From clockwork Biology

To noisy Biology

Biology of the *noisy* gene

Universidad Autónoma de Madrid

Jan 2008

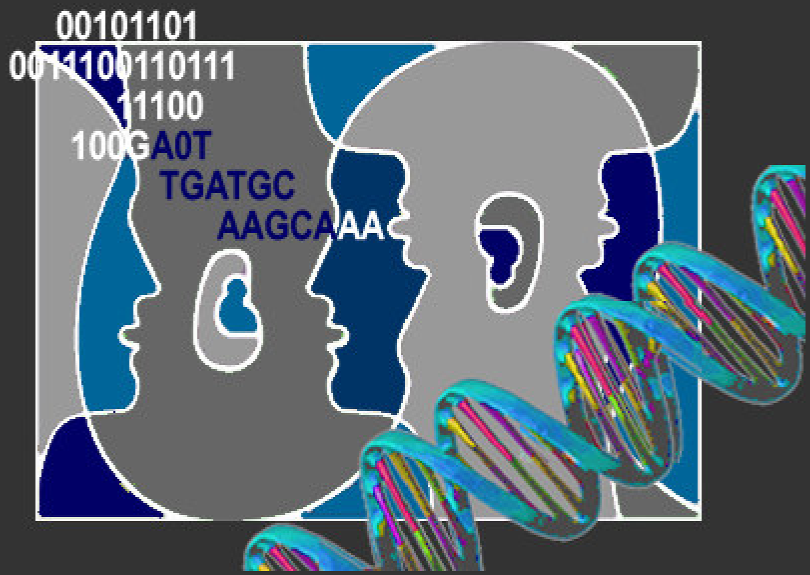
Juan F. Poyatos

Logic of Genomic Systems laboratory

CSIC

Spanish National Biotechnology Centre (CNB)

day I: introducing noise

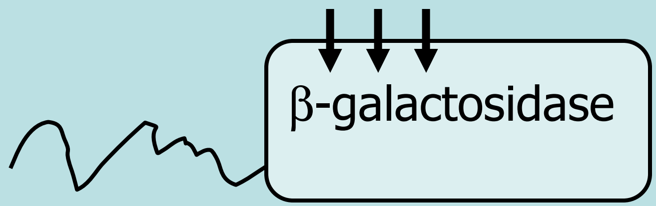


- biochemical noise
- low copy numbers
- simple model of gene expression
- master equation
- the 1/sqrt(n) rule of thumb

Noisy genes: suggestions from early enzyme inductions studies

Enzyme induction (enzymatic adaptation)
Monod's β-galactosidase studies

Lactose/TMG (gratuitous inducer)

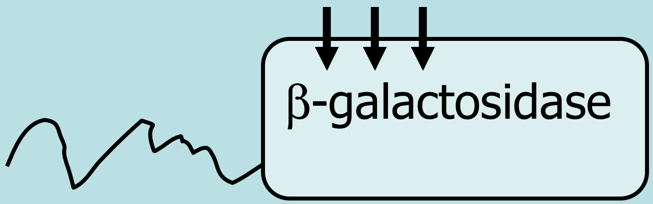


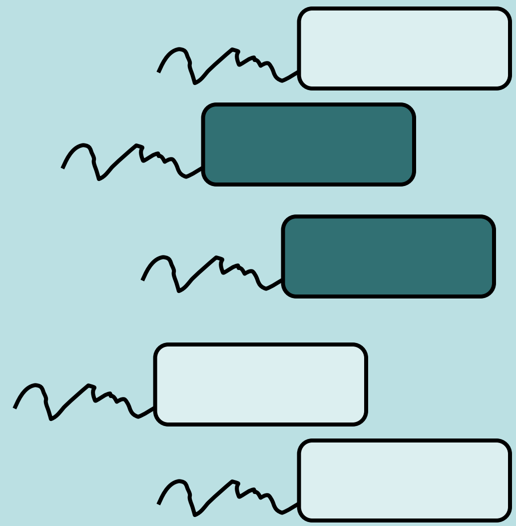
Benzer
Kinetics of induction at **high** [TMG]
Individual kinetics = population kinetics

Novick & Weiner, 1957
Kinetics of induction at **low** [TMG]
Individual kinetics ≠ population kinetics

