Comparison of modeling options for the mRNA Life cycle

*Ivan Mura**
Jorge Enrique González M.**

ABSTRACT

This paper deals with the predictive modeling of cellular processes at the molecular level. We consider the first product of gene expression, the MRNA, and explore different options for the modeling of its life-cycle, from the simplest to the most accurate and complex. For each of the modeling scenarios involved, we implement and analyze a continuous-deterministic and a discrete-stochastic version of the model of MRNA life-cycle. We aim at comparing the representation capabilities of the different models and their implementations, determining which modeling options are the most appropriate ones at the time of matching the available experimental observations. The results of our study indicate that the simple modeling options, which disregard the regulation of gene transcription and MRNA degradation processes, may be inadequate to represent various biological phenomena, in particular those related to the periodic expressions of genes involved in cyclic behaviors.

PALABRAS CLAVE

MRNA life- cycle, cellular processes, molecular levels, gene, modelling, stochastic, degradation, cyclic behavior.

_____

* Ph.D. en Ingeniería Electrónica, Informática y de Telecomunicaciones, Universidad de Pisa. Magister en Ciencias de la Información, Universidad de Pisa y en Information Technology Project Management, George Washington University School of Business.
** Pregrado en Ingeniería de Sistemas, Universidad EAN.
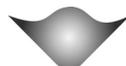
Comparacion de opciones de modelado para el ciclo de vida del ARNm

RESUMEN

Este artículo describe modelos predictivos de procesos celulares a nivel molecular. Se considera el primer producto de expresión genética, MRNA, y explora las diferentes opciones que existen para modelar su ciclo de vida, desde el más sencillo hasta el más complejo y exacto. Para cada uno de los escenarios involucrados, se implementó y analizó una versión determinante, continua, discreta y estocástica del modelo de ciclo de vida MRNA. El objetivo de este trabajo es comparar las capacidades de representación de los diferentes modelos y sus implementaciones para así determinar qué modelos resultan ser los más apropiados al explicar los resultados experimentales obtenidos. Los resultados demostraron que los modelos simples que no toman en cuenta la regulación de la transcripción genética y los procesos de degradación del MRNA pueden ser no apropiados para representar los diferentes fenómenos biológicos, particularmente aquellos relacionados con la periodicidad de los genes involucrados en sus comportamientos cíclicos.

KEY WORDS

Ciclo de vida MRNA, procesos celulares, niveles moleculares, genes, modelaje, estocástico, degradación, comportamiento cíclico.

Diagnostic de la consommation d'énergie électrique d'un iComparaison des modélisations potentielles du cycle de vie mRNA

RÉSUMÉ

Cet article traite de la modélisation prédictive des processus cellulaires au niveau moléculaire. Nous abordons le premier produit de l'expression génique, le mRNA et explorons les différentes possibilités de modélisation de son cycle de vie, de la plus simple à la plus précise et complexe. Pour chacun des scénarios de modélisation étudiés, nous mettons en place et analysons une version continue et déterministe – stochastiques - du modèle de cycle de vie du mRNA. Notre but est de comparer les capacités de représentation des différents modèles et de leur mise en œuvre pour déterminer les modélisations les plus appropriées lors de l'identification des observations expérimentales disponibles. Les résultats de notre analyse indiquent que les modèles les plus simples qui ne tiennent pas compte de la régulation de la transcription génétique et des processus de dégradation des mRNA peuvent être insuffisants pour représenter de manière adéquate les différents phénomènes biologiques, notamment ceux liés aux expressions périodiques de gènes impliqués dans les comportements cycliques.

Mots-clés

Lifecycle ARNm, les processus cellulaires, les niveaux moléculaires, les gènes, modélisation, stochastiques, dégradation, comportement cyclique.

Comparação das opções de modelagem para o ciclo de vida do mRNA

Resumo

Este artigo descreve os modelos preditivos de processos celulares a um nível molecular. É considerado o primeiro produto de expressão genética, MRNA, e explora as diferentes opções existentes para modelar o seu ciclo de vida, desde o mais simples ao mais complexo e preciso. Para cada um dos cenários envolvidos, foi implementado e testado uma versão determinante, contínua, discreto e estocástica do modelo do ciclo de vida MRNA. O objetivo deste trabalho é comparar as capacidades de representação dos modelos e suas implementações para determinar quais os modelos parecem ser os mais adequados para explicar os resultados experimentais obtidos. Os resultados mostraram que os modelos simples que não levam em conta a regulação da transcrição de genes e processos de degradação do MRNA podem não ser apropriados para representar os diferentes fenômenos biológicos, em particular aqueles relacionados com a frequência dos genes envolvidos em seu comportamento cíclico.

Palavras-chave

Ciclo de vida do MRNA, processos celulares, níveis moleculares, genes, modelagem, estocástico, degradação, comportamentos cíclicos.

1. Introduction

Biology is living an era of outstanding technological progresses, which are continuously enhancing our ability to decipher the intricate behavior of living systems. The advances in microscopy, the automation of cellular extract analysis, and the improvements in our ability to modify the genomes of organisms allow uncovering finer and finer details of the biochemistry of the cell.

It is fair to say that biology is moving towards a more exact science, after many centuries of pure empirical observation. On one hand, the insights we are getting provide for more accurate analyses and for deepening the granularity of what is observable, but on the other hand, there is a manifest lack of techniques and tools to integrate and make sense of the huge amount of data that is nowadays available. Bioinformatics can help to identify correlations in huge sets of data, providing for the definition of testable hypotheses; still, it is missing the ability to really integrate the disparate sources of data into coherent models of knowledge.

Mathematical modeling of living systems at the molecular level, which is often referred as computational or systems biology (Kitano, 2002), allow for the creation of modeling abstractions that are able to encapsulate information coming from different levels of the observation, such as genetics, transcriptomics, proteomics and metabolomics. Models of the molecular networks that determine the phenotypes of living systems can be built in a way that are amenable to simulate the future behaviors of systems, starting from an initially known state. Moreover, they allow predicting what the behavior will be under the influences of various types of perturbations, such as genetic mutations or adverse conditions of the environment, which may be hardly reproducible experimentally (Mura, 2013-a).

In this work, we consider the modeling of a basic building block in molecular networks, i.e. the processes at the basis of the genetic expression (Alberts et al., 2002). Genes are expressed through a

transcription process, which produces copies of the genetic information in the form of mRNA molecules. The mRNA molecules are then transcribed into proteins, which are the machines that carry out all the functions in a living organism. The level of the single proteins is what determines the phenotype of a cell. Starting from the same copy of the Deoxyribonucleic Acid (DNA) that is present in all the cells of a multicellular organism, the protein levels determine the appearance and the function of the cell. A tightly controlled gene expression creates the vast repertoire of cellular types and functions. Therefore, a careful representation of the mRNA life cycle is important to set up molecular models that are representative of the biological phenomena of interest.

