



DEA Génomique et Informatique

Université de Rennes 1

Rapport de stage

Année 2003-2004

Modélisation des réseaux de signalisation du facteur de croissance TGF- β 1

Vers une conception plus globale de la cellule

Présenté par Minh-Quang VO

Résumé

Dans ce rapport, nous exploitons un outil de visualisation et d'analyse des réseaux métaboliques et de signalisation nommé GARMeN. Nous avons recueillis par lecture d'articles scientifiques, de nombreuses interactions faisant partie de la signalisation de la cytokine TGF- β 1. Ces données ont permis de construire un réseau de signalisation exploitable à l'aide de GARMeN. Ce réseau s'est révélé utile pour mettre en évidence des effets apparemment contradictoires du TGF- β 1, et pouvant potentiellement orienter les recherches des biologistes. Par la suite, nous avons testé l'application de la modélisation SIGNAL/SIGALI, conçue à l'origine pour d'autres systèmes dynamiques en industrie, au modèle relationnel extrait de GARMeN, afin d'obtenir un modèle comportemental tenant compte de l'aspect fonctionnel des interactions. Ce modèle est basé sur une approche logique à temps discret centrée sur les événements, qui est différente de la modélisation logique des systèmes biologiques déjà décrite dans la littérature. Les expérimentations menées ont montré son efficacité de calcul et son utilité pratique, mais il reste encore un travail approfondi pour l'adapter à la biologie.

Encadré par Michel Le Borgne, Anne Siegel et Nathalie Théret
Laboratoire d'accueil : IRISA, projet SYMBIOSE

Table des matières

1. Introduction	1
2. Matériel et méthode	2
2.1 Pourquoi des modèles <i>in silico</i> ?	2
2.1.1 Modèles de la théorie des graphes	2
2.1.2 Modélisation différentielle	2
2.1.3 Modélisation stochastique	3
2.1.4 Modélisation logique	3
2.2 Formalisation du problème biologique	4
2.2.1 Bases fondamentales de la modélisation des processus biologiques	4
2.2.2 Vision mécanistique	4
2.2.3 Vision globale et concept d'« auto-organisation »	5
2.2.4 Modélisation GARMeN	5
2.2.5 Modélisation SIGNAL/SIGALI	6
3. Résultats	9
3.1 Contexte biologique : le TGF- β 1	9
3.1.1 Voie de signalisation du TGF- β impliquant les protéines SMAD	9
3.1.2 Régulation de l'activité des Smads par les voies kinases	11
3.1.3 Autres voies de signalisation du TGF- β	12
3.2 Recueil des données de la littérature et ses limites	12
3.2.1 Medline, une source hétérogène	12
3.2.2 Les bases de données	13
3.2.3 Bilan	14
3.3 Implémentation GARMeN	15
3.3.1 Le langage de description des interactions	15
3.3.2 La carte du réseau d'interactions	16
3.3.3 Exemple d'exploration utile	17
3.4 Expérimentation SIGNAL/SIGALI	18
3.4.1 Abstraction du modèle GARMeN	18
3.4.2 Passage automatique du modèle GARMeN au modèle SIGNAL	18
3.4.3 Obtention du modèle SIGNAL	20
3.4.4 Manipulation du système dynamique polynomial avec SIGALI	20
4. Conclusion	24
Bibliographie	25

Remerciements

Je tiens tout à d'abord à remercier Michel Le Borgne, Anne Siegel (IRISA, équipe SYMBIOSE), Nathalie Théret (INSERM U456) et Antony Le Béhec (ancien du DEA Génomique et Informatique et actuellement thésard sous la responsabilité de Nathalie) qui m'ont très bien encadré durant mon stage et car ils ont toujours fait de leur mieux pour m'aider dans mes tâches que ce soit en informatique ou en biologie.

Je remercie aussi toutes l'équipe de SYMBIOSE et tous les stagiaires de la salle A126 qui ont su avoir mis de la bonne humeur et de la joie de travailler dans une ambiance inoubliable.

1. Introduction

L'achèvement du séquençage du génome humain et l'avancée de son annotation génèrent des quantités d'informations qu'il faut désormais analyser et organiser dans un contexte fonctionnel. Les biologistes moléculaires, manquant de concepts pour relier l'ensemble des résultats dans une théorie cohérente, ont trouvé dans le domaine de l'informatique une métaphore séduisante mais trompeuse, du programme génétique vu comme un programme informatique, annonçant la guérison prochaine de certaines maladies [1]. Or cette idée prend racine dans la conception réductionniste de la formation de l'organisme adulte. Cette vision suggère donc un finalisme apparent au sens aristotélicien du terme. La notion de programme s'est imposé comme moyen de remplacer ce finalisme par une vision mécanistique selon laquelle les ADN étaient structurés et fonctionnaient comme des programmes d'ordinateur.

Avec le passage de la génomique à la post-génomique, la biologie est en train de sortir de cette période du « tout génétique » ou de la théorie finaliste de programme génétique, pour revenir à une conception plus globale de la cellule [2]. De plus, avec la protéomique, la biologie s'oriente vers la modélisation des systèmes d'interactions et entre dans une nouvelle ère de la complexité. Celle-ci peut se définir de différentes manières mais on peut retenir que d'un point de vue dynamique, ce sont les interactions des protéines, qui dépendent elles-mêmes de leur environnement physico-chimique (pH, potentiel d'oxydoréduction, etc...), qui déterminent la fonctionnalité des gènes. On redécouvre ce qui avait été oublié pendant près de trois décennies : les propriétés fonctionnelles des protéines ne sont pas totalement codées dans la structure des gènes.

Mon stage a consisté à établir une carte du réseau d'interaction fonctionnelles entre molécules et gènes de la signalisation du facteur de croissance TGF- β 1 (transforming growth factor- β 1) chez l'homme, et à construire un premier modèle relationnel des interactions. La compréhension des mécanismes de la signalisation du TGF- β 1 dans le cadre de la progression du carcinome hépatocellulaire est particulièrement cruciale. En effet, cette cytokine est impliquée à la fois dans le développement de la cirrhose, risque majeur dans l'apparition du cancer du foie, et dans le développement de la tumeur [3][4].

Nous avons dans un premier temps établi la carte des réseaux d'interaction de la signalisation TGF- β 1. Pour cette tâche, les ressources bibliographiques des bases de données générales telles que Medline et des bases spécialisées dans les réseaux de gènes telles que TRANSPATH[®] ont été utilisées. D'un point de vue fondamental, cette approche va permettre de rendre compte des propriétés particulières de certains ensembles d'interactions, voire de nouveaux modes de régulation. D'une façon générale, ces réseaux d'interactions nous aideront à mieux comprendre la manière dont l'information contenue dans les gènes s'exprime et se combine à d'autres facteurs épigénétiques (cytoplasmiques) pour déterminer ses caractéristiques réelles (le phénotype).

Nous avons ensuite testé la formalisation du modèle à l'aide des outils SIGNAL et SIGNALI développés au sein de l'INRIA de Rennes par les équipes Expresso¹ et Vertecs² [5]. Ce modèle permet de décrire des systèmes dynamiques à temps discret et multi-horloge qui, en biologie, permettrait d'apporter une méthode d'analyse comportementale des modèles de réseaux d'interactions.

1. <http://www.irisa.fr/expresso>

2. <http://www.irisa.fr/vertecs>

2. Matériel et méthode

2.1 Pourquoi des modèles *in silico* ?

La biologie est en passe de franchir le pas vers une analyse plus globale des mécanismes cellulaires. Cependant la plupart des laboratoires ne se focalisent que sur un type cellulaire, n'étudient que certaines protéines ou certains gènes à des stades précis de développement des tissus sains ou pathologiques. Il y a donc une grande nécessité pour les biologistes de travailler avec informaticiens et mathématiciens afin de chercher une méthodologie d'analyse et de combiner de façon cohérente tous les résultats obtenus.

Dans cette partie, je présenterai quelques modélisations formelles envisageables et nous verrons les avantages et les inconvénients de chacune d'entre elles.

2.1.1 Modèles de la théorie des graphes

Cette première approche fait appel à des mathématiques discrètes permettant d'exploiter de façon formelle les graphes et d'en extraire leurs propriétés (connexité, cycles, etc. . .). L'application aux réseaux d'interactions s'est faite naturellement car elle s'applique parfaitement aux représentations schématiques familières des biologistes.

Pour des systèmes simples, une telle représentation sous forme de graphe permet de répondre à des questions biologiques simples telles que : est-ce que la sur-expression ou la perte de fonction d'une protéine peut avoir des répercussions particulières sur les réseaux ? Existe-t-il un sous-graphe plus simple permettant d'expliquer des propriétés topologiques de régulations de gènes (circuit positif ou négatif, etc. . .) ? Dans ces modèles, il est possible de calculer les chemins entre deux sommets du graphe, de détecter des cycles (algorithme d'Euler), de chercher le plus court chemin en terme de nombre de sommets parcourus ou en terme de poids dans les graphes valués (algorithme de Dijkstra et de Bellman). Ces graphes valués permettent notamment la modélisation de la circulation de flux physiques sur un réseau (analyse de matrices stochastiques).

L'application de cette approche est limitée par le manque de dynamique (décrivant la succession des états du système). Ce modèle est cependant minimal et surtout consitue la base des modèles dynamiques.

Il existe par ailleurs des travaux visant à reconstituer des réseaux métaboliques s'appuyant sur ce formalisme à partir des données de réactions biochimiques à l'INRIA de Grenoble (équipe HELIX, thèse de F. Boyer). Le principe est basé sur l'analyse des flux d'atomes que traduisent les réactions. Leur approche est ainsi de rechercher les chemins du graphe métabolique permettant de transférer le maximum d'atomes avec le minimum d'étapes.

2.1.2 Modélisation différentielle

Dans ce type de modèle, les concentrations ou les activités des espèces moléculaires sont généralement représentées par des grandeurs (ou « variables ») réelles positives, susceptibles de varier de manière continue au cours du temps. La variation de ces grandeurs est formalisée par l'écriture d'un système d'équations différentielles couplées permettant une simulation du modèle. Schoeberl *et al.* [6] ont récemment modélisé la voie de signalisation du facteur de croissance, EGF (« epidermal growth factor ») par des équations différentielles ordinaires, prenant compte les espèces moléculaires et leurs interactions. Les interactions constituant le réseau se traduisent alors en systèmes d'équations cinétiques et le modèle permet de prédire des résultats expérimentaux.

Cette approche est dépendante des données quantitatives disponibles, et présuppose que les concentrations des molécules impliquées varient de manière continue et déterministe. Une autre difficulté de cette approche est la paramétrisation du système : s'il est facile de modifier les paramètres du modèle ou les conditions initiales, il est ardu de faire ressortir des propriétés dynamiques du système intéressantes et pertinentes d'un point de vue biologique.

Le plus grand obstacle de ce type d'analyse reste le manque d'outils mathématiques et informatiques pour étudier ces systèmes de grandes tailles. Les simulations faites sur ces systèmes peuvent en effet durer très longtemps sans donner de résultats significatifs sur des propriétés plus structurales comme la robustesse ou les moyens d'influencer leur fonctionnement.

