Chapter 1

Stochastic modeling of complex macromolecular decay

Angelo Valleriani[*]

1 A short model of gene expression

There are many books and articles summarizing what we know about gene expression. Furthermore, the more we study this aspect of life, the more we find out how complex it is and how variable such concepts like gene, genetic code and protein are. Nevertheless, since this article refers to modeling one aspect of the regulation of gene expression, it is perhaps important to provide the reader with at least a very superficial but quick view of the big framework.

Gene expression is a generic term that is commonly related to what is known as the central dogma of molecular biology. According to the commonly accepted version of this dogma, it all starts with the DNA, which is a chain of four nucleotides A, C, G, T, organized in chromosomes. On the chromosomes, the genes occupy specific regions called loci. In humans, there are 23 different chromosomes, each coming in two copies, one inherited from the mother and one from the father, respectively. The sequence of nucleotides of one gene on a certain locus could be different on the two chromosomes

[*]Max Planck Institute of Colloids and Interfaces, Department of Theory and Bio-Systems; *angelo.valleriani@mpikg.mpg.de*

but in general the number of differences is very small or absent. Population genetics, also a branch of applied stochastic methods, deals with some mechanisms of propagation of gene variants in populations [1, 2]. Certain cells have only one set of chromosomes, e.g. sperm or egg cells. Certain organisms like bacteria have only one single chromosome on which all genes are collected.

The genes found in the DNA are first transcribed into RNA molecules. The majority of the total RNA molecules present in each cell plays a key role in the production of proteins. Some species of RNA molecules become part of ribosomes, which are complex molecular machines made of RNA and proteins. Some RNAs become transport RNA, called also tRNA. Some RNAs are found in the form of small or micro RNA and finally a prominent role in gene expression is played by messenger RNA molecules (mRNA).

Protein synthesis is the final product of gene expression: in this process, the ribosomes read the information encoded in the mRNA and synthesize the proteins using the amino-acid delivered by the tRNAs. This process is called translation. The particular way in which the ribosome reads the mRNA is the basis of what we know as the genetic code. The ribosome reads the nucleotide sequence of the mRNA one triplet of amino-acids per step. To each triplet, called codon, corresponds one amino-acid that will be incorporated into the nascent protein.

The amount of proteins corresponding to a given gene present in the cell will thus depend on several factors. The first factor is the amount of mRNAs of that gene: this is determined by the balance between the synthesis rate of the mRNA (transcription rate) and the degradation rate of the mRNA. The second factor is the amount of ribosomes translating each mRNA molecules, which eventually determines the protein synthesis rate (if we neglect possible translation errors such as ribosome drop-off [3]). The final factor is the degradation rate of the proteins.

The RNA molecules that become part of the ribosome are called rRNA. Both rRNA and tRNA are very stable. Their duty is to provide the machinery of the process of translation, independently of what has to be translated. The mRNA molecules instead are typically not so stable and their lifetime is regulated by some internal cellular mechanisms. Indeed, when the cell needs to change the kind of proteins to be synthesized, due for instance to some stressful condition, it can do that by changing the composition of the cell mRNA population. Beside the important role played by the regulation of transcription, one way to tune the amount of mRNA is to activate or deactivate specific degradation

mechanisms. These degradation mechanisms usually involve several biochemical pathways, which work in parallel or in competition with each other. As an example, micro RNA (miRNA) can bind to a specific mRNA together with some proteins, whose main scope is to recruit other enzymes that start or trigger the digestion of the mRNA [4, 5]. The digestion of the mRNA proceeds usually through several steps until the molecule is not detectable in a fashion that depends on the kind of mRNA and on the organism [6]. This series of steps can be quite complex, since it requires the subsequent participation of several molecules. Since the degradation of mRNA runs over several steps, the lifetime of such molecules is very complex and so we should expect that their decay pattern is not exponential [7, 8, 9, 10]. The degradation of proteins is also a mechanism to regulate gene expression. Also for proteins we know that several factors are needed to degrade them and so we expect their lifetime to be complex and their decay pattern to be non-exponential [11, 12, 13].

Without entering into any of the biochemical mechanisms that determine the degradation of mRNA or proteins, we may still briefly discuss how the decay pattern of such molecules are determined experimentally. There are obviously several experimental techniques. All of them have in common the idea of following the decay in time of the amount of molecules of the species of interest. To achieve this, the experimentalists first cultivate the cells for a certain fixed amount of time during which the molecules of interest are marked, e.g. with radioactive isotopes. This period of labeling with the marker is called the *pulse phase*. Afterwards, the labeling of the molecules is stopped and the experiment proceeds with the harvesting of the molecules at fixed time intervals from the cell extracts of a large amount of cells. This period of harvesting is called the *chase phase*. The measurements at these time points will provide a decaying curve that we call the *decay pattern* of the specific molecule. The curve decays because the amount of labeled molecules cannot increase. In some cases, the pulse phase can be so long that all molecules of interest are labeled. We call this situation the *steady state expression*. In this situation a balanced amount of old and newly synthesized molecules is present in the cell, i.e. we have a stationary age distribution of those molecules. When instead the pulse is extremely short, the labeled molecules are all very young and the age distribution is far away from stationarity. As we shall see in the following, the length of the pulse will make a big difference on how the decay pattern looks like [11, 13]. This is the starting point of our journey towards modeling the process of complex decay.

1.1 Types of decay patterns

Traditionally, decay experiments have been performed under the assumption of an exponential decay. Under this assumption, the decay pattern in a semilog plot looks like a straight line. The fit requires just two point and the liner regression of the data, from which the rate of degradation can be extracted.

As more experimental techniques became available, more datapoints during the chase phase were collected thus revealing that the behavior is often very different from a straight line in a semilog plot (see [9] and references therein). In most cases, the decay pattern could be superficially described as fast at short times and slow at long times thus leading to the idea of a double exponential. Most careful experiments and theory have however shown that the decay pattern changes in shape depending on the length of the pulse phase even if the molecules under examination are the same [11, 13]. From the biochemical point of view, a non-exponential decay pattern means only that the degradation process is complex, i.e. it is composed of several steps [7, 9].

2 Relationship between Poisson processes and uniform distribution

This section states and summarizes a known result concerning the relationship between a Poisson process with constant rate and the uniform distribution [14]. We will show that conditional on a number N of events at a certain time t, the time of origin of the N events is uniformly distributed in $[0,t)$.

Let $X(t)$ be a Poisson process with rate λ and initial condition $X(0)=0$. If $X(t)=n$ we will say that until time t there have been exactly n events. Textbook calculations show that for any fixed t, the number of events n is distributed according to a Poisson distribution

$$\mathsf{P}\{X(t)=n \mid X(0)=0\} = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}, \tag{1.1}$$

which implies also that for $u \leq t$

$$\mathsf{P}\{X(t)=n \mid X(u)=k\} = \frac{[\lambda(t-u)]^{n-k} \exp\big(-\lambda(t-u)\big)}{(n-k)!}, \tag{1.2}$$

holds.