In the literature, the mRNA life cycle is often modeled in a very simplified way, disregarding the regulations that affect its production and degradation. These simplifications may lead to inconsistent predictions and may force the modeler to introduce adjustments in the kinetic parameters of the model that have no biological meaning.

We approach in this paper a comparison of modeling options for the mRNA life cycle, which is guided by the comparison with the types of available experimental data. We consider four different modeling scenarios, which are generated by the incremental inclusion of more and more refined information about the processes that regulate gene expression. Each of the four modeling options is furthermore explored in its continuous-deterministic and discrete-stochastic implementation, which are two main approaches used at the time of creating models that can be analyzed in-silico, through simulation (Mura, 2013-b).

We analyze the output of model simulation with respect to the ability of the various options to match experimental observations, starting with the coarsest type of experimental results, such as averages of molecular abundances taken from populations of cells, and then moving towards more precise data obtained at the single cell level. The results of our analyses indicate that simplistic modeling assumptions, such as the absence or the abstraction of regulatory phenomena, may lead to poor models, which are unable to account for the complex behaviors of cells.
This paper is organized as follows. In section 2 we present an overview of the approaches to the predictive modeling of living systems, introducing

the continuous-deterministic and the discrete-stochastic paradigms, and presenting the two modeling tools that we will be using to implement and simulate the four modeling options. Then, in section 3 we provide a concise introduction to the biology of gene expression and mRNA life cycle. Section 4 is devoted to the definition of the four modeling options, which will be presented by using the abstract language of chemical reactions. Each model variant is implemented and evaluated in section 5, which explores the ability of the models in matching the different types of experimental observations. The comparative analyses of models capabilities and the final remarks are provided in section 6.

2. Predictive modeling and simulation of molecular processes

In this work, we will be defining models of the molecular processes that affect mRNA life cycle. Our modeling is focused on the representation of the dynamically occurring events that change mRNA abundance, such as transcription and degradation. The class of models we will be considering encapsulates enough qualitative and quantitative information to allow determining the future states of the system from an initial one.

To describe models at the most abstract level, we will be using the language of chemical reactions. This choice has been made because chemical reactions do not force the selection of any model analysis paradigm; rather, they convey the minimal amount of information necessary to later implement the models in a modeling and simulation tool of choice. In order to understand and compare the capabilities of the models that will be defined, we will use two widely used modeling and simulation tools, namely COPASI and Möbius.

2.1 COPASI

COPASI is a modeling and simulation tool for biological systems that has been developed from a collaborative research project between the Virginia Bioinformatics Institute, USA, the University of Heidelberg in Germany, and the University of Manchester in the United Kingdom. The features that COPASI offers to the modeler are encapsulated into a comprehensive graphical user interface, through which the user is able to:

- Define the models in terms of biochemical reactions.

- Define the kinetic features of the reactions, i.e. their rate of occurrence, as simple constants or more complex, state-dependent functions.

- Evaluate the steady state of the species abundances and their transient dynamical evolution over time from an initial state of the model.

Models defined in terms of reactions and interpreted as continuous – deterministic representations of the biological system. They are analyzed and simulated through the algorithm for the treatment of systems of ordinary differential equations.

2.2 Möbius

Möbius is a modeling and simulation tool whose development began in the 80's at the University of Illinois at Urbana-Champaign, USA. This software has been used in many research studies and the algorithms it implements have therefore been tested extensively and are very reliable. Möbius is based on the modeling technique of petri nets (Mura, 2010), a graphical modeling formalism that consists of only four elements:

- Places, depicted as empty circles, representing the state of the modeled entities.

- Tokens, graphically represented as black dots (omitted when their number is too high for a proper graphical display), which are contained into places and model the counts of entities in the various states.

- Transitions, depicted as bars that model events and whose firing (occurrence) moves tokens among one place so to represent the state-change of entities.

- Oriented arcs, which connect places to transitions and transitions to places (but not places to places nor transitions to transitions), and represent the ways for token flow.

Möbius is endowed a graphical user interface to define models. The times of occurrence of the events are associated with transitions, and may be selected from various distributions. For the purpose of this study, we will be considering occurrence times of events that follow a negative exponential distribution. This means that Möbius models will always be discrete and stochastic. The defined models are analyzed through discrete event simulation, and the statistical analysis of simulation results is automatically performed by Möbius.

3. The life cycle of mRNA

The mRNAs are sequences of nucleotides, which are produced by a process of transcription of regions of the DNA, called genes. The function of mRNA molecules is to carry DNA genetic information to the ribosome, where proteins are assembled. This protein production process is called translation. The two processes of transcription and translation form what is called the basic paradigm of biology that occurs in all living beings:

DNA → mRNA → Protein

The life of an mRNA molecule begins with transcription and ultimately ends in degradation. In the course of its life, however, mRNA is examined,

modified in various ways, transported and translated into proteins before being eventually eliminated by the cellular degradation machinery. All these processes are performed by proteins and non-coding RNAs whose complex interplay in the cell contributes to determining the proteome changes and the phenotype of cells (Denti, 2013). We shall now provide some details of the transcription, translation and degradation processes of mRNA molecules.

3.1 Transcription

A gene is a DNA sequence that provides the basic information to assembly molecules that are used for the purpose of our body functioning, for instance proteins, enzymes, and non-coding mRNAs. Genes are located on DNA aggregates called chromosomes. The genes (the genotype) have the instructions for creating different kinds of proteins, whose interaction ultimately determine our morphology and physiology (the phenotype). This process is divided into three steps: initiation, elongation and termination, which are described below.

3.1.1 Initiation

The process of transcription is initiated by a set of complex rearrangements of the tightly folded DNA structure that exposes a sequence of the polymer to the access of the molecule RNA polymerase and of other necessary co-transcription factors. Upstream to the protein coding sequence of the gene, there is a region of DNA called promoter, which is where the polymerase and the transcription factors bind. The binding of the transcription factors facilitates RNA polymerase recruitment to the gene. Once the RNA polymerase is bound to the gene, it opens the double strand and begins reading one of the filaments of the DNA chain. This piece is used as a template for RNA polymerase to incorporate nucleotides.

3.1.2 Elongation

The RNA polymerase reads one after the other the nucleotides of the gene, and constructs the mRNA molecule by adding to it, one by one, a nucleotide that matches the one being read on the DNA strand. The

mRNA molecule is a copy of the DNA, except for the substitution of the thymine nucleotide (T) found in DNA with the uracil nucleotide (U) that is used to build mRNA chains. After each transcription step, the RNA polymerase is able to slide along the gene by a sequence of GTP-GTP assisted reactions that lead to allosteric transformation of the molecule and permit it to move.

