Theory of mRNA degradation

Dissertation

zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin theoretische Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

von

Carlus Deneke

Potsdam, Mai 2012

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2012/6199/
URN urn:nbn:de:kobv:517-opus-61998
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-61998

Contents

Abstract

One of the central themes of biology is to understand how individual cells achieve a high fidelity in gene expression. Each cell needs to ensure accurate protein levels for its proper functioning and its capability to proliferate. Therefore, complex regulatory mechanisms have evolved in order to render the expression of each gene dependent on the expression level of (all) other genes. Regulation can occur at different stages within the framework of the central dogma of molecular biology. One very effective and relatively direct mechanism concerns the regulation of the stability of mRNAs. All organisms have evolved diverse and powerful mechanisms to achieve this. In order to better comprehend the regulation in living cells, biochemists have studied specific degradation mechanisms in detail. In addition to that, modern high-throughput techniques allow to obtain quantitative data on a global scale by parallel analysis of the decay patterns of many different mRNAs from different genes.

In previous studies, the interpretation of these mRNA decay experiments relied on a simple theoretical description based on an exponential decay. However, this does not account for the complexity of the responsible mechanisms and, as a consequence, the exponential decay is often not in agreement with the experimental decay patterns.

We have developed an improved and more general theory of mRNA degradation which provides a general framework of mRNA expression and allows describing specific degradation mechanisms. We have made an attempt to provide detailed models for the regulation in different organisms. In the yeast *S. cerevisiae*, different degradation pathways are known to compete and furthermore most of them rely on the biochemical modification of mRNA molecules. In bacteria such as *E. coli*, degradation proceeds primarily endonucleolytically, *i.e.* it is governed by the initial cleavage within the coding region. In addition, it is often coupled to the level of maturity and the size of the polysome of an mRNA. Both for *S. cerevisiae* and *E. coli*, our descriptions lead to a considerable improvement of the interpretation of experimental data. The general outcome is that the degradation of mRNA must be described by an age-dependent degradation rate, which can be interpreted as a consequence of molecular aging of mRNAs. Within our theory, we find adequate ways to address this much debated topic from a theoretical perspective.

The improvements of the understanding of mRNA degradation can be readily applied to further comprehend the mRNA expression under different internal or environmental conditions such as after the induction of transcription or stress application. Also, the role of mRNA decay can be assessed in the context of translation and protein synthesis.

The ultimate goal in understanding gene regulation mediated by mRNA stability will be to identify the relevance and biological function of different mechanisms. Once more quantitative data will become available, our description allows to elaborate the role of each mechanism by devising a suitable model.

1. Introduction

The central dogma of molecular biology describes the core processes in each cell [1, 2]. The genetic information is stored in DNA. The ultimate (and necessary) goal for the viability of each cell is having the optimum level of each protein. Therefore, each gene is transcribed into messenger RNA (mRNA) and proteins are produced from mRNAs in the process of translation. However, both mRNAs and proteins are subject to degradation and dilution, such that the molecules have to be constantly renewed. There are four principal rates that regulate the amount of proteins in a cell: The rates of transcription and translation and the mRNA and protein degradation rates (see figure 1.1). These rates can however be different for each gene and depend also on the physiological state of the cell. Furthermore, many proteins and ribonucleic acids interact such that the presence of one molecule regulates the amount of the others. The cell's biochemistry has evolved specific pathways of regulation. There are enzymes that control the transcription, others stabilize or destabilize mRNAs and also at the level of translation and protein degradation, the corresponding rates are governed by biochemical interactions. The rates themselves are then determined by the concentration of the regulating enzymes, hence, diffusion (or active transport) plays a central role. These are central ideas of systems biology - to view the complex cascades of reactions as a complex network of interactions between all the individual entities. In systems biology one aims at systematically analyzing the pathways, for example by specifically isolating a subset of reactions or by observing the response to an external stimulus. The subsequent response of the network can be quantitatively analyzed and compared to the expected changes from a theoretical model.
Gene expression is fundamentally governed by noise. The noise arises due to the inherent stochasticity of all processes involved in gene expression. One distinguishes between intrinsic noise which is due to the random order of chemical reactions occurring on the

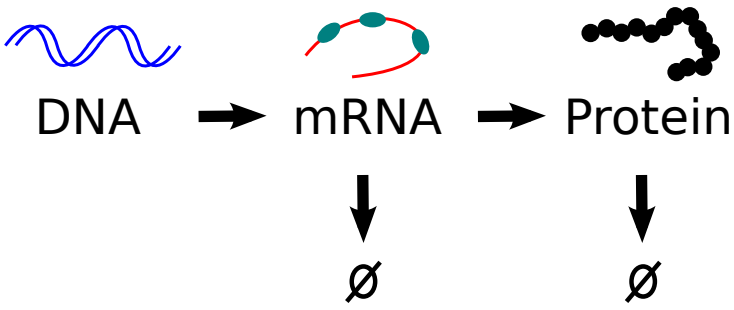

Figure 1.1.: **Central dogma of molecular biology**. Genes are transcribed from DNA to messenger RNA (mRNA). Each mRNA is in turn translated into proteins. Regulation of gene expression can proceed by adjusting the rates of transcription and translation, respectively. Both mRNAs and proteins are actively degraded in cells thus providing another type of efficient mechanisms for regulation.

molecular scale and extrinsic noise which arises from the changing concentrations of enzymes involved in gene expression [3, 4]. Recently, the different sources of noise could be determined experimentally [5].

Regulation of gene expression

Cells of all organisms use regulation of gene expression at all stages to obtain an optimal protein level for their viability and proliferation [2, 6]. At the level of transcription, the adjustment of the transcription rate of a gene is a powerful regulatory mechanism [7, 8, 9]. Recently, much research was devoted to post-transcriptional mechanisms of regulation. At least in eukaryotes, a cellular machinery for RNA processing control exists, *i.e.* regulation via editing (*i.e. splicing*) the *precursor-mRNAs* [10] or by controlling the mRNA export to the cytoplasm [11]. In addition, regulation occurs on the level of translation [12] - for instance by regulating translation initiation by the number of available ribosomes and the specific action of initiation complexes or by regulating translation elongation via the availability of *transfer RNAs* (tRNAs). At the end of the central dogma, the cell can directly control the protein amount by activating proteases or by phosphorylating the proteins and thus changing their enzymatic activity [13].
As another possibility, the number of proteins per mRNA can be changed by altering the stability of mRNA. More stable mRNAs will produce more proteins and a decrease of stability quickly reduces the production of new proteins and saves the metabolic cost of producing them.
In summary, for all regulation mechanisms time and metabolic cost are the crucial factors and all have their respective benefits and disadvantages. Research has addressed all areas of regulation, although only recently the role of regulation at the level of mRNA stability became a major focus of research. In the following, we will sketch some biochemical mechanisms of mRNA degradation. This part is followed by a description of experimental methods and current theoretical models important for mRNA turnover.

Biochemical mechanisms of mRNA degradation

In the past decades, a number of different mechanisms responsible for the degradation of the mRNA have been unveiled [14, 15, 16]. Some mechanisms of degradation are known to affect the decay of all mRNA species and are thus unspecific. On the contrary, other mechanisms are known to affect certain mRNAs more than others depending on different physical and chemical characteristics of the nucleotide chain. For example, micro-RNAs mediate the docking of the degrading enzymes specifically for each mRNA and contribute thus to the large variation of the stability between mRNA species [17, 18, 19, 20].
In all organisms, one can distinguish between two types of degradation mechanisms, the *endonucleolytic* and the *exonucleolytic* degradation pathways. In the endonucleolytic pathways, degradation is initiated by cleavage within the coding region of an mRNA. Once the degradation process has been initiated, it leads to a rapid decay of the attacked mRNA with a sudden interruption of the translation process. In this case, the time scale related to the random encounter between the degradation complex and the mRNA primarily determines the lifetime of the mRNAs. It is commonly believed that eukaryotic

mRNAs are affected to a lesser extent by endonucleolytic degradation than prokaryotic mRNAs [14].
In the exonucleolytic degradation pathways, mRNAs are degraded processively from one end to the other. Here, typically a set of biochemical and physical modifications to the untranslated regions is required for the initiation of degradation. For instance, the *decapping* mechanism, which is thought to be the main degradation pathway in eukaryotic cells, requires deadenylation at the $3'$ region and the destabilization of the $5'$ cap structure before degradation occurs in the $5'$ to $3'$ direction behind the last translating ribosome [21]. Also in bacteria different exonucleolytic degradation pathways exist. For instance in *E. coli*, modification of the $3'$ stem-loop is a prerequisite of exonucleolytic degradation initiation [14]. Moreover, in *B. subtilis* a $5'$ exonuclease has been discovered recently [22, 23]. In summary, in all organisms there are different degradation pathways leading to the destabilization of mRNA. However, the precise contribution of each mechanism to the destabilization of an mRNA remains unknown

Experimental procedures to assess mRNA stability

The standard approach to quantitatively assess mRNA stability consists of impeding the further transcription of a certain gene (or genome-wide) in a populations of cells [24, 25, 26]. This proceeds for instance via the incorporation of a drug that inhibits the initiation or elongation of *RNA polymerases* or by heating up a probe containing heat-sensitive RNA polymerases. Consequently, the subsequent decay of the previously synthesized mRNAs is observable and one can take samples at different points in time after the interruption of transcription. For each sample, the amount of mRNA of a specific gene can by assessed by *northern blotting*. Furthermore, high-throughput techniques such as *microarray*, *qRT-PCR* and *RNA-Seq* allow the simultaneous detection of the mRNA level of all genes separately. One implicit assumption of this method is that the incorporation of the drug occurs sufficiently fast and parallel in all cells of a cell culture. Moreover, the stop of transcription has a strong impact on the entire metabolism of the cell and one might only observe the reaction of the cell due this extreme stress condition. However, instead of blocking transcription one can also introduce heavy or radioactive nucleotides into the cell (*metabolic labeling*). Henceforth, one can distinguish between mRNAs synthesized before the label introduction (*i.e.* mRNAs free of modified nucleotides) and newly synthesized ones. One can now also follow the fraction of previously synthesized mRNAs over time. Also with this method side effects might arise and it remains a challenge to evaluate the quality of each procedure and the associated detection methods.

The current paradigm of mRNA degradation

From the theoretical point of view, the simplest model is to assume that for each gene there is a constant mRNA synthesis and a constant mRNA decay rate (in analogy to radioactive decay). The transcription is zeroth-order, *i.e.* it only depends on the rate, whereas the degradation is first-order, *i.e.* proportional to the amount of mRNA. Mathematically, the mRNA number can be obtained via a Master equation (see chapter 2). According to these assumptions, it follows that after the stop of transcription the decay pattern follows

an exponential decrease (or a straight line on a linear-log scale). It is here that theory and experiment meet and one can fit the theoretical decay function to the experimental data to obtain the decay rate, the mean lifetime and the half-life of an mRNA of a given species.
In many experiments in different organisms it was found that large amounts of the decay patterns are not exponential. One example is given in figure 1.2 A for *S. cerevisiae* (a yeast model organism) [27]. There, only 11 out of 424 (selected) mRNAs obey an exponential decay (as will become more clear in chapter 3 we select only bona-fide curves, all other allow no interpretation). Similarly in *E. coli* only 11 out of 103 and in the marine cyanobacterium *Prochlorococcus* 117 out of 1102 genes resemble an exponential decay (our analysis, experiments from [28, 29]). It is therefore clear that this simple first-order kinetic model does not suffice to describe the experimental situation in detail. Moreover, the assumption of a single, constant rate contradicts the detailed knowledge of the degradation process. Whenever a series of modifications are required for degradation, each process contributes with its specific reaction rate. The assumption of a single rate implies that either all reaction rates are the same (hence also the concentrations of all participating enzymes and their time scales of catalytic activity) or one rates dominates strongly over all other rates (*i.e.* it is much smaller than all other rates). Hence, a description with a single rate constant seems inappropriate in light of the knowledge of the mechanisms leading to degradation.
As a final remark, the assumption of a constant decay rate throughout the lifetime of an mRNA is also counterintuitive. As we will point out in detail in chapter 3, this implies that the molecules do not age. To see this, consider the case that one had hypothetically taken a subsample of only the oldest mRNAs in a population. Under the assumption of a constant decay rate, it would turn out that these mRNAs had the same life expectancy as the entire mRNA population. Though this holds for radioactive decay (the decay constant does presumably not change over billions of years, nor do environmental conditions afflict the stability of a single radionuclide), it is difficult to preserve this idea for molecular reactions in living matter.
Previously, when a large number of non-exponential decay patterns was noticed, the problem was addressed by introducing heuristic fitting functions such as a double exponential [28, 29]. While this might be suitable to decrease the fitting error, it does not establish a link to the mechanisms of degradation. Therefore, we seek a mathematical description that expands the current description, that improves the analysis of the experiments and that allows to link the stability of an mRNA as seen in decay experiments with the underlying mechanisms of its degradation.

1.1. Overview of thesis

In this thesis, we elaborate a theory on mRNA stability and its implications on mRNA and protein turnover. We depart from the classical method of a Master equation to put it in a broader, stochastic framework. Thereby, we expand the possibilities to understand and analyze experiments on mRNA decay. Also, we establish a link between the biochemical mechanisms and decay experiments. Furthermore, this thesis provides a method to plan and analyze more detailed mRNA decay experiments.

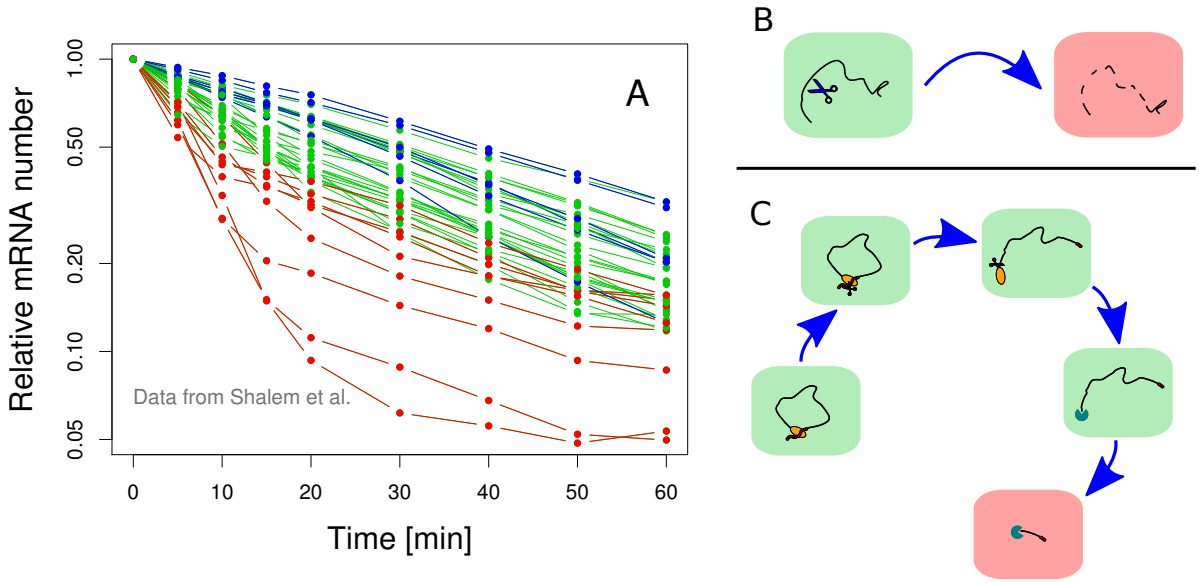

Figure 1.2.: **Non-exponential decay and multi-step mechanisms**. Panel A: A selection of experimental mRNA decay patterns in the yeast *S. cerevisiae* (plotted in log-scale) reveal that the majority of patterns does not follow an exponential decay curve (adapted from [27]). Furthermore, different biochemical pathways of degradation of mRNAs have been discovered. In panel B degradation is depicted as a relatively simple process determined by only a single step, e.g. by unspecific and fast endonucleolytic single-step decay. In panel C, instead, we show a schematic representation of the degradation pathway known as *decapping* which is one of the main degradation mechanisms in eukaryotic cells. The decapping mechanism consists of several biochemical steps, which contribute to destabilize the mRNA until complete degradation takes place, and can be considered as a prototype of multi-step degradation.

In chapter 2 we lay the theoretical foundations of this thesis and derive a general result. Our theory is based on a constant mRNA transcription rate (which corresponds to a Poisson process) and arbitrary lifetime probabilities. We obtain the distribution of the number of mRNAs as well as the age and residual lifetime distribution for all times before and after the interruption of the mRNA transcription process. In the same chapter we introduce further concepts and methods that will be used in the remainder of the thesis. In chapter 3 we will use the general results obtained in chapter 2 to study mRNA decay as measured in two model organisms. Therefore, we develop a concrete model for each of the two organisms. Thereby, we improve the analysis of the experiments but also envision how our theory can be applied for present and future experimental studies on mRNA decay. Furthermore, in this chapter we derive general properties of the decay pattern that hold for any experiment. In chapter 4 we elaborate the role of mRNA degradation in the turnover of mRNA and its transient nature. On the one hand, the results of this chapter are based on the general theory derived in chapter 2. On the other hand, we make use of the knowledge about the degradation process acquired in chapter 3. In the same chapter we analyze an experiment on stress response in *S. cerevisiae* where all methods introduced by then come to application. Chapter 5 puts our theory into a broader setting. Here we study how the transient mRNA turnover afflicts translation and protein synthesis and we discuss implications of the mRNA turnover for the cell cycle. Finally, in chapter 6 we summarize the main findings and give an outlook on theoretical challenges that lie

ahead and where the main results of this thesis could become useful in a broader context. Further details to all chapters and recent lines of research are discussed in the appendix.

Published articles

During the thesis work, two articles were submitted so far to peer-reviewed journals:

1. **Transient Phenomena in Gene Expression after Induction of Transcription**. *PLoS ONE* 7(4): e35044. This article deals with the transient phenomena in mRNA turnover and protein synthesis. This article forms parts of chapters 4 and 5.

2. **Decay patterns and aging of messenger RNA**. (*Under review*). This article introduces the general theory of mRNA decay as presented in chapters 2 and 3. Also, a detailed model was introduced and compared to experimental data on the yeast *S. cerevisiae* which is presented in chapter 3.

2. Mathematical background

In the present chapter, we develop the central mathematical concepts for the description of the mRNA turnover. We begin by discussing the fundamental time scales that are involved in the process. Next, we argue why the transcription can be modeled via a Poisson process. This is followed by the derivation of the stochastic evolution of the number of mRNAs for arbitrary mRNA lifetime distributions (section 2.1.3 and 2.1.4). This result will be of central importance for the entire thesis. Similarly, in section 2.1.5, we find the distributions for the age and residual lifetime of an mRNA. It follows a comparison of our general theoretical results to the standard approach with a Master equation (section 2.2). We also derive the relation between general lifetime distributions and age-dependent rates (section 2.3). The final sections introduce general methods that will prove useful throughout this thesis. We introduce continuous-time Markov chains (section 2.4) and simulation techniques that provide a complementary tool for studying mRNA and protein expression (section 2.5).

2.1. The fluctuation of the mRNA number

Our mathematical description is build upon general concepts of stochastic modeling as introduced in ref. [30]. In this section, we develop the mathematical description for the number of mRNAs as a function of time. It fundamentally depends on two points in time, the start of the transcription process and the stop or inhibition of transcription. Central to our theory is the concept of the random lifetime of an mRNA which follows a statistical distribution.
Furthermore, at any instant a given mRNA has an age and a residual lifetime. We will see how these quantities are statistically related to the lifetime of an mRNA.

2.1.1. Introduction of time scales

Many of the variables that we need in the following can be introduced via a simple visualization of the mRNA turnover process (see figure 2.1). Starting at a time $t = 0$, transcription events occur at random time points. The total number of mRNAs generated until a time t is denoted by the random variable $X(t)$. Each of the transcripts has a random lifetime U, where U is distributed according to the lifetime density ϕ_U. At each point in time a random number of mRNAs is still alive, *i.e.* not degraded. We will denote this quantity by $Y(t)$, and by definition $Y(t) \leq X(t)$ holds for all times t.
For the purpose of studying mRNA decay experiments, we are also interested in the number of mRNAs at a time interval Δt after the interruption of transcription at $t = t_s$. Therefore, we will perform the derivation of the distribution of $Y(t)$ for two different cases:

before the interruption of transcription, *i.e.* $t \leq t_s$, and after the stop of transcription, *i.e.* $t = t_s + \Delta t$.
The time t is a global time and can be seen as a laboratory time scale. Without loss of generality, we set the start of the counting process at $t = 0$ for convenience - note however that one could consider $t = -\infty$ instead. Nevertheless, t can denote also the time elapsed since the start of a physical process such as for example the induction of transcription of a certain gene (or all genes) via an external stimulus. Therefore, we might refer to the time t also as the *time after the start of transcription*.
As mentioned earlier, transcription is often globally stopped via the addition of a drug or other external stimuli. The time elapsed Δt is also a laboratory time scale and it will be referred to mostly as time after *stop of transcription*.
However, both t and Δt can have a broader connotation and can also be interpreted as the time elapsed since an internal signal or the introduction of a label.
Here, we have introduced the terms global or laboratory time scales to emphasize the difference to local or molecular time scales. The latter concern times that are related to the lifetime of an individual mRNA. For instance, at any specific point in time each mRNA has always a well-defined age. According to each age it also has an excess or residual lifetime. Moreover, during the lifetime of an individual mRNA, the transcript might undergo biochemical transformations. To make this more clear, we will denote the molecular time scale as the age of an mRNA.

Definition of lifetime We have already mentioned that the lifetime of a single mRNA is a random variable for each individual mRNA whose distribution is given by the lifetime density ϕ_U. In this paragraph, we seek to obtain a biological definition of the lifetime. Generally, one can distinguish between a *chemical* and a *functional* lifetime.
During the chemical lifetime of an mRNA, the sequence of nucleotides is fully intact. Conversely, functional lifetimes describe the time interval during which an mRNA is functional, *i.e.* during which translation is active. While the first definition corresponds to the quantitative detection in biochemical assays, the second is relevant in protein synthesis.
Both possible approaches are however not unambiguous. For a chemically intact mRNA, the detection via a microarray experiment depends on the choice of the oligonucleotides for hybridization. Additionally, in eukaryotes *precursor* mRNAs (*pre-mRNAs*) are synthesized from DNA which still undergo sequence alterations such as intron splicing and untranslated region (UTR) modification. Hence, the time point of generation of a *mature* mRNA does not coincide with the time point of the transcription of a pre-mRNA. For simplicity, in the following we will, however, term the rate of maturation and delivery to the cytosol also as the transcription rate.
For the functional lifetimes, the lifetime can be defined to range from the first ribosome bound until the unbinding of the last ribosome. Alternatively, the functional lifetime can encompass the time interval of active translation initiation, elongation or termination. Furthermore, in eukaryotic cells translation can be temporarily inactivated and resumed at a later time point.
Therefore, it depends always on the specific aim which definition is more appropriate. Additionally, different organisms and detection methods play an important role. Never-

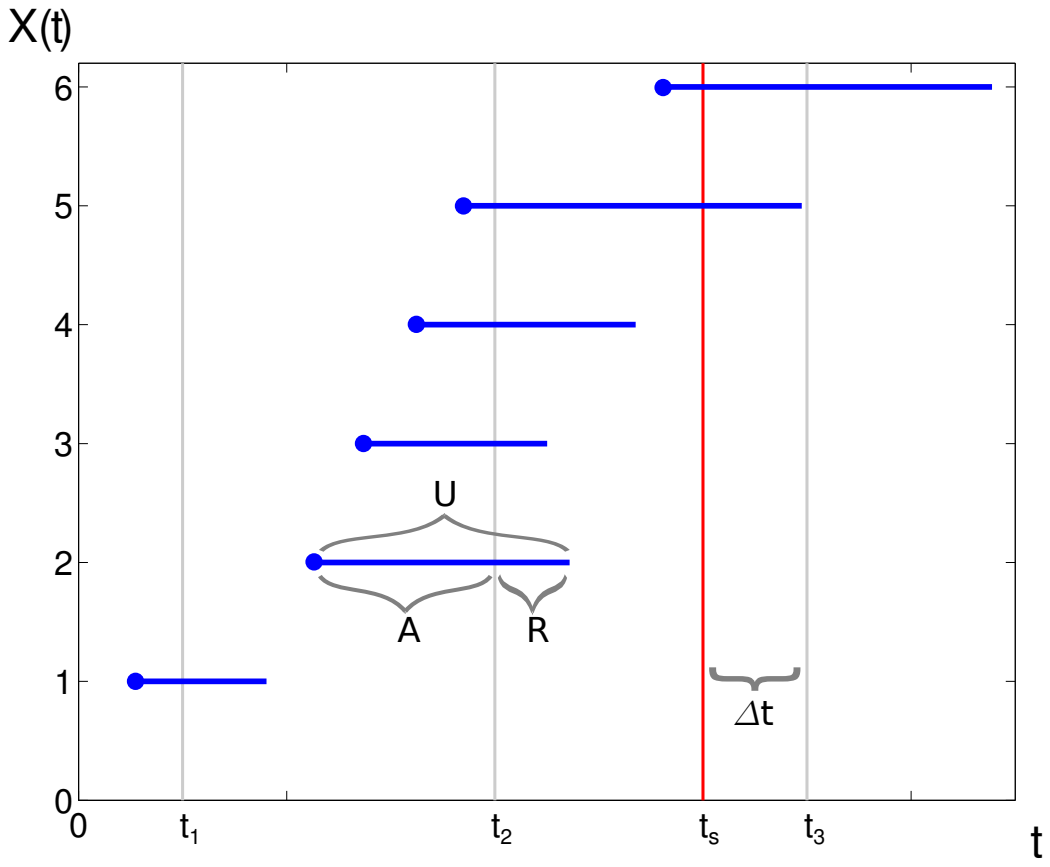

Figure 2.1.: **Illustration of the counting process and definition of time scales.** The function $X(t)$ counts the total number of mRNAs generated in the interval $(0, t]$ according to a Poisson process. Each generation event is symbolized by a blue dot. Additionally, every mRNA has a random lifetime U according to the probability distribution Φ_U, defined in Eq. (2.2), which is symbolized by the length of the blue bars. We want to compute the number of live mRNAs, $Y(t)$, at each time point t. In this visualization, it is given by the number of intersecting mRNA bars with a vertical line at a given t. For instance, $Y(t_1) = 1$, $Y(t_2) = 4$ and $Y(t_3) = 1$. At a definite time $t = t_s$ transcription is interrupted. Hence, $X(t > t_s)$ remains constant and $Y(t > t_s)$ decreases with the time elapsed since the interruption, denoted by Δt. The mRNA highlighted in green possesses a lifetime of u. Furthermore, a (hypothetical) measurement at time t_2 would reveal that at that point in time the mRNA has an age a and a residual lifetime r. This cartoon represents one realization of the counting process - in our computation $X(t)$, $Y(t)$, A, U and R are random numbers and we calculate the corresponding marginal probability distributions.

theless, both definitions need not yield largely different values of a lifetime. For a given mRNA species in a particular organisms, the different definitions should give rise to similar values of the lifetime. Rather, differences between different mRNA species should be manifest in any definition. We will come back to discuss functional lifetimes more specifically in chapter 3 where we will discuss an alternative approach to measure the functional lifetime which goes back to ref. [31].

2.1.2. Transcription modeled as a Poisson process

The process of transcription is a random process determined by the underlying biochemical reactions. Furthermore, also the enzymes involved in transcription, in particular the *RNA polymerase* and *transcription factors* have a low abundance and fluctuate [32, 4]. This gives rise to intrinsic and extrinsic noise [4, 5]. However, we will restrict ourselves to assume a Poisson process for the transcription of each gene [33].

Poisson distribution Let X be the random number of mRNAs generated since a time $t = 0$. Our initial condition is $X(0) = 0$, *i.e.* the transcription process started at $t = 0$, or mRNAs transcribed priorly are not counted. According to a Poisson process the probability of generating k mRNAs until time t reads

$$\Pr\{X(t) = k \mid X(0) = 0\} = \frac{(\omega_{\mathrm{tc}} t)^k \exp(-\omega_{\mathrm{tc}} t)}{k!} \,. \tag{2.1}$$

The parameter of the distribution is $\omega_{\mathrm{tc}}\, t$ where ω_{tc} denotes the transcription rate for that gene. More precisely, it denotes the rate of delivery of the mRNA to the cytosol.
Thus, the total number of transcribed mRNAs is growing with time according to a Poisson distribution with a time dependent parameter $\omega_{\mathrm{tc}}\, t$. The property of the Poisson distribution is that its variance equals the mean value which is the only parameter of the distribution. That means that the Fano factor of a Poisson distribution, defined as the ratio of variance and mean value, is precisely 1. The time between consecutive events is exponentially distributed. Hence, the Poisson process is memoryless, it carries no memory about past occurrences. However, conditioned on a fixed number of total events in $(0, t]$, the origination time of each event is uniformly distributed in $(0, t]$. The last property will be important for the subsequent computation and is derived in the appendix (section A.1).

Single-cells Although the Poisson process is a very powerful mathematical tool that is conventionally applied to model gene expression, it is not known to what extend it accounts for all processes during transcription in single cells. In particular, though transcription events might be infrequent, they might not be independent from another. Fluctuations in the number of available RNA polymerases lead to a fluctuation of the transcription rate. Moreover, different promotor states lead to a random modulation of the transcription rate. Consequently, mRNAs are transcribed in bursts during the active state of the promotor. Furthermore, at different stages of the cell cycle, the transcription rate might not be constant. In these case, one needs to extend the description to a non-homogeneous Poisson process to obtain a more detailed description of the transcription process, see appendix D.
Experiments with single cell resolution could measure the cell-to-cell variation in the expression of mRNAs. In ref. [34], the distribution of the mRNA Fano factor (*i.e.* the ratio of the variance to the mean) over all genes in *E. coli* centered around 1.6 - indicating also non-homogeneous mRNA synthesis production. However, some genes display a Fano factor of one which indicates that these could be described by a Poisson process. In ref. [35] it was found that transcription proceeds via single transcription-initiation events

that are clearly separated in time, rather than by transcriptional bursts. With a more detailed method, no correlation between individual transcripts was found for the construct $POL1p-GLT1$ [36] in *S. cerevisiae*. Hence, the role of different transcription modes on a single-cell level remains open.

Cell populations In spite of this, many experiments investigate different processes of gene expression on the level of cell populations. Typically, one performs a bulk experiment on a large number of unsynchronized cells, ranging from 10^9 up to more than 10^{12} cells. Usually, the total number of cells N_{cells} is held constant by diluting the cell culture and cells of different age and in different stages of the cell cycle are present. Thus, the transcription rate $\omega_i(t)$ of each cell i can adopt a different value and fluctuate in time in a non-synchronal way. The *population* transcription rate is given by $\Omega_{\text{tc}}(t) = \sum_i \omega_i(t)$. If N_{cells} is large, fluctuations of $\omega_i(t)$ will average out and Ω_{tc} will be constant. Therefore, on the scale of a large cell population, the process of transcription is a Poisson process with constant rate Ω_{tc}. In the following, we will use the average rate $\omega_{\text{tc}} = \Omega_{\text{tc}}/N_{\text{cells}}$ to describe the process of generation of new mRNAs in an average cell (of a large population of cells). To some extent, a large cell population resembles a giant cell with many independent sites of transcription that each add up to the global transcription rate Ω_{tc}.
An even more general analytical description including random fluctuations of the transcription rate goes beyond the scope of this work. In this thesis, we concentrate on an improved theoretical description of the process of mRNA degradation. Nevertheless, effects of a varying transcription rate are briefly discussed in chapters 4 and D.

2.1.3. mRNA number after start of transcription

To investigate the mRNA turnover, we have to take the lifetime of the mRNAs into account. Each mRNA of a given gene has a random lifetime U according to the lifetime probability density ϕ_U. Correspondingly, the cumulative distribution function reads

$$\Phi_U(u) \;=\; \int_0^u \mathrm{d}\tau\, \phi_U(\tau)\,, \tag{2.2}$$

and the average lifetime of that mRNA species is given by

$$\langle U \rangle \;=\; \int_0^\infty \mathrm{d}u\, u\, \phi_U(u)\,. \tag{2.3}$$

Let $Y(t)$ be the random number of mRNAs that have been transcribed in the interval $[0, t)$ and are still present at time t, *i.e.* that have not been degraded until time t. Hence $0 \leq Y(t) \leq X(t)$ for all $t \geq 0$.
We aim to calculate the probability that k mRNAs are still present at t, *i.e.*

$$\Pr\{Y(t) = k \mid X(0) = 0\}\,, \tag{2.4}$$

by following the method of ref. [30].

Via the law of total probability, we can rephrase this to

$$\begin{aligned}\Pr\{Y(t) = k \mid X(0) = 0\} \; = \\ \sum_{n=k}^{\infty} \Pr\{Y(t) = k \mid X(0) = 0,\, X(t) = n\} \Pr\{X(t) = n \mid X(0) = 0\}\,.\end{aligned} \tag{2.5}$$

We sum from $n = k$ since $Y(t) \leq X(t)$ and hence all terms $n < k$ are zero. The second term is just the Poisson process for the generation of mRNAs. In the following, we will compute the first term.
First, we calculate the probability p for a single mRNA to be still present at time t. Each mRNA has random origin O and random lifetime U. The mRNA is still present at time t if the random variable $Z \equiv O + U \geq t$. Therefore, we aim to compute

$$p \;\equiv\; \Pr\{Z \geq t\}\,.$$

We rephrase the probability p using the law of total probability and the definition of the marginal probability. Furthermore, we exploit the fact that, in a conditioned Poisson process, the origin of a single mRNA is uniformly distributed in the time interval $(0, t]$ (see section A.1 in appendix). Hence,

$$p \;=\; \frac{1}{t}\int_0^t \mathrm{d}s\; \Pr\{Z \geq t \mid O = s\}\,.$$

Next, we utilize $\Pr\{Z = O + U \geq t \mid O = s\} = \Pr\{U \geq t - s\}$, thus giving

$$p \;=\; \frac{1}{t}\int_0^t \mathrm{d}s\; \Pr\{U \geq t - s\}\,.$$

Finally, we use Eq. (2.2) and substitute $u = t - s$ to find

$$p \;\equiv\; \Pr\{Z \geq t\} = \frac{1}{t}\int_0^t \mathrm{d}u\; (1 - \Phi_U(u))\;. \tag{2.6}$$

In a second step, we exploit that, conditioned on $X(t) = n$, the number of mRNAs is binomially distributed. This becomes clear once one realizes that the probability $p(t)$ defines the outcome of a Bernoulli trial - whether an mRNA is still present at time t or not. The combination of n independent Bernoulli variables leads to the binomial distribution. Thus, we find

$$\Pr\{Y(t) = k \mid X(0) = 0, X(t) = n\} \;=\; \binom{n}{k} p^k (1-p)^{n-k}\,. \tag{2.7}$$

Inserting Eqs. (2.7) and (2.1) into Eq. (2.5) results in

$$\begin{aligned}\Pr\{Y(t) = k \mid X(0) = 0\} \;&=\; \frac{p^k(1-p)^{1-k}}{k!} e^{-\omega_{\mathrm{tc}} t} \sum_{n=k}^{\infty} \frac{[(1-p)\omega_{\mathrm{tc}} t]^n}{(n-k)!} \\ &=\; \frac{(p\,\omega_{\mathrm{tc}} t)^k}{k!} e^{-\omega_{\mathrm{tc}} t} \sum_{n'=0}^{\infty} \frac{[(1-p)\omega_{\mathrm{tc}} t]^{n'}}{(n')!} \\ &=\; \frac{(p\,\omega_{\mathrm{tc}} t)^k}{k!} e^{-p\,\omega_{\mathrm{tc}} t}\end{aligned} \tag{2.8}$$

Finally, using Eq. (2.6) we arrive at

$$\boxed{\Pr\{Y(t) = k \mid X(0) = 0\} \;=\; \frac{[\omega_{\mathrm{tc}} H(t)]^k \exp\left[-\omega_{\mathrm{tc}} H(t)\right]}{k!}} \tag{2.9}$$

where we introduced

$$H(t) \;\equiv\; \int_0^t \mathrm{d}u \, (1 - \Phi_U(u)) \, . \tag{2.10}$$

Hence, we have arrived at an expression for the number of mRNAs at any time t after the start of the transcription process for any lifetime distribution Φ_U. We find that the distribution of the number of mRNAs is a time dependent Poisson distribution with parameter $\omega_{\mathrm{tc}} \, H(t)$ [30]. In chapter 4 we will investigate implications of this finding when we deal with the time evolution of the mRNA amount after induction of transcription and under stress conditions.

2.1.4. mRNA number after stop of transcription

To understand experiments where the transcription is interrupted we want to calculate the number of mRNAs at any time interval Δt after the interruption at time $t = t_s$, *i.e.*

$$\Pr\{Y(t_s + \Delta t) = k \mid X(0) = 0\} \tag{2.11}$$

We can compute this quantity by performing the same steps as in Eqs. (2.5), (2.7) and (2.8).

$$\Pr\{Y(t_s + \Delta t) = k \mid X(0) = 0\} \;=\; \frac{(p_s \, \omega_{\mathrm{tc}} t_s)^k}{k!} e^{-p_s \, \omega_{\mathrm{tc}} t_s} \, . \tag{2.12}$$

This equation differs from Eq. (2.8) since the Poisson process for the synthesis of mRNAs runs from 0 to t_s only and hence the probability p_s of a single mRNA to be still present at $t = t_s + \Delta t$ is different. Nevertheless, we can compute p_s similarly to Eq. (2.6)

$$\begin{aligned} p_s \;&=\; \Pr\{Z \geq t_s + \Delta t\} \\ &=\; \frac{1}{t_s} \int_0^{t_s} \mathrm{d}s \, \Pr\{Z \geq t_s + \Delta t \mid O = s\} \\ &=\; \frac{1}{t_s} \int_{\Delta t}^{t_s + \Delta t} \mathrm{d}u \, (1 - \Phi_U(u)) \, . \end{aligned} \tag{2.13}$$

In the first step, we have exploited again the fact that the generation events are uniformly distributed in the interval $[0, t_s)$. In the second step, we have used $\Pr\{Z = O + U \geq t_s + \Delta t \mid O = s\} = \Pr\{U \geq t_s + \Delta t - s\}$, Eq. (2.2) and a suitable transformation of variables. Thus, inserting Eq. (2.13) into Eq. (2.12) yields

$$\boxed{\Pr\{Y(t_s + \Delta t) = k \mid X(0) = 0\} \;=\; \frac{[\omega_{\mathrm{tc}} H_2(t_s + \Delta t)]^k \exp\left[-\omega_{\mathrm{tc}} H_2(t_s + \Delta t)\right]}{k!}} \tag{2.14}$$

where

$$H_2(t_s + \Delta t) \;\equiv\; \int_{\Delta t}^{t_s + \Delta t} \mathrm{d}u \, (1 - \Phi_U(u)) \, . \tag{2.15}$$

Thus, we conclude that also after the stop of transcription the distribution of the number of mRNAs remains Poissonian with a time-dependent parameter $\omega_{\mathrm{tc}}\, H_2(t_s+\Delta t)$ regardless of the underlying mRNA lifetime distribution Φ_U. In chapter 3 we will use this central result to describe the decreasing amount of mRNAs in decay experiments.
It is convenient to introduce a common abbreviation for both cases, $H(t)$ and $H_2(t_s+\Delta t)$

$$\mathcal{H}_{t_1}^{t_2} \equiv \int_{t_1}^{t_2} \mathrm{d}u\,(1-\Phi_U(u))\,. \tag{2.16}$$

For $t_1 = 0$ and $t_2 = t$ we arrive again at $H(t)$ from Eq. (2.10). Conversely, for $t_1 = \Delta t$ and $t_2 = t_s + \Delta t$ we regain $H_2(t_s + \Delta t)$ from Eq. (2.15). Note that always $t_2 > t_1 \geq 0$ to ensure a positive number of mRNAs.

2.1.5. Age and residual lifetime distribution

A randomly selected mRNA has a random age. Hence, a large ensemble of mRNAs at a given (global) time t will reveal an age distribution (see also figure 2.1).
We are interested in the age of a randomly chosen mRNA at any point in time t. Therefore, in the following, we consider mRNAs that have been created according to a Poisson process in the interval $[0, t_s)$ and have a random lifetime U distributed according to the density ϕ_U. Let A be the random variable that gives the age of a randomly chosen mRNA. The age distribution of the mRNA is given by the distribution of $A = t_s + \Delta t - O$ (see figure 2.1), where O is the random variable giving the time point of origination. As shown in detail in the appendix, section A.2, we can find an analytical expression of the age probability density function

$$\phi_A(a \mid t_s + \Delta t) \;=\; \left[\int_{\Delta t}^{t_s+\Delta t} \mathrm{d}u\,(1-\Phi_U(u))\right]^{-1} (1-\Phi_U(a))\;, \tag{2.17}$$

for $\Delta t \leq a < t_s + \Delta t$ and zero otherwise. Note that Eq. (2.17) is the marginal probability density of the age random variable.
The age distribution at a given time t is an interesting quantity which allows insights into the composition of an ensemble of mRNAs. Note that we can consider different scenarios, before and after the stop of transcription, by choosing t and Δt accordingly. In chapter 4 we will discuss the time dependence of $\phi_A(a \mid t)$ and its average value $\langle A\rangle_t$. Moreover, in chapter 5 we will examine the role of the age distribution for mRNA translation and protein synthesis.

Residual lifetime distribution Similar to its age, a randomly selected mRNA will also have a random residual (or excess) lifetime. Hence, a large ensemble of mRNAs at a given time t will reveal a residual lifetime distribution (see also figure 2.1).
The residual lifetime R of an mRNA is a statistical quantity complementary to the age of the mRNA. A detailed computation (see section A.2 in appendix) reveals that the residual lifetime probability density is given by

$$\phi_R(r \mid t_s + \Delta t) \;=\; \left[\int_{\Delta t}^{t_s+\Delta t} \mathrm{d}u\,(1-\Phi_U(u))\right]^{-1} (\Phi_U(t_s+\Delta t+r) - \Phi_U(\Delta t+r))\;, \tag{2.18}$$

for all $r \geq 0$. Thus, $\phi_R(r)$ is the marginal probability density of the random residual lifetime R. The residual lifetime of an mRNA is a compelling quantity which is capable to characterize the *aging* of the mRNA molecules. In analogy to the age distribution, for the residual lifetime distributions several limit cases of interest exist and will be further discussed in chapters 3 and 4.
Therefore, both the age distribution as well as the residual lifetime distribution depend on the form of the lifetime probability density ϕ_U and on the time after start of transcription t_s as well as the time interval since the interruption of transcription Δt.

Short summary In this section we have performed the central computations used for the remaining parts of the thesis. We found an analytical expression for the number and distribution of mRNAs under different experimental conditions as well as for the age and residual lifetime distributions. To the best of our knowledge, Eqs. (2.14), (2.17) and (2.18) are original results of this work.

