Systems Biology approach

Production of a huge amount of data

Challenge : understanding how complex interaction networks control the cell behavior

Topology of the network : knowledge on the involved components

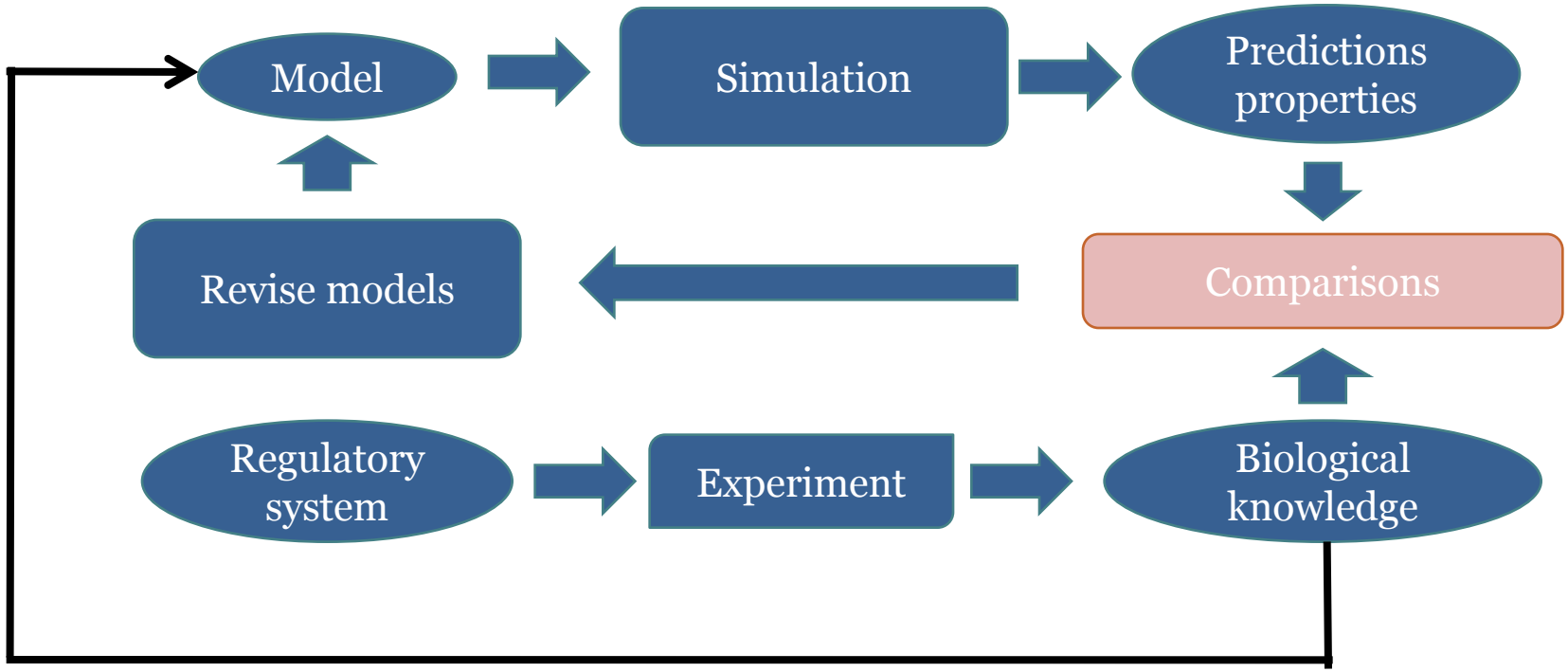
Dynamical modeling : understanding the dynamical properties of the system

Systems Biology approach

•Description of acquired understanding

•Explanatory of biological mechanisms

•Predictive of the behaviour of the system when it is pertubed (mutations, chemical intervention, change in the environnement…)



Formalising understanding

Systems Biology approach

What is a model?

- a simple abstract representation of reality
- can not explained all the details of a biological system
- but can help to understand the structural relationships and time-dependent dynamic behavior
- can reveal missing components or reactions
- allows to test hypotheses

Biological system analysis:

1. Make assumptions and abstraction to keep the model as simple as possible
2. Set boundaries to limit your model (it should stay reasonable)
3. Identify the involved components and their different states if required
4. Identify all actions and any changes occurring in the system
5. Define the relationships between components and actions
6. Define the stoichiometry
7. Define the initial state

It is recommended to start with a simple model and if necessary to refine the model afterwards by adding information.

Systems Biology approach

Two main classes of dynamical models:

- quantitative models:
  - ✓detailed representation of the model
  - ✓ produce quantitative results
  - ✓ require accurate kinetic data (often missing)
  - ✓ mainly based on ordinary differential equation (ODE)

- qualitative models:
  - ✓ require only abstract representation of threshold or concentration (example when the protein X is present at a concentration higher than θ_1 it activates the transcription of gene *a* and when it is present at a concentration higher of θ_2 it represses the transcription of gene *b*. With $\theta_1 < \theta_2$)
  - ✓ defined through discrete formalisms or piecewise linear differential equation

Those modeling techniques are complemented by simulation techniques to make predictions on the behavior of the biological process.

