Mathematical Biology

Stochastic models of regulation of transcription in biological cells

Vincent Fromion[1] · Philippe Robert[2] · Jana Zaherddine[2]

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

In this paper we study an important *global* regulation mechanism of transcription of biological cells using specific macro-molecules, 6S RNAs. The functional property of 6S RNAs is of blocking the transcription of RNAs when the environment of the cell is not favorable. We investigate the efficiency of this mechanism with a scaling analysis of a stochastic model. The evolution equations of our model are driven by the law of mass action and the total number of polymerases is used as a scaling parameter. Two regimes are analyzed: exponential phase when the environment of the cell is favorable to its growth, and the stationary phase when resources are scarce. In both regimes, by defining properly occupation measures of the model, we prove an averaging principle for the associated multi-dimensional Markov process on a convenient timescale, as well as convergence results for "fast" variables of the system. An analytical expression of the asymptotic fraction of sequestrated polymerases in stationary phase is in particular obtained. The consequences of these results are discussed.

Keywords Gene expression · Stochastic model · Regulation · Averaging principle

Mathematics Subject Classification 60G55 · 92D20 · 92C37

✉ Philippe Robert
Philippe.Robert@inria.fr
http://www-rocq.inria.fr/who/Philippe.Robert

Vincent Fromion
Vincent.Fromion@inrae.fr

Jana Zaherddine
Jana.Zaherddine@inria.fr

[1] INRAE, MaIAGE, Université Paris-Saclay, Domaine de Vilvert, 78350 Jouy-en-Josas, France

[2] INRIA Paris, 2 rue Simone Iff, 75589 Paris Cedex 12, France

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. 2007).

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 2018) 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) RNAs used for the building of ribosomes, i.e., rRNAs, *ribosomal RNAs*. 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. 2006). Its second interest is experimental since its 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.

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 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. In particular if there is no idle polymerase for some time, the a free $6S$ will die.

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.

A future work in this domain should be devoted to the experimental validation of such a mathematical model. This can be done in the framework of a complete simulator of the bacterial cell such as in Fischer et al. (2021). This is an ongoing project.

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 the phase when the nucleotids of the RNA are retrieved one by one.

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. 2010) 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 1969) and Zamore and Haley (2005). 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 1967). 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 2014) and Nitzan (2014). 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 10,000 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 2018) and Steuten et al. (2014).

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.

1.3 Mathematical models

Regulation of gene expression has been analyzed with mathematical models for some time now. See (Mackey et al. 2016) and also Chapter 6 of Paul (2014), 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. 2017).

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. 2018; Mehta et al. 2008), and Platini et al. (2011), 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. 2012) and Mitarai et al. (2009). These studies can be seen as extensions of the early works on stochastic models of gene expression, see (Berg 1978; Elowitz et al. 2002) and Rigney and Schieve (1977).

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 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 process associated to the number of such polymerases.
- *Transcription of an rRNA.* A long chain of nucleotides is produced. As it will be seen, it is described by a process with values in $(\{0, 1\} \times \mathbb{N})^J$, where J is the number of types of rRNAs, it is usually small, less than ten. We denote by $(\overline{R}_N(t))$ process of the total number of polymerases in this phase.

– *Sequestered by a 6S RNA.* $(S_N(t))$ is the process associated to the number of such polymerases.

Similarly a 6S RNA can be either free or paired with a polymerase, $(Z_N(t))$ denotes the process of the number of free 6S RNAs. See Sect. 2.1 for more details.

With these notations, the assumption on the constant number of polymerases gives the relation

$$F_N(t) + M_N(t) + S_N(t) + \overline{R}_N(t) = N, \qquad \forall t \geq 0.$$

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 2015). The rate of creation of sequestered polymerases is in particular quadratic with respect to the state, it is proportional to $F_N(t)Z_N(t)$. This is one of the important features of this stochastic model.

Two Regimes

In the following, if $(X_N(t))$ is a sequence of stochastic processes, the notation

$$\lim_{N \to +\infty} (X_N(t)) = \lim_{N \to +\infty} (X_N(t), t \geq 0)$$

refers to the convergence of the distribution of the process $(X_N(t), t \geq 0)$, as in Chapter 2 of Billingsley (1999).

Our mathematical results can be described as follows. See the formal statements in Sect. 5 and 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. Under the condition 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 of processes

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

   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 of processes

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

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 $(\overline{R}_N(t))$ 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.

1.5 Outline of the paper

Section 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 Sect. 3 and 4. The last Sects. 5 and 6 are devoted to the exponential/stationary regimes of our model.

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.

2.1 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. The 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 the rRNA of type $1 \leq j \leq J$, it starts elongation at rate $\alpha_{r,j}$ if there are strictly less than $C_{r,j}^N$ polymerases in the elongation phase of this rRNA. At a given moment there cannot be more than $C_{r,j}^N$ 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 $j \in \{1, \ldots, J\}$ 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 $C_{r,j}^N$. See Sect. 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 min half-life in high-growth rate phase) of each of them. See Sect. 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
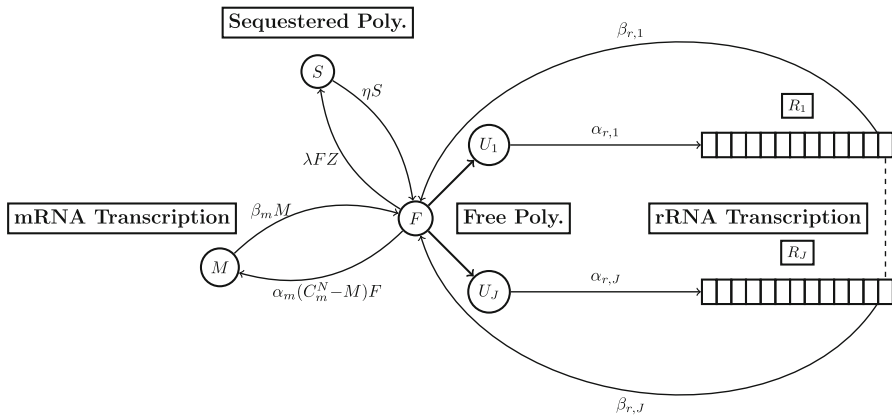
Fig. 1 Polymerases: Transcription of mRNAs/rRNAs and Sequestration

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 1978) and Rigney and Schieve (1977), see also (Paulsson 2005) 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 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.