3.1.3 Termination

This stage is the final one in the transcription of the mRNA. The mRNA polymerase finds a sequence of nucleotides on the gene that signals the end of the protein-coding region. When this sequence is found, the complex RNA polymerase gene becomes unstable. The RNA polymerase detaches from the DNA and the nascent mRNA is cleaved and left free in the cell.

3.2 Translation

This process transforms a molecule of mRNA into the chain of amino acids that form a protein. This process takes place into the cell cytoplasm (in eukaryotes), and is executed by small organelles known as ribosomes. Ribosomes are distributed throughout the cytoplasm of the cell, with the typical eukaryotic cell containing millions of them. They are found in higher concentration along the walls of the endoplasmic reticulum, the region where most of protein synthesis takes place. In the translation process, the message inside the mRNA is read by the ribosome, and interpreted according to the so-called genetic code, to assembly protein molecules. The genetic code is a mapping between the possible sequences of three mRNA nucleotides, called codons, and the 20 amino acids that constitute protein building blocks. There are a total of 64 possible codons, which results from the number of ways it is possible to select 3 objects with repetitions from a set of 4 distinct ones. Since the number of possible codons is higher than the number of amino acids, the genetic code is redundant, meaning that more codons are assigned to the same amino acid. Moreover, out of the 64 codons, 61 of these are used in the genetic code to point to an amino acid, and the remaining 3 are control sequences

that are used in the translation process. The translation process is divided into the following three phases:

3.2.1 Initiation

The synthesis of a protein molecule begins with the binding of the ribosomal translation machinery with the mRNA template. The ribosome molecule consists of a small and a large sub-unit, which before initiation of translation are found in an open configuration. The ribosome contains one specific binding site for the mRNA molecules, and upon binding, the conformation of the small and large sub-units changes to a closed form that tightly engulfs the mRNA. A special sequence of nucleotides is found on the mRNA, the AUG codon, which signals where the ribosome has to align and initiate the translation of the mRNA into a protein.

3.2.2 Elongation

The elongation is a sequential process by which codons are inspected one at the time. Based on the codon sequence, the genetic code identifies the amino acid to be added to the protein. Each amino acid is carried to the site of translation by a specific carrying protein called tRNA. Each tRNA has a binding site complementary to the codon, which allows it to form a complex on the mRNA. Once the mRNA-tRNA complex is formed, the enzyme aminoacyl-tRNA detaches the amino acid from the tRNA and attaches it to the protein chain. At this point, the whole polymerase machinery slides across the mRNA, through a sequence of allosteric changes mediated by GTP-GDP energy provision, and the translation of the next codon begins.

3.2.3 Termination

The end of the protein-coding message encoded in an mRNA molecule is marked by a stop codon, which is UUA, UAG or UGA. None of these three codons has a counterpart amino acid in the genetic code, but rather they serve to signal to the ribosome that translation is over. This causes the

release of the newly formed protein chain into the cytoplasm, and triggers a conformation change of the large and small sub-units of the ribosome, which gets back to open form, therefore releasing the mRNA molecule.

The synthesized protein filament undergoes a complex sequence of rearrangements, whereby the linear structure is modified into a folded one, which provides the proper three-dimensional arrangement that enables the protein to accomplish its molecular functions.

3.3 Degradation

Molecules of mRNA can be degraded by many different processes. Only a fraction of the mRNA molecules that is produced in the nucleus of a eukaryotic cell does reach the cytoplasm, because of many quality control steps that the cell actively employs to degrade any unfit nuclear mRNA.

The mRNA molecules that reach the cytoplasm are mostly degraded through a mechanism called deadenylation. This process results from a gradual shortening of a long sequence (around 200) of A nucleotides found in the mRNA. A specific exonuclease enzyme shortens the poly-A tail, until it reaches a minimal length (around 30), after which the mRNA molecule is rapidly destroyed. This process competes with the translation. Any factor that enhances translation efficiency will have a negative effect on the speed of mRNA degradation.

4. Modeling of mRNA life cycle

All the processes that affect the abundance of the mRNA are tightly regulated by multiple factors. Cells are able to sense the environment and to adapt their behavior, i.e. which genes are transcribed and with what intensity, in response to external and internal signals, so to maintain homeostasis and to react to changing conditions, as appropriate.

Typically, the life cycle of an mRNA molecule is represented through a simple modeling of the transcription and the degradation events that consider them as constant-rate processes (Kar et al., 2009). This means that all of the regulation processes are commonly abstracted. However, various studies demonstrate the importance of an appropriate modeling of the complex phenomena related mRNA life cycle. For example, Decker and Parker (1993) provide experimental evidence that mRNA degradation processes involve multiple steps, and Pedraza and Paulsson (2008) have shown that those processes are multi-stage crucial to reduce fluctuations in the protein levels cells. A recent modeling study (Kuwahara, 2012), shows how the regulation of mRNA degradation can significantly affect the amount of protein.

In the following sub-sections, we consider different scenarios of the modeling of mRNA life cycle of increasing complexity. We start from the simplest case where all the regulation is abstracted, and then we incrementally add a modeling of the regulation of transcription and degradation. We use the naming scheme XY for the scenarios, where both X and Y can take the value U = unregulated, or R = regulated.

4.1 Scenario UU

The UU modeling scenario takes into consideration the simplest process of transcription, which occurs spontaneously without being activated by any particular configurations of DNA. In particular, we consider here that transcription does not need the availability of any transcription factors.

Also, the degradation occurs at a constant rate over time without any regulation or external influences. We can abstractly represent this scenario through the two following biochemical reactions:
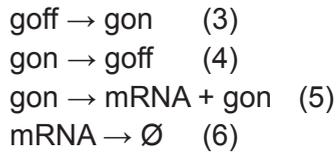
$$\varnothing \rightarrow \textbf{mRNA} \quad \textbf{(1)}$$
$$\textbf{mRNA} \rightarrow \varnothing \quad \textbf{(2)}$$

The syntax we use is borrowed from chemistry, with reactants at the left hand side and products of reactions at the right hand side. The symbol Ø is used to represent a reactant or a product that is outside the scope of the modeling. Hence, with the reaction (1), we are modeling the fact that mRNAs molecules are produced by a process that is using abstract reactants not represented into the model. For instance, the gene the mRNA is transcribed from is not represented, as it is not affected by the reaction. Also, we do not model the RNA polymerase, as we assume that it is very abundant and therefore not limiting the reaction occurrence speed. The form of reaction (1) makes clear that the transcription is not regulated, as we are not considering the availability of any transcription factor nor changes in gene state. The gene (abstracted) is always available for transcription. Reaction (2) represents the degradation reaction, which also happens in an unregulated way. mRNA molecules simply disappear from the model (they reach the outside represented by Ø). The speed of the degradation process is not affected by any changes.

