Alma Mater Studiorum · University of Bologna

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Phd in Physics
SETTORE CONCORSUALE : 02/B3 - FISICA APPLICATA
SETTORE SCIENTIFICO-DISCIPLINARE : FIS/07 - FISICA APPLICATA

Stochastic models and dynamic measures for the characterization of bistable circuits in cellular biophysics

Thesis Advisor:
Chiar.mo Prof.
GASTONE CASTELLANI

Presented by:
ENRICO GIAMPIERI

PhD coordinator:
Chiar.mo Prof.
FABIO ORTOLANI

aa 2011/2012
January 13, 2012

To my family,
whose constant yet discreet support
made all this possible

Introduction

my cells are smarter than me

Lorenzo Farinelli, MD trainee

During the last few years, a great deal of interest has risen concerning the applications of stochastic methods to several biochemical and biological phenomena. This interest comes from the observation, also experimentally founded, that almost every biological process is intrinsically noisy and cannot therefore be described with purely deterministic methods such as Ordinary Differential Equation (ODE). Even taking this into consideration, phenomena like gene expression, cellular memory, bet-hedging strategy in bacterial growth and many others, cannot be described by continuous stochastic models due to their intrinsic discreteness. This feature limits the feasibility of a more standard approach like the Fokker-Plank equation formalism, leading to a progressive introduction of the Master Equation technique in the field of biological modeling due to its capacity of describing discrete stochastic processes. In this thesis I have used the Chemical Master Equation (CME) technique to modelize some feedback cycles and analyzing their properties, including experimental data.

In the first part of this work, the effect of stochastic stability is discussed on a toy model of the genetic switch that triggers the cellular division, which malfunctioning is known to be one of the hallmarks of cancer. Both the presence of a secondary peak without a corresponding deterministic stable point and the reciprocal effect are discussed along, with their biological relevance.

The second system I have worked on is the so-called futile cycle, a closed cycle of two enzymatic reactions that adds and removes a chemical compound, called phosphate group, to a specific substrate. This system is deterministically monostable, but it has been recently shown to be able to exhibit bistable properties when an appropriate level of noise is added to the system. I have thus investigated how adding noise to the enzyme (that is usually in the order of few hundred molecules) modifies the probability of observing a specific number of phosphorylated substrate molecules, and confirmed theoretical predictions with numerical simulations.

In the third part the results of the study of a chain of multiple phosphorylation-dephosphorylation cycles will be presented. This system can exhibit multistability even in the deterministic regime, and we have studied under which conditions this is valid even in the discrete stochastic regime. We will discuss an approximation method for the exact solution in the bidimensional case and the relationship that this method has with the thermodynamic properties of the system, which is an open system far from equilibrium.

In the last section the agreement between the theoretical prediction of the total protein quantity in a mouse cells population and the observed quantity will be shown, measured via fluorescence microscopy. This quantity is observed during the process of cellular senescence and its compared with the behavior of other special proteins like the histones, the protein structure responsible for the chromatin folding, and the laminin-A, a nuclear membrane protein.

Contents

Chapter 1

The master equation

In this chapter I will show where the master equation comes from and how it can be used to analyze biochemical system where the standard, deterministic analysis would give inaccurate results. Starting with a review of the biological basis for the usage of this approach and its mathematical foundations, I will briefly review the most common resolution techniques, both analytical and numerical.

1.1 Noise in biology

In biology there are several processes that cannot be described in term of deterministic evolution. From protein production to the behavior of the whole cell, not only the noise is ever-present, but evolution found several ways to exploit this noise to the advantage of the single cell or the whole population.

1.1.1 Bet hedging

One of the most intriguing example of exploitation of the stochasticity is the so-called "bet hedging strategy" which can be found in several bacteria population. Bacteria are a tiny yet spectacular form of life that can thrive almost everywhere, from ocean-deep and oxygen deprived boiling pits to the human gut, where their function is so crucial that a lot a scientist are starting to refer to our symbiotic partners as *microbiome* and referring to it as an additional organ of our body, not less important for our health than the liver or the lungs.

Most environments are very different from our gut, where our symbiotic bacteria can thrive in a food-filled environment, always with the right conditions of temperature, moisture and pH. Outside our body the conditions are quite harsh, with wide temperature variation, a varying range of pH and moisture, and almost no nutrients. This could severely limit the ability to survive of the bacteria, due to the frequent changes of environmental conditions. The response of a lot of species of bacteria to this problem is the "sporulation" process, where the cell freezes itself in a resistant state waiting for better conditions. Microbiologists have found spores several million of years old and still capable of returning to an active state[13].

The main problem with this approach is that the sporulation process is not instantaneous, but it takes some time to happen (minutes to hours, usually),

and this can lead to a nearly 100% extermination of the population in case of rapidly changing environment. What biologists observed is that in any bacterial population, a fraction of this population is always in the resistant state, whatever the conditions was. This allowed a certain percentage of the population to survive no matter how fast the environment fluctuation were. This was true even for monoclonal population and in general a fixed percentage of the population enters this sporulation phase regardless of the state of the ancestor.

What has been understood is that bacterial cells undergo a transition toward sporulation state and back to the reproductive state with a certain fixed probability. The population that doesn't exploit this method will have a competitive advantage in the short term, but in the end will fall victim to the ambient fluctuation, while the sporulating population, albeit slower in growing, will persist to harder perturbations.[28]

1.1.2 Noise controlling protein production

The production of proteins starting from a gene is now understood as a rich and complex mechanism, even without taking into consideration the interaction of the genes with specialized peptides (small proteins) that are called transcription factors. The concentration of a protein in a cell is crucial in order to perform certain functions, as the regulation of their level is one of the most important function in the cell.

The central dogma in the classic molecular biology is that each gene produces a messenger RNA (mRNA) which is translated into a protein by the ribosomes outside the nucleus. It is known now that the process is more complex than that: the mRNA is first spliced to remove the introns (non translated sequences) and joining the exons, in one or more ways; the resulting mRNA will be translated if not removed by a silencing RNA or a micro RNA (miRNA), two different strands of RNA used by the cell to selectively destroy targets mRNA with a certain grade of specificity: a miRNA can target several different mRNA, and a mRNA can be the target of several miRNA. This second layer of RNA reaction is mostly necessary to keep track of the precise amount of mRNA translated, reducing or amplifying the uncertainty of its distribution. This allows a cell to fine-tune the amount of protein produced to a very precise level or to a very broad distribution. This can allow a cell to obtain memory switches from low cost reactions like the futile cycle (see Chap 3).

A seminal work from the Van Oudenaarden group has shown how it is possible to correctly predict the level of a fluorescent protein in *E. Coli* cells linked to the lactose utilization[71]. Since the lactose utilization network is a very well known and alterable gene group, they altered a colony of *E. Coli* to produce a certain amount of GFP (Green Fluorescent Protein) when the lactose sugar is detected and the whole switch is activated, making it quite easy to control the expression of this ensemble of genes. The actual results matched almost perfectly those predicted with an appropriate master equation, which described the protein production process in term of basic reaction: $gene \rightarrow RNA \rightarrow protein$.

Noise can also be involved in the regulation of the spatial distribution of several substances in the cell. The presence and type of noise found in a cell can create regular oscillations in the position of the concentration of protein in several kind of bacteria[19], from the center to the extrema of the cell. This is a major feature of the stochastic regulation, because chemical gradients inside

a cell are the driving force under the embryonic development of all living thing. In fact, most of our structures, from which side is up and which down to the structure of the spinal chord, can be seen as a macroscopic cascade effect from some chemical gradients in the zygote and down on.

1.1.3 Evolutionary theories and population dynamics

The dynamics of a population of individuals can often be represented with a master equation, as the population size is intrinsically a discrete quantity whose evolution in time is driven by random interaction between individuals. Population growth, epidemic diffusion and the evolution, especially in the formulation of neutral theory of evolution, can all be represented on a discrete stochastic basis.

The simplest population growth model treats the individuals as units whose death is a constant process and reproduction is a simple duplication, and it is often used for bacteria with good approximation. A more detailed model that takes into consideration phenomena like male-female interaction, competition for resources and age groups can be written without special difficulties.

The spreading of a contagious factor in a population, like a disease or a successful social behavior, can be represented as a simple Susceptible-Infected model, with a core dynamic where the encounter of an infected individual with a susceptible one can lead to a transmission: $S + I \to 2I$.

The concept of evolution is well known, even if commonly mis-interpreted, as a combination of random mutations, both with reproductive fitness advantage or disadvantage, and natural selection, i.e. the inter-specific competition of the individual of the population for resources, mating and avoiding predators. Evolution can also be driven by purely stochastic effects, as shown by Motoo Kimura thirty years ago in his book "The Neutral Theory of Molecular Evolution"[57], which launched the concept of neutral evolution. This theory states that in a small population most of the mutation are not fixed in the population by a competitive advantage but rather by mere case, as reproduction can spread a trait among a population and fix by mere fluctuation. The actual probability of fixation for a neutral mutation is in the order of N^{-1} where N is the population size. Advantageous mutations spread easier and faster while disadvantageous ones spread slower with a higher extinction probability, but still can be fixed if the population is small enough. This idea of neutral evolution is becoming more and more important in biology, as it gives a null hypothesis to test against on evolutionary research.

A similar theory has been developed in the ecological niche distribution among several habitats[109], and is based on discrete stochastic process starting from simple population dynamic process[47].

1.2 Master equation

The Master equation is an equation that describes changes in time of the probability of the whole system to be in a specific configuration, driven by a memoryless process of transition between states. This whole idea can understood in the light of the continuous time Markov chain theory. In a different way can be seen

as a diffusion process on a regular graph, whose nodes represent the available states and the links stand for the transitions between states.

1.2.1 Markov processes

The basic concept of the master equation is the idea of markovian process. A Stochastic process is defined as a function of the time t and a stochastic variable X where for each value of X we observe a different realization of the stochastic process.

$$Y_X(t) = f(X, t)$$

The process Y can describe any kind of phenomenon, like the state of a subatomic particle moving through matter, the position of a fluctuating grain of pollen on the surface of water or, as we will see, the number of molecules of each kind in a chemical reaction.

We can describe the probability of observing a specific value of the function Y at a time t as the measure of the ensemble of values of X for which the function Y gives the value y at time t.

$$P(y, t) = \int_{-\infty}^{+\infty} \delta(Y(X, t) - y) dX$$

In the most general mathematical formulation of this problem there is no clear relationship between the value of Y at a certain time and its value at the previous moments, so the probability of observing y at the time $t + dt$ given the observation of the value of Y at previous times can be any function of all the previously observed values.

$$P(Y(t + dt)) = P(Y(t + dt)|Y(t*) \forall t* \leqslant t)$$

Although in a markovian process we have a very simple relationship: the probability distribution of observing the value y at a time $t + dt$ is a function of only the state of the system at the time t.

$$P(Y(t + dt)) = P(Y(t + dt)|Y(t))$$

The system loses any kind of information of its state before the present value, and so the markovian systems are usually called memory-less. In general the probability of a observing y after a certain observation y' is a well behaved measure function in the space of y, with a limit to the deterministic dynamic when the P is a delta-shaped function, thus explicitly determining y given the observation of y'.

We can write the previous relation in a more formal way, that gives the probability of a transition of the system at the state y_2 at the time $t + \tau$ given the observation of the state y_1 at the time t. Under the hypothesis that the transition probability does not depend on the moment in time but only on the elapsed time and calling $P(y_2, t + \tau|y_1, t) = T_\tau(y_2|y_1)$, we obtain the so-called Chapman-Kolmogoroff equation for the transition propensity:

$$T_{\tau+\tau'}(y_3|y_1) = \int T_{\tau'}(y_3|y_2) T_\tau(y_2|y_1) dy_2$$

Taking the first order term of the Taylor serie of the $T_{\tau'}(y_3|y_2)$ integral for small τ', we can write it as:

$$T_{\tau'}(y_3|y_2) = (1 - \alpha_0\tau')\delta(y_3 - y_2) - \tau' W(y_3|y_2) + O(\tau'^2)$$

where the δ function represents the fact that for a brief time the system should not move much and the probability is slightly less than 1 with the term $(1-\alpha_0\tau')$ where $\alpha_0(y2) = \int W(y_3|y_2)dy_3$ is the normalization constant and $W(y_3|y_2)$ is the derivative of $T_\tau(y_3|y_2)$ for $\tau' = 0$.

The last equation, put into the Chapman-Kolmogoroff, divided by τ' under the limit of $\tau' \to 0$ gives us the differential form of the CK equation, which is the master equation:

$$\partial_\tau T_\tau(y_3|y_1) = \int [W(y_3|y_2)T_\tau(y_2|y_1) - W(y_2|y_3)T_\tau(y_3|y_1)]dy_2 \tag{1.1}$$

1.2.2 The Chemical Master Equation

The Eq 1.1 is already a form of master equation, but we can write it into a more understandable form: noticing that the equation has been written for a specific y_1 and t_1, we can remove all the redundant index and write a form for the probability of observing the state y at the time t:

$$\partial_t P(y,t) = \int [W(y|y')P(y',t) - W(y'|y)P(y,t)]dy' \tag{1.2}$$

This can be recognized as an influx of probability to the state y from all the "surrounding" (in the sense of connected) states y' and an efflux from y to every state y' to which it can move to. If the system state space is discrete, as when we work with a system with a discrete number of individuals or molecules, we can write the probability as $P_n(t)$ to represent the discreteness of the state space. In this case the master equation can be called Chemical Master Equation (referring to a chemical environment) and it will be written with sums instead of integrals:

$$\partial_t P_n(t) = \sum_{n=0}^{\infty} [\lambda_{n',n} P_{n'}(t) - \lambda_{n,n'} P_n(t)] \tag{1.3}$$

where the λs are simply the discrete version of the W of the continuous equation, with the origin and destination index exchanged, so that $\lambda_{n,n'}$ represents the probability flux from the state n to the state n'.

We can further simplify Eq 1.3 as a linear dynamic system:

$$\partial_t \vec{P}(t) = \Lambda \vec{P}(t) \tag{1.4}$$

where the matrix Λ is called the transition matrix and is defined as:

$$\Lambda_{i,j} = \begin{cases} \Lambda_{i,j} = \lambda_{i,j} & \forall i \neq j \\ \Lambda_{i,i} = -\sum_{j \neq i} \lambda_{i,j} \end{cases} \tag{1.5}$$

This makes Λ a zero determinant matrix by construction, because

$$\sum_j \Lambda_{ij} = 0 \qquad \forall i$$

The zero determinant matrix represents the conservation of probability: a determinant different from zero means that a certain amount of probability would be generated or destroyed, and this is an absurd, given that we are describing the system as a whole. A zero determinant also means that there is at least one zero eigenvalue, whose corresponding eigenvector is the so-called stationary distribution, the distribution to which the stochastic process always converges, i.e. $\partial_t P_n(t) = 0$, as long as the transition propensities λ are not a function of time. If the system is fully connected (cannot be broken into two non communicating pieces) the stationary distribution is guaranteed to be unique. The stationary distribution will be obviously positive, i.e. all its terms are with positive sign and the sum of all its components is 1 (being a probability distribution). All the other eigenvalues will be with negative module, and the corresponding eigenvectors will have total sum of the components equal to zero, as they can be interpreted as the difference between the present distribution and the stationary one, both having total sums of the components equal to 1. A special role is played by the eigenvalue with the smallest absolute value, which it means that its eigenvector is the longest-standing one. This eigenvector is referred as the *metastable* state and its eigenvalue gives a time-scale of the time of convergence to the stationary distribution. It is worth noticing that albeit each eigenvector components decay exponentially with time, the convergence to the stationary distribution can be slower than exponential if a lot of eigenvalues are closer to the metastable one. If the eigenvalues spectrum is closer to an exponential, it can be shown that the practical convergence time is a power law, i.e the convergence mean-time goes to infinity. One extreme case happens in system which describes ecological systems, where the system is unbounded to infinity and the mean value grows exponentially in time, never actually reaching a stationary distribution.

1.2.3 Van Kampen operators

It is often useful to not write the master equation in its expanded form, which is hard to read, understand and thus to work with. A useful way to write the CME has been proposed by N.G. Van Kampen in his seminal book *Stochastic Processes in Physics and Chemistry*, where he defines a couple of operators inspired by the creation-destruction operators found in quantum physics:

These operators are defined as:

$$\mathbb{E}_n^+ P(n) = P(n+1)$$
$$\mathbb{E}_n^- P(n) = P(n-1)$$

and so:

$$(\mathbb{E}_n^+ - 1)P(n) = P(n+1) - P(n)$$
$$(\mathbb{E}_n^- - 1)P(n) = P(n-1) - P(n)$$

So a generic birth-death process can be written in two different but equivalent forms:

$$\begin{aligned}\partial_t P_n(t) &= -(g_n + r_n)P_n(t) + r_{n+1}P_{n+1}(t) + g_{n-1}P_{n-1}(t) \\ &= (\mathbb{E}_n^- - 1)g_n P_n(t) + (\mathbb{E}_n^+ - 1)r_n P_n(t) \\ &= (\mathbb{E}_n^+ - 1)(r_n - \mathbb{E}_n^- g_n)P_n(t)\end{aligned}$$

It is worth noticing that the $(\mathbb{E}_n^+ - 1)$ term counter-intuitively corresponds to a reaction that reduces the value of n (representing a flux from a high state to a lower one), while the $(\mathbb{E}_n^- - 1)$ corresponds to a process that increases its value.

This formalism comes especially handy when one is writing approximation of the master equation, like the Chemical Langevin Equation[39], where it gives a hint on the correlation matrix of the noise, or the generation of the Fokker-Planck equation, being the $\mathbb{E}_n^+$ approximable with a series of derivatives.

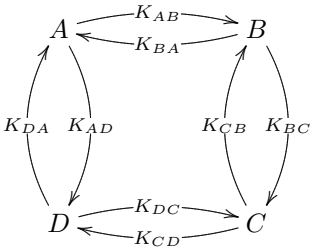
1.2.4 Detailed balance

The condition of detailed balance is one of the most distinguishing property of a system, and corresponds to the thermodynamic equilibrium. The formal definition regards the microscopic probability flux, asserting the microscopy reversibility: $\lambda_{ij}P_i = \lambda_{ji}P_j$. When this relationship holds the resolution of the stationary distribution of the master equation is almost trivial for any dimension, as one can write a "potential" energy of the system and write the solution of the system with the Boltzmann relationship:

$$P_n = P_0 e^{E_n - E_0}$$

This property will be used in Chap 4 to approximate the distribution even when the detailed balance does not hold.

In general this condition can be linked to the Kirchhoff law of fluxes into a network, because the CME can be interpreted as a probability flux on a network generated by the available states and linked by the possible reactions. If we have a circular network of reversible transitions from state A to state B,C and D (so we are working on a system with only 4 possible states) like the following graph:



The condition of detailed balance between fluxes can be rewritten in terms of the reaction propensity alone, recalling that $P_i = P_j \frac{\lambda_{ji}}{\lambda_{ij}}$, and applying it recursively over the cycle, obtaining the following condition:

$$K_{AB}K_{BC}K_{CD}K_{DA} = K_{DC}K_{CB}K_{BA}K_{AD}$$

In a system that has more than one elementary cycle the detailed balance condition should apply to every cycle to be valid for the whole system.

1.3 Resolution methods

The master equation is impossible to solve in the general case, being a huge (or infinite) set of degenerate differential equations. In some special cases it is possible, however, to solve it with analytical and numerical methods. The following is a list of the main methods of analytical solution or numerical integration of the master equation to obtain the corresponding stationary distribution.

In order to choose the appropriate resolution method for a CME, it is useful to distinguish between a set of features that renders one or more solution methods unfeasible.

linearity of the processes: in this case every reaction term is constant or proportional to the first power of only one chemical specie. This make the system easy to solve with almost any method, especially the analytical one, for which it is possible to even try to write the time-dependent solution, see Sec 1.3.4 and Sec 1.3.5

dimensionality of the process: when the system can be seen as a monodimensional system the solution is usually almost trivial with the linear expansion method (see Sec 1.3.1). Note that a system of one specie, but with reactions that allow transitions of several length (like having both $\mathbb{E}_n^+$ and $\mathbb{E}_n^{+2}$) it is not truly linear and thus cannot be solved under the hypothesis of detailed balance. A method to solve bidimensional processes will be explained in Cap 4

stiffness of the processes: Stiff systems are very hard to simulate with the normal SSA method, see Sec 1.3.8, while it is possible to try an approximate numerical method like the τ-leap or the Chemical Langevin Equation, it is generally convenient to take advantage of that stiffness and try to separate the stiff process from the others, determining its stationary distribution independently and plugging that into the resolution method. If the slow process is a linear one it is often possible to substitute the stiff distribution with its mean.

separability of the processes: When in presence of chain processes without feedback it is often possible to solve independently each step of the chain and using that solution for the following processes. For this kind of solution, the most convenient method is the probability generating function explained in Sec 1.3.4

extension of the state space: When the complete state space is of limited size, like when the CME describes the possible states of a set of enzymes, one really useful method is the state enumeration (Sec 1.3.2) followed by an analysis of the eigenvalues of the transition matrix (Sec 1.3.3)

presence of absorbing states: An absorbing state is one where the probability can flow in, but cannot flow out, and thus will end absorbing all the probability, being the only possible state in the system (unless there are more than one absorbing state). In this case the final distribution is rather trivial: if there is only one absorbing state the final distribution will be a δ distribution, while with more absorbing states it will be a sum

of one δ for each state, whose absolute value depend from the starting distribution. In this case can be quite easily calculated the time of extinction by determining the distribution of the meta-stable state as explained in Sec 1.3.9

1.3.1 Linear expansion

Given a CME that can be written as a one-step birth-death process, i.e. has the following form, $\partial_t P_n(t) = -(g_n + r_n)P_n(t) + r_{n+1}P_{n+1}(t) + g_{n-1}P_{n-1}(t)$, we have a process for which the stationary distribution can be easily obtained in the general case with a linear expansion of the coefficient g_n and r_n.

This can be done noting that in this linear case in the stationary condition the flux should be 0 for each step, $P_n g_n = P_{n+1} r_{n+1}$, which corresponds to the condition of detailed balance. This allows to write a simple recursive solution for this system:

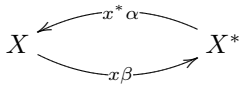
$$P_{n+1} = P_n \frac{g_n}{r_{n+1}}$$

and so, given a certain value of P_0 to maintain the final normalization of probability, to write the solution as:

$$P_n = P_0 \prod_{i=0}^{n-1} \frac{g_i}{r_{i+1}}$$

In some cases this solution can be written in a closed form, but often one can obtain just the series. This method can be used to find the closed form of two important processes, the birth-death and the interconversion.

The interconversion model represents a system whose components can switch between two separate states in a way that is independent from the state of all the others and with a constant probability in time. This is a typical model for a protein which can be in two state, like phosphorylated and non-phosphorylated, and switch back and forth between the two, ignoring the detail on how this transition happens (this hypothesis is sometimes justified by a first-order approximation of an enzymatic reaction, but need to be examined for each case). In this model the transition propensities from the state A and B are in the linear form Ak_a and Bk_b, with the following kinetic scheme:



Given that the probability of $P_{A=0} = C_0$ where C_0 is the normalization constant and N is the total number of present molecules $N = A + B$, the linear expansion

can be written as:

$$\begin{aligned}
& & p_0 &= C_0 \\
p_0 k_b(N-0) &= & p_1 1k_a \rightarrow p_1 &= C_0 \frac{NK_b}{k_a} \\
p_1 k_b(N-1) &= & p_2 2k_a \rightarrow p_2 &= C_0 \frac{NK_b}{k_a}\frac{(N-1)K_b}{2k_a} \\
p_2 k_b(N-2) &= & p_3 3k_a \rightarrow p_3 &= C_0 \frac{NK_b}{k_a}\frac{(N-1)K_b}{2k_a}\frac{(N-2)K_b}{3k_a} \\
& \dots & & \\
p_{i-1} k_b(N-i) &= & p_N ik_a \rightarrow p_i &= C_0 \frac{N!}{i!(N-i)!}\frac{k_b^i}{k_a^i}
\end{aligned}$$

This expression can be recognized as a binomial expression noticing that $\frac{k_a}{k_a+k_b}$ can be interpreted as a probability of the transformation from A to B instead of the opposite happening. With this in mind, we can write the solution as:

$$p_i = C_1 \binom{N}{i} \left(\frac{k_b}{k_a+k_b}\right)^i \left(\frac{k_a}{k_a+k_b}\right)^{n-i}$$

where $\frac{k_a}{k_a+k_b}$ e $\frac{k_b}{k_a+k_b}$ are the p and q of the binomial process and $C_1 = C_0(k_a + k_b)$.

In a similar fashion, we can determine the solution of the birth-death model. This model represents a system where we have a constant income flux of the substance A, which decades with a constant probability. The decay alone would generate an exponential fall of the quantity of A toward 0, that will be balanced by the constant influx. This can arise even from an interconversion system where one the two forms quantity is much greater than the other and can be approximated as constant (a zeroth-order approximation). The kinetic scheme is the following, where the reaction propensity for the birth process is a constant k_c and for the death is a linear one, Ak_d.

$$\varnothing \xrightarrow{\alpha} X \xrightarrow{x\beta} \varnothing$$

As before we write explicitly the solution of the recurrence equation as

$$\begin{aligned}
& & p_0 &= C_0 \\
p_0 k_c &= & p_1 1k_d \rightarrow p_1 &= C_0 \frac{k_c}{k_d} \\
p_1 k_c &= & p_2 2k_d \rightarrow p_2 &= C_0 \frac{k_c}{k_d}\frac{k_c}{2k_d} \\
p_2 k_c &= & p_3 3k_d \rightarrow p_3 &= C_0 \frac{k_c}{k_d}\frac{k_c}{2k_d}\frac{k_c}{3k_d} \\
& \dots & & \\
p_{i-1} k_c &= & p_N ik_d \rightarrow p_i &= C_0 \frac{k_c^i}{k_d^i}\frac{1}{i!}
\end{aligned}$$

where the normalization constant C_0 is equal to $e^{-\frac{k_c}{k_d}}$ (it is sufficient to note that the expression is the power series of an exponential, so the summation

over all P_n is one exponential).This leads to the canonical form of the Poisson distribution:

$$p_i = \frac{e^{-\lambda}\lambda^i}{i!} \text{ with } \lambda = \frac{k_c^i}{k_d^i}$$

This process is valid as long as there are no irregular point in the sequence of the reaction propensities like an attractor point, i.e. a point to which the probability can flow but not flow back: if n is our interest point, we have $g_{n-1} > 0$ and $r_n = 0$. A natural limit, as the one found in the binomial example, is not a problem, because we stop the series when we observe $g_n = 0$.

1.3.2 State enumeration

For small, multidimensional system, one of the best resolution strategies is to explore analytically every possible state determining its links to any other state. This allows us to write the whole transition matrix Λ, which can be used to evaluate any property of the system or to perform a greatly enhanced simulation. If carefully designed, the same algorithm allows us to write the Λ matrix in an analytical fashion, retaining all the dependencies to each parameter, making possible to feed the matrix to a symbolic algebra system like the "Mathematica"®suite to solve it in the general case.

The basic algorithm has been delineated in [15]. Given a starting state x_0, it pushes it in a stack and evaluates for each reaction if it may happen and to which state it would lead. All the reaction propensities will be saved in the Λ matrix (in a sparse format) and the new states will be pushed in the stack, testing that they are not already in it. The algorithm will then moves to the following state in the stack, until all the unexplored states has been visited and added to the stack.

In the end we obtain a list of all accessible states of the system and the transition probability from one to the other in the form of a transition matrix. The algorithm will converge to a finite number of states if the reactions that generate new molecules (and thus make new states accessible) become zero over a certain threshold. If this is not true the system is not limited and so has infinite accessible states. One can always set some threshold on the transition rates, removing those that are smaller than a set value, but this method can be used only when we are certain that it will not remove interesting zones, like secondary peaks of the stationary distribution.

Given this information, we can solve the system in several ways, the most straightforward being the direct integration of the whole CME as a set of linear ODE. This method works fine for limited set of states, but has a very slow convergence due to the determinant of the Λ matrix, which is 0 by design.

1.3.3 Eigenvector solution

Being the CME a linear operator on the probability distribution P_n, represented by the transition matrix Λ, we can solve the master equation like a normal dynamical system. The standard solution would suggest that we can write an evolution law as:

$$P(t + \Delta t) = P(t)e^{\Lambda \Delta t}$$

. This is formally correct, but practically unfeasible due to the exponentiation of the matrix Λ, which is not a solvable problem in the general case and it is not numerically stable when the size of Λ grows, and it often starts from an already infinite size.

The stationary condition $\Lambda\vec{P} = 0$ says that the stationary distribution is the kernel of the matrix Λ, which is always with determinant zero due to the construction of the matrix itself (this is necessary to impose the conservation of the probability). As long as the system is not too big in terms of the number of molecules, we can numerically approximate the value of that eigenvector, truncating the matrix if necessary without too many problems (as long as we set $g_n = 0$ if we truncate at size n to maintain the conservation of probability). This solution has two main problems: it is not analytical and is specific to the exact parameters used in the solution, not giving any insight on the structure of the general solution, and has a very poor convergence as soon as the system has more than few hundreds total states.

For the others eigenvectors very little can be said from a general point of view. The kernel eigenvector is the only one with nonnegative entries, while all the others eigenvectors correspond to negative eigenvalues and their sum over all the probability distribution is zero. In some cases, it is possible to have a hunch on the values of the eigenvalues, which gives information on the convergence time of the CME.