Systems Biology approach

Different modeling approaches among them:

model	Qualitative	Quantitative	Deterministic	Stochastic
Ordinary Differential Equations (ODE)		X	X	
Graph theory	X		X	
Bayesian networks		X		X
Piecewise Linear Differential Equations	X		X	
Boolean/Logical models	X		X	
Petri Nets	X	X	X	X

Piece-wise linear differential equations

When quantitative data are not known, the dynamics of genetic regulatory network can be modeled by a class of piece-wise linear (PL) differential equations originally proposed by Glass and Kauffman (1973) and generalized by Mestl *et al.* (1995).
The model have mathematical properties that favour qualitative analyses of:

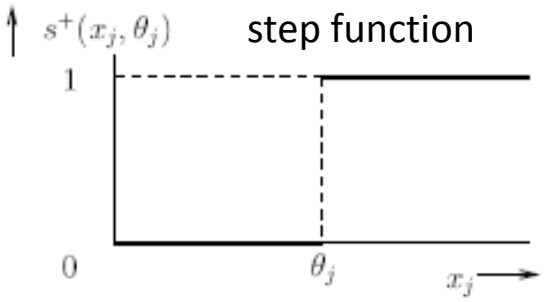
- the steady state
- transient behavior of regulatory systems.

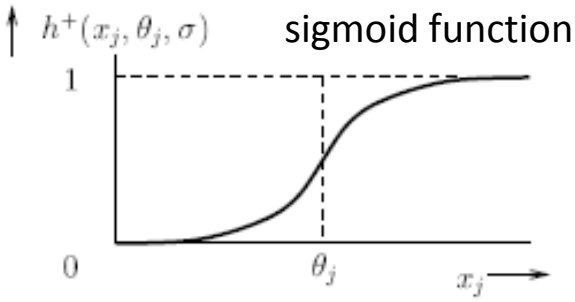
In PL model:

- state variables correspond to concentrations of proteins encoded by genes in the network
- differential equations represent the interactions arising from the regulatory influence of some proteins on the synthesis or degradation of others
- discontinuous step functions modeled the regulatory interactions (approximation of the switch-like behavior of genes whose expression is regulated by continuous sigmoid curves)

⇒ Resulting differential equations are piecewise linear

Piece-wise linear differential equations

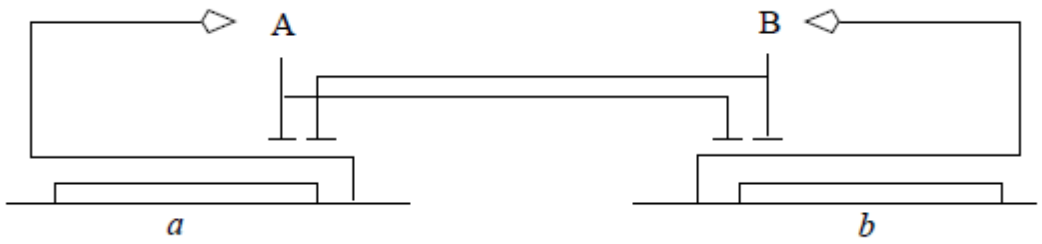




When $\sigma \rightarrow \infty$, the sigmoid functions (also called Hill functions) can be approach by a step function.

Piece-wise linear differential equations

Simple example : two genes *a* and *b*, transcribed from separate promoters, encode proteins A and B. Proteins A and B repress gene *a* and *b* at different concentrations. Repressions of the genes is achieved through the binding of the proteins to regulatory sites overlapping the promoters. One positive and two negative feedback loops.



concentration of A > θa_2 inhibition of gene *a* expression
concentration of A > θa_1 inhibition of gene *b* expression
concentration of B > θb_1 inhibition of gene *a* expression
concentration of B > θb_2 inhibition of gene *b* expression

General form of the state equation represents the difference of the rate of synthesis of protein x and the rate of degradation of protein x:

$$\frac{dx}{dt} = synthesis(x) - \deg radation(x)$$

Piece-wise linear differential equations

State equation: $\frac{dx_i}{dt} = f_i(x) - g_i(x)x_i$ (1) with $f_i(x)$ = rate of synthesis, $g_i(x)x_i$ = rate of degradation $x_i \geq 0,\ 1 \leq i \leq n$

$x = (x_1, \ldots, x_n)'$ is a vector of cellular concentrations.

The function f_i $R^n_{\geq 0} \rightarrow R_{\geq 0}$ is defined as:

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

Where $\kappa_{il} > 0$ is a rate parameter and b_{il} $R^n_{\geq 0} \rightarrow \{0,1\}$ is a regulation function defined in terms of step function. In the simplest case, the step function s+ : $R^2 \rightarrow \{0,1\}$ is defined as follow:

$$s^+(x_j, \theta_j) = \begin{cases} 1, & x_j > \theta_j \\ 0, & x_j < \theta_j \end{cases}$$

In the simple case:

$f_i(x) = \kappa_i s^+(x_j, \theta_j)$ gene *i* is not expressed if the concentration of protein J is below the threshold θ_j and above this threshold it is expressed at the rate κ_i

If protein J is a negative regulator of gene *i* : $f_i(x) = \kappa_i s^-(x_j, \theta_j)$ with $s^-(x_j, \theta_j) = 1 - s^+(x_j, \theta_j)$

The degradation function g_i is expressed analogously with degradation rates note γ instead of κ

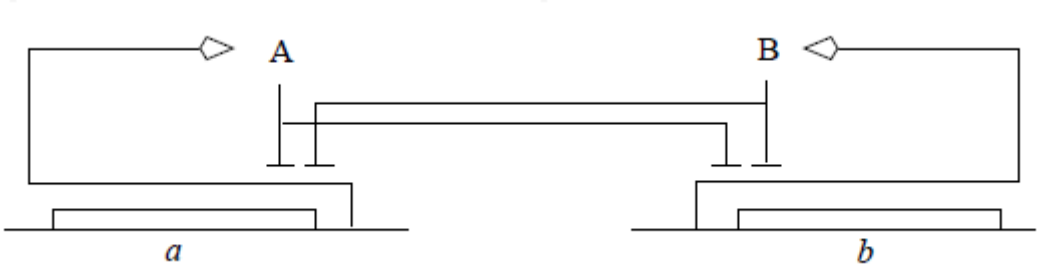
Piece-wise linear differential equations

