Modele stochastycznej ekspresji genów z losowymi burstami i deterministycznym rozpadem białek

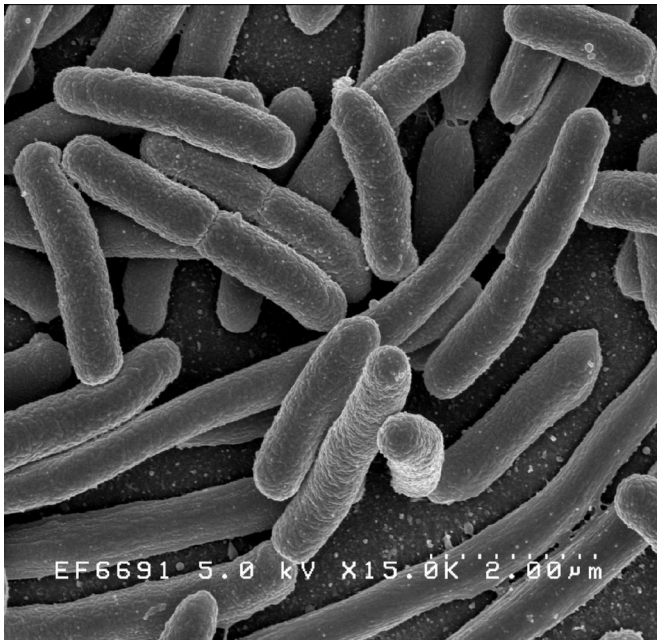
Anna Ochab-Marcinek,
Jakub Jędrak

Institute of Physical Chemistry
Polish Academy od Sciences

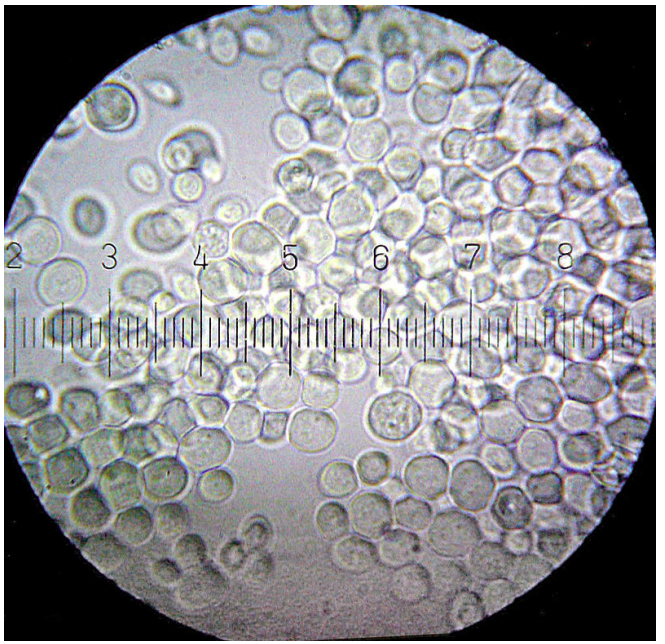
Living cells are systems of a very small volume.

http://bionumbers.hms.harvard.edu

- E. coli volume: ~ 1 μm^3 (Volume occupied by water: ~70%)
- Yeast volume: ~ 40 μm^3 (Volume occupied by water:~70%)
- Yeast nuclear volume: ~ 3 μm^3



E. coli cells, scale 2µm
Rocky Mountain Laboratories, NIAID, NIH - NIAID



Saccharomyces cerevisiae cells, numbered ticks are 11 µm apart
Bob Blaylock CC BY-SA 3.0

The abundance of particular types of proteins can be as low as hundreds, tens, or even single molecules.

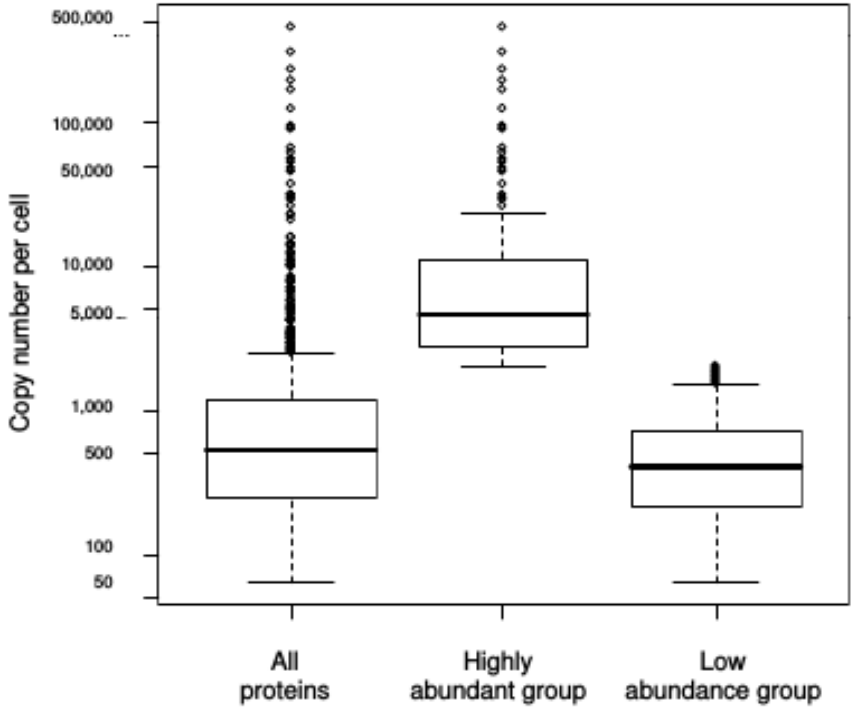


Figure 4
Abundance distribution of all identified proteins. Distributions are shown for the group of highly abundant proteins and the remaining low abundance protein group. Circles show distribution outliers as defined in *Methods*. The lower hinge represents the first quartile (25%) and the upper hinge the third quartile (75%). The high and low group were separated by clustering at a copy number cutoff of 2050 proteins per cell as described in *Methods*.

Ishihama et al. (2008). **Protein abundance profiling of the Escherichia coli cytosol**. BMC Genomics, 9, 102.

The abundance of particular types of proteins can be as low as hundreds, tens, or even single molecules.

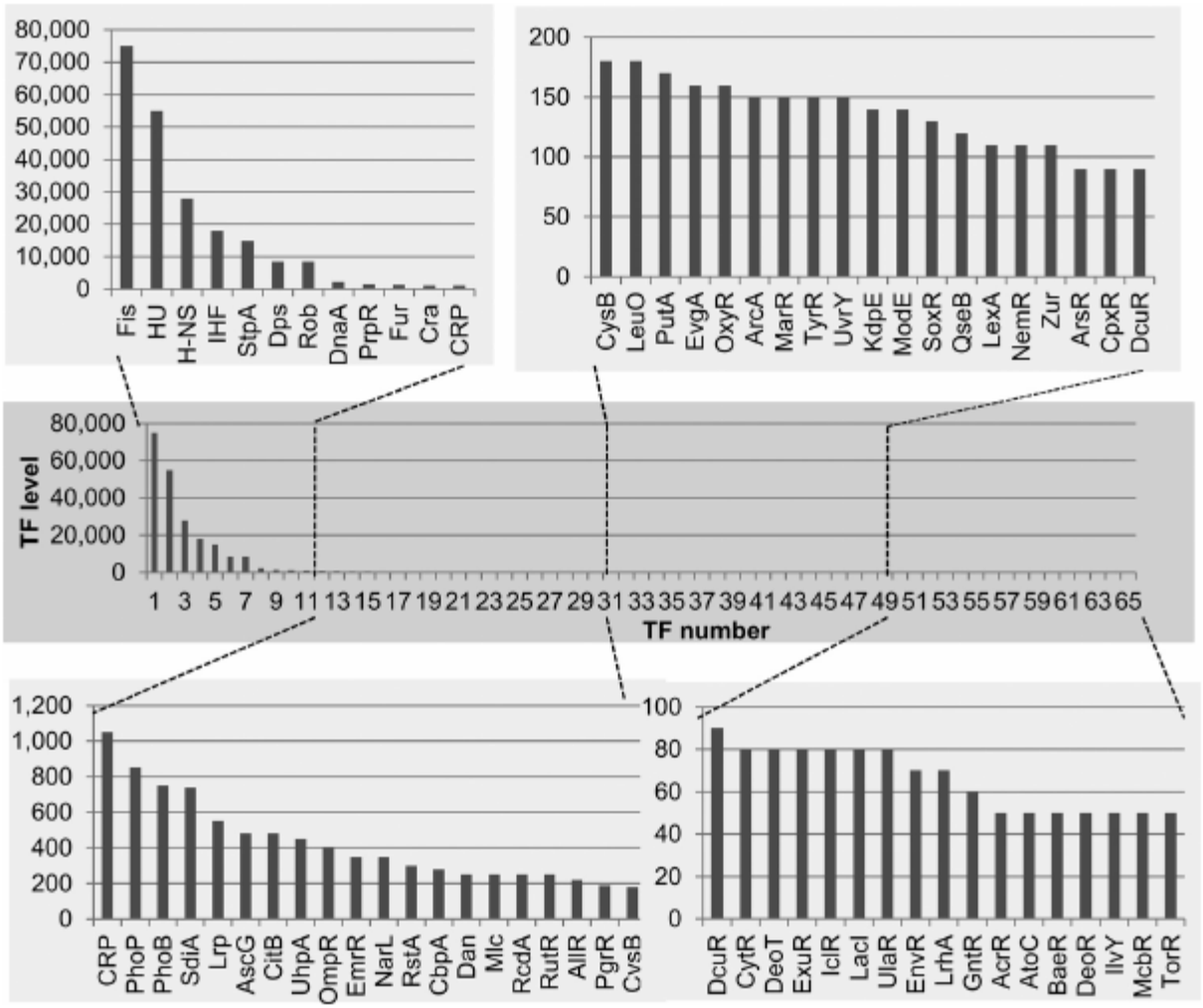
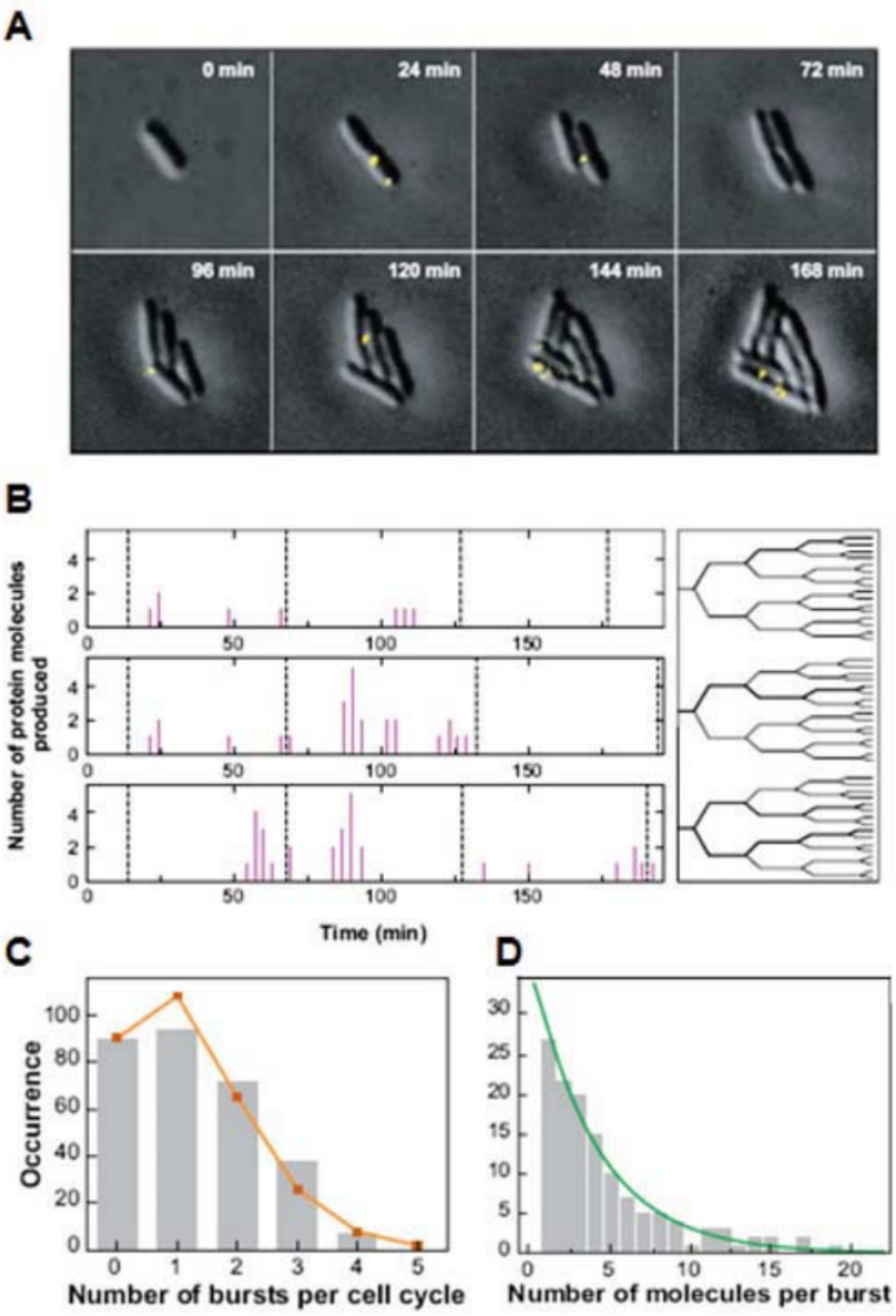


FIG 1 Intracellular concentration of TFs in exponential-phase cells of *E. coli* K-12 W3110. Cells were grown in LB-glucose medium at 37°C with shaking.

Ishihama et al. (2014). **Intracellular concentrations of 65 species of transcription factors with known regulatory functions in Escherichia coli.** Journal of Bacteriology, 196(15), 2718–2727.



- Proteins are produced in random bursts
- Cell division results in uneven distribution of inherited molecules in daughter cells

Xie, X. S. (2010). Enzymology and life at the single molecule level. Springer Series in Chemical Physics, 96, 435–448;
Yu, Ji, et al. Probing gene expression in live cells, one protein molecule at a time. Science 311.5767 (2006): 1600-1603.

Fig. 22.4. (**A**) Time-lapse movie of fluorescence images (*yellow*) overlaid with simultaneous DIC images (*gray*) of *E. coli* cells expressing a membrane protein fused with YFP under the repressed condition. Each yellow spot is due to one YFP generated by gene expression. (**B**) Time traces of the expression of YFP molecules (*left*) along three particular cell lineages (*right*). The vertical axis is the number of protein molecules newly synthesized during the last 3 min. The dotted lines mark the cell division times. Protein production occurs in stochastic bursts, each due to one copy of mRNA and generates variable numbers of YFP molecules. (**C**) Histogram of the number of expression bursts per cell cycle. The fit is a Poisson distribution of an average of 1.2 mRNA per cell cycle. (**D**) Distribution of the number of YFPs in each gene expression burst, which follows an exponential distribution with an average of four molecules per burst. From ref. [16]

Regulation of gene expression by effectors (here: corepressor)

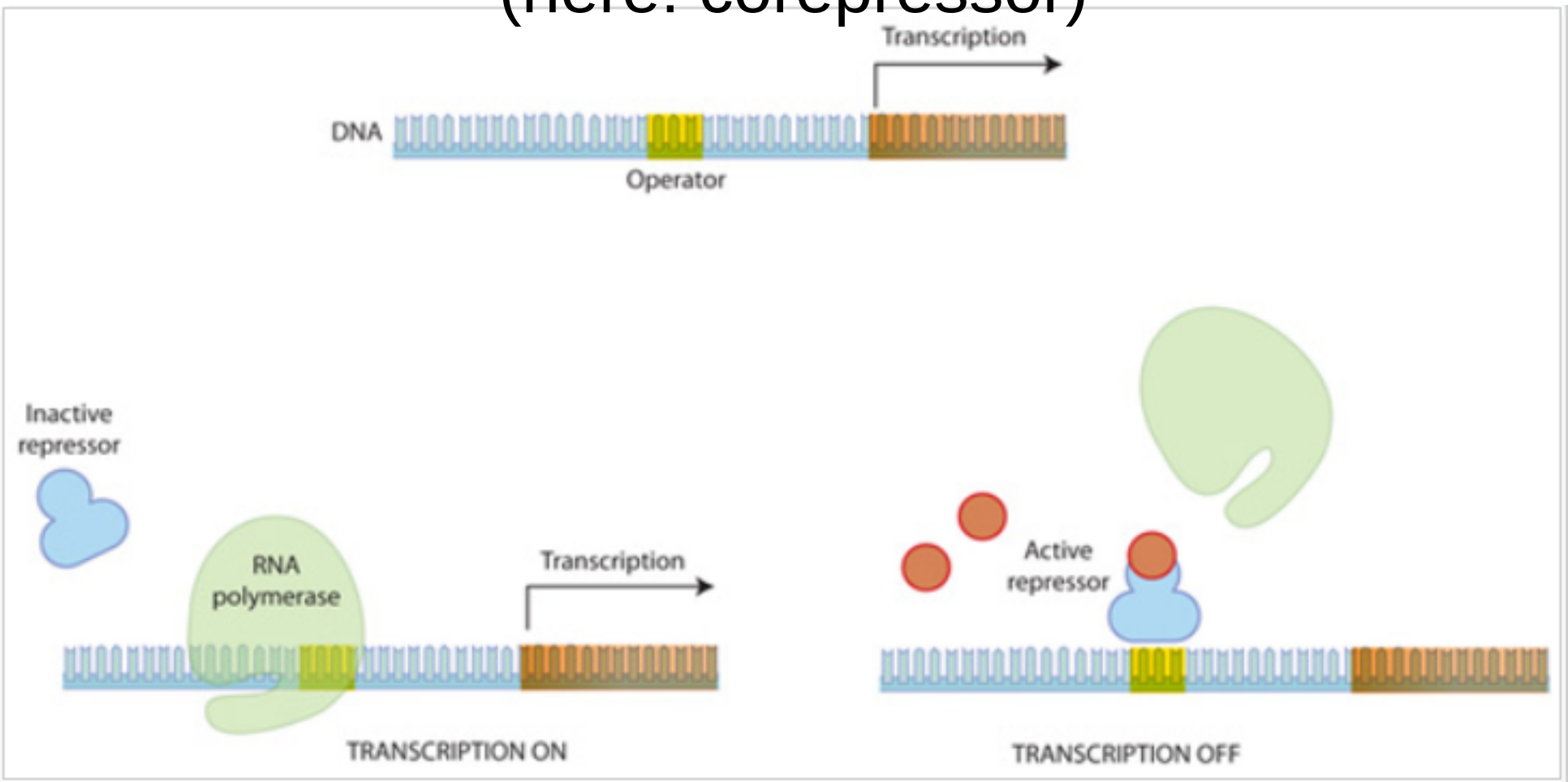


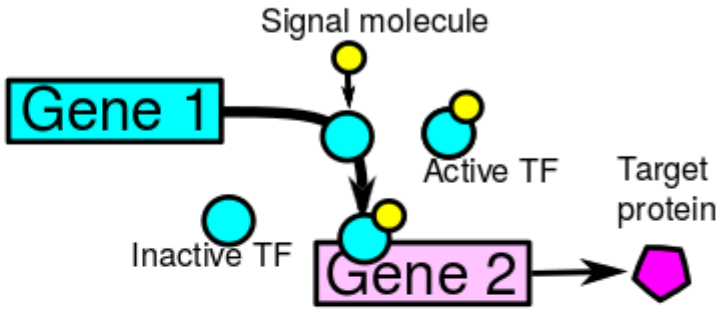
Figure 3: Transcription repression near the promoter region.
Molecules can interfere with RNA polymerase binding. An inactive repressor protein (blue) can become activated by another molecule (red circle). This active repressor can bind to a region near the promoter called an operator (yellow) and thus interfere with RNA polymerase binding to the promoter, effectively preventing transcription.

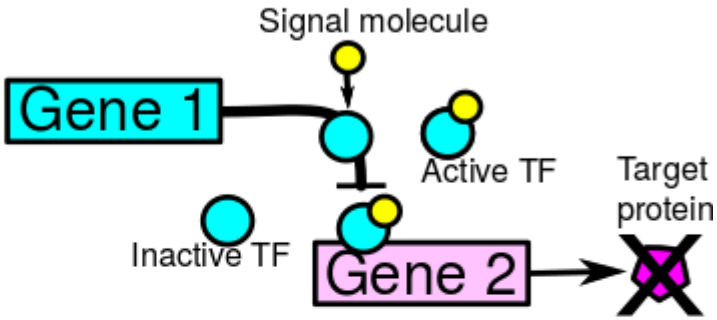
http://www.nature.com/scitable/nated/topicpage/gene-expression-14121669

Activation

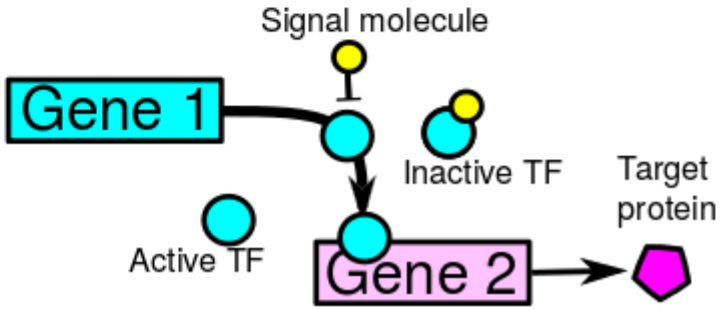
Repression

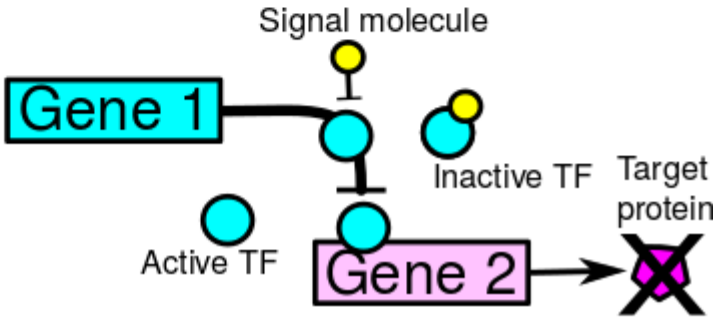
Positive



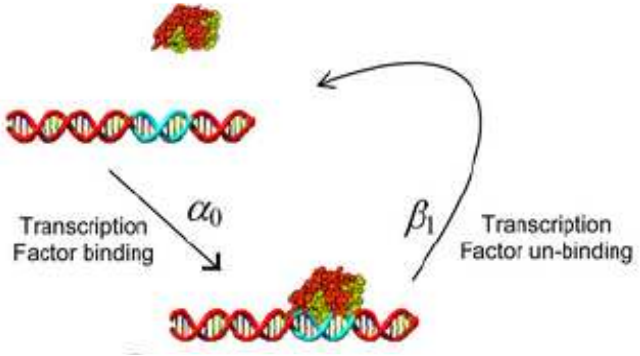


Negative





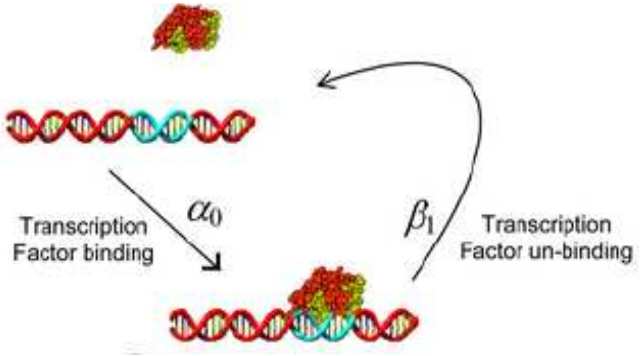
Sources of noise in gene expression



- Transcription factor binds to DNA

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

Sources of noise in gene expression

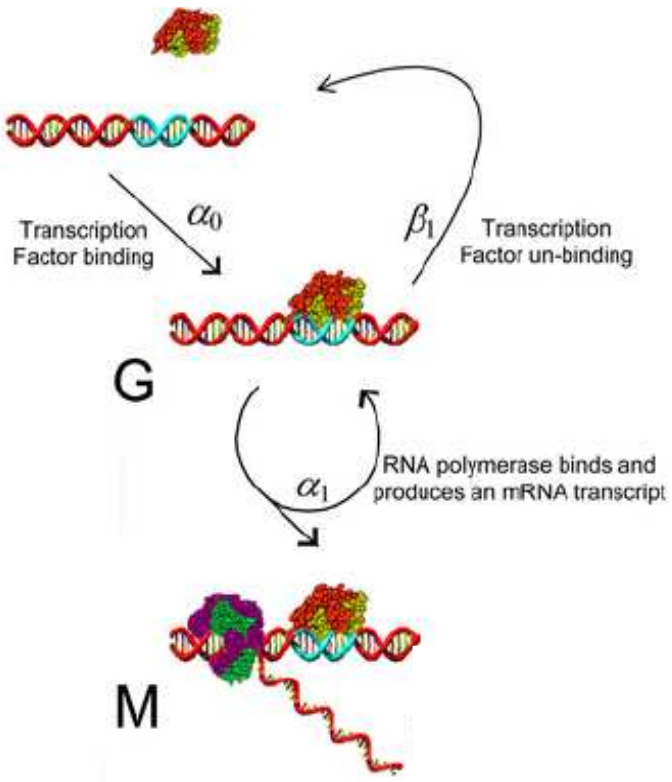


- Transcription factor binds to DNA

We assume fast TF binding/unbinding → Hill kinetics

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

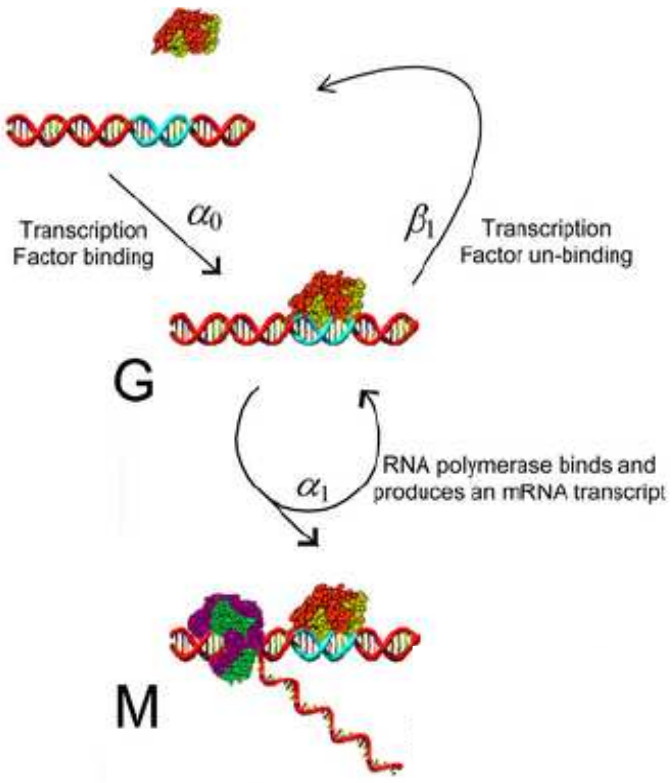
Sources of noise in gene expression



- Transcription factor binds to DNA
- Transcription:
  RNA polymerase reads DNA and produces mRNA

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

Sources of noise in gene expression

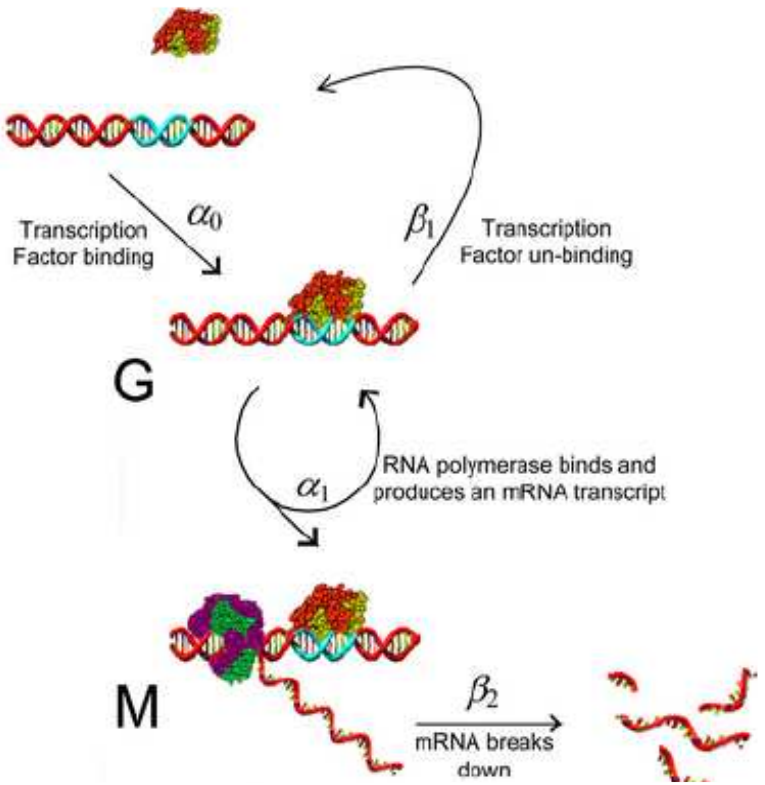


- Transcription factor binds to DNA
- Transcription:
  RNA polymerase reads DNA and produces mRNA

We assume that this step is lumped into the description of protein bursts

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

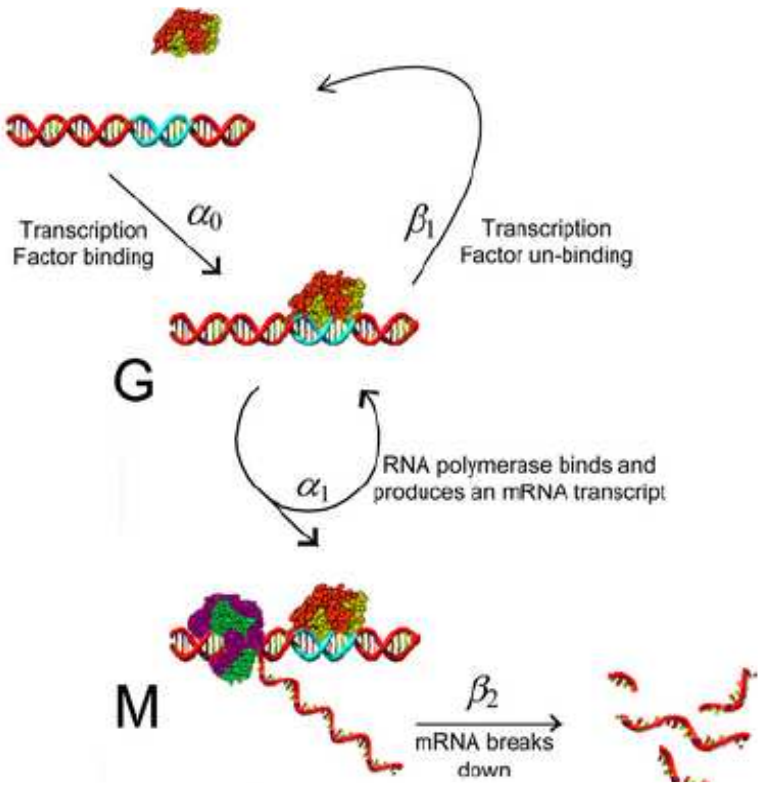
Sources of noise in gene expression



- Transcription factor binds to DNA
- Transcription:
  RNA polymerase reads DNA and produces mRNA
- RNA degradation

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

Sources of noise in gene expression

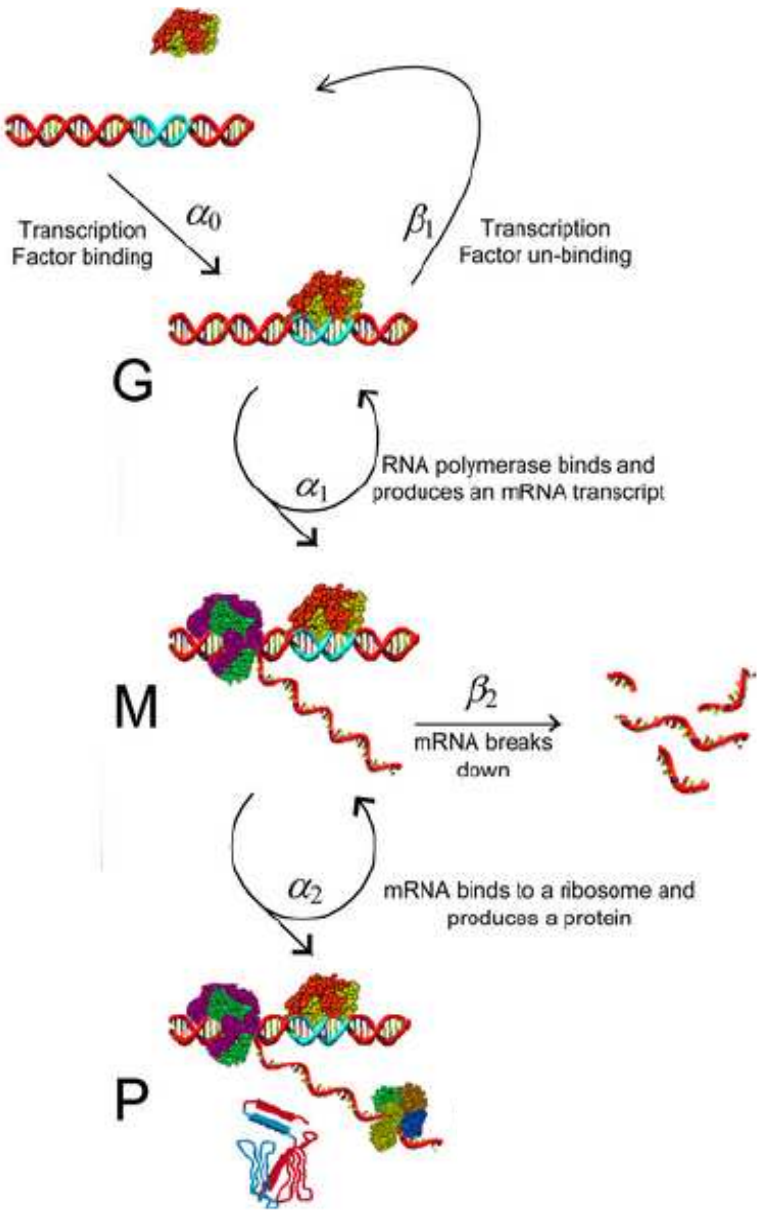


- Transcription factor binds to DNA
- Transcription:
  RNA polymerase reads DNA and produces mRNA
- RNA degradation

We assume that RNA degradation is fast, this step is also lumped into the description of protein bursts

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

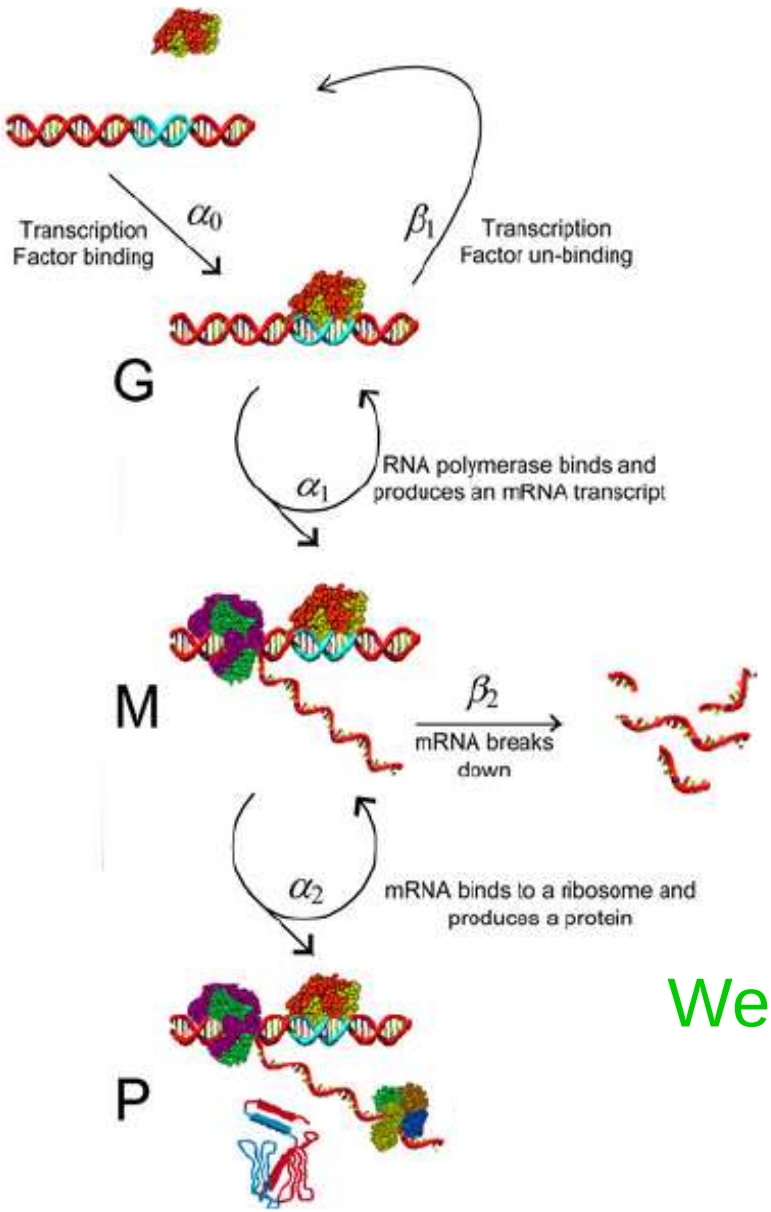
Sources of noise in gene expression



- Transcription factor binds to DNA
- Transcription:
  RNA polymerase reads DNA and produces mRNA
- RNA degradation
- Translation: Ribosome reads mRNA and produces protein

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

Sources of noise in gene expression

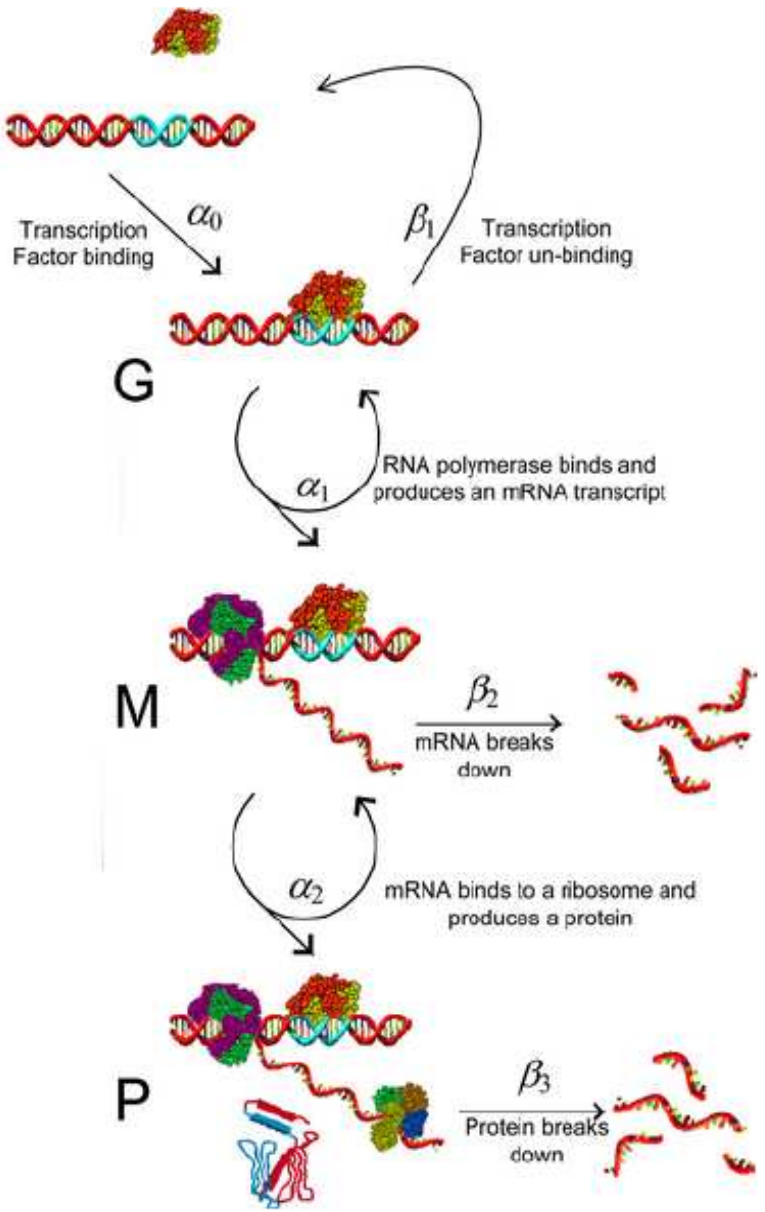


- Transcription factor binds to DNA
- Transcription:
  RNA polymerase reads DNA and produces mRNA
- RNA degradation
- Translation: Ribosome reads mRNA and produces protein

We assume that proteins are produced in exponentially-distributed bursts

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

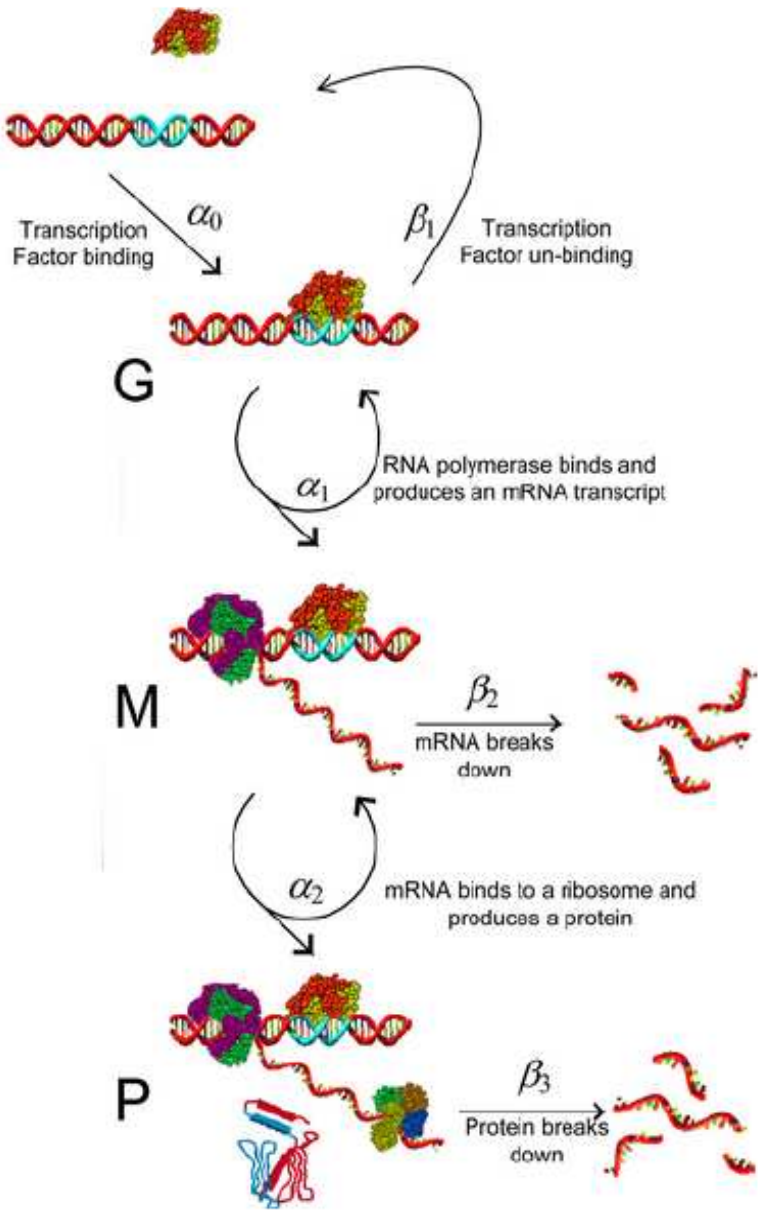
Sources of noise in gene expression



- Transcription factor binds to DNA
- Transcription:
  RNA polymerase reads DNA and produces mRNA
- RNA degradation
- Translation: Ribosome reads mRNA and produces protein
- Protein degradation

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

Sources of noise in gene expression



- Transcription factor binds to DNA
- Transcription:
  RNA polymerase reads DNA and produces mRNA
- RNA degradation
- Translation: Ribosome reads mRNA and produces protein
- Protein degradation

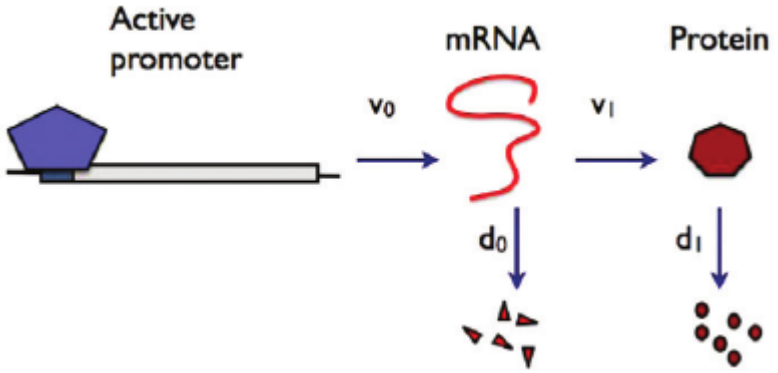
We assume a continuous degradation

P.J. Ingram et al. PloS Comput Biol 2005 e1000192

This is different than
the traditional modeling by
Master equations

Traditionally, gene expression has been modeled using Master equations.

