

Biologie des systèmes

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

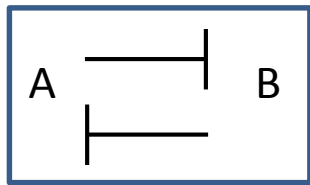
Dans les années 1950, naissance de la biologie moléculaire. Depuis cette discipline a permis d'acquérir de nombreuses données et connaissances sur les processus biomoléculaires, donc à des échelles très petites. Cette approche s'intéressant à des molécules individuelles et/ou à des interactions moléculaires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

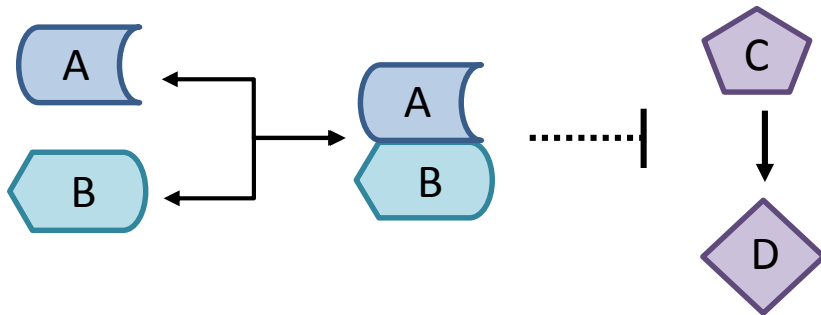
Utilisation de modèle conceptuel utilisant une description verbale du système accompagnée par des diagrammes illustrant comment un ensemble de composés interagissent.

Un exemple simple entre deux composés moléculaires :



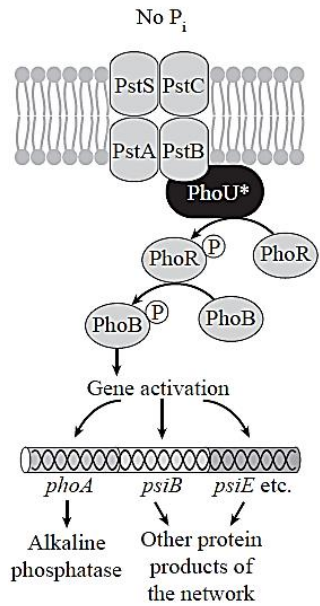
Le produit protéique A du gène *a* va inhiber l'expression du gène *b* et inversement le produit protéique B du gène *b* va inhiber l'expression du gène *a*.

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.

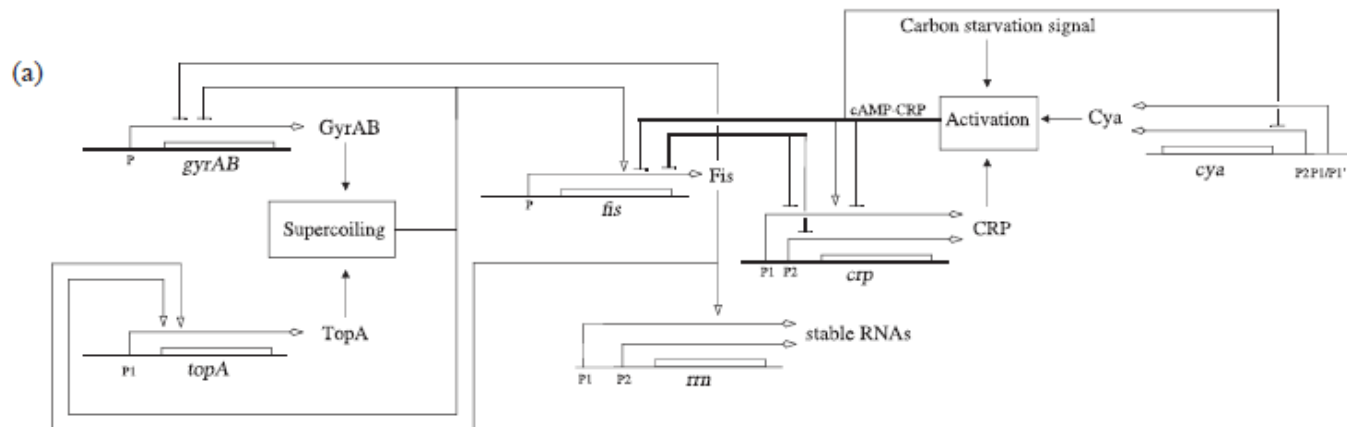
Description des interactions moléculaires par les biologistes et construction de modèle



Représentation d'une cascade de régulation :

La régulation du phosphate chez les entérobactéries

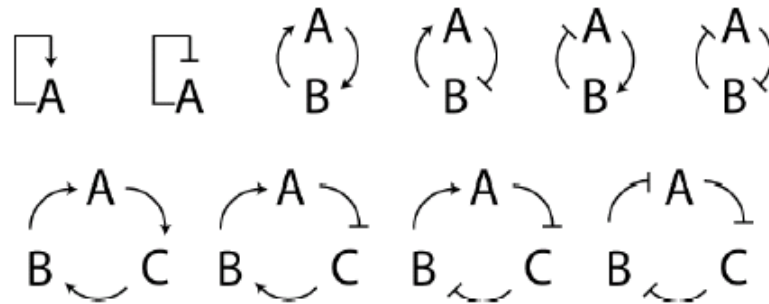
Réseau des gènes, protéines clefs et des interactions de régulation impliqués dans la réponse au stress nutritionnel chez *E.coli*.



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

Pourquoi a-t-on besoin de la modélisation ?

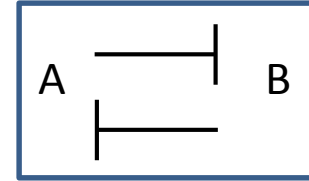
Parmi les réseaux simples ci-dessous impliquant trois protéines, pouvez-vous identifier ceux qui vont conduire à l'oscillation de la concentrations des protéines (A, B et C) ? Pas si facile !



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 :

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

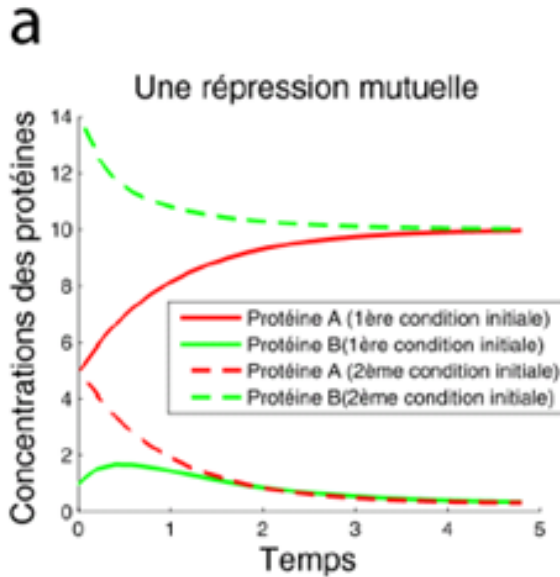
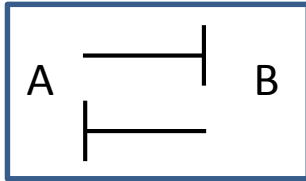
K_d : concentration de B (A) nécessaire pour atteindre la moitié de la répression maximale de A (B) = coefficient de répression.

n : coefficient de Hill

γ : 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) :

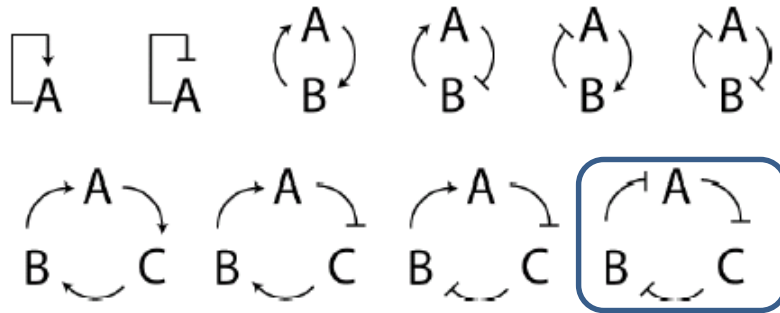


Concentrations initiales des protéines A et B et devenir de l'état du système :

- 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

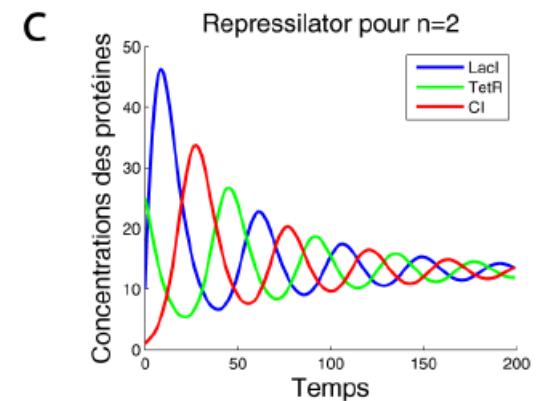
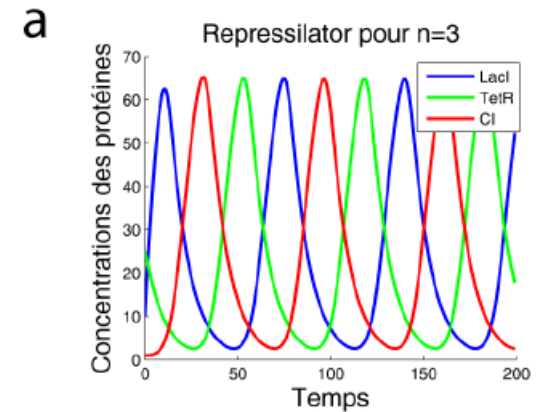
Revenons à notre question : Parmi les réseaux simples ci-dessous impliquant trois protéines, pouvez-vous identifier ceux qui vont conduire à l'oscillation de la concentrations des protéines (A, B et C) ? Pas si facile !



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.