2.1.3 Modélisation stochastique

Ce type de modèle permet de prendre en compte les différentes fluctuations qui peuvent affecter la dynamique du système : les variations temporelles de changement d'état sont considérées comme des transitions affectées d'une probabilité définie. Dans certains cas, il faut faire appel à des méthodes de simulations pour approximer les solutions des « équations maîtresses » de la formalisation rigoureuse. Il est possible d'obtenir un comportement déterministe et/ou non déterministe proche de l'évolution biologique des cellules.

L'expression des gènes peut être analysée comme un processus stochastique car celle-ci peut montrer une hétérogénéité significative alors qu'à l'échelle de la population cellulaire les mesures peuvent être régulières et reproductibles [7]. Ainsi durant l'infection du phage lambda, qui est gouverné par le circuit décisionnel lyse-lysogénie, une seule fraction des phages lysent les cellules (Ptashne 1998). Ces modèles ont été utilisés pour mesurer l'hétérogénéité dans la population des bactéries du sol *Bacillus subtilis* : dans un même environnement de stress, un arsenal de stratégies probables de survie a été observé [8].

L'inconvénient majeur de cette méthode réside dans sa lourdeur à être mise en oeuvre : les réseaux d'interactions doivent être notamment bien caractérisés expérimentalement, et seuls de petits systèmes peuvent être étudiés avec cette modélisation.

2.1.4 Modélisation logique

Les modèles logiques ou qualitatifs sont utilisés lorsqu'il n'y a pas suffisamment de données quantitatives notamment sur les concentrations des espèces moléculaires. Celles-ci sont alors considérées comme des « variables booléennes » prenant les valeurs 0 (absence) ou 1 (présence). L'évolution temporelle des valeurs des variables est définie par des équations logiques, où la valeur de chaque variable (présence ou absence d'une espèce moléculaire) dépend des valeurs précédentes d'une partie des variables du système. Malgré son aspect caricatural des systèmes biologiques, les modèles logiques permettent de mettre en évidence les caractéristiques les plus marquantes du comportement dynamique des réseaux d'interactions.

Deux approches temporelles peuvent être considérées :

- approche « synchrone » : toutes les variables changent de valeur en même temps,
- approche « asynchrone » : les variables peuvent être indépendantes d'un point de vue temporel. L'approche asynchrone est plus complexe, mais elle est plus adaptée à la biologie.

Il existe plusieurs formalismes des modèles basés sur les événements. Ainsi, une autre approche en développement, utilise des réseaux de Petri à la fois de manière synchrone et

asynchrone [9]. Cet outil mathématique très général permettant de modéliser le comportement d'un système dynamique à événements discrets (Fig. 1). Les places d'un réseau de Petri représentent des conditions logiques, des ressources physiques (ex. : molécules) ou les états des ressources physiques d'un système (ex.: présence ou absence). Les transitions d'un réseau de Petri représentent des événements (ex.: interactions). Un jeton dans une place représente la réalisation d'une condition logique, la charge d'une ressource physique (ex.: niveau d'expression d'une molécule) ou l'état actif d'une ressource physique. La distribution des jetons représente l'état du réseau de Petri. La dynamique d'un système modélisé par un réseau de Petri est donnée par les règles de franchissement des transitions validées du réseau. Une transition a lieu s'il y a un nombre suffisant de jetons dans les places entrantes. L'effet de la transition est de consommer les jetons sur les places entrantes et de produire des jetons sur les places sortantes.

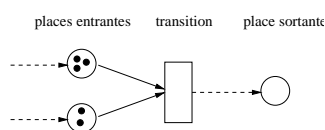


FIG. 1 – *Graphe bipartite de réseau de Petri : la transition est permise si le nombre de jetons (cercle plein) est suffisant dans chaque place*

Au laboratoire de l'IRISA (équipe SYMBIOSE), nous débutons une étude dynamique des réseaux d'interactions moléculaires par la modélisation SIGNAL/ SIGALI, déjà utilisé en programmation de systèmes automatisés et la vérification de leurs propriétés dynamiques. La présentation théorique des modèles développés à l'IRISA fait l'objet de la partie suivante.

2.2 Formalisation du problème biologique

Les systèmes biologiques peuvent être modélisés comme des systèmes dynamiques d'interactions. Pour appréhender les réseaux d'interactions, il est nécessaire d'adopter une vision globale de ces interactions, afin de comprendre comment fonctionnent les processus biologiques.

2.2.1 Bases fondamentales de la modélisation des processus biologiques

Saussure ([10]) est le premier à avoir défini un système biologique comme une unité d'éléments en relation, de façon organisée. L'auto-organisation de ces éléments [11] est un concept clé dans les systèmes complexes. Elle implique l'existence de mécanismes de régulation qui maintiennent la stabilité des systèmes dynamiques. Elle amène aussi à la notion de complexité des processus biologiques selon laquelle, une cellule répond aux stimulus de l'environnement grâce à la présence des éléments moléculaires qui la constituent et à leurs interactions.

2.2.2 Vision mécanistique

Ce paradigme considère les processus biologiques comme une cascade de relations causales. Ces relations entre molécules constituent des processus pouvant être déterministes, mais pas nécessairement. Un processus est une action ou une fonction qui peut se répéter de façon récursive. En d'autres termes, un processus est la modification spatio-temporelle d'un ensemble d'objets identifiables.

Les processus simples peuvent être linéaires ou de type circuit (boucle de rétroaction négative ou positive) (Fig. 2). Ces boucles de rétroaction permettent aux molécules de s'auto-réguler : une molécule A active sur plusieurs étapes un autre molécule B, laquelle régule l'activité de A et ainsi on considère que A est de façon significative son propre activateur ou répresseur. Dans cette vision mécanistique, chaque processus pourrait être décomposé en sous-processus de façon tout à fait arbitraire en fonction de leurs fonction biologique.

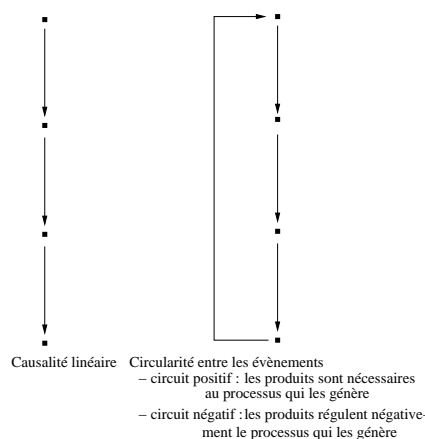


FIG. 2 – *Sous-réseaux d'interactions*

2.2.3 Vision globale et concept d'« auto-organisation »

Les protéines ont des propriétés intrinsèques (activité, structure, etc...) qui dépendent de leur environnement physico-chimique. Ces propriétés indiquent les contraintes sur les interactions entre molécules (localisation, concentrations, etc...) et décrivent une certaine organisation de la « machine cellulaire ». Ainsi, on peut interpréter les phénomènes biologiques comme un scénario d'organisation. Stuart Kauffman [12] dans son livre devenu une référence dans la théorie de l'évolution, relie l'auto-organisation du monde vivant à l'ordre spontané du monde physique. Cette notion est importante car elle défend l'idée que l'évolution darwinienne doit reconnaître d'autres forces, basées sur les contraintes génétiques, les limites structurales et les possibilités des lois physiques générales.

L'« auto-organisation » est l'aptitude des systèmes à créer leurs propres finalités et à pouvoir se structurer eux-mêmes en engendrant leurs propres principes [13]. Cette notion peut se traduire en terme de thermodynamique, mais elle crée un paradoxe du monde vivant : l'auto-organisation est contraire au second principe de la thermodynamique qui veut que tout système physique tende vers le « désordre » (entropie positive). Alors pourquoi les systèmes vivant, eux, se maintiennent dans un état ordonné et fortement improbable (entropie négative)?

Nous n'allons certainement pas résoudre ce paradoxe, mais nous présentons dans les paragraphes suivants, deux types de formalisation visant à simuler l'organisation générale des interactions impliqués dans la signalisation du TGF- β 1.

2.2.4 Modélisation GARMeN

Afin de répertorier et d'analyser les réseaux d'interactions, Michel Le Borgne et Anne Siegel ont développé selon les spécifications des biologistes, l'outil GARMeN (Graphical Analyser for Regulation and Metabolic Network). GARMeN est un outil de visualisation et

d'analyse graphique des réseaux métaboliques et de régulation génique. La description d'un réseau est faite dans un langage déclaratif centré sur les actions :

- catalyses enzymatiques,
- transports de protéines d'un compartiment à un autre,
- réactions d'association/dissociation de complexes,
- expression de gènes.

Le langage est déclaratif au sens où chaque action est décrite indépendamment des autres. La seule contrainte est de regrouper les actions suivant leurs localisations.

A partir de cette description, GARMen génère un modèle du réseau dont la partie affichable est le graphe des flux métaboliques et le graphe des régulations réunis dans un seul graphe orienté.

L'affichage utilise l'outil Grappa³ (Graph Package) qui est un ensemble de méthodes pour construire, manipuler et afficher les graphes dans des applications Java. Les entités biologiques considérées sont les suivantes :

- enzymes,
- molécules de transduction du signal (récepteurs, médiateur, etc...),
- facteurs de transcription,
- et gènes.

Chaque protéine n'est représentée qu'une seule fois et est reliée à toutes les interactions auxquelles elle participe. La protéine possède donc un nom unique comme identifiant dans le graphe. Le programme exécute une séparation automatique entre les différents compartiments déclarés, en affectant à la protéine une marque sur sa localisation. D'un point de vue biologique, la distinction d'une même espèce moléculaire en fonction de sa localisation dans la cellule a un sens car celle-ci influence généralement la fonction des protéines.

Le style du graphe des réseaux d'interactions est neutre par rapport au comportement des molécules (Fig. 3). Le graphe indique seulement le type de l'interaction et toutes les espèces moléculaires impliquées. De plus, les différentes combinaisons possibles entre les molécules ne sont pas indiquées notamment en ce qui concerne les combinaisons entre les facteurs de transcription régulant l'expression des gènes. Des possibilités d'extension du langage, adaptées à différents formalismes sont prévues afin de modéliser la fonction de l'interaction.

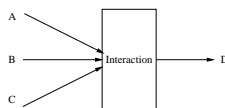


FIG. 3 – Neutralité des interactions. Les entités A, B et C peuvent interagir et donner le produit D. Dans ce modèle, les contraintes d'interactions ne sont pas spécifiées

2.2.5 Modélisation SIGNAL/SIGALI

Le modèle relationnel basé sur les interactions ne permet pas de décrire leur fonctionnalité ou leur comportement. Nous avons donc fait appel à une méthode plus analytique et formelle des interactions entre signaux, le langage SIGNAL utilisé pour décrire les flux de données dans un système : définition des données et la manière dont les signaux interagissent.

3. <http://www.research.att.com/~john/Grappa/>

Ce langage permet de manipuler des objets appelés signaux inspirés de l'électronique. Les traitements des signaux pourront être organisés en sous-systèmes ou sous-processus. Cet outil dispose d'un **compilateur** qui analyse le programme et génère du code utilisable par un **vérificateur** de propriétés du modèle nommé SIGALI. Ce vérificateur est basé sur la manipulation de systèmes d'équations polynomiales dynamiques (PDS) sur $\mathbb{Z}/3\mathbb{Z}$.