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

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

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;
  $(u, r) = ((u_j, r_j), 1 \leq j \leq J)$,
- $u_j \in \{0, 1\}$ to indicate if a polymerase is bound to the promoter of the rRNA of type j or not;
- $0 \leq r_j \leq C_{r,j}^N$, 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

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

Note that the condition that $f > 0$ implies $u_j = 1$ for all $1 \leq j \leq J$ is a consequence of the assumption of the first paragraph of the "Transcription of rRNAs" sub-section at the beginning of this section.

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 $(U_N(t), R_N(t)) = ((U_j^N(t), R_j^N(t)), 1 \leq j \leq J)$. 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 $1 \leq j \leq J$, 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 $1\leq i,\, j\leq J$,

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

– 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.

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

A possible extension for mRNAs

We have chosen to consider C_m^N genes of mRNAs with the same parameters β_m and α_m for the transcription by polymerases, for simplicity essentially. A generalization could be considered for which the C_m^N types of mRNAs can be split into K subgroups $(\mathcal{C}_{m,k}^N)$ of respective sizes $C^N \overset{\text{def.}}{=} (C_{m,k}^N,\, 1\leq k\leq K)$ and with the parameters $(\beta_{m,k},\alpha_{m,k},\, 1\leq k\leq K)$.

It basically amounts to say that mRNAs can be partitioned according to the strengths of the affinity of their promoters and of their lengths in terms of nucleotides. See (Bremer and Dennis 2008) and Ron et al. (2010) for example. A similar assumption for translation of different types of proteins has been done in Fromion and Robert (2022). We denote by $M_{m,j}^N(t)$ the number of polymerases in the elongation phase of an mRNA whose type is in the set $\mathcal{C}_{m,j}^N$ at time t. The state process for this part of the system is

$$
(M_N(t)) \overset{\text{def.}}{=} (M_{m,k}^N(t))\in\mathcal{S}_m^N,\ \text{with}\ \mathcal{S}_m^N \overset{\text{def.}}{=} \left\{x\in\prod_{k=1}^K\left[0,C_{m,k}^N\right]: |x|\leq N\right\}.
$$

Without sequestration and transcription of rRNAs, the model is equivalent to a kind of Ehrenfest urn model, with $K+1$ urns, for $1 \le k \le K$, the urn k has a maximal capacity of $C_{m,k}^N$ and the balls inside move to urn 0 at rate $\beta_{m,k}$. A ball in urn 0 go to a specific empty place of urn k at rate $\alpha_{m,k}(C_{m,k}^N - M_{m,k}^N(t))$

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. 2012).
– The number J of different types of rRNA is small, of the order of 10. See (Bremer and Dennis 2008).
– 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 (Ron et al. 2010; Karp et al. 2019) and Neidhardt and Umbarger (1996) for the estimation of the numerical values of these quantities in various contexts.

Due to the modeling assumptions of Sect. 2.1, we assume that the relations

$$\lim_{N \to +\infty} \frac{C_{r,j}^N}{N} = c_{r,j} > 0, \ 1 \le j \le J, \ \text{and} \ \lim_{N \to +\infty} \frac{C_m^N}{N} = c_m, \tag{2}$$

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

$$\max\left(c_m, \sum_{j=1}^J c_{r,j}\right) < 1 < \sum_{j=1}^J c_{r,j} + c_m. \tag{3}$$

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

Definition 1 (a) The *Exponential Phase* is defined by the relation

$$\min_{1 \leq j \leq J} \frac{\alpha_{r,j}}{\beta_{r,j}} > 1. \tag{4}$$

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

$$\max_{1 \leq j \leq J} \frac{\alpha_{r,j}}{\beta_{r,j}} < 1. \tag{5}$$

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

2.4 An auxiliary model

In Sects. 3 and 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 introduction of the auxiliary process is essentially motivated for the mathematical analysis. It is a technical tool used for the proofs of convergence for the biological system. The intuitive reasons to study this case are two-fold:

(a) Exponential Phase. If Condition (4) holds, as we shall see, "most" of the $J + C_{r,1}^N + C_{r,2}^N + \cdots + C_{r,J}^N$ available places for transcriptions of rRNAs will be occupied by polymerases. Provided that this situation holds on a sufficiently large time scale, under Condition (3), there are A_N available polymerases for sequestration and transcription of mRNAs, with

$$A_N \overset{\text{def.}}{=} N - J - \sum_{j=1}^{J} C_{r,j}^N \sim \gamma N < C_m^N. \tag{6}$$

The system works as if there were A_N polymerases available for the transcription of mRNAs. With Condition (3), we have $A_N < C_m^N$.

(b) Stationary Phase. When Condition (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_m^N under Condition (3).

The mathematical analysis of exponential phase, resp. stationary phase, is done in Sect. 5, resp. Section 6.

We denote $(X_N^0(t)) = (F_N^0(t), S_N^0(t), Z_N^0(t))$ the system defined in Sect. 2.2 but without the part of the model for rRNAs. For $(f, s, z) \in$, the transition rates are:

$$(f, s, z) \longrightarrow \begin{cases} (f-1, s, z) & \alpha_m \left(C_m^N - (N - f - s) \right) f, \\ (f+1, s, z) & \beta_m (N - f - s), \\ (f, s, z+1) & \beta_6, \end{cases}$$

$$(f, s, z) \longrightarrow \begin{cases} (f, s, z-1) & \delta_6 z, \\ (f-1, s+1, z-1) & \lambda f z, \\ (f+1, s-1, z+1) & \eta s. \end{cases} \tag{7}$$

