

Regulation of Gene Expression

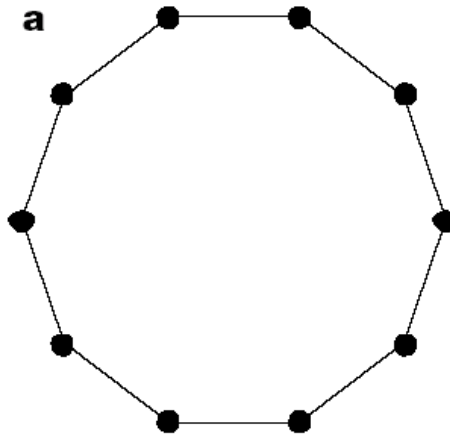
- Hans Geiselmann
hans.geiselmann@ujf-grenoble.fr

Régulation globale de l'expression génique

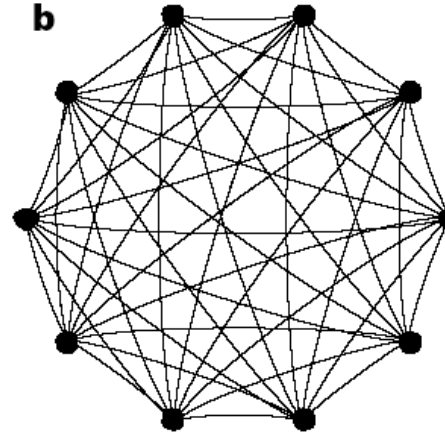
- Les connexions régulatrices dans un réseau de régulation, Topologie, aspect globaux
- Méthodes globales de mesure
- Modules et motifs
 - Fonctionnalité des motifs
 - Analyse de la dynamique d'un motif
- Réseaux de régulation
 - Opéron *lac*, Décision développementale (phage λ)
 - Réseaux artificiels
- Robustesse

Topologie d'un réseau

- La connectivité même d'un réseau affecte profondément son fonctionnement



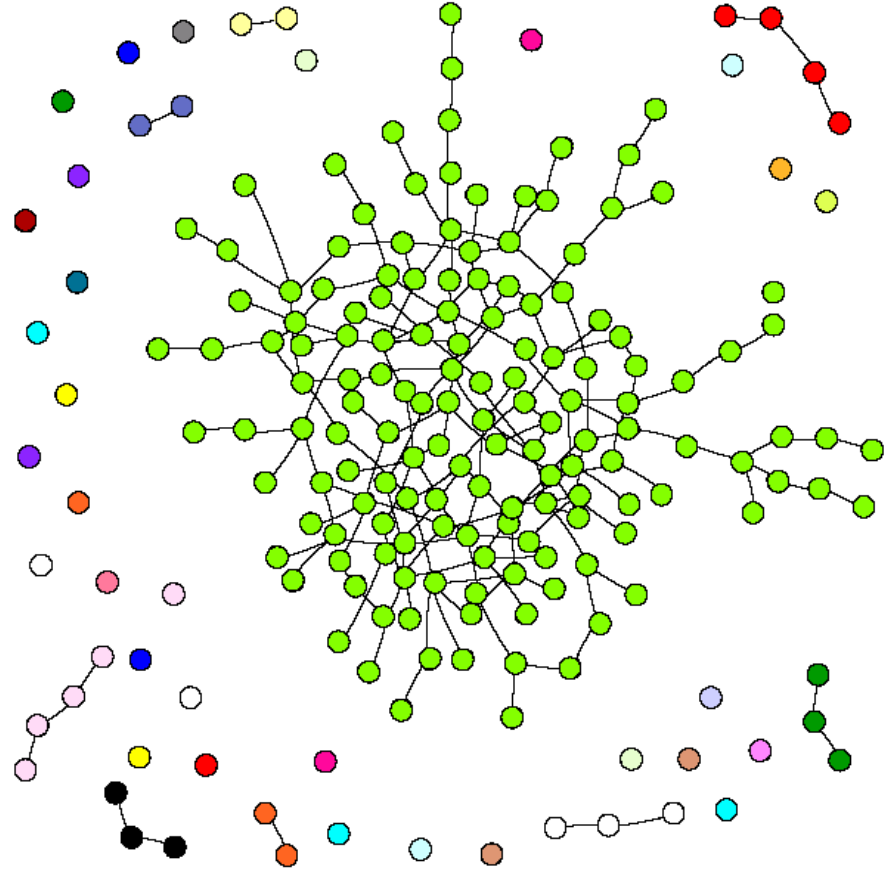
Connexions régulières



Réseau complètement connecté

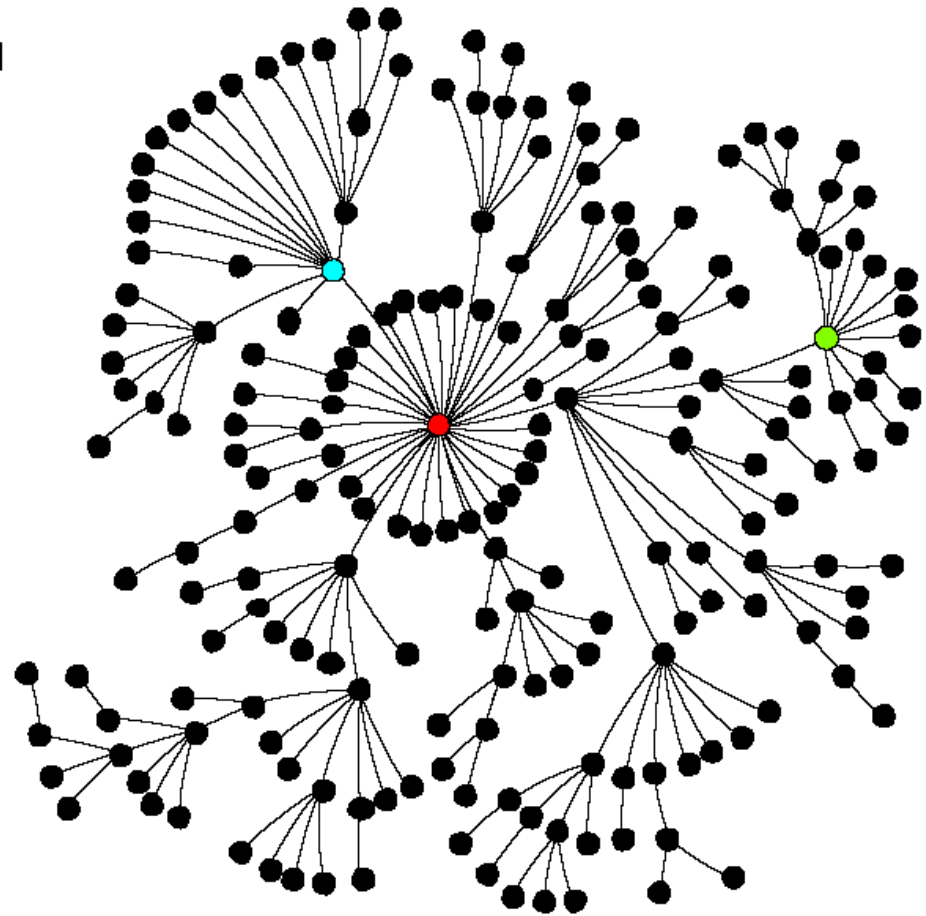
Topologie d'un réseau

- Un réseau connecté^c au hasard produit une composante géante fortement connectée



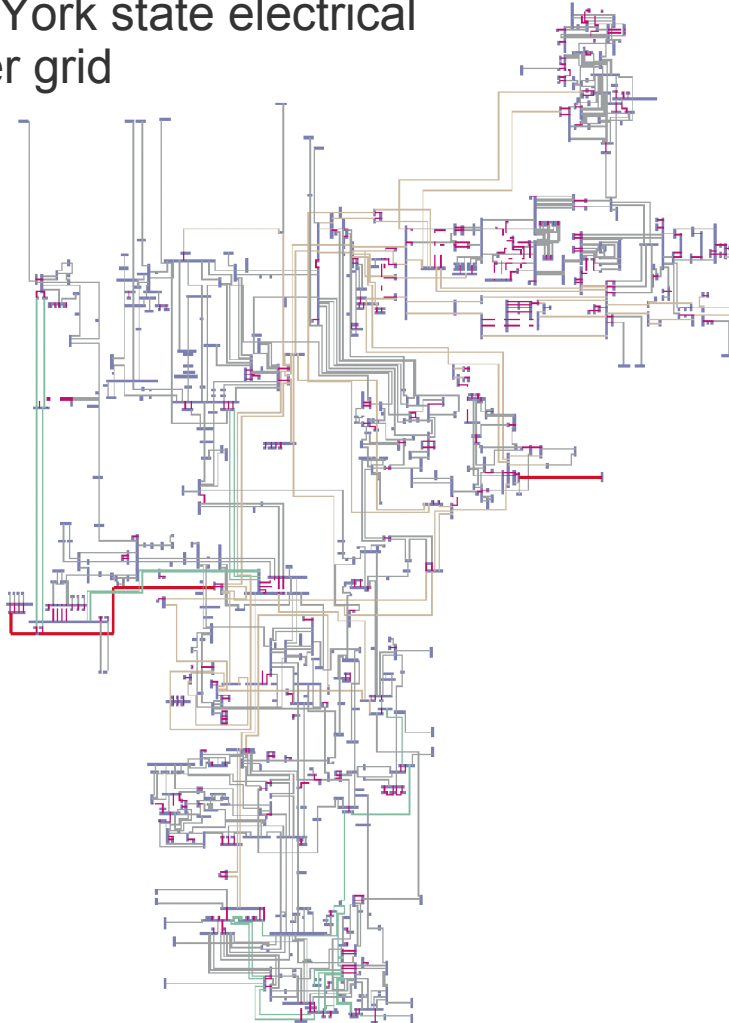
Topologie d'un réseau

- **Scale-free graph**, grown by attaching new nodes at random to previously existing nodes. The probability of attachment is proportional to the degree of the target node; thus richly connected nodes tend to get richer, leading to the formation of hubs and a skewed degree distribution with a heavy tail.

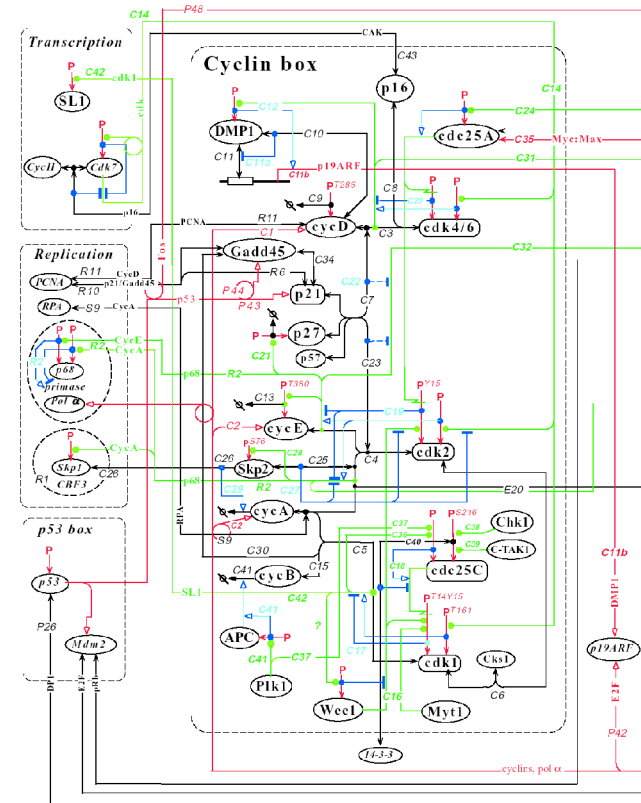


Engineering and biology

New York state electrical power grid

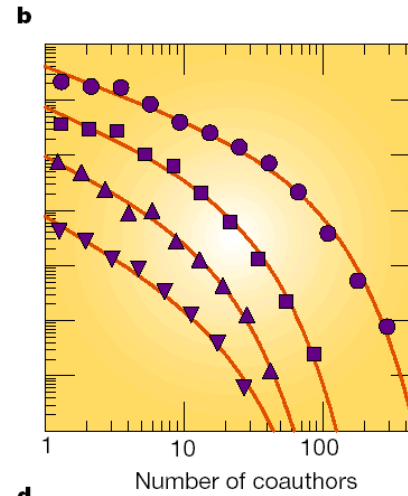
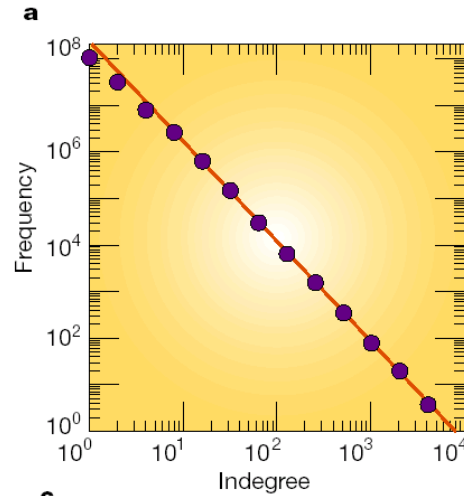


Part of the cell cycle network



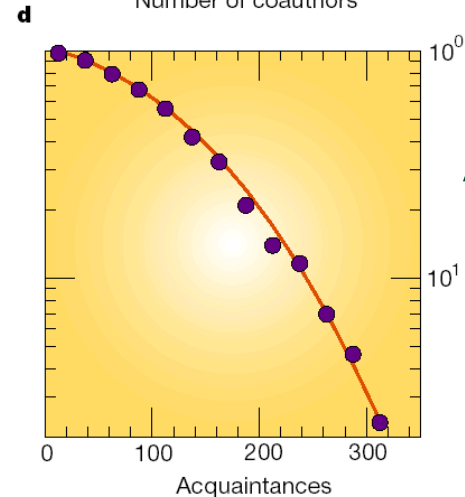
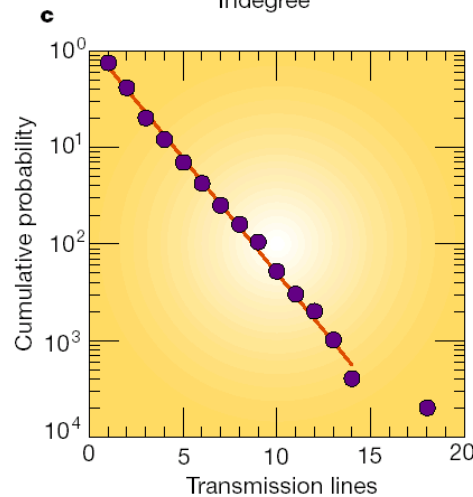
Degree distributions of real networks

Internet
Metabolism



Co-authorship

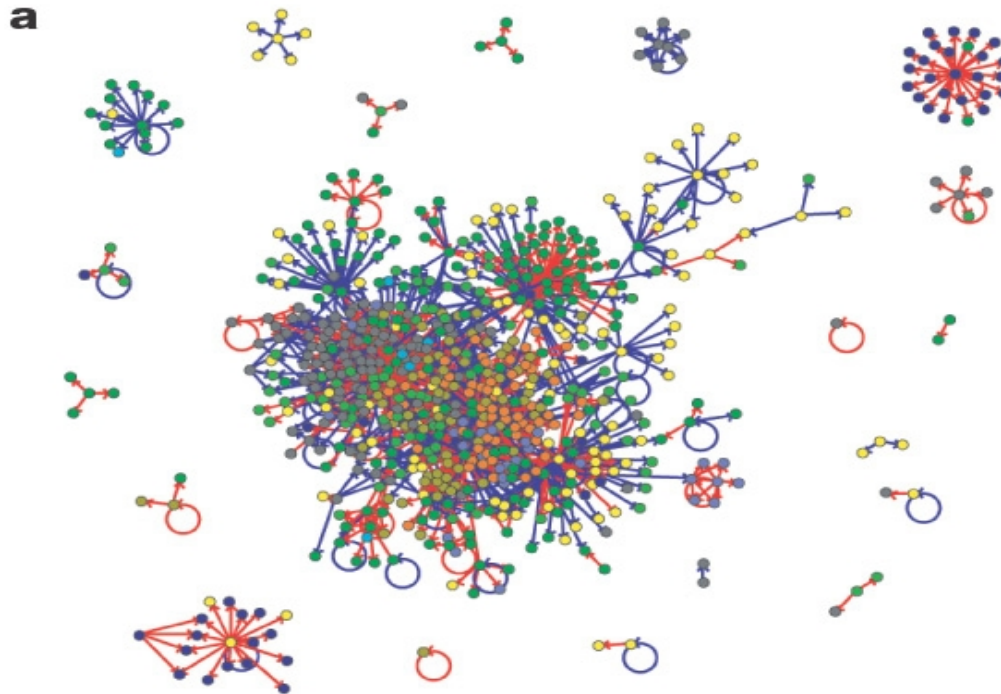
Powergrid of
the Western US
and Canada



Acquaintances

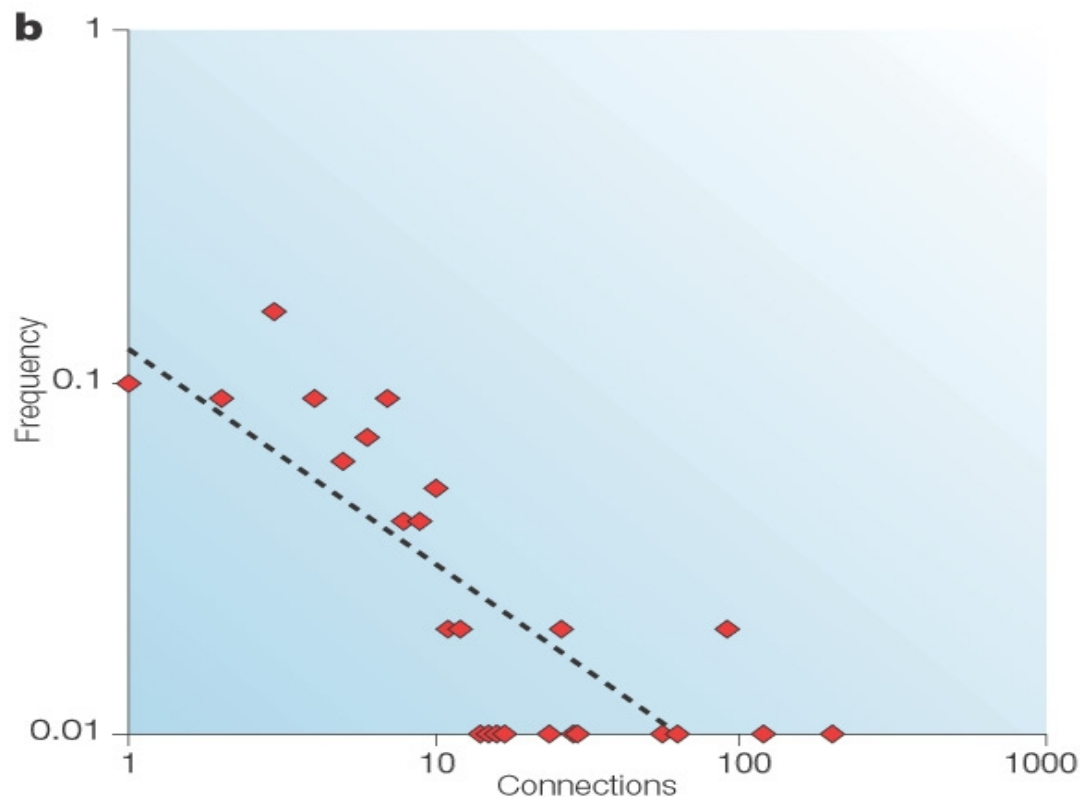
Caractéristiques de réseaux de régulation transcriptionnelle

□ Le réseau d'*Escherichia coli*



Caractéristiques de réseaux de régulation transcriptionnelle

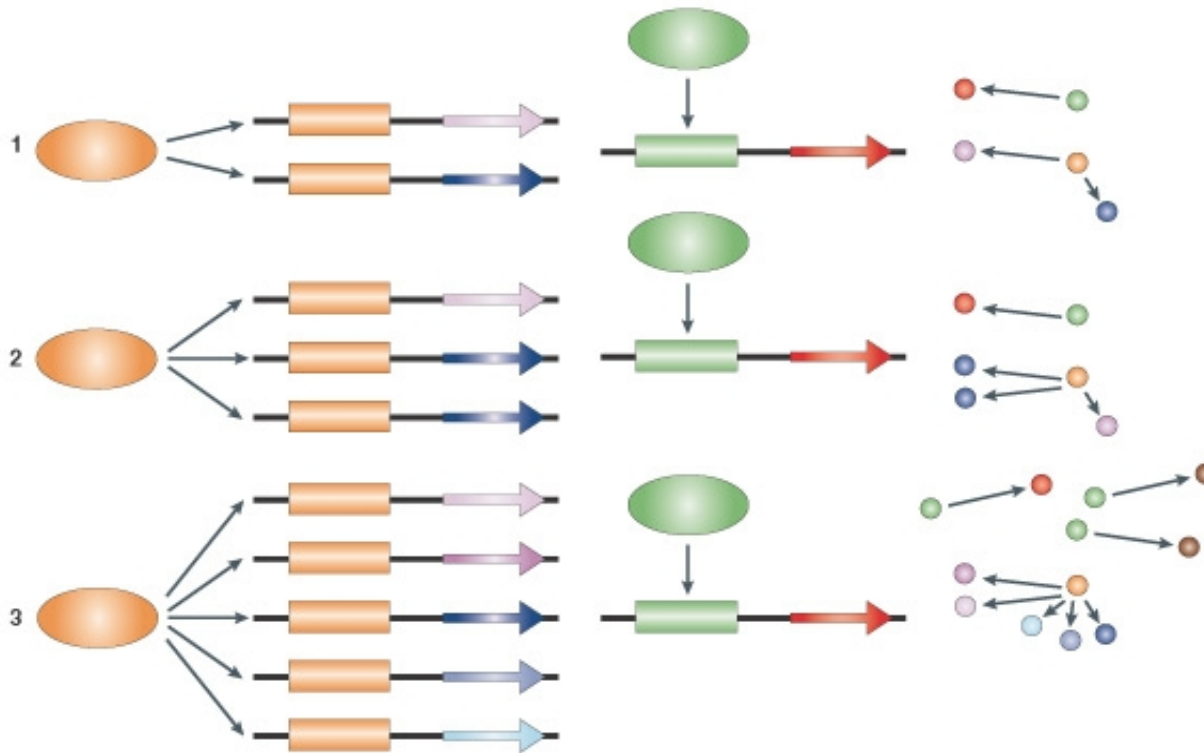
□ Le réseau d'*Escherichia coli* est “scale-free”



Le principe de Mathieu appliqué aux réseaux de transcription

□ On ne prête qu'aux riches

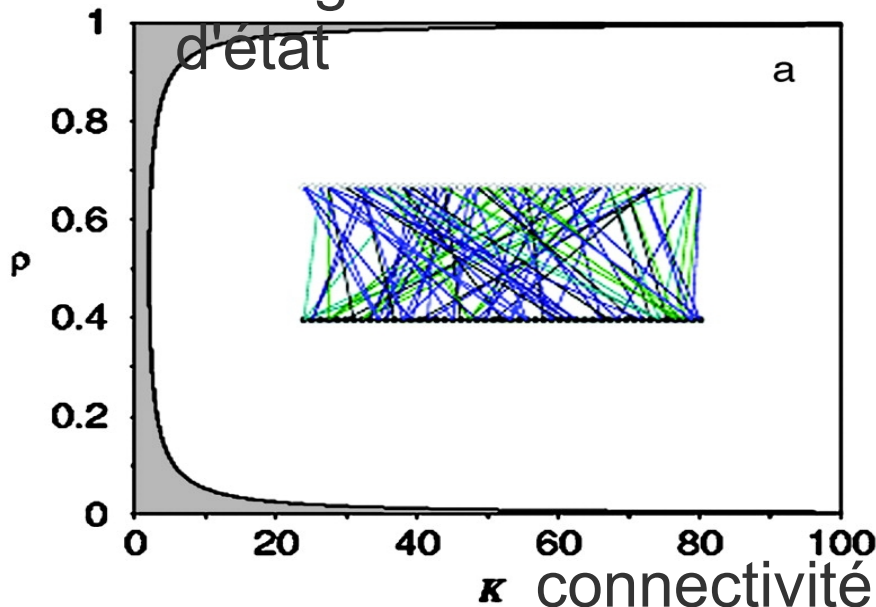
The 'Matthew principle': Matthew 13:12, "...for whosoever hath, to him shall be given, and he shall have more abundance: but whosoever hath not, from him shall be taken away even that he hath..."



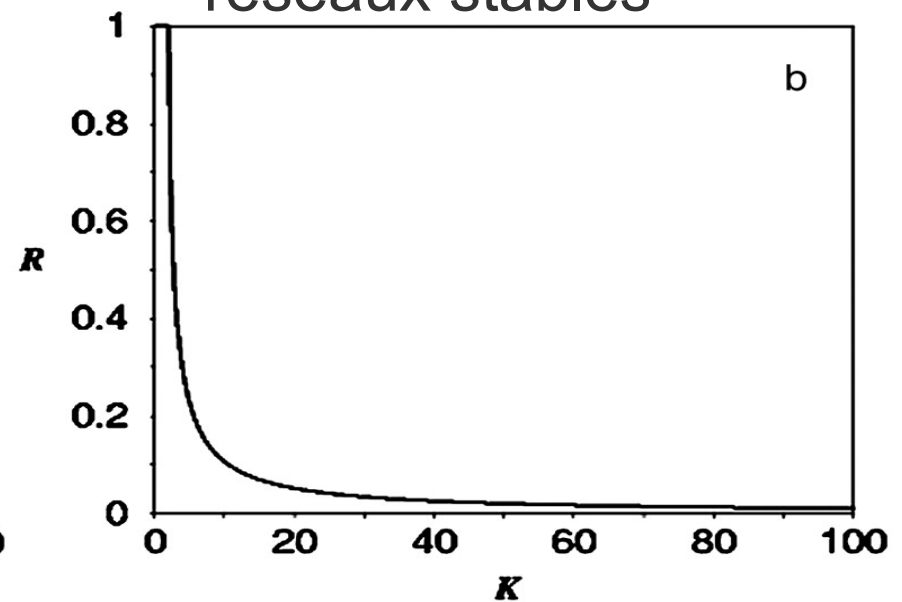
Robustness d'un réseau aléatoire

- Un réseau aléatoire est chaotique
- Seulement un jeu très particulier de paramètres conduit à un comportement stable

Probabilité de
changement
d'état

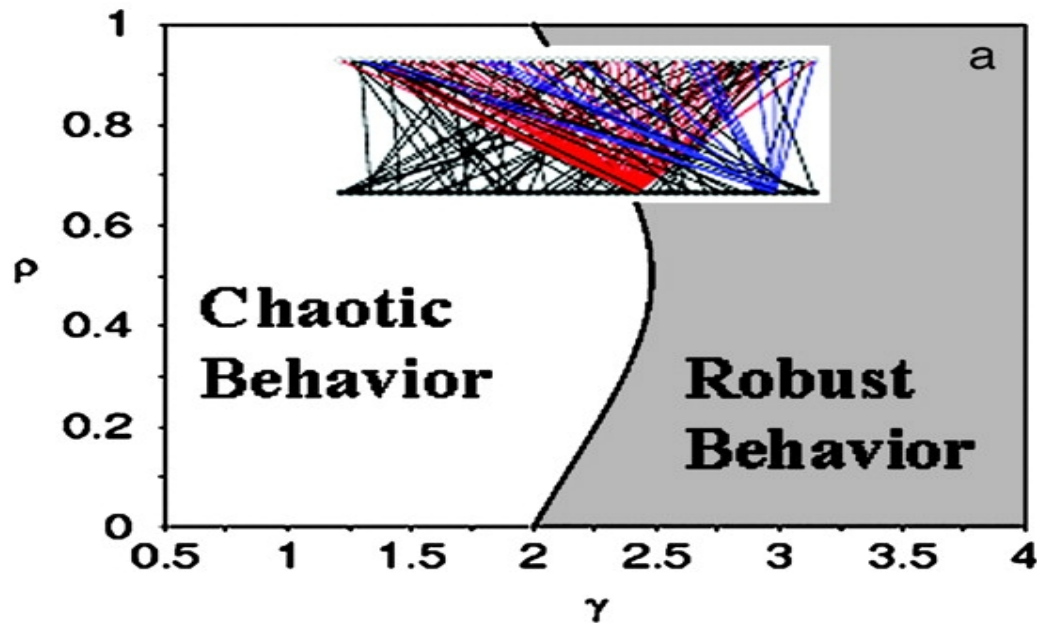


Fraction de
réseaux stables



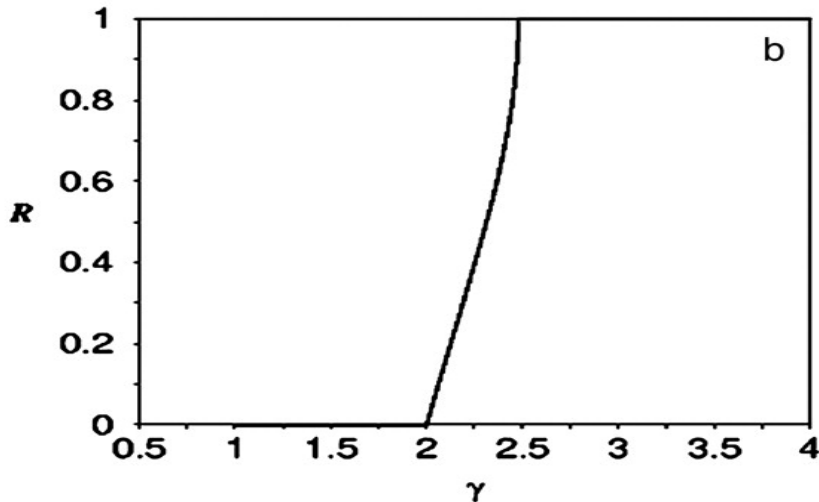
Robustness d'un réseau aléatoire

- Un réseau aléatoire “scale-free” est robuste

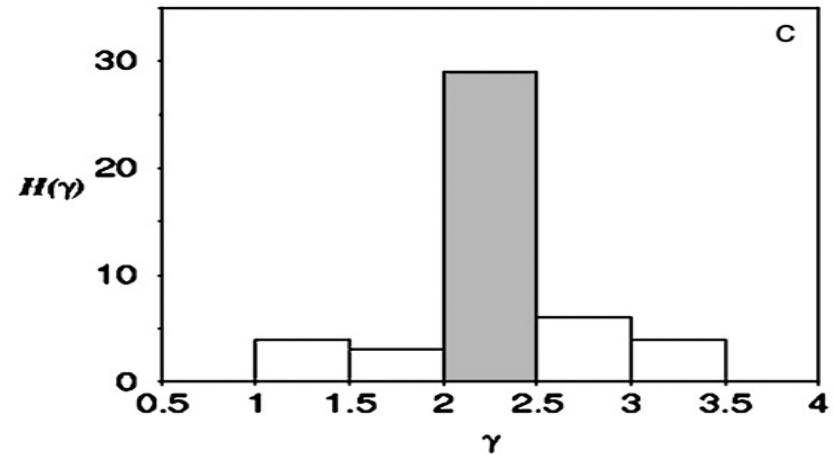


Robustness d'un réseau aléatoire scale-free

□ Un réseau aléatoire “scale-free” est robuste



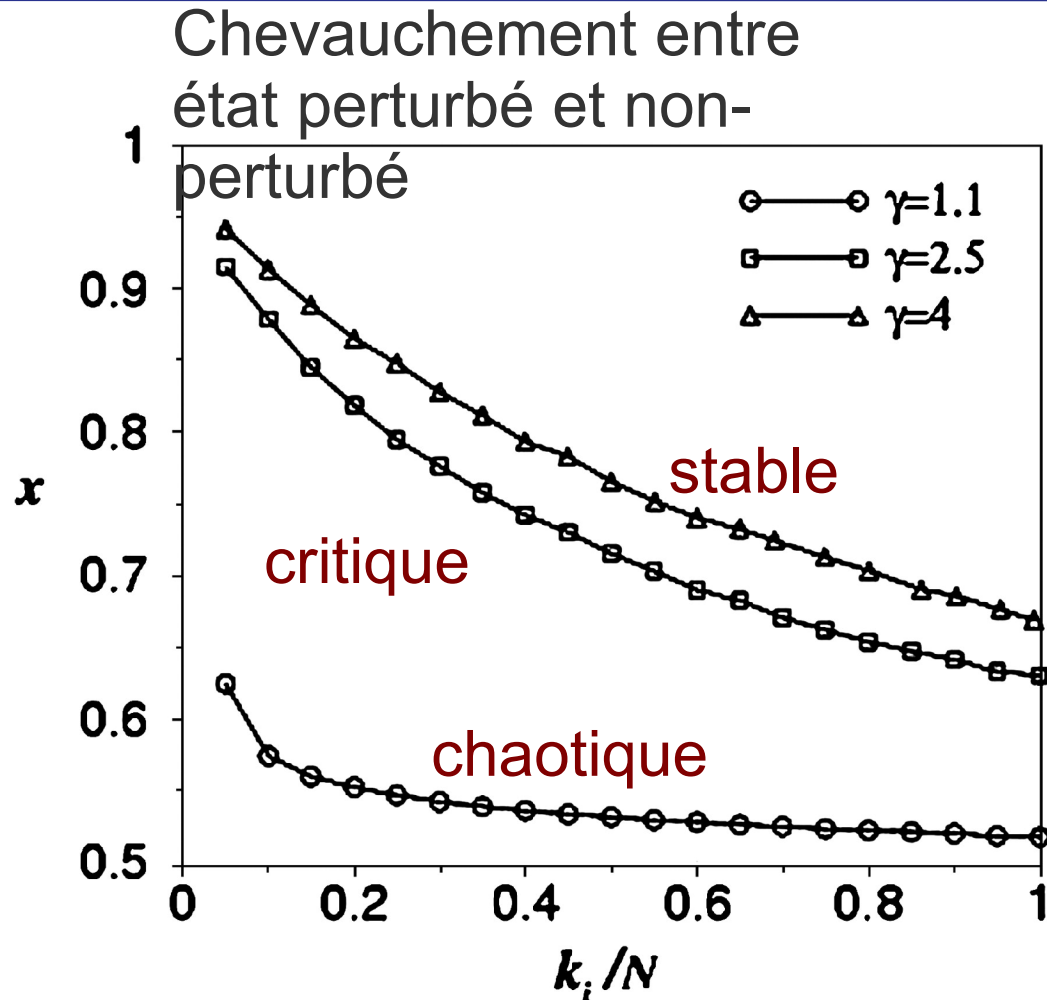
Transition entre chaotique et stable



La plupart des réseaux naturel sont scale-free avec un exposant entre 2 et 2.5

Stabilité dynamique d'un réseau aléatoire

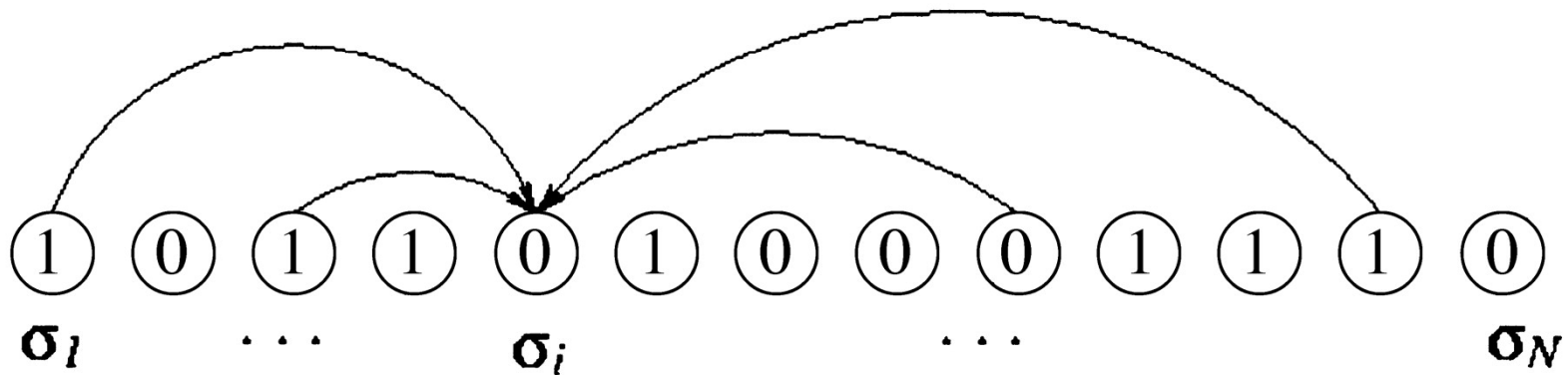
- Un réseau aléatoire "scale-free" est stable
- Plus la connectivité est grande, plus le réseau est instable



