MATHEMATICAL MODELING OF NITROGEN REGULATED BIOLOGICAL SYSTEMS

BY

Pau Casanova Ferrer

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

MATHEMATICAL ENGINEERING

UNIVERSIDAD CARLOS III DE MADRID

ADVISORS:

Saúl Ares García

Javier Muñoz García

TUTOR:

Javier Muñoz García

January 2023

Pau Casanova Ferrer: *Mathematical modeling of nitrogen regulated biological systems*, A systems biology study.
pau.casanova25@gmail.com

This doctoral thesis was supported through the FPI contract BES-2017-079755 associated to the grant FIS2016-78313-P from Ministerio de Ciencia e Innovación (MCIN)/Agencia Estatal de Investigación (AEI)/ 10.13039/501100011033/ and FEDER Una manera de hacer Europa. Additionally the author also acknowledge finantial support from MCIN/AEI/10.13039/501100011033 through grant BADS, no. PID2019-109320GB-100.

Madrid, January 2023

Science is made up of so many things
that appear obvious after they are explained.

Pardot Kynes (10121 AG-10175 AG)
in Frank Herbert's *Dune*

The thing the ecologically illiterate don't realize
about an ecosystem is that it's a system. **A system!**
A system maintains a certain fluid stability that
can be destroyed by a misstep in just one niche.
A system has order, a flowing from point to point.
If something dams the flow, order collapses.
The untrained miss the collapse until too late.
That's why the highest function of ecology
is the understanding of consequences.

Pardot Kynes (10121 AG-10175 AG)
in Frank Herbert's *Dune*

ACKNOWLEDGEMENTS

As it could not be otherwise, I would like to begin this thesis with a disclaimer and an apology. Everybody that knows me, knows that I am terrible at this kind of things. I promise that I am working on it, and I like to think that I naturally convey that I would not have reach this point without any of you. Then, I hope that you all can forgive me for writing this section in my language instead of yours.

Se m'ha fet molt llarg això, és més, encara no m'acabo de creure que ja s'acabi. Però al mateix temps, tinc la sensació el temps m'ha passat volant. Aquesta paradoxa sols s'explica gràcies a vosaltres. Han sigut anys de molt d'estrès i incertesa, però també m'ho he passat molt bé i al final això és el que queda.

No puc començar per un altre lloc que agrair a **Saúl** i **Javier** la seva confiança. Sense ells no estaria escrivint aquesta secció ni tampoc coneixeria a gran part de la gent que ve a continuació. Gràcies per introduir-me al món acadèmic en el meravellós microclima que és el GISC. Especialment, gràcies a Saúl per sempre ficar a disposició la seva xarxa de contactes, sense la teva ràpida intervenció en el canvi per COVID de l'estança doctoral no tindria gens clara la meva pròxima etapa acadèmica.

Una vegada dit això espero que em permeteu girar el protocol i agrair primer als meravellosos companys de despatx que he tingut que són els que tenen més mèrit. He de començar amb **Alberto**, que és clarament l'italiá que més gasta en vols que conec... Encara no entenc com has aguantat la invasió de despatx d'un tio que no pot callar més de 5 minuts seguits, però t'agreixo eternament que ho hagis fet. Ets un crack dins i fora de la universitat. Pablo Lozano, aka **Palozano**[1], mira si tenia clar que volia compartir despatx amb tú que et vaig anar a buscar dos dies després conèixer-te. Et dec, entre moltes altres coses, casi tot el que sé de Linux i Python, gràcies per provar que el que diuen els físics dels enginyers no és sempre cert. I **Jose** MC Mateu vas aconseguir una cosa molt difícil que és que no hi hagués quasi dol al despatx per Palozano. Ets una persona meravellosa amb un humor compulsiu, com no m'has de caure bé Makina, gràcies pels caps de setmana.

[1] *Ho sento, que ja és lleig que sortint abans tinguis tú el apodo pero així vius al meu cap...*

A les dues generacions de la joventut del tupper gràcies per la immediata integració al departament. **Vicky**, **Francy**, **Ignacio**, **Esti**, **Diego** i **Maria** vau fer ràpidament que el viatge diari a Leganés fos tot un plaer. I

en especial a **Pablo** i **Pilar** que a més els vaig donar pràctiques de maternitat avançada. I de la segona fornada, el tàndem de despatx **Juan** (que té la paciència d'un sant) i **Jorge** que junt amb **Bea** (grácies pels concerts), **Miguel** i **Nello** va emplenar de sobra el forat que va deixar el Baby Boom matemàtic.

Una vegada omplert ja l'estómac es pot parlar de ciència. Així que vull agrair a la gent del departament, GISC o no, que sempre ha tingut la porta oberta per a mi, en especial a **Jose** i **Anxo** per amenitzar els matins amb cafè. I de fora de la universitat cal recalcar a **Jacobo**, **Susanna**, **Luiño**, **Iker** i la resta del col·lectiu de les *Evomisas*. Per mostrar-me que no tot és competència en la ciència i també, perquè no, per donar-me una excusa per a menjar bé un dia cada dues setmanes després d'escoltar bona ciència. Sou tots un referent per a mi.

Seguint el viatge cap a fora toca agrair a la barreja de dos grups, de **Pau Formosa** i **George Coupland**, que em va acollir meravellosament durant la meva estada per Colònia. Gràcies per fer-me sentir com a casa, **Martina**, **Gauthier** i **Enric**, ens veiem ben aviat.

I per acabar queda agrair a la família i amics, que són els que s'han menjat el pitjor dels mals moments i als que donaré menys aquí. Realment no tinc paraules d'agraïment que valguin per a vosaltres, ja ho celebrarem i tranquils que no us jutjaré si ja no llegiu res més.

Segur que em deixo algú, així que si estàs llegint això i no t'he donat les gràcies dis-m'ho i ho discutim amb una birra.

Moltíssimes gràcies a tots,

Pau

PUBLISHED AND SUBMITTED CONTENT

PUBLICATIONS IN THIS THESIS

The main results from this thesis are based on the following articles:

Casanova-Ferrer, Pau et al. (Sept. 2022a). "Mathematical Models of Nitrogen-Fixing Cell Patterns in Filamentous Cyanobacteria." In: *Front. Cell Dev. Biol.* 10, p. 959468. ISSN: 2296-634X. DOI: 10.3389/fcell.2022.959468.

Casanova-Ferrer, Pau et al. (Aug. 2022b). "Terminal Heterocyst Differentiation in the Anabaena patA Mutant as a Result of Post-Transcriptional Modifications and Molecular Leakage." In: *PLoS Comput Biol* 18.8. Ed. by Tobias Bollenbach, e1010359. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010359.

The material from these sources included in this thesis is not singled out with typographic means and references. The different chapters are related to the former publications in the following way:

- The revision of the previous models presented in section 2.2 has been already been published in Casanova-Ferrer et al., 2022a.
- The chapter 4 reproduces the study published in Casanova-Ferrer et al., 2022b.

Two more papers are in preparation, each corresponding to the results presented in Chapter 5 and Chapter 6.

OTHER WORKS

Besides the articles that compound this thesis, the author of this thesis has published several articles as a result of collaborations unrelated to the topic of this thesis:

Bernini, Alba et al. (2019). "Evaluating the Impact of PrEP on HIV and Gonorrhea on a Networked Population of Female Sex Workers." In: DOI: 10.48550/ARXIV.1906.09085.

Hernández-del-Valle, Miguel et al. (Dec. 2022). "A Coarse-Grained Approach to Model the Dynamics of the Actomyosin Cortex." In: *BMC Biol* 20.1, p. 90. ISSN: 1741-7007. DOI: 10.1186/s12915-022-01279-2.

Steinegger, Benjamin et al. (May 2022). "Non-Selective Distribution of Infectious Disease Prevention May Outperform Risk-Based Targeting." In: *Nat Commun* 13.1, p. 3028. ISSN: 2041-1723. DOI: 10.1038/s41467-022-30639-3.

Nitrogen is required to form both the amino acids that constitute proteins and the nucleic acids that store the genetic information in the DNA and RNA. This makes it an essential element for life, given that all organisms are constituted by proteins and store their genetic information in either DNA or RNA. The nitrogen cycle, represented in fig. 0.1, is extremely relevant because it controls the amount of organic nitrogen in a given ecosystem. This is because, despite being the largest atmospheric component (almost 78% of the atmosphere is nitrogen), it is mostly present in the form of N_2.

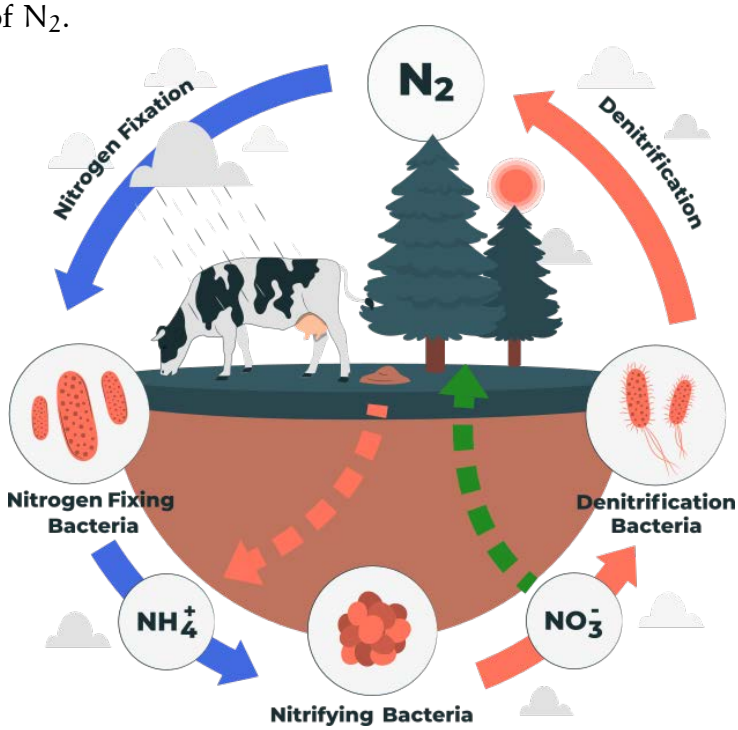

Figure 0.1: **Biological Nitrogen cycle:** This cartoon shows a simplified nitrogen cycle with a particular color for the two processes studied during this thesis. In blue, we mark the nitrogen fixation and in green the nitrogen incorporation into plants.

This chemical form presents a triple covalent bond between the two nitrogen ($N \equiv N$) that requires a lot of energy to break. This makes the compound highly nonreactive and difficult to decompose to any other molecules, and easier to assimilate for all organic forms. This process is called nitrogen fixation and naturally only lightning and a few bacteria and archaebacteria and are capable of it, lightning being by far the less relevant (less than 3% of the total amount) (Fowler et al., 2013). Due to

this, organic nitrogen is usually the major limiting factor of the amount of biomass that a given ecosystem can sustain.

At the beginning of the 20th century, with the discovery of an industrial process to fixate ambient nitrogen into ammonia (the Haber-Bosch process) the use of nitrogen fertilizers skyrocketed. This increase of the global nitrogen fixation capabilities allowed for roughly doubling of the one observed without human intervention (Fowler et al., 2013). Until this discovery, all the processes that required reactive nitrogen were dependent on natural reservoirs of guano (accumulated excrement of seabirds and bats) which greatly impeded its generalization. On the other hand, it has been recently estimated (Erisman et al., 2008) that nitrogen fertilizers are responsible for feeding almost half of the human population (fig. 0.2). This new ammonia source, together with the Green Revolution, allowed for the demographic explosion presented in fig. 0.2.

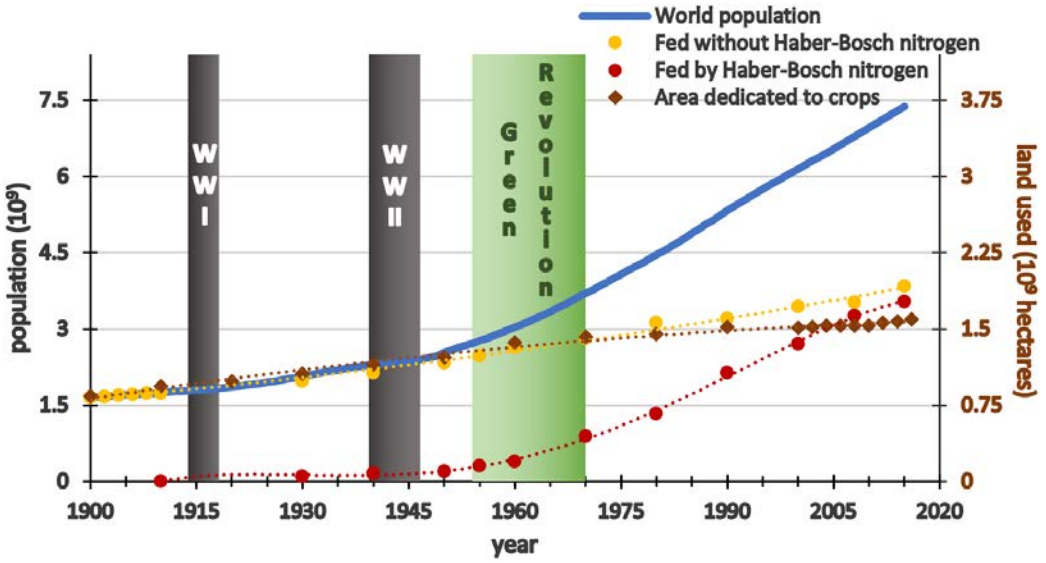

Figure 0.2: **Chronology of the world population dependency with both the Haber-Bosch nitrogen and the land used in agriculture.** The world population has been obtained from Gapminder for the years 1900-1949, and from the UN Population Division from 1950 onwards. The estimates of population sustained with and without the production of Haber-Bosch nitrogen are derived from (Erisman et al., 2008). And finally, the estimate of the cropland area from the History database of the Global Environment (HYDE). Together with these data we represent three time periods relevant to the demography evolution: the two World Wars and the Green Revolution.

One can observe in this same fig. 0.2 that the widespread incorporation of the Haber-Bosch nitrogen in agriculture occurred after the end of World War II, and it coincides with the change of tendency of the population growth. During both war periods and the Interwar period, most of

the ammonia produced was used to fabricate explosives. It is also worth noting that we adjusted the land use axis to evidence the obvious correlation between the evolution of the amount of population fed without the use of "artificial" fertilizer and the area dedicated to crops. Nevertheless, the two magnitudes correlation breaks after the Green Revolution, when the area dedicated to crops remained constant while the efficiency of those crops augmented. With roughly the same crop extension as in 1970 nowadays we can feed a third more of the population fed in 1970.

As already mentioned, this is achieved thanks to the so-called Green Revolution, which is a global agricultural revolution that focused mainly on the intensification of production and the selection of crops to increase production. The modernization of agriculture can be divided into two complementary approaches. On one hand, there was extensive genetic breeding of all the main carbohydrate sources of our diet: wheat, rice, maize, potatoes, and cassava. This selected the most efficient varieties, those that maximize the grain/biomass ratio. It was discovered that dwarf varieties dedicate most of their energy to grain, and due to this they also respond better to fertilization (Evenson and Gollin, 2003). Additionally, the smaller size and sturdier stalks of the plants also help to reduce the lodging (the knockdown of a plant, unearthing the roots) under high wind (Hedden, 2003). This variety search was done by selecting and crossing the plants with the desired characters in a sort of Mendelian genetics. Given that it was not until much later, already in the 90s, that the genetic pathways responsible for these optimal traits were identified (Hedden, 2003). On the other hand, the modernization of agriculture was also due to the extensive use of mechanization, irrigation facilities, and inorganic fertilizers which allowed for the intensification of croplands that were not capable to sustain such high production rates before (Khush, 1999).

While this intensification of agriculture allowed to triplicate the population is not sustainable due to soil erosion and both the depletion and contamination of aquifers (Pingali, 2012). Then a new approach is necessary to make agriculture more sustainable because, as one can observe in fig. 0.2, the increase in production is heavily dependent on nitrogen fertilizers. A deeper understanding of the genetic mechanisms underlying both natural nitrogen fixation in bacteria and nitrogen incorporation in plants could allow for targeted genetic modifications. Recent experimental works, such as (S. Wei et al., 2022; K. Wu et al., 2020), show that targeted genetic modifications could further improve crop efficiency, reducing the need for nitrate fertilization.

Following this point of view, during my thesis, I studied two particular biological systems to provide a mathematical model for two particular

cases of both the nitrogen fixation in bacteria and nitrogen incorporation in plants. We structured the thesis in four parts:

- The part I presents a brief recapitulation of the common theoretical and mathematical framework used in parts II and III which constitute the main body of work of this thesis.
- Part II contains the main results obtained for the analysis of cyanobacteria nitrogen fixation. This part starts with a brief introduction to the biological system in chapter 2. In this introduction, we recapitulate and organize the current state of the art in both the experimental (section 2.1) and modeling (section 2.2) approaches to this system.

  After this introduction, we discuss the stability of the previous minimal model presented in (Muñoz-García and Ares, 2016) in chapter 3. Through this study, we provide an insight over the adaptive advantages that a bistable regime provides over a fluctuating environment. The particular mechanistic deduction of the model is not included in this chapter, and it is presented in appendix A.

  Then, in chapter 4, we investigate the function of *patA* and *hetF* together with their interactions with the main genes responsible for heterocyst pattern formation. The proposed mathematical model reproduces the diverse experimental phenotypes and explains the main function of both *patA* and *hetF* in the gene-regulatory network of heterocyst differentiation. As we did in the previous chapter, the full model deduction is presented in appendix B.

  We close the cyanobacteria part of this thesis with an intercellular correlation study in chapter 5. In this chapter, we present the correlations observed for the model presented in chapter 4 while comparing it with the available experimental data. The obtained intercellular correlation profiles are used to present a new two stage selection mechanism for heterocyst differentiation that reproduces the observed spatial self organization in *Anabaena* filaments.
- Part III contains the main results obtained for the analysis of nitrogen incorporation in rice plants. During this study, we present a model of lateral growth or tillering that works in parallel to the vertical growth regulation through the well-known DELLA proteins. This model provides new insight regarding the maintenance of nitrogen induced yield increases in dwarf rice varieties while reproducing the available experimental observations.
- Finally, part IV recapitulates the main results of the thesis and proposes future research lines to apply the basic biological knowledge in the crop selection and/or design.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND NOMENCLATURE

DNA Deoxyribonucleic acid

RNA Ribonucleic acid

mRNA Messenger RNA

WT Wild type

Δ*genX* Deletion mutant of the gene *genX*

GenX Protein of the gene *genX*

Mch phenotype Multiple continuous heterocysts phenotype

aND after nitrogen deprivation

Part I

SOME KIND OF MANUAL

Here we will present a brief general introduction of both the genetic regulation in biological systems and the mathematical framework used in the rest of the thesis.

1

THE SYSTEMS BIOLOGY PERSPECTIVE

The point of view of complex systems has already a certain antiquity in physics (mainly within the field of non-linear physics). It was already perfectly summarized by Aristotle in the sentence: "The whole is greater than the sum of its parts". In a collective system formed by diverse elements, knowing the behavior of each element separately often does not allow you to deduce the behavior of the entire system. This is due to the so-called emerging properties that appear due to the interactions of individual elements, either with each other or with the environment. But its application to external fields such as sociology, psychology, or biology and therefore its establishment as a scientific field of its own is relatively recent. Its development responds mainly to the technological advancements in computational power, which allow for both the processing of much larger amounts of data and the numerical resolution of much bigger mathematical models. Additionally, the discovery of quantitative high-throughput biological tools together with the creation of globally accessible databases allowed for the systematic research required to develop all the -omics sciences (Ideker et al., 2001). These disciplines (genomics, transcriptomics, proteomics, and metabolomics) provide the huge amount of data necessary to understand a given system as a whole, which is the main objective of the Systems Biology (Kitano, 2002).

The rising of these properties denies the lineal superposition system that ruled in physics at the ages of Newtonian mechanics and electromagnetism. Thus, it is often more effective to develop a simple model that captures these interactions in broad terms. Not only because this is usually easier than knowing in depth the behavior of each of the elements of the system, but also because many times these minimalist models have a much better capacity for predicting dynamics than more extensive models that do not consider these emerging properties. This is because our knowledge of the system is never absolute. Then, it is impossible to know all the interactions between elements and the addition of more elements always makes it easier to artificially "tune" the system to behave as desired. This would make it harder to pinpoint the more relevant interactions. In the case of biological systems, it is impossible to consider every genetic pathway involved tangentially is a given process. Additionally, is it worth mentioning that the typical separation of the molecular regulation in pathways is already a reduction, given that there is only one common pool of

metabolites? This produces a relevant overlap and/or coupling between different regulatory pathways given that the effects that one has over, for example, the ADP/ATP or NAD^+/NADH ratio will affect and condition the other pathways. Then, in order to have both a translatable and faithful model, one should compromise these two perspectives. A perfect model should be representative of the full system, while keeping it as simple as possible to avoid over-fitting and arbitrariness. This is usually accomplished through coarse-graining, which allows for the reduction of the scale of the model through the use of the existent modularities and hierarchies in the system (Ideker et al., 2001; Haiyuan Yu and Mark Gerstein, 2006).

This strategy originated in the molecular dynamics simulations research field (Levitt and Warshel, 1975). There, atomistic models are used to reproduce the behavior of the system of interest, usually a protein, while dissolved in water or embedded in a lipid bilayer or even the interaction between several proteins. These simulations are very demanding computationally, as they require following the trajectories of $10^5 - 10^7$ atoms during $10^{-6} - 10^{-8}$ seconds (Saunders and Voth, 2013). The coarse-graining allows the clustering of a group of neighboring atoms into a new artificial atom that will be used in the simulation. Analogously, through coarse-graining, one can simplify well-known clustered sections of the regulatory network not relevant to our research in such a way that the known external interactions of these sections are maintained but all the internal interactions are ignored. The main example of this proceeding is to consider several genes that mostly interact between them and that react similarly to the studied external conditions as an artificial gene that incorporates all the interactions of those genes with the rest of the network.

This perspective is called holism and complements very well the reductionist approach, which focuses on full detailed descriptions of certain genes or proteins. Traditionally, biologists have taken this reductionist o mechanistic approach and considered that the whole system could be understood through the study of its actors. On the other hand, holistic approaches consider the full regulatory pathway as a whole through the simplification of the already well-known processes. Then it is easy to see how well these two perspectives complement each other fig. 1.1. Detail-oriented and systematic experimental research provides the required information for drafting big-picture models. While, the models constitute a cheaper and faster method to provide predictions of the behavior of the system in new experimental conditions that could be later confirmed through new experiments.

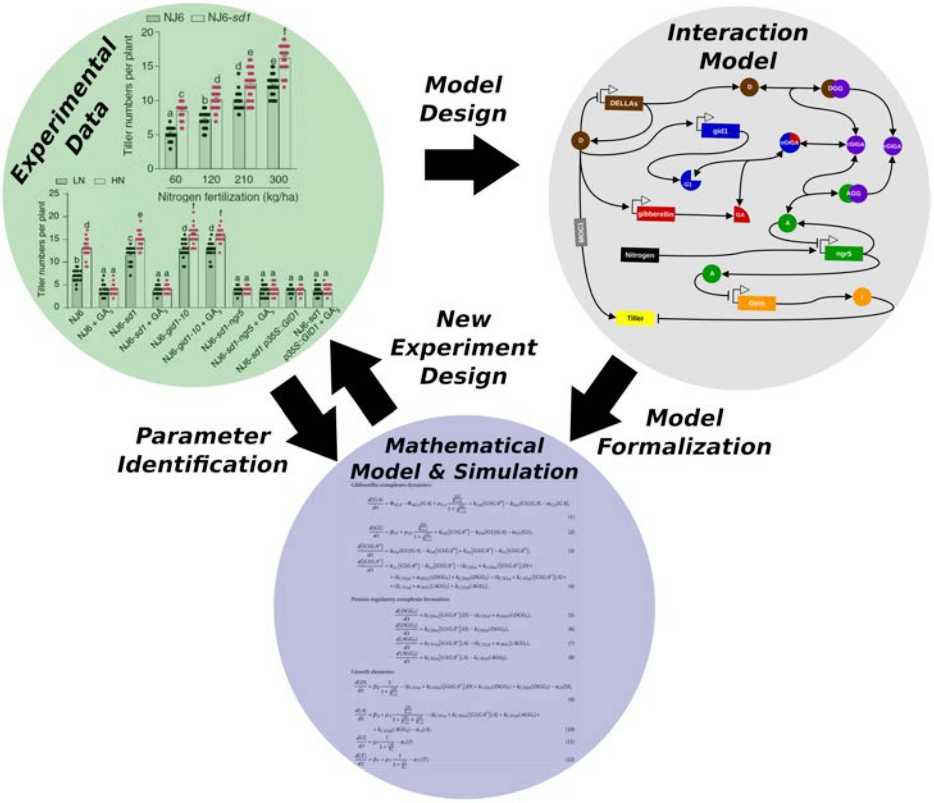

Figure 1.1: **Modeling Loop:** Diagram showing the most basic flowchart for the creation of a biological model. The particular examples on the background of each sphere correspond to the model studied in part III.

There are two main approaches to this interplay between modeling and experiments. The first one, denominated Top-down system's biology, uses system-level behavior information together with the existent partial information of the individual elements of the system to infer new interactions or elements that could explain the observed performance. On the other hand, Bottom-up systems biology uses detailed information on all the known parts to predict the reaction of the full system in new external conditions. While these two approaches are not that different in the system's biology context, where the main objective is the analysis and comprehension of a given system, they start to differ much more when one considers the engineering application of systems biology and synthetic biology. This discipline tries to artificially design biological systems in order to optimize the desired function. In this context, the Top-down approach would be the natural step after the classical genetic modification to produce transgenic with a much bigger scope of modifications in order to obtain new emergent behaviors. While on the other hand, the Bottom-up approach would completely build a new system with known functional parts, creating completely artificial pathways or even organ-

isms capable to mimic natural biological behaviors. The main example of the latter approach, which is also widely considered the starting point of the field, is the repressilator presented in (Elowitz and Leibler, 2000). This artificial oscillator uses a regulatory network of three transcriptional repressor systems that are not part of any natural biological clock to reproduce the behavior of a tunable biological clock.

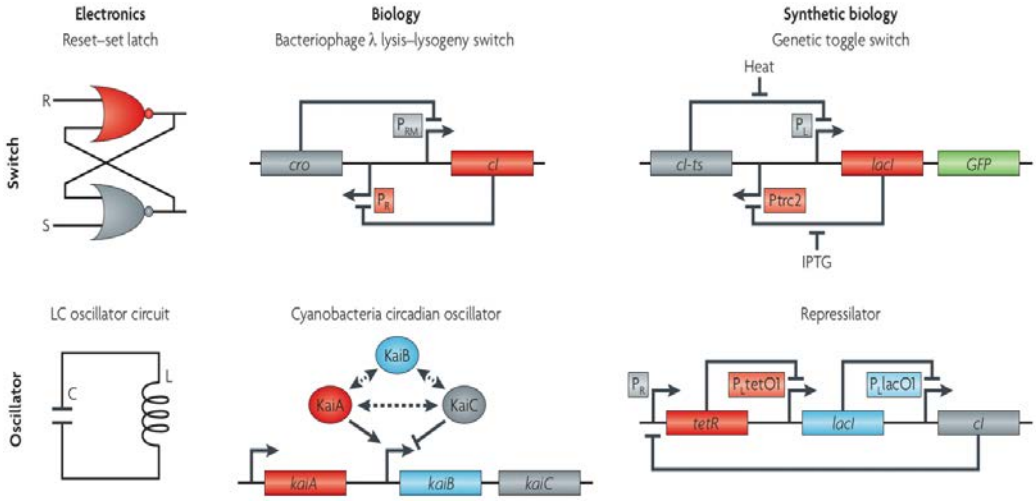

Figure 1.2: **Early synthetic biology designs: switches and oscillators.** This diagram shows an electronic configuration, a natural biological example, and the first synthetic example for both a switch and an oscillator. Reproduced from (Khalil and Collins, 2010).

These minimal network motives are capable to perform a certain general function such as a robust oscillating expression (repressilator) or a controlled deactivation of the gene expression (toggle-switch) in fig. 1.2. Due to this, they could constitute the building blocks for the design or modification of genetic pathways to optimize a certain function in an existent organism. And additionally, given their generality, they could help to predict the function of certain regulatory networks just by their structure.

As a testament to this generality, we will use this systems biology perspective to study a couple of genetic regulatory networks related to nitrogen fixation and intake in cyanobacteria and rice plants, respectively. We will see that, despite the differences in both their function and their organism, the behavior of these two networks can be described with the same mathematical framework. The number of acting elements and their name will be different for each network, but the roles that they perform are limited and common.

In the following section, we will present the theoretical background required for the formulation of mathematical models of any genetic regulatory network.

1.1 GENES, PROTEINS, AND THEIR MODELS

The minimal functional element of a multicellular organism is, as the name implies, the cell. At the cellular level, any possible cellular role is defined by the concentration of proteins, and therefore, it is ultimately codified in the full genetic information of the organism. Then each cell's fate, which defines its functionality, is completely dependent on a particular balance of the expression of the thousands of genes expressed in the cell. The expression of genes is heavily interconnected in such a way that genes tend to be activated in groups, reducing the activity of other genes. The network of interactions between genes is a regulatory system that integrates both intracellular and extracellular signals to control the expression of all the genes. But before presenting the regulatory networks analyzed during this thesis, in this chapter, we will present the theoretical framework used to model them mathematically. We will use both the repressilator and the toggle switch motifs mentioned earlier, rather than general examples of gene regulation modeling.

In order to ensure a common background, we will first present a brief biological description of the mechanisms to regulate gene expression. These mechanisms are used to control the concentration of proteins at the cellular level.

There are two types of molecules capable of storing genetic information, DNA and RNA. Composition differences make the DNA much more stable than the RNA, and this affects the biological function of these two molecules Alberts, 2015. Almost all organisms use DNA as the long-term storage of biological information, the only exception are RNA viruses that use RNA. Alternatively, RNA is mainly used as an intermediary step for protein formation because it is required to build the proteins. Then, the genetic information codified in the DNA is first reproduced into messenger RNA (mRNA) in a process called transcription to build proteins by translating the mRNA in the ribosomes. This process can be observed in fig. 1.3 together with the main difference between prokaryote and eukaryote cells.

Prokaryote cells are much simpler than eukaryotes, given that they have neither a nucleus nor any other membranous organelles (which are functional subunits of a eukaryote cell). While the eukaryotes present different internal compartments, the prokaryotes are a single open space with both the DNA and the ribosomes freely suspended in the cytoplasm.

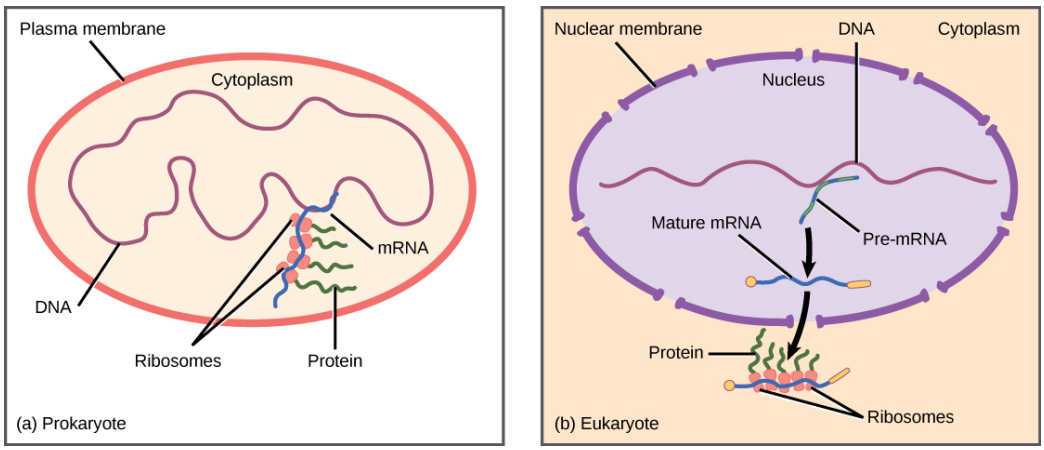

Figure 1.3: **An overview of the flow of information from DNA to protein in both prokaryote and eukaryote cells.** Image provided by OpenStax CNX. License: CC BY: Attribution.

This makes protein production much more complex in the eukaryote cells because the mRNA translated inside the nucleus needs to be exported to the cytoplasm to be translated into the ribosomes. Gene expression regulation can be observed in all the different stages of the process: at both the transcriptional and the translational level modifying the efficiency of each process (by methods described later on) and, at the post-transcriptional and post-translational, with modifications to either the mRNA and the proteins that modify its stability and functional life. While the transcription and translation are almost simultaneous in the prokaryote cells, this transport opens a wider window for post-transcriptional regulation in the eukaryotes and even in some cases requires an additional maturation of the mRNA. Due to this, the regulation of gene expression in prokaryotes is mainly produced either at the transcriptional level or the post-translational level (with targeted protein degradation or protection).

Gene transcription is done by the RNA polymerases, which are large enzymes that build the mRNA chain following the sequence codified in the DNA strand. The enormous relevance of these proteins is shown in both their huge fidelity (on average it makes an error once every 10^{10} base pairs) and the fact that it is one of the most conserved complexes in biology (A. J. F. Griffiths et al., 2020). All the actual RNA polymerases in both prokaryotic and eukaryotes conserve the same functional core. Then, in order for the transcription of a certain gene to be initiated, the RNA polymerase must attach to its promoter, which is a particular sequence that marks its beginning. Given that both the polymerase and the starting are common to all the genes transcribed, it does not favor the expression of any gene by itself. The regulation of the transcription rate is done by other proteins, called transcription factors (A. J. F. Griffiths et al., 2020).

These proteins are usually generalized into two classes. On one side, inhibitors impede the transcription disturbing the polymerase, and, on the other, activators improve its efficiency.

The most straightforward way to negatively regulate a gene is the one presented in fig. 1.4. This scheme shows a typical situation where an inhibitor can attach in a region between the promoter and the proper gene sequence (called an operator in the prokaryote genomes).

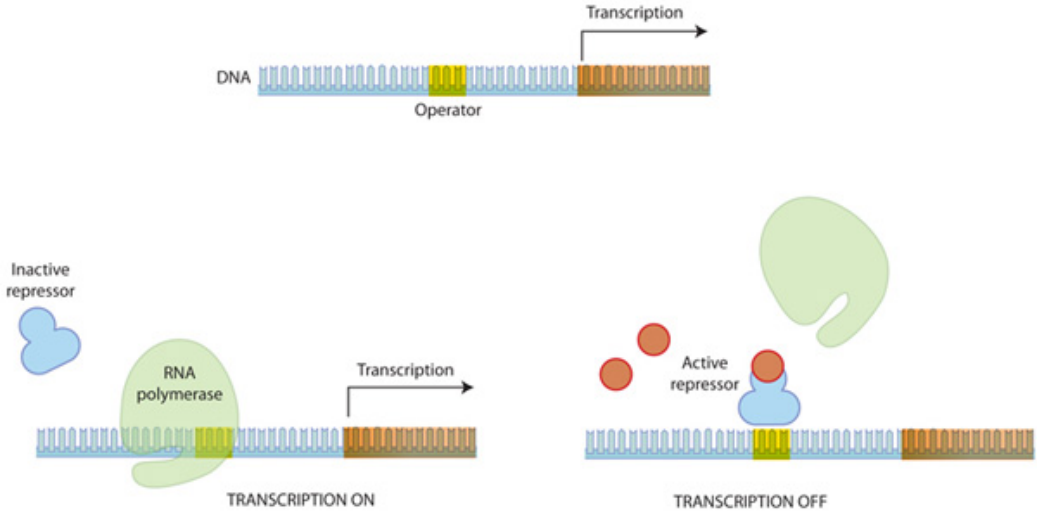

Figure 1.4: **Inhibition of transcription**. Image reproduced from Gene Expression. © 2010 Nature Education

This binding physically blocks the way of the polymerase that cannot initiate the transcription. There are also cases when the inhibitor links two separated regions of the DNA strand looping it on itself and also impeding the transcription of the looped section, or even proteins that impede the opening of the DNA helix (A. J. F. Griffiths et al., 2020). Additional factors such as the number of gene copies and stability of the inhibitor attachment to the DNA mark the intense effect on the transcription of the gene because the polymerase will be blocked for longer times. This allows for a continuous range of inhibition, instead of just a toggle effect with only the on and off options. This can be reinforced additionally with proteins that affect the base polymerase efficiency by altering its conformation.

Similarly, the activators can act directly with analogous strategies but with the opposite effect on gene transcription. Then an activator could link to a preceding region of the promoter in such a way that increases the affinity of the polymerase to the gene promoter. Alternatively, an activator could facilitate the transcription process either by anchoring the polymerase into the desired gene section, as observed in fig. 1.5, or by

binding to the promoter creating a loop that stabilizes the exposure of the transcribed gene (A. J. F. Griffiths et al., 2020).

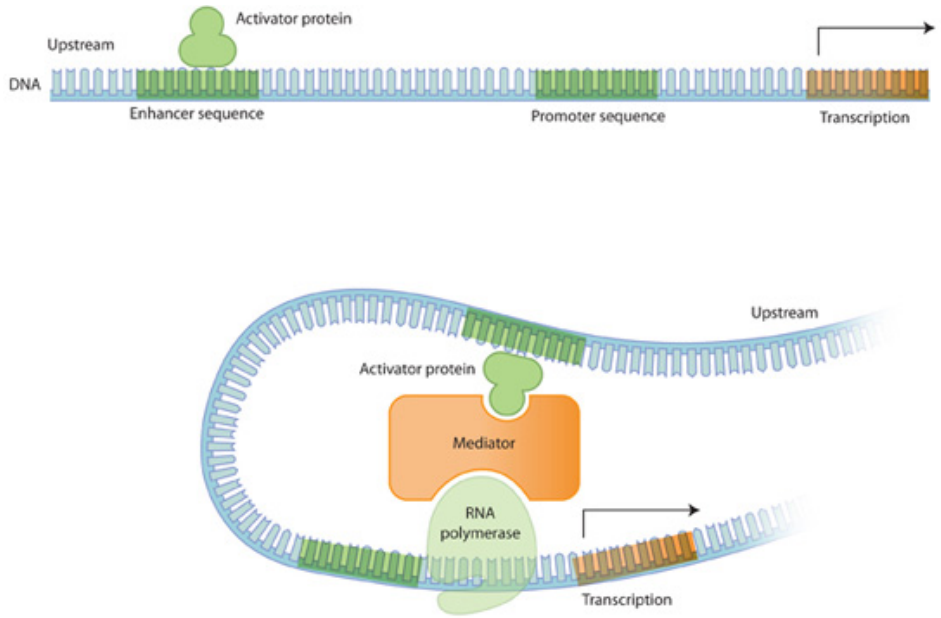

Figure 1.5: **Example of an enhancement of transcription.** Image reproduced from Gene Expression. © 2010 Nature Education

It is also important to also consider that all inhibitors and activators of a given gene usually can either interact or compete with each other. Additionally, they are also affected by other elements whose production is dependent on external signals. This, which is fairly common, further modulates their effect on gene expression and allows the cell to react to extracellular changes. Finally, given that these transcription factors are also proteins, they are also subjected to gene regulation and this creates densely interconnected regulatory networks with abundant close circuits (also denominated feed-back).

It is easy to see that once you start considering the dynamics of a system with a few interconnected genes, the sheer number of cross-interactions makes it increasingly hard to make intuitive predictions. These systems typically present non-linear responses to environmental changes with emergent behaviors that are too complex to allow for an intuitive explanation. If enough quantitative data about the system is known, one can translate the interaction network into a mechanistic model. This mathematical model describes the temporal dynamics of all considered variables through a set of differential equations. This set of differential equations can be then numerically integrated to computationally simulate the system's experimental behavior and should respond to parameter changes analogously to the real system (Ingalls, 2013). Alternatively, one

can solve it to obtain the equilibrium states of the system (if they exist) without having to obtain all the temporal evolution. Nevertheless, this discussion about what to do once you have the model formulated belongs to the next section 1.2. Here, we will present the methodology to translate the interaction network into a mathematical model consisting of a set of differential equations.

In order to study gene expression and protein interaction, we can consider all the species as molecular components and use the well-known mathematical method to model chemical reactions. As an example of this method, we will now consider the system represented in fig. 1.6 which consists of two molecular species.

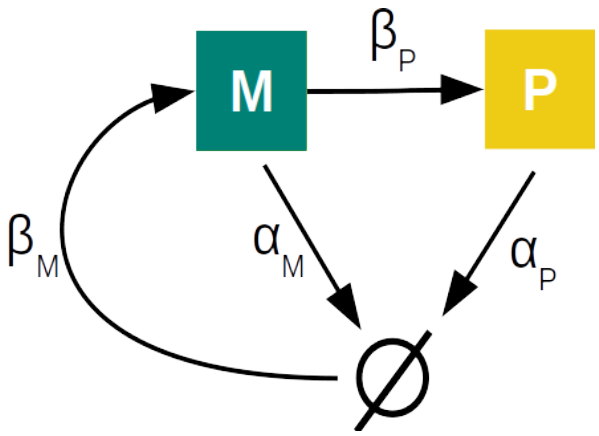

Figure 1.6: **Diagram of a simple open-ended reaction system.**

In this system, the species M is created at a constant rate β_M and degraded at a linear rate α_m, and the species P is created at a rate β_P from m without consuming it and degraded at a linear rate α_p. In classical chemical notation those reactions would be represented as:

$$\varnothing \xrightarrow[\beta_M]{} m, \qquad m \xrightarrow[\alpha_M]{} \varnothing, \tag{1.1a}$$

$$m \xrightarrow[\beta_P]{} m + p, \qquad p \xrightarrow[\alpha_P]{} \varnothing, \tag{1.1b}$$

where the left side elements are the ones producing the reaction, called reactants, and the right ones are the products of the reaction and the ∅ symbol represents the absence of either one. It is easy to see that the gross amount of M is only affected by the reactions in (1.1a) and similarly, the quantity of P is only altered by the reactions in (1.1b).

Up until this point, we have not properly defined either m, p or all the rates. We just know that they are quantities of something that is created and destroyed following (1.1). They could as easily be a really simple model of **m**onkeys and **p**arasites or one of the gene expression of an unregulated gene which is first transcribed into **m**RNA and later translated

into **p**roteins. Given that the genes are neither created nor destroyed during gene expression, there is no need to model them, as they would always remain constant.

To be able to translate this system of reactions into a set of differential equations, we will have to set some restrictions. The law of mass action[1] states that if one assumes homogeneity and continuity in the system, the rate of the reactions is proportional to the concentration of the reactants. These two conditions impose the necessity to consider a fixed volume of our system and a homogeneous distribution of both M and P to make the concentrations m and p good descriptors. As an example of what this means, we will consider again the two descriptions proposed earlier. On one hand, we would have to consider the population density of both monkeys and parasites to have a continuous variable. Additionally, we should only include a certain type of environment where monkeys meet regularly to avoid heterogeneity. On the other hand, we will use molar concentrations for both the mRNA and the proteins and the expression of both should be high enough to ensure a well-mixed system. If this is not true and the expression of them is either too dispersed or too clustered, the comparison of real contacts and the model prediction will be off. If the probability of the reactants meeting in the real system is slightly higher or lower than the one supposed in a homogeneous system, the model dynamics will, respectively, underestimate and overestimate the reaction rate. Before continuing with the modeling, we should address the elephant in the room. Why bother with all of this if we need to use a principle that demands constant volume, and we all know that cells grow? To make it even worse, cells grow quite a lot. As an example, *Anabaena* cell volume changes almost two orders of magnitude during its life (de Tezanos Pinto et al., 2016; Milo and R. Phillips, 2016). The trick here is to realize that all these conditions are not absolute, they are only necessary for the timescale of the reactions. Cellular growth occurs, usually, in a much larger timescale than the one of both mRNA and protein production. As we will soon see, this separation of scales is one of the most important tricks for biological modeling. The main takeaway is that, if you are considering something on a faster scale, everything on a slower scale can be taken as a constant with their initial value. On the other side, if you are studying something on a slower scale, you can suppose that all the faster variables are in equilibrium.

[1] *As both a curiosity and a reflection on the changes to the international scientific communication, it is worth noting that this law was discovered by the chemist Cato Guldberg and the mathematician Peter Waage in 1867. But, as they wrote their research in Norwegian, it was not widely known until its rediscovery ten years later by chemist Jacobus Henricus van't Hoff.*

Now, with these considerations in mind, we can obtain the temporal dynamic of our minimal system by using the law of mass action over the set of reactions (1.1):

$$\frac{dm}{dt} = \beta_m - \alpha_m m, \tag{1.2a}$$

$$\frac{dp}{dt} = \beta_p m - \alpha_p p. \tag{1.2b}$$

This set of equations (1.2) defines the temporal trajectories defined by the system (1.1). Given the initial conditions and the parameter values, we could obtain the time evolution of each of the variables. Furthermore, as this is a trivial case, we could even solve them analytically instead of the regular numerical integration required to obtain the trajectories. But in most cases, this is not feasible and instead one should attempt to simplify the system as much as possible. The most usual strategy is to look for different timescales in the dynamics of the system. Here for example, as the system has an open interaction network the intermediate species m must reach equilibrium before p. Then we can take the eq. (1.2a) to equilibrium by setting $\frac{dm}{dt} = 0$, given that in equilibrium the system will stabilize and no longer change. By doing this, we obtain the value m^{eq} that m will reach $m^{eq} = \frac{\beta_m}{\alpha_m}$. Once you have this value, you can introduce it to the rest of the equations system and reduce the size of the system by an equation and a variable. But again, given the simplicity of this example, the substitution of m^{eq} allows for the direct solution of the system in this case. By also taking, $\frac{dp}{dt} = 0$ we obtain also the equilibrium of the p species $p^{eq} = \frac{\beta_p m^{eq}}{\alpha_p} = \frac{\beta_p \beta_m}{\alpha_p \alpha_m}$.

With this, we have already successfully both modeled the basal expression of an unregulated gene and solved its expression in equilibrium. But as we mentioned earlier, genes with such simple dynamics are quite rare given the huge genetic interconnection. Additionally, a gene like this would produce the same amount of RNA and proteins independently of the external conditions. This can be a huge waste of resources if the RNA and proteins are created without being needed (and is one of the most typical mutation problems (A. J. F. Griffiths et al., 2020)). Due to this, it is thought that only a few housekeeping genes are highly expressed constitutively. The roles of the genes with higher constitutive expression seem to be related to protein regulation (Zhao et al., 2017). Paradoxically, while some of them are ribosomal genes used to both build and maintain the protein factories, the others are related to the ubiquitination proteolytic pathway used to mark and degrade proteins (Zhao et al., 2017). This could seem strange if one thinks of the cell as a system in equilibrium, but that is precisely the error. Cells, and by extension any other organism, are systems profoundly out of equilibrium, constantly changing and adapting

to their environment. Additionally, they are optimized to avoid wasting resources. Due to this, they usually work in a focused way. Then if the external conditions suddenly change, the housekeeping genes ensure that the cell is ready to rapidly change gears by shifting protein concentrations.

Now that we know how to translate into differential equations molecular reactions, it is time to consider a system with a bit more complexity. For this, we will now obtain the mathematical model of one of the functional network motives mentioned earlier (Figure 1.2), the genetic toggle switch. As the name implies, this minimal two-gene system presents a rich dynamical behavior that can be exploited to externally regulate the expression of those genes. In the upper part of fig. 1.7 we present both the

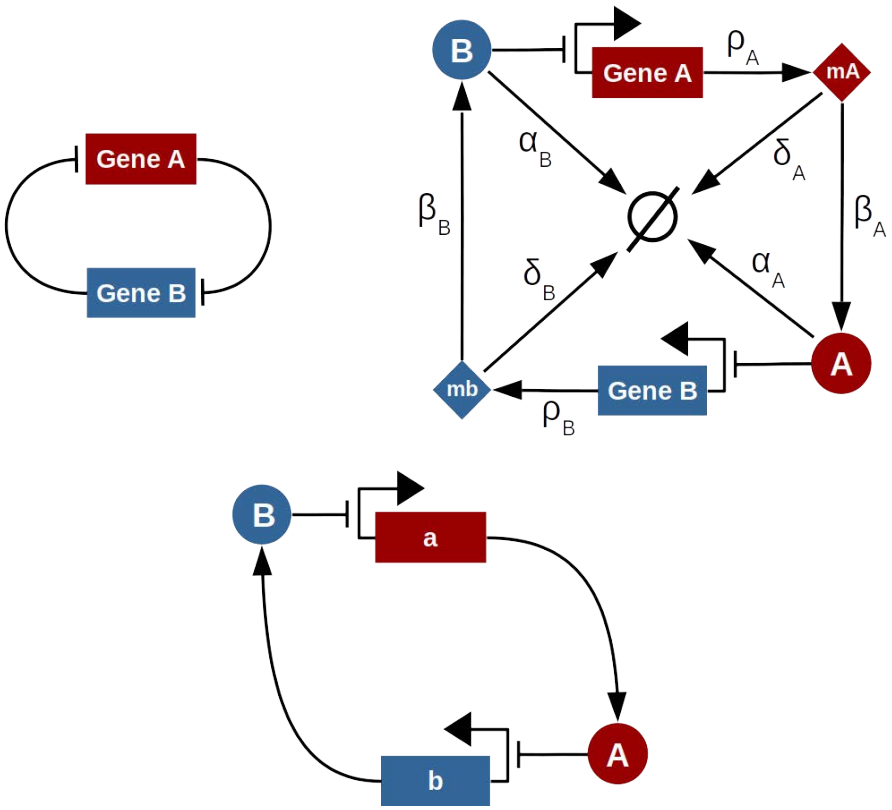

Figure 1.7: **Diagrams of the genetic toggle switch.**

typical gene interaction diagram and a more detailed diagram including genes, mRNAs, and proteins together with all the reactions of the system. While the lower part presents the representation that will be used

throughout the rest of the thesis. We can use the more detailed diagram to enumerate all the reactions of the system:

$$G^A + B \underset{k_{-B}}{\overset{k_B}{\rightleftarrows}} G_o^A, \qquad G^B + A \underset{k_{-A}}{\overset{k_A}{\rightleftarrows}} G_o^B, \tag{1.3a}$$

$$G^A \xrightarrow[\tau_A]{} G^A + mA, \quad G_o^A \xrightarrow[\tau_A \epsilon_A]{} G_o^A + mA, \qquad mA \xrightarrow[\delta_A]{} \varnothing, \tag{1.3b}$$

$$G^B \xrightarrow[\tau_B]{} G^B + mB, \quad G_o^B \xrightarrow[\tau_A \epsilon_A]{} G_o^B + mB, \qquad mB \xrightarrow[\delta_B]{} \varnothing, \tag{1.3c}$$

$$mA \xrightarrow[\beta_A]{} mA + A, \qquad A \xrightarrow[\alpha_A]{} \varnothing, \tag{1.3d}$$

$$mB \xrightarrow[\beta_B]{} mB + B, \qquad B \xrightarrow[\alpha_B]{} \varnothing. \tag{1.3e}$$

The first line (1.3a) describes the binding and unbinding of the A and B to the promoters of the other gene, G_A, and G_B. This attachment occupies the promoter, marked as G_o^A in our model, and reduces the expression of this gene by a factor ϵ_A in (1.3b) with $0 \leqslant \epsilon_A < 1$, given that it is an inhibitor. We have the same for the gene B (1.3c) and all the other elements are equivalent to the ones obtained in the single gene system (1.2). Given that now we have two promoter states, we will have to write dynamic equations for those variables. Despite this, we must remember that the total amount of genes is constant. Then for a gene X, the total number of copies will be the sum of the free and the occupied ones:

$$[G_T^X] = [G^X] + [G_o^X]. \tag{1.4}$$

From this set of reactions, we can obtain the following set of differential equations:

$$\frac{d[G^A]}{dt} = -k_B[B][G^A] + k_{-B}[G_o^A], \tag{1.5a}$$

$$\frac{d[G_o^A]}{dt} = k_B[B][G^A] - k_{-B}[G_o^A], \tag{1.5b}$$

$$\frac{d[G^B]}{dt} = -k_A[G^B][A] + k_{-A}[G_o^B], \tag{1.5c}$$

$$\frac{d[G_o^B]}{dt} = k_A[G^B][A] - k_{-A}[G_o^B], \tag{1.5d}$$

$$\frac{d[mA]}{dt} = \tau_A[G^A] + \tau_A\epsilon_A[G_o^A] - \delta_A[mA], \tag{1.5e}$$

$$\frac{d[mB]}{dt} = \tau_B[G^B] + \tau_B\epsilon_B[G_o^B] - \delta_B[mB], \tag{1.5f}$$

$$\frac{d[A]}{dt} = \beta_A[mA] - \alpha_A[A], \tag{1.5g}$$

$$\frac{d[B]}{dt} = \beta_B[mB] - \alpha_B[B]. \tag{1.5h}$$

Once we have this set of equations, we could finish our analytical work and simulate our system. If we input the values of all the parameters, set the initial concentrations for all the variables, and numerically integrate these expressions in time, we would obtain the proper trajectories of all the variables. But we would be overworking our computer because it is quite clear that there are several timescales in this system that we could use to simplify it and reduce its scale.

The first one is the promoter dynamics because, as stated previously in eq. (1.4), the genes are neither created nor destroyed. As a rule of thumb, one can consider reversible reactions will reach equilibrium faster than irreversible ones. The reasoning behind this is that while irreversibly is usually linked to complex and extensive changes such as conformation changes in protein or the creation and destruction of elements, the reversible processes are usually simple binding and unbinding. The more complex the process, the slower it is because it has to involve more elements. Then if we consider that the binding/unbinding dynamics are much faster than both the gene transcription and translation, we can consider that equations eqs. (1.5a) to (1.5d) are in equilibrium. By taking $\frac{d[G^A]}{dt} = \frac{d[G_o^A]}{dt} = \frac{d[G^B]}{dt} = \frac{d[G_o^B]}{dt} = 0$ we obtain the equilibrium conditions for the gene promoters. For gene A, we have that:

$$k_B[B][G^A]_{eq} = k_{-B}[G_o^A]_{eq} \Rightarrow [G_o^A]_{eq} = \frac{k_B[B][G^A]_{eq}}{k_{-B}} = \frac{[B]}{K_B}[G^A]_{eq}$$

where we have defined the equilibrium constant $K_B \equiv \frac{k_{-B}}{k_B}$ for the binding and unbinding to the promoter. Then, if we use the conservation of the total number of genes eq. (1.4), we can obtain expressions for both $[G^A]_{eq}$ and $[G_o^A]_{eq}$ that depend only of $[G_T^A]$ and $[B]$:

$$[G_o^A]_{eq} = \frac{[B]}{K_B}([G_T^A] - [G_o^A]_{eq}) \Rightarrow \frac{K_B}{[B]} = \frac{[G_T^A]}{[G_o^A]_{eq}} - 1 \Rightarrow [G_o^A]_{eq} = \frac{[G_T^A]}{\frac{K_B}{[B]} + 1} \Rightarrow$$

$$[G_o^A]_{eq} = [G_T^A]\frac{\frac{[B]}{K_B}}{1 + \frac{[B]}{K_B}}, \tag{1.6a}$$

$$[G^A]_{eq} = \frac{K_B}{[B]}([G_T^A] - [G^A]_{eq}) \Rightarrow \frac{[B]}{K_B} = \frac{[G_T^A]}{[G^A]_{eq}} - 1 \Rightarrow [G^A]_{eq} = \frac{[G_T^A]}{\frac{[B]}{K_B} + 1} \Rightarrow$$

$$[G^A]_{eq} = [G_T^A]\frac{1}{1 + \frac{[B]}{K_B}}. \tag{1.6b}$$

Once we have the equilibrium conditions for both promoters, eqs. (1.6) we can introduce them into eq. (1.5e) to obtain:

$$\frac{d[mA]}{dt} = \tau_A[G_T^A]\frac{1}{1+\frac{[B]}{K_B}} + \tau_A\epsilon_A[G_T^A]\frac{\frac{[B]}{K_B}}{1+\frac{[B]}{K_B}} - \delta_A[mA] \Rightarrow$$
$$\frac{d[mA]}{dt} = \tau_A[G_T^A]\frac{1+\epsilon_A\frac{[B]}{K_B}}{1+\frac{[B]}{K_B}} - \delta_A[mA]. \tag{1.7}$$

We can then repeat this process exactly for the gene B and obtain the following expression for eq. (1.5f):

$$\frac{d[mB]}{dt} = \tau_B[G_T^B]\frac{1+\epsilon_B\frac{[A]}{K_A}}{1+\frac{[A]}{K_A}} - \delta_B[mB]. \tag{1.8}$$

With only the consideration of rapid equilibrium for the gene promoter we have already reduced our original model to half its variables:

$$\frac{d[mA]}{dt} = \tau_A\frac{1+\epsilon_A\frac{[B]}{K_B}}{1+\frac{[B]}{K_B}} - \delta_A[mA], \tag{1.9a}$$
$$\frac{d[mB]}{dt} = \tau_B\frac{1+\epsilon_B\frac{[A]}{K_A}}{1+\frac{[A]}{K_A}} - \delta_B[mB], \tag{1.9b}$$
$$\frac{d[A]}{dt} = \beta_A[mA] - \alpha_A[A], \tag{1.9c}$$
$$\frac{d[B]}{dt} = \beta_B[mB] - \alpha_B[B]. \tag{1.9d}$$

It is worth noticing that we also reduced the number of parameters by the same amount, we now have 4 parameters in our model. In this last formulation of our model (1.10) we have redefined the parameters τ_A and τ_B by absorbing the constants $[G_T^A]$ and $[G_T^B]$ into them. If we denominate now the old parameters τ' ($\tau_A \equiv \tau'_A[G_T^A]$ and $\tau_B \equiv \tau'_B[G_T^B]$). This means that while originally the τ parameter represented the maximum expression rate for each gene promoter, each gene now represents the total amount of a particular mRNA produced in the cell (it accumulates all the cell copies). Something similar happened with the promoter binding constants k and k_- in our system now we only have the equilibrium constant $K \equiv \frac{k_-}{k}$. In general, as one reduces a system by using approximations, the resulting parameters are each time less mechanistic because they start to represent the combination of several processes instead of a particular process.

Again, we could stop here and simulate the dynamics of the system if one has the value for all the remaining parameters. In general, one should tailor the deepness of the model to the experimental data available for the studied system. But in most cases, particular experimental rates for molecular reactions are quite rare and one ends up with several free parameters to adjust with the existent data (we will discuss this in more depth in the next section 1.2.1). Due to this it is usually better to simplify the system as much as possible even if this reduces the identifiability of the model. If one does not have enough information, it is better to reduce the number of free parameters to avoid over-fitting (more on this subject in section 1.2.1).

Here we still have one more simplification available. As we observed in the previous model (1.2), the mRNA acts simply as an intermediate state. Additionally, while the transcription and the translation occur at a very similar rate, the degradation of mRNA is much faster (Milo and R. Phillips, 2016). While the median mRNA degradation lifetime is roughly 5 min in E. coli, the fastest actively degraded protein has a half-life of about one hour (Milo and R. Phillips, 2016). This faster turnover would make the mRNA reach equilibrium much earlier than the proteins. Then we can consider fast equilibrium conditions for both eqs. (1.9a) and (1.9b).

