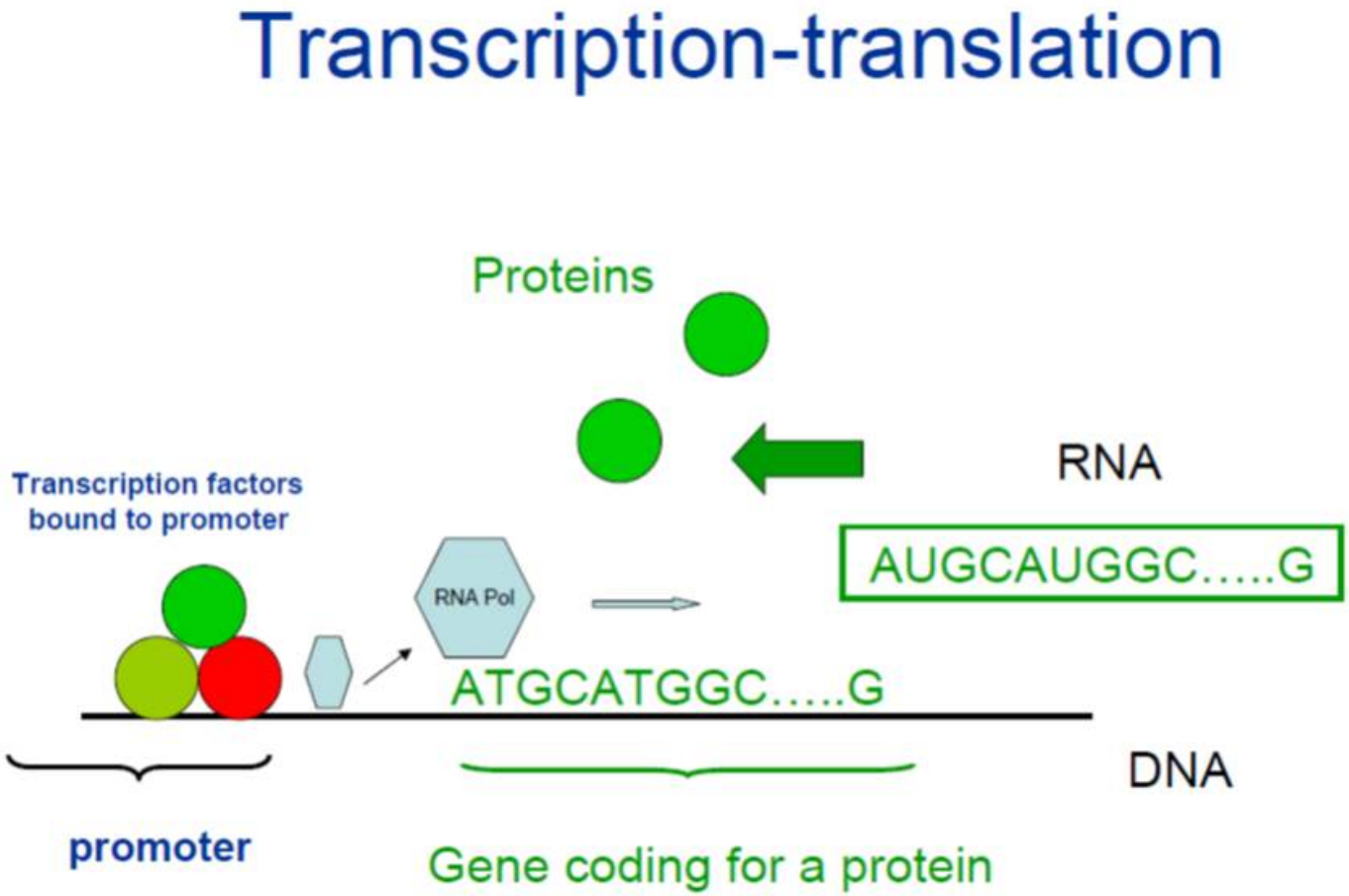
Time-continuous Markov chains and chemical reaction networks

Christian Mazza

Dept. of mathematics, University of Fribourg (CH) and SIB

SCNAT/Rigi 2015

Gene expression



Gene expression

Repress transcription

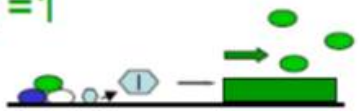
Y=0

The promoter is off, the RNA polymerase can not bind, no transcription

Y=1

The promoter is on, the RNA polymerase can bind and begin transcription

Y=1



Mathematical model

N(t)=# protein present in the cell at time t

Y(t) = state of the promoter

Gene expression

A single gene

Mathematical model

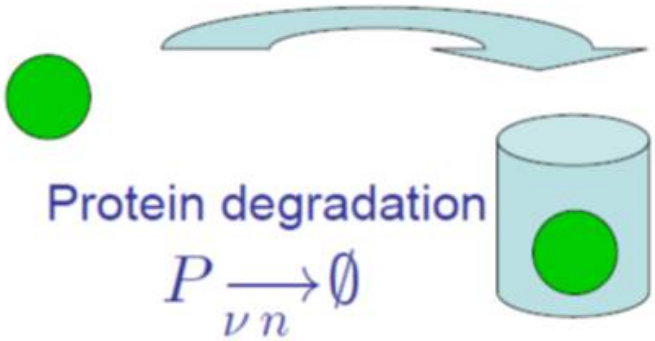
N(t)=# protein present in the cell
at time t
Y(t)= 1, promoter is on, transcription
Y(t)=0, promoter is off, no transcription

Gene expression

The time evolution of the pair (N(t),Y(t)) is modeled as a time continuous Markov chain: The Gillespie algorithm

When the promoter is on, Y(t)=1, N(t) is described as a birth and death process of Poisson steady state:

N(t)=n



Protein degradation

$$P \xrightarrow[\nu\, n]{} \emptyset$$

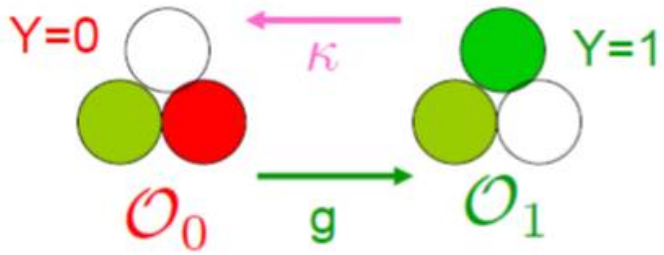
$$\emptyset \xrightarrow{\mu} P$$

Protein production

$$n \to n-1 \text{ at rate } \nu\ n$$

$$n \to n+1 \text{ at rate } \mu$$

Gene expression



Promoter fluctuations

Given (n,y):

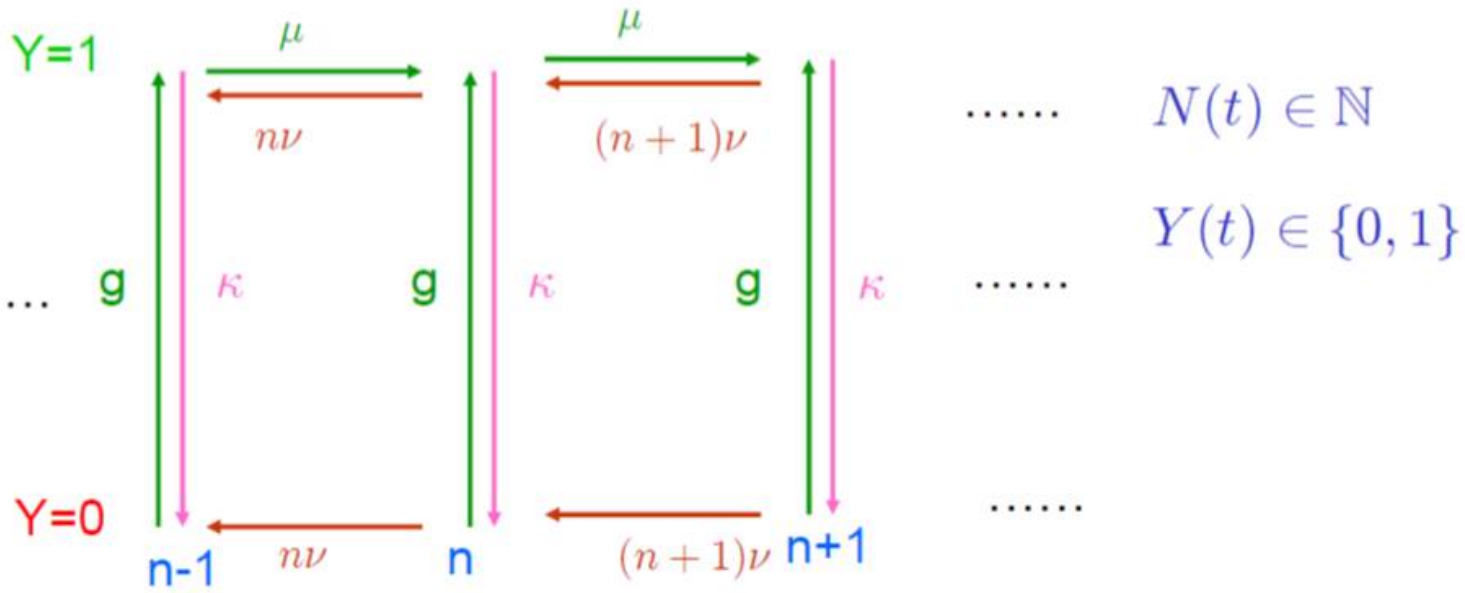
Chemical reactions	Statistical meaning
$P \xrightarrow[\nu\, n]{} \emptyset$	$n \to n-1$ at rate $\nu\, n$
$\emptyset \xrightarrow[\mu]{} P$	$n \to n+1$ at rate $\mu\, y$
$\mathcal{O}_0 \xrightarrow{g} \mathcal{O}_1$	$y=0 \longrightarrow y=1$ at rate g
$\mathcal{O}_1 \xrightarrow{\kappa} \mathcal{O}_0$	$y=1 \longrightarrow y=0$ at rate κ