The PL models can be extended to take into account input variable $u = (u_1, \ldots, u_m)'$ representing the concentration of proteins and small molecules whose synthesis and degradation are regulated outside the system.
The state equation becomes:

$$\frac{dx}{dt} = f(x,u) - g(x,u)x$$

If the input variables are assumed to be constant $\frac{du}{dt} = 0$ and the state equation are reduced to (1)

Piece-wise linear differential equations



concentration of A > θa_2 inhibition of gene *a* expression
concentration of A > θa_1 inhibition of gene *b* expression
concentration of B > θb_1 inhibition of gene *a* expression
concentration of B > θb_2 inhibition of gene *b* expression

State equation for gene a:

$$\frac{dx_a}{dt} = \kappa_a s^-(x_a, \theta a_2) s^-(x_b, \theta b_1) - \gamma_a x_a$$

State equation for gene b:

$$\frac{dx_b}{dt} = \kappa_b s^-(x_a, \theta a_1) s^-(x_b, \theta b_2) - \gamma_b x_b$$

Piece-wise linear differential equations

Dynamical properties of PL of the form (1) could be analyzed in the *n*-dimensional phase space box Ω.

$\Omega = \Omega_1 \times \ldots \times \Omega_n$ where every Ω_i $1 \leq i \leq n$ is defined as:

$\Omega_i = \{x_i \in \mathbb{R}_{\geq 0} \mid 0 \leq x_i \leq max_i\}$ with max_i a parameter corresponding to a maximum concentration for the protein

In general, a protein encoded by a gene will be involved in different interactions at different threshold concentrations. Thus the phase space Ω will be divided into hyper-rectangular regions that are called regulatory domains.

A regulatory domain $D \subseteq \Omega$ id defined by $D = D_1 \times \ldots \times D_n$ where every D_i $1 \leq i \leq n$ is defined by one of the following equations:

$$D_i = \{x_i \mid 0 \leq x_i < \theta_i^1\},$$
$$D_i = \{x_i \mid \theta_i^1 < x_i < \theta_i^2\},$$
$$\ldots$$
$$D_i = \{x_i \mid \theta_i^{p_i} < x_i \leq \max_i\}$$

With protein *i* having p_i thresholds

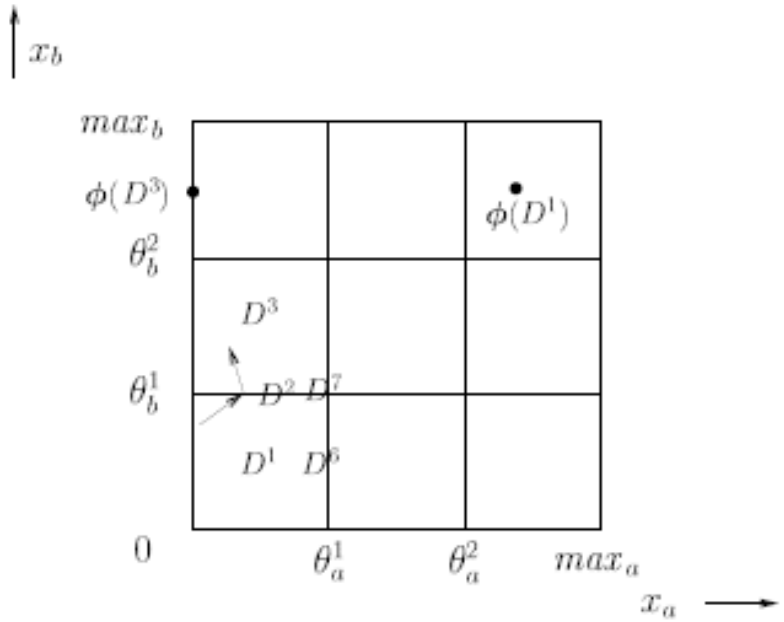
If for a domain *D*, there are some *i,j*, $1 \leq i \leq n$ and $1 \leq j \leq p_i$ such as

$$D_i = \{x_i \mid x_i = \theta_i^j\}$$

Then *D* is called a switching domain

Piece-wise linear differential equations

In our simple example, proteins A and B have two concentration thresholds. Thus, Ω is divided into 9 regulatory domains and 16 switching domains.



D1 is defined by: $0 \le x_a < \theta_a^1$

$0 \le x_b < \theta_b^1$

D2 is defined by: $0 \le x_a < \theta_a^1$

$x_b = \theta_b^1$

D3 is defined by: $0 \le x_a < \theta_a^1$

$\theta_b^1 < x_b < \theta_b^2$

D6 is defined by: $x_a = \theta_a^1$

$0 \le x_b < \theta_b^1$

If max_i is chosen as $\max_i = \max_{x \ge 0} \dfrac{f_i(x)}{g_i(x)}$, it can be shown that all trajectories starting in Ω will stay in it.

Trajectories starting outside will enter Ω at some point.

