

**Représentation des interactions protéine/protéine dans
la cadre d'une méthode de modélisation de réseaux
génomiques**

Grégory Batt

► **To cite this version:**

Grégory Batt. Représentation des interactions protéine/protéine dans la cadre d'une méthode de modélisation de réseaux génomiques. RR-4382, INRIA. 2002. inria-00072206

HAL Id: inria-00072206

<https://hal.inria.fr/inria-00072206>

Submitted on 23 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Représentation des interactions protéine/protéine
dans la cadre d'une méthode de modélisation de
réseaux géniques*

Grégory Batt

N° 4382

Septembre 2001

THÈME 3



*Rapport
de recherche*

Représentation des interactions protéine/protéine dans la cadre d'une méthode de modélisation de réseaux géniques

Grégory Batt

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet HELIX

Rapport de recherche n° 4382 — Septembre 2001 — 21 pages

Résumé : Il est communément accepté que la plupart des propriétés intéressantes d'un organisme biologique émerge des interactions entre ses gènes, protéines, molécules messagères et autres constituants. En général, les interactions entre gènes sont complexes, faisant souvent intervenir des boucles de rétroactions positives et négatives. La simple juxtaposition de ces interactions ne permet pas de connaître les propriétés globales des systèmes étudiés. Il est alors nécessaire d'intégrer ces données dans un formalisme permettant la simulation du comportement dynamique du réseau d'interactions considéré. Dans ce contexte, une méthode de simulation et un outil informatique (GNA) basé sur cette méthode, ont été développés au sein de l'équipe HELIX. Cette méthode a été développée dans le but d'analyser un certain type de réseaux de gènes, ne contenant que des interactions ADN/protéine. Nous montrerons qu'il est également possible d'utiliser cette méthode pour analyser un autre type de réseaux de gènes, contenant également des interactions protéine/protéine. L'analyse d'un système de régulation complexe faisant intervenir ces deux types d'interactions, concernant l'initiation de la sporulation chez *B. subtilis*, a été réalisée.

Mots-clés : bioinformatique, réseau de régulation génique, interaction protéine/protéine, modélisation mathématique, simulation qualitative, phosphorelais, *B. subtilis*, GNA

Representation of Protein/Protein Interactions in the Framework of a Method for Modeling Genetic Networks

Abstract: Most interesting properties of biological organisms emerge from the interactions between its genes, proteins, messenger molecules, and other components. The structure of these networks of interactions is often complex, as they include positive and negative feedback loops. In order to deduce the global properties of the systems under study it is necessary to integrate the interactions in a formalism by means of which the dynamic behavior of the network of interactions can be simulated. To this end, a method for modeling and simulating genetic regulatory networks, as well as a computer tool (GNA) based on this method, have been developed in the HELIX group. This method has been directed at the analysis of a specific kind of genetic networks, in which only DNA/protein interactions occur. We will show that the method can also be used for analysing genetic networks additionally including protein/protein interactions. The analysis of a complex example in which both types of interactions occur -the initiation of sporulation in *B. subtilis*- was performed.

Key-words: bioinformatics, genetic regulatory network, protein/protein interaction, mathematic modeling, qualitative simulation, phosphorelay, *B. subtilis*, GNA

Table des matières

1	Introduction	4
2	Le contexte biologique.	4
2.1	Gènes et protéines.	4
2.2	Réseau de gènes.	5
3	Modélisation de réseaux d'interactions géniques.	6
3.1	Modèles d'équations différentielles.	6
3.2	Problème de modélisation : les interactions protéine/protéine.	7
4	Intégrer des interactions protéine/protéine dans le modèle.	8
4.1	Principe.	8
4.2	Modèle biologique simple du phosphorelais.	8
4.2.1	Modèle biologique des interactions protéine/protéine du phosphorelais.	8
4.2.2	Modèle cinétique.	9
4.2.3	Analyse du modèle cinétique.	9
4.2.4	Modélisation de l'activation de <i>spolIIA</i> en terme de fonctions en escalier.	10
4.3	Etude d'un modèle biologique réaliste du phosphorelais.	11
4.3.1	Modèle biologique des interactions protéine/protéine du phosphorelais.	11
4.3.2	Modèle cinétique.	12
4.3.3	Analyse du modèle cinétique.	13
4.3.4	Modélisation de l'activation de <i>spolIIA</i> en terme de fonctions en escalier.	14
5	Simulation de l'évolution dynamique du réseau de gènes étudié.	14
6	Discussion.	17
A	Calcul de l'état stable du phosphorelais (programme <i>Matlab</i>).	19
B	Fichiers d'entrée de VGNA.	20

1 Introduction

Les évolutions récentes de la biologie laissent entrevoir la possibilité d'accumuler rapidement de très grandes quantités d'informations sur les systèmes biologiques. Toutefois, il est devenu de plus en plus évident que les propriétés de ces systèmes ne découlent pas des propriétés de leurs composants individuels, mais *émergent* des interactions entre les différents composants. Une bonne illustration de ce phénomène est la prévision du comportement dynamique d'un réseau de gènes.

Effectivement, l'expression des gènes est contrôlée par d'autres gènes, via des boucles de rétroaction souvent complexes, ce qui rend difficile l'étude de tels systèmes. Or les réseaux de gènes étudiés sont souvent vastes. Des outils formels, ainsi qu'un support informatique, deviennent alors nécessaires.

Plusieurs techniques de modélisation et de simulation ont été développées [8, 19], mais présentent des problèmes lors de la mise à l'échelle sur des réseaux complexes. Aussi une méthode mathématique, ainsi qu'un outil informatique, permettant la modélisation et la simulation des larges réseaux d'interactions géniques ont été développés par Céline Hernandez, Hidde de Jong et Michel Page au sein du projet Helix à l'INRIA Rhône-Alpes. Cette méthode a été développée dans le but de modéliser un certain type de réseaux de gènes, ne contenant que des interactions ADN/protéine.

L'objectif de ce stage, réalisé au sein de l'équipe Helix est de tester la possibilité d'utiliser cette méthode pour représenter un autre type de réseaux de gènes, contenant également des interactions protéine/protéine.

Quelques notions biologiques simples mais fondamentales, qui permettront au lecteur non biologiste de comprendre la problématique et les solutions proposées seront présentées dans une première partie. Puis, nous étudierons la représentation des réseaux de gènes dans la méthode développée par l'équipe. Enfin, nous aborderons le problème de la représentation des interactions protéine/protéine dans le cadre de la méthode et nous proposerons une solution.

2 Le contexte biologique.

2.1 Gènes et protéines.

Une molécule d'ADN est une longue succession de nucléotides (adénine, cytosine, thymine et guanine). C'est le support de l'information génétique. On y distingue certaines zones, appelées *gènes*, caractérisées par leur séquence nucléotidique. On peut considérer qu'un gène contient l'information nécessaire à la synthèse d'une *protéine*. On dit qu'il *code* pour une protéine. La synthèse protéique se passe en deux étapes principales : la *transcription* -l'information passe de la forme ADN à la forme ARN- et la *traduction* -l'ARN est utilisé pour synthétiser une nouvelle protéine.

Enfin on peut également distinguer, à proximité des gènes, d'autres zones, également caractérisées par leur séquence, appelées *éléments de régulation*, qui contrôlent l'*expression* d'un gène. Effectivement, en fonction des besoins de l'organisme, un gène particulier peut être exprimé -la protéine correspondante est alors présente- ou réprimé -la protéine correspondante est alors absente ou en très faible concentration.

Les protéines sont des macro-molécules de forme globulaire. Elles ont de nombreuses fonctions, comme catalyser une réaction, contrôler l'expression de gènes ou l'activité d'autres protéines, etc...

Les protéines, qui contrôlent l'expression de gènes sont appelées *protéines régulatrices*. Elles se lient à l'ADN au niveau d'*éléments de régulation*, favorisant ou empêchant l'expression du gène concerné. Il s'agit donc d'*interactions ADN/protéine*.

Souvent les protéines existent sous deux formes, active et inactive. La régulation du passage d'une forme à l'autre est souvent contrôlée par une autre protéine. Il s'agit alors d'*interactions protéine/protéine*.

La Fig. 1 illustre, sur un exemple, les différentes interactions pouvant avoir lieu [12]. Lorsque la protéine AbrB se lie sur l'élément de régulation du gène *sinR*¹, celui-ci est activé et la protéine SinR est synthétisée, sous forme active. Si de plus, la protéine SinI est présente, elle se lie à SinR et l'inactive.

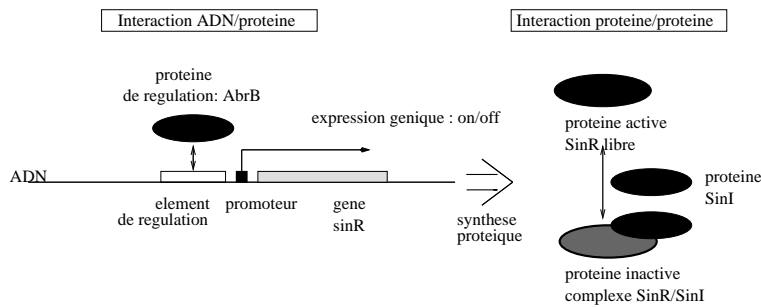


FIG. 1 – Exemple (simplifié) du contrôle du gène *sinR* par la protéine AbrB -par des interactions ADN/protéine- et du contrôle de la protéine SinR par la protéine SinI -par des interactions protéine/protéine.

