

# Réseaux biologiques

A. Siegel

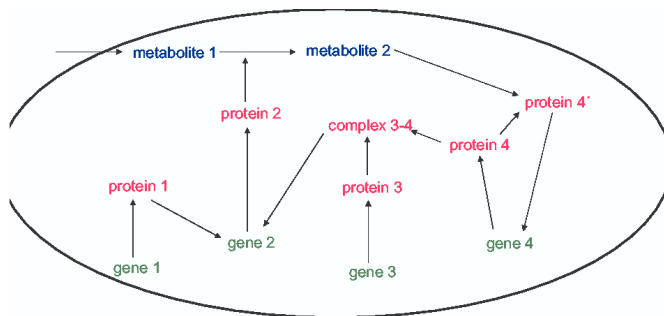
Liffré, novembre 2005

## Contexte

- On dispose d'un catalogue (plus ou moins exact) des **éléments fonctionnels** dans la cellule.
- Des nouvelles technologies fournissent des données en masse
  - **Mécanismes élémentaires**: interactions protéines-protéines, interactions protéines-ADN, réactions biochimiques...
  - **Etat d'une cellule**: expression d'ARN, expression de protéines, phénotype, concentrations de métabolites, flux métaboliques...
- **Le niveau gène/protéine atteint ses limites** pour expliquer le fonctionnement du vivant.

# Réseau biomoléculaire

Réseau d'interactions entre des entités biologiques dans une cellule



Source: D. Ropers (Inria Grenoble)

## Des sommets

- **Gène** (séquence d'ADN)
- **Protéine** (codées à partir d'un gène)
- Complexe protéique
- **Métabolites**
- Autres composés (ATP, ADP, ...)

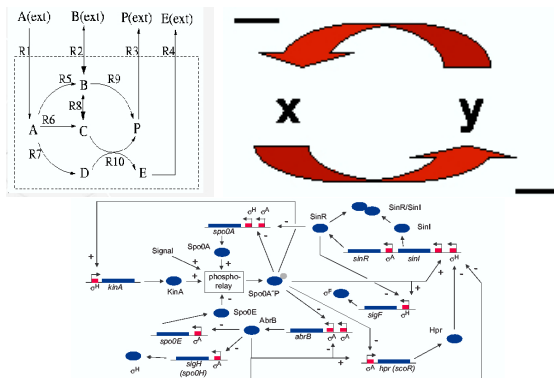
Plus ou moins bien référencés dans des bases de données.

## Des interactions

- Transformations de molécules en autre(s) molécule(s)
  - **Réaction** chimique
  - **Expression** d'un gène
  - Formation d'un complexe
- Contrôle d'une transformation
  - **Catalyse** d'une réaction par une enzyme
  - **Régulation** de l'expression d'un gène
- Divers
  - **Co-occurrence** dans la littérature

Des BDD (plus ou moins fournies) existent pour chaque type d'interaction (Kegg, Amaze, IntAct, Bibliosphere, Bind ...)

# Des graphes... Avec quelle signification ?



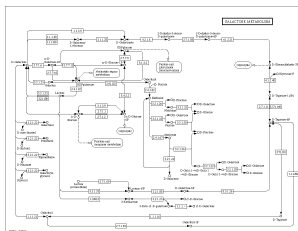
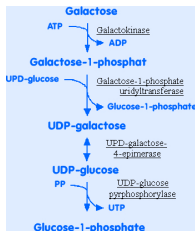
Pour décrire un mécanisme biologique, on utilise **différents types de réseaux** biomoléculaires.

Sources: J. Gagneur (EMBL), D. Thieffry (IBDM, Marseille), H. De Jong (Inria Grenoble)

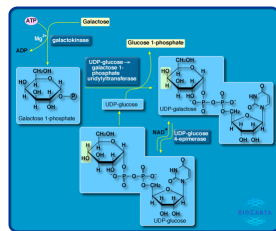
# Réseau métabolique

Série de réactions chimiques contrôlées éventuellement par des enzymes.

Fonction: exploiter et transformer les ressources disponibles en énergie



(glycolyse)

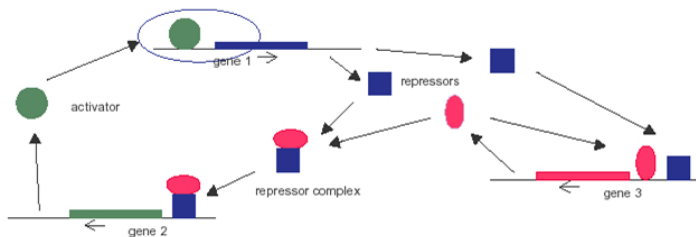


Sources: galactosaemia.com, KEGG, www.biocarta.com

# Réseau de régulation génétique

Interaction de régulation entre des gènes, des protéines et des petites molécules.