Piece-wise linear differential equations

When the expression of the step functions are evaluated in a regulatory domain, f_i and g_i reduce to the sum of rate constants.
More precisely, in every regulatory domain D f_i reduces to some $\mu_i^D \in M_i \equiv \{ f_i(x) \mid 0 \le x \le \max \}$ and g_i to some $\nu_i^D \in N_i \equiv \{ g_i(x) \mid 0 \le x \le \max \}$

M_i and N_i collect the synthesis and degradation rates of the protein in different domains of Ω.
It can be shown that all trajectories in *D* monotonically tend towards a stable steady state $x = \mu^D/\nu^D$, the target equilibrium lying at the intersection of the *n* hyperplans $x_i = \mu_i^D/\nu_i^D$. The target equilibrium of the protein concentration xi gives an indication of the strength of gene expression in D.

In our example, we have the following state equations and the sets M_a, N_a, M_b and N_b :

$$\frac{dx_a}{dt} = \kappa_a s^-(x_a, \theta a_2) s^-(x_b, \theta b_1) - \gamma_a x_a \qquad M_a = \{0, \kappa_a\} \text{ and } N_a = \{\gamma_a\}$$

$$\frac{dx_b}{dt} = \kappa_b s^-(x_a, \theta a_1) s^-(x_b, \theta b_2) - \gamma_b x_b \qquad M_b = \{0, \kappa_b\} \text{ and } N_b = \{\gamma_b\}$$

In domain D^1: $0 \le x_a < \theta_a^1$, $0 \le x_b < \theta_b^1$ — Then, protein *A* does not inhibits either gene *a* or *b,* and protein *B* does not inhibits either gene *a* or *b*.

In D^1, the state equations become:

$$\frac{dx_a}{dt} = \kappa_a - \gamma_a$$

$$\frac{dx_b}{dt} = \kappa_b - \gamma_b$$

Target equilibrium = $(\kappa_a/\gamma_a, \kappa_b/\gamma_b)$ lies outside D^1

Piece-wise linear differential equations

In domain D^3: $0 \le x_a < \theta_a^1$

$\theta_b^1 < x_b < \theta_b^2$

Then, protein *A* does not inhibits either gene *a* or *b,* and protein *B* inhibits gene *a* but not gene *b*.

In D^3, the state equations become:

$$\frac{dx_a}{dt} = -\gamma_a$$

$$\frac{dx_b}{dt} = \kappa_b - \gamma_b$$

Target equilibrium = $(0, \kappa_b/\gamma_b)$ lies outside D^1

Thus, different regulatory domains have different target equilibriums

Piece-wise linear differential equations

Qualitative constraints on parameter values that can be inferred from biological data:

- threshold inequalities
- equilibrium inequalities

Threshold inequalities: they are obtained by ordering the p_i concentration thresholds of the protein encoded by the gene *i*.

$$0 < \theta_i^1 < \theta_i^2 < \ldots < \theta_i^{p_i} < \max_i$$

The threshold inequalities determine the partitioning of Ω into regulatory and switching domains

Equilibrium inequalities: they order the possible target equilibrium levels of xi in different regulatory domains $D \subseteq \Omega$ with respect to the threshold values. They define the strength of gene expression in the domain in a qualitative way, on the scale of ordered concentration thresholds.

For $\mu_i \in M_i$, $\nu_i \in N_i$ and $\mu_i, \nu_i \neq 0$, we can specify:

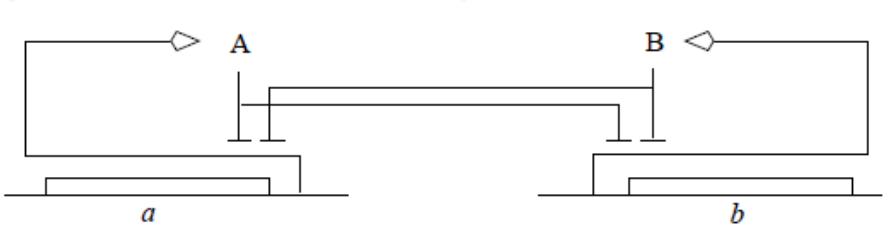
$$0 < \mu_i / \nu_i < \theta_i^1,$$

$$\theta_i^1 < \mu_i / \nu_i < \theta_i^2,$$

$$\ldots$$

$$\theta_i^{p_i} < \mu_i / \nu_i < \max_i .$$

Piece-wise linear differential equations



concentration of A > θa_2 inhibition of gene *a* expression
concentration of A > θa_1 inhibition of gene *b* expression
concentration of B > θb_1 inhibition of gene *a* expression
concentration of B > θb_2 inhibition of gene *b* expression

Equilibrium inequalities for this example :

In the absence of protein B ($s^-(x_b, \theta_b^1) = 1$) , while protein A has not reached the its highest level ($s^-(x_a, \theta_a^2) = 1$) gene *a* is expressed at rate κ_a. The corresponding target equilibrium κ_a / ν_a must be above the threshold θ_a^2, Otherwise, the protein A would not be able to reach or maintain a concentration at which the observed negative autoregulation of gene *a* occurs. The same pertains for protein B.