2.2 Réseau de gènes.

Nous avons vu qu'un gène code pour une protéine, qui peut alors contrôler l'expression d'un autre gène. Les gènes se régulent donc mutuellement. D'où la notion de *réseau de gènes*, que l'on définit comme un ensemble de gènes ainsi que leurs interactions mutuelles.

Nous allons dans la suite, nous intéresser à un réseau de gènes particulier à travers lequel une bactérie peut réagir aux changements de son milieu de vie.

Bacillus subtilis est une bactérie qui, si les conditions de vie sont favorables, se multiplie et, qui dans le cas contraire, peut devenir une spore, qui est une forme de résistance, en attendant que les conditions de vie s'améliorent à nouveau. On parle alors de phénomènes de *multiplication* et de *sporulation* (Fig. 2) [10].

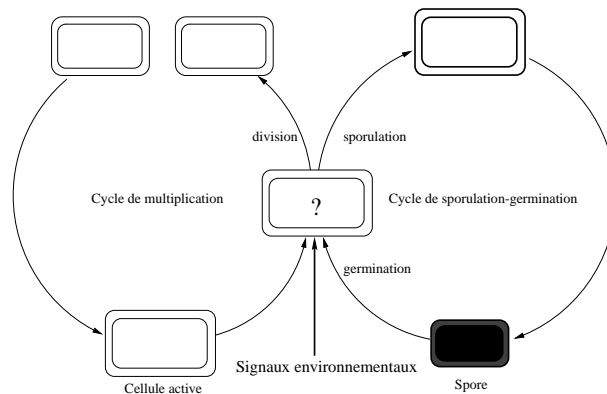


FIG. 2 – Les réactions possibles de *B.subtilis* aux variations du milieu de vie [11].

L'équipe Hélix, en collaboration avec le biologiste J. Geiselmann, s'est intéressée à l'étude de la régulation de l'*initiation* de la sporulation chez *B.subtilis*, phénomène dont la régulation est très complexe, notamment à cause de la présence de nombreuses boucles de rétroaction. De plus, ce système est un système modèle de régulation, qui a fait l'objet de nombreuses études. Un modèle biologique du réseau de gènes impliqués dans la prise de décision est présenté Fig. 3 [3].

Une succession d'interactions protéine/protéine, le *phosphorelais*, qui intègre de nombreux types de signaux, y est centrale (cf Fig. 3). Dans des conditions favorisant l'entrée en sporulation, le phosphorelais transfère des phosphates sur la protéine Spo0A. Lorsque la concentration cellulaire de Spo0A phosphorylée (notée Spo0A~P) atteint un certain seuil, un ensemble de gènes est activés, ce qui entraîne la cellule dans la voie de la sporulation. Une protéine, σ^F , activée par Spo0A~P, peut être considérée comme un *marqueur* de l'entrée

¹Noter la convention d'écriture utilisée permettant de distinguer les gènes des Protéines.

en sporulation, car une concentration cellulaire élevée en celle-ci est caractéristique d'un engagement dans la voie de sporulation

Les signaux qui arrivent au niveau du phosphorelais doivent être amplifiés et stabilisés pour permettre la phosphorylation de suffisamment de Spo0A et déclencher la sporulation. Des boucles de rétroaction positives et négatives sont nécessaires à cette stabilisation. L'entrée en sporulation n'est donc pas déterminée par un seul gène répondant à un seul signal, mais émerge de ce réseau complexe de régulations.

Dans le modèle biologique présenté Fig 3, on peut distinguer de nombreuses interactions ADN/protéine, comme AbrB inhibant *spo0E*, et des interactions protéine/protéine, au niveau du phosphorelais et entre SinR et SinI.

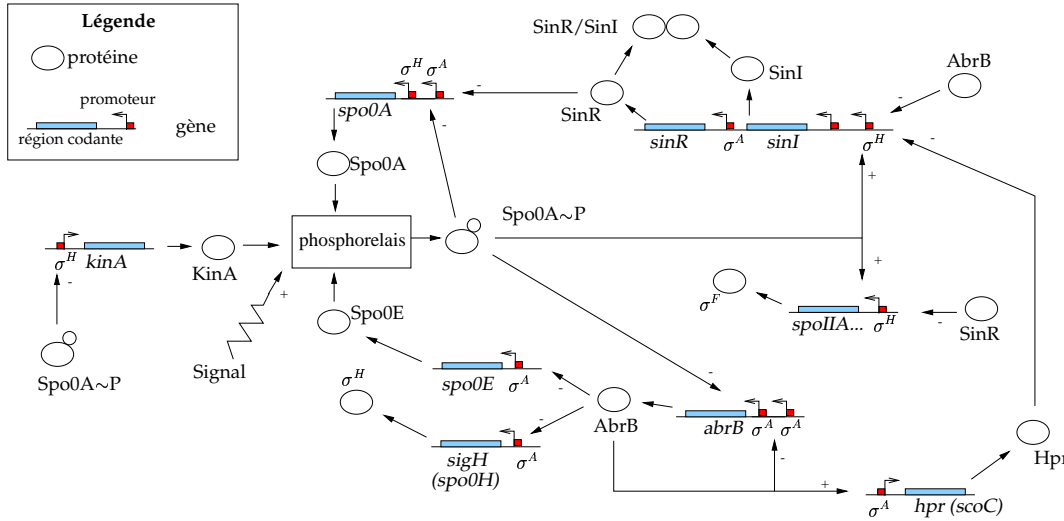


FIG. 3 – Réseau de régulation génique représentant les interactions régulant l'initiation de la sporulation chez *B. subtilis*. L'activation (ou l'inhibition) de l'expression d'un gène, par une protéine est indiqué par '+' (ou '-'). Les interactions ayant lieu dans le phosphorelais ne sont pas détaillées.[3]

Cet exemple illustre bien la complexité des systèmes d'interactions biologiques. La dynamique d'un tel système est très difficile à comprendre. Des outils formels et informatiques sont nécessaires.

3 Modélisation de réseaux d'interactions géniques.

3.1 Modèles d'équations différentielles.

La dynamique d'un réseau d'interactions géniques peut être modélisée par un système d'équations différentielles [13].

Dans la méthode utilisée [4], le système est caractérisé par un ensemble d'équations différentielles de la forme :

$$\dot{x}_i = f_i(\mathbf{x}) - g_i(\mathbf{x}) x_i, \quad x_i \geq 0, \quad 1 \leq i \leq n, \quad (1)$$

où $\mathbf{x} = [x_1, \dots, x_n]'$ est un vecteur de concentrations cellulaires de protéines. Les équations d'état (1) traduisent le fait que la variation de la concentration de la protéine i est la différence entre le taux de production $f_i(\mathbf{x})$ et le taux de dégradation $g_i(\mathbf{x}) x_i$ de cette protéine.

La fonction de production $f_i : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}$ est définie comme

$$f_i(\mathbf{x}) = \sum_{l \in L} \kappa_{il} b_{il}(\mathbf{x}), \quad (2)$$

où $\kappa_{il} > 0$ est une constante de taux, $b_{il}(\cdot)$ est une fonction de régulation, et L un ensemble d'indices de fonctions de régulation.

Une fonction de régulation $b_{il} : \mathbb{R}_{\geq 0}^n \rightarrow \{0, 1\}$ est définie en terme de fonctions en escalier s , et est l'équivalent d'une fonction booléenne exprimant la logique de la régulation du gène [4]. Une fonction en escalier représente l'effet d'une protéine régulatrice sur le gène.

Prenons un exemple. L'expression du gène *hpr* est activée par la présence de la protéine AbrB (cf Fig. 3 et Fig. 4(a)). Le taux de production est égal à (3), où $b_{il}()$ est réduite à une seule fonction en escalier.

$$f_{hr}(\mathbf{x}) = \kappa_{hr} s^+(x_{ab}, \theta_{ab}) \quad (3)$$

avec les abréviations *ab* et *hr* pour *AbrB* et *hpr*, x_{ab} représentant la concentration en *AbrB*, et

$$s^+(x_{ab}, \theta_{ab}) = \begin{cases} 0, & x_{ab} < \theta_{ab}, \\ 1, & x_{ab} > \theta_{ab}. \end{cases} \quad (4)$$

Ceci traduit qu'au dessous d'un certain seuil de concentration $\theta_{AbrB} > 0$ le gène *hpr* n'est pas exprimé, et qu'au dessus de ce seuil, la protéine Hpr est produite avec un taux κ_{hpr} . Pour représenter les inhibitions, on utilise des fonctions en escalier s^- telles que $s^-(x_{ab}, \theta_{ab}) = 1 - s^+(x_{ab}, \theta_{ab})$.

Des fonctions de régulation plus complexes peuvent exprimer des effets combinés de plusieurs protéines régulatrices.

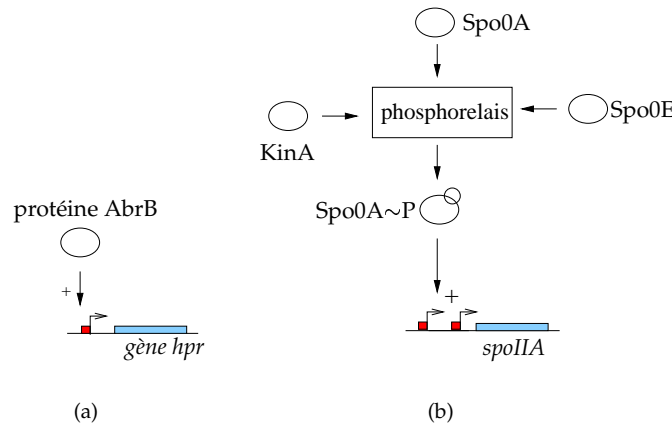


FIG. 4 – (4(a)) Activation du gène *hpr* par la protéine *AbrB*. (4(b)) Activation du gène *spoIIA* par la protéine *Spo0A* activée par phosphorylation, *Spo0A~P*.