**Fonction: production régulée de protéines**



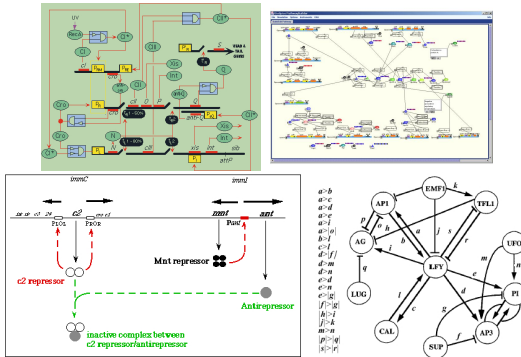
Source: Zhi Xie (Lincoln university)



# Réseau de régulation génétique

Interaction de régulation entre des gènes, des protéines et des petites molécules.

Fonction: production régulée de protéines

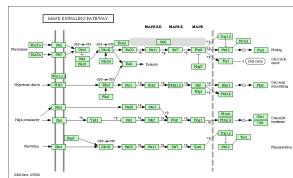
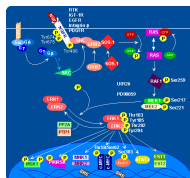
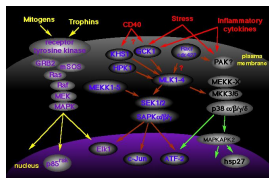


Source: M. E. Mulligan, A. Arkin, S. Malov, mbi DKFZ

# Voie de signalisation

Associations et dissociations de molécules (phosphorylation...).

**Fonction : Transport rapide de l'information**



(Voie de signalisation de MAPK)

Source: J. Woodgett, embiosciences, KEGG

# Différents types de réseaux biomoléculaires

- 1 Réseau de régulation génétique
  - Sommets: gènes
  - Arêtes: régulations
  - Données: séquence, expressions de gènes, puces
- 2 Réseau métabolique
  - Sommets: métabolites
  - Arêtes: consommation, production
  - Données: séquence, biochimie, spectrométrie
- 3 Réseau de signalisation
  - Sommets: protéines et état d'activité
  - Arêtes: interactions modifiant les états
  - Données: mesure de modification post-traductionnelles.
- 4 Réseau d'interaction protéine-protéine
- 5 Réseau de relations

# Différents types de réseaux biomoléculaires

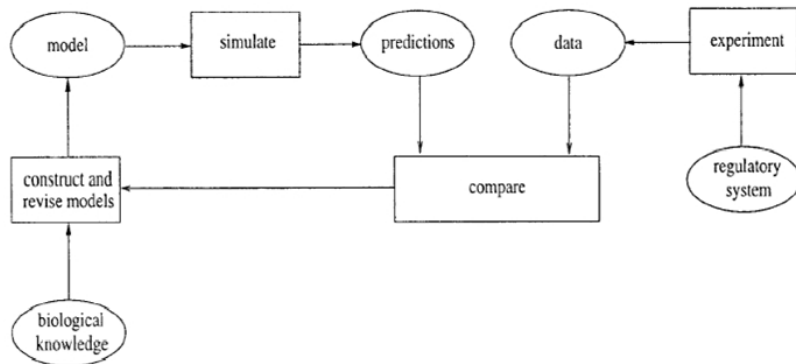
- 1 Réseau de régulation génétique
- 2 Réseau métabolique
- 3 Réseau de signalisation
- 4 Réseau d'interaction protéine-protéine
  - Sommets: protéines
  - Arêtes: interactions physiques
  - double-hybrides, spectrométrie
- 5 Réseau de relations
  - Sommets: gènes
  - Arêtes: relations fonctionnelles
  - Données: séquences, données d'expression, mesures de similarité.

# Questions biologiques motivant l'études des réseaux

Le développement et la fonction d'un organisme émergent des interactions dans et entre ces différents types de réseaux...

- Aider à structurer, **représenter** et interpréter des données expérimentales
  - Intégrer différents types de données.
  - Relier les mécanismes décrits aux états du système.
- Comprendre les processus cellulaires
  - **Prédire** l'état d'une cellule
  - Prédire l'effet d'une modification du système sur ses éléments.
- Aider les **expérimentations**, en produisant des hypothèses vérifiables sur un système.
- Comparer des réseaux pour différents organismes ou tissus, en rapport avec leur évolution.
- Mieux comprendre la **fonction** de gènes donnés.

