

# Biologie des systèmes

**But : comprendre comment les réseaux d'interactions complexes contrôlent le comportement de la cellule.**

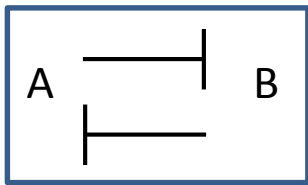
Dans les années 1950, naissance de la biologie moléculaire. Depuis cette discipline a permis d'acquérir de nombreuses données et connaissances sur les processus biomoléculaires, donc à des échelles très petites. Cette approche s'intéressant à des molécules individuelles et/ou à des interactions moléculaires a été appelée « approche réductionniste ».

Début des années 2000, mise en place de nouvelles techniques expérimentales appelées approches à « haut débit » permettant aux chercheurs d'observer simultanément le comportement d'un grand nombre d'espèces moléculaires différentes (différence du taux des ARNm produit par une cellule dans deux conditions différentes par exemple). Ces expériences permettent d'étudier le comportement de systèmes moléculaires dans leur ensemble et donc de prendre en compte les interactions entre composés du système. Ces approches ont conduit au développement de la **Biologie des Systèmes**. Pour comprendre comment le comportement des systèmes évolue au cours du temps, il faut prendre en compte la dynamique des systèmes.

# Description des interactions moléculaires par les biologistes et construction de modèle

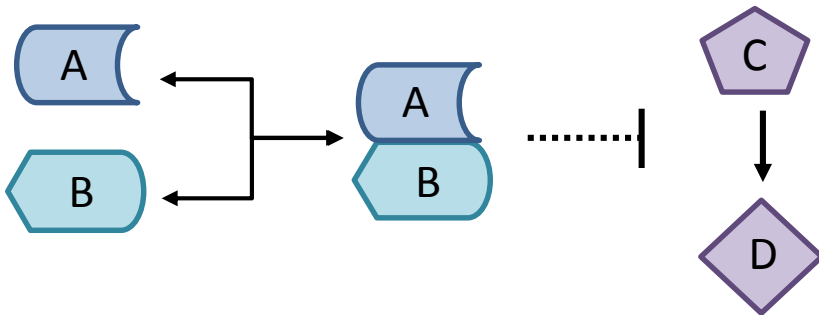
Utilisation de modèle conceptuel utilisant une description verbale du système accompagnée par des diagrammes illustrant comment un ensemble de composés interagissent.

## Un exemple simple entre deux composés moléculaires :



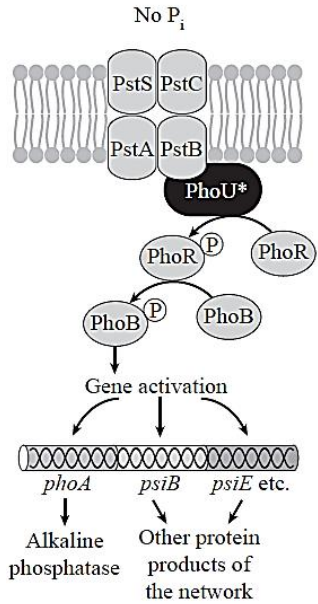
Le produit protéique A du gène *a* va inhiber l'expression du gène *b* et inversement le produit protéique B du gène *b* va inhiber l'expression du gène *a*.

## Un exemple un peu plus compliqué :



Les espèces moléculaires A et B interagissent de façon réversible pour former un complexe. Ce complexe inhibe la réaction de la conversion de la molécule C en D. La ligne en pointillés indique que l'action du complexe est une interaction de type régulation au cours de laquelle le complexe n'est pas consommé.

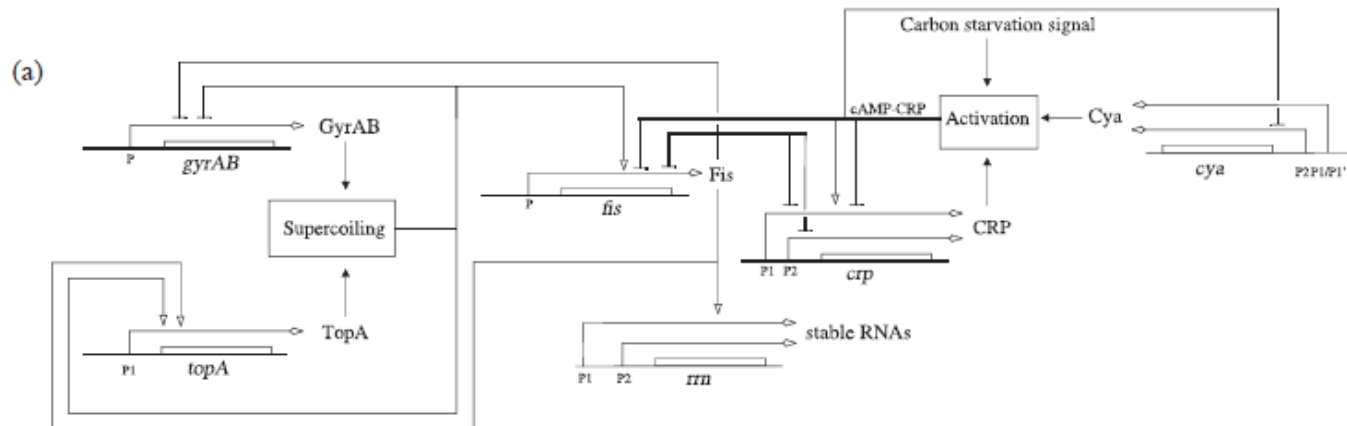
# Description des interactions moléculaires par les biologistes et construction de modèle



Représentation d'une cascade de régulation :

La régulation du phosphate chez les entérobactéries

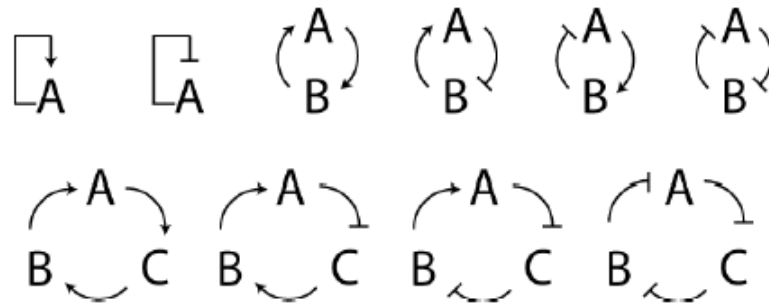
Réseau des gènes, protéines clefs et des interactions de régulation impliqués dans la réponse au stress nutritionnel chez *E.coli*.



# Etude du comportement dynamique du réseau (système) : modélisation mathématique

Pourquoi a-t-on besoin de la modélisation ?

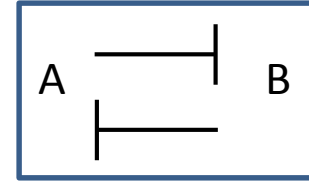
Parmi les réseaux simples ci-dessous impliquant trois protéines, pouvez-vous identifier ceux qui vont conduire à l'oscillation de la concentrations des protéines (A, B et C) ? Pas si facile !



Plus le nombre de composés et des interactions du systèmes augmente, plus il devient difficile d'avoir une compréhension intuitive de son comportement global.

# Etude du comportement dynamique du réseau (système) : modélisation mathématique

Revenons à l'exemple simple de répression mutuelle :



En fonction des concentrations initiales des protéines A et B, le système atteindra deux états stables différents. Comment le savoir ?

Ecriture d'équations différentielles ordinaires permettant d'exprimer la taux de synthèse d'un composé donné en fonction de la concentration des autres composés du système.

Forme générale de l'équation :

$$\frac{dx}{dt} = \text{synthesis}(x) - \text{degradation}(x)$$

Ici le taux de synthèse de A (B) va dépendre de la concentration de B (A) qui a une action d'inhibition. Ceci peut être approximé par une fonction de Hill. On considérera que la dégradation n'est pas régulée et dépend d'une constante de dégradation

$$\frac{d[A]}{dt} = \beta_{\max} \frac{K_d^n}{[B]^n + K_d^n} - \gamma[A]$$

$$\frac{d[B]}{dt} = \beta_{\max} \frac{K_d^n}{[A]^n + K_d^n} - \gamma[B]$$

Avec :

$\beta_{\max}$  : taux maximal de synthèse de A (B) donc en absence de répression

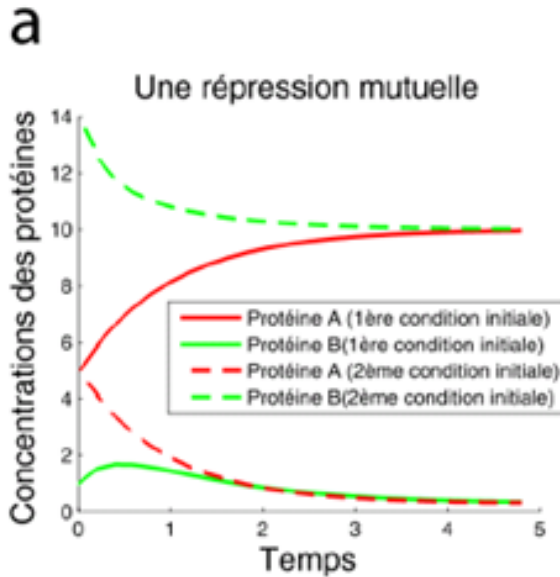
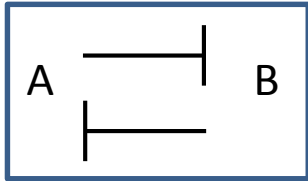
$K_d$  : concentration de B (A) nécessaire pour atteindre la moitié de la répression maximale de A (B) = coefficient de répression.