For gene *a*:

State equation:

$$\frac{dx_a}{dt} = \kappa_a s^-(x_a, \theta a_2) s^-(x_b, \theta b_1) - \gamma_a x_a$$

Threshold inequalities:

$$0 < \theta_a^1 < \theta_a^2 < \max_a$$

Equilibrium inequalities:

$$\theta_a^2 < \kappa_a / \gamma_a < \max_a$$

For gene *b*:

State equation:

$$\frac{dx_b}{dt} = \kappa_b s^-(x_a, \theta a_1) s^-(x_b, \theta b_2) - \gamma_b x_b$$

Threshold inequalities:

$$0 < \theta_b^1 < \theta_b^2 < \max_b$$

Equilibrium inequalities:

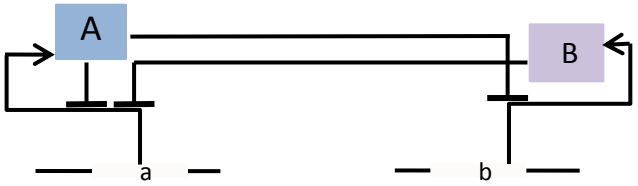
$$\theta_b^2 < \kappa_b / \gamma_b < \max_b$$

Piece-wise linear differential equations

Qualitative simulation:

Given a qualitative PL model and initial conditions, the aim of qualitative simulation is to determine the possible qualitative behaviors of the system.
The simulation algorithm included in the Genetic Network Analyzer (GNA) developed by Hidde de Jong and collaborators results in a transition graph, a directed graph of qualitative states and transitions between qualitative states. This graph contains qualitative equilibrium states of qualitative cycles.
A sequence of qualitative states in the transition graph represents a predicted qualitative behavior of the system.

Genetic Network Analyzer (de Jong *et al.*, 2003)



concentration of A > t_xa2 inhibition of gene a expression
Concentration of B > t_xb inhibition of gene a expression
Concentration of A > t_xa1 inhibition of gene b expression

GNA

File Edit Analysis Results Window

Model (xaxb)
- Variables
  - xa
  - xb
- Initial conditions
  - xaxb1
  - xaxb27
- Atomic propositions
- Properties

Variable xa

New Edit

xa
- Threshold parameters
  - z_xa
  - t_xa1
  - t_xa2
  - max_xa
- Synthesis parameters
  - k_xa
- Degradation parameters
  - g_xa

Variable name: xa Rename

Variable type: State variable / Input variable

Help

State equation:

d/dt xa = k_xa * s-(xa, t_xa2) * s-(xb, t_xb) - g_xa * xa

$$\frac{dxa}{dt} = kxa * s^{-}(xa, t\_xa2) * s^{-}(xb, t\_xb) - g\_xa * xa$$

Variable xb

New Edit

xb
- Threshold parameters
  - z_xb
  - t_xb
  - max_xb
- Synthesis parameters
  - k_xb
- Degradation parameters
  - g_xb

Variable name: xb Rename

Variable type: State variable / Input variable

Help

State equation:

d/dt xb = k_xb * s-(xa, t_xa1) - g_xb * xb

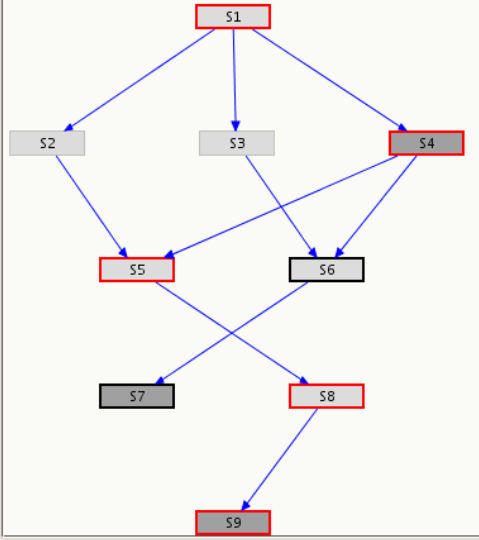
$$\frac{dxb}{dt} = kxb * s^{-}(xa, t\_xa1) - g\_xb * xb$$

Check equation

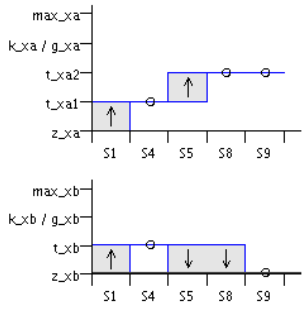
Parameter ordering:

