modélisation et simulation. RIAMS'06, Lyon, France, soumis pour publication dans *Journal of Biological Physics and Chemistry*.

[5] A. Siegel, C. Guziolowski-Vargas, P. Veber, O. Radulescu, M. Le Borgne, Optimiser un plan d'expérience à partir de modèles qualitatifs, *Biofutur*, sous presse.

### **Métabolisme des lipides :**

[6] O. Radulescu, A. Siegel, E. Pecou, S. Lagarrigue, A model for regulated fatty acid metabolism in liver; equilibria and their changes, *arXiv: q-bio.MN/0603021*.

[7] Bourneuf E., Hérault F., Chicault C., Carré W., Assaf S., Monnier A., Mottier S., Lagarrigue S., Douaire M., Mosser J, Diot C. 2006. Microarray analysis of differential gene expression in the liver of lean and fat chickens. *Gene* 372, 162-170.

[8] Lagarrigue S., Pitel F., Carré W., Abasht B., Le Roy P., Neau A., Amigues Y., Sourdioux M., Simon J., Cogburn L., Aggrey S., Leclercq B., Vignal A, Douaire M. 2006. Mapping quantitative trait loci affecting fatness and breast muscle weight in meat-type chicken lines divergently selected for abdominal fatness. *Genet. Sel. Evol.* 38, 85-97.

### **Fonctionnement stochastique de réseaux :**

[9] O. Radulescu, A. Muller, A. Crudu, Théorèmes limites pour des processus de Markov a sauts. Synthèse des résultats et applications en biologie moléculaire, sous presse *Technique et Science Informatiques* 2006.

### **Modularité, hiérarchies, robustesse :**

[10] O. Radulescu, A. Gorban, S. Vakulenko, and A. Zinovyev. Hierarchies and modules in complex biological systems. *proceedings of ECCS'06*, 2006.

[11] A.N. Gorban and O. Radulescu, Dynamical robustness of biological networks with hierarchical distribution of time scales, soumis, *arXiv: q-bio.MN/*.

[12] I.Ndiaye, J.L.Gouzé, F.Grognard, Etude d'un petit réseau métabolico-génétique par des techniques de systèmes monotones, présenté à RIAMS'06.

### **Réduction de modèle :**

[13] E. Pécou, Splitting the dynamics of large biochemical interaction networks, *Journal of Theoretical Biology*, 232, (2005),n°3, pp 375-384.

[14] O.Radulescu, A.Zinovyev, A.Lilienbaum, Hierarchical model reduction and model comparison for NFκB signaling.

### **Réseaux et robustesse de la morphogenèse :**

[15] O. Radulescu and S. Vakulenko. Diffusion and interfaces in pattern formation, soumis. *arXiv: q-bio.MN/0603023*.

[16] S.A.Vakulenko, O.Radulescu, V.V.Gursky, K.N.Kozlov, An effective algorithm of parameter fitting in gene circuits, NanoBio 2006, St.Petersburg.

[17] S.Vakulenko, D.Grigoriev, Algorithms and complexity in biological pattern formation problem. *Ann. Pure Appl. Logic*, 2006, vol. 141, p. 412--428

[18] D.Grigoriev, S.Vakulenko, Complexity of gene circuits, Pfaffian functions and morphogenesis problem. - *Comptes-Rendus Acad.Sci.*, 2003, ser. 1, vol. 337, issue 11, p. 721-724.

[19] Vakulenko S, Grigoriev D., Stable growth of complex systems, *Proceeding of Fifth Workshop on Simulation, Saint Petersburg (2005)* 705-709.

[20] S.Vakulenko, D.Grigoriev, Evolution in random environment and structural instability, *Zapiski seminarov POMI RAN*, V. 325, (2005), p.28 -60, traduit par Springer.

[21] D.Grigoriev, A.Grigorieva, Algorithmic aspects of genetic sequences and relative Kolmogorov complexity. - *Intern. J.Pure Appl.Math.*, 2004, vol.11,3,p.283-292.

### **Signalisation TGFb :**

[22] Zindy P., Andrieux L., Bonnier D., Musso O., Langouet S., Campion J.P., Turlin B., Clement B. and Theret N. (2005), Gene ontology analysis of up-regulated gene clusters in cirrhosis reveals relationships between necroinflammatory activity and DNA repair. *FEBS Let.* 3;579:95-9.

[23] Le Pabic H, L'Helgoualc'h A, Coutant A, Wewer UW, Baffet G, Clément B, Théret N (2005) Involvement of the serine/threonine p70 S6 kinase in TGF- $\beta$ 1-induced ADAM12 expression in activated human hepatic stellate cells. *J Hepatol.* 43(6):1038-44.