$n$  : coefficient de Hill

$\gamma$  : constante de dégradation

# Etude du comportement dynamique du réseau (système) : modélisation mathématique

Intégration numérique des équations différentielles et obtention des valeurs simulées des concentrations de protéines au cours du temps (profil dynamique) :

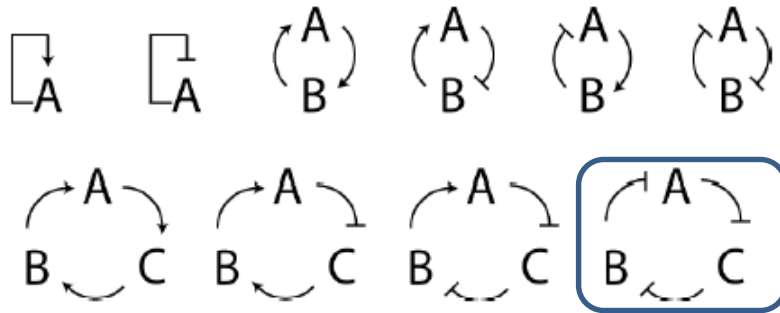


Concentrations initiales des protéines A et B et devenir de l'état du système :

- Si A est présente à haute concentration au début et B à faible concentration, le système atteint un état d'équilibre avec beaucoup de protéines A (courbe rouge trait plein) et peu de protéines B (courbe verte trait plein)
- Si B est présente à une concentration plus élevée que celle de A au début, le système atteint un état d'équilibre inverse avec peu de protéines A (courbe rouge en pointillés) et beaucoup de protéines B (courbe verte en pointillés)
- On dit que le système est **bistable**
- Ce type de motif est appelé un **interrupteur (toggle switch)** car en perturbant/changeant suffisamment la concentration initiale d'une protéine, on peut faire basculer le système vers un état d'équilibre ou vers l'autre

# Etude du comportement dynamique du réseau (système) : modélisation mathématique

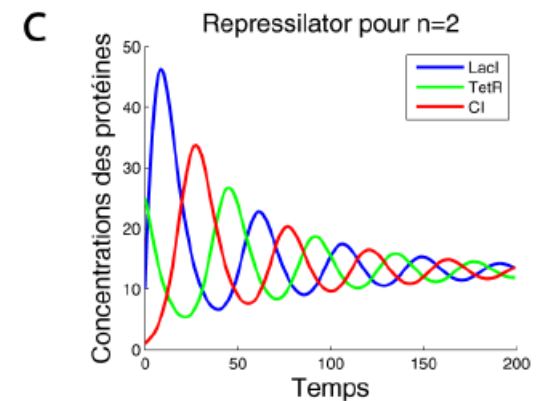
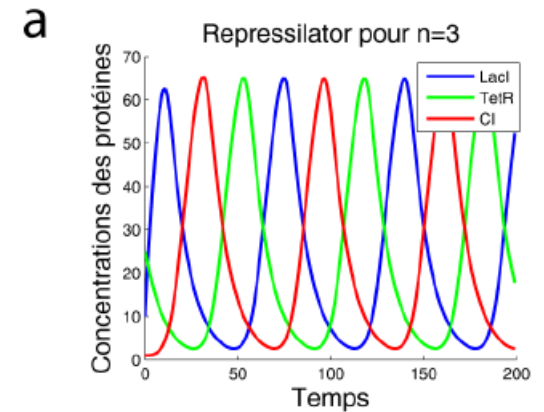
Revenons à notre question : Parmi les réseaux simples ci-dessous impliquant trois protéines, pouvez-vous identifier ceux qui vont conduire à l'oscillation de la concentrations des protéines (A, B et C) ? Pas si facile !



Au moins un, le dernier appelé le « repressilator ».

On observe bien des oscillations des trois protéines en fonction du temps. Cependant, la encore la valeur des paramètres est importante.

- paramètre de Hill  $n = 3$ , les oscillations perdurent et aucun état d'équilibre est atteint.
- paramètre de Hill  $n = 2$ , les oscillations s'atténuent et le système atteint un état d'équilibre stationnaire



# Les réseaux de régulation de l'expression des gènes

La cellule est un dispositif intégré composé de plusieurs milliers de types de protéines interagissant.

La cellule rencontre différentes situations qui requièrent différentes protéines. Par exemple : i) si dans le milieu extérieur une cellule bactérienne « sent » la présence de galactose, elle va commencer à produire les protéines nécessaires à son import dans la cellule et à son catabolisme ou i) si sa membrane est endommagée, elle va produire les protéines pour la réparer.

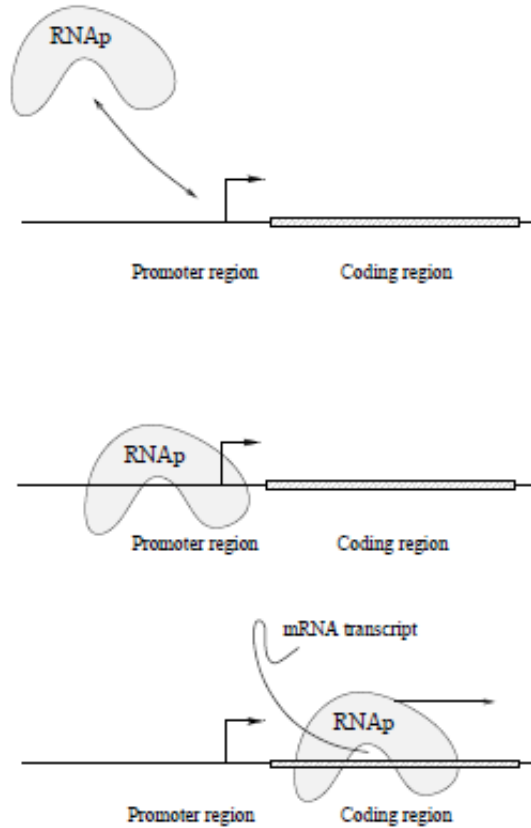
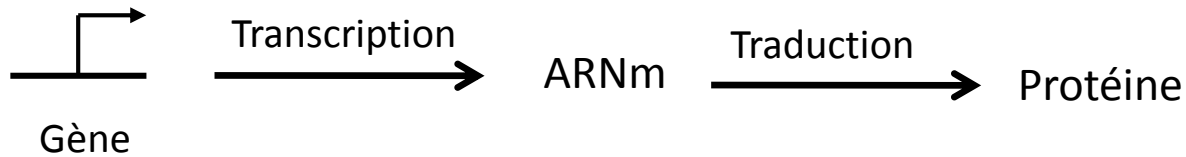
La cellule vit dans un environnement complexe et peut sentir un grand nombre de signaux différents, incluant des paramètres physiques comme la température ou la pression osmotique, la présence de nutriments, la présence de substances chimiques nocives etc. Les informations sur son état interne est également important (niveaux de métabolites clefs par exemple).

La cellule répond à ces signaux en produisant les protéines appropriées à la réponse.

La cellule doit donc surveiller continuellement son environnement et « calculer » la quantité et le type de protéines à synthétiser. La fonction de traitement de l'information qui va déterminer le taux de production de chaque protéine est en majorité réalisé par des réseaux de transcription



# Les éléments des réseaux de transcription



## Absence de régulation de la transcription :

La RNA polymerase (RNAP) va reconnaître à une région spécifique de l'ADN en amont du gène à transcrire appelée promoteur. Elle va se lier au promoteur et initier la copie d'un brin de l'ADN en ARN. Dans le cas d'un gène codant pour une protéine cet ARN est appelé ARNm

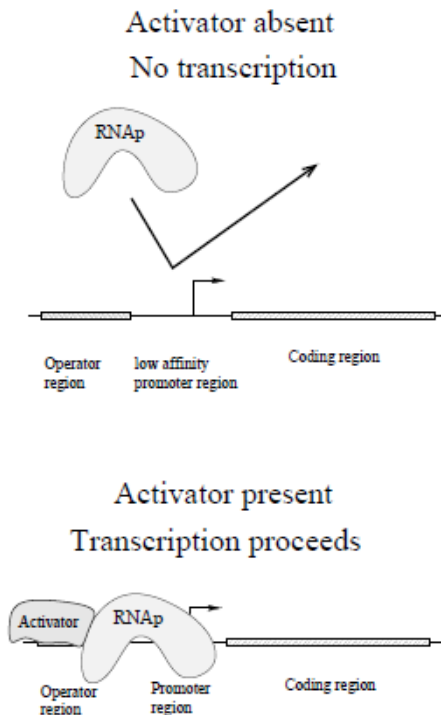
La qualité de cette liaison va spécifier le taux de transcription des gènes cibles

# Les éléments des réseaux de transcription

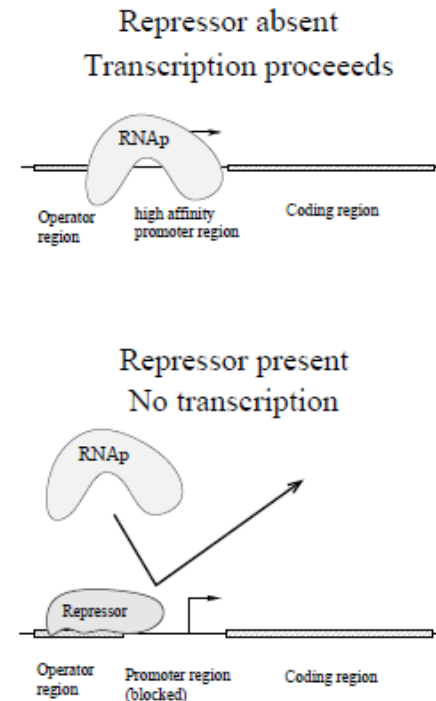
Le taux de transcription d'un gène peut être régulé (modulé) par des protéines appelées **facteurs de transcription**. Ces facteurs affectent le taux de transcription en se fixant sur des sites spécifiques du promoteur ou sur des régions au voisinage du promoteur du moins chez les bactéries (plus compliqué chez les eucaryotes).