Le modèle de Kauffman

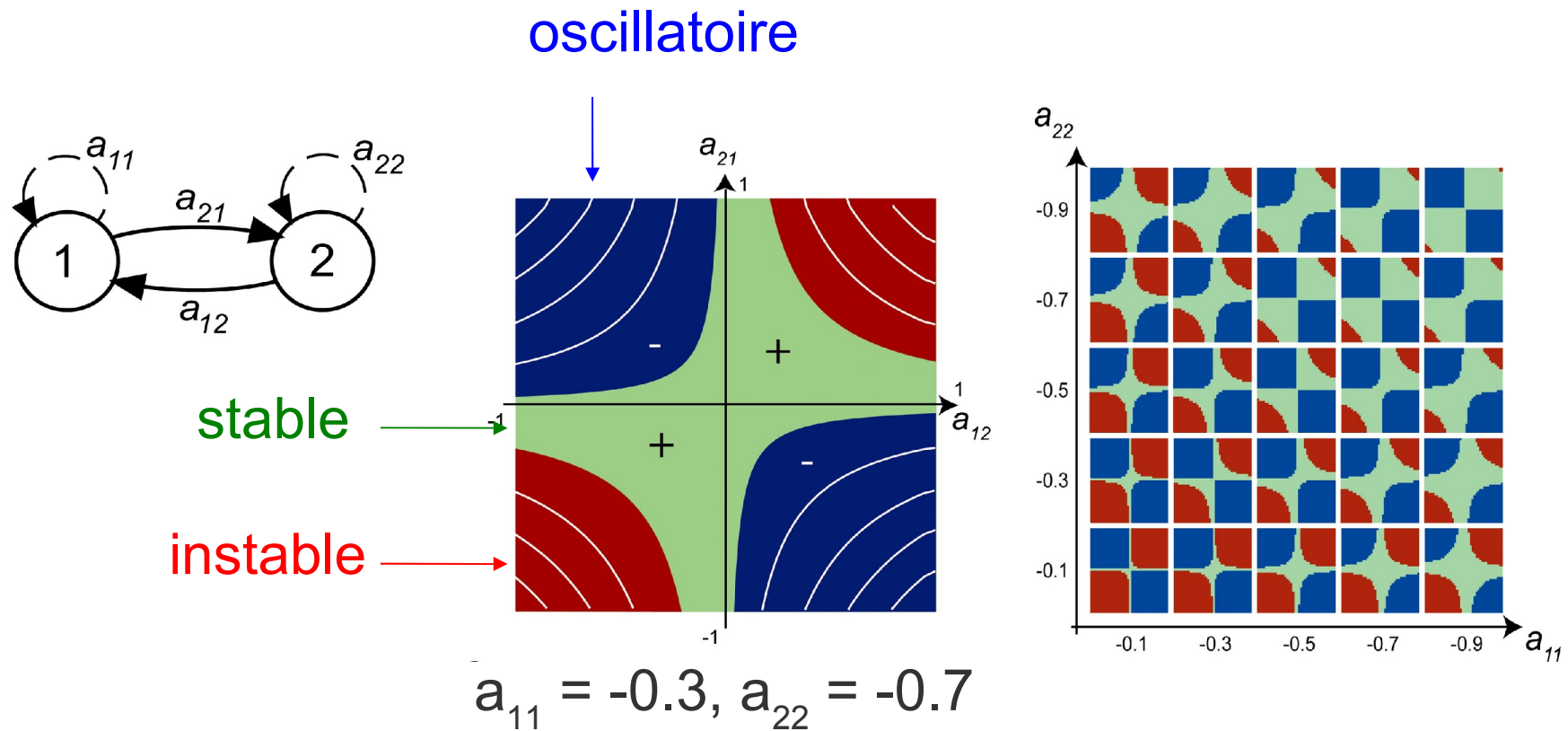
□ Connexions aléatoires

□ $K = 4$



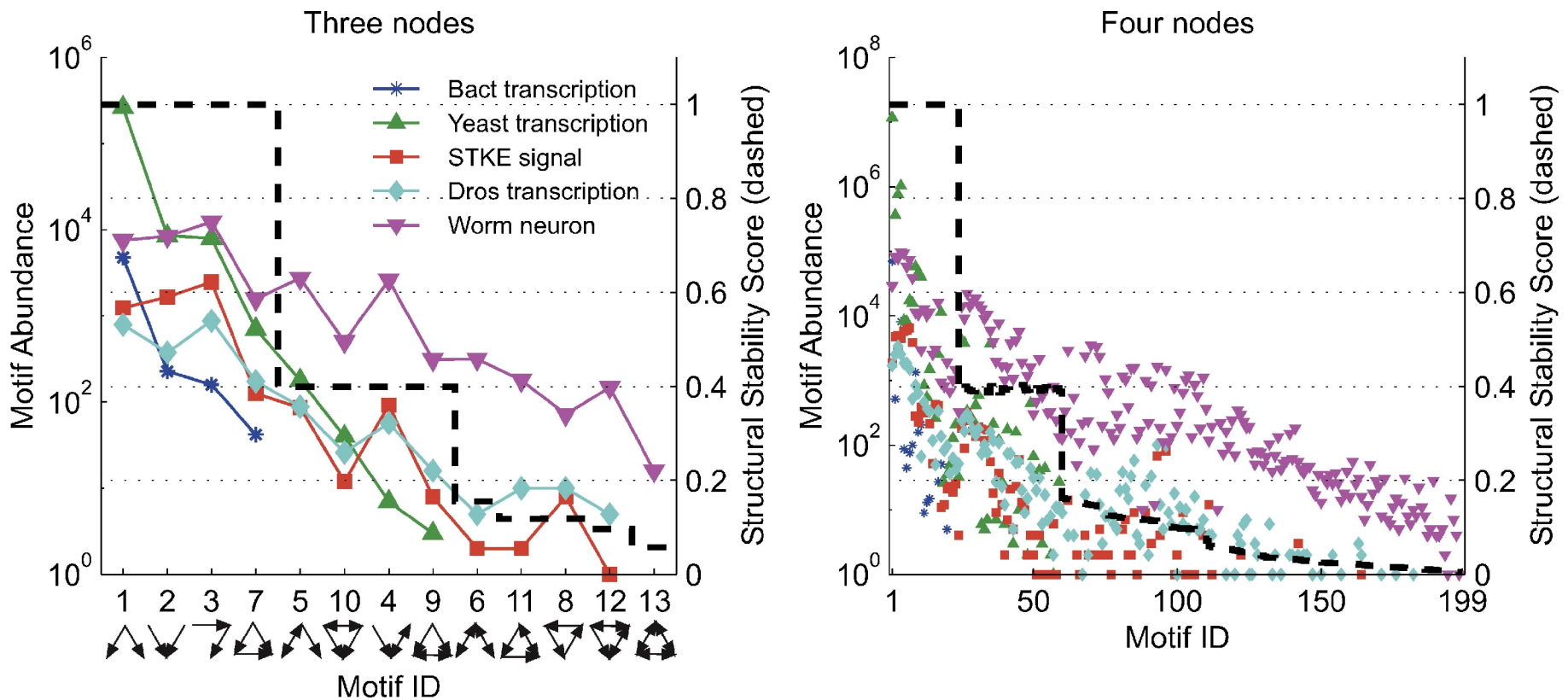
Dynamique d'un module

- Même un module simple peut avoir une dynamique complexe



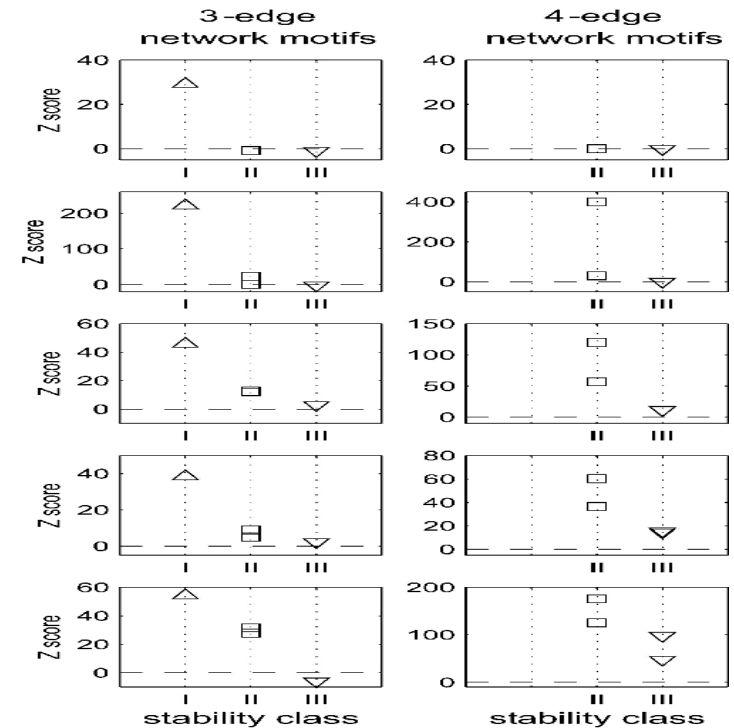
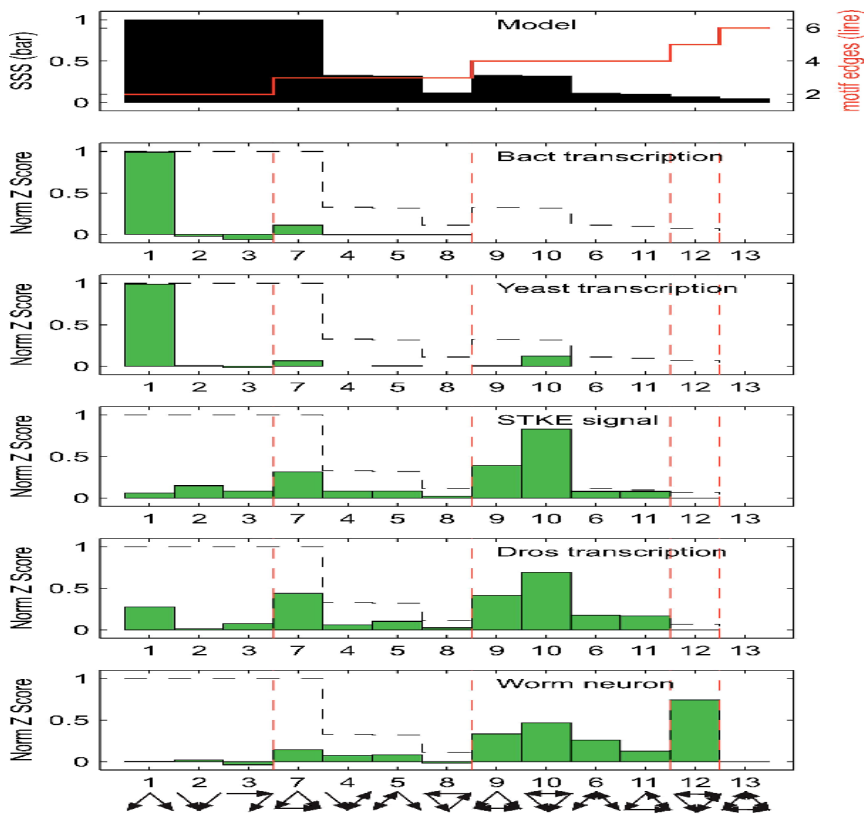
L'abondance d'un motif est corrélé avec son SSS

□ SSS: structural stability score (probabilité de retourner vers l'état stable après une perturbation)

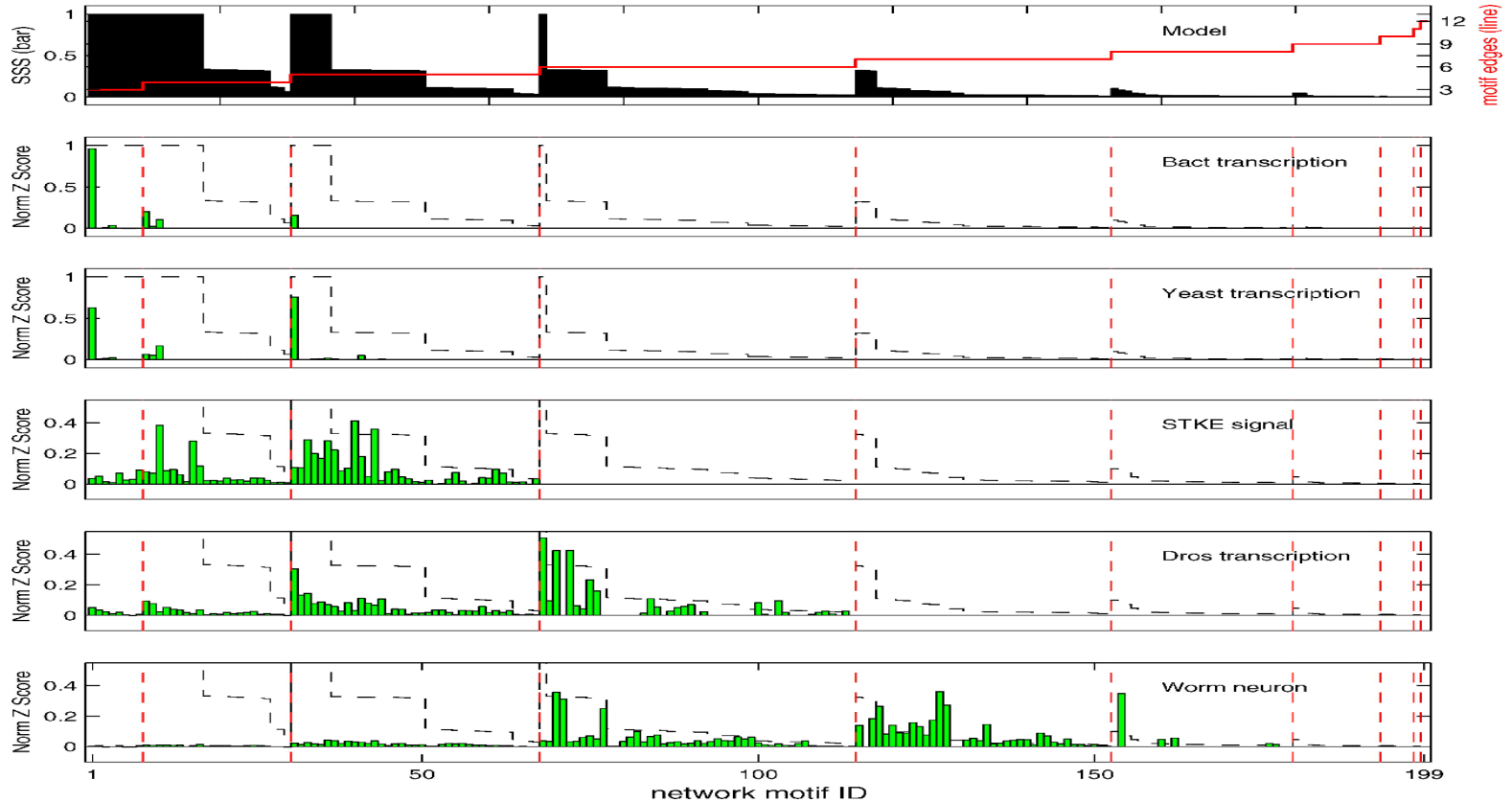


Le motif le plus stable est préféré

□ Pour un nombre de noeuds et de connexions donné, le motif le plus stable est sur-représenté



Le motif le plus stable est préféré (4 noeuds)



Le motif le plus stable est préféré

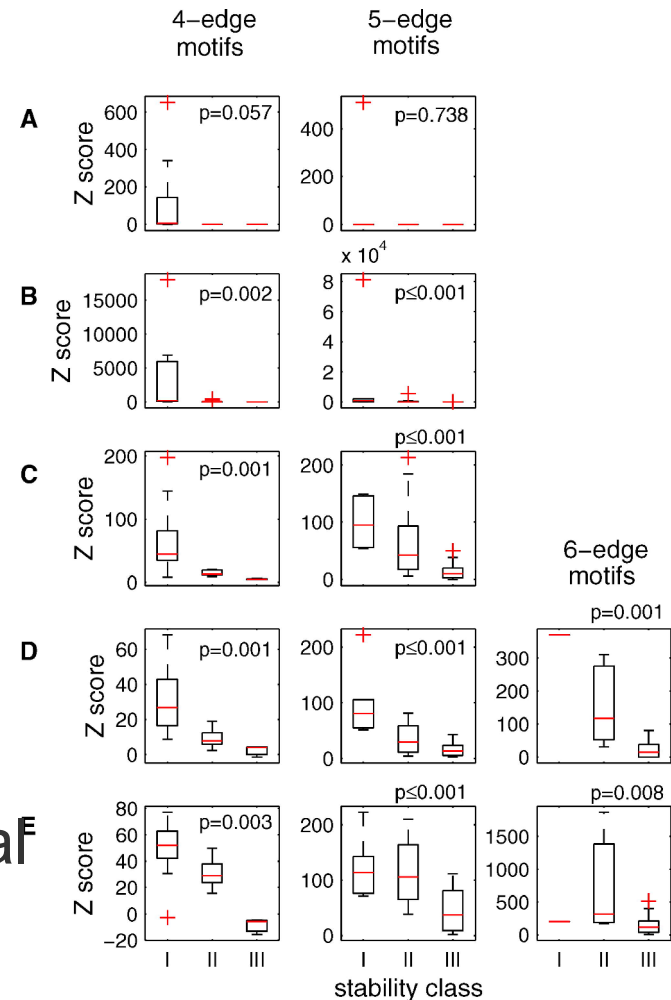
bactéries

levure

Transduction du signal

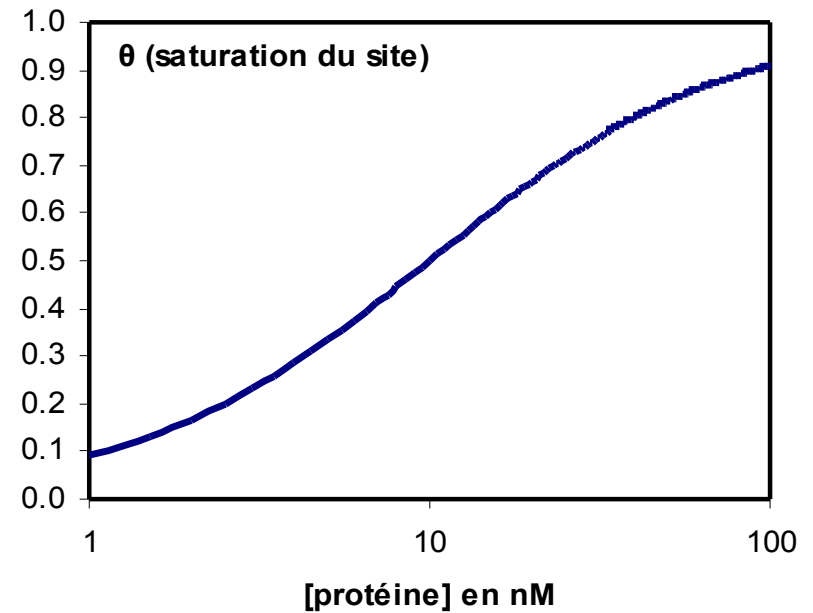
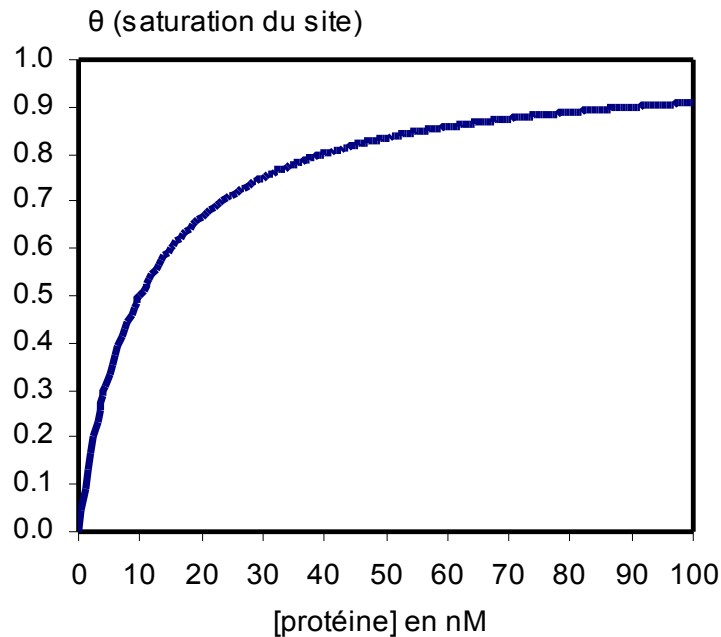
Développement de
D. melanogaster

Développement neuronal
de *C. elegans*



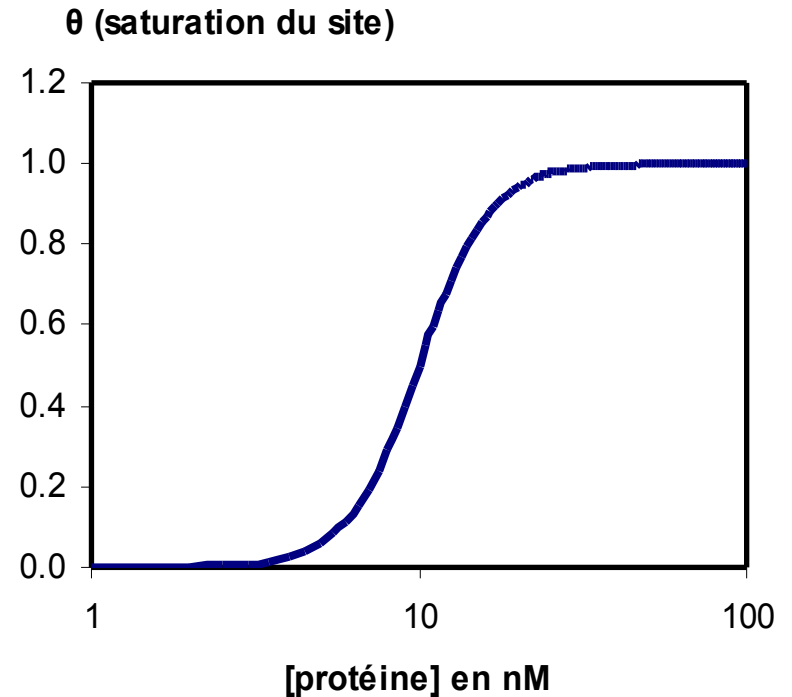
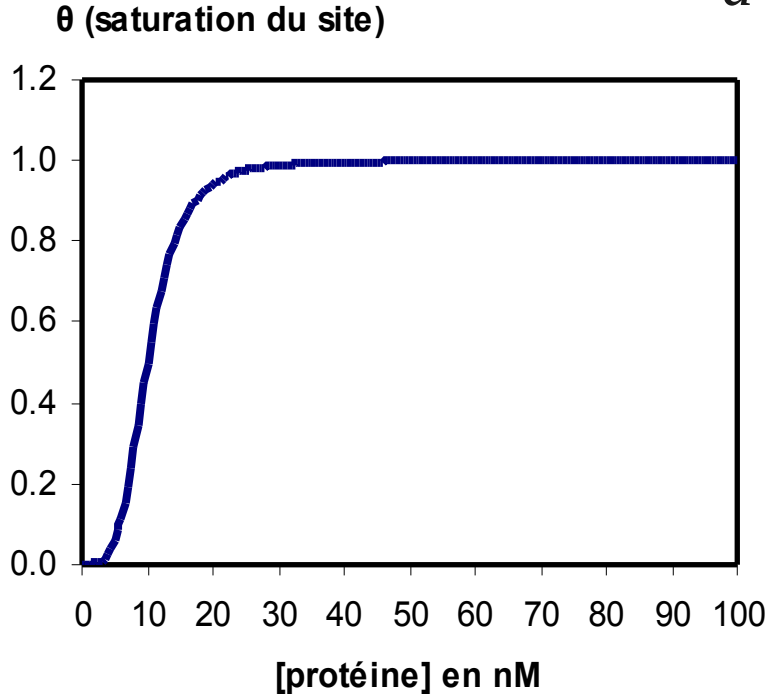
La liaison simple à l'ADN

$$\theta = \frac{[P]}{K_d + [P]}$$



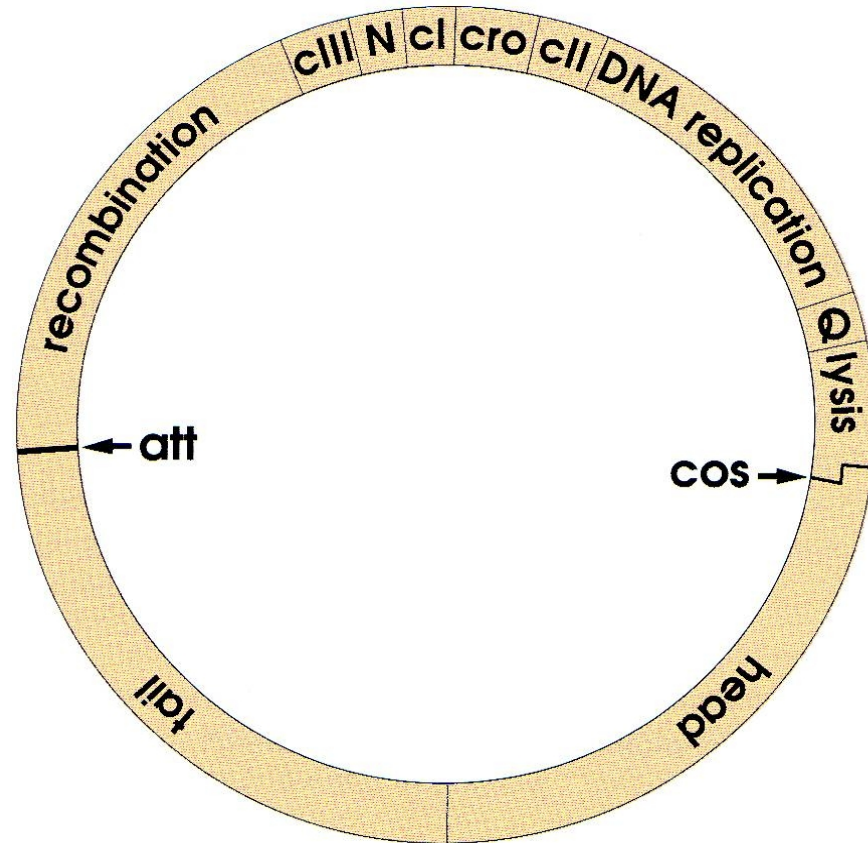
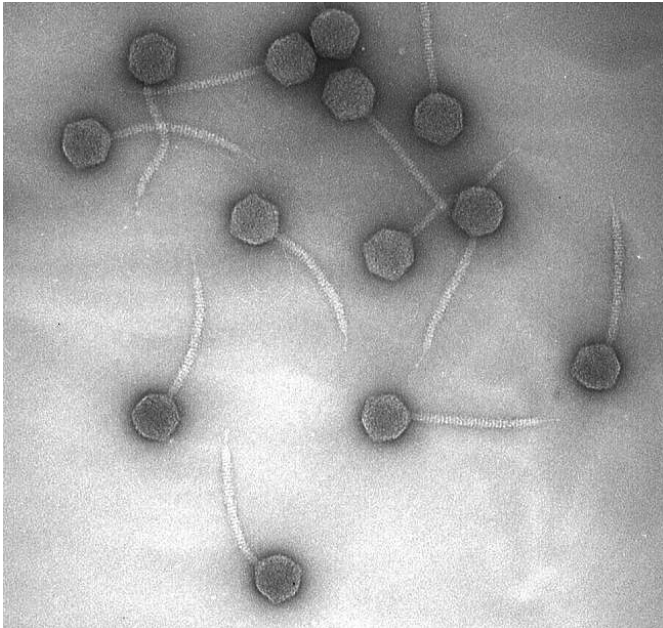
La liaison coopérative à l'ADN

$$\theta = \frac{[P]^n}{K_d + [P]^n}$$



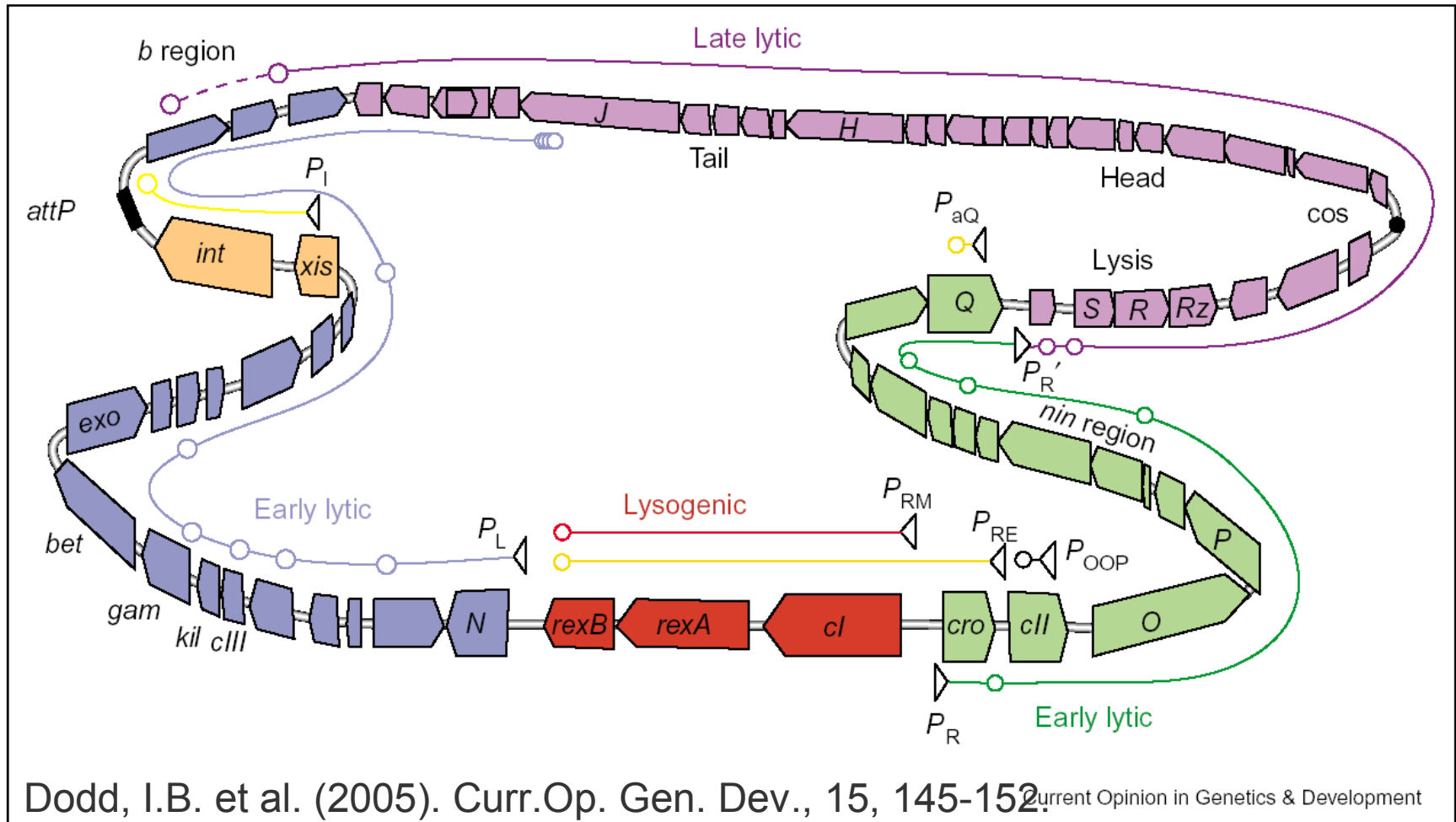
Le bactériophage λ

- Génome circulaire
- Lyse ou lysogénie

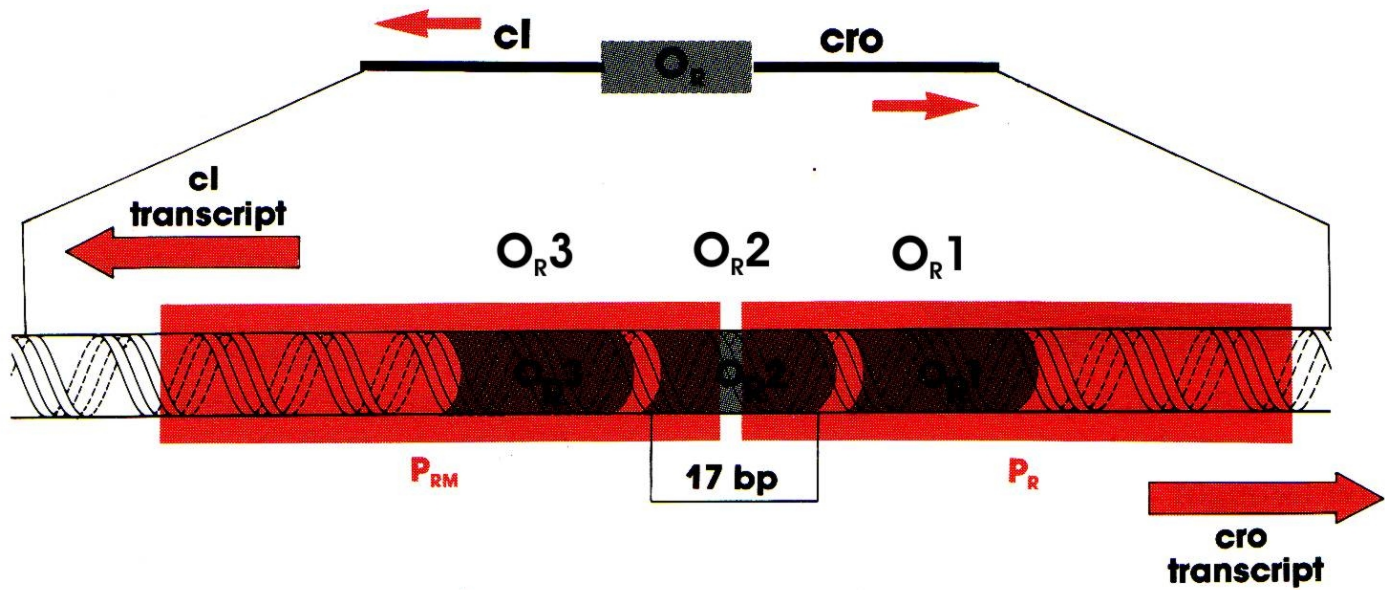


Ptashne, M. (1992). A genetic switch. Cell Press and Blackwell Scientific.

Le génome de λ



La région de contrôle



Ptashne, M. (1992). A genetic switch. Cell Press and Blackwell Scientific.

Liaison de Cro sur les opérateurs

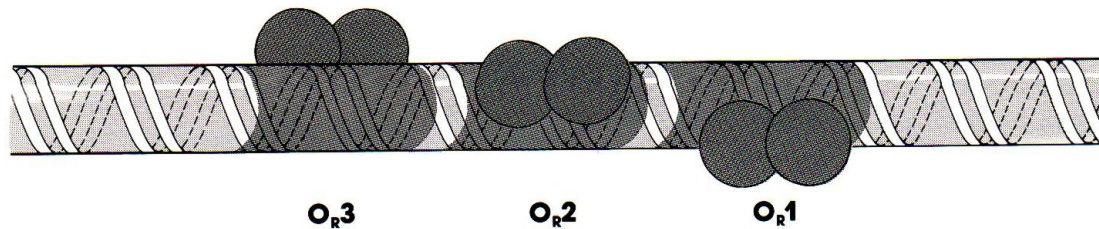
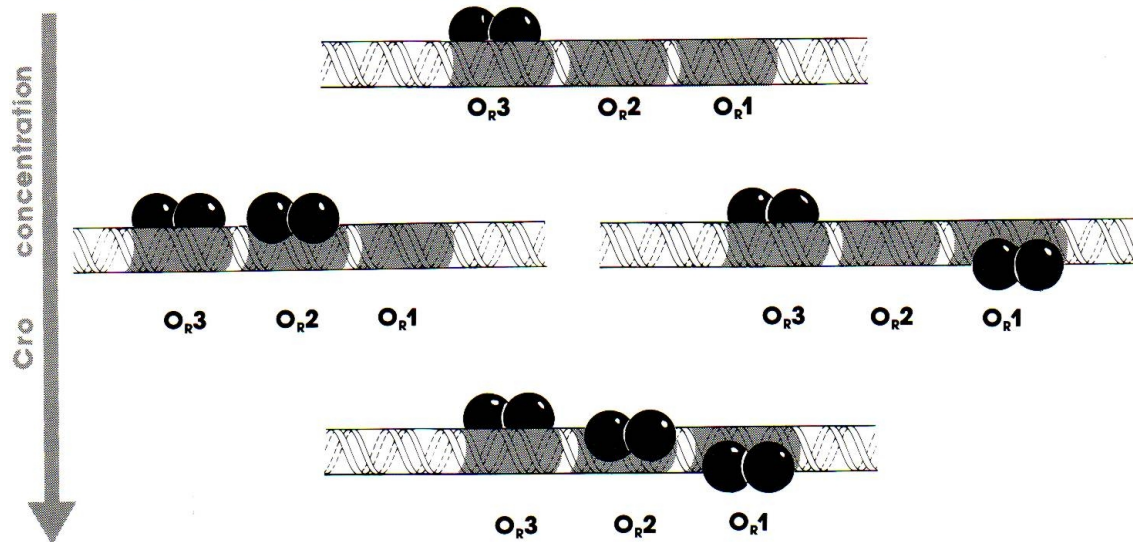
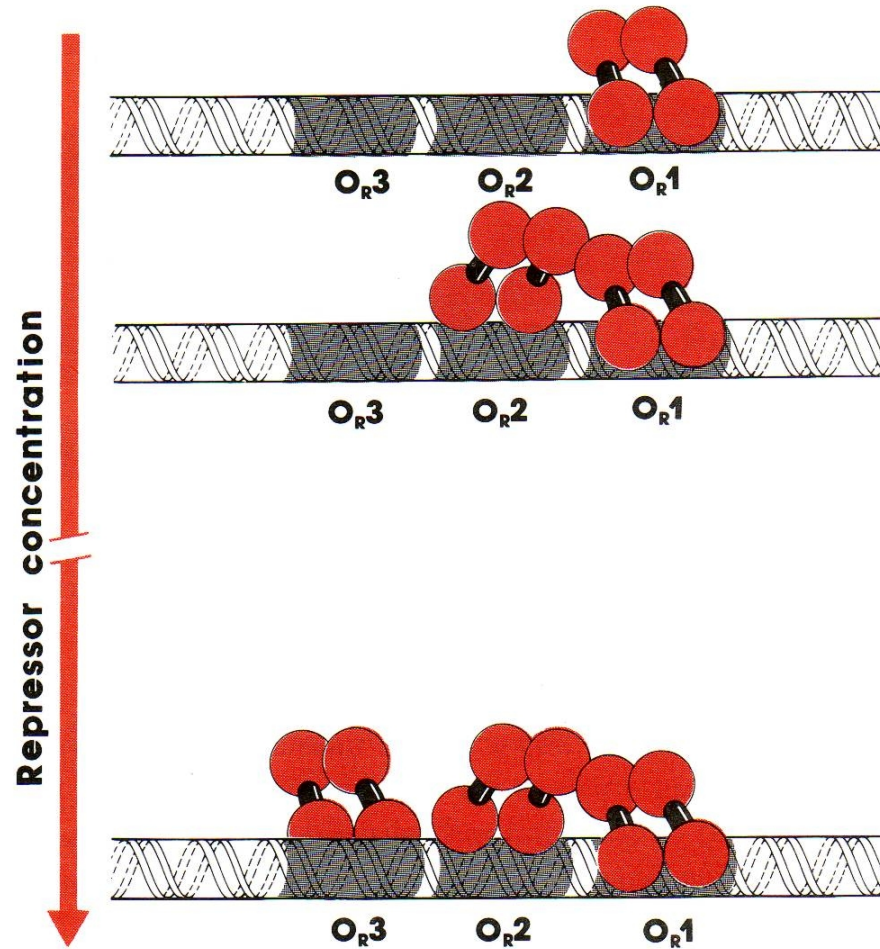


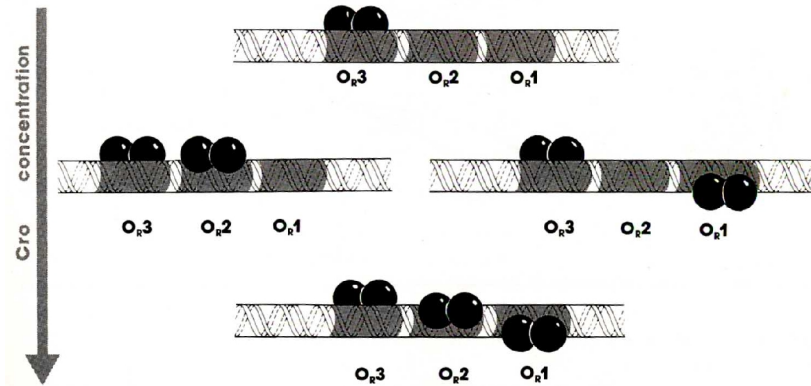
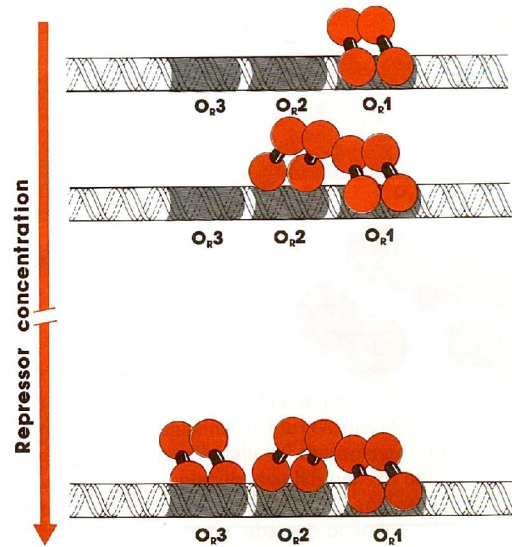
Figure 1.22. Cro bound to O_R . Cro dimers bind independently to each site in the tripartite operator.