L'objectif dans cette partie est de définir les liens possibles entre le modèle GARMeN et le modèle SIGNAL en présentant la formalisation SIGNAL et les possibilités de l'outil SIGALI.

Modèle SIGNAL Le modèle SIGNAL décrit des systèmes à temps discret, c'est-à-dire que les signaux proviennent d'échantillonnages répétés au cours du temps. Un signal est ainsi une séquence de valeurs (de type booléen ou entier, par exemples). Il est toujours associé à une horloge, défini comme l'ensemble des moments (occurrences) où le signal prend une valeur. L'intervalle de temps qui s'écoule entre deux occurrences importe peu.

La composition des signaux s'effectue à l'aide d'opérateurs temporels qui permettent de définir l'ordre dans lequel les signaux doivent prendre une valeur (Fig. 4). Les principaux opérateurs utilisés sont les suivants :

- « when » : extrait des occurrences d'un signal en fonction d'une certaine condition booléenne

$\langle \text{expression signal} \rangle \text{ when } \langle \text{condition} \rangle$.

Ce signal est présent et prend la valeur du membre de gauche quand les deux membres sont présents et la condition est vraie. Au niveau des horloges, cet opérateur réalise une sorte d'intersection des instants de présence.

- « default » : mélange des signaux d'horloge quelconque

$\langle \text{expression signal} \rangle \text{ default } \langle \text{expression signal} \rangle$.

Ce signal prend la valeur du membre de gauche et par défaut celle du membre de droite. Au niveau des horloges, cet opérateur réalise la réunion des instants de présence.

Exemple : considérons les signaux A, B, C et D de type *booléen*.

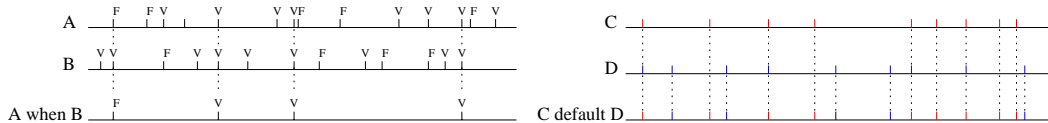


FIG. 4 – Opérateurs temporels sur des signaux booléens

Dans le premier cas, le signal (**A when B**) prend une valeur lorsque les signaux A et B sont présents en même temps, et cette valeur est égale à celle de A. Dans le deuxième cas, le signal (**C default D**) prend une valeur à chaque fois que C ou D en ont une. Mais lorsque C et D apparaissent en même temps, le signal (C default D) prend obligatoirement la valeur de C.

La dynamique est introduite à l'aide de variables appelées **variables d'états**. Comme toute variable, elles ont un type défini (booléen, entier, etc...), mais leur particularité est liée à leur valeur qu'elles continuent à garder tant qu'aucun signal ne permet de la changer. Les variables d'états peuvent avoir plusieurs définitions comme dans l'exemple qui suit.

Exemple : prenons les signaux A et B de type booléen qui influent sur la variable d'états C aussi de type booléen. Voici l'écriture en SIGNAL :

$$C ::= \text{when } A \quad (1)$$

$$C ::= \text{not when } B \quad (2)$$

Dans ce cas, C prend la valeur vrai lorsque A est vrai (1), mais prend la valeur faux lorsque B est vrai et que A est absent à cet instant (2). Implicitement, les deux expressions de C sont **exclusives**. Il n'y a donc qu'un choix possible : dans cet exemple, si A et B ont une occurrence en même temps, alors par défaut, C prend la valeur de A car il est déclaré en premier. Dans ce sens, le langage SIGNAL est déterministe (car il a été conçu pour des circuits électroniques déterministes).

Système dynamique Un système dynamique peut être mathématiquement représenté par un système d'équations polynomiales sur $\mathbb{Z}/3\mathbb{Z}$ (le corps de Galois d'entiers modulo 3) définit ultérieurement, qui a la forme :

$$\begin{cases} Q(G, S) &= 0 \\ G' &= P(G, S) \\ Q_0(G) &= 0 \end{cases}$$

où,

- G est l'ensemble des n variables d'état,
- S est l'ensemble des m variables de type « event »,
- $Q(G, S) = 0$ est l'équation des contraintes entre les variables G et S ,
- $G' = P(G, S)$ est l'équation d'évolution qui définit les règles de transition des variables d'état,
- $Q_0(G) = 0$ est l'équation d'initialisation qui définit l'état initial de chaque variable d'état.

Ce système est généré lors de la compilation du code SIGNAL en code SIGALI. Dans ce système, les valeurs des variables booléennes dont les variables d'états sont codées sur trois valeurs dans le système $\mathbb{Z}/3\mathbb{Z} = \{-1, 0, 1\}$:

$$\begin{cases} \text{présent et vrai} &\Longleftrightarrow +1 \\ \text{présent et faux} &\Longleftrightarrow -1 \\ \text{absent} &\Longleftrightarrow 0 \end{cases}$$

Outil SIGALI SIGALI a été conçu pour la vérification de propriétés des systèmes dynamiques. Les fonctions implémentées dans SIGALI permettent de vérifier les propriétés suivantes :

- le système est-il vivant (« **liveness** »)? il est vivant s'il ne contient pas d'états « puits ». En d'autres termes, dans un système vivant, il existe toujours un ensemble de signaux qui permet de passer d'un ensemble d'états à un autre (Fig. 5).
- un ensemble d'états est-il invariant (« **invariance** »)? Si l'état initial appartient à cet ensemble, l'état appartient-il toujours à cet ensemble indépendamment des événements influençant son évolution?
- un ensemble d'états est-il atteignable (« **reachability** »)? toujours étant données les conditions initiales du système, on pourrait tester l'existence de signaux qui permettent d'atteindre un ensemble d'états donné.
- un ensemble d'états est-il attractif (« **attractivity** »)? selon les conditions initiales, on testerait si un ensemble d'états est la seule évolution possible du système biologique (Fig. 6).

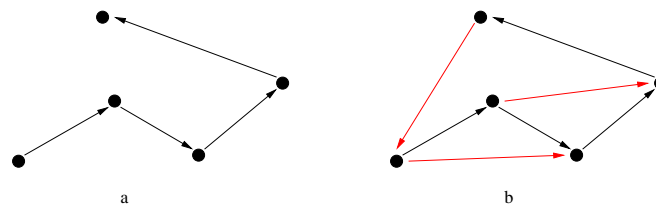


FIG. 5 – Système vivant. Les points représentent les états et les flèches les transitions. Le graphe a représente un système non vivant car au dernier état du chemin il n'y a pas de signaux permettant d'y repartir. Le graphe b est un système vivant chacun de ses éléments

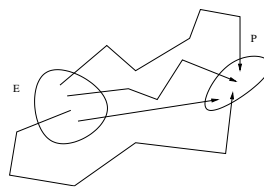


FIG. 6 – Attractivité : l'ensemble P est attractif pour l'ensemble E car tout signal partant de E arrive forcément en P

3. Résultats

3.1 Contexte biologique : le TGF- β 1

Les facteurs de croissance transformants (TGF- β s) forment une super-famille de peptides sécrétés et étroitement apparentés tels que les TGF- β s, les activines, les inhibines, les BMP (Bone Morphogenetic Proteins) et l'hormone anti-müllérienne. Les membres de la famille du TGF- β s sont impliqués dans de nombreuses fonctions biologiques essentielles : l'embryogenèse, la formation de la moelle osseuse, la régulation des systèmes immunitaire et endocrine, ainsi que la cicatrisation [14] [15] [16].

Le facteur de croissance transformant le plus étudié est le TGF- β de type 1. Il s'agit d'un peptide de 25 kDa qui possède de nombreuses fonctions biologiques ; la croissance, la différenciation cellulaire, la régulation de la réponse inflammatoire, la sécrétion de la matrice extracellulaire et l'apoptose [16]. Ces effets diffèrent selon le type cellulaire concerné et selon les conditions physiopathologiques environnementales. C'est notamment un puissant inhibiteur de la prolifération dans de nombreux types cellulaires tels que les hépatocytes, les lymphocytes T et B, les kératinocytes et de nombreux types de cellules épithéliales et endothéliales. Mais le TGF- β 1 peut également être un mitogène capable d'induire la transformation néoplasique de certaines cellules mésenchymateuses et il peut aussi induire ou réprimer l'apoptose selon le type cellulaire ou l'environnement physiologique [17] [14] [16].

3.1.1 Voie de signalisation du TGF- β impliquant les protéines SMAD

Il existe deux types de récepteurs impliqués dans la transduction du message: les récepteurs à activité serine/thréonine kinase de type I et de type II [18] [19].

La signalisation est initiée quand le TGF- β se lie au récepteur II (T β RII) qui possède une activité kinase constitutivement activée. Le Récepteur de type II (T β RII) est alors recruté au niveau du complexe TGF- β /T β RII et phosphoryle le T β RI. La forte augmentation de l'activité kinase du récepteur I permet la transduction du message vers les protéines Smads

dans le cytoplasme [19].

Les Smads forment une classe de protéines qui fonctionnent comme des facteurs de signalisation intracellulaire pour les polypeptides sécrétés de la super famille $TGF\beta$. Les protéines Smads sont des polypeptides de 42 et 60 kDa, avec deux régions homologues en N-terminal et en C-terminal. Ces domaines sont appelés respectivement domaine MH1 et MH2 et sont séparés par un segment « linker » moins bien conservé [15] [16] [20]. Sur la base de considérations structurales et fonctionnelles, les protéines Smads ont été divisées en trois familles :

- les R-Smad (receptor regulated Smads) qui sont les substrats des récepteurs à kinase (Smad2 et 3 pour le récepteur TGF et Smad 1,3,5 pour le récepteur au BMP)
- les co-Smads qui s'associent avec les R-Smad (Smad4);
- les I-Smads (inhibitory Smads, Smad6 et 7) qui inhibent la fonction de signalisation des deux autres groupes en empêchant leur phosphorylation par le récepteur de type I du $TGF\beta$.

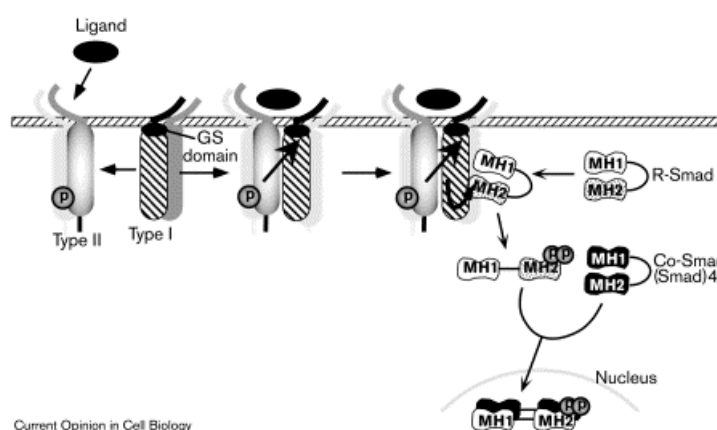


FIG. 8 – Mécanismes de signalisation induite par $TGF\beta$ [21].

Après fixation du $TGF\beta$, le récepteur $TR\beta I$ phosphoryle les protéines Smad 2 et 3 à leur extrémité C-terminale (Fig. 8). L'activation de Smad 2 et 3 permet, après dissociation du complexe récepteur, la formation d'un complexe avec Smad 4 constitutivement phosphorylée qui passe dans le noyau où il régule la transcription [22] [16].