Simulation graph window – Graph3

Graph3



State variables in path of selected states Graph3

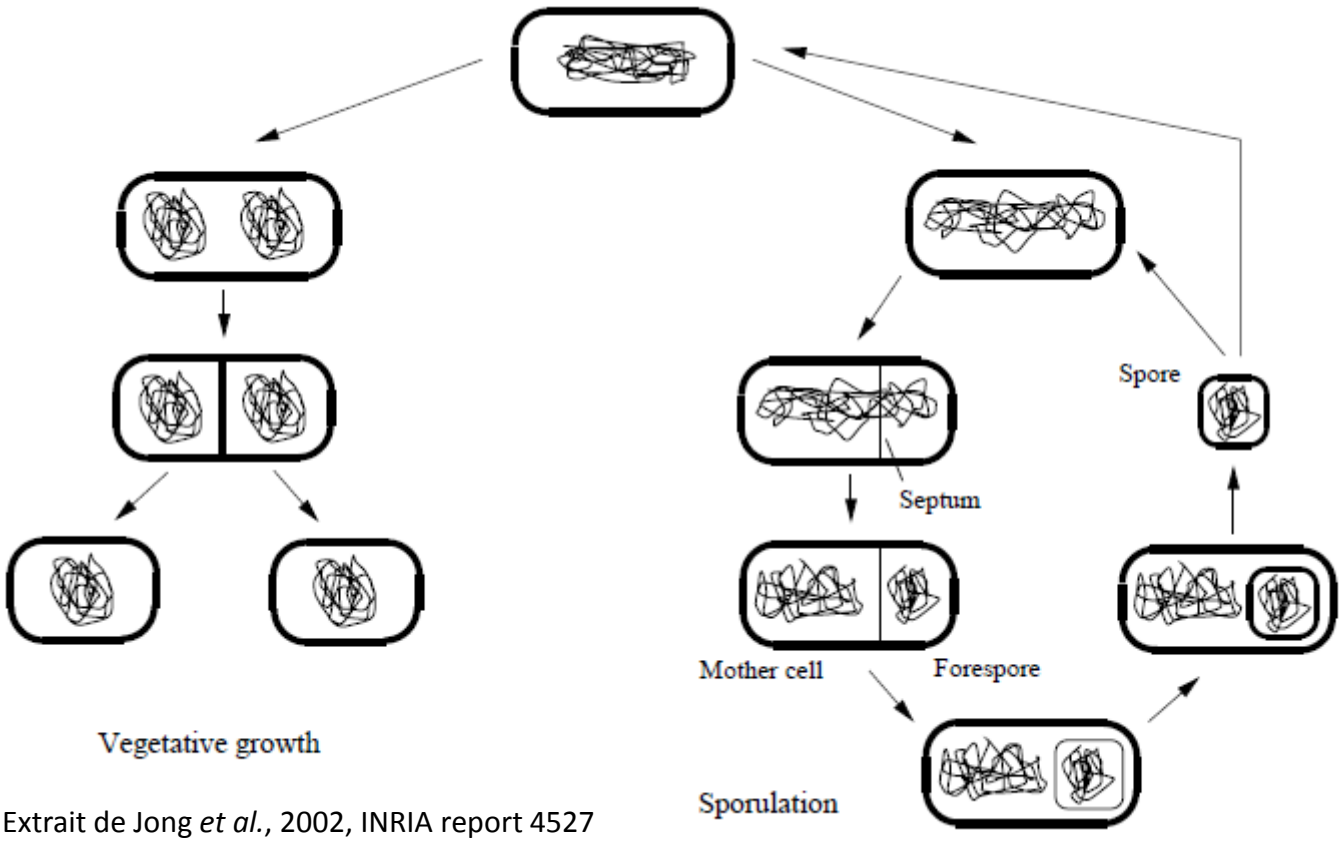


Exemple : modélisation de la régulation de l'initiation de la sporulation chez *B. subtilis*

Adaptation de l'organisme à son environnement de manière à survivre aussi bien sur une courte période de temps qu'à l'échelle de l'évolution.
Les clefs de l'adaptation consistent à répondre au changements environnementaux, comme la disponibilité de nutriments, la densité de cellules, la température etc. en ajustant la synthèse et la dégradation de protéines régulatrices contrôlant la croissance, le métabolisme et le développement.

Quand les conditions environnementales se détériorent, la bactérie *Bacillus subtilis* stoppe sa croissance exponentielle et entre dans la phase stationnaire. Durant la transition en phase stationnaire la cellule initie tout un tas de réponses en vue de survivre dans un environnement de plus en plus hostile. La réponse ultime de la cellule est la sporulation, c'est-à-dire la formation d'une spore remarquablement résistante. La spore peut rester en dormance plusieurs années. Quand les conditions deviennent favorable, la spore germe et la bactérie reprend sa croissance végétative.
Le changement de programme, passage de la croissance végétative à la sporulation implique un changement radical du programme génétique de la cellule. Le switch est contrôlé par un réseau complexe de régulation génétique impliquant plus de 125 gènes.

Exemple : modélisation de la régulation de l'initiation de la sporulation chez *B. subtilis*



Extrait de Jong *et al.*, 2002, INRIA report 4527

Cycle cellulaire en phase végétative : division symétrique, génération de deux cellules identiques
Sporulation : division asymétrique résultant en deux types cellulaires différents, la « forespore » qui formera la spore et la « mother cell » qui aide a dépôt d'une enveloppe résistante autour de la spore et ensuite se désintègre.

Exemple : modélisation de la régulation de l'initiation de la sporulation chez *B. subtilis*

Facteur crucial pour la décision de passer en sporulation : l'état de phosphorylation du facteur de transcription Spo0A en réponse à différents signaux de l'environnement, du cycle cellulaire et de signaux provenant du métabolisme.
Au dessus d'un certain seuil de concentration, Spo0A~P active une cascade de facteur σ dirigeant la transcription de gènes initiant les changements morphologiques ayant lieu pendant la sporulation.

Spo0A~P active la transcription du gène σ^H dont le produit est impliqué dans la formation du septum, dans une boucle de rétroaction négative contrôlant l'accumulation de Spo0A~P et dans l'activation de la transcription des gènes de deux autres facteurs sigma σ^E et σ^F, dont les produits sont respectivement impliqués dans le développement de la « mother cell » et de la « forespore ». Ils activent, entre autre, l'expression des gènes de deux autres facteurs sigma σ^G et σ^K dont les produits contrôlent les étapes tardives de la sporulation.

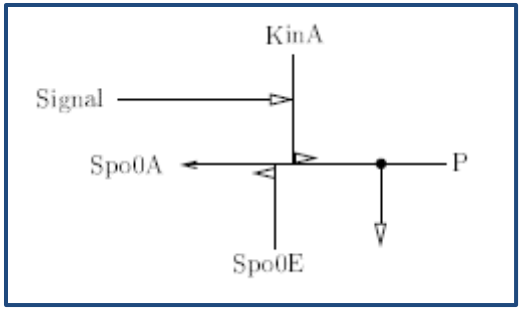
Modélisation de l'étape initiale seulement!

