uc3m | Universidad Carlos III de Madrid

Biomedical Engineering

(2017-2018)

Bahelor thesis

"Synthetic biology of genetic circuits"

Gabriel Rodríguez Maroto

Tutor: Victoria Lucía Doldan Martelli

Leganés, 11th July

Aknowledgements

I would like to express my gratitude to all the people that have contributed to this bachelor thesis offering me their support, advice and time.

ABSTRACT

The combination of positive and negative feedback loops has been shown to increase the robustness of oscillations. Such breakthrough has enabled to understand the importance of that dual control in a few biological systems. The reason is that most biological systems are non-linear. One of the obstacles that must be overcome when dealing with non-linear systems, even if they are simple, is that the use of intuition to predict its behavior is no longer valid. The comprehension of the behavior of the system can only be achieved mathematical modeling and computational simulations

This bachelor thesis aims to develop a mathematical model of the *relaxoscillator*, a gene regulatory network in which two genes with identical promoters are regulated by the same activator and repressor. At the same time, the binding of those depends on the concentrations of two inducers: arabinose and IPTG, which correspond to the control parameters of the system. The obtained model, derived from the chemical reactions, was simulated under different inducer concentrations in an attempt to comprehend the long term behavior of the system. The results show that varying these inducer concentrations allows to tune the period and the amplitude of the oscillations observed in the system. In order to analyze changes in the long term behavior of the system it will be required to include a third control parameter, the transcription repressor rate, so that the system displays different dynamic behaviors. The analysis of the simulations indicates the presence of a supercritical Hopf bifurcation for a given value of the transcription repressor rate that would explain the transition between damped oscillations and persistent oscillations. Nevertheless, due to the theoretical nature of the project, experimental studies as well as two-parameter bifurcation analysis should be performed in order to confirm such hypothesis and gain understanding of the behavior of the system as a function of inducers concentrations.

Contents

List of Figures

Table index

1. Introduction

1.1 Motivation

Less than thirty years ago, the revolution in genomics meant the origins of synthetic biology. Nowadays the breakthroughs in such recent discipline have made possible to talk about *post-genomic* research [1]. One of the future milestones planed by these post-genomic investigations is the comprehension of cellular processes resulting from interactions between genes and protein. Such understanding is mandatory for the construction of synthetic genetic circuits. So far, examples of gene circuits had been successfully understood and implemented [2], [3]. In spite of this, real use of these synthetic circuits for different applications - biofuels, medicine, agriculture purposes- is still far from being achieved. One of the reasons is that these synthetic biological circuits are proof-of-principle designs, what causes them to not be used outside the laboratory [4]. Another limiting factor is the inherent complexity of biological systems. Most of these systems are nonlinear. Such feature avoids to predict their behavior - and therefore comprehending them – intuitively. Those studies can only be performed through computational simulations of the system under different conditions. This project follows that research line: the study of a complex biological system using a mathematical model to simulate its behavior.

Understanding sleeping disorders, insight in Alzheimer's diseases or even the implantation of cells containing synchronized clocks that are modulated to synthesize a desired protein - for example insulin - at regular time intervals in the right dose [5] are some of the aspirations of synthetic biology. The achievements of those goals will only be possible if the given gene or protein regulatory network is successfully understood.

1.2 Fundamental concepts

In the following section, some of the most fundamental concepts needed to the completion of this bachelor thesis will be presented.

1.2.1 Systems biology and synthetic biology

"The whole is greater than the sum of its parts", Aristotle. Someone could use this quote to describe, in an abstract way, what systems and synthetic biology have contributed to biology. These relatively new *"cousin"* - but different - disciplines were responsible for substituting traditional reductionist approaches when studying molecular biology by progressively more global analysis strategies.

Following chronological history to introduce them, systems biology was earlier than synthetic biology. As it names indicates, systems biology was born from the application of systems science to biology. It is true that idea of biology as a single scientific field did not crystallize until the 19th century, but it is also clear that in order to establish the origins of the science of life one has to rewind until the Ancient Greece when Aristotle became one of the firsts to systematically study it. Such past is something systems biology cannot boast about. Systems biology origins point toward the middle of the last century [4] when another discipline of the same field, molecular biology, started to discover some of the networks of interacting molecules driving cellular behavior.

Understanding why the origin of systems biology occurred at that time and not before requires focusing in the "systems" word. A system can be defined as an organized combination of interconnected components forming a whole. In a parallel way, as the French biologist Francis Jacob said, all biological systems are in fact systems of systems. With that in mind, there have always been two ways of analyzing systems: the reductionist and the holistic one. The main exponent of reductionism was René Descartes. The rational philosopher of 17th century believed that the best strategy when dealing with complex problems was its decomposition into smaller problems. In this way, he stated that the sum or the re-assembly of each of those small problems was identical to the original complex problem. Although his postulates were earlier than the development of biology as a formal science and their initial impact was greater in other disciplines like physics, biology adopted this simplistic perspective up to the first half of the XXth century. In fact, reductionist approaches are still used by biology in specific problems in which the division of the problem into hierarchies with sub-problems is the only mechanism to understand it.

With time, the assumption that everything could be explained using reductionist strategies extended in the biological field. However, there was one characteristic of the systems that was not being taken into account, its emergent behavior. This emergent

behavior is the behavior of the system that does not depend on each of its components but on the interactions between them.

Then, reductionist strategies in biological systems were missing the fact that the interaction between the components of the system is as important as each of the components individually. As consequence, systems reductionism was missing or omitting one of the defining features of a system which explains why the beginnings of systems biology can be assumed to come later in time. Such limitations of the mechanistic approach to biological systems prompted inevitable reactions against reductionism [6] at the beginnings of the 20th century. The general objection was the error of considering the whole only as the sum of its parts. Aristotle's ideas in this way reappeared again. Biological systems like cells had emergent properties - life - that could not be explained just as the sum of individual parts. Then, this opposition to the reductionism resulted into a new perspective that opted for analyzing systems as networks of connected individual components and their associated interactions. This was known as *holism.* This change of paradigm affected different disciplines and within molecular biology field it enabled to study the global behavior molecular systems. Eventually, it gave rise to *systems biology*.

The replacement of the reductionism approach by the holistic one does not invalid the former one. Both strategies can be used depending on the specific problem to which they are applied. The point is that analysis of biological systems has proved to demand not only which are the biological components forming it, but also understanding how they interact. This is something reductionist approaches cannot satisfy.

Regarding synthetic biology, its appearance was a consequence of the development of systems biology. If systems biology was developed in an attempt to overcome the limitations of mechanistic approaches when analyzing molecular systems [7], synthetic biology emerged from the need to reproduce or simulate those previously analyzed molecular systems. In other words, first with systems biology the problem is analyzed and later, once the component interactions have been understood, synthetic biology tries to combine those biological parts to achieve a particular goal. The central idea of synthetic biology consists of the application of engineering principles to create, monitor and control cellular behavior [4].

Notice that this inverse relationship between systems biology and synthetic biology can be interpreted as an example of reverse and forward engineering. The latter one refers to the process going of implementing in reality a theoretical (logical) model, and thus, it corresponds to the activities developed by synthetic biology. In contrast, reverse engineering, corresponding to systems biology, is responsible from reconstructing or inferring that theoretical model from the analysis of a real biological system. This explains also why in order for a discipline like synthetic biology to exist, there must exist before a reverse engineering discipline, systems biology, that has previously

analyzed and understood the whole behavior of the system including its interactions, so that now such system can be recreated from the theoretical model.

Synthetic biology appeared associated to the genomic revolution and the already mentioned development of systems biology in 1970 [8]. Similarly to systems biology, this discipline chose global approaches when manipulating cellular behaviors with biological tools. The evolution and the transition from systems biology to synthetic biology were only possible due to the parallel development of mathematical and computational modeling frameworks [9]. Once these two other tools were able to account for the data volume associated to biological systems, both systems and synthetic biology could start to study larger biological systems [10]. Then, in the particular case of synthetic biology, these mathematical and computational breakthroughs made possible the publication of two scientific papers [2][3] considered as two historical landmarks in synthetic biology evolution [11]. These papers [2],[3] were the proof that the forward engineering problem synthetic biology was attempting to solve could effectively be solved. Consequently, this demonstrated that systems biology had successfully addressed the reverse engineering process too. By the first time, design and construction of a synthetic genetic network were detailed. A *"repressilator"* - a type of synthetic oscillator - and a "*toggle switch"* –a bistable gene regulatory network- were the first tangible evidences demonstrating how successful the synergy between engineering and biological systems could potentially be. Once the basis were established, synthetic biology continued evolving becoming one of the hottest emerging areas of biological science research [11]. Nowadays, the synergy between synthetic biology, mathematics and computational science combined with the progress of technology has made it possible to bring synthetic biology closer to its foundational goal, the development of applications in medicine and biotechnology. Biofuels, biomedical compounds or new vaccines are just examples of those applications, but synthetic biology is expected also expected to account also for more theoretical questions like for instance discovering the minimal conditions required to life development.

Synthetic biology can be thought as the last stage in the evolution of biological systems analysis in molecular biology. During this process, the transition from reductionism to holism was responsible for the origins of systems biology. Later on, this discipline would be able to solve the reverse engineering problem, being able to comprehend not only the different components of a real biological system but also successfully inferring the relationships between them. Mathematical and computational breakthroughs were also required to overcome such challenge. At that moment, being able to analyze biological systems meant covering half of the trip. The other half was being able to do inverse process, constructing from the theoretical system an efficient biological system. Attempting to complete the forward problem synthetic biology was developed, so that eventually such designed biological system could be used to achieve a specific goal. Currently, synthetic biology has demonstrated that the application of engineering

principles to molecular biology is the correct strategy to model and control cellular behavior.

1.2.2 Dynamic mathematical modeling

As a forward engineering discipline, systems biology tries to develop models of biochemical and genetic networks. Models are abstractions of reality designed to represent systems. Among the different models biologists can use there are tangible models - molecular ball-and-stick or animal models - and conceptual models [12]. The latter models are the ones systems biology focuses on. Conceptual models are typically represented in the form of diagrams showing the interactions between the components of the system. Although useful when dealing with simple systems, biological models are characterized by the high number of species involved and by the complexity of their interactions. At the same time, molecular processes like gene expression or metabolic routes demand detailed modeling for investigation. These demands joined to the likely ambiguity transmitted by conceptual models when assessing feedback processes make the diagram representation of conceptual models insufficient when analyzing complex biological systems.

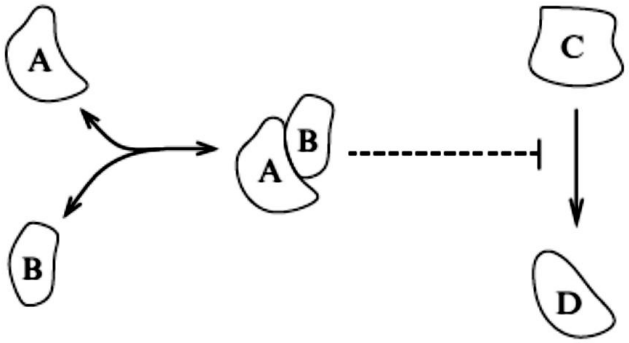

Fig. 1.1 Diagram representation of a conceptual model of cellular processes [12]

These limitations of simple pictographic representations of biological systems can be overcome adding a mathematical description of the system. Doing so, complex biological networks could be described also quantitatively. This is the central idea systems biology revolves around: being able to describe the systems and the components interactions quantitatively. The challenge is found on how to parametrize or mathematically describe such interactions so that it is possible to jump from the conceptual diagram to a dynamic mathematical model. The adjective “challenging” comes from the inherent characteristic of systems biology. The vast data volume biological networks imply is such great and complex, that just the mere attempt to extract from them concise biological formulations with predictive behavior requires in most cases from several years of analysis [13].

When a biological system is modeled, the resulting model would consist of a set of equations able to describe the temporal behavior of the system. To do so, the laws of

physics and chemistry would have to be applied (see 2.2.3 *Chemical reactions modeling*). Those models are called mechanistic models [12] since they are created to describe the mechanism explaining the observed behavior. In addition, there are two possible research options when working with mechanistic models: model simulation or model analysis. Model simulation can be literally defined as the evaluation of the system's model [14]. Under this strategy, models are used as predictive tools. The reason behind such demand deposited on models is the fact human intuition might not be accurate enough when trying to infer the dynamical behavior of complex nonlinear biological systems [15]. It is convenient to mention that although these simulations will never replace real experiments, they offer the possibility to check how the system would behave in situations that could never be replicated in a laboratory. Not being able to predict the exact - literally exact - behavior of the system does not prevent model simulations to be used as tools indicating research lines of investigation.

Alternatively, with model analysis, models can also be directly investigated to better understand why they behave in one way and not in other. Consequently, model simulation [16] would answer how the system temporally behaves depending on the initial conditions, while model analysis would focus on explaining why the system adopts such behavior.
Either way, system modeling is an iterative process. The initial hypotheses that are used to create the first model are tested to see if this one matches the observed data. In case of positive results, it means that the model was successfully built; if inconsistencies are found, these ones will have to be used to refine the initial hypotheses and thus re-compute a new model until the model is able to correctly predict the behavior of the system. Such iterative process contributing to the improvement of the model is known as the virtuous cycle [12].

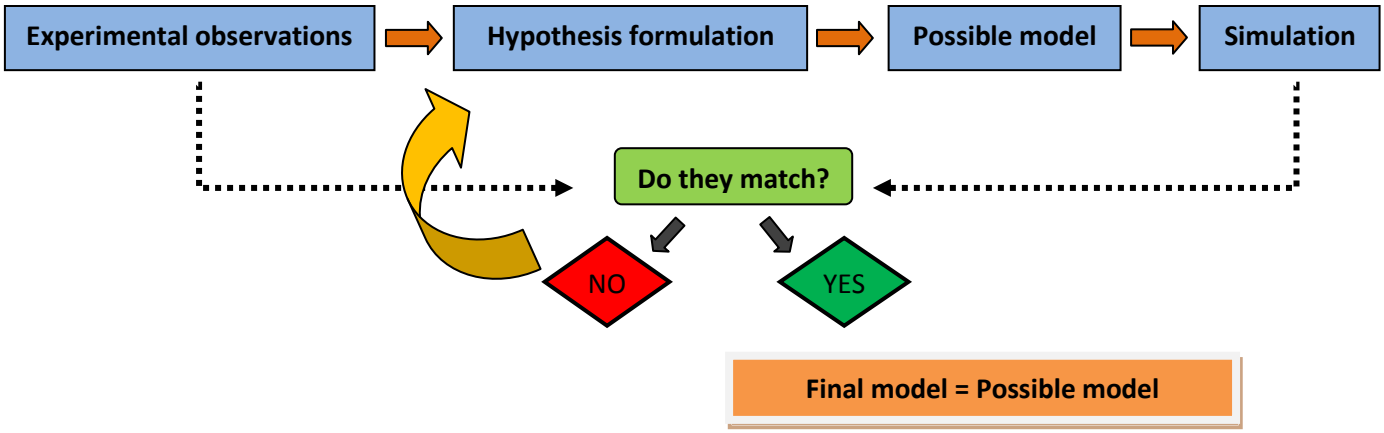

Fig. 1.2 Flow diagram showing system modeling design

As previously stated, the ultimate goal of system modeling consists of building a perfect model capable of predicting the behavior of the real system with maximum precision. In

biological systems this is yet to be achieved. In spite of this model feature susceptible to be improved, systems biology has been able to create good enough models that can be used in synthetic biology.

Finally, when introducing the most basic features defining dynamic systems, there is one feature that needs to be highlighted regarding biological systems: the existence of feedback loops. Those feedback loops can be either negative if systems components inhibit their own activity [12] or positive loops, which are behind processes in which small disturbances of systems components result into a greater perturbation. Although in principle negative feedback loops are associated to stable behaviors, they can induce oscillations when some kind of temporal delay is introduced in the system. Under the same logic, positive feedback processes, traditionally paired with unstoppable increasing divergence, can also be used to force the system to maintain on its long-term state when using appropriately.

1.2.2.1 Key aspects in dynamic mathematical models

One of the most basic components of mathematical models is called variable of state. Associated to each molecular species belonging to the system, it represents the abundance of a given species as a function of time. They should not be confused with the parameters. The latter ones are constant and they settle the environmental conditions as well as systems interactions. This explains why within the same simulation, parameters are kept fixed. The aim of a particular simulation is to recreate the behavior of the systems under a very precise set of environmental conditions. This does not prevent to globally study the behavior of the system under different external conditions provided a simulation for each set of conditions has been performed.
In relation to the distinction between state variables and parameters, these roles will have to be assigned depending on the specific biological system under study. While for some reactions the concentration of a given species will have attached an associated state variable, such species abundance could be assumed to be fixed if the studied biological process has now a different time-scale [12], thus becoming a parameter.

As previously mentioned, another important feature in dynamic mathematical models is at which time those models will be analyzed. For long enough simulations, systems will display their corresponding asymptotic or steady behaviors. However there is another time interval susceptible to be analyzed, the one going from the initial state to the asymptotic behavior. Such time interval is known as transient state.

Models can also be classified according to its linearity into linear or non linear models. Linear models are used whenever the interactions between systems components are linear. Although simpler models, the possibilities these type of models offer are scarce. In contrast in nonlinear models such relationship between systems components is precisely non linear, and thus it is more difficult to be modeled. This is the case of

most biological systems (further information in section 1.2.4 *Analyzing dynamics of mathematical models*.) In the same line, it is also possible to talk about global versus local behavior. Ideally, global behavior analysis is the aim of mathematical modeling. Being able to analyze the global behavior of the system implies that all interactions are understood, up to the smallest detail. Nevertheless, this global behavior analysis is sometimes unfeasible mostly due to the presence of nonlinearities. Consequently, local approaches are accepted as they allow one to consider those nonlinearities as linear, thus, facilitating the analysis. Someone could believe that such strategy oversimplifies too much the model at the cost of missing important information; however, it has been shown that the general behavior of systems is in a high proportion dependent on its behavior around several individual points [12]. In other words the analyzing the local behavior of the systems often provides enough information to infer the global behavior of the system.

Local behavior analysis is different to reductionism. While reductionism divides or compartmentalize the original system in parts to analyze them individually, local approaches analyze individual interactions between the components.

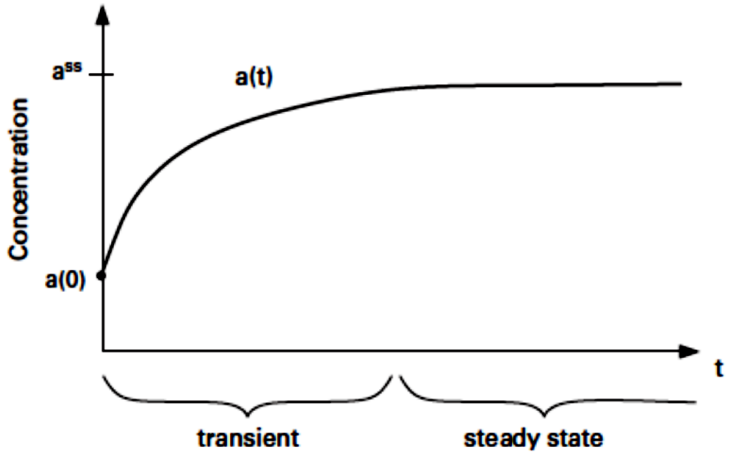

Fig. 1.3 Transient and steady states of a system's variable [12]

Furthermore, mathematical models can be stochastic or deterministic. In deterministic models the behavior of the model is exactly reproducible; this means that no matter how many times the system is simulated, the obtained result will be identical provided the parameters were also identical. On the other hand, stochastic simulations include randomness in such way that they are influenced not only by specified conditions but also by unpredictable factors [12]. The advantage of stochastic models is the possibility to account for such randomness contribution influencing the systems. For certain biological system like gene networks this randomness comes in the real world from the thermal agitation of individual molecules. Gillespie's simulation is an example of a *Monte Carlo* method (stochastic algorithm).

1.2.3 Chemical reactions modeling

As it was previously stated, one way of representing systems is through its interaction diagrams. These will contain the species involved in the system as well as arrows indicating the interactions between them. When dealing with simple systems, one can intuitively infer the behavior of the system just looking at its interaction diagram. However, biological networks do not fall within the group of "simple systems". One of their characteristics is their complexity which hampers elucidating the behavior of the system correctly. Fortunately, this obstacle can be overcome if the system is described not only qualitatively but also quantitatively. Chemical reaction modeling develops dynamical mathematical models of those biological networks. In the following section, some of the fundamental concepts involved in chemical reactions modeling will be introduced.

1. ***Chemical networks:***

   As their own name indicates, chemical networks are set of chemical reactions. From a mathematical point of view, a chemical network can be expressed [17] as:

$$R_i : \sum_{j \in S} \alpha_{ij} S_j \rightarrow \sum_{j \in S} \beta_{ij} S_j \qquad (1.1)$$

   Where R_i is the set of reactions, α_{ij} and β_{ij} are the stoichiometric coefficients and S_j represents each species participating in the reactions. Then, the left hand side components would correspond to the reactant species and the right side to the products. The arrow indicates the direction of the reaction meaning irreversible if its point rightwards. It has to be mentioned that theoretically all chemical reactions are assumed to be reversible [12]; nevertheless, those reactions can be described as irreversible whenever the reaction rates of the inverse process are negligible compared to the forward reaction rate.

   Chemical networks can be considered as closed or opened. In close networks all participating species lie inside the network. Consequently in the long term, these networks will reach a point in which all net reaction rates are zero (thermal equilibrium). In contrast, if a reaction happens spontaneously without reactants or without yielding products, such network will be classified as open, since there will be material exchange with the external environmental. This would be a simple example of an open chemical reaction.

$$A + B \rightarrow C \rightarrow$$

   In open networks there will be a steady flow through the chemical set of reactions when they are in steady state [12]. This is named as *dynamic equilibrium.*

2. Dynamics of chemical reaction networks (CRN)

As stated before, the aim of systems modeling it to obtain a model of the real system that is able to predict its behavior. When saying "behavior", one is referring to its temporal behavior. Then, in order to describe the temporal evolution of the system, it will be mandatory to infer the time-varying behavior of the different component of the system - molecular species - considering their interactions. If system components interactions have been assumed to be described as reactions between different species, in order to know the variation in time of each species, one will need to understand the reaction rates of each reaction. At the same time, these are known to depend on environmental conditions. Eventually, this dependence can be neglected assuming that chemical reactions occur at fixed environmental conditions, allowing reactions rates to depend just on the species involved.

The concept of temporal behavior of a molecular species refers to the concentration - in some cases, instead of concentration discrete number of molecules is used - of that molecular species (reactant or product) as a function of time.

There are two common assumptions made when modeling chemical reaction networks. One of them is that reaction rates are constant in the space [12]. Such assumption seems reasonable if one accepts that the distribution of the reactants is homogeneous. The second assumption refers to the possibility of representing the abundance of each molecular species in terms of the concentration provided that the number of molecules of such species is great enough. There are situations in which these assumptions do not hold. This second assumption, known as *continuum hypothesis* may weaken when the number of molecules per species is relatively low.

Under the previous assumptions, it is possible to use the law of mass action to describe chemical reaction networks. Formulated by Cato M. Guldberg ans Peter Waage, [18] two Norwegian scientists, this law states that *the rate of any chemical reaction is proportional to the product of the concentrations of the reactants* [12]. Considering the following reaction:

$$A + B \rightarrow C$$

According to the law of mass action, the reaction rate can be expressed as:

$$Reaction\ rate = k_i \cdot [A][B] \tag{1.2}$$

Where [A] and [B] are the concentration of the reactants and k_i is the rate constant, the constant of proportionality. The dimensions of this constant will be determined by the number of reactants as well as their kinetic order [12], the exponent to which each reactant appears. In case of a reaction consisting of the uptake of a species from the environment (for instance →A), such reaction is called zero order reaction with a reaction rate equal to the rate constant.

The construction of the dynamic mathematical model of a given biological network will be possible thanks to such law. The law of mass action will allow jumping

from the qualitative description of the system to the quantitative one. At the end, the result will be a set of ordinary differential equations (ODEs) describing the time-varying behavior of each of the biological components belonging to that biological system.

3. CRN example: Modeling a reversible conversion

Most biological systems are complex chemical reaction networks that result into complex systems of differential equations requiring numerical computation. In this section, a basic example will be solved so that the previously mentioned theoretical concepts can be better understood when applied to real examples.

$$A \underset{k_-}{\overset{k_+}{\rightleftharpoons}} B$$

In this example, k_+ and k_- are the rate constants corresponding to each of the production and degradation of species B respectively. Notice that as the there are two reactions, the time variation of each species will depend on two reactions rates, a positive one representing the synthesis of that species and a negative one indicating its degradation.

Rate of change of species A = rate of production of A – rate of consumption of A
Rate of change of species B= rate of production of B – rate of consumption of B

Mathematically, rate of change of the species can be expressed as the derivative of the species concentration with respect to time. Then, using that and applying the law of mass action those equations can be mathematically rewritten as:

$$\frac{dA(t)}{dt} = k_- \cdot B(t) - k_+ \cdot A(t) \tag{1.3}$$

$$\frac{dB(t)}{dt} = k_+ \cdot A(t) - k_- \cdot B(t) \tag{1.4}$$

Where *A(t)* and *B(t)* represent the concentration of each species. Notice that both species appear in both equations. It is possible to simplify such variable dependency by assuming steady state conditions. Accepting this assumption implies that the behavior of the system that will be analyzed will correspond to its long term behavior, not accounting for its transient one. Under the steady state condition, eventually the concentrations of both species will be constant when dynamic equilibrium is reached, meaning that although some A molecule species will be converted into B molecules, at the same time there will be B molecules doing the inverse reaction. This causes at the end, the net rate of change to be zero. From a mathematical point of view, this allows not having to explicitly solve the differential equations. Consequently, the system of equations remains as:

$$0 = k_- \cdot B^{ss} - k_+ \cdot A^{ss} \tag{1.5}$$

$$0 = k_+ \cdot A^{ss} - k_- \cdot B^{ss} \tag{1.6}$$

With A^{ss} and B^{ss} representing the steady state concentrations of each species. When these equations are solved, the following relationship is obtained:

$$\frac{B^{ss}}{A^{ss}} = \frac{k_+}{k_-} = K_{eq} \tag{1.7}$$

This relationship is the concentration ratio between the two species at steady state. Once this condition has been obtained, the initial equations (equations 1.3 and 1.4) can be rewritten using mass conservation principle - the mass is neither created nor destroyed in chemical reactions -. According to this, the total mass of products and reactants together must be constant over time. If such quantity is called C_t, the equation 1.3 becomes:

$$C_t = A_o + B_o = A(t) + B(t) \tag{1.8}$$

$$\frac{dA(t)}{dt} = k_- \cdot (C_t - A(t)) - k_+ \cdot A(t) \tag{1.9}$$

Grouping terms:

$$\frac{dA(t)}{dt} = k_- \cdot C_t - (k_+ + k_-) \cdot A(t) \tag{1.10}$$

Now, applying again the steady-state condition, equation 1.10 becomes:

$$0 = k_- \cdot C_t - (k_+ + k_-) \cdot A^{ss} \tag{1.11}$$

And solving for A^{ss} and repeating the same process with B^{ss} gives:

$$A^{ss} = \frac{k_-}{(k_+ + k_-)} \cdot C_t \qquad B^{ss} = \frac{k_+}{(k_+ + k_-)} \cdot C_t \tag{1.12}$$

These are the concentrations species A and B in their steady state. Notice that such assumptions have allowed not having to compute directly the differential equations. If solved, the solution would be:

$$A(t) = F \cdot e^{-(k_+ + k_-)t} + \frac{k_-}{(k_+ + k_-)} \cdot C_t \tag{1.13}$$

The constant F could be obtained using the initial condition ($A(t=0)=A_o$). This expression for the concentration of A would yield the same result as the steady-

state concentration of A previously computed, when calculating A(t) for long enough times:

$$\lim_{t\to\infty} A(t) = \lim_{t\to\infty}\left\{F\cdot e^{-(k_+ + k_-)t} + \frac{k_-}{(k_+ + k_-)}\cdot C_t\right\} = \frac{k_-}{(k_+ + k_-)}\cdot C_t = A^{ss} \qquad (1.14)$$

This example in which several assumptions have been used to simplify the process is in fact one the most simple examples of chemical networks. Biological networks are by far much more complex. Consequently, the resulting differential equation models are always nonlinear, which contrasts with the previous linear example. These complex systems of differential equations cannot be solved then analytically, but they will require the use of numerical simulation tools.

The issue with biological systems is that defining a priori the best timescale to analyze them is not always easy. Sometimes, one may want to get rid off of variables whose values are almost constant during the whole process. Such model reduction consists of assigning to those variables a constant value. In contrast, sometimes there are processes that occur much faster than the timescale one wants to consider [12]. In those situations the best option is to treat them as instantaneous processes. In order to mathematically describe those assumptions, there are approaches like *rapid equilibrium assumption* or *quasi-steady-state assumption.*

1.2.4 Analyzing dynamic mathematical models:

In this section, different techniques suitable for analyzing dynamics of mathematical models will be introduced. As it has been mentioned several times, ultimately the goal of mathematical models is to predict the temporal behavior of the system. This is achieved by obtaining the time varying concentrations - or discrete molecule population number - of each of components forming that system from a system of ordinary differential equations that is created accounting for the components interactions. If one solves a system of two equations with two species involved, in order to plot the results one can plot the concentration of each species as function of time. Another approach is to plot the concentration of one of the species against that one of the second species. The resulting generated space is called phase plane. The advantage of the phase plane plot is that it shows how concentrations evolve from the initial state at t=0 to the final steady state. Such evolving curve is known as trajectory. Phase plane plots accentuate the relationship between the variables in time, although it de-emphasizes of each of the species with time.

In the following sub-sections, the mathematical-geometrical analysis of systems will be presented starting with linear 2-D systems until reaching more complex systems.