## Construire et étudier un modèle: schéma type



Source: Zhi Xie (Lincoln university)

# Questions informatiques

- 1 (Re)construire/identifier des réseaux à partir de données expérimentales ou théoriques
  - Statistiques
  - Contraintes, diagnostic, ILP...
  - Apprentissage (kernel)
  - Comparaison entre organismes
- 2 Définir des types de **modélisation** en fonction
  - du type de système biologique
  - du type de prédictions recherchées
  - du type de données expérimentales disponibles pour la construction et l'exploitation du modèle.
- 3 Analyser le **comportement** d'un modèle.
- 4 Analyser la **structure** des réseaux
- 5 Applications

# Questions informatiques

- 1 (Re)construire/identifier des réseaux à partir de données expérimentales ou théoriques
- 2 Définir des types de **modélisation** en fonction
- 3 Analyser le comportement d'un modèle.
  - Simulation
  - Etude théorique de propriétés (stabilité...)
  - Prédiction de comportements. Atteignabilité de certains états.
  - Effets du bruit.
- 4 Analyser la **structure** des réseaux
- 5 Applications

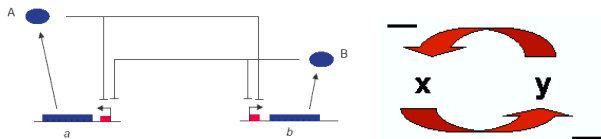


# Questions informatiques

- 1 (Re)construire/identifier des réseaux à partir de données expérimentales ou théoriques
- 2 Définir des types de modélisation en fonction
- 3 Analyser le comportement d'un modèle.
- 4 Analyser la structure des réseaux
  - propriétés structurelles (motifs).
  - Réduction de la taille des modèles; modularité.
  - Comparaison modèles discrets, continus, stochastiques.
  - Entre différents organismes (évolution des réseaux).
- 5 Applications
  - prédire la fonction d'une protéine.
  - Comparer des données aux prédictions d'un modèle
  - Trouver les moyens de contrôler un système ?

## Exemple de base: boucle négative

- Analyser le comportement d'un modèle.
  - Simulation
  - Etude théorique de propriétés (stabilité...)
  - Prédiction de comportements. Atteignabilité de certains états.
  - Effets du bruit.

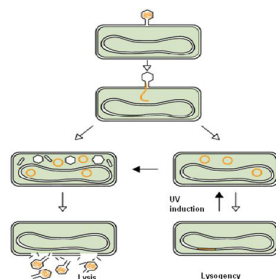


Comment modéliser la dynamique de cette boucle ?

Source: H. De Jong (Inria Grenoble), D. Thieffry (IBDM)

## Contexte biologique: phage lambda

- Le Phage Lambda est un virus qui infecte la bactérie E. Coli
- On dispose de bons modèles de réseaux de régulation
- Lambda P. injecte son ADN dans E. Coli
- Phases de lyse ou lysogénèse en fonction des conditions de concentration.

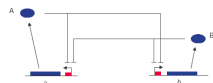


Source: Zhi Xie (Lincoln university)

## Contexte biologique: phage lambda

OR1 Cro OR2 OR3 CI

- Si **Cro domine**, la protéine produite va se lier à OR3.
  - CI ne peut plus être transcrite
  - Cro augmente
  - CI diminue par dégradation
  - **Phase lytique**
- Si **CI domine**, la protéine produite va se lier à OR1.
  - Cro ne peut plus être transcrite
  - CI augmente
  - Cro diminue par dégradation
  - **Croissance lysogénique**



(Modèles complexes pour expliquer la domination de Cro ou CI)

# Différents modèles

## Graphes d'interactions

### Modèles discrets/booléens

Les composants ont un petit nombre d'états qualitatifs.  
Fonctions de transitions.

### Linéaires par morceaux

formalisme différentiel  
les composants ont un nombre fini d'états qualitatifs.  
Fonctions seuils.

### Modèles différentiels

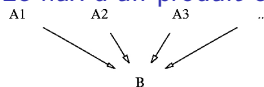
Grand nombre de molécules.  
Coefficients stochiométriques et paramètres.

### Modèles stochastiques

Petit nombre de molécules.  
Lois de probabilités.

## Modélisation différentielle

Le flux d'un produit est la somme des influences qui y arrivent.