4.2 Scenario RU

The RU modeling scenario takes into consideration a regulated process of transcription, which can for instance account for the existence of two different states or configurations of the DNA, or for the availability of transcription factors necessary to initiate the transcription process. In the first gene state, the gene encoding the protein to be transcribed is in an inactive state, and transcription is inhibited. In the second configuration the gene is active, and transcription occurs at a fixed rate > 0.

Gene status switches cyclically from the active to the inactive configuration, so that the model alternates periods of mRNA production to non-production periods. In contrast, the degradation occurs at a constant rate without
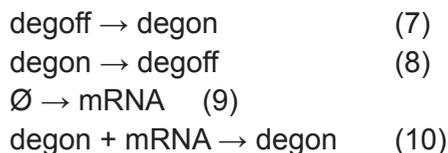
external influences, as in the previous UU scenario case. This modeling scenario can be represented by the following chemical reactions:

$$goff \rightarrow gon \quad (3)$$
$$gon \rightarrow goff \quad (4)$$
$$gon \rightarrow mRNA + gon \quad (5)$$
$$mRNA \rightarrow \varnothing \quad (6)$$

Reactions (3) and (4) model the change of state of the gene. The active state, i.e. the state where transcription is allowed, is called gon (gene on). The state where transcription is inhibited is called goff (gene off). The state changes cyclically from on to off. We do not model the phenomena that lead to this state change, only their effects. Reaction (5) models the transcription reaction. It requires the gene to be in on state to act as a reactant. If the gene is active, the transcription can occur and therefore reaction (5) can happen. The occurrence of a transcription produces an mRNA (first reaction product, right side of 5) and the gene stays in the active configuration (second reaction product, right side of 5). Therefore, additional transcriptions are allowed to happen, until the gene changes its state (reaction 4). Reaction (6) represents the degradation reaction, which happens in an unregulated way, totally equivalent to reaction (2) of the UU modeling scenario.

4.3 Scenario UR

In this scenario we distinguish between an active and an inactive state of the degradation process. In the active state, mRNA molecules are degraded at constant rate > 0, while in the inactive state of degradation mRNA molecules are not degraded. The degradation process alternates cyclically between the two states. This modeling scenario can be represented by the following chemical reactions:

$$degoff \rightarrow degon \quad (7)$$
$$degon \rightarrow degoff \quad (8)$$
$$\varnothing \rightarrow mRNA \quad (9)$$
$$degon + mRNA \rightarrow degon \quad (10)$$

Reactions (7) and (8) model the change of state of the mRNA degradation process. The active state, i.e. the state where degradation is allowed, is called degon (degradation process on), whereas the state where degradation stops is named degoff (degradation process off). The state changes cyclically from on to off. Reaction (9) models the transcription reaction, which is unregulated and therefore equal to reaction (2) of the UU scenario. Reaction (10) represents the mRNA degradation, which is only allowed to occur when the degradation process is active (reactant degon of the reaction). The state of the degradation process is not affected by the occurrence of the mRNA degradation reaction, which is the reason why degon appears as a product of the reaction.

4.4 Scenario RR

In this scenario, both the processes of transcription and degradation are regulated and alternate cyclically between states of activity and inactivity, therefore modulating both the synthesis and destruction of the mRNA molecules. The model can be encoded in terms of biochemical reactions as follows:

$$goff \rightarrow gon \qquad (11)$$
$$gon \rightarrow goff \qquad (12)$$
$$degoff \rightarrow degon \qquad (13)$$
$$degon \rightarrow degoff \qquad (14)$$
$$gon \rightarrow mRNA + gon \qquad (15)$$
$$degon + mRNA \rightarrow degon \qquad (16)$$
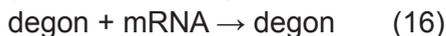
Reactions (11) and (12) model the change of state of the gene, and are the same as reactions (3) and (4) of scenario RU. Reactions (13) and (14) model the change of state of the degradation process of the mRNA molecules, and are the same as reactions (7) and (8) of the UR scenario. Similarly, the transcription reaction (15) models the regulation exerted by the gene state, as in reaction (5) of the RU scenario, and reaction (16) models the regulated degradation of mRNA, same as in reaction (10) of the UR scenario.

5. Analysis of models and discussion

In this section we analyze and compare the results that can be obtained with the four models, for both the continuous-deterministic and discrete-stochastic interpretation. To realize the comparison among the different modeling options, we shall make reference to the ability of the models and of their interpretation to match the observable biological data that can be collected by experimental techniques. Specifically, we shall be considering two different types of observables:

- Average abundance of the mRNA at steady state, a measurement that can be obtained through standard experimental techniques such as transcriptomics, microarray, and RNA sequencing.
- The time-courses of mRNA abundance, measured through single cell analysis over time, for instance through multiple repetitions of single cell Fluorescence In-Situ Hybridization (FISH).

5.1 Match against average steady state values

In this section we assume that a measurement of the steady state abundance of the mRNA is available, which is obtained from a population of cells.

5.1.1 Scenario UU

In this scenario we only have the two reactions (1) and (2). Since COPASI uses a model definition language oriented to chemical reactions, we can easily encode the model (Figure 1). The only variable of the model is the concentration of the mRNA, represented by the species mRNA. The transcription reaction has a rate law of type Constant Flux, meaning that it will constantly occur at a rate k_trscr, whereas the degradation reaction has a rate law of type Mass Action, which means that each mRNA will be degraded at a constant rate k_deg, and the total mRNA degradation rate will be proportional to the concentration of mRNA molecules.

Figure 1. COPASI reactions for scenario UU
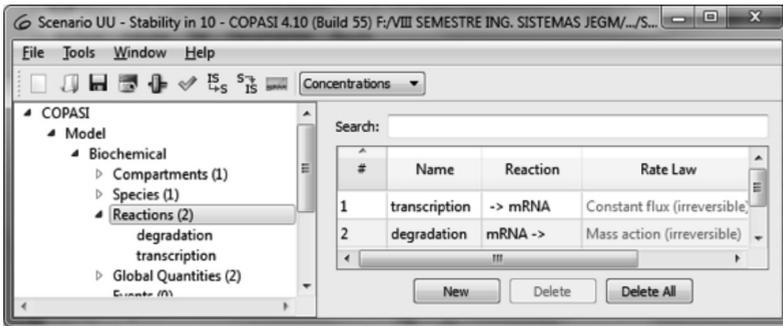
With Möbius, the discrete count of mRNA molecules is represented by the number of tokens contained in the place mRNA of the Petri net (Figure 2). The two reactions are modeled through the transitions t_trscr and t_deg. Transition t_trscr fires at a rate k_trscr, and one token representing one molecule of mRNA is added to place mRNA. When transition t_deg fires (which happens at a rate k_deg), one token is removed from the mRNA place.