For example for the birth-death process can be shown that the eigenvalues spectrum follow a simple rule:

$$\lambda_i = -i(k_a + k_b) \quad \forall i \in \{0, \dots, N\}$$

. The corresponding eigenvectors for a system with $N = 24$ and $k_a = 1.0$,$k_b = 3.1$ is shown in Fig 1.1

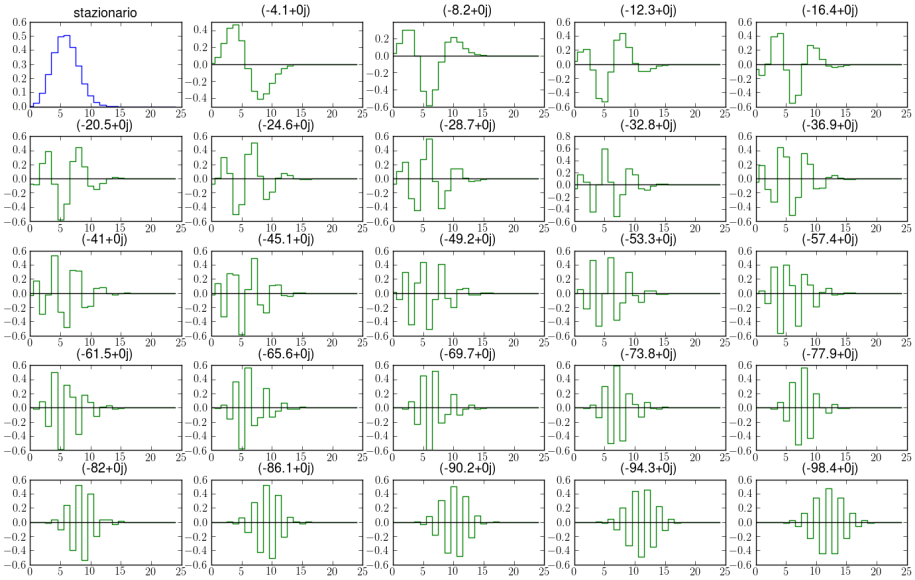


Figure 1.1: eigenvalues and eigenvectors of a birth-death process

1.3.4 Probability generating function

For discrete probability function we have a conjugated function called probability generating function that can be obtained applying a z-transform to the distribution, which is the discrete version of the Laplace transform. Being $p_i = P(X = i)$ the probability distribution associated with the stochastic process X, we define the associated probability generating function (PGF for brevity) G as:

$$G(s) = \sum_{n=0}^{\infty} s^n p_n = E[z^X]$$

so that the sum of n identical, non correlated processes, is simply $G_n X = G_X^n$

The function G is unique for the distribution of X, so that if two unknown distributions X and Y has the same PGF $G_X = G_Y$ we can say that $X = Y$.

If X is the sum of several stochastic processes, the resulting G will be the product of the G of the components. This is a consequence of the convolution theorem of the Laplace transform, being the resulting probability distribution the convolution of the components distributions.

$$X = \sum X_n \Rightarrow G_X = \prod G_{X_n}$$

G can be seen as an alternative definition for the probability of a specific process. In fact, starting from G we can obtain both the distribution of the associated process and his moments.

The distribution terms p_i can be obtained taking the i derivative of G calculated in $s = 0$:

$$p_i(t) = \frac{1}{i!} \frac{d^i G(s,t)}{ds^i}\bigg|_{s=0}$$

which can be seen as the i-th term of the power serie of G. The same term can be equivalently calculated as:

$$p_i(t) = \frac{1}{2\pi i} \oint \frac{G(s,t)}{s^{i+1}} ds$$

If the n-derivative of G is taken with $s = 1$ we obtain the n-th factorial moment of the distribution, $E[X(X-1)(X-2)\dots(X-n)]$. These are a linear combination of the normal moments of the distribution, which can be described as a sum of derivatives of G:

$$\begin{aligned} \mu_0(X) &= 1 = G(1) \\ \mu_1(X) &= E[X] = G'(1) \\ \mu_2(X) &= E[X^2] = G''(1) - G'(1) \\ Var[X] = E[X^2] - E[X]^2 &= G''(1) - G'(1)(1 - G'(1)) \end{aligned}$$

This is true even when G is a multivariate function of two random variable X_1 and X_2. In this case we will have

$$G(s_1, s_2) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} s_1^n s_2^m p(n,m)$$

and the value for the probability p(i,j) will be:

$$p(i,j) = \frac{1}{i!}\frac{1}{j!}\frac{\partial^i}{\partial s_1^i}\frac{\partial^j}{\partial s_2^j}G(s_1,s_2)\bigg|_{s_1=0,s_2=0}$$

and if there is no correlation between X_1 and X_2 the global PGF factorizes as $G(s_1,s_2) = G(s_1)G(s_2)$ and the joint pdf results, as expected, the product of the two distributions on n and m.

If the probability p_n can be written as $np_n^\dagger$ then $G(s) = s\frac{\partial}{\partial s}G^\dagger(s)$. This mean that when we have recurrence relationship like the Chemical Master Equation we can write a differential equation in both time and s which solution is the PGF of the stationary solution of the CME. In fact, if we manage to write the whole time-dependent solution, we can write in closed form the whole evolution of the system. Even when the whole time dependent solution is not available, it is usually possible to write the evolution of the mean and variance of the system.

The above relationship is true whenever we have a master equation driven by polynomial transition rates: the resulting equation is a linear partial differential equation in the form:

$$\frac{\partial G(s)}{\partial t} = \hat{\mathcal{L}}G(s)$$

where $\hat{\mathcal{L}}$ is a generic linear operator, whose order (the highest power of $\frac{\partial}{\partial s}$ present) is equal to the order of the highest reaction rate. Thus, a reaction which behave like n^3 will correspond to a linear operator of order 3, i.e. with a derivative of order $\frac{\partial^3}{\partial s^3}$

We can generally solve this equation in his complete time dependent form only when the maximum order is one, which mean only one body interaction and derivatives of order $\frac{\partial}{\partial s}$. Considering a toy model of a decaying ensemble of n_0 particles with a decaying constant of λ, as pictured by the following scheme.

$$A \underset{k^-}{\xrightarrow{k^+}} \varnothing$$

This system is described by the following master equation:

$$\frac{\partial P_n(t)}{\partial t} = \lambda(\mathbb{E}_n^{+1} - 1)nP_n(t)$$

that has the following PGF partial differential equation:

$$\frac{\partial G(s)}{\partial t} = \lambda(1-s)\frac{\partial G(s)}{\partial s}$$

which leads to the following time-dependent solution using the methods of the characteristics and the initial condition of $G(s,t=0) = s^{n_0}$ which derives from the initial condition $P(n,t=0) = \delta_{n,n_0}$:

$$G(s,t) = [1 + (s-1)e^{(-\lambda t)}]^{n_0}$$

We can immediately recognize from the elevation to the n_0 that we are dealing with n_0 independent processes that evolve independently in an exponential

fashion, as we may expect in a process of independent decay. Transforming back this expression requires some algebra manipulation to obtain the complete time dependent solution:

$$P(n,t) = \frac{n_0!}{n!(n_0-n)!}(e^{-\lambda t})^n(1-e^{-\lambda t})^{n_0-n}$$

The PGF can be used even to determine the i-th moment:

$$< n^i(t) >= \Big[(x\partial_x)^i G(s,t)\Big]_{s=1}$$

This can, in theory, be used to write the equation of each moment in time, but there is a more convenient method that will be discussed in the next subsection.

1.3.5 Moments equation

It's not always possible to solve the PDE associated with a specific PGF, but in several cases, like in a multivariate CME or the ones with non-linear coefficient, it is not feasible to write the PGF at all. This limitation can be avoided if one renounce to determine the whole time dependent distribution ad settle down to just the first and second moment of the distribution. In this case is possible to write a differential equation of these moments that give the evolution in time without solving the whole evolution.

$$< \dot{n}^i >= \sum_{n=0}^{\infty} n^i \dot{P}_n$$

Being able to solve this equation for all i has the same information about the distribution as the solution of the time-dependent CME but is similarly unfeasible. In general, it is fairly easy to write down the equation for mean and variance of the distributions, which are usually independent from the moments of higher order for linear models. considering again the linear decay process we can calculate his moments in time quite easily:

$$\begin{aligned}
< \dot{n} > &= -n\sum_{n=0}^{\infty} nP_n + n\sum_{n=0}^{\infty}(n+1)P_{n+1} \\
&= -\sum_{n=0}^{\infty} n^2P_n + \sum_{n=0}^{\infty}(n+1-1)(n+1)P_{n+1} \\
&= -\sum_{n=0}^{\infty} n^2P_n + \sum_{n=0}^{\infty}(n+1)^2P_{n+1} - \sum_{n=0}^{\infty}(n+1)P_{n+1} \\
&= -\sum_{n=1}^{\infty} n^2P_n + \sum_{n=1}^{\infty}(n)^2P_n - \sum_{n=1}^{\infty}(n)P_n \\
&= - < n >
\end{aligned}$$

where we used the fact that nP_n is 0 for $n = 0$ and so $\sum_{n=1}^{\infty} n^iP_n = \sum_{n=0}^{\infty} n^iP_n$. As we expected from the equivalent deterministic model, the mean

decrease exponentially with time. More interesting is what happen to the second moment (which is related to the variance). With a procedure similar to the previous one we obtain that:

$$< \dot{n^2} > =< n > -2 < n^2 >$$

1.3.6 Fokker Plank approximation

If we look to a generic birth-death process driven by the birth rate g_n and the death rate r_n we can write the master equation as:

$$\frac{\partial P_n(t)}{\partial t} = -(g_n + r_n)P_n(t) + r_{n+1}P_{n+1}(t) + g_{n-1}P_{n-1}(t)$$

If we consider $n \gg 1$ we can approximate this formula treating n as a continuous variable, so that the finite difference can be treated as differentiations.

$$f(n \pm 1) = f(n) \pm \partial_n f(n) + \frac{1}{2}\partial_n^2 f(n)$$

Inserting this result in the previous equation, we get a differential version of that master equation that is equivalent to a Fokker-Planck equation:

$$\partial_t P(n,t) = \partial_n[(g(n) - r(n))P(n,t)] + \frac{1}{2}\partial_n^2[(g(n) + r(n))P(n,t)]$$

where the term $(g(n)-r(n))$ correspond to a drift term, which tends to translate the distribution, and the term $(g(n)+r(n))$ corresponds to a diffusion due purely to the noise presents in the system. This can be rewritten to explicate this terms as:

$$\partial_t P(n,t) = \partial_n[F(n)P(n,t)] + \frac{1}{2}\partial_n^2[D(n)P(n,t)]$$

this form can be recognized as a equivalent to one that writes explicitly the flux of probability between states:

$$\partial_t P(n,t) = \partial_n[[F(n)P(n,t)] + \frac{1}{2}\partial_n[D(n)P(n,t)]]$$

The steady state solution can be found integrating the system to the general solution imposing the flux to be zero. This is always true in the monodimensional case but not in the general sense.

$$P(n) = \frac{1}{ZD(n)} \exp\left[2\int_0^n \frac{F(n')}{D(n')}dn'\right]$$

where Z is a normalization constant. This can be immediately recognized as equivalent to the linear expansion seen in section 1.3.1.

1.3.7 Deterministic equivalent model

Given a CME, it is always possible to write an equivalent ODE for each one of the chemical species described. This is especially true in the case of birth-death processes, where we can simply recognize the g_n as a positive term in the ODE and the r_n as a negative one, leading to the following ODE:

$$\dot{n} = g(n) - r(n)$$

.

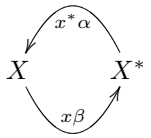
It's very important to notice that the converse, writing the CME given the ODE, although is a common process, is not correct. This can be seen looking to a differential equation in the form $\dot{x} = \alpha - \beta x$.

This differential equation, one of the most simple ones, is afflicted by a severe indetermination when one tries to write a master equation starting from it. There are two main different models which can be approximated to that same deterministic equation: the creation-decadence or the interconversion.

The creation-decadence model hypothesize that the substance X is created with constant probability and decay exponentially, as schematized in the underlying graph:

$$\varnothing \xrightarrow{\alpha} X \xrightarrow{x\beta} \varnothing$$

The interconversion model hypothesizes the existence of a second possible form of X, which we can call X^*, and our molecule oscillates between these two states as shown in the graph below:



While they have the same macroscopic approximation, the underlying distributions are different: in the creation-decadence model we have a Poisson distribution, an unlimited distribution, while in the other case we observe a binomial distribution, a limited one, with generally different moments for every order after the first one, the mean (which is described by the ODE).

1.3.8 Stochastic simulation algorithm

For systems that span more than few molecular species with few hundreds molecules, all the systems listed above fail, one way or the other, due to our limitation in finding analytical solutions in multidimensional system or the limit of the numerical computation, whose rounding errors pile up making any prediction close to meaningless.

In this situation, the only feasible way to analyze a system is through Montecarlo simulations of the system itself. The main system to perform this simulation is the Stochastic Simulation Algorithm[26], which simulates each reaction step in a painstaking way. The solution obtained with this method has been proved to converge to those of the corresponding master equation. Starting from the original formulation, which is a common workhorse in system biology, several others has been proposed by Gillespie himself to overcome the main limitation of the original algorithm, which is a non bounded time of simulation for stiff systems.

The basic Stochastic Simulation Algorithm is strikingly simple: given a state of the system one has to choose which reaction will happen next between the possible ones and how much time will the system stay still before the reaction happens. This process is iterated until the whole time of interest has been simulated.

Without any loss of generality, we can write the algorithm for the monodimensional case. In the scope of this description we will work with a system that is described by the state vector $\vec{x}$, which in general represent the amount of molecules for each chemical specie. Each reaction i will be described by its propensity rate λ_i and by the modification of the state vector that generate: in a birth-death system the variation will be simple ± 1, but in the general case will be a step in the n-dimensional space in which the process happen, which will be represented as $\vec{\nu}_i$.

Given the Markovian hypothesis, each reaction, ignoring the effect of all the others, will have an exponentially distributed wait time between two successive event, whose exponent rate will be described by the propensity λ_i. So the probability of waiting a time Δt between two events will behave like:

$$p_i(\Delta t) = e^{-\lambda_i \Delta t}$$

the waiting time for any reaction to happen, given that any reaction is independent from the others, will be:

$$p(\Delta t) = e^{-\sum_{i=0}^{R} \lambda_i \Delta t}$$

where i is used as index for the R possible reactions. This waiting time can be easily generated with a transformation of a random number between 0 and 1 with the following transform, being

$$\lambda = \sum_{i=0}^{R} \lambda_i$$

:

$$\Delta t = -ln(rand(0,1))\lambda^{-1}$$

After determining the waiting time, we need to know which reaction will happen next. This can be done selecting among the probabilities proportionally to the propensity rate λ_i relative to the total reaction propensity λ. The most convenient method to determine this reaction is described in [61]. The methods they propose is to sum all the propensity λ_i and normalize them by the whole sum λ, obtaining a normed vector with ordered values between 0 and 1. This allows us to perform a binary search of which reaction happens given a random number between 0 and 1 (the basic random number from the Random Number Generator of most languages). This, combined with the bisection method discussed in the appendix, allows the selection step to be done in a logarithmic time in the number of reactions present.

In Fig 1.2 we can see the result of a single realization of a SSA run. The system jumps from a state to the other in a random fashion. The system depicted is a simple creation-destruction process with a mean of 50 molecules. To estimate the mean in a specific moment of time we have to resort to a set of simulation, the more the better. This permits to evaluate the distribution at each moment of time, but it has two severe drawbacks: the number of realization to estimate correctly the distribution can be very high and will be specific for the set of parameters used without any possibility of generalization in the parameters space. This makes the SSA a great tool to perform preliminary analysis on systems with unknown properties or to study the behavior of complex model given a solid theoretical analysis of its expected properties.

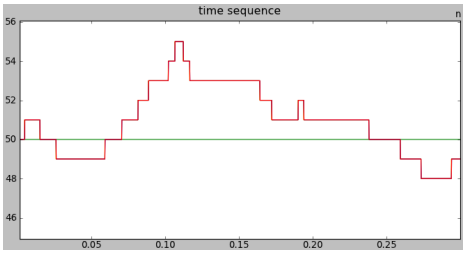


Figure 1.2: short single realization of a SSA run

In Fig 1.3 we analyze the same system of Fig 1.2, confronting the result of a singular realization (in blue) versus the evolution of the mean and standard deviation (red line and orange shade, respectively). In green is represented the expected behavior of the system based on the ODE of the underlying process, which is a correct approximation of the effective mean of the system as long as the system is linear. It is clear that the single realization moves approximately around the mean, but the variation can be substantial.

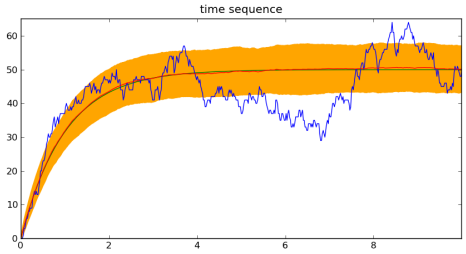


Figure 1.3: single realization of a SSA run versus the expected mean and variance

To evaluate the stationary distribution is crucial to eliminate the first transient time (approximately before $t = 6$ in Fig 1.3) and then proceed to evaluate the probability of being in a certain state as the time spent in each state over the total time of observation. It is necessary to pay attention to the starting time of observation: whatever is the starting position, it is necessary to eliminate from the final distribution all the observed state previous to a at least 5 or 6 time the inverse of the smallest eigenvalue, to give time to the system to forget the his starting point. While in principle it could be possible to use the "Coupling from the Past"[85] method from the MCMC theory, it would be useful only in the monodimensional case, where other, better performing methods can be used to obtain the general solution.

Given the slow convergence of this method, two approximation techniques have been developed by Gillespie itself: the τ-leap[41] and the Chemical Langevin [39] Equation.

In the τ-leap one try to evaluate how many reactions take place in a specific time interval τ given the propensity of each reaction at the time t, so that one can approximately use a Poisson distribution to evaluate how many time

each reaction fires, and update the system correspondingly. With this idea the evolution of the system at the time $t + \tau$ is:

$$\vec{X}(t+\tau) - \vec{X}(t) = \sum_{i=1}^{R} \nu_i P_i(\lambda_i(\vec{X}), \tau)$$

Where ν_i represents the variation in $\vec{X}$ due to the ith-reaction, $\lambda_i(\vec{X})$ is the expected mean of reaction i happening and P_i represents the Poisson distribution of actual event for the ith-reaction. The task of choosing the right τ it is not trivial[14], due to the possibility that the chosen number of reaction events make a specie negative if not carefully checked. In the most delicate point it is often mixed with a standard SSA.

In the Chemical Langevin Equation, one works with a deterministic ODE for the mean, to which is added a Gaussian noise as in the standard Langevin methods, but designed to respect the correlation of the variation due to the various reactions, extracting one normal variable for each reaction instead of one for each specie as the basic Langevin Equation. The corresponding Langevin equation takes the form of a standard Îto integral in the form:

$$d\vec{X}(t) = \sum_{i=1}^{R} \nu_i \lambda_i(\vec{X}(t))dt + \sum_{i=1}^{R} \nu_i \sqrt{\lambda_i(\vec{X}(t))} dW_i(t)$$

Where the $W_i(t)$ are independent scalar Brownian motions. As long as the underlying CME is composed of linear reaction the Chemical Langevin Equation is guaranteed to have the same first two moments[49] and, by definition, the same correlation structure.

1.3.9 Absorbing states

An absorbing state is defined by the property that both g_n and r_n are equal to zero while g_{n-1} or r_{n+1} is different from zero, meaning that there is a net probability influx of probability to the state n, subtracting it to the rest of the system. If it is the only one absorbing point, it will drain all the probability in it, making the stationary distribution simply a δ_n function. This makes any attempt to study the system fruitless. This can be solved in a brutal way simply adding a small term to generate an efflux of probability from the state, but it is hard to guarantee that this will not destroy other features of the system.

This is often true when one is trying to modelize ecological systems, where the point $n = 0$ is obviously an absorbing state as long that there is no immigration in the system. The problem is that the extinction solution is trivial and true only in the limit of infinite time with a very slow convergence time. In fact, in most system the actual probability of extinction is zero for all practical purpose.

A solution for this problem is to explicitly eliminate the absorbing state. For the sake of simplicity we will consider an ecological system with an absorbing state in $n = 0$. In this case we simply impose $r_1 = 0$ and, in fact, the whole state $n = 0$ from the equation. It can be demonstrated that this condition leads to a distribution that is equivalent to the metastable eigenvector of the complete system: in this case this eigenvector is a proper distribution of positive terms because the only negative term appears to be the $P_{n=0}$. This gives a simple

estimation of the characteristic time of extinction, that is proportional to the efflux of probability from $P_{n=1}$ to $P_{n=0}$, and so it is simply the inverse of r_1 before it was ruled out.

Brief note about the computational methods on this thesis

All the computations in this thesis have been done with the scientific framework for python[107]: numpy[77] (basic numerical routines on multidimensional arrays), scipy[30] (the scientific framework for python), matplotlib[53] (plotting library for scipy), sympy[102] (symbolic mathematics), ipython[83] (advanced scipy-ready shell for python) and cython[8] (python compilation as C code).

I would like to give credit to the whole python scientific community for the excellent tools provided, which allowed me to work for years on these topics only with full open-source instruments.

Chapter 2

Analysis of a genetic switch for E2F and Myc

In the following work we examine a genetic toggle switch that underlies the regulation of the cellular cycle. Starting from a previous work that describes a simplified deterministic approximation of that genetic circuit, we develop its stochastic version, studying where the two models were significantly different and studying which information could the stochastic approach adds to the deterministic one. The original system, which involves both gene expression, protein production and miRNA regulation, describes them with a generalized term for the protein and the miRNA levels. It can be bistable under certain conditions, and this was verified as true even for the stochastic equivalent. The main difference was found on the edge between the monomodal and bimodal condition, where secondary peaks could appear without any deterministic equivalent. We then proceeded to write an even more simplified version of the model and empirically shown their equivalence as long as the miRNA half-life is lesser than the protein's one.

REFERENCES:

Title: Stochastic analysis of a miRNA-protein toggle switch

Authors: E. Giampieri, D. Remondini, L. de Oliveira, G. Castellani, P. Lió

Journal: Molecular BioSystems - 2011 (published)

2.1 Motivation of the work

Complex cellular responses are often modeled as switching between phenotype states, and despite the large body of deterministic studies and the increasing work aimed to elucidate the effect of intrinsic and extrinsic noise in such systems, some aspects still remain unclear. Molecular noise, which arises from the randomness of the discrete events in the cell (for example DNA mutations and repair) and experimental studies have reported the presence of stochastic mechanisms in cellular processes such as gene expression [69],[68], [104], decisions of the cell fate [1], and circadian oscillations [5]. Particularly, low copy numbers of important cellular components and molecules give rise to stochasticity in gene expression and protein synthesis, and it is a fundamental aspect to be taken into account for studying such biochemical models [67, 37]. In this paper, we consider a simplified circuit that is known to govern a fundamental step during the eukaryotic cell cycle that defines cell fate, previously studied by means of a deterministic modeling approach [7]. Let set the scene by reminding that "all models are wrong, but some are useful" (said by George Edward Pelham Box, who was the son-in-law of Ronald Fisher). Biologists make use of qualitative models through graphs; quantitative modeling in biochemistry has been mainly based on the Law of Mass Action which has been used to frame the entire kinetic modeling of biochemical reactions for individual enzymes and for enzymatic reaction network systems [43]. The state of the system at any particular instant is therefore regarded as a vector (or list) of amounts or concentrations and the changes in amount or concentration are assumed to occur by a continuous and deterministic process that is computed using the ordinary differential equation (ODE) approach. However, the theory based on the Law of Mass Action does not consider the effect of fluctuations. If the concentration of the molecules is not large enough, we cannot ignore fluctuations. Moreover, biological systems also show heterogeneity which occurs as a phenotypic consequence for a cell population given stochastic single-cell dynamics (when the population is not isogenic and in the same conditions). From a practical point of view, for concentrations greater than about 10 nM, we are safe using ODEs; considering a cell with a volume of 10^{-13} liters this corresponds to thousands of molecules that, under poissonian hypothesis, has an uncertainty in the order of 1%. If the total number of molecules of any particular substance, say, a transcription factor, is less than 1,000, then a stochastic differential equation or a Monte Carlo model would be more appropriate. Similarly to the deterministic case, only simple systems are analytically tractable in the stochastic approach, i.e. the full probability distribution for the state of the biological system over time can be calculated explicitly, becoming computationally infeasible for systems with distinct processes operating on different timescales. An active area of research is represented by development of approximate stochastic simulation algorithms. As commented recently by Wilkinson, the difference between an âapproximateâ and âexactâ model is usually remarkably less than the difference between the âexactâ model and the real biological process [24]. Given we can see this either as an unsatisfactorily state of art or as a promising advancement, we can summarise the methodological approaches as following. Biochemical networks have been modeled using differential equations when considering continuous variables changing deterministically with time. Single stochastic trajectories have been modeled using stochastic differential equations (SDE) for continuous

random variables, and using the Gillespie algorithm for discrete random variables changing with time. Another choice consists in characterizing the time evolution of the whole probability distribution. The corresponding equation for the SDE is the Fokker-Planck equation, and the corresponding equation for the Gillespie algorithm is called the Chemical Master Equation (CME) [54]. Therefore, the CME could be thought as the mesoscopic version of the Law of Mass Action, i.e. it extends the Law of Mass Action to the mesoscopic chemistry and biochemistry, see for example [87, 112].

Here we compare the results of a stochastic versus deterministic analysis of a microRNA-protein toggle switch [34, 44] involved in tumorigenesis with the aim of identifying the most meaningful amount of information to discriminate cancer and healthy states. We show that the stochastic counterpart of such deterministic model has many commonalities with the deterministic one, but some differences arise, in particular regarding the number of stable states that can be explored by the system. The disagreement between the stochastic and deterministic description is observed in a "ghost" effect caused by the proximity to a deterministic bifurcation [99], and in a somehow opposite situation, in which the variance of the stable point can mask the detection of the second peak in the stationary distribution. In this paper we perform a numerical study of the complete two-dimensional model, but we consider also a simplified, biologically meaningful, version of the model that allows to calculate an exact solution, with a numerical characterization of the parameter ranges in which the two systems produce qualitatively similar results. A discussion of the possible implications of our results in a real system are described in the last Section.

2.2 The E2F-MYC toggle switch

Oncogenes and tumor-suppressor genes are two pivotal factors in tumorigenesis. Recent evidences indicate that MicroRNAs (miRNAs) can function as tumor suppressors and oncogenes, and these miRNAs are referred to as âoncomirsâ. MiRNAs are small, non-coding RNAs that modulate the expression of target mRNAs. The biogenesis pathway of miRNAs in animals was elucidated by [80]. MiRNAs undergo substantial processing since the nuclear transcription where two proteins play an essential role: Drosha and Dicer. Most of miRNA are first processed into pre- miRNA by Drosha. After exportated to the cytoplasm, the pre- miRNA is processed by Dicer into a small double strand RNA (dsRNA) called the miRNA: miRNA duplex. The active strand, which is the mature miRNA is incorporated into the RISC and binds to the target mRNA, whereas the inactive strand is ejected and degraded. In normal tissue, proper regulation of miRNAs maintains a normal rate of development, cell growth, proliferation, differentiation and apoptosis. Tumorigenesis can be observed when the target gene is an oncogene, and the loss of the miRNA, which functions as a tumor suppressor, might lead to a high expression level of the oncoprotein. When a miRNA functions as an oncogene, its constitutive amplification or overexpression could cause repression of its target gene, which has a role of tumor suppressor gene, thus, in this situation, cell is likely to enter tumorigenesis. MiRNAs are often part of toggle switches: important examples involve gene pairs built with oncogenes and tumour suppressor genes [65, 66]. Here we focus on the amplification of 13q31-q32, which is the locus of the the miR-17-92.

The miR-17-92 cluster forms a bistable switch with Myc and the E2F proteins [56, 75, 7]. The oncogene Myc regulates an estimated 10% to 15% of genes in the human genome, while the disregulated function of Myc is one of the most common abnormalities in human malignancy [45, 20]. The other component of the toggle switch is the E2F family of transcription factors, including E2F1, E2F2 and E2F3, all driving the mammalian cell cycle progression from G1 into S phase. High levels of E2Fs, E2F1 in particular, can induce apoptosis in response to DNA damage. The toggle switch also interacts with dozens of genes (figure 2.1 depicts a portion of the whole circuitry), particularly with Rb and other key cell-cycle players. A summary of the experiments perturbing miRNA/Myc/E2F and E2F/RB behaviours have suggested the following:

- The Rb/E2F toggle switch is OFF when RB inhibits E2F, i.e. stopping cell proliferation; it is ON when E2F prevails and induces proliferation. Once turned ON by sufficient stimulation, E2F can memorize and maintain this ON state independently of continuous serum stimulation.

- The proteins E2F and Myc facilitate the expression of each other and the E2F protein induces the expression of its own gene (positive feedback loop). They also induce the transcription of microRNA-17-92 which in turn inhibits both E2F and Myc (negative feedback loop) [115].

Moreover, the increasing levels of E2F or Myc drive the sequence of cellular states, namely, quiescence, cell proliferation (cancer) or cell death (apoptosis).