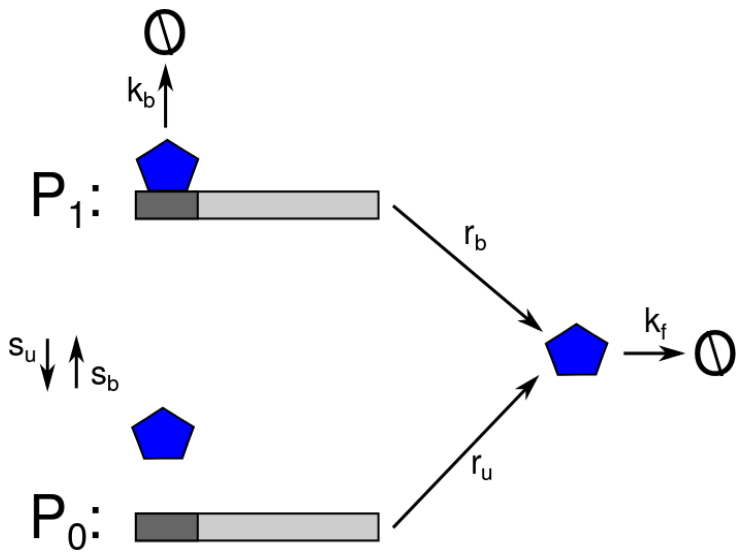
- Difficult; solutions to only a few simplest models.
- Analysis of the solutions is complicated: probability distributions are discrete.



$$\frac{\partial P_{m,n}}{\partial t} = v_0(P_{m-1,n} - P_{m,n}) + v_1 m(P_{m,n-1} - P_{m,n}) + d_0[(m+1)P_{m+1,n} - mP_{m,n}] + d_1[(n+1)P_{m,n+1} - nP_{m,n}]$$

$$\rightarrow \quad P_n = \frac{\Gamma(a+n)}{\Gamma(n+1)\Gamma(a)} \left(\frac{b}{1+b}\right)^n \left(1 - \frac{b}{1+b}\right)^a$$

Shahrezaei, V., & Swain, P. (2008). Analytical distributions for stochastic gene expression. PNAS 105(45), 17256–17261.



$$\frac{d}{dt}P_0(n,t) = r_u(P_0(n-1,t) - P_0(n,t))$$
$$+ k_f((n+1)P_0(n+1,t) - nP_0(n,t))$$
$$+ k_b P_1(n,t) + s_u P_1(n-1,t) - s_b n P_0(n,t),$$

$$\frac{d}{dt}P_1(n,t) = r_b(P_1(n-1,t) - P_1(n,t))$$
$$+ k_f((n+1)P_1(n+1,t) - nP_1(n,t))$$
$$- k_b P_1(n,t) - s_u P_1(n,t) + s_b(n+1)P_0(n+1,t).$$

→ Solution by generating functions

Grima, R., Schmidt, D. R., & Newman, T. J. (2012). Steady-state fluctuations of a genetic feedback loop: an exact solution. The Journal of Chemical Physics, 137(3), 035104.

Our approach:
Continuous variables
&
Hill kinetics

We use an alternative approach:

- Number of molecules assumed to be large enough: continuous variable.

We use an alternative approach:

- Number of molecules assumed to be large enough: continuous variable.

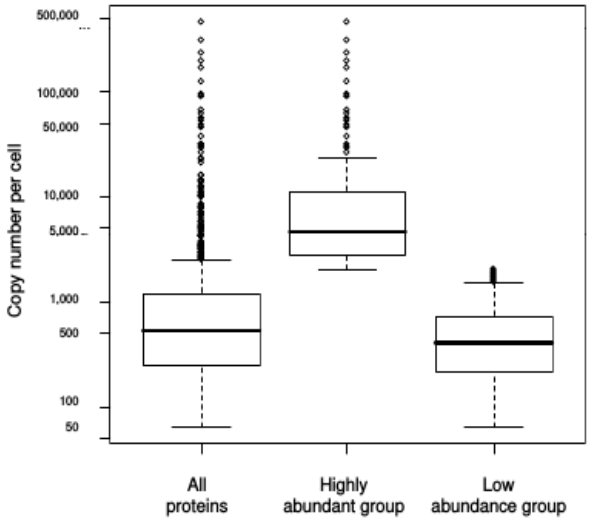


Figure 4
Abundance distribution of all identified proteins. Distributions are shown for the group of highly abundant proteins and the remaining low abundance protein group. Circles show distribution outliers as defined in *Methods*. The lower hinge represents the first quartile (25%) and the upper hinge the third quartile (75%). The high and low group were separated by clustering at a copy number cutoff of 2050 proteins per cell as described in *Methods*.

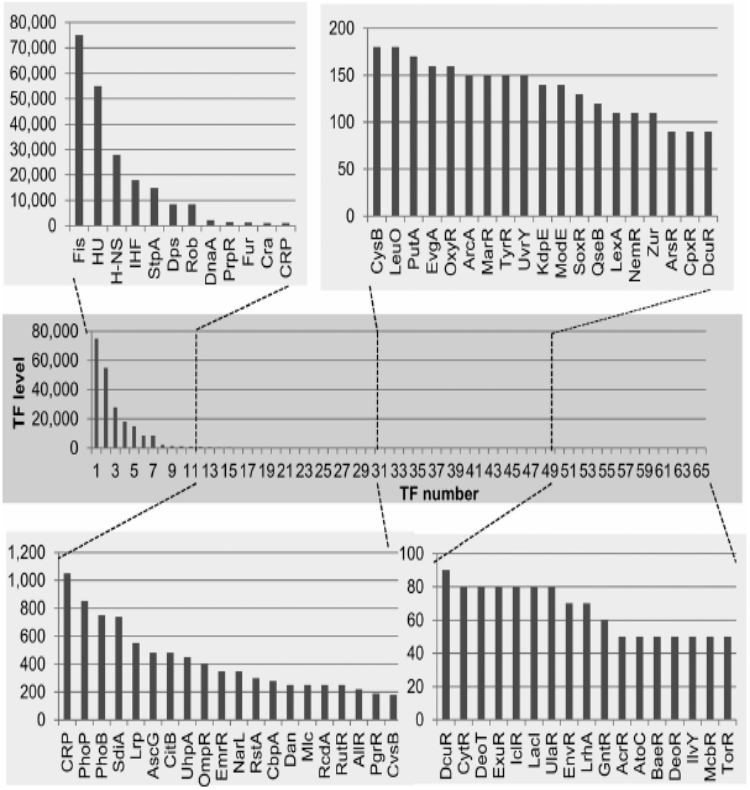


FIG 1 Intracellular concentration of TFs in exponential-phase cells of *E. coli* K-12 W3110. Cells were grown in LB-glucose medium at 37°C with shaking.

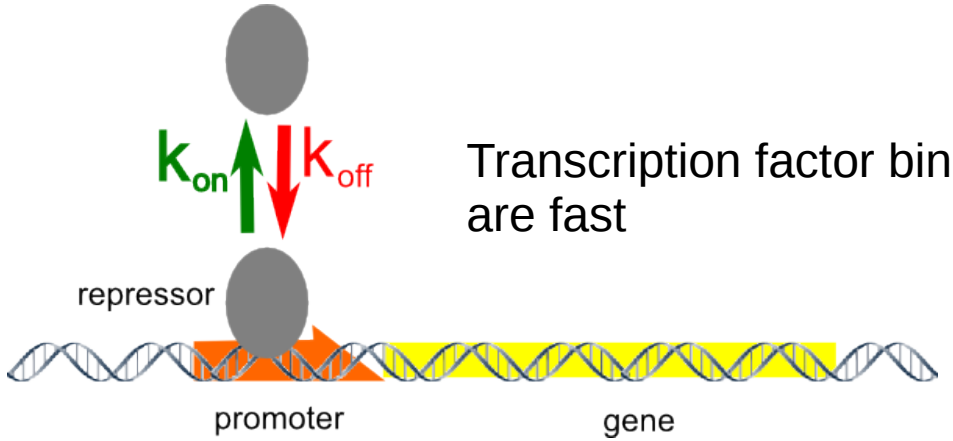
Ishihama et al. (2008). **Protein abundance profiling of the Escherichia coli cytosol**. BMC Genomics, 9, 102.

Ishihama et al. (2014). **Intracellular concentrations of 65 species of transcription factors with known regulatory functions in Escherichia coli.** Journal of Bacteriology, 196(15), 2718–2727.

We use an alternative approach:

- Number of molecules assumed to be large enough: continuous variable.
- Gene-transcription factor binding reactions can be accurately described by Hill kinetics.

Hill / Michaelis-Menten kinetics



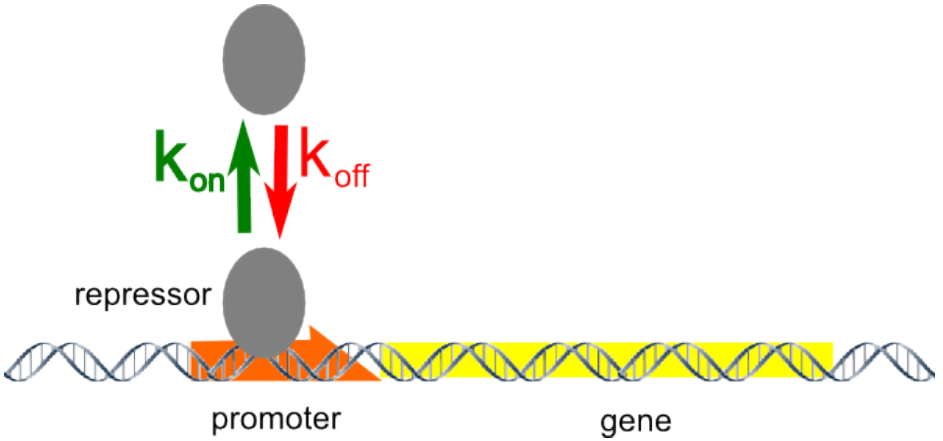
Transcription factor binding and unbinding are fast

Repression:
Transcription when the operator is free

$$\frac{dM}{dt} = k_m O - k_{dm}\ M$$

O: probability that the operator is free

Hill / Michaelis-Menten kinetics



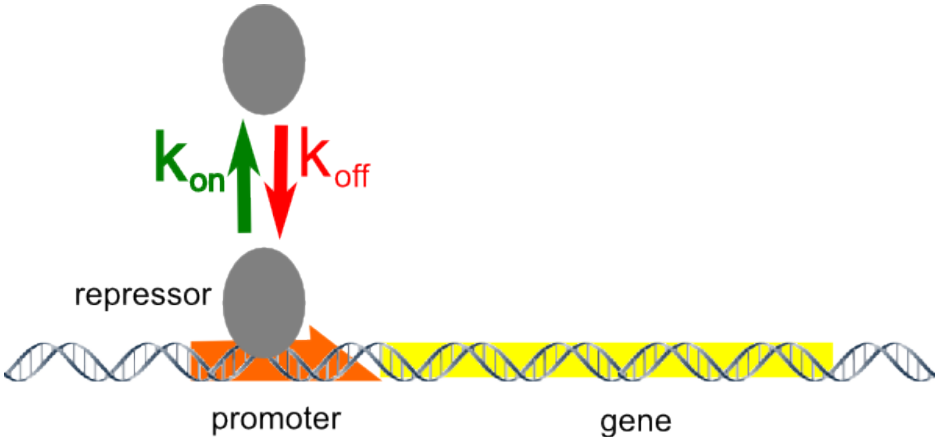
O: probability that the operator is free

O_R: probability that the operator is occupied

At steady state:

$$k_{on} \cdot R \cdot O = k_{off} \cdot O_R \quad \rightarrow \quad O_R = \frac{k_{on}}{k_{off}} R \cdot O$$

Hill / Michaelis-Menten kinetics



$$O + O_R = 1$$

$$O\left(1 + \frac{k_{on}}{k_{off}} R\right) = 1$$

Probability that the operator is free → transcription occurs

$$O \equiv h(R) = \frac{1}{1 + \frac{k_{on}}{k_{off}} R}$$

Hill kinetics, binding of n transcription factors at a time

- Detailed balance

$$k^1_{\text{on}} R \cdot O = k^1_{\text{off}} RO, \quad \ldots, \quad k^n_{\text{on}} R \cdot R_{n-1} O = k^n_{\text{off}} R_n O$$

- Probabilities sum up to 1

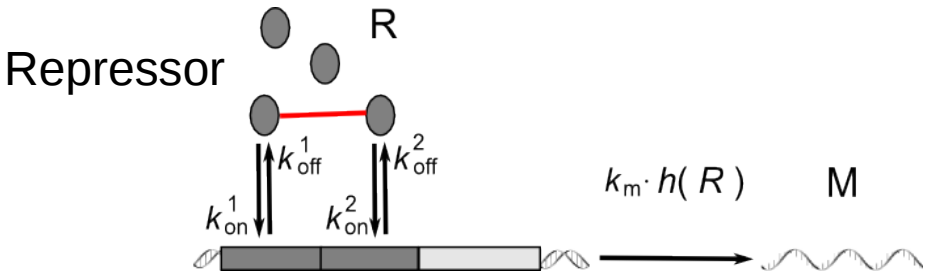
$$O + R \cdot O \frac{k^1_{\text{on}}}{k^1_{\text{off}}} + \ldots + R \cdot R_{n-1} O \frac{k^n_{\text{on}}}{k^n_{\text{off}}} = 1$$

- Full dose-response function

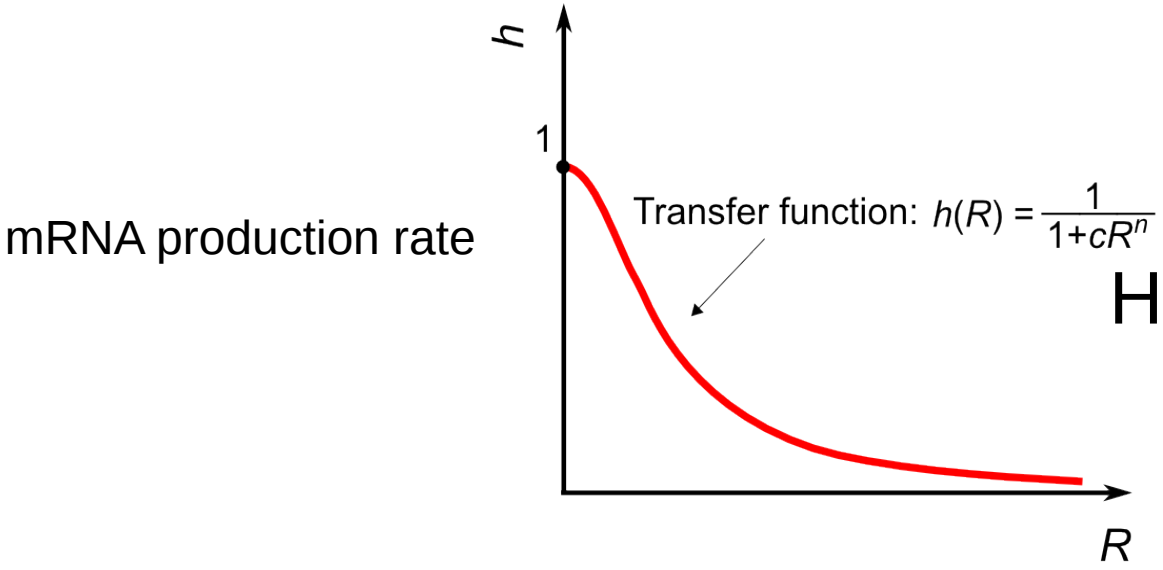
$$h(R) = \left(1 + \frac{k^1_{\text{on}}}{k^1_{\text{off}}} R + \cdots + \frac{k^1_{\text{on}} \cdot \ldots \cdot k^n_{\text{on}}}{k^1_{\text{off}} \cdot \ldots \cdot k^n_{\text{off}}} R^n \right)^{-1}$$

- At strong cooperativity, Hill function:

$$h(R) = \left(1 + \frac{k^1_{\text{on}} \cdot \ldots \cdot k^n_{\text{on}}}{k^1_{\text{off}} \cdot \ldots \cdot k^n_{\text{off}}} R^n \right)^{-1}$$



Cooperative repressor binding

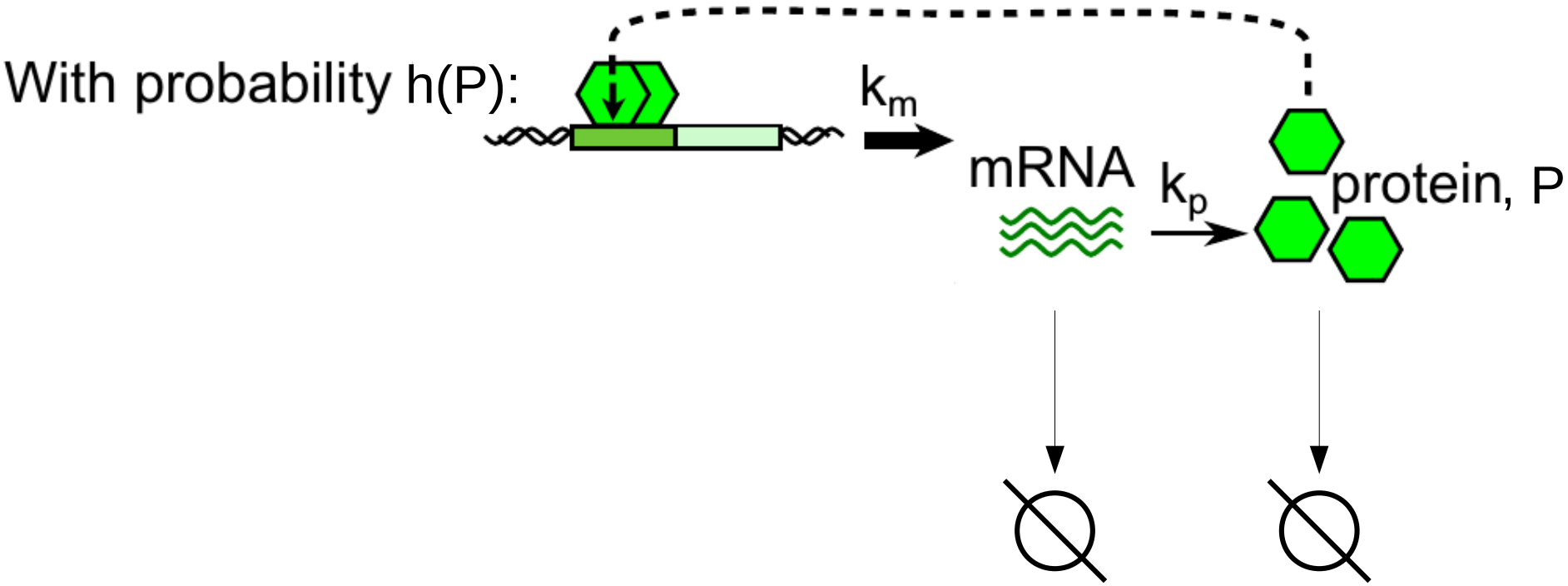


Hill function

Assumptions of Hill kinetics:

- Transcription factor binding/unbinding is fast compared to other time scales in the system
- Cooperativity is strong (only n transcription factors, bound simultaneously, give a non-negligible effect)

Friedman's model of a self-regulating gene



Friedman's model

$$\frac{\partial p(P)}{\partial t} = \frac{\partial}{\partial P}\left[k_{dp}\ P\ p(P)\right] + k_m \int_0^P dP'\ w(P, P')\ p(P')\, h(P')$$

- Hybrid model (Friedman et al., PRL 2006)
- Deterministic degradation
- Production in stochastic bursts
- Exponential distribution of burst sizes

Friedman's model

$$\frac{\partial p(P)}{\partial t} = \frac{\partial}{\partial P}\left[k_{dp}\ P\ p(P)\right] + k_m \int_0^P dP'\ w(P, P')\ p(P')\, h(P')$$

- $x \geq 0$ is a continuous variable
- u: burst size, $w(u) = \nu(u) - \delta(u)$
- Probability distribution of burst sizes

$$\nu(u) = (1/b)\exp(-u/b)$$

Friedman's model

$$\frac{\partial p(P)}{\partial t} = \frac{\partial}{\partial P}\left[k_{dp}\ P\ p(P)\right] + k_m \int_0^P dP'\ w(P, P')\ p(P')\, h(P')$$

- $x \geq 0$ is a continuous variable
- u: burst size, $w(u) = \nu(u) - \delta(u)$
- Probability distribution of burst sizes

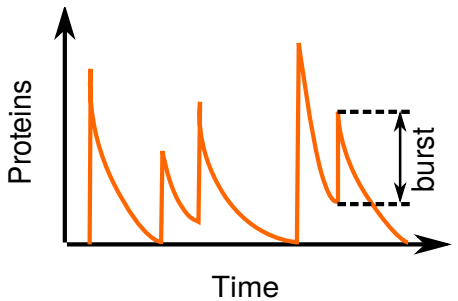
Regulatory function (Hill function)

$$\nu(u) = (1/b)\exp(-u/b)$$

Friedman's model

Steady state solution (Friedman, PRL 2006):

$$p(P) = A\,P^{-1} e^{-P/\beta} e^{\alpha \int dP\, h(P)/P}$$

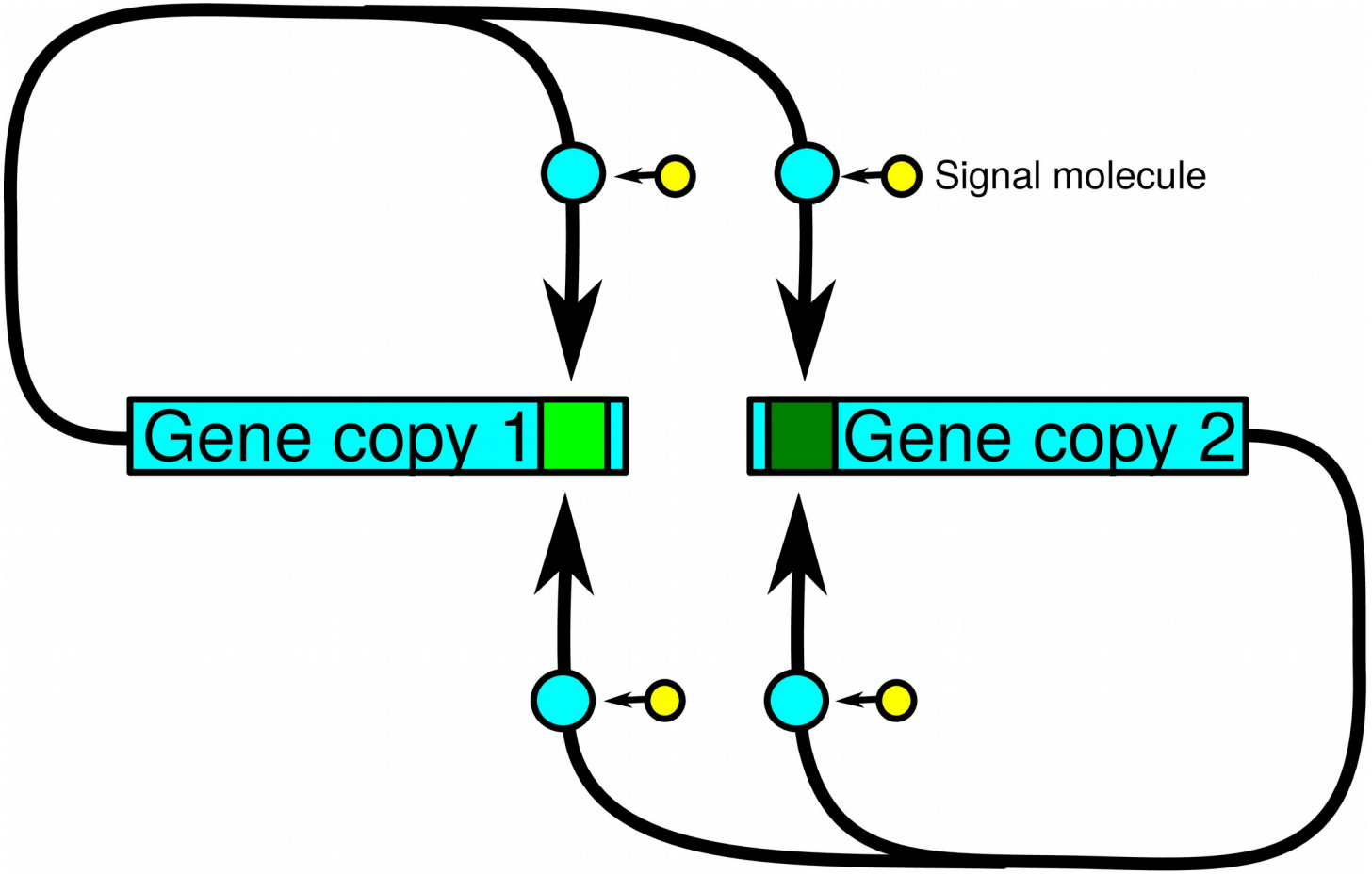


Mean burst frequency at h(P)=1: $\alpha = k_m/k_{dp}$

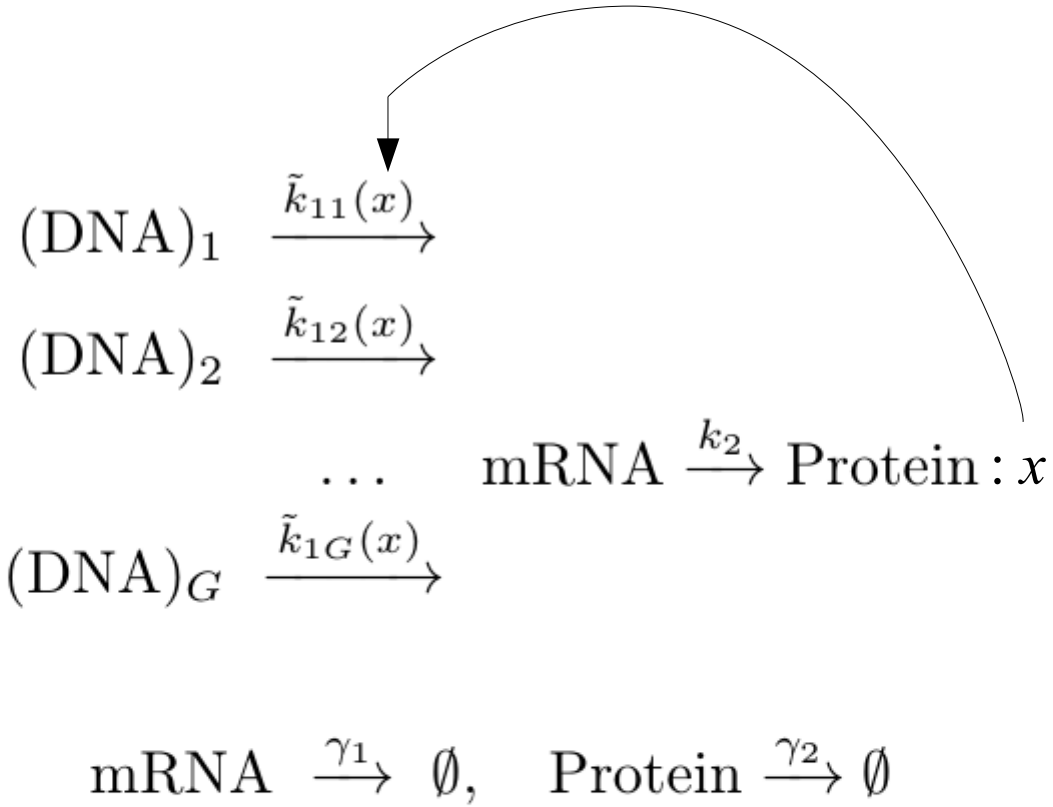
Mean burst size: β

Our analysis of Friedman's model

Multiple copies of a self-regulated gene



J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).



G gene copies.

Transcription rates depend on protein concentration x.

Effective rate constants: $\tilde{k}_{1j}(x) = k_{1j} h_j(x)$.

Transcription factor binding:

$$_{\mathrm{j}}\mathrm{O} + \mathrm{nX} \overset{K_j}{\rightleftharpoons} {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

mRNA synthesis/degradation:

Repressor

$$_{\mathrm{j}}\mathrm{O} \xrightarrow{k_{1\mathrm{j}}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{O}$$

$$_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}} \xrightarrow{k_{1\mathrm{j}\epsilon}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

Activator

$$_{\mathrm{j}}\mathrm{O} \xrightarrow{k_{1\mathrm{j}\epsilon}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{O}$$

$$_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}} \xrightarrow{k_{1\mathrm{j}}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

$$\mathrm{Y} \xrightarrow{\gamma_1} \varnothing$$

Transcription factor synthesis/degradation:

$$\mathrm{Y} \xrightarrow{k_2} \mathrm{X} + \mathrm{Y}$$

$$\mathrm{X} \xrightarrow{\gamma_2} \varnothing$$

TABLE I: Kinetic scheme. X: protein, Y: mRNA, k_{1j}: rate of mRNA synthesis from the operator of the j-th gene copy in the active state, $k_{1j\epsilon}$: rate of mRNA synthesis from the operator of the j-th gene copy in the inactive state (leakage), γ_1: rate of mRNA degradation, k_2: rate of protein synthesis, γ_2: rate of protein degradation.

Transcription factor binding:

$$_{\mathrm{j}}\mathrm{O} + \mathrm{nX} \overset{K_j}{\rightleftharpoons} {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

mRNA synthesis/degradation:

Repressor

$$_{\mathrm{j}}\mathrm{O} \xrightarrow{k_{1\mathrm{j}}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{O}$$

$$_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}} \xrightarrow{k_{1\mathrm{j}\epsilon}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

Activator

$$_{\mathrm{j}}\mathrm{O} \xrightarrow{k_{1\mathrm{j}\epsilon}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{O}$$

$$_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}} \xrightarrow{k_{1\mathrm{j}}} \mathrm{Y} + {}_{\mathrm{j}}\mathrm{OX}_{\mathrm{n}}$$

$$\mathrm{Y} \xrightarrow{\gamma_1} \varnothing$$

Transcription factor synthesis/degradation:

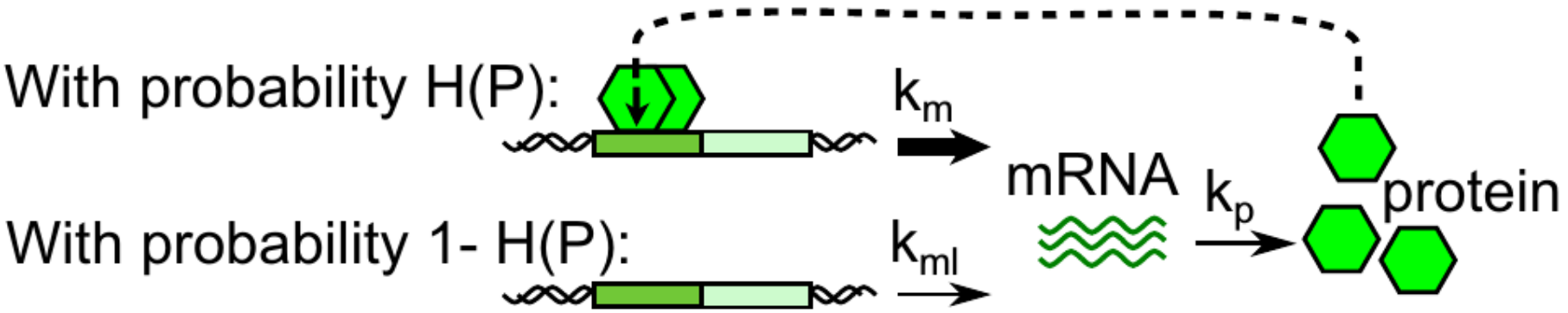
$$\mathrm{Y} \xrightarrow{k_2} \mathrm{X} + \mathrm{Y}$$

$$\mathrm{X} \xrightarrow{\gamma_2} \varnothing$$

Leaky transcription

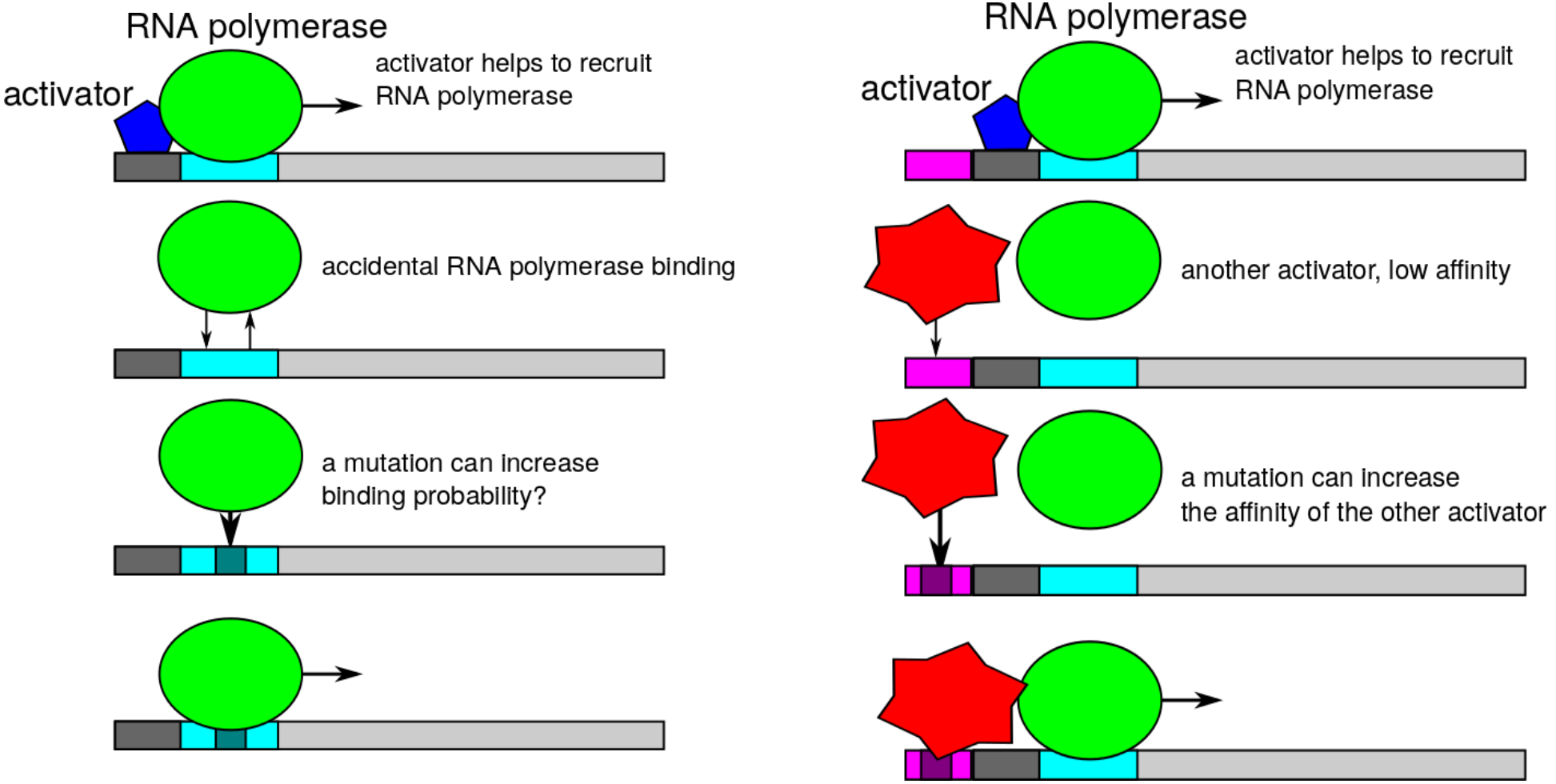
TABLE I: Kinetic scheme. X: protein, Y: mRNA, k_{1j}: rate of mRNA synthesis from the operator of the j-th gene copy in the active state, $k_{1j\epsilon}$: rate of mRNA synthesis from the operator of the j-th gene copy in the inactive state (leakage), γ_1: rate of mRNA degradation, k_2: rate of protein synthesis, γ_2: rate of protein degradation.

Transcriptional leakage (basal expression)



No tight control over the promoter.

- Some level of transcription is maintained even when the promoter is in the *off* state.


RNA polymerase
activator
activator helps to recruit
RNA polymerase
accidental RNA polymerase binding
a mutation can increase
binding probability?
RNA polymerase
activator
activator helps to recruit
RNA polymerase
another activator, low affinity
a mutation can increase
the affinity of the other activator

- Transfer function:

$$h_j(x) = (1 - \epsilon_j)H_j(x) + \epsilon_j, \qquad j = 1, 2, \ldots, G$$

- Transcriptional leakage: $\epsilon_j = k_{1j\epsilon}/k_{1j}$

- Hill kinetics:

$$H_j(x) \;=\; \left[1 + \left(\frac{x}{K_j}\right)^{n_j}\right]^{-1}$$

- Transfer function:

$$h_j(x) = (1 - \epsilon_j)H_j(x) + \epsilon_j, \qquad j = 1, 2, \ldots, G$$

$k_{1j}\, h_j(x)$ = transcription rate dependent on protein concentration x

- Transcriptional leakage: $\epsilon_j = k_{1j\epsilon}/k_{1j}$

- Hill kinetics:

$$H_j(x) = \left[1 + \left(\frac{x}{K_j}\right)^{n_j}\right]^{-1}$$

- Transfer function:

$$h_j(x) = (1 - \epsilon_j)H_j(x) + \epsilon_j, \qquad j = 1, 2, \ldots, G$$

- Transcriptional leakage: $\epsilon_j = k_{1j\epsilon}/k_{1j}$

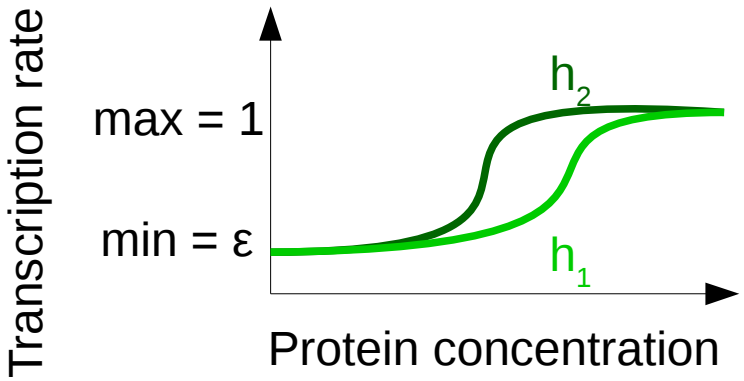
leaky transcription rate / max. regular transcription rate

- Hill kinetics:

$$H_j(x) = \left[1 + \left(\frac{x}{K_j}\right)^{n_j}\right]^{-1}$$

- Transfer function:

$$h_j(x) = (1 - \epsilon_j)H_j(x) + \epsilon_j,$$



- Transcriptional leakage: $\epsilon_j = k_{1j\epsilon}/k_{1j}$

- Hill kinetics:

$$H_j(x) = \left[1 + \left(\frac{x}{K_j}\right)^{n_j}\right]^{-1}$$

$$\frac{\partial p(x,t)}{\partial t} = \gamma_2 \sum_{j=1}^{G} a_j \int_0^x w(x-x') h_j(x') p(x',t) dx' + \gamma_2 \frac{\partial}{\partial x} \left[x p(x,t) \right]$$

$$a_j \equiv \frac{k_{1j}}{\gamma_2}, \qquad b \equiv \frac{k_2}{\gamma_1}$$

- $x \geq 0$ is a continuous variable
- u: burst size, $w(u) = \nu(u) - \delta(u)$ [1]
- Probability distribution of burst sizes (identical for each gene copy):

$$\nu(u) = (1/b) \exp(-u/b)$$

[1] N. Friedman, L. Cai, and X. S. Xie, Phys. Rev. Lett. **97**, 168302 (2006).

$$p(x) = Ax^{-1}e^{-x/b}\prod_{j=1}^{G}\exp\left[a_j\int\frac{h_j(x)}{x}dx\right]$$

$$p(x) = Ax^{-1}e^{-x/b}\prod_{j=1}^{G}x^{a_j}H_j(x)^{\frac{a_j(1-\epsilon_i)}{n_j}}$$

- Bursting of each gene copy is a Poisson process, independent from the bursting of all other copies.
- Thus, their protein production rates are coupled only by the common pool of proteins that regulate the genes as their TFs.

