Stochastic models of protein production in bacterial cells : analysis of regulation mechanisms for transcription and translation phases

Jana Zaherddine

Stochastic models of protein production in bacterial cells: analysis of regulation mechanisms for transcription and translation phases

Jana Zaherddine

Thesis defended on December, 19 2023 for the degree of
Doctor in Mathematics, Sorbonne Université

JURY	
Chairman	Dr. Amandine Veber, Université Paris-Descartes,
Reviewer	Dr. Fabien Campillo, INRIA,
PhD Advisor	Dr. Vincent Fromion, INRAE,
Examiner	Pr. Irina Kourkova, Sorbonne Université,
Examiner	Dr. Olivier Martin, INRAE,
Examiner	Pr. Pascal Moyal, Université de Lorraine,
PhD Advisor	Dr. Philippe Robert, INRIA.

REMERCIEMENTS

Il est difficile de mettre des mots sur l'émotion que je ressens en écrivant ces lignes de remerciements. Depuis mon enfance, j'avais deux rêves qui animaient mes pensées : devenir médecin ou explorer le fascinant monde des mathématiques. Le destin a tracé un chemin pour moi, et aujourd'hui, je suis profondément reconnaissante de préparer ma thèse en mathématiques appliquées à la biologie.

Philippe, par où commencer? Un immense merci pour la constance de ton soutien, la rigueur qui a marqué chacune de nos interactions, et les conseils avisés qui ont façonné ma compréhension des mathématiques. Ces trois années ne sont pas simplement un chapitre de ma vie, mais une aventure que je n'oublierai jamais. Merci pour les discussions qui ont éclairé mon chemin, et pour ta présence rassurante dans les moments où le stress menaçait de prendre le dessus. Ainsi, je tiens à exprimer ma profonde gratitude pour ta contribution essentielle à cette étape cruciale de ma vie académique. Merci d'avoir été bien plus qu'un directeur de thèse, mais un mentor et un guide. Ton impact sur mon parcours académique restera gravé dans ma mémoire.

Vincent, je suis profondément impressionnée par ta passion contagieuse pour la biologie. Tu as ouvert les portes d'un domaine qui n'était pas ma spécialité, et grâce à toi, j'ai non seulement compris, mais j'ai développé un amour pour cette discipline. Merci pour les échanges riches en enseignements, pour ta patience infinie à expliquer les détails, et pour avoir partagé avec moi ta vision passionnée de la biologie. Merci d'avoir partagé ton enthousiasme contagieux, d'avoir été un mentor attentionné, et d'avoir contribué de manière significative à l'enrichissement de mon parcours scientifique.

À mes rapporteurs, Amarjit Budhiraja et Fabien Campillo, merci d'avoir pris le temps d'explorer mon manuscrit avec attention et de me prodiguer des conseils précieux qui ont indéniablement contribué à son amélioration. Un merci particulier aux membres du jury, Fabien Campillo, Irina Kourkova, Olivier Martin, Pascal Moyal et Amandine Veber qui ont accepté de participer à ce moment important de ma vie académique.

L'aventure n'aurait pas été la même sans toute l'équipe Mamba de l'INRIA et l'équipe Maiage de l'INRAE. Votre amabilité, votre collaboration, et l'environnement chaleureux que vous avez créé ont rendu cette expérience encore plus enrichissante. Mes plus sincères remerciements à chacun d'entre vous.

À ma famille, je tiens à exprimer une reconnaissance infinie. Votre présence à chaque étape de mon parcours, vos encouragements constants et le soutien que vous m'avez apporté ont été inestimables. Mon affection pour vous est profonde, et j'espère que notre lien demeure solide tout au long de notre vie. J'adresse également un remerciement particulier à ma mère, sans laquelle je ne serais jamais parvenue à cette étape de ma vie. Merci, maman, pour ton soutien

indéfectible, tes encouragements constants, et ton amour inconditionnel. Cette réussite est aussi la tienne, et je t'aime beaucoup. À mes amis, je ne mentionnerai pas de noms car vous êtes nombreux, mais je tiens à vous remercier du fond du cœur pour tous les moments d'encouragement et le soutien infaillible que vous m'avez accordé dans les moments difficiles. Vous êtes une extension précieuse de ma famille, et je suis reconnaissante d'avoir des amis aussi exceptionnels.

SUMMARY

The process of protein production can be described as a two-step procedure. In the first step, polymerases, macro-molecules, generate RNA molecules from DNA genes. This process is known as transcription. The second step, called translation, involves the synthesis of proteins from messenger RNA (mRNA) molecules using ribosomes, which are large molecular complexes.

Protein production holds immense significance within the cell as it not only facilitates its growth but also consumes a significant portion of its resources. Understanding the mechanisms and regulation of gene expression is a fundamental and complex research topic in the field of biology.

In this thesis, our focus narrows down to two specific regulatory mechanisms. First, there is a transcriptional regulation mechanism that involves a significant macro molecule called 6S RNA. This molecule essentially acts like a regulator, binding to polymerases and inhibiting transcription. Then, we have a translational regulation mechanism, which is governed by a particular nucleotide sequence called (p)ppGpp. inhibits the production of ribosomal RNAs (rRNAs), which are essential components of the ribosomes responsible for protein synthesis. This reduction in ribosomal RNA production limits the availability of functional ribosomes in the cell, leading to a slowdown in protein synthesis, what helps in conserving energy and resources under nutrient-limiting conditions.

Our objective is to develop mathematical models that help us explore these regulation mechanisms. Our main goal here is to create stochastic models that capture both stages of protein synthesis: Transcription and Translation. We are looking to understand some specific regulatory mechanisms and figure out how they impact the activity in the cell. To do this, we have got to create models that not only represent each phase in a reasonably accurate manner but also that give a mathematical interpretation.

Our analysis involves looking at how different macro molecules are distributed during each phase, whether it's transcription or translation, depending on the availability of resources within the cell. Through this analysis, we assess the impact of the regulatory mechanisms on cellular activity.

In order to do all of this, we use scaling methods and given the coexistence of slow and fast processes in our models, e employ the averaging principle using occupation measures as the main mathematical tool. Additionally, a variety of coupling methods are integrated into our approach.

RÉSUMÉ

Le processus de production des protéines peut être décrit comme une procédure en deux étapes. Dans la première étape, les polymérases, de grandes molécules, génèrent des molécules d'ARN à partir des gènes d'ADN. Ce processus est connu sous le nom de transcription. La deuxième étape, appelée traduction, implique la synthèse des protéines à partir des molécules d'ARN messager (ARNm) en utilisant des ribosomes, qui sont de grands complexes moléculaires. La production de protéines revêt une importance immense au sein de la cellule, car elle facilite non seulement sa croissance, mais elle consomme également une part significative de ses ressources. Comprendre les mécanismes et la régulation de l'expression génique est un sujet de recherche fondamental et complexe dans le domaine de la biologie.

Dans cette thèse, notre attention se porte sur deux mécanismes de régulation spécifiques. Tout d'abord, il existe un mécanisme de régulation transcriptionnelle qui implique une macro-molécule significative appelée ARN 6S. Cette molécule agit essentiellement comme un régulateur, se liant aux polymérases et inhibant la transcription. Ensuite, nous avons un mécanisme de régulation translationnelle, qui est gouverné par une séquence nucléotidique particulière appelée (p)ppGpp. Il inhibe la production des ARN ribosomiques (ARNr), qui sont des composants essentiels des ribosomes responsables de la synthèse des protéines. Cette réduction de la production d'ARN ribosomique limite la disponibilité des ribosomes fonctionnels dans la cellule, entraînant un ralentissement de la synthèse des protéines, contribuant ainsi à la conservation de l'énergie et des ressources en conditions de limitation des nutriments.

Notre objectif est de développer des modèles mathématiques qui nous aident à explorer ces mécanismes de régulation. Notre objectif principal ici est de créer des modèles stochastiques qui capturent les deux étapes de la synthèse des protéines : la transcription et la traduction. Nous cherchons à comprendre certains mécanismes de régulation spécifiques et à déterminer leur impact sur l'activité de la cellule. Pour ce faire, nous devons créer des modèles qui représentent non seulement chaque phase de manière raisonnablement précise, mais aussi qui donnent une interprétation mathématique.

Notre analyse implique d'examiner comment différentes macro-molécules sont distribuées pendant chaque phase, que ce soit la transcription ou la traduction, en fonction de la disponibilité des ressources à l'intérieur de la cellule. À travers cette analyse, nous évaluons l'impact des mécanismes de régulation sur l'activité cellulaire. Pour réaliser tout cela, nous utilisons des méthodes de renormalisation, et compte tenu de la coexistence de processus lents et rapides dans nos modèles, nous utilisons le principe d'homogénéisation en utilisant les mesures d'occupation comme principal outil mathématique. De plus, une variété de méthodes de couplage sont intégrées dans notre approche.

Contents

Chapter 1

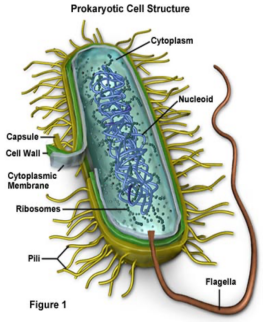
Introduction

1.1 Gene Expression

Gene expression refers to the process of converting genetic information encoded in genes into functional proteins or RNA molecules/non coding RNA molecules. This process involves the transcription of DNA into RNA. When the produced RNA is coding, this first step is followed by a second process called translation. This process translates the coding sequence carried by the coding RNA, called mRNA, into an amino acid polypeptide. In general, after some folding, this polypeptide becomes a protein. Protein production holds immense significance within the cell as it not only facilitates its growth but also consumes a significant portion of its resources. Understanding the mechanisms and regulation of gene expression is a fundamental and complex research topic in the field of biology.

The process of protein production can be described as a two-step procedure. In the first step, polymerases, macro-molecules, generate RNA molecules from DNA genes. This process is known as transcription. The second step, called translation, involves the synthesis of proteins from messenger RNA (mRNA) molecules using ribosomes, which are large molecular complexes.

1.2 A Bacterial Cell



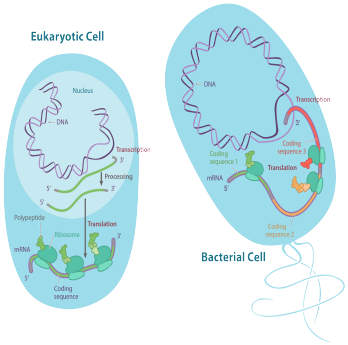
A bacterial cell is a prokaryotic cell, which means it lacks a nucleus and other membrane-bound organelles found in eukaryotic cells. All macro-molecules are colliding with each other in the cytoplasm what gives the process a **stochastic** aspect.

Let us first recall the three main categories of RNAs contained in a bacterial cell:

a. rRNAs, *ribosomal RNAs*, used for the building of ribosomes. A ribosome is a complex assembly of around 50 proteins and, also, of several rRNAs. An rRNA is a long chain of several thousands of nucleotides, it is in particular a costly macro-molecule to produce. Reducing or speeding-up the production of ribosomes, in particular of rRNAs, has therefore a critical impact on resource management of the cell;

b. mRNAs, *messenger RNAs*, used by the translation step (see description below) to produce a protein from mRNAs coding sequences;

c. A large set of RNAs that do not belong to the two previous categories, such as transfer RNAs, tRNAs, or Bacterial small RNAs, sRNAs, often associated to regulation mechanisms.



The production of protein in eukaryotic and bacterial cells involves similar basic steps, transcription and translation. However there are several key differences between these two processes.

For example, in eukaryotic cells, the transcription of DNA into mRNA may take place in the nucleus, while in bacterial cells, in the whole cytoplasm. Another difference is that the transcription and translation take place in the same time, in the sense that mRNA is translated into protein as soon as it is transcribed. This occurs because bacteria do not have a distinct nucleus that separates DNA from ribosomes, so there is no barrier to immediate translation. In addition, unlike eukaryotic cells, in which different types of polymerases exist and each of them is responsible of transcribing a certain type of RNA, in bacterial cells an unique type of polymerase is responsible of the synthesis of mRNA, rRNA and tRNA.

1.2.1 Transcription

During transcription, the DNA is first unwound and a section of it is copied into a complementary DNA strand by the RNA polymerase enzyme. In further detail, it can be described through the three fundamental steps:

a. **Initiation**: the polymerase binds to one of the specificity factors σ to form a *holoenzyme* in order to bind to a specific region on the DNA called the promoter. In our case we focus on the "housekeeping" σ-factor σ^{70} for *E. Coli* and σ^{A} for *B. Subtilis*. This holoenzyme binds to a large set of gene promoters to initialize the transcription. This is the *initiation phase*.This phase is complex at the molecular level and depends on the specific characteristics of the nucleotide sequence defining the promoter region on DNA. This sequence mainly defines the affinity of the polymerase to the promoter (in particular allows to select of a polymerase associated with right sigma factor). In addition, this sequence modulates the specific properties of the initiation, i.e. its rate of success in initiating and the speed of initiation. If this step is successful, the protein σ^{70} is detached and the polymerase completes the elongation of the corresponding RNA.

b. **Elongation**: as RNA polymerase moves along the DNA template, it adds nucleotides to the growing RNA molecule in order to synthesize a complementary RNA sequence strand using the exposed DNA template as a guide.

c. **Termination**: once the RNA polymerase reaches the end of the gene, it encounters a termination signal, which signals the end of transcription. The termination signal causes the RNA polymerase to release the synthesized RNA molecule and dissociate from the DNA template. This termination is organized by a series of specific molecular actors, allowing to modulate and regulate this specific step.

1.2.2 Translation

The mRNA molecule then undergoes translation. It carries the genetic information from the DNA to the ribosomes, where *translation* takes place.

In bacterial cells, ribosomes are large molecular complex composed of two subunits, the small 30S subunit and the large 50S subunit, which combine to form the complete 70S ribosome, which is responsible for the translation of mRNA into proteins in bacterial cells. Both subunits contain several number of ribosomal RNAs (rRNAs) and around 52 proteins, what makes of the ribosome the most demanding macro-molecule to produce in terms of resources.

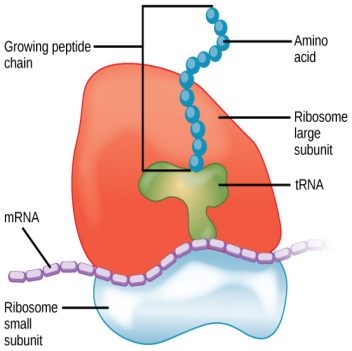


Figure 1.1: Ribosome during translation phase

a. **Initiation**: The *translation* phase starts with the binding of the small ribosomal subunit to the mRNA molecule. The ribosome then scans the mRNA searching for a specific start codon (in general AUG).

b. **Elongation**: Once the start codon is located, the ribosome recruits the large ribosomal subunit and starts the elongation step. During elongation, the ribosome reads the mRNA sequence using the genetic code which is a sequence of 3 letter combinations called *codons*, each of them corresponds to a specific amino-acid. The amino acids are carried to the complex ribosome and mRNA by a macro-molecule called *transfer* RNA (tRNA).

c. **Termination**: The process of elongation continues until the ribosome reaches a stop codon. But sometimes, it can either release the messenger RNA or proceed with the translation of the next gene.

The resulting protein is then folded into its proper shape and can perform its specific function within the bacterium. Overall, the translation phase of protein production in bacterial cells is a highly regulated and sophisticated process that requires the coordinated interaction of many molecular components. The accuracy and efficiency of this process are essential for the proper function of bacterial cells.

Figure 1.2: Protein production: two-step process

1.3 Exponential and Stationary Phases

Bacterial cells experience different regimes during their growth (see Alberts et al. [3] for details). Within the scope of this thesis, our attention will be di-

rected toward the individual cell scale. We will focus on two main regimes: the *exponential* phase and the *stationary* phase. The exponential phase with large growth rate is characterized by an abundance of nutrients (resources) and favorable environmental conditions, allowing the bacterium to grow and reproduce quickly (the production rate of proteins is large for example). However, the stationary phase is when resources become depleted and bacterial growth slows or stops entirely. In this phase, the bacterium enters a state of low metabolic activity because of the environmental conditions.

When the concentrations of different resources in the medium are large enough for some time, the bacterium has the ability to use them efficiently, via its complex regulatory system, to reach a stable exponential growth regime with a fixed growth rate. The growth of a bacterial population in a given medium leads therefore to an active consumption of resources necessary all the molecular components to duplicate the cell.

When resources are scarce, for example when some amino acids are missing, a bacterial cell can adapt, to either exploit differently the available resources or to do without some of them. For *E. coli* or *B. subtilis*, these bacteria use in priority resources maximizing their growth rate. In the context of this adaptation, and for reasons related to the decay of resources, each bacterial cell has to decrease its growth rate, and finally to ultimately stop its growth.

The regulatory network involved in the management of the growth rate to adapt to the environment is complex. In general, the bacterium modifies the concentration of agents in charge of protein production: number of ribosomes, concentrations of proteins in the metabolic network, transporters, ...

In simple terms, when a bacterium needs to survive in a resource-scarce environment, it has to rely on alternative solutions. This, however, means taking away some of the resources it used to use for its growth and making these machines. Because a large part of the system is designed to generate enough spare parts for "copying" the bacterium, this shift in focus reduces the flow of these spare parts that used to be available for everyone. As a result, it reduces the number of each machine that can be made within a certain time frame and decreases the overall quantities of each machine that can produce all the spare parts.

In a first, simplified, description, the decay of a specific resource in the environment leads to a move to a state of the cell where concentrations of several components have been adapted. To study the transition between growth phases, we have chosen to focus on the action of a small RNA, 6S RNA, which plays an important, even essential, role in this domain. Note that, even if this mechanism is central, this description of the transition between growth regimes is nevertheless a simplification in our approach, since the bacterial cell has different ways to modify the steady-state level of its components.

1.4 Mechanisms of Regulation

Regulatory mechanisms are critical for bacterial cells to adapt to changing environments (in terms of availability of resources) and maintain an efficient cellular function. Many studies have been conducted on this topic, such as in Agustino and Collado-Vides [1] where the focus is on examining regulation in *E. coli*.

It is not an overstatement to claim that the intricate biological processes involved in protein synthesis, starting from its DNA sequence, can be precisely regulated to control protein levels. Attempting to provide a comprehensive and concise overview of all potential mechanisms for gene expression regulation is an impractical task. One way the cell regulates its transcriptional processes involves the action of transcription factors, which have the ability to either accelerate or inhibit the cell's activity. This mechanism was investigated by David J. Lee and colleagues (as documented in the study by David J. Lee et al. [18]).

In addition to transcriptional regulation, gene expression control encompasses translational regulation, to modulate protein synthesis. The regulation of gene expression not only controls transcription but also governs translation processes. Translational regulation involves impeding access to the initiation site of mRNA. See Claudio O. Gualerzi [16], thereby influencing the efficiency and timing of protein production.Regulatory mechanisms are critical for bacterial cells to adapt to changing environments (in terms of availability of resources) and maintain an efficient cellular function. Many studies have been conducted on this topic, such as in Agustino and Collado-Vides [1] where the focus is on examining regulation in *E. coli*.

Overall, the complex regulatory mechanisms in bacterial cells ensure that gene expression is tightly controlled and coordinated, allowing the cell to respond to environmental cues and carry out essential functions for survival.

1.5 Transcriptional Regulation

In bacterial cells, transcriptional regulation occurs primarily at the level of initiation, where RNA polymerase binds to the promoter region of a gene and begins the process of transcription. An important mechanism of transcriptional regulation is the use of alternative sigma factors, which can direct RNA polymerase to different sets of genes in response to specific environmental signals.
One of the major post-transcriptional regulators of gene expression is a macromolecule called small RNA (sRNA). The sRNAs are non-coding RNA molecules typically ranging from 50-500 nucleotides in length, that act by base pairing with target mRNAs or proteins to modulate gene expression. Thereby, they influence various cellular processes including metabolism and stress response (see [52]). Numerous sRNAs have been identified and characterized: for example, more than ten sRNAs are known to be encoded in the *E. coli* (represented in Montzka Wassarman et al. [51]) in bacterial genomes, nevertheless the exact number of sRNAs is still not known for other several bacterial cells (see [34]).

The study of sRNAs has led to a greater understanding of bacterial gene regulation and has been used in the development of novel antibiotics and biotechnological tools. This topic has become an important subject to study and analyze, so they started to test the use of sRNAs in the regulation using mathematical models (see for example [23]).

1.5.1 6S RNA: A global regulator of transcription

We will focus on the *6SRNA* regulatory mechanism. 6S RNA is a small RNA discovered in the late 1960*s* in *E. coli* (see Wassarman and Storz [68] for reference) because of its large number during the stationary phase.

Its function had been unknown for a long time until its role is to bind to σ^{70}-RNA polymerase has been shown. The findings of Wassarman and Storz [68] indicate that the association between 6S RNA and RNA polymerase occurs in a precise and effective manner. An observation of 6S RNA making direct contact with the σ^{70} subunit of the polymerase has been done. In addition to that, it was also observed that the stable association between σ and core RNA polymerase only occurs when 6S RNA is present.

In order to understand how the binding of 6S RNA to RNA polymerase is done, the 3D structure of 6S RNA has been investigated. In Chen et al. [15], it was shown that a 6S RNA imitates a DNA promoter, enabling the regulation of transcription by binding and therefore sequestering a free polymerase(see scheme below).

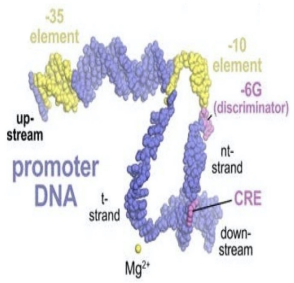


Figure 1.3: Shown is just the promoter DNA. Taken from Chen et al. [15]

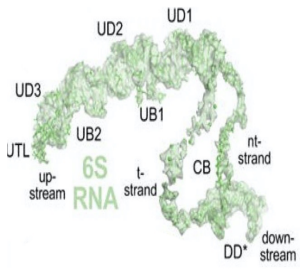
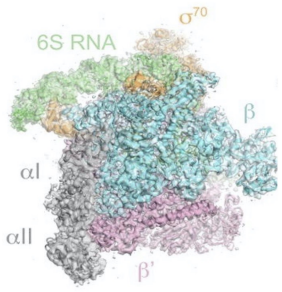


Figure 1.4: Shown is just the model of 6S RNA. Taken from Chen et al. [15]

This is what makes it able to bind to a free polymerase and prevent it from binding to a gene promoter (see scheme).



We call this mechanism, *Sequestration of free polymerases by 6S RNAs*

Figure 1.5: σ^{70}-RNA polymerase bound to a 6S RNA. Adapted from Chen et al. [15]

It plays a crucial role in regulating gene expression during the transition from exponential to stationary phase. During the exponential phase, bacterial cell grows rapidly, and it requires a high level of gene expression to support its growth, the production rate of proteins is therefore large. However, as the cell enters the stationary phase, the availability of nutrients decreases, and the cell begins to experience stress. In response to this stress, 6S RNA accumulates and attains a peak concentration of around 10000 copies per cell and are predominantly bound to σ^{70}-polymerases (detailed by Nitzan et al. [54] and KM [43]). The binding of 6S RNA to RNA polymerase prevents the enzyme from binding to DNA and transcribing genes, leading to a global down-regulation of gene expression.

This mechanism has important consequences on the activity of the cell, as it preserves energy and resources that would otherwise be used for gene expression. In addition, it allows the bacterium to switch its metabolism to alternative conditions in order to be convenient to the stationary phase.

The transition from exponential to stationary phase is not really well understood and many studies have been done in order to characterize it. The simulations shown in the figures below demonstrate that the regulatory mechanism controlled by 6S RNA can endure moderate changes in the affinity between 6S RNA and RNA polymerase.

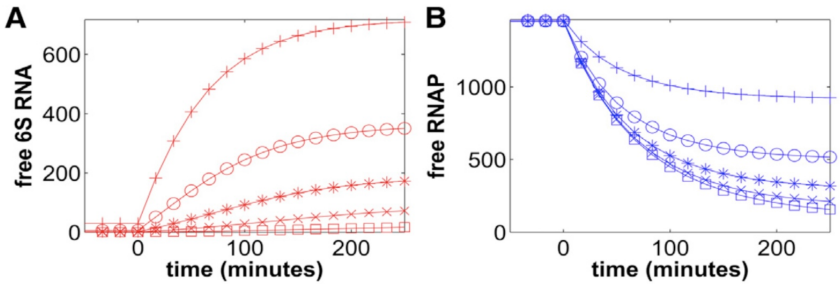


Figure 1.6: The dynamics of the 6S RNA regulation system components at the transition from stationary to exponential phase. Shown are the dynamics analysis results by number of molecules for 6S RNA *(A)*, and free RNA polymerases *(B)*. Adapted from Nitzan et al. [54].

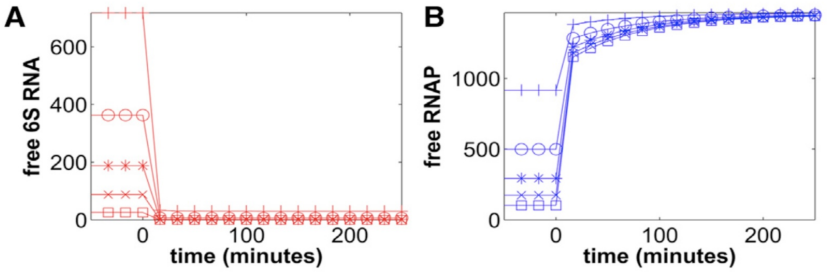


Figure 1.7: The dynamics of the 6S RNA regulation system components at the transition from exponential to stationary phase. Shown are the dynamics analysis results by number of molecules for 6S RNA *(A)*, and free RNA polymerases *(B)*. Adapted from Nitzan et al. [54].

In summary, the 6S RNA regulation in a bacterium is a critical mechanism modulating gene expression during the transition from exponential to stationary phase. It highlights the sophisticated and complex molecular machinery that a bacterial cell uses to survive stressful conditions and challenging environments.

1.6 Translational Regulation

The initiation of mRNA translation is a crucial step in controlling the quantity and accuracy of protein synthesis. It serves as a key point of regulation for gene expression through various post-transcriptional mechanisms. The process starts with the creation of an unstable 30S pre-initiation complex comprising initiation factors (IFs) IF1, IF2, and IF3, along with the translation initiation region of an mRNA and initiator fMet-tRNA. This complex then binds with the 50S subunit, forming a 70S complex (detailed in Gualerzi CO [35]).

1.6.1 The Stringent Response: A regulatory mechanism in bacterial translation

The stringent response has been extensively studied in the model bacterium *E. coli* since the late 1960s. A notable observation in recent years is that although the stringent response has been extensively studied in *E. coli*, the molecular elements involved in its implementation are specific to this bacterium. In reality, in the majority of sequenced and annotated bacteria, the molecular players participating in and defining the stringent response are slightly different from those in *E. coli*.

Although the molecular players may differ, indicating specific biological implementation, the general principles and major actions integrated into our study are conserved across bacteria. Specifically, we refer to the research and results concerning the stringent response in another model bacterium, *Bacillus subtilis*, which belongs to the Gram-positive group (while *E. coli* is a model bacterium for Gram-negative bacteria). For *B. subtilis* and for most bacteria, the molecular organization of the stringent response and its actions on the bacterium vary slightly (which we will discuss further in the final parts of the chapter).

As we will discuss later, the primary objective of the regulatory system that generates the stringent response in bacteria is to ensure a proper balance between the availability of each amino acid and the demand associated with the production of proteins. However, evolution has also exploited this mechanism to address other issues, which we will not consider here. These additional aspects often involve the production or degradation of ppGpp and are directly related to growth rate management rather than amino acid level control during translation.

This regulation is mediated by one protein in *B. subtilis* or two proteins in *E. coli*. The first protein is capable of producing a specific metabolite called pppGpp or ppGpp (referred to as (p)ppGpp hereafter). When the availability of amino acids falls below the bacterium's current protein production needs, this first protein, called RelA in *E. coli*, is complemented by a second protein named SpoT. The second protein degrades (p)ppGpp under specific conditions that activate its degradation. Generally, in most bacteria, the first protein that produces (p)ppGpp (RelA in *E. coli*) also possesses a secondary function to degrade it. Therefore, this protein operates in a dual manner: when it is not activated to produce (p)ppGpp, it can degrade it. Lastly, although this protein is structurally similar to *E. coli*'s RelA, it has recently been named Rel, highlighting its differences from RelA, the protein in *E. coli*.

The presence of (p)ppGpp in the cell triggers a series of downstream effects. It inhibits the production of ribosomal RNAs (rRNAs), which are essential components of the ribosomes responsible for protein synthesis. This reduction in ribosomal RNA production limits the availability of functional ribosomes in the cell, leading to a slowdown in protein synthesis, what helps in conserving energy and resources under nutrient-limiting conditions. In addition, (p)ppGpp interacts with the initiation phase of translation, specifically targeting the formation of the initiation complex. By inhibiting the initiation of translation, (p)ppGpp prevents the binding of aminoacyl-tRNAs to the ribosome, disrupting the in-

corporation of amino acids into growing polypeptide chains. Therefore, it slows down or halts the translation process, leading to a decrease in the overall rate of protein production.

The stringent response, mediated by (p)ppGpp, allows bacterial cells to adapt and respond to adverse environmental conditions by adjusting the rate of protein production in response to environmental conditions, nutrient availability, and other cellular signals. Amino acid starvation poses also a significant risk as it can lead to an increased error rate during protein synthesis. This highlights the critical importance of translational regulation in bacterial cells. By modulating the initiation of translation, bacteria can effectively manage protein synthesis, ensuring accuracy and efficiency, even under conditions of amino acid scarcity.

In *E. coli*, SpoT plays a pivotal role in cellular metabolism and stress response. Also known as (p)ppGpp synthetase/hydrolase, SpoT is an enzyme involved in the stringent response, a regulatory mechanism enabling bacteria to adapt to nutrient scarcity and environmental stressors. SpoT assumes the responsibility of synthesizing and degrading (p)ppGpp signaling molecules, which function as global regulators of gene expression. Through its modulation of (p)ppGpp levels, SpoT exerts influence over diverse cellular processes such as transcription, translation, and metabolism. This capability aids the bacterium in surviving adverse conditions. Conversely, in *B. subtilis*, it is Rel that carries out the production and degradation of (p)ppGpp.
However, the models presented in this thesis are applicable to various bacterial cell types, transcending their inherent differences.

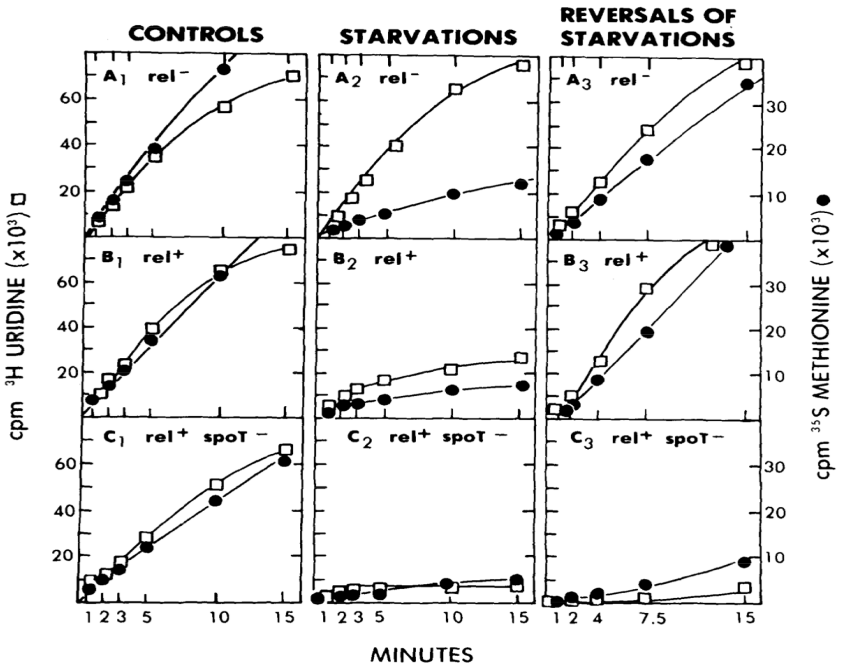


Figure 1.8: Effects of *RelA* and *SpoT* mutations and Amino Acid starvation on the rates of Methionine (•) and Uridine (□) incorporation in *E. Coli.* Adapted from O'Farrell [55].

The experiment represented in the figure 1 aims to distinguish the cell's state and activity under three distinct conditions. The first condition represents normal cellular functioning, while the second condition simulates amino acid starvation. Lastly, the third condition represents the state after readdition of the deficient amino acid. The specific amino acid of interest in this experiment is isoleucine, as it is commonly utilized in protein formation. Throughout these conditions, the availability of methionine and uridine in the medium is adequate, ensuring that the two quantities measured in the experiment are not generally affected by their availability during the course of the experiment. However, their incorporation levels are measured.

The figure represented in this experiment illustrates the impact of amino acid withdrawal and subsequent readdition on the strains *rel- spoT-, rel+ spoT+* and *rel+ spoT-* [1]. Specifically, it is observed that only the *rel+* strains accumulate ppGpp in response to starvation, while the *spoT-* strains exhibit heightened stability of ppGpp, knowing that spoT is not degrading ppGpp.

[1]The + and - symbols associated with gene names identify different strains. A strain name with the gene name associated with a + contains the gene, while a strain name with the gene name associated with a - no longer contains this functional gene. This convention is used in biological articles where manipulated strain-names have a name that explains their main characteristic

This experiment provides valuable insights into the role of ppGpp in cellular activity. Firstly, in terms of the Transcription phase (RNA synthesis), uridine incorporation is measured as a proxy. It is evident that RNA synthesis is inhibited during amino acid starvation solely in *rel+* strains that accumulate ppGpp. Hence, ppGpp acts as an inhibitor of stable RNA synthesis in the absence of amino acids.

Following the readdition of the deficient amino acid, the rate of uridine incorporation swiftly recovers in *rel+ spoT+* strains, where ppGpp degradation occurs due to the presence of spoT. Conversely, in *rel+ spoT-* strains, the increase in uridine incorporation transpires at a slower pace since ppGpp remains active, thereby sustaining the inhibition of transcription.

Regarding the impact of ppGpp on the Translation phase (protein synthesis), protein synthesis rate is measured by methionine incorporation, considering that methionine is the initial amino acid incorporated into the ribosome during protein synthesis. It is noteworthy that amino acid starvation inhibits protein synthesis in both *rel+* and *rel-* cells, which is logical since proteins are composed of amino acid sequences, leading to decreased production regardless of ppGpp presence. Upon the readdition of the deficient amino acid, the rate of protein synthesis experiences a gradual increase in *rel+ spoT-* strains due to the accumulation of ppGpp without spoT-mediated degradation. This highlights that the accumulated levels of ppGpp play a substantial role in the substantial reduction of protein synthesis, resulting in an 86% inhibition.

The regulatory mechanism varies across different bacterial types with regards to how (p)ppGpp influences both transcription and translation phases of protein production. For instance, in *E. coli*, ppGpp binds to the polymerase, leading to the inhibition of ribosomal RNA transcription. Conversely, in *B. subtilis*, GTP governs the polymerases, and ppGpp, in turn, regulates GTP. This implies that in *B. subtilis*, ppGpp indirectly inhibits transcription by controlling GTP levels.

However, the pivotal observation is that the inhibitory effect of ppGpp on translation remains consistent across all bacterial species. This effect is primarily manifested through its impact on IF2, the initiation factor of translation. Consequently, despite the differences in regulatory mechanisms, the variations observed do not undermine the applicability of our model in this chapter. As our study focuses on the influence of ppGpp on the translation phase, our model remains pertinent for diverse bacterial types. (see al. [2] for further details.)

1.7 Probabilistic Models

In this section, we present the mathematical models used to investigate regulation mechanisms introduced in the previous sections. Overall, designing a sufficiently rich class of mathematical models, but still tractable, has been challenging, especially for the translation step.

We give an overview of our models in a simplified manner. We will also introduce the mathematical tools used to analyze these models and discuss the challenges we encountered along the way.

1.7.1 Stochastic Modeling

The objective of the thesis is the development of stochastic models to depict the two stages of protein synthesis: Transcription and Translation. Specifically, we aim to understand some particular regulatory mechanisms, detailed in Sections 1.5.1 and 1.6.1, and to analyze their impact on the activity of the cell. To achieve this, we must design models that not only portray each phase in a reasonably accurate manner but also that give a mathematical interpretation.

1.7.1.1 Averaging Principle

For a given phenomenon, it is convenient to classify stochastic processes into two main types: *Fast* processes and *Slow* processes.

Fast processes evolve rapidly over time, the transition rates between different states are large by definition. Fast processes are in general integer-valued stochastic processes in our analyses. By contrast, The dynamic of the evolution of the state of slow processes is $O(1)$, in general.

Occupation measures are an important mathematical object in the averaging principle. It can be described roughly as follows. For a slow process $(X_N(t))$ and a fast process $(Y_N(t))$, the occupation measure (μ_N) associated to $(Y_N(t))$ is given by

$$\langle \mu_N, g \rangle = \int g\left(s, Y_N(s)\right) ds.$$

The averaging principle consists in proving the convergence in distribution of (μ_N) and the representation of its limit is given by Kurtz [45]

$$\lim_{N \to +\infty} \left(\langle \mu_N, g \rangle \right) = \left(\int g\left(s, y\right) \pi_s(dy) ds \right)$$

where (π_s) is a process with values in the space of probability measures. $(X_N(t))$ can be expressed in the form of a differential equation involving the occupation measure of $(Y_N(t))$.

Occupation measures will be used throughout our upcoming chapters, as it will be seen in Sections 1.7.2.4,1.7.3.4 and 1.7.4.4. However, our approach differs from the traditional one because the slow processes were included into the occupation measures for tightness reasons.
See Kurtz [45], Papanicolaou et al. [56] and Freidlin and Wentzell [29] on this topic.

1.7.2 Chapter 2: Regulation Of Transcription

1.7.2.1 Context

In this chapter, our objective is to analyze the impact of a regulatory process of the Transcription phase. This process involves the 6S RNA macromolecule described in Section 1.5.1, and only polymerases associated with a specific sigma factor: the σ^{70} factor for *E. Coli* (or σ^{A} for *B. subtilis*). In the following, we will

only mention the polymerase for the sake of simplicity instead of the polymerase associated with this single sigma factor .This regulation mechanism is illustrated by the sequestration process of polymerases by 6S RNAs in our model. It is clear that the polymerase plays a pivotal role in the transcription process, thus forming the central focus of our investigation. Throughout this chapter, we will present a mathematical model designed to effectively capture the distribution of polymerases across various states: free, sequestered (bound to a 6S RNA) or in transcription phase. The details of this model will be explored further.

Our starting assumption is that the total number of polymerases remains constant, N. Additionally, we assume that there are $C_m^N \approx c_m N$ different types of mRNAs, and that there are J distinct types of rRNAs. And lastly, we assume that the maximal number of polymerases in elongation phase of an rRNA of type j, $C_{r,j}^N$, is of the order of N.

To understand the allocation of polymerases in this context, we have introduced a Markovian model (dynamics represented in relations (2.8), (2.9) and (2.10)).

$$(F_N(t), S_N(t), M_N(t), Z_N(t), ((U_{j,N}(t), R_{j,N}(t)), 1 \leq j \leq J))$$

where

— $F_N(t)$, the number of free polymerases.

— $S_N(t)$, number of sequestered polymerases.

— $M_N(t)$, number of polymerases in mRNA transcription.

— $Z_N(t)$, number of free 6S RNAs.

— $U_{j,N}(t) \in \{0, 1\}$, to indicate if a polymerase is bound to the promoter of the rRNA of type j or not.

— $R_{j,N}(t) \in \left\{0, \ldots, C_{r,j}^N\right\}$, number of polymerases in elongation phase of an rRNA of type j.

With these notations, the conservation of mass for the ribosomes gives the relation:

$$F_N(t) + S_N(t) + M_N(t) + \sum_{j=1}^{N} (U_{j,N}(t) + R_{j,N}(t)) = N, \forall t \geq 0$$

The figure below shows the model with the distinct states of polymerases along with their corresponding transition rates.

We will now give a brief explanation of the model.

Starting with the transcription process of rRNAs. As mentioned earlier, there are J distinct types of rRNAs, each associated with its own promoter. The transcription of rRNAs occurs through a two-step process. The initial step, referred to as initiation, involves a polymerase binding to a specific promoter. Following this, the elongation phase begins. Once a polymerase is bound to the

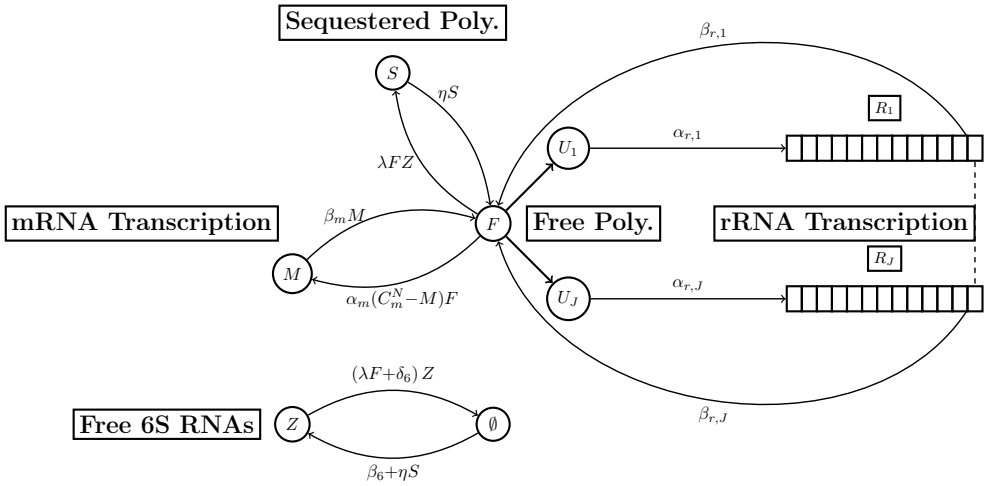


Figure 1.9: Polymerases: Transcription of mRNAs/rRNAs and Sequestration

promoter of an rRNA of type j, it initiates the elongation phase at rate $\alpha_{r,j}$. During this phase, the polymerase actively collects nucleotides at rate $\beta_{r,j}$. This process leads to the formation of an mRNA and subsequently, the polymerase is released from the gene and becomes free again.

It is assumed that there are C_m^N different types of mRNAs as mentioned earlier. A polymerase starts transcribing an mRNA at rate α_m and an mRNA is released at rate β_m and the corresponding polymerase is released at that instant. The process of transcribing mRNA appears simpler in our model compared to the transcription of rRNA. This simplicity arises from the fact that we have combined the initiation and elongation phases into a single step for mRNA transcription. Our primary focus, discussed in Section 1.7.2.2, is to examine the influence of sequestration on rRNA transcription.

A 6S RNA is created at rate $\beta_6 > 0$ and it is in a free state when it's not bound to a polymerase. A given free 6S RNA is degraded at rate $\delta_6 \geq 0$. Only a free 6S RNA can be degraded.

A polymerase is free when it is not bound to a gene or to a 6S RNA. A free polymerase is bound to a given 6S RNA at rate λ. A complex polymerase-6S RNA breaks into a free polymerase and a free 6S RNA at rate η.

1.7.2.2 Objective

In this chapter, we aim to analyze the efficiency of the transcriptional regulatory mechanism outlined in section 1.5.1.

Throughout our study, we work with a large number of polymerases denoted as N (between 2000 and 10000 in E. Coli). In the context of transcription, the purpose is to understand how the bacterium manages its resources (poly-

merases). This involves analyzing how the activity of polymerases between free, actively transcribing (mRNA or rRNA), and sequestered. Our analysis takes into account two key regimes of the bacterial life cycle: the exponential phase and the stationary phase, presented in Section 1.3.

During the stationary phase, defined by a shortage of resources, free polymerases become inactive by the sequestration of free polymerases by macromolecules called 6S RNAs. Therefore, in order to measure the efficiency of this regulatory mechanism, we focus on analyzing the rate of sequestration in each phase, and more specifically the time evolution of the process of the number of sequestered polymerases ($S_N(t)$).

Through this analysis, we have proved the importance of certain parameters and evaluated their impact. Among these parameters, the rate of initiation of transcription of an rRNA of type j, ($\alpha_{r,j}$), holds a significant role. This is because altering the rate of rRNA transcription has an influence on the activity of the cell. As it is showed earlier in figure 1.2.2, ribosomes are the most resource intensive macromolecules, composed of around 52 proteins and numerous ribosomal RNAs (several thousands of nucleotides). Their production occurs when a large number of resources is available in the environment, which corresponds to the exponential phase. Hence, modifying the rate of initiation of rRNA transcription, $\alpha_{r,j}$, plays a pivotal role in transitioning the cell from exponential growth to the stationary phase. Thus, it essentially dictates the phase in which the cell operates.

1.7.2.3 Contributions

We introduce two sets of conditions on the parameters of our model, which define the exponential regime and stationary regime. And we present in this section the outcomes obtained in our study.

a. *Exponential Regime.* It is defined by the condition

$$\min_{1\leq j\leq J} \frac{\alpha_{r,j}}{\beta_{r,j}} > 1.$$

The initiation rate $\alpha_{r,j}$ of the transcription of an rRNA of type j is greater than its production rate.
We proved the convergence in distribution

$$\lim_{N\to+\infty} \left(\frac{\|R_N(t)\|}{N}, \frac{M_N(t)}{N} \right) = (c_r, 1-c_r)$$

where $\|R_N(t)\| \overset{def}{=} \sum_{j=1}^{J} R_{j,N}(t)$.
In this regime, most of the polymerases are actively engaged in the transcription phase, primarily synthesizing rRNAs ($\|R_N(t)\|$) or mRNAs ($M_N(t)$). There are relatively few polymerases that are either free or sequestered by a 6S RNA molecule.

b. *Stationary Regime.* It is defined by the condition

$$\max_{1\leq j\leq J}\frac{\alpha_{r,j}}{\beta_{r,j}}<1.$$

The initiation rate $\alpha_{r,j}$ of the transcription of an rRNA of type j is less than its production rate.
We proved the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{M_N(Nt)}{N},\frac{F_N(Nt)}{N},\frac{S_N(Nt)}{N}\right)=\left(c_m,\overline{f}(t),1-c_m-\overline{f}(t)\right)$$

where $\left(\overline{f}(t)\right)$ is the solution of an ODE.
In this regime, limited number of polymerases engaged in the transcription of rRNAs, in the sense that $\|R_N(t)\|$ is of the order of $O(1)$. Interestingly, a fraction of these polymerases remains free, indicating that the sequestration process does not control all of these seemingly "idle" polymerases. sequestration process does not control all "useless" polymerases.
The fact that sequestration phenomenon of a fraction of N polymerases occurs on the time scale $t\mapsto Nt$ is intuitive, given that the rate of creation of 6S RNAs, β_6, is constant.

1.7.2.4 Methods and Technical Difficulties

In a probabilistic context, using a Markovian model, and with the total number of polymerases denoted as N as our scaling factor, we investigate how the number of polymerases in each state evolves over time when N is large.