Gene expression

Markov chain on a strip (N(t),Y(t))

$$P \longrightarrow \emptyset \qquad \emptyset \xrightarrow{\mu} P$$

$$\mathcal{O}_0 \longrightarrow \mathcal{O}_1 \qquad \mathcal{O}_1 \longrightarrow \mathcal{O}_0$$



Steady state

This simple model for transcription is described by the 4 chemical reactions:

$$P \xrightarrow[\nu\, n]{} \emptyset \qquad \emptyset \xrightarrow[\mu]{} P \qquad \mathcal{O}_0 \xrightarrow[g]{} \mathcal{O}_1 \qquad \mathcal{O}_1 \xrightarrow[\kappa]{} \mathcal{O}_0$$

which describe the transition rates of a time continuous Markov chain (N(t),Y(t))

Basic problems:

Find the steady-state distribution

Find the law of the process

Statistical inference on the parameters g,…, based on $N(t_1), N(t_2), \cdots$

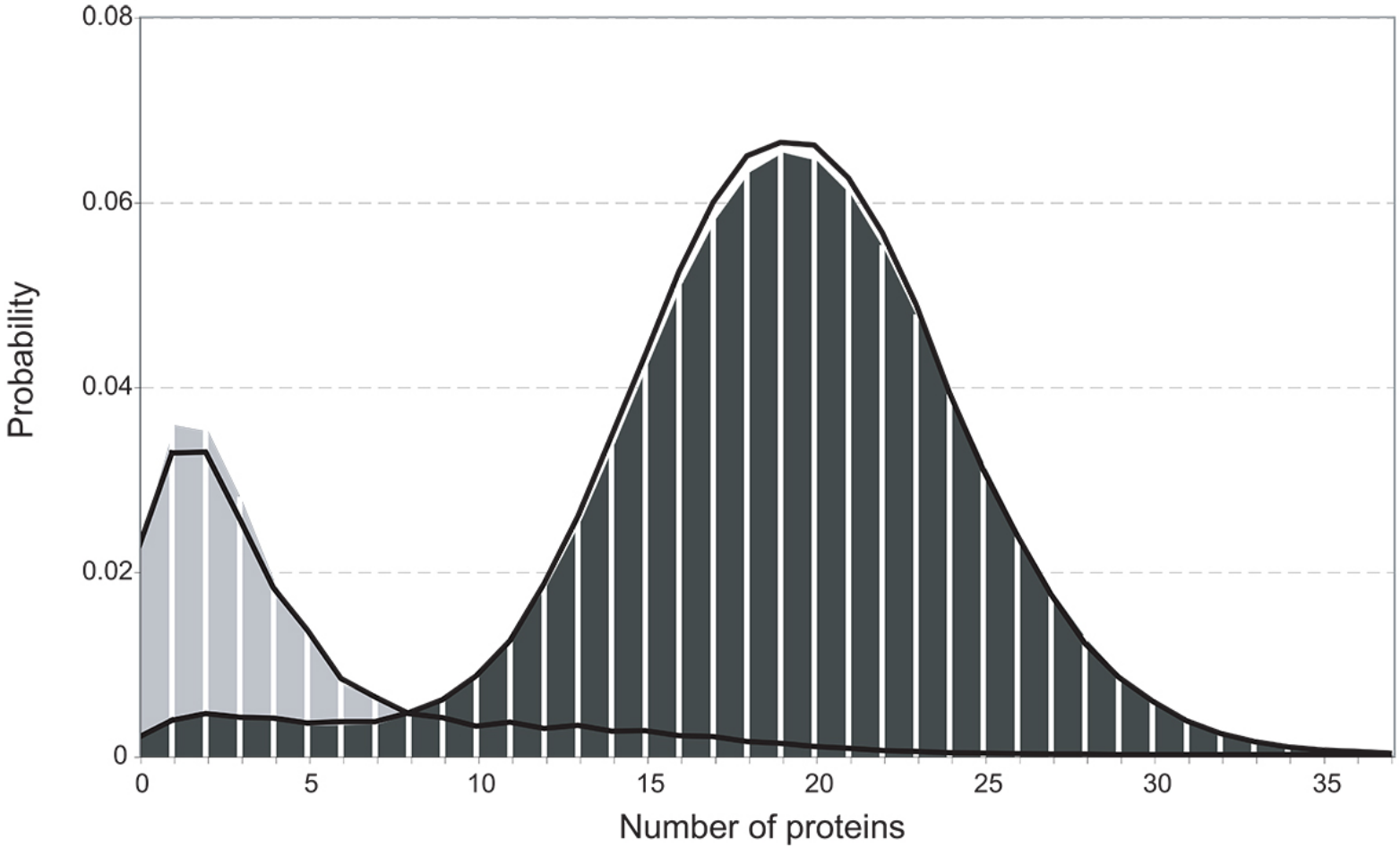


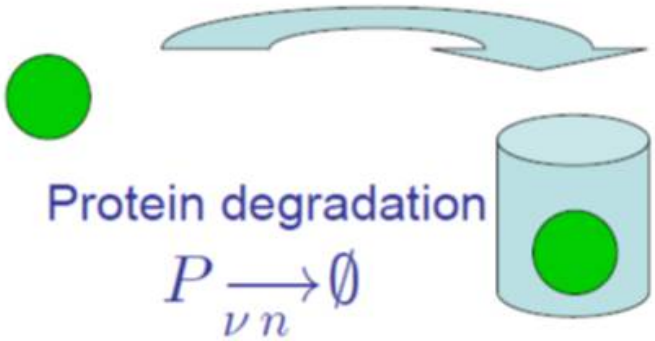
Figure: Grey bars: OFF; Dark bars: ON

A simple model of gene expression

The time evolution of the pair (N(t),Y(t)) is modeled as a time continuous Markov chain: The Gillespie algorithm

When the promoter is on, Y(t)=1, N(t) is described as a birth and death process of Poisson steady state:

N(t)=n



Protein degradation

$$P \xrightarrow[\nu\, n]{} \emptyset$$

$$\emptyset \xrightarrow{\mu} P$$

Protein production

$$n \to n-1 \text{ at rate } \nu\ n$$

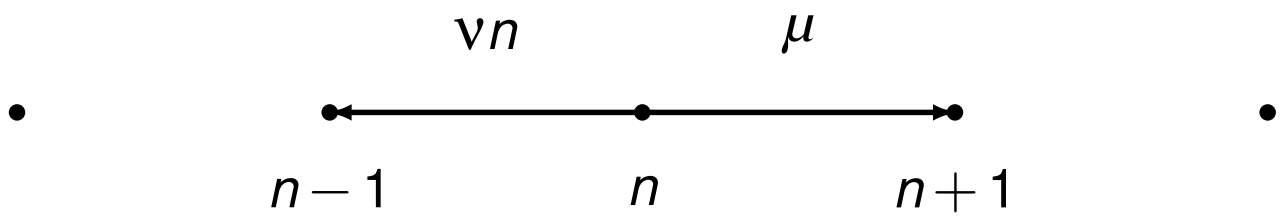
$$n \to n+1 \text{ at rate } \mu$$

Transition rates

The process should be such that

$$P(N(t+h) = n+1 | N(t) = n) \approx \mu h, \quad h \to 0,$$

$$P(X(t+h) = n-1 | X(t) = n) \approx \nu n h, \quad h \to 0,$$



Data defining general birth and death processes

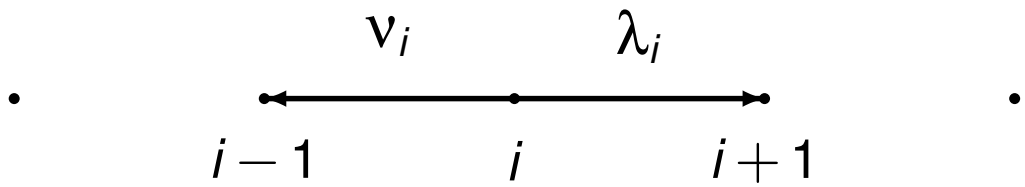


Figure: Nearest neighbours transitions for birth and death processes

$$P(X(t+h) = i+1 | X(t) = i) \approx \lambda_i h, \quad h \to 0,$$
$$P(X(t+h) = i-1 | X(t) = i) \approx \nu_i h, \quad h \to 0.$$

Simulation of general birth and death processes

New parameters

- a sequence of probabilities $(p_i)_{i\in\mathbb{N}}$, p_i (resp. q_i) is the probability to jump to $i+1$ (resp. to $i-1$)
- a sequence of positive real numbers $(a_i)_{i\in\mathbb{N}}$

which are such that

$$\lambda_i = \frac{p_i}{a_i}, \quad \nu_i = \frac{q_i}{a_i}, \quad p_i + q_i = 1,$$

$$\lambda_i + \nu_i = \frac{1}{a_i}, \quad p_i = \frac{\lambda_i}{\lambda_i + \nu_i}, \quad q_i = \frac{\nu_i}{\lambda_i + \nu_i};$$