[24] Zindy P, L'Helgoualc'h A, Bonnier D, Boitin K, Musso O, Zhang CH, Glaise D, Troadec MB, Loréal O, Turlin B, Léger J, Clément B And Théret. N (2006) Up-regulation of the tumour suppressor gene menin in hepatocellular carcinomas and its significance in fibrogenesis. *Hepatology*, 2006, 44:1296-1307.

### **Général, biologie systémique :**

[25] E. Pécou, Mathematical comments on basic topics in Systems Biology, (2005), à paraître aux Actes de l'Ecole CIMPA-UNESCO (Valdivia 2004). "*Mathematical and Computational Methods in Biology*".

Les résultats du projet ont été diffusés aussi par un nombre important de séminaires et communications à des conférences, colloques, écoles.

### **Séminaires, communications, par participant :**

#### **D. Grigoryev**

- Stable growth of complex systems, communication au Fifth Workshop on Simulation, St. Petersburg (2005).
- Communication présentée par S. Vakulenko à "Annual international conference: Days on Diffraction", St. Petersburg (juin-juillet 2005)

#### **C. Guziolowski**

- C. Guziolowski, P. Veber, M. Le Borgne, O Radulescu, and A Siegel, Checking Consistency Between Expression Data and Large Scale Regulatory Networks: A Case Study. Réseaux d'interaction : analyse, modélisation et simulation. RIAMS'06 (novembre 2006).

## **S. Lagarrigue**

- Lagarrigue S., Pitel F., Carré W., Le Roy P., Neau A., Amigues Y., Simon J., Vignal A., Leclercq B., Cogburn L., Douaire M. 2003. An initial QTL scan for abdominal fatness and breast muscle weight in broiler chickens. XI Plant and Animal Genome Conference, San Diego, Californie, Etats-Unis d'Amérique, 11-15 janvier 2003 (communication affichée).
- Bourneuf E., Hérault F., Retout E., Lagarrigue S., Douaire M. et Diot C. 2003. Approche transcriptomique de l'engraissement chez le poulet de chair. 5èmes journées de la Recherche Avicole, 26-27 mars 2003, Tours, p. 407-410 (communication affichée).
- Lagarrigue S., Pitel F., Carré W., Le Roy P., Neau A., Amigues Y., Vignal A., Cogburn L., Leclercq B. et Douaire M. 2003. Identification de QTL pour l'état d'engraissement et le poids de filet chez le poulet de chair. 5èmes journées de la Recherche Avicole, 26-27 mars 2003, Tours, p. 371-373 (communication orale).
- Abasht B., Pitel F., Lagarrigue S., Le Bihan-Duval E., Leroy P., Hérault F., Vignal A., Simon J., Cogburn L., Aggrey S., Douaire M. 2004. Characterisation of quantitative trait loci affecting fatness on chicken chromosome 5. XXII World's Poultry Congress, Istanbul, Turquie, 8-13 juin (communication orale).
- Abasht B., Lagarrigue S., Pitel F., Le Bihan-Duval E., Le Roy P., Simon J., Cogburn L., Aggrey S., Vignal A., Douaire M. 2005. Sex by fatness QTL interaction with transgressive variation in fat and lean lines. Gordon conference on quantitative genetics and genomics, Ventura, Californie, Etats-Unis d'Amérique 20-25 février 2005 (communication affichée, résumé).

## **E. Pécou**

- Novembre 2006 : Conférence « Theoretical Approaches of the Genome », Annecy
- Mars 2006 : Conférence Internationale de Mathématiques Appliquées (ICAM 2006), Santiago, Chili.
- Septembre 2005 : Tercer Taller de Bioinformática de Chile, Santiago.
- Mai 2005 : Rencontre du CIRM « Modélisation dynamique et analyse des réseaux de régulations biologiques », Luminy.
- Mai 2004 : Laboratoire J.-A. Dieudonné de Mathématiques, Université de Nice-Sophia Antipolis, Nice.

- Avril 2004 : I.P.M.C., (Institut de Pharmacologie Moléculaire et Cellulaire) U.M.R. CNRS 6097, Sophia Antipolis.
- Mars 2004 : Institut Mathématique de Luminy, Université de Provence, Marseille.
- Mars 2004 : Centre de Physique Théorique, Université de Provence, Marseille.
- Février 2004 : Château Gombert, Université de Provence, Marseille.
- Mars 2003 : Séminaire Mathématiques pour le Génome, Unité M.I.G. INRA- Jouy-en-Josas.