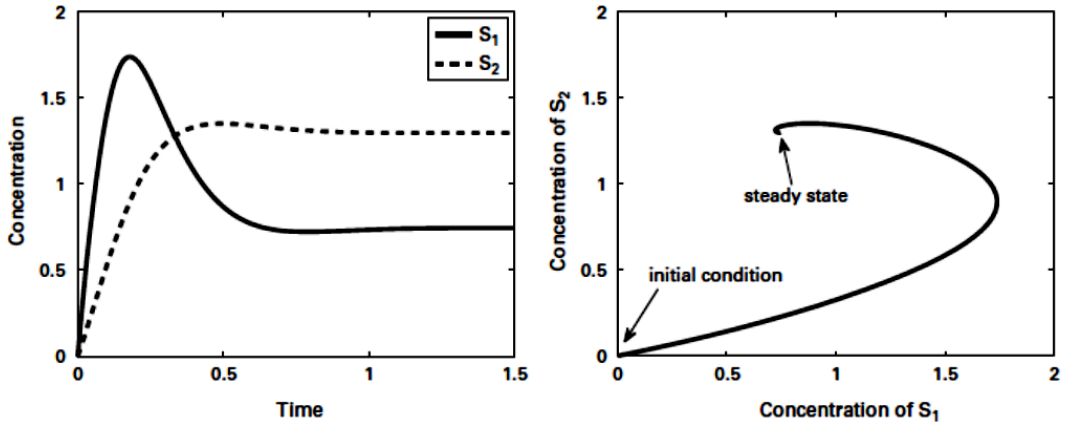

Fig. 1.4 Species concentration against time (left) and species 1 concentration against species 2 concentration or phase plane plot (right) [19].

1.2.4.1 Linear systems

Although biological systems are mainly nonlinear, linear systems' characteristics must be understood first, since the global behavior of systems can be inferred analyzing a set of key points. Studying the behavior of the system at such points is a local analysis. This enables to use linearization at those small regions. Consequently, it is mandatory to comprehend first the behavior of linear systems. Starting with a two-dimensional linear system, it can be written as:

$$\begin{matrix} \dot{x} = ax + by \\ \dot{y} = cx + dy \end{matrix} \quad \rightarrow \quad \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \rightarrow \quad \dot{\mathbf{x}} = A\mathbf{x} \qquad (1.15)$$

Where $\dot{x}$ and $\dot{y}$ are the time derivatives. Such system is linear because provided $\mathbf{x_1}$ and $\mathbf{x_2}$ are solutions, any linear combination of $\mathbf{x_1}$ and $\mathbf{x_2}$ will be a solution of the system.

There are special points that will determine the behavior of the system. They are known as *fixed points*. These points satisfy $\dot{\mathbf{x}} = 0$ (with $\dot{\mathbf{x}}$ being the vector $(\dot{x}, \dot{y})$) and thus, they represent equilibrium solutions of the system [19]. Consequently, whenever the system is at one of those fixed points, it will stay there for all time, explaining why they are also called steady or constant solution. Depending on whether trajectories are attracted to them in the phase plane they will be classified as *stable fixed points* in case they all converge at the fixed point or *unstable fixed points* if they are diverge.

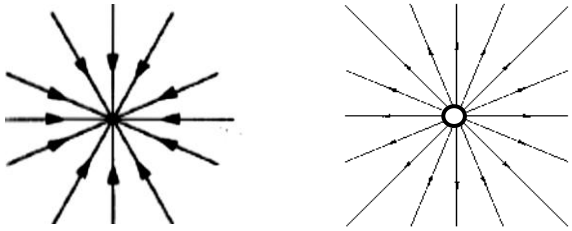

Fig. 1.5 Stable fixed point represented as a black-filled circle (left) and unstable fixed point –white-colored circle- (right) [19].

The stability analysis at these points allows one to infer the stability of the whole system itself. To do so, what needs to be done is to classify the linear system. In order to do it, searching for trajectories with the following form is critical:

$$\mathbf{x}(t) = e^{\lambda t}\mathbf{v} \tag{1.16}$$

In this solution v is a non-zero vector and λ can be understood as the growth rate and need to be determined [19]. If such solution exists, it will mean that there would be an exponential motion along the curve defined by v that will be pointing or "escaping" a given fixed point. Then, defined this, the problem now becomes finding such values. Notice that introducing that solution into the definition of a linear system, it results into:

$$\dot{\mathbf{x}} = A\mathbf{x} \tag{1.15}$$

$$\dot{\mathbf{x}} = \frac{\mathrm{d}}{\mathrm{dt}}\left(e^{\lambda t}\mathbf{v}\right) = \lambda e^{\lambda t}\mathbf{v} \tag{1.17}$$

$$A\mathbf{x} = Ae^{\lambda t}\mathbf{v} = e^{\lambda t}A\mathbf{v} \tag{1.18}$$

Then, the initial equation (1.15) becomes:

$$\lambda e^{\lambda t}\mathbf{v} = e^{\lambda t}A\mathbf{v}$$

$$A\mathbf{v} = \lambda\mathbf{v} \tag{1.19}$$

This is just the typical eigenvalue-eigenvector problem. From that point on, the traditional method for computing the eigenvalues can be used. The determinant of (A-λI) is computed obtaining the characteristic equation:

$$\lambda^2 - \underbrace{(a+d)}_{\tau:\, trace}\lambda + \underbrace{(ad-bc)}_{\Delta:\, delt} = 0 \tag{1.20}$$

The characteristic equation is a second order equation from which the resulting eigenvalues can be obtained as:

$$\lambda = \frac{\tau \pm \sqrt{\tau^2 - 4\Delta}}{2} \tag{1.21}$$

Eventually if one wants to write the general solution, it would resemble to this:

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t}\mathbf{v_1} + c_2 e^{\lambda_2 t}\mathbf{v_2} \tag{1.22}$$

Where c_1 and c_2 are constants that can be computed using the initial conditions. Additionally, the eigenvectors would be computed solving the equation (A- λI)$\mathbf{v}$=*0*.

However, regarding stability analysis is more important to understand what such eigenvalues mean. There are several options:

- Case 1: $\Delta < 0$: both eigenvalues are real and have opposite signs[1]. That fixed point is called a *saddle node*. These types of points are stable - attract neighboring trajectories - in the direction with the negative eigenvalue and unstable –repelling trajectories- in the eigenvector associated to the positive eigenvalue.

- Case 2: $\Delta > 0$: both eigenvalues are real or complex.
  - If $\tau^2 - 4\Delta > 0$, both are real. Those points are called *nodes*.
  - If $\tau^2 - 4\Delta < 0$, both are complex. They are called *spirals*.

  In the case, in order to analyze the stability, one has to see whether one has to analyze τ. If $\tau > 0$, both eigenvalues have positive real parts, thus the fixed point is unstable. If the opposite case $\tau < 0$, it would be a stable spiral or node.

- Case 3: $\Delta = 0$: In that situation, at least one eigenvalue is 0. This means that there is a line or a plane of fixed points [19].

This classification is represented in the following diagram:

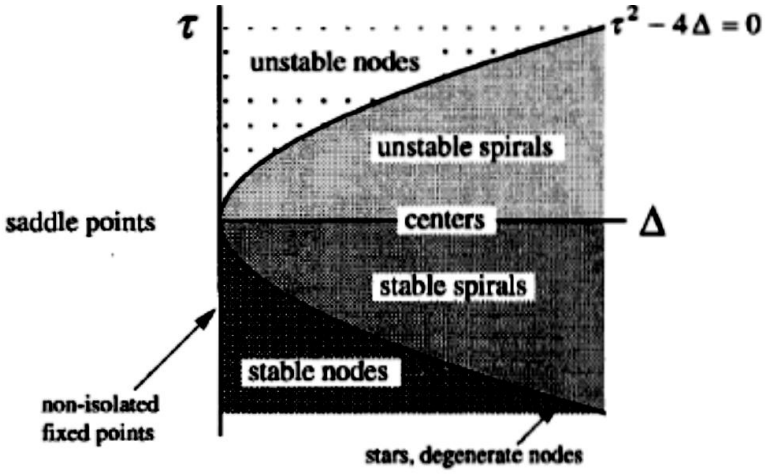

Fig. 1.6 Fixed points classification diagram [19]

Inside this classification, notice that there are limiting cases - the frontier lines between different regions - but, since they will not be used in this bachelor thesis, they will be left for other studies.

Herein the basic tools for the analysis of linear systems - phase plane plots and stability analysis -. Nonetheless, nonlinear biological systems need more tools to be properly studied.

[1] The characteristic equation can be written as $(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - \tau\lambda + \Delta$. Then, $\Delta = \lambda_1\lambda_2$ and $\tau = \lambda_1 + \lambda_2$

1.2.4.2 Linearization

Most solutions of nonlinear systems cannot be extracted analytically. The only way to address such problems is by means of numerical methods. In spite of this, nonlinear systems can be analyzed from a quantitative perspective. This corresponds to obtaining the phase plane without the need of computing the solutions to the system. As said before, when dealing with linear systems, classification of fixed point in the phase plot is almost straightforward, it just requires solving an eigenvector-eigenvalue problem. Nevertheless, the same process cannot exactly be applied to nonlinear systems.

In this way, linearization can be thought as the pre-processing step that must be done when analyzing nonlinear systems' stability. The hypothesis would be that the phase plane plot located close to the fixed points of the nonlinear system is similar enough so that it can be approximated by that of the corresponding linear system [19]. At the end, the algorithm behind linearization results into the substitution of our initial matrix *A* in equation 1.15 by the Jacobian matrix. Once that matrix is calculated, it will have to be evaluated at all fixed points in order to classify them.

Then, starting with a 2D nonlinear system defined as:

$$\begin{aligned} \dot{x} &= f(x, y) \\ \dot{y} &= g(x, y) \end{aligned} \tag{1.23}$$

With (*x**, *y**) as the fixed point,

$$f(x^*, y^*) = g(x^*, y^*) = 0 \tag{1.24}$$

Then, introducing some disturbance close to the fixed point, two additional components are described:

$$u = x - x^* \quad v = y - y^* \tag{1.25}$$

With this change, $\dot{u}$ and $\dot{v}$ can be expressed as:

$$\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \underbrace{\begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix}}_{Jacobian} \begin{pmatrix} u \\ v \end{pmatrix} + quadratic\ terms \tag{1.26}$$

If quadratic terms are neglected, the resulting system is a linearized system whose fixed points can be classified applying the criterion for linear systems to the Jacobian evaluated at the fixed points.

The neglection of the quadratic terms is a reasonable assumption provided that the eigenvalues have a nonzero real part. When this does not apply, *Re(λ)=0*, that

assumption is not valid since the fixed points correspond to those lying on the frontier between different regions. In those situations, other approaches need to be used:

1.2.4.3 Limit cycles

Limit cycles are defined as isolated closed trajectories, meaning that trajectories near them are attracted or repelled by them, so they are not closed [19]. The concept of stability in limit cycles is relatively similar to that of fixed points. Attracting limit cycles are called stables and repulsive ones are known as unstable.

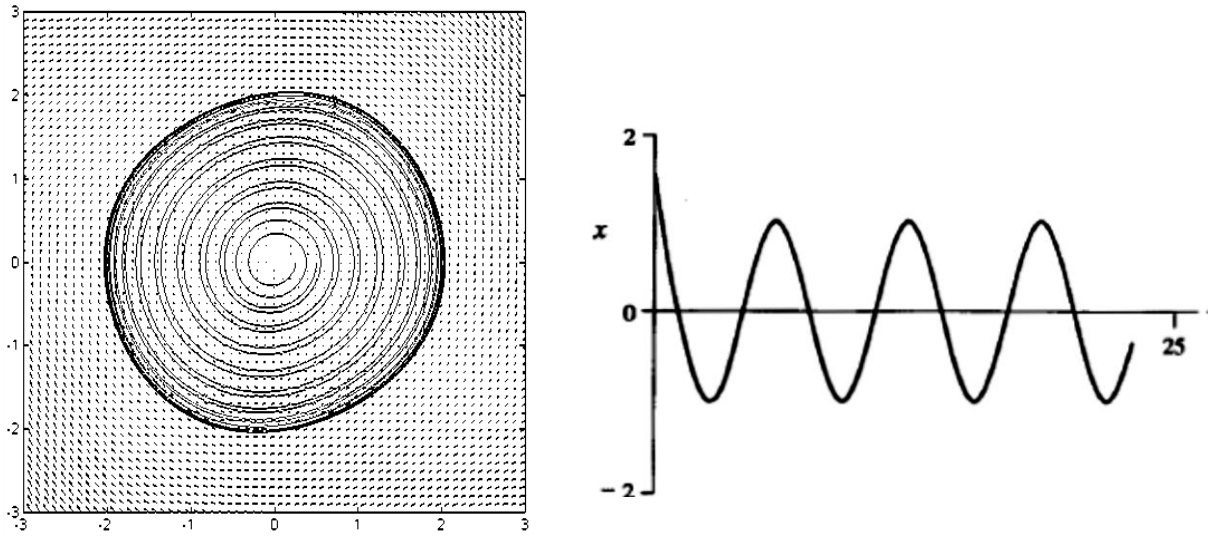

Fig. 1.7: Limit cycle in phase plane plot representation (left) and conventional time visualization (right) [12]

The role of limit cycles in biological systems is critical as it will be explained in the following sections. The reason is that they are the phase plane representation of systems with self-sustained oscillations [19], with periodic solutions, which are characteristics of biological processes like transcription and translation in genetic networks. Other examples of limit cycles are circadian rhythms or heart beating. Another feature of limit cycles is that they are a nonlinear phenomenon.

When analyzing nonlinear systems there are strategies used to discard the existence of limit cycles (*gradient systems, Lyapunov functions* or *Dulac's criterion*) as well as a theorem to determine that closed orbits exist in the given system. Such theorem is known as *Poincaré-Bendixson theorem.*

Then, limit cycles can be described as isolated trajectories in the phase plane causing neighboring trajectories to spiral into it or out of it as time tends to infinity.

1.2.4.4 Bifurcations

So far, when analyzing dynamical systems, all parameters associated to the state variables were assumed to be both fixed. However, as mentioned in section 2.1.2.1, it is possible to gradually change some parameters for each simulation so that the system's behavior can be studied under different environmental conditions. Those parameters are called *control parameters* and play an important role in the stability of the model, as they will be responsible for their creation or destruction of fixed points and limit cycles. These qualitative changes in the stability of the systems are called *bifurcations*. In this epigraph, some of the most important bifurcations will be presented.

- Saddle node bifurcation: it is the basic mechanism by which fixed points are created or destroyed [19]. Imagine a phase plane plot with two fixed points, one stable and another unstable. Then, as the chosen control parameter is changed, both fixed points approach to each other. Eventually, they collide and disappear.

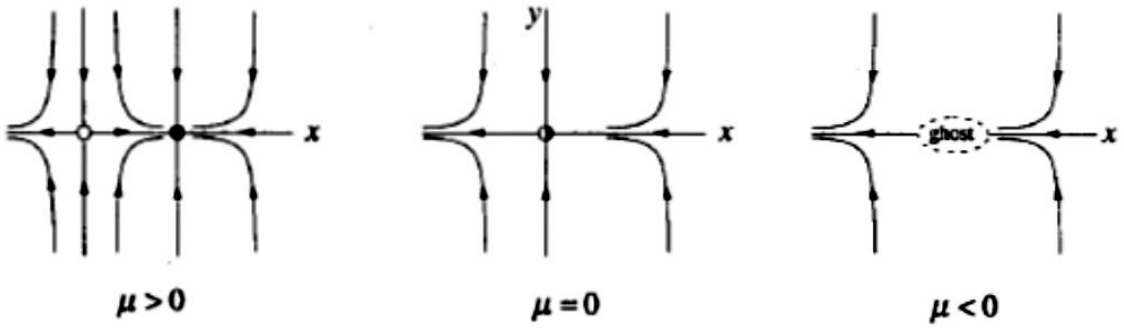

Fig. 1.8 Saddle node bifurcation phase plane plot. As the control parameter decreases, the fixed points become closer to each other until finally disappear. This bifurcation appears in systems with the form: $\dot{x} = \mu - x^2;\ \dot{y} = -y$ [19]

- Transcritical and pitchfork bifurcations: sometimes one fixed point will exist for all values of the parameters. In this case, changing the value of the control parameter will not create nor destroy such fixed point but it will modify its stability. This is what happens in transcritical bifurcations. On the other hand, in pitchfork bifurcations, one fixed point gives rise to three fixed points as the parameter is changed. This bifurcation commonly appears in systems showing symmetry.

  Subcritical and supercritical features are given by the cubic term in pitchfork bifurcation. When it is positive - subcritical bifurcation - there is an abrupt change transition in the fixed point that changes from stable to unstable. Such blow up effect causes *x(t)* to jump to infinite in a finite time. That is the reason why subcritical pitchfork bifurcations are unwanted in biological systems, as they imply that small disturbances can destabilize the whole system.

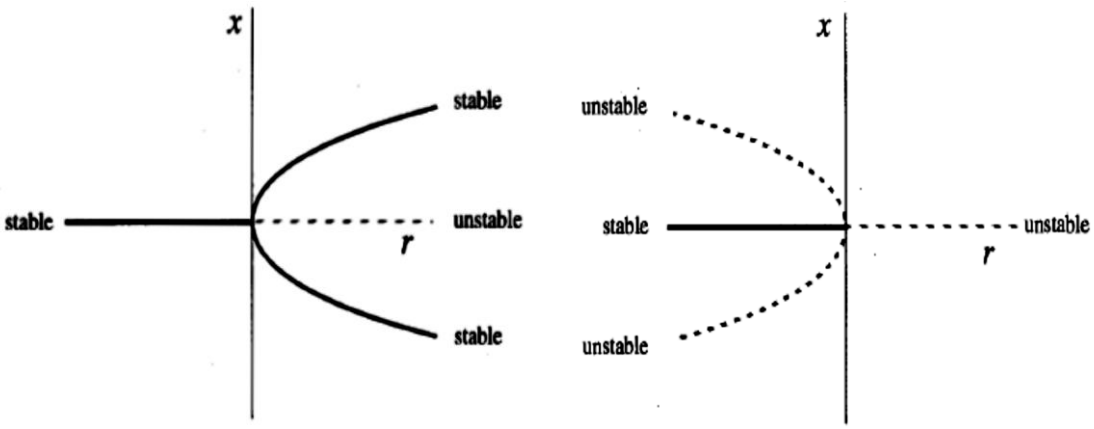

Fig. 1.9 Supercritical pitchfork (left) and Subcritical pitchfork (right) bifurcation diagrams. Their prototype formulas are: $\dot{x} = rx - x^2$; $\dot{x} = rx - x^3$; $\dot{x} = rx + x^3$ $with$ $\dot{y} = -y$ respectively [19]

- Hopf bifurcation: there is another way fixed points can change their stability. Let's assume that the system under study has a stable fixed point. This automatically implies that its eigenvalues have negative real parts. To be more precise, there are only two options: the eigenvalues are both real and negative or they are complex eigenvalues with negative real parts. Either way, the eigenvalues of the fixed points must move to the positive real half plane (right plane) in order to change its stability. The bifurcations explained before correspond to such transition in real eigenvalues. Hopf bifurcation appears when the real part of complex eigenvalues crosses to the positive plane [19] (Fig. 1.10). In this way, similarly to pitchfork bifurcations, there are *supercritical* and *subcritical Hopf bifurcations*.

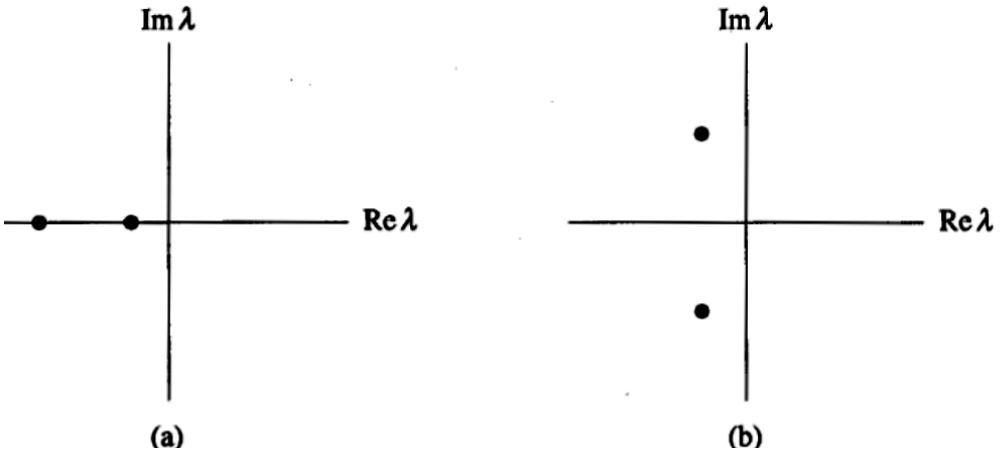

Fig. 1.10 Real eigenvalues in the negative half plane (a) and complex eigenvalues with negative real parts (b) [19].

Supercritical Hopf bifurcation occurs when a stable spiral turns into an unstable spiral surrounded by a limit cycle. In other words, for certain values of the parameter, the fixed point behaves as a sink, attracting neighboring trajectories in a circular way (stable spirals). Initially, they are attracted strongly, which in

mathematical terms means exponentially, but as parameter changes such attraction becomes weaker – algebraic - until eventually when the eigenvalues acquire positive real parts the fixed point reverses becoming into unstable. Then, it behaves as an unstable spiral constricted within a stable limit cycle.

Fig. 1.11 Supercritical Hopf bifurcation. When the bifurcation point is crossed, the stable spiral (left) becomes into a unstable spiral surrounded by a stable limit cycle. Possible system: $\dot{r} = \mu r - r^3$; $\dot{\theta} = \omega + br^2$ [19].

From an oscillatory perspective, supercritical Hopf bifurcations can be understood in the following way: initially, the system exhibits oscillations that progressively become smaller in amplitude until converging into a steady state -damped oscillations-. However, when the control parameter crosses the bifurcation point, the system exhibits the opposite effect: it starts with a constant value and small-amplitude oscillations appear resulting into limit cycle oscillations around the steady state [19].
The example described above represents an ideal case in which the eigenvalues have the following form:

$$\lambda = \mu \pm i\omega \tag{1.27}$$

Consequently they cross the imaginary axis as μ positively increases. Although idealized, it allows to extract important conclusions about supercritical Hopf bifurcations. The first one is that the amplitude of the limit cycle continuously increases from zero, being proportional to $\sqrt{\mu - \mu_c}$ near the bifurcation point (μ) [19]. The second one is that the frequency of oscillations located at the bifurcation point is $\omega = Im(\lambda)$ being the period is $T = {}^{2\pi}/_{Im(\lambda)} + O(\mu - \mu_c)$.
Obviously, the example is ideal but there are some missing points, like the fact that most Hopf bifurcations do not have circular limit but elliptical cycles. In spite of this, those two conclusions can be safely generalized for most situations.
The second type of Hopf bifurcation is the *subcritical Hopf bifurcation*. As stated before, this case can be potentially dramatic when facing it in real systems. In this case, the phase plane plot is different. Initially, for a certain parameter range there is a stable fixed point that co-exists with two concentric

limit cycles, the inner one unstable and the outer one stable. Consequently, trajectories whose initial conditions lie between the fixed point and the unstable limit cycle will be circularly attracted towards the stable fixed point (stable spirals). In contrast, those trajectories starting out of the unstable limit cycle will move to the outer stable limit cycle. In this way, at the same time there would be two stable states: one steady-state and an oscillatory state. However, when the control parameter changes what happens is that the inner limit cycle tightens each time more becoming closer to the stable fixed point. At the end, there is annihilation between them when the amplitude of the unstable limit cycle reaches zero, it collides with the stable fixed point. This event produces a blow-up effect since solutions that used to be close to that fixed point are suddenly forced to become large-amplitude oscillations [19] attracted to the stable limit cycle.

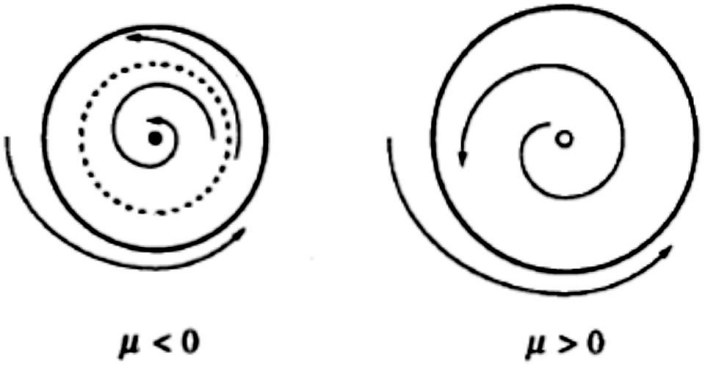

Fig. 1.12 Subcritical Hopf bifurcation phase plant plot [19].

- Other bifurcations: Beside Hopf bifurcations there are other bifurcations involved in the destruction and creation of limit cycles like *saddle-node bifurcation of cycles*, the *infinite period bifurcation* or the *homoclinic bifurcation*, each of them with their associated characteristics and established amplitude and period of the limit cycles.

1.2.5 Gene regulatory networks

Our system is a gene regulatory network, which is defined as a set of genes whose derived proteins regulate the expression of some of each other.
Gene expression is a process divided into two stages: transcription and translation. The first term refers to the process by which DNA is converted into mRNA. This process is performed by an enzyme called *RNA polymerase* that binds to the promoter region in the gene [12]. Translation is the process by which ribosomes read the mRNA synthesizing the corresponding protein. What makes gene regulatory networks different from other chemical reaction networks is the high number of reactions they involve. Additionally, the mentioned *continuum hypothesis* - assuming that the number of

reacting molecules is high enough so that it is possible to described species abundance using concentration - requires to be revisited. These genetic processes affect to a relatively low number of molecules, something that in principle would disable such hypothesis. In addressing the behavior of systems with low molecule counts, we can justify the mass-action formalism by interpreting differential-equation models as descriptions of the average behavior over a large population of cells [12].

In regulated gene expression (Fig. 1.13), such regulation occurs by controlling the initiation of transcription.

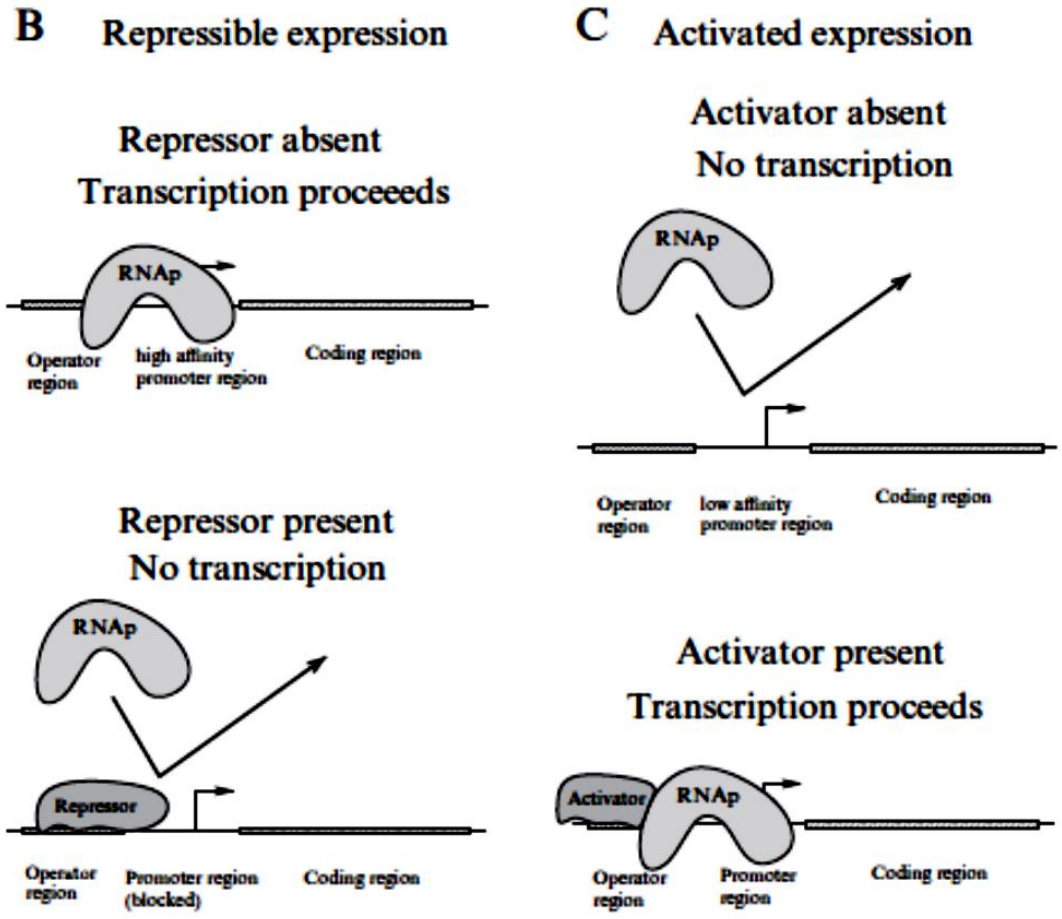

Fig 1.13 Diagram showing regulated gene expression by activator and repressor [12].

Within prokaryotes that happens when the enzyme initiating the transcription, *RNA-polymerase*, binds to promoter region in the gene. Such association event is influenced by a second protein, called transcription factor that binds to the operator regions -nearby promoter region -. That transcription factor can be an *activator* or a *repressor* of the transcription. Even in the absence of the activator, a gene can still be expressed, therefore, a term referring to the basal transcription rate must also be considered.

it is possible to express the binding of a transcription factor as:

$$O + P \underset{d}{\overset{a}{\rightleftharpoons}} OP$$

Where O is the operator, P is the activator or repressor and OP is the combination of both. As the binding-unbinding process occurs much faster than the gene expression itself, it can be assumed to be in equilibrium [12]. In a parallel way, the number of operators that have attached transcription factors is assumed to be proportional to the transcription rate. To model it, it can be assumed to depend on that fraction of bound operators:

$$Transc.rate = \alpha_O + \alpha \cdot Fraction\ of\ OP = \alpha_O + \alpha \frac{[OP]}{[O] + [OP]} \qquad (1.28)$$

Taking into account that:

$$K = \frac{d}{a} = \frac{[O][P]}{[OP]} \rightarrow [OP] = \frac{[O][P]}{K} \qquad (1.29)$$

Introducing it into equation 1.28:

$$Transc.rate = \alpha_O + \alpha \frac{[O][P]/K}{[O] + [O][P]/K} = \alpha_O + \alpha \frac{[P]/K}{1 + [P]/K} \qquad (1.30)$$

Where α_O is the basal transcription rate and α the maximal transcription efficiency.