Dans le noyau, les complexes hétéromériques Smad (Smad2/Smad4 ou Smad3/Smad4) interagissent avec des protéines se fixant à l'ADN, dont les plus étudiés appartiennent à la famille des protéines FAST (forkhead activin signal transducer). Chez le Xénope par exemple, FAST1 est lié de façon constitutive à un élément ARE (Activin Responsive Element) au niveau du promoteur du gène Mix.2. Après stimulation par le $TGF\beta$ ou l'activine il se lie au complexe hétérodimérique Smad2/Smad4 [23]. Smad2 interagit directement avec FAST1 alors que Smad4 participe à la stabilité du complexe au travers de son association avec Smad2. De façon analogue chez les mammifères, des complexes FAST2/Smad2/Smad4 interagissent avec le promoteur goosecoid [24] et les complexes Smad3/4 avec une séquence d'ADN appelée CAGA (AG(C/A)CAGACA présente dans les promoteurs de PAI-1 (plasminogen activator inhibitor) et de la MMP1 [22] [16]. Les complexes hétérodimériques interagissent aussi avec les coactivateurs de la transcription p300/CBP (CREB binding protein) ou avec le corepresseur transcriptionnel TGIF (TG Interacting Protein).

3.1.2 Régulation de l'activité des Smads par les voies kinases

Si la phosphorylation des Smads par le récepteur thérone/kinase T β RI est un élément clé de leur activation [25] [16] [26] d'autres kinases peuvent phosphoryler les protéines Smads (Fig. 9).

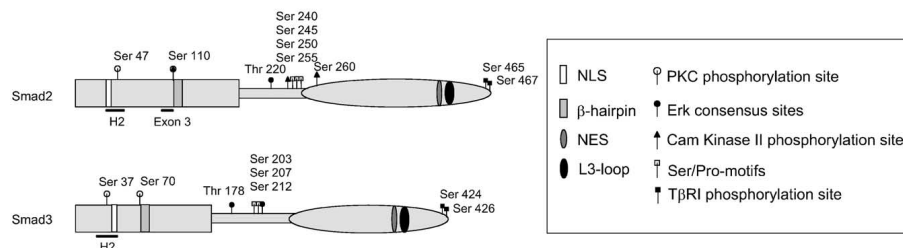


FIG. 9 – Sites de phosphorylation des R-Smad. Les R-Smad sont phosphorylées en carboxy-terminale par T β RI mais d'autres sites de phosphorylation sont la cible de différentes kinases. D'après [27].

Ainsi, la stimulation de récepteurs tyrosine/kinase et/ou de Ras induit la phosphorylation du domaine linker de Smad2 et de Smad3, et du domaine MH1 de Smad2, via l'activation de la voie Erk MAP Kinase (« Mitogen Activated Protein Kinase ») [28] [29]. Ce mécanisme inhibe la translocation nucléaire des Smads activés suite à la stimulation par le TGF β [29] et pourrait expliquer la mauvaise réponse au TGF β dans certaines cellules qui possèdent un signal Ras hyperactif. Cependant, ce phénomène n'a pas été observé dans des cellules transformées par Ras ou dans des cellules ayant un signal MAPK activé [30] [31] [28].

De même, la phosphorylation de Smad2, 3 et 4 induite par CamKII (Ca²⁺/ calmodulin dependent protein kinase II) entraîne l'inhibition de l'hétéro-oligomérisation de Smad2, de la translocation nucléaire et de l'activité transcriptionnelle de Smad2 induites par le TGF β [32]. La phosphorylation des domaines MH1 de Smad2 et 3 par la protéine kinase C (PKC) inhibe les associations Smad-ADN et suggère un rôle de PKC dans la régulation de la transcription par les Smad [33].

A l'opposé, la phosphorylation de Smad2 induite par MEKK1 (« MAPK/Erk Kinase Kinase 1 ») favorise l'hétéro-oligomérisation avec Smad4, la translocation nucléaire et l'activité transcriptionnelle du complexe Smad2-Smad4 [34]. De même, JNK (« c-Jun N-terminal Kinase ») induit la phosphorylation de Smad3, favorisant sa translocation nucléaire et la transcription des gènes cibles du TGF β [31]. Également, l'inhibition de la voie PI3K (« Phosphatidyl Inositol 3 Phosphate ») peut réduire la phosphorylation de Smad2 induite par le TGF β [35].

Le rôle de ces kinases a moins d'impact sur Smad4 car, à l'inverse des R-Smads, l'activité de Smad4 n'est pas dépendante d'une phosphorylation. Cependant, une étude a montré l'existence d'un site de phosphorylation par Erk dans le domaine linker de Smad4 permettant une meilleure translocation nucléaire de Smad4 en réponse au TGF β (Roelen et al., 2003). Quant à Smad6 et 7, elles sont phosphorylées indépendamment de TGF β , et la mutation d'une sérine, normalement phosphorylée dans Smad7, altère son activité transcriptionnelle lorsqu'elle interagit avec un domaine de fixation à l'ADN [36].

3.1.3 Autres voies de signalisation du TGF- β

En plus de la voie Smad, le TGF β active de nombreuses autres cascades de signalisation, dont celle des MAPK (Fig.). Certaines de ces voies régulent l'activation des Smad comme nous l'avons vu précédemment mais d'autres induisent des réponses indépendantes de l'activité transcriptionnelle des Smads.

Le TGF β peut en effet activer les voies Erk, JNK et p38 MAPK indépendamment des Smad en mobilisant les GTPases de type Ras et de la famille Rho (RhoA, Roc1, ...) mais aussi en activant la kinase TAK1, pour les voies JNK et p38.

L'ensemble de ces voies a été largement impliqué dans la régulation de l'expression de gènes induits par le TGF β . Ainsi, l'expression des gènes *PAI-1* [37], *c-jun* [38], *TGF β 1* [39] et du collagène $\alpha 2$ de type I [40], présentent des sites AP1 dans leur promoteurs et répondent au facteur de transcription c-jun et (voie JNK). Également, le TGF β peut induire l'expression du collagèneI via la voie PI3K/Akt/PDK [41].

La **convergence** entre les voies Smad et MAPK activées par le TGF β aboutit souvent vers une coopération dans la régulation transcriptionnelle des gènes cibles [25] [16] [26]. Cependant, il existe de nombreuses cross-régulations conduisant à une diminution du signal TGF-b. Ainsi la phosphorylation des domaines linker des R-Smad par les MAPK conduit à la séquestration des protéines dans le cytoplasme. La liaison de Smad6 à TAK1 diminue l'activité de celle ci [42] tandis que Smad7 peut augmenter l'activation de JNK [43]. Également, c-Jun inhibe le signal de Smad2 via une association avec Ski et cette interaction est régulée par le signal JNK [44]. Ainsi, la **balance** entre l'activation directe des Smad et les voies MAPK définit souvent les réponses de la cellule au TGF β et permet de répondre à la **multiplicité** des fonctions liées au TGF β .

3.2 Recueil des données de la littérature et ses limites

Cette partie ne se préoccupera pas de faire un catalogue de toutes les interactions qui ont pu être détectée dans les bases de données bibliographiques comme Medline. Il s'agira de montrer la complexité des données et les difficultés d'analyser manuellement les articles scientifiques. La recherche des données consiste à consigner les interactions publiées et prouvées *in vitro* et *in vivo*. Il est aussi important de noter l'aspect comportemental des interactions, l'action qu'elle implique sur l'activité d'une protéine ou l'expression d'un gène. Les données purement structurales renseignant par exemple sur l'existence d'un site de fixation pour une kinase ou un facteur de transcription sur un promoteur donné, sont considérées comme des données secondaires.

3.2.1 Medline, une source hétérogène

Tout en se focalisant sur la signalisation du TGF- β 1, plusieurs difficultés d'analyse se sont imposées.

- Les expériences sont souvent réalisées dans des conditions non physiologiques ou *in vitro*, et d'un point de vue quantitatif, les mesures obtenues ne sont donc pas très fiables.
- Dans les expériences *in vivo*, les mesures sont moins précises et elles combinent souvent plusieurs effets difficiles à distinguer. Cependant, avec l'arrivée de la technique de « Knock Out », consistant à priver un système biologique de l'expression d'un

gène ciblé, il est plus facile d'évaluer *in vivo* l'effet d'une molécule sur un processus donné à l'échelle de la cellule.

- Les données spatio-temporelles sont à prendre en compte, car la localisation cellulaire est un très bon moyen pour la cellule de réguler les interactions qui d'une manière temporelle dépendent de l'expression des gènes.
- Une autre difficulté de l'analyse est la présence de contradictions apparentes entre plusieurs publications. En effet, la même voie de signalisation peut-être utilisée pour des effets opposés. Dans l'exemple de la voie PI3K/PDK1/PKB(Akt), la protéine PKB(Akt) peut produire sous l'effet de l'insuline un effet antagoniste sur les protéines Smads dans les cellules épithéliales HEK293T et Hep3B [45]. Par contre, lorsque le TGF- β 1 mobilise lui-même la protéine PKB(Akt), cette voie peut favoriser la signalisation des protéines Smads dans le cas de la régulation du gène *COL1A2* dans les cellules de type myofibroblastique [41]. Dans cet exemple, on s'aperçoit que suivant le type cellulaire, la même voie de signalisation peut moduler soit positivement soit négativement l'effet du TGF- β 1.
- Par ailleurs, ces études portent généralement sur certaines interactions sans regarder le **contexte cellulaire**, c'est-à-dire les signaux déjà présents dans la cellule. Ce point justifie largement le recours à la modélisation de ces systèmes biologiques.

Depuis quelques années, les profils d'expression des gènes obtenus par analyses de puces à ADN constituent une source d'information supplémentaire sur les interactions [46] [47]. En effet certains modèles expérimentaux permettent de dégager des notions de hiérarchie temporelle des régulations géniques. D'autre part, des nouvelles méthodes d'inférence sur les événements de régulation des gènes se sont récemment développées et apportent du sens aux données d'expression. Cependant, ces méthodes sont assez peu précises lorsqu'on dispose de peu d'informations sur la régulation des facteurs de transcription.

3.2.2 Les bases de données

Les données d'interactions sont enfouies dans la littérature scientifique à la fois gigantesque diverse et hétérogène, et ne sont que très partiellement prises en compte dans les bases de données de séquences comme Swiss-Prot et GeneBank et les bases bibliographiques comme Medline. De nouvelles bases de données spécialisées dans les réseaux de signalisation se sont développées plus récemment. Les plus simples d'utilisation et disponibles gratuitement ou en version d'essai sont TRANSPATH[®], TRANSFAC[®] 4, BiblioSphere[®] 5, HPRD "Human Protein Reference Database" 6 et BIND 7.

Ces bases de données disposent d'outils de visualisation et de navigation dans les réseaux d'interactions, et donnent directement les références bibliographiques qui s'y rapportent. BiblioSphere[®] a la particularité de faire de l'analyse automatique de l'information contenue dans les bases bibliographiques comme Medline ("litterature mining network") : elle permet de récupérer et d'analyser toutes les interactions de gènes connues dans les bases de données bibliographiques.