Noisy genes: suggestions from early enzyme inductions studies

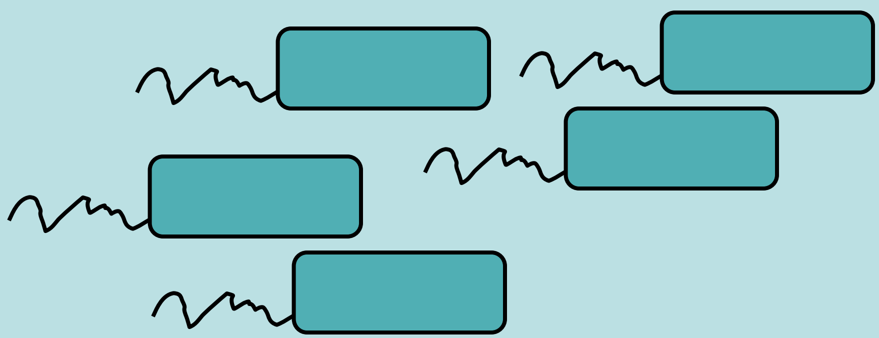
Lactose/TMG (gratuitous inducer)





Novick & Weiner, 1957
Kinetics of induction at **low** [TMG]
Different phenotypes in clonal populations
(no mutations!)

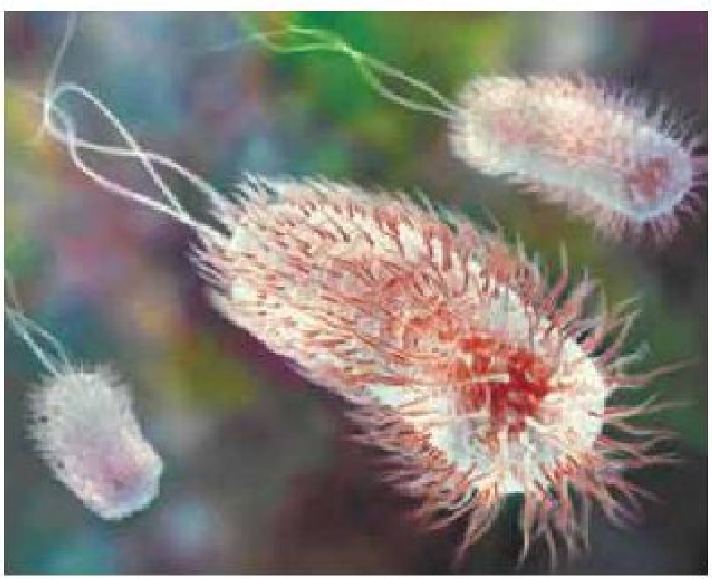
Instead of



Noisy genes: why?

- Many molecules that take part in gene expression (including DNA and important regulatory molecules such as the enzyme polymerase) act at extremely low intracellular concentrations (low copy numbers)

- Gene expression as a series of biochemical reactions experiences "surprising" things when one takes the discreteness of molecule number seriously



Escherichia Coli (*E. coli*) numbers

V ~$\pi/2\ 10^{-15}$ liters

(2μm long, 1μm diameter)

[RNA Polymerase] ~ 100nM = 100 molecules

$100\ 10^{-9}$ mol/lit X 10^{-15} lit X $6\ 10^{23}$ molecules/mol

(1nM ~ 1 molecule)

Biochemical noise

-consider a simple gene expression system
(unregulated gene)

a common approach is to describe these reactions by means of differential reaction-rate equations