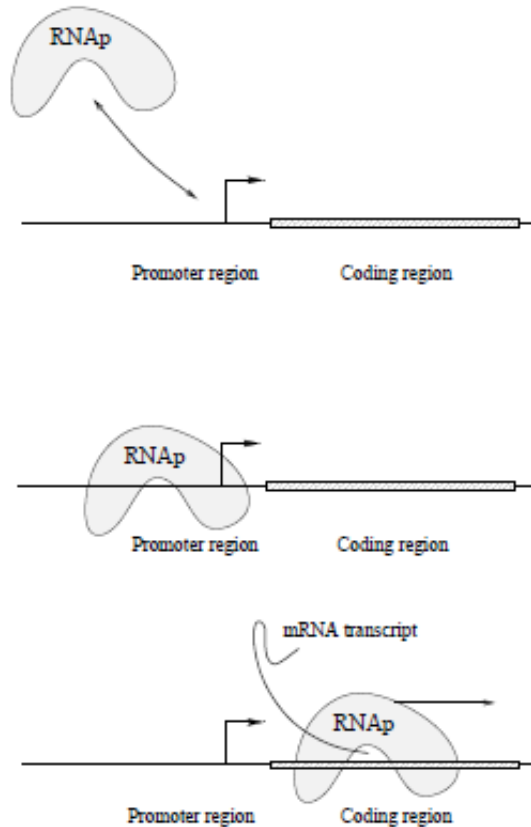
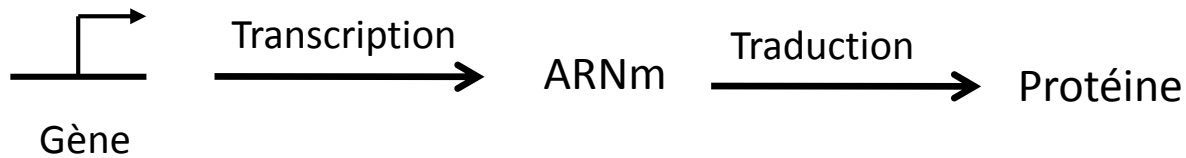
La cellule rencontre différentes situations qui requièrent différentes protéines. Par exemple : i) si dans le milieu extérieur une cellule bactérienne « sent » la présence de galactose, elle va commencer à produire les protéines nécessaires à son import dans la cellule et à son catabolisme ou i) si sa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

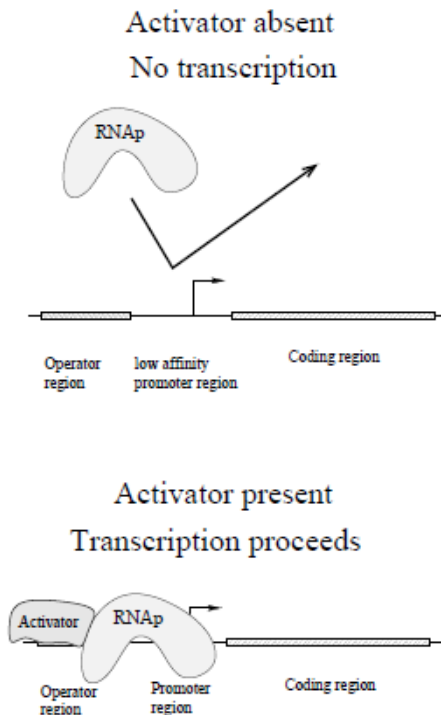
La qualité de cette liaison va spécifier le taux de transcription des gènes cibles

Les éléments des réseaux de transcription

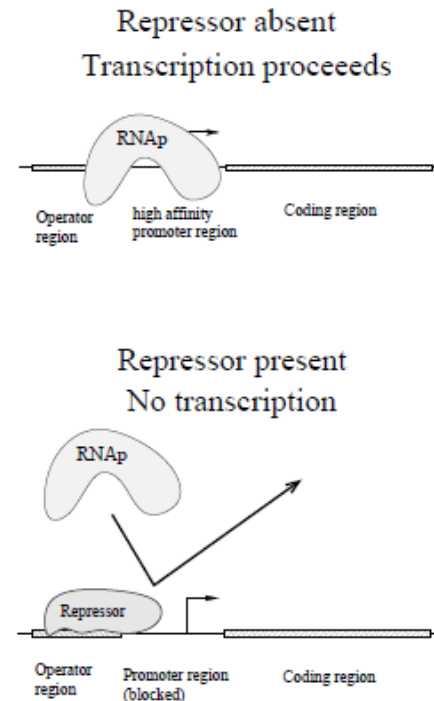
Le taux de transcription d'un gène peut être régulé (modulé) par des protéines appelées **facteurs de transcription**. Ces facteurs affectent le taux de transcription en se fixant sur des sites spécifiques du promoteur ou sur des régions au voisinage du promoteur du moins chez les bactéries (plus compliqué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

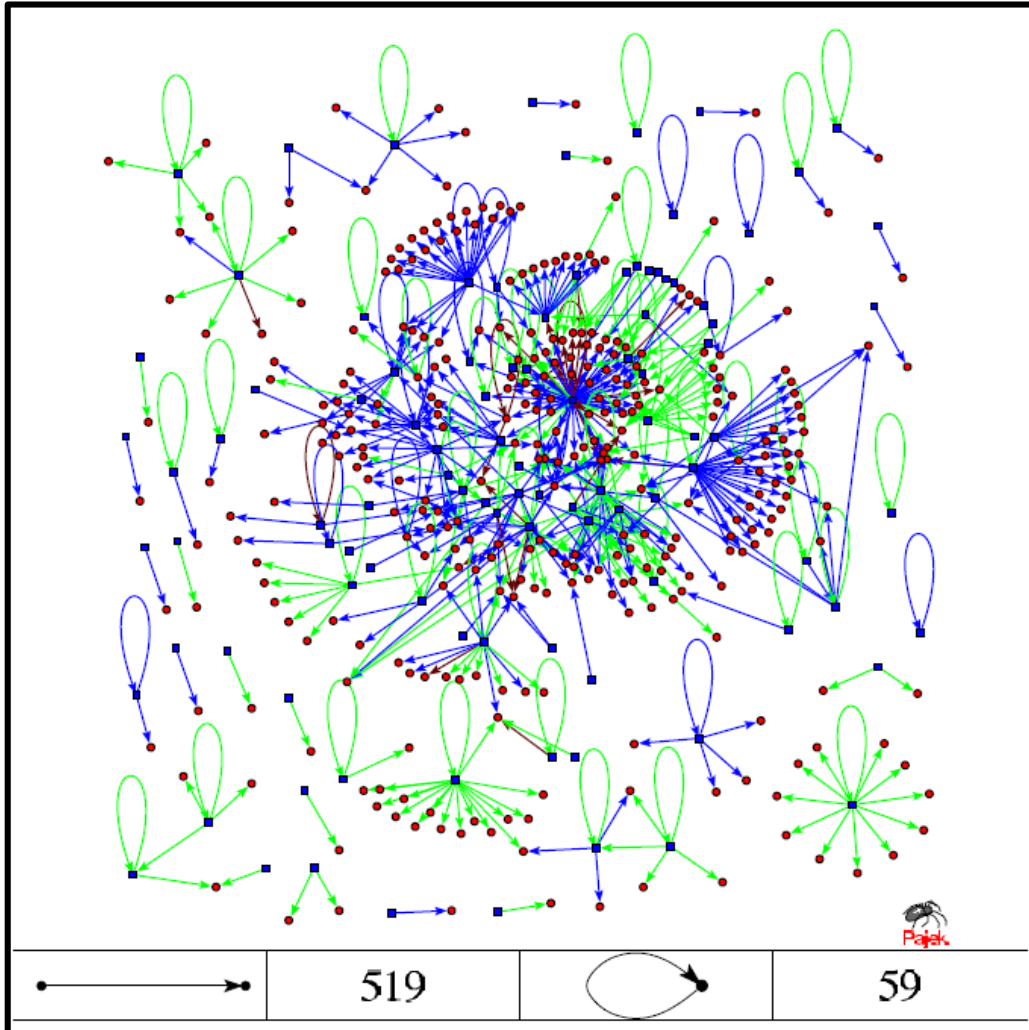
Ces facteurs de transcription sont des protéines codées par des gènes qui peuvent eux-mêmes être régulés par d'autres facteurs de transcriptions qui peuvent eux aussi être régulés par d'autres facteurs de transcriptions etc.

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 colorées en fonction de l'action du TF :

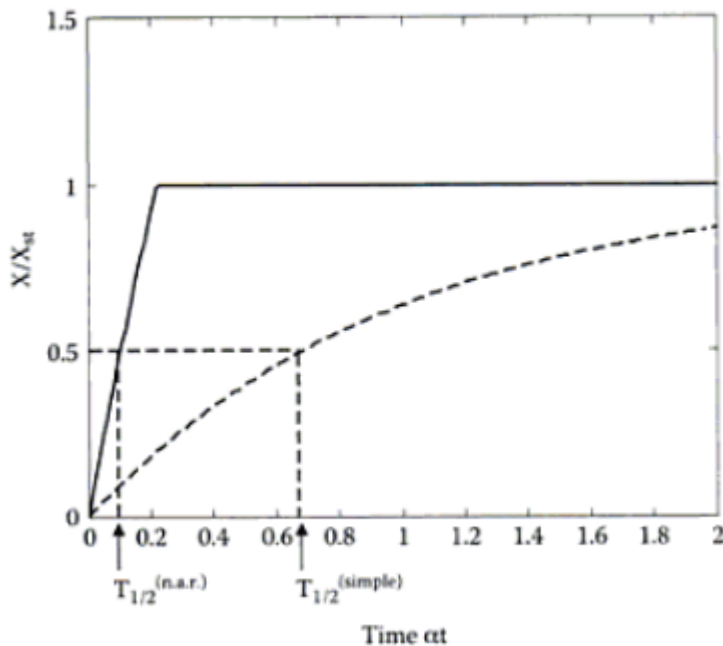
- bleu le TF est un activateur
- vert le TF est un répresseur
- marron le TF a les deux fonctions

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'auto-régulation 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

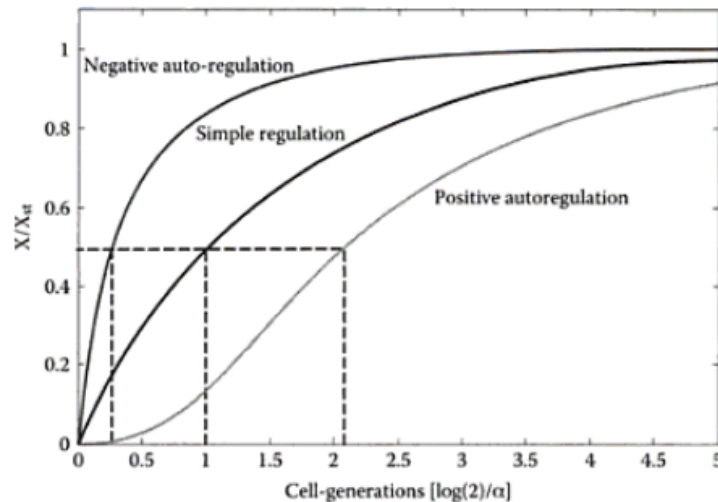


FIGURE 3.5 Dynamics of a negatively autoregulated gene, a simply regulated gene and a positively autoregulated gene. The negatively and positively autoregulated genes have a Hill input function with Hill coefficient $n = 1$. Shown is protein concentration normalized by its steady-state value, X/X_{st}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

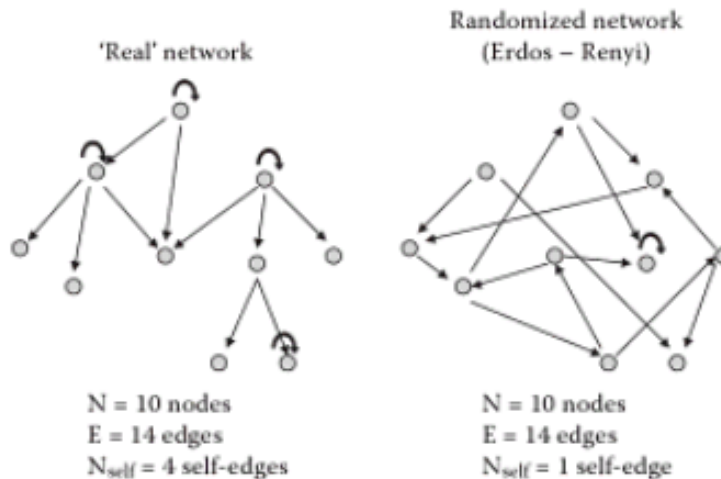
L'autorégulation positive ralentit le temps de réponse par rapport à une simple régulation.