Using the classical formulation in terms of a martingale problem, see Theorem (20.6) in Section IV of Rogers and Williams (1994) for example, the Markov process $(X_N^0(t))$ whose Q-matrix is given by Relation (7), as $(F_N^0(t), S_N^0(t), Z_N^0(t))$, the solution of the SDEs,

$$dF_N^0(t) = \mathcal{P}_1\left(\left(0, \eta S_N^0(t-)\right), dt\right) - \mathcal{P}_3\left(\left(0, \lambda F_N^0(t-)Z_N^0(t-)\right), dt\right) \tag{8}$$
$$+ \mathcal{P}_2\left(\left(0, \beta_m\left(N - F_N^0(t-) - S_N^0(t-)\right), dt\right)\right)$$
$$- \mathcal{P}_4\left(\left(0, \alpha_m F_N^0(t-)\left(C_m^N - \left(N - F_N^0(t-) - S_N^0(t-)\right)\right)\right), dt\right)$$

$$dS_N^0(t) = -\mathcal{P}_1\left(\left(0, \eta S_N^0(t-)\right), dt\right) + \mathcal{P}_3\left(\left(0, \lambda F_N^0(t-)Z_N^0(t-)\right), dt\right) \tag{9}$$

$$dZ_N^0(t) = \mathcal{P}_5\left((0, \beta_6), dt\right) - \mathcal{P}_6\left(\left(0, \delta_6 Z_N^0(t-)\right), dt\right)$$
$$+ \mathcal{P}_1\left(\left(0, \eta S_N^0(t-)\right), dt\right) - \mathcal{P}_3\left(\left(0, \lambda F_N^0(t-)Z_N^0(t-)\right), dt\right), \tag{10}$$

with the convenient initial conditions, where $\mathcal{P}_i, i \in \{1, 2, 3, 4\}$ are independent Poisson point processes on $\mathbb{R}_+^2$ with intensity $ds \otimes dt$, where ds and dt refer to Lebesgue's measure.

An integral representation of these SDEs is given in Section A.1 together with the expression of the previsible increasing processes of the associated martingales.

We will study two regimes of this stochastic model:

- Sub-critical case, when $c_m > 1$, i.e. $N < C_m^N$ 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 filtration common to all our processes, as follows, for $t \geq 0$

$$\mathcal{F}_t = \sigma\left(\mathcal{P}_i(A \times [0, s]) : A \in \mathcal{B}(\mathbb{R}_+), i \in \{1, \cdots, 6\}, s \leq t\right). \tag{11}$$

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.

3 Sub-critical case

In this section we investigate the asymptotic properties of the auxiliary process of Sect. 2.4. 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 $g \in \mathcal{C}_c (\mathbb{R}_+ \times \mathbb{N})$

$$\langle \mu_N, g \rangle \stackrel{\text{def.}}{=} \int_0^{+\infty} g\left(u, F_N^0(u)\right) du. \tag{12}$$

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

Lemma 3 *For $\kappa_0 > 0$ and $\kappa_1 > 0$, let $(Y(t))$ be a birth and death process on $\mathbb{N}$ whose Q-matrix is given by*

$$q(x, x+1) = \kappa_0 \text{ and } q(x, x-1) = \kappa_1 x, \quad x \in \mathbb{N},$$

(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 probability to a constant.
(b) if $Y(0) = 0$, then for any $\delta > 0$, the convergence in distribution of processes

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

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 (2003) for example. Its invariant distribution is Poisson with parameter κ_0 / κ_1.

Proof The first assertion comes directly from Proposition 6.8 of Robert (2003). If $Y(0) = 0$ and, for $p \geq 1$,

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

Proposition 6.10 Robert (2003) gives that, if $\rho \stackrel{\text{def.}}{=} \kappa_0 / \kappa_1$, 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. $\qquad\qquad\square$

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 $(F_N^0(0), S_N^0(0), Z_N^0(0))=(f_N, s_N, z_N)$, 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

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

then the sequence $(\tau_N^0/(\ln N)^2)$ is converging in probability 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 $c_m>1$, there exists $\varepsilon_0>0$ such that, for N sufficiently large, $C_m^N-N>\varepsilon_0 N$. The relations (7) for the transition rates gives that $+1$ jumps of $(F_N^0(t))$ occur at rate less than $N(\beta_m+\eta)$ since $S_N(t)$ is always less than N, and for -1 jumps of $F_N^0(t)$ occur at a rate greater than $\varepsilon_0 N F_N^0(t)$. One can therefore construct a coupling $(F_N^0(t), Y(t))$ such that $Y(0)=F_N^0(0)$ and the relation

$$F_N^0(t) \leq Y(Nt), \qquad \forall t\geq 0, \tag{13}$$

holds almost surely for all $t\geq 0$, where $(Y(t))$ is a process as defined in Lemma 3 with $\kappa_0=\beta_m$ and $\kappa_1=\alpha_m\varepsilon_0$. We conclude the proof by using Lemma 3. $\qquad\square$

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 A.1 of the Appendix.

Proposition 5 (Starting from a Congested State) *If the initial state $(F_N^0(0), S_N^0(0), Z_N^0(0))=(0, s_N, z_N)$ is*

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

and such that $s_0+z_0<1$, then, for the convergence in distribution of processes

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

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

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

See Section A.2 of the appendix.

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 the 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 will remain 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 Sect. 4.1, tightness properties of occupations measures are proved in Sect. 4.2, finally the main convergence results are proved in Sect. 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

$$\left(\overline{X}_N^0(t) \right) = \left(\overline{F}_N^0(t), G_N^0(Nt), Z_N^0(Nt) \right) \text{ with } \left(\overline{F}_N^0(t) \right) \stackrel{\text{def.}}{=} \left(\frac{F_N^0(Nt)}{N} \right). \tag{14}$$

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

$$\left\langle \overline{\Lambda}_N^0, g \right\rangle \stackrel{\text{def.}}{=} \int_{\mathbb{R}_+} g \left(s, \overline{X}_N^0(s) \right) \mathrm{d}s. \tag{15}$$

Note that the, a priori, slow process $(\overline{F}_N^0(t))$ is also included in the definition of the occupation measure $\overline{\Lambda}_N^0$. The reason is that the proof of the tightness of $(\overline{F}_N^0(t))$ (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 $(\overline{F}_N^0(t))$ vanishes does not seem to be straightforward.

The following initial conditions will be assumed,

$$\lim_{N \to +\infty} \frac{F_N^0(0)}{N} = \overline{f}_0 \in (0, 1 - c_m), \ G_N^0(0) = m_0, \ \text{and } Z_N^0(0) = z_0, \tag{16}$$

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

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

As it will be seen in Sect. 4.1 there is no loss of generality to consider these initial conditions.