La fonction de dégradation $g_i()$ est définie de manière analogue à $f_i()$ (cf eq. 2) et permet de modéliser la régulation de la dégradation de la protéine. Les taux de dégradation sont notés γ (et non κ). La dégradation des protéines est rarement régulée. Nous la considérerons dans la suite du travail constante pour toutes les protéines.

Les définitions précédentes de $f_i()$ et de $g_i()$ sont telles que les équations d'état (1) sont *linéaires par morceaux*, puisque $f_i()$ et $g_i()$ sont constantes par morceaux sur le domaine de définition de (1).

Avec ces définitions, il est bien sûr possible de raisonner quantitativement, en instanciant les constantes et les valeurs de seuil. Mais la linéarité par morceaux des équations d'état (1) permet également de raisonner qualitativement, en imposant des contraintes aux constantes et aux valeurs de seuil. Permettre de raisonner qualitativement, dans un domaine où bien peu de valeurs sont connues avec précision, est un des attraits majeurs de cette méthode de modélisation.

3.2 Problème de modélisation : les interactions protéine/protéine.

Nous avons vu comment on peut modéliser, à l'aide d'équations différentielles linéaires par morceau, les interactions géniques.

Dans la présentation faite plus haut de la méthode, une importante ambiguïté demeure toutefois. Effectivement, on s'est gardé de préciser quelle concentration \mathbf{x} représentait. Or, nous l'avons dit, une protéine peut exister sous deux formes, l'une active, l'autre inactive. C'est notamment le cas pour la protéine *Spo0A* qui intervient lors de l'initiation de la sporulation chez *B. subtilis* (cf Fig. 4(b)). Seule la forme phosphorylée *Spo0A~P* est active.

Il faut alors distinguer la concentration en *protéine totale* de la concentration en *protéine active*. Pour clarifier cette situation, on peut adopter les notations $x_{i,total}$ et $x_{i,actif}$ pour préciser si l'on parle des concentrations de la protéine *I totale* ou *active*.

Or une hypothèse implicitement faite était que $\mathbf{x} = \mathbf{x}_{total}$. (1) devient donc explicitement :

$$\dot{x}_{i_{total}} = f_i(\mathbf{x}_{total}) - g_i(\mathbf{x}_{total}) x_{i_{total}}, \quad x_{i_{total}} \geq 0, \quad 1 \leq i \leq n, \quad (5)$$

Or la régulation d'un gène ne dépend bien sûr que des concentrations en protéines actives. Trouver les expressions de f_i et g_i , en fonction des concentrations en protéines totales, tout en conservant la linéarité par morceau de (1), n'est pas immédiat. Il est nécessaire de prendre en compte les interactions protéine/protéine.

Comment représenter, par exemple, l'effet de Spo0A~P sur le gène *spoIIA*, en fonction des concentrations totales en KinA, Spo0E et Spo0A (Fig. 4(b)) ? Trouver une telle représentation pour la fonction f_{spoIIA} dans le cas particulier de l'initiation de la sporulation est justement le sujet de mon stage.

Notons toutefois que connaître une représentation de l'effet de Spo0A~P sur le gène *spoIIA* permettrait également de connaître une représentation de l'effet de Spo0A~P sur tous les gènes que cette protéine contrôle.

4 Intégrer des interactions protéine/protéine dans le modèle.

4.1 Principe.

Commençons par remarquer que l'intégration des interactions protéine/protéine dans le modèle est triviale lorsque la protéine existe sous une seule forme (nécessairement active). On a alors $x_{actif} = x_{total}$. Le problème ne se pose donc que lorsque la protéine régulatrice est modifiée par une (ou plusieurs) autre(s) protéine(s).

Il faut alors faire un modèle cinétique détaillé du système composé des protéines interagissantes. Puis considérer ce système à l'équilibre. Et enfin utiliser ce modèle pour proposer une fonction de production f_i et de dégradation g_i , en fonction de \mathbf{x}_{total} .

Deux modélisations biologiques des interactions ayant lieu dans le phosphorelais ont été réalisées. La démarche présentée a été appliquée sur ces deux modèles du phosphorelais.

4.2 Modèle biologique simple du phosphorelais.

4.2.1 Modèle biologique des interactions protéine/protéine du phosphorelais.

Dans ce premier modèle biologique, certains aspects importants du phosphorelais sont omis. Toutefois malgré les simplifications faites, ce premier modèle conserve le principe de fonctionnement du phosphorelais : le ratio $x_{sa.p}/x_{sa}$, soit le ratio des concentrations en protéine active Spo0A~P et en protéine inactive Spo0A, est modulé par les actions antagonistes d'une kinase et d'une phosphatase, respectivement KinA et Spo0E [2].

L'action de la kinase est de phosphoryler Spo0A tandis que l'action de la phosphatase est de déphosphoryler Spo0A~P. On a donc les réactions :



Une flèche ici représente une *réaction biologique globale*.

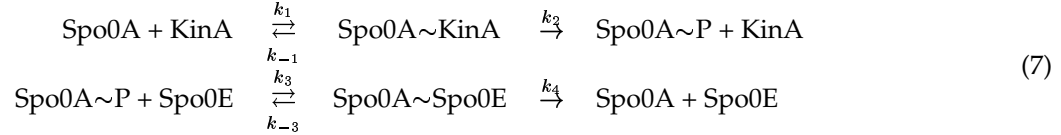
D'après les équations (6), il est évident que la concentration en Spo0A~P est une fonction croissante en fonction de la concentration en KinA_t, une fonction décroissante en fonction de la concentration en Spo0E_t. On peut aussi supposer qu'elle est une fonction croissante en fonction de la concentration en Spo0A_t, car il est probable que plus il y a de Spo0A_t, plus il y a de Spo0A~P.

On peut alors supposer que l'effet de Spo0A~P sur un gène, $s^+(x_{Spo0A~P}, \theta_{Spo0A~P})$, peut être approximé par le produit des fonctions $s^+(x_{Spo0A_t}, \theta_{Spo0A_t})$, $s^+(x_{KinA_t}, \theta_{KinA_t})$ et $s^-(x_{Spo0E_t}, \theta_{Spo0E_t})$ ². Si ceci est vrai, on a alors résolu le problème, puisque la fonction proposée n'est fonction que de concentrations en protéines totales. Testons cette hypothèse.

²ici, bien noter les signes, en exposant, des fonctions en escaliers

4.2.2 Modèle cinétique.

Un modèle cinétique, détaillant le type des interactions moléculaires ayant lieu entre les protéines est construit.



Précisons ici, que dans cette représentation, la sémantique des flèches est différente de (6). Une simple flèche signifie la présence d'une réaction totale et une double flèche signifie la présence d'une réaction réversible entre les entités moléculaires.

Ce modèle cinétique (7) est alors traduit en un système d'équations ((8)-(14)), en utilisant des propriétés élémentaires de cinétique chimique (conservation de la masse et ordre simple des réactions de (7)) [1].

$$\dot{x}_{sa} = -k_1 x_{sa} x_{ka} + k_{-1} x_{sa \cdot ka} + k_4 x_{sa \cdot se} \quad (8)$$

$$\dot{x}_{sa \cdot p} = k_{-3} x_{sa \cdot se} + k_2 x_{sa \cdot ka} - k_3 x_{sa \cdot p} x_{se} \quad (9)$$

$$\dot{x}_{sa \cdot ka} = k_1 x_{sa} x_{ka} - (k_{-1} + k_2) x_{sa \cdot ka} \quad (10)$$

$$\dot{x}_{sa \cdot se} = k_3 x_{sa \cdot p} x_{se} - (k_{-3} + k_4) x_{sa \cdot se} \quad (11)$$

$$x_{ka_t} = x_{ka} + x_{sa \cdot ka} \quad (12)$$

$$x_{se_t} = x_{se} + x_{sa \cdot se} \quad (13)$$

$$x_{sa_t} = x_{sa} + x_{sa \cdot p} + x_{sa \cdot ka} + x_{sa \cdot se} \quad (14)$$

On utilise les notations x_A et x_{A_t} pour représenter les concentrations en protéine A libre et totale et les indices ka, se, sa, sa·p, sa·ka, sa·se représentent les protéines KinA, Spo0E, Spo0A, Spo0A~P, Spo0A~KinA, Spo0A~Spo0E respectivement.

4.2.3 Analyse du modèle cinétique.

Afin de résoudre ce système de 7 équations non-linéaires, des hypothèses simplificatrices sont faites.