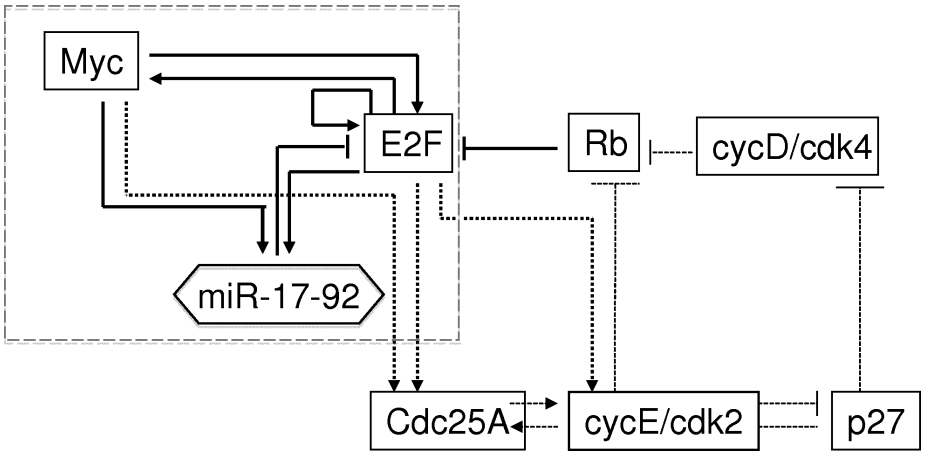


Figure 2.1: The E2F-MYC-miR-17-92 toggle switch with its biochemical environment (derived form [7]). Arrows represent activation, and bar-headed lines inhibition, respectively. The elements inside the dashed box represent the protein compound p (Myc-E2F) and the miRNA cluster m (miR-17-92), modelized in eq. 2.1 and 2.2.

Although there is increasing amount of research on cell cycle regulation, the mathematical description of even a minimal portion of the E2F, Myc and miR-17-92 toggle switch is far from trivial. Aguda and collaborators [7] have developed a deterministic model, which reduces the full biochemical network of

the toggle switch to a protein (representing the E2F-Myc compound) and the microRNA-17-92 cluster (seen as a single element).

It is a 2-dimensional open system, in which p represents the E2f-myc complex and m the miRNA cluster: thus no Mass Action Law holds, and the total p and m concentration is not conserved. The dynamics of p and m concentrations are described by eq. 2.1 and 2.2:

$$\dot{p} = \alpha + \frac{k_1 \cdot p^2}{\Gamma_1 + p^2 + \Gamma_2 \cdot m} - \delta \cdot p \tag{2.1}$$

$$\dot{m} = \beta + k_2 \cdot p - \gamma \cdot m \tag{2.2}$$

The model is the following: constitutive creation-destruction processes for p and m are driven by α,δ,β and γ parameters, while k_1 and k_2 control p and m state-dependent synthesis. The Γ_1 term is a kinetic (enzymatik-like) constant, while Γ_2 modulates miRNA inhibition of p synthesis. The nonlinearity of p in eq 2.1 is a Hill coefficient ($= 2$) representing self-promotion effect driven by a sigmoidal activation curve, a very common behaviour in gene regulation systems. A Hill coefficient > 1 can be justified by a cooperative effect of the terms involved in the compound represented by p: for example due to the E2F trimer, or a more complex aggregate with Myc. Although several experimental results suggest that in some cancer processes, a certain amount of interdependence and interaction among E2Fs exists, a detailed experimental investigation should be needed in order to estimate such parameter correctly [55, 81, 27].

All the results described in this article are very robust with respect to the choice of the specific Hill coefficient (here chosen equal to 2 for continuity with the original model in [7]) as long as it's larger than one (data not shown). We found a good qualitative agreement even if a different functional form was hypotized, as long as it retained its sigmoidal-like structure.

The system can be rewritten in an adimensional form as follows:

$$\epsilon\dot{\phi} = \alpha' + \frac{k \cdot \phi^2}{\Gamma'_1 + \phi^2 + \Gamma'_2 \cdot \mu} - \phi \tag{2.3}$$

$$\dot{\mu} = 1 + \phi - \mu \tag{2.4}$$

Where the parameters are: $\alpha' = \frac{k_2}{\delta \cdot \beta}\alpha$, $k = \frac{k_1 k_2}{\delta\beta}$, $\Gamma'_1 = \frac{k_2^2}{\beta^2}\Gamma_1$, $\Gamma'_2 = \frac{k_2^2}{\beta\gamma}\Gamma_2$, $\epsilon = \frac{\gamma}{\delta}$ and the change of variables is: $\phi = \frac{k_2}{\beta}p$, $\mu = \frac{\gamma}{\beta}m$ and $\tau = \gamma t$.

In the original model [7], the rate of protein synthesis is not a function of the instantaneous concentration (as assumed in eq.2.3) but rather of its concentration at some time Δ in the past:

$$\epsilon\dot{\phi} = \alpha' + \frac{k[\phi(\tau - \Delta)]^2}{\Gamma'_1 + [\phi(\tau - \Delta)]^2 + \Gamma'_2 \cdot \mu(\tau - \Delta)} - \phi(\tau). \tag{2.5}$$

We will not consider such delay in our stochastic realization of the model, since it would increase system dimensionality and it does not seem necessary to obtain the features we want to characterize.

The steady state can be studied in the nondimensionalized system and, therefore, the conditions on the parameters for the existence of multiple steady states. In the resulting cubic equation:

$$\alpha' + \frac{k\phi^2}{\Gamma'_1 + \phi^2 + \Gamma'_2 \cdot (1+\phi)} - \phi = 0 \tag{2.6}$$

the necessary (but not sufficient) conditions for the existence of 3 steady states (and thus a bistable system) are:

$$(\Gamma'_2 - k) < \alpha' < \left(1 + \frac{\Gamma'_1}{\Gamma'_2}\right) \tag{2.7}$$

We took advantage of the deterministic results in [7] in order to consider suitable parameter ranges for our stochastic modelling (as described in the following sections).

2.3 Description of the stochastic model

The system represented by equations 2.1 and 2.2 can be studied as a stochastic system through the Chemical Master Equation (CME) approach [106]. The resulting CME has two variables, the number of p and m molecules, labeled as n and m. The temporal evolution of the probability $p_{n,m}(t)$ to have n and m molecules at time t is described by the following equation:

$$\dot{p}_{n,m} = (\mathbb{E}_n - 1)r^n p_{nm} + (\mathbb{E}_n^{-1} - 1)g^n p_{nm} + (\mathbb{E}_m - 1)r^m p_{nm} + (\mathbb{E}_m^{-1} - 1)g^m p_{nm} \tag{2.8}$$

This CME is derived under the conditions of a one-step Poisson process: $\mathbb{E}$ and $\mathbb{E}^{-1}$ are the forward and backward step operators, g and r the generation and recombination terms for the n and m variables, as shown in superscripts.

The two generation and recombination terms associated with the n and m variables are respectively:

$$g^n = \alpha + \frac{k_1 \cdot n^2}{\Gamma_1 + n^2 + \Gamma_2 \cdot m}; \qquad r^n = \delta \cdot n \tag{2.9}$$

$$g^m = \beta + k_2 \cdot n; \qquad r^m = \gamma \cdot m \tag{2.10}$$

We remark that the molecule influxes into the system (represented by the α and β terms) could be included in different ways in the stochastic equations, since in the deterministic equations they represent a sort of "mean field" value. As an example, molecules could be added in bursts with specific time distributions, that do not appear in the macroscopic continuous deterministic equations. We will consider the simplest approach, but the choice of different influx patterns should deserve further investigation.

2.3.1 The one-dimensional model

We can reduce the problem from two to one dimension, by considering a different time scale for the two reactions (in particular considering m as a fast variable)

and thus considering the steady state solution for the m:

$$m = \frac{\beta + k_2 \cdot p}{\gamma} = \beta' + k' \cdot p \tag{2.11}$$

As a consequence we obtain a deterministic equation for the p only:

$$\dot{p} = \alpha + \frac{k_1 \cdot p^2}{\Gamma' + \Gamma'' \cdot p + p^2} - \delta \cdot p \tag{2.12}$$

with $\Gamma' = \frac{\Gamma_2 \cdot k_2}{\gamma}$ and $\Gamma'' = \Gamma_1 + \frac{\Gamma_2 \beta}{\gamma}$. The stochastic equation for p_n is thus as follows:

$$\dot{p_n} = (\mathbb{E} - 1) r_n \cdot p_n + (\mathbb{E}^{-1} - 1) g_n \cdot p_n \tag{2.13}$$

$$g_n = \alpha + \frac{k_1 \cdot n^2}{\Gamma' + \Gamma'' \cdot n + n^2}; \qquad r_n = \delta \cdot n \tag{2.14}$$

A general solution can be obtained

$$p_n^s = \prod_{i=1}^{N} \frac{g(i-1)}{r(i)} \cdot p_0 = \prod_{i=1}^{N} \frac{\alpha + \frac{k_1 \cdot (i-1)^2}{\Gamma' + \Gamma'' \cdot (i-1) + (i-1)^2}}{\delta \cdot i} \cdot p_0 \tag{2.15}$$

with an adequate normalization factor imposed on p_0:

$$p_0 = \frac{1}{1 + \sum_{i=1}^{N} \prod_{i=1}^{N} p_{n_1}^s} \tag{2.16}$$

We remark that the system is open, thus in theory N is not fixed, but we can truncate the product to a sufficiently high value of N obtaining a good approximation of the whole distribution. This one-dimensional system (for which an analytical solution can be obtained) will be compared to numerical simulations of the exact one-dimensional and two-dimensional systems.

2.4 Analysis of the stochastic model

2.4.1 The stationary distribution

The one-dimensional model can show monomodal as well as bimodal stationary distributions, depending on the parameters considered. As an example, we obtain bistability with a set of parameters as shown in Fig. 2.2.

Thus the qualitative features of the two-dimensional deterministic model (i.e. the possibility of being bistable depending on the parameter range) are recovered for the one-dimensional approximation of the stochastic system. Also the two-dimensional stochastic system shows bistability for the same parameters, and they are in optimal agreement for a range of parameters in which the $\dot{m} \gg \dot{p}$ condition holds

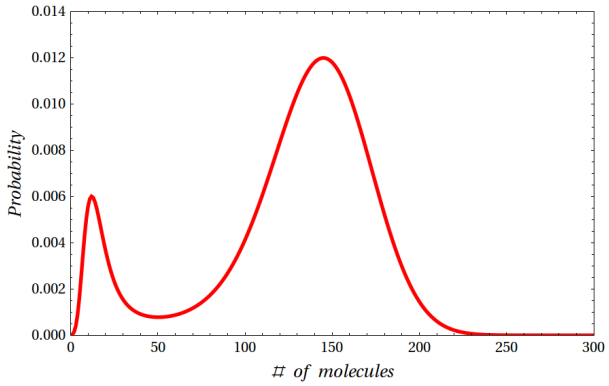


Figure 2.2: The stationary distribution for the one-dimensional space, obtained using the following parameters: $\alpha = 1.68(molecule/h)$, $\beta = 0.202(molecule/h)$, $\delta = 0.2(h^{-1})$, $\gamma = 0.2(h^{-1})$, $\Gamma_1 = 10300(molecule^2)$, $\Gamma_2 = 1006(molecule)$, $k_1 = 90(molecule/h)$ and $k_2 = 0.05(h^{-}1)$.

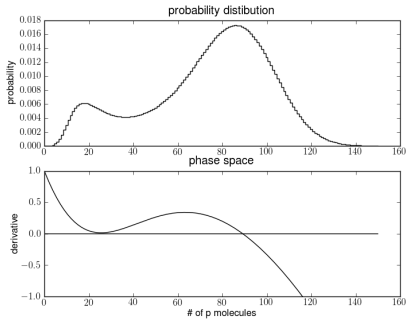


Figure 2.3: Comparison between the deterministic solution (bottom) and the stationary distribution (top) for the parameter set as in Table 2.1, case 3.

We also observe some remarkable differences between the deterministic and the stochastic models: there are regions in parameter space in which the deterministic approach shows only one stable state, but in the stochastic system two maxima in the stationary distribution are observed (see Fig. 2.3). This difference can be explained qualitatively as follows: for the deterministic system, there are parameter values for which the system is monostable but very close to the "transition point" in which the system becomes bistable. It is known that in these situations a "ghost" remains in the region where the stable point has disappeared [99], for which the systems dynamics has a sensible slowing down (i.e. when the system is close to the disappeared fixed point, it remains "trapped" for a longer time close to it, in comparison with other regions). This behaviour results in the presence of a peak in the stationary distribution of the corresponding stochastic systems, that thus remains bistable also when the deterministic system is not anymore.

Another difference is observed: for some parameter values the deterministic system is bistable, but the stochastic distribution shows a clear peak for the

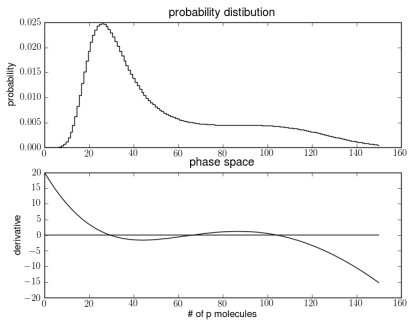


Figure 2.4: Comparison between the deterministic solution (bottom) and the stationary distribution (top) for the parameter set as in Table 2.1, case 4.

maximum with the largest basin of attraction and the smaller peak results "masked" by the tail of the distribution around the first peak (see Fig. 2.4), thus resulting in a monomodal distribution with a long tail. In practice, the highest state behaves like a sort of metastable state, since the states of the system with a high protein level are visited only occasionally.

2.4.2 Numerical analysis

Here we implemented numerical methods to find the stationary distribution of a CME. The most accurate is the Kernel resolution method: given the complete transition matrix of the system, it is possible to solve numerically the eigenvalue problem, obtaining the correct stationary distribution. This method, in this case, has a serious drawback: the system is of non-finite size, preventing a complete enumeration of the possible states. Even with a truncation, the system size rises in a dramatic way: the state space for a bidimensional system is of order N^2 if N is the truncation limit, and thus the respective transition matrix is of order N^4. This means that even for a relatively small system (with a few hundred of molecules) the matrix size explodes well beyond the computational limits. The only feasible resolution strategy is a massive exploration of state space by Montecarlo methods, in which single trajectories of the system are simulated: performing this simulations long enough for several times allows to estimate the stationary distribution.

The Montecarlo method we chose is a modified version of the SSA algorithm (also known as the Gillespie algorithm) named logarithmic direct method [26, 61], which is a statistically correct simulation of an ergodic Markov system. It is not the fastest algorithm available, as compared to other methods like the next-reaction or the τ-leap method, but it produces a correct estimation of the statistical dispersion of the final state.

For each parameter set we performed 10 simulations for about $10^6 - 10^7$ iteration steps each. The multiple simulations were averaged together for a better estimation of the stationary distribution, and they allowed also an estimation of the variance over this average distribution.

In the following we discuss four cases that describe the system behaviour for different parameter settings, shown in Table 2.1.

Table 2.1: Table of the parameter sets for the cases considered.

Par	Case 1	Case 2	Case 3	Case 4
α (molecule/h)	1.0	1.68	1.0	20.0
δ (h^{-1})	1.0	0.20	0.09	1.19
β (molecule/h)	1.0	0.202	0.0	1.0
γ (h^{-1})	100.0	0.20	10.0	1.0
k_1 (molecule/h)	30.0	90.0	12.5	230.0
k_2 ($h^{-}1$)	100.0	0.05	10.0	1.0
Γ_1 (molecule^2)	60.0	10300.0	$(72.8)^2$	$(110.0)^2$
Γ_2 (molecule)	10.0	1006.0	10.0	10.0

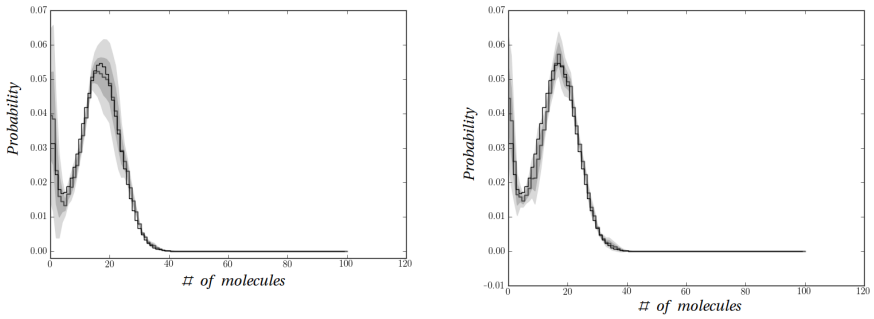


Figure 2.5: Case of good agreement between the theoretical and obtained distribution (see Tab. 2.1, case 1). Left: one-dimensional system, right: two-dimensional system. The thin black line is the theoretical distribution obtained from Eq. 2.15. The thick dark grey line is the average of the various simulations, while the grey and light grey areas represent the range of one and two standard deviations from the average distribution.

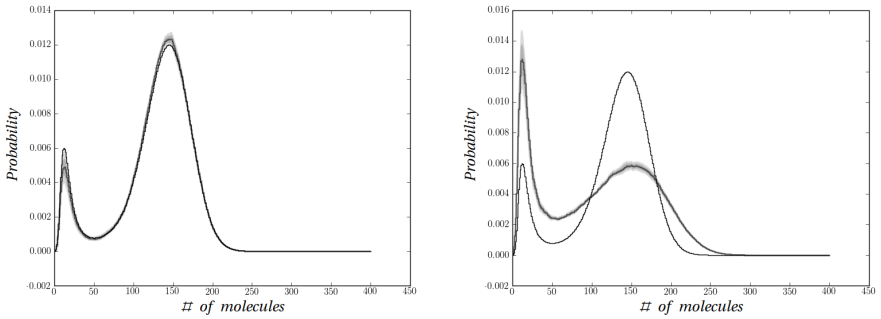


Figure 2.6: Case of poor agreement between the theoretical and obtained distribution (see Tab. 2.1, case 2). Left: one-dimensional system, right: two-dimensional system. The thin black line is the theoretical distribution obtained from Eq. 2.15. The thick dark grey line is the average of the various simulation, while the grey and light grey areas represent the range of one and two standard deviations from the average distribution.

In case 1, we have a system in which the hypothesis of a time-scale separation between m and p is strongly satisfied. The simulation was performed up to a time limit of 10^3: we can see how the two resulting distributions are in good agreement with the theoretical one (see Fig. 2.5), with the regions of higher variance of the histogram around the maxima and minima of the distribution.

In case 2, the time-scale separation assumption does not hold, due to the very low values of γ and k_2: even if this condition doesn't guarantee that the stationary state will be different from the approximate one-dimensional solution, with this set of parameters we can observe a large difference between the two distributions (Fig. 2.6).

In case 3, as defined before, we observe a "ghost" in which, even if a deterministic stable state does not exist, there is clearly a second peak in the distribution (Fig. 2.7). In this system the time-scale separation assumption holds, and we can see how both distributions show similar features.

In this final case (Tab. 2.1, case 4, Fig. 2.8) we observe another effect, in which the peak related to a deterministic stable state is masked by the tail of the stronger peak, becoming just a fat tail. Even without a strong time-scale separation for the m and p variables, both systems give a very similar response, evidencing that this effect is very robust. It is noteworthy that the increase of the γ and k_2 values does not affect the distribution as long as their ratio is kept constant. Note that while there are several computational tools for discrete-state Markov processes such as PRISM [59], APNNtoolbox [11], SHARPE [50], or Mobius [22], there is very little for CMTC (see for instance [23]). Different modeling approaches for toggle switches do exists in the area of formal methods (see for example [36, 35]).

2.5 Discussion of the results

We have studied a stochastic version of a biochemical circuit (the toggle switch) that is supposed to be involved in cell cycle control, with implications for the

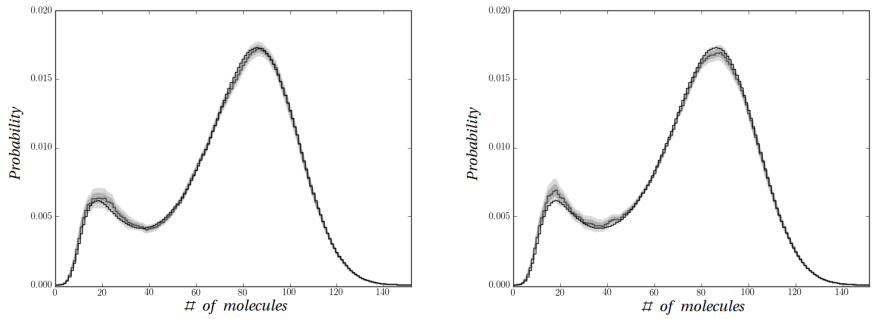


Figure 2.7: Case 3, "ghost effect": only the biggest peak comes from a deterministic stable point. Left: one-dimensional system, right: two-dimensional system. The thick dark gray line is the average of the various simulation, while the gray and light gray areas represent the range of one and two standard deviations from the average distribution.

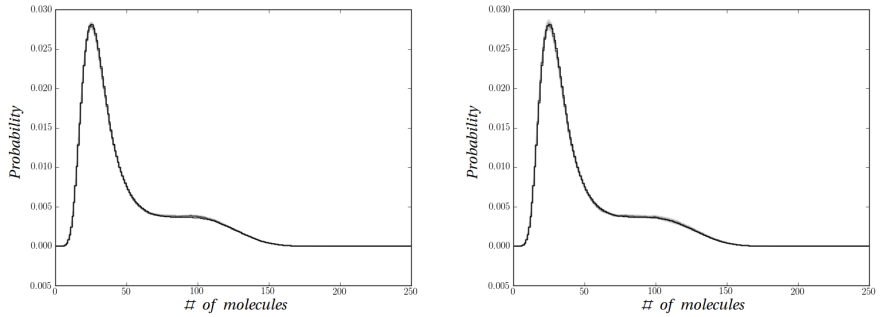


Figure 2.8: Case 4, peak masking effect (parameters as in Tab. 2.1, case 4). The deterministic system has two stable points, but only the peak related to the smallest stable point (with the largest basin of attraction) is visible. Left: one-dimensional system, right: two-dimensional system.

onset of severe diseases such as cancer, consisting of a gene cluster (Myc-E2F) and a miRNA cluster (mir-17-92). This cluster has been reported in a very large number of cancer types: particularly in different types of lymphomas, glioma, non-small cell lung cancer, bladder cancer, squamous-cell carcinoma of head and neck, peripheral nerve sheath tumor, malignant fibrous histiocytoma, alveolar rhabdomyosarcoma, liposarcoma and colon carcinomas. This huge variety of cancer stresses the centrality of this toggle switch and suggests that advancements in its modelling could lead to insights into differences between these cancers. This aim is still far but we are delighted to report that our modeling approach shows important results inching to that direction. First of all, many features are recovered as observed for the deterministic version of the same system, also by means of a further approximation that reduces the system to a unique variable: in this case the system can be treated analytically, and compared to the one- and two-dimensional numerical simulations.

The stochastic approach, that is the exact approach when the number of molecules involved is low, shows a different behaviour than the deterministic one in two situations we have observed. It is noteworthy that the number of molecules involved shows some agreement with the estimates by [16] and by [64] for other miRNA-systems (see also [2]). The cell volume is assumed to be about 10^{-13} liters, then 1 nM = 100 molecules.

First, bistability in the stochastic system (namely, the possibility of having two stable states, one associated to a resting and the other to a proliferative cell state) is observed also in situations in which the corresponding deterministic system is monostable, and this can be explained by the presence of a "ghost" state in the deterministic system that is strong enough to produce a second peak in the stationary distribution of the stochastic model.

Secondly, there are situations in which the peak for the stochastic distribution related to the highest level of expression (with parameter values for which the deterministic system is bistable) is masked by the tail of the distribution of the lowest-expression maximum (that is related to the largest basin of attraction in the deterministic model), making the "proliferative state" appear almost as a scarcely visited metastable state. This is an interesting behaviour, that should be further investigated in real experimental data of protein concentration and gene expression related to the biochemical circuit considered. The "metastable" and the "fully" bimodal distributions could be associated to healthy and tumoral cell states respectively, because the highest "proliferative" state has different properties in the two cases. From a biological point of view, such state, being associated to a dysregulated, disease-related conditions, could actually represent a compendium of several dysregulated states.

We argue that the deterministic approach to this biochemical circuit is not capable to characterize it completely, and the stochastic approach appears more informative: further features unique to the stochastic model could be obtained by considering different time patterns for the molecular influxes to the system, and this point in our opinion should deserve more investigation in a future work. MicroRNAs (miRNAs) express differently in normal and cancerous tissues and thus are regarded as potent cancer biomarkers for early diagnosis. We believe that the potential use of oncomirs in cancer diagnosis, therapies and prognosis will benefit from accurate cancer mathematical models.

Given that MiR-17-92 seems to act as both oncogene and tumor suppressor through decreasing the expression levels of anti-proliferative genes and prolifer-

ative genes, this behavior is suggestive of a cell-type dependent toggle switch. Therefore, fitting experimental data could provide insights into differences among cancer types and on which cell type is behaving differently. The fitting of experimental data with respect to models with different values for the Hill coefficients could also be interesting towards understanding better the chemistry physics of the real microRNA system. Moreover, the comparison between the shape of the expression distributions of the genes/proteins involved in the circuit (and not only the average expression) considering normal and tumoral cells for different cell types, should provide experimental evidence for the different behaviour described from a theoretical point of view in our work, namely the possibility that normal and tumoral cells are in different proliferative "stationary states".

Chapter 3

The futile cycle

In this work we examine a very important biochemical process, the futile cycle, where two enzymes compete to phosphorylate or dephosphorylate a certain substrate. As long as only one phosphorylation site is present on the substrate, this reaction is known to be both deterministically and stochastically monostable. A study of the effects of extrinsic noise is then carried out: this noise is known to lead the system to bistability under the right conditions. These has been studied with an analytical method using the Fokker-Planck limit of the master equation, revealing that is not the level of noise of a single enzyme that drives the transition, but the variance of their proportion. This has been verified empirically comparing two different models of extrinsic noise with different variance-mean proportion, and correlation between the two enzymes. For both models, the significant parameter is shown to be the enzyme proportion's variance with a clear threshold that it is independent from the chosen model, and to be in a good agreement with the analytical prediction.

REFERENCES:

Title: Analysis of noise-induced bistability in Michaelis Menten single-step enzymatic cycle

Authors: Daniel Remondini, Enrico Giampieri, Armando Bazzani, Gastone Castellani and Amos Maritan

Journal: Physical Review E - 2012 (in review)

3.1 Motivation of the work

Many biological phenomena (e.g. memory induction, chromatin remodeling, cell-fate determination) are receiving a great deal of attention in recent years, due to an increasing interest in the description of their complex behaviour by means of basic biochemical circuitry [38, 6, 63]. The signal transduction machinery is mainly based on enzymatic reactions, whose average kinetics can be described within the framework initally proposed by Michaelis and Menten (MM). The steady state approximation of MM model accounts for the majority of known enzymatic reactions, and can be adjusted for the description of regulatory properties such as cooperativity, allostericity and activation/inhibition [74, 88]. The MM equations are still valid at small molecule numbers (as it frequently happens in real cells) if the microscopic interpretation is changed correspondingly [90, 29, 40], but since the discrete stochastic aspects become predominant a deterministic or a stochastic continuous model usually cannot grasp the system features in sufficient detail [92, 25, 105].

A large class of enzymatic reactions controls the reversible addition and removal of phosphoric groups, phosphorylation/dephosphorylation reactions catalyzed by kinases and phosphatases respectively. The phospho/dephosphorylation cycle (PdPC) is thus a post-translational substrate modification that is central for the regulation of several biological processes [58, 95].

How these processes can show a bistable behaviour [79, 37] in the presence of fluctuations [9], reflected by a bimodal stationary distribution of protein number/concentration, is a crucial question for their modeling. In particular, this point is relevant when considering measurements averaged over cell populations with respect to single-cell measurements [105].

The deterministic version of a single PdPC is not bistable in general, but it is hypothesized that external multiplicative noise can trigger such a behavour [95]. The approaches considered previously don't take into account the real nature of the fluctuations due to the finite system size (essentially considering a gaussian noise with arbitrary mean and variance). For this purpose it is necessary to study this cycle by a Chemical Master Equation approach, and consider the continuous approximation only to write explicit conditions on the system parameters that allow bistability. This is the aim of the present paper. A closed form for the stationary distribution of the system (considering only substrate fluctuations) is obtained: we show that the intrinsic noise on the substrates involved in the cycle can not have a bimodal stationary distribution, but additional external noise obtained by plausible biological mechanisms (i.e. the coupling of the system with an enzyme production/activation reaction) can produce such feature, whereas the simple addition of single-enzyme intrinsic noise is not sufficient. Under simplified hypotheses, we compute analytically the conditions in which bimodality occurs, as a function of the reaction parameters (kinetic constants) and system size (number of enzyme and substrate molecules), and we verify our results by numerical simulations with a Gillespie algorithm.

3.2 The model

The PdPC (also referred to as the futile cycle) is composed by one phosphorylation and one dephosphorylation reaction, catalyzed respectively by enzymes

E_1 and E_2:

$$\begin{aligned} E_1 + A &\rightleftharpoons E_1A \rightarrow A^P + E_1 \\ E_2 + A^P &\rightleftharpoons E_2A^P \rightarrow A + E_2 \end{aligned} \tag{3.1}$$

The deterministic dynamics of this cycle can be described via the MM formalism. Assuming a steady-state approximation for both enzymatic reactions $\dot{A}$ and $\dot{A^P}$, we obtain the following equations:

$$\dot{A^P} = v_1 - v_2 \qquad \dot{A} = v_2 - v_1$$

where

$$\begin{aligned} v_1 &= K_{C1} \cdot E_1 \frac{A}{K_{M1} + A} = V_{M1} \frac{A}{K_{M1} + A} \\ v_2 &= K_{C2} \cdot E_2 \frac{A^P}{K_{M2} + A^P} = V_{M2} \frac{A^P}{K_{M2} + A^P} \end{aligned} \tag{3.2}$$

Imposing the conservation of the total substrate concentration, let x be the A molecule concentration, we obtain:

$$\dot{x} = V_{M2} \frac{1 - x}{K_{M2} + 1 - x} - V_{M1} \frac{x}{K_{M1} + x}, \tag{3.3}$$

that can be easily shown to have only one solution in the substrate domain (see [95]).