Before proving the convergence of the sequence of processes $(\overline{X}_N^0(t))$, 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

$$\tau_N(a) \overset{\text{def.}}{=} \inf\left\{t>0 : F_N^0(Nt) \le aN\right\}, \tag{17}$$

and

(a) if $(W(t))$ is a càdlàg process, we denote $(W_N^a(t))=(W(N(t\wedge\tau_N(a))))$;
(b) The "stopped" occupation measure $\overline{\Lambda}_N^{0,a}$ is defined by, if g is non-negative Borelian function on $\mathbb{R}_+^2 \times \mathbb{N}^2$,

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

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

4.1 Technical Lemmas

The two processes $(G_N^0(t)$ and $(Z_N^0(t))$ 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. Using Relations (7), in state $F_N^0=f$, $G_N^0=g$ and $Z_N^0=z$, f, g, $z\in\mathbb{N}$, the jump rates of the process $(G_N^0(t)$ and $(Z_N^0(t))$ are respectively

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

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

$$G_N^0(Nt) \le Y_G(N^2t) \text{ and } Z_N^0(Nt) \le Y_Z(N^2t), \tag{18}$$

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 $Y_G(0)=G_N^0(0)$, resp. $Y_Z(0)=Z_N^0(0)$. It is not difficult, using again Lemma 3, as in Sect. 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 (16) for the initial state can be assumed.

Lemma 9 *Under Conditions (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

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

holds for the convergence in distribution of processes.

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 (18), and Lemma 3. □

4.2 Tightness properties

Throughout this section, conditions (16) and $c_m < 1$ are assumed to hold.

Proposition 10 *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,*

$$\left\langle \overline{\Lambda}_\infty^{0,a}, f \right\rangle = \int_{[0,t_0^a) \times \mathbb{R}_+ \times \mathbb{N}^2} f(s, x, p) \, \pi_s^a(\mathrm{d}x, \mathrm{d}p) \, \mathrm{d}s, \tag{19}$$

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 the space of probability distributions on $\mathbb{R}_+ \times \mathbb{N}^2$.

For an introduction to the convergence in distribution of measure-valued processes, see (Dawson 1993). 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 Diffusions (2000).

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

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

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

$$\mathbb{E}\left(\overline{\Lambda}_N^{0,a}([0, t_0^a] \times [0, 1] \times \mathbb{N} \times [K, +\infty]) \right)$$

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

and, with Relation (18) and Lemma 9, we have

$$
\mathbb{E}\left(\mathbb{1}_{\{\tau_N(a)>t_0^a\}}\int_0^{t_0^a}\mathbb{1}_{\left\{Z_N^{0,a}(s)\geq K\right\}}\,\mathrm{d}s\right)
$$

$$
\leq\int_0^{t_0^a}\mathbb{P}(Y_Z(N^2 s)\geq K)\,\mathrm{d}s=\frac{1}{N^2}\int_0^{N^2 t_0^a}\mathbb{P}(Y_Z(s)\geq K)\,\mathrm{d}s,
$$

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 (1992) shows that the sequence $(\overline{\Lambda}_N^{0,a})$ is tight, and Lemma 1.4 of the same reference gives the representation (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 1993) for example.

Lemma 11 *If $(\overline{\Lambda}_{N_k}^{0,a})$ is a subsequence converging to $\overline{\Lambda}_\infty^{0,a}$ satisfying Relation (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 (1999). 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 (40) of the appendix, we get that, on the event $\{\tau_N(a)>t\}$, the relation

$$
\begin{aligned}
\frac{1}{N^2} f\left(\overline{X}_N^0(t)\right) &= \frac{1}{N^2} f\left(\overline{X}_N^0(0)\right) + \frac{M_{f,N}(t)}{N^2} \\
&+ \lambda \int_0^t \nabla_{-\frac{e_1}{N}-e_3}(f)\left(\overline{X}_N^0(s)\right) \overline{F}_N^0(Ns) Z_N^0(Ns)\, \mathrm{d}s \\
&+ \eta \int_0^t \nabla_{\frac{e_1}{N}+e_3}(f)\left(\overline{X}_N^0(s)\right) \left(1 - \frac{C_m^N}{N} + \frac{G_N^0(Ns)}{N} - \overline{F}_N^0(Ns)\right) \mathrm{d}s \\
&+ \alpha_m \int_0^t \nabla_{-\frac{e_1}{N}-e_2}(f)\left(\overline{X}_N^0(s)\right) G_N^0(Ns)\overline{F}_N^0(Ns)\, \mathrm{d}s \\
&+ \beta_m \int_0^t \nabla_{\frac{e_1}{N}+e_2}(f)\left(\overline{X}_N^0(s)\right) \left(\frac{C_m^N}{N} - \frac{G_N^0(Ns)}{N}\right) \mathrm{d}s \\
&+ \frac{\beta_6}{N} \int_0^t \nabla_{e_3}(f)\left(\overline{X}_N^0(s)\right) \mathrm{d}s + \frac{\delta_6}{N} \int_0^t \nabla_{-e_3}(f)\left(\overline{X}_N^0(s)\right) Z_N^0(Ns)\, \mathrm{d}s \quad (20)
\end{aligned}
$$

holds. The process $(Mf, N(t))$ is a martingale whose previsible increasing process is given by Relation (41). For $i \in \{1, 2, 3\}$, e_i is the ith unit vector of $\mathbb{R}^3$ and, if g is a function on $\mathbb{R}_+ \times \mathbb{N}^2$, for $x, a \in \mathbb{R}_+ \times \mathbb{N}^2$, $\nabla_a(g)(x) = g(x+a) - g(x)$.

Lemma 12 *If f is a continuous bounded function on $\mathbb{R}_+ \times \mathbb{N}$, then the martingale $(M_{f,N}(t), t < t_0^a)$ of Relation (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 (41), the arguments are similar for the others, even easier.