2.2. Formulation with a Master equation

In this section, we confirm the results derived in the previous chapter for the limit case of exponentially distributed lifetimes.
The conventional description of mRNA turnover considers transcription with a constant synthesis rate ω_{tc} and mRNA degradation with a constant decay rate ω_{r}. This can be formulated via the following Master equation

$$\frac{dP_k(t)}{dt} = \omega_{\mathrm{tc}}\, P_{k-1}(t) + \omega_{\mathrm{r}}\,(k+1)\, P_{k+1}(t) - \omega_{\mathrm{r}}\, k\, P_k(t) - \omega_{\mathrm{tc}}\, P_k(t)\,. \tag{2.19}$$

Here, $P_k(t)$ denotes the probability that k mRNAs are present. It holds $k \geq 0$ and for $k = 0$ obvious modifications have to be applied to Eq. (2.19). The first two terms on the right hand side are source terms and the last two are loss terms. The synthesis is *zeroth-order*, *i.e.* just determined by the rate, whereas the degradation *first-order*, *i.e.* proportional to the existing amount of mRNA. The solution can be obtained via generating functions and solving the resulting differential equation. Using the initial condition $P_0(0) = 1$ and $P_i(0) = 0$ $(i > 0)$, we find

$$P_k(t) = \frac{\left[\omega_{\mathrm{tc}}/\omega_{\mathrm{r}}\,(1 - e^{-\omega_{\mathrm{r}}\, t})\right]^k}{k!} \exp\left[-\omega_{\mathrm{tc}}/\omega_{\mathrm{r}}\left(1 - e^{-\omega_{\mathrm{r}}\, t}\right)\right]\,. \tag{2.20}$$

It is a Poisson distribution with a time-dependent parameter $\frac{\omega_{\mathrm{tc}}}{\omega_{\mathrm{r}}}\,(1 - e^{-\omega_{\mathrm{r}}\, t})$. If we insert an exponential lifetime distribution $\Phi_U(t) = 1 - \exp(-\omega_{\mathrm{r}}\, t)$ into Eq. (2.9), we retrieve the following parameter for the Poisson distribution after the start of transcription

$$\begin{aligned} \omega_{\mathrm{tc}}\, H(t) \;&\equiv\; \omega_{\mathrm{tc}} \int_0^t \mathrm{d}u\,(1 - \Phi_U(u)) \\ &= \omega_{\mathrm{tc}} \int_0^t \mathrm{d}u\, \exp(-\omega_{\mathrm{r}}\, u) \\ &= \frac{\omega_{\mathrm{tc}}}{\omega_{\mathrm{r}}}(1 - \exp(-\omega_{\mathrm{r}}\, t))\,. \end{aligned} \tag{2.21}$$

Hence, the formulation with a Master equation corresponds to the solution of the special case of an exponential probability distribution for the lifetime of an mRNA.
The stochastic description of mRNA decay after the interruption of transcription can be discussed in a similar way. The corresponding Master equation for a time interval Δt after the interruption describes a pure death process,

$$\frac{dP_k(\Delta t)}{d(\Delta t)} = \omega_{\mathrm{r}}\,(k+1)\, P_{k+1}(\Delta t) - \omega_{\mathrm{r}}\, k\, P_k(\Delta t)\,. \tag{2.22}$$

Eq. (2.20) at $t = t_s$ serves as the initial condition of Eq. (2.22). Hereby, we can confirm our result for Eq. (2.14) for the limit case of an exponential lifetime distribution with constant decay rate via a Master equation.
The advantage and the principal reason for the choice of the approach in section 2.1 is that it extends the formulation for arbitrary lifetime distributions ϕ_U. This cannot be achieved with a Master equation. Note that in the appendix, section A.4, we scrutinize how partial results of our general approach can be formulated with a Master equation.

2.3. Age-dependent degradation rate

The aim of this section is to show how one can reinterpret the results for general lifetime distributions as an age-dependent degradation rate.
The extension to arbitrary mRNA lifetime densities ϕ_U means that the rate of degradation ω_{deg} is no longer constant. Rather the degradation rate ω_{deg} changes during the lifetime of an mRNA which can be formulated as an age-dependent rate $\omega_{\text{deg}}(a)$. This reflects aging of the mRNA molecules and in chapter 3 we will discuss the origins of the molecular aging.
In the following, we shall derive how the age-dependence is related to the mRNA lifetime densities ϕ_U. Consider a newly synthesized mRNA (with $a = 0$). The probability $\wp(a)$ that the mRNA is still intact at age a obeys

$$\dot{\wp}(a) = -\omega_{\text{deg}}(a)\wp(a)\,. \tag{2.23}$$

We can find the solution via integration and assuming the initial condition $\wp(0) = 1$. Hence we obtain

$$\wp(a) = \exp\left(-\int_0^a d\tau\,\omega_{\text{deg}}(\tau)\right)\,. \tag{2.24}$$

As the cumulative probability is related via $\Phi_U(a) = 1 - \wp(a)$, we have

$$\Phi_U(a) = 1 - \exp\left(-\int_0^a d\tau\,\omega_{\text{deg}}(\tau)\right)\,. \tag{2.25}$$

According to Eq. (2.2) the derivative of Eq. (2.25) gives the probability density

$$\phi_U(a) = \omega_{\text{deg}}(a)\,\exp\left(-\int_0^a d\tau\,\omega_{\text{deg}}(\tau)\right)\,. \tag{2.26}$$

Rearranging finally leads to

$$\omega_{\text{deg}}(a) = \frac{\phi_U(a)}{1-\Phi_U(a)}\,. \tag{2.27}$$

Thus, this well-known result shows that every lifetime probability density can be translated into an age-dependent degradation rate. We will use this concept in chapter 3 when we discuss the degradation of mRNAs and its mechanisms in more detail. It is important to note that the degradation rate depends on its age, *i.e.* on the time elapsed since its synthesis, as opposed to a dependency on a global time. It is this what makes it a very interesting quantity to describe molecular aging.

2.4. Continuous-time Markov processes

In section 2.1.2 we have already introduced the Poisson process - it is a typical example of a continuous-time, discrete-state Markov chain. There, the state $X(t)$ takes discrete values and transitions between states occur at continuous times t. In appendix D we will also discuss a non-homogeneous Cox process, where values of the intensity parameter will fluctuate in time according to a Markov chain. In this section, we will review some more concepts of continuous-time Markov chains that will prove useful in this thesis.

Motivation In the derivation of the distribution of the mRNA number as well as age and residual lifetime distribution, we stressed on the fact that the formulation holds for general lifetime probability densities ϕ_U. Nevertheless, for any concrete comparison of the theory to experimental data, we need to define an adequate, specific form of ϕ_U. As we will see, continuous-time Markov chains will prove useful to describe the stochastic changes between states. In particular, we will define ϕ_U as the probability density of the random time to reach an absorbing state. In this section, we will introduce the concepts necessary to compute ϕ_U. In chapter 3 we will explain in more detail how we relate Markov chains with biochemical processes related to the degradation of mRNA.

General formulation Markov chains consist of discrete states i which are connected via a network of transitions. The Markov property expresses the assumption that the transition probability from one state to another depends only on the corresponding transition rate and not on the history of previous steps. In continuous-time Markov chains, the waiting time at each time is a continuous quantity and the frequency of transition is given by an exponential probability distribution with a corresponding rate.
In order to study Markov chains, we apply the generator formalism that is built on matrix algebra - in the literature also the term *phase-type distributions* is used [37]. Note that one can also obtain a recursive solution from a Master equation.
The transition rate matrix is defined as

$$\mathcal{Q} \equiv \begin{pmatrix} \mathcal{S} & \mathbf{S_0} \\ \mathbf{0} & 0 \end{pmatrix} . \tag{2.28}$$

The matrix $\mathcal{S}$ denotes the inner transitions between the transient states and the matrix $\mathbf{S_0}$ the transitions to the absorbing states. These two matrices obey the condition

$$\mathbf{S_0} \equiv -\mathcal{S}\,\mathbf{1}\,, \tag{2.29}$$

so that all rows sum up to zero. The symbol $\mathbf{1}$ is a column vectors of ones, $\mathbf{1} \equiv (1, 1, \cdots, 1)^\intercal$.
One can compute the cumulative distribution function of the time until absorption [37] as

$$\Phi(t) = 1 - \boldsymbol{\alpha}\, \exp(\mathcal{S}\, t)\, \mathbf{1}\,. \tag{2.30}$$

Here, $\boldsymbol{\alpha}$ is a row vector containing the initial conditions. Hence, the result is a scalar function of t. The probability density function reads

$$\phi(t) = \boldsymbol{\alpha}\, \exp(\mathcal{S}\, t)\, \mathbf{S_0} \tag{2.31}$$

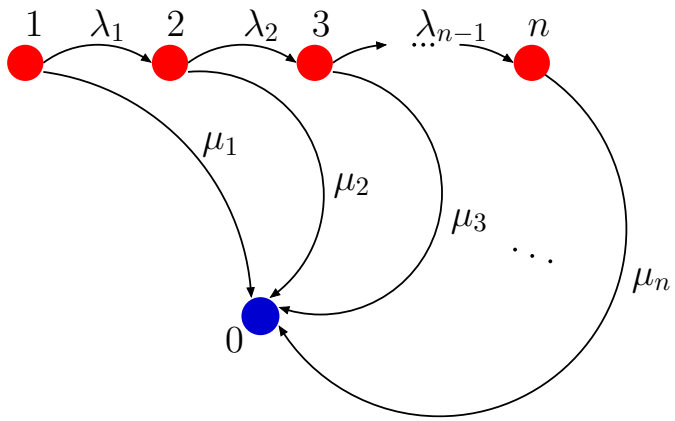

Figure 2.2.: Visualization of a Markov chain of n discrete states. Transitions from a state i to the state $i+1$ are governed by a rate λ_i. Moreover, from each state i a transition to the absorbing state is possible with a rate μ_i. We are interested in the time to absorption which we map to the lifetime probability density ϕ_U (chapter 3).

It is our goal to compute the density $\phi(t)$ for any desired form of a Markov chain. We will identify $\phi(t)$ with the lifetime probability density ϕ_U.

Network of states with one absorbing state In the following, we consider a Markov chain with n transient states and one absorbing state, see figure 2.2. One realization of a Markov chain consists of transitions from each state to the following and a transition from each state to the absorbing state with different rates. In the following, we denote the transition rates from a state i to the state $i+1$ with λ_i. The transitions from a state i to the absorbing state are described by the rate μ_i. Hence, the matrix $\mathcal{S}$ takes the form

$$\mathcal{S} = \begin{pmatrix} -\mu_1 - \lambda_1 & \lambda_1 & 0 & 0 \\ 0 & -\mu_2 - \lambda_2 & \lambda_2 & 0 \\ \vdots & \vdots & \ddots & \ddots \\ 0 & 0 & -\mu_{n-1} - \lambda_{n-1} & \lambda_{n-1} \\ 0 & 0 & 0 & -\mu_n \end{pmatrix} \tag{2.32}$$

and the vector $\mathbf{S_0}$ is given by

$$\mathbf{S_0} = (\mu_1, \mu_2, \ldots, \mu_{n-1}, \mu_n)^\intercal \tag{2.33}$$

For any concrete computation of the absorption probability density $\phi(t)$, we have to define an appropriate initial condition $\boldsymbol{\alpha}$. In view of the later application of this model, the initial conditions generally assumed here reads

$$\boldsymbol{\alpha} = (1, 0, \cdots, 0) \ . \tag{2.34}$$

The main task is to find the matrix exponential $\exp(\mathcal{S}\,t)$ defined as

$$\exp(\mathcal{S}\,t) = \sum_{k=0}^{\infty} \frac{1}{k!} \mathcal{S}^k t^k \ . \tag{2.35}$$

For this simple system, diagonalization is always possible for arbitrary numbers of transitions. For a large number of transitions, the resulting expression will be lengthy but takes a much simpler forms if some rates are the same. Once the matrix exponential is known, the corresponding probability density functions $\phi(t)$ can be computed according to Eq. (2.31).

Example cases The probability density $\phi(t)$ takes the simple form of an exponential with a single rate constant if one rate λ_k is much smaller than all others (e.g. $\lambda_k \ll \lambda_i, \mu_i$) or if $\mu_i = \mu \; \forall \, i$. The same is true for the trivial cases where $\lambda_1 \ll \mu_1$ or $n = 1$. Another simple solution consists in setting $\mu_1 = \mu_2 = \cdots = \mu_{n-1} = 0$, *i.e.* absorption is only possible from the last state. If furthermore $\lambda_i = \lambda \; \forall \, i$, the probability density function takes the form of a Γ-distribution. We will use this case later as an generic example for a multi-step process (see chapter 4). In the case that $\mu_2 = \mu_3 = \cdots = \mu_n = \mu$ the probability of absorption is the same in all states. Therefore, transitions between non-absorbing states play no role and the probability to absorption takes a relatively simple form that is independent of the number of transient states n. Some specific forms of the function $\phi(t)$ that we will apply in chapter 3 are given explicitly in the appendix (see section B.2).

Markov chains and molecular aging In section 2.3, we discussed that general forms of ϕ_U lead to an age-dependent degradation rate. The latter provides a theoretical concept to describe molecular aging. Moreover, as will become more clear in the subsequent chapters of this thesis, Markov chains prove useful to visualize the aging of molecules. As the age of a molecule increases, it will undergo modifications that can be understood as transitions on a Markov chain. Therefore, older mRNAs have a higher probability of being in a more advanced state. Since mRNAs sampled at different time points have a different age distribution (see above), also the probability of finding an mRNA in a specific state is time dependent. In chapter 4, we will see how aging can be related to different states in the Markov chain.

Short summary Markov chains are a useful tool to capture the stochastic behavior of processes consisting of multiple, random steps. Here, we presented an approach where the chain connecting subsequent transient steps is linear and only one absorbing state is present. However, the concept is more general and can be extended to more complicated networks of states. The network should also reflect the existing knowledge of the random process and the desired resolution.

2.5. Simulation techniques

In the previous sections of this chapter we have developed a powerful analytical framework that allows studying the evolution of the number of mRNA under different conditions. Additionally, one can perform stochastic (Monte-Carlo) simulations to study the turnover of mRNA. It is a means to illustrate the mathematically described processes but also a way to confirm our computations. Furthermore, we can use computer simulations in combination with our theory of degradation to study the process of translation. Finally, we will use stochastic simulations when we conceive more detailed models that are no longer analytically tractable.
Monte-Carlo simulations are frequently used to study different stochastic phenomena (in absence of an analytical theory). Gillespie was one of the first to prove that they provide an alternative to analytic solutions of a Master equation [38]. With the advent of more powerful computers, they were readily applied to gene expression [33, 4, 39]. However, the Gillespie algorithm - as it is commonly used in applications - relies on exponentially distributed random variables and our theory goes beyond this by considering arbitrary lifetime distributions. Additionally, we investigate translation with related simulation techniques, commonly referred to as *TASEP* (*Totally antisymmetric exclusion process*) [40, 41, 42]. From the side of analytical theory TASEPs have been described in detail [43].

2.5.1. Simulation of mRNA turnover

One can interpret figure 2.1 as one realization of a stochastic simulation of mRNA turnover. The time points of the origination events are sampled from randomly generated exponential numbers and iterate the time by increments δt. For each origination event, the lifetime is a priori determined from another random number sampled from a lifetime distribution. The variable X counts the transcription events, the array $Start()$ stores the transcription time points and $End()$ the degradation times. Therefore, for each realization one can keep track of the live mRNAs. The simulation ends after a preassigned run length T has been reached.
One can express this more formally via the following algorithm:

Basic mRNA turnover algorithm

set $t = 0$, $X = 0$
while $t < T$
    generate δt
    set $t = t + \delta t$
    set $X = X + 1, \mathrm{Start}(X) = t$
    generate U, **set** $\mathrm{End}(X) = t + U$

This simple algorithm can be used to check our computations. Furthermore, it can be straightforwardly extended to simulate the protein turnover and non-homogeneous transcription (see appendix sections A.6 and D as well as section 5.2). Additionally, it can be adapted for a more detailed model of the cell cycle (section 5.3).

2.5.2. TASEP simulations

A *TASEP* simulation (*Totally antisymmetric exclusion process*) can be used to understand the process of translation. There, ribosomes bind to an mRNA with an initiation rate ω_{on} and walk on a homogeneous chain with an elongation rate v. For each ribosome i on the chain as well as the pool of ribosomes waiting to initiate translation, one can determine the time of the next event by a random number generator. An event consists either in the attempt of a ribosome to progress by one codon or - in the case of initiation - in the attachment of a ribosome to the first site of an mRNA. The attempt is, however, only successful if the site is empty, *i.e.* not occupied by another ribosome. Finally, after reaching the last codon, the ribosome detaches since it has completed the synthesis of a polypeptide. Furthermore, the ribosomes are modeled as extended objects of size l since they typically occupy more than one codon (in fact, typically 10 codons [44, 45]).
An algorithm for a *TASEP* simulation is given in section A.6.

2.6. Chapter summary

In this chapter, we have learned how the process of mRNA turnover can be theoretically described for general lifetime distributions. This extends the standard description with a Master equation and provides a powerful framework for modeling mRNA turnover and decay. We found that the mRNA number distribution before and after the stop of transcription is Poissonian regardless of the choice of the lifetime distribution (see Eqs. (2.9) and (2.14), respectively). Eqs. (2.10) and (2.15) describe the time-dependence of the parameter of the distribution. Moreover, in section 2.1.5 we explained how related quantities such as the age and the residual lifetime of an mRNA can be derived. To find appropriate mRNA lifetime distributions, we introduced continuous-time Markov chains. Here, we map biochemical transformations of the mRNA molecules into a network of states. The computation of the time to absorption gives the desired lifetime probability density. Finally, we have outlined how Monte-Carlo simulations can aid to confirm and extend the analytical approaches introduced here. All of the concepts introduced here will be applied in the subsequent chapters. Most of the methods introduced in this chapter are a review of textbook material. Nevertheless, to the best of our knowledge, Eqs. (2.14), (2.17) and (2.18) are original results of this work. Moreover, the connection of the mRNA lifetime distribution to the absorption process of a Markov chain (figure 2.2) has been formulated for the first time here.

3. Decay of the amount of mRNA

In this chapter, we elaborate how the theory we introduced in chapter 2 can be used to study mRNA decay experiments. Thereby, we can improve the analysis of decay experiments and establish the link between the decay experiments and the underlying biochemical processes leading to decay. In the first part, we provide general implications for mRNA decay (section 3.1). In section 3.2, we apply our theory of multi-step degradation to studies on *S. cerevisiae*. In section 3.3, we develop a concrete model for the degradation in the bacterium *E. coli* (section 3.3).

3.1. General results

As mentioned in the introduction, the stability of mRNA is typically experimentally assessed via the interruption of transcription. An antibiotic such as *rifampicin* or *actinomycin D* which blocks the further synthesis of mRNAs is introduced to the cell culture. Hence, the mRNAs that are present in the cell decay thus allowing to conclude about the stability of the mRNAs for a given gene. An alternative procedure is based on metabolic labeling. When a heavy or radioactive nucleotide is incorporated into the cells, one can distinguish mRNA transcripts synthesized before and after the incorporation of the label (*e.g.* due to their different weight). Therefore, the previously synthesized mRNAs decay similarly to those mRNAs after the complete inhibition. The advantage of the latter method is that it is thought to perturb the cell metabolism to a much lesser extent. In the following, we will refer to all techniques as the *stop of transcription*.
In chapter 2, Eq. (2.14), we found that the number of mRNAs after the stop of transcription is described by a Poisson distribution

$$\Pr\{Y(t_s+\Delta t)=k \mid X(0)=0\} \;=\; \frac{[\omega_{\mathrm{tc}} H_2(t_s+\Delta t)]^k \exp\left[-\omega_{\mathrm{tc}} H_2(t_s+\Delta t)\right]}{k!}\,,$$

with the time-dependent parameter given in Eq. (2.15)

$$H_2(t_s+\Delta t) \;=\; \int_{\Delta t}^{t_s+\Delta t} \mathrm{d}u\,(1-\Phi_U(u))\,.$$

For the analysis of the experiments we are interested in the average mRNA number which is given by $\omega_{\mathrm{tc}} H_2(t_s+\Delta t)$. In the following, we will assume that the cells were growing long enough to have reached a steady state mRNA level prior to the stop of transcription. Hence, we consider the limit case $t_s \to \infty$[1]. Thus, the number of mRNAs after the stop

[1] With our general formalism one could easily adapt the following analysis also to the case when the steady state assumption was not met (see also appendix, section B.1).

of transcription reads

$$N_r(\Delta t) \equiv N_r^{\downarrow}(\Delta t) \equiv \omega_{\mathrm{tc}} \int_{\Delta t}^{\infty} \mathrm{d}u \, (1 - \Phi_U(u)) \, . \tag{3.1}$$

We can normalize this relation to define the relative number of mRNAs

$$\Lambda(\Delta t) \; \equiv \; \frac{1}{\langle U \rangle} \int_{\Delta t}^{\infty} \mathrm{d}u \; (1 - \Phi_U(u)) \; . \tag{3.2}$$

Hence, the decay pattern is governed by the lifetime probability function $\Phi_U(u)$ and its average value $\langle U \rangle$ - note, however, that it is independent of the transcription rate. In Eq. (2.2) we previously defined $\Phi_U(u) = \int_0^u \mathrm{d}t \, \phi_U(t)$, which gives the probability that an mRNA has a lifetime smaller or equal than u.
The half-life $t_{1/2}$ can be computed by solving

$$\Lambda(t_{1/2}) \equiv 1/2 \, . \tag{3.3}$$

We can make some general inferences about the form of the decay pattern regardless of the details of the mRNA lifetime distribution. From the first derivative of Λ we can conclude

$$\frac{\mathrm{d}\Lambda(x)}{\mathrm{d}x} \; = \; -\frac{1 - \Phi_U(x)}{\langle U \rangle} \; \leq \; 0 \, , \tag{3.4}$$

since the cumulative probability function always satisfies $\Phi_U(x) \leq 1$. Hence, the mRNA amount should decay monotonically as should be intuitively clear. Furthermore, an analysis of the second derivative of Λ shows also that

$$\frac{\mathrm{d}^2\Lambda(x)}{\mathrm{d}x^2} \; = \; \frac{\phi_U(x)}{\langle U \rangle} \; \geq \; 0 \, , \tag{3.5}$$

which implies that the decay pattern is always convex.
Given the model independent nature of the relations (3.4) and (3.5), all experimental data about mRNA decay must satisfy the following *bona-fide* criteria:

- the mRNA amount is monotonically decreasing in time
- the decay pattern is always convex

Indeed, when we compare our theory with experimental data, we will use these criteria to exclude implausible data sets. Often, decay patterns show different behavior which might hint to the perturbing nature of the decay experiments including transcription inhibition. Experimental decay data always measure the mRNA abundance at discrete time points. Therefore, to analyze the validity of these necessary conditions, in practice one would have to consider the discrete analogues of the first and second derivative. Given the limited temporal resolution and the noisiness of the experimental setups, one should lift the criteria to some extent and include a certain error tolerance. In any case, from the theoretical point of view, data that do not exhibit both of these properties should involve additional processes that interfere with degradation.
Given a decay pattern, Eq. (3.2) also shows how one can in principle infer the lifetime distribution directly from the decay data. Thus, Eqs. (3.2) provides a method how one can measure indirectly the lifetime probability density ϕ_U of an mRNA species. Practically, however, a very high number of recorded time points is necessary to determine the mRNA lifetime distribution accurately.

3.2. Multi-step degradation in yeast

Any mRNA decay pattern should reflect the molecular mechanism of its degradation. Therefore, the ultimate goal is to build a consistent theoretical description which reflects the degradation pathways and successfully describes the outcome of a decay experiment. In the light of our theory, the quest lies in formulating the correct lifetime distribution function Φ_U.

3.2.1. Idea of multi-step degradation

In the introduction, we have outlined various biochemical mechanisms that are responsible for degradation. In each organism, degradation is a complex process that involves a multitude of different enzymes, many of which appear in degradation complexes. Different mechanisms compete and each one of them may consist of several rate-limiting steps. Hence, for the regulation of the stability of an mRNA of a given gene, various enzymes are responsible. Furthermore, the concentration of each enzyme and their physicochemical characteristics should determine the efficiency and therefore the corresponding rates.
To link these criteria to the lifetime distribution of an mRNA species, we use a Markov chain model which we introduced in section 2.4. We imagine that each mRNAs undergoes a series of biochemical modifications occurring at random time points during its lifetime. Moreover, each modification might alter the efficiency of degradation of that mRNA. The different states of a Markov chain represent the mRNA after certain modifications and the mRNA makes transitions from one state to another with a rate that represents the biochemical process that govern these transitions. Furthermore, at each state the mRNA can be degraded with a rate that is related to its biochemical affinity to the degrading enzymes in that particular state. The lifetime probability density ϕ_U can be computed from the time to absorption (see also figure 2.2). In general, however, the Markov chain does not have to be a linear chain as visualized there.
Therefore, to find an optimal ϕ_U, one would have to identify the relevant modifications and construct an appropriate state network. Next, one would need to estimate the rates of the transitions and degradation from experiment.
To unveil all possible pathways and the corresponding modifications including their rates seems far from practice. Nevertheless, the advantage of this approach is that one can construct different models of degradation and verify or falsify them via the comparison to experiment. One can start with simple assumptions and then increase the accuracy of the description step-by-step.

Decapping as an example

To illustrate these general ideas, we want to analyze how one can map the decapping mechanism to a Markov chain (see figure 3.1).
The *decapping* mechanism is thought to be the most important degradation pathway in eukaryotic organisms such as yeast but also in higher organisms [14]. It consists of several distinguishable steps. First, an mRNA has to be targeted for degradation, thus the enzyme complex (possibly including micro RNAs) must encounter the molecule via diffusion. Next, an enzyme complex (such as the *PAN2-PAN3* complex) digests the

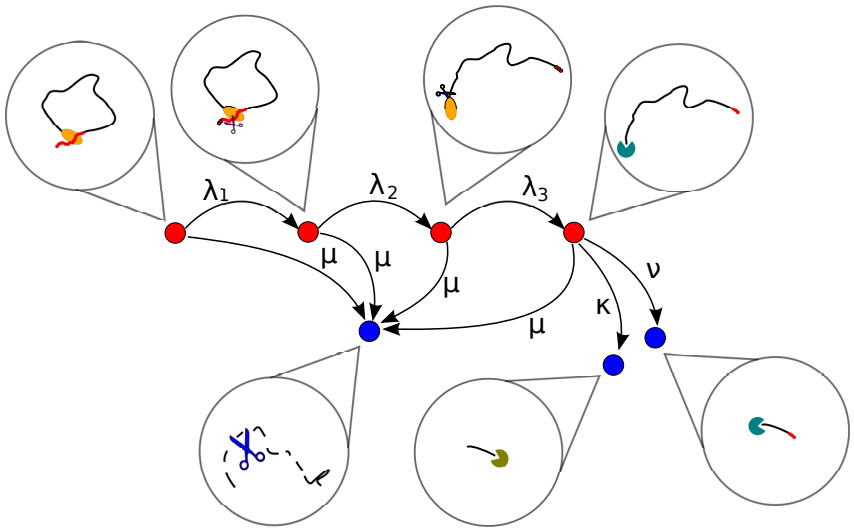

Figure 3.1.: **The decapping mechanism of mRNA degradation**. Shown is a concrete realization of the Markov chain model for the *decapping* mechanism. Each transformation from one state to another occurs with rate λ_i. From the decapped state, the transcripts can be degraded via $5' \to 3'$ exonuclease with rate ν and $3' \to 5'$ exonuclease with rate κ. At all times, there is also a competing degradation pathway which is initiated endonucleolytically with a rate μ. The description contains three absorbing states but can be reduced to one absorbing state with one degradation rate at each state.

poly(A) tail of the mRNA, leading to the *deadenylation* of the mRNA, which occurs at a certain rate. Additionally, also the enzyme needs to encounter the mRNA which is governed by another random waiting time. Once the *poly(A)* tail is short enough, decapping is performed via the action of the decapping enzyme *DCPL* with yet another rate. Finally, exonucleolytic degradation occurs in $5' \to 3'$ direction behind the last translating ribosome [14, 17].

Each of these steps occurs with a certain rate associated with a random waiting time. The waiting time itself is governed by the diffusion time across the cell and the time scale of catalytic activity. Additional steps may occur between each of the steps enlisted here. However, an adequate resolution has to be chosen for the description and only the rate limiting steps have to be considered. An example for different resolution can be given in this case by considering deadenylation as either a single-step process or as the sum of individual A-removing steps.

Furthermore, alternative pathways can influence the stability of the mRNA, for instance in this case $3' \to 5'$ exonuclease after decapping or also endonucleolytic degradation even before decapping [14]. We can include the latter by allowing degradation at all times even before decapping. Therefore, there is a transition to the absorbing state from each state with a certain rate. The $3' \to 5'$ pathway can also be incorporated by adjusting the absorption rate after decapping.

Therefore, to construct such a network of transitions, one has to carefully choose which steps to consider. Ideally, the associated rates can also be inferred from experiment.

3.2.2. Model specification

In the present chapter, we want to illustrate how the Markov chain model can be applied to study decay experiments. Since the exact pathways for each different gene are unknown, we distinguish, in the following, between single-step and multi-step degradation. On the one hand, we consider endonucleolytic degradation, such as depicted in figure 1.2 B, as a prototype of single-step degradation. On the other hand, we regard the decapping mechanism shown in figure 1.2 C as a prototype of multi-step degradation. As outline above, our theory is able to capture the essential features of different pathways of degradation and, thus, allows interpolating between single-step and multi-step degradation pathways. Hence, mRNA degradation can be studied mathematically as a continuous-time Markov chain (see section 2.4). The lifetime distribution will resemble an exponential function if there is only one rate limiting process relevant for degradation, like for instance in figure 1.2 A. Conversely, the lifetime distribution will have a more complicated form if several biochemical modifications are necessary, like for instance in figure 1.2 B, and it will be directly related to the details of the particular degradation pathway.
For the multi-step description, we choose a linear Markov chain with 5 different states. Such a model consists of up to 9 different rate constants that describe the different transitions. However, since the rates are unknown and a large number of free parameters renders a comparison to data trivial, we restrict the number of free parameters in the concrete formulation of the model to three (see figure 3.2). In absence of more detailed knowledge of the rates, we consider that all transitions between states occur with the same rate λ. In the following we distinguish between two possibilities for the degradation rate, namely either setting $\mu_i \equiv \mu$ for $\forall\, i < 5$ and $\mu_5 \equiv \nu$ (case A) or assuming $\mu_1 \equiv \mu$ and $\mu_i \equiv \nu$ for $\forall\, i > 1$ as well as $\mu > \nu$ (case B). Furthermore, in some cases it proved useful to set $\mu = 0$ (in case A) or $\lambda = \nu$ (in case B) to reduce the number of free parameters and thus to improve the search for a global minimum in the fitting algorithm when comparing the theory to experimental data. The analytical result for the lifetime density is given in section B.2 in the appendix.

3.2.3. General results

From the concrete realization of the Markov chain model, we can compute the lifetime probability density $\phi_U(t)$ and the distribution function $\Phi_U(t)$ and insert it into Eq. (3.2). This fixes the functional form of the theoretical decay function $\Lambda(\Delta t)$ after stop of transcription. To determine the unknown parameters λ, μ and ν, we devised a fitting procedure based on non-linear regression.
For each measured time point in the experiment carried out on the yeast *S. cerevisiae* [27], we reduce the root mean square difference between the data and our theoretical curve $\Lambda(\Delta t)$. Thus, we obtain the optimal parameters as well as the fitting error (more precisely the residual sum of squares) for each gene. Furthermore, we can compare the outcome of this analysis with a competing model of single-step decay.
However, large parts of the data bear obvious errors as they are not *bona-fide*, namely they do not satisfy the conditions (3.4) or (3.5). Even after putative interruption of transcription, some mRNAs do not decay at all but their level increases. Besides, many mRNA patterns exhibit non-convex profiles. This might indicate general problems in the

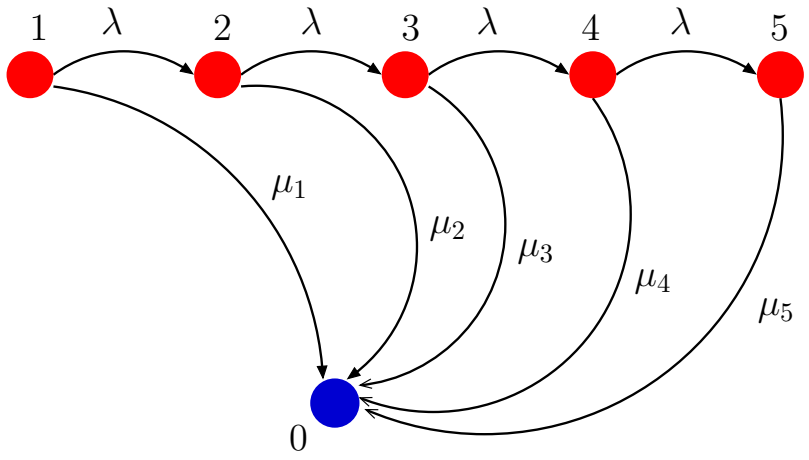

Figure 3.2.: **Markov chain model in practice**. For the practical analysis of multi-step degradation, we use a model with 5 states where, for simplicity, we assume always the same transition rate $\lambda_i \equiv \lambda$. Degradation is possible from all states. To reduce the number of free parameters, we assume moreover either $\mu_i \equiv \mu$ for $\forall\, i < 5$ and $\mu_5 \equiv \nu$ (case A) or $\mu_1 \equiv \mu$ and $\mu_i \equiv \nu$ for $\forall\, i > 1$ as well as $\mu > \nu$ (case B).

quantitative measurement in these types of experiments where transcription is blocked globally. However, the limited temporal resolution of the experiment, the precision of the microarray detector and the inherent fluctuations of the mRNAs should be accounted for. Hence, we restrict the analysis to those data sets that meet these criteria within a reasonable tolerance (see appendix section B.1).

We find 424 nearly bona-fide data sets, 315 of which have a very low fitting error (residual sum of squares $r^2 < 0.005$). In figure 3.3 we display the decay patterns as obtained from our fitting procedure for the subset of the 315 well described genes. The decay patterns can be divided into three different categories. If the fitting of the Markov chain model leads to a negligible improvement of the fitting error compared to a model based on a constant decay rate, the assumption of an exponential decay with constant rate is reasonable. Conversely, if one of the two possible variants of the Markov chain model (see cases A and B above) leads to a substantial refinement of the fitting result, we can assign these data sets to different categories. We find that 36 mRNAs decay exponentially in a good approximation (black curves in figure 3.3). Thus, only in a minority of the cases a description with a constant decay rate is valid. For 69 mRNAs we conclude that they decay slowly in the beginning of the experiment followed by a faster decay at larger times (blue). Moreover, the majority of the mRNAs, namely 210 (red), decay faster than exponential, *i.e.* a very rapid decay at short times is followed by a more moderate one at later times.

All of our theoretical curves improve the fitting error compared to an single-step model. This is clear since we introduced more parameters and the shape of the decay pattern can be more accurately described.

In the following, we want to discuss two exemplary cases, where each one of them represents one of the non-exponential categories in figure 3.3.

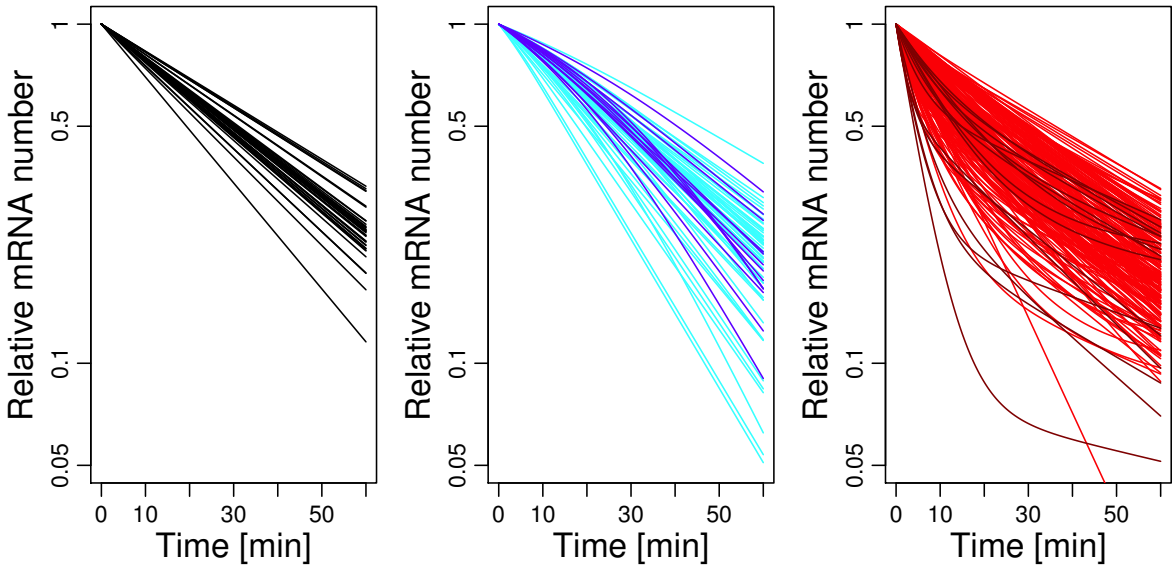

Figure 3.3.: **mRNA decay patterns in *S. cerevisiae*.** The patterns are obtained from a systematic fit of our model, *i.e.* Eqs. (3.2) and (2.31), to the experimental data from ref. [27]. The curves show the theoretical decay patterns that minimize the deviation between theory and experiment. Note that the experimental data points are omitted here for better legibility. The left panel shows 36 fitted curves that decay exponentially in good approximation. Conversely, 69 curves show a moderate decay followed by a fast decay (central panel) and 210 curves decay rapidly at short times and then level off (right panel). Hence, this shows that the purely exponential decay is only one of several possible decay patterns. For this visualization of different categories we considered data that were *bona-fide* (see above) in a good approximation and with a very low residual sum of squares $r^2 < 0.005$ obtained from the model fit. Moreover, in the central and right panel we highlighted a number of decay patterns that display a strong contrast to an exponential decay.

3.2.4. Exemplary decay patterns

In figure 3.4 A, we analyze two instructive examples from the experimental decay patterns in figure 3.3: The mRNA encoding *MGS1* (Maintenance of Genome Stability 1, red) and the constitutive acid phosphatase precursor (*PHO3*, blue). A comparison between the best fit with an exponential function and the best fit with our model shows that the latter is clearly more accurate than a single exponential in describing the decay profiles over the entire time course of the experiment. Furthermore, in the case of *MGS1* mRNA, the fit reveals a half-life $t_{1/2}$ which is substantially smaller than the estimated half-life from an exponential fit. This shows also that a measurement of the half-life $t_{1/2}$ is, in general, not a good measure of the average lifetime $\langle U \rangle$ of the mRNAs and thus fails to predict the decay rate accurately. In fact, a constant decay rate does not exist at all if the lifetime distribution is not exponential (see Eq. (2.27)).

3.2.5. Age-dependent degradation rates

Figure 3.4 A reveals, indeed, that for *MGS1* degradation of the corresponding mRNA becomes less efficient during the lifetime of the molecule. Conversely, for *PHO3* mRNA

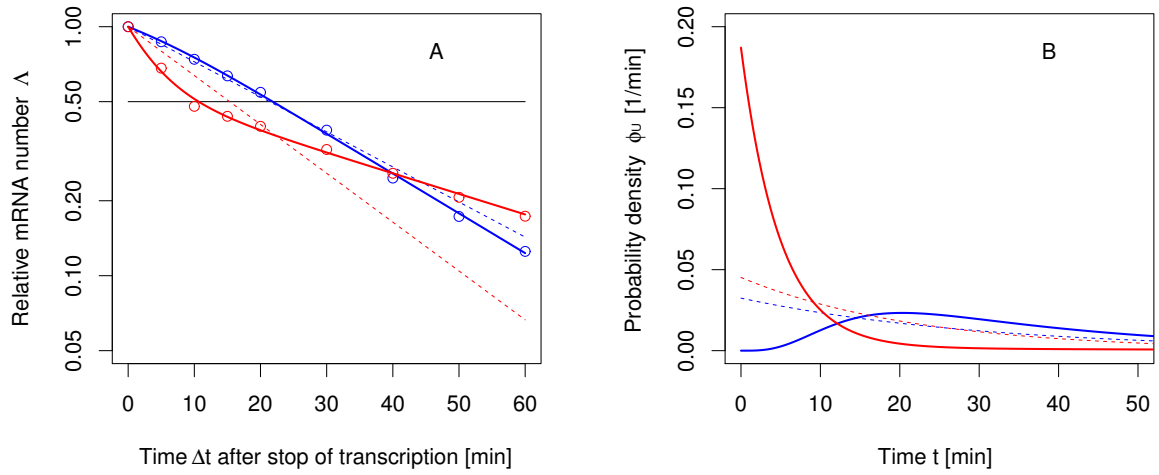

Figure 3.4.: **mRNA decay patterns and lifetime distributions**. (A) Two different experimental decay patterns are reproduced from [27] (circles) corresponding to the genes *MGS1* (red) and *PHO3* (blue) of *S. cerevisiae*. The solid lines represent decay patterns as calculated by the Markov chain model via Eq. (3.2) and the corresponding rates are estimated from a non-linear regression analysis (see section 3.2.3 above). For comparison, we also show a fit with a simple exponential function (dashed lines) which is clearly not suitable to capture the entire information of the degradation process. In particular, the exponential fit for the red data points suggests a half-life (intersection with the horizontal line) that is almost twice as large as the true half-life. (B) The corresponding lifetime densities ϕ_U are derived from the fitted rates by inverting Eq. (3.2). Evidently, both densities differ strongly from an exponential distribution indicated by the dashed lines. While the red line shows a rapidly decaying lifetime distribution, the blue line is broadly distributed, with a clear maximum at an intermediate time.

degradation becomes effective only after a transient time. In both cases we conclude that the effective degradation rate ω_{deg} depends on the age a of the mRNA. This relationship can be expressed as Eq. (2.27) in section 2.2,

$$\omega_{\text{deg}}(a) = \frac{\phi_U(a)}{1 - \Phi_U(a)}\,,$$

which is often referred to as the hazard or failure rate in the literature [30]. Figure 3.5 illustrates the age-dependence of the degradation rates for the curves in figure 3.4. Clearly, the degradation rate varies greatly during the lifetime of the depicted mRNAs. Indeed, it can either decrease or increase, depending on the particular mechanism that is responsible for degradation.

3.2.6. Signature of aging in mRNA decay experiments

In chapter 2 we have introduced the residual lifetime of an mRNA. It allows to characterize the aging of mRNAs sampled from a cell population at any time interval Δt after the

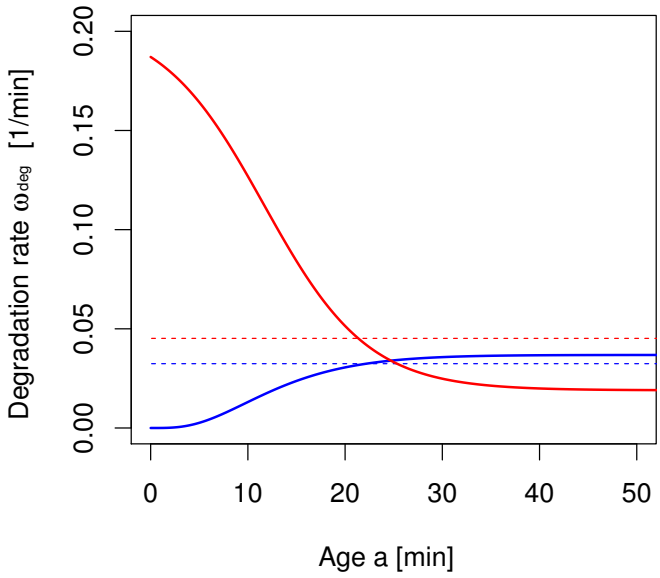

Figure 3.5.: **Effective degradation rate $\omega_{\mathbf{deg}}(\mathbf{a})$ as a function of the age a of an mRNA.** The lifetime distribution of an mRNA can be translated into an age-dependent degradation rate $\omega_{\mathrm{deg}}(a)$ via Eq. (2.27). Here, we illustrate the change of the degradation rate during the lifetime of an mRNA for the two decay patterns shown in figure 3.4. For the mRNA encoding *MGS1* (red), the degradation rate is high for young mRNAs and decreases to a constant value after some transient time. In contrast, for *PHO3* mRNA (blue), the degradation rate is close to zero upon birth of the mRNA and increases gradually to a constant value. For comparison, the constant rates corresponding to a fit of the decay data with purely exponential functions (dashed lines) are also included.

stop of transcription. For the present case, we are interested in the limit case $t_s \to \infty$ of Eq. (2.18),

$$\phi_R\left(r \mid \Delta t\right) = \frac{1 - \Phi_U(\Delta t + r)}{\int_{\Delta t}^{\infty} \mathrm{d}t \; (1 - \Phi_U(t))} \,. \tag{3.6}$$

A simple integration of this quantity shows that if ϕ_U is exponential then $\phi_R = \phi_U$. This is trivial due to the memoryless property of the exponential distribution. Nevertheless, for any other functional form of ϕ_U, we find that ϕ_R is a non-trivial function. The average residual lifetime is determined by the integral

$$\langle R(\Delta t) \rangle \;=\; \int_0^{\infty} \mathrm{d}r \, r \, \phi_R\left(r \mid \Delta t\right) \,, \tag{3.7}$$

for any $\Delta t \geq 0$. As we shall see, this expression enters into the determination of the residual protein synthesis capacity.