Figure 2. Petri net model of scenario UU in Möbius
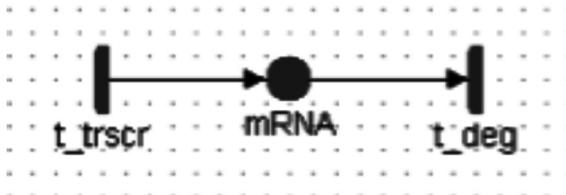
For this simple scenario, the equilibrium value of the abundance of mRNA is known analytically and is equal to k_trscr/k_deg. Therefore, any pair of values of the two parameters that turns out in a ratio equal to the number of observed abundance of mRNA will provide a model that reproduces the measured biological phenomenon. Without any loss of generality, we assume that the measured abundance of mRNA at steady state is 10, to be interpreted as a concentration in the deterministic setting and as a molecular count in the discrete one. We set the values of the two parameters to be k_trscr=1.0 and k_deg=1.0, for both COPASI and

Möbius. With these values of the rates, both models, starting from an initial state where no mRNA is present in the model, achieve a steady state that is equal to 10 (Figure 3).

Figure 3. Graphic of COPASI model (A) and Möbius model (B) simulation results – Scenario UU
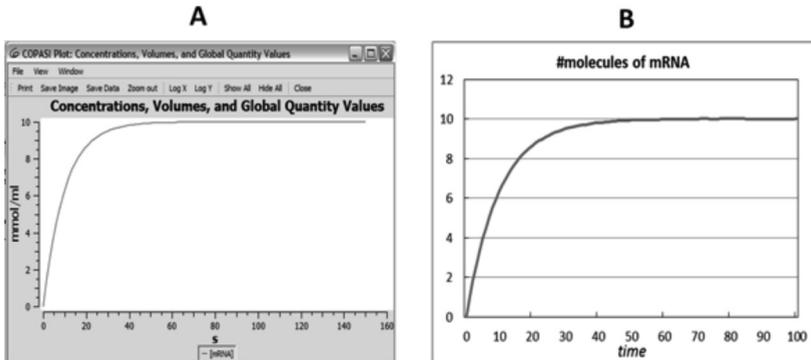
The graphic A in is the solution provided by the COPASI simulator, and the Excel chart B in the same figure is the transient solution returned by the Möbius simulator. As it can be seen, the two model implementations return exactly the same dynamics and steady state values of the mRNA.

5.1.2 Scenario UR

In this scenario we have four reactions to be modeled and therefore we have four parameters. Besides the rate of transcription k_trscr and of degradation k_deg, we need to define the rate of the activation and deactivation of the degradation process. We call these two new parameters k_act2, for the activation, and k_deact2, for the deactivation. Because the degradation process is not continuously active, if we used the same values of k_trscr and k_deg that we defined in the UU scenario, there would be an excess of mRNA abundance at steady state. Therefore, to compensate for the periods of inactive mRNA degradation, we will need either to decrease the mRNA transcription rate k_trscr, or to increase the mRNA degradation rate k_deg (Figure 4).

Figure 4. COPASI reactions for scenario UR
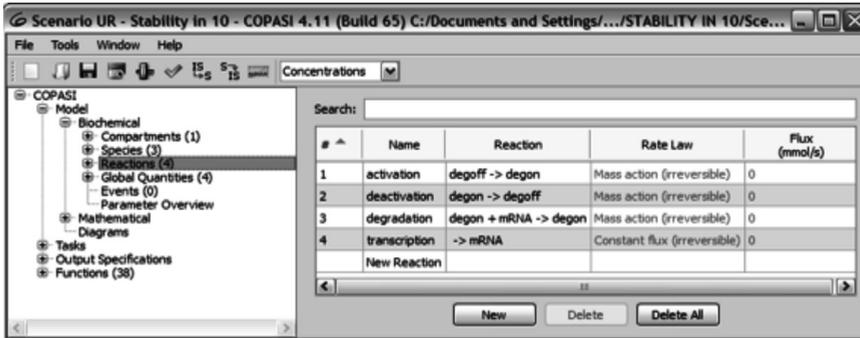
The values of the four parameters for the COPASI implementation are k_trscr=1.0, k_deg=0.3, k_act2=1.0, and k_deact2=2.0.

With Möbius, the discrete count of mRNA molecules is again represented by the number of tokens contained in place mRNA. The mRNA degradation process has now two states. The active state, i.e. the state where degradation is allowed, is called degon (degradation process on), whereas the state where degradation stops is named degoff (degradation process off). We add to the UU model two new corresponding places to model these states. The state changes cyclically from on to off, so we introduce two new transitions, named act_deg and stop_deg, that move one single token from place deg_off to place deg_on, and vice versa (Figure 5). For the degradation transition to fire, the token must be in place deg_on that is the degradation process is in the active state. The token is put back into place deg_on when the degradation transition fires. The firing rate of transition act_deg is k_act2, and the one of transition stop_deg is k_deact2.

Figure 5. Petri net model of scenario UR in Möbius
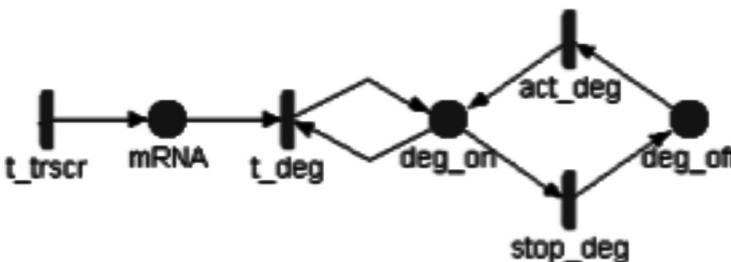
We assign the following values to the firing rates of transitions for the Möbius implementation: k_trscr=1.0, k_deg=0.3214, k_act2=1.0 and k_deact2=2.0. We show in 6 the time-dependent solution for the abundance of mRNA, as obtained from the COPASI model implementation (A) and from the Möbius model implementation (B). If we compare the results with both graphical representations we can see that de behavior on the green line on the COPASI model and the blue line on the Möbius model are the same, i.e. we can see that both graphics show us a steady state value of the mRNA equal to 10 (Figura 6).