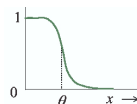


$$\frac{dB}{dt} = K_1 v(A_1) + K_2 v(A_2) + \dots$$

- Une équation pour chaque noeud ayant des arêtes entrantes
- les fonctions cinétiques  $v(A)$  dépendent du type d'interaction
  - paliers si réaction enzymatique;
  - seuil/sigmoïde si interaction génétique;
  - linéaire si réaction chimique usuelle.



$$\begin{cases} \frac{dx}{dt} = k_x F_{s_x}^-(y) - k_{-x} x \\ \frac{dy}{dt} = k_y F_{s_y}^-(x) - k_{-y} y \end{cases}$$



$$F_s^-(x) = \frac{x^n}{x^n + s^n}$$

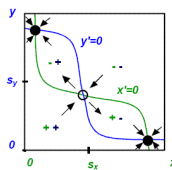
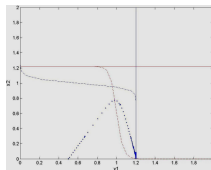
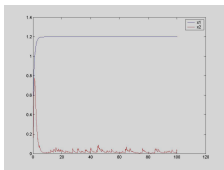
# Dynamique qualitative

- Etats d'équilibre: **Les flux sont constants** :  $\frac{dB}{dt} = 0$ . Système d'équations à résoudre.



$$\begin{cases} \frac{dx}{dt} = k_x F_{s_x}^-(y) - k_{-x}x = 0 \\ \frac{dy}{dt} = k_y F_{s_y}^-(x) - k_{-y}y = 0 \end{cases} \quad \begin{cases} x = K_x F_{s_x}^-(y) \\ y = K_y F_{s_y}^-(x) \end{cases}$$


- Convergence des trajectoires vers un état d'équilibre (sous conditions).
- Plusieurs situations possibles en fonction des valeurs des seuils et de la forme des fonctions.



Sources: O. Radulescu (IRMAR, Rennes), D. Thieffry (IBDM)

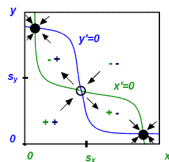
## Dynamique qualitative

- Etats d'équilibre: **Les flux sont constants** :  $\frac{dB}{dt} = 0$ . Système d'équations à résoudre.



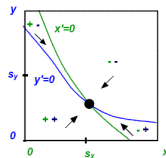
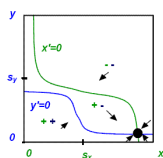
$$\begin{cases} \frac{dx}{dt} = k_x F_{s_x}^-(y) - k_{-x}x = 0 \\ \frac{dy}{dt} = k_y F_{s_y}^-(x) - k_{-y}y = 0 \end{cases} \quad \begin{cases} x = K_x F_{s_x}^-(y) \\ y = K_y F_{s_y}^-(x) \end{cases}$$

- Convergence des trajectoires vers un état d'équilibre (sous conditions).
- Plusieurs situations possibles en fonction des valeurs des seuils et de la forme des fonctions.



Deux attracteurs stables  
Un noeud instable

Sources: D. Thieffry



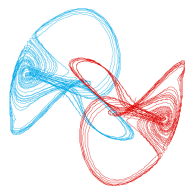
Unique attracteur stable



## Modélisation différentielle: cas général

- Les équations d'équilibres ne sont absolument **pas linéaires**: pas de méthode pour résoudre.
- Très bons outils de *simulation*
- **Diversité des comportements** observés et prédits: multistationnarité, oscillations, en fonction des paramètres et des effets de la non-linéarité.
- **Dynamique chaotique ?**

Question (C. Soulé): est-ce qu'un système différentiel  $\frac{dX}{dt} = F(X) - \lambda X$  avec  $F$  bornée peut avoir une dynamique cahotique ? On n'en sait rien.



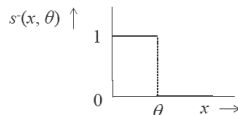
Sources: R. Thomas (ULB, Bruxelles)

## Modélisation linéaire par morceaux

- Les fonctions cinétiques sont approximables par des produits et sommes de fonctions à paliers.
- Les régulations génétiques sont bien approximées par des fonctions seuils.

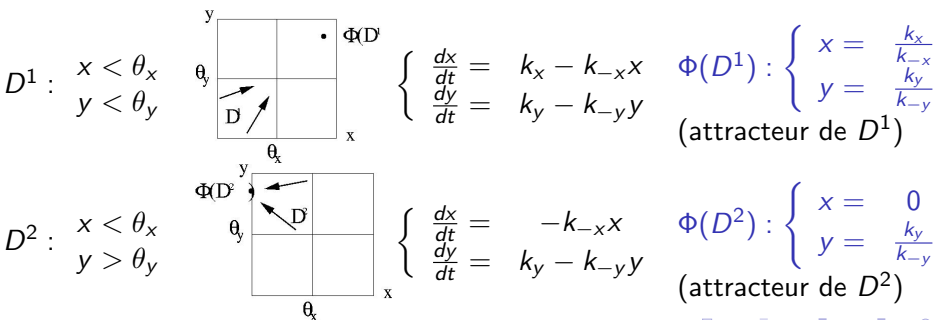


$$\begin{cases} \frac{dx}{dt} = k_x s^-(y, \theta_x) - k_{-x} \\ \frac{dy}{dt} = k_y s^-(x, \theta_y) - k_{-y} \end{cases}$$



