Stochastic models of protein production with feedback

Renaud Dessalles
joint work with Vincent Fromion and Philippe Robert

INRA Jouy-en-Josas - INRIA Rocquencourt
(Fance)

16th. May 2016

Presentation

Biological context

Mathematical framework

Equilibrium results

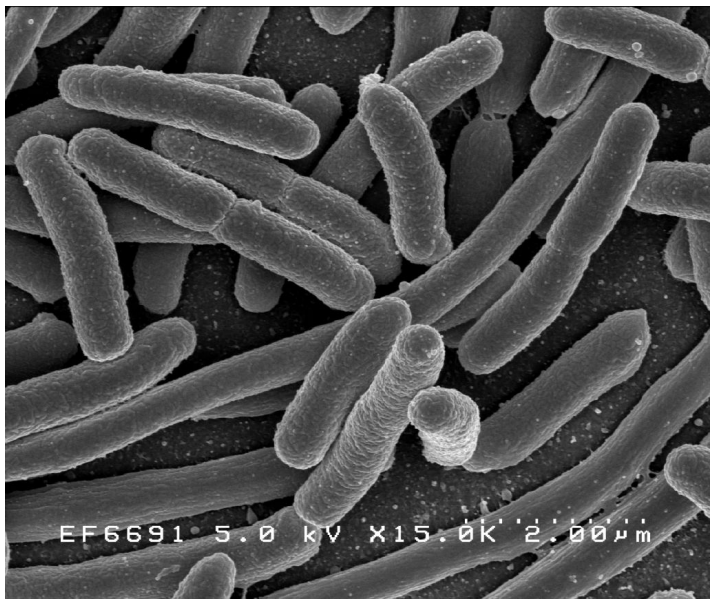
Other aspects of the controlled model

Part 1

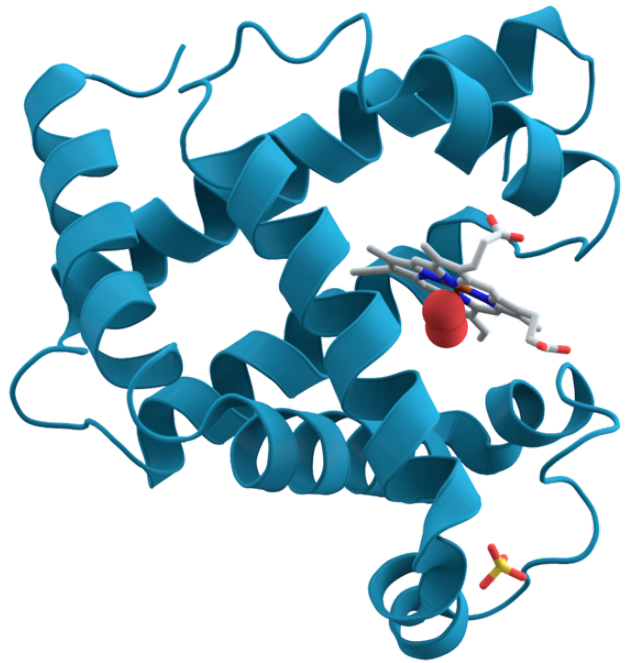
Biological context

Cells and proteins

- Cells: unit of life.
- Its goal: grow and divide.



- Functional molecules: *proteins*
  - enzymes, wall, energy, etc.
- Produced from the genes



Protein production: A central mechanism

Proteins represents:

- 50% of the dry mass
- $\sim$ 3 million molecules
- $\sim$ 2000 different types
- from few dozens up to 10^5 proteins per type

It needs to be duplicated in one cell cycle (approx. 30 *min*)

Protein production: A central mechanism

Proteins represents:

- 50% of the dry mass
- $\sim$ 3 million molecules
- $\sim$ 2000 different types
- from few dozens up to 10^5 proteins per type

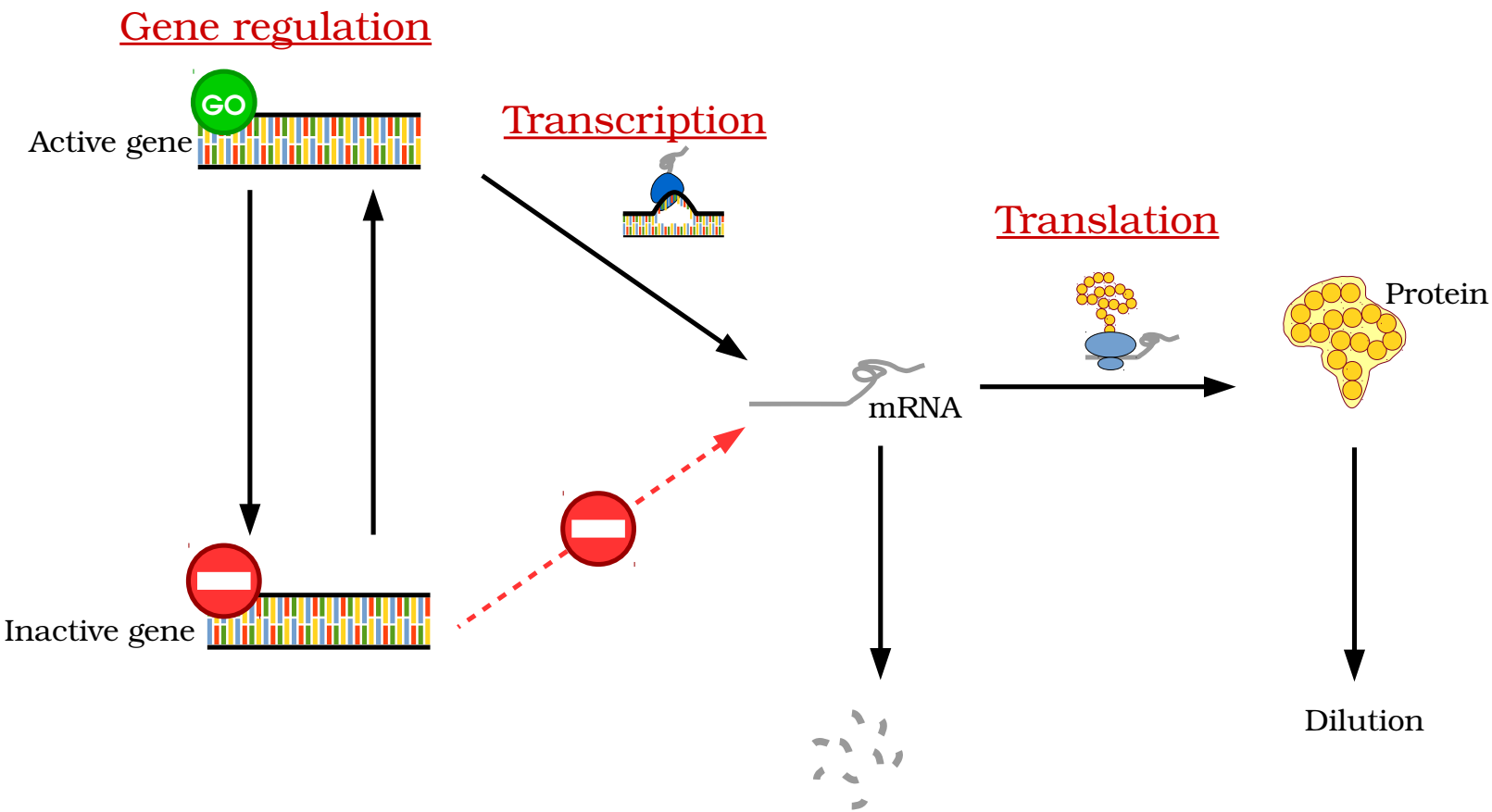
It needs to be duplicated in one cell cycle (approx. 30 *min*)