The residual lifetime of a given mRNA is the remaining time until it is degraded. The average residual lifetime in a sample of mRNA can be easily computed both at the beginning of the experiment, namely at steady state, and during the decay assays if one knows

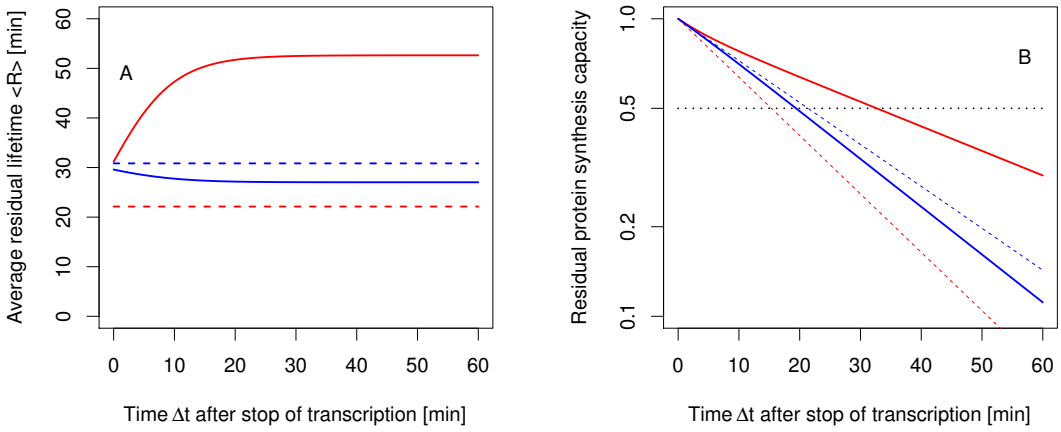

Figure 3.6.: **Average residual lifetime and residual protein synthesis capacity**. In panel A we see the evolution of the residual lifetime as a function of the interval Δt following the interruption of transcription for *MGS1* (red) and *PHO3* mRNA (blue). Under steady state conditions, *i.e.* $\Delta t = 0$, both have similar residual lifetimes. However, when the synthesis of mRNAs is stopped, the remaining mRNA population ages. Hence, the residual lifetime of *MGS1* mRNA increases since the remaining mRNAs have been stabilized (see figure 3.5). On the contrary, for *PHO3* mRNA the average residual lifetime decreases. Only for exponentially distributed lifetimes (dashed lines) the average residual lifetime stays constant. This is due to the special memoryless property of the exponential distribution. In panel B we plot Eq. (3.8), which is proportional to the amount of proteins that will be produced by an average mRNA from the sample. In this semi-log plot we see that the residual protein synthesis capacity decays exponentially if the mRNA has an exponential lifetime distribution (dashed lines) but decays differently if the process of degradation is more complex.

the residual lifetime distribution. In figure 3.6 A we show the behavior of the average residual lifetime $\langle R(\Delta t)\rangle$ as a function of the time interval Δt following the interruption of transcription for the two mRNAs discussed in figure 3.4. We see that the average residual lifetime changes with time due to the aging of the mRNA population after the stop of transcription. This is a consequence of the non-constant degradation rate in figure 3.5. Because the remaining mRNAs are older, their average degradation rate has changed and, thus, the average residual lifetime can increase (*MGS1*) or decrease (*PHO3*). Only mRNAs with exponentially distributed lifetime show no aging and a constant residual lifetime (dashed lines).

The residual lifetime affects the number of proteins that can be synthesized by the remaining mRNAs at any given point in time. Thus, the residual lifetime enters in the assays aimed to measure the so-called *functional lifetime* which is an additional experimental method to determine the stability of mRNAs [31, 46]. If one assumes that aging does not affect the ribosomal traffic, the residual protein synthesis capacity of an average mRNA still present at a time Δt after the stop of transcription can be defined as

$$\mathcal{C}(\Delta t) \equiv \frac{N_r(\Delta t)\,\langle R(\Delta t)\rangle}{N_r(0)\,\langle R(0)\rangle}\,. \tag{3.8}$$

It is proportional to the number of remaining mRNAs and their corresponding average residual lifetime at any time Δt after the interruption of transcription. Note that we normalized Eq. (3.8) such that $\mathcal{C}(0) \equiv 1$.
The plot of the residual protein synthesis capacity is shown in figure 3.6 B. Note that only for exponentially distributed lifetimes the synthesis capacity follows an exponential decay since only in this case $\langle R(\Delta t) \rangle$ is a constant. Conversely, for more complex degradation processes, mRNA stability experiments based on the functional lifetime are affected by the same issues as the other procedures discussed above. Indeed, as figure 3.6 B indicates, the residual protein synthesis capacity does not decay strictly exponential for *MGS1* and *PHO3* mRNA.

Short summary

In this section, we have learned how the general theory of mRNA turnover as introduced in chapter 2 can be linked to mRNA decay experiments. We have devised a strategy based on Markov chains. Transient biochemical modifications of the mRNA molecules are mapped to different states and the degradation rate from each state can be different. We could improve the analysis of an experiment on *S. cerevisiae* by finding more accurate decay patterns. We have obtained different categories of decay patterns that clearly show that the exponential decay is only an exceptional case. Furthermore, we could show how additional information, the residual lifetime and the residual protein synthesis capacity, can be obtained from the analysis of the age-dependent degradation rate.
All the material presented in sections 3.1 and 3.2 is original and is presented in a manuscript currently under review.

3.3. Model of endonuclease in bacteria: Effect of ribosome shielding

In the following, we develop a model of degradation that is designed to match the biochemical mechanisms in *E. coli*. It relies on the general formalism we developed in chapter 2 and on the general findings of section 3.1. However, here we will pursue the inverse approach to define the lifetime distribution function $\Phi_U(t)$. We will build a mechanistic model on the basis of an age-dependent degradation rate that allows us to obtain $\Phi_U(t)$.

3.3.1. Introduction and motivation

In contrast to eukaryotes, in bacteria such as *E. coli* degradation is known to proceed primarily endonucleolytically. Therefore, one might argue that in this case a first-order kinetic model would be sufficient to describe the experiments. However, also in several experiments in bacteria it has been found that exponential decay is only one of many different decay patterns [28, 29]. Furthermore, also in bacteria it is persuasive that the mRNA molecules are subject to aging. In particular, in *E. coli* the enzyme *RNase E* and its homologues are believed to initiate the main pathway of mRNA degradation [14]. The action of *RNase E* consists of cleaving the mRNA in the coding region. Once this has been accomplished, a series of further steps including exonucleolytic digestion and endonucleolytic cleavage follows. Since the subsequent steps are fast, it is the initial step that determines the stability of mRNAs.
The activity of the ribonucleases is, however, not independent of the translational state of the mRNA. First, it was shown experimentally that the stability is dramatically improved when the mRNA is well covered by ribosomes mRNA [47, 48, 49] because the coverage by the ribosomes shields the mRNA from the action of the endonucleolytic enzymes such as *RNase E*. Secondly, ribosomal coverage of mRNA is itself a time dependent process. It takes indeed a certain amount of time until the whole mRNA is covered by ribosomes and thus to build a stable polysome [50, 42, 51]. During this period, which may be affected by the degree of post-transcriptional translation, the stability of the mRNA changes with time.

3.3.2. Model of endonucleolytic decay

A process known to influence the stability of the mRNA is the size of the polysome [47, 48, 52, 53]. In the process of translation, ribosomes bind to the initially empty (or naked) mRNA and start translation, thereby translocating the mRNA one codon at a time. Only after a transient time t_L the first ribosome has reached the stop codon and the mRNA can be seen as fully covered by ribosomes. Hence, when mRNAs are younger than t_L, the leading part of the mRNA is uncovered by ribosomes. In a first approximation, this time is determined by the ratio of the sequence length L of the mRNA and the average elongation rate v of the ribosomes, *i.e.* $t_L = L/v$. Note that in bacteria it is possible that ribosomes bind to the mRNA prior to the completion of transcription [54]. In these cases, the time span t_L to a full coverage would be reduced. In section 3.3.3 we will further discuss this possibility (see below).

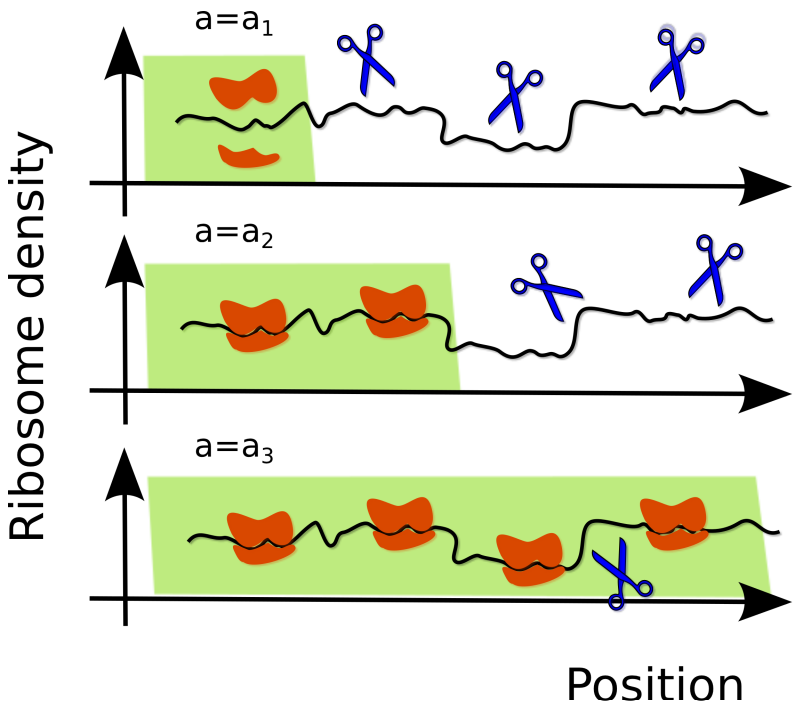

Figure 3.7.: **Illustration of the shielding effect**. The nascent chain is initially not protected by ribosomes and it is therefore more vulnerable to endonucleases. Once the process of translation has started, the ribosomes move along the chain and a larger fraction of the mRNA is (partially) shielded. Only after a transient time the entire chain is covered by ribosomes and hence less vulnerable to endonucleolytic degradation. The three displayed mRNAs are of age $a_1 < a_2 < a_3$, respectively.

The transient ribosomal coverage of an mRNA becomes important when degradation is primarily determined by endonucleolytic attack, such as for example in many bacteria [14], because the ribosomes would make it more difficult for the degrading molecules to bind to the mRNA. In these cases, the degradation rate is not constant during the lifetime of an mRNA but changes according to its age until the ribosomal coverage has reached the steady state (see figure 3.7). The reason is that upstream of the leading ribosome the mRNA is not entirely protected from ribonucleolytic attacks.

To model this phenomenon, we assign a degradation rate ω_a to the more vulnerable, uncovered part in front of the leading ribosome and a rate ω_b to the partially covered part behind the leading ribosome. The effective degradation rate of the mRNA depends therefore on the relative size of each part and in the time interval t_L it evolves from the rate ω_a to a constant rate ω_b. Thus, we have mapped the effect of the dynamic shielding of mRNAs by ribosomes into an age dependent degradation rate

$$\omega_{\text{deg}}(a) = \begin{cases} \omega_a (1 - a/t_L) + \omega_b \, a/t_L & a < t_L \\ \omega_b & a \geq t_L \ , \end{cases} \tag{3.9}$$

where a denotes the age of the mRNA.

In chapter 2 we have already demonstrated how such non-constant rates are related to the

lifetime probability distribution. According to Eq. (2.26) in section 2.3 we can compute

$$\Phi_U(a) = 1 - \exp\left(-\int_0^a d\tau\, \omega_{\deg}(\tau)\right) , \tag{3.10}$$

which, once inserted in Eq. (3.2), provides the decay pattern of the mRNA.

Confirmation of the mechanism with TASEP simulations

One would expect that the two rates ω_a and ω_b are interrelated since in both cases degradation is mediated by the same endonucleolytic enzymes. In a first approximation, the rate in the covered region ω_b is reduced by the fraction ρ of mRNA that is physically covered by ribosomes, *i.e.* $\omega_b \sim \omega_a(1-\rho)$. Experiments suggest that this density is usually relatively low [55], leading to a reduction of the degradation rate by not more than 30 per cent due to the physical presence of ribosomes.
There are, however, a number of features that the theory based on the rates given in Eq. (3.9) implicitly ignores. First, it neglects the dynamics of the translation process and ignores the interaction between ribosomes as for example the possibility of traffic jams. Secondly, the size of the degrading enzymes, its interaction with the ribosomes and the varying distances (*i.e.* the available cleavage area) between the ribosomes are not properly considered. To elucidate the role of these effects, we have performed *TASEP* simulations [42, 41], where the motion of ribosomes along an mRNA is modeled as a stochastic stepping on a linear, homogeneous chain (for more details about the simulation technique see sections 2.5 and A.6). The ribosomes are considered as extended objects occupying 10 codons each and their interaction is governed by a hard core repulsion, *i.e.* two ribosomes cannot occupy the same spot.
The simulation of degradation proceeds as follows. At each instant of the simulation, either a ribosome moves or the mRNA is degraded. Which decision is taken - stepping or degradation - depends on the random occurrence of events associated to the corresponding rates. The stepping rate of the ribosome is given by an elongation rate - just as in a regular TASEP simulation. Furthermore, the degradation rate changes in time and is proportional to the total available cleavage area. Finally, whenever an mRNA is selected for degradation, the simulation is stopped. The potential cleavage area is defined by the part of the mRNA that is not covered by ribosomes. However, only stretches of mRNA of a certain minimal length contribute to the cleavage area, which we set in the simulations to 20 codons. This accounts for the size of the degrading enzymes and the steric repulsion between enzymes and ribosomes.
As pointed out above, in the construction of our mechanistic model via Eqs. (3.9) and (3.10), we had to ignore some of the dynamic properties of ribosome shielding. The more detailed approach based on the computer simulations shows that the degradation rate fluctuates over time with a different time evolution in each realization of this process, whereas each realization corresponds to one single mRNA molecule (figure 3.8 A). The average over a large number of realizations yields a time-dependence of the degradation rate such as in Eq. (3.9), *i.e.* it decreases linearly until at a time t_L it approaches a constant value (see figure 3.8 A). Furthermore, each realization of the simulation has a random run length which represents the lifetime of the mRNA. Thus, from a large sample

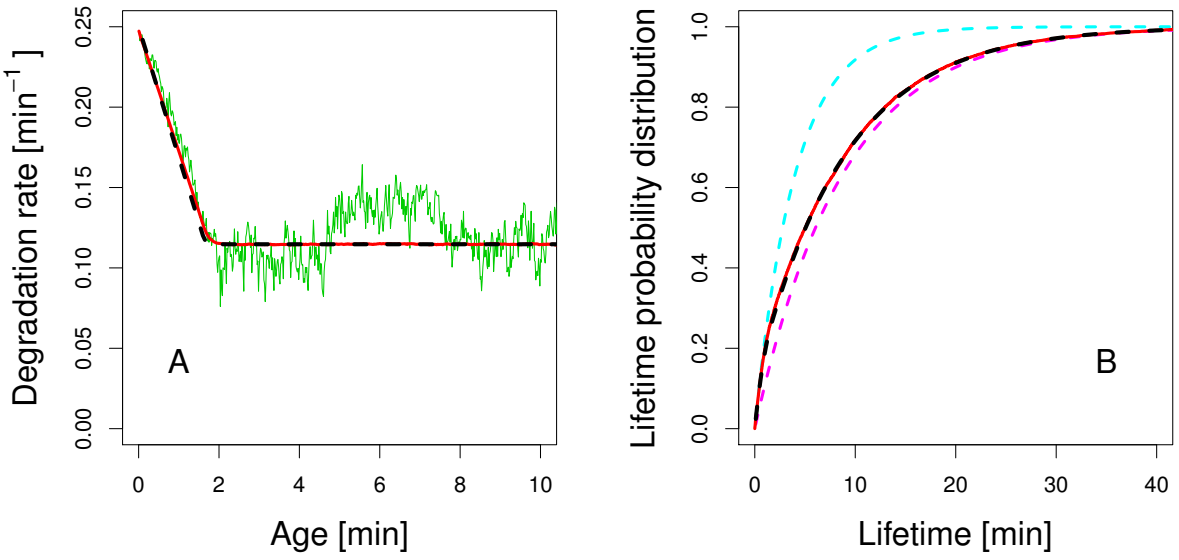

Figure 3.8.: **Time-dependent degradation rate and corresponding lifetime distribution**. During the lifetime of an mRNA its degradation rate is not constant due to the dynamic coverage with ribosomes. At each instant of a TASEP simulation, the positions of the ribosomes are different and hence the shielding of the mRNA changes with time. Consequently, panel A shows that in a single realization of the stochastic process the degradation rate fluctuates (green). Moreover, the rate tends to decrease as more ribosomes bind to the mRNA. An average over 10000 TASEP simulations shows that after a transient time mRNAs reach a steady state ribosome coverage and attain a constant degradation rate (red). Compared is the time-dependent degradation rate as given in Eq. (3.9) with $\omega_a = 0.25\,\mathrm{min}^{-1}$ and $\omega_b = 0.115\,\mathrm{min}^{-1}$ (dashed black line). Both degradation rates are in very good agreement with each other which allows to conclude that an analytic mean field description according to Eq. (3.9) is suitable to describe the average dynamic shielding effect as obtained from the simulations. Here, we consider an mRNA with a length of 1000 codons, an initiation frequency $\omega_{\mathrm{on}} = 0.33\,\mathrm{s}^{-1}$, average ribosome velocity $v = 10$ codons s^{-1} and initial degradation rate $\omega_a = 0.25\,\mathrm{min}^{-1}$. In Panel B, we compare the theoretical lifetime distribution as given by Eq. (3.10) (black dashed) with the random lifetimes obtained from the stochastic simulation (red). For all lifetimes, both curves completely agree. Moreover, for short lifetimes the distribution is described well by the single rate constant ω_a (cyan) whereas for large lifetimes by the rate constant ω_b (magenta).

of realizations we can compute the lifetime probability distribution $\Phi_U(t)$. Figure 3.8 B shows that the theoretical distribution as given by Eq. (3.10) completely agrees with the results obtained from simulation. In the same figure, we show that the cumulative distribution can be approximated by the single rate ω_a for short lifetimes and by ω_b for large lifetimes. These model curves are obtained from Eq. (3.10) with constant decay rates ω_a and ω_b, respectively. Nevertheless, the distribution for all lifetimes can only be understood with a time-dependent degradation rate as given in Eq. (3.9).

Hence, the assumptions made in Eq. (3.9) are justified, as it gives the correct average behavior of the stochastic simulation that was built to mimic the dynamic shielding effect of ribosomes in more detail. Therefore, in the following we will use the mean-field model described by Eq. (3.9) to compare our mechanistic model systematically to experimental data.

3.3.3. Comparison to experiment

With the functional form of the time dependent degradation rate Eq. (3.9), we are able to compute the decaying amount of mRNAs after interruption of transcription, $\Lambda(\Delta t)$, via Eqs. (3.2) and (3.10). The function $\Lambda(\Delta t)$ is a function of the length of the coding sequence L, the average elongation rate v and the two unknown degradation rates ω_a and ω_b.

We compared the theoretical decay curves arising from our model with experimental data from *E. coli* [28] via a nonlinear regression analysis, as described in section 3.2. For the subsequent analysis, we assume an average elongation rate $v = 600$ codons per minute which is a typical value for *E. coli* [56, 57]. The length of the coding sequence L is obtained from the genome analysis of *E. coli* [58]. Similar to the data obtained from yeast, part of the data bears obvious errors - for some genes transcription seems not to be fully inhibited since the mRNA levels are rising. Additionally, as we have pointed out already in section 3.1, the decay pattern should exhibit a positive curvature. Therefore, we restrict the current analysis to data that meet these criteria within a limited margin of error. For each gene, we minimize the residual sum of squares between the experimental decay data and the theoretical curve given by Eq. (3.2). Thereby, we obtain the previously unknown degradation rates ω_a and ω_b. For comparison, we also perform the same analysis with a constant degradation rate.

In figure 3.9, we show three exemplary cases of the decay patterns obtained from a fit to experimental data. Clearly, the model fit based on ribosome shielding, Eqs. (3.9) and (3.10), (solid lines) is much more accurate than a description with constant rate (dashed lines). The three examples represent mRNAs of different length. Hence, they need a different time interval t_L until reaching a steady state ribosome coverage. Only for times larger than t_L the decay follows an exponential curve with constant rate ω_b.

In figure 3.10 we give an overview of the variety of decay patterns that we have obtained through the fitting process. From the 220 mRNAs analyzed here, only 74 decay exponentially. In most cases it was found that a steep decay at short times is followed by a slower decay at later times. Thus, as figure 3.10 reveals, the ribosome shielding model yields a much larger variety of decay patterns as compared to a model based on a constant degradation rate. Moreover, our improved model leads to a considerable reduction of the fitting error. On average, the residual sum of squares decreases by almost one order of magnitude compared to a model based on a constant decay rate.

Furthermore, since the shape of the decay pattern is more accurately accounted for in our theoretical description, we observe also changes in the half-life. Figure 3.11 A reveals that there is a systematic decrease of the half-life if we apply our analysis instead of a simple exponential fit. Consequently, the average half-life of the entire analyzed data set reduces from 4.40 min to 3.99 min. This observation becomes intuitive when we consider figure 3.9 again. In all three cases, the fit with a constant rate systematically overestimates the half-lives.

However, since most decay patterns are not exponential, the half-life of an mRNA alone is not sufficient to capture the entire detail of its decay. Whereas the average lifetime of an mRNA is given by Eq. (2.3), its half-life is given implicitly by Eq. (3.3). Only in the case of a constant decay rate, the average lifetime and half-life are linked via the simple relation $t_{1/2} = \langle U \rangle \cdot \log(2)$. The analysis of the experimental data reveals that the shielding

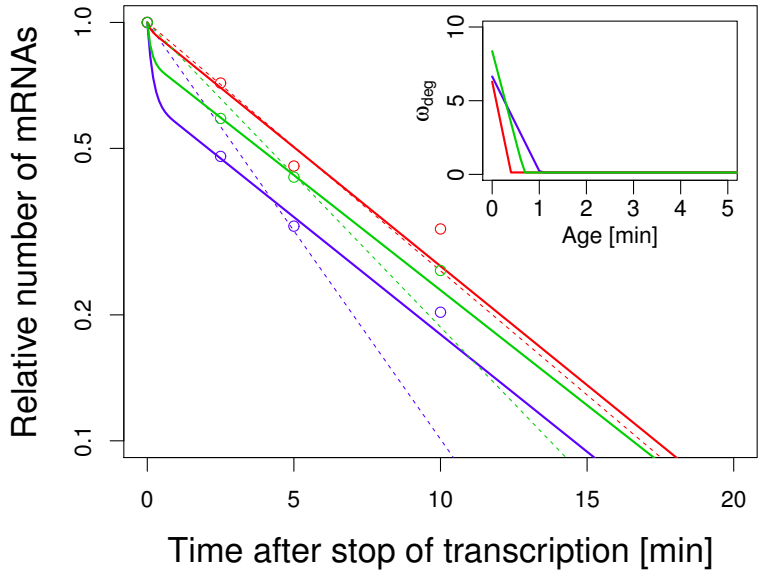

Figure 3.9.: **Three examples of experimental mRNA decay in *E. coli* .** The data points display the relative mRNA amount after inhibition of transcription (on logarithmic scale) adapted from [28] (circles). Theoretically, the decay is described by the function $\Lambda(\Delta t)$ given in Eq. (3.2) together with the age-dependent degradation rate defined in Eqs. (3.9) and (3.10). A nonlinear regression analysis minimizes the deviation between the experimental data and the theoretical description and determines the previously unknown degradation rates ω_a and ω_b. The fitted model curves (solid lines) yield an improved description of the decay data compared to a simple exponential fit (dashed lines). In all three cases, the residual sum of squares is reduced by at least one order of magnitude. The curves show the decay of *atoS* (blue), *fabB* (green) and *ykgE* (red) mRNA. The mRNAs are of different length L (609, 407, 240 codons, respectively) and consequently they remain in the transient period for different time intervals $t_L = L/v$. Therefore, the shielding effect is more noticeable for longer mRNAs. For the present example, this leads to a smaller half-lives for longer mRNAs. In the inset, the corresponding change of the degradation rate ω_{deg} is shown. Starting from a high value ω_a, it drops to a constant, small ω_b at age t_L.

model yields much smaller average lifetimes than a model with constant degradation rate (figure 3.11 B). Indeed, the average of the lifetimes of all mRNAs is 2.6 min, whereas the estimate in the simple exponential model is 6.3 min. Therefore, although in this case an exponential decay law might be able to predict a reasonably accurate half-life, it gives a poor estimate of the average lifetime of an mRNA.

One important result of the fit of the theory to the experimental data is that the uncovered rate ω_a is always larger than the covered rate ω_b. In fact, the majority of the rates differ by at least an order of magnitude. To better appreciate this point, one should recall that the fitting procedure is made without any bias concerning the values of these two rates. If the fitting with the data would give two similar or identical rates, *i.e.* $\omega_a = \omega_b$, one could infer that the suggested transient phase of ribosomal loading plays no role in the process of degradation. This would be the case if the ribosomes followed the RNA polymerase closely during the process of transcription. However, from our analysis we find that only in the minority of the cases the two rates are similar.

The presence of co-transcriptional translation reduces the time span t_L. This time would

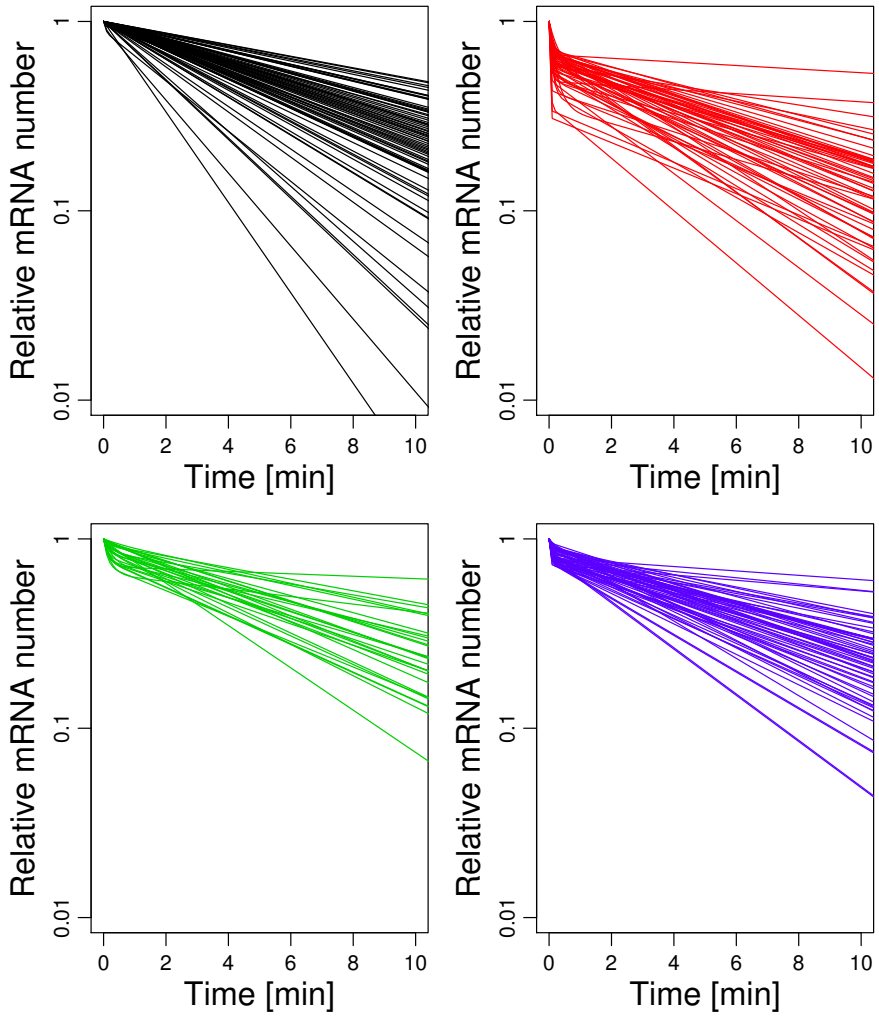

Figure 3.10.: **Categories of decay patterns in *E. coli*** . Shown are the theoretical decay patterns obtained from a systematic fit of the model of ribosomal shielding to experimental data of *E. coli* [28]. For clarity, the experimental data points were omitted. We could distinguish four different categories of decay patterns. For 74 mRNA species one can conclude an exponential decay (black, top left). All other patterns deviate from an exponential decay and display a faster decay for shorter times followed by a more moderate, exponential decay at larger t. In particular, the red patterns have a very short transient followed by a sharp transition to a moderate decay (top right). The 57 mRNAs of this category are mainly short and need thus only a short time t_L before reaching a stationary ribosome coverage. Conversely, for long mRNAs this transient is particularly long (bottom left). The 29 mRNAs of this category are longer than 400 codons. Another typical category of decay patterns is shown in the bottom right corner (60 in total). Many of the mRNAs in this category are of intermediate and some of short length. For the current analysis, we considered all experimental decay patterns that show a nearly *bona-fide* behavior (see above). Furthermore, we chose those fitted curves that have a residual sum-of-squares $r^2 < 0.05$, leading to 220 mRNAs. The choice of categories is by no means unambiguous but should give an overview over different types of patterns. We have discussed some characteristic examples in more detail in figure 3.9.

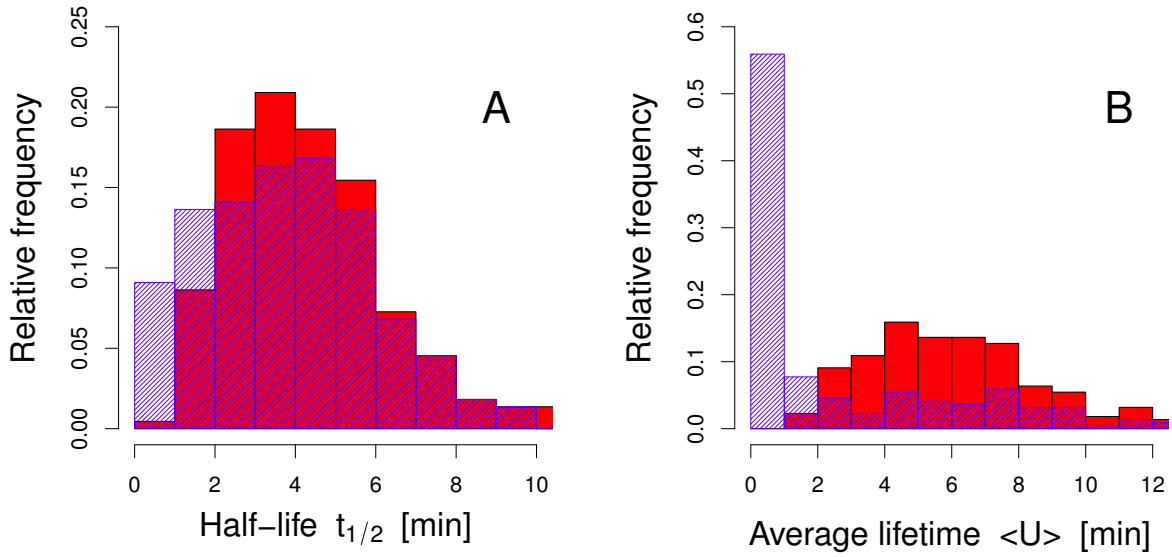

Figure 3.11.: **Distribution of half-lives $t_{1/2}$ and mean lifetimes $\langle U \rangle$.** Panel A compares the distribution of the half-lives when fitting the data with a purely exponential model (red) and the ribosome shielding model (blue). The latter is shifted to smaller times resulting in a reduced mean half-life. Panel B shows a comparison of the distribution of the average lifetime for the two models across the 220 decay data. The ribosome shielding model leads to much smaller average lifetimes because many mRNAs are degraded shortly after their synthesis.

be governed also by the elongation speed of the RNA polymerase and the time delay between the binding of the polymerase and the first ribosome. We can in principle extend our theoretical description to account for that - however at the cost of having to introduce more parameters to the description (see appendix section B.4). When we compare such a theory to the experimental data, the reduction of t_L would essentially lead to a higher rate ω_a. In the extreme case that t_L were close to zero, such a description would fail to reproduce the shape of the experimental decay patterns.

Our theory has important implications for the productivity of mRNAs. It is usually thought that each mRNA contributes to the synthesis of proteins. However, according to the discussion outlined above, only mRNAs that live longer than a time t_L produce proteins. Indeed, the lifetime probability function $\Phi_U(t_L)$ given in Eq. (3.10) determines the fraction of mRNAs that are degraded until time t_L. With the experimentally determined rates we find that for many mRNA species a considerable amount of mRNAs produces no protein at all.

In different experiments and organisms, different correlations between the length of an mRNA and its stability were found [59, 60, 61]. Our analysis of the data from Ref. [28] reveals no simple correlation between length and stability. This might not be too surprising because each mRNA species is regulated individually thus giving rise to the broad range of lifetimes shown in figure 3.11 B. However, those mRNAs that are regulated similarly, *i.e.* they have the same degradation rates ω_a and ω_b, will have different stability due to their length. One such example is shown in figure 3.9. The three example mRNAs vary in length from 240 to 609 codons but have similar degradation rates. Since for longer mRNAs a full ribosome coverage is reached at a later time point t_L, the initial unprotected

phase is also longer. Hence, in this case, longer mRNAs exhibit a shorter lifetime.

Section conclusion

In this section, we have presented another application of the general formalism that we derived in chapter 2. We have translated a dynamic property, the shielded part of an mRNA, to an age-dependent degradation rate, which is linked to the experimentally accessible decay pattern. The analysis of a decay experiment could be substantially improved and the empirical finding could be related to a mechanistic model.

3.4. Chapter summary

In this chapter, we have applied our theory to mRNA decay experiments. In a first part we have formulated some general findings that hold for any particular description of degradation. Key to a successful description is the formulation of a suitable lifetime distribution which reflects the molecular mechanisms of degradation. The two main parts of this chapter deal with degradation in different organisms. The first part is about degradation in eukaryotic cells such as yeast where various biochemical mechanisms compete. Additionally, the mechanisms typically require multiple modifications of the mRNA molecule before degradation. Therefore, we have chosen a description based on a Markov chain that improves the analysis of the experimental data and links the larger variety of mechanisms to the more complex degradation pathway. A systematic analysis revealed that our modeling approach greatly enhances the interpretation of an experiment on *S. cerevisiae* and furthermore determines the previously unknown rates (see figure 3.4).
In the second part, we investigate degradation in *E. coli* bacteria where the rate-limiting step is degradation initiation via endonucleolytic cleavage. Here we have pursued a different approach to refine the description. The process of translation is known to interact with degradation. The dynamics of this interaction is mapped to an age-dependent degradation rate which in turn fixes the lifetime distribution. We found that also this model considerably improves the analysis of experimental data (see figure 3.9). The two different modeling approaches are two interpretations of how one can determine a suitable lifetime distribution. They are both open models that can be adapted to more details of the biochemistry of degradation. Apart from the experimental data, the entire material presented in this chapter is original work.

4. Transient mRNA expression after induction of transcription

In this chapter we will assess the central role of degradation in mRNA expression. We will see that the transient to a steady-state mRNA number depends critically on the degradation pathway. Therefore, we elucidate the cell's capability to react to stress or artificial induction. The foundations for the theoretical description were laid in chapter 2 and we will use the knowledge gained in chapter 3 about the decay patterns and mRNA lifetime distributions. In the first section, we will motivate the theoretical studies and review important findings from chapter 2. It is followed by a discussion of general findings and an outline of methods to experimentally validate the theoretical description. Throughout the main part of this chapter, we will compare the effect of different degradation pathways via two exemplary cases. On the one hand, we show the effect of different lifetime distributions on the mRNA number evolution after the induction of transcription. On the other hand, we discuss how different degradation pathways become manifest in the molecular aging of mRNAs. Next, we will present an extension of the theory that shows how the presented results can be generalized to a more typical situation in a cell where there is a preexisting amount of mRNAs before the induction of transcription. In a final part, we will use the theoretical concepts introduced here to interpret experiments on *S. cerevisiae* cells under stressed conditions.

4.1. Introduction

In cells, the transcription of a gene can be induced due to internal or external signals [2]. It is also common practice to artificially induce transcription of previously untranscribed genes located on a recombinant plasmid [62]. Furthermore, different environmental conditions can trigger a stress response of the cell which leads to a possible alteration of the transcription rate. Additionally, also the mRNA stability can change due to external factors. All these alterations of the transcription process lead to a change in the mRNA expression. This evolution is governed by the rate of transcription on the one hand and by the stability of the mRNAs on the other hand.

Starting from zero amount of mRNA of a given gene, after the induction of transcription there is a growth of the number of mRNA molecules. This process eventually leads to a stationary state, which reflects the balance between synthesis and degradation of mRNA. However, even if the average transcription rate per cell ω_{tc} is constant, the patterns of growth of the number of mRNAs depend on the choice of the lifetime density ϕ_U.

We will see in a later part of this chapter how we can formulate our theory for general initial conditions. In order to learn about the general transition to a steady-state which

gives the general response kinetics of cells, we will first consider the starting condition of zero amount of mRNA of a given gene.
As we have shown in chapter 2, according to Eq. (2.9) the distribution of the number of mRNAs at any time t after the start of transcription is given by

$$\Pr\{Y(t) = k \mid X(0) = 0\} \;=\; \frac{[\omega_{\mathrm{tc}} H(t)]^k \exp\left[-\omega_{\mathrm{tc}} H(t)\right]}{k!}\,,$$

where in Eq. (2.10) we introduced

$$H(t) \;=\; \int_0^t \mathrm{d}u\,(1 - \Phi_U(u))\,. \qquad ((2.10))$$

and $\Phi_U(u)$ is the probability distribution of U related to the lifetime density via Eq. (2.2). Thus, the average number N_r of mRNA molecules per cell can be written as

$$N_r(t) \equiv N_r^{\uparrow}(t) \equiv \omega_{\mathrm{tc}} \int_0^t \mathrm{d}u\,(1 - \Phi_U(u))\,. \qquad (4.1)$$

It describes the evolution of the average number of mRNAs in time and depends on the transcription rate ω_{tc} and the lifetime probability function $\Phi_U(t)$. The notation $N_r^{\uparrow}(t)$ should indicate that we consider the growing amount of mRNA after the induction of transcription.

4.2. General observations

Similar to the analysis for the decaying amount, we can infer directly that the pattern of mRNA evolution, Eq. (4.1), must obey certain constraints. First, the number of mRNAs must be continuously increasing and secondly the curvature of the pattern must be concave (negative).

$$\frac{dN_r(t)}{dt} \geq 0 \qquad \text{and} \qquad \frac{d^2N_r(t)}{dt^2} \leq 0\,. \qquad (4.2)$$

This results directly from the formulation given in Eq. (4.1) and the positivity of probability measures. For short times after the induction of transcription, the amount of mRNAs grows proportionally to the transcription rate, *i.e.* is not influenced by the degradation process. However, for all cases where the average lifetime $\langle U \rangle$ is finite, a stationary mRNA amount is always attained. Since (at least in all models assumed here) the lifetime density $\phi_U(t)$ decreases exponentially in the limit of large times, the steady state is also approached exponentially. As $t \to \infty$, it follows from Eq. (2.10) that the number of mRNAs reaches a steady state and its probability distribution is given by

$$p_k^{\mathrm{st}} \;=\; \frac{[\omega_{\mathrm{tc}} \langle U \rangle]^k \exp\left[-\omega_{\mathrm{tc}} \langle U \rangle\right]}{k!}\,, \qquad (4.3)$$

which depends only on the transcription rate ω_{tc} and the average lifetime $\langle U \rangle$ which was defined in Eq. (2.3). Hence, the average mRNA level in steady state reads

$$N_r^{\mathrm{st}} = \omega_{\mathrm{tc}} \langle U \rangle\,. \qquad (4.4)$$

In the following we will be interested in the transient to steady state.
There is an interesting symmetry between the growing mRNA number after induction of transcription and the decay of the mRNA after interruption of transcription. Consider an experiment where a metabolic label allows distinguishing between mRNAs generated before and after the incorporation of the label [63]. Since - on average - the number of mRNAs stays constant, the decay of the old mRNAs, $N_r^{\downarrow}(t)$, is exactly matched by newly generated mRNAs, $N_r^{\uparrow}(t)$. In our theoretical framework this can be expressed as $N_r^{\uparrow}(t) + N_r^{\downarrow}(t) = N_r^{\mathrm{st}}$ for all times t. Indeed, a closer inspection of Eqs. (2.10) and (2.15) (with $t_s \to \infty$) reveals

$$H(t)+H_2(t) = \int_0^t du\,(1-\Phi_U(u))+\int_t^\infty du\,(1-\Phi_U(u)) = \int_0^\infty du\,(1-\Phi_U(u)) = \langle U\rangle\,. \quad (4.5)$$

Thus, the two processes of transcription induction and mRNA decay are complimentary. This symmetry relation is illustrated in figure 4.1 B.

Remark on experimental observation Experimentally, the dynamics following the induction of transcription can be studied in different ways. On the one hand, many experiments are based on biochemical assays where the mRNA expression level in a population of cells is analyzed via a *microarray* or *RNA-Seq*. After an external stimulus or perturbation of the cells, the kinetic response of different genes is revealed by a time-course analysis [27, 64, 65]. However, one can only measure relative changes of the mRNA expression, *i.e.* the fold-change due to the perturbation. Nevertheless, the incorporation of a metabolic label allows distinguishing mRNAs generated before and after introduction of the label [66, 67, 63]. Thus one can directly follow the induction kinetics of mRNAs synthesized after a certain point in time.
On the other hand, optical methods based on fluorescently labeled mRNAs allow a real-time observation of mRNAs in single-cells [68, 69, 70]. Here, the sequence of an mRNA is altered such that it provides binding sites for *green fluorescent protein* (GFP). The induction kinetics can be followed for instance by *fluorescence recovery after photobleaching* (FRAP) [71, 68]. However, the attachment of the markers have been found to modify the stability of the mRNAs [72] - thus it is not an optimal technique to investigate the role of mRNA degradation in gene expression.
For the comparison of our theoretical model to experimental data in the final part of this chapter we analyze an experiment that is based on the first technique which reveals the relative change of the mRNA level after a stimulus. In section 4.6 we will therefore extend the theoretical apparatus to be able to understand also relative changes of the mRNA amount.