La quantité de substrat (ici Spo0A) lié à une enzyme (ici KinA et Spo0E) est négligeable par rapport à la quantité de substrat totale -*hypothèse 1*. Cette hypothèse permet de simplifier l'expression (14) en (15), puisque $x_{sa \cdot ka} + x_{sa \cdot se} \ll x_{sa} + x_{sa \cdot p}$.

$$x_{sa_t} = x_{sa} + x_{sa \cdot p} \quad (15)$$

Une enzyme, son substrat et le complexe enzyme/substrat sont à l'équilibre -*hypothèse 2*. Donc $d[\text{complexe enzyme/substrat}]/dt = 0$. On peut donc simplifier (10) et (11) en (16) et (17), puisque $\dot{x}_{sa \cdot ka} = 0$ et $\dot{x}_{sa \cdot se} = 0$.

$$0 = k_1 x_{sa} x_{ka} - (k_{-1} + k_2) x_{sa \cdot ka} \quad (16)$$

$$0 = k_3 x_{sa \cdot p} x_{se} - (k_{-3} + k_4) x_{sa \cdot se} \quad (17)$$

Ces deux hypothèses sont standards en cinétique enzymatique [1, 15]. On obtient un premier système simplifié (18)-(24).

$$\dot{x}_{sa} = -k_1 x_{sa} x_{ka} + k_{-1} x_{sa \cdot ka} + k_4 x_{sa \cdot se} \quad (18)$$

$$\dot{x}_{sa \cdot p} = k_{-3} x_{sa \cdot se} + k_2 x_{sa \cdot ka} - k_3 x_{sa \cdot p} x_{se} \quad (19)$$

$$0 = k_1 x_{sa} x_{ka} - (k_{-1} + k_2) x_{sa \cdot ka} \quad (20)$$

$$0 = k_3 x_{sa \cdot p} x_{se} - (k_{-3} + k_4) x_{sa \cdot se} \quad (21)$$

$$x_{ka_t} = x_{ka} + x_{sa \cdot ka} \quad (22)$$

$$x_{se_t} = x_{se} + x_{sa \cdot se} \quad (23)$$

$$x_{sa_t} = x_{sa} + x_{sa \cdot p} \quad (24)$$

On remarque alors qu'en utilisant les équations (20) et (21), on obtient $\dot{x}_{sa} + \dot{x}_{sa \cdot p} = 0$. On peut donc éliminer une des deux équations (18) ou (19). On élimine (18).

Egalement, on suppose que le système est à l'équilibre -hypothèse3. Effectivement la synthèse protéique fait intervenir un très grand nombre d'interactions protéine/protéine. Elle est donc lente par rapport aux interactions protéine/protéine étudiées. On suppose alors que les perturbations du système, dues à la synthèse protéique, sont suffisamment lentes pour que celui-ci atteigne très rapidement l'équilibre après chaque perturbations. On a donc : $\dot{x}_{sa \cdot p} = 0$. Donc (19) devient :

$$0 = k_{-3} x_{sa \cdot se} + k_2 x_{sa \cdot ka} - k_3 x_{sa \cdot p} x_{se} \quad (25)$$

On obtient finalement le système (26)-(31) :

$$0 = k_{-3} x_{sa \cdot se} + k_2 x_{sa \cdot ka} - k_3 x_{sa \cdot p} x_{se} \quad (26)$$

$$0 = k_1 x_{sa} x_{ka} - (k_{-1} + k_2) x_{sa \cdot ka} \quad (27)$$

$$0 = k_3 x_{sa \cdot p} x_{se} - (k_{-3} + k_4) x_{sa \cdot se} \quad (28)$$

$$x_{ka_i} = x_{ka} + x_{sa \cdot ka} \quad (29)$$

$$x_{se_i} = x_{se} + x_{sa \cdot se} \quad (30)$$

$$x_{sa_i} = x_{sa} + x_{sa \cdot p} \quad (31)$$

On peut alors montrer que ce système implique que la concentration en Spo0A~P vérifie l'équation du second degrés (32).

$$P(x_{sa}) = (k_4 x_{se_i} - k_2 x_{ka_i}) x_{sa}^2 + (k_2 x_{ka_i} (x_{sa_i} - K_{M_3}) - x_{se_i} k_4 (x_{sa_i} + K_{M_1})) x_{sa} + K_{M_3} k_2 x_{ka_i} x_{sa_i} = 0, \quad (32)$$

avec les constantes $K_{M_1} = (k_{-1} + k_2)/k_1$ et $K_{M_3} = (k_{-3} + k_4)/k_3$.

Il faut sélectionner parmi les deux racines du polynôme (32) une seule valeur, qui correspond effectivement à la concentration de Spo0A~P. Ceci n'a pas été possible analytiquement car aucun critère n'a été trouvé permettant de ne retenir qu'une valeur. Les constantes sont doncinstanciées à partir de données bibliographiques[5, 15].

constante	K_{M_1}	K_{M_3}	k_2	k_4
valeur	$0.13 \cdot 10^{-6} M$	$530 \cdot 10^{-6} M$	$29.5 s^{-1}$	$3.8 s^{-1}$

Deux critères sont utilisés pour sélectionner la solution correcte. Tout d'abord, la concentration en Spo0A~P trouvée doit bien sûr être dans l'intervalle $[0, x_{sa_i}]$ (Propriété 1). Egalement, seuls les points d'équilibre stable sont retenus, car ce sont les seuls intéressants d'un point de vue biologique (Propriété 2). La stabilité du système au voisinage du point d'équilibre est testée en utilisant la propriété que les valeurs propres de la matrice jacobienne, associée au système et évaluée en un point d'équilibre, doivent être toutes négatives ou nulles pour que ce point d'équilibre soit un point d'équilibre *stable* [9].

On applique alors la propriété précédente à notre système. Soit x_{sa}^* une solution de (32)-c'est donc un point d'équilibre-, alors la valeur propre de la matrice jacobienne -réduite ici à $(\frac{\partial \dot{x}_{sa \cdot p}}{\partial x_{sa \cdot p}})_{(x_{sa \cdot p}^*)}$ - doit être négative pour que cet équilibre soit stable.

Le système est résolu numériquement pour différentes valeurs de concentration en protéines KinA, Spo0E et Spo0A *totales* (voir programme en Annexe A). On trouve toujours une seule solution ayant les propriétés 1 et 2.

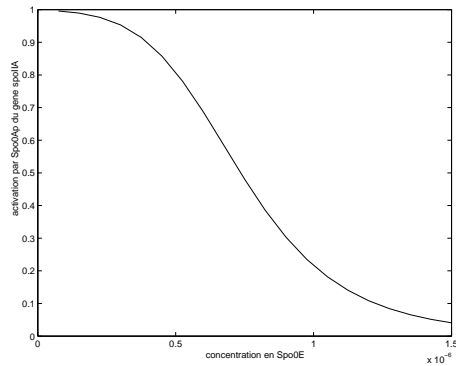
On cherche à exprimer l'activité du gène *spoIIA* en fonction des concentrations en protéines totales. On sait calculer la concentration de Spo0A~P, en fonction des concentrations en protéines totales. On cherche donc à exprimer l'activité du gène *spoIIA* en fonction de Spo0A~P, $s^+(x_{sa \cdot p}, \theta_{sa \cdot p})$. On a montré que l'activation d'un gène par une protéine régulatrice était le plus souvent fonction sigmoïde de la concentration en cette protéine [20]. On applique donc un filtre sigmoïde sur la valeur trouvée pour $x_{sa \cdot p}$ (33), obtenant ainsi une mesure de l'activation par Spo0A~P du gène *spoIIA*.

$$f_{spoIIA}(x_{sa \cdot p}) = \kappa_{spoIIA} \frac{x_{sa \cdot p}^n}{x_{sa \cdot p}^n + \theta_{spoIIA}^n} \quad (33)$$

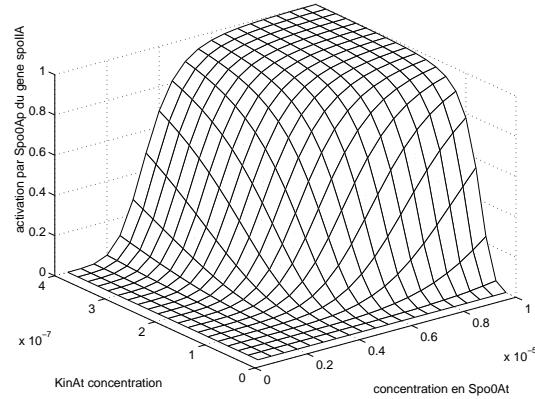
avec θ_{spoIIA} , une concentration seuil.

4.2.4 Modélisation de l'activation de *spoIIA* en terme de fonctions en escalier.

On utilise alors une représentation graphique, pour étudier l'influence des variations d'une concentration en protéine (graphes en deux dimensions) et l'influence des variations conjuguées de plusieurs concentrations en protéine (graphes en trois dimensions) sur cette mesure d'activation.



(a) graphe en 2D, montrant l'influence de la variation de concentration en $Spo0E_t$ sur l'activation du gène $SpoIIA$ par $Spo0A\sim P$.



(b) graphe en 3D, montrant l'influence de la variation conjuguée des concentrations en $KinA_t$ et $Spo0A_t$ sur l'activation du gène $SpoIIA$ par $Spo0A\sim P$.

FIG. 5 – Analyse de l'activation par $Spo0A\sim P$ d'un gène. (a) en fonction la concentration en protéine $Spo0E$ totale, (b) en fonction des concentrations en protéines $KinA$ et $Spo0A$ totales.

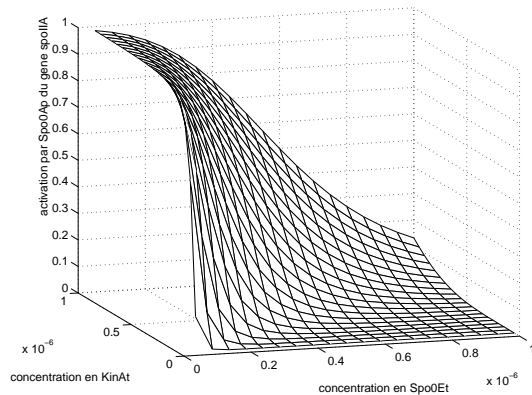


FIG. 6 – graphe montrant une coopérativité entre les protéines $KinA$ et $Spo0E$.