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

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

$\square$

Proposition 13 *If $\overline{\Lambda}_\infty^{0,a}$ is a limiting point of $\overline{\Lambda}_n^{0,a}$ with the representation (19) of Proposition 10, then, if $\pi_t^{1,a} = \pi_t^{0,a}(\cdot, \mathbb{N}^2)$, for any $t < t_0^a$ and any continuous function g on $\mathbb{R}_+ \times \mathbb{N}^2$ we have*

$$
\begin{aligned}
\int_0^t \int g(x, p)\pi_s^a(\mathrm{d}x, \mathrm{d}p)\, \mathrm{d}s = \int_0^t \int_{\mathbb{R}_+} &\mathbb{E}\left[g\left(x, \mathcal{N}_1\left(\left[0, \rho_m \frac{c_m}{x}\right]\right),\right.\right. \\
&\left.\left. \mathcal{N}_2\left(\left[0, \rho_1 \frac{1-c_m-x}{x}\right]\right)\right)\right] \pi_s^{1,a}(\mathrm{d}x)\, \mathrm{d}s, \quad (21)
\end{aligned}
$$

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

$$
\rho_1 = \frac{\eta}{\lambda} \text{ and } \rho_m = \frac{\beta_m}{\alpha_m}. \quad (22)
$$

Relation (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 $(\overline{\Lambda}_{N_k}^0)$ be a subsequence of $(\overline{\Lambda}_N^0)$ converging to some $\overline{\Lambda}_\infty^0$ of the form (19). By letting k go to infinity in Relation (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

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

< holds for all $t \le T$ 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. See Corollary 11.2.5 of Richard (2018) for example. 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 $\kappa_0 = \beta_m c_m$, $\kappa_1 = \alpha_m x$ for $(Y_1(t))$ and $\kappa_0 = \eta(1-c_m-x)$, $\kappa_1 = \lambda x$ for $(Y_2(t))$.

Consequently, for almost all $t \le T$, 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 $\pi_t^{1,a}(dx)$ almost surely, for all functions f_2 with finite support on $\mathbb{N}^2$. Consequently, $\pi^{1,a}(dx)$ almost surely, $\widetilde{\pi}_t^a(dp|x)$ is the invariant distribution associated to the Q-matrix $\Omega[x]$. The proposition is proved. $\qquad\square$

We fix (N_k) an increasing sequence such the sequence $(\overline{\Lambda}_{N_k}^{0,a})$ is converging in distribution to the law of $\overline{\Lambda}_\infty^{0,a}$ with a representation given by Relations (19) and (21).

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 (9) and (10), we have

$$\frac{\widetilde{Z}_N^0(Nt)}{N} = M_{Z,N}(t) + \frac{\widetilde{Z}_N^0(0)}{N} + \beta_6 t - \delta_6 \int_0^t Z_N^0(Ns)\, ds, \tag{23}$$

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

$$\left(\langle M_{Z,N} \rangle(t) \right) = \left(\frac{1}{N} \left(\beta_6 t + \delta_6 \int_0^t Z_N^0(Ns)\, ds \right) \right). \tag{24}$$

Proposition 14 *Under Conditions* (16) *and* $c_m < 1$, *for the convergence in distribution of processes*

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

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 (23) and (24), and of Doob's Inequality. For $0 \leq s \leq t$, the coupling (18) and Cauchy–Schwartz' Inequality give

$$\mathbb{E}\left(\left(\mathbb{1}_{\{\tau_N(a) > t\}} \int_s^t Z_{N_k}^0(Ns)\, ds \right)^2 \right) \leq (t - s)\mathbb{E}\left(\mathbb{1}_{\{\tau_N(a) > t\}} \int_s^t Z_{N_k}^0(s)^2\, ds \right)$$

$$\leq (t - s)\mathbb{E}\left(\int_s^t Y_Z(Ns)^2\, ds \right)$$

$$\leq (t - s)^2 \sup_{u \geq 0} \mathbb{E}\left(Y_Z(u)^2 \right).$$

We now use the Kolmogorov-Čentsov's criterion, see Theorem 2.8 and Problem 4.11, page 64 of Karatzas and Shreve (1998) and Lemma 9 to show that the sequence of stochastic processes

$$\left(\int_0^t Z_{N_k}^0(Ns) \, \mathrm{d}s, \, t < t_0^a \right)$$

is tight for the convergence in distribution.

Lemma 11 and Relation (21) give the convergence in distribution of processes

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

By using again Relation (18), we have

$$\mathbb{E} \left(\int_0^{t_0^a} Z_{N_k}(N_k s) \mathbb{1}_{\left\{ Z_{N_k}^0(N_k s) \geq K \right\}} \, \mathrm{d}s \right) \leq \mathbb{E} \left(\int_0^{t_0^a} Y_Z(N_k^2 s) \mathbb{1}_{\left\{ Y_Z(N_k^2 s) \geq K \right\}} \, \mathrm{d}s \right)$$

and the convergence in distribution of processes 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

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

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

The proposition is proved.   □

Relation (23) therefore shows that, on the time interval $I_a = [0, t_0^a)$, the sequence of processes

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

is converging in distribution. Since

$$\left(\frac{\widetilde{Z}_{N_k}^0(N_k t)}{N_k} \right) = \left(1 - \frac{F_{N_k}^0(N_k t)}{N_k} - \frac{C_m^{N_k}}{N_k} + \frac{G_{N_k}^0(N_k t)}{N_k} + \frac{Z_{N_k}^0(N_k t)}{N_k} \right)$$

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

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

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

Relation (23) gives that, on the time interval I_a and under the initial conditions (16), then the sequence of processes $(\overline{F}_{N_k}^0(t))$ of Relation (14) is converging in distribution to $(\overline{f}(t))$ such that

$$
\begin{aligned}
1 - \overline{f}(t) &= 1 - \overline{f}_0 + \beta_6 t - \delta_6 \rho_1 \int_0^t \int_{\mathbb{R}_+} \frac{1 - c_m - x}{x} \pi_s^{1,a}(\mathrm{d}x)\,\mathrm{d}s \\
&= 1 - \overline{f}_0 + \beta_6 t - \delta_6 \rho_1 \int_0^t \frac{1 - c_m - \overline{f}(s)}{\overline{f}(s)}\,\mathrm{d}s
\end{aligned}
\tag{25}
$$

with Proposition 14 and Notation (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 of processes,*

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

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

$$\frac{\mathrm{d}\overline{f}}{\mathrm{d}t}(t) = -\delta_6(\rho_6 + \rho_1) + \delta_6\rho_1(1 - c_m)\frac{1}{\overline{f}(t)},\tag{26}$$

where ρ_6 and ρ_1 are defined by Relation (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*

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

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 (26) with $\overline{f}(0) = \overline{f}_0$.

5 Exponential phase

Throughout this section, Conditions (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 Sect. 2.2.