Now we would like to show that under the condition of a fixed number n of events until time t, the time points at which the n events arose are uniformly distributed in $[0,t)$. To proceed with this comparison, we shall first state one of the elementary properties of the uniform distribution. Imagine to throw n points at random on a line of length t such that each throw is independent and the probability density function of the position of each point is uniform. Let now be u an arbitrary point in $[0,t)$ and let Y_u be the number of points, out of the n points thrown in total over the line, that fall into the segment $[0,u)$. Since each point has probability u/t to fall over this segment, the probability distribution for Y_u is given by the binomial distribution

$$\mathsf{P}\{Y_u = k\} = \binom{n}{k} \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k}. \tag{1.3}$$

Let us turn our attention back to the Poisson process and let us consider the ensemble of Poisson processes with rate λ conditioned that $X(t) = n$, namely conditioned that the number of events up to time t is n. Under this condition, we would like to know what is the distribution of the number of events of the process occurring during the interval of time $[0,u)$ for $u \in [0,t)$. This is equivalent to ask about the probability distribution of the stochastic variable $X(u)$. This leads to

$$\begin{aligned}
&\mathsf{P}\{X(u) = k \mid X(0) = 0, X(t) = n\} \\
&\qquad = \frac{\mathsf{P}\{X(u) = k, X(t) = n \mid X(0) = 0\}}{\mathsf{P}\{X(t) = n \mid X(0) = 0\}} \\
&\qquad = \frac{\mathsf{P}\{X(t) = n \mid X(u) = k\}\mathsf{P}\{X(u) = k \mid X(0) = 0\}}{\mathsf{P}\{X(t) = n \mid X(0) = 0\}} \\
&\qquad = \frac{n!}{k!(n-k)!} \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k},
\end{aligned} \tag{1.4}$$

where the last line arises after using (1.1) and (1.2). The comparison between (1.3) and (1.4) should be enough to convince ourselves that the distribution of the number of events falling into any interval of time in the conditioned Poisson process is identical to the distribution of the number of points falling into the same interval. We conclude that the position in time of the events of the conditioned Poisson process follows a uniform distribution.

3 The time dependent distribution of the number of molecules

In this section we will assume that the pulse phase starts at time zero and that the molecules under consideration have an arbitrary life time distribution. We will first derive the distribution of the number of molecules at time t after the start of the pulse. Later we will assume that the pulse phase is interrupted at time t_p and derive the time dependent distribution of the number of molecules at time $t_p + \Delta t$ during the chase phase. Finally, we will derive the time dependent age and rest-life distributions of the molecules as a function of time. To simplify the discussion, in the rest of these notes we will concentrate on the analysis of mRNA decay as an example of application. The extension of the analysis and modeling to protein decay is then easy.

3.1 Basic assumptions about transcription and degradation

As specified earlier, mRNA molecules are synthesized from their DNA template through a process called transcription. For the whole content of this section we will assume that transcription occurs at a fixed constant rate ω_{tc} and that at each transcription event only one mRNA is synthesized. To be precise, in eukaryotic cells the process of transcription does not directly deliver mRNA molecules. In this kind of cells, our ω_{tc} is then the rate of mRNA delivery in the cytoplasm of the cell. Let $R(t)$ be the stochastic variable that gives the number of labeled mRNA found at time t after the start of the pulse phase. We assume that the pulse phase starts at time zero and that thus the initial condition of our process is given by $R(0) = 0$. Thus, in the absence of mRNA degradation, the process of transcription is a Poisson process with constant rate ω_{tc}. Another assumption that we make concerns the life time distribution of each mRNA molecule. Let $U \in [0,\infty)$ be the random life time of an mRNA molecule. We will assume that the probability function $G(t_u)$ of U is given by

$$G(t_u) \equiv \mathsf{P}\{U \leq t_u\} = \int_0^{t_u} \phi_U(\tau)\mathrm{d}\tau\,, \tag{1.5}$$

where ϕ_U is the probability density function of U. We will not make any special restriction concerning the form of the density ϕ_U apart from the fact of being integrable in $[0,\infty)$ and having well defined moments. In particular, we will not impose that ϕ_U is an exponential

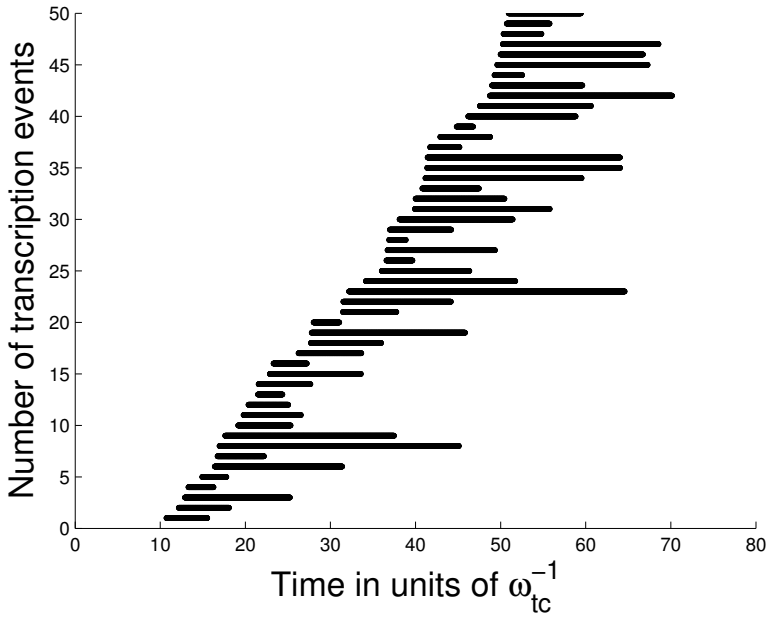

Figure 1.1: This figure shows the process of origination of the mRNA chains at random times starting from an arbitrary point in time, $t = 10$. The units of time have been arbitrarily fixed to be equal to ω_{tc}^{-1}. In this illustration, we have set $\langle U \rangle = 10\omega_{tc}^{-1}$. We see that the mRNA chains are generated at random points and have a random duration. In this example, just to fix the ideas, we have considered the gamma distribution $\phi_U(t) = (\lambda^2 t/2)\exp(-\lambda t)$. At any arbitrary point in time, one can count the number of chains that are still intact at that point. By following this number in time, one obtains the trajectory of the number of mRNA molecules originating from one gene in one cell. This statistics is shown in Figure 1.2.

distribution. As mentioned earlier, the exponential distribution has been considered so far as a suitable approximation but the available experimental data show that for most mRNAs there is no simple exponential fall off and we should consider more complex processes. Therefore, in order to leave the modeling open to any possible functional form of ϕ_U, we will henceforth consider the generic form given in (1.5). On the basis of these assumptions we have therefore that transcription at a fixed rate generates new mRNAs while degradation eliminates them. During the pulse phase, these new mRNAs are labeled. We should expect that after a certain amount of time, typically longer than

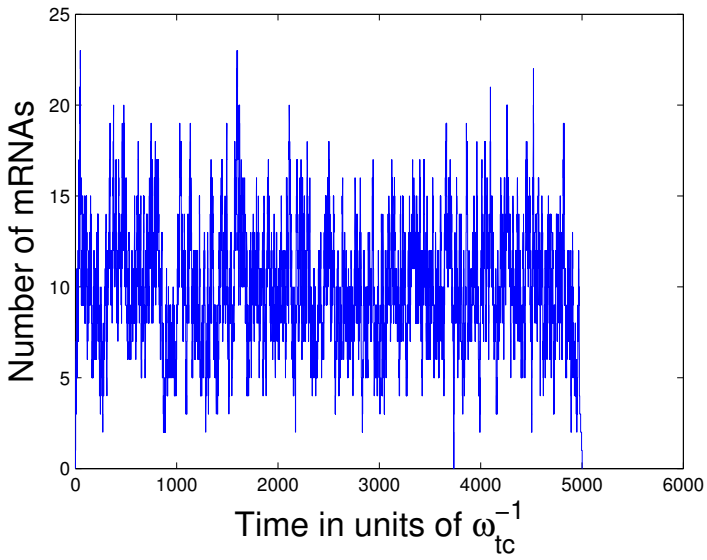

Figure 1.2: Number of mRNA molecules over time. From Figure 1.1 one can derive the number of mRNA chains present in the cell at any given point in time. By following this process over a long period of time, one can obtain a pictorial view of the kind of fluctuations that characterize the number of mRNA molecules. We see in this plot that the number of molecules attains a stationary value until, at the end of the process, the pulse phase is interrupted. Despite the stationarity of the process during a long period of time, the figure shows that the fluctuations are particularly strong. In this work we will provide a detailed analytical and exact description of such fluctuations.

the pulse duration in any common experimental set-up, the processes of synthesis and degradation reach a steady state and that the number of labeled mRNA R attains a stable distribution.