On souhaite obtenir des courbes d'allures sigmoïdales dans tous les graphes, en deux et trois dimensions.

Les graphes obtenus sont conformes à ce que l'on attendait d'après notre hypothèse de départ (Fig. 5). Notons toutefois une exception. Pour un couple de protéines ($KinA$ et $Spo0E$), la surface obtenue, bien que d'allure sigmoïde, n'a pas la forme d'un produit de deux sigmoïdes, ce qui n'est pas tout à fait cohérent avec l'hypothèse de départ (Fig. 6). Ceci traduit une certaine coopérativité des protéines, non prise en compte dans l'hypothèse proposée.

La solution initialement proposée semble donc effectivement être une approximation satisfaisante de la fonction cherchée.

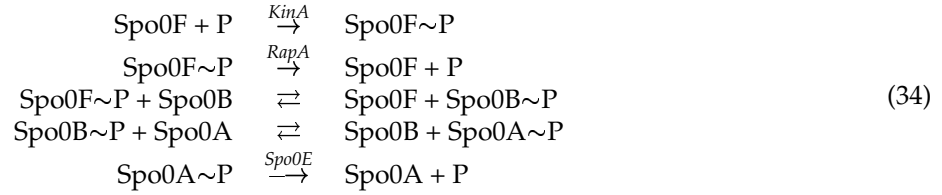
Toutefois, le modèle biologique des interactions protéine/protéine du phosphorelais présenté n'est pas satisfaisant du point de vue biologique. Avec les simplifications faites lors de la modélisation biologique, on ne peut être sûr que tous les aspects importants du fonctionnement du phosphorelais sont pris en compte. Un modèle plus détaillé a alors été fait et la démarche présentée précédemment a, à nouveau, été mise à l'épreuve.

4.3 Etude d'un modèle biologique réaliste du phosphorelais.

4.3.1 Modèle biologique des interactions protéine/protéine du phosphorelais.

Le phosphorelais est en fait une cascade de phosphorylations. Une kinase phosphoryle un premier substrat ($Spo0F$), puis celui-ci transmet son nouveau phosphate à un substrat intermédiaire ($Spo0B$), qui le transmet

à nouveau au substrat terminal (Spo0A). D'où le nom de *phosphorelais*. Des phosphatases (RapA et Spo0E) déphosphorylent les substrats aux deux extrémités de la cascade, permettant ainsi une stabilisation du flux de phosphates[6, 14, 11]. Le modèle suivant est donc un modèle réaliste du phosphorelais :

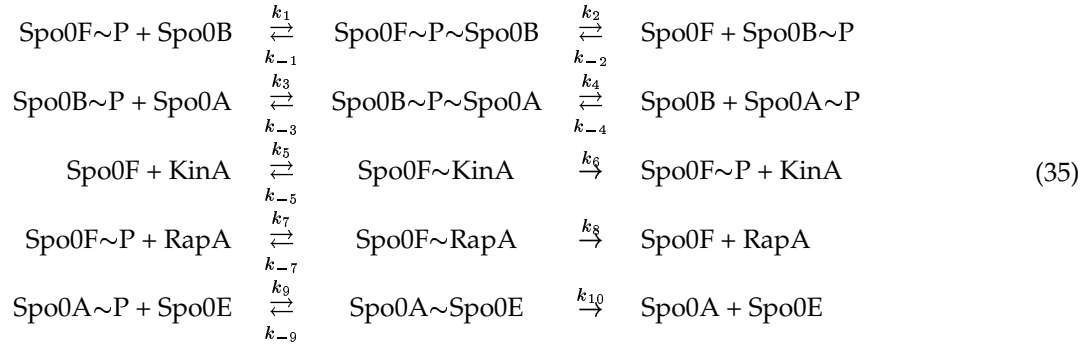


Ces équations montrent donc que dans le modèle du phosphorelais précédent, les trois espèces Spo0F, Spo0B et Spo0A sont confondues. On ne peut donc pas non plus distinguer les deux phosphatases rapA et Spo0E.

A partir des effets des protéines sur le système, et en s'inspirant des résultats précédents, on peut faire l'hypothèse que la concentration en Spo0A~P est très certainement une fonction croissante en fonction des concentrations en KinA_t, Spo0F_t, Spo0B_t et Spo0A_t, et une fonction décroissante en fonction des concentrations en RapA_t et Spo0E_t. On peut alors supposer que l'effet de Spo0A~P sur un gène, $s^+(x_{\text{Spo0A}\sim\text{P}}, \theta_{\text{Spo0A}\sim\text{P}})$, peut être approximé par le produit des fonctions $s^+(x_{\text{KinA}_t}, \theta_{\text{KinA}_t})$, $s^+(x_{\text{Spo0F}_t}, \theta_{\text{Spo0F}_t})$, $s^+(x_{\text{Spo0B}_t}, \theta_{\text{Spo0B}_t})$, $s^+(x_{\text{Spo0A}_t}, \theta_{\text{Spo0A}_t})$, et $s^-(x_{\text{RapA}_t}, \theta_{\text{RapA}_t})$, $s^-(x_{\text{Spo0E}_t}, \theta_{\text{Spo0E}_t})$. Testons cette hypothèse.

4.3.2 Modèle cinétique.

Comme précédemment, un modèle cinétique détaillé est construit à partir des équations biologiques globales (34), puis traduit en un système de 17 équations non-linéaires, donné ci-dessous ((36)-(52)).



$$\dot{x}_{sf} = -k_5 x_{sf} x_{ka} + k_{-5} x_{sf \cdot ka} + k_8 x_{sf \cdot ra} - k_{-2} x_{sf} x_{sb \cdot p} + k_2 x_{sf \cdot p \cdot sb} \quad (36)$$

$$\dot{x}_{sf \cdot p} = k_{-7} x_{sf \cdot ra} - k_7 x_{sf \cdot p} x_{ra} + k_6 x_{sf \cdot ka} - k_1 x_{sf \cdot p} x_{sb} + k_{-1} x_{sf \cdot p \cdot sb} \quad (37)$$

$$\dot{x}_{sb} = k_{-1} x_{sf \cdot p \cdot sb} - k_1 x_{sf \cdot p} x_{sb} + k_4 x_{sb \cdot p \cdot sa} - k_{-4} x_{sb} x_{sa \cdot p} \quad (38)$$

$$\dot{x}_{sb \cdot p} = k_2 x_{sf \cdot p \cdot sb} - k_{-2} x_{sf} x_{sb \cdot p} + k_{-3} x_{sb \cdot p \cdot sa} - k_3 x_{sb \cdot p} x_{sa} \quad (39)$$

$$\dot{x}_{sa} = k_{-3} x_{sb \cdot p \cdot sa} - k_3 x_{sb \cdot p} x_{sa} + k_{10} x_{sa \cdot se} \quad (40)$$

$$\dot{x}_{sa \cdot p} = k_4 x_{sb \cdot p \cdot sa} - k_{-4} x_{sb} x_{sa \cdot p} - k_9 x_{sa \cdot p} x_{se} + k_{-9} x_{sa \cdot se} \quad (41)$$

$$\dot{x}_{sf \cdot p \cdot sb} = k_1 x_{sf \cdot p} x_{sb} + k_{-2} x_{sf} x_{sb \cdot p} - (k_{-1} + k_2) x_{sf \cdot p \cdot sb} \quad (42)$$

$$\dot{x}_{sb \cdot p \cdot sa} = k_3 x_{sb \cdot p} x_{sa} + k_{-4} x_{sb} x_{sa \cdot p} - (k_{-3} + k_4) x_{sb \cdot p \cdot sa} \quad (43)$$

$$\dot{x}_{sf \cdot ka} = k_5 x_{sf} x_{ka} - (k_{-5} + k_6) x_{sf \cdot ka} \quad (44)$$

$$\dot{x}_{sf \cdot ra} = k_7 x_{sf \cdot p} x_{ra} - (k_{-7} + k_8) x_{sf \cdot ra} \quad (45)$$

$$\dot{x}_{sa \cdot se} = k_9 x_{sa \cdot p} x_{se} - (k_{-9} + k_{10}) x_{sa \cdot se} \quad (46)$$

$$x_{ka_t} = x_{ka} + x_{sf \cdot ka} \quad (47)$$

$$x_{ra_t} = x_{ra} + x_{sf \cdot ra} \quad (48)$$

$$x_{se_t} = x_{se} + x_{sa \cdot se} \quad (49)$$

$$x_{sf_t} = x_{sf} + x_{sf \cdot p} + x_{sf \cdot ka} + x_{sf \cdot ra} \quad (50)$$

$$x_{sb_t} = x_{sb} + x_{sb \cdot p} \quad (51)$$

$$x_{sa_t} = x_{sa} + x_{sa \cdot p} + x_{sa \cdot se} \quad (52)$$

On utilise, en plus des notations précédentes, les indices sf, sb, ra, p pour les protéines Spo0F, Spo0B, RapA, et le groupement phosphate, respectivement, ainsi que la notation $a \cdot b$, pour le complexe $A \sim B$.