Simulation: Gillespie Algorithm

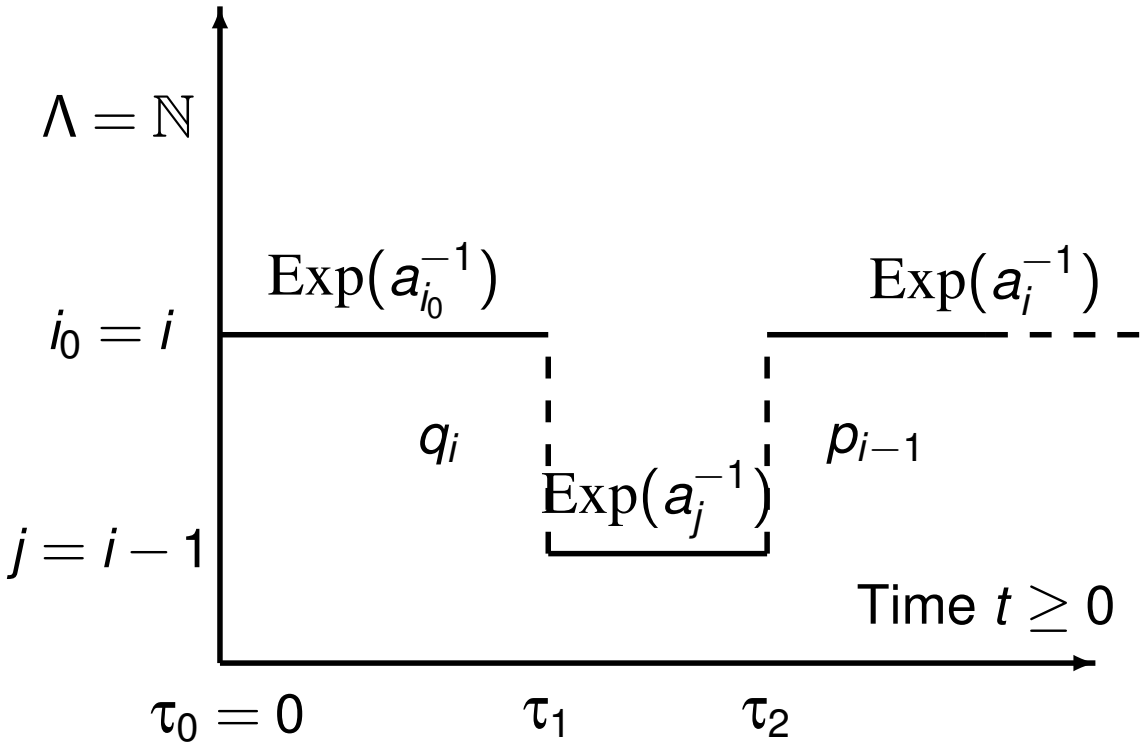


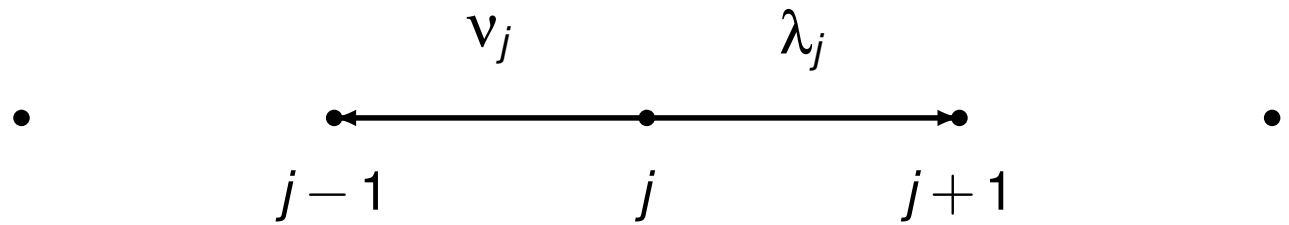
Figure: The Markov chain starts with $X(0) = i_0$. It waits there an exponential time of parameter $a_{i_0}^{-1}$, and then jumps at time τ_1 to the new state $j = i_0 - 1$ with probability q_{i_0}, so that $X(\tau_1) = j$.

Master equation

$$P_{ij}(t) = P(X(t) = j | X(0) = i).$$

Assume that $\inf_i a_i > 0$. Then the transition probabilities satisfy the Kolmogorov equation

$$\frac{\mathrm{d}P_{ij}(t)}{\mathrm{d}t} = \lambda_{j-1} P_{ij-1}(t) + \nu_{j+1} P_{ij+1}(t) - (\lambda_j + \nu_j) P_{ij}(t). \tag{1}$$



Steady state

Natural question: what is the probability that $\{X(t) = j\}$ when t is large ?

$$\frac{\mathrm{d}P_{ij}(t)}{\mathrm{d}t} = \lambda_{j-1}P_{ij-1}(t) + \nu_{j+1}P_{ij+1}(t) - (\lambda_j + \nu_j)P_{ij}(t). \tag{2}$$

One sets $\frac{\mathrm{d}P_{ij}(t)}{\mathrm{d}t} = 0$ to get the linear equation

$$0 = \lambda_{j-1}\pi_{j-1} + \nu_{j+1}\pi_{j+1} - (\lambda_j + \nu_j)\pi_j, \tag{3}$$

where, under mild conditions like $\sup_i(\nu_i + \lambda_i) < +\infty$,

$$\pi_j = \lim_{t\to\infty} P(X(t) = j | X(0) = i).$$

Substrates and enzymes

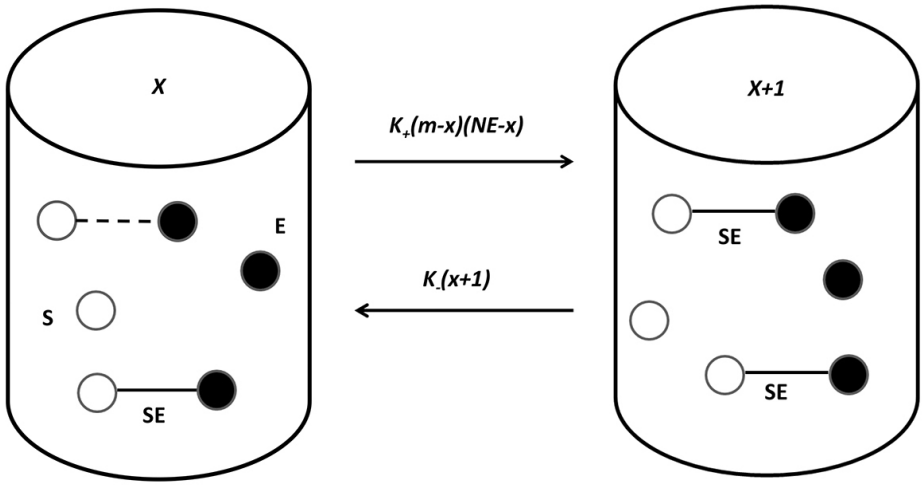


Figure: Transitions associated with the reactions $\mathrm{S} + \mathrm{E} \underset{\kappa_-}{\overset{\kappa_+}{\longleftrightarrow}} \mathrm{SE}$

Let $X(t) = x$: the number of ways of creating a new complex is given by $(m-x)(N_E-x)$. The rate of production of such complexes is hence $(m-x)(N_E-x)\kappa_+$. On the other hand, each of the x complexes has a probability of dissociating during small time intervals, so that the rate of dissociation is $\kappa_- x$.

Let

$$K = \frac{\kappa_+}{\kappa_-}.$$

This is a birth and death process of rates given by

$$\lambda_x = \kappa_+(m-x)(N_E-x) \text{ and } \nu_x = \kappa_- x.$$

The related master equation describing the time evolution of the probability $P_x(t) = P(X(t) = x)$ is

$$\begin{aligned}\frac{\mathrm{d}P_{x+1}(t)}{\mathrm{d}t} &= (m-x)(N_E-x)\kappa_+ P_x(t) + \kappa_-(x+2)P_{x+2}(t) \\ &\quad -(\kappa_-(x+1) + \kappa_+(m-(x+1))(N_E-(x+1)))P_{x+1}(t).\end{aligned}$$

$$\lambda_x = \kappa_+(m-x)(N_E-x) \text{ and } \nu_x = \kappa_- x.$$

A direct computation shows that the related steady state is

$$\pi(x) = P_x(\infty) = \frac{m!N_E!K^x}{x!(m-x)!(N_E-x)!}\frac{1}{Z_{m,N_E}}, \tag{4}$$