By taking both $\frac{d[mA]}{dt} = \frac{d[mB]}{dt} = 0$ we can easily obtain a system with only protein dynamics:

$$\delta_A [mA]_{eq} = \tau_A \frac{1 + \epsilon_A \frac{[B]}{K_B}}{1 + \frac{[B]}{K_B}} \Rightarrow [mA]_{eq} = \frac{\tau_A}{\delta_A} \frac{1 + \epsilon_A \frac{[B]}{K_B}}{1 + \frac{[B]}{K_B}}, \tag{1.10a}$$

$$\delta_B [mB]_{eq} = \tau_B \frac{1 + \epsilon_B \frac{[A]}{K_A}}{1 + \frac{[A]}{K_A}} \Rightarrow [mB]_{eq} = \frac{\tau_B}{\delta_B} \frac{1 + \epsilon_B \frac{[A]}{K_A}}{1 + \frac{[A]}{K_A}}. \tag{1.10b}$$

Finally, before introducing those equilibrium values for mA and mB we will introduce a couple of new parameters:

$$\rho_A = \frac{\tau_A}{\delta_A} \beta_A, \tag{1.11a}$$

$$\rho_B = \frac{\tau_B}{\delta_B} \beta_B. \tag{1.11b}$$

These new ρ parameters incorporate both the rates of transcription and translation together with the mRNA linear degradation term, and they represent the rate of protein creation over time from a given gene. With

this final arrangement, we obtain the toggle switch in its more simplified form, which is the one typically presented in the literature:

$$\frac{d[A]}{dt} = \rho_A \frac{1 + \epsilon_A \frac{[B]}{K_B}}{1 + \frac{[B]}{K_B}} - \alpha_A [A], \tag{1.12a}$$

$$\frac{d[B]}{dt} = \rho_B \frac{1 + \epsilon_B \frac{[A]}{K_A}}{1 + \frac{[A]}{K_A}} - \alpha_B [B]. \tag{1.12b}$$

As we kept the regulatory effect general, we could also see the activator role for, $\epsilon > 1$ given that the protein binding to the promoter would increase its efficiency. It is worth mentioning that the resulting regulatory function for the activation mode is the well-known Michaelis-Menten equation (Michaelis, Menten, et al., 1913).

In the paper that presented the construction of the genetic toggle switch, (Gardner et al., 2000) they have the maximum inhibition with both $\epsilon_A = \epsilon_B \equiv 0$ and cooperative effects on the protein binding. Here we have not considered any cooperative effects, but we could add them with an extra reaction for the binding to the promoter for each additional site. To ease the visualization we will just consider a second-order cooperative action for the B, then in this case the A promoter reactions would be

$$G^A + B \underset{k_{-B}}{\overset{2k_B}{\rightleftarrows}} G_1^A, \tag{1.13a}$$

$$G_1^A + B \underset{2k_{-2B}}{\overset{k_{2B}}{\rightleftarrows}} G_2^A, \tag{1.13b}$$

where we have changed the notation of the promoter occupied by a protein from G_o^A to G_1^A where the number indicates the number of proteins attached. We presented here the most complete case where the two promoter sites are equivalent and there is cooperative binding. The indistinct consideration causes the 2s in both the k_B rate in eq. (1.13a) (because there are two possible sites to bind) and the k_{-2B} rate in eq. (1.13b) (because there are two attached proteins equally possible to unbind). And the cooperative binding causes that $k_B \neq k_{2B}$ and $k_{-B} \neq k_{-2B}$ which means that the affinity of the protein to the promoter changes if there is already a protein attached. Typically, the affinity increases with the number of ligands attached to the promoter. But these characteristics are not true for all the gene promoters, so their incorporation into the model depends on the information already available in the system. But again, if one does not have concrete experimental information, it is typically better to opt for the most simple case to reduce the risk of over-fitting. In this case, it would be to consider distinctive sites (the proteins can only attach directly to one

site and this opens the next one) and, a bit paradoxically, that the affinity is approximately the same for all ligands. With these considerations, both the multiplicative factors and two parameters disappear because $k_{2B} \equiv k_B$ and $k_{-2B} \equiv k_{-B}$.

Then, one can repeat the same process of recursively considering equilibrium for all of them, analogously to eqs. (1.6), to obtain the new dynamics for the mRNA:

$$\frac{d[mA]}{dt} = \tau_A[G_T^A]\frac{1 + \epsilon_{1A}\frac{[B]}{K_{1B}} + \epsilon_{2A}\frac{[B]^2}{K_{1B}K_{2B}}}{1 + 2\frac{[B]}{K_B} + \frac{[B]^2}{K_{1B}K_{2B}}} - \delta_A[mA]. \tag{1.14}$$

This general expression considers indistinct cooperative binding sites. As such, we could adjust both the affinities and the effect on the transcription of the full and partially occupied promoters. If one instead considers the general case with n distinct binding sites (to avoid the multiplicity factor and have a more succinct expression) we would obtain the following regulatory function:

$$f(x) = \tau\frac{1 + \sum_{i=1}^{n} \epsilon_i \frac{x^i}{\prod_{j=1}^{i} K_j}}{1 + \sum_{i=1}^{n} \frac{x^i}{\prod_{j=1}^{i} K_j}} \tag{1.15}$$

where τ is the transcription factor, ϵ_i is the regulatory effect on the transcription rate (as before $0 \leqslant \epsilon_i < 1$ for inhibition, $\epsilon_i = 1$ no effect and $\epsilon_i > 1$ for activation) and K_i is the equilibrium constant for the binding of the i promoter site.

But once one starts to consider more binding sites, it starts to become a bit unrealistic to suppose that we would be able to determine experimentally all the particular binding affinities. This, together with the fact that the last ligand has typically a much higher affinity $K_n << K_{n-1}$, allows us to simplify eq. (1.15) to simply:

$$f(x) = \tau\frac{1 + \epsilon_n\frac{x^n}{K_n}}{1 + \frac{x^n}{K_n}} \tag{1.16}$$

The expression that we have obtained through the mechanistic model is the discrete exponent version of the named function, the Hill equation. This function was proposed empirically to describe the binding of oxygen to hemoglobin[2] by Archibald Hill in 1913 (A. V. Hill, 1913). It is worth noting that for the cases where the multimer structure is formed before attaching to the promoter (as it will happen for our *Anabaena* model described in part II) only the final form would have a regulatory function. In those cases, this is expression eq. (1.16), not a simplification given

[2] *For that particular case, the optimal exponent is 4 given that the hemoglobin presents 4 sites where the oxygen can attach itself.*

that the intermediate stages of the multimer do not affect the promoter. These experimental differences decide the proper definition of the equilibrium constant. Depending on the experimental data, you could have that $K_N \equiv K_n$ when only the last element attaches to the promoter. Alternatively, $K_N \equiv \prod_{i=1}^{n} K_i \simeq (K_n)^n$ when the multimerization occurs in the promoter, and therefore, we only could have experimental values of the monomer affinity.

The only difference between this final form eq. (1.16) and the original equation eq. (1.7) is in the exponent on the regulatory transcription factor. As one can observe in fig. 1.8 the exponent defines the sensitivity of the regulation. To be able to plot together activation and inhibition as complete opposites, we have supposed an activation mechanism that does not present constitutive production. This type of activation cannot be described as a multiplicative effect through an ϵ. Then, we can rewrite these expressions considering instead an additive effect where we use β for the constitutive rate $\beta \equiv \tau$ and ρ for the modification effects after the regulation. For the activators (as $\epsilon > 1$) we will have $\rho \equiv \tau(\epsilon - 1)$ and $\rho \equiv \tau(1 - \epsilon)$ for the inhibitors (given that $\epsilon < 1$). With these variable changes, we obtain:

$$f(t) = \beta + \rho \frac{\frac{x^n}{K_n}}{1 + \frac{x^n}{K_n}}, \tag{1.17a}$$

$$f(y) = \beta + \rho \frac{1}{1 + \frac{y^n}{K_n}}, \tag{1.17b}$$

This is the notation that will be used throughout the thesis. The behavior of both an activator eq. (1.17a) and an inhibitor eq. (1.17b) for several values of n is represented in fig. 1.8 in solid and dashed lines, respectively.

One can observe in fig. 1.8 that for lower values of the exponent, or Hill coefficient, the regulation is much more gradual, needing a much higher concentration of the ligand to affect the transcription. As n increases, the function resembles more and more a step function and behaves more like a threshold around the value K_n. This value is the only common point in all distributions, and the value of the function on this point will always be half its maximum value.

The promoter structures of genes are complex and there are a lot of possible mechanisms to regulate the expression of a gene. Due to this, it is possible that several transcription factors regulate the same gene. If those transcription factors act independently of each other, one can prove that

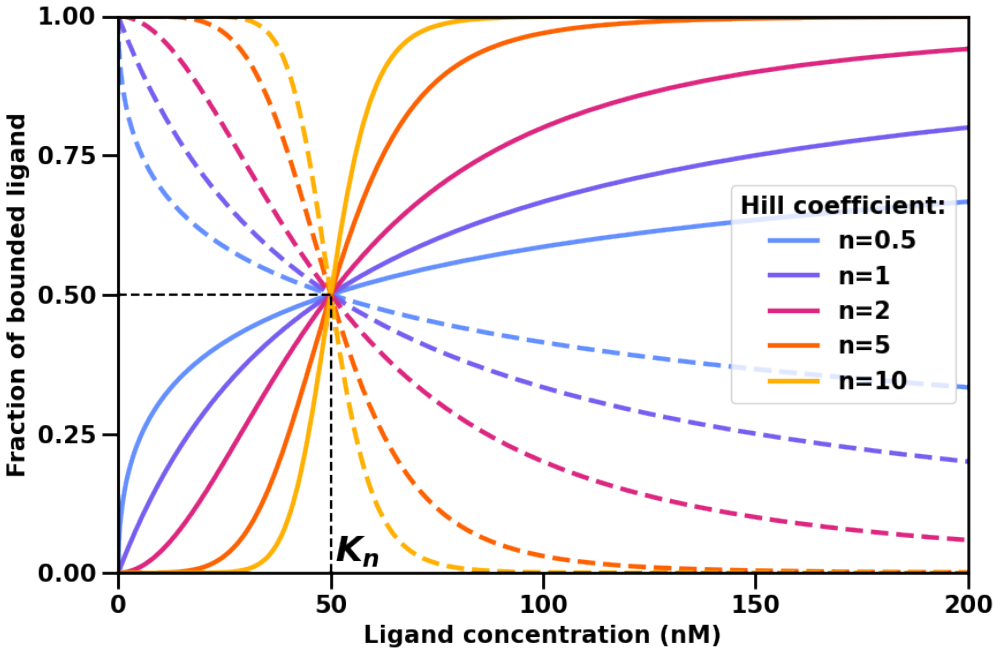

Figure 1.8: The figure shows the effect of the hill coefficient over the fraction of bounded ligand for a certain ligand concentration. Solid lines show the activation function eq. (1.17a) and dashed lines the inhibitory function eq. (1.17b).

their effects accumulate following the same methodology. The general regulatory term with varied transcription factors interactions is

$$f(x) = \frac{\sum_{p \in P} \frac{p^{n_p}}{K_p}}{1 + \sum_{p \in P} \frac{p^{n_p}}{K_p} + \sum_{i \in I} \frac{i^{n_i}}{K_i}}, \tag{1.18}$$

where P are the group of all the activators that intervene in the process with its respective equilibrium constant K_p and cooperative exponent n_p and I is the group of all the inhibitory transcription factors with their particular parameters.

With all this, we now have all the tools to translate both the regulatory network that controls pattern formation in *Anabaena* (studied in Part II and the one that controls lateral growth (or tillering) in Rice Part III into mathematical models of differential equations. Then we will now move to present the main strategies and methods used to analyze the models resulting from these proceedings in the following section.

1.2 MODEL ANALYSIS AND SIMULATION

Once one has obtained the final expression for the model, one has to think if it wants to describe the full dynamics of the system or only requires the equilibrium state. If we only require the equilibrium state, it is usually not necessary to integrate the dynamics until the system no longer changes.

As the example system for this section, we will use a simpler version with of the toggle switch with multiplicity presented in (Gardner et al., 2000)

$$\frac{d[A]}{dt} = \rho \frac{1}{1 + [B]^n} - [A], \tag{1.19a}$$

$$\frac{d[B]}{dt} = \rho \frac{1}{1 + [A]^n} - [B], \tag{1.19b}$$

which defined both the equilibrium constants and the degradation rates to 1 $K_A = K_B = \alpha_A = \alpha_B \equiv 0$. Additionally, we further simplified that expression by considering the same production rate ρ and multiplicity factor n for the two species.

A system is in equilibrium if it does not change with time. Then, to obtain the equilibrium state of this system, one only needs to solve $\frac{d[A]}{dt} = \frac{d[B]}{dt} = 0$. As this system is really simple, we can even solve it graphically. To do so we have to find the isoclines which are the set of point where one of the components of the derivative is null. This set of point will always have a dimension less than the original system. Then in this case we will have the two lines:

$$[A] = \rho \frac{1}{1 + [B]^n}, \tag{1.20a}$$

$$[B] = \rho \frac{1}{1 + [A]^n}. \tag{1.20b}$$

We represented these isoclines in the fig. 1.9 top for two different values of the multiplicity[3] $n = 1$ and 10. Along the red dashed line eq. (1.20a) we have that, $\frac{d[A]}{dt} = 0$ and in the blue dashed line eq. (1.20b) we have $\frac{d[A]}{dt} = 0$. Then the crossing points of these two lines will be the equilibrium states of the system. We can see that for $n = 2$, we only have one equilibrium state (to which all the trajectories in light green go). However, for $n = 10$ we have three equilibrium states with two stable ones, represented in blue, and one saddle-node represented in yellow color. As a result of this, the final state of this system with $n = 10$ depends on its initial conditions. The saddle-node point is only accessible through the black dashed line $[A] = [B]$, and any other trajectory will never cross this

[3] *These two particular curves have been already presented in fig. 1.8.*

line. Consequently, all the trajectories starting from a set of initial concentrations on a certain side of this line will end on the stable equilibrium point located on the same side.

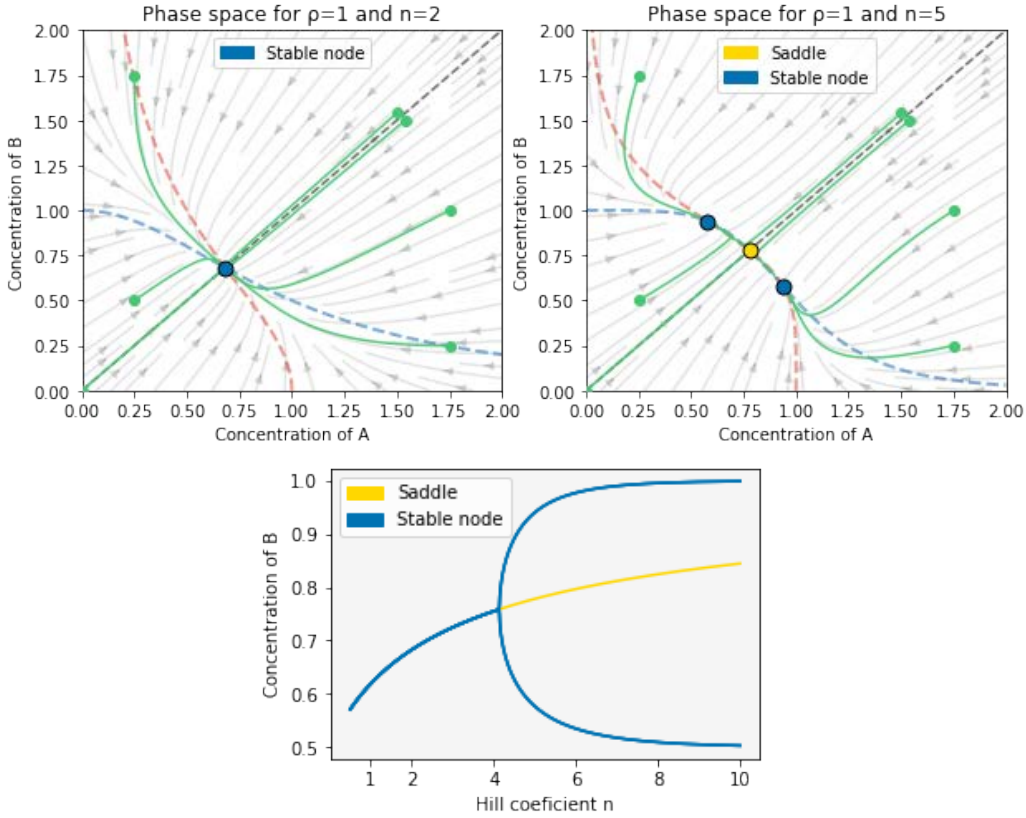

Figure 1.9: The upper plots show the phase space for two different values of the multiplicity $n = 1\&10$. The gray lines represent the flow of the system, several trajectories are drawn in green and the equilibrium points are represented with blue and yellow dots. To ease the visualization, we also included the isoclines of the system, represented with red and blue dashed lines, and the bisection with a dashed black line. The lower plot presents the bifurcation diagram for the multiplicity factor, also known as Hill coeficient, n. The color of the line define the stability of the equilibrium state. These images are obtained with a code adapted from the collected lecture notes by Guilhem Doulcier. License: CC BY-SA 4.0.

To obtain the nature of the equilibrium points, we have performed a linear perturbation analysis. But as we will not directly use it, we will not provide further details on this procedure. The main idea behind it is to study how the system reacts over a linear fluctuation over a steady state. If this fluctuation is always amplified the state is unstable, and if it is dampened it is a stable state. But it can also depend on the fluctuation characteristics, then the type of state options broaden. An example of this is the saddle equilibrium in fig. 1.9 which is unstable from all directions except the $A = B$ one. We only formally studied the stability

of our system in chapter 3, and in that case, this information was automatically provided by the XPP-Auto software. This is a typical software to numerically solve systems of differential equations. For more complex systems, it is usually easier to numerically solve $\frac{d[A]}{dt} = \frac{d[B]}{dt} = 0$. However, it is worth noting that this would provide the equilibrium points but, typically, not the flux. Due to this, we would have no information regarding the dynamics of the system. If we are interested on the dynamics of the system (as we will be in chapter 4), we should numerically integrate its system of equation from a set of initial conditions. In that case, the equilibrium states should be obtained through the repetition of different initial conditions to check for bistabilities with different trajectories.

This study can be repeated for a continuum array of n values to obtain the bifurcation diagram presented in fig. 1.9 bottom. This diagram shows the effect that the multiplicity has over the toggle-switch system and exemplify the richness in behavior that nonlinear effects provide even on a simple system. This sort of parameter sweep is really useful and informative when the number of parameters is small and one can individually study all of them (as we will do in chapter 3) or if we have some information regarding which are the most relevant ones.

For all the other cases when the system is too big for a systematic analysis, we will typically use optimization algorithms that autonomously reduce the parameter range to explore.

1.2.1 *Model adjustment and parameter optimization*

In order to design a systematic exploratory algorithm to search the best set of parameters, we need to find a method to quantify the quality of a parameter set. To do so, we will define a function that compares the expected experimental values of all the known observable with the ones obtained with each set of parameters. This function is commonly known either as loss or cost function and must be well-behaved, strictly positive, and present the minimum value when those two values coincide. For this, quadratic differences such as the mean squared error such as

$$E(P) = \sum_{x \in Obs} \left[\omega_x (x_{exp} - x_P)^2 \right] \tag{1.21}$$

are typically used. We have added an ω_x weight that can modify the relative relevance of an observable x. Then x_{exp} would be the experimental value of that observable and x_P the value produced by the model with a set P of parameters.

Once we have obtained this loss function, $E(P)$ we can treat it as a Hamiltonian that defines an energy of our model in the parameter space.

Then we have translated our parameter choice into an optimization problem of this eq. (1.21). Then, if our objective is only to reach a minimum of this function, the most efficient way to explore the parameter space is the so-called gradient descent strategy. Depending on the parameter sampling method, you have a huge array of specific protocols. But at its core, this strategy only contains one policy: to only accept parameter changes that reduce the energy of the system. This makes obvious sense, if you only move downwards, you will for sure end up with a better parameter than the one you started with. But if the energy landscape is rough with a lot of "peaks" and "valleys" is a really poor exploratory algorithm. It will be completely captured by the first local minima that it finds, even if it is the shallowest one. To solve this issue, one could compromise a bit the acceptation policy and decide to accept some parameters that increase the energy of the system only to expand the exploration range. This is the main idea behind the Metropolis–Hastings algorithm (Metropolis et al., 1953).

To further the statistical mechanics approach, we can borrow the Boltzmann distribution to establish a more permissive acceptance policy. This distribution

$$p_i = \frac{e^{\frac{-E_i}{kT}}}{\sum_i^N e^{\frac{-E_i}{kT}}} \tag{1.22}$$

gives the probability to observe a state p_i with an energy E_i when the system is at a temperature T. k is the Boltzmann constant that relates proportionality between the energy of the particles of an ideal gas and its temperature and $\sum_i^N e^{\frac{-E_i}{kT}}$ (the Canonical partition function) is the normalization factor required to obtain that the probability of the system to be in some considered states is $1\sum_i^N p_i \equiv 1$. This probability can be easily expanded to consider the probability of a transition between two states. If we want to go from a state i to j and i is much more probable than j this transition should be really rare. Then this transition probability must be proportional to the relative probabilities of each state

$$P_{ij} = \frac{p_j}{p_i} = e^{\frac{-(E_j - E_i)}{kT}}. \tag{1.23}$$

[4] *This expression is called the Boltzmann factor because is no longer normalized and therefore is no longer a distribution.*

Then we can use this expression[4] as the acceptance protocol for our algorithm. Then for each proposed parameter change, we calculate the P_{ij} and obtain a random value from a uniform distribution between 0 and 1. Then if the obtained value is lower than P_{ij} we will accept the transition. Is it easy to see that with this criterion we will still accept all the transitions to lower energy parameter sets, for those $p_j > p_i$, and therefore $P_{ij} > 1$. But now we can tune the algorithm to also accept some transitions to higher energy through the temperature value. The physical in-

terpretation for this is that stochastic thermal fluctuations could provide the extra energy for the system to reach less optimal states. Then, the higher the temperature is, the more different is the behavior from the gradient descent. For really high temperatures, it can ve considered almost a completely stochastic exploration that could have convergence problems. And for lower temperatures, the Metropolis–Hastings algorithm acts as a slightly noisy gradient descent. This strategy has the potential to avoid the local minima capture, but with a slight caveat. The parameter trajectories obtained with the algorithm greatly depend on the chosen temperature. Then we incorporated an extra optimization problem, which is the temperature range that will provide the most efficient exploration of the parameter space to avoid the initial conditions bias without compromising too much of its convergence. There are several strategies to deal with this problem, such as adaptable parameter sampling or reset strategies (Newman and Barkema, 1999). We will not discuss extensively here this options, but we will incorporate some of them to our algorithm in each particular study, such as reset strategies in chapter 4 or shifting/oscillating Hamiltonian weights chapter 6. Here, we will focus on the common idea of concatenating an initial exploration on high temperature with a posterior refinement of the parameter value with a much lower temperature. This is the idea at the core of the simulated annealing algorithm presented in (Kirkpatrick et al., 1983) that we will extensively use through this thesis.

This algorithm takes its name from the annealing procedure in condensed matter physics (Laarhoven and Emile H. L. Aarts, 1987). The annealing consist in first a rapid heating of a substance, typically a metal, to melt it completely and therefore randomly arrange its particles. Then the heat bath is progressive and slowly cooled to allow the arrangement of all the particles in the lowest energy state available. As one could expect from the name, the algorithm can easily be explained through an analogy of this process. In its minimum expression, represented in fig. 1.10, the simulated annealing is a Metropolis–Hastings algorithm with an initial high temperature and a continuous cooling. Then, when the algorithm initiates its exploration (the top window plot in fig. 1.10), the high temperature allows for an almost random and exploration of the parameter space almost without taking into account the energy of the system. This allows for a more diverse initiations of the trajectories, even without altering the initial conditions. Once the temperature starts to decrease (the middle window plot in fig. 1.10), the algorithm will progressively reject more and more proposed parameter changes. This will increase the deterministic drift towards lower energy parameters while still allowing for fluctuations that avoid the local minima trapping. And finally during the last low temperature regime (the low window plot in fig. 1.10) the al-

gorithm preforms a pure local gradient descent to select the parameters with the lowest energy. This same representation in fig. 1.10 shows an exaggerated example that even with this methodology one cannot ensure that the minimum located is the global one. Due to this, it is typical to concatenate several consecutive runs starting with the previous best fit to ensure consistency. But in most cases, it is not necessary to ensure that a given parameter set is unique or the best one. We will typically stop the parameter search once we believe that the agreement between the model and the experimental observable.

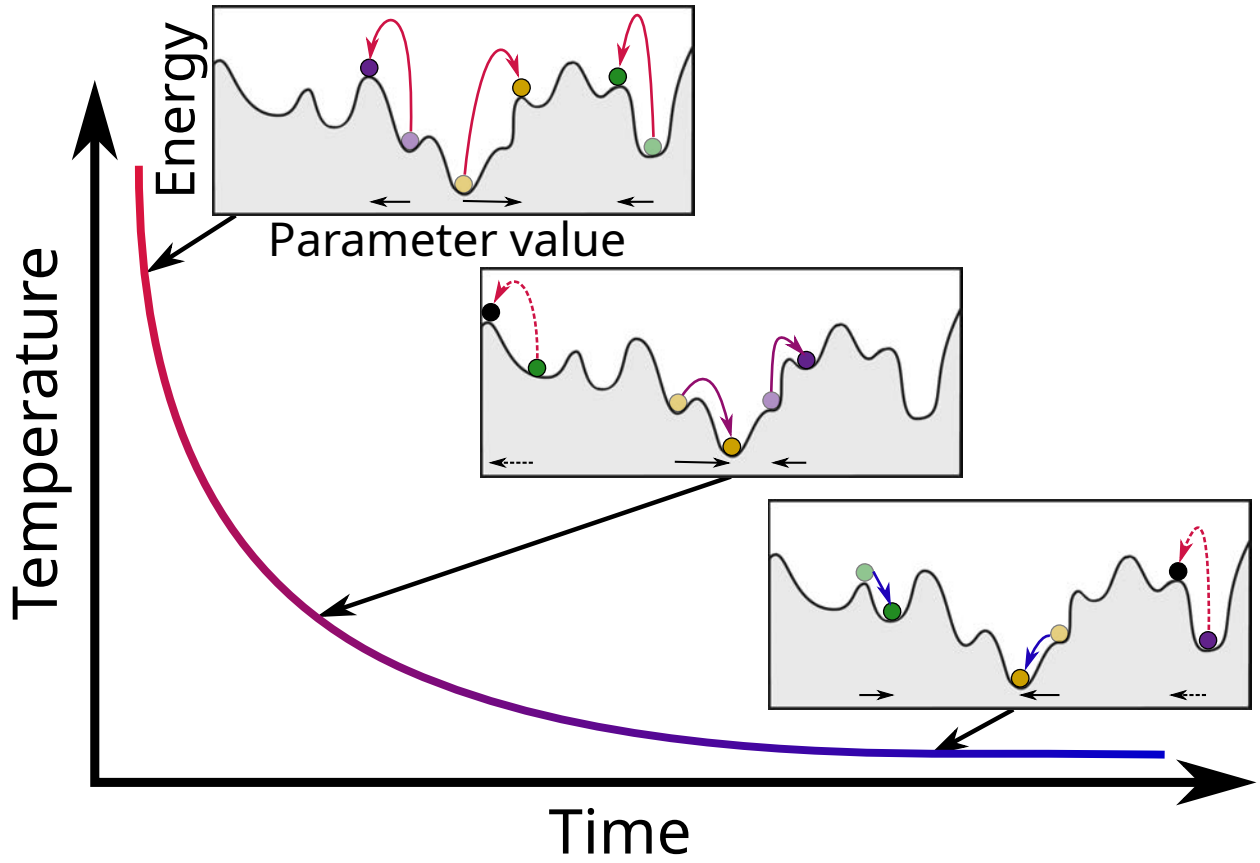

Figure 1.10: The figure shows three different imagined instances of a simulated annealing algorithm exploration in purple, orange, and green for three different temperature regimes. The energy of a given parameter value is represented by the greyed area in the window plots. The parameter changes are represented with black lines and the energy transitions with arrows colored with the temperature typically required to observe it. Brighter colors represent the final parameter value, the soft ones the initial one and black dots represent proposed parameter changes, represented with dashed lines, that the algorithm did not accept. Image adapted from Simulated Annealing Evolution by Sergio Ledesma, Jose Ruiz and Guadalupe Garcia (CC BY 3.0.)

There are numerous cooling strategies in the bibliography (E. H. L. Aarts and Korst, 1989; Laarhoven and Emile H. L. Aarts, 1987) with different degree of elaboration. But through this thesis, we used one of the simplest ones

$$T(t) = \frac{T_{ini}}{\sqrt{t}}. \tag{1.24}$$

We initially set the T_{ini} such that we would always accept a doubling of the energy in the first simulation step of the first run. For posterior analyses, we decreased a bit to have an 80% probability to accept a doubling in the first step. The more concrete configuration of each optimization algorithm will be discussed in each one of the chapters where we have used this technique.

Part II

NITROGEN FIXATION IN HETEROCYSTS

Here we will study the patterning of nitrogen-fixing cells (heterocysts) in cyanobacteria filaments. As the amount of heterocysts controls the filament fixing capabilities, if one controls the pattern one controls the filament's dedication to nitrogen fixation.

2

INTRODUCTION AND STATE OF THE ART

Cyanobacteria is the name of a phylum of prokaryote organisms that presents the common characteristic of performing photosynthesis. Ancestral cyanobacteria-like prokaryotes developed photosynthesis, which converts energy and oxygen from sunlight, water, and CO2, around 3 billion years ago. This gives them an advantage over the existent anaerobic prokaryotes that had to search for their nutrients. As a result of this, they proliferated and started to oxygenate the atmosphere, which until that point was completely devoted to oxygen (anaerobic) and formed primarily by nitrogen (N_2), carbon dioxide (CO_2) and methane (CH_4). This liberation of the oxygen stored in both the water HO_2 and the CO_2 through photosynthesis produced huge environmental changes oxidizing most of the superficial iron to hematite forming numerous continental Red beds. This massive liberation caused the great oxygenation event, which is the name of the saturation of the atmosphere with oxygen, up to more than a 30% concentration in oxygen. This change was supposed the first known mass extinction by killing most of the anaerobic organisms to which the oxygen was toxic. Additionally, the oxygen also reduced the amount of methane (CH_4) in the atmosphere by converting it to both water HO_2 and CO_2. This caused a thinning of the atmosphere, destroying a methane greenhouse on a timescale as short as 1 million years, that led to a global glaciation (the Makganyene snowball) that lasted several hundred million years (Kopp et al., 2005).

These two related events produced a niche for aerobic organisms that consume oxygen to grow and evolve, ultimately producing all the breathing biodiversity observed today. Not only this, but it is also hypothesized that the chloroplasts, the organelle where the photosynthesis is realized, originated through endosymbiosis of a cyanobacterium by an aerobic eukaryote organism (Archibald, 2015). This would make the cyanobacteria the cellular ancestor of all plant organisms.

Cyanobacteria are still extremely successful today and can be found in a very diverse range of environments with very different colony sizes. Huge cyanobacteria colonies, called blooms, arise suddenly in both fresh fig. 2.1A and salt water, fig. 2.1B covering as much as 200,000 km^2 of sea surface some summers (Kahru et al., 2018). On land, cyanobacteria can easily live in ponds and wet terrain, but small colonies are capable of even sustaining extreme environments such as drylands (both cold and hot)

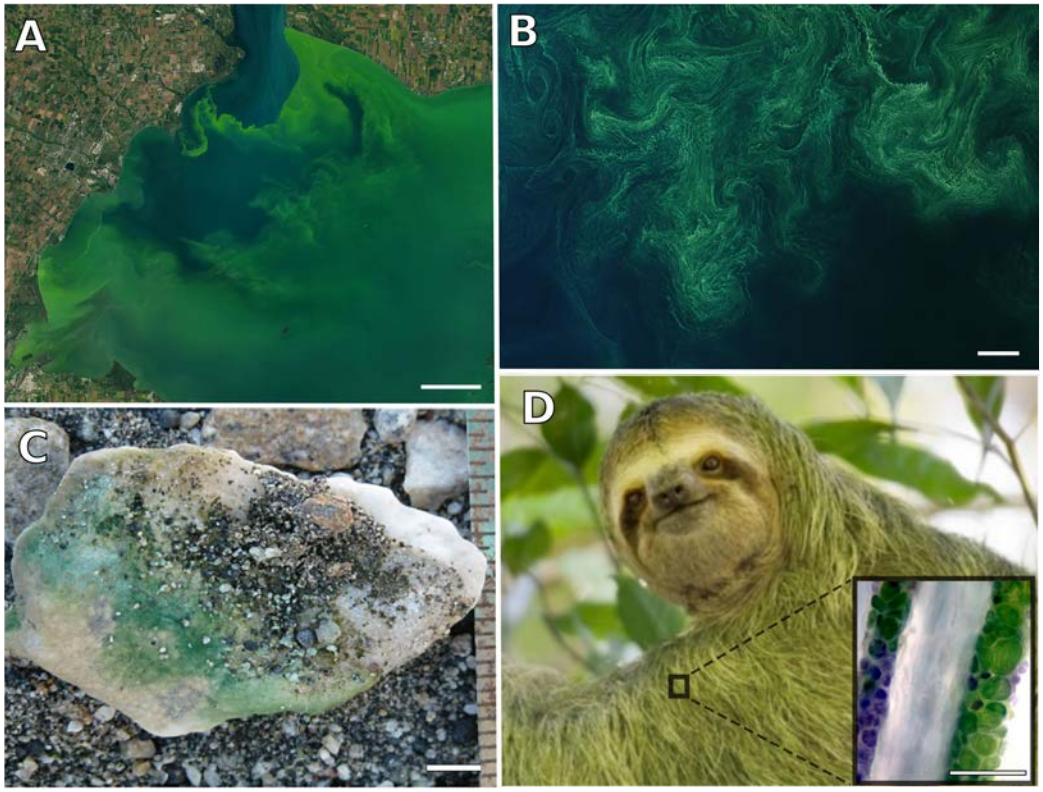

Figure 2.1: **Cyanobacteria environmental range: A** Freshwater bloom on Erie lake on September 26, 2017. **B** Saltwater bloom in the Baltic Sea on August 15, 2020. Both present the same 10 km scale bar and the images are reproduced from NASA Earth Observatory images by Joshua Stevens, using Landsat data from the U.S. Geological Survey (A and B). **C** Unearthed hypolithic cyanobacteria biofilm intimately associated with a quartz substrate. The scale bar measures 2 cm and the image is reproduced from (Chan et al., 2012). **D** Sloth fur as a mobile ecosystem for green algae and cyanobacteria (Kaup et al., 2021). The scale bar measures 20µm and the image is composed by a sloth photograph of Johner Images/Alamy and a zoomed hair modified from a figure in (Suutari et al., 2010).

and are capable to grow underground attached to quartz rocks fig. 2.1C. And finally, cyanobacteria also form symbiotic relationships with other organisms. The most eye-catching of those are represented in fig. 2.1D that shows that colonies of both green algae and cyanobacteria are capable to grow in the sloth fur. This imbues the fur with a green hue that facilitates the sloth camouflage, and it is also speculated in (Kaup et al., 2021) that there is also a direct exchange of substances between the organisms. Less striking, but much more relevant for this thesis, are the extensive examples of cyanobacteria-plant symbiosis (Lee and C.-M. Ryu, 2021). Cyanobacteria are capable to grow in both roots and leaves of plants to

both protect and contribute with fixed nitrogen to the plant while they receive carbon-based nutrients. This relationship and the possibilities that it opens will be more extensively discussed later in part IV.

This ubiquity shows huge adaptability and resilience, but is also probably helped by the fact that cyanobacteria are a pretty general biological term. It includes a varied biological group with quite diverse cellular lifestyles (Komárek, 2016).

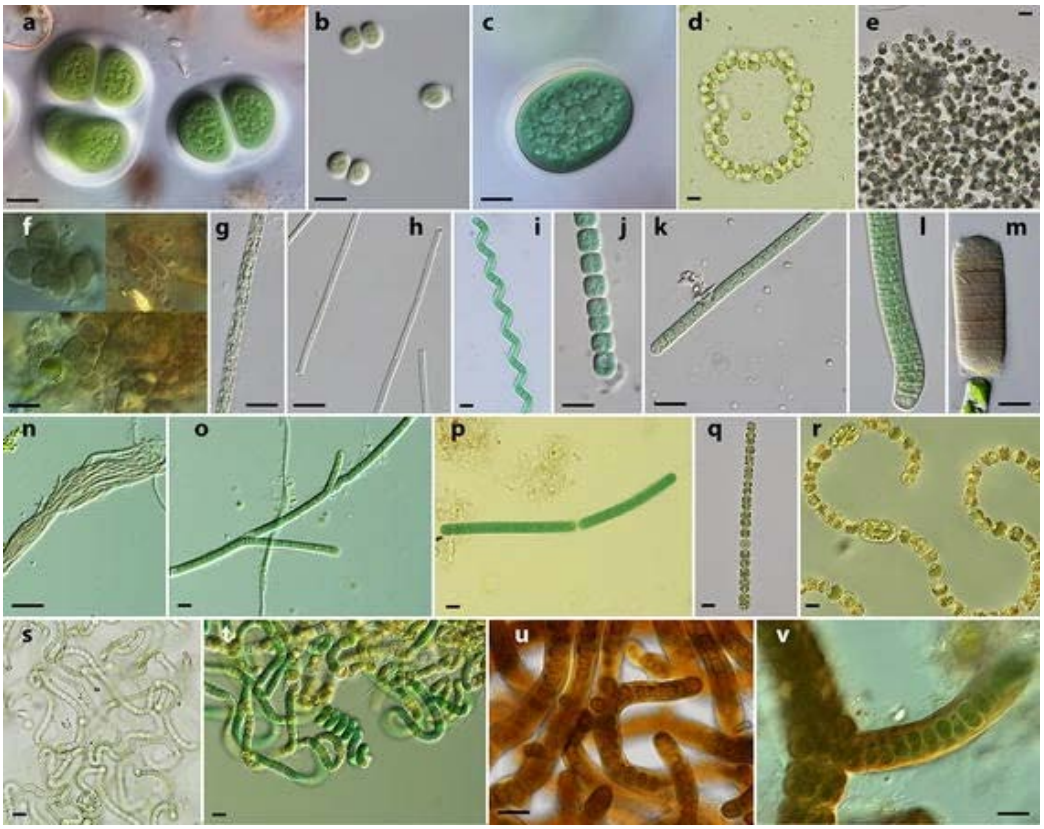

Figure 2.2: **Illustration of morphological diversity in cyanobacteria.** Families (morphological Orders) follow (Rippka et al., 1979). **I**. ***Chroococcales***: **a** *Chroococcus subnudus*, **b** *Ch. limneticus*, **c** *Cyanothece aeruginosa*, **d** *Snowella litoralis*, **e** *Microcystis aeruginosa*. **II**. ***Pleurocapsales***: **f** *Pleurocapsa minor*. **III**. ***Oscillatoriales***: **g** *Planktothrix agardhii*, **h** *Limnothrix redekei*, **i** *Arthrospira jenneri*, **j** *Johanseninema constricum*, **k** *Phormidium sp.*, **l**, **m** *Oscillatoria sp.*, **n** *Schizothrix sp.*, **o** *Tolypothrix sp.*, **p** *Katagnymene accurata.*, **IV**. ***Nostocales***: **q** *Dolichospermum planctonicum*, **r** *Dolichospermum sp.*, **s** *Nostoc sp.*, **t** *Nodularia moravica.* **V**. ***Stigonematales*** con : **u**, **v** *Stigonema sp.* Scale bar **a**–**u** = 10 µm, **v** = 20 µm. Reproduced from (Dvořák et al., 2015), CC BY 2.0 license.

As one can see in fig. 2.2 there are five big "classical families" distinguished through morphological criteria. Both the *Chroococales*, fig. 2.2a-e, and the *Pleurocapsales*, fig. 2.2e-f form aggregated bacterial colonies. This

would be a clear unicellular behavior. Others, the *Oscillatoriales*, form continuous filaments but without specialized cell types, constitute an example of borderline multicellular organism fig. 2.2g-p. And finally families, such as the *Nostocales*, fig. 2.2q-t, and the *Stigonematales*, fig. 2.2u,v, that present both a filamentous multicellular structure and specialized cell types. These specialized cellular forms are akinetes, enlarged reinforced cells that act as resource reservoirs, and heterocysts. These last ones will be the object of study of this part II. They are cells irreversibly dedicated to fixing ambient nitrogen into an organic form. This process is catalyzed by nitrogenase, and this enzyme is easily degraded by oxygen. In order to avoid this degradation, some cyanobacteria have developed a mechanism to protect nitrogenase from the oxygen produced by vegetative cells. This mechanism is to stop carrying out photosynthesis, develop an external glycolipid cell to impede oxygen diffusion from the medium, and finally degrade all remaining oxygen and photosynthesis-related proteins while it increases the production of nitrogenase. This nitrogen fixation is indispensable to the environment because only a few bacteria and archaea are capable of fixing atmospheric nitrogen. Then, both their ubiquity and adaptability and this unique ability make them crucial for the viability of all living beings on Earth. They provide the nitrogen necessary to sustain the vegetative biomass that constitutes the base of the trophic chain.

The biology of cyanobacteria has been the subject of intensive work during the last decades. A great number of studies have focused on a specific type of cyanobacteria that forms colonies consisting of one-dimensional filaments, the strain PCC 7120 of the genus *Anabaena*, to such an extent that this has become a model organism in the field (Harish and Seth, 2020; Herrero and Flores, 2019; A. M. Muro-Pastor and Hess, 2012; X. Zeng and C.-C. Zhang, 2022; C. C. Zhang et al., 2018). Under nitrogen-rich conditions, these filaments are composed only of vegetative cells carrying photosynthesis. However, as a response to different environmental stresses, vegetative cells can differentiate into specialized cell types that can fix atmospheric nitrogen. These individual conversions allow the survival of the colony and represent a paradigmatic example of a prokaryotic multicellular life form with differentiated cell types. When external nitrogen sources are scarce, heterocysts appear in a quasi-regular pattern, with intervals of roughly ten vegetative cells between heterocysts. Fixed nitrogen produced by the heterocysts reaches the vegetative cells and sustains their growth, as can be observed in fig. 2.3.

Reciprocally, nutrients produced by photosynthesis in vegetative cells are also shared to maintain the production of nitrogenous compounds in heterocysts, which require high energy consumption (Flores and Herrero, 2010; A. M. Muro-Pastor and Hess, 2012). Upon differentiation, hetero-

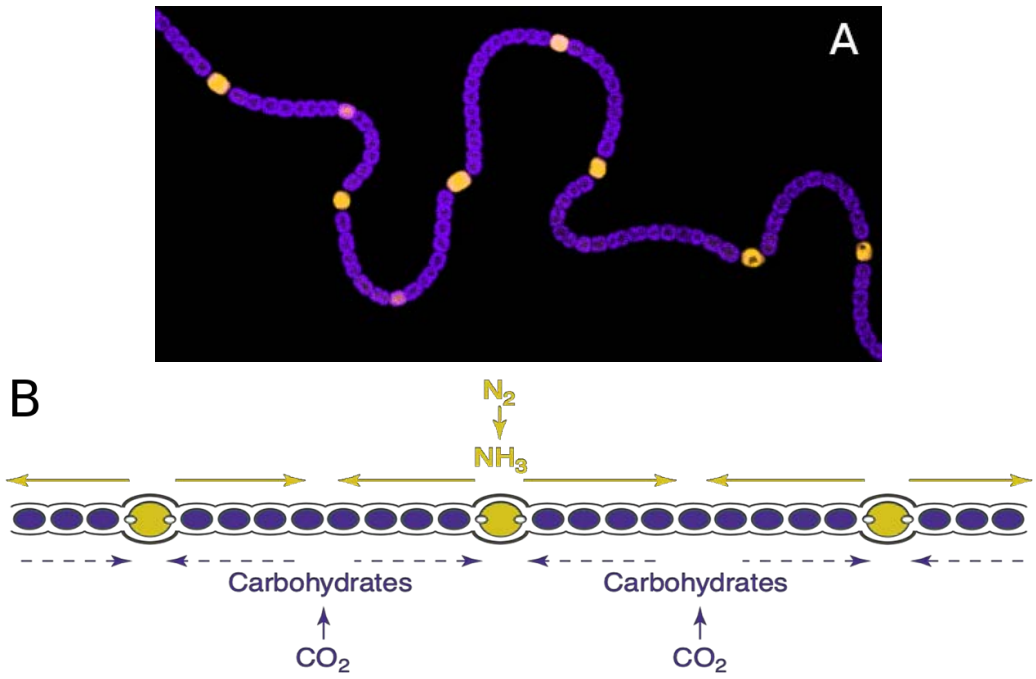

Figure 2.3: **Nutrient exchange in an *Anabaena* PCC 7120 filament under nitrogen deprivation:** The A image shows a fluorescence image of an *Anabaena* PCC 7120. The purple fluorescence is associated with the thylakoid membranes (and therefore to the photosynthesis) while the yellow fluorescence marks NsiR1 expression, which is described as an early marker of heterocyst differentiation in A. M. Muro-Pastor, 2014. Image courtesy of Alicia Muro-Pastor. The B diagram shows a sketch of an *Anabaena* filament with the same color code and the extracellular nutrient exchanges along the filament. Adapted from: James W. Golden and Ho Sung Yoon, 2003.

cysts lose the possibility to undergo cell division, but vegetative cells continue dividing, producing filament growth and increasing the distance between consecutive heterocysts. Due to this, new heterocysts differentiate in the middle of the intervals between previously existing heterocysts in order to not diminish the supply of nitrogen to distant vegetative cells. In this way, the dynamic process of differentiation allows the overall pattern of heterocyst location to conserve its properties over time.

There are many processes involved in the regulation of heterocyst pattern formation and its maintenance. In addition to nitrogen levels and other environmental aspects, many genes and transcription factors play a role (Herrero et al., 2013). In the next section, we will present a brief chronology of the identification of the known elements of the regulatory network.

2.1 EXPERIMENTAL CHRONOLOGY OF THE REGULATORY NETWORK

The first link between nitrate (organic nitrogen) and heterocyst formation appeared in (Fogg, 1949). In this work, it was shown that the presence of nitrate in the medium inhibits the formation of heterocysts in *Anabaena Cylindrica Lemm*. Despite this, the involvement of heterocysts in nitrogen fixation was not proposed until a few years later in (Fay et al., 1968) (even though they still did not provide any proof of their hypothesis). Roughly at the same time, it was hypothesized in (C. P. Wolk, 1967) that existent heterocysts condition the apparition of subsequent heterocysts through the inhibition of nearby cells. In their work, the authors show that fragmented filaments of *Anabaena Cylindrica Lemm.* (separating vegetative cells from heterocysts) produce more heterocysts than intact filaments. A similar result in (Water and Simon, 1982) unequivocally linked the heterocysts to nitrogen fixation. In this experiment, after the fragmentation of *Cylindrospermum licheniforme* filaments, a synchronous round of heterocyst differentiation was induced. This round of differentiation was coupled with a spike in nitrogenase activity between 13 and 26 hours that mirrors the production of heterocyst envelope components (glycolipid). This equivalent behavior implies a close relationship between heterocysts and nitrogen fixation.

Once the close relationship between heterocysts and nitrogen processing was established, the main objective in the field moved to the screening for the different genes that control the differentiation process. The first gene linked to this process was *ntcA*. This gene was first identified in (Vega-Palas et al., 1990) (during the analysis of mutants of *Synechococcus sp.Strain PCC 7942* incapable to assimilate nitrate) and properly characterized in (Vega-Palas et al., 1992), (T. F. Wei et al., 1994), (Frías et al., 1994), and (Luque et al., 1994) as an essential gene for the atmospheric nitrogen assimilation and heterocyst development. After depletion of the nitrate present in the media, the role of *ntcA* would be to start the differentiation process by up-regulating both itself and the master regulator of the heterocyst transition, *hetR*. The gene *hetR* was initially identified in (Buikema and Haselkorn, 1991) by the isolation of *Anabaena* sp. strain PCC 7120 mutants unable to grow aerobically on media lacking fixed nitrogen and proposed as a regulatory protein crucial for the differentiation process. This regulatory function and its capability to positively self-regulate itself was confirmed in (Todd A. Black et al., 1993).

On the other hand, the two main inhibitors of the heterocyst differentiation process, *hetN* and *pats*, were also identified in the 1990s. Curiously enough, they were discovered through the same research line that was initiated with (Ernst et al., 1992), where they generated mutants of *An-*

abaena sp. strain PCC 7120 incapable of sustained growth on media lacking fixed nitrogen using TnS-derived transposons. First, *hetN* was identified in (T. A. Black and C. P. Wolk, 1994) while studying a particular insertion location that was capable of arresting the production of heterocysts and multiple contiguous heterocysts (Mch phenotype) depending on the inserted gene. This insertion location was found to be the locus of *hetN*, whose deletion mutant also produces a Mch phenotype. A following work (Bauer et al., 1997) showed that *hetN* transcription starts after at least 6 hours of nitrogen deprivation and also located another cosmid (8E11) with the ability to suppress heterocyst formation. Finally, the authors of (Ho Sung Yoon and James W. Golden, 1998) studied this cosmid and identified the gene responsible for the heterocyst suppression, which they denominated *patS*.

In *Anabaena* sp. strain PCC 7120 this gene blocks heterocyst differentiation when overexpressed and its deletion mutant shows an increased frequency of heterocysts and a Mch phenotype. Moreover, it was also shown that the last five peptides of this gene (PatS-5) are enough to inhibit heterocyst formation. This led the authors to propose that a processed PatS peptide, originating from differentiating protoheterocysts, diffuses along the filament by neighboring cells, creating a gradient of an inhibitory signal.

Also during this timeframe, the gene *patA* was identified in (Liang et al., 1992) by the study of a mutation whose phenotype presented heterocysts mostly at the filament ends. Additionally, the deletion of *patA* suppressed the multiheterocyst phenotype produced by extra copies of the wild-type *hetR* gene. Just a bit after the turn of the century, another essential gene for the heterocyst formation in *Nostoc punctiforme* was presented in

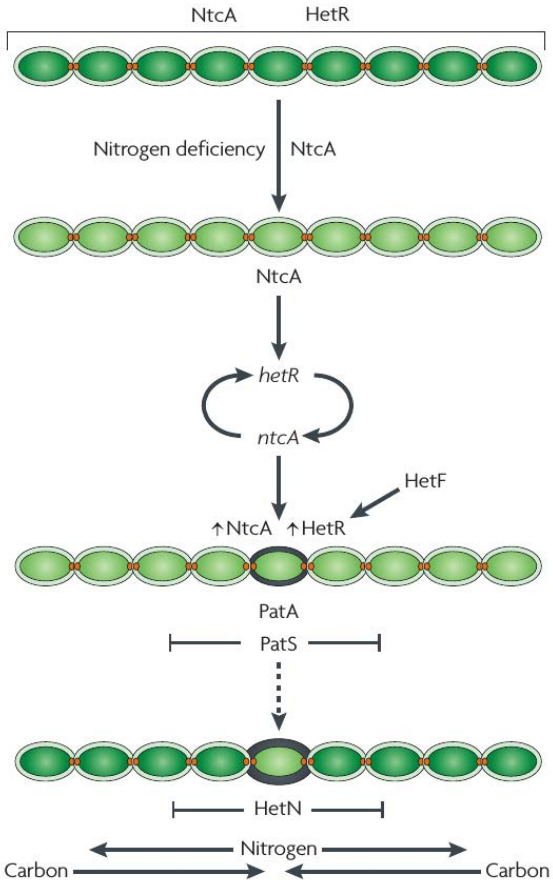

Figure 2.4: **Progress of heterocyst differentiation:** Scheme of the process of differentiation of heterocyst-forming cyanobacteria, with the roles of the main genes involved. Darker green means more fixed nitrogen in a cell, and the darker border represents the glycolipid cell wall that is formed once the cell is committing to differentiation. Reproduced from (Flores and Herrero, 2010).

(Wong and J. C. Meeks, 2001), where it was shown that the Δ*hetF* deletion mutant is not able to form heterocysts while presenting a higher concentration of HetR than the wild type. Advances in the study of this gene will not arrive until much further ahead, at this point it was only hypothesized that they belong to the same regulatory network controlled by *hetR*.

At this point, at the beginning of the 21st century, the main players in the regulatory network of heterocyst differentiation were already identified, albeit with different levels of detail. Besides *patA* and *hetF*, which gave the impression of being relevant but without a clear role, the rest of the mentioned genes seemed to have a well-defined role presented in fig. 2.4. Therefore, the experimental research focus shifted from the "Who?" to the "How?". Then the researchers started to study the interactions between those genes, mainly studying the epistasis between the several mutants.

For example, a more detailed description of the mechanism of nitrogen sensing used by *ntcA* was presented in (M. I. Muro-Pastor et al., 2001), (Tanigawa et al., 2002), and (Maria Félix Vázquez-Bermúdez et al., 2003). It was shown that the nitrogen status is detected through the internal level of the 2-Oxoglutarate metabolite and therefore closely relates the heterocyst formation with the nitrogen metabolism of the vegetative cells. The close relationship between *ntcA* and *hetR* was further studied in (A. M. Muro-Pastor et al., 2002), (Olmedo-Verd et al., 2005), and (Ehira and Ohmori, 2006), where the positive feedback loop between those genes is fully fleshed out.

Regarding the inhibitory gradients in the filament, (H. S. Yoon and J. W. Golden, 2001) showed that while the deletion mutant of *patS* presents clusters of heterocysts and shorter intervals of vegetative cells between heterocysts, this difference with the wild type phenotype gets smaller with time due to the inhibitory gradients of nitrogen products. Due to this, they hypothesized that the main role of PatS is the resolution of the local dispute by the vegetative cells to differentiate, and they situated the commitment window between 9 and 13 hours. A full characterization of the inhibitory role of *hetN* over *hetR* is presented in (S. M. Callahan and Buikema, 2001) where the native *hetN* promoter was substituted with a copper-inducible petE promoter, which allowed to tune the activation of *hetN* by controlling the concentration of copper in the media. Additionally, they observed that the *hetN* transcription is produced mainly in the heterocysts. Due to the previous, the system presents two distinct inhibitory timescales, the one of *patS*, which is produced early after nitrogen deprivation, and the one of *hetN*, which activates after the differentiation commitment. It was hypothesized that *patS* is relevant for the de novo pattern

formation, while *hetN* is more relevant for the pattern maintenance. Parallel, independent work by (B. Li et al., 2002) also reports inhibitory role evidence by showing that plasmid-mediated over expression of *hetN* is capable to avoid the up-regulation of *hetR* on nitrogen depletion.

It was not until a few years later that the first evidence of the inhibitory mechanism was presented in (Xu Huang et al., 2004). Here, the authors prove that HetR forms homodimers capable to bind to DNA to up-regulate the production of both itself and *patS*, and that the pentapeptide PatS-5 inhibits this up-regulation. These results showed that the inhibitory mechanism of *patS* is to impede the regulatory functions of HetR. This implication was later confirmed in two independent studies, (I. Y. Khudyakov and James W. Golden, 2004) and (Borthakur et al., 2005). In the first one, the authors identified a point mutation (a C to T transition) on the *hetR*, providing the organism with resistance to the over expression of both inhibitors and producing a completely random heterocyst location pattern when those genes are not over expressed. In the second study, the authors showed that both inhibitory pathways are independent of each other and therefore can function in the mutant background of the other gene. Additionally, they observe that the double mutant Δ*patS*Δ*hetN* leads to an almost complete differentiation of the filament that can be reduced by the continued supply of ammonia to the system. Therefore, these two works presented evidence that, while the pathways of both inhibitors seem independent and act at different time points of the differentiation, they present a certain similarity in the inhibitory mechanism. The final proof of the effect of the inhibitors over HetR concentration came in (Risser and Sean M. Callahan, 2009). In this study, the authors showed gradients of GFP-tagged HetR that increase as one moves away from a cell that produces inhibitors and reported that this gradient is caused by inhibitor-mediated degradation of HetR: *"Is it also worth noting that this work was realized both in mutants that presented a higher concentration coupled with a lower amount of heterocysts: Two modifications of the hetR and deletion of patA mutants which share a common phenotype: a higher concentration coupled with a lower amount of heterocysts."(direct quotation from Risser and Sean M. Callahan, 2009)*.

Regarding the relationship of the gene *patA* with the rest of the regulatory network, first (Makarova et al., 2006) studied the sequence of PatA and detected a domain (PATAN) related to protein-protein interactions, suggesting a post-transcriptional regulatory role. And afterward, a couple of studies, (Orozco et al., 2006) and (Risser and Sean M. Callahan, 2008), covered the study of the epigenetic relationship of this gene. In (Orozco et al., 2006) the relationship between *hetR*, *hetN*, *patS*, and *patA* was studied using multiple mutants. The authors showed that, while the

deletion phenotype is epistatic to any other deletion, the over expression of *patS* is epistatic to *hetR* over expression. This indicates that *hetR* acts upstream of all the other genes, given that all of them require a functional *hetR*. They also observed that while the Δ*patA* phenotype (a filament with heterocysts mostly on the filament ends) gets more extreme on the Δ*patA*Δ*hetN* double mutant (which presents multiple heterocysts on the ends), it is not present on both the Δ*patA*Δ*patS* double mutant and Δ*patA*Δ*hetN*Δ*patS* (Orozco et al., 2006). The first, Δ*patA*Δ*patS*, is more similar to the patS single mutant with longer vegetative intervals than to the Δ*patA*. And Δ*patA*Δ*hetN*Δ*patS* presents fewer heterocysts than the hetNpatS double mutant.

These results indicated that a functional *patS* is required to observe the *patA* phenotype and led the authors to suggest that PatA reduces the effect of HetN and PatS while independently promoting differentiation through HetR (Orozco et al., 2006). On the other hand, in (Risser and Sean M. Callahan, 2008) the relationship between *hetF* and *patA* was studied. They reported that the HetR concentration increase observed in Δ*hetF* and Δ*patA* mutants were caused by a reduction of the HetR protein turnover. Additionally, they also observed that *hetF* is capable to recover the Δ*patA* mutant and is also necessary for the HetR mediated transcriptional promotion of both itself and *patS*. These two facts suggest an unknown shared role of these genes in a post-transcriptional activation of HetR that allows its regulatory function.

A bit more recently, it was also shown in (Hou et al., 2015) that the expression of *patA* is also regulated by *hetR*. And finally, a very recent study (Valladares et al., 2020) showed a relationship of this gene with several components of the Divisome and, therefore, with the control of the cell division.

Also, during this first decade of the 21st century, two new genes that had previously received less experimental focus were identified. The first one, *hetL*, was presented in (D. Liu and James W. Golden, 2002). Its overexpression relieves inhibition by PatS-5 pentapeptide in *Anabaena* PCC 7120 and, therefore, it is affecting heterocyst regulation downstream of PatS production. This has been recently confirmed for *Nostoc* PCC 7120 in (Xiaomei Xu et al., 2020), where it was shown that HetL provides immunity to HetR to the inhibitory action of PatS. The second gene, *hetZ*, was studied in (W. Zhang et al., 2007). The authors showed that its mutants present delayed or no heterocyst differentiation at all. A posterior work by (Videau et al., 2018) showed that *hetZ* over-expression can fully bypass hetR to produce functional heterocysts and hypothesized a role in the commitment to differentiation.