85% of the resources for protein production

Classic protein production mechanism

Protein production in 3 steps:

1. Gene regulation
2. Transcription: to produce mRNA
3. Translation: to produce proteins



Highly variable process

The protein production subject to high variability:

- Interior of bacteria: non-organized medium
- Mobility of compounds: through random diffusion
- Cellular mechanism: random collision between molecules

Highly variable process

The protein production subject to high variability:

- Interior of bacteria: non-organized medium
- Mobility of compounds: through random diffusion
- Cellular mechanism: random collision between molecules

Problem: 85% of the resources for the protein production, impacted by a large variability.

Highly variable process

The protein production subject to high variability:

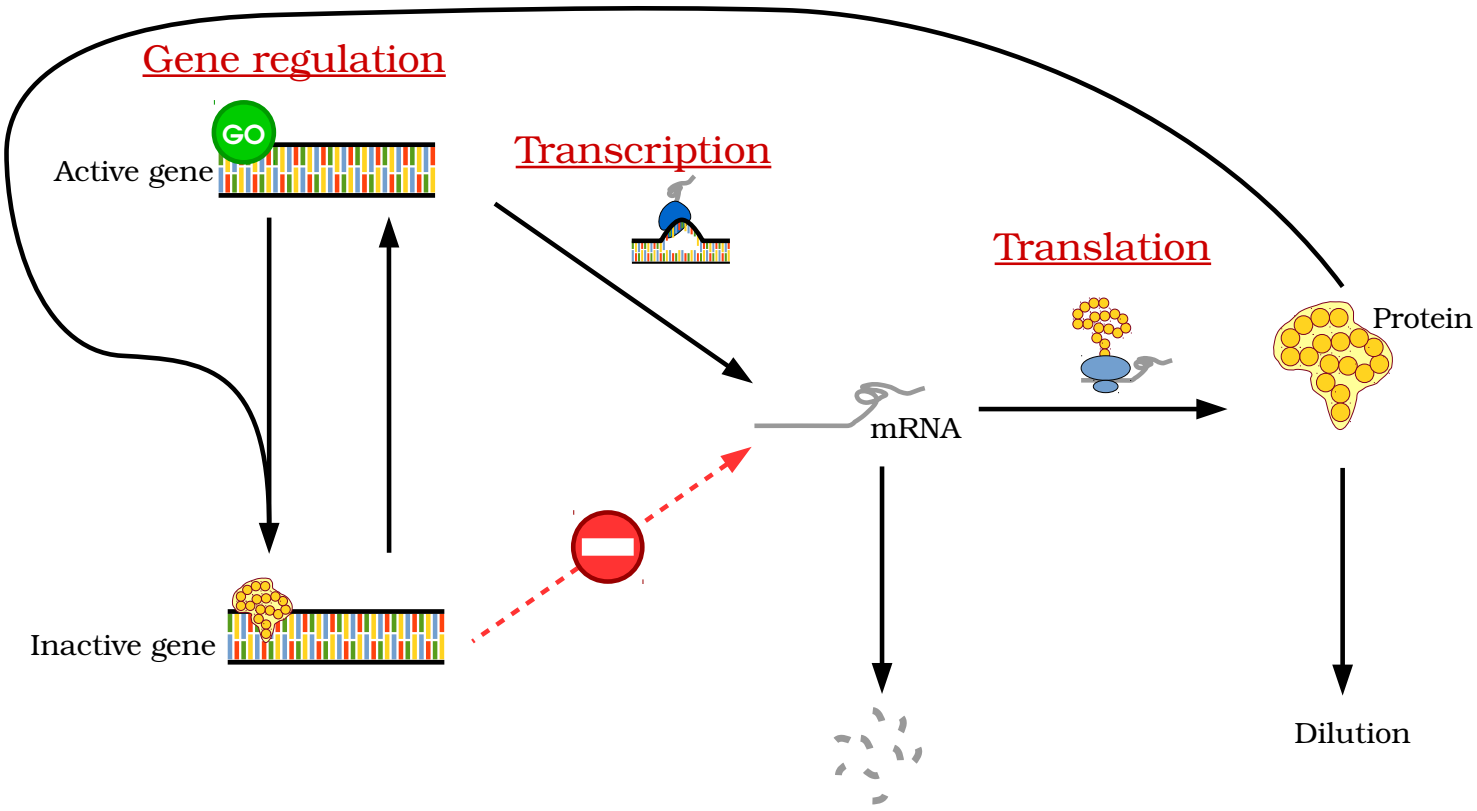
- Interior of bacteria: non-organized medium
- Mobility of compounds: through random diffusion
- Cellular mechanism: random collision between molecules

Problem: 85% of the resources for the protein production, impacted by a large variability.