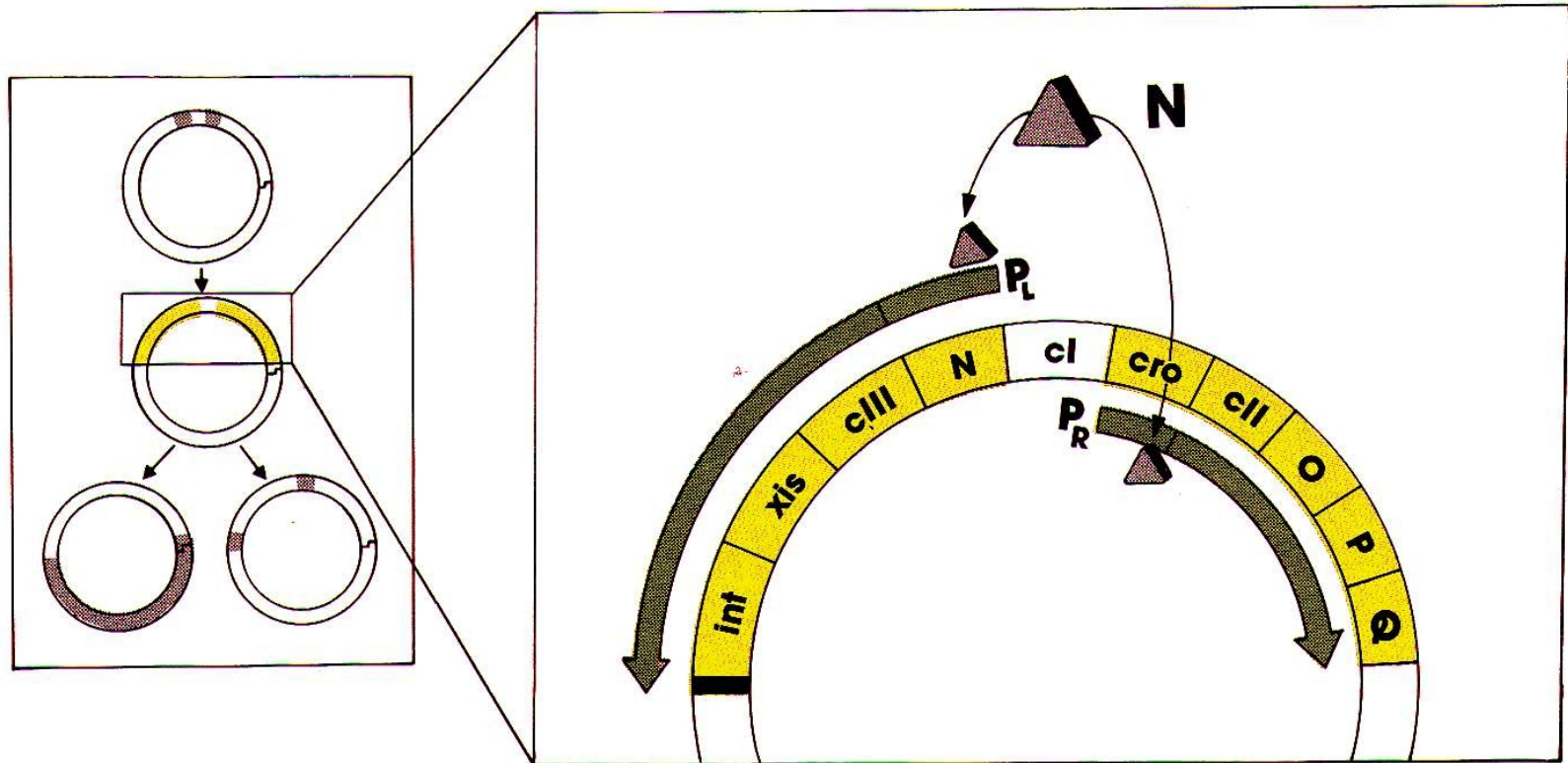
Liaison de CI sur les opérateurs



La décision lyse-lysogénie

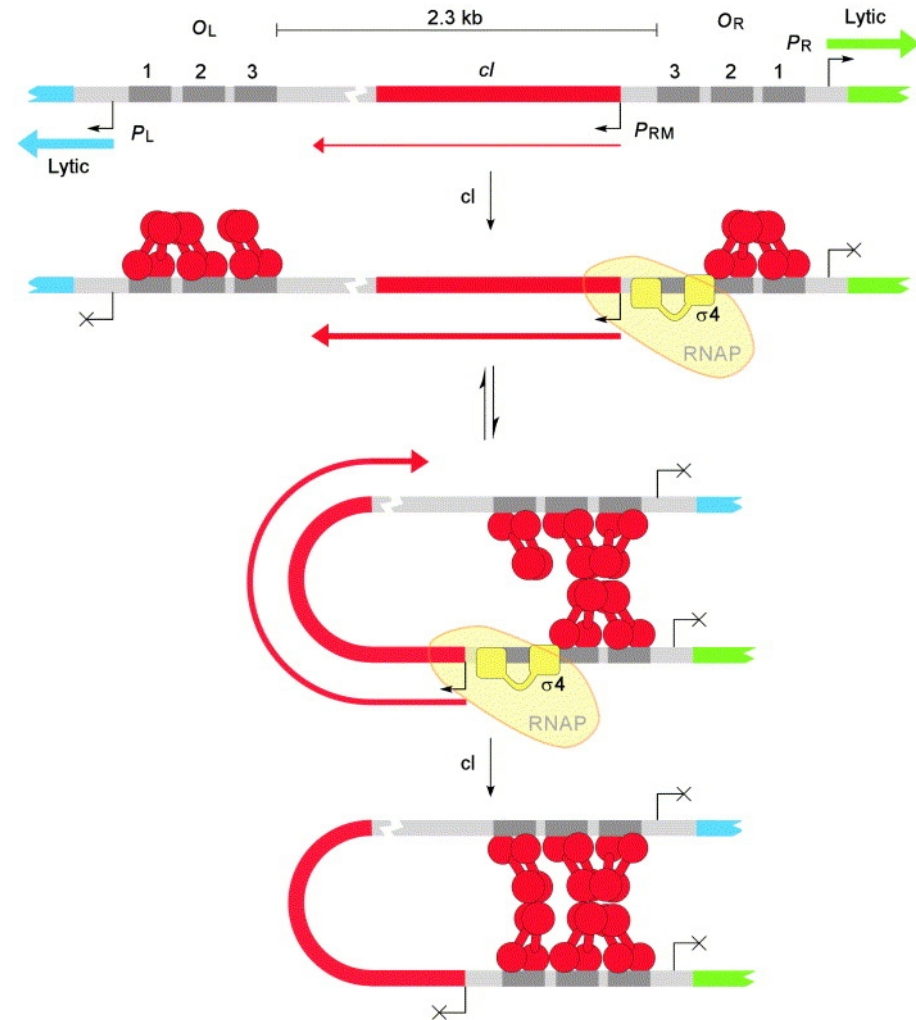


Evénements précoces après l'infection



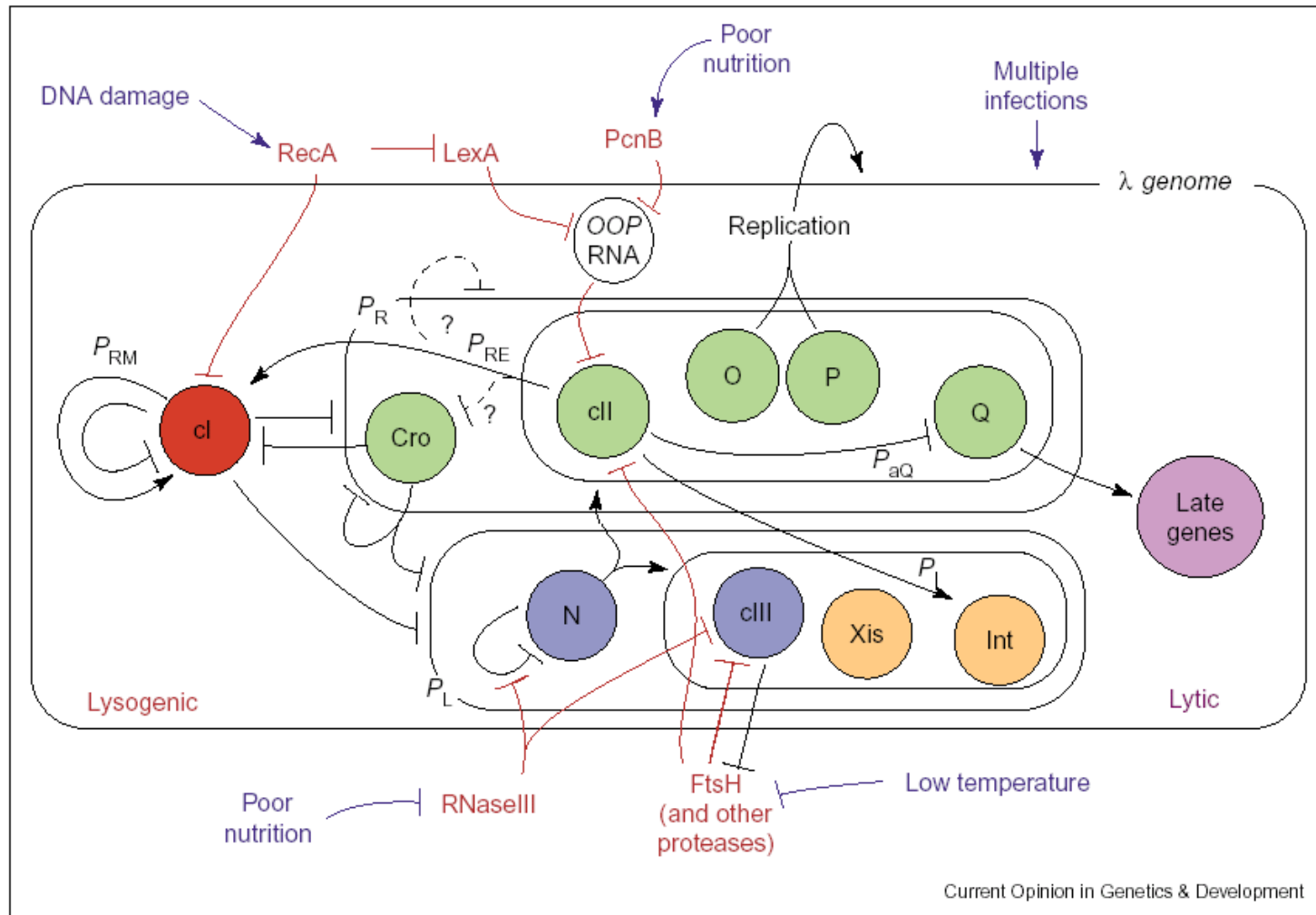
Régulation de la transcription par CI

- CI participe à la régulation de la transcription aux promoteurs p_R , p_L et p_{RM} .
- Coopérativité entre les dimères de CI



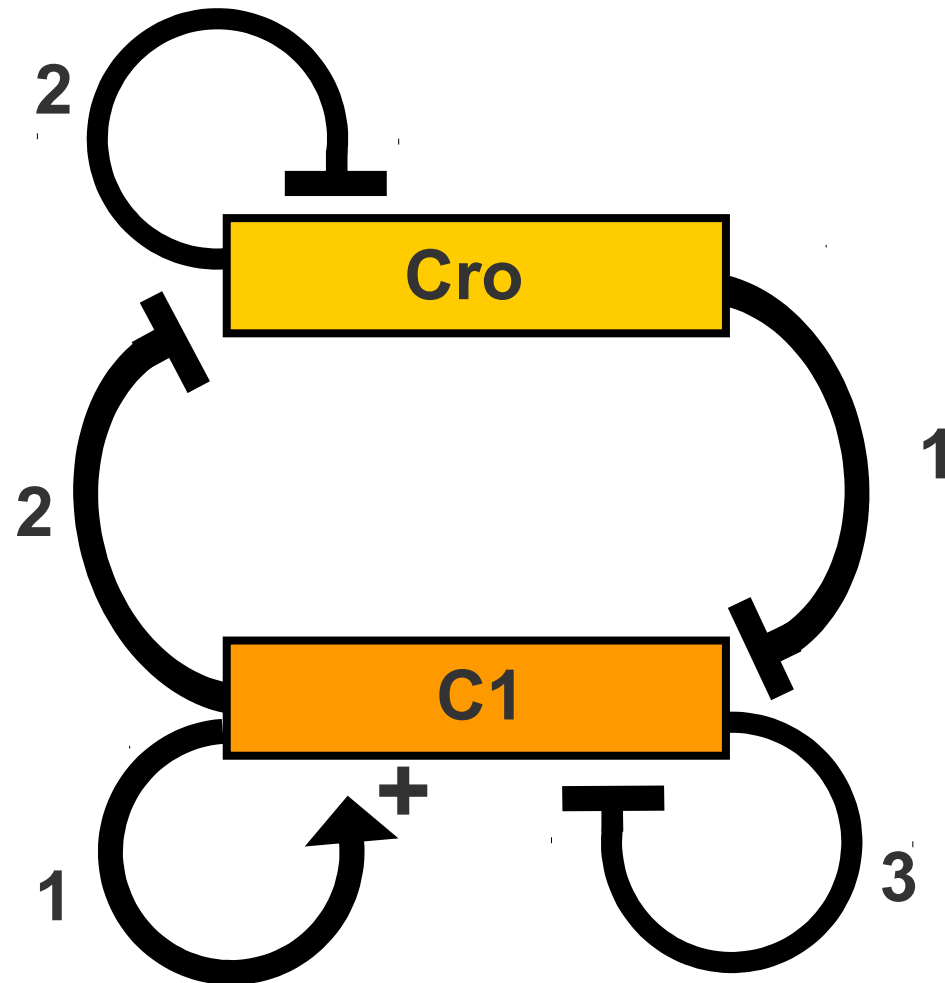
Dodd, I.B. et al. (2005).
Curr.Op. Gen. Dev., 15, 145-152.

Complexité du système de régulation du bactériophage λ

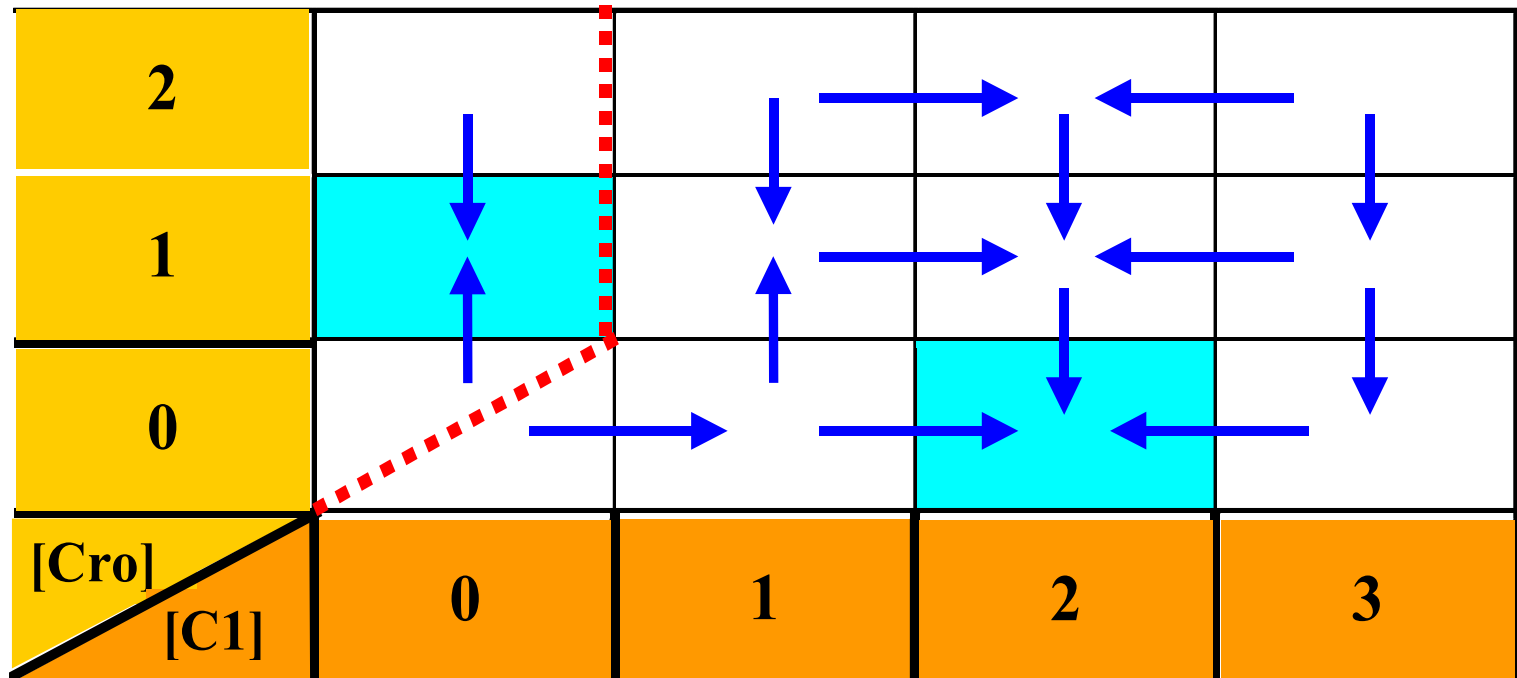


Modèle simple de la décision lyse-lysogénie

- La connectivité du réseau de régulation

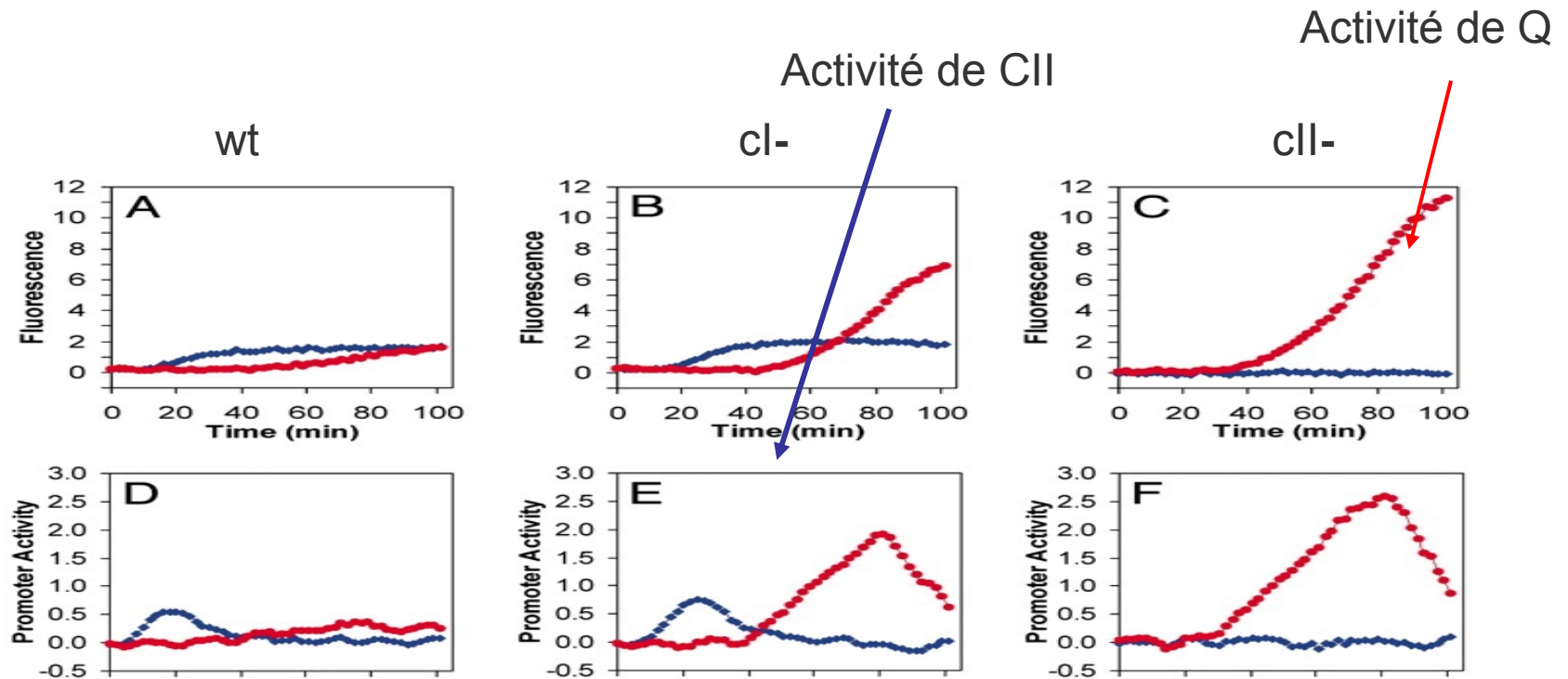


Dynamique qualitative du réseau de régulation



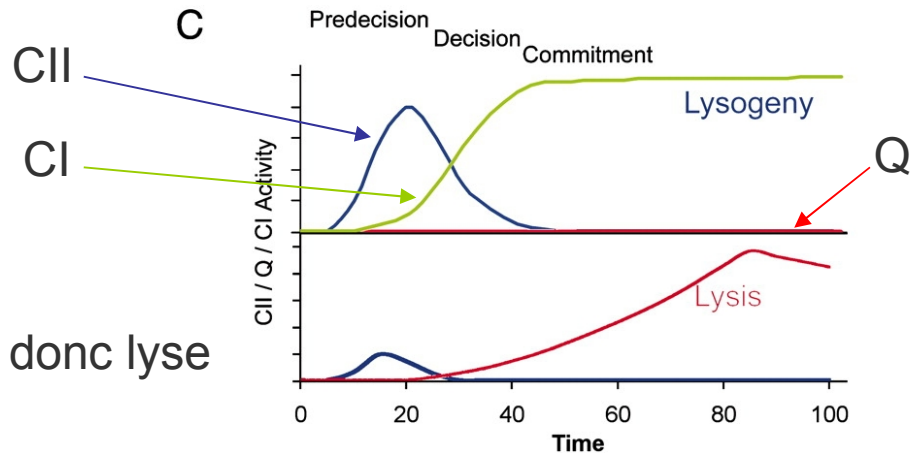
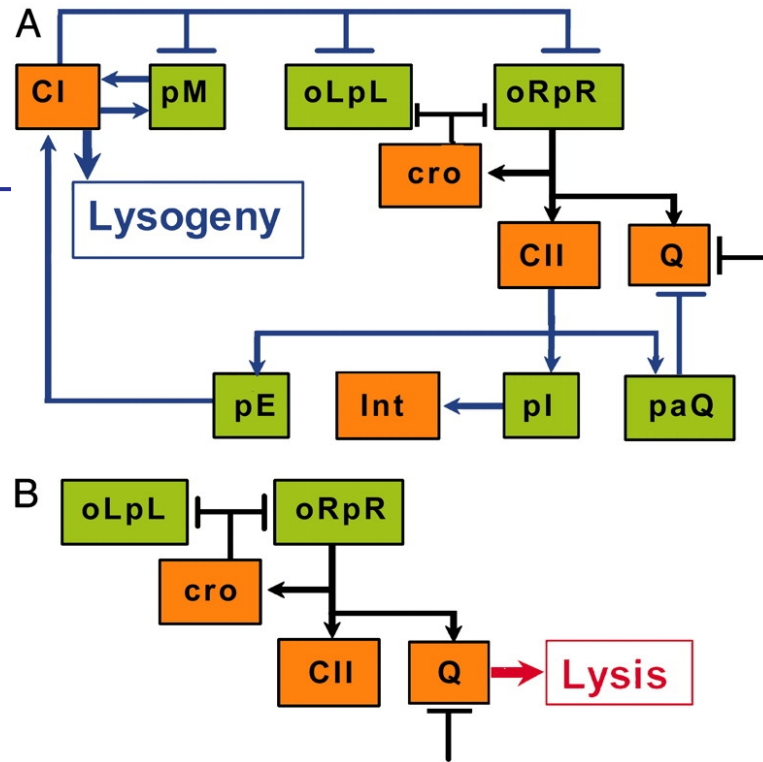
CII active CI et inhibe Q

- Délai entre activation de CII et de Q



Kobiler, Oren et al. (2005) Proc. Natl. Acad. Sci. USA 102, 4470-4475

Décision lyse-lysogénie

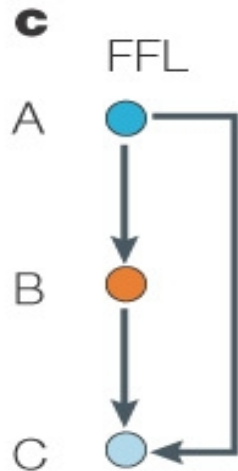


Trop peu de CII, donc lyse

Caractéristiques de réseaux de régulation transcriptionnelle

□ Le motifs de connexion

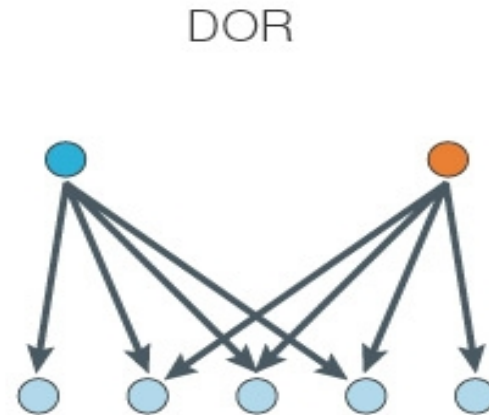
Feed forward loop



Single input module

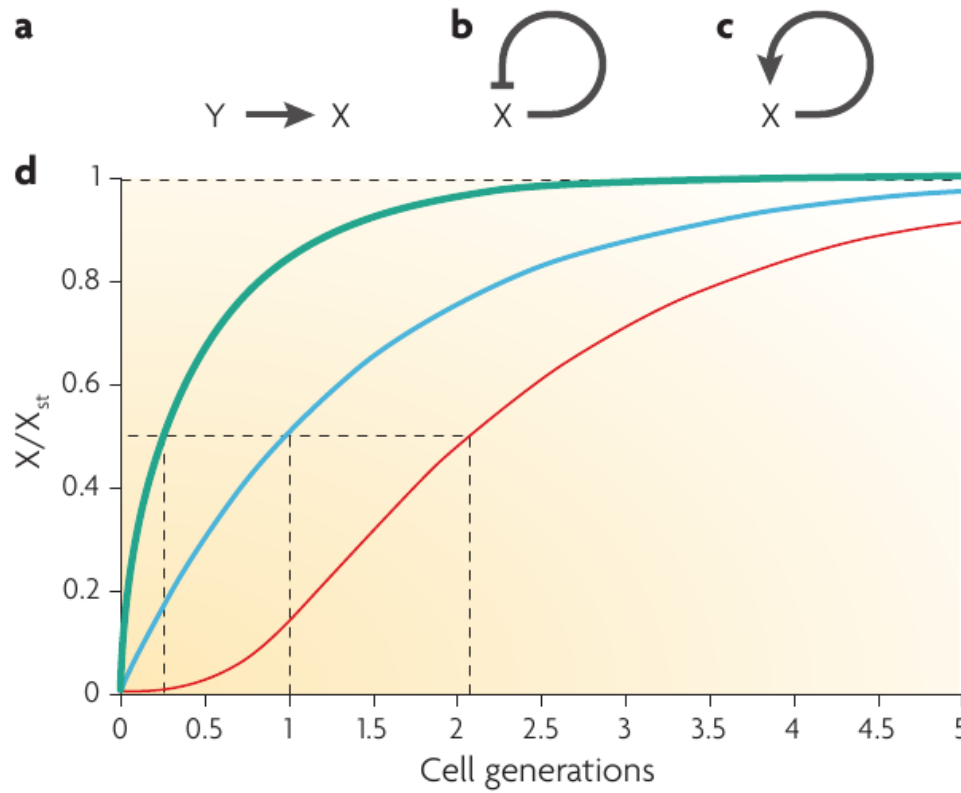


Dense overlapping regulon



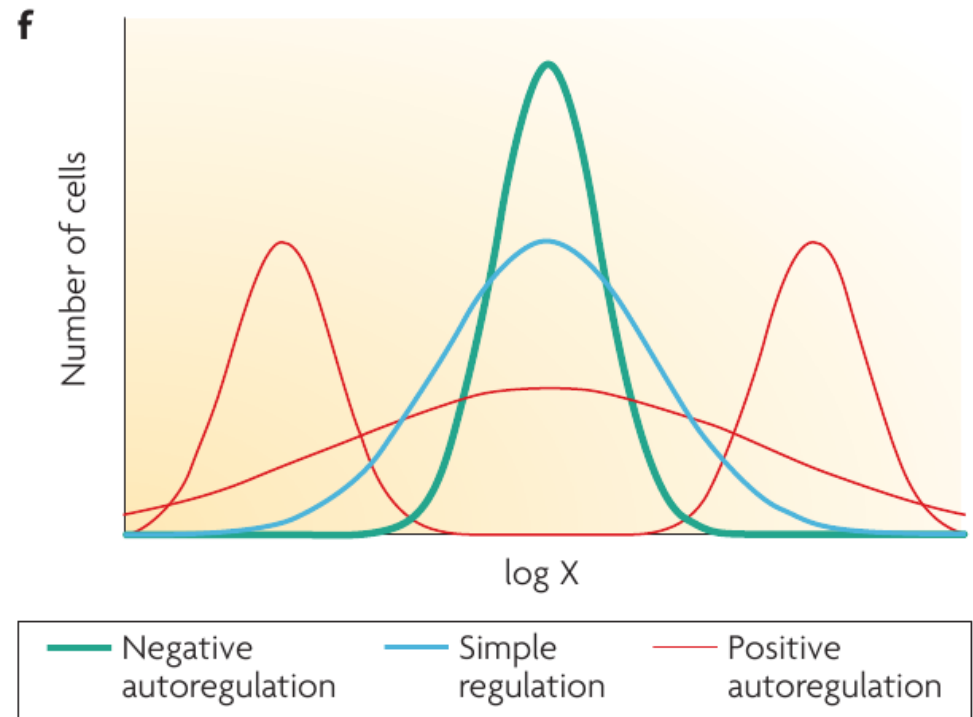
Motifs de régulation simples

□ La vitesse de réponse dépend de la connectivité



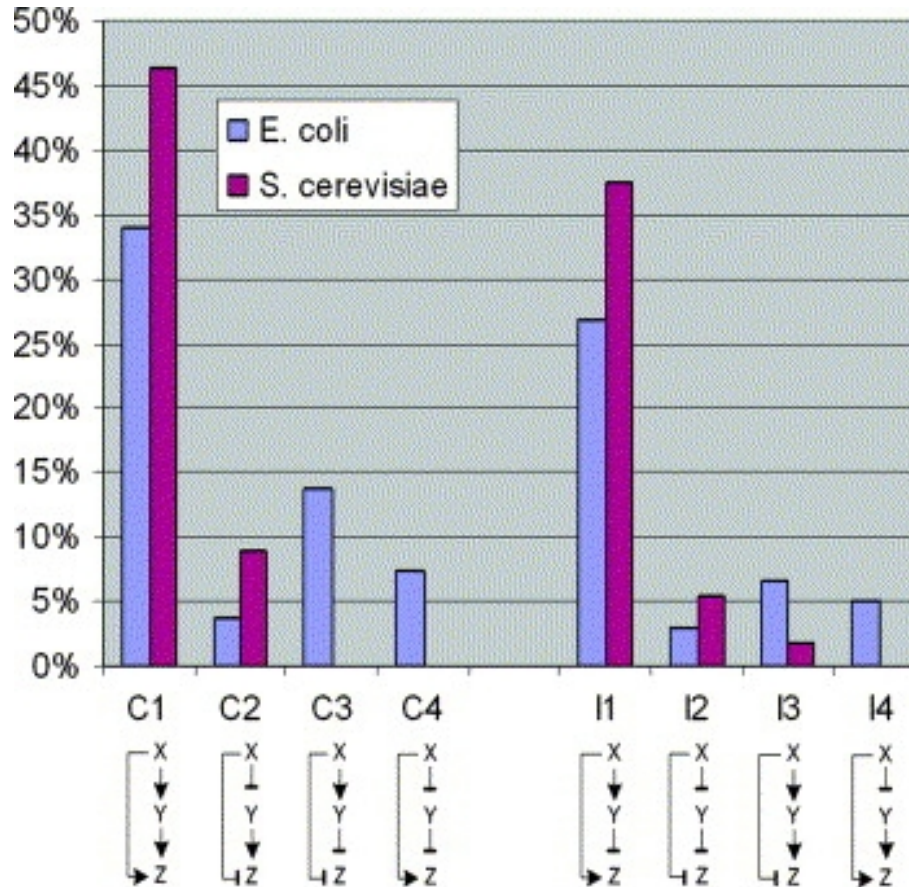
Distribution des régulateurs

- La régulation négative produit des distributions étroites
- La régulation positive produit des distributions large (même bi-modales)



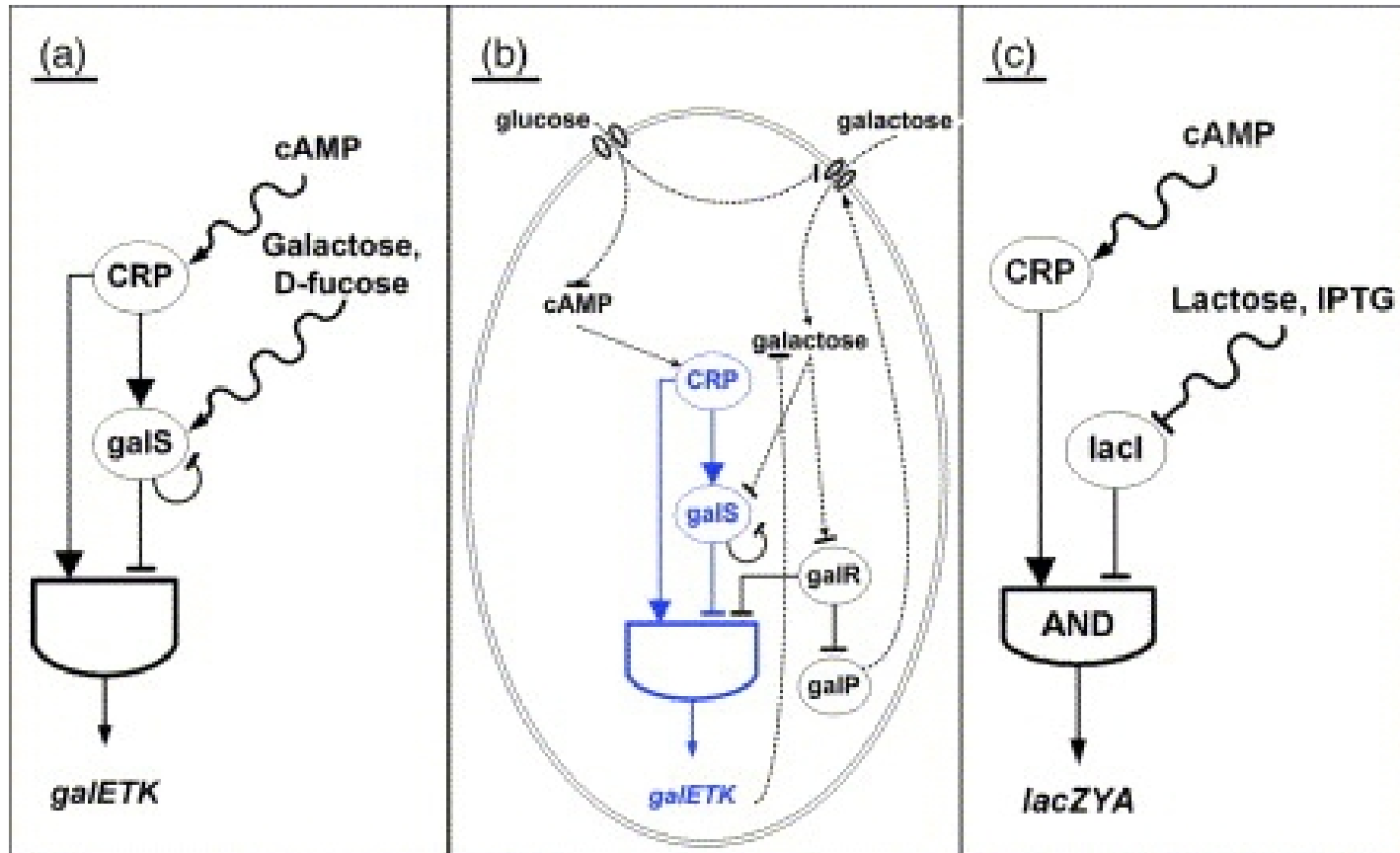
Les huit types de FFL

□ Distribution similaire chez *E. coli* et *S. cerevisiae*



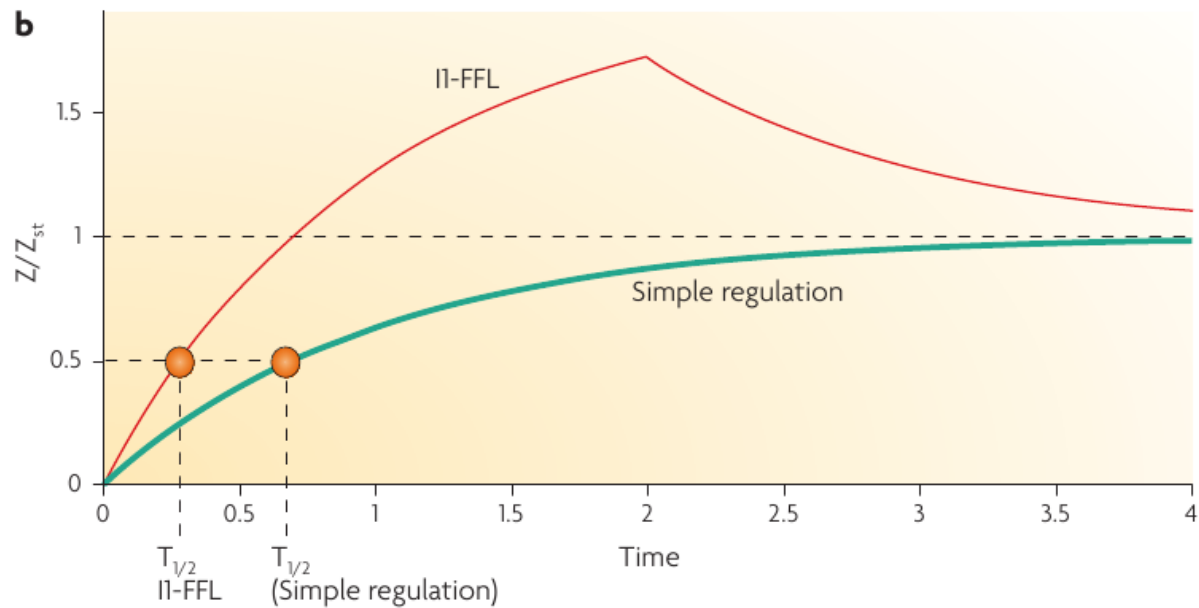
The galactose and lactose systems of E. coli

□ Les FFL



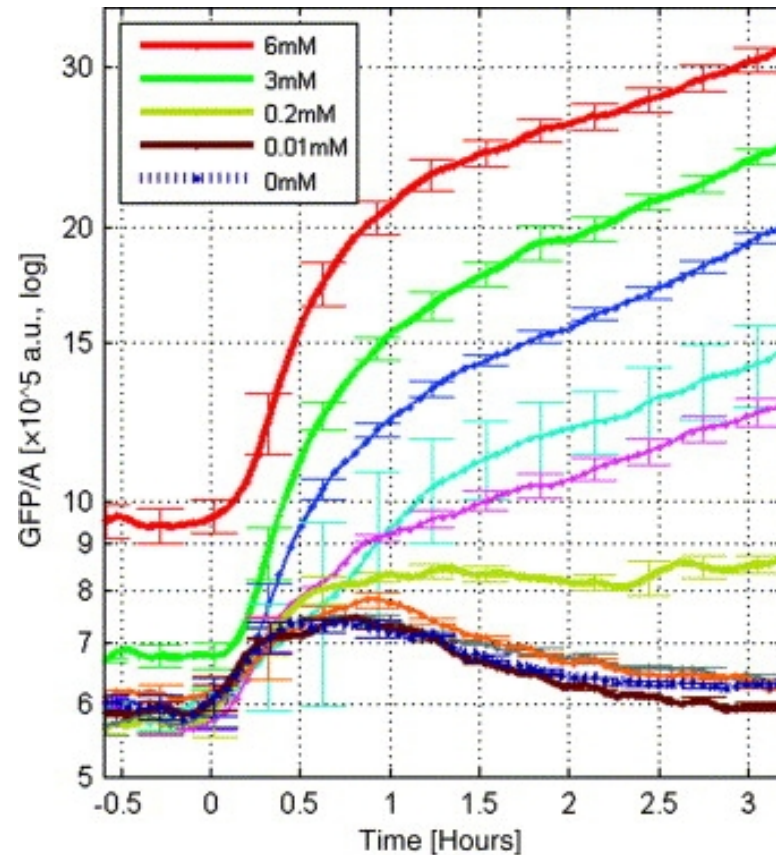
Feed forward loop

□ Réponse rapide (théorie)



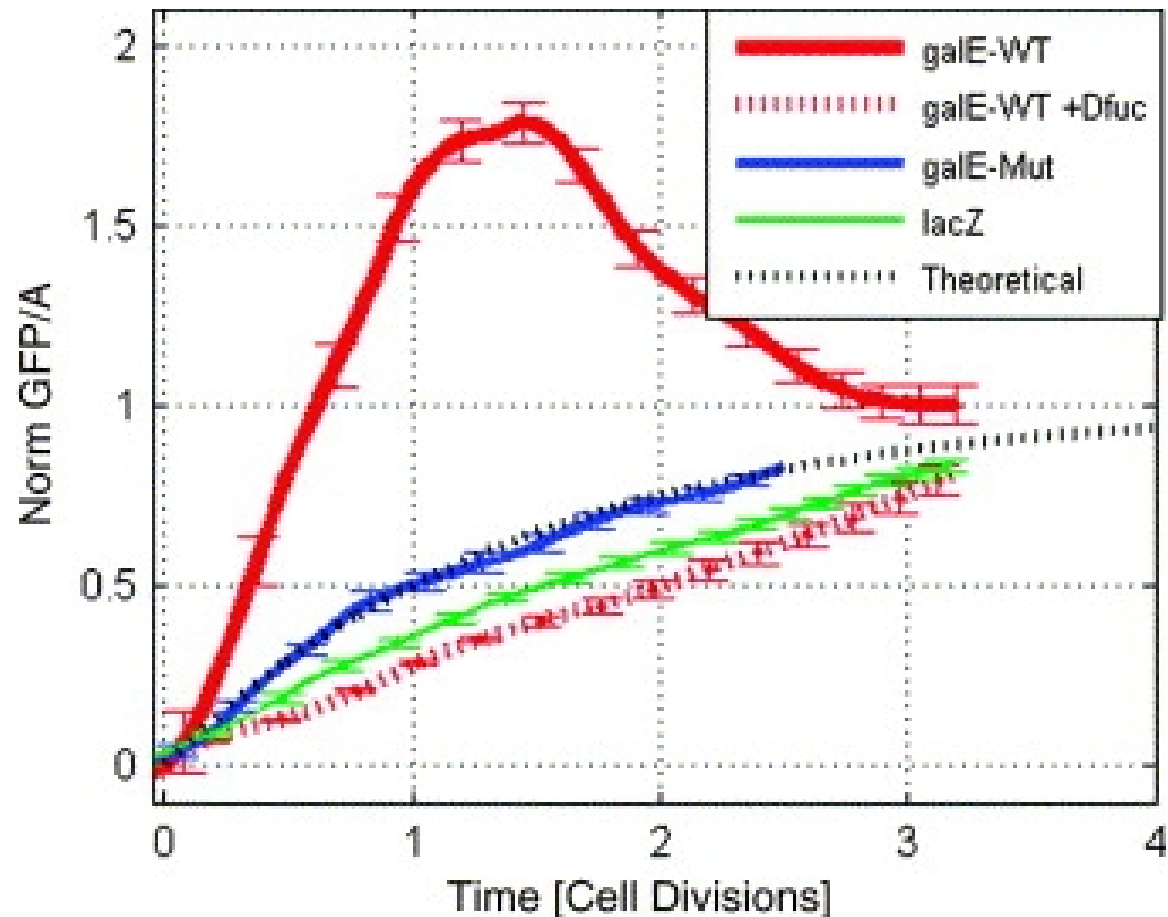
Dynamique du promoteur *galE*

- Induction du promoteur *galE* par épuisement de glucose à temps zéro
- Différentes concentrations de l'inducteur non-métabolisable, d-fucose