During more recent years, experimental efforts have focused on the mechanics of the interactions of the regulatory genetic network through the study of the proteins codified by the previously presented genes, with the aim to obtain the functional form of those proteins and the regulatory pathways in which they intervene. The first example is a study of the HetR structure with the use of crystallography. In (Kim et al., 2011) it was shown that HetR forms dimers capable of interacting with DNA and also that this interaction is impeded by the binding of PatS-5 (the RGSGR sequence shown to be capable to disrupt heterocyst formation (Xu Huang et al., 2004)) with HetR. Posterior research in (Valladares et al., 2016) also evidenced the existence of a tetrameric functional form of HetR with a regulatory effect on its phosphorylation (which, in turn, affects its regulatory function). This effect of the phosphorylation on the heterocyst formation was expanded in (Roumezi et al., 2020) through the study of the relevance of the Pkn22 kinase on the *hetR* regulation. Here, the authors showed that the phosphorylation of HetR is essential for its regulatory function.

Focusing now on the inhibitors, during the last decade a couple of studies (Feldmann et al., 2011, 2012) quantified the biochemical constants of the binding of the peptide sub-products of PatS, especially PatS-5 (the RGSGR sequence) and PatS-6 (which contains an additional peptide, ERGSGR), with HetR. As the affinity of PatS-6 with HetR is more than 30 times bigger than the one of PatS-5, the authors hypothesized that PatS-6 could be the actual functional form of the PatS inhibitor. At the same time, it was also shown in (Higa et al., 2012) that the functional form of HetN is also an RGSGR sequence equivalent to one of PatS. Slightly later arrived the consolidation of a peptide chain as the diffusible inhibitor, with the first evidence for intercellular transfer of peptide inhibitory products of *patS* presented in (Corrales-Guerrero et al., 2013). It is also known that despite having the same functional form, both the processing and export of this polypeptide (Rivers et al., 2014) and the promoter configuration of the genetic transcription (Rivers et al., 2018) are different between *patS* and *hetN*. And finally, the existence of redundancy in the early inhibition has been shown in (I. Khudyakov et al., 2020) through the identification of *patX*, a complementary gene of *patS* with a much lower regulatory effect.

Regarding more recent studies about *hetR* and the inhibitors *patS* and *hetN*, it has been reported the first evidence of their transcriptional dynamics during the heterocyst differentiation in (Di Patti et al., 2018). By tagging the proteins with GFP, the authors managed to characterize the activation of these three genes during the transition. While *hetR* shows a gradual activation that spans over more than 30 hours until reaching around 10 times the original transcription, both inhibitors have a sudden increase in transcription. For *patS* the signal resembles a spike that

reaches a maximum increase of 4-6 times the original transcription after around 5 hours and recedes to a higher stable transcription of 3-4 the original one after another 5 hours. Alternatively, the *hetN* signal increase suddenly over around 10 hours to reach a level of transcription 1.3 times the original one. By defining the commitment time as the point where the auto-florescence (a proxy for the photosynthesis) decays to half the maximum value, the authors were also able to observe that the *hetR* activation inversely correlates with the auto-fluorescence decay, that the maximum of the *patS* expression is located on this commitment time, and that the increase of *hetN* starts right after this commitment time.

It is worth mentioning that very recently some other genes with a relevant function during the heterocyst differentiation have been identified. For example, the gene *hetP* has been presented in (Videau et al., 2016) to work together with its homologous in a redundant switch-like mechanism that controls commitment. Posterior works by (H. Zhang et al., 2018) and (Yaru Du et al., 2020) have shown a functional overlap of this gene with *hetZ* for the control of heterocyst differentiation in response to *hetR* up-regulation. Another example is the gene *patD*, presented in (Li Wang et al., 2019), which is regulated by NtcA. Its inactivation increases the heterocyst frequency (mainly through a *hetN* down-regulation), without losing the heterocyst patterning, and reduces growth efficiency.

Experimental results, such as the previously described, have allowed advancement in the understanding of the mechanisms and interactions between the known main players: *hetR*, *patS*, and *hetN* that give rise to the appearance and maintenance of heterocyst patterns. However, in addition to these, other transcription factors such as *patA* and *hetF* have been shown to play an important role in the early steps of differentiation, regulating the transcriptional activity of *hetR*. All this complex network of interactions, represented in fig. 2.5, where other heterocyst related genes, such as *hetC*, *hetP*, *hetL*, *patN*, and *hetZ*, also play a role, has made the complete understanding of heterocyst differentiation a challenge during the last two decades.

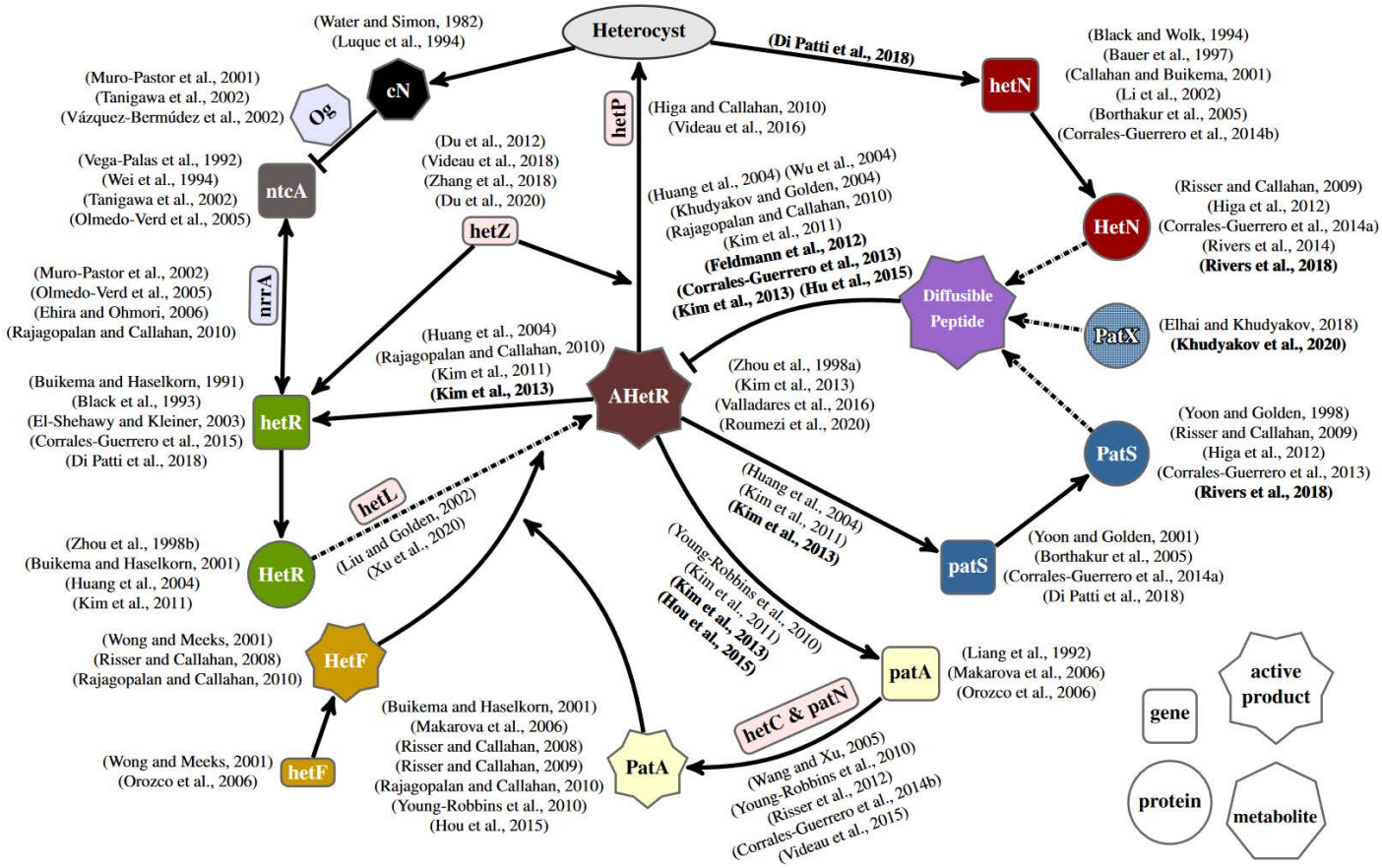

Figure 2.5: **Gene regulatory network of heterocyst differentiation:** The main elements and their interactions are depicted schematically, together with other relevant elements with either more dubious (in the light red background) or oversimplified roles (in the light blue background) in the scheme. AHetR stands for the active form of HetR. The ellipse represents the differentiation into a heterocyst. Arrows with solid lines represent interactions between elements. Arrows with dashed-dotted lines represent post-transcriptional changes. Regular and bold formatted references to indicate phenotypically inferred and observed molecular interactions, respectively.

Now that we have a functional regulatory genetic network, represented in fig. 2.5, we will also revise the different perspectives taken by previous modelization attempts over the system in the next section.

2.2 REVIEW OF THE EXISTENT MATHEMATICAL MODELS

Pattern formation is extremely relevant in embryonic development because it allows for precise periodic spatial differentiation of certain cells or groups of cells. An important question is how a pattern, and therefore heterogeneity, is produced from a homogeneous state, given that embryos develop from a single cell. Another intriguing feature is that patterning

must be robust enough to ensure reliability, given that embryo development is a highly reproducible process. Additionally, the widespread action of pattern formation in all organisms and different levels of development seems to point to the existence of simple intrinsic mechanisms capable to act with widely different elements.

The reaction-diffusion system, presented by Turing in his seminal work (Turing, 1952), constitutes a simple model capable of forming spatial patterns starting from a homogeneous state. Turing considers a ring of equivalent cells that generate a couple of diffusible morphogens whose production depends on the concentrations of both of them. He realized, through a linear perturbation analysis, that, while the system starts homogeneous, slight perturbations in the diffusion of morphogens are reinforced and create "waves" of morphogens in the cell ring. This reinforcement is caused because, when a cell sends more inhibitors to its neighboring cells than what it receives, the neighboring cells produce fewer inhibitors. This further reduces the flux of inhibitor that enters the cell, which in turn, increases inhibitor production and its flux to the neighboring cells. This feedback loop produces waves of morphogens that can drive the system to a heterogeneous state if system parameters are capable to sustain the perturbation out of the linear regime. Furthermore, if there are more than two diffusible morphogens, the heterogeneous state can be oscillatory. The general condition that allows these instabilities to form is the combination of an activator and a more diffusible inhibitor. The particular ratio between the diffusion rates is highly dependent on the reaction system that regulates these morphogens (Gierer and H. Meinhardt, 1972). This fine-tuning required for the pattern fixation questions the biological feasibility of this mechanism because it makes the system susceptible to small changes in parameter values that would alter greatly its behavior.

These types of biological pattern-forming systems were further extensively studied by Gierer and Meinhardt (Hans Meinhardt, 2008), and fully theoretically fledged out in Murray's textbook (Murray, 2003). A state-of-the-art discussion on Turing's ideas, their development, and some system examples can be found on a paper in this same Special Topic issue (Lacalli, 2022). Subsequently, Murray's analysis has been expanded to consider reaction-diffusion systems in continuously growing domains, observing that depending on the characteristics of the growth, it can produce more robust pattern formation, or add difficulties to it (Barrass et al., 2006; Crampin et al., 1999). Finally, the limiting case in which the activator does not diffuse cannot create a stable stationary pattern and, therefore, the emergent patterns are always of a dynamical nature (Marciniak-Czochra et al., 2017). The incorporation of mechano-chemical feedback can me-

diate the reinforcement and consequent fixation of the pattern through a morphological change that affects the diffusion of the inhibitory morphogen (Brinkmann et al., 2018).

When talking about these reaction-diffusion systems, it is important to remember that the insights from linear stability analysis, usually invoked to determine whether a system can form a stable pattern or not, can be deceiving: the dispersion relation close to a homogeneous fixed point can sometimes be very helpful, but also deceiving once full non-linearities kick in. For this reason, classical rules for pattern *formation* based on linear analysis are better understood as applying to pattern *inception*. Given that it studies linear perturbations and the stabilization of a final pattern is a process where non-linearities cannot in general be neglected (Smith and Dalchau, 2018). For instance, against classical thinking, systems with equally diffusing signals can make stable patterns (Marcon et al., 2016). In this framework, it is clear that events like domain growth (Raspopovic et al., 2014), discrete nature of the system (Nakamasu et al., 2009), or separation of timescales for the action of different molecular species, can all play a role to shape the formation and maintenance of patterns.

All these characteristics are relevant for the study of pattern formation in the filamentous cyanobacterium *Anabaena* (fig. 2.6). The cells of the filament exchange nutrients and react as a whole to environmental changes. Due to this, one can consider the filament as a whole multicellular organism. This is especially evident when the filament is placed in conditions of nitrogen deprivation. Under these conditions, the filament undergoes a dynamical differentiation process that differentiates roughly one in every ten cells into nitrogen-fixing heterocysts, in a quasi-regular pattern that is maintained as the filament keeps growing (Flores and Herrero, 2010). This patterned differentiation constitutes an example of specialization, cooperation, and distribution of labor because, while the vegetative cells keep producing carbon through photosynthesis, the heterocysts fix environmental nitrogen into organic forms that can be assimilated by all cells. Thus, for the filament to subsist, both end products must be shared and diffuse through the filament to the cells that are not capable of synthesizing them. While previous reviews have already compiled the current theories about heterocyst pattern formation (Flores and Herrero, 2010; Harish and Seth, 2020; Herrero et al., 2016; X. Zeng and C.-C. Zhang, 2022), in this chapter we will systematically discuss the different mathematical and computational frameworks that have been used to model the physics of cell differentiation and pattern formation in this system.

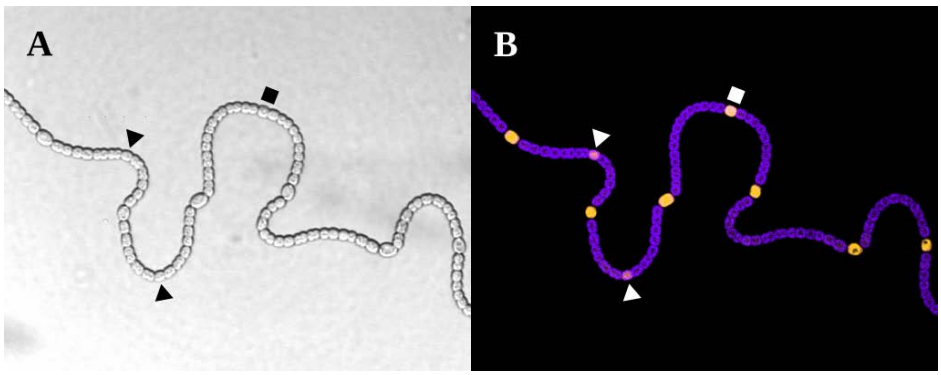

Figure 2.6: A. Wild type *Anabaena* PCC 7120 filament grown in conditions of nitrogen deprivation, showing vegetative cells and heterocysts. B. Fluorescence images that characterize the differentiation stage of each cell. The purple fluorescence is associated with the thylakoid membranes (and therefore to the photosynthesis) while the yellow fluorescence marks NsiR1 expression, which is described as an early marker of differentiation in (A. M. Muro-Pastor, 2014). The intermediate stages of developing heterocysts are indicated by polygonal shapes. Images courtesy of Alicia Muro-Pastor.

2.2.1 *Diffusion models of an inhibitor exported from heterocysts*

First, we present works that attempt to model the patterned distribution of heterocysts without any explicit genetic regulation. With this aim, these models only consider gradients in nitrogen concentration or some inhibitory signal originating from heterocysts.

An early attempt at modeling the patterning of heterocyst differentiation in cyanobacteria filaments came at a time when, while the biological role of heterocysts was not well-defined (Fay et al., 1968), it was already stated that heterocysts seem to inhibit the formation of new ones (C. P. Wolk, 1967). The model presented in (Baker and Herman, 1972) consisted of an integer linear cell array simulator which allowed cell-to-cell diffusion of an inhibitory product and division of vegetative cells. Due to computational limitations at the time, concentrations were modeled as integer numbers, as a result, set a discrete minimum change in concentration. This model was used to test the hypothesis that cell division and differentiation are two competing processes in which, at the end of each cell cycle (quantified by a countdown), cells have to choose a fate depending on the inhibitor concentration. This simple model can obtain feasible distributions for heterocyst placement, but the code was heavily limited by having to work with integer concentrations. The model predicted a low threshold of the inhibitor to avoid differentiation, causing integer rounding to be comparable with concentration values. This low

threshold was probably caused by the unrealistic assumption of an equal rate for cell-cell and cell-media diffusion, which impeded the formation of a well-defined inhibitory gradient in the filament.

Just a year after this first study, the group responsible for one of the initial experimental studies (C. P. Wolk, 1967) presented also theoretical results using a simulation code (C. Peter Wolk, 1975). The authors considered that heterocyst placement was defined by a diffusible inhibitor whose concentration dynamic was expressed as:

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2} - k \cdot C(x,t). \tag{2.1}$$

where $C(x,t)$ is the concentration of the diffusible inhibitor at the point x in the time t, D is the diffusion constant and k is the decay rate.

From this equation, the authors obtained an inhibitor diffusion root-mean-square distance for the closed ($k = 0$) and general systems, by considering a discrete approximation with cells as distance units and the inhibitor generated from a point source. These two distances were used as alternative ways to define the range of inhibition that a heterocyst has over neighboring cells in the simulation. This simulation was a sequential random pick of non-inhibited vegetative cells that continues until all cells are inhibited or heterocysts. The solution for the general system is shown to agree with the experimental distribution of distances between heterocysts better than the closed system. However, the closed system produces a slightly more uniform distribution, while presenting much longer intervals than the experimental data. This leads the authors to propose two diffusion-based spacing mechanisms, in which a heterocyst would appear on a cell sufficiently distant from preexisting heterocysts so that it has a concentration of the activator higher than some critical level. At the heart of this work is the initial idea that heterocyst differentiation is a purely stochastically driven process. Then, control is only exerted through desensitization that protects the cells that are close to existent heterocysts against differentiation.

The same diffusive eq. (2.1) was studied in (De Koster and Lindenmayer, 1987) obtaining two different analytical solutions (one continuous and another discrete). These solutions were compared with an improved version of the integer linear cell array simulator discussed previously. This version avoids some problems in (Baker and Herman, 1972) by storing the concentration as a floating-point variable and eliminating the environment with the initialization of the filament already in equilibrium with two heterocysts in the extremes. Through this comparison two biologically reasonable estimations were made: $D = 0.14 - 0.39\ \mu m^2/s$ for the inhibitor diffusion constant and $k = 2.7 - 7.5 \cdot 10^{-4}\ s^{-1}$ for the degrada-

tion rate, and an inaccurate estimation for the cell cycle, 7.25 h, which is known to be around 24 h.

Much later on, the group of researchers of (Allard et al., 2007) proposed a series of models. While the first three models will be discussed here, the last one will be considered in the following section as it includes some genetic interactions. The initial work (Allard et al., 2007) compares the distribution of heterocysts obtained through random placement with one obtained with a model of nitrogen propagating over a filament with a continuous periplasm. In this model, vegetative cells consume nitrogen to grow, while heterocysts produce nitrogen that diffuses through the filament. When the nitrogen level of a vegetative cell reaches 0 the cell irreversibly commits to differentiation. Additionally, the cells grow at a constant rate and divide at a certain fixed division time for each cell. The model is initialized with a couple of heterocysts at the ends of the filaments and a randomly distributed growth rate for each cell. For this model to be able to reproduce the experimental distributions, the authors have to consider an immediate release of nitrogen after commitment in order to avoid the formation of multiple heterocysts. This work presents an opposing paradigm to the earlier ideas in (C. P. Wolk, 1967): while in the oldest work there was a deterministic system of inhibition with a stochastic initiation of differentiation, this work includes a deterministic drive that starts the differentiation to explain *de novo* heterocyst formation. Nevertheless, the need to include a sizable immediate release of nitrogen once a cell is committed to differentiation to avoid the formation of clusters of heterocysts shows that some level of stochasticity is necessary. This stochasticity is represented here by the random distribution of growth rates along the filament. The heterogeneity of growth rates will decide which one of the cells, located in a nitrogen-deprived area, will consume faster its reserves and therefore become a heterocyst. This interplay between deterministic dynamics on a random heterogeneous system seems necessary to recover the observed experimental heterocyst spacing distributions and will be a common trait of most of the models presented below.

This model was expanded in (Brown and Rutenberg, 2012a,b) with the addition of a coupling between the growth and the available nitrogen in the cell and the possibility of nitrogen leakage into the media. Additionally, the commitment condition is also modified and cells have to remain in complete nitrogen deprivation for a set time before they differentiate into heterocysts. This model is capable of reproducing the experimental placement of heterocysts (with a commitment time of 8 hours) considerably better than a random placement and a partially random one where positions adjacent to heterocysts cannot differentiate. Nevertheless, the

assumption that a heterocyst is capable of releasing a sizable amount of fixed nitrogen right after commitment is not biologically feasible and will be substituted by genetic regulation in later work (Brown and Rutenberg, 2014) (described in section **Genetic regulation models**). The authors also obtained a relationship between filament growth rate and heterocyst frequency, and found that growth rate presents a maximum for a certain value of heterocyst frequency (Brown and Rutenberg, 2012a). This maximal growth is similar for different placement strategies if nitrogen leakage is not considered in the model. However, if leakage over 1% is considered, the differences in the growth rate between strategies are relevant: the strategy of differentiation by nitrogen-starved cells, which produced the most realistic heterocyst distributions, is also the one that produces maximal growth (Brown and Rutenberg, 2012a).

Alternatively, (Ishihara et al., 2015) considers a paracrine inhibitory signal originating from the heterocysts instead of considering the nitrogen dynamics of the filament. Experimental data obtained from a mutant strain harboring a P*hetR::gfp* reporter cassette (Asai et al., 2009) presents delayed heterocyst differentiation, observing the first heterocysts at 63–65 hours after nitrogen deprivation instead of the typical 18 to 24 hours (Flores and Herrero, 2010), indicating that the differentiation process is somehow altered in this strain. In their model, the authors continue the idea, first presented in (Baker and Herman, 1972), that cell division and heterocyst differentiation are two competing mechanisms. They propose a cellular automaton model where cells have the capacity of aging, divide and differentiate into heterocysts (that are immediately functional), and dynamics are simulated with a Gillespie algorithm (Gillespie, 1976). The division and differentiation probabilities are represented by sigmoid Hill functions of the cell age. Additionally, the differentiation rate is affected by a lateral inhibition produced by existent heterocysts. This effect decreases as the number of vegetative cells in the source heterocysts increases. The initial condition for the simulation is a filament of a random number of cells with random ages flanked by two heterocysts. The model reproduces the experimental distribution of segments between heterocysts, but not the age distribution of the cells that differentiate. However, it is worth noting that the filaments in which all vegetative cells differentiate into heterocysts before the filament has grown up to 5000 cells are discarded. The model predicts that most cells differentiate at an older age, while experimentally, the differentiation happens at a younger age. From this, it is inferred that the model does not properly capture early pattern formation. To solve this disagreement, *hetR* transcription was studied, observing that it was not immediately perturbed by cell division and remained active at the early stage, concluding that *hetR* activity should be considered independent of cell age. Following this, a model was pre-

sented where differentiation is independent of cell age, obtaining a more realistic age distribution of the commitment time. Finally, both early (defined as 63–65 hours after nitrogen deprivation) and late (more than 69 hours after nitrogen deprivation) differentiation could be explained with the same kinetic parameters, by altering the differentiation dependency with cellular age. Given that the commitment time to differentiation is around 7-8 hours (A. M. Muro-Pastor and Hess, 2012; H. S. Yoon and J. W. Golden, 2001), it is evident that the reporter strain used introduces artifacts, and any conclusion based on its observation has to be taken with extreme caution.

All the models discussed until this point are remarkably capable of reproducing the overall experimental interval length distribution of heterocysts, however, they fail to capture the early pattern formation in the filament. Besides, the initial conditions for almost all these studies are filaments with functional heterocysts in the extremes. Therefore, since inhibitors will only reach the cells close to a heterocyst, only the regions far from these heterocysts, if long enough filaments are considered, would properly reflect *de novo* pattern formation. Additionally, all these models only consider an inhibitory signal originating from the heterocysts without including the well-known competitive lateral inhibition between vegetative cells through PatS (Corrales-Guerrero et al., 2013; Yaru Du et al., 2020; H. S. Yoon and J. W. Golden, 2001). Given that the only selection mechanism acting over the vegetative cells during the first round of differentiation is the initial heterogeneity, the authors are forced to add arbitrary mechanisms to avoid the excessive simultaneous differentiation of contiguous cells. In (Allard et al., 2007; Brown and Rutenberg, 2012a,b) the mechanism is an immediate big release of nitrogen from the heterocysts that stops the differentiation of the close neighbors of committed cells. Alternatively, in (Ishihara et al., 2015) the model is fitted with a strain with an apparent differentiation impairment in which the first round of differentiation appears almost three times later than the typical appearance time.

As a result of these limitations, it seems necessary to include an inhibitory signal originating from vegetative cells in order to fix the heterocyst's pattern. Thus, once this initial pattern is formed, an inhibitory signal originating from the heterocysts, that could be due to the fixed nitrogen(Fogg, 1949; Water and Simon, 1982), to a paracrine inhibitor identified as HetN in (S. M. Callahan and Buikema, 2001) or to a combination of both, could be enough to maintain the preexisting pattern.

2.2.2 *Genetic regulatory models*

The role of the main genes involved in heterocyst differentiation is depicted in fig. 2.4. The differentiation mechanism is initiated by the upregulation of *ntcA* in nitrogen deficiency conditions. This increase of *ntcA* causes an increase of *hetR* that initiates the production of *patS*. This gene codifies a lateral inhibitor that avoids the differentiation of several contiguous cells into heterocysts. Once the cell has already differentiated it starts producing both fixed nitrogen and *hetN* which is another inhibitor of heterocyst formation. *hetR* is the master regulator of the process: in its absence, there is no heterocyst differentiation, consistent with observations in *Cylindrospermopsis*, which is the only Nostocal that lost the ability to develop heterocysts and fix nitrogen.

In (Gerdtzen et al., 2009) a deterministic compartmental model was introduced with three genes represented by a vector, with values in the interval $[0, 1]$, that interact between them through an interaction matrix. The genes considered are *ntcA*, *hetR*, and *patS*. *ntcA* is considered to be activated by nitrogen depletion (Vega-Palas et al., 1992) and, in turn, activates *hetR* (A. M. Muro-Pastor et al., 2002). *hetR* is considered to activate both itself and *patS* (Xu Huang et al., 2004). Finally, *patS* inhibits *hetR* production (Ho Sung Yoon and James W. Golden, 1998). All these interactions are considered to have the same relative strength, except the *hetR* activation of *patS* that is defined to have half of this strength. An representative diagram of the model is included in fig. 2.7.

This model also includes a proxy for *patS* and fixed nitrogen diffusion through a multiplicative factor D^n ($D < 1$). This factor reduces the inhibitory effect of *patS* over the *hetR* expression of a cell located n cells away from the *patS* source. The inhibitory effect of *ntcA* through fixed nitrogen is characterized as an inhibition from *hetR* expression, given that the cells with $hetR = 1$ will be considered heterocysts.

The simulation is initialized from random conditions, and then state transitions are considered to occur asynchronously, with one gene state on a given cell being updated using the interaction matrix at a time in random order for the whole array of cells. After a certain time, the system converges to a patterned filament where $\overline{L_H}$, the average interval between heterocysts (cells expressing all genes in the model at the maximum possible level, 1) depends on the value of the diffusion constant D. Increments of D up to a critical value of 0.7 produce an almost linear increase in $\overline{L_H}$ due to the creation of fewer heterocysts. However, from this point onward, the behavior of $\overline{L_H}$ stops being linear and the system saturates to a state without any heterocyst for $D \geqslant 0.92$.

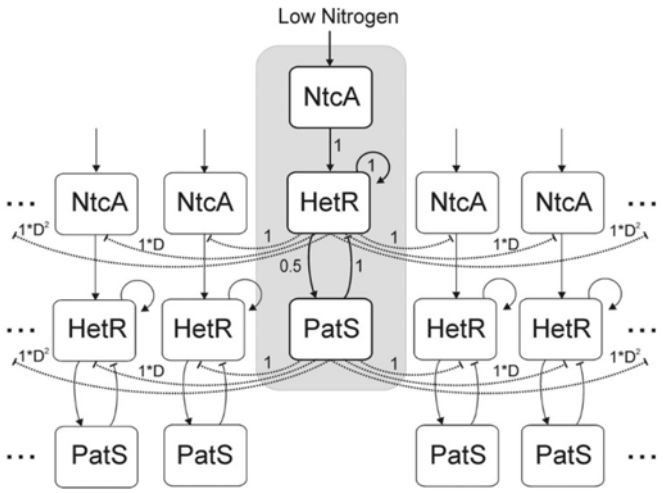

Figure 2.7: **Diagram of the network considered in (Gerdtzen et al., 2009)**. Cells are organized cyclically. Direct interactions are represented by solid lines and indirect interactions by dashed lines. Arrowheads indicate activation and vertical lines indicate inhibition. Numbers indicate the strength of the interactions considered among the elements of the network. Reproduced from (Gerdtzen et al., 2009), CC BY 2.0 license.

The authors set the value of D that produced an $\overline{L_H} = 10 \pm 2$ cells, which is similar to the experimental value observed in (Ho Sung Yoon and James W. Golden, 1998) to study the system. They presented the histograms for intervals between heterocysts in the case of the wild type, the Δ*patS* deletion mutant, and *hetR* over-expressed condition. These results show that, while the means $\overline{L_H}$ are compatible with the experimental data, the simulation produces histograms much more skewed towards larger intervals for both the wild type and the over-expression of *hetR*, and a strictly decreasing distribution of interval length for the Δ*patS* mutant. The first discrepancy could be caused due to the reinterpretation of the fixed nitrogen inhibition of *ntcA* through *hetR*. This change would produce an additional inhibitory signal originating from developing cells instead of only from mature heterocysts as it should be. On the other hand, the discrepancy in the Δ*patS* could be just a problem of interpretation. Given that the expression of the variable *patS* is never shut off, one could argue that this variable is a joint representation of the two main inhibitory genes, *patS* and *hetN*. Then the deletion of this variable should result in the complete differentiation observed in the double Δ*patS*Δ*hetN* mutant (Borthakur et al., 2005), but with the additional artificial inhibition of *hetR* described above. This inhibition partially rescues this mutant because it fulfills the same dual role. The rescue is not full because it targets *ntcA* instead of *hetR*, which reduces its efficiency.

A continuous representation of a linearly growing one-dimensional filament was presented in (M. Zhu et al., 2010). The system of equations that defines its dynamics is:

$$\frac{dr}{dt} = \alpha_r + \beta_r F(r,s) + G(r,s,n) - \kappa_r r \tag{2.2}$$

$$\frac{ds}{dt} = \alpha_s + \beta_s F(r,s) + D_s \frac{\partial^2 s}{\partial x^2} - \kappa_s s \tag{2.3}$$

$$\frac{dn}{dt} = \beta_n F(r,s) + D_n \frac{\partial^2 n}{\partial x^2} - \kappa_n n, \tag{2.4}$$

where r is the concentration of HetR, s of PatS, and n of HetN. The spatial domain, that is, the filament length L, grows at a constant rate ρ:

$$\frac{dL}{dt} = \rho L. \tag{2.5}$$

This model considers linear degradation rates for all the genes ($\kappa_r, \kappa_s, \kappa_n$) and diffusion of the two inhibitors with rates D_s and D_n. Regarding protein production, the authors considered basal production for both HetR (α_r), PatS (α_s), and regulated production for all genes through the function

$$F(r,s) \equiv \frac{r^2}{(K_s + s)(K_r^2 + r^2)}, \tag{2.6}$$

and an additional production term for *hetR*

$$G(r,s,n) \equiv (r_e - r)^2 (n_c - n - \eta s). \tag{2.7}$$

Both eqs. (2.6) and (2.7) include the HetR homodimer formation described in (Xu Huang et al., 2004) through a quadratic *hetR* variable. Equation (2.6) models activation of HetR, PatS, and HetN by HetR dimer, and inhibition by PatS. Equation (2.7) is a phenomenological term affecting HetR: its strength depends on the difference between HetR concentration and an ad hoc level r_e, and its sign is set by the parameter n_c: when the combination of HetN and Pats concentrations given by $n + \eta s$ is larger than n_c, function $G(r,s,n)$ has the effect of degradation, otherwise, it promotes the production of HetR. Through this term, low levels of inhibitors have the effect of an extra activation, that disappears only when the concentration of HetR is r_e. With this model, the goal is to study pattern maintenance; due to this, the initial condition simulates the presence of heterocysts in the borders of the system. This condition is translated into a uniform initial distribution of both HetR and PatS, set to their equilibrium concentration based only on the constitutive production and degradation terms; in the heterocysts, the concentration of HetR is set to the equilibrium value r_e. HetN is initially set to a diffusion-mediated 'bowl-shaped' distribution, with the maxima at the heterocysts. In a way akin to Turing

patterning (Turing, 1952), the apparition of only one heterocyst in the middle is heavily conditioned by the difference in the two inhibitors' diffusive rates. Particularly, the diffusion of HetN should be lower than the filament growth rate, so that there can be HetN depletion in the middle of the filament to induce HetR production. The diffusive rate of PatS must be higher than the one of HetN to reduce the length of the induced region. With these conditions, the model properly reflects the rise of HetR in the middle of the filament that is hypothesized to originate the new heterocyst (Todd A. Black et al., 1993) and the reported inhibitory gradients produced by it (Risser and Sean M. Callahan, 2009).

The low robustness of the pattern to modification of the diffusion parameters is characteristic of Turing-like continuous models. It stems from the requirement that the pattern is an equilibrium state of the overall regulatory system, and consequently the interplay of the two inhibitors must be tuned in such a way that the range of the inhibitors is different enough to create steady spatial differences in gene expression that originate the pattern. However, discrete systems like *Anabaena* filaments can fixate an unstable pattern through the irreversible commitment of a cell that presents a sustained high expression of a given gene, even if that expression is transient and would be reversed without the differentiation. For this reason, in *Anabaena*, the dynamical stability of the pattern is much less relevant than its establishment.

A few years later, (Brown and Rutenberg, 2014) presented the last model of the series discussed in the previous section on diffusion models. In this paper, the authors incorporate a mechanism of genetic inhibition to the nitrogen diffusion model presented in (Brown and Rutenberg, 2012a). This lateral inhibition through *patS* and *hetN* substitutes the immediate release of nitrogen and allows a more biologically realistic maturation of the heterocysts. Both genes are modeled as Boolean variables that directly prevent the commitment to differentiation of a fixed range of contiguous cells. To replicate the experimental observations, the authors set this range to five cells. The expression of both *patS* and *hetN* is in turn modeled as deterministic switches. On one hand, *patS* inhibition starts right after commitment until the complete maturation of the heterocyst (10 hours after commitment), and a time τ_S (set to 1 hour) after this point the heterocyst starts producing fixed nitrogen. On the other hand, *hetN* inhibition starts a certain time τ_N (also set to 1 hour) after commitment and is never shut off.

The initial condition considered is a lonely cell that grows for over 7 days in nitrogen-rich conditions in order to get a heterogeneous filament, which will be put in nitrogen-deprived conditions. The model properly reproduces the vegetative interval histogram tendencies for all

the mutants, but with less noise and without the experimental preference for even-numbered vegetative intervals. Additionally, the authors observe that younger cells are more likely to differentiate, especially on the first round of differentiation (24 hours), suggesting an indirect effect of the cell cycle on heterocyst commitment. This work shows that a deterministic model whose only random variable is the growth rate can reproduce some pattern features observed experimentally.

While the Boolean switch-like genetic model is able to reproduce the experimental mutant behaviors, it does it artificially with an immediate complete inhibition over a fixed range. This is arguably hard to justify experimentally. Despite incorporating both *patS* and *hetN* in the model, their roles are completely equivalent to the immediate release of nitrogen presented in their previous works (Allard et al., 2007; Brown and Rutenberg, 2012a,b). Instead of having distinct roles in the pattern formation, one in the pattern formation and the other in its maintenance, as hypothesized in (S. M. Callahan and Buikema, 2001), they are modeled with the same function (which is to avoid the formation of multiple heterocysts). Additionally, the design of the switch-like dynamics forces the mutant phenotype by providing a window of a duration τ_S (after the cell commitment) and τ_N (after *patS* deactivation) in which there is no inhibition of differentiation in the system for the mutants Δ*patS* and the Δ*hetN*.

A year after this work, a model was presented that used the systems biology framework to study both the stable states of a unicellular system and the pattern formation in a filament (Torres-Sánchez et al., 2015). This model incorporates the nitrogen sensing module of the genetic network with the inclusion of the *nctA* dependence of the GS/GOAT cycle (M. I. Muro-Pastor et al., 2001) and *patS* mediated inhibition. Particularly, *ntcA* production is increased by both HetR and NtcA and inhibited by fixed nitrogen. *hetR* transcription has the same regulation as *ntcA* plus the inhibition from *patS*. Besides, both *patS* and fixed nitrogen are positively regulated by HetR. This model considers the dynamics of fixed nitrogen, NtcA, HetR, and PatS concentrations. The authors use both the biological information of the genetic network and statistical mechanics analysis to obtain the regulatory equations of the system. After obtaining a set of parameters that reproduces heterocyst differentiation, the authors assume that there are two temporal scales: one fast, formed by HetR and ntcA, which relaxes to its steady-state much earlier than the slow one, which is due to the dynamics of PatS and fixed nitrogen. Assuming this, one can use an adiabatic argument and consider that the fast variables are in equilibrium when considering the dynamics of the slow ones.

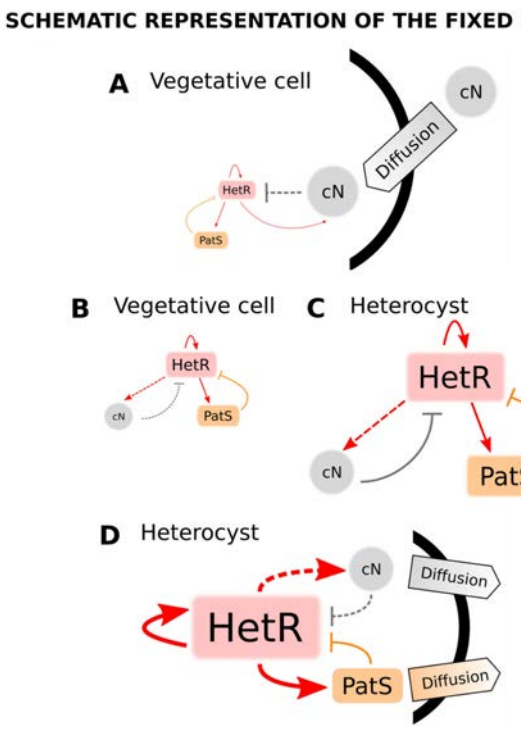

Figure 2.8: **States of a cyanobacterium when subjected to different conditions of nitrogen and diffusion in the model in (Torres-Sánchez et al., 2015).** When combined Nitrogen (cN) is provided to the cell, there is only one stable fixed point (A), which corresponds to a state in which the production of both HetR and PatS is minimum (vegetative state). When subjected to nitrogen deprivation, there are two stable fixed points (B and C). In those, it exists an equilibrium between the production of HetR, PatS, and cN. The one with higher expression (B) corresponds to the heterocyst steady state. When the cell is exposed to nitrogen stress its trajectory evolves from A to the steady state B, and thus it remains vegetative. But if one includes diffusion of cN and PatS from the cell, the only stable state (D) corresponds to a heterocyst state with high levels of production of HetR, cN, and PatS. Adapted from (Torres-Sánchez et al., 2015), CC BY 4.0 license.

A bifurcation analysis for this reduced system is presented, and the equilibrium states are shown in fig. 2.8. For nitrogen-rich conditions, the systems only present one stable solution, which corresponds to the vegetative state (fig. 2.8A: low expression of both *hetR* and *patS*). On the other hand, for nitrogen-deprived conditions, the system presents bistability, with a vegetative stage (fig. 2.8B: with equivalent expressions of *hetR* and *patS*, but higher expression than in nitrogen-rich conditions) and a heterocyst stage (fig. 2.8C: with high expressions of *hetR*, *patS*, and fixed nitrogen). Additionally, the system presents hysteresis, so after nitrogen deprivation, the system will stay in the vegetative state unless a perturbation pushes it into the heterocyst state. Such a perturbation occurs when considering the diffusion of fixed nitrogen and PatS, which is enough to destabilize the vegetative state and push the dynamics to the heterocyst state (fig. 2.8D). The study is later expanded to a discrete filament of cells to show that by adding uniform white noise and diffusion of both PatS and fixed nitrogen, the model is capable to form a patterned differentiation. It is also stated that the appearance of differentiation is considered a pure stochastic event, and also that the biological parameters of the model can be tuned to observe the same pattern with different amplitudes of the white noise.

The high dependency on the noise to differentiate seems to contradict previous works that considered deterministic models of nitrogen-mediated inhibition (Brown and Rutenberg, 2012a, 2014) or even the same regulatory network (Gerdtzen et al., 2009). Those models produced a comparable agreement with the experimental data with much lower relevance of the noise; the stochasticity is only present in the initial conditions of the systems and not in their dynamics. In our opinion, this discrepancy can be attributed to the overlapping of the roles of both PatS and fixed nitrogen, which saturates the system with inhibitors that stop the increase in HetR production. Particularly, given that the filament model is able to produce fixed nitrogen at a low level once HetR concentration rises in vegetative cells, there is no need for a transition to the high nitrogen production cellular state (heterocyst). Instead, the biological system is not capable of fixing nitrogen until the transition has already occurred, given that nitrogen fixation cannot coexist with photosynthesis. Therefore, the activation of *hetR* transcription through *ntcA* cannot be shut off until nitrogen is provided to the system. Due to this, the system is forced to maintain the individually unstable state of high expression of both *hetR* and *patS* until a heterocyst is formed in the filament. Additionally, recent experimental works (Corrales-Guerrero et al., 2013; Higa et al., 2012; Rivers et al., 2014, 2018) seem to indicate that both *patS* and *hetN* require a post-transcriptional modification to produce the inhibitory peptide. It has been suggested (Corrales-Guerrero et al., 2014b; Rivers et al., 2018) that this transformation occurs at the cell membrane during cell-to-cell trafficking. This would avoid self-inhibition from the *patS* and *hetN* produced in a given cell and, therefore, would make impossible a unicellular stability study.

A similar system's biology approach was considered in (Muñoz-García and Ares, 2016), where an alternative three-gene minimal model was presented fig. 2.9. This model also included both *hetR* as the main non-diffusive regulator of the system and *patS* as an inhibitor of HetR-mediated activation. Instead of the nitrogen sensing module, the model included *hetN* as an inhibitor produced in the heterocysts. Then *hetR* activates both *patS* and its own expression, while *patS* and *hetN* (which is produced at a basal level in the heterocysts) inhibit this activation. Fixed nitrogen is included as a direct inhibitor of HetR regulation. As a substitute for the *ntcA* role as the trigger of *hetR* expression, the model includes a low basal expression of *hetR*. Using mass-action kinetics, the authors obtained a deterministic set of differential equations from the mechanistic information of these interactions.

The model considers that, while HetR needs to form a homodimer to promote expression, this activation can also be inhibited with the at-

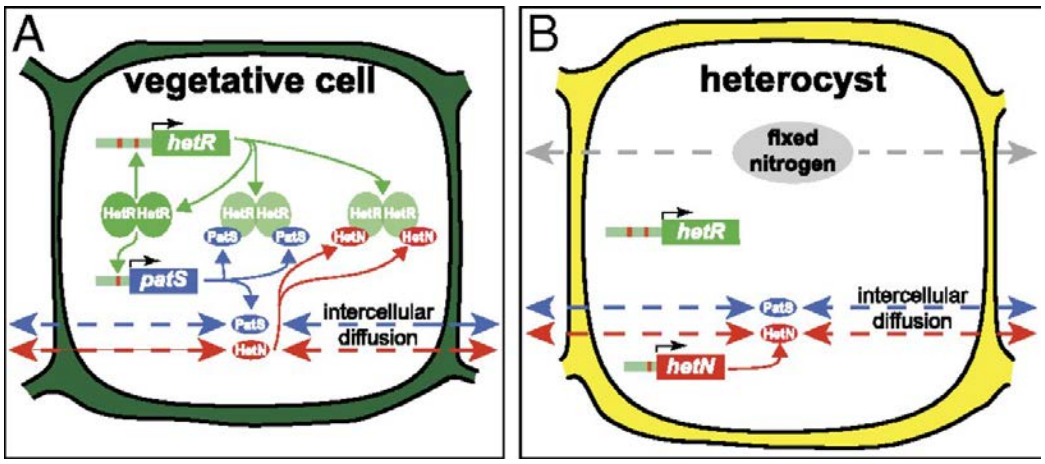

Figure 2.9: **Minimal model of the genetic network in (Muñoz-García and Ares, 2016).** (A) In vegetative cells, HetR dimers can activate the expression of hetR and patS. (B) In heterocysts, hetN is expressed constitutively and fixed-nitrogen products are produced. Active products of PatS and HetN, possibly the RGSGR pentapeptide, can diffuse between cells of any kind in the filament and bind HetR, preventing it from binding DNA. Fixed-nitrogen products also diffuse to other cells and contribute weakly to inhibit differentiation. Image reproduced from (Muñoz-García and Ares, 2016)

tachment of just one inhibitor. The stochastic nature of gene expression was considered, adding noise to the equations using Langevin dynamics (Gillespie, 2000). This genetic model was introduced in a agent based simulation of a filament with inhibitor diffusion where each cell has its own noisy dynamical variables, growth rate and thresholds for both differentiation and cell division. The model was able to reproduce the experimental distribution of vegetative intervals between heterocysts up to the third moment of the distribution, for both the wild type and the Δ*patS* mutant, and gives a reasonable prediction for the Δ*hetN* mutant, for which it made no comparison with experimental data.

The phenotypical reproduction by the model of the deletion mutants reinforces the role of the two inhibitory genes proposed in (S. M. Callahan and Buikema, 2001). This model also provides additional insight into the interplay between cell division and heterocyst differentiation. Due to the similar timescale between these two processes, the noise on the cell division defines the overall behavior of the filament. If there is low noise and cells divide in a quasi-synchronous way, the filament pattern has an oscillatory behavior with enlargement and posterior shortening of the mean distance between heterocysts. In this low noise regime, the model also recovers the larger appearance of even-numbered vegetative cell intervals characteristic of heterocyst patterns (John C. Meeks and Elhai, 2002). In-

stead, for a noisier cell division, the percentage of even intervals remains always close to 50% and the oscillatory behavior of the mean vegetative cell's interval disappears.

Following the Turing-like characterization in (Xu Huang et al., 2004), (Di Patti et al., 2018) presents the same three gene system (*hetR* as an activator and both *patS* and *hetN* as diffusible inhibitors) but, in this case, the inhibitory effect is produced through degradation of HetR dimers mediated by PatS and HetN. The model assumes a basal production in all cells and a linear degradation for the three genes, and an increase in the production of both HetR and PatS activated by HetR. As the model does not enforce any distinction between vegetative and heterocyst cells, all cells actively produce both inhibitors simultaneously. Thus, it does not reflect the temporal differences in the onset of production of PatS and HetN. A bit surprisingly, the same work presents very nice experimental evidence of this difference through GFP reporters of transcription.

With this set of interactions, the authors obtain a set of differential equations through the van Kampen expansion. Initially, the authors study the linear stability around the homogeneous fixed point of the mean field approximation. Through this analysis, the authors located a set of parameters that allow the formation of instabilities that could originate a pattern in the mean-field conditions. As already mentioned, during the analysis of (Xu Huang et al., 2004), the most relevant parameters are the diffusion constants for both inhibitors. The smaller the ratio $\frac{D_S}{D_N}$ is, the narrower the instability region. Subsequently, the authors introduce the same interactions in a Gillespie algorithm with the same set of parameters to check how stochasticity affects pattern formation. The authors show that the presence of noise promotes the spontaneous selection of a leading wavelength in the emerging pattern. Due to this, the parameter region where the system presents instabilities, and therefore pattern seeding, is considerably larger in the noisy system. The addition of filament growth (cellular division) to the model increases the amount of available unstable modes of the system. Despite this similarity, the patterning is much more stable in the system with deterministic growth than in the noisy one. Without noise, a new high HetR expression region (heterocyst) appears in the midsection of the existing pattern when the filament elongates enough. On the other hand, the addition of growth to the noisy system destabilizes the patterning and allows for the transition between the different unstable modes that arise with the filament growth. This implies that while the pattern formation is enhanced by the addition of noise to the system, its maintenance in a growing domain requires an irreversible fixation of the heterocyst state. This result is in clear agreement with the theoretical result on (Marciniak-Czochra et al., 2017) that stated

that Turing systems with stationary activators cannot autonomously fix the dynamical pattern. Additionally, this model shows that it is possible to form the pattern through the regulation of degradation instead of the previously studied regulator inhibition. Nevertheless, it should be tested if this alternative inhibition through HetR degradation reproduces the experimental data for a model with a more realistic temporal separation between the two inhibitors.

After this systematic analysis of the existent genetic models, one can extract some common key ideas. First, the realization that several configurations of a minimal three-gene network with an activator, *hetR*, and a couple of inhibitors, *patS* and typically *hetN*, but it could also be the fixed nitrogen through *ntcA* regulation, as in (Gerdtzen et al., 2009), are capable of reproducing the wild type behavior. Due to this, it seems indispensable to consider other phenotypes, especially the deletion ones, in order to properly evaluate the regulatory mechanisms proposed. Another clear common argument is that there must be a certain temporal separation between the inhibitory effects to originate a pattern. This difference could be either produced due to the relationship between the diffusion coefficients (Di Patti et al., 2018; M. Zhu et al., 2010). Directly imposed, as observed in (Brown and Rutenberg, 2014) and also (Muñoz-García and Ares, 2016) (where *hetN* is exclusively produced in heterocysts). Or alternatively, through the use of fixed nitrogen (Gerdtzen et al., 2009; Torres-Sánchez et al., 2015) acting indirectly through *ntcA* and therefore presenting a certain delay.

2.2.3 *Cyanobacteria population models*

An alternative point of view to the study of spatial pattern formation is to consider the cyanobacteria culture as a population problem where the percentage of each cell type is defined by external conditions.

This approach is used in (Hense and Beckmann, 2006), presenting a deterministic model of the life cycle of cyanobacteria dependent on energy, mainly in the form of light and nitrogen availability. In this formulation, the heterocyst would be the stage with high energy (abundant light) and low nitrogen availability. The model is capable of reproducing the seasonal changes in the cyanobacteria population composition and infers a correlation between summer blooms and cycle velocity, where previous summer conditions strongly affect the possibility of explosive growth. The scope of this work is mostly ecological and does not provide extensive insight into the mechanisms controlling the vegetative-heterocyst transition.

Alternatively, (Pinzon and Ju, 2006) takes the same culture level population approach but with a more biomolecular focus on the cellular processes that modulate the transition from a vegetative cell to a heterocyst. The deterministic model proposed includes the photosynthetic growth of vegetative cells, heterocyst differentiation, self-shading effect on light penetration, and nitrogen fixation. The authors hypothesize that heterocyst differentiation is driven by the difference between the required fixed nitrogen to support maximal growth and the available nitrogen. The model describes experimental profiles well and gives reasonable predictions even for the transition from growth over external nitrogen sources to self-sustained growth.

This population point of view was taken again later in (Grover et al., 2019). In this work, the authors present a deterministic model where the transition between vegetative and heterocyst cells is controlled by the relationship between the processed and free concentration of both nitrogen and phosphorus in the cells. The model predicts a relationship between the heterocyst-to-vegetative ratio with the nitrogen-to-phosphorus ratio of the environment. The authors use this to discuss an evolutionary reason for the regulation of heterocyst differentiation. Given that phosphorus-limited habitats are much more common than nitrogen-limited ones, the costly investments in nitrogen fixation are tightly regulated.

As one can see, this kind of the mean-field point of view is more useful for an ecological and evolutionary perspective but does not provide much insight into the regulation of the heterocyst differentiation. The patterned differentiation of heterocysts seems relevant to the mechanism controlling the differentiation decision, therefore the population point of view is less optimal because the pattern information is lost.

2.2.4 *Modeling overview*

In the section about inhibitor diffusion models, we have discussed examples of models where diffusible inhibitory signal produced in the heterocysts is enough to maintain an existing pattern in a filament. If one considers the genetic regulatory system (fig. 2.4) it is easy to see that the role of this inhibitory signal originating from the heterocysts would be taken by HetN and fixed nitrogen. HetN acts directly over HetR and fixed nitrogen indirectly through *ntcA*. With this minimal structure, a new heterocyst would arise in the space between heterocysts roughly when the interval doubles its length. Then, as observed when discussing genetic regulatory models, if one also considers *patS*, which is a lateral inhibitor expressed in vegetative cells, the system is capable to create *de novo* pattern formation.

This regulatory system (fig. 2.10) would be coupled with a switch-like genetic mechanism that initiates differentiation when the HetR concentration is higher than a certain threshold. This three-gene system of an activator and two inhibitors could seem a Turing pattern, but it presents a key difference, one of the inhibitors *hetN* has its production restricted to the heterocysts. Moreover, this differentiation to heterocysts entails a morphological change and therefore is irreversible. This ensures the stability of the pattern that would not be possible in a Turing system.

Different strategies to model these three genes can simulate a heterocyst pattern, so more biological information is necessary to properly define the differentiation mechanism. A powerful tool is constraining models by comparison with different genetic backgrounds. In the last section, section 2.1 we show a simplified regulatory network of heterocyst differentiation obtained from surveying experimental literature fig. 2.5. From it, it is evident that the mechanism controlling heterocyst differentiation is quite more complex than any model discussed in this work. Then, by incorporating more genes into the models one would, on one hand, deepen

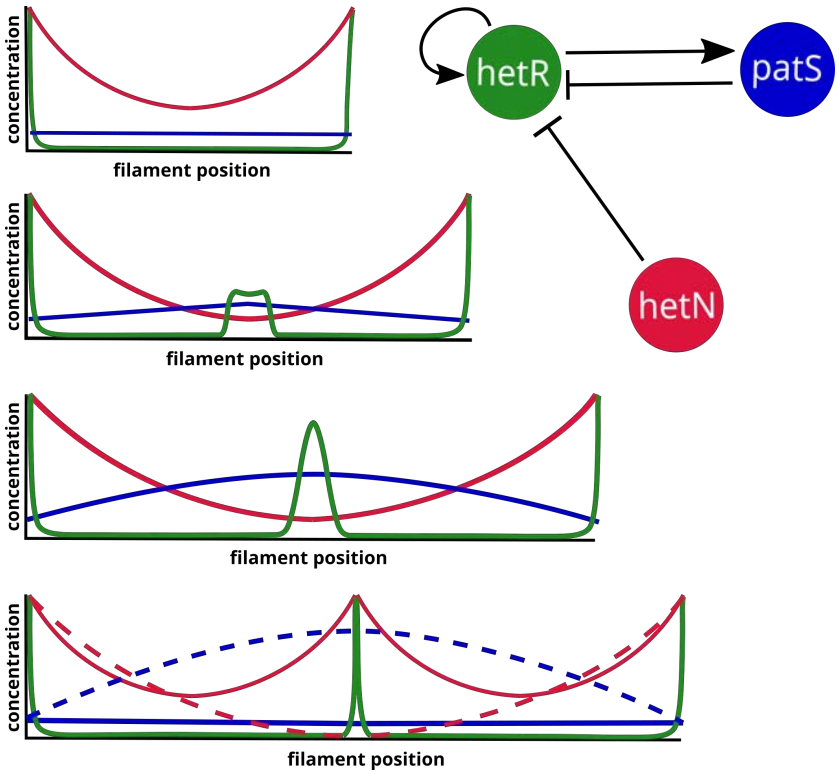

Figure 2.10: Three genes, simplified regulatory system and expected protein concentration dynamics for the emergence of a new intercalary heterocyst in a fully deterministic system. The heterocysts are represented as peaks of HetR that produce hetN. The dashed profiles represent the state right before the transition to a heterocyst.

the understanding of the regulatory network, and on the other hand, open the possibility to compare with a wider range of genetic backgrounds. Therefore, the way forward is to incorporate into model genes that still have more dubious roles but with enough experimental information, such as *patA* and *hetF*. The research of the genes represented in light red rectangles in fig. 2.5 is still quite brief and there is not enough information to justify their inclusion in models. There is evidence that both *hetC* and *patN* are connected to *patA* regulation, but there is not enough information to assign a proper role to them. On the other hand, the function of *hetL* seems quite clear: it appears to be involved with HetR activation, but there is no clear link to *hetF* and other genes in the system, besides *hetR*. Also with an apparently clear function but without a clear relation with the other genes are both *hetP* and *hetZ*, which are heavily linked to the heterocyst commitment but without a clear explanation on how they affect the commitment.

This patched information is natural given that, as we have already shown in the previous section 2.1, the first experimental evidence is usually the effects of the knock-out mutants of these genes over the known network. Posterior studies that provide experimental information regarding protein translation, such as (Corrales-Guerrero et al., 2015; Di Patti et al., 2018), could be really useful to properly characterize such genes in a mathematical model. The modeling of putative mechanisms for these genes is also a useful source of predictions, helping to focus on what experiments should look for. As an example of this, our new model that includes *patA* and *hetF*, which will be presented in chapter 4 predicts that inhibitor leakage at the filament extremes is necessary to explain the *patA* phenotype (Casanova-Ferrer et al., 2022). This expansion of the genetic scope of the models would bridge the gap that now exists between this kind of reduced model focusing on heterocyst differentiation and the more general genome-scale frameworks such as (Malatinszky et al., 2017), which model the full metabolism of the *Anabaena* cell and how it changes after the differentiation into a heterocyst.

3

STABILITY STUDY OF THE *ANABAENA* MINIMAL MODEL

In this chapter, we will expand the modeling perspective presented in (Muñoz-García and Ares, 2016) that only considered a minimal three-gene system (represented in red in fig. 3.1) to reproduce both the pattern formation and maintenance. We will study the stability of our model by obtaining the bifurcation diagrams for several parameters of the model. But before that we will expand a bit on the core biology of the system, to properly argue the modeling choices.

It has been shown that, when nitrogen stress is perceived, the transcription regulator *ntcA* is important to trigger heterocyst differentiation (Herrero et al., 2004; C. C. Zhang et al., 2006), by directly or indirectly controlling the expression of several genes including *hetR* (Valladares et al., 2008).