A main issue for the bacteria:
control the variability in protein production.

Protein production mechanism with feedback

Production with feedback: the protein binds to its own gene.



More proteins ⇒ Gene more inactive

A way to reduce variability?

Comparison of models

Classical production vs *Feedback production*

- Conjecture: less variability in proteins with feedback production.

Comparison of models

Classical production vs *Feedback production*

- Conjecture: less variability in proteins with feedback production.

Our Goal

Comparison of distributions of proteins in the two models.

Part 2

Mathematical framework

Markovian description

Framework for protein production modeling:

- Rigney and Schieve (1977)
- Berg (1978)
- Paulsson (2005)

Three types of events:

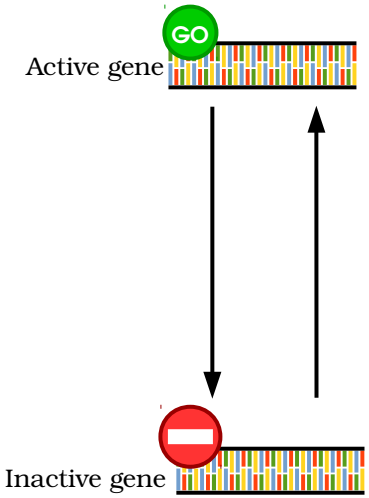
- Encounter between molecules
- Elongation of molecules
- Lifetime of molecules

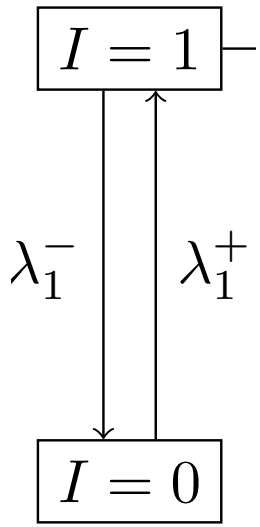
Assumption: Exponential times

Each event occurs at exponential time.