3.2.1 The CME approach

Starting from the previous equations, a Chemical Master Equation (CME) approach [106] is introduced to account for intrinsic noise on the substrate A (p_n is the A-molecule distribution function over the possible states $n \in [0 : N]$, $D_+ f(n) = f(n+1) - f(n)$):

$$\dot{p_n} = D_+ J; \quad J = r_n p_n - g_{n-1} p_{n-1} \tag{3.4}$$

where

$$r_n = V'_{M1} \frac{n}{K'_{M1} + n} \qquad g_n = V'_{M2} \frac{N - n}{K'_{M2} + N - n}$$

N is the total number of the substrate molecules, n is the the number of A molecules and the MM constants have been accordingly scaled: $K'_M = N \cdot K_M$ and $V'_M = N \cdot V_M$. In the hypothesis of fast relaxation times, the stationary solution of eq. (3.4) describes the statistical properties of the reaction in Fig. 3.1. The stationary distribution p^s_n is derived by imposing $\dot{p_n}(t) = 0$; excluding the existence of a constant current in the system, we get the condition

$$\frac{p^s_n}{p^s_{n-1}} = \frac{g_{n-1}}{r_n} \quad \Rightarrow \quad D_+ \ln p^s(n) = \ln \frac{g_n}{r_{n+1}}$$

If we define a potential $V(n)$, such that $D_+ V(n) = -\ln(g_n / r_{n+1})$, the stationary solution has the Boltzmann form

$$p^s_n = F \cdot e^{-V(n)}, \tag{3.5}$$

where F is a normalizing constant. According to (3.5), the maxima and minima of the distribution are obtained by imposing $D_+V(n) = 0$, extending the n domain to the set of real numbers. This leads to $g_n/r_{n+1} = 1$, similar to (3.3) and with an unique solution inside the $[0 : N]$ domain. This result is also confirmed by Gillespie simulations of the dynamical process.

3.2.2 Bimodality induced by enzyme noise

Relaxing the assumption of fixed enzyme concentration, we characterize the effect of enzyme fluctuations on the substrate stationary distribution. In the symmetric case $K'_{M1} = K'_{M2} = K'_M$and $K'_{C1} = K'_{C2}$, the equilibrium points of the average equation (3.2) corresponding to maxima of p^s_n can be calculated explicitly as a function of the ratio between enzymes γ:

$$\gamma = \frac{E_1}{E_2} = \frac{N-n+1}{n}\frac{K'_M+n}{K'_M+N-n+1} \tag{3.6}$$

If one introduces the variable u:

$$u = \frac{n}{N} - \frac{N+1}{2N} = \frac{n}{N} - a \qquad u \in \left[-a, +a+\frac{1}{N}\right] \tag{3.7}$$

where $a \simeq 1/2$ for $N \gg 1$, the condition (3.6) reads

$$\gamma = \frac{a-u}{a+u}\cdot\frac{K_M+a+u}{K_M+a-u},$$

where $K_M = K'_M/N$. Assuming that $K_M \ll 1$, so that the critical point is quite sensitive to the enzyme concentration, and performing a perturbative approach over K_M, the previous equation can be rewritten as

$$\left(\frac{1}{a+u} - \frac{1}{a-u}\right) = \frac{\gamma-1}{K_M} \tag{3.8}$$

When $\gamma = 1$ (i.e. $E_1 = E_2$) we have the trivial solution $u = 0$ (an unique maximum with $n = (N+1)/2$, $x = 1/2$), whereas for $\gamma - 1 > 0$ (resp. < 0) u shifts towards $-a$ (resp. a).
Supposing that enzyme concentration can fluctuate around the average value, given $\xi = (\gamma-1)/K_M$ and $p(\xi)$ the corresponding probability distribution, the probability distribution $p(u)$ reads

$$p(u) = p(\xi(u))\left|\frac{d\xi}{du}\right| = p(\xi(u))\left(\frac{1}{(a+u)^2} + \frac{1}{(a-u)^2}\right). \tag{3.9}$$

Under the hypotheses that ξ fluctuates around zero and $p(\xi)$ tends sufficiently fast to zero at the boundaries (natural boundary condition), we study the conditions for bimodality of $p(u)$. The critical points of $p(u)$ must satisfy

$$\frac{dp(u)}{du} = \frac{d^2\xi}{du^2}p(\xi(u)) + \left(\frac{d\xi}{du}\right)^2\frac{dp}{d\xi} = 0 \tag{3.10}$$

If we approximate $p(\xi)$ with a Gaussian distribution (justified for a sufficiently large enzyme molecule number and K sufficiently small, see Fig. 3.1) so that

$$p(u) = \left(\frac{1}{(a+u)^2} + \frac{1}{(a-u)^2}\right)e^{-\frac{1}{2\sigma^2_\xi}\left(\frac{1}{a+u}-\frac{1}{a-u}\right)^2}$$

the equation (3.10) reads

$$\frac{1}{\sigma_\xi^2}\left(\frac{1}{(a+u)^2}+\frac{1}{(a-u)^2}\right)^2\left(\frac{1}{a+u}-\frac{1}{a-u}\right) \\ -2\left(\frac{1}{(a+u)^3}-\frac{1}{(a-u)^3}\right)=0 \tag{3.11}$$

If we exclude the symmetric solution $u = 0$, we get the following condition for bimodality (recalling that $\sigma_\xi = \sigma_\gamma/K_M$)

$$\sigma_\xi^2 > \frac{2}{3a^2} \qquad \Rightarrow \qquad \sigma_\gamma^2 > \frac{2K_M^2}{3a^2} \tag{3.12}$$

The condition (3.12) can thus be realized if the enzymatic concentrations fluctuate largely enough around the mean, and as it will be shown in the next section this fluctuation depends on enzyme number.

3.3 Numerical simulations for two biological circuits

As previously shown, substrate intrinsic noise cannot induce bimodality in PdPC, but enzyme fluctuations can produce this effect for K_M sufficiently small by coupling the initial system (3.1) with a set of reactions involving the enzymes of the cycle. We have considered two cases: 1) uncorrelated Poisson fluctuations on both enzymes, and 2) two enzymes whose fluctuation is coupled by a mechanism of competitive induction (e.g. by an effector molecule [97]). The stochastic dynamics of these systems has been implemented numerically by means of the Gillespie algorithm. We considered an identical variance for the stationary distribution of single enzymes in both systems (that can be calculated analytically) that produced a different variance in the distribution of enzyme ratio (calculated numerically from the simulations with the Gillespie algorithm), and thus resulted in different degrees of bimodality as predicted by Eq. (3.12).

In system 1), we have two (identical) equations for enzyme creation/destruction $\dot{E}_i = K_C - K_D \cdot E_i$, that can be described by a Master equation for each enzyme with $g_m = K_C$ and $r_m = K_D \cdot m$: these equations produce the fluctuating number of E_1, E_2 enzymes contained in eq. (3.2,3.3). The stationary distribution is a Poisson distribution, with the average number of molecules and the variance given by $\langle E\rangle = \sigma_E^2 = K_C/K_D$.

For system 2), defining as E_1^* and E_2^* the inactive (or alternatively, located externally to the reaction region) enzyme concentrations, and E_A the concentration of activating molecules, the full kinetic reaction scheme becomes:

$$\begin{aligned} E_A + E_1^* &\rightleftharpoons E_1 \\ E_A + E_2^* &\rightleftharpoons E_2 \\ E_1 + A \rightleftharpoons E_1A &\rightarrow A^P + E_1 \\ E_2 + A^P \rightleftharpoons E_2A^P &\rightarrow A + E_2 \end{aligned}$$

A simplified version of the enzyme competition can be obtained by considering a direct interchange between the two active enzymes: $E_1 \rightleftharpoons E_2$, with $E_2 =$

$E_T - E_1$, E_T the total enzyme concentration. The kinetic equation for this system thus reduces to $\dot{E}_1 = K_1 \cdot (E_T - E_1) - K_2 \cdot E_1$, thus the corresponding Master equation has $g_m = K_1 \cdot (M_T - m)$ and $r_m = K_2 \cdot m$. The stationary distribution of this equation is of Binomial type, with a mean value $\langle E \rangle = E_T/2$ and a variance $\sigma_E^2 = E_T/4$ (considering $K_1 = K_2$). In the limit of large enzyme molecule number, these systems coincide with the enzymatic cycle in eq. (3.1), and the variance of the enzyme ratio distribution $p(\gamma)$ tends to zero; thus there is a limit system size in order to observe bimodality.

Applying the calculation performed previously, we have shown that the stationary distribution is a bimodal distribution if the condition (3.12) is satisfied under the simplified assumption that $p(\gamma)$ can be appoximated by a Gaussian function. In Fig. 3.1 we numerically check our approximation for a moderate number of enzymes. For even lower enzyme molecule numbers (in our simulations $\langle E \rangle$ ranges from unity to hundreds) the gaussianity condition is not strictly satisfied, but our criterion is in qualitative agreement with our simulations anyway (see Fig. 3.2).

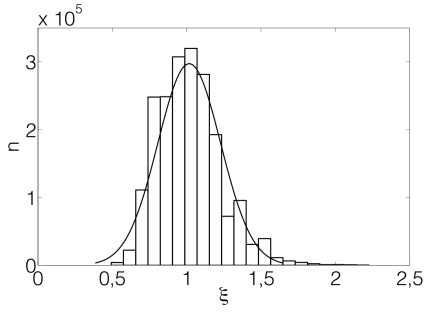


Figure 3.1: Gaussian approximation of $p(\gamma)$ ($\gamma = E_1/E_2$) for $E_1 + E_2 = 100$. Bars: empirical distribution of γ; continuous line: gaussian distribution with same mean and variance.

In Fig. 3.2 we show the results of our simulations for both systems. In order to characterize the effect of system size on bimodality, we fixed the chemical reaction constants of substrate reactions in (3.2) $K_{M1} = K_{M2} = 0.1$, $K_{C1} = K_{C2} = 10$ and we considered an increasing number of total substrate molecules N_T ranging from 100 to 1000. Moreover, we scaled the enzyme-substrate system by fixing the ratio between average enzyme and total substrate molecule number, with a different ratio for the two systems in order to have the same variance in single enzyme distribution: $\langle E \rangle / X_T = 0.3$ (remembering that $E_T = 2 \cdot \langle E \rangle$), $K_1 = K_2 = 1$ for system 1. For system 2, we adjusted the kinetic constants accordingly: $K_C = X_T \cdot 0.15$, $K_D = 1$.

We have numerically checked the agreement with our theoretical results by looking at the value for which bimodality becomes negligible, that corresponds approximately to $N > 400$ for the binomial system and to $N > 900$ for the Poisson system. In Fig. (3.2, top) the corresponding σ_γ^2 values are crossed by the same horizontal line (estimated from the picture below) confirming the existence of a unique threshold related only to this parameter, and not for example to the substrate or enzyme molecule number. Numerically, the horizontal line sets a

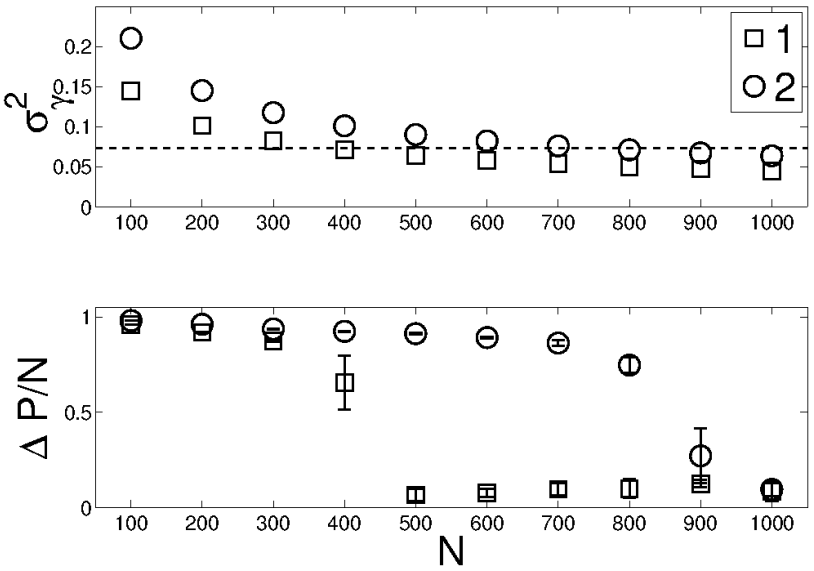


Figure 3.2: Results of simulations for system 1 (circles) and 2 (squares). Top: plot of σ_γ^2 as a function of substrate total number of molecules N; the dashed line shows approximately the critical value of σ_γ^2 for which bimodality is lost (see bottom plot) in both systems. Bottom: plot of peak distance (rescaled over the total substrate molecule number N), calculated as the distance between the maxima of the two half distributions $N \in [0 : N/2]$ and $N \in [N/2 : N]$, by taking advantage of the symmetry of the system.

threshold for σ_γ^2 approximatively equal to 0.07, that corresponds to a value of $K_M = 0.16$, in good agreement with the real $K_M = 0.1$ used in our simulations.

3.4 Discussion of the results

In this paper we study the conditions that result in a noise-induced bimodal stationary distribution of a single phosphorylation/dephosphorylation cycle. The system is described by Michaelis-Menten equations in the quasi-steady state assumption, and studied with a finite number of molecules by a Master Equation approach. We show that the effect of an intrinsic noise on the substrate molecules is not enough to achieve bimodality in this system. Therefore, we propose a different approach to generate stochastic bimodality (that does not change the number of deterministic stable states, thus a purely noise-driven phenomenon) by coupling the cycle with a reaction scheme for enzyme turnover. For this system we clarify the conditions for which stochastic fluctuations in enzyme concentration can lead to bimodality in substrate concentration in a continuous approximation to the model. We show that it depends on the distribution of the ratio between the two enzymes involved in the reaction, and on one of the kinetic constants related to the substrate reaction, K_M, that determines the susceptibility of the system to enzyme fluctuations. In the final section we implement, by means of Gillespie algorithm, two different reaction schemes that can be considered biologically plausible for modeling enzyme turnover. The simulation results are in good agreement with our simplified analytical calculations, and moreover they emphasize that bimodality depends indirectly on the system size (i.e. the number of molecules involved). We remark that the number of molecules used in our simulations are of the same order of magnitude than some molecules known to have a central role in cell regulation (such as Myc [3, 38] and P53 [60] oncogenes). These results define in more detail the conditions for which this phenomenon could occur in real biological systems, stating that if noise has to be exploited to achieve a bimodal behaviour there must be a relationship between the chemical parameters of the system and its size.

Chapter 4

The double phosphorylation cycle

In the following work we study a system composed by two connected futile cycles, i.e. the substrate has two different phosphorylation sites. This system is known to be deterministically bistable, and we verify that this is true even for the stochastic version of the model. Then we proceed to study a perturbative method to evaluate the stationary distribution, starting from a detailed balance approximation of the process and using this to evaluate the non-equilibrium probability fluxes. This approximation is always possible as long as the system is bidimensional, and is based on a discrete equivalent of the Helmholtz field decomposition theorem, which allows us to write a vector field as a sum of a scalar potential (the basis for the detailed balance principle) and a vector potential (which breaks the microscopical reversibility and so the detailed balance of the system).

REFERENCES:

Title: Bistability in the Chemical Master Equation for Dual Phosphorylation Cycles

Authors: Armando Bazzani, Gastone C. Castellani, Enrico Giampieri, Daniel Remondini, and Leon N Cooper

Journal: Journal of Chemical Physics - 2012 (in review)

4.1 Motivation of the work

One of the most important aspects of biological systems is their capacity to learn and memorize patterns and to adapt themselves to exogenous and endogenous stimuli by tuning signal transduction pathways activity. The mechanistic description of this behavior is typically depicted as a "switch" that can drive the cell fate to different stable states characterized by some observables such as levels of proteins, messengers, organelles or phenotypes [114]. The biochemical machinery of signal transduction pathways is largely based on enzymatic reactions, whose average kinetic can be described within the framework of chemical kinetics and enzyme reactions as pioneered by Michaelis and Menten [98, 72]. The steady state velocity equation accounts for the majority of known enzymatic reactions, and can be adjusted to the description of regulatory properties such as cooperativity, allostericity and activation/inhibition[74]. Theoretical interest in enzymatic reactions has never stopped since Michaelis-Menten's work and has lead to new discoveries such as zero-order ultrasensitivity [42, 9]. Among various enzymatic processes, a wide and important class comprises the reversible addition and removal of phosphoric groups via phosphorylation and dephosphorylation reactions catalyzed by kinases and phosphatases. The phospho/dephosphorylation cycle (PdPC) is a reversible post-translational substrate modification that is central to cellular signalling regulation and can play a key role in the switch phenomenon for several biological processes [58, 95]. Dual PdPC's are classified as homogeneous and heterogeneous based on the number of different kinases and phosphatases [52]; the homogeneous has one kinase and one phosphatase, while the heterogeneous has two kinases and two phosphatases. Variants of homogeneous and heterogeneous dual PdPC's may only have a non-specific phosphatase and two specific kinases [91] or, symmetrically, a non-specific kinase and two specific phosphatases. The PdPC with the non-specific phosphatase controls the phosphorylation state of AMPA receptors that mediates induction of Long Term Potentiation (LTP) and Long Term Depression (LTD) in vitro and in vivo[91, 37, 48]. Recently, several authors have reported bistability in homogeneus pPdPC [79, 52]) as well as those with a non-specific phosphatase and two different kinases [37]. The bistable behaviour of the homogeneous system is explained on the basis of a competition between the substrates for the enzymes. The majority of studies on biophysical analysis of phospho/dephosphorylation cycle have been performed in a averaged, deterministic framework based on Michaelis-Menten (MM) approach, using the steady state approximation. However, recently some authors [70] have pointed to the role of fluctuations in the dynamics of biochemical reactions. Indeed, in a single cell, the concentration of molecules (substrates and enzymes) can be low, and thus it is necessary to study the PdPC cycle within a stochastic framework. A "natural" way to cope with this problem is the so-called Chemical Master Equation (CME) approach [106], that realizes in an exact way the probabilistic dynamics of a finite number of molecules, and recovers the chemical kinetics of the Law of Mass Action, which yields the continuous Michaelis-Menten equation in the thermodynamic limit ($N \to \infty$,) using the mean field approximation. In this paper we study a stochastic formulation of enzymatic cycles that has been extensively considered by several authors [52, 79]. The deterministic descriptions of these models characterize the stability of fixed points and give a geometrical interpretation of the observed steady states, as the intersection of conic

curves[79]. The stochastic description can in fact provide further information on the relative stability of the different steady states in terms of a stationary distribution. We propose a perturbative approach for computing the stationary solution out of the thermodynamics equilibrium. We also point out the role of currents in the transition from a mono-modal distribution to a bimodal distribution; this corresponds to bifurcation in the deterministic approach. The possibility that chemical fluxes control the distribution shape suggests a generic mechanism used by biochemical systems out of thermodynamic equilibrium to obtain a plastic behavior. Moreover, we show that at the bimodal transition there exists a diffusion region in the configuration space where a Fokker-Planck equation can be introduced to approximate the stationary solution. Analogous models have been previously studied [9, 86, 89] for single-step PdPC.

4.2 Dual phosporylation/dephosphorylation enzymatic cycles

The process shown in Fig. (4.1) is a two-step chain of addition/removal reactions of chemical groups and may, in general, model important biological processes such as phenotype switching (ultrasensitivity) and chromatine modification by histone acetylation/deacetylation as well as phospho/dephosphorylation reactions. Without loss of generality, we perform a detailed study of the homogeneous phospho/dephosphorylation two-step cycles (PdPC cycles) where two enzymes drive phosphorylation and dephosphorylation respectively. Thus, there is a competition between the two cycles for the advancement of the respective reactions.

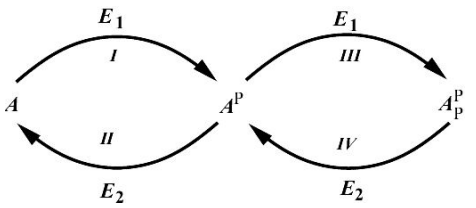


Figure 4.1: Scheme of the double enzymatic cycle of addition/removal reactions of chemical groups via Michelis-Menten kinetic equations as shown in eq (4.1) in the case of phosphoric groups.

The deterministic Michaelis-Menten (MM) equations of the scheme (4.1) with the quasi-steady-state hypothesis reads:

$$\begin{aligned}
\frac{dA}{dt} &= \frac{k_{M_4} v_{M_2}(1 - A - A_P^P)}{k_{M_2}k_{M_4} + k_{M_4}(1 - A - A_P^P) + k_{M_2}A_P^P} \\
&- \frac{k_{M_3} v_{M_1} A}{k_{M_1}k_{M_3} + k_{M_1}(1 - A - A_P^P) + k_{M_3}A} \\
\frac{dA_P^P}{dt} &= \frac{k_{M_1} v_{M_3}(1 - A - A_P^P)}{k_{M_1}k_{M_3} + k_{M_1}(1 - A - A_P^P) + k_{M_3}A} \\
&- \frac{k_{M_2} v_{M_4} A_P^P}{k_{M_2}k_{M_4} + k_{M_4}(1 - A - A_P^P) + k_{M_2}A_P^P}
\end{aligned} \tag{4.1}$$

where A and A_P^P are the concentrations of the non-phosphorylated and double phosphorylated substrates, k_{M_i} denote the MM constants and v_{M_i} are the maximal reaction velocities ($i = 1, \cdots 4$). Let n_1 and n_2 denote the molecules number of the substrates A and A_P^P respectively, the corresponding CME for the probability distribution $\rho(n_1, n_2, t)$ is written

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &= g_1(n_1 - 1, n_2)\rho(n_1 - 1, n_2, t) - g_1(n_1, n_2)\rho(n_1, n_2, t) \\
&+ r_1(n_1 + 1, n_2)\rho(n_1 + 1, n_2, t) - r_1(n_1, n_2)\rho(n_1, n_2, t) \\
&+ g_2(n_1, n_2 - 1)\rho(n_1, n_2 - 1, t) - g_2(n_1, n_2)\rho(n_1, n_2, t) \\
&+ r_2(n_1, n_2 + 1)\rho(n_1, n_2 + 1, t) - r_2(n_1, n_2)\rho(n_1, n_2, t)
\end{aligned} \tag{4.2}$$

where $g_j(n_1, n_2)$ and $r_j(n_1, n_2)$ are the generation and recombination terms respectively, defined as :

$$\begin{aligned}
r_1(n_1, n_2) &= \frac{K_{M_3} v_{M_1} n_1}{K_{M_1}K_{M_3} + K_{M_1}(N_T - n_1 - n_2) + K_{M_3}n_1} \\
g_1(n_1, n_2) &= \frac{K_{M_4} v_{M_2}(N_T - n_1 - n_2)}{K_{M_2}K_{M_4} + K_{M_4}(N_T - n_1 - n_2) + K_{M_2}n_2} \\
r_2(n_1, n_2) &= \frac{K_{M_2} v_{M_4} n_2}{K_{M_2}K_{M_4} + K_{M_4}(N_T - n_1 - n_2) + K_{M_2}n_2} \\
g_2(n_1, n_2) &= \frac{K_{M_1} v_{M_3}(N_T - n_1 - n_2)}{K_{M_1}K_{M_3} + K_{M_1}(N_T - n_1 - n_2) + K_{M_3}n_1}
\end{aligned} \tag{4.3}$$

N_T is the total number of molecules, and we have introduced scaled constants $K_M = N_T k_M$. The biochemical meaning is that the enzyme quantities should scale as the total number of molecule N_T, to have a finite thermodynamic limit $N_T \to \infty$, in the transition rates (4.3). As it is known from the theory of one-step Markov processes, the CME (4.2) has a unique stationary solution $\rho_s(n_1, n_2)$ that describes the statistical properties of the system on a long time scale. The CME recovers the Mass Action-based MM equation (4.1) in the thermodynamic limit when the average field theory approach applies. Indeed it can be shown that the critical points of the stationary distribution for the CME can be approximately computed by the conditions (cfr. eq. (4.15))

$$g_1(n_1, n_2) = r_1(n_1 + 1, n_2) \quad g_2(n_1, n_2) = r_2(n_1, n_2 + 1) \tag{4.4}$$

whose solutions tend to the equilibrium points of the MM equation when the fluctuation effects are reduced in the thermodynamic limit as $O(1/\sqrt{N})$. As a consequence, one would expect that the probability distribution becomes singular, being concentrated at the fixed stable points of the equations (4.1), and that the transition rate among the stability regions of attractive points is negligible. However, when a phase transition occurs due to the bifurcation of the stable solution, fluctuations are relevant even for large N_T and the CME approach is necessary. In the next section we discuss the stationary distribution properties for the CME (4.2).

4.3 The Stationary Distribution

The stationary solution $\rho_s(n_1, n_2)$ of the CME (4.2) can be characterized by a discrete version of the zero divergence condition for the current vector $\vec{J}$ components(see Appendix)

$$\begin{aligned} J_1^s &= g_1(n_1-1,n_2)\rho_s(n_1-1,n_2) - r_1(n_1,n_2)\rho_s(n_1,n_2) \\ J_2^s &= g_2(n_1,n_2-1)\rho_s(n_1,n_2-1) - r_2(n_1,n_2)\rho_s(n_1,n_2) \end{aligned} \tag{4.5}$$

and the CME r.h.s. reads

$$D_1 J_1^s(n_1,n_2) + D_2 J_2^s(n_1,n_2) = 0 \tag{4.6}$$

where we have introduced the difference operators

$$\begin{aligned} D_1 f(n_1,n_2) &= f(n_1+1,n_2) - f(n_1,n_2) \\ D_2 f(n_1,n_2) &= f(n_1,n_2+1) - f(n_1,n_2) \end{aligned} \tag{4.7}$$

Due to the commutative properties of the difference operators, the zero-divergence condition for the current is equivalent to the existence of a current potential $A(n_1, n_2)$ such that

$$\begin{aligned} J_1^s(n_1,n_2) &= D_2 A(n_1,n_2) \quad n_1 \geq 1 \\ J_2^s(n_1,n_2) &= -D_1 A(n_1,n_2) \quad n_2 \geq 1 \end{aligned} \tag{4.8}$$

We remark that the r.h.s. of eq. (4.8) is a discrete version of the curl operator. The potential difference $A(n_1', n_2') - A(n_1, n_2)$ defines the chemical transport across any line connecting the states (n_1, n_2) and (n_1', n_2'). At the stationary state the net transport across any closed path is zero and we have no current source in the network. As discussed in [106, 96] we distinguish two cases: when the potential $A(n_1, n_2)$ is constant (the so called "detailed balance condition") and the converse case corresponding to a non-equilibrium stationary state. In this case the stationary solution ρ_s is characterized by the condition $J_1 = J_2 = 0$ over all the states, whereas in the other case we have macroscopic chemical fluxes in the system. When detailed balance holds, simple algebraic manipulations (see Appendix) result in the following conditions for the stationary solutions

$$D_1 \ln \rho_s(n_1,n_2) = \ln a_1(n_1,n_2) \tag{4.9}$$

$$D_2 \ln \rho_s(n_1,n_2) = \ln a_2(n_1,n_2) \qquad n_1 + n_2 < N_T \tag{4.10}$$

where the discrete drift vector field components a_i are defined by:

$$a_1(n_1, n_2) = \frac{g_1(n_1, n_2)}{r_1(n_1 + 1, n_2)} \quad a_2(n_1, n_2) = \frac{g_2(n_1, n_2)}{r_2(n_1, n_2 + 1)}. \tag{4.11}$$

Equations (9) and (4.10) imply the existence of a potential $V(n_1, n_2)$ such that

$$\begin{aligned} \ln a_1(n_1, n_2) &= -D_1 V(n_1, n_2) \\ \ln a_2(n_1, n_2) &= -D_2 V(n_1, n_2) \end{aligned} \tag{4.12}$$

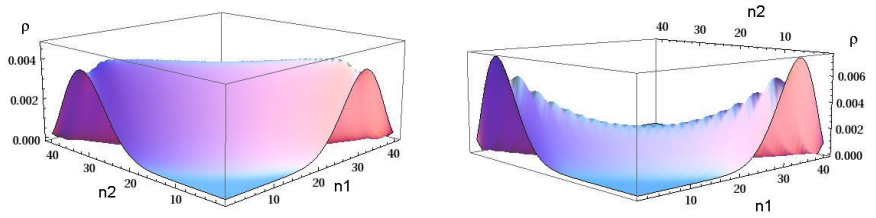


Figure 4.2: Stationary distributions for the A and A_P^P states in the double phosphorylation cycle when detailed balance (4.13) holds with $K_{M_1} = K_{M_4} = 1$ and $K_{M_2} = K_{M_3} = 2$. In the top figure we set the reaction velocities $v_{M_1} = v_{M_2} = 1$ and $v_{M_2} = v_{M_3} = 1.05$ (symmetric case), whereas in the bottom figure we increase the v_{M_2} and v_{M_3} value to 1.15. The number of molecules is $N_T = 40$. The transition from a unimodal distribution to a bimodal distribution is clearly visible.

Using definition (4.3) it is possible to explicitly compute a set of parameter values for the PdPC cycle, that satisfy the detailed balance condition (4.12) according to the relations

$$2K_{M_1} = K_{M_2} = K_{M_3} = 2K_{M_4} \tag{4.13}$$

where the reaction velocities V_M are arbitrary. The stationary distribution is given by the Maxwell-Boltzmann distribution

$$\rho_s(n_1, n_2) = \exp(-V(n_1, n_2)) \tag{4.14}$$

where the potential $V(n_1, n_2)$ is computed by integrating equation (4.12) and choosing the initial value $V(0, 0)$ to normalize the distribution (4.14). Using a thermodynamical analogy, we can interpret the potential difference $V(n_1, n_2) - V(0, 0)$ as the chemical energy needed to reach the state (n_1, n_2) from the initial state $(0, 0)$. As a consequence, the vector field (4.11) represents the work for one-step transition along the n_1 or n_2 direction. Definition (4.14) also implies that the critical points of the stationary distribution are characterized by the conditions

$$\frac{g_1(n_1, n_2)}{r_1(n_1 + 1, n_2)} = \frac{g_2(n_1, n_2)}{r_2(n_1, n_2 + 1)} = 1 \tag{4.15}$$

and coincides with the critical points of the MM equations. In figure 4.2 we plot the stationary distributions (4.14) in the case $N_T = 40$ with $K_{M_1} = K_{M_4} = 1$ and $K_{M_2} = K_{M_3} = 2$; we consider two symmetric cases: $v_{M_1} = v_{M_2} = 1$ with $v_{M_2} = v_{M_3} = 1.05$ or $v_{M_2} = v_{M_3} = 1.15$ (all the units are arbitrary). In the first case, the probability distribution is unimodal, whereas in the second case the transition to a bimodal distribution is observed. Indeed, the system has a phase transition at $v_{M_2} = v_{M_3} \simeq 1.1$ that corresponds to a bifurcation of the critical point defined by the condition (4.15).