Malgré tous ces avantages, ces bases de données spécialisées ne fournissent pas de données contextuelles des interactions : fonctions et effet des interactions dans un type cellulaire

4. <http://www.biobase.de>

5. <http://www.genomatix.de>

6. <http://www.hprd.org>

7. <http://www.blueprint.org/bind/bind.php>

donné, données spatio-temporelles, par exemples. Seule la HPRD fournit des informations sur le type d'expérience, *in vitro* ou *in vivo*.

3.2.3 Bilan

Le recueil de ces informations bibliographiques ont permis de construire une carte des réseaux d'interactions du TGF- β 1 comportant :

- 361 éléments biologiques (avec distinction de la localisation cellulaire);
- 250 interactions;
- et plus de 101 références bibliographiques.

Ces articles ont été lus manuellement par des recherches sur PubMed et sur les bases de données spécialisées sur les réseaux d'interaction. Cependant, il serait envisageable d'extraire ces informations de façon automatique à condition de définir toutes les manières d'exprimer dans le langage naturel, une interaction ou une régulation identifiée dans la littérature. Un problème majeur que l'on a pu rencontré dans la littérature du TGF- β 1, c'est la grande diversité des noms attribués aux protéines, notamment pour les protéines MAP kinase. Par exemple, la protéine MKK4 qui est son nom le plus courant peut aussi être retrouvé sous le nom de MAP2K4 (MITOGEN-ACTIVATED PROTEIN KINASE KINASE 4) ou SERK1, SEK1 (SAPK/ERK KINASE 1) ou JNKK1 (JNK-ACTIVATED KINASE 1) ou MEK4 (MAPK/ERK KINASE 4). Un autre problème majeur est l'ambiguïté des textes concernant les noms des protéines, des gènes. Par exemple, dans certains articles, il est mentionné de l'effet de la protéine ERK, mais en fait il existe au moins deux types d'ERK (1 et 2). De même, lorsqu'il s'agit du TGF- β , parle-t-on de la molécule (et laquelle, car il en existe trois isoformes) ou bien de la super-famille? En conclusion, si la collecte d'informations sur les interactions décrites dans la littérature peut être envisagée en utilisant des outils de fouille de données, la lecture et l'interprétation par un expert reste indispensable pour extraire l'information qui nous intéresse.

Nous avons réuni toutes les interactions protéine-protéine, protéine-gène et leur type de régulation dans un même réseau, mais il serait intéressant de prendre compte le type cellulaire dans laquelle a été identifiée l'interaction. En effet, selon le type cellulaire, il existe des différences entre les mêmes voies de signalisation, car les signaux présents peuvent être spécifiques des cellules. Il serait aussi intéressant de noter le nombre de publication référant chaque interaction. Cela pourrait donner des indications sur la valeur biologique de ces résultats.

Toutes ces points soulèvent des questions intéressantes comme « est-il possible de modéliser les connaissances? » ou bien « peut-on apprendre sur des connaissances et faire de l'inférence et émettre des hypothèses plausibles sur les réseaux d'interaction? ». Ces quelques points posent le problème de la publication de résultats mettant en évidence des interactions nouvelles. Un effort de normalisation devra probablement être fait pour faciliter les intégrations dans des modèles qui nécessitent la collecte d'un grand nombre d'informations de ce type. Indépendamment du format adopté (XML par exemple), il y a un problème de représentation des connaissances d'autant plus ardu qu'on ne connaît pas, à l'heure actuelle, un formalisme unique permettant de modéliser les réseaux métaboliques de signalisation et géniques.

3.3 Implémentation GARMeN

Dans ce chapitre, nous présentons comment peuvent être structurées les informations biologiques collectées, dans l'environnement GARMeN, et un exemple d'exploration intéressante que l'on peut obtenir.

3.3.1 Le langage de description des interactions

Les interactions sont répertoriées dans un fichier qui est traité par GARMeN. Le principe est de déclarer toute espèce moléculaire avant son utilisation dans une réaction. Il n'y a ainsi pas de règles strictes sur la structuration du code. Le corps du fichier peut être segmenté en sous-compartiments **imbriqués**, permettant de placer les réactions et les réactants dans un même compartiment « cellulaire ». Le choix de ces compartiments est laissé au biologiste. Ils peuvent être modéliser de la manière suivante :

- « Extracellular Matrix » qui représente la matrice extracellulaire;
- « External_layer_cell », la membrane externe des cellules ou surface de contact des cellules avec la matrice extracellulaire;
- « cytosol » le cytoplasme;
- « nucleus » le noyau.

En langage GARMeN, cela donne l'emboîtement suivant :

```
In ExtracellularMatrix {
  In External_layer_cell {
    In cytosol {
      In nucleus {
      }
    }
  }
}
```

Dans chaque compartiment, il est permis de construire des sous-réseaux (« Subnet ») dans lesquels est réuni un ensemble de réactions d'une même voie de signalisation ou métabolique. L'intérêt de ce sous-réseau est purement pratique afin d'alléger visuellement le graphe. Chaque « Subnet » peut ensuite être visualisé plus en détail. Ces sous-réseaux sont totalement arbitraires et ne font pas appel à des algorithmes de détermination automatique de sous-réseaux appliqués ou en cours de développement dans le cadre de la théorie des graphes. Chaque réaction est documentée par au moins une référence MedLine (PMID) documentant l'interaction en question. En langage GARMeN, nous avons par exemple pour l'activation des protéines Smads :

```
Subnet Smad_Activation {
  // HETERODIMERISATION : Smad3 & Smad4 interaction
  Reaction (name = SMAD3_SMAD4_INTERACTION,
    input = Smad3_P,           // --> REACTIF
    input = Smad4,             // --> REACTIF
    output = Smad3_P_Smad4,    //--> PRODUIT
    RefMedline = PMID:12809600) // --> REFERENCE Medline
} // end of Subnet
```

Concernant l'expression des gènes, la réaction prend pour argument le nom du gène, la protéine codée et tous les facteurs de transcription, activateurs et répresseurs sans distinction sous le label « actor ». L'exemple suivant montre l'action des protéines Smads sur le gène

mmp2, codant la collagénase A dont l'activité a été corrélée avec les stades précoces de la fibrose et l'invasivité des cellules tumorales (Yoshizaki *et al.*).

```

/* MMP2 collagenase type IV-A or gelatinase A
   Activator : Smad (Smad2 or Smad3), p53
*/
GeneExpression (gene = mmp2,
                 actor = Smad2_P_Smad4,
                 actor = Smad3_P_Smad4,
                 actor = p53,
                 protein = MMP2,
                 RefMedline = PMID:14502566,
                 RefMedline = PMID:7642529,
                 RefMedline = PMID:9343394)

Transport To cytosol (target = MMP2)

```

Les protéines peuvent migrer d'un compartiment à un autre grâce à la réaction « Transport » qui prend en compte le compartiment de destination et le nom de la protéine.

3.3.2 La carte du réseau d'interactions

La Fig. 11 montre le réseau entier des interactions dans l'interface graphique de GARMeN. Cette interface facilite la navigation dans le réseau en permettant à l'utilisateur de sélectionner et d'afficher les sous-réseaux ou les produits. A l'aide de simples clics de souris, il est par exemple possible d'avoir accès aux références PubMed de chaque interaction.

L'interface permet à l'utilisateur de sélectionner des espèces moléculaires et d'afficher toutes les autres molécules qui lui sont en relation, en recherchant leurs prédécesseurs ou leurs successeurs. Pour faciliter la navigation, le programme enregistre les sous-réseaux déjà explorés. Il est ainsi facile de retrouver des sous-graphes intéressants d'un point de vue biologique.

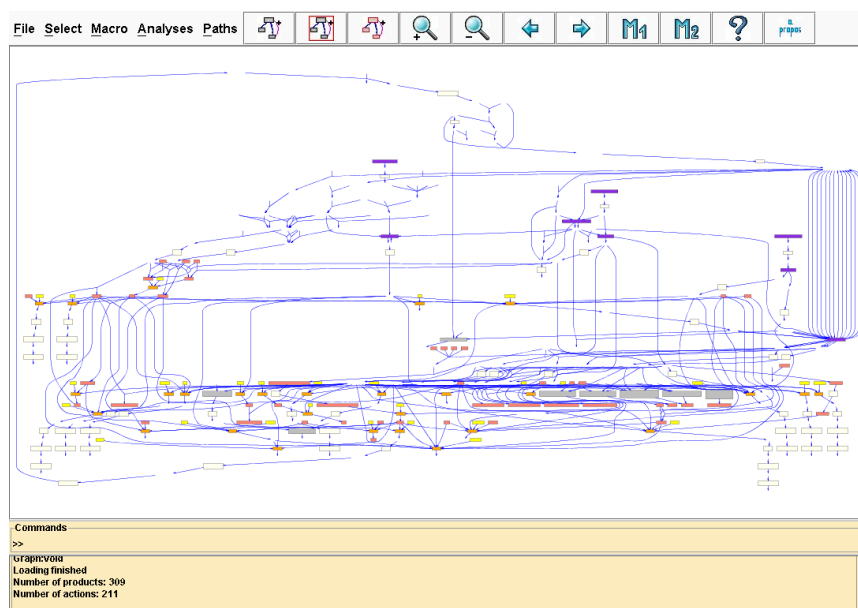


FIG. 11 – Réseau d'interactions affiché par GARMeN

3.3.3 Exemple d'exploration utile

Nous avons choisis de présenter la voie de la kinase S6 ribosomale ($p70^{S6K}$) / mTOR (« mammalian target of rapamycin ») activée par le TGF- β 1 et le facteur de croissance EGF via la protéine PI3K (Fig. 12). Dans la littérature, la régulation de la protéine $p70^{S6K}$ est complexe, car elle active les processus de transformations, de croissance cellulaire et d'entrée dans le cycle cellulaire. Cette voie de signalisation est aussi fortement impliquée dans de nombreuses tumeurs malignes, incluant le cancer du sein, de la prostate, du colon et du cerveau.