Expression du promoteur galE par rapport au steady state

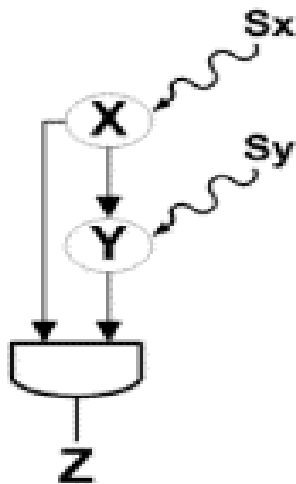
□ Le glucose est épuisé à temps zéro



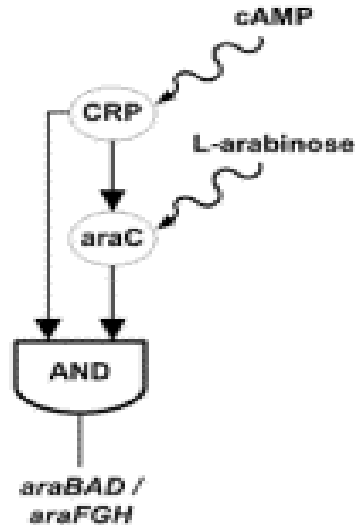
Les FFL cohérents

□ Le FFL cohérent avec ces inducteurs (ara)

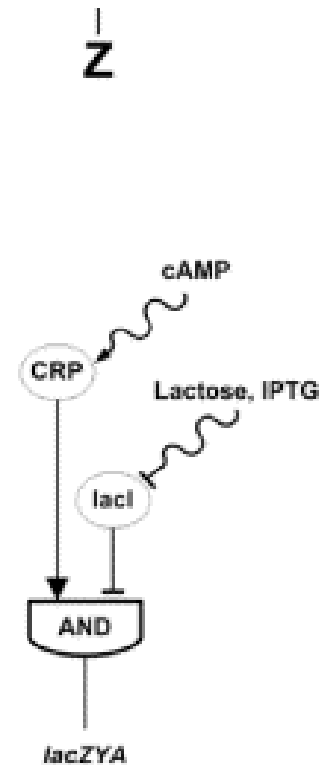
a)



b)



c)



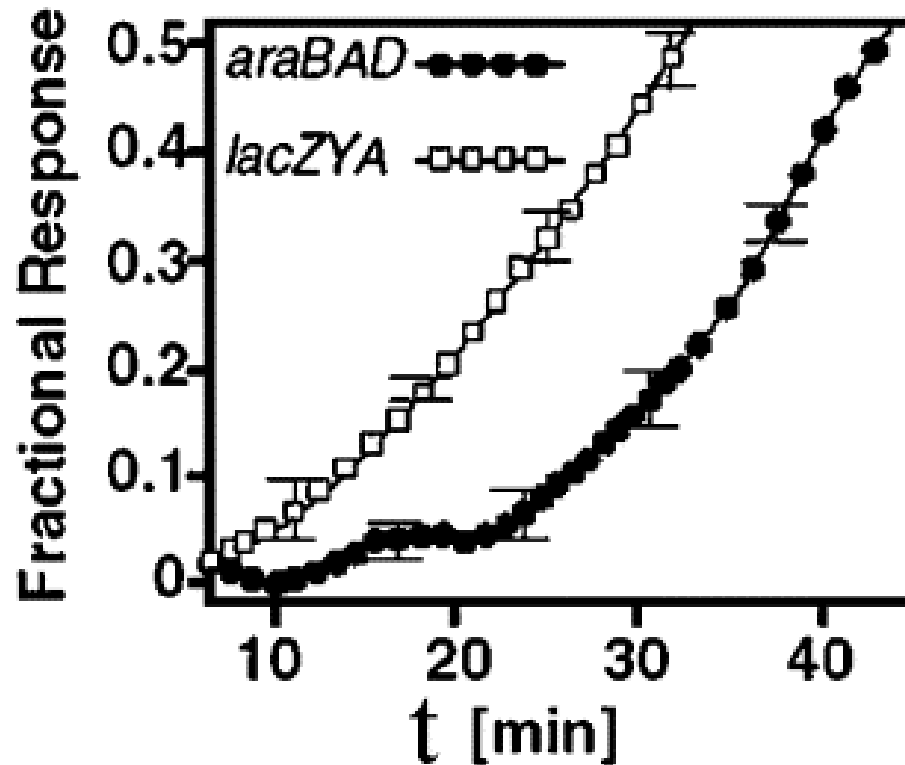
d)

c)

d)

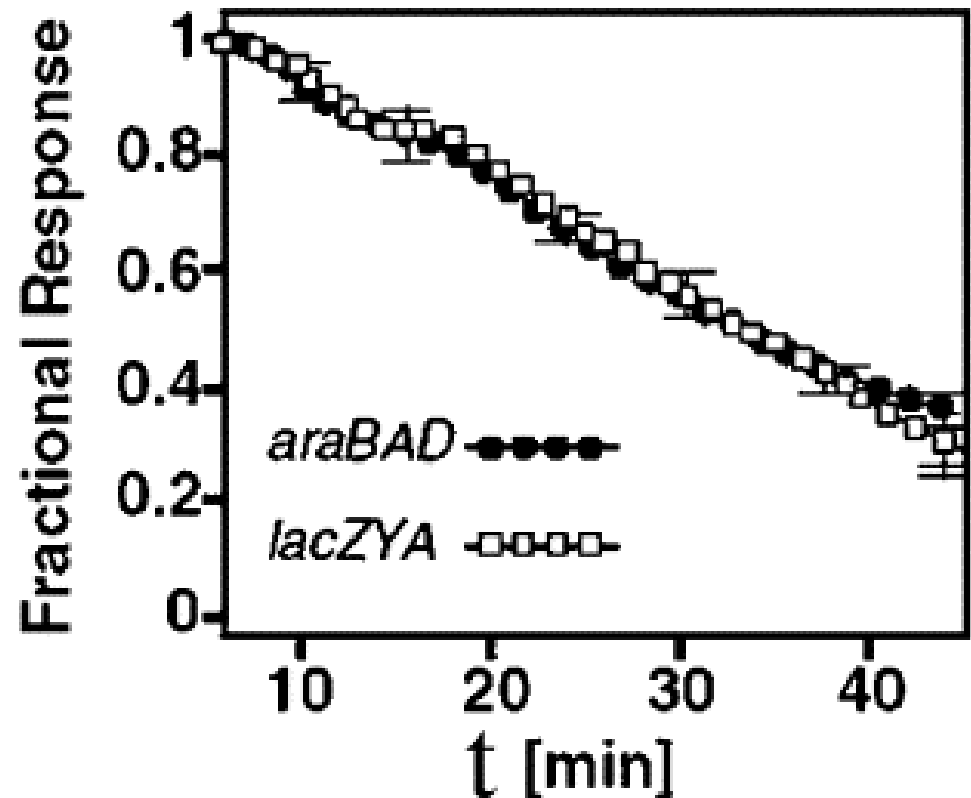
Mesure de l'activité par la GFP

- 10 mM AMPc sont ajoutés à temps zéro
- Délai d'activation pour le FFL cohérent



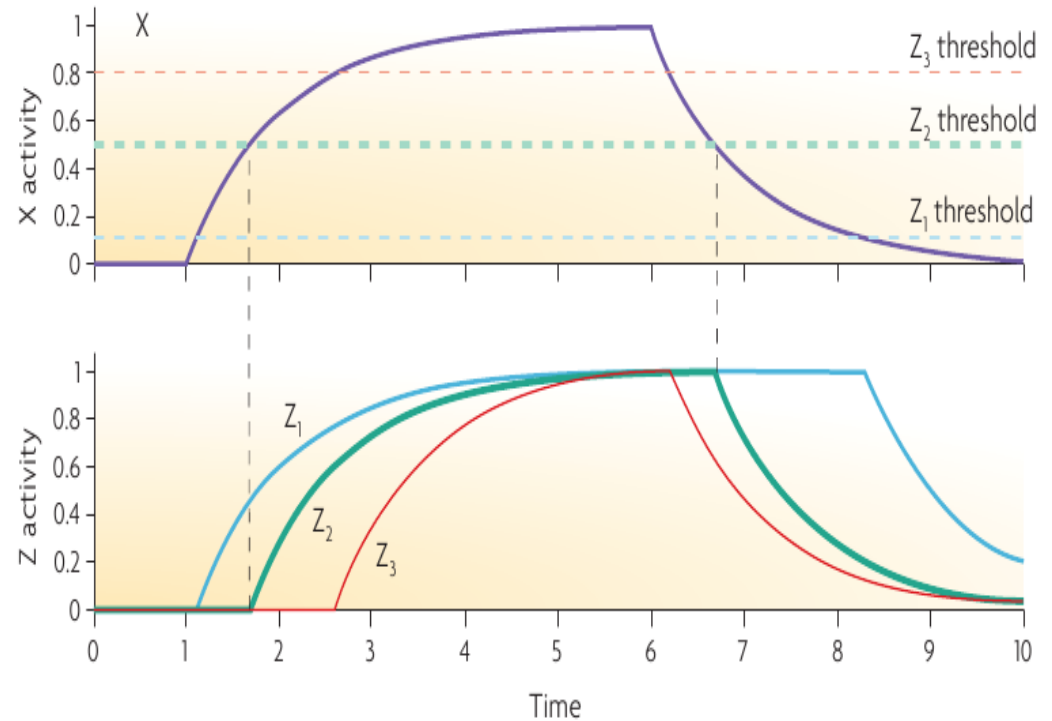
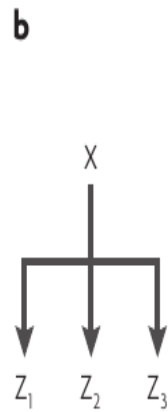
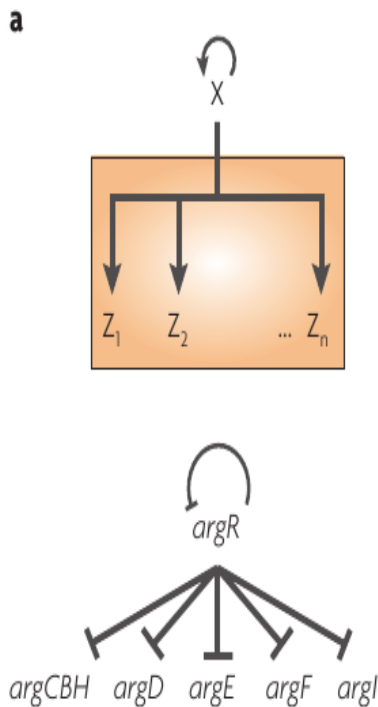
Enlever l'AMPc

- Les bactéries poussent sur glycérol et glucose est rajouté à temps zéro
- Pas de délai pour arrêt de l'induction



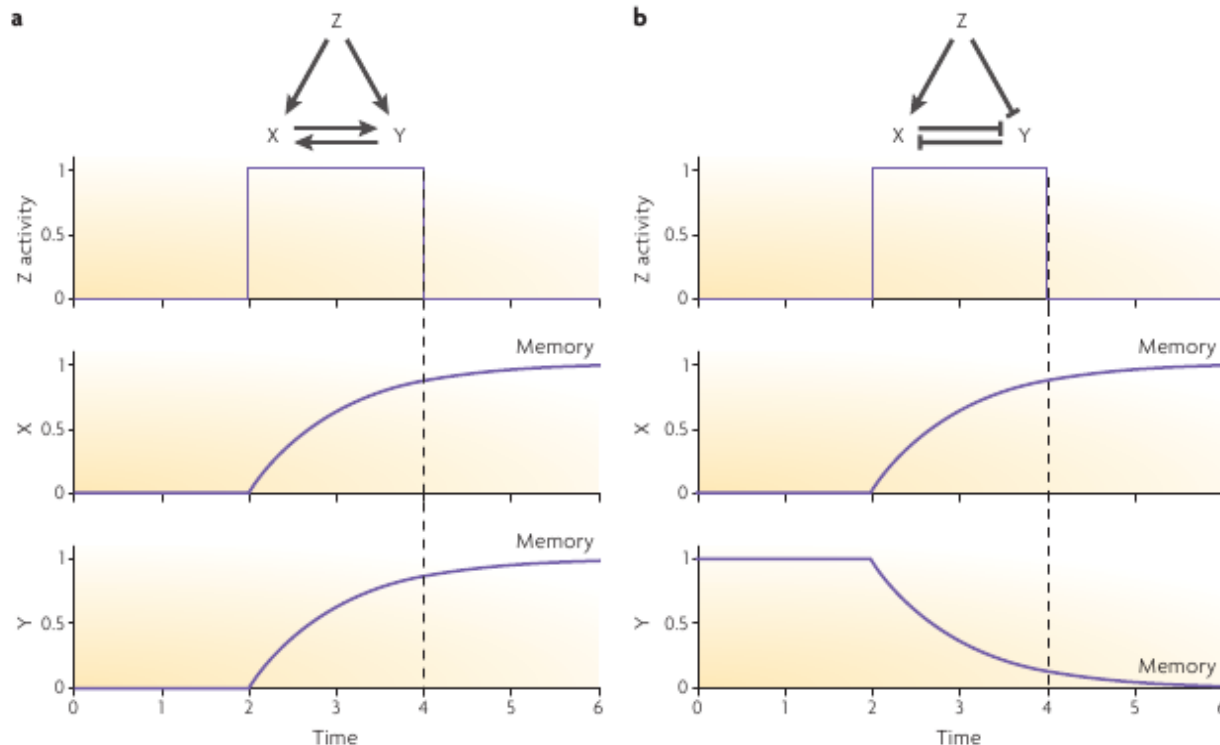
Single Input Module

□ Activation séquentielle



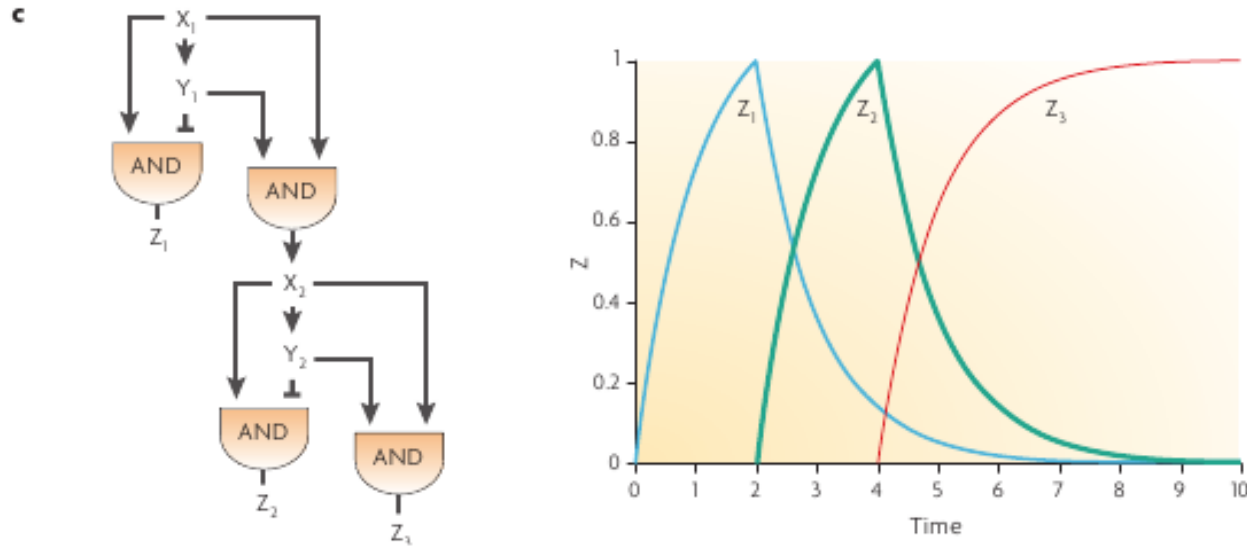
Les réseaux de régulation dans le développement

□ Un signal initial est mémorisé

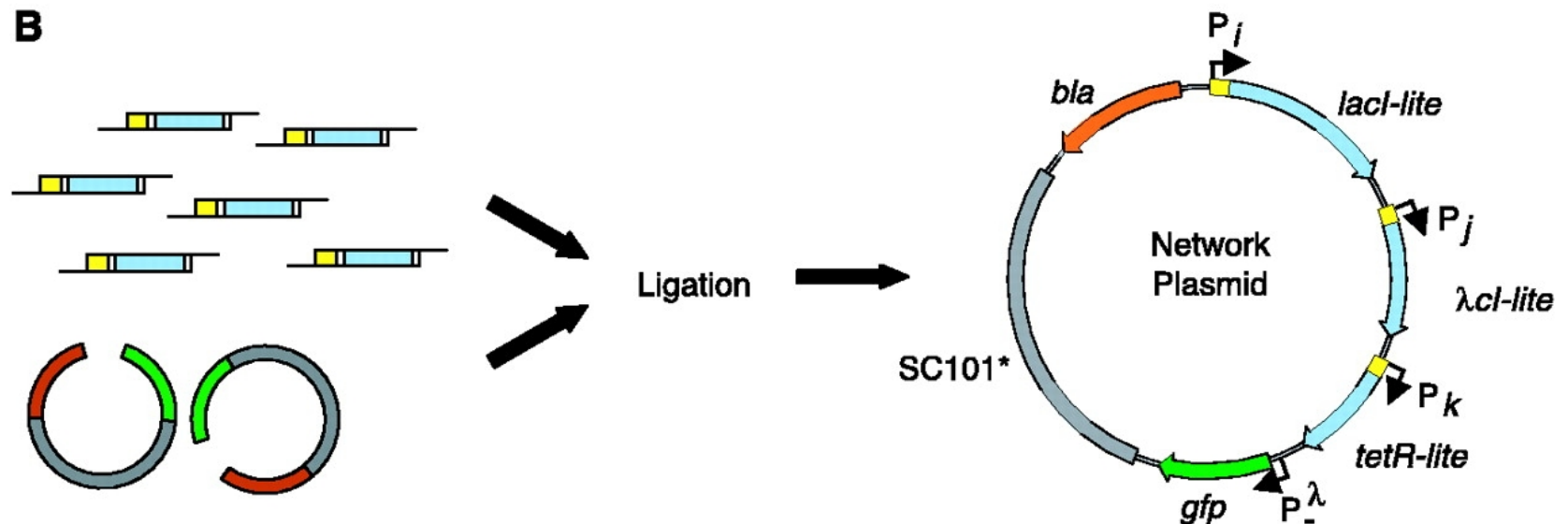
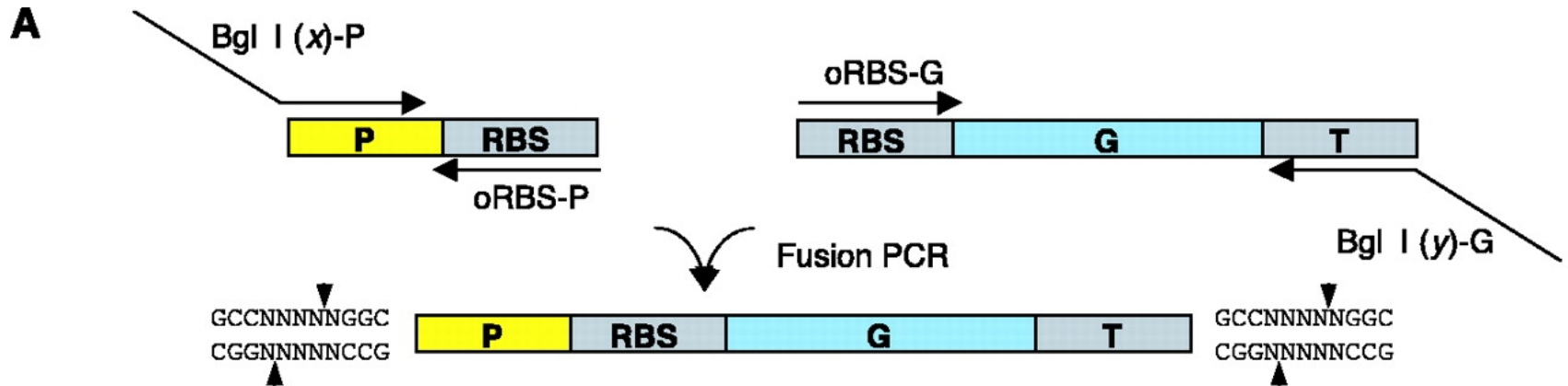


Les réseaux de régulation dans le développement

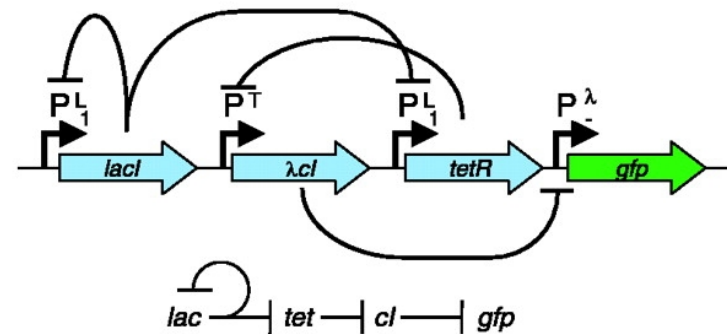
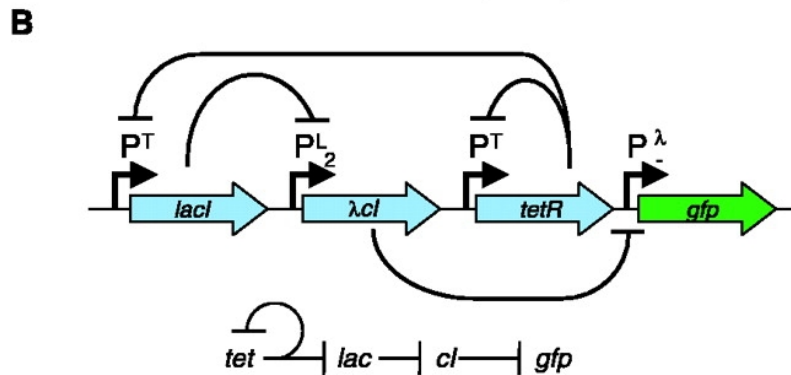
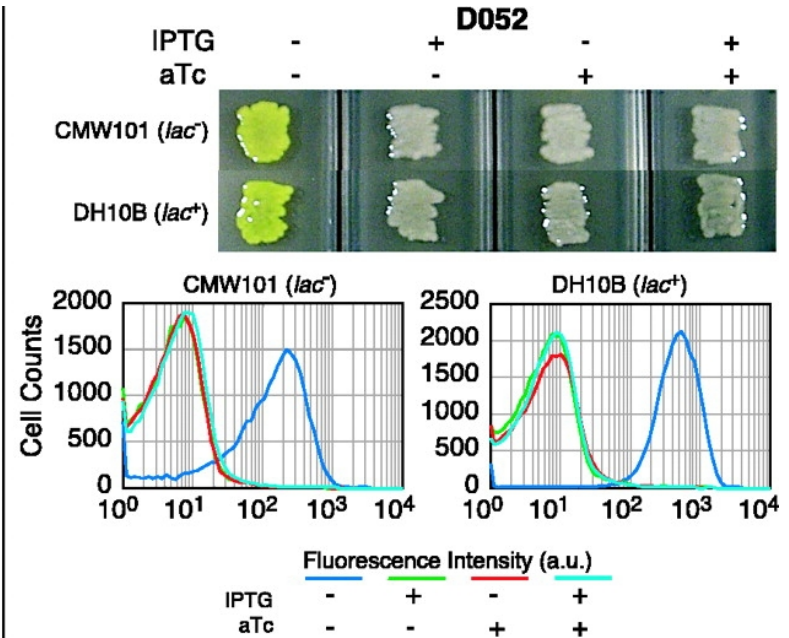
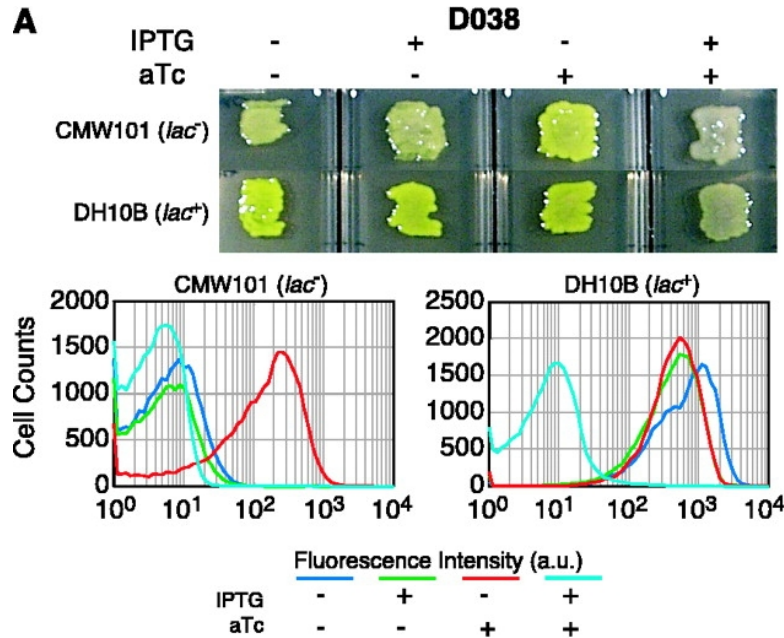
- Les connectivités plus compliquées conduisent à une activation séquentielle



Combinatorial construction of gene regulatory networks



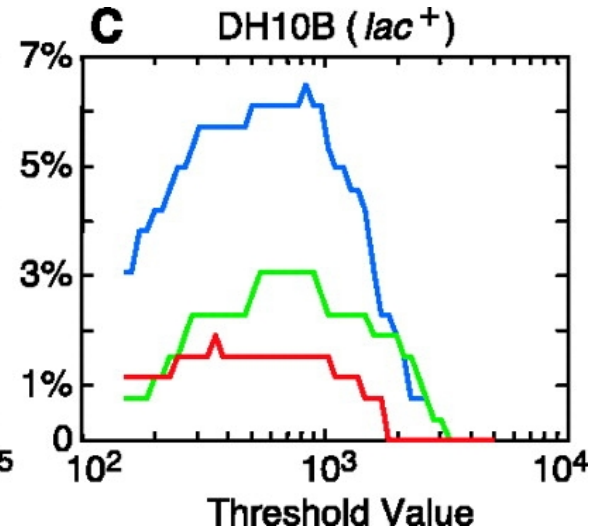
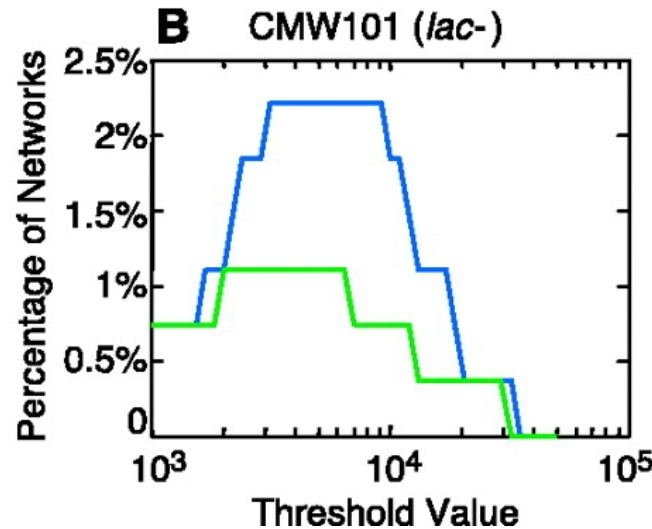
Two examples of logical circuits: NOR NAND



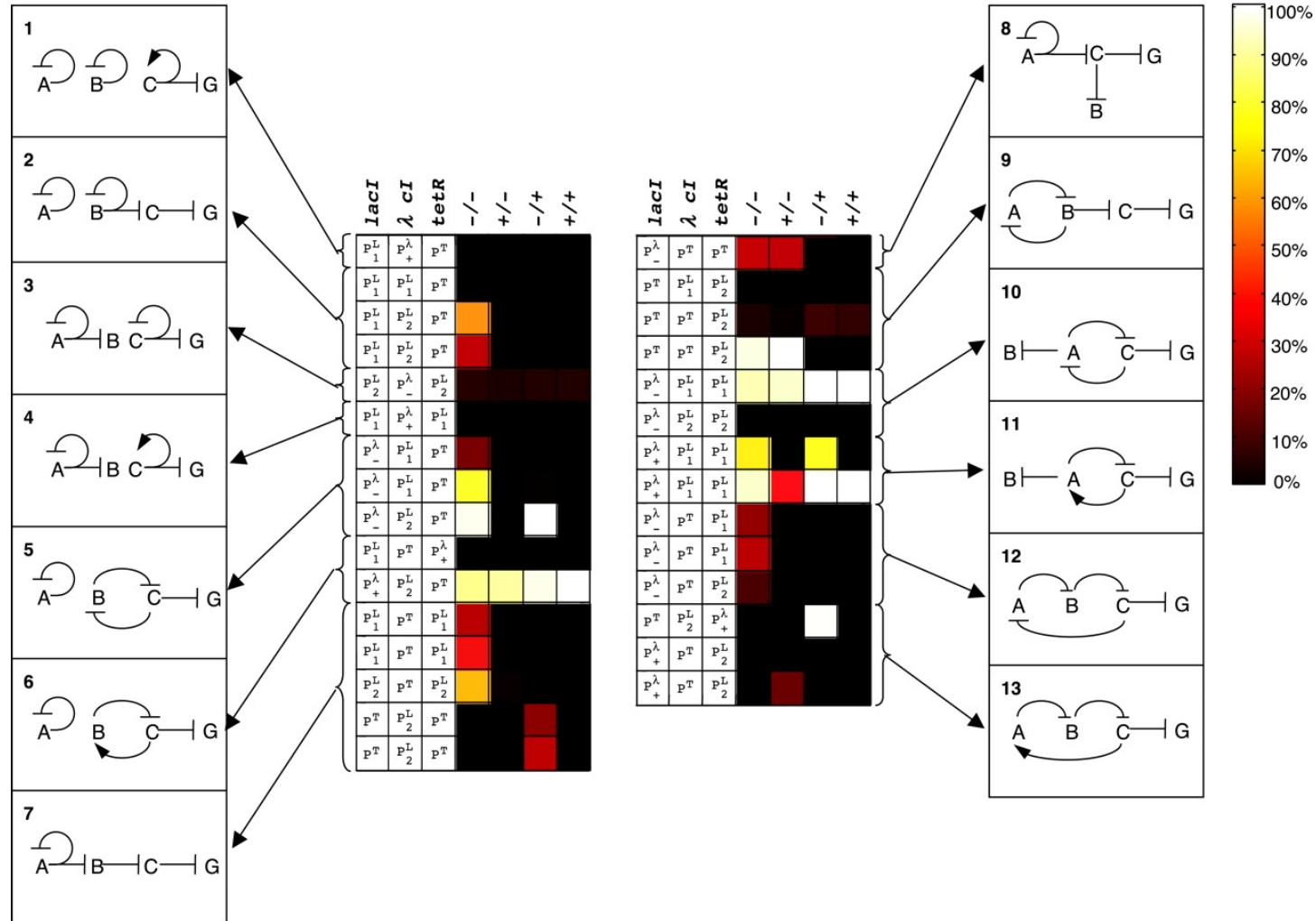
Proportion of binary logical networks in the two different cell types

A

	IPTG	-	+	-	+
	aTc	-	-	+	+
—	NOR	On	Off	Off	Off
—	NOT IF	Off	On	Off	Off
—	NAND	On	On	On	Off

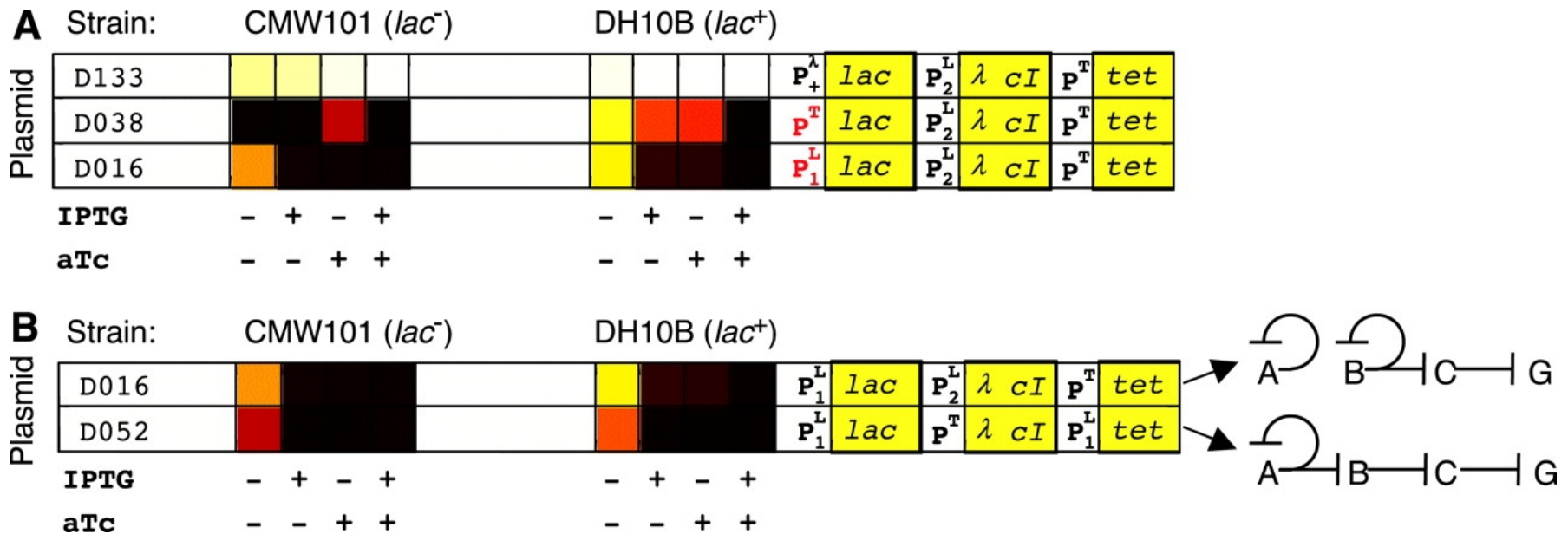


Dependence of phenotypic behavior on network connectivity

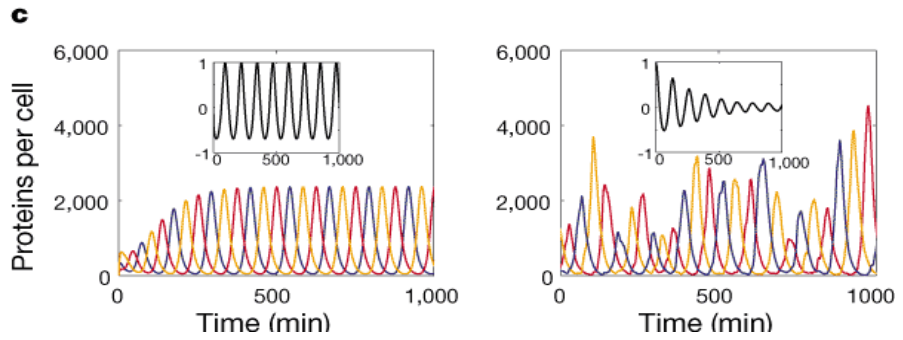
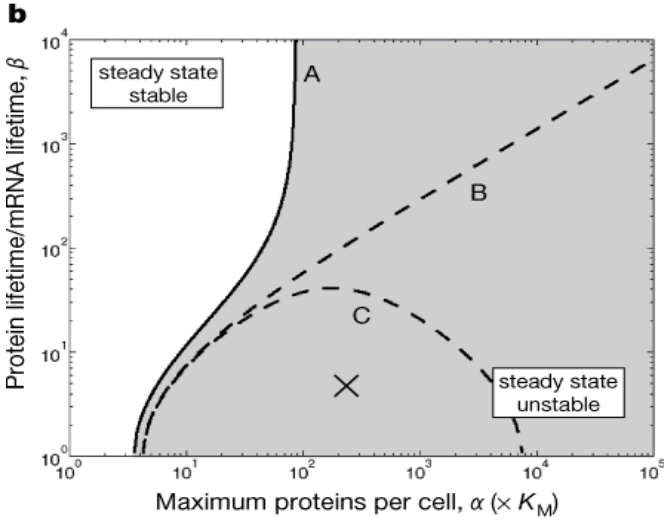
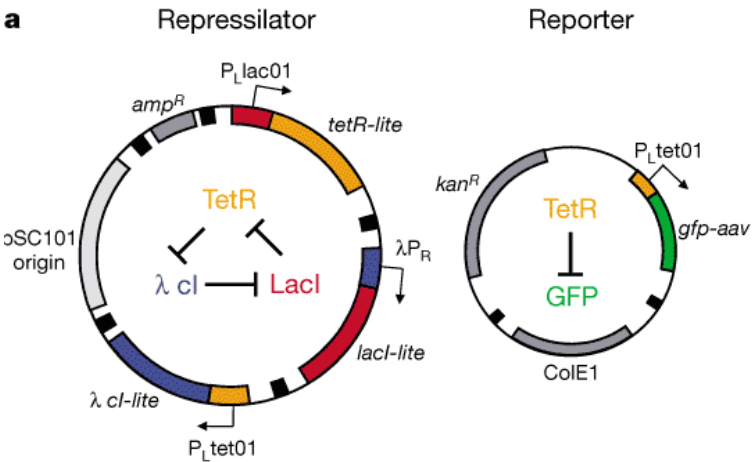


Comportement dynamique des réseaux

- Small changes in the network lead to very different behaviors
- The same behavior can be achieved by different networks