Otherwise, if $F_N(t)=0$, there is a total of at least $A_N \overset{\text{def.}}{=} N - C_{r,1}^N - \cdots - C_{r,J}^N - J$ 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

$$\frac{1}{\alpha_{r,j}} + \frac{J}{\delta N_0} < \frac{1}{\beta_{r,j}}, \qquad \forall 1 \leq j \leq J \text{ with } \delta = \min(\eta, \beta_m). \tag{27}$$

We are interested in the behavior of $(Q_j^N(t)) \overset{\text{def.}}{=} (C_{r,j}^N(t) - R_j^N(t))$, $1 \leq j \leq J$, 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 $\beta_{r,j}$. If $R_j^N(t) \neq 0$, the rate of jumps of size $+1$ of $Q_j^N(t)$ is $\beta_{r,j}$.
- The service of a customer is expressed as the sum of two independent exponential random variables with respective parameters $\alpha_{r,j}$ and $\delta N_0 / J$.

To summarize, the instant jumps $+1$ of $(Q_j(t))$ are a subset of the $+1$ jumps of $(\overline{Q}_j(t))$ and the time between two jumps -1 is greater for $(\overline{Q}_j(t))$.

The process $(\overline{Q}_j(t))$ is the process of the number of customers of an $M/G/1$ queue, see Chapter 2 of Robert (2003). 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 (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* (3) *and* (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*

$$\tau_N^e \overset{\text{def.}}{=} \inf\{t > 0 : R_{N,j}(t) = C_{r,j}^N, \forall 1 \leq j \leq J\}, \tag{28}$$

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

$$\mathbb{P}_{((1,0))} \left(\sup_{0 \le t \le T} \max_{1 \le j \le J} C_{r,j}^N - R_j^N(t) \ge K \right) \le \varepsilon \tag{29}$$

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 (4) holds.

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

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

if g is a continuous function with compact support on $\mathbb{R}_+ \times \mathbb{N}$, where A_N defined by Relation (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 (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 (29), we know that for K sufficiently large, the probability of the event

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

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* (3) *and* (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,*

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

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

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

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 $a > 0$,

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

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 (2003).

6 Stationary phase

Conditions (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 Sect. 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

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

hold. See Chapter 5 of Robert (2003) 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 Robert (2003)

Proposition 20 *Under Conditions* (3) *and* (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 *defined by*

$$\tau_N^s \overset{def.}{=} \inf\{t>0 : R_{N,j}(t) = 0, \forall 1 \le j \le J\}, \tag{30}$$

is almost surely finite and

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

Lemma 21 *Under Condition* (5) *then, for any* $K>0$*,*

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

Proof This is a simple consequence of the independence of the $(Q_j(t))$ and of Proposition 5.11 of Robert (2003). $\qquad\square$

The above result shows that few polymerases are in transcription of an rRNA, hence the results of Sect. 4 on the auxiliary process can be used, in particular Theorem 15.

Theorem 22 (Asymptotic Behavior in Stationary Phase) *Under Conditions* (3) *and* (5)*, and the initial state such that*

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

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

$$\frac{d\overline{f}}{dt}(t) = -\delta_6(\rho_6+\rho_1)+\delta_6\rho_1(1-c_m)\frac{1}{\overline{f}(t)}, \tag{31}$$

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,

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

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.

Declarations

Conflict of interest The authors state that they have no conflicts of interest regarding the publication of this article.

Appendix A: 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 Sect. 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 Sect. 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 (12). For $K>0$, with the same notations as in the proof of Lemma 4, Relation (13) gives the inequality

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

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 (1992) 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 $\mathbb{N}$.

A.1 Proof of Proposition 5

By integrating Relations (9) and (10), we obtain the identities, for $t \geq 0$,

$$S_N^0(t) = S_N^0(0) + M_6^N(t) - \eta \int_0^t S_N^0(s)\,\mathrm{d}s + \lambda \int_0^t F_N^0(s)Z_N^0(s)\,\mathrm{d}s, \qquad (32)$$

$$Z_N^0(t) = Z_N^0(0) + M_Z^N(t) + \beta_6 t - \delta_6 \int_0^t Z_N^0(s)\,\mathrm{d}s$$

$$+ \eta \int_0^t S_N^0(s)\,\mathrm{d}s - \lambda \int_0^t F_N^0(s)Z_N^0(s)\,\mathrm{d}s, \qquad (33)$$

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

$$\left\langle M_6^N \right\rangle(t) = \eta \int_0^t S_N^0(s)\,\mathrm{d}s + \lambda \int_0^t F_N^0(s)Z_N^0(s)\,\mathrm{d}s, \qquad (34)$$

$$\left\langle M_Z^N \right\rangle(t) = \beta_6 t + \delta_6 \int_0^t Z_N^0(s)\,\mathrm{d}s + \eta \int_0^t S_N^0(s)\,\mathrm{d}s + \lambda \int_0^t F_N^0(s)Z_N^0(s)\,\mathrm{d}s. \qquad (35)$$

Relations (34) and (35), Relation (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 (10). Relations (32) and (33), and the criterion of the modulus of continuity, see (Billingsley 1999), 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 $f \in C_c \left(\mathbb{N} \times \mathbb{R}_+^2 \right)$, Relation (8) gives the identity

$$
f \left(F_{N_k}^0(t), \frac{S_{N_k}^0(t)}{N_k}, \frac{Z_{N_k}^0(t)}{N_k} \right) = f \left(f_{N_k}, s_{N_k}, z_{N_k} \right) + M_f^{N_k}(t)
$$

$$
+ \beta_m \int_0^t \nabla_{e_1}(f) \left(X_{N_k}^0(s) \right) \left(N_k - F_{N_k}^0(s) - S_{N_k}^0(s) \right) ds
$$

$$
+ \alpha_m \int_0^t \nabla_{-e_1}(f) \left(X_{N_k}^0(s) \right) \left(C_m^{N_k} - N_k + F_{N_k}^0(s) + S_{N_k}^0(s) \right) F_{N_k}^0(s) ds
$$

$$
+ \lambda \int_0^t \nabla_{-e_1 + \frac{e_2}{N_k} - \frac{e_3}{N_k}}(f) \left(X_{N_k}^0(s) \right) F_{N_k}^0(s) Z_{N_k}^0(s) ds
$$

$$
+ \eta \int_0^t \nabla_{e_1 - \frac{e_2}{N_k} + \frac{e_3}{N_k}}(f) \left(X_{N_k}^0(s) \right) S_{N_k}^0(s) ds
$$

$$
+ \beta_6 \int_0^t \nabla_{\frac{e_3}{N_k}}(f) \left(X_{N_k}^0(s) \right) ds + \delta_6 \int_0^t \nabla_{-\frac{e_3}{N_k}}(f) \left(X_{N_k}^0(s) \right) Z_{N_k}^0(s) ds,
$$