3.2 Formal derivation of the distribution of the amount of mRNA

Consider an experimental set-up in which the pulse duration is t_p. For a time $t \leq t_p$, let $X(t)$ be the underlying Poisson process that describes the amount of labeled mRNA molecules synthesized until time t with transcription rate ω_{tc}. The stochastic variable $R(t)$ is instead the random number of labeled mRNA molecules present in the cell at time

t. Therefore, while $X(t)$ just counts the number of synthesized molecules, $R(t)$ takes into account also the possibility that some molecules synthesized until time t might have been already degraded. Thus, it holds that $0 \leq R(t) \leq X(t)$ for all $t \geq 0$. Under the general assumptions made before, we wish to compute the distribution of $R(t)$, which we can formally write as

$$\mathsf{P}\{R(t) = k \mid X(0) = 0\}, \tag{1.6}$$

for any $k \geq 0$ and $t < t_p$. To approach this problem, we first write this probability by using the law of total probability as [7]

$$\begin{aligned}&\mathsf{P}\{R(t) = k \mid X(0) = 0\}\\ &\qquad = \sum_{n=k}^{\infty} \mathsf{P}\{R(t) = k \mid X(0) = 0, X(t) = n\}\mathsf{P}\{X(t) = n \mid X(0) = 0\}, \end{aligned}\tag{1.7}$$

and consider first $\mathsf{P}\{R(t) = k \mid X(0) = 0, X(t) = n\}$.

We have learned in Section 2 that conditioned on the number of events up to time t, the time position of the events is uniformly distributed in $[0,t)$. Let now be O the origination time of a randomly chosen mRNA and let U be its random lifetime. This mRNA molecule will be present at time t only if the variable $Z = O + U$ satisfies $Z \geq t$. The probability p of this event gives the probability per mRNA to be present at time t. We have therefore

$$p = \mathsf{P}\{Z \geq t\} = \frac{1}{t}\int_0^t \mathsf{P}\{Z \geq t \mid O = s\}\,\mathrm{d}s = \frac{1}{t}\int_0^t (1 - G(u))\,\mathrm{d}u, \tag{1.8}$$

where $G(t_u)$ is defined in (1.5) and we have made use of the fact that $\mathsf{P}\{Z \geq t \mid O = t_o\} = \mathsf{P}\{U \geq t - t_o\}$. Finally, it results that conditioned on $X(t) = n$ the number of labeled mRNA still present at time t is binomially distributed according to

$$\mathsf{P}\{R(t) = k \mid X(0) = 0, X(t) = n\} = \binom{n}{k} p^k (1-p)^{n-k}. \tag{1.9}$$

At this point, we are ready to consider again (1.7) by plugging in both (1.1) and (1.9). This simple computation delivers the time dependent distribution of the number of labeled mRNA at time t after the start of pulse phase, given by

$$\mathsf{P}\{R(t) = k \mid X(0) = 0\} = \frac{[\omega_{\mathrm{tc}} A(t)]^k \exp[-\omega_{\mathrm{tc}} A(t)]}{k!}, \tag{1.10}$$

where

$$A(t) = \int_0^t \big(1 - G(u)\big)\,\mathrm{d}u\,.$$

One can notice that (1.10) is a time dependent Poisson distribution with parameter $\omega_{\mathrm{tc}} A(t)$. Nevertheless, one can easily see that $A(t) \to \langle U \rangle$ as $t \to \infty$ with

$$\langle U \rangle = \int_0^\infty u \phi_U(u)\,\mathrm{d}u\,,$$

being the average life time of the mRNA molecules. Thus, for very long pulse durations ($t_p \to \infty$) the limit $t \to \infty$ leads to the stationary distribution

$$\mathsf{P}\{R^{\mathrm{st}} = k\} = \frac{[\omega_{\mathrm{tc}}\langle U \rangle]^k \exp\left[-\omega_{\mathrm{tc}}\langle U \rangle\right]}{k!}\,, \tag{1.11}$$

which depends only on the average life time and not anymore on the details of the degradation process. For the sake of illustration, a comparison between the theoretical distribution and its numerical estimation for a particular choice of $\langle U \rangle$ is shown in Figure 1.3.

Thus we see that the distribution of the amount of mRNA molecules after the start of the pulse phase and before time t_p is Poisson at each time point with a parameter that depends on time. In particular, the *transient time to stationarity* depends critically on the life time distribution of the mRNA [7]. Moreover, the transient time is large if the life distribution is very broad. This must be particularly long in eukaryotic cells, if we interpret ω_{tc} as the rate of delivery of the mRNA. We know indeed that in eukaryotic cells the average life time of the mRNA can be particularly long. In prokaryotic cells instead, the life time of most mRNA molecules is relatively short so that in most of the cases the stationary state is reached in about 20 minutes. The limiting (stationary) distribution given in (1.11) depends instead only on the mean life time of the mRNA chains and not on any other details of the degradation process as long as this process does not modify the average lifetime $\langle U \rangle$.

3.2.1 Time dependent life time distributions

This subsection can be omitted on a first reading. In this section we derive the distribution of the number of mRNA molecules for those cases when the time of origin of the mRNA is relevant. This is usually the case if the experimental conditions change with time.

It is of a certain interest at this point to have a brief look at the special case in which

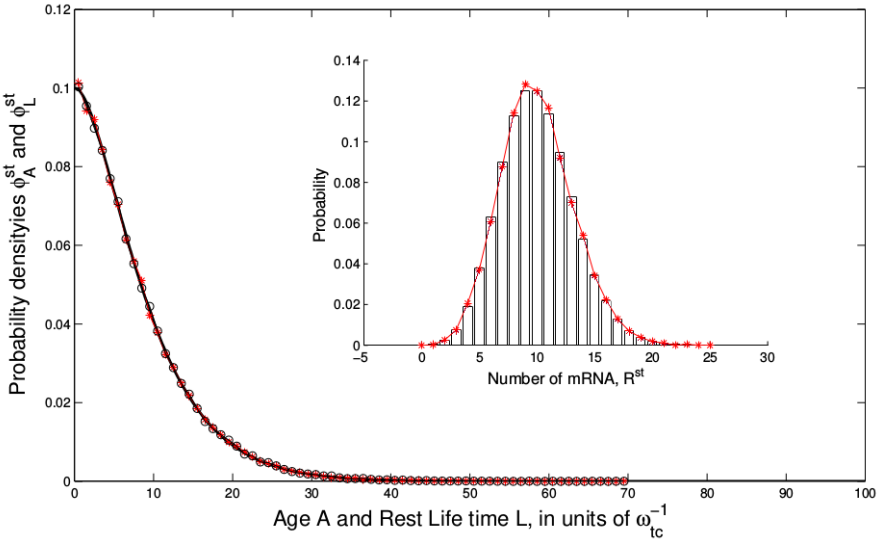

Figure 1.3: Age and rest life time distributions at steady state. Inset: stationary distribution of the number of mRNA. Under the same conditions of Figures 1.1 and 1.2 the main plot shows a comparison between the numerical estimation at steady state of the age of the mRNA taken at any random point in time (circles), the rest life time distribution (stars) and the theoretical predictions given in (1.28) and (1.29), respectively. In the inset, we see the comparison between the statistics from a computer simulation for the number of mRNA at stationary state with $\langle U \rangle = 10\,\omega_{tc}^{-1}$ with the corresponding theoretical distribution derived in (1.11).