Identical gene copies

- $h_j(x) = h(x),\ a_j = a$
- The maximum burst frequency scales linearly with gene copy number

$$a \to Ga$$

$$p_G(x) \quad = \quad Ax^{Ga-1}e^{-x/b}\left[1+\left(\frac{x}{K}\right)^n\right]^{-\frac{Ga(1-\epsilon)}{n}}$$

Result 1

Ambiguity of the measures of noise

Two standard quantitative measures of gene expression noise are used interchangeably in literature:

- Fano factor $F = \sigma^2/\mu_1$
- Coefficient of variation $\eta = \sigma/\mu_1$

Non-regulated gene – Volfson et al. [2]:

$$\langle x \rangle \sim G^1, \quad \eta \sim G^{-\frac{1}{2}}, \quad F \sim G^0$$

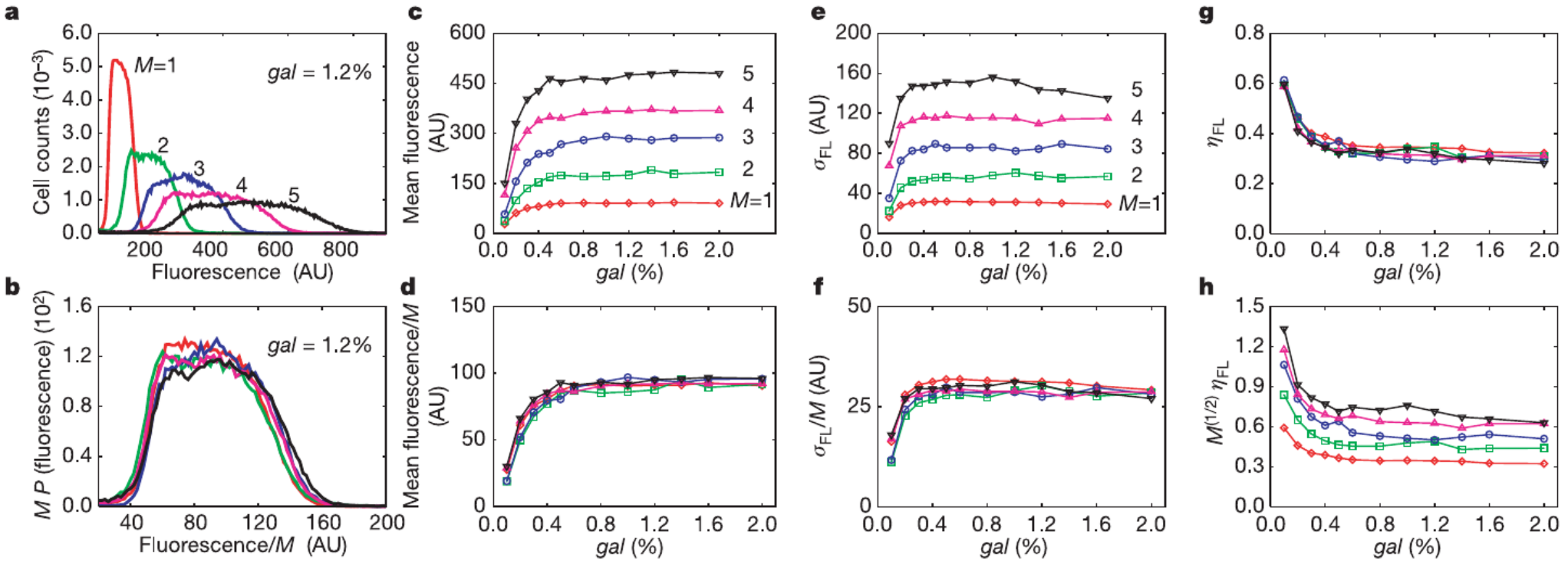
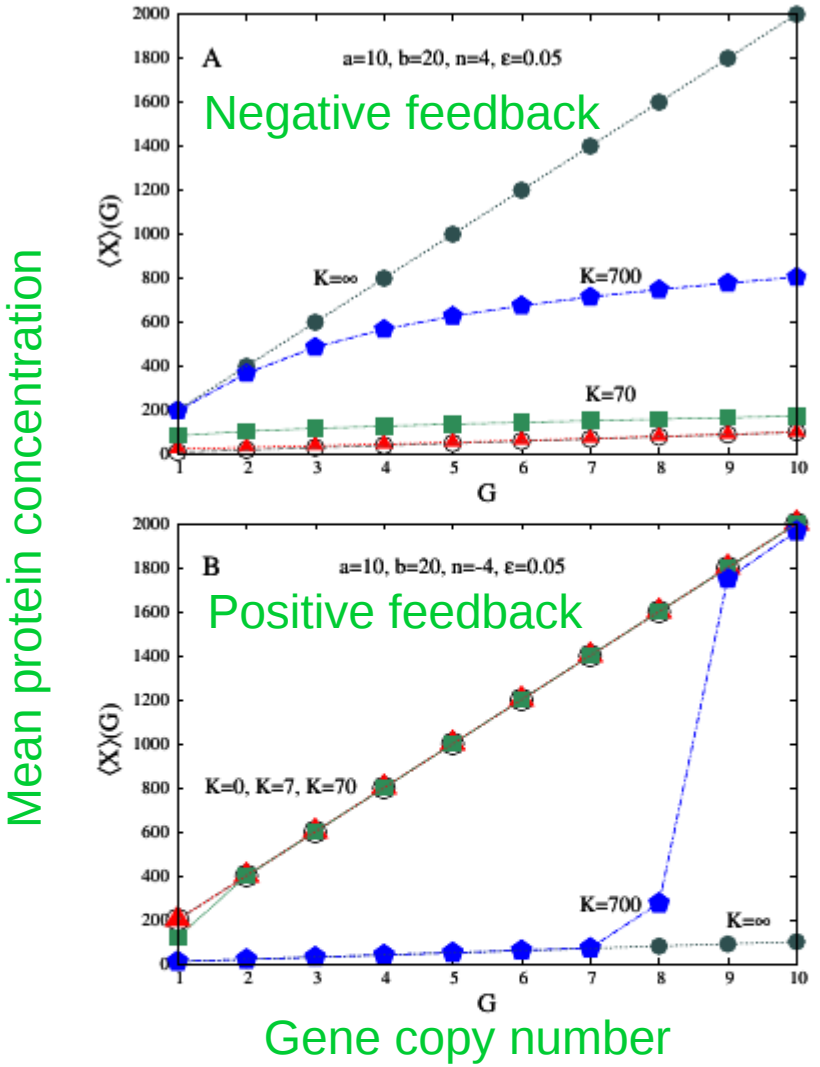


Figure 1 | Experimental results for GFP expression for different copy numbers and galactose concentrations. **a**, Histogram of GFP measurements for copy numbers from $M = 1$ to $M = 5$ above the saturation ($gal = 1.2\%$). **b**, The collapse of GFP distributions under the transformation $F \rightarrow F/M$, $P(F) \rightarrow MP(F/M)$ implies an extrinsic source of variability. AU, arbitrary units. **c**, Induction curves for copy numbers from $M = 1$ to $M = 5$. **d**, Collapse of the induction curves implies that transcription from each promoter is independent. **e**, Standard deviations of GFP corresponding to induction curves. **f**, The collapse of the standard deviation implies an extrinsic source of variability. **g**, The collapse of the coefficient of variation for different copy number implies an extrinsic source of variability. **h**, Lack of collapse implies that the variability is not of intrinsic origin.

[2] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty, Nature **439**, 861 (2006).

- Self-regulating gene:

  Mean does not depend linearly on the number of gene copies G.

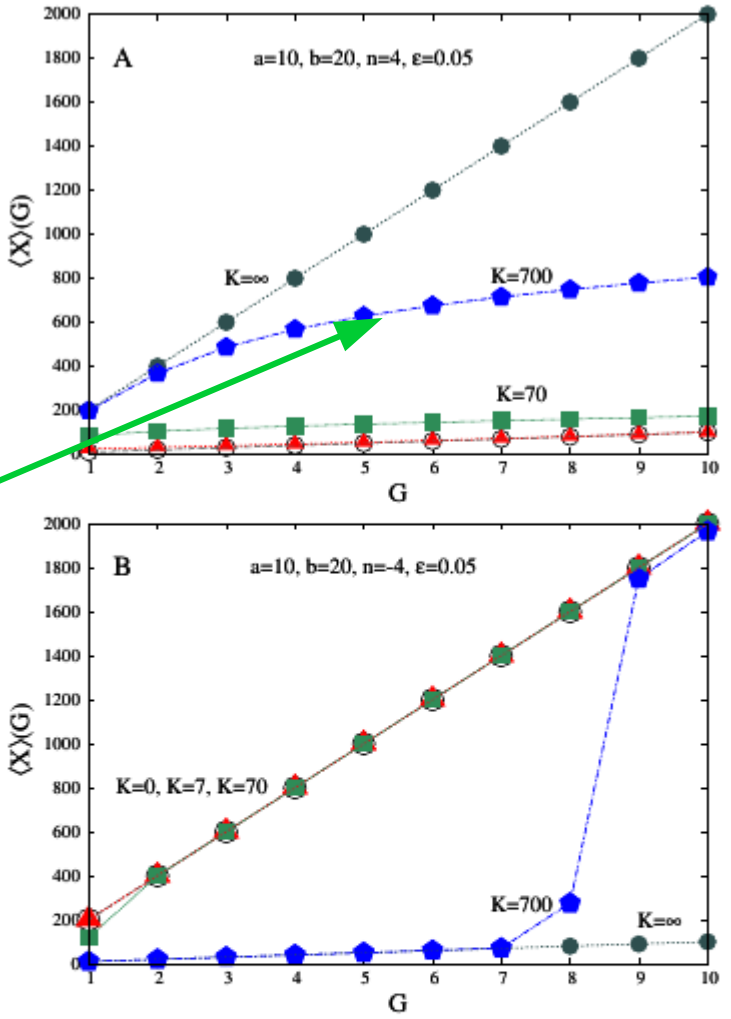


Average protein number may depend on gene copy number in a nonlinear manner in self-regulating genes. A: Negative auto-regulation ($n = 4$). B: Positive auto-regulation ($n = -4$). The abrupt increase for $K = 700$ and $G = 8$ is due to the transition of the protein number distribution through bimodality Feedback strength parameter $K = 0$ (empty circles), $K = 7$ (triangles), $K = 70$ (squares), $K = 700$ (pentagons), and $K = \infty$ (full circles). Maximum mean burst frequency $a = 10$. Mean burst size $b = 20$. Leakage $\epsilon = 0.05$. Lines provide guide for the eye only.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

- Self-regulating gene:

  Mean does not depend linearly on the number of gene copies G.

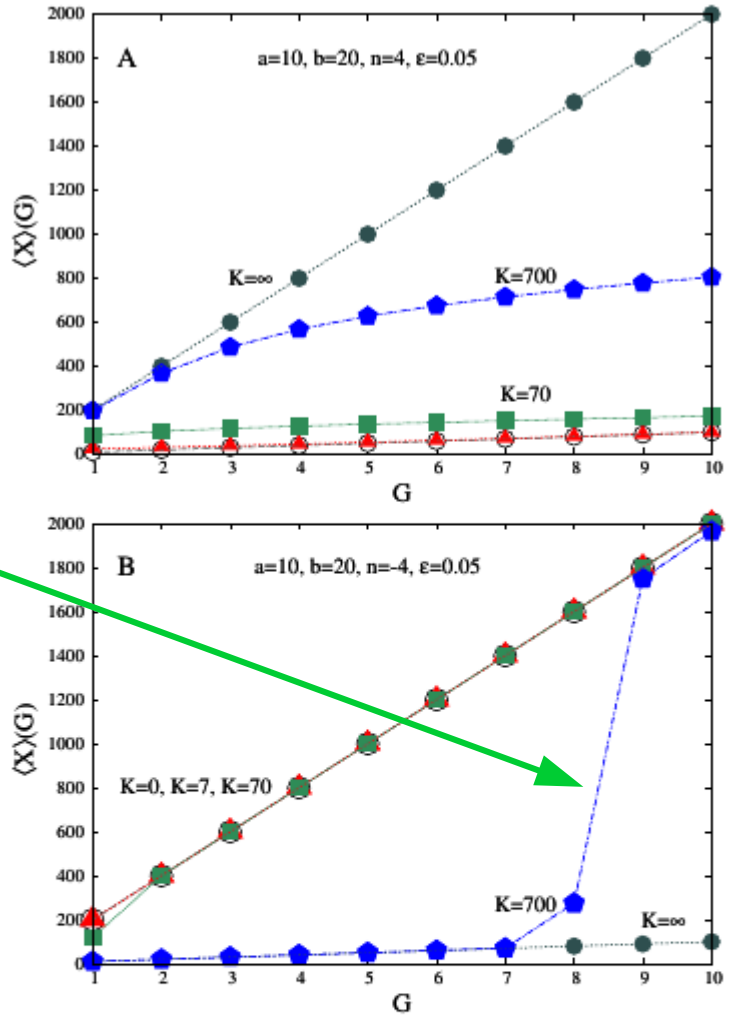


Average protein number may depend on gene copy number in a nonlinear manner in self-regulating genes. A: Negative auto-regulation ($n = 4$). B: Positive auto-regulation ($n = -4$). The abrupt increase for $K = 700$ and $G = 8$ is due to the transition of the protein number distribution through bimodality Feedback strength parameter $K = 0$ (empty circles), $K = 7$ (triangles), $K = 70$ (squares), $K = 700$ (pentagons), and $K = \infty$ (full circles). Maximum mean burst frequency $a = 10$. Mean burst size $b = 20$. Leakage $\epsilon = 0.05$. Lines provide guide for the eye only.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

- Self-regulating gene:

  Mean does not depend linearly on the number of gene copies G.



Average protein number may depend on gene copy number in a nonlinear manner in self-regulating genes. A: Negative auto-regulation ($n = 4$). B: Positive auto-regulation ($n = -4$). The abrupt increase for $K = 700$ and $G = 8$ is due to the transition of the protein number distribution through bimodality Feedback strength parameter $K = 0$ (empty circles), $K = 7$ (triangles), $K = 70$ (squares), $K = 700$ (pentagons), and $K = \infty$ (full circles). Maximum mean burst frequency $a = 10$. Mean burst size $b = 20$. Leakage $\epsilon = 0.05$. Lines provide guide for the eye only.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

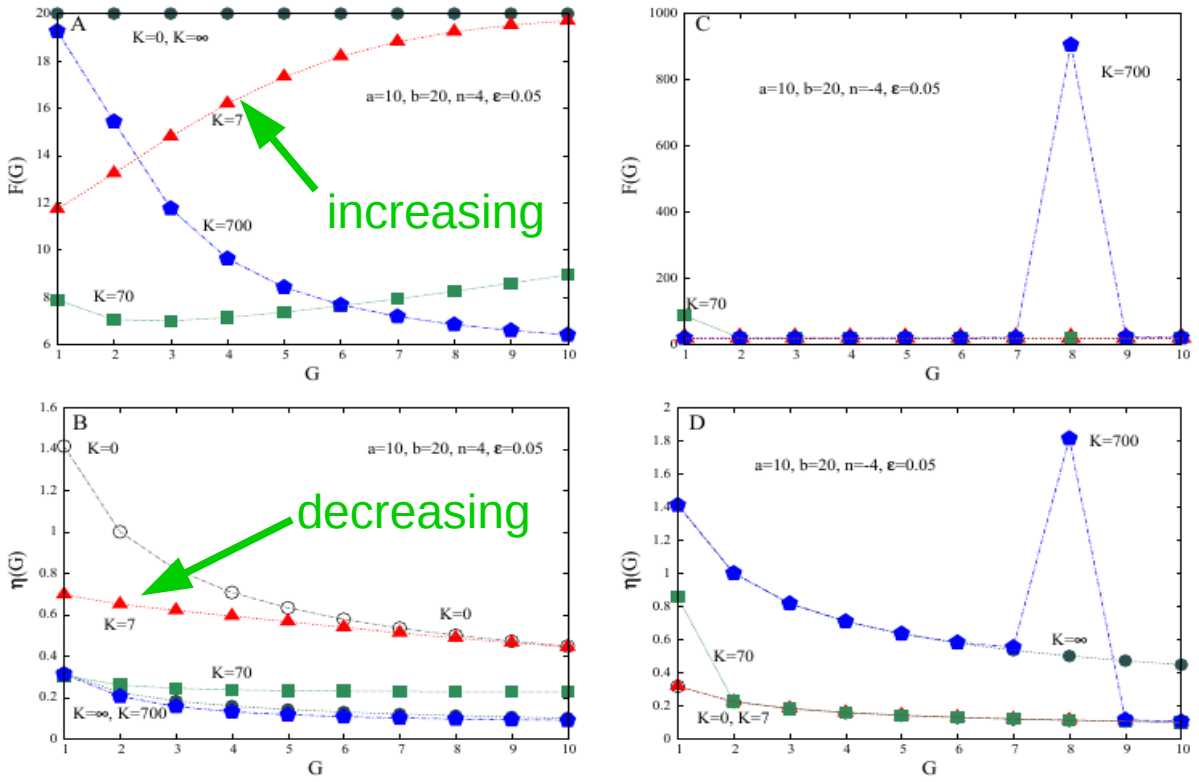
Self-regulating gene: Fano Factor and CV vary in a different manner as G is varied.

Negative feedback

Positive feedback

Fano factor, variance/mean

Coefficient of variation, variance/mean^2



In self-regulating genes, Fano factor and by coefficient of variation may depend on gene copy number in a qualitatively different manner. A, B: Negative auto-regulation, $n = 4$. A: Depending on the feedback strength parameter K, Fano factor $F = \sigma^2/\langle x \rangle$ may both decrease, increase or vary in a non-monotonous manner as gene copy number G is varied. B: Coefficient of variation $\eta = \sigma/\langle x \rangle$ is a monotonically decreasing function of gene copy number G. C, D: Positive auto-regulation, $n = -4$. Here, for $K = 700$, Fano factor $F(G)$ has just one maximum (C), whereas the coefficient of variation $\eta(G)$ has two clear maxima (D). The sharp maximum for $K = 700$ and $G = 8$ is due to the transition of the protein number distribution through bimodality. In absence of gene regulation ($K = 0$ and $K = \infty$), $F = b$ and $\eta \sim G^{-1/2}$. For negatively self-regulating genes, $F(G) < b$ and for positive auto-regulation, $F(G) > b$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

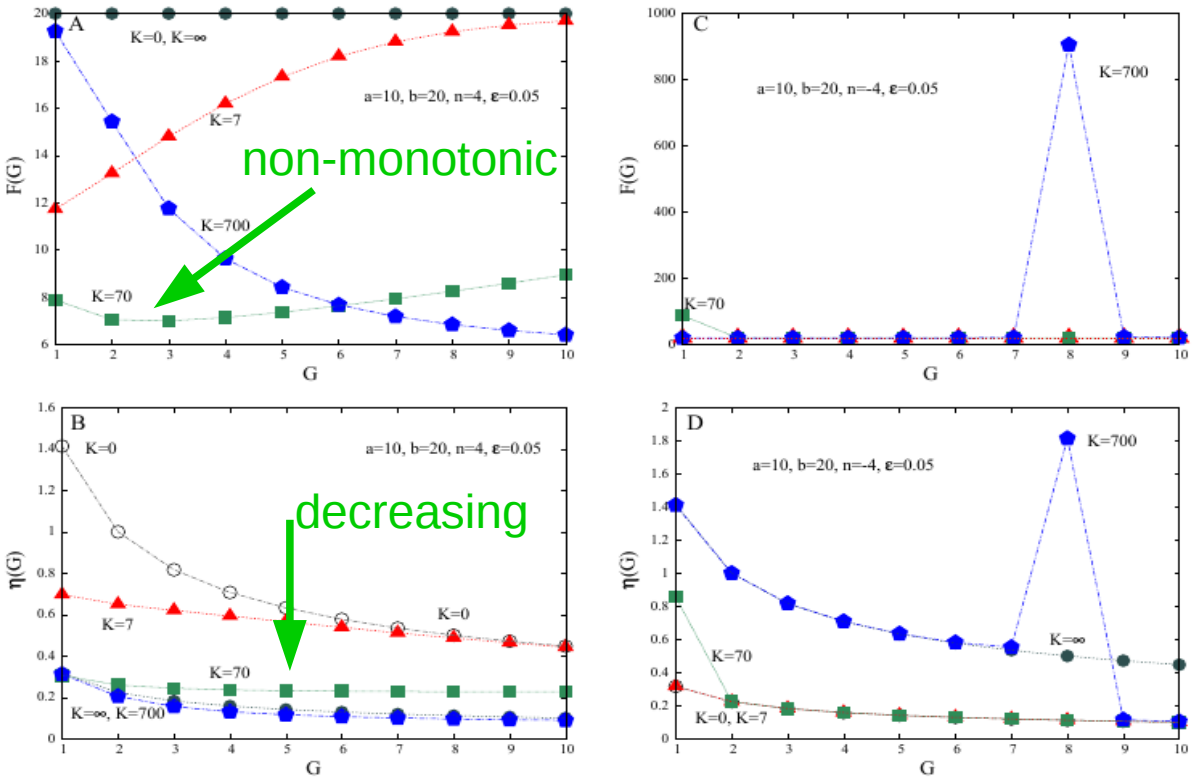
Self-regulating gene: Fano Factor and CV vary in a different manner as G is varied.

Negative feedback

Positive feedback

Fano factor, variance/mean

Coefficient of variation, variance/mean^2



In self-regulating genes, Fano factor and by coefficient of variation may depend on gene copy number in a qualitatively different manner. A, B: Negative auto-regulation, $n = 4$. A: Depending on the feedback strength parameter K, Fano factor $F = \sigma^2/\langle x \rangle$ may both decrease, increase or vary in a non-monotonous manner as gene copy number G is varied. B: Coefficient of variation $\eta = \sigma/\langle x \rangle$ is a monotonically decreasing function of gene copy number G. C, D: Positive auto-regulation, $n = -4$. Here, for $K = 700$, Fano factor $F(G)$ has just one maximum (C), whereas the coefficient of variation $\eta(G)$ has two clear maxima (D). The sharp maximum for $K = 700$ and $G = 8$ is due to the transition of the protein number distribution through bimodality. In absence of gene regulation ($K = 0$ and $K = \infty$), $F = b$ and $\eta \sim G^{-1/2}$. For negatively self-regulating genes, $F(G) < b$ and for positive auto-regulation, $F(G) > b$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

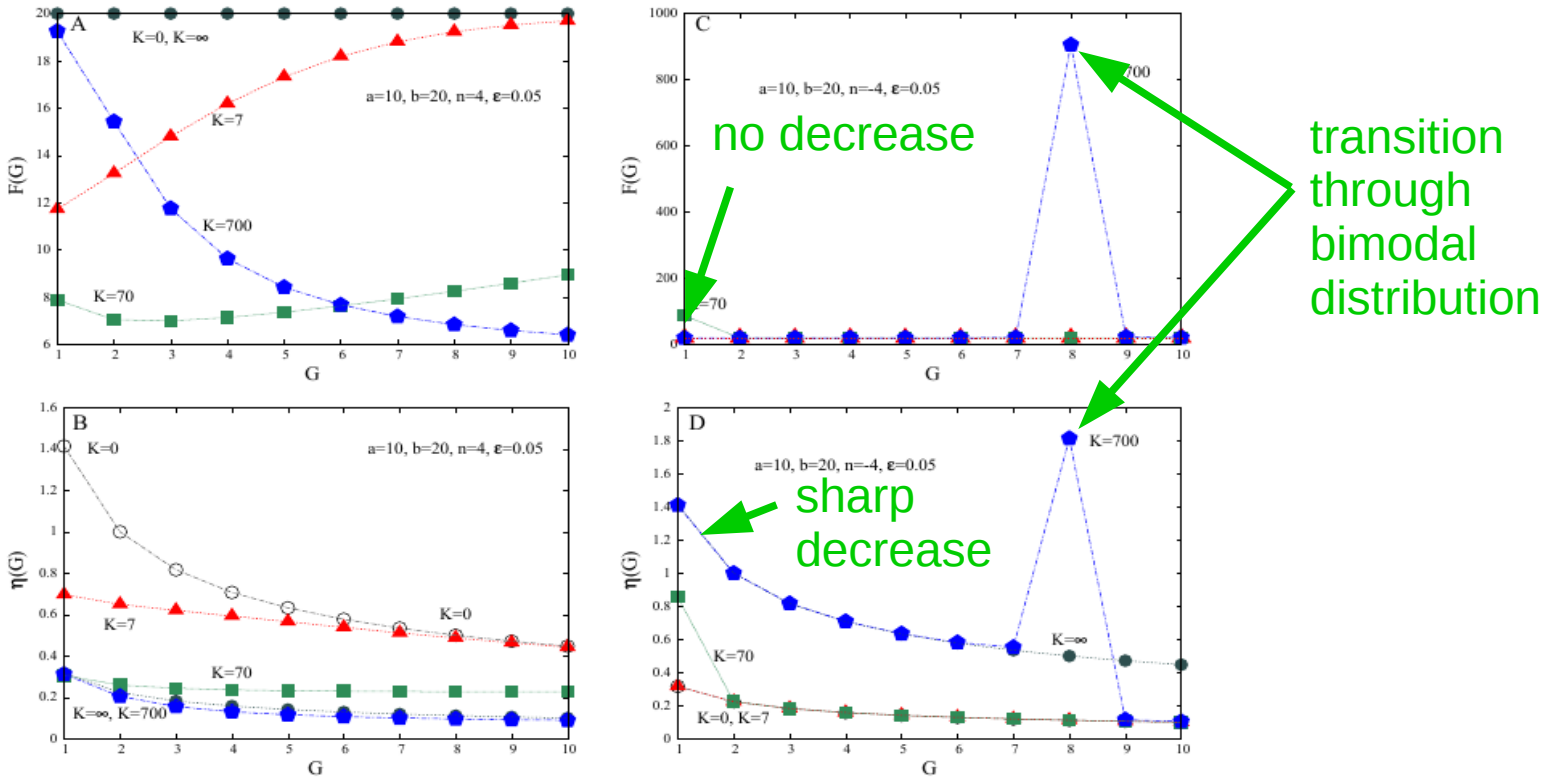
Self-regulating gene: Fano Factor and CV vary in a different manner as G is varied.

Negative feedback

Positive feedback

Fano factor, variance/mean

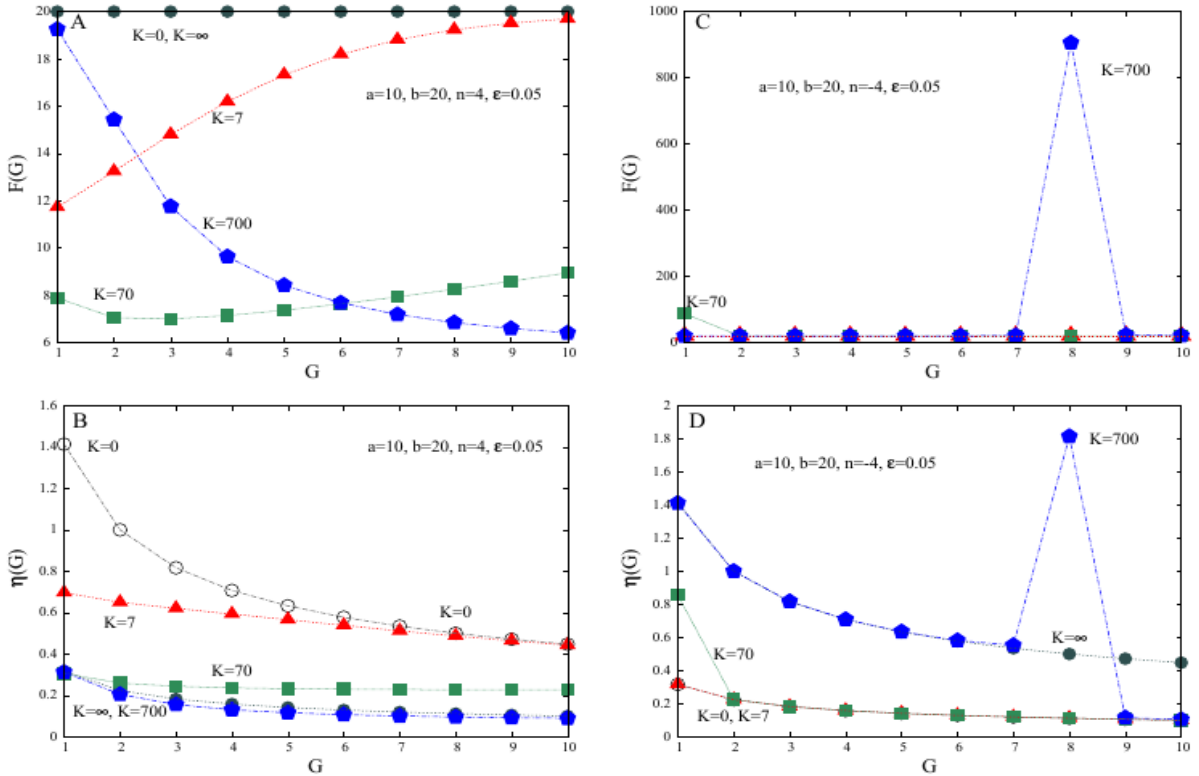
Coefficient of variation, variance/mean^2



In self-regulating genes, Fano factor and by coefficient of variation may depend on gene copy number in a qualitatively different manner. A, B: Negative auto-regulation, $n = 4$. A: Depending on the feedback strength parameter K, Fano factor $F = \sigma^2/\langle x \rangle$ may both decrease, increase or vary in a non-monotonous manner as gene copy number G is varied. B: Coefficient of variation $\eta = \sigma/\langle x \rangle$ is a monotonically decreasing function of gene copy number G. C, D: Positive auto-regulation, $n = -4$. Here, for $K = 700$, Fano factor $F(G)$ has just one maximum (C), whereas the coefficient of variation $\eta(G)$ has two clear maxima (D). The sharp maximum for $K = 700$ and $G = 8$ is due to the transition of the protein number distribution through bimodality. In absence of gene regulation ($K = 0$ and $K = \infty$), $F = b$ and $\eta \sim G^{-1/2}$. For negatively self-regulating genes, $F(G) < b$ and for positive auto-regulation, $F(G) > b$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Self-regulating gene: Fano Factor and CV vary in a different manner as G is varied.



This demonstrates that experimental assessments of the influence of gene expression noise on cell fitness may be ambiguous because they are dependent on the particular function used to quantify noise.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

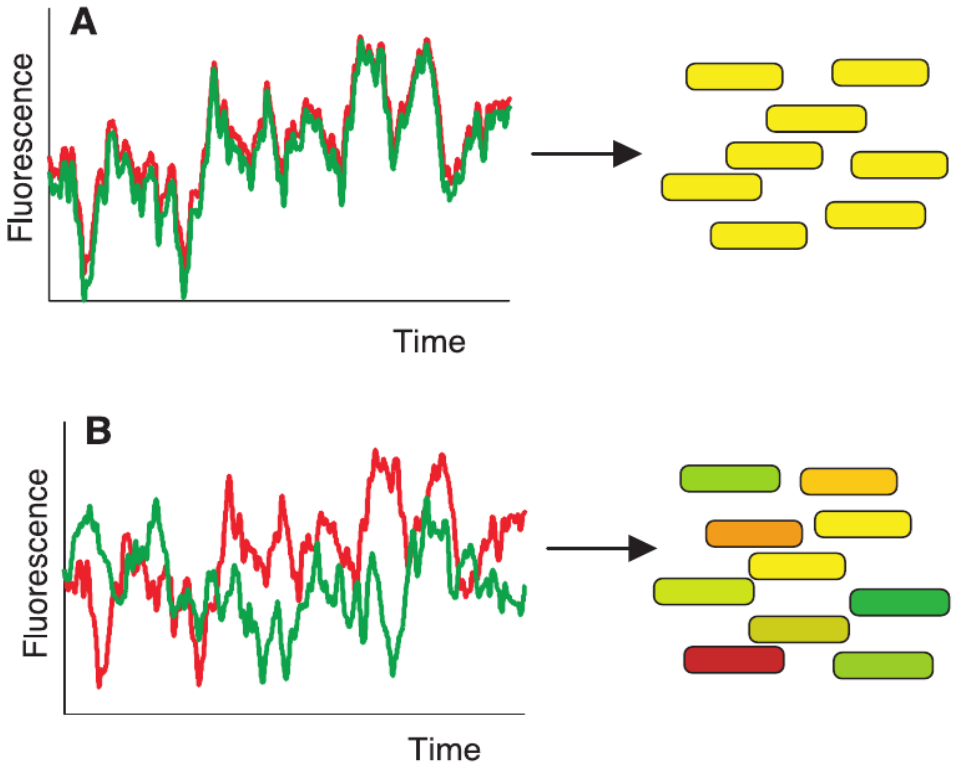
Result 1

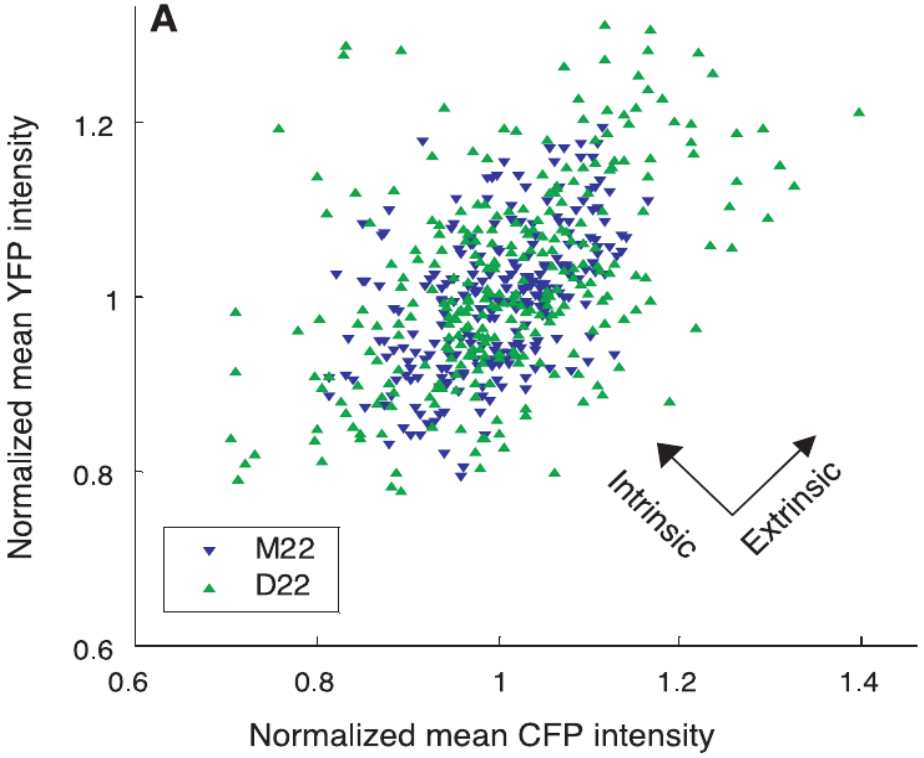
Ambiguity of the measures of noise

Result 2

One-reporter assay cannot be used
for experimental measurement of noise
in self-regulated genes

2-reporter assay





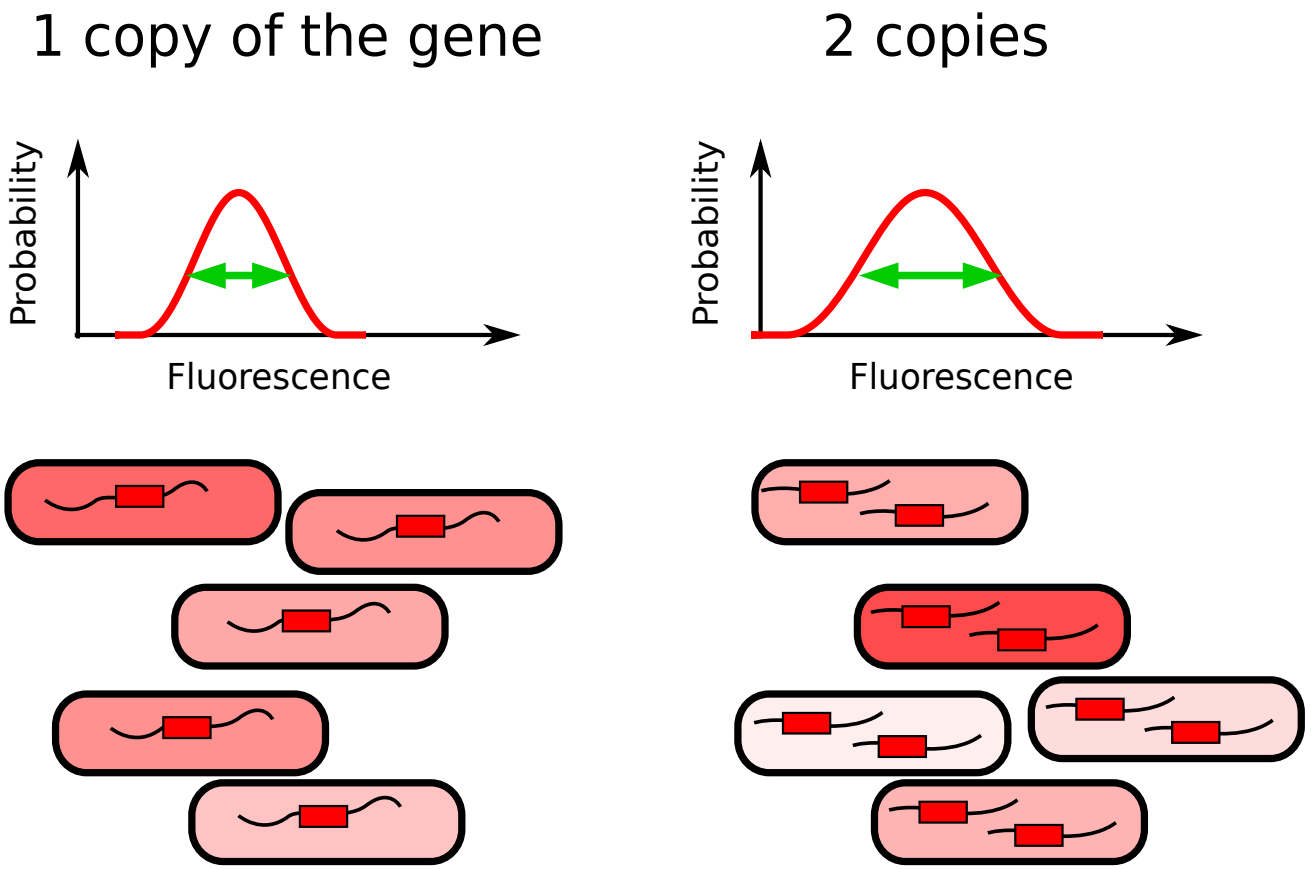
Compares the variability of expression of two reporter genes

This method is often used to determine intrinsic and extrinsic contributions to gene expression noise.

Problem: Equivalence between the two fluorescent proteins

[3] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).