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

$$
\int_0^t \nabla_{e_1}(f)(x, s(u), z(u)) \, (\beta_m - (\beta_m - \eta) \, s(u)) \, \pi_u(dx) \, du
$$

$$
+ \int_0^t \nabla_{-e_1}(f)(x, s(u), z(u)) \, (\alpha_m (c_m - 1 + s(u)) + \lambda z(u)) \, x \pi_u(dx) \, du = 0,
$$

and therefore

$$
\int_0^t \int_{\mathbb{N}} \Omega_{s(u), z(u)}(g)(x) \pi_u(dx) \, du = 0, \tag{36}
$$

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

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

$\Omega_{s,z}$ 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 (36) and with the same methods as in Sect. 4, we obtain that, almost surely,

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

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 Sect. 4, for $T \geq 0$ such that $s(t)+z(t)<1$ holds for all $t \leq T$, we obtain that the identities

$$s(t) = s_0 - \eta \int_0^t s(u)\, du + \lambda \int_0^t z(u) \frac{\beta_m - (\beta_m - \eta)\, s(u)}{\alpha_m(c_m - 1 + s(u)) + \lambda z(u)}\, du, \qquad (37)$$

$$s(t) + z(t) = s_0 + z_0 - \delta_6 \int_0^t z(u)\, du, \qquad (38)$$

hold almost surely, for $t \leq T$. From Relation (38) we obtain that $(s(t)+z(t))$ is a non-increasing function, hence $0 \leq s(t)+z(t) \leq s_0+z_0 < 1$, 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 (38) shows that $(z(t))$ is an integrable function on $\mathbb{R}_+$, hence the second integral of the right-hand side of Relation (37) is converging as t goes to infinity. Since $(s(t))$ is bounded, the other integral of (37) also converges at infinity. We conclude that $(s(t))$ has a limit at infinity and therefore also $(z(t))$. By dividing Relation (38) by t and let t go to infinity, we obtain that the limit of $(z(t))$ is 0. With a similar method in Relation (37), we get that both limits are zero. Proposition 5 is proved.

A.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 $(Y_N(t)) \overset{\text{def.}}{=} (S_N(t), Z_N(t))$ 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 $t \geq 0$,

$$\mathbb{P}\left(A_1^N \geq t \mid (F_N^0(s))\right) = \exp\left(-\lambda z \int_0^t F_N^0(s)\, ds\right); \qquad (39)$$

(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 (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

$$\mathbb{P}_{(s,z)}\left(Y_N\left(t_1^N\right) = (s+1, z-1), t_1^N \geq t\right) = \mathbb{P}\left(A_1^N \geq t, A_1^N \leq A_2 \wedge A_3 \wedge A_4\right)$$

$$= \mathbb{E}\left(\mathbb{1}_{\{A_1^N \geq t\}} \exp\left(-(\eta s + \beta_6 + \delta_6 z) A_1^N\right)\right),$$

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 1999).

Super-critical case

The assumption $c_m < 1$ holds throughout this section. Technical results used in Sect. 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 (8), (9), and (10) give directly the relations

$$f\left(\overline{X}_N^0(t)\right) = f\left(\overline{X}_N^0(0)\right) + M_{f,N}(t)$$

$$+ \lambda N \int_0^t \nabla_{-\frac{e_1}{N} - e_3}(f)\left(\overline{X}_N^0(s)\right) F_N^0(Ns) Z_N^0(Ns)\, \mathrm{d}s$$

$$+ \eta N \int_0^t \nabla_{\frac{e_1}{N} + e_3}(f)\left(\overline{X}_N^0(s)\right) \left(N - C_m^N + G_N^0(Ns) - F_N^0(Ns)\right) \mathrm{d}s$$

$$+ \alpha_m N \int_0^t \nabla_{-\frac{e_1}{N} - e_2}(f)\left(\overline{X}_N^0(s)\right) G_N^0(Ns) F_N^0(Ns)\, \mathrm{d}s$$

$$+ \beta_m N \int_0^t \nabla_{\frac{e_1}{N} + e_2}(f)\left(\overline{X}_N^0(s)\right) \left(C_m^N - G_N^0(Ns)\right) \mathrm{d}s$$

$$+ \beta_6 N \int_0^t \nabla_{e_3}(f)\left(\overline{X}_N^0(s)\right) \mathrm{d}s + \delta_6 N \int_0^t \nabla_{-e_3}(f)\left(\overline{X}_N^0(s)\right) Z_N^0(Ns)\, \mathrm{d}s, \quad (40)$$

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

$$
\begin{aligned}
\langle M_{f,N} \rangle (t) = {} & \lambda N \int_0^t \nabla_{-\frac{e_1}{N}-e_3}(f) \left(\overline{X}_N^0(s)\right)^2 F_N^0(Ns) Z_N^0(Ns)\, \mathrm{d}s \\
& + \eta N \int_0^t \nabla_{\frac{e_1}{N}+e_3}(f) \left(\overline{X}_N^0(s)\right)^2 (N - C_m^N + G_N^0(Ns) - F_N^0(Ns))\, \mathrm{d}s \\
& + \alpha_m N \int_0^t \nabla_{-\frac{e_1}{N}-e_2}(f) \left(\overline{X}_N^0(s)\right)^2 G_N^0(Ns) F_N^0(Ns)\, \mathrm{d}s \\
& + \beta_m N \int_0^t \nabla_{\frac{e_1}{N}+e_2}(f) \left(\overline{X}_N^0(s)\right)^2 \left(C_m^N - G_N^0(Ns)\right)\, \mathrm{d}s \\
& + \beta_6 N \int_0^t \nabla_{e_3}(f) \left(\overline{X}_N^0(s)\right)^2 \mathrm{d}s \\
& + \delta_6 N \int_0^t \nabla_{-e_3}(f) \left(\overline{X}_N^0(s)\right)^2 Z_N^0(Ns)\, \mathrm{d}s.
\end{aligned}
\tag{41}
$$