## **O. Radulescu**

- Novembre 2004, "Information positionnelle et réaction-diffusion dans l'organisation spatiale", séminaire Mathématiques appliquées à la génomique, Génopole Marseille-Nice.
- Février 2005, séminaire dans le groupe mathématiques appliquées, ENS Paris.
- Mars 2005, séminaire "Diffusion and Morphogenesis: from wormlike micelles to Drosophila" au Centre de Modélisation Mathématique, Université de Leicester, Angleterre.
- Septembre 2005, communication "Robustness and Concentration in Biological Networks", école d'été "Robustness and Noise in Gene Regulatory Networks", Coquelles, France.
- Octobre 2005, communication "Concentration and spectral robustness of biological networks with hierarchical distribution of time scales", ECCS'05, Paris, France.
- Novembre 2005, séminaire "Raisonnement qualitatif sur réseaux: applications en génomique fonctionnelle" à l'Institut Curie, Paris.
- Novembre-Décembre 2005, 2 séminaires au Centre de Modélisation Mathématique, Université de Chili.
- Avril 2006, communication "Hasard et robustesse en biologie moléculaire" à la conférence annuelle de la Société Roumaine de Probabilités et Statistiques, Bucarest, Roumanie.
- Juin 2006, communication "New qualitative approaches in molecular biology", Third Indo-French Bioinformatics Meeting (IFBM 2006), Bangalore, Inde.
- Juillet 2006 : séminaire au Center for Developmental Genetics, Stony Brook, Etats Unis.

- Septembre 2006, communication "Hierarchical models and robustness of complex biological models", European Conference on Complex Systems, ECCS'06, Oxford, Angleterre.
- Octobre 2006 : poster "Qualitative, Integrated Model of Lipid Metabolism in Liver", Conférence Internationale en Biologie Systémique, Yokohama, séminaire dans le groupe Kaneko à l'Université de Tokyo, séminaire au centre de bioinformatique de l'Université de Kyoto.
- Novembre 2006 : communication "An effective algorithm of parameter fitting in gene circuits", présentée par Vitaly Gursky à NanoBio 2006, St.Petersburg, Russie.

### **A. Siegel**

- S. Lagarrigue et A. Siegel, Modélisation in silico de la régulation génétique du métabolisme des lipides: Problèmes et méthodes, Troisième carrefour OUEST-genopole, Brest (janvier 2006).
- A. Siegel, Etude des déplacements d'équilibre d'un modèle différentiel par décomposition en modules, réunion ACI VicAnne, Paris, IHP, Abstraction, modularity, and compositionality of genetic and protein interaction networks, Paris, IHP (février 2006).
- S. Lagarrigue et A. Siegel, Modélisations (étendue et abstraite) de la régulation génétique du métabolisme des lipides , Journée Inra-Inria, Lyon (décembre 2006).

### **N. Théret**

- Zindy P-J, Andrieux L., Bonnier D., Musso O., Langouet S., Champion Jp, Turlin B., Clement B , Théret N. (2004). Differential gene expression in underlying livers of hepatocellular carcinomas : increased DNA Repair network in cirrhosis. American Association of Cancer Research Meeting (AACR) 27-31 mars 2004.
- Le Bécheuc A, Zindy P, Bihouée A, Léger J, Clément B, Théret N (2004) "M@LIMS & M@IA: Microarray Experiments Management & Data Mining", à JOBIM 2004 : 5èmes Journées Ouvertes Biologie Informatique Mathématiques, 28-30 juin 2004, Montréal, Canada.
- Théret N , Zindy P-J, Andrieux L., Bonnier D., Musso O., Langouet S., Champion Jp, Turlin B., Clement B (2005) Identification of a new set of genes involved in DNA repair and damage in underlying livers of hepatocellular carcinomas: a combination of

gene expression profiling and gene ontology bioinformatic tool HUGO's 10th Human Genome Meeting Kyoto, Japan -18-21 April 2005 .

- Le Béhec A, Mary V, Zindy P, Cuinet N, Touzeau N, Clément B, Théret N (2005) Meta-analysis of differentially expressed genes during liver injury by using a Cross Microarray Synthesis tool Integrative Post Genomic Lyon 1- 2 dec 2000.
- Zindy P, L'Helgoualc'h A, Bonnier D, Boitin K, Musso O, Zhang CH, Glaise D, Troadec MB, Loréal O, Turlin B, Léger J, Clément B And Théret. N (2006) Up-regulation of the tumour suppressor gene menin in hepatocellular carcinomas and its significance in fibrogenesis. 57th Annual Meeting of the AASLD Boston, Massachusetts, October 27 - 31, 2006.
- Bourd-boittin K, Basset L, Samson M, L'helgoualc'h A, Bonnier D, Clement B, Théret N. (2006) Involvement of activated hepatic stellate cells in expression and activation of CX3CL1 (fractalkine) in liver injury. 57th Annual Meeting of the AASLD Boston, Massachusetts, October 27 - 31, 2006.

#### **P. Veber**

- Décembre 2006, Vérifier la consistance entre données d'expression et réseaux d'interactions, séminaire au Laboratoire de Biométrie et Biologie Evolutive, Lyon.
- Octobre 2005, Ph.Veber, M. Le Borgne, A. Siegel, O. Radulescu, Complex Qualitative Models in Biology: a new approach, ECCS'O5, Paris, France.