Exemple : modélisation de la régulation de l'initiation de la sporulation chez *B. subtilis*

Initiation :

Centrée sur un phospho-relai. Il transfère une phosphate à partir d'une kinase, KinA, KinB ou KinC à un régulateur de réponse Spo0F qui à son tour transfère le phosphate à Spo0A par le biais de la phospho-transférase Spo0B. Les phosphatases Spo0E, RapA et RapB, agissant sur Spo0A~P ou Spo0F~P peuvent inverser le flux de phosphate au travers du phospho-relai, inactivant ainsi *spo0A*. L'activation des kinases et phosphatases est réalisée par des signaux informant entre autre sur la disponibilité de nutriment, la densité cellulaire, la progression du cycle cellulaire et l'activité des réseaux métaboliques.

Phospho-relai : système intégrant des paramètres environnementaux et physiologique pour prendre sa décision. Décision importante, la cellule doit balancer le risque d'entrer sans nécessité en sporulation créant ainsi un désavantage sérieux de croissance contre le risque d'une entrée trop tardive en sporulation. Ceci explique la complexité du phospho-relai.

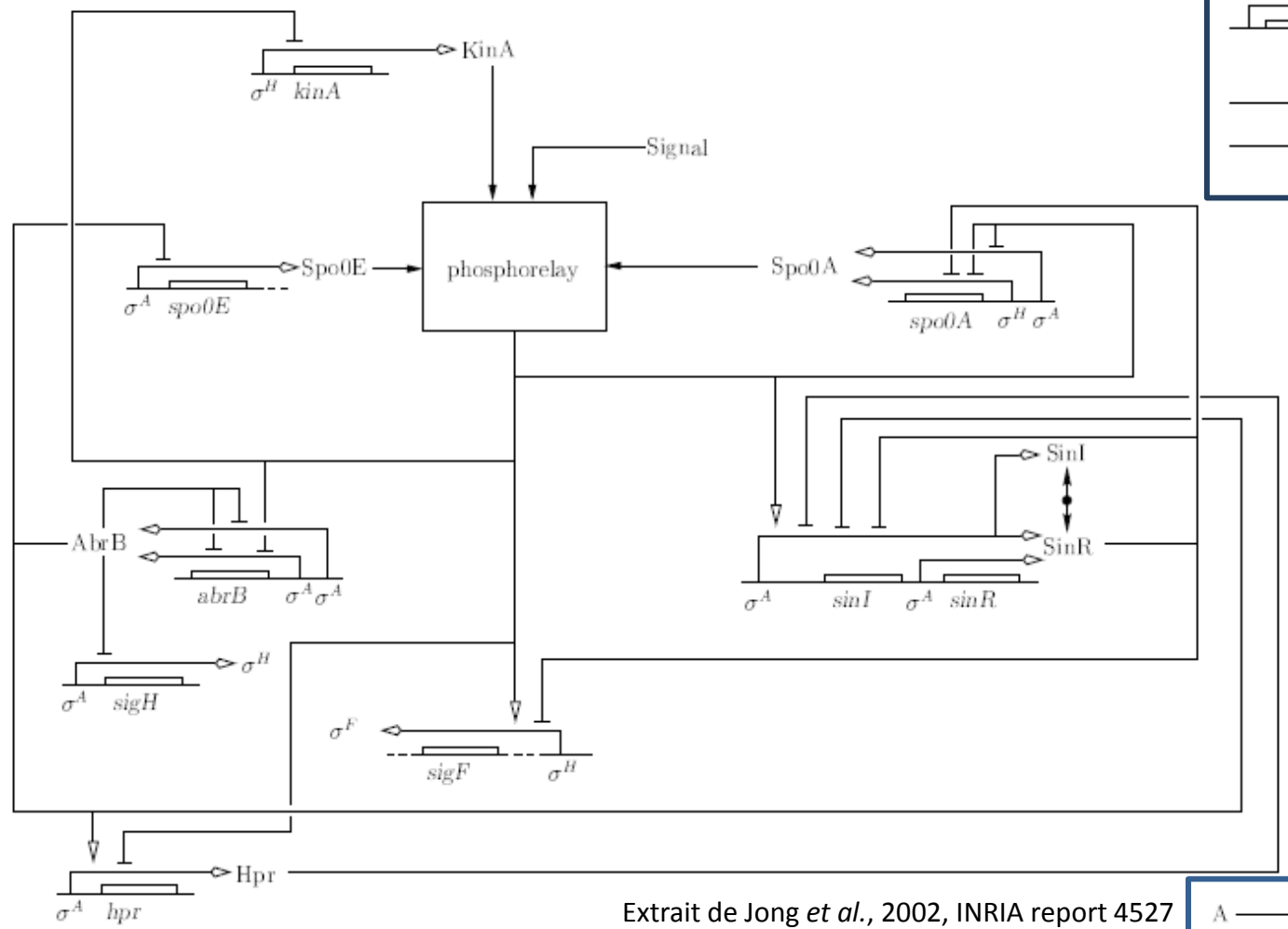


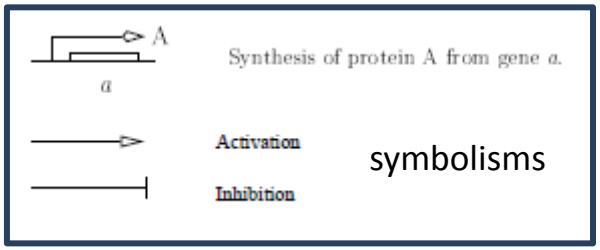
Modélisation simplifiée :

-n'affecte pas la fonction du phospho-relai (modulation du flux de phosphate au travers de la compétition kinase- phosphatase).