the life time distribution of the mRNA depends on time t after the initiation of the experiment. This dependence may be due to changing environmental conditions during the experiments both because of intrinsic and extrinsic origins. Putting it in simple terms, this means the the time of origin of an mRNA cannot be forgotten because this determines the history of that chain.

Formally, let us define the random variable $C(\tau,t)$ that gives the state of a mRNA chain of age τ at time t. This stochastic variable takes values in the state space $\{0,1\}$ such that 0 is the degraded (absorbing) state and 1 is the intact state. We assume to have

$$\mathrm{P}\{C(\tau+h,t+h) = 0 \,|\, C(\tau,t) = 1\} = \omega(\tau,t)h + o(h)\,, \tag{1.12}$$

for $h \to 0$. In this case, it is an easy calculation to find out that

$$\mathsf{P}\{U > t_u \mid O = t_o\} = \exp\left(-\int_0^{t_u} \omega(x, t_o + x)\,\mathrm{d}x\right), \tag{1.13}$$

where t_o is the time at which the mRNA chain was originated. Finally, this leads to the probability function

$$\mathsf{P}\{U \leq t_u \mid O = t_o\} \equiv G(t_u, t_o) = 1 - \exp\left(-\int_0^{t_u} \omega(x, t_o + x)\,\mathrm{d}x\right), \tag{1.14}$$

which depends both on t_o and on t_u and not only on their difference. Therefore, to evaluate p we cannot follow the same route that led to (1.8). In this case, we need to consider that

$$\mathsf{P}\{Z \geq t \mid O = t_o\} = \mathsf{P}\{U \geq t - t_o \mid O = t_o\} = 1 - G(t - t_o, t_o), \tag{1.15}$$

and thus the probability p becomes

$$p = \frac{1}{t}\int_0^t \big(1 - G(t - s, s)\big)\,\mathrm{d}s = \frac{H(t)}{t}, \tag{1.16}$$

with the additional property that the limit of large t for $H(t)$ may not be known or even be unbounded. At this point we impose that

$$\lim_{x \to \infty} G(x, y) = 1 \tag{1.17}$$

for any value of y. Note that Eq. (1.17) is not contained in the definition of the degradation rate given in (1.12). By using the same formalism set in (1.12) we can then rewrite $H(t)$ as

$$H(t) = \int_0^t \exp\left(-\int_0^{t-s} \omega(x, s + x)\,\mathrm{d}x\right)\mathrm{d}s = \int_0^t \exp\left(-\int_s^t \omega(x - s, x)\,\mathrm{d}x\right)\mathrm{d}s, \tag{1.18}$$

which should converge for $t \to \infty$ if $\omega(x, y) \geq \omega_{\min} > 0$ for any $x, y \geq 0$. Finally, the time dependent distribution of the number of mRNA at time t after the start of transcription is given by

$$\mathsf{P}\{R(t) = k \mid X(0) = 0\} = \frac{[\omega_{\mathrm{tc}} H(t)]^k \exp(-\omega_{\mathrm{tc}} H(t))}{k!}. \tag{1.19}$$

3.3 The number of mRNA molecules after the pulse

In this section we consider the case in which the pulse is stopped (or interrupted) at some time t_p and we observe the number of mRNA molecules an interval Δt during the chase phase.

Let us assume that pulse phase is terminated at time t_p after having started at time zero. We have seen that the distribution of the number of labeled mRNAs for the transcribed gene is given by (1.10) for $t = t_p$. Since labeling is terminated, we expect that no further labeled mRNAs are produced and that, therefore, there is no increase in the number of labeled mRNAs after time t_p and, rather, we expect to witness a decrease of those mRNAs due to the effect of the degradation mechanisms.

We would like to find out what is the distribution of the number of mRNAs at time $t = t_p + \Delta t$, namely Δt time units after the interruption of labeled synthesis [9]. By using the same notation as in the previous calculations, we shall find

$$\mathsf{P}\{R(t_p+\Delta t) = k \mid X(0) = 0, \omega_{\mathrm{tc}} > 0 \text{ for } t \in [0,t_p), \omega_{\mathrm{tc}} = 0 \text{ for } t \in [t_p,\infty)\}, \quad (1.20)$$

where, in the condition, we have explicitly written the change in the synthesis (labeling) rate ω_{tc} underlying the interruption of labeling. To simplify the notation, we will henceforth omit the condition on the synthesis rate ω_{tc} of labeled molecules. In order to compute (1.20) we shall proceed in a fashion similar to what was done previously. The important information that has to be kept in mind is that here the time origin O of the mRNAs is in the interval $[0,t_p)$.

Rephrasing the formalism used in (1.7), we find

$$\begin{aligned}&\mathsf{P}\{R(t_p+\Delta t) = k \mid X(0) = 0\}\\ &\quad = \sum_{n=k}^{\infty} \mathsf{P}\{R(t_p+\Delta t) = k \mid X(0) = 0, X(t_p) = n\}\mathsf{P}\{X(t_p) = n \mid X(0) = 0\}, \quad (1.21)\end{aligned}$$

and consider first $\mathsf{P}\{R(t_p+\Delta t) = k \mid X(0) = 0, X(t_p) = n\}$. From the equivalence between Poisson processes and uniform distribution introduced in Section 2, conditioned on the number of events up to time t_p, the time position of the origin of each mRNA is uniformly distributed in $[0,t_p)$. Let now be O the random origination time of a randomly chosen mRNA and let U be its random life time. This mRNA molecule will be present at time $t_p + \Delta t$ only if the variable $Z = O + U$ satisfies $Z \geq t_p + \Delta t$. The probability p_c of this

event gives the probability per mRNA to be present at time $t_p + \Delta t$. We have therefore

$$p_c = \mathsf{P}\{Z \geq t_p + \Delta t\} = \frac{1}{t_p}\int_0^{t_p} \mathsf{P}\{Z \geq t_p + \Delta t \mid O = s\}\,\mathrm{d}s = \frac{1}{t_p}\int_{\Delta t}^{t_p+\Delta t} \big(1 - G(u)\big)\,\mathrm{d}u\,, \tag{1.22}$$

where $G(t_u)$ is defined in (1.5). Finally, it results that conditioned on $X(t_p) = n$ the number of labeled mRNA still present at time $t_p + \Delta t$ is binomially distributed according to

$$\mathsf{P}\{R(t_p + \Delta t) = k \mid X(0) = 0, X(t_p) = n\} = \binom{n}{k} p_c^k (1 - p_c)^{n-k}\,. \tag{1.23}$$

