

# Regulation of Gene Expression

---

- Hans Geiselmann  
[hans.geiselmann@ujf-grenoble.fr](mailto:hans.geiselmann@ujf-grenoble.fr)

# Measuring DNA-protein interactions

---

- Gel retardation
- Spectroscopic methods
- Footprinting

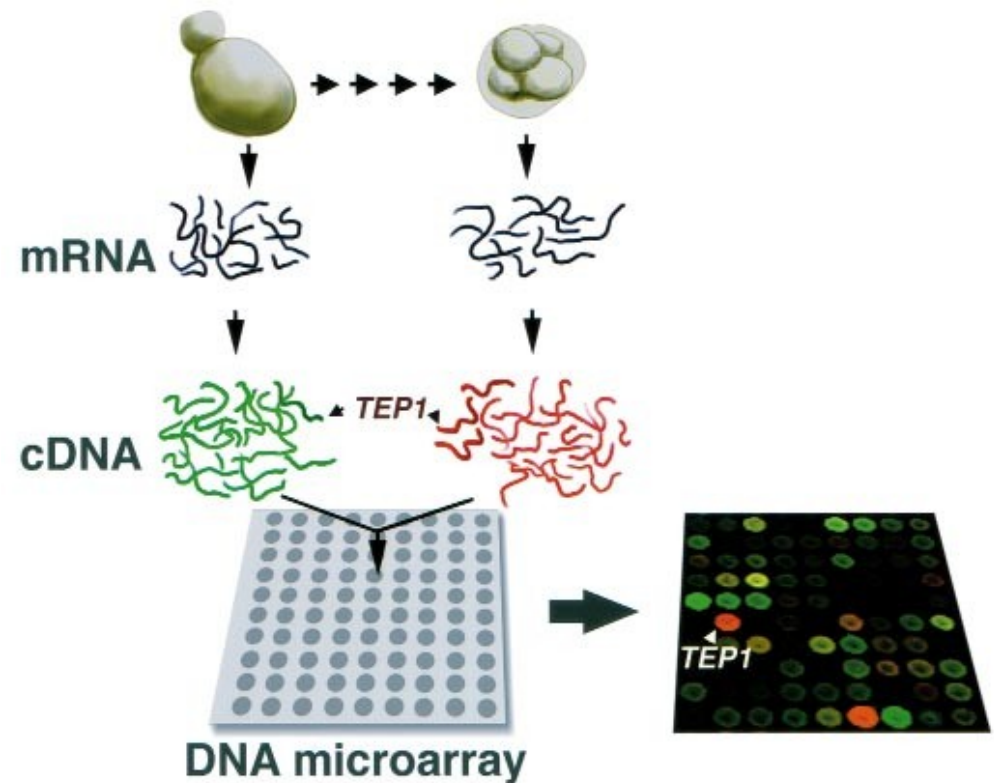
# Global methods

---

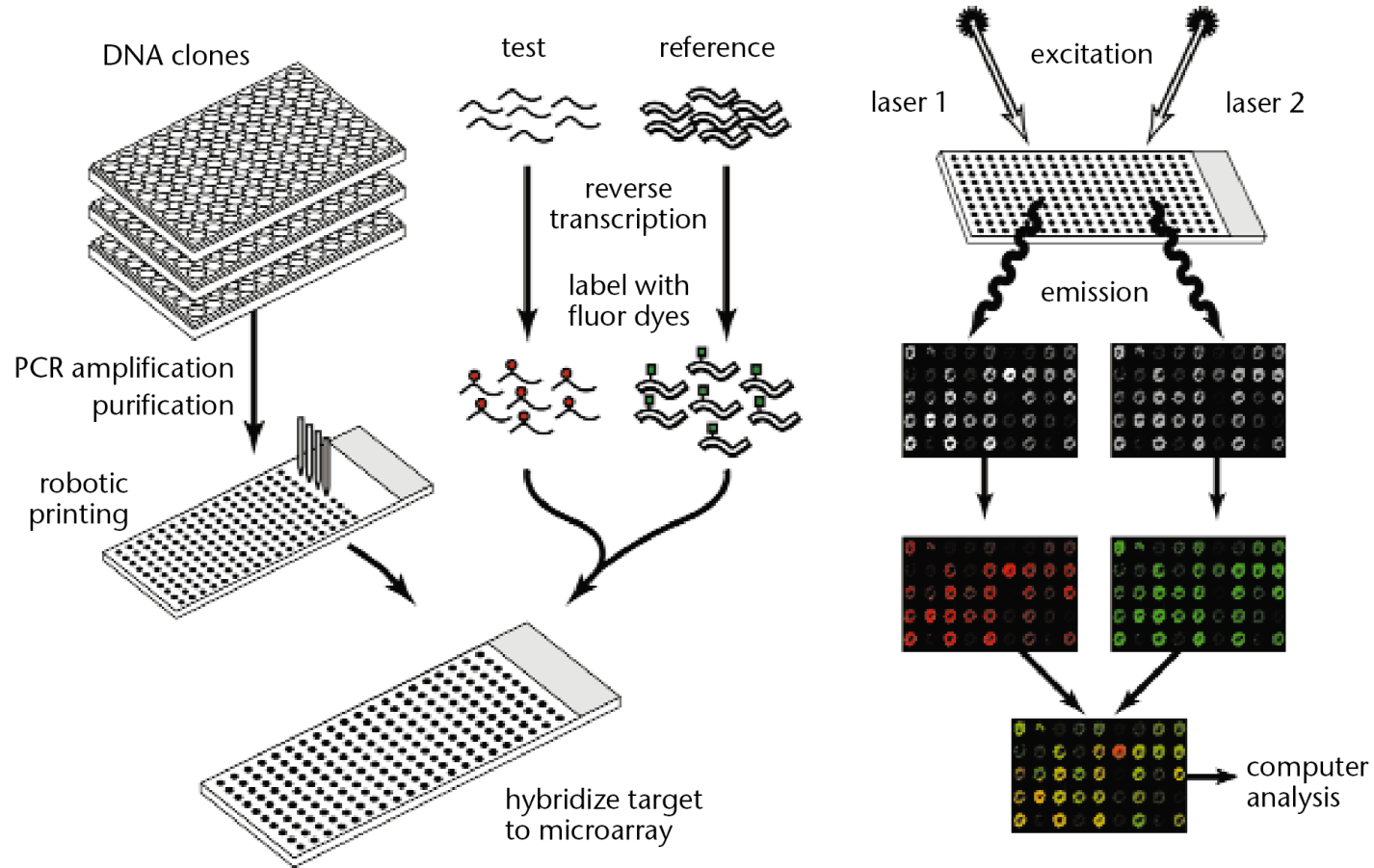
- DNA chips
- ChIP – Chip
- proteomics

# Détection de changements d'expression génique au cours du développement

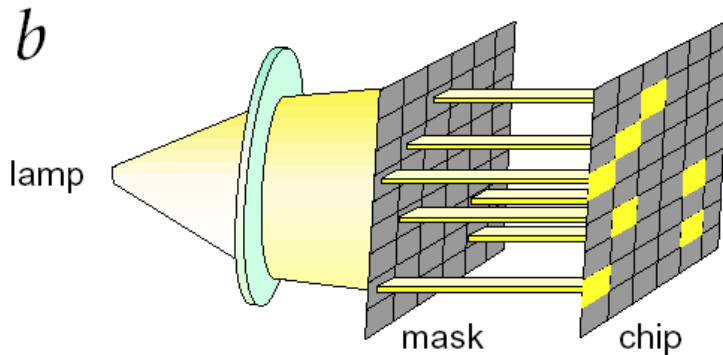
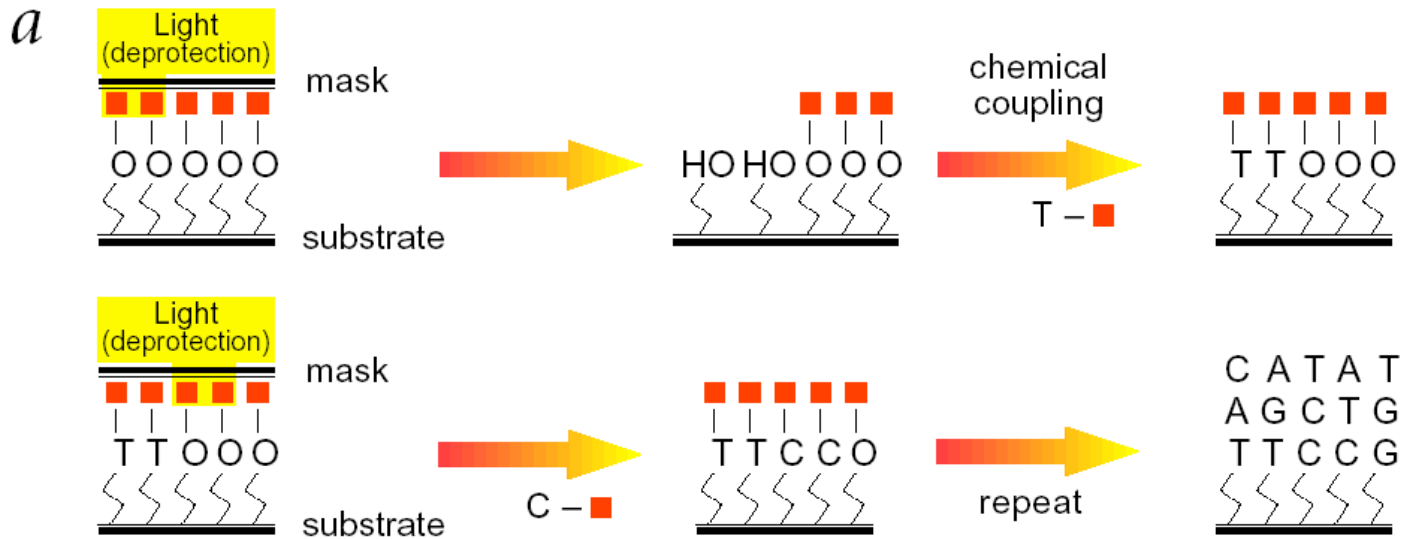
- Immobiliser une sonde sur un support solide
- Les cDNA des différents échantillons sont marqués avec deux fluorophores différentes
- Hybrider avec des cDNA
- Lecture de la fluorescence
- Le rapport de fluorescence correspond au facteur d'induction



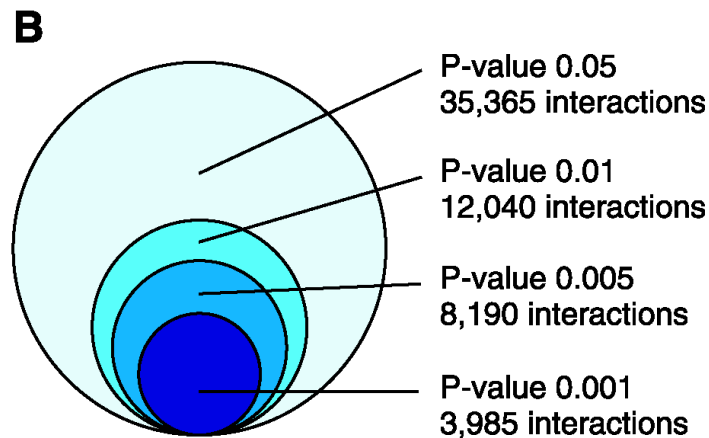
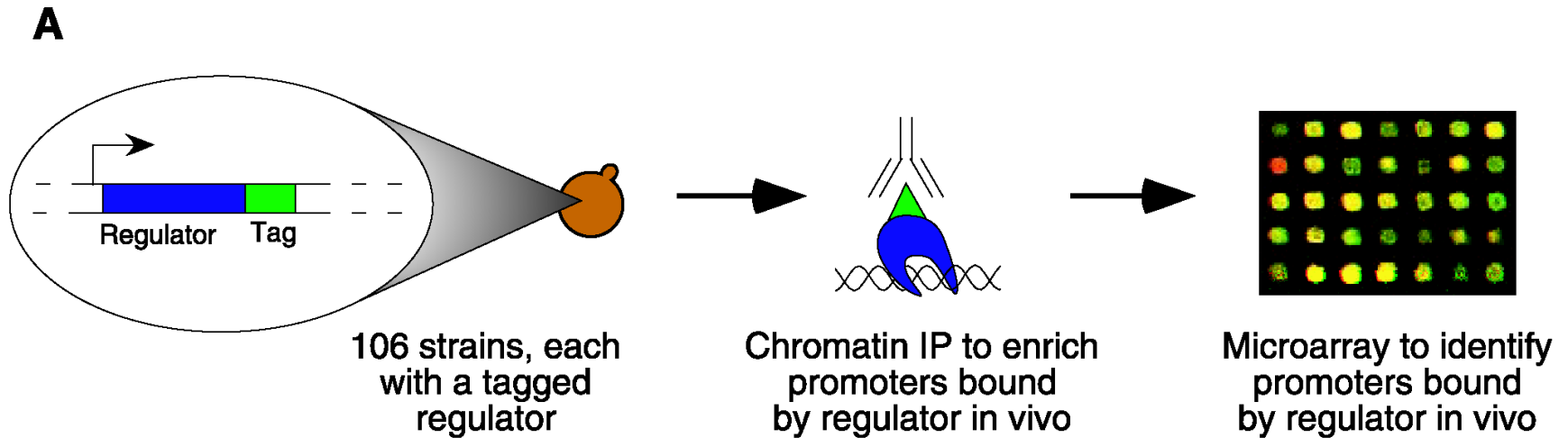
# Principe des puces à ADN



# Puces à ADN par synthèse in situ

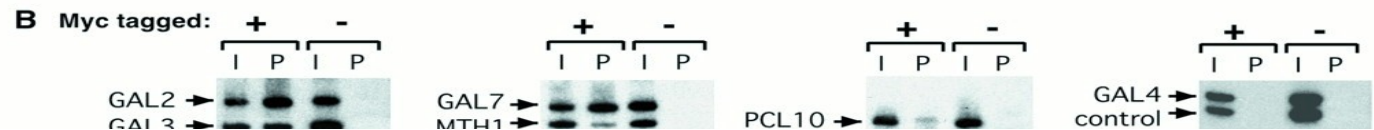
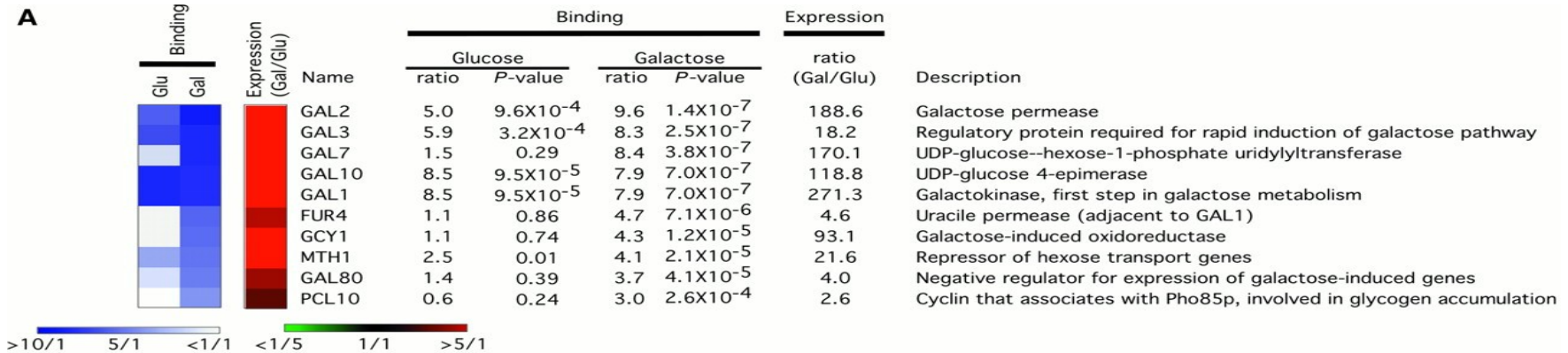
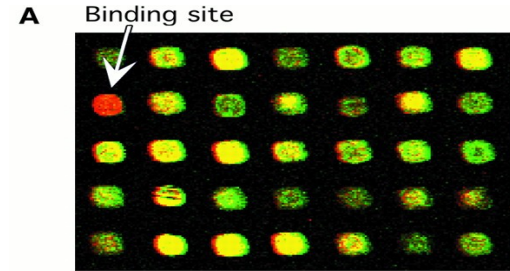


# Identification des sites de liaison de régulateurs par ChIP (Chromatin Immuno-Precipitation)



# Variations sur l'analyse du transcriptome

- Localiser les sites de liaison de facteurs de transcription





# Variations sur l'analyse du transcriptome

---

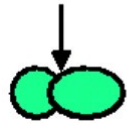
- Utilisation de données d'expression pour déterminer l'ordre d'assemblage d'un complexe macromoléculaire
- Il y a une forte corrélation entre l'ordre temporelle d'expression des gènes et l'ordre d'assemblage du complexe

# Assembly of the flagellar motor

## Class 1

*flhDC*

FlhD-FlhC

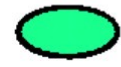


Basal body + hook  
(BBH)

## Class 2

*fliFGHJK, fliLMNOPQR, fliE, flhBAE,*  
*flgBCDEFGHIJ, fliAZY, flgAMN,*

FlhA



FlgM

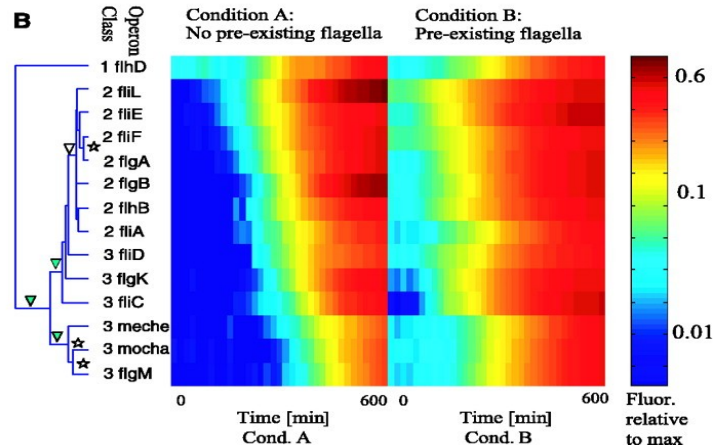
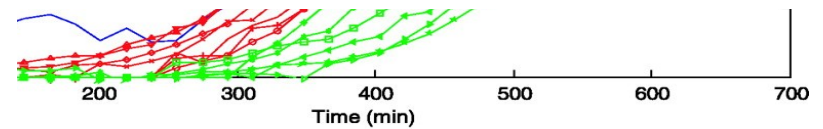
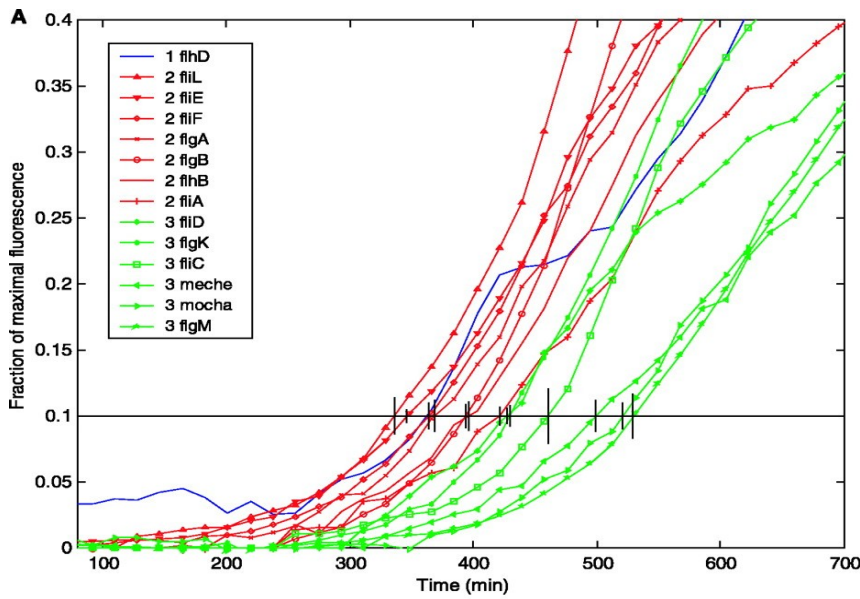


## Class 3

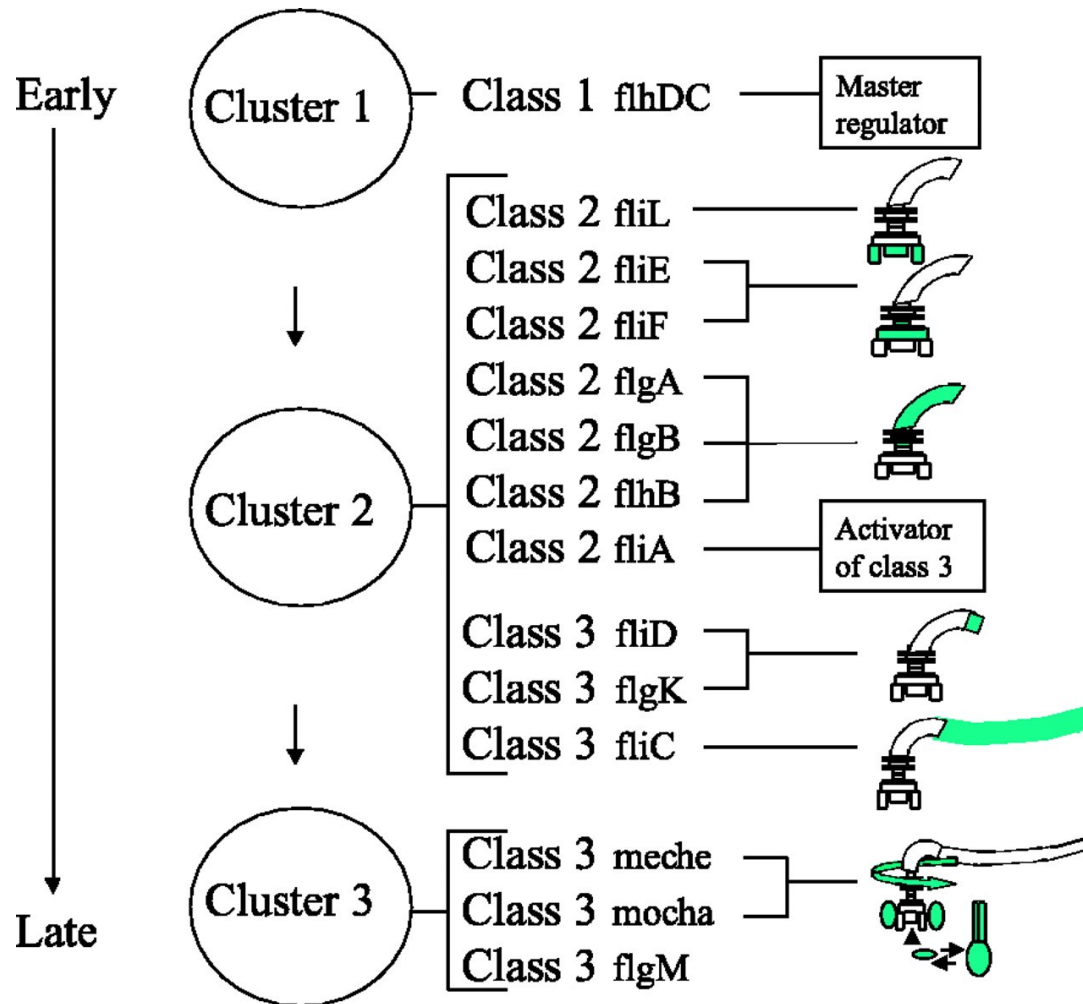
3a: *flgKL, fliDST, flgMN*  
3b: *fliC, meche (tar tap cheRBYZ),*  
*mocha (motABcheAW)*

Flagellum +  
chemotaxis system

# Temporal pattern of gene expression correlates with order of assembly



# Temporal pattern of gene expression correlates with order of assembly



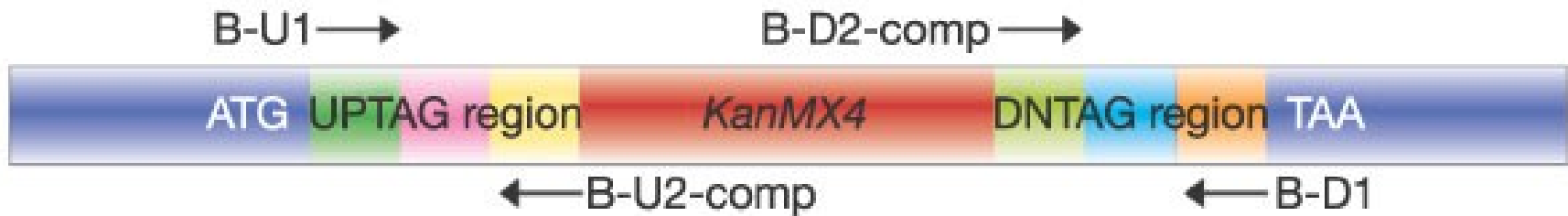
# Analyse fonctionnel systématique

---

- Exemple de la levure
  - Déleter systématiquement tous les gènes
  - Mesurer l'effet sur la transcription
  - Mesurer un paramètre fonctionnel