Les facteurs de transcription peuvent agir comme des **activateurs** et augmenter le taux de transcription du gène ou comme des **répresseurs** qui réduisent le taux de transcription.

## Régulation par un activateur



## Régulation par un répresseur



# Les éléments des réseaux de transcription

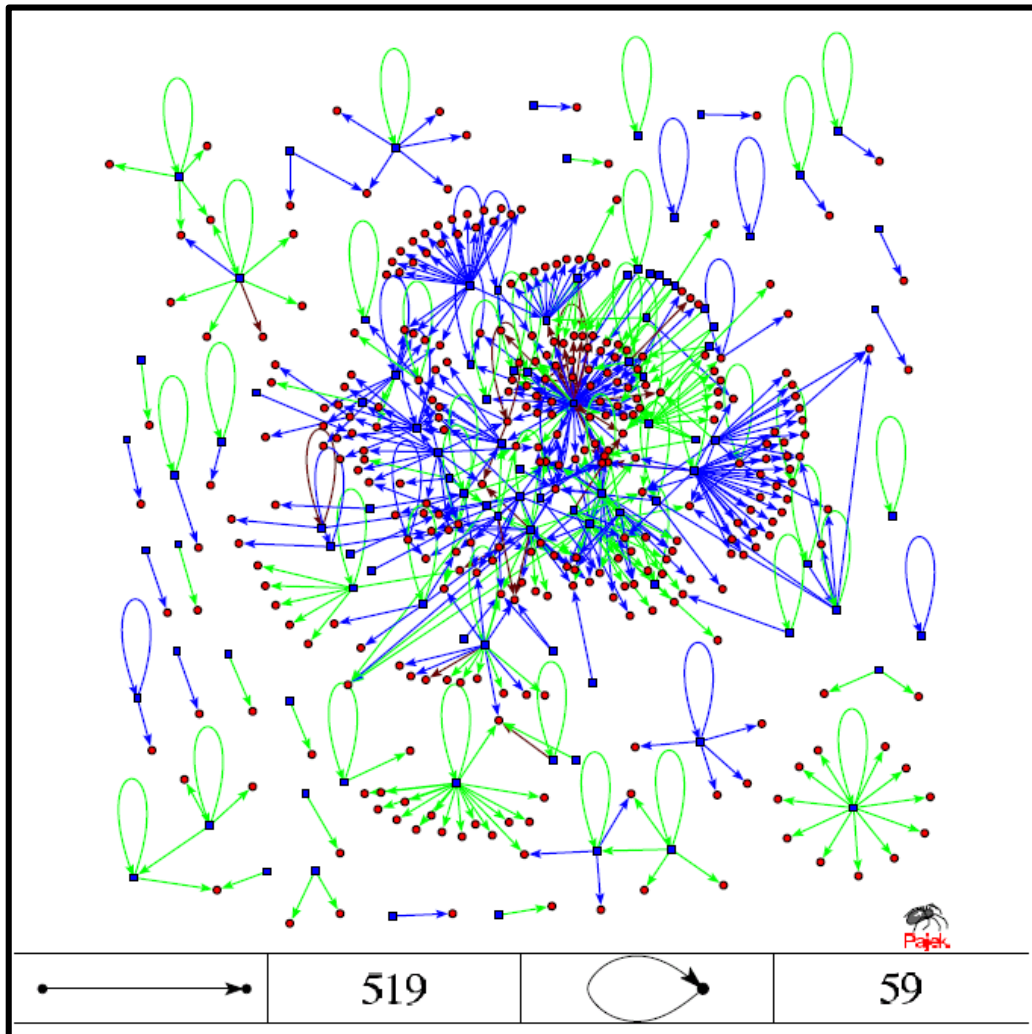
Ces facteurs de transcription sont des protéines codées par des gènes qui peuvent eux-mêmes être régulés par d'autres facteurs de transcriptions qui peuvent eux aussi être régulés par d'autres facteurs de transcriptions etc.

Cette cascade de régulation est appelée réseaux de transcription.

L'environnement est perçu par la cellule par l'intermédiaire de signaux. Ces stimuli extérieurs activent des voies de transduction du signal qui à la suite de cascades de régulation vont permettre la réponse de la cellule (transcription des gènes).

Le réseau représente donc un système dynamique.

# Exemple des réseaux de transcription chez *E. coli*



Les carrés bleus représentent les facteurs de transcription (TF).

Les cercles rouges identifient les opérons (ensemble de gènes co-transcrits) qui sont régulés par les TFs.

Les liens (flèches) sont coloriées en fonction de l'action du TF :

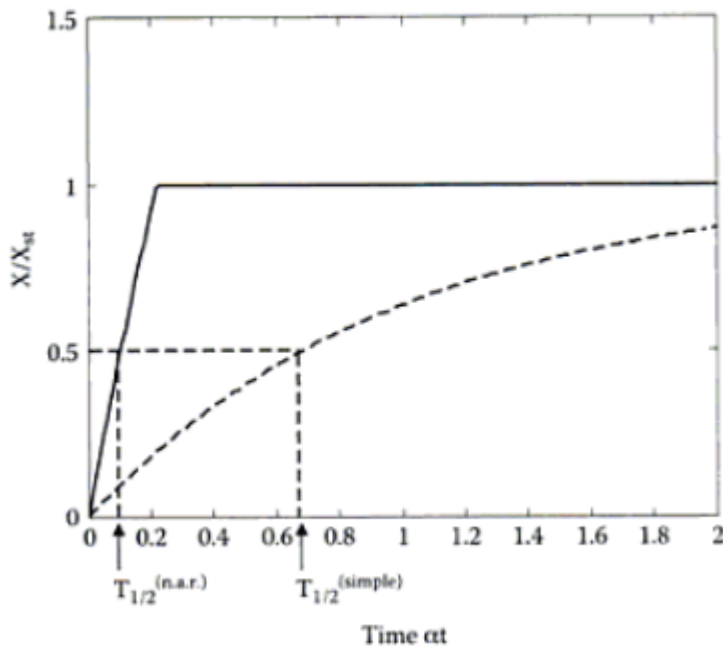
- bleu le TF est un activateur
- vert le TF est un répresseur
- marron le TF a les deux fonctions

Le nombre de deux types de liens élémentaires sont données en bas de la figure :

- boucles d'auto-régulation (le TF régule sa propre synthèse) : 59 qui correspond à environ la moitié des TFs d'*E. coli*
- Lien direct (le TF régule un autre TF ou un opéron) : 519

# Boucles d'autorégulation négative

Autorégulation négative : quand le facteur de transcription réprime sa propre transcription en se liant à son propre promoteur pour inhiber la synthèse de son ARNm  
⇒ Plus élevée est la concentration de X, plus faible est son taux de synthèse



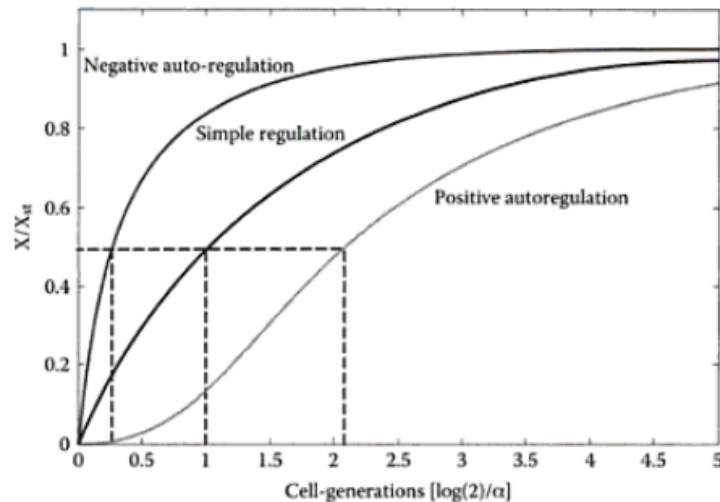
Comparaison de la dynamique du produit protéique d'un gène autorégulé négativement (ligne pleine) et de celui d'un gène régulé simplement (ligne en pointillés) qui atteignent un même état stable et qui possèdent les mêmes constantes de dégradation  
Le temps de réponse correspond au temps nécessaire pour que le niveau de concentration de la protéine atteigne 50% du taux correspondant à son état d'équilibre ( $T_{1/2}$ ).