In figure 4.2 we distinguish two regions: a drift dominated region and a diffusion dominated region. In the first region the chemical reactions mainly follow the gradient of the potential $V(n_1, n_2)$ and tend to concentrate around the stable critical points, so that the dynamic is well described by a Liouville equation[51]. In the second region the drift field (4.11) is small and the fluctuations due to the finite size introduce a diffusive behaviour. Then the distribution can be approximated by the solution of a Fokker-Planck equation[106]. As discussed in the Appendix, the diffusion region is approximately determined by the conditions

$$\begin{aligned} g_1(n_1, n_2) - r_1(n_1 + 1, n_2) &\simeq g_2(n_1, n_2) - r_2(n_1, n_2 + 1) \\ &\simeq O(1/N_T) \end{aligned} \tag{4.16}$$

To illustrate this phenomenon, we outline in fig. 4.3 the region where condition (4.16) is satisfied (i.e. the gradient of the potential $V(n_1, n_2)$ is close to 0 (4.12)). This is the region where the fixed points of the MM equation are located, and comparison with fig. 4.2 shows that it defines the support of the stationary distribution.

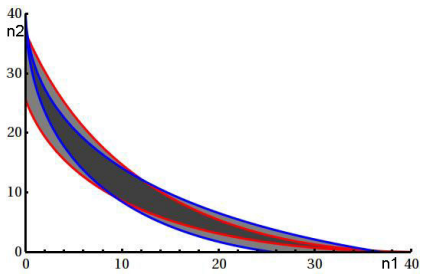


Figure 4.3: In grey we show the region where components of the vector field (4.11) are $\simeq 1$ using the parameter values of fig. 4.2 (bottom). The blue lines enclose the region where the first component is nearby 1, whereas the red ones enclose the corresponding region for the second component.

In the diffusion dominated region a molecule can undergo a transition from the dephosphorylated equilibrium to the double phosphorylated one. At the stationary state the transition probability from one equilibrium to the other can be estimated by the Fokker-Planck approximation; but this does not imply that the Fokker-Planck equation allows us to describe the transient relaxation process toward the stationary state. Indeed, due to the singularity of the thermodynamics limit, the dynamics of transient states may depend critically on finite size effects not described by using the Fokker-Planck approximation[108]. To cope with these problems in the PdPC model further studies are necessary.

When the current (4.8) is not zero the CME (4.2) relaxes toward a Non-Equilibrium Stationary State (NESS) and the field (4.11) is not conservative. We shall perform a perturbative approach to point out the effects of currents on the NESS nearby the transition region to the bimodal regime. We consider the following decomposition for the vector field

$$\begin{aligned} \ln a_1(n_1,n_2) &= -D_1V_0(n_1,n_2) + D_2H(n_1,n_2) \\ \ln a_2(n_1,n_2) &= -D_2V_0(n_1,n_2) - D_1H(n_1,n_2) \\ & \quad n_1+n_2 \leq N_T-1 \end{aligned} \tag{4.17}$$

where the rotor potential $H(n_1,n_2)$ takes into account the irreversible rotational part. The potential $H(n_1,n_2)$ can be recursively computed using the discrete Laplace equation

$$(D_1D_1+D_2D_2)H = D_2\ln\left(a_1(n_1,n_2)\right) - D_1\ln\left(a_2(n_1,n_2)\right) \tag{4.18}$$

where $n_1+n_2 \leq N_T-2$, with the boundary conditions $H(n,N-n) = H(n,N-1-n) = 0$ (see Appendix). Assuming that the potential H is small with respect to V, we can approximate the NESS by using the Maxwell-Boltzmann distribution (4.14) with $V = V_0$. However, as we shall show in the next section, at the phase transition even the effect of small currents becomes critical, and the study of higher perturbative orders is necessary. To point out the relation among the rotor potential H, the NESS ρ_s and the chemical flux J, we compute the first perturbative order letting $\rho_s(n_1,n_2) = \exp(-V_0(n_1,n_2) - V_1(n_1,n_2))$. From definition (4.5) the condition (8) reads:

$$\begin{aligned} &\exp\left(-V_0(n_1,n_2) - V_1(n_1,n_2)\right) r_1(n_1,n_2) \\ &\cdot \left(\exp\left(D_2H(n_1-1,n_2)\right) - \exp\left(-D_1V_1(n_1-1,n_2)\right)\right) \\ &= D_2a(n_1,n_2)) \\ &\exp\left(-V_0(n_1,n_2) - V_1(n_1,n_2)\right) r_2(n_1,n_2) \\ &\cdot \left(\exp\left(-D_1H(n_1,n_2-1)\right) - \exp\left(-D_2V_1(n_1,n_2-1)\right)\right) \\ &= -D_1a(n_1,n_2) \end{aligned} \tag{4.19}$$

$V_1(n_1,n_2)$ turns out to be an effective potential that simulates the current's effect on the unperturbed stationary distribution by using a conservative force. We note that if the rotor potential H is zero, then both the current potential A and the potential correction V_1 are zero, so that all these quantities are of the same perturbative order, and the first perturbative order of eqs. (4.19) reads

$$\begin{aligned} &\exp\left(-V_0(n_1,n_2)\right) r_1(n_1,n_2) \cdot \\ &\left(D_1V_1(n_1-1,n_2) + D_2H(n_1-1,n_2)\right) = D_2A(n_1,n_2) \\ &\exp\left(-V_0(n_1,n_2)\right) r_2(n_1,n_2) \cdot \\ &\left(D_2V_1(n_1,n_2-1) - D_1H(n_1,n_2-1)\right) = -D_1A(n_1,n_2) \end{aligned} \tag{4.20}$$

From the previous equations we see that the currents depend both on the rotor potential H and the potential correction V_1, which is unknown; thus they cannot

be directly computed from (4.20). We obtain an equation for V_1 by eliminating the potential A from (4.20)

$$\begin{aligned} &D_1\left(\exp\left(-V_0(n_1,n_2)\right) r_1(n_1,n_2) D_1 V_1(n_1-1,n_2)\right) \\ &+D_2\left(\exp\left(-V_0(n_1,n_2)\right) r_2(n_1,n_2) D_2 V_1(n_1,n_2-1)\right) = \\ &D_2\left(\exp\left(-V_0(n_1,n_2)\right) r_2(n_1,n_2) D_1 H(n_1,n_2-1)\right) \\ &-D_1\left(\exp\left(-V_0(n_1,n_2)\right) r_1(n_1,n_2) D_2 H(n_1-1,n_2)\right) \end{aligned} \tag{4.21}$$

Eq. (4.21) is defined for $n_1 \geq 1$, $n_2 \geq 1$ and $n_1 + n_2 \leq N_T - 1$, and we can solve the system by introducing the boundary conditions $V_1(n,0) = V_1(0,n) = V_1(n, N_T - n) = 0$. It is interesting to analyze equation (4.21) in the phase transition regime. When the recombination terms r_1,r_2 are almost equal and their variation is small (for our parameter choice this is true for $n_1 \simeq n_2$ and $n_1 + n_2 \gg 1$) the r.h.s. can be approximated by

$$\begin{aligned} \exp(-V_0(n_1,n_2)\,&[D_2 V_0(n_1,n_2) D_1 H(n_1,n_2-1) \\ &-D_1 V_0(n_1,n_2) D_2 H(n_1-1,n_2)] \end{aligned} \tag{4.22}$$

As a consequence, in the transition regime this term is negligible since both $D_1 V_0(n_1,n_2)$ and $D_2 V_0(n_1,n_2)$ tend toward zero in the diffusion region where the bifurcation occurs; thus the first perturbative order is not enough to compute the stationary distribution correction, but higher orders should be considered.

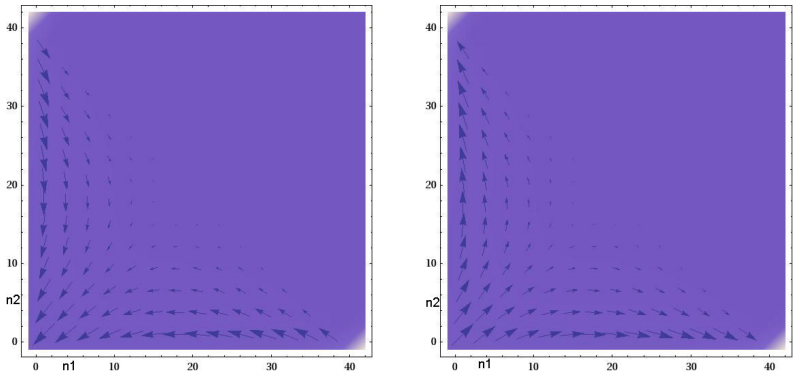


Figure 4.4: Plot of the rotor field for the potential H using the following parameter values: $v_{M1} = v_{M2} = 1$, $v_{M_2} = v_{M_3} = 1.15$ $K_{M_1} = K_{M_4} = 1$, $N_T = 40$ and $K_{M_2} = K_{M_3} = 1.8$ (left picture) or $K_{M_2} = K_{M_3} = 2.2$ (right picture).

4.4 Numerical simulations

In order to study the non equilibrium stationary conditions in the double phosphorylation cycle (4.1) we have perturbed the detailed balance conditions considered in figure (4.2) by changing the value of the MM constants K_{M_2} and K_{M_3}. In figure (4.4) we show the rotor field of the potential H in the cases

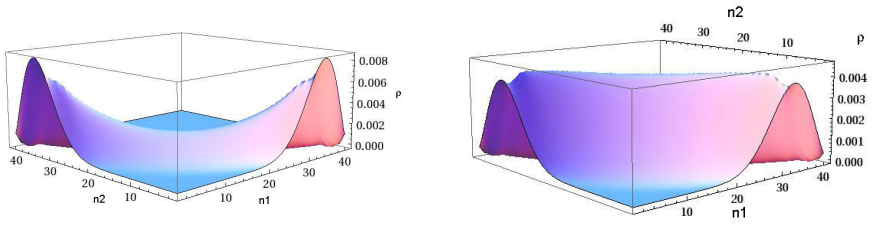


Figure 4.5: (Left picture): plot of the zero-order approximation for the probability distribution using the decomposition (4.17) for the vector field associated to the CME. (Right picture): plot of the stationary distribution computed by directly solving the CME (4.2). We use the parameter values of the case I in the table 4.4.

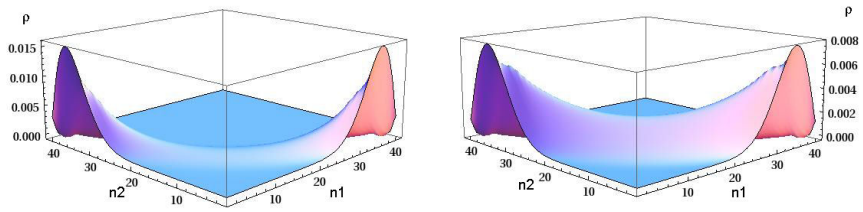


Figure 4.6: The same as in fig. 4.5 using parameter values of case II in table 4.4.

$K_{M_2} = K_{M_3} = 2.2$ and $K_{M_2} = K_{M_3} = 1.8$ when the detailed balance condition (4.13) does not hold. We see that in the first case (left picture) the rotor field tends to move the particles from the borders toward the central region $n_1 = n_2$, so that we expect an increase of the probability distribution in the center, whereas in the second case the rotor field is directed from the central region to the borders and we expect a decrease of the probability distribution in this region. To illustrate the effect of the potential H, we compare the zero-order approximation of the probability distribution (4.14), where the potential $V_0(n_1, n_2)$ is computed using decomposition (4.17) with the stationary solution of the CME (4.2). The main parameter values are reported in table 4.4

Case	K_{M_2}	K_{M_3}	v_{M_2}	v_{M_3}
I	1.8	1.8	1.05	1.05
II	1.8	1.8	1.15	1.15
III	2.2	2.2	1.05	1.05
IV	2.2	2.2	1.15	1.15

whereas the other parameters values are: $v_{M_1} = v_{M_2} = 1$,$K_{M_1} = K_{M_4} = 1$ and $N_T = 40$. For the first parameter set, the zero-order approximation is a bimodal

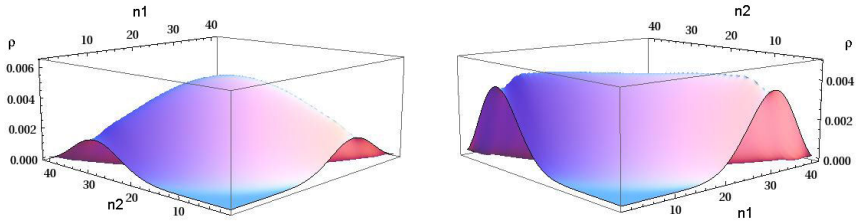


Figure 4.7: The same as in fig. 4.5 using the parameter values of the case III in table 4.4.

distribution, but the effect of the currents induced by the rotor potential H (see fig. 4.4) left) are able to destroy the bimodal behaviour by depressing the two maximal at the border (see fig. 5).

If we increase the value of the K_{2M} and K_{3M} (case II), the exact stationary distribution becomes bimodal and the effect of currents is to introduce a strong transition probability between the two distribution maxima (fig. 4.6).

Therefore, when the MM constants K_{M2} and K_{M3} are < 2 (we note that for $K_{M_2} = K_{M_3} = 2$, the detailed balance holds), the non-conservative nature of the field (4.17) introduces a delay in the phase transition from a mono-modal to a bimodal distribution. However, when we consider $K_{M2} = K_{M3} > 2$ (cases III and IV) the rotor potential H moves the particle towards the borders and the central part of the distribution is depressed. This is shown in the figure 4.7 where we compare the zero-order approximation of the stationary distribution and the solution of the CME using the case III parameters of the table 4.4.

Finally, in case IV of table 4.4, the CME stationary solution undergoes a transition to a bimodal distribution, whereas the zero-order approximation is still mono-modal (fig. 4.9), so that the effect of currents is to anticipate the phase transition.

Using the stationary solution one can also compute the currents according to definition (4.5). In figure (9) we plot the current vector in case IV parameters to show that the current tends to become normal to the distribution gradient near the maximal value.

This result can be also understood using the perturbative approach (4.21), where one shows that the main effect of the V_1 potential correction is to compensate the rotor field of H along the distribution gradient directions. As a consequence, the current is zero at the maximal distribution value and condition (4.15) defines the critical points of the stationary distribution even in the non-conservative case.

4.5 Discussion of the results

The CME approach we present here is a powerful method for studying complex cellular processes, even with significant simplifications such as spatial homogeneity of volumes where the chemical reactions are taking place. The CME theory is attractive for a variety of reasons, including the richness of aspects (the

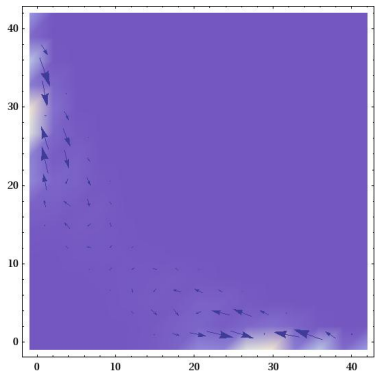


Figure 4.8: Current vector field computed by using definiton (4.5) and the stationary solution of the CME with case IV parameters. The distribution is bimodal (cfr. figure 8) and the current lines tend to be orthogonal to the distribution gradient near the maximal value.

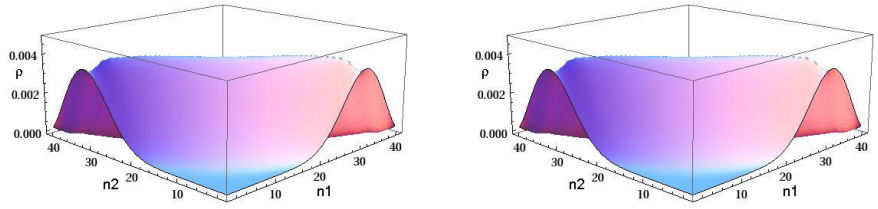


Figure 4.9: The same as in fig. 4.5 using parameter values of case IV in table 4.4.

capability of coping with fluctuations and chemical fluxes) and the possibility of developing thermodynamics, starting from the distribution function. The violation of detailed balance gives information on the "openness" of the system and on the nature of the bistable regimes, which are induced by the external environment; in contrast, it is a free-energy equilibrium when detailed balance holds. This statement can be expressed in a more rigorous form by introducing the vector field generated from the ratio between the generation and recombination terms, by decomposing it into a sum of "conservative" and "rotational" fields (Helmholtz decomposition) and by relating the chemical fluxes to the non-conservative field. The magnitude of deviations from detailed balance influences the form of the stationary distribution at the transition to a bistable regime, which may be driven by the currents. An interesting test for the prediction of the PdPC CME model would be to perform one experiment with the parameter values chosen to satisfy DB and compare it with another set of parameters where DB is not fulfilled. Our results show that the PdPC can operate across these two regions, and that the transition regime can be explained by the role of the currents, that, within a thermodynamic framework, can be interpreted as the effect of an external energy source. A full thermodynamic analysis of

this cycle is beyond the scope of this paper, but we can surmise that this approach might be extended to other cycles in order to quantify if, and how much energy, is required to maintain or create bistability. Another way to extend this analysis would be the generalization to n-step phospho/dephosphorylation cycles, where the stationary distribution will be the product of n independent one-dimensional distributions. In conclusion, our results could be important for a deeper characterization of biochemical signaling cycles that are the molecular basis for complex cellular behaviors implemented as a "switch" between states.

4.6 Mathematical results

The master equation describes the evolution of one-step Markov Processes according to

$$\begin{aligned}
&\frac{\partial \rho}{\partial t}(n_1, n_2, t) = \\
&g_1(n_1 - 1, n_2)\rho(n_1 - 1, n_2, t) - g_1(n_1, n_2)\rho(n_1, n_2, t) \\
&+r_1(n_1 + 1, n_2)\rho(n_1 + 1, n_2, t) - r_1(n_1, n_2)\rho(n_1, n_2, t) \\
&+g_2(n_1, n_2 - 1)\rho(n_1, n_2 - 1, t) - g_2(n_1, n_2)\rho(n_1, n_2, t) \\
&+r_2(n_1, n_2 + 1)\rho(n_1, n_2 + 1, t) - r_2(n_1, n_2)\rho(n_1, n_2, t)
\end{aligned} \tag{4.23}$$

with the boundary conditions for the coefficients

$$\begin{aligned}
&g_1(n, N - n) = g_2(n, N - n) = 0 \quad \text{and} \\
&r_2(n, 0) = r_1(0, n) = 0 \qquad n \in [0, N_T]
\end{aligned} \tag{4.24}$$

so that $n_1 + n_2 \leq N_T$. By introducing the difference operators (4.7), eq. (4.23) can be written in the form of a continuity equation

$$\frac{\partial \rho}{\partial t}(n_1, n_2, t) = -D_1 J_1(n_1, n_2, t) - D_2 J_2(n_1, n_2, t) \tag{4.25}$$

where we introduce the current vector J of components:

$$\begin{aligned}
J_1(n_1, n_2, t) &= g_1(n_1 - 1, n_2)\rho(n_1 - 1, n_2, t) \\
&\quad -r_1(n_1, n_2)\rho(n_1, n_2, t) \\
J_2(n_1, n_2, t) &= g_2(n_1, n_2 - 1)\rho(n_1, n_2 - 1, t) \\
&\quad -r_2(n_1, n_2)\rho(n_1, n_2, t)
\end{aligned} \tag{4.26}$$

The stationary solution $\rho_s(n_1, n_2)$ is characterized by the zero divergence condition for the current (25). Detailed balance holds when the current is zero and

ρ_s satisfies

$$\begin{array}{ll} & r_1(n_1,n_2)\rho_s(n_1-1,n_2) \\ \cdot & \left(\dfrac{\rho_s(n_1,n_2)}{\rho_s(n_1-1,n_2)} - \dfrac{g_1(n_1-1,n_2)}{r_1(n_1,n_2)}\right) = 0 \\ & r_2(n_1,n_2)\rho_s(n_1,n_2-1) \\ \cdot & \left(\dfrac{\rho_s(n_1,n_2)}{\rho_s(n_1,n_2-1)} - \dfrac{g_2(n_1,n_2-1)}{r_2(n_1,n_2)}\right) = 0 \end{array} \tag{4.27}$$

for $0 < n_1$ and $0 < n_2$. The previous equations can be written in the form

$$\begin{aligned} D_1 \ln(\rho_s(n_1-1,n_2)) &= \ln\left(\frac{g_1(n_1-1,n_2)}{r_1(n_1,n_2)}\right) \\ D_2 \ln(\rho_s(n_1,n_2-1)) &= \ln\left(\frac{g_2(n_1,n_2-1)}{r_2(n_1,n_2)}\right) \end{aligned} \tag{4.28}$$

and, if one introduces the the vector field

$$\begin{aligned} a_1(n_1,n_2) &= \frac{g_1(n_1,n_2)}{r_1(n_1+1,n_2)} \\ a_2(n_1,n_2) &= \frac{g_2(n_1,n_2)}{r_2(n_1,n_2+1)} \end{aligned} \tag{4.29}$$

due to the commutative property of the difference operators D_i, detailed balance implies an irrotational character for the vector field $\ln(a(n_1,n_2))$

$$D_2 \ln(a_1(n_1,n_2)) - D_1 \ln(a_2(n_1,n_2)) = 0 \tag{4.30}$$

If we have no singularities in the domain, eq. (4.30) is a sufficient condition for the existence of a potential $V(n_1,n_2)$ (cfr. eq, (4.12)) and the distribution $\rho_s(n_1,n_2)$ can be computed using the recurrence relations

$$\begin{aligned} \rho_s(n_1+1,n_2) &= a_1(n_1,n_2)\rho_s(n_1,n_2) \\ \rho_s(n_1,n_2+1) &= a_2(n_1,n_2)\rho_s(n_1,n_2) \end{aligned} \tag{4.31}$$

Therefore, the components $a_1(n_1,n_2)$ and $a_2(n_1,n_2)$ can also be interpreted as creation operators according to relations (4.31) and detailed balance condition (4.30) is equivalent to the commutativity property for these operators. The stationary distribution can be written in the Maxwell-Boltzmann form (4.14) and the potential $V(n_1,n_2)$ is associated with an "energy function". We finally observe that the critical points of the stationary distribution ρ_s are defined by the condition

$$a_1(n_1,n_2) = a_2(n_1,n_2) = 1 \tag{4.32}$$

For the double phosphorylation cycle (4.1) it is possible to derive explicit expressions for the stationary distribution $\rho_s(n_1,n_2)$ by applying recursively the

relations (4.31) in a specific order: for example, first moving along the n_2 direction and then along n_1, we obtain an expression for $\rho_s(n_1, n_2)$ as a function of $\rho_s(0,0)$

$$\rho_s(n_1,n_2) = \prod_{i=1}^{n_1} \frac{g_1(i-1,n_2)}{r_1(i,n_2)} \prod_{l=1}^{n_2} \frac{g_2(0,l-1)}{r_2(0,l)} \rho_s(0,0) \tag{4.33}$$

A direct substitution of the coefficients (4.3) in the relation (4.33) gives

$$\begin{aligned}\rho_s(n_1,n_2) = \prod_{i=1}^{n_1} & \frac{K_{M4}V_{M2}(N_T-i-n_2+1)}{K_{M2}K_{M4}+K_{M4}(N_T-i-n_2+1)+K_{M2}n_2} \cdot \\ & \frac{K_{M1}K_{M3}+K_{M1}(N_T-i-n_2)+K_{M3}i}{K_{M3}V_{M1}i} \\ & \cdot \prod_{l=1}^{n_2} \frac{K_{M1}V_{M3}(N_T-l+1)}{K_{M1}K_{M3}+K_{M1}(N_T-l+1)} \\ & \frac{K_{M2}K_{M4}+K_{M4}(N_T-l)+K_{M2}l}{K_{M2}V_{M4}l}\rho_s(00)\end{aligned}$$

We can further simplify this expression by using the definition of multinomial coefficients and the rising and falling factorial symbols, defined as $x^{(n)} = x(x+1)(x+2)\cdots(x+n-1) = \frac{(x+n-1)!}{(x-1)!}$ and $x_{(n)} = x(x-1)(x-2)\cdots(x-n+1) = \frac{x!}{(x-n)!}$) respectively.

$$\begin{aligned}&\rho_s(n_1,n_2) = \left(\frac{V_{M2}}{V_{M1}}\right)^{n_1} \left(\frac{V_{M3}}{V_{M4}}\right)^{n_2} \binom{N_T-n_2}{n_1} \binom{N_T}{n_2} \\ &\left(\frac{K_{M3}-K_{M1}}{K_{M3}}\right)^{n_1} \left(\frac{K_{M2}-K_{M4}}{K_{M2}}\right)^{n_2} \cdot \\ &\frac{(K_{M1}(1+n_2-K_{M3}-N_T)-K_{M3})^{(n_1)}}{(K_{M1}-K_{M3})^{(n_1)}(K_{M2}(1+\frac{n_2}{K_{M4}})+N_T-n_2)_{(n_1)}} \cdot \\ &\frac{(K_{M2}(1+K_{M4})+K_{M4}(N_T-1))^{(n_2)}}{(K_{M2}-K_{M4})^{(n_2)}(K_{M3}+N_T)_{(n_2)}}\rho_s(0,0)\end{aligned}$$

Finally, it is interesting to go to a continuous limit that is equivalent to $N_T \to \infty$. First we introduce the population densities $A = n_1/N_T$ and $A_P^P = n_2/N_T$ and use the fact that the generation and recombination rates are invariant by substituting n_1 and n_2 with A and A_P^P. Then we approximate

$$\begin{aligned}D_1V(A,A_P^P) &= V(A+1/N_T,A_P^P) - V(A,A_P^P) \\ &= \frac{1}{N_T}\frac{\partial V}{\partial A}(A+\frac{1}{2N_T},A_P^P) + O(1/N_T^3)\end{aligned}$$

and a similar expression holds for $D_2V(A,A_P^P)$. According to eq. (4.12), the partial derivatives of $V(A,A_P^P)$ are bounded when $N_T \to \infty$ only in the domain where the following approximation holds (diffusion dominated region)

$$\frac{g_i(A,A_P^P)}{r_i(A,A_P^P)} = 1 + O(1/N_T) \qquad i = 1,2 \tag{4.34}$$

and we can estimate

$$\begin{aligned} & \ln\left(\frac{g_1(A,A_P^P)}{r_1(A+1/N_T,A_P^P)}\right) \\ \simeq\ & -2\frac{r_1(A+1/N_T,A_P^P)-g_1(A,A_P^P)}{r_1(A+1/N_T,A_P^P)+g_1(A,A_P^P)}+O(1/N_T^3) \\ & \ln\left(\frac{g_2(A,A_P^P)}{r_2(A,A_P^P+1/N_T)}\right) \\ \simeq\ & -2\frac{r_2(A,A_P^P+1/N_T)-g_2(A,A_P^P)}{r_2(A,A_P^P+1/N_T)+g_2(A,A_P^P)}+O(1/N_T^3) \end{aligned} \tag{4.35}$$

Then we may approximate (we use the convention of leaving out the dependence on A_P^P)

$$\frac{r_1(A+1/N_T)-g_1(A)}{r_1(A+1/N_T)+g_1(A)} = \frac{r_1(A+1/2N_T)-g_1(A+1/2N_T)+1/2N_T\left(\partial r_1/\partial A(A+1/2N_T)+\partial g_1/\partial A(A+1/2N_T)\right)}{r_1(A+1/N_T)+g_1(A)} \tag{4.36}$$

up to an error of order $O(1/N_T^2)$ (a similar expression is obtained for the second equation). Therefore, detailed balance in the continuous limit reads

$$\begin{aligned} \frac{r_1(A,A_P^P)-g_1(A,A_P^P)+1/2N_T\left(\partial r_1/\partial A(A,A_P^P)+\partial g_1/\partial A(A,A_P^P)\right)}{r_1(A+1/2N_T,A_P^P)+g_1(A-1/2N_T,A_P^P)} &= -\frac{1}{2N_T}\frac{\partial V}{\partial A}(A,A_P^P) \\ \frac{r_2(A,A_P^P)-g_2(A,A_P^P)+1/2N_T\left(\partial r_2/\partial A_P^P(A,A_P^P)+\partial g_2/\partial A_P^P(A,A_P^P)\right)}{r_2(A,A_P^P+1/2N_T)+g_2(A,A_P^P-1/2N_T)} &= -\frac{1}{2N_T}\frac{\partial V}{\partial A_P^P}(A,A_P^P) \end{aligned} \tag{4.37}$$

The limit $N_T \to \infty$ turns out to be singular since

$$\begin{aligned} -\frac{\partial V}{\partial A}(A,A_P^P) &= \frac{2N_T(r_1(A,A_P^P)-g_1(A,A_P^P))+\partial r_1/\partial A(A,A_P^P)+\partial g_1/\partial A(A,A_P^P)}{r_1(A,A_P^P)+g_1(A,A_P^P)} \\ -\frac{\partial V}{\partial A_P^P}(A,A_P^P) &= \frac{2N_T(r_2(A,A_P^P)-g_2(A,A_P^P))+\partial r_2/\partial A_P^P(A,A_P^P)+\partial g_2/\partial A_P^P(A,A_P^P)}{r_2(A,A_P^P)+g_2(A,A_P^P)} \end{aligned} \tag{4.38}$$