The expressions of *ntcA* and *hetR* are mutually dependent, although the latter seems to be sufficient for heterocyst development (Buikema and Haselkorn, 1991). Thus, positive auto-regulation of *hetR* is required for differentiation and is particularly significant in developing heterocysts (Todd A. Black et al., 1993; Rajagopalan and Sean M. Callahan, 2010). *hetR* expression is the main positive regulatory factor in heterocyst development, (Todd A. Black et al., 1993; Buikema and Haselkorn, 1991) and this

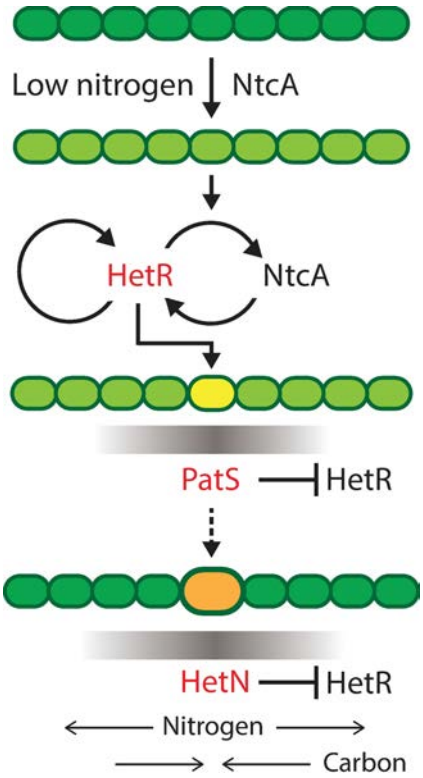

Figure 3.1: **Minimal mechanism of heterocyst differentiation:** The scheme represents the process of differentiation of a filament of heterocyst-forming cyanobacteria, with the roles of the main genes involved. Darker green means more fixed nitrogen in a cell. Yellow marks a cell committing to differentiation, and orange is a functional heterocyst. Reproduced from (Di Patti et al., 2018), CC BY 4.0 license.

gene is epistatic to the others involved in heterocyst differentiation (Orozco et al., 2006). These two genes constitute the main activation hub of heterocyst differentiation, and their dynamics are equivalent while in nitrogen deprivation.

The gene *patS* is a negative regulator of *hetR*, suppressing differentiation when over expressed and inducing multiple contiguous heterocysts, the so-called Mch phenotype, when deleted (H. S. Yoon and J. W. Golden, 2001; Ho Sung Yoon and James W. Golden, 1998). The expression of *patS* produces a short peptide PatS, predicted to be formed by 13 or 17 amino acids, which contains a carboxyl-terminal that prevents DNA binding activity of HetR (Feldmann et al., 2012; Xu Huang et al., 2004) and inhibits differentiation when added to culture medium (Ho Sung Yoon and James W. Golden, 1998). The expression of *patS* in small groups of cells was shown to diminish the levels of HetR in adjacent cells (Risser and Sean M. Callahan, 2007), suggesting that a PatS-dependent signal can be trafficked along the filament (Corrales-Guerrero et al., 2013).

Although lack of *patS* expression initially produces a pattern with groups of contiguous heterocysts and short intervals of vegetative cells between those clusters, this pattern tends to recover the characteristics of a wild type-like pattern later on (H. S. Yoon and J. W. Golden, 2001). This suggests the presence of additional patterning signals operating long after nitrogen deprivation. The most relevant player that leads to this late inhibitory effect is the *hetN* gene, expressed only in heterocysts. Similarly to *patS*, the product codified by *hetN* also contains an ERGSGR sequence of amino acids (Glutamate-Arginine-Glycine-Serine-Glycine-Arginine), raising the possibility that an ERGSGR-containing peptide derived from the full protein goes from cell to cell (Higa et al., 2012; Rivers et al., 2018). However, in contrast to the Δ*patS* mutant phenotype, the Δ*hetN* mutant phenotype has a heterocyst pattern similar to the wild type at the initial stages of nitrogen depletion and a multiple contiguous heterocyst phenotype after 48 hours (S. M. Callahan and Buikema, 2001). This indicates that *hetN* expression is activated later than that of *patS*. Additional proof of the inhibitory function of *patS* and *hetN* is that, when both genes are suppressed, almost all cells eventually differentiate, causing lethal levels of heterocysts (Borthakur et al., 2005). For these reasons, we have included both inhibitors in our model. But, given the new evidence presented in (Rivers et al., 2018), we have chosen to consider a unique inhibitory element produced from both *patS* and *hetN* during their extracellular export.

3.1 MODEL DEDUCTION AND ADJUSTMENT

From the biological information presented and using the ideas and procedures introduced in section 1.1, we can obtain a system of differential equations that reproduces the system behavior. The full model deduction is presented in appendix A, here we only present the final set of equations:

$$\frac{dR_j(t)}{dt} = \beta_r + \phi(r_j,i_j,F_j)\rho_r - \alpha_r r(R_j,i_j) - \frac{r(R_j,i_j)^2}{K_b}\left[\alpha_d + \alpha_{Id}\frac{i_j^2}{K_d^2}\right] \tag{3.1a}$$

$$\frac{dp_j(t)}{dt} = (1-\delta_{hc,j})\phi(r_j,i_j,F_j)\rho_p - 2d_p p_j - \alpha_p p_j \tag{3.1b}$$

$$\frac{dn_j(t)}{dt} = \delta_{hc,j}\rho_n - 2d_n n_j - \alpha_n n_j \tag{3.1c}$$

$$\frac{di_j(t)}{dt} = d_p(p_{j-1}+p_{j+1}) + d_n(n_{j-1}+n_{j+1}) + d_i(i_{j-1}-2i_j+i_{j+1}) - \alpha_i i_j \tag{3.1d}$$

$$\frac{dF_j(t)}{dt} = \delta_{hc,j}\rho_F - \alpha_F F_j + d_F(F_{j-1}-2F_j+F_{j+1}) \tag{3.1e}$$

$$\phi(r_j,i_j,F_j)^* = \frac{\frac{r(R_j,i_j)^2}{K_b K_r}}{1+\frac{r(R_j,i_j)^2}{K_b K_r}+\frac{F_j(t)}{K_d}}, \tag{3.2a}$$

$$r(R_j,i_j) = \frac{-K_b K_d^2 + \sqrt{K_b^2 K_d^4 + 8K_b K_d^2 R_j\left(K_d^2+i_j^2\right)}}{4\left(K_d^2+i_j^2\right)}, \tag{3.2b}$$

* *To be able to reproduce the HetN deletion mutant we add, phenomenology, a lineal inhibition from the fixed nitrogen products (F), if we do not add this term the deletion mutant of hetN shows a completely differentiated filament due to the fact that we do not have considered the nitrogen sensing system (which is mediated through ntcA)*

where the subindex j indicates that the variable refers to the cell number j in the filament (being then $j-1$ and $j+1$ its neighboring cells) and t denotes time. The variables $R_j(t)$, $p_j(t)$, $n_j(t)$, $i_j(t)$, $F_j(t)$ are the concentration of HetR, PatS, HetN, diffusible pentapeptide, and fixed nitrogen respectively. And $\delta_{hc,j}$ specifies whether the cell is vegetative or a heterocyst: its value is 1 if the cell j is a heterocyst, 0 if it is vegetative. The rest of the parameter's biological description is presented in table 3.1 together with its value. In addition, we define the boundary condition that if $j+1$ or $j-1$ is outside the filament, then $X_{j+1}=0$ or $X_{j-1}=0$, with X representing an arbitrary variable.

Once we have the complete model of our system, studying the stability of this system according to the parameters would be as easy as analytically resolving the stationary state, annulling all the temporal derivatives, and observing how the stability of the solutions is altered in relation to the value of the parameters. But this straightforward method is only possible for systems with much simpler temporal evolution, such as those

Table 3.1: wild type parameters of the model.

Parameter	Description	Value	Units
b_r	HetR basal production rate	40	nM h^{-1}
ρ_r	HetR maximum regulated production rate	1000	nM h^{-1}
α_r	HetR monomer degradation rate	4	h^{-1}
α_d	HetR free dimer degradation rate	0.25	h^{-1}
α_{Id}	HetR inhibited dimer degradation rate	0.5	h^{-1}
ρ_p	PatS maximum regulated production rate	4000	nM h^{-1}
α_p	PatS degradation rate	2	h^{-1}
d_p	PatS-Inhibitor conversion rate	10	h^{-1}
ρ_n	HetN production rate in heterocysts	10000	nM h^{-1}
α_n	HetN degradation rate	2	h^{-1}
d_n	HetN-Inhibitor conversion rate	5	h^{-1}
d_i	Inhibitor diffusion rate	25	h^{-1}
α_i	Inhibitor degradation rate	2	h^{-1}
ρ_F	fixed nitrogen production rate in heterocysts	8000	nM h^{-1}
α_F	fixed nitrogen degradation rate	4	h^{-1}
d_f	fixed nitrogen diffusion rate	100	h^{-1}
K_r	HetR-Promoter equilibrium constant	100	nM
K_b	Monomer binding equilibrium constant	75	nM
K_d	Affinity of RGSGR and fixed nitrogen to HetR	227 (Feldmann et al., 2011)	nM
ρ	Cellular growth rate	0.06	$\mu m\ h^{-1}$
σ	Variance of noise in cellular growth	0.2	$\mu m\ h^{-1}$

studied in section 1.1. As our case is much more complex than those few examples, we need to solve numerically our system of equations and in order to do so we need to find a set of parameters that adequately reproduce the experimental results. In order to adjust this new model, we modified the C++ simulation code from (Muñoz-García and Ares, 2016), as the final systems are quite similar, and we wanted to maintain our parameters as close as possible to the ones presented in that study. Then, through manual changes from their set of parameters, we have obtained the adjustment with experimental data presented in (Corrales-Guerrero et al., 2014a).

As one can see while the means adjust quite well to the experimental data, the only serious discrepancy comes with the last experimental value for the Δ*patS* deletion mutant[5]. But this point seems quite conflictive because while the experimental data show a clear reduction of the spacing between heterocysts.

[5] *Which means that is a strand without a* patS *or with a faulty copy that has been rendered unusable. In our system is equivalent to* $\rho_p = 0$

Other experimental results such as (H. S. Yoon and J. W. Golden, 2001) affirm textually that *"A patS mutant grown for several days under nitrogen-fixing conditions showed partial restoration of the normal heterocyst pattern,*

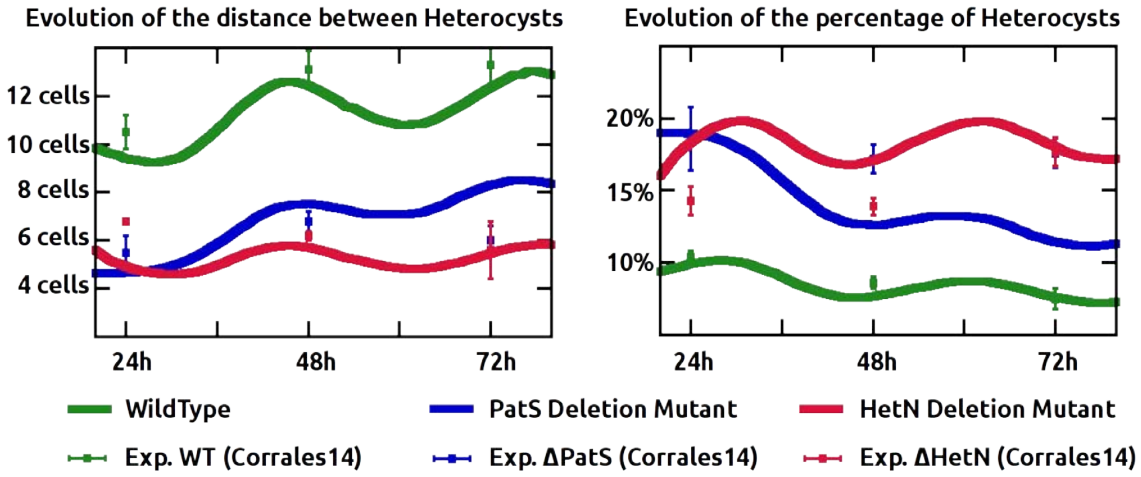

Figure 3.2: Mean number of vegetative cells between heterocysts and the total percentage of heterocysts in the filament. Symbols represent experimental values from (Corrales-Guerrero et al., 2014a) and lines are simulation results.

presumably because of a gradient of nitrogen compounds supplied by the heterocysts"(direct quotation from H. S. Yoon and J. W. Golden, 2001).

The percentages of heterocyst do not present an agreement as good as the means. This happens because the amount of contiguous heterocysts is lower in the model than in the experimental data. But, given that we are going to focus on the pattern formation for this study, we consider that is a good enough fit. Additionally, the phenotypes observed during the different mutants simulations, fig. 3.3, closely follow the ones described in the bibliography for those two mutants.

It is important to realize that even after obtaining all the parameters, this is still a Many-Body complex problem. We would have this set of six equations (eqs. (3.1)) for every considered cell. Then the size of our system will increase by 5-fold, the number of new cells in the system. Due to this, in order to be able to obtain the bifurcation diagrams, we will have to limit the number of cells in the system. In this case, we will reduce the system to two cells, which is the minimum required to observe heterogeneity in HetR concentration as an early stage of pattern formation in the filament. As we are interested in the pattern inception, we will not consider either filament growth or actual differentiation in the system. Then we can safely ignore both the HetN and the fixed nitrogen. It is also worth noting that this focus on the early pattern inception is the reason for not further simplifying the model. We could consider, as it was previously done in (Muñoz-García and Ares, 2016), that the dimerization equilibrium is faster than the inhibition. This consideration would elimi-

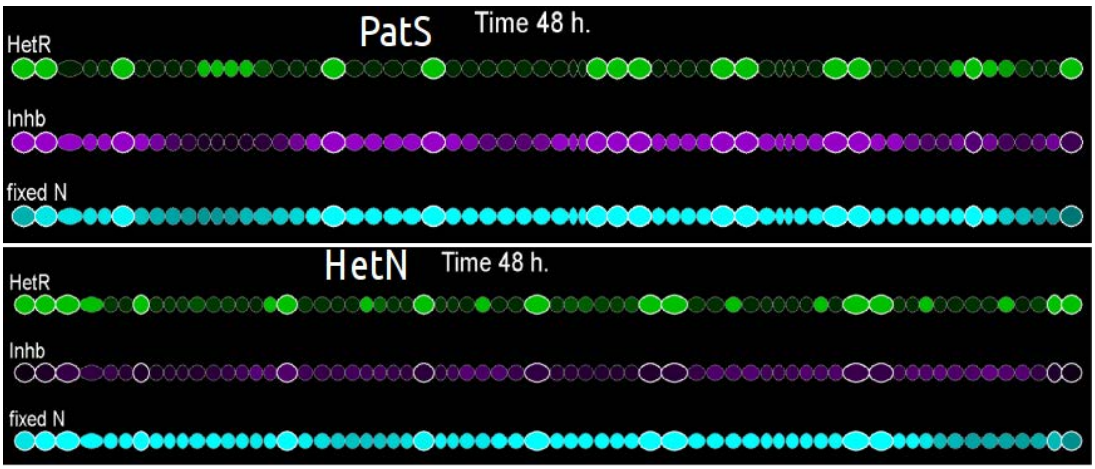

Figure 3.3: Example of the the Δ*patS* and Δ*hetN* Mch phenotypes observed in our simulation. In the Δ*patS* mutant, the multiple heterocysts are formed simultaneously. On the other hand, in the Δ*patS* mutant the multiple heterocysts appear consecutively.

nate the competition between these two and therefore eq. (3.2a). But, as we have already shown in section 1.1, it would reduce the interpretability of the model by combining several biological processes in the same parameter. As we will discuss particular changes to the parameters, and we are reducing the scale of the system, we have chosen to keep the parameters as mechanistic as possible. For the next chapters, we will instead see that with that consideration of sequential interactions, these systems can be further simplified.

With all this in mind, can reduce our model to obtain the following equations for a two cells system:

$$\frac{dR_1(t)}{dt} = \beta_r + \rho_r g_1(r_1,i) - \alpha_r r_1 - \frac{r_1^2}{K_b}\left[\alpha_d + \alpha_{Id}\frac{i_1^2}{K_d^2}\right] \tag{3.3a}$$

$$\frac{dp_1(t)}{dt} = \rho_p g_1(r_1,i_1) - d_p p_1 - \alpha_p p_1 \tag{3.3b}$$

$$\frac{di_1(t)}{dt} = d_p p_2 + d_i(-i_1 + i_2) - \alpha_i i_1 \tag{3.3c}$$

$$\frac{dR_2(t)}{dt} = \beta_r + \rho_r g_2(r_2,i_2) - \alpha_r r_2 - \frac{r_2^2}{K_b}\left[\alpha_d + \alpha_{Id}\frac{i_2^2}{K_d^2}\right] \tag{3.3d}$$

$$\frac{dp_2(t)}{dt} = \rho_p g_2(r_2,i_2) - d_p p_2 - \alpha_p p_2 \tag{3.3e}$$

$$\frac{di_2(t)}{dt} = d_p p_1 + d_i(-i_2 + i_1) - \alpha_i i_2 \tag{3.3f}$$

$$r_1(R_1,i_1) = \frac{-K_b K_d^2 + \sqrt{K_b^2 K_d^4 + 8K_b K_d^2 R_1 (K_d^2 + i_1^2)}}{4(K_d^2 + i_1^2)}, \tag{3.4a}$$

$$g_1(r_1,i_1) = \frac{r(R_1,i_1)^2}{K_b K_r + r(R_1,i_1)^2}, \tag{3.4b}$$

$$r_2(R_2,i_2) = \frac{-K_b K_d^2 + \sqrt{K_b^2 K_d^4 + 8K_b K_d^2 R_2 (K_d^2 + i_2^2)}}{4(K_d^2 + i_2^2)}, \tag{3.4c}$$

$$g_2(r_2,i_2) = \frac{r(R_2,i_2)^2}{K_b K_r + r(R_2,i_2)^2}. \tag{3.4d}$$

3.2 STABILITY STUDY OF THE 2 CELL CASE

Now we could directly introduce this reduced system (eqs. (3.3)) in XPP-AUTO (an open-source numerical integrator software) to obtain bifurcation diagrams. But our main objective is to observe if there is pattern formation on the filament for a given value of the parameters. In our two cells system, this is equivalent to studying if the solutions of the HetR concentration for the two cells are homogeneous or not. Then, if they are homogeneous the system would be in a non-pattern state while if they are heterogeneous the system would theoretically form a pattern over the filament. As our objective is to evaluate the relative values of the two cells, it would be much easier to consider both the difference and the addition of the two cell concentrations rather than their individual values:

$$\Delta X \equiv X_1 - X_2,$$
$$\sum X \equiv X_1 + X_2,$$

and therefore

$$X_1 = \frac{\sum X - \Delta X}{2}, \tag{3.5a}$$

$$X_2 = \frac{\Delta X - \sum X}{2}. \tag{3.5b}$$

If we apply this change of variables eqs. (3.5) over eqs. (3.3). With this change and setting the correct configuration of the Runge-Kutta from XPP-AUTO we obtain the following results for our first parameter β_r (fig. 3.4):
These two first graphics will be used as examples for all the ones to come. In this plot, fig. 3.4, we have not been cut through the black dotted lines that mark that do not make sense physically to show the whole diagram

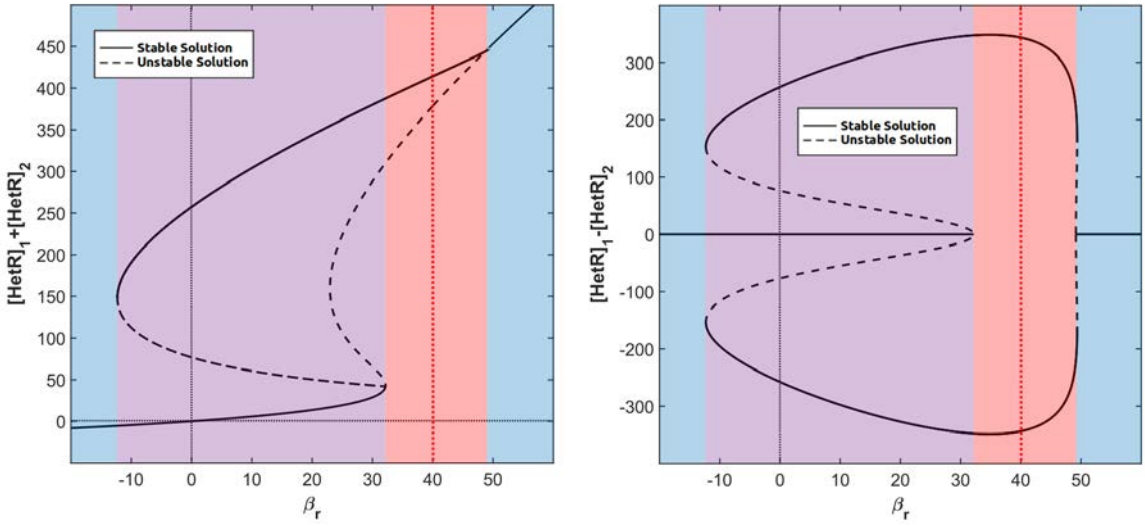

Figure 3.4: Bifurcation diagram of the sum and difference of HetR concentration for the parameter of basal production β_r. To ease the visualization, the homogeneous regime is colored in blue, the bistable one in purple, and finally the heterogeneous one in red. We will also show the value of the WT parameters, table 3.1, with a red dotted line.

provided by XPP-AUTO. We also present in fig. 3.4 both the ΔX and the $\sum X$ diagrams, while from now on we will only present the ΔX as it is the most useful to study the behavior of the system. These diagrams present the stable solutions with a solid black line, while the unstable ones will be plotted using dashed lines. In addition, it also represented the regime of the stable solutions for each value of the parameter: blue background for the stable homogeneous solution, purple for the coexistence between homogeneous and heterogeneous solutions, and red for the heterogeneous solutions. As one would expect, the value for the WT parameter (that forms patterns in the complete model simulations) is always in the heterogeneous region.

Now we will present the bifurcation diagrams for all the reduced model parameters[6] obtained, maintaining all the other parameters in the wild type values. We will comment on similarities and differences between them once we have shown them all. As we are presenting, $|\Delta[HetR]|$ we have to consider that all the heterogeneous solutions are double, as can be observed in fig. 3.4 right.

[6] *Except* d_i *whose integration never converged for some reason that I do not really know*

Most of the bifurcation diagrams, as the first two on fig. 3.5 (with the two parameters that control HetR production), present 2 bistable zones (the second one in β_r is barely visible) and therefore two hysteresis zones like the ones observed in the (Torres-Sánchez et al., 2015) model. That

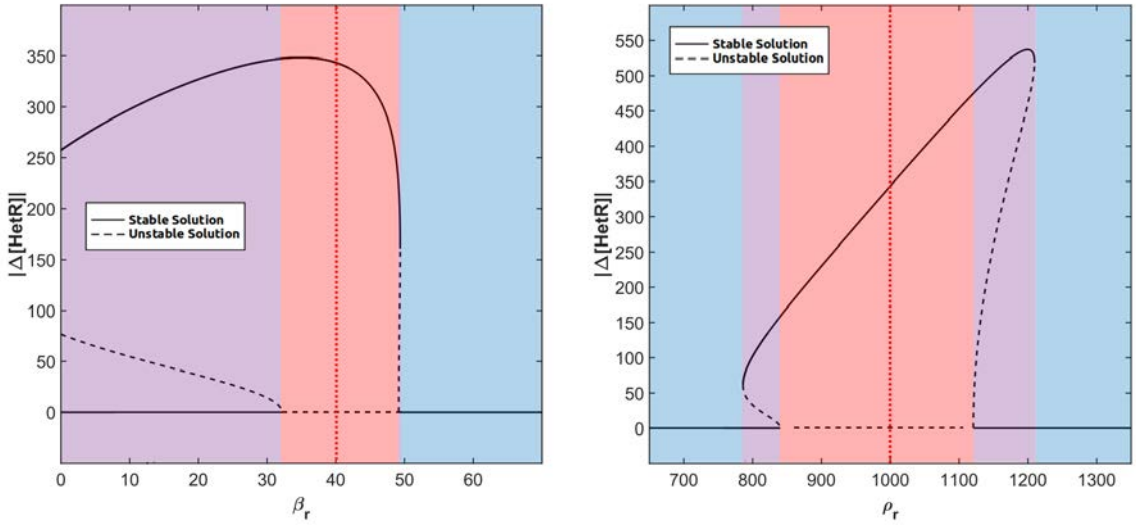

Figure 3.5: Bifurcation diagram for the parameters of HetR production (β_r and ρ_r). The homogeneous regime is colored in blue, the bistable one in purple, and finally the heterogeneous one in red. The WT value is presented with a red dotted line.

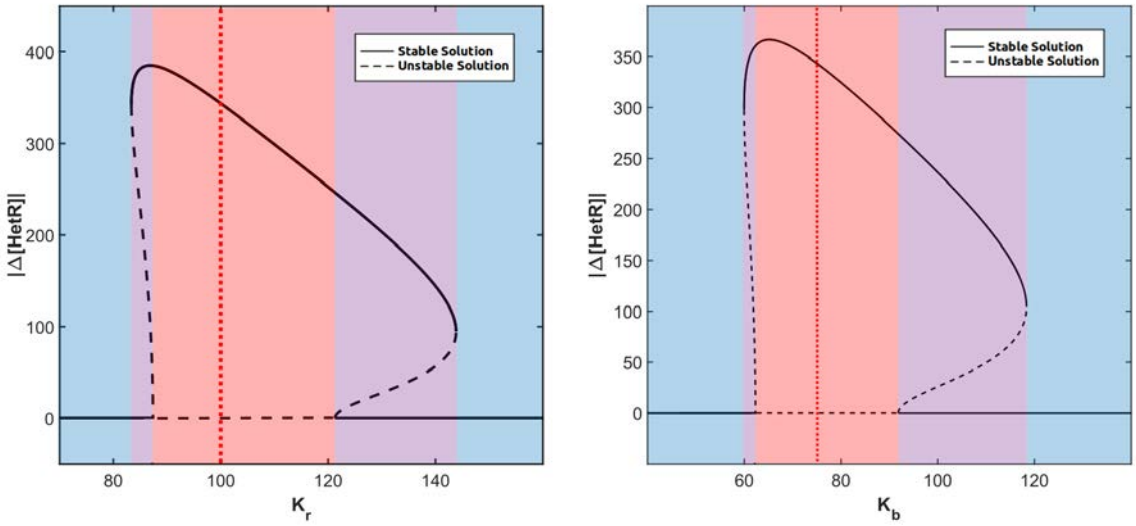

Figure 3.6: Bifurcation diagram for the parameters of HetR interaction: K_r of the promoter induction and K_b of the dimer formation. The homogeneous regime is colored in blue, the bistable one in purple, and finally the heterogeneous one in red. The WT value is presented with a red dotted line.

means that there are two regions of the phase diagram for those parameters where you could move, with very small variations of the initial pa-

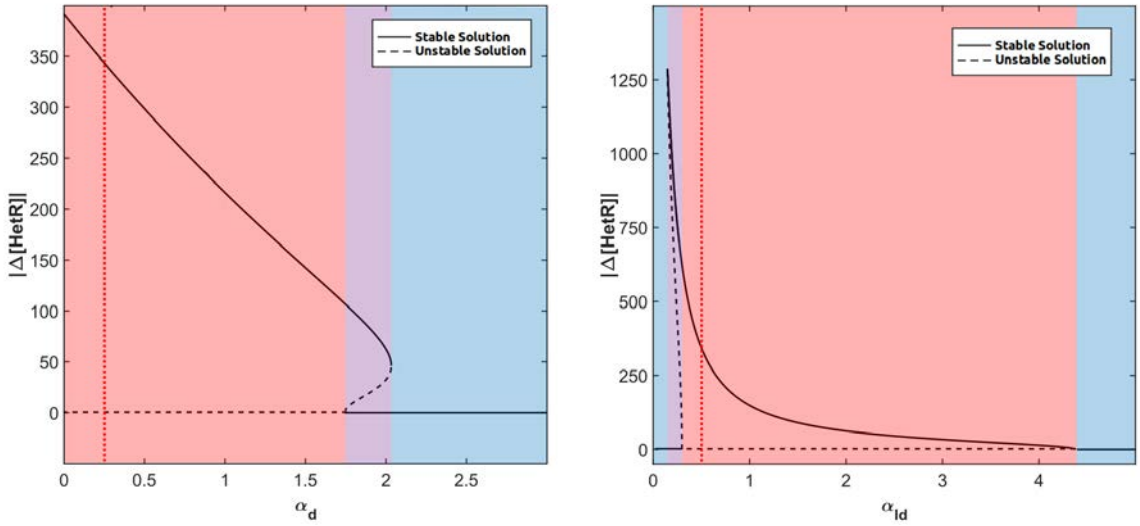

Figure 3.7: Bifurcation diagram for the parameters of HetR dimer degradation (α_d and α_{Id}). The homogeneous regime is colored in blue, the bistable one in purple, and finally the heterogeneous one in red. The WT value is presented with a red dotted line.

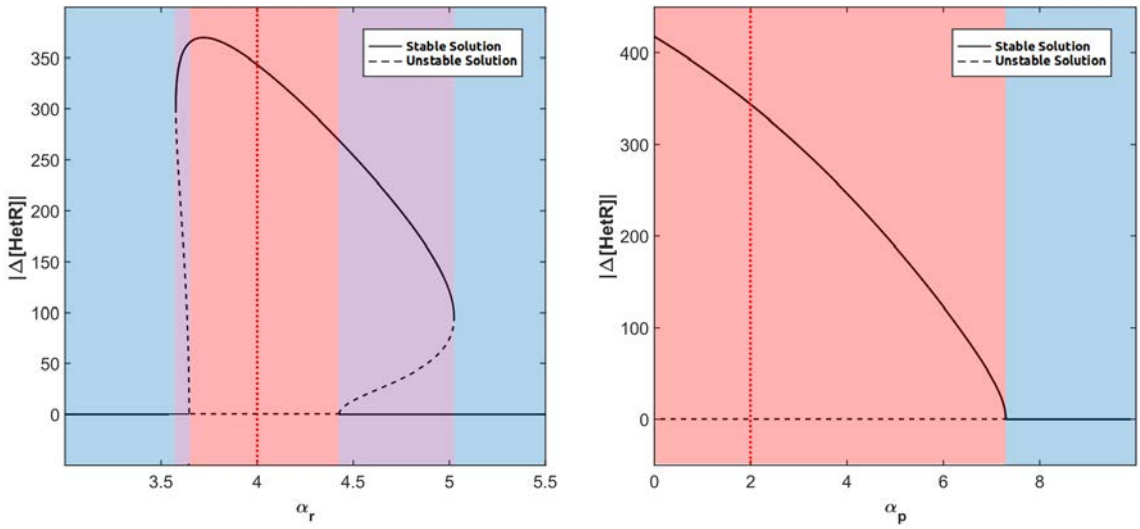

Figure 3.8: Bifurcation diagram for the parameters of HetR monomers and PatS degradation (α_r and α_p). The homogeneous regime is colored in blue, the bistable one in purple, and finally the heterogeneous one in red. The WT value is presented with a red dotted line.

rameter, through two different stable solutions depending on if you travel from the homogeneous phase to the heterogeneous or vice-versa. Those bifurcations look like two sub-critical pitchfork bifurcations opposed to each other for values near the branching point but the existence of a much

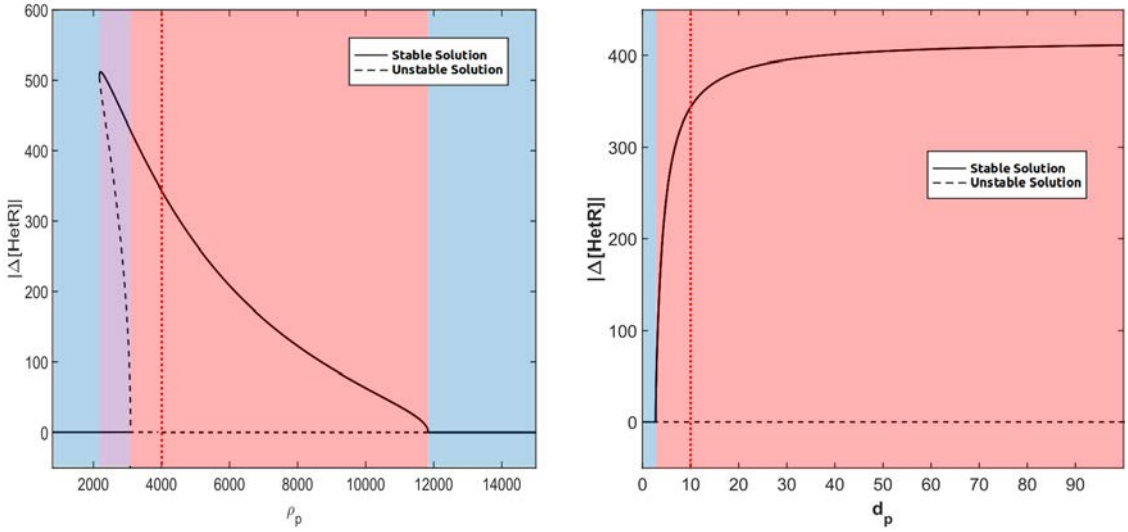

Figure 3.9: Bifurcation diagram for the parameters of PatS production (ρ_p) and conversion to the inhibitory form (d_p). The homogeneous regime is colored in blue, the bistable one in purple, and finally the heterogeneous one in red. The WT value is presented with a red dotted line.

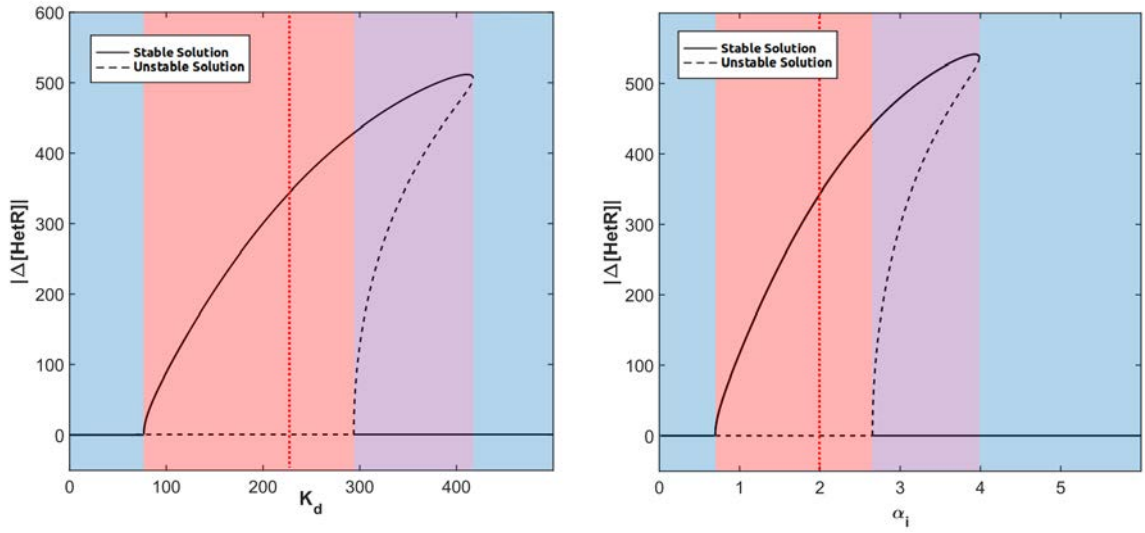

Figure 3.10: Bifurcation diagram for the parameters of the inhibitory interaction (K_d) and inhibitor degradation (α_i). The homogeneous regime is colored in blue, the bistable one in purple, and finally the heterogeneous one in red. The WT value is presented with a red dotted line.

higher order term stabilizes the unstable branches for values of the parameter bigger than the furthest from the critical value of the parameter that is still in the bifurcation region (Strogatz, 2015). This effect is what causes

the bistability and therefore the hysteresis phenomenon. Nevertheless, it is worth noting that we have not considered here the noisy character of gene expression.

The only other type of bifurcation observed in our model is a supercritical pitchfork bifurcation, where from one single homogeneous stable solution we obtain two symmetric heterogeneous stable solutions and the homogeneous one becomes unstable. Like the one in the leftmost bifurcation point of the two graphs in fig. 3.10.

Therefore, all bifurcations presented are either sub-critical fork bifurcations, if a bistable region can be observed, or supercritical fork bifurcations if none exist. The fact that all bifurcations are pitchfork type is quite logical since the phase change occurs between homogeneous and heterogeneous solutions and therefore a single homogeneous solution always has to appear as two heterogeneous as the cells are indistinguishable and then each one of them could be the one with higher concentration.

This 2-cell system reproduces the behavior observed in the sensibility analysis of (Muñoz-García and Ares, 2016). In this previous work, it was observed that the most sensible parameters were the ones related to HetR production, interactions, and degradation. One can observe in figs. 3.5 to 3.7 and fig. 3.8 left, that these same processes are also the ones with smaller heterogeneous regions with two surrounding bistable regions.

It is also worth noting that while the dimer degradation parameters were not especially sensible in (Muñoz-García and Ares, 2016) this 2-cell system, clearly distinguishes the effect over inhibited and non-inhibited dimers. One can see in fig. 3.7 that both present a huge heterogeneous region, but both the location of the wild-type parameters and the bistable region differs. For the degradation of uninhibited dimers, the behavior would only change if the value increases six or seven-fold, but one could consider that there is no degradation and the system would remain in the pattern regime. On the other hand, the inhibited dimer degradation presents two homogeneous regions. One, equivalent to the uninhibited inhibition, for high degradation values and another, much smaller one, for really low degradation values with a bistable transition. Even if it is interesting that the bistable region seems to coincide with the non-inhibited degradation value, the particular values of these transitions are not relevant because they would change when other parameters are modified. The fact that the non-degradation point is homogeneous is important. This indicates that while there is no need for the degradation of free dimers one should consider degradation of the inhibited ones in order to observe the pattern. This observation aligns with the previous experimental hypothesis that attributed a proteolytic function to both *patS* and *hetN* (Risser and Sean M. Callahan, 2009).

3.3 NUMERICAL STOCHASTIC SIMULATION OF 100 CELLS

In order to check if this 2-cells system is representative of a much larger one. For this, we adapted the same code used to adjust the parameters in fig. 3.2 to eliminate both growth and the differentiation into heterocyst in a static 100 cells filament.

First, we studied deterministic evolution where one does not consider any noise in gene expression. Without noisy dynamics, the initial conditions completely define the available equilibrium states. One can observe in fig. 3.11 that the system requires a certain initial heterogeneity to establish a pattern along the filament. This initial dispersion is represented in our system by taking the initial value from a Gaussian distribution with mean IC and variance σ_sIC. In fig. 3.11 we present the effect of both this multiplicative noise σ_s and the mean value IC.

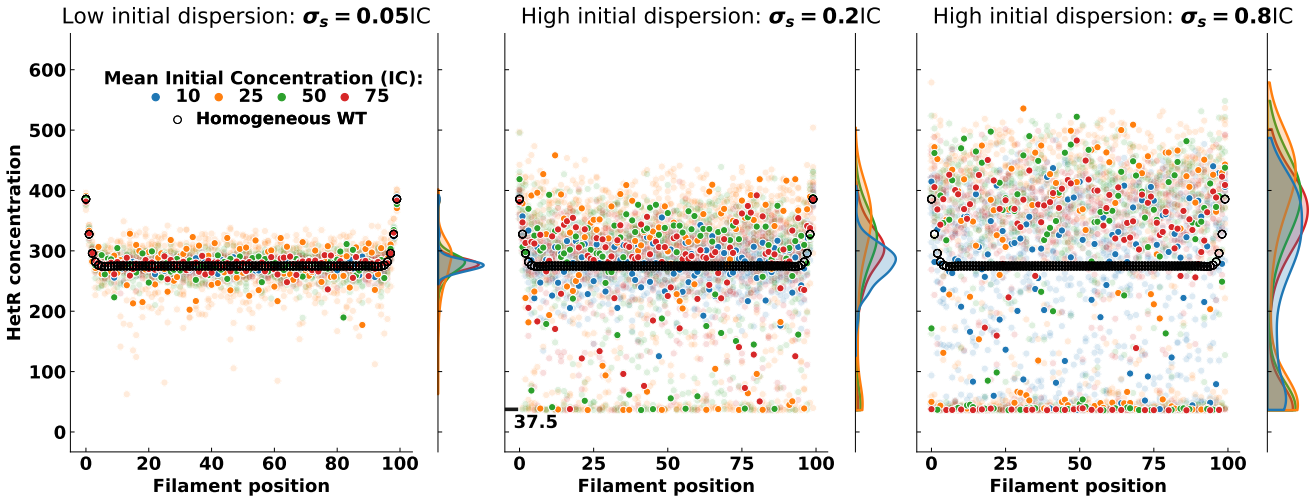

Figure 3.11: **Effect of initial heterogeneity on pattern formation:** The main plots present the concentration of HetR along the filament for 10 filaments initiated with four different initial concentrations of HetR and three different levels of dispersion. The translucent points are all the concentration values, the brighter ones are from one particular filament, and the empty black dots represent the situation with completely homogeneous initial conditions. The lateral plots present the marginal distributions of the concentration of HetR in the filament cells.

To study the effect of the multiplicative noise σ_s, one should compare the three subplots between them. For lower values of noise, the system only presents an equilibrium state that is analogous to the one observed in a completely homogeneous situation. We only present one case for the completely homogeneous situation because we obtained this same equilibrium state for all the different IC. But as the initial noise increases,

this equilibrium value increases and diverges from the homogeneous one and a new low HetR concentration equilibrium appears.

This general behavior is common for all the initial conditions, but the sensibility of the system is not equivalent for all the initial concentrations of HetR (IC). Already in the leftmost marginal plot, one can see that the initial distribution around [HetR]=25 nM presents a much larger equilibrium dispersion despite having a much lower initial dispersion (as we are using a multiplicative noise). Additionally, this same case (IC $=$ 25 nM) is the first to present two equilibrium states (central marginal plot) and also presents the higher distance between the two equilibrium points for the high noise regime. It is also worth noting that while the higher concentration equilibrium changes with the election of (IC) the one with a lower HetR concentration is always located around [HetR]=37.5 nM. This particular value is not that relevant, as is defined by the particular set of parameters chosen, but the fact that is common for all the IC indicates that is an absorbent state and its value probably defines the best choice for IC. This state probably corresponds to a deactivated cell state, where the regulated production is completely inhibited ($g_j(r_j, i_j) \approx 0$ in eqs. (3.3)). But further work would be necessary to analytically prove this hypothesis. As IC $= 25nM$ seems to be the best initial concentration to originate patterning, we will use this value with the higher static noise $\sigma_s = 0.8$IC for all the consecutive studies.

If one considers some noisy gene expression, fig. 3.12, the relevance of both the IC and σ_s is lost. The dynamical noise completely dominates the system and defines if a pattern forms along the filament. For low values of noise, $\sigma_d = 0.05$, the system presents the same two clear and separated attractors for all the considered initial conditions. We should not denominate this state equilibrium states because this filament with only *hetR* and *patS* never reaches equilibrium and the patterns have a dynamical nature. Despite having a different nature from the deterministic equilibrium states, these attractors are located at roughly the same concentration values. This distribution of "on" and "off" cells along the filament is broken for higher values of noise, $\sigma_d = 0.1$. In this noise regime, the separation between the two attractors is lost, and we cannot observe a "valley" in the marginal plot.

Once we studied the WT parameter's behavior to ensure that we have a heterogeneous state for this extended system, we will study both two different parameters and two levels of noise. This will show if the predictions of the 2-cell reduced system hold for two noise regimes. We have chosen the two production parameters of HetR given that are the ones with the narrowest heterogeneous regions. We have chosen the parameter values as central to study as central as possible on their representative

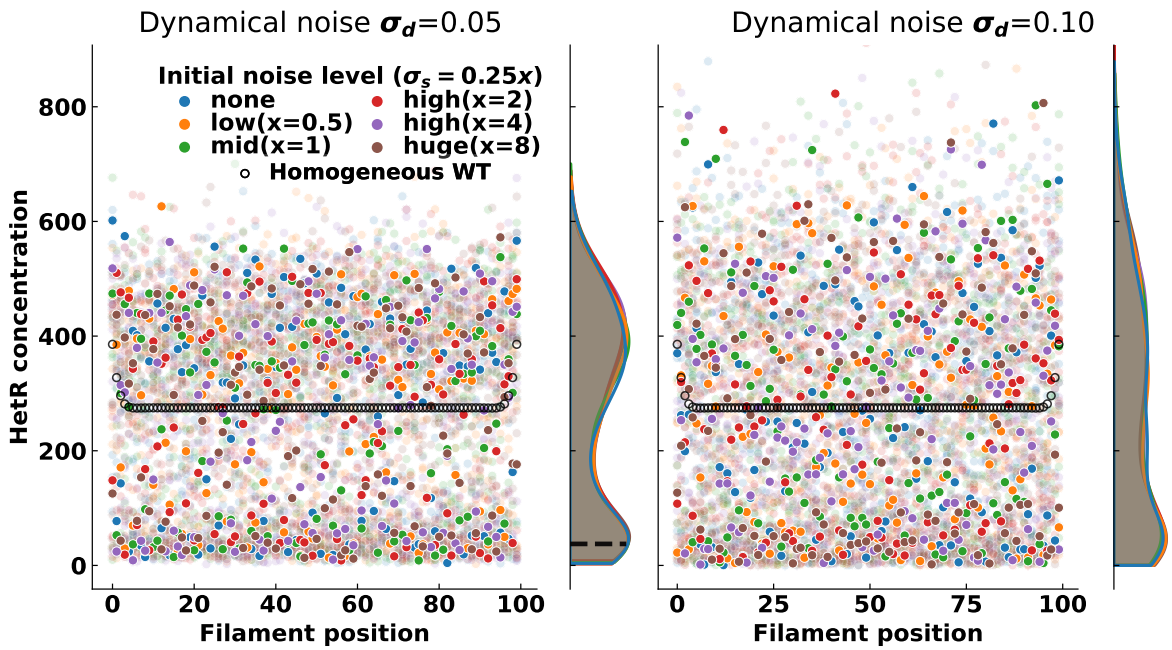

Figure 3.12: **Effect of the noisy gene expression on pattern formation**: The plots show the HetR concentration along 10 filaments for each of the 7 initial conditions and two regulatory noise σ_d regimes. The brighter dots are from one particular filament, and the empty black dots represent the situation with completely homogeneous initial conditions and no dynamical noise. The marginal plots show the distribution of HetR for all the initial conditions, and the dashed line marks the 37.5 nM maximum observed in fig. 3.11.

region in fig. 3.5. For β_r, fig. 3.13, we studied two values for the lower bistable region, $\beta_r = 10\,\&\,20$ nM h^{-1}, the wild type value, $\beta_r = 40$ nM h^{-1}, for the heterogeneous region and $\beta_r = 80$ nM h^{-1} for the homogenous region. We chose two values in the same bistable region to study if the distance to transition at the heterogenous one, in the parameter space, affected the observed state. One can first observe that for this parameter, low levels of noise affect very little the bistable regimes. This is probably because this parameter is the only basal production of our model eqs. (3.3) and, as such, its reduction greatly cools down the dynamics of the system. For a much slower system, the dynamic noise is much less relevant. All the non-heterogeneous regimes present a concentration close to one of the homogeneous solutions for most of the cells of the filament. Alternatively, in the heterogeneous regime, this homogeneous solution becomes unstable and is located in the "valley" area between the two equally populated attractors.

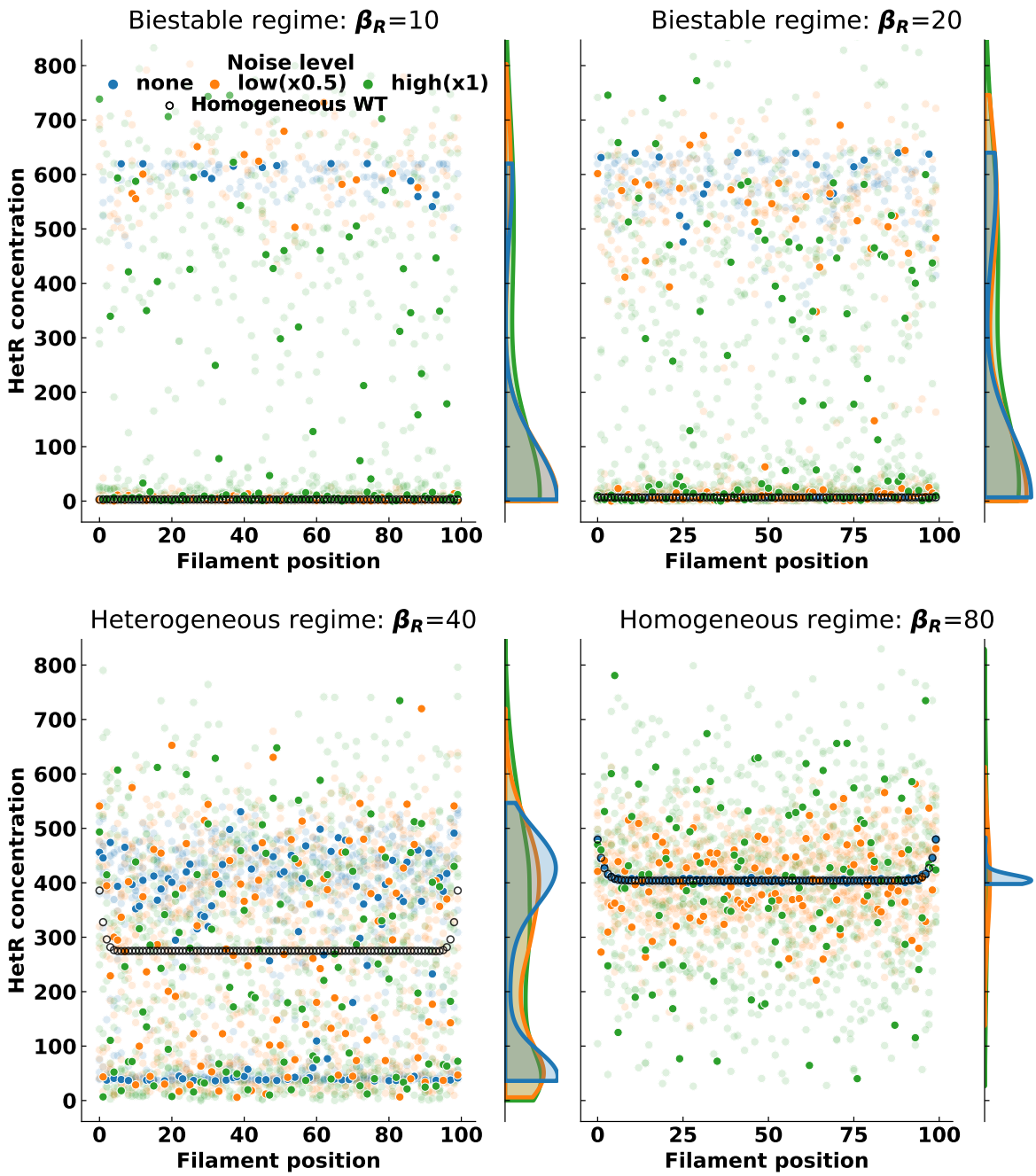

Figure 3.13: **Effect of the basal production of HetR on the pattern formation**: The plots show the HetR concentration along 10 filaments for 4 different β_r and three regulatory noise σ_d regimes. The brighter dots are from one particular filament, and the empty black dots represent the situation with completely homogeneous initial conditions and no dynamical noise. Each plot title marks the expected regime on the 2-cell analysis 3.5 left.

Alternatively, for ρ_r, fig. 3.14, we considered only one value for each region because the system presents two transitions (3.5 right). Due to this, we studied the lower concentration homogeneous state with $\rho_r = 700$ nM h^{-1}, the lower bistable region with $\rho_r = 810$ nM h^{-1}, the heterogeneous WT with $\rho_r = 1000$ nM h^{-1}, the higher bistable region with $\rho_r = 1170$ nM h^{-1}, and the high homogeneous state $\rho_r = 1300$ nM h^{-1}.

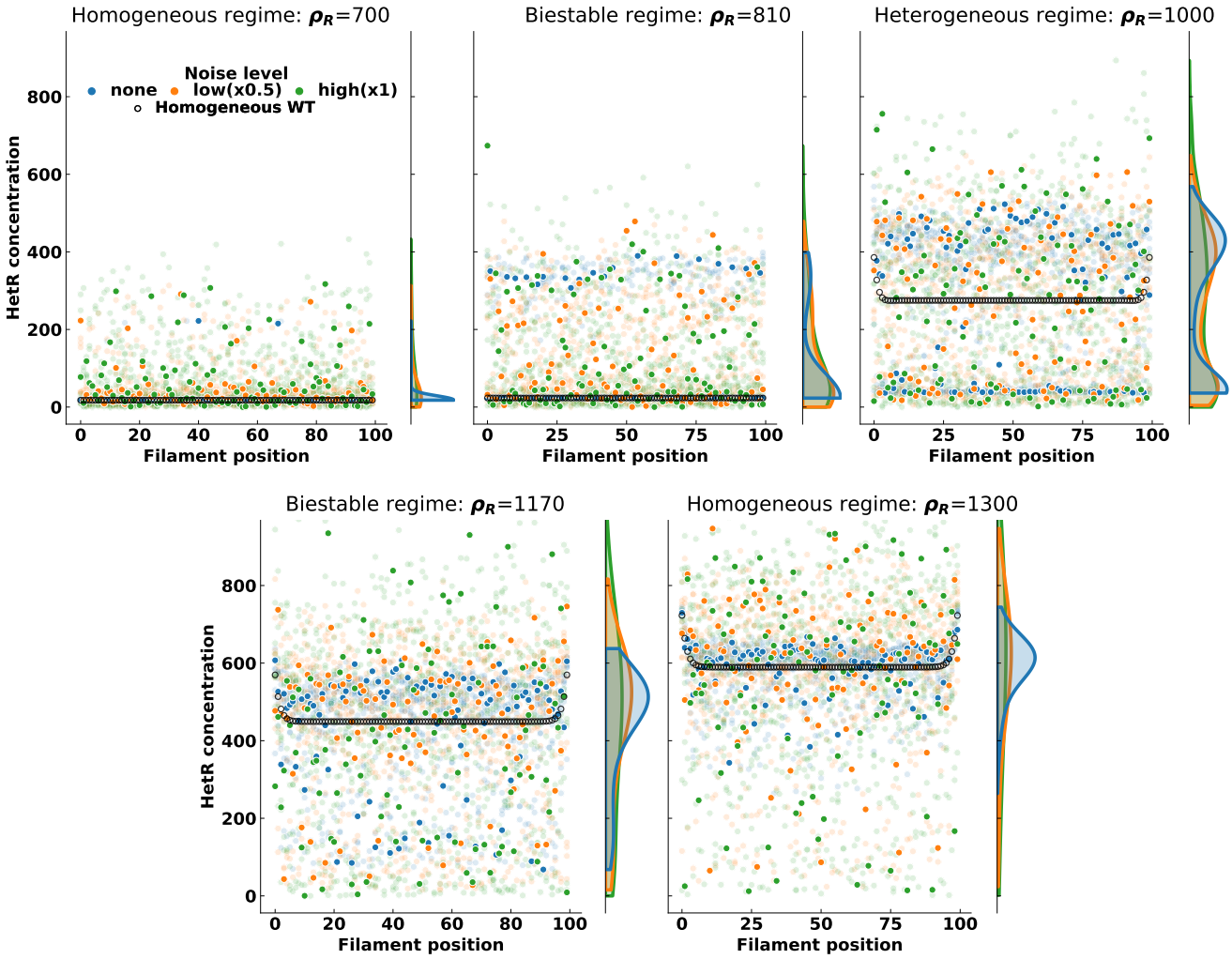

Figure 3.14: **Effect of the regulated production of HetR on the pattern formation**: The plots show the HetR concentration along 10 filaments for 4 different ρ_r and three regulatory noise σ_d regimes. The brighter dots are from one particular filament, and the empty black dots represent the situation with completely homogeneous initial conditions and no dynamical noise. Each plot title marks the expected regime on the 2-cell analysis 3.5 right.

While for β_r we observed a lower effect of the noise in the bistable areas, here we see that the addition of noise in all the non-heterogeneous

regimes is much more disruptive than its increase. It is really easy in fig. 3.14 that the presence of noise produces a discontinuous activation of the cells in the filament. In the homogeneous case without any noise, we observe a continuous progressive activation of all the cells of the filament. Alternatively, once one introduces noise in the system, the filament distributes its cells between the "on" and "off" states in a proportion that depends on the regimes described in 3.5 right.

As shown in both figs. 3.13 and 3.14, the 2-cell prediction works perfectly for noise regimes in which the system presents heterogeneity for the WT parameters. One can observe that the bistable attractors are really asymmetrical. The one closer to the homogeneous solution is always more populated than the other one, independently of the initial conditions. We could not find hysteresis in this system. But a more extensive exploration[3] reports that while the introduction of noise reduces the bistable area, the region does not disappear and therefore the transition is always abrupt and with a hysteresis effect *(unpublished data from Doldan-Martelli's study)*. The overall behavior described in section 3.3 can be expected for all the other parameters considered in section 3.2, given that all the observed bifurcations are of the pitchfork type.

[3] *The extended study done by my colleague Victoria Doldan-Martelli where pairs of parameters are modified at the same time and noise is incorporated in a 50 cells equally simplified system is still in preparation*

3.4 BIOLOGICAL INTERPRETATION OF THE RESULTS

One could interpret modification of the parameters as possible biological adaptations to the media. As such, the presence of bistable regimes close to the parameter space provides the system with higher adaptability. The discontinuous nature of the transition between the rate of cells with low and high concentrations of HetR (denominated "off" and "on" states during this chapter) in the filament facilitates sudden changes in the filament strategy if the environmental conditions change suddenly. As we have already discussed in section 2.1, a high concentration of HetR is known to be required for the differentiation into a heterocyst. Then having a wide array of different strategies to distribute the level of HetR production in the cells is really convenient. This would allow the cell to remain in the most energetically favorable "off" homogeneous state for stable environmental conditions. And, once it reacts to environmental changes through parameter changes, has the possibility of a partial commitment to the "on" strategy through a bistable regime where the reversion to the homogeneous state is still much easier. In our particular case, the advantage of the discontinuous transition can be easily seen in both the β_r diagram 3.5 left and fig. 3.13. This is due to the fact that β_r is an artificial parameter of our model constructed to ease the study of the pattern formation environmental conditions. We substituted the *ntcA* initialization with a basal production of *hetR* through β_r. Then, our virtual filaments are always in

nitrogen-deprived conditions as *hetR* production is always on (albeit at a much lower level than through the self-regulated production).

With this in mind, one could simulate the detection of nitrogen deprivation in our model as a modification of β_r value. Its value would progressively increase from the $\beta_r = 0$ which corresponds to an [HetR]$= 0$ homogenous states until reaching the heterogeneous $\beta_r = 40$ nM h^{-1}. With this framing, it is easy to see that the lower bistable region in 3.5 left allows for a much faster adaptation to the new conditions. Even with tiny values of β_r there would be some cells with a high enough HetR concentration to allow for a potential heterocyst formation in the filament. Then, if nitrogen deprivation is maintained, our filament would gradually increase the proportion of "on" cells until reaching the heterogeneous regime. In that regime, the heterocyst formation would be much easier because roughly half of the cells present a HetR concentration compatible with the initiation of the differentiation.

4

INCORPORATION OF *PATA* AND *HETF* TO THE MODEL

In the last chapter, we extended the study of the minimal model presented in (Muñoz-García and Ares, 2016) observing the conditions that give rise to the multistability required to originate a pattern. We realized that even with a smaller system of just 2 elements, *patS* and *hetR*, is capable to show a phase transition between a homogeneous and a heterogeneous filament configuration. But, as it was already discussed in the original work (Muñoz-García and Ares, 2016), it is necessary to include hetN in the system to have proper maintenance of the quasi-regular pattern of heterocysts along the filament. But, as one can observe in fig. 2.5, the regulatory network that controls heterocyst differentiation is much larger. In this chapter, we will expand the model through the inclusion of *hetF* and *patA* whose role is still not clearly defined.

The most striking feature of the Δ*patA* mutant is that it forms heterocysts almost exclusively at the terminal cells of the filament, while the Δ*hetF* cannot form heterocysts. Then, these two genes seem to have a clear role in pattern formation, given the heterocyst distribution on the filament is heavily distorted in their deletion mutants.