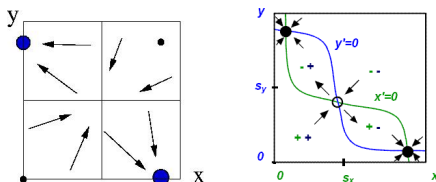
## Modélisation linéaire par morceaux

- Les équations sont **linéaires sur chaque domaine**
- Dans chaque domaine, il y a un **unique attracteur**
- **Simulation** du comportement qualitatif
- Etude des **attracteurs**
- Etude de tous les comportements possibles: **model checking**



## Modélisation linéaire par morceaux

- Les équations sont **linéaires sur chaque domaine**
- Dans chaque domaine, il y a un **unique attracteur**
- **Simulation** du comportement qualitatif
- Etude des **attracteurs**
- Etude de tous les comportements possibles: **model checking**



Convergence vers un des deux attracteurs stables

(on retrouve qualitativement le comportement général du système différentiel)

## Modélisation booléenne

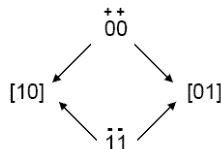
- Les variables sont supposées prendre une valeur 0 ou 1.
- Dynamique asynchrone: on ne bouge qu'une variable à la fois.



$$\begin{cases} X = \bar{y} \\ Y = \bar{x} \end{cases}$$

xy	XY
00	11
01	01
10	10
11	00

Table des attracteurs



Graphe de transition

Deux états d'équilibre

## Modélisation booléenne multivaluée

- **Abstractions logiques**: les concentrations  $(\hat{x}, \hat{y})$  prennent un nombre fini de valeurs.
- $\hat{x}$  prend autant de valeur que le nombre d'arêtes sortant de  $x$ .

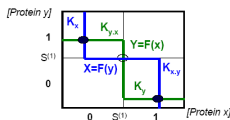


$$\hat{x} = 0 \quad \text{si} \quad x < k_x$$

$$\hat{x} = 1 \quad \text{si} \quad x > k_x$$

- L'**attracteur**  $(X, Y)$  d'un état est une fonction  $F$  qui décrit la position de l'**attracteur vis à vis des seuils**.

$$X = F_x(\hat{y}), \quad Y = F_y(\hat{x}).$$

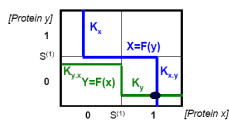


$$F_x(0) = 1 = K_x,$$

$$F_x(1) = 0 = K_{x,y}.$$

$$F_y(0) = 1 = K_y,$$

$$F_y(1) = 0 = K_{y,x}.$$



$$F_x(0) = 0,$$

$$F_x(1) = 1.$$

$$F_y(0) = 0,$$

$$F_y(1) = 0.$$

- **Etat d'équilibre logique**:  $X = \hat{x}$ .
- Prédiction des états stables et simulations dynamiques.

## Modélisation stochastique

- $X$  désigne le **nombre de molécules** (et pas la concentration)
- $p(X, t)$  est la proba que, à l'instant  $t$ , la cellule contienne  $X_1$  molécules du premier produit, etc...
- $\alpha_j \delta t$  est la proba que la réaction  $j$  se produise entre  $t$  et  $t + \delta t$  sachant que la cellule est dans l'état  $X$  en  $t$ .
- $\beta_j \delta t$  est la proba que la réaction  $j$  amène la cellule en l'état  $X$  entre  $t$  et  $t + \delta t$ .

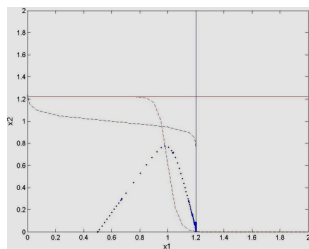
$$p(X, t + \delta t) = p(X, t)(1 - \sum \alpha_j \delta t) + \sum \beta_j \delta t$$

- Simulations
- **Richesse des comportements**
- Adaptés aux petits nombres de molécules et aux phénomènes de bruits.