Transcription of a gene can also be regulated by a repressor –now the transcription rate would be proportional to the number of unbound operators, by multiple transcription factors. The level of complexity can even increase more if one accounts for cooperativity.

The simplest example of genetic regulatory network is that one of gene whose translated protein auto-inhibits or auto-activate its expression.

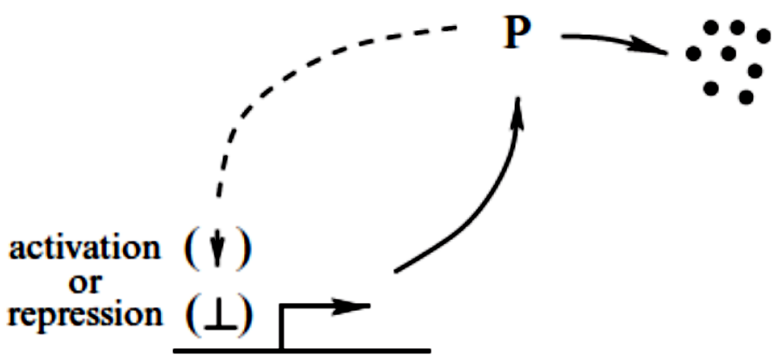

Fig. 1.14: Auto-inhibition/activation diagram of a genetic circuit [12].

The auto-inhibition can be modeled as:

$$\frac{dp(t)}{dt} = \alpha \frac{1}{1 + p(t)/K} - \delta_p p(t) \tag{1.31}$$

Where *p(t)* is the concentration of transcription factor, K is the equilibrium constant - assuming the binding is much faster than the gene expression itself - and δ_p accounts for the degradation of the repressor.

Other examples of important gene regulatory networks include genetic switches like the *Lac Operon* or oscillatory gene networks like the *Goodwin oscillator*. Within the last group, it is possible to find synthetic oscillatory networks over which this thesis will revolve around.

1.3 Project background

The work developed in this bachelor thesis is focused on the article *"A fast, robust and tunable synthetic gene oscillator"* (2008) [20]. This publication, together with *"Robust, tunable biological oscillations from interlinked positive and negative feedback"* (2008) [21] constitute the main research lines adopted in this bachelor thesis. Additionally, the third pillar of this thesis comes from the studies and experiments, Saúl Ares had performed related to those scientific articles.
Articles [20] and [21] highlighted the idea that the combination of positive and negative feedback loops in gene regulatory networks enabled to increase the robustness of oscillations. [21] studies such hypothesis in real biological oscillators like circadian rhythms or mammalian heart rate, [20] applies that hypothesis to the construction of a synthetic oscillator. The fact that sustained oscillations could be generated by negative feedback loops was already known in systems biology, however, there were many biological oscillators that not only had a negative feedback loop but also a positive one [21]. The advantages and the implications of such characteristic were poorly understood until both articles were published. Both concluded that the advantage of mixed feedback regulatory networks over only negative-feedback networks was found on the greater tunability of systems with positive feedback loops. While frequency was difficult to adjust in negative feedback loops without dramatically affecting the amplitude, the addition of the positive feedback diminished such dependency. This advantage would contribute to explain the predominant existence of mixed feedback loops in biological processes required to operate at a different range of frequencies but giving the same output, like the heart beating.
In [20], researchers arrived to the same conclusion by showing that the deletion of the positive feedback loop avoided the possibility of tuning the oscillation period - and thus, frequency -. Thus, the contribution of both investigations is fundamental for the future

of synthetic biology, the design of a genetic “clock” - a synthetic oscillator - that would release drugs when needed or express a given protein, like insulin, every certain time period [22].

1.4 Objectives

The main objective of this bachelor thesis is to analyze the synthetic oscillator developed in the article “*A fast, robust and tunable synthetic gene oscillator”* [20]. This general objective can be divided into several subgoals:

1. To understand from a biological point of view all the components and interactions within this biological system.

2. To develop a mathematical model for this synthetic oscillator.

3. To study the changes in the qualitative behavior of the system when varying the control parameters.

4. To analyze the roles of the positive and negative feedback loops in the system and comparing those results with the original article.

5. To develop a mathematical interpretation of the results.

2. Methods:

This chapter describes how the mathematical model is built as well as how the simulations are implemented. Initially, the biological network to be modeled will be described and then the derivation of the differential equations will be presented. The following reactions, diagrams and ODEs are based on the original scientific article [20].

2.1 Step 1: Biological network to be modeled

As stated in the project background, the biological system to be modeled corresponds to a relaxation synthetic genetic oscillator in *Escherichia coli*, one of the most widely used model organisms in molecular biology. This oscillator involves two genes whose resulting proteins act as a repressor and an activator for gene expression.

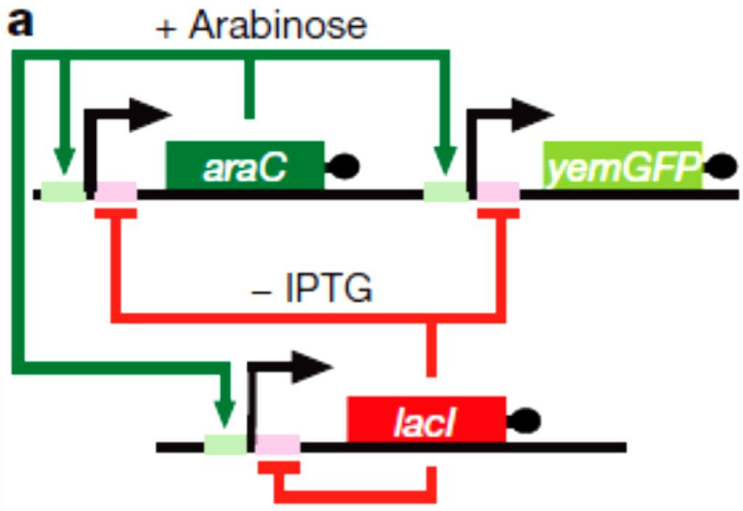

Fig. 2.1 Network diagram of relaxation oscillator [20].

As it can be seen in the diagram (Fig. 2.1), the synthetic gene network is governed by two genes *AraC* and *LacI*, each of them with the same hybrid promoter ($p_{lac/ara-1}$) formed by the activator operator site - from *araBAD promoter* - and the repressor operator site - from the *lacZYA* promoter - [20]. *AraC* and *LacI* products - *AraC* and *LacI* proteins - act as the transcription factors of that hybrid promoter: while *AraC* protein activates the expression of both genes in the presence of arabinose, *LacI* protein represses the expression in the absence of IPTG (Fig. 2.1).

This dual-regulation of the promoter represents the co-existence of both negative and positive feedback loops.

On one hand, AraC protein production together reinforces promoter activity - as long as there is arabinose and IPTG -. This process represents the positive feedback loop. At the same time, there is a parallel increment in LacI protein that induces promoter repression when IPTG diminishes. This second process is the negative feedback loop. The difference in activity between both loops will be responsible for the oscillatory behavior of the network [20]. Additionally, there is a third gene with the same promoter whose expression will produce a fluorescent protein that was used by the authors to experimentally measure the amplitude and period of oscillations.

From an experimental perspective, notice that arabinose is a monosaccharide used as a culture medium for certain bacteria, while IPTG is a molecular mimic of *allolactose.* The latter one is a lactose metabolite that activates transcription of the *lac operon* by binding to the *lac* repressor, which explains why it can be used to regulate gene expression (Fig. 2.15)

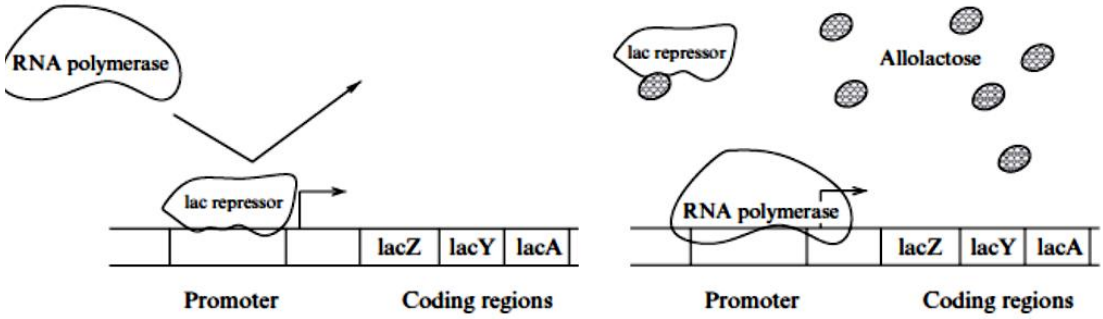

Fig. 2.2 Diagram showing the effect of allolactose –IPTG- on gene expression. If it is present, it inhibits the binding of the repressor to the promoter. If it is absent, lac repressor can bind to the promoter inhibiting gene expression [12].

2.2 Step 2: Identification of chemical reactions

In the original research *araC* gene and the gene coding for the fluorescent protein were cloned in the same plasmid –activator/reporter plasmid while the *lacI* gene was placed in a different plasmid - the repressor plasmid - . In both cases the cloned genes were controlled by the hybrid promoter $p_{lac/ara-1}$. Although it can be done in a different way, for instance Ares cloned them within the same plasmid. The mathematical model was built according to this two-plasmid experimental design, although the separation of the genes into two plasmids is not strictly necessary as commented with by Ares.
In our system, it is possible to establish four different groups: reactions involved in the binding of the transcription factors to the promoter, transcription reactions, translation and post-translational reactions and degradation processes.

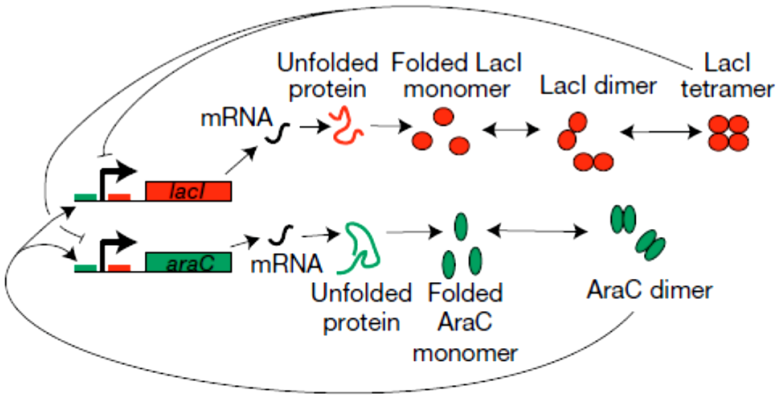

Fig. 2.3 Higher level diagram showing intermediate processes involved in the gene circuit [20].

The reactions responsible for the binding of the transcription factors to the hybrid promoters are the following ones:

$$P_{0,j}^{a/r} + a_2 \underset{k_{-a}}{\overset{k_a}{\rightleftharpoons}} P_{1,j}^{a/r} \qquad j \in \{0,1,2\} \tag{I}$$

$$P_{i,0}^{a/r} + r_4 \underset{k_{-r}}{\overset{2k_r}{\rightleftharpoons}} P_{i,1}^{a/r} \qquad i \in \{0,1\} \tag{II}$$

$$P_{i,1}^{a/r} + r_4 \underset{2k_{-r}}{\overset{k_r}{\rightleftharpoons}} P_{i,2}^{a/r} \qquad i \in \{0,1\} \tag{III}$$

$$P_{1,2}^{a/r} \overset{k_l}{\rightarrow} P_{L,2}^{a/r} + a_2 \tag{IV}$$

$$P_{0,2}^{a/r} \overset{k_l}{\rightarrow} P_{L,2}^{a/r} \tag{V}$$

$$P_{L,0}^{a/r} \overset{k_{ul}}{\rightarrow} P_{0,0}^{a/r} \tag{VI}$$

Where, $P_{i,j}^{a/r}$ refers to the promoter state in each of the plasmids - $P_{i,j}^{a}$ corresponds to the activator and $P_{i,j}^{r}$ to the repressor -. The sub-indexes coupled to each promoter define the state of the promoter where "i" indicates the number of AraC dimers, a_2 (only one dimer can bind the promoter), while the "j" represents the number of LacI tetramers, r_4, bound to the promoter (up to two tetramers can bind the promoter). Notice that reactions I,II and III represent a set of reactions. Nevertheless, the latter derived differential equations will have to explicitly consider all possible promoter states. On the other hand, $P_{L,j}^{a/r}$ accounts for DNA in its loop form [23]. Whenever in that form, AraC dimers are unable to bind the promoter. At the same time, these DNA loops are assumed

to occur just when the promoter has bound two LacI tetramers. Then, reaction VI corresponds to the unlooping reaction by which DNA would come to its normal state, and the reactions IV and V represent the looping processes. Contrary to the rest of reactions, looping and un-looping reactions are assumed to be irreversible. This assumption holds since reverse reaction rates were much slower than the forward rate, explaining why the researchers decided to neglect them.

Regarding the rate constants, k_a and k_r are defined as the forward binding rates of the activator and repressor to the operator site [23]. In the same way, k_{-a} and k_{-r} represent the rate of unbinding of the activator and repressor from the promoter. Finally, k_l and k_{ul} are the looping and unlooping reaction rates, respectively.

So far, the reactions I-VI described the binding processes between the transcription factors and the promoter. The following reactions describe the transcription from DNA to mRNA, $m_{a/r}$.

$$P_{0,0}^{a/r} \xrightarrow{b_{a/r}} P_{0,0}^{a/r} + m_{a/r} \quad \text{(VII)}$$

$$P_{1,0}^{a/r} \xrightarrow{\alpha b_{a/r}} P_{1,0}^{a/r} + m_{a/r} \quad \text{(VIII)}$$

Reaction VII represents the basal transcription, occurring in the absence of activator (AraC protein), while reaction VIII corresponds to the gene expression increase due to the activator-promoter binding (the promoter state is $P_{1,0}^{a/r}$ indicates that it has attached an AraC dimer). b_a and b_r are the basal transcription rates of the activator and repressor genes, while the product of the previous constants with α gives the transcription rates when the promoter has bound to the activator transcription factor -a_2-.

$$m_a \xrightarrow{t_a} m_a + a_{uf} \quad \text{(IX)}$$

$$m_r \xrightarrow{t_r} m_r + r_{uf} \quad \text{(X)}$$

$$a_{uf} \xrightarrow{k_{fa}} a \quad \text{(XI)}$$

$$a + a \underset{k_{-da}}{\overset{k_{da}}{\rightleftharpoons}} a_2 \quad \text{(XII)}$$

$$r_{uf} \xrightarrow{k_{fr}} r \quad \text{(XIII)}$$

$$r + r \underset{k_{-dr}}{\overset{k_{dr}}{\rightleftharpoons}} r_2 \quad \text{(XIV)}$$

$$r_2 + r_2 \underset{k_{-tr}}{\overset{k_t}{\rightleftharpoons}} r_4 \quad \text{(XV)}$$

Reactions IX-XV correspond to the translational and post-translational modification of the *LacI* and *AraC* proteins. Initially, mRNA is translated by ribosome into proteins. However, these initial proteins are unfolded - a_{uf} and r_{uf} -. Consequently, as it was described in the diagram of Fig. 2.3 the products of the translation are folded so that they can properly perform the corresponding function. Notice that proteins are firstly synthesized in its primary structure, which is just a linear sequence of aminoacids. In order to become functional they must at least adopt their secondary structure. Thus, such event is described by the two folding reactions (XI and XIII). After that, the monomers of LacI and AraC proteins (*a* and *r)* are functional. Nevertheless, they still require additional reactions. AraC monomers will undergo a reversible dimerization reaction - a_2- , while LacI monomers will be first grouped into dimers -r_2- and later into tetramers -r_4-.

Within this set of reactions, t_a and t_r are the translation rates of the AraC and LacI mRNA into AraC and LacI proteins and k_{fa} and *kfr* represent the folding rates of the previous protein precursors. Moreover, forward and inverse dimerization rates must be established in both the activator and the repressor proteins (k_{da}, k_{-da}, k_{dr} and k_{-dr}). In the case of LacI dimers, they can also be grouped into tetramers with tetramerization rates k_t and k_{-t}.

The degradation reactions are described with the following terms:

$$m_{a/r} \xrightarrow{d_{a/r}} \emptyset \quad \text{(XVI)}$$

$$a_{uf} \xrightarrow{\lambda f(X)} \emptyset \quad \text{(XVII)}$$

$$r_{uf} \xrightarrow{f(X)} \emptyset \quad \text{(XVIII)}$$

$$a \xrightarrow{\lambda f(X)} \emptyset \quad \text{(XIX)}$$

$$r \xrightarrow{f(X)} \emptyset \quad \text{(XI)}$$

$$a_2 \xrightarrow{\lambda f(X)} \emptyset \quad \text{(XXII)}$$

$$r_2 \xrightarrow{f(X)} \emptyset \quad \text{(XXIII)}$$

$$r_4 \xrightarrow{f(X)} \emptyset \quad \text{(XXIV)}$$

$$P_{1,j}^{a/r} \xrightarrow{f(X)} P_{0,j}^{a/r} \qquad j \in \{0,1,2\} \qquad \text{(XXV)}$$

$$P_{i,1}^{a/r} \xrightarrow{f(X)} P_{i,0}^{a/r} \qquad i \in \{0,1\} \qquad \text{(XXVI)}$$

$$P_{i,2}^{a/r} \xrightarrow{2f(X)} P_{i,1}^{a/r} \qquad i \in \{0,1\} \qquad \text{(XXVII)}$$

$$P_{L,2}^{a/r} \xrightarrow{2\epsilon f(X)} P_{L,1}^{a/r} \qquad \text{(XXVIII)}$$

$$P_{L,1}^{a/r} \xrightarrow{\epsilon f(X)} P_{L,0}^{a/r} \qquad \text{(XIX)}$$

Notice that this system is considered to be open because of the degradation reactions (XVI-XXIV). mRNA transcripts decay exponentially with degradation rates d_a and d_r. The different protein forms are degraded through enzymatic reactions (XVIII-XXIV). These degradation reactions are crucial when modeling the behavior of the gene network. Proteins functioning as transcription factors - AraC and LacI proteins - are characterized by rapid degradation reactions. The reason behind such speed is enabling cells to respond quickly to external stimuli [24]. In an attempt to improve cell's responsiveness to the changing environment, researchers added degradation tags to the proteins in the network - *ssrA* tags [23] -, peptide sequences genetically grafted onto these proteins. These degradation tags are responsible for the variable *f(X)* appearing as a rate constant in the previous degradation reactions. That constant can be defined as:

$$f(X) = \frac{\gamma}{C_e + X} \qquad (2.1)$$

Where X is the total number of *ssrA* tags in the system [23], γ represents the maximum active degradation rate of the tag and C_e is the number of *ssrA* tags so that degradation rate is reduced by half. Additionally, there are two other parameters that need to be established, λ which is the maximum degradation rate of activator protein forms and ϵ, a correcting factor included when the degradation of tetramers occurs in any of the DNA loops forms.

2.3 Step 3: Obtaining differential equations

Once all the reactions have been introduced, the next step is the development of the mathematical model. At stated in the introduction, the two possible options were deterministic or stochastic simulations. The former one was the chosen one. To do so, the resulting deterministic ordinary differential equations were constructed on the grounds of the law of mass action where all equations are deterministic. According to this law, *the rate of any chemical reaction is proportional to the product of the concentrations of the reactants*. Additionally, the second assumption made will be the

continuum hypothesis. Under this second condition, it will be possible to work with molecule concentrations as long as the number of molecules is high enough. It is true that this assumption sometimes moves away from reality when modeling genetic networks; however it can still apply if one considers that the resulting differential equations do not just model the behavior in a single cell, but over a large population of cells (see 1.2.5 *Gene Regulatory Networks*).

With those concepts in mind, it is possible to write the differential equations that will describe the temporal behavior of the system. Although the hybrid promoter is the same for both genes, since the genes have been cloned into two different plasmids, there will be two set of equations describing the variation in the concentration of the promoter on each of the plasmids. The equations describing the time variation of promoters in a given state are:

$$\frac{dP^a_{0,0}(t)}{dt} = -k_a P^a_{0,0}(t)a_2(t) + k_{-a}P^a_{1,0}(t) - 2k_r P^a_{0,0}(t)r_4(t) + k_{-r}P^a_{0,0}(t) + k_{ul}P^a_{L,0}(t) + f(X)P^a_{1,0}(t) + f(X)P^a_{0,1}(t) \quad (2.2)$$

$$\frac{dP^a_{1,0}(t)}{dt} = k_a P^a_{0,0}(t)a_2(t) - k_{-a}P^a_{1,0}(t) - 2k_r P^a_{1,0}(t)r_4(t) + k_{-r}P^a_{1,1}(t) - f(X)P^a_{1,0}(t) + f(X)P^a_{1,1}(t) \quad (2.3)$$

$$\frac{dP^a_{0,1}(t)}{dt} = -k_a P^a_{0,1}(t)a_2(t) + k_{-a}P^a_{1,1}(t) + 2k_r P^a_{0,0}(t)r_4(t) - k_{-r}P^a_{0,1}(t) - k_r P^a_{0,1}(t)r_4(t) + 2k_{-r}P^a_{0,2}(t) + f(X)P^a_{1,1}(t) - f(X)P^a_{0,1}(t) + 2f(X)P^a_{0,2}(t) \quad (2.4)$$

$$\frac{dP^a_{1,1}(t)}{dt} = k_a P^a_{0,1}(t)a_2(t) - k_{-a}P^a_{1,1}(t) + 2k_r P^a_{1,0}(t)r_4(t) - k_{-r}P^a_{1,1}(t) - k_r P^a_{1,1}(t)r_4(t) + 2k_{-r}P^a_{1,2}(t) - f(X)P^a_{1,1}(t) - f(X)P^a_{1,1}(t) + 2f(X)P^a_{1,2}(t) \quad (2.5)$$

$$\frac{dP^a_{1,2}(t)}{dt} = -k_a P^a_{0,2}(t)a_2(t) - k_{-a}P^a_{1,2}(t) + k_r P^a_{1,1}(t)r_4(t) - 2k_{-r}P^a_{1,2}(t) - k_l P^a_{1,2}(t) - 2f(X)P^a_{1,2}(t) - f(X)P^a_{1,2}(t) \quad (2.6)$$

$$\frac{dP^a_{0,2}(t)}{dt} = -k_a P^a_{0,2}(t)a_2(t) + k_{-a}P^a_{1,2}(t) + k_r P^a_{0,1}(t)r_4(t) - 2k_{-r}P^a_{0,2}(t) - k_l P^a_{0,2}(t) + f(X)P^a_{1,2}(t) - 2f(X)P^a_{0,2}(t) \quad (2.7)$$

$$\frac{dP^a_{L,0}(t)}{dt} = -k_{ul}P^a_{L,0}(t) + \varepsilon f(X)P^a_{L,1}(t) \quad (2.8)$$

$$\frac{dP^a_{L,1}(t)}{dt} = 2\varepsilon f(X)P^a_{L,2}(t) - \varepsilon f(X)P^a_{L,1}(t) \quad (2.9)$$

$$\frac{dP_{L,2}^{a}(t)}{dt} = k_l P_{0,2}^{a}(t) + k_l P_{1,2}^{a}(t) - 2\varepsilon f(X) P_{L,2}^{a}(t) \tag{2.10}$$

These first nine equations (2.2 – 2.10) represent the change in the concentration of promoters at a given state. Notice also that they correspond to promoters found in activator plasmids. There is an identical set of equations for the promoters in repressor plasmids (see Appendix A).

Those equations corresponding to mRNAs and the different protein forms are written as:

$$\frac{dm_a(t)}{dt} = b_a P_{0,0}^{a}(t) + \alpha b_a P_{1,0}^{a}(t) - d_a m_a(t) \tag{2.11}$$

$$\frac{dm_r(t)}{dt} = b_r P_{0,0}^{r}(t) + \alpha b_r P_{1,0}^{r}(t) - d_r m_r(t) \tag{2.12}$$

$$\frac{da_{uf}(t)}{dt} = t_a m_a(t) - k_{fa} a_{uf}(t) - \lambda f(X) a_{uf}(t) \tag{2.13}$$

$$\frac{dr_{uf}(t)}{dt} = t_r m_r(t) - k_{fr} r_{uf}(t) - f(X) r_{uf}(t) \tag{2.14}$$

$$\frac{da(t)}{dt} = k_{fu} a_{uf}(t) - k_{da}(a(t))^2 - \lambda f(X) a(t) + k_{-da} a(t) \tag{2.15}$$

$$\frac{dr(t)}{dt} = k_{fr} r_{uf}(t) - k_{dr}(r(t))^2 - \lambda f(X) r(t) + k_{-dr} r(t) \tag{2.16}$$

$$\frac{dr_2(t)}{dt} = k_{dr}\big(r(t)\big)^2 - k_{-dr} r_2(t) - k_t\big(r_2(t)\big)^2 + k_{-t} r_4(t) - \lambda f(X) r_2(t) \tag{2.17}$$

So far, none of the equations accounts at the same time for promoters in both the activator and the repressor plasmids. The next two equations do include such relationship, and thus, can are critical equations to model.

$$\begin{aligned}\frac{da_2(t)}{dt} = {} & -k_a P_{0,0}^{a}(t) a_2(t) + k_{-a} P_{1,0}^{a}(t) - k_a P_{0,1}^{a}(t) a_2(t) + k_{-a} P_{1,1}^{a}(t) \\ & - k_a P_{0,2}^{a}(t) a_2(t) + k_{-a} P_{1,2}^{a}(t) + k_l P_{1,2}^{a}(t) - k_a P_{0,0}^{r}(t) a_2(t) \\ & + k_{-a} P_{1,0}^{r}(t) - k_a P_{0,1}^{r}(t) a_2(t) + k_{-a} P_{1,1}^{r}(t) - k_a P_{0,2}^{r}(t) a_2(t) \\ & + k_{-a} P_{1,2}^{r}(t) + k_l P_{1,2}^{r}(t) + k_{da}(a(t))^2 - k_{-da} a_2(t) \\ & - \lambda f(X) a_2(t)\end{aligned} \tag{2.18}$$

$$\begin{aligned}\frac{dr_4(t)}{dt} = {} & -2k_r P_{0,0}^{a}(t) r_4(t) + k_r P_{0,1}^{a}(t) - 2k_r P_{1,0}^{a}(t) r_4(t) + k_{-r} P_{1,1}^{a}(t) \\ & - k_r P_{0,1}^{a}(t) r_4(t) + 2k_{-r} P_{0,2}^{a}(t) - k_r P_{1,1}^{a}(t) r_4(t) + 2k_{-r} P_{1,2}^{a}(t) \\ & - 2k_r P_{0,0}^{r}(t) r_4(t) + k_{-r} P_{0,1}^{r}(t) - 2k_r P_{1,0}^{r}(t) r_4(t) \\ & + k_{-r} P_{1,1}^{a}(t) - k_r P_{0,1}^{r}(t) r_4(t) + 2k_{-r} P_{0,2}^{r}(t) - k_r P_{1,1}^{r}(t) r_4(t) \\ & + 2k_{-r} P_{1,2}^{r}(t) + k_t\big(r_2(t)\big)^2 - k_{-t} r_4(t) - \lambda f(X) r_4(t)\end{aligned} \tag{2.19}$$

This set of twenty seven equations (including the nine equations regarding the repressor plasmid in the Appendix A) will form the system of ordinary differential equations that will have to be solved in order to simulate the behavior of the system.

2.4 Step 4: Solving and simulating the system

This system of deterministic ordinary differential equations cannot be solved analytically. Because they are nonlinear and they include several time dependent variables, explaining why numerical computational algorithms will have to be used. Before solving the system there is a still a previous task to performed, we need to establish the values of the rate constants governing the chemical reactions. The parameters used for the simulations were the following ones:

TABLE 2.1: PARAMETER'S VALUE USED FOR SIMULATION [23]

Basal transcription rate activator/repressor[2]:	$b_a = b_r = 0.36\ min^{-1}$
Transcription rate with one activator bound to promoter:	$\alpha = 20$
Rate of unbinding of activator/repressor from promoter:	$k_{-a} = k_{-r} = 1.8\ min^{-1}$
Translation rates of activator/repressor transcripts:	$t_a = t_b = 90\ min^{-1}$
Degradation rate activator/repressor transcripts:	$d_a = d_r = 0.54\ min^{-1}$
Folding rates:	$k_{fa} = k_{fr} = 0.9\ \ min^{-1}$
Dimerization rates:	$k_{da} = k_{dr} = 0.018\ \ min^{-1}\ molecules^{-1}$
Tetramerization rate of r_2:	$k_t = 0.018\ \ min^{-1}\ molecules^{-1}$
Dimer breaking rate:	$k_{-da} = k_{-dr} = 0.00018\ \ min^{-1}$
Tetramer breaking rate:	$k_{-t} = 0.018\ \ min^{-1}$
Looping rate:	$k_l = 0.36\ \ min^{-1}$
Un-looping rate:	$k_{ul} = 0.18\ \ min^{-1}$
Activator/repressor active degradation rate:	$\gamma = 1080\ \ molecules/min$
ssrA tags that half the degradation rate:	$C_e = 0.1\ molecules$
Maximum degradation rate of activator proteins forms:	$\lambda = 2.5$
Active degradation factor in looped/un-looped promoters:	$\epsilon = 0.2$

All these parameters are assumed to be the same for all the simulations, except for b_r, whose value will be modified later on. Furthermore, there are three missing parameters in table 2.1: two of them correspond to the forward binding rates of the activator dimers and repressor tetramers to the operator site [23]. These two parameters are kept apart because they do depend on environmental conditions and as consequence they will change depending on the chosen simulation conditions. They are expressed as:

[2] The rate of transcription in the repressor will be later modified (see 3.Results)

$$k_r = k_{-r}\left(\left(C_r^{max} - C_r^{min}\right)\cdot\frac{1}{1+\left(\frac{[IPTG]}{k_{r1}}\right)^{b_1}} + C_r^{min}\right) \tag{2.20}$$

$$k_a = k_{-a}\left(\left(C_a^{max} - C_a^{min}\right)\cdot\frac{[ara]^{c_1}}{\left(k_{a_1}\right)^{c_1}+[ara]^{c_1}}\cdot\frac{1}{1+\left(\frac{[IPTG]}{k_{r2}}\right)^{b_2}} + C_a^{min}\right) \tag{2.21}$$

Where k_{-r} and k_{-a} are the unbinding rates of the repressor and the activator, respectively. Notice also that these two parameters depend at the same time on more parameters such as the maximum and minimum affinities of transcription factors to their binding sites - $C_{a/r}^{max}$ and $C_{a/r}^{min}$- (table 2.2). But more importantly, for the first time in the modeling steps, the two inductors appear. They are the arabinose and IPTG concentrations - [*ara*] and [*IPTG*] -. In fact, they can be classified as the control parameters, since they represent environmental conditions that can be physically controlled.