for a normalization constant

$$Z_{m,N_E} = \sum_{x\geq 0} \frac{m!N_E!K^x}{x!(m-x)!(N_E-x)!}.$$

Metabolic networks

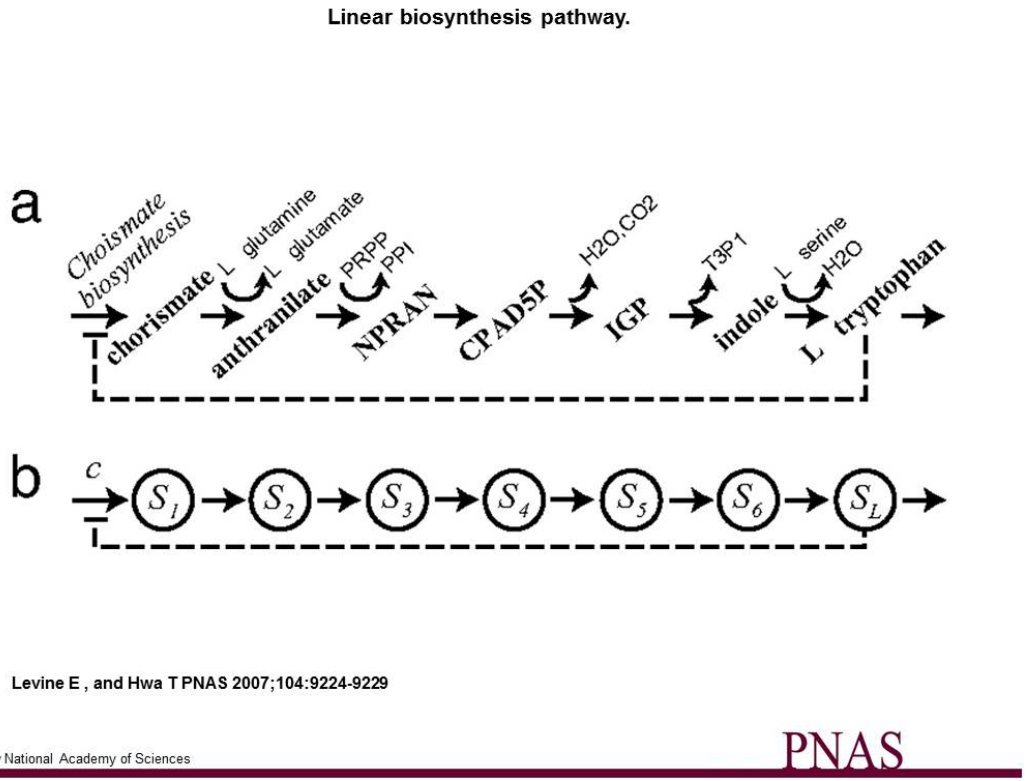


Figure: (a) Linear biosynthesis pathway: Tryptophan pathway in *E. Coli*. (b): linear chain of enzymatic reactions with a negative feedback loop from end product to chorismate.

Metabolic networks

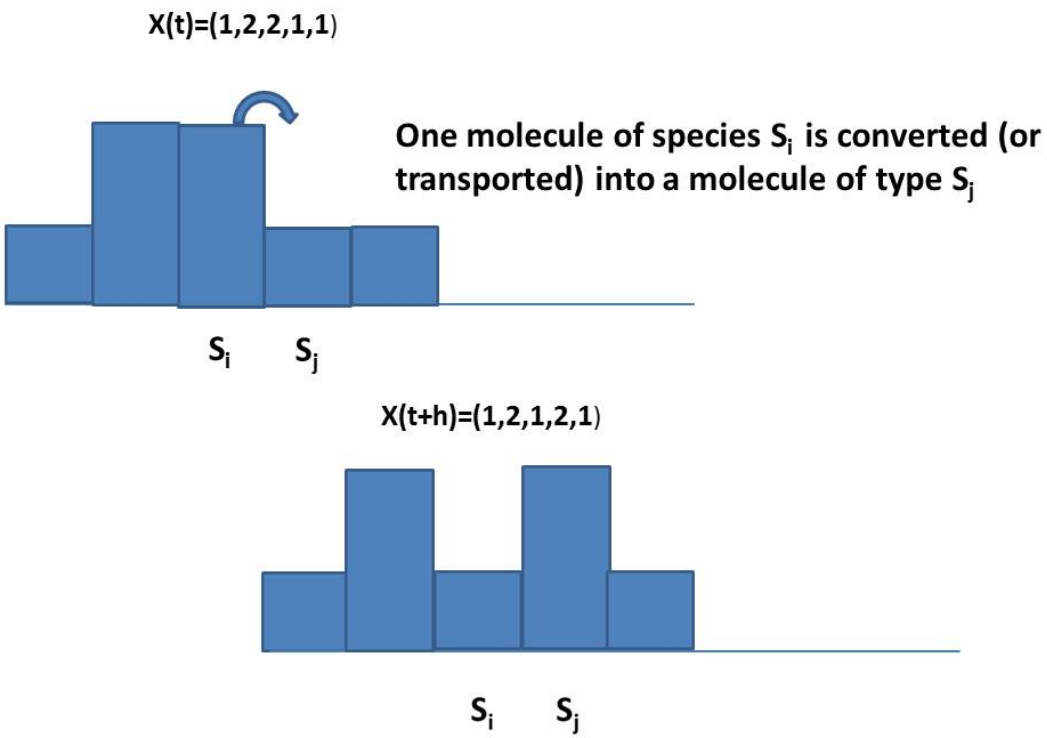


Figure: Markov chain $X(t) = (X_1(t), \cdots, X_n(t))$ describing the metabolic dynamics

Gene network

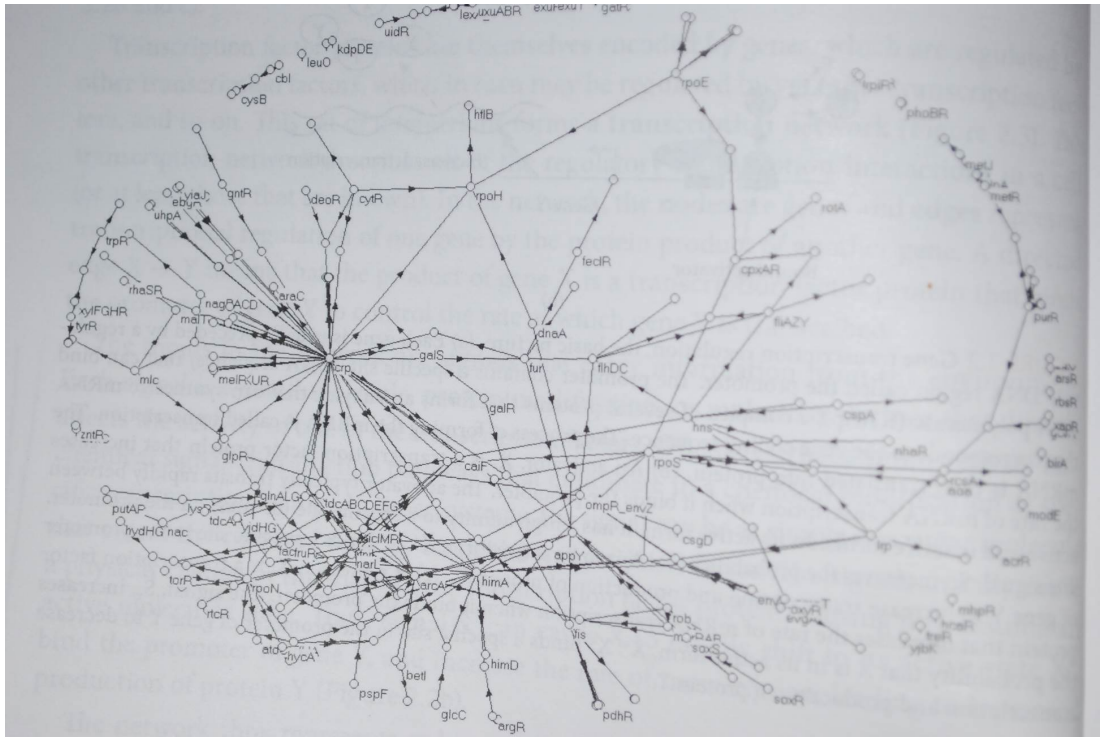


Figure: Markov chain $X(t) = (X_1(t), \cdots, X_n(t))$ giving the abundances of the various protein species

E Coli network

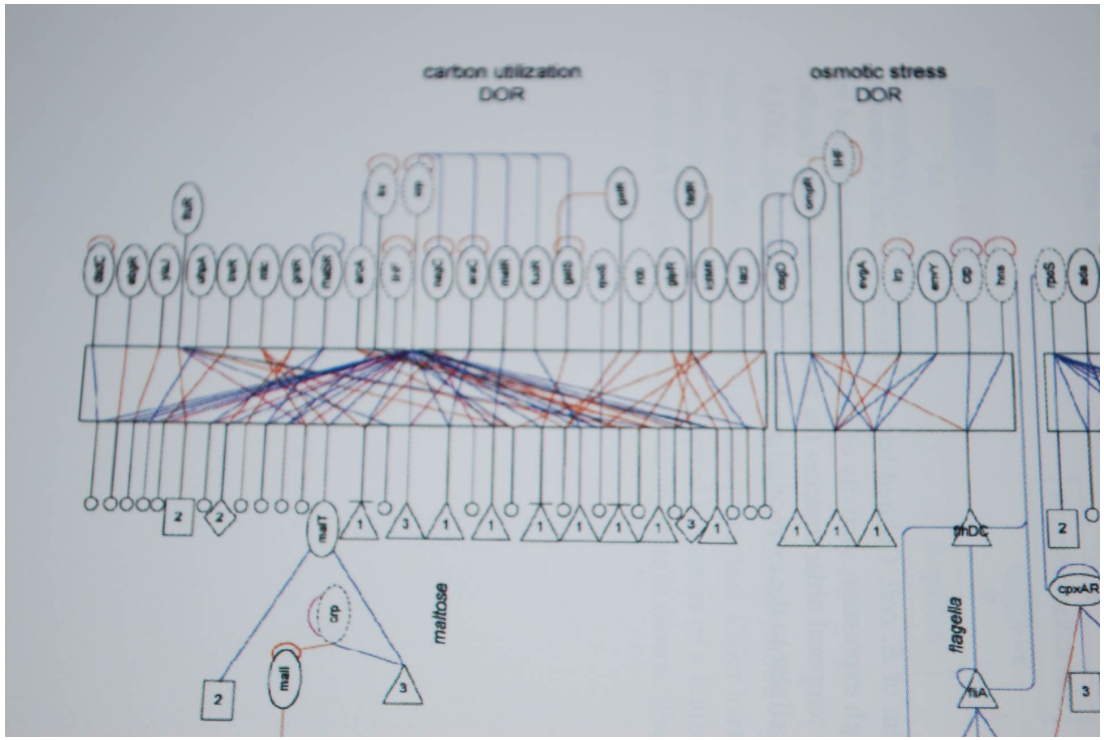


Figure: Markov chain $X(t) = (X_1(t), \cdots, X_n(t))$ giving the abundances of the various protein species. Here the logical gates associated with flagella formation in *E Coli*