1-reporter assay



Stewart-Ornstein et al:
Compares the variability in gene expression between these two cell populations

[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

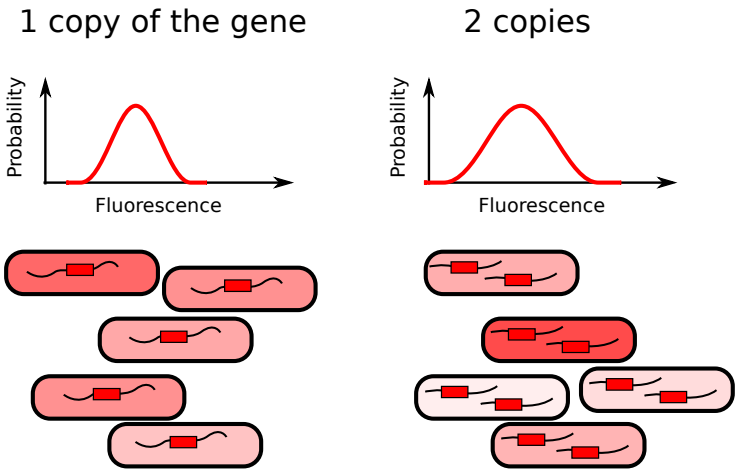
1-color assay

Extrinsic noise was calculated from the one- and two-FP strain measurements using the following formula:

$$Var(a_1 + a_2) = Var(a_1) + Var(a_2) + 2^{*}Cov(a_1, a_2)$$
$$Cov(a_1, a_2) = [Var(a_1 + a_2) - 2^{*}Var(a_1)]/2$$

Here, a_1 and a_2 are indistinguishable alleles of the same gene. $Var(a_1) = Var(a_2)$ is measured in the one-FP strains. $Var(a_1 + a_2)$ is measured in the two-FP strains. Extrinsic noise is then given by the normalized covariance

$$\sqrt{\frac{Cov(a_1, a_2)}{\bar{a}_1 {}^{*} \bar{a}_2}} = Noise_{ext}$$



[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

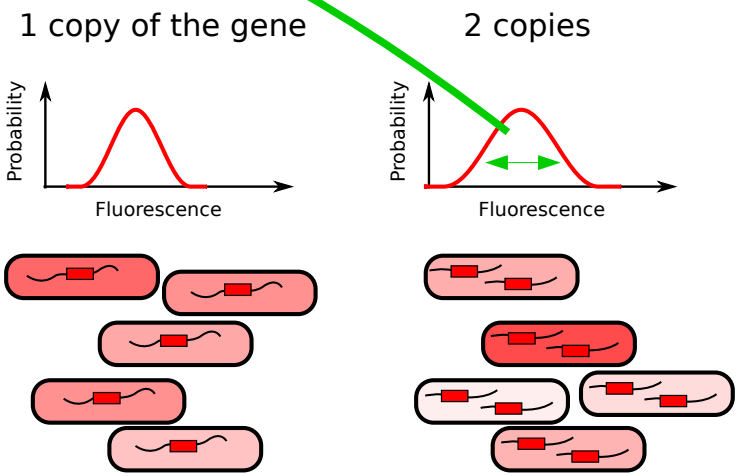
1-color assay

Extrinsic noise was calculated from the one- and two-FP strain measurements using the following formula:

$$Var(a_1 + a_2) = Var(a_1) + Var(a_2) + 2^{*}Cov(a_1, a_2)$$
$$Cov(a_1, a_2) = [Var(a_1 + a_2) - 2^{*}Var(a_1)]/2$$

Here, a_1 and a_2 are indistinguishable alleles of the same gene. $Var(a_1) = Var(a_2)$ is measured in the one-FP strains. $Var(a_1 + a_2)$ is measured in the two-FP strains. Extrinsic noise is then given by the normalized covariance

$$\sqrt{\frac{Cov(a_1, a_2)}{\bar{a}_1 {}^{*} \bar{a}_2}} = Noise_{ext}$$



[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

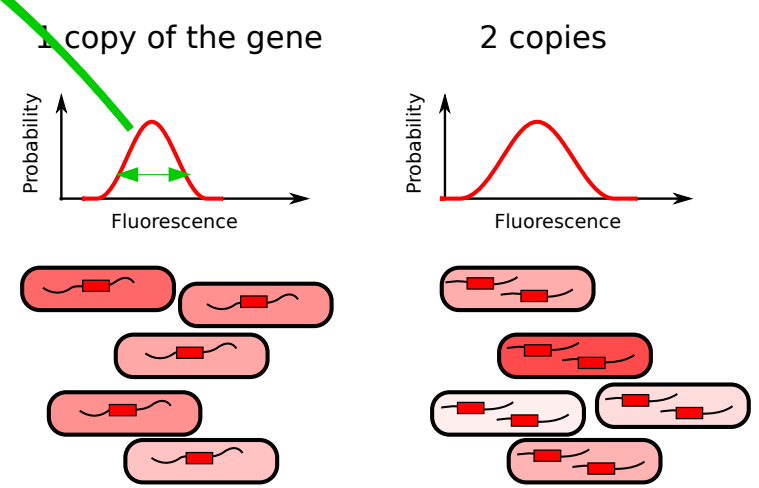
1-color assay

Extrinsic noise was calculated from the one- and two-FP strain measurements using the following formula:

$$Var(a_1 + a_2) = Var(a_1) + Var(a_2) + 2^{*}Cov(a_1, a_2)$$
$$Cov(a_1, a_2) = [Var(a_1 + a_2) - 2^{*}Var(a_1)]/2$$

Here, a_1 and a_2 are indistinguishable alleles of the same gene. $Var(a_1) = Var(a_2)$ is measured in the one-FP strains. $Var(a_1 + a_2)$ is measured in the two-FP strains. Extrinsic noise is then given by the normalized covariance

$$\sqrt{\frac{Cov(a_1, a_2)}{\bar{a}_1 {}^{*} \bar{a}_2}} = Noise_{ext}$$



[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

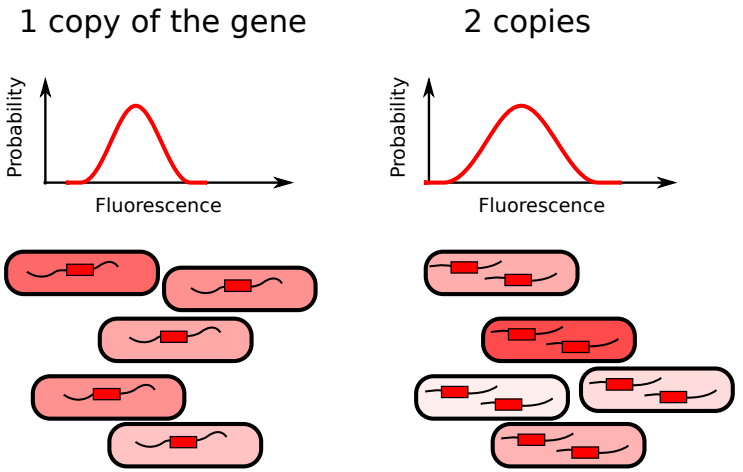
1-color assay

Extrinsic noise was calculated from the one- and two-FP strain measurements using the following formula:

$$Var(a_1 + a_2) = Var(a_1) + Var(a_2) + 2^{*}Cov(a_1, a_2)$$
$$Cov(a_1, a_2) = [Var(a_1 + a_2) - 2^{*}Var(a_1)]/2$$

Here, a_1 and a_2 are indistinguishable alleles of the same gene. $Var(a_1) = Var(a_2)$ is measured in the one-FP strains. $Var(a_1 + a_2)$ is measured in the two-FP strains. Extrinsic noise is then given by the normalized covariance

$$\sqrt{\frac{Cov(a_1, a_2)}{\bar{a}_1 {}^{*} \bar{a}_2}} = Noise_{ext}$$



[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

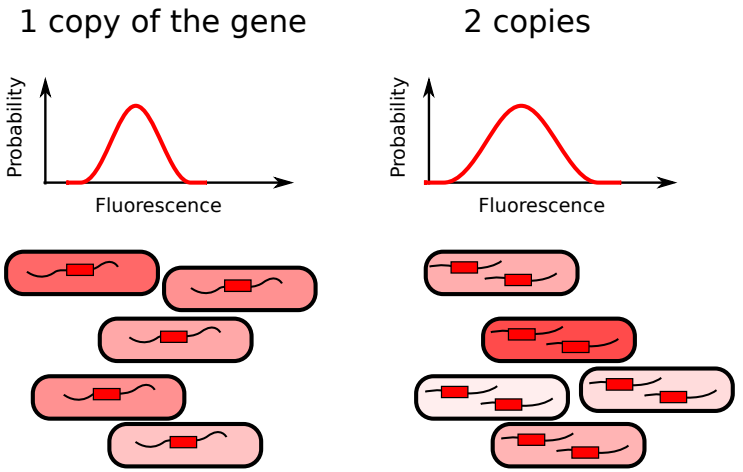
1-color assay

Extrinsic noise was calculated from the one- and two-FP strain measurements using the following formula:

$$Var(a_1 + a_2) = Var(a_1) + Var(a_2) + 2^{*}Cov(a_1, a_2)$$
$$Cov(a_1, a_2) = [Var(a_1 + a_2) - 2^{*}Var(a_1)]/2$$

Here, a_1 and a_2 are indistinguishable alleles of the same gene. $Var(a_1) = Var(a_2)$ is measured in the one-FP strains. $Var(a_1 + a_2)$ is measured in the two-FP strains. Extrinsic noise is then given by the normalized covariance

$$\sqrt{\frac{Cov(a_1, a_2)}{\bar{a}_1 {}^{*} \bar{a}_2}} = Noise_{ext}$$



[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

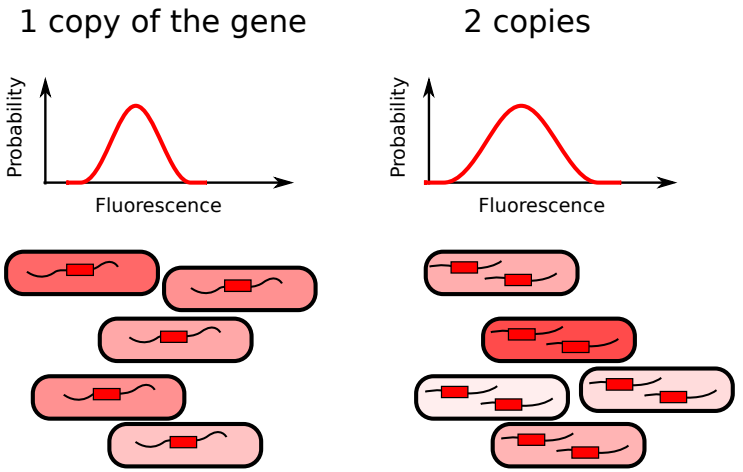
1-color assay

Extrinsic noise was calculated from the one- and two-FP strain measurements using the following formula:

$$Var(a_1 + a_2) = Var(a_1) + Var(a_2) + 2^{*}Cov(a_1, a_2)$$
$$Cov(a_1, a_2) = [Var(a_1 + a_2) - 2^{*}Var(a_1)]/2$$

Here, a_1 and a_2 are indistinguishable alleles of the same gene. $Var(a_1) = Var(a_2)$ is measured in the one-FP strains. $Var(a_1 + a_2)$ is measured in the two-FP strains. Extrinsic noise is then given by the normalized covariance

$$\sqrt{\frac{Cov(a_1, a_2)}{\bar{a}_1 {}^{*} \bar{a}_2}} = Noise_{ext}$$



[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

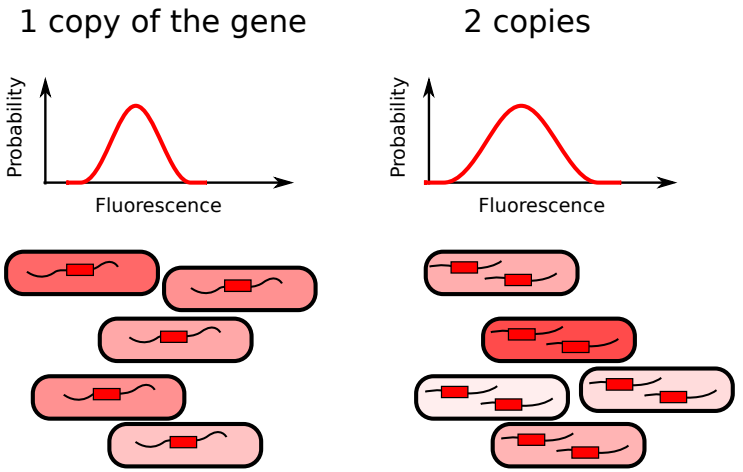
1-color assay

Extrinsic noise was calculated from the one- and two-FP strain measurements using the following formula:

$$Var(a_1 + a_2) = Var(a_1) + Var(a_2) + 2^{*}Cov(a_1, a_2)$$
$$Cov(a_1, a_2) = [Var(a_1 + a_2) - 2^{*}Var(a_1)]/2$$

Here, a_1 and a_2 are indistinguishable alleles of the same gene. $Var(a_1) = Var(a_2)$ is measured in the one-FP strains. $Var(a_1 + a_2)$ is measured in the two-FP strains. Extrinsic noise is then given by the normalized covariance

$$\sqrt{\frac{Cov(a_1, a_2)}{\bar{a}_1 {}^{*} \bar{a}_2}} = Noise_{ext}$$



Something is wrong with this interpretation of noise:
Negative number under square root...

[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

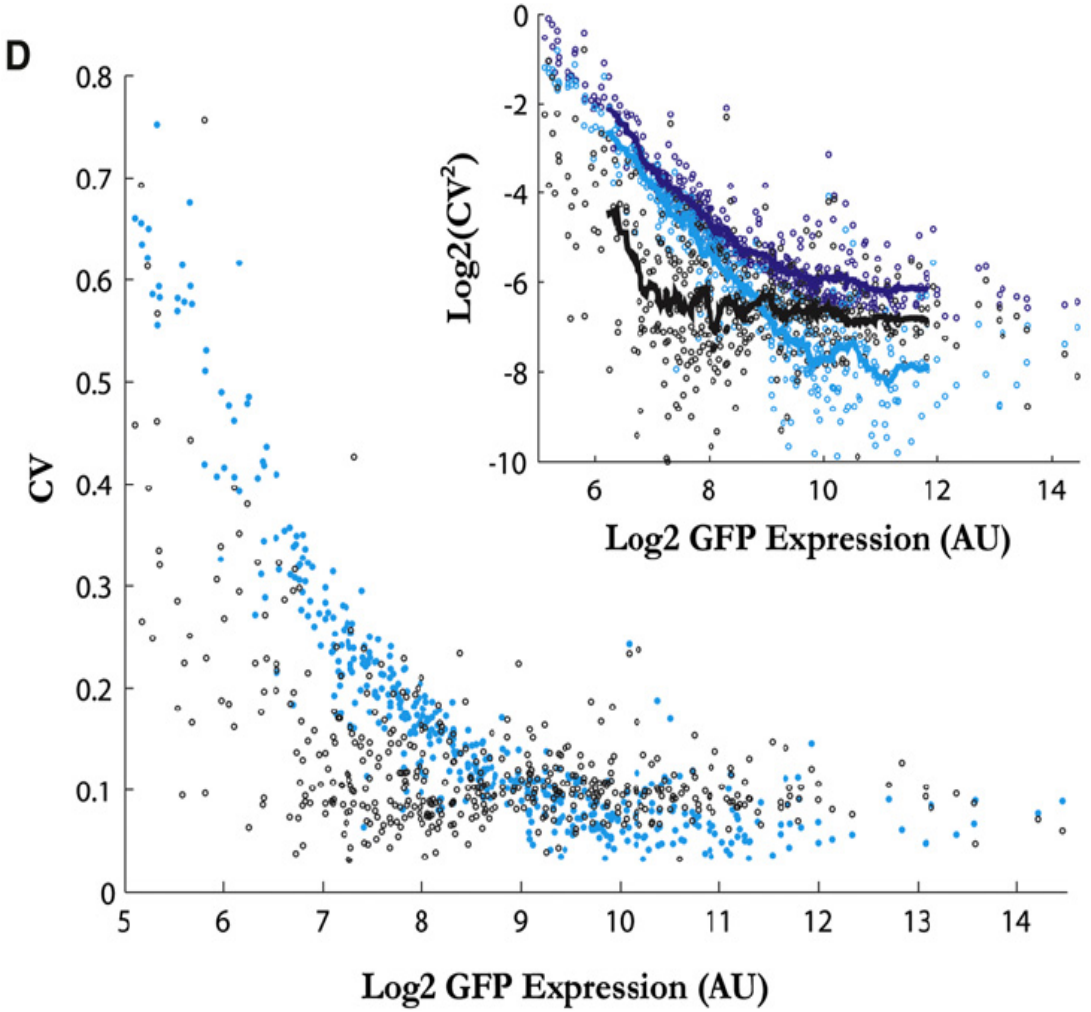
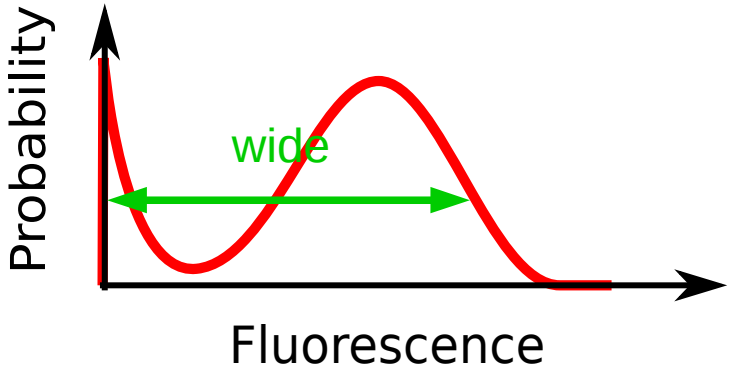


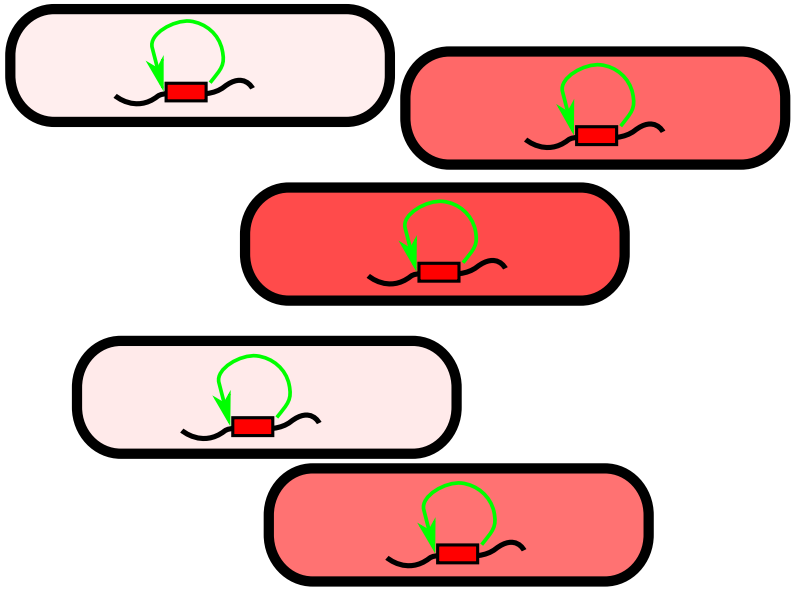
Figure 1. Extrinsic Versus Intrinsic Noise Decomposition across the Proteome

(D) Intrinsic (cyan) and extrinsic(black) noise plotted against log2 mean expression for 465 genes. Inset: log2(CV^2) plotted against log2(mean), running means (smoothing window of 30) for intrinsic(cyan), extrinsic(black), total (dark blue) noise.

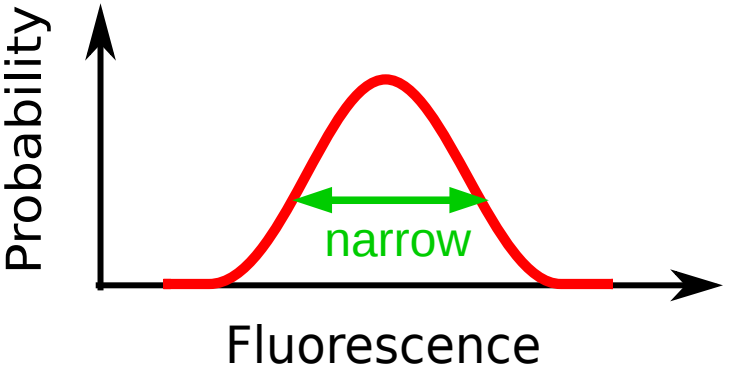
[4] J. Stewart-Ornstein, J. S. Weissman, and H. El-Samad, Molecular cell **45**, 483 (2012).

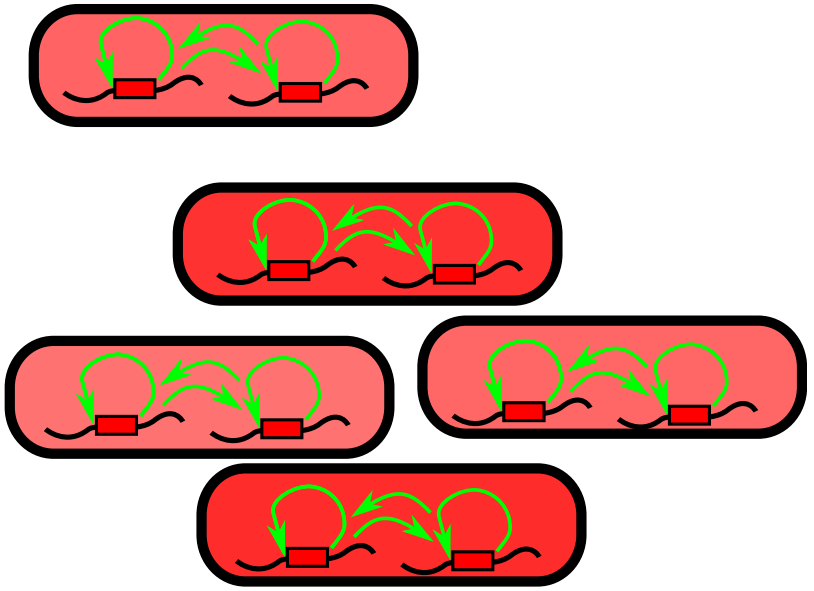
1 copy of the gene





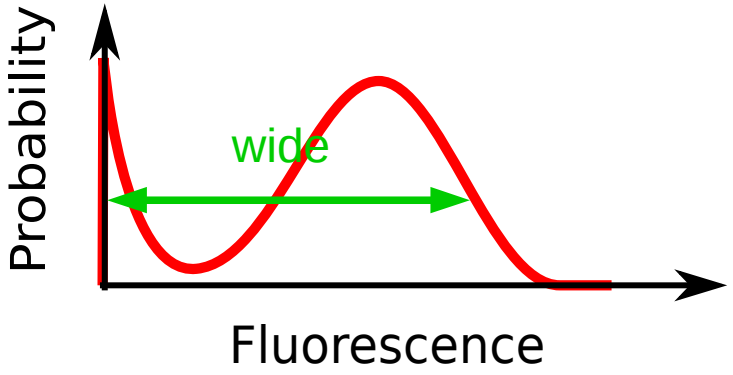
2 copies

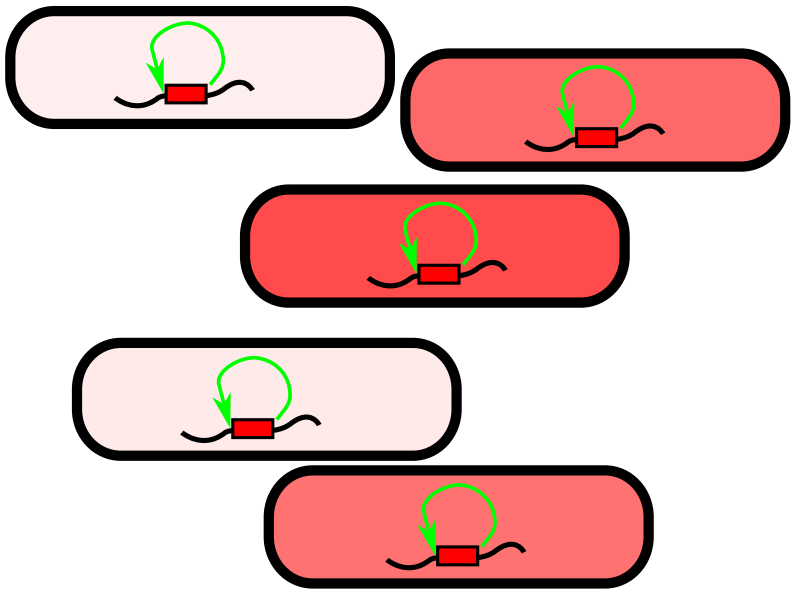




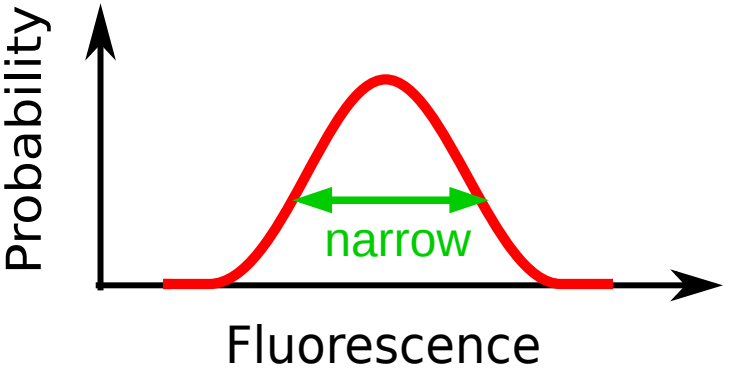
If the gene is self-regulated, the histogram of its expression in the 1-copy strain can be wider than the histogram for the 2-copy strain

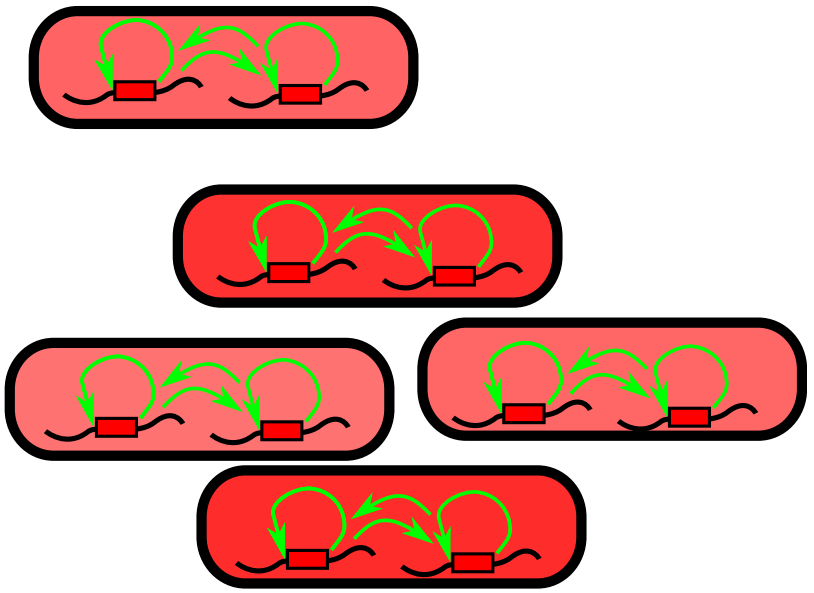
1 copy of the gene





2 copies

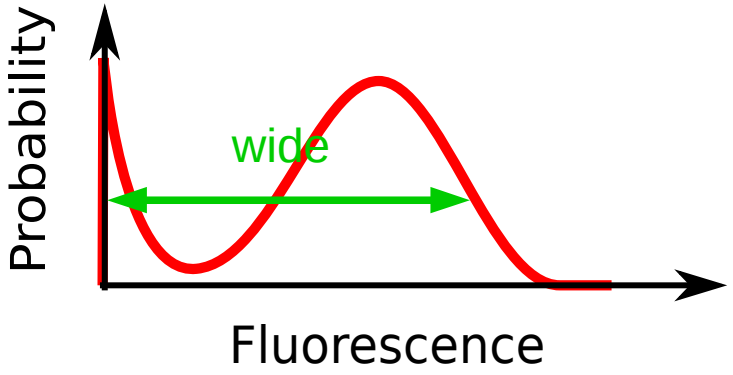


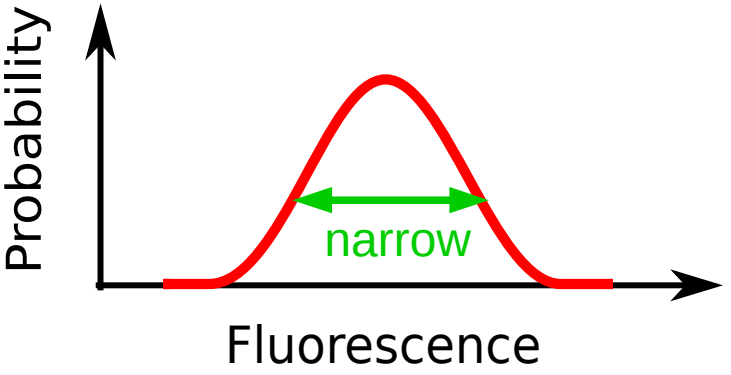


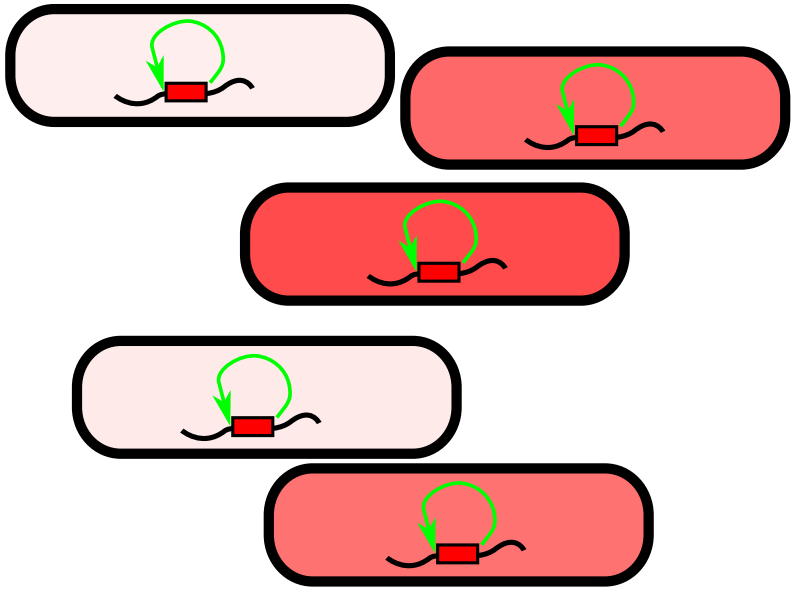
It can happen that *var(2 copies) - 2*var(1 copy) < 0*