**L'autorégulation négative accélère, entre autre, le temps de réponse.**

# Boucles d'autorégulation positive

Autorégulation positive : quand le facteur de transcription active sa propre transcription en se liant à son propre promoteur

⇒ La dynamique est initialement lente, puis avec l'accumulation de X, le taux de production augmente



**FIGURE 3.5** Dynamics of a negatively autoregulated gene, a simply regulated gene and a positively autoregulated gene. The negatively and positively autoregulated genes have a Hill input function with Hill coefficient  $n = 1$ . Shown is protein concentration normalized by its steady-state value,  $X/X_{st}$ , following an increase in production rate. Time is in cell generations, or for actively degraded proteins,  $\log(2)/\alpha$ , where  $\alpha$  is the protein degradation/dilution rate. The response time is found by the intersect of the dynamics with a horizontal line at  $X/X_{st} = 0.5$ .

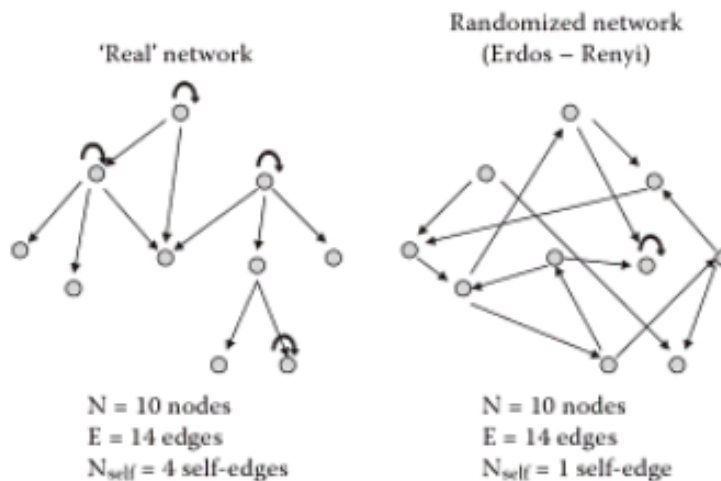
L'autorégulation positive ralentit le temps de réponse par rapport à une simple régulation.

La dynamique commence lentement et le taux de production augmente à mesure que X s'accumule.

Si la valeur de  $\beta_{max}$  (taux maximal de synthèse de X) est plus élevé que la valeur de  $\gamma$  (taux de dégradation), le système devient bistable. Une fois que le gène est activé, il est verrouillé dans un état de forte expression et il se maintient dans l'état ON, même si le signal à l'origine de l'activation a disparu (utilisé par exemple au cours du développement pour assurer l'irréversibilité d'une décision comme le type de cellule)

# Analyse des réseaux de transcription : recherche de motifs dans le réseau

On compare le réseau réel à un réseau aléatoire. Le réseau aléatoire garde les mêmes caractéristiques que le réseau réel, *i. e.*, même nombre de nœuds et d'arêtes (liens) mais les connections entre les nœuds sont faites au hasard.

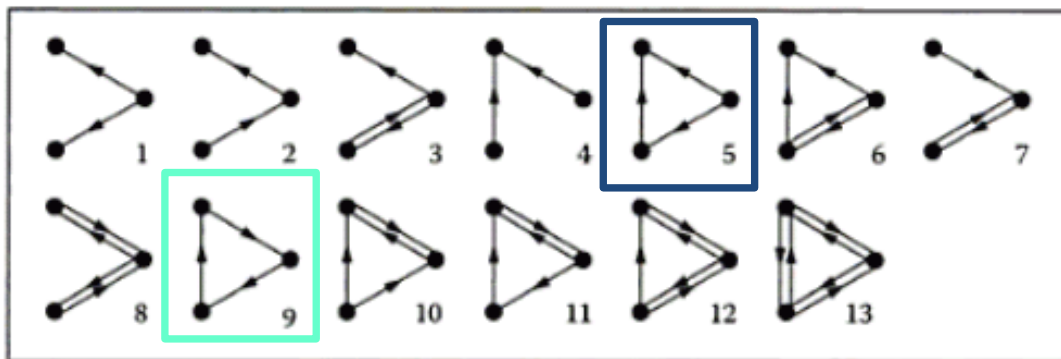


Un motif dans le réseau = un pattern que l'on retrouve beaucoup plus souvent dans le réseau réel que dans le réseau aléatoire

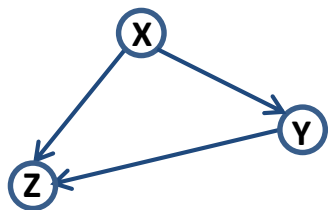
Idée de base : les motifs qui apparaissent dans le réseau réel plus souvent que dans le réseau aléatoire ont du être préservés au cours de l'évolution contre des mutations qui changent les liens de façon aléatoire (par exemple, une mutation dans la séquence d'un promoteur peut abolir la liaison du facteur de transcription et conduire à la perte du lien (de l'arête) dans le réseau. Pression de sélection pour maintenir le lien).

# Analyse des réseaux de transcription : recherche de motifs dans le réseau

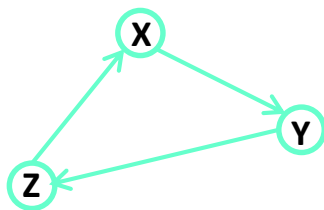
On s'intéresse aux motifs composés de trois nœuds : Il en existe 13 possibles



Seulement le 5<sup>ème</sup> est un motif dans le réseau (parfois le 9<sup>ème</sup>).



Le 5<sup>ème</sup> est appelé une feed-forward loop (FFL)



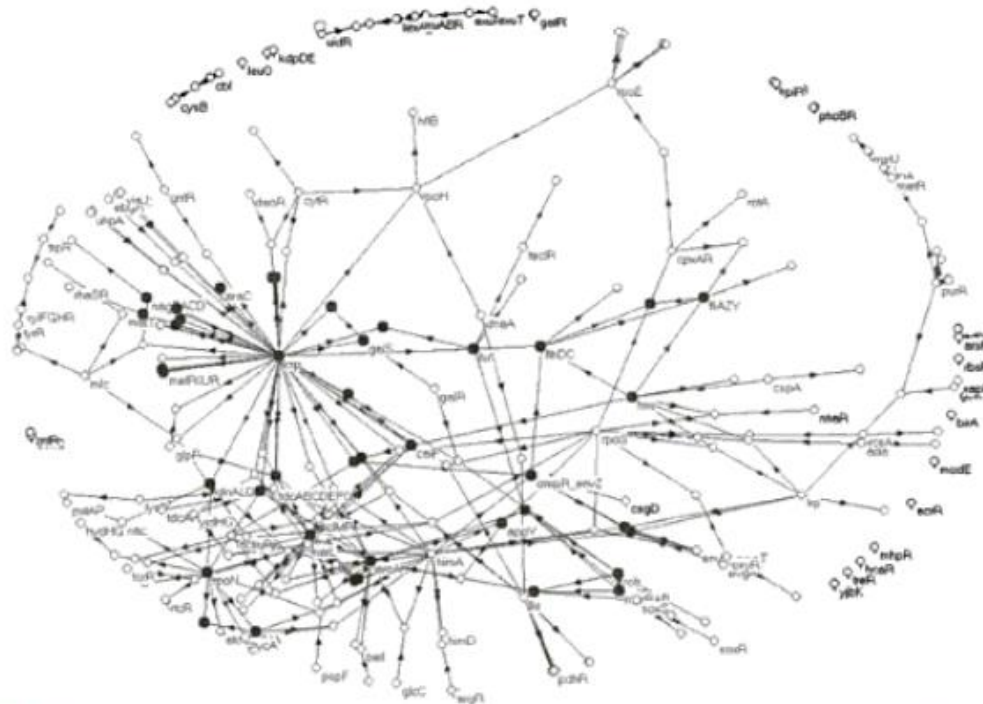
Le 9<sup>ème</sup> est appelé une feedback loop à 3 nœuds (boucle de rétroaction)

Analyse du réseau de transcription d'*E. coli* : 42 FFL identifiées et 0 boucle de rétroaction à 3 nœuds



# Feed-forward loop

Analyse du réseau de transcription d'*E. coli* utilise dans Alon (2007, *An introduction to systems Biology*), : 42 FFL identifiées et 0 boucle de rétroaction à 3 nœuds