A synthetic oscillator



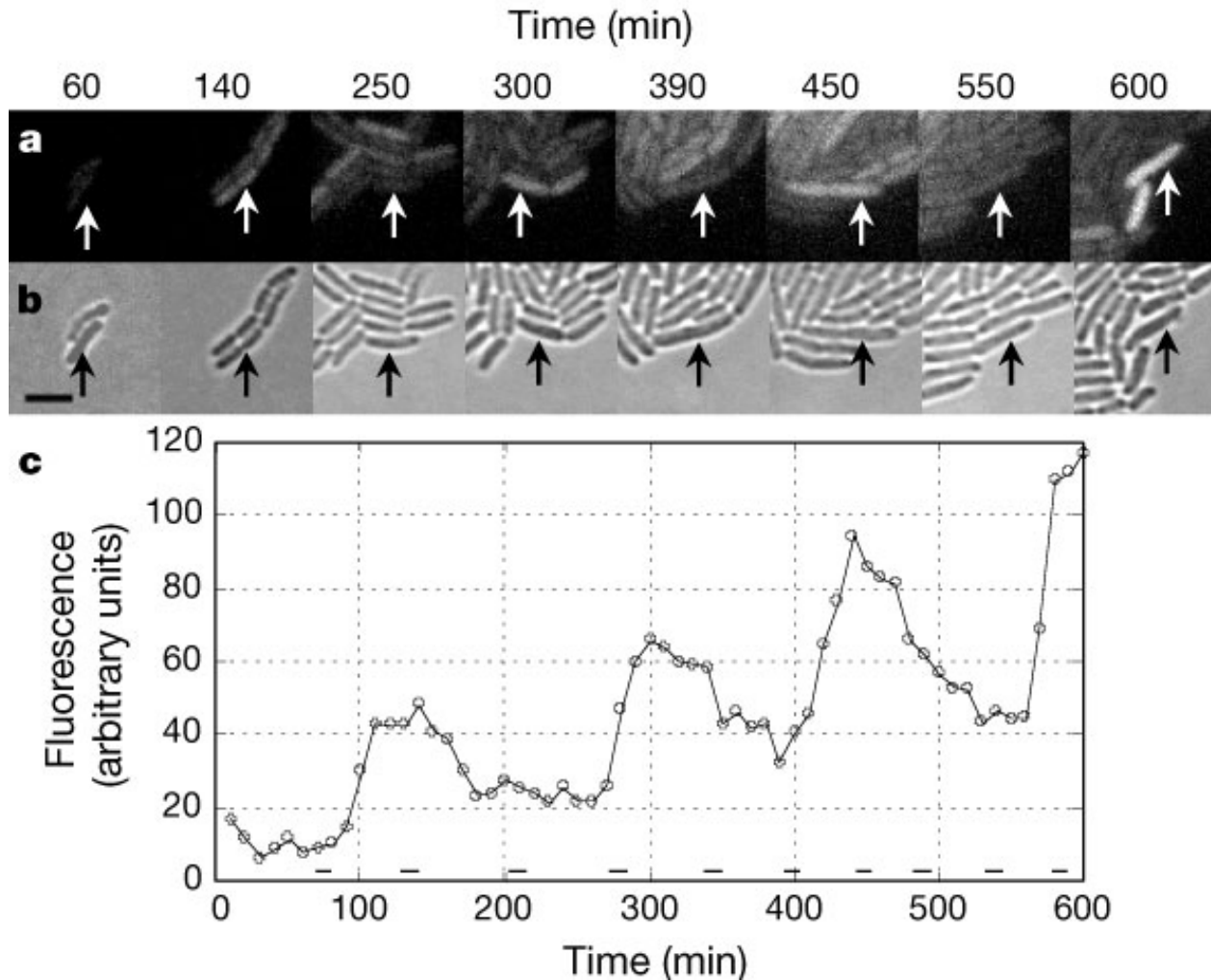
Synthetic oscillator

$$\frac{dm_i}{dt} = -m_i + \frac{\alpha}{1 + p_j^n} + \alpha_0$$

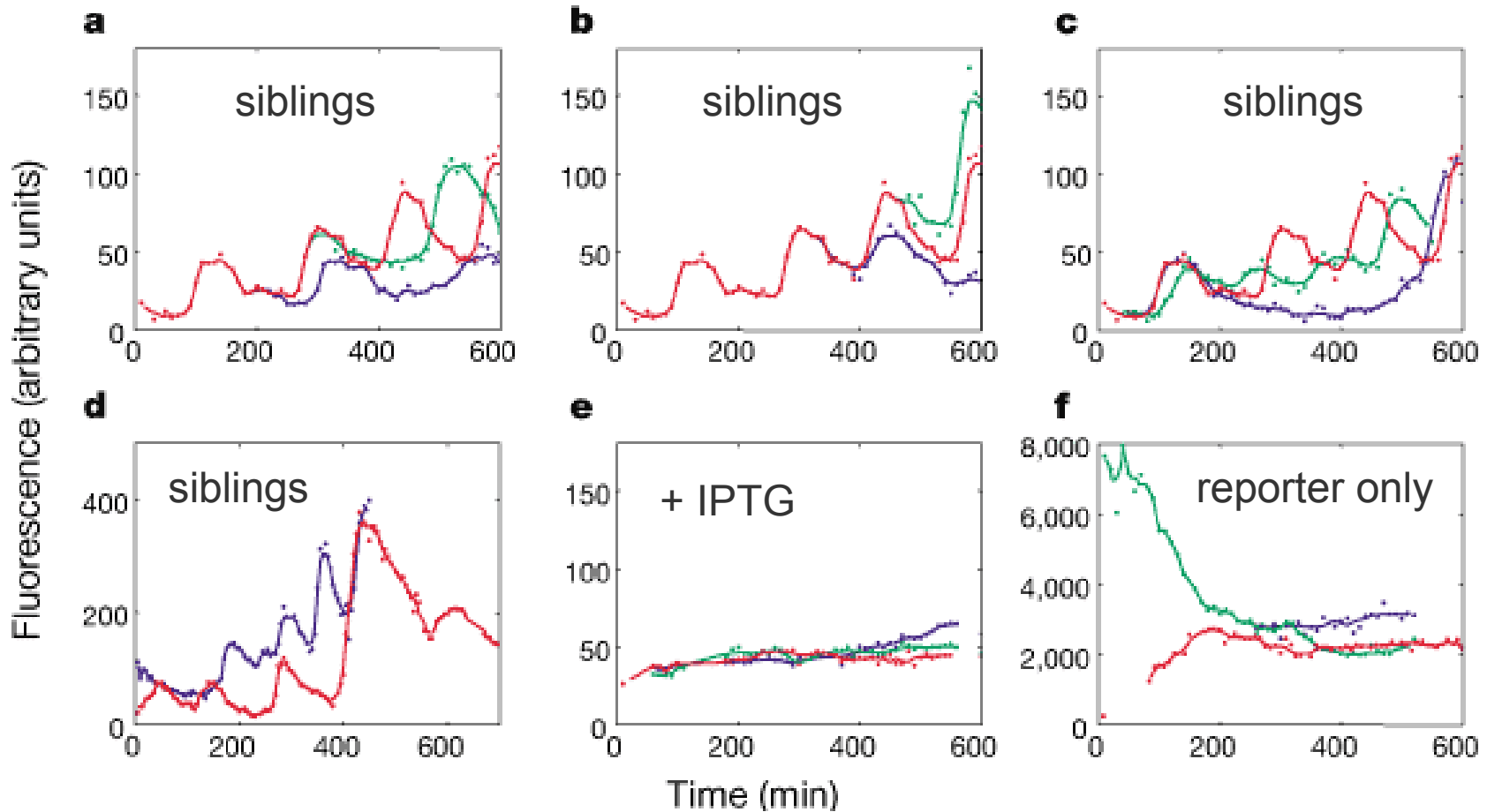
$$\frac{dp_i}{dt} = -\beta \times (p_i - m_i)$$

$$\left(\begin{array}{l} i = lacI, tetR, cl \\ j = cl, lacI, tetR \end{array} \right)$$

Oscillations are maintained through cell division

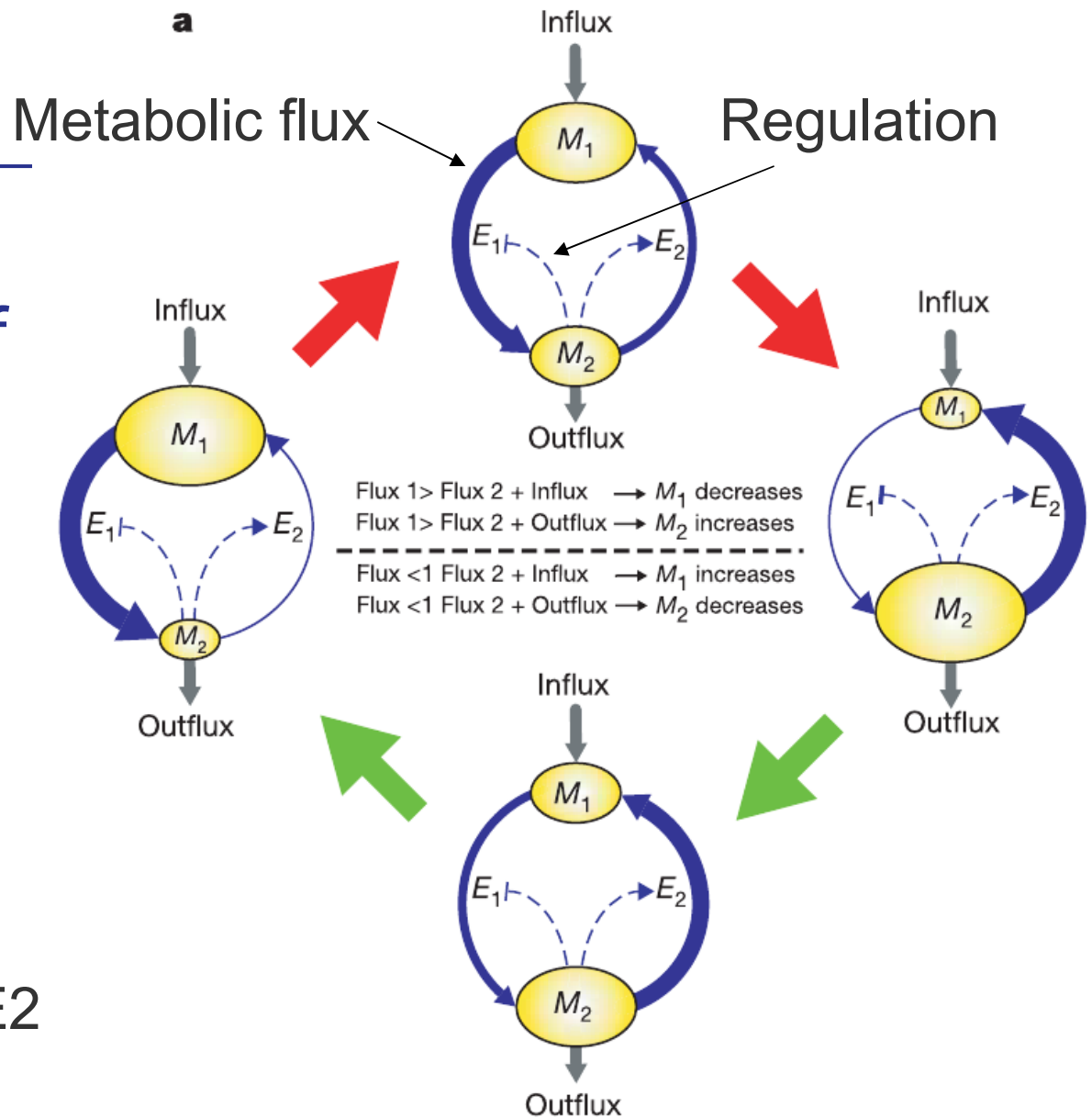


Examples of oscillatory behavior



Construction of a metabolic oscillator

Two metabolites, M1 and M2, and their control by enzymes E1 and E2



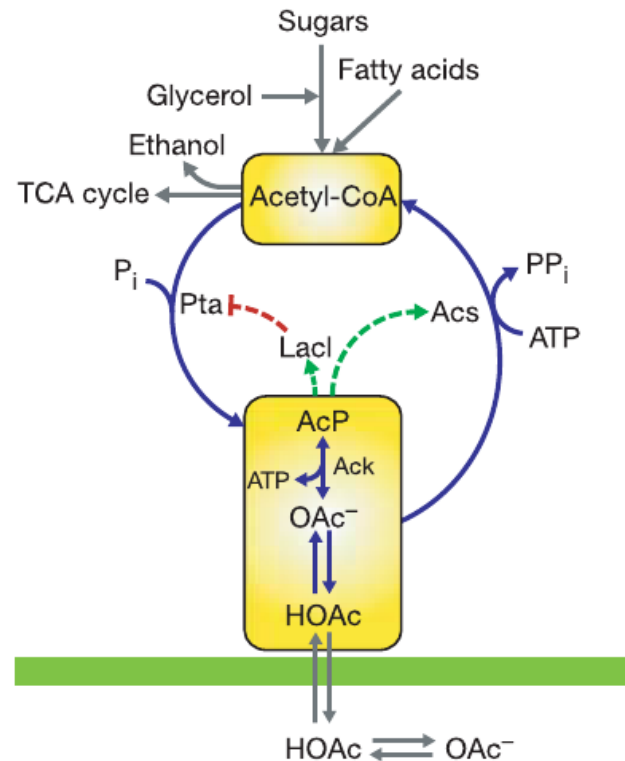
Construction in Escherichia coli

- Use acetyl-CoA and acetyl-phosphate as metabolites

Pta: phosphate acetyltransferase

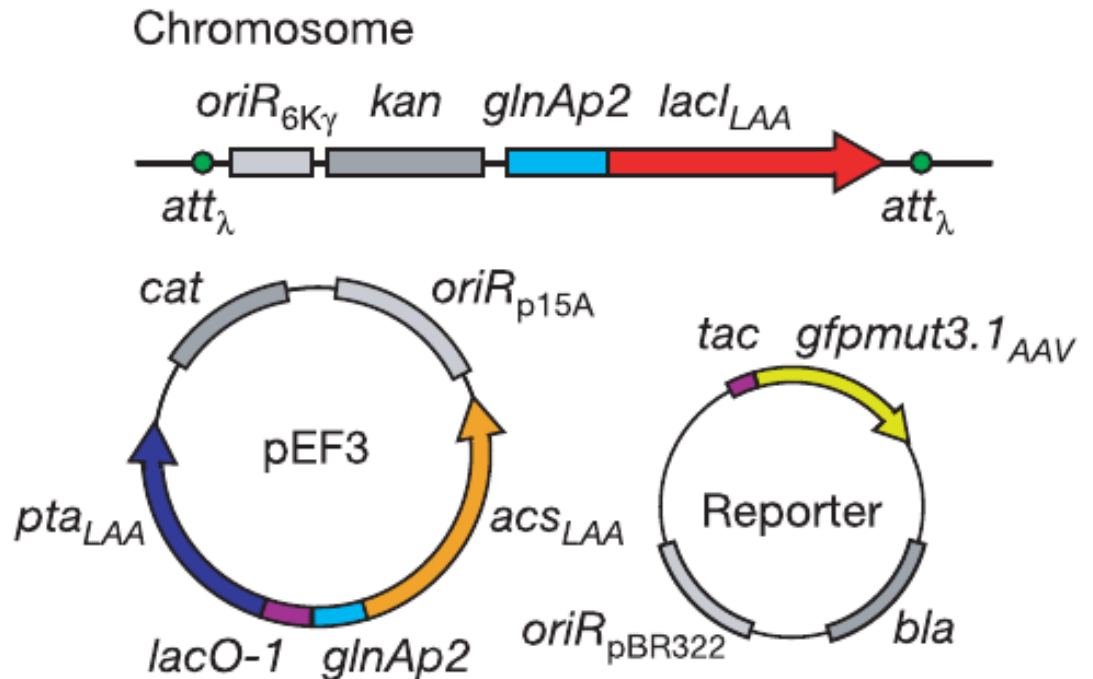
Acs: Acetyl-CoA synthetase

Ack: Acetate kinase



Change the regulation of the enzymes

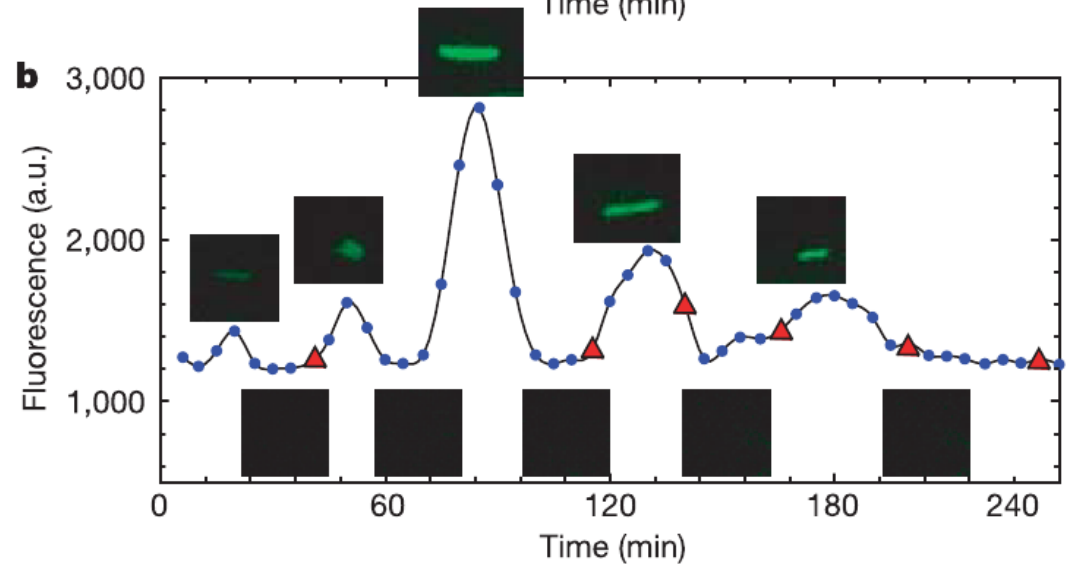
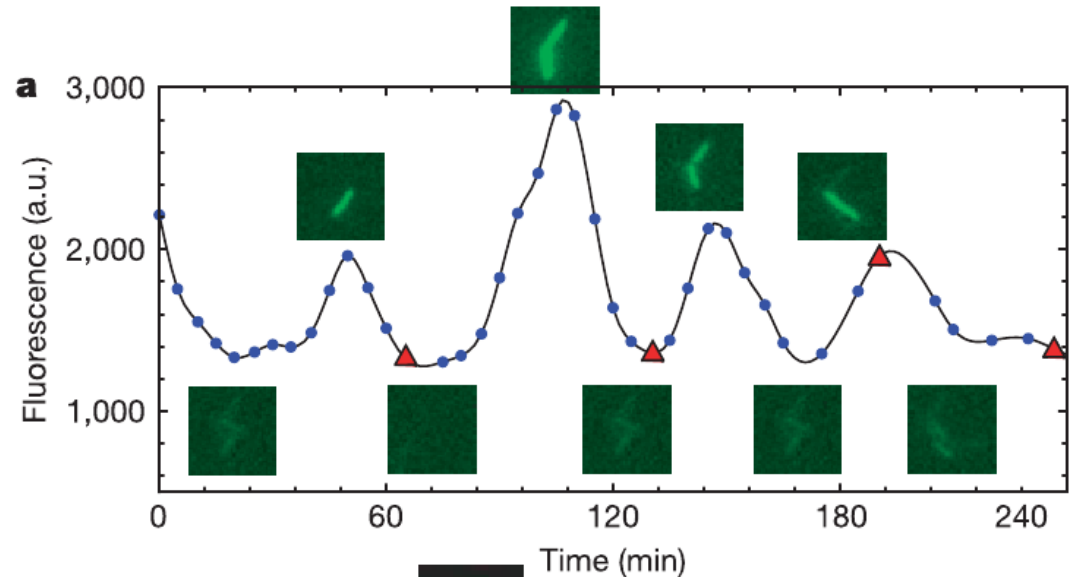
- **Ac-P**
phosphorylates NRI
and activates
glnAp2. Put ***acs***
downstream.
- **Pta** is expressed
from a lac promoter.
- **Lac repressor** is
under the control of
glnAp2.



Enzyme half-lives are reduced by an *ssrA*-tag

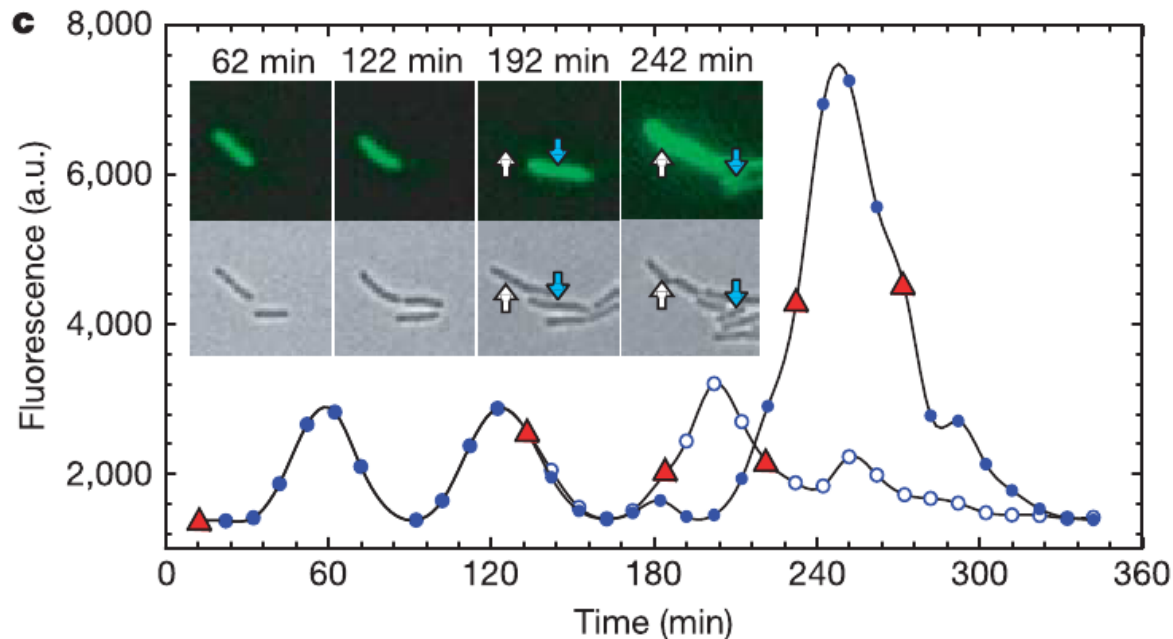
Oscillations in single cells

- Oscillation in 60% of cells
- Period: 40 min.
- Doubling time: 60 min.



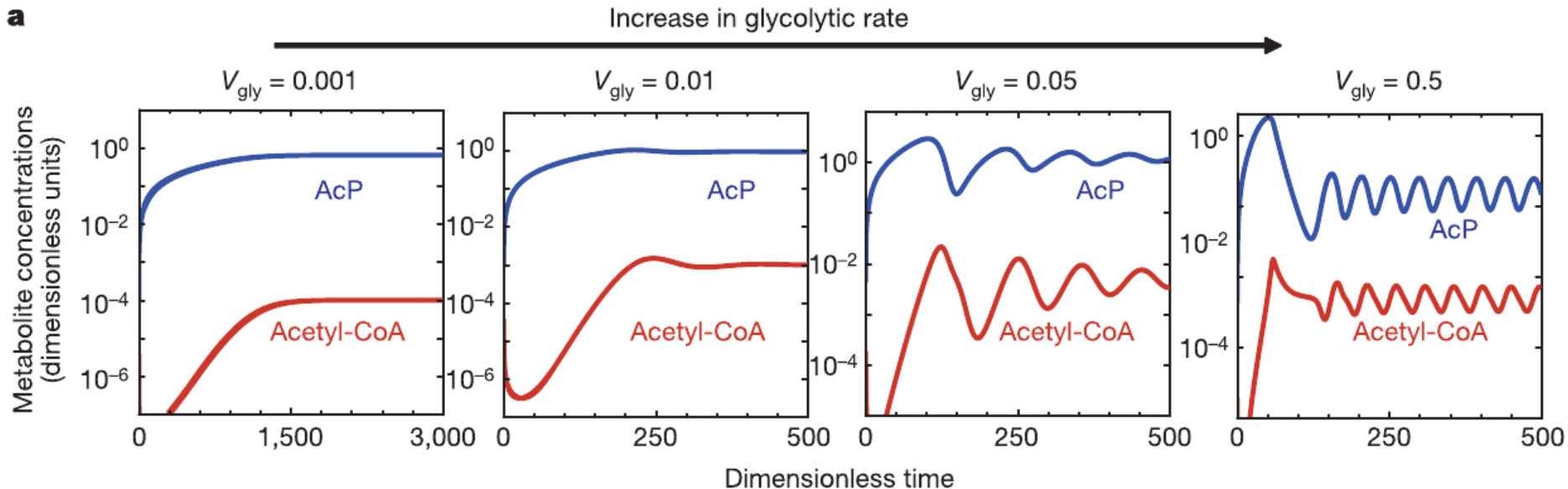
Oscillations in single cells

- The oscillation is transmitted (more or less faithfully) to daughter cells



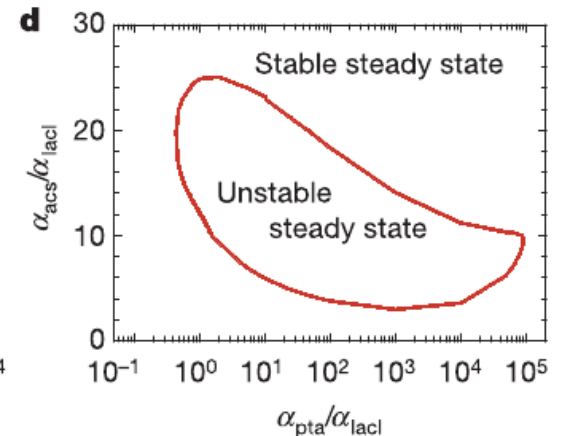
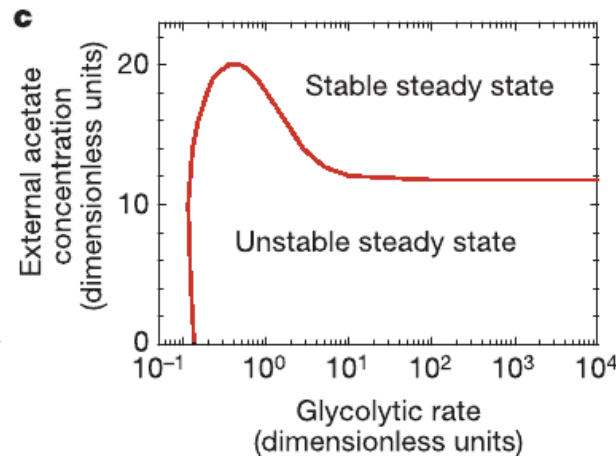
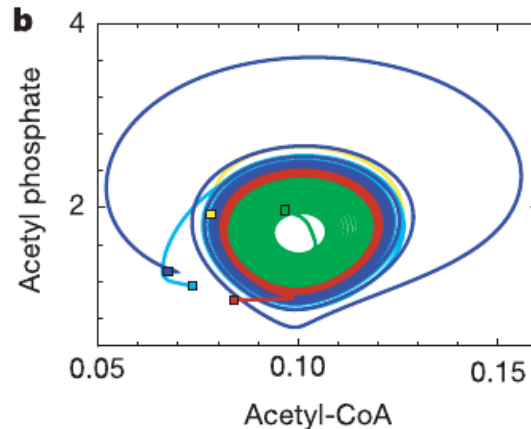
Modeling the oscillator

- Use Michaelis-Menten for the enzyme kinetics and Hill functions for the gene regulation
- Oscillations are predicted at high glucose flux



Modeling the oscillator

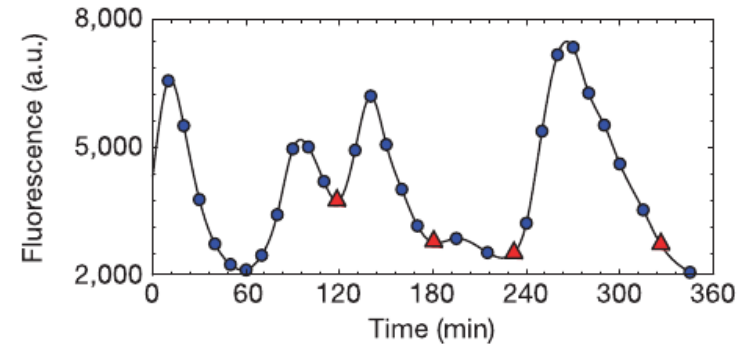
- The oscillator is a limit cycle
- External acetate stabilizes the system
- The enzyme concentrations have to be in the right range



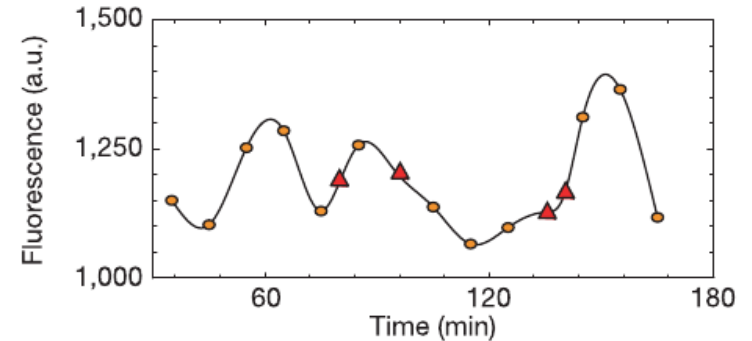
Test theoretical predictions

- Change the glycolytic flux by changing carbon source

fructose

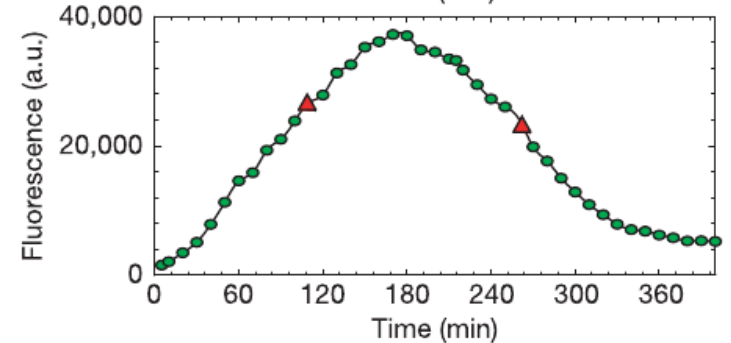


mannose



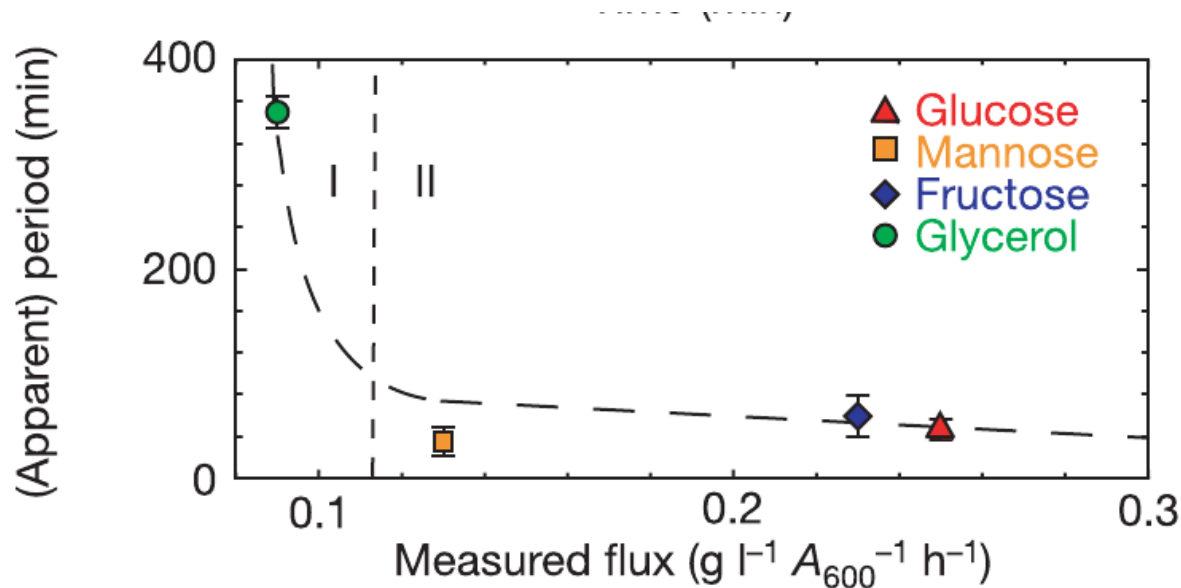
Red triangle:
cell division

glycerol



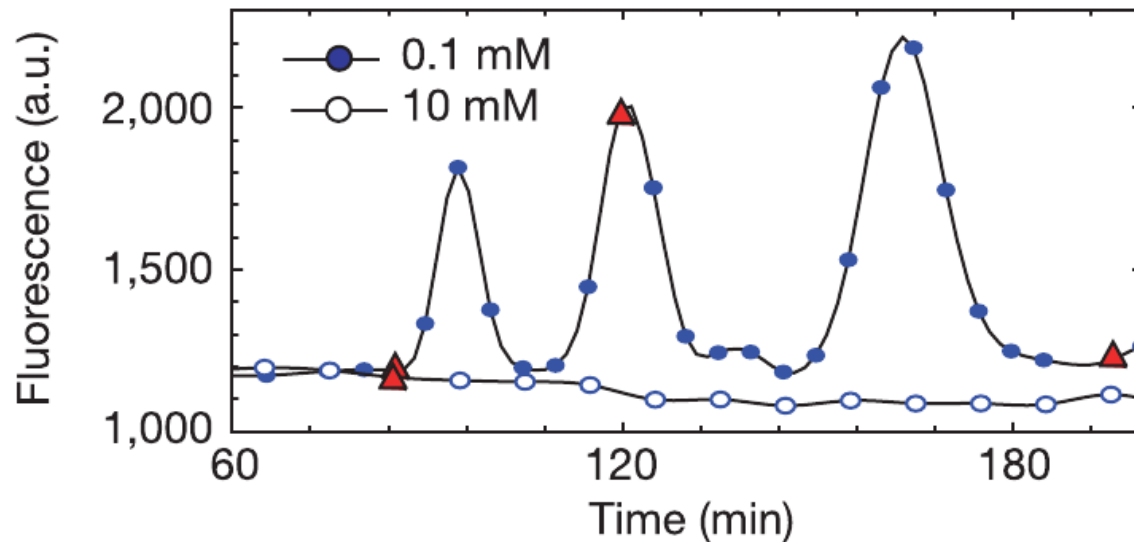
Test theoretical predictions

- Different carbon flux leads to different periods of the oscillation (apparent oscillation for glycerol)



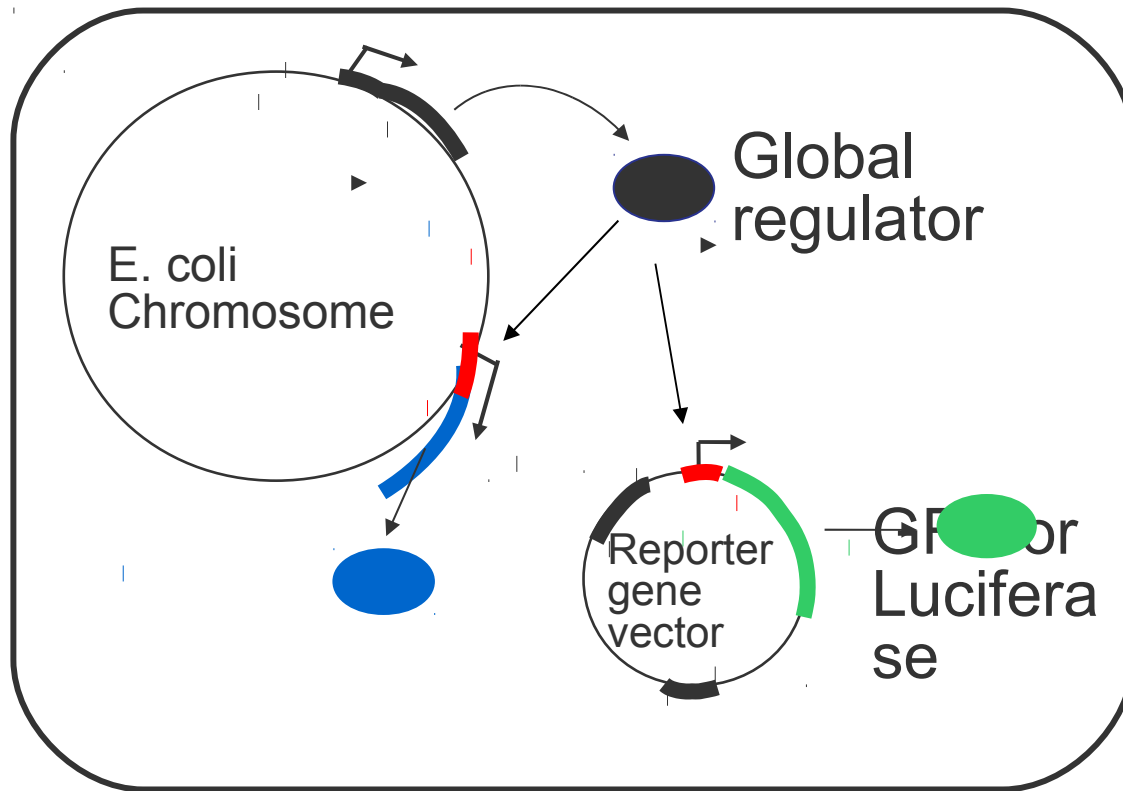
Test theoretical predictions

- External acetate suppresses the oscillation



Experimental validation of model predictions

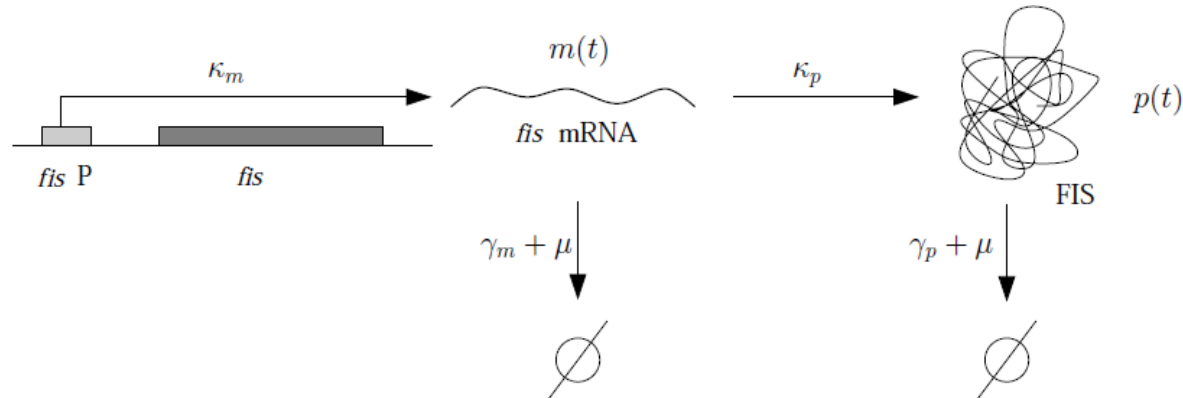
- Experimental validation of model predictions requires real-time monitoring of gene expression



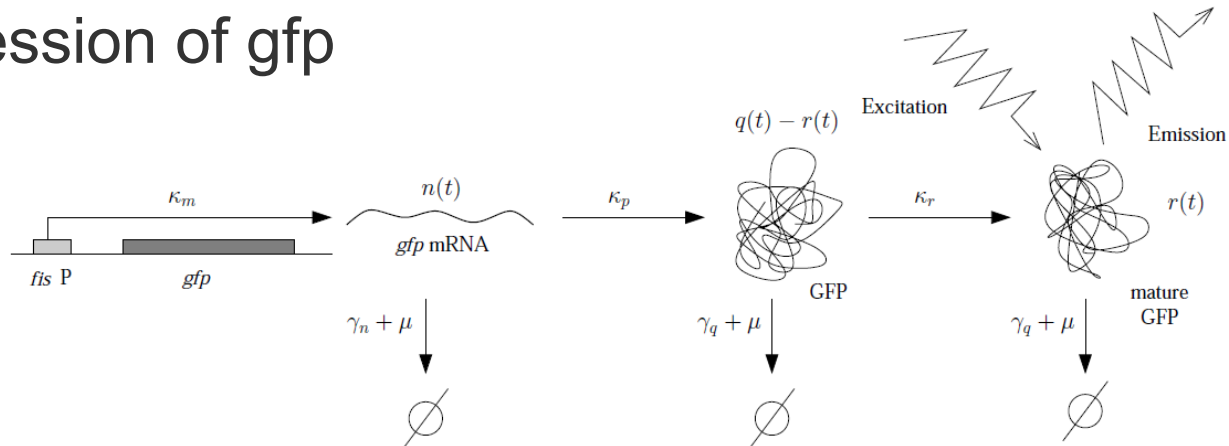
- The expression of the **reporter gene** reflects the expression of the **target gene**.