Using now (1.1) and (1.21) the time dependent distribution of the number of labeled mRNA at time $t_p + \Delta t$ after the start of the pulse and Δt time units after the interruption of the pulse phase and beginning of the chase phase is given by

$$\mathsf{P}\{R(t_p + \Delta t) = k \mid X(0) = 0\} = \frac{[\omega_{\mathrm{tc}} A_c(t_p + \Delta t)]^k \exp\big(-\omega_{\mathrm{tc}} A_c(t_p + \Delta t)\big)}{k!}\,, \tag{1.24}$$

where

$$A_c(t_p + \Delta t) = \int_{\Delta t}^{t_p+\Delta t} \big(1 - G(u)\big)\,\mathrm{d}u\,.$$

This relation is similar to (1.10) but not identical. In fact, in (1.24) we have that $A_c(t_p + \Delta t) \to 0$ as $\Delta t \to \infty$, expressing the fact that in the chase phase the number of mRNAs cannot increase. An important point from the experimental point of view is to realize that in (1.24) there is an explicit dependence on the length of the pulse t_p. This means that the distribution will be different for the same target mRNA depending on the value of t_p. This point plays a very important role in the analysis of decay data [13].

To describe the expression of a given gene after a very long pulse phase, it is sufficient to take the limit $t_p \to \infty$ in (1.24). This simple limit gives an expression for the number of mRNA Δt units of time after the start of the chase phase, which is again a Poisson distribution with parameter

$$A_c(\Delta t) = \omega_{\mathrm{tc}} \int_{\Delta t}^{\infty} \big(1 - G(u)\big)\,\mathrm{d}u\,, \tag{1.25}$$

where we see that the life time distribution determines also the behavior of the distribution during the chase phase.

3.4 The decay law of the mean number of mRNA

In experiments aimed at investigating the decay of the number of mRNA in order to deduce its half time t_h, it is not the distribution of the number of mRNA that is measured but the mean number over an ensemble of cells. This measurement is performed by interrupting the transcription and by measuring the relative amount of mRNAs at different time points after interruption.

To compare our results with the experiments we therefore need the average number N_r of mRNA as a function of time after the interruption of transcription [13]. This is simply given by the average performed over the distribution (1.24) and gives

$$N_r(t_p+\Delta t) = \omega_{\rm tc} A_c(t_p+\Delta t) = \omega_{\rm tc}\int_{\Delta t}^{t_p+\Delta t}\big(1-G(u)\big)\,{\rm d}u\,, \tag{1.26}$$

where the half time t_h measured in the experiments is then given by solving

$$N_r(t_p+t_h) = \frac{1}{2}N_r(t_p)\,, \tag{1.27}$$

as a function of t_h. By substituting (1.26) into (1.27) we find that t_h must satisfy

$$\int_{t_h}^{t_p+t_h}\big(1-G(u)\big)\,{\rm d}u = \frac{1}{2}\int_0^{t_p}\big(1-G(u)\big)\,{\rm d}u\,,$$

which, in the case in which $t_p\to\infty$ leads to

$$\int_{t_h}^{\infty}\big(1-G(u)\big)\,{\rm d}u = \frac{1}{2}\langle U\rangle\,.$$

To fix the ideas about this relationship, we will consider the example in which the life time probability density is a Gamma function of order $n=0$ and $n=1$, where

$$\phi_U(t) = \frac{\lambda(\lambda t)^n\exp(-\lambda t)}{n!}\,,$$

with $n=0,1,\ldots$ and $\lambda = {}^{(n+1)}\!/\!_{\langle U\rangle}$. For $n=0$, namely for an exponentially distributed life time, the half time t_h is related to the mean life time $\langle U\rangle$ through

$$t_h^{(n=0)} = \langle U\rangle\log 2 \sim 0.69\langle U\rangle\,,$$

as is well known. On the other hand, for $n = 1$ the relationship is

$$t_h^{(n=1)} \sim 0.57\langle U \rangle ,$$

for the same mean value. If one would now ignore that the result for $n = 1$ does not come from an exponential density and one would assume that it comes from an exponential, one would try to derive the mean value $\langle U \rangle'$ such that $t_h = \langle U \rangle' \log 2$. This would lead to

$$\langle U \rangle' = 0.83 \langle U \rangle ,$$

namely it would lead to an underestimation of the mean life time by about 20%.

3.5 Age and rest life time distributions

Given that there is a turn-over of the mRNA, there is an age distribution of the molecules. This age distribution depends on the length of the pulse phase and on the time point of measurement after the start of the chase phase. The same holds also for the rest life time. There is a connection with renewal theory here. In renewal theory the age is called *current life* and the rest life time is called *excess life*.

3.5.1 The age distribution of one chain

We shall start by considering the age distribution of one mRNAs at any arbitrary point in time $t + \Delta t$ after the start of the pulse phase[1]. At this point in time, the only age distribution that can be computed is the one of mRNAs that at this time are not yet degraded. In other words, one can compute the age distribution of an mRNA only under the condition that this molecule still exists at the moment of measurement. Using the same notation as before, let us call O the time of origination of a randomly chosen labeled mRNA and U its random life time. The random time of origination, O, must be contained in the interval $[0,t)$. The given mRNA will be present at time $t + \Delta t$ only if the variable $Z = O + U$ satisfies $Z > t + \Delta t$. Then, the age distribution of the mRNA is given by the distribution of $A = t + \Delta t - O$ under the condition $Z > t + \Delta t$.

[1] The choice of $t + \Delta t$ for the measurement time may seem to be a bit unusual at this point. We shall see that this choice allows for a natural derivation of a number of useful limits

In order to compute this quantity we shall first compute the probability density for O conditional that $Z > t + \Delta t$. Thus, we have

$$\begin{aligned}
\mathsf{P}\{O \le x \mid Z > t+\Delta t\} &= \frac{\mathsf{P}\{O \le x, Z > t+\Delta t\}}{\mathsf{P}\{Z > t+\Delta t\}} \\
&= \int_{t+\Delta t-x}^{\infty} \frac{\mathsf{P}\{O \le x, Z > t+\Delta t \mid U = u\}}{\mathsf{P}\{Z > t+\Delta t\}} \phi_U(u)\,\mathrm{d}u \\
&= \int_{t+\Delta t-x}^{\infty} \frac{\mathsf{P}\{t+\Delta t-u < O \le x\}}{\mathsf{P}\{Z > t+\Delta t\}} \phi_U(u)\,\mathrm{d}u
\end{aligned}$$

for $0 \le x < t$ by making a simple use of the definition of conditional probability and of the law of total probability. We shall now use the fact that the random variable O is uniformly distributed in $[0,t)$ (because we have conditioned that there is one mRNA alive at time $t+\Delta t$) and thus that $\mathsf{P}\{Z > t+\Delta t\}$ is given by a formula similar to (1.22). After canceling the normalization factors $t+\Delta t$ in the numerator and denominator, this leads to

$$\begin{aligned}
&\int_{t+\Delta t-x}^{\infty} \frac{\mathsf{P}\{t+\Delta t-u < O \le x\}}{\mathsf{P}\{Z > t+\Delta t\}} \phi_U(u)\,\mathrm{d}u \\
&\qquad = \left[\int_{\Delta t}^{t+\Delta t} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \int_{t+\Delta t-x}^{\infty} \big[x - \max(0, t+\Delta t-u)\big] \phi_U(u)\,\mathrm{d}u\,,
\end{aligned}$$

from which we can compute the distribution of the age $A = t+\Delta t - O$ under the condition $Z > t+\Delta t$. This distribution is given by

$$\begin{aligned}
\mathsf{P}\{A \le a \mid Z > t+\Delta t\} &= 1 - \mathsf{P}\{O \le t+\Delta t-a \mid Z > t+\Delta t\} \\
&= 1 - \left[\int_{\Delta t}^{t+\Delta t} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \int_{a}^{\infty} \big[\min(u, t+\Delta t) - a\big] \phi_U(u)\,\mathrm{d}u\,,
\end{aligned}$$

which finally leads to the probability density function after differentiation

$$\phi_A(a \mid t+\Delta t) = \left[\int_{\Delta t}^{t+\Delta t} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \big(1-G(a)\big)$$

for $\Delta t \le a < t+\Delta t$ and zero otherwise. One can discuss the following limiting cases:

- $\diamond$ $\Delta t > 0$ $t = t_p > 0$. In this case $\phi_A(a \mid t_p + \Delta t)$ gives the age distribution of the mRNAs at time $t_p + \Delta t$ after the start of labeling given that the pulse phase has

been interrupted at time t_p.