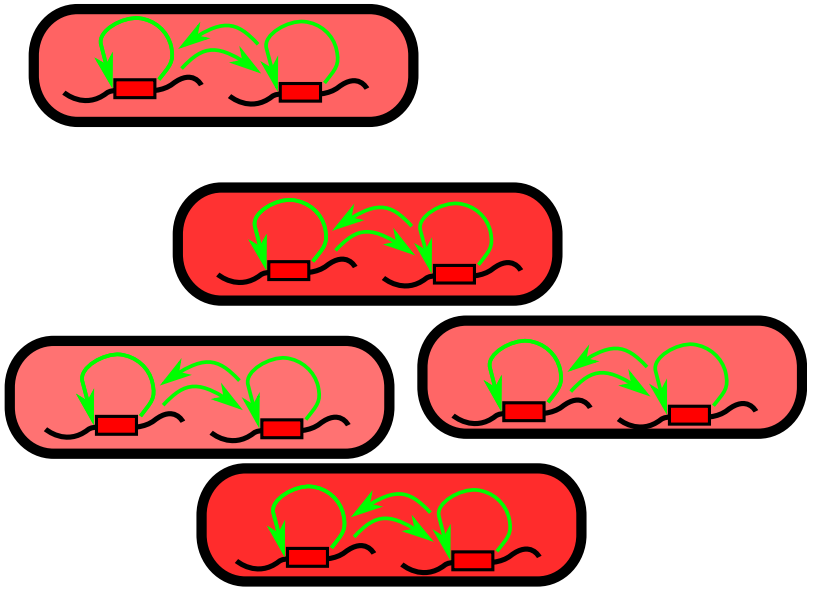
1 copy of the gene

2 copies

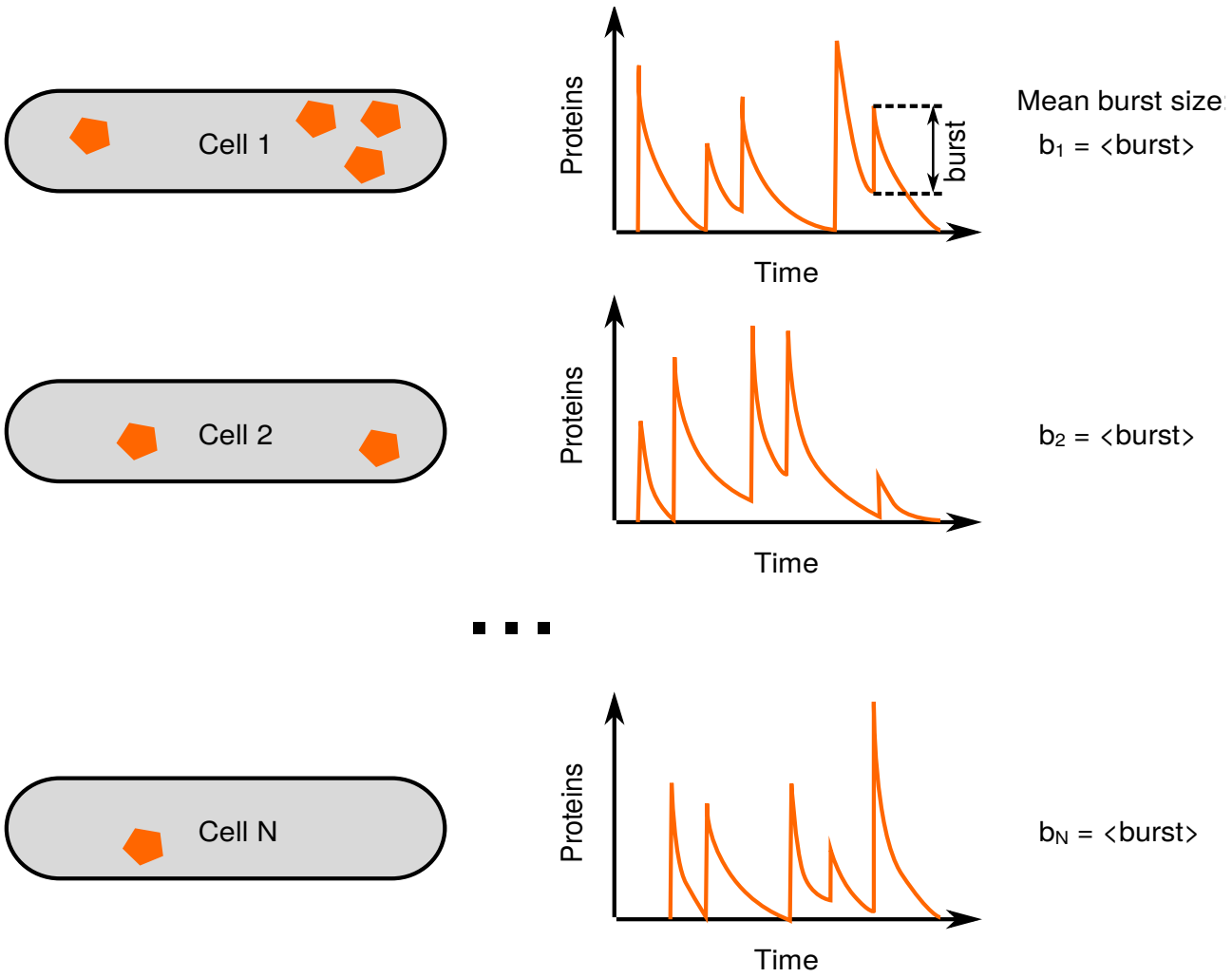


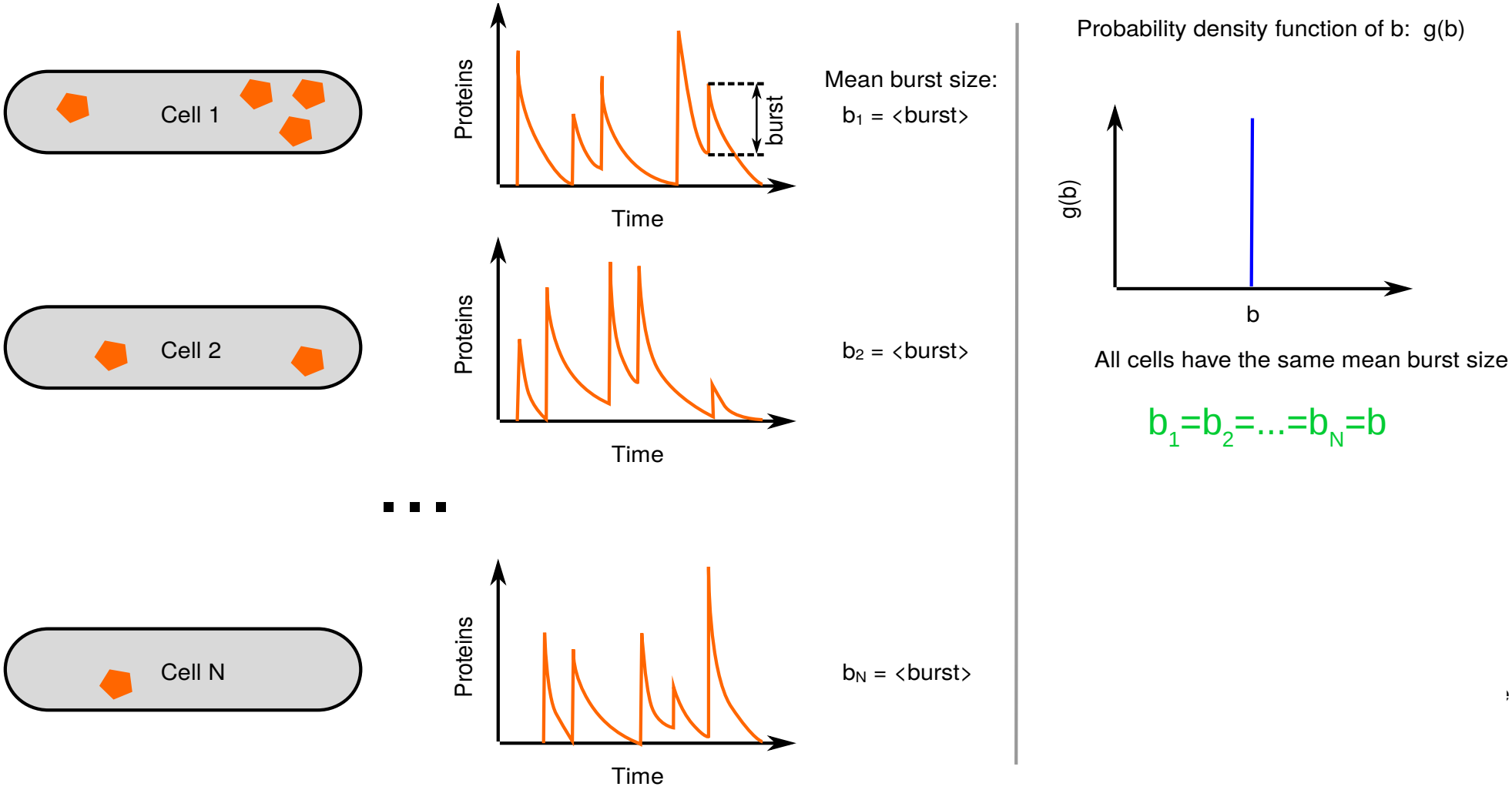


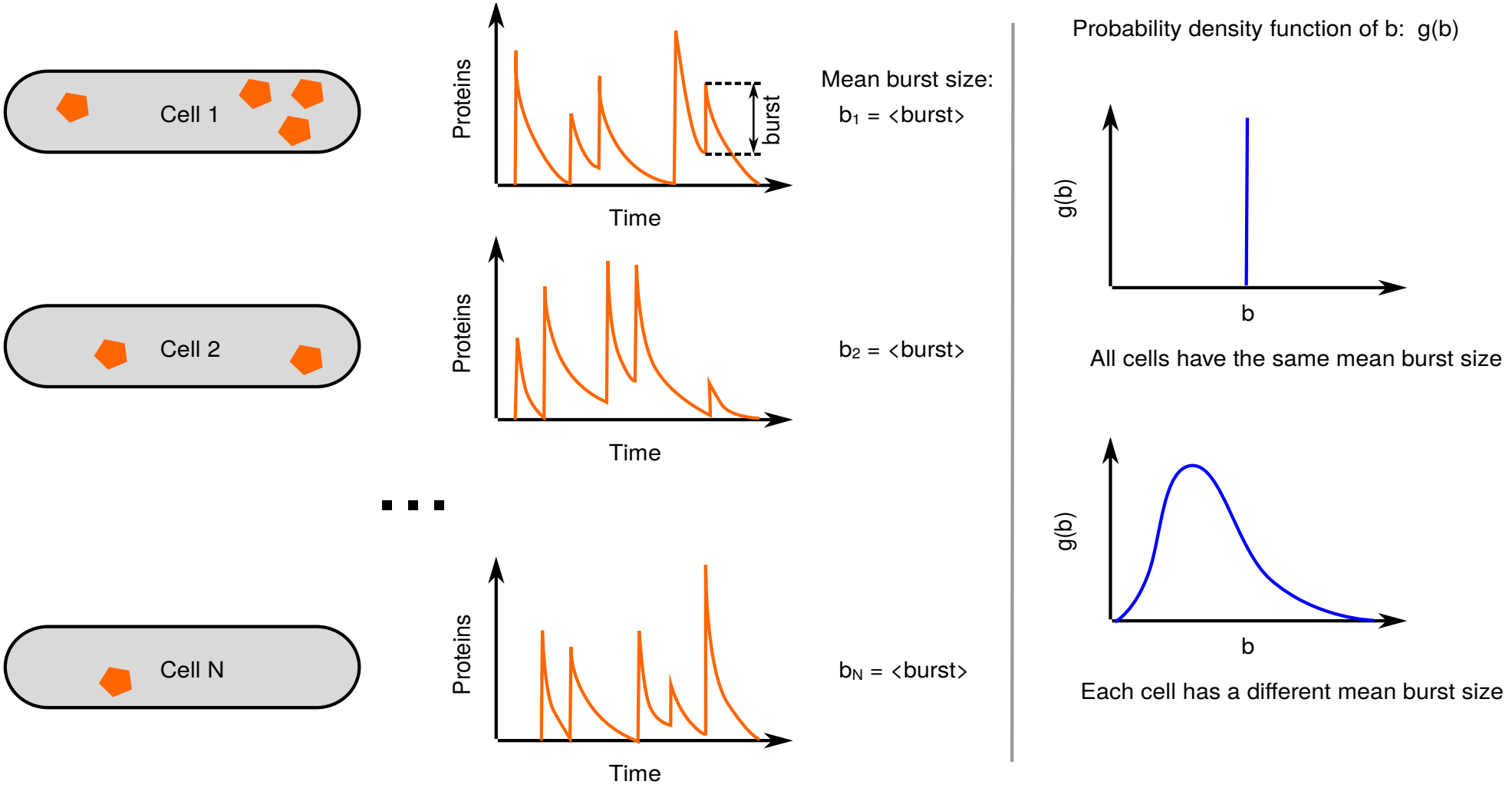


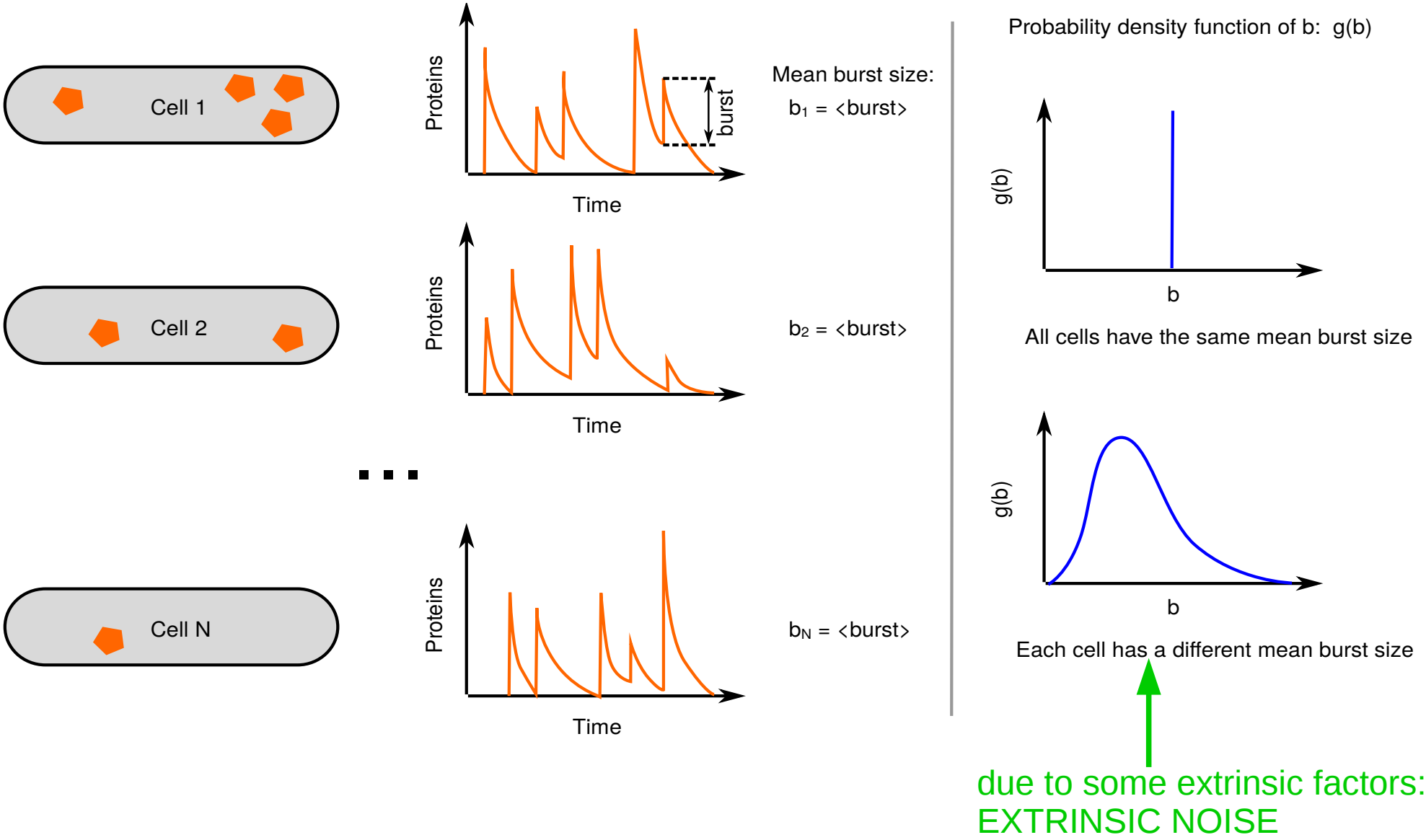


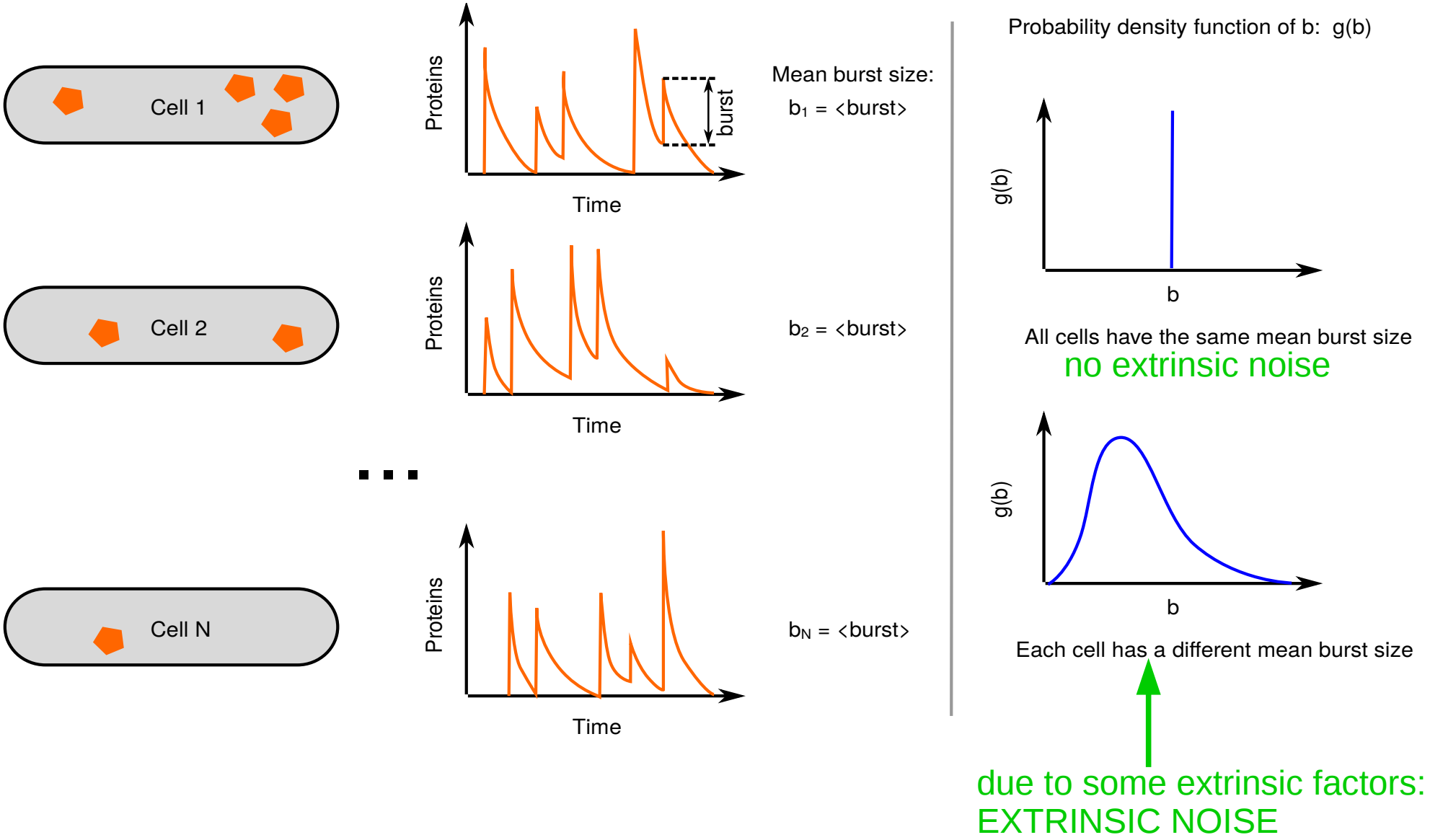
It can also happen that *var(2 copies) - 2*var(1 copy) = 0*
but extrinsic noise is present


Cell 1
Proteins
burst
Time
Mean burst size:
b1 = <burst>
Cell 2
Proteins
Time
b2 = <burst>
Cell N
Proteins
Time
bN = <burst>

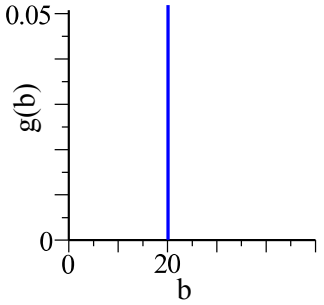

Cell 1
Proteins
burst
Time
Mean burst size:
b_1 = <burst>
Cell 2
Proteins
Time
b_2 = <burst>
Cell N
Proteins
Time
b_N = <burst>
Probability density function of b: g(b)
g(b)
b
All cells have the same mean burst size
$b_1=b_2=...=b_N=b$


Cell 1
Cell 2
Cell N
Proteins
Time
burst
Mean burst size:
b_1 = <burst>
b_2 = <burst>
b_N = <burst>
Probability density function of b: g(b)
g(b)
b
All cells have the same mean burst size
Each cell has a different mean burst size


Cell 1
Cell 2
Cell N
Proteins
Time
burst
Mean burst size:
b_1 = <burst>
b_2 = <burst>
b_N = <burst>
Probability density function of b: g(b)
g(b)
b
All cells have the same mean burst size
Each cell has a different mean burst size
due to some extrinsic factors:
EXTRINSIC NOISE

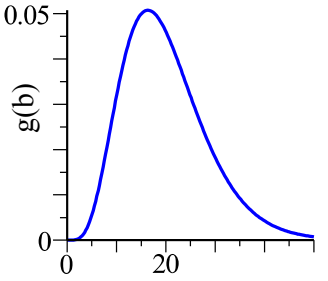

Cell 1
Cell 2
Cell N
Proteins
Time
burst
Mean burst size:
b_1 = <burst>
b_2 = <burst>
b_N = <burst>
Probability density function of b: g(b)
g(b)
b
All cells have the same mean burst size
no extrinsic noise
Each cell has a different mean burst size
due to some extrinsic factors:
EXTRINSIC NOISE

Mean burst size is the same in each cell (no extrinsic noise)



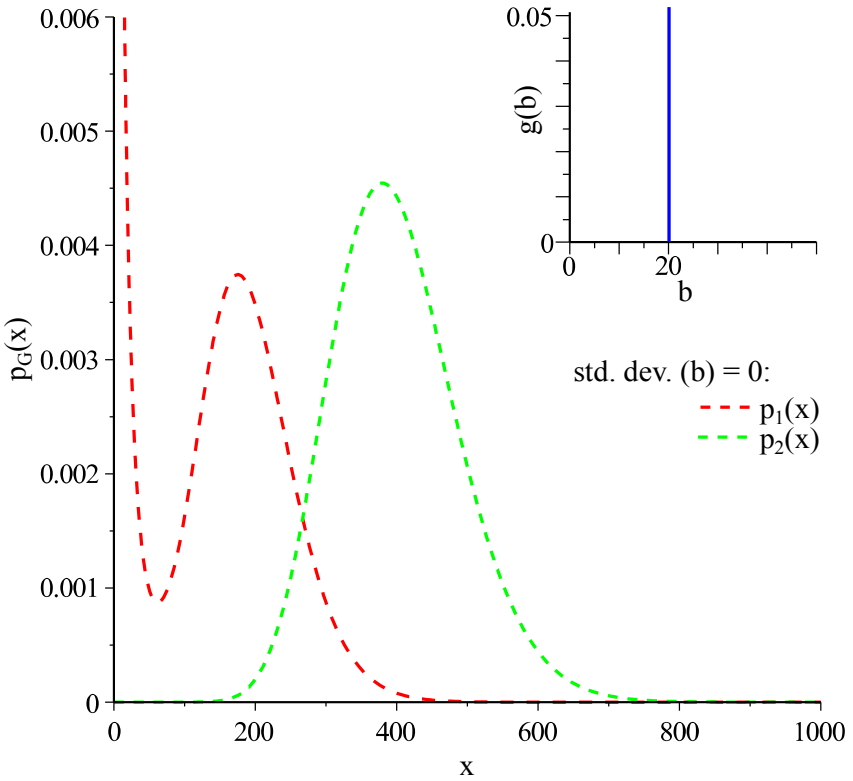
std. dev. (b) = 0:

Mean burst size differs from cell to cell (extrinsic noise)

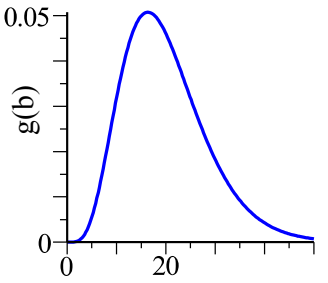


std. dev. (b) = 8.528:

Mean burst size is
the same in each cell
(no extrinsic noise)



Mean burst size
differs from cell to cell
(extrinsic noise)



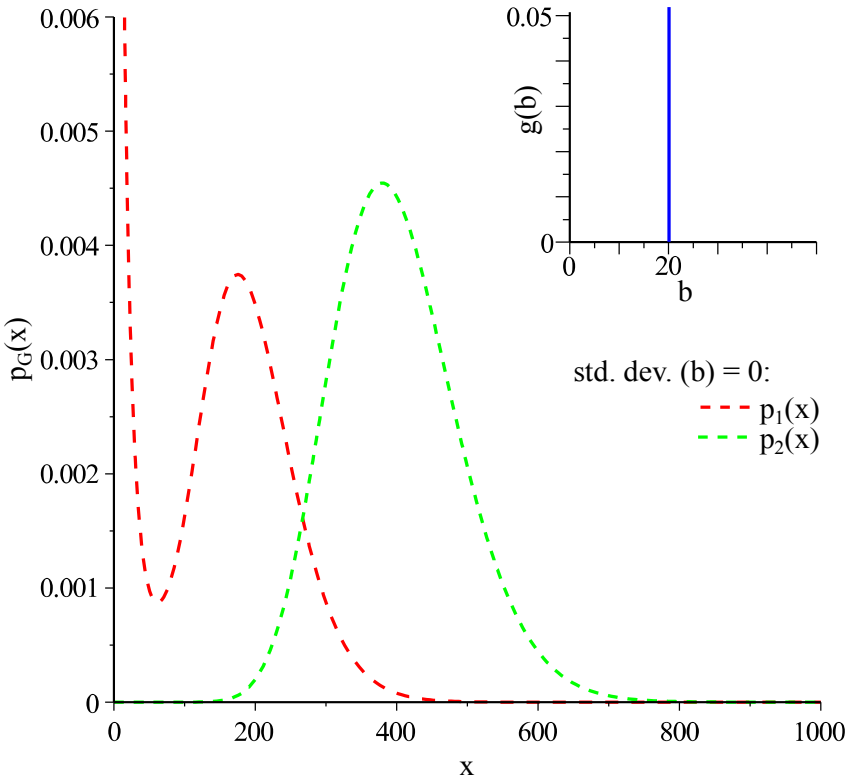
std. dev. (b) = 8.528:

$$\frac{\mathrm{var}(x_1 + x_2) - 2\ \mathrm{var}(x_1)}{\langle x_1^2 \rangle} < 0$$

„measure of noise”
according to Stewart-Ornstein et al.

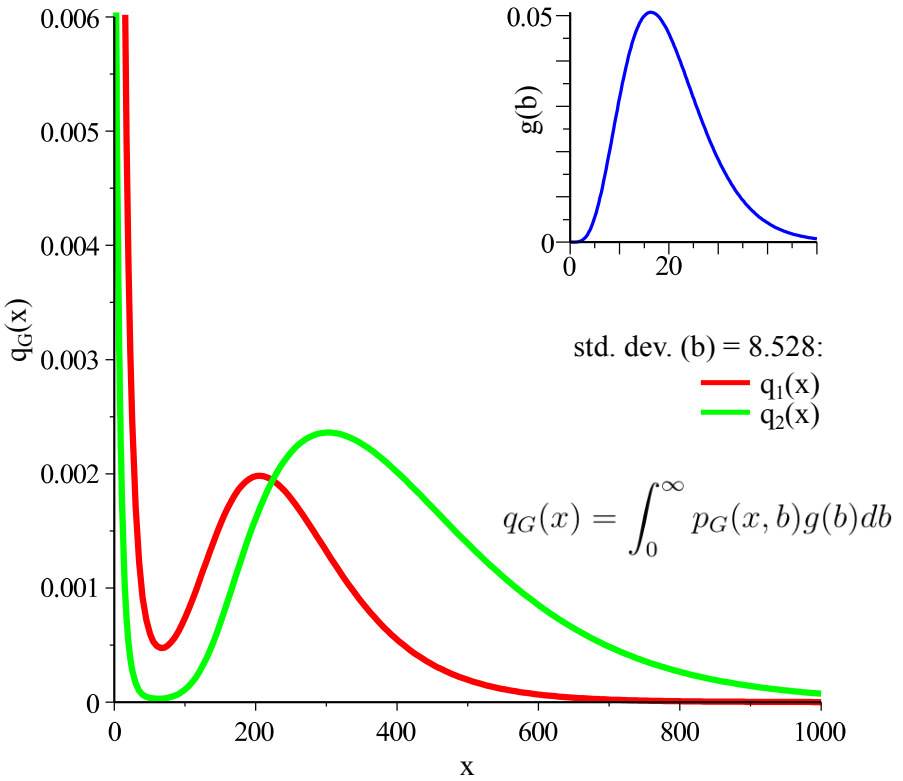
J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Mean burst size is
the same in each cell
(no extrinsic noise)



$$\frac{\mathrm{var}(x_1+x_2)-2\ \mathrm{var}(x_1)}{\langle x_1^2\rangle}<0$$

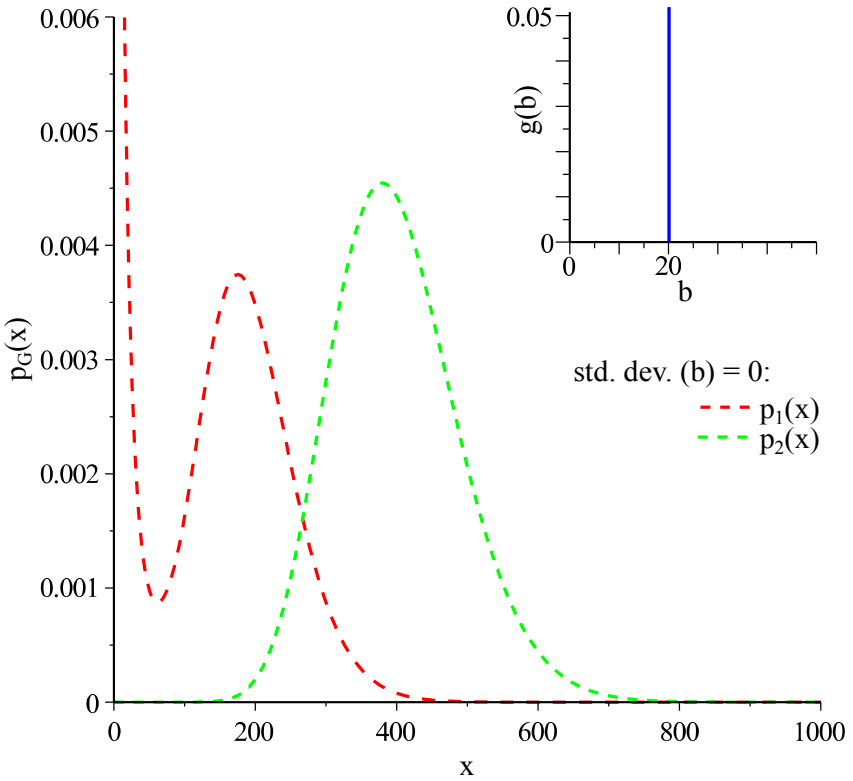
Mean burst size
differs from cell to cell
(extrinsic noise)



$$\frac{\mathrm{var}(x_1+x_2)-2\ \mathrm{var}(x_1)}{\langle x_1^2\rangle}=0$$

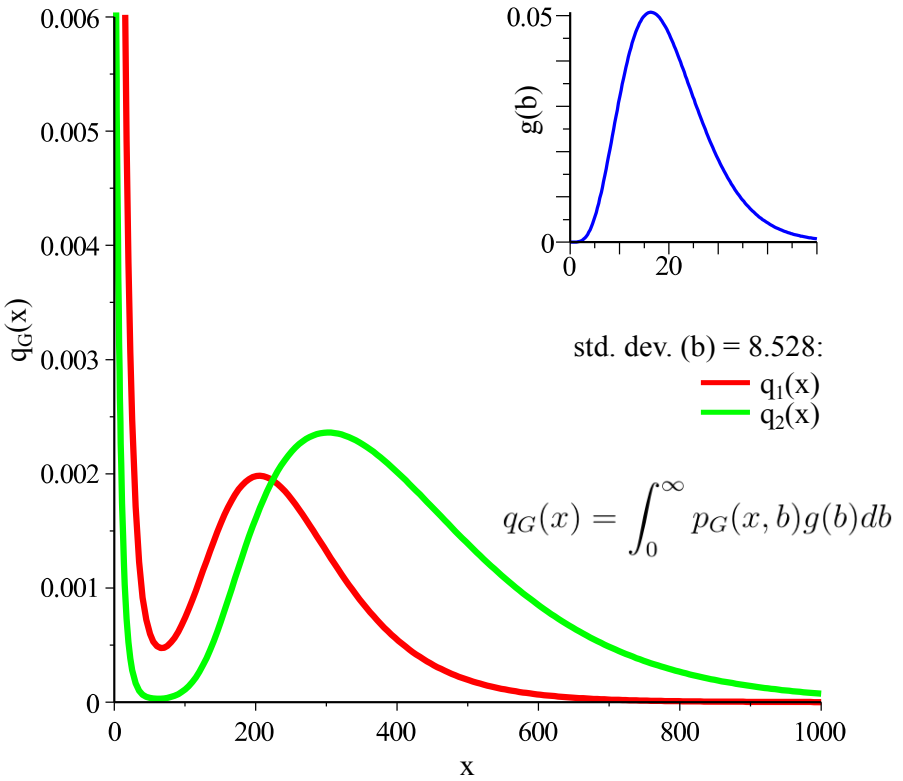
J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Mean burst size is
the same in each cell
(no extrinsic noise)



$$\frac{\mathrm{var}(x_1+x_2)-2\ \mathrm{var}(x_1)}{\langle x_1^2\rangle}<0$$

Mean burst size
differs from cell to cell
(**extrinsic noise**)



$$\frac{\mathrm{var}(x_1+x_2)-2\ \mathrm{var}(x_1)}{\langle x_1^2\rangle}=0$$

If interpreted according to Stewart-Ornstein et al.,
this would mean that extrinsic noise is zero,
which is not true!

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

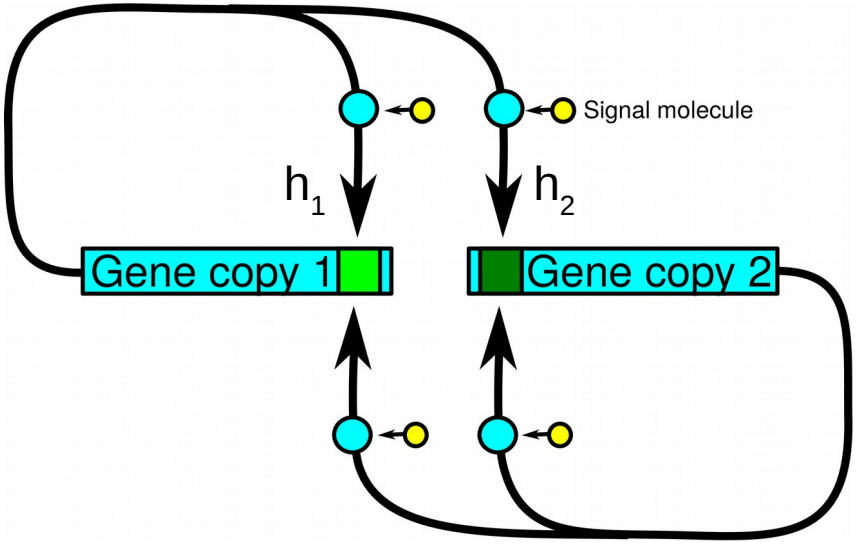
Result 2

One-reporter assay cannot be used
for experimental measurement of noise
in self-regulated genes

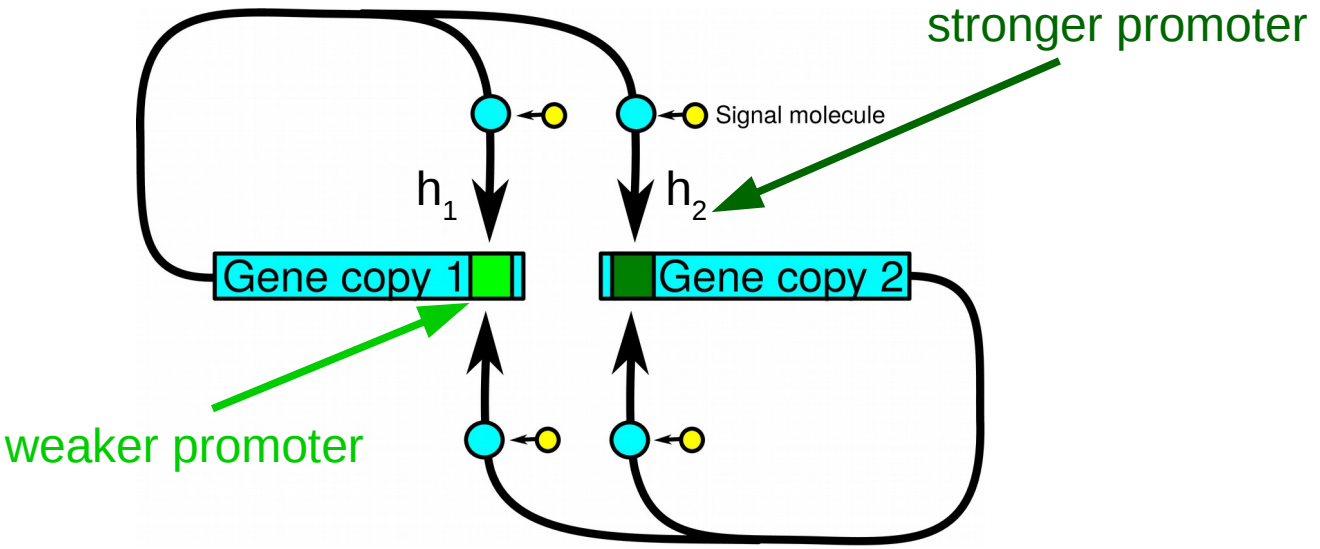
Result 3

Imperfect gene duplication may lead to
mixed, binary+graded response
of the gene system to a signal

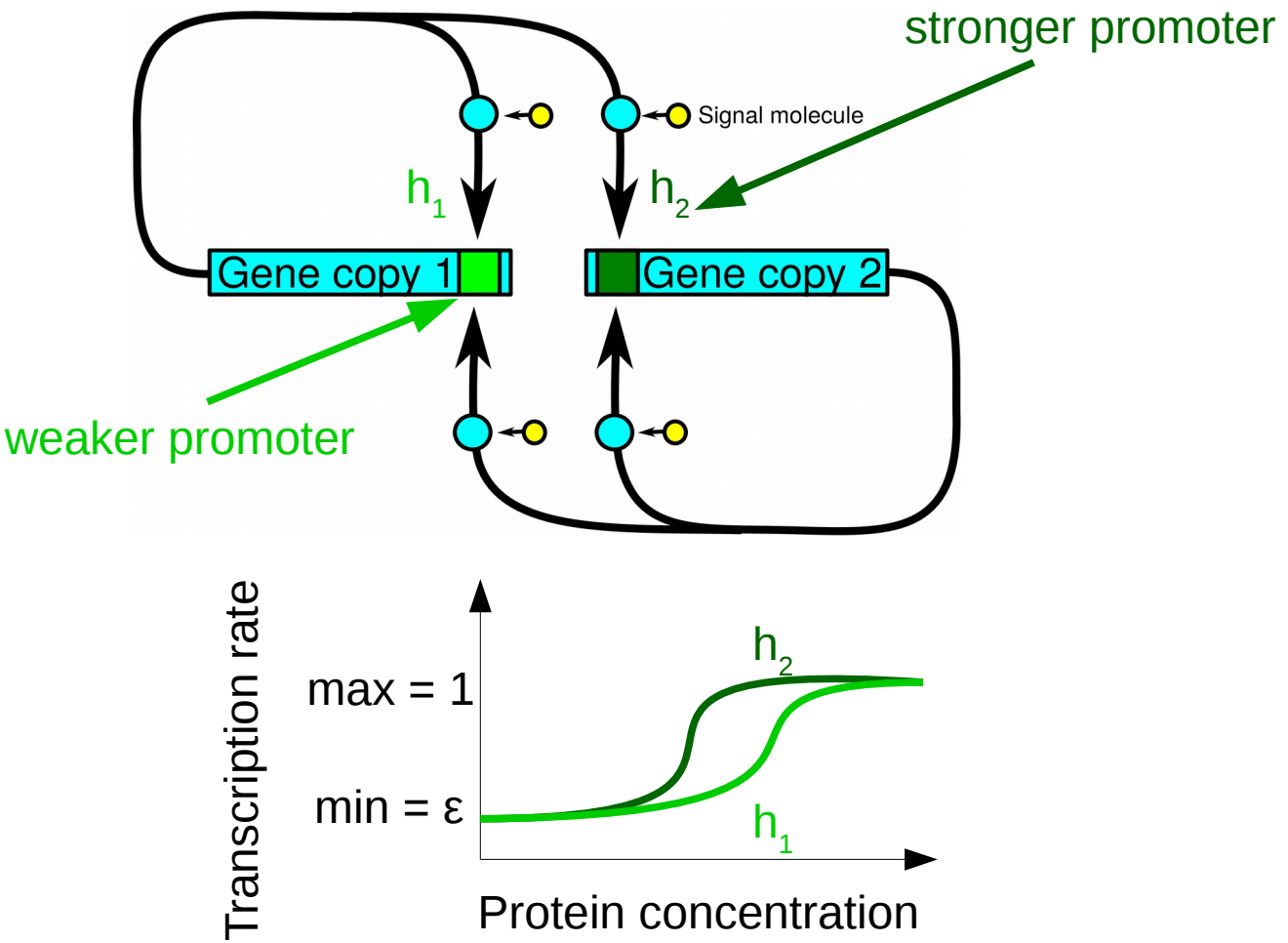
Two non-identical gene copies (imperfect duplication)



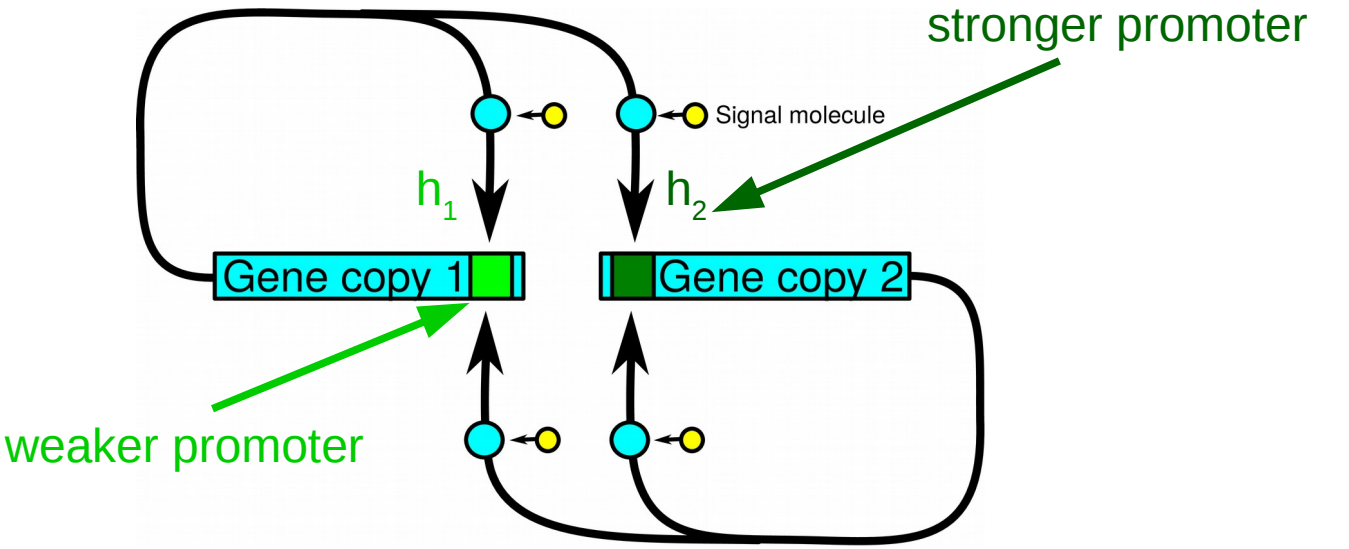
Two non-identical gene copies (imperfect duplication)

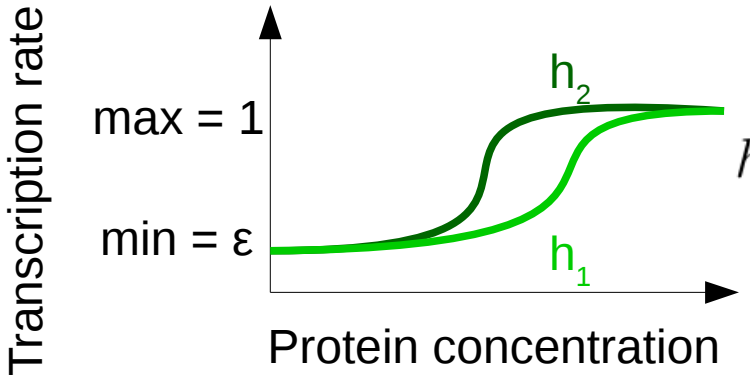


Two non-identical gene copies (imperfect duplication)



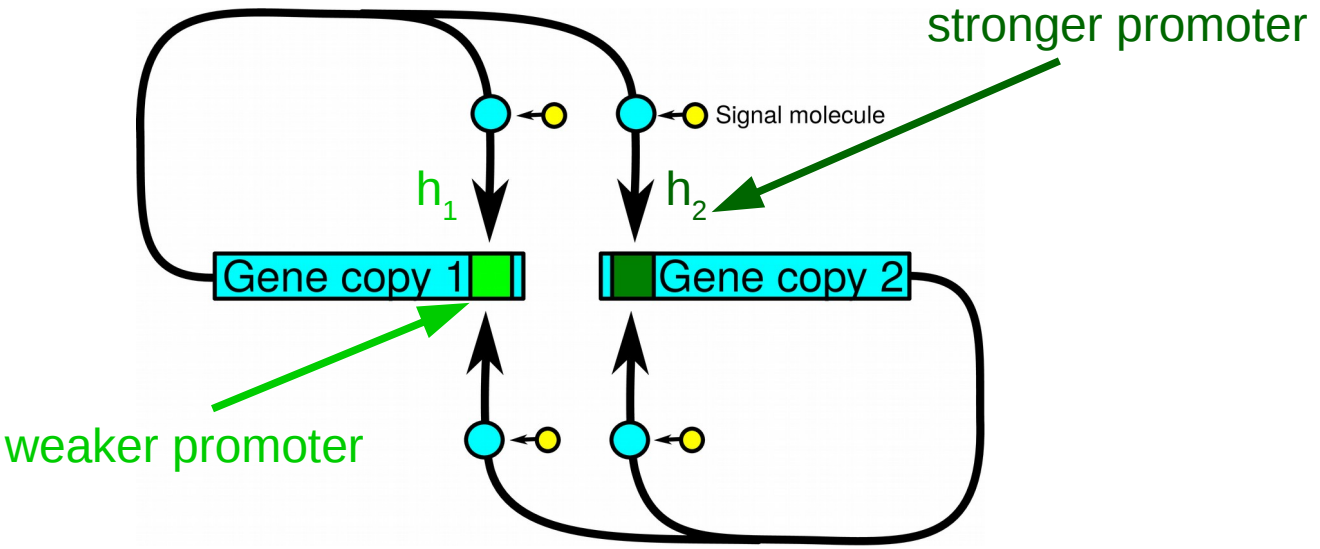
Two non-identical gene copies (imperfect duplication)





$$h_j(x) = (1 - \epsilon_j)H_j(x) + \epsilon_j$$

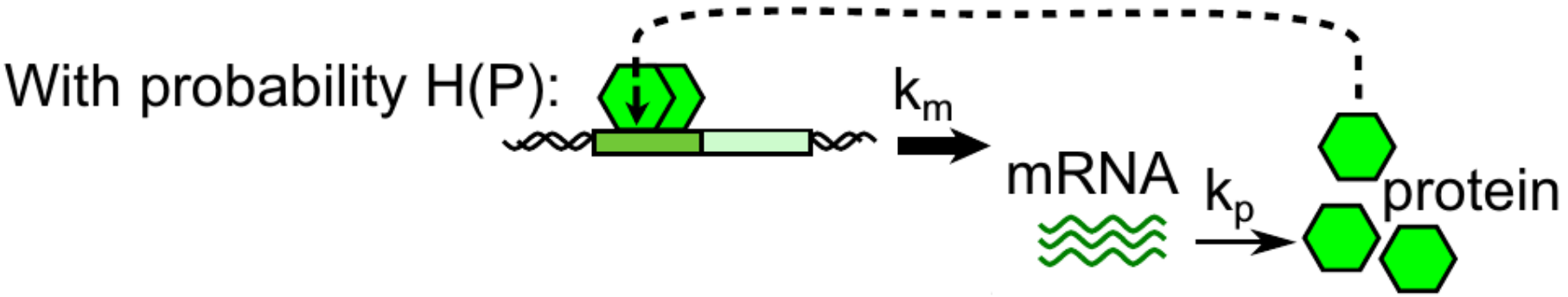
Two non-identical gene copies (imperfect duplication)



Probability density function for x proteins:

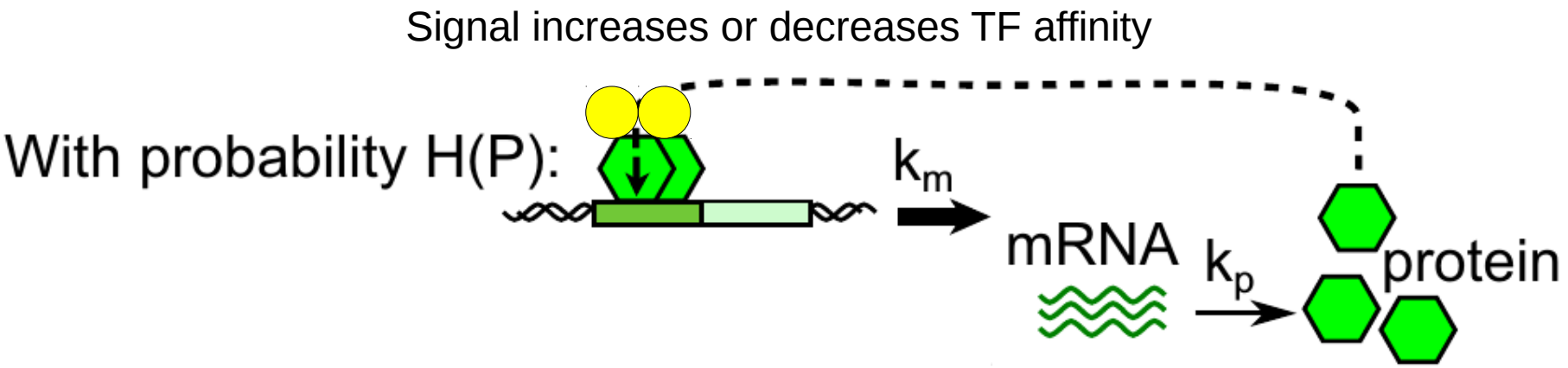
$$p(x) = Ae^{-x/b}x^{a_1+a_2-1}[H_1(x)]^{\frac{a_1(1-\epsilon_1)}{n_1}}[H_2(x)]^{\frac{a_2(1-\epsilon_2)}{n_2}}$$

Response to a signal



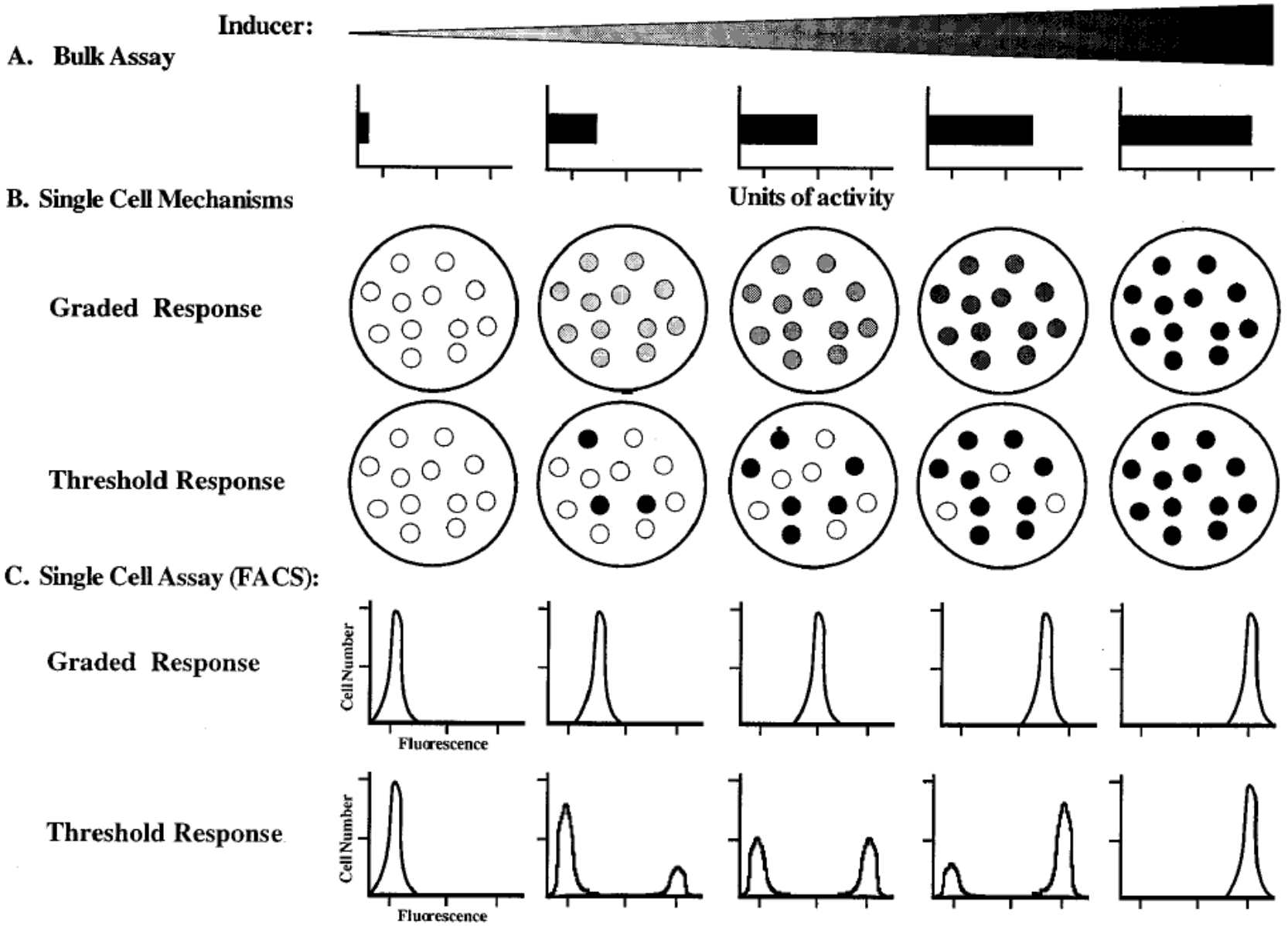
Ochab-Marcinek, Tabaka (2015) Phys Rev E, 91, 012704