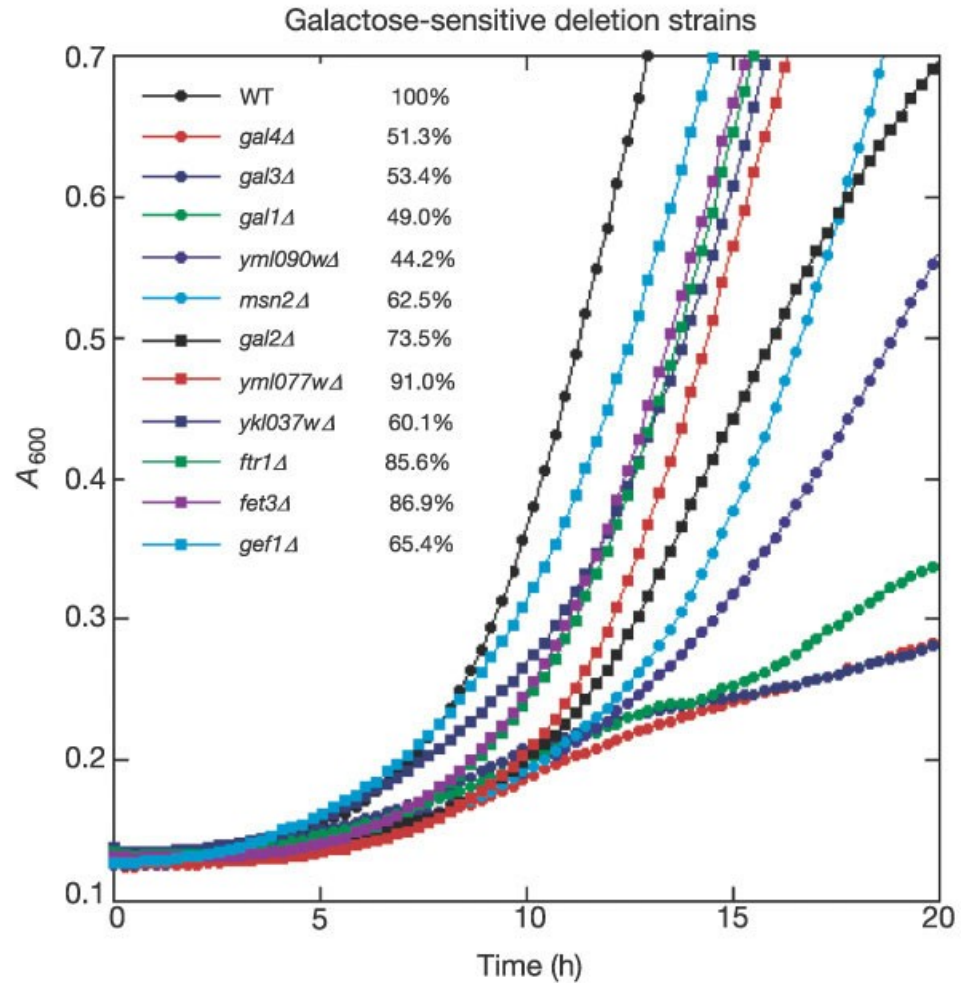
# La délétion du gène introduit également une étiquette unique

---

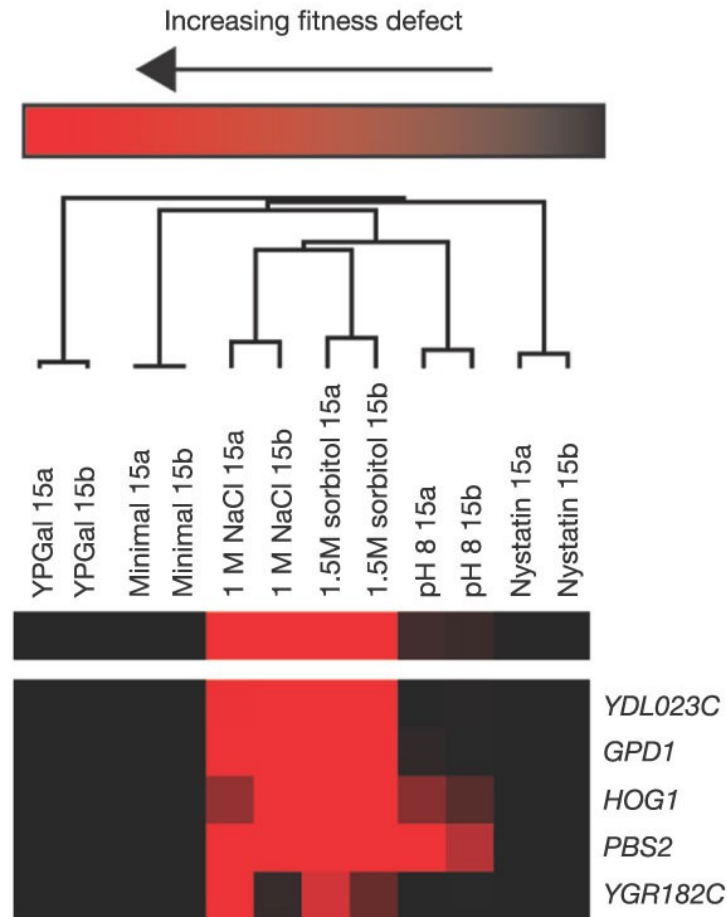


# Mesure de l'expression et d'un phénotype

- **Délétion systématique de gènes**
- **Mesure de l'expression génique par puces à ADN**
- **Mesure d'un paramètre fonctionnel (croissance)**



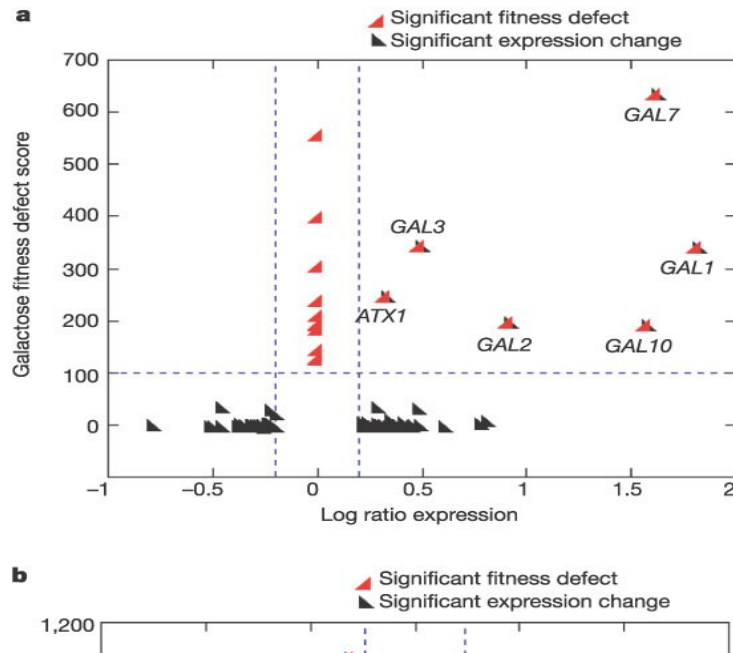
# Regrouper les gènes qui provoquent un même phénotype



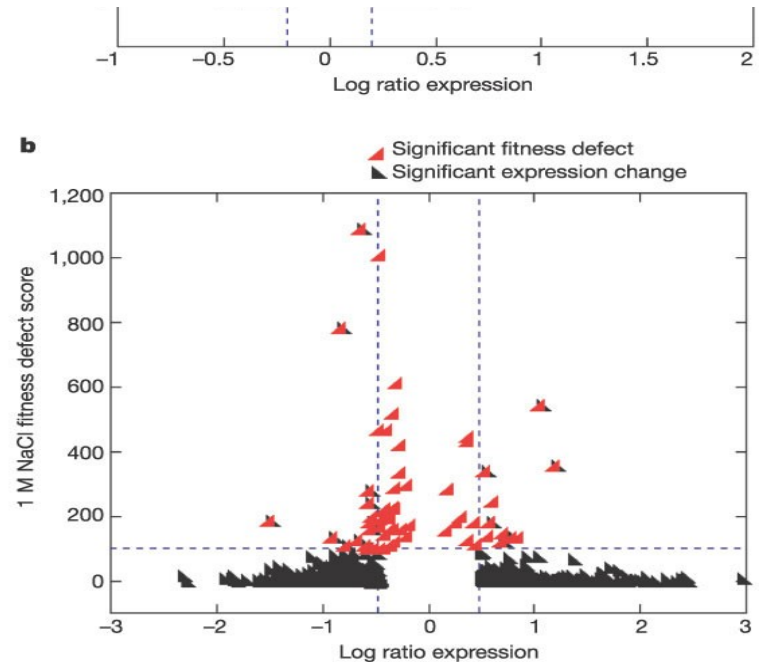


# Corrélation entre expression et phénotype

## Croissance sur galactose

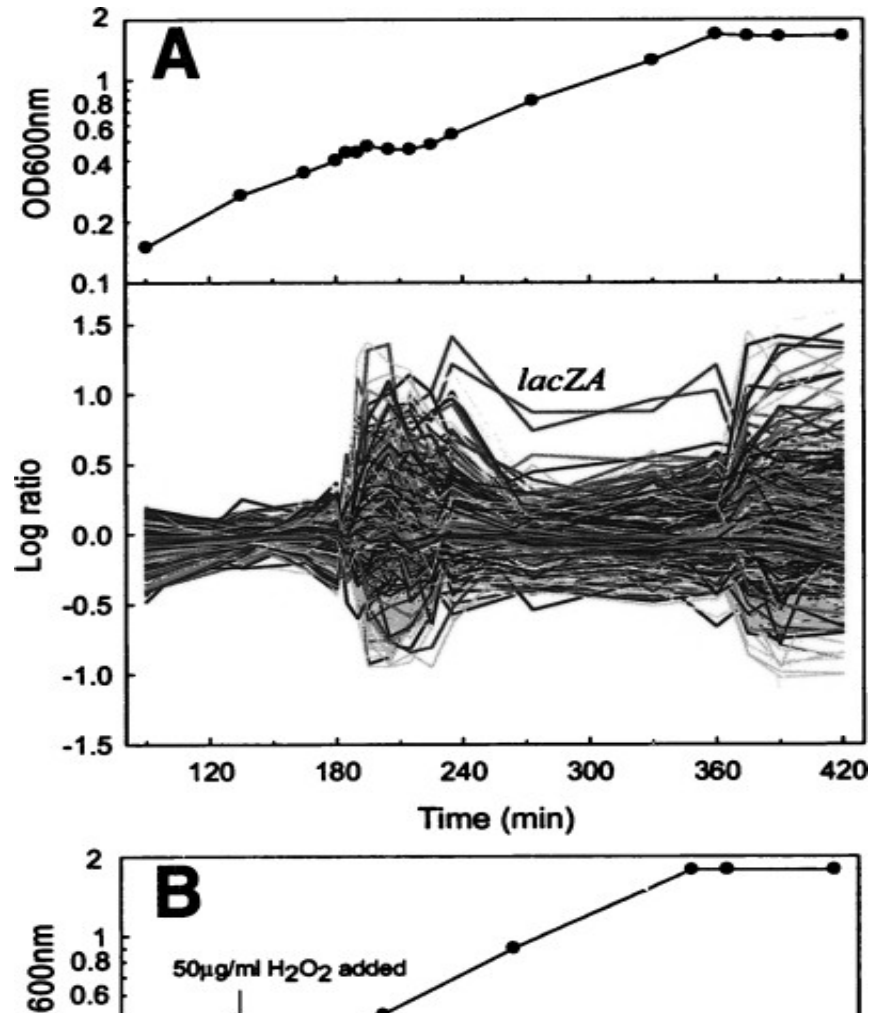


## Croissance à haute osmolarité



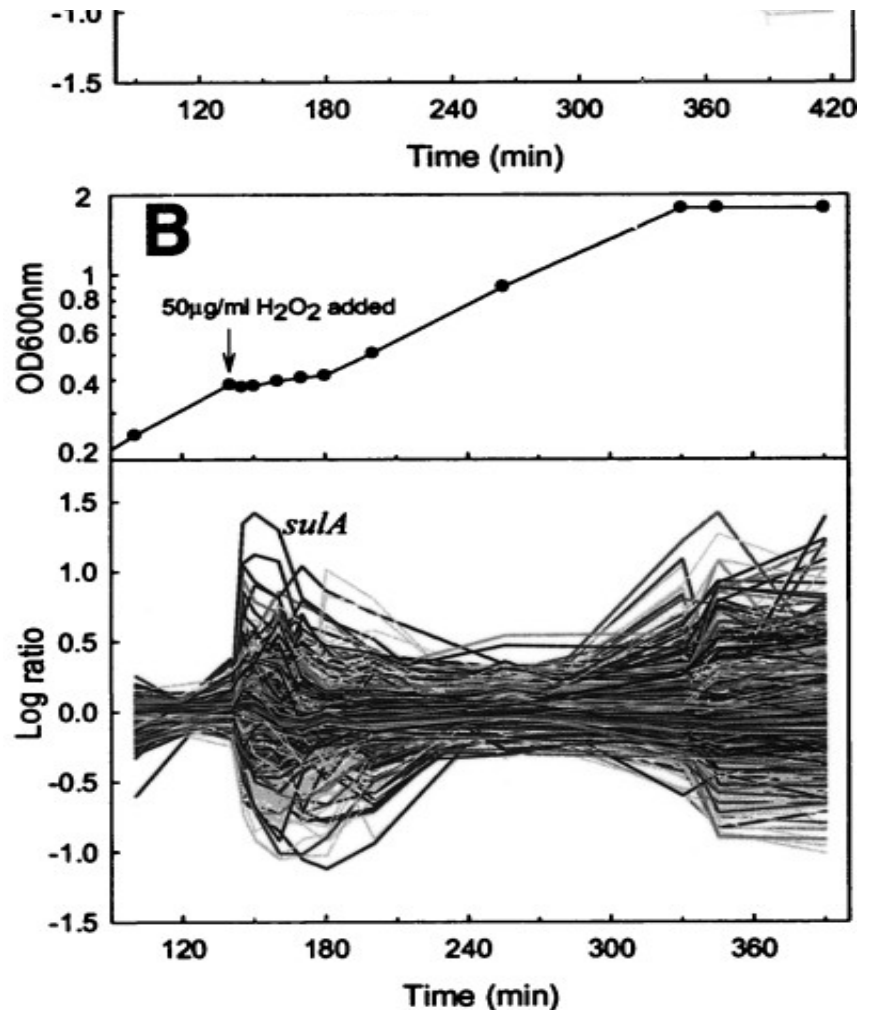
# Croissance d'*Escherichia coli* analysée par puces à ADN

- Diauxie glucose – lactose
  - Courbe de croissance
  - Expression de tous les gènes d'*Escherichia coli*



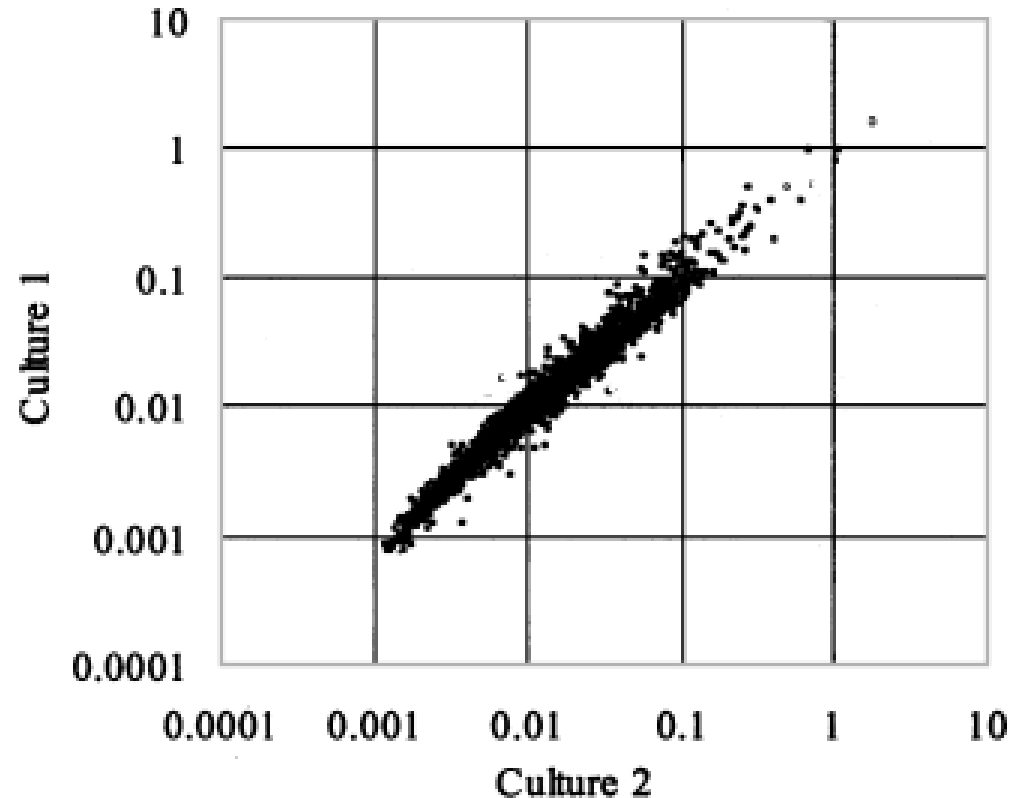
# Croissance d'*Escherichia coli* analysée par puces à ADN

- Stress oxydatif
  - Courbe de croissance
  - Expression de tous les gènes d'*Escherichia coli*
- Réponse similaire à celle de la diauxie



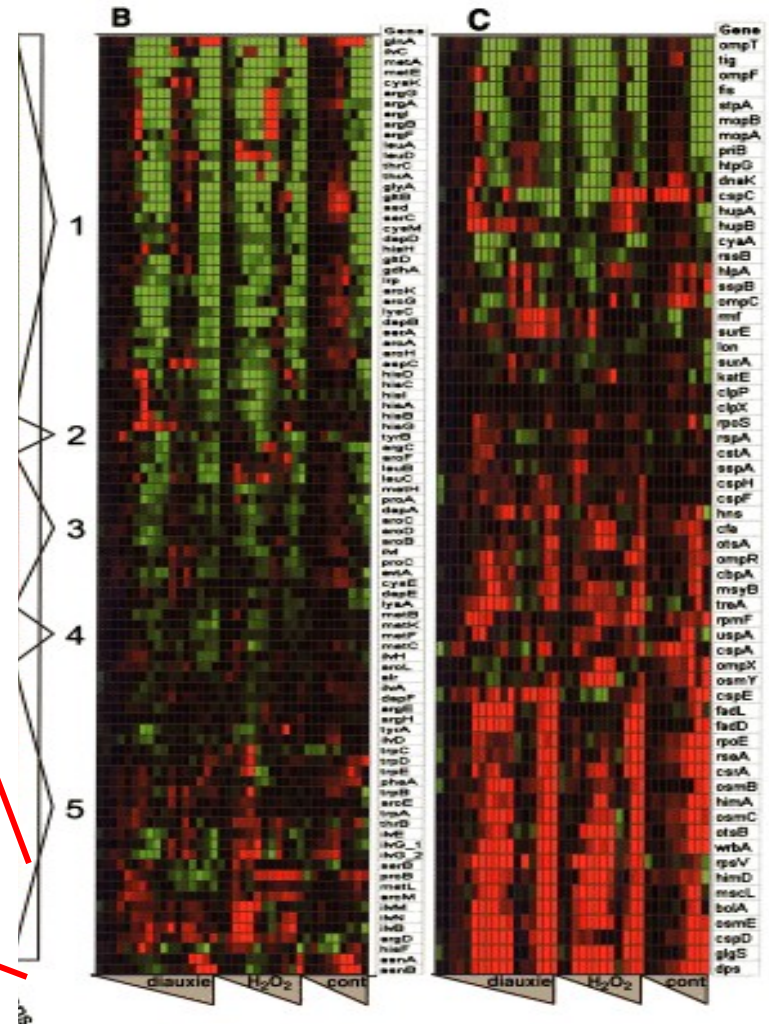
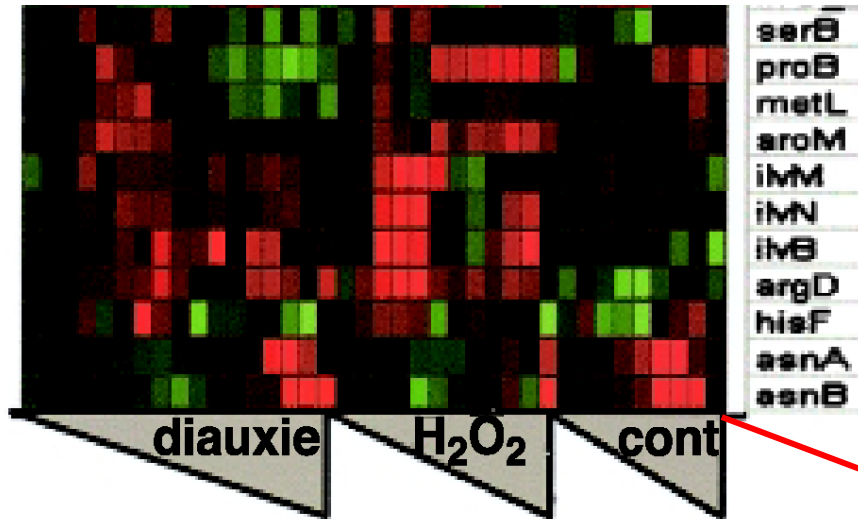
# Reproductibilité de la mesure

- Comparaison de la culture en diauxie et en stress oxydant



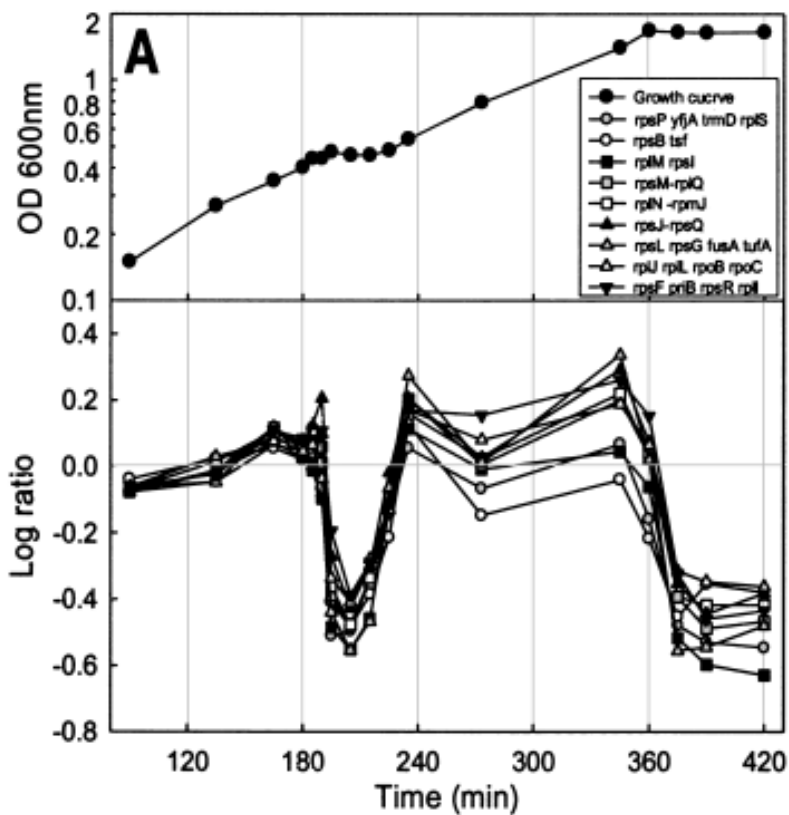
# Regroupement des gènes possédant un pattern d'expression similaire

- La biosynthèse de seulement quelques acides aminés est activée pendant l'arrêt de la croissance

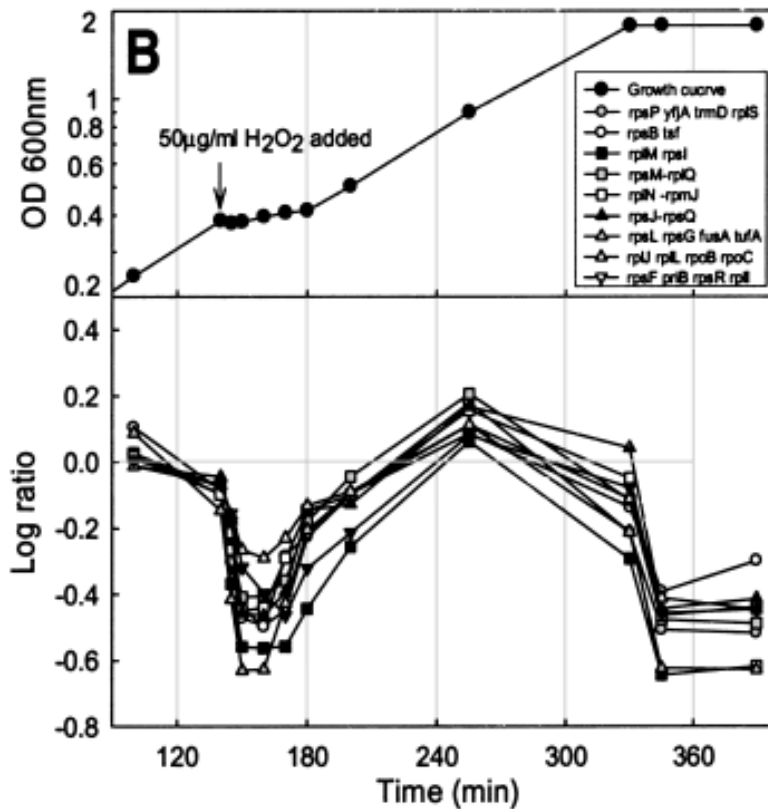


# L'expression de l'appareil de traduction est fortement diminuée à l'arrêt de la croissance

## Diauxie



## Stress oxydant

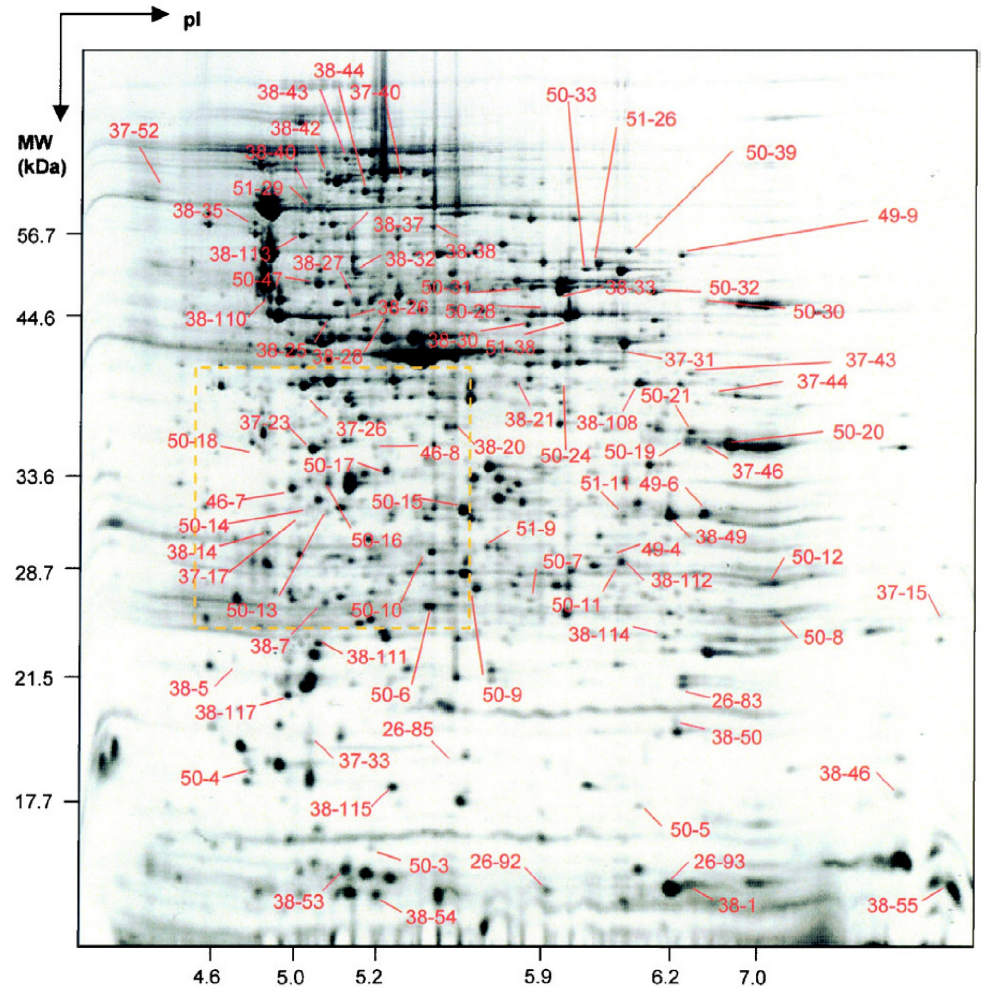




# Protéomique

- Séparation des protéines sur un gel 2D
- Identification des spots par MALDI-TOF MS