#### **2.5 Autres résultats : structuration, création d'équipe, nouvelles collaborations, thèses....**

HDR de Elisabeth Pécou (Octobre 2005, l'Institut de Mathématiques de Bourgogne)

« *Exemples d'applications des systèmes dynamiques : topologie des variétés de dimension 3 et mécanismes de régulations biomoléculaires.* »

Septembre 2006 : Nomination de Elisabeth Pécou en tant que professeure au Laboratoire de Mathématiques J.-A. Dieudonné de l'université de Nice – Sophia Antipolis (UNSA). Responsable d'un projet de structuration et de développement d'une collaboration entre mathématiciens et biologistes de l'UNSA à travers la demande d'un PPF (Projet pluri-formation). Mise en place d'une nouvelle collaboration scientifique avec l'équipe de J.

Pouysségur de l'Institut de Signalisation, Biologie du Développement et du Cancer (UMR CNRS 6543) pour l'étude de la régulation génétique de la voie hypoxique.

HDR de Ovidiu Radulescu (Décembre 2006, Rennes)

«Modèles mathématiques de la complexité en biologie moléculaire et en mécanique des fluides».

Projet d'activité transversale "Mathématiques et modélisation pour la biologie moléculaire" au sein de l'IRMAR. L'activité transversale proposée à IRMAR est possible grâce à des compétences développées au cours des quatre dernières années par Dima Grigoriev, Dimitri Petritis et Ovidiu Radulescu, notamment dans le cadre du projet MathResoGen. Elle est basée sur un réseau de collaborations interdisciplinaires rennaises, nationales et internationales. Elle fonctionnera par la participation de personnes appartenant à des équipes différentes de IRMAR et se propose la résolution d'un certain nombre de problèmes transversaux à l'algèbre, à l'analyse (EDP, systèmes dynamiques), à la théorie des probabilités et à la modélisation. Le projet d'activité transversale sera autour de deux grands problèmes biologiques (1) la régulation génétique et la signalisation en relation avec la physiologie et la pathologie 2) la morphogenèse et le développement) et trois classes de méthodes mathématiques (1) Réduction de modèles différentiels 2) Modélisation stochastique 3) Méthodes algébriques: étude de la complexité, analyse formelle, équations qualitatives).

Deux thèses (Thomas Sierocinski et Alina Crudu) ont démarré en 2005 à IRMAR sur ces sujets. Une ATER (Aurelie Müller) recrutée en 2006 par ENS Cachan est associée aux projets de modélisation. Dans le cadre d'un projet ANR accepté pour 2006-2009 nous allons accueillir un post-doc. D'autres collègues de IRMAR s'intéressent à nos activités et souhaiteraient y participer.

Le projet a été soumis à la direction de IRMAR et sera présenté dans le conseil de laboratoire en mars 2007.

Une autre conséquence structurante du projet est représentée par la consolidation de l'axe "systèmes et modèles biologiques" au sein de l'équipe Symbiose de l'IRISA Rennes. Cette activité se propose l'étude des réseaux biologiques, par l'analyse de données diverses et par la modélisation. Cet axe est représenté par A.Siegel et M.LeBorgne, ainsi que par O.Radulescu, temporairement en délégation INRIA au sein du projet Symbiose. Deux thèses (Philippe Veber et Carito Guziolowski) sont en cours dans l'équipe sur des sujets issus de

MathResoGen. Une troisième thèse (Pierre Blavy) vient de commencer sur la modélisation du métabolisme des lipides, financée par l'INRA.

Les résultats de MathResoGen ont ouvert des nouvelles collaborations. En plus des collaborations internationales mentionnées au 2.3, on peut citer l'extension du réseau de collaborations nationale dans deux directions importantes.

Collaboration avec l'Institut Curie.

En 2005, nous avons commencé à travailler avec Andrei Zinovyev (membre de l'équipe bioinformatique de Curie) sur un des thèmes de MathResoGen, notamment la réduction de modèles. Suite à des prises de contacts multiples nous avons défini un projet commun sur la biologie systémique du cancer avec le service de bioinformatique et l'équipe d'Olivier Delattre. Il s'agit de la modélisation de la signalisation induite par la molécule chimérique EWS/FLI dans le sarcome d'EWING. Le projet, accepté par l'ANR Biologie Systémique, vient de commencer en décembre 2006. Les partenaires de ce projet sont l'Institut Curie, IRMAR et IRISA Rennes.

Collaboration avec l'INRA.