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

Si la valeur de β_{max} (taux maximal de synthèse de X) est plus élevé que la valeur de γ (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.

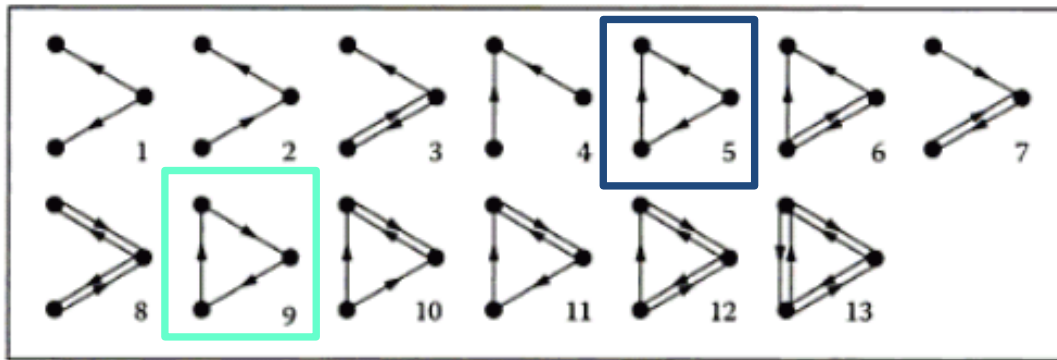


Un motif dans le réseau = un pattern que l'on retrouve beaucoup plus souvent dans le réseau réel que dans le réseau aléatoire

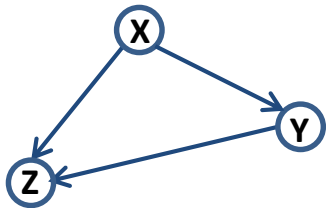
Idée de base : les motifs qui apparaissent dans le réseau réel plus souvent que dans le réseau aléatoire ont du être préservés au cours de l'évolution contre des mutations qui changent les liens de façon aléatoire (par exemple, une mutation dans la séquence d'un promoteur peut abolir la liaison du facteur 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

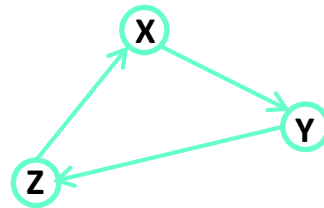
On s'intéresse aux motifs composés de trois nœuds : Il en existe 13 possibles



Seulement le 5^{ème} est un motif dans le réseau (parfois le 9^{ème}).



Le 5^{ème} est appelé une feed-forward loop (FFL)



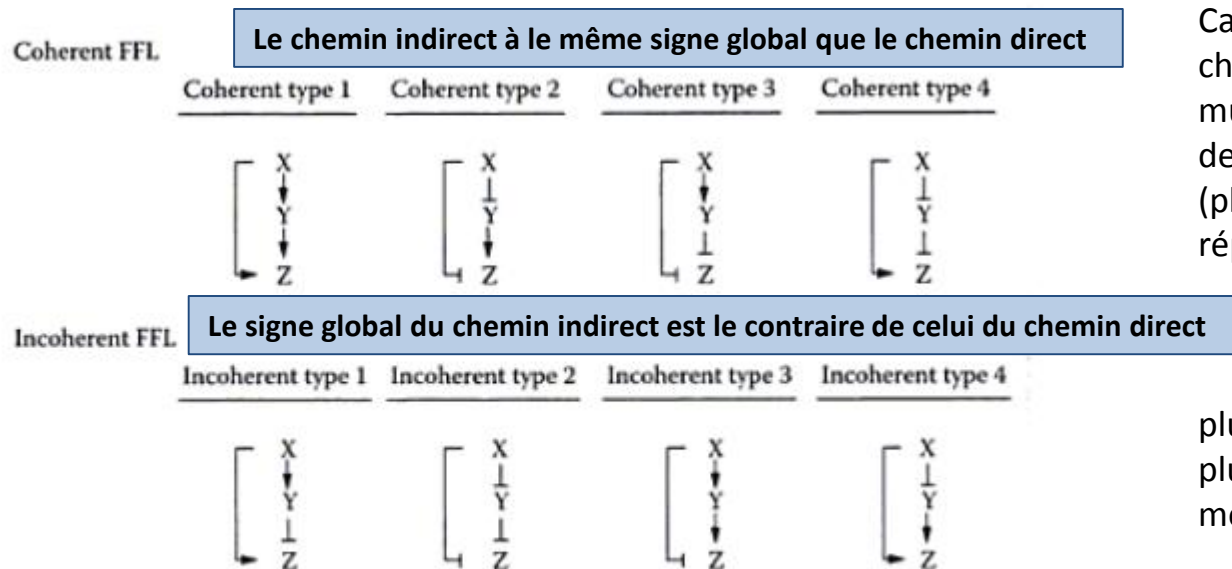
Le 9^{ème} est appelé une feedback loop à 3 nœuds (boucle de rétroaction)

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

Feed-forward loop

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

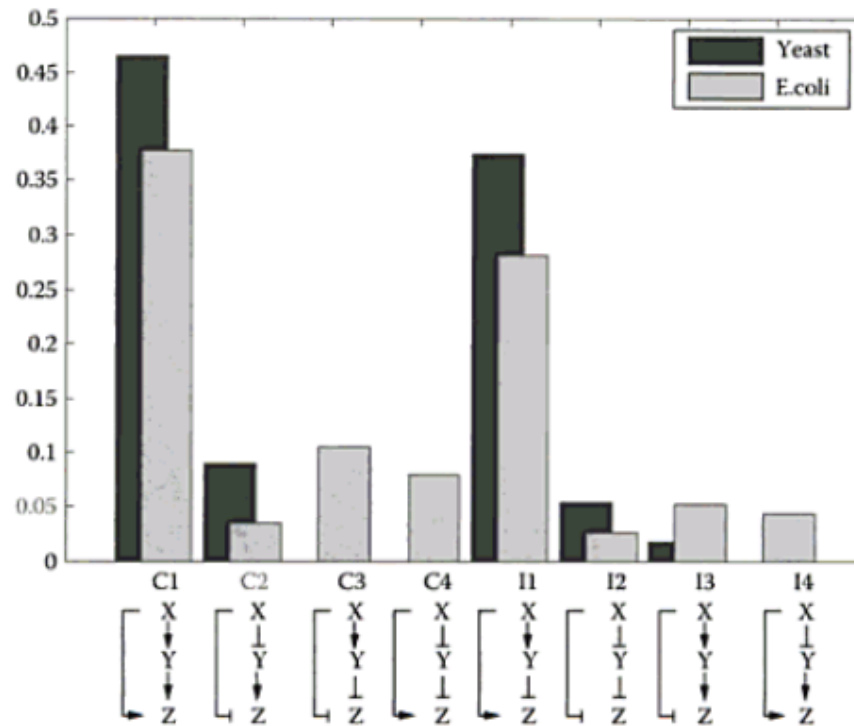


FIGURE 4.4 Relative abundance of the eight FFL types in the transcription networks of yeast and *E. coli*. FFL types are marked C and I for coherent and incoherent. The *E. coli* network is based on the Ecocyc and RegulonDB databases and has about twice as many edges as in the network of Figure 2.3.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

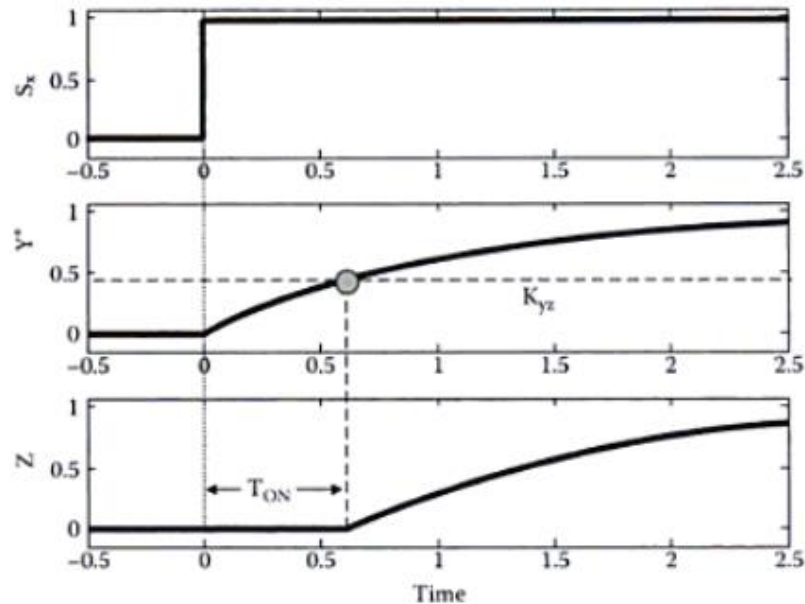
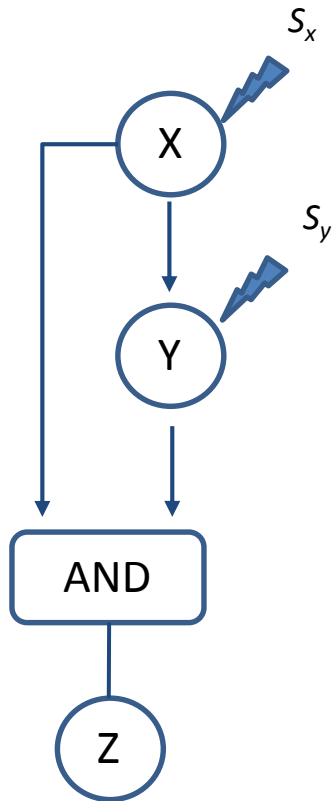