Figure 6. Graphic of COPASI model (A) and Möbius model (B) simulation results – Scenario UR
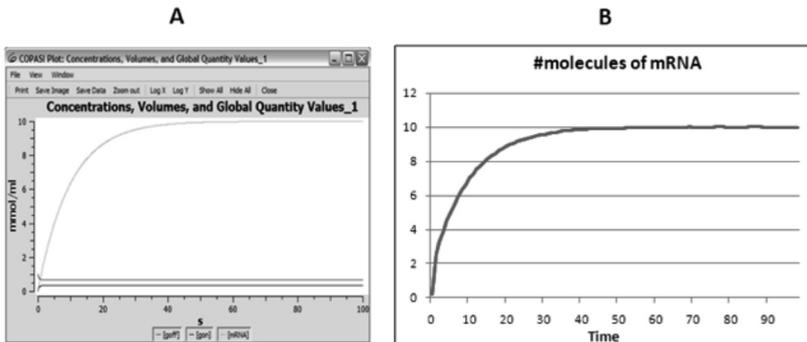
5.1.3 Scenario RU

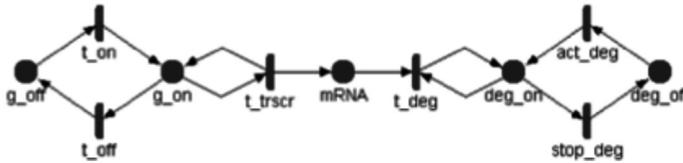In this scenario we have again four reactions to be modeled and hence four parameters. Besides the rate of transcription k_trscr and of degradation k_deg, we need to define the rate of the activation and deactivation of the transcription process. We call these two new parameters k_act1, for the activation, and k_deact1, for the deactivation. Because the transcription process is not continuously active, if we used the same values of k_trscr and k_deg that we defined in the UU scenario, there would be a minor mRNA abundance at steady state. Therefore, to compensate for the periods of inactive mRNA transcription, we will need either to increase the mRNA transcription rate k_trscr, or to decrease the mRNA degradation rate k_deg. (Figure 7).

Figure 7. COPASI reactions for scenario RU
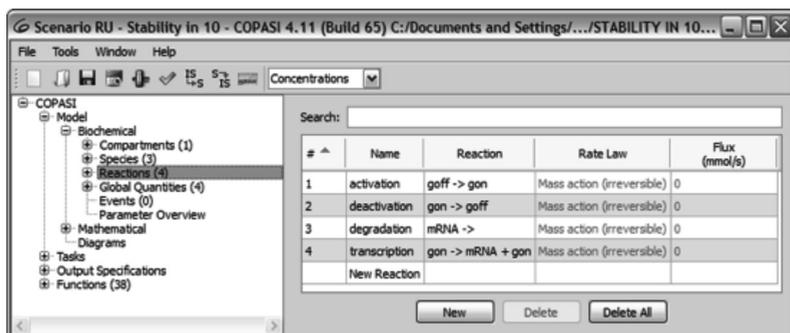
The values assigned to the model parameters are k_trscr=3.0, k_deg=1.0, k_act1=1.0, k_deact1=2.0.

In the Möbius model (Figure 8), the firings of transitions t_on (firing rate k_act1) and t_off (firing rate k_deact1) represent the activation and the deactivation of the transcription process, and the state of the transcription process is modeled by the content of the places g_off and g_on. A single token circulated between these two places: when the token is in place g_off, the transcription is inactive, and when the token is in place g_on, the transcription is active.

Figure 8. Petri net model of scenario RU in Möbius

We assign to the Möbius model parameters the same values we use for the COPASI version. We show in Figure 9 the time-dependent solution for the abundance of mRNA, as obtained from the COPASI model

implementation (A) and from the Möbius one (B). We can see that both models achieve a steady-state value of the mRNA equal to 10 (Figura 9).

Figure 9. Graphic of COPASI model (A) and Möbius model (B) simulation results – Scenario RU
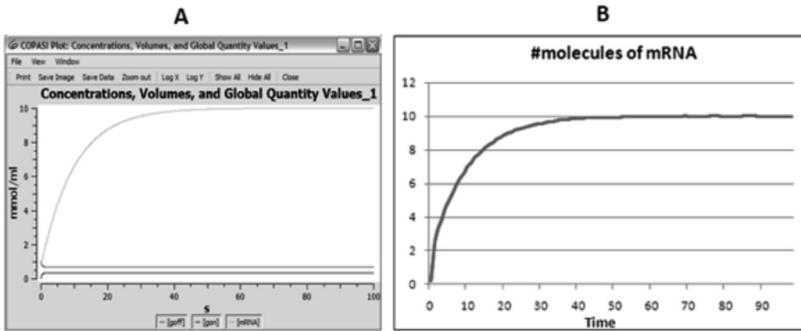
5.1.4 Scenario RR

In this scenario, we combine the regulated transcription and the regulated degradation. We have therefore six reactions to be modeled, two for the regulation of the transcription, two for the regulation of the degradation and two for the synthesis and degradation of the mRNA. The model is easily encoded in COPASI, (Figure 10), and in Möbius (Figura 11).

Figure 10. COPASI reactions for scenario RR

Figure 11. Petri net model of scenario RR in Möbius.

The values of the rates used to instanciate the two models are as follows: for COPASI, k_trscr=3.0, k_deg=0.3, k_act1=k_act2=1.0, k_deact1=k_deact2=2.0, for Möbius, k_trscr=3.0, k_deg=0.3214, k_act1=k_act2=1.0, k_deact1=k_deact2=2.0. Figure 12 provides the time-dependent solution for the abundance of mRNA, as obtained from COPASI (A) and from Möbius (B). As we can see, both graphics show us a steady state value of the mRNA equal to 10 (Figura 12).

Figure 12. Graphic of COPASI model (A) and Möbius model (B) simulation results – Scenario RR
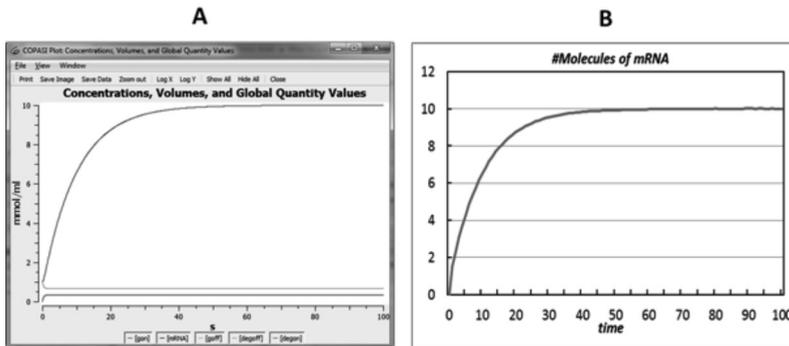
5.2 Match against time courses

We consider here how the abundance of mRNA varies over time. This type of observable can be experimentally obtained by taking multiple snapshots of the amount of mRNA in single cells, or on a population of synchronized cells. It is interesting to look at two very different patterns of mRNA time course that have been measured, and that reflect the two radically different situations of gene expression regulation:

- Genes that are constitutively expressed over time, without changes in the required level of their products, for instance genes that code for metabolism required enzymes;
- Genes that are expressed with precise time-dependent patterns, for instance the genes that are used to regulate cell-cycle progress (Ball et al. 2013).