TABLE 2.2: PARAMETER'S VALUE IN FORWARD BINDING RATES [23]

Maximum affinity of LacI tetramers to binding site:	$C_r^{max} = 0.2\ molecules^{-1}$
Minimum affinity of LacI tetramers to binding site:	$C_r^{min} = 0.01\ molecules^{-1}$
IPTG half strength for k_r:	$k_{r1} = 0.035\ mM$
IPTG cooperativity for k_r:	$b_1 = 2$
Maximum affinity of AraC dimers to binding site	$C_a^{max} = 1\ molecules^{-1}$
Minimum affinity of AraC dimers to binding site	$C_a^{min} = 0\ molecules^{-1}$
Activator half strength:	$k_{a1} = 2.5\%\ (\%\ w/v)$
IPTG half strength for k_a:	$k_{r2} = 1.8\ mM$
IPTG cooperativity for k_a:	$b_2 = 2$
Activator –arabinose- cooperativity:	$c_1 = 2$

All parameters were obtained from the original research [21]. For the sake of simplicity, we wanted to keep the values of the original paper. Considering the original article works with discrete variables and in this work, the continuum hypothesis is applied; by an abuse of notation we will use "number of molecules" as an equivalent to "concentration" in our simulation plots. As the goal of this thesis was to analyze the qualitative behavior of the system, this equivalency does not affect the results.

The general name of those coefficients is Hill coefficient. The Hill coefficient is used to when describing reactions in which there exists some kind of cooperation between the

events. Just mentioned in section *1.2.5 Gene Regulatory Networks*, Hill coefficient is used typically used when genes are regulated by transcription factors acting on multiple binding sites, like in this case, in which there are two operator sites, one for the activator - AraC dimer - and another two for the repressor - LacI tetramer -. The existence of this multiple binding makes transcription factors already bound to the operator site to influence the binding process of further transcription factors. Such influence is known as cooperativity and it can be modeled using the Hill function [12]:

$$Y = \frac{[X]^n}{K^n + [X]^n} = \frac{1}{\left(\frac{K_i}{[X]}\right)^n + 1} \tag{2.22}$$

Where *Y*, in this case represents the fraction of the transcription factor already bound to the ligand, *[X]* is the unbound ligand concentration, K_i is the ligand concentration needed to occupy half of the binding sites and *n* is the so called Hill coefficient.

The third parameter missing in table 2.1 refers to the total number of ssrA tags.

$$\begin{aligned} X = {} & \left(a_{uf}(t) + r_{uf}(t) + a(t) + r(t)\right) \cdot 1 + \left(a_2(t) + r_2(t)\right) \cdot 2 + r_4(t) \cdot 4 \\ & + \left(P^a_{1,0}(t) + P^a_{1,1}(t) + P^a_{1,2}(t) + P^r_{1,0}(t) + P^r_{1,0}(t) + P^r_{1,0}(t)\right) \cdot 2 \\ & + \left(P^a_{0,1}(t) + P^a_{1,1}(t) + P^r_{0,1}(t) + P^r_{1,1}(t)\right) \cdot 4 \\ & + \left(P^a_{0,2}(t) + P^a_{1,1}(t) + P^r_{0,2}(t) + P^r_{1,2}(t)\right) \cdot 4 \\ & + \left(P^a_{L,1}(t) + P^r_{L,1}(t)\right) \cdot 2 + \left(P^a_{L,2}(t) + P^r_{L,2}(t)\right) \cdot 4 \end{aligned} \tag{2.23}$$

This parameter, affecting nearly all the degradation reactions, was not placed in that table because it is expected to change. The number of ssrA tags depends on each molecule, in such a way that there is one for each monomeric version, two for dimers and four for tetramers, including proteins bound to the operator sites.

Once all the parameters have been defined, the next step was to solve the system of equations. As it has been mentioned several times, a system of twenty seven nonlinear ordinary differential equations cannot be solved analytically; instead, numerical methods are the most appropriate option. Within the computational field, there exist different options for solving differential equations. XPPAUT, Matlab, Mathematica or Mapple are only some examples of available computational software packages [12]. In this thesis, Matlab was the selected option as it is one of the programs learnt and used in this degree.

In an *initial value problem*, like this one, the system of ODEs is solved starting from an initial state. From that initial condition and establishing a time interval over which the simulation is wanted to be obtained, the solution is obtained through iterations. In this

specific case, the initial conditions correspond to the initial species concentrations. All of them were set to zero, except the number of free hybrid promoters.

$$P_{0,0}^{a}(t=0) = 50 \tag{2.24}$$

$$P_{0,0}^{r}(t=0) = 25 \tag{2.25}$$

As there was just one hybrid promoter cloned in each plasmid, the number of activator/repressor plasmids used determined these initial conditions.
Once the initial conditions were established, the next step was the selection of a numerical ODEs solver. Matlab offers a wide variety of ODEs solvers, depending on the problem type they are aimed to solve [25]. An ODE problem would be stiff if the searched solution varies slowly, but at the same time, there were nearby solutions that do it rapidly. As a consequence, some numerical methods would have to use a extremely small step size - explicit methods - while others could perform quite well – implicit -. [25] [26]. As one can infer from this description, the concept of stiffness is related to computational efficiency. When modeling big systems of equations, one key variable determining the efficiency of the numerical method - solver - used is the computational time it needs to reach the solution. One of the factors affecting this time is the step size used. Ideally, in regions were the solution curve is nearly flat one would like the step size to be increased; the opposite behavior would be desired when solution curve considerably varies. Stiff problems force some numerical methods to use such small step size that they become exceedingly difficult to solve.

Determining *a priori* if a given problem is stiff or not is not trivial, therefore, the common approach is to initially consider the problem as non-stiff, and then, depending on whether the solver is able or not to solve or the time it takes, a decision can be made. Initially, *ode45* was used. According to Matlab's documentation, this non-stiff ODEs solver should be the first one to try. However, experience showed that *ode23* was more efficient in this case, especially in terms of computational time. This could occur because the problem was closer to be mildly stiff than non-stiff [27]. This solver consists of an implementation *Bogacki-Shampine* method [27]. This numerical method is a type *Runge-Kutta* method of order three. Moreover, it has added an extra second-order method that makes possible to have an adaptive step size.

Suppose the equation to be solved is:

$$y' = f(t, y) \text{ with initial condition } y_o \tag{2.26}$$

Then, *Bogacki-Shampine* method uses:

$$k_1 = f(t_n, y_n) \tag{2.27}$$

$$k_2 = f(t_n + \frac{1}{2}h_n, y_n + \frac{1}{2}h_n k_1) \tag{2.28}$$

$$k_3 = f(t_n + \frac{3}{4}h_n, y_n + \frac{3}{4}h_n k_2) \tag{2.29}$$

$$y_{n+1} = y_n + \frac{2}{9}h_n k_1 + \frac{1}{3}h_n k_2 + \frac{4}{9}h_n k_3 \tag{2.30}$$

Where y_n represents the numerical solution at time t_n, and h_n is the step size defined as:

$$h_n = t_{n+1} - t_n \tag{2.31}$$

Equations 26-29 correspond to the third-order approximation [28]. However, this type of *Runge-Kutta* method has another method of order two:

$$k_4 = f(t_n + h_n, y_{n+1}) \tag{2.32}$$

$$z_{n+1} = y_n + \frac{7}{24}h_n k_1 + \frac{1}{4}h_n k_2 + \frac{1}{3}h_n k_3 + \frac{1}{8}h_n k_4 \tag{2.33}$$

Where z_{n+1} is the mentioned second order approximation. The use of this second method is what actually allows adapting the step size, since the difference between z_{n+1} and y_{n+1} is used for that adaptation purpose. Additionally, another property of *Bogacki-Shampine* is the so-called FSAL - first same as last - meaning that k_4 at a certain step is equal to k_1 in the following step, therefore, it just needs three function evaluations per step [29].

Assuming all variables have been defined, the line of code calling that solver would be the following one:

```
[t,yd] = ode23 (odefun, tspan, yd0, options);
```

The solver receives four input arguments. `Odefun` input argument refers to the function to solve. In this case, it would correspond to a matrix containing on each row one of the differential equations describing the system. The second input parameters is `tspan`, which is the time interval of integration. This time span has to be defined as a vector with two entries for initial and final times. Ideally, the final time would be a high number, so that long term behavior of the system can be observed. However, the

greatest the time interval, the more computational time the program will take. Therefore, this tradeoff between computational time and desired output makes necessary to adjust these input parameters depending on the specific simulation to be performed. The third parameter `yd0` corresponds to the initial conditions used to compute the first numerical solution. As mentioned above, all promoters, mRNA transcripts and protein forms were set to zero except the concentrations of free promoters, which received the same values as the number of activator/repressor plasmids used by the researchers. Finally, the fourth input parameter is an optional one allows the user to specify parameters like relative error tolerance [25], turn on solver statistics… Although available during this simulation that parameter was left as default.

The output arguments of solver are `t` and `yd`. `t` is an array containing all the intermediate evaluation points between the initial and the final times introduced as input arguments and `yd` is a matrix with as many columns as number variables the system has and with each row corresponding to the solution value at a time located in that same row of the time vector - `t` -.

The measurement and amplitude of oscillations was done using a Matlab function called *findpeacks.* This function receives an input vector with data values and then returns a vector with the maxima of the input vector and another vector with the location of those maxima. The period was extracted by averaging the distance between the position of consecutive maxima. To compute the minima, the initial data was inverted (multiplied by -1) and then, *findpeaks* again was used to compute the maxima that in this case correspond to the original minima. Finally, the amplitude was calculated by subtracting to the maxima the subsequent minima. *Findpeaks* offers the possibility to adjust several parameters. In this case, one of those options, *MinPeakProminence,* was used to discard oscillations caused by the integrator itself and not by the system.

```
[peaks, locs]=findpeaks(signal,'MinPeakProminence');
```

Up to this point, what it has been done is to go from the chemical reactions describing a gene regulatory network to obtain the differential equations required to analyze that system qualitatively and quantitatively. The next step was to obtain the temporal variation of each molecular species involved in the reaction network as a function of time. Those results will be presented in the following section (*3. Results*) and then will be analyzed and discussed (See 4. *Discussion and conclusion).*

3. Results:

In this section, the different results obtained from the simulation of the system will be presented. These results will be developed from a more general to a more specific analysis.

3.1 General analysis

The key point about systems modeling is the possibility of testing and simulating the behavior of the system under different conditions. In this case, the different conditions correspond to the different values given to the control parameters: the forward binding rates of the transcription factors to the operator sites - k_a and k_r - which in turn depend on both arabinose and IPTG concentrations (see equations 2.20 and 2.21). Thus, we can study the system under different concentrations of each of the inducers and see how variations in their concentrations affect the behavior of the system.

Initially, the simulations were performed using some of the values of inducers of the original model. First of all, the system was simulated keeping constant one of the inducers and varying the other one.

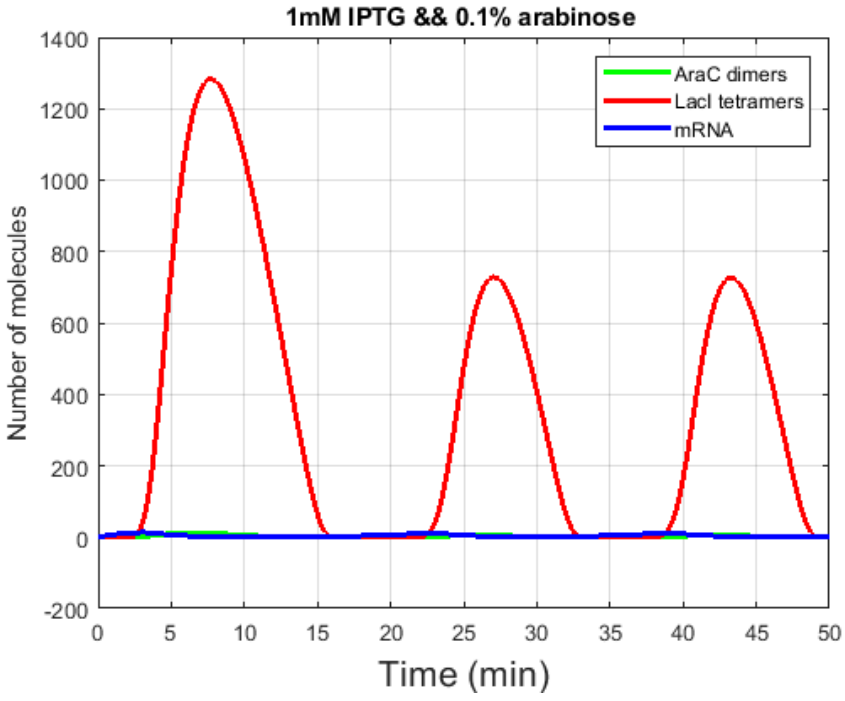

Fig. 3.1 Simulation example. Time dynamics of AraC dimers, LacI tetramers and mRNA of LacI and AraC as a function of time for 0.1% arabinose concentrations and 1mM IPTG.

In order to perform these simulations, the arabinose values were 0.1%, 0.2%, 0.4%, 0.7% and 1% while the IPTG concentrations varied between 0mM, 1mM, 2mM, 4mM, 7mM and 10mM. As it can be seen in Fig. 3.1, it is a difficult task to compare the absolute values of the variables: i.e., the range of the variable AraC dimers is very small compared to the range of LacI tetramers. Changes in AraC dimers values are small due to the low number of molecules involved. In order to solve that, we decided to normalize the variables by dividing each species population by the maximum value of that species during the simulation.

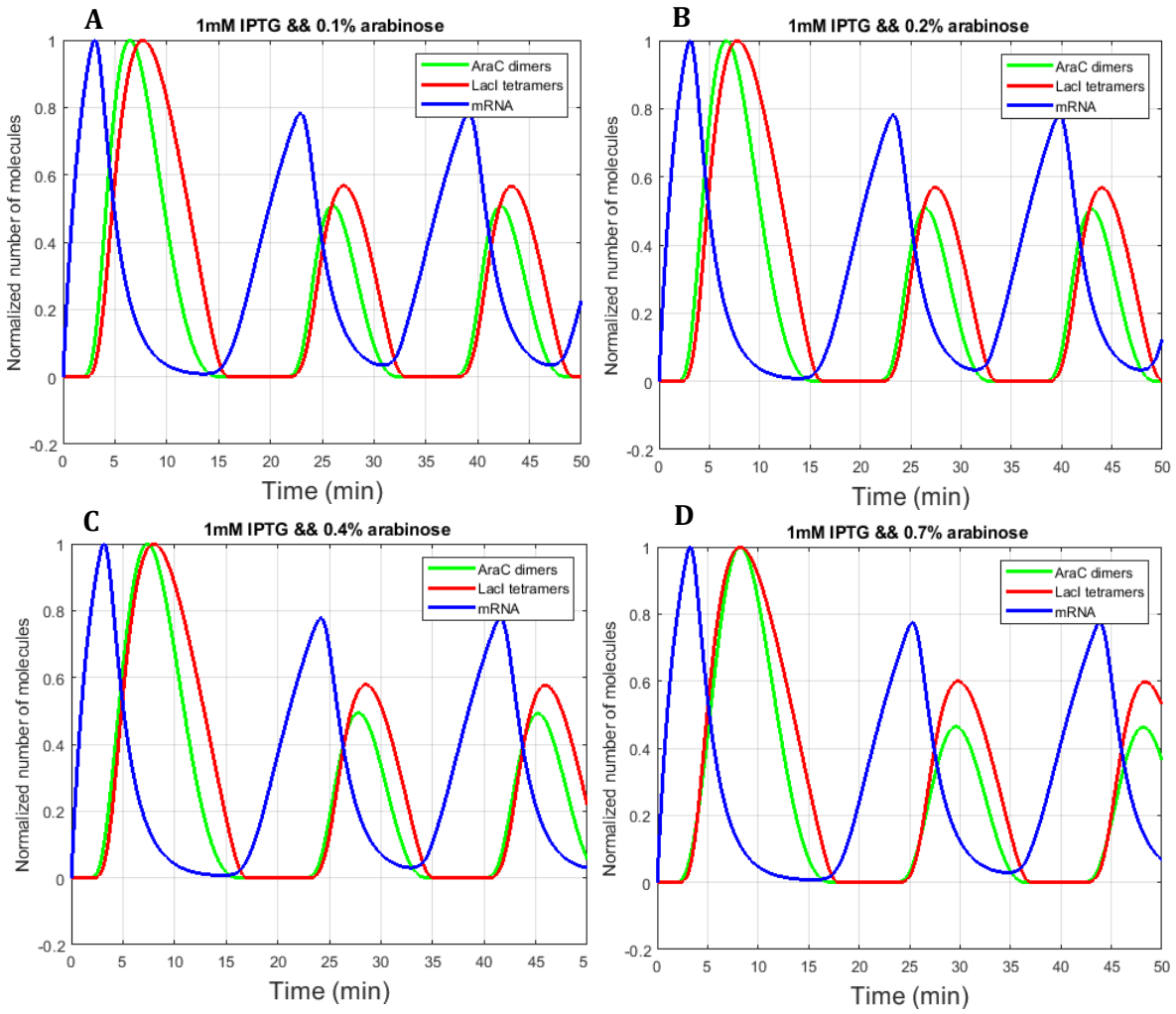

Fig. 3.2 Time dynamics of normalized AraC dimers, LacI tetramers and mRNA of both genes as a function of time at constant fixed 1mM IPTG and at 0.1% (A), 0.2% (B), 0.4% (C) and 0.7% (D) arabinose concentration.

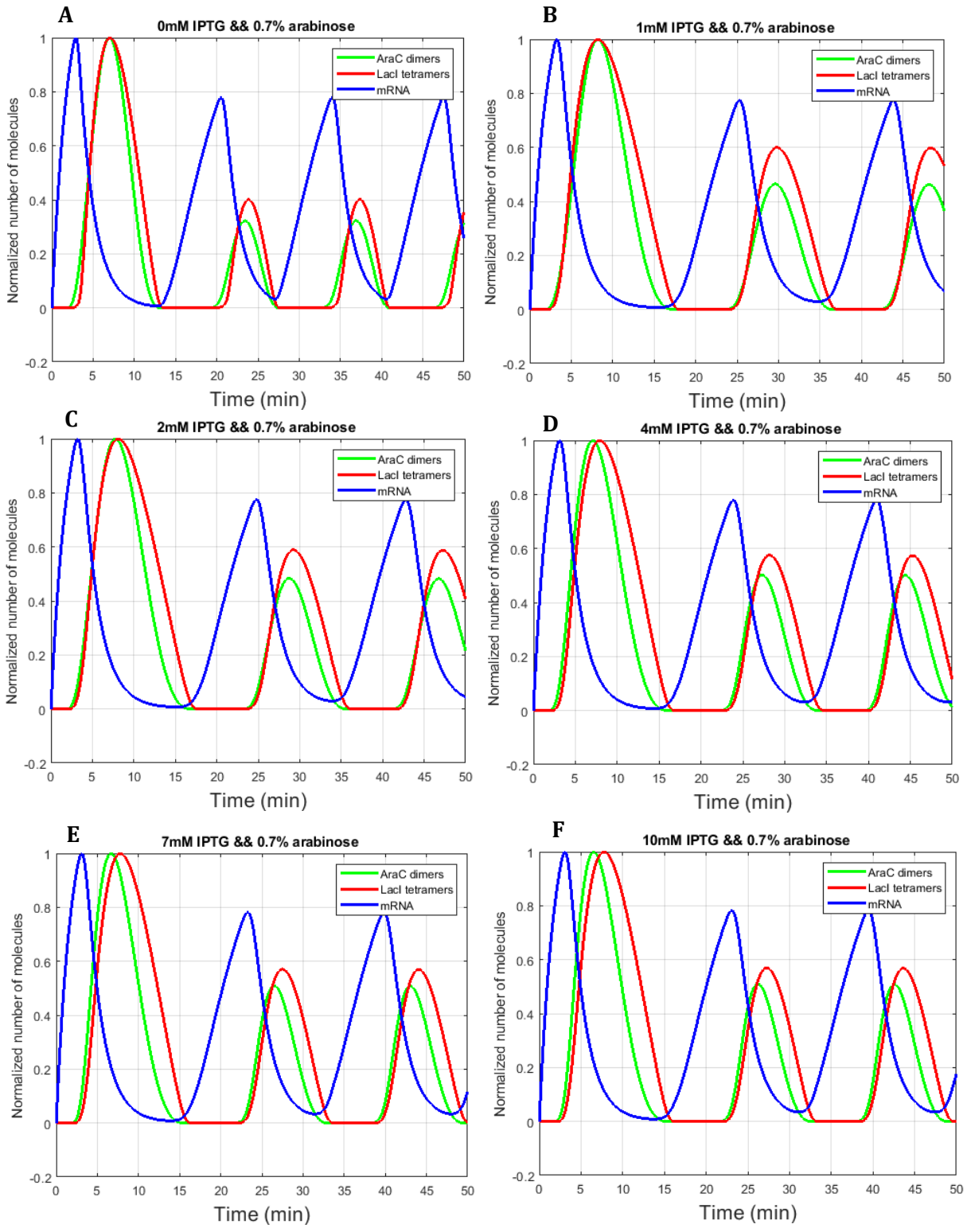

Fig. 3.3 Time dynamics of normalized AraC dimers, LacI tetramers and mRNA as a function of time at constant fixed 0.7% arabinose and at 0 (A), 1 (B), 2 (C) 4 (D), 7(E) and 10(F) mM of IPTG.

This normalization was applied to the 30 simulations (see in the Appendix B). The results obtained were qualitatively similar. Although the specific variable values of the system changed for different simulations, in all cases the same key points can be appreciated. Interpreting or even describing this type of graph is a challenging task that requires knowledge about the specific gene regulatory network modeled.

I. There is a burst beginning with the basal transcription of mRNA of both promoters.
II. After a short delay, the amount of functional transcription factors increases. Depending on the specific concentration of arabinose and IPTG, such increase will be faster for AraC dimers or LacI tetramers (see Fig. 3.4). As it was mentioned when describing the system, AraC dimer binding to operator site will reinforce transcription in both genes, while LacI tetramers will have the opposite effect.
III. As mRNA is translated into protein, transcription is already repressed by the repressor factor (LacI tetramers) resulting into a decrease of mRNA transcripts (blue curve).
IV. AraC dimers and LacI tetramers population increases until the fall in mRNA levels produces a decrease in translation, and degradation dominates protein dynamics
V. Eventually, almost at the same time, AraC dimers and LacI tetramers extinguish and mRNA transcripts start increasing again, repeating the cycle.

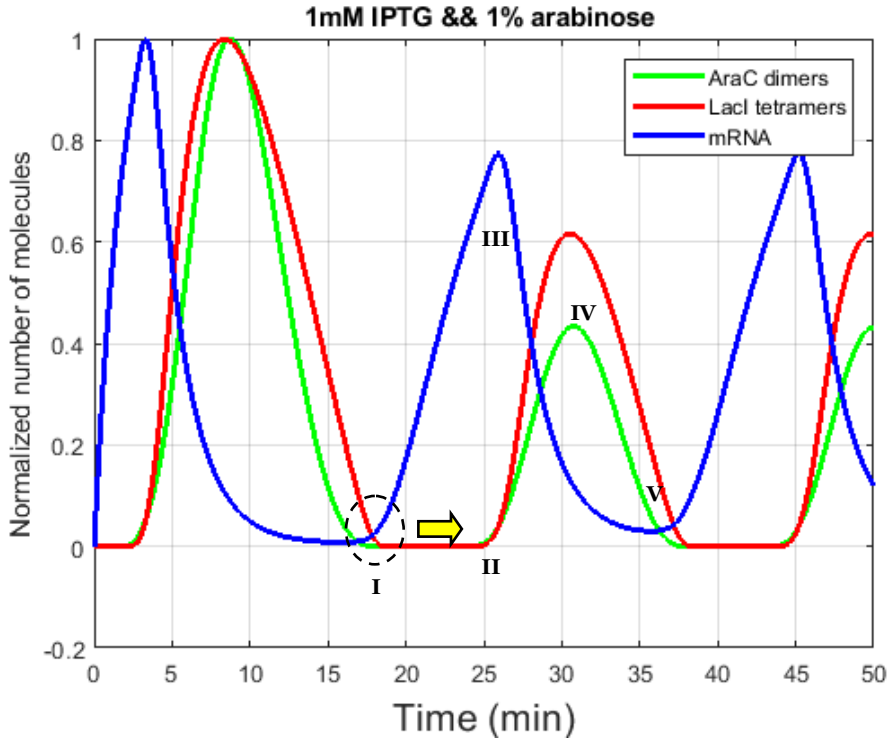

Fig. 3.4 Time dynamics of AraC dimers, LacI tetramers and mRNA at 1 mM of IPTG and 1% arabinose.

As it can be seen, these general simulations were performed over a time interval of 50 minutes.

Having normalized the plots, it was still difficult to understand from a quantitative point of view how the system behaved in response to changes in inducer concentrations. Although changes in amplitude and period of oscillations could be inferred if all graphs were observed at the same time, we plot the evolution of each molecule population against time individually for different inductor values within the same graph. Therefore, according to that criterion, there would be two types of graphs: one for AraC dimers and another one for LacI tetramers. Moreover, for each molecule, there would be two possible situations, one in which IPTG is kept constant and arabinose is changed and vice versa. The graphs associated to AraC dimers are in Fig. 3.5 and Fig. 3.6:

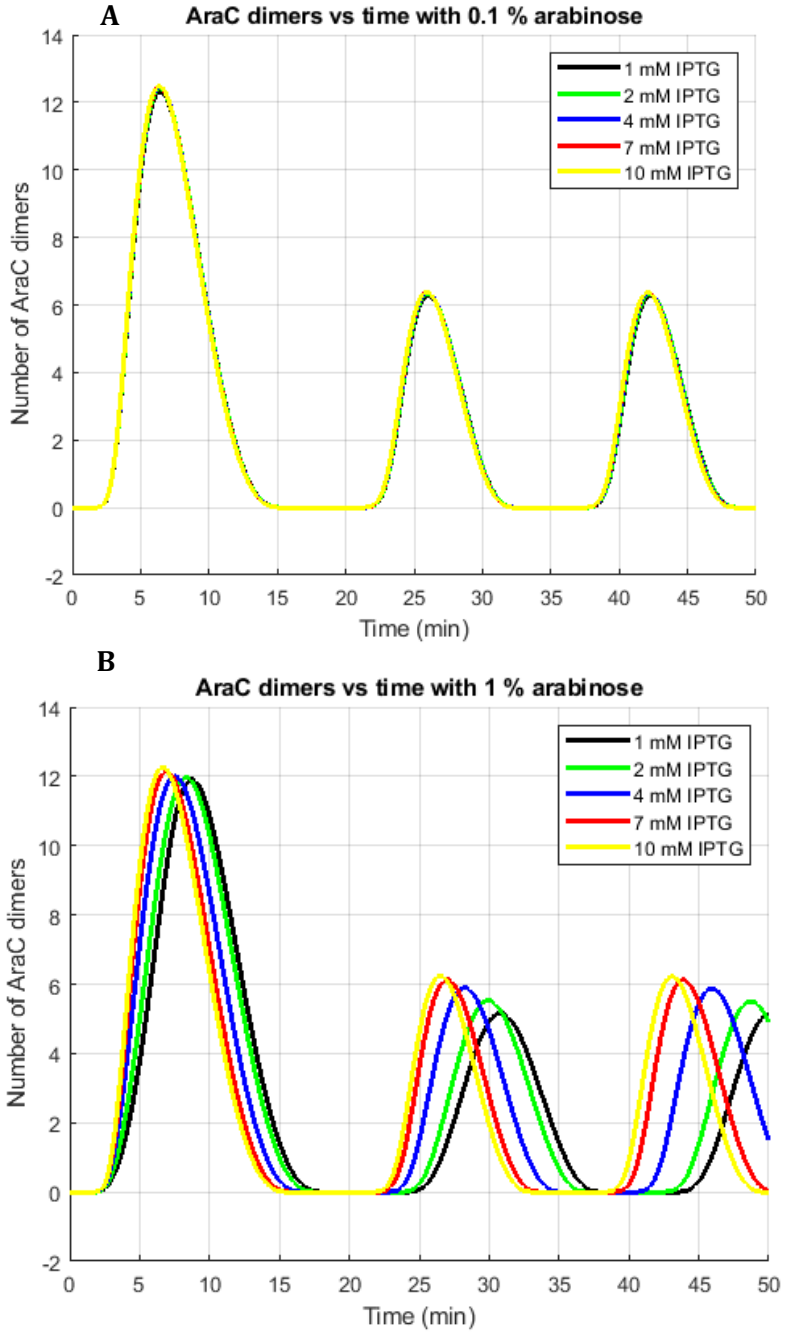

Fig. 3.5 Time dynamics of AraC dimers varying IPTG at fixed arabinose concentration 0.1% (A) and 1% (B) versus time.