Gel 2D d'*Escherichia coli*  
Croissance en LB

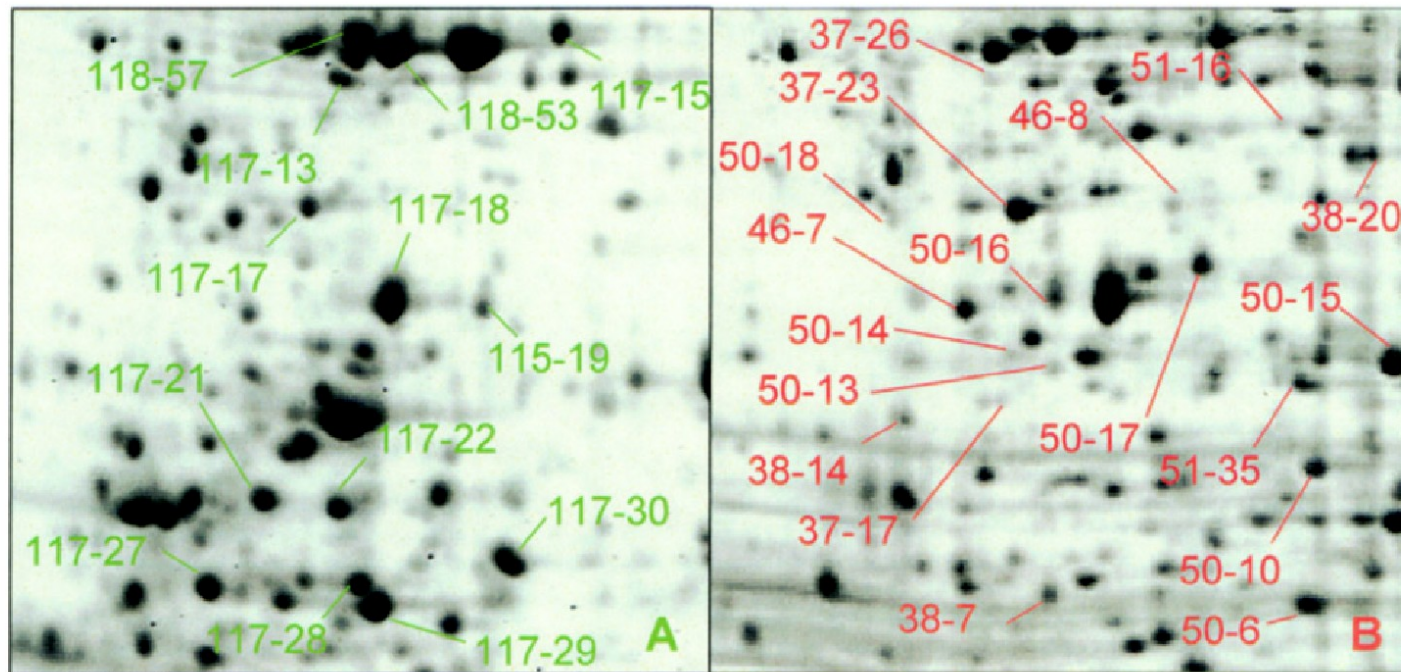


# Expression différentielle mesurée par la protéomique

- Modification de l'expression de protéines chez *Escherichia coli* en fonction du milieu de culture

Milieu minimal

Milieu riche





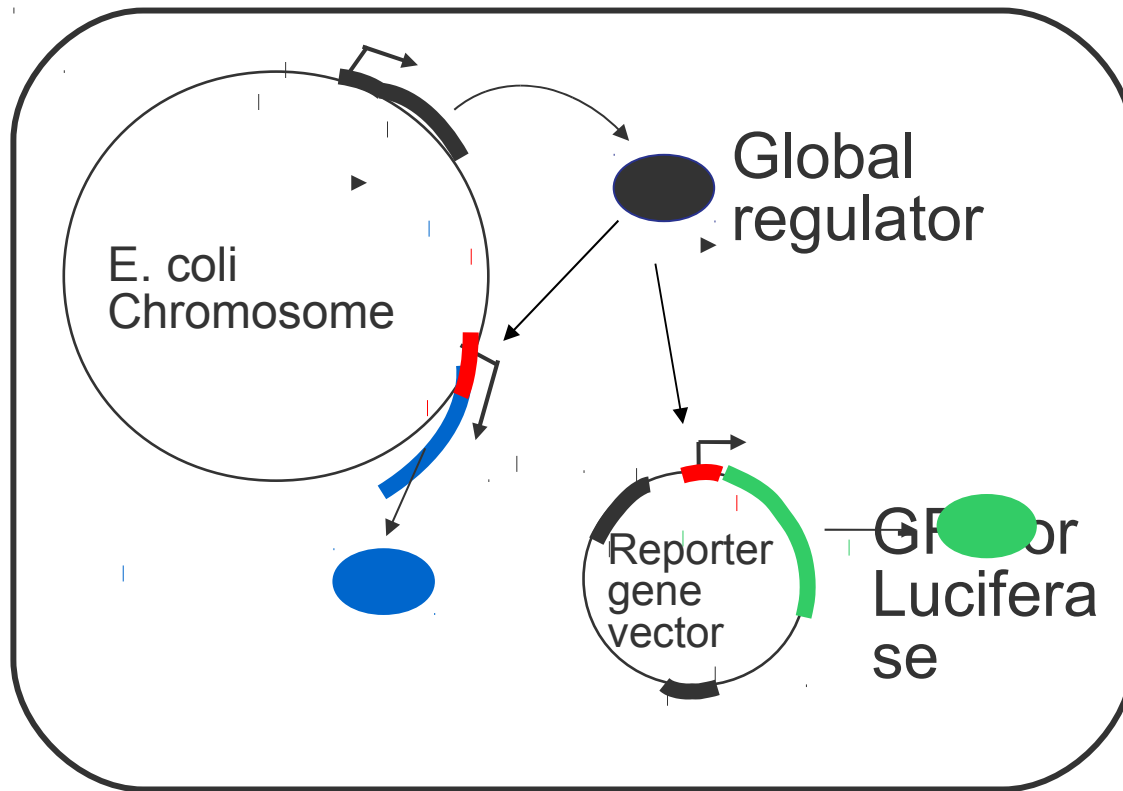
# Dynamical measurements

---

- Reporter genes
  - population measurements
  - single-cell measurements

# Experimental validation of model predictions

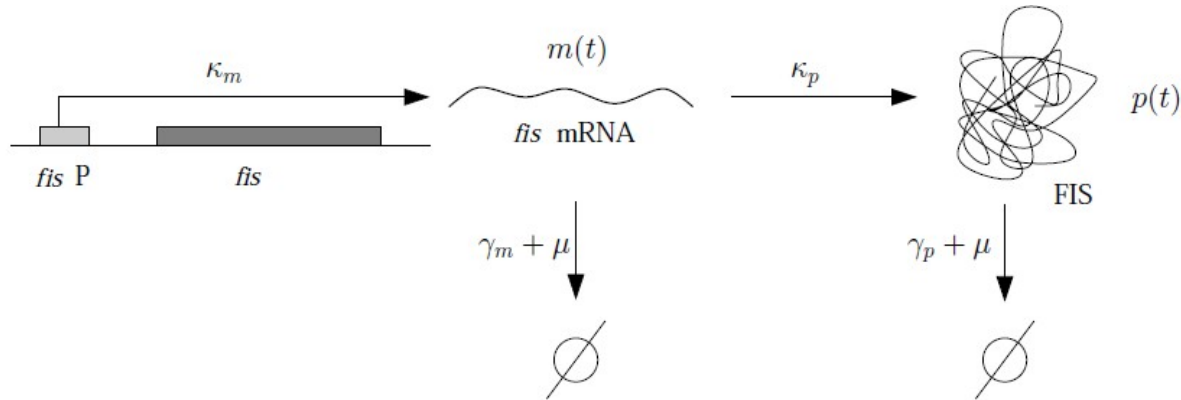
- Experimental validation of model predictions requires real-time monitoring of gene expression



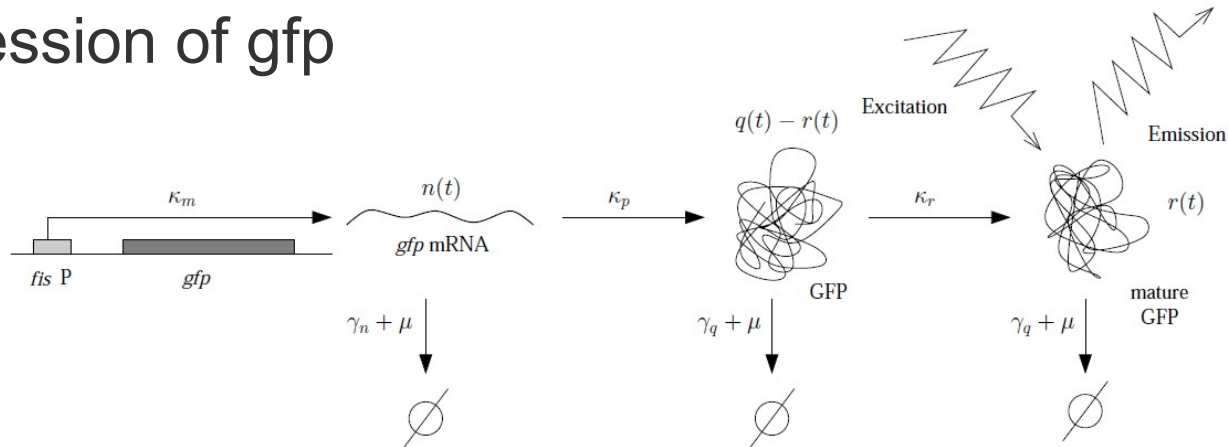
- The expression of the **reporter gene** reflects the expression of the **target gene**.

# Host protein and reporter genes

- Expression of the host protein

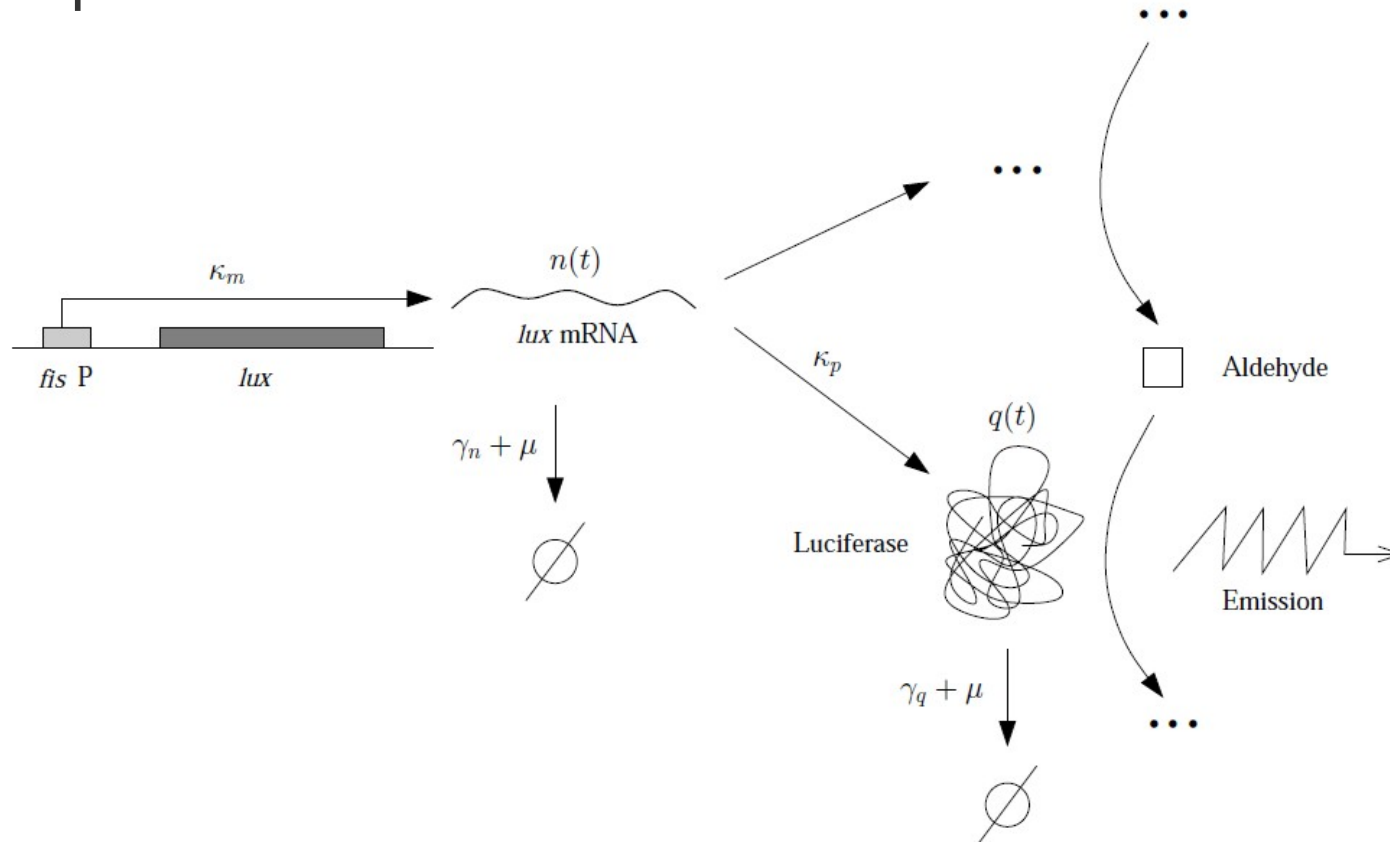


- Expression of gfp



# Host protein and reporter genes

- Expression of luciferase

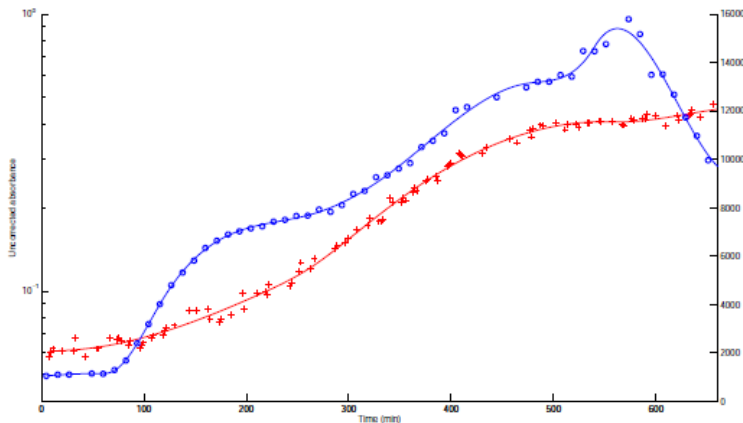


# Acquisition and treatment of raw data

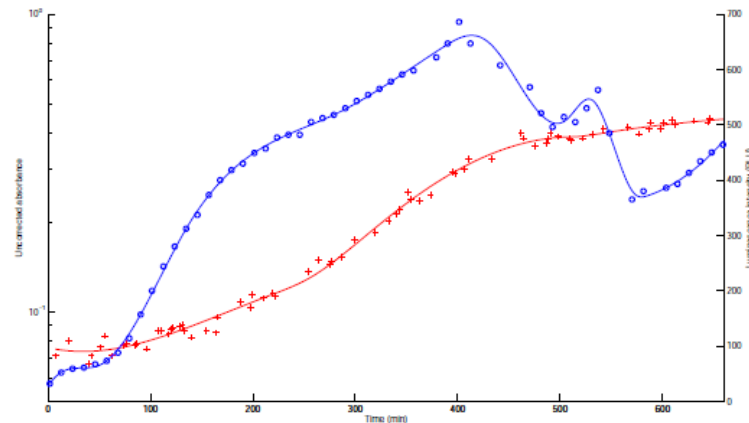
## □ Data acquisition with automated microplate reader

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

gfp



Luciferase



## □ Treatment of raw data

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

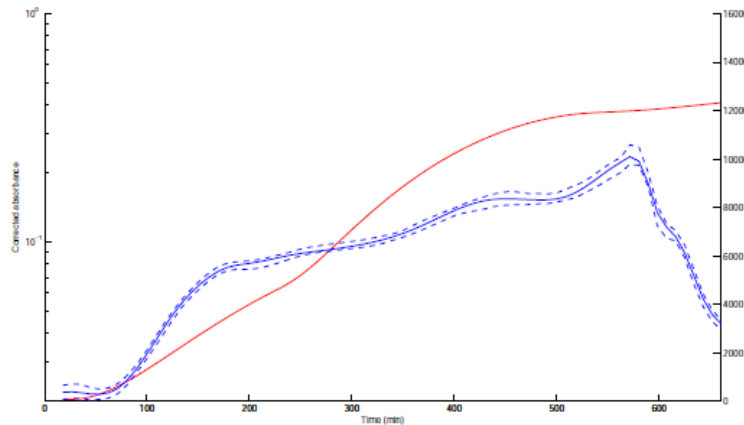
de Jong et al, BMC Systems Biology, 2010

# Confidence intervals

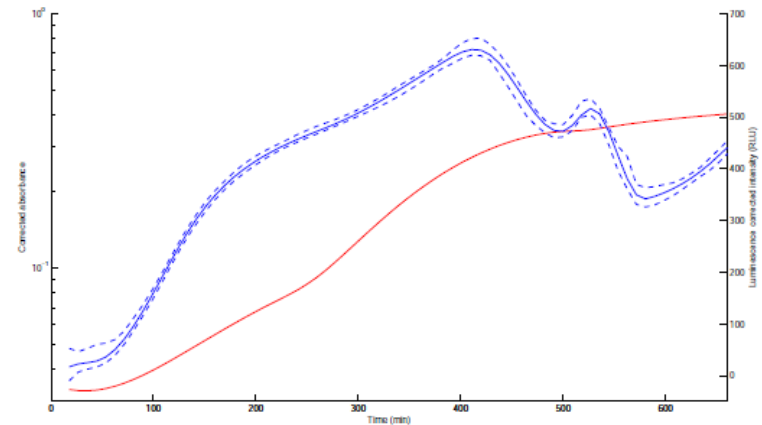
□ Calculate the confidence intervals of the corrected data

□ **Bootstrap method**

gfp



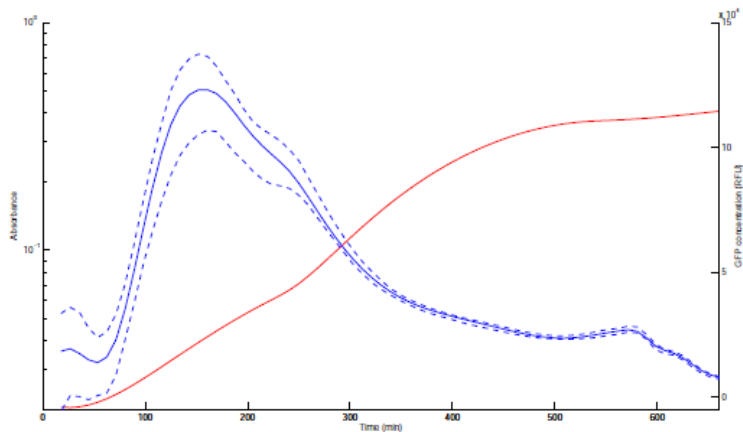
Luciferase



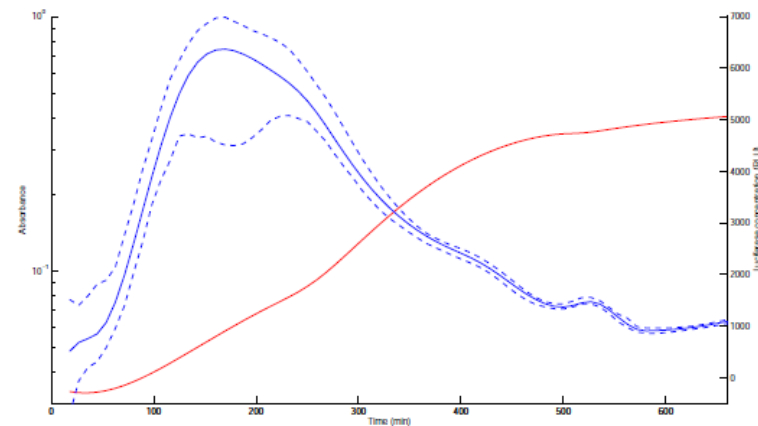
# Concentration of the reporter proteins

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

gfp



Luciferase

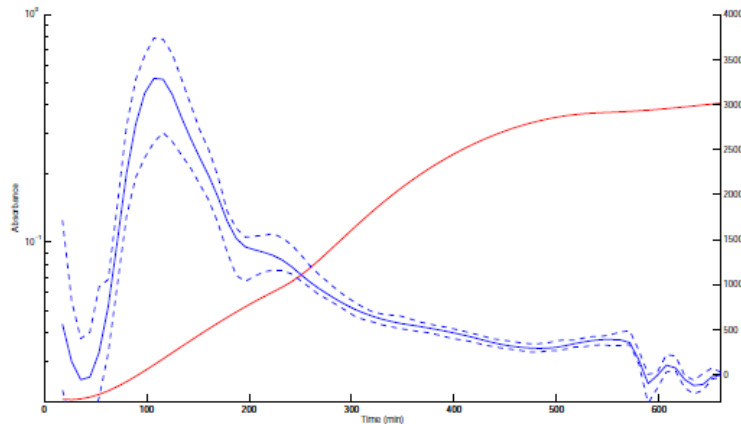


Reporter concentrations

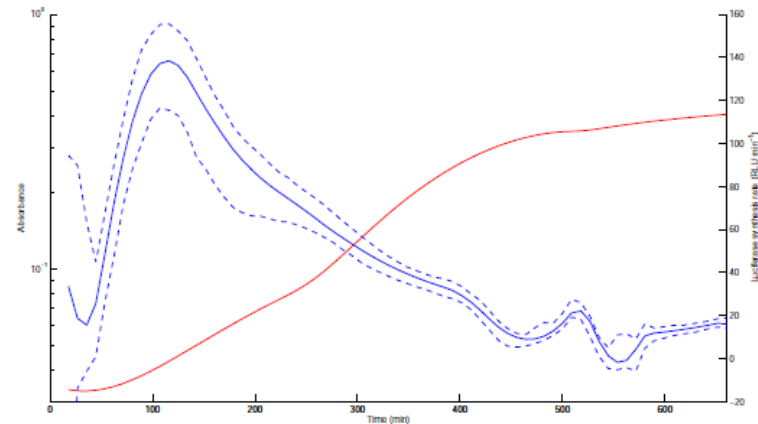
# Concentration of the reporter proteins