The two cases give raise to very distinct patterns of time courses of the mRNA. In case 1, the abundance of mRNA oscillates around its average value, without any specific pattern. In case 2, it is possible to appreciate "oscillations" of the abundance of mRNAs, with clear peaks of expression over time.

5.2.1 Constitutively expressed genes

In the case of constitutively expressed genes, we can easily reproduce the pattern of the typical time course with the UU model scenario. Indeed, for such genes there is a perpetual equilibrium of the mRNA number in the cell with at most a random fluctuation around such equilibrium. If we look at the steady state time courses obtained with the continuous-deterministic and discrete-stochastic version of the UU model, we find patterns (Figure 13). Similar curves, apart for the lack of noise, can be obtained with COPASI.

Figure 13. Graphic of time courses of the mRNA in scenario UU, obtained with Möbius.
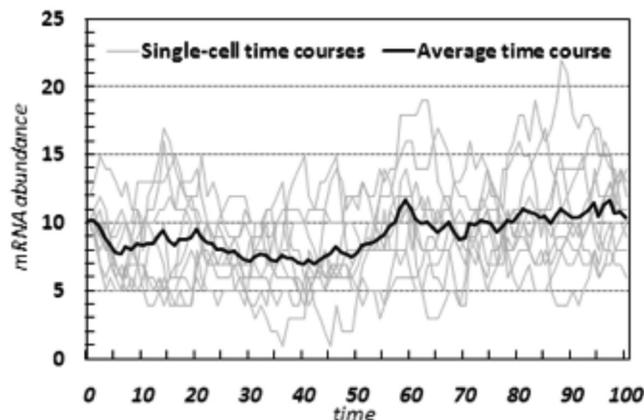
With the models of the scenarios UR, RU and RR, the situation is somehow more complex. In the stochastic versions of the models, any run of the RU and UR models will provide curves with clear indications of the regulations, qualitatively different patterns such as those shown previously. We show the results of simulations for the RU scenario (Figure 14). Similar results are obtained for the UR scenario.

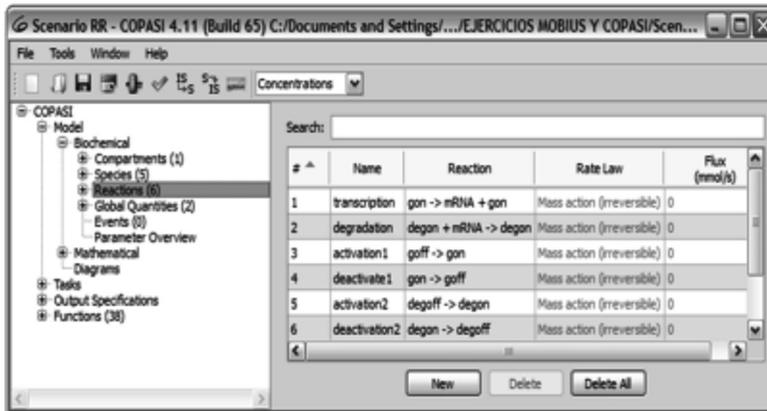**Figure 14. Graphic of the simulated time courses of mRNA in scenario RU**
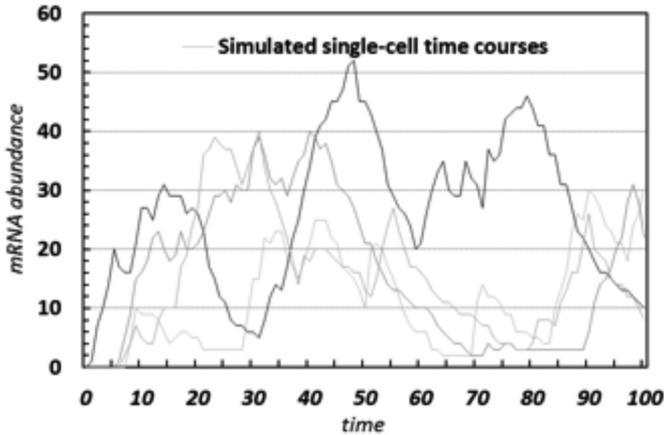
These considerations apply for the stochastic versions of the RU, UR and RR models. The continuous-deterministic ones that we built in COPASI are consistently able to reproduce the pattern of the constitutive gene expression time course because the regulating processes quickly go to equilibrium (which means they stop changing their state) and their effect on the transcription and degradation disappears.

5.2.2 Genes with patterns of expression

In the case of actively regulated genes, we expect the scenarios UR, RU and RR to be those of practical use. Let us first focus on the discrete-stochastic versions of these models, which are easier to interpret. For instance, let us look at the set of RR time courses (Figure 15). The repeated switches between the on and off states of the transcription and degradation process give raise to the oscillating pattern of the total number of mRNAs.

We would like to mention the fact that, if the time courses of each individual cell are not aligned in terms of the switch times, the average value of the time course may lose the oscillation characteristic patterns due to the asynchronous pulsing. This is indicating that collecting observable of mRNA time courses from a population of cells may in fact hide the characteristics of regulation of the gene.

**Figure 15. Graphic of the simulated time courses of mRNA in the scenario RR-synchronized**

As for the continuous-deterministic interpretation of models, as already mentioned, each of the UR, RU and RR scenarios goes to a stable steady state that cannot reproduce the pattern of time course of a regulated gene. To provide for an oscillating time course, we need to modify the COPASI version of those model scenarios by replacing the simple on-off modeling of the state of the process with a non-linear driving pulse that does not have a steady state. This can be accomplished by introducing a molecular oscillator, for instance a simple 3-molecules oscillator with autocatalytic reactions such as the one in Ballarini, Mardare, and Mura (2009). With this addition we may have the transcription and/or degradation to be driven in a synchronized way, obtaining patterns of mRNA abundance that are oscillating as shown in the COPASI results for the RU modeling scenario (Figure 16).

Figure 16. Graphic of the simulated time courses of mRNA in the scenario RU-synchronized
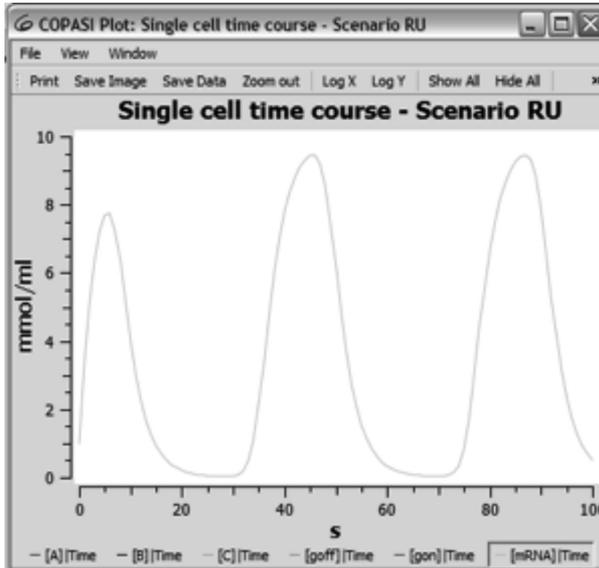
6. Final comments and conclusions

The results we obtained by the analysis of the capabilities of the various modeling options offer some interesting points for reflection. First of all, we point out that all the proposed modeling scenarios, independently of their continuous or discrete implementation, are able to reproduce any biological observation that only consists of a measurement of the average value of the mRNA.