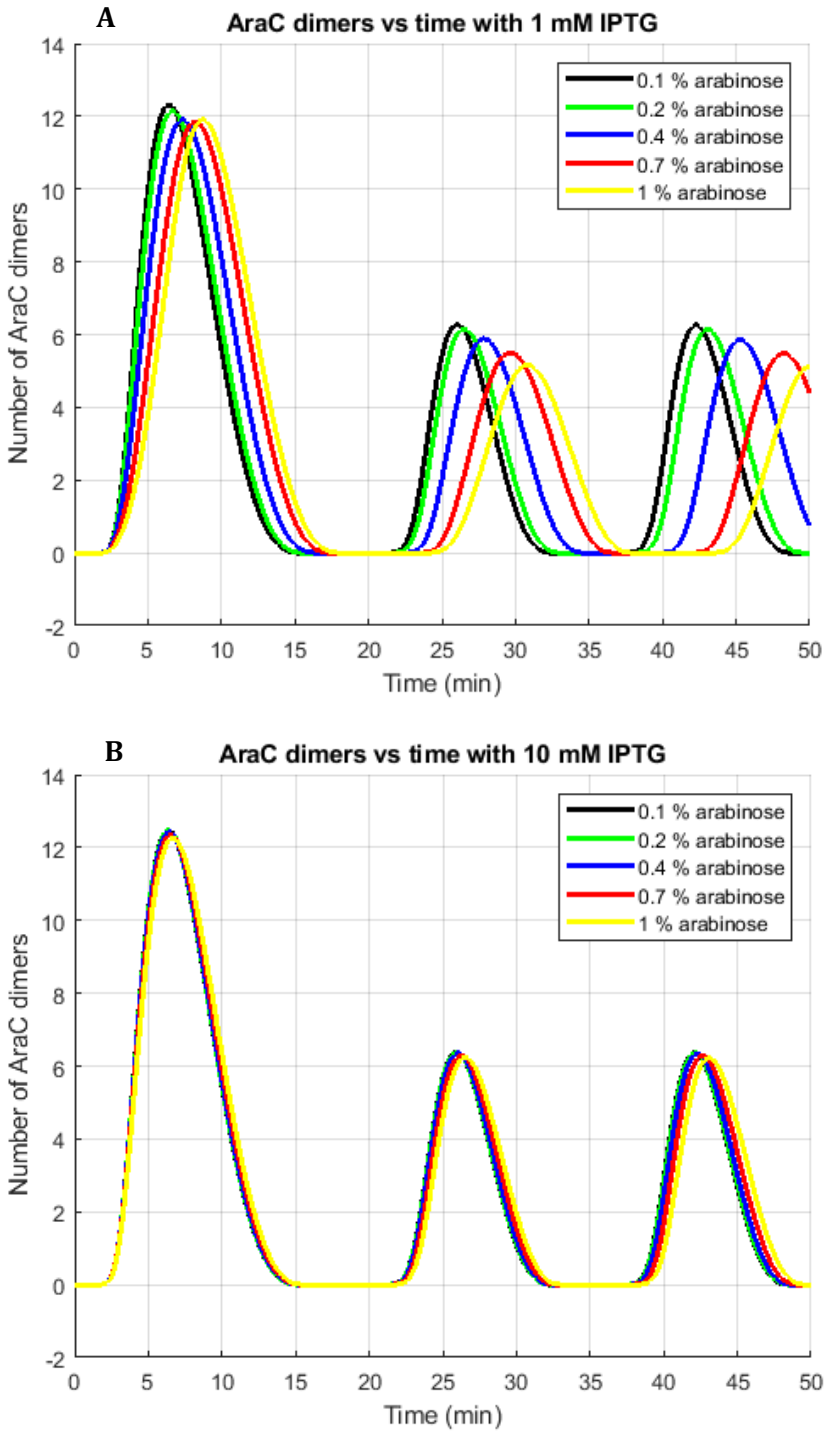

Fig. 3.6 Time dynamics of AraC dimers at fixed IPTG concentrations of 1 (A) and 10 (B) mM with varying arabinose concentration versus time.

Fig. 3.5A shows how at low arabinose concentrations the oscillations are almost identical for different IPTG concentration values. For high IPTG values, oscillations also look similar at different arabinose conditions (Fig. 3.6B). When arabinose concentration is increased to 1%, oscillations increase in amplitude and reduce in period when IPTG increases (Fig. 3.5B). In contrast, at low IPTG values (1mM), AraC dimers' oscillations decrease in amplitude and increase in period for increasing arabinose (Fig. 3.6A).

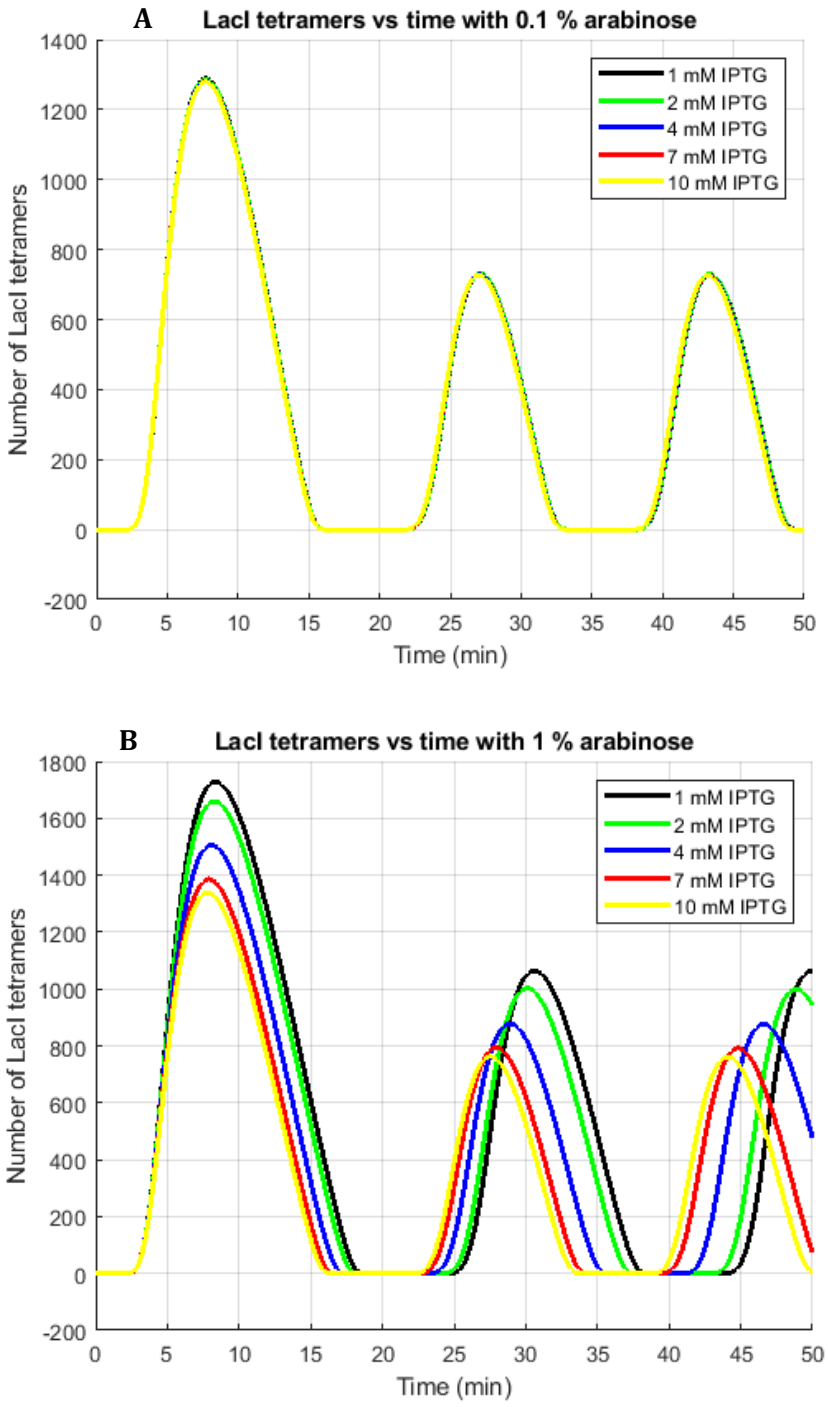

Fig. 3.7 Time dynamics of LacI tetramers at fixed arabinose concentrations of 0.1% (A) and 1% (B) with varying IPTG concentration versus time.

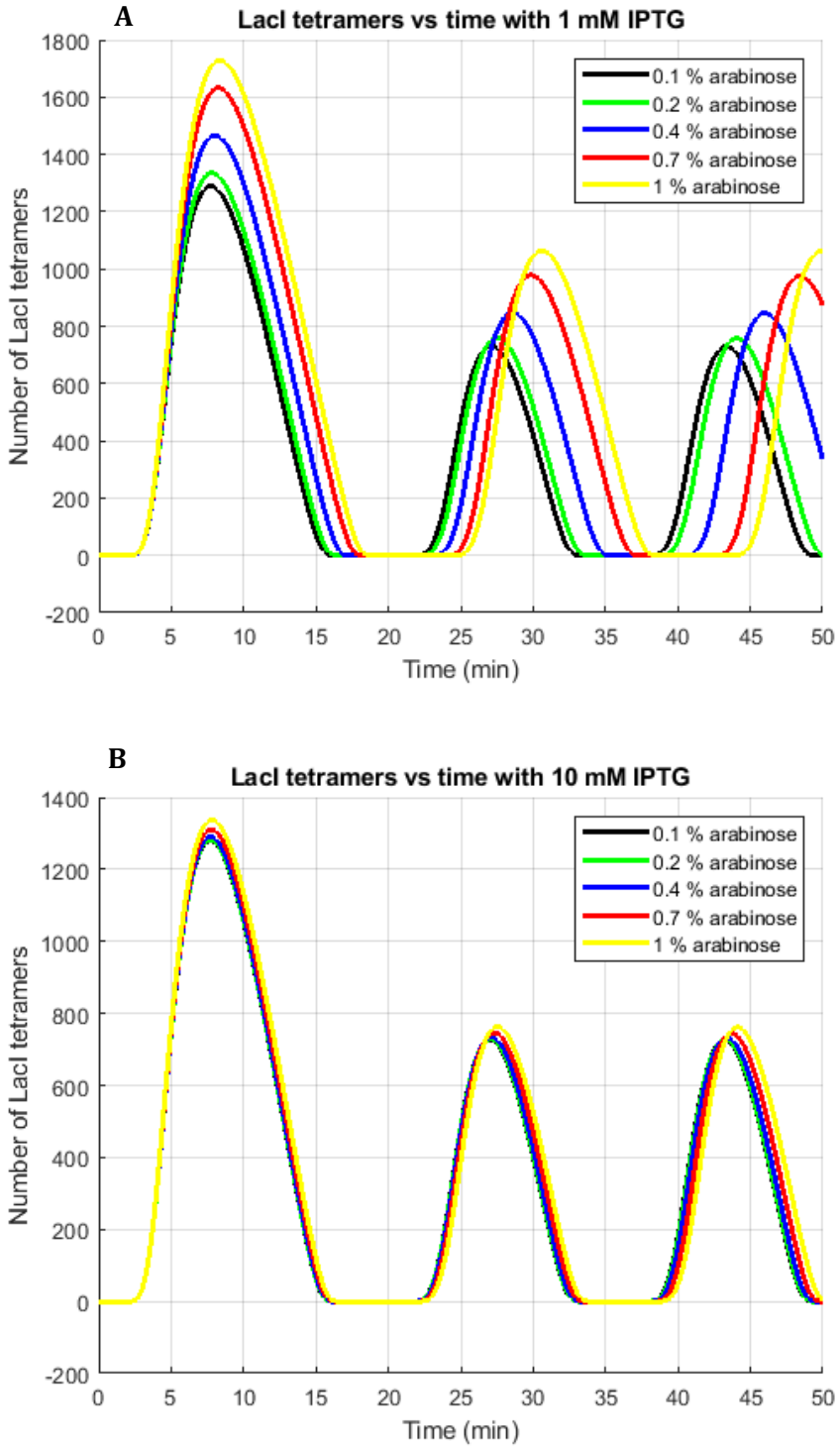

Fig. 3.8 Time dynamics of LacI tetramers at fixed IPTG concentrations of 1 (A) and 10 (B) mM with varying arabinose concentration versus time.

Similarly to AraC dimers, changes in amplitude and period of oscillations can also be appreciated in the dynamics of LacI tetramers. On one hand, the behavior of LacI tetramers is identical to that of the AraC dimers when arabinose concentration is low

(0.1% Fig. 3.7A) or IPTG concentration is high enough (10Mm Fig. 3.8B). However, the other two figures (Fig. 3.7B and Fig. 3.8A) display different behaviors compared to that of AraC dimers plotted under the same conditions. In this case, LacI tetramers oscillations appear to decrease both in amplitude and period when IPTG increases (Fig. 3.7B). At the same time, these oscillations seem to increase in amplitude and period when arabinose concentration increases at low levels of IPTG (Fig. 3.8A).

These results will be further analyzed in section - 4. *Discussion and conclusion* -. However, from the point of view of the system's behavior, all the simulations show that the system oscillated independently of the inducer concentrations. One of the initial goals of this thesis was to analyze qualitative changes in the system, something that it was not observed in the previous results. In an attempt to see more complex behaviors, we decided to modify the transcription rate of LacI. We hypothesized that that decreasing this transcription rate would diminish the influence of the negative feedback loop, which is responsible for the oscillation behavior.

3.2 Analysis with a third control parameter:

As previously stated, in this section we introduced a third control variable: the transcription rate of LacI - b_r -. This parameter, originally 0.36 min^{-1}, was varied in order to modulate the negative feedback loop. Then, the oscillatory behavior of the system was explored. These analysis were only performed at relatively low IPTG concentrations (1 mM and 2 mM) since higher concentrations had been shown to interfere with arabinose-AraC binding and consecutive activation of promoter [23].
The system was simulated for the following values:

TABLE 3.1 VALUES CONTROL PARAMETERS:

b_r [min^{-1}]:	0.015	0.025	0.05	0.1	0.36
Arabinose [%]:	0.1	0.2	0.4	0.7	1
IPTG [mM]:	1	2			

Obtaining the results shown in Fig. 3.9:

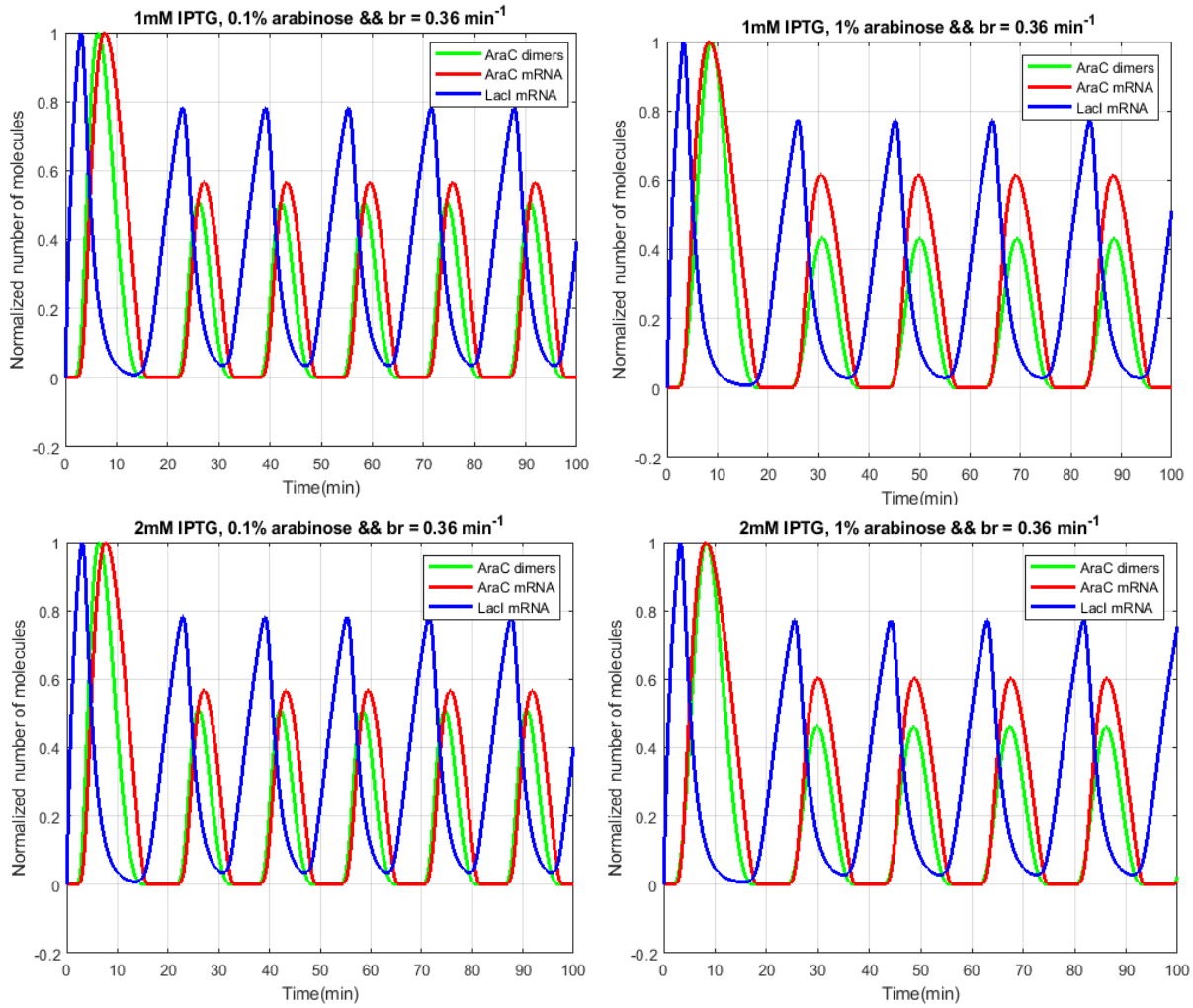

Fig. 3.9 Dynamics of AraC dimers, LacI tetramers and mRNA with original transcription rate at 1 mM and 2 mM of IPTG and 0.1% and 0.2% arabinose concentrations.

As it can be observed in Fig. 3.9, for the original value of b_r there are oscillations independently on the value of the other two control parameters. This situation changes when b_r was reduced.

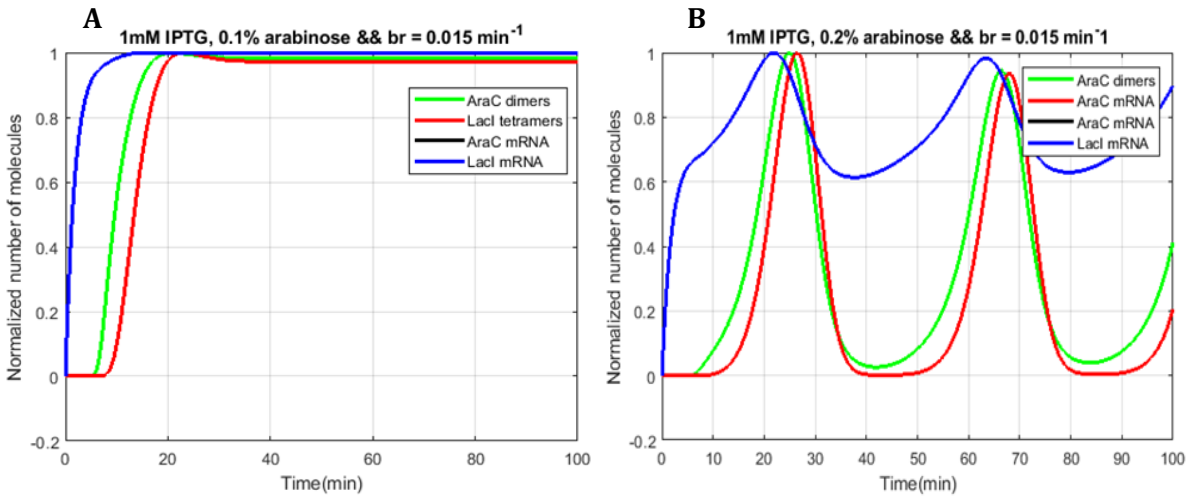

Fig. 3.10 Time dynamics AraC dimers, LacI tetramers and mRNA at b_r=0.015 min^{-1} at 1 mM and 0.1% (A) and 0.2% arabinose (B).

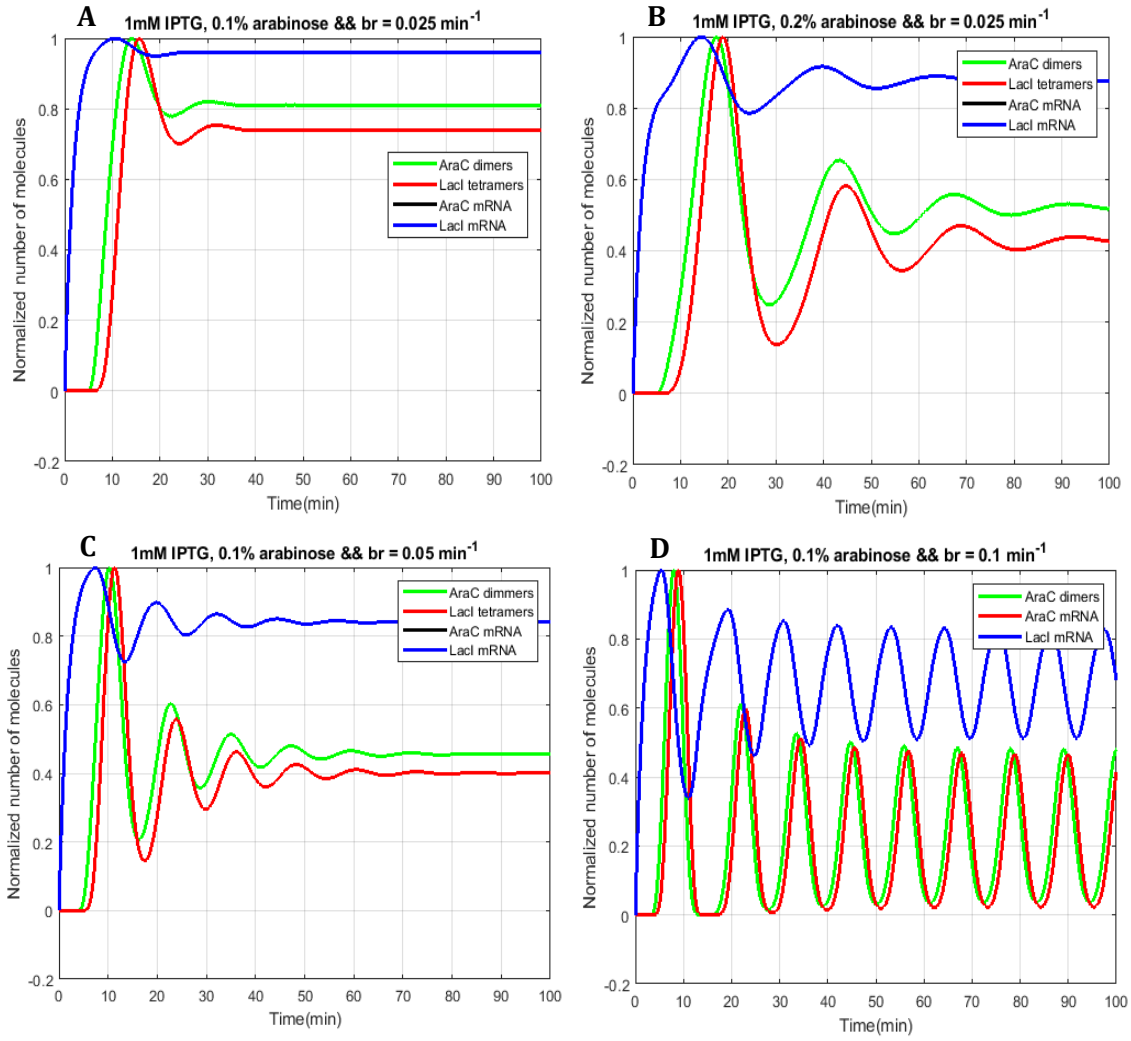

Fig. 3.11 Time dynamics of AraC dimers, LacI tetramers and mRNA at 1 mM IPTG and different arabinose and b_r values.

Now, when the transcription rate of LacI is reduced, changes in the behavior of the system for different inducer concentrations can be appreciated. For some conditions the variable quickly reaches the steady state (Fig. 3.10A or Fig. 3.11A). For others, there is an initial set of damped oscillations (Fig. 3.11B or 3.11C) or these oscillations persist over time (Fig. 3.11D). Although all these simulations were performed for 1mM of IPTG, variability in the results was also obtained when using 2mM of IPTG. This variability can be observed in Fig. 3.12, where AraC dimers are plotted at constant IPTG and arabinose concentration while varying LacI transcription rates. Depending on b_r values, the system displays oscillations or not (Fig. 3.12A).

Although these results will be analyzed in detail in the following section, it can be observed that the oscillatory behavior of the system depends on the values of b_r.

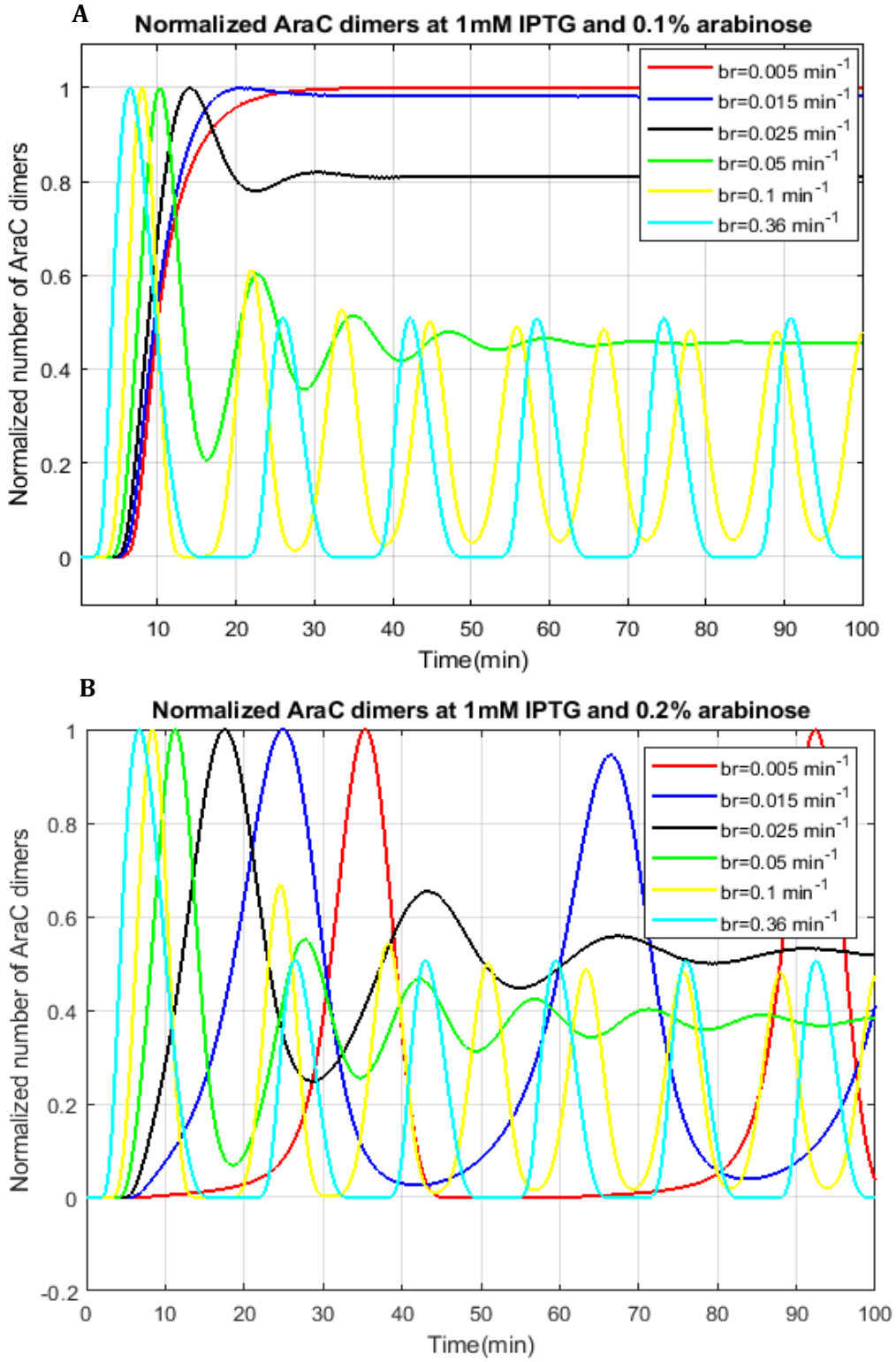

Fig.3.12 Time dynamics of AraC dimers at fixed 1 mM IPTG, 0.1% arabinose (A) and 0.2% arabinose (B) and varying b_r .