Hence in the diffusion domain defined by condition (4.34), we recover detailed balance for a Fokker-Planck equation with drift and diffusion coefficients defined as:

$$c_i(A,A_P^P) = 2N_T\left(r_i(A,A_P^P)-g_i(A,A_P^P)\right) \qquad i=1,2$$

and

$$b_i(A,A_P^P) = r_i(A,A_P^P)+g_i(A,A_P^P) \qquad i=1,2$$

In the diffusion region the drift and the diffusion coefficients are of the same order, otherwise $c_i(A,A_P^P) \gg b_i(A,A_P^P)$ when $N_T \gg 1$. As a consequence, the stationary solution of F.P. equation is an approximation of the stationary

distribution of the CME in the diffusion region, but the approximation of the transient state dynamics using the F.P. equation requires further studies due to the singularity of the thermodynamic limit.
In the generic case of eq. (4.23), we represent the r.h.s. of eq. (4.28) as a sum of a rotational and a gradient vector fields

$$\begin{aligned} \ln a_1(n_1,n_2) &= D_1V(n_1,n_2) + D_2H(n_1,n_2) \\ \ln a_2(n_1,n_2) &= D_2V(n_1,n_2) - D_1H(n_1,n_2) \end{aligned} \tag{4.39}$$

Taking into account the condition $n_1 + n_2 \leq N_T - 1$, from the eqs. (4.39) we get the discrete Poisson equations

$$\begin{aligned} (D_1D_1 + D_2D_2)V(n_1,n_2) = \\ D_1 \ln(a_1(n_1,n_2)) + D_2 \ln(a_2(n_1,n_2)) \\ (D_1D_1 + D_2D_2)H(n_1,n_2) = \\ D_2 \ln(a_1(n_1,n_2)) - D_1 \ln(a_2(n_1,n_2)) \end{aligned} \tag{4.40}$$

We remark that the r.h.s. of eqs. (4.40) is defined only if $n_1 + n_2 \leq N - 2$ and corresponds to $N(N-1)/2$ independent equations, whereas we have $(N+2)(N+1)/2$ unknown values $H(n_1,n_2)$. As a consequence from the explicit form of the discrete Poisson operator

$$\begin{aligned} (D_1^2 + D_2^2)H(n_1,n_2) = H(n_1+2,n_2) - 2H(n_1+1,n_2) + \\ H(n_1,n_2) + H(n_1,n_2+2) - 2H(n_1,n_2+1) + H(n_1,n_2) \end{aligned} \tag{4.41}$$

we can set the boundary conditions $H(n,N-n) = H(n,N-1-n) = 0$ and recursively solve the system setting

$$\begin{aligned} 2H(n,N-2-n) = \\ \ln\left(\frac{a_1(n,N-1-n)a_2(n,N-2-n)}{a_1(n,N-2-n)a_2(n+1,N-2-n)}\right)) \end{aligned} \tag{4.42}$$

and successively using the equations

$$\begin{aligned} 2H(n,N-j-n) = -H(n+2,n-j-n) - H(n,N-j-n-2) \\ +2H(n+1,N-j-n) + 2H(n,N-j-n-1) \\ +\ln\frac{a_1(n,N-j-n+1)a_2(n,N-j-n)}{a_1(n,N-j-n)a_2(n+1,N-j-n)} \end{aligned} \tag{4.43}$$

for $N \geq j > 2$. Once $H(n_1,n_2)$ is computed, we define the "potential" $V(n_1,n_2)$ by using eq. (4.39). The recursion relations (4.43) can be written in an exponential form by defining

$$R(n_1,n_2) = \exp(H(n_1,n_2))$$

$$
\begin{aligned}
&R(n, N-j-n)) = \\
&\sqrt{\frac{a_1(n, N-j-n+1)a_2(n, N-j-n)}{a_1(n, N-j-n)a_2(n+1, N-j-n)}} \\
&\cdot \frac{R(n+1, N-j-n)R(n, N-j-n-1)}{\sqrt{R(n+2, n-j-n)R(n, N-j-n-2)}}
\end{aligned}
\tag{4.44}
$$

As a consequence, the recurrence (4.31) reads

$$
\begin{aligned}
\rho_s(n_1+1, n_2) &= a_1(n_1, n_2)\frac{R(n_1, n_2)}{R(n_1, n_2+1)}\rho_s(n_1, n_2) \\
\rho_s(n_1, n_2+1) &= a_2(n_1, n_2)\frac{R(n_1+1, n_2)}{R(n_1, n_2)}\rho_s(n_1, n_2)
\end{aligned}
\tag{4.45}
$$

for all $n_1 + n_2 \leq N-1$.
The current componentss (25) turn out to be proportional to the rotational part of the field (4.39) (i.e. to $H(n_1, n_2)$)[78], so that the current vanishes at the points where condition (4.32) is satisfied. One can prove that the critical points of the stationary distribution are still defined by eq. (4.32). Indeed, if one computes the formal expansion of the generation and recombination rates around a solution of eqs. (31)

$$
\begin{aligned}
g(n) &= 1 + \frac{\partial g}{\partial n}(n^*)\cdot\Delta n + ... \\
r(n) &= 1 + \frac{\partial r}{\partial n}(n^*)\cdot\Delta n + ...
\end{aligned}
\tag{4.46}
$$

(for the sake of simplicity we have normalized the value of the generation and recombination rate to 1 at the critical point) the current components can be approximated by the expressions

$$
\begin{aligned}
&J_1^s(n_1, n_2) \simeq \rho_s(n_1-1, n_2) - \rho_s(n_1, n_2) \\
&+\frac{\partial g_1}{\partial n}(n^*)\cdot\Delta n\rho_s(n_1-1, n_2) - \frac{\partial r_1}{\partial n}(n^*)\cdot\Delta n\rho_s(n_1, n_2) \\
&J_2^s(n_1, n_2) \simeq \rho_s(n_1, n_2-1) - \rho_s(n_1, n_2) \\
&+\frac{\partial g_2}{\partial n}(n^*)\cdot\Delta n\rho_s(n_1, n_2-1) - \frac{\partial r_2}{\partial n}(n^*)\cdot\Delta n\rho_s(n_1, n_2)
\end{aligned}
\tag{4.47}
$$

At the critical point n_* we get

$$
\rho_s(n_1^*-1, n_2^*) = \rho_s(n_1^*, n_2^*-1) = \rho_s(n_1^*, n_2^*) \tag{4.48}
$$

since $J^s(n_1^*, n_2^*) = 0$. The condition (4.48) means the n^* is a critical point for the stationary distribution ρ_s.
When $H(n_1, n_2)$ is small, the detailed balance solution (4.14) is a good approximation of the stationary solution $\rho_s(n_1, n_2)$ and a perturbative approach can

be applied. Let us write the stationary condition (4.6) in the form

$$\begin{aligned} D_1 r_1(n_1,n_2)\rho_s(n_1,n_2)\left(1-a_1(n_1-1,n_2)\frac{\rho_s(n_1-1,n_2)}{\rho_s(n_1,n_2)}\right) &+ \\ D_2 r_2(n_1,n_2)\rho_s(n_1,n_2)\left(1-a_2(n_1,n_2-1)\frac{\rho_s(n_1,n_2-1)}{\rho_s(n_1,n_2)}\right) &= 0 \end{aligned} \tag{4.49}$$

By using the definitions (4.39), we assume that the rotational field is associated with a potential $\epsilon H(n_1,n_2)$, with $\epsilon \ll 1$ perturbation parameter and we write the stationary solution in the form

$$\rho_s(n_1,n_2) = C\exp(V(n_1,n_2)+\epsilon V_1(n_1,n_2)) \tag{4.50}$$

From a direct calculation we get

$$\begin{aligned} D_1 r_1(n_1,n_2)e^{V(n_1,n_2)}\left(1-\exp(\epsilon(D_2H(n_1-1,n_2)-D_1V_1(n_1-1,n_2)))\right) \\ +D_2 r_2(n_1,n_2)e^{V(n_1,n_2)}\left(1-\exp(-(\epsilon D_1H(n_1,n_2-1)+D_2V_1(n_1,n_2-1))\right) \simeq \\ \epsilon D_1 r_1(n_1,n_2)e^{V(n_1,n_2)}\left(D_1V_1(n_1-1,n_2)-D_2H(n_1-1,n_2)\right) \\ +\epsilon D_2 r_2(n_1,n_2)e^{V(n_1,n_2)}\left(D_2V_1(n_1,n_2-1)+D_1H(n_1,n_2-1)\right)=0 \end{aligned} \tag{4.51}$$

for all the values $n_1+n_2 \leq N-1$ and $n_i \geq 1$. The correction potential $V_1(n_1,n_2)$ has to be computed from the previous equation, and it enters in the definition of the stationary currents.

Chapter 5

Protein concentration during cellular senescence

In this chapter we will give the outlines of a work aimed to model the total number of proteins in the cell nucleus and their concentration as the cells undergo cellular senescence. We will see that while the cells become senescent they increase in size and the total amount of nuclear proteins increases. One surprising factor is that also their density increases, meaning that the nucleus is being clogged with proteins.

A basic model for protein production will be taken in consideration and its prediction will be confronted with the experimental data provided by Dott. Marco De Cecco during its PhD, under the guidance of Prof. John Sedivy at Molecular Biology Department at Brown University. A model for the whole cell division will be discussed, with its prevision for a new set of observation of the same phenomena.

5.1 Fluorescence microscopy

Fluorescence microscopy is a technique that in the past ten years has revolutionized the field of biology, combined with the introduction of the GFP (Green Fluorescent Protein). The problem with standard microscopy is that observing a biological sample under transmitted white light gives a very poor response: most of the fundamental structures of the cell have more or less the same optical refraction index of the water, so they appear uniformly transparent with just few little local variations; in addiction, this microscopy form suffers from the *Abbe diffraction limit*, that affirms that is not possible to distinguish two point objects that a closer than the distance limit $d = \frac{\lambda}{2n\sin(\theta)}$, where λ is the wavelength of light, n is the refraction index of the observed medium and θ is half of the microscope angular aperture.

With modern equipment the diffraction limit can reach roughly half of the wavelength of the incident light which, for a green light (the best for human perception), is in the order of 250 nm, while most cellular structures have a size at least one order of magnitude less. This limit can be stretched using UV microscopy, which on the other end has a lower contrast factor for biological tissues and damages the sample rather quickly.

In the fluorescence microscopy one does not observe the transmitted light from the sample but rather the emission from specific molecules called fluorophores, which get excited by a fine tuned light source and emit back with a different wavelength, so that one can discriminate between the source light and the fluorescence light with a dichroic mirror.

Several fluorophores exist that permit to cover almost the whole luminous spectrum, and can be used for various purposes. Fluorescent proteins like the GFP can be put in sequence to a target gene to monitor its expression. Similar to the GFP (also called FITC sometimes) we can find the DAPI, which is a blue fluorescent protein which bind to the nucleus and is commonly used to locate and identify it, and the TRITC, a red fluorescent one. The combination of the previous three is common due to the very little superposition in absorbing and emitting spectra of the three proteins, which allows to perform quite easily a single experiment with all these fluorophores. Also, being non-toxic to the cell, they allow to follow biological processes with as little interference as possible.

For a general review on this topic, see: [62, 73, 84, 101, 46, 113, 10, 17, 111, 76, 110, 93, 18, 32]

5.1.1 Experimental methods

Usage of this fluorophores in live cell imaging, combined with an incubator, allows prolonged periods of time of observation. We are utilizing a Nikon Ti Eclipse microscope, with a completely functional incubating unit that allows continuous observation up to three days, fully automated with a computer interface. This allows not only for a continuous experiment under constant monitoring, but even real time numerical analysis to guide the observation.

However, the data shown here has been produced by Dott. Marco De Cecco during its PhD, under the guidance of Prof. John Sedivy, of the Molecular Biology Department of the Brown University, RI. The cells are a standard line of murine fibroblast cells, cultivated from a young specimen. Every two duplication steps (every few days near convergence on the plate) the plate was transferred and samples were taken for analysis. This is referred as one passage. The sample analyzed afterward are taken from passage 3 (still young cells) and from 9,10,11,12 and 13 (late cellular senescence). These cells, due to the cellular senescence, were significantly bigger and slower in the reproduction, near to a complete stop.

Each sample was then prepared with an immunofluorescence treatment, where the biological samples are killed and fixated by chemical means, then a specific developed fluorescent antibody is added to the mixture, binding it to the target. In this experiment the cells and nucleus membranes were made permeable and then all the nuclear proteins were denatured with a high temperature bath, rendering them susceptible to a generic antibody staining that binds to the hydrophobic core of the proteins. This permits the direct observation of the quantity of the nuclear proteins amount. A second antibody marked the chromatin, to ease the separation of the nuclear matter from the rest. These fixated plates were then observed with a confocal fluorescence microscope, a high precision microscope that acquires the image with a laser scansion to obtain great precision on the observation of the fluorescence, allowing to select only the focal plane of interest removing all the light from other sources.

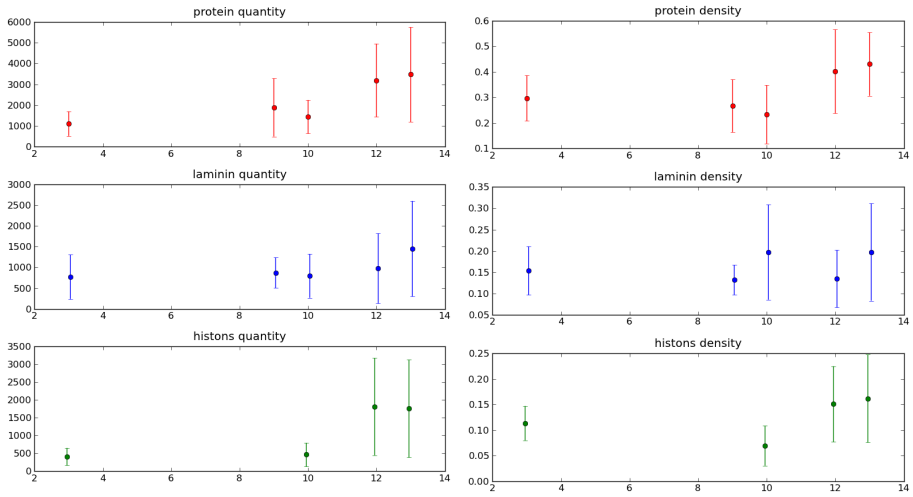


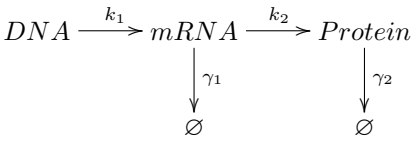
Figure 5.1: time-dependences of total quantity and density for the nuclear protein, laminin-A and histones

The resulting images were then numerically treated to select only the protein signal from the nucleus (thanks to the nuclear basal staining), calculating the size of the nucleus (in pixels) and the total fluorescence from the nuclear proteins. Others staining has been done to select the laminin-A, a protein that binds to the nucleus membrane and is thought to be unchanged during the cellular senescence, and the nuclear histones which, being linked to the amount of DNA present in the cell, are almost assured to be constant during the stationary phase of the cell cycle.

In the following image we can see a summary of the results of the analysis of the first set of data: the lines represented are the change in time of the total quantity and density of the three measured quantities: nuclear proteins, laminin-A and histones. Even with a great variance we can see how both the quantity and density of proteins increase with time, while the other two remains approximately constants.

5.2 Mechanistic model

Modeling protein production is not a trivial task at first glance, due to the complexity of the underlying phenomena. Recent experimental[12, 103, 116] and theoretical advances give a different hint: even if the protein production process is a rather complex one, it is possible to approximate it with the most basilar model, where one schematize just the mRNA and protein production, neglecting everything else. In this kind of model the protein production system is represented by the following reaction scheme:



Where the DNA quantity is assumed being constant and the production of mRNA scarce. The reaction constants k_1 and k_2 represent the production rate of mRNA and protein respectively, and the γ_1 and γ_2 their degradation rate. For each mRNA molecule, several proteins are produced, generating the so-called protein production burst. This burst has been experimentally observed and the distribution was compatible with an exponential one, as expected from the model above in the limit of short-lived mRNA.

5.2.1 A basic model for protein concentration

The above model has been solved in the concentration limit by Friedman et Al[33], which started from a generic monodimensional continuous master equation for the concentration of protein $p(x)$ with $x = n/V$, where V is the volume of the cell and n the protein number. In this equation the first term represents the dilution of the protein due to both the degradation of the protein and the augmentation of cell volume due to the cellular duplication. γ_2 can thus be seen as a composition of the this two components,$\gamma_2 = (\ln 2/T + \ln 2/T_1)$, where T the protein half life and T_1 the cell cycle duration (cell volume doubling time). The term $w(x, x')$ is a transition constant as discussed in Sec 1.2.1 about Markov process.

$$\frac{\partial p(x)}{\partial t} = \frac{\partial}{\partial x}[\gamma_2 x p(x)] + k_1 \int_0^x dx' w(x, x') p(x')$$

This equation can be solved in the stationary limit, where $\frac{\partial p(x)}{\partial t} = 0$ and thus:

$$-\frac{\partial}{\partial x}[xp(x)] = aw * p(x)$$

where $a = k_1/\gamma_2$ represents the medium number of production burst per cycle.

To solve this equation we can work with its Laplace transform, which is

$$s\frac{\partial \hat{p(s)}}{\partial s} = a\hat{w}\hat{p(s)} \tag{5.1}$$

where $\hat{p(s)}$ and $\hat{w}$ are the Laplace transforms of the probability distribution and the burst distribution respectively.

Under the hypothesis of burst dimension independent from the present concentration and described by a distribution $\nu(x - x')$ we obtain $w(x, x') = w(x - x') = \nu(x - x') - \delta(x - x')$, where the delta guarantees the probability flux conservation. Remembering the exponential approximation for the burst size, we can write $\nu(x) = (1/b)exp(-x/b)$, where $b = k_2/\gamma_1$ is the average number of proteins produced on each burst. Its Laplace transform is $\hat{w(s)} = -s/[s + (1/b)]$.

Inserting this formula into Eq 5.1 we have a defined equation that can be solved to obtain the transform of the stationary distribution:

$$\hat{p(s)} = [s + (1/b)]^{-a}$$

This can be recognized as the transform conjugate of a gamma distribution with shape a and scale b:

$$p(x) = \frac{x^{a-1}e^{-x/b}}{b^a\Gamma(a)}$$

The interesting property of the Gamma distribution is that, like the Gaussian, is closed under sum as long as the b parameter is the same. Being b an index of the performance of the protein translation mechanism by the ribosomes and not of the protein *per se*, it is plausible that it should be shared among all the different kind of proteins present in the cell, allowing to write a gamma for the total protein quantity in the cell:

$$\sum_i \Gamma(a_i, b) = \Gamma(\sum_i a_i, b)$$

.

An other interesting property of this distribution is that has a Bayesian conjugate prior quite easy to work with, allowing for a good parameter estimation based on the likelihood method alone. Given n observation x_i, the conjugate prior is:

$$\Gamma(a, b|p, q, r) = \frac{1}{Z}\frac{p^{a-1}e^{-b^{-1}}}{\Gamma(b)^r b^{ar}}$$

Where Z is a normalization constant (that can be usually neglected for practical purposes). The parameters p, q, r are defined as follow:

$$\begin{aligned} p' &= p\textstyle\prod_i x_i \\ q' &= q + \textstyle\sum_i x_i \\ r' &= r + n \end{aligned}$$

The gamma distribution can also be seen as the continuous limit of a negative binomial distribution, as demonstrated by Paulsson et Al[82]. The negative binomial is a commonly observed distribution among the solution of the master equation. Two examples of this are the following two master equations defined by their g_n and r_n terms. The first one represents a population in an open environment with birth and death proportional to the population and a constant immigration term, and converges to a negative binomial a long as $\gamma > \beta$, diverges to infinity otherwise:

$$\begin{aligned} g_n &= \alpha + \beta n \\ r_n &= \gamma n. \end{aligned} \tag{5.2}$$

The second one, that is completely equivalent as stationary distribution, represents a chemical environment where the chemical specie in consideration is constantly created, but decreased with a nonlinear term that can be interpreted as a Michaelis-Menten kinetic reaction. A different possible interpretation will be used in the next section, where it will represent the actual density over a

volume that is filled with the substance of interest and an independent, constant, inert background. A convergence condition is required as before, $\gamma > \alpha$, otherwise a divergence to infinity happens:

$$\begin{aligned} g_n &= \alpha \\ r_n &= \gamma \frac{n}{\theta + n}. \end{aligned} \tag{5.3}$$

5.2.2 The model used for the cell growth

We worked on a basic model to represent the quantity of visible proteins in the cell nucleus. The protein of interest is called Pv_a, which stands for visible protein in the cell A. The cell index A is used only for modeling the cell division, and is a mute index otherwise. We have then the protein Pd_a, which stands for denatured protein, which is the protein that is not fluorescent due to degradation or simply because it is of a different specie. These two species have two specular variables, Pv_b and Pd_b which will represent the quantity of protein in the daughter cell during the division. A third quantity, taken as constant to represent the bulk materials of the nucleus, is called DNA, being this non fluorescent part mainly composed of chromatin. We define the concentration of the two kind of proteins as $C_v = \frac{Pv_a}{Pv_a+Pd_a+DNA}$ and $C_d = \frac{Pd_a}{Pv_a+Pd_a+DNA}$.

Both kind of proteins are driven by a immigration-reproduction-death as described on Eq 5.2. This is a plausible assumption, remembering that the production of proteins can be enhanced even by the protein itself and, being Pv_a not a single specie but rather a mixture of cellular proteins, the basic hypothesis is a linear self replication and a linear degradation term. A small quantity of Pv_a can decade into Pd_a, and we could safely ignore the reverse transition (it is common that a fluorescent protein stop being so after degradation, but the converse is quite implausible). The status of the cell, growing or dividing, is driven by the *growth* and *divide* factors, which can be 1 or 0 and switch back and forth given the right condition.

A set of reactions is bounded to happen only during the growth phase, like the production and degradation of the proteins. During the division regime, which is very short-lived in respect to the stationary phase of the cell cycle, we have a balanced interconversion between Pv_a and Pv_b to represent the exchange of materials between the mother and daughter cell. An identical reaction happens to Pd_a, which balances with Pd_b. The net effect of this reactions is that after the division the cell is roughly split in half with a certain dishomogeneity between the cells. These reactions are very fast and reach the equilibrium in a time scale much shorter than the others, so that the equilibrium is assured. No reaction has been implemented for the DNA because it will be conserved before and after the duplication and any variation during the duplication phase can be ignored.

The last variable is called *clock*, and represents the cellular "off" switch: each cell starts with a certain amount of tolerance, which is removed after each duplication. Eventually the loss of this clock molecule slows down the reproduction process. This term represents the core idea of an active aging effect for the cell, some sort of programmed "shutdown". We are working even on a model with a passive form of aging due to a clogging effect of the damaged proteins, but it is still in the first phase of development.

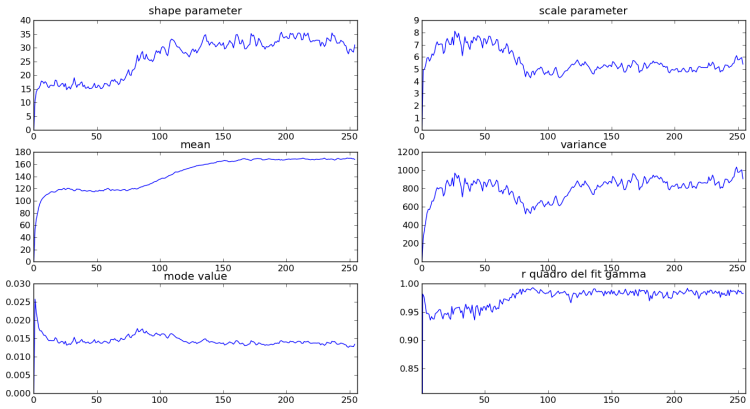


Figure 5.2: model prediction for the resulting gamma distribution

$$
\begin{aligned}
growth + clock &\to 10^2 \tfrac{C_v^2 clock^2}{k+clock^2} \to & divide \\
divide &\to 10^0 \to & growth \\
Pv_a + growth &\to 10^3 C_v^2 \to & 2Pv_a + growth \\
Pv_a + growth &\to 10^- 3 Pv_a \to & growth \\
growth &\to 10^0 \to & Pv_a + growth \\
Pv_a + growth &\to 10^0 Pv_a \to & Pd_a + growth \\
growth &\to 10^2 \to & Pd_a + growth \\
Pd_a + growth &\to 10^0 C_v C_d \to & growth \\
Pd_a + growth &\to 10^1 Pd_a \to & growth \\
Pv_b + growth &\to\gg 10^1 Pv_b \to & growth \\
Pd_b + growth &\to\gg 10^1 Pd_b \to & growth \\
Pv_b + divide &\to\gg 10^1 Pv_b \to & Pv_a + divide \\
Pv_a + divide &\to\gg 10^1 Pv_a \to & Pv_b + divide \\
Pd_b + divide &\to\gg 10^1 Pd_b \to & Pd_a + divide \\
Pd_a + divide &\to\gg 10^1 Pd_a \to & Pd_b + divide
\end{aligned}
$$

Due to the great number of non-specified parameters, this model is not yet strong enough to make discriminating predictions, but in most of the cases it predicts that the distribution of the protein density will be a Gamma distribution both in the young and old cells, and maintains this distribution even during the aging process, where a group of still active cells and inactive one coexist. An example of this can be seen in Fig. 5.2, that is the fit of the time dependent solution of the model for a generic set of parameters. It shows both parameters of the gamma distribution with its mean, variance and maximum peak value of the distribution. We can see that the mean and variance are at a glance compatible with the experimentally observed. On the right-low panel we can see the evaluated r^2 between the observed distribution and the Gamma distribution fitted with Maximum Likelihood Method. We can see that the r^2 is high on each phase of the aging, as it will be confirmed by the observations.

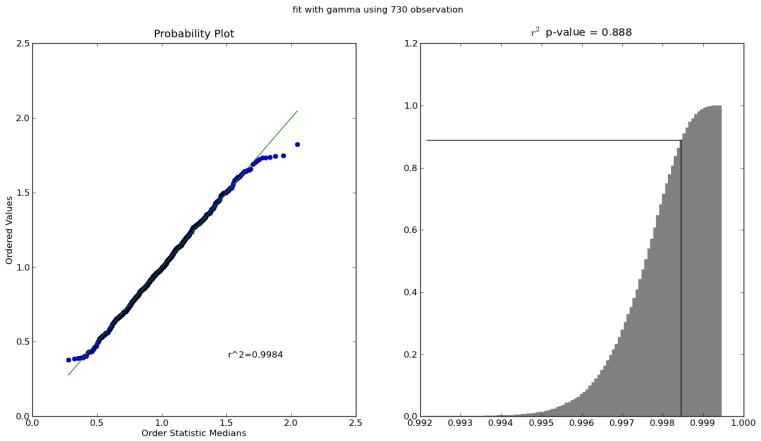


Figure 5.3: example of the resulting quantile-quantile plot and the relative distribution of the expected r^2 values

5.3 Fit of the obtained distributions

The fits are performed with a standard maximum likelihood estimation on the raw data. The distributions studied are the Gaussian, the Lognormal and the Gamma distribution. The Gaussian distribution has been chosen as a control distribution, to attest that the data do not conform to it. The Lognormal has been chosen being a common choice for the cellular size[21]. To avoid spurious defects from the image processing technique used to evaluate the protein quantity, like the fusion of two proximal cells or the division of a misshaped nucleus, the dataset was trimmed by 1% on each extreme.

The goodness of fit has been evaluated with a Montecarlo method based on the r^2 value obtained from the quantile-quantile plot: given the fit parameters the distribution of possible r^2 observable given the null hypothesis has been evaluated, operating on the simulated dataset with the same process of trimming of the original one. The result of these tests for the three distributions in the various passages, for the total protein quantity, density and nucleus size, are shown below.

Following the three tables with the fits results there are six images which depict the distribution of the protein total quantity, density and nuclear size for the first serie of data (passage number 3) and for the last one (passage number 13). For each distribution the partial distribution of each of the four separate experiment is shown below, with the small caption of the quantile quantile plot and respective r^2 evaluation.