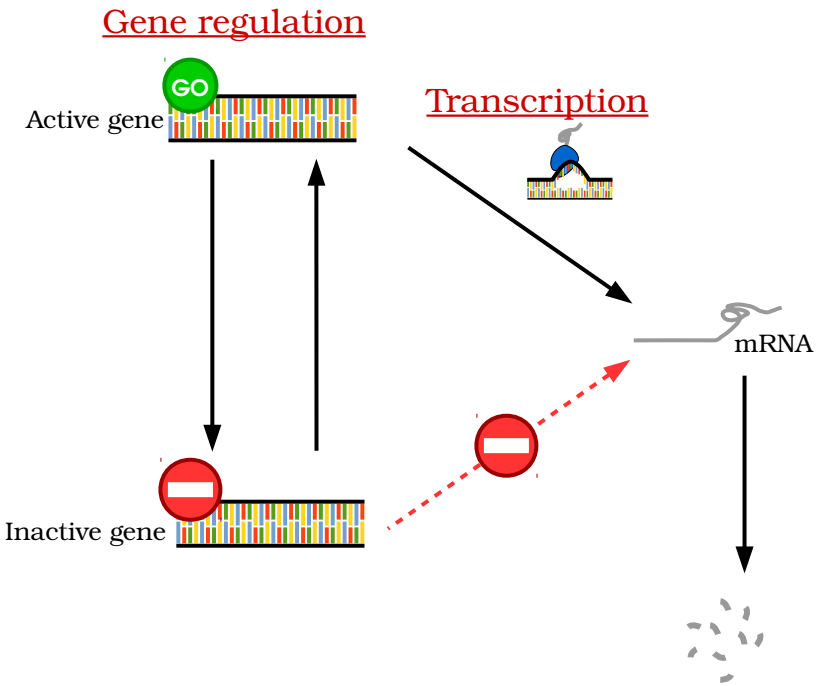
The classical model

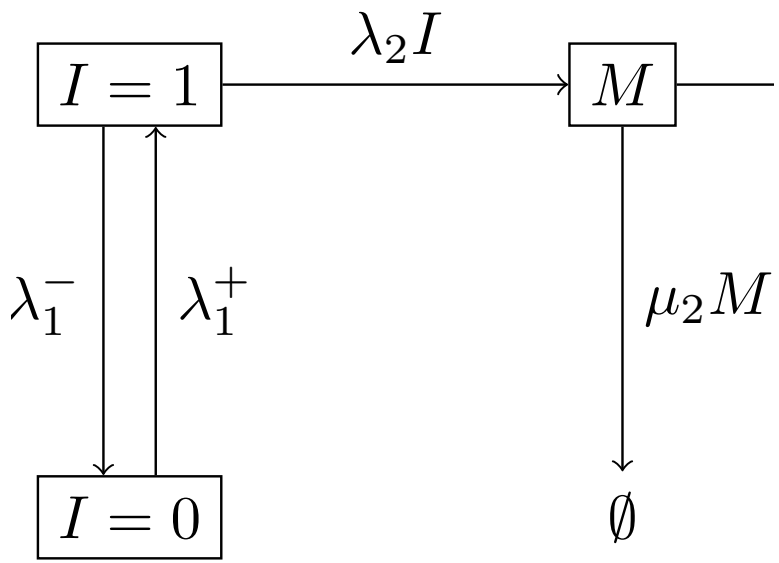
Gene regulation



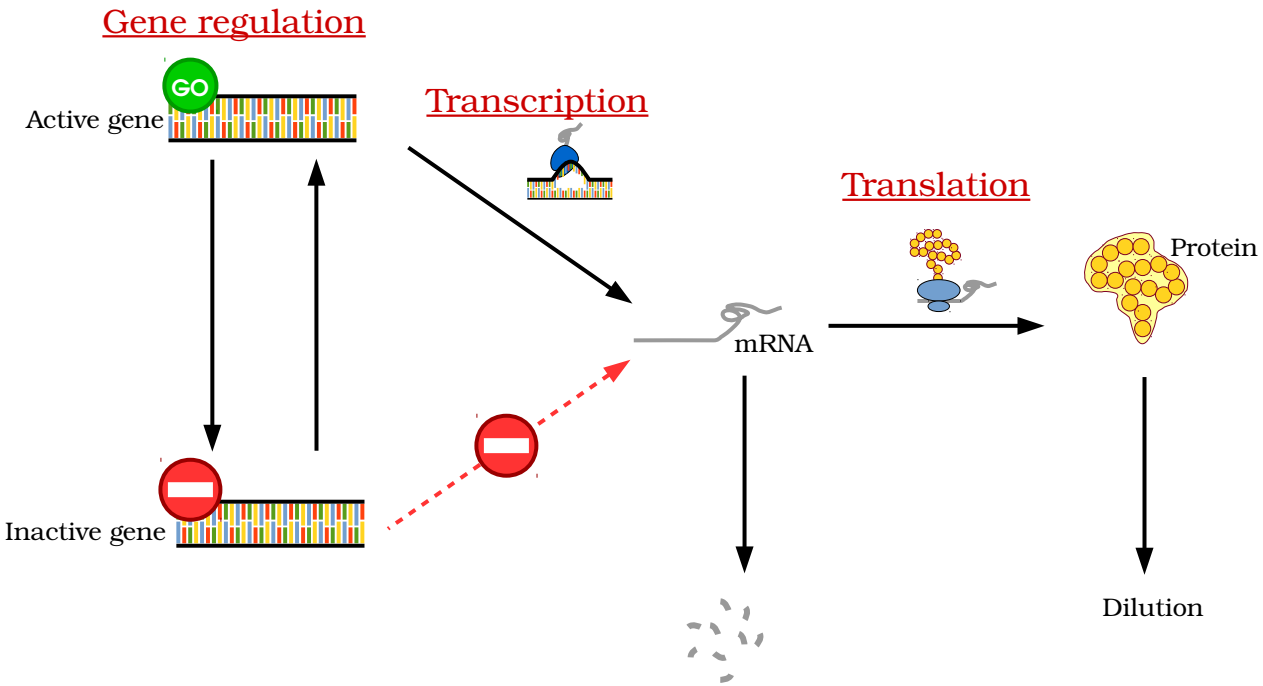


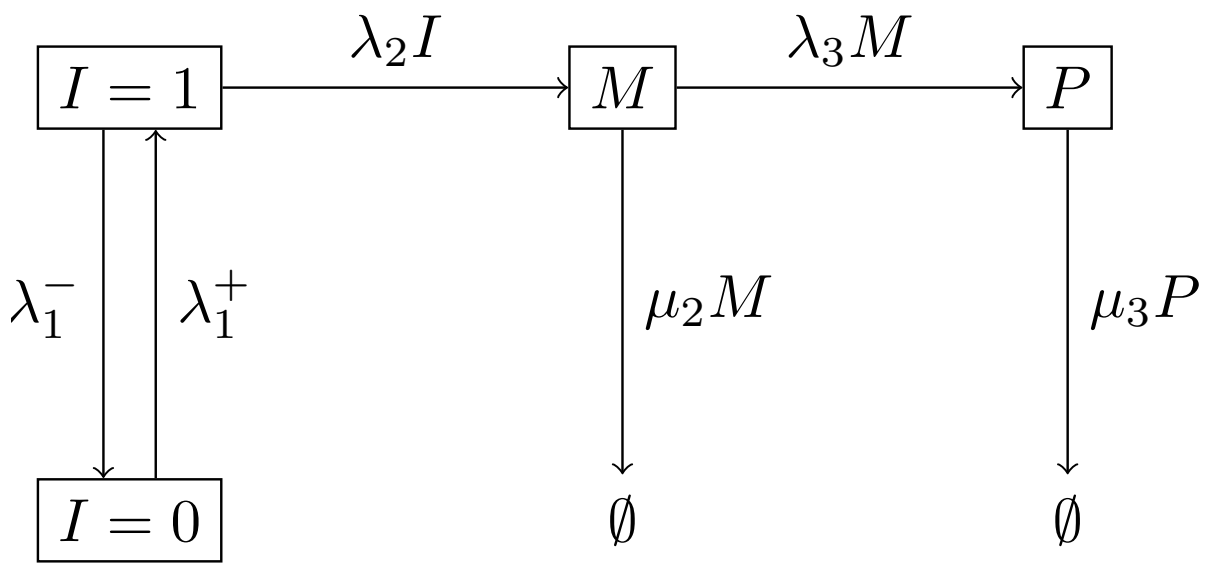
The classical model





The classical model





Mean and variance for the classical model

For the classical model, the mean and the variance are known Paulsson (2005):