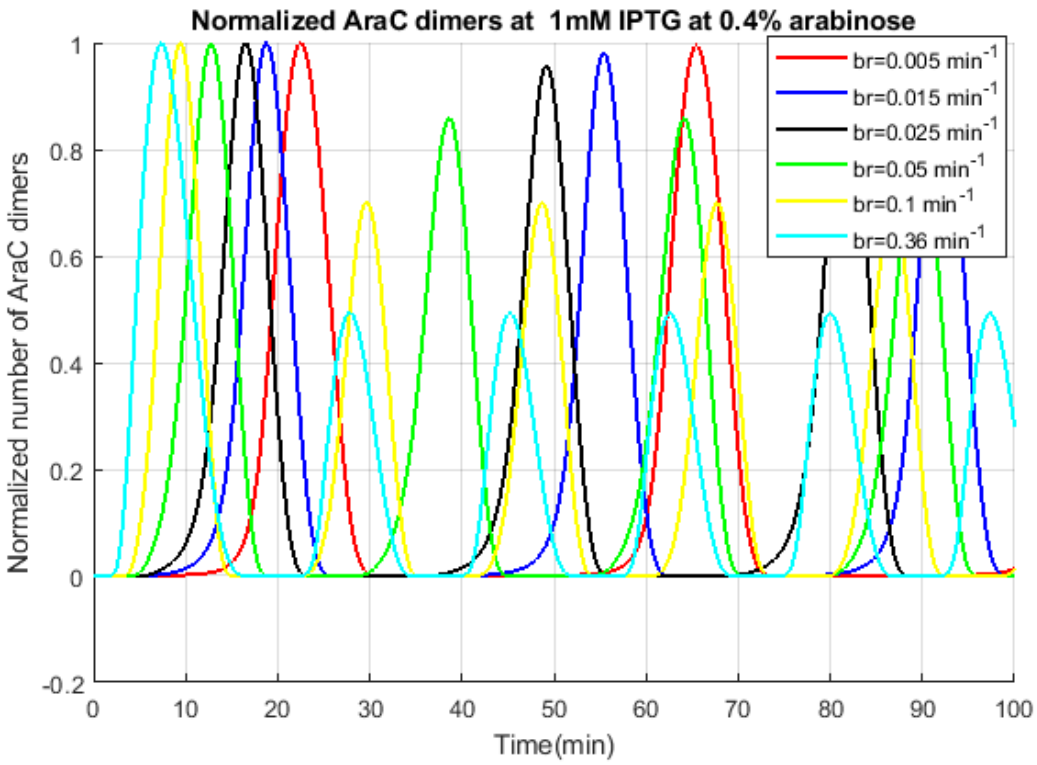

Fig. 3.13 Time dynamics of AraC dimers at 1 mM IPTG and 0.4% arabinose and varying and b_r .

3.3 Restricting the interval:

The simulations above indicate that there are values of the LacI transcription rate for which different behaviors can be observed. Following that, the next step was to study with higher accuracy the point at which such change in the behavior of the system occurs. This change in the behavior (and then in the stability) of the system corresponds to a bifurcation.
Among the different possible inducer inducers concentrations and b_r values over which the system had shown persistent oscillations and convergence to a steady state, the following conditions were chosen:

$$b_r = 0.015\ min^{-1} \ \&\& \ [IPTG] = 1\ mM \ \&\& \ [arabinose] \in (0.1\% - 0.2\%)$$

Simulations were performed at fixed IPTG and b_r and varying the arabinose concentration. As shown in Fig. 3.10, another valid approach could have been fixing the arabinose concentration and changing the repressor transcription rate. The reason why varying arabinose was preferred is that it is a parameter that it is easier to control experimentally.

Another important parameter to consider was time. In this case, the closer arabinose concentration was to the bifurcation point, the more time would be needed to simulate the system. Very long simulation times will be the only way to observe if the oscillations are maintained in time or in contrast are damped very slowly to reach a steady state. The main drawback derived from these long simulation times was the associated long computational times.

Having said that, initially the behavior of the system was simulated with the above parameters using increments of 0.005% in arabinose concentration.

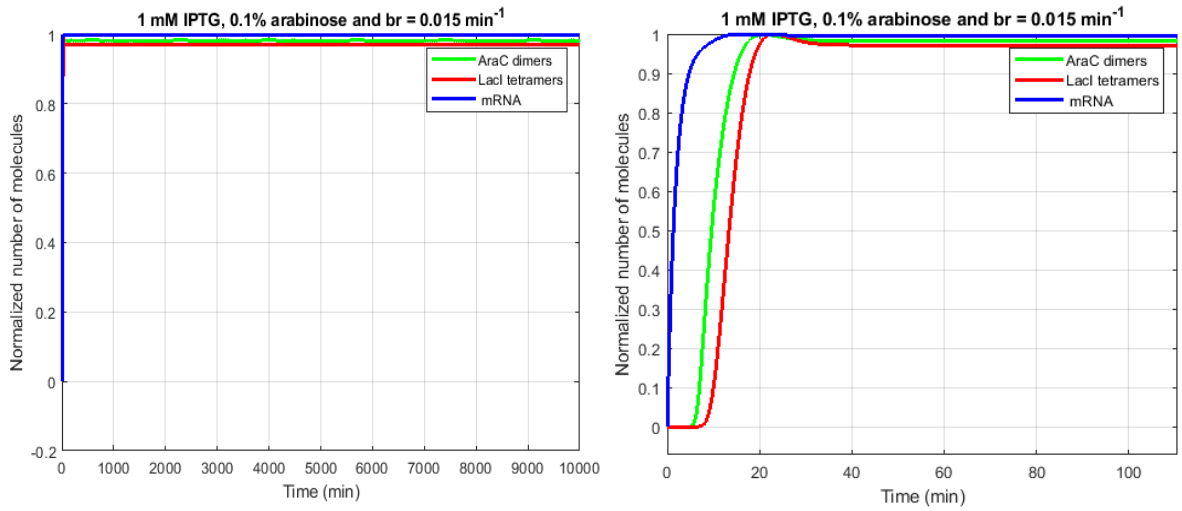

Figure 3.14 Time dynamics of AraC dimers, LacI tetramers and mRNA at 1 mM IPTG, b_r=0.015min^{-1} and varying arabinose versus time. Right panel represents amplified regions of the left panel.

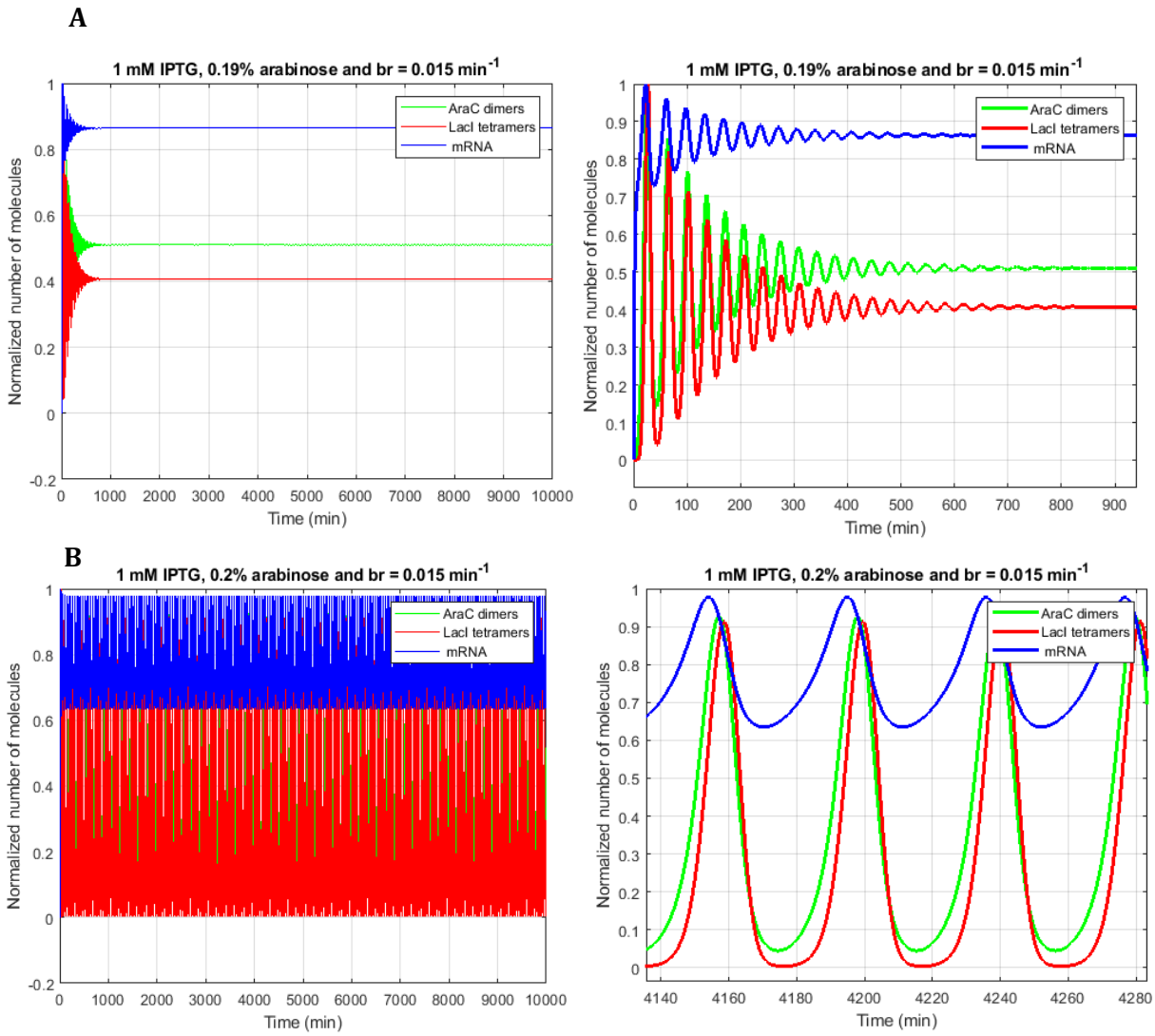

Figure 3.15. Time dynamics of AraC dimers, LacI tetramers and mRNA at 1 mM IPTG, b_r=0.015min^{-1} and 0.19% (A) and 0.20% (B) arabinose. Graphs at the right column are zoomed in regions of graphs in the left.

Thanks to these simulations, it was possible to restrict the interval of interest between 0.19 and 0.2 % arabinose concentration. Additionally, the period and the amplitude of the oscillations were also computed. For those simulations in which oscillations were damped, the period together with the amplitude were set to zero. Fig. 3.16 confirms that the bifurcation point is located inside the interval 0.19 to 0.20% of arabinose. Then, the next simulations were performed between 0.19 and 0.20% arabinose concentrations using increments of 0.0005% and increasing the simulation time up to 10^5 minutes.

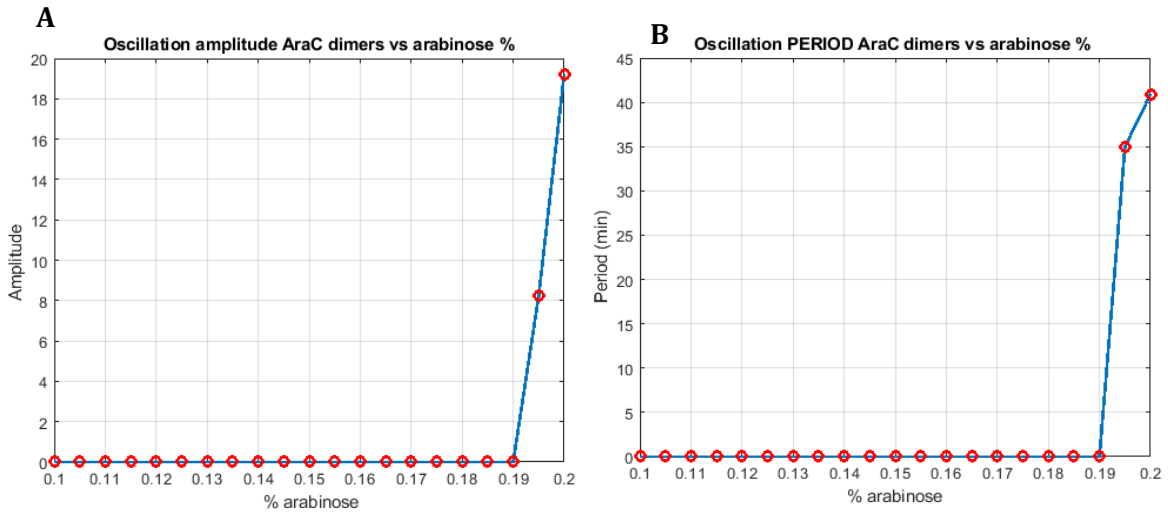

Fig. 3.16 Amplitude (A) and period (B) of oscillations of AraC dimers at 1 mM IPTG, b_r=0.015min^{-1} and varying arabinose.

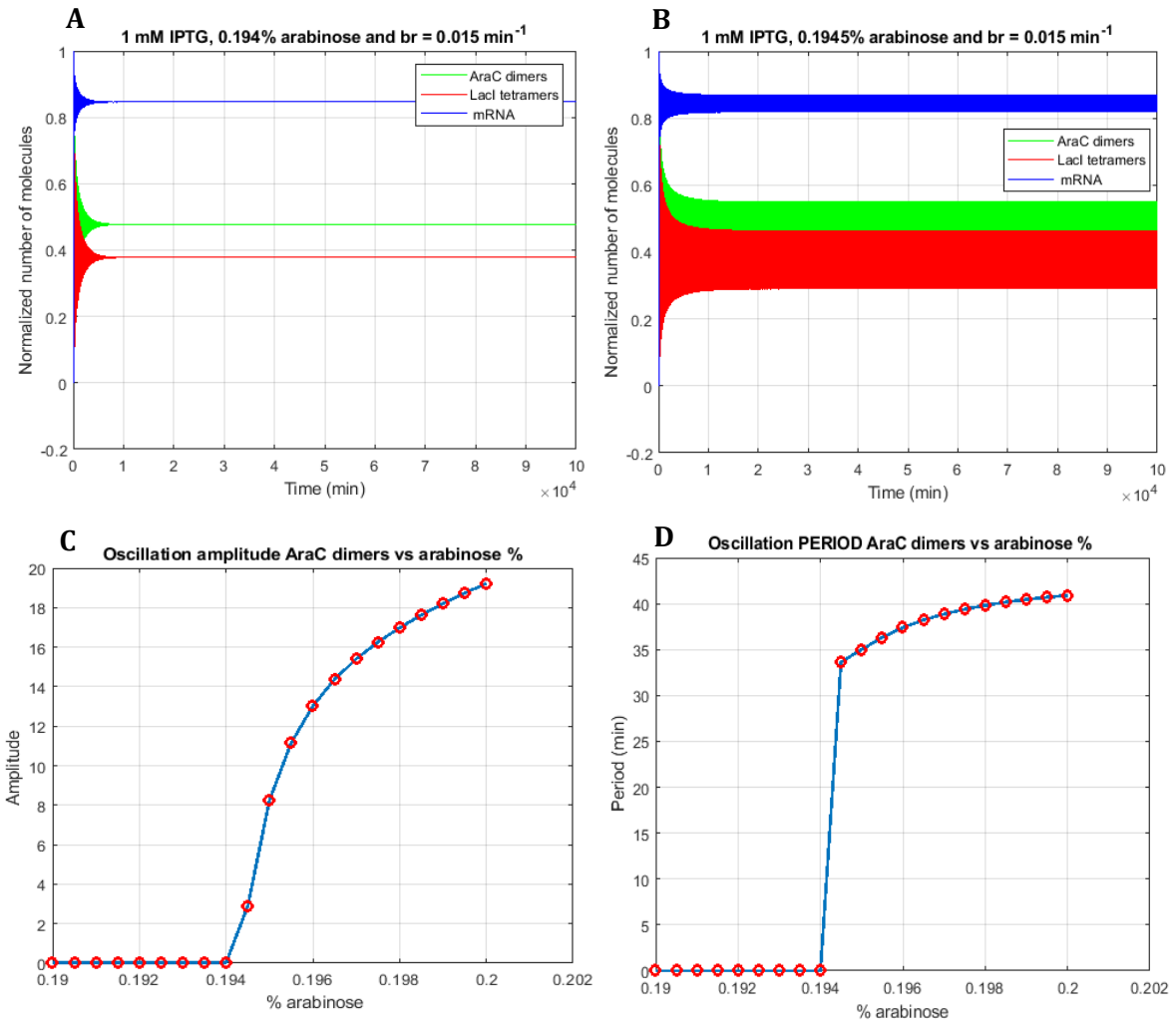

Fig. 3.17 Time dynamics of AraC dimers, LacI tetramers and mRNA at 1 mM IPTG, b_r=0.015min^{-1} and 0.194% (A) and 0.1945% (B) arabinose. Amplitude (C) and period (D) of oscillations of AraC dimers at 1 mM IPTG, b_r=0.015min^{-1} with 0.0005% arabinose increments.

As it can be seen in Fig. 3.17, the change in system's behavior occurred between 0.194% and 0.1945%, passing from damped oscillations to persistent oscillations. This sharp transition can also be inferred in the graphs showing period variation (Fig. 3.18). As consequence, the interval of interest was restricted to 0.194 - 0.196% arabinose concentrations. This resulting interval was again analyzed by reducing the step size of arabinose concentration to $5 \cdot 10^{-5}$ (10 times smaller than the simulations before).

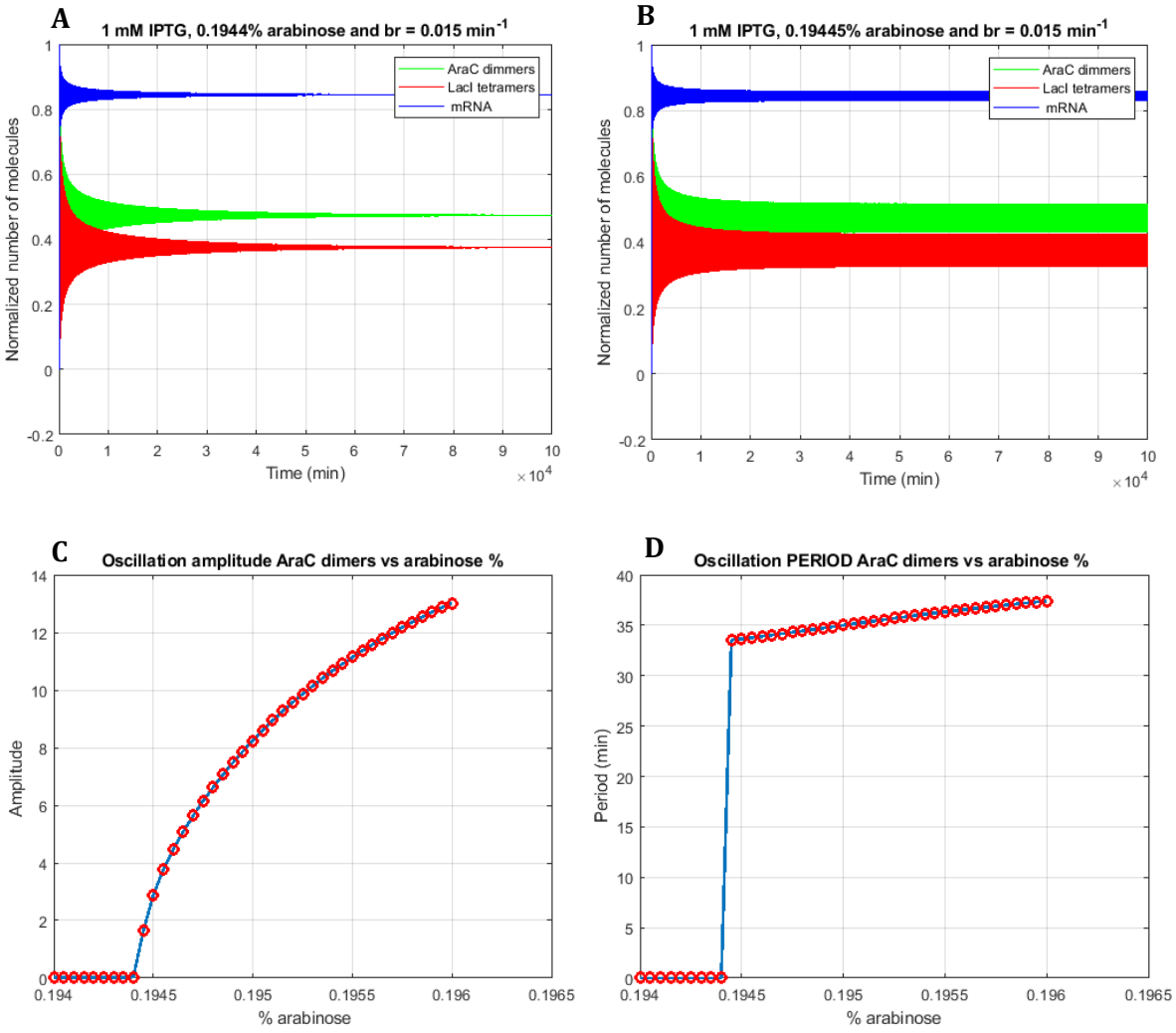

Fig. 3.18 Time dynamics of AraC dimers, LacI tetramers and mRNA at 1 mM IPTG, b_r=0.015min^{-1} and 0.1944% (A) and 0.19445% (B) arabinose. Amplitude (C) and period (D) of oscillations of AraC dimers at 1 mM IPTG, b_r=0.015min^{-1} with 0.00005% arabinose increments.

These simulations allowed to restrict the interval of arabinose concentration values containing the bifurcation point to 0.1944% to 0.1945% arabinose. Such interval was analyzed in the following simulations, reducing the arabinose increments to $5 \cdot 10^{-6}$ and increasing the time up to 500,000 minutes.

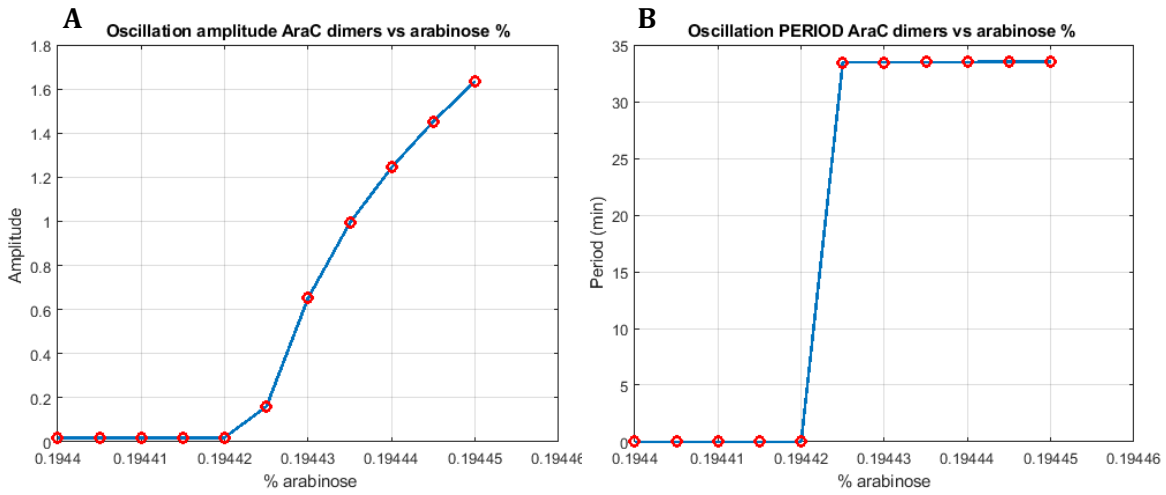

Fig. 3.19 Amplitude (A) and period (B) of oscillations of AraC dimers at 1 mM IPTG, b_r=0.015min^{-1} with 0.000005% arabinose increments.

Fig. 3.19 shows how the location of the bifurcation point is somewhere between 0.19442% and 0.194425% arabinose. Interestingly, as we get closer to the bifurcation point, the amplitudes become progressively smaller. Finally, a last interval restriction was performed by simulating the system within that interval with smaller arabinose increments. These results are displayed below:

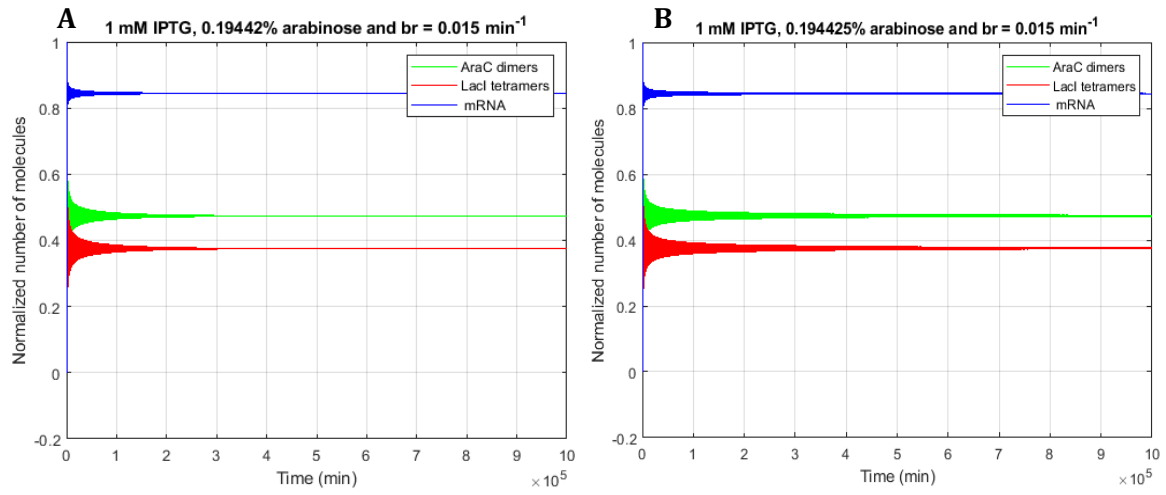

Fig. 3.20 Time dynamics of AraC dimers, LacI tetramers and mRNA at 1 mM IPTG, b_r=0.015min^{-1} and 0.199442% (A) and 0.194425% (B) arabinose concentrations.

Although apparently similar, those two arabinose concentrations are responsible for inducing different system's behaviors. If the graphs in the Fig. 3.18 are amplified, the difference becomes more evident.

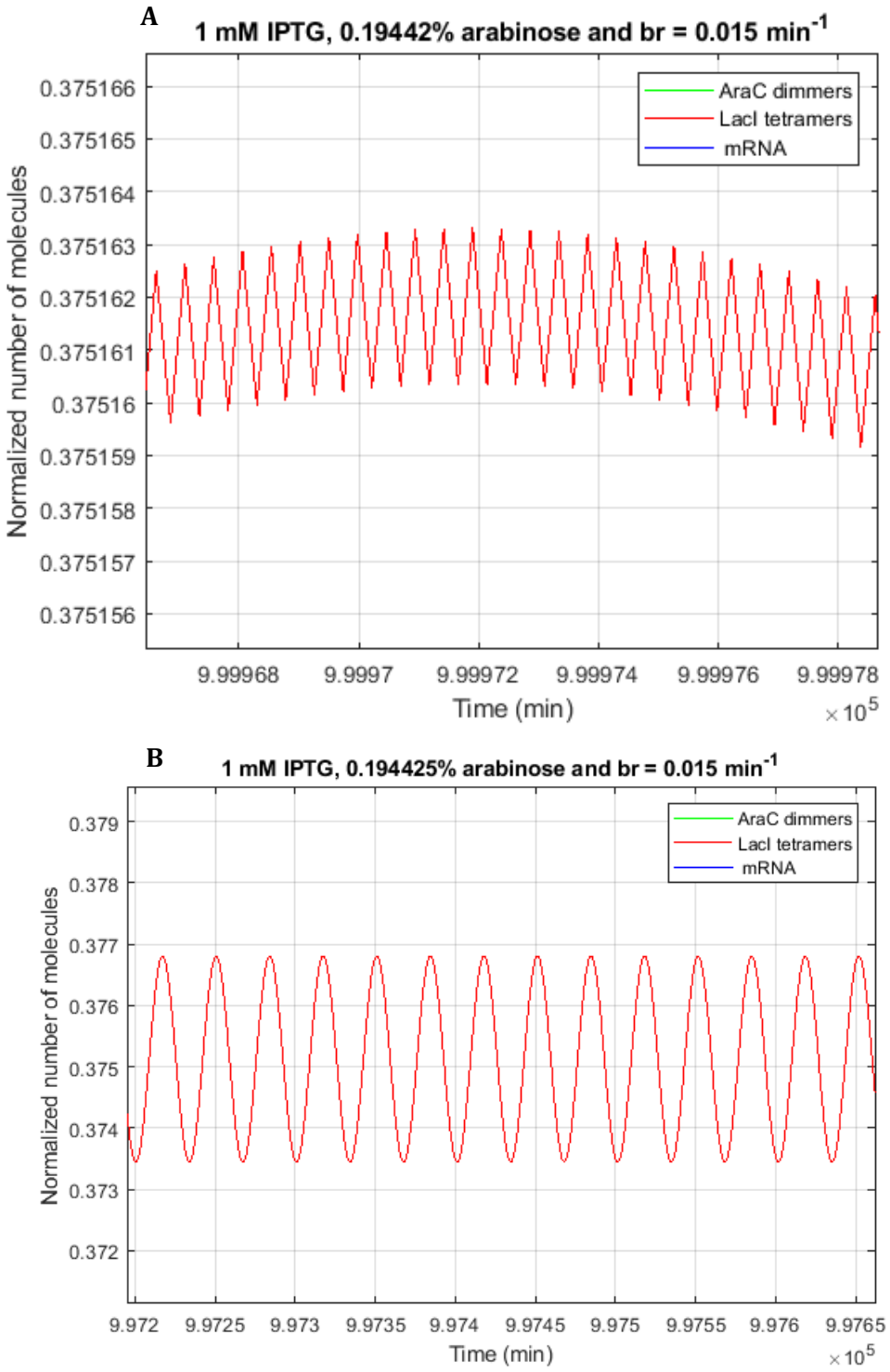

Fig 3.21 Amplification of fig. 3.20. LacI tetramer variation.

Fig. 3.21 confirms such thought: the arabinose concentrations induce different behaviors. At 0.19442% arabinose the oscillations do not come from the system itself but from the integrator used to solve the differential equations. Consequently, the resulting oscillations do not have a sinusoidal but a triangular shape and are much smaller in amplitude than real oscillations observed at 0.194425% arabinose. Thus, we

consider 0.19442% of arabinose as a steady solution in which oscillations are assumed to extinguish while the system does oscillate at 0.194425% arabinose.