GAMMA DISTRIBUTION:							
observation		protein density		nuclei dimension		protein quantity	
passage	N	r^2	pval	r^2	pval	r^2	pval
03	730	0.997	0.371	0.998	0.961	0.996	0.75
09	251	0.982	0.001	0.974	0.001	0.994	0.403
10	193	0.962	0.001	0.995	0.27	0.982	0.004
11	101	0.996	0.963	0.992	0.436	0.993	0.766
12	180	0.991	0.075	0.992	0.155	0.996	0.727
13	672	0.998	0.673	0.999	0.996	0.994	0.309

LOGNORMAL DISTRIBUTION:							
observation		protein density		nuclei dimension		protein quantity	
passage	N	r^2	pval	r^2	pval	r^2	pval
03	730	0.998	0.871	0.998	0.871	0.995	0.884
09	251	0.99	0.197	0.99	0.197	0.998	1.0
10	193	0.981	0.084	0.981	0.084	0.968	0.001
11	101	0.993	0.733	0.993	0.733	0.976	0.248
12	180	0.992	0.201	0.992	0.201	0.99	0.269
13	672	0.999	0.978	0.999	0.978	0.996	0.983

GAUSSIAN DISTRIBUTION:							
observation		protein density		nuclei dimension		protein quantity	
passage	N	r^2	pval	r^2	pval	r^2	pval
03	730	0.996	0.018	0.954	0.001	0.946	0.003
09	251	0.931	0.002	0.933	0.001	0.923	0.001
10	193	0.866	0.001	0.986	0.001	0.953	0.001
11	101	0.97	0.002	0.967	0.001	0.919	0.001
12	180	0.955	0.001	0.979	0.001	0.954	0.001
13	672	0.995	0.004	0.969	0.002	0.927	0.001

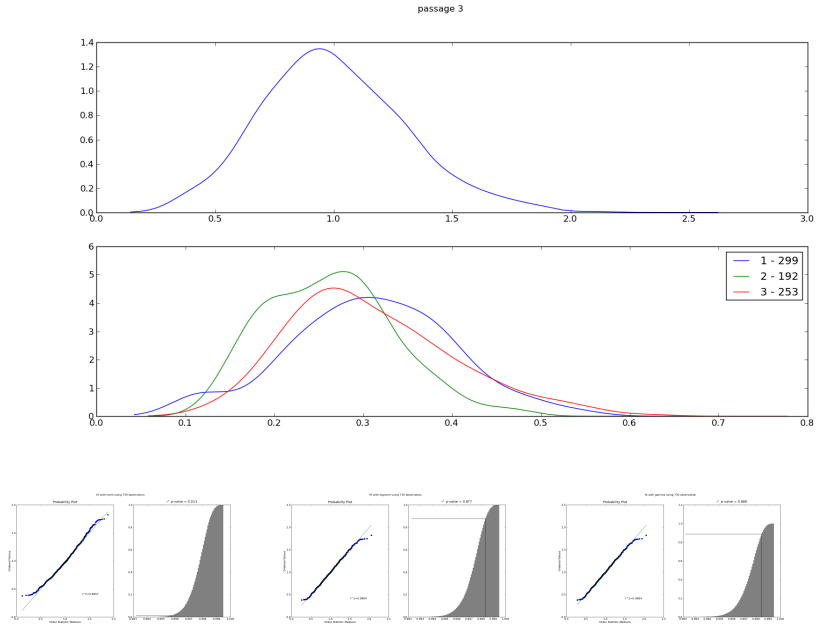


Figure 5.4: protein density of the third passage group and fit with Gaussian, lognormal and Gamma (from left to right)

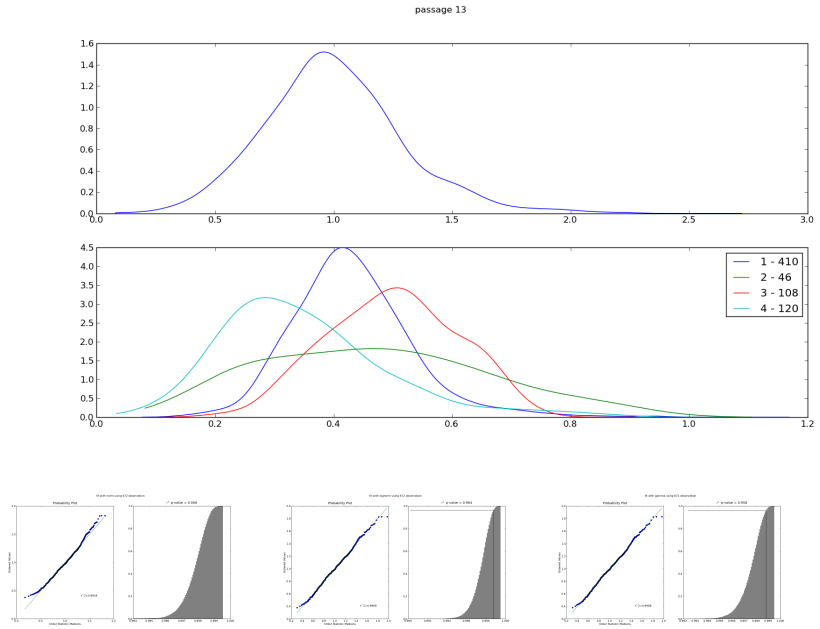


Figure 5.5: protein density of the thirteenth passage group and fit with Gaussian, lognormal and Gamma (from left to right)

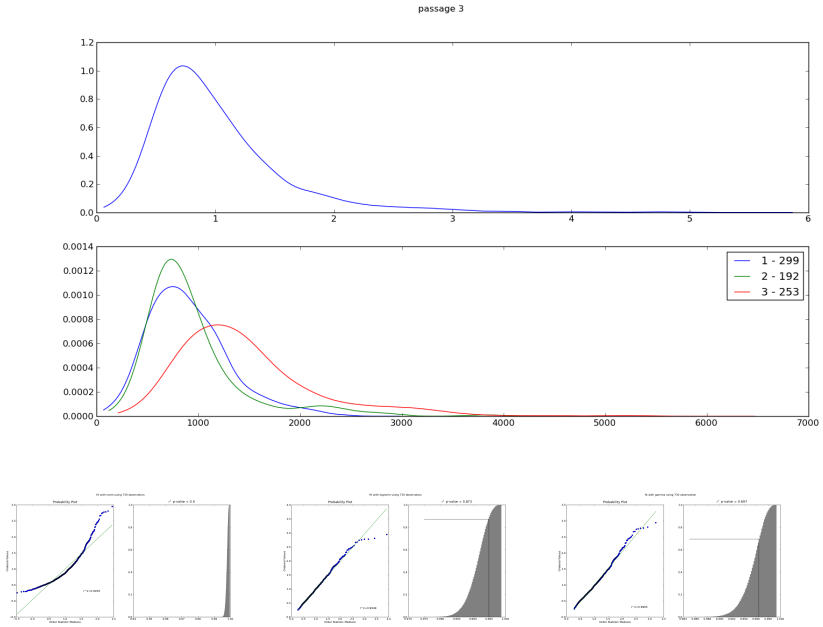


Figure 5.6: protein quantity of the third passage group and fit with Gaussian, lognormal and Gamma (from left to right)

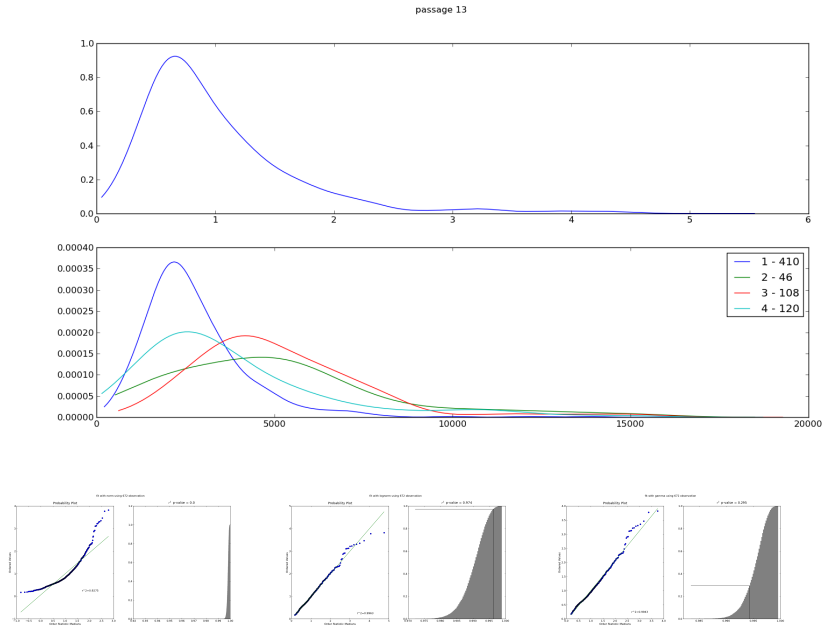


Figure 5.7: protein quantity of the thirteenth passage group and fit with Gaussian, lognormal and Gamma (from left to right)

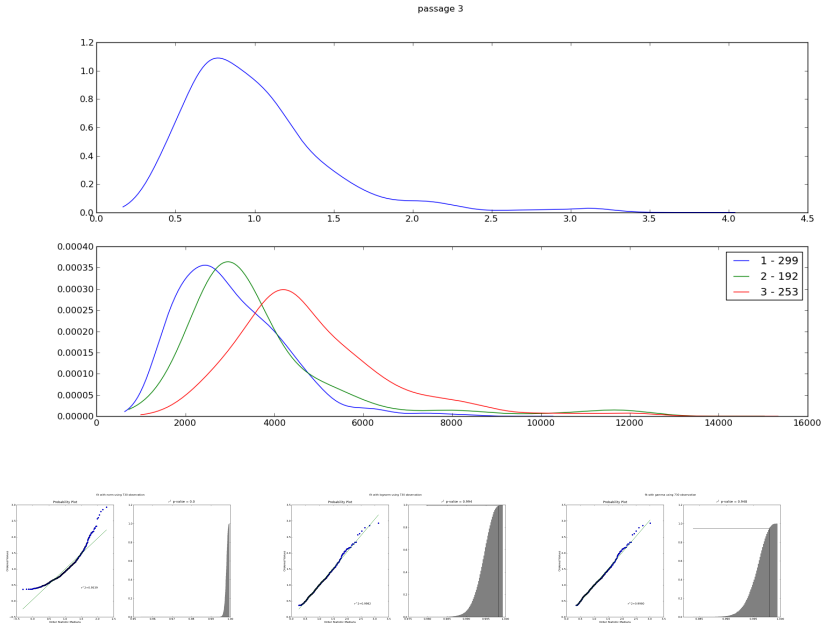


Figure 5.8: protein dimension of the third passage group and fit with Gaussian, lognormal and Gamma (from left to right)

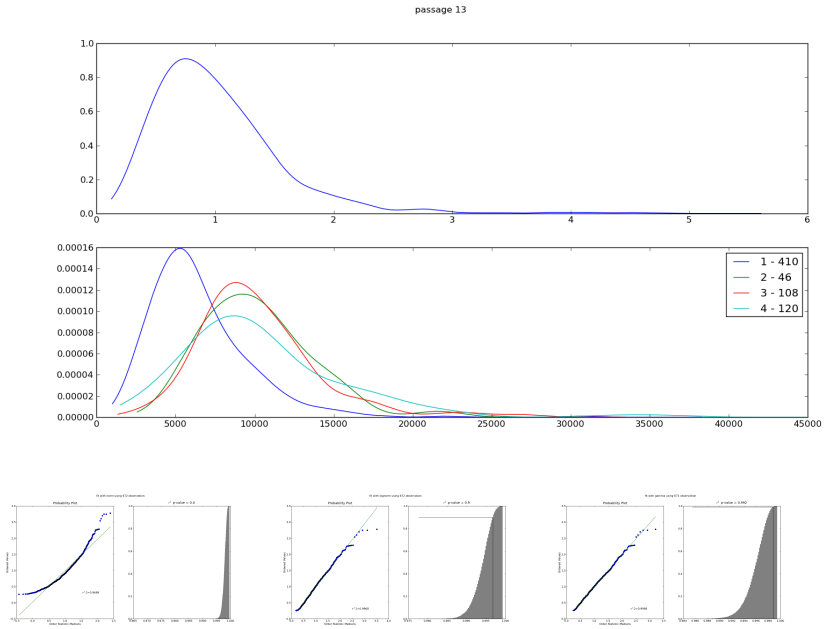


Figure 5.9: protein dimension of the thirteenth passage group and fit with Gaussian, lognormal and Gamma (from left to right)

5.3.1 Parameters evaluation for the data

Here we present the parameters of the gamma distributions resulting from the fit performed in the previous section. In each plot four quantities will be shown versus the number of passage from which it was observed. The first two are the shape and scale parameters of the distribution, here called α and β, that in the case of the protein quantity are linked to the number of burst per cell cycle and the number of proteins for each burst. Then starting from these two parameters we evaluate the mean of the distribution as $\alpha\beta$ and its variance as $\alpha\beta^2$. For the mean will be shown the standard error due to the sampling.

First we have the estimated nucleus size utilizing every set of data (being the only common measure), that shows how the cells grow steadily until they reach the senescence at the thirteenth passage, where they stop increasing their size. Being the quantity of DNA constant in the senescence process, we expect an increase of the quantity of the proteins and an increase also in their density. This trend can be clearly seen in the data means.

The laminin-A should grow alongside the nucleus border length, so their linear density, estimated as the quantity of laminin-A divided by the square root of the area of the nucleus, is shown with their density and total quantity. This linear density should be constant, and the data support this theory. The deviations from the constant value are probably due to the rough estimation of the nucleus border length.

The histones are expected to be constant, being the structural skeleton of the chromatin, whose quantity should not change with the senescence. Instead, they show a very sharp increase in the total quantity in the last two passages, especially in the total amount. It is not clear if this increase is an actual augmentation of the number of molecules or it is due to a form of un-packing of the DNA strands, leaving more histones exposed to the immunofluorescence staining.

Several studies have shown that a reduction in the number of histones is a normal sign of cellular aging, while its increase has been related to the transformation into neoplastic cells[31, 94, 4, 100]. Murine cells reach the normal stasis after the senescence, but are known to transform often into immortalized lines (tumoral cells), so it is possible a relationship between the increase of the histones and the abnormal fate of these cells.