We derive functional limiting results, with respect to the scaling parameter, of the time evolution of several stochastic processes. The slow and fast processes are dependent on varying parameter conditions.
An important feature of our model is that the transition rates of the main Markov process exhibit a quadratic dependence on the state of the process, due to the use of the law of mass action for the dynamic of our model.

We will now outline the slow and fast processes within each of the scenarios discussed in section 1.7.2.3.

In Scenario (a), the slow processes are the scaled number of polymerases in rRNA(resp. mRNA) transcription phase $(R_{j,N}(t)/N)$ for all $1\leq j\leq J$ (resp. $(M_N(t)/N)$). The fast processes are the number of free polymerases $(F_N(t))$, the number of sequestered polymerases $(S_N(t))$ and the number of free 6S RNAs $(Z_N(t))$.
In order to prove our results in this regime, we introduce a coupling to study the occupancy of the places for transcription of rRNAs. We then prove an averaging principle. The corresponding occupation measure is given by

$$\langle\mu_N,g\rangle=\int_{R_+}g\left(s,F_N(s)\right)ds.$$

In Scenario (b), the slow processes are the scaled number of free, sequestered, and in mRNA transcription $(F_N(t)/N)$, $(S_N(t)/N)$ and $(M_N(t)/N)$. And the

fast processes are the number of polymerases in rRNA transcription ($R_{j,N}(t)$) for all $1 \leq j \leq J$ and the number of free 6S RNAs ($Z_N(t)$).
In this scenario, we employ a coupling method and an averaging principle. The key component in this context is the associated occupation measure, which is expressed as:

$$\langle \mu_N, g \rangle = \int_{R_+} g\left(s, \frac{F_N(Ns)}{N}, G_N(Ns), Z_N(Ns)\right) ds$$

where $G_N(t) \stackrel{def}{=} C_m^N - (N - F_N(t) - S_N(t))$.
What is quite unusual, is that, in the definition of the occupation measure, we include the slow process ($F_N(Nt)/N$) as well. The reason behind this choice lies in the fact that proving the tightness of ($F_N(Nt)/N$) is not straightforward. Due to the fast time scale, proving that the martingale component of ($F_N(Nt)/N$) converges to zero is not clear.

1.7.3 Chapter 3: Pairing Mechanisms

1.7.3.1 Context

Regulation mechanisms, in essence, function by engaging with the macro-molecules of the cell. We refer to this interaction between macro-molecules as *Pairing Mechanisms* within the context of our study. These Pairing Mechanisms play a crucial role in the cell as they control the growth rate based on the environment of the cell. They can either decrease the usage of specific macro-molecules or accelerate their activity. Small RNAs (sRNAs) are the global regulators, as they have various effects on the functions of the cell, depending on its surroundings (resources availability). They can reduce the usage of specific macro-molecules, a phenomenon termed *Sequestration* in Chapter 1.7.2, or they can enhance activity. The binding between these molecules eventually breaks due to thermal noise after a certain period of time. The cell produces agents on a regular basis and when the cell environment is normal and does not require regulation, free agents are degraded quickly resulting in only a few of them remaining. Agents that are paired with particles cannot be degraded.

In the previous chapter 1.7.2, a sequestration process, which acts as a pairing mechanism, has been investigated. However, in this chapter we analyze it differently. Essentially, we separate it from the transcription processes outlined earlier and we focus only on the sequestration process. The key distinction here is that we approach this topic from a more general perspective. Unlike Chapter 1.7.2, where we primarily dealt with a single type of particles which were the polymerases and agents 6S RNAs, in this chapter, a total of J different types of particles is considered.
Assuming that the environment has limited resources, the efficiency of the regulation is measured by the number of paired particles (sequestered in Chapter 1.7.2).

To do this, we introduce a mathematical model describing the interaction between different types of particles and specific particles, denoted as agents.

We assume that there are a constant total number of particles, N. Among these particles, there are J different types of particles and a total number, $C_{j,N} \approx c_j N$, of particles of type j. Thus $N = C_{1,N} + \cdots + C_{J,N}$.
There is just one type of agent in our model.

To investigate the sequestration process in a general context, a Markovian model has been introduced (with dynamics represented in relations (3.7) and (3.8)),

$$(X_N(t)) \stackrel{def}{=} (F_N(t), Z_N(t)) \stackrel{def}{=} ((F_{j,N}(t), j = 1, \ldots, J), Z_N(t))$$

where

— $F_{j,N}(t), j \in \{1, \ldots, J\}$, the number of free particles of type j.

— $Z_N(t)$, number of free agents.

And we denote by $S_{j,N}(t) \stackrel{def}{=} C_{j,N} - F_{j,N}(t)$ the number of paired particles of type j.

With these notations, the conservation of mass for the particles gives the relation:

$$F_{j,N}(t) + S_{j,N}(t) = C_{j,N}, \ \forall t \geq 0, \ \forall j \in \{1, \ldots, J\}$$

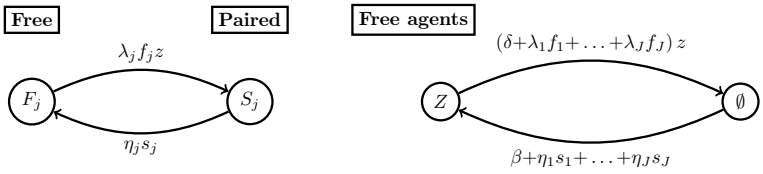


Figure 1.10: Transitions of Pairing Mechanisms

We will now provide a brief overview of the model depicted in the figure.

— A couple of free particle of type j and free agent binds at rate λ_j to form a pair. And since there are $f_j z$ possible combinations of free particle/free agent, this process exhibits a quadratic dependence, resulting in a rate of $\lambda_j f_j z$.

— A pair of particles of type j and an agent split at rate η_j and gives a free particle of type j and a free agent.

— An agent is created at rate β and a free agent is degraded at rate δ. An agent paired with a particle cannot die.

1.7.3.2 Objective

In this chapter, our goal is to explore the process $(X_N(t))$ involving the interaction of various types of particles and agents within a cell's environment. These interactions are highly dependent on the specific conditions present in the cell.

Throughout our study, we work with a large number of particles, N, which is the scaling parameter. Our goal is to examine the impact of the creation and degradation of agents on the activity of the cell, as described in Section 1.7.3.3.

The efficiency of a pairing mechanism is discussed in this chapter in a general way. We focus on analyzing the rate of pairing, more specifically, we investigate the time evolution of the number of paired particles $(S_{j,N}(t))$ of type j for all $j \in \{1, \dots, J\}$, knowing that the goal of regulation mechanism is to maximize this number.

1.7.3.3 Contributions

Investigating the efficiency of these pairing mechanisms led us to examine two distinct situations. First, we looked at a situation where there is a fixed number of agents, of the order of N, and there is no creation or degradation of agents. Second, we explored a scenario where agents can be created and degraded.

In the scenario with a fixed number of agents, we studied two cases. One where the number of agents, $rN, r < 1$, is strictly less than the total number of particles, N. Therefore we analyzed three different scenarios in our model. Each is characterized by a specific condition regarding the number of agents present.

a. *No creation nor degradation of agents*: Fixed total number of agents of the order of $rN, 0 < r \leq 1$:

   (a) $r < 1$: In the scenario where there are less agents than particles, the number of free particles of type j, $(F_{j,N}(t))$ is of the order of N, for all $j \in \{1, \dots, J\}$. The convergence in distribution

   $$\lim_{N \to +\infty} \left(\frac{F_{j,N}(t)}{N} \right) = (f_j(t)) \,, \ \forall j \in \{1, \dots, J\}$$

   holds, with $(f_j(t))$ is the solution of an ODE.

   (b) $r = 1$ In the scenario where the number of agents is equal to the number of particles, it is shown that the number of free particles of type j is of the order of $\sqrt{N}$ for all $j \in \{1, \dots, J\}$. We proved the convergence in distribution

   $$\lim_{N \to +\infty} \left(\frac{F_{j,N}(t/\sqrt{N})}{\sqrt{N}} \right) = (f_j(t)) \,, \ \forall j \in \{1, \dots, J\}$$

   where $(f_j(t))$ is the solution of an ODE.
   And a central limit theorem is proved. The convergence in distribu-

tion

$$\lim_{N\to+\infty}\left(\frac{F_{j,N}\left(t/\sqrt{N}\right)-\sqrt{N}f_j(t)}{\sqrt[4]{N}}\right)=\left(\widehat{F}_N(t)\right)$$

where $\left(\widehat{F}_N(t)\right)$ is the solution of a SDE. This shows that the fluctuations are of the order of $\sqrt[4]{N}$.

b. *With creation and degradation of agents*: Agents are created at rate β and degraded at rate δ. We start initially with few agents in the medium. An agent creates is paired right away with a particle via the successive steps of pairing and splitting mechanisms, as long as the number of free particles is large enough in order to have the pairing that happens before the degradation of agents.
In order to study the time evolution of the number of free/paired particles in the cell, a timescale $t \mapsto Nt$ must be employed due to the rate of creation of agents β which is bounded.
In this context, the number of free particles remains of the order of N.
Unlike Scenarios (aa) and (ab), it turns out in this case that the sequence of processes $(F_{j,N}(Nt)/N)$ converges in distribution but only in a weak form, via occupation measures. Only the scaled process of the total number of free particles $(\|F_N(Nt)\|/N)$, (where $\|F_N(t)\| = \sum_{j=1}^{J} F_{j,N}(t)$), converges in distribution to a continuous process,

$$\lim_{N\to+\infty}\left(\frac{\|F_N(Nt)\|}{N}\right)=(H(t))$$

where $H(t) \in (0,1)$ is the unique solution of an equation presented in Chapter 4. And the equilibrium point of $(H(t))$ is strictly positive.

Scenarios (aa) and (ab) are employed to explore situations where the environment remains relatively stable and when there is already a significant number of agents in the system. And the Scenario (b) is used to understand the impact of creation and degradation of agents.
We first notice the existence of three different timescales, illustrating the speed at which the pairing mechanism operates in each scenario.
The difference between Scenario (ab) and Scenario (b) shows the impact of the dynamical arrivals and departures of agents. The difference between the results of these two scenarios, in addition to the employed timescale is that in Scenario (b), the fraction of paired particles is strictly less than 1. This indicates that in this case, there is a maximal pairing rate and that the number of free particles is large of the order of N.
In biological terms, the Scenario (b) is more realistic, and our results shed light on the transition between the exponential and stationary phase. Therefore, based on our results, despite the existence of the regulation mechanism, the number of free/useless particles remains large in the medium. In addition, this transition happens at a timescale proportional to N.

1.7.3.4 Methods and Technical Difficulties

In a stochastic context, using a Markovian model, our focus here is on understanding the asymptotic evolution of the number of paired particles in different states, with the total number of particles denoted as N serving as a scaling factor.

In all situations, we establish a first order theorem supported by an averaging principle. The slow and fast processes are dependent on the scenario. Naturally, our methodology involves averaging principles proving convergences in distribution of occupation measures. We also employ coupling methods in our analysis.

The main difficulty was proving convergence in distribution results through the averaging principle, dealing with occupation measures, especially in Scenario (b). This was tricky because we had to include the slow processes in the occupation measure as we were unable to prove its tightness.

We will now outline the slow and fast processes for each scenario discussed in section 1.7.3.3. Additionally, we will present the associated occupation measures.

a. In Scenario (aa), the slow processes are the scaled number of free particles of type j, $(F_{j,N}(t)/N)$ for all $j \in \{1, \ldots, J\}$, and the fast process is the number of free agents, $(Z_N(t))$. The associated occupation measure is, classically, given by

$$\langle \mu_N, g \rangle = \int g\left(s, Z_N(s)\right) ds.$$

b. In Scenario (ab), the slow processes are $\left(F_{j,N}(t/\sqrt{N})/\sqrt{N}\right)$ for all $j \in \{1, \ldots, J\}$. A coupling is proved in this case

$$F_{N,j}(t) \leq F_{j,N}(0) + X_j^N(t)$$

where $\left(X_j^N(t)\right)$ is solution of SDE.

c. In Scenario (b), the slow processes are $(F_{N,j}(Nt)/N)$ for all $j \in \{1, \ldots, J\}$, and the fast process is $(Z_N(Nt))$. The associated occupation measure is given by

$$\langle \mu_N, g \rangle = \int g\left(s, \frac{F_{j,N}(Ns)}{N}, Z_N(Ns)\right) ds.$$

In this case, the definition of occupation measure is extended to include the slow processes in order to overcome the lack of tightness properties of $(F_{j,N}(Nt)/N)$.

1.7.4 Chapter 4: Regulation Of Translation

1.7.4.1 Context

The final chapter is dedicated to the translation phase, a critical stage in the protein production. The ribosome plays a central role in this particular phase. Within this chapter, we introduce a mathematical model that effectively captures how ribosomes are distributed among various states, which will be detailed further.

To simplify things, our model is devised to illustrate the translation of a protein composed of just a single amino acid. However, within the chapter, we also present a model with two different amino acids. In fact, the existence of four different regimes, listed in Section 1.7.4.3, does not depend on the number of amino acids in the model. Adding amino acids in our model, surely will give a more realistic description from a biological point of view but does not seem to have an impact from a mathematical point of view. For this reason, the chapter is essentially devoted to the study of a model with just one amino acid.

We make the assumption that the overall number of ribosomes remains constant, denoted as N, alongside a total number of tRNAs denoted by $C_q^N \approx c_q N$ and a total number of messenger RNAs denoted by $C_m^N \approx c_m N$

In order to study the state of ribosomes in this case, a Markovian model of dimension 5 has been introduced,

$$\left(R_N^F(t), R_N^M(t), R_N^E(t), R_N^S(t), Q_N(t)\right)$$

where

— $R_N^F(t)$, the number of free ribosomes.

— $R_N^M(t)$, number of ribosomes in the initiation step.

— $R_N^E(t)$, number of ribosomes in the elongation step.

— $R_N^S(t)$, number of sequestered ribosomes.

— $Q_N(t)$, the number of charged tRNAs, i.e. carrying an amino acid.

With these notations, the conservation of mass for the ribosomes gives the relation:

$$R_N^F(t) + R_N^M(t) + R_N^E(t) + R_N^S(t) = N, \forall t \geq 0$$

The figure below illustrates the model featuring distinct states along with their corresponding transition rates.

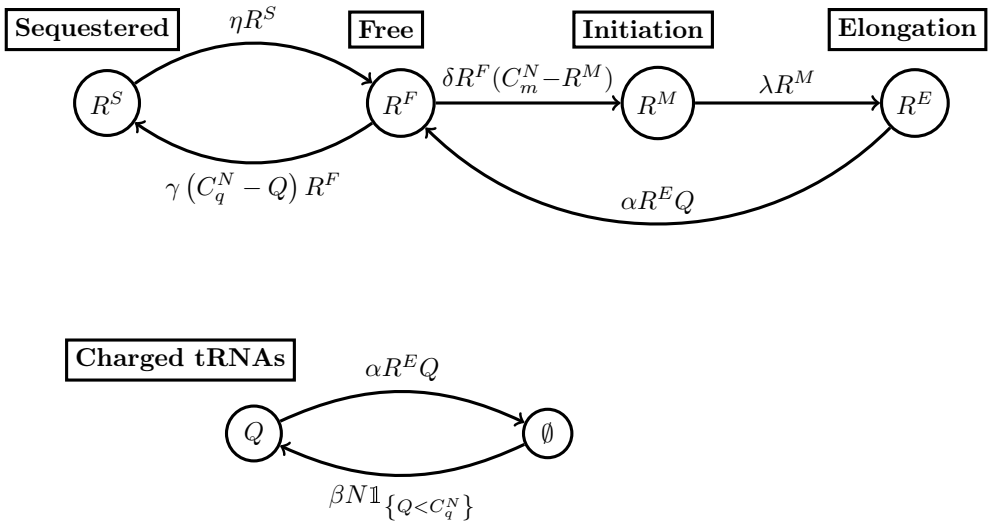


Figure 1.11: Ribosomes: Translation and Sequestration with 1 amino acid

We will now provide a brief overview of the model depicted in the figure (details in Chapter 4).

The initiation step represents the binding of the ribosome to an mRNA. It is assumed that there are C_m^N different types of mRNAs, but all of these various types share the same rate of binding with a ribosome, denoted as δ. The initiation step happens at rate $\delta\left(C_m^N - R^M\right)$. And once an mRNA is bound to a ribosome, the ribosome initiates the elongation step at rate λ and a protein is formed at rate αQ, marking the point when the ribosome detaches from the mRNA.

An empty tRNA carries an amino acid at rate βN and it becomes discharged (empty) when it transfers the amino acid it carries to a ribosome awaiting it in the elongation phase. Therefore, this discharging event occurs at rate αR^E.

Finally concerning the sequestration of free ribosomes, it represents the regulatory mechanism that occurs during amino acid starvation. Hence, the rate of sequestration is proportional to the number of empty tRNAs, denoted as $C_q^N - Q$. The sequestration happens at rate $\gamma\left(C_q^N - Q\right)$ and a ribosome is released and becomes free again at rate η.

1.7.4.2 Objective

In this chapter, we analyze the efficiency of the regulatory mechanism outlined in Section 1.6.1.

Throughout our study, we work with a large number of ribosomes denoted as N (between 10000 and 30000 in E. Coli). Our goal is to understand how

the bacterium manages its resources. In the context of translation, this involves analyzing how the state of ribosomes between free, initiation, elongation and sequestration phase, is affected by the flow of amino acids.

When there are variations in the flow of amino acids, it directly impacts the rate of proteins production. In the situation where there is a shortage of amino acids, ribosomes get blocked in the translation process (elongation more specifically). This prolonged delay can eventually result in selecting incorrect amino acids and thus introducing errors during protein elongation. To avoid errors, ribosomes become inactive (as described in Section 1.6.1). This process is represented as a sequestration process in our model.

In order to analyze the efficiency of the regulatory mechanism, the most important criterion is the rate of sequestration. Therefore, we study the time evolution of the number of sequestered ribosomes $\left(R_N^S(t)\right)$ depending on the cell environment. By doing that, we showed the importance of certain parameters and assessed their impact, among which the rate of loading tRNAs with amino acids (β) certainly plays a significant part. And so under certain conditions on the parameter β we get four different regimes. These regimes and their implications will be detailed in Section 1.7.4.3.

1.7.4.3 Contributions

Identifying the conditions that influence each state we observed led our study into two distinct situations: one where there is a lack of amino acids and another where there is an adequate supply of them. What is driving these situations is how quickly tRNAs are loaded with amino acids, what means that the rate of the charging process of tRNAs with amino acids, β, is the pivotal parameter.

In the case of a shortage of amino acids, the scenario is defined by a certain relationship: $\beta < \lambda$, where λ can be thought of as the rate at which proteins are being produced. On the other side, when there are enough amino acids around, the situation is defined by $\beta > \lambda$.

In the amino acid shortage scenario, we actually proved the existence of two distinct regimes within it. One involved a maximal rate of sequestration, while in the other the sequestration is not fully activated. Therefore, we proved the existence of four different regimes in our model. Each of these patterns is defined by specific conditions related to the availability of resources as well as a condition concerning the maximal sequestration rate.

It should be noted that we have only partial results in this chapter. The scaling results associated to each of the four regimes described below depend on a technical result on some hitting time. Basically it states that, on a "small" time interval, the order of magnitude of the coordinates of the coordinates state variable do not change for some specific class of initial states. This is achieved completely in Section 4.6 for the first regime. The key result is Lemma 51 which relies on a not completely trivial coupling idea. For the other regimes, the corresponding lemma is stated but as a conjecture. These are Lemma 58 of Section 4.7, Lemma 68 of Section 4.8 and Conjecture 74 of Section 4.9.

a. In the scenario where there's a sufficient amount of amino acids available, a large fraction of ribosomes is in the Translation phase, specifically in the Initiation step (whose number is denoted as $\left(R_N^M(t)\right)$), while the tRNAs ($(Q_N(t))$) become saturated. Mathematically, we have shown that there exists $T > 0$ such that the convergence in distribution

$$\lim_{N\to+\infty}\left(\left(\frac{R_N^M(t)}{N}, \frac{Q_N(t)}{N}\right), t \geq T\right) = (1, c_q)$$

holds.
To interpret this, when there's an ample supply of amino acids, it means that ribosomes possess the necessary resources to synthesize proteins effectively. This explains why they're all actively engaged, particularly in the Initiation step of the Translation phase. Additionally, the saturation of tRNAs suggests that there are enough "carriers" for amino acids, reinforcing the productive protein synthesis process.

b. In the scenario where amino acids are lacking and the sequestration mechanism is not active, we have shown that ribosomes predominantly exist in two states: initiation, denoted as $\left(R_N^M(t)\right)$, and elongation, denoted as $\left(R_N^E(t)\right)$. Other states contain only a limited number of ribosomes, and the number of charged tRNAs $(Q_N(t))$ is also relatively low. We proved that for the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{R_N^M(t)}{N}, \frac{R_N^E(t)}{N}\right) = \left(r^M(t), 1 - r^M(t)\right)$$

where $\left(r^M(t)\right)$ is the solution of an ODE.

This outcome is due to the shortage of resources, which results in ribosomes being essentially stuck at the elongation stage, waiting for the necessary amino acids to proceed.

c. In the scenario where the sequestration is present but not fully active, we proved that most of the ribosomes are essentially distributed among three states: initiation $\left(R_N^M(t)\right)$, elongation $\left(R_N^E(t)\right)$ and sequestration $\left(R_N^S(t)\right)$. The remaining states contain only a limited number of ribosomes, and the number of charged tRNAs $(Q_N(t))$ is relatively low. We proved that for the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{R_N^M(t)}{N}, \frac{R_N^E(t)}{N}, \frac{R_N^S(t)}{N}\right) = \left(r^M(t), r^E(t), 1-r_N^M(t)-r_N^E(t)\right)$$

where $\left(r^M(t)\right)$ and $\left(r^E(t)\right)$ are the solutions of two ODEs.
Therefore in this case, ribosomes exist in either the initiation or elongation phase, waiting for an amino acid. However, we also get sequestered ribosomes, which lowers the amount of blocked ribosomes during translation.

d. In this scenario where the sequestration is fully activated, the majority of ribosomes is present in two states: initiation $\left(R_N^M(t)\right)$ and sequestration $\left(R_N^S(t)\right)$. In addition, in this case we have a considerable number of charged tRNAs $(Q_N(t))$. We prove the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{R_N^M(t)}{N},\frac{R_N^S(t)}{N},\frac{Q_N(t)}{N}\right)=\left(r^M(t),1-r_N^M(t),q(t)\right)$$

where $\left(r^M(t)\right)$ and $(q(t))$ are the solutions of two ODEs.
Under these conditions of maximal sequestration, the majority of ribosomes are either engaged in the initiation phase or sequestered, while only a small fraction remains unoccupied or awaiting amino acids. And a substantial number of charged tRNAs, roughly of the order of N, is detected.

The situations described as Scenarios (c) and (d) are actually determined by a specific condition related to the total number C_q^N of tRNAs. This particular condition will be elaborated upon in detail in Chapter 4, and it will be interpreted as the criterion for the maximal sequestration rate.

1.7.4.4 Methods and Technical Difficulties

In a stochastic context, using a Markovian model, and by taking the total number of ribosomes, N, as a scaling parameter, we study the asymptotic behavior of the time evolution of the number of ribosomes in each state. In all situations, we establish a first order theorem supported by an averaging principle. The slow and fast processes are dependent on varying parameter conditions. Naturally, our methodology involves proving convergences in distribution of occupation measures and representing their limits by using Kurtz [45] as a reference point.

We will now outline the slow and fast processes within each of the scenarios discussed in section 1.7.4.3. Additionally, we will present the associated occupation measures.

Regardless of the scenario, the process of the number of free ribosomes $\left(R_N^F(t)\right)$ is always a fast process while the scaled process of the number of ribosomes in initiation phase $\left(R_N^M(t)/N\right)$ is a slow process. The other fast and slow processes depend on the specific scenario.
In Scenario (a) and (d), the other fast process is the number of ribosomes in the elongation phase $\left(R_N^E(t)\right)$ and the slow processes are the scaled number of sequestered ribosomes $\left(R_N^S(t)/N\right)$ and the scaled number of charged tRNAs $(Q_N(t)/N)$.
The occupation measure considered in this case is given by

$$\langle\mu_N,g\rangle=\int_{\mathbb{R}_+}g\left(s,R_N^E(s),R_N^F(s)\right)ds.$$

In Scenario (b) and (c), the other fast process is the number of charged tRNAs $(Q_N(t))$. In both situations, $\left(R_N^E(t)/N\right)$ is a slow process. However, in scenario

(c), $\left(R_N^S(t)/N\right)$ is an additional slow process.
The occupation measure is

$$\langle \mu_N, g \rangle = \int_{\mathbb{R}_+} g\left(s, Q_N(s), R_N^F(s)\right) ds.$$

First, working with a model of dimension five and proving the previously reported convergence results were the technical obstacles in this chapter. Lastly, the other challenge was to interpret the results paying particular attention to the existence of two regimes under the condition $\beta < \lambda$ where there is a shortage of amino acids: one regime with maximal sequestration and the other with partial sequestration.

Chapter 2

Regulation Of Transcription

2.1 Introduction

The central dogma of molecular biology asserts for biological cells that genetic information flows mainly in one direction, from DNA to RNAs, and to proteins. For the two most studied bacteria *Escherichia coli* and *Bacillus subtilis*, production of proteins is a central process which can be described as a process in two main steps. In the first step, macro-molecules *polymerases* produce RNAs with genes of DNA. This is the *transcription* step. The second step, *translation*, is the production of proteins from mRNAs, *messenger RNAs*, with macro-molecules *ribosomes*. See Watson et al. [70].

In bacterial cells, protein production uses essentially most of cell resources: a large number of its macro-molecules such as polymerases and ribosomes, biological bricks of proteins, i.e., amino acids, and the energy necessary to build proteins during the translation step, such as GTP.

In this paper we study an important regulation mechanism of transcription using specific RNA macro-molecules, 6S RNAs, common to a large number of bacteria. See Wassarman [67] for example. The functional property of this RNA is of blocking/sequestering free polymerases from producing RNAs. The general context of this regulation is related to complex mechanisms of the cell to finely tune the production of a large set of RNAs. Let us first recall the three main categories of RNAs:

a. rRNAs, *ribosomal RNAs*, used for the building of ribosomes. A ribosome is a complex assembly of around 50 proteins and, also, of several rRNAs. An rRNA is a long chain of several thousands of nucleotides, it is in particular a costly macro-molecule to produce. Reducing or speeding-up the production of ribosomes, in particular of rRNAs, has therefore a critical impact on resource management of the cell;

b. mRNAs, *messenger RNAs*, used by the translation step to produce a protein from mRNAs coding sequences;

c. A large set of RNAs that do not belong to the two previous categories, such as *transfer RNAs*, tRNAs, or *Bacterial small RNAs*, sRNAs, often associated to regulation mechanisms. This class includes 6S RNAs.

When the concentrations of different resources in the medium are high enough for some time, the bacterium has the ability to use them efficiently, via its complex regulatory system, to reach a stable exponential growth regime with a fixed growth rate. The growth of a bacterial population in a given medium leads therefore to an active consumption of resources necessary to produce new cells.

When resources are scarce, each bacterium of the population can adapt, to either exploit differently the available resources, or to do without some of them, as for example when some amino-acids are missing. For *E. coli* or *B. subtilis*, these bacteria use in priority resources maximizing their growth rate. In the context of this adaptation, and for reasons related to the decay of resources, each bacterial cell has to decrease its growth rate, and finally to ultimately stop its growth.

The regulatory network involved in the management of the growth rate to adapt to the environment is complex. The important point in this domain is that the bacterium has to modify the concentration of most of the agents in charge of it: number of ribosomes, concentrations of proteins in the metabolic network, transporters, ... In a first, simplified, description, the decay of a specific resource in the environment leads to a move to a state of the cell where concentrations of several components have been adapted. To study the transition between growth phases, we have chosen to focus on the action of a small RNA, 6S RNA, which plays an important, even essential, role in this domain. Note that, even if this mechanism is central, this description of the transition between growth regimes is nevertheless a simplification in our approach, since the bacterial cell has different ways to modify the steady-state level of its components.

In this article, we investigate a simplified scenario where transitions occur between two phases: an exponential growth phase and the stationary phase, where the growth rate is equal to 0. The first interest of this scenario lies in the sharp transition of the polymerase management by the cell, via the strong effect of the *stringent response* on the production of rRNAs: the transcription of rRNAs is completely stopped. This is where the action of 6S RNAs is crucial. See Gottesman et al. [33]. Its second interest is experimental since it is possible in practice to create this transition by the addition of a convenient product in the medium of cell populations to induce a stringent response. Our general goal is to investigate if, with this simplified framework, the regulatory system organized around 6S RNA has the desired qualitative properties to ensure a convenient transition between these regimes. In this paper, we analyze the efficiency of the regulation by 6S RNAs with stochastic models. We investigate in particular the time evolution of the activity of polymerases in the cell under different regimes.

2.1.1 A Simple Description as a Particle System

In order to explain the basic principle of the regulation mechanism investigated in this paper, we describe a simplified version in terms of a particle system. Section 2.1.2 describes in more depth and detail the biological context of this class of models.

We consider two types of particles P and $6S$. There is a fixed number of particles of type P and there are random arrivals of particles of type $6S$. A particle of type P can be in three states: busy, idle, or paired with a $6S$ particle. Similarly, a $6S$ particle is either idle or paired. The possible events are:

— an idle, resp. busy, P particle becomes busy, resp. idle;

— a couple of an idle P and an idle $6S$ is paired;

— a pair $P{-}6S$ is broken giving two idle P and $6S$;

— an idle $6S$ arrives/dies.

Note that only an idle $6S$ can die. A statistical assumption is that each couple of free P particle and free $6S$ particle is paired at some fixed rate and each free $6S$ dies at a fixed rate too.

We present a heuristic description of the phenomena we are interested in:

a. If the parameters of the P particles are such that, on average, most of particles of type P are busy. Therefore, few of them are idle, the arriving $6S$ particles will very likely die before they can be paired with a P particle. In this case there will be few $6S$ particles in the system.

b. Otherwise, if, on average, a significant fraction of particles of type P are idle, the arriving $6S$ particles will very likely pair with one of them. In particular, as long as there are many idle P particles, $6S$ will be quickly paired so that few of them will die. In this manner, the dynamic arrivals of $6S$ progressively decrease the number of idle P particles.

A pair $P{-}6S$ is seen as a sequestration of a P particle, the purpose of $6S$ particles is of storing "useless" P particles. The case a) corresponds to the case when most of P particles are efficient so that no regulation is required. This corresponds to the *exponential growth phase* of our biological process. The case b) is when there is a need of sequestration of P particles, this is the *stationary growth phase* of our model.

The nice feature of this mechanism is its adaptive property due to the dynamic arrivals of $6S$: if they are useless, they disappear after some time. Otherwise, as it will be seen, their number builds up until some threshold of sequestration is reached.

The main goal of the paper is of understanding under which conditions on the parameters the cases a) or b) may occur. To assess the efficiency of the regulation mechanism in the case a), we study the time evolution of the number

of $6S$ particles. In the case b), we investigate the number of sequestered P particles to determine the maximal sequestration rate of the regulation.

The model investigated in the paper is in fact a little more complicated in the sense that P particles can be "busy" in two ways: either it remains busy during a random amount of time before being idle again. The other busy state is that it joins a queue where only the particle at the head of the queue becomes idle again after a random amount of time. In our model, this is related to mRNAs and rRNAs production. The next section gives a detailed description of these aspects.

2.1.2 Biological Background

Transcription

In a bacterial cell, a polymerase may be associated to several specific proteins, called σ-factor to form a *holoenzyme* Eσ. In our case we focus on the "housekeeping" σ-factor σ^{70}. This holoenzyme binds to a large set of gene promoters to initialize the transcription. This is the *initiation phase*. If this step is successful, the protein σ^{70} is detached and the polymerase completes the elongation of the corresponding RNA.

This is a simplified description of course. The precise description of the mechanisms are dependent on the bacterium, it is nevertheless sufficiently accurate from our modeling perspective. Throughout the paper we do the slight abuse of using the term polymerase instead of the more biologically correct term holoenzyme. Another important aspect is that the initiation phase may fail due to random fluctuations within the cell, or to a low level of nucleotides needed for the initiation of transcription, i.e. ATP, GTP, UTP, CTP, ... When this happens the transcription is aborted. The level of GTP, for example, has an impact on the modulation of initiation of transcription with respect to the growth rate for bacterium B. subtilis, and, similarly, the level of ppGPpp for bacterium *Escherichia coli*. See Geißen et al. [32] in the case of an rRNA.

Regulation by small RNAs

A subset of RNAs whose sizes in nucleotides is less than 100, small RNAs or sRNAs has been shown to play an important role to regulate gene expression. The first such sRNAs were identified in the late 1960's. See Britten and Davidson [13] and Zamore and Haley [72]. They were shown to turn in or turn off specific genes under convenient conditions.

Among them the sRNA 6S RNA was first discovered because of its abundance in E. coli in some circumstances. See Hindley [37]. This has been one of the first sRNAs to be sequenced. Nevertheless, it took three decades to understand its role in the regulation of transcription.

Experimental studies have shown that 6S RNA acts in fact as a *global regulator* of transcription and not only for the regulation of a reduced subset of genes as most of small RNAs. A 6S RNA has a three-dimensional structure

similar to a DNA promoter, so that the holoenzyme" $E\sigma^{70}$ may be bound to it and is, in some way, sequestered by it. See Cavanagh and Wassarman [14] and Nitzan et al. [54]. It has been shown that during *stationary phases*, when the growth rate is null, the 6S RNAs accumulate to a high level, with more than 10000 copies. During an *exponential phase*, when the growth is steady, the average duration time of cell division is around 40min for E. coli, there are less than 1000 copies. See Wassarman [67] and Steuten et al. [66].

The fluctuations of the number of copies of 6S RNAs is therefore an important indicator of the growth rate of the cell. An important question is to assess the efficiency of the regulation mechanism operated by the 6S RNAs. The impact of several parameters of the cell are investigated: The total number of polymerases, the production rate of 6S RNA and their degradation rate, initiation rates of polymerases for rRNAs and mRNAs and the sequestration rate, i.e. the binding rate of a couple 6S RNA and polymerase.

2.1.3 Mathematical Models

Regulation of gene expression has been analyzed with mathematical models for some time now. See Mackey et al. [46] and also Chapter 6 of Bressloff [12], and the references therein. The lac operon model is one of the most popular mathematical models in this domain, for its bistability properties in particular. See also Dessalles et al. [21].

Specific stochastic models of regulation by RNAs are more scarce. The regulation of mRNAs by sRNAs in a stochastic framework has been the subject of several studies recently. In Kumar et al. [44], Mehta et al. [47], and Platini et al. [59], the authors study regulation mechanisms of mRNAs by sRNAs with a two-dimensional Markov chain for the time evolution of the number of sRNAs and mRNAs. Some limiting regimes of the corresponding Fokker-Planck evolution equations are investigated and discussed. The difficulty is the quadratic dependence on the number of mRNAs and sRNAs. See also Baker et al. [6] and Mitarai et al. [50]. These studies can be seen as extensions of the early works on stochastic models of gene expression, see Berg [9], Elowitz et al. [22] and Rigney and Schieve [60]. See also Fromion et al. [30].

2.1.4 The Main Results

In this paper, we will study the efficiency of the sequestration of polymerases by 6S RNAs. Recall that this is in fact the holoenzyme which is sequestered. We investigate the behavior of several variables associated to the regulation of the transcription phase: Number of free/sequestered polymerases and number of polymerases in the elongation phase of mRNAs and rRNAs.

Technical Challenges

We assume that there are N polymerases with N large. We derive functional limiting results, with respect to this scaling parameter, of the time evolution

of several stochastic processes. An important feature of our model is that the main Markov process exhibits a quadratic dependence of the state of the process, due to the use of the law of mass action for the dynamic of our model. One of the main technical difficulties is in the proof of Theorem 15 of an averaging principle. Several preliminary results have to be established as well as a convenient definition of occupation measures. This is due to the (very) fast underlying timescale, $t \mapsto N^2 t$, used. Formally, the diffusion component is of the order of N but should vanish for this first order result. For this reason, in a first step, the "slow" processes are included in the definition of occupation measures and not only the "fast" processes as it is done in general. In our proofs we use several coupling arguments, estimates of hitting times of rare events, stochastic calculus for stochastic differential equations driven by Poisson processes, and the framework of averaging principles.

A Chemical Reaction Network Description

For simplicity, the number of total polymerases is assumed to be constant. There is also a production of 6S RNAs which we will distinguish from the production of other RNAs. From the point of view of our model, polymerases can be in several states

— *Free.* The polymerase may bind to a gene of an mRNA, or of an rRNA, or be sequestered by a 6S RNA, $(F_N(t))$ denotes the process of the number of free polymerases.

— *Transcription of an mRNA.* A chain of nucleotides is produced, $(M_N(t))$ is the number of such polymerases. There is a large number of types of mRNAs.

— *Transcription of an rRNA.* A long chain of nucleotides is produced. As it will be seen, it is described by a process $((U_j^N(t), R_j^N(t)), 1 \leq j \leq J)$. The number J of types of rRNAs is usually small, less than ten. We denote by $\|R(t)\|$ the total number of polymerases in this situation.

— *Sequestered by a 6S RNA.* The associated process is $(S_N(t))$.

Similarly a 6S RNA can be either free or paired with a polymerases, $(Z_N(t))$ denotes the process of the number of free 6S RNAs. See Section 2.2.1 for more details.

With these notations, the assumption on the conservation of mass for the polymerases gives the relation

$$F_N(t) + M_N(t) + S_N(t) + \|R_N\|(t) = N, \qquad \forall t \geq 0.$$

The dynamic of this stochastic system is governed by the analogue of the law of the mass action in this context. See Anderson and Kurtz [5]. The rate of creation of sequestered polymerases is in particular quadratic with respect to the state, it is proportional to $F_N Z_N$. This is one of the important features of this stochastic model.

Two Limiting Regimes

Our mathematical results can be described as follows. See the formal statements in Section 2.5 and 2.6. In Definition 1, we introduce two sets of conditions on the parameters of our model, which define the exponential regime and the stationary regime, our cases a) and b) above. We do not detail them here. Assuming that the maximum number of polymerases simultaneously in transcription of rRNAs, resp. mRNAs, is of the order of $c_r N$, resp. $c_m N$, under some scaling conditions and appropriate initial conditions, we have:

1) *Exponential Phase.*
For the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{\|R_N\|(t)}{N},\frac{M_N(t)}{N}\right)=(c_r,1-c_r),$$

and, for any $t_0>0$, the random variable $(F_N(t_0))$ converges in distribution to a Poisson distribution and the sequence of process $(S_N(t),Z_N(t))$ is converging in distribution to a positive recurrent Markov process. See Theorem 19.

In this case, the polymerases are mostly doing transcription of rRNAs or mRNAs, few of them are free or sequestered by a 6S RNA.

2) *Stationary Phase.*
For the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{M_N(Nt)}{N},\frac{F_N(Nt)}{N},\frac{S_N(Nt)}{N}\right)=(c_m,\overline{f}(t),1-c_m-\overline{f}(t)),$$

where $(\overline{f}(t))$ is the solution of an ODE, such that

$$\lim_{t\to+\infty}\overline{f}(t)=\overline{f}(\infty)>0.$$

The process $(\|R_N(t)\|)$ is stochastically upper-bounded by a positive recurrent Markov process. See Theorem 22.

In the stationary phase there are few polymerases doing transcription of rRNAs. A fraction of them remains free, asymptotically $\overline{f}(\infty)$, i.e. the sequestration process does not control all "useless" polymerases. This is in fact a non-trivial consequence of the dynamic creation and destruction of 6S RNAs, even if an 6S RNA paired with a polymerase cannot be degraded. The fact that sequestration phenomenon of a fraction of the N polymerases occurs on the time scale $t\mapsto Nt$ is intuitive given that the rate of creation of 6S RNAs is constant.

In all cases the process $(Z_N(t))$ is stochastically upper-bounded by a positive recurrent Markov processes.

2.1.5 Outline of the Paper

Section 2.2 introduces in detail the complete model of transcription and also an important model, the auxiliary process. The exponential/stationary phases corresponds to super/sub critical condition for this auxiliary process. They are investigated respectively in Section 2.3 and 2.4. The last sections 2.5 and 2.6 are devoted to the exponential/stationary regimes of our model.

2.2 Stochastic Model

The chemical species involved in the regulation process are the genes of different types of mRNAs and rRNAs and of $6S$ RNA, and polymerases. The products are different types of mRNAs and of rRNAs and also 6S RNAs. We first describe our main assumptions of our stochastic model.

2.2.1 Modeling Assumptions

— Transcription of rRNAs.
There are J types of rRNAs and there is a promoter (binding site for polymerases) for each of them. The transcription of an rRNA of type j, $1 \leq j \leq J$, is in two steps. Promoters of rRNAs are assumed to have a high affinity during the growth phase: If one of these promoters is empty and if there is at least one free polymerase, then the promoter is occupied right away by a polymerase.

Once a polymerase is bound to the promoter of an rRNA of type $1 \leq j \leq J$, it starts elongation at rate $\alpha_{r,j}$ if there are strictly less than $C^N_{r,j}$ polymerases in the elongation phase of this rRNA. At a given moment there cannot be more than $C^N_{r,j}$ polymerases in elongation of an rRNA of type j.

For each polymerase in elongation, nucleotides are collected at rate $\beta_{r,j}$. The simplification of the model is that the polymerases in elongation are moving closely on the gene so that the duration of time to get the last nucleotide for the oldest polymerase in elongation is enough to describe the time evolution of the number of polymerases producing rRNA of type j. The polymerases associated to an rRNA of type $j \in \{1, \dots, J\}$ can then be represented as a couple (u_j, R_j), where $u_j \in \{0, 1\}$ indicates if a polymerase is on the promoter or not, and $R_j \in \mathbb{N}$ is the number of polymerases in elongation: If $R_j \geq 1$, an rRNA of type j is therefore created at rate $\beta_{r,j}$.

The assumption is reasonable in the exponential phase, since in this case the number of polymerases producing rRNA of type j is maximal, of the order of $C^N_{r,j}$. See Section 2.5. The rRNA part of the system is therefore saturated. In the stationary phase, this assumption has little impact since, as we shall see, the total number of polymerases in the elongation phase of rRNAs is small with high probability and, therefore, negligible for our scaling analysis.

— Transcription of mRNAs.
It is assumed that there are C_m^N different types of mRNAs and that at a given time, for any $1 \leq i \leq C_m^N$ there is at most one polymerase in the elongation phase of an mRNA of type i. When the promoter of an mRNA of type i is free, a free polymerase may bind to this promoter at a rate α_m. If the promoter of an mRNA of type i is occupied, an mRNA is released at rate β_m and the corresponding polymerase leaves the promoter at that instant. The production of mRNAs have simplified in the sense that the initiation phase and the elongation phase are merged into one step. The results obtained in this paper could be obtained without too much difficulty for a model distinguishing them, but at the expense of a more complicated state variable.

The main difference in our model between the rRNAs and the mRNAs is on the number of polymerases in elongation of the corresponding gene. At a given moment, under favorable growth conditions, there will be many polymerases in the elongation phase of an rRNA, due in particular to the high initiation rate of these genes. For the mRNAs, the number of polymerases in elongation phase of a given mRNA type should be small in general. Indeed, there are in each cell few copies of each messenger (from 1 to 100). Furthermore, the rate of production of each messenger is such that its small number remains on average constant during growth or stationary phases and despite the regular degradation (average of 2 minutes half-life in high-growth rate phase) of each of them. See Section 2.2.3. In our model we have set the maximal number of polymerases in elongation phase of a given mRNA type to 1 for simplicity, but it is not difficult to adapt our results with a maximum number D. Similarly, the initiation rates and production rate, α_m and β_m are taken equal for all species of mRNA, also for the sake of simplicity. We have simplified the description of the production of mRNAs to focus mainly on the sequestration mechanisms that regulate the transcription. From our point of view, the production of mRNAs holds/stores a subset of polymerases and releases each of them after some random amount of time. It should be noted that this is in fact the usual mathematical setting to investigate the fluctuations of the production of mRNAs and proteins. See Berg [9] and Rigney and Schieve [60], see also Paulsson [57] for a review of these models.

— Creation/Degradation of 6S RNAs.
The creation of 6S RNAs involves, of course, polymerase. As in the case of mRNAs, it is assumed that there is at most one polymerase in elongation phase of this sRNA. A 6S RNA is created at rate $\beta_6 > 0$. A 6S RNA is free when it is not bound to a polymerase. A given free 6S RNA is degraded at rate $\delta_6 \geq 0$. Only a free 6S RNA can be degraded.

— Sequestration/de-Sequestration of Polymerases.
A polymerase is free when it is not bound to a gene or to a 6S RNA. In our study the total number of polymerases is assumed to be constant equal to N. A free polymerase is bound to a free 6S RNA at rate λ. A complex polymerase-6S RNA breaks into a free polymerase and a free 6S RNA at rate η.

2.2.2 The Markov Process and its Q-Matrix

The vector $(\alpha_{r,j})$ introduced is the vector of *initiation* rates of transcription of the different types of rRNAs. The difference between a slow growth (stationary phase) and a steady growth (exponential phase) will be expressed in terms of the comparison, coordinate by coordinate, of the vectors $(\alpha_{r,j})$ and $(\beta_{r,j})$. We now give a Markovian description of our system. Convenient limiting results will be obtained for the associated Markov process in both phases.

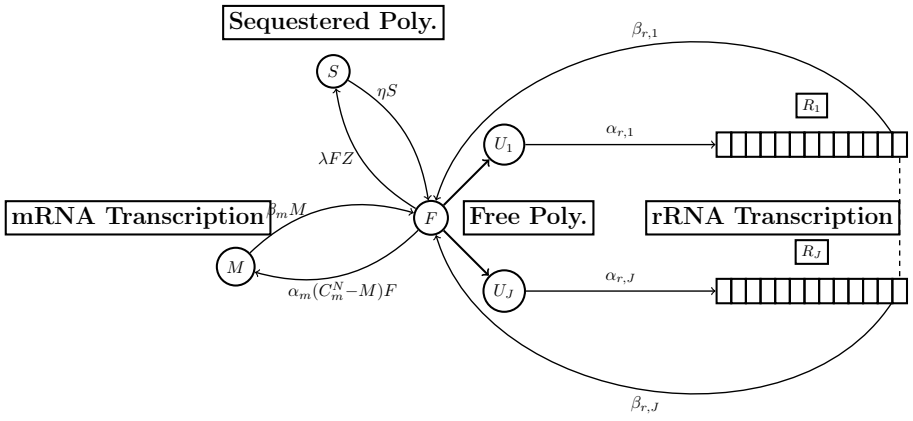


Figure 2.1: Polymerases: Transcription of mRNAs/rRNAs and Sequestration

State Space

All transitions described in the last section occurs after a random amount of time with an exponential distribution. With this assumption, there is a natural Markov process to investigate the regulation of transcription. The state space

is given by

$$\mathcal{S}_N \stackrel{\text{def.}}{=} \left\{ x=(f,s,z,(u_j,r_j))\in\mathbb{N}^3\times\prod_{j=1}^{J}\Big(\{0,1\}\times\{0,\dots,C^N_{r,j}\}\Big) : \right.$$
$$\left. f+s+\sum_{j=1}^{J}(u_j+r_j)\le N \text{ and if } f>0, \text{ then } u_j=1, \forall 1\le j\le J \right\},$$

if the state of the system is $x=(f,s,z,(u,r))\in\mathcal{S}_N$, then

— f is the number of free polymerases;

— s, the number of sequestered polymerases;

— z, the number of free 6S RNAs;

$(u,r)=((u_j,r_j),1\le j\le J)$,

— $u_j\in\{0,1\}$ to indicate if a polymerase is bound to the promoter of the rRNA of type j or not;

— $0\le r_j\le C^N_{r,j}$, number of polymerases in elongation phase of an rRNA of type j.