Response to a signal



Ochab-Marcinek, Tabaka (2015) Phys Rev E, 91, 012704

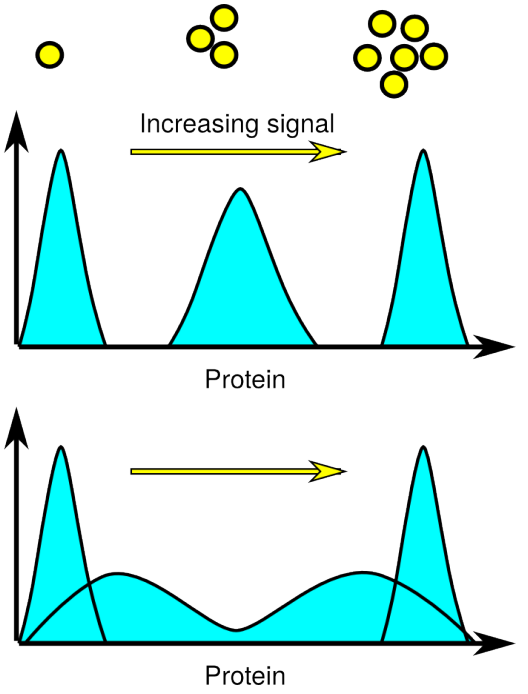
Binary / graded response



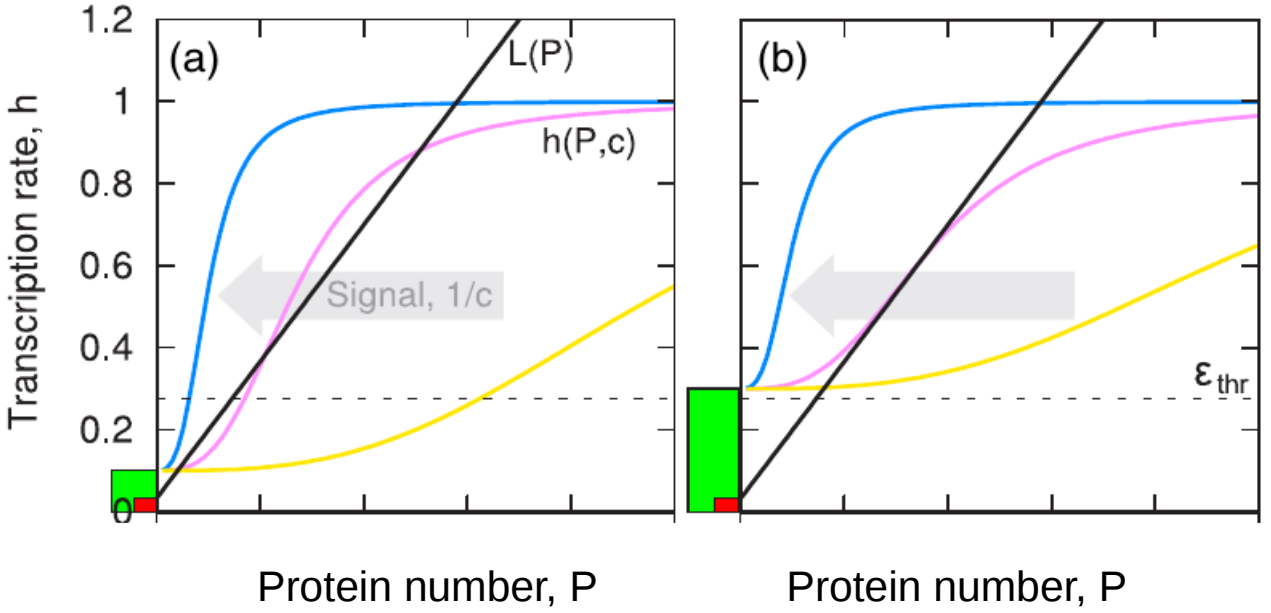
Kringstein, AM (1998) PNAS 95(23), 13670–5

Possible functions of binary/graded responses

- Graded: Precise
- Binary: Bet hedging



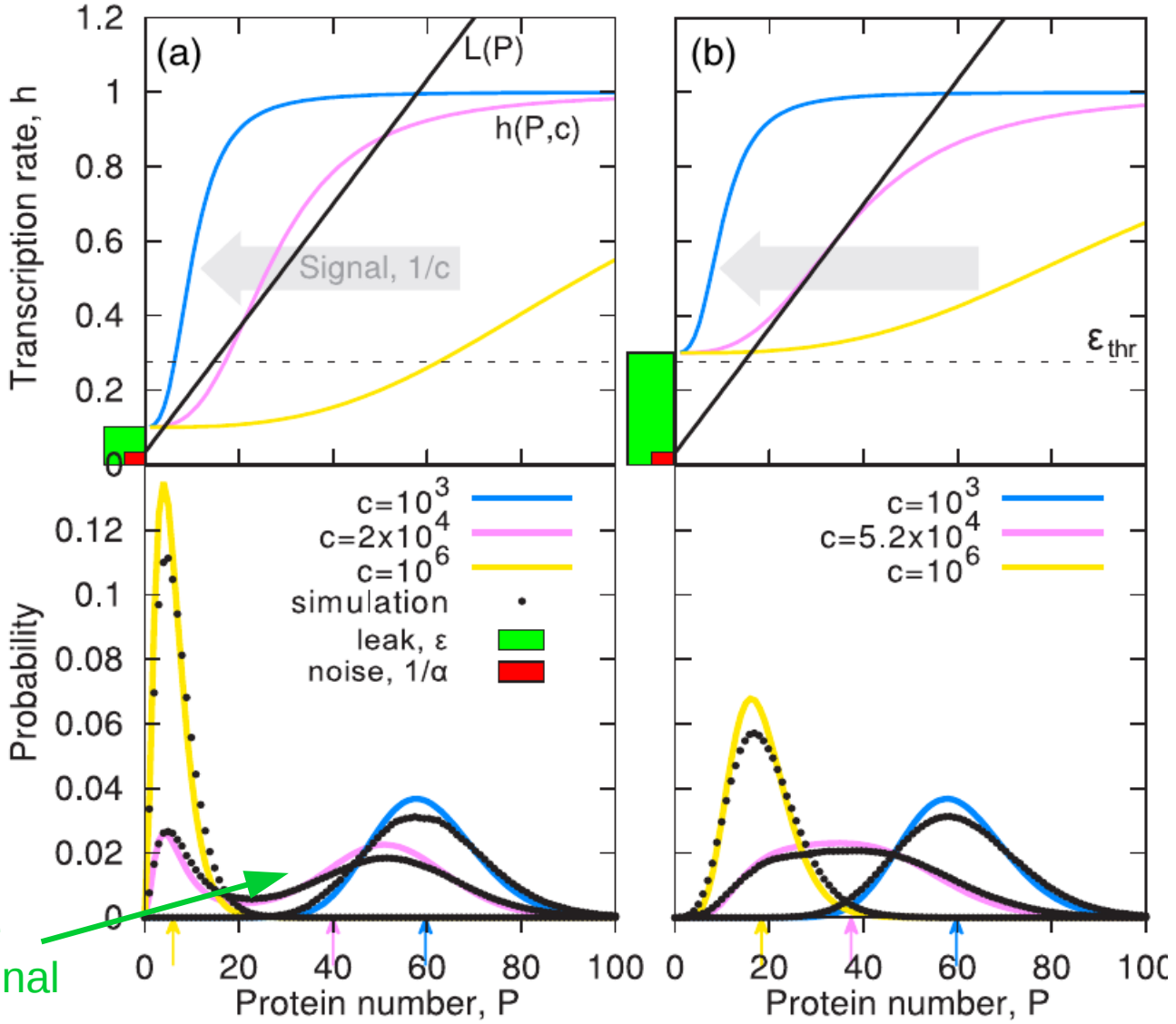
It can be easily shown that extrema of distribution (for a single gene) are given by geometric construction



$$H(P)(1-\epsilon)+\epsilon = \frac{1}{\alpha\beta}P + \frac{1}{\alpha}$$

Ochab-Marcinek, Tabaka (2015) Phys Rev E, 91, 012704

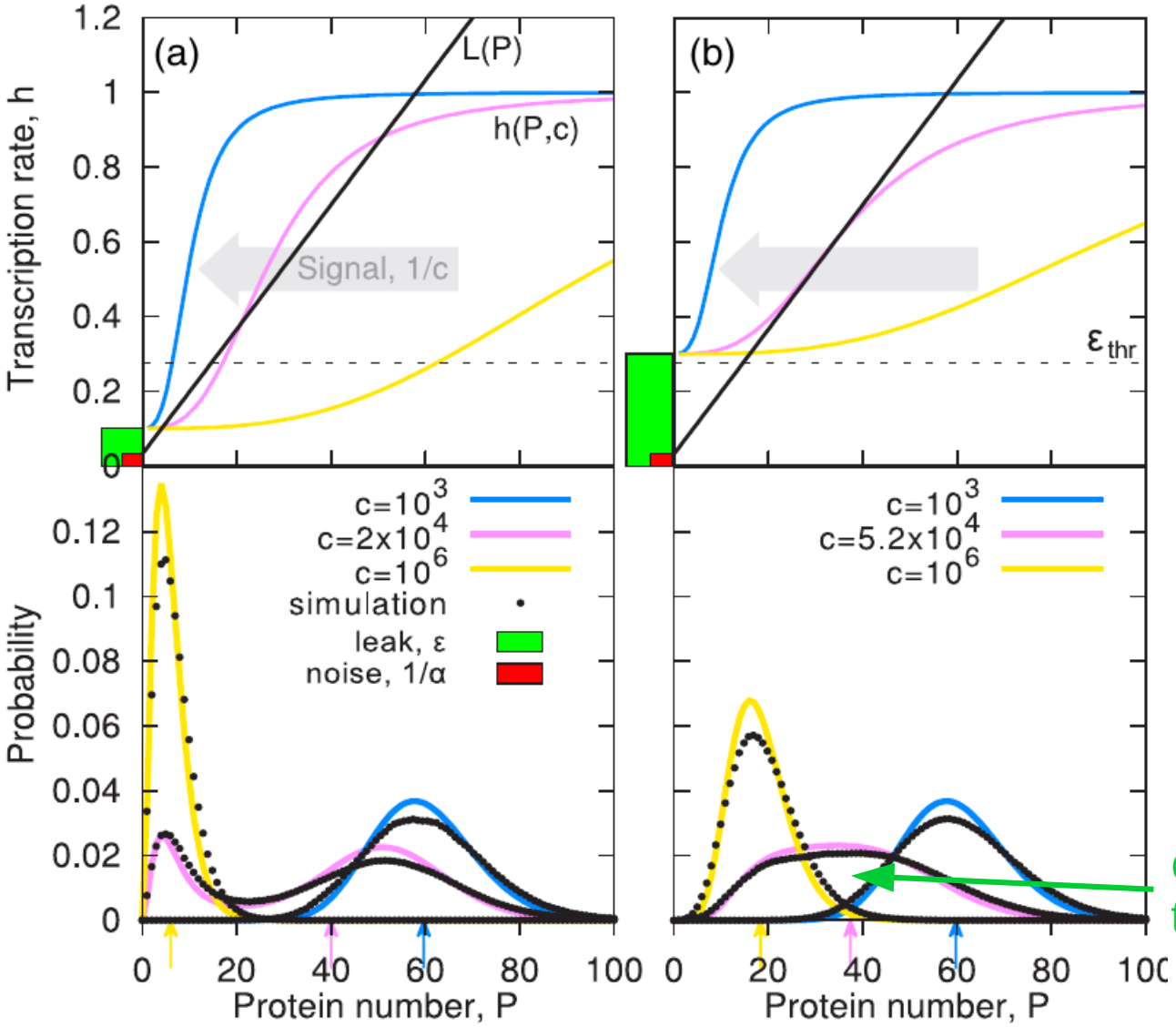
It can be easily shown that extrema of distribution (for a single gene) are given by geometric construction



Binary response to increasing signal

Ochab-Marcinek, Tabaka (2015) Phys Rev E, 91, 012704

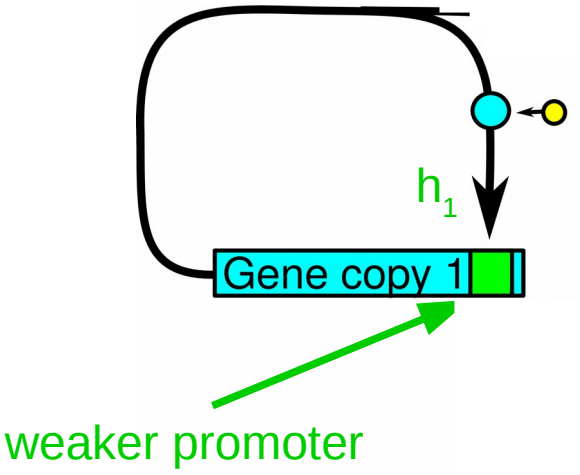
It can be easily shown that extrema of distribution (for a single gene) are given by geometric construction



Graded response to increasing signal

Ochab-Marcinek, Tabaka (2015) Phys Rev E, 91, 012704

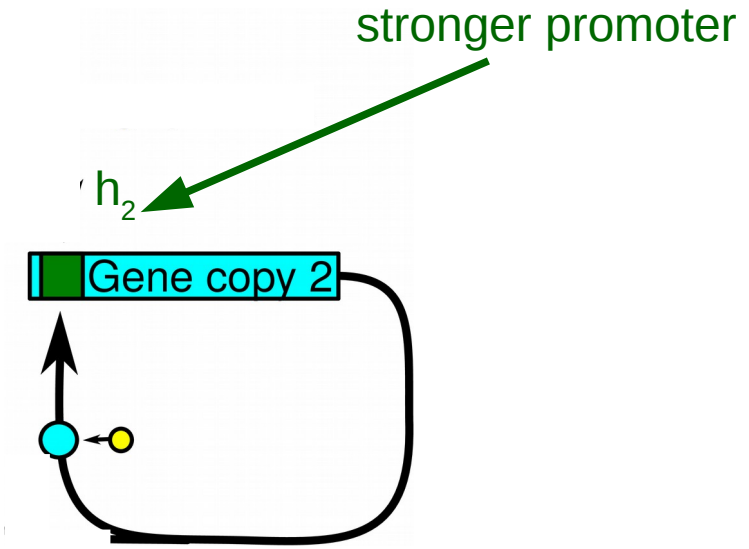
Two non-identical gene copies (imperfect duplication)



For a single gene:

$$\frac{1}{ab}x + \frac{1}{a} \quad = \quad h_1(x)$$

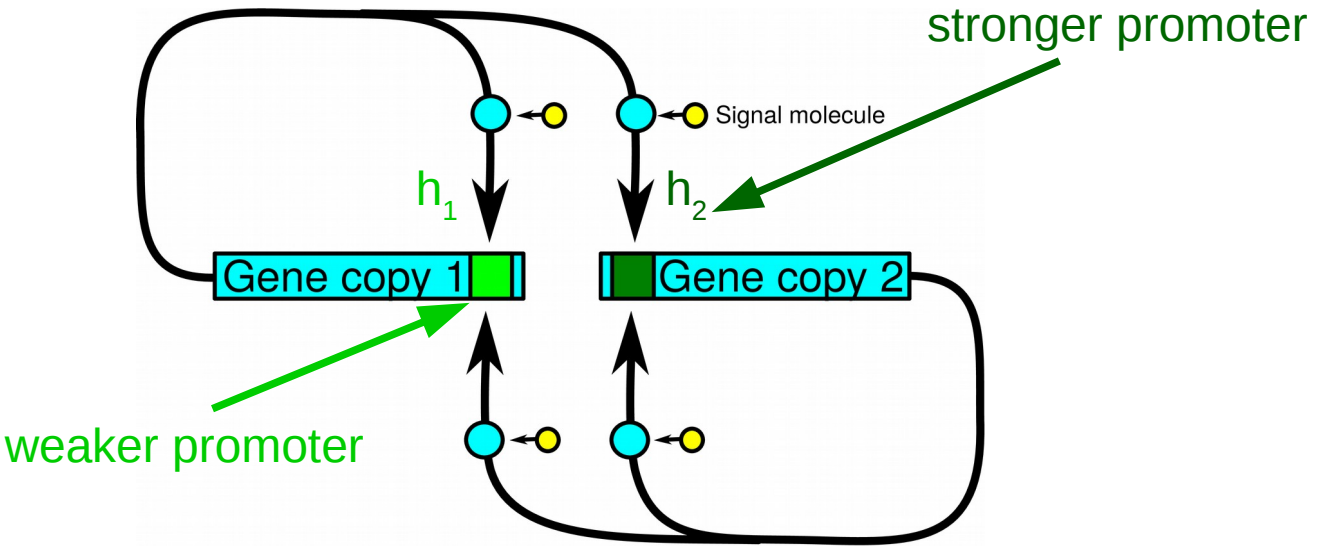
Two non-identical gene copies (imperfect duplication)



For a single gene:

$$\frac{1}{ab}x + \frac{1}{a} = h_2(x)$$

Two non-identical gene copies (imperfect duplication)



Probability density function for x proteins:

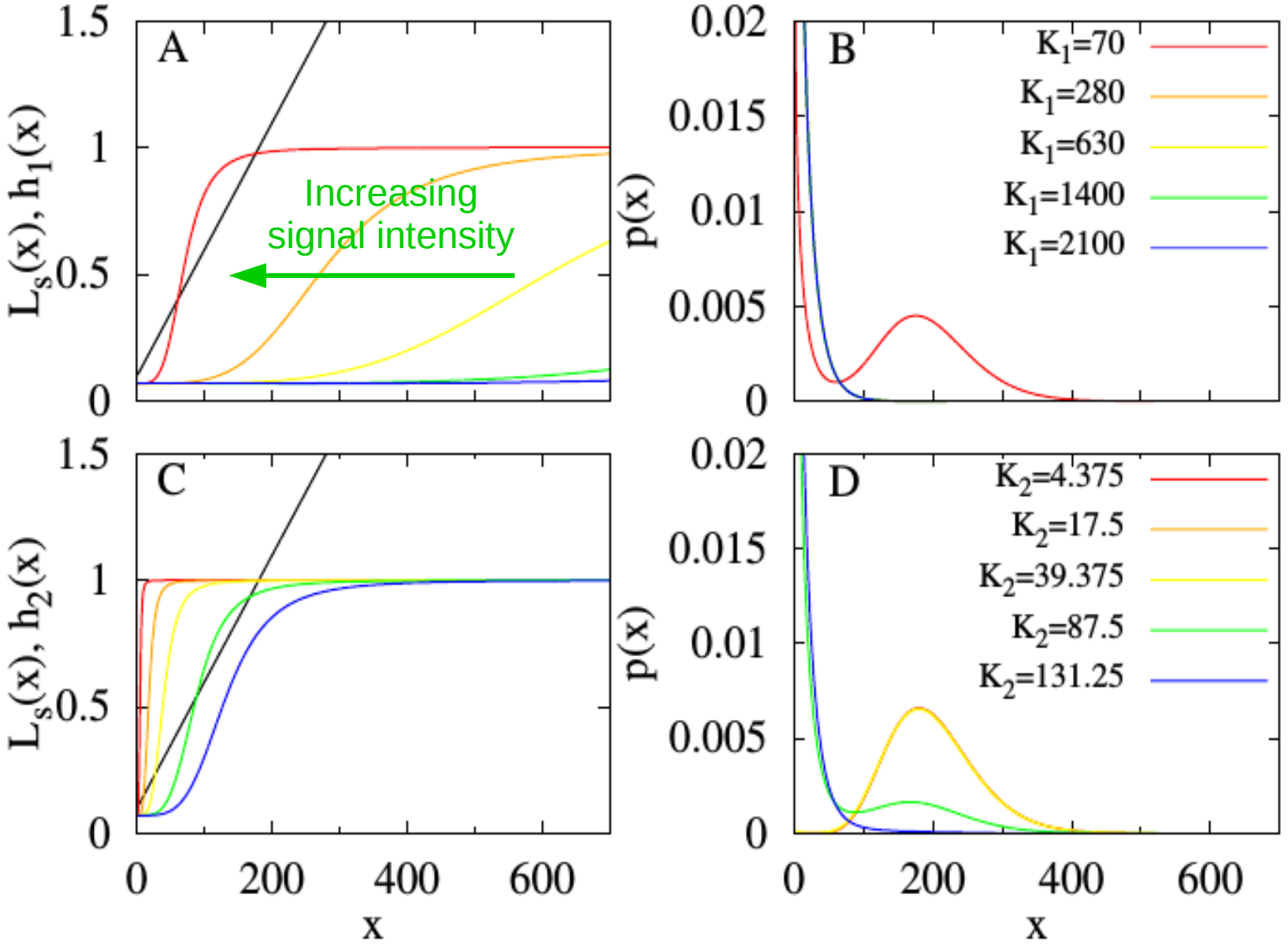
$$p(x) = Ae^{-x/b}x^{a_1+a_2-1}[H_1(x)]^{\frac{a_1(1-\epsilon_1)}{n_1}}[H_2(x)]^{\frac{a_2(1-\epsilon_2)}{n_2}}$$

When a_1=a_2=a, its extrema are given by the geometric construction:

$$\frac{1}{ab}x + \frac{1}{a} = h_1(x) + h_2(x)$$

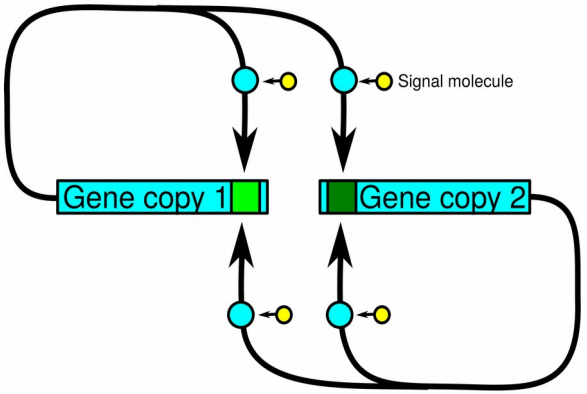
weaker promoter

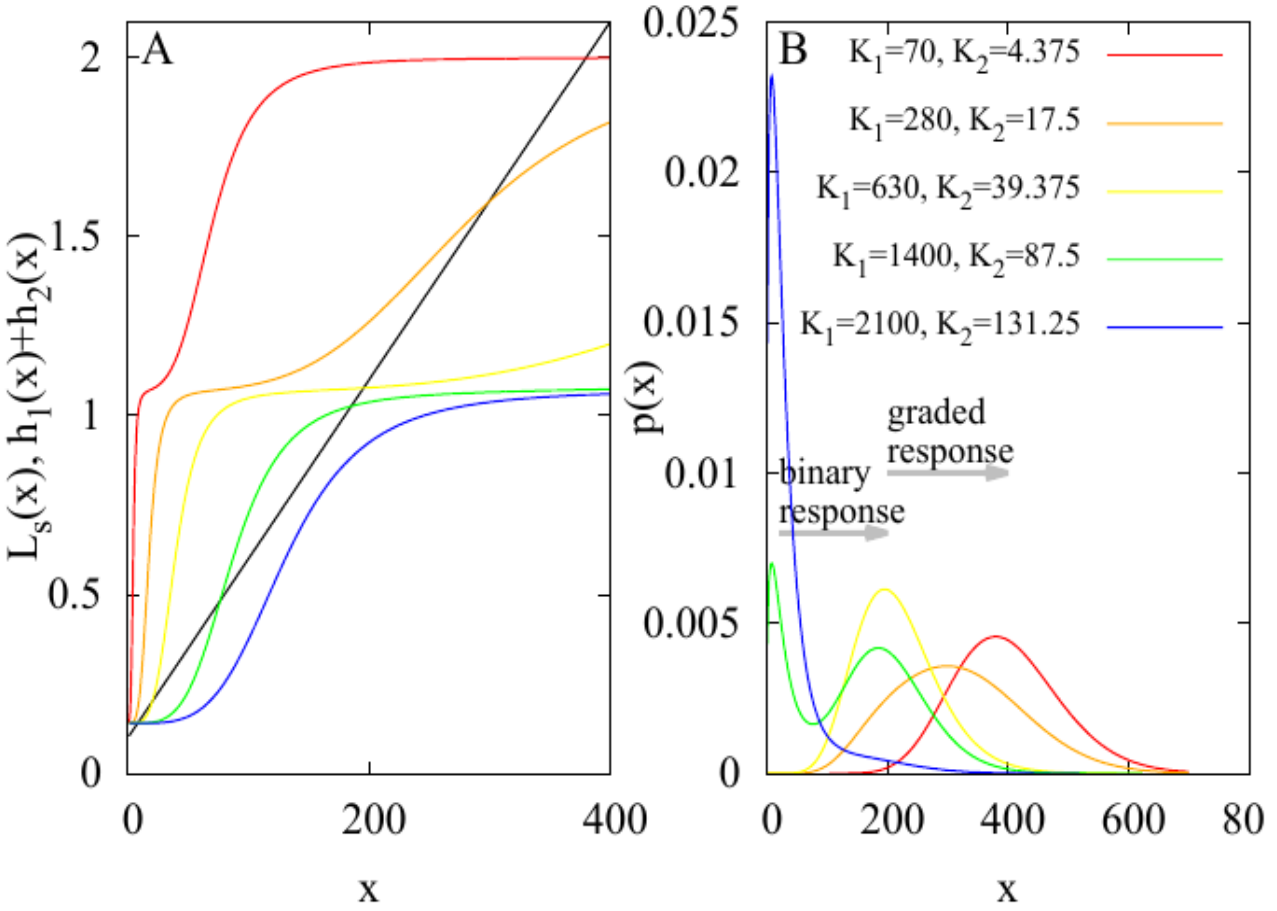
stronger promoter



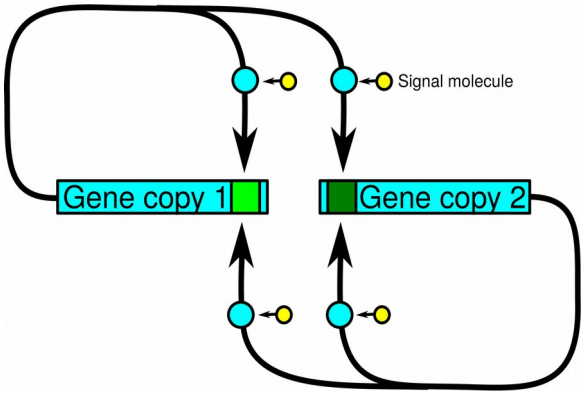
Both genes, taken separately, show a binary response

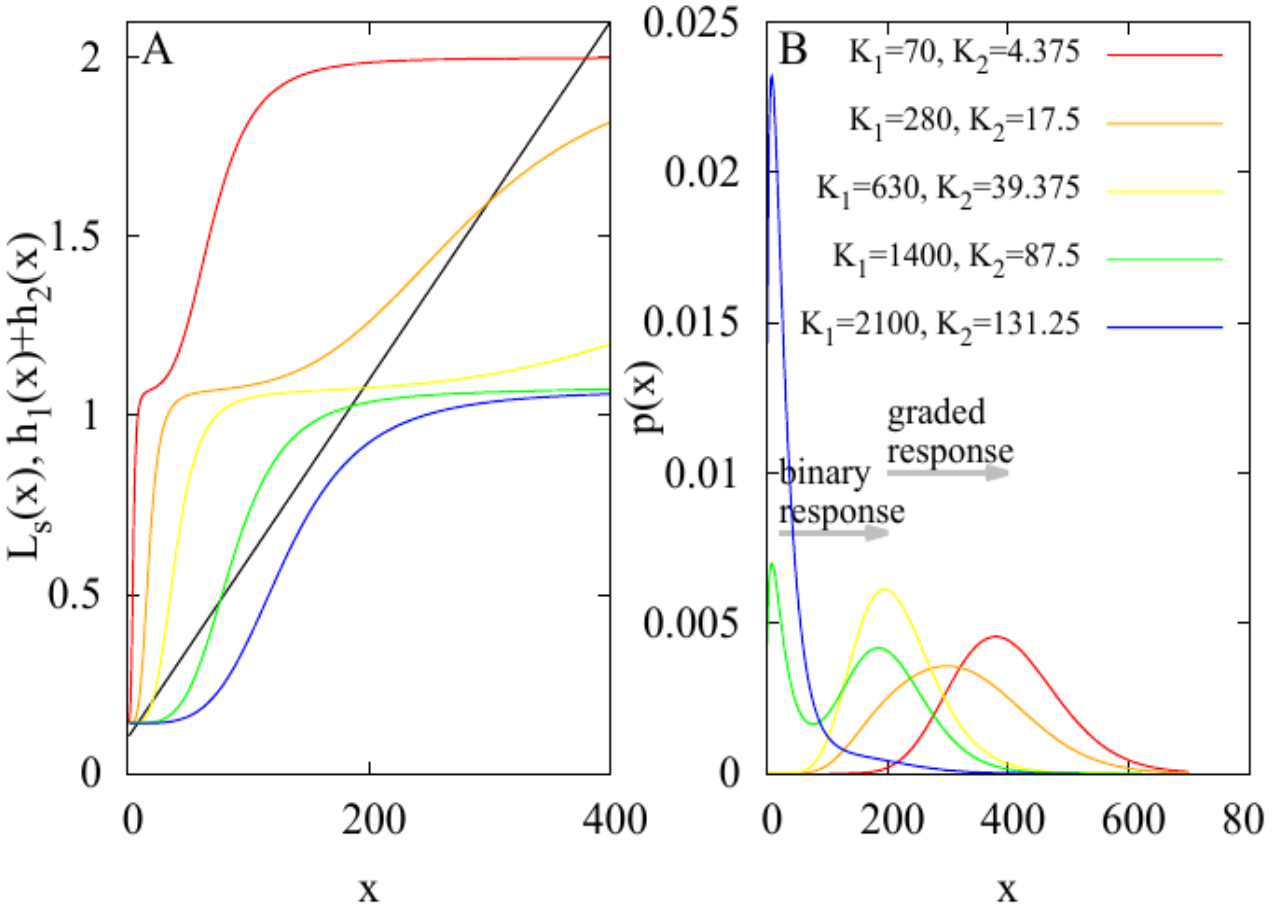
Two gene copies together: Mixed, binary+graded response





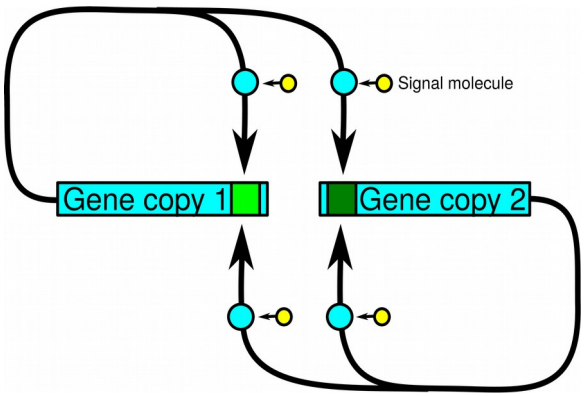
Two gene copies together: Mixed, binary+graded response





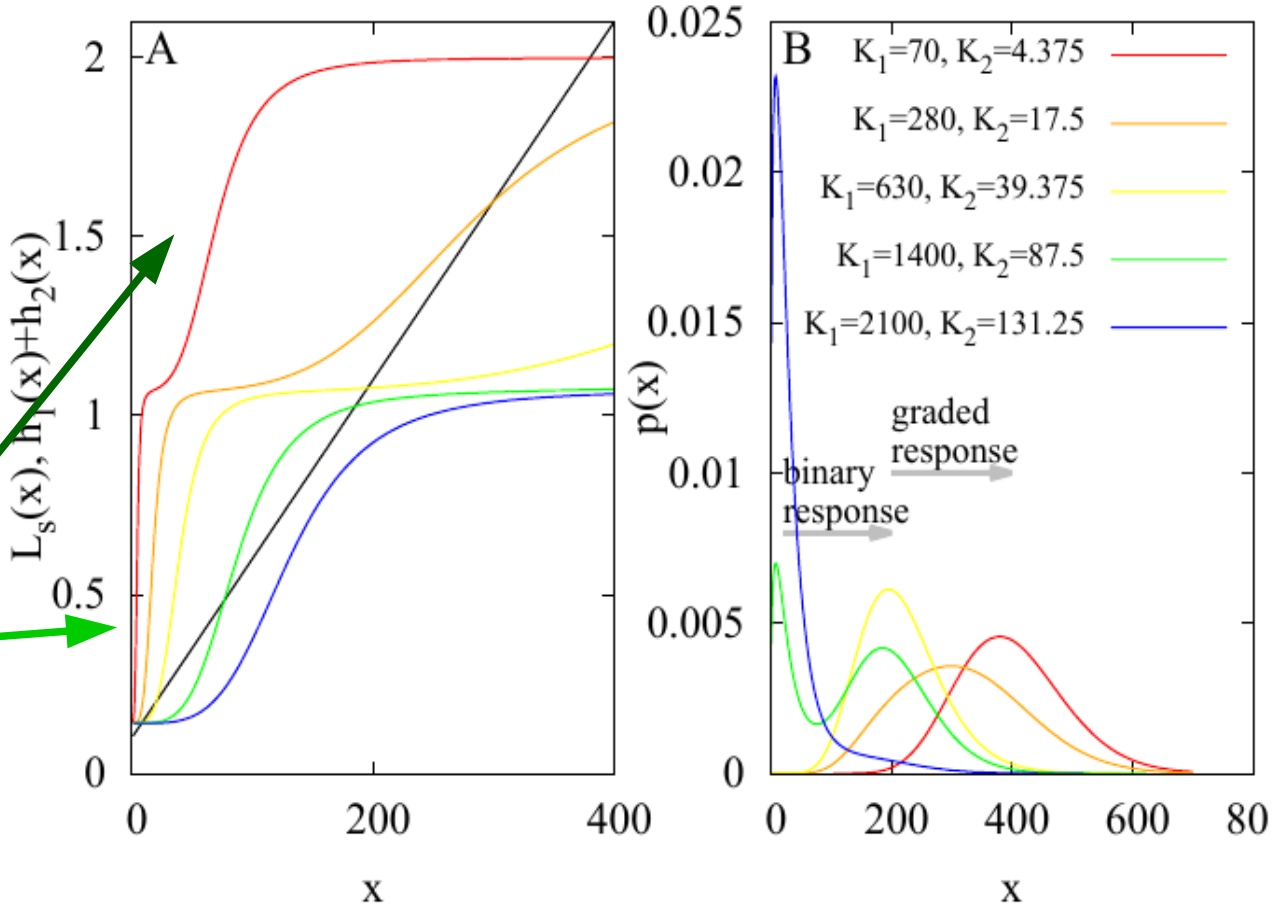
$$\frac{1}{ab}x + \frac{1}{a} = h_1(x) + h_2(x)$$

Two gene copies together: Mixed, binary+graded response



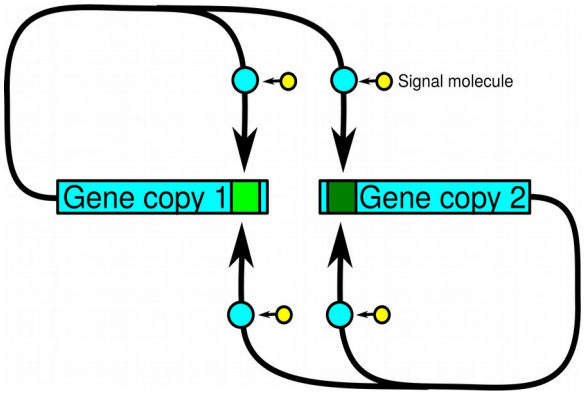
stronger promoter

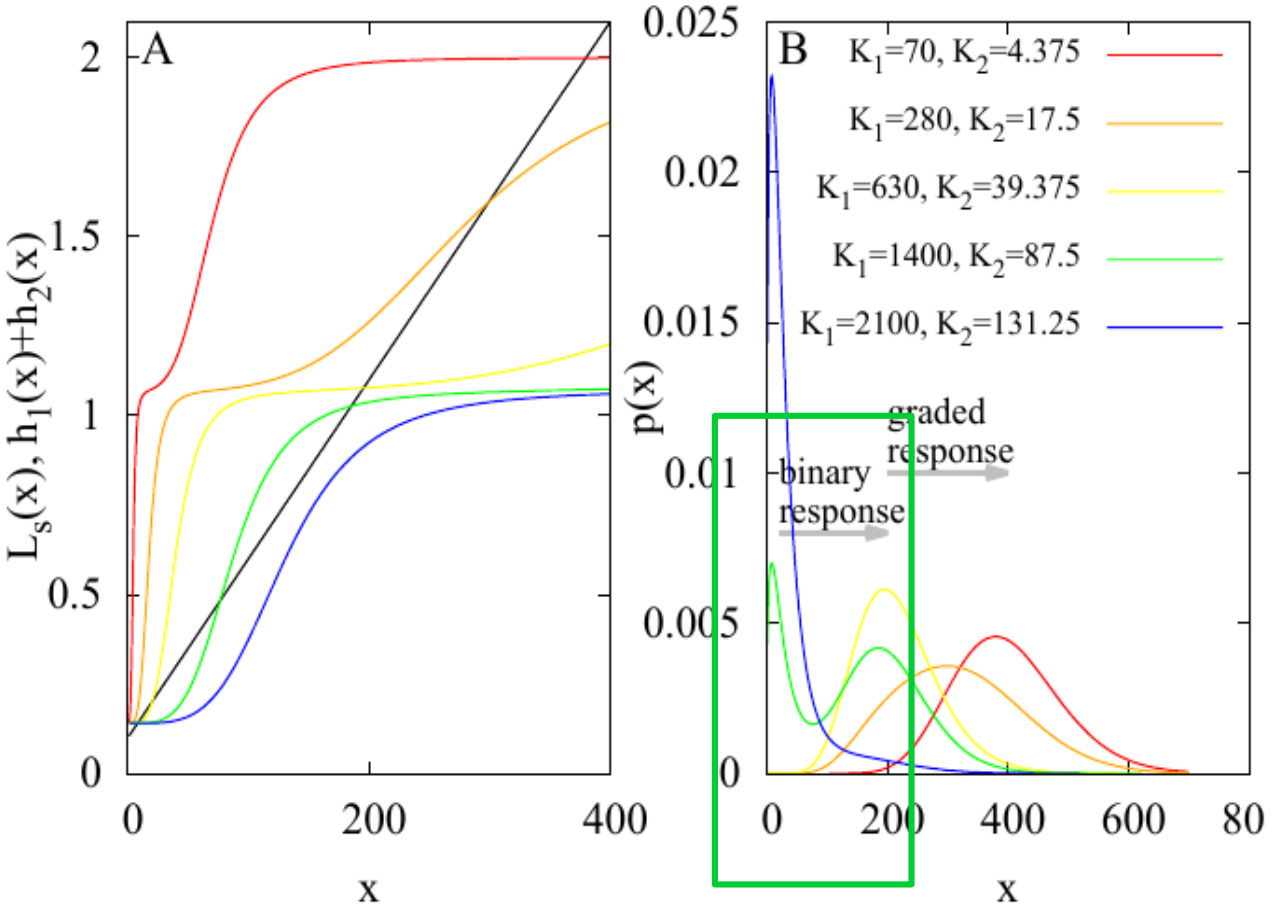
weaker promoter



$$\frac{1}{ab}x + \frac{1}{a} = h_1(x) + h_2(x)$$

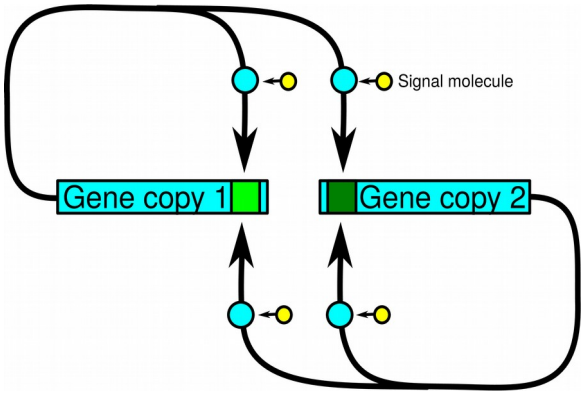
Two gene copies together: Mixed, binary+graded response