General Markov chains

$$X(t) \in \Lambda$$

is defined using transition rates $q_{xy} \geq 0$, x, $y \in \Lambda$, which are encoded in a **generator matrix**

$$Q: \ \Lambda \ \mathrm{x} \ \Lambda \longrightarrow \mathbb{R}, \ \ Q = \{q_{xy}, \ x, \ y \in \Lambda\},$$

such that

$$q_{xx} = -\sum_{y \neq x} q_{xy} \text{ and } \sum_{y \in \Lambda} q_{xy} \equiv 0. \tag{5}$$

$$P(X(t+h) = y | X(t) = x) \approx q_{xy} h, \tag{6}$$

as $h \approx 0$

Kolmogorov equation

The transition function

$$P_{xy}(t) = P(X(t) = y | X(0) = x)$$

solves the master equation

$$\frac{\mathrm{d}P_{xy}(t)}{\mathrm{d}t} = \sum_{z \neq y} \Big(P_{xz}(t) q_{zy} - P_{xy}(t) q_{yz} \Big). \tag{7}$$

Let

$$P(t) = (P_{xy}(t))_{x,y\in\Lambda}.$$

Then

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = P(t)Q, \quad P(0) = \mathrm{id}.$$

The limiting behaviour of $P(t)$ is obtained by stationary probability measures π solving the linear equation

$$\pi Q = 0.$$

Convergence to steady state

Theorem[Convergence to steady state] Let Q be an irreducible generator defined on the countable set Λ, of unique invariant probability measure π. Assume that $\sup_x \sum_{z \neq x} q_{xz} < +\infty$. Then

$$P_{xy}(t) = P(X(t) = y | X(0) = x) \longrightarrow \pi(y), \ t \to \infty. \tag{8}$$

Substate-Enzyme-Product

To model metabolic pathways, one needs reaction involving biochemical reactions like

$$\emptyset \xrightarrow{\mu} \mathrm{S_i},\ \mathrm{S_i} + \mathrm{E_i} \underset{\kappa_-^i}{\overset{\kappa_+^i}{\longleftrightarrow}} \mathrm{SE_i} \xrightarrow{\kappa_2^{ij}} \mathrm{S_j}. \tag{9}$$

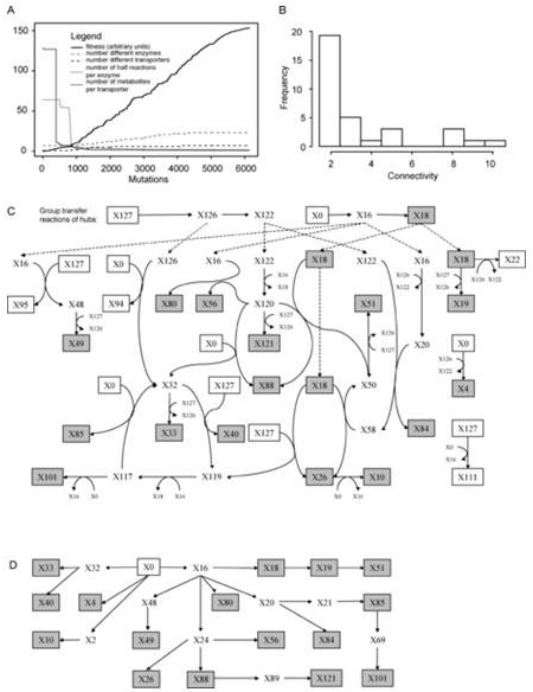
Markov chain $X(t)$ where

$$X(t) = (X_1(t), \cdots, X_n(t)),$$

where $X_i(t)$ gives the number of substrate molecules of type S_i at time t.

Metabolic networks

Figure 2. Sample Simulation for the Evolution of Metabolic Networks



Pfeiffer T, Soyer OS, Bonhoeffer S (2005) The Evolution of Connectivity in Metabolic Networks. PLoS Biol 3(7): e228. doi:10.1371/journal.pbio.0030228
http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0030228

PLOS | BIOLOGY

Figure: Directed graph associated with a metabolic pathway

Metabolic networks

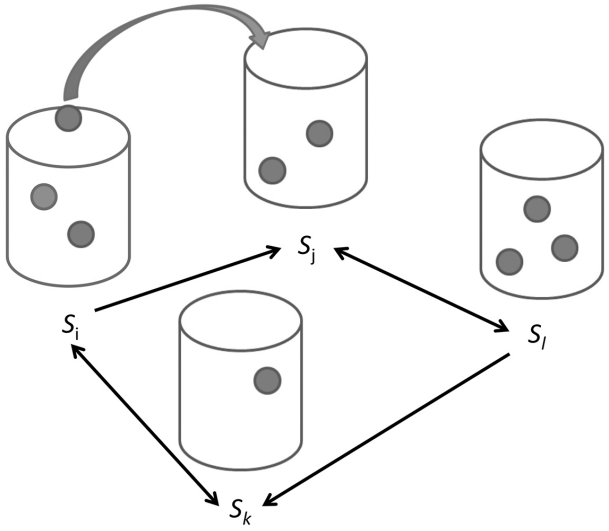


Figure: A substrate molecule of type *i* is being transformed into a type *j* substrate molecule

The actual knowledge on such stochastic processes is very limited at present time.

Let e_i denote the ith unit vector of $\mathbb{R}^M$, that is, the vector having all components equal to 0 except the ith one, which is equal to 1. The transitions of the particle system are of the form

$$(x \to x - e_i + e_j) \text{ when } (S_i \to S_j) \in \mathcal{E},$$

meaning that a particle of species S_i has been converted into a molecule of type S_j, at rate

$$q_{x,x-e_i+e_j} = \kappa_{ji} \nu_{x_i},$$

where

$$\nu_x = \kappa_2 \frac{x}{x + (K - N_E - 1)}, \tag{10}$$

which is the so-called **Michaelis Menten** law.

It turns out that, for certain graph structures, the steady state limiting distribution of this particle system factorizes as

$$\begin{aligned} \lim_{t\to\infty} P(X_1(t) = x_1, \cdots, X_n(t) = x_n) &= \pi(x_1, \cdots, x_n) \\ &= \pi_1(x_1)\cdots\pi_n(x_n), \end{aligned}$$

where the marginal laws are of the form (for Michaelis Menten kinetics)

$$\pi(m) = \binom{m+K+N_E-1}{m}(1-z)^{K+N_E m}, \tag{11}$$

where $z = \mu/\nu_{\max}$.

- Factorization implies independence: the random fluctuations of the abundance of some species do not influence the other species
- The marginal distribution of S_i only depends on enzyme E_i: this allows regulation of species S_i by fine tuning the related enzyme
- Perhaps the effect of evolution to adapt to fluctuating environments
- Information on the metabolic network can't be obtained from steady state data