L'année 2006 a marqué l'extension de notre collaboration avec l'INRA. A partir de résultats que nous avons obtenus dans la modélisation du métabolisme des lipides, nous avons proposé un programme de recherche d'une plus large envergure. Il s'agit de partir du modèle abstrait de métabolisme de lipides par une hépatocyte proposé dans le projet MathResoGen et de lui ajouter de niveaux de complexité supplémentaires : biochimique et génétique en allant vers un modèle étendu et physiologique en allant vers un modèle multi-organes. De même, nous sommes intéressés par les différences entre les espèces animales. Une telle modélisation multi-organes et multi-espèces du métabolisme des lipides, ne serait pas possible sans le support de plusieurs laboratoires expérimentaux. Plusieurs laboratoires de INRA ont été attirés par ce projet: le Laboratoire de Pharmacologie et Toxicologie (UR66) de Toulouse du département INRA Santé Animale, puis le laboratoire SENAH de St.Gilles du département INRA PHASE (et se sont ajoutés à l'UMR de Génétique animale de Rennes du département INRA de génétique animale. Un projet commun a permis le recrutement d'un thésard de type « Attaché Scientifique Contractuel de l'INRA » (contrat de 3 ans renouvelable une fois) (Pierre Blavy) dans le cadre d'un appel à sujets de thèse en Biologie intégrative (AgroBI) La

collaboration avec le Laboratoire de Pharmacologie et Toxicologie a permis l'obtention des données nécessaires pour la validation de notre modèle de hépatocyte. Les résultats de cette collaboration ont été présentés à la conférence internationale en biologie systémique de Yokohama.

On peut également mentionner la participation de plusieurs membres de MathResGen (M. Le Borgne, O.Radulescu, A.Siegel) à l'ARC MOCA 2006-2007 (Modularité, Compositionnalité et Abstraction dans les réseaux géniques et protéiques) coordonnée par le projet CONTRAINTES de INRIA Rocquencourt (Sylvain Soliman). Le thème central de cet ARC rejoint les lignes du projet MathResoGen. En effet, nous avons développé de techniques mathématiques de réduction de modèle, hiérarchiques et modulaires. Ces méthodes peuvent s'appliquer aux modèles de cycle cellulaire étudiés dans le cadre de l'ARC.

## **2.6 Valorisation : colloques, formation, expositions....**

Les membres de MathResoGen ont contribué à l'organisation de colloques, ainsi qu'à la popularisation des méthodes du nouveau domaine de la biologie systémique parmi différentes communautés scientifiques en France.

On peut mentionner quelques événements, organisés avec l'aide ou par des membres de MathResoGen :

- Janvier 2004 (29-30) : Groupe de travail « Méthodes pour l'analyse des réseaux de régulations génétiques et métaboliques » - Dijon (France), organisé par Elisabeth Pécou, MathResoGen a contribué à son financement.
- Janvier 2005, Atelier à Nice sur les aspects stochastiques de la modélisation des réseaux de régulation. Organisé par J.A. Sepulchre et A. Siegel. O. Radulescu a animé un groupe de travail de cette réunion.
- Janvier 2004 (5-16) : Ecole d'été CIMPA-UNESCO «Mathematical and Computational Methods in Biology » - Valdivia (Chili), a été organisée avec la participation d'un membre de MathResoGen, E.Pécou. Un étudiant chilien, Guillermo Espinoza, qui a suivi les cours de cette école va commencer sa thèse en co-tutelle avec E.Pécou et A. Maass en janvier 2007.
- Les membres du projet symbiose participent à la gestion de l'ACI VicAnne dont le but est de proposer des réunions pour la communauté française de biologie systémique. La

prochaine réunion est organisée à Rennes les 1<sup>er</sup> et 2 février 2007 par Ph. Veber et A. Siegel.

Une importance particulière a été accordée à la sensibilisation de la communauté de biologistes à l'approche de la modélisation, en particulier dans le cadre de la génopole Ouest. Ainsi, des membres de MathResoGen (Anne Siegel et Sandrine Lagarrigue) ont intervenu pour présenter l'activité de modélisation et son impact sur la recherche biologique dans les forums suivants :

- Troisième carrefour OUEST-genopole, Brest (janvier 2006).
- Journée INRA-INRIA, Lyon (décembre 2006).

## **Annexe : Résumés des papiers les plus significatifs**

### **[1] Topology and static response of interaction networks in molecular biology**

O.Radulescu et al

We introduce a mathematical framework describing static response of networks occurring in molecular biology. This formalism has many similarities with the Laplace-Kirchhoff equations for electrical networks. We introduce the concept of graph boundary and we show how the response of the biological networks to external perturbations can be related to the Dirichlet or Neumann problems for the corresponding equations on the interaction graph. Solutions to these two problems are given in terms of path moduli (measuring path rigidity with respect to the propagation of interaction along the graph). Path moduli are related to loop products in the interaction graph via generalized Mason-Coates formulae. We apply our results to two specific biological examples: the lactose operon and the genetic regulation of lipogenesis. Our applications show consistency with experimental results and in the case of lipogenesis check some hypothesis on the behaviour of hepatic fatty acids on fasting.

### **[2] Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks**

A.Siegel et al.