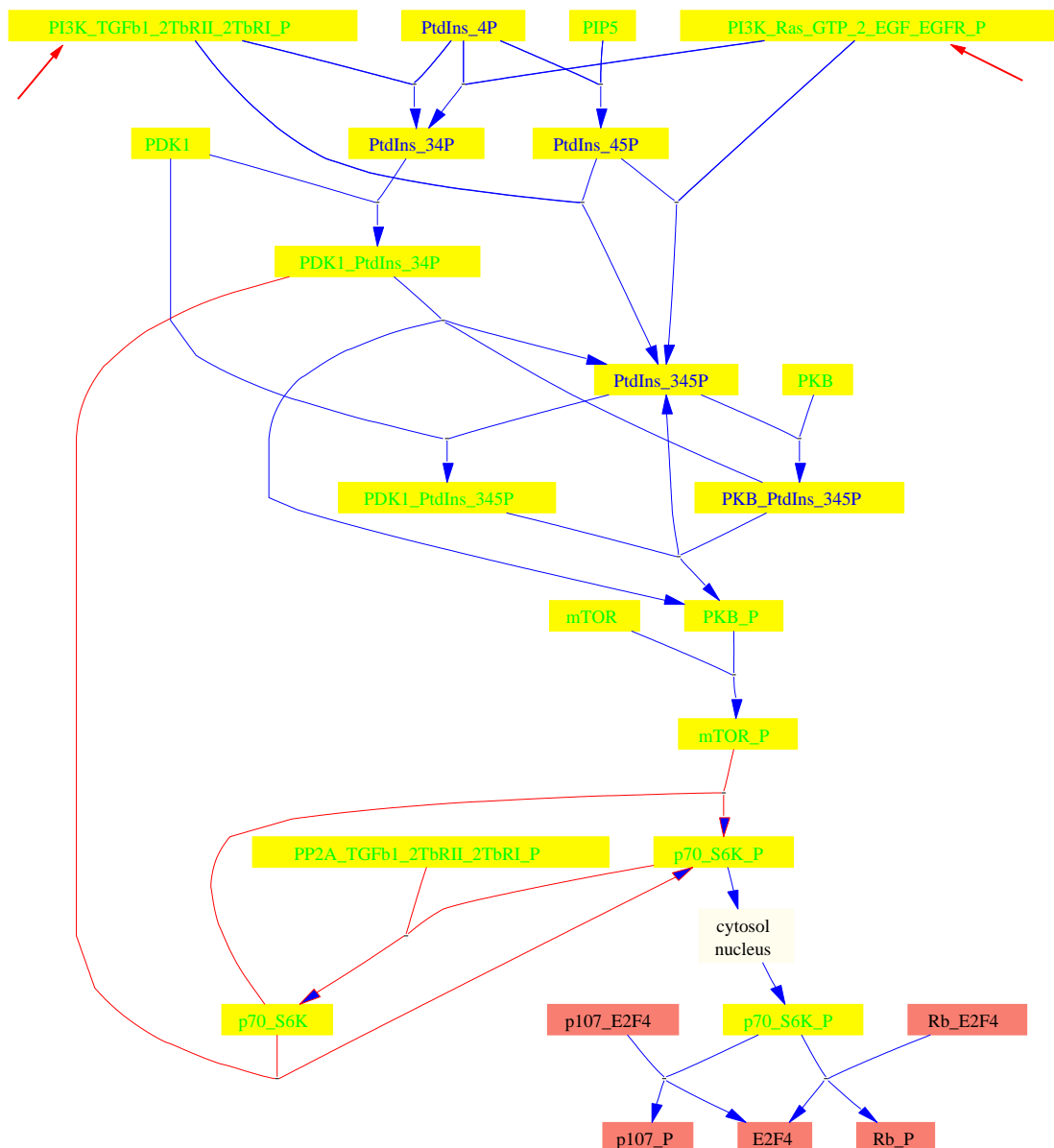


FIG. 12 – Activation de $p70^{S6K}$ par la voie Pi3K et inhibition par le TGF- β 1 induisant l'arrêt du cycle cellulaire [48] [49] [50]

Dans le graphe de la Fig. 12, nous pouvons remarquer que le TGF- β 1 peut à la fois favoriser la phosphorylation (état activé) et la déphosphorylation (état inactif) de la $p70^{S6K}$

(flèches rouges). GARMen permet donc de mettre en évidence des effets de régulations opposées que nous rencontrons de façon dispersée dans la littérature.

3.4 Expérimentation SIGNAL/SIGALI

Un premier modèle relationnel du réseau d'interactions étant construit, le temps restant du stage a permis de tester la modélisation formelle SIGNAL/SIGALI décrite précédemment.

L'objectif a été de tester si nous pouvions coder de façon pertinente les interactions de signalisation et les régulations des gènes à l'aide du langage SIGNAL (expressivité) et ensuite, d'analyser les réponses du modèle SIGNAL à des questions biologiques.

3.4.1 Abstraction du modèle GARMen

Nous allons maintenant essayer d'adapter le modèle SIGNAL au modèle relationnel. Une protéine (enzyme, facteur de transcription, molécule signal) qui joue un rôle dans la transduction du signal est représentée par un signal au sens de SIGNAL/SIGALI. Ce signal sera de type « **event** », c'est-à-dire qu'il pourra prendre une seule valeur toujours égale à 1 ou *true*, contrairement à un signal booléen. Cette représentation permettra de raisonner sur les présences relatives au niveau des signaux biologiques.

Ce type de modèle semble adapté à la réalité biologique des voies de signalisation induites par les cytokines pour les raisons suivantes :

- les données très peu quantitatives justifient le choix de modèle logique,
- l'échelle de temps des interactions est assez courte pour les considérer instantanées, par rapport à l'activité des gènes,
- les gènes ont au contraire une échelle de temps plus longue à cause des nombreuses molécules impliquées dans leur expression, et on va supposer qu'ils gardent leur activité ou leur inactivité d'expression tant qu'il n'y a pas d'activateur ou de répresseur de transcription; on pourra ainsi les modéliser comme des variables d'états,
- les protéines ont par hypothèse, des horloges différentes, car elles ne sont pas exprimées et n'agissent pas toutes en même temps; c'est pour cela que l'on utilise les opérateurs temporelles polychrones (ou multi-horloge) vus dans la sous-section 2.2.5 (page 7).

Les gènes sont codés dans le langage SIGNAL en variables d'états de type booléen (vrai ou faux). Les activateurs d'expression des gènes modifient leur valeur à vrai (v) et les répresseurs la mettent à faux (f).

3.4.2 Passage automatique du modèle GARMen au modèle SIGNAL

La formalisation a consisté à représenter chaque interaction protéine-protéine et protéine-gène par une relation SIGNAL (Fig. 14).

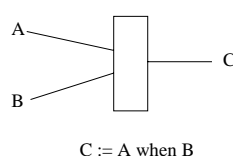


FIG. 14 – Formalisation des interactions

Cependant nous avons été confronté à plusieurs difficultés qui sont les suivantes :

- cycles repérés dans le graphe qui ne sont pas des vrais cycles au sens de la régulation. Il y a dans ce cas, un problème de définition de signaux où un signal se définit par lui-même. Prenons l'exemple de la Fig. 15, placé dans le contexte où il n'y a que des interactions d'association et de dissociation de protéines A, B, C et D.

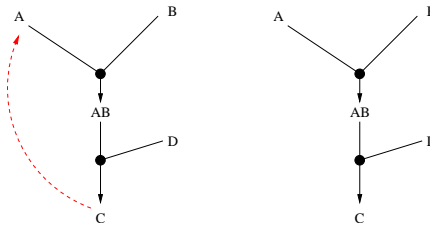


FIG. 15 – Cycles $A/A-B/C/A$: A s'associe avec B pour former le complexe AB qui interagissant avec D donne le complexe C et fait se dissocier la protéine A

Dans cet exemple nous avons :

$$\begin{aligned} AB &:= A \text{ when } B \\ C &:= AB \text{ when } D \\ \rightarrow A &:= C \end{aligned}$$

ce qui pose un problème de définition théorique des signaux et donc un problème de compilation du code à la troisième ligne fléchée, car la présence de C dépend de A qui dépend elle-même de C ! La solution est de faire abstraction de la flèche en tirets rouge pour obtenir le graphe de droite de la Fig. 15. Le problème majeur sur ce point est de savoir distinguer ces types de cycle avec d'autres circuits comme ceux impliqués dans la réversibilité des réactions chimiques (voir Fig. 12 pour la phosphorylation et la déphosphorylation de la protéine $p70^{S6K}$)

- la compétition entre activateurs et répresseurs de transcription des gènes pose un problème, car formellement les signaux permettant le changement d'une variable d'états doivent être exclusifs. La solution est d'introduire une **variable artificielle** de type booléen qui est présente lorsque les compétiteurs sont présents en même temps et qui arbitrairement selon la valeur de cette variable donne la prédominance à l'un des deux compétiteurs. Ici le problème est de pouvoir généraliser ces compétitions à plusieurs antagonistes sans introduire trop de variables libres dans le système alourdissant les calculs.
- le codage de certains gènes pose un problème biologique car les gènes, pour lesquels nous connaissons leurs activateurs, mais pas leurs répresseurs, restent dans l'état activé sans revenir à leur niveau basal. Dans ce cas, il faut alors de nouveau introduire une variable qui arbitrairement au bout d'un certain nombre d'occurrences de l'horloge la plus rapide bascule le gène dans l'état inactif. Mais ce choix arbitraire peut être remis en cause. Cependant dans d'autres cas, nous connaissons l'existence de complexes spécialisés dans la dégradation des facteurs de transcription assurant un recyclage de ces protéines, mais aussi soit l'activation soit la répression de la transcription des gènes selon le type des facteurs de transcription considérés.

3.4.3 Obtention du modèle SIGNAL

Les signaux moléculaires sont organisés en processus selon qu'ils appartiennent ou non à la même voie de signalisation comme dans l'exemple ci-dessous.

```
% ----- Smad Activation in the membrane plasma ----- %
process SMAD_ACTIVATION =
(? event TGFR, SMAD2_SARA,
   SMAD7, Smurf2;  //--> signaux d'ENTREE
! event SMAD2P, SMAD3P, SARA; )  //--> signaux de SORTIE
(| % active receptor complex and TURN over by Smad7/Smurf2 complex %
  TGFR_Actif := TGFR ^- SMAD7 ^- Smurf2

% ACTIVATION de Smad2 %
| SMAD2_SARA_TGFR := (SMAD2_SARA when TGFR_Actif)

% PHOSPHORYLATION DE SMAD2 %
| SMAD2P := SMAD2_SARA_TGFR ^+ SMAD2_SMAD3_SARA_TGFR

| SARA := SMAD2P ^+ SMAD3P |)
% ----- %
where event TGFR_Actif, SMAD2_SARA_TGFR, SMAD3_SARA_TGFR,
           SMAD2_SMAD3_SARA_TGFR, SMAD3_TGFR;  //--> signaux INTERMEDIAIRES
end
```

Plusieurs modèles SIGNAL ont été écrits. Un premier modèle décrit la signalisation des protéines Smads et celles des autres voies de signalisation. Un deuxième modèle plus judicieux ne traite que la signalisation du TGF- β 1 en tenant compte des points d'interférence avec d'autres voies de signalisation qui peuvent l'affecter. Ce dernier modèle, permet en effet de réduire considérablement le nombre de paramètres dans le système. Dans la suite, nous ne parlerons que de ce deuxième modèle.

La compilation de ce code génère ensuite un système dynamique polynômiale (PDS) que l'on peut manipuler avec SIGALI.

3.4.4 Manipulation du système dynamique polynômial avec SIGALI

Nous avons testé le comportement du système dynamique sans contraintes (où les signaux en entrée et les variables d'états sont libres), et avec contraintes sur certains signaux :

- absence du récepteur au TGF- β 1,
- protéines MAPK sont toujours présentes et actives.

La littérature a montré que les protéines MAPK (p38, JNK, PI3K et ERK) sont très importantes dans la signalisation du TGF- β 1. En effet, ces MAPK sont directement liées à la régulation de l'activation et la translocation des protéines Smads dans le noyau, et permettent au TGF- β 1 d'agir sur un très grand nombre de gènes [51].

L'objectif est de regarder si à partir des conditions initiales et des contraintes du système, un ensemble particulier d'états de gènes est atteignable, c'est-à-dire s'il existe des signaux permettant d'atteindre cet ensemble d'états.