4.3. Time evolution of the number of mRNAs

From Eq. (4.1) it is clear that the evolution of the amount of mRNA after induction depends critically on the stability regulation of the particular mRNA species which becomes manifest in its lifetime density ϕ_U. In the previous chapter we pointed out that the complexity of the stability regulation leads generally to a more complex functional form of ϕ_U. It is directly related to the underlying mechanism of degradation of that mRNA species. In the following, we will consider two different exemplary forms of $\phi_U(t)$, namely the exponential lifetime density

$$\phi_U^{(\exp)}(t) \;=\; \omega_\mathrm{r} \exp(-\omega_\mathrm{r} t)\,, \tag{4.6}$$

with average lifetime $\langle U\rangle = \omega_\mathrm{r}^{-1}$. A straightforward extension of Eq. (4.6) is the gamma density

$$\phi_U^{(\Gamma)}(t) \;=\; \frac{\lambda(\lambda t)^{n-1}}{(n-1)!} \exp(-\lambda t)\,, \tag{4.7}$$

with shape parameter n and average value $\langle U\rangle = n/\lambda$. Note that for $n = 1$ the gamma density reconstitutes the exponential density. Whereas the lifetime density $\phi_U^{(\exp)}$ describes the decay of mRNA species in a simple first-order kinetic model, the density $\phi_U^{(\Gamma)}$ is related to a more refined model of mRNA decay where multiple successive biochemical steps are required for degradation. In the following, to be able to better compare the two degradation modes, we arbitrarily set $n = 5$ in Eq. (4.7) and fix λ and ω_r such that the average lifetimes $\langle U\rangle$ are identical for both distributions. Furthermore, for simplicity we will assume $\omega_\mathrm{tc} = 1\,\mathrm{min}^{-1}$ in most cases. Note however that our results do not depend on our particular choices of n, $\langle U\rangle$ and ω_tc. The distributions are shown in figure 4.1 and we will use the same color code to distinguish mRNAs with $\phi_U^{(\exp)}$ (red) and $\phi_U^{(\Gamma)}$ (blue) throughout this chapter. Later, in section 4.7 we will consider other forms of $\phi_U(t)$ in connection to the analysis of experimental data.

In figure 4.2 we see the mRNA number distribution for the two generic mRNAs, $\phi_U^{(\exp)}$ and $\phi_U^{(\Gamma)}$, at different times after the start of transcription. The distribution is given by a Poisson distribution at all times, however, the parameter of the distribution changes with time and is different for the two exemplary mRNA species. Note that in steady state it is solely given by the transcription rate ω_tc and average lifetime $\langle U\rangle$.

In the following, we analyze the evolution of the average number of mRNAs $N_r(t)$ over time as defined in Eq. (4.1). Figure 4.3 A shows the evolution of mRNAs with a gamma lifetime distribution as given in Eq. (4.7). The different curves represent mRNAs determined by $\omega_\mathrm{tc} = 1\ \mathrm{min}^{-1}$, $\langle U\rangle = 4$ min (solid), $\omega_\mathrm{tc} = 1\ \mathrm{min}^{-1}$, $\langle U\rangle = 8$ min (dotted) and $\omega_\mathrm{tc} = 2\ \mathrm{min}^{-1}$, $\langle U\rangle = 4$ min (dashed), respectively. The assumed transcription rates and mRNA average lifetimes resemble typical values for organisms such as *E. coli* but were chosen without loss of generality. Note that the steady state mRNA number is given by $N_r^\mathrm{st} = \omega_\mathrm{tc}\langle U\rangle$, *i.e.* a two-fold increase can be reached either by doubling the transcription rate (dashed lines) or the mRNA stability (dotted lines). Furthermore, in this example the transient time to steady state depends critically on the mean lifetime. Conversely, in figure 4.3 B, we compare mRNAs with different forms of the lifetime distribution, $\phi_U^{(\Gamma)}$ (blue) and $\phi_U^{(\exp)}$ (red), but same $\omega_\mathrm{tc} = 1\,\mathrm{min}^{-1}$ and average lifetime $\langle U\rangle = 4$

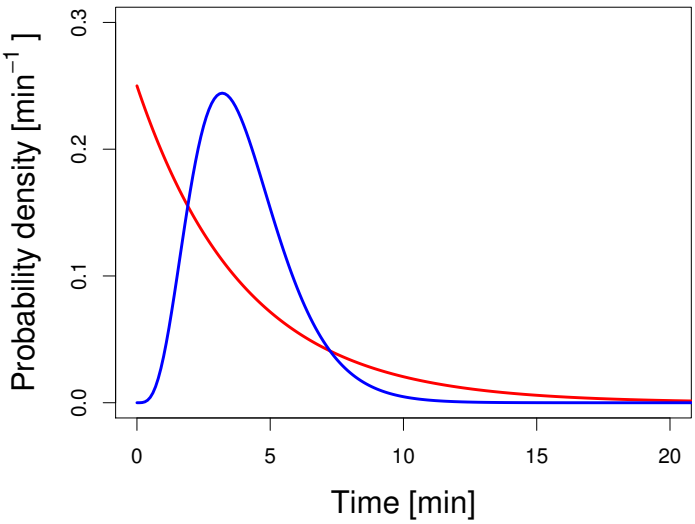

Figure 4.1.: **Model lifetime distributions**. Shown are two exemplary lifetime distributions, $\phi_U^{(\mathrm{exp})}(t)$ (red) and $\phi_U^{(\Gamma)}(t)$ with $n = 5$ (blue), as given in Eqs. (4.6) and (4.7). Both have the same average value $\langle U \rangle = 4$ min but different shape parameters n, giving rise to a higher variance of the exponential distribution (red) as compared to the gamma-distribution (blue). In the following, we will frequently use these two distribution to illustrate the effect of non-exponential distributions in transient phenomena of gene expression. Moreover, we will distinguish the two cases by using consistently the same color code.

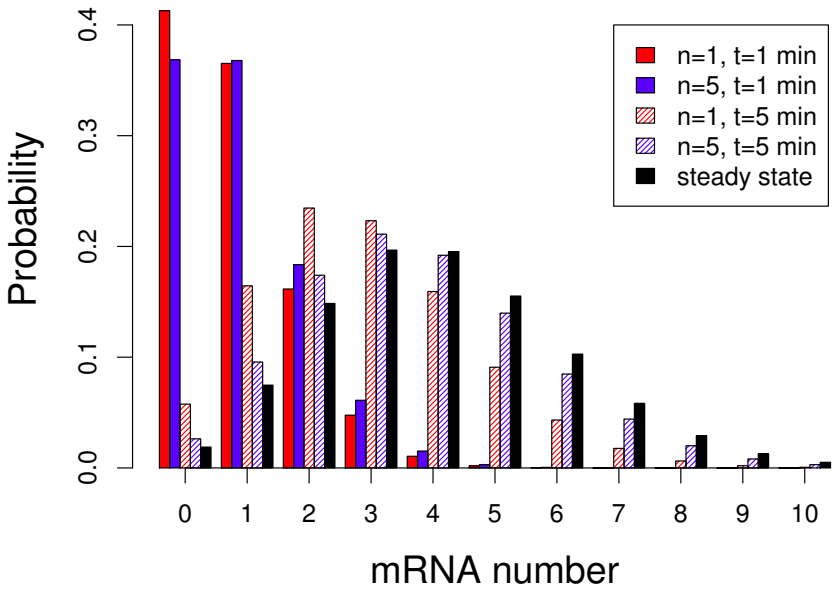

Figure 4.2.: **mRNA number distribution**. Shown are the mRNA number distributions, $\phi_U^{(\mathrm{exp})}(t)$ (red) and $\phi_U^{(\Gamma)}(t)$, with $\langle U \rangle = 4$ for different times after start of transcription (see figure legend). At all times, the mRNA number is described by a Poisson distribution - however with different parameter - as given by Eq. (2.10). Nevertheless, when approaching steady state, mRNAs with different lifetime distributions but same average lifetime give rise to the same mRNA number distribution (black bars).

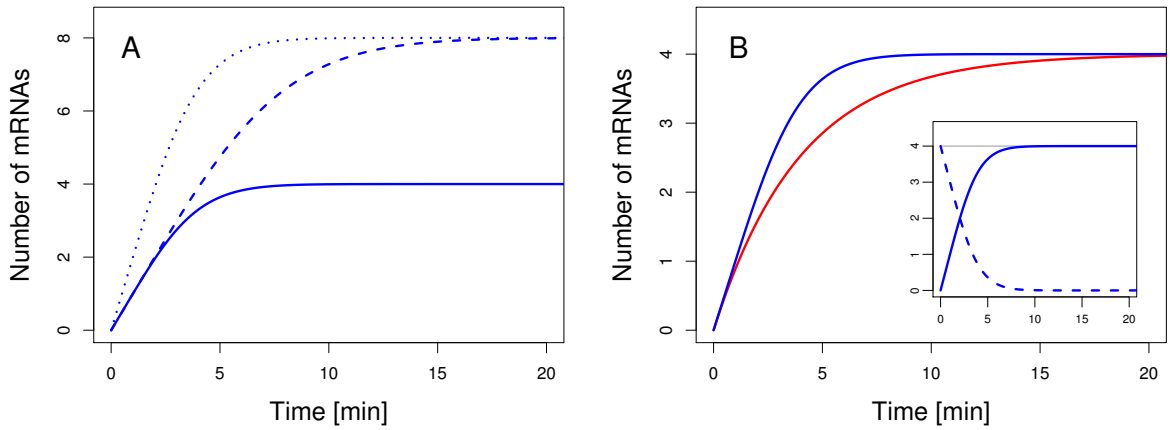

Figure 4.3.: **Time evolution of mRNA number after start of transcription**. In the left panel, we discuss the evolution of mRNAs with a gamma lifetime distribution according to Eq. (4.7). The curves represent mRNAs determined by $\omega_{\mathrm{tc}} = 1\,\mathrm{min}^{-1}$, $\langle U \rangle = 4$ min (solid), $\omega_{\mathrm{tc}} = 1\,\mathrm{min}^{-1}$, $\langle U \rangle = 8$ min (dotted) and $\omega_{\mathrm{tc}} = 2\,\mathrm{min}^{-1}$, $\langle U \rangle = 4$ min (dashed), respectively. Note that the steady state mRNA number is given by $N_r^{\mathrm{st}} = \omega_{\mathrm{tc}} \langle U \rangle$, *i.e.* a two-fold increase can be reached either by doubling the transcription rate (dashed line) or the mRNA stability (dotted line). Furthermore, in this example the transient time to steady state depends on the average lifetime alone. Conversely, in the right panel, mRNAs with different forms of the lifetime distribution, $\phi_U^{(\Gamma)}$ (blue) and $\phi_U^{(\mathrm{exp})}$ (red), but same $\omega_{\mathrm{tc}} = 1\,\mathrm{min}^{-1}$ and average lifetime $\langle U \rangle = 4$ min are shown. Hence, the length of the transient depends critically on the form of the lifetime distribution. The inset emphasizes the symmetry between the decay of the mRNA amount after the stop of transcription (dashed) and the evolution of the mRNA number after start of transcription (solid). At each time interval after start and stop of transcription, the average mRNA number of the two processes add up to $\omega_{\mathrm{tc}} \langle U \rangle$, the decrease in the one process is exactly balanced by the increase in the other as is shown theoretically in Eq. (4.5).

min. The mRNAs with a gamma lifetime distribution have a shorter transient to a steady state level. This is due to the fact that the gamma distribution (with $n = 5$) has a smaller variance than the exponential distribution. Generally, the transient time is large if the lifetime density is very broad, *i.e.* has a high variance.
In conclusion, the length of the transient depends on both the form of the lifetime distribution Φ_U and the average lifetime $\langle U \rangle$. This fact has several implications that will be investigated in the next sections.

4.4. Transient evolution of age and residual lifetime

Due to the turnover of mRNA, there is an age distribution which reflects the age composition of the mRNA pool. Also the age distribution of the mRNAs expressed after the induction evolves in time. We consider again a gene that was not transcribed before the induction.
In chapter 2 we derived a general result for the age probability density ϕ_A. Here we are interested in the limit case where transcription was not stopped, *i.e.* the limit case of Eq. (2.17) where $\Delta t = 0$ and $t = t_s$. The age probability density function at time t after the induction of transcription is given by

$$\phi_A(a \mid t) \;=\; \left[\int_0^t \mathrm{d}\tau \,(1 - \Phi_U(\tau))\right]^{-1} (1 - \Phi_U(a)) \;, \tag{4.8}$$

for $a < t$ and zero otherwise. In the limit $t \to \infty$ also the age distribution becomes stationary and its expression is given by

$$\phi_A^{\mathrm{st}}(a) \;=\; \frac{1 - \Phi_U(a)}{\langle U \rangle} \,, \tag{4.9}$$

which is the stationary distribution of the age of a renewal process [73]. In figure 4.4 A we compare the stationary age distribution given by a gamma lifetime density $\phi_U^{(\Gamma)}$ defined in Eq. (4.7) to the exponential case $\phi_U^{(\mathrm{exp})}$ defined in Eq. (4.6). Clearly, in the former more young mRNAs are present, whereas in the exponential case there is a higher proportion of older mRNAs. This follows directly from the fact that the exponential distribution has a higher variance as was shown in figure 4.1. The inset shows the age distributions five minutes after the start of transcription. Hence, no mRNAs are older than 5 min - however the shape of the distribution at smaller ages $a < t = 5$ min is identical to the stationary case except for a rescaling due to the renormalization.
The average age of the mRNAs at time t after the induction of transcription is given by

$$\langle A \rangle_t \;=\; \int_0^t \mathrm{d}a \, a \, \phi_A(a \mid t) \,, \tag{4.10}$$

and its time evolution can be followed in figure 4.4 B. While shortly after the induction both mRNAs have a similar average age, the effect of the different lifetime densities becomes more pronounced at larger times. The average age at steady state is lower for gamma-like mRNAs which follows from the different shape of the age distribution as shown in figure 4.4 A.
Similarly, one can derive a result for the residual lifetime of a given mRNA species at any time after the induction. For $\Delta t = 0$ it follows immediately from Eq. (2.18) that

$$\phi_R(r \mid t) \;=\; \left[\int_0^t \mathrm{d}u \,(1 - \Phi_U(u))\right]^{-1} (\Phi_U(t + r) - \Phi_U(r)) \;. \tag{4.11}$$

Figure 4.5 A shows the residual lifetime probability density of mRNAs with a gamma lifetime distribution $\phi_U^{(\Gamma)}$ according to Eq. (4.7) at different time points after the start of

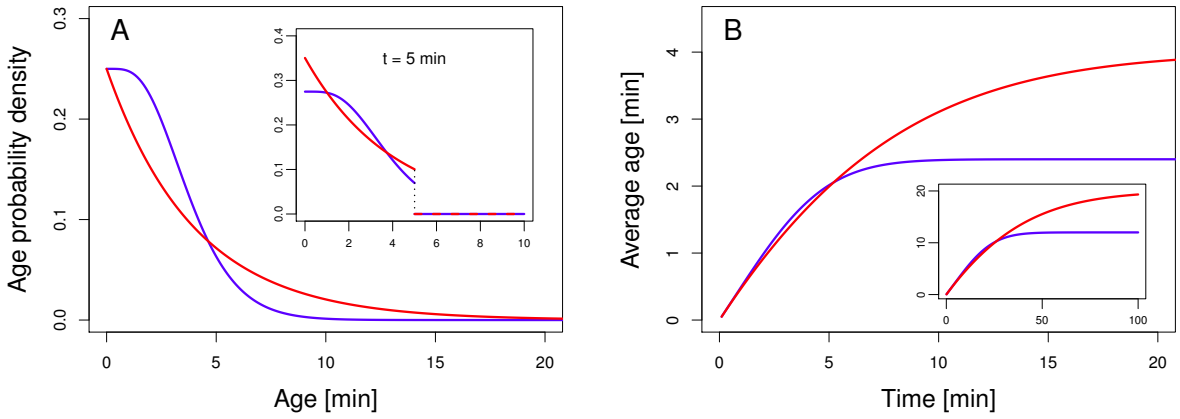

Figure 4.4.: **Time evolution of the age distribution**. The left panel shows the mRNA age probability density $\phi_A(a \mid t)$ under steady state conditions for the two exemplary lifetime distributions, $\phi_U^{(\Gamma)}$ (blue) and $\phi_U^{(\exp)}$ (red) - see figure 4.2 A and Eqs. (4.6) and (4.7). Despite the fact that the stationary mRNA number distributions are identical (see figure 4.2 B), the stationary age distributions are different. The inset shows the mRNA age distribution at a fixed time $t = 5$ min after the induction of transcription. Clearly, no mRNA can be older than 5 minutes - for smaller ages the age distributions remain the same as for the stationary case apart from a rescaling. In the right panel, the evolution of the average age as given in Eq. (4.10) is plotted versus the time after the start of transcription for $\langle U \rangle = 4$ min. During the transient to stationarity the two average ages are very similar to each other. On the other hand, at stationarity the average age under the lifetime density (4.7) is clearly much smaller than the average age under the exponential distribution (4.6). The inset shows the evolution for mRNAs with mean lifetime $\langle U \rangle = 20$ min.

transcription. Whereas shortly after the start of transcription the distribution is peaked around the average lifetime $\langle U \rangle = 4$ min (solid blue), the distribution becomes more skewed to shorter residual lifetimes at later time points (the dashed-dotted line represents $t = 5$ min). In fact, for small values of t, $\phi_R(r \mid t) \to \phi_U^{(\Gamma)}$, *i.e.* the residual lifetime distribution has the same form as the lifetime distribution. When a steady state of the mRNA turnover is reached, the residual lifetime distribution is also stationary (black line). Moreover, for $t \to \infty$ - namely, at steady state - the residual lifetime distribution $\phi_R^{\mathrm{st}}(r) = \phi_A^{\mathrm{st}}(a)$ for all lifetime distributions Φ_U. Conversely, for mRNAs with exponential lifetime distribution according to Eq. (4.6) the residual lifetime distribution is the same at all points after the start of transcription and takes the form of Eq. (4.6).
The average residual lifetime of the mRNAs at time t after the induction of transcription is given by

$$\langle R \rangle_t = \int_0^t \mathrm{d}r \, r \, \phi_R(r \mid t) \tag{4.12}$$

and its time evolution can be followed in figure 4.5 B. Also the average residual lifetime evolves in time and attains a steady state distribution $\phi_R^{\mathrm{st}}(r)$ for $t \to \infty$. As the average age of gamma regulated mRNAs increases (as we noted figure 4.4 B), the residual lifetime decreases with time. Note, however, that in the case of an exponential lifetime distribution $\phi_U^{(\exp)}$, the mRNAs stay forever young, *i.e.* they have the same life expectancy at all times. This is due to the Markov property of the exponential distribution as we discussed above

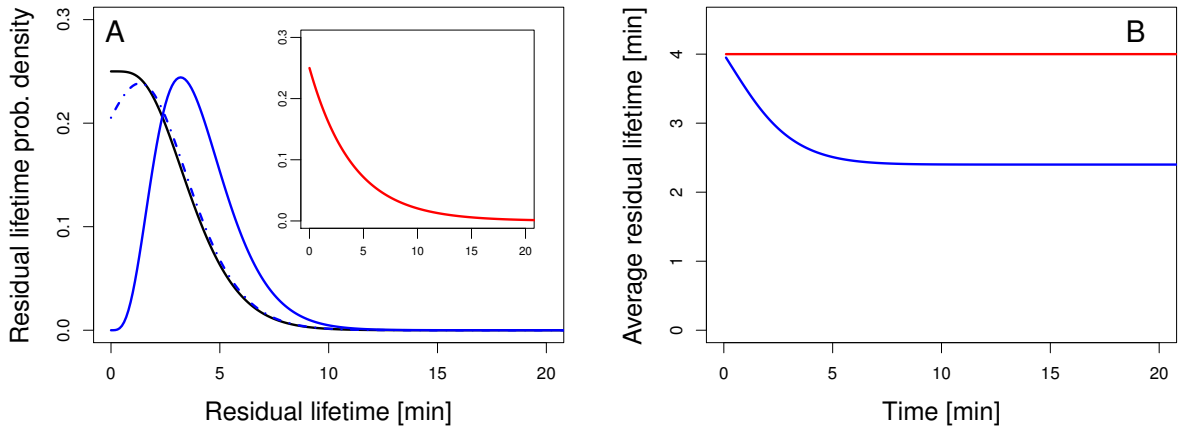

Figure 4.5.: **Time evolution of the residual lifetime distribution**. The left panel shows the residual lifetime probability density $\phi_R(r \mid t)$ of mRNAs with a gamma lifetime distribution $\phi_U^{(\Gamma)}$ according to Eq. (4.7) at different time points after the start of transcription. Whereas shortly after the start of transcription the distribution is peaked around the average lifetime $\langle U \rangle = 4$ min (solid blue), the distribution becomes more skewed to shorter residual lifetimes at later time points (the dashed-dotted line represents $t = 5$ min). When a steady state of the mRNA turnover is reached, the residual lifetime distribution is also stationary (black line) and has the same form as the stationary age distribution as shown in figure 4.4 A. In the inset, the analysis is repeated for mRNAs with exponential lifetime distribution $\phi_U^{(\exp)}$ according to Eq. (4.6). Here, the residual lifetime distribution is the same at all points after the start of transcription and is described by an exponential distribution. The right panel displays the temporal evolution of the average residual lifetime. Clearly, mRNAs with exponential lifetime distributions show no indication of aging, their average residual lifetime is constant no matter how old the mRNAs are. Conversely, for the mRNAs with a gamma lifetime distribution the average residual lifetime decreases with the time evolved since the start of transcription until finally reaching a stationary value. This becomes clear since also the average age increases over time (see figure 4.4 B) - hence, older chains have a smaller life expectancy.

in chapter 3.
Hence, both the age and the residual lifetime distributions as well as their average values are important criteria to describe the molecular aging of mRNAs. However, they are not directly measurable objects. Nevertheless, they indirectly control the processes of translation and protein synthesis. We will discuss these implications in the next chapter. Furthermore, we learned that mRNAs with exponentially distributed lifetimes show very peculiar properties as they show no indication of aging, *i.e.* they have a constant residual lifetime independent of the their age.

4.5. Aging of states

We can also investigate the aging of the mRNA molecules in the framework of the Markov chain model introduced above. In this description, different states reflect different biochemical states of the mRNA molecule. Depending on its age, each randomly chosen mRNA can be found in one of its n different states with a certain probability. If K is the random state of a randomly chosen mRNA, we have to compute the probability to be in each of the states in a Markov chain conditioned that the mRNA has not yet been degraded, *i.e.*

$$\xi_k(a) \equiv \Pr\{K(a) = k \mid K(0) = 1,\, K(a) \neq 0\} = \frac{\Pr\{K(a) = k,\, K(a) \neq 0 \mid K(0) = 1\}}{\Pr\{K(a) \neq 0 \mid K(0) = 1\}}\,. \tag{4.13}$$

Clearly, the probability depends on the age of each mRNA and initially all mRNAs are in state 1. One can compute the probabilities in the nominator of Eq. (4.13) via a Master equation. Moreover, the denominator is just the complementary distribution to the probability lifetime distribution, *i.e.* the *survival probability* $1 - \Phi_U(a)$. The probabilities ξ_k for different ages a are depicted in figure 4.6 A.
Since mRNAs of different age are present at each time of the experiment, we have to multiply the probabilities $\xi_k(a)$ with the age distribution $\phi_A(a \mid t)$ and integrate over all

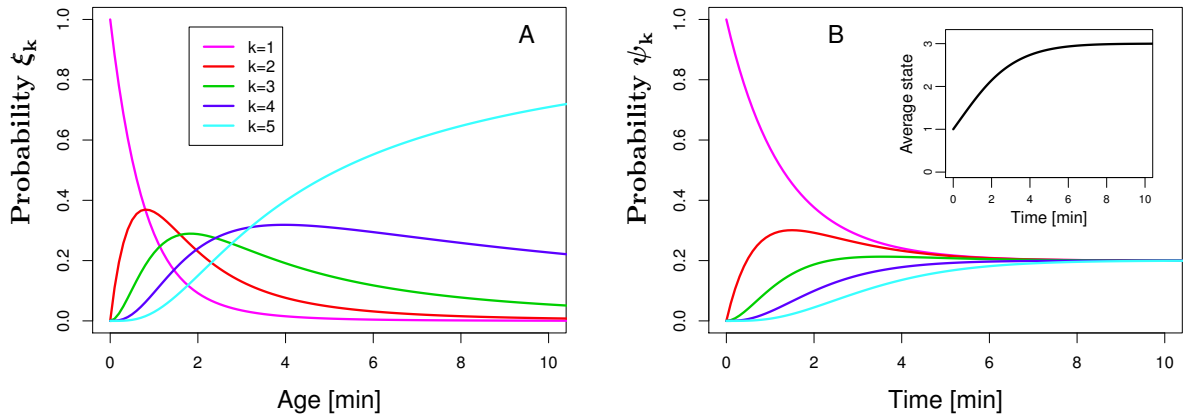

Figure 4.6.: **Manifestation of aging in Markov chain model.** If we interpret the lifetime distribution $\phi_U^{(\Gamma)}$ in the framework of the Markov chain model, different states reflect a different biochemical state of the mRNA molecule. As a molecule becomes older, it is more probable to be found in an advanced state. In the left panel, we follow the probability $\xi_k(a)$, defined in Eq. (4.13), for an mRNA with lifetime distribution $\phi_U^{(\Gamma)}$ and transition rates $\lambda = 5/4\,\mathrm{min}^{-1}$ to be in each of the states $k = 1, 2, \ldots 5$ (see figure legend). Clearly, at large ages the most advanced state $k = 5$ is the most populated. In the right panel, the probability $\psi(t)$, given in Eq. (4.14), of an ensemble of mRNAs with age distribution $\phi_A(a \mid t)$ for different times t after the induction of transcription is shown. After a transient time, the probabilities attain all the same stationary value, $\psi_k^{\mathrm{st}} = 1/5$ for $\forall\, k$. Note that this is a particular result for Γ-distributed lifetimes where all intermediate degradation rates $\mu_i \equiv 0$ $(i < 5)$ and all transition rates λ_i as well as the degradation rate μ_5 have the same value. From the population probability of the individual states, we can infer the average state of a randomly chosen mRNA at any time t via Eq. (4.15) (see inset). Clearly, the aging is indicated by an increase of the average state.

possible ages a. Thus, the probability of an ensemble of mRNAs to be in state k in a Markov chain reads

$$\psi_k(t) \equiv \int_0^\infty da\, \xi_k(a)\, \phi_A(a \mid t)\,. \tag{4.14}$$

Figure 4.6 B shows how the occupation probability $\psi_k(t)$ for each state k in the Markov chain evolves with time after the start of transcription. Since the average age increases with time, more mRNAs are found at later states in the Markov chain at larger t. Moreover, when the age distribution approaches a stationary distribution, also the occupation probability becomes stationary. The aging can also be observed in the average state of an mRNA at time t,

$$\langle\, k(t)\,\rangle = \sum_{k=1}^{n} k\, \psi_k(t) \tag{4.15}$$

Figure 4.6 B shows that the average state increases over time and reaches a stationary value after some time. This is a clear manifestation of aging.

4.6. Modulation of the transcription rate

In the preceding analysis of the evolution of the number of mRNAs after the induction of transcription, we always assumed that the initial condition at $t = 0$ was $N_r(0) = 0$. This was a reasonably choice of an initial condition when we reconsider the theoretical framework developed in chapter 2. We treated the creation of mRNAs as a Poisson process that starts at a time t and in this formalism there is no knowledge contained about preexisting mRNAs. In this section we will extent our formalism to describe the change in the mRNA amount with arbitrary amount of preexisting mRNAs.
Furthermore, we have assumed so far that the transcription rate ω_{tc} stays constant throughout an experiment. In this section, we will study the effect when the transcription rate changes at a distinct time point $t = 0$ from $\omega_{\mathrm{tc}}^{(1)}$ to $\omega_{\mathrm{tc}}^{(2)}$. We will see that this modulation of the transcription rate is related to the consideration of preexisting mRNAs.
We consider a first Poisson process with rate $\omega_{\mathrm{tc}}^{(1)}$ which starts at a time $t = t_1 < 0$. The number of mRNAs evolves in time as discussed in the previous sections and is visualized in figure 4.7 A (red curve). At time $t = 0$ this first Poisson process is interrupted and the mRNAs subsequently decay according to their lifetime distribution (red dotted curve). At the same time a second Poisson process is started with rate $\omega_{\mathrm{tc}}^{(2)}$ and the number of mRNAs evolves again in time (green curve). The two Poisson processes describe the evolution of mRNAs of the same gene with transcription rates but the same degradation pathway. Hence, the evolution of the total number of mRNAs is given by the sum of the two sub-processes (black curve in figure 4.7 A) what we can formally describe as

$$N_r^{\mathrm{tot}}(t) = N_r^{(1)}(t) + N_r^{(2)}(t)\,. \tag{4.16}$$

More precisely,

$$N_r^{(1)}(t) = \begin{cases} \omega_{\mathrm{tc}}^{(1)} \displaystyle\int_0^{t_1} d\tau\,(1 - \Phi_U(\tau)) & t < 0 \\ \omega_{\mathrm{tc}}^{(1)} \displaystyle\int_t^{t_1+t} d\tau\,(1 - \Phi_U(\tau)) & t \geq 0 \end{cases} \tag{4.17}$$

and

$$N_r^{(2)}(t) = \begin{cases} 0 & t < 0 \\ \omega_{\mathrm{tc}}^{(2)} \displaystyle\int_0^{t} d\tau\,(1 - \Phi_U(\tau)) & t \geq 0 \end{cases} \tag{4.18}$$

For illustrative purposes, we have depicted the evolution of the number of mRNAs in figure 4.7 A. There, we have chosen $t_1 = -10$ min, $\omega_{\mathrm{tc}}^{(1)} = 1\,\mathrm{min}^{-1}$, $\omega_{\mathrm{tc}}^{(2)} = 2\,\mathrm{min}^{-1}$ and $\phi_U(t) = \phi_U^{(\Gamma)}(t)$ with $n = 5$ and $\langle U \rangle = 4$ min. The mRNA number evolves to a first steady state $N_r^{(1),\mathrm{st}} = \omega_{\mathrm{tc}}^{(1)} \langle U \rangle$ until $t = 0$ and a second steady state $N_r^{(2),\mathrm{st}} = \omega_{\mathrm{tc}}^{(2)} \langle U \rangle$. Note that in both cases the transient to a steady state is solely governed by ϕ_U. Moreover, the arbitrary choice of t_1 is made without loss of generality.
In the following, we will assume that the first Poisson process started sufficiently long before $t = 0$ to have reached a steady state and focus on the evolution for $t > 0$. Hence, using the notation introduced in Eq. (2.16), Eq. (4.16) simplifies to

$$N_r^{\mathrm{tot}}(t) = \omega_{\mathrm{tc}}^{(1)} \mathcal{H}_t^{\infty} + \omega_{\mathrm{tc}}^{(2)} \mathcal{H}_0^t\,. \tag{4.19}$$

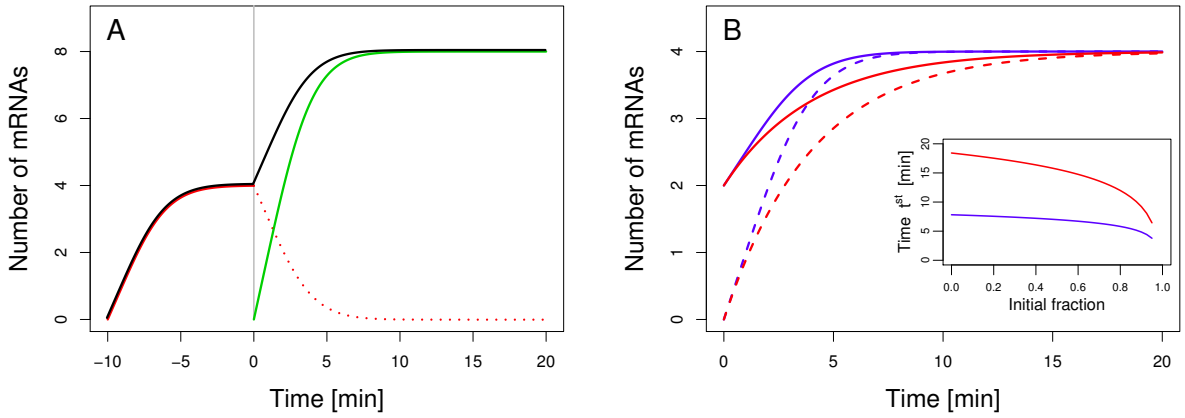

Figure 4.7.: **Modulation of the transcription rate and transient to steady state from different initial conditions**. Left panel: A first process with rate $\omega_{\mathrm{tc}}^{(1)}$ is started at time $t = t_1 < 0$ and the number of mRNAs evolves according to Eq. (4.1) until $t = 0$ (red line). There, the first process is stopped and the remaining mRNAs decay according to Eq. (3.1) (red dotted line). Furthermore, a second process is started with rate $\omega_{\mathrm{tc}}^{(2)}$ which subsequently also leads to the evolution of the mRNA number in time (green line). The two Poisson processes represent the evolution of the same type of mRNAs with different transcription rates. The evolution of the total amount of mRNA is given by the sum of the two sub-processes (black line) and can be described via Eq. (4.16). We use this formalism to describe the evolution of the amount of mRNA for different initial conditions at $t = 0$ (right panel). We compare the time evolution to a steady state mRNA number for two mRNA species regulated according to $\phi_U^{(\mathrm{exp})}(t)$ (red) and $\phi_U^{(\Gamma)}(t)$ (blue). Moreover, we compare both cases to the corresponding evolution of mRNAs with no preexisting mRNAs at $t = 0$ (dashed lines). In the inset, we estimate the time to steady state for a large variety of initial values.

In this case, the initial number of mRNAs $N_r^{\mathrm{tot}}(0)$ can be non-zero - it is determined by $\omega_{\mathrm{tc}}^{(1)}\langle U\rangle$. Thus, Eq. (4.19) extends the formalism and relaxes the restriction on the initial condition $N_r(0)$.

Figure 4.7 B shows the time evolution of the average mRNA number when the initial amount is already half the steady state value. We compare the evolution for mRNAs regulated according to $\phi_U^{(\mathrm{exp})}(t)$ (red) and $\phi_U^{(\Gamma)}(t)$ (blue). Moreover, we compare both cases to the corresponding evolution of mRNAs with no preexisting mRNAs at $t = 0$ (dashed lines). Similar to the discussion in section 4.3, the transient to a steady state depends critically on the lifetime distribution. Additionally, the time to steady state does not differ much if there is an initial amount of mRNAs at $t = 0$.

To investigate the duration of the transient further, we have performed the same analysis for a wide range of initial values. Therefore, we define implicitly the approximate time to steady state t^{st} as $N_r(t^{\mathrm{st}})/N_r^{\mathrm{st}} \equiv 0.99$. In figure 4.7 B we follow the dependency of t^{st} on the fraction of the initial amount compared to the steady state amount. In particular for mRNAs regulated by $\phi_U^{(\Gamma)}$, the transient time depends only weakly on the initial value (blue). Also for mRNAs regulated by $\phi_U^{(\mathrm{exp})}$, an initial fraction of 0.8 reduces the transient time by less than a factor of two (red). Only when the initial fractions approaches 1, the transient time decreases substantially.

In this section we have discussed average mRNA numbers. In the same way we can also

describe the fluctuations of the mRNA amount. The number of mRNAs pertaining to each of the two transcription phases are described by a Poisson distribution with parameter $N_r^{(1)}(t)$ and $N_r^{(2)}(t)$, respectively. A fundamental property of the Poisson distribution states that the sum of two Poisson-distributed random variables also follows a Poisson distribution [30]. Its parameter is given by the sum of the parameters of the two primary Poisson distributions. Thus, in the case of modulated transcription the random number of mRNAs also follows a Poisson distribution with parameter N_r^{tot}.

In conclusion, we have found a formalism to describe the evolution of mRNAs for arbitrary initial conditions. This initial number of mRNAs can represent the preexisting amount of mRNAs before the induction of a gene. In the subsequent section we will apply this formalism to see how the transcription rate is altered when yeast cells are put under stress. Moreover, we will further exploit this formalism in section 5.3 where we turn to the implications of the transient for the cell cycle.

In this section, we have learned how the mRNA number evolves when the transcription rate is modulated at a fixed point in time. Certainly, the description can be extended to include more modulation events. In the appendix, section C.1, we will elaborate a more detailed theory where the rate of transcription changes gradually from one value to another. Furthermore, in single cells the transcription rate may vary in complicated stochastic patterns. This requires a description of transcription via the Cox process. In the appendix D we will approach this problem by means of Monte-Carlo simulation.

4.7. Response to stress application in yeast

To show how one can use the theoretical framework describing the evolution of mRNAs presented in the previous sections, we will apply it to explain experimental data on stress response in the yeast *S. cerevisiae* [27]. In the experiment, yeast cells were exposed to either hydrogen peroxide (H_2O_2) or methyl methane-sulfonate (*MMS*) which induces DNA damage. For each gene, the time evolution of the relative number of mRNAs after stress application was followed via a microarray measurement. Additionally, the decay of the amount of mRNA was measured before and 25 min (40 min for *MMS*) after the stress application, *i.e.* at a time when one can assume that the stress became fully manifest in the decay pattern.
The stress induction may result in a change of transcription and mRNA degradation for each gene and in the following we will apply the formalism of mRNA turnover developed so far in this thesis to further understand the contribution of each of the two processes.
First, we assume that before stress application, the yeast cells were grown sufficiently long to attain a steady state mRNA number at $t = 0$. Hence, a first Poisson process with rate $\omega_{\rm tc}^{(1)}$ leads to a steady state mRNA number $N_r^{(1),{\rm st}}$. Following the reasoning of the previous section, the first Poisson process is stopped at the time of stress application, $t = 0$, and the remaining mRNAs decay according to the stressed decay pattern $\Lambda^{(2)}(t)$. At the same time, a second Poisson process with rate $\omega_{\rm tc}^{(2)}$ is started. The experiment measures the response of the mRNA expression for 180 min and we will consider the case that within that period of time also for the second process a steady state $N_r^{(2),{\rm st}}$ is reached.
In the framework of our general theory, the steady state mRNA amount is related to the transcription rate and average lifetime via $N^{\rm st} = \omega_{\rm tc}\langle U\rangle$, see Eq. (4.4) in section 4.2. Thus, the ratio of the transcription rates before and after stress application reads

$$\frac{\omega_{\rm tc}^{(2)}}{\omega_{\rm tc}^{(1)}} = \frac{\langle U^{(2)}\rangle}{\langle U^{(1)}\rangle}\frac{N^{(2),{\rm st}}}{N^{(1),{\rm st}}} \tag{4.20}$$

Since the experimental expression levels at $t = 0$ have been normalized, *i.e.* $N^{(1),{\rm st}} \equiv 1$, we can only infer the relative change of the transcription rate. It is given by the average lifetimes obtained from the decay experiments and from the (relative) steady state level of mRNAs, $N^{(2),{\rm st}}$, after the stress application.
The total number of mRNAs at each time t is given by the contributions of the mRNAs originating before and after the application of the stress. The mRNAs originating before the stress application undergo a change of their lifetime distribution at $t = 0$. To account for this, we have to extend the general theoretical description developed in chapter 2. The fundamental assertion is that for $t < 0$ the mRNAs have a random lifetime U_1 given by $\Phi_U^{(1)}$ and for $t \geq 0$ a random lifetime U_2 given by $\Phi_U^{(2)}$. In the appendix section C.3 we give detailed account for the necessary changes in our theoretical description. As the final result, Eq. (C.14) describes the evolution of the total mRNA amount for cells that were exposed to a stress at $t = 0$,

$$\begin{aligned} N_r^{\rm tot}(t) &= N_r^{(1)}(t) + N_r^{(2)}(t) \\ &= \omega_{\rm tc}^{(1)} \int_0^\infty {\rm d}u\, \frac{1-\Phi_U^{(1)}(u)}{1-\Phi_U^{(2)}(u)} \left(1-\Phi_U^{(2)}(u+t)\right) + \omega_{\rm tc}^{(2)} \int_0^t {\rm d}u \left(1-\Phi_U^{(2)}(u)\right) . \end{aligned} \tag{4.21}$$

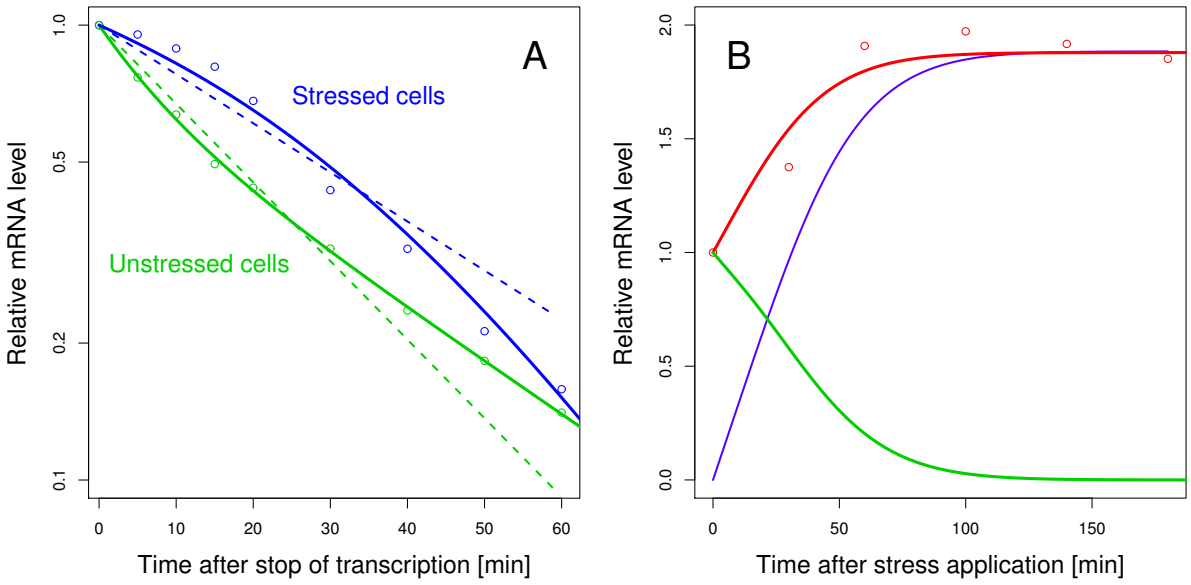

Figure 4.8.: **Decay patterns and stress response of *HAT2* after stress activation via exposure to DNA damaging *MMS***. In the left panel, the experimental decay pattern before (green circles) and after (blue circles) the stress application are shown [27]. We fit both patterns according to the Markov chain model developed in section 3.2 (solid lines). We find that the average lifetime changes from $\langle U^{(1)} \rangle = 17.8$ min to $\langle U^{(2)} \rangle = 56.8$ min and obtain the corresponding lifetime densities $\phi_U^{(1)}$ and $\phi_U^{(2)}$. For comparison a fit with a simple exponential function is shown (dashed lines). For the relative response of the mRNA number (right panel) we consider two subsequent Poisson processes. Process one is thought to start sufficiently long before the stress application with rate $\omega_{\mathrm{tc}}^{(1)}$ such that at the time point of stress application at $t = 0$ it has reached a steady state value $N_r^{(1),\mathrm{st}}$. At $t = 0$ the first process is stopped and the remaining mRNAs decay according to the stressed degradation pathway $\Lambda^{(2)}(t)$ (green line). Furthermore, a second Poisson process starts with rate $\omega_{\mathrm{tc}}^{(2)}$ and the mRNA number evolves to a new steady state $N_r^{(2),\mathrm{st}}$ (blue line). According to Eq. (4.20) we can obtain the previously unknown transcription rates from $N_r^{(1),\mathrm{st}}$, $\langle U^{(1)} \rangle$, $N_r^{(2),\mathrm{st}}$ and $\langle U^{(2)} \rangle$. Since the experimental data were normalized to $N_r^{(1),\mathrm{st}} \equiv 1$, we can only infer the relative change of the transcription rate, $\omega_{\mathrm{tc}}^{(2)}/\omega_{\mathrm{tc}}^{(1)} = 0.498$. Thus, for *HAT2*, the stress results in an increase in the stability of the mRNAs which is partially balanced by a decrease in the transcription rate. The combined changes in transcription and degradation lead to a higher mRNA expression level (red circles) which can be theoretically predicted by the sum of the two Poisson processes given in Eq. (4.21) (red line).

Note that in the present experiment the expression data were normalized to $N^{(1),\mathrm{st}} = 1$. Therefore, the transcription rates $\omega_{\mathrm{tc}}^{(1)}$ and $\omega_{\mathrm{tc}}^{(2)}$ in Eq. (4.21) are relative quantities in respect to the normalization condition.

To probe our theory, we conduct a case study for the response of the expression of gene *HAT2* (Histone Acetyl Transferase 2) to the exposure to *MMS*. Figure 4.8 A shows the measured patterns of decay before and after the stress application as well as the change of the mRNA expression level.