Signalling cascades

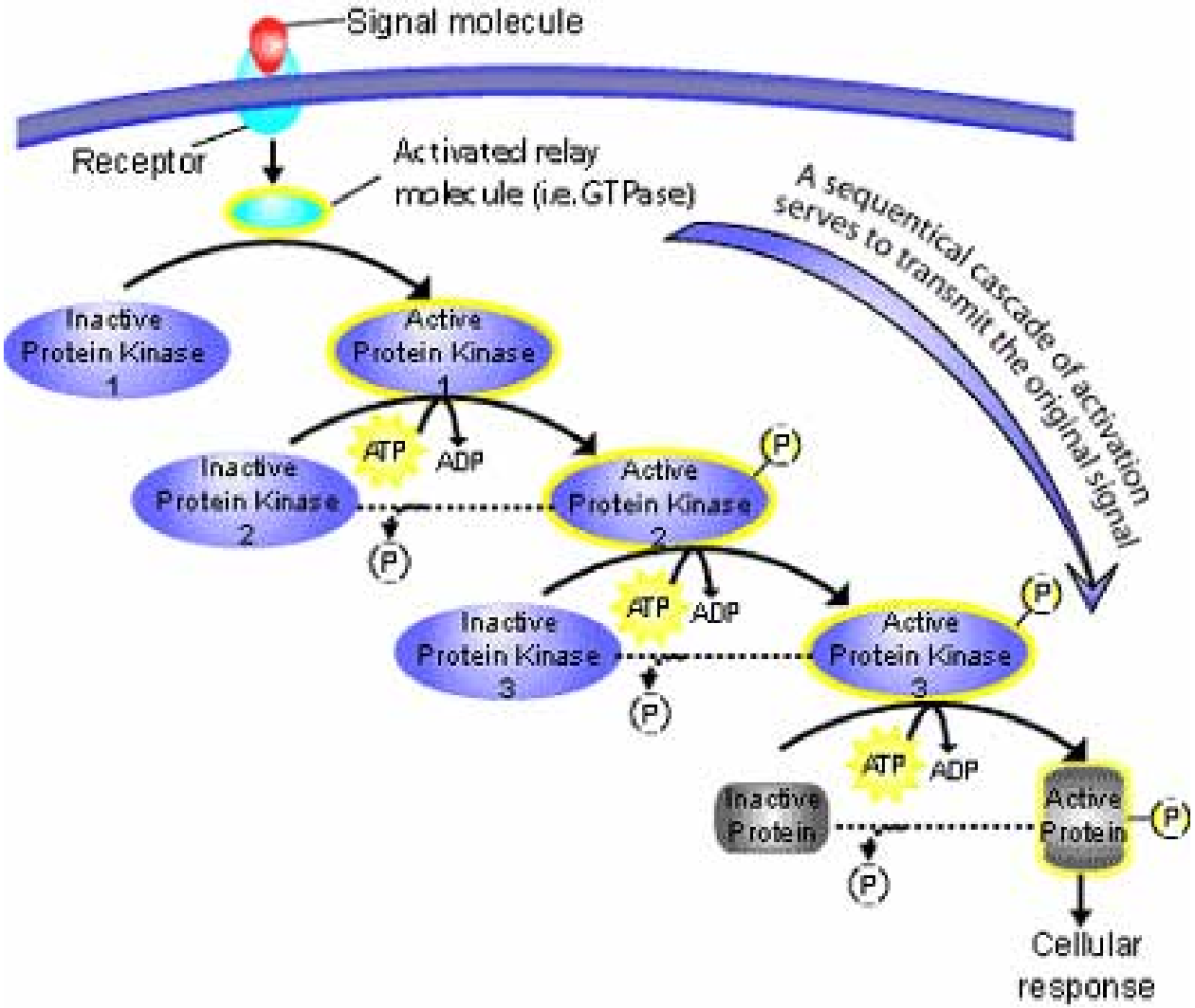


Figure: A sequential signalling cascade

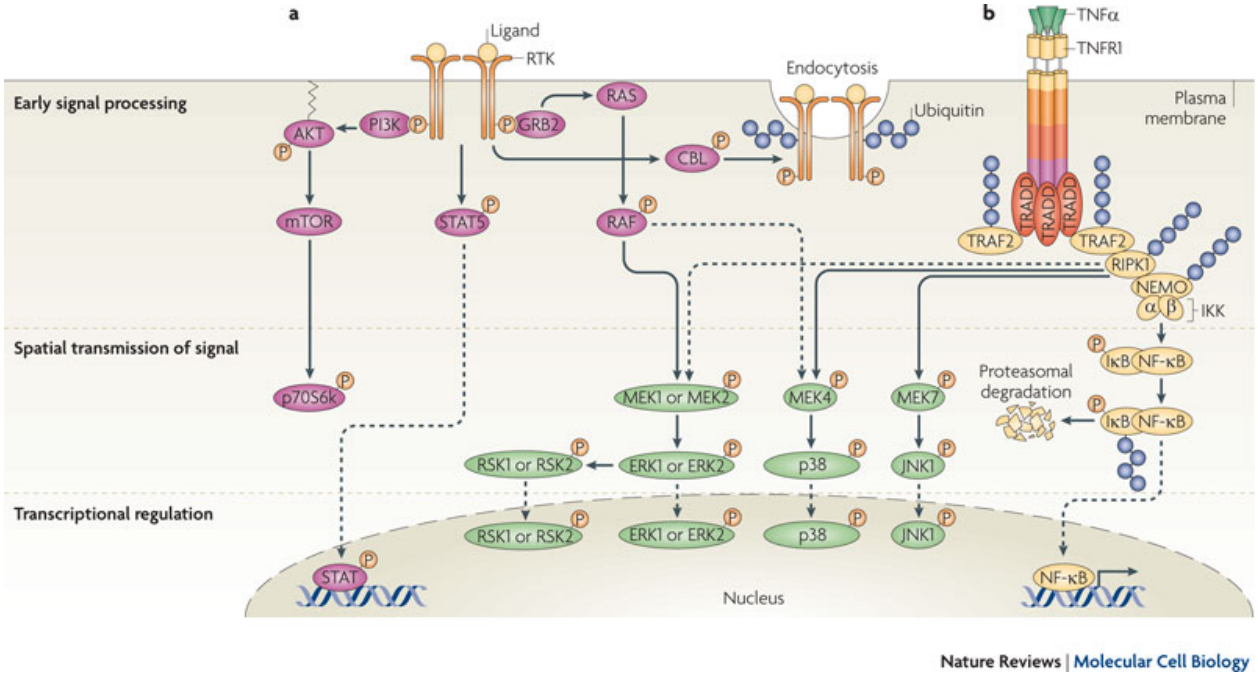


Figure: Signalling cascade

Biochemical signal processing

An input signal $R(t)$ arrives at the cell membrane (pheromone, growth factors, chemical agent...) This information will be transduced or sent toward **transcription factors (TF), which will induce the expression of some target genes in response to this signal.**

- Usually, the signal concentration is weak. Nature found a way of amplifying this input in order to activate properly TF
- This amplification is realized using kinases
- These kinases switch randomly between active and silent states
- Active protein kinases induce the activation of other kinases
- Complex graph structures

Biochemical signal processing

Main research questions

- Amplification
- Speed of transmission
- Crosstalk between antennas, interferences
- Noise filtering

Exercises

A stochastic model

The first kinase is activated by some external signal $R(t)$, $t \geq 0$. Let $Y_1(t)$ be a Bernoulli random variable describing the activity of the first kinase at time t, with $Y_1(t) = 1$ when it is active, or phosphorylated, and $Y_1(t) = 0$ otherwise. The related dynamic might be described by a time non-homogeneous two-state Markov chain. The transition rate from the OFF state to the ON state is proportional to the signal amplitude, of the form $\tilde{\alpha}_1 R(t)$. The reverse transition is described by a constant rate β_1. This reaction is then modelled by the relation

$$O_0^1 \underset{\beta_1}{\overset{\tilde{\alpha}_1 R(t)}{\longleftrightarrow}} O_1^1. \tag{12}$$

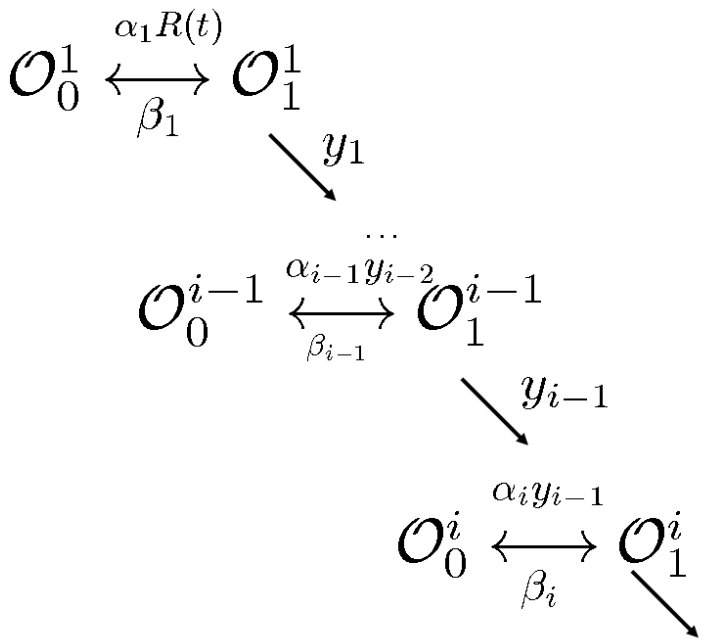


Figure: The signalling cascade. The phosphorylation rate of the first kinase depends on the external signal $R(t)$. $\mathcal{O}_0^i$ denotes the OFF state, and $\mathcal{O}_1^i$ the ON state.

Chemical reactions/Markov chain

The state of the system at time t is described by the random binary vector $Y(t) = (Y_1(t), \cdots, Y_M(t))$. Let $E_i(t) = \mathbb{E}(Y_i(t)) = P(Y_i(t) = 1)$. The related master equation yields that

$$\frac{\mathrm{d}E_i(t)}{\mathrm{d}t} = \tilde{\alpha}_i \mathbb{E}((1 - Y_i(t))Y_{i-1}(t)) - \beta_i E_i(t),$$

when $i \geq 2$. Due to statistical correlations,

$$\mathbb{E}((1 - Y_i(t))Y_{i-1}(t)) \neq \mathbb{E}((1 - Y_i(t)))\mathbb{E}(Y_{i-1}(t)) = (1 - E_i(t))E_{i-1}(t),$$

in general. If these random variables are not correlated, one obtains the o.d.e.

$$\frac{\mathrm{d}E_i(t)}{\mathrm{d}t} = \tilde{\alpha}_i (1 - E_i(t))E_{i-1}(t) - \beta_i E_i(t).$$

Simplified model

Heinrich, Neel and Rapoport(2002) proposed to study this last o.d.e., using variables $X_i(t)$, describing kinase abundances: let C_i denote the total number of type i kinases, which can be active or not. Their model is given by the set of differential equations

$$\frac{\mathrm{d}X_1(t)}{\mathrm{d}t} = \tilde{\alpha}_1 R(t)(C_1 - X_1(t)) - \beta_1 X_1(t),$$

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = \tilde{\alpha}_i X_{i-1}(t)(C_i - X_i(t)) - \beta_i X_i(t),\ i \geq 2.$$