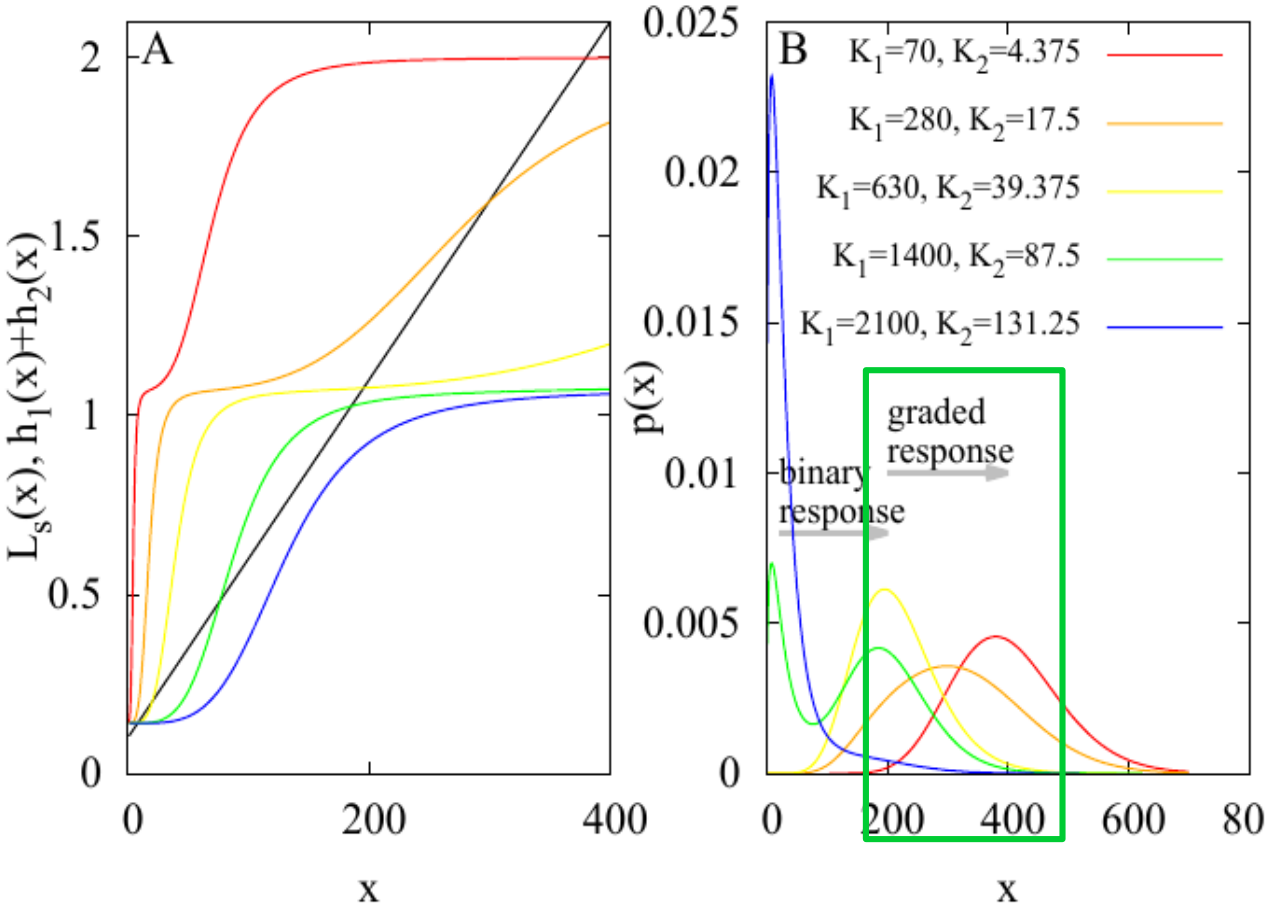




$$\frac{1}{ab}x + \frac{1}{a} = h_1(x) + h_2(x)$$

Two gene copies together: Mixed, binary+graded response





$$\frac{1}{ab}x + \frac{1}{a} = h_1(x) + h_2(x)$$

Result 3

Imperfect gene duplication may lead to
mixed, binary+graded response
of the gene system to a signal

Result 4

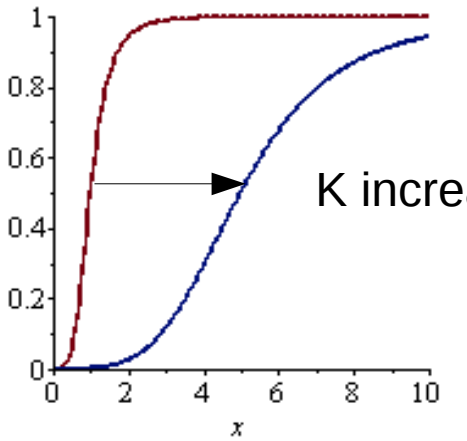
Accumulation of gene duplication depends on the fold-change of mean gene expression before and after duplication.
The fold-change depends non-monotonically on gene's noisiness

Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

- $1/K_1$, $1/K_2$: regulation strength

$$h_j(x) = (1 - \epsilon_j) H_j(x) + \epsilon_j$$

$$H_j(x) = \left[1 + \left(\frac{x}{K_j} \right)^{n_j} \right]^{-1}$$



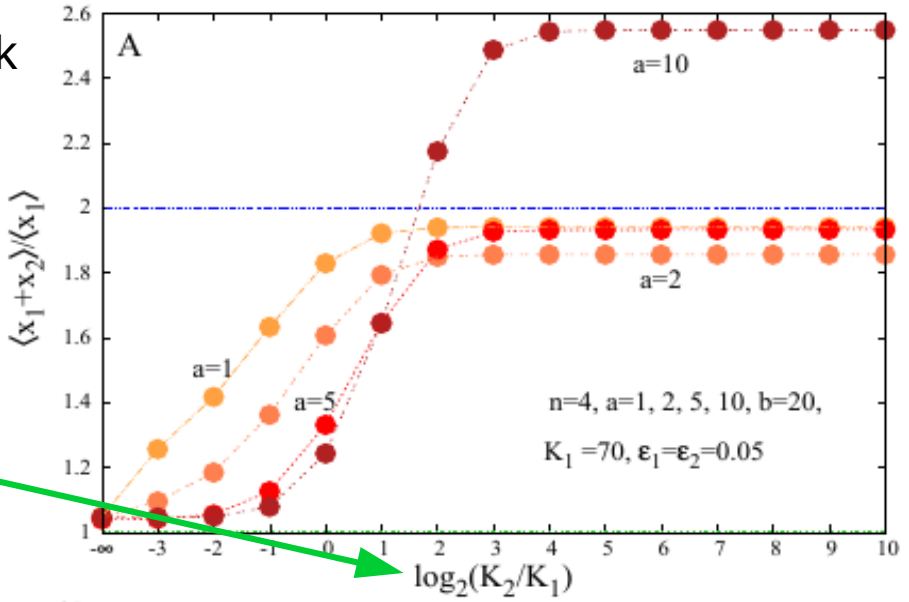
K increasing: weaker regulation

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

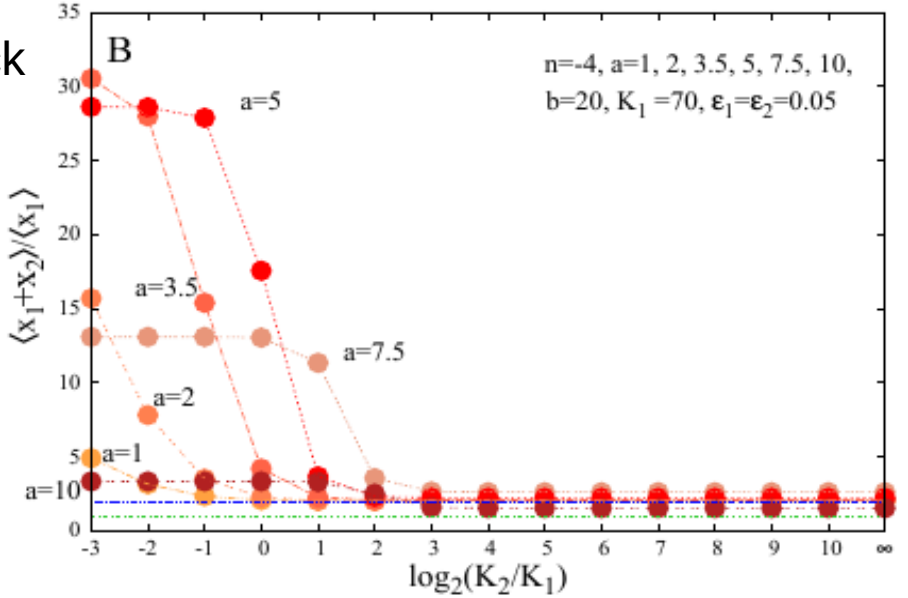
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

Negative feedback

We compare strengths of the two promoters: K_2/K_1



Positive feedback

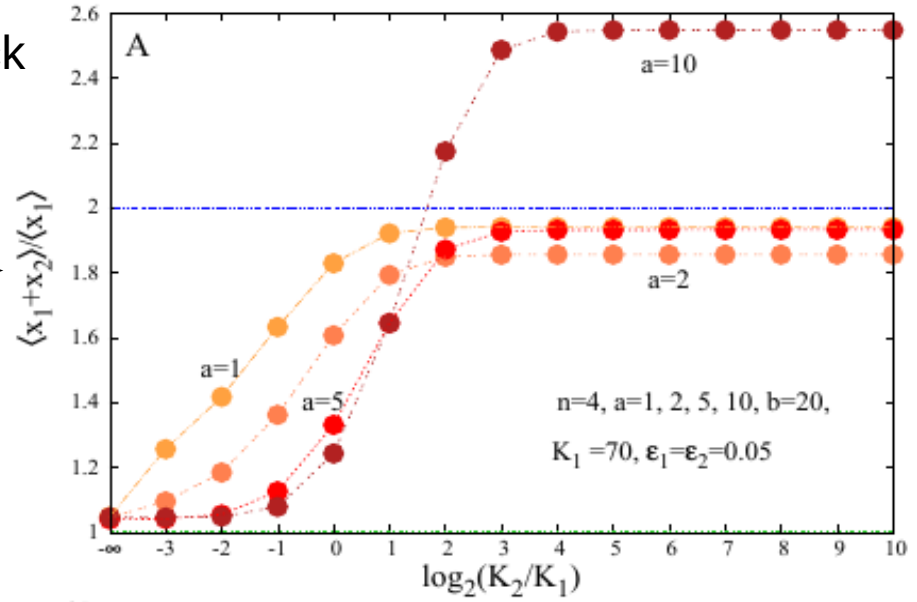


J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

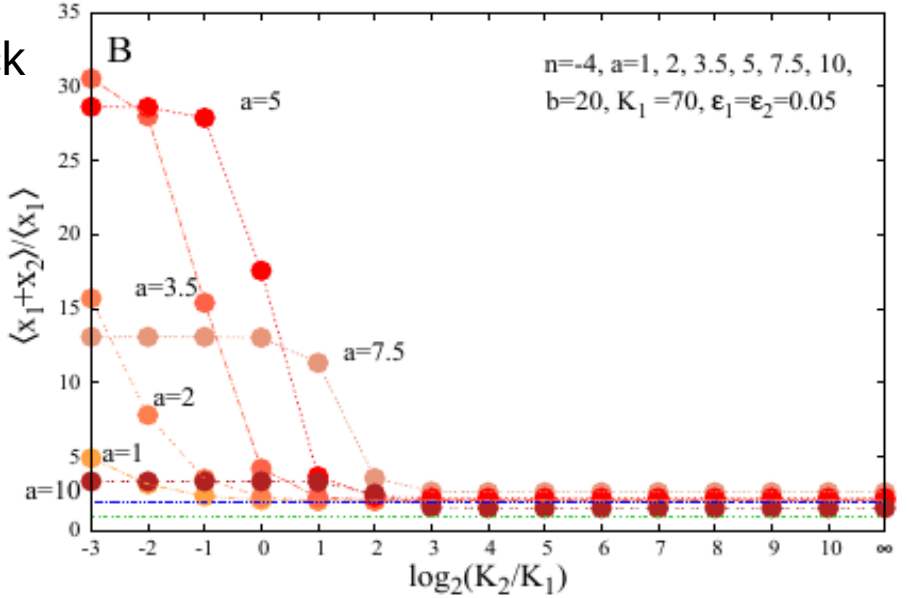
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

Negative feedback

Mean protein concentration:
After gene duplication / before duplication



Positive feedback

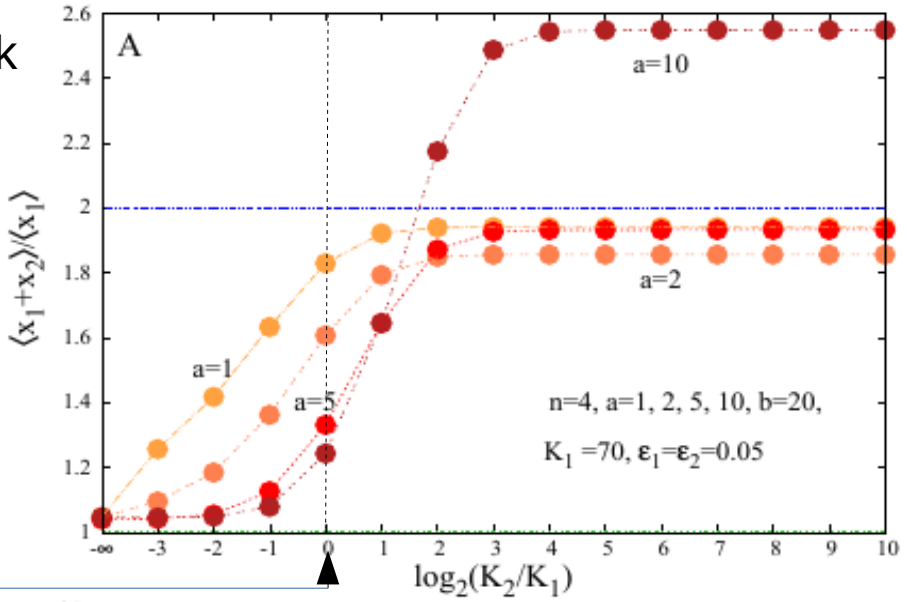


J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

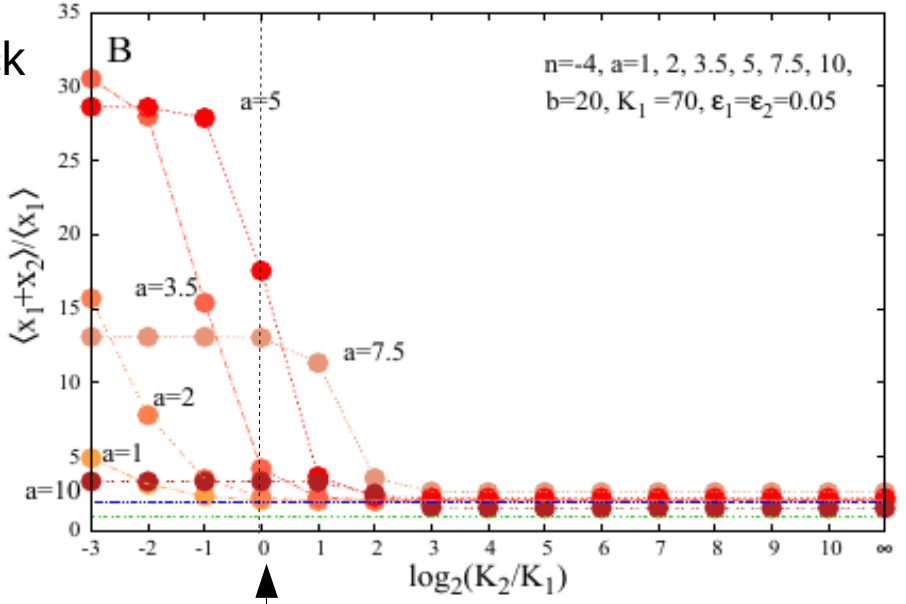
Negative feedback

Mean protein concentration:
After gene duplication / before duplication



Positive feedback

Identical gene copies



J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

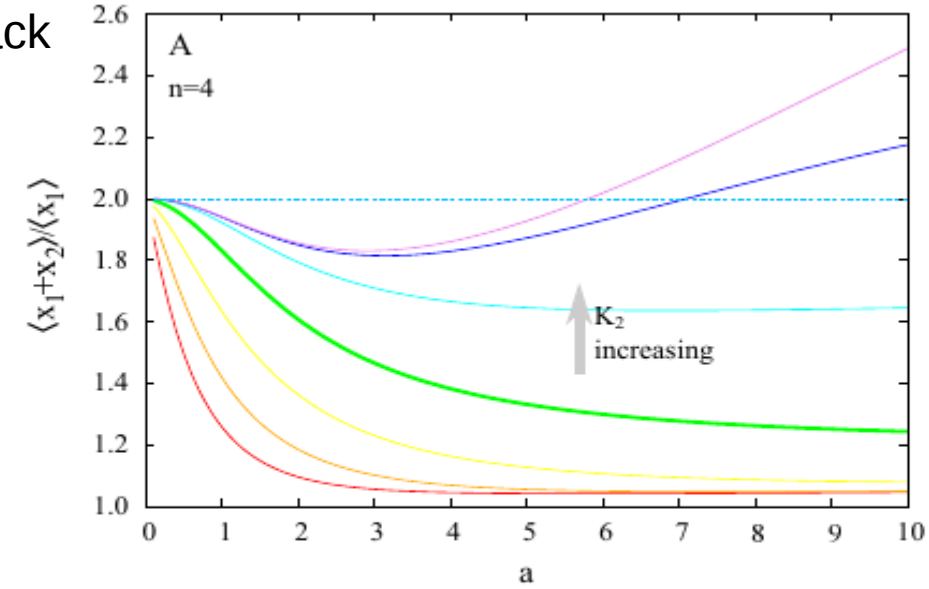
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

- After duplication, gene expression is not necessarily twice as high as before duplication
- Large fold-change may be detrimental
- Mutants with small changes have a greater chance to survive
- The relative change depends on the burst frequency, *a*, in a non-monotonic manner

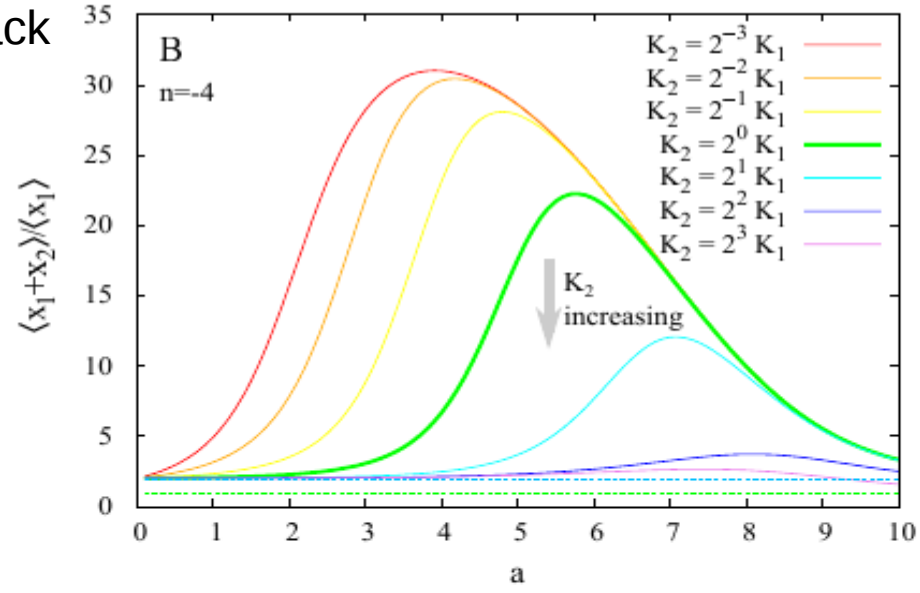
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

Negative feedback

Mean protein concentration:
After gene duplication / before duplication



Positive feedback

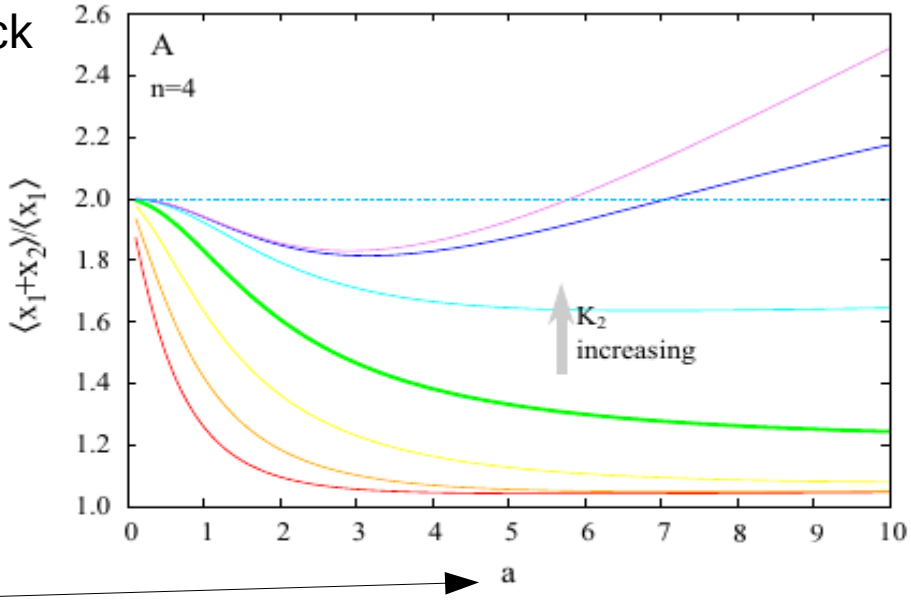


J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

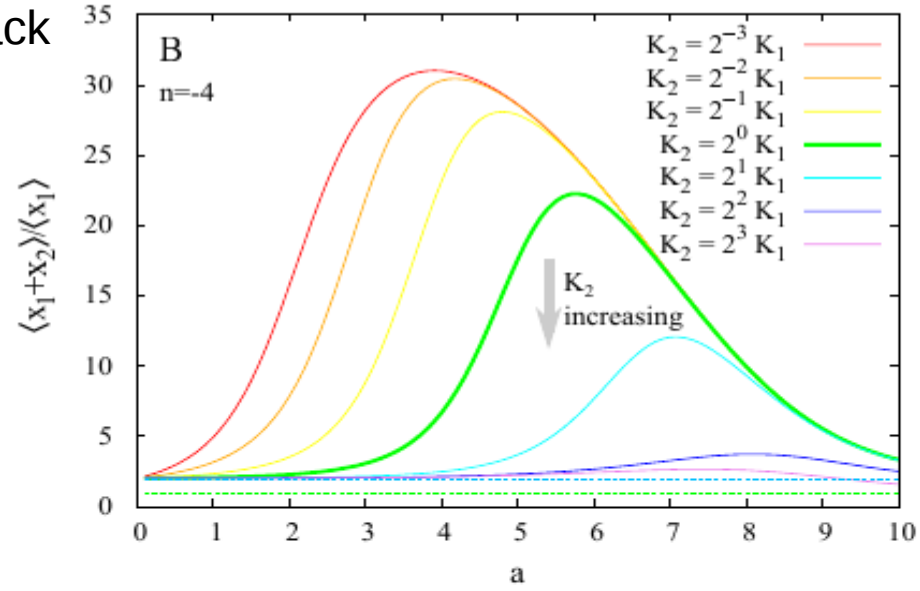
Negative feedback

Mean protein concentration:
After gene duplication / before duplication



Maximal
mean burst frequency,
measure of noisiness

Positive feedback



J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

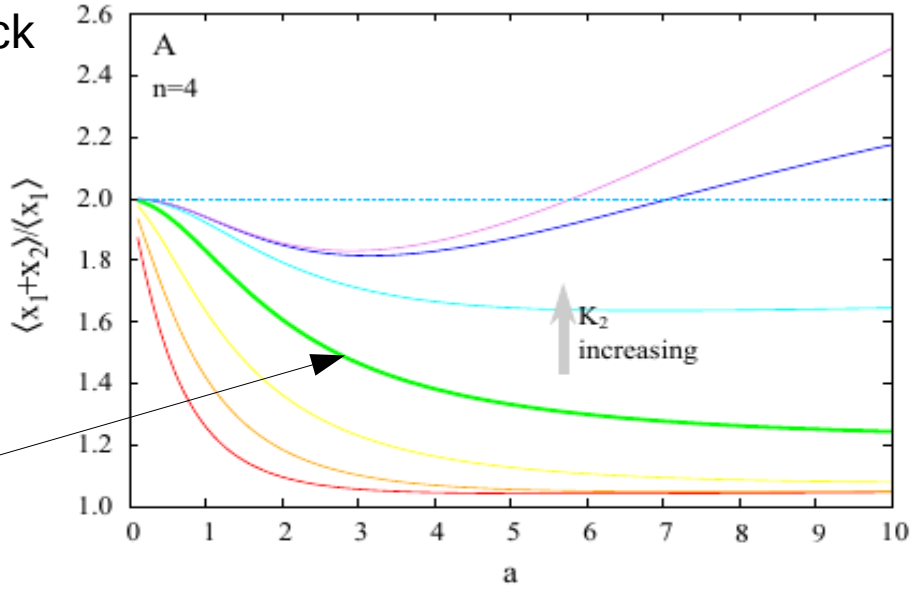
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

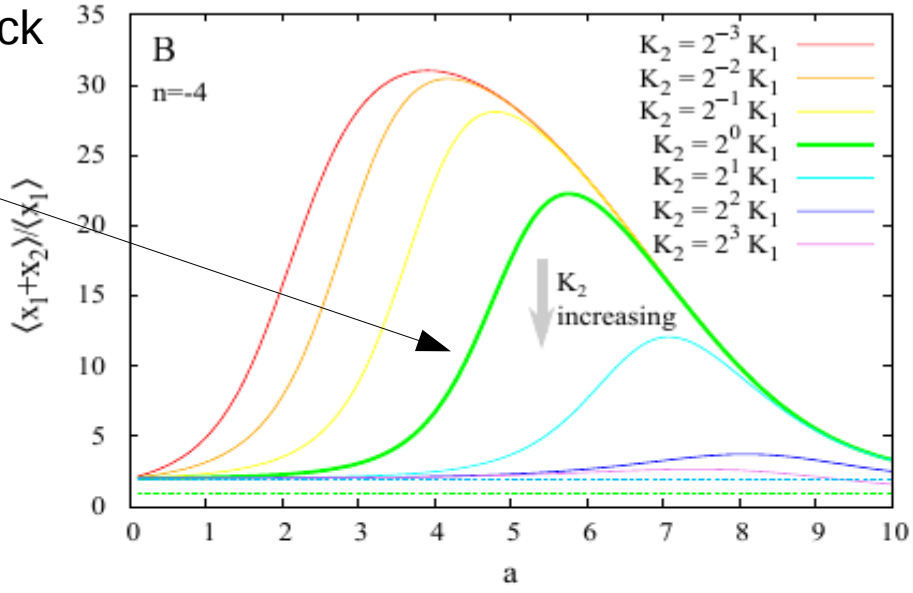
Negative feedback

Mean protein concentration:
After gene duplication / before duplication

Identical gene copies

Positive feedback





J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

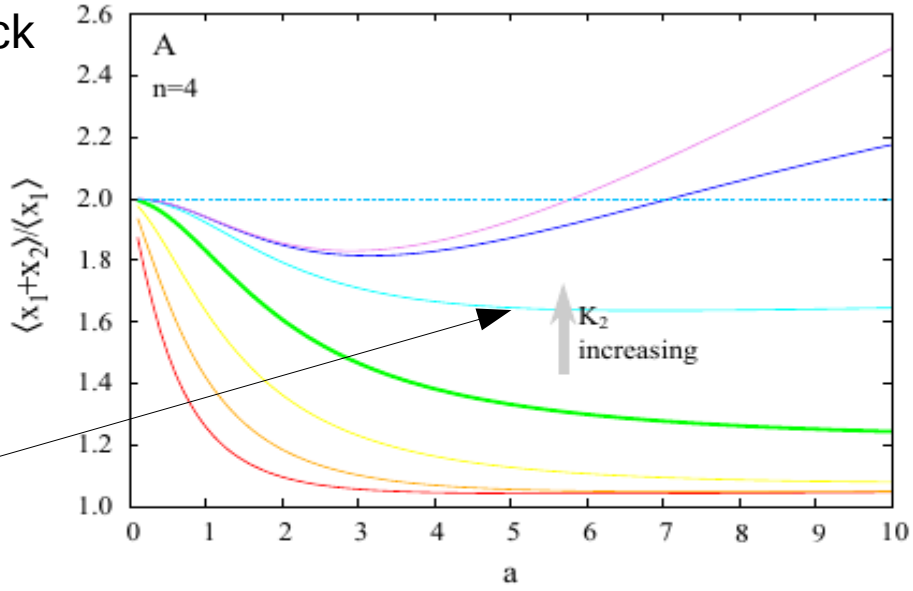
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

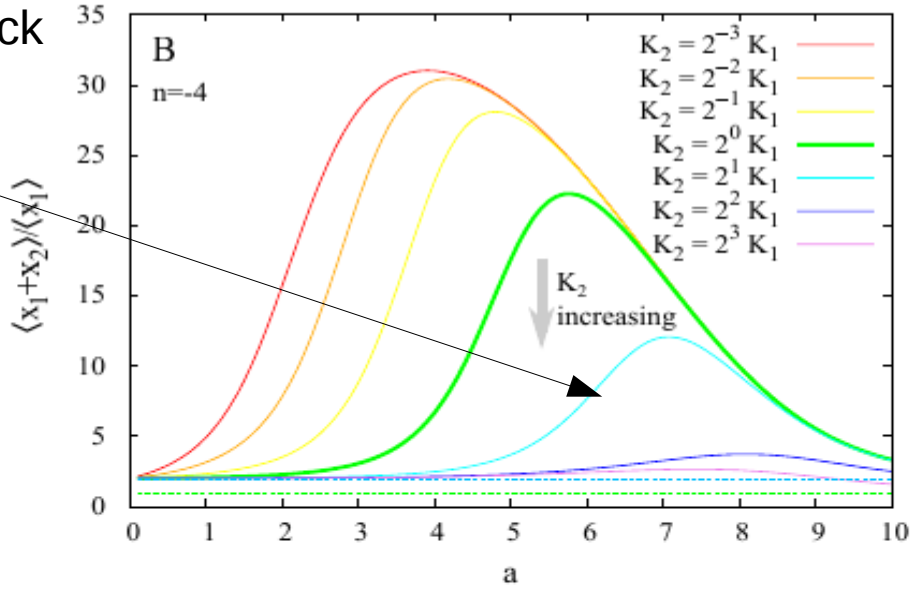
Negative feedback

Mean protein concentration:
After gene duplication / before duplication

Promoter of the new gene is weaker than the old one

Positive feedback





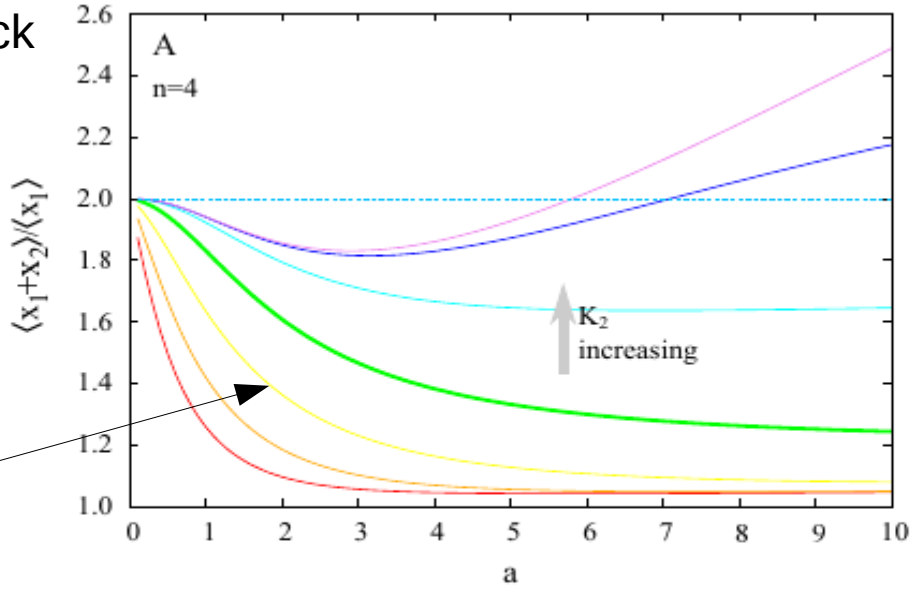
J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

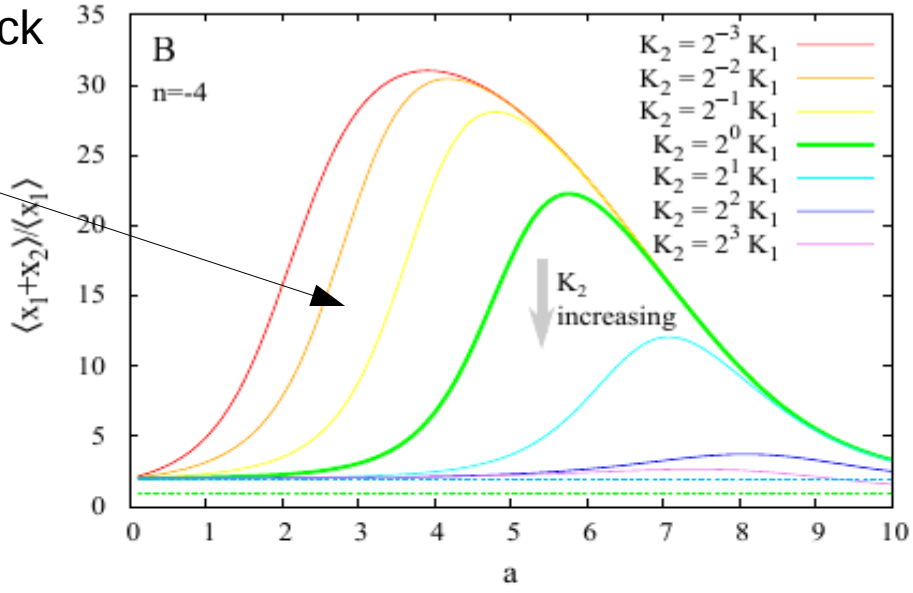
Negative feedback

Mean protein concentration:
After gene duplication / before duplication

Promoter of the new gene is stronger than the old one



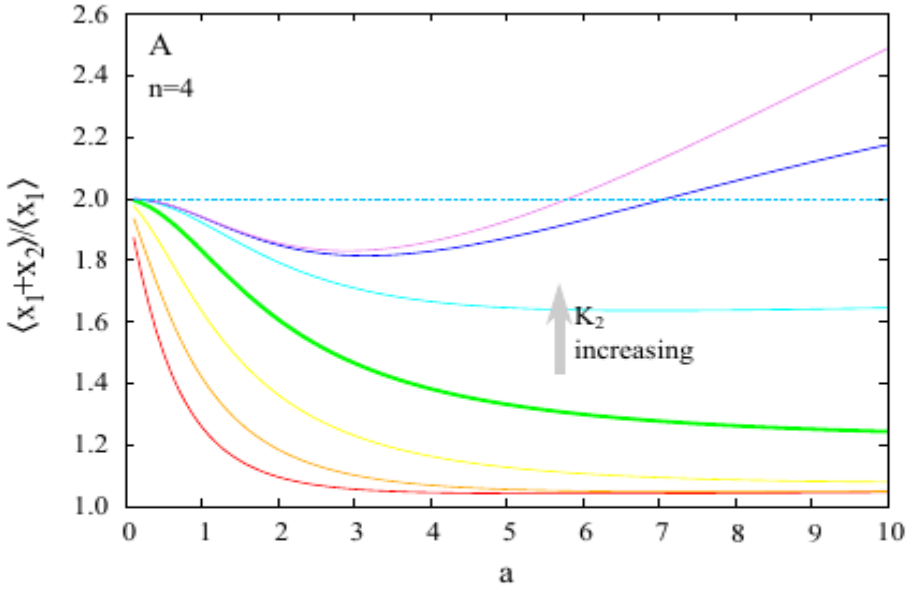
Positive feedback

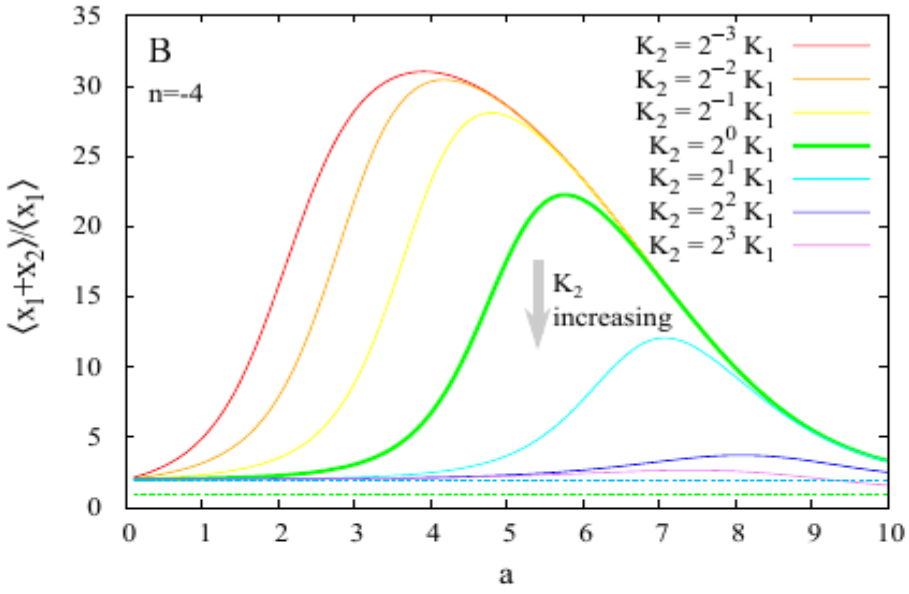


J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

Mean protein concentration:
After gene duplication / before duplication





If large change in gene expression is detrimental, then the gene duplication can survive as long as it introduces a small change.