## Modélisation stochastique: Exemple

$$\begin{cases} \frac{dx}{dt} = -x + \frac{\alpha_1}{1+\gamma y^n} \\ \frac{dy}{dt} = -y + \frac{\alpha_2}{1+\gamma x^n} \end{cases}$$

$\gamma$  est un processus de Markov



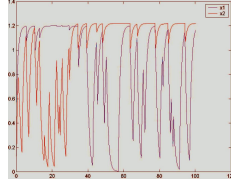
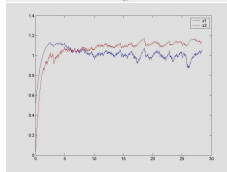
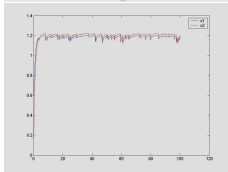
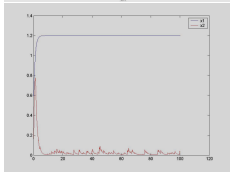
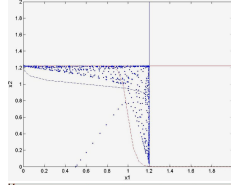
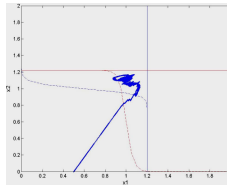
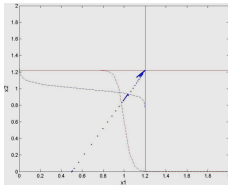
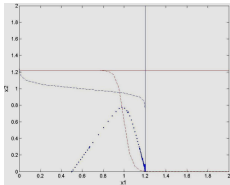
Sources: O. Radulescu (IRMAR, Rennes)

- Deux états:  $\gamma(t) \in \{0, 1\}$
- Un saut après un temps moyen  $\tau = \exp[(T_1 + T_2)^{-1}]$ .
- Proba de saut  $0 \rightarrow 1$ :  $\pi_1 = \frac{T_2}{T_1 + T_2}$
- Proba de saut  $1 \rightarrow 0$ :  $\pi_2 = \frac{T_1}{T_1 + T_2}$



# Modélisation stochastique: Exemple

*Beaucoup plus de comportements que dans le cas différentiel*



convergence vers  
(1, 0) (attracteur  
diff.)

convergence vers  
(1, 1) (non  
attracteur diff.)

oscillation autour  
d'un point  
quelconque

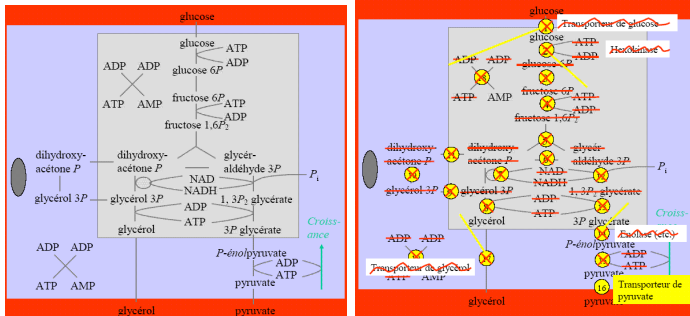
comportement  
erratique

# Application différentielle

- Les modèles différentiels sont une **idéalisation** qui permettent de générer des simulations et des prédictions qualitatives. Mais ne permettent pas nécessairement une description du point de vue biologique.
- Principalement utilisé pour l'étude des **réseaux métaboliques**.
- **Simulations** de comportements (Gepasi; P. Mendes)
- **Analyse convexe**: décomposition des réseaux en voies élémentaires (Flux Balance Analysis)
- Etude et descriptions de trajectoires.

## Application différentielle: round-up

Le seul moyen d'amener une cellule à la mort par **surproduction non contrôlée** lors de la glycolyse est d'agir sur les transporteurs du pyruvate

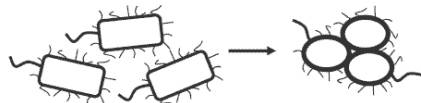
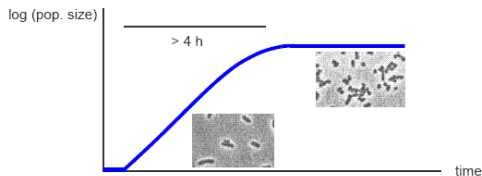


Sources: A. Cornish-Bowden (IBSM, Marseille)

## Application PL: E. Coli

- Sous l'effet d'un **stress nutritionnel**, E. Coli passe en phase stationnaire (croissance lente)
- La transition est contrôlée par un **réseau de régulation génétique** intégrant divers signaux environnementaux

- Peu de données qualitatives
- Modèle de la réponse au stress
  - lecture de la biblio
  - **modèle linéaire par morceaux (7 équations)**
  - modélisation des boîtes noires à l'aide de **simulations différentielles**