Seize gènes ont été sélectionnés pour faire partie du PDS. La plupart d'entre eux codent pour des facteurs de transcription et les autres pour des protéines de la matrice extracellulaire et d'autres jouent un certain rôle dans la réparation et l'inflammation des tissus (Tab. 1).

L'initialisation des gènes est un **choix arbitraire** en fonction de l'effet activateur ou répresseur du TGF- β 1 sur ces gènes : si le TGF- β 1 a un effet activateur, alors le gène est initialisé à l'état inactif et dans l'autre cas, le gène est à l'état actif. Cette vision peut être simpliste mais la modélisation ne tient compte que de deux états de gène (actif et non actif).

Nome du gène	Initialisation	Protéine codée
<i>MADH7</i>	inactif	Smad7
<i>cmYC</i>	inactif	cMyc
<i>cfos</i>	inactif	cFos
<i>tgif</i>	inactif	TGIF
<i>cjun</i>	inactif	cJun
<i>TP53</i>	inactif	p53
<i>junB</i>	inactif	JunB
<i>atf3</i>	inactif	ATF3
<i>timp1</i>	inactif	TIMP1
<i>mmp1</i>	actif	MMP1
<i>mmp2</i>	inactif	MMP2
<i>mmp13</i>	inactif	MMP13
<i>COL1A1</i>	inactif	collagène $\alpha 1$ I
<i>COL1A2</i>	inactif	collagène $\alpha 2$ I
<i>thbs1</i>	inactif	thrombospondine
<i>et1</i>	inactif	endothéline

TAB. 1 – Liste de gènes pris en compte dans le modèle

Étude statique de la phosphorylation de $p70^{S6K}$ Une des caractéristiques du TGF- $\beta 1$ est d’avoir des effets dépendants du contexte cellulaire. Par inspection à l’aide de GARMEn, nous avons mis en évidence une partie du modèle pouvant donner des indications sur des effets opposés du TGF- $\beta 1$ sur le cycle cellulaire (Fig. 12). La protéine $p70^{S6K}$ joue un rôle dans le passage G1-S du cycle cellulaire et dans ce modèle peut être phosphorylée soit par mTOR_P soit par PDK1_PtdIns_34P sous l’action de PI3K activée indépendamment par le TGF- $\beta 1$ ou par l’EGF. Mais le TGF- $\beta 1$ n’a pas qu’un effet de phosphorylation sur $p70^{S6K}$, il peut via la protéine PP2A induire sa déphosphorylation. Plusieurs questions se posent alors :

- peut-il y avoir une seule phosphorylation à la fois ?
- quelles sont les conditions pour qu’il n’y ait qu’une seule ou les deux phosphorylations à la fois ?

Pour la première question, la démarche à suivre est de créer deux signaux fictifs représentant les différences réciproques des signaux PDK1_PtdIns_34P et mTOR_P :

$$\begin{aligned} \text{mtorp_pdk1} &:= \text{mTOR_P} \wedge \neg \text{PDK1_PtdIns_34P} \\ \text{pdk1_mtorp} &:= \text{PDK1_PtdIns_34P} \wedge \neg \text{mTOR_P} \end{aligned}$$

L’analyse par SIGALI nous indique qu’indépendamment du TGF- $\beta 1$ et de la voie Ras induite par l’EGF, nous pouvons avoir une seule ou deux phosphorylation. Le signal `mtorp_pdk1` est nul quelque soit les signaux d’entrée, alors que le signal `pdk1_mtorp` est non nul. Cela signifie que les instants du signal mTOR_P sont compris dans ceux de PDK1_PtdIns_34P. D’un point de vue biologique, la phosphorylation de mTOR_P ne pourrait se produire que si celle de PDK1_PtdIns_34P a lieu. Par ailleurs, la seule manière d’avoir une seule phosphorylation à la fois est de passer par la PDK1.

Pour la deuxième question, la démarche consiste à analyser les solutions calculées par SIGALI pour le signal `pdk1_mtorp` qui est présent lorsqu’il n’y a qu’une phosphorylation comme dans la figure 16.

Ce graphe représente les données fournies par SIGALI. Il indique que :

- la phosphorylation de $p70^{S6K}$ nécessite indifféremment l’induction par le TGF- $\beta 1$ ou

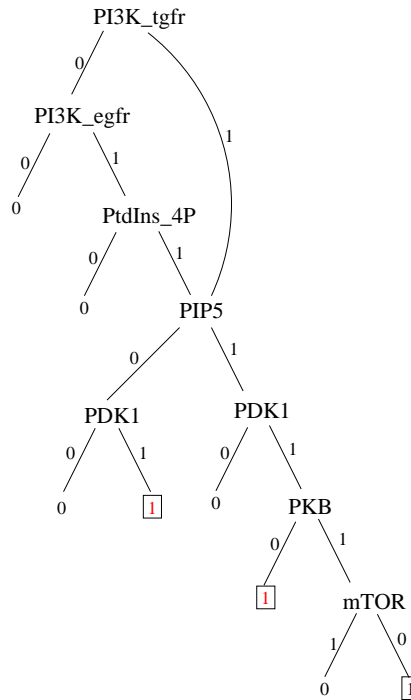


FIG. 16 – Solutions calculées par SIGALI pour le signal *pdk1_mtorp*

bien par l'EGF,

- la phosphorylation de $p70^{S6K}$ a lieu lorsque la protéine PKB est absente ou bien lorsque la protéine PIP5 est absente et que la PDK1 est présente.

D'un point de vue biologique, il resterait à vérifier la pertinence de ces résultats, mais ceux-ci ont été obtenus dans un modèle simplifié de la réalité ne tenant compte que de quelques acteurs de la régulation de la voie $p70^{S6K}$. Cependant, cette analyse montre que ce modèle est capable de répondre à des questions simples, mais qui peuvent donner des pistes de recherches par rapport aux connaissances actuelles. Il serait intéressant de vérifier si le fait qu'il n'y ait qu'une phosphorylation de $p70^{S6K}$ par la PDK1 a un impact significatif sur la progression dans le cycle cellulaire, et de rechercher si les inhibiteurs pharmacologiques connus de mTOR et de PDK1 ont un effet souhaitable dans ce système de régulation, comme par exemples UCN-01 (Kyowa Hakko Kogyo) qui cible la PDK1 ou le CCI-779 (Wyeth) qui cible mTOR [52].

Étude dynamique de la suppression du TGF- β 1 Nous testons l'effet de cette cytokine sur l'évolution des gènes. Il suffit de modifier à 0 (absent continue), la valeur des signaux influençant la libération du TGF- β 1, dans les équations des contraintes ($Q(G,S)$) du système.

SIGALI nous répond avec la fonction « Reachable » ou atteignabilité qu'aucun des gènes ne peut changer d'état. De plus, les états de départ sont des **états attractifs** (fonction « Attractivity »). Autrement dit, la présence du TGF- β 1 est indispensable pour faire modifier l'état de tous les gènes du système.

Étude dynamique de la présence continue du signal Ras-GTP Dans les cellules épithéliales Ras-oncogène, cette protéine hyperactive interagit avec la signalisation du TGF- β 1 en

inhibant son effet sur les gènes (Kretzschmar *et al.* 1999). En mettant la valeur des signaux ERK1_P et ERK2_P actif à 1 (présent à tous les instants) dans les équations de contraintes du système, SIGALI répond que les gènes ne changent pas d'états. Autrement dit, les protéines ERKs phosphorylées sont dans le modèle, des remparts forts à la signalisation du TGF- β 1.

Etude dynamique de la présence continue du signal p38-P Dans le système actuel, nous avons fixé la protéine p38 à la valeur 1 (présent à toutes les occurrences), ainsi que d'autres MAKKK (MKK3, MKK6 et MKK4) dont dépend l'activation de la protéine p38. Dans ce système, nous obtenons de façon surprenante que tous les gènes ne peuvent pas changer d'états, même les gènes qui dépendent directement de p38 comme le gène cfos.

L'objectif premier était de tester les outils informatiques. Nous avons constaté que ces outils supportaient des modèles de taille significative par rapport aux phénomènes biologiques que l'on souhaite étudier. En termes biologiques, les trajectoires des systèmes dynamiques représentent des évolutions de profils d'expression de gènes et sont potentiellement intéressantes. Par contre, le temps a manqué pour cibler un problème biologiquement pertinent et l'intérêt biologique de la modélisation proposée demande à être évalué.

4. Conclusion

Nous avons manuellement collecté et analysé de nombreux articles afin de réunir des informations sur les interactions protéine-protéine et protéine-gène impliqués dans la signalisation du TGF- β 1. Nous avons ainsi construit un réseau d'interaction de signalisation du TGF- β 1 visualisable, et exploitable par le logiciel GARMeN. Ce langage de description peut être écrit et modifié à souhait par le biologiste. Ce modèle présente notamment l'intérêt de mettre en relation des interactions identifiées indépendamment dans la littérature et de révéler des systèmes de régulations probables. Mais il ne permet pas d'étudier les phénomènes de façon dynamique (évolution des états de gènes par exemple).

Nous avons choisis le langage SIGNAL pour faire un modèle comportemental, car il semble à priori un bon choix pour décrire les interactions biologiques. Il permet de décrire comment se combinent les différents signaux et de vérifier des propriétés du modèle biologique, comme par exemple l'atteignabilité d'un ensemble d'états correspondant à un profil d'expression de gènes. Seulement ce langage n'est pas tout à fait adapté à la dynamique des systèmes biologiques. L'obstacle majeur est le déterminisme du langage qui fait que pour un signal donné, un seul état de gène peut être atteint. Nous avons vu que pour les compétitions entre différentes entités biologiques, il fallait entrer de nouveaux paramètres de type booléen pour simuler l'indéterminisme. Cela augmente la complexité des calculs car pour chaque variable libre du système, la taille du système d'équations polynomiales est multiplié par trois (3^n où n est le nombre de variables libres).

Plusieurs modèles SIGNAL ont été écrit à la main et nous avons été confronté à plusieurs difficultés de codage des interactions en langage SIGNAL. Le principe dans la description des interactions est de s'intéresser aux **événements** que l'on veut étudier (comme par exemple la phosphorylation d'une molécule correspondant à son activation et à la transmission du signal vers d'autres voies de signalisation).

Les perspectives de recherche consistent à vérifier, à confirmer le modèle relationnel, et de le compléter si nécessaire. Le souci sur ces modèles est de savoir s'il est utile de distinguer les interactions identifiées dans différents types cellulaires. Concernant le modèle comportemental, les premières expériences ont prouvé que le modèle SIGNAL pouvait potentiellement être un outil puissant pour étudier les propriétés dynamiques d'un réseau en terme de fonctionnalité des interactions inhérente à la modélisation par événements (c'est-à-dire non quantitative). Ce modèle permettrait aussi d'expérimenter *in silico* des stratégies d'inhibition et de stimulation des voies de signalisation pour guider celles qui sont réalisées dans le domaine biomédical. Cependant, il faudra mener des recherches pour l'adapter à la logique du vivant car l'étude dynamique doit tenir compte de l'indéterminisme des interactions. Il serait en plus envisageable de ne pas seulement tenir compte des interactions pertinentes, mais aussi de leur poids relatif. La possibilité de générer automatiquement el modèle SIGNAL à partir du modèle GARMeN pourrait aussi faire l'objet d'une étude.