- $\diamond$ $\Delta t = 0$ and $t \leq t_p$. In this case $\phi_A(a \mid t)$ gives the age distribution of the mRNAs at time t after the start of pulse.

- $\diamond$ $\Delta t = 0$ and $t_p \to \infty$. In the case

$$\phi_A^{\mathrm{st}}(a) = \lim_{t\to\infty} \phi_A(a \mid t) = \frac{1-G(a)}{\langle U \rangle} \tag{1.28}$$

gives the stationary age distribution of the mRNAs assuming that labeling was active since very long time. A comparison between this result and its numerical estimation is given in Figure 1.3.

- $\diamond$ $\Delta t > 0$ $t_p \to \infty$. In this case

$$\lim_{t\to\infty} \phi_A(a \mid t+\Delta t) = \left[\int_{\Delta t}^{\infty} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \big(1-G(a)\big)$$

gives the age distribution of the mRNAs Δt time units after the beginning of the chase phase, assuming that the pulse duration was so long that all the mRNAs are labeled.

- $\diamond$ $\Delta t = 0$ and $t \ll 1$ leads to $\phi_A(a \mid t) = t^{-1}$ and thus gives an age distribution that is asymptotically flat between 0 and t.

3.5.2 The rest life time distribution

The determination of the distribution of the rest life time L proceeds in a similar fashion. Let $Z = O + U$ be the sum of the time origin and of the life time of a given mRNA and let $t+\Delta t$ the time of observation or measurement after the start of the pulse phase. The rest life time is clearly given by $L = Z - (t+\Delta t)$. We start then by looking at the statistics of Z. We have

$$\begin{aligned}\mathsf{P}\{Z \leq t+\Delta t+x \mid Z > t+\Delta t\} &= \frac{\mathsf{P}\{t+\Delta t < Z \leq t+\Delta t+x\}}{\mathsf{P}\{Z > t+\Delta t\}} \\ &= \left[\int_{\Delta t}^{t+\Delta t} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \left(\int_{\Delta t}^{t+\Delta t} \big(1-G(y)\big)\,\mathrm{d}y - \int_{\Delta t+x}^{t+\Delta t+x} \big(1-G(y)\big)\,\mathrm{d}y\right),\end{aligned}$$

where one can easily verify that this expression satisfies the conditions to be a probability function. Using this result, one can easily find out that the following relationship holds

$$\mathsf{P}\{L \leq r \mid Z > t+\Delta t\} = \mathsf{P}\{Z \leq t+\Delta t+r \mid Z > t+\Delta t\},$$

which leads to the probability density for L upon derivation by r, thus leading to

$$\phi_L(r \mid t+\Delta t) = \left[\int_{\Delta t}^{t+\Delta t} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \big(G(t+\Delta t+r) - G(\Delta t+r)\big),$$

for $r \geq 0$. One can discuss the following limiting cases:

- ◇ $\Delta t > 0$ $t = t_p > 0$. In this case $\phi_L(r \mid t_p+\Delta t)$ gives the rest life time distribution of the mRNAs at time $t_p+\Delta t$ after the start of labeling given that the pulse duration was t_p.

- ◇ $\Delta t = 0$. In this case $\phi_L(r \mid t)$ gives the rest life time distribution of the mRNAs at time t after the start of the pulse phase.

- ◇ $\Delta t = 0$ and $t \to \infty$. In the case

$$\phi_L^{\mathrm{st}}(r) = \lim_{t\to\infty} \phi_L(r \mid t) = \frac{1-G(r)}{\langle U \rangle} \tag{1.29}$$

  gives the stationary rest life time distribution of the mRNAs assuming that the pulse duration was so long to reach steady state. A comparison between the theoretical prediction and its numerical estimation is given in Figure 1.3. One can notice that the stationary age and stationary rest life time distributions are identical.

- ◇ $\Delta t > 0$ and $t \to \infty$. In this case

$$\lim_{t\to\infty} \phi_L(r \mid t+\Delta t) = \left[\int_{\Delta t}^{\infty} \big(1-G(y)\big)\,\mathrm{d}y\right]^{-1} \big(1-G(\Delta t+r)\big),$$

  gives the rest life time distribution of the mRNAs Δt time units after the start of the chase phase assuming that labeling was active since very long time and the number of labeled mRNAs had reached steady state.

◇ $\Delta t = 0$ and $t \to 0$ leads to $\phi_L(r \mid t \to 0) = \phi_U(r)$ as it is intuitively clear since the rest life of a new born mRNA is given by its whole life.

3.5.3 Heuristic derivation of the age and rest life time distributions

At any given moment of observation far away from the initial condition we will detect an mRNA with a probability that is proportional to its life time U. This means that the probability to catch an mRNA of total life time between u and $u + \mathrm{d}u$ is proportional to $u\phi_U(u)\,\mathrm{d}u$. Thus, normalizing accordingly we obtain that this probability is governed by the density $u\phi_U(u)/\langle U\rangle$.

Now, conditional that the life time of an mRNA is $U = u$, both the age and the rest life of this mRNA are uniformly distributed in $[0,u)$ with density $1/u$, namely

$$\phi_A(x \mid U = u) = \phi_L(x \mid U = u) = \frac{1}{u},$$

with $0 \le x < u$. Finally,

$$\phi_A(x) = \phi_L(x) = \int_x^\infty \frac{1}{u}\frac{u\phi_U(u)}{\langle U\rangle}\,\mathrm{d}u = \frac{1-G(x)}{\langle U\rangle},$$

in agreement with the previous calculations.

4 Time dependent transcription rates

In this section we abandon the experimental set-up of pulse and chase and we describe more closely a situation in which the synthesis (or transcription) rate changes with time. This situation is biologically meaningful because the life of a single cell is made of different phases (cell-cycle phases) in which the expression of certain genes is turned on or off. Here we will assume that the changes of the transcription rate follow a known temporal, deterministic pattern. As before, we will describe the distribution of the number of mRNA molecules and other statistical properties of the system under the condition that the transcription rate is not constant. Here, we will consider the simple problem of a deterministic series of changes in transcription rate.