$$\emptyset \xrightarrow{k} P, \quad \text{production}$$

$$P \xrightarrow{\delta} \emptyset, \quad \text{first order degradation}$$

$$\frac{d[P]}{dt} = k - \delta[P]$$

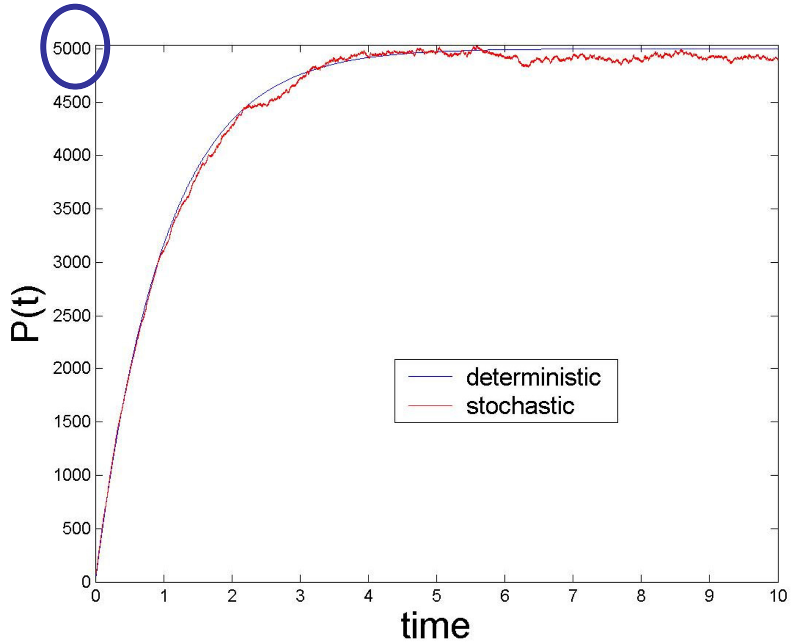
This approach assumes that the time evolution of such reaction is both continuous and deterministic

Biochemical noise

continuous? molecule number changes in discrete ways

deterministic? impossible to predict the motion of (classical) molecules due to the ignorance of positions and velocities of all components of the system

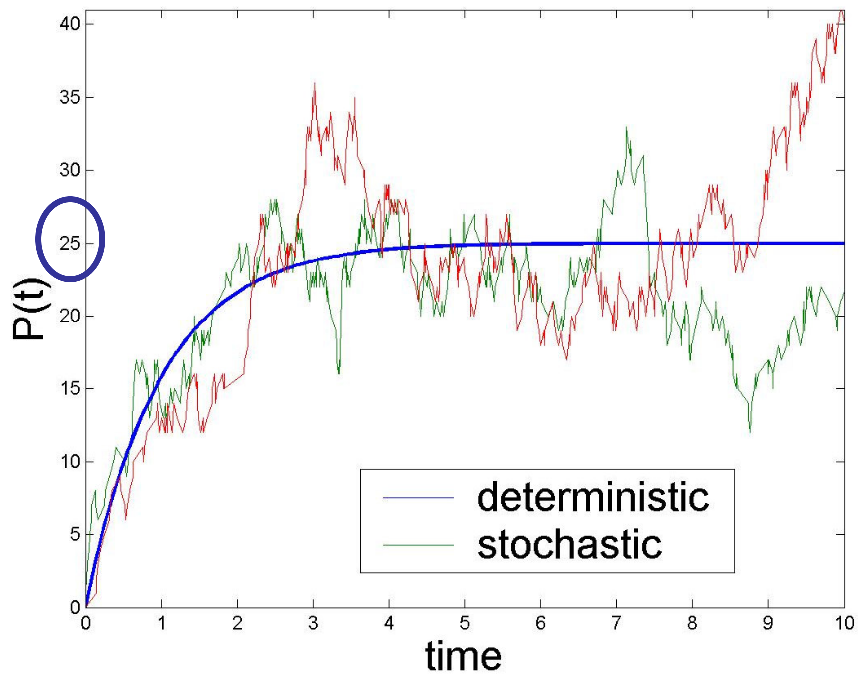
however in many cases of course the time evolution of a chemically reacting system can, to a very acceptable degree of accuracy, be treated as a continuous, deterministic process



large number of molecules
deterministic approximation works

small number of molecules
deterministic approximation fails

large protein fluctuations



Stochastic motion

- motion generated by random forces, e.g., forces randomly applied in time

- to describe a stochastic system we need probabilities

- chemical systems are intrinsically stochastic (noisy), specially when a small pool of reactants is involved

Stochastic description of chemical reactions

Recall:

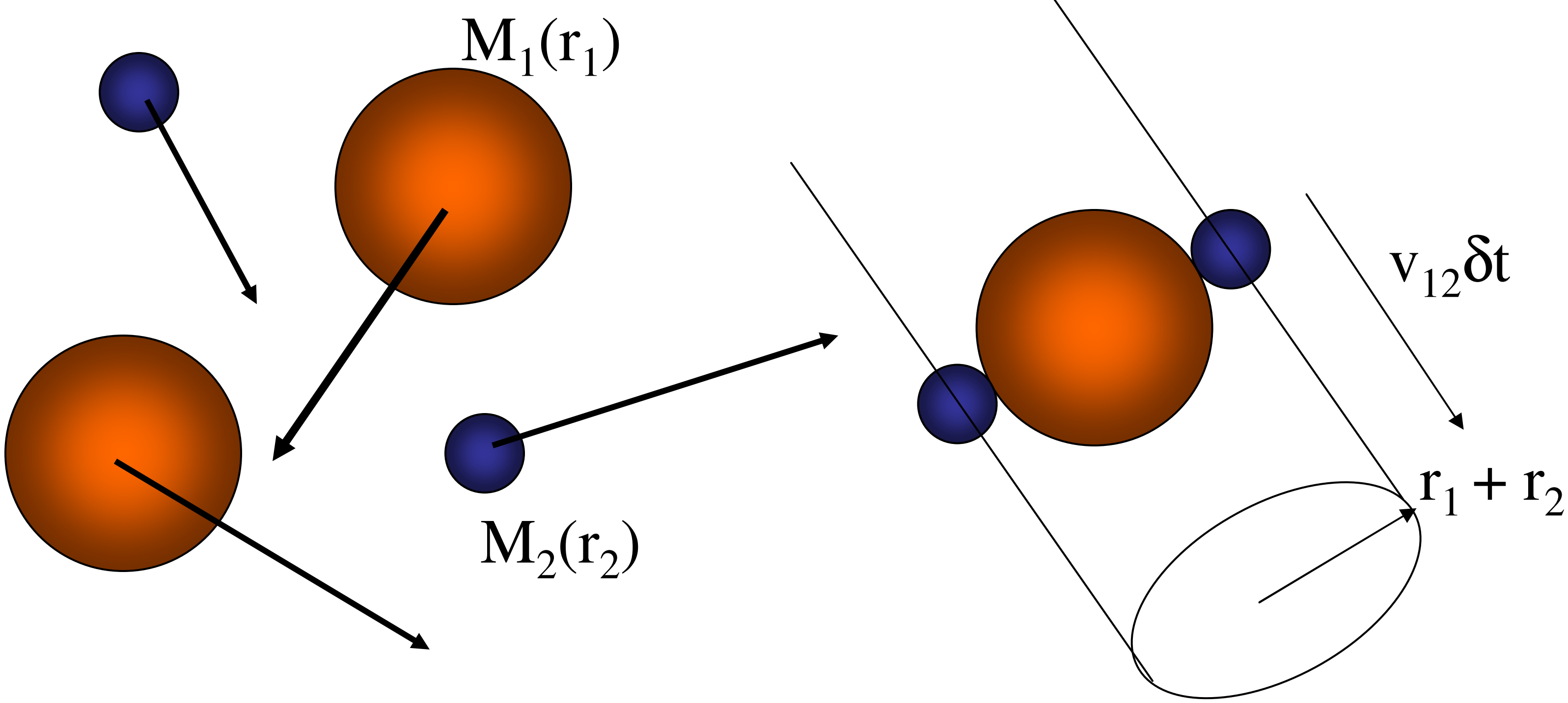
For a stochastic system it is not possible to determine exactly the state of the system at later times given its state at the current time.
We must thus deal with probabilities.

Basis of the stochastic formulation: a chemical reaction occurs when molecules collide in an appropiate way

- Molecular collisions: random microscopic events

Stochastic description of chemical reactions

V_{coll} – Collision volume. The molecules M_2 which are within collision volume will be hit by a particular molecule M_1 in the next time interval δt.



$$V_{coll} = v_{12}\delta t\, \pi\, (r_1 + r_2)^2$$

$$P\text{ (a given } M_1 \text{ and } M_2 \text{ collide)} = \frac{\overline{v_{12}}\,\delta t\,\pi\,(r_1 + r_2)^2}{V}$$

$$P\text{ (a } M_1 \text{ and } M_2 \text{ molecule collide)} = n_1\,n_2\,\frac{\overline{v_{12}}\,\delta t\,\pi\,(r_1 + r_2)^2}{V}$$

and finally

diffusion-limited R close to one always

$$P\text{ (a } M_1 \text{ and } M_2 \text{ react)} = n_1\,n_2\,\boxed{\frac{\overline{v_{12}}\,R\,\pi\,(r_1 + r_2)^2}{V}}\,\delta t = n_1\,n_2\,\mathbf{r}\,\delta t$$

probability that a given M_1 and M_2 react in unit time (**r**)

units of inverse time

this is the fundamental hypothesis from which we derive both the Master Equation and the Stochastic Simulation approaches.