We introduce a mathematical framework that allows to test the compatibility between differential data and knowledge on genetic and metabolic interactions. Within this framework, a behavioral model is represented by a labeled oriented interaction graph; its predictions can be compared to experimental data. The comparison is qualitative and relies on a system of linear qualitative equations derived from the interaction graph. We show how to partially solve the qualitative system, how to identify incompatibilities between the model and the data, and how to detect competitions in the biological processes that are modeled. This approach can be used for the analysis of transcriptomic, metabolic or proteomic data.

### **[3] Complex Qualitative Models in Biology: a new approach**

Veber et al.

We advocate the use of qualitative models in the analysis of large biological systems. We show how qualitative models are linked to theoretical differential models and practical

graphical models of biological networks. A new technique for analyzing qualitative models is introduced, which is based on an efficient representation of qualitative systems. As shown through several applications, this representation is a relevant tool for the understanding and testing of large and complex biological networks.

#### **[4] Checking Consistency Between Expression Data and Large Scale Regulatory Networks: A Case Study**

Carito Guziolowski et al.

We proposed in previous articles a qualitative approach to check the compatibility between a model of interactions and gene expression data. The purpose of the present work is to validate this methodology on a real-size setting. We study the response of *E.coli* regulatory network to nutritional stress, and compare it to publicly available DNA microarray experiments. We show how the incompatibilities we found reveal missing interactions in the network, as well as observations in contradiction with available literature.

#### **[6] A model for regulated fatty acid metabolism in liver; equilibria and their changes**

O.Radulescu et al.

We build a model for the hepatic fatty acid metabolism and its metabolic and genetic regulations. The model has two functioning modes: synthesis and oxidation of fatty acids. We provide a sufficient condition (the strong lipolytic condition) for the uniqueness of its equilibrium. Under this condition, modifications of the glucose input produce equilibrium shifts, which are gradual changes from one functioning mode to the other. We also discuss the concentration variations of various metabolites during equilibrium shifts. The model can explain a certain amount of experimental observations, assess the role of poly-unsaturated fatty acids in genetic regulation, and predict the behavior of mutants. The analysis of the model is based on block elimination of variables and uses a modular decomposition of the system dictated by mathematical global univalence conditions.

#### **[7] Microarray analysis of differential gene expression in the liver of lean and fat chickens.**

Bourneuf E, et al *Gene* 372, 162-170.

Unite Mixte de Recherche Genetique Animale, INRA-Agrocampus Rennes, IFR140-GFAS, 65 rue de Saint-Brieuc, CS 84215, 35042 Rennes cedex, France.

Emmanuelle.Bourneuf@gmail.com

Excessive adiposity has become a major drawback in meat-type chicken production. However, few studies were conducted to analyze the liver expression of genes involved in pathways and mechanisms leading to adiposity. A previous study performed by differential display on RNAs extracted from chicken livers from lean and fat lines allowed us to isolate cDNA products of genes with putative differential expression. In this study, a cDNA microarray resource was developed from these products together with cDNAs from genes involved in or related to lipid metabolism. This resource was used to analyze gene expression in the liver from lean and fat chickens. Some genes were found with a difference in expression between lean and fat animals and/or correlated to adipose tissue weight. Cytochrome P450 2C45, thought to play a role in biotransformation of steroids and poly-unsaturated fatty acids, was more expressed in lean chickens whereas fatty acid synthase, stearoyl-CoA desaturase, sterol response element binding factor 1 and hepatocyte nuclear factor 4, respectively involved in lipogenesis and its regulation, were more expressed in fat chickens. These results indicate that mechanisms involved in the expression and regulation of lipogenic genes could play a key role in fatness ontogenesis in chickens from lean and fat lines.

### **[8] Mapping quantitative trait loci affecting fatness and breast muscle weight in meat-type chicken lines divergently selected on abdominal fatness.**

Lagarrigue S, et al , UMR Inra-Agrocampus genetique animale, 35042 Rennes, France.

Quantitative trait loci (QTL) for abdominal fatness and breast muscle weight were investigated in a three-generation design performed by inter-crossing two experimental meat-type chicken lines that were divergently selected on abdominal fatness. A total of 585 F2 male offspring from 5 F1 sires and 38 F1 dams were recorded at 8 weeks of age for live body, abdominal fat and breast muscle weights. One hundred-twenty nine microsatellite markers, evenly located throughout the genome and heterozygous for most of the F1 sires, were used for genotyping the F2 birds. In each sire family, those offspring exhibiting the most extreme values for each trait were genotyped. Multipoint QTL analyses using maximum likelihood methods were performed for abdominal fat and breast muscle weights, which were corrected for the effects of 8-week body weight, dam and hatching group. Isolated markers were

assessed by analyses of variance. Two significant QTL were identified on chromosomes 1 and 5 with effects of about one within-family residual standard deviation. One breast muscle QTL was identified on GGA1 with an effect of 2.0 within-family residual standard deviation.