References

Anderson DF, Kurtz TG (2015) Stochastic analysis of biochemical systems, Mathematical Biosciences Institute Lecture Series, Springer Publishing Company, Incorporated

Baker C, Jia T, Kulkarni RV (2012) Stochastic modeling of regulation of gene expression by multiple small RNAs. Phys Rev E 85(6):061915

Bakshi S, Siryaporn A, Goulian M, Weisshaar JC (2012) Superresolution imaging of ribosomes and RNA polymerase in live escherichia coli cells. Mol Microbiol 85(1):21–38

Berg OG (1978) A model for the statistical fluctuations of protein numbers in a microbial population. J Theor Biol 71(4):587–603 (**(eng)**)

Billingsley P (1999) Convergence of probability measures, 2nd edn. Wiley series in probability and statistics: probability and statistics. Wiley, New York, A Wiley-Interscience Publication

Bremer H, Dennis PP (2008) Modulation of chemical composition and other parameters of the cell at different exponential growth rates. EcoSal Plus 3(1)

Britten RJ, Davidson EH (1969) Gene regulation for higher cells: a theory. Science 165(3891):349–357

Cavanagh AT, Wassarman KM (2014) 6s RNA, a global regulator of transcription in *Escherichia coli*, bacillus subtilis, and beyond. Annu Rev Microbiol 68:45–60

Dawson DA (1993) Measure-valued Markov processes, École d'Été de Probabilités de Saint-Flour XXI–1991. Lecture Notes in Math., Springer, Berlin 1541:1–260

Dessalles R, Fromion V, Robert P (2017) A stochastic analysis of autoregulation of gene expression. J Math Biol 75(5):1253–1283

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297(5584):1183–1186

Fischer S, Dinh M, Henry V, Robert P, Goelzer A, Fromion V (2021) BiPSim: a flexible and generic stochastic simulator for polymerization processes. Sci Rep 11(1)

Fromion V, Robert P, Zaherddine J (2022) Stochastic models of regulation of transcription in biological cells

Geissen R, Steuten B, Polen T (2010) E. coli 6s RNA: a universal transcriptional regulator within the centre of growth adaptation. RNA Biol 7(5):564–568

Gottesman S, McCullen CA, Guillier M, Vanderpool CK, Majdalani N, Benhammou J, Thompson KM, FitzGerald PC, Sowa NA, FitzGerald DJ (2006) Small RNA regulators and the bacterial response to

stress, Cold Spring Harbor symposia on quantitative biology, vol 71, Cold Spring Harbor Laboratory Press, pp 1–11

Hindley JT (1967) Fractionation of 32p-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. J Mol Biol 30(1):125–136

Karatzas I, Shreve SE (1998) Brownian motion and stochastic calculus, 2 edn. Graduate Texts in Mathematics, vol 113. Springer, New York

Karp PD, Kothari A, Fulcher CA, Keseler IM, Latendresse M, Krummenacker M, Subhraveti P, Ong Q, Billington R, Caspi R, Paley SM, Ong WK, Midford PE (2019) The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform 20(4):1085–1093

Kumar N, Zarringhalam K, Kulkarni RV (2018) Stochastic modeling of gene regulation by noncoding small RNAs in the strong interaction limit. Biophys J 114(11):2530–2539

Kurtz TG (1992) Averaging for martingale problems and stochastic approximation. Appl Stoch Anal 177:186–209

Mackey MC , Moisés S, Marta T-K, Zeron ES (2016) Simple mathematical models of gene regulatory dynamics, Lecture Notes on Mathematical Modelling in the Life Sciences. Springer, Cham

Mehta P, Goyal S, Wingreen NS (2008) A quantitative comparison of sRNA-based and protein-based gene regulation. Mol Syst Biol 4:221

Milo R, Jorgensen P, Moran U, Weber GM, Springer M (2010) Bionumbers: the database of key numbers in molecular and cell biology. Nucleic Acids Res 38:750–753

Mitarai N, Benjamin J-AM, Krishna S, Semsey S, Csiszovszki Z, Massé E, Sneppen K (2009) Dynamic features of gene expression control by small regulatory RNAs. Proc Natl Acad Sci 106(26):10655–10659

Neidhardt FC, Umbarger HE (1996) Chemical composition of Escherichia coli, and Salmonella: cellular and molecular biology, 2 edn. ASM Press

Nitzan M, Wassarman KM, Biham O, Margalit H (2014) Global regulation of transcription by a small RNA: a quantitative view. Biophys J 106(5):1205–1214

Paul C (2014) Bressloff. Stochastic processes in cell biology, Interdisciplinary applied mathematics. Springer

Paulsson J (2005) Models of stochastic gene expression. Phys Life Rev 2(2):157–175

Platini T, Jia T, Kulkarni RV (2011) Regulation by small RNAs via coupled degradation: mean-field and variational approaches. Phys Rev E 84(2):021928

Richard M (2018) Dudley. CRC Press, Real analysis and probability

Rigney D, Schieve W (1977) Stochastic model of linear, continuous protein synthesis in bacterial populations. J Theor Biol 69(4):761–766

Robert P (2003) Stochastic networks and queues, stochastic modelling and applied probability series, vol 52. Springer, New-York

Rogers LCG, David W (2000) Markov Processes and Martingales: Volume 2, Itô Calculus, Cambridge University Press, September (English)

Rogers LCG, Williams D (1994) Diffusions, Markov processes, and martingales, vol 1: Foundations, 2nd edn. Wiley, Chichester

Steuten B, Schneider S, Wagner R (2014) 6s RNA: recent answers-future questions. Mol Microbiol 91(4):641–648

Wassarman KM (2018) 6S RNA, a global regulator of transcription. Microbiol Spectrum 6(3)

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R, CSHLP I (2007) Molecular biology of the gene, 6th ed., Pearson/Benjamin Cummings; Cold Spring Harbor Laboratory Press, San Francisco; Cold Spring Harbor, NY

Zamore PD, Haley B (2005) Ribo-gnome: the big world of small RNAs. Science 309(5740):1519–1524