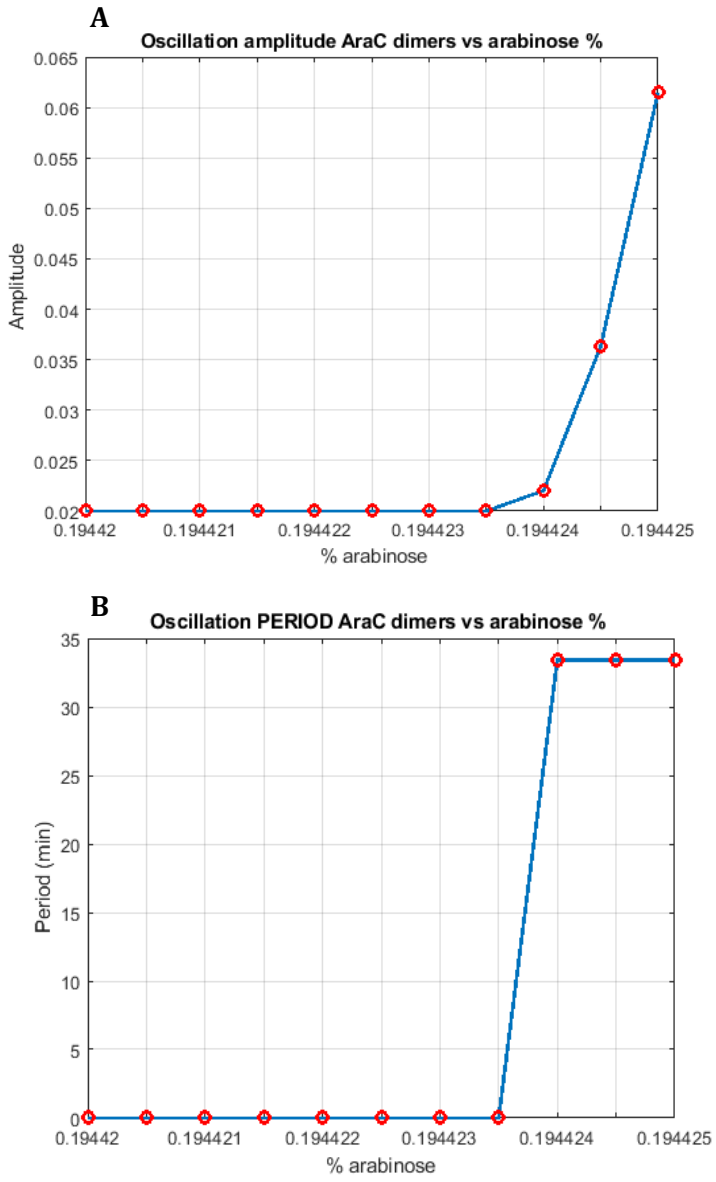

Fig. 3.22 Amplitude (A) and period (B) of oscillations at 1 mM IPTG, b_r=0.015min^{-1} with 0.0000005% arabinose increments.

These final simulations made possible to restrict the location to the bifurcation point to the interval going from 0.1944235-0.1994425 % arabinose concentrations.

4. Discussion and conclusion

4.1 Interpretation of the results

The system under study is called a relaxation oscillator or *relaxoscillator* which codifies for two different transcription factors: one activator and a repressor. Both genes have the same promoter which can be regulated by both transcription factors (the activator and the repressor). This system is able to produce persistent oscillations. Initially, the activator and the repressor are expressed together, but then, when there is enough repressor, this cuts off gene expression of both transcription factors. Consequently, the number of activators and repressors is reduced. After some time, repression is relieved and gene expression of both transcription factors starts, initiating again the cycle. The name of relaxoscillator is due to the alternation of periods of gene expression and periods without expression. One of the features of this system that enables oscillations is the time delay between reactions. If there was no time delay between gene expression and gene repression processes, there would be no oscillations. As soon as the expression of repressor was different from zero, it would automatically repress gene expression. The time delay between reactions was achieved by the inclusion of intermediate reactions such as folding of protein precursors, protein multimerization, translation or DNA looping. Hasty and coworkers [20] exposed it was mandatory to consider these intermediate processes in order to reproduce the experimental behavior of the system.

The coexistence of an activator and a repressor acting on the same promoter requires to be examined. As it was shown in Fig. 2.1, this dual-opposite regulation of the genes results into two feedback loops. The negative feedback loop is represented by the inhibition of the transcription in both genes by LacI tetramers. On the other hand, the positive feedback loop is caused by the activation of transcription via AraC dimers. The strength of these feedback loops can be controlled with two parameters: arabinose and IPTG. The negative feedback loop will be weakened if there is enough IPTG. At the same time, the increase in arabinose concentration will reinforce the positive feedback loop.

Initially, the goal of the simulations was to observe the behavior of the system under different conditions. Arabinose and IPTG concentrations were the two control parameters allowing for different simulations. As described in equations 2.20 and 2.21, these inducers affect the forward binding rate of the transcription factors - AraC dimers and LacI tetramers - to the operator sites. Their association to the activator or repressor promotes or hinders the successive binding process between the transcription factor and the promoter. In case an AraC dimer binds to arabinose, the binding rate of the activator - AraC dimer - to the promoter will be faster and then, the transcription rate will also be

faster due to the presence of the activator. Therefore, more arabinose will result into higher transcription rates of both genes and thus higher protein expression. In the simulations, increasing the arabinose concentration keeping fixed IPTG concentration resulted into increased amplitude of LacI tetramers and a decrease in AraC dimer amplitude with a combined increase in period (Fig. 3.6A and Fig 3.8A). Increasing arabinose at constant IPTG means higher binding association rate of AraC dimers to the promoter. The binding of the activator causes faster transcription rates in both genes. The increase in the number of mRNA transcripts eventually causes an increase in AraC dimers. However, as it can be seen in Fig. 3.6A, the number of AraC dimers appears to be reduced with higher arabinose. The reason is that those AraC dimers refer to the number of unbound AraC dimers. Therefore, a decrease in the amplitude as it can be seen in the figure indicates in fact an increase in the number of AraC dimers bound to the promoter. This makes sense since the direct effect of arabinose was precisely favoring that binding.

Regarding LacI tetramers, they increase in amplitude because of the previously mentioned increase of transcription rate. An increase of arabinose reinforces the positive feedback loop which in turns causes higher number of transcription factors. The number of unbound AraC dimers decreases because such reinforcement of the positive feedback loop promotes the binding of these dimers to the promoter.

When IPTG is high, independently of the arabinose concentrations, the period and the amplitude of oscillations is almost identical, as IPTG has interfering effects on AraC-arabinose binding, and therefore, on the activation of the promoter [20].In the second type of simulation, arabinose is kept constant while IPTG concentration varies. When arabinose concentration is very low, similarly to what occur at very high concentrations IPTG, the amplitude and the period of oscillations remain almost constant independently on IPTG variations (Fig. 3.5A and Fig. 3.7A).

When there is enough arabinose and IPTG is increased, AraC dimers increase in amplitude while LacI tetramers decrease in amplitude. IPTG binds to LacI tetramers - the repressor – and to AraC dimers – the activator -, avoiding their binding to the promoter. According to the results, AraC dimers grow in amplitude (Fig. 3.5B) while LacI tetramers decrease in amplitude (Fig. 3.7B) for increasing IPTG. Therefore, IPTG increase will diminish the strength of the negative feedback loop by decreasing the number of LacI tetramers. Regarding AraC dimers, IPTG will cause unbound dimers to grow in number –since it will lower the binding rate to the promoter – as it can be seen in figure 3.5B.

This initial set of results allowed to understand the behavior of the system under different conditions. The system's behavior could be considered to be uniform, meaning that independently on the inducer values, the system always displayed sustained oscillations. From a mathematical point of view, this oscillating long term behavior of

the system corresponds to the existence of a limit cycle at those values. One of the goals of this thesis was the mathematical interpretation of the behavior of system. However, the lack of variability in system's behavior oversimplified too much the analysis. Therefore, in an attempt to find those changes in the long term behavior of the system we decided to simulate the system under the control of a third parameter: the transcription rate of the repressor. From a control point of view, the negative feedback - provided there is time delay - is the one responsible for the oscillating behavior. Therefore, from a theoretical point of view, decreasing the strength of the negative feedback loop should eventually cause the extinction of the oscillations. The influence of the negative feedback loop in the system was reduced by decreasing the transcription rate of the repressor gene - LacI gene -. Several simulations were performed with reduced b_r values (Fig. 3.12, Fig. 3.13). From these simulations several ideas could be extracted:

- ✓ For arabinose concentrations greater than 0.4% the system displayed oscillations for low values of *br* (Fig. 3.13). Arabinose reinforces gene expression of both genes. Therefore, more arabinose indirectly causes also to reinforce the negative feedback loop as there will be more LacI expression, and then, it will be more likely that a LacI tetramer binds to the operator inducing repression and then oscillations even if *br* has been reduced.

- ✓ There were no remarkable differences between simulations at 1mM and 2 mM IPTG, due to the proximity between these concentrations.

- ✓ Even at low arabinose concentrations, convergence to a steady state (a non oscillating behavior) was only observed for relatively low values of *br* (less than a third of the original value).

Therefore, the main conclusion obtained from this second set of simulations was that in order to appreciate changes in the long term system's behavior, two conditions had to be satisfied: low arabinose concentrations and low repressor transcription rate. Those conditions made possible to see transitions from a steady state to persistent oscillations. From a mathematical point of view, the transitions between those states are the confirmation of the existence of a bifurcation. The presence of persistent oscillations is an indicator of a limit cycle while its absence or convergence to a steady solution indicates the existence of a fixed point. This change in the dynamics of the system was hypothesized to be due to a Hopf bifurcation. According to the theory, Hopf bifurcations occur whenever the behavior of the system changes from damped oscillations decaying progressively to the steady state as the parameter is varied, to persistent limit cycle oscillations around the previous steady state. If the limit cycle oscillations initially start from zero amplitude, such Hopf bifurcation is called *supercritical*. However, a discontinuous change of oscillations' amplitude will make the bifurcation to be classified as *subcritical* (see Fig. 1.11). This jump in amplitude is

specifically one of the features that was used to infer which type of Hopf bifurcation the system was going through.

Although there is an analytical criterion to determine the type of Hopf bifurcation, this can become difficult to apply [19]. Consequently, in order to take that decision, we decided to perform simulations within a parameter interval whose end values were known to cause different long term behaviors. To do so we decided to perform the simulations between 0.1% and 0.2% arabinose concentrations, at 1 mM IPTG and $b_r = 0.015\ min^{-1}$. As it had been checked before, the simulation of the system at the lower limit of arabinose resulted into the convergence to a steady state, while the system had persistent oscillations for the upper limit of arabinose concentrations. An important factor to consider in this third set of simulations was the time. As previously stated, starting from damped oscillations, the closest the system is to the bifurcation point, the more time these oscillations will take to decay. Therefore, these simulations will require much longer times to confirm whether there is an effective decay or not.

Firstly, the simulations were performed between 0.1% and 0.2% choosing increments of 0.005% in arabinose concentration. The results proved that at 0.19% arabinose concentration oscillations were damped to the steady state. Starting at 0.1% arabinose, successive values induced damped oscillations longer time (Fig. 3.14 and 3.15A). This is progressively longer ringing before reaching the steady state was one of the features of Hopf bifurcations above described. The amplitude and period of oscillations was also computed in such a way that if the oscillations decayed, both were considered to be zero. Although the amplitude seemed to increase drastically (Fig. 3.16A), this increase was not due to the bifurcation itself but due to the big step size within the interval of arabinose. Whenever analyzing the amplitude of the oscillations once the bifurcation point has been crossed, one must be sure that the control parameter is effectively close to that bifurcation point, something that could not be confirmed when using such big step size.

Consequently, the simulations between 0.1% and 0.2% allowed to restrict the parameter range containing the bifurcation point. In the following simulations, the interval was restricted to 0.19% and 0.20%, the step size of arabinose increments would be reduced ten times and the simulation time would increase ten times more, in prevision of a longer oscillation decay (Fig. 3.17A and B). The same process was repeated again. Analyzing the amplitude and the period of the oscillations allowed to restrict the interval to 0.194% and 0.196% of arabinose (Fig. 3.17C and D). As it can be seen when the amplitude is plotted against arabinose increments (Fig. 3.17C and D), the amplitude seemed to grow fast. At this point, there were two possible reasons explaining that sudden increase: the Hopf bifurcation was subcritical or the step size used as arabinose increments. Too big arabinose increments cause consecutive arabinose concentrations to be too far from each other to consider significant the increase in amplitude.

In an attempt to solve that doubt, a third set of simulations was performed in that interval [0.194% - 0.196% arabinose] with smaller arabinose increments ($5 \cdot 10^{-5}$ %) (Fig. 3.18A and B). This last set of simulations successfully restricted even more the interval of interest, but more importantly, they show a curve with a peculiar shape as it can be observe in Fig.3.18C. One of the conditions of a supercritical Hopf bifurcation is that the amplitude of oscillations - the size of the limit cycle – has to increase continuously from zero, proportionally to $\sqrt{\mu - \mu_c}$ provided that the system is close to the bifurcation point. In that expression, μ_c represents the searched bifurcation point and μ is the varying arabinose. The fact of having that shape implies the satisfaction of one of the characteristics that supercritical bifurcation has [19].

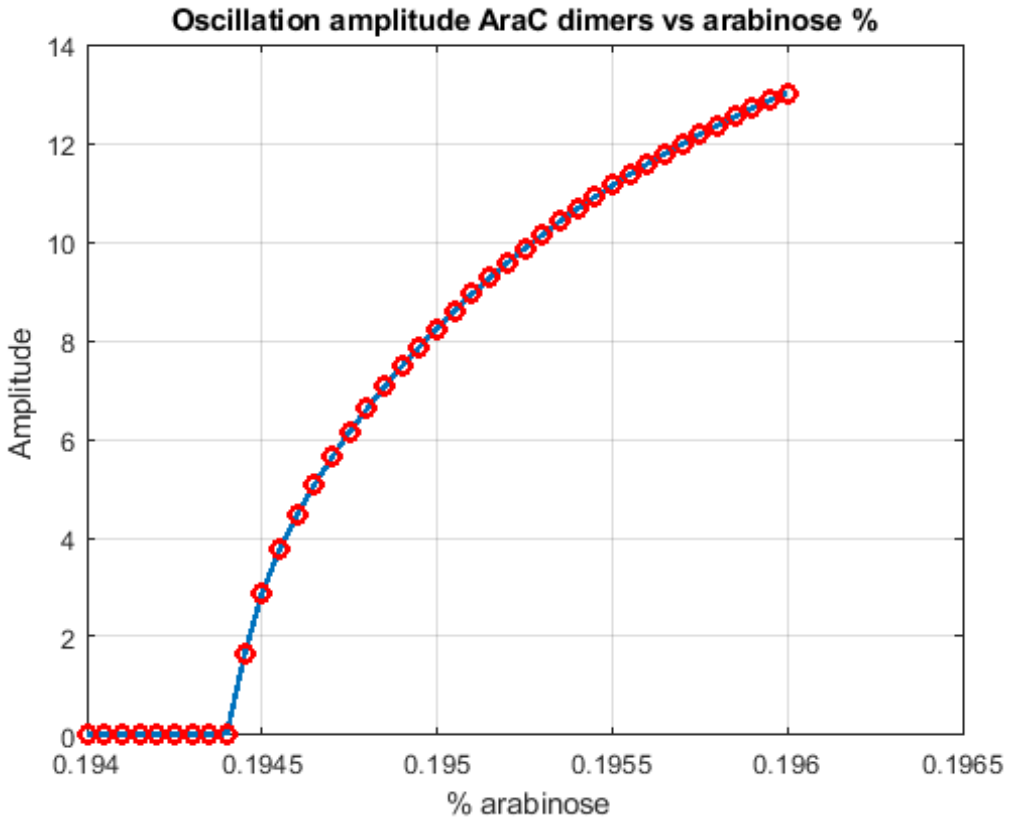

Fig. 4.1 Amplitude at 1 mM IPTG and b_r=0.015min^{-1} for varying arabinose.

In order to confirm if our system follows that condition, amplitude variation against arabinose increments was compared to a function proportional to $\sqrt{\mu - \mu_c}$:

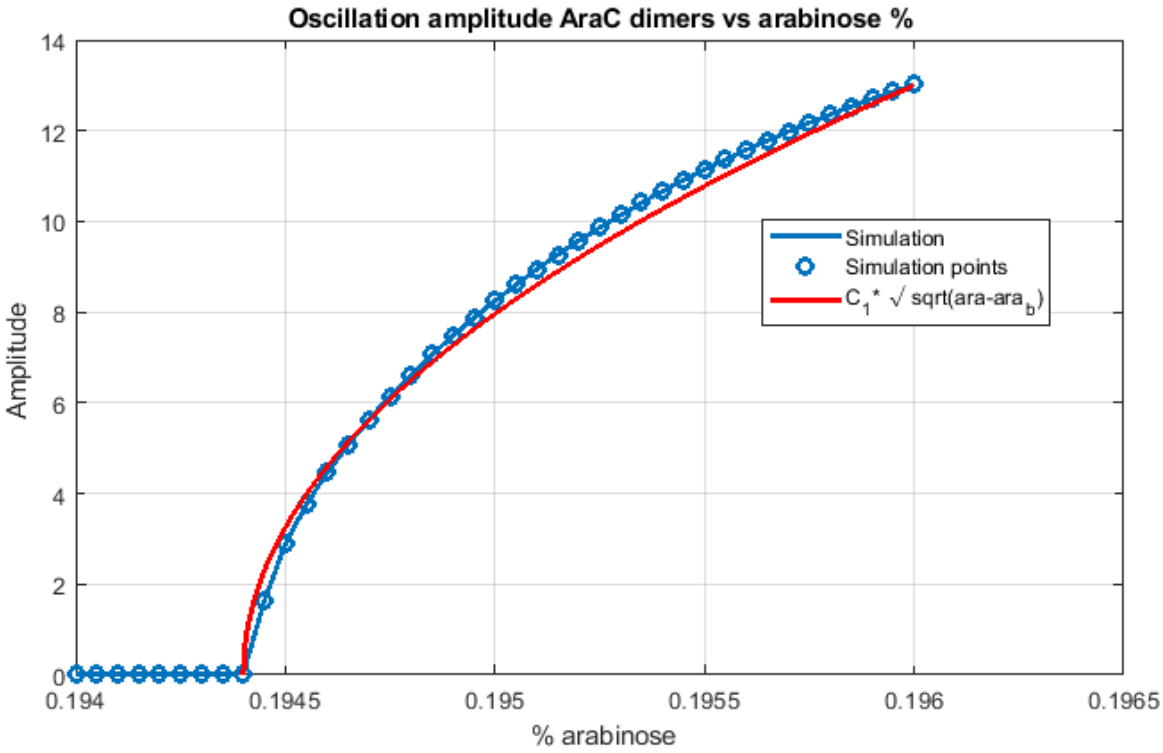

Fig. 4.2: Comparison between measured amplitude at different arabinose concentrations (in percentage) and theoretical prediction. *Ara* corresponds to arabinose values (*ara* ϵ (0.1944 – 0.196 % arabinose) and Ara_b refers to the bifurcation point, in this prediction it was taken ara_b=0.1944% arabinose concentration. C_1 is a constant of proportionality (C_1=*325*).

As it can be observed, the results clearly matched the theoretical prediction (Fig. 4.2), and therefore, it satisfies one of the characteristics features of supercritical Hopf bifurcations. This discovery indicated the possible existence of a supercritical Hopf bifurcation around 0.1944% and 0.1945%. In spite of this, two more precise sets of simulations were performed to discard or reduce the possibility of a sudden increase in amplitude that would refute the hypothesis of a supercritical bifurcation. Simulations between (0.1944%-0.1945%) first (Fig. 3.19) and then (0.19442%-0.1994425%) arabinose concentrations (Fig. 3.20A and B) did not show sudden grows in the amplitude of the oscillations (Fig. 3.19A and Fig. 3.22A), reinforcing the hypothesis of the existence of a supercritical Hopf bifurcation. Notice that although the slope of the amplitude curve in Fig. 3.22A seems to be high, the increase in amplitude is small according to the values in the y-axis.

Therefore, from the results obtained in section *3.3 Restricting the interval*, two main general conclusions can be extracted. The first one is that the specific change in system's behavior was very likely due to a supercritical bifurcation. This bifurcation causes damped oscillations converging to the steady state to decay progressively slower

as they approach the bifurcation point. In contrast, once the bifurcation point has been crossed, there will be persistent oscillations of increasing amplitude (starting at zero amplitude). From a stability perspective, for smaller values than the bifurcation point, the fixed point would consist of a stable spiral that would attract neighboring trajectories to the steady solution. Such stable spiral would lose its stability when the bifurcation point is crossed, becoming into an unstable spiral that would be attracted and surrounded by a stable limit cycle. Second conclusion is that the bifurcation point should be somewhere between 0.194423% and 0.194424% of arabinose concentration.

4.2 Conclusion

This bachelor thesis can be divided into three different parts. The first one corresponds to the comprehension of both components and interactions of the gene regulatory network from a biological point of view. The second part consists of the development of a mathematical model to simulate the system. Finally the third part includes the implementation of multiple simulations of the system for later evaluation of the results.

Without considering knowledge acquisition process performed prior to the beginning of this thesis, the first part of this bachelor thesis consists of the comprehension of the system from a general perspective. Understanding the basis of the components and interactions of the genetic network – *relaxoscillator* - before building the model was crucial. The coexistence of two genes regulated by the same promoter whose expressed proteins act as transcription factors enhancing or inhibiting gene expression was the key concept. It is important to emphasize the following: in some biological systems, the direct observation of the interactions' diagram is enough to successfully predict the long term behavior of the system. In that case, the mathematical model will only confirm the initial predictions. However, in some other cases, when dealing with non-linear systems, human intuition can fail when trying to predict system's behaviors under different conditions. In those systems, mathematical models are the only valid tools for analyzing the system. The relaxoscillator is a complex nonlinear system so in order to understand and predict the behavior of the system we need to use mathematical modeling tools.

This genetic network is included among the latter systems explaining why the mathematical model and simulations are needed to understand and predict the behavior of the system with higher accuracy.

The development of the mathematical model was the second objective of this thesis. The chemical reactions describing the interactions were available [23], and the differential equations could be extracted from them. The most important feature to be considered when modeling this specific gene network was time delay. The time delay in the negative feedback loop is the one responsible for the oscillations exhibited by the

system. As stated in the previous section, in the reference article [20], researchers realized about the need of including the time delay once they simulated the system and checked that its behavior did not match with the measured data. This time delay was introduced in the form of intermediate reactions like folding or dimerization.

With the mathematical model built, the second objective was satisfied. The last goal of this thesis was to perform the analysis of the system under different conditions from a both a biological and a mathematical perspective. In order to do so, the different parameters were established and the system was simulated in Matlab. In order to analyze the system it was decided to move from the general to the particular case. Therefore, multiple simulations were performed varying the two control parameters: the inducers (IPTG and arabinose). The influence of inducers was not trivial to interpret. IPTG was shown to decrease the strength of the negative feedback loop while arabinose reinforced the positive one. Additionally, this initial set of simulations also enabled to observe how the period and the amplitude of oscillations were changed due to inducer variations, matching the statements that had been made by the researchers [20].

These simulations is did not show the any variability in system's behavior. Although they were different in amplitude and period, all simulations displayed sustained oscillations. This lack of variability in system's behavior clashed with the last proposed objective. In front of this situation, the analysis of the behavior of the system was reduced to the existence of a limit cycle responsible for the oscillations. Then, the next step had to be obtaining some variations in the behavior of the system, like a non-oscillating behavior. In this type of systems, the reduction of the negative feedback loop is known to negatively control the oscillatory behavior. To do so, the repressor transcription rate was added as the third control parameters. The system was simulated over different conditions with lower repressor transcription rate than the original case. The results showed a suppression of the oscillations.

The transition between damped oscillations that converged into a steady state to a limit cycle oscillatory behavior suggested the existence of a Hopf bifurcation. The third set of simulations was preformed to evaluate such hypothesis as well as restricting the interval containing the bifurcation point. The obtained results, especially the shape of the amplitude curve for increasing arabinose concentrations (Fig. 4.2) reinforced it. It suggested the existence of a supercritical bifurcation near to 0.19442% arabinose concentration. Such finding supposed achievement of the last objective proposed.

5. Future work:

Although in general terms the objectives were achieved, there were several areas susceptible to be improved and therefore, candidates for future research lines. One of the main limitations found during the project was the theoretical nature of itself. Ideally, the study of a biological system like this one should encompass both theoretical and practical studies. Although the system under study had been previously experimentally analyzed by the researchers, such analysis was performed in based to their needs.

Therefore, two main research lines can be followed after this project:

- ✓ To develop an experimental study to find the predicted Hopf bifurcation. The experimental confirmation of its presence would reinforce the validity of the developed mathematical model. One of the main obstacles this experimental study would have to overcome would be the extremely accuracy needed to find the bifurcation point. As shown in the results, the bifurcation could only be inferred when using very small arabinose increments (up to $5 \cdot 10^{-7}$%). Therefore, the experimental study would require such level of precision in the measurements.

- ✓ To perform a theoretical study using two-parameter bifurcation analysis to gain a global understanding of the qualitative behavior of the system as a function of the inducer concentrations.

6. Socio-economic impact and budget:

6.1 Socio-economic impact

As mentioned in the introduction, synthetic biology was born with the aim to develop new applications primarily in medicine and biotechnology. In the future, synthetic genetic circuits will be designed for biofuel production or vaccine generation among other purposes. Since the beginning, synthetic biology has progressed in the development of synthetic circuits. The successful development of this discipline has been favored by the parallel progress of mathematics, computational science and technology. In spite of this, the contributions of these disciplines would have been useless if those biological systems had not been understood and modeled before. This means that in order to achieve its foundational goal, synthetic biology demands a previous comprehension of the biological system to be modeled. Such prior knowledge acquisition can be through experimental measures and mathematical model construction. Therefore, modeling biological systems will be needed before the construction of the synthetic circuits. The impact of this bachelor thesis must be understood as the contribution to the acquisition of knowledge about the given system. Although the validation of the model requires its comparison to the real system, once the model is validated, it can be used to predict the behavior of the system under conditions that could be never replicated *in vitro* but that could be possible in the real system. These simulations would allow one to determine critical unstable operating points that should be avoided when using the synthetic circuit for different applications.

This bachelor thesis can be understood as one of those intermediate stages needed by synthetic biology to appropriately design the corresponding gene circuit. The design of a genetic circuit capable of expressing insulin every day at regular interval may sound utopist yet, but when it is achieved, it will be thanks to prior study and analysis of that biological system. The impact of this bachelor thesis is included as a part of that initial stage.

6.2 Budget

TABLE 6.1 BUDGET BREAKDOWN

Project description	
Author:	Gabriel Rodríguez Maroto
Department:	Mathematics Biomedical and Aerospace Engineering
Title	Synthetic biology of genetic circuits
Duration	7 months

Budget breakdown			
Processing cost			
Description	**Number of units**	**Cost per unit (€)**	**Total cost (€)**
Computer	1	1,000.00	1,000.00
MATLAB license	1	6,000.00	6,000.00
		Total	**7,000.00**

Human resources			
Description	**Number of hours**	**Cost per hour (€):**	**Total cost (€)**
Student	600	8.00	4,800.00
Supervisor	250	15.00	3,750.00
Tutor	100	30.00	3,000.00
		Total	**12,550.00**

Other direct costs		
Description	Company	Cost (€)
Internet	Telefónica	420.00
Windows 7	Microsoft	85.00
	Total	**505.00**

TOTAL PROJECT COST	**19,055.00 €**

7. Regulatory framework

Regarding this section, the theoretical nature of this bachelor thesis causes the absence of regulatory framework ruling this project. The experiments in this research work were done *in silico*; therefore no guidelines regulating cell manipulation have to be presented and no bioethical controversy arises. The license of the computational software used, Matlab, was provided by the Universidad Carlos III de Madrid. In relation to intellectual property, original ideas are properly cited and referred when addressing the corresponding scientific topic.

8. Bibliography

[1] J. Hasty, D. Mcmillen, and J. J. Collins, “Engineered gene circuits” , *Nature*, vol. 420, no. 6912, pp. 224–230, 2002.

[2] M. B. Elowitz and S. Leibler, “A synthetic oscillatory network of transcriptional regulators”, *Nature,* vol. 403, no. 6767, pp. 335–338, 2000.

[3] T. S. Gardner, C. R. Cantor, and J. J. Collins, “Construction of a genetic toggle switch in Escherichia coli”, *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.

[4] D. E. Cameron, C. J. Bashor, and J. J. Collins, “A brief history of synthetic biology”, *Nature Reviews Microbiology*, vol. 12, no. 5, pp. 381–390, Jan. 2014

[5] M. Fussenegger, “Synchronized bacterial clocks”, *Nature*, vol. 463, no. 7279, pp. 301–302, 2010.

[6] J. C. Smuts, *Holism and evolution*. La Vergne, TN: Kessinger Publ., 2010

[7] P. D. Backer, D. D. Waele, and L. V. Speybroeck, “Ins and Outs of Systems Biology vis-à-vis Molecular Biology: Continuation or Clear Cut?”, *Acta Biotheoretica*, vol. 58, no. 1, pp. 15–49, 2009.

[8] M. A. Omalley, A. Powell, J. F. Davies, and J. Calvert, “Knowledge-making distinctions in synthetic biology”, *BioEssays*, vol. 30, no. 1, pp. 57–65, 2008.

[9] F. C. Boogerd, F. J. Bruggeman, J.-H. S. Hofmeyr, and H. V. Westerhoff, “Towards philosophical foundations of Systems Biology: introduction”, *Systems Biology*, pp. 3–19, 2007.

[10] W. Bechtel and A. A. Abrahamsen, “Thinking Dynamically About Biological Mechanisms: Networks of Coupled Oscillators”, *Foundations of Science*, vol. 18, no. 4, pp. 707–723, 2012.

[11] R. D. Sleator, “The synthetic biology future”, *Bioengineered*, vol. 5, no. 2, pp. 69–72, 2014.