Références

- [1] E.F. Keller. *Le rôle des métaphores dans les progrès de la biologie*. Institut d'édition Sanofi-Synthélabo, 1999.
- [2] H. Atlan. *La fin du tout génétique*. Inra Editions, 1999.
- [3] D.M. Bissell. Chronic liver injury, TGF- β , and cancer. *Experimental and Molecular Med*, 33:79–190, 2001.
- [4] A.M. Gressner, R. Weiskirchen, K. Breitkopf, and S. Dooley. Roles of TGF- β in hepatic fibrosis. *Frontiers in Bioscience*, 7:793–807, 2002.
- [5] H. Marchand, P. Bournai, M. Le Borgne, and P. Le Guernic. Synthesis of Discrete-Event Controllers based on the Signal Environment. *Kluwer Academic Publishers*, 2000.
- [6] B. Schoeberl, Eichler-Jonsson .C, Dieter Gilles E, and Müller G. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nature*, 2002.
- [7] M.S. Ko, Nakauchi H., and Takahashi N. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J.*, 9:2835–2842, 1990.
- [8] T. Msadeck. When the going gets tough: survival strategies and environmental signaling networks in *Bacillus subtilis*. *Trends Microbiology*, 7:201–207, 1999.
- [9] Bassano V. and Thieffry D. and Bernot G. Marked Regulatory aphs : A formal framework for biological regulatory networks. *Interdisciplinary school on Imaging, modeling and manipulating regulatory networks, Ambleteuse 2002*, 2002.
- [10] F. De Saussure. *Cours de Linguistique Générale*. Payot, Genève, 1931.
- [11] J. Monod. *Le hasard et la nécessité*. Point Seuil, 1970.
- [12] S. Kauffman. *The Oirgins of the orders: Self-Organisation and Selection in Evolution*. Oxford University Press, 1993.
- [13] H. Atlan. On a formal definition of Organisation. *J. Theor. Biol.*, 45:1–9, 1974.
- [14] R. Derynck and X.H. Feng. TGF-beta receptor signaling. *Biochim Biophys Acta*, 1333(2):105–50, 1997.
- [15] C.H. Heldin, K. Miyazono, and P. ten Dijke. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature*, 390:465–71, 1997.
- [16] J. Massague and Y.G. Chen. Controlling TGF-beta signaling. *Genes Dev.*, 14(16):627–44, 2000.
- [17] R. Derynck, R.J. Akhurst, and A. Balmain. TGF-beta signaling in tumor suppression and cancer progression. *Nat Genet.*, 29(2):117–29, 2001.
- [18] J.L. Wrana, L. Attisano, J. Carcamo, A. Zentella, J. Doody, M. Laiho, X.F. Wang, and J. Massague. TGF beta signals through a heteromeric protein kinase receptor complex. *Cell*, 71(6):1003–14, 1992.
- [19] J.L. Wrana, L. Attisano, R. Wieser, Ventura F., and J. Massague. Mechanism of activation of the TGF-beta receptor. *Nature*, 370(6488):341–7, 1994.
- [20] J.L. Wrana and L. Attisano. The Smad pathway. *Cytokine Growth Factor Rev.*, 11(1-2):5–13, 2000.
- [21] L. Attisano and J.L. Wrana. Smads as transcriptional co-modulators. *Curr Opin Cell Biol*, 12(2):235–43, 2000.
- [22] R. Derynck, Y. Zhang, and XH. Feng. Smads: transcriptional activators of TGF-beta responses. *Cell*, 95(6):737–40, 1998.

- [23] X. Chen, E. Weisberg, V. Fridmacher, M. Watanabe, G. Naco, and M. Whitman. Smad4 and FAST-1 in the assembly of activin-responsive factor. *Nature*, 389(6646):85–9, 1997.
- [24] E. Labbe, C. Silvestri, P.A. Hoodless, J.L. Wrana, and L. Attisano. Smad2 and Smad3 positively and negatively regulate TGF beta-dependent transcription through the forkhead DNA-binding protein FAST2. *Mol Cell.*, 2(1):109–20, 1998.
- [25] S. Itoh, F. Itoh, M.J. Goumans, and Ten Dijke P. Signaling of transforming growth factor-beta family members through Smad proteins. *Eur J Biochem.*, 267(24):6954–67, 2000.
- [26] A. Moustakas, S. Souchelnytskyi, and C.H. Heldin. Smad regulation in TGF-beta signal transduction. *J Cell Sci.*, Pt 24:4359–69, 2001.
- [27] M. Lutz and P. Knaus. Integration of the TGF-beta pathway into the cellular signalling network. *Cell Signal*, 12:977–88, 2002.
- [28] M. Funaba, C.M. Zimmerman, and L.S. Mathews. Modulation of Smad2-mediated signaling by extracellular signal-regulated kinase. *J Biol Chem.*, 277(44):41361–8, 2002.
- [29] M. Kretzschmar, J. Doody, I. Timokhina, and Massagué J. A mechanism of repression of TGF- β /smad signaling by oncogenic Ras. *Genes Dev.*, 13:804–816, 1999.
- [30] M.P. de Caestecker, W.T. Parks, C.J. Frank, P. Castagnino, D.P. Bottaro, A.B. Roberts, and R.J. Lechleider. Smad2 transduces common signals from receptor serine-threonine and tyrosine kinases. *Genes Dev.*, 12(11):1587–92, 1998.
- [31] M.E. Engel, M.A. McDonnell, B.K. Law, and H.L. Moses. Interdependent SMAD and JNK signaling in transforming growth factor-beta-mediated transcription. *J Biol Chem.*, 274(52):37413–20, 1999.
- [32] S.J. Wicks, S. Lui, N. Abdel-Wahab, R.M. Mason, and A. Chantry. Inactivation of smad-transforming growth factor beta signaling by Ca(2+)-calmodulin-dependent protein kinase ii. *Mol Cell Biol*, 20(21):8103–11, 2000.
- [33] I. Yakymovych, P. Ten Dijke, C.H. Heldin, and S. Souchelnytskyi. Regulation of smad signaling by protein kinase C. *FASEB J.*, 15(3):553–5, 2001.
- [34] J.D. Brown, M.R. DiChiara, K.R. Anderson, M.A. Gimbrone, and JN. Jr, Topper. Mekk-1, a component of the stress (stress-activated protein kinase/c-Jun N-terminal kinase) pathway, can selectively activate Smad2-mediated transcriptional activation in endothelial cells. *J Biol Chem*, 274(13):8797–805, 1999.
- [35] A.V. Bakin, A.K. Tomlinson, N.A. Bhowmick, H.L. Moses, and CL. Arteaga. Phosphatidylinositol 3-kinase function is required for transforming growth factor beta-mediated epithelial to mesenchymal transition and cell migration. *J Biol Chem*, 275(47):36803–10, 2000.
- [36] L. Pulaski, M. Landstrom, C.H. Heldin, and S. Souchelnytskyi. Phosphorylation of Smad7 at Ser-249 does not interfere with its inhibitory role in transforming growth factor-beta-dependent signaling but affects Smad7-dependent transcriptional activation. *J Biol Chem*, 276(17):14344–9, 2001.
- [37] M.R. Keeton, S.A. Curriden, A.J. van Zonneveld, and D.J. Loskutoff. Identification of regulatory sequences in the type 1 plasminogen activator inhibitor gene responsive to transforming growth factor beta. *J Biol Chem*, 266(34):23048–52, 1991.
- [38] P. Angel, K. Hattori, T. Smeal, and M. Karin. The jun proto-oncogene is positively autoregulated by its product, Jun/AP-1. *Cell*, 55(5):875–85, 1988.

-
- [39] S.J. Kim, A. Glick, M.B. Sporn, and A.B. Roberts. Characterisation of the promoter region of the human transforming growth factor- β 1 gene. *J. Biol. Chem.*, 264:402–408, 1989.
 - [40] K.Y. Chung, A. Agarwal, J. Uitto, and A. Mauviel. An AP-1 binding sequence is essential for regulation of the human α 2(I) collagen (COL1A2) promoter activity by transforming growth factor- β . *J Biol Chem*, 271:3272–3278, 1996.
 - [41] C.E. Runyan, H.W. Schnaper, and A.C. Poncelet. The phosphatidylinositol 3-kinase/Akt pathway enhances Smad3-stimulated mesangial cell collagen I expression in response to transforming growth factor- β 1. *J Biol Chem.*, 279(4):2632–9, 2004.
 - [42] N. Kimura, R. Matsuo, H. Shibuya, K. Nakashima, and T. Taga. BMP2-induced apoptosis is mediated by activation of the TAK1-p38 kinase pathway that is negatively regulated by Smad6. *J Biol Chem*, 275(23):17647–52, 2000.
 - [43] A. Mazars, F. Lallemand, C. Prunier, J. Marais, N. Ferrand, M. Pessah, G. Cherqui, and A. Atfi. Evidence for a role of the JNK cascade in Smad7-mediated apoptosis. *J Biol Chem*, 276(39):36797–803, 2001.
 - [44] M. Pessah, J. Marais, C. Prunier, N. Ferrand, F. Lallemand, A. Mauviel, and A. Atfi. c-Jun associates with the oncoprotein Ski and suppresses Smad2 transcriptional activity. *J Biol Chem*, 277(32):29094–100, 2002.
 - [45] I. Remy, A. Montmarquette, and SW. Michnick. Pkb/akt modulates tgf- β signalling through a direct interaction with smad3. *Nat Cell Biol.*, 6(4):358–65, 2004.
 - [46] E. Segal, M. Shapira, A. Regev, D. Pe’er, Botstein D., and Daphne K. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature genetics*, 34(2):166–176, 2003.
 - [47] Y.C. Yang, Piek E., J. Zavadil, D. Liang, D. Xie, J. Heyer, P. Pavlidis, R. Kucherlapati, Roberts A.B., and E Böttinger. Hierarchical model of gene regulation by transforming growth factor β . *PNAS*, 100(18):10269–10274, 2003.
 - [48] V.P. Krymskaya, R. Hoffman, A. Eszterhas, V. Ciocca, and R.A. Jr Panettieri. TGF- β 1 modulates EGF-stimulated phosphatidylinositol 3-kinase activity in human airway smooth muscle cells. *Am J Physiol*, 273(L1220-7), 1997.
 - [49] I. Griswold-Prenner, C. Kamibayashi, E.M. Maruoka, M.C. Mumby, and R. Derynck. Physical and functional interactions between type i transforming growth factor β receptors and B α , a WD-40 repeat subunit of phosphatase 2A. *Molecular and Cellular Biology*, 18(11):6595–6604, 1998.
 - [50] C. Petritsch, H. Beug, A. Balmain, and M. Oft. TGF- β inhibits p70 S6 kinase via protein phosphatase 2A to induce G1 arrest. *Genes & Development*, 14(3093-3101), 2000.
 - [51] R. Derynck and YE. Zhang. Smad-dependent and Smad-independent pathways in TGF- β family signalling. *Nature*, 425(6958):577–84, 2003.
 - [52] J.A. Fresno Vara, E. Casado, J. de Castro, P. Cejas, C. Belda-Iniesta, and Gonzalez-Baron M. PI3K/Akt signalling pathway and cancer. *Cancer Treat Rev*, 30(2):193–204, 2004.