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

gfp



Luciferase



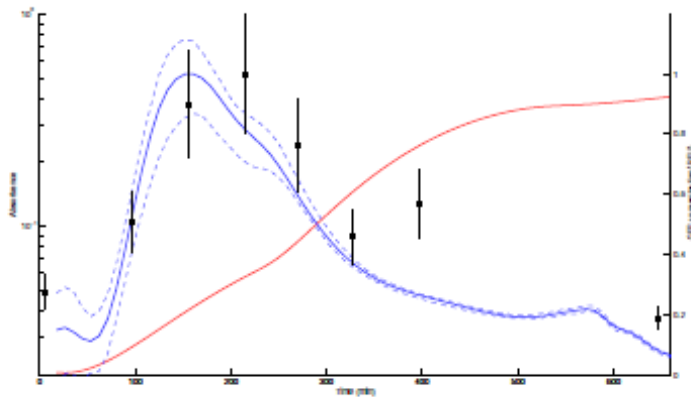
Synthesis rates



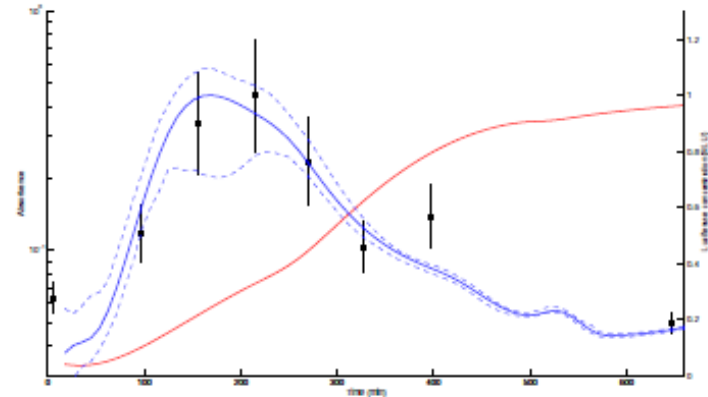
# Concentration of the host protein

Comparison between the predicted and measured concentration of the host protein

gfp

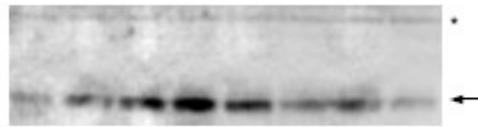


Luciferase

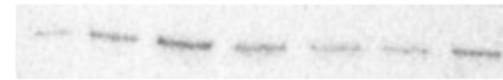


# Concentration of the mRNA

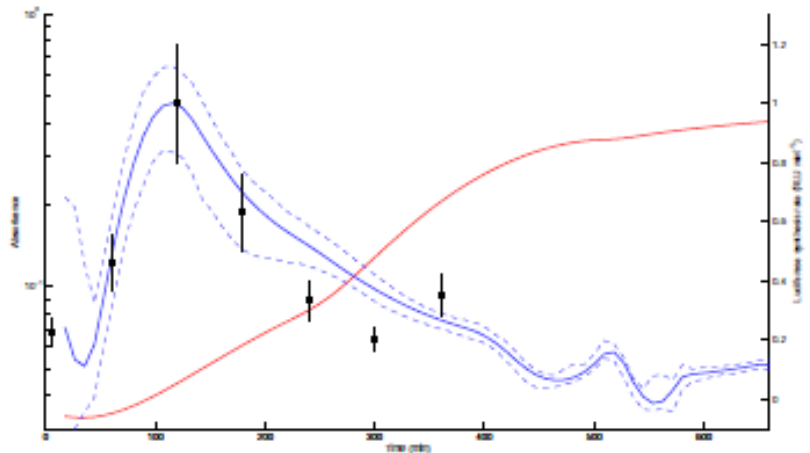
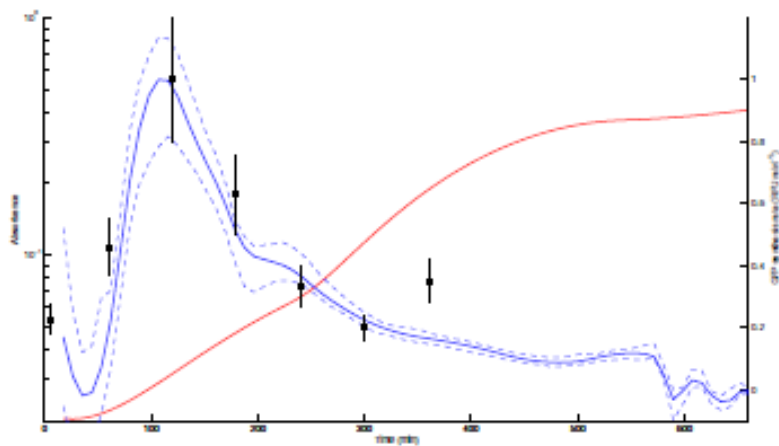
Predicted and measured concentration of the mRNA of the host protein and reporter genes



gfp



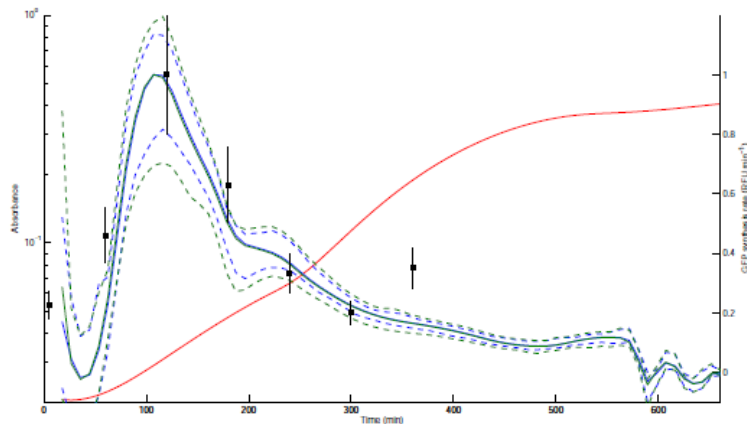
Luciferase



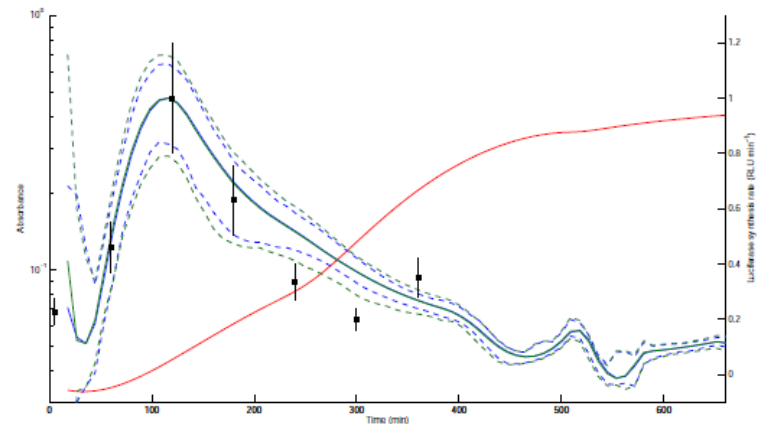
# Influence of the half-life of mRNA

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



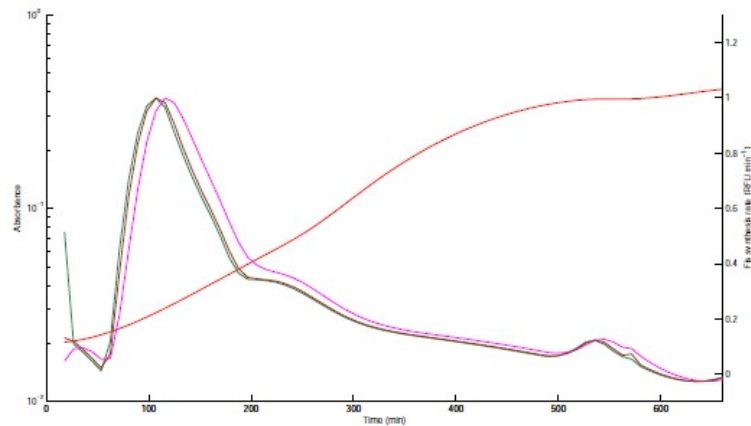
green: host protein

blue: gfp or luciferase

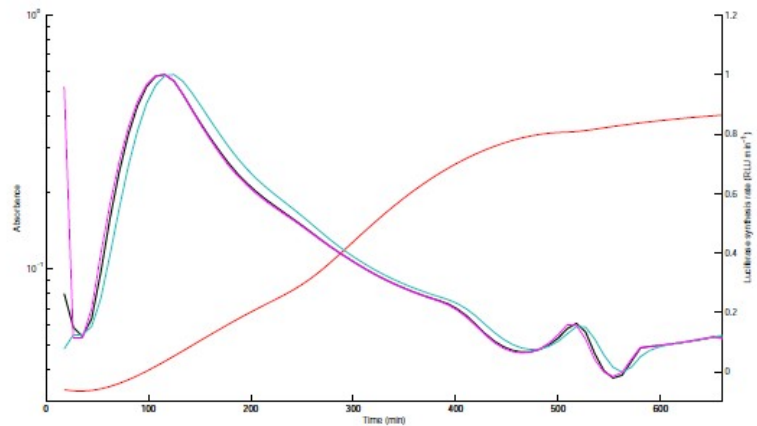
# Robustness of the synthesis rate

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

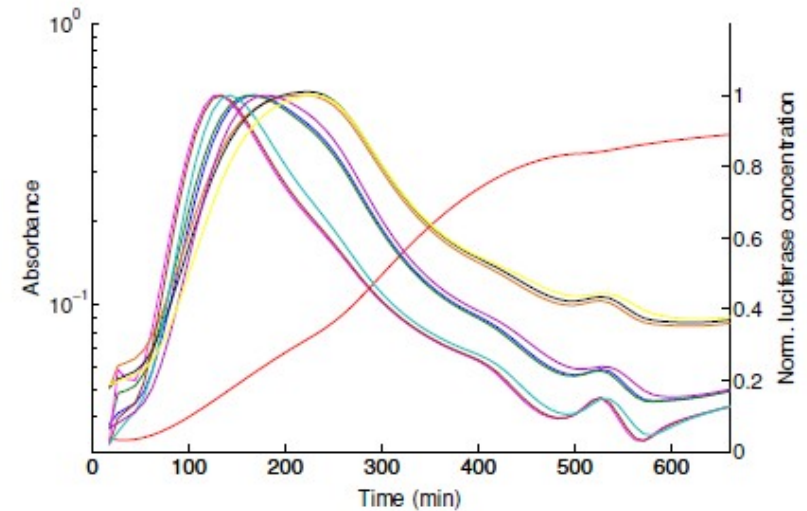
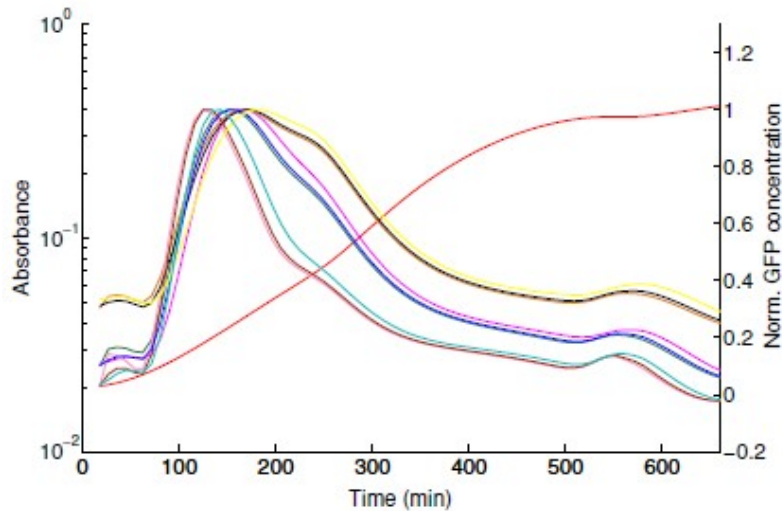


Luciferase



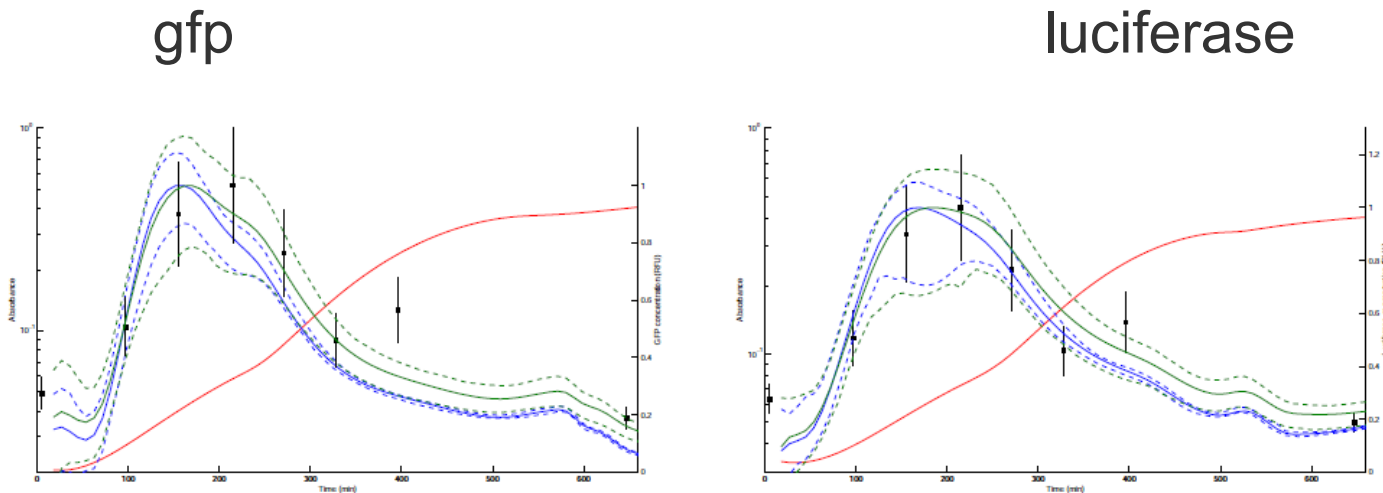
# Influence of the protein half-life

- More pronounced influence



# Correction for half-lives of reporter and host protein

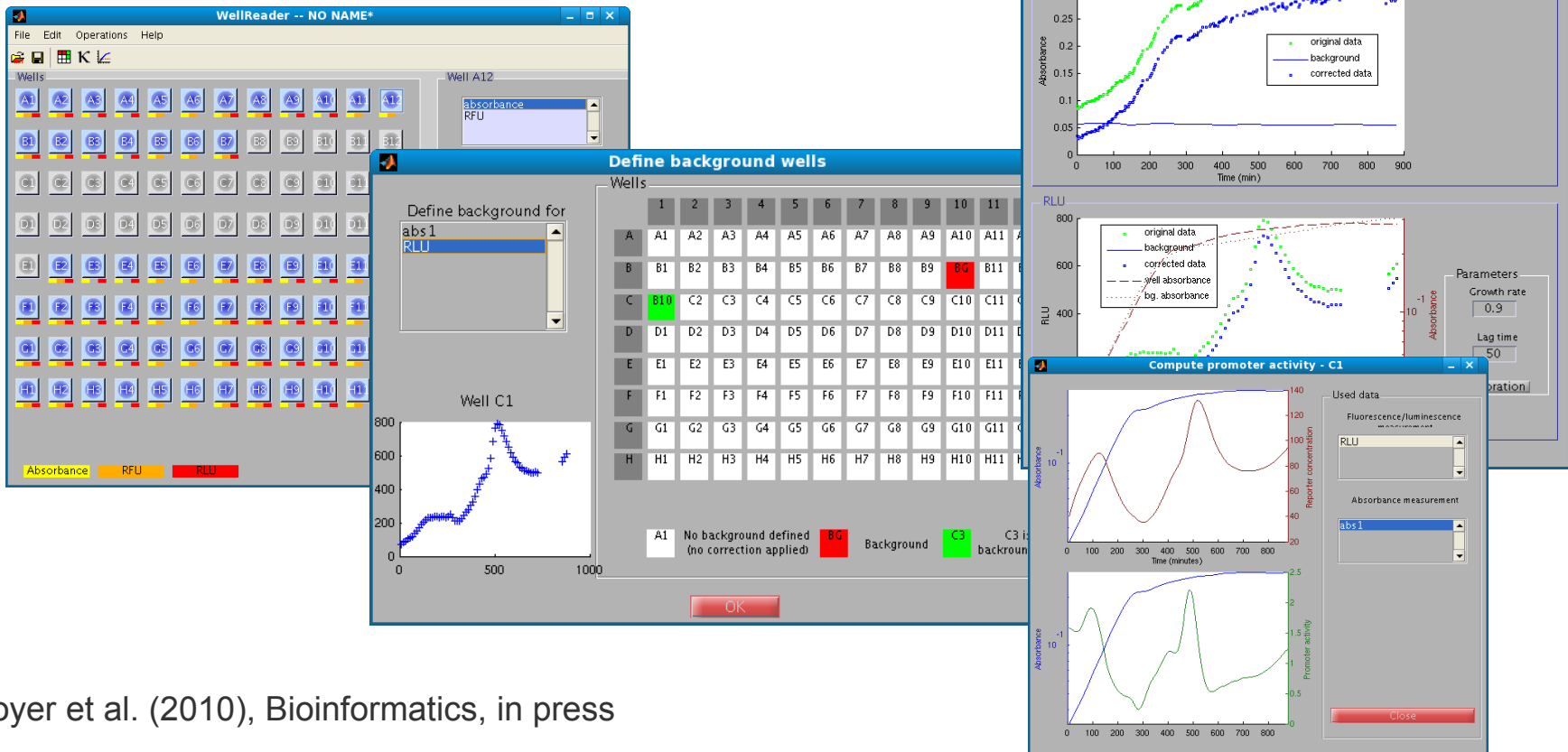
The calculated concentration of the host protein is corrected for the differences in half-lives of reporter gene and host protein



green: host protein      blue: gfp or luciferase

# Analysis of reporter gene expression data

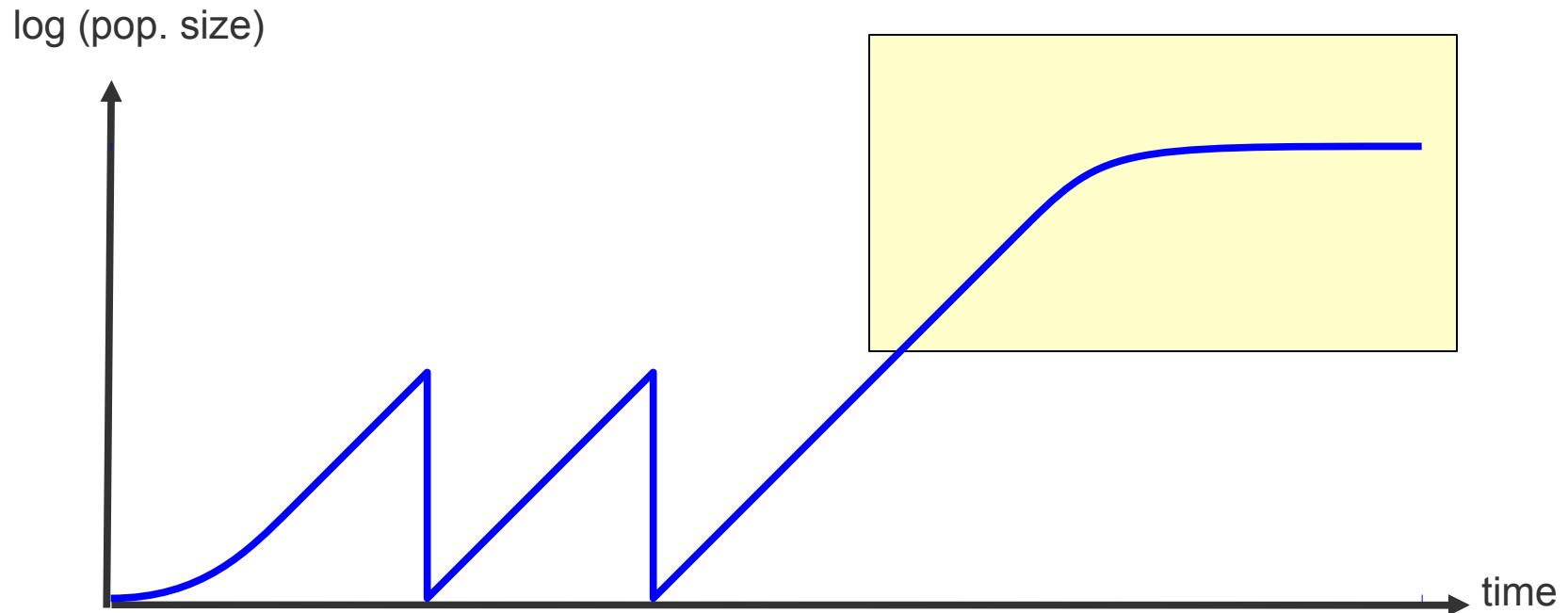
Wellreader: Matlab program for analysis of reporter gene expression data