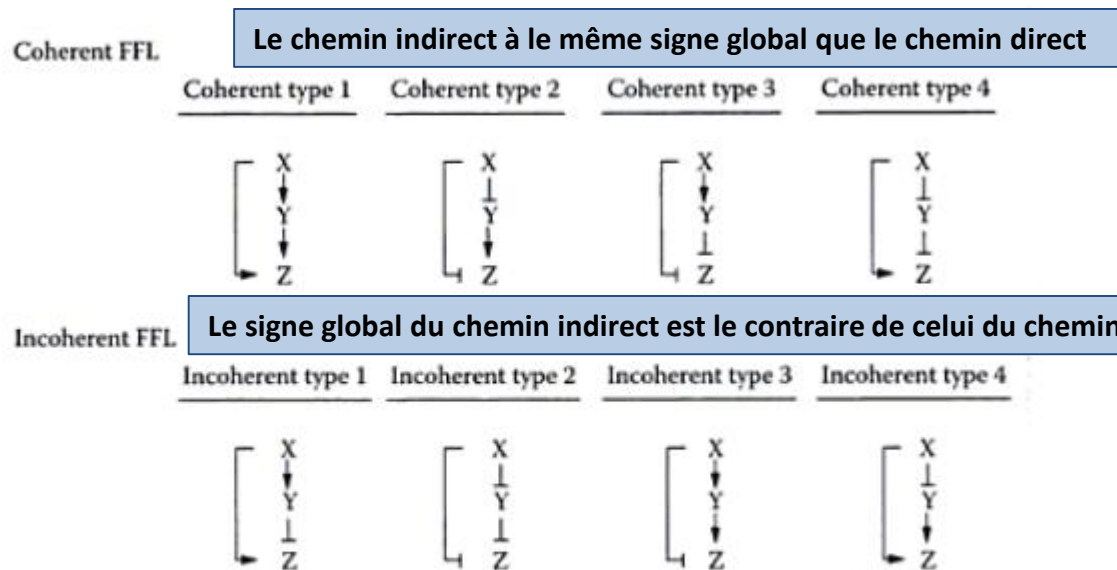
**FIGURE 4.2** Feed-forward loops in the *E. coli* transcription network. Black nodes participate in FFLs.

Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

# Feed-forward loop

Question : pourquoi une telle boucle a été sélectionnée ?

Une FFL est composé d'un facteur de transcription X qui régule le gène d'un second facteur de transcription Y, les deux facteurs X et Y régulent le gène Z. Chaque régulation peut être une activation ou bien une répression. Il y a donc 8 types possibles de FFLs qui peuvent être classés en deux groupes : les boucles cohérentes et les boucles incohérentes.

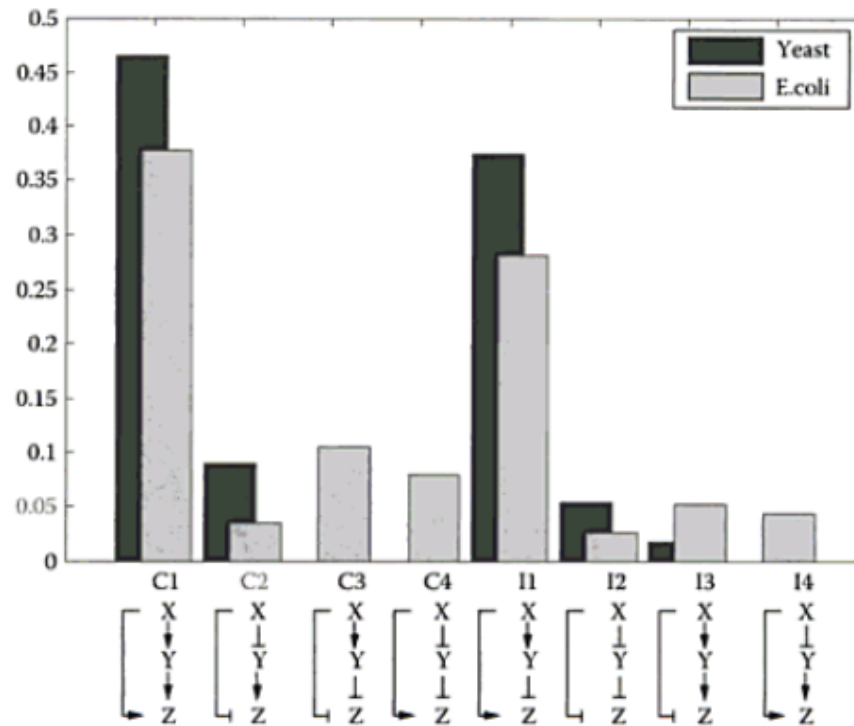


Calcul du signe global du chemin indirect = multiplication des signes des deux régulations (plus : activation, moins répression)

plus-plus → plus  
 plus-moins → moins  
 moins-moins → plus

**FIGURE 4.3** The eight sign combinations (types) of feed-forward loops. Arrows denote activation and  $\perp$  symbols denote repression.

# Feed-forward loop



**FIGURE 4.4** Relative abundance of the eight FFL types in the transcription networks of yeast and *E. coli*. FFL types are marked C and I for coherent and incoherent. The *E. coli* network is based on the Ecocyc and RegulonDB databases and has about twice as many edges as in the network of Figure 2.3. (From Mangan et al., 2006.)

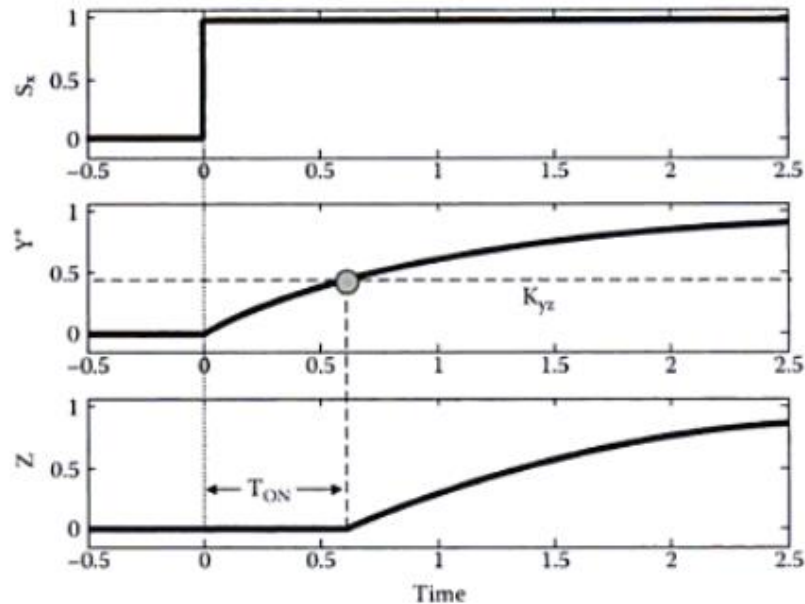
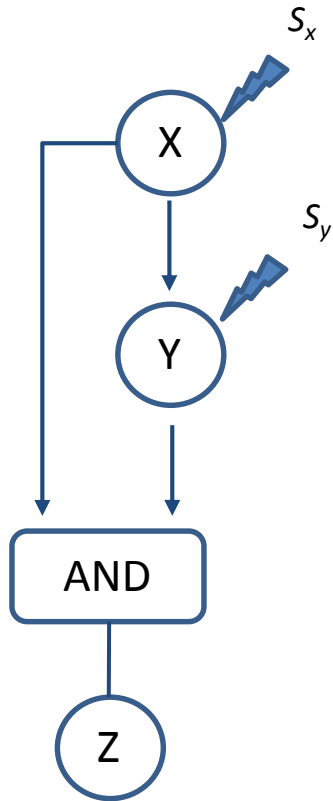
Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

Les différents types de FFLs pas trouvés avec le même fréquence.

La boucle cohérente C1-FFL où les trois régulations sont des activations (positives) est la plus fréquente suivie de la boucle incohérente I1-FFL.

# Dynamics of the C1-FFL with AND gate

L'activation of Z requiert la liaison des deux formes actives de X et Y ( $X^*$  et  $Y^*$ )