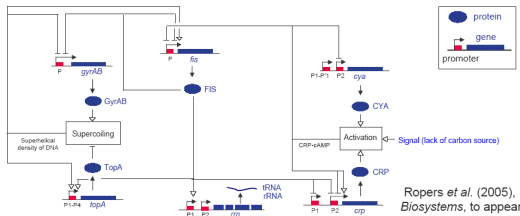


Sources: D. Ropers, H. De Jong (Inria, Grenoble)

## Application PL: E. Coli

- Sous l'effet d'un **stress nutritionnel**, E. Coli passe en phase stationnaire (croissance lente)
- La transition est contrôlée par un **réseau de régulation génétique** intégrant divers signaux environnementaux

- Peu de données qualitatives
- Modèle de la réponse au stress
  - lecture de la biblio
  - **modèle linéaire par morceaux (7 équations)**
  - modélisation des boîtes noires à l'aide de **simulations différentielles**



Sources: D. Ropers, H. De Jong (Inria, Grenoble)

Prédiction d'oscillations en cours de vérification

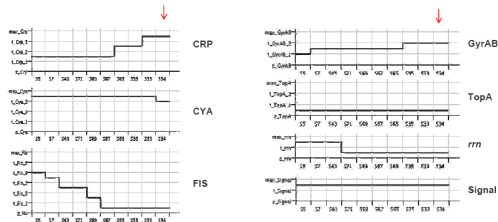
## Application PL: E. Coli

- Sous l'effet d'un **stress nutritionnel**, E. Coli passe en phase stationnaire (croissance lente)
- La transition est contrôlée par un **réseau de régulation génétique** intégrant divers signaux environnementaux

- Peu de données qualitatives
- Modèle de la réponse au stress
  - lecture de la biblio
  - **modèle linéaire par morceaux (7 équations)**
  - modélisation des boites noires à l'aide de **simulations différentielles**

Simulation of transition from exponential to stationary phase

State transition graph with 26 states generated in < 1 s, single **equilibrium state**

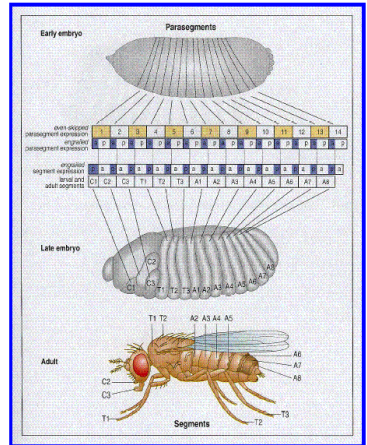


Sources: D. Ropers, H. De Jong (Inria, Grenoble)

Prédiction d'oscillations en cours de vérification

## Application booléen : drosophile

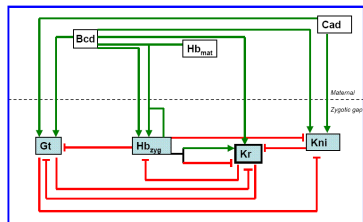
- Formation des **motifs dorso-ventraux** chez la Drosophile
- Expressions de gènes différentes dans les zones de l'embryon qui donneront des motifs chez l'adulte
- **Modèle**: différents modules de régulation imbriqués
- Le gap module: modèle(s) **multivalué**
- **simulations compatibles avec les XP sur mutants**



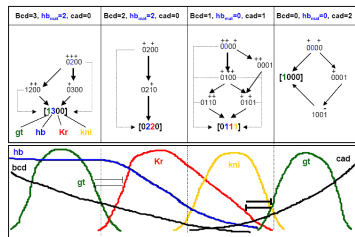
Sources: Wolpert et al. (1998), D. Thieffry (IBDM)

## Application booléen : drosophile

- Formation des **motifs dorso-ventraux** chez la Drosophile
- Expressions de gènes différentes dans les zones de l'embryon qui donneront des motifs chez l'adulte
- **Modèle**: différents modules de régulation imbriqués
- Le gap module: modèle(s) **multivalué**
- simulations compatibles avec les XP sur mutants



Gap Module - Simulation (gt, hb<sub>zyg</sub>, Kr, kni)



Sources: Wolpert et al. (1998), D. Thieffry (IBDM)