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

# Transition from growth on glucose to acetate

- Serial dilution assures steady state at the beginning of the experiment

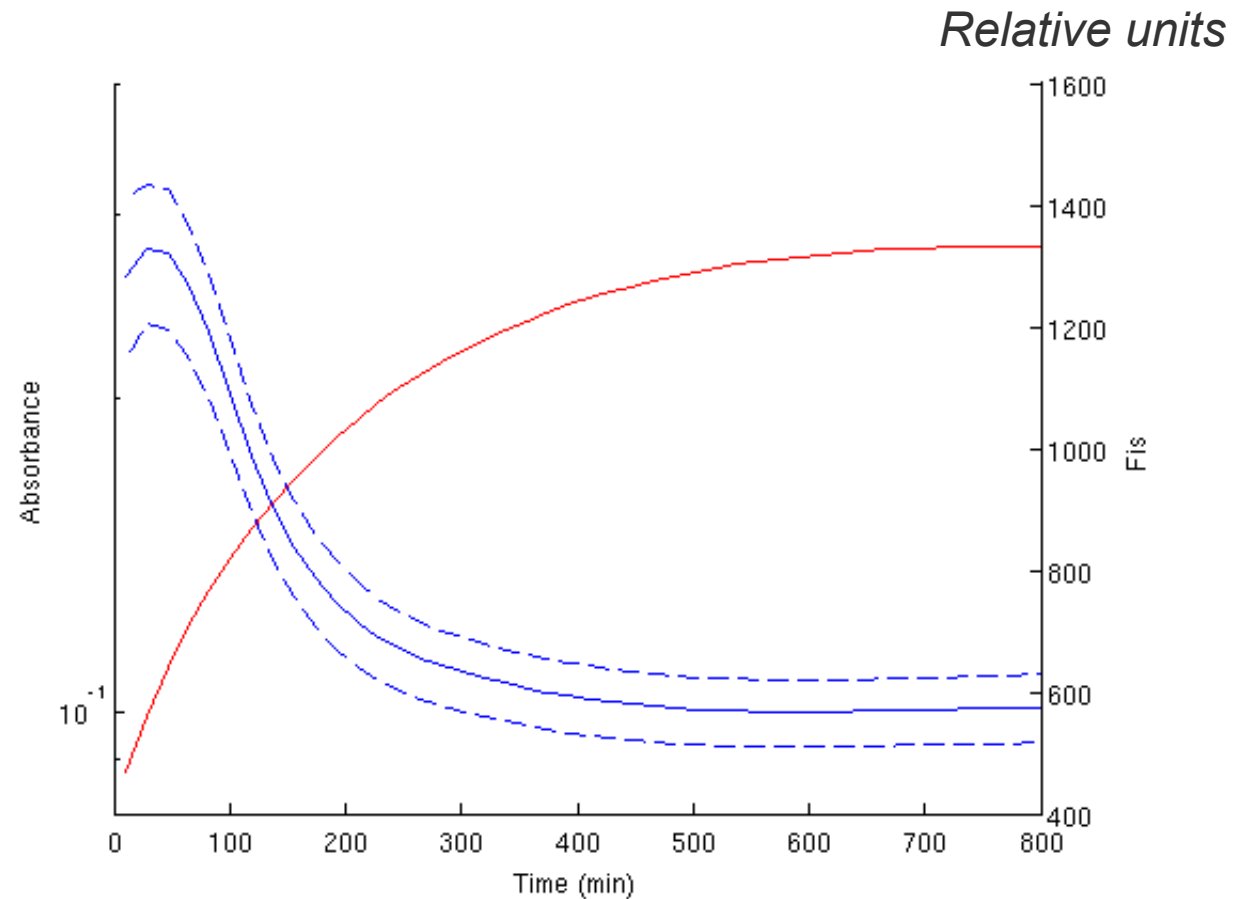




# Fis

## □ Protein concentration at growth transition

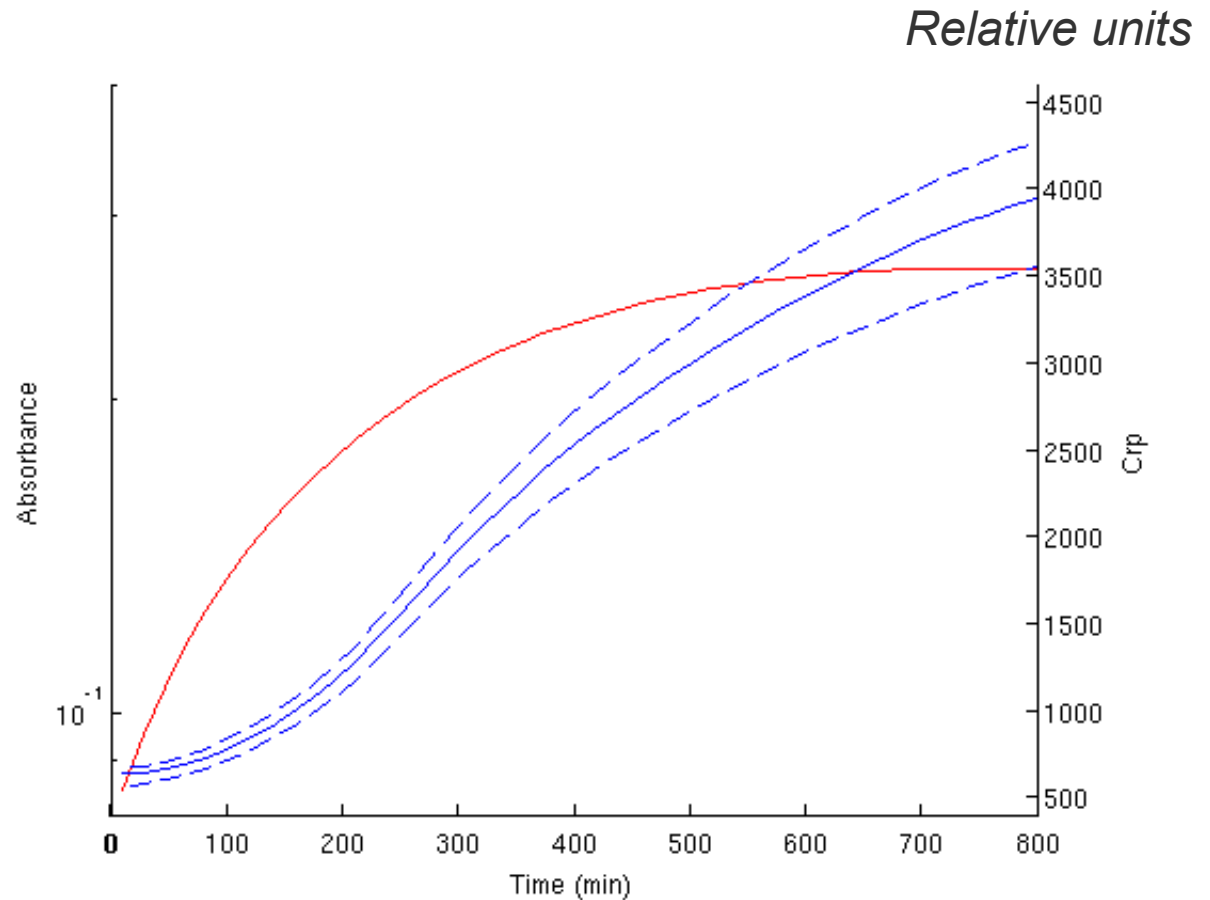
*Rapid decrease to new steady state*



# Crp

## □ Protein concentration at growth transition

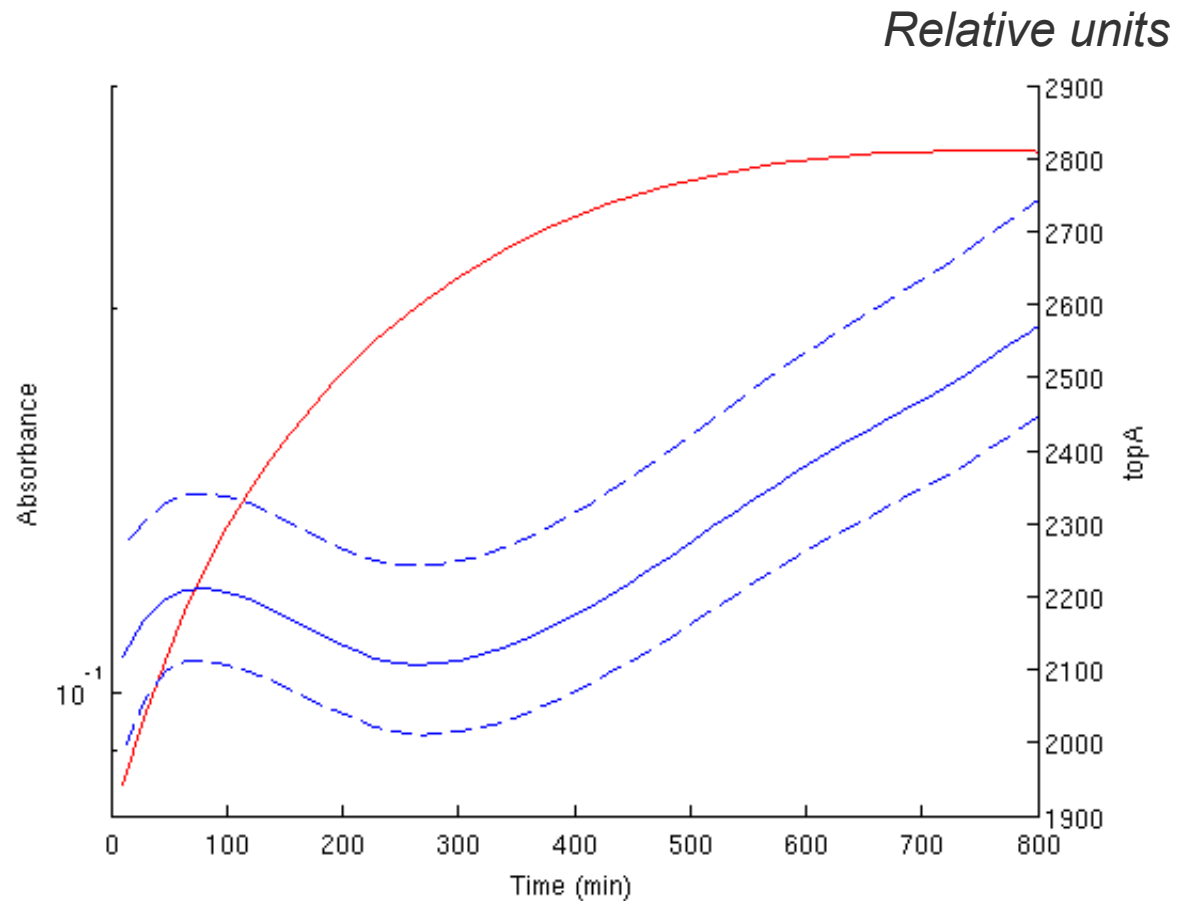
*Increase to new steady state*



# TopA

## □ Protein concentration at growth transition

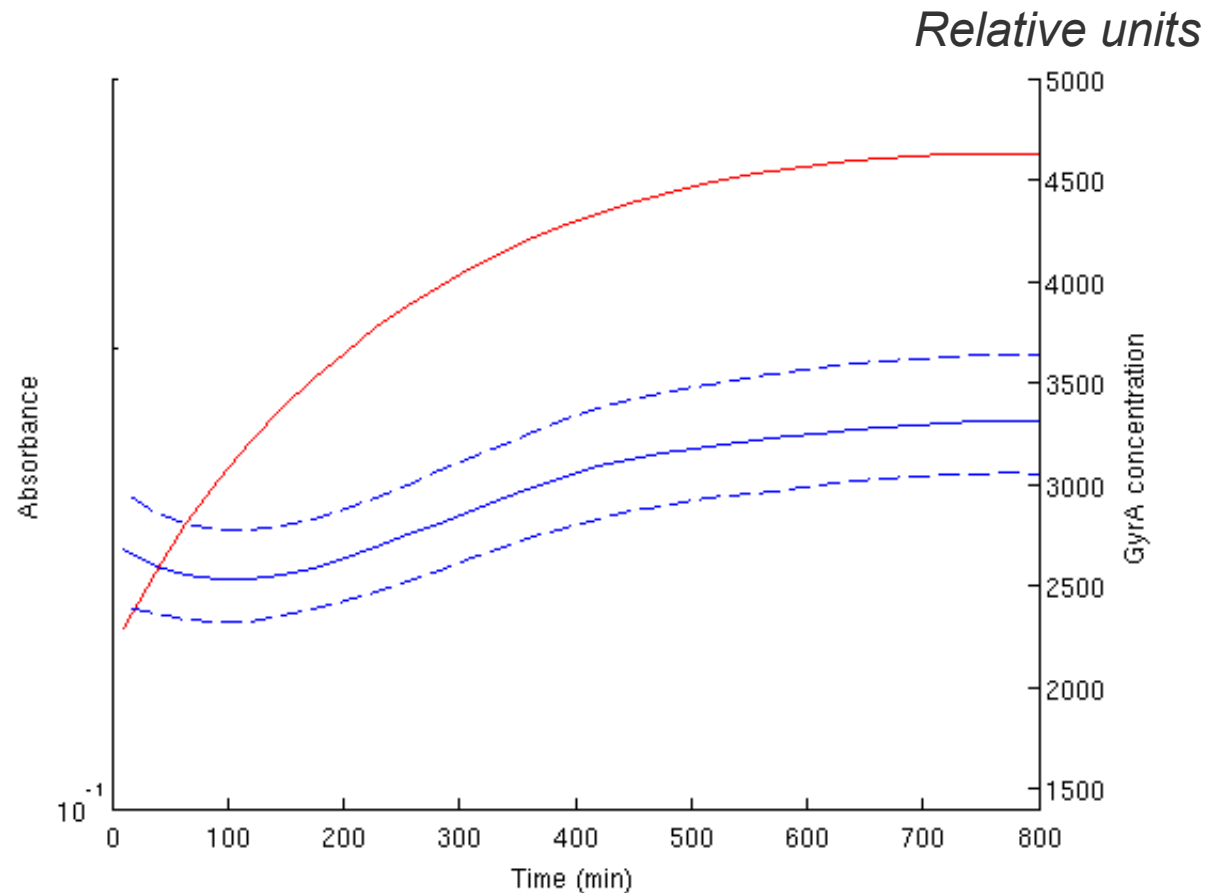
*Slow  
increase to  
new steady  
state*



# GyrA

## □ Protein concentration at growth transition

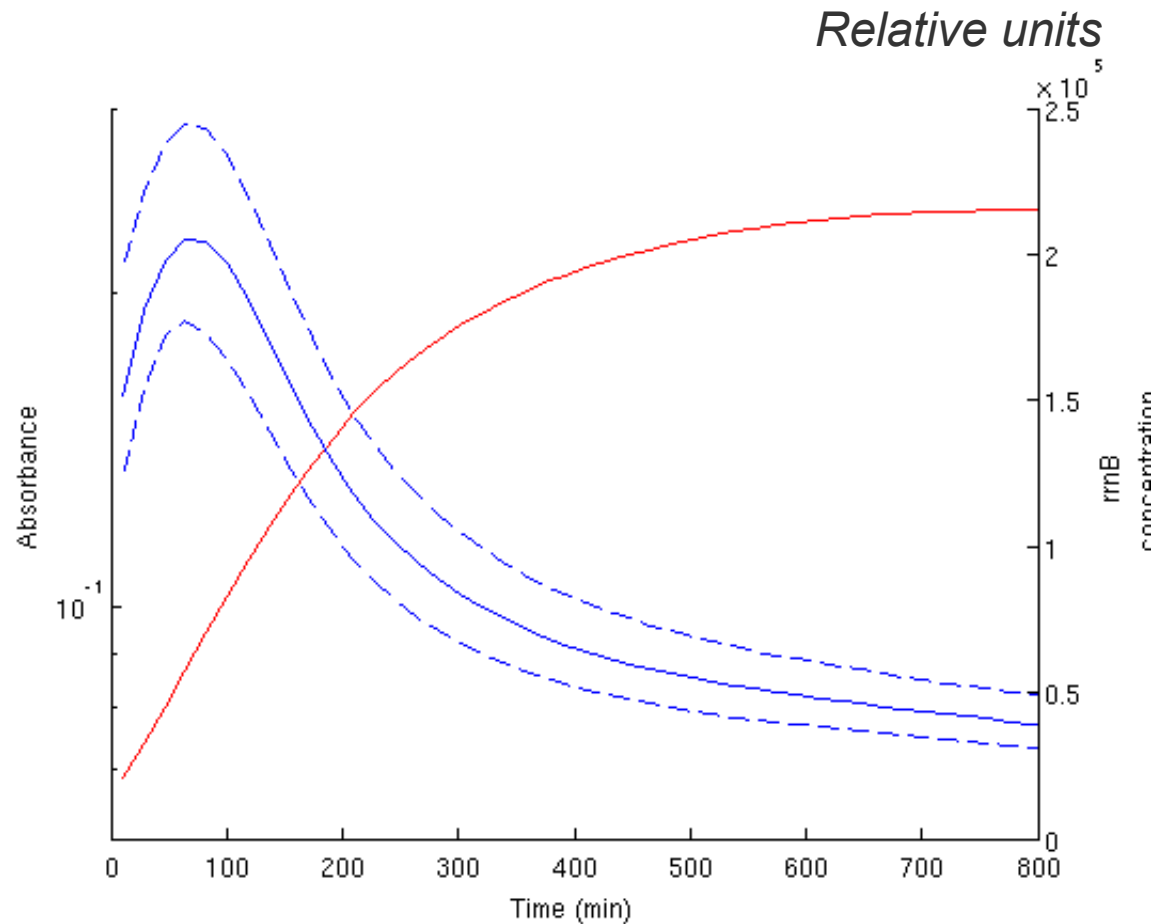
*Slow  
increase to  
new steady  
state*



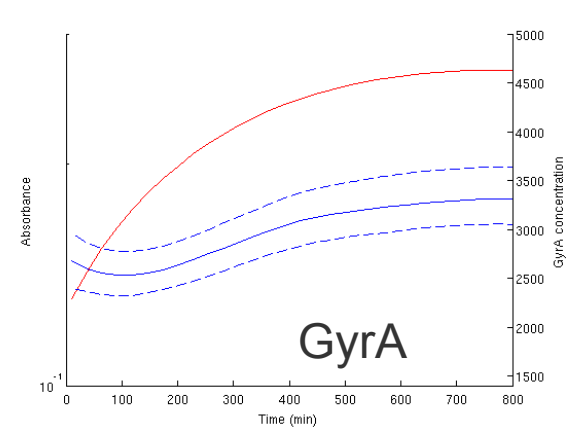
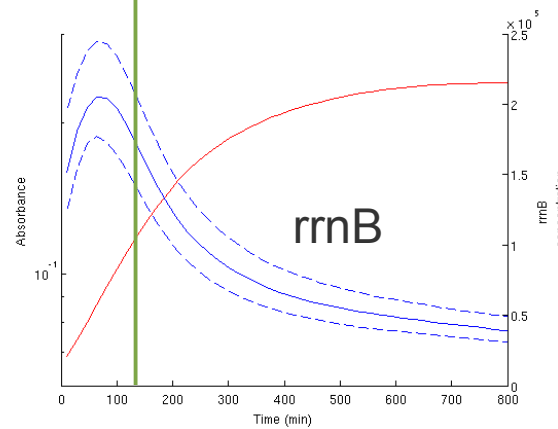
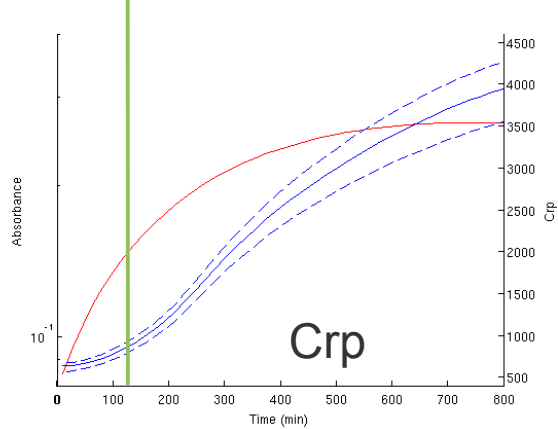
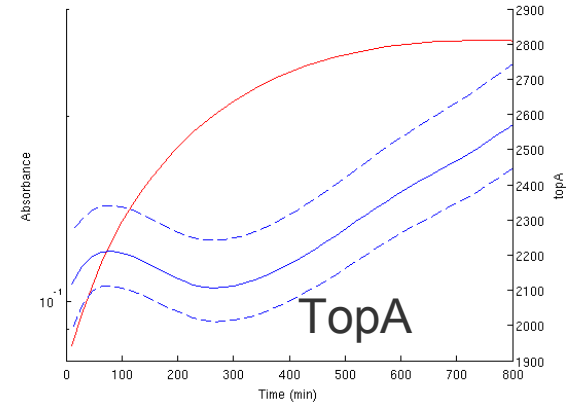
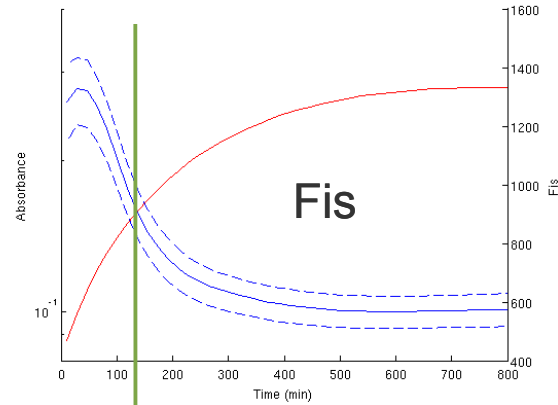
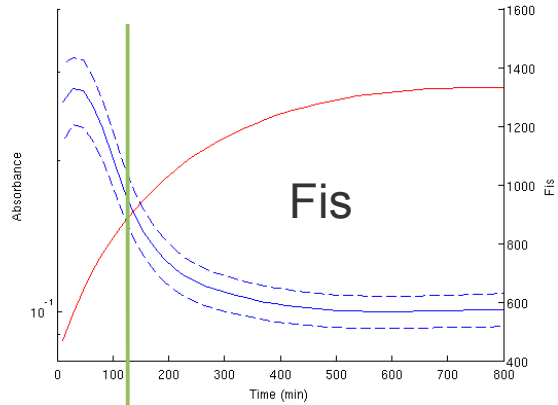
# rrnB

## □ Protein concentration at growth transition

*Rapid decrease to new steady state*



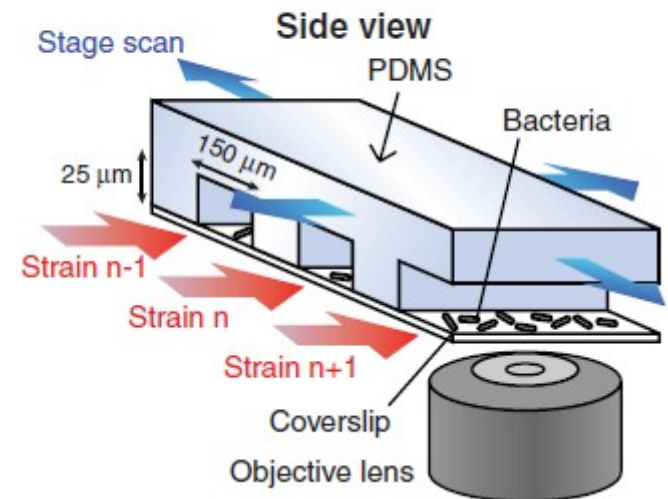
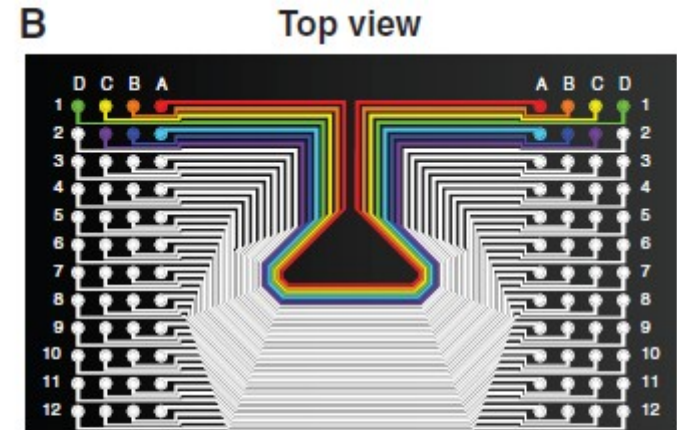
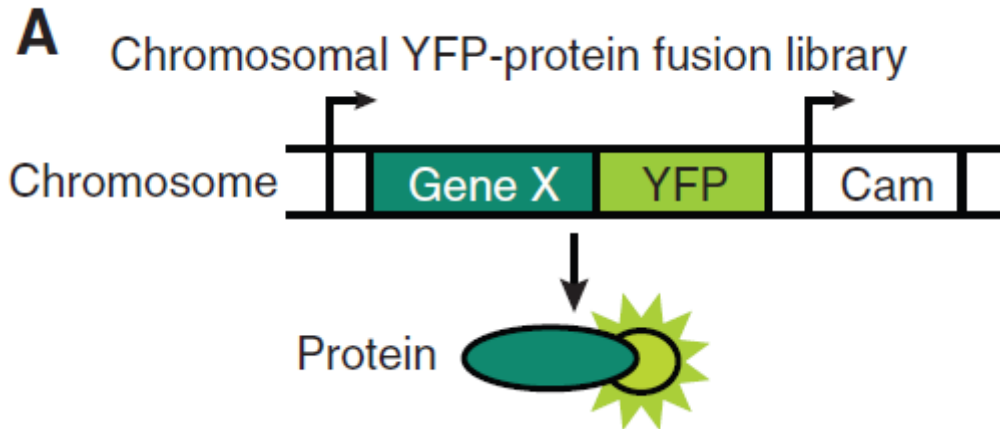
# Expression timing





# Single cell imaging of fusion proteins

- Measure fluorescence of fusion proteins

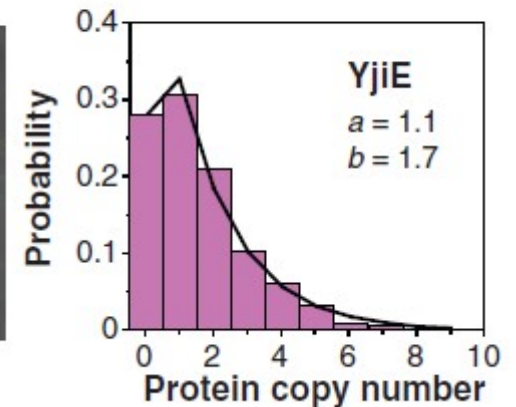
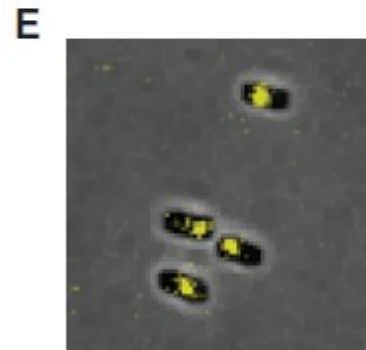
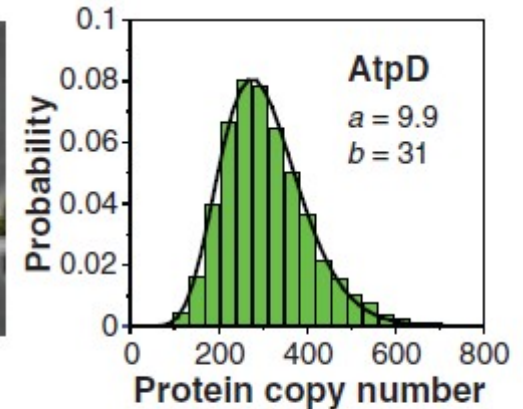
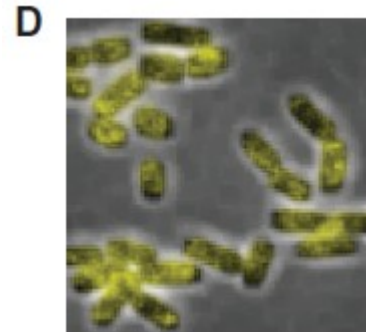
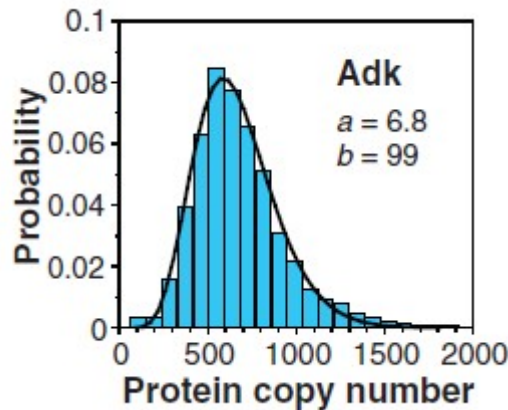
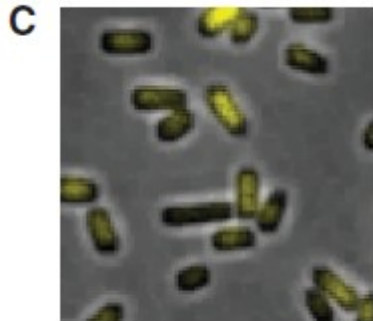