The Master Equation

The stochastic framework considers the discrete number of molecules whose state changes probabilistically

Recall our previous simple gene expression model

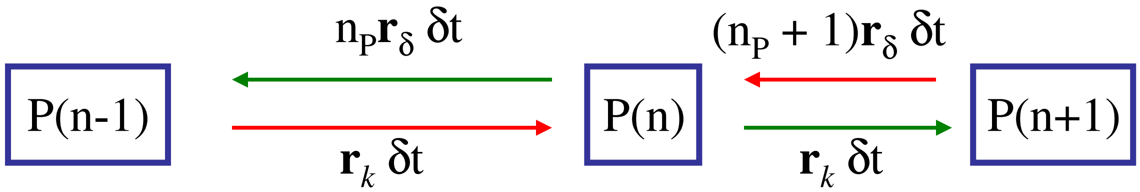
$$\emptyset \xrightarrow{k} P, \qquad \text{P}(k \text{ reaction}) = \mathbf{r}_k\, \delta t$$

$$P \xrightarrow{\delta} \emptyset, \qquad \text{P}(\delta \text{ reaction}) = n_P\, \mathbf{r}_\delta\, \delta t$$

$$\frac{d[P]}{dt} = k - \delta[P]$$

Thus, we go from reaction rates to reaction probabilities per unit time

How does the probability of having, say, n P molecules, p(n), change with time?



$$\begin{aligned} p(n, t+\delta t) &= p(n,t) \\ &+ p(n-1,t) r_k \delta t \\ &+ p(n+1,t)(n_P+1) r_\delta \delta t \\ &- p(n,t) n_P r_\delta \delta t \\ &- p(n,t) r_k \delta t. \end{aligned}$$

and thus we get in the limit $\delta t \rightarrow 0$

$$\frac{dp(n)}{dt} = -p(n)(r_k + n_P r_\delta) + p(n-1) r_k + p(n+1)(n_P+1) r_\delta$$

Which distribution?

Steady State

$$\frac{dp_n}{dt} = 0 = -p_n(r_k + nr_\delta) + p_{n-1}r_k + p_{n+1}(n+1)r_\delta$$

and

$$-p_n r_k + p_{n+1} r_\delta (n+1) = -p_{n-1} r_k + p_n r_\delta n$$

then

$-p_n r_k + p_{n+1} r_\delta (n+1)$ is constant (independent of n).

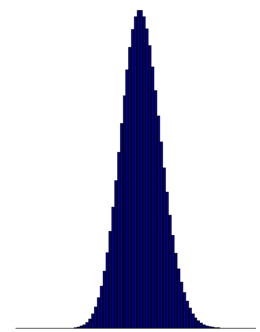
further, considering that $\langle n \rangle_{ss} = \frac{r_k}{r_\delta}$ and that the probability is normalizable → constant = 0

Which distribution?

thus $$p_n = \frac{\langle n \rangle_{ss}}{n} p_{n-1} = \ldots = \frac{\langle n \rangle_{ss}^n}{n!} p_0 .$$