— In state x, the number of polymerases in elongation phase of an mRNA is given by

$$\Psi(x) \stackrel{\text{def.}}{=} N-f-s-\sum_{j=1}^{J}(u_j+r_j). \tag{2.1}$$

The associated Markov process is denoted by

$$(X_N(t)) \stackrel{\text{def.}}{=} (F_N(t),S_N(t),Z_N(t),(U_N(t),R_N(t)))\,,$$

with $(U_N(t),R_N(t))=((U^N_j(t),R^N_j(t)),1\le j\le J)$. The number of polymerases at time t in elongation phase of an mRNA is defined by $M_N(t)=\Psi(X_N(t))$.

If $w=(w_j)\in\mathbb{N}^J$, we define $\|w\|=w_1+\cdots+w_J$ and, for $1\le j\le J$, e_j denotes the jth unit vector of $\mathbb{N}^J$. It is easily checked that $(X_N(t))$ is an irreducible Markov process on $\mathcal{S}_N$. Its transition rates are given by

— Transcription of rRNAs. For $1\le i,j\le J$,

$$(f,s,z,(u,r)) \longrightarrow \begin{cases} (f-1,s,z,(u,r+e_j)) & \alpha_{r,j}\mathbb{1}_{\{f>0,r_j<C^N_{r,j}\}},\\ (0,s,z,(u-e_j,r+e_j)) & \alpha_{r,j}\mathbb{1}_{\{u_j=1,r_j<C^N_{r,j},f=0\}},\\ (f+1,s,z,(u,r-e_j)) & \beta_{r,j}\mathbb{1}_{\{r_j>0,u_k>0,\forall 1\le k\le J\}},\\ (0,s,z,(u+e_i,r-e_j)) & \dfrac{\beta_{r,j}}{J-\|u\|}\mathbb{1}_{\{r_j>0,u_i=0\}}. \end{cases}$$

— Transcription of mRNAs.

$$(f,s,z,(u,r)) \longrightarrow \begin{cases} (f{-}1,s,z,(u,r)) & \alpha_m f\left(C_m^N{-}\Psi(x)\right), \\ (f{+}1,s,z,(u,r)) & \beta_m \Psi(x). \end{cases}$$

— Creation/Degradation of 6S RNAs.

$$(f,s,z,(u,r)) \longrightarrow \begin{cases} (f,s,z{+}1,(u,r)) & \beta_6, \\ (f,s,z{-}1,(u,r)) & \delta_6 z. \end{cases}$$

— Sequestration/de-Sequestration of Polymerases.

$$(f,s,z,(u,r)) \longrightarrow \begin{cases} (f{-}1,s{+}1,z{-}1,(u,r)) & \lambda f z, \\ (f+1,s{-}1,z{+}1,(u,r)) & \eta s. \end{cases}$$

2.2.3 Orders of Magnitude and Scaling Assumptions

We now discuss the orders of magnitude of the main parameters of the biological process.

— The scaling variable used in our analysis is N, the total number of polymerases in the cell. It is assumed that this number is constant during the growth phase investigated, this quantity is quite large, between 2000 and 10000 for E. coli, depending of the environment. See Bakshi et al. [7].

— The number J of different types of rRNA is small, of the order of 10. See Bremer and Dennis [11].

— We shall assume that the maximal number of polymerases in transcription of an rRNA of type j, $C_{r,j}^N$, is of the order of N, the total number of polymerases. Indeed, in a steady growth phase a given rRNA gene can accommodate a significant number of polymerases. Recall that the length in nucleotides of an rRNA is large, of the order of 5000.

— Similarly, the total number of different types of mRNAs is also of the order of N, several thousands, of the order of 3500 for E. coli.

See also Milo et al. [49], Karp et al. [42] and Neidhardt and Umbarger [53] for the estimation of the numerical values of these quantities in various contexts.

Due to the modeling assumptions of Section 2.2.1, we assume that the relations

$$\lim_{N\to+\infty} \frac{C_{r,j}^N}{N} = c_{r,j} > 0,\ 1{\leq}j{\leq}J,\ \text{and}\ \lim_{N\to+\infty} \frac{C_m^N}{N} = c_m, \tag{2.2}$$

hold and that, in order to cope with the production of rRNAs during a steady growth phase, the total number of polymerases N is larger than the total maximal number of polymerases in elongation phase of rRNAs, i.e. that $C_{r,1}^N{+}\cdots{+}C_{r,J}^N$

and, also that there are not too many polymerases for the transcription, i.e.

$$\max\left(C_m^N, \sum_{j=1}^{J} C_{r,j}^N\right) < N < C_m^N + \sum_{j=1}^{J} C_{r,j}^N.$$

In view of (2.2), these assumptions are expressed by the following conditions on the scaled parameters $(c_{r,j})$ and c_m,

$$\max\left(c_m, \sum_{j=1}^{J} c_{r,j}\right) < 1 < \sum_{j=1}^{J} c_{r,j} + c_m. \tag{2.3}$$

We can now introduce the two regimes of interest in our paper.

Definition 1.

a. The Exponential Phase *is defined by the relation*

$$\min_{1\le j\le J} \frac{\alpha_{r,j}}{\beta_{r,j}} > 1. \tag{2.4}$$

The initiation rate $\alpha_{r,j}$ of type j rRNAs is greater than its production rate.

b. The stationary phase *is defined by the relation*

$$\max_{1\le j\le J} \frac{\alpha_{r,j}}{\beta_{r,j}} < 1. \tag{2.5}$$

The initiation rate $\alpha_{r,j}$ of type j rRNAs is less than its production rate.

It should be noted that Relations (2.4) and (2.5) are not complementary but this is not a concern for the following reason. If there exists a subset S of $\{1, \ldots, J\}$ such that

$$\max_{j\notin S} \frac{\alpha_{r,j}}{\beta_{r,j}} < 1 < \min_{j\in S} \frac{\alpha_{r,j}}{\beta_{r,j}},$$

we will express it as a model for which the rRNAs are defined by the subset S and the remaining nodes $S^c = \{1, \ldots, J\} \backslash S$ are included in the mRNAs. It can be shown that the addition of a finite number of nodes to the mRNA does not change the first order in N of the number of polymerases in transcription of an mRNA. See Section 2.6. With this change and if Condition (2.3) holds for this modified system, this is still an exponential phase.

2.2.4 An Auxiliary Model

In Sections 2.3 and 2.4, we study a process which can be interpreted as a model similar to $(X_N(t))$ but with only transcription of mRNAs and sequestration by 6S RNAs but without rRNAs. The reasons to study this case are two-fold:

a. Exponential Phase. If Condition (2.4) holds, as we shall see, "most" of the $J+C^N_{r,1}+C^N_{r,2}+\cdots+C^N_{r,J}$ available places for transcriptions of rRNAs will be occupied by polymerases. Provided that this situation holds on a sufficiently large time scale, under Condition (2.3), there are A_N available polymerases for sequestration and transcription of mRNAs, with

$$A_N \stackrel{\text{def.}}{=} N-J-\sum_{j=1} C^N_{r,j} \sim \gamma N < C^N_m. \tag{2.6}$$

The system works as if there were A_N polymerases available for the transcription of mRNAs. With Condition (2.3), we have $A_N<C^N_m$.

b. Stationary Phase. When Condition (2.5) holds, then, roughly speaking, the total number of polymerases in the elongation phase of rRNAs is $O(1)$, so that this part of the system is in some way negligible. In this case the total number of polymerases available for transcription of mRNAs is essentially N and thus greater than C^N_m under Condition (2.3).

The precise definition of exponential phase, resp. stationary phase, is in Section 2.5, resp. Section 2.6.

We denote $(X^0_N(t))=(F^0_N(t), S^0_N(t), Z^0_N(t))$ the system defined in Section 2.2.2 but without the part of the model for rRNAs. From a state (f,s,z), the transition rates are:

$$\begin{cases} (f-1,s,z) & \alpha_m \left(C^N_m-(N-f-s)\right) f, \\ (f+1,s,z) & \beta_m(N-f-s), \\ (f,s,z+1) & \beta_6, \end{cases} \qquad \begin{cases} (f,s,z-1) & \delta_6 z, \\ (f-1,s+1,z-1) & \lambda f z, \\ (f+1,s-1,z+1) & \eta s. \end{cases} \tag{2.7}$$

Using the classical formulation in terms of a martingale problem, see Theorem (20.6) in Section IV of Rogers and Williams [63] for example, the Markov process $(X^0_N(t))$ whose Q-matrix is given by Relation (2.7), as $(F^0_N(t), S^0_N(t), Z^0_N(t))$, the solution of the SDEs,

$$\begin{aligned} \mathrm{d}F^0_N(t) &= \mathcal{P}_1\left(\left(0,\eta S^0_N(t-)\right),\mathrm{d}t\right) - \mathcal{P}_3\left(\left(0,\lambda F^0_N(t-)Z^0_N(t-)\right),\mathrm{d}t\right) \\ &\quad + \mathcal{P}_2\left(\left(0,\beta_m\left(N-F^0_N(t-)-S^0_N(t-)\right)\right),\mathrm{d}t\right) \\ &\quad - \mathcal{P}_4\left(\left(0,\alpha_m F^0_N(t-)\left(C^N_m-\left(N-F^0_N(t-)-S^0_N(t-)\right)\right)\right),\mathrm{d}t\right) \end{aligned} \tag{2.8}$$

$$\mathrm{d}S^0_N(t) = -\mathcal{P}_1\left(\left(0,\eta S^0_N(t-)\right),\mathrm{d}t\right) + \mathcal{P}_3\left(\left(0,\lambda F^0_N(t-)Z^0_N(t-)\right),\mathrm{d}t\right) \tag{2.9}$$

$$\begin{aligned} \mathrm{d}Z^0_N(t) &= \mathcal{P}_5\left((0,\beta_6),\mathrm{d}t\right) - \mathcal{P}_6\left(\left(0,\delta_6 Z^0_N(t-)\right),\mathrm{d}t\right) \\ &\quad + \mathcal{P}_1\left(\left(0,\eta S^0_N(t-)\right),\mathrm{d}t\right) - \mathcal{P}_3\left(\left(0,\lambda F^0_N(t-)Z^0_N(t-)\right),\mathrm{d}t\right), \end{aligned} \tag{2.10}$$

with the convenient initial conditions, where $\mathcal{P}_i$, $i\in\{1,2,3,4\}$ are independent Poisson processes on $\mathbb{R}^2_+$ with intensity $\mathrm{d}s\otimes\mathrm{d}t$.

We will study two regimes of this stochastic model:

— Sub-critical case, when $c_m>1$, i.e. $N<C^N_m$ for N sufficiently large.

— Super-critical case, when $c_m<1$.

As it will be seen these two regimes are respectively associated to the exponential and stationary phases.

Notations

We define a common filtration common to all our processes, as follows, for $t \geq 0$

$$\mathcal{F}_t = \sigma\left(\mathcal{P}_i(A \times [0,s]) : A \in \mathcal{B}(\mathbb{R}_+), i \in \{1, \cdots, 6\}, s \leq t\right). \tag{2.11}$$

From now on, all notions of stopping time, adapted process, martingale, refer to this (completed) filtration. A càdlàg process is an adapted process such that with probability one, it is right-continuous process with left limits at any positive real number.

If H is a locally compact metric space, we denote by $\mathcal{C}_c(H)$ the set of continuous functions with compact support on H. It is endowed with the topology of the uniform norm. The set $\mathcal{P}(H)$ is the space of Borelian probability distributions on H.

2.3 Sub-critical Case

It is assumed throughout this section that $c_m > 1$ holds, the total number of possible sites for transcription of mRNAs is much larger than the total number of polymerases.

Definition 2 (Occupation measure of $(F_N^0(t))$)**.** *For $g \in \mathcal{C}_c(\mathbb{R}_+ \times \mathbb{N})$*

$$\langle \mu_N, g \rangle \stackrel{\text{def.}}{=} \int_0^{+\infty} g\left(u, F_N^0(u)\right) \mathrm{d}u. \tag{2.12}$$

We start with a technical result on a birth and death process.

Lemma 3. *For $\kappa_i > 0$ and $\kappa_o > 0$, let $(Y(t))$ be a birth and death process on $\mathbb{N}$ whose Q-matrix is given by*

$$q(x, x+1) = \kappa_i \text{ and } q(x, x-1) = \kappa_o x, \quad x \in \mathbb{N},$$

a. if $Y(0) = N$ and

$$H_Y^N \stackrel{\text{def.}}{=} \inf\{t > 0 : Y(t) = 0\},$$

then $(H_Y^N / \ln N)$ is converging in distribution to a constant.

b. if $Y(0) = 0$, then for any $\delta > 0$, the convergence in distribution

$$\lim_{N \to +\infty} \left(\frac{Y\left(N^\delta t\right)}{\ln(N)^2} \right) = 0$$

holds.

The process $(Y(t))$ can be thought as a kind of discrete Ornstein-Uhlenbeck process on $\mathbb{N}$. In a queueing context, this is the process of the number of jobs of an $M/M/\infty$ queue. See Chapter 6 of Robert [61] for example. Its invariant distribution is Poisson with parameter κ_i/κ_o.

Proof. The first assertion comes directly from Proposition 6.8 of [61]. If $Y(0)=0$ and, for $p\geq 1$,

$$T_p = \inf\{t>0 : Y(t) > p\},$$

Proposition 6.10 [61] gives that, if $\rho \stackrel{\text{def.}}{=} \kappa_i/\kappa_o$, the sequence

$$\left(\rho^p \frac{T_p}{p!}\right)$$

is converging in distribution to an exponential distribution. In particular for any $K>0$,

$$\lim_{N\to+\infty} \mathbb{P}\left(T_{\ln(N)^2} < KN^\delta\right) = 0,$$

since, by Stirling's Formula,

$$\lim_{N\to+\infty} \frac{(\ln(N)^2)!}{\rho^{\ln(N)^2} N^\delta} = +\infty.$$

The lemma is proved. □

We begin with a lemma showing that the initial state of $(X_N^0(t))$ can be taken with few free polymerases.

Lemma 4. *If $c_m>1$ and $(F_N^0(0), S_N^0(0), Z_N^0(0))=(f_N, s_N, z_N)$, such that*

$$\lim \frac{1}{N}(f_N, s_N, z_N) = ((f_0, s_0, z_0)\in\mathbb{R}_+^3,$$

with $f_0+s_0<1$, and, if

$$\tau_N^0 \stackrel{\text{def.}}{=} \inf\left\{t : F_N^0(t) = 0\right\},$$

then the sequence $(\tau_N^0/(\ln N)^2)$ is converging in distribution to 0.

The Condition $f_0+s_0<1$ is to take into account the fact that $F_N^0(0)+S_N^0(0)\leq N$.

Proof. Because of the assumption $c_m>1$, for N sufficiently large, there exists $\varepsilon_0>0$ such that $C_m^N-N>\varepsilon_0 N$, the relations (2.7) for the transition rates show that one can construct a coupling $(F_N^0(t), Y(t))$ such that $Y(0)=F_N^0(0)$ and the relation

$$F_N^0(t) \leq Y(Nt), \qquad \forall t\geq 0, \tag{2.13}$$

holds almost surely for all $t\geq 0$, where $(Y(t))$ is a process as defined in Lemma 3 with $\kappa_i=\eta+\beta_m$ and $\kappa_o=\alpha_m\varepsilon_0$. We conclude the proof by using Lemma 3. □

We can now state the main result of this section. It shows that for the asymptotic system, when N is large, all polymerases are eventually in the transcription phase of mRNAs, i.e. the fraction of sequestered polymerases is close to 0. A sketch of the proof is given in Section 2.7.1.

Proposition 5 (Starting from a Congested State)**.** *Under the condition* $c_m>1$ *and if the initial state is* $(F_N^0(0), S_N^0(0), Z_N^0(0))=(0, s_N, z_N)$ *and*

$$\lim \frac{1}{N}(s_N, z_N) = (s_0, z_0) \in \mathbb{R}_+^2,$$

such that $s_0+z_0<1$*, then, for the convergence in distribution*

$$\lim_{N\to+\infty} \left(\frac{S_N^0(t)}{N}, \frac{Z_N^0(t)}{N} \right) = (s(t), z(t)),$$

where $(s(t), z(t))$ *is the unique solution of the system of ODEs,*

$$\dot{s}(t) = -\eta s(t)+\lambda z(t)\frac{\beta_m-(\beta_m-\eta)\, s(t)}{\alpha_m(c_m-1+s(t))+\lambda z(t)}, \quad \dot{s}(t)+\dot{z}(t) = -\delta_6 z(t),$$

with $(s(0), z(0)) = (s_0, z_0)$*.*

It is not difficult to see that the function $(s(t), z(t))$ is converging to $(0, 0)$ at infinity.

To study the asymptotic behavior of the model in the exponential regime, we investigate the occupation measure associated to free polymerases when the initial state is "small". In this case, contrary to the last proposition, the processes $(S_N^0(t), Z_N^0(t))$ should be "slow", i.e. their transition rate are of the order of $O(1)$, only $(F_N^0(t))$ is "fast".

Proposition 6 (Fixed Initial Point)**.** *Under the condition* $c_m>1$ *and the initial state is such that* $(F_N^0(0), S_N^0(0), Z_N^0(0))=(f_0, s_0, z_0)\in\mathbb{N}^3$*, then, for the convergence in distribution*

$$\lim_{N\to+\infty} \langle \mu_N, g\rangle = \int_0^{+\infty} \mathbb{E}\left(g\left(u, \mathcal{N}_1\left(0, \rho_m\right)\right)\right) \mathrm{d}u,$$

for any $g\in\mathcal{C}_c\left(\mathbb{R}_+\times\mathbb{N}\right)$*, where* $\rho_m=\beta_m/(\alpha_m(c_m-1))$*,* μ_N *is the occupation measure defined by Relation* (2.12)*, and* $\mathcal{N}_1$ *is a Poisson process with rate* 1*.*

The sequence of processes $(S_N^0(t), Z_N^0(t))$ *converges in distribution for the Skorohod topology to a jump process* $(Y(t))$ *on* $\mathbb{N}^2$ *whose transition rates are given by*

$$(s, z) \longrightarrow (s, z)+\begin{cases}(1,-1) & \lambda\rho_m z,\\ (-1,1) & \eta s,\end{cases} \quad \begin{cases}(0,1) & \beta_6,\\ (0,-1) & \delta_6 z.\end{cases}$$

See Section 2.7.2.

2.4 Super-critical Case

In this section we study the auxiliary process under the condition $c_m<1$, so that $C_m^N<N$ for N sufficiently large. In this case the places for transcription of

mRNAs are likely to be saturated quickly. Consequently, there should remain many free polymerases and the sequestration mechanism has to play a role.

If there are no 6S RNAs initially, since the creation of 6S RNAs is constant, the sequestration of a significant fraction of these polymerases will occur after a duration of time at least of the order of N. In this case, when a 6S RNA is created, it is right away paired with a free polymerase and will paired again and again that after the successive steps of sequestration/desequestration, as long as the number of free polymerases is sufficiently "large". The sequestration occurs always before a possible degradation of the 6S RNAs takes place. The precise result is in fact more subtle than that. It will be shown that, on the fast time scale $t \mapsto Nt$, the sequestration of polymerases increases but, due to the degradation of 6S RNAs there remains a positive fraction of free polymerases.

The goal of this section is of proving an averaging principle for the process $(F_N^0(t), Z_N^0(t))$. A coupling and a technical lemma are presented in Section 2.4.1, tightness properties of occupations measures are proved in Section 2.4.2, finally the main convergence results are proved in Section2.4.3.

Definition 7. *For $N>0$ and $t\geq 0$, we define*

$$G_N^0(t) \stackrel{\text{def.}}{=} C_m^N - \left(N - F_N^0(t) - S_N^0(t)\right),$$

the number of "empty" places for transcription of mRNAs at time t.

The scaled process is defined by

$$\left(\overline{X}_N^0(t)\right) = \left(\overline{F}_N^0(t), G_N^0(Nt), Z_N^0(Nt)\right) \text{ with } \left(\overline{F}_N^0(t)\right) \stackrel{\text{def.}}{=} \left(\frac{F_N^0(Nt)}{N}\right). \tag{2.14}$$

If g is non-negative Borelian function on $\mathbb{R}_+^2 \times \mathbb{N}^2$, we define the occupation measure

$$\left\langle \overline{\Lambda}_N^0, g \right\rangle \stackrel{\text{def.}}{=} \int_{\mathbb{R}_+} g\left(s, \overline{X}_N^0(s)\right) \mathrm{d}s. \tag{2.15}$$

Note that the, a priori, slow process $(\overline{F}_N(t))$ is also included in the definition of the occupation measure $\overline{\Lambda}_N^0$. The reason is that the proof of the tightness of $(\overline{F}_N(t))$ (for the topology of the uniform norm on càdlàg functions) is not clear. Due to the fast time scale, the proof that the martingale component of $(\overline{F}_N(t))$ vanishes does not seem to be straightforward.

The following initial conditions will be assumed,

$$\lim_{N\to+\infty} \frac{F_N^0(0)}{N} = \overline{f}_0 \in (0, 1-c_m),\ G_N^0(0) = m_0, \text{ and } Z_N^0(0) = z_0, \tag{2.16}$$

with m_0 $z_0 \in \mathbb{N}$. A fraction $\overline{f}_0$ of the polymerases are initially free and there are z_0 6S RNAs and $C_m^N - m$ polymerases in the transcription phase of mRNAs and the number of sequestered polymerases $S_N^0(0)$ is therefore such that

$$\lim_{N\to+\infty} \frac{S_N^0(0)}{N} = 1 - \overline{f}_0 - c_m.$$

As it will be seen in Section 2.4.1 there is no loss of generality to consider these initial conditions.

Before proving the convergence of the sequence of processes $(\overline{X}^0_N(t))$, we analyze the convergence of a "stopped" version of it. In several technical arguments we will need that the fraction of free polymerases is not too small. A second step is of showing that, essentially, the stopped process does not differ from the original process.

Definition 8. *For $a>0$, the stopping time $\tau_N(a)$ is defined by*

$$\tau_N(a) \stackrel{\text{def.}}{=} \inf\left\{t>0 : F^0_N(Nt) \leq aN\right\}, \tag{2.17}$$

and

a. *if $(W(t))$ is a càdlàg process, we denote $(W^a_N(t))=(W(N(t\wedge\tau_N(a))))$;*

b. *The "stopped" occupation measure $\overline{\Lambda}^{0,a}_N$ is defined by, if g is non-negative Borelian function on $\mathbb{R}^2_+\times\mathbb{N}^2$,*

$$\left\langle \overline{\Lambda}^{0,a}_N, g\right\rangle \stackrel{\text{def.}}{=} \int_0^{\tau_N(a)} g\left(s, \overline{X}^0_N(s)\right) \mathrm{d}s.$$

With a slight abuse, the notation $(\overline{F}^a_N(t))=(F^0_N(N(t\wedge\tau_N(a)))/N)$ will be used in the following.

2.4.1 Technical Lemmas

The two processes $(G^0_N(t)$ and $(Z^0_N(t))$ are in fact in a neighborhood of 0 quickly. They will be the fast processes (on the timescale $t\mapsto Nt$) of our averaging principle. In state $F^0_N=f$, $G^0_N=g$ and $Z^0_N=z$, f, g, $z\in\mathbb{N}$, the jump rates of the process $(G^0_N(t)$ and $(Z^0_N(t))$ are respectively

$$\begin{cases} +1, & \beta_m(C^N_m-g), \\ -1, & \alpha_m f g, \end{cases} \quad \text{and} \quad \begin{cases} +1, & \left(\beta_6+\eta(N-f-C^N_m+g)\right), \\ -1, & (\lambda f+\delta_6)\, z. \end{cases}$$

If $\eta_0>\eta$ and $\eta_1>\beta_m c_m$, and N sufficiently large, up to time $\tau_N(a)$, a simple coupling shows that there exist independent processes $(Y_G(t))$ and $(Y_Z(t))$ such that

$$G^0_N(Nt) \leq Y_G(N^2t) \text{ and } Z^0_N(Nt) \leq Y_Z(N^2t), \tag{2.18}$$

holds for all $t\in(0,\tau_N(a))$. The process $(Y_G(t))$, resp. the process $(Y_Z(t))$, is as in Lemma 3 with $\kappa_{i,G}=\eta_1$ and $\kappa_{o,G}=\alpha_m a$ (resp. $\kappa_{i,Z}=\eta_0$ and $\kappa_{o,Z}=\lambda a$), and $Y_G(0)=G^0_N(0)$, resp. $Y_Z(0)=Z^0_N(0)$. It is not difficult, using again Lemma 3, as in Section 2.3, that the hitting time of $(0,0)$ by $(Y_G(t),Y_Z(t))$ is of the order of $\ln N$ so that Condition (2.16) for the initial state can be assumed.

Lemma 9. *Under Conditions* (2.16) *and* $c_m<1$*, and if* $a\in(0,\overline{f}_0)$*, then*

$$\lim_{N\to+\infty}\mathbb{P}\left(\tau_N(a)<t_0^a\right)=0,$$

with $t_0^a=(\overline{f}_0-a)/\beta_6$*, and the relation*

$$\lim_{N\to+\infty}\left(\frac{G_N^{0,a}(Nt)}{\ln(N)^2},\frac{Z_N^{0,a}(Nt)}{\ln(N)^2}\right)=(0).$$

holds for the convergence in distribution.

In the following, we will use the notation t_0^a, where $a\in(0,\overline{f}_0)$ is fixed.

Proof. The first relation is clear since, for $x\leq 1$ and $t>0$, on the event $\{\overline{F}_N^0(t)<x\}$ there are at least $F_N^0(0)-\lfloor Nx\rfloor-z_0$ new 6S RNAs which have been created up to time t. The rest of the proof follows from the coupling with $(Y_G(t),Y_Z(t))$, Relation (2.18), and Lemma 3. □

2.4.2 Tightness Properties

Proposition 10. *Under Conditions* (2.16) *and* $c_m<1$*, the sequence of measure-valued processes* $(\overline{\Lambda}_N^{0,a})$ *on the state space* $[0,t_0^a)\times\mathbb{R}_+\times\mathbb{N}^2$ *is tight for the convergence in distribution and any limiting point* $\overline{\Lambda}_\infty^{0,a}$ *can be expressed as,*

$$\left\langle\overline{\Lambda}_\infty^{0,a},f\right\rangle=\int_{[0,t_0^a)\times\mathbb{R}_+\times\mathbb{N}^2}f(s,x,p)\,\pi_s^a(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s,\tag{2.19}$$

for any function f *with compact support on* $[0,t_0^a)\times\mathbb{R}_+\times\mathbb{N}^2$*, where* $t_0^a=(\overline{f}_0-a)/\beta_6$ *and* (π_s^a) *is an optional process with values in* $\mathcal{P}(\mathbb{R}_+\times\mathbb{N}^2)$*.*

For an introduction to the convergence in distribution of measure-valued processes, see Dawson [19]. The optional property is just used to have convenient measurability properties to define random variables as integrals with respect to $(\pi_s^a,s>0)$. See Section VI.4 of Rogers and Williams [62].

Proof. Note that, for $K>0$ and $t<t_0^a$, since

$$\int_0^t\mathbb{1}_{\left\{Z_N^0(Ns)\geq K\right\}}\,\mathrm{d}s=\int_0^t\mathbb{1}_{\left\{Z_N^{0,a}(s)\geq K\right\}}\,\mathrm{d}s$$

holds on the event $\{\tau_N(a)\geq t_0^a\}$, then

$$\begin{aligned}\mathbb{E}\left(\overline{\Lambda}_N^{0,a}([0,t_0^a]\times[0,1]\times\mathbb{N}\times[K,+\infty])\right)&\\ \leq\mathbb{E}\left(\mathbb{1}_{\left\{\tau_N(a)>t_0^a\right\}}\int_0^{t_0^a}\mathbb{1}_{\left\{Z_N^{0,a}(s)\geq K\right\}}\,\mathrm{d}s\right)&+t_0^a\mathbb{P}\left(\tau_N(a)<t_0^a\right),\end{aligned}$$

and, with Relation (2.18) and Lemma 9, we have

$$\mathbb{E}\left(\mathbb{1}_{\{\tau_N(a)>t_0^a\}}\int_0^{t_0^a}\mathbb{1}_{\{Z_N^{0,a}(s)\geq K\}}\,\mathrm{d}s\right)$$
$$\leq\int_0^{t_0^a}\mathbb{P}(Y_Z(N^2s)\geq K)\,\mathrm{d}s=\frac{1}{N^2}\int_0^{N^2t_0^a}\mathbb{P}(Y_Z(s)\geq K)\,\mathrm{d}s,$$

since the Markov process $(Y_Z(t))$ converges in distribution to a Poisson distribution with parameter $\kappa_{i,Z}/\kappa_{o,Z}$, the ergodic theorem for Markov processes and Lemma 9 give therefore the inequality

$$\limsup_{N\to+\infty}\mathbb{E}\left(\overline{\Lambda}_N^{0,a}([0,t_0^a]\times[0,1]\times\mathbb{N}\times[K,+\infty])\right)\leq t_0^a\mathbb{P}(\mathcal{N}_1(0,\eta/(\lambda a))\geq K),$$

where $\mathcal{N}_1$ is a Poisson process on $\mathbb{R}_+$ with rate 1. One can choose K sufficiently large such that $\mathbb{E}\left(\overline{\Lambda}_N^{0,a}([0,t_0^a]\times[0,1]\times\mathbb{N}\times[K,+\infty])\right)$ is arbitrarily small uniformly in N. Similarly, by replacing (Z_N^0,Y_Z) by (G_N^0,Y_G) the same property can be proved for $\mathbb{E}\left(\overline{\Lambda}_N^{0,a}([0,t_0^a]\times[0,1]\times[K,+\infty]\times\mathbb{N}\right)$ for K and N sufficiently large. For any $\varepsilon>0$, there exists some K_0 such that

$$\sup_N\mathbb{E}\left(\overline{\Lambda}_N^{0,a}([0,t_0^a]\times[0,1]\times[0,K_0]^2\right)\geq(1-\varepsilon)t_0^a.$$

Lemma 1.3 of Kurtz [45] shows that the sequence $(\overline{\Lambda}_N^{0,a})$ is tight, and Lemma 1.4 of the same reference gives the representation (2.19). □

Proposition 10 has established tightness properties $(\overline{\Lambda}_N^{0,a})$. The following simple lemma extends this result in terms of the convergence of stochastic processes. It will be used repeatedly, in particular to identify the possible limits of $(\overline{\Lambda}_N^{0,a})$. See Dawson [19] for example.

Lemma 11. *Under Conditions* (2.16) *and* $c_m<1$, *if* $(\overline{\Lambda}_{N_k}^{0,a})$ *is a subsequence converging to* $\overline{\Lambda}_\infty^{0,a}$ *satisfying Relation* (2.19), *then for any* $g\in\mathcal{C}_c(\mathbb{R}_+\times\mathbb{N}^2)$, *for the convergence in distribution of processes associated to the uniform norm,*

$$\lim_{k\to+\infty}\left(\int_0^t g\left(\overline{X}_{N_k}^0(s)\right)\mathrm{d}s\right)=\left(\int_0^t\int_{\mathbb{R}_+\times\mathbb{N}^2}g\,(x,p)\,\pi_s^a(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s\right).$$

Proof. The tightness of the sequence of stochastic processes is obtained by the use of the criterion of the modulus of continuity. See Theorem 7.3 of Billingsley [10]. The identification of the limit is a straightforward consequence of the convergence of $(\overline{\Lambda}_{N_k}^{0,a})$ □

If we divide by N^2 Relation (2.40), we get that, on the event $\{\tau_N(a)>t\}$, the relation

$$\begin{aligned}\frac{1}{N^2}f\left(\overline{X}^0_N(t)\right) &= \frac{1}{N^2}f\left(\overline{X}^0_N(0)\right)+\frac{M_{f,N}(t)}{N^2} \\ &+\lambda\int_0^t \nabla_{-\frac{e_1}{N}-e_3}(f)\left(\overline{X}^0_N(s)\right)\overline{F}^0_N(Ns)Z^0_N(Ns)\,\mathrm{d}s \\ &+\eta\int_0^t \nabla_{\frac{e_1}{N}+e_3}(f)\left(\overline{X}^0_N(s)\right)\left(1-\frac{C^N_m}{N}+\frac{G^0_N(Ns)}{N}-\overline{F}^0_N(Ns)\right)\mathrm{d}s \\ &+\alpha_m\int_0^t \nabla_{-\frac{e_1}{N}-e_2}(f)\left(\overline{X}^0_N(s)\right)G^0_N(Ns)\overline{F}^0_N(Ns)\,\mathrm{d}s \\ &+\beta_m\int_0^t \nabla_{\frac{e_1}{N}+e_2}(f)\left(\overline{X}^0_N(s)\right)\left(\frac{C^N_m}{N}-\frac{G^0_N(Ns)}{N}\right)\mathrm{d}s \\ &+\frac{\beta_6}{N}\int_0^t \nabla_{e_3}(f)\left(\overline{X}^0_N(s)\right)\mathrm{d}s+\frac{\delta_6}{N}\int_0^t \nabla_{-e_3}(f)\left(\overline{X}^0_N(s)\right)Z^0_N(Ns)\,\mathrm{d}s\end{aligned} \tag{2.20}$$

holds. Recall that, for $i\in\{1,2,3\}$, e_i is the ith unit vector of $\mathbb{R}^3$.

Lemma 12. *Under Conditions* (2.16) *and* $c_m<1$, *if* f *is a continuous bounded function on* $\mathbb{R}_+\times\mathbb{N}$, *then the martingale* $(M_{f,N}(t)/N^2, t<t^a_0)$ *of Relation* (2.20) *converges in distribution to* 0.

Proof. We take care of one of the six terms of $(\langle M_{f,N}/N^2\rangle(t))$ of Relation (2.41), the arguments are similar for the others, even easier.

$$A_{1,N}(t)\overset{\text{def.}}{=}\frac{\lambda}{N^2}\int_0^t\left[\nabla_{-\frac{e_1}{N}-e_3}(f)\left(\overline{X}^0_N(s)\right)\right]^2\overline{F}^0_N(s)Z^0_N(Ns)\,\mathrm{d}s$$

We note that for $t\geq 0$, $0\leq Z^0_N(t)\leq N+\mathcal{P}_5((0,\beta_6)\times(0,t])$. Consequently, Doob's Inequality shows the convergence of $(M_{f,N}(t)/N^2)$ to 0. The lemma is proved. □

Proposition 13. *Under Conditions* (2.16) *and* $c_m<1$, *and if* $\overline{\Lambda}^{0,a}_\infty$ *is a limiting point of* $\overline{\Lambda}^{0,a}_n$ *with the representation* (2.19) *of Proposition 10, then, if* $\pi^{1,a}_t=\pi^{0,a}_t(\cdot,\mathbb{N}^2)$, *for any* $t<t^a_0$ *and any continuous function* g *on* $\mathbb{R}_+\times\mathbb{N}^2$ *we have*

$$\begin{aligned}&\int_0^t\int g(x,p)\pi^a_s(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s \\ &=\int_0^t\int_{\mathbb{R}_+}\mathbb{E}\left[g\left(x,\mathcal{N}_1\left(\left[0,\rho_m\frac{c_m}{x}\right]\right),\mathcal{N}_2\left(\left[0,\rho_1\frac{1-c_m-x}{x}\right]\right)\right)\right]\pi^{1,a}_s(\mathrm{d}x)\,\mathrm{d}s,\end{aligned} \tag{2.21}$$

where $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *are two independent Poisson processes on* $\mathbb{R}_+$ *with rate* 1 *and*

$$\rho_1=\frac{\eta}{\lambda}\ \textit{and}\ \rho_m=\frac{\beta_m}{\alpha_m}. \tag{2.22}$$

Relation (2.21) states that, for almost all $t<t_0^a$, π_t conditioned on the first coordinate x is a product of two Poisson distributions with respective parameters $\rho_m c_m/x$ and $\rho_1(1-c_m-x)/x$.

Proof. Let $(\overline{\Lambda}^0_{N_k})$ be a subsequence of $(\overline{\Lambda}^0_N)$ converging to some $\overline{\Lambda}^0_\infty$ of the form (2.19). By letting k go to infinity in Relation (2.20), with Lemmas 9, 11 and 12, we obtain that there exists an event $\mathcal{E}_1$ with $\mathbb{P}(\mathcal{E}_1)=1$ on which the relation

$$\int_0^t \int_{\mathbb{R}_+\times\mathbb{N}^2} (\eta\,(1-c_m-x)\,\nabla_{e_3}(f)(x,p)+\lambda x p_2 \nabla_{-e_3}(f)(x,p))\,\pi_s^a(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s$$
$$+\int_0^t \int_{\mathbb{R}_+\times\mathbb{N}^2} (\beta_m c_m \nabla_{e_2}(f)(x,p)+\alpha_m \nabla_{-e_2}(f)(x,p) p_1 x)\,\pi_s^a(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s = 0,$$

holds for all $t\leq T$ and for all functions $f\in\mathcal{C}_c(\mathbb{R}_+\times\mathbb{N}^2)$, by using the separability property of this space of functions for the uniform norm. If $f(x,p)=f_1(x)f_2(p)$, this relation can be rewritten as

$$\int_0^t \int_{\mathbb{R}_+\times\mathbb{N}^2} f_1(x)\Omega[x](f_2)(p)\pi_s^a(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s = 0$$

where, for $h:\mathbb{N}^2\to\mathbb{R}_+$ and $p=(p_1,p_2)\in\mathbb{N}^2$,

$$\Omega[x](h)(p) = \beta_m c_m \nabla_{e_1}(h)(p)+\alpha_m p_1 x \nabla_{-e_1}(h)(p)$$
$$+\eta(1-c_m-x)\nabla_{e_2}(h)(p)+\lambda x p_2 \nabla_{-e_2}(h)(p).$$

$\Omega[x]$ is the jump matrix of two independent birth and death processes $(Y_1(t))$ and $(Y_2(t))$ as in Lemma 3 with parameters $\kappa_i=\beta_m c_m$, $\kappa_o=\alpha_m x$ for $(Y_1(t))$ and $\kappa_i=\eta(1-c_m-x)$, $\kappa_o=\lambda x$ for $(Y_2(t))$.

Consequently, for almost all $t\leq T$, the relation

$$\int_{\mathbb{R}_+\times\mathbb{N}^2} f_1(x)\Omega[x](f_2)(p)\pi_t^a\,\mathrm{d}x,\mathrm{d}p) = 0$$

holds. Hence, if $\widetilde{\pi}_t^a(\cdot|x)$ is the conditional probability on $\mathbb{N}^2$ of $\pi_t^a(\mathrm{d}x,\mathrm{d}p)$ given x, we have

$$\int_{\mathbb{R}_+} f_1(x) \int_{\mathbb{N}^2} \Omega[x](f_2)(p)\widetilde{\pi}_t^a(\mathrm{d}p|x)\pi^{1,a}(\mathrm{d}x) = 0,$$

we deduce that the relation

$$\int_{\mathbb{N}^2} \Omega[x](f_2)(p)\widetilde{\pi}_t^a(\mathrm{d}p|x) = 0$$

holds $\pi_t^{1,a}(\mathrm{d}x)$ almost surely, for all functions f_2 with finite support on $\mathbb{N}^2$. Consequently, $\pi^{1,a}(\mathrm{d}x)$ almost surely, $\widetilde{\pi}_t^a(\mathrm{d}p|x)$ is the invariant distribution associated to the Q-matrix $\Omega[x]$. The proposition is proved. □

We fix (N_k) an increasing sequence such the sequence $(\overline{\Lambda}^{0,a}_{N_k})$ is converging in distribution to the law of $\overline{\Lambda}^{0,a}_\infty$ with a representation given by Relations (2.19) and (2.21).

2.4.3 Averaging Principle

We define, for $t \geq 0$,

$$\widetilde{Z}_N^0(t) = S_N^0(t) + Z_N^0(t),$$

$\widetilde{Z}_N^0(t)$ is in fact the total number of 6S RNAs (free or paired) of the system at time t. Using the SDEs (2.9) and (2.10), we have

$$\frac{\widetilde{Z}_N^0(Nt)}{N} = M_{Z,N}(t) + \frac{\widetilde{Z}_N^0(0)}{N} + \beta_6 t - \delta_6 \int_0^t Z_N^0(Ns)\,\mathrm{d}s, \tag{2.23}$$

where $(M_{Z,N}(t))$ is a local martingale whose previsible increasing process is given by

$$(\langle M_{Z,N} \rangle (t)) = \left(\frac{1}{N} \left(\beta_6 t + \delta_6 \int_0^t Z_N^0(Ns)\,\mathrm{d}s \right) \right). \tag{2.24}$$

Proposition 14. *Under Conditions* (2.16) *and* $c_m < 1$*, for the convergence in distribution*

$$\lim_{k \to +\infty} \left(\int_0^t Z_{N_k}(N_k s)\,\mathrm{d}s, t < t_0^a \right) = \left(\rho_1 \int_0^t \int_{\mathbb{R}_+} \frac{1 - c_m - x}{x} \pi_s^1(\mathrm{d}x)\,\mathrm{d}s, t < t_0^a \right),$$

with $t_0^a = (\overline{f}_0 - a)/\beta_6$ *and* $\rho_1 = \eta/\lambda$*. Furthermore,* $(M_{Z,N}(t), t < t_0^a)$ *is converging to* 0*.*

Proof. The convergence of the sequence of stochastic processes $(M_{Z,N}(t), t < t_0^a)$ to 0 is a consequence of Relations (2.23) and (2.24), and of Doob's Inequality. For $0 \leq s \leq t$, the coupling (2.18) and Cauchy-Schwartz' Inequality give

$$\begin{aligned}
\mathbb{E}\left(\left(\mathbb{1}_{\{\tau_N(a) > t\}} \int_s^t Z_{N_k}(Ns)\,\mathrm{d}s \right)^2 \right) &\leq (t-s)\mathbb{E}\left(\mathbb{1}_{\{\tau_N(a) > t\}} \int_s^t Z_{N_k}^a(s)^2\,\mathrm{d}s \right) \\
&\leq (t-s)\mathbb{E}\left(\int_s^t Y_Z(Ns)^2\,\mathrm{d}s \right) \leq (t-s)^2 \sup_{u \geq 0} \mathbb{E}\left(Y_Z(u)^2 \right).
\end{aligned}$$

We now use the Kolmogorov-Čentsov's criterion, see Theorem 2.8 and Problem 4.11, page 64 of Karatzas and Shreve [41] and Lemma 9 to show that the sequence of stochastic processes

$$\left(\int_0^t Z_{N_k}(Ns)\,\mathrm{d}s, t < t_0^a \right)$$

is tight for the convergence in distribution.

Lemma 11 gives the convergence in distribution

$$\begin{aligned}
&\lim_{k \to +\infty} \left(\int_0^t Z_{N_k}(N_k s) \wedge K\,\mathrm{d}s, t < t_0^a \right) \\
&\qquad = \left(\int_0^t \int_{\mathbb{R}_+} \mathbb{E}\left(\mathcal{N}_1\left(0, \rho_1 \frac{1 - c_m - x}{x} \right) \wedge K \right) \pi_s^1(\mathrm{d}x)\,\mathrm{d}s, t < t_0^a \right).
\end{aligned}$$

By using again Relation (2.18), we have

$$\mathbb{E}\left(\int_0^{t_0^a} Z_{N_k}(N_k s)\mathbb{1}_{\left\{Z_{N_k}(N_k s)\geq K\right\}}\,\mathrm{d}s\right) \leq \mathbb{E}\left(\int_0^{t_0^a} Y_Z({N_k}^2 s)\mathbb{1}_{\left\{Y_Z(N_k^2 s)\geq K\right\}}\,\mathrm{d}s\right)$$

and the convergence in distribution of $(Y_Z(t))$, as t goes to infinity, to $Y_Z(\infty)$ a random variable with a Poisson distribution with parameter $\rho_Z{=}\kappa_{i,Z}/\kappa_{o,Z}$ give

$$\lim_{k\to+\infty}\mathbb{E}\left(\int_0^{t_0^a} Y_Z(N_k^2 s)\mathbb{1}_{\left\{Y_Z(N_k^2 s)\geq K\right\}}\,\mathrm{d}s\right) = t_0^a\mathbb{E}\left(Y_Z(\infty)\mathbb{1}_{\{Y_Z(\infty)\geq K\}}\right).$$

It is then easy to obtain the first convergence by letting K go to infinity.

The proposition is proved. □

Relation (2.23) therefore shows that, on the time interval $I_a{=}[0,t_0^a)$, the sequence of processes

$$\left(\frac{\widetilde{Z}^0_{N_k}(N_k t)}{N_k}\right)$$

is converging in distribution. Since

$$\left(\frac{\widetilde{Z}^0_{N_k}(N_k t)}{N_k}\right) = \left(1-\frac{F^0_{N_k}(N_k t)}{N_k}-\frac{C^{N_k}_m}{N_k}+\frac{G^0_{N_k}(N_k t)}{N_k}+\frac{Z^0_{N_k}(N_k t)}{N_k}\right)$$

with Lemma 9, we therefore obtain that the sequence of processes $(F_{N_k}(N_k t)/N_k)$ is converging in distribution to some process $(\overline{f}(t))$ on I_a. In particular, for $t{<}t_0^a$ and $g{\in}\mathcal{C}_c(\mathbb{R}_+)$, we have

$$\int_0^t\int g(x)\pi^1_s(\mathrm{d}x)\,\mathrm{d}s = \int_0^t g\left(\overline{f}(s)\right)\mathrm{d}s,$$

hence, π^1_s is in fact the Dirac measure at $\overline{f}(s)$ for $s{<}t_0^a$.

Relation (2.23) gives that, on the time interval I_a and under the initial conditions (2.16), then the sequence of processes $(\overline{F}_{N_k}(t))$ is converging in distribution to $(\overline{f}(t))$ such that

$$\begin{aligned}1-\overline{f}(t) = 1-\overline{f}_0+\beta_6 t-\delta_6\rho_1\int_0^t\int_{\mathbb{R}_+}\frac{1-c_m-x}{x}\pi^1_s(\mathrm{d}x)\,\mathrm{d}s\\ = 1-\overline{f}_0+\beta_6 t-\delta_6\rho_1\int_0^t\frac{1-c_m-\overline{f}(s)}{\overline{f}(s)}\,\mathrm{d}s \quad (2.25)\end{aligned}$$

with Proposition 14 and Notation (2.22).