# The number of proteins is gamma distributed

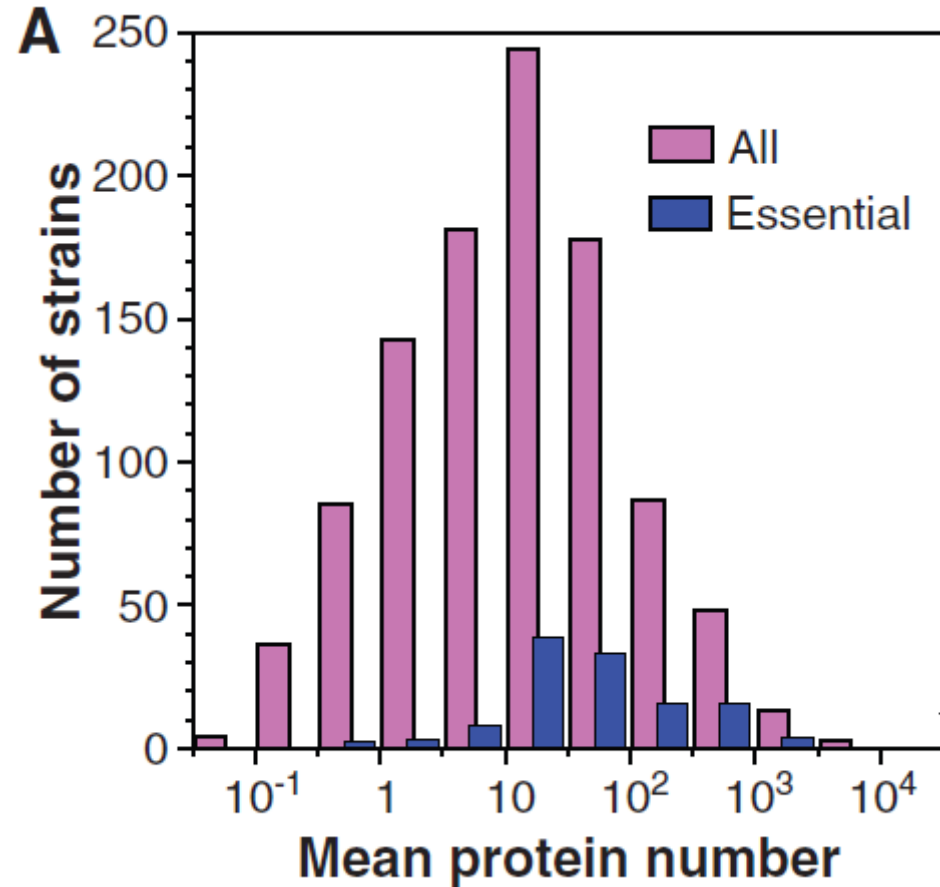
- Examples of protein distributions

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



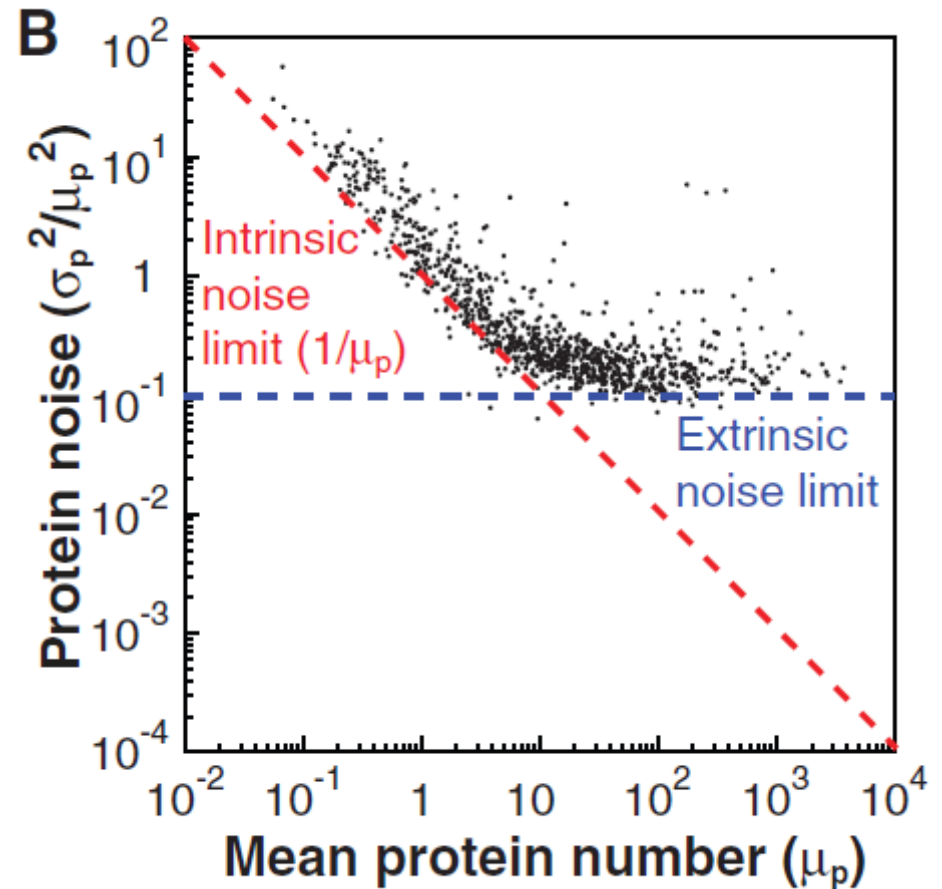
# Essential proteins are more abundant

- Many proteins are present in low abundance



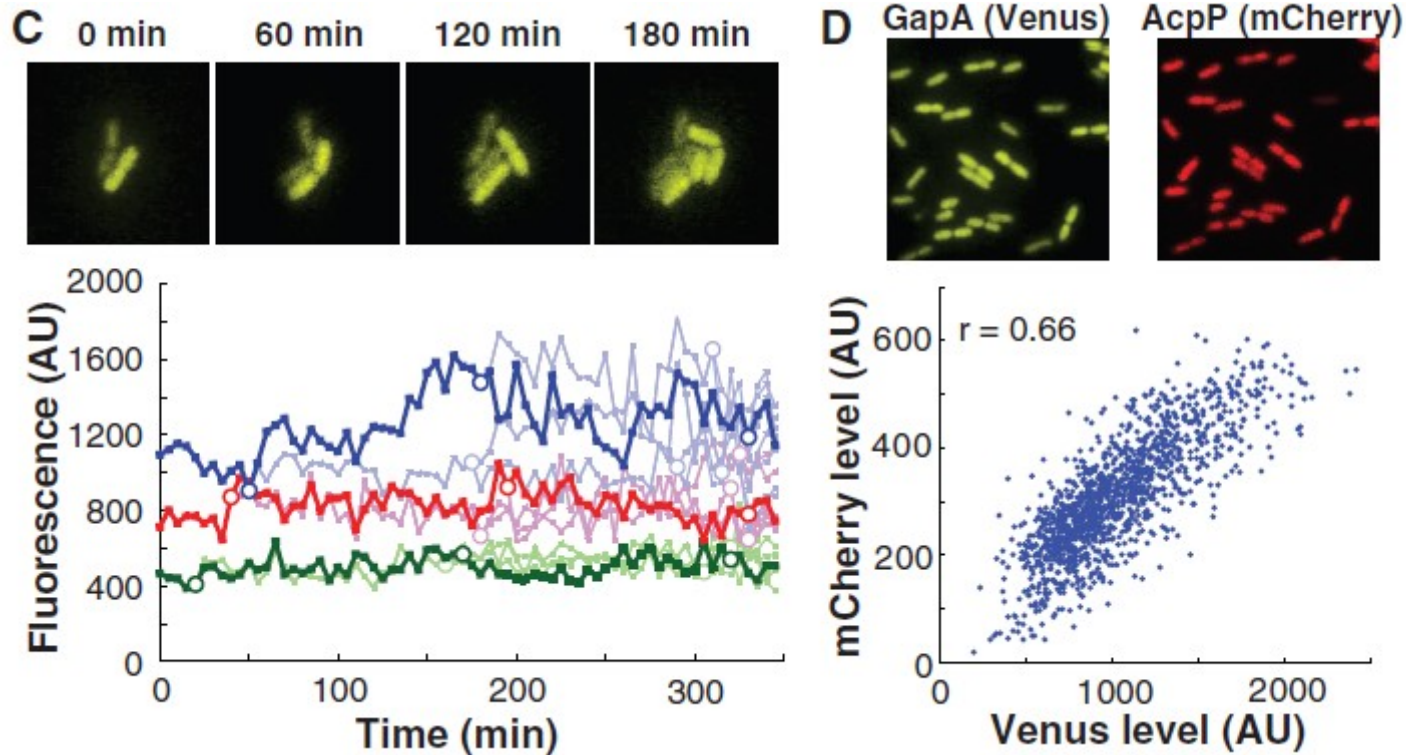
# Noise of protein expression

- 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



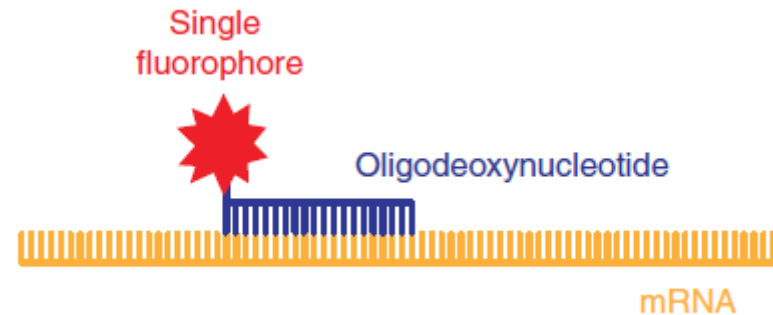
# Fluctuations in protein levels are slow

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



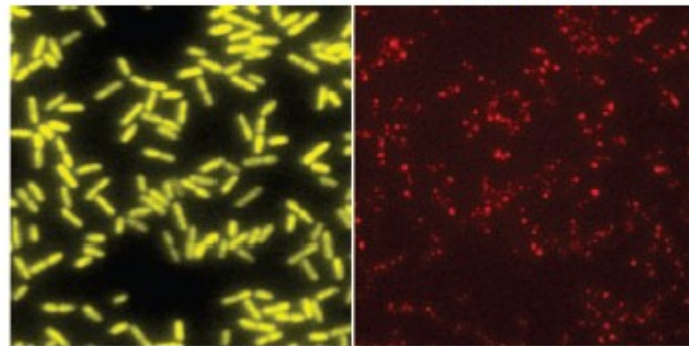
# Measuring RNA molecules by FISH

- target *yfp* RNA

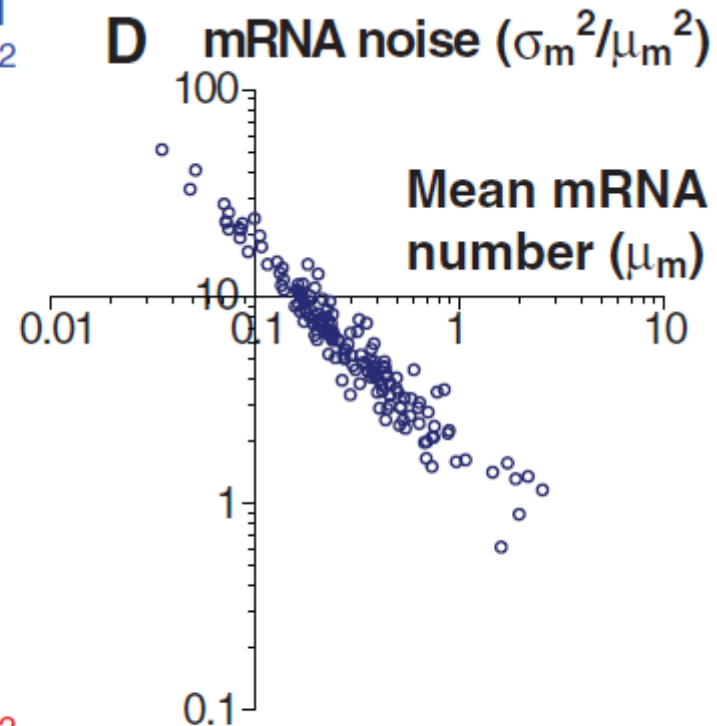
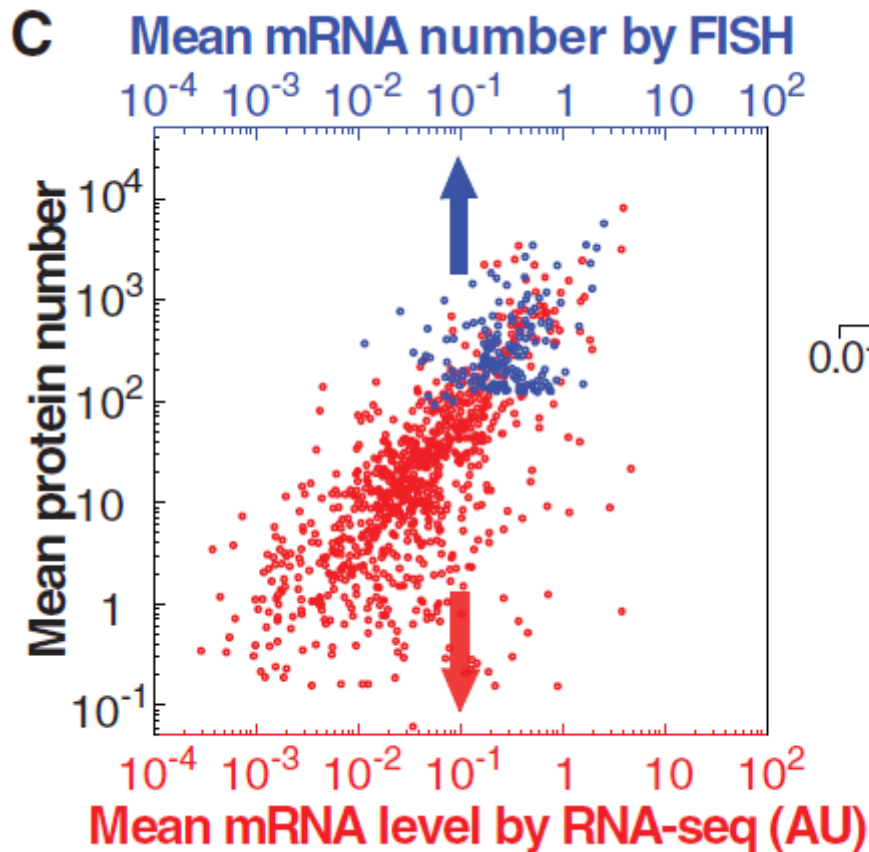


YFP fluorescence  
514 nm ex.

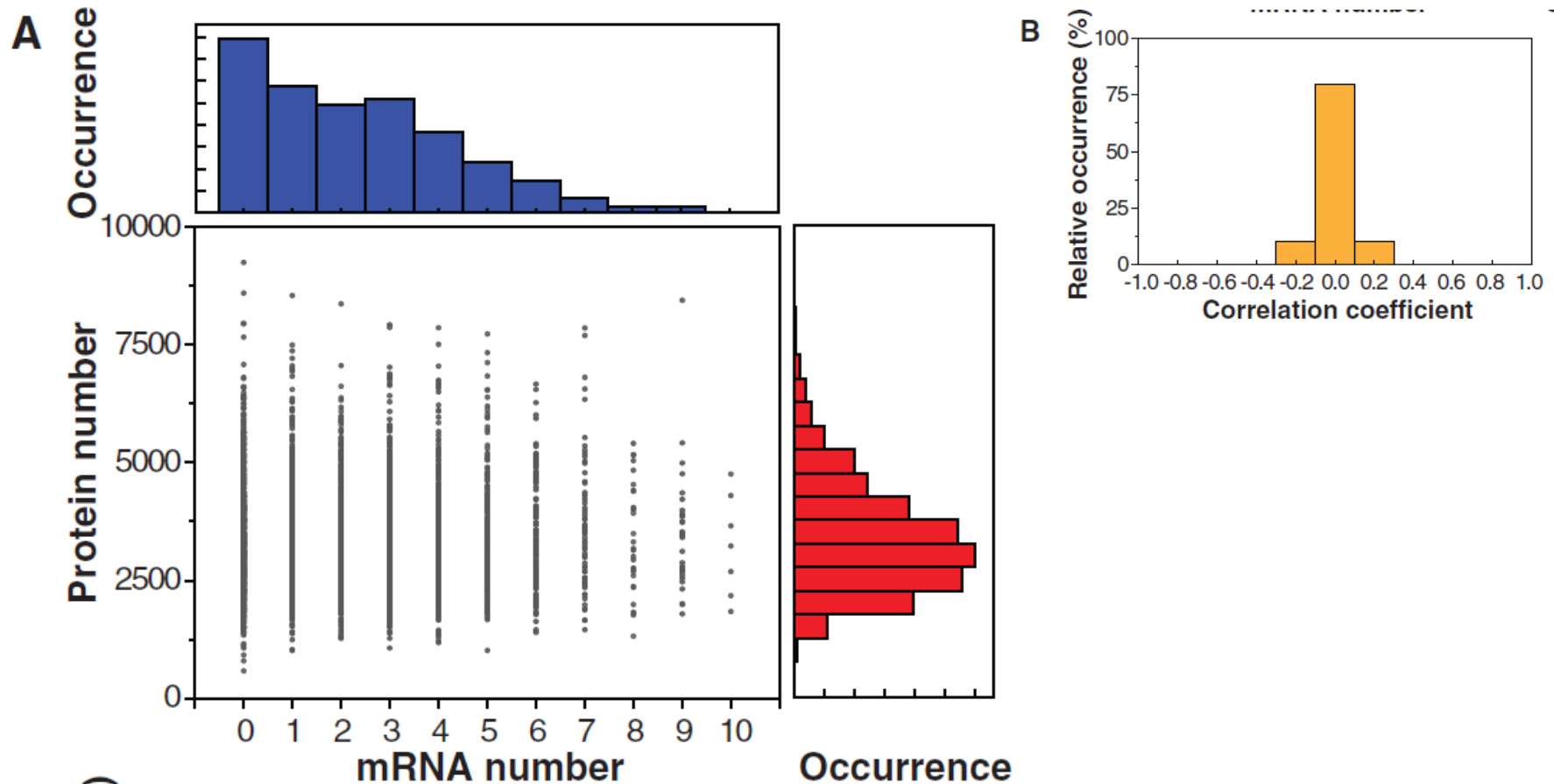
*yfp* mRNA  
with Atto 594 probe  
580 nm ex.



# General correlation between mRNA and protein concentrations and random noise

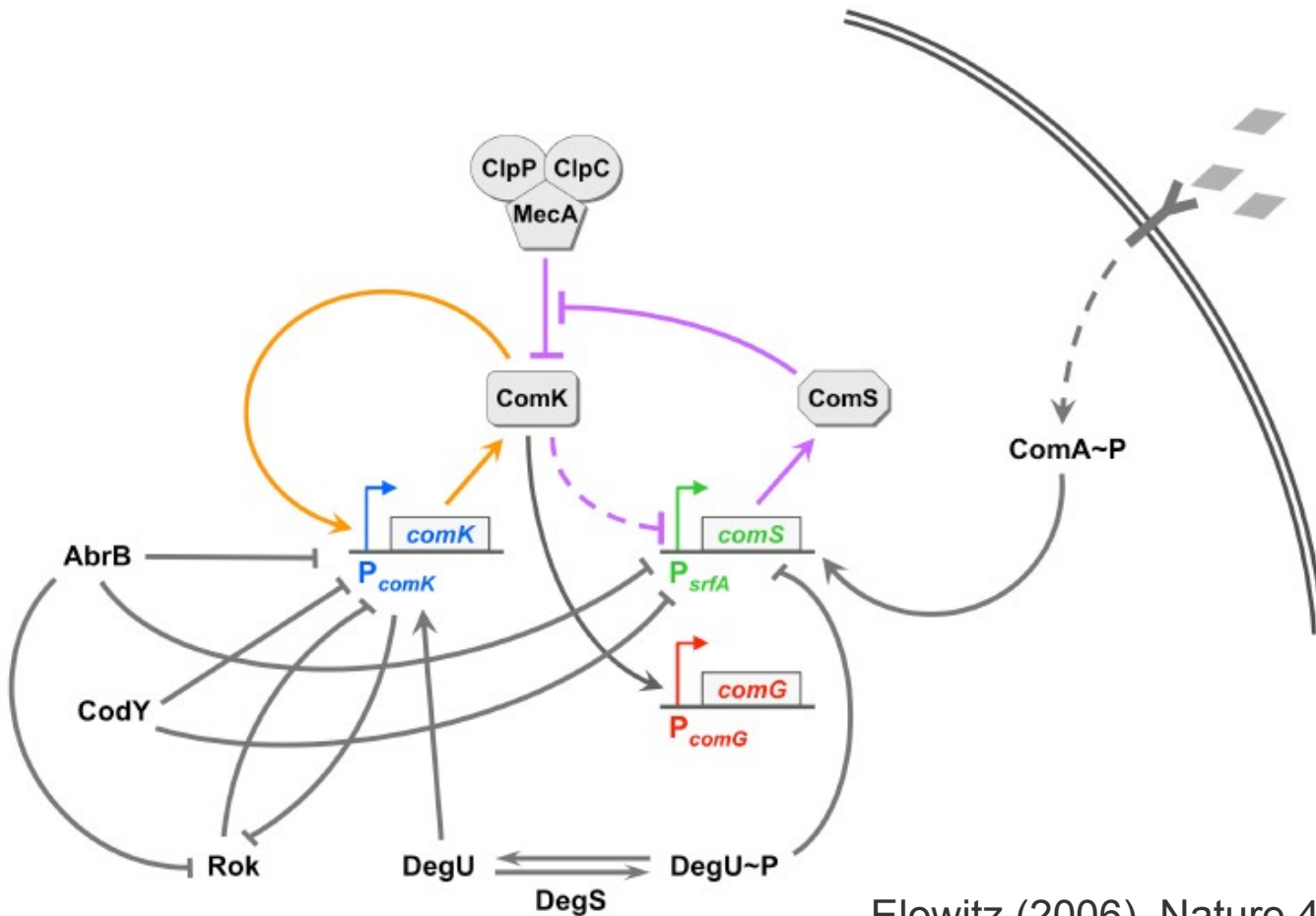


# No correlation between mRNA and protein level in a single cell





# Regulation of competence

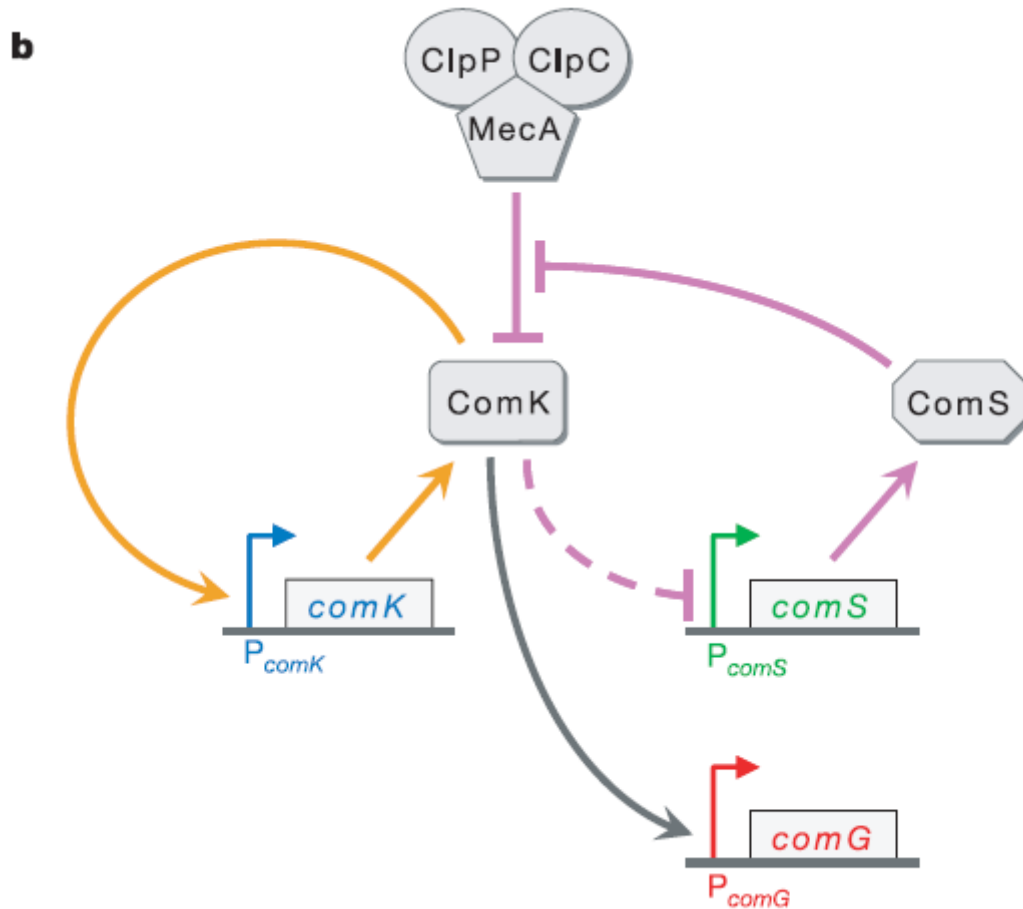


Elowitz (2006), Nature 440, 545-550



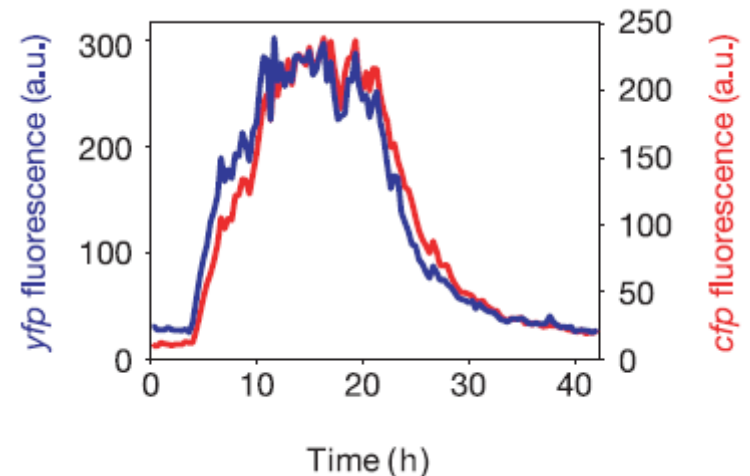
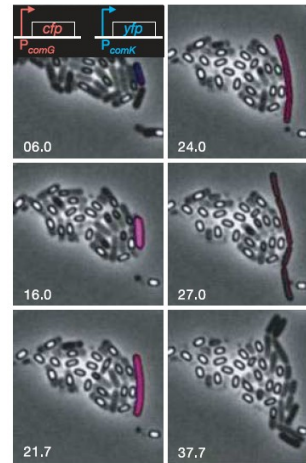
# Simplified MeKS model

- Focus on the essential system components



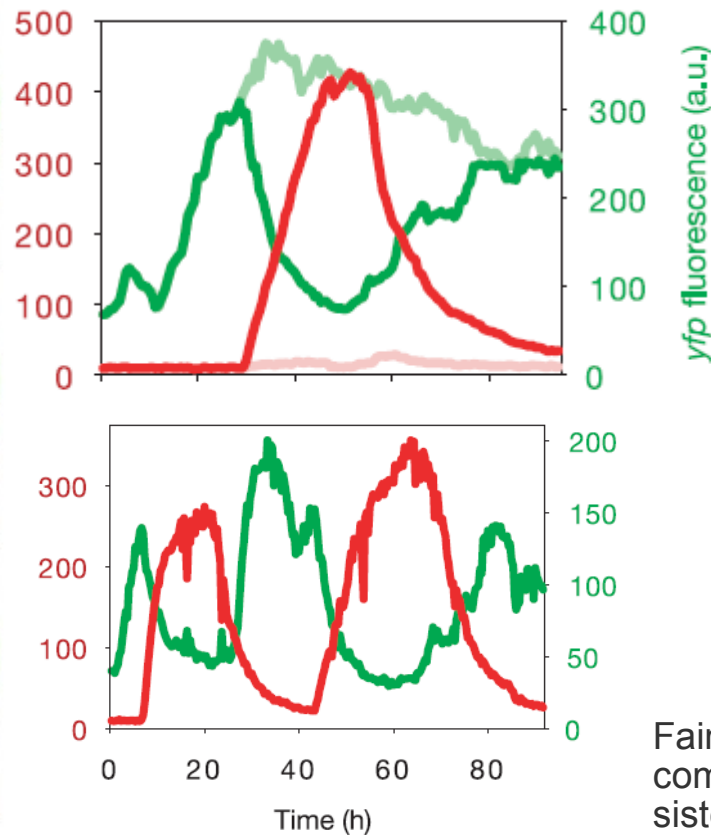
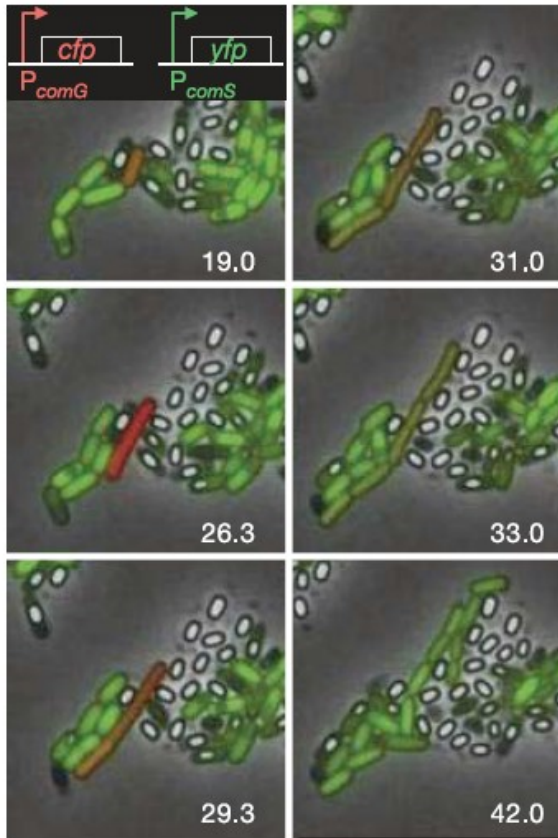
# $P_{comG}$ and $P_{comK}$ activities are highly correlated