Host protein and reporter genes

- Expression of the host protein

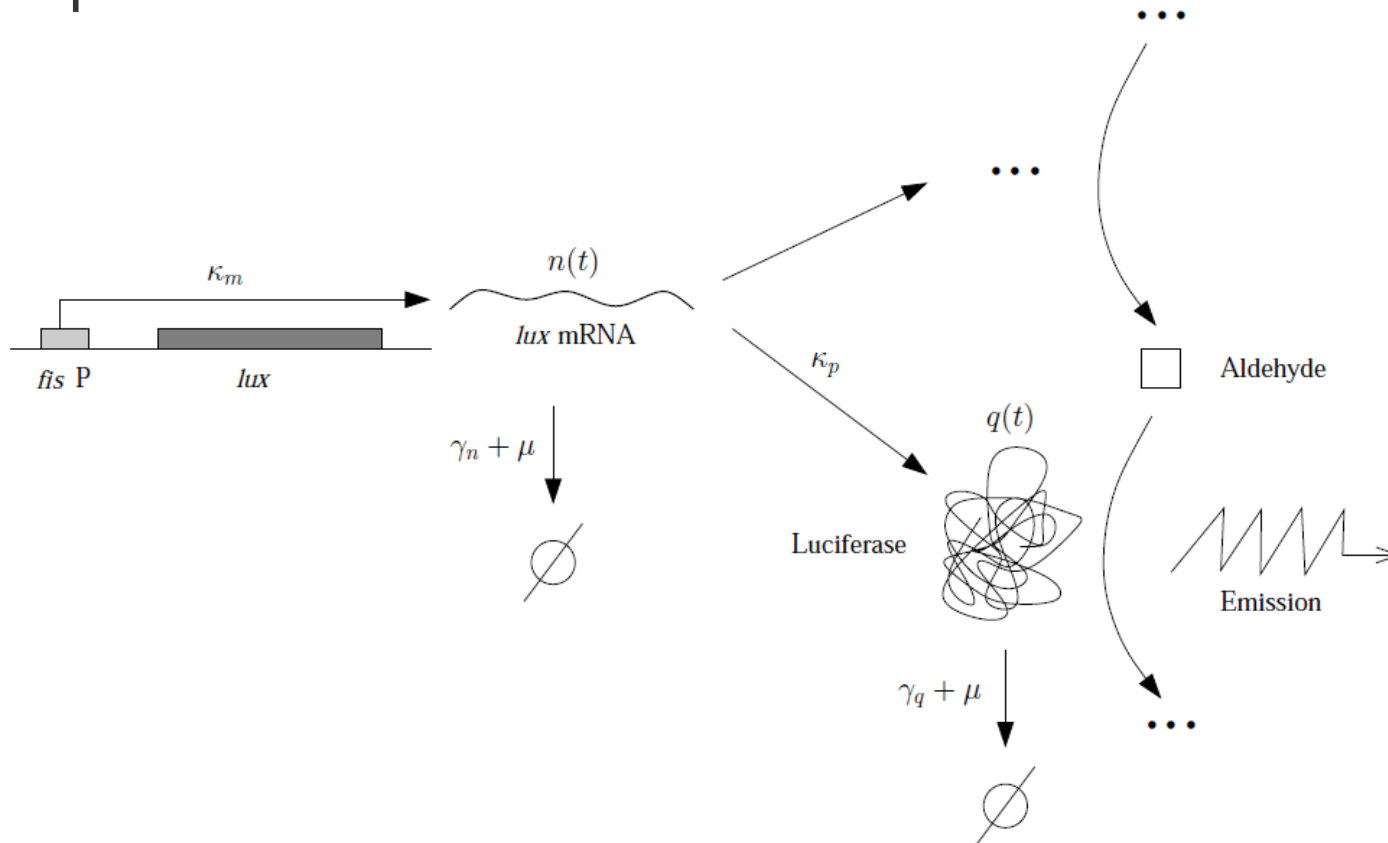


- Expression of *gfp*



Host protein and reporter genes

- Expression of luciferase

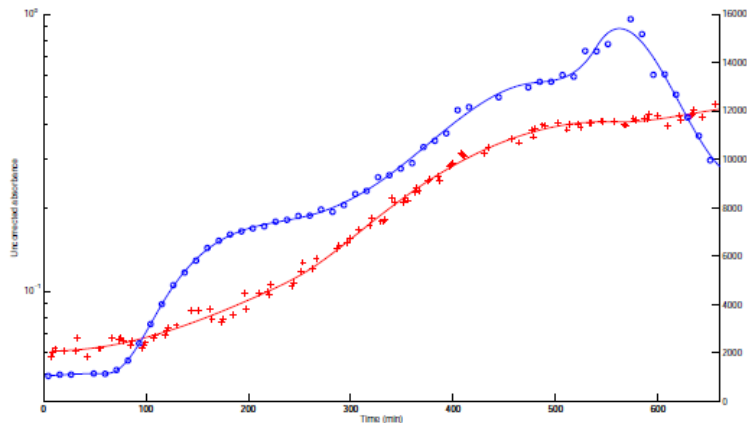


Acquisition and treatment of raw data

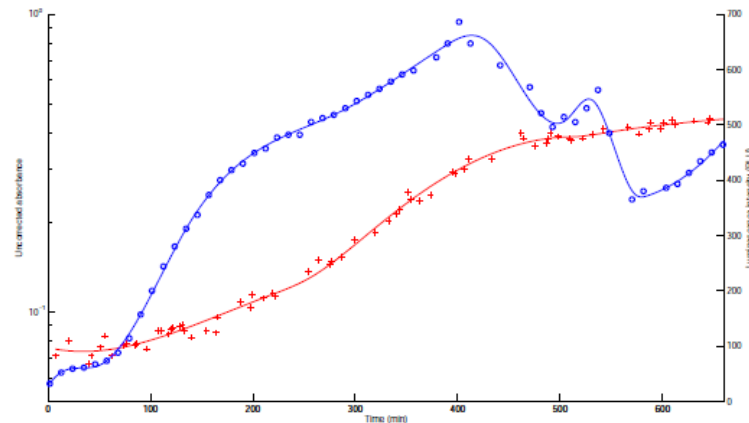
□ Data acquisition with automated microplate reader

- **About 100 data points over an interval of 10 hours**

gfp



Luciferase



□ Treatment of raw data

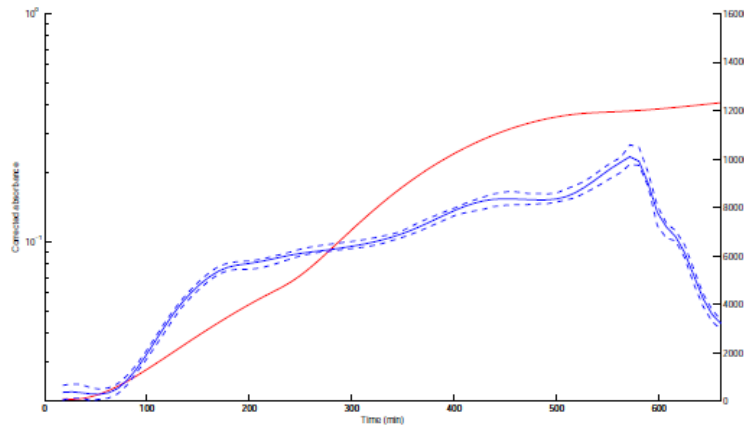
- **Outlier detection, spline fitting, background subtraction**

de Jong et al, BMC Systems Biology, 2010

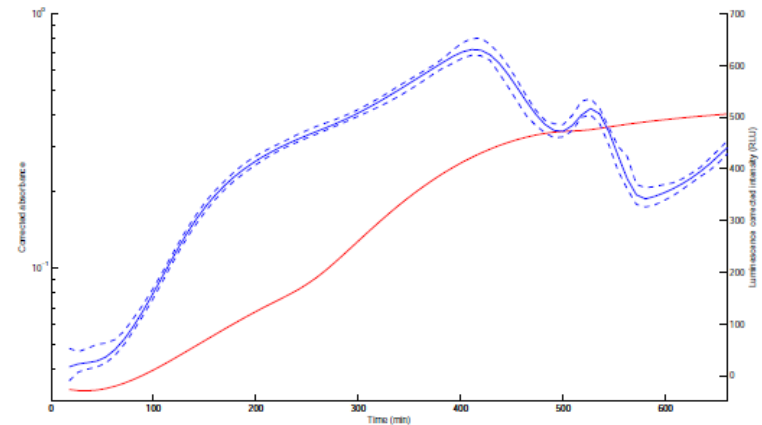
Confidence intervals

- Calculate the confidence intervals of the corrected data
 - **Bootstrap method**

gfp



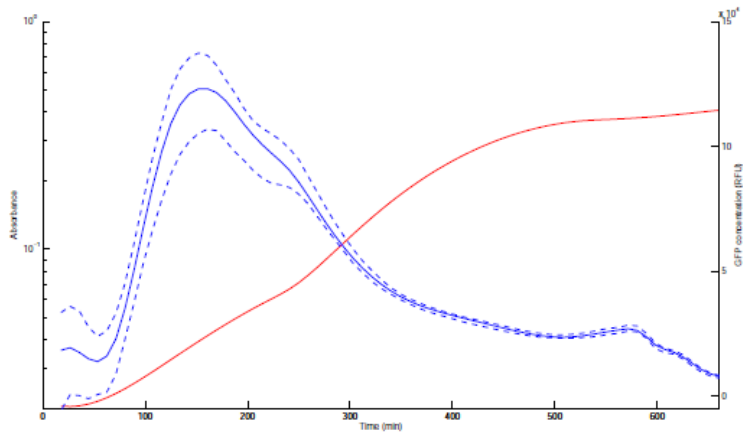
Luciferase



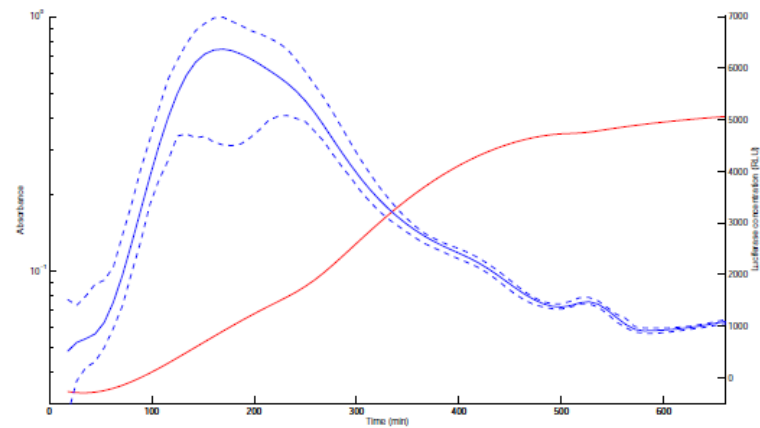
Concentration of the reporter proteins

- Compute derived quantities
 - *Reporter concentrations and synthesis rates*
 - *Relative units*

gfp



Luciferase



Reporter concentrations

Concentration of the reporter proteins

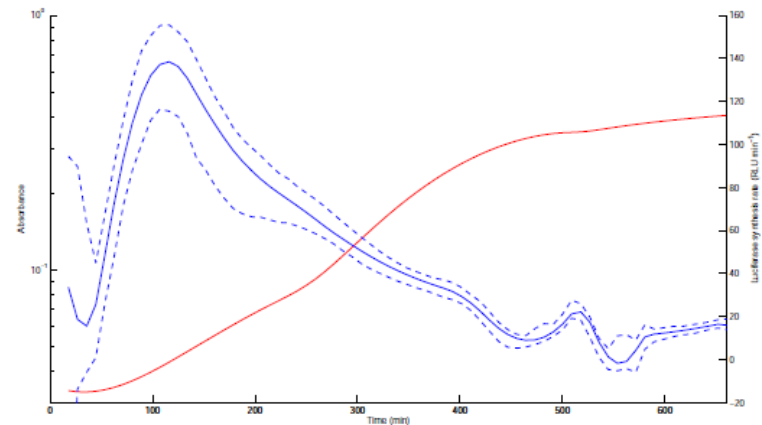
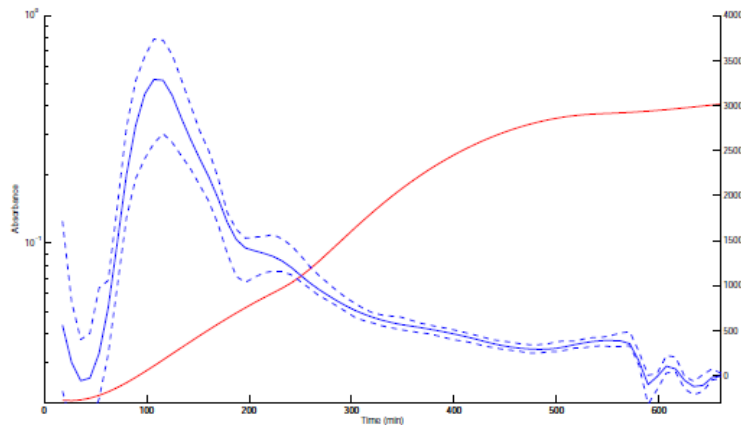
□ Compute derived quantities

□ *Reporter concentrations and synthesis rates*

□ *Relative units*

gfp

Luciferase

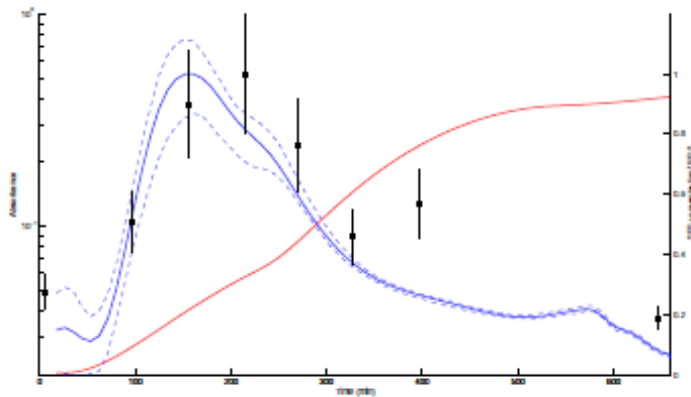


Synthesis rates

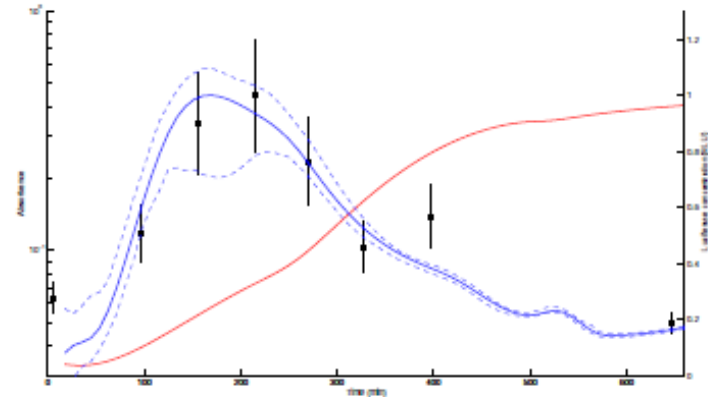
Concentration of the host protein

Comparison between the predicted and measured concentration of the host protein

gfp

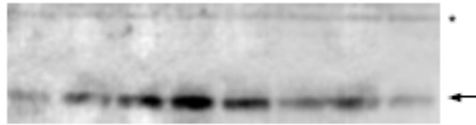


Luciferase

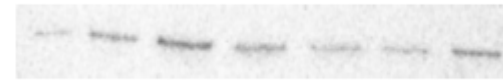


Concentration of the mRNA

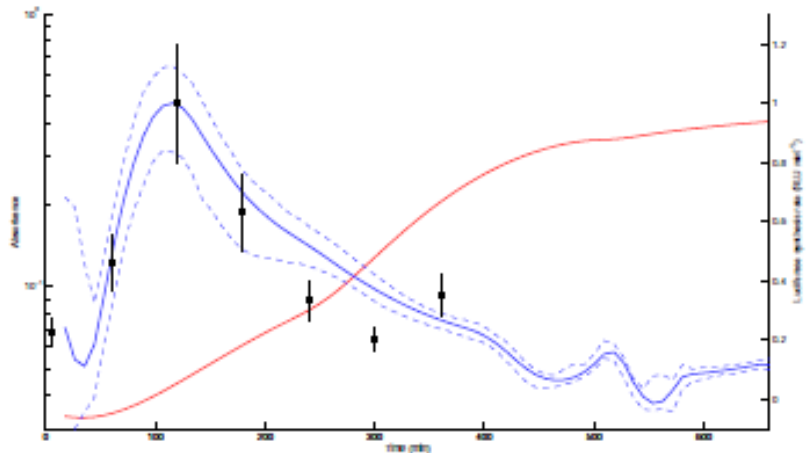
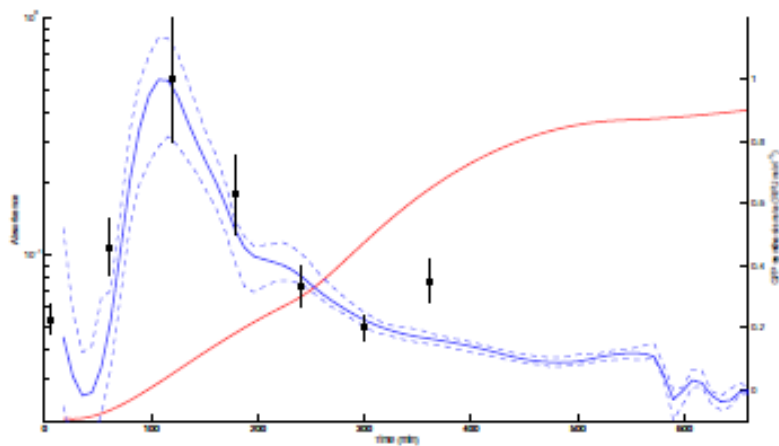
Predicted and measured concentration of the mRNA of the host protein and reporter genes



gfp



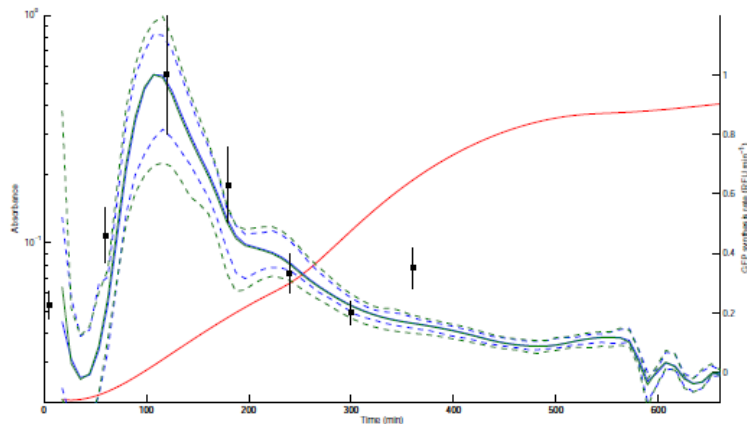
Luciferase



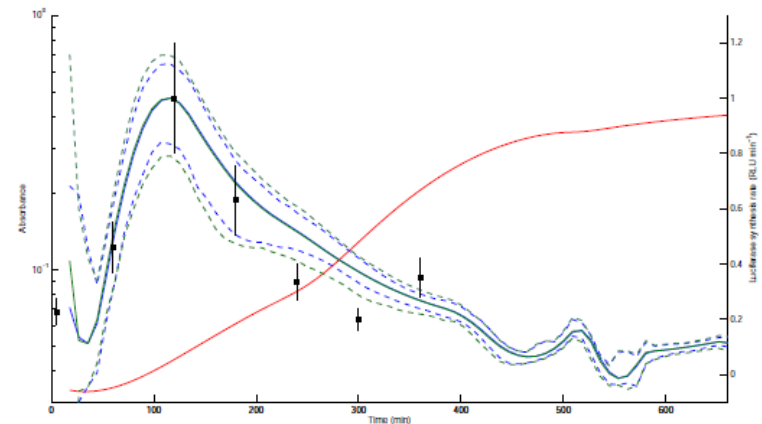
Influence of the half-life of mRNA

Taking into account the differences in the half-lives of the mRNAs of the reporter and host proteins has little influence on the profile of the mRNAs

gfp



Luciferase



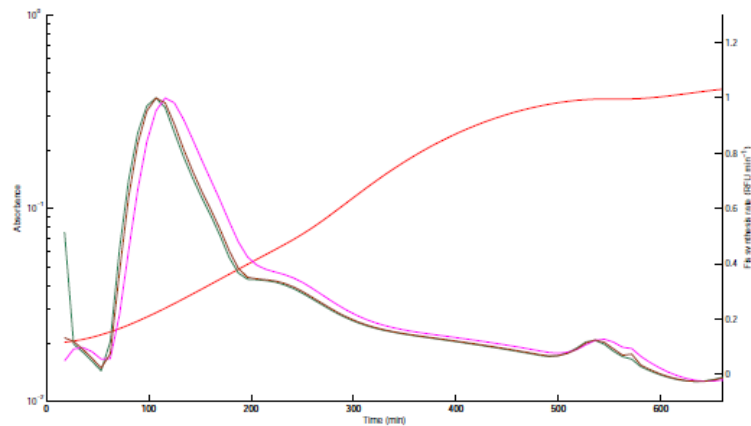
green: host protein

blue: gfp or luciferase

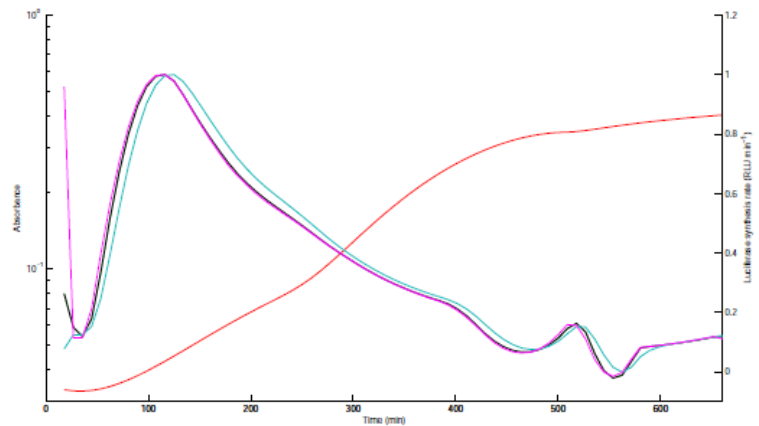
Robustness of the synthesis rate

The profile of mRNA expression changes very little when the half-life of the host protein mRNA is varied by a factor of 16

gfp

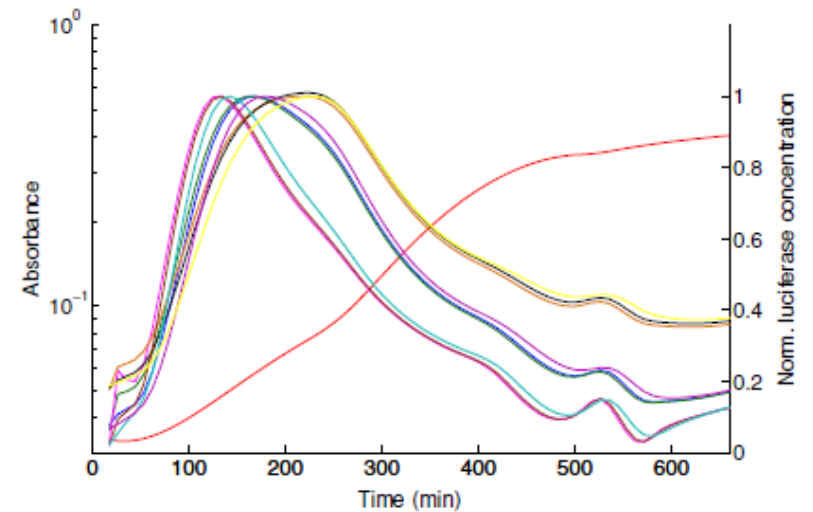
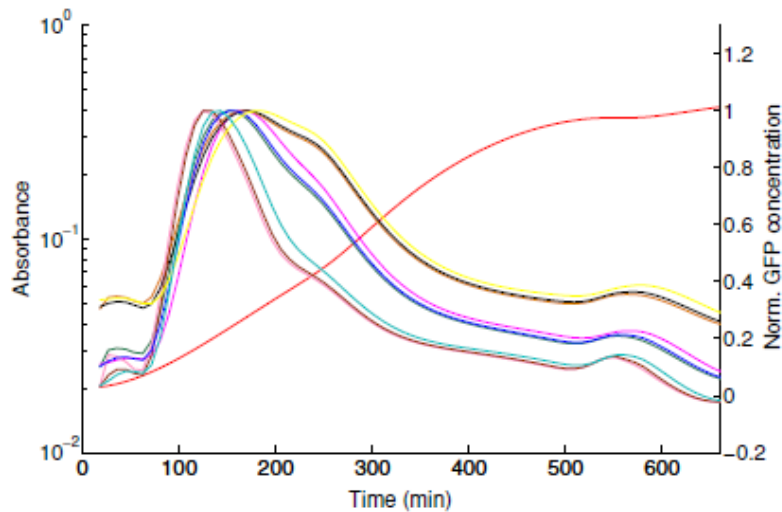


Luciferase



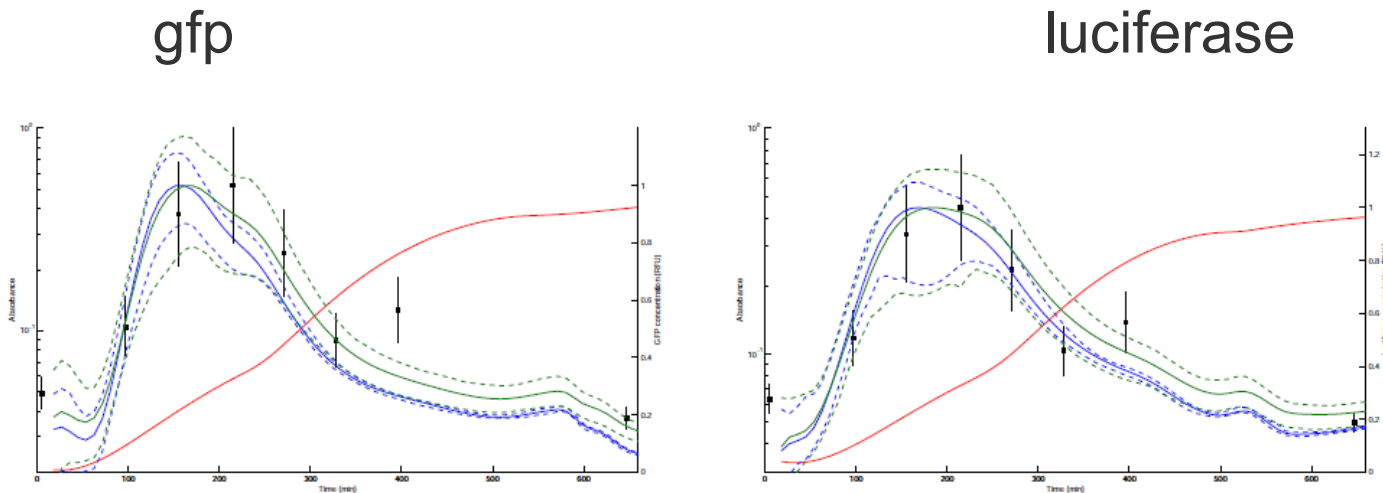
Influence of the protein half-life

- More pronounced influence



Correction for half-lives of reporter and host protein

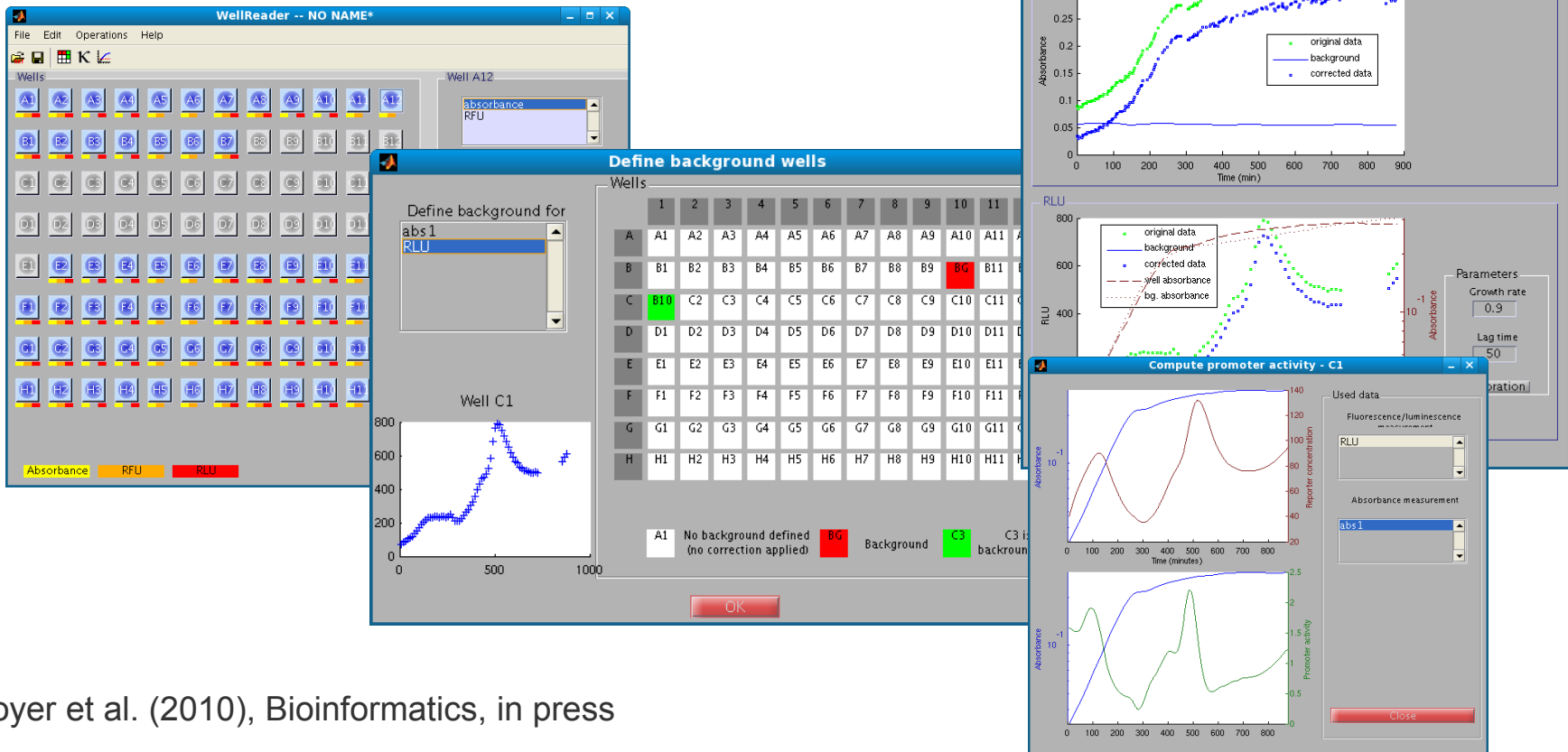
The calculated concentration of the host protein is corrected for the differences in half-lives of reporter gene and host protein



green: host protein blue: gfp or luciferase

Analysis of reporter gene expression data

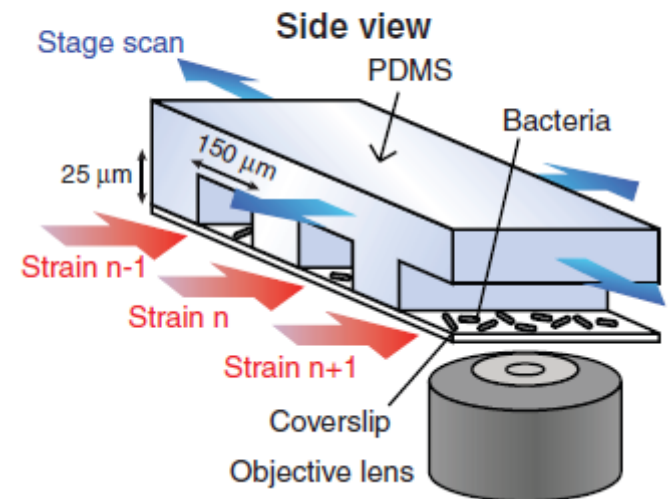
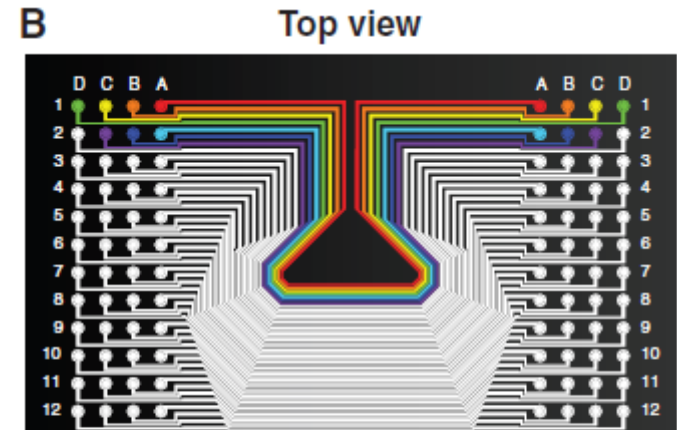
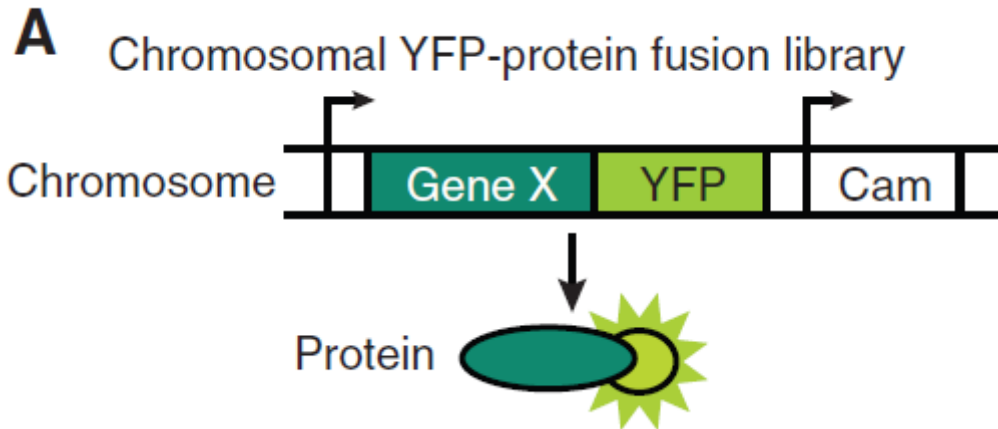
Wellreader: Matlab program for analysis of reporter gene expression data



Boyer et al. (2010), Bioinformatics, in press

Single cell imaging of fusion proteins

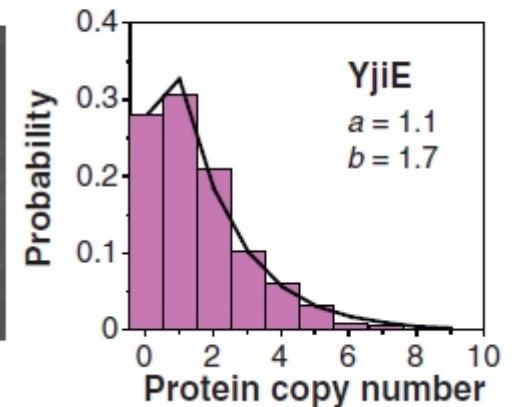
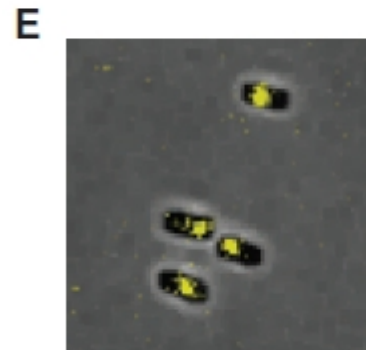
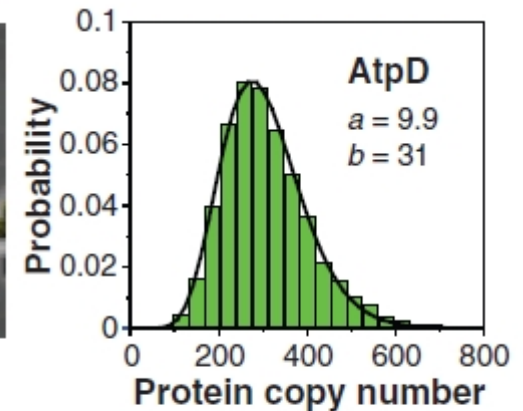
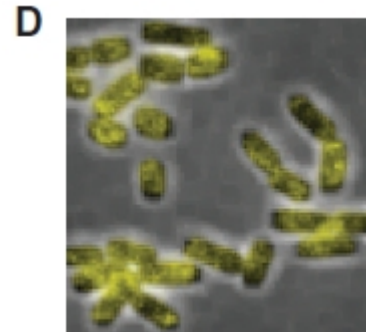
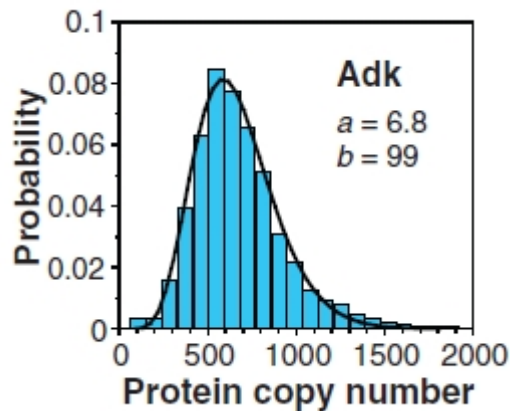
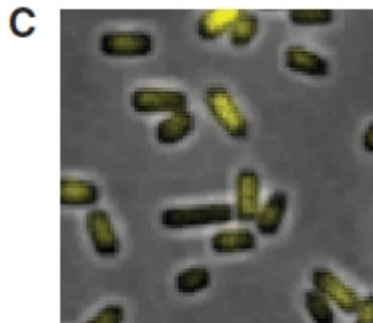
- Measure fluorescence of fusion proteins



The number of proteins is gamma distributed

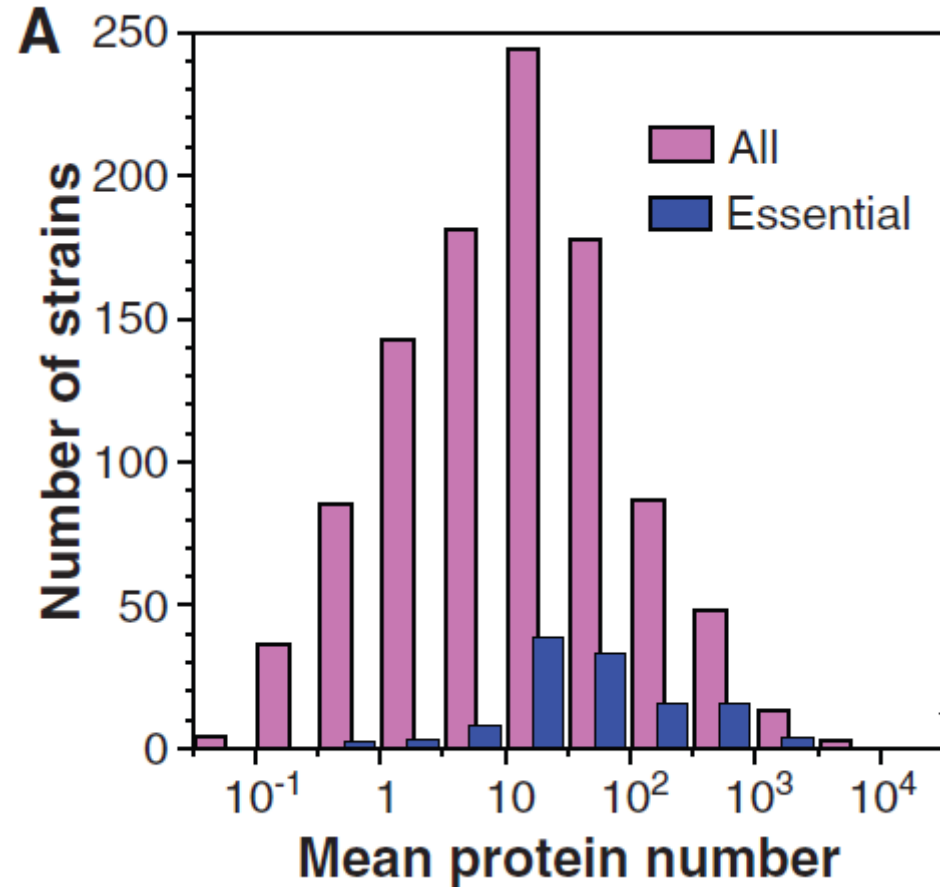
- Examples of protein distributions

$$p(x) = \frac{x^{a-1} e^{-x/b}}{\Gamma(a)b^a}$$



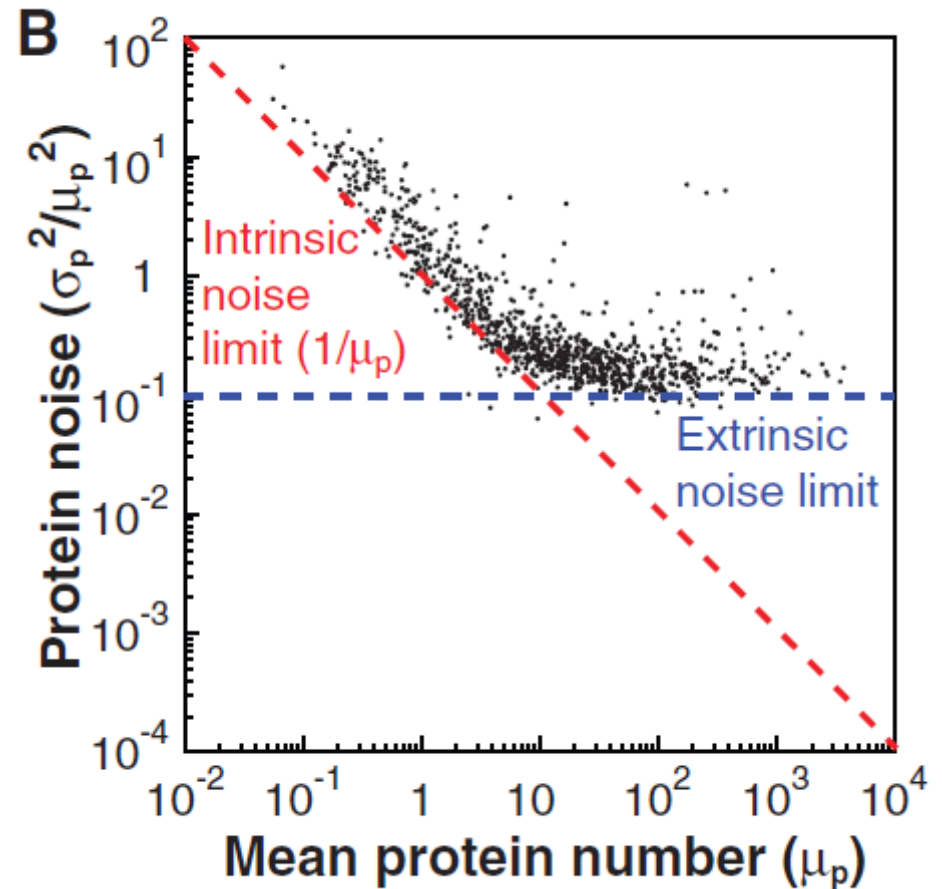
Essential proteins are more abundant

- Many proteins are present in low abundance



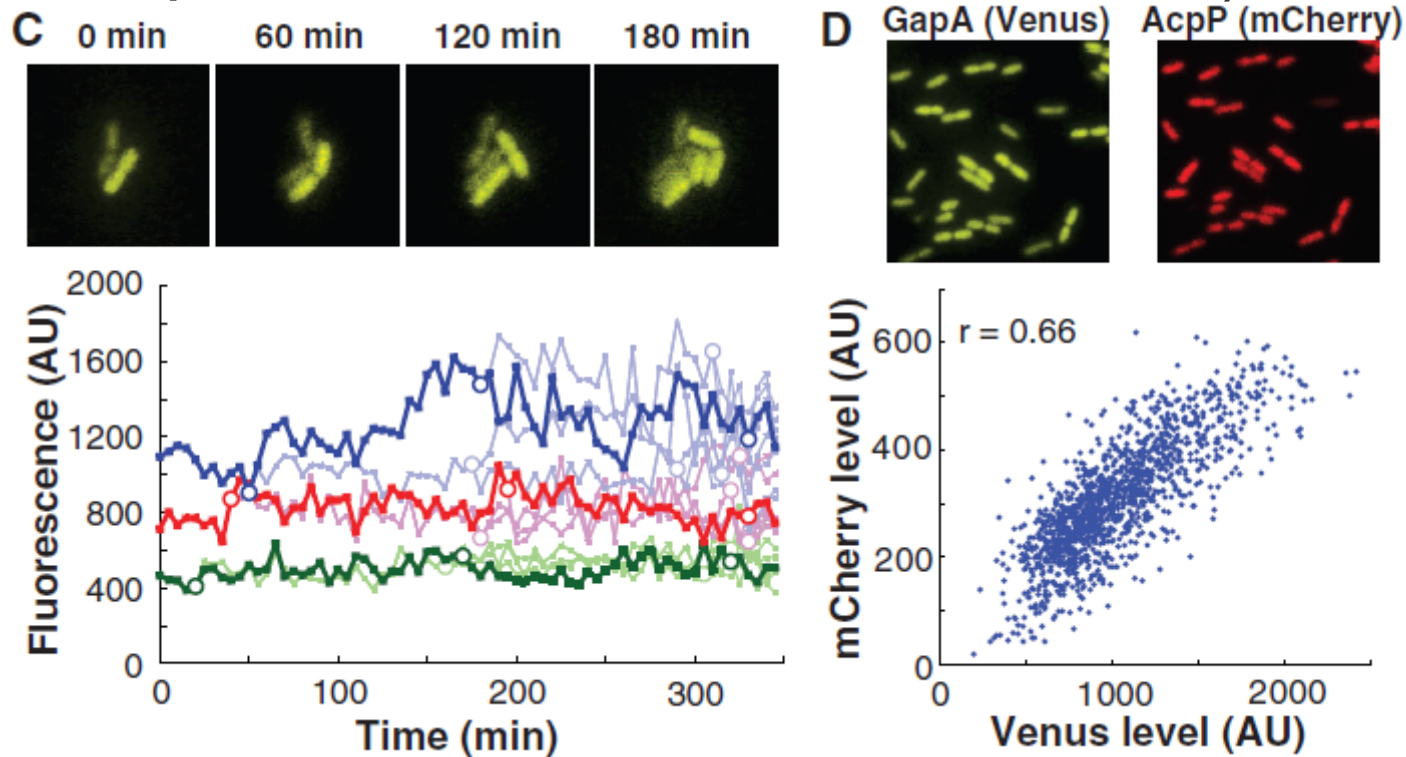
Noise of protein expression

- Noise $\eta_p^2 \equiv \sigma_p^2 / \mu_p^2$
- is equal to the gamma parameter a
- proportional to protein abundance for $N < 10$ proteins per cell. This indicates intrinsic noise



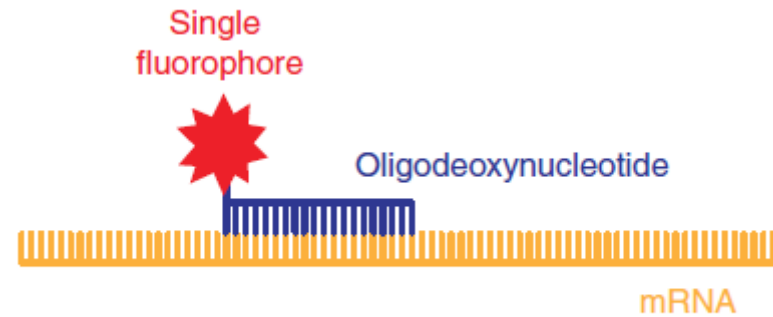
Fluctuations in protein levels are slow

- The extrinsic noise of highly expressed proteins is correlated. Therefore global factors (cellular component, such as ribosomes, etc.)