Mutations introducing small changes are more probable to be accumulated.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

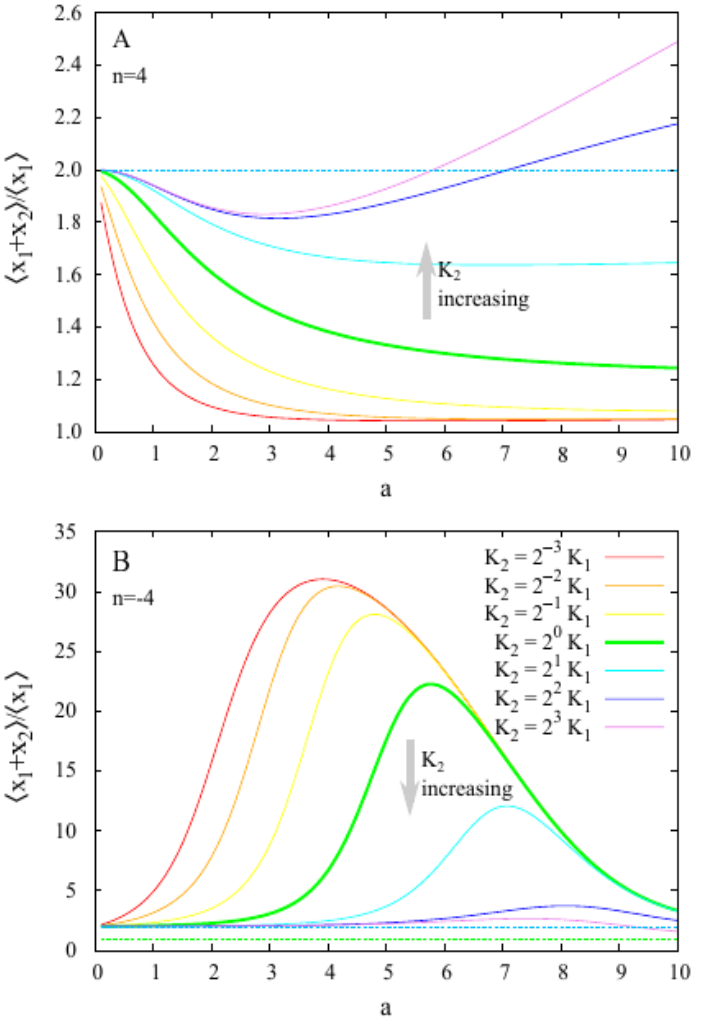
Relative change in average protein number due to gene duplication, depending on how the genes differ in the regulation strength

Negative autoregulation:

- Gene 2 has a stronger regulation than Gene 1 : Accumulation of gene duplications is more probable at larger *a* (smaller noise)
- Gene 2 has a weaker regulation than Gene 1 : Accumulation of gene duplications may be most probable at an optimal *a*

Positive autoregulation:

- Intermediate values of *a* give the greatest change in gene expression (one gene is uninduced, but two genes induce each other)
- Accumulation of gene duplications more probable for very small or very large *a*



Relative change in the average protein concentration before and after gene duplication, as a function of maximum mean burst frequency a, for various values of K_2/K_1. A: Negative auto-regulation, $n = 4$. B: Positive auto-regulation, $n = -4$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Result 4

Accumulation of gene duplication depends on the fold-change of mean gene expression before and after duplication.
The fold-change depends non-monotonically on gene's noisiness

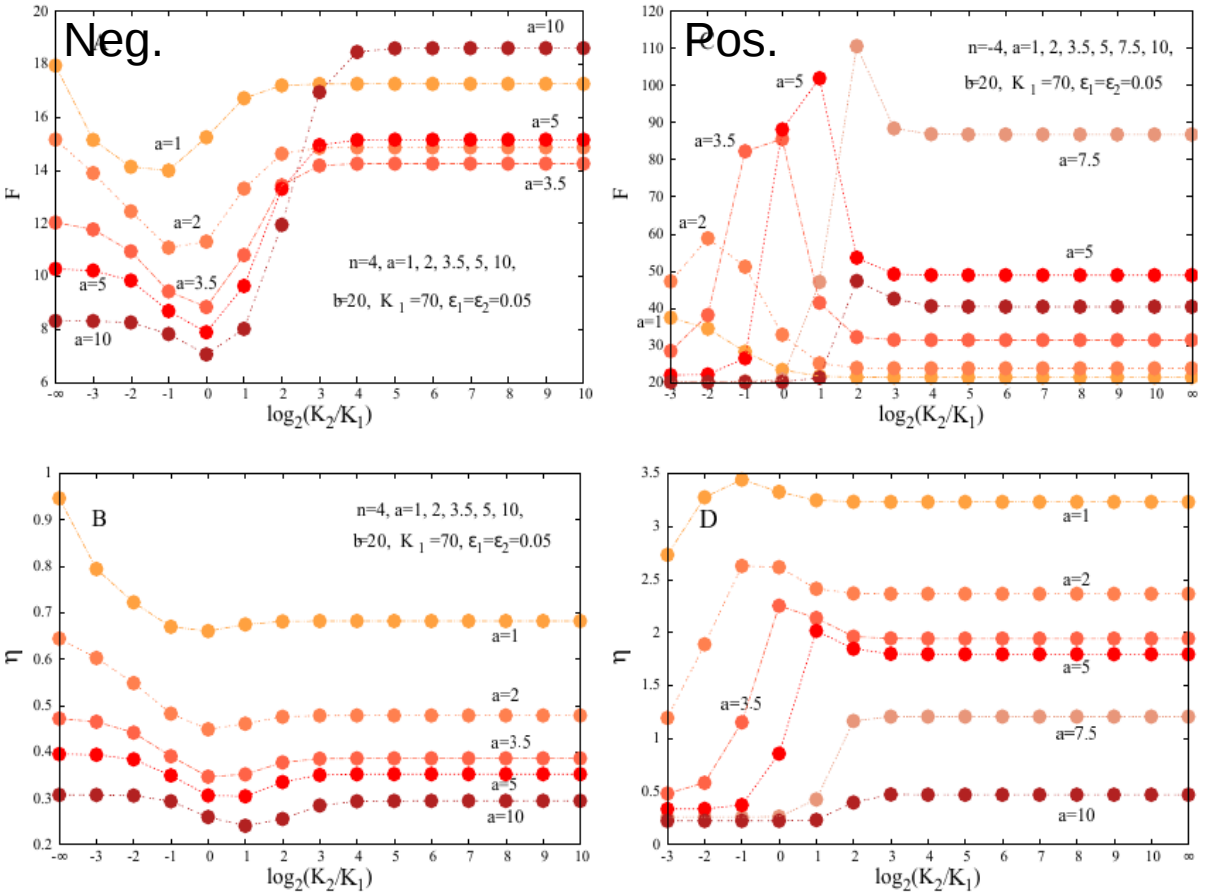
Result 5

Fano factor and CV vary differently.
Therefore, it does not make sense to say that
„evolution optimizes a gene system
to minimize noise”
(unless we really know
how evolution measures noise)

Fano factor and CV vary in a different manner as the relative difference in regulation strengths is varied

Fano factor

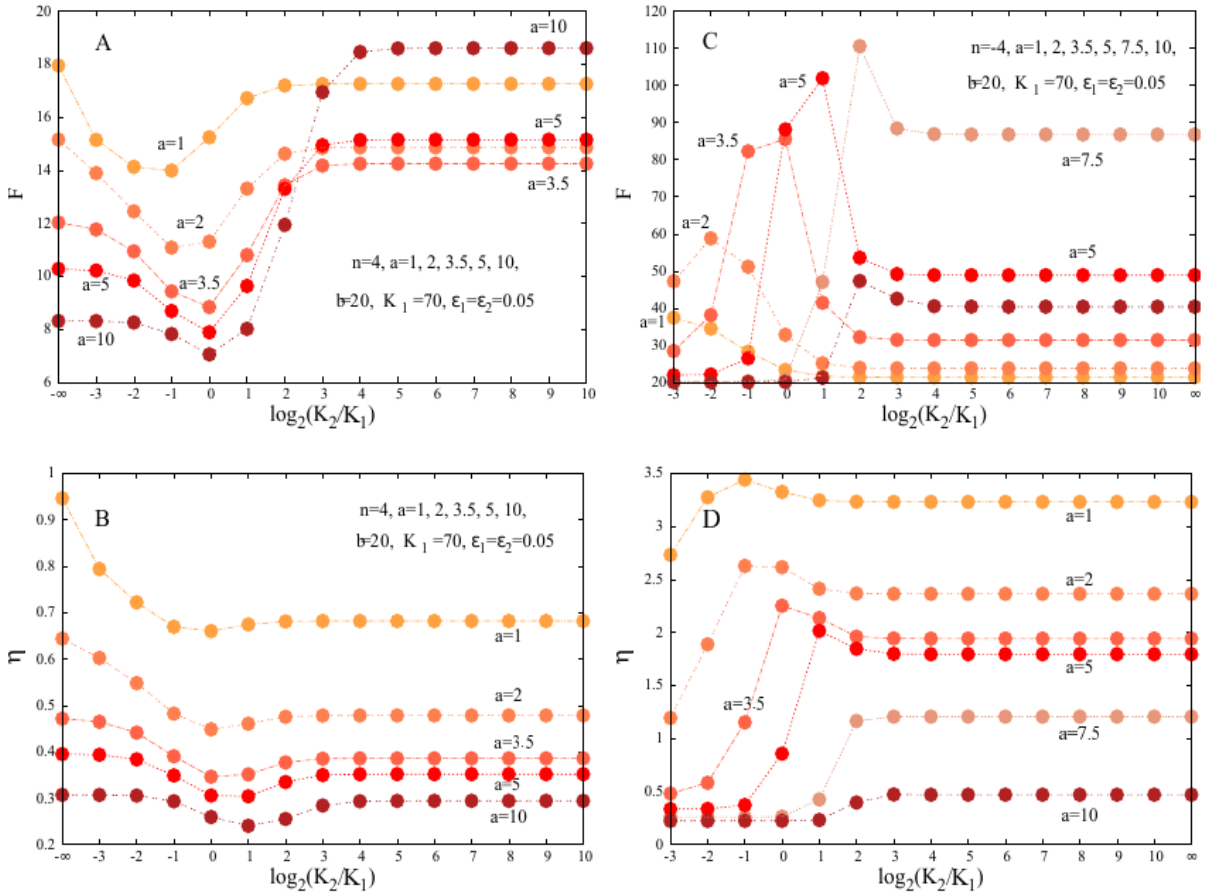
Coefficient of variation



Two non-equivalent copies of a negatively (A,B) and positively (C,D) self-regulating gene: Different measures of noise, Fano factor F and coefficient of variation η, may show differences in their behaviour as functions of the relative sensitivity K_2/K_1 of both promoters to auto-regulation. For negative auto-regulation, $n = 4$ (A,B), the positions and depth of minima are different for F and η. For positive auto-regulation, $n = -4$ (C,D), the maxima of both measures of noise roughly correspond to the transition through bimodal distributions. The exact positions and height of the maxima are, however, different for F and η. Additionally, for both positive and negative auto-regulation, F varies non-monotonically with a, whereas the dependence of η on a is monotonic. Parameters: $b = 20$, $K_1 = 70$, $\epsilon_1 = \epsilon_2 = 0.05$.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Fano factor and CV vary in a different manner as the relative difference in regulation strengths is varied



This demonstrates that experimental assessments of the influence of gene expression noise on cell fitness may be ambiguous because they are dependent on the particular function used to quantify noise.

J. Jędrak and A. Ochab-Marcinek, Journal of Theoretical Biology **408**, 222 (2016).

Result 5

Fano factor and CV vary differently.
Therefore, it does not make sense to say that
„evolution optimizes a gene system
to minimize noise”
(unless we really know
how evolution measures noise)

Model of a single, non-regulated gene

Time-dependent solutions

Model of a single, non-regulated gene

Time-dependent solutions

PHYSICAL REVIEW E **94**, 032401 (2016)

Time-dependent solutions for a stochastic model of gene expression with molecule production in the form of a compound Poisson process

Jakub Jędrak* and Anna Ochab-Marcinek

Stochastic Model

(Friedman et al. [Phys. Rev. Lett. **97**, 168302 (2006)])

$$\frac{\partial p(x,t)}{\partial t} = \gamma(t)\frac{\partial}{\partial x}\left[xp(x,t)\right] + k(t)\int_0^x w(x-x',t)p(x',t)dx' \tag{3}$$

where $w(x-x') = \nu(x-x') - \delta(x-x')$, $\nu(u)$ is the burst-size PDF, $u = x - x'$; remaining model parameters (functions of time): $\gamma(t)$, $k(t)$. (In Mathematics similar models were studied by Lajos Takács).

Laplace transform $p(x,t) \to \hat{p}(s,t) = \mathcal{L}\{p(x,t)\}$ applied to (3) yields

$$\frac{\partial \hat{p}(s,t)}{\partial t} + \gamma(t) s \frac{\partial \hat{p}(s,t)}{\partial s} + k(t)\hat{w}(s,t)\hat{p}(s,t) = 0. \qquad (10)$$

Method of characteristics gives

$$\hat{p}(s,t) = \Phi(\Omega(t)s) \exp\left(F(t, \Omega(t)s)\right).$$

where

$$\Phi(z) = \hat{p}(z,t_0) = \mathcal{L}\{p(x,t_0)\} \qquad (12)$$

$$F(t,z) = \int_{t_0}^{t} k(t')\hat{w}\left(\frac{z}{\Omega(t')}, t'\right) dt'. \qquad (13)$$

$$\Omega(t) = e^{-\int_{t_0}^{t} \gamma(t')dt'}. \qquad (14)$$

Result 1

All parameters time-dependent:
Solution for time evolution of cumulants

Moment-generating function

using well-known relations

$$\hat{p}(s,t) = M(-s,t) = \sum_{m=0}^{\infty} \mu_m(t) \frac{(-s)^m}{m!}, \tag{16}$$

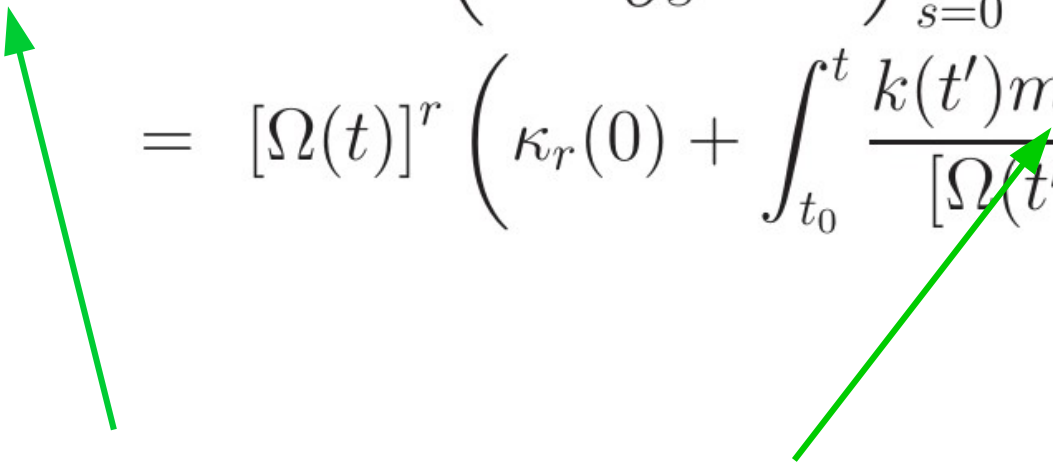
$$\ln[\hat{p}(s,t)] = K(-s,t) = \sum_{m=1}^{\infty} \kappa_m(t) \frac{(-s)^m}{m!}, \tag{17}$$

we obtain time-dependence of moments $\mu_r(t)$ and cumulants $\kappa_r(t)$ of $p(x,t)$

Cumulant-generating function

we obtain time-dependence of moments $\mu_r(t)$ and cumulants $\kappa_r(t)$ of $p(x,t)$,

$$\begin{aligned}\kappa_r(t) &= (-1)^r \left(\frac{\partial^r \ln[\hat{p}(s,t)]}{\partial s^r} \right)_{s=0} \\ &= [\Omega(t)]^r \left(\kappa_r(0) + \int_{t_0}^{t} \frac{k(t')m_r(t')}{[\Omega(t')]^r} dt' \right). \end{aligned} \quad (18)$$



r-th cumulant of the distribution of protein concentration depends only on the r-th moment of burst size distribution

Result 1

All parameters time-dependent:
Solution for time evolution of cumulants

Result 2

Only transcription rate time-dependent:
Oscillating transcription rate:
The amplitude of cumulants depends on frequency

Oscillatory solutions

Let us now analyze the case of molecule production rate (burst frequency) $k(t)$ of the form

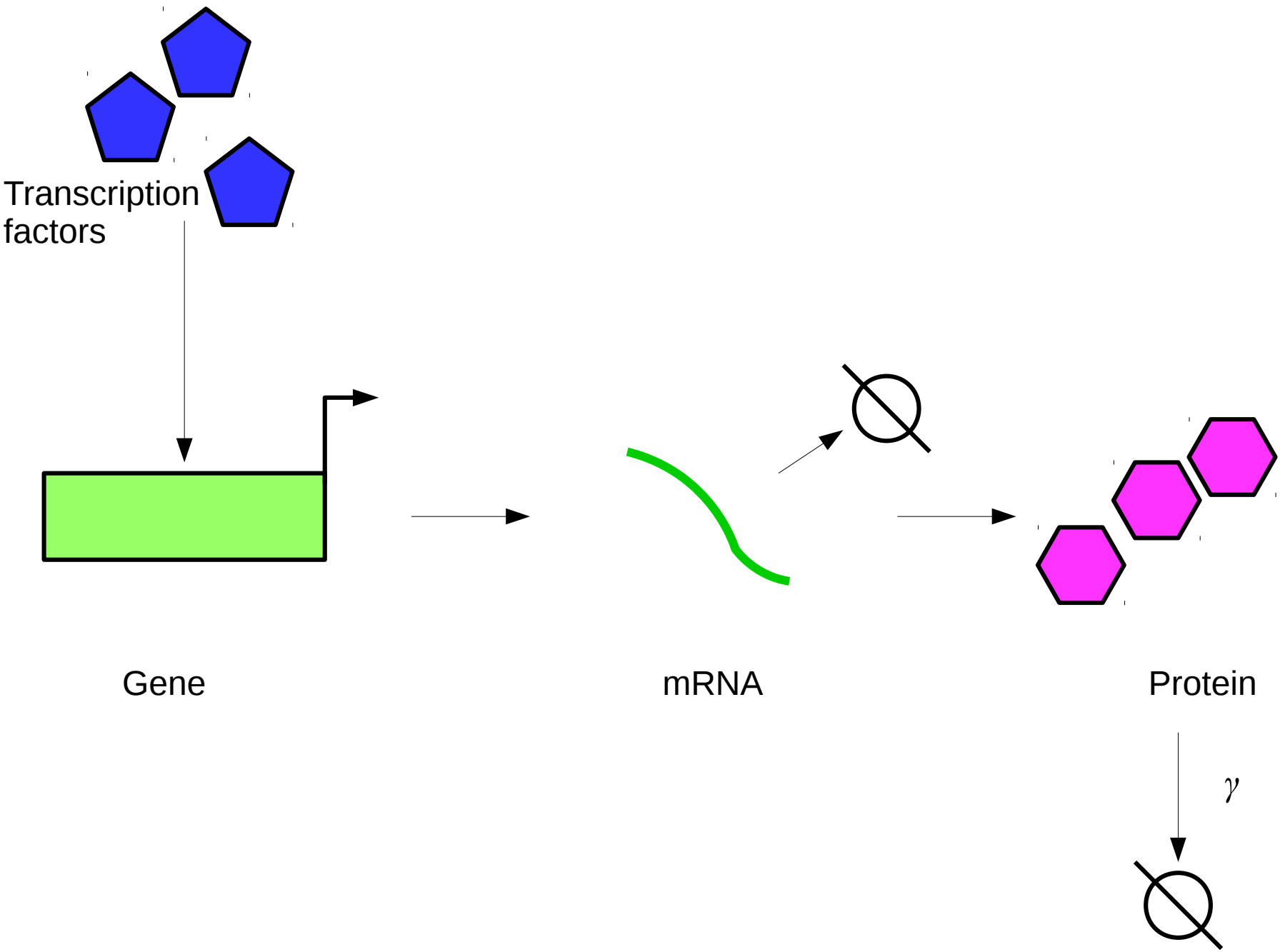
$$k(t) = C_1 \sin(\omega_f t + \varphi) + C_2, \tag{33}$$

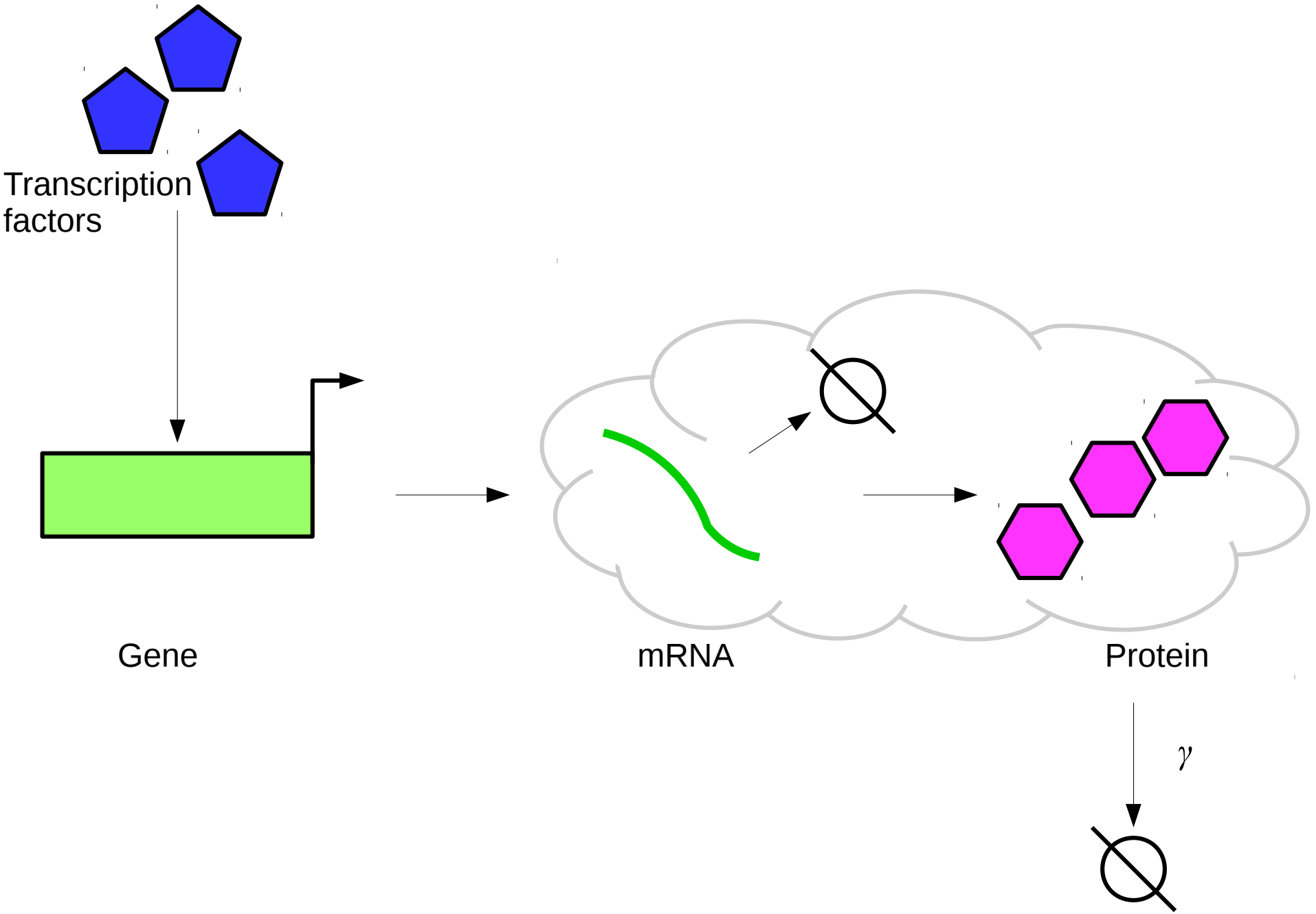
where $0 \leq C_1 < C_2$, i.e., gene is periodically driven with a single angular frequency,

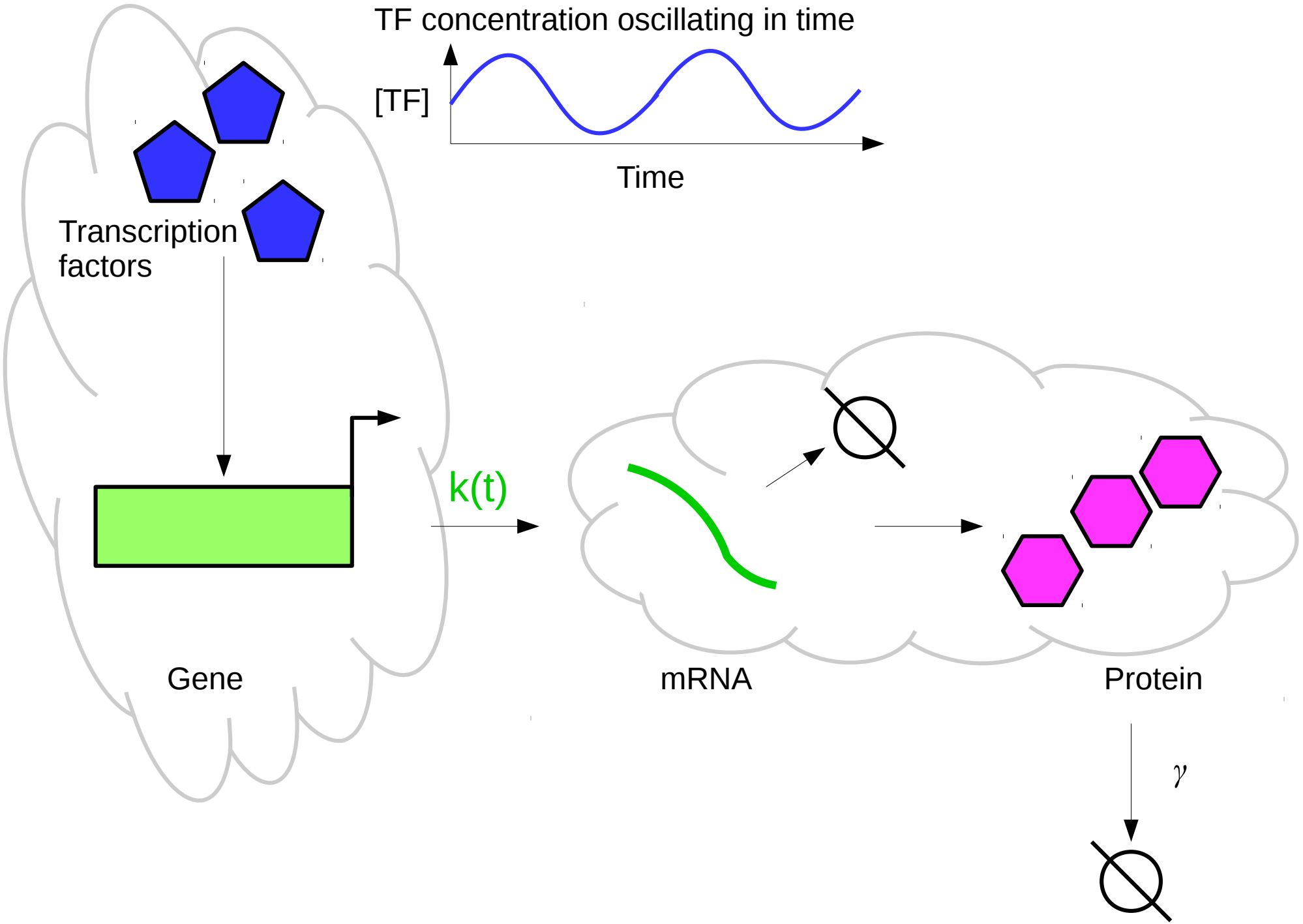
$$\omega_f = 2\pi/T, \tag{34}$$

where T is an oscillation period; φ is the initial phase.

Time-independent, but otherwise arbitrary burst size probability distribution *v(u)*, constant decay rate *y* and molecule production rate (burst frequency) *k(t)*


Transcription factors
Gene
mRNA
Protein
γ


Transcription factors
Gene
mRNA
Protein
γ


TF concentration oscillating in time
[TF]
Time
Transcription factors
k(t)
Gene
mRNA
Protein
γ

Oscillatory solutions

$$\begin{aligned}\kappa_r(t) &= \kappa_r(0)e^{-r\gamma t} + \frac{C_2 m_r}{r\gamma}\left(1 - e^{-r\gamma t}\right) \\ &+ \frac{C_1 m_r \sin(\omega_f t + \varphi + \beta)}{\sqrt{r^2\gamma^2 + \omega_f^2}} \\ &- \frac{C_1 m_r \sin(\varphi + \beta)e^{-r\gamma t}}{\sqrt{r^2\gamma^2 + \omega_f^2}},\end{aligned} \tag{35}$$

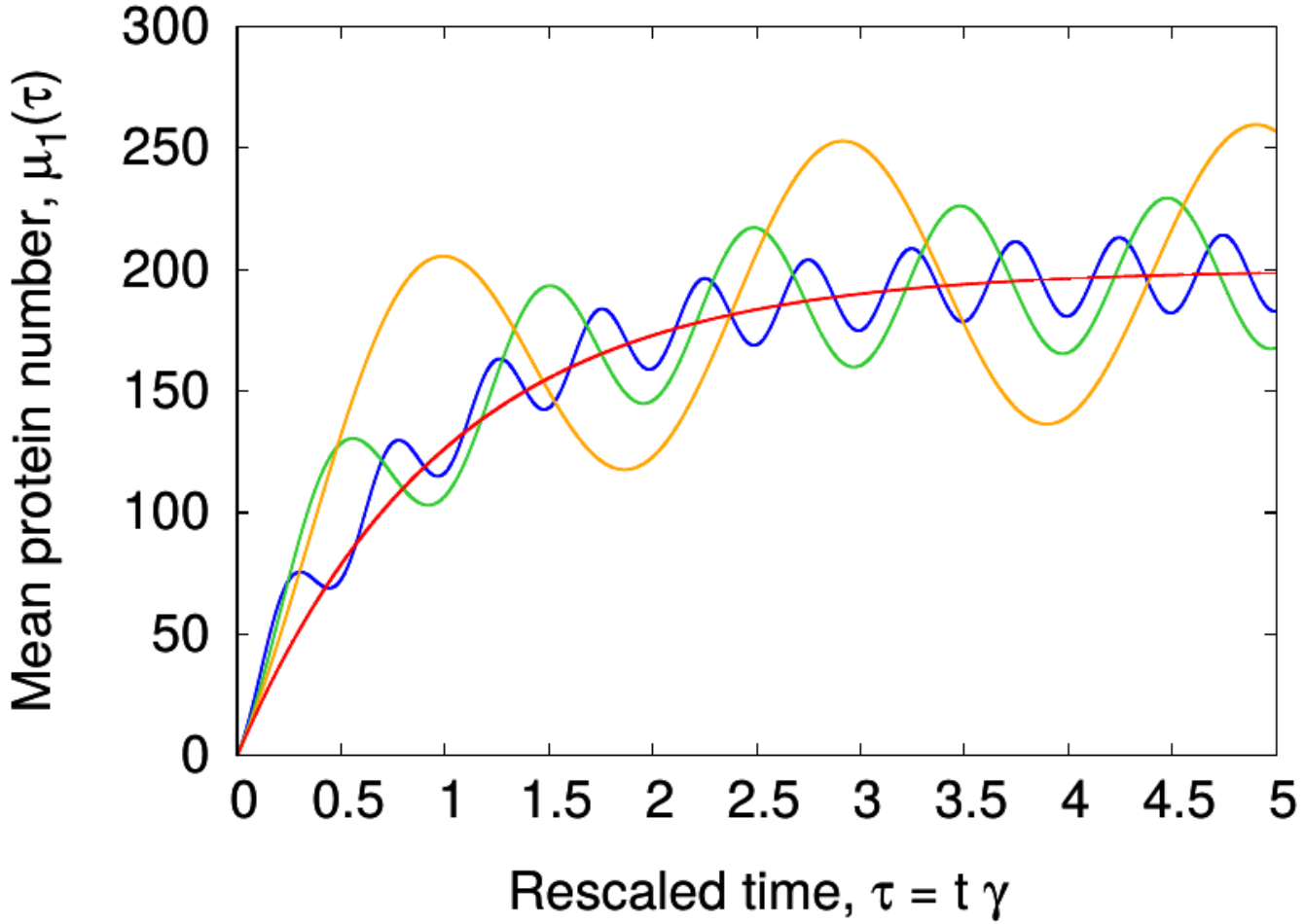
where

$$\beta = \arctan\left(\frac{-\omega_f}{r\gamma}\right). \tag{36}$$

$$A_r(\gamma, \omega_f) = \frac{C_1}{\sqrt{r^2\gamma^2 + \omega_f^2}}. \tag{37}$$

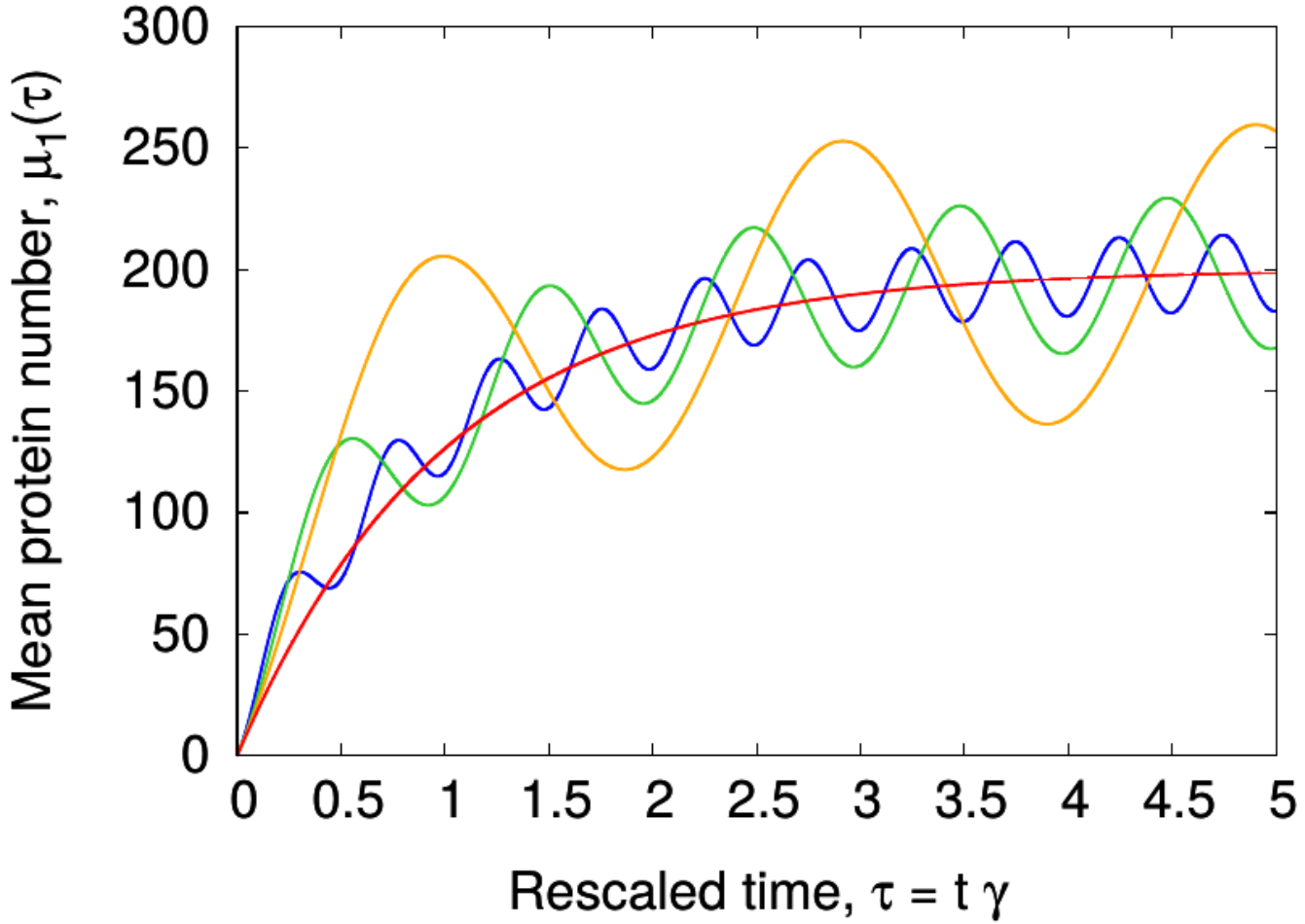
$A_r(\gamma, \omega_f)$ (37) is a monotonically decreasing function of ω_f, therefore no resonant behavior should be expected.

Oscillatory solution - mean protein number



Physical analogy 1: Resistor-capacitor (RC) low-pass filter:
Fast oscillations of the external driving of gene expression have less effect than the slow ones

Oscillatory solution - mean protein number



Physical analogy 2: motion of a particle moving with velocity $v = \kappa_r$ in a viscous medium under the influence of both the drag force [$-r\gamma\kappa_r(t)$, with constant γ] and the external periodic force $m_r(t)k(t)$.

Result 2

Only transcription rate time-dependent:
Oscillating transcription rate:
The amplitude of cumulants depends on frequency

Result 3

All parameters time-independent:
Simple form of cumulants.
Measures of noise are again ambiguous.

All parametes time-independent

Time evolution of cumulants:

$$\kappa_r(t) = \kappa_r(0)e^{-r\gamma t} + a(1 - e^{-r\gamma t})\frac{m_r}{r}$$

r-th cumulant of the distribution of protein concentration depends only on the r-th moment of burst size distribution

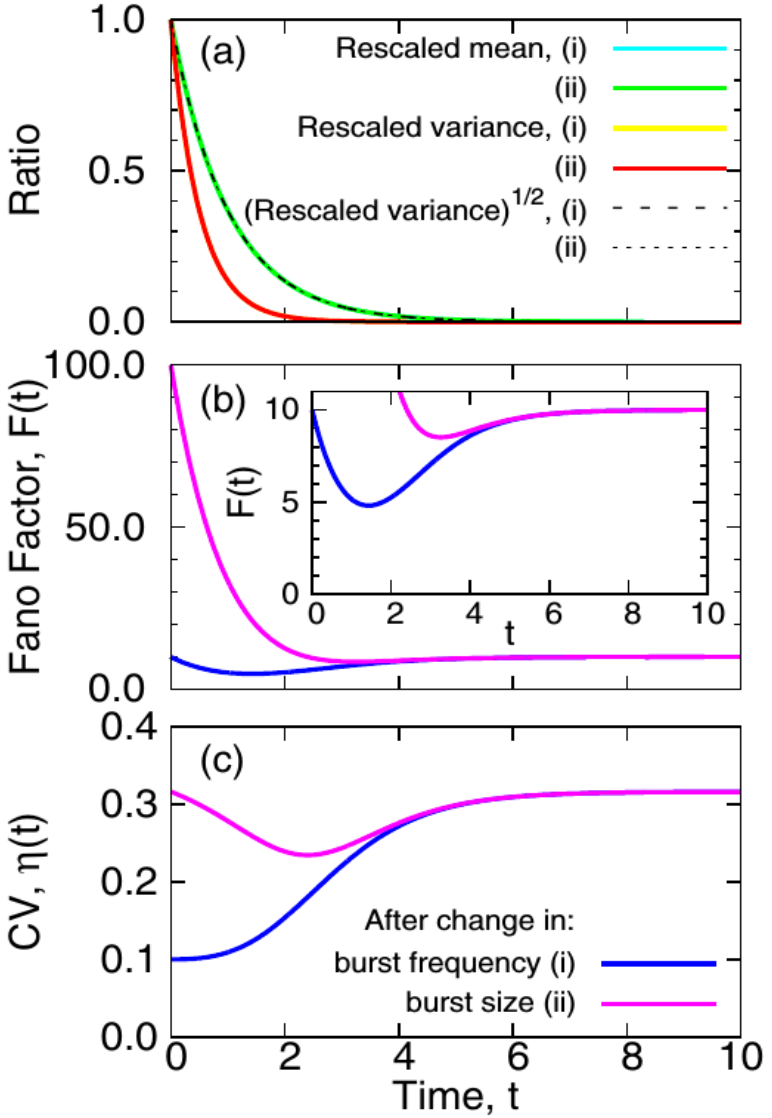


FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

The gene was at a certain level of expression

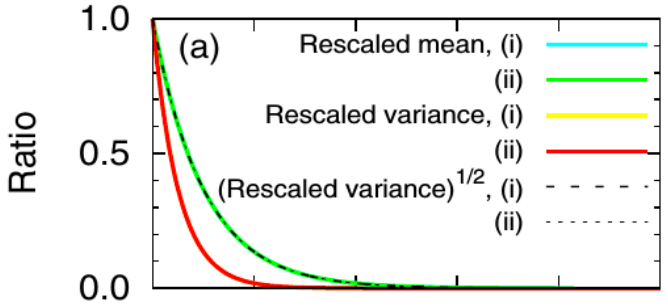
Then, the expression is reduced:

- By the reduction of mean burst size

  or

- By the reduction of mean burst frequency

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).



Relative change in mean/variance:

$$K_r(t) \equiv \frac{\kappa_r(t) - \kappa_r(\infty)}{\kappa_r(0) - \kappa_r(\infty)} = e^{-r\gamma t}$$

where cumulants are:

$$\kappa_r(t) = \kappa_r(0)e^{-r\gamma t} + a(1 - e^{-r\gamma t})\frac{m_r}{r}$$

FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

Fano factor and CV vary in a different manner

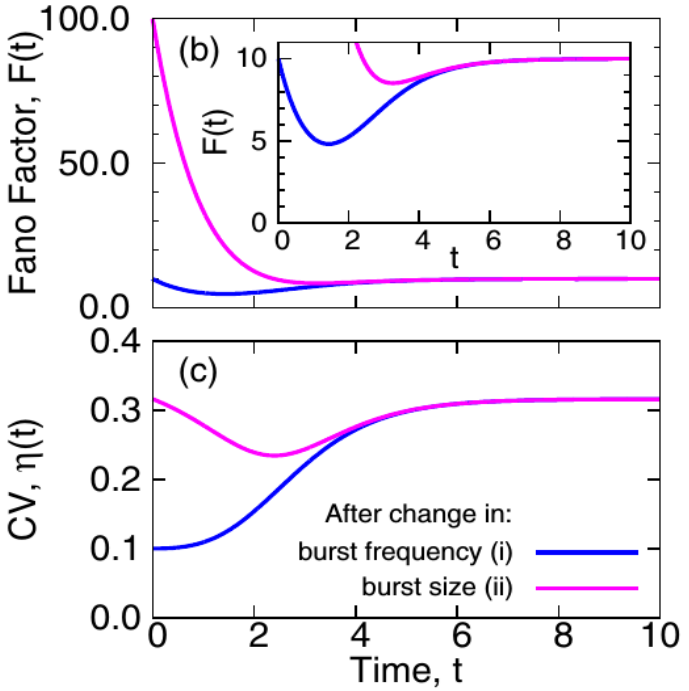


FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

If we propose any biological hypothesis about optimization with respect to noise, we need to justify why the specific measure of noise has been chosen.

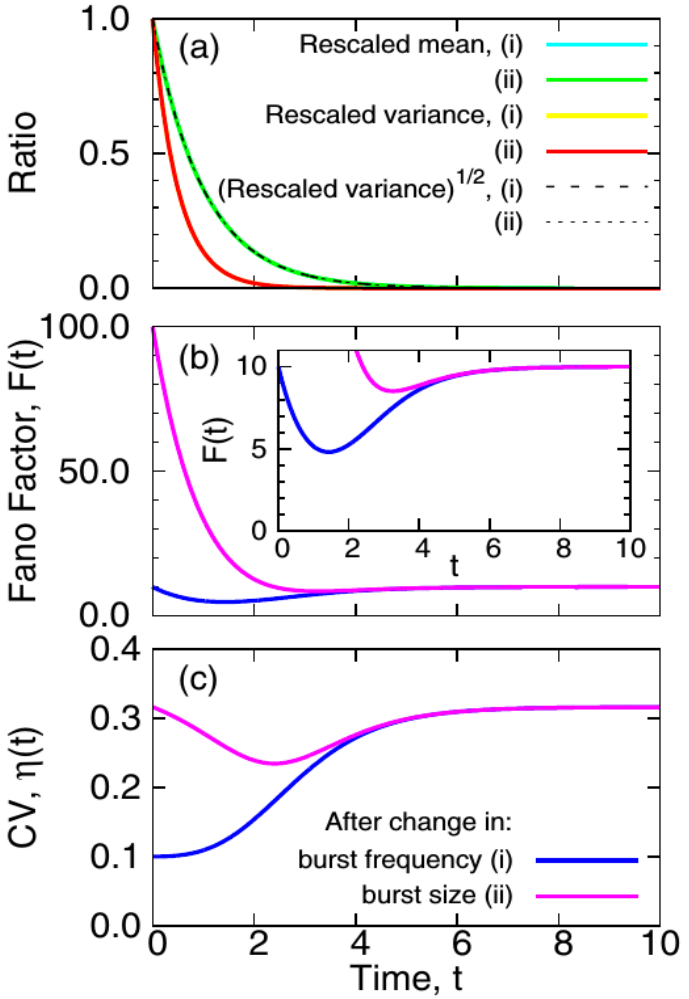


FIG. 3. Different measures of noise have different transient behaviors in time after an abrupt reduction of the mean burst frequency a (i), or the mean burst size b (ii). The mean protein concentration $\langle x \rangle = ab$ was decreased from 10^3 to 10^2. (a) Fractional change of mean, $K_1(\tau)$, fractional change of variance, $K_2(\tau)$, and its square root, $K_2(\tau)^{1/2}$. The curves for the cases (i) and (ii) overlap. (b) Fano factor $F(\tau)$, inset: zoom to show the minima. (c) Coefficient of variation, $\eta(\tau)$.

J. Jędrak and A. Ochab-Marcinek, Physical Review E **94**, 032401 (2016).

Result 3

All parameters time-independent:
Simple form of cumulants.
Measures of noise are again ambiguous.

Conclusions 1

- Stochastic model of an autoregulated gene, present in multiple copies.
- One-reporter assay may not measure correctly the extrinsic noise in self-regulated genes.
- Imperfect duplication of auto-activated gene: mixed, binary+graded response possible.
- Accumulation of gene duplications: non-trivial dependence on inherent noisiness of gene.
- Measurement of noise using Fano Factor or coefficient of variation is ambiguous.

Conclusions 2

- Stochastic model of a single non-regulated gene, arbitrary burst distribution, time-dependent parameters
- Time-dependent solutions for cumulants of probability distribution of protein concentrations
- r-th cumulant depends on the r-th moment of burst distribution
- Oscillatory gene regulation: Low-pass filter
- Measurement of noise using Fano Factor or coefficient of variation is ambiguous.

Open problem

- If the amount of noise in gene expression is optimized by the evolution, how is it measured?