Hence, by uniqueness of the solution of the integral equation,

$$\overline{f}(t) = \overline{f}_0-\delta_6(\rho_6+\rho_1)t+\delta_6\rho_1(1-c_m)\int_0^t\frac{1}{\overline{f}(s)}\,\mathrm{d}s$$

holds on I_a, for the convergence in distribution, we thus have

$$\lim_{N\to+\infty}\left(\overline{F}_N(t), t\in I_a\right) = \left(\overline{f}(t), t\in I_a\right).$$

The equilibrium point of $(\overline{f}(t))$ is $\overline{f}_\infty=\rho_1(1-c_m)/(\rho_6+\rho_1)$, if $\overline{f}_0<\overline{f}_\infty$, then $\overline{f}(t)\geq\overline{f}_0$ for all $t\geq 0$, and otherwise $\overline{f}(t)\geq\overline{f}_\infty$. By induction, this implies that the convergence in distribution of $(\overline{F}_N(t))$ can be extended on time intervals $(0, nt_0^a)$, for all $n\geq 1$ and, consequently, on $\mathbb{R}_+$. We summarize our results.

Theorem 15 (Law of Large Numbers)**.** *If*

$$\lim_{N\to+\infty}\frac{F_N^0(0)}{N} = \overline{f}_0\in(0, 1-c_m),$$

and $(G_N(0)=, Z_N(0)=(m_0, z_0)$*, then, for the convergence in distribution,*

$$\lim_{N\to+\infty}\left(\frac{F_N^0(Nt)}{N}\right) = (\overline{f}(t)),$$

where $(\overline{f}(t))$ *is the solution of the ODE*

$$\frac{\mathrm{d}\overline{f}}{\mathrm{d}t}(t) = -\delta_6\left(\rho_6+\rho_1\right)+\delta_6\rho_1(1-c_m)\frac{1}{\overline{f}(t)}, \tag{2.26}$$

where ρ_6 *and* ρ_1 *are defined by Relation* (2.22)*.*

We summarize the results obtained for the convergence in distribution of the occupation measures $(\overline{\Lambda}_N^0)$. This is an extension of Proposition 13.

Corollary 16. *Under the conditions of Theorem 15, the sequence of empirical distributions* $(\overline{\Lambda}_N^0)$ *converge in distribution to the measure* $\overline{\Lambda}^0$ *such that*

$$\left\langle\overline{\Lambda}^0, g\right\rangle = \int_{\mathbb{R}_+}\mathbb{E}\left(g\left(s, \overline{f}(s), \mathcal{N}_1\left(\left[0, \rho_m\frac{c_m}{\overline{f}(s)}\right]\right), \mathcal{N}_2\left(\left[0, \rho_1\frac{1-c_m-\overline{f}(s)}{\overline{f}(s)}\right]\right)\right)\right)\mathrm{d}s,$$

for any continuous function g *on* $\mathbb{R}_+^2\times\mathbb{N}^2$*, where* $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *are two independent Poisson processes on* $\mathbb{R}_+$ *with rate* 1 *and* $(\overline{f}(t))$ *is the solution of Relation* (2.26) *with* $\overline{f}(0)=\overline{f}_0$*.*

2.5 Exponential Phase

Throughout this section, Conditions (2.3) and of exponential phase of Definition 1 hold. Heuristically, if there are sufficiently many polymerases, there will be an accumulation of them in the elongation phase of rRNAs and, therefore, the output rate of all types of rRNAs is maximal. The goal of this section is to prove precise results for this assertion.

Under this condition, for any $1\leq j\leq J$, the initiation rate $\alpha_{r,j}$ of rRNA of type j, is larger than $\beta_{r,j}$, the rate at which an rRNA of type j grows.

A coupling

We introduce a coupling to study the occupancy of the places for transcription of rRNAs. The idea is quite simple: for $1 \leq j \leq J$, as long as $R_j^N(t)$ is strictly less than $C_{r,j}^N$, when $U_j^N(t)=1$, a new polymerase is added for transcription after an exponential with parameter $\alpha_{r,j}$ and if at that time $F_N(t)$ is positive, then the variable $U_j^N(t)$ remains at 1. See the part of transcription of rRNAs in the Q-Matrix of our process in Section 2.2.2.

Otherwise, if $F_N(t)=0$, there is a total of at least $A_N \stackrel{\text{def.}}{=} N - C_{r,1}^N - \cdots - C_{r,J}^N - J$ polymerases either in transcription of mRNAs or sequestered. If $\delta = \min(\eta, \beta_m)$, the duration of time after which there will be a free polymerase which can be accommodated by the jth promoter of rRNAs, with probability at least $1/J$, is stochastically bounded by an exponential random variable with parameter $2\delta A_N$. Hence, if $F_N(t)=0$ and $U_j^N(t)=0$, then $U_j(t)$ returns to 1 after a duration whose distribution is stochastically bounded by an exponential random variable with parameter $2\delta A_N/J$.

We choose N_0 sufficiently large, so that

$$\frac{1}{\alpha_{r,j}} + \frac{J}{\delta N_0} < \frac{1}{\beta_{r,j}}, \qquad \forall 1 \leq j \leq J \text{ with } \delta = \min(\eta, \beta_m). \tag{2.27}$$

We are interested in the behavior of $(Q_j^N(t)) \stackrel{\text{def.}}{=} (C_{r,j}^N(t) - R_j^N(t)), 1 \leq j \leq J$, which measures the congestion of the transcription of the rRNAs. The above coupling shows that if $N \geq N_0$, it can be stochastically bounded by independent queueing processes $(\overline{Q}_j(t))$, $1 \leq j \leq J$ characterized as follows: for $1 \leq j \leq J$,

— the arrivals of customers is a Poisson process with rate $\beta_{r,j}$.

— The distribution of the service of a customer is the distribution of the sum of two independent exponential random variables with respective parameters $\alpha_{r,j}$ and $\delta N_0/J$. The service will be seen as the sum of the duration of a phase $\alpha_{r,j}$ and a phase $\delta N_0/J$.

This is an $M/G/1$ queue, see Chapter 2 of Robert [61]. It has a Markovian representation as $(I_j(t), \overline{Q}_j(t))$ where $I_j(t) \in \{1,2\}$, $I_j(t)=1$ indicates that the customer being served is in phase $\alpha_{r,j}$ and $I_j(t)=2$ when it is in the phase $\delta N_0/J$.

Under Condition (2.27), $(\overline{Q}(t)) = ((I_j(t), \overline{Q}_j(t)), 1 \leq j \leq J)$ is a positive recurrent Markov process, since the coordinates are independent positive recurrent Markov processes. In particular if $\overline{Q}(0) \in (\{0,1\} \times \mathbb{N})^J$, then

$$\inf \left\{ t>0 : \overline{Q}(t) = ((1,0), j=1, \ldots, J) \right\}$$

is almost surely finite and integrable and for any $\varepsilon>0$ and $T>0$, there exists K such that

$$\mathbb{P}\left(\sup_{0 \leq t \leq T} \max_{1 \leq j \leq J} \overline{Q}_j(t) \geq K \right) \leq \varepsilon.$$

Furthermore if

$$\overline{\tau}_j^N = \inf\{t{>}0 : \overline{Q}_j(t){=}0\}, \text{ with } \overline{Q}_j(0){=}C_{r,j}^N,$$

then it is not difficult to show, with the classical law of large numbers, that, if $i{\in}\{0,1\}$,

$$\lim_{K\to+\infty} \frac{\mathbb{E}_{(i,K)}(\overline{\tau}_j^N)}{N} = c_{r,j} \Big/ \left(\frac{1}{1/\alpha_{r,j} + J/(\delta N_0)} - \beta_{r,j} \right).$$

We have thus proved the following proposition which shows that in the exponential phase, the transcription of rRNAs is essentially congested.

Theorem 17 (Saturation of Transcription of rRNAs)**.** *If Conditions* (2.3) *and* (2.4) *hold and if* $F_N(0){=}N$, $Z_N(0){=}0$ *and* $(U_N(0), R_N(0)){=}(0,0)$, *i.e. all polymerases are initially free, then the variable* τ_N^e *defined by*

$$\tau_N^e \stackrel{\text{def.}}{=} \inf\{t{>}0 : R_{N,j}(t){=}C_{r,j}^N, \forall 1{\le}j{\le}J\}, \tag{2.28}$$

is almost surely finite and

$$\sup_N \frac{\mathbb{E}(\tau_N^e)}{N} < +\infty.$$

For any $\varepsilon{>}0$ *and* $T{>}0$, *there exists* K *such that*

$$\mathbb{P}_{((1,0))} \left(\sup_{0\le t\le T} \max_{1\le j\le J} C_{r,j}^N - R_j^N(t) \ge K \right) \le \varepsilon \tag{2.29}$$

The variable τ_N^e is the first time when all places for transcription of rRNAs are occupied, i.e. the first instant when this part of the system is saturated. Our proposition gives an upper bound linear in N for the average value of this random variable when Condition (2.4) holds.

Now we investigate the asymptotic behavior of the remaining part of the system after time τ_N^e. We introduce

$$\left\langle \Lambda_N^F, g \right\rangle \stackrel{\text{def.}}{=} \int g\left(s, F_N(s)\right)) \, \mathrm{d}s \text{ and } \left\langle \Lambda_N^{0,F}, g \right\rangle \stackrel{\text{def.}}{=} \int g\left(s, F_{A_N}^0(s)\right)) \, \mathrm{d}s,$$

if g is a continuous function with compact support on $\mathbb{R}_+{\times}\mathbb{N}$, where A_N defined by Relation (2.6) is the number of polymerases available when transcription of rRNA is saturated. The process $(F_{A_N}^0(t))$ is the solution of Relation (2.8) whose initial condition is the same as the process $(F_N(t), S_N(t), Z_N(t))$.

Lemma 18 (Coupling with the Auxiliary Process)**.**
If $(F_N(0), S_N(0), Z_N(0)){=}(f,s,z){\in}\mathbb{N}^3$ *and if* $(U_N(0), R_N(0)){=}((1, C_{r,j}^N))$, *then for any* $g{\in}\mathcal{C}_c(\mathbb{R}_+{\times}\mathbb{N})$,

$$\lim_{N\to+\infty} \left| \mathbb{E}\left(\left\langle \Lambda_{A_N}^{0,F}, g \right\rangle \right) - \mathbb{E}\left(\left\langle \Lambda_N^F, g \right\rangle \right) \right| = 0.$$

Proof. From Relation (2.29), we know that for K sufficiently large, the probability of the event

$$\mathcal{E}_K \stackrel{\text{def.}}{=} \left\{ \sup_{t \leq T} A_N - \left(N - \sum_1^J U_j^N(t) + R_j^N(t) \right) \leq K \right\}$$

is close to 1.

Given our initial state, at time 0 there are A_N polymerases either sequestered, free or in transcription of an mRNA. On the event $\mathcal{E}_K$, on the time interval $[0,T]$, there may be at most K additional polymerases. Since they enter this part of the system as free, at rate at least $C_m^N - (N - C_{r,1}^N \cdots - C_{r,J}^N)$, they go into transcription of an mRNA. Note that, almost surely, any of these K polymerases may return a finite number of times as free on $[0,T]$. Hence, with high probability, their contribution to the integral defining the occupation measure is arbitrarily small as N gets large, and so is their impact on the random variable $(F_N(t), S_N(t), Z_N(t))$. □

We can now state convergence results for the number of free and sequestered polymerases. It is a direct consequence of the arguments of the proof of the last lemma and Proposition 5. It shows that in this case, basically, the number of free polymerases has a Poisson distribution and the process of the number of sequestered polymerases and free 6S RNAs is a positive recurrent Markov process on $\mathbb{N}^2$.

Theorem 19 (Free/Sequestered Polymerases and 6S RNAs)**.** *Under Conditions* (2.3) *and* (2.4) *and if* $(F_N(0), S_N(0), Z_N(0)) = (f,s,z) \in \mathbb{N}^3$ *and* $(U_N(0), R_N(0)) = ((1, C_{r,j}^N))$*, then, for the convergence in distribution,*

$$\lim_{N \to +\infty} \int g(s, F_N(s)))\, \mathrm{d}s = \int_0^{+\infty} \mathbb{E}\left(g\left(u, \mathcal{N}_1\left(0, \rho_m\right)\right)\right) \mathrm{d}u,$$

for any $g \in \mathcal{C}_c(\mathbb{R}_+ \times \mathbb{N})$*, where* $\mathcal{N}_1$ *is a Poisson process with rate* 1 *and*

$$\rho_m = \frac{\beta_m(1-c_r)}{\alpha_m(c_m + c_r - 1)}, \quad c_r \stackrel{\text{def.}}{=} \sum_{j=1}^J c_{r,j}.$$

Furthermore, the sequence of processes $(S_N(t), Z_N(t))$ *converges in distribution for the Skorohod topology to a jump process* $(S(t), Z(t))$ *on* $\mathbb{N}^2$ *whose transition rates are given by*

$$(s,z) \longrightarrow (s,z) + \begin{cases} (1,-1) & \lambda \rho_m z, \\ (-1,1) & \eta s, \end{cases} \quad \begin{cases} (0,1) & \beta_6, \\ (0,-1) & \delta_6 z. \end{cases}$$

Note that the process $(S(t), Z(t))$ is a positive recurrent Markov process. Indeed, if, for $a > 0$,

$$H_a(s,z) \stackrel{\text{def.}}{=} as + z,$$

then it is easily seen that H_a is a Lyapunov function for this Markov process if $a \in \mathbb{R}_+$ is chosen so that

$$1 < a < 1 + \frac{\delta_6}{\lambda \rho_m},$$

see Proposition 8.14 of Robert [61].

2.6 Stationary Phase

Conditions (2.3) and of stationary phase of Definition 1 now hold. For any type j of rRNA, the initiation rate $\alpha_{r,j}$ is less than its production rate.

A coupling

As in Section 2.5 we introduce a simple coupling to study the occupancy of the slots for transcription of rRNAs. Since a polymerase enters in elongation phase of an rRNA of type $j \in \{1, \ldots, J\}$ at rate at most $\alpha_{r,j}$, it is easy to construct a coupling with J independent $M/M/1$ processes $(Q_j(t))$ with respective input rate $\alpha_{r,j}$ and service rate $\beta_{r,j}$, so that the relations

$$R_j^N(t) \leq Q_j^N(t), \quad \forall t \geq 0, 1 \leq j \leq J,$$

hold. See Chapter 5 of Robert [61] for example. The following proposition is a direct consequence of this coupling and the fact that, for the convergence in distribution, the hitting time of p starting from a fixed initial state is exponential with respect to p, for p large. See Proposition 5.16 of [61]

Proposition 20. *Under Conditions* (2.3) *and* (2.5)*, and if* $F_N(0)=N$*,* $Z_N(0)=0$ *and* $(U_N(0), R_N(0))=(0,0)$*, all polymerases are initially free, then the variable* τ_N^s *defined by*

$$\tau_N^s \stackrel{\text{def.}}{=} \inf\{t>0 : R_{N,j}(t) = 0, \forall 1 \leq j \leq J\}, \tag{2.30}$$

is almost surely finite and

$$\sup_N \frac{\mathbb{E}(\tau_N^s)}{N} < +\infty.$$

Lemma 21. *Under Condition* (2.5) *then, for any* $K>0$*,*

$$\lim_{N \to +\infty} \mathbb{P}_{(u,r)} \left(\sup_{t \leq NT} \frac{R_j^N(t)}{\ln(N)^2} > K \right) = 0.$$

Proof. This is a simple consequence of the independence of the $(Q_j(t))$ and of Proposition 5.11 of Robert [61]. □

The above result shows that few polymerases are in transcription of an rRNA, hence the results of Section 2.4 on the auxiliary process can be used, in particular Theorem 15.

Theorem 22 (Asymptotic Behavior in Stationary Phase)**.** *Under Conditions* (2.3) *and* (2.5)*, and the initial state such that*

$$\lim_{N\to+\infty}\left(\frac{F_N(0)}{N},\frac{S_N(0)}{N}\right)=\left(\overline{f},\overline{s}\right)\in[0,1]^2,\ \text{with } \overline{f}+\overline{s}=1-c_m,$$

and $(U_N(0),R_N(0))=(u,r)\in(\{0,1\}\times\mathbb{N})^J$ *then, for the convergence the sequence of processes*

$$\lim_{N\to+\infty}\left(\frac{F_N(t)}{N},\frac{S_N(t)}{N})\right)=\left(\overline{f}(t),1-c_m-\overline{f}(t)\right),$$

where $(\overline{f}(t))$ *is the solution of the ODE*

$$\frac{\mathrm{d}\overline{f}}{\mathrm{d}t}(t)=-\delta_6\left(\rho_6+\rho_1\right)+\delta_6\rho_1(1-c_m)\frac{1}{\overline{f}(t)},\tag{2.31}$$

with $\rho_1=\eta/\lambda$ *and* $\rho_6=\beta_6/\delta_6$.

If $g\in\mathcal{C}_c(\mathbb{R}_+\times\mathbb{N})$ *then, for the convergence in distribution,*

$$\lim_{N\to+\infty}\left(\int_{\mathbb{R}_+}g(t,Z_N(t))\,\mathrm{d}t\right)=\int_{\mathbb{R}_+}\mathbb{E}\left[g\left(t,\mathcal{N}_1\left(\left[0,\rho_1\frac{1-c_m-\overline{f}(t)}{\overline{f}(t)}\right]\right)\right)\right]\mathrm{d}t,$$

where $\mathcal{N}_1$ *is a Poisson processes on* $\mathbb{R}_+$ *with rate* 1.

In particular, the asymptotic fraction of free polymerases is

$$\frac{\rho_1}{\rho_6+\rho_1}(1-c_m),$$

and, in this state, the number of free 6S RNAs has a Poisson distribution with parameter ρ_6.

2.7 Sub-critical Case

It is assumed throughout this section that $c_m>1$ holds. We give a sketch of the proof of the averaging principle at the basis of the proof of Proposition 5 for the sake of completeness. The analogue of this result in the super-critical case in Section 2.4 is quite different and more challenging. The corresponding tightness property is less clear in this case, in particular the definition of occupation measures has to include the slow processes. The arguments of the proofs of Section 2.4 can be used in the same way. As it will be seen, it is easy to show that the sequences of "slow" processes $(S_N^0(t)/N)$ and $(Z_N^0(t)/N)$ are tight.

Recall that μ_N is the occupation measure defined by Relation (2.12). For $K>0$, with the same notations as in the proof of Lemma 4, Relation (2.13) gives the inequality

$$\mathbb{E}\left(\langle\mu_N,[0,t]\times[0,K]\rangle\right)\geq\int_0^t\mathbb{P}(Y(Ns)\leq K)\,\mathrm{d}s=\frac{1}{N}\int_0^{Nt}\mathbb{P}(Y(s)\leq K)\,\mathrm{d}s.$$

Since $(Y(t))$ is converging in distribution to a Poisson distribution with parameter a/b, for any $\varepsilon>0$ and $t>0$, there exists K_0 and N_0 such that if $K\geq K_0$ and $N\geq N_0$, then $\mathbb{E}\left(\langle\mu_N,[0,t]\times[0,K]\rangle\right)>(1-\varepsilon)t$. Lemma 1.3 and 1.4 of Kurtz [45] show that the sequence (μ_N) of random measures is tight and any limiting point μ_∞ can be expressed as

$$\langle\mu_\infty,g\rangle=\int_{\mathbb{R}_+\times\mathbb{N}}g\,(u,x)\,\pi_u(\mathrm{d}x)\,\mathrm{d}u$$

where (π_u) is a previsible process with values in the state space of probability distributions on $\mathbb{N}$.

2.7.1 Proof of Proposition 5

By integrating Relations (2.9) and (2.10), we obtain the identities, for $t\geq 0$,

$$S_N^0(t)=S_N^0(0)+M_6^N(t)-\eta\int_0^t S_N^0(s)\,\mathrm{d}s+\lambda\int_0^t F_N^0(s)Z_N^0(s)\,\mathrm{d}s, \tag{2.32}$$

$$\begin{aligned}Z_N^0(t)=Z_N^0(0)+M_Z^N(t)+\beta_6 t-\delta_6\int_0^t Z_N^0(s)\,\mathrm{d}s\\ +\eta\int_0^t S_N^0(s)\,\mathrm{d}s-\lambda\int_0^t F_N^0(s)Z_N^0(s)\,\mathrm{d}s,\end{aligned} \tag{2.33}$$

where $(M_6^N(t))$ and $(M_Z^N(t))$ are martingales whose previsible increasing processes are given by

$$\left\langle M_6^N\right\rangle(t)=\eta\int_0^t S_N^0(s)\,\mathrm{d}s+\lambda\int_0^t F_N^0(s)Z_N^0(s)\,\mathrm{d}s, \tag{2.34}$$

$$\left\langle M_Z^N\right\rangle(t)=\beta_6 t+\delta_6\int_0^t Z_N^0(s)\,\mathrm{d}s+\eta\int_0^t S_N^0(s)\,\mathrm{d}s+\lambda\int_0^t F_N^0(s)Z_N^0(s)\,\mathrm{d}s. \tag{2.35}$$

Relations (2.34) and (2.35), Relation (2.13), and Doob's Inequality show that, for convergence in distribution, then

$$\lim_{N\to+\infty}\left(\frac{M_6^N(t)}{N}\right)=\lim_{N\to+\infty}\left(\frac{M_Z^N(t)}{N}\right)=0.$$

We note that, for $t\geq 0$, $S_N^0(t)\in[0,N]$ and $0\leq Z_N^0(t)\leq N+\mathcal{P}_5((0,\beta_6)\times(0,t])$ by Relation (2.10). Relations (2.32) and (2.33), and the criterion of the modulus of continuity, see Billingsley [10], give that the sequence of processes $\left(S_N^0(t)/N,Z_N^0(t)/N\right)$ is tight for the convergence in distribution associated to the uniform norm on compact sets of $\mathbb{R}_+$.

We can therefore take a subsequence of $\left(\mu_N,\left(S_N^0(t)/N\right),\left(Z_N^0(t)/N\right)\right)$ with indices (N_k) converging in distribution to $(\mu_\infty,(s(t)),(z(t)))$, where $(s(t))$ and $(z(t))$ are continuous processes.

If $f \in \mathcal{C}_c\left(\mathbb{N}\times\mathbb{R}_+^2\right)$, Relation (2.8) gives the identity

$$
\begin{aligned}
f\left(F_{N_k}^0(t), \frac{S_{N_k}^0(t)}{N_k}, \frac{Z_{N_k}^0(t)}{N_k}\right) &= f\left(f_{N_k}, s_{N_k}, z_{N_k}\right) + M_f^{N_k}(t) \\
&\quad + \beta_m \int_0^t \nabla_{e_1}(f)\left(X_{N_k}^0(s)\right)\left(N_k - F_{N_k}^0(s) - S_{N_k}^0(s)\right) \mathrm{d}s \\
&\quad + \alpha_m \int_0^t \nabla_{-e_1}(f)\left(X_{N_k}^0(s)\right)\left(C_m^{N_k} - N_k + F_{N_k}^0(s) + S_{N_k}^0(s)\right) F_{N_k}^0(s)\,\mathrm{d}s \\
&\quad + \lambda \int_0^t \nabla_{-e_1+\frac{e_2}{N_k}-\frac{e_3}{N_k}}(f)\left(X_{N_k}^0(s)\right) F_{N_k}^0(s) Z_{N_k}^0(s)\,\mathrm{d}s \\
&\quad + \eta \int_0^t \nabla_{e_1-\frac{e_2}{N_k}+\frac{e_3}{N_k}}(f)\left(X_{N_k}^0(s)\right) S_{N_k}^0(s)\,\mathrm{d}s \\
&\quad + \beta_6 \int_0^t \nabla_{\frac{e_3}{N_k}}(f)\left(X_{N_k}^0(s)\right)\mathrm{d}s + \delta_6 \int_0^t \nabla_{-\frac{e_3}{N_k}}(f)\left(X_{N_k}^0(s)\right) Z_{N_k}^0(s)\,\mathrm{d}s,
\end{aligned}
$$

with the notation $\nabla_a(f)(x) = f(x+a) - f(x)$, for a and $x \in \mathbb{N}\times\mathbb{R}_+^2$.

With the same arguments as for the martingales $(M_S^{N_k}(t))$ and $(M_Z^{N_k}(t))$, the process $(M_f^{N_k}(t))$ is converging in distribution to 0. By dividing by N_k the last relation, and by letting k go to infinity, we get

$$
\begin{aligned}
&\int_0^t \nabla_{e_1}(f)(x, s(u), z(u))\left(\beta_m - (\beta_m - \eta)\, s(u)\right)\pi_u(\mathrm{d}x)\,\mathrm{d}u \\
&\quad + \int_0^t \nabla_{-e_1}(f)(x, s(u), z(u))\left(\alpha_m(c_m - 1 + s(u)) + \lambda z(u)\right) x \pi_u(\mathrm{d}x)\,\mathrm{d}u = 0,
\end{aligned}
$$

and therefore

$$
\int_0^t \int_{\mathbb{N}} \Omega_{s(u),z(u)}(g)(x)\pi_u(\mathrm{d}x)\,\mathrm{d}u = 0, \tag{2.36}
$$

with, for s, $z \geq 0$, $s+z<1$ and $x \in \mathbb{N}$,

$$
\begin{aligned}
\Omega_{s,z}(g)(x) &= \Big(\beta_m - (\beta_m - \eta)\, s\Big)\left(g(x+1) - g(x)\right) \\
&\quad + \Big(\alpha_m(c_m - 1 + s) + \lambda z\Big)\left(g(x-1) - g(x)\right),
\end{aligned}
$$

$\Omega_{s,z}$ is the infinitesimal generator of the Markov process $(Y(t))$ of Lemma 3 with $a = a(s,z) = (\beta_m - (\beta_m - \eta)\, s)$ and $b = b(s,z) = \alpha_m(c_m - 1 + s) + \lambda z)$. From Relation (2.36) and with the same methods as in Section 2.4, we obtain that, almost surely,

$$
\int_0^t \int_{\mathbb{N}} g(x)\pi_u(\mathrm{d}x)\,\mathrm{d}u = \int_0^t \int_{\mathbb{N}} g(x)\pi_u(\mathrm{d}x)\,\mathrm{d}u = \int_0^t \mathbb{E}\left(g\left(P_u\right)\right)\mathrm{d}u
$$

holds for all $t>0$ and all functions g with finite support on $\mathbb{N}$, where P_u is a Poisson random variable with parameter $a(s(u), z(u))/b(s(u), z(u))$, $u \geq 0$.

Hence, with similar arguments as in Section 2.4, for $T \geq 0$ such that $s(t)+z(t)<1$ holds for all $t \leq T$, we obtain that the identities

$$s(t) = s_0 - \eta \int_0^t s(u)\,\mathrm{d}u + \lambda \int_0^t z(u) \frac{\beta_m - (\beta_m - \eta)\, s(u)}{\alpha_m(c_m - 1 + s(u)) + \lambda z(u)}\,\mathrm{d}u, \tag{2.37}$$

$$s(t)+z(t) = s_0 + z_0 - \delta_6 \int_0^t z(u)\,\mathrm{d}u., \tag{2.38}$$

hold almost surely, for $t \leq T$. From Relation (2.38) we obtain that $(s(t)+z(t))$ is a non-increasing function, hence $s(t)+z(t) \leq s_0+z_0<1$, for $t \geq 0$, the above system has therefore a unique solution defined on $\mathbb{R}_+$. Since the function $(s(t)+z(t))$ is converging at infinity, Equation (2.38) shows that $(s(t))$ converges at infinity too. By dividing both sides of Relations (2.37) and (2.38) by t and by letting t got to infinity, we deduce that both limits are zero. Proposition 5 is proved.

2.7.2 Proof of Proposition 6

The first assertion on the convergence of the occupation is obtained in the same way but with $(s, z)=(0, 0)$, hence for $u \geq 0$, $s(u)=z(u)=0$, and the operator is

$$\Omega_{(s(u),z(u))}(g)(x) = \beta_m(g(x+1)-g(x)) + \alpha_m(c_m - 1)(g(x-1)-g(x)).$$

Therefore P_u is a Poisson random variable with parameter ρ_m.

Let, for $k \geq 1$, t_k^N be the kth jump of $(Y_N(t)) \stackrel{\text{def.}}{=} (S_N(t), Z_N(t))$ when the initial state is (s, z), there are four random variables A_i, $i \in \{1, 2, 3, 4\}$, to trigger a change of state of $(Y_N(t))$,

a. A_1^N is a random variable such that, for $t \geq 0$,

$$\mathbb{P}\left(A_1^N \geq t \mid (F_N^0(s))\right) = \exp\left(-\lambda z \int_0^t F_N^0(s)\,\mathrm{d}s\right); \tag{2.39}$$

b. A_2, A_3, A_4 are independent exponential random variables with respective parameters ηs, β_6 and $\delta_6 z$,

and, conditionally on $(F_N^0(t))$, the random variables A_1^N, A_i, $i \in \{2, 3, 4\}$ are independent.

Relation (2.39) and the convergence of the sequence (μ_N) of occupation measure of $(F_N^0(t))$ given that A_1^N is converging in distribution to an exponential distribution with parameter $\lambda z \rho_m$.

For $t \geq 0$, we have

$$\begin{aligned}\mathbb{P}_{(s,z)}\left(Y_N\left(t_1^N\right) = (s+1, z-1), t_1^N \geq t\right) &= \mathbb{P}\left(A_1^N \geq t, A_1^N \leq A_2 \wedge A_3 \wedge A_4\right) \\ &= \mathbb{E}\left(\mathbb{1}_{\left\{A_1^N \geq t\right\}} \exp\left(-(\eta s + \beta_6 + \delta_6 z) A_1^N\right)\right),\end{aligned}$$

hence,

$$\lim_{N\to+\infty} \mathbb{P}_{(s,z)} \left(Y_N\left(t_1^N\right) = (s{+}1, z{-}1), t_1^N {\geq} t \right) \\ = \frac{\lambda\rho_m z}{(\lambda\rho_m z{+}\eta s{+}\beta_6{+}\delta_6 z)} e^{-(\lambda\rho_m z+\beta_6+\delta_6 z+\eta s)t},$$

and this last quantity is $\mathbb{P}_{(s,z)}(Y(t_1) = (s{+}1, z{-}1), t_1 \geq t)$, where $(Y(t))$ is the jump process defined in Proposition 6 and (t_i) is the non-decreasing sequence of its instants of jumps. A similar convergence result is obtained in the same manner for the other possibilities for the first jump of $(Y_N(t))$. By induction, one can show that for $k{\geq}1$ and any sequence $(a_i){\in}\mathbb{N}^2$,

$$\lim_{N\to+\infty} \mathbb{P}\left(Y_N(t_i^N) = a_i, 1{\leq}i{\leq}k \right) = \mathbb{P}(Y(t_i) = a_i, 1{\leq}i{\leq}k).$$

We conclude the proof of the convergence by using directly the very definition of the Skorohod topology. See Billingsley [10].

2.8 Super-critical Case

The assumption $c_m{<}1$ holds throughout this section. Technical results used in Section 2.4 are presented here.

Recall that $(\overline{X}_N^0(t)){=}(\overline{F}_N^0(t), G_N^0(Nt), Z_N^0(Nt))$, with

$$G_N^0(t) \stackrel{\text{def.}}{=} C_m^N - \left(N{-}F_N^0(t){-}S_N^0(t)\right).$$

If f be a non-negative Borelian function on $\mathbb{R}_+{\times}\mathbb{N}^2$, the SDEs (2.8), (2.9), and (2.10) give directly the relations

$$\begin{aligned} f\left(\overline{X}_N^0(t)\right) &= f\left(\overline{X}_N^0(0)\right) + M_{f,N}(t) \\ &+ \lambda N \int_0^t \nabla_{-\frac{e_1}{N}-e_3}(f)\left(\overline{X}_N^0(s)\right) F_N^0(Ns) Z_N^0(Ns)\,\mathrm{d}s \\ &+ \eta N \int_0^t \nabla_{\frac{e_1}{N}+e_3}(f)\left(\overline{X}_N^0(s)\right)\left(N{-}C_m^N{+}G_N^0(Ns){-}F_N^0(Ns)\right)\mathrm{d}s \\ &+ \alpha_m N \int_0^t \nabla_{-\frac{e_1}{N}-e_2}(f)\left(\overline{X}_N^0(s)\right) G_N^0(Ns) F_N^0(Ns)\,\mathrm{d}s \\ &+ \beta_m N \int_0^t \nabla_{\frac{e_1}{N}+e_2}(f)\left(\overline{X}_N^0(s)\right)\left(C_m^N{-}G_N^0(Ns)\right)\mathrm{d}s \\ &+ \beta_6 N \int_0^t \nabla_{e_3}(f)\left(\overline{X}_N^0(s)\right)\mathrm{d}s + \delta_6 N \int_0^t \nabla_{-e_3}(f)\left(\overline{X}_N^0(s)\right) Z_N^0(Ns)\,\mathrm{d}s, \end{aligned} \tag{2.40}$$

where, for $i{\in}\{1,2,3\}$, e_i is the ith unit vector of $\mathbb{R}^3$, and $(M_{f,N}(t))$ is a local

martingale and its previsible increasing process is given by

$$\begin{aligned}\langle M_{f,N}\rangle(t) &= \lambda N\int_0^t \nabla_{-\frac{e_1}{N}-e_3}(f)\left(\overline{X}_N^0(s)\right)^2 F_N^0(Ns)Z_N^0(Ns)\,\mathrm{d}s \\ &+ \eta N\int_0^t \nabla_{\frac{e_1}{N}+e_3}(f)\left(\overline{X}_N^0(s)\right)^2 (N-C_m^N+G_N^0(Ns)-F_N^0(Ns))\,\mathrm{d}s \\ &+ \alpha_m N\int_0^t \nabla_{-\frac{e_1}{N}-e_2}(f)\left(\overline{X}_N^0(s)\right)^2 G_N^0(Ns)F_N^0(Ns)\,\mathrm{d}s \\ &+ \beta_m N\int_0^t \nabla_{\frac{e_1}{N}+e_2}(f)\left(\overline{X}_N^0(s)\right)^2 \left(C_m^N-G_N^0(Ns)\right)\mathrm{d}s \\ &+ \beta_6 N\int_0^t \nabla_{e_3}(f)\left(\overline{X}_N^0(s)\right)^2 \mathrm{d}s \\ &+ \delta_6 N\int_0^t \nabla_{-e_3}(f)\left(\overline{X}_N^0(s)\right)^2 Z_N^0(Ns)\,\mathrm{d}s.\end{aligned} \tag{2.41}$$

Chapter 3

Pairing Mechanisms

3.1 Introduction

In this paper we investigate a general mechanism of interaction between different populations of particles and specific particles, agents, in some environment. Assuming that each of the particles follows a random path in the medium, when a particle and an agent meet, they may form a pair which has a specific functional property in the medium. Such a pair is also subject to random events, it splits after some random amount of time. The efficiency of the pairing mechanism is analyzed with the time evolution of the number of paired particles of each type.

3.1.1 Motivation

The initial motivation comes from molecular biology where this is an almost ubiquitous phenomenon occurring in biological cells. It can be (roughly) described as follows: different types of macro-molecules (ribosomes, or polymerases for example), referred to as *particles*, are in charge of producing some of the functional components necessary to the development of the cell (mRNAs, proteins). Specific macro-molecules, referred to as *agents* in the paper, like small RNAs, have a regulation role in the cell. Agents can pair/bind with particles to block, or to speed-up, their activity. Due to thermal noise, a pair agent-particle splits after some time. The dynamic behavior of the systems investigated are described in terms of binding/unbinding operations of agents and particles. See Section 3.5 for a more detailed presentation of these aspects.

3.1.2 Literature

A typical representation of pairing mechanisms in the literature, written as a chemical reaction, is of the type,

$$\mathcal{Z}+\mathcal{F}_j \rightleftharpoons \mathcal{Z}\mathcal{F}_j \rightharpoonup \mathcal{G}_j+\mathcal{Z} \tag{3.1}$$

where the chemical species are as follows: $\mathcal{Z}$ is associated to what we call agents (enzymes, small RNAs, ...), $\mathcal{F}_j$ is for particles of type $j \in \{1, \ldots, J\}$ (RNAs, polymerases, ...). The species $\mathcal{Z}\mathcal{F}_j$ is for pairs of $\mathcal{Z}$ and $\mathcal{F}_j$ and $\mathcal{G}_j$ is for a "product" of type j, it can be $\mathcal{F}_j$. In a deterministic setting the leads to a set of ODEs for a dynamical system $(X_A(t), (A \in Z), F_j, ZF_j, G_j)$, for example, for $(X_{ZF_j}(t))$ it gives

$$\frac{\mathrm{d}}{\mathrm{d}t} X_{ZF_j}(t) = \kappa_j^+ X_Z(t) X_{F_j}(t) - \kappa_j^- X_{XF_j}(t) \tag{3.2}$$

for some constants $\kappa_j^{\pm} \geq 0$. Note the quadratic term on the right hand side. Investigations are generally on the stability of these dynamical systems. See Petrides and Vinnicombe [58], Del Giudice et al. [20] and Jayaprakash and Das [40]. See Section 3.2 for a brief presentation of this formalism.

In a stochastic context, this is represented as a Markov process whose state descriptor is the vector of the number of copies of the different chemical species. Simulations and numerical analysis of the associated Fokker-Planck equations have been used to study these phenomena, see Petrides and Vinnicombe [58].

The technical context is related to the celebrated Michaelis-Menten kinetics. These chemical reactions involve enzyme, substrate and product macromolecules, whose associated chemical species are denoted respectively as $\mathcal{E}$, $\mathcal{S}$ and $\mathcal{P}$. The chemical reaction

$$\mathcal{E} + \mathcal{S} \rightleftharpoons \mathcal{E}\mathcal{S} \rightharpoonup \mathcal{P} + \mathcal{E},$$

has been investigated for some time now. The basic assumption for these models is that there are few copies of chemical species $\mathcal{E}$ but a large number of copies of substrate, so that the reaction rate is large (for the chemical reaction on the left). In a deterministic setting it leads to a system of non-polynomial ODEs. In a stochastic context, these ODEs can be obtained via the proof of an averaging principle. See Michaelis and Menten [48] and for a general overview Sanft et al. [65] and Cornish-Bowden [17].

Averaging principles also play an important role in our paper. For the mathematical point of view, agents may be seen as playing the role of enzymes in our model. Nevertheless our framework is not really that of Michaelis-Menten. Their number is nevertheless not fixed in our main model of Section 3.4. As we will see, the production of agents has a strong impact on the qualitative behavior of the system. As it can be expected, the quadratic expressions due to pairing mechanisms, like in Relation (3.2), are at the origin of some technical difficulties in the proof of limit theorems.

3.1.3 Stochastic Model

There are J types of particles. For $1 \leq j \leq J$, N_j is the total number of particles of type j, this quantity is assumed to be fixed. The total number of particles is $N = N_1 + \cdots N_J$, it is our scaling parameter. A Markovian stochastic model is considered, each event occurs after an amount of time with an exponential

distribution and the corresponding random variables are assumed to be independent.

There is only type of agent. An agent and a particle of type $j \in \{1, \ldots, J\}$ bind/pair at rate λ_j, and, in a reverse operation, such a pair split into an agent and a particle of type j at rate η_j. An agent or a particle which is not paired is said to be *free*.

The variables of interest are

$$(F_N(t), Z_N(t)) \stackrel{\text{def.}}{=} ((F_{N,j}(t), j=1, \ldots, J), Z_N(t))$$

where, for $1 \leq j \leq J$ and $t \geq 0$, $F_{N,j}(t)$ is the number of free particles of type j, i.e. not paired with an agent, and $(Z_N(t))$ is the process for the number of free agents. When the goal of pairing mechanism is of reducing the activity of the particles, this will be referred to as *sequestration* of particles, the objective is of minimizing

$$\left(\sum_{j=1}^{J} \frac{F_{N,j}(t)}{N} \right)$$

the process of the fraction of the number of free particles. We analyze the asymptotic behavior, when N goes to infinity, of the time evolution of the J-dimensional process $(F_{j,N}(t)/N)$ associated to the free particles. An appropriate timescale for a non-trivial asymptotic evolution when N goes to infinity has to be determined.

In Fromion et al. [31], a related model of sequestration has been analyzed, to study the regulation of transcription. It also includes additional variables which are not considered in this paper. A component of the stochastic model is a related Markov process but in dimension 1, i.e. for $J=1$. As it will be seen, compared to the case $J=1$, the multi-dimensional aspect of our model has a significant impact on the scaling properties of the associated stochastic processes.

Two types of models for agents are analyzed.

a. Agents are neither created nor removed: the number of agents is fixed, of the order of N.

b. An agent is created at rate β and, only when it is not paired with a particle, it dies at rate δ.

Case a) is used to investigate the case when the environment does not change significantly and when there is already a large number of agents to regulate the system. Two cases are considered. In Section 3.3.1 the total number of agents is of the order of rN with $r \in (0, 1)$, there are much more particles than agents. It is shown that the process $(F_{j,N}(t)/N)$ is converging in distribution to the solution of an ODE. The equilibrium point of this ODE is unique and its coordinates are positive. For this system the number of free particles of type $j \in \{1, \ldots, J\}$ is, of course, of the order of N.

In Section 3.3.2 the total number of agents is N the same as the total number of particles. It is shown that with, appropriate initial conditions, the process $(F_{j,N}(t/\sqrt{N})/\sqrt{N})$ is converging in distribution to the solution of an ODE and a central limit theorem is proved, it shows that the fluctuations are of the order of $\sqrt[4]{N}$. In this case the impact of stochasticity on the pairing mechanism is minimal since there is a fraction of the order of $1/\sqrt{N}$ of free particles.

Case b) is investigated in Section 3.4 for the case when, initially, there few agents (free or paired) are in the system, the goal is of investigate the growth of the number of paired particles. The proof of an averaging principle in this context is challenging for several reasons.

Since a paired agent does not die (it is not degraded), one can expect an asymptotic situation as in Section 3.3.2 with a negligible fraction of free particles. We show that this is not the case, in fact, formally, the behavior is similar to that of Section 3.3.1, but on a faster time scale and with important qualitative and technical differences.

If the system starts with few agents, in this case most of N particles are initially "free", all agents created will pair with a free particle right away and will keep doing that, via the successive steps of pairing/splitting, as long the number of free particles is "large" so that, with high probability, pairing occurs before degradation for agents. Given the rate of creation of agents, the natural timescale to study this problem is (Nt).

It can be expected that the multi-dimensional process

$$\left((\frac{F_N(Nt)}{N}\right) = \left(\frac{F_{j,N}(Nt)}{N}\right)$$

converges in distribution to a continuous process reflecting the asymptotic degree of pairing of the system. Due to their large transition rates, the integer-valued processes $(Z_N(Nt))$ and $(F_N(Nt))$ are "fast" processes. Because of the space scaling, $(F_N(Nt)/N)$ is an a priori "slow" process. Following the classical approach in this domain, see Papanicolaou et al. [56] in a stochastic calculus context and Kurtz [45] for its formulation for jump process. For $T>0$, one has to consider the occupation measure associated to $(Z_N(Nt))$, i.e. this is the functional on non-negative Borelian functions on $[0,T]\times\mathbb{N}$,

$$g \longrightarrow \int_0^T g(s, Z_N(Ns))\,\mathrm{d}s.$$

If this approach allows us to derive the results of Section 3.3.1 for case a), where an averaging principle is proved, it does not work for case b). The sequence of processes $(F_{j,N}(Nt)/N, j{=}1,\ldots,J)$ does *not* converge in distribution in fact. It is not tight for the topology associated to uniform convergence if the initial state does not converge to some one-dimensional curve of $[0,1]^J$. The main convergence result of this case is Theorem 35 of Section 3.4. It shows that the process associated to the total number of free particles,

$$(\|F_N(Nt)\|) \stackrel{\text{def.}}{=} \left(\sum_{j=1}^{J} \frac{F_{N,j}(Nt)}{N}\right),$$

converges in distribution to a continuous process. The sequence of $[0,1]^J$-valued processes $(F_{j,N}(Nt)/N, 1 \leq j \leq J)$ converges in distribution, but in a weak form, via its associated occupation measure. It turns out that the process $(\|F_N(Nt)\|)$ determines, in some way, the behavior of the coordinates of $(F_N(Nt))$. For N large $(F_{N,j}(Nt))$ can in fact be represented as a curve of $[0,1]^J$ determined by $\|F_N(Nt)\|$. In Fromion et al. [31], no such difficulty shows up since $J=1$.

Intuitively, it is shown that, in the limit, the number of free particles is of the order of N, as in case a) but for some specific $r<1$. These results stress the impact of dynamical arrivals and departures of agents. In particular the fraction of paired particles is asymptotically strictly less than 1.

Technical difficulties are related to the lack of tightness properties of the process $(F_{j,N}(Nt)/N)$. For this reason the definition of the occupation measure is extended to include also "slow" processes and not only the fast processes as it is classical in the context of averaging principles. As a functional on Borelian functions on $[0,T]\times[0,1]^J\times\mathbb{N}$, the occupation measure is expressed as

$$g \longrightarrow \int_0^T g\left(s, \frac{F_N(Ns)}{N}, Z_N(Ns)\right) \mathrm{d}s.$$

The investigation of the limiting behavior of this sequence of occupation measures is the main topic of Section 3.4, including the identification of possible limiting points.

The reason of this behavior is essentially due to the interaction of several fast time scales. At the normal time scale (t), if the components of the vector $F_N(t)$ are already of the order of N, the pairing/splitting events occur at a rate proportional to N. Since the natural time scaling for case b) is sped-up as (Nt), roughly speaking, the pairing/splitting events will be instantaneously at equilibrium, at the first order, at any "time" t for the current "mass" $\|F_N(Nt)\|$. In particular, if the initial point $\overline{F}_N(0)$ does not converge to the equilibrium associated to the mass $\|F_N(0)\|$, there cannot be a convergence in a neighborhood of $t=0$, this is the one-dimensional curve mentioned above.

Outline of the Paper

Section 3.2 introduces notations and the Markovian process used to investigate pairing mechanisms. Section 3.3 analyzes the static case when the number of agents is fixed and in Section 3.4 a stochastic averaging principle is proved when agents are created and degraded. To motivate the design of such stochastic models, Section 3.5 presents several examples of regulation mechanisms in biological cells. Section 3.6 is a quick reminder of classical limit results for $M/M/1$ and $M/M/\infty$ queues. These queues play an important role in the design of couplings used in the proofs of our limit theorems.

3.2 Stochastic Model

Definitions and Notations

If H is a locally compact metric space, $\mathcal{C}_c(H)$ is the space of continuous functions with compact support endowed with the topology of uniform convergence. We denote by $\mathcal{M}^+(H)$ the set of non-negative Radon measures on H and $\mathcal{M}_1(H)$, the set of probability distributions on H, both spaces are endowed with the weak topology. See Rudin [64]. Throughout the paper convergence in distribution of a sequence of jump processes $(U_N(t))$ to a process $(U(t))$ is understood with respect to the topology of uniform convergence on compact sets for càdlàg functions. See Chapter 2 of Billingsley [10] for example.