- une seule kinase KinA
- une seule phosphatase Spo0E
- signal environnemental agissant sur KinA
- éléments du phospho-relai sont régulés au niveau transcriptionel par Spo0A~P et par un nombre de protéines dont les gènes sont directement ou indirectement contrôlées par Spo0A~P

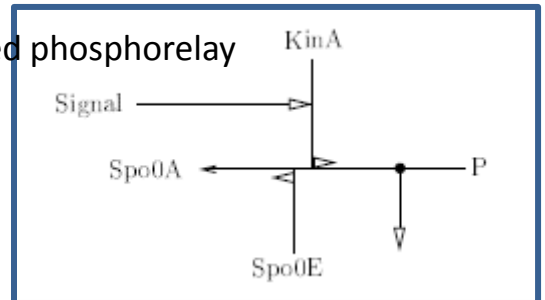
Simplified regulatory network of initiation of sporulation in *B. subtilis*

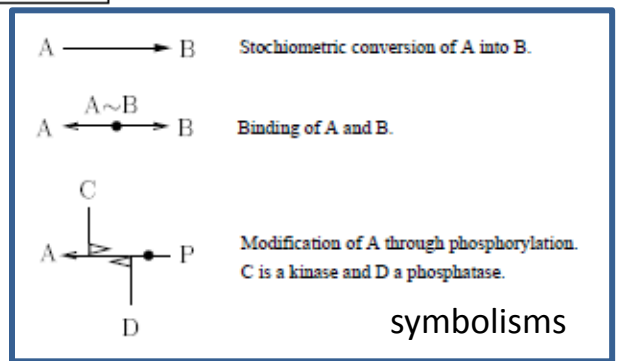




Extrait de Jong *et al.*, 2002, INRIA report 4527

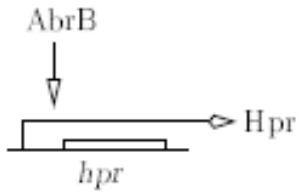
Simplified phosphorelay

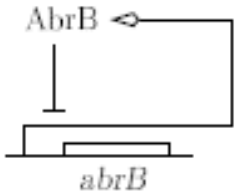




Modélisation simple pour commencer

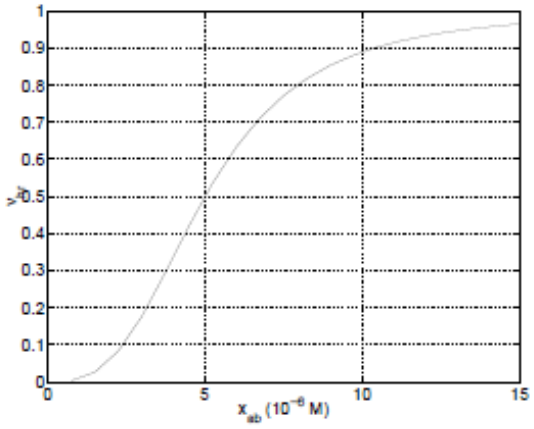
Régulateur de transcription:



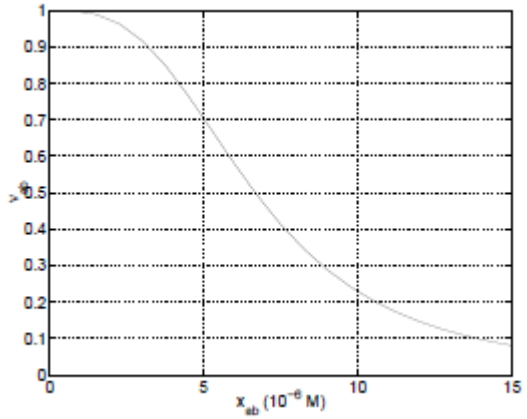


AbrB présente en concentration x_{ab} active le gène *hpr*
Modélisation par une courbe de Hill

AbrB présente en concentration x_{ab} réprime son propre gène
Modélisation par une courbe sigmoide mais fonction décroissante en fonction de la concentration x_{ab}



Extrait de Jong *et al.*, 2002, INRIA report 4527



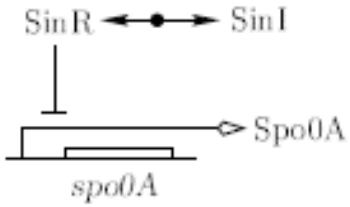
$x_{ab} > \theta_{ab}^{1}$: activation du gène *hpr*

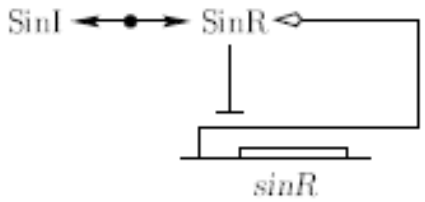
$x_{ab} > \theta_{ab}^{2}$: répression du gène *abrB*

avec $0 < \theta_{ab}^{1} < \theta_{ab}^{2} < \max_{ab}$

Modélisation simple pour commencer

Régulateur de transcription inactivé par un anti-répresseur:





SinR réprime l'expression de spo0A et de son propre gène. Liaison avec SinI inhibe la répression.
SinR existe sous deux formes :

- libre → active
- complexée → inactive

Association/dissociation SinI-SinR se réalisant à une échelle de temps bien inférieure à celle de la synthèse/dégradation des protéines, on peut supposer que le premier processus est quasiment à l'équilibre par rapport au second. Simplifie le modèle.
On peut donc calculer l'activité de spo0A en fonction des concentrations de SinI et SinR
Modélisation : surface sigmoïdale