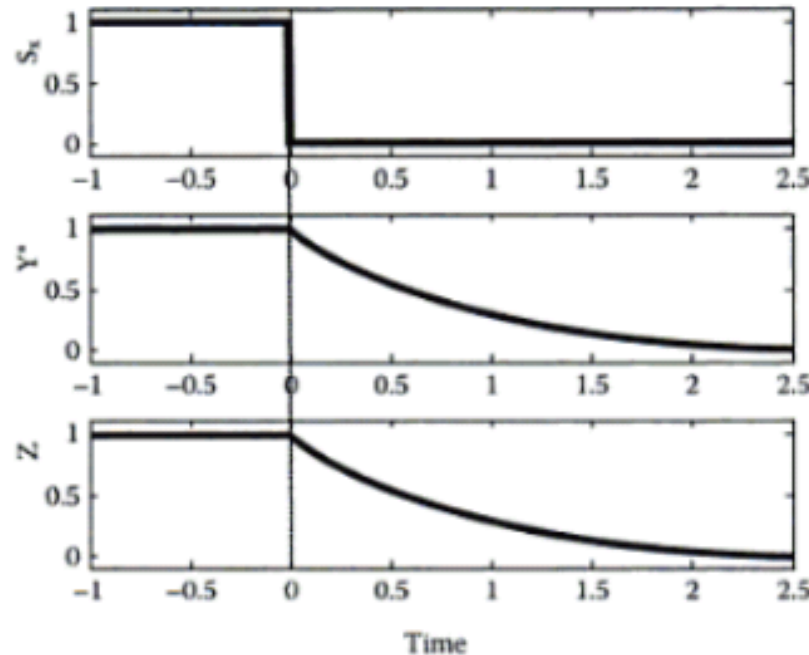
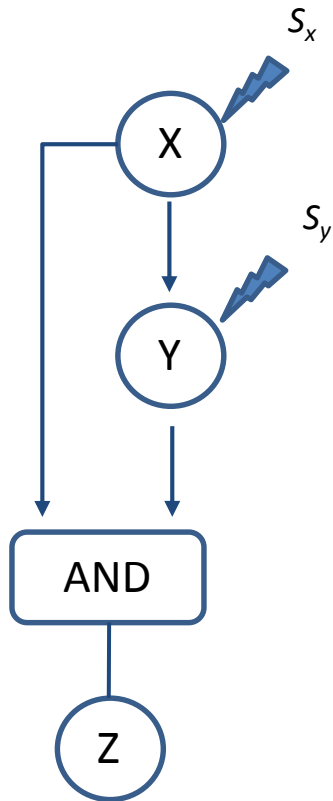
FIGURE 4.7 Dynamics of the coherent type-1 FFL with AND logic following an ON step of S_x at time $t = 0$ in the presence of S_y . The activation threshold of Z by Y is K_{yz} (horizontal dashed line). The production and degradation rates are $\alpha_y = \alpha_z = 1$, $\beta_y = \beta_z = 1$. The delay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



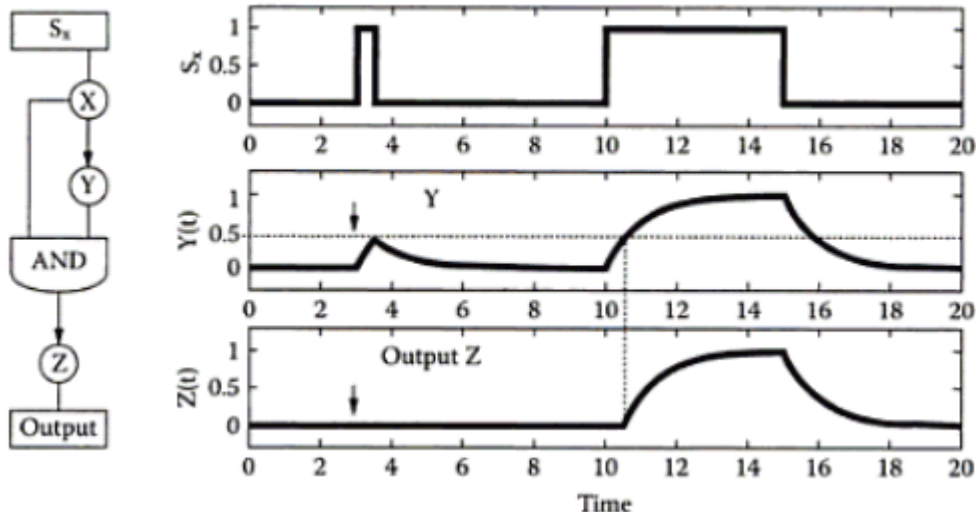
Extracted from : *An introduction to systems Biology* (2007) U. Alon Ed, Chapman & Hall/CRC Mathematical and Computational Biology Series

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

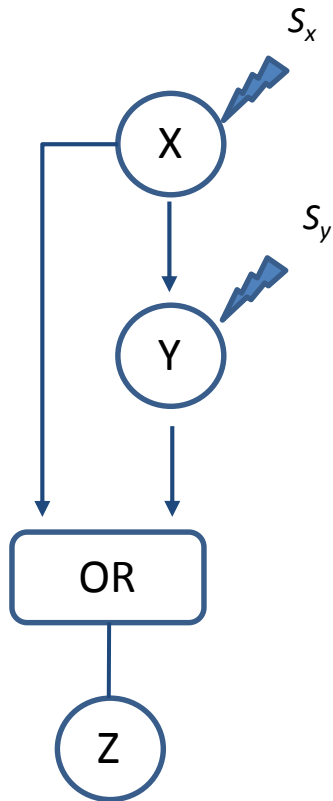
- Si la persistance du signal S_x est plus courte que le temps nécessaire pour que la concentration de la protéine Y atteigne la valeur K_{yz} , la protéine Z n'est pas produite.
- La protéine Z sera produite uniquement si le signal S_x persiste sur une durée qui excède le délai pour que la concentration 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*.

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^{ème} 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.

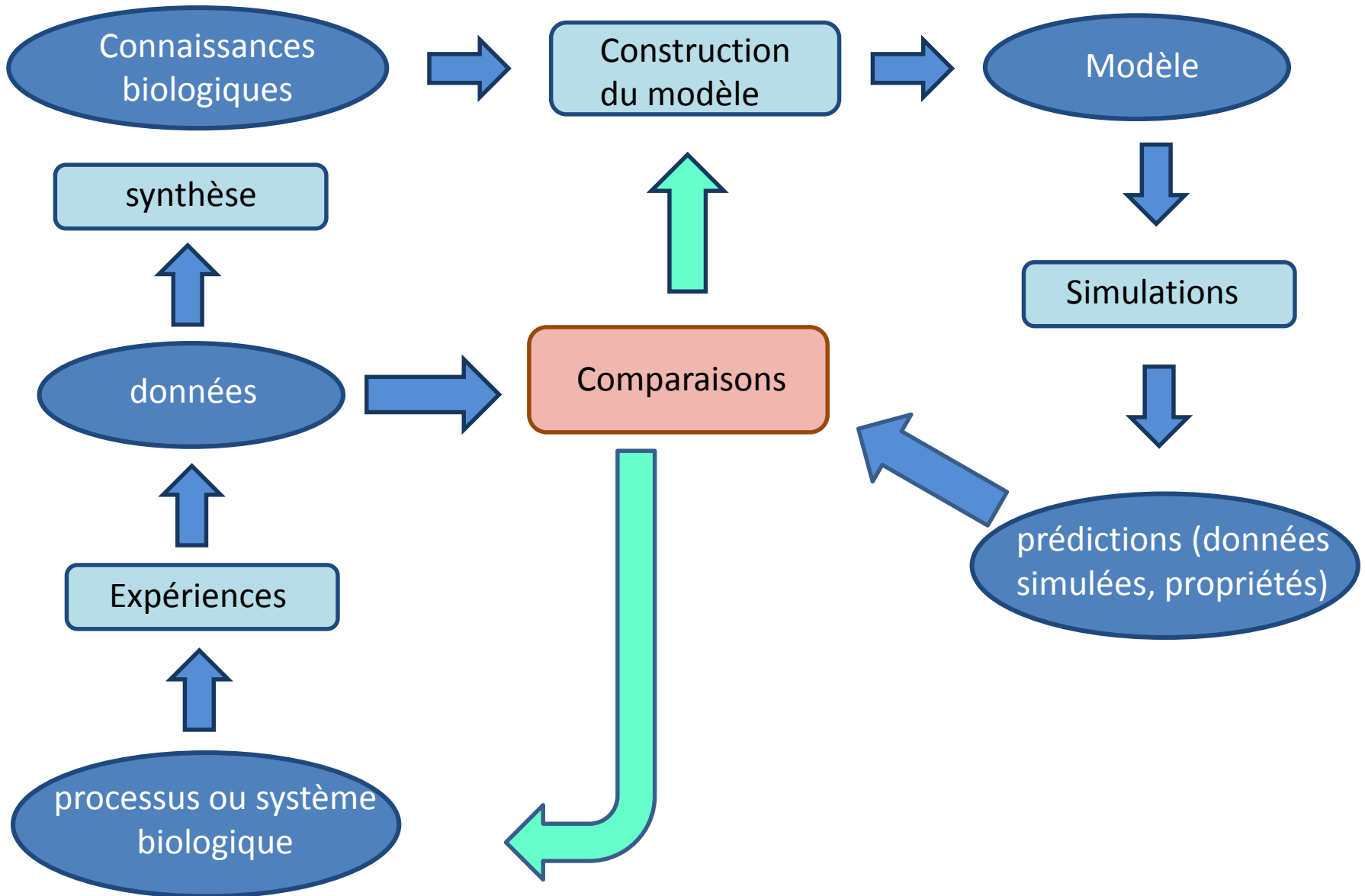
Les modèles :

- 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

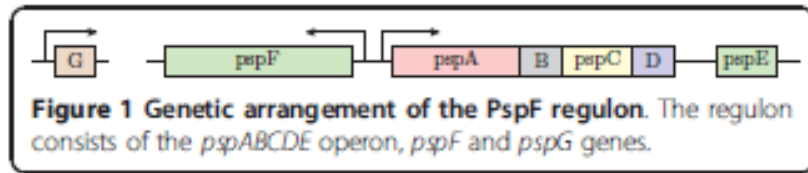
Il est conseillé de commencer avec un modèle simple et de le complexifier par la suite si cela est nécessaire en ajoutant des informations supplémentaires.

Modélisation d'un processus biologique : exemple

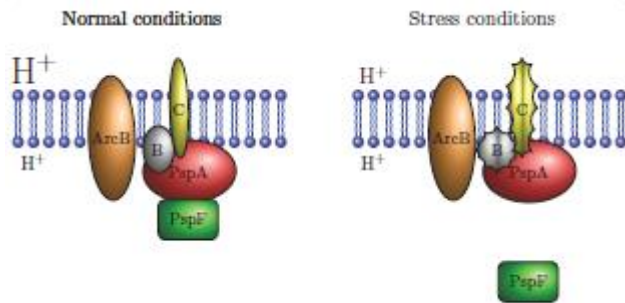
Analysis of the phage shock protein stress response in *Escherichia Coli* : The Psp response that responds to alterations in the bacterial cell envelope (Toni *et al.*, BMC Systems Biology, 5: 69)

Connaissances biologiques :

- les gènes *psp* chez *E. coli* forment le régulon PspF qui inclut l'opéron *psp* (les gènes *pspA*, *pspB*, *pspC*, *pspD* and *pspE*), et les gènes *pspF* et *pspG*.



PspF est un facteur de transcription qui active la transcription des opérons *pspA-E* et *pspG*. Le gène *pspF* est transcrit via le facteur σ de ménage.