4.3.3 Analyse du modèle cinétique.

Ce système est alors simplifié, en utilisant les hypothèses 1, 2 et 3 présentées section 4.2. On obtient le système ((53)-(66)), réduit à 14 équations indépendantes. Les hypothèses utilisées sont simplement citées en face des équations.

$$0 = -k_5 x_{sf} x_{ka} + k_{-5} x_{sf \cdot ka} + k_8 x_{sf \cdot ra} - k_{-2} x_{sf} x_{sb \cdot p} + k_2 x_{sf \cdot p \cdot sb} \quad (\text{hypothese3}) \quad (53)$$

$$0 = k_{-1} x_{sf \cdot p \cdot sb} - k_1 x_{sf \cdot p} x_{sb} + k_4 x_{sb \cdot p \cdot sa} - k_{-4} x_{sb} x_{sa \cdot p} \quad (\text{hypothese3}) \quad (54)$$

$$0 = k_4 x_{sb \cdot p \cdot sa} - k_{-4} x_{sb} x_{sa \cdot p} - k_9 x_{sa \cdot p} x_{se} + k_{-9} x_{sa \cdot se} \quad (\text{hypothese3}) \quad (55)$$

$$0 = k_1 x_{sf \cdot p} x_{sb} + k_{-2} x_{sf} x_{sb \cdot p} - (k_{-1} + k_2) x_{sf \cdot p \cdot sb} \quad (\text{hypothese2}) \quad (56)$$

$$0 = k_3 x_{sb \cdot p} x_{sa} + k_{-4} x_{sb} x_{sa \cdot p} - (k_{-3} + k_4) x_{sb \cdot p \cdot sa} \quad (\text{hypothese2}) \quad (57)$$

$$0 = k_5 x_{sf} x_{ka} - (k_{-5} + k_6) x_{sf \cdot ka} \quad (\text{hypothese2}) \quad (58)$$

$$0 = k_7 x_{sf \cdot p} x_{ra} - (k_{-7} + k_8) x_{sf \cdot ra} \quad (\text{hypothese2}) \quad (59)$$

$$0 = k_9 x_{sa \cdot p} x_{se} - (k_{-9} + k_{10}) x_{sa \cdot se} \quad (\text{hypothese2}) \quad (60)$$

$$x_{ka_t} = x_{ka} + x_{sf \cdot ka} \quad (61)$$

$$x_{ra_t} = x_{ra} + x_{sf \cdot ra} \quad (62)$$

$$x_{se_t} = x_{se} + x_{sa \cdot se} \quad (63)$$

$$x_{sf_t} = x_{sf} + x_{sf \cdot p} \quad (\text{hypothese1}) \quad (64)$$

$$x_{sb_t} = x_{sb} + x_{sb \cdot p} \quad (65)$$

$$x_{sa_t} = x_{sa} + x_{sa \cdot p} \quad (\text{hypothese1}) \quad (66)$$

Ce système est alors réduit en un système de 3 équations ((67)-(69)).

$$K_{M_1} (x_{sf_t} - x_{sf}) x_{sb} - K_{M_{-1}} x_{sf} (x_{sb_t} - x_{sb}) - \frac{K_{M_5} x_{ka_t} x_{sf}}{1 + \frac{K_{M_5} x_{sf}}{k_6}} + \frac{K_{M_7} x_{ra_t} (x_{sf_t} - x_{sf})}{1 + \frac{K_{M_7} (x_{sf_t} - x_{sf})}{k_8}} = 0 \quad (67)$$

$$K_{M_1} (x_{sf_t} - x_{sf}) x_{sb} - K_{M_{-1}} x_{sf} (x_{sb_t} - x_{sb}) - K_{M_3} (x_{sa_t} - x_{sa \cdot p}) (x_{sb_t} - x_{sb}) + K_{M_{-3}} x_{sa \cdot p} x_{sb} = 0 \quad (68)$$

$$-K_{M_3} (x_{sa_i} - x_{sa_p}) (x_{sb_i} - x_{sb}) + K_{M_{-3}} x_{sa_p} x_{sb} + \frac{K_{M_9} x_{sa_p} x_{se_i}}{1 + \frac{K_{M_9} x_{sa_p}}{k_{10}}} = 0 \quad (69)$$

où l'on définit les constantes K_{M_i} comme valant $(k_{-i} + k_{i+1})/k_i$.

La résolution de ce système est alors possible en utilisant le logiciel *Maple*. On obtient un (énorme) polynôme du cinquième degrés en x_{sa_p} , la concentration en Spo0A~P.

Une résolution analytique n'étant pas possible, on cherche une solution numérique, après avoir instancié les constantes utilisées, en fonction des données bibliographiques disponibles [5, 15].

La sélection de "la" bonne racine est faite comme précédemment, à partir de critères portant sur l'appartenance au domaine de concentration et la stabilité du système au point d'équilibre. Le polynôme est résolu pour différentes valeurs de concentrations en protéines totales et les solutions vérifiant les deux conditions ci-dessus sont sélectionnées. Pour tous les polynômes étudiés, une seule solution est toujours retenue. Cette valeur est alors passée dans un filtre sigmoïde (33) pour obtenir une mesure de l'activation par Spo0A~P d'un gène.

4.3.4 Modélisation de l'activation de *spoIIA* en terme de fonctions en escalier.

Afin d'étudier cette activation, une étude graphique est faite en deux dimensions pour toutes les concentrations en protéines totales et en trois dimensions pour tous les couples de concentrations, et pour différentes valeurs de concentration en protéines non représentées.

Comme attendu d'après l'hypothèse de départ, on obtient bien des courbes sigmoïdes, pour les graphes en deux dimensions (non représenté).

L'analyse des influences des variations conjuguées de plusieurs concentrations en protéine (graphes en trois dimensions) met en évidence que la majorité des courbes obtenues sont cohérentes avec l'hypothèse proposée (Fig 7). Ainsi, approximer le comportement global du système par le produit de fonction en escalier proposé semble tout à fait raisonnable. Toutefois, le système présente deux caractéristiques non prises en compte dans l'hypothèse proposée. Il y a comme, observé précédemment, des coopérativités entre certaines protéines, mais également, on observe une certaine redondance entre les actions des deux phosphatases, RapA et Spo0E (Fig 8).

5 Simulation de l'évolution dynamique du réseau de gènes étudié.

Le contrôle transcriptionnel des trois gènes supplémentaires introduits dans la seconde modélisation du phosphorelais a été intégré dans le modèle biologique global de l'initiation de la phosphorylation précédemment construit (cf Fig. 9) [18, 17, 7, 16].

Les résultats obtenus section 4 permettent alors de modéliser complètement le réseau de gènes proposé, en respectant le formalisme de la méthode présentée. Il est possible de simuler l'évolution dynamique du système.

Effectivement un outil, *Visual Genetic Network Analyser* a été développé au sein de l'équipe, permettant, à partir d'un modèle d'un système biologique et de conditions initiales, de suivre les évolutions possibles des différentes concentrations en protéines et d'analyser le comportement dynamique du système (identification d'attracteurs, de leurs bassins, ...).

Ces simulations donnent des résultats tout à fait satisfaisant, puisque les variations des concentrations en protéines trouvées sont en accord avec les données bibliographiques disponibles, que les conditions initiales soient des conditions déclenchant la sporulation ou non, et ce quelle que soit la modélisation biologique du phosphorelais utilisée. Cela valide donc, sur un exemple assez complexe, la méthode présentée.

Le second modèle, plus complet permet toutefois d'obtenir d'autres résultats intéressants. Effectivement *B. subtilis* peut présenter un autre comportement lorsque les conditions de vie se détériorent, appelé *acquisition de compétence*, que nous n'avons encore pas évoqué. C'est, comme la sporulation, une réaction à un changement du milieu de vie. Les deux voies de décision (sporulation et compétence) ne sont pas indépendantes [6]. Donc le comportement réel n'est pas simplement binaire, croissance ou sporulation, comme nous l'avons présenté jusqu'à maintenant. De plus, l'expression d'un des gènes ajouté au modèle plus complexe, *rapA*, est contrôlée par un acteur principal de la décision d'acquisition de la compétence, *comA*.

Il est ainsi possible de simuler, de manière élémentaire, une prise de décision entre *sporulation* et *compétence*. Quand les conditions initiales sont des conditions déclenchant soit la sporulation uniquement, soit l'acquisition de la compétence uniquement, le comportement du système est cohérent avec les données bibliographiques.