### **[9] Théorèmes limites pour processus de Markov à sauts ; Synthèse de résultats et applications en biologie moléculaire**

O.Radulescu et al.

Les processus de Markov à sauts du genre systèmes de réactions chimiques permettent la modélisation des phénomènes stochastiques en biologie moléculaire. Néanmoins, il y a peu de résultats mathématiques sur la dynamique de ces processus. Aussi, leur simulation sur ordinateur rencontre des difficultés liées au temps d'exécution. Nous présentons des résultats permettant la réduction de la complexité de la dynamique stochastique. Ces méthodes utilisent des théorèmes limites probabilistes.

### **[10] Hierarchies and modules in complex biological systems**

O.Radulescu et al.

We review several mathematical methods allowing to identify modules and hierarchies with several levels of complexity in biological systems. These methods are based either on the properties of the input-output characteristic of the modules or on global properties of the dynamics such as the distribution of timescales or the stratification of attractors with variable dimension. We also discuss the consequences of the hierarchical structure on the robustness of biological processes. Stratified attractors lead to Waddington's type canalization effects. Successive application of the many to one mapping relating parameters of different levels in an hierarchy of models (analogue to the renormalization operation from statistical mechanics) leads to concentration and robustness of those properties that are common to many levels of complexity. Examples such as the response of the transcription factor NF $\kappa$ B to signalling, and the segmentation patterns in the development of *Drosophila* are used as illustrations of the theoretical ideas.

### **[11] Dynamical robustness of biological networks with hierarchical distribution of time scales**

A.N.Gorban, O.Radulescu

We propose the concepts of distributed robustness and  $r$ -robustness, well adapted to functional genetics. Then we discuss the robustness of the relaxation time using a chemical reaction description of genetic and signalling networks. First, we obtain the following result for linear networks: for large multiscale systems with hierarchical distribution of time scales the variance of the inverse relaxation time (as well as the variance of the stationary rate) is much lower than the variance of the separate constants. Moreover, it can tend to 0 faster than  $1/n$ , where  $n$  is the number of reactions. We argue that similar phenomena are valid in the nonlinear case as well. As a numerical illustration we use a model of signalling network that can be applied to important transcription factors such as NF $\kappa$ B.

### **[13] Splitting the dynamics of large biochemical interaction networks**

E. Pécou

Cet article s'inscrit dans le cadre de l'étude de la dynamique des réseaux d'interactions biochimiques (métaboliques et génétiques) à grande échelle. La caractéristique fréquente de la modélisation dans le cadre de la biochimie est le nombre élevé de paramètres et variables intervenant dans une fonction donnée. C'est pourquoi il est intéressant de mettre au point des méthodes de partage de systèmes permettant une étude simplifiée des sous-systèmes, et surtout en gardant la possibilité de reconstruire la dynamique globale. L'utilisation des échelles de temps en se basant sur la théorie des perturbations singulières est un outil puissant qui est particulièrement adapté dans le cadre des systèmes biologiques qui présentent une hiérarchie temporelle assez marquée. Nous proposons également un autre concept, la théorie de la synchronisation maître-esclave, qui peut être employé pour déceler les partitions variables/paramètres ne permettant pas la reconstruction de la dynamique globale.

### **[15] Diffusion and interfaces in pattern formation**

Ovidiu Radulescu, Sergei Vakulenko

We discuss several qualitative properties of the solutions of reaction-diffusion systems and equations of the form  $u_t = \epsilon^2 D \Delta u + f(u, x, \epsilon t)$ , that are used in

modeling pattern formation. We analyze the diffusion neutral and the diffusion dependent situations that, in the time autonomous case, are distinguished by considering the attractors of the shorted equation  $u_t = f(u,x)$ . We discuss the consequences of being in one or in the other of the two situations and present examples from developmental biology and from fluid mechanics.

**[18] Complexity of gene circuits, Pfaffian functions and the morphogenesis problem.**

D.Grigoryev, S.Vakulenko

(Rapport MathSciNet)

In this paper, the authors announce some results related to the problem of biological morphogenesis, i.e., the emergence of patterns from an initially homogeneous structure. Among the several examples of these phenomena, we can recall the colour pattern of some animals, e.g. butterflies, as the final result of the process, or the cellular differentiation in the early embryo as the ongoing process. In the work under review a class of gene circuits of neural type is considered. They model them as time-discrete dynamical systems, whose dynamical quantities are the spatial profiles of the gene concentrations. The results are the following: given any spatio-temporal pattern for a gene, it is possible to construct a circuit such that one of its genes behaves in the prescribed way. Moreover it is possible to estimate the minimal number of genes of the circuit needed to achieve such a result; in such a way a suitable connection is given between the pattern complexity and the circuit complexity.