since $\sum_n p_n = 1$ we get

$$p_n = \frac{\langle n \rangle_{ss}^n}{n!} e^{-\langle n \rangle_{ss}}$$



$\langle n \rangle_{ss}$

the steady state distribution is the Poisson Distribution

A simple model of gene expression, summary

Poisson distribution

mean $\langle n \rangle = \langle n \rangle_{ss}$

variance $\sigma^2 = \langle n \rangle_{ss}$

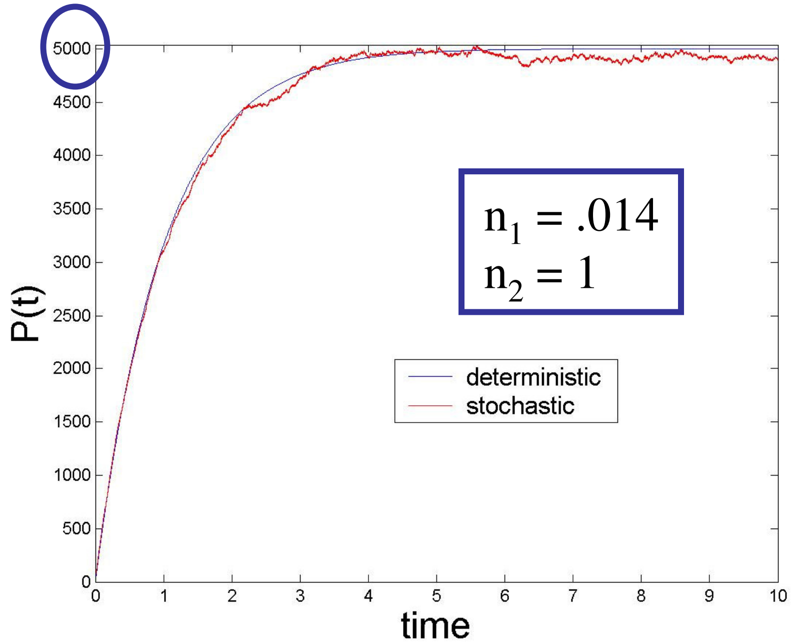
Macroscopic statistics

What is noise then?

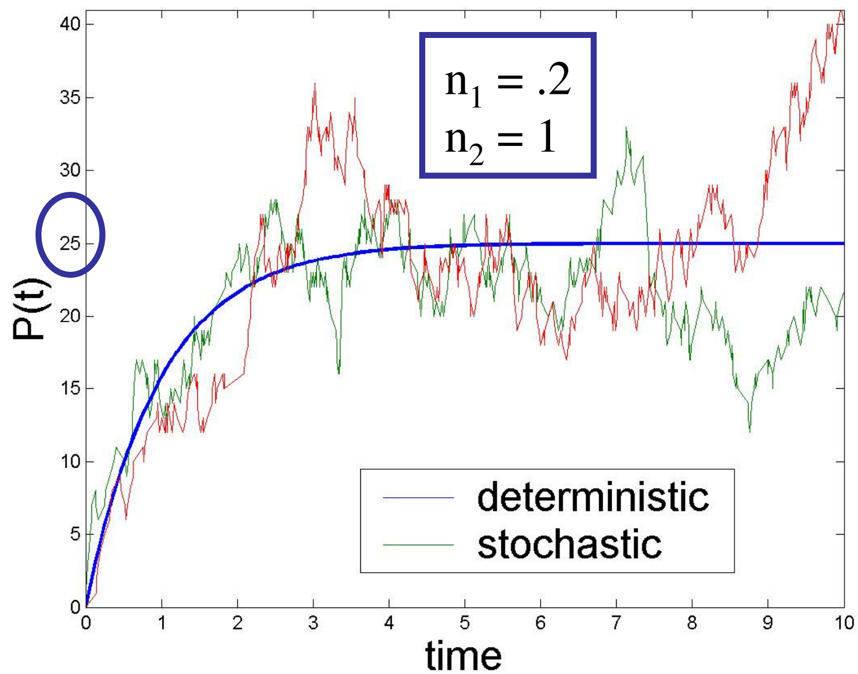
standard deviation

definition-1 (coe. variat.) = $n_1 = \frac{\sigma}{\langle n \rangle}$ (= $1/\sqrt{\langle n \rangle}$. Poisson distribution, noise increases as the number of molecules decreases)

definition-2 (Fano factor) = $n_2 = \frac{\sigma^2}{\langle n \rangle}$ (= 1, Poisson distribution mean = variance)



large number of molecules
deterministic approximation works
spread of 1/sqrt(500)~**1%** around mean



small number of molecules
deterministic approximation fails

large protein fluctuations
spread of 1/sqrt(25)=**20%** around mean

MATLAB code 1

```
% .. code1.m
% .. simple gene expression deterministic equations

clear all
k = 10;
delta = 1;

tspan = [0 10];
P0 = 0;
options = [];
[t P] = ode23(@code1equations,tspan,P0,options,k,delta);
```

```
% .. code1equations.m
% .. rate equations for code1

function dPdt = code1equations(t,P,k,delta)

dPdt = [k - delta*P(1)];
```