En l'absence de stress PspA se lie à PspF ce qui inhibe l'activité ATPase de PspF. La transcription des opérons *pspA-E* et *pspG* est basale.

En présence de stress (membrane endommagée), un stimulus est converti en un signal qui est transduit au travers de PspB et PspC. Le signal déstabilise la liaison PspA-PspF et permet à PspF, sous forme d'hexamère, d'activer la transcription conduisant à une augmentation de la concentration de plusieurs protéines Psp.

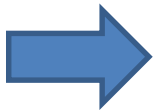
Modélisation d'un processus biologique : exemple

Rôles connus des protéines Psp :

- PspA, PspD and PspG jouent un rôle majeur dans le switch du métabolisme de la cellule de la respiration anaérobie à la fermentation
- PspA and PspD sont impliqués dans la réparation de la membrane endommagée, PspA sous la forme d'un oligomère de 36 protéines
- PspA, PspD and PspG régule à la baisse la motilité de la cellule qui à son tour régule à la baisse la consommation de la force proton motrice et maintient l'usage de l'énergie.

Questions sur le comportement du système, comme par exemple :

- comment la réponse évolue au cours du temps ?
- à quelle vitesse les cellules répondent au stress quand il est induit ?
- à quelle vitesse la membrane est réparée ?
- comment la cellule répond à la suppression du stress ?



Modélisation du réseau d'interactions dans un formalisme mathématique pour analyser le comportement du système et interpréter les résultats en terme d'implications biologiques.

Modélisation d'un processus biologique : exemple

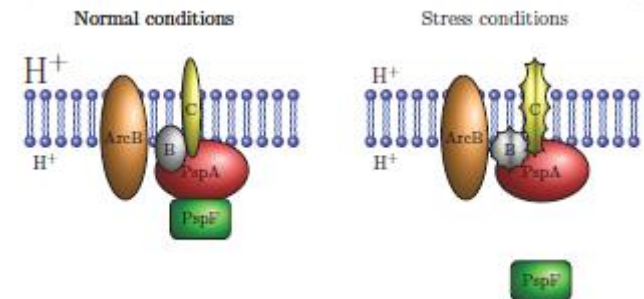
Construction du modèle : hypothèses et choix (quels sont les éléments biologiques importants qu'il faut retenir pour rendre compte de la dynamique de la réponse au stress) → construction d'un modèle simplifié.

- PspD, PspE and PspG : impliquées dans la réponse physiologique mais pas impliquées directement dans la régulation de la réponse au stress → pas prises en compte dans le réseau
- Seules PspA, PspB, PspC et PspF sont retenues. De plus, PspB et PspC seront représentées directement sous forme d'un complexe BC dont la formation ne sera pas modélisée.
- Membrane : Elle peut être intacte ou endommagée quand le stress agit sur la membrane. Pour pouvoir discrétiser la mesure de l'endommagement de la membrane, la membrane sera modélisée comme étant formée d'une partie intacte et d'une partie endommagée. La partie endommagée sera exprimée en pourcentage (maximum étant 100).

Modélisation d'un processus biologique : exemple

Composants du système qui doivent être pris en compte :

- stress
- la membrane endommagée
- la membrane intacte
- PspA
- PspB et PspC modélisés sous forme de complexe
- le complexe BCA
- le complexe BCAF
- le complexe BCA ayant subi un changement de conformation du au stress ($B_cC_cA_c$)
- PspF
- l'hexamère de PspF agissant comme facteur de transcription
- l'oligomère de PspA (36 protéines) impliqué dans la réparation de la membrane



Modélisation d'un processus biologique : exemple

Listes des réactions entre les composants du système qui doivent être modélisée :

- stress + membrane intacte \rightarrow stress + membrane endommagée
- membrane endommagée + oligomère de PspA \rightarrow membrane intacte + oligomère de PspA
- membrane endommagée + complexe BCAAF \rightarrow membrane endommagée + PspF + complexe $B_cC_cA_c$
- 6 PspF \rightarrow facteur de transcription
- facteur de transcription \rightarrow 6 PspF
- facteur de transcription \rightarrow PspA (100) + complexe BC (60 or 40) + facteur de transcription
- 36 PspA \rightarrow oligomère
- PspA + complexe BC \rightarrow complexe BCA
- complexe BCA + PspF \rightarrow complexe BCAAF
- BCA + membrane endommagée \rightarrow membrane endommagée + complexe $B_cC_cA_c$
- membrane intacte + complexe $B_cC_cA_c$ \rightarrow membrane intacte + complexe BCA
- dégradation de BCA
- dégradation de $B_cC_cA_c$
- dégradation de l'oligomère
- dégradation du complexe BC
- dégradation de PspA

Le nombre de protéines a été déduit de mesures expérimentales de la production des ARNm de PspA, PspB et PspC (100:60:40) . Comme on modélise directement le complexe BC, le nombre de complexes peut être 60 ou 40. Une partie des protéines PspA produites est impliquée dans la formation du complexe BCA et l'autre partie forme l'oligomère de 36 protéines.

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 θ_1 elle active la transcription du gène a et quand X est présente à une concentration supérieure à θ_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.

Les modèles basés sur les réseaux de Petri

Les réseaux de Petri (Petri nets) :

Nommé d'après Carl Adam Petri qui, au début des années 1960, a proposé un formalisme graphique et mathématique permettant la modélisation et l'analyse de systèmes distribués asynchrones concurrents.

Fortement utilisé pour la modélisation des systèmes biologiques.

Sous sa forme la plus simple : un graphe dirigé bipartite

Deux types de sommets :

- places représentent les conditions ou les ressources (ex: l'oligomère de PspA)
- transitions représentent les activités, *i.e.*, les événements qui peuvent changer l'état des ressources (ex: synthèse des protéines PspA, du complexe BC)

Des arcs dirigés connectent les places et les transitions

- les places sont exclusivement connectées à des transitions
- les transitions sont exclusivement connectées à des places

Les jetons (tokens) sont placés dans les places et définissent l'état du réseau de Petri

Un arc peut être pondéré : nombre de jetons qui vont être consommés lors de la transition (ex: 36 protéines PspA pour la formation de l'oligomère)

Les modèles basés sur les réseaux de Petri

- **Les places** : sommet passifs du graphe, représentés par des cercles. Dans un contexte biologiques, les places peuvent représenter : des molécules, des ions, des protéines, des complexes multicellulaires, des cellules, des organismes, des espèces, des populations. Seules les places possèdent des jetons.
- **Les transitions** : sommets actifs, représentés par des carrés, décrivent les activités du réseaux. Dans un contexte biologiques, les transitions peuvent représenter des réactions biochimiques, des interactions moléculaires ou des changements intramoléculaires. Les transitions consomment des jetons présents dans les pré-places et produisent des jetons dans les post-places en fonction de la pondération de l'arc la reliant à ses places.
- **Les jetons** : éléments variables du réseau de Petri, représentés par des points ou des chiffres dans les places, produits et consommés par les transitions. Pour les systèmes biologiques, les jetons peuvent représenter le niveau de concentration ou le nombre d'un composé, *e.g.*, protéines, ions, molécules organiques ou inorganiques. Ils peuvent aussi représenter des valeurs de quantités physiques comme la température, des valeurs de pH. Un réseau de Petri sans jeton est dit « vide ».
- **Les arcs dirigés** : éléments inactifs visualisés par des flèches. Ils spécifient les relations causales entre transitions et places. En biologie, ils relient les réactants/substrats et produits d'une réaction biochimiques. Chaque arc est pondéré et décrit le nombre de composés (jetons) consommés et produit par la transition. Ces poids représentent la stœchiométrie des réactions biochimiques .

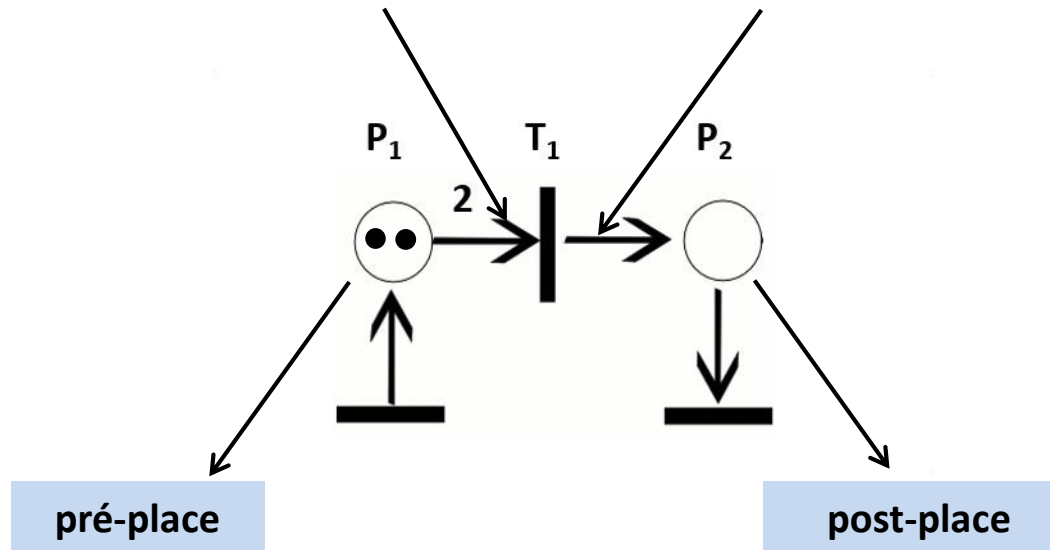
Un exemple de réseaux de Petri très simple

Un réseau de Petri avec deux place P_1 et P_2 et une transition T_1

Le poids de l'arc reliant P_1 à T_1 indique le nombre de jetons consommés par la réaction. Donc pour que la transition puisse avoir lieu (faire feu) il faut qu'au moins deux jetons soient présents dans P_1 . Quand aucun poids ne figure sur un arc, cela veut dire que le poids est égal à 1. Donc la transition T_1 va enlever (consommer) deux jetons de la place P_1 et ajouter (produire) un jeton dans la place P_2 .