However, it is interesting to look at the values of the rate parameters that have been used in the different situations analyzed in Section 5.1 to instantiate models so that they can consistently produce the same steady state value of the mRNA (10) in all considered cases. We summarize the parameter values (Table 1).

Table 1. Summary of values assigned to the model rate parameters

	UU		RU		UR		RR	
Rate	COPASI	Möbius	COPASI	Möbius	COPASI	Möbius	COPASI	Möbius
k_trscr	1.0	1.0	3.0	3.0	1.0	1.0	3.0	3.0
k_deg	0.1	0.1	0.1	0.1	0.3	0.3214	0.3	0.3214
k_act1	-	-	1.0	1.0	-	-	1.0	1.0
k_deact1	-	-	2.0	2.0	-	-	2.0	2.0
k_act2	-	-	-	-	1.0	1.0	1.0	1.0
k_deact2	-	-	-	-	2.0	2.0	2.0	2.0

We can notice that for the UU and UR scenarios, the continuous-deterministic and the discrete-stochastic models share the same values of rate parameters. The representation of the model is not important for the sake of the steady state value that the model achieves.

When the degradation is regulated (scenarios UR and RR), there is a difference between the set of values that need to be assigned for achieving the desired steady state value. The influence of the inactive degradation process is more important in the discrete-stochastic implementations of the models, which is demonstrated by the need of having a higher value of the degradation rate k_deg. Intuitively, in the continuous-discrete implementations, the switch between states of active and inactive degradation is influencing the model only through the relative frequency of occupation of the two states. As we are using a rate of 1.0 for the activation and a rate that is twice (2.0) for the inactivation, at the equilibrium all the model versions will spend 1/3 of the time in the active and 2/3 of the time in the inactive degradation state.

In the continuous-deterministic model, the net effect of the degradation switch off is totally compensated by a corresponding proportional increase of the k_deg rate with respect to the baseline value we used in the UU scenario, so we set k_deg = 0.1 * 3 = 0.3. In the discrete-stochastic version, the value of the rate k_deg must be set to value higher than 0.3. The reason is that, beyond sensing the effect of the degradation switch off periods, the discrete-stochastic model is also sensitive to the speed of

the regulation process, which is much faster than the degradation itself. Indeed, the rates of the degradation process state change are 2.0 and 1.0, whereas the speed of degradation of each single mRNA molecule is about one order less. The degradation reaction will be competing with the state change of the degradation process, and degradations may be stopped before completing due to a transition to the inactive degradation state. To better explain it, suppose we set the rates of state change of the degradation process to 200 and 100. At equilibrium, the degradation process would again be active for 1/3 of the time and inactive for 2/3 of the time, but it would be very hard for a degradation to complete before the process moves to the inactive state. This interference, which becomes important when very different time scales of event occurrence times are used, see for instance Kuwahara and Schwartz (2012), is an additional effect that is not visible in the continuous-discrete model implementation. The important conclusion that can be drawn is that, even if all the model versions are able to reproduce the biological observation of the steady state value of the mRNA abundance, the determination of the model parameters needs to take the regulation into account. Whenever regulation processes exist in the mRNA life cycle, if they are not properly represented into the models, then the values used for the rate parameters would have no correspondence with reality.

Finally, when matching the observed time courses of mRNAs, we found that both the discrete-stochastic and the continuous-deterministic implementations of models are able to match the time courses of unregulated genes. However, the continuous-deterministic ones consistently return a pattern of mRNA abundance variation that can only match the one of unregulated gene expression. This is because to create oscillating behaviors in the continuous setting requires introducing non-linearity in the models, whereas all the modeling scenarios we defined in the paper are linear. In this respect, the discrete-stochastic model interpretations turn out to be much more flexible. With a minimum level of complexity, the stochastic models can represent both types of gene expression time courses.

7. Bibliography

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Molecular Biology of the Cell (4th ed.). New York: Garland Science.

Ball, D. A., Adames, N. R., Reischmann, N., Barik, D., Franck, C. T., Tyson, J.J., Peccoud, J. (2013). Measurement and modeling of transcriptional noise in the cell cycle regulatory network. Cell Cycle (12) 19, pp. 3203-3218. doi: 10.4161/cc.26257.

Ballarini, P., Mardare, R., & Mura, I. (2008). Analysing Biochemical Oscillations through Probabilistic Model Checking. Electronic Notes in Theoretical Computer Science. (229) -1, pp. 3–19. doi: 10.1016/j.entcs.2009.02.002

Decker, C. J., & Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. Genes Dev. (7) 8, pp 1632-1643. doi: 10.1101/gad.7.8.1632

Denti, M., Viero, G., Provenzani, A.,Quattrone, A., & Macchi, P. (2013). mRNA fate: life and death of the mRNA in the cytoplasm. RNA Biology (10) 3, pp. 360–366. doi:10.4161/rna.23770

Kar, S., Baumann, W., Paul, M. R., & Tyson, J. J. (2009). Exploring the roles of noise in the eukaryotic cell cycle. PNAS (Vol. 106:16, pp. 6471–6476). doi: 10.1073/pnas.0810034106

Kitano, H. (2002). Systems biology: a brief overview. Science (295) 5560, pp. 1662-1664. doi: 10.1126/science.1069492

Kuwahara, H., & Schwartz, R. (2012). Stochastic steady state gain in a gene expression process with mRNA degradation control. Journal of the Royal Society Interface. (9) 72, pp. 1589-1598. doi: 10.1098/rsif.2011.0757

Mura, I. (2013-a). Modeling noise in gene expression. Revista de Investigación Ontare, 1 (1) pp. 151-166.

Mura, I. (2013-b). On modeling approaches for the predictive simulation of living systems dynamics, Revista de Investigación Ontare, (1) 2, pp. 101-124)

Mura, I. (2010). Stochastic modeling. In I. Koch, W. Reisig, & F. Schreiber (Eds.) Modeling in Systems Biology, The Petri Net Approach, Computational Biology (Vol. 16). London: Springer-Verlag London Limited.

Pedraza, J. M., & Paulsson, J. (2008). Effects of molecular memory and bursting on fluctuations in gene expression. Science (319) 5861, pp. 339-343. doi: 10.1126/science.1144331