For $J{\in}\mathbb{N}$, if $x{=}(x_i)$, $y{=}(y_i){\in}\mathbb{R}^J$, define

$$\|x\| \stackrel{\text{def.}}{=} |x_1|{+}\cdots{+}|x_J|, \quad \langle x,y\rangle = x_1y_1{+}\cdots{+}x_Jy_J, \tag{3.3}$$

and,

$$\overline{x}{=}\max(x_j, 1{\leq}j{\leq}J) \text{ and } \underline{x}{=}\inf(x_j, 1{\leq}j{\leq}J). \tag{3.4}$$

We now introduce the main definitions for our stochastic model. There are J different types of particles. The total number of particles of type $j{\in}\{1,\ldots,J\}$ is $C_{j,N}$, and $N{=}C_{1,N}{+}\cdots{+}C_{J,N}$, the total number of particles is a fixed number, it is also our scaling parameter. It is assumed that,

$$\lim_{N\to+\infty}\left(\frac{C_{j,N}}{N}\right) = c{=}(c_j) \tag{3.5}$$

holds, for some $c{\in}(0,1)^J$ such that $c_1{+}c_2{+}\cdots{+}c_J{=}1$.

State Space

The state space is

$$\mathcal{S}_N{=}\left\{x=(f,z)=((f_j),z){\in}\mathbb{N}^{J+1} : f_j{\leq}C_{j,N}, \forall 1{\leq}j{\leq}J\right\},$$

and, if $t{\geq}0$,

— for $1{\leq}j{\leq}J$, $F_{j,N}(t)$ denotes the number of free particles of type j at time t and $F_N(t){=}(F_{j,N}(t), 1{\leq}j{\leq}J)$;

— The number of free agents at time t is $Z_N(t)$;

— The state of the process at time t is $X_N(t){=}(F_N(t), Z_N(t)){\in}\mathcal{S}_N$.

The number of agents paired with a particle of type j at time t is therefore $S_{j,N}(t){=}C_{j,N}{-}F_{j,N}(t)$. In state $(F_N(t), Z_N(t))$, the total number of free particles is $\|F_N(t)\|$.

Transitions

The dynamical behavior of $(X_N(t))$ is driven by several types of transitions.

a. A given particle of type j and a given agent are paired at rate λ_j;

b. A pair (particle of type j, agent) is split at rate $\eta_j{>}0$ to give a particle of type j and a free agent;

c. Agents are created at rate $\beta{\geq}0$ and a *free* agent dies, is degraded, at rate $\delta{>}0$. An agent paired to a particle cannot die.

The state process $(X_N(t)){=}(F_N(t), Z_N(t))$ is almost surely a *càdlàg function*, i.e. a right-continuous function with left limits at any point of $(0, +\infty)$. It is described as an irreducible Markov process on $\mathcal{S}_N$ whose Q-matrix Q_F is given by

$$(f,z){=}((f_j),z) \longrightarrow (f,z){+}\begin{cases}(-e_j,-1) & \lambda_j f_j z,\\ (+e_j,+1) & \eta_j(C_{j,N}{-}f_j),\\ (0,+1) & \beta,\\ (0,-1) & \delta z,\end{cases}$$

where e_j is the jth unit vector of $\mathbb{N}^J$.

Note that the pairing mechanism induces quadratic transition rates in the Q-matrix.

Definition 23.

$$c{=}(c_j),\ \eta{=}(\eta_j),\ \lambda{=}(\lambda_j), \quad \rho_0{=}\frac{\beta}{\delta}\ \textit{and}\ \rho_j{=}\frac{\eta_j}{\lambda_j}, j{=}1,\ldots,J.$$

For $y{\in}(0,1)$, $\phi(y)$ is defined as the unique solution of the equation

$$\sum_{j=1}^{J} \frac{\rho_j}{\rho_j{+}\phi(y)} c_j = y. \tag{3.6}$$

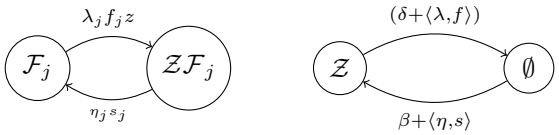


Figure 3.1: Transitions of Pairing Mechanism, with $s{=}(s_j)\stackrel{\text{def.}}{=}(C_{j,N}{-}f_j)$.

Stochastic Differential Equations

The process $(X_N(t))=(F_N(t), Z_N(t))$ is represented as the solution of a system of SDEs (Stochastic Differential Equations). On the probability space there are $2(J+1)$ independent Poisson processes on $\mathbb{R}_+^2$ with intensity measure $\mathrm{d}x\otimes\mathrm{d}t$, $\{\mathcal{P}_z^+, \mathcal{P}_z^-, \mathcal{P}_j^+, \mathcal{P}_j^-,\ j=1,\ldots,J\}$. See Rogers and Williams [62] for example. The underlying filtration $(\mathcal{F}_t)$ is defined by, for $t>0$,

$$\mathcal{F}_t = \sigma\left\langle \mathcal{P}_{j/z}^{\pm}([a,b]\times[0,s], j=1,\ldots,J\}, a\le b, s\le t\right\rangle.$$

In the following, measurability properties are assumed to be with respect to this filtration.

Let $(F_N(t), Z_N(t))$ be the solution of the SDE, for $j=1,\ldots,J$,

$$\mathrm{d}F_{j,N}(t) = \mathcal{P}_j^+((0,\eta_j(C_{j,N}-F_{j,N}(t-))),\mathrm{d}t) - \mathcal{P}_j^-((0,\lambda_j F_{j,N}(t-)Z_N(t-)),\mathrm{d}t), \tag{3.7}$$

$$\mathrm{d}Z_N(t) = \mathcal{P}_z^+((0,\beta),\mathrm{d}t) - \mathcal{P}_z^-((0,\delta Z_N(t-)),\mathrm{d}t) + \sum_{j=1}^{J}\mathrm{d}F_{j,N}(t), \tag{3.8}$$

where $U(t-)$ denotes the left-limit of the càdlàg process $(U(s))$ at $t>0$ and with the usual notation, if $A\ge 0$ and $\mathcal{P}$ is a Poisson point process on $\mathbb{R}_+^2$,

$$\mathcal{P}((0,A),\mathrm{d}t) = \int \mathbb{1}_{\{x\le A\}}\mathcal{P}(\mathrm{d}x,\mathrm{d}t). \tag{3.9}$$

By integrating these relations, we obtain, for $j=1,\ldots,J$,

$$F_{j,N}(t) = F_{j,N}(0) + M_{j,N}(t) + \eta_j\int_0^t (C_{j,N}-F_{j,N}(s))\,\mathrm{d}s - \lambda_j\int_0^t F_{j,N}(s)Z_N(s)\,\mathrm{d}s, \tag{3.10}$$

and

$$Z_N(t) = Z_N(0) + M_{z,N}(t) + \beta t - \delta\int_0^t Z_N(s)\,\mathrm{d}s + \sum_{j=1}^{J}\eta_j\int_0^t (C_{j,N}-F_{j,N}(s)\,\mathrm{d}s - \lambda_j\int_0^t F_{j,N}(s)Z_N(s)\,\mathrm{d}s, \tag{3.11}$$

where $(M_{j,N}(t))$ and $(M_{z,N}(t))$ are square integrable martingales whose previsible increasing processes are given by

$$(\langle M_{j,N}\rangle(t)) = \left(\eta_j\int_0^t (C_{j,N}-F_{j,N}(s))\,\mathrm{d}s + \lambda_j\int_0^t F_{j,N}(s)Z_N(s)\,\mathrm{d}s\right), \tag{3.12}$$

$$(\langle M_{z,N}\rangle(t)) = \left(\beta t + \delta\int_0^t Z_N(s)\,\mathrm{d}s + \sum_{j=1}^{J}\langle M_{j,N}\rangle(t)\right). \tag{3.13}$$

Invariant Distribution

As explained in the introduction, our model can be expressed in the framework of chemical reaction networks (CRN). Since we are mainly interested in the transient behavior of our system, we just give a quick sketch. It is only mentioned as an interesting aspect of our system. See Feinberg [25] for a general introduction on CRNs.

The corresponding chemical reactions are represented as

$$\emptyset \underset{\delta}{\overset{\beta}{\rightleftharpoons}} \mathcal{Z}, \qquad \mathcal{Z} + \mathcal{F}_j \underset{\eta_j}{\overset{\lambda_j}{\rightleftharpoons}} \mathcal{F}\mathcal{Z}_{|}, \quad 1{\leq}j{\leq}J. \tag{3.14}$$

The associated dynamical system $((f_j(t), g_j(t)), z(t))$ is defined by the ODEs

$$\begin{cases} \dot{f}_j(t) = \lambda_j f_j(t) z(t) - \eta_j g_j(t), \quad 1{\leq}j{\leq}J, \\ \dot{f}_j(t) + \dot{g}_j(t) = 0, \quad 1{\leq}j{\leq}J, \\ \dot{z}(t) = \beta - \delta z(t). \end{cases}$$

Its fixed point is given by $((u_j, v_j), w)$ with

$$w = \rho_0, \quad u_j = \frac{C_j^N}{1+\rho_0\rho_j} = C_j^N - v_j, \quad 1{\leq}j{\leq}J, \tag{3.15}$$

with $\rho_0{=}\beta/\delta$ and $\rho_j{=}\lambda_j/\eta_j$, for $1{\leq}j{\leq}J$.

The characteristics of this CRN are:

— $m{=}2J{+}2$ chemical species: $\mathcal{Z}$, $\mathcal{F}_j$, $\mathcal{F}\mathcal{Z}_{|}$, $j{=}1, \ldots, J$;

— $\ell{=}J{+}1$ cycles, these are the *single linkage classes* of the CRN;

— The range, the dimension of the stochiometric space, is $s{=}J{+}1$.

This is a CRN with deficiency $\delta{=}m{-}\ell{-}s{=}0$. A standard result, see Anderson et al. [4], gives an explicit expression of the invariant distribution π_N of $(X_N(t))$ on $\mathcal{S}_N$.

Proposition 24. *The invariant distribution of $(X_N(t))$ on $\mathcal{S}_N$ is given by,*

$$\pi(f, s) = \frac{1}{Z_N} \frac{w^z}{z!} \prod_{j=1}^{J} \frac{u_j^{f_j}}{f_j!} \frac{v_j^{C_j^N - f_j}}{(C_j^N - f_j)!}, \quad (f, z){=}((f_j), z){\in}\mathcal{S}_N,$$

where Z_N is the normalization constant and (u_j) and (v_j) are defined by Relation (3.15).

3.3 Fixed Number of Agents

Throughout this section the number of agents C_Z^N is fixed, of the order of rN, with $r<1$ in Section 3.3.1, and is exactly N, the total number of particles, in Section 3.3.2. There are no creations or degradation of agents. Only pairing and splitting mechanisms operate in these cases. As explained in the introduction, the purpose is of understanding the behavior of the system when the total number of agents does not change. Section 3.4 investigate a much more dynamic version of the system.

The state space of the system is $\mathcal{S}_N \stackrel{\text{def.}}{=} \prod_{j=1}^{J}\{0,\dots,C_j^N\}$. For a state $x=(x_j)\in\mathcal{S}_N$, the total number of paired particles with an agent is $N-x_1-\cdots-x_J$. The associated process in $\mathcal{S}_N$ is denoted as $(F_N^r(t))$ has the Markov property, its Q-matrix is given by, for $x\in\mathcal{S}_N$,

$$x \to \begin{cases} x+e_j & \eta_j\left(C_j^N-x_j\right), \\ x-e_j & \lambda_j x_j\left(C_Z^N-(N-x_1-\cdots-x_J)\right), \end{cases}$$

where e_j is the jth unit vector of $\mathbb{N}^J$.

3.3.1 Overloaded Case

In this section the total number of agents if of the order of rN, with $r<1$,

$$\lim_{N\to+\infty}\frac{C_Z^N}{N}=r. \tag{3.16}$$

Since there are not enough agents to handle all particles, it is clear that the number of free particles of type j, $1\le j\le J$, should be of the order of N. The SDE (3.10) for $(F_N^r(t))$ becomes, for $1\le j\le J$,

$$\begin{aligned}\mathrm{d}F_{N,j}^r(t) = {}& \mathcal{P}_{S_j}((0,\eta_j(C_j^N-F_{N,j}(t-))),\mathrm{d}t) \\ & -\mathcal{P}_{F_j}((0,\lambda_j F_{N,j}^r(t-)Z_N^r(t-)),\mathrm{d}t),\end{aligned} \tag{3.17}$$

where $(Z_N^r(t))$ is the process of free agents,

$$(Z_N^r(t)) = \left(C_Z^N-\sum_{j=1}^{J}\left(C_j^N-F_{N,j}^r(t)\right)\right) = \left(C_Z^N+\|F_N^r(t)\|-N\right). \tag{3.18}$$

We assume that the initial conditions are such that $Z_N(0)=z_0$, for some fixed $z_0\in\mathbb{N}$, and

$$\lim_{N\to+\infty}\frac{F_N^r(0)}{N}=\overline{f}_0=(\overline{f}_{0,j})\in[0,1]^J, \tag{3.19}$$

such that

$$\sum_{j=1}^{J}\overline{f}_{0,j}=1-r. \tag{3.20}$$

The last condition expresses simply that most of the agents are initially paired with particles. We will see that the number of free agents remains a finite random variable.

Definition 25. *For $N>0$, the scaled process is defined as*

$$\left(\overline{F}^r_N(t)\right) \stackrel{\text{def.}}{=} \left(\frac{F^r_{N,j}(t)}{N}, j=1,\dots,J\right). \tag{3.21}$$

If g is non-negative Borelian function on $\mathbb{R}_+\times\mathbb{N}$, we define the occupation measure

$$\langle \Lambda^r_N, g\rangle \stackrel{\text{def.}}{=} \int_{\mathbb{R}_+} g\,(s, Z^r_N(s))\,\mathrm{d}s. \tag{3.22}$$

Since $F^r_{N,j}(t)\le C^N_j\le N$, $1\le j\le J$, for $t\ge 0$, the state space of the process $(\overline{F}^r_N(t))$ is included in $[0,1]^J$.

Lemma 26. *If N is sufficiently large, there exists a coupling of the process $(Z^r_N(t))$ with $(L(Nt))$, where $(L(t))$ is an $M/M/\infty$ queue with input rate $\overline{\eta}$ and service rate $\underline{\lambda}r/2$, with*

$$\overline{\eta}=\max_j \eta_j, \ \textit{and}\ \underline{\lambda}=\min_j \lambda_j,$$

such that $L(0)=z_0$ and $Z^r_N(t)\le L(Nt)$ holds for all $t\ge 0$.

See Section 3.6.2 on the $M/M/\infty$ queue.

Proof. This is a simple consequence of the fact that if $Z^r_N(t)=z\in\mathbb{N}$, the rate at which there is a jump of size $+1$, resp. -1, is

$$\sum_{j=1}^N \eta_j\left(C^N_j - F^r_{N,j}(t)\right) \le \overline{\eta}\sum_{j=1}^N C^N_j = \overline{\eta}N,$$

resp.

$$\sum_{j=1}^N \lambda_j F^r_{N,j}(t) \ge \underline{\lambda}(N-C^N_Z) \ge \underline{\lambda}\,(1-\varepsilon)\,N,$$

for some $\varepsilon\in(0,1)$ if N is large enough. It is then straightforward to construct the desired coupling. □

The integration of the SDE (3.17) gives the relation

$$\begin{aligned}\overline{F}^r_{j,N}(t) = \overline{F}^r_{j,N}(0) + M^r_{j,N}(t)\\ - \lambda_j\int_0^t \overline{F}^r_{N,j}(s)Z^r_N(s)\,\mathrm{d}s + \eta_j\int_0^t\left(\frac{C^N_j}{N}-\overline{F}^r_{N,j}(s)\right)\mathrm{d}s,\end{aligned} \tag{3.23}$$

The process $(M_{j,N}^r(t))$ is a martingale whose previsible increasing process is

$$\left(\left\langle M_{j,N}^r\right\rangle(t)\right) = \left(\frac{\lambda_j}{N}\int_0^t \overline{F}_{j,N}^r(s)Z_N^r(s)\,\mathrm{d}s + \frac{\eta_j}{N}\int_0^t\left(\frac{C_j^N}{N} - \overline{F}_{N,j}^r(s)\right)\mathrm{d}s\right). \tag{3.24}$$

Proposition 27. *Under the assumptions* (3.5), (3.19), *and* (3.20)*for the initial state, then for the convergence in distribution the relation*

$$\lim_{N\to+\infty}\left(\frac{F_N^r(t)}{N}\right) = (f^r(t)) = (f_j^r(t)),$$

holds, where $(x(t))$ *is the solution of the ODEs, for* $1{\leq}j{\leq}J$ *and* $t{>}0$*,*

$$\frac{\mathrm{d}}{\mathrm{d}t}f_j^r(t) = \lambda_j f_j^r(t)\frac{\langle\eta, c{-}f^r(t)\rangle}{\langle\lambda, f^r(t)\rangle} - \eta_j\left(c_j{-}f_j^r(t)\right), \tag{3.25}$$

with Definition 23.

Proof. By using the notations of Lemma 26 and the ergodic theorem for positive recurrent Markov processes, it is not difficult to prove that the sequence of processes

$$\left(\int_0^t L(Ns)\,\mathrm{d}s\right)$$

is tight with the criterion of the modulus of continuity, see Theorem 7.3 of Billingsley [10], and that its limiting point is necessary $(\overline{\eta}/\underline{\lambda}\cdot t)$.

Since $F_{N,j}^r(t){\leq}C_j^N{\leq}N$, for $1{\leq}j{\leq}J$ and $t{\geq}0$, with Relation (3.24), we obtain therefore that the process $(\langle M_{j,N}^r\rangle(t))$ is converging in distribution to 0, Doobs' Inequality gives that the same result holds for the martingale $(M_{j,N}^r(t))$.

For $T{>}0$, with Relation (3.23), the modulus of continuity of $(\overline{F}_{j,N}^r(t))$ on the time interval $[0,T]$ is

$$\begin{aligned}\omega_{F_N,T}(\delta) &\stackrel{\text{def.}}{=} \sup_{\substack{s,t\leq T\\ |s-t|\leq\delta}}\left|\overline{F}_{j,N}^r(t){-}\overline{F}_{j,N}^r(s)\right| \\ &\leq \sup_{s\leq T}|M_{j,N}^r(s)| + \lambda_j\sup_{\substack{s\leq t\leq T\\ |s-t|\leq\delta}}\int_s^t L(Nu)\,\mathrm{d}u + \delta\eta_j\frac{C_j^N}{N}.\end{aligned}$$

Again with Theorem 7.3 of Billingsley [10], we deduce that the sequence of processes $(\overline{F}_{N,j}^r(t))$ is tight.

For $K{>}0$,

$$\mathbb{E}(\Lambda_N^r([0,T]{\times}[K,+\infty))) \leq \int_0^T \mathbb{P}\left(L(Ns){\geq}K\right)\mathrm{d}s = \frac{1}{N}\int_0^{NT}\mathbb{P}(L(s){\geq}K)\,\mathrm{d}s,$$

since $(L(s))$ converges in distribution to a Poisson distribution, see Section 3.6.2, the last term can be made arbitrarily small for K sufficiently large. Lemma 1.3

of Kurtz [45] gives the tightness of the sequence of random measures (Λ_N^r) and any of its limiting points Λ_∞^r can be represented as,

$$\langle \Lambda_\infty^r, f\rangle = \int_{\mathbb{R}_+\times[0,1]^J\times\mathbb{N}} f\,(s,x)\,\mu_s(\mathrm{d}x)\,\mathrm{d}s,$$

if f is a non-negative Borelian function on $[0,T]\times\mathbb{N}$, where (μ_s) is an optional process on $\mathcal{M}_1(\mathbb{N})$.

Hence the sequence of random variables $((\overline{F}_N^r(t)), \Lambda_\infty^r)$ is tight, we denote by $((f^r(t)), \Lambda_\infty^r)$ one of its limiting points. Since Proposition 41 establishes a similar result, but in a more difficult technical framework. The analogue of the sequence of processes $(\overline{F}_N^r(t))$ is *not* tight in this case. For this reason, we skip the proof of the fact that, in the above representation of Λ_∞^r, μ_s can be expressed as a Poisson distribution with parameter

$$\frac{\langle \eta, c-f^r(s)\rangle}{\langle \lambda, f^r(s)\rangle},$$

and that $(f^r(t))$ satisfies the ODE (3.25). The proposition is proved. □

Corollary 28. *With the notations of Proposition 27, the equilibrium point f_∞^r of $(f^r(t))$ is given by*

$$f_\infty^r = \left(\frac{\eta_j c_j}{\eta_j+\lambda_j h_\infty}\right),$$

where $h_\infty=\phi(1-r)$ and ϕ is defined by Relation (3.6).

3.3.2 Critical Case

The number of agents is exactly N, the total number of particles and there are still no creation or degradation of agents. We prove that the process of the number of free particles of type $j\in\{1,\ldots\}$, $(F_{N,j}^1(t))$, is of the order of $\sqrt{N}$, with fluctuations of the order of $\sqrt[4]{N}$. See Theorems 31 and 32.

Note that, for $t\geq 0$, the total number of free agents at time t is

$$N-\sum_{j=1}^J \left(C_j^N - F_{N,j}^1(t)\right) = \sum_{j=1}^J F_{N,j}^1(t) = \|F_N^1(t)\|.$$

The Q-matrix Q_f of $(F_{N,j}^1(t))$ is thus given by, for $x\in\mathcal{S}_N$,

$$x \to \begin{cases} x+e_j & \eta_j\left(C_j^N - x_j\right), \\ x-e_j & \lambda_j x_j \|x\|. \end{cases}$$

Lemma 29. *If, for η, $\lambda>0$ and $N\geq 1$, if $(X_N(t))$ is the solution of the SDE,*

$$\mathrm{d}X_N(t) = \mathcal{P}_{S_1}((0,\eta N),\mathrm{d}t) - \mathcal{P}_{F_1}\left(\left(0,\lambda X_N(t-)^2\right),\mathrm{d}t\right),$$

with $X(0)=0$, then for any $T\geq 0$, there exists $K_1>0$ such that,

$$\lim_{N\to+\infty}\mathbb{P}\left(\sup_{t\leq NT}\frac{X_N(t)}{\sqrt{N}}\geq K_1\right)=0,$$

and

$$\sup_{t\geq 0}\mathbb{E}\left(X_N(t)^2\right)<+\infty.$$

Proof. We fix K_0 such that $\lambda K_0^2>\eta$. If we define the process $(Y(t))$ by the SDE

$$\mathrm{d}Y(t)=\mathcal{P}_{S_1}((0,\eta),\mathrm{d}t)-\mathbb{1}_{\{Y(t-)>0\}}\mathcal{P}_{F_1}((0,\lambda K_0^2),\mathrm{d}t),$$

with $Y(0)=0$. As in the proof of Proposition 30, by induction on the successive jumps of $(X_N(t))$, it is easy to show that the relation

$$X_N(t)\leq K_0\sqrt{N}+Y(Nt)$$

holds almost surely for all $t>0$. The process $(Y(t))$ is a reflected random walk on $\mathbb{N}$, it is usually associated to the $M/M/1$ queue. See Chapter 5 of Robert [61]. Proposition 5.11 of this reference gives that if T_A is the hitting time of A by $(Y(t))$ then the random variable

$$\left(\frac{\eta}{\lambda K_0^2}\right)^A T_A$$

is converging in distribution to an exponential random variable when A goes to infinity. This shows in particular that (T_A/A^2) is converging in distribution to infinity, hence, for any $K>0$,

$$\lim_{N\to+\infty}\mathbb{P}\left(\sup_{t\leq T}\frac{Y(tN)}{\sqrt{N}}\geq 1\right)=\lim_{N\to+\infty}\mathbb{P}\left(T_{\lceil\sqrt{N}\rceil}\leq TN\right)=0.$$

This gives the first part of the lemma. We conclude the proof by setting $K_1=K_0+1$ and remarking that the invariant distribution of $(Y(t))$ is a geometric distribution with parameter $\eta/(\lambda K_0^2)$ and that, with a simple coupling argument, the mapping $t\to\mathbb{E}(Y(t)^2)$ is non-decreasing. □

The next result shows that all coordinates of $(F_N^1(t))$ are at most of the order of $\sqrt{N}$ very quickly independently of the initial point. Theorem 31 completes this result by showing that the order of magnitude of its coordinates is exactly $\sqrt{N}$.

Proposition 30 (Coupling)**.** *For all $j\in\{1,\ldots,J\}$,*

$$F_{N,j}^1(t)\leq F_{N,j}^1(0)+X_{N,j}(t),\quad t\geq 0,\tag{3.26}$$

where the $(X_{N,j}(t))$ are the solutions of the SDEs

$$\mathrm{d}X_{N,j}(t)=\mathcal{P}_{S_j}((0,\overline{\eta}N),\mathrm{d}t)-\mathcal{P}_{F_j}\left(\left(0,\underline{\lambda}X_{N,j}(t-)^2\right),\mathrm{d}t\right),\quad 1\leq j\leq J,\tag{3.27}$$

with $(X_j(0))=0$ *and* $\overline{\eta}$ *and* $\underline{\lambda}$ *are defined by Relation* (3.4).

There exists $\ell_0>0$ *such that if*

$$\tau_N \stackrel{\text{def.}}{=} \inf\left\{t>0 : F^1_{N,j}(t)\leq \left\lceil \ell_0\sqrt{N}\right\rceil, \forall j\in\{1,\ldots,J\}\right\},$$

then

$$\sup_{N\geq 1}\sup_{x\in\mathcal{S}_N}\mathbb{E}_x(\tau_N)<+\infty.$$

Proof. The Q-matrix Q_X of the Markov process $(X_{N,j}(t))$ defined by the SDEs (3.27) is

$$x\to\begin{cases} x+e_j & \overline{\eta}N,\\ x-e_j & \underline{\lambda}x_j^2,\end{cases}$$

clearly $q_f(x,x+e_j)\leq q_X(x,x+e_j)$ and $q_f(x,x-e_j)\geq q_X(x,x-e_j)$.

A simple coupling, by induction on the successive jumps of $(F^1_{N,j}(t))$, gives that the relation

$$F^1_{N,j}(t)\leq F^1_{N,j}(0)+X_{N,j}(t)$$

holds for all $t\geq 0$.

To prove the last assertion, in view of Relation (3.26), it is enough to prove it for the "maximal" initial state, i.e. $(F^1_{N,j}(0))=(C_j^N)$. If, for $A>0$, $\|x\|>JA\sqrt{N}$, then, if $g(x)=\|x\|$, for $x\in\mathbb{N}^J$,

$$Q_f(g)(x)\leq J\overline{\eta}N-\underline{\lambda}J^2A^2N.$$

If we choose $\ell=JA$ such that $\gamma=\underline{\lambda}\ell^2-J\overline{\eta}>0$, by using Proposition 8.14 and Theorem 8.13 of Robert [61], we obtain

$$\mathbb{E}(\tau_N)\leq\frac{g(F^1_N(0))}{N\gamma}=\frac{1}{\gamma}.$$

The proposition is proved. □

Theorem 31 (Law of Large Numbers)**.** *If*

$$\lim_{N\to+\infty}\frac{F^1_N(0)}{\sqrt{N}}=\overline{f}^1_0=\left(\overline{f}^1_{0,j}\right),\ \textit{and}\ \left(\overline{F}_N(t)\right)\stackrel{\text{def.}}{=}\left(\frac{F^1_{N,j}\left(t/\sqrt{N}\right)}{\sqrt{N}}\right),$$

then the sequence of processes $\left(\overline{F}_N(t)\right)$ *is converging in distribution to the solution* $(\overline{f}^1(t))=(\overline{f}^1_j(t))$ *of the ODE,*

$$\left(\overline{f}^1_j\right)'(t)=c_j\eta_j-\lambda_j\overline{f}^1_j(t)\left\|\overline{f}^1(t)\right\|, \tag{3.28}$$

with $\overline{f}^1(0)=\overline{f}^1_0$.

The equilibrium point of the ODE (3.28) *is given by*

$$\overline{f}^1_{j,\infty}=\left(c_j\rho_j\,/\sqrt{c_1\rho_1+\cdots+c_J\rho_J}\right), \tag{3.29}$$

where (ρ_j) *is given by Definition 23.*

Proof. By integration of Relation (3.34), we obtain, for $t \geq 0$,

$$\overline{F}_{N,j}(t) = \overline{F}_{N,j}(0) + M^0_{N,j}(t) + \eta_j \int_0^t \left(\frac{C^N_j}{N} - \frac{\overline{F}_{N,j}(s)}{\sqrt{N}} \right) \mathrm{d}s - \lambda_j \int_0^t \overline{F}_{N,j}(s) \sum_{k=1}^{J} \overline{F}_{N,k}(s)\, \mathrm{d}s, \quad (3.30)$$

where $(M^0_N(t)) = (M^0_{N,j}(t), 1 \leq j \leq J)$ is the martingale defined by, for $1 \leq j \leq J$,

$$M^0_{N,j}(t) \stackrel{\text{def.}}{=} \frac{1}{\sqrt{N}} \int_0^{t/\sqrt{N}} \left[\mathcal{P}_{S_j}((0, \eta_j(C^N_j - F^1_{N,j}(s-))), \mathrm{d}s) - \eta_j(C^N_j - F^1_{N,j}(s))\, \mathrm{d}s \right] - \frac{1}{\sqrt{N}} \int_0^{t/\sqrt{N}} \left[\mathcal{P}_{F_j}\left(\left(0, \lambda_j F^1_{N,j}(s-) \sum_{k=1}^{J} \overline{F}_{N,k}(s-) \right), \mathrm{d}s \right) - \lambda_j F^1_{N,j}(s) \sum_{k=1}^{J} \overline{F}_{N,k}(s)\, \mathrm{d}s \right]. \quad (3.31)$$

Its previsible increasing process is given by

$$\left(\left\langle M^0_{N,j} \right\rangle (t) \right) = \left(\frac{\eta_j}{\sqrt{N}} \int_0^t \left(\frac{C^N_j}{N} - \frac{\overline{F}_{N,j}(s)}{\sqrt{N}} \right) \mathrm{d}s + \frac{\lambda_j}{\sqrt{N}} \int_0^t \overline{F}_{N,j}(s) \sum_{k=1}^{J} \overline{F}_{N,k}(s)\, \mathrm{d}s \right), \quad (3.32)$$

and $\left\langle M^0_{N,j}, M^0_{N,k} \right\rangle (t) = 0$, for $1 \leq j \neq k \leq J$. Lemma 29 shows the convergence

$$\lim_{N \to +\infty} \left(\mathbb{E}\left(\left\langle M^0_{N,j}(t), M^0_{N,k} \right\rangle (t) \right), 1 \leq j, k \leq J \right) = 0,$$

and, with Doob's Inequality, the martingale $(M^0_N(t))$ converges to 0, and also that, for the convergence in distribution

$$\lim_{N \to +\infty} \left(\frac{\overline{F}_{N,j}(t)}{\sqrt{N}} \right) = 0.$$

Standard arguments, using the criterion of the modulus of continuity, see Theorem 7.3 Billingsley [10] for example, give that the sequence of processes $(\overline{F}_N(t))$ is tight and that any limiting point $(\overline{f}^1(t)) = (\overline{f}^1_j(t))$ satisfies the identity

$$\overline{f}^1_j(t) = \overline{f}^1_j(0) + \eta_j c_j t - \lambda_j \int_0^t \overline{f}^1_j(s) \sum_{k=1}^{J} \overline{f}^1_k(s)\, \mathrm{d}s. \quad (3.33)$$

The theorem is proved. □

The fluctuations of $(F_N^1(t))$ on the timescale $(t/\sqrt{N})$ are now developed in the following theorem.

Theorem 32 (Central Limit Theorem)**.** *Under the assumption on the initial state of Theorem 31, if*

$$\left(\widehat{F}_N^1(t)\right) = \left(\widehat{F}_{N,j}^1(t)\right) \stackrel{\text{def.}}{=} \left(\frac{F_{N,j}^1\left(t/\sqrt{N}\right) - \sqrt{N}\,\overline{f}_j^1(t)}{\sqrt[4]{N}} \right),$$

where $\left(\overline{f}^1(t)\right)$ *is defined by Relation* (3.28) *and*

$$\lim_{N\to+\infty} \widehat{F}_N^1(0) = \widehat{f}_0^1 \in \mathbb{R}^J,$$

then the sequence of processes $(\widehat{F}_N^1(t))$ *is converging in distribution to* $(\widehat{F}^1(t))$*, the solution of the SDE*

$$\begin{aligned} \mathrm{d}\widehat{F}_j^1(t) = \sqrt{-\left(\overline{f}_j^1\right)'(t) + 2\eta_j c_j}\ \mathrm{d}B_j(t) \\ -\lambda_j\left(\widehat{F}_j^1(t)\left\|\overline{f}^1(t)\right\| + \overline{f}_j^1(t)\left\|\widehat{F}^1(t)\right\|\right)\mathrm{d}t, \quad (3.34) \end{aligned}$$

with $\widehat{F}^1(0)=\widehat{f}_0^1$*, where* $(B_j(t))$ *is the standard Brownian motion in* $\mathbb{R}^J$*.*

Proof. Relations (3.34) and (3.33) give the identity,

$$\begin{aligned} \widehat{F}_{N,j}^1(t) = \sqrt[4]{N}\left(\overline{F}_{N,j}^1(t) - \overline{f}_j^1(t)\right) = \widehat{F}_{N,j}^1(0) + \sqrt[4]{N} M_{N,j}^0(t) \\ + \eta_j \frac{C_j^N - c_j N}{N^{3/4}} t - \eta_j \int_0^t \frac{\overline{F}_{N,j}^1(s)}{\sqrt[4]{N}}\,\mathrm{d}s \\ -\lambda_j \int_0^t \widehat{F}_{N,j}^1(s) \sum_{k=1}^J \overline{F}_{N,k}^1(s)\,\mathrm{d}s - \lambda_j \int_0^t \overline{F}_{N,j}^1(s) \sum_{k=1}^J \widehat{F}_{N,k}^1(s)\,\mathrm{d}s, \quad (3.35) \end{aligned}$$

and, with Relation (3.32), for $1\le j\le J$,

$$\begin{aligned} \left(\left\langle \sqrt[4]{N} M_{N,j}^0 \right\rangle(t)\right) = \left(\eta_j \int_0^t \left(\frac{C_j^N}{N} - \frac{\overline{F}_{N,j}^1(s)}{\sqrt{N}} \right) \mathrm{d}s \right. \\ \left. + \lambda_j \int_0^t \overline{F}_{N,j}^1(s) \sum_{k=1}^J \overline{F}_{N,k}^1(s)\,\mathrm{d}s \right). \end{aligned}$$

From Lemma 29 and Theorem 31, we obtain that, for the convergence in distribution, the relation

$$\begin{aligned} \lim_{N\to+\infty}\left(\left\langle \sqrt[4]{N} M_{N,j}^0 \right\rangle(t)\right) = \left(\eta_j c_j t + \lambda_j \int_0^t \overline{f}_j^1(s) \sum_{k=1}^J \overline{f}_k^1(s)\,\mathrm{d}s \right) \\ = \left(\overline{f}_j^1(0) - \overline{f}_j^1(t) + 2\eta_j c_j t\right), \end{aligned}$$

holds, by Relation (3.33). Recall that

$$\left(\left\langle \sqrt[4]{N}M^0_{N,j}, \sqrt[4]{N}M^0_{N,k}\right\rangle(t)\right)=0$$

holds for $1\leq j\neq k\leq J$. Theorem 1.4 page 339 of Ethier and Kurtz [24] shows that the sequence of martingales $(\widehat{M}^0_N(t))$ converges in distribution to the distribution of the process

$$\left(\int_0^t \sqrt{-(\overline{f}^1_j)'(s)+2\eta_j c_j}B_j(\mathrm{d}s)\right),$$

where $(B_j(t))$ is a standard Brownian motion on $\mathbb{R}^J$. Using again Lemma (29), we have

$$\lim_{N\to+\infty}\left(\int_0^t \frac{\overline{F}^1_{N,j}(s)}{\sqrt[4]{N}}\,\mathrm{d}s\right)=(0).$$

The rest of the proof is standard, first by showing the tightness of $(\widehat{F}^1_N(t))$ and secondly by identifying it as the solution of an SDE. See the proof of Theorem 6.14 of [61] for example. The theorem is proved. □

The following proposition shows that the invariant distribution of the Markov process $(F^1_N(t))$ has in fact a simple expression. This is a consequence of Proposition 24, the reversibility property is in fact the additional (simple) result.

Proposition 33 (Invariant Distribution)**.** *The Markov process $(F^1_N(t))$ is reversible, and its invariant distribution π_N*

$$\pi_N(x)=\frac{1}{Z_N}\frac{1}{\|x\|!}\prod_{j=1}^J \rho_j^{x_j}\frac{C^N_j!}{(C^N_j-x_j)!x_j!},\qquad x\in\mathcal{S}_N, \tag{3.36}$$

where Z_N is a normalizing constant.

A version of Theorems (31) and (32) could probably be considered via a saddle-point analysis of the constant Z_N. This is not done in this paper.

3.4 Dynamical Arrivals

If the systems starts with few agents so that most of N particles are "free", when an agent created, it is paired with a free particle right away, at a rate proportional to N. This will happen repeatedly, via the successive steps of sequestration/de-sequestration, as long the number of free particles is sufficiently "large" so that sequestration occurs always before the degradation/death of an agent. The precise result is in fact a little more subtle than that. We show that, in the limit, on the timescale $t\mapsto Nt$, there remains a positive fraction of free particles of the order of N.

The state descriptor of the pairing process is in this case

$$(X_N(t))=(F_N(t),Z_N(t))=((F_{j,N}(t),j=1,\ldots,J),Z_N(t)).$$

It can be expressed as the solution of the SDEs (3.7) and (3.8), the initial conditions are assumed to satisfy the following scaling relations

$$\lim_{N\to+\infty}\frac{F_N(0)}{N}=\overline{f}_0\neq 0,\overline{f}_0{=}(\overline{f}_{0,j})\in\prod_{j=1}^{J}[0,c_j]^J,\text{ and }Z_N(0){=}z_0{\in}\mathbb{N},\qquad(3.37)$$

where $c{=}(c_j)$ is defined by Relation (3.5).

Initially, a fraction $\overline{f}_{0,j}$ of particles of type $j{\in}\{1,\ldots,J\}$ are free and there are z_0 free agents. Since the external input rate of agents is constant and equal to β, in order to have a positive fraction in N of particles paired with an agent, the natural time scale to consider is, at least, $t{\mapsto}Nt$.

The setting of the analysis will be that of averaging principles as presented in Kurtz [45]. As it will be seen there are specific technical difficulties related to the scaling framework which we introduce now.

Definition 34 (Scaled Processes)**.** *For $N{>}0$, $(\overline{X}_N(t))\stackrel{\text{def.}}{=}(\overline{F}_N(t),Z_N(Nt))$, with*

$$\left(\overline{F}_N(t)\right)\stackrel{\text{def.}}{=}\left(\frac{F_N(Nt)}{N}\right)=\left(\frac{F_{j,N}(Nt)}{N},j{=}1,\ldots,J\right)\in\prod_{j=1}^{J}\left[0,\frac{C_{j,N}}{N}\right].\quad(3.38)$$

For $t{\geq}0$, we have $\|F_N(t)\|{\leq}1$ since $F_{j,N}(t){\leq}C_j^N$,for all $1{\leq}j{\leq}J$.

The occupation measure *is the random measure on $H\stackrel{\text{def.}}{=}\mathbb{R}_+{\times}[0,1]^J{\times}\mathbb{N}$ defined by*

$$\langle\Lambda_N,g\rangle\stackrel{\text{def.}}{=}\int_0^{+\infty}g\left(s,\left(\frac{F_{j,N}(Ns)}{N}\right),Z_N(Ns)\right)\mathrm{d}s,\qquad(3.39)$$

for a continuous function g with compact support on H,

Note that the "slow" process $(\overline{F}_N(t))$ is included in the definition of the occupation measure Λ_N. The reason is that the timescale is too fast, of the order of N^2 in fact, to get directly convenient tightness properties for the sequence of processes $(\overline{F}_{j,N}(t))$.

We can have a glimpse of this problem as follows. If $(\overline{M}_{j,N}(t))$ is the martingale of Relation (3.10), it does not clearly converges in distribution to 0 as N gets large as it could be expected if a "standard" averaging principle were true. Indeed Relation (3.12) gives, for $1{\leq}j{\leq}J$,

$$\left(\left\langle\overline{M}_{j,N}\right\rangle(t)\right)=\left(\eta_j\int_0^t\left(\frac{C_{j,N}}{N}{-}\overline{F}_{j,N}(s)\right)\mathrm{d}s{+}\lambda_j\int_0^t\overline{F}_{j,N}(s)Z_N(Ns)\,\mathrm{d}s\right),$$

which does not seem to vanish.

We state the main result of this paper.

Theorem 35 (Averaging Principle)**.** *Under the scaling assumption* (3.5) *and if $(F_N(0)/N)$ converges to $\overline{f}_0{\neq}0$, then the sequence of processes $(\|F_N(Nt)\|/N)$*

converges in distribution to $(H(t))$*, defined by, for* $t{\geq}0$*,* $H(t){\in}(0,1)$ *is the unique solution of the relation*

$$\int_{\|\overline{f}_0\|}^{H(t)} \frac{1}{\delta\phi(u){-}\beta}\,\mathrm{d}u = t, \tag{3.40}$$

where ϕ *is defined by Relation* (3.6)*.*

Furthermore the sequence (Λ_N) *is converging in distribution to the measure* Λ_∞ *on* $H{=}\mathbb{R}_+{\times}[0,1]^J{\times}\mathbb{N}$*, such that*

$$\langle \Lambda_\infty, g\rangle \stackrel{\text{def.}}{=} \int_{\mathbb{R}_+\times\mathbb{N}} g\left(s,\left(f_j(s)\right),x\right) P_{\phi(H(s))}(\mathrm{d}x)\,\mathrm{d}s, \tag{3.41}$$

for a non-negative Borelian function g *on* H*, where, for* $1{\leq}j{\leq}J$*,*

$$f_j(t) = \frac{\rho_j}{\rho_j{+}\phi(H(t))}c_j, \tag{3.42}$$

and, for $y{>}0$*,* P_y *is a Poisson distribution with parameter* y*.*

Note that we have a convergence in distribution $(\|F_N(Nt)\|/N)$, but not of the processes $(F_{j,N}(Nt)/N)$, $j{=}1,\ldots,J$.. The convergence in distribution for this J-dimensional process is weaker, it is expressed through the sequence of occupation measures (Λ_N). See Dawson [19] for general definitions and results for the convergence in distribution of random measures.

It is not difficult to see that, under Condition (3.37) for the initial conditions, one cannot have a convergence in distribution of $(F_{j,N}(Nt)/N)$. Otherwise, its limit would be $(f_j(t))$, but this would imply that the asymptotic initial point $(\overline{f}_{0,j})$ would satisfy the relation

$$\overline{f}_{0,j} = \frac{\rho_j}{\rho_j{+}\phi(\overline{f}_0)}c_j,$$

which is not the case a priori. Asymptotically the vector $(F_{j,N}(Nt)/N)$ lives in a one dimensional curve of the state space, this is due to the fast processes which lead to a kind of state space collapse. See Propositions 44 and 45.

Corollary 36 (Equilibrium)**.** *Under the assumptions, and with the notations of, Theorem* (35)*, for* $1{\leq}j{\leq}J$*,*

$$\lim_{t\to+\infty} H(t) = H_\infty = \sum_{j=1}^{J} \frac{\rho_j}{\rho_j{+}\beta/\delta}c_j,$$

The quantity H_∞ is the asymptotic fraction of free particles. The proof of this theorem is achieved in several steps. The general picture is that nevertheless a kind of averaging principle holds, the slow process being $(\overline{F}_N(t))$ and the "fast" process is $(Z_N(Nt))$. The general method in this domain is described in Kurtz [45], see also Papanicolaou et al. [56] and Freidlin and Wentzell [29]. It turns

out that, due to the very fast timescale mentioned above, the slow process has to be included in the definition of the occupation measure, see Definition 34. This situation leads to technical difficulties, to identify the invariant measures of fast processes in particular.

Definition 37. *For $a>0$, the stopping time $\tau_N(a)$ is defined by*

$$\tau_N(a) \stackrel{\text{def.}}{=} \inf\left\{ t>0 : \|F_N(Nt)\| = \sum_{j=1}^{J} F_{j,N}(Nt) \leq aN \right\}. \tag{3.43}$$

To prove convenient tightness properties of a scaled version of $(X_N(t))$, we first derive some technical results. In a first step, we fix some $a_0 \in (0, \|\overline{f}_0\|)$ and we will work with a "stopped" *occupation measure*, it is the random measure on H defined by

$$\left\langle \Lambda_N^0, g \right\rangle \stackrel{\text{def.}}{=} \int_0^{\tau_N(a_0)} g\left(s, \left(\frac{F_{j,N}(Ns)}{N}\right), Z_N(Ns)\right) \mathrm{d}s, \tag{3.44}$$

for a continuous function g with compact support on H. The motivation of the stopped occupation is due to a technical argument used for the identification of the invariant distributions of fast processes. See Proposition 45.

Lemma 38. *If $\|\overline{f}_0\|>0$, for $a \in (0, \|\overline{f}_0\|)$, we have*

$$\lim_{N\to+\infty} \mathbb{P}\left(\tau_N(a) < \ell(a)\right) = 0,$$

with $\ell(a) \stackrel{\text{def.}}{=} \|\overline{f}_0\| - a/(2\beta)$, and the relation

$$\lim_{N\to+\infty} \left(\frac{Z_N(Nt)}{\sqrt{N}}\right) = (0)$$

holds for the convergence in distribution.

Proof. For $t>0$ and N sufficiently large, on the event $\{\|F_N(t)\| < aN\}$ there are at least $(\|F_N(0)\| - \lceil aN \rceil - z_0)$ new agents created up to time Nt. Consequently for $y>0$,

$$\{\tau_N(a) \leq y\} \subset \left\{ \mathcal{P}_z^+((0,\beta)\times(0,yN)\} \geq \|F_N(0)\| - \lceil aN \rceil - z_0 \right\},$$

by Relation (3.8). The first assertion follows from the law of large numbers for Poisson processes.

We now show that there exists a coupling of $(Z_N(t))$ with $(L_0(t))$, the state process of an $M/M/\infty$ queue such that the relation $Z_N(t) \leq (L_0(N^2 t))$ holds for all $t < \tau_N(a)$. See Section 3.6.2 on the $M/M/\infty$ queue.

