Stochastic models of protein production with cell division and gene replication

Renaud Dessalles
joint work with Vincent Fromion and Philippe Robert

INRA Jouy-en-Josas - INRIA Paris
(France)

29th. August 2016
Presentation

Biological context

Classical models for protein production

Model with cell division and gene replication

Results and further work
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
- ~3 million molecules
- ~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)

67% of the resources for protein production
Classic protein production mechanism

Two main steps in protein production:

1. Transcription: to produce mRNA
2. Translation: to produce proteins

Transcription

Translation

Gene mRNA Protein

Gene

mRNA

Protein
Highly variable process

The protein production subject to high variability:

- Thermal noise (random collision between molecules)
- Cell events (division, gene replication)
- Fluctuations in common resources
Highly variable process

The protein production subject to high variability:

- Thermal noise (random collision between molecules)
- Cell events (division, gene replication)
- Fluctuations in of common resources

Problem: the main mechanism of the cell, impacted by a large variability.
Highly variable process

The protein production subject to high variability:

- Thermal noise (random collision between molecules)
- Cell events (division, gene replication)
- Fluctuations in of common resources

Problem: the main mechanism of the cell, impacted by a large variability.

“How the cell deals with this variability?”
A main topic for experimental research.
Taniguchi et al. (2010) experimental measures

Population of cells
- Measure volume $v_i$
- Measure of prot. number $p_i$

Interest in concentrations

*Empirical mean:*

$$\mu_p = \frac{1}{N} \sum_{i=1}^{N} \frac{p_i}{v_i}$$

*Empirical variance:*

$$\sigma_p^2 = \frac{1}{N} \sum_{i=0}^{N} \left( \frac{p_i}{v_i} \right)^2 - \mu_p^2$$
Taniguchi et al. (2010) experimental measures

Two regimes in the protein variability:

\[ \frac{\sigma_p^2}{\mu_p^2} \]

\[ \mu_p \text{ [copies/µm}^3\text{]} \]
Goal: modelling the protein production

- Models to describe the stochastic protein variability.
- Confront the models to real experiments (two regimes)
Part 2

Classical models for protein production
Markovian description

Framework for protein production modeling:

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

Three types of events:

- Encounter between molecules
- mRNA and protein creation
- Lifetime of molecules

Assumption: Exponential times

Each event occurs at exponentially distributed time.
The classical model
The classical model

Transcription

Gene → mRNA → Protein

mRNAs

$\lambda_1 M$ → $\sigma_1 M$ → $\emptyset$
The classical model

Transcription

Gene → mRNA

Translation

mRNA → Protein

mRNAs

$\lambda_1 M \xrightarrow{+1} \lambda_2 M \xrightarrow{+1} P \xrightarrow{-1} \sigma_2 P$

Proteins

Dilution

$\sigma_1 M \xrightarrow{-1} \emptyset$
Limitations of classical models

Classical model, at equilibrium mean $\mathbb{E}[P]$ and the variance $\text{Var}[P]$ are known Paulsson (2005).

But this model has some limitations:

- it does not take into account the division
Limitations of classical models

Classical model, at equilibrium mean $\mathbb{E}[P]$ and the variance $\text{Var}[P]$ are known Paulsson (2005).

But this model has some limitations:

- it does not take into account the division
- it does not consider gene replication
Limitations of classical models

Classical model, at equilibrium mean $\mathbb{E}[P]$ and the variance $\text{Var}[P]$ are known Paulsson (2005).

But this model has some limitations:

- it does not take into account the division
- it does not consider gene replication
- it represents numbers and not concentrations
Limitations of classical models

Classical model, at equilibrium mean $\mathbb{E}[P]$ and the variance $\text{Var}[P]$ are known Paulsson (2005).

But this model has some limitations:

- It does not take into account the division
- It does not consider gene replication
- It represents numbers and not concentrations

We need to have a model with the notion of cell cycle.
Part 3

Model with cell division and gene replication
Features of the model

A model with cell cycle:
- Considering a growing cell
- Gene replication at $\tau_R$
- Division at $\tau_D$

Times $\tau_R$ and $\tau_D$ are considered deterministic.
Presentation of the model

Before replication:

After replication:

Volume growth:

$$V(s) = V(0)2^{s/\tau_D}$$
Presentation of the model

Before replication:
\[ \lambda_1 \]
\[ 2\lambda_1 \]

After replication:
\[ M \]
\[ \sigma_1 M \]
\[ \emptyset \]

Proteins

Periodic divisions every \( \tau_D \)

Volume growth:
\[ V(s) = V(0)2^{s/\tau_D} \]

Concentrations can be considered:
\[ P_s/V(s) \] and \[ M_s/V(s) \]
Explicit solution for the number of mRNAs

For any time $s$ of the cell cycle the distribution of $M_s$ is known.