We can fit both decay patterns according to the Markov chain model developed in section 3.2 (solid lines). We find that the average lifetime increases from $\langle U^{(1)} \rangle = 17.8$

min to $\langle U^{(2)} \rangle = 56.8$ min and obtain the corresponding lifetime densities $\phi_U^{(1)}$ and $\phi_U^{(2)}$. Furthermore, according to Eq. (4.20), we can infer the ratio of the transcription rates $\omega_{\mathrm{tc}}^{(2)}/\omega_{\mathrm{tc}}^{(1)} = 0.498$. Thus, we find that for *HAT2* the DNA damage reduces the transcription rate almost by a factor of 2.
Hence, all parameters of Eq. (4.21) are derived from experimental data. Figure 4.8 B demonstrates that our theory approximates the measured evolution of the mRNA amount after stress application very well. The increase of the average mRNA lifetime and the decrease of the transcription rate due to DNA damage counteract but lead to an overall increase of the mRNA expression level for the gene *HAT2*. Note that for the deduction of the theoretical stress response curve no fitting needed to be done. With our theory it is well defined by $\phi_U^{(1)}$, $\phi_U^{(2)}$ and $N_r^{(2),\mathrm{st}}$.
We developed a formalism to describe the response of the mRNA expression due to an external stress. It allows discriminating between altered transcription and degradation in the experiment. The case study for *HAT2* demonstrates the principle scope of our theory. Nevertheless, a number of assumptions were made in the formulation of Eq. (4.21). First, we assumed that the stress becomes effective instantly after the stress application. Therefore, we have considered the possibility of a delayed change of the mRNA transcription and degradation in a more detailed model (see section C.2 in the appendix). At least for the present case of the gene *HAT2*, the more elaborate model does not lead to a more accurate description of the experimental response pattern (see appendix section C.3). However, this is not necessarily true for the response of all genes.
Additionally, it is desirable to find a description which considers gradual changes of the transcription rate. In the appendix section C.1, we have reformulated the theoretical description to allow for the definition of the change of the transcription rate with an arbitrary function such as for instance a linear ramp.
Secondly, for mRNAs that were synthesized before and decay after the stress application, the lifetime distribution changes during the experiment. It is a good example for the interference of global time scales (the time of stress application) and molecular time scales (such as the age of an mRNA at stress application). In the light of different possible biological interpretations, diverse theoretical approaches can be followed to account for this change of the lifetime distribution. In the appendix, section C.3, we elaborate different possibilities of a time-dependent lifetime distribution.

4.8. Chapter summary

In this chapter we have elucidated the role of mRNA degradation in the context of transient phenomena in mRNA expression. The average lifetime determines the steady state level of the number of mRNAs. However, the form of the mRNA lifetime distribution determines the time to a steady state mRNA level, even when the average lifetime is held constant (see figure 4.3). Hence, it sets an important time scale for the response of the cell to external stimuli. Furthermore, we generalized our theory to arbitrary mRNA initial conditions. This was particularly useful to interpret experiments in which the response of the mRNA level to an external stress was measured (see figure 4.8). Thus, we could differentiate between contributions of altered modes of transcription and degradation un-

der the stressed conditions. Additionally, we have introduced the effect of aging of the mRNAs during this transient which we will utilize in the next chapter where we discuss its implications on translation and protein synthesis. The entire material of this chapter is original research work. Parts of the results in sections 4.2 4.3 and 4.4 are already published [74].

5. Effect of mRNA decay on translation, protein synthesis and the cell cycle

This chapter aims at setting the concepts learned so far into the broader context of gene expression. In chapter 3 we have learned that typically mRNA decay cannot be described with a single decay constant. However, with a more accurate description of the biochemical degradation mechanisms we could find suitable lifetime distributions. Moreover, in chapter 4 we found that this leads to more intricate transient phenomena. On the one hand, the growth of the mRNA amount after induction depends on the lifetime distribution Φ_U. On the other hand, non-exponential decay implies aging of the mRNA molecules which becomes manifest in the residual lifetime distribution. Since the efficiency of translation and protein synthesis depends strongly on the stability of mRNAs, these findings will have important implications on these processes. In sections 5.1 and 5.2 we will assess the role of these findings. In the final two sections we will discuss some implications when we consider single cells. Therefore, in section 5.3 we envision how mRNAs evolve in the cell cycle under continuous division and dilution. Besides, all sections of this chapter are subject to very intense research itself. Therefore, we focus on aspects that are directly linked to mRNA stability.

5.1. Implications for translation

5.1.1. Interaction of time scales

In section 3.3 we have already discussed the interplay of the timescales of the lifetime of mRNAs and translation and its implications for the stability of mRNAs in bacteria. Here we focus on the effect on translation.

On average the time needed for a ribosome to transverse the coding sequence of an mRNA is given by $t_L = L/v$, *i.e.* the length of the coding sequence L divided by the elongation speed v. This *elongation* time is however a stochastic quantity which can be considered explicitly in computer simulations. This time determines one of the time scales of translation. As we pointed out in chapter 4, randomly sampled mRNAs can be described by an age distribution $\phi_A(a)$. The time scale of translation t_L interacts with the time scale of mRNA turnover if they are of the same order of magnitude. More precisely, mRNAs younger than t_L are not in a steady state ribosome coverage - rather the coverage changes in time, giving rise to dynamic effects. Thus, the age distribution $\phi_A(a)$ of an mRNA determines the strength of these dynamic effects. Conversely, if the time scale t_L is much shorter than the typical age, most mRNAs are in a steady state

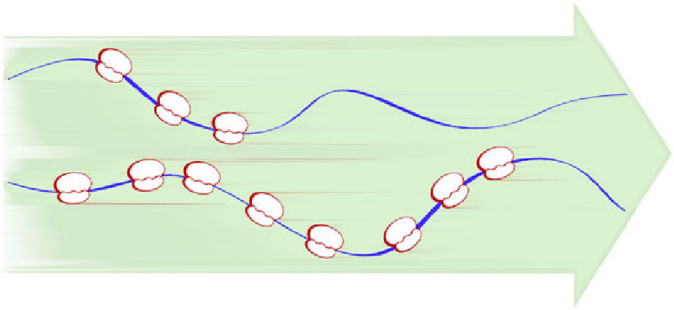

Figure 5.1.: **Illustration of two polysomes of different age**. The upper mRNA transcript is younger, hence less ribosomes have bound to the mRNA. In any experiment one would measure a mixed sample of old and young mRNAs of different polysome sizes, respectively. From a theoretical point of view, one has to sample from the age distribution ϕ_A to account for such a heterogeneous composition. [Adapted from [50]]

ribosome coverage and no transient phenomena occur. If the lifetime of an mRNA were much shorter than t_L, translation would be too slow to yield proteins and is hence rather unlikely to occur in healthy cells.

Note that the precise definition of the time scales may be different for different organisms. However, regardless of the organism it will always take a certain time for a ribosome to translate the entire mRNA.

Figure 5.1 illustrates the role of the age distribution. In every sample obtained from experiment one would collect mRNAs of different age. However, according to the age of each mRNA, the transcript is occupied by a different number of ribosomes. Moreover, we learned in chapter 4 that the age distribution depends critically on the time after the induction of transcription. Therefore, to fully understand the process of translation, the underlying age distribution has to be taken into account.

5.1.2. Experimental methods

From centrifugation experiments one can deduce the distribution of the number of ribosomes on a given mRNA. The experiments were pioneered by Arava *et al.* [55]. After lysing the cells, one can fix the ribosomes on the mRNAs by adding the drug *cycloheximide*. Next, one can separate mRNAs of different ribosomal coverage by velocity sedimentation through a sucrose gradient. Hence, one can retrieve a *spectrum*, *i.e.* a mass vs. intensity graph, where the individual peaks correspond to different ribosome numbers. Furthermore, individual mRNA species can be identified via standard methods such as *northern blotting* or *qRT-PCR*.

This method has been extended to perform *ribosomal profiling* experiments [75, 76]. Here, after the fixation of the ribosomes, the mRNA stretches not covered by ribosomes are removed by adding a ribonuclease. After further purification, the remaining fragments that were previously covered by ribosomes can be analyzed by *RNA-Seq*. Thus one can identify the position of the ribosomes on the mRNA for each gene, *i.e.* retrieve the *footprints* of the ribosomes on the mRNAs.

5.1.3. Theory of polysome and ribosomal profiles

We have developed a theory that is capable of modeling the two kinds of experiments. Our theory simplifies the biology since it provides only a sequence-independent mean field description. To incorporate effects of individual sequences, one would have to consider more details of the translation process such as codon specific waiting times. This is part of active research and goes beyond the scope of this analysis. Nevertheless, our theory is able to predict general trends in profiling experiments that are due to a dynamic age composition of the mRNAs in the cell. Parts of the theory were already derived in [50]. We extend this description by considering general mRNA lifetimes and non-stationary age distributions as introduced in chapter 4.

Polysome size distribution In the following, we aim to compute the probability q_n that n ribosomes are bound to an mRNA of a given species, where n can assume any integer value up to the maximal packing. That means if n ribosomes are engaged in translating an mRNA, the polysome size is n. In a first approximation, for a length of the coding sequence L and an average elongation speed v, it takes a time t_L for a ribosome to transverse the mRNA. Hence, on average only mRNAs of an age larger than t_L have full ribosome occupancy. The equation that describes the evolution of the polysome size with initiation rate ω_{on} is

$$\frac{d\,q_n}{d\,a} = \omega_{\text{on}}\, q_{n-1} - \omega_{\text{on}}\, q_n \quad \text{for} \quad a < t_L \tag{5.1}$$

and $q_n(a) = q_n(t_L)$ for $a \geq t_L$. The initial condition is that no ribosomes are bound to the mRNA, thus $q_0(0) = 1$ and $q_i(0) = 0$ for $i > 0$. Eq. (5.1) describes a Poisson process which approximates well the polysome dynamics if ω_{on} is sufficiently small (in that case the ribosome density stays sufficiently low so that interaction between adjacent ribosomes can be neglected) [50]:

$$q_n(a) = \begin{cases} \frac{(\omega_{\text{on}}\, a)^n}{n!} \exp\left(-\omega_{\text{on}}\, a\right) & a < t_L \\ \frac{(\omega_{\text{on}}\, t_L)^n}{n!} \exp\left(-\omega_{\text{on}}\, t_L\right) & a \geq t_L \end{cases} \tag{5.2}$$

Hence, as the age of an mRNA evolves, more ribosomes bind to the chain until reaching a steady state in this average description. To obtain the polysome size of a population of mRNAs we have to take into account the corresponding age distribution of the mRNAs in the cell population. Therefore, we have to multiply Eq. (5.2) with the age distribution and integrate over all possible ages. However, in chapter 4 we have found that the age distribution depends on the time after the start of the experiment t (*i.e.* after the induction of transcription). Consequently, the polysome size distribution $Q_n(t)$ of an ensemble of mRNAs with age distribution $\phi_A(a \mid t)$ is time-dependent, *i.e.*

$$Q_n(t) = \int_0^\infty da\, q_n(a)\, \phi_A(a \mid t)\,. \tag{5.3}$$

Ribosome density From the polysome size distribution, we can easily derive the average polysome size by summing over all possible values of n. From Eq. (5.2) we find that the

average polysome size of an mRNA of age a reads

$$M(a) = \begin{cases} \omega_{\text{on}}\, a & a < t_L \\ \omega_{\text{on}}\, t_L & a \geq t_L \end{cases} . \tag{5.4}$$

Similarly, from Eq. (5.3) we obtain the average polysome size of an ensemble of mRNAs at time t after the induction of transcription, *i.e.*

$$\langle M \rangle_t = \int_0^\infty da\, M(a)\, \phi_A(a \mid t)\,. \tag{5.5}$$

Hence, we have obtained the mean ribosome density $\langle M \rangle_t / L$ for a species of mRNAs of length L at any point in time after the induction of transcription t.
The ribosomes are distributed along the mRNA transcript according to a profile density $m(z)$. Here, z denotes the codon position on the mRNA. Starting from the initial condition of no ribosomes bound to the mRNA, ribosomes progress on the mRNA with constant average elongation speed v. On average, after a time interval a the foremost ribosome has reached codon position $z = a\, v$. Upstream of the first ribosome, the average density can be assumed to be constant and it is given by the ratio of initiation rate ω_{on} and elongation speed v. Furthermore, the stretch of mRNA occupied by a single ribosome, *i.e.* the ribosomal *footprint* length l, has to be taken into account. As mentioned earlier, the size of a footprint is typically around 10 codons [44]. Thus, depending on the age a of a given mRNA of length L, its density profile along z reads

$$m(z \mid a) = \frac{\omega_{\text{on}}\, l}{v} \qquad\qquad a > z/v \tag{5.6}$$

and zero for $a \leq z/v$. Moreover, in each cell and at each observation point t after the induction of transcription, mRNAs of different age are present. Therefore, to obtain the average mRNA profile density we have to multiply Eq. (5.6) with the age distribution $\phi_A(a \mid t)$ and integrate over all possible ages a, *i.e.*

$$\langle m(z) \rangle_t = \int_0^t da\, m(z \mid a)\, \phi_A(a \mid t) = \frac{\omega_{\text{on}}\, l}{v} \int_{z/v}^t da\, \phi_A(a \mid t)\,. \tag{5.7}$$

It gives the average profile density along the mRNA for a given species for any age distribution and corresponds to the average ribosomal profile of that mRNA species. Note that specific effects such as codon dependent elongation rates are ignored here - Eq. (5.7) gives the profile of a homogeneous mRNA.

Computer simulations This simple theory of translation does not take into account the stochastic movement of the ribosomes and interactions between ribosomes. This becomes important when the ribosome density - which is governed by the translation initiation rate ω_{on} - is high. To include these effects, we have performed *TASEP* simulations. The basic principle of this simulation technique has been introduced in section 2.5. There, the stepping of each ribosome is a stochastic process that proceeds with a certain stepping probability. The ribosomes are modeled as finite size objects that have a hard-core repulsion. To account for the different ages of the mRNAs, the simulations have a random run length drawn from the age distribution $\phi_A(a \mid t)$. As a result, one can obtain the polysome and ribosomal profiles at each time t after the induction of transcription.

5.1.4. Results

In both the analytical description and the simulations, we have considered two different lifetime densities $\phi^{(\mathrm{exp})}$ and $\phi^{(\Gamma)}$ defined in (4.6) and (4.7), respectively, with an average lifetime $\langle U \rangle = 4$ minutes and an mRNA of $L = 1025$ codons, corresponding to the length of the lacZ gene in *E. coli*, see figure 5.2.

For a translation initiation rate $\omega_{\mathrm{on}} = 1/5\,\mathrm{sec}^{-1}$ [77] and an average velocity $v = 10$ codons per second [78], we observe that the theoretical description and results obtained by computer simulations agree very well (see figure 5.2). Hence, at least for the ribosome densities obtained from experiment, the analytical description is sufficient.

Moreover, we find that the polysome statistics and the profile densities depend only weakly

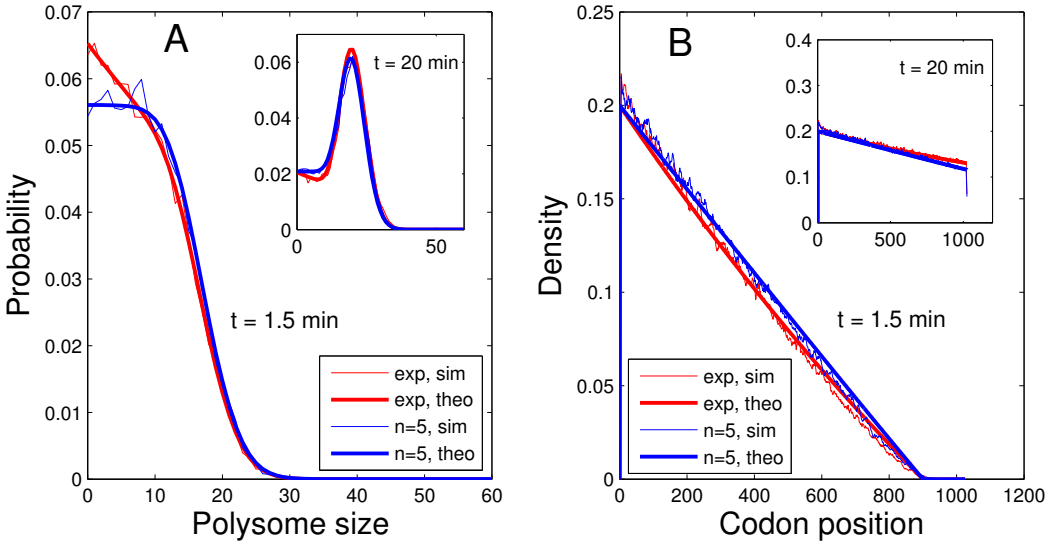

Figure 5.2.: **Time dependent polysome and ribosomal profiles.** Panel A shows the distribution of the number of ribosomes on an mRNA chain of 1025 codons at two different points in time, namely after $t = 1.5$ minutes and after $t = 20$ minutes. The two curves correspond to simulations based on the lifetime probability densities $\phi^{(\mathrm{exp})}$ (red) and $\phi^{(\Gamma)}$ (blue), respectively. Apart from the region around small polysome sizes, both distributions lead to similar polysome profiles. However, at different time points after induction there is a notable change of the profiles. Thus, this demonstrates that the outcome of such an experiment depends critically on the time of measurement after the induction of that particular gene. In panel B we display the profile density of the ribosomes along the mRNA. Shown is the probability that the corresponding codon is found covered by a ribosome at time t. Similar to the polysome profiles, the ribosome profile densities depend on the time of measurement t because the age composition of the sample changes with time during the transient. However, the ribosome profiles depend only weakly on the form of the underlying lifetime density ϕ_U in the present case. For the simulations, we have used typical parameters determined in experiments on *E. coli* . The rate of translation initiation has been fixed to $\omega_{\mathrm{on}} = 1/5\,\mathrm{sec}^{-1}$ [77] and the average velocity of ribosomes is 10 codons per second [78]. Each curve represents the average over 15000 independent realizations. For all plots shown here we have taken the average lifetime $\langle U \rangle = 4$ minutes. Both plots show also the predictions of the simple theory developed above (solid lines) that are in remarkably good agreement with the computer simulations. Hence, for the experimental parameters considered here, the simplifications of our analytical theory do not matter and give equally good results as the more detailed computer simulations.

on the underlying lifetime distribution (see figure 5.2). However, both quantities depend strongly on the time of measurement following the induction of transcription. This is due to the fact that both quantities depend on the age distribution of the mRNAs in the sample, which changes with time as we have seen. Therefore, it is important to take into consideration the time after the start of induction when performing an experiment on polysome profiles.
An additional effect of the heterogeneous age composition of the samples is given by the relatively large plateau in the polysome statistics at small polysomes. This plateau depends on the form of the lifetime densities. This implies that the polysome statistics and in particular the relative amount of mRNA with small polysomes carries a signature of the degradation process of the mRNA.

5.2. Effects of aging on protein synthesis

5.2.1. Rate of protein synthesis

Both the age distribution and the distribution of the amount of mRNA change in time depending on the shape of the lifetime probability density ϕ_U. This finding has implications on the rate of protein synthesis. The protein synthesis rate in a cell is determined by the amount of mRNA $N_r(t)$ and the ribosome flux on each mRNA. As pointed out in the previous section, mRNAs are polysomes on which several ribosomes are engaged in translation with an average elongation speed v. Thus, at low ribosomal densities, such as those found *in vivo* [55, 76], the average ribosome flux is given by the constant density ρ of ribosomes upstream of the leading ribosome and their average elongation speed v. However, in the process of translation there is a transient time t_L between initiation of translation and the time until the leading ribosome completes the synthesis of a protein. In eukaryotic cells and for certain prokaryotic organisms [79, 80], translation can initiate only after the whole mRNA has been synthesized. In contrast, whenever translation occurs co-transcriptionally [81] translation can initiate during the synthesis of the mRNA. Therefore, a functional definition of the lifetime of an mRNA is more appropriate here and the age $a = 0$ defines the moment when the first ribosomes binds to an mRNA. Consequently, the transient time t_L is proportional to the length of the mRNA and inversely proportional to the average elongation speed v, such that $t_L = L/v$. The consequence of this transient time t_L is that at any time t after the induction of transcription only those mRNAs that are older than t_L can contribute to the rate of protein synthesis. Under this perspective, t_L acts like a delay time that affects the rate of protein synthesis, such that

$$\omega_{\mathrm{ps}}(t) \;=\; v\,\rho\,N_r(t)\int_{t_L}^{t} \mathrm{d}a\,\phi_A(a \mid t) \text{ for } t > t_L \tag{5.8}$$

and zero otherwise, with ϕ_A defined in (4.8). The rate of protein synthesis, thus, depends both on the average lifetime of the mRNA and on the form of the lifetime density.
Figure 5.3 compares the time-dependence of the rate of protein synthesis for the lifetime distributions $\phi_U^{(\mathrm{exp})}$ (red) and $\phi_U^{(\Gamma)}$ (blue), as well as $\langle U\rangle = 4$ min (left panel) and $\langle U\rangle = 20$ min (right panel), respectively. For small average lifetimes, the difference in the age distribution at $a \sim t_L$ leads to a different stationary protein synthesis rate for different lifetime distributions. Conversely, for more stable mRNAs the time scale of translation is less relevant and the two example cases assume very similar stationary rates. Moreover, the protein synthesis rate per mRNA, $\omega_{\mathrm{ps}}(t)/N_r(t)$, reveals that the different transient of the synthesis rates is due to the different evolution of the mRNA number $N_r(t)$ (see insets in figure 5.3).

5.2.2. Protein amount

The turnover of proteins is described by their synthesis rate $\omega_{\mathrm{ps}}(t)$ and the protein degradation rate ω_{p} which we assume to be constant. Moreover, in case of very stable proteins, the degradation rate is effectively given by the rate of dilution due to cell division. Currently, it seems not to be feasible to find an analytical expression that accounts for the

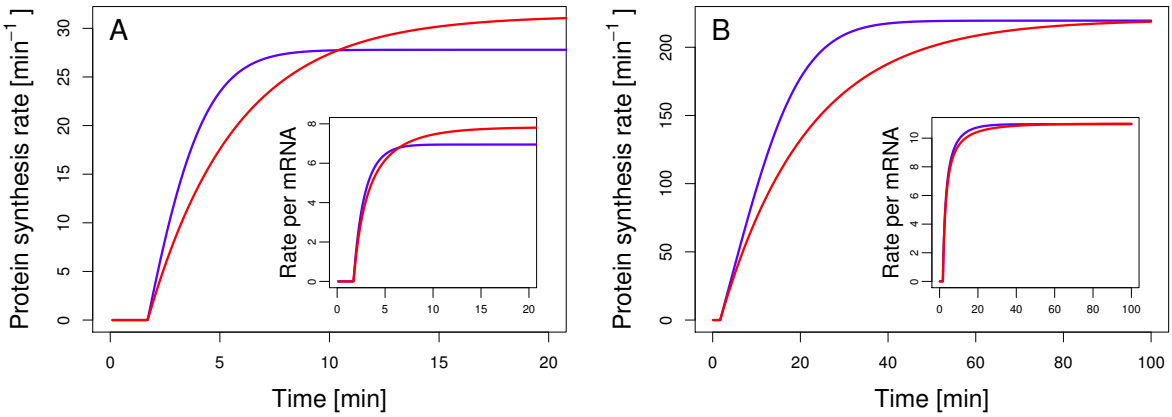

Figure 5.3.: **Average protein synthesis rate $\omega_{\rm ps}(t)$ over time after start of transcription**. The time evolution of $\omega_{\rm ps}(t)$ arises from the translation of the average number of mRNAs $N_r(t)$ at each point in time t, as derived in Eq. (5.8). The red lines represent $\omega_{\rm ps}(t)$ with a lifetime density $\phi^{(\rm exp)}$ defined in (4.6). The blue lines depict $\omega_{\rm ps}(t)$ with lifetime density $\phi^{(\Gamma)}$ defined in (4.7). In both cases we have assumed a ribosome density equal to 20% of the maximal packing [82], corresponding to $\rho = 0.02$ ribosomes per codon and a velocity of 10 codons per second. Panel A shows the average rate of protein synthesis under the two different lifetime densities when the average lifetime is $\langle U \rangle = 4$ min. For an initial interval of time, the rate of protein synthesis is smaller if the underlying lifetime density is exponential. At steady state, the exponential lifetime leads to a larger protein synthesis rate. Eq. (5.8) demonstrates that the total protein synthesis rate is given by both the number of mRNAs and their lifetime. The inset shows the evolution of the protein synthesis per mRNA $\omega_{\rm ps}(t)/N_r(t)$. The different total synthesis rates shortly after the start of transcription are governed by the different mRNA numbers of the two mRNA species (see figure 4.3 B) whereas the stationary rate follows from the different lifetime distributions. In panel B, instead, an average lifetime of $\langle U \rangle = 20$ min was chosen. Here, the two rates attain similar values only after about two hours. The final rates are closer than in the left panel because the contribution of t_L is smaller with respect to the average life time. In both plots we have fixed the length of the coding sequence $L = 1025$ codons, corresponding to the length of the lacZ gene in *E. coli* .

full stochasticity of both mRNA and protein turnover. However, we can derive the mean protein number N_p resulting from the population protein synthesis rate given in Eq. (5.8). The corresponding differential equation reads

$$\frac{d}{dt} N_p = \omega_{\rm ps}(t) - \omega_{\rm p} N_p \tag{5.9}$$

which can be solved by variation of the constant. Hence, the protein amount under the initial condition $N_p(0) = 0$ evolves in time according to

$$N_p(t) = e^{-\omega_{\rm p} t} \int_{t_L}^{t} d\tau\, \omega_{\rm ps}(\tau)\, e^{\omega_{\rm p} \tau} \tag{5.10}$$

and it remains zero as long as $t < t_L$.
In steady state, the protein number is governed by the equilibrium of synthesis and degradation. It is given by the limit $t \to \infty$ and it takes the following form

$$N_p^{\rm st} = \frac{\omega_{\rm ps}^{\rm st}}{\omega_{\rm p}}\,, \tag{5.11}$$

where $\omega_{\rm ps}^{\rm st} \equiv \omega_{\rm ps}(t \to \infty)$ according to Eq. (5.8).
Figure 5.4 A shows the time evolution of the protein amount after start of transcription. The time to a steady state is mainly governed by the protein degradation rate $\omega_{\rm p}$.
Note that the steady state protein level is partially determined from the lifetime densities since they give rise to different stationary age distributions as was pointed out in the previous section. However, this difference is small if the average lifetime is large compared to t_L (inset).
We can compare the result obtained from the time-dependent protein synthesis rate, Eq. (5.8), with a more simple model with a constant protein synthesis rate throughout the experiment. It is an effective model for protein synthesis because it ignores the turnover of mRNA. The evolution of the protein number is given by

$$\tilde{N}_p(t) = \frac{\omega_{\rm ps}^{\rm st}}{\omega_{\rm p}} \left(1 - \exp\{-\omega_{\rm p} t\}\right) , \tag{5.12}$$

as can be easily deduced from a Master equation. Here, the effective synthesis rate $\omega_{\rm ps}^{\rm st}$ is just given by the steady state value of Eq. (5.8). Therefore, both models lead to the same steady state protein level (see figure 5.4 A). Since the transient build-up of the mRNA level and the delayed protein synthesis is ignored here, this simplified description gives rise to a faster increase of the protein amount. This effect is more pronounced for longer mRNA lifetimes (see inset in figure 5.4 A). We can conclude that the mRNA turnover process becomes only negligible for the description of the average protein number when the stability of the mRNAs is much smaller than the stability of the proteins. In this case, the processes of mRNA and protein turnover occur on different time scales.

5.2.3. Protein distribution

It is desirable to also find a theoretical description for the distribution of the number of proteins. Let $\mathcal{P}(n_R, n_P, t)$ be the probability to have n_P proteins and n_R mRNAs at time t. The time evolution is given by

$$\begin{aligned} \frac{\partial \mathcal{P}(n_R, n_P, t)}{\partial t} &= \hat{\omega}_{\rm ps}\, n_R \left[\mathcal{P}(n_R, n_P - 1, t) - \mathcal{P}(n_R, n_P, t)\right] \\ &\quad + \omega_{\rm deg}(t) \left[(n_R + 1)\,\mathcal{P}(n_R + 1, n_P, t) - n_R\,\mathcal{P}(n_R, n_P, t)\right] \\ &\quad + \omega_{\rm tc} \left[\mathcal{P}(n_R - 1, n_P, t) - \mathcal{P}(n_R, n_P, t)\right] \\ &\quad + \omega_{\rm p} \left[(n_P + 1)\,\mathcal{P}(n_R, n_P + 1, t) - n_P\,\mathcal{P}(n_R, n_P, t)\right] . \end{aligned} \tag{5.13}$$

The first line describes the effect of protein synthesis with rate $\hat{\omega}_{\rm ps}$ per mRNA, the second line accounts for mRNA degradation with a time dependent rate $\omega_{\rm deg}(t)$, the third line the synthesis of mRNAs with rate $\omega_{\rm tc}$ and the last line the protein degradation with rate $\omega_{\rm p}$. Note that $\omega_{\rm deg}(t)$ describes the time-dependence of degradation resulting from an arbitrary lifetime distribution $\phi_U(t)$. Therefore, as should have become clear from chapter 2, $\omega_{\rm deg}(t)$ also depends on the age of the mRNAs (for more details see section A.4 in the appendix). It seems not to be possible to find an analytical solution of Eq. (5.13). However, in ref. [4] the moments of the protein number distribution are computed for some limit cases. There, the degradation of mRNA was considered with a constant decay rate.

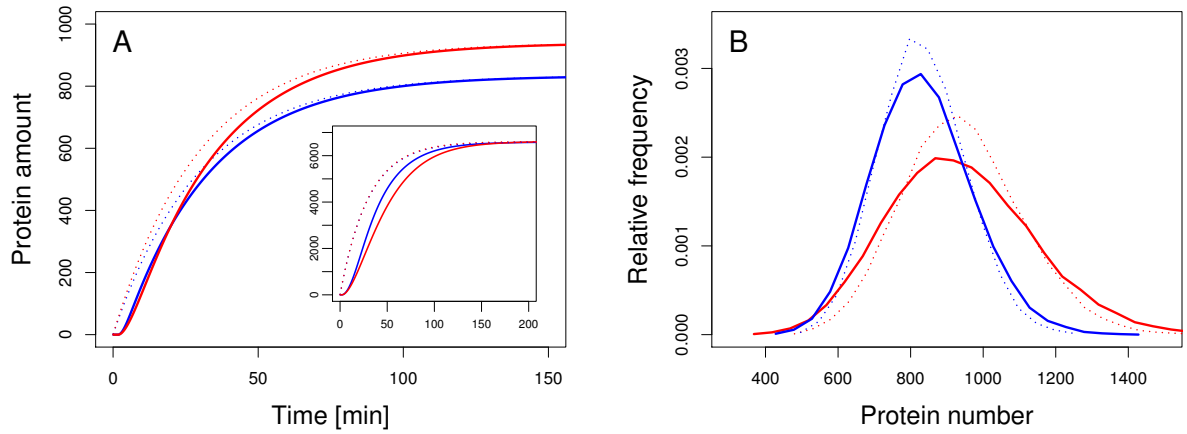

Figure 5.4.: **Time evolution of the protein amount and protein number distribution**. The left panel shows the temporal evolution of the average protein number as given by the time-dependent protein synthesis rate, Eq. (5.8), and a protein degradation rate $\omega_{\mathrm{p}} = 1/30\,\mathrm{min}^{-1}$. We compare the evolution of the protein turnover arising from mRNAs with the same average lifetime $\langle U \rangle = 4$ but different lifetime distributions $\phi_U^{(\mathrm{exp})}$ (red) and $\phi_U^{(\Gamma)}$ (blue). A stationary level is not reached before 150 min and the mRNAs with exponentially distributed lifetimes imply a higher steady state protein number due to their higher stationary synthesis rate, see figure 5.3. In the inset, the evolution is depicted for mRNAs with the same lifetime distributions but $\langle U \rangle = 20$ min. Hence, their steady state amount differs less since also their stationary synthesis rates are close, see figure 5.3. For comparison, the dotted lines show the time evolution of the average protein amount under the assumption of a constant rate of protein synthesis and thus ignoring the transient mRNA turnover, see Eq. (5.12). In this case, the evolution of the protein number is faster. This becomes more pronounced when the mean lifetime of the mRNAs is 20 min (see inset). In the right panel we depict the stationary distribution of the protein amount as found in a stochastic simulation of combined mRNA and protein turnover. We compare the distributions arising from the same exemplary mRNA lifetime distribution considered before with $\langle U \rangle = 4$ min. The distributions are relatively broad which result in a Fano factor of 43.1 (red) and 22.1 (blue), respectively. The dotted lines show the distributions without age effect, *i.e.* the numerical solution of Eq. (5.13). Here, the width of the distribution is smaller, giving rise to smaller Fano factors (28.6 and 16.8). Thus, the consideration of a delayed protein synthesis due to a constant time lag t_L leads to an increase of the inherent noise in gene expression.

We have simulated the protein turnover via a Monte-Carlo simulation (see section 2.5). The mRNAs are generated according to a Poisson process and each mRNA has a random lifetime according to a probability distribution (such as Eq. (4.7), with mean lifetime $\langle U \rangle$ and shape n fixed). Throughout the lifetime of an mRNA, proteins are synthesized at a constant rate corresponding to the steady state value of Eq. (5.8) divided by the number of mRNAs. However, the protein synthesis only starts after a time delay t_L which accounts for the time scale of translation introduced above. Finally, every protein has a random but finite lifetime which is determined by its degradation rate ω_{p} and at each instant of the simulation the number of live proteins can be determined. A large number of independent realizations of this process provide thus the distribution and average value of the amount of protein.

In figure 5.4 B, the results from the simulation is shown for $\langle U \rangle = 4$ min and $n = 1$ (red) and $n = 5$ (blue). The distributions are broad and centered around the mean

values we obtained from the analytical calculation, Eq. (5.11). We can compare the simulation results to a numerical solution of Eq. (5.13). In this case, the transient time t_L until the first protein is expressed, is ignored. Figure 5.4 B reveals that in this case the protein number distributions are more narrow which also results in a smaller Fano factor - defined as the variance divided by the mean - of the protein distribution. Thus, the additional incorporation of the time scale t_L increases the *noisiness* of the stochastic protein turnover.

5.3. Time evolution of the mRNA amount during the cell cycle

In most parts of the thesis, we have considered the situation that mRNAs are synthesized via a Poisson process from an initial amount $N_r(0) = 0$. We have seen that, depending on the lifetime distribution and average lifetime of an mRNA, the transient to steady state can be relatively long, which leads to the question if the mRNA level of a given gene can reach a steady state at all in the limited period of a cell cycle. In sections 4.6 and 4.7 we have already discussed how our formalism can be expanded to more realistic situations where there is a non-vanishing amount of mRNAs present at the time point of induction or stress applications. Here we want to advance the methodology to consider a simple model of the cell cycle.
The understanding of the mechanisms of the cell cycle is in itself a non-trivial matter. Many processes like the growth and division events are of stochastic nature [83, 84, 85, 86, 87]. Furthermore, eukaryotic cells undergo distinct phases and in each of the phases transcription and degradation of the mRNAs may be differently regulated. Therefore, in this thesis we can only consider a simple model of cell division. The constituents of each cell, e.g. its mRNA and protein content, increase during the growth phase of each cell. At the time of division, the molecules are split between the two daughter cells and the growth process starts anew. In this section we will first develop a deterministic division model that is based on our general formalism developed in chapter 2. We will compare these results with a more realistic model of cell division obtained from computer simulations. In a final part, we will contemplate the changes of the concentration of the mRNAs.

5.3.1. Deterministic model of cell division

In this model, we consider the evolution of mRNAs as the consequence of continuous transcription and mRNA degradation. During one cell cycle the conditions that determine transcription and degradation remain constant. Each cell cycle has a fixed duration t_{cc} and at division the mRNA content of a cell is equally distributed to its two daughter cells. At an arbitrary time $t = 0$, transcription of mRNAs starts and proceeds until the end of the first cycle at $t = t_{cc}$. Hence the growth of the amount of mRNA is described by Eq. (4.1),

$$N_r^{(1)}(t) = \omega_{\mathrm{tc}} \mathcal{H}_0^t \,. \tag{5.14}$$

Here and in the following we use the notation $\mathcal{H}_{t_1}^{t_2} \equiv \int_{t_1}^{t_2} \mathrm{d}u \, (1 - \Phi_U(u))$ which we introduced in Eq. (2.16). For a generic example, the growth of the average amount of mRNA is illustrated in figure 5.5 A (red line). The evolution of the mRNA amount during the second cycle can by understood as the sum of the contribution of the newly synthesized mRNAs and the decaying amount of the first cycle,

$$N_r^{(2)}(t) = \omega_{\mathrm{tc}} \mathcal{H}_0^{t-t_{cc}} + \frac{\omega_{\mathrm{tc}}}{2} \mathcal{H}_{t-t_{cc}}^{t} \tag{5.15}$$

for $t_{cc} \leq t < 2\, t_{cc}$. The first term describes the growth of newly transcribed mRNAs which result from a transcription process started at $t = t_{cc}$ (blue line in figure 5.5 A). The second term describes the decay of the remaining mRNAs from the first cycle where

the factor of 1/2 arises from the division at $t = t_{cc}$ (red line in figure 5.5 A). In the third cycle ranging from $2\,t_{cc} < t \leq 3\,t_{cc}$, a third transcription process is started (green line). Moreover, at $t = 2\,t_{cc}$ only half of the mRNA amount prior to division is still present. Hence, the mRNA population generated during the second cycle is halved (blue) while the mRNAs originating in the first cycle (red) underwent division twice - thus the factor of 1/4. Consequently, the mRNA amount evolves according to

$$N_r^{(3)}(t) = \omega_{\mathrm{tc}} \mathcal{H}_0^{t-2\,t_{cc}} + \frac{\omega_{\mathrm{tc}}}{2} \mathcal{H}_{t-t_{cc}}^{t} + \frac{\omega_{\mathrm{tc}}}{4} \mathcal{H}_{t-2\,t_{cc}}^{t-t_{cc}} \,. \tag{5.16}$$

This scheme can be straightforwardly generalized to the kth cycle which ranges from $(k-1)\,t_{cc} < t \leq k\,t_{cc}$

$$N_r^{(k)}(t) = \omega_{\mathrm{tc}} \mathcal{H}_0^{t-(k-1)\,t_{cc}} + \omega_{\mathrm{tc}} \sum_{j=0}^{k-2} 2^{-(j+1)} \mathcal{H}_{t-(j+1)t_{cc}}^{t-j\,t_{cc}} \,. \tag{5.17}$$

The evolution of the mRNA amount during the first 4 cycles is visualized in figure 5.5 A. Different colors indicate the contribution of mRNAs originating from different cycles. The sum of the contributions gives the evolution of the total amount (black) as described by Eq. (5.17).

5.3.2. Stochastic cell cycle simulations

To grasp some more details of the inherent stochasticity of the cell division process, we have complementarily studied the problem with computer simulations. The general concept of stochastic simulations has been pointed out in section 2.5 and an algorithm is given in the appendix (section A.6). First, we extend the description of cell division by considering a stochastic distribution of the mRNA molecules. At division each mRNAs goes with probability 1/2 into the first daughter cell whereas with the same probability it terminates in the second daughter cell. Hence, the number of mRNAs staying in one cell is binomially distributed. Secondly, the duration of a cell cycle is also random. Therefore, we have considered a log-normally distributed cell cycle duration with the same mean value t_{cc} as in the deterministic model. The variance of the cell cycle duration distribution was set to $t_{cc}/2$.
In figure 5.5 B we compare results from the deterministic model and the computer simulations. We consider an mRNA with a lifetime distribution $\phi_U^{(\Gamma)}$ with shape parameter $n = 5$ and average lifetime $\langle U \rangle = 30$ min. Furthermore, the transcription rate $\omega_{\mathrm{tc}} = 1\,\mathrm{min}^{-1}$ and the average cell cycle duration $t_{cc} = 30$ min. In the deterministic case, the evolution of the average mRNA number reaches a stable pattern after the second cycle (red). Thus, in all subsequent cycles the time evolution is identical. One realization of the simulation leads to strong fluctuations due to the inherent stochasticity (gray). However, an average over 1000 realizations yields an oscillating pattern of the number of mRNAs with the period of the average cell cycle duration (blue). After several cycles the oscillation decays and approaches a constant, stationary mRNA level.
We can understand this value by considering the deterministic model again. During each cycle the mRNA amount evolves from a minimum to a maximum value according to Eq.

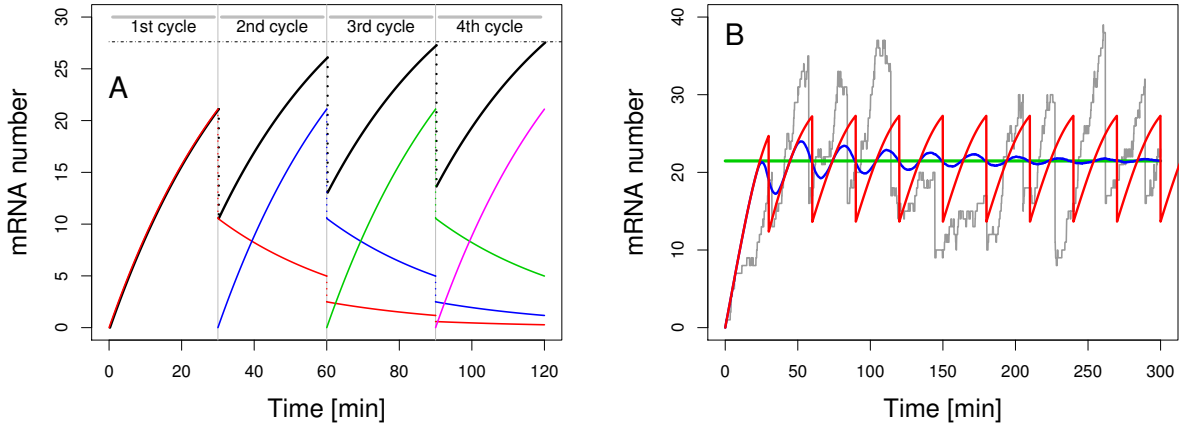

Figure 5.5.: **mRNA evolution and cell division**. The left panel shows a scheme how one can find an analytical expression for the mRNA evolution in a deterministic model of cell division. Each cycle has a fixed duration t_{cc} and the mRNA population is split in half at each division. Each color marks the mRNA created within one cell cycle and the black line gives the sum of all mRNAs at each time. After a few cycles, the maximum amount is reached (as indicated by the black dashed line). The right panel shows a comparison of the deterministic cell division model with $n = 5$, $\langle U \rangle = 30$ min and $t_{cc} = 30$ min (red), see Eq. (5.17) and results obtained from Monte-Carlo simulations with stochastic division and random cycle duration. A single realization is governed by large fluctuations in the mRNA amount due to stochastic division and varying cycle duration (gray). The average mRNA amount obtained from a large number of realizations shows only an oscillating behavior which decays to a stationary value when sufficiently far from the initial configuration (blue). This stationary average value obtained by computer simulations can be understood with the help of the deterministic model. The green line gives the average mRNA number over one deterministic cycle, Eq. (5.18), long after the division process was started. Clearly, the average from many simulations (blue) agrees with this value.

(5.17). The average mRNA amount during one cycle is given by

$$\langle N_r^{(k)} \rangle = \frac{1}{t_{cc}} \int_0^{t_{cc}} d\tau \, N_r^{(k)}(\tau) \,. \tag{5.18}$$

After several cycles, *i.e.* far from the initial configuration, Eq. (5.18) becomes constant. Conversely, in a single realization of the stochastic model there is no stationary cell cycle average. However, the average over many realization leads to the same *average* cell cycle duration and to the same *average* dilution. Indeed, figure 5.5 B reveals that the cell cycle average as given by Eq. (5.18) (green) agrees with the long term limit of the stochastic simulation (blue).