As we did in the previous chapter 3, we will first recap and expand the most relevant biological information to properly justify our modeling choices.

4.1 EXPERIMENTAL EVIDENCE ABOUT THE ROLE OF *hetF* AND *patA*

The fact that *patA* seems to be required for the differentiation of intercalary heterocysts but not for terminal heterocysts has made some authors think that a different differentiation process in which *patA* is not involved could occur depending on cell position (Orozco et al., 2006). Surprisingly, even though rare intercalary heterocysts are formed in strains lacking the *patA* gene, high levels of HetR (higher than in the wild type) are measured 18 hours after removing combined nitrogen (Risser and Sean M. Callahan, 2008, 2009). Regarding the connection between *patA* and the master regulator in heterocyst differentiation, *hetR*, multiple contiguous heterocysts appear when *hetR* is over expressed (Liang et al., 1992). However, this differentiation is mostly suppressed in the *patA* mutant, for which the same *patA* phenotype with only terminal heterocysts is

obtained even under nitrogen starvation. Additionally, in (Young-Robbins et al., 2010), it was observed that the *patA* transcription is greatly reduced in strains for which the expression of *hetR* is blocked. Therefore, the activation of *patA* expression seems to be directly up regulated by HetR.

More insights into the functional relationship between *patA* and other genes involved in heterocyst differentiation are presented in (Orozco et al., 2006). This work studies the connections between *patA* and *hetR*, *patS*, and *hetN*, analyzing the single, double, and triple mutant phenotypes. For the *patA* mutant background with *hetN* inactivated, they obtain a phenotype indistinguishable from the *patA* mutant with single terminal heterocysts at 24 hours post-induction. However, an increasing number of contiguous heterocysts are formed, mostly at the ends of the filament, after that time. In the case of the Δ*patA*Δ*patS* double mutant, its phenotype is identical to the Δ*patS* single mutant during short times after induction. This seems to imply that a functional *patS* gene is required to obtain a *patA* phenotype. However, after 48 hours, the average distance between heterocysts for the Δ*patA*Δ*patS* double mutant is larger than in the Δ*patS* single mutant and the wild type.

Based on the previous results, it is suggested that PatA might reduce the efficiency of the inhibitory function of both PatS and HetN. This effect could be achieved in two general ways. The first one would be that PatA interacts with PatS and HetN to reduce its inhibitory potential through a post-transcriptional modification that could be forced degradation, a conformation change, or sequestration. On the other hand, PatA could also interact with HetR to protect it by reducing its sensitivity to inhibition. Nevertheless, given that the Δ*patA*Δ*patS* and the Δ*patA*Δ*patS*Δ*hetN* mutants do not present the same phenotype of the Δ*patS* and the Δ*patS*Δ*hetN* mutants, PatA must have another effect besides the protection of HetR to the inhibition through PatS and HetN. The Δ*patA*-like phenotype obtained for an isolated allele of *hetR* made the authors in (Orozco et al., 2006) suggest that *patA* might also promote differentiation independently of its effects on *patS* and *hetN*.

A study by Risser and Sean M. Callahan, 2008, shows that the deletion of *hetF* in *Anabaena PCC 7120* produces enlarged vegetative cells (with a morphology similar to *patA* over expression) that do not differentiate into heterocysts even after several days of nitrogen starvation. When *hetF* is over expressed, vegetative cells become significantly smaller than those in the wild type, and multiple contiguous heterocysts (the Mch phenotype) are induced 24h after nitrogen step-down.

Their deletion mutants (and the double mutant) present similar high levels of HetR. However, the addition of an ectopic functional HetF reverts the phenotype of all these mutants to the wild type. Furthermore, the

regulatory effect of those genes on *hetR* seems to be post-transcriptional. Both *patA* and *hetR* is necessary for the aberrant cell morphology of the Δ*hetF* mutant, and the addition of extra copies of *hetF* can functionally bypass the deletion of *patA* without requiring a direct interaction of PatA with *hetF*. Finally, *hetR* self-regulation and *patS* up regulation through HetR depend on *hetF*. These results lead the authors of (Risser and Sean M. Callahan, 2008) to suggest the existence of an activation process of HetR controlled by *hetF* which induces the *hetR* regulatory function.

4.2 GENE REGULATORY NETWORK

Using the information presented in section 4.1, we propose a regulatory network, depicted in fig. 4.1, that includes the main genes involved in heterocyst differentiation. This genetic network[7] modifies and expands a previous minimal model for the interaction of *hetR*, *patS*, and *hetN* (Muñoz-García and Ares, 2016).

[7] *Which is a reduction from fig. 2.5*

This proposal includes the novel gene *patX*. This gene has recently been described as a redundant gene for the inhibitory mechanism of *patS* (Elhai and I. Khudyakov, 2018; I. Khudyakov et al., 2020). Despite acting analogously to *patS*, *patX* seems to play a secondary role, complementing the main signal produced by *patS*. As shown in (I. Khudyakov et al., 2020), the Δ*patX* single mutant does not present an altered phenotype. But on the other hand, the Δ*patX*Δ*patS* double mutant displays a much higher percentage of heterocysts than the Δ*patS* single mutant with a lethal almost complete Mch phenotype (Elhai and I. Khudyakov, 2018; I. Khudyakov et al., 2020). To consider this in a simple way, we consider that the variable in our model for PatS represents the combined effects of PatS and PatX. The knock-out of *patS* will be modeled reducing 90% the value of the production rate for this variable; the remaining 10% represents the redundant effect of *patX* expression.

Additionally, we consider the same functional mobile form of the inhibitor for all the inhibitory genes: *patS*, *hetN*, and *patX*. This inhibitor is the hexapeptide ERGSGR (Rivers et al., 2018). We consider that the hexapeptide is produced as a modification of PatS, HetN, or PatX at the cell membrane, with characteristic rates for each protein. The resulting product is then exported to the neighbor cell. This hexapeptide has shown to have a higher affinity with HetR than the pentapeptide originally proposed (Feldmann et al., 2012).

The most relevant novel inclusion in the model we propose is the requirement of a post-transcriptional transformation of HetR to act as a genetic regulator. This active form of HetR, which we term AHetR, is probably obtained through a phosphorylation process (Makarova et al.,

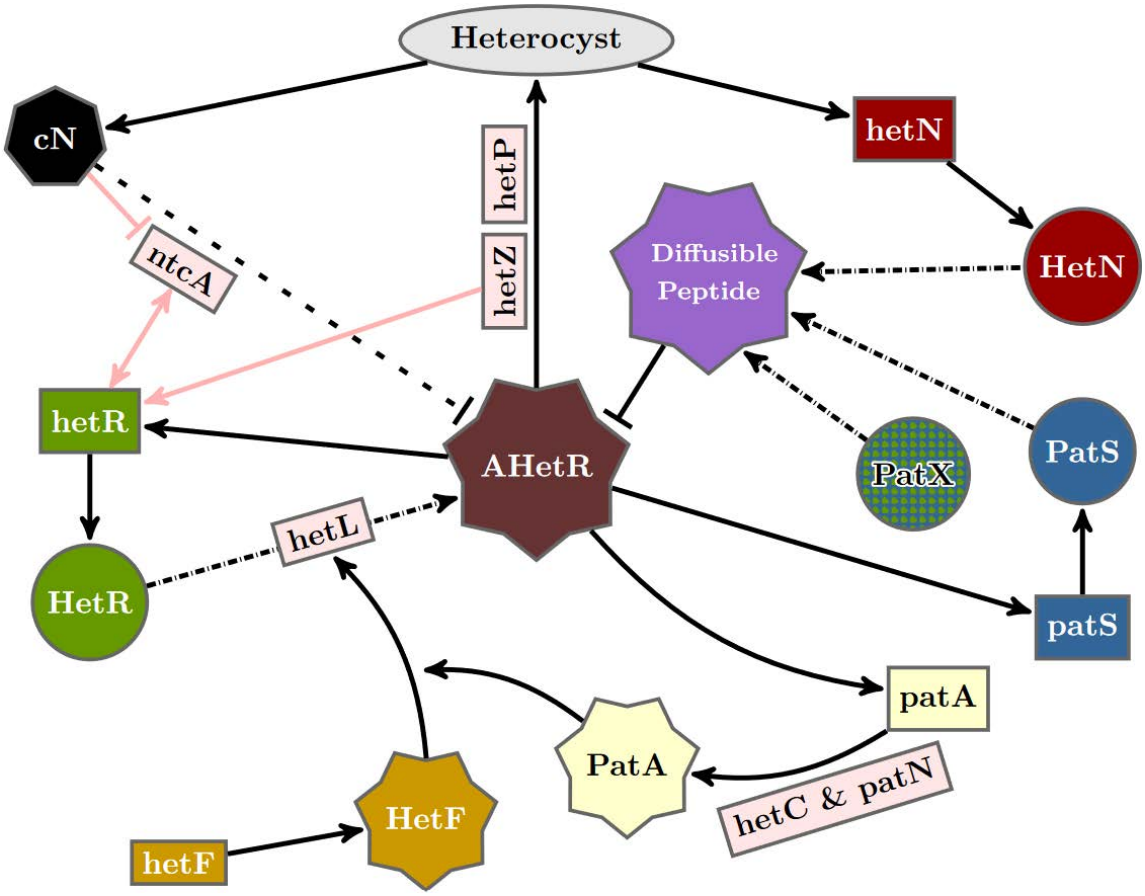

Figure 4.1: **Gene regulatory network of heterocyst differentiation.** The main elements involved and their interactions for heterocyst differentiation are depicted schematically together with other relevant genes (light red rectangles) not included in our model. Rectangles, circles, and polyhedral forms represent genes, inactive proteins, and active products, respectively. AHetR stands for the active form of HetR. The ellipse represents differentiation into a heterocyst. Arrows with solid lines represent interactions between elements. Arrows with dashed-dotted lines represent post-transcriptional changes. The dashed line represents a simplified interaction of nitrogen sensing through *ntcA*.

2006; Valladares et al., 2016; Zhou et al., 1998b) and only the active fraction of HetR would contribute to heterocyst differentiation. This could explain the apparent paradox of a higher concentration of HetR with less heterocyst formation in both Δ*hetF* and Δ*patA*. The high concentration of HetR in these mutants would be explained through a higher turnover rate for the activated HetR protein (Zhou et al., 1998b).

A possible active form of HetR has been recently suggested in (Roumezi et al., 2020), where the authors present evidence for the phosphorylation of HetR as crucial for its activity in *Nostoc PCC 7120*. This phosphorylation is shown to require the presence of the Pkn22 kinase, but no

more information regarding the regulation of this phosphorylation is provided. Here we hypothesize that this would be the role of the genetic pathway controlled by *hetF* which has already been presented as a protease (Risser and Sean M. Callahan, 2008) and therefore can be expected to have a role in post-transcriptional modifications. Thus, we assign to *hetF* the role of activator of HetR with the mediation of PatA through a post-transcriptional interaction that would change HetR into its enzymatic form.

This hypothesis is inferred by observing changes in the phenotypes of several mutant backgrounds and has been considered in fig. 4.1. However, the nature of this interaction cannot be confirmed, since the observed phenotype could also be explained through indirect interactions mediated by HetF.

To complete the mechanistic model, one must take into account that different genes are expressed in heterocysts or vegetative cells, as depicted in fig. 4.2.

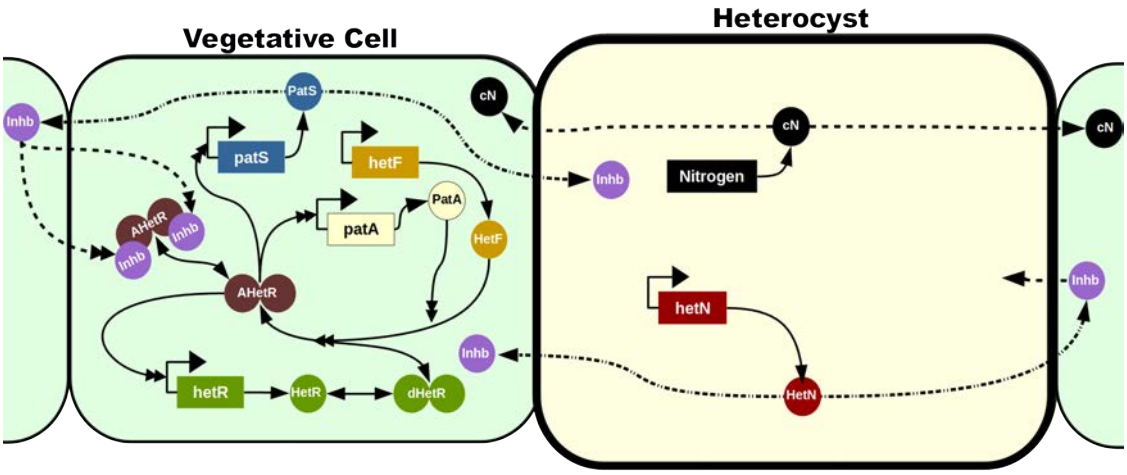

Figure 4.2: **Mechanistic Model:** The vegetative cells are represented with a soft green background, while the heterocyst has a soft yellow background and a thicker cell wall. Genes are represented in rectangles and proteic elements with circles. The dimers are represented with two attached circles and can be inactivated (in green), activated (in brown), and activated and inhibited (in brown with two attached purple inhibitors). Solid lines represent production (with only one simple arrowhead), transformations (with a simple arrowhead on both ends), and interactions (with a double arrowhead). Dashed lines represent intercellular traffic, and dashed-doted lines represent a transformation when exported to a neighboring cell.

As shown in fig. 4.2, while *hetF* is produced only constitutively at a low-level (Wong and J. C. Meeks, 2001), *patA* only has a regulated expression that depends on the active form of the dimeric transcription factor

AHetR, which also activates its own expression. At the protein level: the HetR dimer needs to be activated by HetF (whose enzymatic activity can be enhanced by PatA) to become AHetR. PatS become an inhibitor of the transcription factor by protein transformation during cell-to-cell transport. The inhibitor thus produced is a small molecule that can move along the filament. *hetN* is expressed basally in heterocysts and becomes an inhibitor of the transcription factor, similar to the PatS product, by protein transformation during cell-to-cell transport. The fixed nitrogen products produced by the heterocyst can also move to act as an inhibitor of AHetR.

4.3 MATHEMATICAL MODEL

Here, we have used the interactions explained below in the previous section and depicted in fig. 4.1 to design our model schematically presented in fig. 4.2. We formulated a set of differential equations for the evolution of the species involved in heterocyst differentiation.

The full description of the resulting mechanistic model is presented in appendix B. First, we assumed that protein interactions are much faster than the production of those proteins, equilibrium states were considered for the reactions with a shorter timescale. Then, we incorporated information about the expected relative abundance of each protein to further simplify the system through approximations. In that same appendix, we also detail the simplifications and approximations taken to reduce both the size and the computational load. Applying these simplifications, we get a more manageable mathematical model to describe the temporal evolution of the concentration of the main protein monomers.

Thus, the temporal evolution equation for the main species are

$$\frac{dR_j}{dt} = \beta_R + g(R_j,A_j,I_j,G_i)\rho_R - \alpha_R R_j\left(1+2\mu R_j\right) \tag{4.1}$$

$$\frac{dA_j}{dt} = g(R_j,A_j,I_j,G_i)\rho_A - \alpha_A A_j, \tag{4.2}$$

$$\frac{dS_j}{dt} = (1-\delta_{Hc,j})g(R_j,A_j,I_j,G_i)\rho_S - 2c_S S_j - \alpha_S S_j, \tag{4.3}$$

$$\frac{dN_j}{dt} = \delta_{Hc,j}\rho_N - 2c_N N_j - \alpha_N N_j, \tag{4.4}$$

$$\frac{dI_j}{dt} = c_S(S_{j-1}+S_{j+1}) + c_N(N_{j-1}+N_{j+1}) + d_I(I_{j-1}-2I_j+I_{j+1}) - \alpha_I I_j, \tag{4.5}$$

$$\frac{dG_j}{dt} = \delta_{Hc,j}\rho_G + d_G(G_{j-1}-2G_j+G_{j+1}) - \alpha_G G_j, \tag{4.6}$$

$$g(R_j,A_j,I_j,G_i) = \frac{F_R R_j^2\left(1+\frac{A_j}{\tau_A}\right)}{1+F_R R_j^2\left(1+\frac{A_j}{\tau_A}\right)+\frac{I_j^2}{K_d^2}+\frac{G_j}{K_G}}, \tag{4.7}$$

where the subindex j indicates that the variable refers to the cell number j in the filament (being then $j-1$ and $j+1$ its neighboring cells) and t denotes time.

The concentration of the protein monomers is represented by R_j, A_j, S_j, and N_j, which stand for the concentration of HetR, PatA, the addition of both PatS and PatX, and HetN, respectively, in the cell j. We consider the same functional mobile form of the inhibitor for all the inhibitory genes considered: *patS*, *hetN*, and *patX*. This inhibitor is the hexapeptide ERGSGR (Rivers et al., 2018), represented by I_j in our model. We consider that the hexapeptide is produced as a modification of PatS, HetN, or PatX at the cell membrane, with characteristic rates for each protein. The product of the modification is exported to the neighbor cell. Finally, G_j represents the concentration of fixed nitrogen in the cell j. The factor $\delta_{Hc,j}$ specifies whether the cell is vegetative or a heterocyst: its value is 1 if the cell j is a heterocyst, and 0 if it is vegetative. This value is used as a switch for the production of both HetN and fixed nitrogen which are only produced in heterocysts and the production of PatS which is only produced in heterocysts.

Assuming that HetF is produced at a constant basal rate and degraded linearly, we have simplified the model considering an equilibrium concentration of HetF (F_{eq}) following

$$\frac{dF_j}{dt} = \beta_F - \alpha_F F_j = 0,$$

from where we get a constant concentration of HetF at any cell: $F_{eq} = \beta_F/\alpha_F$. Given the experimental observation that vegetative cells in *hetF* mutants have morphological phenotypes, it is probable that *hetF* has an additional role in cell morphology regulation. However, since we are focused on studying heterocyst differentiation, regulation of both HetR and HetF in heterocysts is irrelevant to our purposes and is not considered in the model.

For simplicity, we have considered basal production only for HetR monomers and HetF proteins with rates β_R and β_F respectively. The maximum regulated production rates are represented by ρ_R, ρ_A, ρ_S, ρ_N, and ρ_G, for HetR, PatA, both PatS and PatX, HetN, and fixed nitrogen, respectively. The linear degradation and dilution rates are α_R, α_F, α_A, α_S, α_N, α_I, and α_G, since dilution and degradation act in the same direction, these constants combine the effects of these two mechanisms. The active transport rates or diffusion between adjacent cells are c_S, c_N, d_I, and d_G.

We assume the border cells at the filament's ends leak both the inhibitor and the fixed nitrogen to the exterior at a lower rate than the communication between neighboring cells. Since molecular trafficking

between neighboring cells may depend on elements sitting on both cell membranes, something not possible in the border cells, we have reasoned that it might be possible that leakage to the exterior is less efficient than intercellular trafficking. To allow for this possibility, we have modeled this by multiplying the rates d_I, d_G for molecular trafficking between the exterior and the first and last cells of the filament by a factor d_{border}. The exact value of this factor does not have a qualitative effect on the model, but can affect the number of terminal heterocysts in some conditions. The resulting equations for I and G in the first and last cells of the filament are:

$$\frac{dI_1}{dt} = c_S S_2 + c_N N_2 - d_{border} d_I I_1 + d_I(I_2 - I_1) - \alpha_I I_1 \tag{4.8}$$

$$\frac{dG_1}{dt} = \delta_{Hc,1}\rho_G - d_{border} d_G G_1 + d_G(G_2 - G_1) - \alpha_G G_1 \tag{4.9}$$

$$\frac{dI_N}{dt} = c_S S_{N-1} + c_N N_{N-1} - d_{border} d_I I_N + d_I(I_{N-1} - I_N) - \alpha_I I_N \tag{4.10}$$

$$\frac{dG_N}{dt} = \delta_{Hc,N}\rho_G - d_{border} d_G G_N + d_G(G_{N-1} - G_N) - \alpha_G G_N. \tag{4.11}$$

[8] *Introduced in the seminal paper of this thesis (Muñoz-García and Ares, 2016).*

The μ parameter[8] refers to the nonlinear degradation of HetR mediated through its dimerization and could be further expressed as a function of the rates of binding (k_b) and unbinding (k_u) of monomers to form dimers and the degradation rates for both monomers (α_R) and dimers (α_d) of HetR as

$$\mu = \frac{k_b}{\alpha_R}\left(1 - \frac{k_u}{k_u + \alpha_d}\right).$$

This parameter is easily obtainable considering equilibrium over the binding dynamics and one assumes that all dimers degrade at the same rate. We decided to keep the dimmer inhibition presented in (Muñoz-García and Ares, 2016) and did not add a differentiated mode of inhibition for the inhibited dimers.

The main argument for this choice is simplicity. It was already shown in (Muñoz-García and Ares, 2016) that there is no need for a higher degradation for inhibited dimers to obtain a pattern. But there is some additional rationale behind this choice if one remembers the bifurcation diagrams for the 2-cell minimal model fig. 3.7. As we already mentioned in chapter 3, fig. 3.8 right shows that a certain amount of inhibited dimer degradation is necessary to observe a pattern. But, at the same time, once you reach that point further degradation will not affect the system because the heterogeneous region is quite large. On the other hand, the degradation of the uninhibited dimers fig. 3.8 left works exactly in the opposite way. While it is possible to reduce it as much as you want without affecting the pattern, the upper limit to observe a pattern is much

lower than the one of the inhibited one. Then, these two degradation rates, which have a substantial overlapping range, as only one would further reduce the parameter space to explore easing the adjustment of the model to experimental data.

We can simplify the system assuming that the promoter regulation through HetR is equivalent for *hetR*, *patA*, and *patS*. This regulation is modeled using the factor $g(R_j,A_j,I_j,G_i)$, an example of the HetR self-regulation for several conditions is shown in fig. 4.3. This eq. (4.7) represents the equilibrium state of the processes of dimerization, activation, and inhibition of HetR. To obtain this expression we have considered the following equilibrium constants: K_R for the dimerization of HetR, K_F for the activation of the HetR dimers by HetF, τ_A for the activation mediated through PatA, and K_d and K_G for the inhibition through the hexapeptide and the fixed nitrogen respectively. Assuming an equilibrium concentration of HetF, the expression for this regulatory term is reduced to

$$g(R_j,A_j,I_j,G_i) = \frac{\frac{F_{eq}}{K_F}\frac{R_j^2}{K_R}\left(1+\frac{A_j}{\tau_A}\right)}{1+\frac{F_{eq}}{K_F}\frac{R_j^2}{K_R}\left(1+\frac{A_j}{\tau_A}\right)+\frac{I_j^2}{K_d^2}+\frac{G_j}{K_G}},$$

where we have defined

$$F_R \equiv \frac{F_{eq}}{K_F K_R^2}$$

in order to obtain the simplified version in eq. (4.7).

4.3.1 *Model Implementation*

We have implemented a code in an object-oriented platform to model both the biochemical interactions which give rise to heterocyst differentiation and filament growth. Each cell of the filament has its own variables representing the cellular size and concentration of considered species. The dynamical equations for the concentrations of ERGSGR inhibitor and fixed nitrogen in each cell are coupled with its adjacent neighbors.

The resulting set of equations that controls the filament evolution is the noisy extension of the deterministic system eqs. (4.1) to (4.7). This system of equations has been expanded to the Langevin dynamics in the Itô interpretation (Gardiner et al., 1983). This expansion adds a stochastic term of the form $\omega_i^x(t)\sqrt{P_i^x + |D_i^x|}$ for each cell i and species x, where P_i^x and D_i^x are the sum of production (synthesis) terms and the sum of degradation terms respectively, and $\omega_i^x(t)$ is an uncorrelated Gaussian white noise (Frigola et al., 2012). This noise has zero mean and variance $\langle \omega_i^x(t)\omega_j^x(t')\rangle = \Omega_\Phi \delta_{ij}\delta(t-t')$ and models the intrinsic fluctuations in the genetic dynamics.

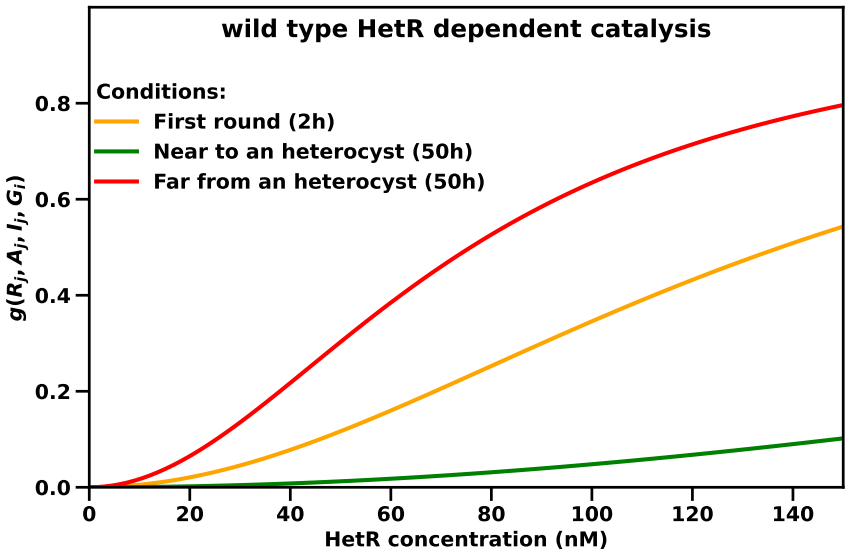

Figure 4.3: **Example of HetR self-regulation for several cell conditions:** The graph shows the value of $g(R_j,A_j,I_j,G_i)$ depending on the concentration of HetR for three different cell conditions. The first one, in yellow, represents a random cell in the initial times of the simulation before the rise of the pattern. The other ones are instead cells in a particular position of a filament with a heterocyst pattern already formed. The green one represents a cell near a heterocyst and therefore with a high concentration of both inhibitors and nitrogen. Alternatively, the red one represents a cell located in the middle of a vegetative interval. This one presents a higher response to HetR concentration due to the lower concentration of both inhibitors and nitrogen.

To differentiate into a heterocyst, a vegetative cell has to accumulate up to a certain level of HetR. This has been implemented with an integration of the value of HetR concentration over time for each vegetative cell, once the value of R_j is above a threshold, T_R^j.

This threshold is cell-specific, being drawn from a Gaussian distribution with mean T_R and variance $T_R^2\Omega_\Phi$. If at any point, the value of R_j drops below T_R^j, the integral is reset to zero. Otherwise, if the integral ever reaches a value M_R^j, also a cell-specific Gaussian distributed parameter with mean M_R and variance $M_R^2\Omega_\Phi$, the vegetative cell differentiates into a heterocyst. Given the integrative nature of this process, we have set a minimum time T_{min} necessary to avoid unrealistic sudden differentiation due to spikes in HetR production and properly reflect the extensive biological changes required to obtain a mature functional heterocyst. For simplicity, we have also used Ω_Φ to parametrize variability of T_R and M_R, to avoid defining too many noise-related parameters.

Vegetative cell growth was modeled by a stochastic differential equation for each cell:

$$\frac{d\Lambda_i}{dt} = \lambda\left[1 + \omega_i^{\Lambda}(t)\right], \tag{4.12}$$

where Λ_i is the size of the cell i, λ is a constant growth rate and $\omega_i^{\Lambda}(t)$ is an uncorrelated Gaussian white noise with zero mean and variance $\langle\omega_i^{\Lambda}(t)\omega_j^{\Lambda}(t')\rangle = \Omega_{\Lambda}\delta_{ij}\delta(t-t')$, which models the intrinsic fluctuations in the growth process. Starting from an initial size, each cell evolves following eq. (4.12) up to a maximum size M_{Λ}^i, which is a noisy value drawn for each cell from a Gaussian distribution of mean M_{Λ} and variance $M_{\Lambda}^2\Omega_{\Lambda}$. We have used Ω_{Λ} to parametrize this variance for simplicity, again to avoid having too many parameters describing noisy magnitudes.

When this size M_{Λ}^i is reached, the vegetative cell divides, producing two new vegetative cells with one-half of its current size and identical protein concentrations. Heterocysts follow the same growth, but once they have reached their maximum size, M_{Λ}^i, they do not divide and stop growing.

The code to simulate the model is available at `https://github.com/PauCasanova/AnabaenaPLOS`

4.3.2 *Parameter Estimation*

The parameter values employed in simulation results shown in this work can be found in table 4.1.

To simulate loss-of-function conditions we have considered the production rates equal to zero, except for of the *patS* loss-of-function, where we have reduced the production rate of *patS* by 90%. The remaining 10% represents the redundant effect still present through the expression of *patX* (Elhai and I. Khudyakov, 2018; I. Khudyakov et al., 2020).

In order to adjust our model, a custom simulated annealing algorithm (Kirkpatrick et al., 1983) was employed to obtain the parameters above the double line in table 4.1, employing as initial values for the optimization those in the model of (Muñoz-García and Ares, 2016) when an equivalent parameter exists. The last six parameters ($K_d, \lambda, M_{\Lambda}, T_R, M_R, T_{min}$) were manually set. The value of the affinity of ERGSGR to HetR, represented by K_d, was taken from (Feldmann et al., 2012). Additionally, we choose both the mean maximum cell size to $M_{\Lambda} = 4$ µm and the cellular growth rate to $\mu = 0.08\mu m \cdot h^{-1}$ in agreement with filament growth data (Asai et al., 2009). Finally, the values of T_R, M_R, and T_{min} are fixed beforehand in order to obtain a commitment time of around 8-9h (A. M. Muro-Pastor and Hess, 2012). The selection of T_R would be dependent on the typical concentration of HetR on a vegetative cell that is differentiating into a

Table 4.1: Parameter values used for the wild type simulations. The parameters over the double line are adjusted with a simulated annealing algorithm, and the ones under it have been fixed from the bibliography.

Parameter	Description	Value	Units
β_R	Basal production rate of HetR monomers	66.665	$nM \cdot h^{-1}$
ρ_R	Maximum regulated production rates of HetR monomers	763.33	$nM \cdot h^{-1}$
α_R	Linear degradation rate (including dilution) of HetR monomers	2.1293	h^{-1}
μ	Nonlinear degradation rate of HetR dimers	$1.6354 \cdot 10^{-3}$	nM^{-1}
ρ_A	Maximum regulated production rates of PatA	3862.3	$nM \cdot h^{-1}$
α_A	Linear degradation rate (including dilution) of PatA	3.823	h^{-1}
τ_A	Equilibrium constant of the PatA enhancement of the activation process of HetR dimers	81.132	nM
F_R	Fraction of HetF activated HetR dimers	$1.2578 \cdot 10^{-4}$	DL
ρ_S	Maximum regulated production rates of PatS	337.5	$nM \cdot h^{-1}$
α_S	Linear degradation rate (including dilution) of PatS	2.0827	h^{-1}
c_S	Rate of conversion of PatS to the ERGSGR hexapeptide through cellular transport	8.1485	h^{-1}
ρ_N	Maximum regulated production rates of HetN	527.2	$nM \cdot h^{-1}$
α_N	Linear degradation rate (including dilution) of HetN	2.6204	h^{-1}
c_N	Rate of conversion of hetN to the ERGSGR hexapeptide through cellular transport	5.7671	h^{-1}
α_I	Linear degradation rate (including dilution) of ERGSGR hexapeptide	2.3684	h^{-1}
d_I	Diffusion rate of the ERGSGR hexapeptide	9.0777	h^{-1}
ρ_G	Maximum regulated fixation rates of nitrogen	7533.7	$nM \cdot h^{-1}$
α_G	Linear degradation rate (including dilution) of fixed nitrogen	4.32	h^{-1}
d_G	Diffusion rate of the fixed nitrogen	100.04	h^{-1}
K_G	Equilibrium constant for the inhibitory reaction between HetR and the fixed nitrogen	162.03	nM
d_{border}	Rate of diffusion through the border cells	0.381	DL
Ω_A	Noise strength in cellular growth and size	0.081	DL
Ω_Φ	Noise strength in gene expression and initial concentration	0.21	DL
K_d	Equilibrium constant for the inhibitory reaction between HetR and the ERGSGR hexapeptide (Feldmann et al., 2012)	7.36	nM
λ	Cellular growth rate	0.08	$\mu m \cdot h^{-1}$
M_A	Average maximum cell size	4	μm
T_R	Average minimum HetR concentration threshold to differentiate	110	nM
M_R	Average accumulated HetR concentration required to form a heterocyst	1320	nM
T_{min}	Minimum time to differentiate	5	h

heterocyst. However, given that this experimental information is not available, we set a value taken directly from (Muñoz-García and Ares, 2016), and adjust the rest of the parameters. If one changes this value and runs again the optimization algorithm one obtains a new set of parameters, for which always the concentration of HetR for the *patA* mutant is uniformly maintained under the threshold, but close enough to allow the stochastic formation of heterocysts due to fluctuations. Once the value of T_R is set, M_R is the equivalent of maintaining the HetR concentration T_R over 12h, which is the maximum commitment time reported, and T_{min} is set to 5h to ensure that the concentration is at least this time over T_R and therefore eliminates the excessive relevance that sudden bursts could have on an integrative decision process, given the extensive morphological changes that the differentiation to heterocysts requires.

We have used a custom simulated annealing algorithm that selected a set of data that minimizes the following energy function:

$$\begin{aligned} E \equiv \sum_{O} \left[w_1 \cdot \Delta_{KS}(hist_{Exp}, hist_{Sim})^2 + w_2 \cdot \left((m_{Exp} - m_{Sim})^2 + (p_{Exp} - p_{Sim})^2 \right) \right] + \\ + \sum_{N} \left[w_3 \cdot \Delta_{KS}(hist_{Exp}, hist_{Sim})^2 + w_4 \cdot p_{Sim} \right] + \delta_{WT,NoHc} \left([\bar{HetR}]^{WT}_{Sim} - T_R \right)^2 \end{aligned} \tag{4.13}$$

where $\Delta_{KS}(hist_{Exp}, hist_{Sim})$ is the Kolmogorov-Smirnov distance (Massey, 1951) between the experimental and simulation histograms, m is the mean distance between heterocysts and p is the mean percentage of heterocysts. These quantities are compared between experiments and simulations. The comparison with (Corrales-Guerrero et al., 2014b) is done in the summatory with O which are the datasets of the wild type, Δ*patS* and Δ*hetN* for 24h, 48h, and 72h. While the comparison with (Orozco et al., 2006) is done in the summatory N which are the datasets of the Δ*patA* and Δ*patA*Δ*hetN* for 24h, 48h, 72h, and 96h. The inclusion of both the means and the distributions of vegetative length intervals allows the algorithm to set the noise in order to imitate the variance of the experimental data. Finally, $\delta_{WT,NoHc}$ is a factor that is 0 except if the wild type does not present heterocysts where 1. Given that all the other energy terms suppose heterocyst formation, we included a term that compares the HetR concentration for the wild type with the threshold value T_R in order to drive the system towards a set of parameters compatible with heterocyst formation. The set of parameters presented in this paper has been obtained using the weights $(w_1, w_2, w_3, w_4) = (1000, 10, 500, 1)$, which after some tests produced the best results.

One can observe in fig. 4.4, that the high stochasticity of the system forced us to have multiple realizations of each simulation in the simulated annealing algorithm in order to obtain a reliable energy for a parameter

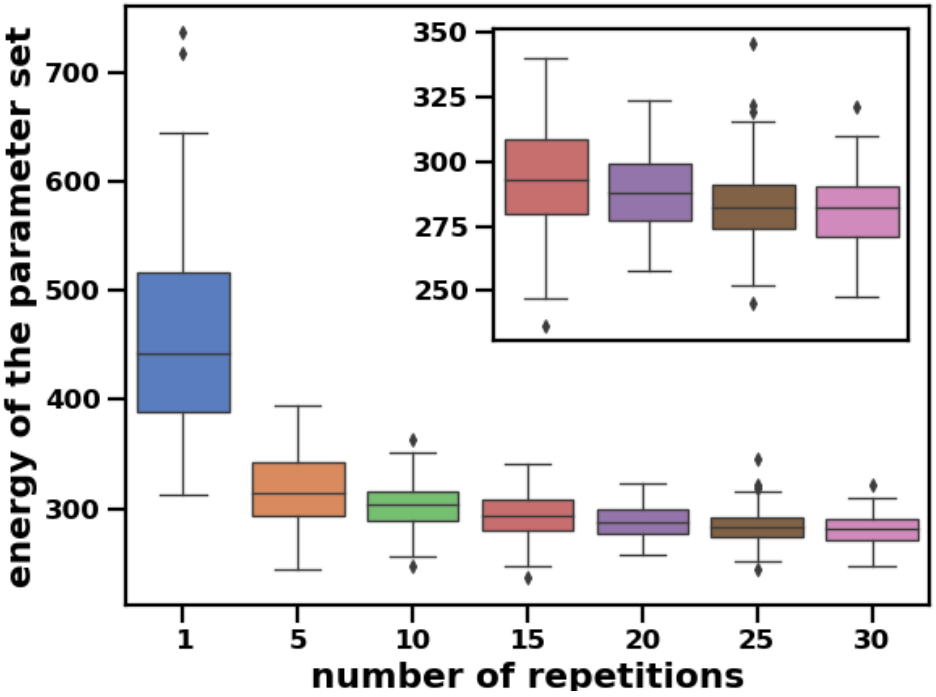

Figure 4.4: Energy dispersion of a given parameter set depending on number of repetitions

We settled on 20 realizations for each parameter set to ensure proper validation of the parameters. But as we are considering five different mutants over a span of at least 72 virtual hours, this increased quite a lot the computational cost of the exploratory algorithm. Due to this, we did not realize an extensive exploration of the parameter space with widely diverging initial parameter estimations and simply settled with the best set of data produced after 3 months of running time.

4.4 STUDY OF THE WILD TYPE AND THE $\Delta patS$ AND $\Delta hetN$ MUTANTS

In order to compare the results obtained from our numerical simulations with the experimental data from (Corrales-Guerrero et al., 2014b; Orozco et al., 2006), we have replicated the statistical analysis of these works. In both cases, all the data is aggregated for each experiment and then the averages and standard deviation are calculated between the experimental replicates. Thus, a certain amount of filaments has been aggregated in

batches and then averaged to obtain the standard deviation. On the experimental side, in (Corrales-Guerrero et al., 2014b), for each strain, 300 cells, or 100 intervals, were counted in three or four independent experiments. In (Orozco et al., 2006), the number of contiguous heterocysts at the ends of 50 filaments is averaged in three experiments. Alternatively, our data were obtained from 15 batches of 10 simulations for filaments with an initial size of 30 cells (consistent with experiments in (Burnat et al., 2014)) which grow to have around 50, 100, and 200 cells at 24h, 48h, and 72h respectively. Therefore, we present a bigger data sample than the one considered in (Corrales-Guerrero et al., 2014b) and of the same order of magnitude as the one in (Orozco et al., 2006).

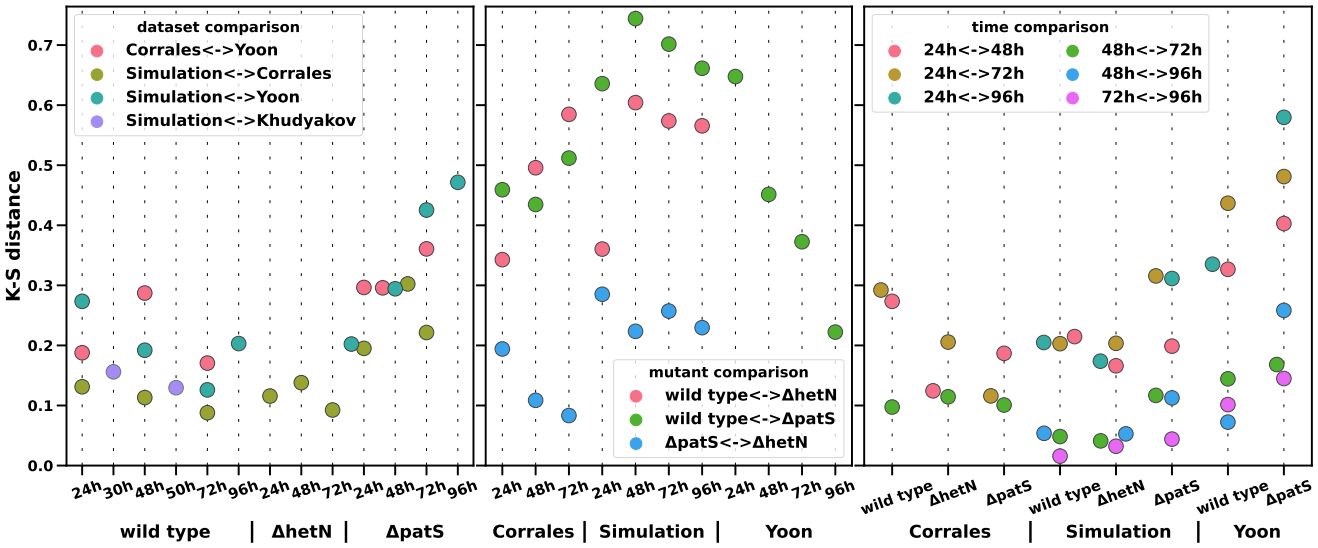

Figure 4.5: Kolmogorov–Smirnov distance between histograms of vegetative intervals length obtained with our model and the experimental datasets from (H. S. Yoon and J. W. Golden, 2001), (Corrales-Guerrero et al., 2014b) and (I. Y. Khudyakov and James W. Golden, 2004). The X-axis presents the common variables between the histograms and the legend the variable that we are comparing. In the first subplot, we compare the histograms of a given condition at a given time from two different datasets to evaluate the agreement of our model with the available experimental data. In the second, we obtain the histogram K-S distance between conditions from the same dataset at a given time. And finally, in the third subplot, we present the temporal change of a given condition for all the datasets considered.

The statistical distribution of vegetative cell intervals between heterocysts may differ from one experiment to another, as one can notice comparing the results from different authors (Borthakur et al., 2005; Corrales-Guerrero et al., 2013, 2014b; I. Y. Khudyakov and James W. Golden, 2004; Orozco et al., 2006; Risser and Sean M. Callahan, 2009; X. Wu et al., 2004; H. S. Yoon and J. W. Golden, 2001; Ho Sung Yoon and James W. Golden, 1998). For consistency, to compare our results with the experi-

mental data for the wild type and both the Δ*patS* and the Δ*hetN* mutants, we consider the relative frequency of vegetative intervals of a given length presented in (Corrales-Guerrero et al., 2014b), which is the most recent dataset available and one of the most comprehensive. Despite this, we also show in fig. 4.5 a systematic comparison of the Kolmogorov-Smirnov distance between the distribution of the vegetative interval distribution that our model produces, and the ones observed in (Corrales-Guerrero et al., 2014b; I. Y. Khudyakov and James W. Golden, 2004; H. S. Yoon and J. W. Golden, 2001) (which are the experimental datasets considered in (Muñoz-García and Ares, 2016) and in this same paper).

The agreement between the model and experimental data is very good, as one can also observe in fig. 4.5. There, it is shown that the differences between simulation data and experimental data are of the same order as temporal differences of the same mutant and considerably smaller than the differences between the wild type and both Δ*patS* and Δ*hetN*. Additionally, a closer inspection even reveals that the model produces histograms more similar to the ones presented in (H. S. Yoon and J. W. Golden, 2001) with respect to the ones from (Corrales-Guerrero et al., 2014b).

As already mentioned, in order to simplify the description we have modeled both genes *patX*, *patS* using a single variable. Thus, a complete loss of function of this variable represents the experimental Δ*patX*Δ*patS* mutant. This double mutant induces considerably more heterocysts than the single Δ*patS* mutant, fig. 4.6 top, as observed experimentally (I. Khudyakov et al., 2020) (see the Δ*patS*Δ*patX* double mutant movie). We have also simulated the Δ*patX* mutant (see the Δ*patX* mutant movie), where the production rate is only 10% of the combined PatS + PatX variable.

We have also simulated the Δ*patX* mutant (see the Δ*patX* mutant movie), where the production rate is only 10% of the combined PatS+PatX variable. The phenotype observed in fig. 4.6 down is still compatible with the wild type data, as reported in (I. Khudyakov et al., 2020), with both slightly shorter vegetative intervals and a higher percentage of heterocysts.

In fig. 4.7, we observe the agreement between the simulations and the experimental data from (Corrales-Guerrero et al., 2014b) for wild type and *patS* and *hetN* mutants. For the wild type, it is well known that roughly one of every ten cells differentiates, with a slight increase of this interval length with time (see the wild type movie). In the Δ*hetN* mutant, the initial pattern is similar to the wild type, except for the contiguous heterocysts appearing at a sequential pace (see the Δ*hetN* mutant movie). The Δ*patS* mutant shows a cluster of heterocysts, which appear simultaneously at short times (see the Δ*patS* mutant movie). For longer times,

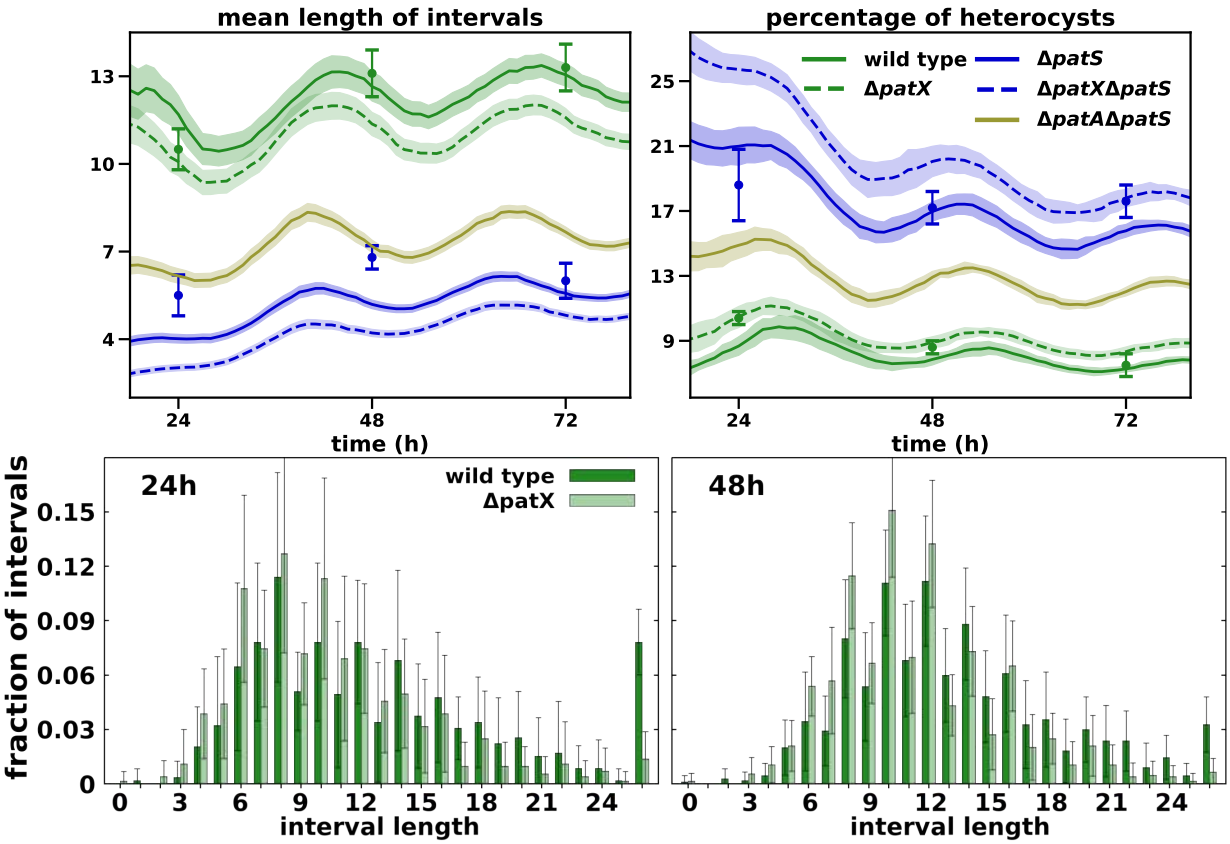

Figure 4.6: Upper plots: Mean number of vegetative cells between heterocysts and the total percentage of heterocysts in the filament for different conditions, as indicated. Symbols represent experimental values from (Corrales-Guerrero et al., 2014b), and lines are simulation results with their standard deviation as a shadowed area.
Lower plots: Comparison at different times after nitrogen deprivation (as indicated) between simulated histograms of the number of vegetative cells between heterocysts for wild type and Δ*patX*, as indicated. Bars are means of interval lengths, errors are standard deviations.

the pattern of heterocysts is more similar to the wild type, but with a higher incidence of contiguous heterocysts, in agreement with the experimental results reported in (H. S. Yoon and J. W. Golden, 2001) with a larger statistical sample.

These typical phenotype observations are also shown in fig. 4.8) where one can observe that both Δ*patS* and Δ*hetN* present multiple clusters of heterocysts. While the distribution for Δ*patS* remains the same through 24 to 72h, the amount of clusters longer than 3 heterocysts increases with time for the Δ*hetN* mutant.

The comparison of the distributions of the number of vegetative cells between heterocysts of the simulated filaments and the experimental data in (Corrales-Guerrero et al., 2014b) is shown in fig. 4.9.

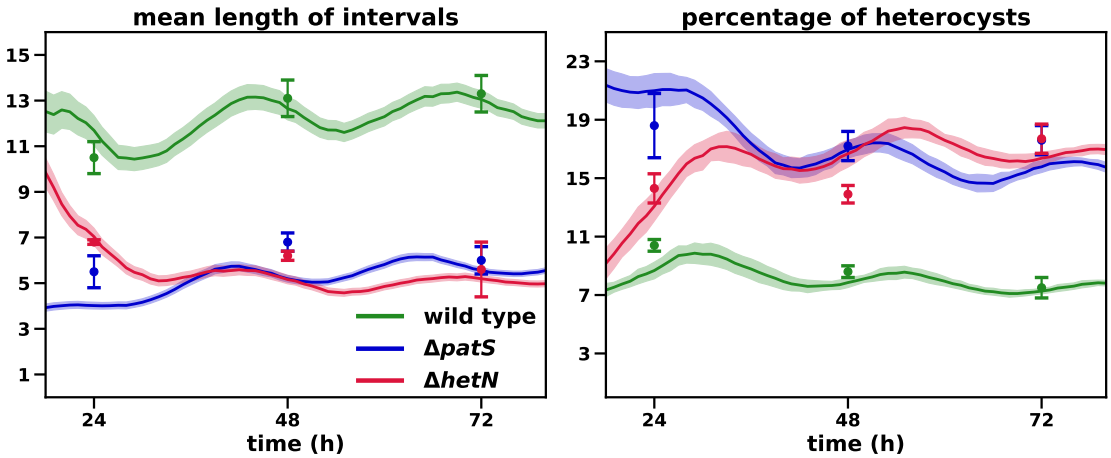

Figure 4.7: Mean number of vegetative cells between heterocysts and the total percentage of heterocysts in the filament. Symbols represent experimental values from (Corrales-Guerrero et al., 2014b) and lines are simulation results with their standard deviation as a shadowed area.

A small deviation appears for the early phenotype of the Δ*patS* mutant, especially at 24h. The simulations Δ*patS* to present less contiguous heterocysts and shorter intervals than both the experimental data (fig. 4.9) and the Δ*hetN* mutant simulation (fig. 4.8) at this time. This difference could be due to the effect of not considering a protoheterocyst phase. Without this phase, the differentiation of adjacent cells is strongly reduced because, once a cell differentiates, it immediately starts producing both HetN and nitrogen products, which inhibit differentiation. Thus, an artificial surplus of one and two-cell intervals is observed in the first round of differentiation. After the first round of division, this causes the observed peaks of two and four-cell intervals observed at 24h, as observed fig. 4.9 and the Δ*patS* mutant movie.

The low amount of contiguous heterocysts observed in wild type simulations in comparison with experimental data (Corrales-Guerrero et al., 2014b) can be explained using the same argument. A protoheterocyst phase would ease the stochastic formation of multiple simultaneous heterocysts in all genetic backgrounds. On top of that, it is worth noting that these contiguous heterocysts are seldom described in the literature. Their appearance on the wild type can be explained by stochastic fluctuations of the genetic expression that get fixed through the irreversible process that is the differentiation into a heterocyst. In any case, the model also allows for the formation of these contiguous heterocysts, albeit in a much smaller proportion.

These results also improve the phenotypes obtained in the minimal model from (Muñoz-García and Ares, 2016). In that model, the Δ*hetN*

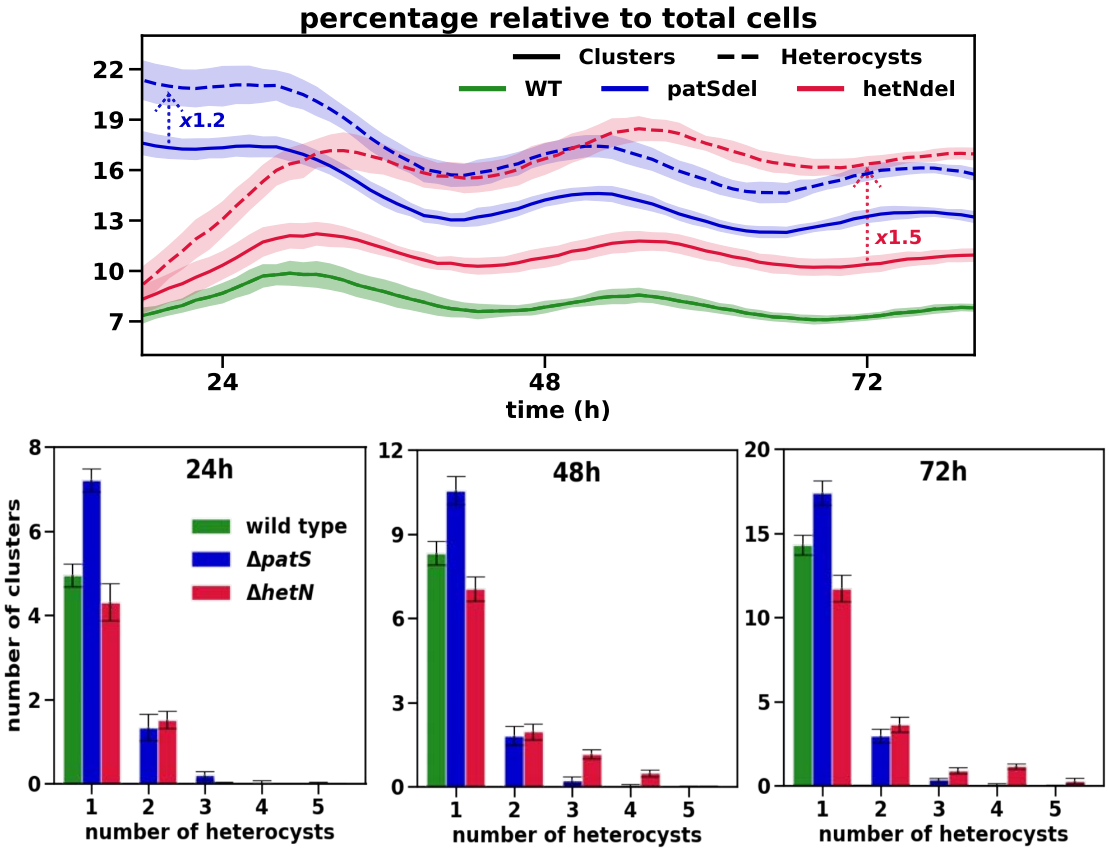

Figure 4.8: The upper plot shows the evolution of the percentage of clusters of heterocysts (solid lines) and individual heterocysts (dashed lines). The lower plots show the histogram of the number of clusters of a given size observed at 24h, 48h, and 72h. Bars are the mean number of clusters of each size at each time, error bars are standard deviation.

mutant showed both smaller clusters of heterocysts and shorter intervals between them. The frequency of the vegetative intervals for longer times (48 and 72 hours) has a much stronger inverse dependency with the interval longer than the one observed experimentally (Corrales-Guerrero et al., 2014b). While the experimental data presents a Bell-shaped distribution around the 6-8 cells intervals with an abnormally increased frequency for contiguous heterocysts, the model distribution is more akin to two different decays (one for even intervals and another for odd intervals that get more different as the time increase) of the interval frequency in regard to their length.

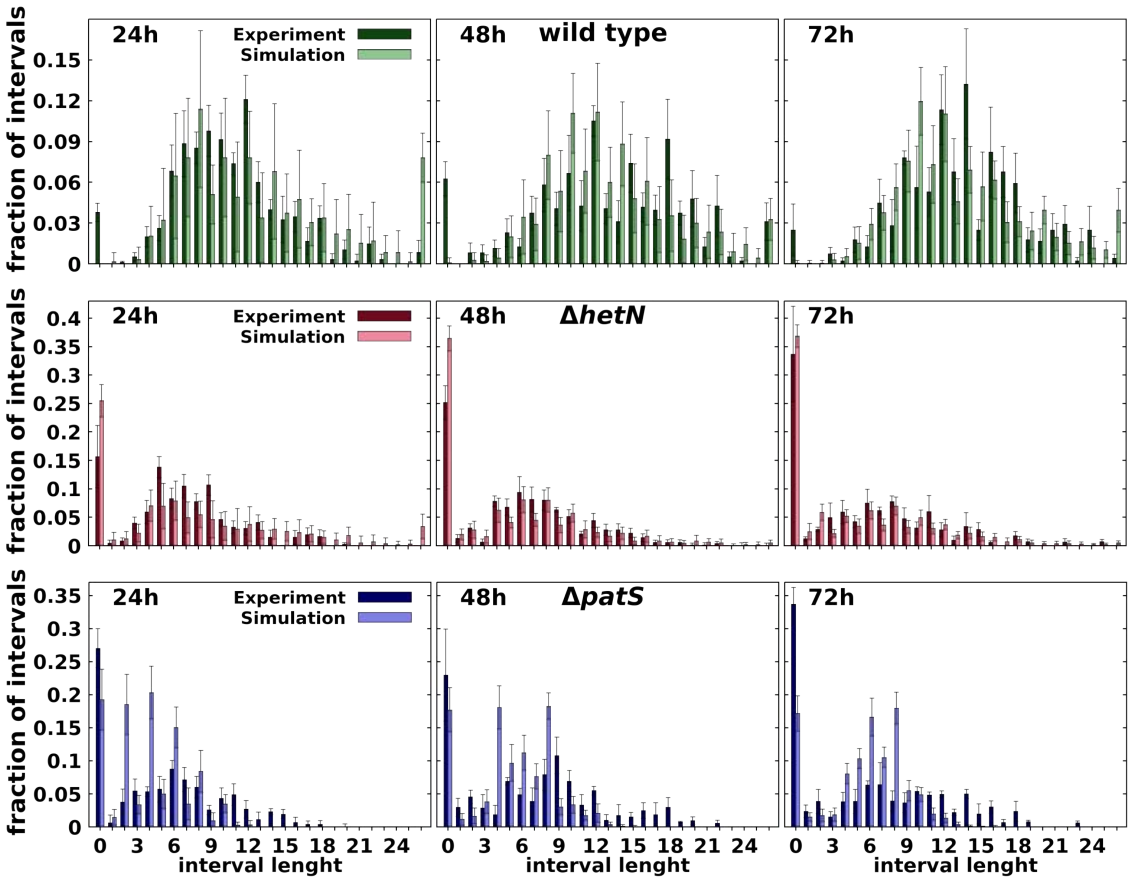

Figure 4.9: Comparison at different times after nitrogen deprivation (as indicated) between experimental (Corrales-Guerrero et al., 2014b) and simulated histograms of the number of vegetative cells between heterocysts for wild type (top row), Δ*hetN* (middle row), Δ*patS* (bottom row). Bars are means of interval lengths, errors are standard deviations.