**[20] Evolution in random environment and structural instability.**

S.Vakulenko, D.Grigoryev

(Rapport MathSciNet)

Typical systems that stem from applications and exhibit a complicated large time behavior are not structurally stable. The goal of the paper is to consider stability and evolution of complex biological systems, in particular genetic networks. In the presence of fluctuations of an external medium, the problem of homeostasis is tackled with the help of a natural measure of stochastic stability. A generic system with fixed parameters is shown to be unstable. However, stability can be obtained for populations which are capable of changing their parameters. Evolution algorithms that provide stability of populations are not trivial, but the mathematical results are consistent with experimental data on genetic evolution.

**[22] Upregulation of DNA repair genes in active cirrhosis associated with hepatocellular carcinoma.**

Zindy et al.

Phenotypic changes in injured livers involve complex network of genes whose interplays may lead to fibrosis and cirrhosis, a major risk of hepatocellular carcinoma. Gene expression profiles in fibrotic livers were analyzed by using cDNA microarray, hierarchical clustering and gene ontology. Analyses of a major cluster of upregulated genes in cirrhosis identified a new set of genes involved in DNA repair and damage. The upregulation of DNA repair genes was confirmed by real-time quantitative polymerase chain reaction and associated with necroinflammatory activity ( $P < 0.001$ ). Increased DNA repair activity in cirrhosis with inflammatory activity may reflect increased DNA damages as a consequence of chronic liver injury.

**[23] Involvement of the serine/threonine p70S6 kinase in TGF-beta1-induced ADAM12 expression in cultured human hepatic stellate cells.**

Le Pabic et al.

In chronic liver injury, quiescent hepatic stellate cells change into proliferative myofibroblast-like cells, which are a main source of fibrosis. We have recently reported that these cells synthesize ADAM12, a disintegrin and metalloprotease whose expression is up-regulated by TGF-beta1 in liver cancers. Here, we studied the role of the serine/threonine p70S6 kinase (p70S6K) in regulating TGF-beta1-induced ADAM12 expression. RESULTS: The phosphatidylinositol 3-kinase (PI3K) inhibitor LY294002 and the mitogen-activated protein kinase inhibitor, UO126, decreased the TGF-beta1-dependent ADAM12 expression and prevented the phosphorylation of p70S6K. In addition, TGF-beta1-induced ADAM12 up-regulation was blocked by the Frap/mTOR inhibitor rapamycin, which abrogated the phosphorylation of p70S6K. In untreated cells, LY294002 but not rapamycin diminished the basal ADAM12 expression related to inhibition of Akt and the glycogen synthase kinase-3 phosphorylation. CONCLUSIONS: The data suggest that TGF-beta1 induces ADAM12 gene expression through both the PI3K/Frap-mTOR/p70S6K and MEK/ERK pathways. In addition, activation of the PI3 pathway might be involved in the basal ADAM12 expression in cultured hepatic stellate cells. The involvement of PI3K in ADAM12 expression, similar to that previously observed for collagen I and fibronectin, suggests common pathways for gene

up-regulation in hepatic stellate cells that occur during liver fibrogenesis and contribute to tumor progression.

**[24] Upregulation of the tumor suppressor gene menin in hepatocellular carcinomas and its significance in fibrogenesis.**

Zindy et al.

The molecular mechanisms underlying the progression of cirrhosis toward hepatocellular carcinoma were investigated by a combination of DNA microarray analysis and literature data mining. By using a microarray screening of suppression subtractive hybridization cDNA libraries, we first analyzed genes differentially expressed in tumor and nontumor livers with cirrhosis from 15 patients with hepatocellular carcinomas. Seventy-four genes were similarly recovered in tumor (57.8% of differentially expressed genes) and adjacent nontumor tissues (64% of differentially expressed genes) compared with histologically normal livers. Gene ontology analyses revealed that downregulated genes ( $n = 35$ ) were mostly associated with hepatic functions. Upregulated genes ( $n = 39$ ) included both known genes associated with extracellular matrix remodeling, cell communication, metabolism, and post-transcriptional regulation gene (e.g., ZFP36L1), as well as the tumor suppressor gene menin (multiple endocrine neoplasia type 1; MEN1). MEN1 was further identified as an important node of a regulatory network graph that integrated array data with array-independent literature mining. Upregulation of MEN1 in tumor was confirmed in an independent set of samples and associated with tumor size ( $P = .016$ ). In the underlying liver with cirrhosis, increased steady-state MEN1 mRNA levels were correlated with those of collagen alpha2(I) mRNA ( $P < .01$ ). In addition, MEN1 expression was associated with hepatic stellate cell activation during fibrogenesis and involved in transforming growth factor beta (TGF-beta)-dependent collagen alpha2(I) regulation. In conclusion, menin is a key regulator of gene networks that are activated in fibrogenesis associated with hepatocellular carcinoma through the modulation of TGF-beta response.