In state $z \in \mathbb{N}$, the jump rates of the process $(Z_N(t))$ in state $((f_j), z)$ at time t are given by

$$\begin{cases} +1, & \beta + \displaystyle\sum_{j=1}^{J} \eta_j \left(C_j^N - f_j\right), \\ -1, & \delta z + \displaystyle\sum_{j=1}^{J} \lambda_j f_j z. \end{cases}$$

Let $(L_0(t))$ the process of the $M/M/\infty$ queue with input rate $2\overline{\eta}$ and service rate $\delta+a\underline{\lambda}$, with Definition (3.4), with $L_0(0)=Z_N(0)=z_0$. We take N sufficiently large so that $\beta\leq\overline{\eta}N$. By comparing the jump rates, one can construct a version of $(Z_N(t), L_0(t))$ such that the relation

$$Z_N(Nt) \leq L_0(N^2t) \tag{3.45}$$

holds for all $t<\tau_N(a)$. For $\varepsilon>0$, let $T_N(\varepsilon)=\inf\left\{t : L_0(t)\geq\varepsilon\sqrt{N}\right\}$. Proposition 47 shows that the sequence of random variables

$$\left(\left(\frac{2\overline{\eta}}{a\underline{\lambda}}\right)^{\lceil\varepsilon\sqrt{N}\rceil}\frac{T_N(\varepsilon)}{(\lceil\varepsilon\sqrt{N}\rceil-1)!}\right)$$

converges in distribution to an exponential random variable. In particular, for any $t>0$,

$$\lim_{N\to+\infty}\mathbb{P}(T_N(\varepsilon)\leq N^2t)=0.$$

The proof of the lemma follows from the relation

$$\mathbb{P}\left(\sup_{s\leq t}\frac{Z_N(N(s\wedge\tau_N(a_0)))}{\sqrt{N}}>\varepsilon\right)\leq\mathbb{P}\left(\sup_{s\leq t}\frac{L_0(N^2s)}{\sqrt{N}}>\varepsilon\right)=\mathbb{P}\left(T_N\leq N^2t\right).$$

□

Proposition 39. *The sequence of random measures (Λ_N^0) defined by Relation (3.44) is tight and any limiting point Λ_∞^0 can be expressed as*

$$\left\langle\Lambda_\infty^0, f\right\rangle=\int_{\mathbb{R}_+\times[0,1]^J\times\mathbb{N}}f\,(s,x,p)\,\pi_s(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s, \tag{3.46}$$

for any function $f\in\mathcal{C}_c(\mathbb{R}_+\times[0,1]^J\times\mathbb{N})$, where (π_s) is an optional process with values in the space of probability distributions on $[0,1]^J\times\mathbb{N}$.

If $(\Lambda_{N_k}^0)$ is a sub-sequence of (Λ_N^0) converging to Λ_∞^0, then with the convention of Relation (3.9), for the convergence in distribution of processes,

$$\lim_{k\to+\infty}\left(\int g(x,z)z\Lambda_{N_k}^0([0,t],\mathrm{d}x,\mathrm{d}z)\right)=\left(\int g(x,z)z\Lambda_\infty^0([0,t],\mathrm{d}x,\mathrm{d}z)\right), \tag{3.47}$$

for any bounded continuous function g on $[0,1]^J\times\mathbb{N})$ and the limit is integrable for all $t\geq0$.

Proof. For $K>0$, with Relation (3.45) in the proof of Lemma 38, and with the same notations, the relation

$$\int_0^t\mathbb{1}_{\{Z_N(Ns)\geq K\}}\,\mathrm{d}s\leq\int_0^t\mathbb{P}(L_0(N^2s)\geq K)\,\mathrm{d}s$$

holds on the event $\{\tau_N(a_0)\geq t\}$. Consequently, for $t<\ell(a_0)$,

$$\mathbb{E}\left(\Lambda_N^0([0,t]\times[0,1]^J\times[K,+\infty])\right) = \mathbb{E}\left(\int_0^{t\wedge\tau_N(a_0)} \mathbb{1}_{\{Z_N(Ns)\geq K\}}\,\mathrm{d}s\right)$$
$$\leq \mathbb{E}\left(\mathbb{1}_{\{\tau_N(a_0)>t\}}\int_0^t \mathbb{1}_{\{L_0(N^2s)\geq K\}}\,\mathrm{d}s\right) \leq \frac{1}{N^2}\int_0^{N^2t}\mathbb{P}(L_0(s)\geq K)\,\mathrm{d}s.$$

Since the Markov process $(L_0(t))$ converges in distribution to a Poisson distribution with parameter $2\overline{\eta}/(\underline{\lambda}a_0)$, see Section 3.6.2. With Lemma 38, we obtain the relation

$$\limsup_{N\to+\infty}\mathbb{E}\left(\Lambda_N^0([0,t]\times[0,1]^J\times[K,+\infty])\right) \leq \mathbb{P}(\mathcal{N}_1(0,2\overline{\eta}/(\underline{\lambda}a_0))\geq K)\,t,$$

where $\mathcal{N}_1$ is a Poisson process on $\mathbb{R}_+$ with rate 1. In particular, one can choose K sufficiently large such that

$$\sup_N \mathbb{E}\left(\Lambda_N^0([0,t]\times[0,1]^J\times[K,+\infty])\right)$$

is arbitrarily small. Lemma 1.3 of Kurtz [45] shows that the sequence (Λ_N^0) is tight, and Lemma 1.4 of the same reference gives the representation (3.46).

For the second part of the proposition, Relation (3.45) in the proof of Lemma 38 and the Cauchy-Schwartz' Inequality give, for $s\leq t$,

$$\mathbb{E}\left(\left(\int g(x,z)z\Lambda_N^0([s,t],\mathrm{d}x,\mathrm{d}z)\right)^2\right) = \mathbb{E}\left(\left(\int_{s\wedge\tau_N(a_0)}^{t\wedge\tau_N(a_0)} g\left(\overline{X}_N(s)\right)Z_N(Ns)\,\mathrm{d}s\right)^2\right)$$
$$\leq \|g\|_\infty\mathbb{E}\left(\left(\int_{s\wedge\tau_N(a_0)}^{t\wedge\tau_N(a_0)} L_0(N^2s)\,\mathrm{d}s\right)^2\right) \leq \|g\|_\infty\mathbb{E}\left(\left(\int_s^t L_0(N^2s)\,\mathrm{d}s\right)^2\right)$$
$$\leq (t-s)\|g\|_\infty\int_s^t \mathbb{E}\left(L_0(N^2s)^2\right)\mathrm{d}s \leq (t-s)^2\|g\|_\infty\sup_{u\geq 0}\mathbb{E}\left(L_0(u)^2\right).$$

Kolmogorov-Čentsov's criterion, implies that the sequence of processes

$$\left(\int g(x,z)z\,\Lambda_N^0([0,t],\mathrm{d}x,\mathrm{d}z)\right)$$

is tight for the convergence in distribution and any of its limiting points is a continuous process. See Theorem 2.8 and Problem 4.11, page 64 of Karatzas and Shreve [41] for example.

For $t>0$ and $C>0$, for the convergence in distribution we have

$$\lim_{k\to+\infty}\int g(x,z)\,(z\wedge C)\;\Lambda_{N_k}^0([0,t],\mathrm{d}x,\mathrm{d}z) = \int g(x,z)\,(z\wedge C)\;\Lambda_\infty^0([0,t],\mathrm{d}x,\mathrm{d}z).$$

Using again Relation (3.45), with the same argument as before,

$$\mathbb{E}\left(\int g(x,z)\,(z\wedge C)\;\Lambda^0_{N_k}([0,t],\mathrm{d}x,\mathrm{d}z)\right)$$
$$\leq \|g\|_\infty \int_0^t \mathbb{E}\left(L_0(N^2u)\wedge C\right)\mathrm{d}u \leq t\|g\|_\infty \sup_{u\geq 0}\mathbb{E}\left(L_0(u)^2\right) < +\infty,$$

by letting first k and then C go to infinity, we obtain the relation

$$\mathbb{E}\left(\int g(x,z)z\,\Lambda^0_\infty([0,t],\mathrm{d}x,\mathrm{d}z)\right) < +\infty,$$

for all $t\geq 0$. Similarly, we have

$$\mathbb{E}\left(\int g(x,z)z\mathbb{1}_{\{z\geq C\}}\,\Lambda^0_{N_k}([0,t],\mathrm{d}x,\mathrm{d}z)\right)$$
$$\leq \|g\|_\infty \int_0^t \mathbb{E}\left(L_0(N_k^2u)\mathbb{1}_{\left\{L_0(N_k^2u)\geq C\right\}}\right)\mathrm{d}u \leq \frac{t}{C}\|g\|_\infty \sup_{u\geq 0}\mathbb{E}\left(L_0(u)^2\right),$$

and therefore the convergence in distribution

$$\lim_{k\to+\infty}\int g(x,z)z\,\Lambda^0_{N_k}([0,t],\mathrm{d}x,\mathrm{d}z) = \int g(x,z)z\,\Lambda^0_\infty([0,t],\mathrm{d}x,\mathrm{d}z),$$

for $t>0$. For $p\geq 1$ and $0\leq t_1\leq\cdots\leq t_p$, this convergence also clearly holds for finite marginals at (t_i). The proposition is proved. □

If f is a non-negative Borelian function on $\mathbb{R}_+^J\times\mathbb{N}$, the SDEs (3.7) and (3.8) give directly the relations

$$\begin{aligned} f\left(\overline{F}_N(t),Z_N(Nt)\right) &= f\left(\overline{F}_N(0),Z_N(0)\right)+M_{f,N}(t) \\ &+\sum_{j=1}^J \lambda_j\int_0^t N\Delta_{-e_j/N,-1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)F_{j,N}(Ns)Z_N(Ns)\,\mathrm{d}s \\ &+\sum_{j=1}^J \eta_j\int_0^t N\Delta_{e_j/N,1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)(C_j^N-F_{j,N}(Ns))\,\mathrm{d}s \\ &+\beta N\int_0^t \Delta_{0,1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)\mathrm{d}s \\ &+\delta N\int_0^t \Delta_{0,-1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)Z_N(Ns)\,\mathrm{d}s, \end{aligned} \tag{3.48}$$

with the notation, for x, $u\in\mathbb{R}_+^J$ and y, $b\in\mathbb{N}$,

$$\Delta_{u,v}(f)(x,y) \stackrel{\text{def.}}{=} f(x+u,y+v)-f(x,y),$$

and, for $1\leq j\leq J$, e_j is the jth unit vector of $\mathbb{R}^J$.

The process $(M_{f,N}(t))$ is a square integrable martingale whose previsible increasing process is given by

$$\begin{aligned}\langle M_{f,N}\rangle(t) &= \beta N\int_0^t \Delta_{0,1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)^2\mathrm{d}s \\ &+\delta N\int_0^t \Delta_{0,-1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)^2 Z_N(Ns)\,\mathrm{d}s \\ &+\sum_{j=1}^J \lambda_j\int_0^t N\Delta_{-e_j/N,-1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)^2 F_{j,N}(Ns)Z_N(Ns)\,\mathrm{d}s \\ &+\sum_{j=1}^J \eta_j\int_0^t N\Delta_{e_j/N,1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)^2 (C_j^N-F_{j,N}(Ns))\,\mathrm{d}s.\end{aligned} \tag{3.49}$$

If we divide by N Relation (3.48) taken at Nt, we get

$$\begin{aligned}\frac{1}{N}\left(f\left(\overline{F}_N(t),Z_N(Nt)\right)-f\left(\overline{F}_N(0),Z_N(0)\right)\right) &= \frac{1}{N}M_{f,N}(t) \\ &+\sum_{j=1}^J \lambda_j\int_0^t N\Delta_{-e_j/N,-1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)\overline{F}_{j,N}(s)Z_N(Ns)\,\mathrm{d}s \\ &+\sum_{j=1}^J \eta_j\int_0^t N\Delta_{e_j/N,1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)\left(\frac{C_j^N}{N}-\overline{F}_{j,N}(s)\right)\mathrm{d}s \\ &+\beta\int_0^t \Delta_{0,1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)\mathrm{d}s \\ &+\delta\int_0^t \Delta_{0,-1}(f)\left(\overline{F}_N(s),Z_N(s)\right)Z_N(Ns)\,\mathrm{d}s.\end{aligned} \tag{3.50}$$

Lemma 40. *If f is a bounded $\mathcal{C}_1$ function with compact support on $\mathbb{R}_+^J\times\mathbb{N}$, then the sequence of martingales $(M_{f,N}(t)/N, t<\ell(a_0))$ of Relation* (3.50) *converges in distribution to* 0.

Proof. We take care of the third term $(A_N(t))$ of $(\langle M_{f,N}/N\rangle(Nt))$ in Relation (3.49), the arguments are similar for the others, even easier. For $1\le j\le J$, denote by $(A_{j,N}(t))$ the jth term of $(A_N(t))$, then

$$A_{j,N}(t)\overset{\text{def.}}{=}\frac{\lambda_j}{N}\int_0^t N\Delta_{-e_j/N,1}(f)\left(\overline{F}_N(s),Z_N(Ns)\right)^2\overline{F}_{j,N}(s)Z_N(Ns)\,\mathrm{d}s,$$

then, again with Relation (3.45) of the proof of Lemma 38, since $(\overline{F}_{j,N}(t))$ is bounded by 1,

$$\mathbb{E}(A_{j,N}(t))\le\frac{\lambda_j}{N}\left\|\frac{\partial f}{\partial x_j}\right\|_\infty^2\left(\mathbb{E}\left(\int_0^t L_0(N^2s)\,\mathrm{d}s\right)+t\mathbb{P}(\tau_N(a_0)\le\ell(a_0))\right),$$

since $(\mathbb{E}(L_0(t))$ is converging as t goes to infinity, we have

$$\lim_{N\to+\infty}\mathbb{E}(A_{j,N}(t))=0.$$

Doob's Inequality shows the convergence of $(M_{f,N}(t))/N, t<\ell(a_0))$ to 0. $\square$

Proposition 41. *If Λ^0_∞ is a limiting point of (Λ^0_N) with the representation* (3.46) *of Proposition 39 then, for any continuous function g on $\mathbb{R}_+^J\times\mathbb{N}$, the relation*

$$\left(\int_0^t\int_{[0,1]^J\times\mathbb{N}} g(x,p)\pi_s(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s, t<\ell(a_0)\right)$$
$$=\left(\int_0^t\int_{[0,1]^J}\mathbb{E}\left(g\left(x,\mathcal{N}_1\left(\left[0,\frac{\langle\eta,c-x\rangle}{\langle\lambda,x\rangle}\right]\right)\right)\right)\pi_s^1(\mathrm{d}x)\,\mathrm{d}s, t<\ell(a_0)\right), \quad (3.51)$$

holds almost surely, where π_t^1 is the marginal of π_t on $\mathbb{R}_+^J$, λ, η and $c\in\mathbb{R}_+^J$ are given by Definition 23 and $\ell(a_0)$ in Lemma 38, and $\mathcal{N}_1$ is a Poisson process on $\mathbb{R}_+$ with rate 1.

Furthermore, almost surely,

$$\int_0^{\ell(a_0)}\pi_s^1\left(x\in[0,1]^J:\sum_{j=1}^J x_j<a_0\right)\mathrm{d}s=0. \quad (3.52)$$

Relation (3.51) states that, for almost all $t<\ell(a_0)$, π_t conditioned on the first coordinate x is a Poisson distribution with parameter $\langle\eta,c-x\rangle\,/\,\langle\lambda,x\rangle$. Note that Relation (3.52) shows that the denominator $\langle\lambda,x\rangle$ is lower bounded by $\underline{\lambda}a_0$ almost surely for π_t, for $t<\ell(a_0)$.

Proof. Let $(\Lambda^0_{N_k})$ be a subsequence of (Λ^0_N) converging to some Λ^0_∞ of the form (3.46). By letting k go to infinity in Relation (3.50), with Lemma 38, Relation (3.47) of Proposition 39, and Lemma 40, for any continuous function g with compact support on $[0,1]^J\times\mathbb{N}$,

$$\int_0^t\int_{[0,1]^J\times\mathbb{N}}(g(x,p-1)-g(x,p))\left(\sum_{j=1}^J\lambda_j x_j\right)p\pi_s(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s$$
$$+\int_0^t\int_{[0,1]^J\times\mathbb{N}}(g(x,p+1)-f(x,p))\left(\sum_{j=1}^J\eta_j(c_j-x_j)\right)\pi_s(\mathrm{d}x,\mathrm{d}p)\,\mathrm{d}s=0,$$

holds almost surely for all $t<\ell(a_0)$. Hence we have

$$\int_{[0,1]^J\times\mathbb{N}}\lambda(g(x,p-1)-g(x,p))\,\langle\lambda,x\rangle\,p\pi_t(\mathrm{d}x,\mathrm{d}p)$$
$$+\int_{[0,1]^J\times\mathbb{N}}\eta(g(x,p+1)-f(x,p))\,\langle\eta,c-x\rangle\,\pi_t(\mathrm{d}x,\mathrm{d}p)=0,$$

almost everywhere on $t\in\mathbb{R}_+$, or if $g(x,p)=g_1(x)g_2(p)$,

$$\int_{\mathbb{R}_+\times\mathbb{N}}g_1(x)\Omega[x](g_2)(p)\pi_t(\mathrm{d}x,\mathrm{d}p)=0,$$

where, for $h : \mathbb{N} \to \mathbb{R}_+$,

$$\Omega[x](h)(p) \stackrel{\text{def.}}{=} \langle \eta, c-x \rangle \, (h(p+1)-h(p)) + \langle \lambda, x \rangle \, p(h(p-1)-h(p)).$$

Let $\pi_t(\cdot \mid x)$ be the conditional probability measure $\pi_t(\mathrm{d}x, \mathrm{d}p)$ given $x \in \mathbb{R}_+$, then we have, $\pi_t(\mathrm{d}x, \mathbb{N})$ almost everywhere

$$\int \Omega[x](g_2)(p) \pi_t(\mathrm{d}p \mid x) = 0,$$

for all functions g_2 with finite support. Since, for $x>0$, $\Omega[x]$ is the Q-matrix of an $M/M/\infty$ queue, the last relation gives that $\pi_t(\mathrm{d}p|x)$ is its invariant distribution, i.e. it is a Poisson distribution with parameter $\langle \eta, c-x \rangle \, / \, \langle \lambda, x \rangle$.

Relation (3.52) is simple a consequence of Lemma 38.

The proposition is proved.

□

From now on, we fix (N_k) an increasing sequence such the sequence $(\Lambda^0_{N_k})$ is converging to Λ^0_∞ with a representation given by Relations (3.46) and (3.51). The following corollary is a direct consequence of Propositions 39 and 41.

Corollary 42. *If g is a continuous bounded function on $[0,1]^J$, then, for the convergence in distribution,*

$$\lim_{k \to +\infty} \left(\int_0^t g\left(\overline{F}_{N_k}(s)\right) Z_{N_k}(N_k s) \, \mathrm{d}s, t<\ell(a_0) \right) = \left(\int_0^t \int_{[0,1]^J} g(x) \frac{\langle \eta, c-x \rangle}{\langle \lambda, x \rangle} \pi^1_s(\mathrm{d}x) \, \mathrm{d}s, t<\ell(a_0) \right). \quad (3.53)$$

The following proposition is a convergence result for the scaled process $(\overline{F}_N(t))$. It is weaker that the convergence stated in Theorem 35 clearly, but this is a crucial ingredient to establish the theorem in fact.

Proposition 43. *The sequence of processes $(\|\overline{F}_{N_k}(t)\|, t<\ell(a_0))$ is converging in distribution to*

$$(H(t), t<\ell(a_0)) \stackrel{\text{def.}}{=} \left(\|\overline{f}_0\| + \int_0^t \int_{[0,1]^J} \left(\delta \frac{\langle \eta, c-x \rangle}{\langle \lambda, x \rangle} - \beta \right) \pi^1_s(\mathrm{d}x) \, \mathrm{d}s, t<\ell(a_0) \right).$$

Proof. We define, for $t \geq 0$,

$$\widetilde{Z}_N(t) = Z_N(t) + \sum_{j=1}^J (N_j - F_{j,N}(t)) = N - \|F_N(t)\| + Z_N(t), \quad (3.54)$$

$\widetilde{Z}_N(t)$ is the total number of agents (free or paired) of the system at time t. Using the SDEs (3.10) and (3.11), we have

$$\frac{\widetilde{Z}_N(Nt)}{N} = M_{Z,N}(t) + \frac{\widetilde{Z}_N(0)}{N} + \beta t - \delta \int_0^t Z_N(Ns) \, \mathrm{d}s, \quad (3.55)$$

where $(M_{Z,N}(t))$ is a local martingale whose previsible increasing process is given by

$$(\langle M_{Z,N}\rangle(t)) = \left(\frac{1}{N}\left(\beta t+\delta\int_0^t Z_N(Ns)\,\mathrm{d}s\right)\right). \tag{3.56}$$

Using again Doob's Inequality, Lemma 38 and Relation (3.54), the proposition is proved. □

The next proposition gives a characterization of the process (π_s^1) which will be elucidated in Proposition 45.

Proposition 44. *If g is a $\mathcal{C}_1$-function on $[0,1]^J$, then, almost surely,*

$$\left(\int_0^t\int_{[0,1]^J}\sum_{j=1}^J\frac{\partial g}{\partial x_j}(x)\left(\lambda_j x_j\frac{\langle\eta,c-x\rangle}{\langle\lambda,x\rangle}-\eta_j(c_j-x_j)\right)\pi_s^1(\mathrm{d}x)\,\mathrm{d}s, t<\ell(a_0)\right)=(0). \tag{3.57}$$

Proof. For $t>0$, let g be a $\mathcal{C}_2$-function on $[0,1]^J$, Relation (3.50) gives,

$$\begin{aligned}\frac{1}{N}g\left(\overline{F}_N(t)\right) &= \frac{1}{N}g\left(\overline{F}_N(0)\right)+M_{g,N}(t)\\ &\quad+\sum_{j=1}^J\lambda_j\int_0^t N\nabla_{-e_j/N}(g)\left(\overline{F}_N(s)\right)\overline{F}_{j,N}(s)Z_N(Ns)\,\mathrm{d}s\\ &\quad+\sum_{j=1}^J\eta_j\int_0^t N\nabla_{e_j/N}(g)\left(\overline{F}_N(s)\right)\left(\frac{C_j^N}{N}-\overline{F}_{j,N}(s)\right)\mathrm{d}s,\end{aligned} \tag{3.58}$$

and, since

$$\begin{aligned}(\langle M_{g,N}\rangle(t)) &= \left(\frac{1}{N^2}\sum_{j=1}^J\lambda_j\int_0^t\left(N\nabla_{-e_j/N}(g)\left(\overline{F}_N(s)\right)\right)^2\overline{F}_{j,N}(s)Z_N(Ns)\,\mathrm{d}s\right.\\ &\quad\left.+\frac{1}{N^2}\sum_{j=1}^J\eta_j\int_0^t\left(N\nabla_{e_j/N}(g)\left(\overline{F}_N(s)\right)\right)^2\left(\frac{C_j^N}{N}-\overline{F}_{j,N}(s)\right)\mathrm{d}s\right),\end{aligned}$$

Relation (3.47) shows that, for $t<\ell(a_0)$, $\mathbb{E}(\langle M_{g,N_k}\rangle(t))$ is converging to 0, the martingale $(M_{g,N_k}(t))$ converges therefore in distribution to 0 as k gets large. By using again Relation 3.47 and the differentiability properties of g, it is easy to conclude the proof of the proposition. □

For $1\le j\le J$, by taking a function $g(x)=h(x_j)$, $x=(x_i)\in[0,1]^J$, where h is $\mathcal{C}_2$, we obtain, that the relation

$$\left(\int_0^t\int_{[0,1]^J}h'(x_j)\left(\lambda_j x_j\frac{\langle\eta,c-x\rangle}{\langle\lambda,x\rangle}-\eta_j(c_j-x_j)\right)\pi_s^1(\mathrm{d}x)\,\mathrm{d}s, t<\ell(a_0)\right)=(0)$$

holds almost surely. Hence, almost surely, for any f in a dense subset of Borelian functions on $[0,1]$, the identity

$$\int_{[0,1]^J} f(x_j)\left(\lambda_j x_j \frac{\langle \eta, c-x\rangle}{\langle \lambda, x\rangle} - \eta_j(c_j - x_j)\right)\pi_t^1(\mathrm{d}x)$$

holds almost everywhere for $t\in\mathbb{R}_+$, with respect to Lebesgue's measure.

If $U(t)=(U_j(t))$ is a random variable with distribution π_t^1, the last relation can be translated in terms of conditional expectation, almost surely

$$\lambda_j U_j(t)\mathbb{E}\left(\frac{\langle \eta, c-U(t)\rangle}{\langle \lambda, U(t)\rangle}\middle| U_j(t)\right) = \eta_j(c_j - U_j(t)),$$

almost everywhere for $t\geq 0$. The following proposition is the key step to identify completely the limit points of $((\|\overline{F}_N(t)\|), \Lambda_N^0)$.

Proposition 45. *Let $U=(U_j)$ be a random variable on $\prod_1^J[0,c_j]$, such that , almost surely, $\|U\|\geq\eta$, for some $\eta>0$, and, for $1\leq j\leq J$,*

$$\lambda_j U_j \mathbb{E}\left(\frac{\langle \eta, c-U\rangle}{\langle \lambda, U\rangle}\middle| U_j\right) = \eta_j(c_j - U_j), \tag{3.59}$$

then, almost surely,

$$U_j = \frac{\rho_j c_j}{\phi(\|U\|)+\rho_j},$$

where for $y\in(0,1)$, $\phi(y)$ is the unique solution of the equation

$$\sum_{j=1}^J \frac{\rho_j}{\rho_j+\phi(y)} c_j = y,$$

and (ρ_j) is given by Definition 23 and ϕ in Theorem 35.

Proof. Define, for $1\leq j\leq J$,

$$A_j \stackrel{\text{def.}}{=} \frac{\eta_j(c_j-U_j)}{\lambda_j U_j}, \quad B_j \stackrel{\text{def.}}{=} \lambda_j U_j,$$

and $\mathcal{F}_j$ is the σ-field associated to U_j. Note that, because of the assumptions the variables A_j and B_j are bounded by Relation (3.59), this identity can be re-written as

$$\mathbb{E}\left(\frac{\sum_k A_k B_k}{\sum_k B_k}\middle|\mathcal{F}_j\right) = A_j,$$

or, since A_j and B_j are $\mathcal{F}_j$-measurable

$$\mathbb{E}\left(\frac{\sum_k (A_k - A_j)A_j B_j B_k}{\sum_k B_k}\middle|\mathcal{F}_j\right) = 0,$$

this identity gives therefore the relation

$$\mathbb{E}\left(\frac{\sum_{j,k}(A_j-A_k)^2B_jB_k}{\sum_k B_k}\right)=2\sum_{j=1}^{J}\mathbb{E}\left(\frac{\sum_k(A_j-A_k)A_jB_jB_k}{\sum_k B_k}\right)=0,$$

since the variables B_j, j=1,...,J are almost surely positive, this implies that, almost surely, A_j=A_1 for j=1,...,J, and therefore

$$U_j=\frac{\rho_j c_j}{A_1+\rho_j},$$

the proposition is proved. □

Proof of Theorem 34. We have only to gather all the results obtained up to now. Proposition 43 shows that, on the time interval $[0,\ell(a_0))$, the sequence of processes $(\|F_{N_k}(t)\|)$ converges in distribution to $(H(t))$ given by

$$(H(t))=\left(\|\overline{f}_0\|+\int_0^t\int_{[0,1]^J}\left(\delta\frac{\langle\eta,c-x\rangle}{\langle\lambda,x\rangle}-\beta\right)\pi_s^1(\mathrm{d}x)\,\mathrm{d}s\right)$$
$$=\left(\|\overline{f}_0\|+\int_0^t(\delta\phi(H(s))-\beta)\,\mathrm{d}s\right),\quad(3.60)$$

by Proposition 45. This determines completely $(H(t))$ as the solution of Relation (3.40), and therefore the convergence of the sequence $(\|F_N(t)\|)$. Relation (3.57) of Proposition 44 and Proposition 45 gives the desired expression. Relation (3.41) for the limit of the sequence (Λ_N^0) of occupation measures.

It is easily seen that the solution of Relation (3.40) satisfies Corollary 36, in particular, if $H(0)$>0, there exists w>0 such that $H(t)\geq w$ for all $t\geq 0$. The convergence in distribution obtained can be therefore extended to any finite interval of $\mathbb{R}_+$. The theorem is proved. □

3.5 Biological Background

In bacterial cells, protein production uses an important number of cell resources: macro-molecules such as *polymerases* and *ribosomes*, biological bricks of proteins, i.e., amino acids, and the energy necessary to build proteins, such as GTP. The two main steps associated with protein and RNA production are

— *Transcription.* When an RNA polymerase is bound to an active gene, it starts to make a copy of this gene. The product which is a sequence of nucleotides is an *RNA*. If the gene is associated to a protein, it is a *messenger* RNA, an *mRNA*. When the full sequence of nucleotides of the RNA has been successively assembled, the RNA is released in the cytoplasm.

— *Translation*. The step is achieved through another large macro-molecule: a *ribosome*. When a ribosome is bound to an mRNA, it builds a chain of amino-acids using the mRNA as a template to produce a linked chain of amino-acids, a *protein*.

There are several types of RNAs, outside mRNAs, tRNAs to carry amino-acids for the translation phase, rRNAs which are (large) building blocks of ribosomes. The average size of an mRNA is of the order of 300 nucleotids (nt), the size of an rRNA is of the order of 5000nt. Another class of RNAs, the small RNAs, or sRNAs, has been discovered in the 1970's, the 6S sRNAs in particular. The size of a sRNA is of the order of 50-100nt and their functional role of regulation has been identified around 2000, quite recently in fact. See Hindley [37] and Beisel and Storz [8].

In a biological context, the internal dynamics of cells can be, essentially, expressed in terms of pairing mechanisms of various couples of macro-molecules. The quite recent discovery of the existence of small RNAs and of their functional role in the regulation of the protein production has shed a new light on pairing mechanisms as a general tool to control gene expression. Pairing mechanisms are to regulate the growth of cells in order to adapt to a varying environment. Depending on external conditions, it may be desirable to reduce or speed-up the growth of the cell, and so the use of these macro-molecules. A specific regulation mechanism relies on a type of macro-molecules, sRNAs, small RNAs, which we will refer to as *agents* in the following. Their functional role has been discovered only recently in fact, at the end of the 20th century. They may bind/pair with one of these macro-molecules and the pairing may have several effect, depending on the type of sRNA: it can sequester a macro-molecule/particle and, therefore, reduce significantly its interactions with other components of the cell. Or it may speed-up the activity of the macro-molecule by increasing its interaction rate (affinity) with other components of the cell.

Due to thermal noise inside the cell, a particle/agent pair will split after a certain time. These separation mechanisms are as important as binding events are. The quantity of agents presents within the cell is also highly variable, and their quantity is actively regulated by the cell through a variety of regulatory mechanisms. In most prokaryotes, when a cell grows, its size increases. Without any action, the concentrations of agents/particles inside the cell begin to decrease by dilution. In the case of a limited number of agents, the cell has specific means of degrading agents to rapidly reduce their quantity if dilution is not enough. When the cell environment changes, the number of active agents may have to increase to adapt. It can be done by either generating new agents and particles or, more rapidly, by altering the quantity of active agents, i.e., agents bound to particles, by enhancing couplings. The duration of these transitions is an important characteristic of the control mechanism under consideration. In general it has two terms. The first term is linked to the production of new agents (and is therefore rather slow), while the second term is linked to a (quick) change in the number of active machines via bindings of agents with particles. The rate of adaptation depends on the combination of these two mechanisms,

and is therefore a crucial issue in the case of cells. Assuming that the environment is such that regulation is necessary, the fraction of particles that are bound to an agent is a measure of the efficiency of these mechanisms. We refer to Section 3.5 for references and further details on the biological context.

We describe several examples of regulation via pairing mechanisms. For the sake of simplicity, we give a rough description, not all macro-molecules, mechanisms involved are not mentioned. The main goal is of showing that pairing mechanisms play a major role concerning the regulation of the cell. References with much more details are mentioned.

Regulation of Transcription

The 6S sRNA is in charge of regulating the transcription phase, via a pairing with a σ-factor, a macromolecule necessary to the initiation of transcription. See Fromion et al. [31] for more details. With a slight abuse, by ignoring the σ-factor, the mechanism can be represented as a chemical reaction with three chemical species particles,

$$\text{RNAP}+\text{sRNA} \rightleftharpoons \text{RNAP}-\text{sRNA},$$

where RNAP stands for RNA polymerase. In such a context the pairing of RNAP and sRNA is seen as a sequestration of polymerases. See Waters and Storz [69], Wassarman and Storz [68] and Felden and Augagneur [26]. See Nitzan et al. [54].

Regulation of Translation

There is a similar regulation mechanism for ribosomes. The pairing of a macromolecule denoted as (p)ppGpp with the ribosome has the main effect of interfering with the initiation phase of translation and leading to abort the operation. See Yang et al. [71], Fer et al. [27] and Hauryliuk et al. [36] for more details.

Regulation of mRNAs by sRNAs

The translation can be also controlled via the mRNAs in the following way. The pairing of an sRNA and an mRNA modifies the translation efficiency of the mRNA. It can repress or enhance its activity. The pairing of an sRNA and mRNA . See Waters and Storz [69], Jagodnik et al. [38] and Beisel and Storz [8] for prokaryotic cells, and Flynt and Lai [28] for eukaryotic cells, with micro RNAs, miRNAs, acting on mRNAs. The references Jayaprakash and Das [40], Baker et al. [6] and Del Giudice et al. [20] study the Fokker-Planck equations for Markovian models of this mechanism.

3.6 Technical Results

3.6.1 The $M/M/1$ queue

We introduce a birth and death process on $\mathbb{N}$ which is in fact a reflected random walk on $\mathbb{N}$. An $M/M/1$ queue with input rate γ and service rate μ on $\mathbb{N}$ is a Markov process $(L_1(t))$ with Q-matrix matrix given by

$$z \mapsto \begin{cases} z+1, & \gamma, \\ z-1, & \mu \mathbb{1}_{\{z \geq 1\}}. \end{cases}$$

If $\gamma<\mu$, its invariant distribution is a geometric distribution with parameter γ/μ. Define

$$T_K = \inf \{t : L_\infty(t) \geq K\},$$

Proposition 5.11 of Chapter 5 of Robert [61] gives the asymptotic behavior of T_K when K is large.

Proposition 46. *If $\gamma<\mu$ and $L(0)=z_0$, then, as K goes to infinity, the random variable*

$$\left(\frac{\gamma}{\mu}\right)^K T_K$$

converges in distribution to an exponential random variable.

3.6.2 The $M/M/\infty$ queue

This is another classical birth and death process on $\mathbb{N}$. An $M/M/\infty$ queue with input rate γ and service rate μ on $\mathbb{N}$ is a Markov process with Q-matrix given by

$$z \mapsto \begin{cases} z+1, & \gamma, \\ z-1, & \mu z. \end{cases}$$

Its invariant distribution is a Poisson distribution with parameter γ/μ.

It can be seen in fact as a discrete Ornstein-Uhlenbeck process. See Chapter 6 of Robert [61] for example. For $K \in \mathbb{N}$, the hitting time of level K by the process $(L(t))$ of an $M/M/\infty$ input rate γ and service rate μ is defined by

$$T_K = \inf \{t : L_\infty(t) \geq K\}.$$

The following result, see Proposition 6.10 of [61] for example, is used to establish several tightness results.

Proposition 47. *If $L(0)=z_0$, as K goes to infinity, the random variable*

$$\left(\frac{\gamma}{\mu}\right)^K \frac{T_K}{(K-1)!}$$

converges in distribution to an exponential random variable.

Chapter 4

Regulation Of Translation

4.1 Gene Expression

Gene expression refers to the process of converting genetic information encoded in genes into functional proteins or RNA molecules/non coding RNA molecules. This process involves the transcription of DNA into RNA. When the produced RNA is coding, this first step is followed by a second process called translation. This process translates the coding sequence carried by the coding RNA, called mRNA, into an amino acid polypeptide. In general, after some folding, this polypeptide becomes a protein. Protein production holds immense significance within the cell as it not only facilitates its growth but also utilizes a significant portion of its resources. Understanding the mechanisms and regulation of gene expression has been a fundamental and complex pursuit in the field of biology, with substantial progress made recently.

This chapter will specifically focus on the translation phase. The mRNA molecule carries genetic information from DNA to the ribosomes, where translation takes place. In bacterial cells, ribosomes are composed of two sub-units: the small 30S sub-unit and the large 50S sub-unit. These sub-units combine to form a complete 70S ribosome, responsible for translating mRNA into proteins. Ribosomes consist of multiple ribosomal RNAs (rRNAs) and around 50 proteins (52 in *Bacillus subtilis*), making them the most resource-demanding macro-molecules to produce.

a. **Initiation:** The translation phase starts with the small ribosomal sub-unit binding to the mRNA molecule. The ribosome then scans the mRNA for a specific start codon (AUG), which marks the beginning of translation.

b. **Elongation:** Once the start codon is identified, the ribosome recruits the large ribosomal sub-unit and initiates the elongation step. During elongation, the ribosome reads the mRNA sequence using the genetic code, which consists of three-letter combinations known as codons. Each codon corresponds to a specific amino acid. Transfer RNA (tRNA) molecules carry the corresponding amino acids to the ribosome-mRNA complex.

c. **Termination:** Elongation continues until the ribosome encounters a stop codon. At this point, the ribosome releases the synthesized protein chain and dissociates from the mRNA molecule.

In summary, gene expression involves the transcription of DNA into RNA and the subsequent translation of RNA into proteins. The translation phase comprises initiation, elongation, and termination steps, where ribosomes read the mRNA sequence, synthesize proteins, and release them upon encountering stop codons.

4.2 Biological Background

Regulatory mechanisms are critical for bacterial cells to adapt to changing environments (in terms of availability of resources) and maintain an efficient cellular function. Many studies have been conducted on this topic, such as in Agustino and Collado-Vides [1] where the focus is on examining regulation in *E. coli*.

It is not an overstatement to claim that the intricate biological processes involved in protein synthesis, starting from its DNA sequence, can be precisely regulated to control protein levels. Attempting to provide a comprehensive and concise overview of all potential mechanisms for gene expression regulation is an impractical task. One way the cell regulates its transcriptional processes involves the action of transcription factors, which have the ability to either accelerate or inhibit the cell's activity. This mechanism was investigated by David J. Lee and colleagues (as documented in the study by David J. Lee et al. [18]).

In addition to transcriptional regulation, gene expression control encompasses translational regulation, to modulate protein synthesis. The regulation of gene expression not only controls transcription but also governs translation processes. Translational regulation involves impeding access to the initiation site of mRNA. See Claudio O. Gualerzi [16], thereby influencing the efficiency and timing of protein production.

This chapter focuses on the translational regulation associated with the utilization of a stringent response, which is a mechanism that happens when a cell encounters an unusual situation, such as a shortage of amino acids (defined and elaborated upon in section 4.2.1).

Overall, the complex regulatory mechanisms in bacterial cells ensure that gene expression is tightly controlled and coordinated, allowing the cell to respond to environmental cues and carry out essential functions for survival.

The goal of this chapter is to examine the influence of a regulatory mechanism that operates during the Translation phase, involving the sequestration of ribosomes in response to amino acid scarcity. Thus, our aim is to construct a stochastic model that captures the diverse ribosome states in relation to amino acid arrivals. This model will enable us to analyze how the distribution of ribosomes evolves over time under various environmental conditions.

4.2.1 The Stringent Response: A regulatory mechanism in bacterial translation

The stringent response has been extensively studied in the model bacterium *E. coli* since the late 1960s. Even though the stringent response has been extensively studied recently in *E. coli*, the molecular elements involved in its implementation are specific to this bacterium. In reality, in the majority of sequenced and annotated bacteria, the molecular players participating in and defining the stringent response are slightly different from those in *E. coli*.

Although the molecular players may differ, indicating specific biological implementation, the general principles and major actions integrated into our study are similar for most bacteria. Specifically, we refer to the research and results concerning the stringent response in another model bacterium. For *B. subtilis* and most bacteria, the molecular organization of the stringent response and its actions on the bacterium varies slightly.

As we will discuss later, the primary objective of the regulatory system that generates the stringent response in bacteria is to ensure a proper balance between the availability of each amino acid and the demand associated with the production of proteins. Evolution has exploited this mechanism to address other issues, which we will not consider here. These additional aspects often involve the production or degradation of a small nucleotide molecule called (p)ppGpp, and are directly related to growth rate management rather than amino acid level control during translation.

This regulation is mediated by one protein in *B. subtilis* or two proteins in *E. coli*. The first protein is capable of producing a specific metabolite called pppGpp or ppGpp (referred to as (p)ppGpp hereafter). When the availability of amino acids falls below some threshold, this first protein, called RelA in *E. coli*, is complemented by a second protein named SpoT. The second protein degrades (p)ppGpp under specific conditions that activate its degradation. Generally, in most bacteria, the first protein that produces (p)ppGpp (RelA in *E. coli*) also possesses a secondary function to degrade it. Therefore, this protein operates in a dual manner: when it is not activated to produce (p)ppGpp, it can degrade it. Lastly, although this protein is structurally similar to *E. coli*'s RelA, it has recently been named Rel, highlighting its differences from RelA, the protein in *E. coli*.

The presence of (p)ppGpp in the cell triggers a series of downstream effects. It inhibits the production of ribosomal RNAs (rRNAs), which are essential components of the ribosomes responsible for protein synthesis. This reduction in ribosomal RNA production limits the availability of functional ribosomes in the cell, leading to a slowdown in protein synthesis, what helps in conserving energy and resources under nutrient-limiting conditions. In addition, (p)ppGpp interferes with the initiation phase of translation, specifically targeting the formation of the initiation complex. By inhibiting the initiation of translation, (p)ppGpp prevents the binding of aminoacyl-tRNAs to the ribosome, disrupting the incorporation of amino acids into growing polypeptide chains. Therefore, it slows down or halts the translation process, leading to a decrease in the overall rate

of protein production.

The stringent response, mediated by (p)ppGpp, allows bacterial cells to adapt and respond to adverse environmental conditions by adjusting the rate of protein production in response to environmental conditions, nutrient availability, and other cellular signals. Amino acid starvation poses a significant risk as it can lead to an increased error rate during protein synthesis. This highlights the critical importance of translational regulation in bacterial cells. By modulating the initiation of translation, bacteria can effectively manage protein synthesis, ensuring accuracy and efficiency, even under conditions of amino acid scarcity.

In *E. coli*, SpoT plays an important role in cellular metabolism and stress response. Also known as (p)ppGpp synthetase/hydrolase, SpoT is an enzyme involved in the stringent response, a regulatory mechanism enabling bacteria to adapt to nutrient scarcity and environmental stress. SpoT assumes the responsibility of synthesizing and degrading (p)ppGpp signaling molecules, which function as global regulators of gene expression. Through its modulation of (p)ppGpp levels, SpoT exerts influence over diverse cellular processes such as transcription, translation, and metabolism. This capability helps the bacterium in surviving adverse conditions. Conversely, in *B. subtilis*, it is Rel that carries out the production and degradation of (p)ppGpp.

However, the models presented in this chapter are applicable to various bacterial cell types, transcending their inherent differences.

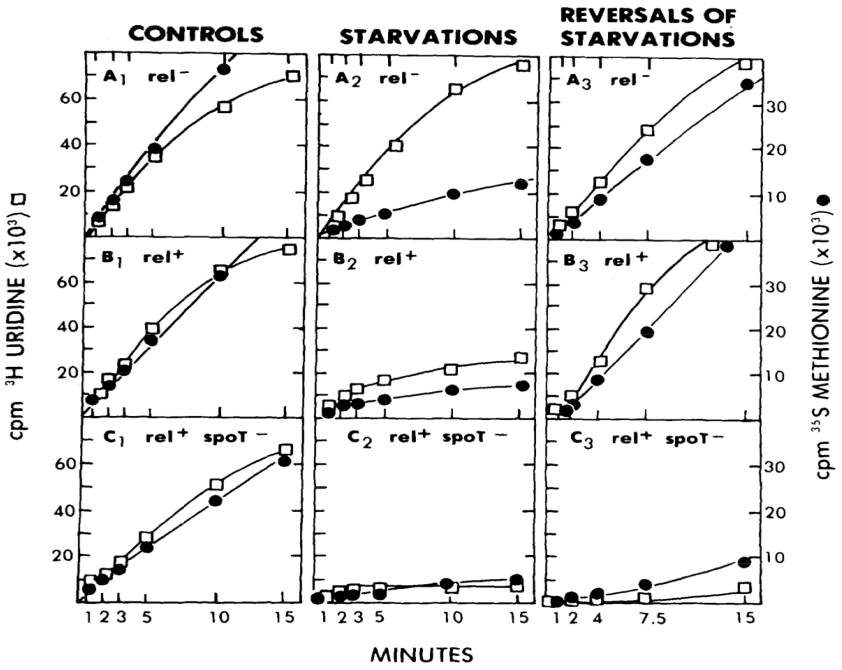


Figure 4.1: Effects of *RelA* and *SpoT* mutations and Amino Acid starvation on the rates of Methionine (•) and Uridine (□) incorporation in *E. Coli.* Adapted from O'Farrell [55].

The experiment represented in Figure 4.1 distinguishes the cell state and activity under three distinct conditions. The first condition represents normal cellular functioning, while the second condition simulates an amino acid starvation. Lastly, the third condition represents the state after addition of the deficient amino acid. The specific amino acid of interest in this experiment is isoleucine, as it is commonly utilized in protein formation. Throughout these conditions, the availability of methionine and uridine in the medium is adequate, ensuring that the two quantities measured in the experiment are not generally affected by their availability during the course of the experiment. However, their incorporation levels are measured.

The figure represented in this experiment illustrates the impact of amino acid withdrawal and subsequent addition on the strains *rel- spoT-*, *rel+ spoT+* and *rel+ spoT-*. Specifically, it is observed that only the *rel+* strains accumulate ppGpp in response to starvation, while the *spoT-* strains exhibit heightened stability of ppGpp, knowing that spoT is not degrading ppGpp.

This experiment provides valuable insights into the role of ppGpp in cellular activity. Firstly, in terms of the Transcription phase (RNA synthesis), uridine incorporation is measured as a proxy. It is evident that RNA synthesis is inhibited during amino acid starvation solely in *rel+* strains that accumulate ppGpp. Hence, ppGpp acts as an inhibitor of stable RNA synthesis in the absence of

amino acids.

Following the addition of the deficient amino acid, the rate of uridine incorporation swiftly recovers in *rel+ spoT+* strains, where ppGpp degradation occurs due to the presence of spoT. Conversely, in *rel+ spoT-* strains, the increase in uridine incorporation transpires at a slower pace since ppGpp remains undegraded, thereby sustaining the inhibition of transcription.

Regarding the impact of ppGpp on the Translation phase (protein synthesis), protein synthesis rate is measured by methionine incorporation, considering that methionine is the initial amino acid incorporated into the ribosome during protein synthesis. It is noteworthy that amino acid starvation inhibits protein synthesis in both *rel+* and *rel-* cells, which is logical since proteins are composed of amino acid sequences, leading to decreased production regardless of ppGpp presence. Upon the addition of the deficient amino acid, the rate of protein synthesis experiences a gradual increase in *rel+ spoT-* strains due to the accumulation of ppGpp without spoT-mediated degradation. This highlights that the accumulated levels of ppGpp play a substantial role in the substantial reduction of protein synthesis, resulting in an 86% inhibition.

The regulatory mechanism varies across different bacterial types with regards to how (p)ppGpp influences both transcription and translation phases of protein production. For instance, in *E. coli*, ppGpp binds to the polymerase, leading to the inhibition of ribosomal RNA transcription. Conversely, in *B. subtilis*, GTP governs the polymerases, and ppGpp, in turn, regulates GTP. This implies that in *B. subtilis*, ppGpp indirectly inhibits transcription by controlling GTP levels.