4.1 Two consecutive transcription phases

In the previous sections we have considered always the initial condition in which the number of mRNA chains present at time zero (or at any initial arbitrary time) was zero. This was useful in order to single out the pattern of growth of the number of mRNA chains without the complication of a different initial condition. It is perfectly possible to perform the computation with a different initial condition but it is certainly more convenient to deal with this question according to the following scheme. We start by considering here the case in which at time $t_i^{(1)}$ starts a transcription process with rate $\omega_{\rm tc}^{(1)}$ which is concluded at time $t_s^{(1)}$ followed by a second transcription process starting at time $t_i^{(2)} = t_s^{(1)}$ with rate $\omega_{\rm tc}^{(2)}$ and concluded at time $t_s^{(2)}$. We assume here that the life time distribution ϕ_U is unaffected by these changes and that it is not dependent on the time at which the measurements are made. We also assume that at time $t_i^{(1)}$ there are no mRNA chains of this kind, so that $X(t_i^{(1)}) = 0$.

We consider now a measurement time t at some point in the middle of the second transcription phase. Thus, $t_i^{(2)} < t < t_s^{(2)}$ and therefore $t > t_s^{(1)}$. We want to know what is the distribution of the number of mRNA chains at this time t. The first observation to make is that this distribution has two contributions. The first contribution comes from the mRNAs that originated in the first period and did not degrade before time t. The second contribution comes from the mRNAs that originated in the second period, before time t, and did not degrade before time t. These two contributions are independent of each other and thus the number of mRNA that we observe at time t is made of the sum of two independent random numbers.

The contribution from the mRNAs that originated in the first period leads to a Poisson distribution that was derived in Section 3.3 Eq. (1.24). In the present contest, this is a Poisson distribution with parameter

$$\lambda_1(t) = \omega_{\rm tc}^{(1)} \int_{t-t_s^{(1)}}^{t-t_i^{(1)}} \big(1 - G(u)\big)\, {\rm d}u\,, \tag{1.30}$$

where we have substituted the time intervals in (1.25) in explicit time differences. The contribution arising from the second period, instead, can be derived using the results of Section 3.2 Eq. (1.10). This leads to a Poisson distribution with parameter

$$\lambda_2(t) = \omega_{\rm tc}^{(2)} \int_{0}^{t-t_i^{(2)}} \big(1 - G(u)\big)\, {\rm d}u\,, \tag{1.31}$$

where the lower integration limit is zero because $t < t_s^{(2)}$. Although the present status of the derivation may look discouraging, especially if one thinks about its generalization to more than two periods, here it follows a result that saves the day.

Let X_1 and X_2 two independent Poisson distributed random numbers, whose distributions have parameters λ_1 and λ_2, respectively. Let Y be the random number given by the sum of X_1 and X_2,

$$Y = X_1 + X_2 \,,$$

then the probability distribution of Y is also Poisson with parameter $\lambda_1 + \lambda_2$. The simplest way to see this, is by means of the generating functions. Indeed, the generating function $g_i(s)$ of the variables X_i is given by

$$g_i(s) = \exp\left[-\lambda_i(1-s)\right]$$

and thus the generating function $g(s)$ of their sum is the product of the g_i, given by

$$g(s) = g_1(s)g_2(s) = \exp\left[-(\lambda_1+\lambda_2)(1-s)\right] \,,$$

which concludes the demonstration.

Therefore, the distribution of the number of mRNA at time t at any point in the middle of the second transcription period is a Poisson distribution with parameter

$$\lambda(t) = \omega_{\mathrm{tc}}^{(1)} \int_{t-t_s^{(1)}}^{t-t_i^{(1)}} \left(1-G(u)\right) \mathrm{d}u + \omega_{\mathrm{tc}}^{(2)} \int_0^{t-t_i^{(2)}} \left(1-G(u)\right) \mathrm{d}u \,, \tag{1.32}$$

with the first term from (1.30) and the second from (1.31).

4.1.1 Change of transcription and degradation

The switching point between two different transcription rates can be the time point at which also the degradation changes. This would imply that before the switching point the mRNA are transcribed with rate $\omega_{\mathrm{tc}}^{(1)}$ and have a lifetime probability density $\phi_U^{(1)}$ while after the switching they are generated with a rate $\omega_{\mathrm{tc}}^{(2)}$ and have a lifetime probability density $\phi_U^{(2)}$. Under these conditions, the parameter of the Poisson distribution for the

mRNAs generated after the switching point $t_i^{(2)}$ is given by (1.31) with the modification

$$\lambda_2(t) = \omega_{\rm tc}^{(2)} \int_0^{t-t_i^{(2)}} \big(1 - G_2(u)\big)\, {\rm d}u\,, \tag{1.33}$$

where $G_2(t)$ is the probability function. The fate of the mRNAs generated before the switching, i.e. before time $t_i^{(2)}$ is more complicated and could be characterized in different ways depending on the biological processes of aging. The simplest assumption is that the age accumulated by each mRNA during the first period is carried over or recognized by the environment in the second period as well (for instance, only some of the microscopic rates are changed). Under this scenario, we have that the conditional probability given in Eq. (1.10) must be modified as follows

$$\mathsf{P}\{U \geq t-s\} = \mathsf{P}\{U_1 \geq t_s^{(1)} - s\}\mathsf{P}\{U_2 \geq t-s \,|\, U_2 \geq t_i^{(2)} - s\}\,,$$

where the conditional probability in the rhs expresses the fact that also according to the degradation process that determines U_2 the mRNAs have the same age as under the process that determines U_1. Notice that in all these equations the identity $t_s^{(1)} \equiv t_i^{(2)}$ holds. Thus, now one obtains

$$\lambda_1(t) = \omega_{\rm tc}^{(1)} \int_{t-t_s^{(1)}}^{t-t_i^{(1)}} \frac{1 - G_1\big(t_s^{(1)} - s\big)}{1 - G_2\big(t_i^{(2)} - s\big)} \big(1 - G_2(t-s)\big)\, {\rm d}s\,, \tag{1.34}$$

which obeys the obvious property that it goes back to (1.30) when the two degradation processes are identical. Finally, at any time $t > t_i^{(2)}$ the number of mRNA is distributed according to a Poisson distribution with parameter

$$\lambda(t) = \lambda_1(t) + \lambda_2(t)\,,$$

given by (1.34) and (1.33).

4.2 A long series of transcription phases

It is not difficult to generalize the result derived in (1.32) to accomplish a row of $n+1$ such phases, each characterized by its own transcription rate $\omega_{\rm tc}^{(j)}$. Indeed, also the sum

on many independent Poisson distributed random numbers is Poisson distributed with parameter given by the sum of the parameters.

Thus let us assume that there are $n+1$ phases. Each phase j starts at time $t_i^{(j)}$ and terminates at time $t_s^{(j)}$ with transcription rate $\omega_{\rm tc}^{(j)}$. We make a measurement at time t such that it is in the middle of the $n+1$ phase, namely $t_i^{(n+1)} < t < t_s^{(n+1)}$. Thus, the distribution of the number of mRNA molecules at time t, will be a Poisson distribution with parameter

$$\lambda(t) = \sum_{j=1}^{n} \omega_{\rm tc}^{(j)} \int_{t-t_s^{(j)}}^{t-t_i^{(j)}} \big(1-G(u)\big)\,{\rm d}u + \omega_{\rm tc}^{(n+1)} \int_0^{t-t_i^{(n+1)}} \big(1-G(u)\big)\,{\rm d}u\,, \qquad (1.35)$$

which in the limit of large n would give a distribution that does not depend on the initial conditions at time $t_i^{(1)}$ and where we still assume that the sequence of all $\omega_{\rm tc}^{(j)}$ is known and deterministic.