Hence we can conclude that although a stationary mRNA level might not be reached within one cycle, the pattern of mRNA evolution within one cycle does become stationary. The additional stochastic aspects considered in the simulations prolongs the transient regime but also eventually reaches a stationary cell cycle average. More stretched mRNA lifetime distributions and larger average lifetimes increase the number of required cycles, however the mRNA turnover will always reach a stationary regime. It should be clear that a similar analysis can also be performed for the evolution of the protein amount. There, one would expect that even longer lifetimes of the proteins and a non-constant

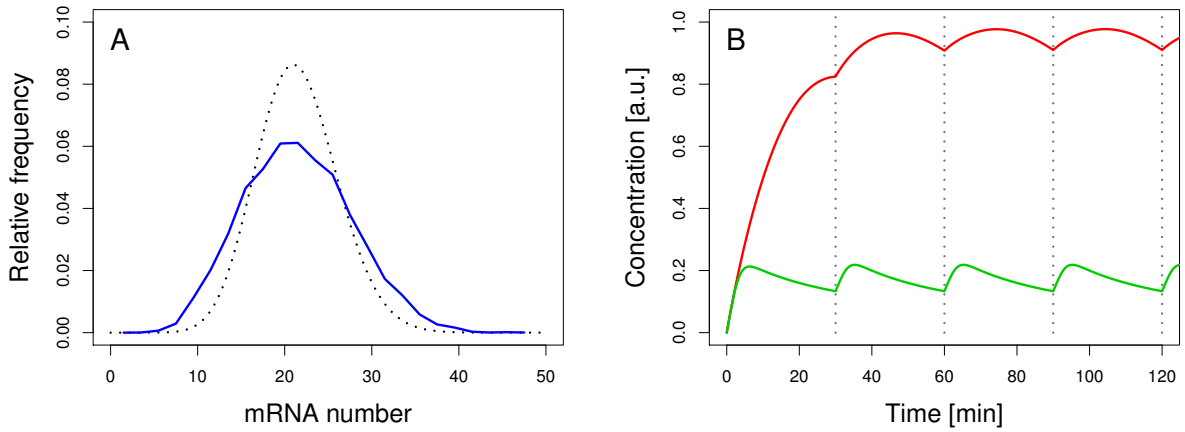

Figure 5.6.: **Cell to cell variability and mRNA concentrations**. In the left panel, the cell-to-cell variability of the mRNA amount is shown for the stochastic cell division model introduced above (blue). The corresponding mRNA number distribution is broader than a Poisson distribution (black) and the Fano factor reads 1.8. Thus, the cell division process is an additional source of stochasticity in mRNA turnover. In the right panel, we follow the mRNA concentration, *i.e.* the number divided by the cell volume, over several cell cycles. For the calculation of the deterministic cell cycle model, Eq. (5.17), we have assumed a cell cycle duration $t_{cc} = 30$ min and linear volume growth. Whereas the green line represents mRNAs with $\phi_U^{(\Gamma)}$ and $\langle U \rangle = 4$ min, the red line gives the evolution for $\langle U \rangle = 30$ min. In both cases, the maximum of the concentration is obtained at an early time in the cell cycle and the concentration decreases thereafter. This becomes evident once one realizes that the mRNA amount approaches a steady state depending on the lifetime distribution and average lifetime whereas in the simple model of volume growth considered here the volume increases continuously.

synthesis rate would lead to an even longer transient. However, if the number of cycles is sufficiently large, one will always attain a stationary mRNA and protein expression pattern.
We can also assess the cell-to-cell variability of the mRNA amount for the present stochastic cell division model (see figure 5.6 A). The corresponding mRNA number distribution is broader than a Poisson distribution and the Fano factor reads 1.8. Thus, the cell division process is an additional source of stochasticity in mRNA turnover.

5.3.3. Effects of volume growth and molecule concentrations

In a cell, not only the absolute number of molecules of each kind is important but also its concentration. The concentration is simply given by the ratio of number vs. cell volume and both are subject to change during the cell cycle. As long as we consider cell populations only, the population is in an equilibrium situation with regard to the cell volume. The average volume is held constant. Hence, the concentration is just the average molecule number divided by the average volume. However, once one considers single cells, the volume of the cell increases from an initial value after division to its final value before division. In a simple deterministic model, one starts with a fixed initial volume V_0 which expands during the cell cycle period t_{cc} to twice its value, *i.e.* $V(t_{cc}) = 2\,V_0$. Different growth models have been considered in the literature [88, 89]. In particular, *E. coli*

bacteria elongate in one dimension only and divide somewhere close to the center [86].
In the following, we will investigate the change of concentration during one cell cycle in a linear growth model,

$$V(t) = V_0 \left(1 + \frac{t}{t_{cc}}\right) \tag{5.19}$$

Here, the cycle spans from $0 \leq t \leq t_{cc}$ and the volume is divided by two at $t = t_{cc}$. It is however straightforward to devise other growth models and follow the change of concentration in time.
In a more realistic model, the cell volume before and after division as well as the cell division time are random quantities. One can account for this randomness in Monte-Carlo simulations similar to those described earlier in this section.
We can follow the change of concentration during the cell cycle (see figure 5.6 B). Both molecule number and volume increase in time and the concentrations grow to a maximum value at a specific time during the cell cycle. After some time the mRNA level approaches a steady state while the cell volume grows continuously. Hence, there is a maximum of the concentration followed by a decrease. Once the mRNA number has reached its maximal level, further cycles reveal the same concentration pattern. In a cell such as *E. coli*, the DNA duplicates during the cell cycle so that more than one copy of each gene is present. This can be understood as a doubling of the transcription rate but has not been considered here.

5.4. Chapter summary

In this chapter, we have elucidated implications of the extended theory of mRNA degradation for the processes of translation and protein synthesis. We developed a theoretical description for low density ribosome traffic and examined how ribosome profiles depend on the non-stationary mRNA number and age distribution. Moreover, using simulations we could validate these results and extend the analysis for any ribosome flux. In addition, for the rate of protein synthesis and the evolution of the amount of proteins, an analytical description was found (see section 5.2). Conversely, the corresponding protein number distributions have been obtained by Monte-Carlo simulations. One of the main finding of this chapter is that the processes of translation and protein synthesis bear transient phenomena that originate from the turnover of the mRNAs. Furthermore, we have addressed the question what this non-stationarity implies to the cell cycle. There, we have developed an analytical and deterministic model of cell division and compared it to stochastic simulations. We found that after a limited number of cycles the mRNA level approaches a stationary expression pattern.
Section 5.1 reviews and expands some published results [50], whereas sections 5.2 and 5.3 are largely original research, part of which already published [74].

6. Summary and outlook

6.1. Summary

In this thesis, we have developed a general theoretical formalism of mRNA expression. Our focus was on an improved description of the mRNA degradation. It extends previous theoretical approaches by considering general mRNA lifetime distributions. This was motivated by the failure of existing models to accurately describe mRNA decay experiment. Moreover, it provides a framework to incorporate the knowledge of the underlying biochemical mechanisms of degradation.
In chapter 2 we have developed most of the theoretical tools required for the remaining part of the thesis. Under relatively general and mild mathematical conditions, we have found that the number of mRNAs can be described by a Poisson distribution with time dependent parameter regardless of the lifetime distribution. We could show that this includes the findings of a simpler theory, where the lifetimes are restricted to be exponentially distributed, as a limiting case.
In chapter 3 we used the theoretical findings to improve the understanding of mRNA decay experiments. We find some general results which hold for any decay mechanism. Thus we can identify experimental decay patterns that are *bona-fide*. This is important because most experiments reveal data that are difficult to interpret - presumably because of the perturbing nature of the decay experiments. Furthermore, we developed two specific models that describe the degradation in two model organisms, *S. cerevisiae* and *E. coli*. In the first model, we could map the complexity of different decay mechanisms into a model based on Markov chains. Thus, we could improve the interpretation of experiments considerably. We find different categories of decay patterns which are strictly related to the underlying decay mechanism. In a second model, we have pursuit a different strategy where we modeled the relation of degradation to the state of translation of a given mRNA. This resembles the situation in bacteria such as *E. coli*. Again, our approach led to an improved description of experimental data and to a more accurate evaluation of the average mRNA lifetimes. Moreover, we could thus devise a strategy to assess the role of co-transcriptional translation in bacteria.
Chapter 4 dealt with the role of degradation in the expression levels of mRNAs. On the one hand, we have found that the transient evolution of the mRNA amount after induction of a gene is governed by the lifetime distribution of an mRNA. Non-exponential lifetime distributions become manifest in the aging of mRNA molecules. We have found powerful criteria such as the age and residual lifetime distribution to assess the aging of mRNAs. On the other hand, we have considered the situation that the rate of transcription changes during an experiment. This is particularly useful when trying to understand experiments where cells where subjected to stressed conditions. With the theoretical tools developed in this chapter, we can understand some aspects of the mRNA expression changes under

stressed conditions. There, our theory allows to differentiate between effects due to a changed mode of transcription and degradation, respectively.
In chapter 5 we addressed the implications of our improved description of mRNA turnover in a larger setting. We have investigated its role for the processes of translation and protein synthesis. Since all processes in gene expression are tightly coupled, the transient evolution of the mRNA amount governs the time scales of these processes as well. Furthermore, we have pointed out how the mRNA amount evolves in single cells by devising an appropriate cell division model.

6.2. Further perspectives

All quantitative experiments in modern biology require a proper model for their interpretation. Since previous models lead to unsatisfactory results, our theory provides an enlarged toolbox for the interpretation of further experiments. Moreover, our theoretical description hints how the design of the experiments can be improved by stressing the fact that only a high temporal resolution can yield an accurate derivation of the lifetime distribution of an mRNA.
Furthermore, our theory establishes a relation between mRNA decay experiments and the underlying biochemical mechanisms of decay. In experiments, the role of a specific mechanism is often studied by merely looking at mRNA half-lives. Conversely, we provide a tool how different mechanisms can be validated by the experimental decay patterns. Therefore, our formalism based on a Markov chain can be readily expanded to the current knowledge of the biochemistry of mRNA stability regulation. An advantage of our description is that it is general and we provide straightforward ideas how mechanisms can be incorporated into falsifiable models. Furthermore, it should have become clear that the modeling is open for alternative strategies to obtain a suitable lifetime distribution. These can be easily incorporated into the general formalism introduced in this thesis.
Hence, our theoretical framework and the strategies related to find the appropriate lifetime distributions lead to a more complete picture of gene expression. It opens a possibility to better understand and design experiments on translation and protein synthesis.
A better comprehension of the role of transcription and degradation in the role of gene expression is mandatory to fully understand the level of mRNA expression. Hence, our description should be able to enhance the interpretation of many experiments that pursuit the better understanding of mRNA expression under aberrant conditions.

List of Symbols

t	Time	9
$X(t)$	Random number of mRNAs generated in a time interval t	9
U	Random lifetime of an mRNA	9
ϕ_U	Lifetime probability density of an mRNA of a given gene	9
$Y(t)$	Random number of mRNAs not degraded at time t	9
Δt	Time elapsed after the interruption of transcription	9
t_s	Time point of interruption of transcription: $t_s = t - \Delta t$	9
ω_{tc}	Transcription rate of a given gene	12
N_{cells}	Number of cells	13
Ω_{tc}	Population transcription rate	13
$\omega_i(t)$	Time-dependent transcription rate of cell i	13
Φ_U	Cumulative probability distribution function of ϕ_U	13
$\langle U \rangle$	Average lifetime of an mRNA as given by ϕ_U	13
p	Survival probability of a single mRNA	14
O	Random time point of origination of an mRNA	14
Z	Sum of random variables $O + U$	14
$H(t)$	Parameter of the mRNA distribution at time t after start of transcription	15
p_s	Survival probability of a single mRNA after stop of transcription	15
$H_2(\Delta t)$	Parameter of the mRNA distribution at time Δt after stop of transcription	15
$\mathcal{H}$	Common abbreviation for combination of $H(t)$ ans $H_2(t)$	16
A	Random age of an mRNA at time t: $A = t - O$	16
$\phi_A(a \mid t)$	Age probability density at time t	16
R	Random residual lifetime of an mRNA at time t: $R = O + U - t$	16
$\phi_R(r \mid t)$	Residual lifetime probability density at time t	16
ω_{r}	mRNA decay rate constant	18
$P_k(t)$	Probability of having k mRNAs at time t	18
$\omega_{\text{deg}}(a)$	Age-dependent degradation rate	19
$\wp$	Survival probability of an mRNA	19
$\mathcal{Q}$	Transition rate matrix of a Markov chain	20
$\mathcal{S}$	Matrix of internal transitions of a Markov chain	20
$\mathbf{S_0}$	Matrix of transitions to the absorbing states of a Markov chain	20
$\Phi(t)$	Cumulative probability of time to absorption	20
$\boldsymbol{\alpha}$	Vector of initial values of Markov chain	20
$\phi(t)$	Probability density of time to absorption	20
λ_i	Transition rate to next non-absorbing state from state i	21
μ_i	Transition rate to absorbing state from state i	21
δt	Time increments in Monte-Carlo simulations	23
T	Total run time in Monte-Carlo simulations	23

v	Elongation speed of a ribosome	24
l	Number of codons a ribosome occupies on an mRNA	24
$N_r(t)$	Time-dependent number of mRNAs	26
$N_r^{\downarrow}(\Delta t)$	Number of mRNAs at Δt after the interruption of transcription	26
$\Lambda(\Delta t)$	Relative number of mRNAs at Δt after the interruption of transcription	26
$t_{1/2}$	mRNA half-life	26
$\langle R(\Delta t)\rangle$	Average residual lifetime of an mRNA at time Δt	33
$\mathcal{C}(\Delta t)$	Residual protein synthesis capacity	34
t_L	Time a ribosome takes to transverse the coding sequence of an mRNA	36
L	Length of the coding sequence of an mRNA	36
a	Age of an mRNA, here considered as a time parameter	37
ω_a	Degradation rate of mRNAs in absence of ribosomes	37
ω_b	Degradation rate of mRNAs fully covered by ribosomes	37
ρ	Ribosome density on an mRNA	38
$N_r^{\uparrow}(t)$	Time-dependent mRNA amount after the induction of transcription	46
$\phi_U^{(\mathrm{exp})}(t)$	Exponential lifetime probability density	48
$\phi_U^{(\Gamma)}(t)$	Gamma lifetime probability density	48
n	Shape parameter of the gamma distribution	48
λ	Rate parameter of the gamma distribution	48
$\phi_A(a \mid t)$	Age probability density after induction of transcription	51
$\phi_A^{\mathrm{st}}(a)$	Age probability density at steady state	51
$\langle A\rangle_t$	Average age after induction of transcription	51
$\phi_R(r \mid t)$	Residual lifetime probability density	51
$\langle R\rangle_t$	Average residual lifetime after induction of transcription	52
K	Random variable that describes the state in a Markov chain	54
$\xi_k(a)$	Probability for an mRNA of age a to be in state k	54
$\psi_k(t)$	Probability of an ensemble of mRNAs to be in state k at time t	55
$\langle k(t)\rangle$	Average state of an ensemble of mRNAs at time t	55
t^{st}	Approximate time to steady state mRNA amount	57
q_n	Probability that n ribosomes are bound to an mRNA	65
n	Number of ribosomes bound to an mRNA	65
$Q_n(t)$	Polysome size distribution of an ensemble of mRNAs at time t	20
$M(a)$	Average polysome size of an mRNA of age a	66
$\langle M\rangle_t$	Average polysome size of an ensemble of mRNAs at time t	66
z	Codon position on an mRNA	66
$m(z \mid a)$	Ribosome profile density of an mRNA of age a	66
$\langle m(z)\rangle_t$	Ribosome profile density of an ensemble of mRNAs at time t	66
ω_{on}	Translation initiation rate	67
$\omega_{\mathrm{ps}}(t)$	Protein synthesis rate at time t	69
$N_p(t)$	Protein amount at time t after the start of induction	70
ω_{p}	Protein degradation rate	69
N_p^{st}	Steady-state protein amount	70
$\omega_{\mathrm{ps}}^{\mathrm{st}}$	Steady-state protein synthesis rate	70
$\tilde{N}_p(t)$	Time-dependent protein amount in effective model without mRNA turnover	71
$\mathcal{P}$	Probability to have n_p Proteins and n_R mRNAs at time t	71

n_R	Number of mRNAs in one realization	71
n_P	Number of proteins in one realization	71
t_{cc}	Average cell cycle duration	74
$N_r^{(k)}(t)$	mRNA amount in the kth cycle	75
$\langle N_r^{(k)} \rangle$	Average mRNA amount during cycle k	76
$V(t)$	Cell volume at time t in the cell cycle	77
V_0	Initial cell volume	77

Glossary of biological terms

3' UTR
Part of the mRNA sequence following the stop codon

5' CAP
Protective end on the 5′ side of an mRNA consisting of a methylated guanine nucleotide

5' UTR
Part of the mRNA sequence upstream of the start codon

B. subtilis
Bacillus subtilis. A bacterium frequently studied in cell biology

Coding sequence
Sequence of nucleotides of an mRNA that codes for a protein

Codon
Triplet of nucleotides that codes for an amino acid (or termination)

Cytosol
Solution that contains all parts of a cell inside the cell membrane

Deadenylation
Process of removing the Poly(A)-tail of an mRNA

Decapping
Process of removal of the phosphate-CAP upstream of the 5′ UTR

Degradosome
Complex of enzymes in *E. coli* that is responsible for degradation of mRNAs

E. coli
Escherichia coli, the prokaryotic *standard model* of cell biology

Elongation
Process of elongating the mRNA transcripts in transcription or the polypeptide chains in translation

Endonuclease
Class of enzymes that cleave an mRNA between two nucleotides

Eukaryote
Domain of organisms that have a cell nucleus

Exonuclease
Class of enzymes that processively degrade mRNA from one of its ends

Exosome
Complex of enzymes in eukaryotes that is responsible for degradation of mRNAs

Footprint
The position of ribosomes on a target mRNA

Homologue
Group of genes, RNAs or proteins with similar or identical function

Initiation
Start of process of transcription or translation

Northern blot
Experimental method based on gel electrophoreses to quantify mRNA expression in a cell assay

Metabolic labeling
Experimental method in which nucleotides are labeled via heavy or radioactive isotopes

miRNA
microRNA: Short sequence of nucleotides that binds to complementary parts of mRNAs resulting in gene silencing

Microarray
Experimental method to quantify the genome wide mRNA expression

mRNA
messenger RNA

Poly(A)
Sequence of adenosine nucleotides following the 3′ UTR

Plasmid
DNA molecules that are separate from the chromosomal DNA. Recombinant plasmids are used to express artificial genes in an organism

Phosphorylation
The addition of phosphate groups to a protein that can alter the function and activity of a protein

Polypeptide chain
Sequence of amino acids that makes up a protein

Polysome
A cluster of ribosomes that are attached to an mRNA

Pre-mRNA
Precursor mRNA denotes the transcript in eukaryotes before *splicing* and *UTR* addition

Prokaryote
Group of organisms that lack a cell nucleus, *i.e.* bacteria and archaea

Protease
Enzyme or enzyme complex which degrades proteins

qRT-PCR
Real-time reverse-transcription Polymerase Chain Reaction. A quantitative method for measuring gene expression

Ribosome
Complex that translates information encoded in mRNA into polypeptide chains

RNA polymerase
Enzyme complex that transcribes DNA to mRNA

RNase E
Primary endonuclease in *E. coli*

RNA-seq
Experimental high-throughput method to quantify the mRNA expression in a cell assay - also known as *deep sequencing*, *next-generation sequencing*, *Whole Transcriptome Shotgun Sequencing*

S. cerevisiae
Baker's yeast. One of the simplest eukaryotic organisms

Splicing
A process which modifies RNAs in eukaryotes: Introns are removed and exons are joined.

Transcription
The process of transcribing the genetic information from DNA to mRNA

Transcription factors
Proteins that can activate or repress the recruitment of RNA polymerases

Translation
The process of translating the genetic information from mRNA into a polypeptide chain

tRNA
transfer RNA

List of Figures

Bibliography

[1] Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, August 1970.

[2] Bruce Alberts, editor. *Molecular biology of the cell*. Garland Science Taylor & Francis, New York, 5. ed. edition, 2008.

[3] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, August 2002.

[4] Peter S Swain, Michael B Elowitz, and Eric D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, October 2002.

[5] Jonathan M. Raser and Erin K. O'Shea. Noise in gene expression: Origins, consequences, and control. *Science*, 309(5743):2010 –2013, 2005.

[6] H.F. Lodish. *Molecular cell biology*. W.H. Freeman, 2008.

[7] Douglas F. Browning and Stephen J. W. Busby. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2(1):57–65, January 2004.

[8] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, Thomas Kuhlman, and Rob Phillips. Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development*, 15(2):125–135, April 2005.

[9] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, and Rob Phillips. Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*, 15(2):116–124, April 2005.

[10] Rob Benne. RNA editing: how a message is changed. *Current Opinion in Genetics & Development*, 6(2):221–231, April 1996.

[11] Susana Rodríguez-Navarro and Ed Hurt. Linking gene regulation to mRNA production and export. *Current Opinion in Cell Biology*, 23(3):302–309, June 2011.

[12] Marilyn Kozak. Initiation of translation in prokaryotes and eukaryotes. *Gene*, 234(2):187–208, July 1999.

[13] C. Walsh. *Posttranslational modification of proteins: expanding nature's inventory*. Roberts and Co. Publishers, 2006.

[14] Joel G. Belasco. All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat. Rev. Mol. Cell Biol.*, 11(7):467–478, July 2010.

[15] Sidney R. Kushner. mRNA Decay in Escherichia coli Comes of Age. *J. Bacteriol.*, 184(17):4658–4665, September 2002.

[16] Murray P. Deutscher. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Research*, 34(2):659 –666, 2006.

[17] Joerg E. Braun, Eric Huntzinger, Maria Fauser, and Elisa Izaurralde. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Molecular Cell*, 44(1):120–133, October 2011.

[18] Eric Huntzinger and Elisa Izaurralde. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, 12(2):99–110, February 2011.

[19] Vlad Elgart, Tao Jia, and Rahul Kulkarni. Quantifying mRNA synthesis and decay rates using small RNAs. *Biophys. J.*, 98(12):2780–2784, June 2010.

[20] Erel Levine, Peter McHale, and Herbert Levine. Small regulatory RNAs may sharpen spatial expression patterns. *PLoS Comput Biol*, 3(11):e233, November 2007.

[21] Wenqian Hu, Thomas J. Sweet, Sangpen Chamnongpol, Kristian E. Baker, and Jeff Coller. Co-translational mRNA decay in Saccharomyces cerevisiae. *Nature*, 461(7261):225–229, 2009.

[22] Sergine Even, Olivier Pellegrini, Lena Zig, Valerie Labas, Joelle Vinh, Dominique Bréchemmier-Baey, and Harald Putzer. Ribonucleases J1 and J2: two novel endoribonucleases in B.subtilis with functional homology to E.coli RNase E. *Nucleic Acids Research*, 33(7):2141 –2152, 2005.

[23] Ciarán Condon. What is the role of RNase J in mRNA turnover? *RNA Biology*, 7(3):316–321, 2010.

[24] Roy Parker, David Herrick, Stuart W. Peltz, Allan Jacobson, and Gerald R. Fink Christine Guthrie. Measurement of mRNA decay rates in saccharomyces cerevisiae. In *Guide to Yeast Genetics and Molecular Biology*, volume Volume 194, pages 415–423. Academic Press, 1991.

[25] L. Maquat and C. Arraiano. *RNA Turnover in Bacteria, Archaea and Organelles*. Methods in Enzymology. Academic Press, 2008.

[26] Pei-Hsun Lin, Dharam Singh, Jonathan A Bernstein, and Sue Lin-Chao. Genomic analysis of mRNA decay in e. coli with DNA microarrays. *Methods in Enzymology*, 447:47–64, 2008.

[27] Ophir Shalem, Orna Dahan, Michal Levo, Maria Rodriguez Martinez, Itay Furman, Eran Segal, and Yitzhak Pilpel. Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Molecular Systems Biology*, 4, October 2008.

[28] Douglas W. Selinger, Rini Mukherjee Saxena, Kevin J. Cheung, George M. Church, and Carsten Rosenow. Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation. *Genome Research*, 13(2):216 –223, February 2003.

[29] Claudia Steglich, Debbie Lindell, Matthias Futschik, Trent Rector, Robert Steen, and Sallie W Chisholm. Short RNA half-lives in the slow-growing marine cyanobacterium Prochlorococcus. *Genome Biology*, 11(5):R54, 2010.

[30] Howard M. Taylor and Samuel Karlin. *An introduction to stochastic modeling*. Academic Press, San Diego, 3rd edition, 1999.

[31] A. Kepes. Kinetic of induced enzyme synthesis determination of the mean life of galactosidase-specific messenger RNA. *Biochim. Biophys. Acta*, 76:293–309, 1963.

[32] Purnananda Guptasarma. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of escherichia coli? *BioEssays*, 17(11):987–997, November 1995.

[33] Mukund Thattai and Alexander van Oudenaarden. Intrinsic noise in gene regulatory networks. 98(15):8614 –8619, July 2001.

[34] Yuichi Taniguchi, Paul J. Choi, Gene-Wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X. Sunney Xie. Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science*, 329(5991):533 –538, July 2010.

[35] Daniel Zenklusen, Daniel R Larson, and Robert H Singer. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature Structural & Molecular Biology*, 15(12):1263–1271, December 2008.

[36] Daniel R. Larson, Daniel Zenklusen, Bin Wu, Jeffrey A. Chao, and Robert H. Singer. Real-Time observation of transcription initiation and elongation on an endogenous yeast gene. *Science*, 332(6028):475 –478, April 2011.

[37] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling*. ASA-SIAM series on statistics and applied probability. Society for Industrial and Applied Mathematics, 1999.

[38] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[39] Nir Friedman, Long Cai, and X. Sunney Xie. Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Physical Review Letters*, 97(16):168302, October 2006.

[40] Carolyn T MacDonald and Julian H Gibbs. Concerning the kinetics of polypeptide synthesis on polyribosomes. *Biopolymers*, 7(5):707–725, May 1969.

[41] Leah B. Shaw, R. K. P. Zia, and Kelvin H. Lee. Totally asymmetric exclusion process with extended objects: A model for protein synthesis. *Physical Review E*, 68(2):021910, 2003.

[42] Apoorva Nagar, Angelo Valleriani, and Reinhard Lipowsky. Translation by Ribosomes with mRNA Degradation: Exclusion Processes on Aging Tracks. *Journal of Statistical Physics*, 145:1385–1404, 2011.

[43] B. Derrida. An exactly soluble non-equilibrium system: The asymmetric simple exclusion process. *Physics Reports*, 301(1-3):65–83, July 1998.

[44] J. A. Steitz. Polypeptide chain initiation: Nucleotide sequences of the three ribosomal binding sites in bacteriophage r17 rna. *Nature*, 224:957–964, December 1969.

[45] Changwon Kang and Charles R. Cantor. Structure of ribosome-bound messenger RNA as revealed by enzymatic accessibility studies. *Journal of Molecular Biology*, 181(2):241–251, January 1985.

[46] Carsten Petersen. The functional stability of the lacZ transcript is sensitive towards sequence alterations immediately downstream of the ribosome binding site. *MGG Molecular & General Genetics*, 209(1):179–187, 1987.

[47] Frederique Braun, Jacques Le Derout, and Philippe Regnier. Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of escherichia coli. *EMBO Journal*, 17(16):4790–4797, 1998.

[48] Atilio Deana and Joel G. Belasco. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes & Development*, 19(21):2526–2533, November 2005.

[49] Margit Pedersen, Søren Nissen, Namiko Mitarai, Sine Lo Svenningsen, Kim Sneppen, and Steen Pedersen. The functional Half-Life of an mRNA depends on the ribosome spacing in an early coding region. *Journal of Molecular Biology*, 407(1):35–44, March 2011.

[50] Angelo Valleriani, Zoya Ignatova, Apoorva Nagar, and Reinhard Lipowsky. Turnover of messenger RNA: Polysome statistics beyond the steady state. *EPL*, 89(5):58003, 2010.

[51] Namiko Mitarai, Kim Sneppen, and Steen Pedersen. Ribosome collisions and translation efficiency: Optimization by codon usage and mRNA destabilization. *Journal of Molecular Biology*, 382(1):236–245, September 2008.

[52] Daniel E. Morse and Charles Yanofsky. Polarity and the degradation of mRNA. *Nature*, 224(5217):329–331, October 1969.

[53] Joel G. Belasco and Christopher F. Higgins. Mechanisms of mRNA decay in bacteria: a perspective. *Gene*, 72(1-2):15–23, December 1988.

[54] Gunther S. Stent. Genetic transcription. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 164(995):181–197, March 1966.

[55] Yoav Arava, Yulei Wang, John D. Storey, Chih Long Liu, Patrick O. Brown, and Daniel Herschlag. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 100(7):3889 –3894, April 2003.

[56] Frederick C. Neidhardt, editor. *Escherichia coli and salmonella typhimurium*. American Society for Microbiology, Washington, DC, 1987.

[57] Michael A. Sørensen, C. G. Kadsurland, and Steen Pedersen. Codon usage determines translation rate in escherichia coli. *Journal of Molecular Biology*, 207(2):365–377, May 1989.

[58] Frederick R. Blattner, Guy Plunkett, Craig A. Bloch, Nicole T. Perna, Valerie Badsurland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson Wayne Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, and Ying Shao. The complete genome sequence of escherichia coli k-12. *Science*, 277(5331):1453–1462, September 1997.

[59] Liang Feng and Deng-Ke Niu. Relationship between mRNA stability and length: An old question with a new twist. *Biochemical Genetics*, 45(1-2):131–137, 2007.

[60] Jonathan A Bernstein, Arkady B Khodursky, Pei-Hsun Lin, Sue Lin-Chao, and Stanley N Cohen. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences*, 99(15):9697–9702, July 2002.

[61] Anders F. Andersson, Magnus Lundgren, Stefan Eriksson, Magnus Rosenlund, Rolf Bernander, and Peter Nilsson. Global analysis of mRNA stability in the archaeon Sulfolobus. *Genome Biology*, 7(10):R99, 2006.

[62] S.C. Makrides. Strategies for achieving high-level expression of genes in escherichia coli. *Microbiological Reviews*, 60(3):512–538, 1996.

[63] Lars Dölken, Zsolt Ruzsics, Bernd Rädle, Caroline C. Friedel, Ralf Zimmer, Jörg Mages, Reinhard Hoffmann, Paul Dickinson, Thorsten Forster, Peter Ghazal, and Ulrich H. Koszinowski. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, 14(9):1959 –1972, 2008.

[64] Ran Elkon, Eitan Zlotorynski, Karen I Zeller, and Reuven Agami. Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics*, 11:259, April 2010.

[65] José E. Pérez-Ortín, Paula M. Alepuz, and Joaquín Moreno. Genomics and gene transcription kinetics in yeast. *Trends in Genetics*, 23(5):250–257, May 2007.

[66] Christian Miller, Björn Schwalb, Kerstin Maier, Daniel Schulz, Sebastian Dumcke, Benedikt Zacher, Andreas Mayer, Jasmin Sydow, Lisa Marcinowski, Lars Dölken, Dietmar E. Martin, Achim Tresch, and Patrick Cramer. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular Systems Biology*, 7, January 2011.

[67] Caroline C. Friedel, Lars Dölken, Zsolt Ruzsics, Ulrich H. Koszinowski, and Ralf Zimmer. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Research*, 37(17):e115–e115, September 2009.

[68] Ido Golding and Edward C Cox. RNA Dynamics in Live Escherichia Coli Cells. *Proceedings of the National Academy of Sciences of the United States of America*, 101(31):11310–11315, August 2004.

[69] E Bertrand, P Chartrand, M Schaefer, S M Shenoy, R H Singer, and R M Long. Localization of ASH1 mRNA particles in living yeast. *Molecular Cell*, 2(4):437–445, October 1998.

[70] Thuc T. Le, Sébastien Harlepp, Călin C. Guet, Kimberly Dittmar, Thierry Emonet, Tao Pan, and Philippe Cluzel. Real-time RNA profiling within a single bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, 102(26):9160 –9164, June 2005.

[71] Ya'ara Ben-Ari, Yehuda Brody, Noa Kinor, Amir Mor, Toshiro Tsukamoto, David L. Spector, Robert H. Singer, and Yaron Shav-Tal. The life of an mRNA in space and time. *Journal of Cell Science*, 123(10):1761 –1774, May 2010.

[72] Ido Golding, Johan Paulsson, Scott M. Zawilski, and Edward C. Cox. Real-Time kinetics of gene activity in individual bacteria. *Cell*, 123(6):1025–1036, December 2005.

[73] David R. Cox. *Renewal theory*. Methuen's monographs on applied probability and statistics. Methuen, 1967.

[74] Carlus Deneke, Sophia Rudorf, and Angelo Valleriani. Transient phenomena in gene expression after induction of transcription. *PLoS ONE*, 7(4):e35044, April 2012.

[75] Nicholas T. Ingolia, Sina Ghaemmaghami, John R. S. Newman, and Jonathan S. Weissman. Genome-Wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924):218–223, April 2009.

[76] Nicholas T. Ingolia, Liana F. Lareau, and Jonathan S. Weissman. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4):789–802, November 2011.

[77] N Jacques and M Dreyfus. Translation initiation in Escherichia-coli - old and new questions. *Molecular Microbiology*, 4(7):1063–1067, July 1990.

[78] S.-T. Liang, Y.-C. Xu, P. Dennis, and H. Bremer. mRNA composition and control of bacterial gene expression. *Journal of Bacteriology*, 182(11):3037 –3044, June 2000.

[79] Peter J. Lewis, Shail D. Thaker, and Jeffrey Errington. Compartmentalization of transcription and translation in bacillus subtilis. *EMBO J*, 19(4):710–718, February 2000.

[80] Keren Nevo-Dinur, Anat Nussbaum-Shochat, Sigal Ben-Yehuda, and Orna Amster-Choder. Translation-Independent localization of mRNA in e. coli. *Science*, 331(6020):1081 –1084, February 2011.

[81] I Iost and M Dreyfus. The stability of escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. *The EMBO Journal*, 14(13):3252–3261, July 1995.

[82] Florian Brandt, Stephanie A. Etchells, Julio O. Ortiz, Adrian H. Elcock, F. Ulrich Hartl, and Wolfgang Baumeister. The native 3D organization of bacterial polysomes. *Cell*, 136(2):261–271, January 2009.

[83] Naama Brenner and Yair Shokef. Nonequilibrium statistical mechanics of dividing cell populations. *Physical Review Letters*, 99(13):138102, 2007.

[84] Naama Brenner, Keren Farkash, and Erez Braun. Dynamics of protein distributions in cell populations. *Physical Biology*, 3(3):172–182, September 2006.

[85] Matthew Scott, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007):1099–1102, November 2010.

[86] S Cooper. The escherichia coli cell cycle. *Research in Microbiology*, 141(1):17–29, January 1990.

[87] M. Rading, T. Engel, R. Lipowsky, and A. Valleriani. Stationary size distributions of growing cells with binary and multiple cell division. *Journal of Statistical Physics*, 145(1):1–22, October 2011.

[88] J. M. Mitchison. Single cell studies of the cell cycle and some models. *Theoretical Biology and Medical Modelling*, 2(1):4, February 2005.

[89] Stephen Cooper. Distinguishing between linear and exponential cell growth during the division cycle: Single-cell studies, cell-culture studies, and the object of cell-cycle research. *Theoretical Biology and Medical Modelling*, 3(1):10, February 2006.

[90] William Feller. *An introduction to probability theory and its applications*. Wiley, 1971.

[91] Mingoo Kim, Sung Bum Cho, and Ju Han Kim. Mixture-Model based estimation of gene expression variance from public database improves identification of differentially expressed genes in small sized microarray data. *Bioinformatics*, 26(4):486–492, February 2010.

[92] Daehee Hwang, William A Schmitt, George Stephanopoulos, and Gregory Stephanopoulos. Determination of minimum sample size and discriminatory expression patterns in microarray data. *Bioinformatics*, 18(9):1184–1193, September 2002.

[93] Mei-Ling Ting Lee, Frank C Kuo, G. A Whitmore, and Jeffrey Sklar. Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. *Proceedings of the National Academy of Sciences*, 97(18):9834–9839, August 2000.

[94] Paul Pavlidis, Qinghong Li, and William Stafford Noble. The effect of replication on gene expression microarray experiments. *Bioinformatics*, 19(13):1620–1627, September 2003.

[95] Tatjana Trcek, Daniel R Larson, Alberto Moldón, Charles C Query, and Robert H Singer. Single-Molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell*, 147(7):1484–1497, December 2011.

[96] O. L. Miller, Barbara A. Hamkalo, and C. A. Thomas. Visualization of bacterial genes in action. *Science*, 169(3943):392–395, July 1970.

A. Appendix to chapter 2: Mathematical background

In this part, we elaborate some more findings related to chapter 2.

A.1. Properties of Poisson process

First we state some properties of a Poisson process. For any $0 < t < T$ and $0 \leq k, \leq n$, it holds

$$\Pr\{X(t) = k \mid X(T) = n\} = \frac{n!}{k!(n-k)!}(\frac{t}{T})^k(1-\frac{t}{T})^{n-k}\,. \tag{A.1}$$

The probability of no event occurring in the interval $t_{i+1} - t_i$ reads

$$\Pr\{X(t_{i+1}) \,-\, X(t_i) \,=\, 0\} = \exp(-\omega_{\mathrm{tc}}\,(t_{i+1} - t_i)) \tag{A.2}$$

Furthermore, consider $S \equiv t_{i+1} - t_i$ as the random sojourn time between subsequent events. Its probability density reads

$$f(S = s) = \omega_{\mathrm{tc}}\,\exp(-\omega_{\mathrm{tc}}\,s) \tag{A.3}$$

Thus, the sojourn time is exponentially distributed.

A.1.1. Relation of uniform distribution and Poisson process

In the following, we demonstrate the relation of uniform distribution and Poisson process. We used the result of this computation in the derivation of our main results from chapter 2, in particular, the distribution of the mRNA number before and after stop of transcripts. The following derivation proceeds analogous to [30].
Consider a Poisson process where until time t exactly n events are to happen. Each event i occurs at a random time point W_i in the (infinitesimal) interval $w_i < W_i < w_i + \Delta w_i$. Hence, no event occurs in $(w_i + \Delta w_i, w_{i+1}]$. Thus, the events occur in disjoint intervals. One can distinguish between intervals with no event

$$Pr\{X(w_i) - X(w_{i-1} + \Delta w_{i-1}) = 0\} = e^{-\lambda(w_i - w_{i-1} - \Delta w_{i-1})}\,, \tag{A.4}$$

which follows directly from the homogeneity of the process and obvious alterations for $i = 1$ and $i = n$.
And intervals where exactly one event occurs

$$Pr\{X(w_i + \Delta w_i) - X(w_i) = 1\} = \lambda \Delta w_i\,, \tag{A.5}$$

i.e. they are proportional to the rate of the Poisson process.
Putting the pieces together, the probability density reads

$$\begin{aligned}
&f_{W_1,\dots,W_n|X(t)=n}(w_1,\dots,w_n)\Delta w_1 \dots \Delta w_n \\
&= Pr\{w_1 < W_1 < w_1 + \Delta w_1, \dots, w_n < W_n < w_n + \Delta w_n \mid X(t) = n\} \\
&= Pr\{w_1 < W_1 < w_1 + \Delta w_1, \dots, w_n < W_n < w_n + \Delta w_n, X(t) = n\} / Pr\{X(t) = n\} \\
&= \prod_{i=1}^{n} \Big(Pr\{X(w_i + \Delta w_i) - X(w_i) = 1\} \cdot Pr\{X(w_i) - X(w_{i-1} + \Delta w_{i-1}) = 0\} \Big) / Pr\{X(t) = n\} \\
&= \frac{e^{-\lambda t} \lambda^n e^{\lambda(\Delta w_1 + \dots + \Delta w_n)}}{e^{-\lambda t}(\lambda t)^n / n!} \Delta w_1 \dots \Delta w_n
\end{aligned} \tag{A.6}$$

Thus, dividing both sides by $\Delta w_1 \dots \Delta w_n$ and taking the limit $\Delta w_1 = \dots = \Delta w_n = 0$ finally leads to

$$f_{W_1,\dots,W_n|X(t)=n}(w_1,\dots,w_n) = n!\, t^{-n} \tag{A.7}$$

This is exactly the probability distribution of n uniformly distributed events in an interval $(0, t]$ where the factor of $n!$ arises from the number of possibilities of rearranging the n individual events.
Thus we can conclude that in the conditional Poisson process the events are uniformly distributed.

A.2. Derivation of the age and residual lifetime distribution

In chapter 2 we have given the results of the distribution of the age A and residual lifetime R of an mRNA at arbitrary time before and after the interruption of transcription. In the following, we will give a detailed description of the derivation of these results.

A.2.1. Age distribution

A randomly selected mRNA will also be of a random age. Hence, a large ensemble of mRNAs at a given (global) time t will reveal an age distribution (see also Fig. 2.1). In this section, we will derive a general analytical result for the age probability density ϕ_A of an mRNA.

In the following, we will not distinguish between scenarios before and after the interruption of transcription. Instead we derive the more general formulation directly and certain limit cases can be easily obtained from the final result. We are interested in the age of a randomly chosen mRNA at any time point t. Therefore, in the following, we consider a single mRNA that has been created according to a Poisson process in the interval $[0, t_s)$ and has a random lifetime U distributed according to the density ϕ_U. Using the same notation as before, the given mRNA will be present at an observation time $t_s + \Delta t$ only if the variable $Z = O + U$ satisfies $Z \geq t_s + \Delta t$. Let A be the random variable that gives the age of a randomly chosen mRNA. Then, the age distribution of the mRNA is given by the distribution of $A = t_s + \Delta t - O$ under the condition $Z \geq t_s + \Delta t$ (see Fig. 2.1). In order to compute this quantity we shall first realize that

$$\Pr\{A \leq a \mid Z \geq t_s + \Delta t\} \;=\; 1 - \Pr\{O < t_s + \Delta t - a \mid Z \geq t_s + \Delta t\}\,, \tag{A.8}$$

and thus compute the probability density for O conditional that $Z \geq t_s + \Delta t$. To compute this quantity, recall that, in this case, we condition that the number of transcribed mRNAs until time t_s is just one. Therefore, the random variable O is uniformly distributed in $[0, t_s)$. Since the transcription events are independent from another, we can thus compute the age distribution of a sample of mRNAs. We obtain

$$\begin{aligned}
\Pr\{O < x \mid Z \geq t_s + \Delta t\} \;&=\; \frac{\Pr\{O < x, Z \geq t_s + \Delta t\}}{\Pr\{Z \geq t_s + \Delta t\}} \\
&=\; \int_{t_s+\Delta t-x}^{\infty} \mathrm{d}u\, \frac{\Pr\{O < x, Z \geq t_s + \Delta t \mid U = u\}\, \phi_U(u)}{\Pr\{Z \geq t_s + \Delta t\}} \\
&=\; \int_{t_s+\Delta t-x}^{\infty} \mathrm{d}u\, \frac{\Pr\{t_s + \Delta t - u \leq O < x\}\, \phi_U(u)}{\Pr\{Z \geq t_s + \Delta t\}}\,,
\end{aligned} \tag{A.9}$$

where $0 \leq x < t_s$ because we implicitly conditioned that the origination time point is before time t_s. We shall now use the fact that the random variable O is uniformly distributed in $[0, t_s)$ (because we have conditioned that there is one mRNA alive at time

$t = t_s + \Delta t$) and thus that $\Pr\{Z \geq t_s + \Delta t\}$ is given by (2.13). This leads to

$$\begin{aligned} &\int_{t_s+\Delta t-x}^{\infty} \mathrm{d}u \, \frac{\Pr\{t_s + \Delta t - u \leq O < x\}\phi_U(u)}{\Pr\{Z \geq t_s + \Delta t\}} \\ &= \left[\int_{\Delta t}^{t_s+\Delta t} \mathrm{d}y \, (1 - \Phi_U(y))\right]^{-1} \int_{t_s+\Delta t-x}^{\infty} \mathrm{d}u \, [x - \max(0, t_s + \Delta t - u)] \, \phi_U(u) \,, \end{aligned} \tag{A.10}$$

from which we can compute the distribution of the age $A = t_s + \Delta t - O$ under the condition $Z \geq t_s + \Delta t$. This distribution is given by (A.8) and reads

$$\begin{aligned} \Pr\{A \leq a \mid Z \geq t_s + \Delta t\} &= 1 - \Pr\{O \leq t_s + \Delta t - a \mid Z \geq t_s + \Delta t\} \\ &= 1 - \left[\int_{\Delta t}^{t_s+\Delta t} \mathrm{d}y \, (1 - \Phi_U(y))\right]^{-1} \int_{a}^{\infty} \mathrm{d}u \, [\min(u, t_s + \Delta t) - a] \, \phi_U(u) \,, \end{aligned} \tag{A.11}$$

which upon differentiation with respect to a finally leads to the probability density function

$$\phi_A(a \mid t_s + \Delta t) = \left[\int_{\Delta t}^{t_s+\Delta t} \mathrm{d}u \, (1 - \Phi_U(u))\right]^{-1} (1 - \Phi_U(a)) \,, \tag{A.12}$$

for $\Delta t \leq a < t_s + \Delta t$ and zero otherwise.
Note that Eq. (A.12) is the marginal probability density of the age random variable A. We can consider different scenarios, before and after the stop of transcription, by choosing t and Δt accordingly.