However, the main observation is that the inhibitory effect of ppGpp on translation remains consistent across all bacterial species. This effect is primarily manifested through its impact on IF2, the initiation factor of translation. Consequently, despite the differences in regulatory mechanisms, the variations observed do not undermine the applicability of our model in this chapter. As our study focuses on the influence of ppGpp on the translation phase, our model remains pertinent for diverse bacterial types. See al. [2] for further details.

4.3 Mathematical Model

The objective of this chapter is to introduce a mathematical model that captures the influence of ppGpp, a regulatory mechanism, on cellular activity. Due to the intricate nature of bacterial cells and the multitude of activities they achieve, selecting a specific cellular function to focus on presents a challenge. However, given the significance and costly nature of ribosomes within the cell, we have chosen to center our investigations on their activity. Consequently, we have developed a model that characterizes the different states of ribosomes within the cell under varying environmental conditions. By doing so, we aim to gain insights into the broader implications of ppGpp regulation on cellular processes.

4.3.1 Technical Challenges

We assume that there are N ribosomes with N typically large. We derive functional limiting results, with respect to this scaling parameter, of the time evolution of several stochastic processes. The primary challenges encountered involved two key aspects. Firstly, designing a model that effectively describes the intricate mechanism under investigation, encompassing protein production, with diverse ribosome states, various amino acids and tRNAs. This is crucial to obtain meaningful insights on the system. Additionally, another significant challenge involved formulating a mathematical characterization that differentiates between two cases: one where amino acid starvation is absent, and the other where amino acid starvation is present. The primary difficulty is the system's high dimensionality. Specifically, when dealing with two amino acids, the system had seven dimensions, and when dealing with only one amino acid, it still had five dimensions. This dimension added complexity to our analysis. Finally, an important step involved providing a biological interpretation of the obtained results. Addressing these challenges was crucial for gaining a comprehensive understanding of the system's dynamics and its implications.

4.3.2 Description

For simplicity, the number of ribosomes N is assumed to be constant.
In this model, we consider the diverse states of ribosomes. Two types of transfer RNAs (tRNAs) are considered. By accounting for the various states of ribosomes and distinguishing between the different types of tRNAs, our model offers a more comprehensive representation of the complex dynamics involved in protein synthesis. Employing a Markovian model, we were able to explore the interplay between ribosome states, tRNA availability, and their impact on the overall functioning of the cellular system.
From the point of view of our model, ribosomes can be in several states:

— *Free.* The ribosome may bind to a messenger RNA (mRNA), or be sequestered, or in the translation phase (initiation/elongation). $\big(R_N^F(t)\big)$ denotes the process of the number of free ribosomes.

— *Initiation.* The ribosome is bound to a mRNA, $\big(R_N^M(t)\big)$ is the number of such ribosomes.

— *Elongation.* During the elongation phase, the ribosome initiates protein synthesis by awaiting an amino acid of either type 1 $\big(R_{1,N}^E(t)\big)$ or type 2 $\big(R_{2,N}^E(t)\big)$, depending on whether the protein chain starts with an amino acid of type 1 or type 2, respectively.

— *Sequestered.* The associated process is $\big(R_N^S(t)\big)$.

With these notations, the conservation of mass for the ribosomes gives the relation

$$R_N^F(t) + R_N^M(t) + R_{1,N}^E(t) + R_{2,N}^E(t) + R_N^S(t) = N,\ \forall t \geq 0.$$

Similarly, we also represent the number of charged tRNAs of type 1 (respectively type 2) by $(Q_{1,N}(t))$ (respectively $(Q_{2,N}(t))$).
The dynamic of this stochastic system is governed by the analogue of the law of the mass action in this context.

4.3.3 Orders of Magnitude and Scaling Assumptions

We now discuss the orders of magnitude of the main parameters of the biological process.

— The scaling variable used in our study is N, the total number of ribosomes in the cell. It is assumed that this number is constant during the growth phase investigated. This number is quite large, between 10000 and 30000 for *E. Coli*, depending on the environment.

— The number C_m^N of different types of mRNAs is large, of the order of N, of the order of 40000.

— Similarly, the total number of each type of tRNA is also of the order of N, several thousands, with a mean value of 3200 for E. Coli (see Jakubowski H [39]).

We assume that the relations

$$\lim_{N\to+\infty} \frac{C_m^N}{N} = c_m{>}1, \quad \lim_{N\to+\infty} \frac{C_{q,1}^N}{N} = c_{q,1}{>}0 \ \textit{and} \ \lim_{N\to+\infty} \frac{C_{q,2}^N}{N} = c_{q,2}{>}0$$

hold.

And since we have these assumptions, we can deduce that there exists $a > 0$ such that

$$C_m^N - R_N^M(t) \geq aN \ \ \forall t \geq 0 \tag{4.1}$$

4.3.4 A sketch of the contributions: Four Limiting Regimes

In a probabilistic context, using a Markovian model, we investigate how the number of ribosomes in each state evolves asymptotically when N goes to infinity. Each regime corresponds to a set of conditions specifically related to the parameter β, which is the rate of production of amino acids, as will be detailed in the chapters to follow.

Our mathematical analysis considered the existence of four distinct regimes, each based upon specific parameter conditions. In our research, we put a condition on β_2, representing the rate at which tRNAs of type 2 are charged, across all four regimes. This condition serves to confirm that there is an ample supply of amino acids of type 2. In all these four regimes, we proved identical outcomes related to type 2 amino acids. Specifically that the charged tRNAs, referred to as Q_2, have reached a saturation point. As a result, our analysis will exclusively concentrate on amino acids of type 1.

For the moment we do not state the specific conditions and details of these three regimes here (see Section 4.6, 4.9, 4.8 and 4.7), We provide a brief overview of results obtained for each distinct regime.

Assuming that we have a constant total number of mRNAs $\left(c_m^N \approx c_m N\right)$, tRNAs of type 1 $\left(c_{q,1}^N \approx c_{q,1} N\right)$ and tRNAs of type 2 $\left(c_{q,2}^N \approx c_{q,2} N\right)$.

To assess the effectiveness of the regulatory mechanism, we focus on a crucial criterion: the rate of sequestration. To do this, we examine how the time evolution of the number of sequestered ribosomes, denoted as $R_N^S(t)$, depending on the cellular environment. Through this analysis, we showed the importance of certain parameters and evaluated their impact. Notably, the rate at which tRNAs are loaded with amino acids, represented as β, plays a crucial role in this context. The driving factor behind these situations is the speed at which tRNAs are loaded with amino acids, showing the role of the charging process rate, denoted as β.

Under specific conditions on the parameter β along with certain scaling conditions and appropriate initial conditions, we have the following classification. It should be noted that we have only partial results in this chapter. The scaling results associated to each of the four regimes described below depend on a technical result on some hitting time. Basically it states that, on a "small" time interval, the order of magnitude of the coordinates of the coordinates state variable do not change for some specific class of initial states. This is achieved completely in Section 4.6 for the first regime. The key result is Lemma 51 which relies on a not completely trivial coupling idea. For the other regimes, the corresponding lemma is stated but as a conjecture. These are Lemma 58 of Section 4.7, Lemma 68 of Section 4.8 and Conjecture 74 of Section 4.9.

a. *Adequate supply of amino acids.* The first regime corresponds to a normal cellular state, where there is an adequate supply of amino acids. In this regime, cellular processes operate without any constraints.
   For the convergence in distribution

$$\lim_{N\to\infty} \left(\frac{R_N^M(t)}{N}, \frac{Q_{1,N}(t)}{N}, \frac{Q_{2,N}(t)}{N} \right) = (1, c_{q,1}, c_{q,2})$$

   Under normal conditions within the cellular environment, the majority of ribosomes are engaged in the initiation phase, while only a small fraction of them remain free, either sequestered or awaiting amino acids.
   In this context, we observe the saturation of charged tRNAs of type 1. Crucially, all ribosomes are engaged in the initiation phase, signifying the absence of ribosomes sequestered or obstructed in the elongation phase. This configuration characterizes the cell's typical operational state.

b. *Regulation-independent deficiency of type* 1 *amino acids.* The second regime manifests when there is a lack of type 1 amino acids, but where the regulatory mechanism of sequestration is absent.

For the convergence in distribution

$$\lim_{N\to\infty}\left(\frac{R_N^M(t)}{N},\frac{R_{1,N}^E(t)}{N},\frac{Q_{2,N}(t)}{N}\right)=(r^M(t),1-r^M(t),c_{q,2})$$

where $\left(r^M(t)\right)$ is the solution of an ODE, such that

$$\lim_{t\to\infty} r^M(t)=r_\infty^M \text{ with } 0<r_\infty^M<1$$

and for any $t>0$, the random variables $\left(R_N^F(t)\right)$, $\left(R_{2,N}^E(t)\right)$ and $(Q_{1,N}(t))$ converge to Poisson distributions.

In this scenario, characterized by a deficiency of type 1 amino acids but the absence of the regulatory mechanism (referred to as sequestration in our model), most of the ribosomes exist in either the initiation or elongation phase, awaiting an amino acid of type 1. Only few ribosomes are free or awaiting an amino acid of type 2.
We deduce that a limited number of charged tRNAs of type 1 are available. Notably, the count of ribosomes obstructed in the elongation phase while awaiting type 1 amino acids is of the order of N. This observation shows the abnormal behavior of the cell.

c. *Partial Sequestration in the Deficiency of Type* 1 *Amino Acids.* The third regime occurs when there is a deficiency of amino acids of type 1, but the maximal sequestration rate (represented in Definition 48) is not attained due to the failure to satisfy certain conditions. Here, the cellular system experiences a shortage of specific amino acids, yet the sequestration mechanism does not reach its maximum capacity.
For the convergence in distribution

$$\lim_{N\to\infty}\left(\frac{R_N^M(t)}{N},\frac{R_{1,N}^E(t)}{N},\frac{R_N^S(t)}{N},\frac{Q_{2,N}(t)}{N}\right)=(r^M(t),r_1^E(t),1-r^M(t)-r_1^E(t),c_{q,2})$$

where $\left(r^M(t)\right)$ and $\left(r_1^E(t)\right)$ are the solutions of two ODEs, such that

$$\begin{aligned}\lim_{t\to\infty} r^M(t) &= r_\infty^M \text{ with } 0<r_\infty^M<1\\ \lim_{t\to\infty} r_1^E(t) &= r_{1,\infty}^E \text{ with } 0<r_{1,\infty}^E<1-r_\infty^M\end{aligned}$$

and for any $t>0$, the random variables $\left(R_N^F(t)\right)$, $\left(R_{2,N}^E(t)\right)$ and $(Q_{1,N}(t))$ converge to Poisson distributions.

In this case, where there is a deficiency of type 1 amino acids and the regulatory mechanism partially interferes but does not reach its maximum capacity, we observe similar outcomes as in the previous case. However, the subsequent sections 4.9 and 4.8 will provide a detailed analysis of the distinctions between these two situations.

The model incorporating sequestration during amino acid deficiency, although not fully activated at its maximum rate, Sections 4.8 and 4.9 reveal the persistence of a small count of charged tRNAs of type 1. Additionally, despite the ribosomes blocked in the elongation phase remaining of the order of N level, the sequestered ribosomes now also are of the order of N. This leads to a reduction in the quantity of blocked ribosomes compared to the sequestration-free model. Consequently, this result outlines the influence of sequestration on ribosome distribution.

d. *Maximal Sequestration in the Deficiency of Type* 1 *Amino Acids.* The fourth regime arises when there is a scarcity of amino acids of type 1, and the sequestration mechanism reaches its maximal rate. In this case, the cellular system faces a pronounced lack of amino acids, and the sequestration process is fully activated.
For the convergence in distribution

$$\lim_{N\to\infty}\left(\frac{R_N^M(t)}{N},\frac{R_N^S(t)}{N},\frac{Q_{1,N}(t)}{N},\frac{Q_{2,N}(t)}{N}\right)=(r^M(t),1-r^M(t),q_1(t),c_{q,2})$$

where $\left(r^M(t)\right)$ and $(q_1(t))$ are the solutions of two ODEs, such that

$$\begin{aligned}\lim_{t\to\infty} r^M(t) &= r_\infty^M \ with\ 0<r_\infty^M<1\\ \lim_{t\to\infty} q_1(t) &= q_{1,\infty}\ with\ 0<q_{1,\infty}<c_{q,1}\end{aligned}$$

and for any $t>0$, the random variables $\left(R_N^F(t)\right)$, $\left(R_{1,N}^E(t)\right)$ and $\left(R_{2,N}^E(t)\right)$ converge to Poisson distributions.

Under these conditions of maximal sequestration, the majority of ribosomes are either engaged in the initiation phase or sequestered, while only a small fraction remains unoccupied or awaiting amino acids.

In this scenario where maximal sequestration is implemented, we have two notable outcomes. Firstly, the presence of blocked ribosomes reduces entirely, as these previously blocked ribosomes are now sequestered. Furthermore, a substantial number of charged tRNAs of type 1, roughly of the order of N, is detected. This indicates that maximal sequestration not only reduces the number of blocked ribosomes but also increases the number of charged amino acids, gradually returning the environmental conditions back to normal.

Based on this classification, we have noticed that the primary objective of sequestration is to decrease the number of blocked ribosomes in the elongation phase. It follows that ribosomes, when blocked while awaiting an amino acid, may ultimately select the wrong one, leading to errors in protein production (see O'Farrell [55] for reference). Consequently, the sequestration has the effect of reducing the rate of errors in protein production.

4.4 Stochastic Model

In this section we introduce the state space description of the regulation of translation. We first describe our main assumptions in the design of the stochastic model.

4.4.1 Modeling Assumptions

The chemical species involved in the regulation process are the ribosomes in different states and the tRNAs.

Introducing the concept of ribosome dynamics, our model operates under the assumption of a constant total number of ribosomes, denoted as N.

We will start by describing our model.

— INITIATION OF TRANSLATION.
The initiation step represents the binding of the ribosome to an mRNA. It is assumed that there are C_m^N different types of mRNAs and that at a given time, there is at most one ribosome is bound to a mRNA. When a mRNA is free, a free ribosome may bind to this mRNA at rate $\delta\left(C_m^N - R^M\right)$, since $\left(C_m^N - R^M\right)$ represents the free places (mRNAs) for initiation. In the case where an mRNA is already bound to a ribosome, the ribosome initiates the elongation step at a rate of λ_1 if the protein chain it is constructing begins with an amino acid of type 1. Similarly, if the protein chain starts with an amino acid of type 2, the ribosome initiates elongation at a rate of λ_2.

— ELONGATION OF TRANSLATION.
Once the ribosome enters the elongation phase of a protein that begins with an amino acid of type $i \in \{1, 2\}$, it transitions into a waiting state for an amino acid of the same type i. Once the amino acid is being carried by a tRNA, three possible scenarios can occur. Firstly, the ribosome may continue to wait for another amino acid of type i, which transpires at a rate of ν_i. Alternatively, it may switch its waiting state to anticipate an amino acid of the other type, transpiring at a rate of ψ_i. Lastly, if the protein is fully synthesized, the ribosome detaches from the mRNA, reverting to a free state once again. This detachment occurs at a rate of α_i.
Concerning the amino acids part in this model: In order to simplify the model, we consider the formation of a protein using two types of amino acids, labeled as 1 and 2. However, it's important to note that the outcomes obtained can be generalized to a greater number, such as the typical 20 amino acids.

— CHARGING/DISCHARGING OF tRNAs BY AMINO ACIDS.
In our model, which considers proteins composed of two distinct types

of amino acids, we have introduced two corresponding types of tRNAs. We assume that the total number of tRNAs for each type, denoted as $C^N_{q,i} = c_{q,i}N$, is determined. The charging and discharging process of tRNAs with amino acids directly involves ribosomes during the elongation phase.
An empty tRNA of type $i \in \{1, 2\}$ transports an amino acid of type i at rate $\beta_i N$. It becomes discharged (empty) when it transfers the amino acid it carries to a ribosome awaiting it in the elongation phase. Therefore, this discharging event occurs at a rate of $\alpha_i+\nu_i+\psi_i$.

— SEQUESTRATION/DESEQUESTRATION OF RIBOSOMES.
In our model the sequestration part represents the regulatory mechanism that occurs during amino acid starvation. Hence, it is crucial for the rate of sequestration to be dependent on the number of empty tRNAs, denoted as $C^N_{q,i}-Q_i$ for $i \in \{1, 2\}$. The sequestration occurs at rate γ_i when there is a deficiency of amino acid of type i, and a ribosome is released and becomes free again at rate η.

4.4.2 The Markov Process and $\mathcal{Q}$-Matrix

In our model, the rate of amino acid production β_i of type $i \in \{1, 2\}$ emerges as the pivotal parameter. This significance is attributed to our investigation of the regulatory mechanism during amino acid starvation. The differences among the regimes represented in section 4.3.4 will be characterized by conditions on the parameter β_i. We now give a Markovian description of our system.

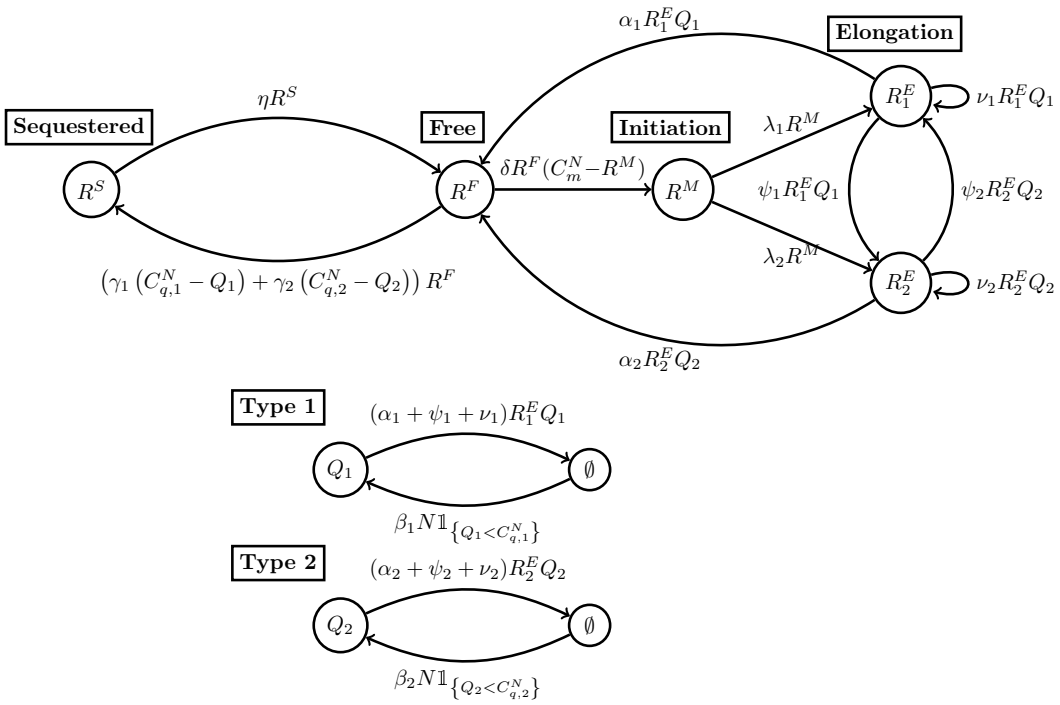


Figure 4.2: Ribosomes: Translation and Sequestration

State Space. All transitions described in the last section occur after a random amount of time with an exponential distribution. With this assumption, there is a natural Markov process to investigate the regulation of translation. The state space is given by

$$\mathcal{S}_N \stackrel{def}{=} \left\{ x = (r^F, r^M, r_1^E, r^S, q_1, q_2) \in \mathbb{N}^4 \times \prod_{j=1}^{2} \left\{0, \ldots, C_{q,j}^N\right\} : r^F + r^M + r_1^E + r^S \leq N \right\}$$

If the state of the system is $x = (r^F, r^M, r_1^E, r^S, q_1, q_2) \in \mathcal{S}_N$, then

— r^F is the number of free ribosomes;

— r^M, the number of ribosomes in the initiation step of the translation phase;

— r_1^E, the number of ribosomes in the elongation step waiting for an amino acid of type 1;

— r^S, the number of sequestered ribosomes;

— q_1, the number of tRNAs carrying an amino acid of type 1;

— q_2, the number of tRNAs carrying an amino acid of type 2;

— In state x, the number of ribosomes in the elongation step waiting an amino acid of type 2 is given by

$$\Psi(x) \stackrel{def}{=} N - r^F - r^M - r_1^E - r^S$$

The associated Markov process is denoted by

$$X_N(t) \stackrel{def}{=} \left(R_N^F(t), R_N^M(t), R_{1,N}^E(t), R_N^S(t), Q_{1,N}(t), Q_{2,N}(t)\right)$$

The number of ribosomes at time t in elongation phase waiting an amino acid of type 1 is defined by $R_{2,N}^E(t) = \Psi(X_N(t))$.

It is easily checked that $(X_N(t))$ is an irreducible Markov process on $\mathcal{S}_N$. Its transition rates are given by

— INITIATION OF TRANSLATION.

$$(r^F, r^M, r_1^E, r^S, q_1, q_2) \longrightarrow \begin{cases} (r^F-1, r^M+1, r_1^E, r^S, q_1, q_2) & \delta r^F (C_m^N - r^M), \\ (r^F, r^M-1, r_1^E+1, r^S, q_1, q_2) & \lambda_1 r^M, \\ (r^F, r^M-1, r_1^E, r^S, q_1, q_2) & \lambda_2 r^M. \end{cases}$$

— ELONGATION OF TRANSLATION.

$$(r^F, r^M, r_1^E, r^S, q_1, q_2) \longrightarrow \begin{cases} (r^F, r^M, r_1^E, r^S, q_1-1, q_2) & \nu_1 r_1^E q_1, \\ (r^F, r^M, r_1^E-1, r^S, q_1-1, q_2) & \psi_1 r_1^E q_1, \\ (r^F+1, r^M, r_1^E-1, r^S, q_1-1, q_2) & \alpha_1 r_1^E q_1, \\ (r^F, r^M, r_1^E, r^S, q_1, q_2-1) & \nu_2 r_2^E q_2, \\ (r^F, r^M, r_1^E+1, r^S, q_1, q_2-1) & \psi_2 r_2^E q_2, \\ (r^F+1, r^M, r_1^E, r^S, q_1, q_2-1) & \alpha_2 r_2^E q_2. \end{cases}$$

— CHARGING AND DISCHARGING OF tRNAs BY AMINO ACIDS.

$$(r^F, r^M, r_1^E, r^S, q_1, q_2) \longrightarrow \begin{cases} (r^F, r^M, r_1^E, r^S, q_1+1, q_2) & \beta_1 N \mathbb{1}_{\{q_1 < C_{q,1}^N\}}, \\ (r^F, r^M, r_1^E-1, r^S, q_1, q_2+1) & \beta_2 N \mathbb{1}_{\{q_2 < C_{q,2}^N\}}. \end{cases}$$

The discharging of tRNAs is represented in elongation of translation part.

— SEQUESTRATION/DESEQUESTRATION OF RIBOSOMES.

$$(r^F, r^M, r_1^E, r^S, q_1, q_2) \longrightarrow \begin{cases} (r^F-1, r^M, r_1^E, r^S+1, q_1, q_2) & \left(\gamma_1 \left(C_{q,1}^N - q_1\right) + \gamma_2 (C_{q,2}^N - q_2)\right) r^F, \\ (r^F+1, r^M, r_1^E, r^S-1, q_1, q_2) & \eta r^S. \end{cases}$$

Just to clarify, the transitions in the figure 4.4.2 represent simultaneous changes happening in both the ribosome states and the charging and discharging of tRNAs. These dynamics are represented in the two separate graphs within the figure.

4.5 Models with 1 amino acid

To simplify things, our model is devised to illustrate the translation of a protein composed of just a single amino acid. However, within the chapter, we also present a model with two different amino acids. In fact, the existence of four different regimes, listed in Section 1.7.4.3, does not depend on the number of amino acids in the model. Adding amino acids in our model, surely will give a more realistic description from a biological point of view but does not seem to have an impact from a mathematical point of view. For this reason the chapter is essentially devoted to the study of a model with just one amino acid.

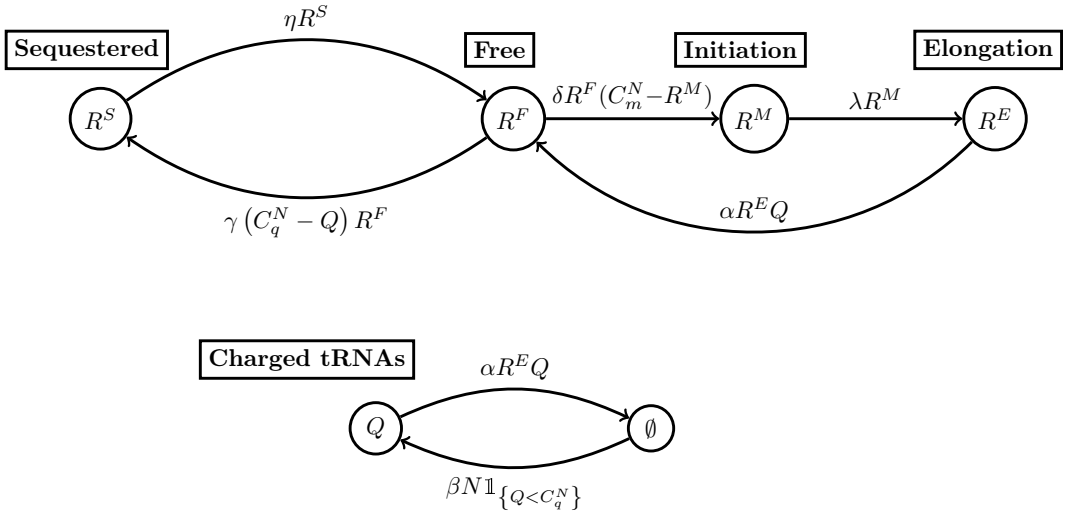


Figure 4.3: Ribosomes: Translation and Sequestration with 1 amino acid

Definition 48. — *We introduce $\mathcal{S}$ as*

$$\mathcal{S} \stackrel{def}{=} \frac{\eta\delta}{\beta\gamma}\left(c_m - \frac{\beta}{\lambda}\right)\left(1 - \frac{\beta}{\lambda}\right) \tag{4.2}$$

$\mathcal{S}$ interpreted as the maximal sequestration rate in this model with 1 amino acid.

We can now introduce the four regimes of interest in our study.

Definition 49. *a. The regime of Adequate supply of amino acids is defined by the relation*

$$\beta > \lambda \tag{4.3}$$

The production rate β of amino acids is greater than its consumption rate.

b. The regime of Regulation independent deficiency of amino acids is defined by the relation

$$\beta < \lambda \tag{4.4}$$

The production rate β of amino acids is less than its consumption rate.

c. The regime of Partial Sequestration in the deficiency of amino acids is defined by the relations

$$\beta < \lambda \ , \ c_q < \mathcal{S} \tag{4.5}$$

Same conditions 4.4 with the fraction of tRNAs of the lacking amino acid less than the maximal sequestration rate.

d. The regime of Maximal Sequestration in the deficiency of amino acids is defined by the relations

$$\beta < \lambda \ , \ c_q > \mathcal{S} \tag{4.6}$$

Same conditions 4.4 with the fraction of tRNAs of the lacking amino acid greater than the maximal sequestration rate.

4.6 Adequate supply of amino acids

In this section, we focus on the scenario where there is an adequate supply of amino acids and all conditions are considered normal. Thus, our study centers on the distribution of ribosomes under these standard conditions. The objective is to establish a basis for comparison with the other sections, where we examined the effects of amino acid deficiencies in various conditions. Throughout this section, we will be operating under condition (4.3).

We will present a new variable, denoted as $V_N(t)$, which signifies the count of unoccupied tRNAs. It is calculated as the difference between the total number of tRNAs, denoted as C_q^N, and the number of tRNAs engaged in the translation process, represented as $Q_N(t)$. Thus, the formula for $V_N(t)$ is:

$$V_N(t) = C_q^N - Q_N(t)$$

Definition 50. *If g is non-negative Borelian function on $\mathbb{R}_+ \times \mathbb{N}^2$, we define the occupation measure*

$$\langle \mu_N, g \rangle \stackrel{def}{=} \int_{\mathbb{R}_+} g\left(s, R_N^E(s), R_N^F(s)\right) ds$$

In this section, we maintain identical initial conditions as presented in Section 4.7. Additionally, we use the same stopping time $\tau_Q^N(c)$ and occupation measure as in definition 57. Consequently, the couplings employed here remain consistent with relation (4.17).

Lemma 51. *Under conditions* (4.3) *and* (4.15)*, there exist* $t_0^c > 0$ *such that*

$$\underset{N\to+\infty}{\mathbb{P}}\left(\tau_Q^N(c) < t_0^c\right) = 0$$

Proof. We provide a detailed step-by-step explanation of the demonstration. The Markov process considered here is $\left(R_N^E(t), Q_N(t)\right)$, for $x = (r^E, q)$,

$$x \longrightarrow x \begin{cases} e_1 & \lambda r^M, \\ -e_1 - e_2 & \alpha r^E q, \\ e_2 & \beta N. \end{cases}$$

with initial state:

$$R_N^E(0) = r_0^E \in \mathbb{N}\, and \lim_{N\to+\infty} \frac{Q_N(0)}{N} = q_0 > 0$$

For $0 < \eta_1 < q_0$, we define the stopping time

$$\tau_N = \inf\left\{t > 0 : \frac{Q_N(t)}{N} \geq \eta_2\right\}$$

We introduce the auxiliary Markov process $\left(\tilde{R}_N(t), \tilde{Q}_N(t)\right)$ with $\mathcal{Q}-$matrix

$$x \longrightarrow x \begin{cases} +e_1 & \lambda, \\ -e_1 - e_2 & \alpha\eta_1\tilde{r}, \\ -e_2 & \alpha(\eta_2 - \eta_1)\tilde{r}, \\ +e_2 & \beta. \end{cases}$$

with the same initial state (r_0^E, q_0) and

$$\tilde{\tau}_N = \inf\left\{t > 0 : \frac{\tilde{Q}_N(Nt)}{N} \leq \eta_1\right\}$$

Since $\beta > \lambda$, $\left(\tilde{Q}_N(t)\right)$ represents a transient $M/M/1$ queue, thus we have:
If η_1, η_2 are chosen such that $\eta_1 < q_0 < \eta_2$ with $\eta_2/\eta_1 \leq \beta/\lambda$, then, for the convergence in distribution

$$\lim_{N\to+\infty}\left(\frac{\tilde{Q}_N(Nt)}{N}, \int_0^t \tilde{R}_N(Ns)ds\right) = \left(q_0 + \left(\beta - \frac{\eta_2}{\eta_1}\lambda\right)t, \frac{\lambda}{\alpha\eta_1}t\right) \tag{4.7}$$

and for any $T \geq 0$

$$\lim_{N\to+\infty} \mathbb{P}\left(\tilde{\tau}_N \leq T\right) = 0$$

Therefore, we can prove the couplings

$$R_N^E(t) \leq \tilde{R}_N(Nt) \text{ and } \tilde{Q}_N(Nt) \leq Q_N(t) \tag{4.8}$$

for all $t \leq T$ on the event $\{\tilde{\tau}_N \wedge \tau_N > T\}$.

From all this, we prove the convergence in distribution

$$\lim_{N \to +\infty} \left(\frac{Q_N(t)}{N} \right) = (q_0 + (\beta - \lambda)t)$$

In fact, firstly, since $(Q_N(t))$ increases at rate at most βN, we can find $t_0 < (\eta_2 - q_0)/\beta$ such that

$$\lim_{N \to +\infty} \mathbb{P}(\tau_N \leq t_0) = 0$$

Now we define the occupation measure

$$\langle \mu_N, f \rangle = \int_0^{t_0} f\left(s, R_N^E(s)\right) ds$$

(μ_N) is tight for the convergence in distribution.
For $K > 0$, with Relation (4.8), we get

$$\begin{aligned}\int_0^{t_0} \mathbb{P}\left(R_N^E(s) \geq K\right) ds &\leq t_0 \left(\mathbb{P}(\tau_N \leq t_0) + \mathbb{P}(\tilde{\tau}_N \leq t_0)\right) + \int_0^{t_0} \mathbb{P}\left(R_N^E(s) \geq K, \tau_N \wedge \tilde{\tau}_N \geq t_0\right) ds \\ &\leq t_0 \left(\mathbb{P}(\tau_N \leq t_0) + \mathbb{P}(\tilde{\tau}_N \leq t_0)\right) + \int_0^{t_0} \mathbb{P}\left(\tilde{R}(Ns) \geq K\right) ds\end{aligned}$$

We then prove the tightness of the sequence $(Q_N(t)/N, t \leq t_0)$, actually,

$$\overline{Q}_N(t) \stackrel{def}{=} \frac{Q_N(t)}{N} = \frac{Q_N(0)}{N} + \frac{M_{Q,N}(t)}{N} + \beta t - \alpha \int_0^t R_N^E(s) \frac{Q_N(s)}{N} ds$$

$$w_N(\delta) = \sup_{\substack{0 \leq s \leq t \\ |t-s| \leq \delta}} \left|\overline{Q}_N(t) - \overline{Q}_N(s)\right|$$

For $\eta > 0$, we take δ such that $\beta\delta < \eta/3$, then the couplings in Relation (4.8), give the relation

$$\begin{aligned}&\mathbb{P}(w_N(\delta) \geq \eta) \leq \mathbb{P}(\tau_N \leq t_0) + \mathbb{P}(\tilde{\tau}_N \leq t_0) \\ &+ \mathbb{P}\left(\sup_{s \leq t_0} |M_N(s)| \geq \frac{\eta}{6}, \tau_N \wedge \tilde{\tau}_N \geq t_0\right) + \mathbb{P}\left(\sup_{\substack{0 \leq s \leq t \\ |t-s| \leq \delta}} \alpha\eta_2 \int_s^t \tilde{R}(Nu) du \geq \frac{\eta}{3}\right)\end{aligned}$$

Relation (4.7) and the convergence of the martingale $(M_N(t \wedge \tau_N \wedge \tilde{\tau}_N)/N)$, to 0 give the required tightness property.

From there, by looking at the choice of η_1 and η_2, we prove the desired convergence in distribution on the time interval $[0, 1/\lambda)$. One can proceed by induction to establish the convergence on $\mathbb{R}_+$. □

4.6.1 Tightness Properties

Proposition 52. *Under conditions* (4.3) *and* (4.15)*, the sequence of processes*

$$\left(\overline{R}_N^{M,c}(t)\right),\ \left(\overline{R}_N^{S,c}(t)\right) \text{ and } \left(\overline{V}_N^{c}(t)\right)$$

are tight for the convergence in distribution.

Proof. Same method as in the proof of proposition 59. □

We denote their limits, resp.

$$\left(r^M(t)\right),\ \left(r^S(t)\right) \text{ and } \left(v(t)\right)$$

Proposition 53. *Under conditions* (4.3) *and* (4.15)*, and if* μ_∞^c *is a limiting point of* μ_N^c*, then for any* $t < t_0^c$ *and any continuous function* g *on* $\mathbb{N}^2$*, we have*

$$\begin{aligned}\lim_{N\to+\infty} &\left(\int_0^t g\left(R_N^E(s), R_N^F(s)\right) ds\right) \\ &= \int_0^t \mathbb{E}\left(g\left(\mathcal{N}_1\left(0, \frac{\lambda r^M(s)}{\alpha(c_q - v(s))}\right), \mathcal{N}_2\left(0, \frac{(\lambda-\eta)r^M(s)) + \eta}{\gamma v(s) + \delta(c_m - r^M(s))}\right)\right) ds\right)\end{aligned} \tag{4.9}$$

where $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *are two Poisson processes on* $\mathbb{R}_+$ *with rate* 1*.*

Proof. Same method as in the proof of proposition 63. □

4.6.2 Saturation of Q

The objective of this section is to prove that under the conditions $\beta > \lambda$, the charged tRNAs are saturated. We will prove that there exists $T > 0$ such that for $t \geq T$,

$$\left(\frac{Q_N(t)}{N}\right) \xrightarrow{\mathcal{D}} (c_q) \tag{4.10}$$

Since our study is done on $[0, \tau_Q^N(c))$, there exists $\delta_0 > 0$ such that

$$V_N(t) \stackrel{def}{=} C_q^N - Q_N(t) < \delta_0 N$$

where $\delta_0 = c_q - c$.
We have,

$$= F_N(t) + R_N(t)$$

where

$$F_N(t) = \frac{V_N^c(0)}{N} + \frac{M_N(t)}{N} + \alpha \int_0^t \frac{C_q^N - V_N^c(s)}{N} R_N^{E,c}(s) ds - \beta t$$

$$R_N(t) = \beta \int_0^t \mathbb{1}_{\left\{V_N^c(s)=0\right\}} ds$$

Referring to the convergence in distribution in Proposition 53, we have

$$(F_N(t)) \xrightarrow{\mathcal{D}} (f(t))$$

where

$$f(t) = v_0 + \lambda \int_0^t r^M(s)ds - \beta t$$

and $v_0 = \lim_{N\to\infty} \frac{\tilde{V}_N(0)}{N}$.

Since $r^M(s) < 1 \ \forall s$ and $\beta > \lambda$, we get

$$f(t) \leq v_0 + (\lambda - \beta)t$$

The couple $\left(\frac{V_N^c(t)}{N}, R_N(t)\right)$ is the solution of the Skorokhod problem associated to the free process $(F_N(t))$.

We have, the convergence of the process $\left(\frac{V_N^c(t)}{N}, R_N(t)\right)$ to $(f(t)^+, f(t)^-)$ where for $a \in \mathbb{R}$, $a^+ = \max(a, 0)$ and $a^- = \max(-a, 0)$.

Thus,

$$\lim_{N\to\infty} \frac{\tilde{V}_N(t)}{N} \leq (v_0 + (\lambda - \beta)t)^+$$

and since $\beta > \lambda$, there exists T such that for all $t \geq T$, $(v_0 + (\lambda - \beta)t)^+ = 0$.

Finally, Q is saturated.

As a consequence, we get the corollary.

Corollary 54. *There exists $T>0$ such that for all $t\geq T$, the convergence of distribution below holds,*

$$\begin{aligned}\lim_{N\to+\infty} &\left(\int_0^t g\left(R_N^E(s), R_N^F(s)\right) ds\right) \\ &= \int_0^t \mathbb{E}\left(g\left(\mathcal{N}_1\left(0, \frac{\lambda r^M(s)}{\alpha c_q}\right), \mathcal{N}_2\left(0, \frac{(\lambda-\eta)r^M(s)) + \eta}{\delta(c_m - r^M(s))}\right)\right) ds\right)\end{aligned} \tag{4.11}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ are two Poisson processes on $\mathbb{R}_+$ with rate 1.

Theorem 55 (Law of Large Numbers)**.** *Under conditions* (4.3)*, and* (4.15)*, this convergence in distribution holds.*

$$\lim_{N\to+\infty} \left(\frac{R_N^M(t)}{N}\right) = \left(r^M(t)\right)$$

where $\left(r^M(t)\right)$) is the solution of the ODE

$$\frac{dr^M}{dt}(t) = \eta - \eta r^M(t) \tag{4.12}$$

Thus, we get the equilibrium point for $\left(r^M(t)\right)$,

$$r_\infty^M = 1$$

Proof. Same method as in the proof of theorem 65. □

At equilibrium, all ribosomes are observed to be in the initiation phase, signifying that under regular conditions, they are actively involved in the translation process. There is no sequestration or blocking of ribosomes in the elongation phase. This state represents the normal activity of the cell without any constraints. In other words, in the absence of external limitations, the cell's ribosomes are fully occupied and productively engaged in translation, reflecting the standard functioning state of the cell.

4.7 Maximal Sequestration in the Deficiency of Amino Acids

Throughout this section, conditions (4.6) hold. In the context of amino acid starvation, a protein known as RelA is expected to initiate a stringent response by producing (p)ppGpp, which leads to the inhibition of translation. This process is implicitly represented in our model through the sequestration component, resulting in a substantial number of sequestered ribosomes. However, we are interested in understanding the effects of maximal sequestration on the quantity of deficient amino acids, as represented in our model by the charged tRNAs, and the number of ribosomes in the elongation phase awaiting these deficient amino acids.

The objective of this section is twofold: first, to analyze the sequestration process by focusing on the time evolution of the number of sequestered ribosomes $\left(R_N^S(t)\right)$, and second, to address the question regarding the evolution of the number of deficient amino acids under conditions of maximal sequestration by investigating the time evolution of the process $(Q_N(t))$.

Definition 56. *The scaled process is defined by*

$$\left(\overline{X}_N(t)\right) = \left(R_N^F(t), \overline{R}_M^N(t), \overline{R}_N^S(t), \overline{Q}_N(t)\right), \tag{4.13}$$

with

$$\overline{R}_N^M(t) = \frac{R_N^M(t)}{N},\ \overline{R}_N^S(t) = \frac{R_N^S(t)}{N},\ \overline{Q}_N(t) = \frac{Q_N(t)}{N}.$$

If g is non-negative Borelian function on $\mathbb{R}_+ \times \mathbb{N}^2$, we define the occupation measure

$$\langle \mu_N, g \rangle \stackrel{def}{=} \int_{\mathbb{R}_+} g(s, R_N^E(s), R_N^F(s)) ds. \tag{4.14}$$

The following initial conditions will be assumed,

$$\begin{aligned} R_N^E(0) = r_0^E \in \mathbb{N}\,,\ R_N^F(0) = r_0^F \in \mathbb{N}, \\ \lim_{N\to\infty} \frac{R_N^M(0)}{N} = r_0^M\ ,\ \lim_{N\to\infty} \frac{Q_N(0)}{N} = q_0, \end{aligned} \tag{4.15}$$

with $r_0^M < 1$ and $q_0 \in (0, c_q)$.

And the number of sequestered ribosomes $R_N^S(0)$ is therefore such

$$\lim_{N\to\infty} \frac{R_N^S(0)}{N} = 1 - r_0^M.$$

Before proving the convergence of the sequence of processes $\left(\overline{X}_N(t)\right)$, we analyze the convergence of a "stopped" version of it. In several technical arguments we will first need to have that the fraction of charged tRNAs is not too small. A second step is of showing that, essentially, the stopped process does not differ from the original process.

Definition 57. *For $c > 0$, the stopping time $\tau_Q^N(c)$ is defined by*

$$\tau_Q^N(c) \stackrel{def}{=} \inf\{t > 0; Q_N(t) \leq cN\} \tag{4.16}$$

and

a. If $(W(t))$ is a càdlàg process, we denote $(W_N^c(t)) = \left(W\left(N\left(t \wedge \tau_Q^N(c)\right)\right)\right)$;

b. The "stopped" occupation measure μ_N^c is defined by, if g is non-negative Borelian function on $\mathbb{R}_+ \times \mathbb{N}^2$,

$$\langle \mu_N^c, g\rangle \stackrel{def}{=} \int_0^{\tau_Q^N(c)} g(s, R_N^{E,c}(s), R_N^{F,c}(s))ds$$

With a slight abuse, the notation $\left(\overline{W}_N^c(t)\right) = \left(\frac{W_N^c(t)}{N}\right)$ will be used in the following.

4.7.1 Technical Lemmas

The two processes $\left(R_N^E(t)\right)$ and $\left(R_N^F(t)\right)$ are in fact in a neighborhood of 0 quickly. They will be the fast processes of our averaging principle. In state $R_N^E = r^E$, $R_N^F = r^F$, $R_N^M = r^M$ and $Q_N = q$, the jump rates of the processes $\left(R_N^E(t)\right)$ and $\left(R_N^F(t)\right)$ are respectively

$$\begin{cases} +1 & \lambda r^M, \\ -1 & \alpha r^E q. \end{cases}$$

$$\begin{cases} +1 & \eta(N - r^F - r^M - r^E) + \alpha r^E q \\ -1 & \left(\delta\left(C_m^N - r^M\right) + \gamma\left(C_q^N - q\right)\right) r^F. \end{cases}$$

For N sufficiently large and up to time $\tau_Q^N(c)$, a simple coupling shows that there exist independent processes $(Y(t))$ and $(Z(t))$ such that

$$R_N^E(t) \leq Y(Nt) \text{ and } R_N^F(t) \leq Z(Nt) \tag{4.17}$$

holds.

— The process $(Y(t))$ is a birth and death processes on $\mathbb{N}$ whose $\mathcal{Q}$-matrix is given by

$$q(y, y+1) = \lambda \ \textit{and} \ q(y, y-1) = \alpha c$$

— The process $(Z(t))$ is a birth and death processes on $\mathbb{N}$whose $\mathcal{Q}$-matrix is given by

$$q(z, z+1) = \kappa_0 \ \textit{and} \ q(z, z-1) = \delta a z \ \textit{with} \ \kappa_0 > 0$$

It is not difficult to prove that the hitting time of $(0,0)$ by $(Y(t), Z(t))$ is of the order of $\ln(N)$ so that conditions (4.15) can be assumed. See Chapter 6 of Robert [61].

Conjecture 58. *Under Conditions* (4.6)*, and* (4.15)*, there exist* $t_0^c > 0$ *such that*

$$\lim_{N \to +\infty} \mathbb{P}\left(\tau_Q^N(c) < t_0^c\right) = 0$$

We will adopt this assumption as it is essential for proving the tightness properties of the processes as it will be seen in the next section.

4.7.2 Tightness Properties

Proposition 59. *Under Conditions* (4.6)*, and* (4.15) *and Conjecture 58, the sequences of processes*

$$\left(\overline{R}_N^{M,c}(t)\right), \ \left(\overline{R}_N^{S,c}(t)\right), \ \textit{and} \ \left(\overline{Q}_N^c(t)\right)$$

are tight for the convergence in distribution.