- *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

---

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

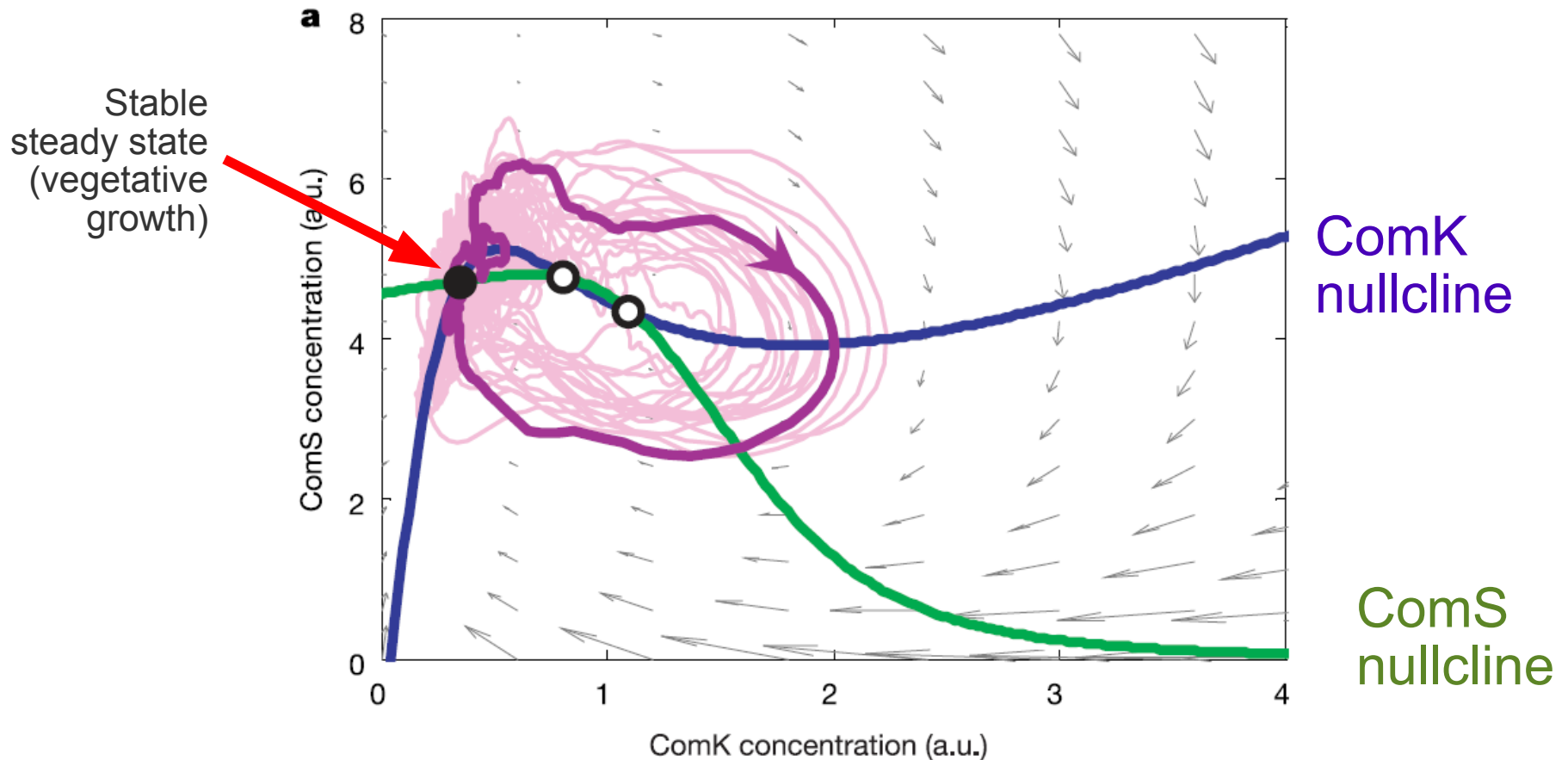
Auto-activation                      Competitive inhibition

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

Inhibition    noise

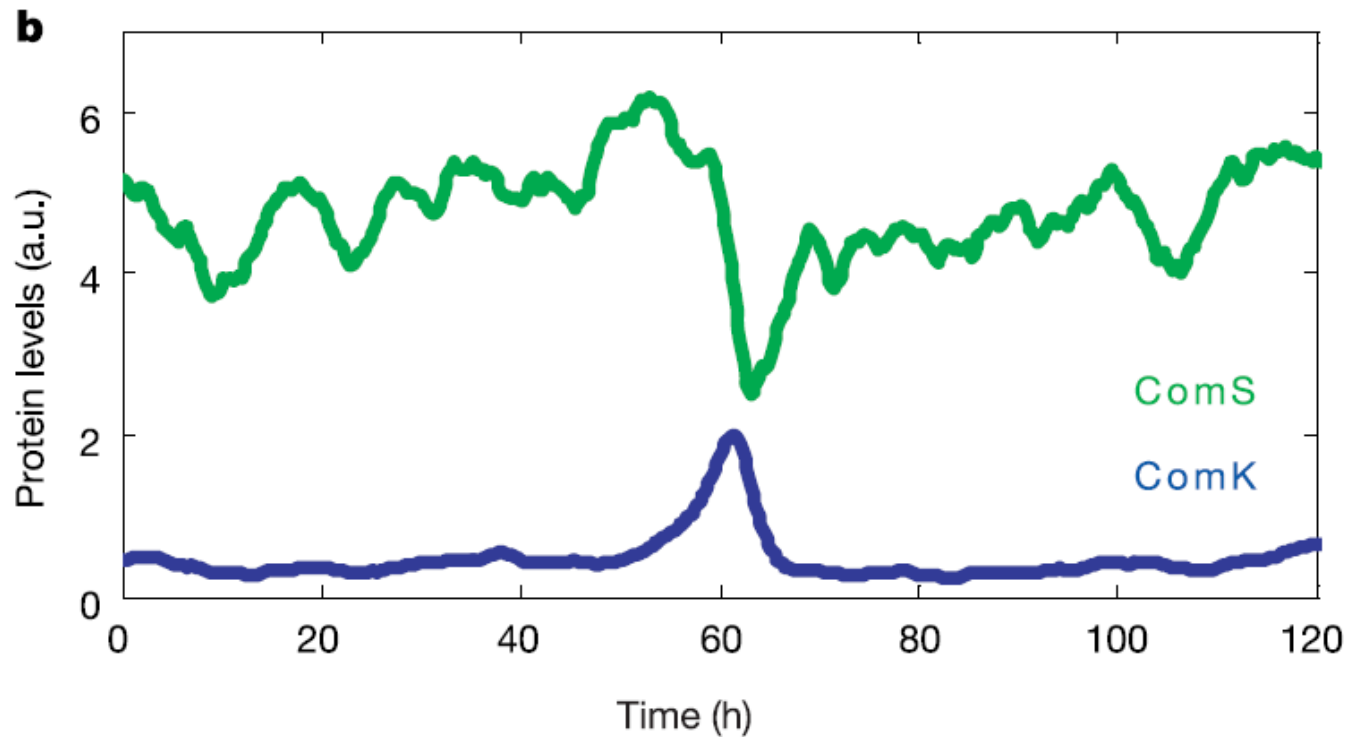
# Analysis of the dynamical system

- Phase plane diagram for ComK and ComS



# Simulation of ComK and ComS trajectories

- Excursion triggered by noise
- Anticorrelation between ComK and ComS



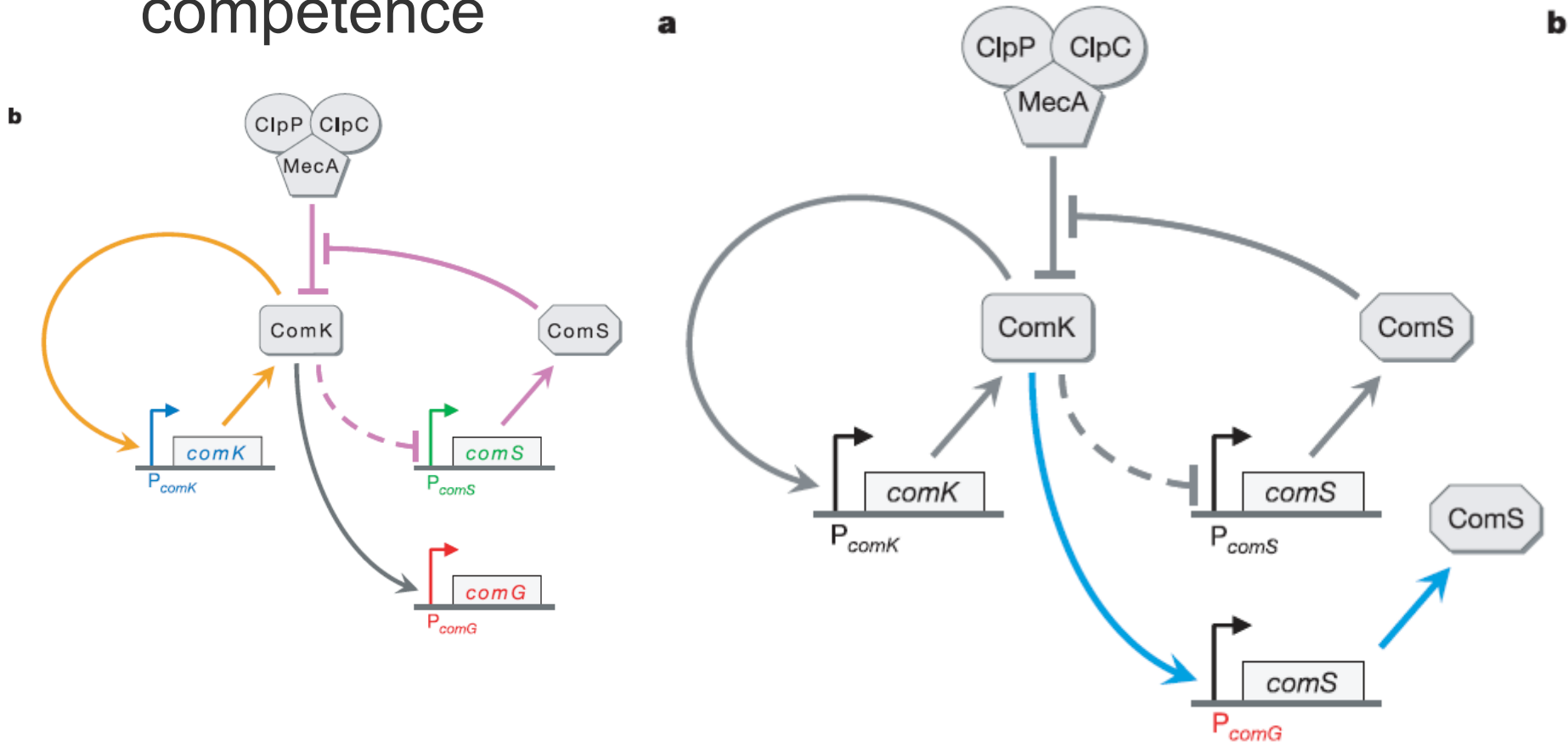
# Dynamics of system functioning

---

- 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

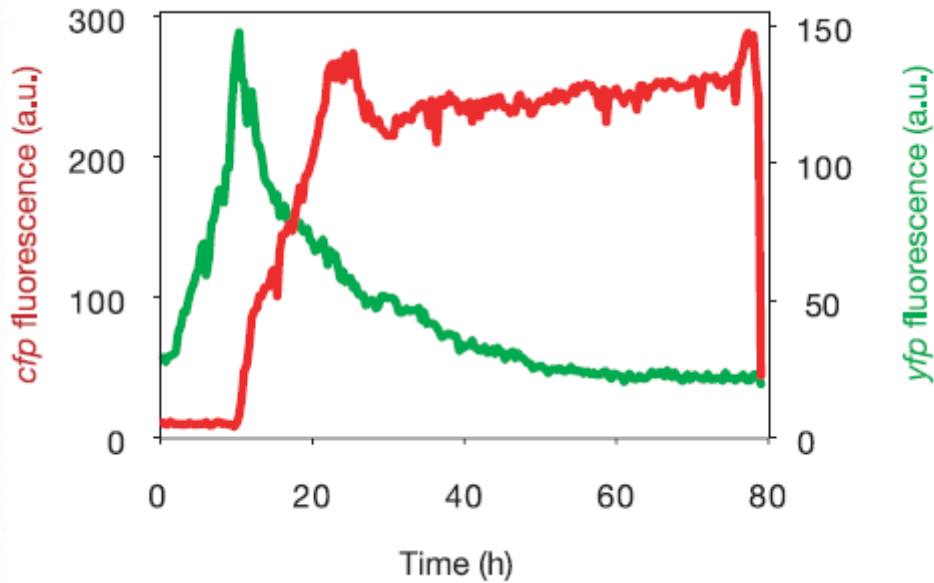
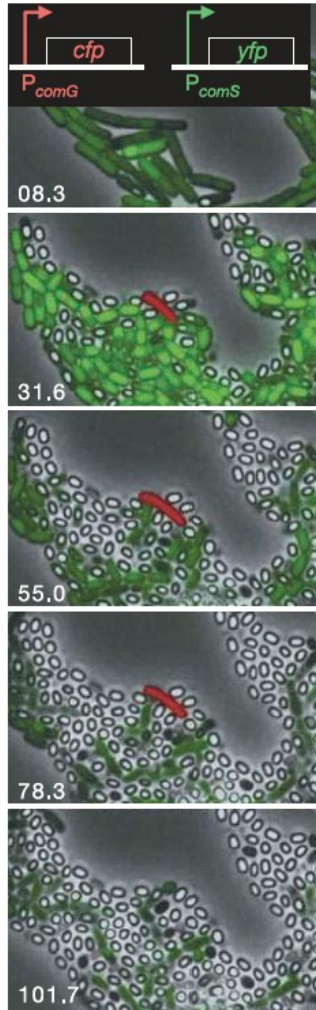
# Locking the system into competence mode

- Modify a connection to keep ComS during competence





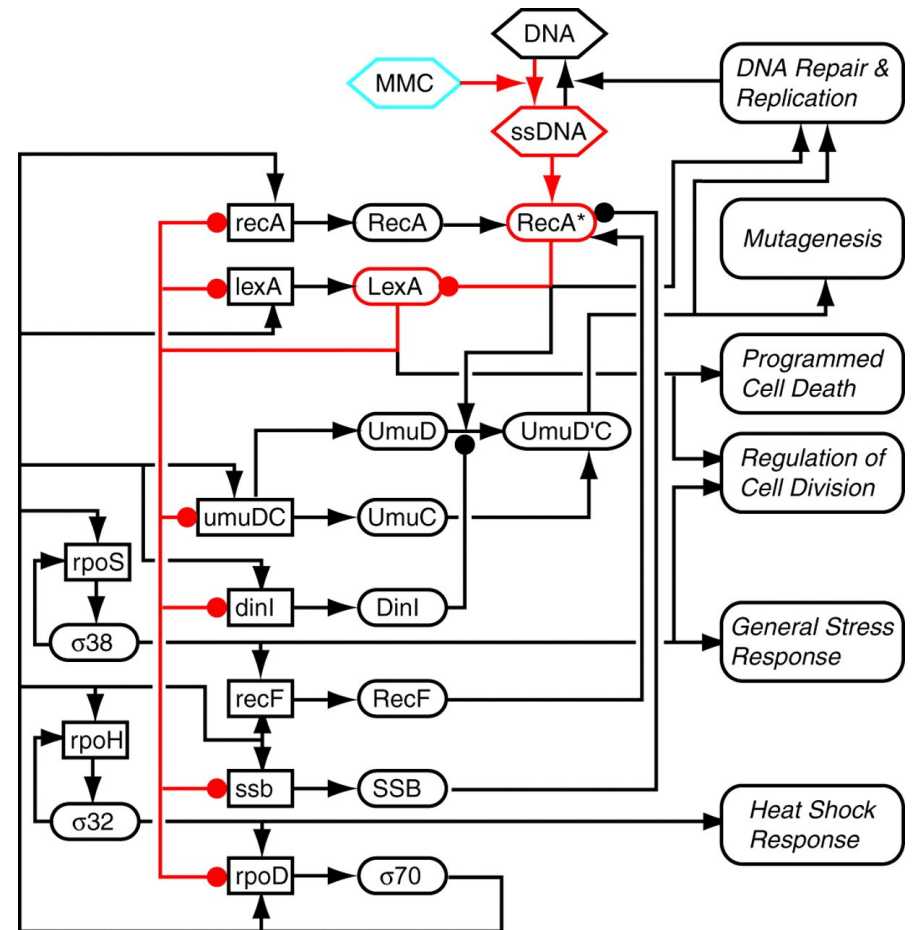
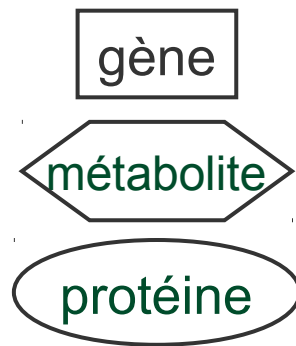
# The modified system can not escape from competence



# 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

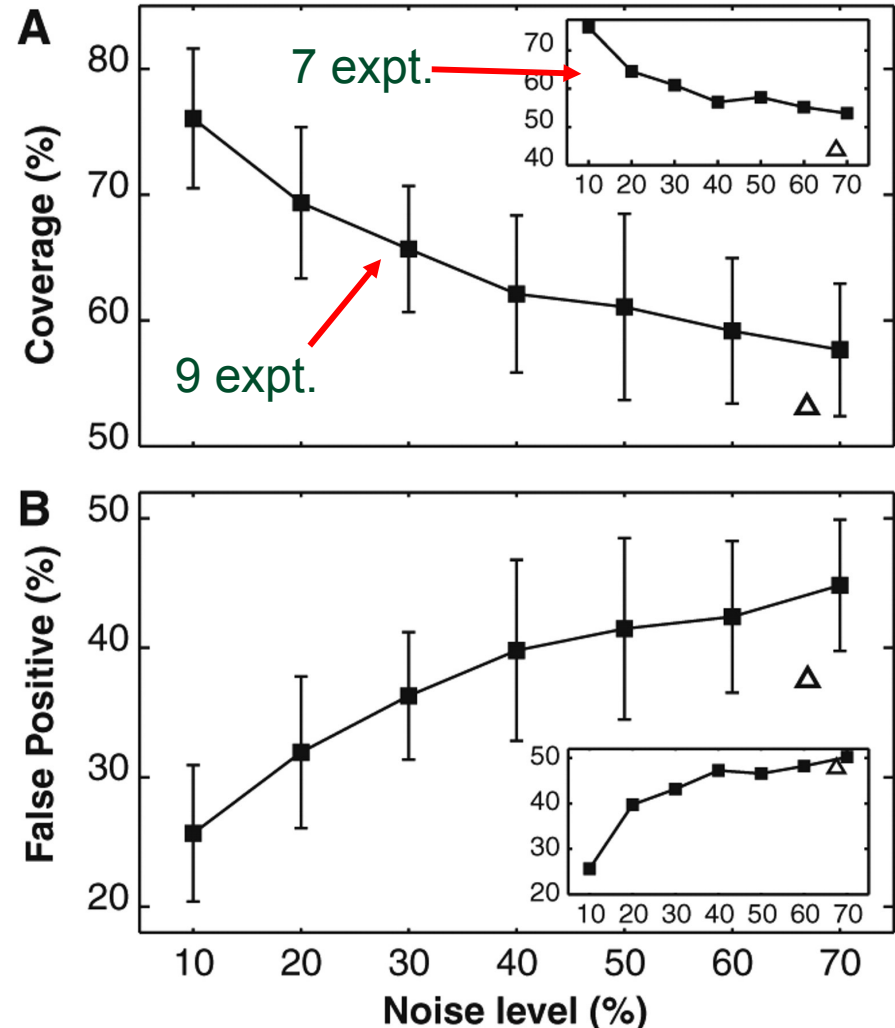
---

- 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

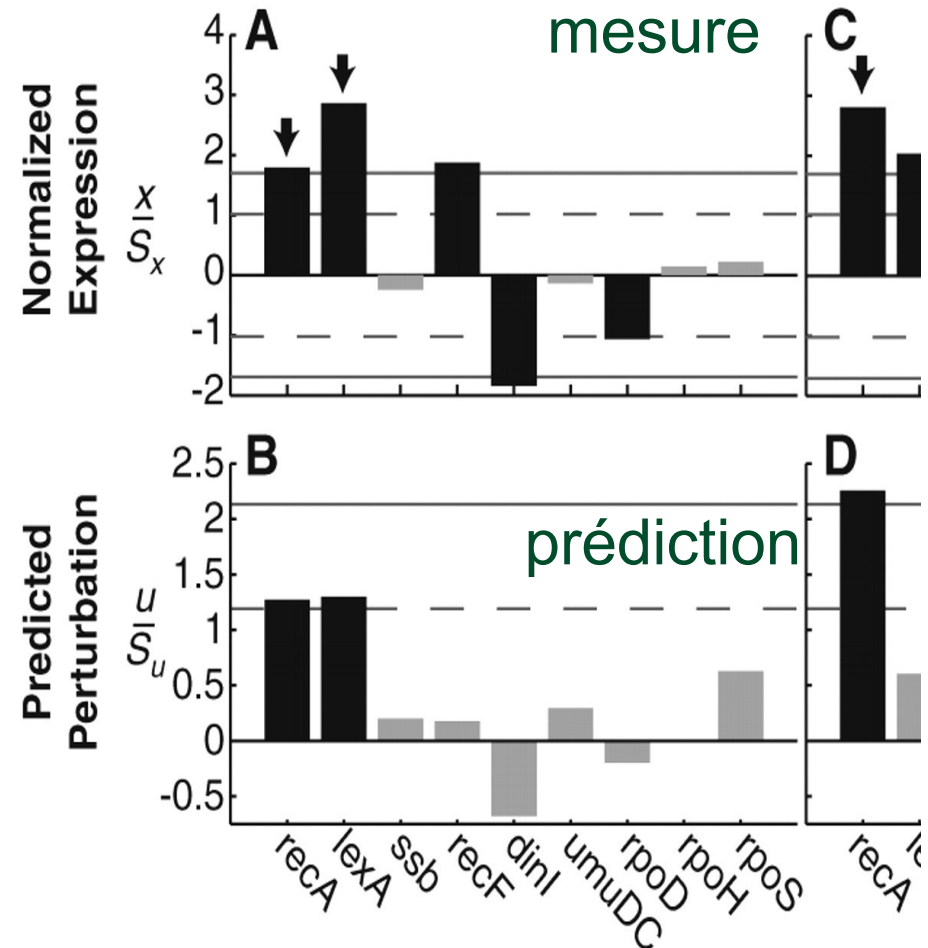
- Simulation d'un réseau de neuf gènes
- Rajouter un bruit expérimental
- Reconstruire le réseau
- Evaluer le nombre de bonnes et mauvaises connexions trouvées

**Bruit expérimental : 68%**



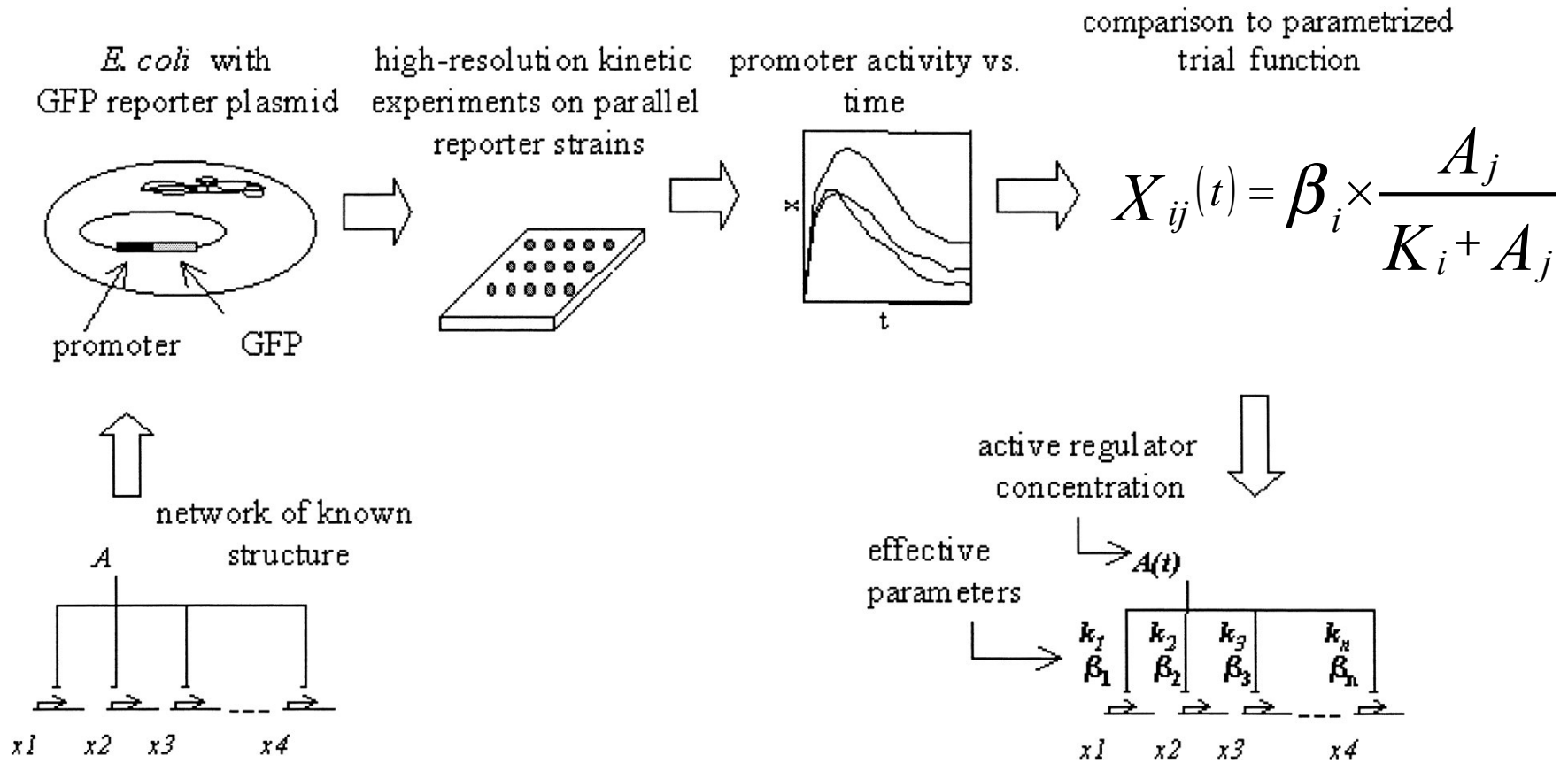
# Prédiction de l'effet d'une double perturbation