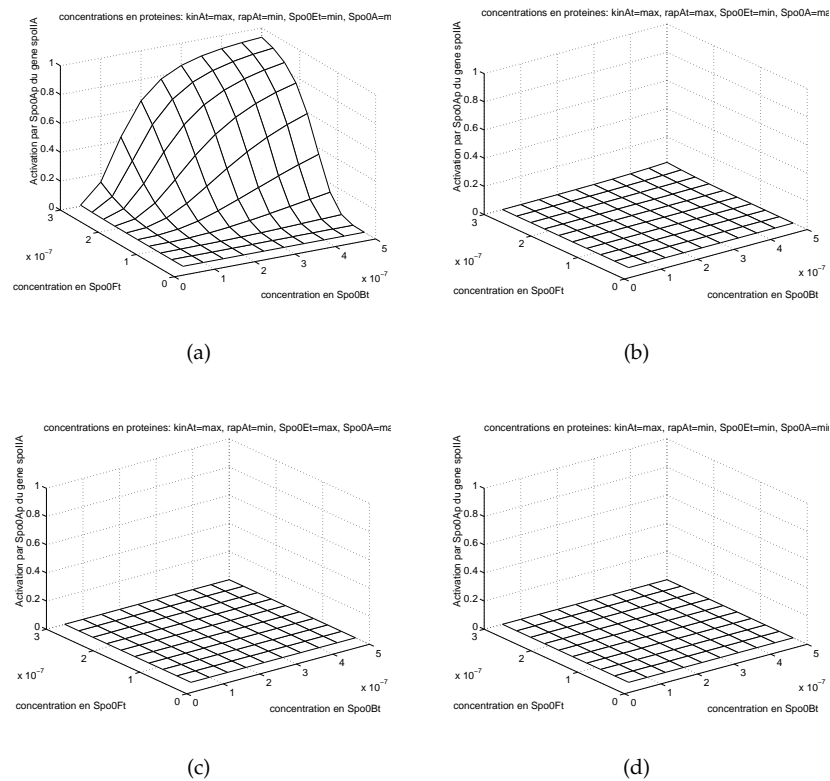


FIG. 7 – Exemple de l'activation par Spo0A~P du gène *SpoIIA*, en fonction des concentrations en Spo0F_t et Spo0B_t. Les quantités des autres protéines sont données au-dessus de chaque graphe, avec *max* pour maximale, et *min* pour minimale. 7(a) : toutes les protéines autres que Spo0F et Spo0B sont en quantité requise pour pouvoir observer une activation du gène. 7(b), 7(c) and 7(d) : la phosphatase rapA_t est présente, la phosphatase Spo0E_t est présente, et le substrat Spo0A_t est absent, respectivement ; aucune activation n'est observée.

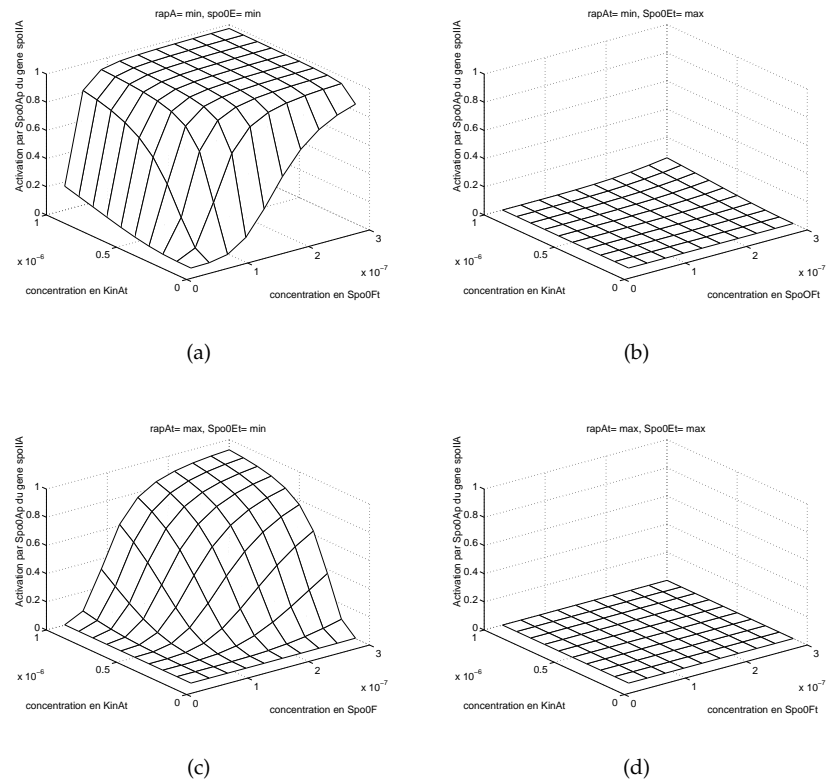


FIG. 8 – Exemple de l’activation par $Spo0A\sim P$ du gène $SpoIIA$, en fonction des concentrations en $KinA_t$ et $Spo0F_t$. 8(a), 8(b), 8(d) : graphes cohérents avec l’hypothèse émise. 8(c) : graphe montrant une certaine redondance entre les deux phosphatases, puisque l’absence d’une phosphatase ($Spo0E$) n’entraîne pas l’absence d’activation. Les concentrations en $RapA_t$ et $Spo0E_t$ sont données au-dessus de chaque graphe. Les concentrations en $Spo0B_t$ and $Spo0A_t$ sont : $[Spo0B_t] = max$ and $[Spo0A_t] = max$.

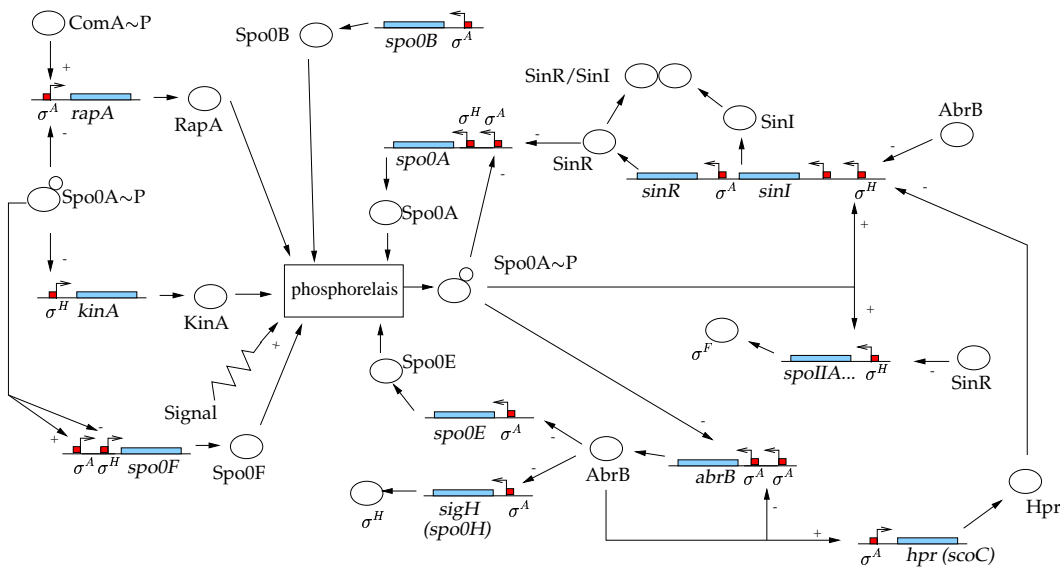


FIG. 9 – Modèle biologique étendu, construit à partir de celui représenté Fig. 3. Les interactions ayant lieu dans le phosphorelais ne sont pas détaillées.[18, 17, 7, 16]

Quand les conditions initiales sont telles que les deux comportements sont a priori possibles, aucune décision n'est semblé-t'il prise, puisque les deux évolutions sont possibles, d'après les prédictions du simulateur. Aucune donnée bibliographique, décrivant le comportement réellement observé dans ces conditions, n'a été trouvée.

6 Discussion.

Il convient tout d'abord de revenir sur la validité de certaines hypothèses.

Le formalisme de modélisation ne permet de prendre en compte que les gènes et les produits de leur expression, les protéines. C'est bien sûr extrêmement réducteur car de nombreuses autres molécules peuvent intervenir dans les réactions biologiques. Toutefois, les situations les plus intéressantes biologiquement sont celles où a lieu une *intégration* de l'information, comme dans le cas présenté. Or l'intégration d'information se fait essentiellement par le jeu d'interactions ADN/protéine et protéine/protéine. On ne peut donc pas tout représenter, mais on peut certainement représenter l'essentiel.

Biologiquement, l'activation (ou l'inhibition) d'un gène en fonction de la concentration en protéine régulatrice est souvent une fonction sigmoïdale [20]. L'approximer par une fonction en escalier semble alors cohérente. Mais des interactions protéine/protéine complexes ne sont représentables que par des fonctions sigmoïdales complexes. Or l'approximation de fonctions sigmoïdales complexes par des fonctions en escalier est assez grossière. Les approximations seront donc de plus en plus grossières avec la complexité croissante des systèmes. Ceci peut éventuellement devenir une limitation de la complexité des interactions protéine/protéine représentables.

Enfin, les résultats obtenus lors de la simulation de l'évolution du système, lorsque les conditions de vie simulées favorisent la sporulation *et* l'acquisition de compétence, sont tout particulièrement intéressants. Effectivement, le simulateur n'a pas donné une réponse unique et a priori, les deux comportements sont possibles. On peut émettre trois hypothèses. Si biologiquement, on observe effectivement les deux comportements dans ces conditions, la réponse du simulateur est pertinente. Si un seul comportement est observé, on peut alors à remettre en cause le modèle (rappelons que le modèle réalisé était centré sur la prise de décision entre croissance et sporulation, et non pas entre compétence et sporulation). Mais on peut aussi supposer que la *précision* de la méthode, *qualitative*, est trop faible pour pouvoir trancher et prédire seulement le comportement réel.

L'objectif principal de ce travail était de montrer la possibilité de prendre en compte des interactions protéine/protéine complexes dans le cadre de la méthode de modélisation développée dans l'équipe Helix, en conservant le formalisme. Des résultats encourageants ont été obtenus, bien que ce travail ne prouve pas que cela sera toujours possible. De plus, un travail similaire à celui-ci doit être fait à chaque fois que des interactions protéine/protéine devront être prises en compte. Toutefois, la démarche utilisée dans ce travail pourra elle être réutilisée.

Remerciements :

Je tiens à remercier l'INRIA, et particulièrement François Rechenmann, de m'avoir accueilli dans ses murs.