## Application: mucoviscidose.

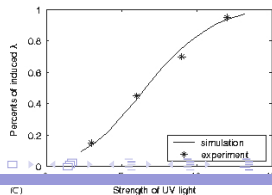
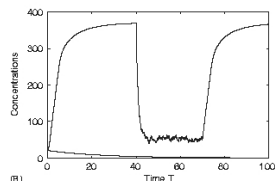
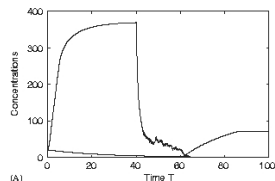
- production de mucus chez la bactérie *Pseudomonas Aeruginosa*
- cause principale de mortalité chez les malades atteints de mucoviscidose
- explication admise: **mutation** sélectionnée dans les poumons malades (irréversible)
- **logique booléenne**: un comportement **multistationnaire (et réversible)** est cohérent avec les connaissances biologiques
- nouveaux traitements médicaux ?
- plan expérimentaux en cours

Sources: J. Guespin-Michel et al (2005)

## Application stochastique: le phage lambda

- **Systeme bistable**: phases lytique et lisogénique
- **Switch**: les cellules passent de la lysogène à la lyse sous l'action d'UV
- Les modèles différentiels n'expliquent pas les switchs observés
- Il faut introduire du **bruit intrinsèque**, dû au faible nombre de molécules
- le taux de switch prédit par le modèle est très proche des courbes expérimentales

Sources: Tian and Burrage (2004)



## Comparaison de modèles

Les différentes modélisations ne produisent pas les mêmes simulations et prédictions.

Le choix d'un formalisme est la première étape de la modélisation.

- **Différentiel**: Grand nombre de variables. Bonnes simulations. Surtout métabolisme (besoin de coefficients)
- **Piecewise linear**: Surtout génétique. Etude de tous les comportements possibles. Dégradation obligatoire.
- **Booléen**: Plus de liberté pour définir les attracteurs. Moins de flou possible. Introduction du métabolisme possible.
- **Stochastique**: Petit nombre de molécules. Grande liberté de modélisation. Calculs horriblement longs.

# Un point commun à tous les formalismes

## Graphe d'interaction

- **Sommets**: noeuds du modèle
- **Arêtes**:  $A \rightarrow B$  si une perturbation de la concentration de  $A$  a un effet sur la concentration de  $B$ . On garde le signe de l'effet.



$$\begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}$$

(-1 sur la diagonale:  
retorégulations négatives;  
dégradations)

- Différentes définitions en fonction du type de modélisation
- Modèles différentiels: la matrice d'adjacence du graphe est donnée par le **Jacobien** de la dynamique

# Multistabilité

- **Système multistationnaire**: existence d'au moins deux points fixes stables.
- Les **switchs** (changement d'état d'équilibres) sont cruciaux pour expliquer les phénomènes de **différentiation**.
- Règle de Thomas. *Si un système de régulation génétique a deux états d'équilibre, alors son graphe d'interaction admet au moins un circuit positif.*
  - Preuve par Snoussi (95) dans un cas différentiel
  - Preuve par Soulé (03) dans le cas différentiel général et linéaire par morceaux.
  - Preuve par De Jong (98), puis Rémy et al. (05) dans le cas booléen.
- **Condition pas suffisante**: la boucle positive ne doit pas être affectée par d'autres circuits.
- **Validité** dans le cas stochastique? pour les systèmes métaboliques?

## Motifs

## Synthesis of auto-regulated gene circuits

	Gardner <i>et al.</i> (2000) <i>Nature</i> <b>403</b> : 339-342	Elowitz & Leibler (2000) <i>Nature</i> <b>403</b> : 335-338	Becskei & Serrano (2000) <i>Nature</i> <b>405</b> : 590-593
Construction			
Logical scheme	<b>Positive circuit</b> 	<b>Negative circuit</b> 	<b>Negative circuit</b> 
Main properties	<ul style="list-style-type: none"> <li>- Stable and exclusive expression of one of the two repressors</li> <li>- Memorisation of induction</li> <li>- Stability and robustness against biochemical fluctuations</li> </ul>	<ul style="list-style-type: none"> <li>- Cyclic expression of the repressors and reporter gene</li> <li>- Transmission of this oscillating behaviour through bacterial divisions</li> </ul>	<ul style="list-style-type: none"> <li>- Increased stability and decreased variability of the repressor expression</li> <li>- Compensation of dosage effects due to the variation of the number of copies</li> </ul>

Sources: D. Thieffry (IBDM, Marseille)

## Sources

- A. Cornish-Bowden (Métabolisme)
- G. Bernot (Booléen)
- K. Burrage (Phage Lambda)
- H. De Jong (Piecewise Linear)
- J. Guespin-Michel (Mucoviscidose)
- E. Pécou (Différentiel)
- O. Radulescu (Stochastique)
- E. Sontag (Différentiel, Contrôle)
- C. Soulé (Règle de Thomas)
- D. Thieffry (Booléen multivalué)
- J. Tyson (Différentiel)
- Z. Xie (Lambda Phage, stochastique)
- (...)