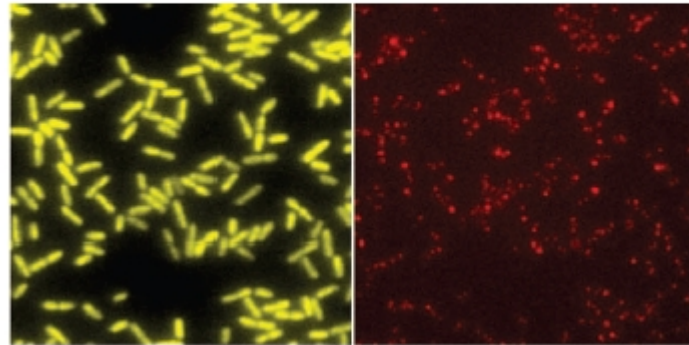
Measuring RNA molecules by FISH

- target *yfp* RNA

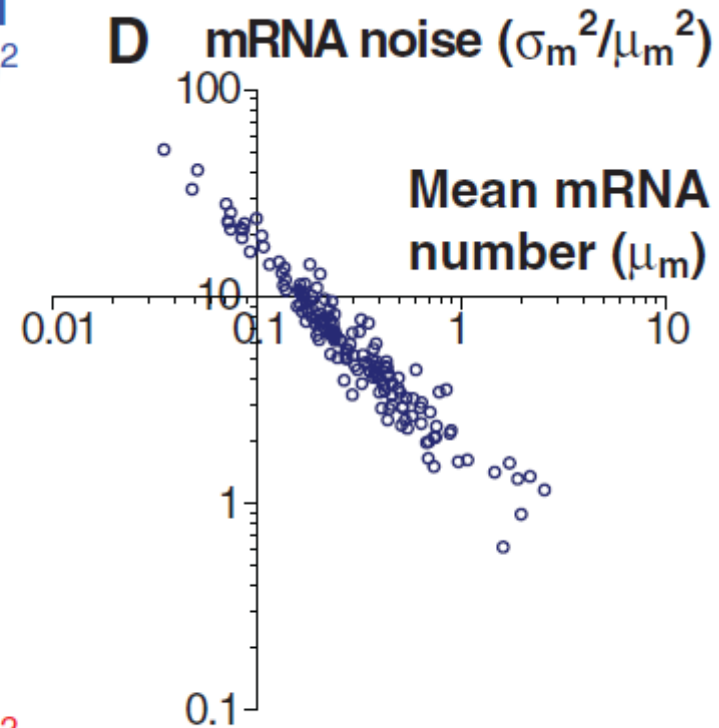
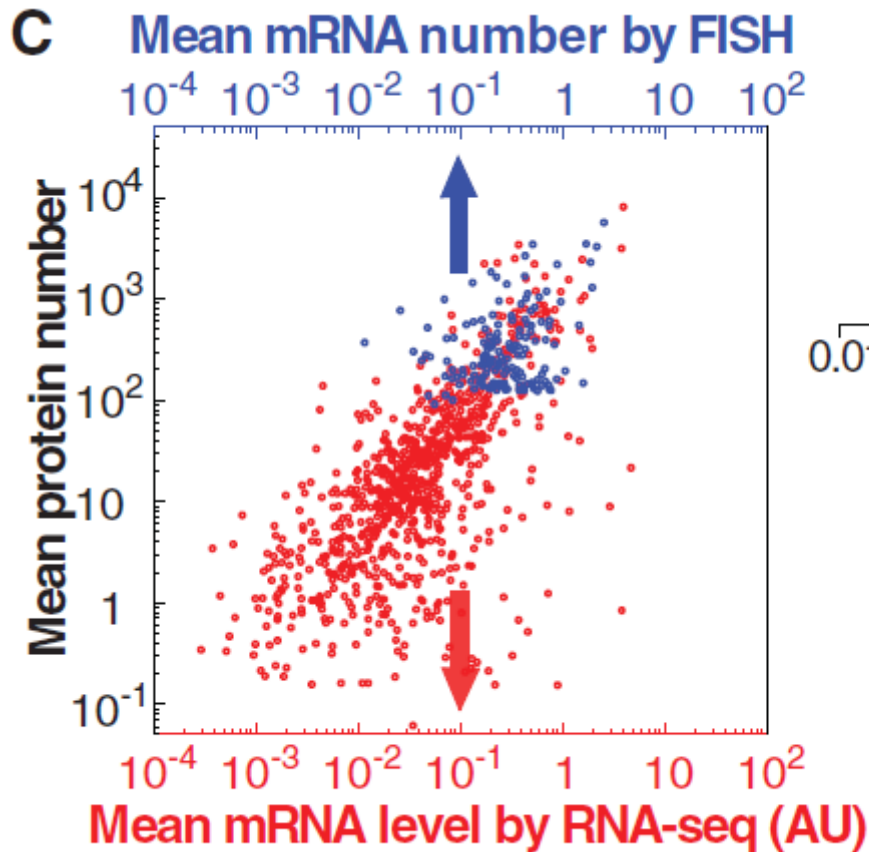


YFP fluorescence
514 nm ex.

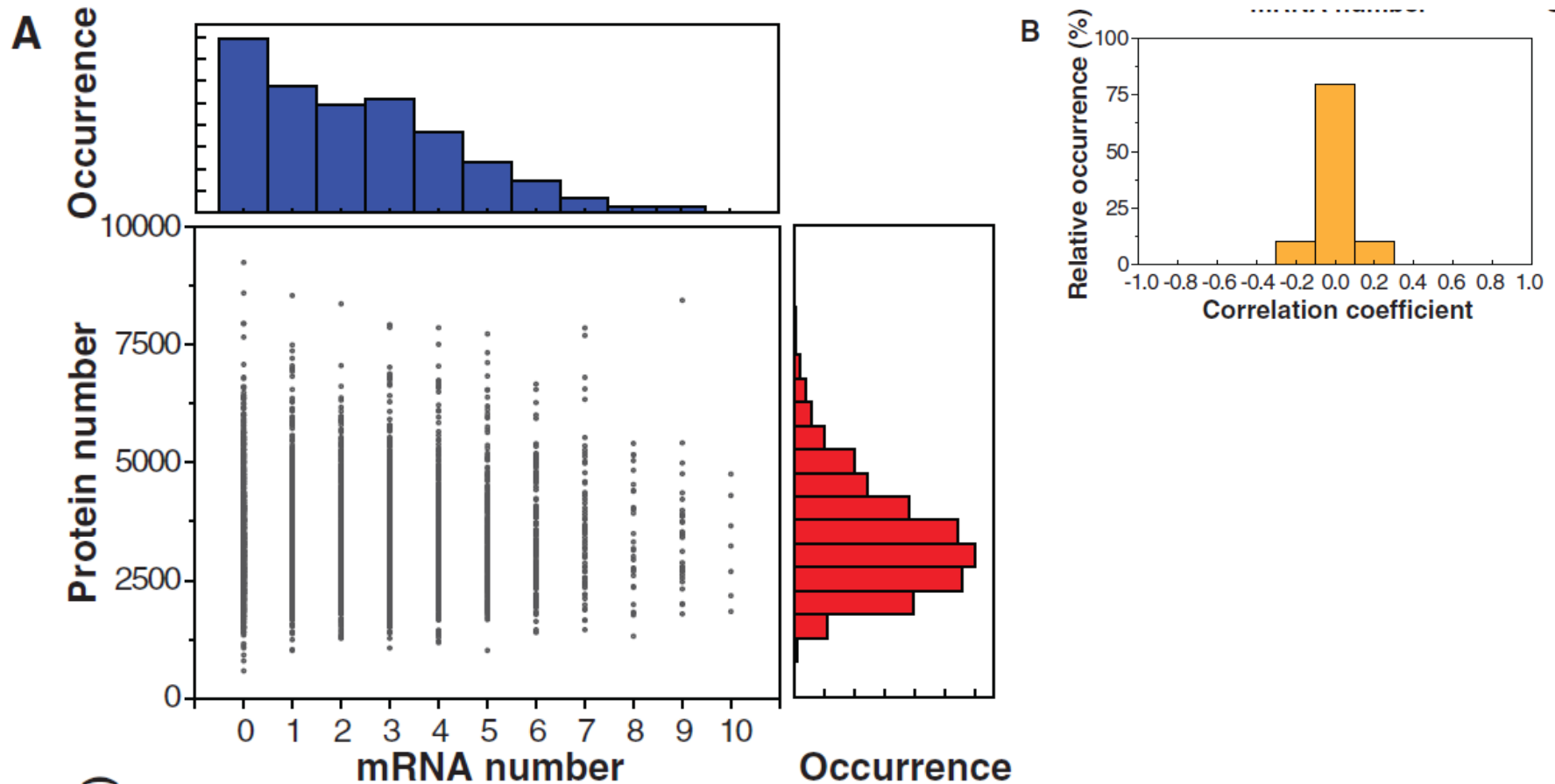
yfp mRNA
with Atto 594 probe
580 nm ex.



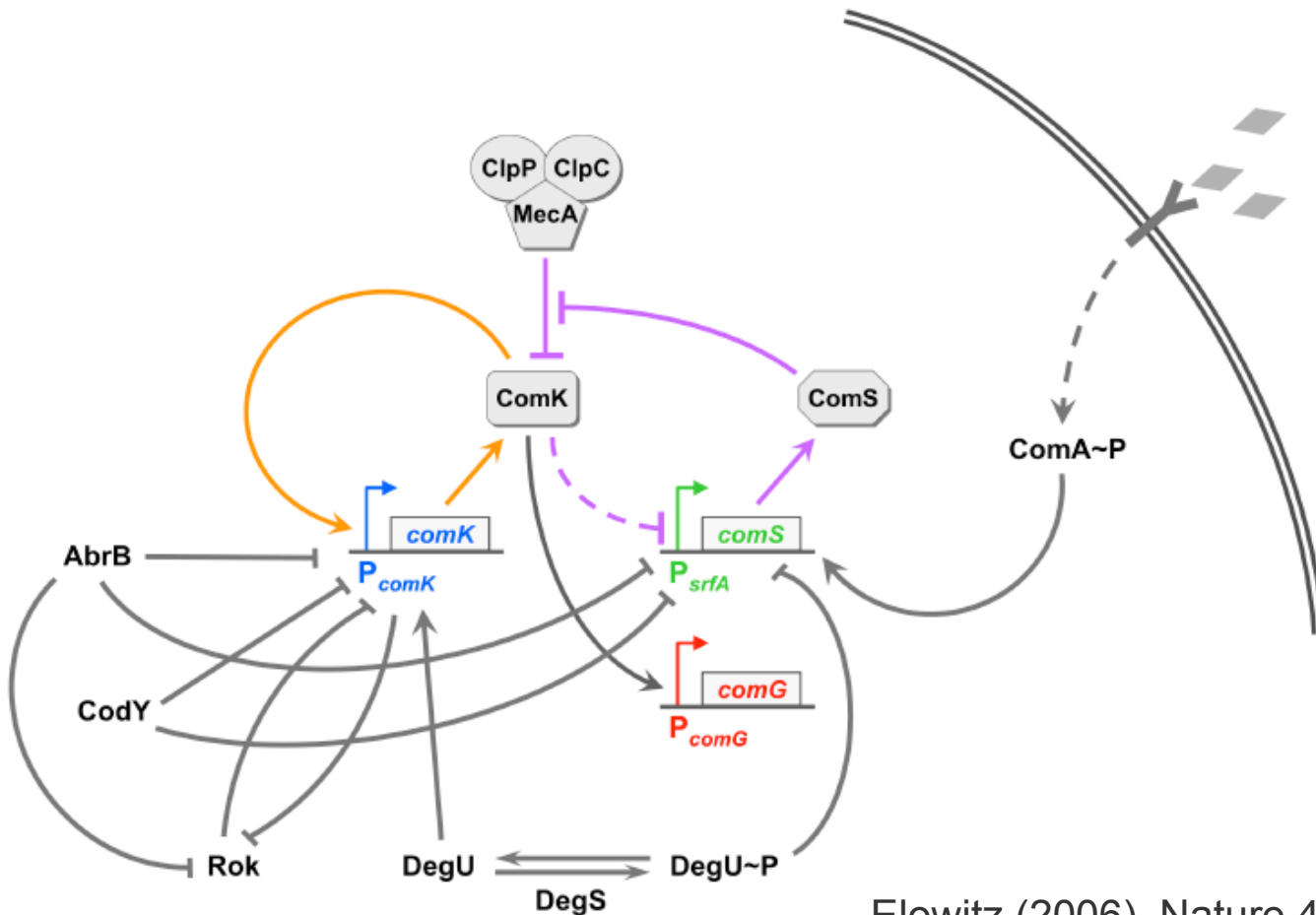
General correlation between mRNA and protein concentrations and random noise



No correlation between mRNA and protein level in a single cell



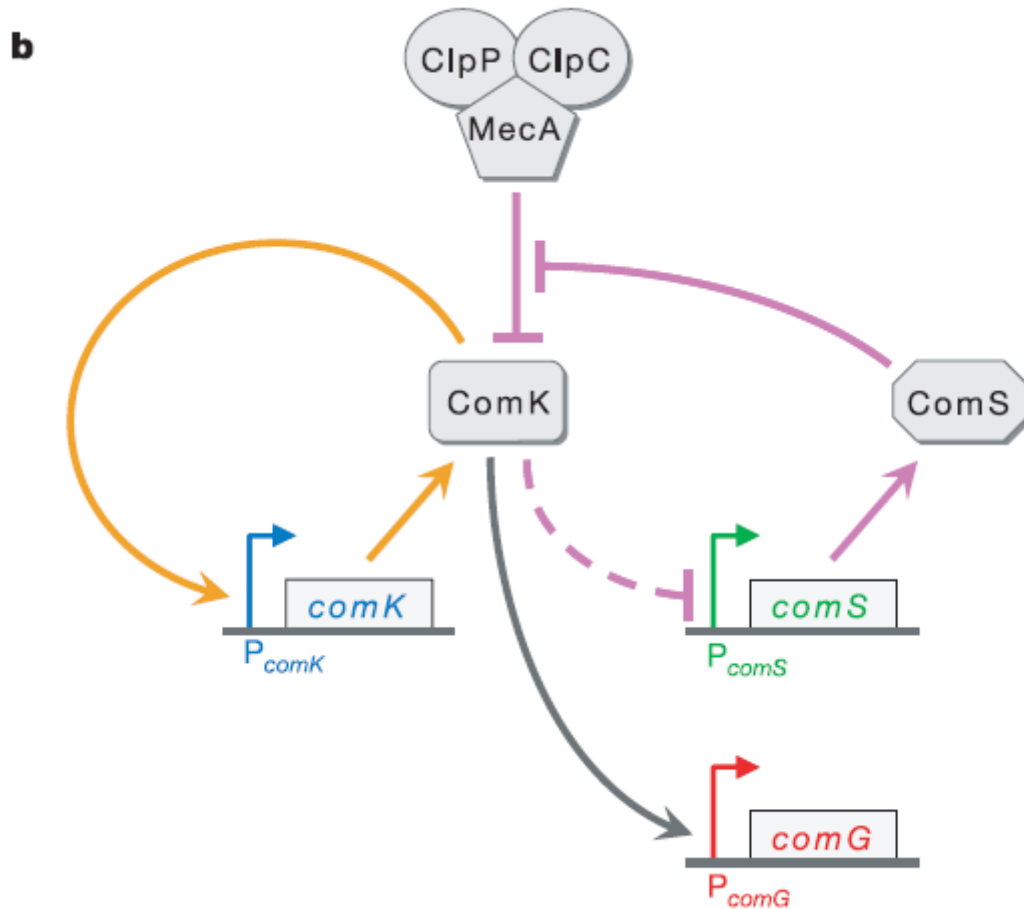
Regulation of competence



Elowitz (2006), Nature 440, 545-550

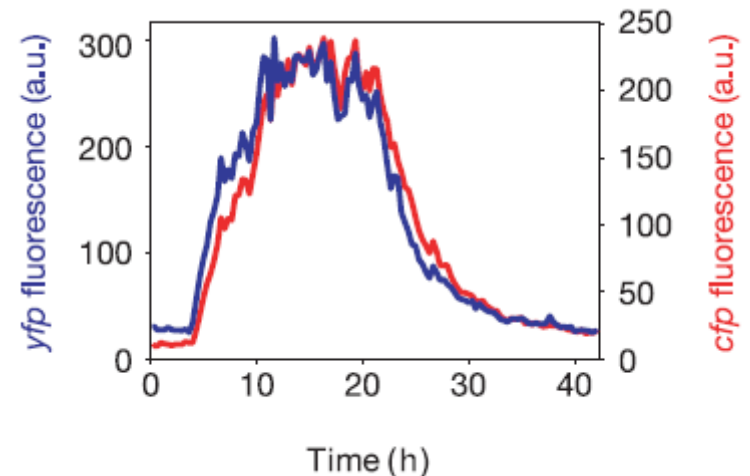
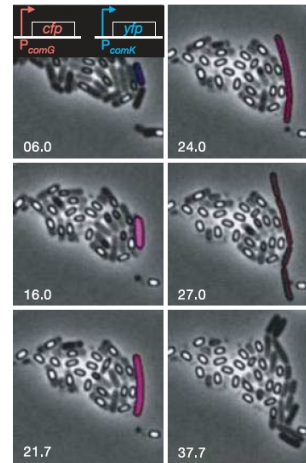
Simplified MeKS model

- Focus on the essential system components



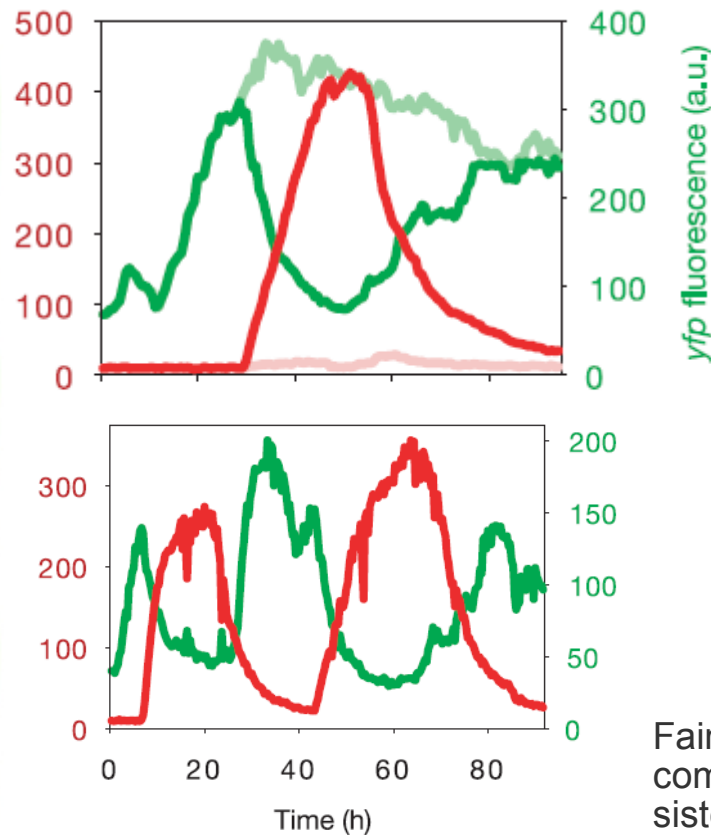
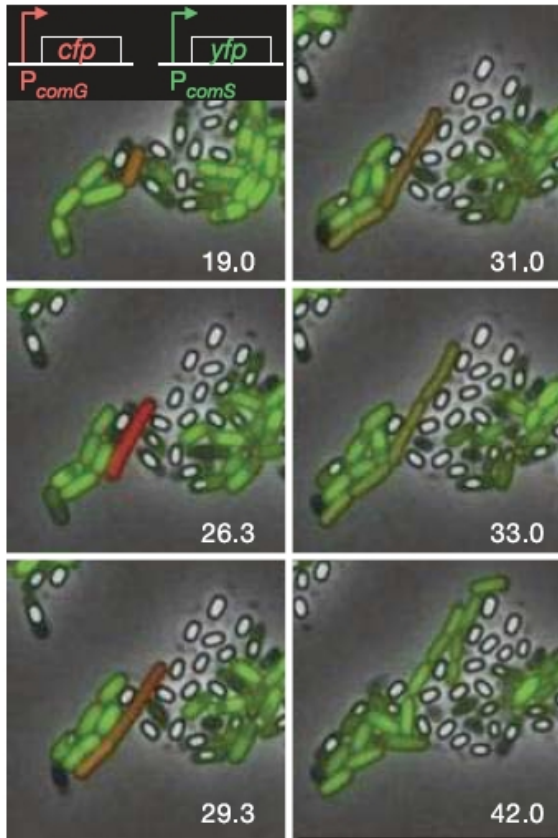
P_{comG} and P_{comK} activities are highly correlated

- *comG* is a target of ComK
- *comK* is a target of ComK, but many other genes regulate *comK*
- If the activities of P_{comK} and P_{comG} are correlated, the other influences on *comK* are less important



Activities of the *comG* and *comS* promoters

- Negative correlation between *comG* and *comS* activities confirms the “simple” MeKS model



Faint: non-competent sister cell

Model of the MeKS system

$$\frac{dK}{dt} = a_k + \frac{b_k K^n}{k_0^n + K^n} - \frac{K}{1 + K + S}$$

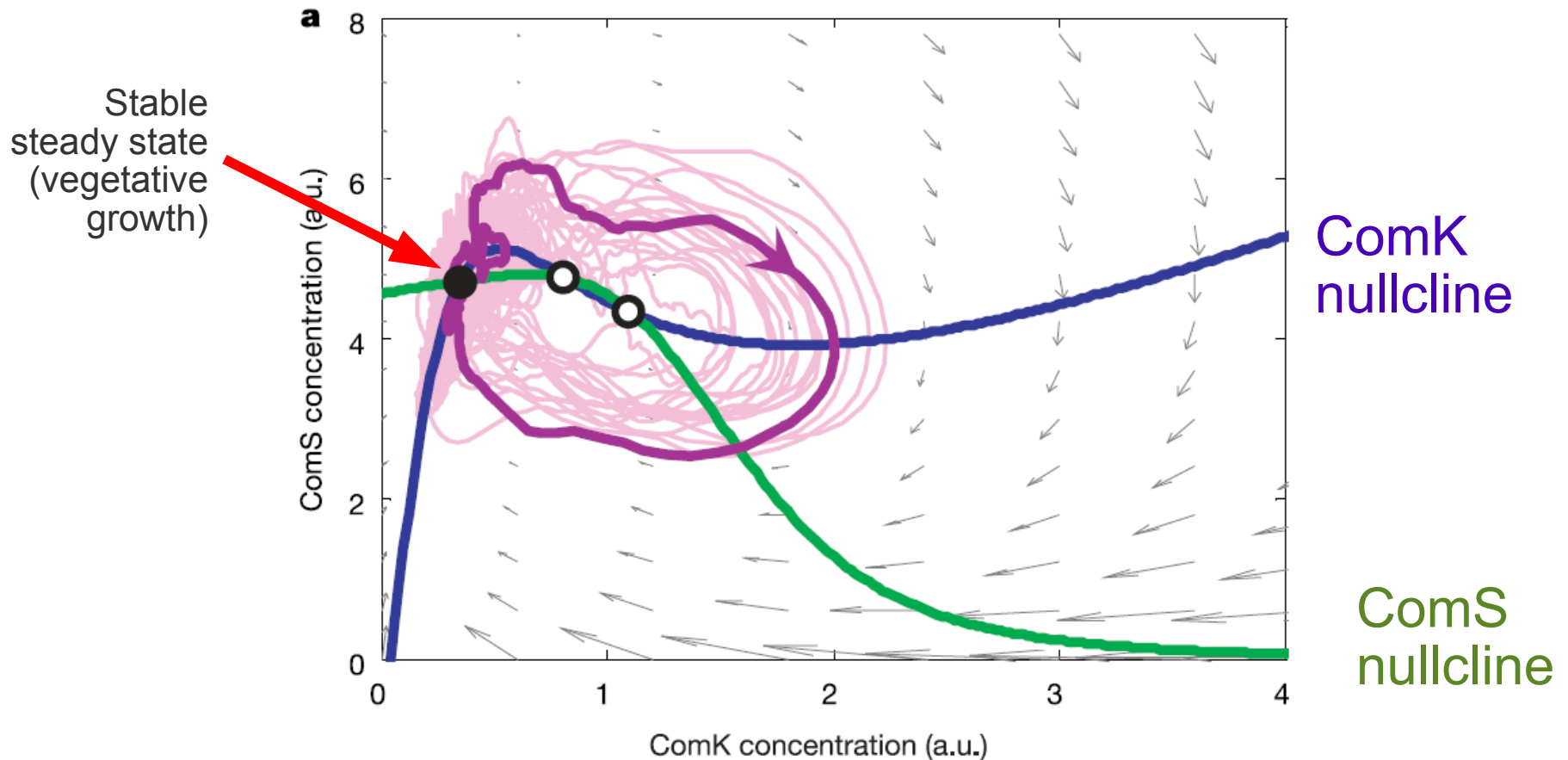
Auto-activation Competitive inhibition

$$\frac{dS}{dt} = \frac{b_s}{1 + (K/k_1)^p} - \frac{S}{1 + K + S} + \xi(t)$$

Inhibition noise

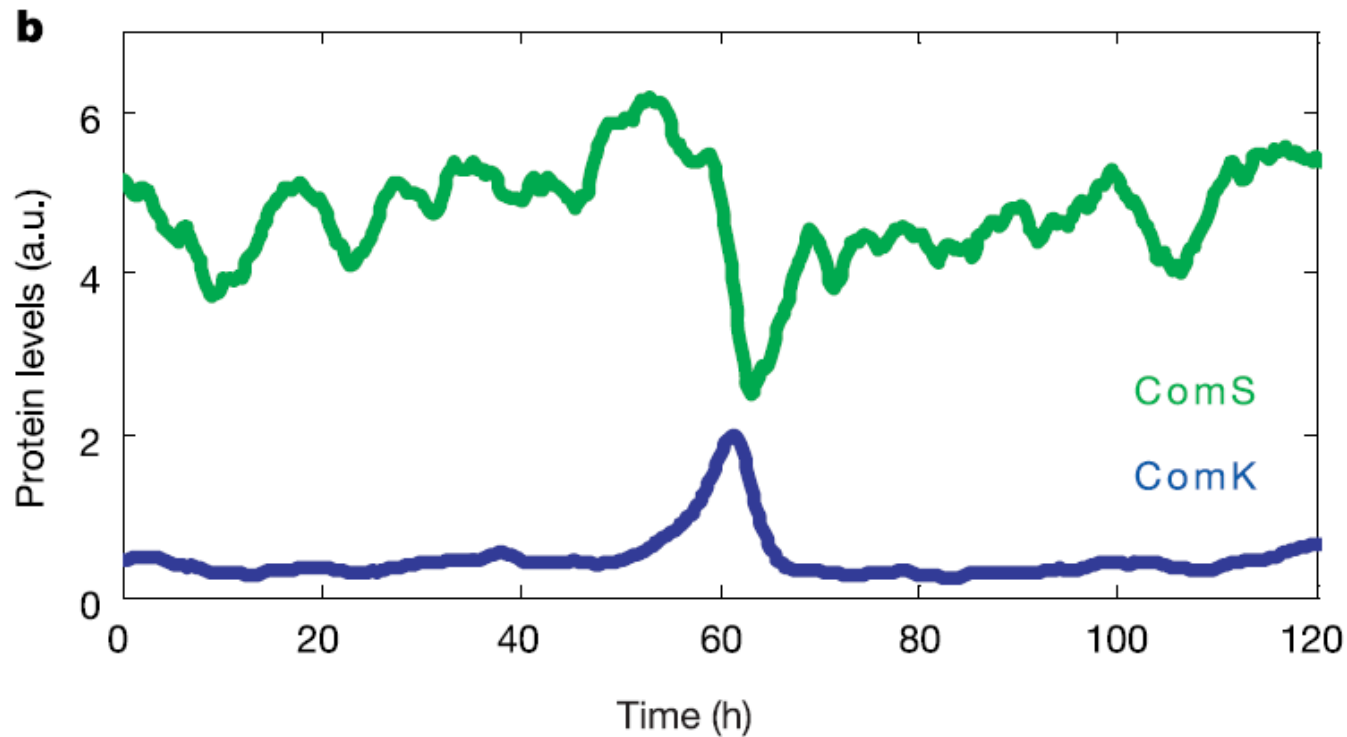
Analysis of the dynamical system

- Phase plane diagram for ComK and ComS



Simulation of ComK and ComS trajectories

- Excursion triggered by noise
- Anticorrelation between ComK and ComS

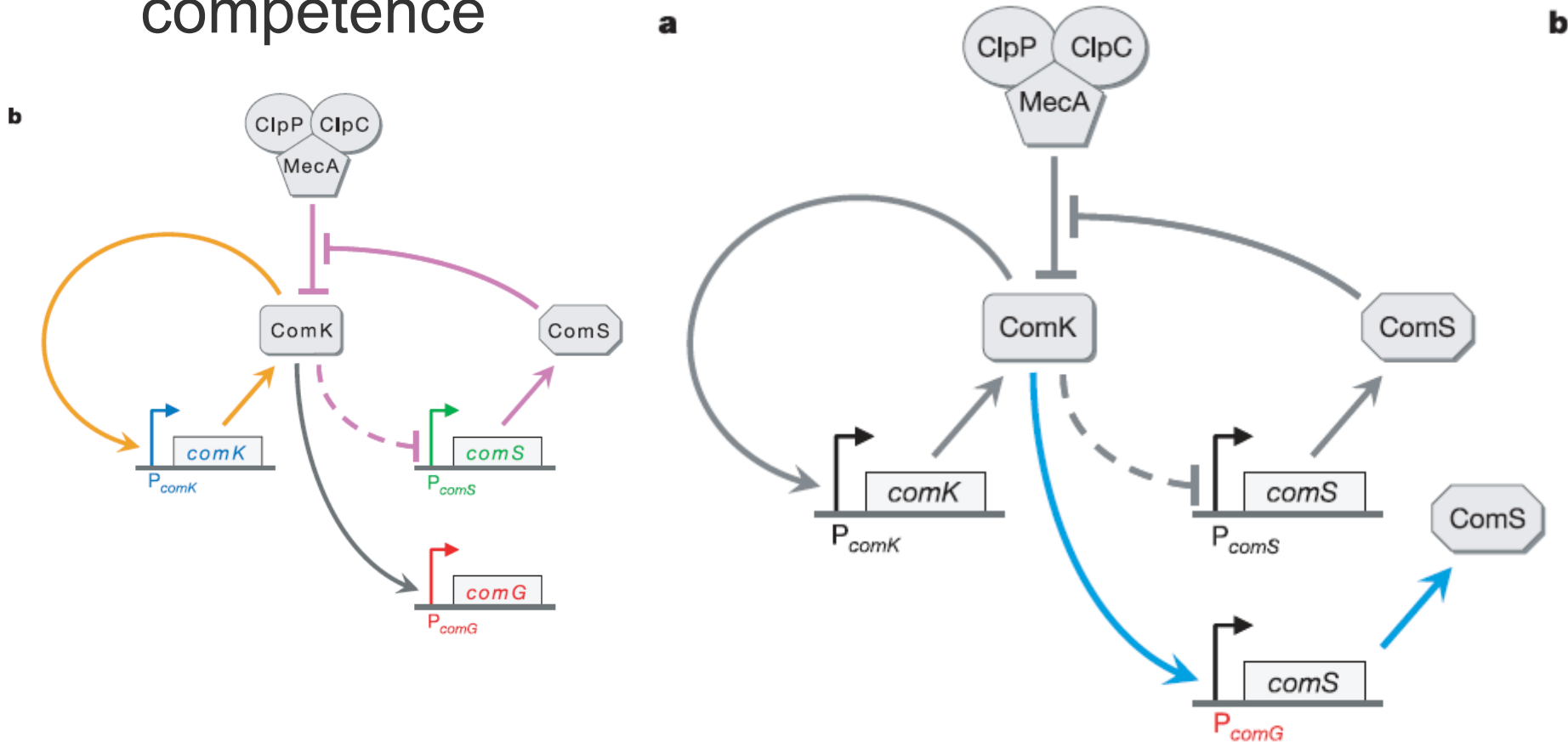


Dynamics of system functioning

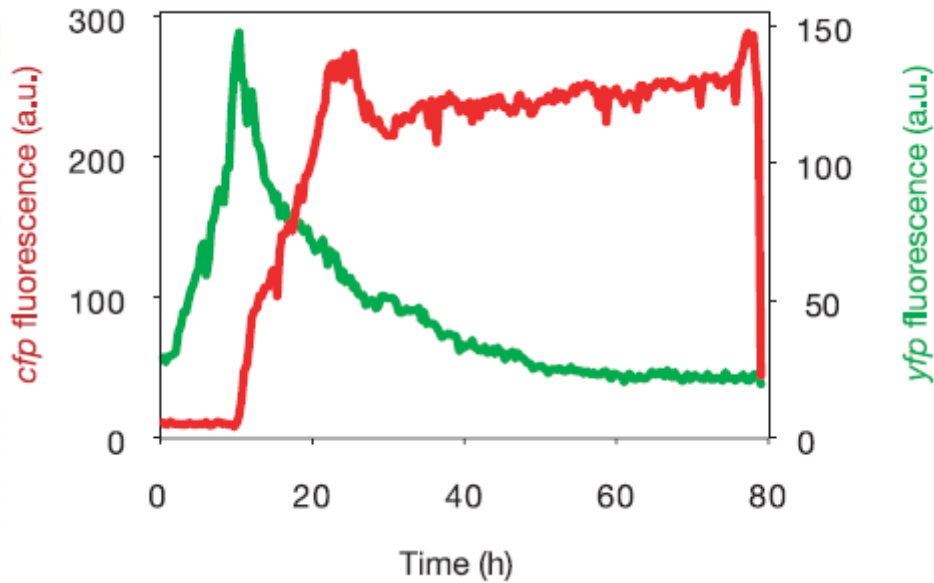
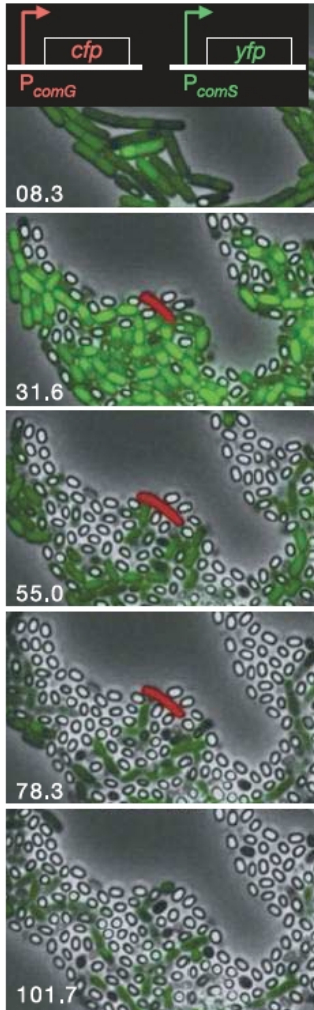
- After a transient excursion, the **fast**, positive feedback loop of ComK on itself leads to competence
- On a slower timescale, this initiates the ComS mediated negative feedback loop
- Reduction in ComS concentration eventually shuts the system back “off” through increased ComK degradation