- 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

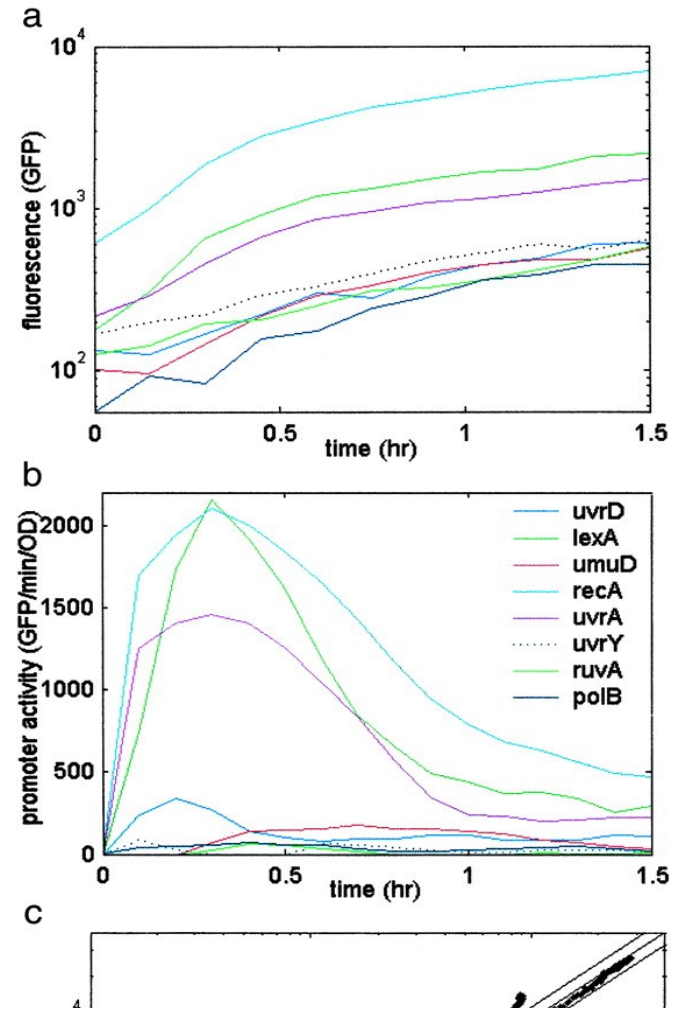
- Approche expérimentale



# 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

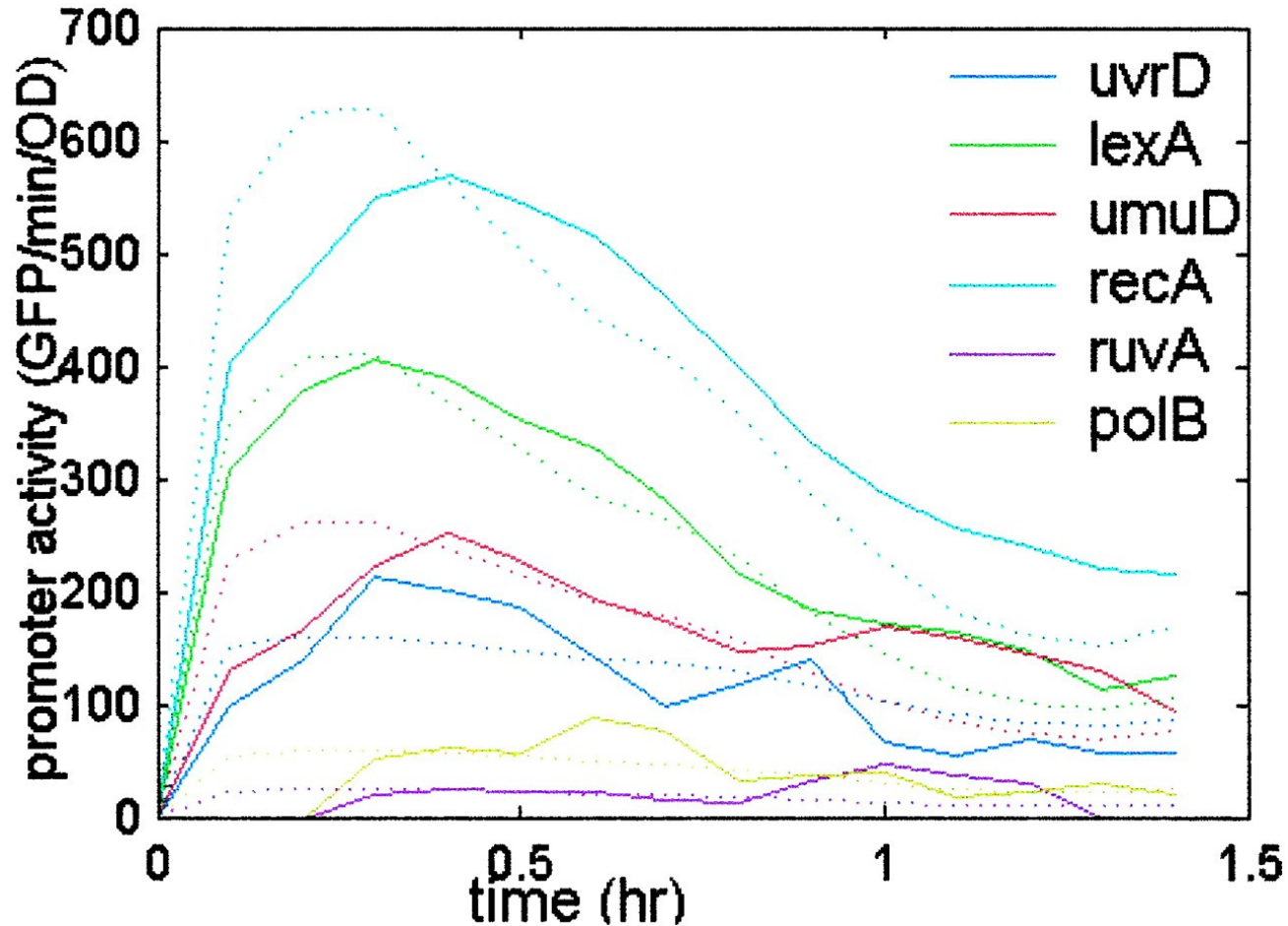
Table 1. The effective kinetic parameters for the SOS system ( $\pm$ SD)

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

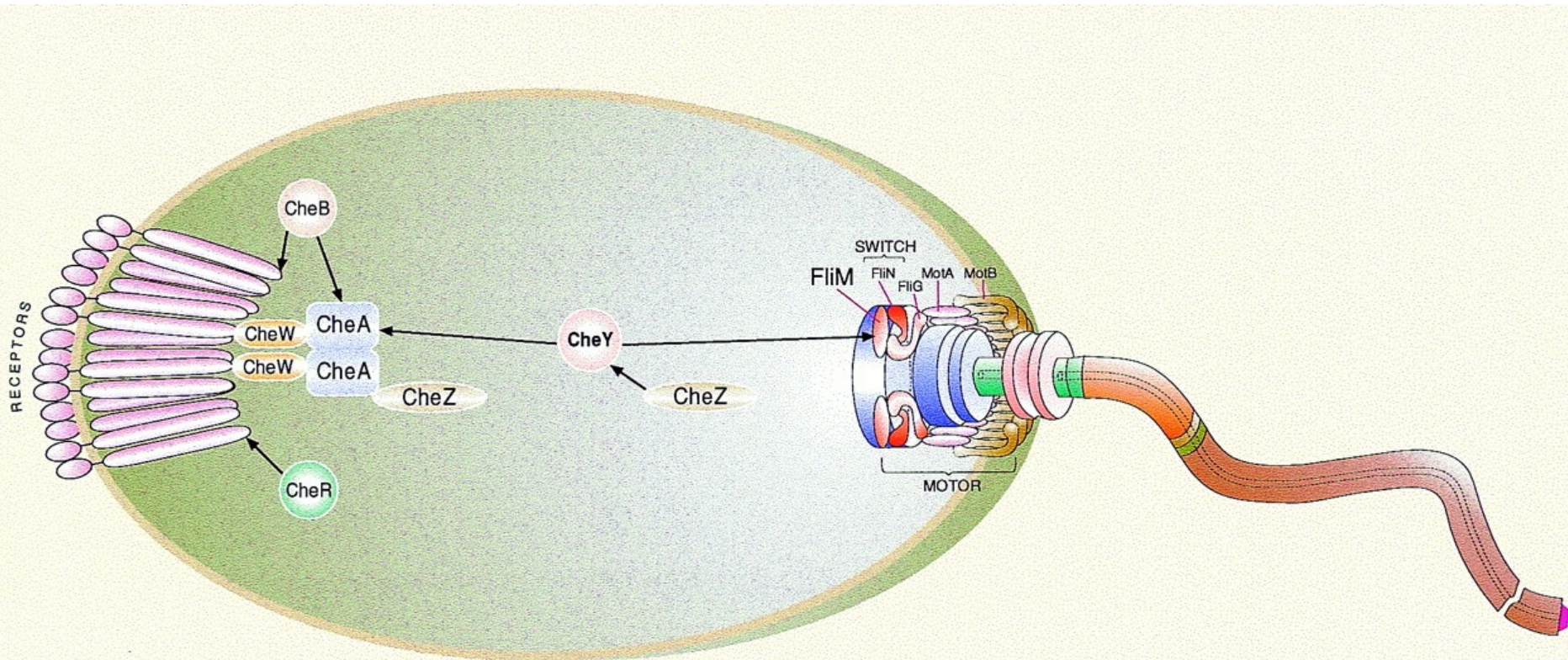
$E$  is the mean error for the promoter activity prediction (see *Methods*).



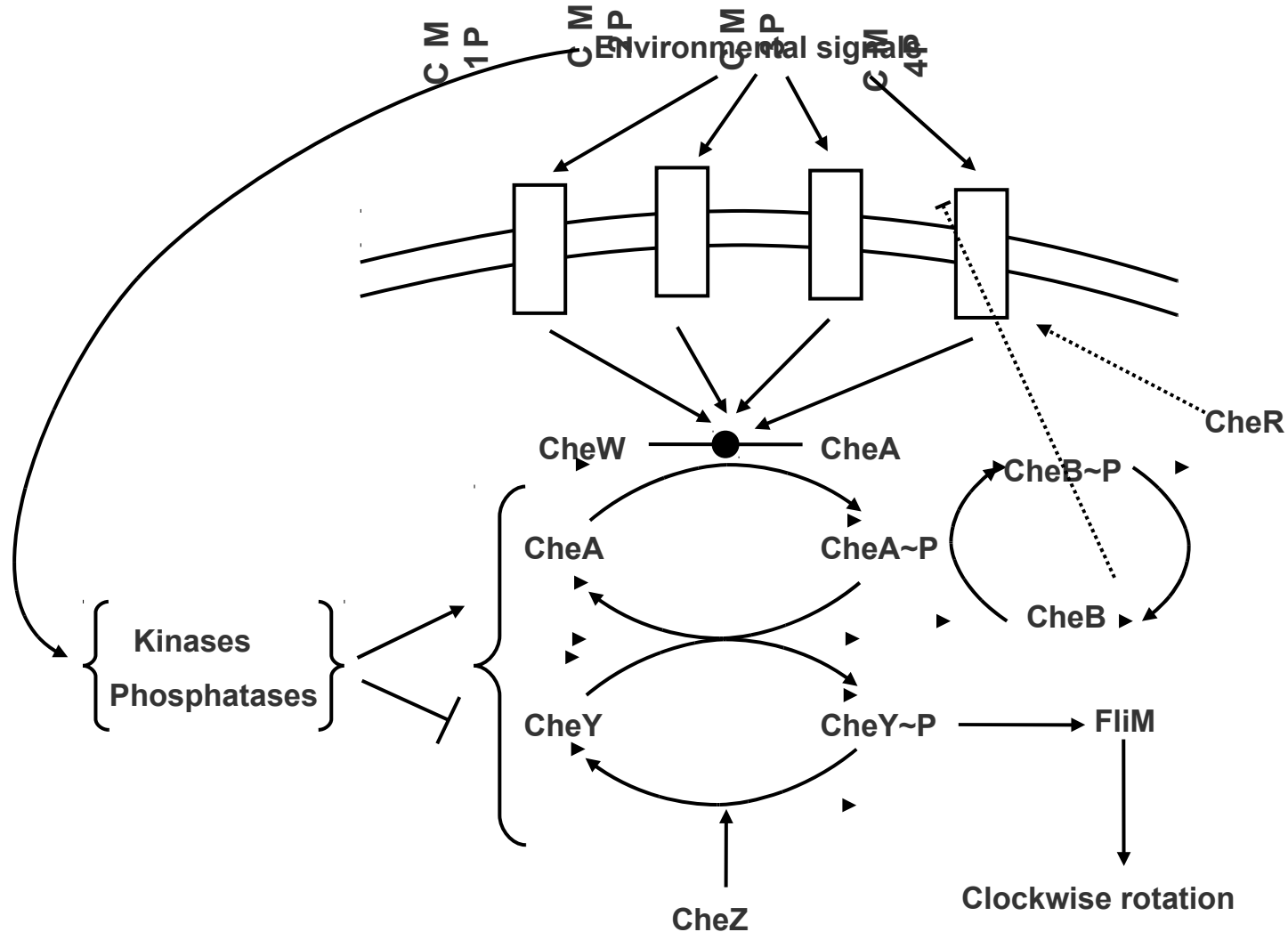
# Les paramètres prédisent correctement le comportement du système



# Bacterial chemotaxis

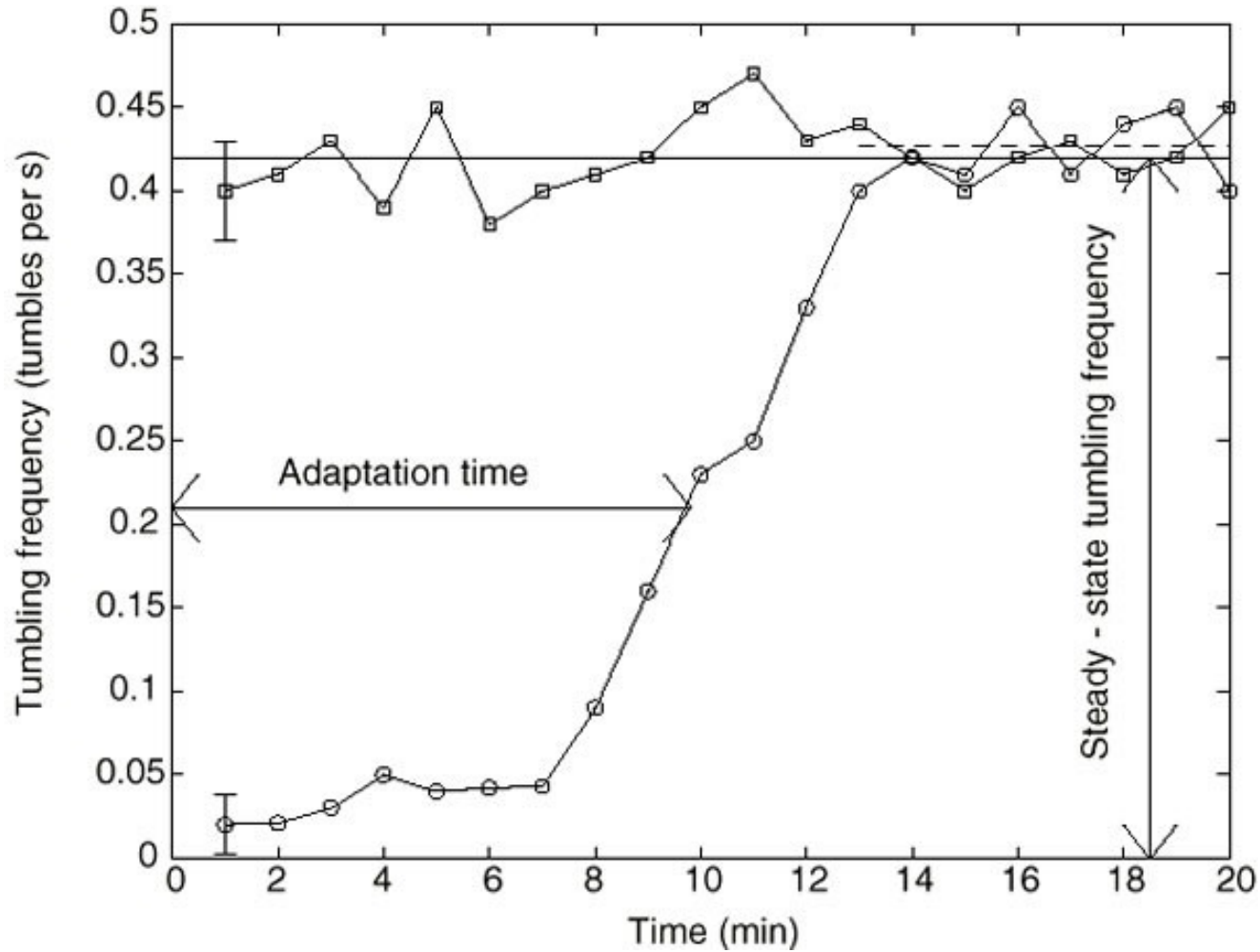


# Basic chemotaxis regulation network

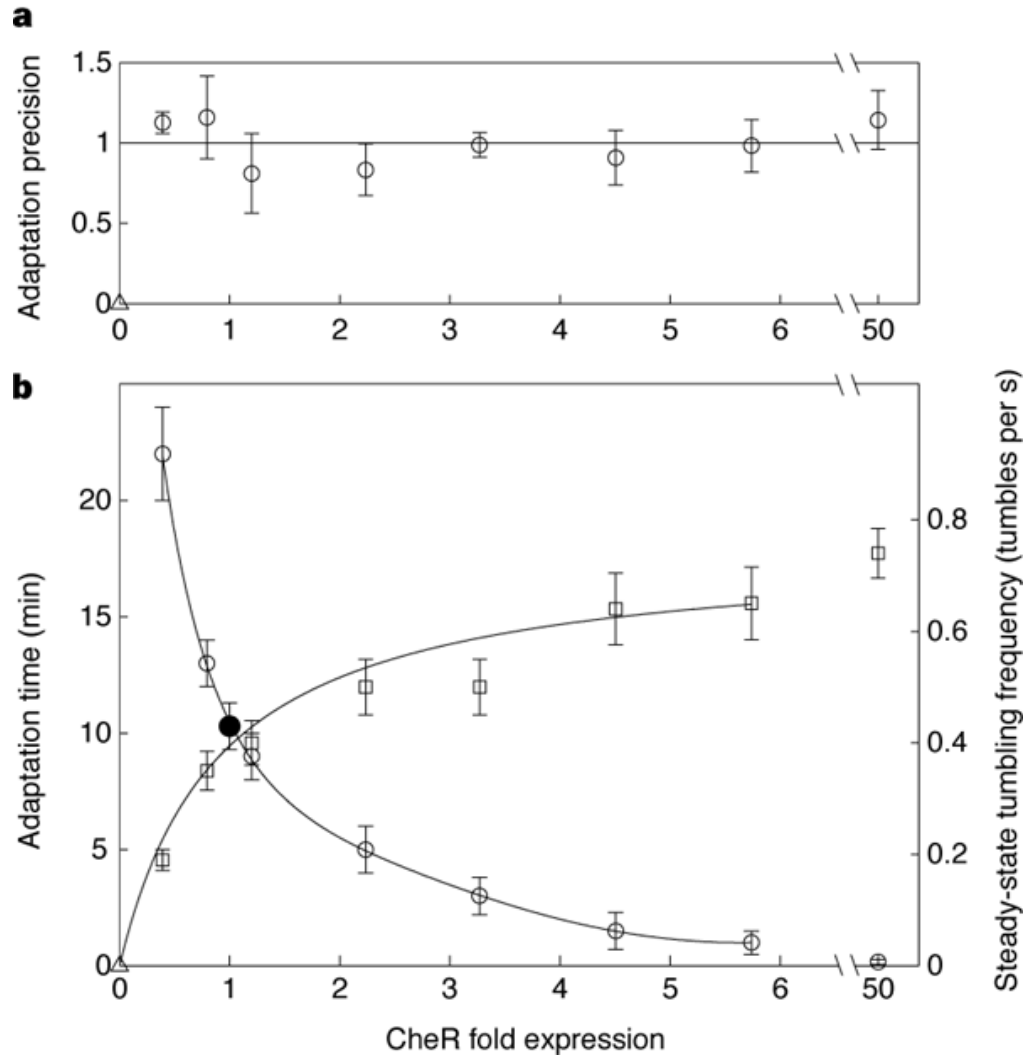




# Robustness in bacterial chemotaxis



# Robustness in bacterial chemotaxis

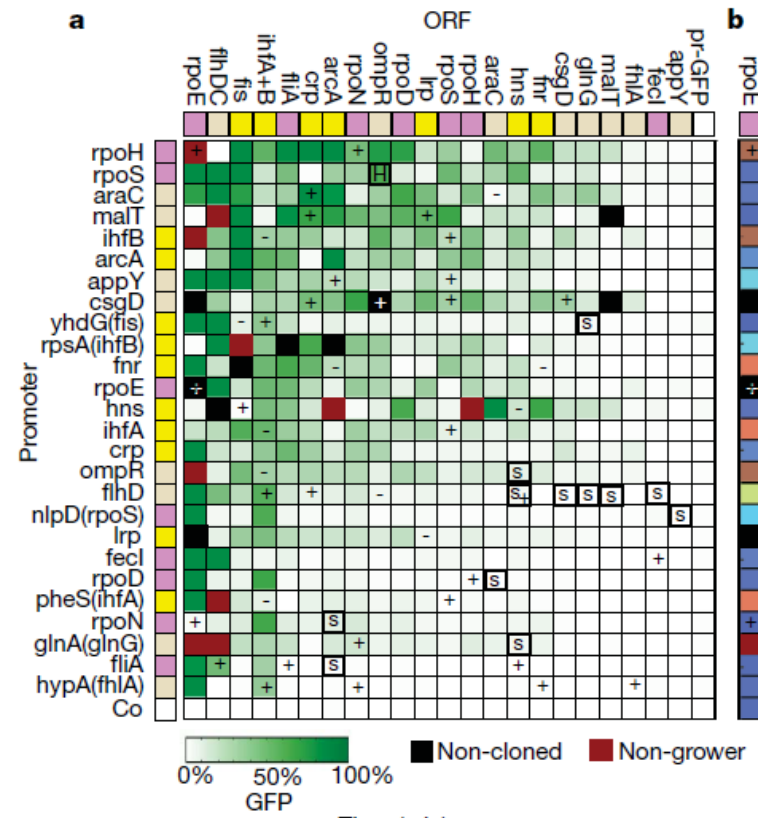
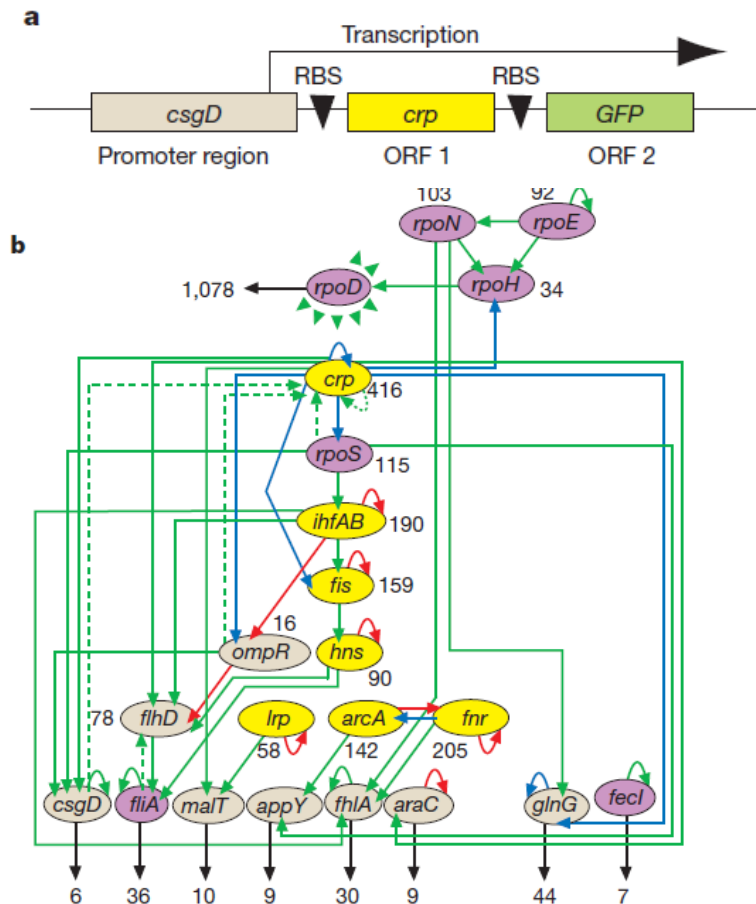


# Robustness of adaptation precision

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

# 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