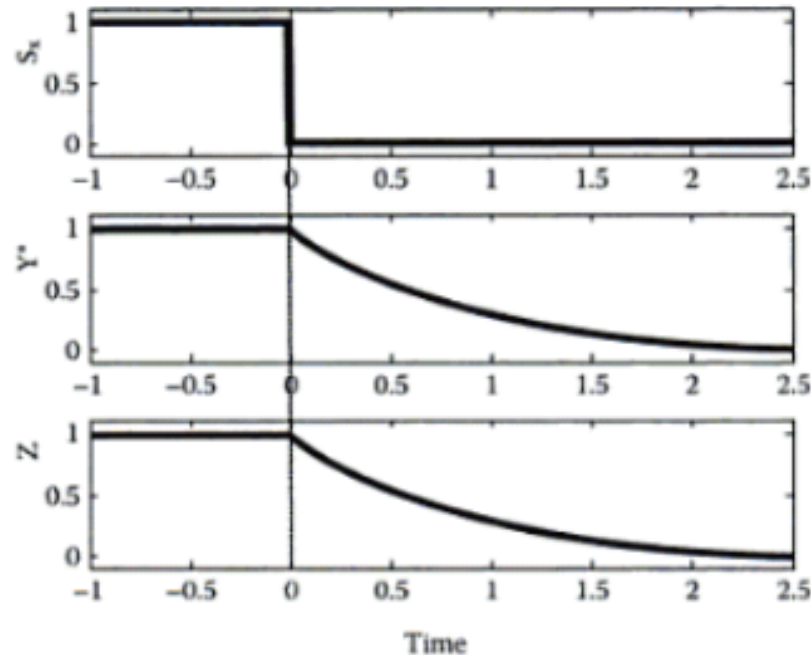
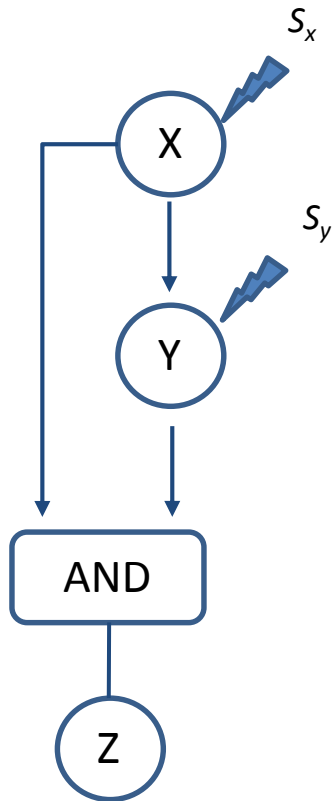
**FIGURE 4.7** Dynamics of the coherent type-1 FFL with AND logic following an ON step of  $S_x$  at time  $t = 0$  in the presence of  $S_y$ . The activation threshold of Z by Y is  $K_{yz}$  (horizontal dashed line). The production and degradation rates are  $\alpha_y = \alpha_z = 1$ ,  $\beta_y = \beta_z = 1$ . The delay in Z production is  $T_{ON}$ .

Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

Un signal fort  $S_x$  déclenche l'activation de X (simulation de type escalier) qui transite rapidement vers sa forme active  $X^*$ .  $X^*$  se lie au promoteur du gène Y qui initie la synthèse de la protéine Y. La protéine Y doit dépasser un niveau spécifique avant de pouvoir activer Z (seuil d'activation  $K_{yz}$ ) et le signal  $S_y$  doit être présent pour que Y passe sous sa forme active  $Y^*$ . Il en résulte un retard dans la production de Z

# Dynamics of the C1-FFL with AND gate

Quand le signal  $S_x$  stoppe, il n'y a pas de délai et la production de Z s'arrête aussitôt.



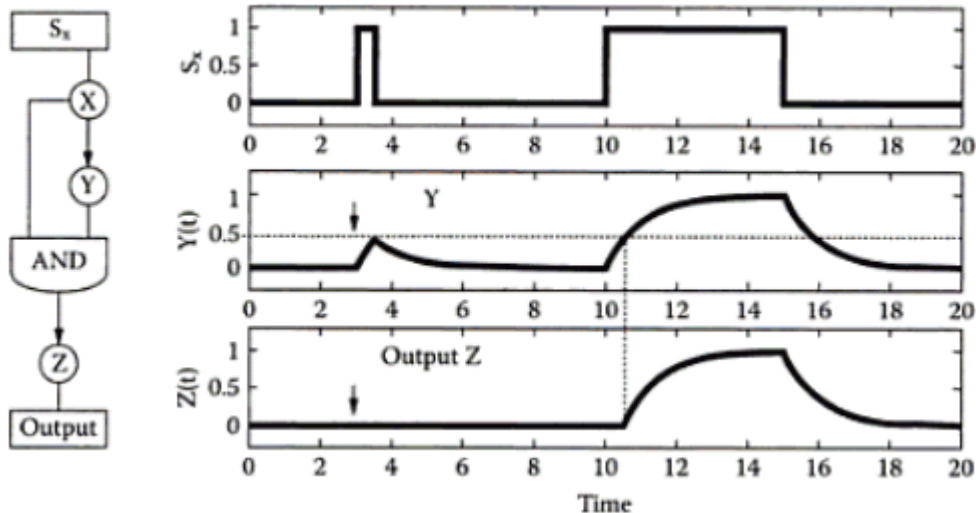
Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

Quand on enlève le signal  $S_x$  d'activation de X, X devient inactif et se décroche des promoteurs des gènes codant pour Y et Z. Comme l'activation de Z nécessite les formes actives  $X^*$  et  $Y^*$  des deux facteurs de transcription, l'activation de Z est stoppée.

# Dynamics of the C1-FFL with AND gate

Ce type de boucle est un détecteur de persistance d'une impulsion du signal

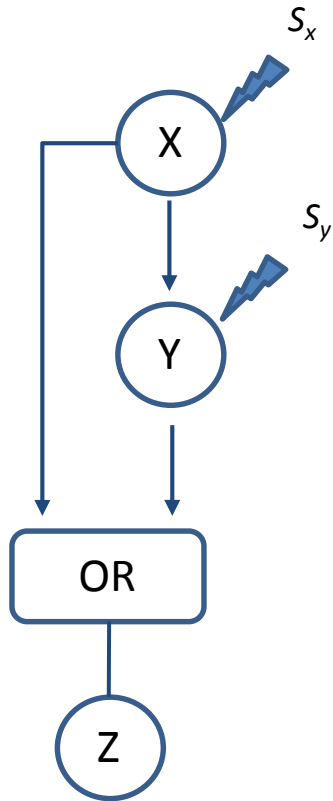
- Si la persistance du signal  $S_x$  est plus courte que le temps nécessaire pour que la concentration de la protéine Y atteigne la valeur  $K_{yz}$ , la protéine Z n'est pas produite.
- La protéine Z sera produite uniquement si le signal  $S_x$  persiste sur une durée qui excède le délai pour que la concentration de Y atteigne  $K_{yz}$ .
- Par contre, la boucle réagit immédiatement quand l'impulsion s'arrête.



Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

L'évolution peut avoir sélectionné les boucles C1-FFL quand la cellule a besoin d'une telle fonction de protection. En effet, l'environnement de la cellule est souvent soumis à des fluctuations. Parfois, les stimuli peuvent être présents sous forme d'impulsions courtes qui ne nécessitent pas le déclenchement de la réponse cellulaire.

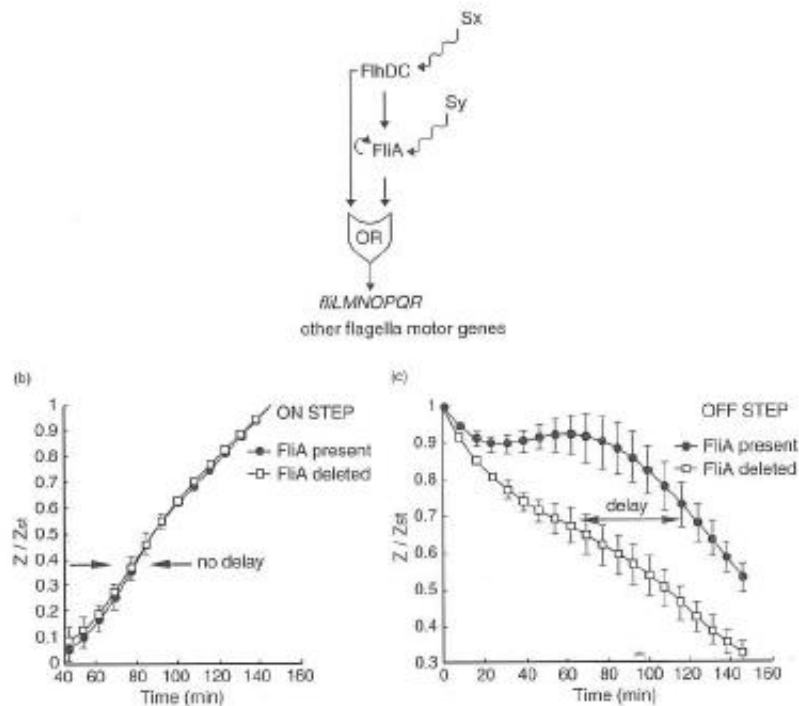
## Dynamics of the C1-FFL with OR gate



L'activation de Z nécessite la liaison d'une seule des deux formes actives de X ou de Y ( $X^*$  ou  $Y^*$ )

- Après une impulsion de  $S_x$  (ON), la transcription du gène codant pour Z est immédiatement activée car X acquiert rapidement sa forme active  $X^*$  et qu'une seule des deux formes actives des facteurs de transcription est nécessaire. Donc il n'y a pas de délai pour l'activation de Z.
- Par contre, quand l'impulsion de  $S_x$  cesse, il y aura un délai avant l'arrêt de l'activation de Z car  $Y^*$  peut activer Z en absence de  $X^*$ . Il faut que  $Y^*$  chute à une concentration inférieure à  $K_{yz}$  pour qu'il ne puisse plus activer Z. Le délai correspondra au temps nécessaire à  $Y^*$  d'atteindre une concentration  $< K_{yz}$ .
- une boucle C1-FFL avec une OR gate peut maintenir l'expression de Z même si le signal est perdu momentanément.
- Une telle dynamique a été démontré expérimentalement dans le cas du système du flagelle chez *E. coli*.