Now we notice that we can rewrite (1.35) by using the following definition. Let

$$\omega_{\rm tc}(\tau) = \omega_{\rm tc}^{(j)}\,,$$

if $t_i^{(j)} \le \tau < t_s^{(j)}$. Then from (1.35) we obtain

$$\lambda(t) = \sum_{j=1}^{n} \int_{t-t_s^{(j)}}^{t-t_i^{(j)}} \omega_{\rm tc}(t-u)\big(1-G(u)\big)\,{\rm d}u + \int_0^{t-t_i^{(n+1)}} \omega_{\rm tc}(t-u)\big(1-G(u)\big)\,{\rm d}u\,, \quad (1.36)$$

which then, taking into consideration that $t_i^{(j+1)} = t_s^{(j)}$ leads to

$$\lambda(t) = \int_0^{t-t_i^{(1)}} \omega_{\rm tc}(t-u)\big(1-G(u)\big)\,{\rm d}u\,. \qquad (1.37)$$

If we, without loss of generality, arbitrarily set now $t_i^{(1)} = 0$ and apply a rather intuitive change of variable, we obtain that the distribution of the number of mRNA during a process with any arbitrary number of transcription phases can be expressed as a Poisson distribution with parameter

$$\lambda(t) = \int_0^{t} \omega_{\rm tc}(u)\big(1-G(t-u)\big)\,{\rm d}u\,, \qquad (1.38)$$

where we see that the major contribution arises from the recent events, close to t, if we remind that $\big(1-G(x)\big) \to 0$ as $x \to \infty$. For the present and future reference we write then the full form of the distribution as

$$\mathsf{P}\{R(t) = k \mid X(0) = 0,\, \omega_{\text{tc}}(u)\; 0 \le u < t\} = \frac{[\lambda(t)]^k \exp\big(-\lambda(t)\big)}{k!}\,, \tag{1.39}$$

with $\lambda(t)$ given in (1.38). We shall notice that until one does not know the complete behavior of $\omega_{\text{tc}}(t)$ for large t it is not possible to discuss the question of the stationary distribution.

• • • • • • • • • •

5 Conclusions

The previous sections have provided a general framework to describe any kind of non-exponential decay pattern. We have found out that as long as the synthesis rate is either constant or varies deterministically in time, the distribution of the number of molecules in the cell is a time-dependent Poisson distribution. Nevertheless, when dealing with experimental data we would like to find out some properties associated to the molecule that has been measured, such as the mean lifetime. The framework outlined above, thus, has to be complemented with a parametric way to obtain the lifetime distribution of the molecules.

An important hint in this direction comes from biochemistry. The degradation mechanisms are characterized by the interaction of the target molecule with other molecules, complexes or enzymes present in the cell environment. For instance, in miRNA mediated degradation, the target mRNA is first recruited by a complex made of miRNA and other molecules. Afterwards, once the miRNA and its partners are bound to the target mRNA, more enzymes are recruited that start modifying the structure of the mRNA and start degrading the mRNA one nucleotide per step. In general, thus, the degradation process can be described at the single molecule level as a network of biochemical interactions. In this network, the vertices are biochemical state of the target mRNA. One or more of these states then lead to the degradation, which is obviously irreversible. If we take this idea, we can generalize it by describing the target molecule as visiting a network of states during its life until a final absorbing state is reached, which is the complete degradation [15]. For our purposes, a molecule is degraded when it is no longer detectable by the

experimental technique. In an even more abstract step, we assume that it is possible to describe the single-molecule process of degradation as a continuous time Markov chain which has a well defined initial condition (the newly synthesized mRNA) and has one well defined absorbing state (the degraded state). In this way, the life time probability density $\phi_U(t)$ is just the probability density of the absorption time on this network.

When the details of the biochemical process of degradation are known, the network of states, or at least its starting structure, can be taken from it. In that case, the data are used to estimate the parameters (the transition rates) of the network and the analysis can be used to validate the hypothesized biochemical process [15]. When the biochemical network is unknown, the network of states has to be chosen among the most parsimonious networks that are able to fit the data in a satisfactorily manner. Here, the analysis provides important kinematic properties of the degradation process and some hint to further experimental clarification of the underlying biochemistry.

Alternatively, one can define an age dependent degradation rate by using results from hazard rate theory [8], whenever there is a mechanistic model of degradation that cannot be cast into the framework of a stochastic process with a finite number of discrete states.

• • • • • • • • • • • •

Acknowledgments

I would like to thank all those people who contributed to the understanding of this topic, both from the theoretical and from the biological point of view, by providing fruitful contributions, useful criticisms, and inspiring comments: J. Belasco, G. Blanchard, D. Chiarugi, C. Deneke, Z. Ignatova, R. Lipowsky, E. McShane, J. Pedraza, S. Rœlly, S. Rudorf, M. Selbach, C. Sin, C. Vogel.

• • • • • • • • •

Bibliography

[1] Gillespie, J. H.: *Population Genetics: A concise guide*. John Hopkins University Press (2004).

[2] Ewens, W. J.: *Mathematical population genetics: I. Theoretical introduction*. Springer, (2004).

[3] Sin, C., Chiarugi, D., Valleriani, A.: *Quantitative assessment of ribosome drop-off in E. coli*, Nucl. Acid Res. **44**(6), 2528–2537 (2016).

[4] Braun, J. E., Huntzinger, E., Fauser, M., Izaurralde, E.: *GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets*, Mol. Cell. **44**(1), 120–133 (2011).

[5] Huntzinger, E., Izaurralde, E.: *Gene silencing by microRNAs: Contributions of translational repression and mRNA decay*, Nat. Rev. Gen. **12**(2), 99–110 (2011).

[6] Belasco, J. G.: *All things must pass: Contrasts and commonalities in eukaryotic and bacterial mRNA decay*, Mat. Rev. Mol. Cell Biol. **11**, 467–478 (2010).

[7] Deneke, C., Rudorf, S., Valleriani, A.: *Transient phenomena in gene expression after induction of transcription*, PLoS one **7**(4), e35044 (2012).

[8] Deneke, C.: *Theory of mRNA degradation*, PhD Thesis University of Potsdam (2012).

[9] Deneke, C., Lipowsky, R., Valleriani, A.: *Complex degradation processes lead to non-exponential decay patterns and age-dependent decay rates of messenger RNA*, PLoS one **8**(2), e55442 (2013).

[10] Deneke, C., Lipowsky, R., Valleriani, A.: *Effect of ribosome shielding on mRNA stability*, Phys. Biol. **10**, 046008 (2013).

[11] Wheatley, D. N., Giddings, M. R., Inglis, M. S.: *Kinetics of degradation of "short-" and "long-lived" proteins in cultured mammalian cells*, Cell Biol. Int. Rep. **4**(12), 1081–1090 (1980).

[12] McShane, E., Sin, C., Zauber, H., Wells, J. N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchova, Z., Marsh, J.A., Valleriani, A., Selbach, M.: *Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation*, Cell **167**, 803–815 (2016).

[13] Sin, C., Chiarugi, D., Valleriani, A.: *Degradation parameters from pulse-chase experiments*, PLoS one **11**(5), e0155028 (2016).

[14] Taylor, H. M., Karlin, S.: *An introduction to stochastic modeling*. Academic Press, San Diego (1998).

[15] Sin, C., Chiarugi, D., Valleriani, A.: *Single-molecule modeling of mRNA degradation by miRNA: Lessons from data*, BMC Sys. Biol. **9**(S3), S2 (2015).