The **signalling time** τ_i is defined to be the average time needed to activate a kinase of type i.

$$f_i(t) = \frac{X_i(t)}{\int_0^\infty X_i(s)\mathrm{d}s}, t > 0.$$

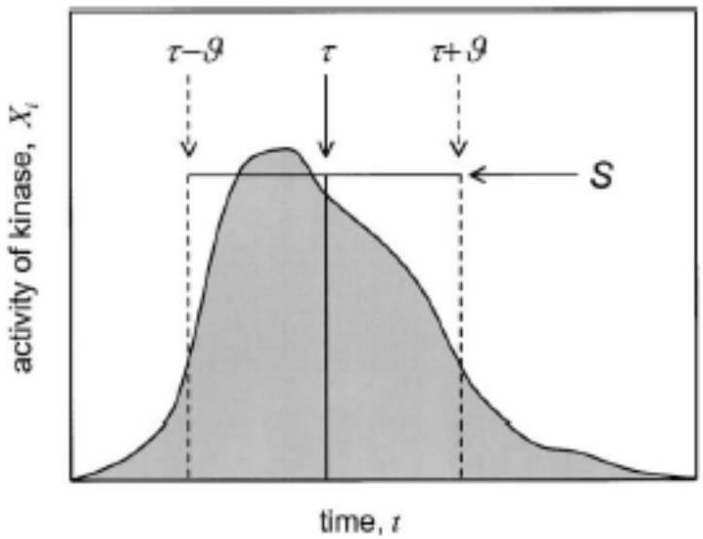


Figure: Density $f_i(t)$ associated with the ith kinase

Let $T_i > 0$ be a random variable of density f_i. The signalling time τ_i is the expected value of T_i

$$\tau_i = \mathbb{E}(T_i) = \int_0^\infty t f_i(t) \mathrm{d}t.$$

The **signal duration** is the related standard deviation

$$\sigma_i = \sqrt{\mathrm{Var}(T_i)} = \sqrt{\mathbb{E}(T_i^2) - (\mathbb{E}(T_i))^2}.$$

The **signal amplitude** A_i,

$$A_i = \frac{\int_0^\infty X_i(s)\mathrm{d}s}{2\sigma_i},$$

which is the height of a rectangle of length $2\sigma_i$ of area $I_i = \int_0^\infty X_i(s)\mathrm{d}s$.

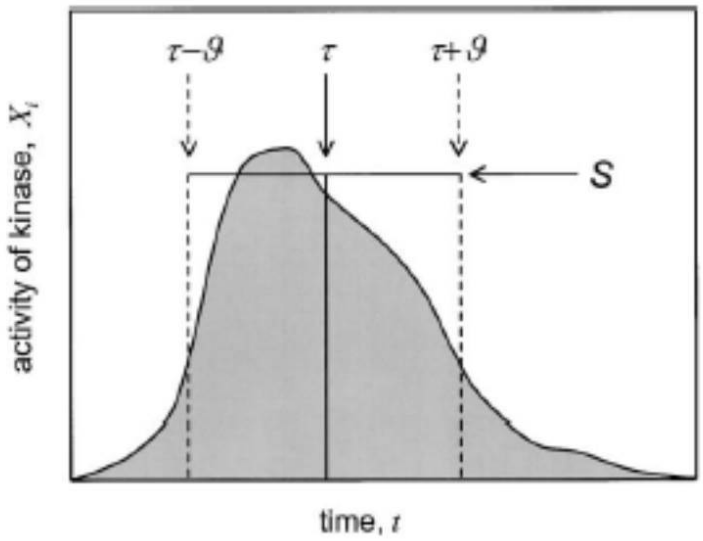


Figure: Signal amplitude

Weakly activated pathways

Assume that $X_i(t)/C_i \approx 0$ and $X_i(0) = 0$, $\forall i$, and that $R(t) = R(0)\exp(-\lambda t)$, $\lambda > 0$, Heinrich et al. have proven that

$$\tau_M = \frac{1}{\lambda} + \sum_{i=1}^{M} \frac{1}{\beta_i}, \tag{13}$$

which does not depend on the phosphorylation rate α_i. Moreover, all the phosphatases have the same effect on the signalling time, regardless of their position in the pathway. They also observed the same phenomenon for the signal duration,

$$\sigma_M = \sqrt{\frac{1}{\lambda^2} + \sum_{i=1}^{M} \frac{1}{\beta_i^2}}. \tag{14}$$

In contrast, the signal amplitude depends on all the pathway parameters

$$A_M = \frac{R(0)\prod_{i=1}^{M}\frac{\alpha_i}{\beta_i}}{2\sqrt{1+\lambda^2\sum_{i=1}^{M}\frac{1}{\beta_i^2}}}, \tag{15}$$

where we set

$$\alpha_i = C_i\tilde{\alpha}_i.$$

The signal can be amplified, that is, $A_{i-1} < A_i$ when

$$\beta_i < \alpha_i\sqrt{1-\frac{1}{\alpha_i^2\sigma_{i-1}^2}},$$

so that there is some amplification when the dephosphorylation rate β_i is small relative to α_i.

Optimal cascade length

Given some amplification level, there is an optimal cascade length that minimizes signalling time:

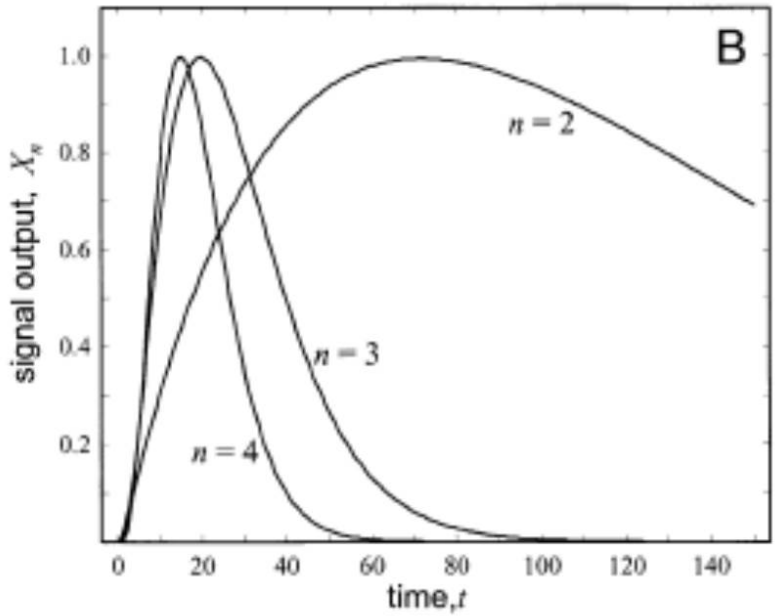


Figure: Parameters are chosen to achieve a given amplification level. Long cascades can lead to fast signal transmission

Exercise 1

The differential system associated with weakly activated linear activation cascades is

$$\frac{\mathrm{d}X_1(t)}{\mathrm{d}t} = \alpha_1 R(t) - \beta_1 X_1(t),$$

$$\frac{\mathrm{d}X_i(t)}{\mathrm{d}t} = \alpha_i X_{i-1}(t) - \beta_i X_i(t),\ i \geq 2,$$

where we assume that all the kinases are inactive at time $t = 0$, that is, we set $X_i(0) = 0$, $\forall i$. Consider an exponentially decreasing input signal $R(t) = R(0)\exp(-\lambda t)$, $R(0) > 0$, for some parameter $\lambda > 0$, which is such that $\lambda \neq \beta_i$, $\forall i$.

- Use the method of variation of constants to deduce that

$$X_1(t) = \alpha_1 R(0) \frac{e^{-\lambda t} - e^{-\beta_1 t}}{\beta_1 - \lambda} \longrightarrow 0,$$

  exponentially fast as $t \to \infty$.

Exercise 1

- Consider the density function f_i on $\mathbb{R}_+$,

$$f_i(t) = \frac{X_i(t)}{Z_i},$$

where

$$Z_i = \int_0^\infty X_i(s)\mathrm{d}s.$$

Show that

$$Z_i = \frac{\alpha_i}{\beta_i} Z_{i-1},\ i \geq 1.$$

- Deduce from this that

$$Z_i = \frac{R(0)}{\lambda} \prod_{k=1}^{i} \frac{\alpha_k}{\beta_k}.$$

- Use the definition of f_i to show that the signalling times τ_i satisfy the recursion

$$\tau_i = \tau_{i-1} + \frac{1}{\beta_i},$$

and deduce (13)

Exercise 2: Steady state of birth and death processes

Consider a birth and death processes of parameters λ_j and ν_j which starts within the interval $\Lambda = \{0, \cdots, M\}$. Assume that

$$\lambda_j = 0,\ j \geq M \text{ and that } \nu_0 = 0,\ \nu_j > 0,\ 1 \leq j \leq M.$$

- Show that the process will stay in Λ forever.
- Show that the steady state distribution $\pi_j,\ 0 \leq j \leq M$ which solves (3) is given by

$$\pi_j = \frac{\prod_{i=0}^{j-1} \lambda_i \prod_{i=j+1}^{M} \nu_i}{\sum_{j=0}^{M} \prod_{i=0}^{j-1} \lambda_i \prod_{i=j+1}^{M} \nu_i}.$$

- In the special case where $\lambda_j \equiv \mu$ and $\nu_j \equiv \nu j$, show that the steady state is

$$\pi_j = \frac{\frac{\lambda^j}{j!}}{\sum_{j=0}^{M} \frac{\lambda^j}{j!}},$$

  where $\lambda = \mu/\nu$.
- Deduce from this that π is a Poisson distribution of parameter λ when $M \to \infty$ (see Exercise 4 for a verification through simulations).