Post-transition → consomme des jetons

Pré-transition → produit des jetons



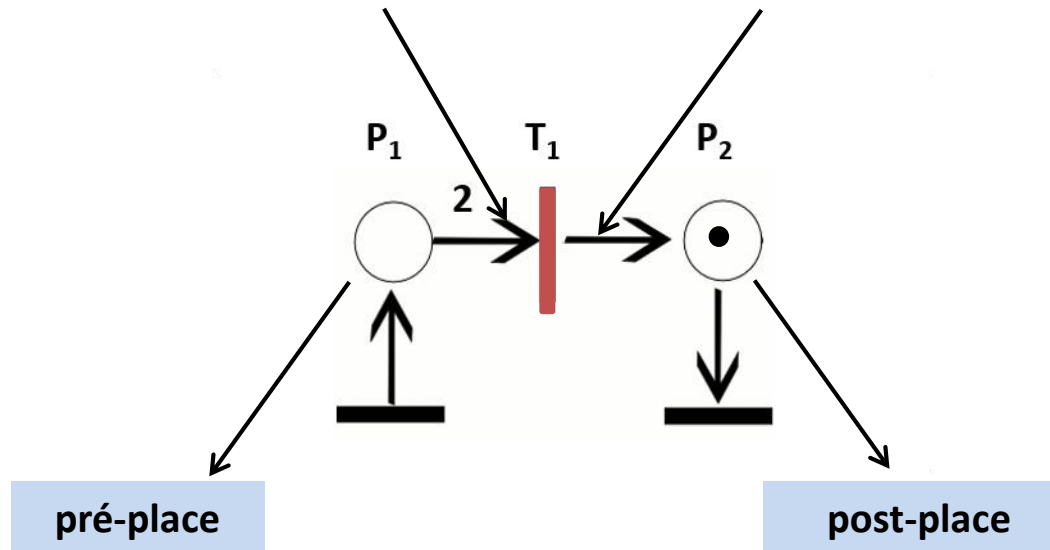
Un exemple de réseaux de Petri très simple

Un réseau de Petri avec deux place P_1 et P_2 et une transition T_1

Le poids de l'arc reliant P_1 à T_1 indique le nombre de jetons consommés par la réaction. Donc pour que la transition puisse avoir lieu (faire feu) il faut qu'au moins deux jetons soient présents dans P_1 . Quand aucun poids ne figure sur un arc, cela veut dire que le poids est égal à 1. Donc la transition T_1 va enlever (consommer) deux jetons de la place P_1 et ajouter (produire) un jeton dans la place P_2 .

Post-transition → consomme des jetons

Pré-transition → produit des jetons

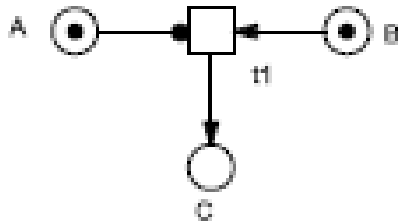


réseaux de Petri étendu

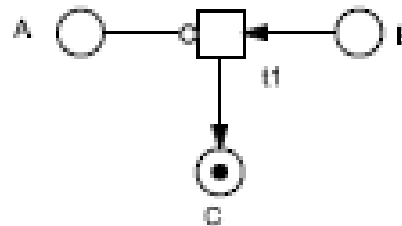
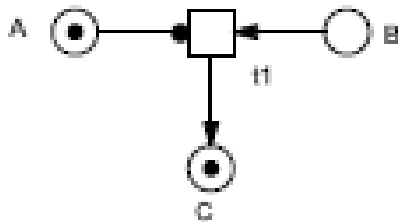
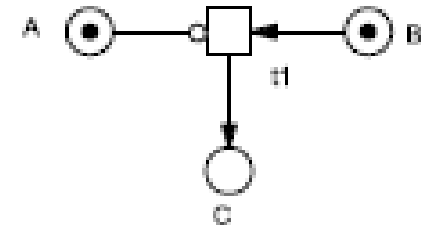
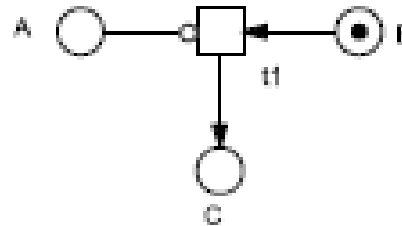
Deux types d'arcs sont ajoutés :

- arc test : active la transition sans consommé de jetons :  ou 
- arc inhibiteur : inhibe la transition : 

A. Read Arc **Arc test**



B. Inhibitor Arc **Arc inhibiteur**



t_1 est possible les places A et B ont suffisamment de jetons. Quand la transition a fait feu, les jetons sont consommés de la place B mais pas ceux de la place A.

t_1 est possible si la place B contient assez de jetons mais que la place A n'en contient pas assez. Quand la transition a fait feu, les jetons sont consommés de la place B mais pas ceux de la place A.

réseaux de Petri et métabolisme

Les réseaux de Petri standard permettent la représentation des composants essentiels des voies métaboliques et ils peuvent être utilisés pour réaliser des analyses qualitatives

Voie métabolique = interconnexion de réseaux de réactions enzymatiques (produit d'une réaction est le substrat de l'enzyme qui catalyse la réaction suivante).

Modélisation de 5 types de réactions :

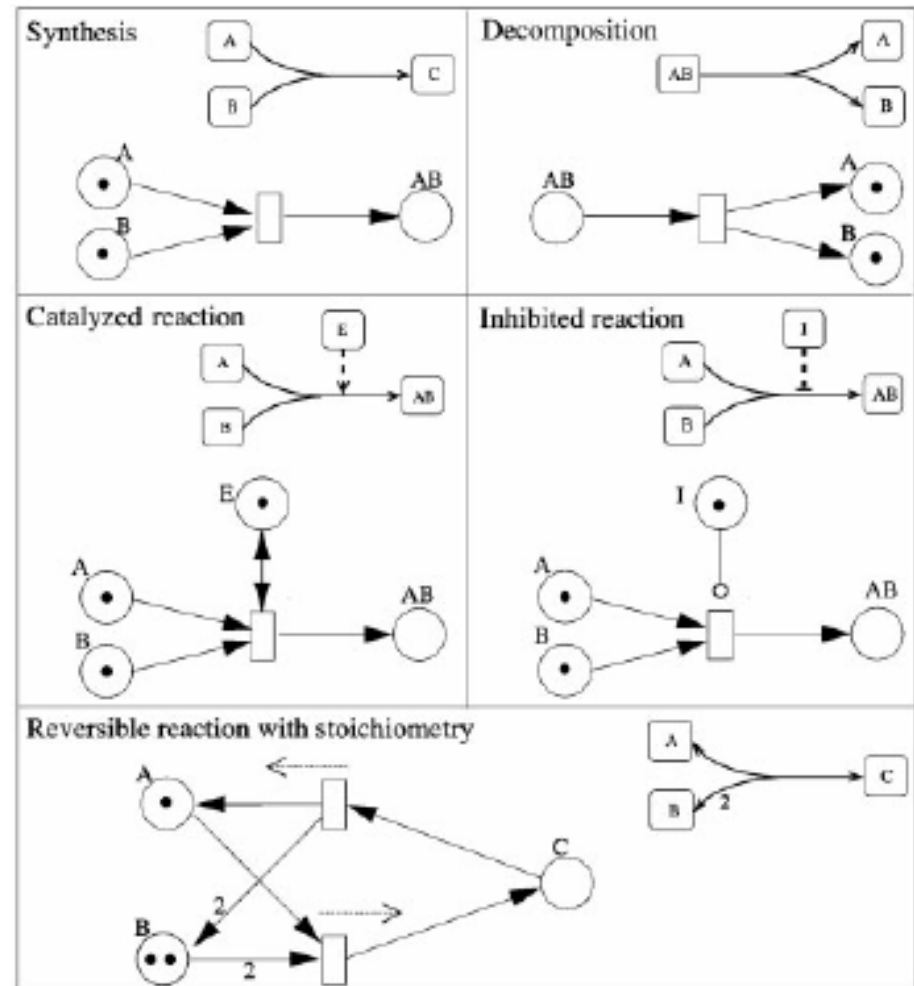
- Places = substrats, produits ou enzyme
- Transitions = réactions
- Poids des arcs = coefficients stœchiométriques des réactions

Réaction de catalyse :

L'enzyme doit être présente et n'est pas consommée. La place est reliée à la transition par un arc test.

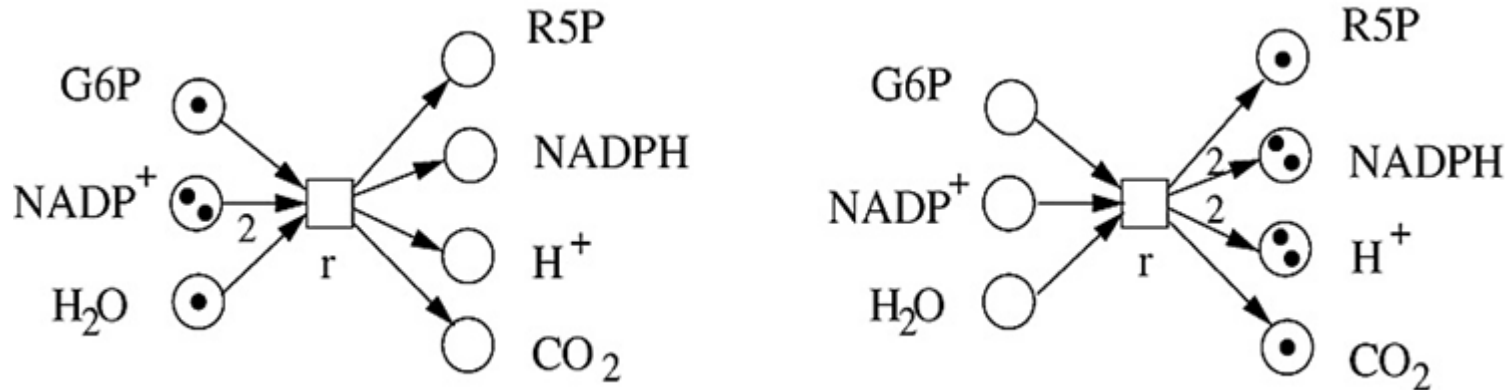
Réaction d'inhibition :

Quand l'enzyme est présente, la réaction (transition) ne peut pas avoir lieu. La place est reliée à la transition par un arc inhibiteur.



réseaux de Petri et métabolisme

Exemple de la modélisation de la réaction :



Grunwald *et al.*, 2008

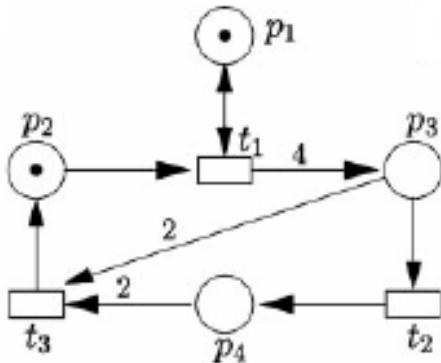
Pour que la transition soit possible et puisse faire feu, il faut que les pré-places soient suffisamment marquées, *i. e.*, contenir un nombre de jetons \geq au poids des arcs les reliant à la transitions.

Quand la transition a fait feu: les jetons des pré-places sont consommés (en fonction du poids des arcs entrant de la transition) et de nouveaux jetons sont produits dans les post-places. Le nombre de jetons produits est déterminé par le poids des arcs sortant de la transition.

Le « jeu des jetons » ou « token game » représente l'évolution dynamique du système

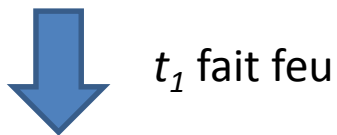
réseaux de Petri : exemple d'un « token game »

Marquage initial M_0



p_1 et t_1 sont connectées par un arc test. Le marquage de p_1 va gouverner le fait que la transition puisse faire feu ou pas. Ce marquage n'est pas modifié une fois que la transition a fait feu. Ici t_1 peut faire feu car p_2 et p_1 contiennent un nombre suffisant de jetons.