4.5 LOSS OF HETEROGENEITY IN THE HETR CONCENTRATION PROFILE IN A Δ*patA* MUTANT BACKGROUND

The simulations for the Δ*patA* single mutant show a similar behavior to experimental results (Liang et al., 1992; Orozco et al., 2006; Risser and Sean M. Callahan, 2008; Risser et al., 2012; Young-Robbins et al., 2010), with heterocysts forming mostly on the filament ends despite higher global HetR concentration in the filament (approximately 24% higher in our simulations, fig. 4.10) in Δ*patA* than in the wild type.

One can also see in fig. 4.10 that HetR concentration in vegetative cells is only oscillates in those conditions that present heterocyst differentiation mainly in the central part of existing vegetative intervals. This is because the increase of HetR concentration that leads to heterocyst differ-

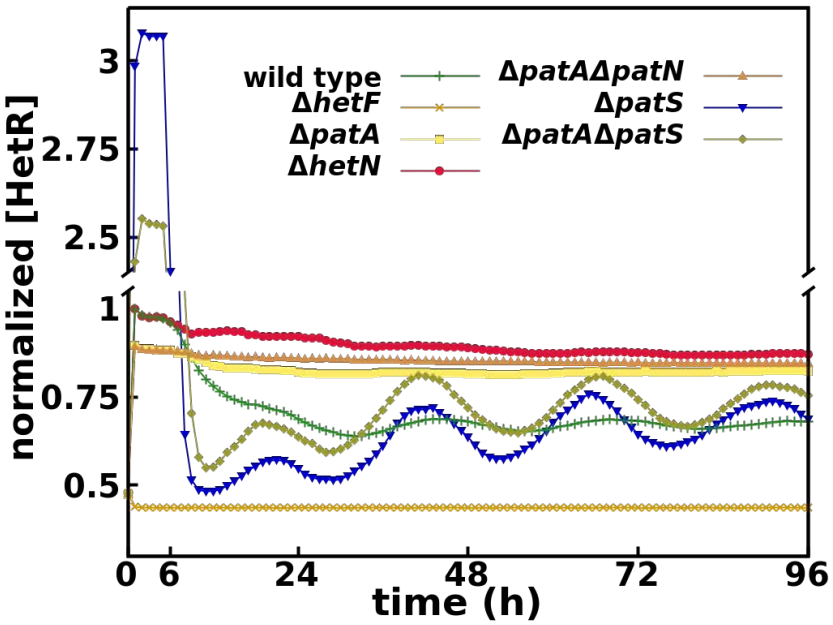

Figure 4.10: Time evolution of HetR concentration on vegetative cells for all mutants considered, normalized by the maximum wild type concentration.

entiation is then dependent on an elongation of an existing interval that is coupled to the cell cycle synchronicity. A cell has to be far enough from a heterocyst to be able to self catalyze HetR production, as can be observed in the shape of the regulatory function for HetR fig. 4.3. This is the main reason why both the Δ*patS* and Δ*patA*Δ*patS* mutants, which present the formation of multiple simultaneous heterocysts, present a much higher oscillation than the wild type, which usually just produces one heterocyst. The rest of mutants does not form internal heterocysts (Δ*hetF*, Δ*patA*, Δ*patA*Δ*hetN* mutants) or present heterocysts mainly next to already existent heterocysts (Δ*hetN* mutant), which, in either case, are uncoupled from the cell cycle.

The absence of internal heterocysts in Δ*patA* is caused by the loss of a distinct HetR concentration profile, fig. 4.11 and the Δ*patA* mutant movie.

Without PatA in the system, the conversion to active HetR is slowed to a minimum level, and this also slows down the regulatory effect of HetR over both itself and *patS*. This produces a homogenization of the production of PatS and also of the inhibitory hexapeptide along the filament. Then, in the absence of a pronounced inhibitory gradient, the levels of HetR increase uniformly to levels close to the threshold for differentiation (fig. 4.11). In these conditions, the selection of the few internal cells that will differentiate is exclusively due to stochastic fluctuations in the

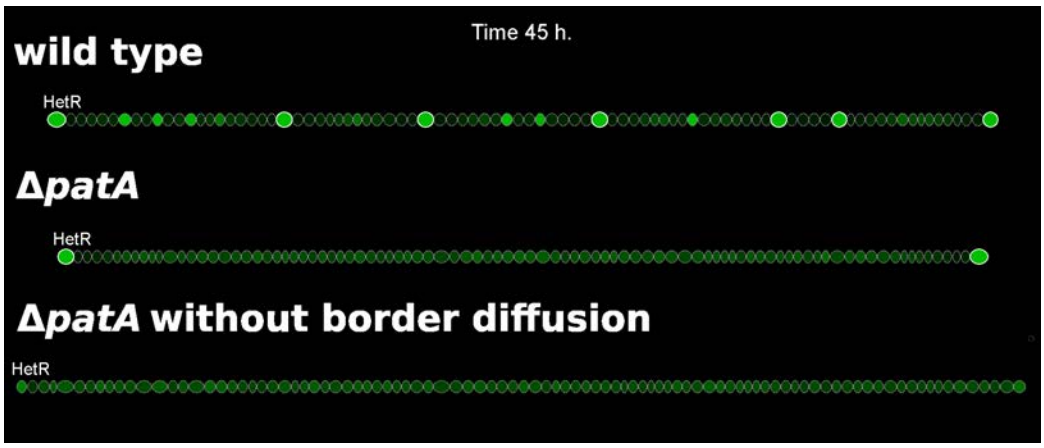

Figure 4.11: Three simulations showing HetR profiles in filaments of wild type, Δ*patA*, and Δ*patA* with no inhibitor leakage from the terminal cells ($d_{broder} = 0$). HetR concentration is represented by the brightness of green, and the heterocysts present an additional white cell wall. For the wild type, one can observe cells with much higher levels of HetR, those are the candidates to differentiate into heterocysts.

protein production of both HetF and PatS. On the other hand, the model assumes a certain passive diffusion of both the inhibitory hexapeptide and fixed nitrogen through the filament ends, which causes a local reduction of the inhibitory signals, allowing the differentiation of the boundary cells. If one does not allow the diffusion through the border cells, heterocysts do not form in the filament ends as HetR does not accumulate enough on those cells, fig. 4.11.

A simple physical analogy of the importance of the boundary condition to understand the *patA* phenotype can be made using the following continuous reaction-diffusion model defined on a filament of length L where the position is denoted by the coordinate $x \in [0, L]$:

$$\frac{\partial r(x,t)}{\partial t} = \beta_r + f(r,s)\rho_r - \alpha_r r \tag{4.14}$$

$$\frac{\partial s(x,t)}{\partial t} = f(r,s)\rho_s - \alpha_s s + d\frac{\partial^2 s}{\partial x^2}, \tag{4.15}$$

where r is the concentration of a non-diffusing activator (HetR), s the concentration of a diffusible inhibitor, β_r a basal production for the activator ($\beta_r > 0$), $f(r,s)$ a smooth regulatory function, monotonically increasing in r and decreasing in s, and the other symbols are parameters.

These equations need to be supplemented with boundary conditions for s. If the inhibitor cannot diffuse across the filament's boundaries, the condition is:

$$\left.\frac{\partial s}{\partial x}\right|_{x=0,L} = 0. \tag{4.16}$$

With this condition, we assume that the system is such that a parameter regime exists where there is a stable homogeneous solution. However, if leakage of the inhibitor through the boundaries is possible, the relevant boundary condition is:

$$s|_{x=0,L} = 0. \tag{4.17}$$

With this alternative condition and under the same parameter regime, there would be a gradient of inhibitor decreasing from the center to the boundaries of the filament, which in turn would produce a profile of activator with maxima in the boundaries, making them the favored location for heterocyst differentiation. This simple analogy explains the physical mechanism behind the boundary-induced pattern observed in the *patA* mutant, and agrees qualitatively with the observation in the discrete model depicted in fig. 4.11.

The results of the simulation for the Δ*patA*Δ*hetN* double mutant also show a phenotypic agreement with observations (Orozco et al., 2006). The filaments present multiple terminal heterocysts with only the occasional internal heterocyst (see the Δ*patA*Δ*hetN* double mutant movie). Additionally, the model predicts a higher HetR concentration (approximately 27% higher in our simulations, fig. 4.10) in Δ*patA*Δ*hetN* than in the wild type, equivalent to the one observed in Δ*patA*. Here the border effect of the inhibitor diffusion through the ends of the filament gets propagated to multiple cells because the Δ*patA*Δ*hetN* double mutant, besides the homogenization of the HetR concentration, does not present the inhibitory gradient around existent heterocysts produced by *hetN*.

In fig. 4.12, we present the temporal evolution of the amount of terminal heterocyst for both Δ*patA* and Δ*patA*Δ*hetN* mutants. These results are in agreement with experimental data (Orozco et al., 2006) as shown in fig. 4.12. The model seems to have a small delay in the formation of heterocysts (also present in the Δ*patA* mutant), but except for early times the model shows a good agreement with the experimental data. This delay could be related to the mechanism of commitment to the differentiation of a given cell. In our model, this decision is exclusively linked to

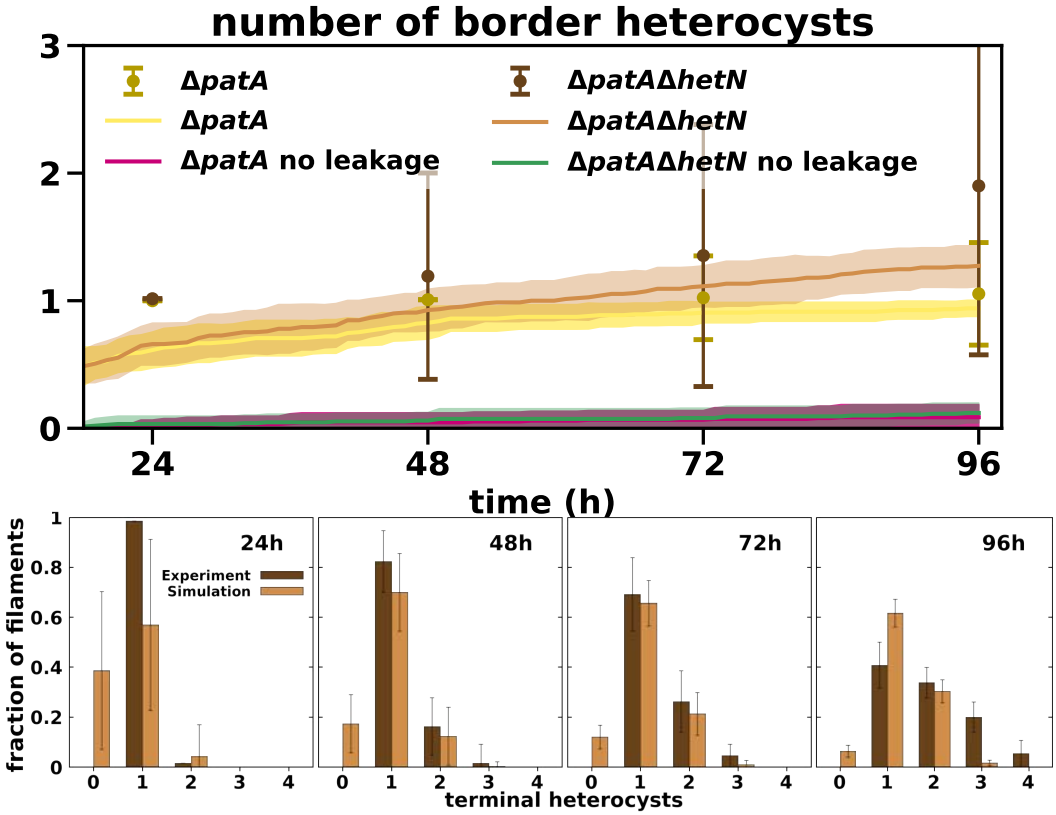

Figure 4.12: The upper plot shows the temporal evolution of the number of heterocysts in the Δ*patA* and Δ*patA*Δ*hetN* mutants. Symbols represent experimental values from (Orozco et al., 2006) and lines are simulation results with their standard deviation as a shadowed area.
The lower plots present the comparison at different times after nitrogen deprivation (as indicated) between experimental Orozco et al., 2006 and simulated histograms of the number of heterocysts in the filaments end for the Δ*patA*Δ*hetN* double mutant. Bars are means of the number of birder heterocysts, errors are standard deviations.

a sustained high concentration of HetR and not to the expression of a supplementary gene (*hetP* and/or *hetZ*) as recent publications (Yaru Du et al., 2020; Ye Du et al., 2012; Higa and Sean M. Callahan, 2010; Videau et al., 2016, 2018; H. Zhang et al., 2018) seem to indicate. Hence, incorporating a gene-controlled differentiation commitment would surely improve these results, because the self-regulation of that gene could amplify the differentiation signal.

Finally, we studied the Δ*patA*Δ*patS* double mutant, of which to our knowledge there are only phenotypical observations (Orozco et al., 2006). Its phenotype is described as similar to the single Δ*patS* mutant phenotype, but with longer distances between heterocysts.

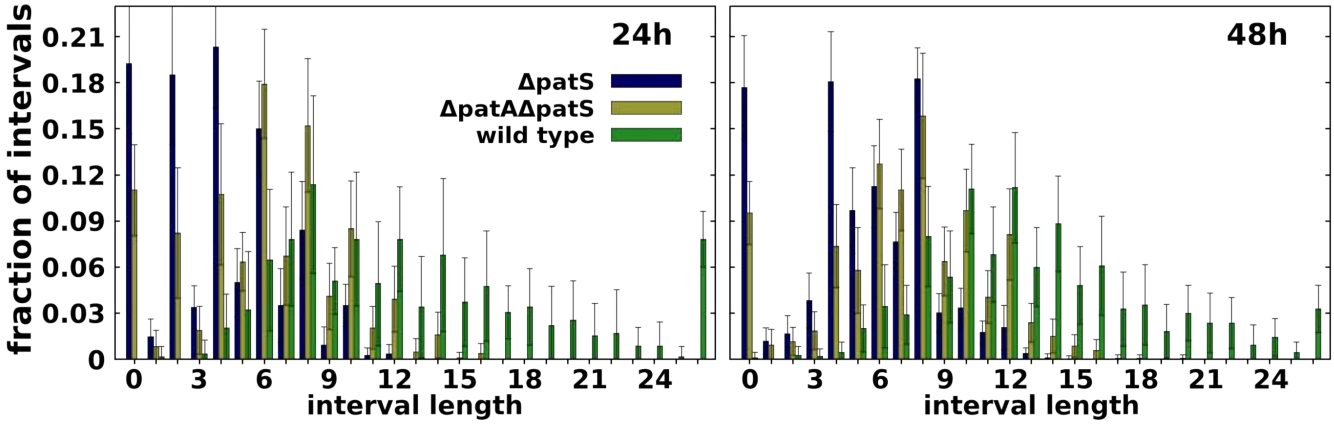

Figure 4.13: Comparison at different times after nitrogen deprivation (as indicated) between simulated histograms of the number of vegetative cells between heterocysts for wild type, Δ*patS* and Δ*patA*Δ*patS*, as indicated. Bars are means of interval lengths, errors are standard deviations.

The simulated filament (see the Δ*patA*Δ*patS* double mutant movie) fits this behavior, fig. 4.6 and fig. 4.13. One can notice that the Δ*patA*Δ*patS* mutant presents a higher amount of contiguous heterocysts, but with a smaller frequency and larger intervals of vegetative cells between heterocysts. Additionally, the simulated filament does not present an increase in HetR concentration typical to other *patA* mutants. This also confirms that the full Δ*patA* phenotype, which shows almost no internal heterocysts and a high concentration of HetR, requires a functional *patS* gene (Orozco et al., 2006).

As we discussed previously, the increase in the HetR concentration is produced due to the homogenization of the PatS production along the filament, which produces a state where all the cells have a homogeneous HetR concentration lower than the decision threshold. Therefore, if *patS* is not functional this effect would not be observed and the weak activation of HetR due to the absence of PatA precludes the formation of as many clusters of heterocysts as one observes in the Δ*patS* mutant.

One can observe this in fig. 4.10, where the HetR concentration for this mutant does not behave like the Δ*patA* single mutant or the Δ*patA*Δ*hetN* double mutant, with high constant values of HetR. It behaves similarly to the simple *patS* mutant, with a slightly higher concentration than that mutant after the first outburst of HetR that marks the first differentiation round (Wong and J. C. Meeks, 2001). On the other hand, the concentration is higher in the Δ*patS* single mutant during the first round of differentiation. This inversion is due to the faster production of HetR with a functional *patA*. Therefore, the decision to differentiate is reached in a shorter time than in the Δ*patA*Δ*patS* mutant. Hence, on a homogeneous initial condition (the first round of differentiation), the concentration would be

higher in the Δ*patS* mutant because all cells start producing HetR at a much faster rate. But starting from heterogeneous initial conditions (all the following differentiation rounds), the concentration would be higher in the double mutant. Due to the slower commitment, more cells (closer to the heterocysts) start producing HetR before being shut down by the newly formed heterocysts. This effect can also be observed by comparing the Δ*patS* mutant movie and the Δ*patA*Δ*patS* double mutant movie.

4.6 COMPLETE LOSS OF HETR REGULATORY FUNCTION IN THE Δ*hetF* MUTANT BACKGROUND

If HetF is necessary to produce the active form of HetR, in its absence HetR should lose its regulatory function. This prediction is in agreement with experimental observations in *Nostoc punctiforme* (Wong and J. C. Meeks, 2001), which presents a similar transcriptional induction pattern of *hetR* than *Anabaena* PCC 7120. There, the induction of *hetR* is dependent on the presence of an intact copy of *ntcA*. Moreover, the induction of *hetR* is still present in the Δ*hetF* background, but with an altered induction pattern that eliminates the 0 to 6h burst.

Our simulated Δ*hetF* successfully eliminates the initial burst of HetR production observed in all other mutants, fig. 4.10. The absence of this burst in Δ*hetF* implies that it is exclusively produced by the positive self-regulation of *hetR* (Risser and Sean M. Callahan, 2008; Wong and J. C. Meeks, 2001). Due to this, the introduction of an additional *hetR* promoter activated as a response to nitrogen deprivation would improve the agreement between our model and experiments by increasing only the overall HetR concentration in our simulations (especially on the Δ*hetF* and Δ*patA* simple and multiple mutants) without altering much of the dynamics.

4.7 NOISE EFFECT AND SENSITIVITY ANALYSIS

The effect of the noise amplitude can be observed in fig. 4.14. For low noise in cellular growth Ω_Λ, oscillations appear in the time evolution of the mean number of vegetative cells between heterocysts. This effect is due to the effect of nearly synchronized cell divisions and was discussed in detail in (Muñoz-García and Ares, 2016).

As this noise increases, the system loses synchronization between cells, and therefore both the mean vegetative interval and the percentage of heterocysts are more stabilized to roughly the mean value that one observes in the less noisy regimes.

The genetic regulatory noise Ω_Φ affects the heterocyst pattern more than the noise in cell growth. This is because a certain genetic noise is

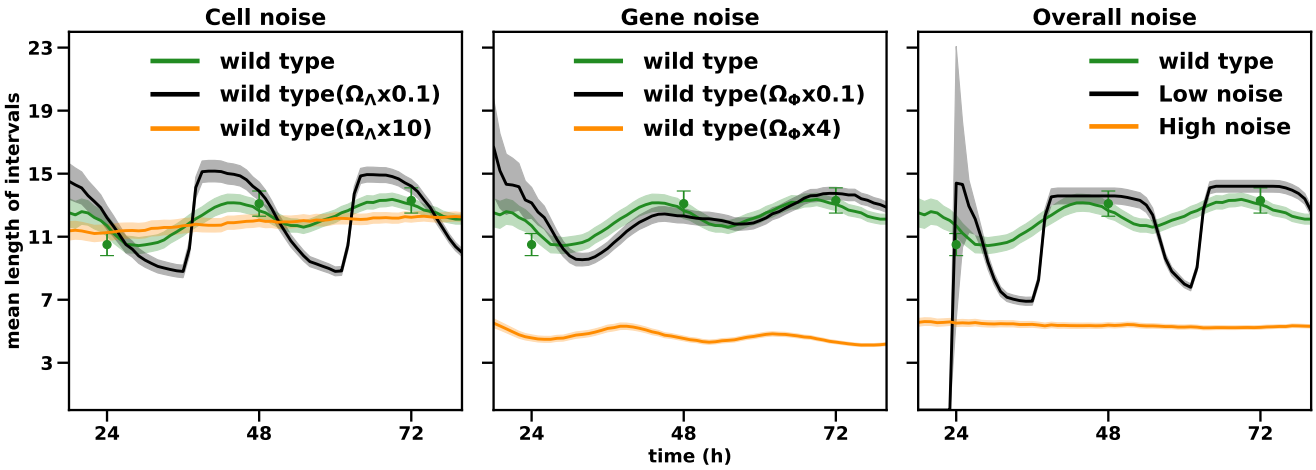

Figure 4.14: **Effect of noise on the mean interval of vegetative cells between heterocyst for wild type parameters.** Symbols represent experimental values from (Corrales-Guerrero et al., 2014b), and lines are simulation results with their standard deviation as a shadowed area.

required for the pattern formation. Therefore, reducing the noise does not seem to affect the long-term stability of the pattern, but it delays its inception to much later on. Alternatively, increasing the noise does not modify the early pattern, but it produces considerably more heterocysts which reduces the overall length of the intervals. Finally, if one increases the two noises by the same rate, one observes that the growth noise Ω_Λ effect dominates for the low overall noise regime while the regulatory noise Ω_Φ effect is the dominant one.

We also studied how the effect of the gene expression noise changes with its intensity.

One can see in fig. 4.15 that while the deterministic gene regulation dominates the noise, up to the point marked with the dotted line, the noise facilitates the formation of the pattern. Given that, the heterocysts form much easier due to higher fluctuations in protein production. Then, more cells differentiate into heterocysts, and the length of vegetative intervals decreases. But once the noise dominates over the deterministic gene expression, the number of heterocysts decreases and the vegetative intervals increase in both their mean length and variance. Once the noise dominates, the cells are no longer capable of effectively responding to the external signaling. This produces an uncoupling of the filament cells, and the differentiation to a heterocyst becomes a completely independent stochastic process.

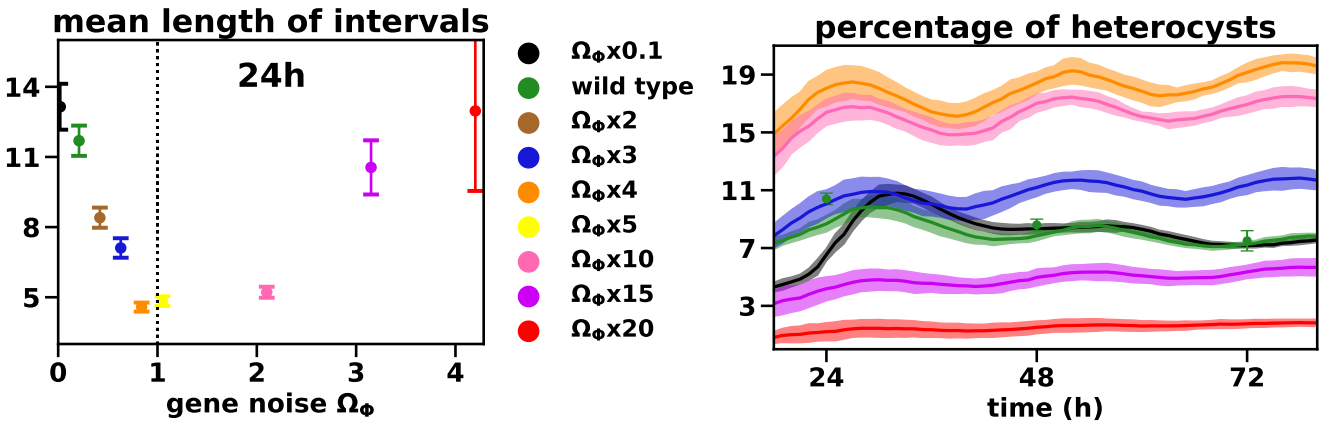

Figure 4.15: **Effect of gene expression noise on the pattern:** The left plot shows the effect of several levels of gene expression noise on the mean interval of vegetative cells between heterocysts for wild type parameters at 24 hours. The dotted line marks the transition when the noise dominates over the deterministic gene expression. The right plot shows the evolution of the effects of the same levels of gene expression noise on the percentage of heterocysts.

Sensitivity analysis

To assess the robustness of our results, we have performed a sensitivity analysis following the approach in (Muñoz-García and Ares, 2016). We calculate the sensitivity of the model to a given parameter X by evaluating the observable Y at two points, the wild type value X_0 and the perturbation $X = X_0 + \Delta X$. Using the resulting change in the observable, ΔY, we calculate the sensitivity S_{YX} of the observable to the parameter as:

$$S_{YX} = \frac{\Delta Y/Y}{\Delta X/X}. \tag{4.18}$$

In fig. 4.16, we show the sensitivity of the mean distance between contiguous heterocysts to changes of 10% in the parameters.

The results are qualitatively similar to our previous work (Muñoz-García and Ares, 2016), however, we find that the extension of the model has made it even more robust to variations in individual parameters. The largest sensitivity is found when modifying HetR production and degradation, followed by inhibitor degradation, the strength of the inhibition, and PatS production. For all other parameters, the relative changes in mean interval length are much smaller than the relative change in the parameter.

Additionally, our results are robust to the details of eq. (4.7). For instance, it has also been observed that the tetramers of HetR are capable to

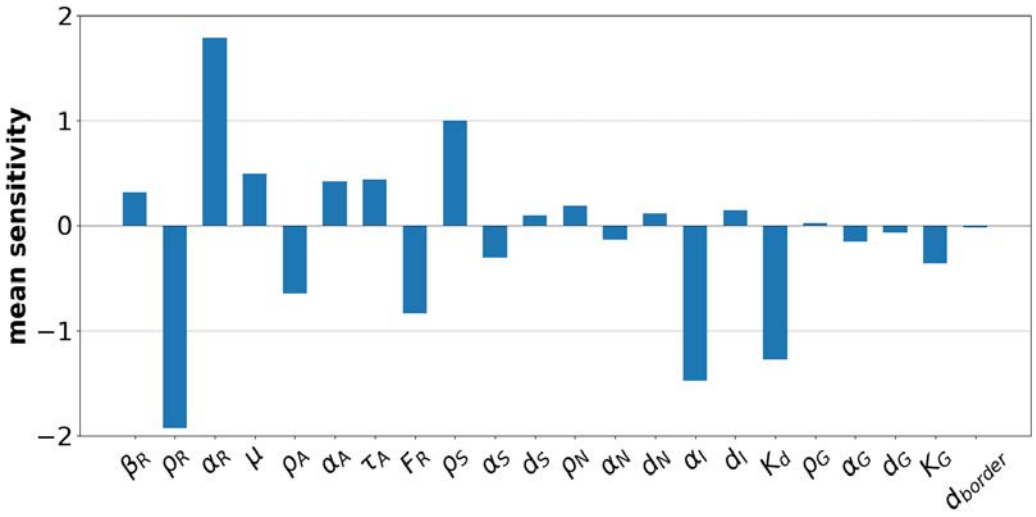

Figure 4.16: Sensitivity of the mean distance between heterocysts after 72h with respect to 10% changes in the indicated parameters. Changes are with respect to the wild type values in table 4.1.

bind to DNA (Hu et al., 2015; Kim et al., 2013) and its abundance increase during the transition (Valladares et al., 2016). But it is still not clear how the ERGSGR interact with them and therefore if it has a clear role during the transition. Despite this, we have checked that this alteration does not change the qualitative system behavior and the same dynamics can be recovered assuming tetrameric binding with an alternative set of parameter values

$$r(R_j,A_j,I_j,G_i) = \frac{13F_R R_j^4\left(1+\frac{A_j}{\tau_A}\right)}{1+13F_R R_j^4\left(1+\frac{A_j}{\tau_A}\right)+\frac{I_j^2}{K_d^2}+\frac{G_j}{K_G}}, \tag{4.19}$$

allows for a similar fit to experimental data fig. 4.17 with only one extra 13 factor multiplying the F_R.

Allowing details like affinities or exponents to be different for the different species regulated by HetR would only increase the number of free parameters of the model, so for simplicity and to reduce the risk of overfitting we have made the choice of using the same function in all cases.

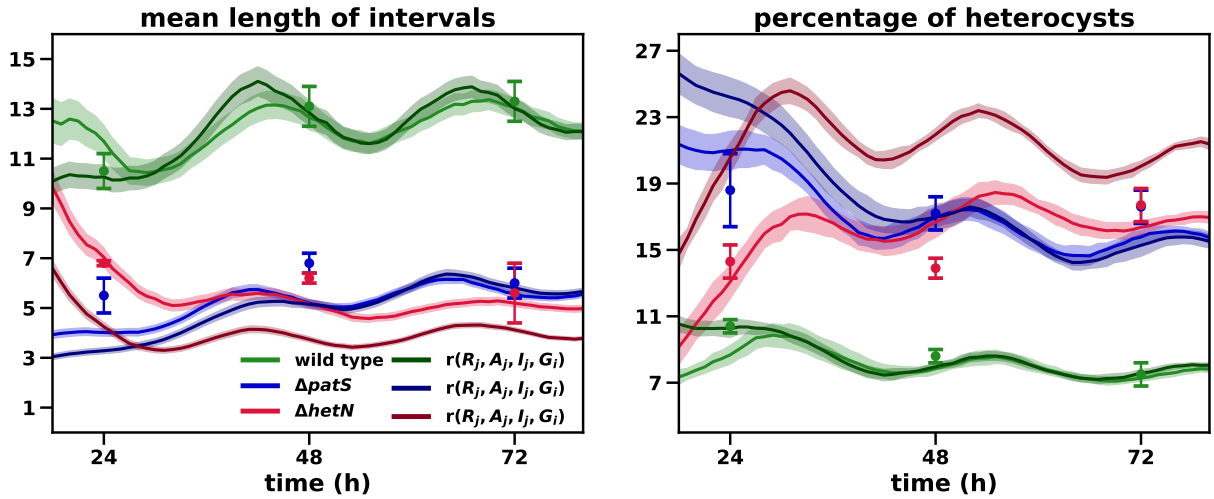

Figure 4.17: Mean number of vegetative cells between heterocysts and the total percentage of heterocysts in the filament for both the dimer and the tetramer active form presented in eq. (4.19). Symbols represent experimental values from (Corrales-Guerrero et al., 2014b), and the lines are simulation results with their standard deviation as a shadowed area.

4.8 CONCLUSION

The formation and maintenance of heterocyst patterns is a paradigmatic example in which many processes, such as complex gene regulatory networks, interactions at different time scales, molecular trafficking, and cell growth, act together to give rise to a multicellular pattern. All these aspects form an intriguing puzzle for which a complete understanding is still elusive. A practical way to expand our knowledge about this problem is to investigate what are the functions of some of its pieces. Thus, increasing the complexity of a previous minimal model, we have been able to gain insight into the functions of the players involved. We have focused on the *patA* gene, and based on experimental phenotypical evidence, we have formulated a mathematical model that includes the interactions of this gene with the key genes responsible for heterocyst pattern formation.

Our model is not identifiable because several sets of parameters produce similar behaviors, and most of the phenotypic changes that the modification of a given parameter value produces can be restored by tuning a subset of the rest of the parameters. Nevertheless, we never intended to have an identifiable model because the data available is quite scarce and is always phenotypic (distances between heterocysts at a given time) while these parameters act on the molecular scale. In order to have an identifiable model, we would need much more data and ideally of the molecular scale (concentration of a given protein or a quantification of

the gene transcription). By working with just phenotype data with a time step of 24h, we consider that one cannot extract conclusions regarding the parameter set and therefore only argue about the validity of the model as a whole. Our exploratory optimizations indicate that good fits lay on a manifold of the parameter space, producing equivalent phenotypical predictions for different parameter sets. As the model is robust to parameter changes, if in the future molecular data is available we could readjust the parameters of our model and then use it to extract predictions about the values of the unknown parameters.

The model shows that considering PatA as a collaborator of the activation process of HetR directed by *hetF* is capable of explaining all the phenotypes of the genes considered in our genetic network. This agreement suggests that there is some interaction, direct or indirect, between *hetF* and *patA* that has not been reported experimentally. This consideration, together with the existence of an active form of HetR, is also enough to account for the paradoxical changes in HetR concentration in the *patA* mutant, which seemed to question the central role of *hetR* in heterocyst differentiation.

New experiments are required to confirm the validity of the interactions proposed. In particular, experimental information regarding protein translation (Corrales-Guerrero et al., 2015; Di Patti et al., 2018) could be useful in order to have more detailed information regarding the effects of a given gene on the regulatory network. The roles of many other actors are still to be elucidated and could be included in the core processes to obtain a more extensive mathematical description. For example, recent work (Xiaomei Xu et al., 2020) presents evidence that *hetL*, a gene previously shown to alter PatS mediated inhibition of heterocyst differentiation (D. Liu and James W. Golden, 2002), interacts with HetR without inhibiting its DNA-binding activity in *Nostoc PCC 7120*. This interaction protects HetR from the inhibitory effects of the Pat-derived hexapeptide and seems to be essential for the proper function of HetR as a genetic regulator.That would be the role that we have assigned to hetF based on the phenotypic evidence on *Anabaena PCC 7120*.

A recent study (W.-Y. Xing et al., 2022), published almost at the same time that this model, already confirmed experimental evidence of an indirect activator role for HetF over *hetR*. This paper shows that HetF regulates cell division and heterocyst differentiation by controlling the inhibitory effects of PatU3. The information regarding this gene is still pretty scarce, and therefore, it is difficult to locate this gene in the network presented in fig. 2.5. Despite this, a previous work by (Yaru Du et al., 2020) situates it in close relationship with *hetZ* to actively regulate *hetR*

and *patS*/*patX*. This indirect effect could account for the one predicted by the model if *patA* is also involved in this PatU3 regulation.

Finally, our work shows that it is possible to reproduce the *patA* mutant phenotype without considering a differentiation mechanism depending on a cell's position on the filament. The model also expands the characterization presented in (Orozco et al., 2006) by directly linking the increase of the HetR concentration in all the cells with the absence of internal heterocysts in both Δ*patA* and Δ*patA*Δ*hetN* mutants. This is obtained by a slowing of the *hetR* transcription, which produces a homogenization of HetR concentration through the filament. Then it is easy to see why this phenotype is not present in the Δ*patA*Δ*patS* mutant, where this reduction in the transcription rate is completely compensated by the absence of the early inhibitor PatA. The intriguing differentiation of only terminal heterocysts appears as a consequence of the boundary conditions of the system: leakage of inhibitors out of the filament through the terminal cells. Hence, despite the apparent simplicity of *Anabaena* compared to other developmental systems, it is already clear that genetic and metabolic interactions result in patterns shaped by physical constraints.

5

FILAMENT CORRELATION IN THE *patAhetF* MODEL

After the validation of the *patAhetF* model with the available experimental data in the previous chapter. Here we will continue the analysis of the model studying the correlation between cells of a filament. We decided to study the correlations because they should reflect the coordination between cells required to self-organize into a pattern.

We will use the same model formulation and parameter set presented in section 4.3 with the only exception that we will initiate the filament with much more cells. Instead of starting with the more realistic 30 cells filaments, we will start the simulations with 5000 cells. This change reduces the border effect on our analysis and will ensure that we have enough statistics for all the cell distances considered, from 1 up to 20 cells. Then for the lowest data sample, which would correspond to the initial time, before the first cell division, for the 20 cell intervals, we would have 250 intervals to calculate the correlation factor. Additionally, we will consider 10 independent filaments to calculate the correlation values, and then we will average the correlation coefficient obtained for each filament.

We tried several correlation factors, the linear correlation with Pearson and also rank correlation with both Spearman and Kendall. But since we did not observe notable differences between them, we will only present here the results for the Pearson coefficient. The Pearson correlation coefficient between two datasets x and y is defined as the covariance between x and y normalized over the product of their variances:

$$Pc(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y}.$$

In our particular case, we will have that the Pearson correlation coefficient for a cell distance d we will compare all the cells with the ones at a distance d with

$$Pc(d) \equiv \left\langle \frac{Cov\left(v|^{-d}, v|_d\right)}{\sigma_v^2} \right\rangle_{filaments}$$

where v is the vector of concentrations in each cell and the $|_x^y$ is used to represent a subvector from the x index until the y index. Then $v|^{-d}$ is the vector v without the last d cells and $v|_d$ is the vector v without the first d cells. Finally, σ_v^2 is the variance of the concentration in the whole filament.

5.1 CORRELATION LENGTH OF THE INHIBITOR GRADIENT DEFINES PATTERN LENGTH

First, we will study the correlation of the ERGSGR inhibitor. As it is the only diffusible element of our model, we expect that is the element with the highest correlation.

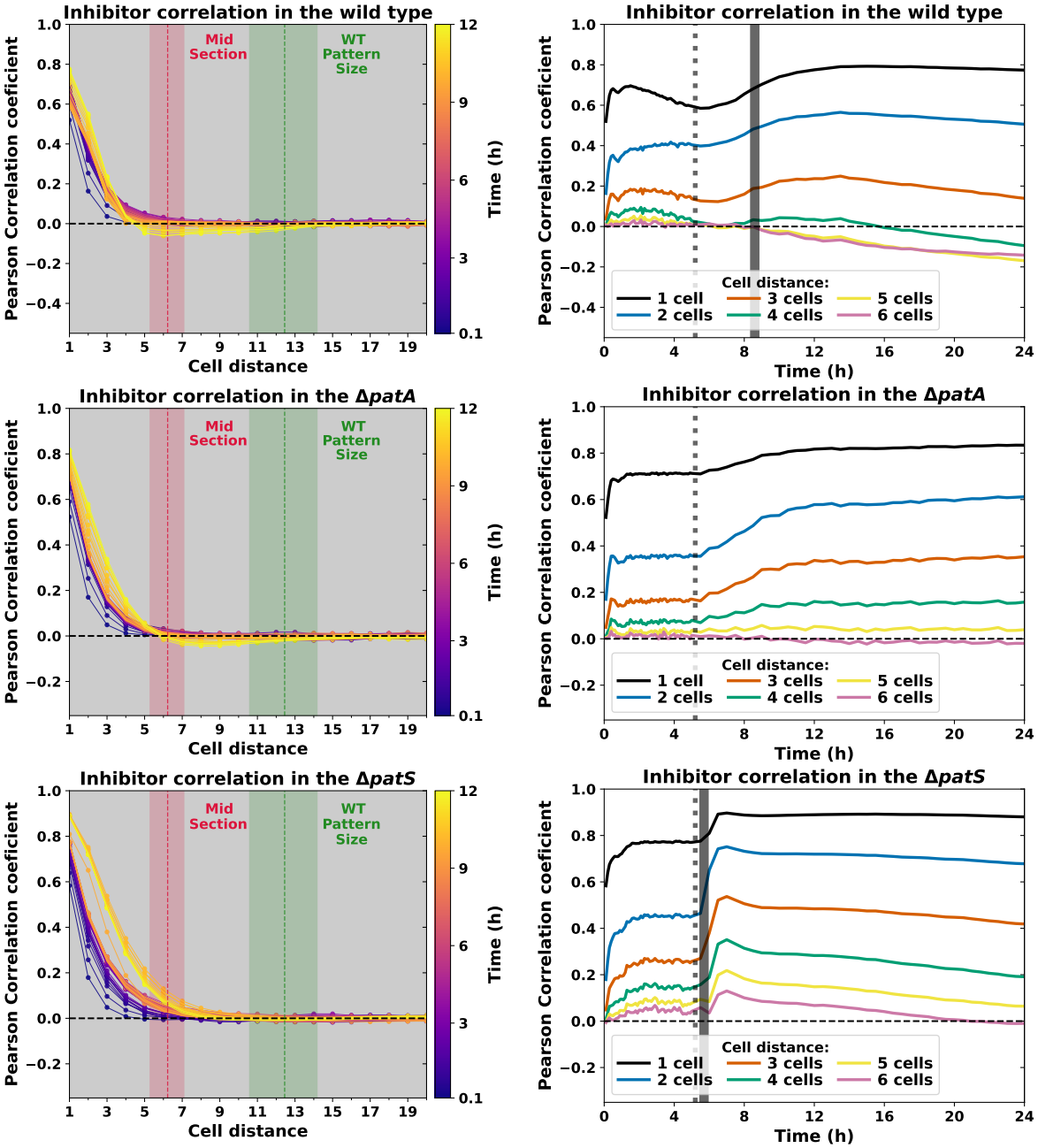

Figure 5.1: **Correlation of ERGSGR inhibitor concentration for several genetic backgrounds:** The graphs on the left show the evolution of the relationship between the correlation between two cells and their relative distance in the filament. We colored in green and red all the cell distances and midpoints compressed between the lowest and the maximum average vegetative intervals between heterocysts in the wild type for 10h to 96h, with a doted line for the global average. The graphs on the right show the evolution of the correlation between cells located at six particular cell distances in time. To ease visualization, we marked with a gray doted line the time when the first heterocyst appears in our simulation and a gray band when the pattern arises.

In fig. 5.1 one can see that in fact the correlation of the diffusible inhibitor is really high for close neighbors and decays up to the distance equivalent to half the pattern length. Additionally, this correlation does not seem to be significantly affected by the Δ*patA* and Δ*patS* mutations that greatly alter the wild type pattern at short times. This leads us to argue that this correlation length of around 6 cells is produced by the combination of the diffusion coefficient and the degradation rate. Specially the degradation, as we saw that it affects the mean interval much more than variations in the diffusion rate (fig. 4.16). Moreover, we think that this correlation length defines the size of the pattern that can arise in the system. In the end, the diffusible ERGSGR is the only element that diffuses between cells during the pattern inception phase. Then the correlation length marks a perception limit in the filament. Cells do not "see"/sense other cells located further than 6-7 positions.

We can also see that once heterocyst start to appear in the filament, the correlations are distorted in a manner proportional to the number of heterocyst. This is specially visible in the first differentiation round that goes from the apparition of the first heterocyst (the gray doted line) up to the grey band that marks the rise of statistically well-defined vegetative cell intervals. The sudden appearance of functional heterocysts floods the filament with inhibitors, homogenizing its concentration, this enhances all the correlations already present in the filament. This sudden increase could be an artifact caused by the absence of a protoheterocyst stage. The addition of a progressive activation of HetN would stagger the production of inhibition and maybe reduce considerably this effect.

5.2 TWO STAGED COMPETITION WITH 5 CELLS GROUPING REGULATE HETEROCYST SELECTION

During this chapter, and specially this section, we will extensively reference (Corrales-Guerrero et al., 2015), as it is the only experimental work concerning a spacial characterization of HetR expression. This work characterizes the correlations of hetR-gfp expression along filaments in conditions of both nitrogen repletion and deprivation. They observed that the correlation along the filament are stationary in nitrogen replete conditions. In those conditions, the authors report a low positive correlation between neighboring cells that recedes with cell distance, with an expected spacial range of 2-3 cells. These correlations are bigger for the Δ*patS* mutant background (around a Pearson coefficient of 0.30) than for the wild type (around a Pearson coefficient of 0.10).

Given that our model only considers nitrogen deprivation conditions, we cannot validate our model in nitrogen replete conditions. Additionally,

as we exposed in section 4.3, our model does not consider HetR and PatA dynamics in heterocysts. As a result of this, when a cell differentiates in our model, we freeze the HetR and PatA concentration value. Then this value will remain constant for the rest of the simulation. This causes a problem in the level of HetR activity (eq. (4.7)), it is not properly defined in heterocysts given that this expression uses the concentration of both HetR and PatA. To avoid this issue, we decided to artificially fix the activity to the maximum of 1 in the heterocysts. This is in agreement with the experimental observations, because, as we mentioned before, HetR activity in the heterocyst is supposed to be high (Herrero and Flores, 2019). We chose 1 because it is the highest possible value and therefore most different to the maximum observed value in vegetative cells, which in the wild type is around 0.6. Setting the same value for all the heterocysts introduces an artificial higher correlation between heterocysts, but this correlation is reduced by the fairly disperse vegetative interval distribution fig. 4.9 and the lower amount of heterocysts fig. 4.8. Additionally, as the value is much higher to the one observed in vegetative, one could use the HetN correlation as a guide.

The first realization presented in (Corrales-Guerrero et al., 2015) is that the intercellular correlation in the filament is no longer stationary. They reported that up to three hours after nitrogen deprivation, the distribution of the fluorescence along the filament still resemble the ones from nitrogen replete conditions. But from seven hours after nitrogen step-down onwards, the filament shows a marked tendency for an intercalation between high and low fluorescence values in adjacent cells. As a result of this, they suggest that the fluctuations that originate the pattern build from the existent intrinsic correlations in nitrogen replete conditions. The authors use this observation, together with their observation in nitrogen replete conditions, to endorse the two stage development model presented by (John C. Meeks and Elhai, 2002). This model hypothesizes that on the early stage, only certain cells are susceptible to initiate the differentiation process once the filament is under nitrogen deprivation. And then the second stage where the cells of this primered clusters compete to be the one that differentiates.

We have quite a different initial condition in our model, we initiate the filament with a homogeneous initial condition (all the cells deactivated and empty). But this should not affect the pattern inception mechanism. We have already shown in chapter 3 that the initial heterogeneity is irrelevant for the formation of the pattern if one considers a noisy gene expression. This work alone already suggests that existent heterogeneity of the filament, while could affect which of the cells are more susceptible to differentiate, cannot be essential to originate a pattern. Additionally,

our model considers instantaneous differentiation to a heterocyst. This absence of a protoheterocyst stage will probably alter the timescale of our model with respect to the real system, as described in (Corrales-Guerrero et al., 2015). Despite these differences, the model presents a remarkably similar overall behavior of the cell correlations in a wild type filament.

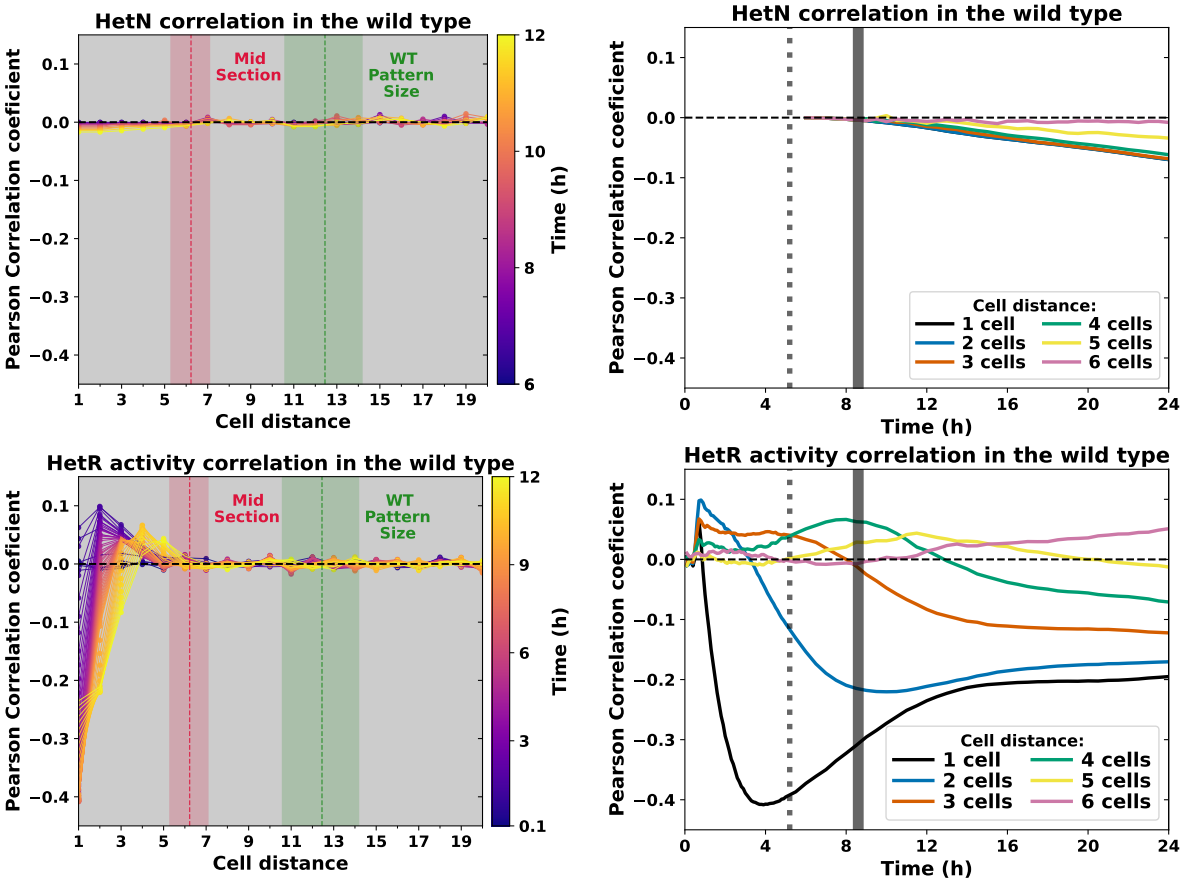

Figure 5.2: **Correlation of the HetN concentration and the HetR activity for the wild type:** The graphs on the left show the evolution of the relationship between the correlation between two cells and their relative distance in the filament. We colored in green and red all the cell distances and midpoints compressed between the lowest and the maximum average vegetative intervals between heterocysts in the wild type for 10h to 96h, with a doted line for the global average.
The graphs on the right show the evolution of the correlation between cells located at six particular cell distances in time. To ease visualization, we marked with a gray doted line the time when the first heterocyst appears in our simulation and a gray band when the pattern arises

First, we can see in fig. 5.2 top that in the interval from the apparition of the first heterocyst to the pattern rising, the effect of the heterocysts over the filament correlation is completely negligible. We will see later that when the amount of heterocyst increases, this is no longer true. Due to this we will limit ourselves to the discussion of the pat-

tern inception. In this regard, one can observe in fig. 5.2 bottom that the model correlation between neighboring cells also increase until a value comparable to the one observed experimentally, around 0.1, before starting to anticorrelate (as described experimentally in (Corrales-Guerrero et al., 2015)). This indicates that our system also presents an initial build up of the HetR concentration until the lateral inhibition of PatS initiates. This inhibition progressively increases the difference of the HetR activity between consecutive cells until reaching a -0.4 Pearson correlation coefficient. This rapid increase of the first neighbor anticorrelation shows that, at least in our model, the lateral inhibition acts really fast over neighboring cells. This creates a noisy alternation between on and off cells in the filament (represented on the second line of fig. 5.3).

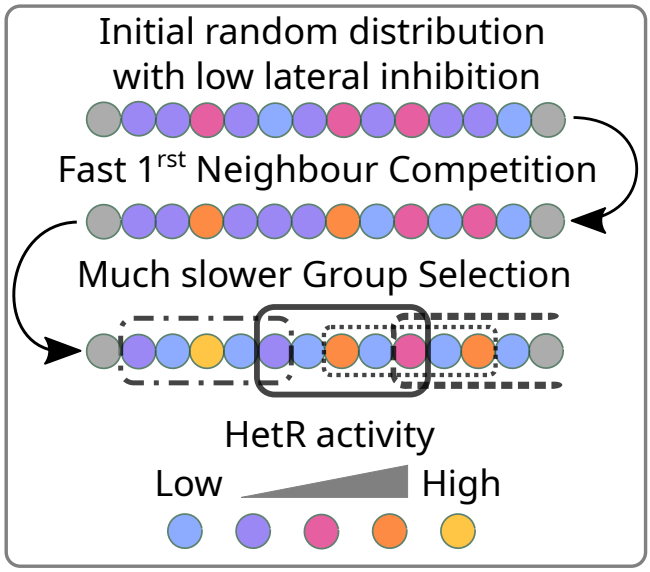

Figure 5.3: **Diagram of the cell selection for heterocyst differentiation:** The scheme shows the same filament fragment at three different times after nitrogen deprivation, with the cells colored with their respective HetR activity (gray cells represent the rest of the filament). The first line represents a filament right before the start of the contiguous anticorrelation, around 1h after nitrogen deprivation (aND). The second line correspond to around 2h aND where the contiguous anticorrelation is high, but there is still some 2-cell correlation. And finally the third row presents a filament at around 4-5h aND when the contiguous anticorrelation reaches its maximum and the 2-cell intervals start to anticorrelate. The boxes represent the expected "safe spaces" that an existent candidate needs to stabilize and differentiate.

But in our model, in agreement with the observations in (Corrales-Guerrero et al., 2015), the 2 cell distance correlation is clearly the most relevant to the formation of the pattern. We have already seen that the inhibitor correlation between two cells roughly halves with each extra cell between them. Then the interaction between contiguous and 2 cell range cells will dominate over the rest, especially in such a noisy system. Additionally, it is easy to see in fig. 5.2 bottom that the evolution of the correlation of cells at a 2 cell distance reaches its minimum right before the completion of the pattern.

To present the first stage of our proposed method of selection is represented in fig. 5.3, specially on the right side of the filament fragment. But first, on the left side of the fragment you can see a yellow cell which would be an example of a candidate for early differentiation. That cell had the "good luck" of being the only one with signi-

ficative HetR expression in its neighborhood (the 5 cell interval centered in itself, represented as a dashed doted box). Due to this, if this situation is not perturbed due to stochastic fluctuations, this cell would differentiate. On the other side, we will now focus the group of 5 cells surrounded by a dotted box. If one observes the "initial" configuration on the first row, only the leftmost orange and the central pink cells of this group presented significantly higher activity. But as they were located only 2 cells apart, they inhibited each other. This allowed the rightmost cell of the group to increase its activity and start inhibiting the central cell. This can be imagined as if the most secure or lucky of a group of two cells with higher expression with a lower one between them pushes the increased expression of the other 2 cells apart. When and if this cell recedes, this fragment would reach an organized semi stable state that maximizes 4 cell distance correlation and marks the rise of the pattern fig. 5.2 bottom. To ease the visualization of this process, we have considered a much more static situation with lower noise. In the proper modeled filament, this selection strategy would be complemented with stochastic fluctuations that would give a more fluid character to the spacial organization.

But if we are saying that the cells "compete" in groups of 5 cells, and we do not observe any correlation for intervals larger than 5 cells until the pattern has been established, how can the mean pattern length it be 10-12 cells? This due to the second stage alluded in the section title. In our model, once a cell differentiates its inhibitory range increases, given that it produces both HetN and fixed nitrogen and the nitrogen diffuses faster along the filament (fig. 5.4). This allows the recently formed heterocyst to completely inhibit the differentiation of neighboring groups or tho push them, analogously as exposed earlier for the intragroup competition. This can be observed in the wild type movie as it happens regularly that a light up cell is rapidly shut off by a new neighboring heterocyst. Additionally, this second stage would explain the increase of the correlation in activity at 5 cells distance that starts right when the first heterocyst appears in the system. Then, while the cells compete inside 5

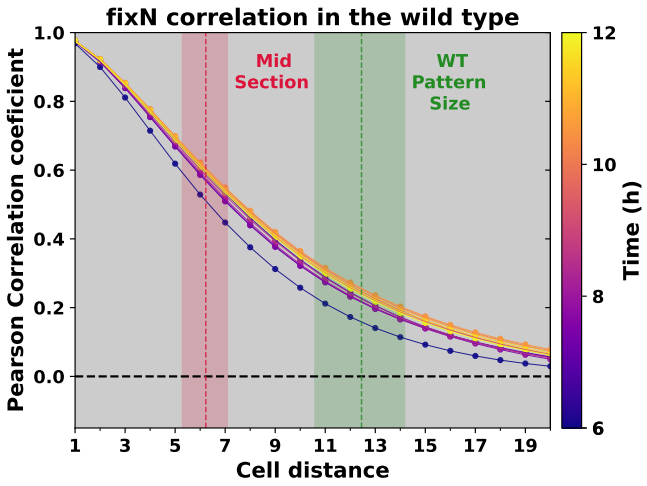

Figure 5.4: **Correlation of the fixed nitrogen concentration for the wild type:** The graphs on the left show the evolution of the relationship between the correlation between two cells and their relative distance in the filament. We colored in green and red all the cell distances and midpoints compressed between the lowest and the maximum average vegetative intervals between heterocysts in the wild type for 10h to 96h, with a doted line for the global average.

cells groups for the first stage, for the second stage, the groups compete as a whole to differentiate first.

This second stage, as exposed here, is a bit artificial because it depends greatly on the sudden HetN production right after differentiation. But we think that with the incorporation of a proper nitrogen sensing in the model, this kind of strategy could still work with a more gradual HetN and nitrogen activation. The substitution of the HetR basal production with a regulated one dependent on the nitrogen concentration in a cell would increase the responsivity to nitrogen of the cells. Additionally, the cellular growth should be coupled to nitrogen consumption given that it is an essential nutrient for the cells. These two effects combined could allow for a more realistic second stage of pattern inception. Due to this, we are currently working on a more realistic depiction of both the heterocyst differentiation and the nitrogen regulation that would be necessary to verify and further this study.

5.3 $\Delta patS$ AND $\Delta patA$ MUTATIONS GREATLY INCREASE INTERCELLULAR CORRELATION

We repeated the previous study for both the $\Delta patS$ and $\Delta patA$ and observed, as expected, that in both cases the selection mechanism is similarly distorted, albeit for completely different reasons.