The Gillespie algorithm

For a general Markov chain $X(t)$ of generator $Q = (q_{xy})$

$$P(X(t+h) = y | X(t) = x) \approx q_{xy} h.$$

The Gillespie Algorithm uses Q to simulate the random trajectories of $X(t)$.
One defines the jump matrix $P = (p_{xy})_{x,y \in \Lambda}$ associated with the generator Q is the (stochastic) matrix defined by

$$p_{xy} = \frac{q_{xy}}{\sum_{z \neq x} q_{xz}} \text{ when } x \neq y \text{ and } \sum_{z \neq x} q_{xz} > 0.$$

Let also

$$q(x) = \sum_{y \neq x} q_{xy} = -q_{xx}.$$

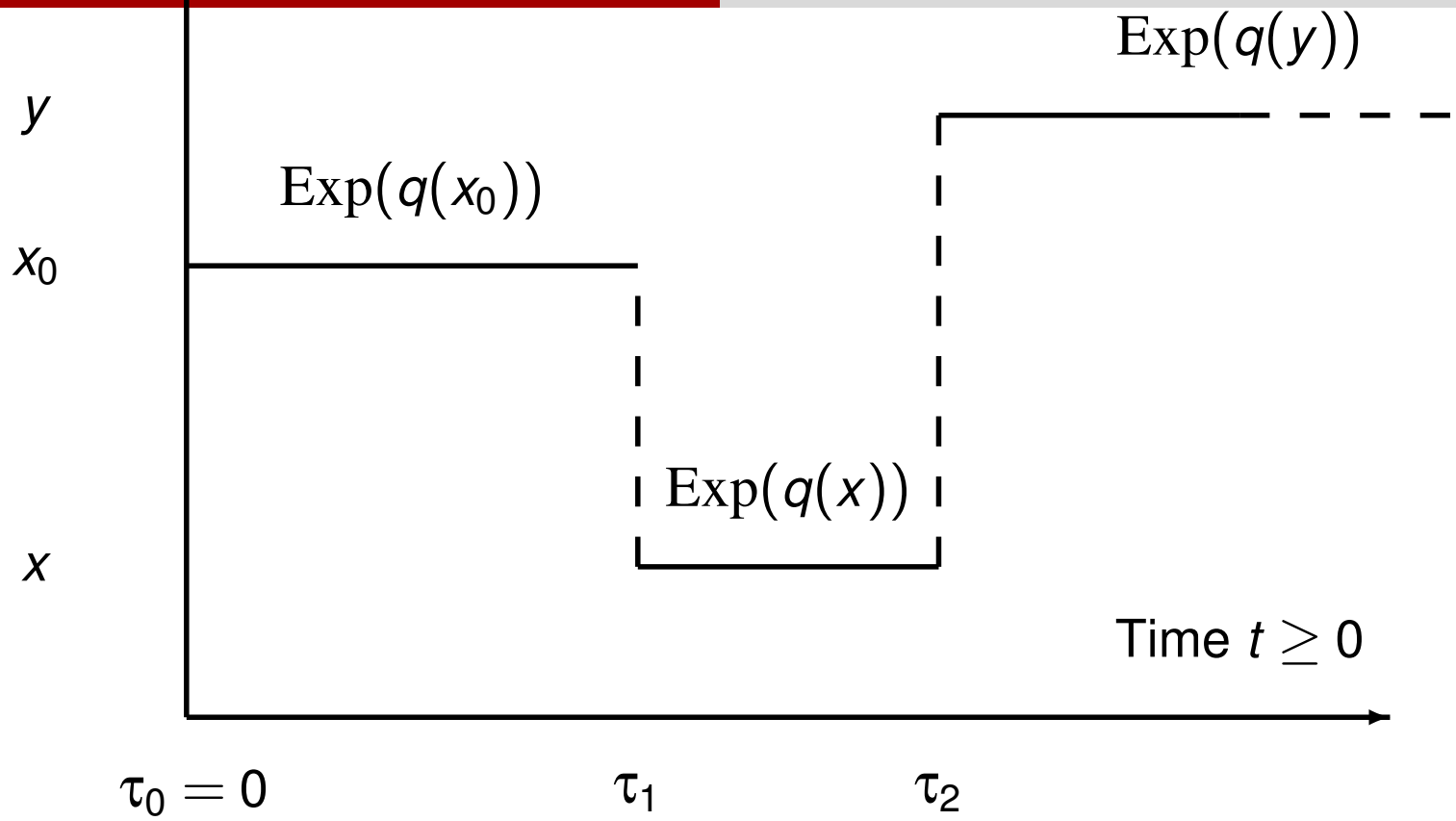


Figure: The Markov chain starts with $X(0) = x_0$. It waits there an exponential time of parameter $q(x_0)$, and then jumps with probability $p_{x_0 x}$ at time τ_1 to the new state x, so that $X(\tau_1) = x$. It then waits an exponential time of parameter $q(x)$ to jump to the new state y, with $X(\tau_2) = y$ with probability p_{xy}

Exercise 3: The first reaction method

Prove the following result:

Theorem: Let T_k be independent exponential random variables of parameters q_k, such that $q = \sum_k q_k < \infty$ and $q > 0$. Let

$$T = \inf_k T_k.$$

The above infimum is attained at a unique index k_* with probability 1. T and k_* are independent and

- T is exponential of parameter q,
- and $P(k_* = k) = \frac{q_k}{q}$.

Exercise 3: The first reaction method

Given that the process arrives at x at time τ, the process stays an exponential time of parameter $q(x)$ at x, and then jumps to a new state y with probability $p_{xy} = q_{xy}/q(x)$, where $q(x) = \sum_{z \neq x} q_{xz}$.

Use the Theorem to check that the above procedure is equivalent to, setting $q_k = q_{xz}$ and $q = q(x)$,

- Draw independent exponential random variables T_{xz} of parameter q_{xz}
- Look for the index z_* such that $T_{xz_*} = \inf_z T_{xz}$,
- Set the waiting time at x as T_{xz_*},
- and let the process jump from x to z_*.

Exercise 4: Gene expression

Consider the following set of chemical reactions (or transitions)

$$\emptyset \xrightarrow{\mu} A, \tag{16}$$

which corresponds to the creation of a molecule of type A, and, for degradation,

$$A \xrightarrow{\nu n} \emptyset. \tag{17}$$

when the system contains n molecules of type A. Let $X(t)$ be the number of type A molecules at time t. Assume that the process arrives at $x = n$ at time τ. The possible transitions are (see Exercise 2)

- $n \to n+1$ at rate $q_{nn+1} = \mu$,
- and $n \to n-1$ at rate $q_{nn-1} = \nu n$.

Exercise 4: Gene expression

Check that the Gillespie algorithm is equivalent to the following pseudo-code:

- Generate two random numbers r_1 and r_2 uniformly in the unit interval $[0,1]$.
- Set $q(x) = x\nu + \mu$.
- Stay at x the random time $\triangle$, where

$$\triangle = \frac{1}{q(x)} \ln(\frac{1}{r_1}).$$

- Jump then according to the following rule:

$$X(t+\triangle) = \begin{cases} X(t)+1 & \text{when } r_2 < \frac{\mu}{q(x)}, \\ X(t)-1 & \text{when } r_2 > \frac{\mu}{q(x)}. \end{cases}$$

Implement this pseudo-code and check through simulations that the steady state distribution is Poisson of parameter μ/ν.

Exercise 5: Second order reactions

Consider a system composed of type A and type B molecules, and let us denote by $X_A(t)$ and $X_B(t)$ the number of these molecules present in the system at time t. The set of chemical reactions is

$$A + A \xrightarrow{\kappa_1 X_A(X_A-1)} \emptyset,$$

$$A + B \xrightarrow{\kappa_2 X_A X_B} \emptyset,$$

$$\emptyset \xrightarrow{\kappa_3} A,$$

$$\emptyset \xrightarrow{\kappa_4} B.$$

Set

$$\alpha_1 = \kappa_1 X_A(X_A - 1),\ \alpha_2 = \kappa_2 X_A X_B,\ \alpha_3 = \kappa_3,\ \alpha_4 = \kappa_4.$$

Exercise 5: Second order reactions

Check that the Gillespie algorithm for this set of reactions is equivalent to the following pseudo-code:

- Generate two random numbers r_1 and r_2 uniformly in the unit interval $[0,1]$.
- Set $\alpha_0 = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = q(X)$, where $X(t) = (X_A, X_B)$.
- Stay at X the random time $\triangle$, where

$$\triangle = \frac{1}{q(x)} \ln(\frac{1}{r_1}).$$

- Jump then according to the following rule:

$$X(t+\triangle) = \begin{cases} (X_A - 2, X_B) & \text{when } 0 < r_2 < \frac{\alpha_1}{\alpha_0}, \\ (X_A - 1, X_B - 1) & \text{when } \frac{\alpha_1}{\alpha_0} < r_2 < \frac{\alpha_1+\alpha_2}{\alpha_0}, \\ (X_A + 1, X_B) & \text{when } \frac{\alpha_1+\alpha_2}{\alpha_0} < r_2 < \frac{\alpha_1+\alpha_2+\alpha_3}{\alpha_0}, \\ (X_A, X_B + 1) & \text{when } r_2 > \frac{\alpha_1+\alpha_2+\alpha_3}{\alpha_0}. \end{cases}$$