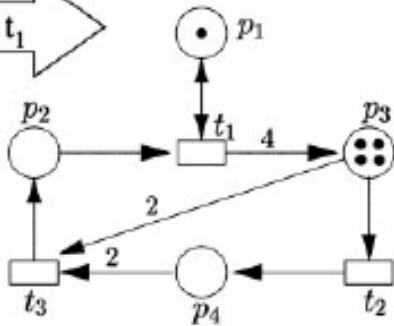
Example from Chaouiya, 2007



t_1 fait feu



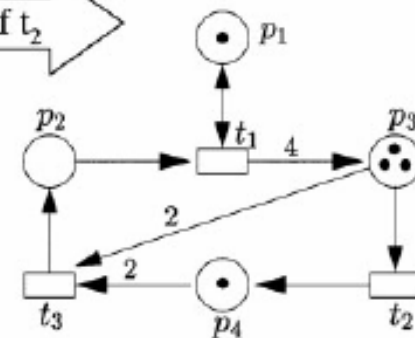
Nouveau marquage M1



t_2 fait feu



Nouveau marquage M2



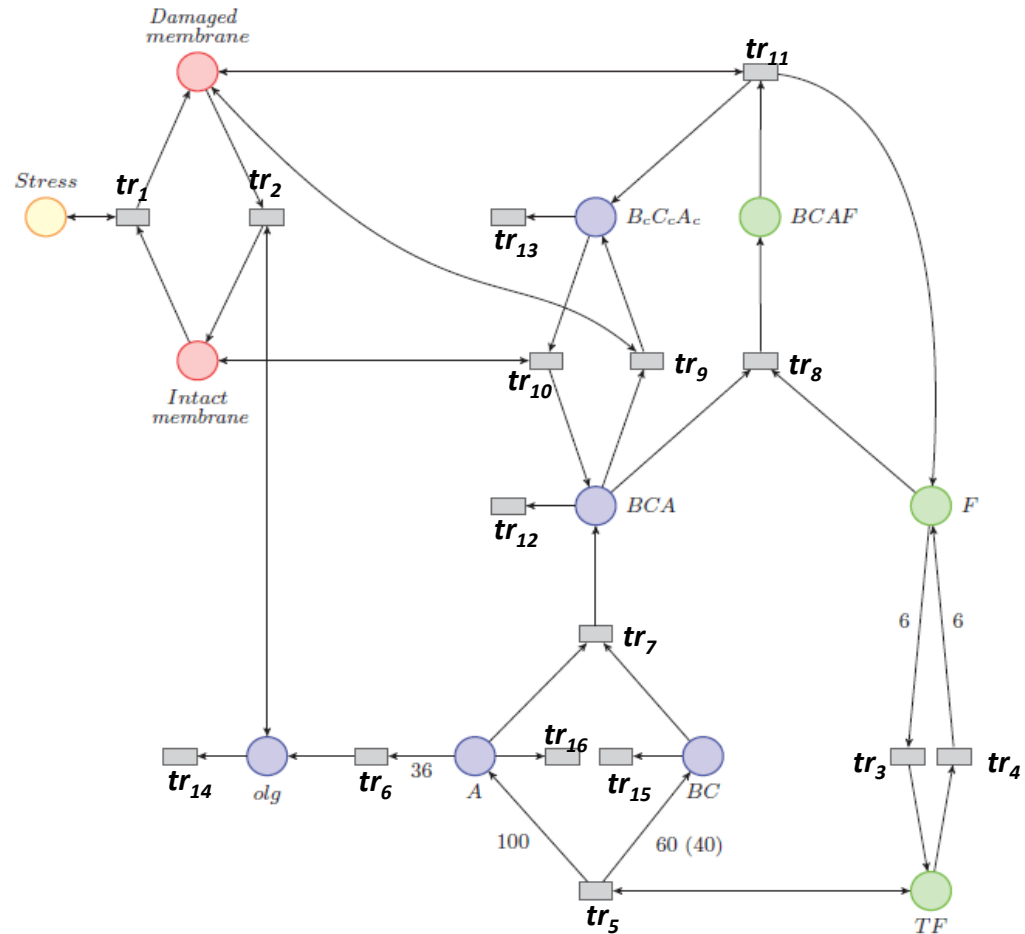
Le jeton de p_2 est consommé. Quatre jetons sont produits dans p_3 . Le nouveau marquage M_1 permet à t_2 de faire feu.

Un jeton de p_3 est consommé et un jeton est produit dans p_4 . Le nouveau marquage M_2 ne permet pas à la transition t_3 de faire feu. Le système est dans un état mort car aucune transition ne peut être activée.

réseaux de Petri correspondant à la modélisation de la réponse PsP

Des réactions aux transitions:

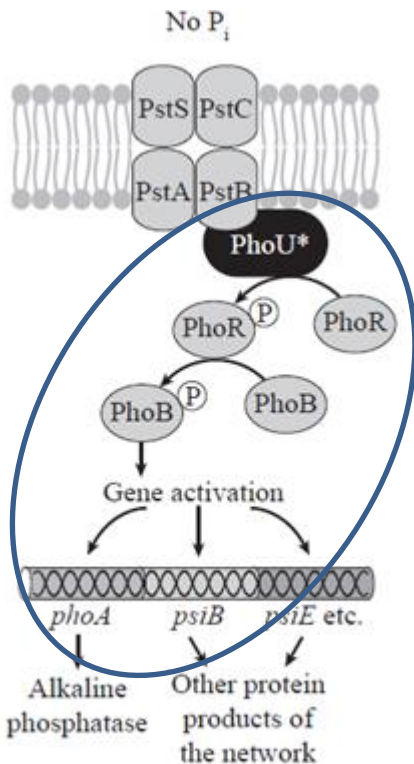
- stress + membrane intacte \rightarrow stress + membrane endommagée (tr_1)
- membrane endommagée + oligomère de PspA \rightarrow membrane intacte + oligomère de PspA (tr_2)
- membrane endommagée + complexe BCAF \rightarrow membrane endommagée + PspF + complexe $B_cC_cA_c$ (tr_{11})
- 6 PspF \rightarrow facteur de transcription (tr_3)
- facteur de transcription \rightarrow 6 PspF (tr_4)
- facteur de transcription \rightarrow PspA (100) + complexe BC (60 or 40) + facteur de transcription (tr_5)
- 36 PspA \rightarrow oligomère (tr_6)
- PspA + complexe BC \rightarrow complexe BCA (tr_7)
- complexe BCA + PspF \rightarrow complexe BCAF (tr_8)
- BCA + membrane endommagée \rightarrow membrane endommagée + complexe $B_cC_cA_c$ (tr_9)
- membrane intacte + complexe $B_cC_cA_c$ \rightarrow membrane intacte + complexe BCA (tr_{10})
- dégradation de BCA (tr_{12})
- dégradation de $B_cC_cA_c$ (tr_{13})
- dégradation de l'oligomère (tr_{14})
- dégradation du complexe BC (tr_{15})
- dégradation de PspA (tr_{16})



(adapté de Toni *et al.*, BMC Systems Biology, 5: 69)

réseaux de Petri et réseaux de régulation génétique

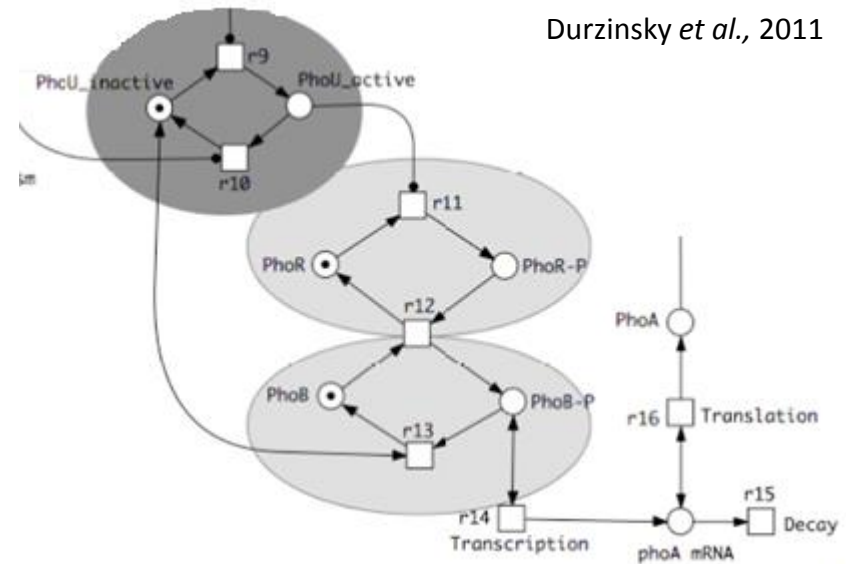
Régulation du phosphate dans les entérobactéries : le gène *phoA* est active par une cascade de régulation