Proof. (Modulus of continuity.) We will establish the tightness specifically for the process $\left(\overline{R}_N^{M,c}(t)\right)$, and the identical method will be employed to demonstrate the tightness of the other processes.

$$\frac{R_N^{M,c}(t)}{N} = \frac{R_N^M(0)}{N} + \frac{M_N(t)}{N} + \delta \int_0^t R_N^{F,c}(s) \frac{C_m^N - R_N^{M,c}(s)}{N} ds - \lambda \int_0^t \frac{R_N^{M,c}(s)}{N} ds$$

The increasing process is given by:

$$\left\langle \frac{M_N(t)}{N} \right\rangle = \delta \int_0^t R_N^{F,c}(s) \frac{C_m^N - R_N^{M,c}(s)}{N^2} ds + \lambda \int_0^t \frac{R_N^{M,c}(s)}{N^2} ds$$

By using the coupling (4.17), and since $(\mathbb{E}(Z(u)))$ converges as u goes to infinity, we have

$$\lim_{N \to \infty} \mathbb{E}\left(\left\langle \frac{M_N(t)}{N} \right\rangle\right) = 0$$

Doob's Inequality shows then the convergence of $\left(\frac{M_N(t)}{N}\right)$ to 0. Thus we have this inequality of the modulus of continuity:

$$w_N(\delta_0) \stackrel{def}{=} \sup_{\substack{0\le u,v\le t\\ |u-v|\le\delta_0}} \left|\frac{R_N^{M,c}(u)-R_N^{M,c}(v)}{N}\right| \le \sup_{\substack{0\le u,v\le t\\ |u-v|\le\delta_0}} \left|\frac{M_N(u)-M_N(v)}{N}\right| + \delta \sup_{\substack{0\le u,v\le t\\ |u-v|\le\delta_0}} \left|\int_u^v R_N^{F,c}(s)\frac{C_m^N - R_N^{M,c}(s)}{N}ds\right| + \lambda \sup_{\substack{0\le u,v\le t\\ |u-v|\le\delta_0}} \left|\int_u^v \frac{R_N^{M,c}(s)}{N}ds\right|$$

By using the couplings (4.17) and the convergence in distribution of the martingale, we get for all $\eta>0, \varepsilon>0$, there exists δ_0 such that $\mathbb{P}(w_N(\delta_0)\ge\eta)\le\varepsilon$ for N sufficiently large. As a result, we observe the convergence of sub-sequences, and due to the uniqueness of the limit, this proposition is proved. □

We denote their limits, resp.

$$\left(r^M(t)\right),\ \left(r^S(t)\right)=\left(1-r^M(t)\right)\ \textit{and}\ \ (q(t))$$

Proposition 60. *Under the conditions of Proposition 59, the sequence of measure-valued processes (μ_N^c) on the state space $[0,t_0^c)\times\mathbb{N}^2$ is tight for the convergence in distribution and any limiting point μ_∞^c can be expressed as,*

$$\langle\mu_\infty^c, f\rangle = \int_{[0,t_0^c)\times\mathbb{N}^2} f(s,y,z)\pi_s^c(dy,dz)ds, \tag{4.18}$$

for any function f with compact support on $[0,t_0^c)\times\mathbb{N}^2$, where (π_s^c) is an optional process with values in $\mathcal{P}\left(\mathbb{N}^2\right)$.

Proof. Note that, for $K>0$ and $t<t_0^c$, since

$$\int_0^t \mathbb{1}_{\left\{R_N^F(s)\ge K\right\}}ds = \int_0^t \mathbb{1}_{\left\{R_N^{F,c,d}(s)\ge K\right\}}ds$$

holds on the event $\left\{\tau_Q^N(c)\ge t_0^c\right\}$, then

$$\begin{aligned}&\mathbb{E}\left(\mu_N^c\left([0,t_0^c]\times\mathbb{N}\times[K,+\infty[\right)\right)\\ &\le \mathbb{E}\left(\mathbb{1}_{\left\{\tau_Q^N(c)\ge t_0^c\right\}}\int_0^{t_0^c}\mathbb{1}_{\left\{R_N^{F,c}(s)\ge K\right\}}ds\right) + t_0^c\mathbb{P}\left(\tau_Q^N(c)<t_0^c\right),\end{aligned}$$

and with Relation (4.17) and Conjecture 58, we have

$$\begin{aligned}&\mathbb{E}\left(\mathbb{1}_{\left\{\tau_Q^N(c)\ge t_0\right\}}\int_0^{t_0^c}\mathbb{1}_{\left\{R_N^{F,c}(s)\ge K\right\}}ds\right)\\ &\le \int_0^{t_0^c}\mathbb{P}\left(Z(s)\ge K\right)ds = \frac{1}{N}\int_0^{Nt_0^c}\mathbb{P}\left(Z(s)\ge K\right)ds,\end{aligned}$$

since the Markov process $(Z(t))$ converges in distribution to a Poisson distribution with parameter $\delta a/\kappa_0$, the ergodic theorem for Markov processes and Conjecture 58 give therefore the inequality

$$\limsup_{N\to+\infty}\mathbb{E}\left(\mu_N^c\left([0,t_0^c]\times\mathbb{N}\times[K,+\infty[\right)\right)\leq t_0^c\mathbb{P}\left(\mathcal{N}_1\left(0,\delta a/\kappa_0\right)\geq K\right)$$

where $\mathcal{N}_1$ is a Poisson process on $\mathbb{R}_+$ with rate 1. One can choose K sufficiently large such that $\mathbb{E}\left(\mu_N^c\left([0,t_0^c]\times\mathbb{N}\times[K,+\infty[\right)\right)$ is arbitrarily small uniformly in N. Similarly, by replacing $\left(R_N^F,Z\right)$ by $\left(R^E,Y\right)$, the same property can be proved for $\mathbb{E}\left(\mu_N^c\left([0,t_0^c]\times[K,+\infty[\times\mathbb{N}\right)\right)$ for K and N sufficiently large. Therefore, for any $\varepsilon>0$, there exists some K_0 such that

$$\sup_N\mathbb{E}\left(\mu_N^c\left([0,t_0^c]\times[0,K_0]^2\right)\right)\geq(1-\varepsilon)t_0^c$$

Lemma 1.3 of Kurtz [45] shows that the sequence (μ_N^c) is tight, and Lemma 1.4 of the same reference give the representation (4.18). □

Proposition 60 has established tightness properties of (μ_N^c). The following lemma extends this result in terms of the convergence of stochastic processes. It will be used repeatedly, in particular to identify the possible limits of (μ_N^c).

Lemma 61. *Under the conditions of Proposition 59, if $\left(\mu_{N_k}^c\right)$ is a subsequence converging to μ_∞^c satisfying relation (4.18), then for any $g\in\mathcal{C}_\rfloor\left(\mathbb{N}^2\right)$, for the convergence in distribution of processes associated to the uniform norm,*

$$\lim_{k\to+\infty}\int_0^t\left(g\left(R_N^{E,c}(s),R_N^{F,c}(s)\right)ds\right)=\left(\int_0^t\int_{N^2}g(y,z)\pi_s(dy,dz)ds\right)$$

Proof. The tightness of the sequence of stochastic processes is obtained by the use of the criterion of the modulus of continuity. See Theorem 7.3 of Billingsley [10]. The identification if the limit is straightforward consequence of the convergence of (μ_N^c). □

In the following, we will denote by $(W_N(s))$ the sequence $\left(R_N^E(s),R_N^F(s)\right)$. We have

$$\begin{aligned}\frac{1}{N}f\left(W_N(t)\right)=&\frac{f\left(W_N(0)\right)}{N}+\frac{M_{f,N}(t)}{N}\\&+\lambda\int_0^t\nabla_{e_1}(f)\left(W_N(s)\right)\frac{R_N^M(s)}{N}ds\\&+\alpha\int_0^t\nabla_{-e_1+e_2}(f)\left(W_N(s)\right)R_N^E(s)\frac{Q_N(s)}{N}ds\\&+\delta\int_0^t\nabla_{-e_2}(f)\left(W_N(s)\right)\frac{C_m^N-R_N^M(s)}{N}R_N^F(s)ds\\&+\gamma\int_0^t\nabla_{-e_2}(f)\left(W_N(s)\right)\frac{C_q^N-Q_N(s)}{N}R_N^F(s)ds\\&+\eta\int_0^t\nabla_{e_2}(f)\left(W_N(s)\right)\frac{R_N^S(s)}{N}ds\end{aligned}\tag{4.19}$$

with the notation, for $(y,z) \in \mathbb{N}^2$, and $u, v \in \mathbb{N}$

$$\nabla_{u,v}(f)(y,z) \stackrel{def}{=} f(y+u, z+v) - f(y,z)$$

and for $1 \leq j \leq 2$, e_j is the jth unit vector of $\mathbb{R}^2$.

And $\left(\frac{M_{f,N}(t)}{N}\right)$ is a local martingale and its previsible increasing process is given by

$$\begin{aligned}
\left\langle \frac{M_{f,N}}{N} \right\rangle (t) = & \frac{\lambda}{N} \int_0^t \nabla_{e_1}(f) \left(W_N(s)\right)^2 \frac{R_N^M(s)}{N} ds \\
& + \frac{\alpha}{N} \int_0^t \nabla_{-e_1+e_2}(f) \left(W_N(s)\right)^2 R_N^E(s) \frac{Q_N(s)}{N} ds \\
& + \frac{\delta}{N} \int_0^t \nabla_{-e_2}(f) \left(W_N(s)\right)^2 \frac{C_m^N - R_N^M(s)}{N} R_N^F(s) ds \\
& + \frac{\gamma}{N} \int_0^t \nabla_{-e_2}(f) \left(W_N(s)\right)^2 \frac{C_q^N - Q_N(s)}{N} R_N^F(s) ds \\
& + \frac{\eta}{N} \int_0^t \nabla_{e_2}(f) \left(W_N(s)\right)^2 \frac{R_N^S(s)}{N} ds
\end{aligned} \tag{4.20}$$

Lemma 62. *Under the conditions of Proposition 59, if f is a continuous bounded function on $\mathbb{N}^2$, then the martingale $\left(\frac{M_{f,N}(t)}{N}, t<t_0^c\right)$ of Relation* (4.19) *converges in distribution to* 0.

Proof. We take care of one of the five terms of $(\langle M_{f,N}/N \rangle (t))$ of Relation (4.20), the arguments are similar for the others, even easier.

$$A_N(t) \stackrel{def}{=} \frac{\alpha}{N} \int_0^t \nabla_{-e_1+e_2}(f) \left(W_N(s)\right)^2 R_N^E(s) ds \frac{Q_N(s)}{N} ds$$

By using the coupling (4.17), and since $(\mathbb{E}(Y(u)))$ converges as u goes to infinity, we have

$$\lim_{N\to\infty} \mathbb{E}\left(\left\langle \frac{M_{f,N}}{N} \right\rangle (t) \right) = 0$$

Doob's Inequality shows then the convergence of $(M_{f,N}(t)/N, t<t_0^c)$ to 0. The lemma is proved. □

Proposition 63. *Under the conditions of Proposition 59, and if μ_∞^c is a limiting point of μ_N^c with the representation* (4.18) *of Proposition 60, then, for any $t<t_0^c$ and any continuous function g on $\mathbb{N}^2$, we have*

$$\begin{aligned}
& \int_0^t \int_{N^2} g(y,z) \pi_s(dy,dz) ds \\
& \quad = \int_0^t \mathbb{E}\left(g\left(\mathcal{N}_1\left(0, \frac{\lambda}{\alpha} \frac{r^M(s)}{q(s)}\right), \mathcal{N}_2\left(0, \frac{(\lambda-\eta) r^M(s) + \eta}{\delta\left(c_m - r^M(s)\right) + \gamma\left(c_q - q(s)\right)}\right)\right)\right) ds
\end{aligned} \tag{4.21}$$

where $\mathcal{N}_1$, $\mathcal{N}_2$ and $\mathcal{N}_3$ are three independent Poisson processes on $\mathbb{R}_+$ with rate 1.

Proof. Referring to relation (4.19), and by using Lemmas 61 and 62 and Relation (4.17) and the convergence of the martingale to 0, we get:

$$\int_0^t \int_{\mathbb{N}^2} \lambda r^M(s) \nabla_{e_1}(f)(y,z) + \alpha \nabla_{-e_1+e_2}(f)(y,z) q_1(s) y \\ + \nabla_{-e_2}(f)(y,z) \left(\delta(c_m - r^M(s)) + \gamma(c_q - q(s))\right) z \\ + \eta \nabla_{e_2}(f)(y,z)(1 - r^M(s)) \pi_s(dx,dy,dz) ds = 0$$

Consequently, for almost all t, the relation

$$\int_{\mathbb{N}^2} \Omega(f)(y,z) \pi_t(dy,dz) = 0$$

holds.
Thus, $\pi_t(dy,dz)$ is the invariant distribution associated to the $\mathcal{Q}-$matrix Ω. However, Ω is the jump matrix of two birth and death processes $(U_1(t))$ and $(U_2(t))$ with parameters $\frac{\lambda}{\alpha}\frac{r^M(s)}{q_1(s)}$ for $(U_1(t))$ and

$$\frac{(\lambda - \eta) r^M(s) + \eta}{\delta\left(c_m - r^M(s)\right) + \gamma\left(c_q - q(s)\right)}$$

for $(U_2(t))$. The proposition is proved. □

Proposition 64. *Under the conditions of Proposition 59, for the convergence in distribution*

$$\lim_{k \to +\infty} \left(\int_0^t R^E_{N_k}(s) ds, t < t_0^c \right) = \left(\int_0^t \frac{\lambda}{\alpha} \frac{r^M(s)}{q(s)} ds, t < t_0^c \right)$$
$$\lim_{k \to +\infty} \left(\int_0^t R^F_{N_k}(s) ds, t < t_0^c \right) = \left(\int_0^t \frac{(\lambda - \eta) r^M(s) + \eta}{\delta(c_m - r^M(s)) + \gamma(c_q - q(s))} ds, t < t_0^c \right)$$

Proof. For $0 \leq s \leq t$, the coupling (4.17) and Cauchy-Schwartz' Inequality give

$$\mathbb{E}\left(\left(\mathbb{1}_{\left\{\tau_Q^N(c) > t\right\}} \int_s^t R^{E,c}_{N_k}(s) ds \right)^2 \right) \leq (t-s) \mathbb{E}\left(\mathbb{1}_{\left\{\tau_Q^N(c) > t\right\}} \int_s^t R^{E,c}_{N_k}(s)^2 ds \right) \\ \leq (t-s) \mathbb{E}\left(\int_s^t Y(Ns)^2 ds \right) \leq (t-s)^2 \sup_{u \geq 0} \mathbb{E}\left(Y(u)^2\right).$$

Therefore, according to the Kolmogorov-Centsov's criterion, the sequence of stochastic processes

$$\left(\int_0^t R^E_{N_k}(s) ds, t < t_0^c \right)$$

is tight for the convergence in distribution.
For $K > 0$, proposition 63 gives the convergence in distribution

$$\lim_{k \to +\infty} \left(\int_0^t R^E_{N_k}(s) \wedge K ds, t < t_0^c \right) = \left(\int_0^t \mathbb{E}\left(\mathcal{N}_1\left(0, \frac{\lambda}{\alpha} \frac{r^M(s)}{q(s)}\right) \wedge K \right) ds, t < t_0^c \right)$$

By using Relation (4.17), we have

$$\mathbb{E}\left(\int_0^{t_0^c} R^F_{N_k}(s)\mathbb{1}_{\left\{R^F_{N_k}(s)\geq K\right\}}ds\right) \leq \mathbb{E}\left(\int_0^{t_0^c} Y(N_k s)\mathbb{1}_{\{Y(N_k s)\geq K\}}ds\right)$$

and the convergence in distribution of $(Y(t))$, as t goes to infinity, to $Y(\infty)$ a random variable with a Poisson distribution with parameter $\frac{\lambda}{\alpha c}$ give

$$\lim_{k\to+\infty}\mathbb{E}\left(\int_0^{t_0^c} Y(N_k s)\mathbb{1}_{\{Y(N_k s)\geq K\}}ds\right) = t_0^c\mathbb{E}\left(Y(\infty)\mathbb{1}_{\{Y(\infty)\geq K\}}\right).$$

It is then easy to obtain the first convergence by letting K go to infinity. And same method for $\left(\int_0^t R^F_{N_k}(s)ds, t<t_0^c\right)$.

The proposition is proved. □

Theorem 65 (Law of Large Numbers)**.** *Under the conditions of Proposition 59, the following convergence in distribution holds,*

$$\lim_{N\to+\infty}\left(\frac{Q_N(t)}{N}, \frac{R^M_N(t)}{N}\right) = \left(q(t), r^M(t)\right),$$

where $(q(t))$ (resp. $\left(r^M(t)\right)$) is the solution of the ODE

$$\frac{dq_1}{dt}(t) = \beta - \lambda r^M(t) \tag{4.22}$$

$$\begin{aligned}\frac{dr^M}{dt}(t) &= -\lambda r^M(t)\\ &\quad + \frac{\delta(c_m - r^M(t))}{\delta(c_m - r^M(t)) + \gamma\,(c_q - q(t))}\left(\eta + r^M(t)(\lambda-\eta)\right)\end{aligned} \tag{4.23}$$

Using relation (4.22), we get the equilibrium point for $\left(r^M(t)\right)$,

$$r^M_\infty = \frac{\beta}{\lambda}$$

If $r^M_0 < r^M_\infty$ then $r^M(t) \geq r^M_0$ for all $t\geq 0$, and if $r^M_0 \geq r^M_\infty$ then $r^M(t)\geq r^M_\infty$ for all $t\geq 0$. Hence, we can extend our analysis for all $t\geq 0$.

We observe that the equilibrium point of $\left(r^S(t)\right)$ in this model coincides with the equilibrium point of $\left(r^E(t)\right)$ in the section 4.9 where sequestration was not present. As a result, the maximal sequestration has effectively relocated all ribosomes that were blocked in elongation phase to become sequestered ribosomes.

And by replacing r^M_∞ in relation (4.23) by its value, we get

$$q_\infty = c_q - \mathcal{S} > 0$$

where $\mathcal{S}$ defined by relation (4.2).
That means, that the role of maximum sequestration is not only about "unblocking" ribosomes in Elongation phase but also to control the number of charged tRNAs and to keep it as large as possible.

This indicates that maximal sequestration reduces the number of ribosomes blocked during translation, leading to a decrease in potential errors in protein production (see O'Farrell [55]). Moreover, it regulates the quantity of empty tRNAs, facilitating a gradual return of cellular activity to its usual state.

4.8 Partial Sequestration in the Deficiency of Amino Acids

Conditions (4.5) now hold. In this section, all propositions are demonstrated using the same approach as in the preceding section 4.7.

Definition 66. *The scaled process is defined by*

$$\left(\overline{X}_N(t)\right) = \left(\overline{R}^N_M(t), \overline{R}^E_N(t), \overline{R}^S_N(t), Q_N(t)\right). \tag{4.24}$$

If g is non-negative Borelian function on $\mathbb{R}_+ \times \mathbb{N}^2$, we define the occupation measure

$$\langle \mu_N, g \rangle \stackrel{def}{=} \int_{\mathbb{R}_+} g(s, Q_N(s), R^F_N(s))ds. \tag{4.25}$$

The following initial conditions will be assumed,

$$\begin{gathered} Q_N(0) = q_0 \in \mathbb{N}\,,\ R^F_N(0) = r^F_0 \in \mathbb{N}, \\ \lim_{N\to\infty} \frac{R^M_N(0)}{N} = r^M_0\ ,\ \lim_{N\to\infty} \frac{R^E_N(0)}{N} = r^E_0. \end{gathered} \tag{4.26}$$

with $r^M_0 + r^E_0 < 1$ and $q_0 \in (0, c_q)$.

And the number of sequestered ribosomes $R^S_N(0)$ is therefore such

$$\lim_{N\to\infty} \frac{R^S_N(0)}{N} = 1 - r^M_0 - r^E_0.$$

As it will be seen in Section 4.28, there is no loss of generality to consider these initial conditions.

As in Section 4.7, before proving the convergence of the sequence of processes $\left(\overline{X}_N(t)\right)$, we analyze the convergence of a "stopped" version of it. In several technical arguments we will need that the number of ribosomes in the elongation phase waiting an amino acid is not too small. A second step is of showing that, essentially, the stopped process does not differ from the original process.

Definition 67. *For $b > 0$, we define the stopping time $\tau^N_{R^E}(b)$ as*

$$\tau^N_{R^E}(b) \stackrel{def}{=} \inf\left\{t > 0; R^E_N(t) \le bN\right\} \tag{4.27}$$

and

a. *If* $(W(t))$ *is a càdlàg process, we denote* $\left(W_N^b(t)\right) = \left(W\left(N\left(t \wedge \tau_{R^E}^N(b)\right)\right)\right)$*;*

b. *The "stopped" occupation measure* μ_N^b *is defined by, if* g *is non-negative Borelian function on* $\mathbb{R}_+ \times \mathbb{N}^2$*,*

$$\left\langle \mu_N^b, g \right\rangle \stackrel{def}{=} \int_0^{\tau_{R_1^E}^N(b)} g(s, Q_N^b(s), R_N^{F,b}(s))ds$$

With a slight abuse, the notation $\left(\overline{W}_N^b(t)\right) = \left(\frac{W_N^b(t)}{N}\right)$ will be used in the following.

4.8.1 Technical Lemmas

The two processes $(Q_N(t))$ and $\left(R_N^F(t)\right)$ are in fact in a neighborhood of 0 quickly. They will be the fast processes of our averaging principle. In state $R_N^E = r^E$, $R_N^F = r^F$, $R_N^M = r^M$ and $Q_N = q$, the jump rates of the processes $\left(R_N^F(t)\right)$ are as represented in Section 4.7.1. And the jump rates of the process $(Q_N(t))$ are

$$\begin{cases} +1 & \beta N \mathbb{1}_{\{q < C_q^N\}}, \\ -1 & \alpha r^E q. \end{cases}$$

Same as in Section 4.7.1, for N sufficiently large and up to time $\tau_{R^E}^N(b)$, a simple coupling shows that there exist birth and death processes $(X(t))$ and $(Z(t))$ such that

$$Q_{1,N}(t) \leq X(Nt) \text{ and } R_N^F(t) \leq Z(Nt) \tag{4.28}$$

holds.

Conjecture 68. *Under Conditions* (4.5)*, and* (4.26)*, there exist* t_0^b *such that*

$$\lim_{N \to +\infty} \mathbb{P}\left(\tau_{R^E}^N(b) < t_0^b\right) = 0$$

4.8.2 Tightness Properties

Proposition 69. *Under Conditions* (4.5)*, and* (4.26) *and Conjecture 68, the sequences of processes*

$$\left(\overline{R}_N^{M,b}(t)\right), \ \left(\overline{R}_N^{E,b}(t)\right) \text{ and } \left(\overline{R}_N^{S,b,c}(t)\right)$$

are tight for the convergence in distribution.

Proof. Same method as in the proof of proposition 59. □

We denote their limits, resp.

$$\left(r^M(t)\right), \ \left(r^E(t)\right), \text{ and } \left(r^S(t)\right) = \left(1 - r^M(t) - r^E(t)\right)$$

We establish the tightness of the occupation measure and represent its limit using the same arguments presented in the previous section. Additionally, we prove the convergence in distribution of the sequence of processes in the following propositions.

Proposition 70. *Under the conditions of Proposition 69, the sequence of measure-valued processes* $\left(\mu_N^b\right)$ *on the state space* $[0,t_0^b)\times\mathbb{N}^2$ *is tight for the convergence in distribution and any limiting point* μ_∞^c *can be expressed as,*

$$\left\langle \mu_\infty^b, f\right\rangle = \int_{[0,t_0^b)\times\mathbb{N}^2} f(s,y,z)\pi_s^b(dy,dz)ds, \tag{4.29}$$

for any function f *with compact support on* $[0,t_0^b)\times\mathbb{N}^2$*, where* $\left(\pi_s^b\right)$ *is an optional process with values in* $\mathcal{P}\left(\mathbb{N}^2\right)$*.*

Proof. Same method as in the proof of proposition 60. □

Proposition 71. *Under the conditions of Proposition 69, and if* μ_∞^b *is a limiting point of* μ_N^b*, then, for any* $t<t_0^b$ *and any continuous function* g *on* $\mathbb{N}^2$*, we have*

$$\begin{aligned}\lim_{N\to+\infty}&\left(\int_0^t g\left(Q_N^b(s),R_N^{F,b}(s)\right)ds\right)\\ &=\int_0^t \mathbb{E}\left(g\left(\mathcal{N}_1\left(0,\frac{\beta}{\alpha r^E(s)}\right),\mathcal{N}_2\left(0,\frac{\eta(1-r^M(s)-r^E(s))+\beta}{\delta(c_m-r^M(s))+\gamma c_q}\right)\right)\right)ds\end{aligned} \tag{4.30}$$

where $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *are three independent Poisson processes on* $\mathbb{R}_+$ *with rate* 1*.*

Proof. Same method as in the proof of proposition 63. □

Theorem 72 (Law of Large Numbers)**.** *Under the conditions of Proposition 69, the following convergence in distribution holds,*

$$\lim_{N\to+\infty}\left(\frac{R_N^E(t)}{N},\frac{R_N^M(t)}{N}\right)=\left(r^E(t),r^M(t)\right)$$

where $\left(r^E(t)\right)$ *(resp.* $\left(r^M(t)\right)$*) is the solution of the ODE*

$$\frac{dr^E}{dt}(t)=\lambda r^M(t)-\beta \tag{4.31}$$

$$\frac{dr^M}{dt}(t)=-\lambda r^M(t)+1+\delta\frac{\beta+\eta(1-r^M(t)-r^E(t))}{\delta(c_m-r^M(t))+\gamma c_q} \tag{4.32}$$

Proof. Same method as in the proof of theorem 65. □

Using relation (4.31), we get the equilibrium point for $\left(r^M(t)\right)$,

$$r_\infty^M=\frac{\beta}{\lambda}$$

So it is the same equilibrium point as in the model without sequestration in section 4.9.
And by replacing r_∞^M in relation (4.32) by its value and using that $r_\infty^S = 1 - r_\infty^M - r_\infty^E$, we can find the value of $r_\infty^S > 0$.
Finally, $r_\infty^E = 1 - r_\infty^M - r_\infty^S < 1 - r_\infty^M$, we notice that it is smaller than the one found in the model without sequestration represented in Section 4.9. What shows that the regulation decreases the number of "blocked" ribosomes in elongation.

4.9 Regulation-independent deficiency of amino acids

In this section, our focus will be on investigating the model in the absence of sequestration but with a deficiency of amino acids. This allows us to understand the cellular behavior when the regulatory mechanism is not in effect. To achieve this, we will first introduce the model representation, and then we will present the results obtained using the same methodology as in the previous section 4.7.

4.9.1 Model

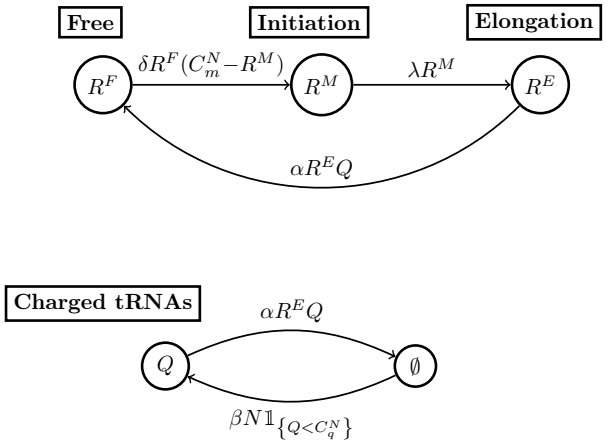


Figure 4.4: Ribosomes: Translation without Sequestration

Conditions (4.4) hold throughout this section.

Definition 73. *The scaled process is defined by*

$$\left(\overline{X}_N(t)\right) = \left(\overline{R}_M^N(t), \overline{R}_N^E(t), Q_N(t)\right). \tag{4.33}$$

If g is non-negative Borelian function on $\mathbb{R}_+ \times \mathbb{N}^2$, we define the occupation measure

$$\langle \mu_N, g \rangle \stackrel{def}{=} \int_{\mathbb{R}_+} g(s, Q_N(s), R_N^F(s)) ds. \tag{4.34}$$

The following initial conditions will be assumed,

$$\begin{gathered} Q_N(0) = q_0 \in \mathbb{N}\,,\; R_N^F(0) = r_0^F \in \mathbb{N}, \\ \lim_{N\to\infty} \frac{R_N^M(0)}{N} = r_0^M, \end{gathered} \tag{4.35}$$

with $r_0^M < 1$ and $q_0 \in (0, c_q)$.

And the number of ribosomes in elongation phase waiting an amino acid, $R_N^E(0)$ is therefore such

$$\lim_{N\to\infty} \frac{R_N^E(0)}{N} = 1 - r_0^M.$$

Similar to section 4.8, we will examine the convergence of a "stopped" version of the model. We will sketch the method used. By defining the same stopping time used in section 4.8(Definition 67) and couplings of the type (4.28) are proved, as well as an analog of Conjecture 68, with the same method, in this section.

4.9.2 Tightness Properties

Conjecture 74. *Under Conditions* (4.4) *and* (4.35)*, the sequences of processes*

$$\left(\overline{R}_N^{M,b}(t)\right) \text{ and } \left(\overline{R}_N^{E,b}(t)\right)$$

are tight for the convergence in distribution.

We denote their limits, resp.

$$\left(r^M(t)\right) \text{ and } \left(1 - r^M(t)\right)$$

Additionally, we get the convergence in distribution of the sequence of processes in the following proposition.

Proposition 75. *Under Conjecture 74, and if μ_∞^b is a limiting point of μ_N^b, then, for any $t<t_0^b$ and any continuous function g on $\mathbb{N}^2$, we have*

$$\begin{aligned} &\lim_{N\to+\infty} \left(\int_0^t g\left(Q_N^b(s), R_N^{F,b}(s)\right) ds \right) \\ &\qquad = \int_0^t \mathbb{E}\left(g\left(\mathcal{N}_1\left(0, \frac{\beta}{\alpha(1 - r^M(s))}\right), \mathcal{N}_2\left(0, \frac{\beta}{\delta(c_m - r^M(s))}\right)\right)\right) ds \end{aligned} \tag{4.36}$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ are three independent Poisson processes on $\mathbb{R}_+$ with rate 1.

Proof. Same method as in the proof of proposition 63. □

Theorem 76 (Law of Large Numbers)**.** *Under the conditions of Proposition 75, the following convergence in distribution holds,*

$$\lim_{N\to+\infty}\left(\frac{R_N^M(t)}{N}\right)=\left(r^M(t)\right)$$

where $\left(r^M(t)\right)$ *is the solution of the ODE*

$$\begin{aligned}\frac{dr^M}{dt}(t)&=\beta-\lambda r^M(t)\\ r^M(t)&=\left(r_0^M-\frac{\beta}{\lambda}\right)e^{-\lambda t}+\frac{\beta}{\lambda}\end{aligned}\tag{4.37}$$

Proof. Same method as in the proof of Theorem 65. □

Using relation (4.37), we get the equilibrium point for $\left(r^M(t)\right)$,

$$r_\infty^M=\frac{\beta}{\lambda}$$

4.10 Generalizing Results to a Model with Multiple Amino Acids

In this chapter, we proved our results using a simplified model with only one amino acid for clarity and ease of understanding. However, all the results obtained from this one amino acid model can be extended to the more complex model introduced in the beginning 4.4.2, which involves two types of amino acids.
By doing so, the results discussed in the previous sections on the model with one amino acid, along with their interpretation concerning the impact of sequestration on ribosome distribution, can be applied to the model with two amino acids, with scenarios involving deficient amino acids. The same methods employed to prove the earlier results can be effectively used in this case in order to prove the outcomes in Section 4.3.4 concerning the model with two types of amino acids.

The only distinction lies in a few definitions that we are about to present.

Definition 77. — *For* $i,j\in\{1,2\}$*, with* $i\neq j$ *we define* $\mathcal{B}_i$ *as*

$$\mathcal{B}_i\overset{def}{=}\frac{(\alpha_i+\psi_i+\nu_i)(\lambda_i\alpha_j+\lambda_i\psi_j+\lambda_j\psi_j)}{\alpha_i\alpha_j+\alpha_i\psi_j+\alpha_j\psi_i}$$

$\mathcal{B}_i$ interpreted as the maximal consumption rate of amino acids of type i.

— *We also introduce $\mathcal{C}$ as*

$$\mathcal{C}\overset{def}{=}\frac{\mathcal{B}_1}{\beta_1}\frac{\eta\delta\left(1-\frac{\beta_1}{\mathcal{B}_1}\right)\left(c_m-\frac{\beta_1}{\mathcal{B}_1}\right)}{\gamma_1(\lambda_1+\lambda_2)}\tag{4.38}$$

$\mathcal{C}$ interpreted as the maximal sequestration rate.

In the model with 1 amino acid : $\mathcal{B}_1 = \lambda$.
We can now introduce the four regimes of interest in our study.

Definition 78. *a. The regime of Adequate supply of amino acids is defined by the relations*

$$\beta_i > \mathcal{B}_i, \ \ \forall i \in \{1, 2\} \tag{4.39}$$

The production rate β_i of amino acids of type i is greater than its consumption rate.

b. The regime of Regulation independent deficiency of type 1 *amino acids is defined by the relations*

$$\begin{aligned} \beta_1 &< \mathcal{B}_1 \\ \beta_2 &> \mathcal{B}_2 \end{aligned} \tag{4.40}$$

The production rate β_1 of amino acids of type 1 *is less than its consumption rate and the production rate β_2 of amino acids of type* 2 *is greater than its consumption rate.*

c. The regime of Partial Sequestration in the deficiency of type 1 amino acids is defined by the relations

$$\begin{aligned} \beta_1 &< \mathcal{B}_1 \\ \beta_2 &> \mathcal{B}_2 \\ c_{q,1} &< \mathcal{C} \end{aligned} \tag{4.41}$$

Same conditions (4.40) *with the fraction of tRNAs of the lacking amino acid less than the maximal sequestration rate.*

d. The regime of Maximal Sequestration in the deficiency of type 1 amino acids is defined by the relations

$$\begin{aligned} \beta_1 &< \mathcal{B}_1 \\ \beta_2 &> \mathcal{B}_2 \\ c_{q,1} &> \mathcal{C} \end{aligned} \tag{4.42}$$

Same conditions (4.40) *with the fraction of tRNAs of the lacking amino acid less than the maximal sequestration rate.*

The generalization of our results to this comprehensive model is a significant step towards a more comprehensive understanding of the intricate dynamics underlying a translational regulation mechanism and its implications for protein synthesis.

Bibliography

[1] M.-A. Agustino and J. Collado-Vides. "Identifying global regulators in transcriptional regulatory networks in bacteria". In: *Curr Opin Microbiol.* (2003), pp. 482–489.

[2] P. M. et al. "The nucleotide-binding site of bacterial translation initiation factor 2 (IF2) as a metabolic sensor". In: *Proceedings of the National Academy of Sciences of the United States of America.* 103 (2006), pp. 13962–13967.

[3] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* 4th edition. New York, NY: Garland Science, 2002.

[4] D. F. Anderson, G. Craciun, and T. G. Kurtz. "Product-form stationary distributions for deficiency zero chemical reaction networks". In: *Bulletin of mathematical biology* 72.8 (2010), pp. 1947–1970.

[5] D. F. Anderson and T. G. Kurtz. *Stochastic Analysis of Biochemical Systems.* Mathematical Biosciences Institute Lecture Series. Springer Publishing Company, Incorporated, 2015.

[6] C. Baker, T. Jia, and R. V. Kulkarni. "Stochastic modeling of regulation of gene expression by multiple small RNAs". In: *Physical Review E* 85.6 (2012), p. 061915.

[7] S. Bakshi, A. Siryaporn, M. Goulian, and J. C. Weisshaar. "Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells". In: *Molecular Microbiology* 85.1 (2012), pp. 21–38.

[8] C. L. Beisel and G. Storz. "Base pairing small RNAs and their roles in global regulatory networks". In: *FEMS Microbiology Reviews* 34.5 (Sept. 2010), pp. 866–882.

[9] O. G. Berg. "A model for the statistical fluctuations of protein numbers in a microbial population". In: *Journal of theoretical biology* 71.4 (Apr. 1978), pp. 587–603.

[10] P. Billingsley. *Convergence of probability measures.* Second. Wiley Series in Probability and Statistics: Probability and Statistics. New York: John Wiley & Sons Inc., 1999, pp. x+277.

[11] H. Bremer and P. P. Dennis. "Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates". In: *EcoSal Plus* 3.1 (2008).

[12] P. C. Bressloff. *Stochastic processes in cell biology*. Interdisciplinary applied mathematics. Springer, 2014.

[13] R. J. Britten and E. H. Davidson. "Gene regulation for higher cells: a theory". In: *Science* 165.3891 (July 1969), pp. 349–357.

[14] A. T. Cavanagh and K. M. Wassarman. "6S RNA, a global regulator of transcription in Escherichia coli, Bacillus subtilis, and beyond". In: *Annual review of microbiology* 68 (2014), pp. 45–60.

[15] J. Chen, K. M. Wassarman, S. Feng, K. Leon, A. Feklistov, J. T. Winkelman, Z. Li, T. Walz, E. A. Campbell, and S. A. Darst. "6S RNA mimics B-form DNA to regulate Escherichia coli RNA polymerase". In: *Molecular cell* 68.2 (2017), pp. 388–397.

[16] C. L. P. Claudio O. Gualerzi. "Initiation of mRNA translation in bacteria: structural and dynamic aspects". In: *Cellular and Molecular Life Sciences* 72 (2015), pp. 4321–4367.

[17] A. Cornish-Bowden. "One hundred years of Michaelis–Menten kinetics". In: *Perspectives in Science* 4 (2015), pp. 3–9.

[18] S. J. W. B. David J Lee Stephen D Minchin. "Activating transcription in bacteria". In: *Annu Rev Microbiol* (2012), pp. 092611–150012.

[19] D. A. Dawson. "Measure-valued Markov processes". In: *École d'Été de Probabilités de Saint-Flour XXI—1991*. Vol. 1541. Lecture Notes in Math. Berlin: Springer, Nov. 1993, pp. 1–260.

[20] M. Del Giudice, C. Bosia, S. Grigolon, and S. Bo. "Stochastic sequestration dynamics: a minimal model with extrinsic noise for bimodal distributions and competitors correlation". In: *Scientific Reports* 8 (2018), p. 10387.

[21] R. Dessalles, V. Fromion, and P. Robert. "A Stochastic Analysis of Autoregulation of Gene Expression". In: *Journal of Mathematical Biology* 75.5 (Nov. 2017), pp. 1253–1283.

[22] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. "Stochastic gene expression in a single cell". In: *Science* 297.5584 (Aug. 2002), pp. 1183–1186.

[23] T. H. Erel Levine. "Small RNAs establish gene expression thresholds". In: *Curr. Opin. Microbiol.* 11 (2008), pp. 574–579.

[24] S. N. Ethier and T. G. Kurtz. *Markov processes: Characterization and Convergence*. New York: John Wiley & Sons Inc., 1986, pp. x+534.

[25] M. Feinberg. *Foundations of chemical reaction network theory*. Vol. 202. Applied Mathematical Sciences. Springer, Cham, 2019, pp. xxix+454.

[26] B. Felden and Y. Augagneur. "Diversity and Versatility in Small RNA-Mediated Regulation in Bacterial Pathogens". In: *Frontiers in Microbiology* 12 (2021).

[27] H. Fer et al. "The alarmones (p)ppGpp are part of the heat shock response of Bacillus subtilis". In: *PLoS Genetics* 16.3 (Mar. 2020), e1008275.

[28] A. S. Flynt and E. C. Lai. "Biological principles of microRNA-mediated regulation: shared themes amid diversity". In: *Nature Reviews Genetics* 9.11 (Nov. 2008), pp. 831–842.

[29] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*. Second. New York: Springer-Verlag, 1998, pp. xii+430.

[30] V. Fromion, E. Leoncini, and P. Robert. "Stochastic Gene Expression in Cells: A Point Process Approach". In: *SIAM Journal on Applied Mathematics* 73.1 (Jan. 2013), pp. 195–211.

[31] V. Fromion, P. Robert, and J. Zaherddine. "Stochastic Models of Regulation of Transcription in Biological Cells". In: *Journal of Mathematical Biology* (2023).

[32] R. Geißen, B. Steuten, T. Polen, and R. Wagner. "E. coli 6S RNA: A universal transcriptional regulator within the centre of growth adaptation". In: *RNA Biology* 7.5 (2010), pp. 564–568.

[33] S. Gottesman, C. McCullen, M. Guillier, C. Vanderpool, N. Majdalani, J. Benhammou, K. Thompson, P. FitzGerald, N. Sowa, and D. FitzGerald. "Small RNA regulators and the bacterial response to stress". In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 71. Cold Spring Harbor Laboratory Press. 2006, pp. 1–11.

[34] S. Gottesman and G. Storz. "Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations". In: *Cold Spring Harb Perspect Biol* (2011).

[35] P. C. Gualerzi CO. "Initiation of mRNA translation in bacteria: structural and dynamic aspects". In: *Cell Mol Life Sci* 72 (2015), pp. 4341–4367.

[36] V. Hauryliuk, G. C. Atkinson, K. S. Murakami, T. Tenson, and K. Gerdes. "Recent functional insights into the role of (p)ppGpp in bacterial physiology". In: *Nature Reviews Microbiology* 13.5 (May 2015), pp. 298–309.

[37] J. T. Hindley. "Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting". In: *Journal of molecular biology* 30.1 (1967), pp. 125–136.

[38] J. Jagodnik, A. Brosse, T. N. Le Lam, C. Chiaruttini, and M. Guillier. "Mechanistic study of base-pairing small regulatory RNAs in bacteria". In: *Methods* 117 (2017), pp. 67–76.

[39] G. E. Jakubowski H. "Quantities of individual aminoacyl-tRNA families and their turnover in Escherichia coli." In: *Bacteriol.* 3 (1984), pp. 769–775.

[40] C. Jayaprakash and J. Das. "Stochastic Sequestration Promotes Specificity in Decision Making in Single Cells". In: *The Journal of Physical Chemistry B* 123.49 (2019), pp. 10323–10330.

[41] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. 2nd ed. Vol. 113. Graduate Texts in Mathematics. New York, NY: Springer, 1998.

[42] P. D. Karp et al. "The BioCyc collection of microbial genomes and metabolic pathways". In: *Briefings in Bioinformatics* 20.4 (Aug. 2019), pp. 1085–1093.

[43] W. KM. "6S RNA: a small RNA regulator of transcription". In: *Curr Opin Microbiol* 10 (2007), pp. 164–168.

[44] N. Kumar, K. Zarringhalam, and R. V. Kulkarni. "Stochastic Modeling of Gene Regulation by Noncoding Small RNAs in the Strong Interaction Limit". In: *Biophysical Journal* 114.11 (2018), pp. 2530–2539.

[45] T. G. Kurtz. "Averaging for martingale problems and stochastic approximation". In: *Applied Stochastic Analysis*. Ed. by I. Karatzas and D. Ocone. Vol. 177. Springer-Verlag, 1992, pp. 186–209.

[46] M. C. Mackey, M. Santillán, M. Tyran-Kamińska, and E. S. Zeron. *Simple mathematical models of gene regulatory dynamics*. Lecture Notes on Mathematical Modelling in the Life Sciences. Springer, Cham, 2016, pp. xiv+124.

[47] P. Mehta, S. Goyal, and N. S. Wingreen. "A quantitative comparison of sRNA-based and protein-based gene regulation". In: *Mol Syst Biol* 4 (2008), p. 221.

[48] L. Michaelis and M. L. Menten. "The kinetics of invertin action". In: *Dept of Biochemistry* 49.333-369 (1913), p. 352.

[49] R. Milo, P. Jorgensen, U. Moran, G. M. Weber, and M. Springer. "BioNumbers - the database of key numbers in molecular and cell biology." In: *Nucleic Acids Res.* 38.Database-Issue (2010), pp. 750–753.

[50] N. Mitarai, J.-A. M. Benjamin, S. Krishna, S. Semsey, Z. Csiszovszki, E. Massé, and K. Sneppen. "Dynamic features of gene expression control by small regulatory RNAs". In: *Proceedings of the National Academy of Sciences* 106.26 (2009), pp. 10655–10659. eprint: `https://www.pnas.org/doi/pdf/10.1073/pnas.0901466106`.

[51] K. Montzka Wassarman, A. Zhang, and G. Storz. "Small RNAs in Escherichia coli". In: *Trends in Microbiology* 7.1 (1999), pp. 37–45.

[52] R. R. Mor Nitzan and H. Margalit. "Integration of Bacterial Small RNAs in Regulatory Networks". In: *Annu Rev Biophys* (2017), pp. 131–148.

[53] F. C. Neidhardt and H. E. Umbarger. "Chemical composition of *Escherichia coli*". In: Escherichia coli *and* Salmonella*: cellular and molecular biology*. 2 edition. ASM Press, May 1996.

[54] M. Nitzan, K. M. Wassarman, O. Biham, and H. Margalit. "Global regulation of transcription by a small RNA: a quantitative view". In: *Biophysical journal* 106.5 (2014), pp. 1205–1214.

[55] P. H. O'Farrell. "The suppression of defective translation by ppGpp and its role in the stringent response". In: *Cell* 14.3 (1978), pp. 545–557.

[56] G. Papanicolaou, D. W. Stroock, and S. R. S. Varadhan. "Martingale approach to some limit theorems". In: *Proc. 1976. Duke Conf. On Turbulence*. III. Duke Univ. Math, 1977.

[57] J. Paulsson. "Models of stochastic gene expression". In: *Physics of Life Reviews* 2.2 (June 2005), pp. 157–175.

[58] A. Petrides and G. Vinnicombe. "Enzyme sequestration by the substrate: An analysis in the deterministic and stochastic domains". In: *PLOS Computational Biology* 14.5 (May 2018), pp. 1–23.

[59] T. Platini, T. Jia, and R. V. Kulkarni. "Regulation by small RNAs via coupled degradation: mean-field and variational approaches". In: *Physical Review E* 84.2 (2011), p. 021928.

[60] D. Rigney and W. Schieve. "Stochastic model of linear, continuous protein synthesis in bacterial populations". In: *Journal of Theoretical Biology* 69.4 (Dec. 1977), pp. 761–766.

[61] P. Robert. *Stochastic Networks and Queues*. Vol. 52. Stochastic Modelling and Applied Probability Series. New-York: Springer, June 2003, pp. xix+398.

[62] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales: Volume 2, Itô Calcul us*. Cambridge University Press, Sept. 2000.

[63] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1: Foundations*. Second. Chichester: John Wiley & Sons Ltd., 1994.

[64] W. Rudin. *Real and complex analysis*. Third. New York: McGraw-Hill Book Co., 1987, pp. xiv+416.

[65] K. R. Sanft, D. T. Gillespie, and L. R. Petzold. "Legitimacy of the stochastic Michaelis–Menten approximation". In: *IET systems biology* 5.1 (2011), pp. 58–69.

[66] B. Steuten, S. Schneider, and R. Wagner. "6S RNA: recent answers–future questions". In: *Molecular microbiology* 91.4 (2014), pp. 641–648.

[67] K. M. Wassarman. "6S RNA, a Global Regulator of Transcription". In: *Microbiology Spectrum* 6.3 (May 2018).

[68] K. M. Wassarman and G. Storz. "6S RNA regulates E. coli RNA polymerase activity". In: *Cell* 101.6 (2000), pp. 613–623.

[69] L. S. Waters and G. Storz. "Regulatory RNAs in bacteria". In: *Cell* 136.4 (2009), pp. 615–628.

[70] J. D. Watson, T. A. Baker, S. P. Bell, A. Gann, M. Levine, R. Losick, and I. CSHLP. *Molecular biology of the gene*. 6th. San Francisco; Cold Spring Harbor, N.Y.: Pearson/Benjamin Cummings; Cold Spring Harbor Laboratory Press, Dec. 2007.

[71] J. Yang, B. W. Anderson, A. Turdiev, H. Turdiev, D. M. Stevenson, D. Amador-Noguez, V. T. Lee, and J. D. Wang. "The nucleotide pGpp acts as a third alarmone in Bacillus, with functions distinct from those of (p) ppGpp". In: *Nature Communications* 11.1 (Oct. 2020), p. 5388.

[72] P. D. Zamore and B. Haley. "Ribo-gnome: the big world of small RNAs". In: *Science* 309.5740 (Sept. 2005), pp. 1519–1524.