**Theorem**

At equilibrium, at a time $s$ in the cell cycle, the mRNA number $M_s$ follows a Poisson distribution of parameter

$$x_s = \frac{\lambda_1}{\sigma_1} \left[ 1 - \frac{e^{-(s+\tau_D-\tau_R)\sigma_1}}{2 - e^{-\tau_D\sigma_1}} + \mathbb{1}_{s \geq \tau_R} \left( 1 - e^{-(s-\tau_R)\sigma_1} \right) \right].$$

We need to use Marked Poisson Point Process for the proof.
Explicit solution for the mean and the variance

With more calculus, the first two moments of $P_s$ are known.
Explicit solution for the mean and the variance

With more calculus, the first two moments of $P_s$ are known.

**Theorem**

*At equilibrium, at any time $s$ of the cell cycle, the mean and the variance of the protein number $P_s$ are*

\[
\mathbb{E}[P_s] = \lambda_2 (f_1(\tau_R) + f_2(\tau_D) + f_1(\tau_R \wedge s) + \mathbb{1}_{s \geq \tau_R} f_2(s))
\]

\[
\text{Var}[P_s] = \text{Var}[P_0] + 2\lambda_2 \frac{1 - e^{-\sigma_1 s \wedge \tau_R}}{\sigma_1} \text{Cov}[P_0, M_0] + g_1(s \wedge \tau_R)
\]

\[
+ \mathbb{1}_{s \geq \tau_R} \left( 2\lambda_2 \frac{1 - e^{-\sigma_1 (s - \tau_R)}}{\sigma_1} \text{Cov}[P_{\tau_R}, M_{\tau_R}] + g_2(s) \right)
\]

*with $f_1, f_2, g_1, g_2, \text{Var}[P_0], \text{Cov}[P_0, M_0]$ and $\text{Cov}[P_{\tau_R}, M_{\tau_R}]$ explicitly depending on $\lambda_1, \sigma_1, \lambda_2, \tau_R$ and $\tau_D$.***
Part 4

Results and further work
Parameters

We use the empirical mean and variance of proteins in Taniguchi et al. (2010) to fit the parameters.
Parameters

We use the empirical mean and variance of proteins in Taniguchi et al. (2010) to fit the parameters. For each type of protein:

\[ \mu_p = \frac{1}{\tau_D} \int_0^{\tau_D} \frac{\mathbb{E} [P_s]}{V(s)} \, ds. \]
Protein profile

The previous theorem can predict the protein variability:

Simulations

Experiments adapted from fig 4.b of Walker et al. (2016)
Protein profile

The previous theorem can predict the protein variability:

Simulations

Experiments

adapted from fig 4.b of Walker et al. (2016)
Protein noise

Direct comparison with Taniguchi et al. (2010)

Simulations

Experiments

A more complex model is needed
Protein noise

Direct comparison with Taniguchi et al. (2010)

Simulations

Experiments

A more complex model is needed
Multi-protein model

Model with a sharing of common resources: RNA-polymerases and ribosomes:
Conclusions

In this work:

- A model with division and replication
- Analytical results for protein mean and variances
- On average, coherent with experiments

But it does not reproduce all of the protein variability.
Thank you for your attention

PhD work supervised by

- Vincent Fromion
- Philippe Robert
For each gene, Taniguchi et al. (2010) gives:

- empirical mean of mRNA concentration: $\mu_m$
- empirical mean of protein concentration: $\mu_p$
- mRNA lifetime $\sigma_1$
- gene position (from which $\tau_R$ can be deduced)
Main idea of the proof

Question: How many mRNAs $X_s$
  ▶ created since the birth of the cell
  ▶ still present at time $s$ (with time $s$ before replication)

Use of a Marked Poisson Point Process of intensity

$$\nu(dx, dy) = \lambda_1 dx \otimes \sigma_1 e^{-\sigma_1 y} dy.$$
Main idea of the proof

Question: How many mRNAs $X_s$
  ▶ created since the birth of the cell
  ▶ still present at time $s$ (with time $s$ before replication)

Use of a Marked Poisson Point Process of intensity

$$\nu(dx, dy) = \lambda_1 dx \otimes \sigma_1 e^{-\sigma_1 y} dy.$$
Main idea of the proof

Question: How many mRNAs $X_s$
  ▶ created since the birth of the cell
  ▶ still present at time $s$ (with time $s$ before replication)

Use of a Marked Poisson Point Process of intensity

$$\nu(dx, dy) = \lambda_1 dx \otimes \sigma e^{-\sigma y} dy.$$
Main idea of the proof

Question: How many mRNAs $X_s$

- created since the birth of the cell
- still present at time $s$ (with time $s$ before replication)

Use of a Marked Poisson Point Process of intensity

$$\nu(dx, dy) = \lambda_1 dx \otimes \sigma_1 e^{-\sigma_1 y} dy.$$