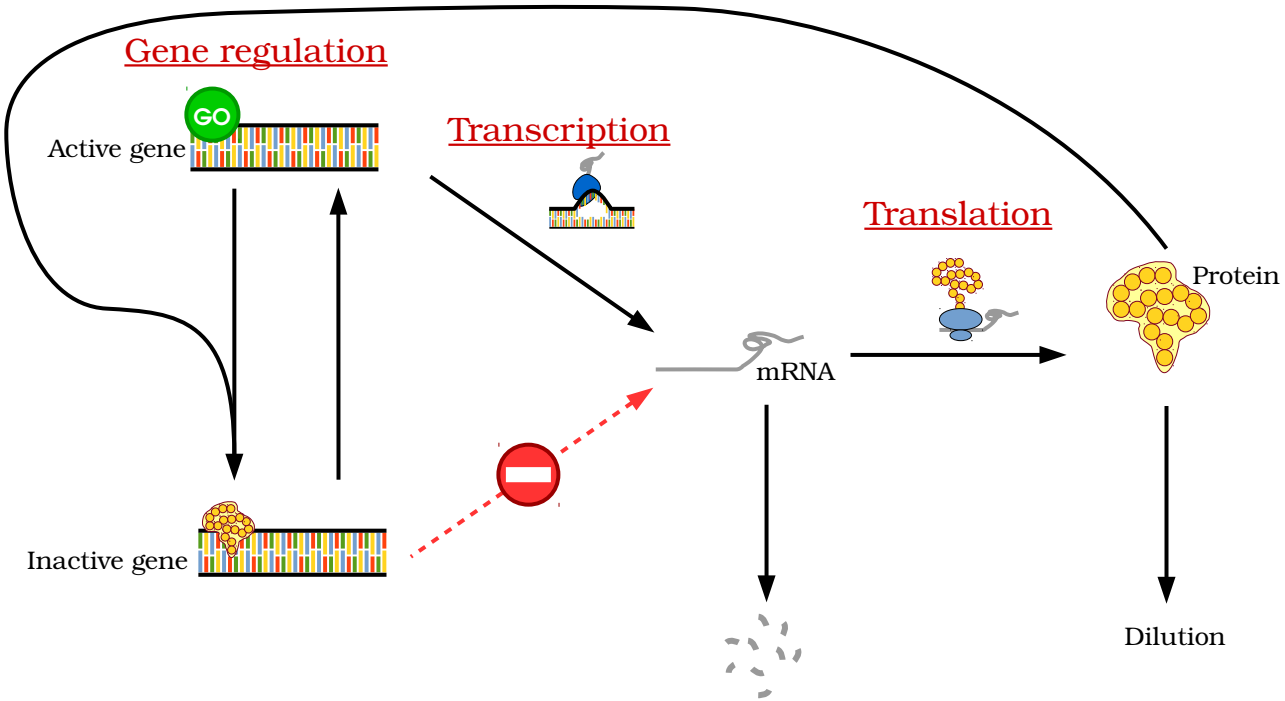
- Equality of flows gives

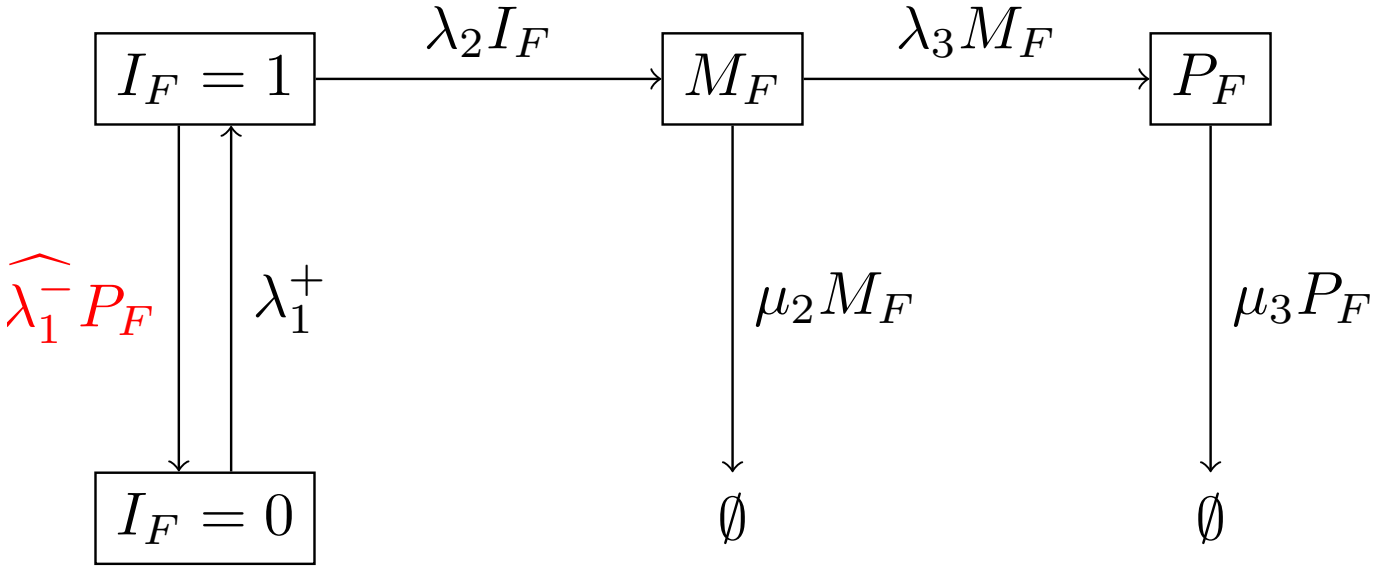
$$\mathbb{E}[P] = \frac{\lambda_1^+}{\lambda_1^+ + \lambda_1^-} \cdot \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_3}{\mu_3}$$

- Equilibrium equations give:

$$\begin{aligned}\mathbb{V}ar[P] &= \mathbb{E}[P]\left(1 + \frac{\lambda_3}{\mu_2+\mu_3}\right. \\ &\left. + \frac{\lambda_2\lambda_3\left(1-\lambda_1^+/\left(\lambda_1^+ + \lambda_1^-\right)\right)\left(\lambda_1^+ + \lambda_1^- + \mu_2 + \mu_3\right)}{\left(\mu_2+\mu_3\right)\left(\lambda_1^+ + \lambda_1^- + \mu_2\right)\left(\lambda_1^+ + \lambda_1^- + \mu_3\right)}\right).\end{aligned}$$