# Dynamics of the C1-FFL with OR gate

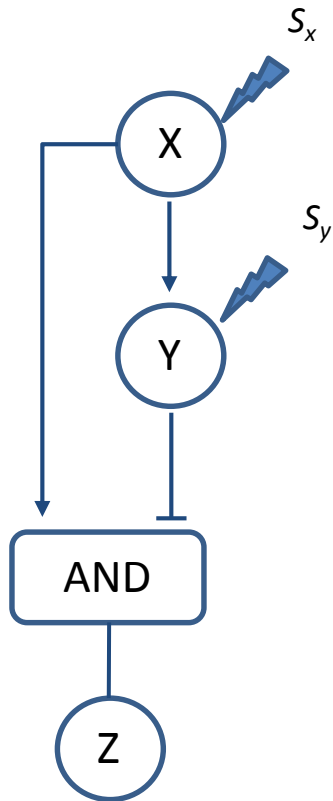


From Kalir *et al.* (2005) *Molecular System Biology*, doi: 10.1038:mbs4100010

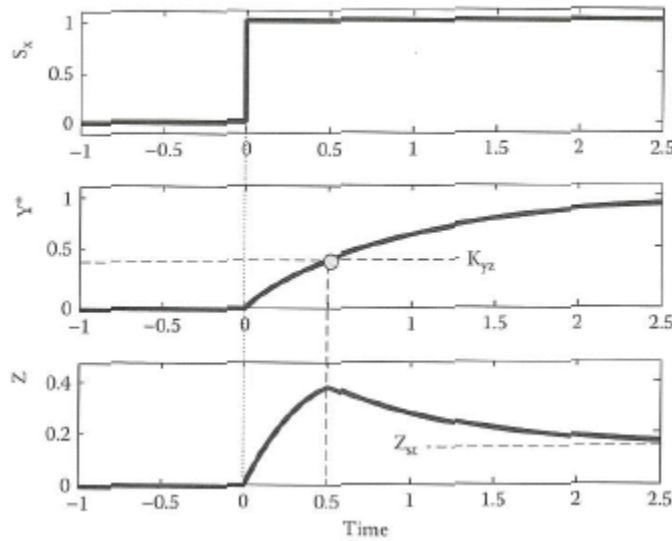
**Example of the C1-FFL OR gate in the flagella system of *E. coli*.** The input signals  $S_x$  are environmental signals like glucose limitation, osmotic pressure, temperature. The second signal  $S_y$  is a check point that is triggered when the first motors are completed. In b), after an ON step on  $S_x$ , the activity of the promoter  $fliL$  is measured by means of a green-fluorescent protein used as a reporter. In c) the promoter dynamics is reported after an OFF step of  $S_x$  in the presence of  $S_y$ . Black circles: wild type bacterium, white circles: mutant in which the gene  $fliA$  was depleted



# Dynamics of the I1-FFL with AND gate



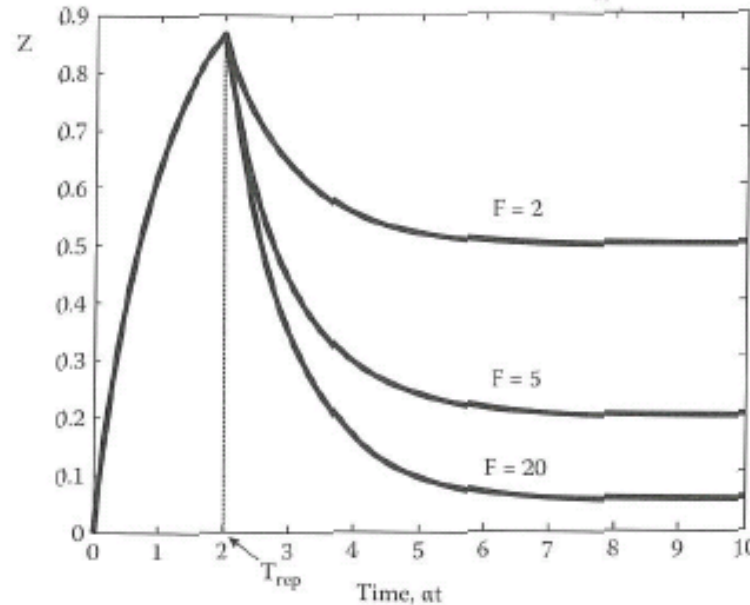
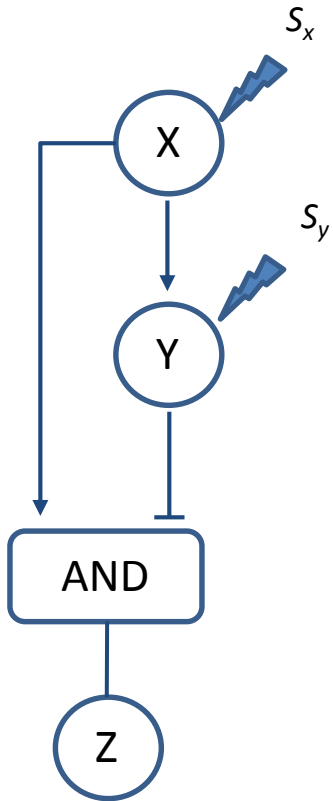
La boucle est constituée de deux chemins de régulation antagonistes : X active Z mais aussi Y qui en retour réprime Z.



Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

Après une impulsion de  $S_x$  (ON), le régulateur X devient actif ( $X^*$ ), se lie au promoteur du gène Z dont la transcription démarre conduisant à la synthèse de la protéine Z. La forme active de X ( $X^*$ ) se lie aussi au promoteur du gène Y et active la production de la protéine Y. Quand la concentration de la protéine  $Y^*$  atteint le seuil  $K_{yz}$  (coefficient de répression),  $Y^*$  commence à réprimer l'expression du gène Z et le niveau de la protéine Z décroît. Donc, les boucles I1-FFL AND gate peuvent générer un pulse de production de la protéine Z.

# Dynamics of the I1-FFL with AND gate

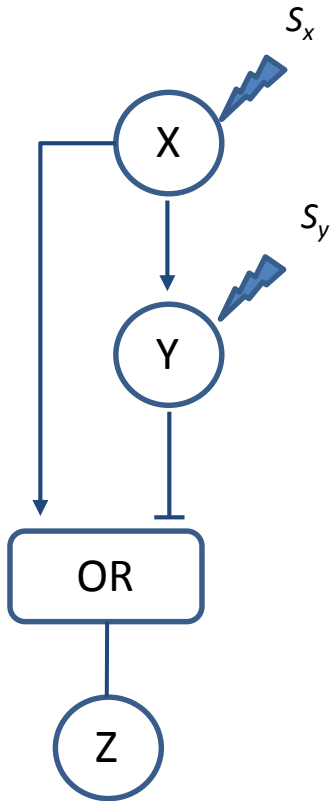


F is the repression factor

Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

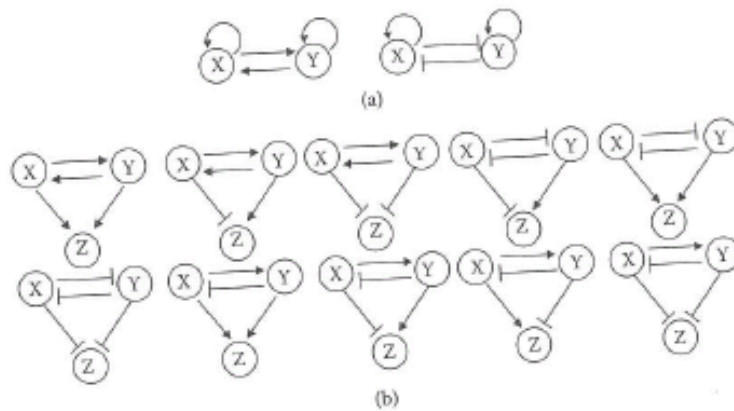
- Quand le répresseur Y a un fort effet inhibiteur sur la production de Z ( $F \gg 1$ ), la dynamique de Z a une forme analogue à une impulsion.
- De plus, il peut accélérer le temps de réponse du système compare à un simple circuit de régulation (not shown).
- Quand l'impulsion de  $S_x$  stoppe (OFF), il y a un shut-off immédiat de la production de Z .
- Thus, I1-FFL with AND gate is a sign-sensitive response regulator

## Dynamics of the I1-FFL with ORgate



Cette boucle va avoir la même fonction que celle avec une AND gate mais comme nous l'avons déjà vu pour la boucle CC1-FFL OR gate, elle peut maintenir l'expression de Z même si le signal d'input est momentanément perdu.

# Feed-back loop



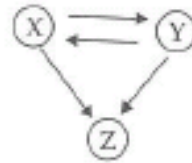
- (a) Two-nodes feedback loops with autoregulation. A common network motif in developmental transcriptional networks
- (b) The 10 different types of regulating feedback motifs

Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

Une boucle de rétroaction peut être soit positive soit négative en fonction du nombre d'interactions négatives. Si ce nombre est pair, alors la boucle de rétroaction est positive. Sinon, la boucle de rétroaction est négative (nombre impair d'interactions négatives).

# Feed-back loop

Double-positive feedback loop:

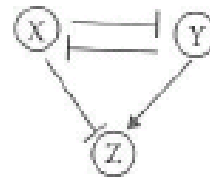


The gene Z is regulated as shown by the two transcriptional factors that activates each other. This loop will have two stable steady states:

A signal that causes the production of proteins X and Y can irreversibly lock the system into a state where X and Y are ON and activate each other. This case of bi-stable switch is called a lock-on mechanism.