First we will consider the $\Delta patS$ mutant, the vegetative cells of this mutant are not capable to produce enough lateral inhibitor to avoid the simultaneous differentiation of multiple contiguous cells. The much lower expression of redundant genes such as *patX* limits the extension of the clusters of cells with a high expression, avoiding a mortal complete differentiation of the filament (Elhai and I. Khudyakov, 2018; I. Khudyakov et al., 2020). In this mutant, the differentiation is merely a race between almost all the cells of the filament, given that it is really homogeneous. The average activity in the filament 5h aND is 0.83 ± 0.02. Additionally, as can be observed in fig. 5.5 top, the filament only present positive correlations that decay with the cell distance up to the apparition of the first heterocyst. And from that point the high presence of heterocysts completely dominates the correlation of activity, which is exactly equal than the HetN one. This decaying profile (fig. 5.5 top left) is really similar to the one observed for the inhibitor in fig. 5.1 bottom left. But in this case, we are not sure of what is causing it, given that the only interaction between cells is both inhibitory and heavily reduced in this mutant. We expected to observe an uncorrelated filament without much spacial organization. But instead, we found activation clusters of roughly the pattern size. To ensure that we were not seeing an artifact, we checked the more proba-

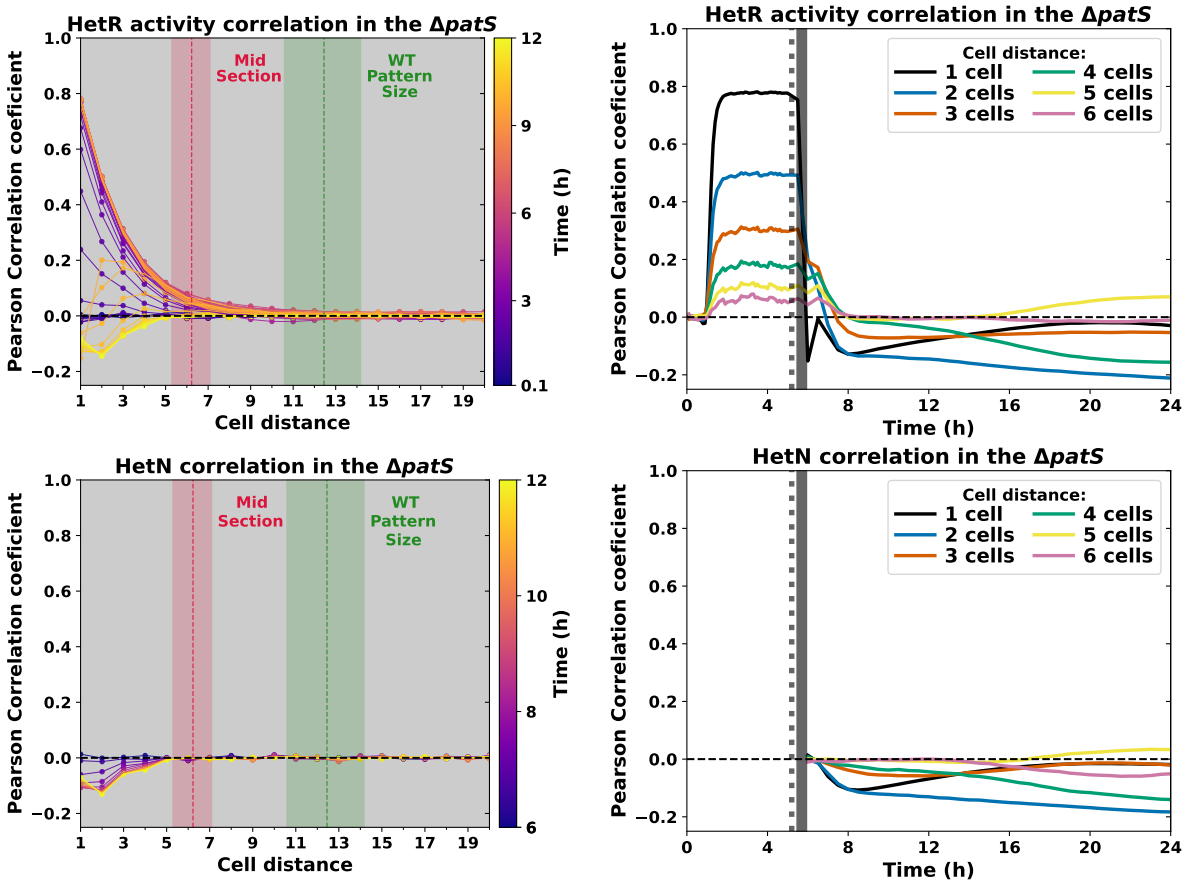

Figure 5.5: **Correlation of HetR activity and HetN concentration for the Δ*patS* mutant:** The graphs on the left show the evolution of the relationship between the correlation between two cells and their relative distance in the filament. We colored in green and red all the cell distances and midpoints compressed between the lowest and the maximum average vegetative intervals between heterocysts in the wild type for 10h to 96h, with a doted line for the global average.
The graphs on the right show the evolution of the correlation between cells located at six particular cell distances in time. To ease visualization, we marked with a gray doted line the time when the first heterocyst appears in our simulation and a gray band when the pattern arises

ble causes of this. First, we checked if it was the cell division, given that the daughter cells inherit the cellular concentration of its progenitor. And then also the border conditions, given that those cells are the ones with the highest activity (higher than 0.9 at 5h aND). But there is no difference in the correlation if we completely stop the division, or we take out 50 cells from the filament ends. Indicating that neither of those are causing for this positive correlation. With our current data, we do not have a proper explanation for this behavior. We are now running simulations with different range of Δ*patS* partial mutation severity to characterize the transition between the wild type and the Δ*patS* behavior. It could be an

artifact due to the enormous homogeneity in the filament. But we want to check if the reduction of the inhibitor production rate could somehow modulate the resolution range of the lateral inhibition. We have seen that, at full strength in the wild type setting, the lateral inhibition is capable to completely suppress neighboring cells. And for really low inhibition we have a larger correlation length which defines activation clusters. Then maybe this behavior is gradual, and once a cell is not capable to completely inhibit their immediate neighbors they constitute a cluster that is capable of inhibiting the neighboring cluster. If this happens, we should observe a gradual increase of the correlation length with the mutant severity.

A similar behavior is also observed for the HetR activity correlation in the Δ*patS* mutant presented in fig. 5.6.

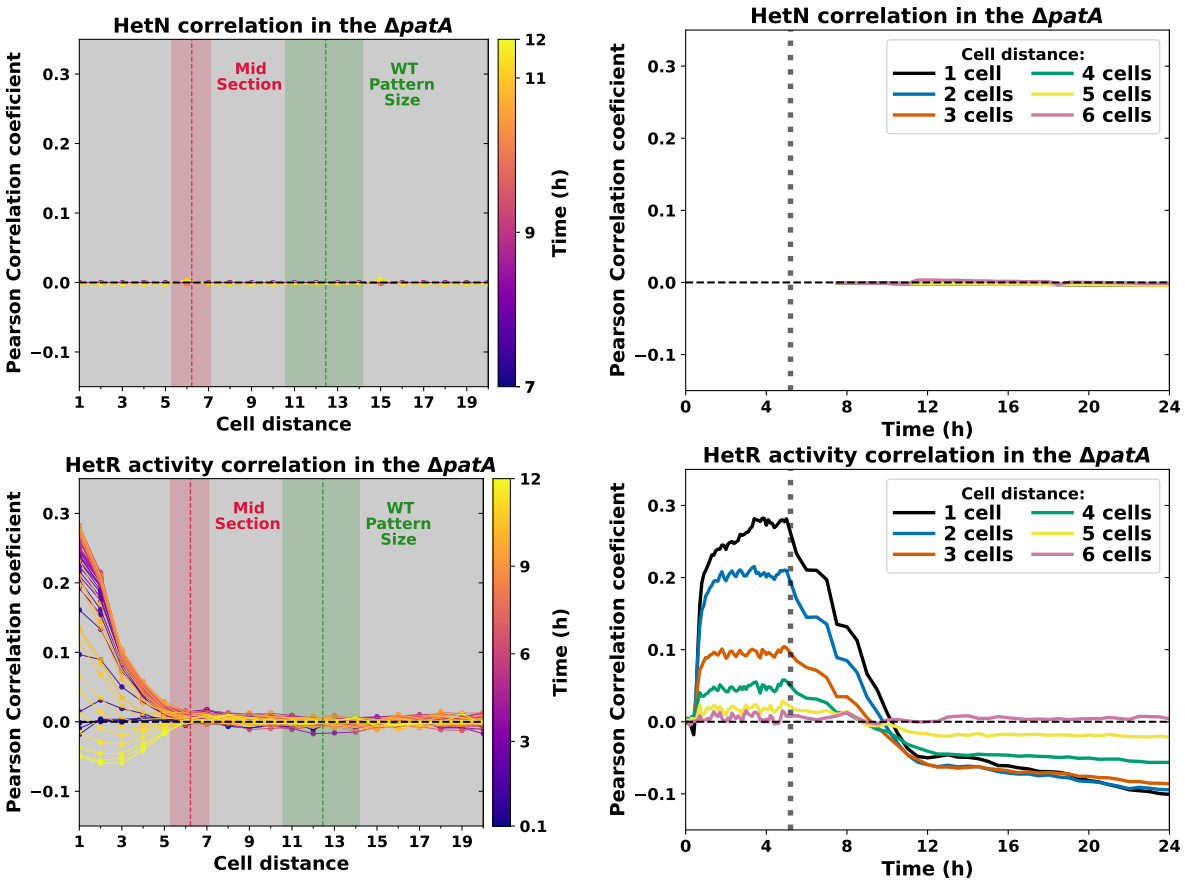

Figure 5.6: **Correlation of the HetN concentration and the HetR activity for the Δ*patA*:** The graphs on the left show the evolution of the relationship between the correlation between two cells and their relative distance in the filament. We colored in green and red all the cell distances and midpoints compressed between the lowest and the maximum average vegetative intervals between heterocysts in the wild type for 10h to 96h, with a doted line for the global average.
The graphs on the right show the evolution of the correlation between cells located at six particular cell distances in time. To ease visualization, we marked with a gray doted line the time when the first heterocyst appears in our simulation and a gray band when the pattern arises.

We can see that for this mutant there is no pattern formation and, therefore, the filament do not present any correlation on the HetN production. We also have a really homogeneous filament, but at a much lower activity level. At 5h aND we have an average activity of 0.10 ± 0.02, with the border cells also being the ones with the highest activity around 0.4. This higher difference reduces the correlation value observed before the apparition of the first heterocyst. The low average activity is also the reason why the few heterocysts that appear are capable to completely eliminate the correlations in the filament. As we have set the activity in those to 1 the filament variance increases enormously. This is remedied if we consider the correlations in the HetR concentration instead, as shown in fig. 5.7.

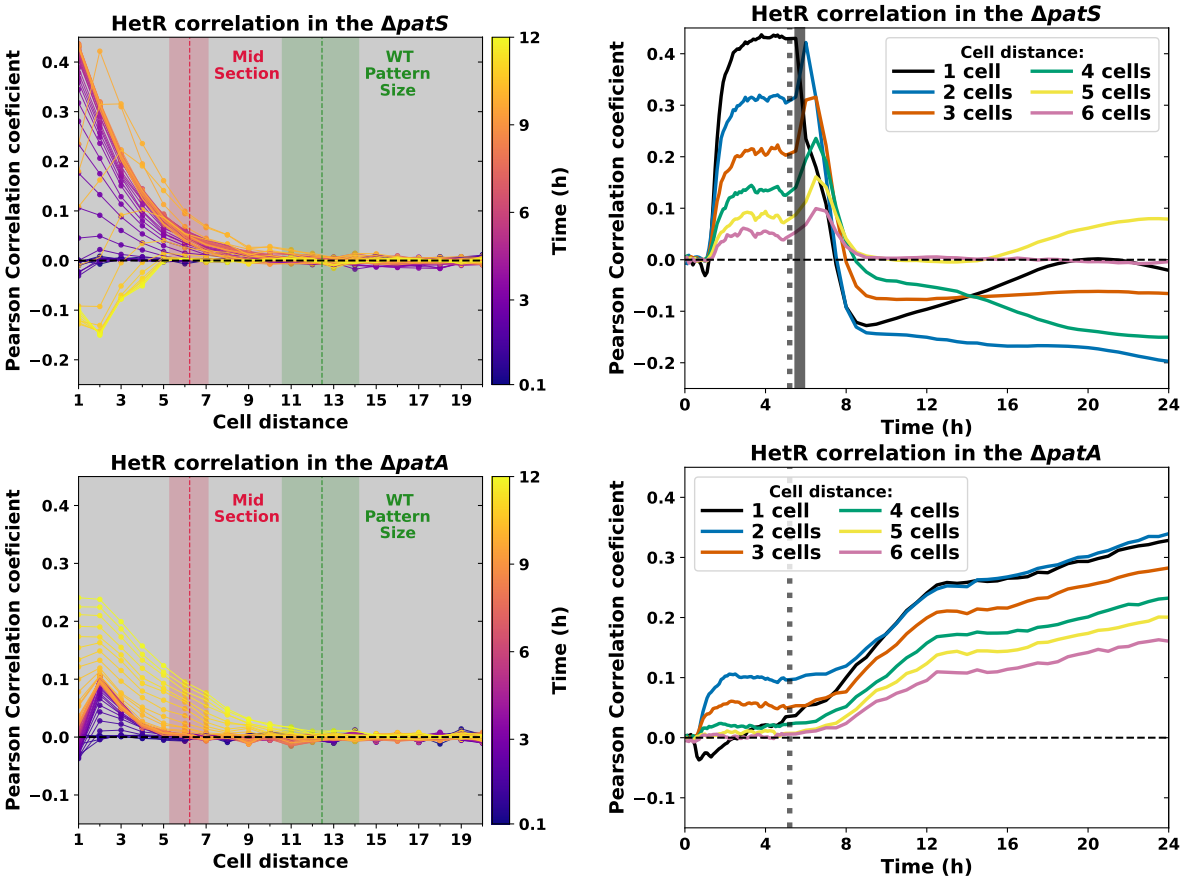

Figure 5.7: **Correlation of HetR concentration for the $\Delta patS$ and $\Delta patA$ mutants:**
The graphs on the left show the evolution of the relationship between the correlation between two cells and their relative distance in the filament. We colored in green and red all the cell distances and midpoints compressed between the lowest and the maximum average vegetative intervals between heterocysts in the wild type for 10h to 96h, with a doted line for the global average.
The graphs on the right show the evolution of the correlation between cells located at six particular cell distances in time. To ease visualization, we marked with a gray doted line the time when the first heterocyst appears in our simulation and a gray band when the pattern arises

It is easy to see that the freezing of the HetR concentration in heterocysys does not affect the correlation over HetR that much. The reason for this is that we use an integrative threshold to mark a cell for differentiation, and this is already a fairly noisy system. Due to this, the frozen value would be different for each cell but always higher than the threshold value, which is what one would expect on young heterocysts. It is well known that the expression of HetR is high in heterocyst (Herrero and Flores, 2019).

These two mutants represent the two sides of the same coin. If the lateral inhibition is distorted, the self organized heterogeneity does not appear and the filament remains in a homogeneous regime with all the cells with a similar level of HetR. This is exactly equivalent to the observations done in chapter 3 for the minimal 3 genes model. As these two mutations can be understood as a reduction (the Δ*patA* one) or an increase (the Δ*patS* one) of the HetR production. And as such the filament states before the first heterocyst appearance are analogous to the ones presented in fig. 3.14. The only difference is that for the high concentration homogeneous state, the value is over the differentiation threshold. Then, if some cells differentiate before due to stochastic fluctuations, the functional HetN and fixed nitrogen are capable to enforce the spacial organization and rescue the pattern.

5.4 CORRELATION OF ERGSGR INHIBITOR THE DURING THE PATTERNING REGIME

This is specially evident if one observe the correlations of inhibitor concentrations for longer times after nitrogen deprivation. In the wild type (fig. 5.8 top) once the pattern appears, there is a transition regime in which the regularity of the heterocysts morph the correlations along the filament. Once the correlations reach an oscillatory equilibrium, we see the same periodicity observed in fig. 4.10. As we discussed in the previous chapter, his periodicity is defined by the coupling between the differentiation rounds and the cell division. In this genetic background, the inhibitor correlations in the patterned regime clearly show the pattern with a high (for the system standards) anticorrelation length equivalent to half the pattern length. This constitutes a differentiation valley between two existing heterocysts where a new heterocyst will appear.

If we now focus on the Δ*patA* (fig. 5.8 bottom) we see that the overall profile of inhibitor correlations does not change because, as the pattern does not arise, the random heterocyst that appear do not have a significant effect over the spatial distribution of inhibitor along the filament. The

concentration of inhibitor is still largely homogeneous in the filament, and there is no significant long range correlation between cells of the filament.

Finally, the Δ*patS* (fig. 5.8 middle) shows an intermediate correlation state between those two.

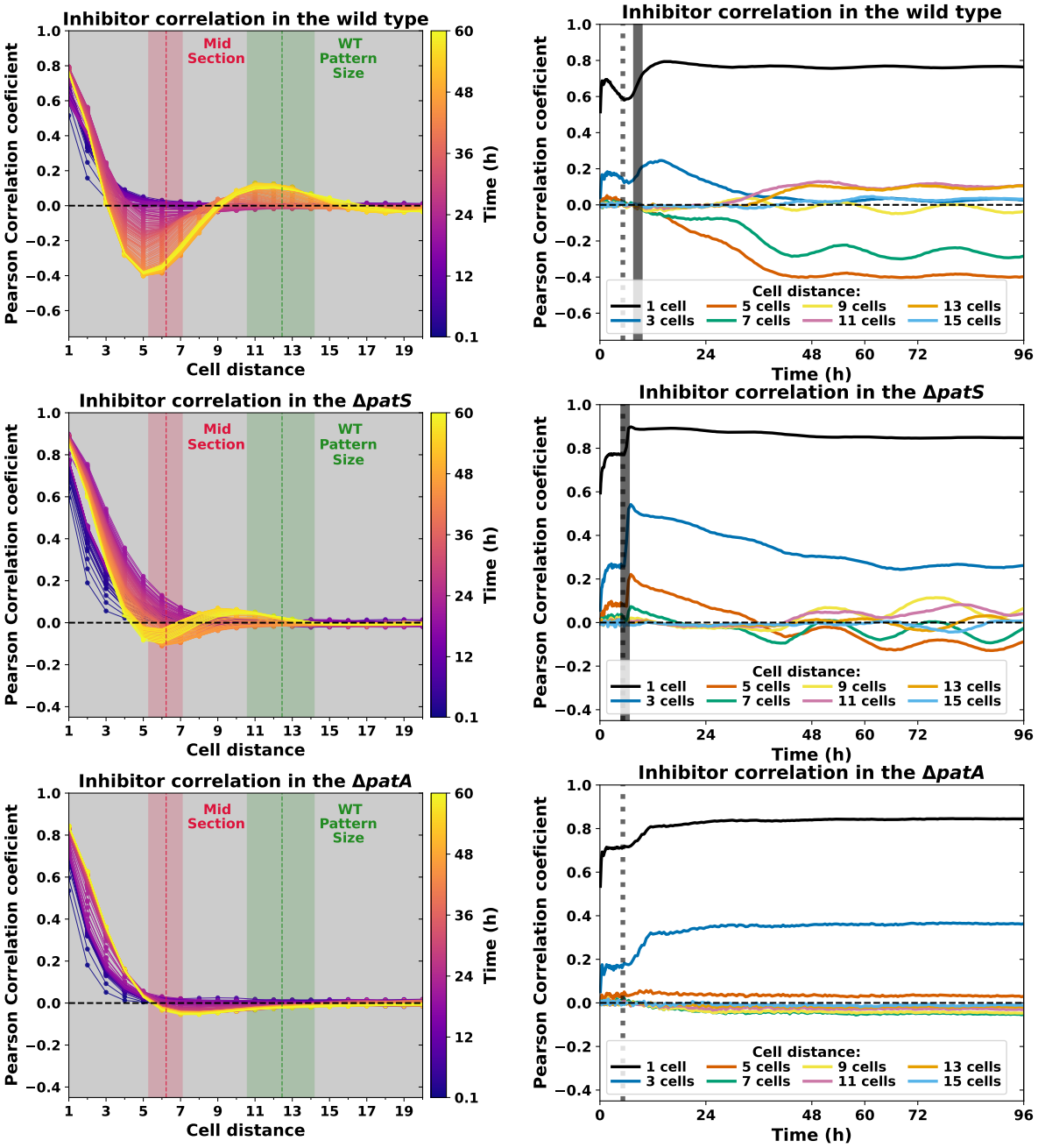

Figure 5.8: **Correlation of ERGSGR inhibitor concentration for the wild type and the Δ*patS* and Δ*patA* mutants:** The graphs on the left show the evolution of the relationship between the correlation between two cells and their relative distance in the filament. We colored in green and red all the cell distances and midpoints compressed between the lowest and the maximum average vegetative intervals between heterocysts in the wild type for 10h to 96h, with a doted line for the global average. The graphs on the right show the evolution of the correlation between cells located at six particular cell distances in time. To ease visualization, we marked with a gray doted line the time when the first heterocyst appears in our simulation and a gray band when the pattern arises.

The overall correlation evolution in this mutant closely mimics the one of the wild type. After the pattern rise, there is a transition regime until the correlations reach an oscillatory equilibrium with the same wild type periodicity. But in this case, the amplitude of the correlation oscillations is much lower than the wild type's with a correlation length profile much more similar to the Δ*patA* one.

5.5 CONCLUSIONS AND FURTHER WORK

In this chapter, we use intercellular correlation data to propose an alternative two staged self organization process to select the heterocyst localization in the filament. Unlike in the previously proposed one (John C. Meeks and Elhai, 2002), this mechanism does not require the existence of spacial heterogeneity in the filament to explain the apparent discrepancy between pattern length and intercellular correlation characteristic length. In (Risser et al., 2012), the authors suggest that a biased inheritance of *PatN*, could favor the differentiation of those cells with less concentration of this protein. This is due to the inhibitory effect that this protein has over the cellular expression of patA. Despite this, the fact that some cells are more susceptible to differentiate would not produce by itself a longer spatial organization, but it will produce an intrinsic heterogeneity. Then, some cells will be capable to differentiate much faster than others, staggering the differentiation round, and therefore easing the intergroup competition.

We are aware of the model limitations, specially with the instantaneous maturation in our heterocysts given the pivotal role that we assign to the nitrogen signaling in the selection mechanism. This stems for the simplification of not considering a transitional stage, the protoheterocysts after the cell commitment. This seems to contradict the extensive cellular reorganization necessary to transition from an internal oxygen production to an anaerobic condition to allow nitrogen fixation. But recent gene expression studies (Di Patti et al., 2018) show that while *hetN* expression starts after commitment, it rises rapidly and coexists with high *patS* expression, which decays much slowly after commitment. Additionally, they report that this occurs while the autofluorescence from photosynthetic pigments is still receding, with the HetN activation occurring less than 5 hours after half of the autofluorescence. This clearly indicates, while we are clearly considering shorter maturation times for a heterocyst to be able to produce nitrogen, we are probably considering much longer times for the HetN expression. And this two difference could completely compensate each other.

We are currently working on a more realistic heterocyst differentiation with a clear timescale separation between HetN and fixed nitrogen production. This would require to first incorporate a protoheterocyst stage where the *patS* expression decays while the *hetN* one rises. But also to renounce to the basal HetR production simplification incorporating a dual regulation of *hetR* and *patA* (Bastet et al., 2010; A. M. Muro-Pastor et al., 2002). The existent one without the phenomenological fixed nitrogen inhibition and the *ntcA* one that would only respond to the nitrogen concentration in the cell. Then the Δ*patS* mutant will be less intense because its effect over *hetR* and *patA* expression on nitrogen deprived cells would reduce.

And finally, more experimental studies with gene expression reporters are required. The mechanism proposed here agrees with the experimental observations presented in (Corrales-Guerrero et al., 2015) but the experimental work studying the spatial self organization in the filament is still really rare. A longitudinal study considering three times corresponding to the three filament regimes presented here could easily test our hypothesis. One, a few hours after nitrogen deprivation but before the first heterocyst appearance, a second one right after that appearance before the full round of differentiation and a last one as late as experimentally possible. This could easily validate if there is in fact a transformation of the correlation-cell distance profile. Even a more detailed phenotypical description of the heterocyst differentiation rounds could provide additional insight about the expected cell correlation length. Given that our mechanism requires at least two differentiation waves in each round, the differentiation should not be sudden along all the filament and appear more as a cascading effect.

Part III

NITROGEN INCORPORATION IN PLANTS

Here we will study a particular example of how nitrogen availability regulates plant growth. We analyze the regulatory network that controls the plant allocation of resources on both vertical and lateral growth. The knowledge of how this allocation is decided allows for targeted modifications that would increase the yield of the plant.

6

TILLERING REGULATION IN RICE PLANTS

The green revolution rice presented a mutation in the allele Semi-dwarf1 (*sd1*) that reduces gibberellin (GA) production/maturation (Monna et al., 2002; Sasaki et al., 2002; Spielmeyer et al., 2002). This decrease of the GA abundance allows for an increase in the activity of the SLENDER RICE1 (SLR1) of the growth-repressing DELLA protein family (Itoh et al., 2002). This increase in SLR1 concentration reduces the overall height of the plant, while additionally maintaining its sensibility to nitrogen fertilization. This reduction of size increases the harvest index (grain–straw ratio) to values near 50% (from an original 30%-79%)(Khush, 1999). Additionally, it allows for higher fertilization because rates of fertilization that produced exceedingly tall inefficient crops that were easily lodged (flattened by the wind) now could even double the gross yield of the crop without any height modification (Khush, 1999). This is possible because GA seems to control the number of stem internodes that elongate upon bolting (which defines the number of branches) independently of its effect on internode elongation (that will result in the plant height)(Rieu et al., 2008). This existence of parallel regulatory mechanisms (Hedden, 2003) allowed for an optimization of the crops such that the additional nutrients are directed mainly to the increase of the seeds that a plant produces without affecting much of the biomass production.

The widespread augment of fertilization to sustain intensive agriculture caused progressive soil degradation. In recent years, the main focus of the genetic screening of the green revolution varieties has shifted from the gross increase of yield with nitrogen fertilization of dwarf varieties (Khush, 1999) to the search for more specialized sustainable varieties that reduce soil degradation due to intensive agriculture (Pingali, 2012). This ideal variety would have optimal productivity with a perfect configuration of short panicles (main branches) with many secondary branches that contain more and bigger grains. In order to be capable of engineering such an ideal crop, it is necessary to understand the regulatory pathways that intervene in each biological feature and how interconnected those pathways are. Several genes that control particular traits associated with rice productivity such as grain number, grain size, panicle size and tillering have been identified (Y. Xing and Q. Zhang, 2010). Both GN1A (Ashikari et al., 2005) and APO1 (Ikeda-Kawakatsu et al., 2009) regulate grain number, GS3 (Fan et al., 2006), GW2(X.-J. Song et al., 2007),

qSW5(Shomura et al., 2008) and OsSPL16 (Shaokui Wang et al., 2012) regulate grain size and DEP1 (Xianzhong Huang et al., 2009) controls panicle size. But we will focus on the genetic regulatory network that controls the lateral branching in rice (tillering).

It is already well known that the SLR1 also affects tiller formation through the protection of MOC1 to degradation (Komatsu et al., 2003; X. Li et al., 2003; Liao et al., 2019; Oikawa and Kyozuka, 2009). But a recent paper, (K. Wu et al., 2020), presented NGR5 that reduces the transcription of several genes known to inhibit lateral branching and tiller number (D14 (Arite et al., 2009; Yao et al., 2016), D3 (Ishikawa et al., 2005; Jiang et al., 2013), OsTB1 (Takeda et al., 2003) and OsSPL14 (Jiao et al., 2010; Miura et al., 2010)). Specifically, the authors of (K. Wu et al., 2020) found that both D14 and OsSPL14 function downstream of NGR5, and NGR5 mediates a nitrogen-promoted increase in tiller number by repressing the inhibitory functions of D14 and OsSPL14 (and likely of other) branching-regulatory genes.

This duality in the regulation of the tillering allows the system to allow to both links the lateral growth with the reduction of the overall height (through the SLR1 pathway) but not compromise fully to this coupling by having an alternative pathway (through NGR5) that is not directly related to the height regulation.

Here, we translate this new information into a model that relates the branching regulation to the well-characterized regulation of the DELLA proteins to study how these two pathways interplay.

6.1 TILLER REGULATION MODEL

We developed our model through the incorporation of the GA-DELLA interaction presented in (Murase et al., 2008) and later modeled in (Middleton et al., 2012) into the tillering regulation described earlier. This allows us to reduce the number of parameters to adjust in our model by taking the equivalent ones from (Middleton et al., 2012).

For simplicity, we chose to model the tillering as a continuous protein species with a global multiplicative factor called T_{factor} to allow higher flexibility on the parameters' adjustment. This choice may seem too reductive, but any more realistic alternative would have to consider both the full branching regulatory system and the sequential and spatially localized character of a branch initiation. This hugely detailed description of the system is completely out of the scope of this work, given that our objective is to present a deterministic mean-field model at the plant level. To be capable to reproduce the changes in the number of formed branches

for different mutants and experimental conditions, this simple framing is enough.

We consider both a direct activation of the tillering of the DELLA proteins (referred to as D during this paper), through MOC1 (which is not included in the model) and an indirect one through NGR5 (referred to as A during this paper). This indirect activation of NRG5 is attained by reducing the production of tiller inhibitors. And both these proteins are regulated by the GID1(G1)-Gibberellin(GA) complex, which is capable of both sequestering and degrading those proteins.

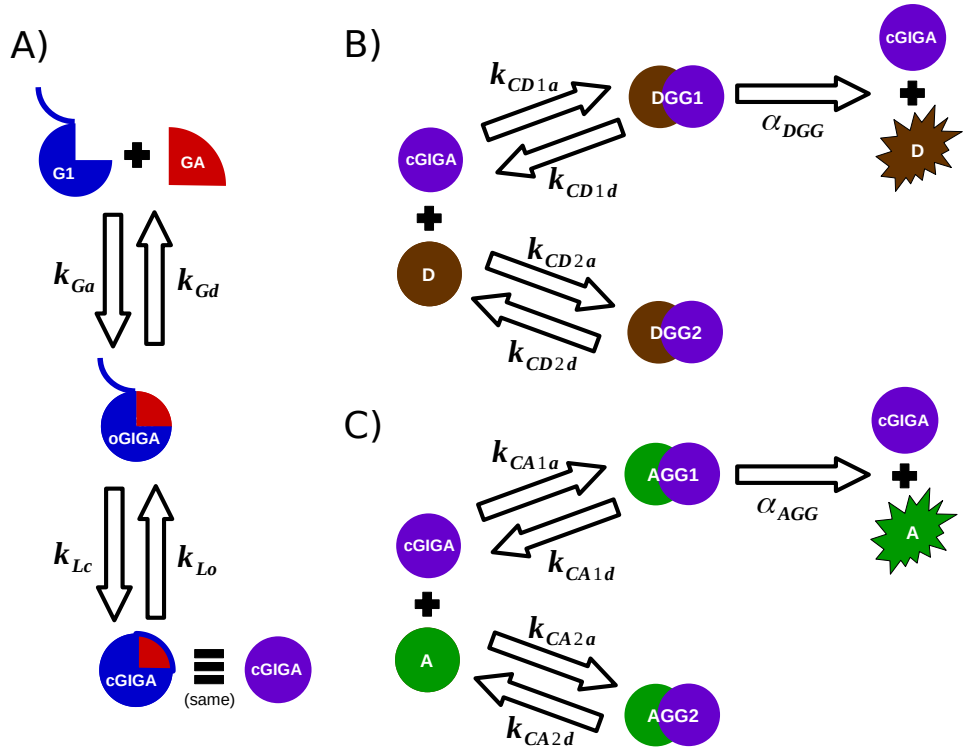

Figure 6.1: **GA-GID1 interactions:** A) Two-step process to form the functional GID1-GA complex that involves a GA adsorption and the closing of a "lid". B-C) GID1-GA complex interactions with the DELLA proteins and NGR5 (named A during this work) respectively.

This complex works in three steps. First, the GA attaches to the GID1, forming the complex oGIGA (represented graphically in fig. 6.1 A). This produces a conformation change that closes the lid enclosing the GA, as described in (Murase et al., 2008), converting it to cGIGA. Finally, this GID1-GA complex is now active and capable to link with both NGR5(A) and DELLA(D). This link could be strong, obtaining the complexes AGG1 and DGG1, and leading to the degradation of A and D respectively. Or alternatively, it could be weak with a higher dissociation constant, which

results in temporal sequestration of both A and D in the complexes AGG2 and DGG2. These interactions are depicted graphically in fig. 6.1 B-C. As one can easily see, we have chosen to consider the same interactions with the GID1-GA complex for NGR5 as the ones described in (Middleton et al., 2012). This is justified due to the evidence of both GID1-GA-NGR5 interaction and a competitive relationship between SLR1 DELLAs and NGR5 with respect to GID1 reported in (K. Wu et al., 2020).

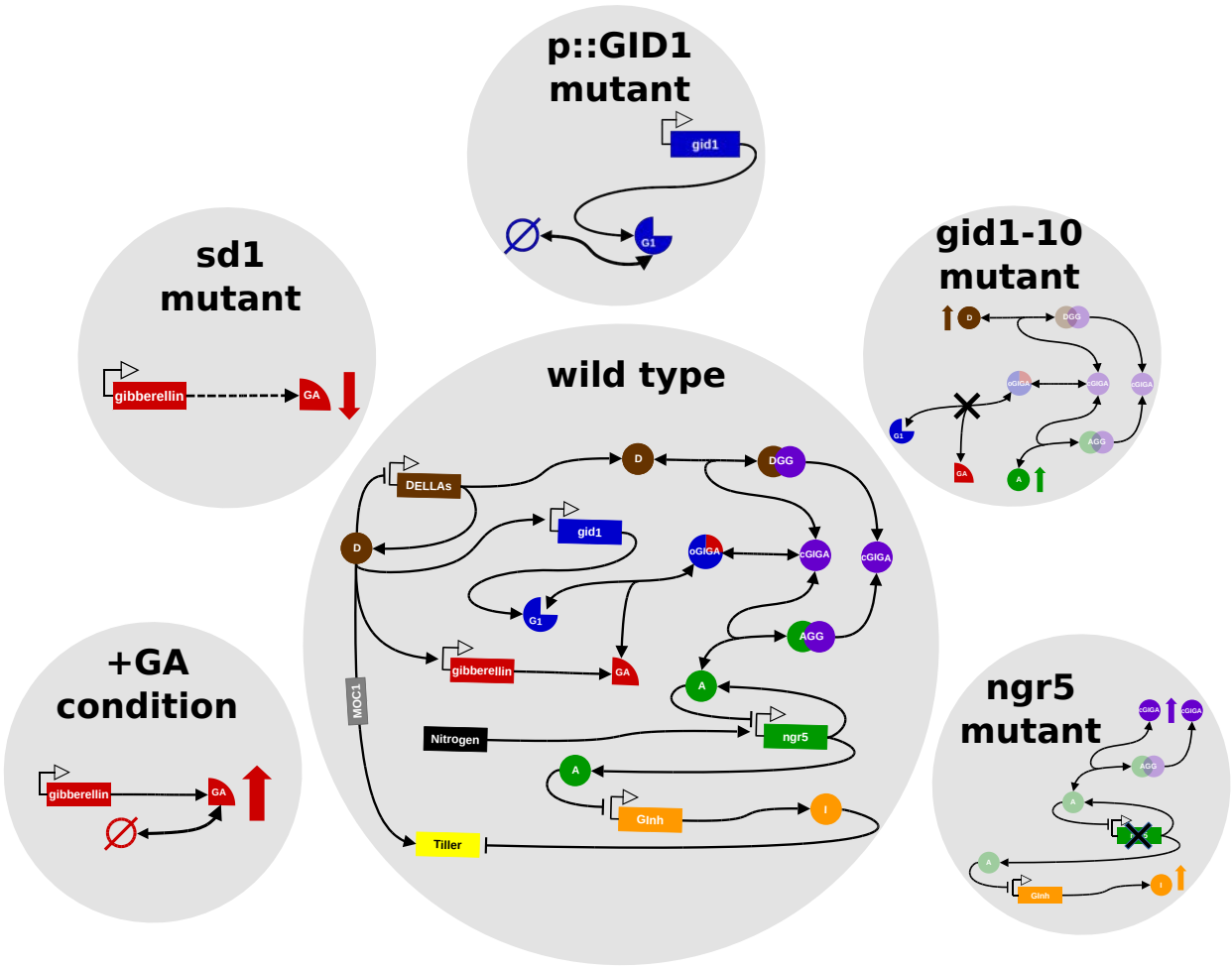

Figure 6.2: Interaction model with a graphical representation of how the individual mutations affect the system regulation.

The full regulation model is represented in the fig. 6.2 together with the graphical representation of the individual effect on the model of the experimental conditions considered in (K. Wu et al., 2020). The +GA condition is an exogenous treatment of 100 µM Gibberellin, represented as an external constant flux of GA to the system. *sd1* mutation reduces gibberellin abundance and is modeled as a reduction of GA production. *p::GID1* are plants over expressing GID1 under the control of the

cauliflower mosaic virus (CaMV) 35S promoter, represented as an increase of G1 production. *gid1-10* is a *gid1* loss-of-function mutant that impossibilities the formation of all the different G1-GA complexes. And finally, *ngr5* is a mutant with an inactivated NGR5 gene.

6.1.1 *Diferential equations for the model dynamics*

Using the scheme of interactions presented in fig. 6.2, we obtained the set of differential equations to obtain the equilibrium states of the system. If one does not consider the RNA dynamics, one can define the protein dynamics by applying the strategy shown in eq. (1.18). The temporal evolution equations for all the modeled concentrations are

$$\frac{d[GA]}{dt} = \Phi_{GA} + \rho_{GA}\frac{\frac{[D]}{K_{GAD}}}{1+\frac{[D]}{K_{GAD}}} - \alpha_{GA}[GA] - k_{Ga}[G1][GA] + k_{Gd}[G1GAo], \quad (6.1)$$

$$\frac{d[G1]}{dt} = \Phi_{G1} + \rho_{G1}\frac{\frac{[D]}{K_{G1D}}}{1+\frac{[D]}{K_{G1D}}} - \alpha_{G1}[G1] + k_{Gd}[G1GAo] - k_{Ga}[G1][GA], \quad (6.2)$$

$$\frac{d[G1GAo]}{dt} = k_{Ga}[G1][GA] - k_{Gd}[G1GAo] + k_{Lo}[G1GAc] - k_{Lc}[G1GAo], \quad (6.3)$$

$$\begin{aligned}\frac{d[G1GAc]}{dt} =& k_{Lc}[G1GAo] - k_{Lo}[G1GAc] - (k_{CD1a}+k_{CD2a})[G1GAc][D] - \\ &- (k_{CA1a}+k_{CA2a})[G1GAc][A] + (k_{CD1d}+\alpha_{DGG})[DGG1] + \\ &+ (k_{CA1d}+\alpha_{AGG})[AGG1] + k_{CD2d}[DGG2] + k_{CA2d}[AGG2], \end{aligned} \quad (6.4)$$

$$\begin{aligned}\frac{d[D]}{dt} =& \rho_D\frac{1}{1+\frac{[D]}{K_{DD}}} - \alpha_D[D] + k_{CD1d}[DGG1] + k_{CD2d}[DGG2] - \\ &- (k_{CD1a}+k_{CD2a})[G1GAc][D], \end{aligned} \quad (6.5)$$

$$\begin{aligned}\frac{d[A]}{dt} =& \rho_A\frac{\frac{[N]}{K_{AN}}}{1+\frac{[N]}{K_{AN}}+\frac{[A]}{K_{AA}}} - \alpha_A[A] - (k_{CA1a}+k_{CA2a})[G1GAc][A] + \\ &+ k_{CA1d}[AGG1] + k_{CA2d}[AGG2], \end{aligned} \quad (6.6)$$

$$\frac{d[DGG1]}{dt} = k_{CD1a}[G1GAc][D] - (k_{CD1d}+\alpha_{DGG})[DGG1], \quad (6.7)$$

$$\frac{d[DGG2]}{dt} = k_{CD2a}[G1GAc][D] - k_{CD2d}[DGG2], \quad (6.8)$$

$$\frac{d[AGG1]}{dt} = k_{CA1a}[G1GAc][A] - (k_{CA1d}+\alpha_{AGG})[AGG1], \quad (6.9)$$

$$\frac{d[AGG2]}{dt} = k_{CA2a}[G1GAc][A] - k_{CA2d}[AGG2], \quad (6.10)$$

$$\frac{d[I]}{dt} = \rho_I \frac{1}{1 + \frac{[A]}{K_I}} - \alpha_I[I], \tag{6.11}$$

$$\frac{d[T]}{dt} = \beta_T + \rho_T \frac{\frac{[D]}{K_{MOC}}}{1 + \frac{[D]}{K_{MOC}} + \frac{[I]}{K_T}} - \alpha_T[T], \tag{6.12}$$

where [GA], [G1], [G1GAo], [G1GAc], [DGG1], [DGG2], [AGG1], [AGG2], [D], [A], [I] and [T] are respectively the concentrations of gibberellin, GID1 proteins, GA-GID1 open complex, GA-GID1 close complex, DELLA-GA-GID1 stable complex, DELLA-GA-GID1 unstable complex, NGR5-GA-GID1 stable complex, NGR5-GA-GID1 unstable complex, DELLA proteins, NGR5 proteins, tiller inhibitory proteins and the tillers.

Regarding the production, we have β_T for the basal production of the tillers, and ρ_{GA}, ρ_{G1}, ρ_D, ρ_A, ρ_I, ρ_T for defining the regulated production of gibberellin, GID1, DELLA, NGR5, inhibitors and tillering. The linear degradation rate of each specie is α_{GA}, α_{G1},α_D,α_{DGG}, α_A, α_{AGG}, α_I, α_T with the specie marked in the subindex.

The affinity constants of the promoter regulation are K_{GAD}, K_{G1D}, and K_{MOC} for the activation of GA, GID1, and tillers through DELLA. K_{DD} and K_{AA} define the self-inhibition of DELLA and NGR5. K_{AN} sets de affinity of the nitrogen to the NGR5 promoter, K_I the affinity of NGR5 to the inhibitor promoter, and K_T the affinity of the inhibitors to the tiller "promoter".

The rates of transformation of the different stages of the GAG1 complex are defined by: kGa and kGd for the association and dissociation rates of the GAG1 complex; kLo and kLc for the rates for the lid opening and closing, and finally k_{CD1a} and k_{CD2a} for the formation of the stable (DGG1) and unstable (DGG2) DELLA complexes while k_{CA1a} and k_{CA2a} define the NGR5 ones (AGG1 and AGG2).

Finally, Φ_{GA} and Φ_{G1} are the rates of introduction of exogenous gibberellin and GID1 (used for the +GA and p::GID1 experimental conditions).

To simulate the different mutants and experimental conditions, we will modify a certain parameter to reproduce the experimental effect over the system. For the +GA condition, we will set a value of $\Phi_{GA} \neq 0$ (presented in table 6.1) while for the rest of the conditions this external input will not exist ($\Phi_{GA}^{WT} \equiv 0$). Analogously, for the *p::GID1* mutant, we will set a $\Phi_{G1} \equiv \varphi_{G1}\rho_{G1}$ for those mutants and $\Phi_{G1}^{WT} \equiv 0$ for all the rest. The *sd1* mutation, a reduction of GA production, is modeled with a modification of $\rho_{GA}^{sd1} \equiv \varphi_{sd1}\rho_{GA}$. And both *gid1-10* and *ngr5* are both loss-of-function mutants and as such are modeled by setting a certain rate to

0. For $gid1 - 10$ that parameter is the association rate of GID1 and Gibberellin ($k_{Ga} \equiv 0$) and for *ngr5* is the production rate of NGR5 ($\rho_A \equiv 0$).

Given the complexity of our system, we could directly solve it neither analytically nor numerically. We decided to integrate numerically its dynamics until it reached an equilibrium state. We set a maximum integration time of 4000 minutes (a bit less than 3 days). This is an arbitrary decision chosen to allow the maximum number of possible parameter sets while being a reasonable maximum value for a growth development homeostasis in all the experimental conditions. We chose something in the range of a few days given some transformations integral to the system (the opening and closing of the lid act, see value in table 6.1) in the hour scale. Nevertheless, the model could be readjusted to work in a shorter or longer timescale just by changing the restriction over the adjusted parameters. This is actually easier to do with this full integration of the dynamics than with the direct solving for steady states.

As we have already mentioned earlier, part of the parameters of our model, presented in table 6.1, are taken from a previous model (Middleton et al., 2012) of the DELLA regulation of Gibberellin. With the particular case of Φ_{GA} that has been adapted to properly reflect the treatment 100 μM of GA presented in (K. Wu et al., 2020).

Table 6.1: Subset of the parameter values used for the wild type that we fixed from the bibliography (Middleton et al., 2012).

Parameter	Description	Value	Units
α_{GA}	Lineal degradation rate of Gibberellin	0.2921	min^{-1}
α_{G1}	Lineal degradation rate of GID1	3.51	min^{-1}
α_{DGG}	Lineal degradation rate of the DELLA-GID1-Gibberellin complex	6.92	min^{-1}
α_{AGG}	Lineal degradation rate of the NGR5-GID1-Gibberellin complex	6.92	min^{-1}
k_{Ga}	GID1-Gibberellin complex association rate	$1.35 \cdot 10^{-3}$	$nM^{-1} \cdot min^{-1}$
k_{Gd}	GID1-Gibberellin complex dissociation rate	2.84	min^{-1}
k_{Lo}	GID1 lid opening rate	0.076	min^{-1}
k_{Lc}	GID1 lid closing rate	0.0251	min^{-1}
k_{CD1a}	Binding rate for the stable GID1-GA-DELLA complex	0.01	$nM^{-1} \cdot min^{-1}$
k_{CD1d}	Dissociation rate for the stable GID1-GA-DELLA complex	0.133	min^{-1}
k_{CD2a}	Binding rate for the unstable GID1-GA-DELLA complex	0.31622	$nM^{-1} \cdot min^{-1}$
k_{CD2d}	Dissociation rate for the unstable GID1-GA-DELLA complex	2.82	min^{-1}
K_{GAD}	Equilibrium constant for the inhibition of Gibberellin production through DELLA	638	nM
K_{G1D}	Equilibrium constant for the inhibition of GID1 production through DELLA	0.56	nM
K_{DD}	Equilibrium constant for the DELLA negative self-regulation	10	nM
Φ_{GA}	Additional basal production rate of Gibberellin the +GA condition	8187.076	$nM \cdot min^{-1}$

6.1.2 *Parameter adjustment*

For the rest of the parameters, presented in table 6.2, we used several rounds of custom simulated annealing algorithm exploration to set them.

Table 6.2: Subset of parameter values used for the wild type simulations that we adjusted with a simulated annealing algorithm.

Parameter	Description	Value	Units
β_T	Basal production rate of tillering	$2.582 \cdot 10^{-2}$	$nM \cdot min^{-1}$
ρ_{GA}	Maximum regulated production rates of Gibberellin	33736	$nM \cdot min^{-1}$
ρ_{G1}	Maximum regulated production rates of GID1	0.6803	$nM \cdot min^{-1}$
ρ_D	Maximum regulated production rates of DELLA	$1.298 \cdot 10^{-2}$	$nM \cdot min^{-1}$
ρ_A	Maximum regulated production rates of NGR5	0.9832	$nM \cdot min^{-1}$
ρ_I	Maximum regulated production rates of tillering inhibitors	$7.989 \cdot 10^{-3}$	$nM \cdot min^{-1}$
ρ_T	Maximum regulated production rates of tillering	3.173	$nM \cdot min^{-1}$
α_D	Lineal degradation rate of DELLA	$9.823 \cdot 10^{-4}$	min^{-1}
α_A	Lineal degradation rate of NGR5	$7.980 \cdot 10^{-4}$	min^{-1}
α_I	Lineal degradation rate of the tillering inhibitors	$4.040 \cdot 10^{-3}$	min^{-1}
α_T	Lineal degradation rate of the tillering	$7.172 \cdot 10^{-3}$	min^{-1}
k_{CA1a}	Binding rate for the stable GID1-GA-NGR5 complex	$2.705 \cdot 10^{-2}$	$nM^{-1} \cdot min^{-1}$
k_{CA1d}	Dissociation rate for the stable GID1-GA-NGR5 complex	$4.367 \cdot 10^{-3}$	min^{-1}
k_{CA2a}	Binding rate for the unstable GID1-GA-NGR5 complex	$8.867 \cdot 10^{-2}$	$nM^{-1} \cdot min^{-1}$
k_{CA2d}	Dissociation rate for the unstable GID1-GA-NGR5 complex	1.169	min^{-1}
K_{MOC}	Equilibrium constant for the DELLA mediated promotion of tillering	0.8695	nM
K_{AN}	Equilibrium constant for the promotion of NGR5 production through nitrogen	$4.043 \cdot 10^{-6}$	kg/ha
K_{AA}	Equilibrium constant for the NGR5 negative self-regulation	$4.302 \cdot 10^{-9}$	nM
K_I	Equilibrium constant for the NGR5 mediated inhibition of the production of tillering inhibitors	0.950	nM
K_T	Equilibrium constant for the tillering inhibitors inhibition of tillering	$5.667 \cdot 10^{-4}$	nM
φ_{G1}	Multiplicative factor for the additional basal production rate of GID1 in the *p::GID1* mutants	91.60	$nM \cdot min^{-1}$
φ_{sd1}	Multiplicative factor for the reduction of Gibberellin production in *sd1* mutants	0.1256	DL
T_{factor}	Multiplicative factor to rescale the tillering	0.7814	nM^{-1}

The only exception is the tillering rescaling factor T_{factor}. This factor will be set for each parameter to have the same dynamical range for all the adjustments:

$$\Lambda \equiv \frac{T_{Exp}(gid1\text{-}10, 180) - T_{Exp}(sd1\text{-}ngr5{+}GA, 180)}{T_{Sim}(gid1\text{-}10, 180) - T_{Sim}(sd1\text{-}ngr5{+}GA, 180)}$$

where $T_{Exp}(M, N)$ and $T_{Sim}(M, N)$ the tillering observed for a certain experimental condition and nitrogen amount for the experimental data in (K. Wu et al., 2020) and our model, respectively.

Then we set T_{factor} as

$$T_{factor} \equiv \begin{cases} \Lambda & \text{when} \quad \Lambda < 1 \\ 1 & \text{when} \quad \Lambda \geqslant 1 \end{cases} \tag{6.13}$$

With this restriction, eq. (6.13), we would be only rescaling downwards to shorter dynamical ranges of the tillering. This is a conservative safety assumption to ensure that we do not accept parameters with a dynamic range too low. Those parameters would be much less robust to the eventual addition of genetic noise to the system. It is also worth noting that the addition of this factor completely eliminates the identifiability of our model. But given that was never our intention, together with the huge parameter space to explore, we decided to prioritize having more leeway in the parameter selection.

But for this study, we only considered the deterministic mean-field approach. This approach allows for a much faster exploration of the parameter space through the Simulated Annealing than the one presented in section 4.3.2. This allowed us to attempt bolder adjustment strategies, such as allowing multiple changes of parameters and modifying the energy function during the search to avoid local minima.

As we are considering only the equilibrium state, the algorithm only accepted parameters whose full state changed less than 1% in the last 160min of our integration time of 4000min. This is only the automatic cutoff for the final parameter, we further check it by doing a much longer simulation for all the experimental conditions to ensure is really in equilibrium.

The algorithm explores the parameter space, as described in section 1.2.1, and evaluates each set of parameters using this energy function:

$$\begin{aligned} E \equiv \sum_M \omega_M \cdot \Big[\big[T_{Exp}(M,90) - T_{Sim}(M,90)\big]^2 + \big[T_{Exp}(M,180) - T_{Sim}(M,180)\big]^2 \Big] + \\ + \sum_N \Big[\frac{\omega_{WT}}{4} \cdot \big[T_{Exp}(WT,N) - T_{Sim}(WT,N)\big]^2 + \frac{\omega_{sd1}}{4} \cdot \big[T_{Exp}(sd1,N) - T_{Sim}(sd1,N)\big]^2 + \\ + \omega_{SFT} \cdot \Big(\big[T_{Exp}(WT,N) - T_{Exp}(sd1,N)\big] - \big[T_{Sim}(WT,N) - T_{Sim}(sd1,N)\big] \Big)^2 \Big], \end{aligned} \tag{6.14}$$

where M is the list of experimental conditions considered in (K. Wu et al., 2020)

$$\mathrm{M} \equiv \left\{ \begin{array}{cccccc} \mathrm{WT}, & \mathrm{WT{+}GA}, & \textit{sd1}, & \textit{sd1}{+}\mathrm{GA}, & \textit{gid1-10}, & \textit{gid1-10}{+}\mathrm{GA}, \\ \textit{sd1-ngr5}, & \textit{sd1-ngr5}{+}\mathrm{GA}, & \textit{sd1-p::GID1}, & \textit{sd1-p::GID1}{+}\mathrm{GA} & & \end{array} \right\},$$

and N is the list of the different nitrogen conditions considered in (K. Wu et al., 2020), besides 90 and 180 kg/ha (which are considered in the first term)

$$\mathrm{N} \equiv \Big\{ 60, \quad 120, \quad 210, \quad 300 \Big\} \text{ kg/ha},$$

and finally the weights ω_x are initially set as

$$\begin{array}{cccc} \omega_{\mathrm{WT}} = 5, & \omega_{\mathrm{WT{+}GA}} = 6, & \omega_{\textit{sd1}} = 4, & \omega_{\textit{sd1}{+}\mathrm{GA}} = 4.8, \\ \omega_{\textit{gid1-10}} = 3, & \omega_{\textit{gid1-10}{+}\mathrm{GA}} = 3.6, & \omega_{\textit{sd1-ngr5}} = 4.8, & \omega_{\textit{sd1-ngr5}{+}\mathrm{GA}} = 5,76, \\ \omega_{\textit{sd1-p::GID1}} = 3.6, & \omega_{\textit{sd1-p::GID1}{+}\mathrm{GA}} = 4,32, & \omega_{\mathrm{SFT}} = 3, & \end{array}$$

but would evolve along the exploration of parameters.

We initiated the parameter search with the previous weights excepting for $\omega'_{\mathrm{WT}} = 2\omega_{\mathrm{WT}}$ and $\omega'_{\mathrm{WT{+}GA}} = 2\omega_{\mathrm{WT{+}GA}}$ which in this case emphasizes the wild type phenotype. Then, if the system strays too far from the current best adjustment for too long, we will change the predominant phenotype. Specifically, we chose that if the algorithm stays for 2500 consecutive steps with energy bigger than $1.1\mathrm{E}_{\mathrm{Best}}$ we would change its focus. At that point, we would reset $\omega'_{\mathrm{WT}} = \omega_{\mathrm{WT}}$ and $\omega'_{\mathrm{WT{+}GA}} = \omega_{\mathrm{WT{+}GA}}$ and chose which weights with be doubled stochastically between

$$\left\{ \begin{array}{ccc} \mathrm{WT}, & \textit{sd1}, & \textit{gid1-10}, \\ \textit{ngr5}, & \textit{p::GID1}, & \mathrm{SFT} \end{array} \right\}.$$

All the weights with the selected characteristics in their subindex will be doubled until the next change. The exploration will continue changing its focus until it reaches the 1000000 steps or has an acceptance rate lower than 40% whatever comes first. At that point, we would restart the simulated annealing with the best parameter obtained.

The election of both the energy function and the weights is arbitrary and heavily influenced by the adjustment of the preliminary parameters. They changed slightly during the adjustment, but the parameter presented is the best one from, at least, 10 consecutive explorations with the methodology presented here. As we will mention later, the third term was added later in the study specifically to try to solve the biggest deviation of the model with the experimental data.

6.2 REPRODUCTION OF THE EXPERIMENTAL RESULTS

We first consider the experimental data shown in fig. 3B of (K. Wu et al., 2020) that studies how changes in the provided nitrogen affect the tillering for several experimental conditions. The comparison between our model and the observed behavior is shown in fig. 6.3.

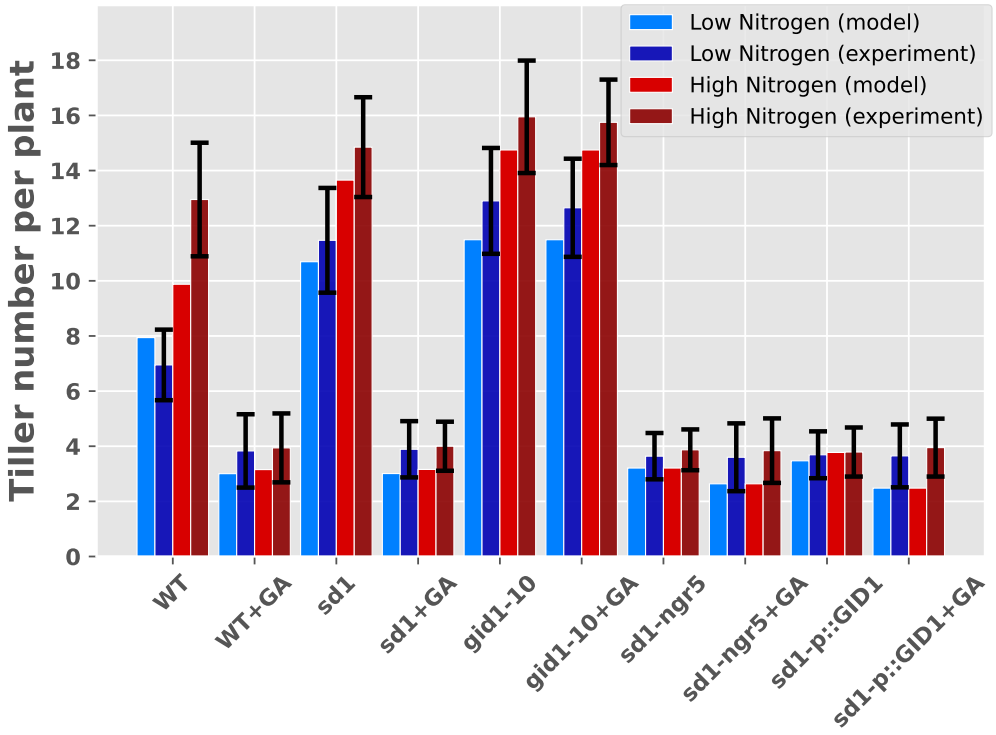

Figure 6.3: Comparison between the modeled and experimental tillering (fig. 3B of (K. Wu et al., 2020)) for low (90 kg/ha) and high (180 kg/ha) nitrogen fertilization in several experimental conditions.

The model is capable of reproducing perfectly the observed behavior described in (K. Wu et al., 2020), with the only exception of the wild type behavior. We tried several interaction configurations in which the NGR5 is the only responding to nitrogen changes, and none presented a higher difference for the WT than for the sd1 and gid-1 mutants. This was expected because those mutants increase both the NGR5 and the DELLA concentration and therefore have a double effect on the tillering.

One can observe that the sensibility of the tillering to all the different modifications is completely captured. Both the inactivation of *ngr5* or the over expression of either gibberellin (in the +GA condition) or GID1 (in the *p::GID1* condition) produce the same effect, a complete insensitivity to nitrogen with a minimal tillering number. This is easily explained because all this modifications produce the same result, a complete inactivation of the NGR5 promotion of tillering. The only exception to this phenotype

is the *gid-10*+GA condition, where the inactivation of the GID1 negates the effect of the gibberellin over NGR5. We also reproduce the increase of tillering in the *sd1* mutant due to a reduction of the gibberellin concentration in the system.

The smaller difference between the low and high nitrogen conditions for the wild type in the model could indicate that maybe there is more influence of the nitrogen through NGR5 through some other pathway that is not mediated by the GID1-GA complex. The reasoning behind this hypothesis is the following. We know that the extra wild type tillering must be mediated through NGR5 because it is not observed in the *ngr5* mutant. Additionally, this tillering response to nitrogen cannot be mediated through the GID1-GA complex because then both the *sd1* and, specially, the *gid1-10* mutants would have comparably or even larger increases. But there is probably some more nuisance with the experimental setups of (K. Wu et al., 2020) that we do not properly understand. Because if one aggregates the data from fig. 1A of (K. Wu et al., 2020) and the two columns from the wild type and the sd1 mutant in the same plot (fig. 6.4) it is easy to see that there is something strange happening. The varieties studied are the same, but their reaction to nitrogen differs considerably between the datasets of the two figures.