The feedback model





Mean and variance for the Feedback model

- Equality of flows gives

$$\mathbb{E}[P_C] = \mathbb{E}[I_C] \cdot \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_3}{\mu_3}.$$

- Problem : no known expression for $\mathbb{E}[I_C]$:
  - the equality of flows on I_C:

$$\widehat{\lambda_1^-}\,\mathbb{E}[I_C P_C] = \lambda_1^+ (1 - \mathbb{E}[I_C]).$$

Difficulties to make comparisons between the two models

Part 3

Equilibrium results

Scaling

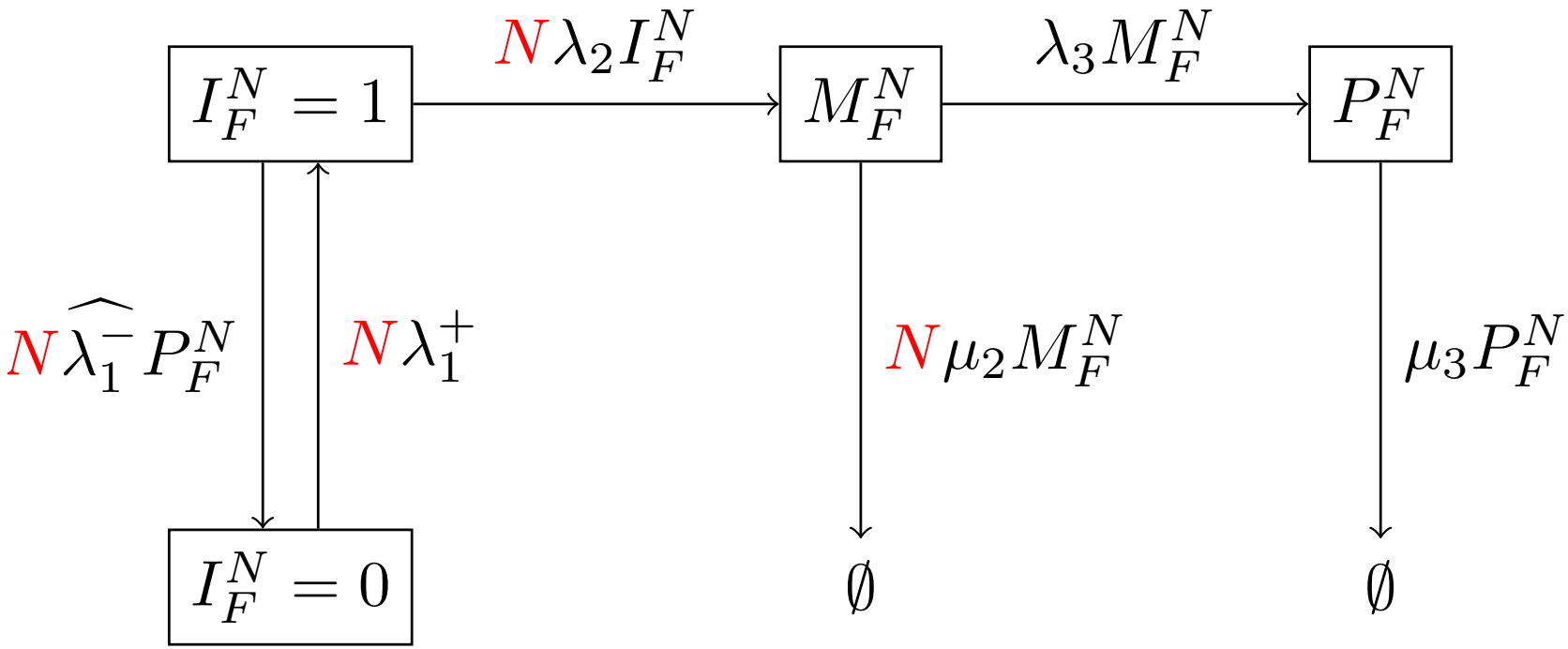
- Introduction of a scaling:

$$\left.\begin{array}{l}\text{Gene regulation timescale}\\ \text{Messenger RNA timescale}\end{array}\right\} \text{faster than the protein time scale}$$

Scaling

- Introduction of a scaling:

Gene regulation timescale
Messenger RNA timescale
} faster than the protein time scale



N scaling parameter

Effects of the scaling

Example: Feedback model

- State of the model:

$$\left(I_F^N(t), M_F^N(t), P_F^N(t)\right)$$

  - I_F^N and M_F^N on a quick timescale.
  - P_F^N on a slow timescale.

I_F^N and M_F^N reach some equilibrium depending on the slow current $P_F^N(t)$ state

Convergence of the gene regulation and the messengers

τ_1^N: the first time of jump of P_F^N;
Starting at number of proteins $x = P_F^N(0)$;

Convergence of the gene regulation and the messengers

τ_1^N: the first time of jump of P_F^N;
Starting at number of proteins $x = P_F^N(0)$;

- $\left(I_F^N(t)\right)$ reaches its equilibrium quickly:

$$\mathbb{E}\left[I_F^N(t) | 0 < t < \tau_1^N, P_F^N(0) = x\right] \xrightarrow{N\to\infty} \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-} x}.$$

Convergence of the gene regulation and the messengers

τ_1^N: the first time of jump of P_F^N;
Starting at number of proteins $x = P_F^N(0)$;

- $\left(I_F^N(t)\right)$ reaches its equilibrium quickly:

$$\mathbb{E}\left[I_F^N(t) | 0 < t < \tau_1^N, P_F^N(0) = x\right] \xrightarrow{N\to\infty} \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-}x}.$$

- $\left(M_F^N(t)\right)$ reaches its equilibrium quickly:

$$\mathbb{E}\left[M_F^N(t) | 0 < t < \tau_1^N, P_F^N(0) = x\right] \xrightarrow{N\to\infty} \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-}x}.$$