A.2.2. Residual lifetime distribution

Similar to its age, a randomly selected mRNA will also have a random residual lifetime. Hence, a large ensemble of mRNAs at a given (global) time t will reveal a residual lifetime distribution (see also Fig. 2.1). In this section, we will derive a general analytical result for the residual lifetime density ϕ_R.
The residual (or excess) lifetime R of an mRNA is a statistical quantity complementary to the age of the mRNA. The derivation of its distribution proceeds in a similar fashion as for the age distribution. Once again, we consider a single mRNA that is generated in the interval $[0, t_s)$ and at the time $t_s + \Delta t$ a measurement is performed. The residual lifetime is given by $R = Z - (t_s + \Delta t)$, *i.e.* by the difference of the time of degradation and the observation time point (see Fig. 2.1). Hence, the probability distribution of the residual lifetime under the condition that the mRNA is still alive at $Z \geq t_s + \Delta t$ is given by

$$\begin{aligned} \Pr\{R \leq r \mid Z \geq t_s + \Delta t\} &= \Pr\{Z \leq t_s + \Delta t + r \mid Z \geq t_s + \Delta t\} \\ &= \frac{\Pr\{Z \geq t_s + \Delta t\} - \Pr\{Z \geq t_s + \Delta t + r\}}{\Pr\{Z \geq t_s + \Delta t\}} \,, \end{aligned} \tag{A.13}$$

where we have used the relation for R and the definition of the conditional probability. We can proceed by using (2.13) and upon derivation by r we obtain the residual lifetime

probability density

$$\phi_R(r \mid t_s + \Delta t) = \left[\int_{\Delta t}^{t_s+\Delta t} \mathrm{d}u\,(1 - \Phi_U(u))\right]^{-1} (\Phi_U(t_s + \Delta t + r) - \Phi_U(\Delta t + r)) \ . \quad \text{(A.14)}$$

It holds for all $r \geq 0$. Similarly as for the age distribution, also for the residual lifetime distributions several limit cases of interest exist and will be further discussed in chapters 3 and 4.

Hence, both the age distribution as well as the residual lifetime distribution depend on the form of the lifetime probability density ϕ_U and on the time after start of transcription t_s as well as the time interval since the interruption of transcription Δt.

In this section we have performed the central computations used for the remaining parts of the thesis. We found an analytical expression for the number and distribution of mRNAs under different experimental conditions as well as for the age and residual lifetime distributions.

A.3. mRNA fluctuations in cell populations

Each measurement via a microarray yields the mRNA level of each gene. However, as we pointed out, the process of mRNA turnover is inherently stochastic. Therefore, a single measurement only reveals one realization of possible values and to obtain the distribution and mean value, repeated measurements under the same conditions are necessary.

From a theoretical point of view, each cell in a population has a random mRNA signal (*i.e.* mRNA number) M. If we consider unsynchronized cells, these random numbers are furthermore independent. The fluctuation in the cell population would thus be defined by the sum of all random numbers pertaining to individual cells. Under the assumptions that the mRNA fluctuations are Poissonian in each cell, one can compute the sum of independent, Poisson distributed random variables. A standard calculation reveals that the sum gives a random number which is also Poisson distribution and the parameter of the distribution is given by the sum of the single cell parameters [30]. In this scenario, the distribution for each random variable in a single cell can have different distribution parameters μ. This would reflect different phases in the cell cycle that become manifest in a variation of the transcription rate ω_{tc} and mRNA average lifetime $\langle U \rangle$. Since experiments typically assess the mRNA expression in a large number of cells, *i.e.* more than 10^9, the resulting Poisson distribution can be approximated with a normal distribution (according to the central limit theory in probability theory [90]). Hence, each measurement would reveal only one possible value according to a normal distribution with mean $\Omega_{\mathrm{tc}}\langle U \rangle$ and a variance equal to the mean (see figure A.1) and need not necessarily reproduce the mean expression level itself.

Indeed, in a systematic analysis of many data sets, either a normally distributed variation was found or a multi-normal distribution in cases where cells in different expression states are present [91]. It has been pointed out in the bioinformatics community that a minimum of 5 replicate experiments is necessary to be able to estimate the parameters of the distribution with maximum-likelihood methods [92, 93, 94]. Thus, the lower number of

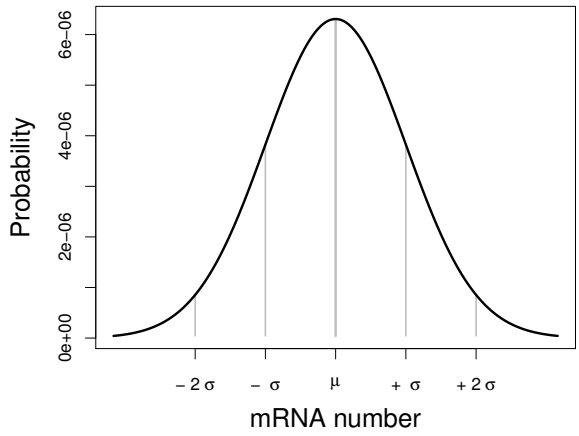

Figure A.1.: **Gaussian distribution of large population of cells**. If one considers a large number of unsynchronized cells, where the fluctuations in a single cell are described by a Poisson distribution, the distribution of mRNAs is given by a normal distribution.

replicas performed in many experiments might explain part of the weak reproducibility, although this analysis only concerns the systematic, inherent variation and not possible additional experimental uncertainties. Moreover, if single cell fluctuations cannot be described by a Poisson distribution as for example in a Cox process (see section D) the sum of the random mRNA number in each cell is not easy to calculate (in fact, as we have pointed out above there is no analytical solution for the fluctuations of the mRNA number available). However, also in this case the fluctuations in single cells would add up to describe the mRNA fluctuation in cell populations and also here the central limit theorem would lead to a normal distribution (albeit with larger variance).

A.4. Master equation and phenomenological descriptions

Details of calculation, start and stop of transcription, mean values
Similar to the findings for the fluctuations of the mRNA number for constant rates, one directly compute the evolution of the mean mRNA number after start and stop of transcription.
For the first case, the Master equation reads

$$\frac{dN_r}{dt} = \omega_{\mathrm{tc}} - N_r(t)\omega_{\mathrm{r}}\,, \tag{A.15}$$

with the solution described by $N_r(t) = \frac{\omega_{\mathrm{tc}}}{\omega_{\mathrm{r}}}(1 - \exp(-\omega_{\mathrm{r}}t)$. If transcription is stopped for a populations that was previously in a steady state, the Master equation reads

$$\frac{dN_r}{dt} = -N_r(t)\omega_{\mathrm{r}}\,. \tag{A.16}$$

The decay of the relative number is given by $\Lambda(t) = \exp(-\omega_{\mathrm{r}}t)$.
It would be interesting to see if a description with a Master equation is also possible for the more general theory with arbitrary lifetimes discussed in this thesis. However, as we pointed out in section 2.2 the degradation rate depends on the age of the mRNA, and a possible formulation would have the following form

$$\frac{dN_r}{dt} = \omega_{\mathrm{tc}} - N_r(t)\int_0^t da\,\omega_{\mathrm{r}}(a)\phi_A(a \mid t)\,. \tag{A.17}$$

Clearly, a solution to this equation seems not trivial and the full stochastic formulation we have pursuit advantageous.

Phenomenological model

The function $\Phi_U(t)$ denotes the probability that an mRNA has been degraded until a time Δt after its transcription. Hence, one can define the survival probability

$$S(\Delta t) \equiv 1 - \Phi_U(\Delta t) \tag{A.18}$$

that an mRNA is still alive after a time Δt. However, we are not interested in the survival probability of a single mRNA with precisely known origin, but the fraction of remaining

mRNAs originating from random transcription points in the interval $[\Delta t, \infty)$ before the interruption of transcription. We can define the average survival probability as

$$\hat{S}(\Delta t) \equiv \int_{\Delta t}^{\infty} d\tau \, (1 - \Phi_U(\tau)) \, . \tag{A.19}$$

It describes the fraction of remaining mRNAs at at time Δt after the interruption of transcription. A comparison to Eq. (3.2) reveals that Eq. (A.18) takes the same form as $\Lambda(\Delta t)$. One can extend this description by accounting for non-homogeneous description with a transcription density term $\rho(t)$, *i.e.*

$$\hat{S}(\Delta t) \equiv \int_{\Delta t}^{\infty} d\tau \, (1 - \Phi_U(\tau)) \, \rho(\Delta t - \tau) \, . \tag{A.20}$$

The special cases, $\rho(\Delta t - \tau) = \delta(\Delta t - \tau)$ and $\rho(\Delta t - \tau) = 1$ lead to Eqs. (A.18) and (A.19), respectively. It would be interesting however to study more variants of $\rho(t)$, e.g. random spikes of transcription activity or modulated transcription rates.

A.5. Non-linear Markov chains

The general concept of a Markov chain to describe the degradation process as a multi-step process can also be expanded to non-linear chains with several absorbing states. Here, degradation from an initially intact mRNA (black) can proceed via miRNA mediated decay (lower branch) or another, independent pathway (upper branch). In the miRNA mediated pathway (red), two enzymes are known to trigger deadenylation and degradation and the sub-pathways can be differentially analyzed. Finally, in this picture degradation can proceed in at least three different ways resulting in three different absorbing states (blue). This model is motivated by a study of biochemical degradation mechanisms in mice [17].

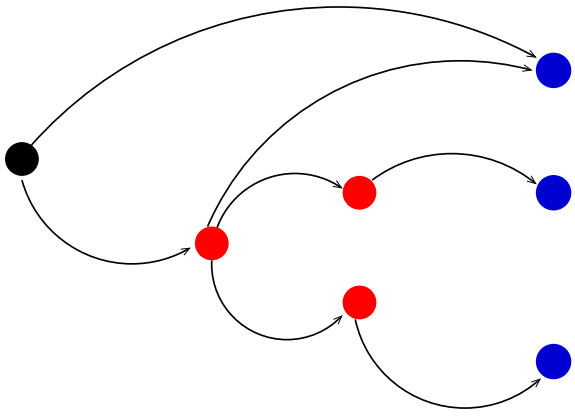

Figure A.2.: **Example of non-linear Markov chain**.The general concept of a Markov chain to describe the degradation process as a multi-step process can also be expanded to non-linear chains with several absorbing states.

A.6. More details of the computer simulations

In chapter 2 we have outlined possible strategies to model mRNA turnover via stochastic simulations. Here, we give algorithms for more detailed simulations which we used in this thesis.

Basic mRNA and protein turnover algorithm The mRNA turnover simulation proceeds as before, however during the lifetime of each mRNA, proteins are produced with a constant rate. For each mRNA, protein events are a random time interval da apart. Counting the total number of proteins as well as the random time points of their generation and degradation allows retrieving the number of live proteins at any time of the simulation. A large number of realizations gives furthermore the distribution of proteins.

set $t = 0$, $X = 0$
while $t < T$
    generate dt
    set $t = t + dt$
    set $X = X + 1, \mathrm{Start}(X) = t$
    generate U, **set** $\mathrm{End}(X) = t + U$
    while $a < U$
        generate da
        set $a = a + da$
        set $p = p + 1, \mathrm{Start}(p) = a$
        generate U_p, **set** $\mathrm{End}(p) = a + U_p$

Thinning algorithm for simulating the Non-homogeneous Poisson Process (NHPP) The intensity $\lambda(t)$ is a randomly varying quantity. However, for each realization we can determine $\lambda(t)$ beforehand. Thus, during the simulation of the Poisson process of generation of mRNA it is a time-dependent but deterministic quantity.

set $t = 0$, $\lambda = \omega_{\mathrm{tc}}$
while $t < T$
    generate κ
    generate r
    if $r > 0.5$
        set $\lambda = \lambda + \chi$
    else
        set $\lambda(t + \kappa) = \lambda(t) - \chi$
    if $\lambda > \omega_{\mathrm{tc}} + \epsilon$
        set $\lambda = \omega_{\mathrm{tc}} + \epsilon$
    if $\lambda < \omega_{\mathrm{tc}} - \epsilon$
        set $\lambda = \omega_{\mathrm{tc}} - \epsilon$
    set $t = t + \kappa$

Hence, we can obtain the random change of the intensity $\lambda(t)$. This will be the input for the non-homogeneous Poisson process

set $t = 0$, $X = 0$
while $t < T$
    generate dt
    set $t = t + dt$
    generate r
    if $r \leq \lambda(t)/\lambda$ **then**
        set $X = X + 1, \mathrm{Start}(X) = t$
        generate U
        set $\mathrm{End}(X) = t + U$

Simulation of cell division Additionally to the mRNA turnover, at random intervals t_{cc} (drawn from a log-normal distribution) the cell undergoes cell division and the mRNAs stay in the mother cell with a probability 0.5.

set $t = 0$, $X = 0$
while $t < T$
    generate dt
    set $t = t + dt$
    set $X = X + 1, \mathrm{Start}(X) = t$
    generate U, **set** $\mathrm{End}(X) = t + U$

    determine time of division t_{cc}
        for each X Eliminate X with prob. 0.5

TASEP algorithm for protein synthesis In a *TASEP* simulation, ribosomes bind to an mRNA with an initiation rate ω_{on} and walk on a homogeneous chain with an elongation rate v. For each ribosome i on the chain as well as the pool of ribosomes waiting to initiate translation, one can determine the time of the next event by a random number generator, *i.e.* the waiting times of $\delta t(i)$. An event consists of an attempt of a ribosome to progress by one codon (or - in the case of initiation - to attach to the first site of an mRNA). The attempt is, however, only successful if the site is empty, *i.e.* not occupied by another ribosome. Finally, after reaching the last codon, the ribosome detaches since it has completed the synthesis of a polypeptide. Furthermore, the ribosomes are modeled as extended objects of size l since they typically occupy more than one codon (in fact, typically 10 codons [44, 45]).

set $i = 0$, $t = 0$
while $t < T$
    determine $\forall i$ $(\delta t(i))$
    find $k = min_{\forall i}(\delta t(i))$
    if $position(k-1) - position(k) < l$
        set $position(k) = position(k) + 1$
        set $t = t + \delta t(k)$
    update i

In the algorithm, the run length T is given by the lifetime of an mRNA. Since any measurement would reveal mRNAs of different age, one can sample T from the age distribution ϕ_A. We use this algorithm to study polysome and ribosomal profiles (see section 5.1). In section 3.3 we use a variant were the run length is not determined externally but rather is an intrinsic quantity (see below).

Degrading TASEP simulations This algorithm proceeds similar to the regular TASEP except that it has no a priori determined run length T but rather the degradation time is a stochastic quantity inherent of the simulation. During the simulation, a base degradation rate ω_{base} is multiplied by the available cleavage area. The latter is the sum over all ribosome free mRNA stretches of a certain minimum length δl which accounts for the size of ribosomes and enzyme and their steric repulsion). In the TASEP simulation the position of the ribosomes changes randomly and hence also the degradation rate. To include degradation, at each instant of the simulation a probability criterion which relies on the time-dependent degradation rate determines if the mRNA is degraded and the simulation is stopped. We used this algorithm to evaluate the role of fluctuations for our model of ribosome shielding (see section 3.3).

B. Appendix to chapter 3: Decay models

B.1. Additional general implications

B.1.1. Decay from non-stationary samples

If in a decay experiment the mRNAs were not in a steady state prior to the interruption of transcription, we can define the relative remaining mRNA number analogously:

$$\tilde{\Lambda}(\Delta t) \equiv \frac{\omega_{\mathrm{tc}} \mathcal{H}_{\Delta t}^{t_s+\Delta t}}{\omega_{\mathrm{tc}} \mathcal{H}_0^{t_s}} \tag{B.1}$$

In the case that the function $\mathcal{H}$ is determined by an exponential lifetime distribution, the decay pattern is just the same for any value of t_s. For all other lifetime distributions, this is not exactly the case - however the difference is small and negligible in most cases. For lifetime distributions given by a sum of exponentials (such as in a Gamma-model or Markov chain model) the difference scales with $exp(-\lambda t_s)$. Hence, if there is reason to assume that the average mRNA amount was not in a steady state prior to the interruption of transcription, one can analyse the decay pattern with Eq. (B.1). Nevertheless, before starting the decay experiments, cells are grown for several order of magnitudes longer than the approximate time to steady state.

B.1.2. Bona-fide criterion in practice

In chapter 3 we found two general bona-fide criteria that hold for any experimental decay patterns regardless of the organism and degradation mechanism. These criteria followed directly from the analysis of the first and second derivative of the probability density function ϕ_U. In practice, the experimental decay patterns are given by the mRNA amount at discrete time points. Therefore, to assess if a given data set is *bona-fide*, we also have to reformulate the criteria into a discrete language. Hence, the following inequalities should hold for any time points t_1, t_2 and t_3:

$$N_r(t_2) < N_r(t_1) \;\; \forall\, t_2 > t_1 \tag{B.2}$$

and

$$\frac{N_r(t_3) - N_r(t_2)}{t_3 - t_2} > \frac{N_r(t_2) - N_r(t_1)}{t_2 - t_1} \;\; \forall\, t_3 > t_2 > t_1 \tag{B.3}$$

Given the limited temporal resolution and the intrinsic and experimental noise in mRNA expression data, for the practical analysis performed in chapter 3, we allowed a certain

tolerance to this criteria. Therefore, we term a data set nearly bona-fide if the slope between any two data points does not increase by more than 10% and the curvature between three subsequent data points is convex within an error of up to 25 %.

B.1.3. Cell cycle dependent degradation rate

Eukaryotic cells undergo different phases in the cell cycle. In principle the mode of transcription and degradation can vary in different phases. In an experiment on yeast, one could differentially measure the decay pattern of cells in S and M phase [95]. From the three native genes analyzed there, two had a changed decay pattern while the third was approximately unchanged. Moreover, the decay patterns during one cell cycle were not always exponential. While broader conclusions might be premature at the moment, it would be interesting to see a genome wide analysis. One might learn that our theory is more accurate in describing the decay for different phases of the cell cycle and that the global decay patterns can be described by the sum over cells in different stages of the cell cycle.

B.2. Appendix to multi-step degradation

In this section, we will give some more details of the particular lifetime density functions we obtained from the Markov chain model in section 3.2.

Details of Markov chain model in analysis of yeast data

We have shown two example decay patterns pertaining to different decay pattern categories in figure 3.4. The fitting functions have been obtained by using the Markov chain model (see figure 2.2) with $n = 5$ states and two independent rates, as follows. For the red curve in figure 3.4 we have set $\mu_1 \equiv \mu$ and all other $\mu_i \equiv \nu$ as well as the transition rate $\lambda = \nu$ which corresponds to a decreasing degradation rate during the lifetime of an mRNA if $\mu > \nu$. We find the lifetime distribution to be

$$\phi_U(t) = \frac{1}{\mu}\left(e^{-(\nu+\mu)t}\,(\nu+\mu)\,(\mu-\nu) + e^{-\nu t}\,\nu^2\right)\,, \tag{B.4}$$

i.e. independent of the number of states n. For the blue curve in figure 3.4 we have set $\mu_5 \equiv \nu$ and $\mu_i \equiv 0$ otherwise, while the transitions between the states occur with rate λ. This situation is inspired by the multi-step degradation depicted in figure 1.2C and leads to

$$\phi_U(t) \;=\; \nu\left(\frac{\lambda}{\lambda-\nu}\right)^4\left[e^{-\nu t} \;- e^{-\lambda t}\left(\sum_{j=0}^{3}\frac{(\lambda-\nu)^j t^j}{j!}\right)\right]\,. \tag{B.5}$$

All unknown rates for the decay patterns can be obtained via a non-linear regression analysis by minimizing the root mean-square difference r^2 between the measured data and Eq. (3.2). For the analysis of all mRNA decay patterns, we have compared the fitting results of Eqs. (B.4) (B.5) with different starting values as well a fit based on an exponential lifetime distribution. For the two example cases in figure 3.4, the rates read $\mu = 0.187\,\mathrm{min}^{-1}$ and $\nu = 0.019\,\mathrm{min}^{-1}$ (red, *MGS1*), whereas in the latter $\lambda = 0.313\,\mathrm{min}^{-1}$ and $\nu = 0.037\,\mathrm{min}^{-1}$ (blue, *Pho3*).

Probability density for arbitrary number of states

In section 3.2 we have used a Markov chain model to obtain the mRNA lifetime probability density $\phi_U(t)$. We can calculate the probability densities for arbitrary number of states n and rates λ_i and μ_i. However, in absence of the knowledge of the rates, we have restricted the analysis to 3 independent parameters: λ is the rate between transient states and μ as well as ν denote rates to the absorbing state.
In a first variant we have set $\mu_i \equiv \mu \;\forall\; i < n$ and $\mu_n = \nu$. In this case the lifetime probability density reads

$$\phi_U(t) = \mu\, e^{-\mu\, t}\,\frac{\Gamma_{n-1}(\lambda\, t)}{\Gamma(n-1)} \;+\; \nu\, e^{-\nu\, t}\left(\frac{\lambda}{\lambda+\mu-\nu}\right)^{n-1}\left(1 - \frac{\Gamma_{n-1}((\lambda+\mu-\nu)\, t)}{\Gamma(n-1)}\right) \tag{B.6}$$

In some cases, the fitting result was improved if degradation was only possible from last state n, *i.e.* $\mu_i \equiv 0 \;\forall\; i < n$ and $\mu_n = \nu$. Then Eq. (B.6) becomes

$$\phi_U(t) = \nu\, e^{-t\,\nu}\left(\frac{\lambda}{\lambda-\nu}\right)^{n-1}\left(1 - \frac{\Gamma_{n-1}((\lambda-\nu)\, t)}{\Gamma(n-1)}\right)\,.$$

The advantage of this equation is that it contains only two independent parameters. Conversely, we also considered a second variant where all but the first absorption rates were the same, *i.e.* $\mu_1 = \mu$ and $\mu_i = \nu \; \forall \; i > 1$. Hence, the absorption probability density reads

$$\phi_U(t) = \frac{e^{-t(\lambda+\mu)}\left((\lambda+\mu)(\mu-\nu) + e^{t(\lambda+\mu-\nu)}\lambda\nu\right)}{\lambda+\mu-\nu} \tag{B.7}$$

In this scenario, the equation is independent of the number of states n (as long as $n > 2$). In some cases, the fitting result could be improved by reducing the number of free parameters to 2. Thus, we have set $\nu \equiv \lambda$.

Comparison to experimental data

As pointed out in the main text, section 3.2, we have inserted the results from Eqs. (B.6) and (B.7) (for $n = 5$) into Eq. (3.2) to obtain the theoretical description of the relative amount of mRNA after interruption of transcription, $\Lambda(\Delta t)$. Via a non-linear regression analysis we could find the optimal parameters that reduce the residual sum of squares between theoretical prediction and experimental data. We compared the different variants given above, *i.e.* Eqs. (B.6) and (B.7), with different initial parameters to an exponentially distributed lifetime density, $\phi_U^{(\mathrm{exp})}$. Depending on which function gave the best fitting result, we could classify the results for mRNA decay in yeast into different categories (see figure 3.3). For the two example cases in figure 3.4, the rates read $\lambda = 0.313\,\mathrm{min}^{-1}$, $\mu = 0$ and $\nu = 0.037\,\mathrm{min}^{-1}$ (variant 1 from Eq. (B.6)). and $\mu = 0.187\,\mathrm{min}^{-1}$ and $\nu = \lambda = 0.019\,\mathrm{min}^{-1}$ (variant 2 from Eq. (B.7)).
Note that one can classify the experimental decay patterns also independently from this concrete model. The two main categories differ characteristically in their slopes. While the mRNA level in the first case decreases relatively slowly in the beginning and more rapid at later times, for the second category the decay is fast in the beginning and levels off after some time. This becomes in particular clear when observing the experimental decay patterns on a semi-log scale. Even more, one can devise simple algorithms to distinguish the patterns pertaining to the different categories.

Probability density for independent transition and absorption rates

As mentioned, we can straightforwardly also compute the probability density if all rates are different One arrives at lengthy expressions that cannot generally be formulated for arbitrary n. E.g. for $n = 3$ and all λ_i and μ_i different, the probability density reads

$$\begin{aligned}\phi_U(t) = &\left(\mu_1 + \frac{-\lambda_1(\lambda_1+\mu_1)\mu_2 + \lambda_1(\lambda_2+\mu_2)\mu_3}{(\lambda_1-\lambda_2+\mu_1-\mu_2)(\lambda_1+\mu_1-\mu_3)}\right)e^{-(\lambda_1+\mu_1)\,t} \\ &+ \frac{\lambda_1(\lambda_2+\mu_2)(\mu_2-\mu_3)}{(\lambda_1-\lambda_2+\mu_1-\mu_2)(\lambda_2+\mu_2-\mu_3)}e^{-(\lambda_2+\mu_2)\,t} \\ &+ \frac{\lambda_1\lambda_2\mu_3}{(\lambda_1+\mu_1-\mu_3)(\lambda_2+\mu_2-\mu_3)}e^{-\mu_3\,t}\end{aligned} \tag{B.8}$$

Note however that all lifetime densities are of the form

$$\phi_U(t) = \sum_{k=1}^{n-1} c_k \, e^{-(\lambda_k+\mu_k)\,t} + c_n \, e^{-\mu_n\,t} \tag{B.9}$$

That implies that when approaching $t \to \infty$ only one term dominates, namely the one with the smallest coefficient in the exponent, *i.e.* $\min\left((\lambda_k + \mu_k),\, \mu_n\right)$. Hence in the limit of large times, the lifetime density decreases like a single exponential function. Therefore, the evolution of the mRNA number always approaches a stationary level in an exponential manner.

Aging of states

In the Markov chain model, the mRNAs are in different states according to their age a (see figure 3.2). After the interruption of transcription, no new mRNAs are synthesized and thus the typical age of an mRNA increases. Consequently, more mRNAs can be found in a more advanced state, *i.e.* more mRNAs have undergone biochemical transformation. As we have already pointed out in section 2.4, the changing population of states can be interpreted as a sign of aging of the mRNA molecules.
In section 4.5, we have introduced how one can compute the population of states of an ensemble of mRNAs at any time t. Moreover, from the population probabilities we can deduce the average state of an ensemble of mRNAs. It is straightforward to extend this

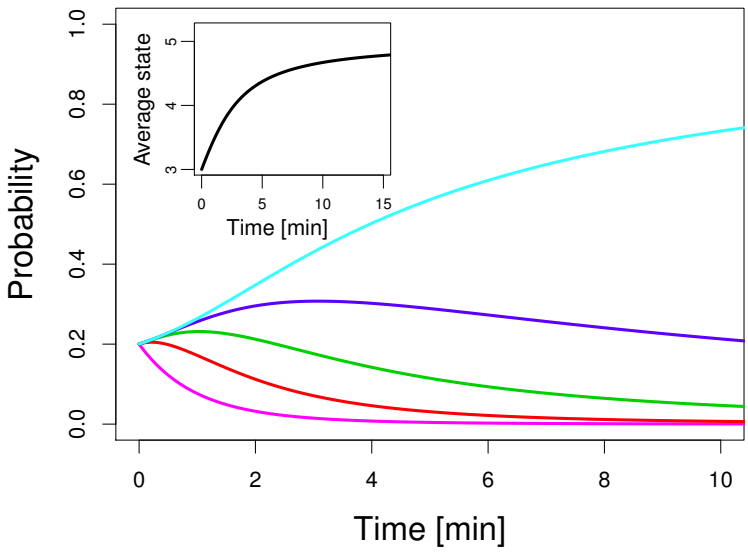

Figure B.1.: **Aging of decaying mRNAs in Markov chain models**. Starting from a steady state mRNA population given by $\phi_U^{(\Gamma)}$ where all of the 5 states are equally populated, the probability to be in each of the 5 states evolves in time after the interruption of transcription. Eventually, mRNAs can be only found in the most advanced state 5. This can also be understood in terms of the evolution of the average state, *i.e.* the first moment of these probabilities (see inset). From an average state $\langle k \rangle = 3$ it evolves to 5. The different states are denoted with 1 (magenta), 2 (red), 3 (green), 4 (blue), 5 (cyan).

analysis to mRNA populations after the interruption of transcription. In figure B.1 we show the time evolution of the probabilities to be in each of the states in a Markov chain as well as the corresponding average state after stop of transcription.

B.3. Model of pathway knock-out

We can use the formalism with a Markov chain to study the role of different degradation pathways when analyzing a detailed experiment of the biochemistry in *drosophila melanogaster* [17]. There, the decay of *F-luc-Nerfin* mRNA was studied under different conditions. The gene *F-luc-Nerfin* is not part of the endogenous genome of *drosophila melanogaster* but was introduced via a plasmid. A second plasmid was introduced that expresses *miR-9b* primary transcripts, a micro RNA that is thought to be important for the deadenylation of the *F-luc-Nerfin* mRNAs. Its action consists to form complexes with enzymes involved in the deadenylation process and through its specific sequence it is designated to dock to *F-luc-Nerfin* mRNA. Two deadenylation complexes are known, the *PAN2-PAN3* complex and the *CCR4-CAF1-NOT* complex. Both are known to lead independently to the deadenylation and subsequent degradation of the mRNAs. The authors suggest that degradation proceeds as a multi-step process consisting of deadenylation by one of the complexes, decapping via *DCPL* and $5' \to 3'$ degradation via *XRN1*. The introduction of the *miR-9b* vector allows studying the role of miRNA-mediated mRNA degradation.
In the experiment they compared the decay pattern under the following conditions

1. Control 1: No *miR-9b* was expressed at all (green)
2. Control 2: *miR-9b* was expressed but neither the enzyme *NOT* nor *PAN2* was expressed (black)
3. *miR-9b* was expressed but only *NOT* was expressed (knock-down of *Pan2*) (yellow)
4. *miR-9b* was expressed but only *Pan2* was expressed (knock-down of NOT) (blue)
5. *miR-9b* was expressed as well as *Pan2* and *NOT* (red)

Thus they can study the role of miRNA-mediated mRNA degradation and the contribution of the complex involving *NOT* and *Pan2*. The resulting experimental decay patterns after the interruption of transcription via incorporation of *actinomycin D* is shown in figure B.2 (circles). Clearly, all patterns decay rapidly in the beginning and more moderately at later times - thus they belong to the category 3 described in chapter 3. Moreover, the slope at later times is similar for all conditions. The most stable mRNAs are under condition (2) and the most unstable under condition (5). Furthermore, *Pan2* (blue) contributes by far more to the destabilization of mRNA than *NOT* (yellow).
To understand the experiment we apply the Markov chain model introduced above. Here, we assume 3 different states which denote mRNAs prior to deadenylation (1), prior to decapping (2) and prior to $5' \to 3'$ degradation (3). We also consider the possibility that the mRNAs are degraded independently of deadenylation and decapping. The biochemical mechanism of these pathways are not entirely clear but could denote degradation by endonucleolytic cleavage. Thus, we there are 5 independent rates. On the one hand, the transition rates λ_1 and λ_2 that describe the processes of deadenylation and decapping, respectively. On the other hand, degradation from each of the three different states is described by the rates μ_1,μ_2 and μ_3. Most of the rates should not be affected from the

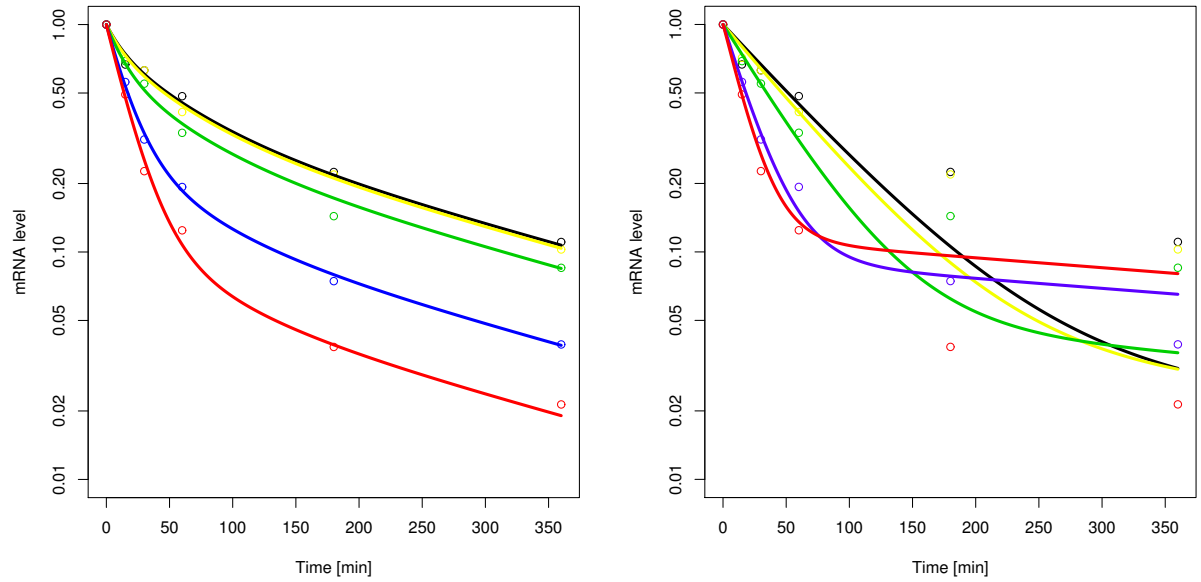

Figure B.2.: **Model of pathway knockout in *drosophila***. The different data points (circles) denote different experimental conditions of an mRNA decay experiment: No *miR-9b* was expressed (green), *miR-9b* was expressed but none of the deadenylation complexes (black), all deadenylation complexes were expressed (red), only *PAN2* (blue) and only *NOT* was expressed (yellow). Data adapted from [17]. The lines correspond to a fit of our theory to the data. Thereby, we obtain rates of decapping, degradation and deadenylation for each of the physiological conditions. Left Panel: Model with free parameter μ_1, right panel: Model with $\mu_1 \equiv 0$ (see text).

differing conditions in the experiment. Therefore, we focus on a description in which only the rate λ_1 is different across the set of conditions.
We compare our theoretical description to the experimental data via non-linear regression. First, we compute the lifetime probability function Φ_U from the Markov chain model and hence obtain the theoretical decay pattern $\Lambda(\Delta t)$. We find the unchanged rates λ_2, μ_1,μ_2 and μ_3 as well as the altering rates under the 5 different conditions $\lambda_1^{(1)}$, $\lambda_1^{(2)}$, $\lambda_1^{(3)}$, $\lambda_1^{(4)}$ and $\lambda_1^{(5)}$ via minimizing the root-mean square difference between the experimental data points and theoretical prediction at all time points tp of the experiment:

$$min \sum_{cond=1}^{5} \sum_{tp=1}^{6} \left(\mathrm{Exp}^{cond}(tp) - \mathrm{Th}^{cond}(tp)\right)^2 \tag{B.10}$$

Hence, there are 5 data sets with 6 points each and 9 independent parameters. To find the global minimum we have used 10000 different initial conditions.
We find the rates to be $\lambda_2 = 0.003\,\mathrm{min}^{-1}$, $\mu_1 = 0.06\,\mathrm{min}^{-1}$, $\mu_2 = 0.015\,\mathrm{min}^{-1}$ and $\mu_3 = 0.004\,\mathrm{min}^{-1}$. Clearly, degradation is initially dominated by μ_1 which exceeds the other degradation rates. This follows directly from the shape of the experimental decay pattern (first fast then slow decay). However, the nature of this rapid initial and decapping-pathway-independent mechanism is unclear. One might speculate that it is related to degradation of pre-mRNAs in the nucleus or via endonucleolytic decay in the cytoplasm. The deadenylation rates under the different conditions read $\lambda_1^{(1)} = 0.0127\,\mathrm{min}^{-1}$, $\lambda_1^{(2)} = 0.0238\,\mathrm{min}^{-1}$, $\lambda_1^{(3)} = 0.0216\,\mathrm{min}^{-1}$, $\lambda_1^{(4)} = 0.0036\,\mathrm{min}^{-1}$, $\lambda_1^{(5)} = 0.0016\,\mathrm{min}^{-1}$.

Figure B.2 shows that our model can describe the decay data very accurately. In particular, the model grasps the non-exponential decay of the mRNAs. However, since we have found a very large initial degradation rate μ_1, the magnitude of the rates for the transition $\lambda_1^{(i)}$ are in reverse order than expected. In the case when both deadenylation complexes are present, the deadenylation rate is lowest. Thus, we conclude that we have to be careful in interpretation of the fitting results.
In an improved attempt we set the degradation rate $\mu_1 \equiv 0$ to prevent this artifact. When we fit each pattern independently, we obtain good fitting results. However, we seek a description where the rates λ_2, μ_1, μ_2 and μ_3 are the same for all decay patterns and only the $\lambda_1^{(i)}$ vary (*i.e.* only the rate of deadenylation). A systematic fitting procedure similar to the one described above (however with $\mu_1 = 0$) yields the corresponding rates $\lambda_1^{(i)}$. However, the resulting fitted decay patterns show only a relatively poor agreement with the experimental data points.

B.4. Appendix to ribosome shielding model

Figure B.3 shows the fitting error of the comparison of the ribosome shielding model to experimental data, as discussed in section 3.3. Moreover, it displays the distribution of the degradation rates, ω_a and ω_b, obtained from the analysis.

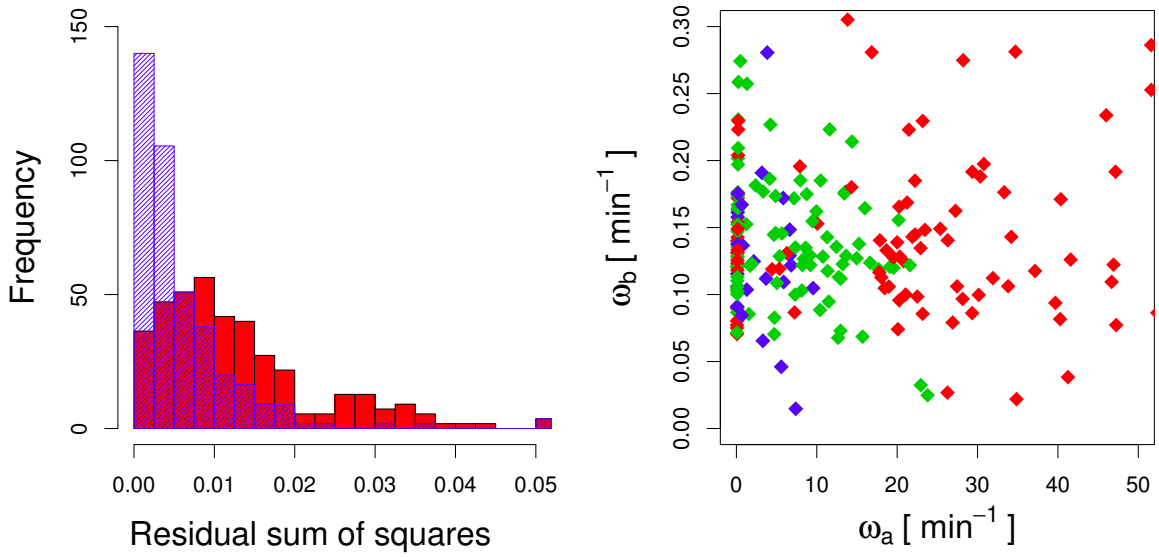

Figure B.3.: **Error reduction and degradation rates in the ribosome shielding model**. The left panel displays a histogram of the fitting error (residual sum of squares between model function and experimental data). It shows that the error pertaining to the ribosome shielding model (blue) is considerably lower than in a model based on an exponential lifetime distribution (red) - on average by almost an order of magnitude. The right panel shows the fitted degradation rates in the shielding model, ω_a vs ω_b. Clearly, the rate describing the unprotected mRNA, ω_a, is much larger than the rate ω_b. Nevertheless, the distribution of the rates for the different genes is widely scattered indicating the individual regulation of each gene. The different colors denote mRNAs of different length (red $L \leq 250$ codons , green $250 \leq L < 500$ codons , blue $L > 500$ codons). Apparently, longer mRNAs have a smaller initial degradation rate ω_a.

Coupling to transcription

In prokaryotes, there is no cell nucleus. Both, the ribosomes and the RNA polymerase move in $5' \rightarrow 3'$ direction. That means that, in principle, mRNAs can be translated even before transcription has terminated [54, 81]. Electro-graph images of the lacZ gene gave visual examples of this mechanism [96], although it was also found that at least some mRNAs in prokaryotes were translated in the vicinity of the membrane [79, 80]. One can distinguish between a ribosome elongation speed v and a polymerase elongation speed w. Equivalently, on can define the basic time scale of translation $t_L = L/v$ and the basic time scale of transcription $t_c = L/w$. To avoid that ribosomes clash into the polymerase, it is reasonable to assume $v < w$. Experimentally, different relations ranging form $v \sim w$ and $v \ll w$ were found. Additionally, one might think that translation starts only after a certain lag time t_0 corresponding to a minimum transcript length $l_0 = w\,t_0$. This

corresponds to the length before the mRNA is sufficiently separated from the polymerase and the average waiting time before the first ribosomes binds to the mRNA. One can now think of the following extended model for the time-dependence of the degradation rate:

$$\omega_{\mathrm{deg}}(t) = \begin{cases} \omega_b vt/L + \omega_a(wt - vt)/L & t < t_c \\ \omega_b vt/L + \omega_a(L - vt)/L & t_c \leq t < t_L \\ \omega_b & t \geq t_L \end{cases} \tag{B.11}$$

Transcription and translation start at the same time. The part of the mRNA covered by ribosomes grows with speed v while the total length of the transcript grows with speed w. At time t_c, the transcript reaches its full length L and in the following the uncovered part decreases linearly proportional to v until time t_L the mRNA is fully covered by ribosomes. Note that we can additionally include the initial lag time t_0 where the mRNA is transcribed but not yet translated.
To explain the experimental data in *E. coli* with this extended model, we would either have to assume a higher transcription elongation speed w than translation elongation speed v or a very high parameter ω_a.
Hence, the implications of this model depend critically on the difference between v and w - however both can be different for each gene under consideration and such detailed experimental knowledge is currently not available.