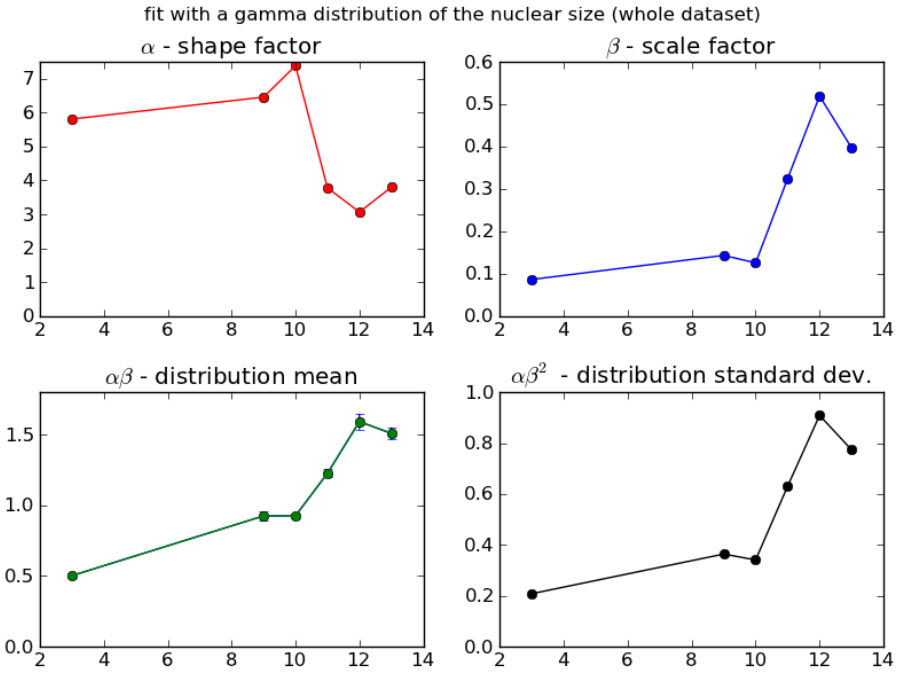


Figure 5.10: nucleus size across all experiments

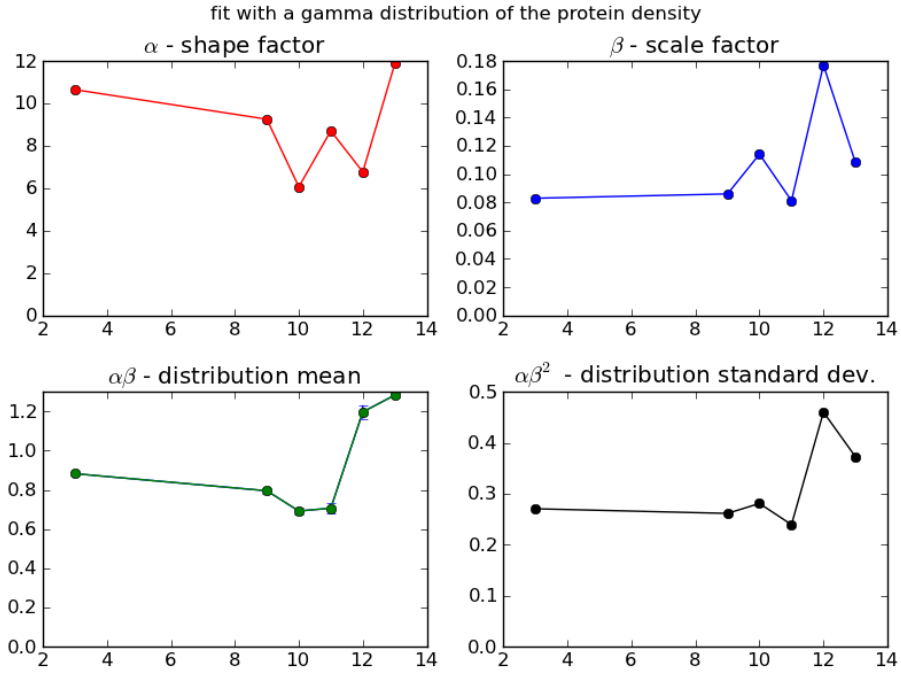


Figure 5.11: density of the proteins during passages

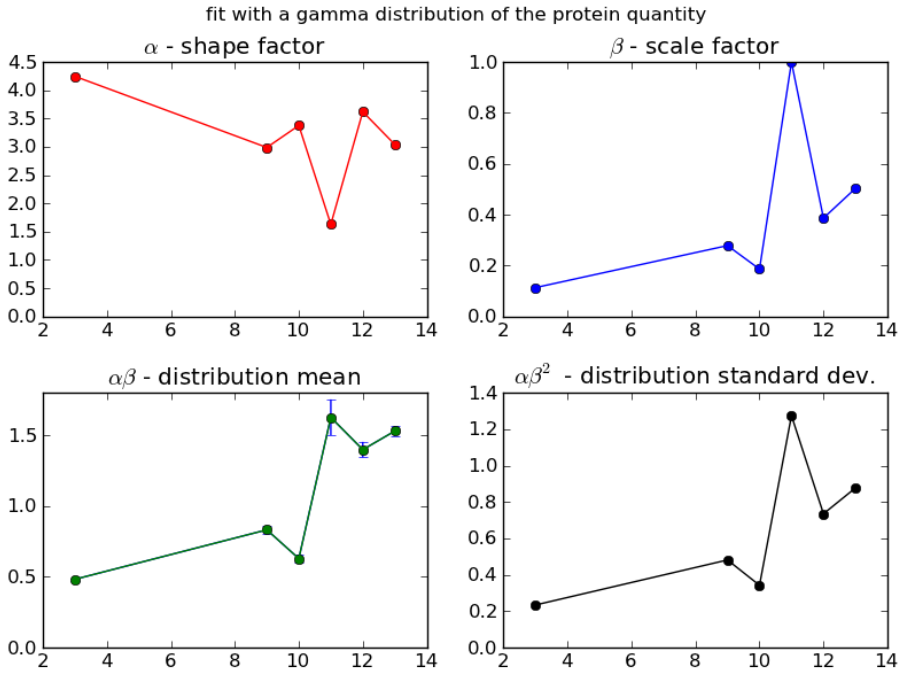


Figure 5.12: total quantity of proteins

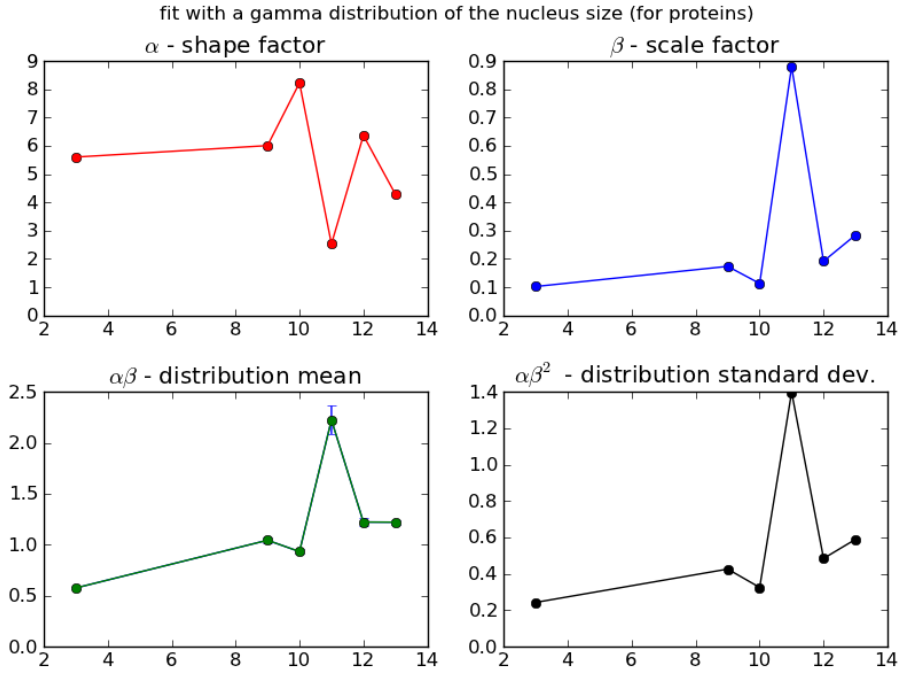


Figure 5.13: nucleus size on the histones experiment

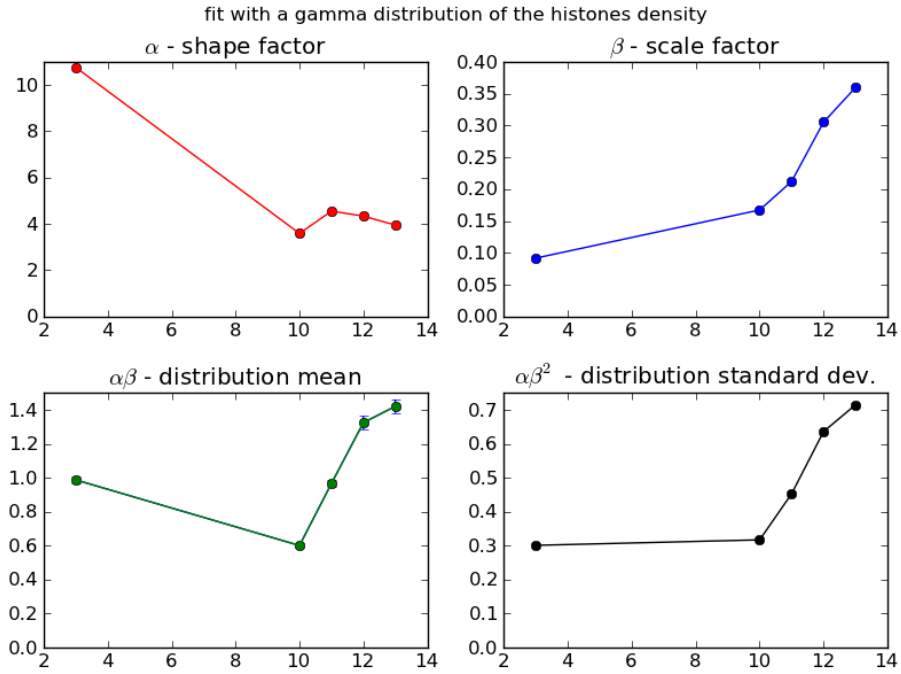


Figure 5.14: density of the histones during passages

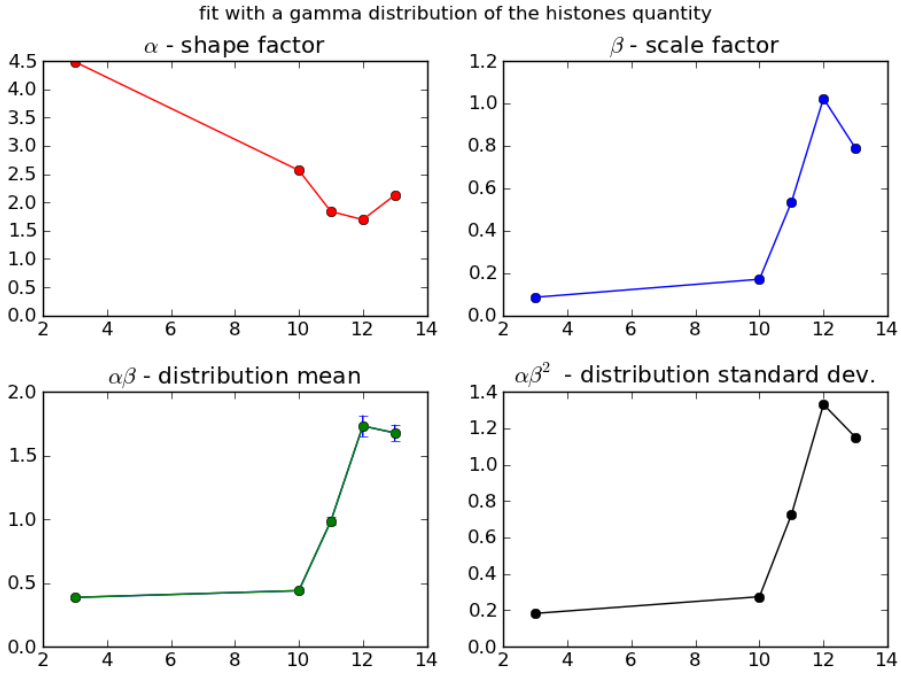


Figure 5.15: total quantity of histones

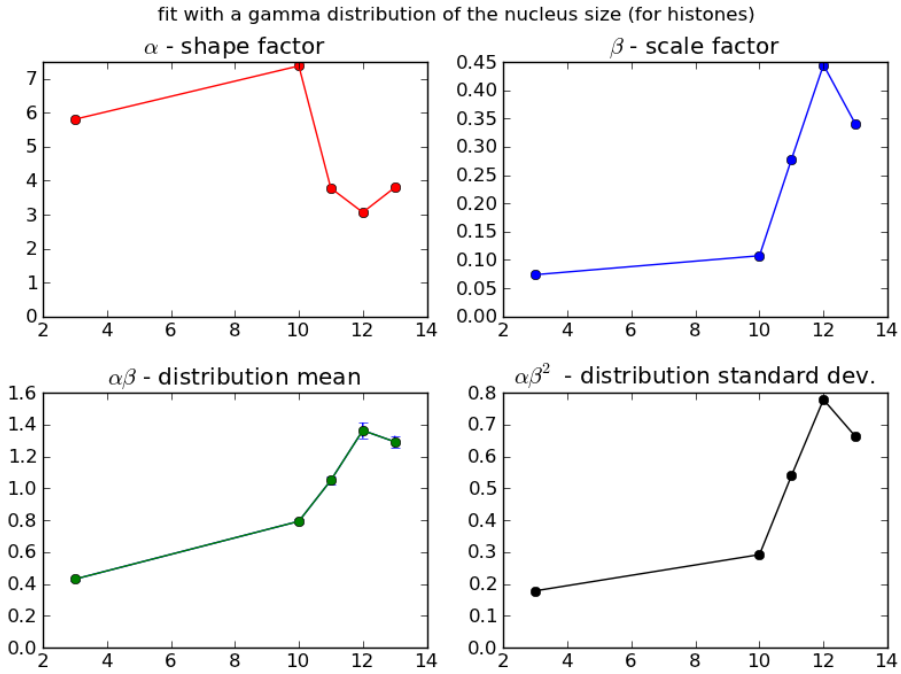


Figure 5.16: nucleus size on the histones experiment

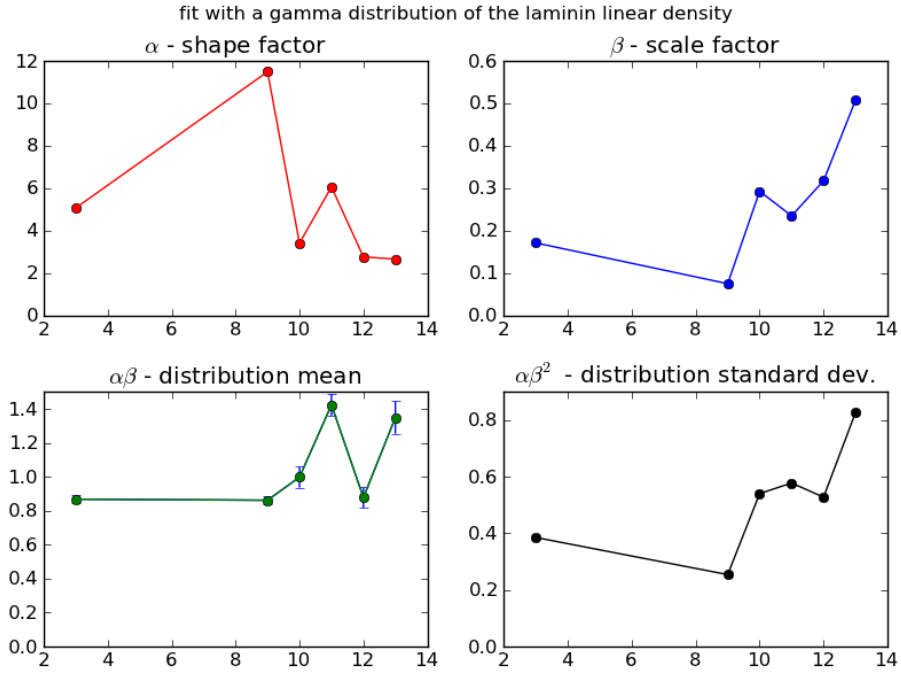


Figure 5.17: linear density of the laminin-A during passages

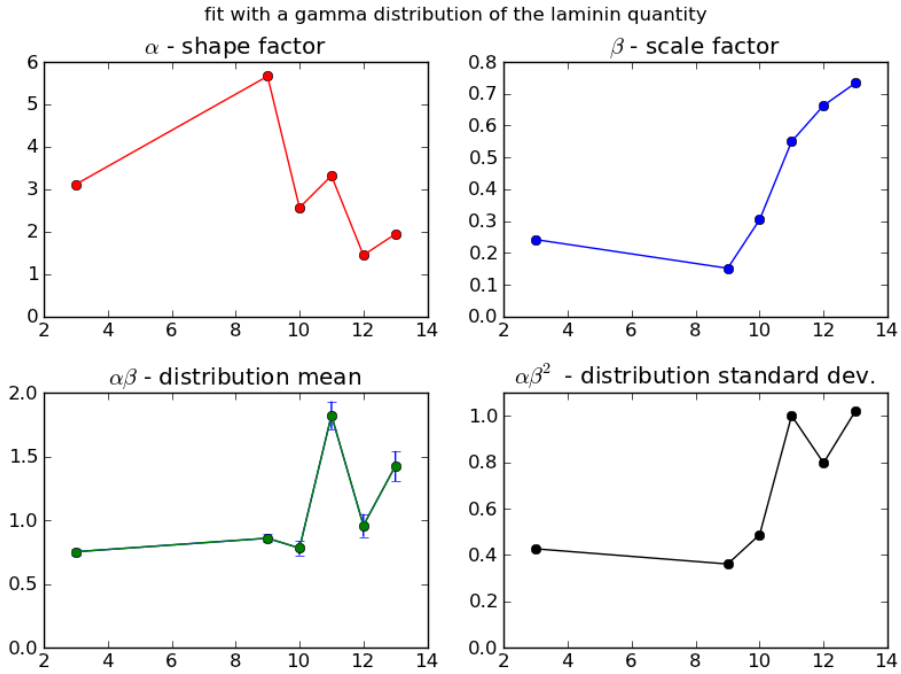


Figure 5.18: total quantity of laminin-A

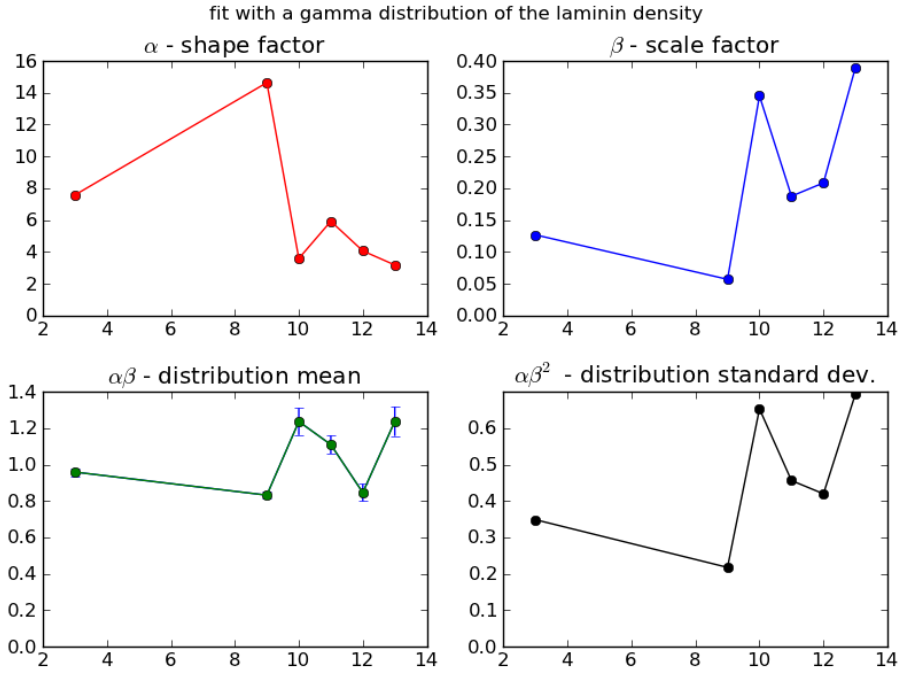


Figure 5.19: density of the laminin-A during passages

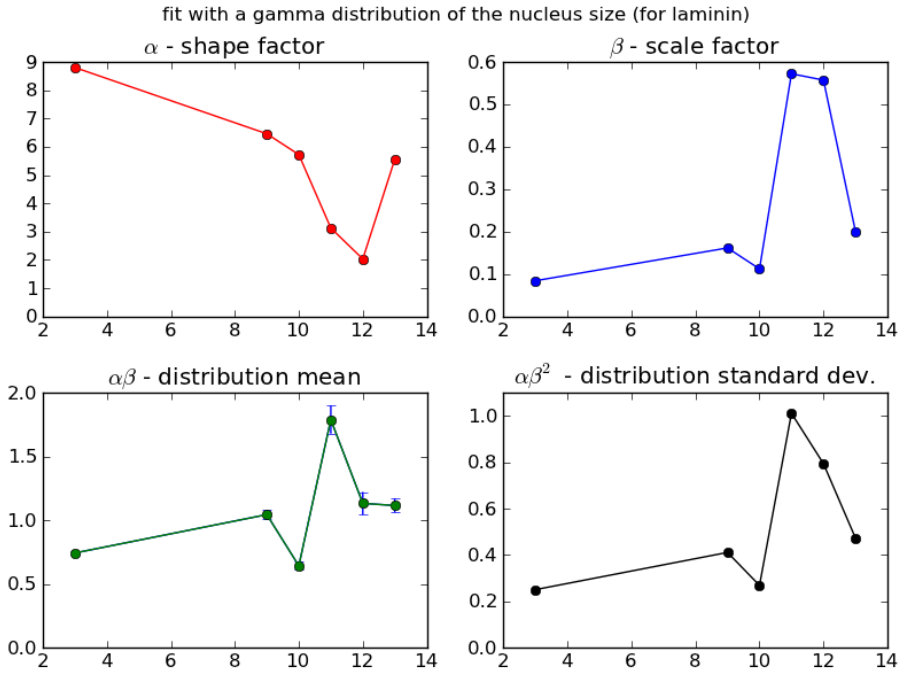


Figure 5.20: nucleus size on the laminin-A experiment

Chapter 6

Conclusions

The aim of the present thesis was to show how the master equation formalism can successfully enter in the description of biological phenomena. After a first introduction to the theoretical fundamentals of this technique, three works have been presented.

In the first one we developed a simplified model for a genetic toggle switch of cardinal importance in molecular biology and for cancer development, and showed how this model react to the presence of the intrinsic noise in the system. This research confirmed that for nonlinear system like the genetic regulations mechanisms, the introduction of the intrinsic noise, due to the discreteness and finiteness of the system, cannot be neglected. The model is interesting on its own, leading to the identification of features that we hope will be further characterized, also experimentally, as possible links with the self sustained replication of tumoral cells.

In the second one we worked on the conceptual structure of an enzymatic futile cycle, an ubiquitous phenomena in both bacterial and eukaryotic cells. This cycle, whose usefulness was not clear, is without any interesting property when analyzed from a deterministic point of view, but it is known to be sensitive to the presence of external noise through a zeroth-order sensitivity which allows the system to amplify and rectify the incoming noise, becoming a bistable switch and so capable of holding a bit of information for the cellular memory. We showed how this sensitivity works and which parameters of this external noise allow the transition from monostable to bistable, linking it to the variance of the balance between the two concurrent enzymes. Being these futile cycles often arranged in a cascade which amplifies the noise variance at each step, this can be one piece of the puzzle to explain the exceptional frequency of this biochemical pattern.

In the third work we studied two coupled futile cycles driven by the same couple of enzymes (one kinase one phosphatase), i.e a double modification of the substrate. This reaction is quite different from the single futile cycle, allowing the existence of a double stable state even when treated under deterministic hypothesis. Double, triple and so-on modification cycles (phosphorylation or methylation are the most common ones) are a *leitmotif* in the cellular regulation pathways. The methylation is responsible for the chromatin's folding and unfolding, indirectly silencing or expressing whole part of the genome, and it is thought to be one of the main driver under the non-genetic inheritance,

the epigenetic phenomena, which transmits information on the genome state of activation between the parents and the children without any alteration of the underlying genetic data. We showed how to analyze this bidimensional system with an approximation to the detailed balance case equivalent to a discrete version of the Helmholtz theorem for the vector field decomposition, and we used this method to solve some simple cases and discussed the thermodynamic properties of the system.

In the fourth chapter, a work in progress, we studied several parameters of a mouse cell line with fluorescence microscopy to analyze the morphological changes that an eukaryotic cell undergo while it reaches the replicative senescence. The total quantity of proteins in the nucleus is studied from a theoretical point of view, starting from a simplified model and writing a more detailed one to take into consideration how the cell reproduction influences the protein amount and how this change can be modeled in respect to the senescence. Extensive fit of several dataset has been made with an *ad hoc* technique to evaluate the goodness of fit, showing how the previsions of the model are compatible with the observed data, giving a hint on a possible minimal model for the total protein distribution inside the cell.

Summarizing, my opinion is that the stochastic framework used in this thesis and the biochemical modeling described by it, will be of ever-growing importance in the field of biology and biotechnology, as the growing number of works on this topic confirms. The power to predict and modelize, not only the expected values of the observed parameters, but also their whole probability distribution, is being recognized as a cardinal development in a field that, a couple of decades ago, was still thought as far from the theoretical physics as possible.

Bibliography

[1] A. Arkin, J. Ross, and H. H. McAdams, *Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected e. coli cells*, Genetics, 149 (1998), p. 1633â1648.

[2] A. Arvey, E. Larsson, C. Sander, C. S. Leslie, , and D. S. Marks, *Target mrna abundance dilutes microrna and sirna activity*, Molecular Systems Biology, 6, p. 363.

[3] B. Lüscher and L-G Larsson, *The world according to MYC*, EMBO reports, 8 (2007), pp. 1110–1114.

[4] C. Barbara, W. Assaf, D. F. Francesca, M. F. M., C. Elisa, R. R. L., G. Lorna, B. Dilair, R. Grazisa, G. Paolo, P. Massimiliano, B. Tiziana, R. Jiannis, F. Nir, C. Giorgio, B. M. E., and A. Alessandra, *Substantial histone reduction modulates genomewide nucleosomal occupancy and global transcriptional output*, PLoS Biol, 9(6) (2011), p. 1218â1225.

[5] N. Barkai and S. Leibler, *Biological rhythms: Circadian clocks limited by noise*, Nature, 403 (2000), p. 267â268.

[6] E. Batchelor, A. Loewer, C. Mock, and G. Lahav, *Stimulus-dependent dynamics of p53 in single cells*, Molecular Systems Biology, 7; 488 (2011).

[7] A. BD, K. Y, P.-H. MG, F. A, and M. CB, *Microrna regulation of a cancer network: consequences of the feedback loops involving mir-17-92, e2f, and myc*, Proc Natl Acad Sci U S A., 105(50) (2008), pp. 19678–83.

[8] Behnel, S. and Bradshaw, R. and Citro, C. and Dalcin, L. and Seljebotn, D.S. and Smith, K., *Cython: The Best of Both Worlds*, Computing in Science Engineering, 13 (2011), pp. 31 –39.

[9] O. Berg, J. Paulsson, and M. Ehrenberg, *Fluctuations and quality of control in biological cells: zero-order ultrasensitivity reinvestigated*, Biophysical Journal, 79 (2000), pp. 1228–1236.

[10] C. M. Brown, *Fluorescence microscopy: avoiding the pitfalls*, Journal of Cell Science, 120 (2007), pp. 1703–1705.

[11] P. Buchholz, J.-P. Katoen, P. Kemper, , and C. Tepper., *Model-checking large structured markov chains*, J. Log. Algebr. Program., 56(1-2) (2003), p. 69â97.

[12] L. Cai1, N. Friedman, and X. S. Xie, *Stochastic protein expression in individual cells at the single molecule level*, nature, 440 (2006), pp. 358–363.

[13] R. Cano and M. Borucki, *Revival and identification of bacterial spores in 25- to 40-million-year-old dominican amber*, Science, 268 (1995), pp. 1060–64.

[14] Y. Cao, D. T. Gillespie, and L. R. Petzold, *Efficient step size selection for the tau-leaping simulation method*, THE JOURNAL OF CHEMICAL PHYSICS, 124 (2006), pp. 044109 1–11.

[15] L. J. Cao Y, *An optimal algorithm for enumerating state space of stochastic molecular networks with small copy numbers of molecules.*

[16] H.-M. Chan, L.-S. Chan, R. N.-S. Wong, , and H.-W. Li, *Direct quantification of single-molecules of microrna by total internal reflection fluorescence microscopy*, Anal. Chem., 82 (2010), p. 6911â6918.

[17] D. Coling and B. Kachar, *Principles and application of fluorescence microscopy*, Current Protocols in Molecular Biology, 14.10 (2003), pp. 1–44.

[18] ——, *Theory and application of fluorescence microscopy*, Current Protocols in Neuroscience, 2 (2007), pp. 1–24.

[19] F. D and E. J, *Noise-induced min phenotypes in e. coli*, PLoS Comput Biol, 2 (6) (2006), p. 80.

[20] R. D., O. B., I. N., S. J.M., N. N., C. G.C., and C. L.N., *Targeting c-myc-activated genes with a correlation method: detection of global changes in large gene expression network dynamics*, PNAS, 102(19) (2005), p. 6902â6906.

[21] R. D, B. C, C. G, and C. R, *Characterization of spinal ganglion neurons in horse (equus caballus). a morphometric, neurochemical and tracing study*, Neuroscience, (2011).

[22] D. Daly, D. D. Deavours, J. M. Doyle, P. G. Webster, , and W. H. Sanders., *Mobius: An extensible tool for performance and dependability modeling.*, Computer Performance Evaluation / TOOLS, (2000), p. 332â336.

[23] F. Didier, T. A. Henzinger, M. Mateescu, , and V. Wolf, *Approximation of event probabilities in noisy cellular processes*, Proc. of CMSBâ09, 5688 of LNCS (2009), p. 173â18.

[24] W. D.J., *Stochastic modelling for quantitative description of heterogeneous biological systems*, Nature Review Genetics, 10 (2009), pp. 122–133.

[25] D.Levens and A. Gupta, *Reliable noise*, Science, 370 (2010), pp. 1088–1089.

[26] G. DT, *Exact stochastic simulation of coupled chemical reactions*, The Journal of Physical Chemistry, 81(25) (1977), p. 2340â2361.

[27] L. D. E and H. K., *E2f1 is crucial for e2f-dependent apoptosis*, EMBO Rep., 6(7) (2005), pp. 661–668.

[28] J. Elf, K. Nilsson, T. Tenson, , and M. Ehrenberg, *Bistable bacterial growth rate in response to antibiotics with low membrane permeability*, Phys. Rev. Lett., 97 (2006), pp. 258104 1–4.

[29] B. English, W. Min, A. van Oijen, K. Lee, G. Luo, H. Sun, B. Cherayil, S. Kou, and X. Xie, *Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited.*, Nat Chem Biol, 2 (2006), pp. 87–94.

[30] Eric Jones and Travis Oliphant and Pearu Peterson and others, *SciPy: Open source scientific tools for Python*, 2001–.

[31] J. Feser, D. Truong, C. Das, J. J. Carson, J. Kieft, T. Harkness, and J. K. Tyler, *Elevated histone expression promotes life span extension*, Mol Cell, 39 (2010), p. 724â735.

[32] M. Fricker, J. Runions, and I. Moore, *Quantitative fluorescence microscopy: From art to science*, Annual Review of Plant Biology, 57 (2006), pp. 79–107.

[33] N. Friedman, L. Cai, and X. S. Xie, *Linking stochastic dynamics to population distribution:an analytical framework of gene expression*, PHYSICAL REVIEW LETTERS, 97 (2006), pp. 168302 1–4.

[34] C. C. Gardner TS and C. JJ, *Construction of a genetic toggle switch in escherichia coli.*, Nature, 403 (2000), pp. 339–342.

[35] G.Bella and P.Lio, *Analysing the microrna-17-92/myc/e2f/rb compound toggle switch by theorem proving*, NETTAB, 9th Workshop on Network Tools and Applications in Biology, (2009), pp. 59–62.

[36] ——, *Formal analysis of the genetic toggle*, CMSB 2009, 7th Conference on Computational Methods in Systems Biology, (2009), pp. 96–110.

[37] C. G.C., B. A., Cooper, and L.N., *Toward a microscopic model of bidirectional synaptic plasticity.*, PNAS, 106(33) (2009), pp. 14091–14096.

[38] E. Giampieri, D. Remondini, L. de Oliveira, G. Castellani, and P. Lió, *Stochastic analysis of a miRNA-protein toggle switch*, Molecular Biosystems, 7 (2011), pp. 2796–2803.

[39] D. Gillespie, *The chemical Langevin equation*, Journal of Chemical Physics, 113(1) (2000), pp. 297–306.

[40] D. Gillespie, *The Chemical Langevin and Fokker-Planck Equations for the Reversible Isomerization Reaction*, J. Phys. Chem. A, 106 (20) (2002), pp. 5063–5071.

[41] D. T. Gillespie and L. Petzold, *Approximate accelerated stochastic simulation of chemically reacting systems*, Journal of Chemical Physics, 115 (2001), p. 1716â1733.

[42] A. Goldbeter and D. Koshland, *An amplified sensitivity arising from covalent modification in biological systems*, Proceedings of the National Academy of Sciences of the United States of America, 78 (1981), p. 6840.

[43] L. E. Guldberg and waage, *the law of mass action*, 1965, 42 (J. Chem. Ed), p. 548â550.

[44] K. H, K. M, A. M, C. K, G. TS, C. CR, and C. JJ, *Programmable cells: Interfacing natural and engineered gene networks*, PNAS, 101(22) (2004), pp. 8414–8419.

[45] C. HA, F. JJ, and L.-M. A, *Mycâed messages: Myc induces transcription of e2f1 while inhibiting its translation via a microrna polycistron*, PLoS Genet, 3 (2007), p. e146.

[46] E. Haustein and P. Schwille, *Trends in fluorescence imaging and related techniques to unravel biological information*, HFSP Journal, 1 (2007), pp. 169–180.

[47] F. He, *Deriving a neutral model of species abundance from fundamental mechanisms of population dynamics*, Functional Ecology, 19 (2005), p. 187â193.

[48] A. Heynen, M. Shuler, and M. B. J.R. Whitlock, *Learning induces long-term potentiation in the hippocampus*, Science, 313 (2006), pp. 1093–1097.

[49] D. J. Higham and R. Khanin, *Chemical Master versus Chemical Langevin for First-Order Reaction Networks*, the open applied mathematics journal, 2 (2008), pp. 59–79.

[50] C. Hirel, R. A. Sahner, X. Zang, , and K. S. Trivedi, *Reliability and performability modeling using sharpe 2000*, Computer Performance Evaluation / TOOLS, (2000), p. 345â349.

[51] K. Huang, *Statistical Mechanics*, Wiley, 1987.

[52] Q. Huang and H. Qian, *Ultrasensitive dual phosphorylation dephosphorylation cycle kinetics exhibits canonical competition behavior*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 19 (2009), p. 033109.

[53] Hunter, John D., *Matplotlib: A 2D graphics environment*, Computing In Science and Engineering, 9 (2007), pp. 90–95.

[54] L. J. and Q. H, *Computational cellular dynamics based on the chemical master equation: A challange for understanding complexity*, J. Comput. Sci. Tech., 25 (2010), p. 154â168.

[55] F. K, Y. I, H. DP, and A. ER., *Prediction and genetic demonstration of a role for activator e2fs in myc-induced tumors*, Cancer Res, 71 (2011), pp. 1924–1932.

[56] O. KA, W. EA, Z. KI, D. CV, and M. JT, *c-myc-regulated microrna modulate e2f1 expression*, Nature, 435(7043) (2005), pp. 839–43.

[57] M. Kimura, *The Neutral Theory of Molecular Evolution*, Cambridge University Press, 1983.

[58] E. Krebs, A. Kent, and E. Fischer, *The muscle phosphorylase b kinase reaction.*, J Biol Chem, 231 (1958), pp. 73–83.

[59] M. Z. Kwiatkowska, G. Norman, and D. Parker., *Prism 2.0: A tool for probabilistic model checking.*, QEST, (2004), p. 322â323.

[60] A. Levine, *p53, the cellular gatekeeper for growth and division*, Cell, 88 (1997), pp. 323–31.

[61] H. Li and L. Petzold, *Logarithmic direct method for discrete stochastic simulation of chemically reacting systems*, Technical Report, (2006).

[62] J. W. Lichtman and J. A. Conchello, *Fluorescence microscopy*, Nature Methods, 2 (2005), pp. 910–919.

[63] H. Lim and A. van Oudenaarden, *A multistep epigenetic switch enables the stable inheritance of dna methylation states*, Nature Genetics, 39 (2007), pp. 269–275.

[64] L. P. Lim, N. C. Lau, E. G. Weinstein, A. Abdelhakim, S. Yekta, M. W. Rhoades, C. B. Burge, , and D. P. Bartel, *The microrna of caenorhabditis elegans*, Genes Dev., 17(8) (2003), p. 991â1008.

[65] L. LP, L. NC, G.-E. P, G. A, S. JM, and et al., *Microarray analysis shows that some microrna downregulate large numbers of target mrna.*, Nature, 433 (2005), p. 769â773.

[66] Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., and et al., *Microrna expression profiles classify human cancers*, Nature, 435 (2005), p. 834â838.

[67] S. M, P. S, and A. AP, *Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations*, PNAS, 102(7) (2005), pp. 2310–2315.

[68] E. MB, L. MJ, S. ED, and S. PS, *Stochastic gene expression in a single cell*, Science, 297 (2002), p. 1183â1186.

[69] H. H. McAdams and A. Arkin, *Stochastic mechanisms in gene expression*, PNAS, USA, 94 (1997), p. 814â819.

[70] J. Mettetal and A. van Oudenaarden, *Necessary noise*, Science, 317 (2007), p. 463.

[71] J. T. Mettetal, D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. van Oudenaarden, *Predicting stochastic gene expression dynamics in single cells*, PNAS, 103-19 (2006), pp. 7304–7309.

[72] L. Michaelis and M. Menten, Biochem. Z., 49 (1913), pp. 333–369.

[73] X. Michalet, A. N. Kapanidis, T. Laurence, F. Pinaud, S. Doose, M. Pflughoefft, and S. Weiss, *The power and prospects of fluorescence microscopies and spectroscopies*, Annual Review of Biophysics and Biomolecular Structure, 32 (2003), pp. 161–182.

[74] W. Min, I. Gopich, B. English, S. Kou, X. Xie, and A. Szabo, *When does the Michaelis-Menten equation hold for fluctuating enzymes?*, J Phys Chem B, 110 (2006), pp. 20093–7.

[75] B. MJ, de Castro IP, and M. M, *Control of cell proliferation pathways by microrna*s, Cell Cycle, 7(20) (2008), pp. 3143–8.

[76] A. J. North, *Seeing is believing? a beginners' guide to practical pitfalls in image acquisition*, Journal of Cell Biology, 172 (2006), pp. 9–18.

[77] T. E. Oliphant, *Python for scientific computing*, Computing in Science and Engineering, 9 (3) (2007), pp. 10–20.

[78] L. Onsager, *Reciprocal relations in irreversible processes. i.*, Phys. Rev., 37 (1931), pp. 405–426.

[79] F. Ortega, J. Garcés, F. Mas, B. Kholodenko, and M. Cascante, *Bistability from double phosphorylation in signal transduction*, FEBS JOURNAL, 273 (2006), p. 3915.

[80] B. D. P., *Microrna*s*: genomics, biogenesis, mechanism, and function*, Cell, 116 (2004), pp. 281–297.

[81] T. P, S. N, O. R, R. A, P. C, O. MC, R. MF, and L. G., *E2f1-3 are critical for myeloid development*, J Biol Chem, 286 (2011), pp. 4783–4795.

[82] J. Paulsson and M. Ehrenberg, *Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks*, Phys. Rev. Lett., 84 (23) (2000), p. 5447.

[83] Pérez, Fernando and Granger, Brian E., *IPython: a System for Interactive Scientific Computing*, Comput. Sci. Eng., 9 (2007), pp. 21–29.

[84] H. R. Petty, *Fluorescence microscopy: established and emerging methods, experimental strategies, and application in immunology*, Microscopy Research and Technique, 70 (2007), pp. 687–709.

[85] Propp, James Gary and Wilson, David Bruce, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures and Algorithms, 9 (1996), pp. 223–252.

[86] H. Qian, *Thermodynamic and kinetic analysis of sensitivity amplification in biological signal transduction*, Biophysical chemistry, 105 (2003), pp. 585–593.

[87] H. Qian and L. M. Bishop, *The chemical master equation approach to nonequilibrium steady-state of open biochemical systems: Linear single-molecule enzyme kinetics and nonlinear biochemical reaction networks*, Int. J. Mol. Sci., 11 (2010), pp. 3472–3500.

[88] H. Qian, J. Cooper, et al., *Temporal Cooperativity and Sensitivity Amplification in Biological Signal Transduction*, Biochemistry, 47 (2008), pp. 2211–2220.

[89] ——, *Temporal Cooperativity and Sensitivity Amplification in Biological Signal Transduction*, Biochemistry, 47 (2008), pp. 2211–2220.

[90] H. Qian and E. Elson, *Single-molecule enzymology: stochastic Michaelis-Menten kinetics.*, Biophys Chem, 101-102 (2002), pp. 565–76.

[91] E. Quinlan, L. Cooper, H. Shouval, and G.C. Castellani, *A biophysical model of bidirectional synaptic plasticity: dependence on AMPA and NMDA receptors.*, Proceedings of the National Academy of Sciences, 98(22) (2001), pp. 12772–7.

[92] A. Raj and A. van Oudenaarden, *Nature, nurture, or chance: stochastic gene expression and its consequences.*, Cell, 135 (2008), pp. 216–26.

[93] U. Resch-Genger, K. Hoffmann, and A. Hoffmann, *Standardization of fluorescence measurement: Criteria for the choice of suitable standards and approaches for fit-for-purpose calibration tools*, Annals of the New York Academy of Sciences, 1130 (2008), pp. 35–43.

[94] O. RJ, K. S, S. SL, and K. J, *Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres*, Nat Struct Mol Biol, 17 (2010), p. 1218â1225.

[95] M. Samoilov, S. Plyasunov, and A. Arkin, *Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations.*, Proc Natl Acad Sci U S A, 102 (2005), pp. 2310–5.

[96] J. Schnakenberg, *Network theory of microscopic and macroscopic behavior of master equation systems*, Rev. Mod. Phys., 48 (1976), pp. 571–585.

[97] L. A. Segel, *Mathematical models in molecular and cellular biology*, Cambridge Univ Pr, 1980.

[98] ——, *Modeling dynamic phenomena in molecular and cellular biology*, Cambridge University Press, 1984.

[99] S. SH, *Nonlinear dynamics and chaos*, Perseus Books, 1994.

[100] G. S. Stein, W. E. Criss, and H. P. Morris, *Properties of the genome in experimental hepatomas: Variations in the composition of chromatin*, Life Sciences, 14(1) (1974), p. 95â105.

[101] E. H. K. Stelzer, *Contrast, resolution, pixelation, dynamic range and signal-to-noise ratio: fundamental limits to resolution in fluorescence light microscopy*, Journal of Microscopy, 189 (1998), pp. 15–24.

[102] SymPy Development Team, *SymPy: Python library for symbolic mathematics*, 2009.

[103] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, *Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells*, Science, 329 (2010), p. 333.

[104] K. T.B. and E. T.C., *Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations*, Biophys. J., 81 (2001), p. 3116â3136.

[105] T. To and N. Maheshri, *Noise can induce bimodality in positive transcriptional feedback loops without bistability*, Science, 327 (2010), pp. 1142–1145.

[106] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North Holland, third ed., May 2007.

[107] G. van Rossum, *Python tutorial*, Technical Report CS-R9526 Centrum voor Wiskunde en Informatica, (1995).

[108] M. Vellela and H. Qian, *A quasistationary analysis of a stochastic chemical reaction: Keizers paradox.*, Bull. Math. Biol., 69 (2007), pp. 1727–1746.

[109] I. Volkov1, J. R. Banavar1, S. P. Hubbell, and A. Maritan, *Patterns of relative species abundance in rainforests and coral reefs*, Nature, 450 (2007), pp. 45–49.

[110] J. C. Waters, *Accuracy and precision in quantitative fluorescence microscopy*, Journal of Cell Biology, 185 (2009), pp. 1135–1148.

[111] D. E. Wolf, *Fundamentals of fluorescence and fluorescence microscopy*, Methods in Cell Biology, 81 (2007), pp. 63–91.

[112] V. Wolf, R. Goel, M. Mateescu, and T. A. H. R. nartgicl e, *the chemical master equation using sliding windows*, BMC Systems Biolog, (2010), p. 4:42.

[113] F. S. Wouters, *The physics and biology of fluorescence microscopy in the life sciences*, Contemporary Physics, 47 (2006), pp. 239–255.

[114] W. Xiong and J. Ferrell, *A positive-feedback-based bistable memory module that governs a cell fate decision*, Nature, 426 (2003), pp. 460–465.

[115] S. Y, D. G. V, Q. E, M. UK, B. V, M. F, F. G, and C. P, *An e2f/mir-20a autoregulatory feedback loop*, J Biol Chem, 282(4) (2007), pp. 2135–43.

[116] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie, *Probing gene expression in live cells, one protein molecule at a time*, science, 311 (2006), p. 1600.

Ringraziamenti

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aenean ipsum sem, porta ut imperdiet vitae, varius in ipsum. Mauris posuere vehicula sem in vestibulum. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Donec adipiscing rutrum nisl, eget malesuada velit ultrices a. Maecenas tempus felis ut massa pulvinar adipiscing a non felis. Quisque facilisis lobortis nibh, sit amet pulvinar nulla ornare sed. Vivamus vehicula viverra ante id convallis. Nam ante diam, ornare vel luctus non, pellentesque eget libero. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed facilisis convallis diam, eget ullamcorper turpis dictum pretium. Cras sit amet arcu eget nisi consectetur pulvinar eget ut nunc. Maecenas ac odio neque, sed tempus nisi. aggiunge la bibliografia come elemento del TOC