Convergence of the gene regulation and the messengers

τ_1^N: the first time of jump of P_F^N;
Starting at number of proteins $x = P_F^N(0)$;

- $\left(I_F^N(t)\right)$ reaches its equilibrium quickly:

$$\mathbb{E}\left[I_F^N(t) | 0 < t < \tau_1^N, P_F^N(0) = x\right] \xrightarrow{N\to\infty} \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-}x}.$$

- $\left(M_F^N(t)\right)$ reaches its equilibrium quickly:

$$\mathbb{E}\left[M_F^N(t) | 0 < t < \tau_1^N, P_F^N(0) = x\right] \xrightarrow{N\to\infty} \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-}x}.$$

- Rate of production of proteins tends to:

$$\lambda_3 \cdot \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-}x}.$$

Convergence of the models

Theorem

The process $\left(P_F^N(t)\right)$ *converges in distribution to a birth and death process with (x number of proteins):*

$$\beta_x = \lambda_3 \cdot \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-} x} \qquad \textit{and} \qquad \delta_x = \mu_3 x.$$

Convergence of the models

Theorem

The process $\left(P_F^N(t)\right)$ converges in distribution to a birth and death process with (x number of proteins):

$$\beta_x = \lambda_3 \cdot \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-} x} \qquad \text{and} \qquad \delta_x = \mu_3 x.$$

Idem for uncontrolled model:

Theorem

The process $\left(P^N(t)\right)$ converges in distribution to a birth and death process with (x number of proteins):

$$\beta_x = \lambda_3 \cdot \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_1^+}{\lambda_1^+ + \lambda_1^-} \qquad \text{and} \qquad \delta_x = \mu_3 x.$$

Feature of the scaled models

- Equilibrium distributions.
  - Classical model: P^{∞} follow a *Poisson distribution*:

$$P^{\infty} \sim \mathcal{P}\left(\frac{\lambda_3}{\mu_3} \cdot \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_1^+}{\lambda_1^+ + \lambda_1^-}\right)$$

  - Feedback model: P^{∞} follow the limit distribution

$$\pi_F(x) = \frac{1}{Z \cdot x!}\left(\frac{\lambda_3}{\mu_3} \cdot \frac{\lambda_2}{\mu_2}\right)^x \prod_{i=0}^{x-1} \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-}\, i}$$

Feature of the scaled models

- Equilibrium distributions.
  - Classical model: P^∞ follow a *Poisson distribution*:

$$P^\infty \sim \mathcal{P}\left(\frac{\lambda_3}{\mu_3} \cdot \frac{\lambda_2}{\mu_2} \cdot \frac{\lambda_1^+}{\lambda_1^+ + \lambda_1^-}\right)$$

  - Feedback model: P^∞ follow the limit distribution

$$\pi_F(x) = \frac{1}{Z \cdot x!}\left(\frac{\lambda_3}{\mu_3} \cdot \frac{\lambda_2}{\mu_2}\right)^x \prod_{i=0}^{x-1} \frac{\lambda_1^+}{\lambda_1^+ + \widehat{\lambda_1^-}\, i}$$

Variance comparison

$$\mathbb{V}ar\left[P^\infty\right] = \mathbb{E}\left[P^\infty\right] \qquad \text{and} \qquad \mathbb{V}ar\left[P_F^\infty\right] \leq \mathbb{E}\left[P_F^\infty\right]$$

Asymptotic behaviour of the Feedback model

Introducing:

$$\rho := \frac{\lambda_1^+}{\lambda_1^-}\frac{\lambda_3}{\mu_3}\cdot\frac{\lambda_2}{\mu_2} \quad \text{and} \quad \eta := \frac{\lambda_1^+}{\lambda_1^-} - 1$$

it comes:

$$\pi_F(x) = \frac{1}{Z\cdot x!}\rho^x \prod_{i=1}^{x}\frac{1}{\eta+i}.$$

Asymptotic behaviour

Increase ρ while keeping η constant:
Increase protein production while keeping the gene regulation constant

Asymptotic behaviour of the Feedback model

With Laplace method:

Theorem

Convergence in distribution:

$$\lim_{\rho\to\infty} \frac{P_F^\infty - a_\rho}{\sqrt{a_\rho}} = \mathcal{N}\left(0, 1/\sqrt{2}\right)$$

with $a_\rho = \left(\sqrt{\eta^2 + 4\rho} - \eta\right)$.

Corollary

$$\lim_{\rho\to\infty} \frac{\mathbb{E}\left[P_F^\infty\right]}{\sqrt{\rho}} = 1 \quad \text{and} \quad \lim_{\rho\to\infty} \frac{\mathbb{V}ar\left[P_F^\infty\right]}{\mathbb{E}\left[P_F^\infty\right]} = \frac{1}{2}$$

Conclusion for the noise control

For the scaled model:

$$\mathbb{V}ar\left[P^{\infty}\right] = \mathbb{E}\left[P^{\infty}\right] \qquad \text{and} \qquad \mathbb{V}ar\left[P_F^{\infty}\right] \leq \mathbb{E}\left[P_F^{\infty}\right]$$

Asymptotic behaviour:

$$\mathbb{V}ar\left[P^{\infty}\right] = \mathbb{E}\left[P^{\infty}\right] \qquad \text{and} \qquad \mathbb{V}ar\left[P_F^{\infty}\right] \underset{\rho\to\infty}{\sim} \frac{1}{2}\mathbb{E}\left[P_F^{\infty}\right]$$

The reduction of variance is limited in feeback model.

Part 4

Other aspects of the controlled model

Dynamical aspects

Equilibrium reaching

Which model go faster to reach the equilibrium?

Biological example: need for a quick activation of the protein production.