Neidhardt *et al.* 1990

Modélisation par réseau de Petri

Durzinsky *et al.*, 2011

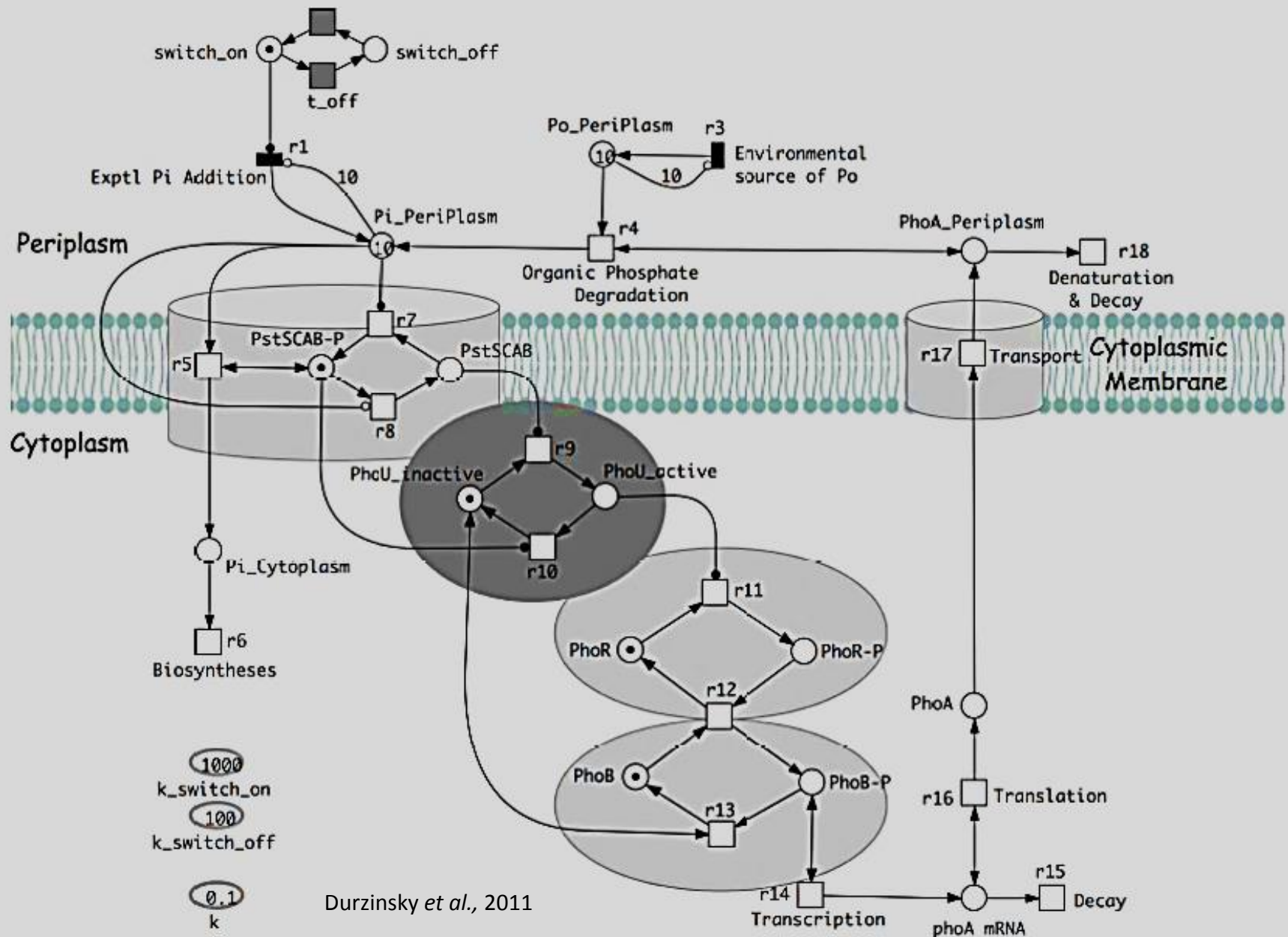


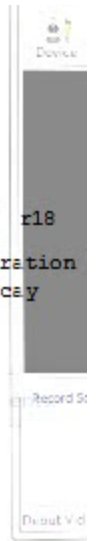
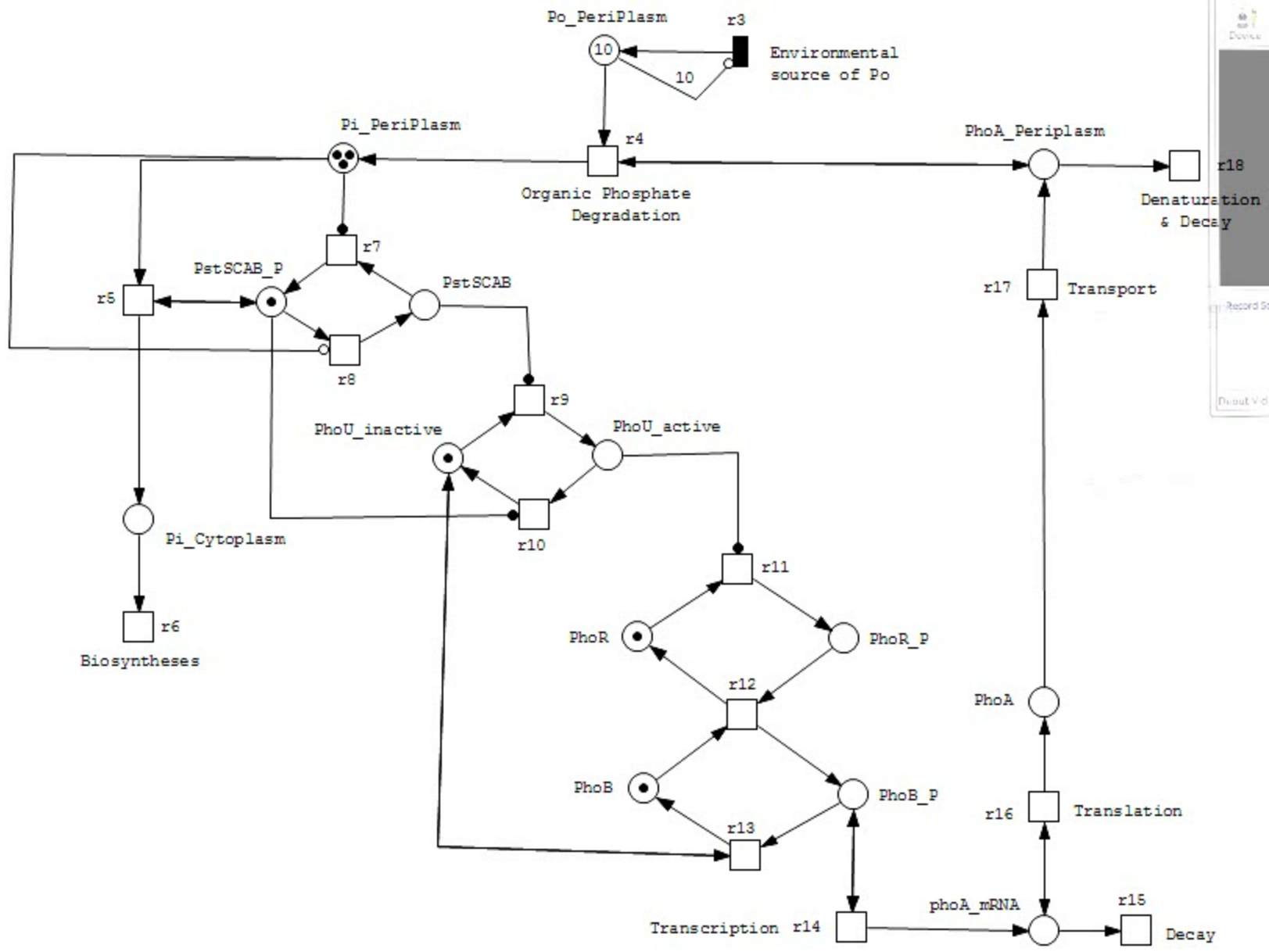
PhoR-PhoB: système à deux composants

PhoR: histidine kinase

PhoB: régulateur de réponse

Petri net and genetic regulatory network: a second example





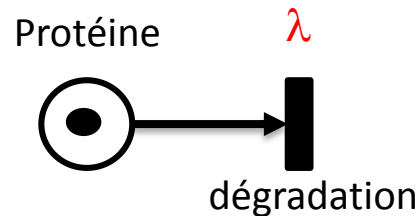
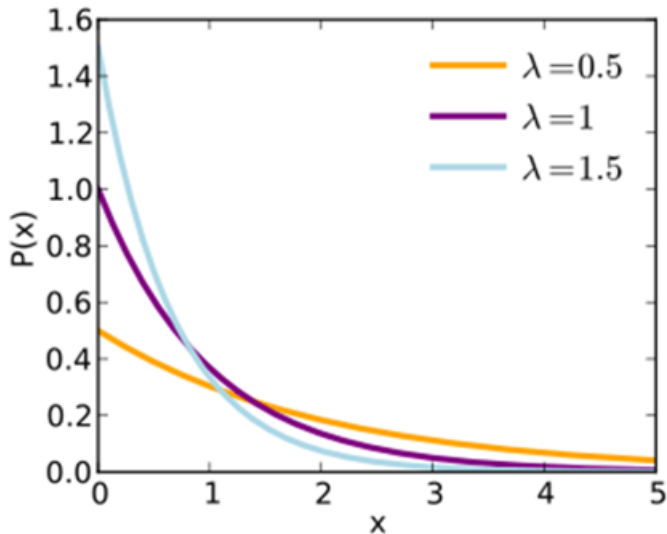
réseaux de Petri stochastique

Dans un réseau de Petri standard, pas de notion de temps.

Prise en compte du temps : deux possibilités stochastique ou continu.

Réseaux de Petri stochastique :

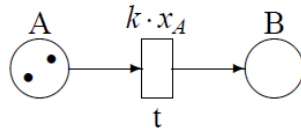
La vitesse à laquelle chaque réaction va avoir lieu est déterminée par une fonction de densité de probabilité, et cette fonction de densité de probabilité suit une distribution exponentielle négative, qui va représenter le délai d'attente avant l'activation d'une transition lorsque celle-ci est possible. Et donc plus la valeur du paramètre sera élevée, plus la transition aura une probabilité élevée d'être exécutée, et un temps d'attente réduit. Processus de Poisson -> la distribution de probabilité du temps d'attente entre 2 occurrence suit une loi exponentielle



réseaux de Petri stochastique

Fonction choisie → Mass Action : La forme la fonction de densité de probabilité va dépendre du temps pour que la réaction (en terme moléculaire) puisse avoir lieu, donc cela dépend de la constante de vitesse associée à la réaction et du nombre de jetons présents dans les pré-places de la transition. Plus le nombre de jetons sera élevé, plus la vitesse de réaction sera rapide et plus la probabilité d'exécution de la transition sera élevée.

Exemple 1:



k est la constante de la réaction

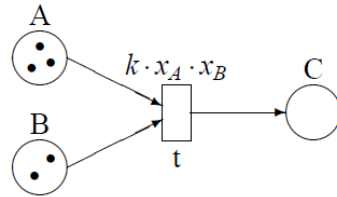
La transition est réalisable car la pré-place A contient suffisamment de jetons. Un délai avant l'exécution de la transition τ_1 est choisi pour la transition t tirée de la distribution exponentielle négative de paramètre $k x_A$ avec $k x_A = 2k$. Une horloge commence un compte à rebours de τ_1 à 0. Quand l'horloge atteint 0, la transition est exécutée.

On obtient un nouveau marquage du réseau avec 1 jeton dans A et 1 jeton dans B. La transition est donc toujours réalisable mais le délai avant son exécution a changé et vaut maintenant k car $x_A = 1$.

Donc le nouveau temps de délai d'exécution τ_2 sera sélectionné à partir d'une variable aléatoire exponentielle différente de celle de τ_1 . τ_2 sera plus long que τ_1 . De nouveau le compte à rebours est enclenché à partir de τ_2 . Le nouveau marquage est $x_A = 0$, $x_B = 2$. La transition n'est plus réalisable.

réseaux de Petri stochastique

Exemple 2:



k est la constante de la réaction

La transition est réalisable car les deux pré-places A et B sont suffisamment marquées

Dans le marquage initiale, il y a 6 possibilités indépendantes permettant à la réaction de se produire, chacune associée à une sélection spécifique des paires de molécules A et B qui vont interagir. La vitesse de réaction est donc : $k x_A x_B = 6k$.

Après la première exécution, le marquage change $x_A = 2, x_B = 1, x_C = 1$

La nouvelle vitesse de la réaction change donc et devient $k x_A x_B = 2k$.

réseaux de Petri stochastique

Une simulation décrit un comportement du système. Comme, nous sommes dans un modèle stochastique, une deuxième simulation décrira un comportement qui sera différent. En réalisant plusieurs simulations, la moyenne des résultats sur l'ensemble des simulations est calculé et une cinétique moyenne des composants du système sera calculée (comportement moyen du système)

Plus on réalisera de simulations, plus la cinétique moyenne sera précise. Chaque simulation individuelle fluctuera autour de cette cinétique moyenne.