	X	Y	Z
Steady state 1	ON	ON	ON
Steady state 2	OFF	OFF	OFF

Double-negative feedback loop:



The gene Z is regulated as shown by the two transcriptional factors that repressed each other. This loop will have two stable steady states:

Here, the loop expresses X either Y.

	X	Y	Z
Steady state 1	ON	OFF	OFF
Steady state 2	OFF	ON	ON

# Feed-back loop

## **Maintenance of homeostasis**

Almost all homeostatic control mechanisms are **negative feedback mechanisms**. These mechanisms change the variable back to its original state or “ideal value”.

Example: Control of blood sugar (glucose) by insulin:

When blood sugar rises, receptors in the body sense a change . In turn, the control center (pancreas) secretes insulin into the blood effectively lowering blood sugar levels. Once blood sugar levels reach homeostasis, the pancreas stops releasing insulin.

## **Positive feedback mechanisms**

A positive feedback mechanism is the exact opposite of a negative feedback mechanism. With negative feedback, the output reduces the original effect of the stimulus. In a positive feedback system, the output enhances the original stimulus.

Example of a positive feedback system is child birth. During labor, a hormone called oxytocin is released that intensifies and speeds up contractions. The increase in contractions causes more oxytocin to be released and the cycle goes on until the baby is born. The birth ends the release of oxytocin and ends the positive feedback mechanism.

# Modélisation d'un processus biologique

Question : quel est le comportement dynamique d'un processus biologique d'intérêt pour lequel on a accumulé de nombreuses données et connaissances expérimentales dont la synthèse est résumée sous forme de diagramme par les biologistes (cf les exemples de la 3<sup>ème</sup> diapositive)

Pourquoi établir des modèles mathématiques ?

Approches complémentaires des approches expérimentales permettant d'obtenir des indications importantes sur le processus qui peuvent être très difficiles d'obtenir expérimentalement.

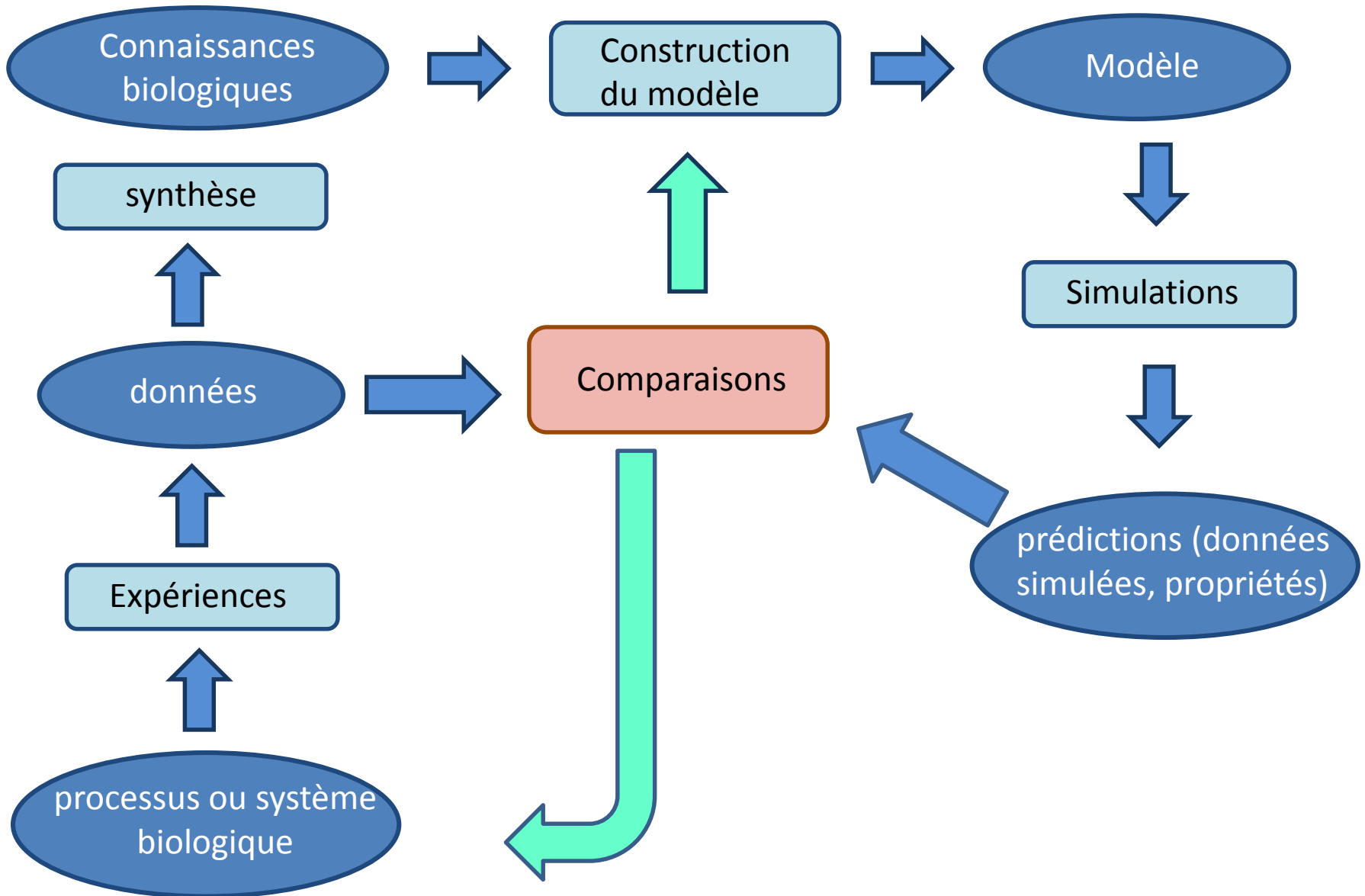
Les modèles :

- consolident les connaissances acquises au travers de différents types d'expérience en en faisant la synthèse,
- permettent de tester les conséquences des hypothèses sous-jacentes (ou tester de nouvelles hypothèses),
- peuvent mettre en évidence des incohérences ou des incomplétudes des connaissances.

Après validation du modèle, il peut être utilisé pour :

- faire des prédictions, par exemple prédire le comportement dynamique du processus quand le système est perturbé (mutations, changement de l'environnement, perturbation chimique etc.),
- prédire des propriétés du système difficilement observables expérimentalement.

# Modélisation d'un processus biologique : démarche





# Modélisation d'un processus biologique : démarche

Qu'est ce qu'un modèle et comment le construire ?

- une abstraction de la réalité
- ne peut pas expliquer tous les détails du système biologique
- faire des hypothèses et des abstractions pour garder le modèle aussi simple que possible
- établir les limites du modèle (il doit rester de taille raisonnable)
- identifier les espèces moléculaires impliquées et leurs différents états si nécessaire (exemple une protéine dans un état non phosphorylé et dans un état phosphorylé)
- identifier les réactions et/ou les changements qui se produisent dans le processus
- définir les relations entre les composés et les actions
- définir la stœchiométrie des réactions
- définir les conditions initiales

Il est conseillé de commencer avec un modèle simple et de le complexifier par la suite si cela est nécessaire en ajoutant des informations supplémentaires.

# Modélisation d'un processus biologique : formalisme

Choisir son type de modélisation, *i. e.*, son formalisme mathématique : dépend de la nature des données dont on dispose sur le processus, données quantitatives ou pas : données sur la valeur des paramètres cinétiques, sur les constantes d'association/dissociation, données d'expression (par exemple obtenues par fusion transcriptionnelle d'un gène rapporteur (GFP ou luciférase) permettant de suivre l'expression des gènes d'intérêt au cours du temps)

Deux classes majeures de modèles dynamiques :

- les modèles quantitatifs :
  - ✓ représentation détaillées du modèle
  - ✓ requiert des valeurs précises des paramètres cinétiques (très rarement le cas)
  - ✓ produit des résultats quantitatifs (par exemple quantité de protéines produites par le modèle)
  - ✓ Essentiellement basé sur des équations différentielles ordinaires
  
- les modèles qualitatifs :
  - ✓ requiert seulement une représentation abstraite des seuils de concentration des protéines (par exemple quand la protéine X est présente à une concentration supérieure à  $\theta_1$  elle active la transcription du gène  $a$  et quand X est présente à une concentration supérieure à  $\theta_2$  elle réprime la transcription du gène  $b$ . Avec  $\theta_1 < \theta_2$ )
  - ✓ défini au travers de formalisme discret ou par des équations différentielles linéaires par morceaux.

# Modélisation d'un processus biologique : formalisme

Différentes approches de modélisation parmi lesquelles :

modèle	Qualitatif	Quantitatif	Déterministe	Stochastique
Equations différentielles ordinaires		X	X	
Théorie des graphes	X		X	
Réseaux bayésiens		X		X
Equations différentielles linéaires par morceaux	X		X	
Modèles booléens/logiques	X		X	
Réseaux de Petri	X	X	X	X

Ces techniques de modélisation sont accompagnées de techniques de simulation pour réaliser des prédictions sur le comportement dynamique du processus biologique étudié.