Dynamical aspects

Equilibrium reaching

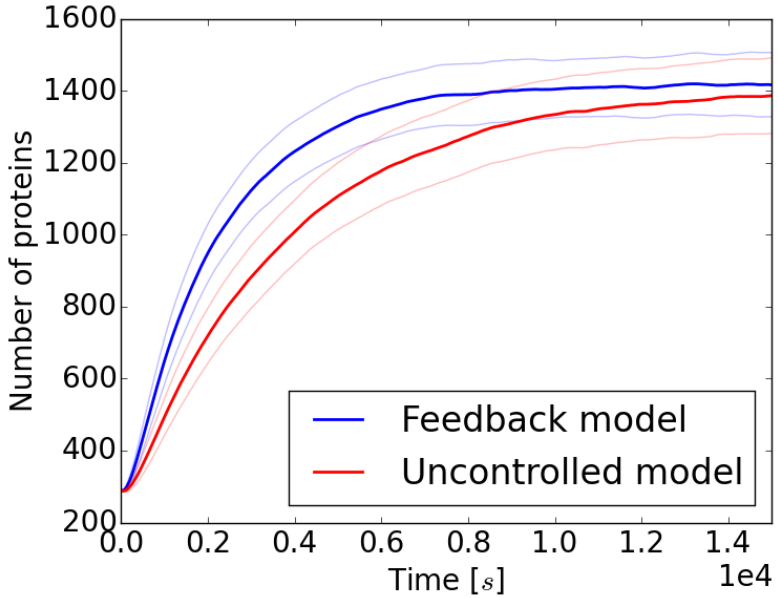
Which model go faster to reach the equilibrium?

Biological example: need for a quick activation of the protein production.

Simulations: comparison for the two models:

- Starting at a low protein production
- Evolution to a high level of protein production

Simulation for dynamical aspects



- 1000 simulations
  - Thick lines: average of trajectories
  - Fine lines: $\pm$ standard deviation
- Here, controlled model is 20% faster

Thank you for you attention

Article:

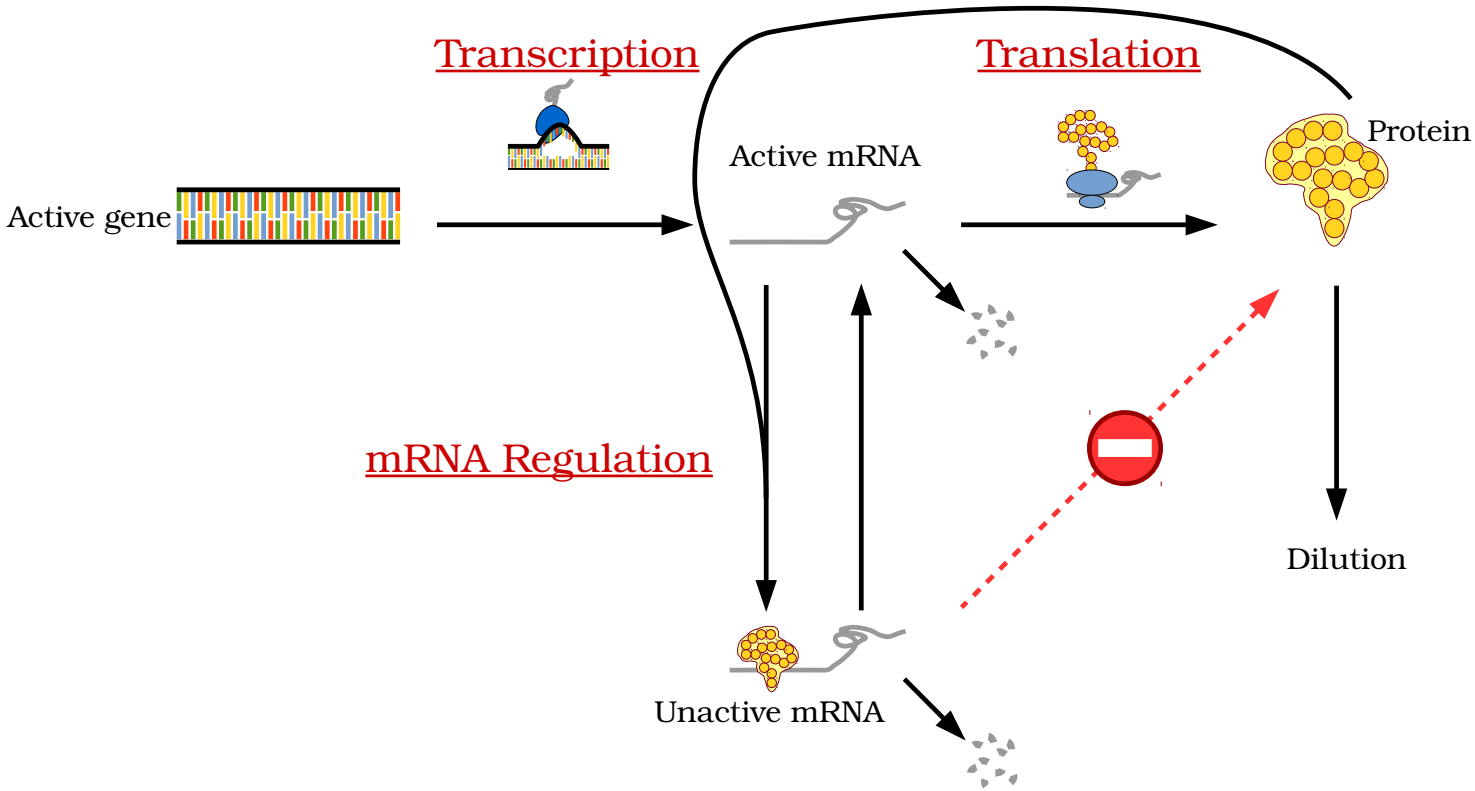
Dessalles, R., Fromion, V., and Robert, P. (2015). arXiv:1509.02045

PhD work supervised by

- Vincent Fromion
- Philippe Robert

Other work

- Regulation on the mRNA rather than on the gene



- Intermediate metabolite step in regulation