[12] B. P. Ingalls, Mathematical modeling in systems biology: an introduction. Cambridge (MA),: M.I.T. Press, 2013.

[13] Y. N. Kaznessis, “Models for synthetic biology”, *BMC Systems Biology*, vol. 1, no. 1, p. 47, Nov. 2007.

[14] A. Maria, “Introduction to modeling and simulation”, Proceedings of the 29th conference on Winter simulation - WSC 97, 1997.

[15] J. C. Wooley and H. Lin, Catalyzing inquiry at the interface of computing and biology. Washington, D.C.: National Academies Press, 2005.

[16] B. P. Zeigler, Prähofer Herbert, and T. G. Kim, Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems. Amsterdam: Academic Press, 2010.

[17] D. Angeli, “A Tutorial on Chemical Reaction Network Dynamics”, *European Journal of Control*, vol. 15, no. 3-4, pp. 398–406, 2009.

[18] K. J. Laidler, “Law of mass action,” Encyclopædia Britannica, 26-Oct-2016. [Online]. Available: https://www.britannica.com/science/law-of-mass-action. [Accessed: 18-Mar-2018].

[19] S. Strogatz, Nonlinear dynamics and Chaos: with applications to physics, biology, chemistry, and engineering. Reading, MA: Perseus Books, 1998.

[20] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, “A fast, robust and tunable synthetic gene oscillator”, *Nature*, vol. 456, no. 7221, pp. 516–519, 2008.

[21] T. Y.-C. Tsai, Y. S. Choi, W. Ma, J. R. Pomerening, C. Tang, and J. E. Ferrell, “Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops”, *Science*, vol. 321, no. 5885, pp. 126–129, Apr. 2008.

[22] M. Tigges, T. T. Marquez-Lago, J. Stelling, and M. Fussenegger, “A tunable synthetic mammalian oscillator”, *Nature*, vol. 457, no. 7227, pp. 309–312, 2009.

[23] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, “Supplementary information: A fast, robust and tunable synthetic gene oscillator”, *Nature*, vol. 456, no. 7221, pp. 516–519, 2008.

[24] G. M. Cooper and R. E. Hausman, The cell a molecular approach. Washington, DC: ASM Press, 2003.

[25] “odeset,” Probabilistic Roadmaps (PRM) - MATLAB & Simulink - MathWorks España. [Online]. Available: https://es.mathworks.com/help/matlab/math/choose-an-ode-solver.html. [Accessed: 20-Apr-2018].

[26] M. Spijker, “Stiffness in numerical initial-value problems”, *Journal of Computational and Applied Mathematics*, vol. 72, no. 2, pp. 393–406, Oct. 1996.

[27] “Stiff Differential Equations,” Probabilistic Roadmaps (PRM) - MATLAB & Simulink - MathWorks España. [Online]. Available: https://es.mathworks.com/company/newsletters/articles/stiff-differential-equations.html. [Accessed: 18-Jun-2018].

[28] P. Bogacki and L. Shampine, “A 3(2) pair of Runge - Kutta formulas”, *Applied Mathematics Letters*, vol. 2, no. 4, pp. 321–325, 1989.

[29] L. F. Shampine and M. W. Reichelt, “The MATLAB ODE Suite”, *SIAM Journal on Scientific Computing*, vol. 18, no. 1, pp. 1–22, 1997.

Appendix A: ODEs of repressor plasmid

$$\frac{dP^r_{0,0}(t)}{dt} = -k_a P^r_{0,0}(t)a_2(t) + k_{-a}P^r_{1,0}(t) - 2k_r P^r_{0,0}(t)r_4(t) + k_{-r}P^r_{0,0}(t) + k_{ul}P^r_{L,0}(t) + f(X)P^r_{1,0}(t) + f(X)P^r_{0,1}(t) \quad (A.1)$$

$$\frac{dP^r_{1,0}(t)}{dt} = k_a P^r_{0,0}(t)a_2(t) - k_{-a}P^r_{1,0}(t) - 2k_r P^r_{1,0}(t)r_4(t) + k_{-r}P^r_{1,1}(t) - f(X)P^r_{1,0}(t) + f(X)P^r_{1,1}(t) \quad (A.2)$$

$$\frac{dP^r_{0,1}(t)}{dt} = -k_a P^r_{0,1}(t)a_2(t) + k_{-a}P^r_{1,1}(t) + 2k_r P^r_{0,0}(t)r_4(t) - k_{-r}P^r_{0,1}(t) - k_r P^r_{0,1}(t)r_4(t) + 2k_{-r}P^r_{0,2}(t) + f(X)P^r_{1,1}(t) - f(X)P^r_{0,1}(t) + 2f(X)P^r_{0,2}(t) \quad (A.3)$$

$$\frac{dP^r_{1,1}(t)}{dt} = k_a P^r_{0,1}(t)a_2(t) - k_{-a}P^r_{1,1}(t) + 2k_r P^r_{1,0}(t)r_4(t) - k_{-r}P^r_{1,1}(t) - k_r P^r_{1,1}(t)r_4(t) + 2k_{-r}P^r_{1,2}(t) - f(X)P^r_{1,1}(t) - f(X)P^r_{1,1}(t) + 2f(X)P^r_{1,2}(t) \quad (A.4)$$

$$\frac{dP^r_{1,2}(t)}{dt} = -k_a P^r_{0,2}(t)a_2(t) - k_{-a}P^r_{1,2}(t) + k_r P^r_{1,1}(t)r_4(t) - 2k_{-r}P^r_{1,2}(t) - k_l P^r_{1,2}(t) - 2f(X)P^r_{1,2}(t) - f(X)P^r_{1,2}(t) \quad (A.5)$$

$$\frac{dP^r_{0,2}(t)}{dt} = -k_a P^r_{0,2}(t)a_2(t) + k_{-a}P^r_{1,2}(t) + k_r P^r_{0,1}(t)r_4(t) - 2k_{-r}P^r_{0,2}(t) - k_l P^r_{0,2}(t) + f(X)P^r_{1,2}(t) - 2f(X)P^r_{0,2}(t) \quad (A.6)$$

$$\frac{dP^r_{L,0}(t)}{dt} = -k_{ul}P^r_{L,0}(t) + \varepsilon f(X)P^r_{L,1}(t) \quad (A.7)$$

$$\frac{dP^r_{L,1}(t)}{dt} = 2\varepsilon f(X)P^r_{L,2}(t) - \varepsilon f(X)P^r_{L,1}(t) \quad (A.8)$$

$$\frac{dP^r_{L,2}(t)}{dt} = k_l P^r_{0,2}(t) + k_l P^r_{1,2}(t) - 2\varepsilon f(X)P^r_{L,2}(t) \quad (A.9)$$

Appendix B: Simulations under different inducer conditions

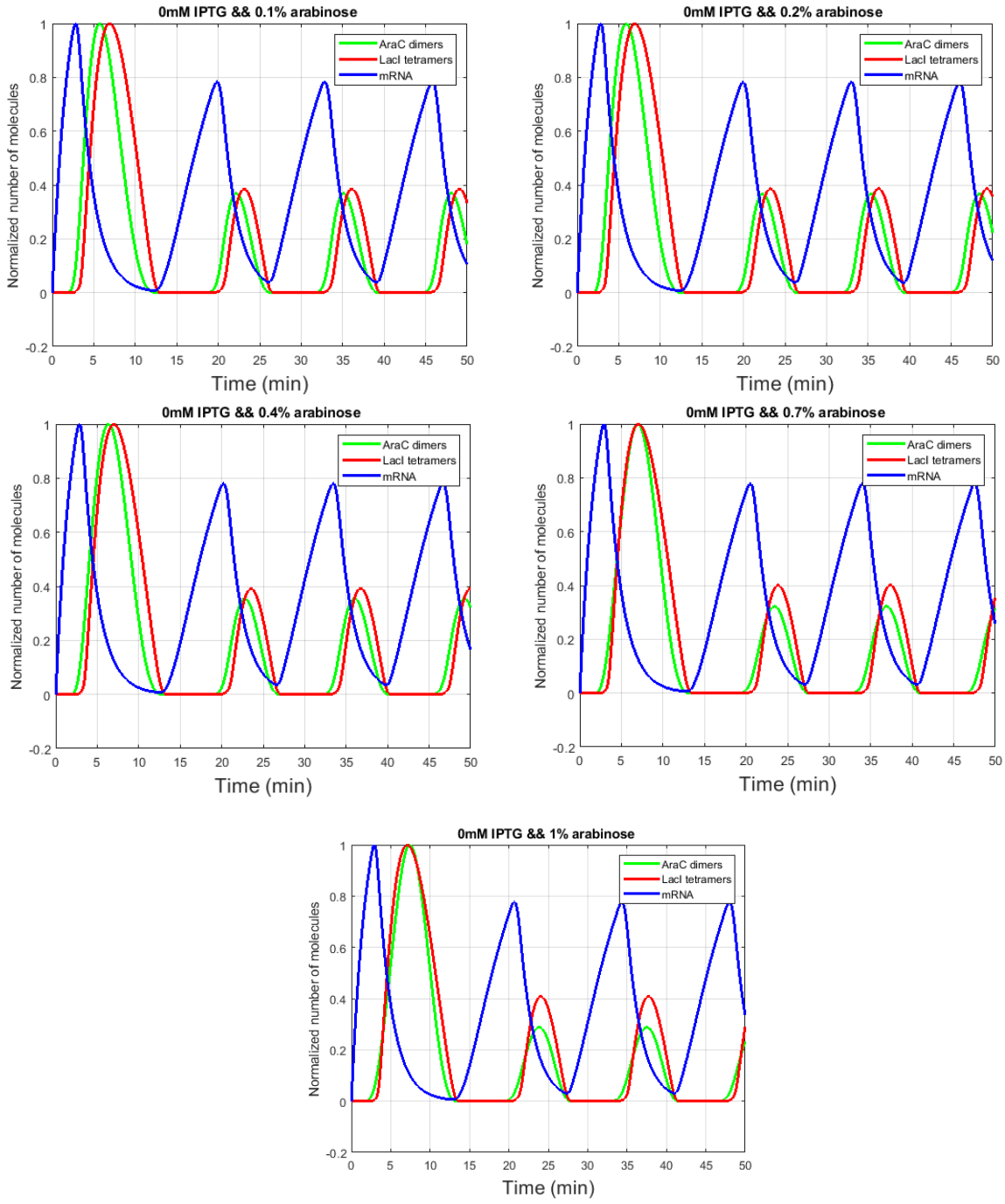

Figure 9.1 Time dynamics of AraC dimers, LacI tetramers and mRNA at fixed 0 Mm of IPTG and at varying arabinose concentrations.

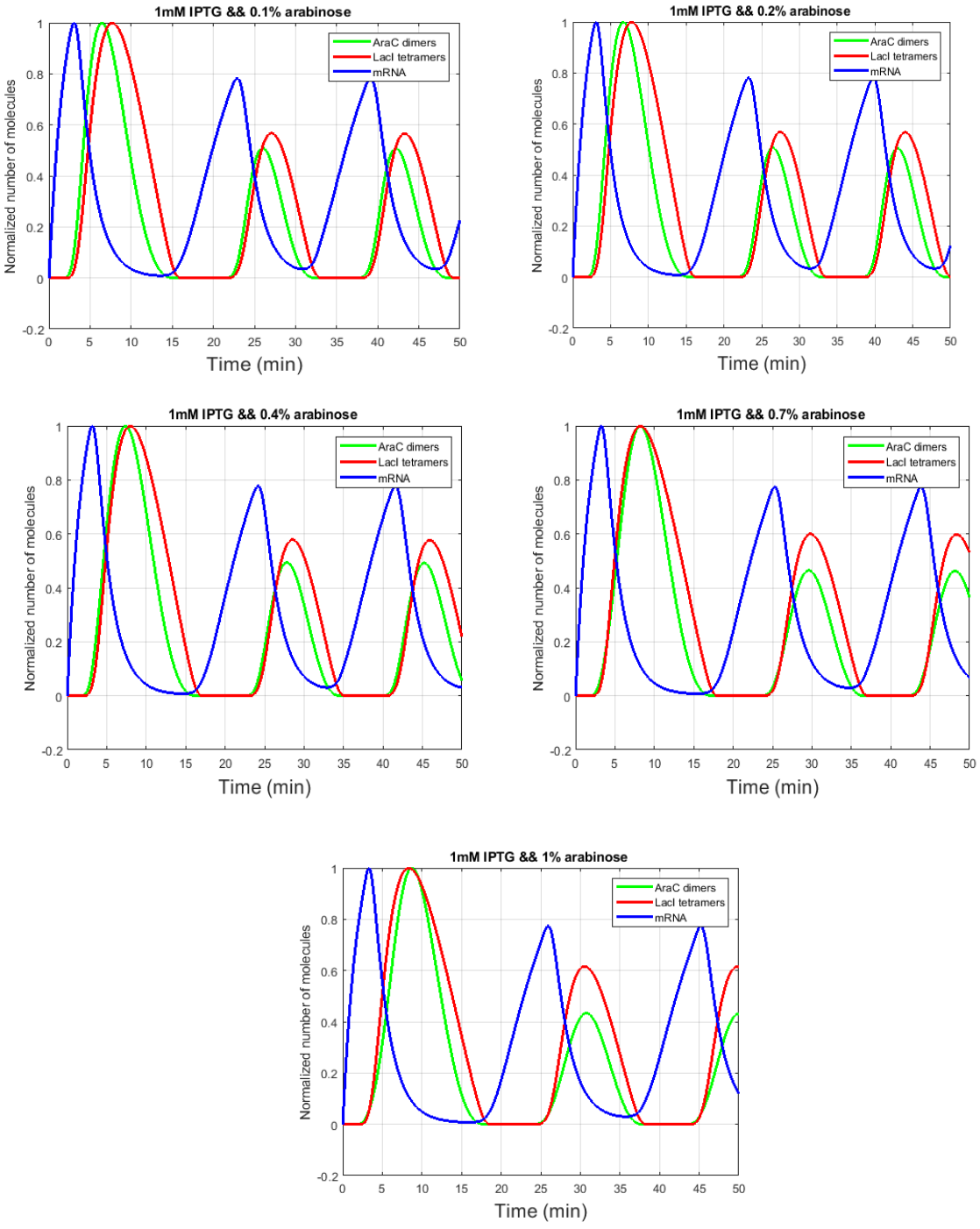

Figure 9.2 Time dynamics of AraC dimers, LacI tetramers and mRNA at fixed 1 Mm of IPTG and at varying arabinose concentrations.

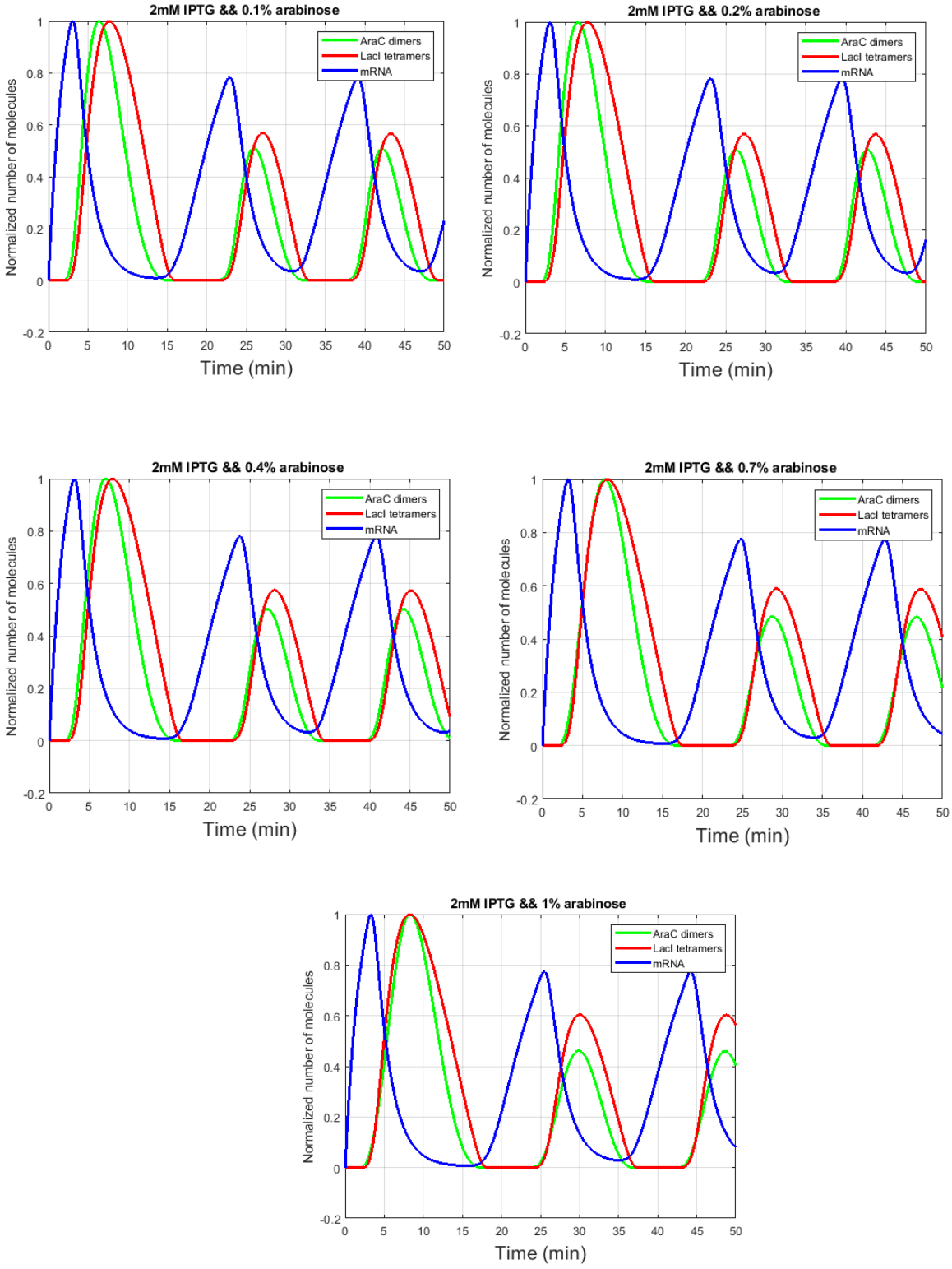

Fig. 9.3 Time dynamics of AraC dimers, LacI tetramers and mRNA at fixed 2 Mm of IPTG and at varying arabinose concentrations.

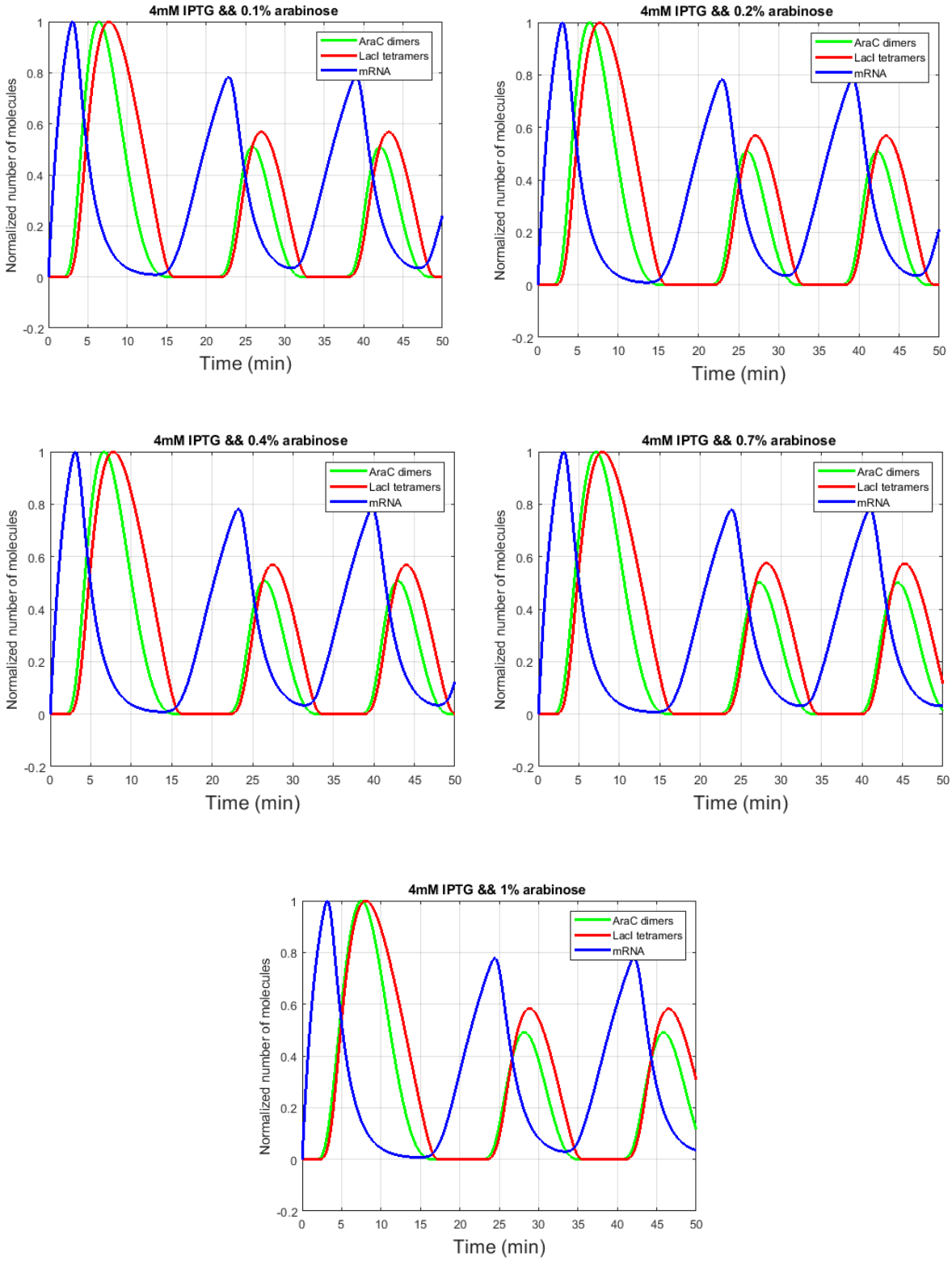

Fig. 9.4 Time dynamics of AraC dimers, LacI tetramers and mRNA at fixed 4 mM of IPTG and at varying arabinose concentrations.

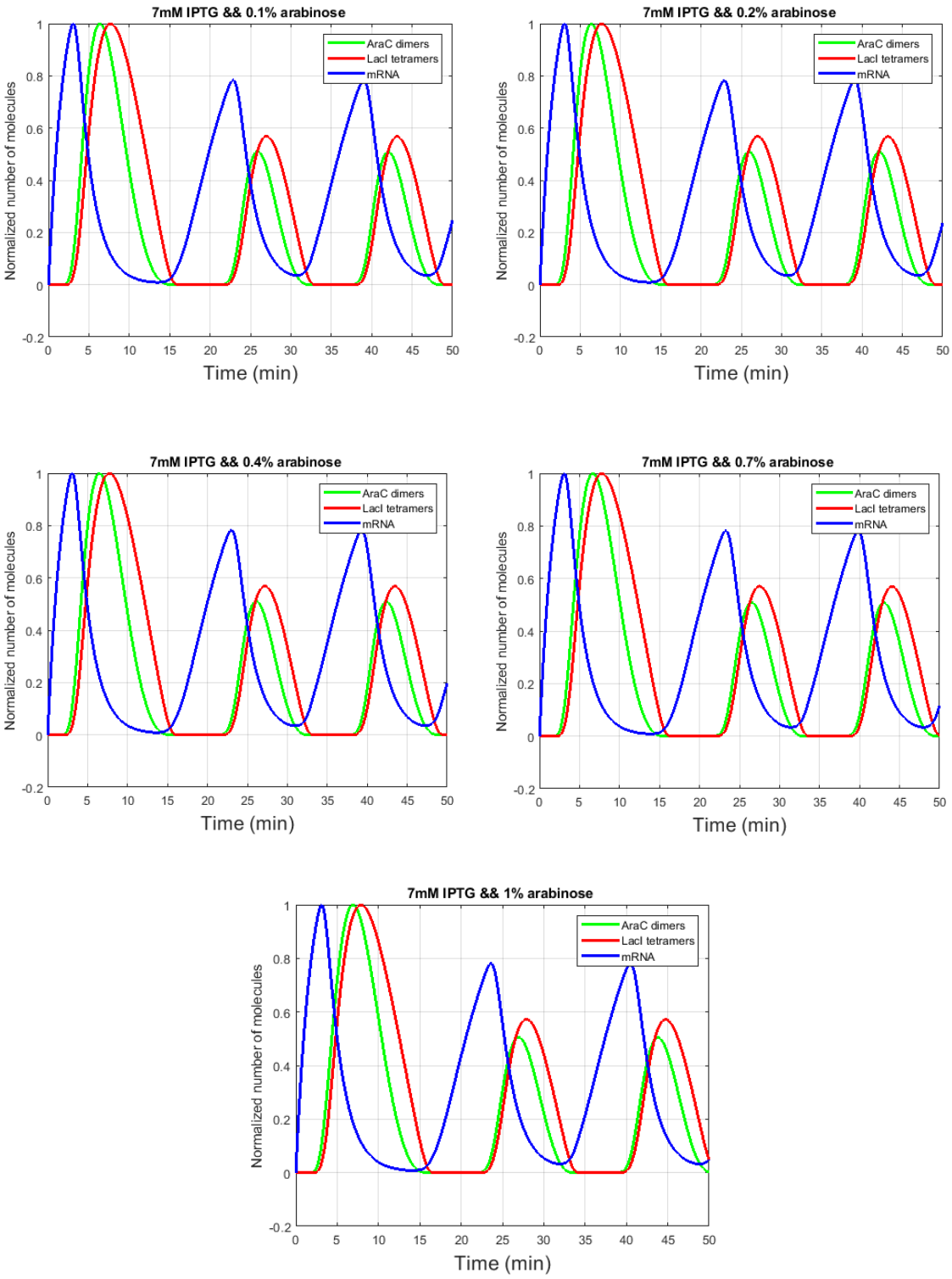

Fig. 9.5 Time dynamics of AraC dimers, LacI tetramers and mRNA at fixed 7 mM of IPTG and at varying arabinose concentrations.

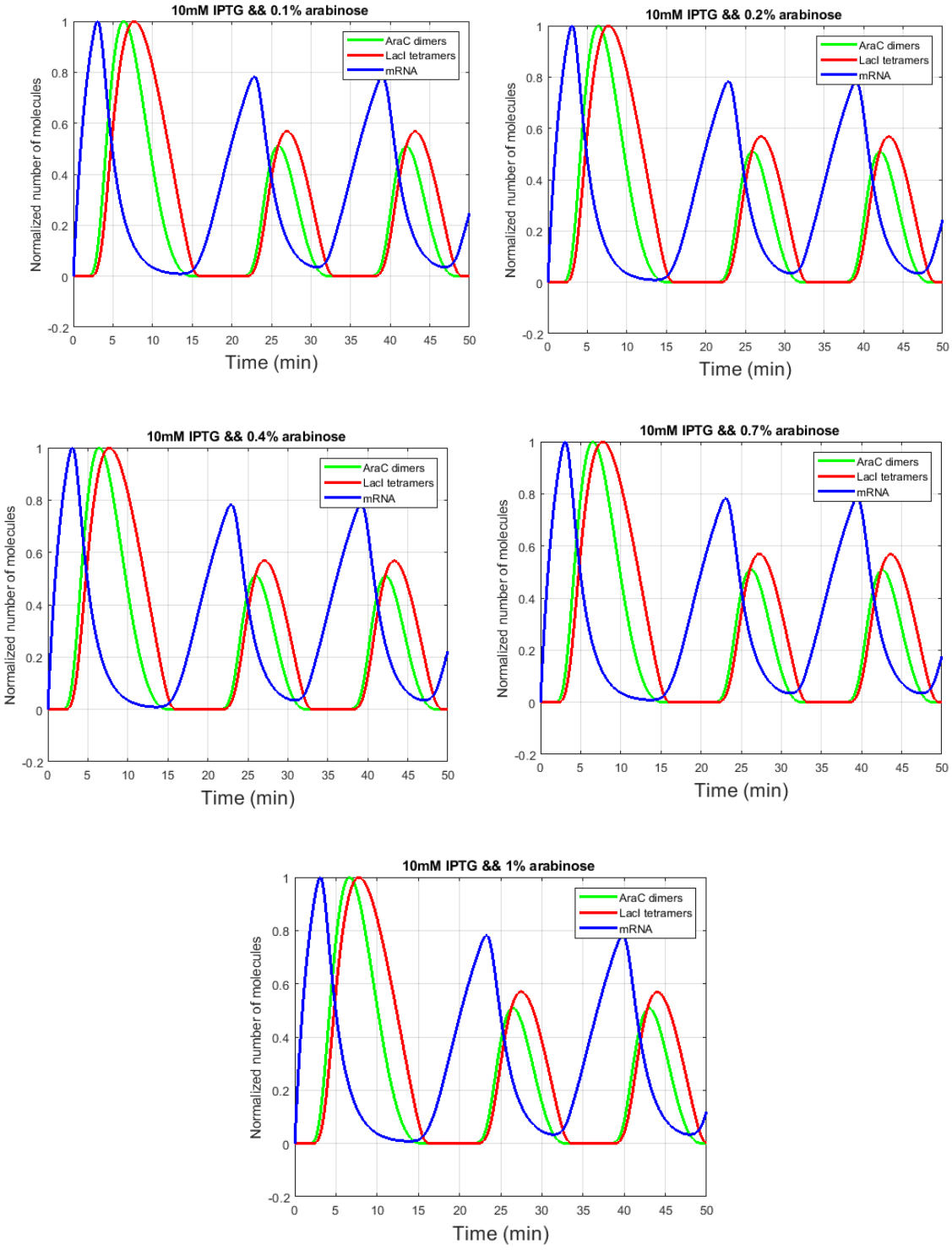

Fig. 9.6 Time dynamics of AraC dimers, LacI tetramers and mRNA at fixed 10 mM of IPTG and at varying arabinose concentrations.