Locking the system into competence mode

- Modify a connection to keep ComS during competence



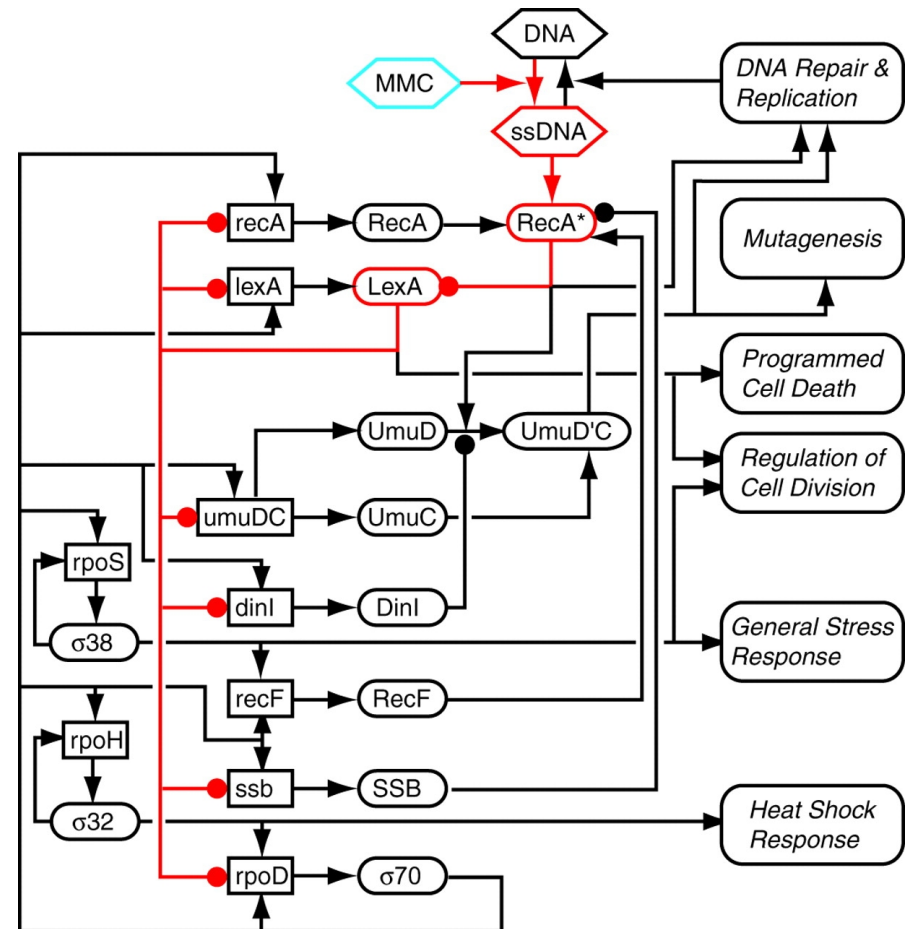
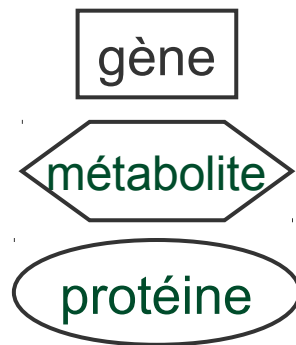
The modified system can not escape from competence



Le régulon SOS (version un peu plus complète)

- Les dommages à l'ADN induisent l'opéron
- Le réseau comprend des boucles de rétroaction

Interactions principales



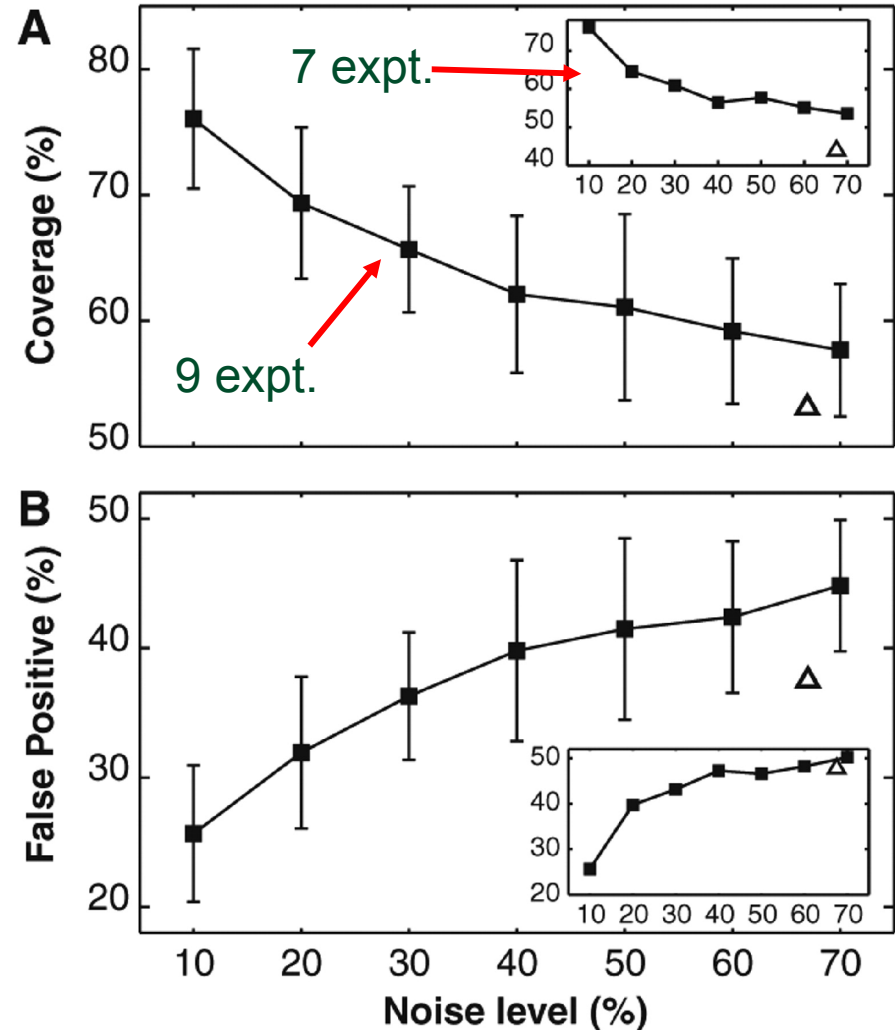
Reconstruction du réseau de régulation à partir de données d'expression

- Mesurer la concentration d'ARNm des 9 principaux composants du réseau de régulation
- Perturber le réseau : surexpression d'un des neuf gènes
- Mesurer le nouvel état stationnaire du système
- Simplifier les interactions entre gènes à des interactions linéaires
- Trouver le jeu de paramètres qui décrit le mieux les changements observés

La reconstruction des interactions dans un réseau de régulation est très sensible aux erreurs expérimentales

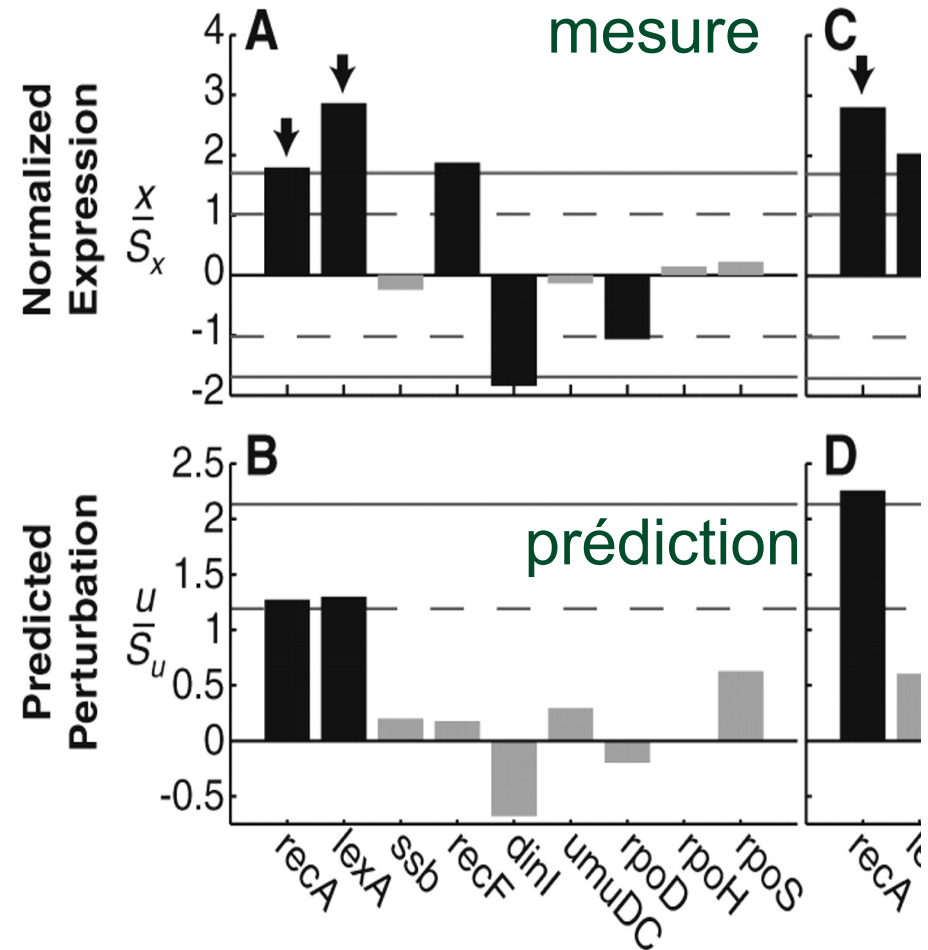
- Simulation d'un réseau de neuf gènes
- Rajouter un bruit expérimental
- Reconstruire le réseau
- Evaluer le nombre de bonnes et mauvaises connexions trouvées

Bruit expérimental : 68%



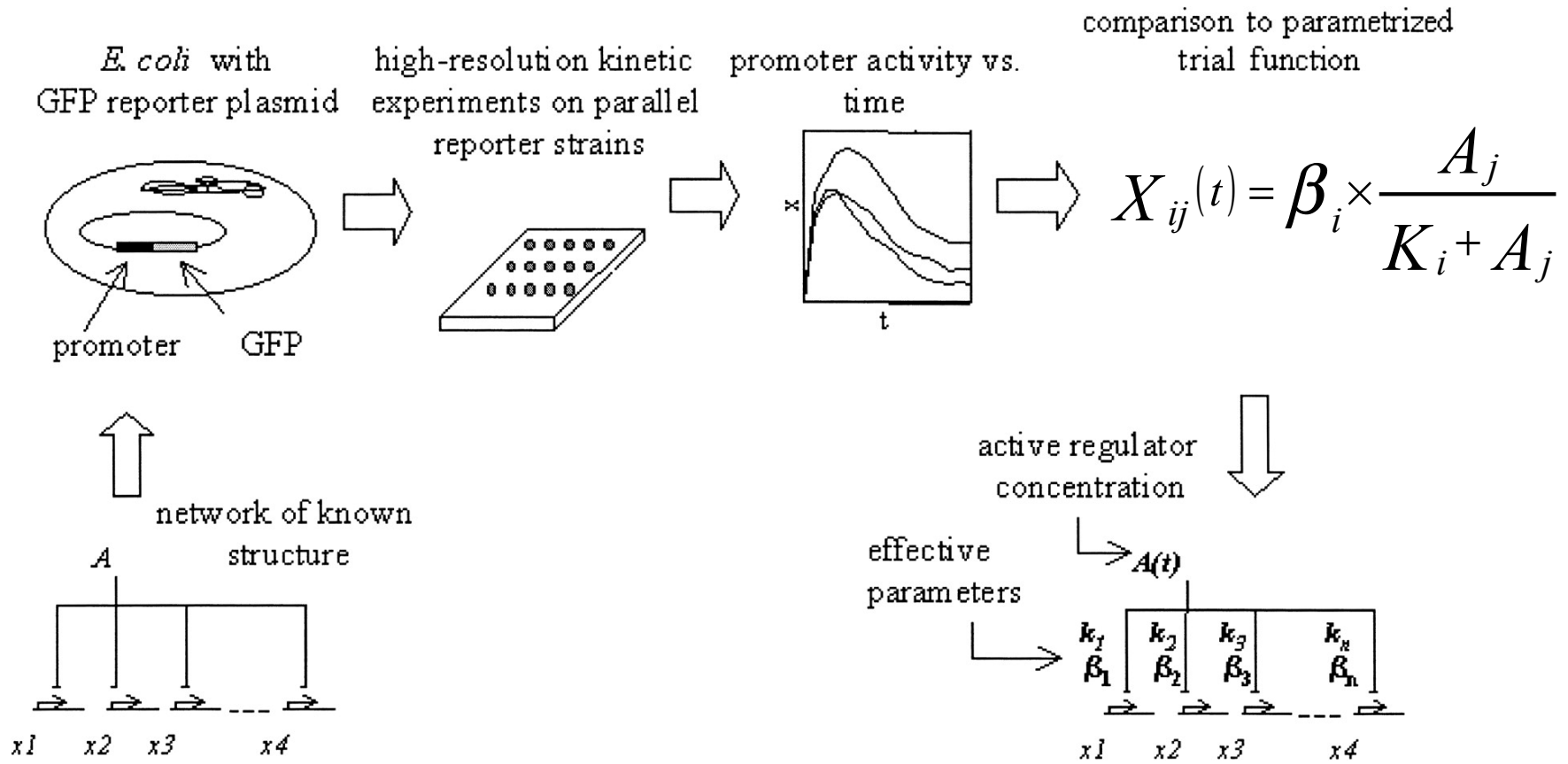
Prédiction de l'effet d'une double perturbation

- Les gènes *recA* et *lexA* sont surexprimés simultanément
- Le réseau d'interaction est utilisé pour prédire le changement d'expression des neuf gènes
- Le résultat compare bien avec les mesures



Déterminer les paramètres cinétiques d'un réseau d'interactions connu

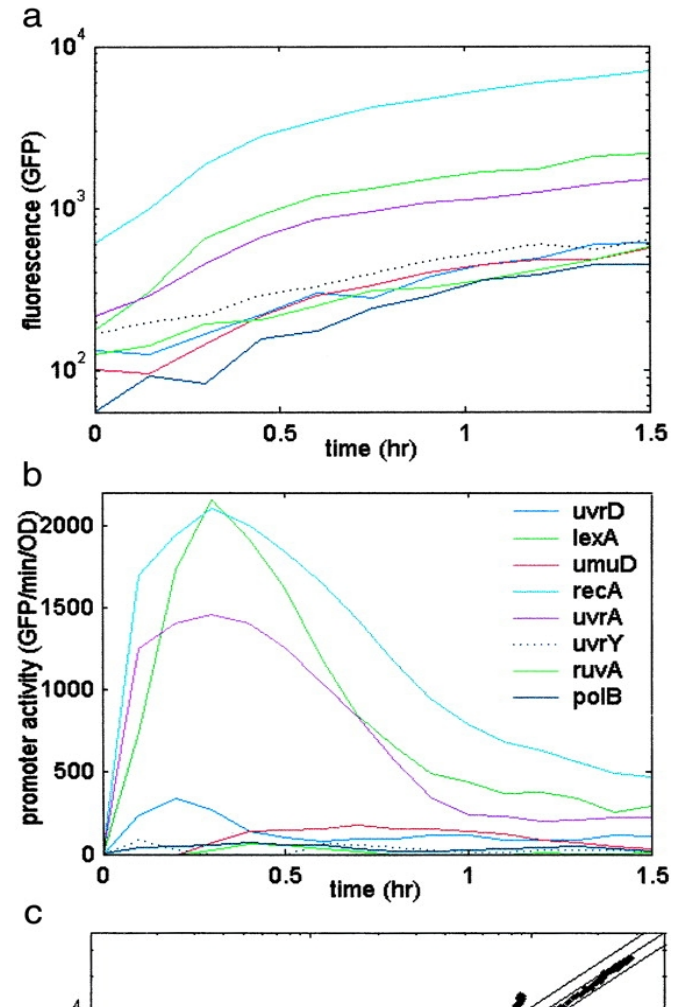
- Approche expérimentale



L'utilisation d'un gène rapporteur permet de mesurer une série temporelle avec une bonne précision

- Mesure de la fluorescence tous les minutes pour tous les gènes

L'activité du promoteur est la dérivé de la courbe de fluorescence



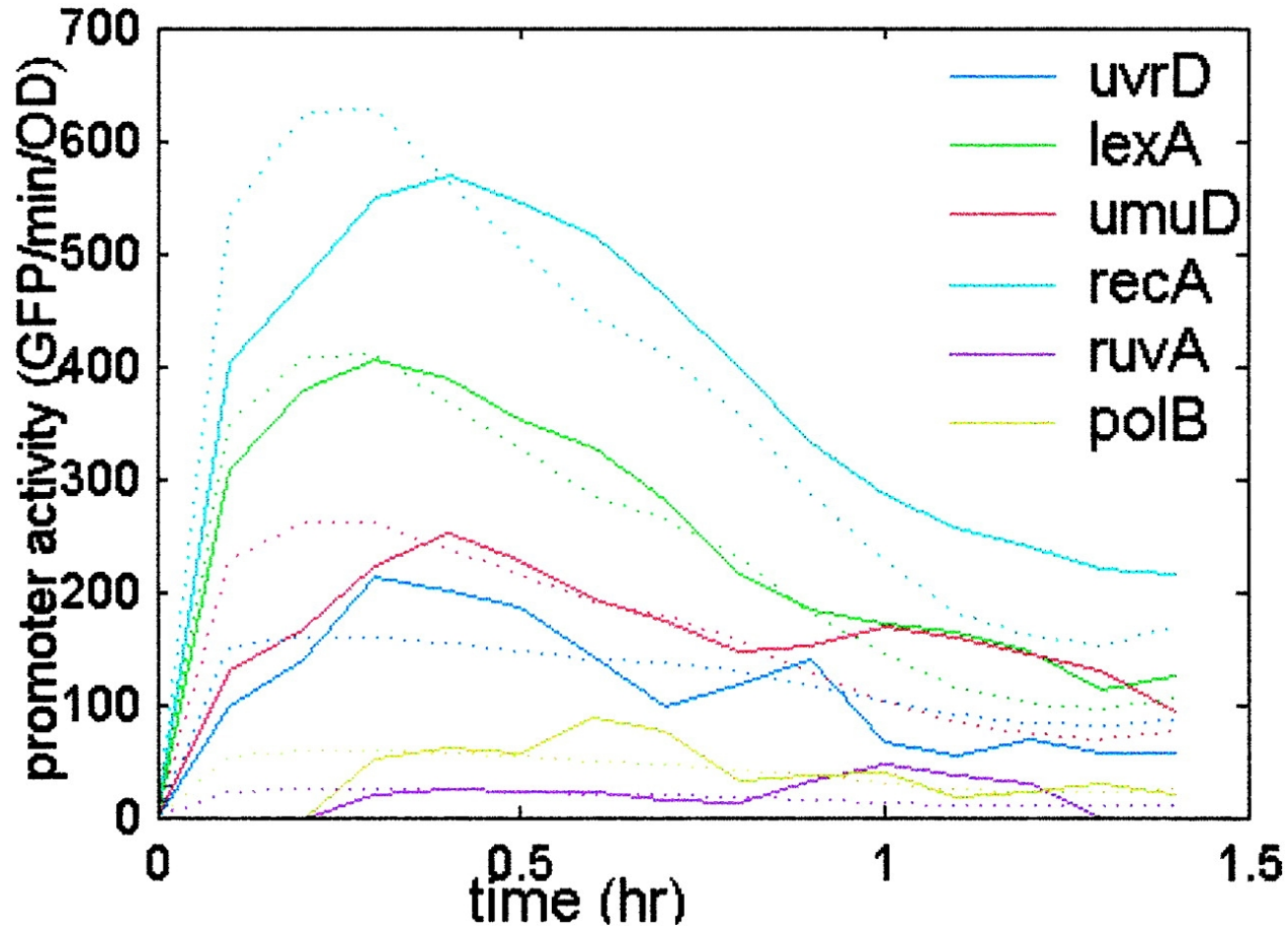
Le fitting permet de déduire les paramètres

Table 1. The effective kinetic parameters for the SOS system (\pm SD)

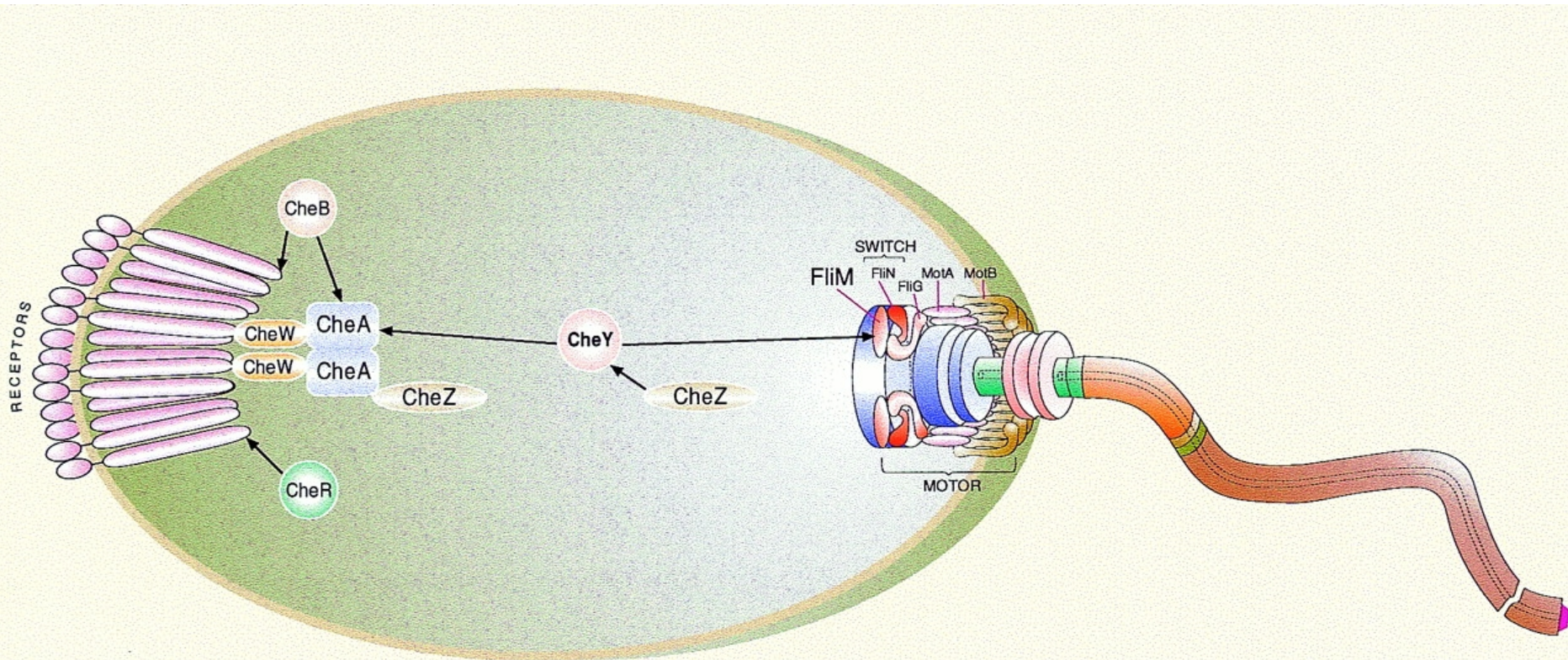
Gene	k	β	E	Function
<i>uvrA</i>	0.09 ± 0.04	$2,800 \pm 300$	0.14	Nucleotide excision repair
<i>lexA</i>	0.15 ± 0.08	$2,200 \pm 100$	0.10	Transcriptional repressor
<i>recA</i>	0.16 ± 0.07	$3,300 \pm 200$	0.12	Mediates LexA autocleavage, blocks replication forks
<i>umuD</i>	0.19 ± 0.1	330 ± 30	0.21	Mutagenesis repair
<i>polB</i>	0.35 ± 0.15	70 ± 10	0.31	Trans-lesion DNA synthesis, replication fork recovery
<i>ruvA</i>	0.37 ± 0.1	30 ± 2	0.22	Double-strand break repair
<i>uvrD</i>	0.65 ± 0.3	170 ± 20	0.20	Nucleotide excision repair, recombinational repair
<i>uvrY</i>	0.51 ± 0.25	300 ± 200	0.45	SOS operon of unknown function, additional roles in two-component signaling
<i>lacZ</i>	—	—	1.53	Unrelated to SOS system

E is the mean error for the promoter activity prediction (see *Methods*).

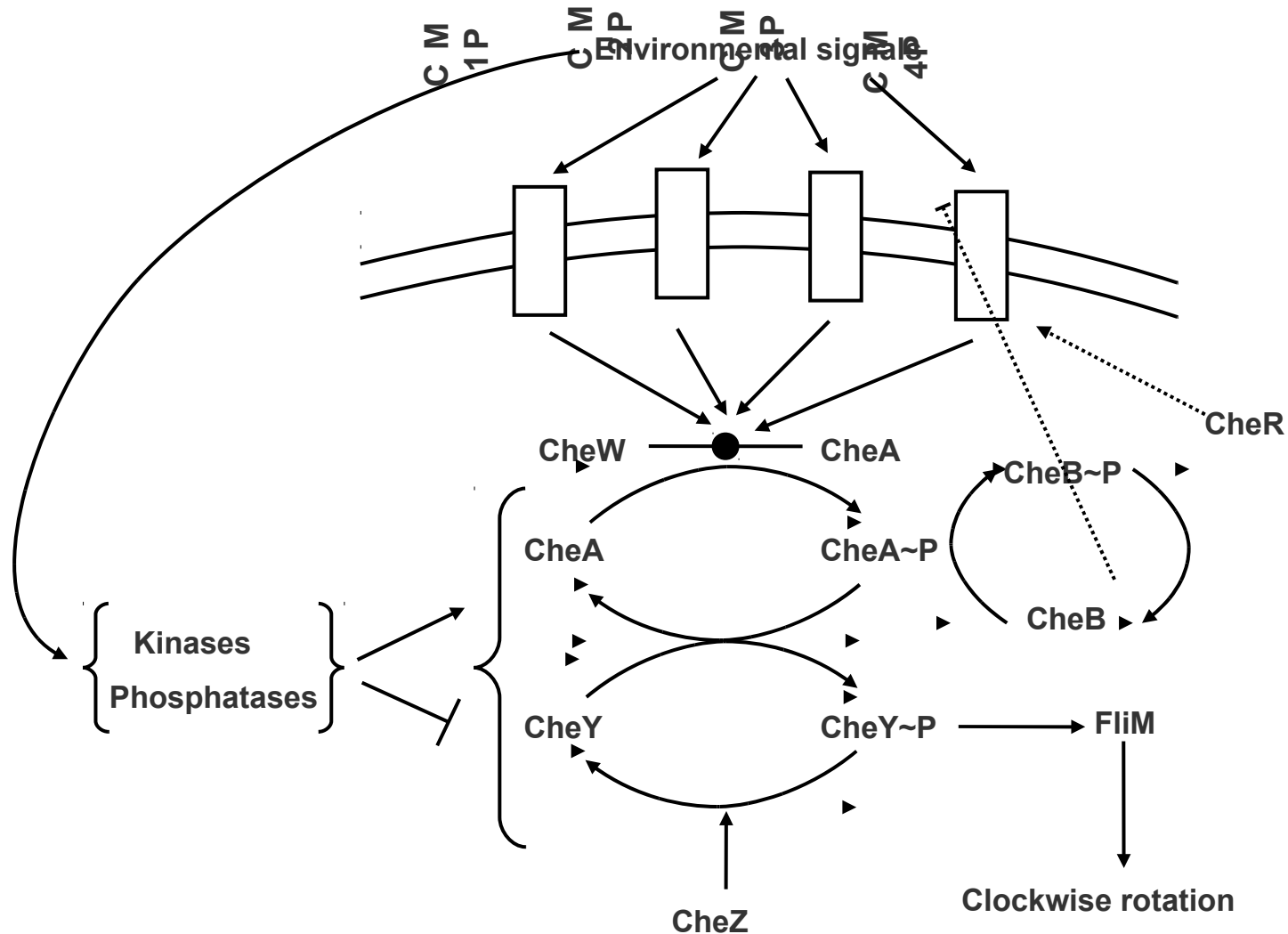
Les paramètres prédisent correctement le comportement du système



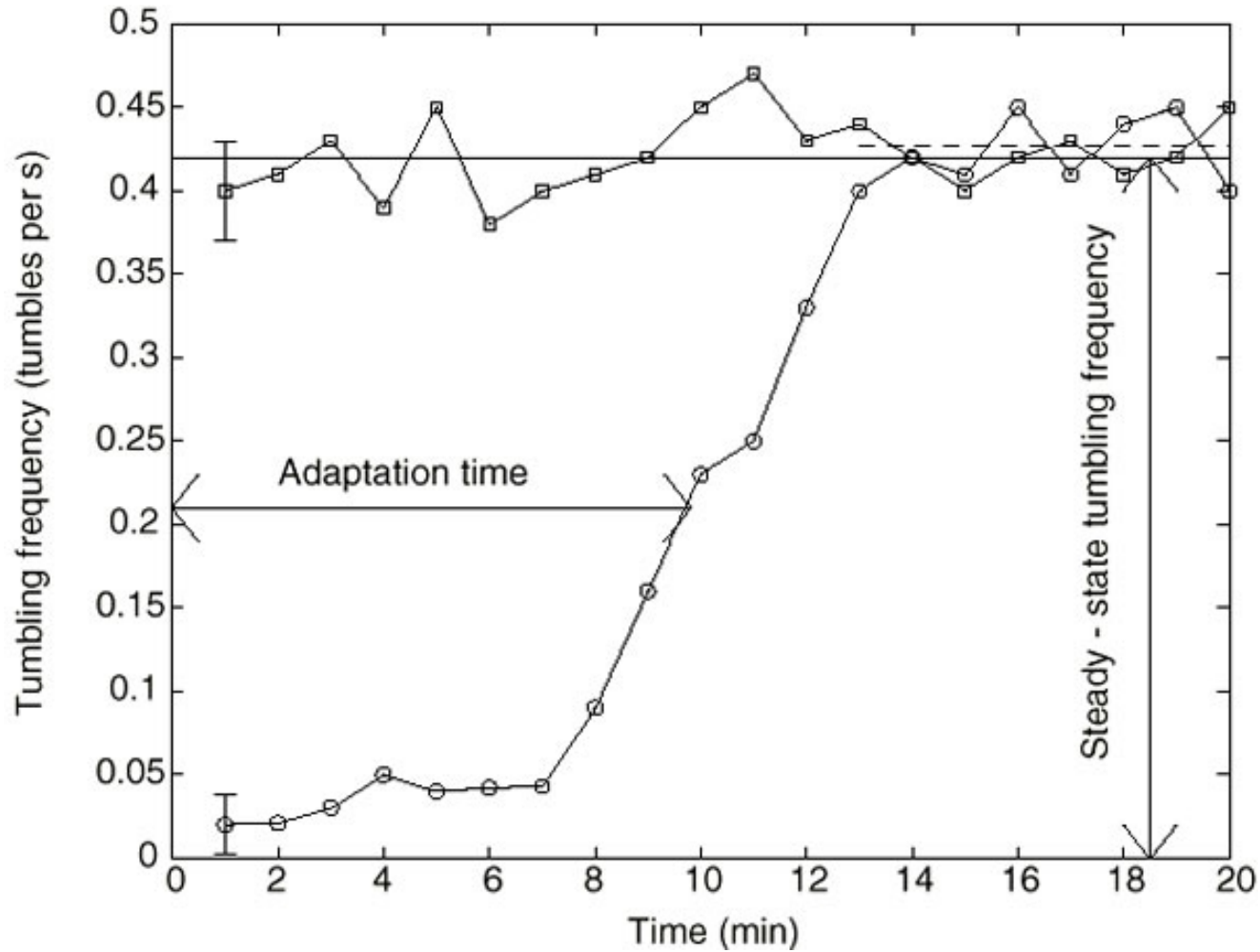
Bacterial chemotaxis



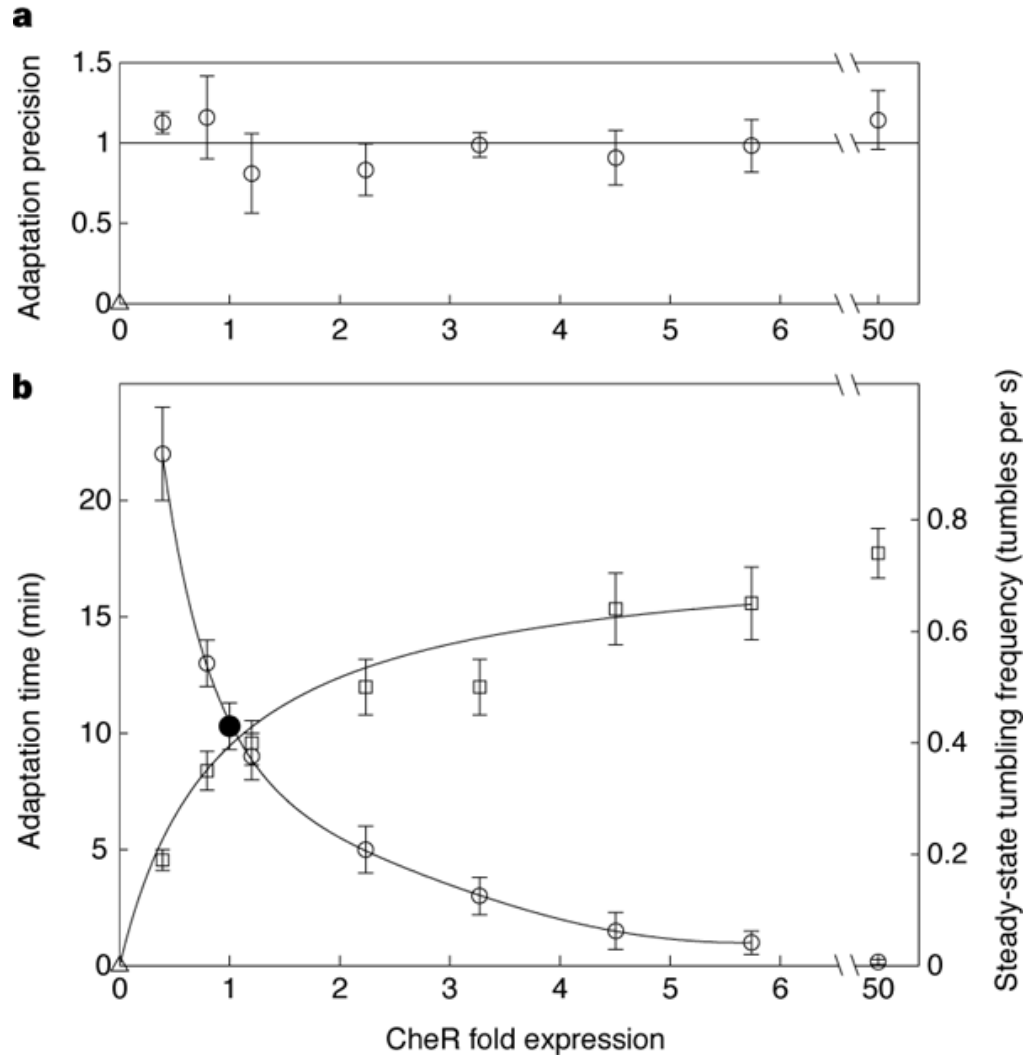
Basic chemotaxis regulation network



Robustness in bacterial chemotaxis



Robustness in bacterial chemotaxis

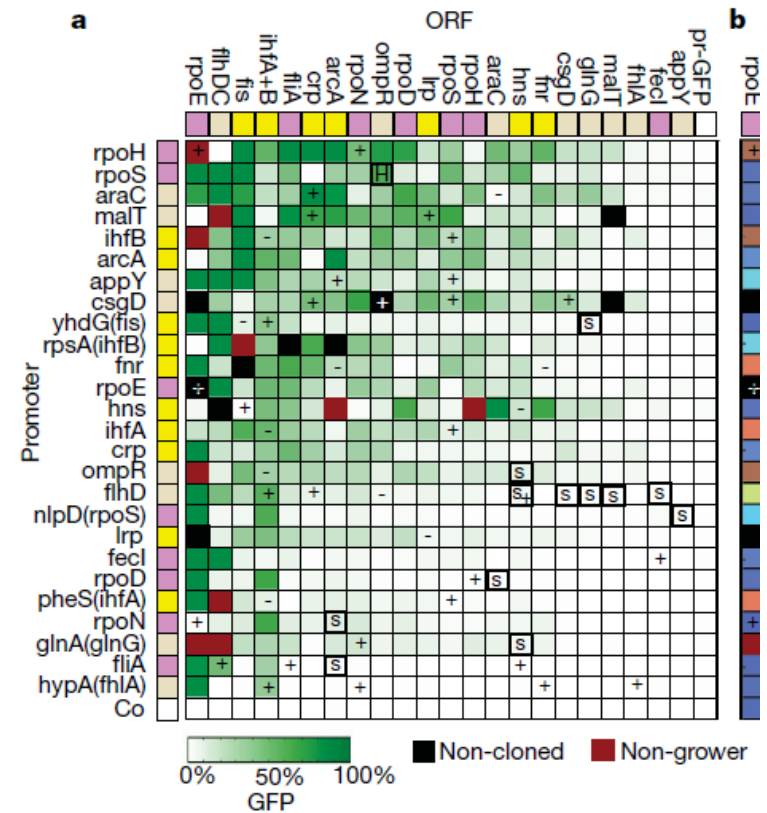
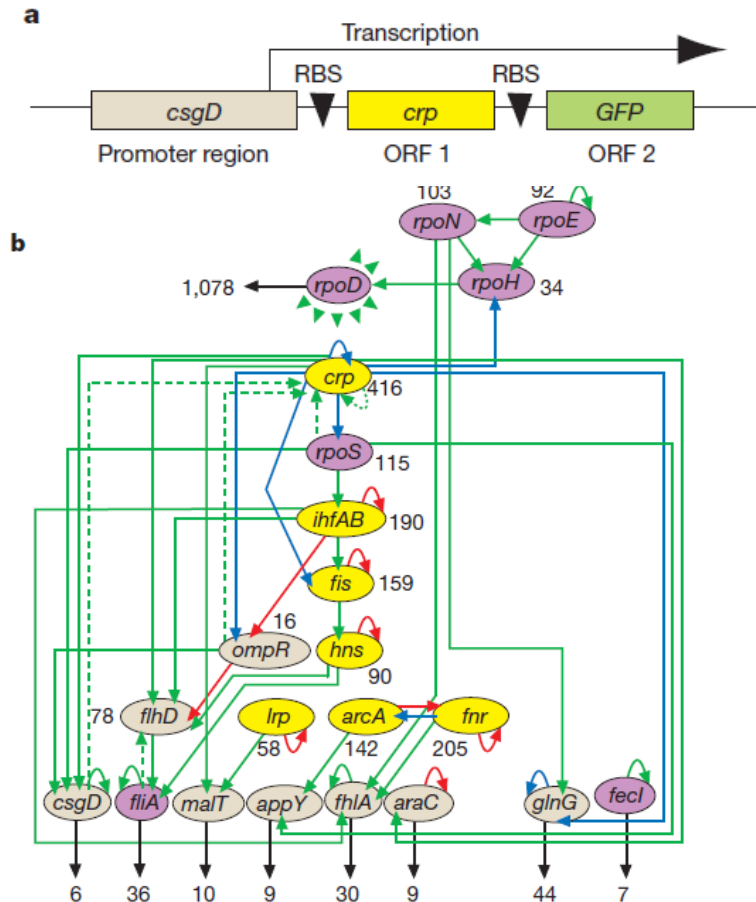


Robustness of adaptation precision

Protein varied	Fold expression	Strain background	Steady-state tumbling frequency (s^{-1})	Adaptation time (min)	Precision of adaptation
Wild type	1.0	Wild type	0.44 ± 0.03	10 ± 1	0.98 ± 0.05
CheB*	0.4 ± 0.1	$\Delta cheB$	0.66 ± 0.05	7 ± 1	0.98 ± 0.12
CheB*	12 ± 3	$\Delta cheB$	0.14 ± 0.02	15 ± 1	1.09 ± 0.11
CheBc	~ 1	$\Delta cheB$	0.74 ± 0.06	9 ± 2	0.90 ± 0.13
CheY	0.2 ± 0.1	$\Delta cheY,Z$	0.24 ± 0.04	11 ± 3	1.04 ± 0.08
CheZ	0	$\Delta cheZ$	1.6 ± 0.1	10 ± 2	1.1 ± 0.14
Tar, Tap, CheR,B,Y,Z	5 ± 2	Wild type	0.30 ± 0.06	3 ± 1	1.04 ± 0.07

Rewiring a cellular network

- Almost all network variants grow well



New connections provide “improved” properties

- Selection of advantageous connections by repeated exposure to a particular stress