C. Appendix to chapter 4: Transient mRNA expression

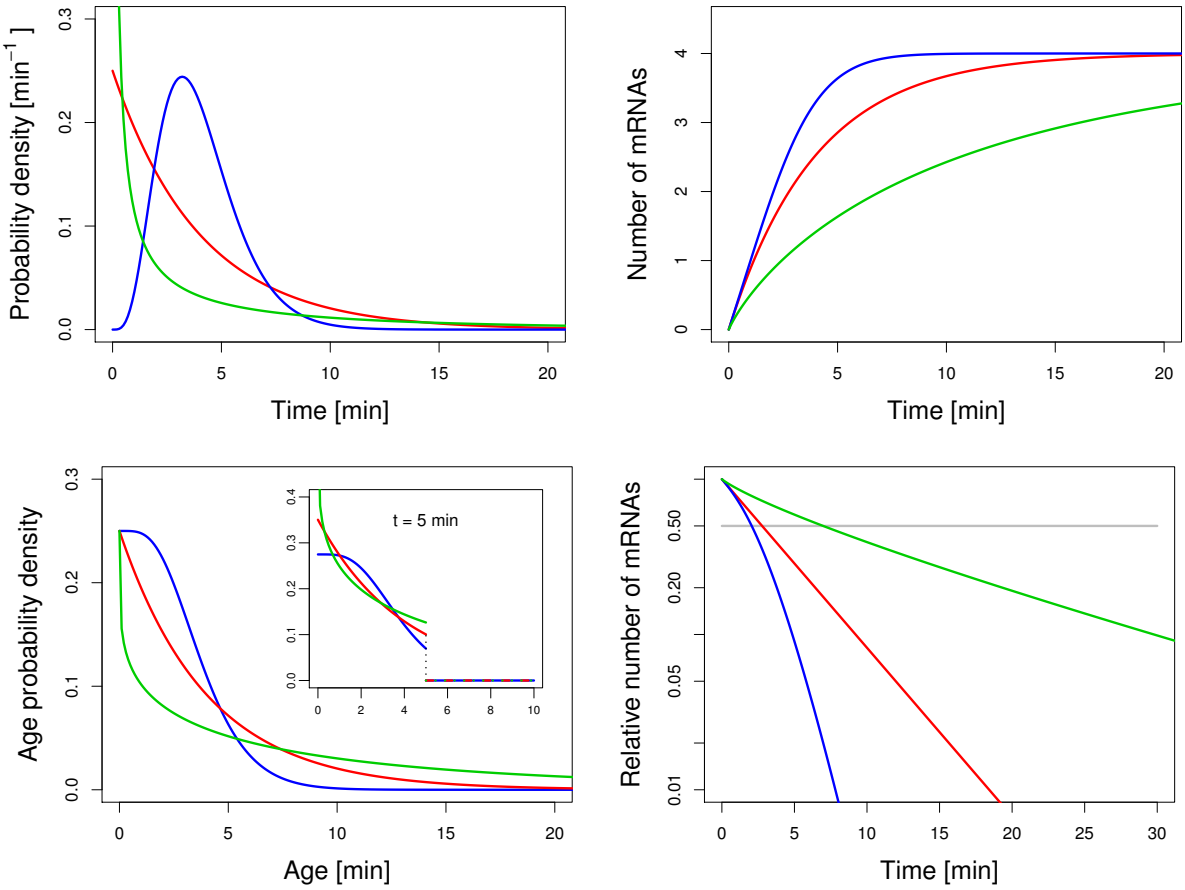

Figure C.1.: **Time evolution of mRNA amount after start of transcription for more exemplary cases**. Upper-left Panel: The mRNA lifetime probability densities ϕ_U is a gamma distribution with shape parameters $n = 1$ (red), $n = 5$ (blue) and $n = 1/5$ (green) and average lifetime $\langle U \rangle = 4$ min. The upper-right panel shows the evolution of the mRNA after the start of transcription, the lower-left panel the shape of the age distribution and the lower-right panel the decay pattern for all three example cases. Clearly, the variation of ϕ_U determines the shape of the other curves. One can see these three cases as generic examples for the different lifetime distributions obtained from experimental decay patterns.

C.1. Gradual modulation

In section 4.6 we discussed a possibility how we can account for a modulation of the transcription rate. We proceeded via two subsequent Poisson processes with different transcription rates. Certainly, one can extend this to an arbitrary number k of transcription rates, *i.e.*

$$N_r^k(t) = \sum_{j=1}^{k-1} \omega_j \mathcal{H}_{t-t_{stop}^{(j)}}^{t-t_{start}^{(j)}} + \omega_k \mathcal{H}_0^{t-t_{start}^{(k)}} \,. \tag{C.1}$$

where $t_{start}^{(j)}$ and $t_{stop}^{(j)}$ are the time points of start and the stop of each transcription phase j. However, it is desirable to find a continuous description where the transcription rate is given by

$$\omega(\tau) \equiv \omega_j \qquad \text{for} \quad t_{start}^{(j)} \leq \tau < t_{stop}^{(j)} \tag{C.2}$$

Hence,

$$\begin{aligned} N_r^k(t) &= \sum_{j=1}^{k-1} \omega_j \int_{t-t_{stop}^{(j)}}^{t-t_{start}^{(j)}} du\,(1-\Phi_U(u)) + \omega_k \int_0^{t-t_{start}^{(k)}} du\,(1-\Phi_U(u)) \\ &= \sum_{j=1}^{k-1} \int_{t-t_{stop}^{(j)}}^{t-t_{start}^{(j)}} du\,\omega(t-u)\,(1-\Phi_U(u)) + \int_0^{t-t_{start}^{(k)}} du\,\omega(t-u)\,(1-\Phi_U(u)) \\ &= \int_0^t du\,\omega(t-u)\,(1-\Phi_U(u))\,, \end{aligned} \tag{C.3}$$

where we have exploited Eq. (C.2), $t_{start}^{(1)} = 0$ and the fact that $t_{start}^{(j+1)} \equiv t_{stop}^{(j)}$. If we assume an instant switching of the rate, *i.e.*

$$\omega(\tau) = \begin{cases} \omega_1 & 0 < \tau \leq t_1 \\ \omega_2 & t_1 < \tau \leq t \end{cases} \tag{C.4}$$

we find (after transforming $\tau \to t-u$)

$$N_r(t) = \int_{t-t_1}^{t} du\,\omega_1\,(1-\Phi_U(u)) + \int_0^{t-t_1} du\,\omega_2\,(1-\Phi_U(u)) = \omega_1\,\mathcal{H}_{t-t_1}^{t} + \omega_2\,\mathcal{H}_0^{t-t_1}\,. \tag{C.5}$$

Thus, we have recapitulated the result from chapter 4 (note the different definition of the starting point of the time t). We can also find a solution for a gradual, linear change of the transcription rate from ω_1 to ω_2 in the interval (t_1, t_2), *i.e.*

$$\omega(\tau) = \begin{cases} \omega_1 & \tau \leq t_1 \\ \omega_1 + \dfrac{\omega_2 - \omega_1}{t_2 - t_1}\,(\tau - t_1) & t_1 < \tau < t_2 \\ \omega_2 & \tau \geq t_2 \end{cases} \tag{C.6}$$

Hence, after transforming $\tau \to t - u$ we find

$$\begin{aligned} N_r(t) = &\int_{t-t_1}^{t} du\, \omega_1 \left(1 - \Phi_U(u)\right) \\ &+ \int_{t-t_2}^{t-t_1} du\, \left(\omega_1 + \frac{\omega_2 - \omega_1}{t_1}\left(t - u - t_1\right)\right) \left(1 - \Phi_U(u)\right) \\ &+ \int_{0}^{t-t_2} du\, \omega_2 \left(1 - \Phi_U(u)\right). \end{aligned} \tag{C.7}$$

The resulting evolution of the mRNA amount is depicted in figure C.2 A. Finally, we can also consider the case of an oscillating transcription rate $\omega(\tau) = \omega_0 \sin(\tau/t_0)$

$$N_r(t) = \omega_0 \int_0^t du\, \left(1 + sin((t-u)/t_0)\right) \left(1 - \Phi_U(u)\right) \tag{C.8}$$

The resulting fluctuation of the mean mRNA number is given in figure C.2 B.

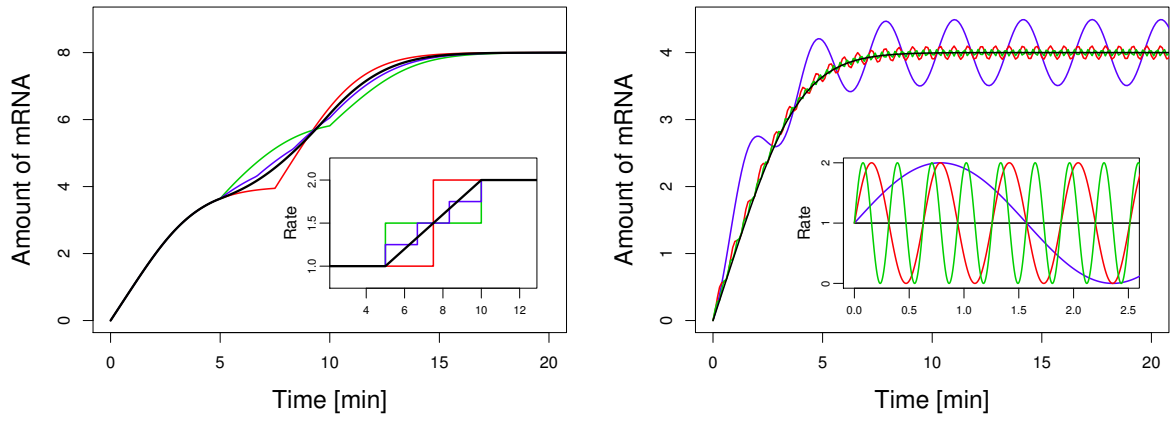

Figure C.2.: **Temporal evolution of the mRNA amount for a gradual change of the transcription rate**. The left panel shows a linear change of the transcription rate from $\omega_1 = 1\,\mathrm{min}^{-1}$ to $\omega = 2\,\mathrm{min}^{-1}$ in the interval $(5, 10]$ min (black). The colored lines show the evolution for different discrete approximations and the inset displays the corresponding rate changes. In the right panel we depict a sinusoidal modulation of the transcription rate, *i.e.* $\omega\tau = \omega_0(1 + sin((t-u)/t_0))$ (red: $t_0 = 0.5$ min, green: $t_0 = 0.1$ min, blue: $t_0 = 0.05$ min). The evolution is similar to a constant transcription rate (black) - however the oscillations persist throughout the time evolution.

C.2. Delayed stress response

The application of a stress such as the exposure of the yeast cells to DNA damaging *MMS* does not necessarily mean that it triggers an immediate response. Rather, the complex network of regulation means that for each gene the altered mode of transcription and degradation can be delayed by a time constant. Here we consider the possibility that either one of them or both are delayed by a constant δ. These extension of the theoretical description allows to describe the stress response patterns with greater flexibility, however on the cost of introducing a free parameter. Although this parameter has a clear interpretation, it would be more satisfactory to have a better understanding of the origin of the delay from the perspective of regulation. One can describe the evolution of mRNAs before and after stress application at $t = t_s$ with

$$N_r(t) \equiv \begin{cases} f(\omega_1, \Phi_U^{(1)}) & t \leq t_s \\ f(\omega_1, \Phi_U^{(2)}) \text{ or } f(\omega_2, \Phi_U^{(1)}) \text{ or } f(\omega_1, \Phi_U^{(1)}) & t_s < t < t_s + \delta \\ f(\omega_2, \Phi_U^{(2)}) & t \geq t_s + \delta \end{cases} . \tag{C.9}$$

where for brevity we have used a function $f(\omega, \Phi_U)$ to describe the product of different ω_j and $\mathcal{H}_{t_1}^{t_2}$. For example, in the first line it reads $N_r(t) \equiv \omega_1 \mathcal{H}_0^{t_s} \equiv \omega_1 \int_0^{t_s} d\tau \, (1 - \Phi_U(\tau))$. Figure C.3 shows the stress response for two exemplary cases in the experiment on stressed

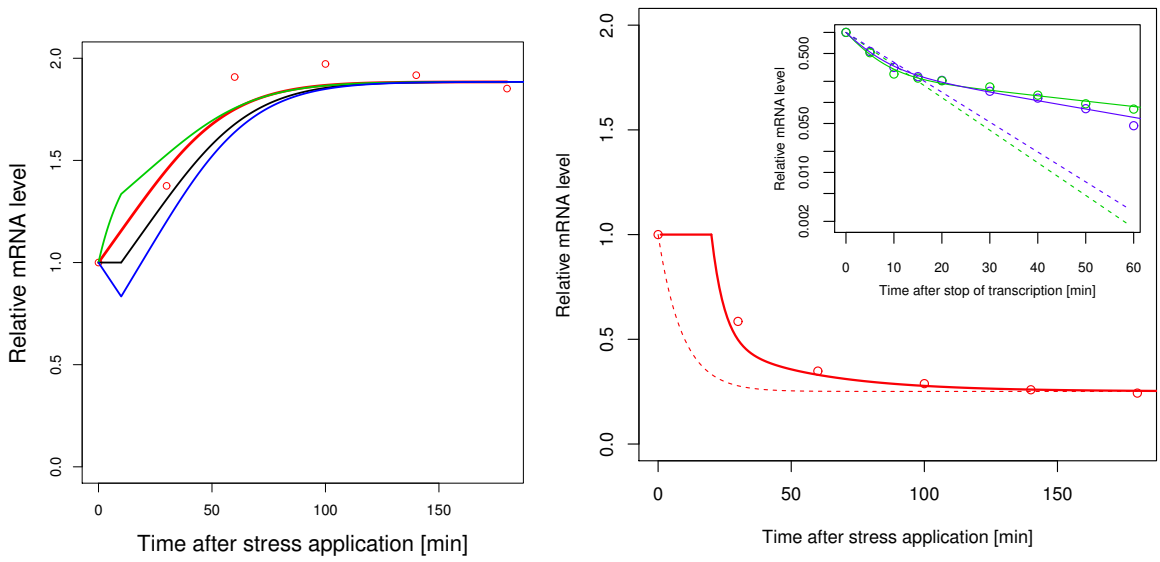

Figure C.3.: **Delayed stress response**. Left panel: Example of stress response of mRNA expression of *HAT2*. Additional to the analysis in the main section (red), here we also consider a possible delay of either the change of transcription (blue), of degradation (green) or both (black) by 10 min. For the given example, it does not yield an improved description of the experimental data. Right panel: Example of negative response after stress application. In this example, a delay of both, transcription and degradation by 20 min improves the description of the mRNA stress response pattern.

yeast cells [27]. In the case of *HAT2* is does not improve the description of the experimental stress response pattern, thus indicating that a possible delay would be rather short. Conversely, for the second example in figure C.3, the mRNA level decreases after the stress application and without a delay of both, transcription and degradation response, the resulting pattern cannot be described in a satisfactory manner.

C.3. Interaction of global and molecular time scale

In the description of stress response given in section 4.7, we mentioned that due to the stress the lifetime distribution of the mRNAs is affected.
In the most simple description, the lifetime distribution changes instantanously from $\Phi_U^{(1)}$ to $\Phi_U^{(2)}$ at the moment of stress application. Hence, the total number of mRNAs at each time t is given by the contributions of the mRNAs originating before and after the application of the stress,

$$N_r^{\mathrm{tot}} = N_r^{(1)} + N_r^{(2)} = \omega_{\mathrm{tc}}^{(1)} \int_t^\infty d\tau \, (1 - \Phi_U^{(2)}(\tau)) + \omega_{\mathrm{tc}}^{(2)} \int_0^t d\tau \, (1 - \Phi_U^{(2)}(\tau)) \qquad \text{(C.10)}$$

where $\Phi_U^{(2)}$ denotes the lifetime distribution of the cells under stressed conditions.
In the simplified description given by Eq. (C.10), we made the assumption of a instant change of the lifetime distribution from $\Phi_U^{(1)}$ to $\Phi_U^{(2)}$. If we consider figure 2.1 in chapter 2 again, we can assert that such instant change is not clearly defined in our theoretical framework. The question arises how the lifetime distribution of an mRNA generated prior to the stress is affected due to the stress application. This is a good example of the interaction of molecular time scale (the age of each individual mRNA) and a global time scale (the stress). If the mRNAs had no memory of the past, *i.e.* if they were described by an exponential distribution, one could just adapt the degradation rate from $\omega_r^{(1)}$ to $\omega_r^{(2)}$. However, for general lifetime distributions Φ_U this is not the case and it is not entirely clear how one can treat this problem exactly (see below). However, also from the perspective of a biological interpretation, one does not know what a change of the degradation mechanism means. One could argue that the rates under the two conditions changed, hence the age (as defined as the time elapsed since the transcription) of an mRNA would determine the life expectancy. However, the stress could also lead to a change of the degradation mechanism, e.g. other regulating enzymes are produced. Therefore, biochemical modifications of the mRNA prior to the stress might not be related to the degradation pathway under stressed conditions. In this sense, the accumulated age of an mRNA prior to the stress would not be relevant for its degradation state. Thus it should be clear that a full solution to this problem is non-trivial from the perspective of mathematics but also from the point of view of biochemistry.

Improved theory of mRNA fluctuations

In chapter 2, Eq. (2.7), we have found that

$$\Pr\{Y(t) = k \mid X(0) = 0, X(t) = n\} \; = \; \binom{n}{k} p^k (1-p)^{n-k} \,,$$

This lead to Eq. (2.8), *i.e.*

$$\Pr\{Y(t) = k \mid X(0) = 0\} \;=\; \frac{(p\,\omega_{\mathrm{tc}} t)^k}{k!} e^{-p\,\omega_{\mathrm{tc}} t}\,,$$

where p is the probability of a single mRNA to be still alive at time $t = t_s + \Delta t$. Here, t_s denotes the time of the stress addition and Δt the time elapsed since then. Hence, we want to compute p in a similar fashion as in Eq. (2.13).

$$\begin{aligned} p \;&=\; \Pr\{Z \geq t_s + \Delta t\} \\ &=\; \frac{1}{t_s}\int_0^{t_s} \mathrm{d}s\; \Pr\{Z = O + U \geq t_s + \Delta t \mid O = s\} \\ &=\; \frac{1}{t_s}\int_0^{t_s} \mathrm{d}s\; \Pr\{U \geq t_s + \Delta t - s\} \\ &=\; \frac{1}{t_s}\int_0^{t_s} \mathrm{d}s\; \Pr\{U \geq t_s - s\}\, \Pr\{U \geq t_s + \Delta t - s \mid U \geq t_s - s\}\,. \end{aligned} \tag{C.11}$$

In the following we want to introduce the stress in our formalism. Therefore, we introduce two distinct and independent random variables U_1 and U_2 which denote the mRNA lifetime before and after the stress application, respectively. They represent different modes of mRNA stability regulation under the different conditions and are described by two independent lifetime distributions $\Phi_U^{(1)}$ and $\Phi_U^{(2)}$. Hence, Eq. (C.11) takes the following form

$$p \;=\; \frac{1}{t_s}\int_0^{t_s} \mathrm{d}s\; \Pr\{U_1 \geq t_s - s\}\, \Pr\{U_2 \geq t_s + \Delta t - s \mid U_2 \geq t_s - s\}\,. \tag{C.12}$$

The first term dictates that the mRNA is still alive at $t = t_s$, *i.e.* it has not been degraded under the unstressed conditions. The second term gives the probability that the mRNA is still alive an interval Δt after the stress application - now under the stressed condition. We can reformulate Eq. (C.12), using the definition of the conditional probability, the definition of the lifetime probability distributions $\Phi_U^{(i)}$ and a suitable transformation of the variable s,

$$\begin{aligned} p \;&=\; \frac{1}{t_s}\int_0^{t_s} \mathrm{d}u\; \frac{1 - \Phi_U^{(1)}(u)}{1 - \Phi_U^{(2)}(u)} \Big(1 - \Phi_U^{(2)}(u + \Delta t)\Big) \\ &=\; \frac{1}{t_s}\int_{\Delta t}^{t_s + \Delta t} \mathrm{d}u\; \frac{1 - \Phi_U^{(1)}(u - \Delta t)}{1 - \Phi_U^{(2)}(u - \Delta t)} \Big(1 - \Phi_U^{(2)}(u)\Big)\,. \end{aligned} \tag{C.13}$$

The two forms correspond to different substitutions and the form can be chosen for best convenience. Note that for most cases, the integral in Eq. (C.13) cannot be solved analytically. Nevertheless, for every lifetime distribution $\Phi_U(t)$ we can compute the distribution and average mRNA number after the stress application. Eq. (C.13) together with Eq. (2.8) gives the contribution of the decaying amount arising from mRNAs generated before the stress. The increase of the number of mRNAs generated after the stress can be

computed according to Eqs. (2.9) and (2.10). Hence, we find the time evolution of the mRNA evolution after the stress application

$$\begin{aligned} N_r^{\mathrm{tot}}(\Delta t) &= N_r^{\downarrow}(\Delta t) + N_r^{\uparrow}(\Delta t) \\ &= \omega_{\mathrm{tc}}^{(1)} \int_0^{\infty} \mathrm{d}u \, \frac{1-\Phi_U^{(1)}(u)}{1-\Phi_U^{(2)}(u)} \left(1-\Phi_U^{(2)}(u+\Delta t)\right) + \omega_{\mathrm{tc}}^{(2)} \int_0^{\Delta t} \mathrm{d}u \left(1-\Phi_U^{(2)}(u)\right) . \end{aligned} \tag{C.14}$$

Thus, we have found an exact description of the number of mRNAs under stressed conditions. The advantage of this approach is that it is fairly general, it holds independent of the underlying model of $\Phi_U(t)$. However, this renders the biological interpretation of the transition at $t = t_s$ more difficult. Some examples of the effect of this description are given in figures C.5 and C.4. Furthermore, in the main text, section 4.7, the experimental data to stress response were analyzed with this extended description.

Stress response in the framework of the Markov chain model

We have already seen in section 4.5 that the aging of an mRNA becomes manifest in the population probabilities of the states of a Markov chain (see Eq. (4.14)). Under the assumption that the mRNA turnover has reached a steady-state before the stress application we have

$$\psi_k = \int_0^{\infty} da \, \xi_k(a) \, \phi_A^{\mathrm{st}}(a) \, .$$

Here, the $\xi_k(a)$ denote the probability that an mRNA of age a is in state k of a Markov chain conditioned that the mRNA has not yet been degraded - as defined in Eq. (4.13). Thus, at the time of stress application the mRNAs are found in each of the states $k = 1, \ldots, n_{\max}$ with probability ψ_k. Correspondingly, for an mRNA in state k the Markov chain describing its degradation pathways is reduced to the last $n_{\max} - (k-1)$ states. In this interpretation, we assume that the underlying description of degradation, *i.e.* the number of states in the Markov chain model, does not change due to the stress application. Therefore, only the transition and degradation rates are influenced by the stress and their values can be obtained by a fit of the Markov chain model to the experimental decay data (as has been done in section 4.7). Consequently, the decaying amount of mRNAs that have been generated prior to the stress is given by the contributions of the mRNAs in each state k and their corresponding reduced degradation pathway. It reads

$$N_r^{\downarrow}(\Delta t) = N_r(0) \left(\sum_{k=1}^{n_{\max}} \psi_k \int_{\Delta t}^{\infty} du \, \left(1 - \tilde{\Phi}_U^{(n=n_{\max}-(k-1))}(u)\right) \right) . \tag{C.15}$$

Here, $\tilde{\Phi}_U^{(n)}$ denotes the mRNA lifetime probability distribution with n states, *i.e.* a Markov chain model where only the last n states of the original model with $n_{\max}$ states are remaining. The tilde in $\tilde{\Phi}_U$ should emphasize that the decay under stressed conditions occurs with rates $\tilde{\lambda}_i$ and $\tilde{\mu}_i$ (obtained from the decay pattern long after the stress application). The evolution of the newly generated mRNAs after the stress application is just given by $N_r^{\uparrow}(\Delta t)$ as outlined in section 4.7.

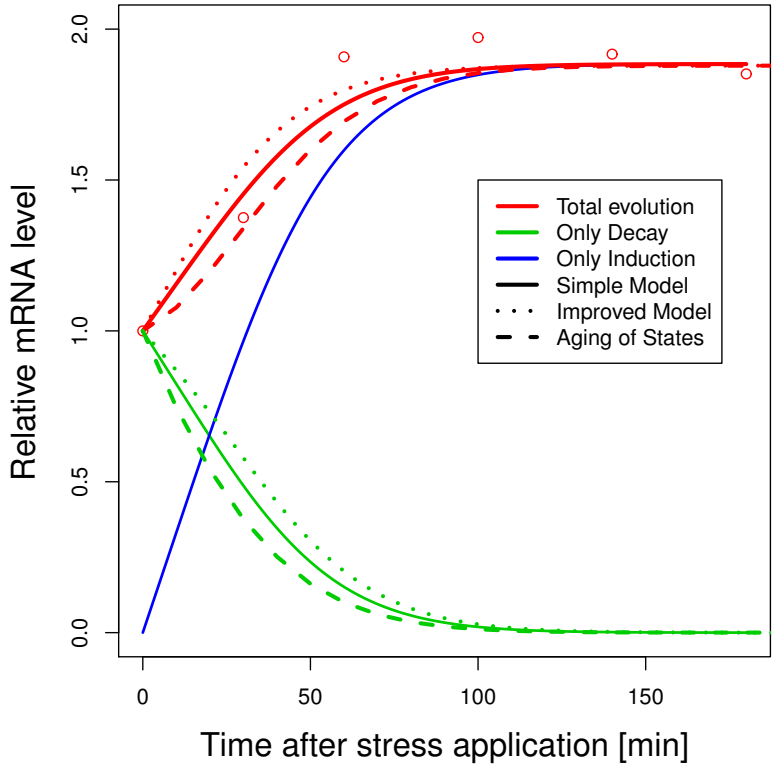

Figure C.4.: **Stress response with refined models**. Shown is the response to stress application for the gene *HAT2* in *S. cerevisiae* (circles). We compare different improvements of the theoretical description to the experimental data (see text). While the theory based on Eq. (C.14) (dotted) leads to a slight improvement of the description, the model based on aging of states as given in Eq. (C.15) (dashed) is not more successful than the simple description (solid).

This approach is limited to the interpretation of $\Phi_U(t)$ in the framework of a Markov chain model. It gives exact results with a clear biological interpretation. However, from the perspective of the biochemistry occurring due to the stress it remains unclear if this is the correct description. It limits the action of the stress only to a change of the rates in the Markov model - however the topology of the state network in the Markov chain remains unchanged. Some examples of the effect of this description are given in figures C.5 and C.4.

Non-constant degradation rate

A different approach considers the change of the degradation rate. In section 2.2 we demonstrated how general lifetime distributions $\Phi_U(t)$ can be mapped to an age dependent degradation rate $\omega_{\text{deg}}(a)$ related to an exponential distribution. Thus, each mRNA has a degradation given by its current age. If at the time of stress application an mRNA has a age a and degradation rate $\omega^{(1)}_{\text{deg}}(a)$, we assume that when the stress becomes manifest it changes to $\omega^{(2)}_{\text{deg}}(a)$. Hence, it carries the memory of its past and takes its age (*i.e.* state of modification) into account for its excess lifetime under the stressed conditions.
To further follow this approach, we have performed stochastic simulations. There, at each

instant of the simulation the life expectancy of an mRNA is determined by its current degradation rate (via an exponential distribution), however, it is constantly updated according to the increasing age. The stress is considered in a change of the underlying function $\omega_{\text{deg}}(a)$ given by Φ_U. Thus, we can follow the number of mRNAs on time and observe the changes near the stress application (see figure C.5 A). Furthermore, we can also determine the time-dependent lifetime probability function $\Phi_{U,t}(u)$ for mRNAs originating in different intervals close before and after the stress application (see figure C.5 B).

Outlook These are three promising approaches how one can solve the interaction of global and local timescales mathematically. A broader ansatz consists of defining an time-dependent age $\phi_A(a,t)$ (or residual lifetime) distribution and derive the time-dependent lifetime distribution $\Phi_{U,t}$ [87]. However, as mentioned earlier, it should be clarified first from the perspective of biology which idea is best to follow.

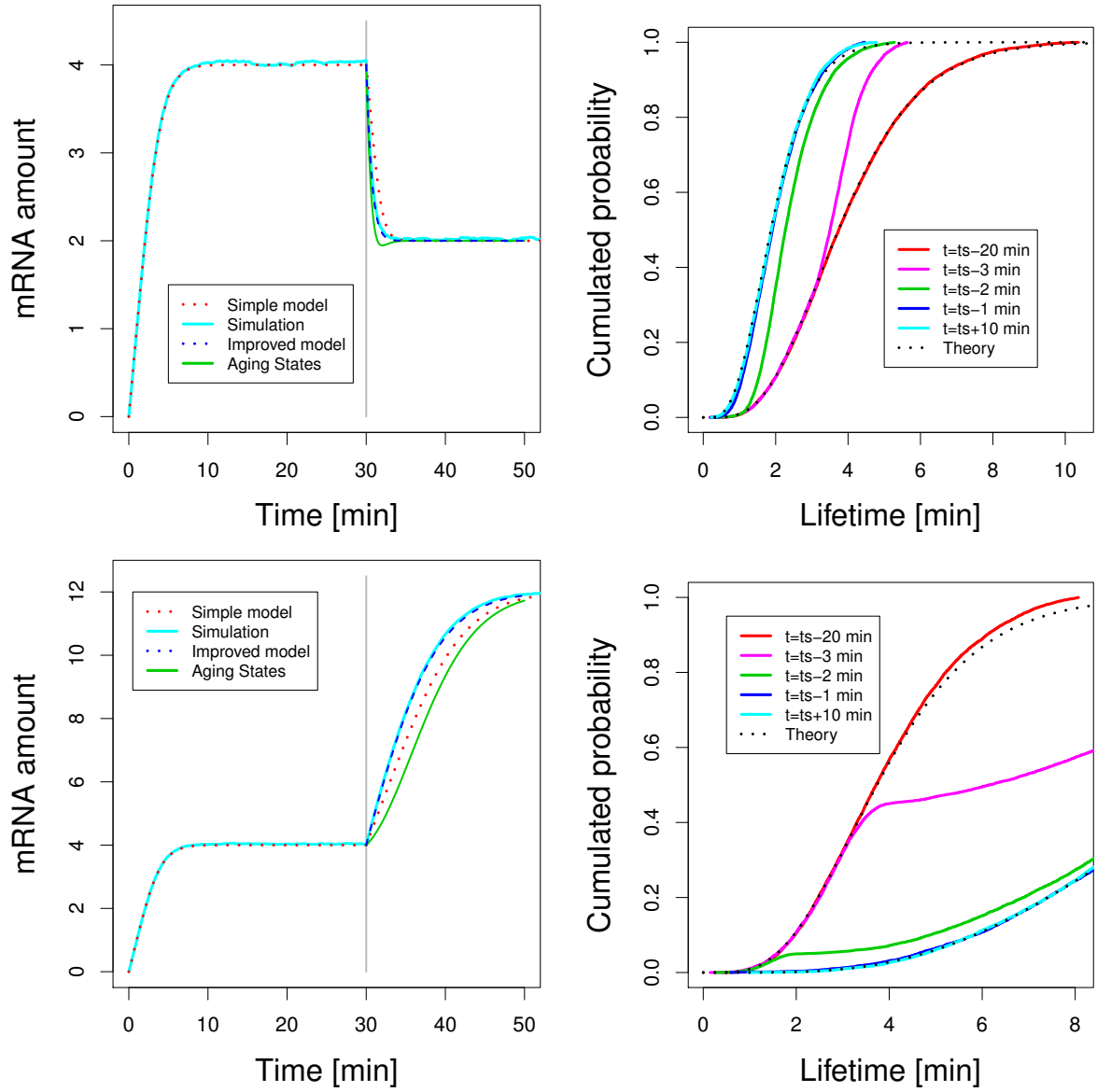

Figure C.5.: **Simulation of stress response**. mRNAs with lifetime density $\phi_U^{(\Gamma)}$ and shape parameter $n = 5$ are synthesized with transcription rate $\omega_{\text{tc}} = 1$ min^{-1}. At $t = 30$ min the scaling parameter of the lifetime distribution is changed such that the average lifetime changes from 4 min to 2 min (top) or from 4 to 12 min (bottom), respectively. In the left panels, we compare the simple model with sudden change of the lifetime distribution (red dotted), results from the simulation of the mRNA turnover process (cyan, see text) as well as the two distinct analytical approaches discussed above (blue and green, respectively). The simulation allows furthermore to extract the mRNA lifetime distribution near the transition at $t = 30$ min, *i.e.* shortly before the stress application (right panels). Shown are the cumulative distributions for mRNAs generated 20 (red), 3 (magenta), 2 (green) and 1 (blue) minute before as well as 10 (cyan) min after the stress application (see also figure legend). For comparison, theoretical curves for $\Phi_U^{(\Gamma)}$ with $\langle U \rangle 4$ and 2 min (top) as well as with $\langle U \rangle = 4$ and 12 min are depicted (black dotted lines). Note that in the first case (top panels), all three different attempts to find an improved description yield similar results. Conversely, for the second case (bottom), the results based on the model of aging Markov chains differs from the simulation and the improved general theory.

D. Different models of transcription

As mentioned earlier, the use of a Poisson process for the transcription of mRNA is well justified when one considers the description on a level of cell populations. There, the average population transcription rate is approximately constant. In single cells, some genes were found to be transcribed according to independent transcription events with a constant rate [36, 35]. However, from other experiments it was inferred that a bursting behavior is noticeable [34, 35], *i.e.* phases with high transcription activity alternate with phases of low activity.

A more detailed model of the interplay of transcription and mRNA degradation also includes a varying transcription rate. This variation is linked to different gene activity by the random binding of various transcription factors. In particular, if the number of available transcription factors and RNA polymerases is low, this gives rise to an additional stochasticity of the rate of transcription.

Therefore, a straightforward extension is to consider a non-homogeneous Poisson process with a time-dependent intensity $\omega(t)$. Moreover, since the variation arises due to the non-trivial stochasticity of the transcription process, the intensity varies in a stochastic fashion, *i.e.* it is a *Cox* process [73]. To extend our analytical formalism that we developed in chapter 2 to a Cox-process is not trivial but will be the focus of future research. Here, we will study effects of a randomly fluctuating transcription rate (*i.e.* intensity) via Monte-Carlo simulations.

The random fluctuation of $\omega(t)$ can be modeled via a continuous-time Markov chain. Starting from the mean intensity $\omega(0) = \omega_{\mathrm{tc}}$, we allow random excursions of a fixed step size χ that occur at random time points according to a rate κ. At each step event i, the intensity changes to $\omega(t_{i+1}) = \omega(t_i) \pm \chi$ with probability 0.5 each. Hence, the intensity parameter performs a symmetric random walk around its mean value ω_{tc}. We are interested in studying the effect of different variability of the intensity on mRNA turnover. Therefore, we additionally introduce an upper bound for the excursions of the random walk. The intensity can vary between $\omega_{\mathrm{tc}}(1-\epsilon) \leq \omega(t) \leq \omega_{\mathrm{tc}}(1+\epsilon)$. The parameter ϵ can be chosen in the range $[0, 1]$, leading to the limit cases of a homogeneous Poisson process for $\epsilon = 0$ and the highest possible variation for $\epsilon = 1$. In this case, transcription can be shut off completely for some time intervals. The temporal fluctuation of $\omega(t)$ is exemplary visualized in figure D.1 A.

To study the effect of a stochastically varying intensity on mRNA turnover, we perform Monte-Carlo simulations. For each realization, we first determine the time course of $\omega(t)$. This, in turn, serves as input for the time-dependent transcription rate in the simulation of mRNA turnover (see chapter 2 and section A.6 in the appendix for more details of the simulation technique).

The parameters χ and κ should reflect the time scale of the fluctuations in complex transcription process. We have chosen $\kappa = 2\,\mathrm{min}^{-1}$ and $\chi = 0.1 \cdot \omega_{\mathrm{tc}}$ which roughly

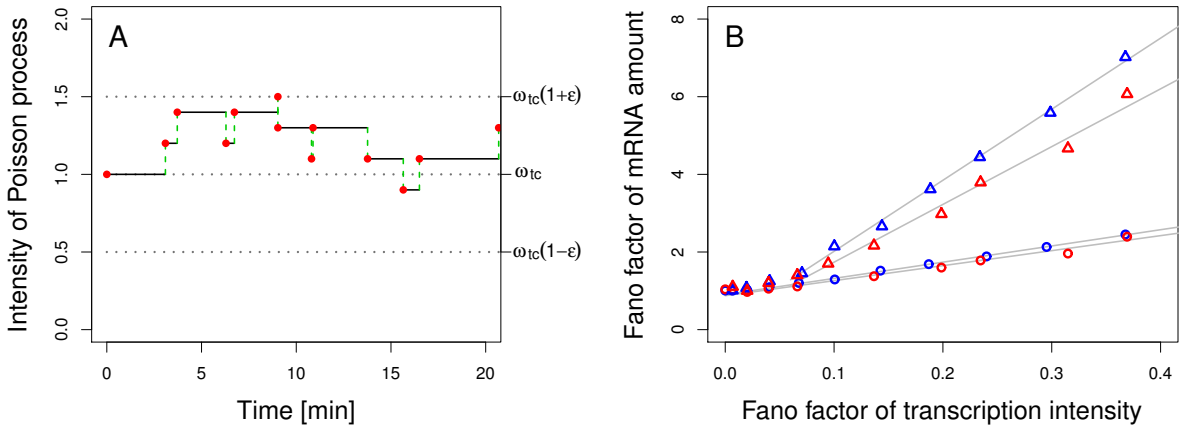

Figure D.1.: **Cox Process for the description of mRNA turnover**. Panel A shows an illustration of the Cox process for the random variation of the intensity of the Poisson process. Starting from an initial value ω_{tc} at $t = 0$ it performs a random walk of fixed step size χ with step events occurring according to a rate κ. The random walk is bound by maximal excursions of $\pm\epsilon$ which determines the variance of ω. According to the intensity parameter $\omega(t)$, mRNAs are generated at random time points and have a random lifetime U which is given by ϕ_U. Panel B shows the evolution of the mRNA Fano factor F^{RNA} vs. the Fano factor of the intensity $F^{(\omega)}$, for different lifetime distributions ϕ_U. The red data points arise from $\phi_U^{(\mathrm{exp})}$ whereas the blue points come from $\phi_U^{(\Gamma)}$ with $n = 5$. The mRNAs represented by the circles have an average lifetime $\langle U\rangle = 4$ min and those shown as triangles $\langle U\rangle = 20$ min. From $F^{(\omega)} = 1$ follows $F^{\mathrm{RNA}} = 1$, which represents the limit case of a homogeneous Poisson process. A Fano factor of 1 is unique for the Poisson process and the deviation from the Poisson process becomes evident in the growth of F^{RNA} with $F^{(\omega)}$. The evolution is different for different values of the shape parameter n and mean lifetime $\langle U\rangle$, however the Fano factor value scales in remarkable good agreement with a linear fit (gray lines).

resembles the time scale of transcription and the order of magnitude of the number of free transcription factors in a cell. After equilibration of the simulation, the variance of ω always attains its maximal value determined by ϵ for a large variation of different parameters χ and κ, *i.e.* it is robust under variation of χ and κ.
In Fig D.1 B we show the dependence of the mRNA Fano factor $F^{\mathrm{RNA}} \equiv \sigma_{N_r}/N_r$ on the Fano factor of the intensity parameter $F^{(\omega)} \equiv \sigma_\omega/\omega_{\mathrm{tc}}$. We can control the variance of $\omega(t)$ via the choice of the upper bound ϵ. As $F^{(\omega)}$ increases, also F^{RNA} departs steadily from the value of a Poisson distribution. Moreover, apart from small $F^{(\omega)}$, the relation is approximately linear.
Figure D.2 displays the time evolution of the average number of mRNAs resulting from non-homogeneous transcription following the induction and interruption of transcription, respectively. Apart from small fluctuations, simulations with different parameters ϵ completely agree with the theoretically predicted evolution corresponding to a homogeneous Poisson process as discussed in chapters 3 and 4. This indicates that our theoretical formalism based on a homogeneous Poisson process is also capable to model the evolution of the average mRNA amount under non-homogeneous transcription conditions.
These simulation results demonstrate how our general theory of mRNA turnover can be expanded. Moreover, our ansatz seems to provide a valid description for the mean field behavior of non-homogeneous *Cox* processes as well. The challenge in a mathematical

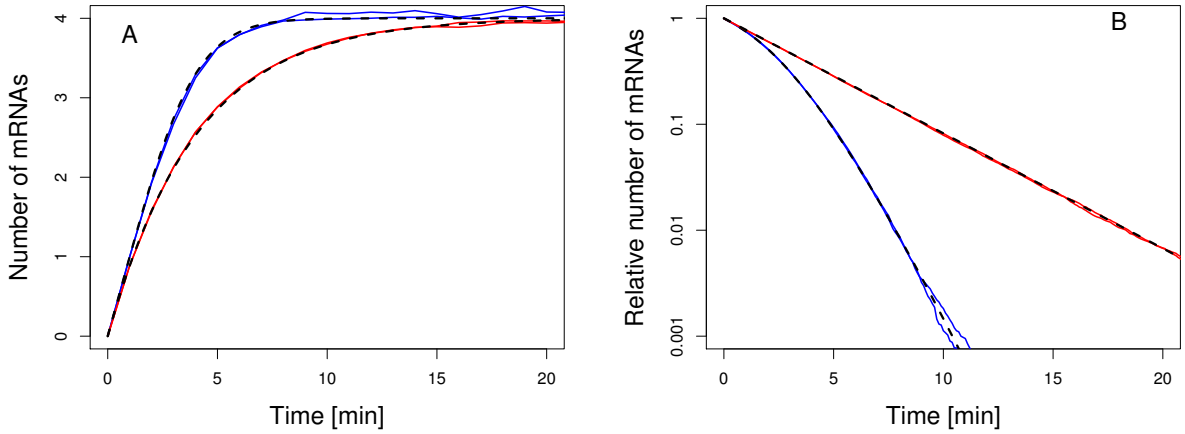

Figure D.2.: **Induction and decay in the Cox Process**. Panels A and B show the patterns of mRNA induction and decay according to the simulation of a Cox process, respectively. In both figures, mRNAs given by $\phi_U^{(\mathrm{exp})}$ (red) and $\phi_U^{(\Gamma)}$ (blue) but identical mean lifetime $\langle U \rangle = 4$ min are depicted. The simulations have been repeated for different parameters ϵ and yield very similar results. The dashed black lines show the analytical results as derived in chapters 3 and 4. Apart from small fluctuations the simulation results completely overlap with the theory of a homogeneous Poisson process. Thus, for the description of the behavior of the average number of mRNAs our general results obtained for the homogeneous Poisson process are applicable for non-homogeneous processes as well.

description lies in combining non-homogeneous transcription with arbitrary lifetime distributions Φ_U. In a Cox process, the distance between mRNA generation events is no longer uniformly distributed which was one of the key assumptions of the derivation in chapter 2. One possible approach consists of introducing a path integral formalism which is not tractable analytically for the cases discussed here.

Abschließende “Gedanken”

> Der Fahrgast, als der Taxichauffeur das hohe Trinkgeld wortlos wegsteckt: “Sagt man eigentlich in Berlin nicht *Danke*?” - “Det ist untaschiedlich”, erwidert der Taxifahrer, “ manche saren et, manche saren et nich.” (*unbekannt*)

Dieser Berliner Unentschlossenheit kann man Folgendes entgegenhalten:

“Es ist ein lobenswerter Brauch,
Wer was Gutes bekommt,
Der bedankt sich auch.”
(*Wilhelm Busch*)

Daher möchte ich hier die Gelegenheit nutzen, mich zu bedanken.

Danke.

Zunächst vielen Dank an Reinhard Lipowsky für das Ermöglichen der Doktorarbeit und die Auswahl des interessanten Themas.
Insbesondere bedanke ich mich bei meinem Betreuer Angelo Valleriani für die hervorragende Betreuung und die harmonische Zusammenarbeit.
Die vergangenen Jahre wären nur halb so erfreulich verlaufen ohne die angenehme und freundschaftliche Arbeitsatmosphäre der Abteilung am MPIKG. Danke liebe Freunde und Kollegen für diese schöne Zeit!
Hervorzuheben ist noch die aufmerksame Lektorenarbeit meiner Kollegen Sophia und Thomas.
Natürlich auch einen besonderen Dank an meine Eltern, Geschwister, Freunde und an Lena - dafür, dass sie da sind.