Je remercie aussi toute l'équipe Helix pour son accueil chaleureux.

Merci également à tout ceux qui ont participé à l'élaboration de VGNA, et à Johannes Geiselmann, pour avoir répondu à toutes mes sollicitations.

Un grand merci enfin à Hidde de Jong pour toute l'attention qu'il a bien voulu me consacrer et pour tous ses conseils.

Références

- [1] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, London, revised edition, 1995.
- [2] H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in *B. subtilis*. In L. Duret, C. Gaspin, and T. Schiex, editors, *Recueil des Actes des Journée Ouvertes Biologie Informatique Mathématiques, JOBIM 2001*, pages 187–194, Toulouse, 2001.
- [3] H. de Jong, M. Page, C. Hernandez, and J. Geiselmann. Qualitative simulation of genetic regulatory networks : Method and application. In G. Biswas, editor, *Proceedings of the Fifteenth International Workshop on Qualitative Reasoning, QR-01*, pages 134–141, San Antonio, TX, 2001.

- [4] H. de Jong, M. Page, C. Hernandez, and J. Geiselmann. Qualitative simulation of genetic regulatory networks : Method and application. In B. Nebel, editor, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI-01*, pages 67–73, San Mateo, CA, 2001. Morgan Kaufmann.
- [5] C. Grimshaw, S. Huang, C. Hanstein, M. Strauch, D. Burbulys, L. Wang, J. Hoch, and J. Whiteley. Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in *Bacillus subtilis*. *Biochemistry*, 37 :1365–1375, 1998.
- [6] A. Grossman. Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annual Review of Genetics*, 29 :477–508, 1995.
- [7] W. Haldenwang. The sigma factors of *Bacillus subtilis*. *Microbiological Reviews*, 59(1) :1–30, 1995.
- [8] K. Heidtke and S. Schulze-Kremer. Design and implementation of a qualitative simulation model of λ phage infection. *Bioinformatics*, 14(1) :81–91, 1998.
- [9] M. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, San Diego, CA, 1974.
- [10] J. Hoch. The phosphorelay signal transduction pathway in the initiation of *Bacillus subtilis* sporulation. *Journal of Cellular Biochemistry*, 51 :55–61, 1993.
- [11] J. Hoch. Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. *Annual Review of Microbiology*, 47 :441–465, 1993.
- [12] B. Lewin. *Genes VI*. Oxford University Press, Cambridge, 1997.
- [13] T. Mestl, E. Plahte, and S. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *Journal of Theoretical Biology*, 176 :291–300, 1995.
- [14] M. Perego. Kinase-phosphatase competition regulates *Bacillus subtilis* development. *Trends in Microbiology*, 6(9) :366–370, 1998.
- [15] E. Shacter, P. Chock, and E. Stadtman. Regulation through phosphorylation/dephosphorylation cascade systems. *Journal of Biological Chemistry*, 259(19) :12252–12259, 1984.
- [16] A. Sonenshein. Control of sporulation initiation in *Bacillus subtilis*. *Current Opinion in Microbiology*, 3 :561–566, 2000.
- [17] P. Stragier and R. Losick. Molecular genetics of *Bacillus subtilis*. *Annual Review of Genetics*, 30 :297–341, 1996.
- [18] M. Strauch, J.-J. Wu, R. Jonas, and J. Hoch. A positive feedback loop controls transcription of the *spo0F* gene, a component of the sporulation phosphorelay in *Bacillus subtilis*. *Molecular Microbiology*, 7(6) :967–974, 1993.
- [19] R. Trelease, R. Henderson, and J. Park. A qualitative process system for modeling NF- κ B and AP-1 gene regulation in immune cell biology research. *Artificial Intelligence in Medicine*, 17 :303–321, 1999.
- [20] G. Yagil and E. Yagil. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophysical Journal*, 11 :11–27, 1971.

Appendix

A Calcul de l'état stable du phosphorelais (programme *Matlab*).

```

% /home/bacillus/batt/stage/matlab/simple_model_calcul.m
% calculates stable steady state of phosphorylation cascade
% system: S+kin; Sp+phos

'entering calculus'

% output format
format long e;

%plot resolution: number of points per dimension
Res = 10;

%substrate limit concentrations
Stmax = 10 * 10^(-8);
Stmin = 0;
Stdiff = Stmax - Stmin;

%enzymes limit concentrations
kin_max = 10 * 10^(-11);
kin_min = 0;
kin_diff = kin_max - kin_min;
rap_max = 10 * 10^(-7);
rap_min = 0;
rap_diff = rap_max - rap_min;

% kinetic constants
Kmf =0.13 * 10^(-6);
Kmr =530 * 10^(-6);
kf =29.5;
kr =3.8;
K0= Stmax;

%concentrations matrix initialisations
kin = [kin_min+kin_diff/Res : kin_diff/Res : kin_max];
rap = [rap_min+rap_diff/Res : rap_diff/Res : rap_max] ;
s = [Stmin+Stdiff/Res : Stdiff/Res : Stmax ];
K0tab = [K0/Res : K0/Res : K0];
Activation = zeros(Res,Res,Res,Res);

%stable steady-state calculations
for i = 1:Res
    'external loop done'
    for j = 1:Res
        for k = 1:Res
            % steady state value of S
            % calculation of the polynom roots (2 values)
            Sppoly = [rap(j)*kr-kf*kin(i)
                    kin(i)*kf*(-Kmr+s(k))-rap(j)*kr*(Kmf+s(k))
                    kf*kin(i)*s(k)*Kmr ];
            Sproots = roots(Sppoly);

            % selection of a valid solution: concentrations are in the range
            % and the system is stable;
            valid_Sproots = zeros(2,1); %stores valid solutions
            for q=1:2 %both values
                stable= 0; % indicates if the point is stable or not (cf HdJ814, p187)
            end
        end
    end
end

```

```

lambda= 2*Sproots(q)*rap(j)*kr-2*Sproots(q)*kf*kin(i)
        -kf*kin(i)*Kmr+kf*kin(i)*s(k)-rap(j)*kr*Kmf-rap(j)*kr*s(k);
if (lambda<=0)
    stable= 1;
end %else stable= 0
if ((Sproots(q)>= Stmin) & (Sproots(q)<= Stmax) & (stable==1))
    valid_Sproots(q)= Sproots(q);
end
end
end

count=0; %count the number of good solutions(hopefully only one)
for q=1:2
    if (valid_Sproots(q) ~= 0) % !\: dangerous condition
        Sp_concentration= valid_Sproots(q); %stores "the" good solution
        count=count+1;
    end
end
if count ~= 1
    'gros probleme'
    pause(500);
    exit
else
    % only one good solution
    for u=1:Res % various thresholds
        Activation(i,j,k,u)= Sp_concentration^6/(K0tab(u)^6 + Sp_concentration^6);
    end
end
% "for loops" end
end
end
end
end

save simple.mat;

```

B Fichiers d'entrée de VGNA.

Exemple de la modélisation qualitative complète du gène *kinA*.

```

state-variable: kinA
zero-parameter: z_kinA
box-parameter: max_kinA
production-parameters: k_kinA_1, k_kinA_2
degradation-parameters: g_kinA
threshold-parameters: t_kinA_1, t_kinA_2, t_kinA_3
state-equation:
    d/dt kinA =  k_kinA_2
                * s+(sigH, t_sigH_1)
                * ( 1 - s+(spo0A, t_spo0A_3)
                    * s+(kinA, t_kinA_3)
                    * s-(spo0E, t_spo0E_1)
                    * s+(signal, t_signal_1) )
                + k_kinA_1 * s+(sigA, t_sigA_1)
                - g_kinA * kinA
threshold-inequalities:
    t_kinA_1 < t_kinA_2;
    t_kinA_2 < t_kinA_3;
nullcline-inequalities:
    k_kinA_2 > g_kinA * t_kinA_3;
    k_kinA_2 < g_kinA * max_kinA;
    k_kinA_2 + k_kinA_1 > g_kinA * t_kinA_3;
    k_kinA_2 + k_kinA_1 < g_kinA * max_kinA;

```

```

k_kinA_1 > g_kinA * t_kinA_1;
k_kinA_1 < g_kinA * t_kinA_2;

```

Différentes modélisations de la régulation de l'expression du gène *KinA*, mettant en évidence les modifications apportées.

– modèle simple du phosphorelais :

```

d/dt kinA = k_kinA_2
            * s+(sigH, t_sigH_1)
            * ( 1 - s+(spo0A, t_spo0A_3)
                * s+(kinA, t_kinA_3)
                * s-(spo0E, t_spo0E_1)
                * s+(signal, t_signal_1) )
            + k_kinA_1 * s+(sigA, t_sigA_1)
            - g_kinA * kinA

```

– modèle plus complexe du phosphorelais :

```

d/dt kinA = k_kinA_2
            * s+(sigH, t_sigH_1)
            * ( 1- s+(spo0F, t_spo0F_3)
                * s+(spo0B, t_spo0B_3)
                * s+(spo0A, t_spo0A_3)
                * s+(kinA, t_kinA_3)
                * s-(rapA, t_rapA_1)
                * s-(spo0E, t_spo0E_1)
                * s+(signal, t_signal_1) )
            + k_kinA_1 * s+(sigA, t_sigA_1) /* autres kinases */
            - g_kinA * kinA

```




Unité de recherche INRIA Rhône-Alpes

655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique

615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur

INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)

<http://www.inria.fr>

ISSN 0249-6399