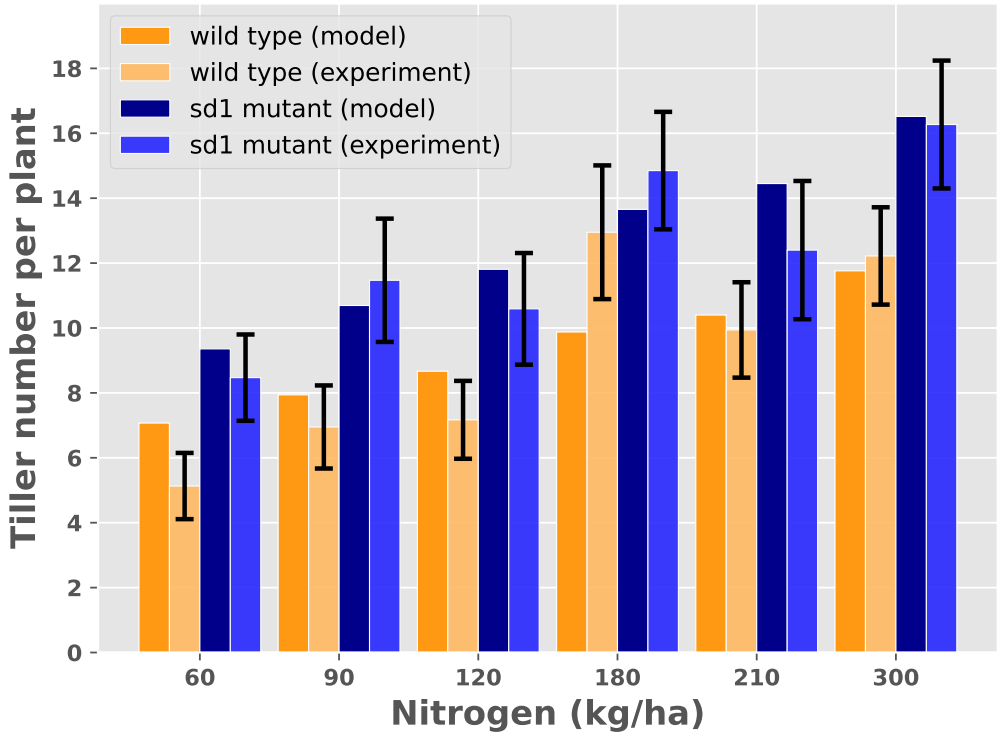

Figure 6.4: Comparison between the modeled and experimental tillering (fig. 1A and 3B of (K. Wu et al., 2020)) for a range of nitrogen fertilization in the wild type and *sd1* conditions.

As one can clearly see in fig. 6.4, the experimental values for the 90 and 180 kg/ha do not follow the same trend that the rest of the data points. This is specially evident for the 180 kg/ha wild type, which is also the more conflicting point in fig. 6.3. The rest of data point show a gradual increase of tillering in response to increases in nitrogen. Additionally, the difference between wild type and *sd1* conditions remain roughly constant. It is only in the particular 180 kg/ha dataset where these two mutants tillering almost overlap. For the rest of the datasets, the model reproduces remarkably well the experimental observations.

6.3 EVOLUTION OF THE NITROGEN EFFICIENCY

To further the study of our model, we checked the variation of the nitrogen dependence of the tillering for all the mutants in a much larger range. At the same time, we considered random initial conditions to check if the system presented hysteresis.

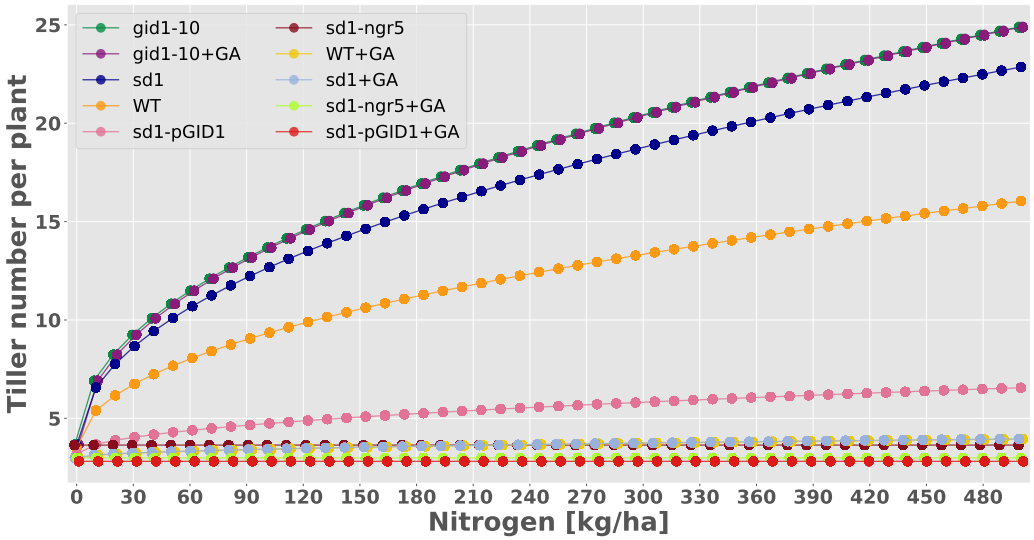

Figure 6.5: Tillering observed in the model for different experimental conditions over a large nitrogen fertilization range. To ease visualization given the amount of concentrated data points, the legend is ordered from highest to lower tillering. Each point represents the tiller equilibrium state of 100 simulations for each particular mutant. The initial conditions of each simulation are taken stochastically from a uniform distribution from 0 to 100 times the equilibrium state for each variable in the wild type equilibrium.

One can observe in fig. 6.5 that the system does not present hysteresis for any condition (all the 50 points coincide perfectly) and the overall behavior described in the previous section. Additionally, all conditions

with nitrogen sensibility present the same kind of behavior, a logarithmic growth of the tillering with respect to the nitrogen. This means that, at least for our model, nitrogen fertilization is a clear example of a diminishing returns in all the considered strains (even without considering negative effects of intensive nitrogen fertilization in the soil). There is an initial superlineal growth which is longer for the most efficient strains, *gid-10* and *sd1*, but even for those this region ends around the low nitrogen conditions (90 kg/ha). Despite this, it is worth noting that the diminishing returns effect strength is proportional to the strain efficiency over nitrogen. Due to this, if one compares those strains to the wild type, their efficiency increases with nitrogen.

6.4 EPISTASIS AND ROBUSTNESS OF THE MUTATIONS

Finally, we studied intermediate states in a linear transition from each experimental to the wild type to observe how critical are those conditions. Here, we present only the high nitrogen fertilization (180 kg/ha) conditions because the behavior for the low nitrogen is equivalent.

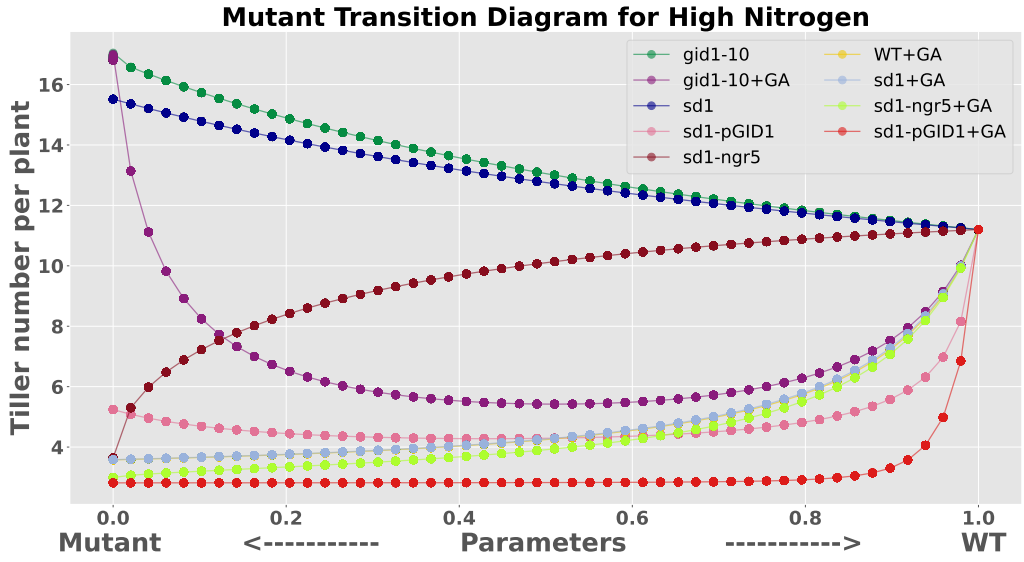

Figure 6.6: Tillering observed in the model for the intermediate states of a linear transition from the different experimental conditions to the wild type. To ease visualization given the amount of concentrated data points, the legend is ordered from highest to lower tillering. Each point represents the tiller equilibrium state of 100 simulations for each particular mutant. The initial conditions of each simulation are taken stochastically from a uniform distribution from 0 to 100 times the equilibrium state for each variable in the wild type equilibrium.

To obtain the figure presented in, fig. 6.6 we have considered the vector in the parameter space that goes from each particular mutant to the wild type. Then we obtain 50 parameter sets linearly distributed along that vector. With each parameter set, we will calculate 100 equilibrium states, initiating the system with a random initial configuration. This initial condition was taken stochastically from a uniform distribution from 0 to 100 times the equilibrium state for each variable in the wild type equilibrium. One can see that we do not observe bistability or hysteresis in the system either, given that all the 100 simulation perfectly stack over each other.

With respect to the transitions, one can see that the advantageous unique mutations *gid1-10* and *sd1* have a fairly linear behavior. This is not the case for the only unique negative one, the +GA treatment. By looking at the WT+GA curve, one can see that even with only a 20% of the original treatment, the model halves its tillering. This could indicate that there is no need to apply a treatment of 100 μM GA because with one of at least 60 μM GA we would observe similar results.

As we already saw in fig. 6.3, the +GA behavior dominate over the *sd1* in *sd1*+GA. But contrary to what one could expect, the interplay between the +GA and the *gid1-10* effect in the *gid1-10*+GA condition is more interesting. To discuss it properly, we should first define what a partial *gid1-10* mutation means. As the *gid1-10* entails a loss of function of GID1, a partial mutation could either be that a fraction of the proteins produced is completely faulty or that all of them has a defect that reduces its efficiency in that same fraction. Both of this mechanistic explanations are equivalent once one consider concentrations and the mean field approach. Then, by looking at the *gid1-10*+GA curve, we can see that a reduction of at least the 60% on the GID1 efficiency is necessary to start to nullify the GA treatment. And it isn't until less than 4% of its initial efficiency where the tillering improves with respect to the wild type. This means that, while a perfect disabling of GID1 is not necessary to observe an increase of tillering, it is necessary to protect such increase from GA treatments.

As could be expected by observing fig. 6.3 the rest of conditions with only disadvantageous modifications behave quite similarly to WT+GA curve. Only a similar tradeoff effect, albeit at a much lower degree, is observed in the *sd1-p::GID1* condition. In this condition, we have the advantageous effect of a reduction of GA production by *sd1* mixed with the negative effect of an over expression of GID1 through *p::GID1*. Overall, the overexpresion of *p::GID1* dominates, but its reduction of tillering is attenuated with higher reductions of GA production. This could indicate that the system is saturated of GID1, and due to this the concentration

of Gibberellin becomes the bottleneck for the formation of the GID1-GA complex.

6.5 SENSITIVITY ANALYSIS

More information regarding which element is the most sensitive to perturbation would inform us about the limiting elements for both the tiller promotion and inhibition.

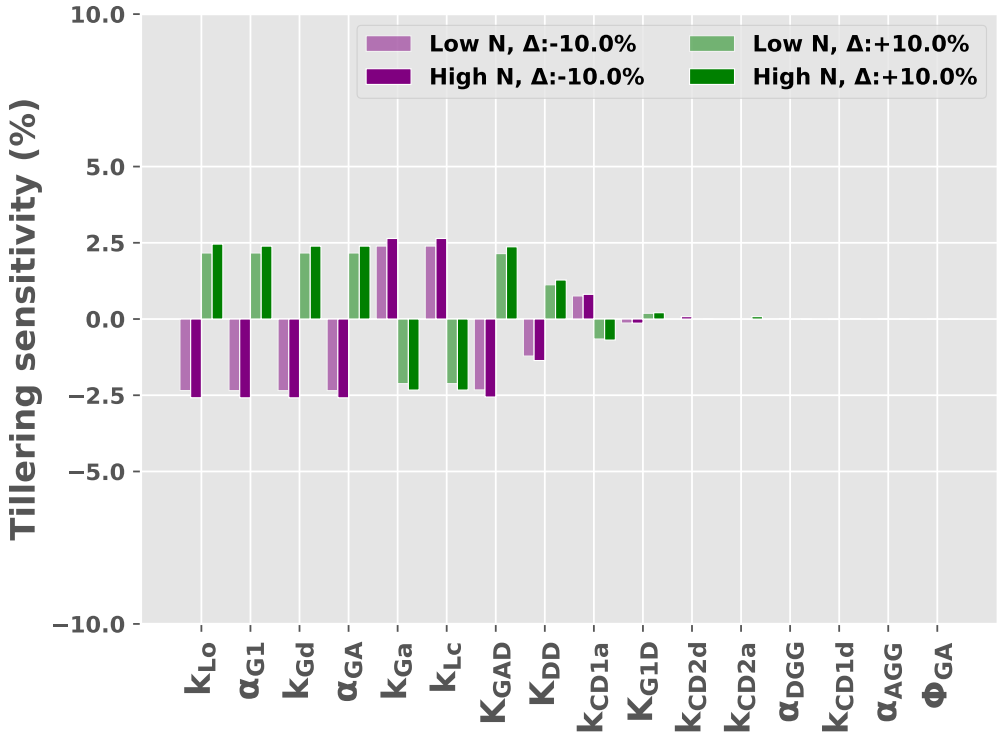

Figure 6.7: **Sensibility analysis for the fixed parameters:** Sensitivity of the number of tillers observed in a plant with respect to 10% positive (purple) and negative (green) changes in the indicated parameters. Changes are with respect to the wild type. We present the effect for both low (bright colors) and high (dark colors) nitrogen fertilization.

First, we studied the parameters adapted from the previous model (Middleton et al., 2012) in fig. 6.7. Those parameters are quite robust to perturbation, with variations of less than the 2.5% in the tillering. The ones with the higher effect are the GID1-GA complex related, and the degradation of both GA and GID1 and the affinity of DELLA to promote GA production.

On the other hand, the adjusted parameters present a higher effect to the tillering. But the first seven ones, which are the ones that surpass the 5% modification of the tiller number, are all related directly to tiller or

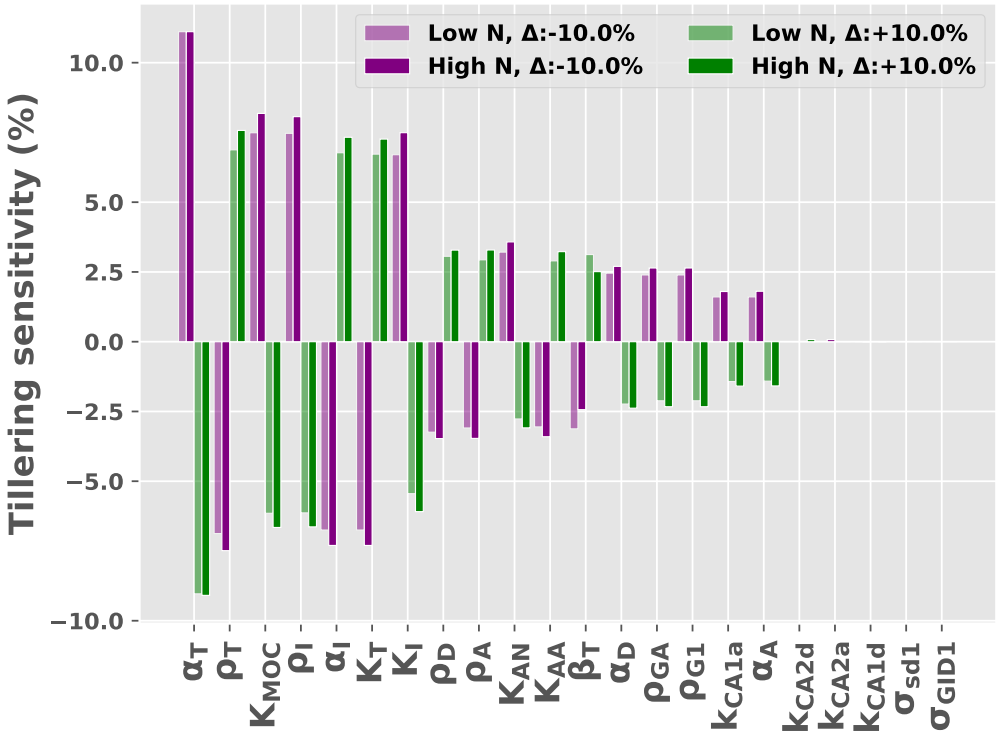

Figure 6.8: **Sensibility analysis for the adjusted parameters:** Sensitivity of the number of tillers observed in a plant with respect to 10% positive (purple) and negative (green) changes in the indicated parameters. Changes are with respect to the wild type. We present the effect for both low (bright colors) and high (dark colors) nitrogen fertilization.

inhibitor production. This was to be expected, and all of them have more or less the same impact over the tillering. The only thing that surprised us of this set is that the system is slightly more sensible to the affinity of the DELLA proteins to promote tillering (K_{MOC}) than to both the affinity of the inhibitors to the tillering (K_T) and the NGR5 to the inhibitors (K_I). This would mean that the system is less saturated with DELLA than with NGR5. But if one observes the production rates of DELLA and NGR5, the effect over the tillering is almost the same.

With these results, we are not able to properly identify the bottlenecks. Either there isn't a clear limiting reactant o the model is too simple to identify it.

6.6 PARAMETER ROBUSTNESS

Finally, we explored the parameter space, trying to obtain other sets of parameters that presented a similar agreement with the experimental data. To do so, we reexplored the parameter space with the same simulated an-

nealing presented previously. But, in this case, we started each parameter from a Gaussian distribution around the chosen parameters P_{chosen} with a variance of 1. Additionally, we do not consider any weight (all $\omega_x \equiv 0$) and allow for $T_{factor} > 1$ and we stop the exploration once we reach an energy lower than $1.05E_{chosen}$ (where E_{chosen} is the energy of the selected parameters).

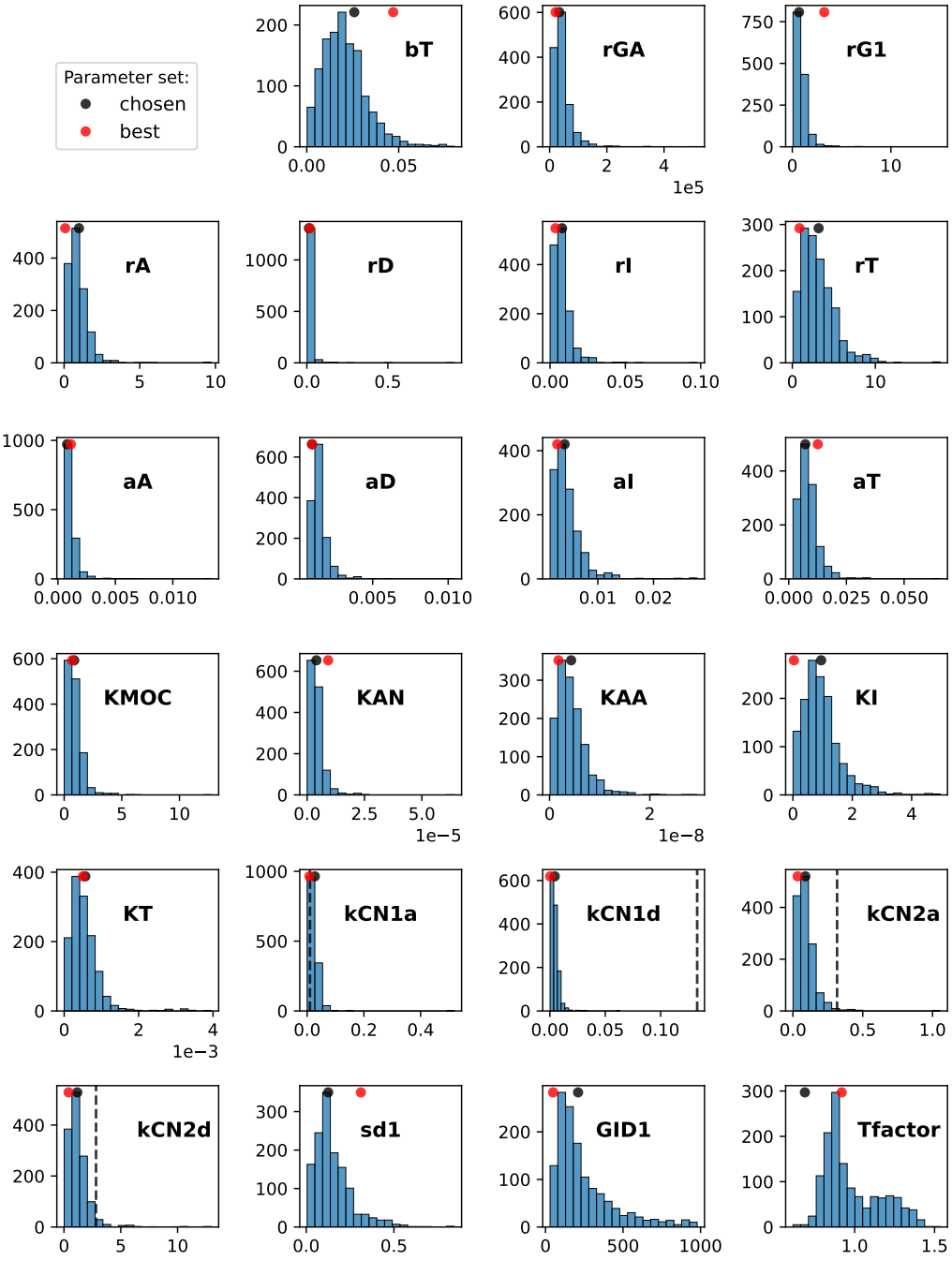

Figure 6.9: Histograms of all parameters in the model made from 1000 parameter sets that accurately fit the available data. y axes represent histogram counts and x axes the parameter values. Black dots show the parameter value in table 6.2, red dots mark the value for the set with the minimal energy, and the black dashed lines show the equivalent value for the DELLA-GID1-GA complex.

With these conditions, we obtained 1350 distinct parameter sets. But, as one can observe in fig. 6.9, all the parameter values are tightly packed around the original parameter values (marked with a black dot in the figure). Additionally, we represented the parameters value for the dynamics of the DELLA-GID1-GA complex, taken from (Middleton et al., 2012), with a black dashed line. If one compares the adjusted values with those, it is easy to see that the GID1-GA mediated degradation in our model is more important for the NGR5 regulation than for the DELLA's one. The affinity of the stable complex k_{CN1a} is the only one with a higher value than the DELLA equivalent k_{CD1a}. In addition, the value for the dissociation of the complex without degrade the protein, k_{CN1d}, is much lower than the DELLA's one,k_{CD1d}. Then, given that we supposed the same active degradation rate for the two species $\alpha_{DGG} \equiv \alpha_{DGG}$, NGR5 is being degraded with priority in our model.

Of those, there are several sets of parameters that adjusted better the experimental data. As a reference, we present in fig. 6.10 the set with the lowest difference with the experimental data.

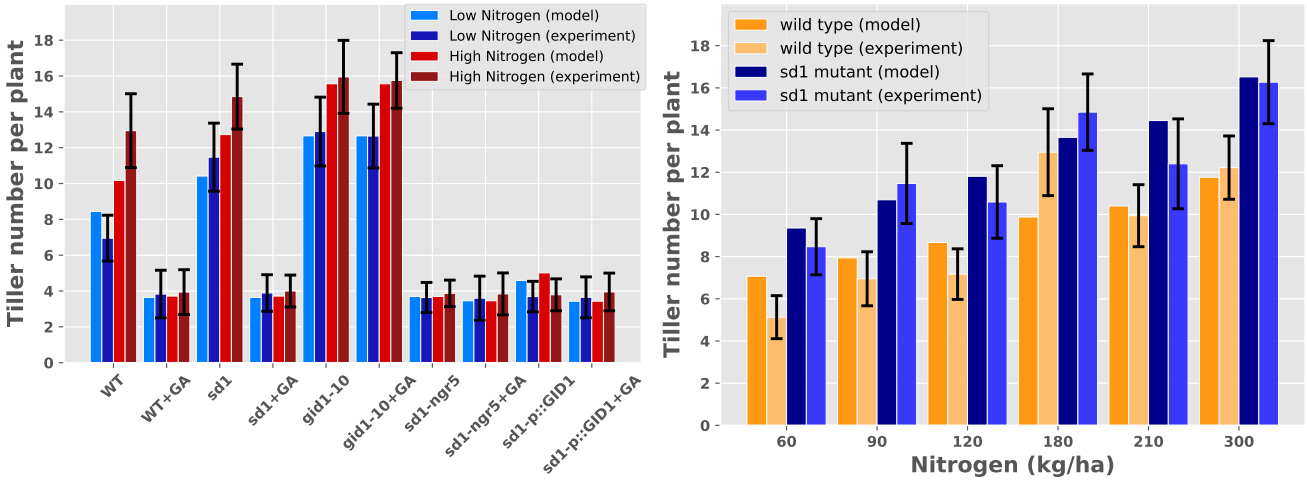

Figure 6.10: Comparison between the best parameter set and the experimental data available. This is a reproduction of both fig. 6.3 and fig. 6.4 with the best set of parameters.

One can see that, despite being a better fit, the main characteristics of the model remain the same. It does not improve the wild type discrepancy, which is the main discrepancy of the previous set. Given this, together with the small dispersion observed in fig. 6.9, one would expect really similar results for this parameter set. Additionally, the new set of parameter presents a T_{factor} much closer to 1 which would make the dynamical range of the model much smaller.

To study if there is any unexpected relation between parameters, we performed a pairwise correlation analysis between the parameters over the full 1350 distinct parameter sets.

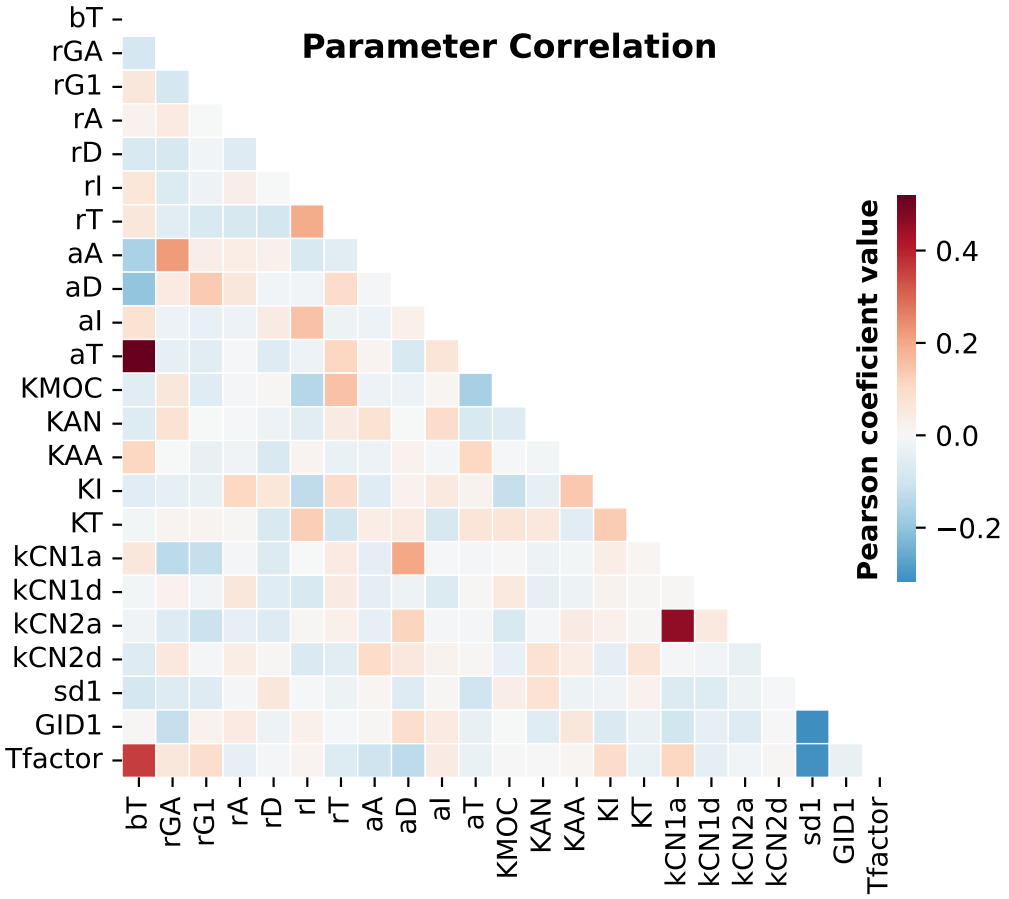

Figure 6.11: Pairwise correlation matrix between the parameters of the different 1350 parameter sets.

The correlation matrix presented in fig. 6.11 shows that the pairwise correlations between parameters is in general really low. The highest pairings, which we represent in fig. 6.12, are either obvious (such as the β_T, α_T correlation) or directly artificial. The ones that include the T_{factor} are all directly artificial, given that this parameter is a rescaling of the tillering. Likewise, the (k_{CN1a}, k_{CN2a}) correlation is also artificial because we force the association rate of the unstable GA-GID1-NGR5 to be higher than the stable one. This is a typical consideration and imitates de GA-GID1-DELLA behavior described in (Middleton et al., 2012).

It is easy to see that almost all the scatter plots in fig. 6.12 present an L shape. This indicates that the available range of one parameter is conditioned by the other, but not linearly and probably not even directly. This suggests that our model is not overfitting the data, given that we do not

observe tradeoff relationships between pairs of parameters (fig. 6.11) and most of them seem to have a similar sensibility to perturbations (figs. 6.7 and 6.8).

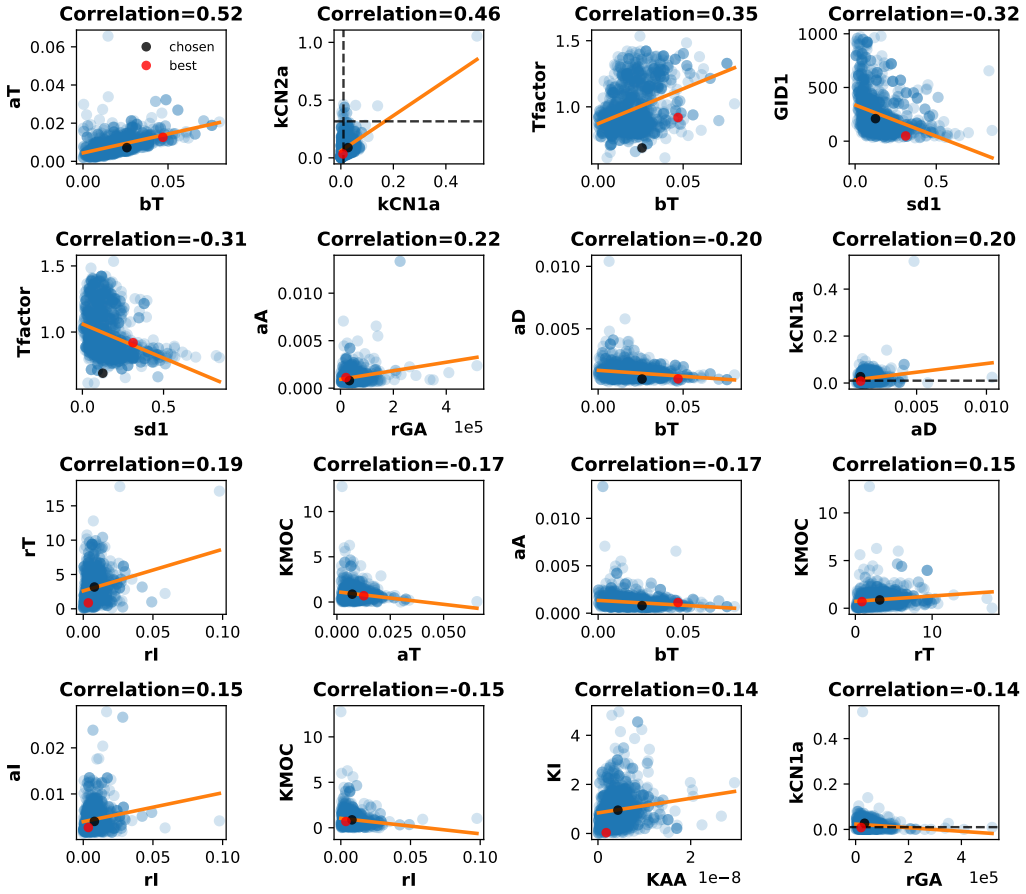

Figure 6.12: Pairwise scatter plot for the 16 parameters with the higher correlation with a linear fit in orange. Red dots mark the value of the parameter corresponding to the set with the minimal energy, while black dots represent the value of the parameter corresponding to the set used in this study.

The large number of parameters makes a completely blind initialization of the parameters when integrating the dynamics with a maximum time almost impossible for convergence reasons. Because of this, we cannot rule out with the current data that there is no other set of parameters that do not belong to the parameter manifold shown in fig. 6.9. We are currently repeating this analysis with a larger initial variance to see if we continue to see this tight distribution of parameters, or obtain a more diverse set of parameters.

6.7 CONCLUSIONS AND FUTURE WORK

In (K. Wu et al., 2020), the authors presented NGR5, a new gene that directly controls the number of tillers that the plant will develop. It is also shown that this protein is targeted by the Gibberellin-GID1 complex, which also works in the regulation of the DELLA protein through proteasomal destruction. We have shown here a minimal model that incorporates this new gene, NGR5, to the well known DELLA Gibberellin pathway, and it is capable to perfectly reproduce all the experimental data.

The model backs the hypothesis that NGR5 acts in parallel to the DELLA proteins. The same mechanism that reduces vertical growth through the accumulation of the growth repressor DELLA proteins directly induces horizontal growth through the tillering protection of NGR5. Then, this regulatory structure would allow the plant to direct the overall growth through the GID1-GA complex. This could explain the maintenance of the nitrogen sensibility in the *sd1* mutants. In the search for the most efficient dwarf varieties during the Green Revolution, the breeders inadvertently optimized the regulation of NGR5 alongside the DELLA's one.

With this interpretation, the different affinities of the complex with the NGR5 and the DELLA would mark the priority of the system. The model suggest that natively, the NGR5 is degraded with priority to the DELLA. This is reasonable from an evolutionary point of view, the plant prioritizes to stop the formation of new tillers before inducing vertical growth. Typically, plants grow to offset neighboring plants and receive more sunlight. Then is more energy efficient to delay the tillering until reaching the desired height. But what was a favorable trait in the wild, it is no longer useful in an intensive monoculture. In this situation, all the plants are equal, and the crops don't need to compete for sunlight. Due to this, should it be possible to genetically modify those affinities, it would be more efficient to prioritize DELLA degradation over NGR5.

Additionally, the model shows that the epistasis between the several modifications changes considerably with defective modifications. This is specially relevant in the more nitrogen efficient strain *gid1-10*. It is shown that to be insensitive to Gibberellin regulation the disabling of the GID1 proteins must be perfect and robust because even a small quantity of GID1-GA complex is capable to negate the increase in tillering of this mutant.

Further work is necessary to confirm these results. First, we should confirm that there is not another set of parameters outside the observed attractor. We are already working on that, starting the parameter exploration from a much disperse parameter set. If the final dispersion of parameter sets do not change, we intend to study how would the system

react to changes in the Gibberellin-GID1 complex prioritization. And secondly, we should include a more detailed description of the DELLA dynamics. As we could not find any evidence of direct promotion of DELLA by nitrogen fertilization, we considered a stable self regulated production. But, if the DELLA production is affected by nitrogen intake, the model adjustment could change, reducing their difference with the NGR5 proteins.

Part IV

FUTURE PERSPECTIVES

Here we recapitulate the main results over the two biological systems. Additionally, we will breifly discuss how the basic understanding over this biological systems can be incorporated in the current agricultural crop design.

7

CONCLUSIONS AND FUTURE PERSPECTIVES

There are two main common threads along the two topics discussed during this thesis. The first one is their framing, in both cases we model the genetic regulatory network underlying the responses of the organisms to environmental changes. This theoretical and mathematical approach allow us to reduce the system to the minimal expression capable of reproducing the observed behavior. Additionally, the results are usually more generalizable, given that the model is not a detailed description of a given system. And the second link is one whose full exploration is beyond the scope of this thesis. The two regulatory systems respond to the same type of environmental signal, albeit in completely different regimes, the abundance of nitrogen in the medium. Therefore, in this chapter we will first summarize the main takeaway messages of each study, and we will later discuss their possible joint application in crop design.

For the nitrogen fixation cyanobacteria, we knew by previous studies section 2.2 that several Turing-like minimal 3 gene models were capable to capture the formation of the pattern even in a discretized growing domain such as this. Here, we provide additional insight over this minimal models through the stability study presented in chapter 3. We show that a 2-element system with simple diffusible inhibitor (*patS*) and a localized activator (*hetR*) from (Muñoz-García and Ares, 2016) is already capable to shift the filament between homogeneous and pattern-like heterogeneous regimes with small changes in the parameter values of the model.

Then, in order to reproduce the $\Delta patA$ phenotype, we presented in chapter 4 a model that includes the requirement of maturation of HetR in order to act as a transcription factor. We hypothesize that HetF is necessary for this maturation, and PatA would enhance it. The particular role of this genes is still not well-defined in the literature, but recent experimental studies seem to confirm both the existence of an HetR maturation process (Xiaomei Xu et al., 2020) and an indirect activator role of *hetF* (W.-Y. Xing et al., 2022). There is still no evidence connecting these two mechanisms, given that they are mediated by different genes, *hetL* and *patU3*, whose role is still somewhat obscure. Despite this, our work focuses more on *patA*'s role in heterocyst regulation, so it should not be greatly modified by the particular configuration of the HetR regulation. This analysis of the *patA* mutant is specially relevant because it is a clear example that one can disrupt the formation of the pattern by affecting

the intensity of the feedback loops controlled by HetR. Here, *patA* is hypothesized to have a reinforcing role to *hetR* regulation. Then, without *patA* the fraction of HetR that gets activated is reduced with respect to the wild type, this mutant seems to lose the compounding effect that allowed the formation of the pattern. This mutant is much less susceptible to sudden spikes of HetR production, and therefore most of the stochastic fluctuations get buffered without affecting the overall homogeneity of the filament. This recovers the idea presented in chapter 3: through modulation of *patA* expression the filament is capable to transition between the pattern forming and homogeneous regimes.

This work is later expanded in chapter 5, where we observe a similar initial homogenization effect for the Δ*patS* mutant that is corrected with the apparition of heterocysts and the HetN production. Additionally, we use the correlations of both the diffusible inhibitor and the level of regulatory activity of HetR to propose a new two staged selection method of heterocysts. This novel mechanism is capable to explain the existence of a pattern with a characteristic length much larger than the experimentally observed cell to cell correlation, without the need for an initial intrinsic spacial organization.

Finally, we present our simple model of the lateral growth (tillering) in rice, chapter 6, where we study the interplay between two growth strategies. Our model suggest that the regulatory mechanism that controls both the inhibition of vertical growth and the induction of tillering is evolutionary tuned to prioritize the inhibition of vertical growth over the increased tillering. This happens due to the preferential targeting for degradation of the tillering inducer NGR5 over the growth inhibitors DELLA. If we were capable to invert this preference, we could have crops that prioritize the formation of new tillers over the reduction of height. If validated experimentally, this could open a new possibility to further enhance tiller response to increase of nitrogen fertilization. This is specially relevant because we also observed in our model that the nitrogen fertilization seems to have diminishing returns in all the strains considered.

As we have already proposed expansions for each particular study in their respective chapter, we will not reiterate these ideas here. Instead, we would like to present a more general perspective, that would constitute a new focus in our research line.

One can see that our plant and cyanobacteria studies are largely independent during this thesis, but this does not have to be like this. It is well characterized that there are several instances of plant-cyanobacteria symbiosis (Kollmen and Strieth, 2022), a well studied example being Azolla-Anabaena (D. J. Hill, 1975; Peters and J C Meeks, 1989). It has been shown that, besides the obvious growth advantages of the autonomous supple-

ment of nitrogen fixated by the cyanobacteria, the plant typically shows also increased resistance to plant diseases (Kollmen and Strieth, 2022).

The Azolla-Anabaena symbiosis is not of direct agricultural interest, given that Azolla is a fern, but it has been used as a complementary crop and manure to enricher the soil almost exclusively in traditional agriculture of Asian countries as far as 17th century (end of Ming dynasty) (Pereira, 2018; Watanabe, 1982). But the Azolla fertilization is still not suitable for intensive agriculture: it cannot sustain high yield rice crops and has to be tightly controlled to avoid cross contamination of the rice (Pereira, 2018). Despite these caveats, it can be a used as a complement to avoid soil degradation. This degradation of the soil is the main caveat of modern intensive monoculture, that has been shown to accelerate desertification, and more sustainable practices such as crop rotation techniques have to be established (Grzebisz et al., 2022).

Besides the application of already existent plant-bacteria symbiosis as a biological fertilizer, one could imagine the engineering of new symbiosis through the combination of the basic knowledge in both systems. We have already modified cyanobacteria for industrial use, in applications such as wastewater treatment or the fabrication of components such as ethanol and biomass (Abed et al., 2009; Ehira et al., 2018; Möllers et al., 2014). We could then expand this genetic manipulation to enforce the creation of advantageous biofilms in high yield crops. This is already an open area of research, with recent studies that show promising results in cotton crops (Triveni et al., 2015) and rice (Álvarez et al., 2020). There are also recent advances into the alternative option, which is to incorporate the nitrogen fixation capabilities to an organism that already presents a symbiotic relationship with a certain crop (M.-H. Ryu et al., 2020).

We believe that knowledge regarding the genetic networks that control both the nitrogen fixation and incorporation in cyanobacteria and plants are key to this more applied genetic modification field. Then studies such as the ones presented here could provide the insight necessary to successfully engineer a more sustainable crop.

APPENDIX

A

STABILITY STUDY MODEL DEDUCTION

With the brief biological introduction of the system, presented in chapter 3 along with some more details that will be exposed as the model approach develops, we can now present the elements that will make up our model as well as the interactions between them.

MODELED ELEMENTS AND REACTIONS:

As we already said our model is composed by just three genes, but we will also have to consider: the promoters[9] of those gens that can be positively regulated by HetR: HetR itself and PatS, the several forms of the proteins and the Fixed Nitrogen. Then variables considered in this model are:

- Promoters of hetR (H^x): empty promoters (H_0^x) and occupied promoters (H_A^x)
  Then by mass conservation of the promoters: $H^x = H_0^x + H_A^x$.
- HetR: monomers (r) and dimers (D) that can be inhibited (rrI) or not (rr).
  By mass conservation: $R = r + D = r + 2rr + 2rrI$.
- Inhibitors: PatS (s), HetN (n) and difusible pentapeptide (i).
- Fixed Nitrogen: F

[9] *As explained in section 1.1, a promoter is a region of the gene that has some affinity for a certain protein and controls the genetic activity.That activity can be boosted or reduced depending on the protein that binds to the promoter.*

Once we have defined our variables now it is time to enumerate all the interactions that will be considered in this model:

- Occupation of the promoters:o
  We only consider the promoter's dynamics of the genes that are regulated by HetR: *hetR* itself and *patS* (Herrero et al., 2013).

$$H_0^r + rr \underset{k_{-r}}{\overset{k_r}{\rightleftarrows}} H_A^r, \tag{A.1a}$$

$$H_0^s + rr \underset{k_{-s}}{\overset{k_s}{\rightleftarrows}} H_A^s. \tag{A.1b}$$

  The first approximation that we realize is to consider that the activated dimer have the same affinity for all the promoters of the several genes regulated by HetR $K_r = K_s$.

- Basal production of HetR (Herrero et al., 2013):

$$H_0^r \xrightarrow[\beta_r]{} H_0^r + r \tag{A.2}$$

- Regulated production of HetR and PatS (Herrero et al., 2013) in vegetative cells:

$$H_A^r \xrightarrow[\beta_r+\rho_r]{} H_A^r + r, \tag{A.3a}$$

$$H_A^s \xrightarrow[\rho_s]{} H_A^s + s. \tag{A.3b}$$

- Regulated production of HetN (S. M. Callahan and Buikema, 2001) and fixed Nitrogen in heterocysts:

$$\mathrm{Heterocyst} \xrightarrow[\rho_n]{} \mathrm{Heterocyst} + n, \tag{A.4a}$$

$$\mathrm{Heterocyst} \xrightarrow[\rho_F]{} \mathrm{Heterocyst} + F. \tag{A.4b}$$

- Conversion of PatS and HetN into the inhibitory form (Rivers et al., 2018):

$$s \xrightarrow[\delta_s]{\text{diffusion}} i, \tag{A.5a}$$

$$n \xrightarrow[\delta_n]{\text{diffusion}} i. \tag{A.5b}$$

We assume that both PatS and HetN inhibit the HetR regulation in a similar mechanistic manner (concretely in the form of a diffusible hexapeptide (Corrales-Guerrero et al., 2013). This hypothesis is supported by the evidence that both genes require the same hexapeptide (ERGSGR) for being able to inhibit heterocyst formation (Rivers et al., 2018).

- Degradation and dilution of proteins and fixed Nitrogen:

$$r \xrightarrow[\alpha_r]{} \varnothing \tag{A.6a}$$

$$rr \xrightarrow[\alpha_d]{} \varnothing \tag{A.6b}$$

$$rrI \xrightarrow[\alpha_{Id}]{} \varnothing \tag{A.6c}$$

$$s \xrightarrow[\alpha_s]{} \varnothing \tag{A.6d}$$

$$n \xrightarrow[\alpha_n]{} \varnothing \tag{A.6e}$$

$$i \xrightarrow[\alpha_i]{} \varnothing \tag{A.6f}$$

$$F \xrightarrow[\alpha_F]{} \varnothing \tag{A.6g}$$

The degradation terms in fact combine both degradation and dilution, since terms for both this effect would have the same form, they can be combined. As we do not have experimental degradation rates, the most permissive option is to suppose that each "species" is degraded with a different rate.

- HetR dimer formation:

$$r + r \underset{k_{-b}}{\overset{k_b}{\rightleftarrows}} rr \tag{A.7}$$

- Inhibition of the HetR dimers (Rivers et al., 2018):

$$rr + i + i \underset{K_d^2}{\rightleftarrows} rrI \tag{A.8}$$

  This equation supposes that the inhibition of the dimers is cooperative and that the affinity of the second inhibitor is much bigger than the affinity of the first one. We also define K_d with the quadratic form to maintain the units of concentration and then allow the direct comparison between $[i]$ and ${K_d}^2$.

We will now use the Mass Action Law to obtain the time evolution of each of our variables. This proposition assumes that the system is in thermodynamic equilibrium to affirm that the probability or rate at which a reaction occurs is proportional to the concentration of the reactants in that reaction. More specifically, this approximation assumes that there is a sufficient quantity of perfectly diluted reagents in the system, achieving that the concentration is homogeneous in the whole cell.

This approach does not have too many problems for chemical reactions carried out in test tubes since there the reagents are in complete dissolution isolated from everything that does not interest you for the specific reaction you are studying. But for in-vivo processes inside the cell this assertion is already more than questionable due to the high occupation of the cytosol, in the same way the existence of any free diffusion process inside the cells could also be questioned. There are somewhat more sophisticated alternatives that consider the possibility of local concentrations that we are not going to consider, since for a minimum model this approximation is sufficient to find coherent results. It should also be noted that most of the results of the current biochemistry have been obtained using this assumption.

ORIGINAL DIFFERENTIAL EQUATIONS:

Promoters dynamics:

$$\frac{d[H_0^r]}{dt} = -k_r[H_0^r][rr] + k_{-r}[H_A^r], \tag{A.9a}$$
$$\frac{d[H_A^r]}{dt} = k_r[H_0^r][rr] - k_{-r}[H_A^r], \tag{A.9b}$$
$$\frac{d[H_0^s]}{dt} = -k_s[H_0^s][rr] + k_{-s}[H_A^s], \tag{A.9c}$$
$$\frac{d[H_A^s]}{dt} = k_s[H_0^s][rr] - k_{-s}[H_A^s]. \tag{A.9d}$$

HetR dynamics:

$$\frac{d[r]}{dt} = \beta_r[H_0^r] + (\beta_r + \rho_r)[H_A^r] + 2k_{-b}[rr] - 2k_b[r]^2 - \alpha_r[r], \tag{A.10a}$$
$$\frac{d[rr]}{dt} = k_b[r]^2 - k_{-b}[rr] - k_d[rr][i]^2 + k_{-d}[rrI] - \alpha_d[rr], \tag{A.10b}$$
$$\frac{d[rrI]}{dt} = k_d[rr][i]^2 - k_{-d}[rrI] - \alpha_d[rrI]. \tag{A.10c}$$

PatS dynamics:

$$\frac{d[s]}{dt} = (1 - \delta_{Hcist})\rho_s[H_A^s] - 2d_s[s] - \alpha_s[s]. \tag{A.11}$$

HetN dynamics:

$$\frac{d[n]}{dt} = \delta_{Hcist}\rho_n - 2d_n[n] - \alpha_n[n]. \tag{A.12}$$

Inhibitor dynamics (of a cell j):

$$\frac{d[i]_j}{dt} = d_s([s]_{j+1} + [s]_{j-1}) + d_n([n]_{j+1} + [n]_{j-1}) + d_i([i]_{j-1} - 2[i]_j + [i]_{j+1}) - \\ - k_d[rr]_j[i]_j^2 + k_{-d}[rrI]_j - \alpha_i[i]_j. \tag{A.13}$$

Fixed Nitrogen dynamics (of a cell j):

$$\frac{d[F]_j}{dt} = \delta_{Hcist}\rho_F + d_F([F]_{j+1} - [F]_j + [F]_{j-1}) - \alpha_F[F]_j. \tag{A.14}$$

MODEL SIMPLIFICATIONS:

Our next task is to reduce the number of variables of the system by supposing that any reversible process is much faster than the transcription and translation required to produce a new protein. Then, as the simulation will be carried out for the slow time of the production of proteins,

the other processes will already have arrived at the stationary situation. where the proportion of the reactants is equal to the constant of equilibrium of the reaction.

First, we use that the promoter occupation dynamics is much faster than the *hetR* transcription and translation necessary to obtain a HetR monomer to consider that both eqs. (A.9a) and (A.9b) are in equilibrium $\left(\frac{d[H_0^r]}{dt} = \frac{d[H_A^r]}{dt} = 0\right)$.

With that, we can obtain the following expression for K_r:

$$K_r = \frac{[H_0^r][rr]}{[H_A^r]} \quad \text{where} \quad K_r = \frac{k_{-r}}{k_r}, \tag{A.15}$$

and $[rr]$ is the free, non-inhibited, dimer concentration.
Then using that the total amount of promoter sites $[H^r]$ will be constant

$$[H^r] = [H_0^r] + [H_A^r], \tag{A.16}$$

we can transform eq. (A.15) using eq. (A.16)

$$[H_A^r] = K_r^{-1}[rr][H_0^r] = K_r^{-1}[rr]\Big([H^r] - [H_A^r]\Big) \Rightarrow \frac{K_r}{[rr]} = \frac{[H^r]}{[H_A^r]} - 1 \Rightarrow [H_A^r] = [H^r]\frac{1}{1 + \frac{K_r}{[rr]}},$$

to obtain an expression of both $[H_A^r]$ and $[H_0^r]$ that only depends on the variables $[H^r]$ and $[rr]$:

$$[H_A^r] = [H^r]\frac{\frac{[rr]}{K_r}}{1 + \frac{[rr]}{K_r}}. \tag{A.17}$$

$$[H_0^r] = [H^r]\frac{1}{1 + \frac{[rr]}{K_r}}. \tag{A.18}$$

After that if we introduce eqs. (A.16) and (A.17) into eq. (A.10a) we obtain

$$\frac{d[r]}{dt} = \beta_r[H^r] + \rho_r[H^r]\frac{\frac{[rr]}{K_r}}{1 + \frac{[rr]}{K_r}} + 2k_{-b}[rr] - 2k_b[r]^2 - \alpha_r[r]$$

As we suppose that the number of total promoters of *hetR* is constant we can absorb the $[H^r]$ into the production constants β_r and ρ_r to reduce the number of parameters. The resultant expression of the evolution of the concentration of monomers is:

$$\frac{d[r]}{dt} = \beta_r + \rho_r g\big([rr]\big) + 2k_{-b}[rr] - 2k_b[r]^2 - \alpha_r[r], \tag{A.19}$$

where we define:

$$\phi([rr]) \equiv \frac{\frac{[rr]}{K_r}}{1+\frac{[rr]}{K_r}}. \tag{A.20}$$

Analogously, using the approximation that $K_r = K_a = K_s$, we can obtain the following expression for eq. (A.11):

$$\frac{d[s]}{dt} = (1-\delta_{Hcist})\rho_s\phi([rr]) - 2d_s[s] - \alpha_s[s], \tag{A.21}$$

Secondly, we are going to consider that both the HetR dimers formation and inhibition (which in both cases consist on the binding/unbinding between existent element) is much faster that the protein synthesis. As a result of this, we can consider that the reactions (A.7) and (A.8) are in equilibrium and therefore:

$$[rrI] = \frac{[rr][i]^2}{K_d^2} \quad \text{where} \quad K_d^2 = \frac{k_{-d}}{k_d}, \tag{A.22}$$

$$[rr] = \frac{[r]^2}{K_b} \quad \text{where} \quad K_b = \frac{k_{-b}}{k_b}. \tag{A.23}$$

If we introduce eq. (A.22) on eq. (A.13) we obtain:

$$\begin{aligned}\frac{d[i]_j}{dt} = {} & d_s([s]_{j+1} + [s]_{j-1}) + d_n([n]_{j+1} + [n]_{j-1}) + \\ & + d_i([i]_{j-1} - 2[i]_j + [i]_{j+1}) - \alpha_i[i]_j.\end{aligned} \tag{A.24}$$

Additionally the total amount HetR ([R]) in the system will be distributed as:

$$[R] = [r] + 2[rr] + 2[rrI] = [r] + 2\frac{[r]^2}{K_b} + 2\frac{[r]^2}{K_b}\frac{[i]^2}{K_d^2} \tag{A.25}$$

thus,

$$[r] = \frac{-K_b K_d^2 \pm \sqrt{K_b^2 K_d^4 + 8K_b K_d^2 [R](K_d^2 + [i]^2)}}{4(K_d^2 + [i]^2)}. \tag{A.26}$$

This expression eq. (A.26) is the solution of the second grade equation eq. (A.25) and represents the concentration of free monomers of HetR. Is it worth noting that we obtained this second grade equation eq. (A.25) due to the competition between the binding of the inhibitors and the unbinding of the HetR dimers. We could have considered instead, and we will in the next section, chapter 4, that the dimerization reaches equilibrium much faster than the inhibition. In that case we would lose this competition and the two processes can be considered to act independently of each other.

Once we have the expression for the free monomers, we can transform eq. (A.20) to only depend on the monomers by using eq. (A.23):

$$\phi([r],[i]) = \frac{\frac{[r]^2}{K_b}}{1+\frac{[r]^2}{K_r}} = \frac{\frac{[r]^2}{K_b K_r}}{1+\frac{[r]^2}{K_b K_r}}. \tag{A.27}$$

and likewise eq. (A.24) changes to

$$\begin{aligned}\frac{d[i]_j}{dt} = {} & d_s\left([s]_{j+1}+[s]_{j-1}\right) + d_n\left([n]_{j+1}+[n]_{j-1}\right)+ \\ & + d_i\left([i]_{j-1}-2[i]_j+[i]_{j+1}\right) - \alpha_i[i]_j.\end{aligned} \tag{A.28}$$

Additionally we can obtain the dynamics of this new variable R using eqs. (A.19), (A.25), (A.10b) and (A.10c):

$$\begin{aligned}\frac{d[R]}{dt} = \frac{d[r]}{dt}+\frac{d[rr]}{dt}+\frac{d[rrI]}{dt} = {} & \beta_r + \rho_r g\left([rr]\right) + \cancel{2k_{-b}[rr]} - \cancel{2k_b[r]^2} - \alpha_r[r]+ \\ + \cancel{k_b[r]^2} - \cancel{2k_{-b}[rr]} - \cancel{k_d[rr][i]^2} + \cancel{k_{-d}[rrI]} - \alpha_d[rr] + \cancel{k_d[rr][i]_j^2} - \cancel{k_{-d}[rrI]} - \alpha_d[rrI] = \\ = {} & \beta_r + \rho_r g\left([rr]\right) - \alpha_r[r] - \alpha_d[rr] - \alpha_d[rrI] = \\ = {} & \beta_r + \rho_r g\left([rr]\right) - \alpha_r[r] - \frac{[r]^2}{K_b}\left(\alpha_d + \alpha_{Id}\frac{[i]^2}{K_d^2}\right).\end{aligned} \tag{A.29}$$

COMPLETE MODEL:

With all the simplifications presented before, we obtain the final set of differential equations that we will use to simulate the behavior of the system.

$$\frac{dR_j(t)}{dt} = \beta_r + \phi(r_j,i_j,F_j)\rho_r - \alpha_r r(R_j,i_j) - \frac{r(R_j,i_j)^2}{K_b}\left[\alpha_d + \alpha_{Id}\frac{i_j^2}{K_d^2}\right] \tag{A.30}$$

$$\frac{dp_j(t)}{dt} = (1-\delta_{hc,j})\phi(r_j,i_j,F_j)\rho_p - 2d_p p_j - \alpha_p p_j \tag{A.31}$$

$$\frac{dn_j(t)}{dt} = \delta_{hc,j}\rho_n - 2d_n n_j - \alpha_n n_j \tag{A.32}$$

$$\frac{di_j(t)}{dt} = d_p(p_{j-1}+p_{j+1}) + d_n(n_{j-1}+n_{j+1}) + d_i(i_{j-1}-2i_j+i_{j+1}) - \alpha_i i_j \tag{A.33}$$

$$\frac{dF_j(t)}{dt} = \delta_{hc,j}\rho_F + d_F(F_{j-1}-2F_j+F_{j+1}) - \alpha_F F_j \tag{A.34}$$

with the boundary condition that if $j+1$ or $j-1$ is outside the filament then $X_{j+1}=0$ or $X_{j-1}=0$ and where

$$r_j(R_j,i_j) = \frac{-K_b K_d^2 + \sqrt{K_b^2 K_d^4 + 8K_b K_d^2 R_j\left(K_d^2+i_j^2\right)}}{4\left(K_d^2+i_j^2\right)}, \tag{A.35}$$

$$\phi(r_j,i_j,F_j)^* = \frac{\frac{r(R_j,i_j)^2}{K_b K_r}}{1+\frac{r(R_j,i_j)^2}{K_b K_r}+\frac{F_j(t)}{K_d}} \tag{A.36}$$

* To be able to reproduce the HetN deletion mutant we add, phenomenology, a lineal inhibition from the fixed nitrogen products (F), if we do not add this term the deletion mutant of *hetN* shows a completely differentiated filament due to the fact that we do not have considered the nitrogen sensing system (which is mediated through *ntcA*)

B

FULLY MECHANISTIC DEDUCTION OF THE PATA/HETF MODEL

In this section, we will present the rigorous full mechanistic approach to the modeling of the system as an example of what we could use in a system with more biochemical information available. This is also why we chose to not absorb any of the parameters besides the promoters and the equilibrium constants. But, given that we only have biochemical information regarding the affinity between the inhibitor and the HetR dimers, we will heavily simplify the model in chapter 4 to a more phenomenological version without competition/interferences between the different processes.

MODELED ELEMENTS:

With the brief biological introduction of the system, presented in chapter 4, we have enough information to present the elements that will make up our model and the interactions between them. The variables considered in this model are:

- Gene promoters of the gen x (H^x): empty promoters (H_0^x) and occupied promoters (H_A^x)
  Then by mass conservation of the promoters: $H^x = H_0^x + H_A^x$
- HetR: monomers (r) and dimers (d) which in turn we will also divide into two types:
  - Not activated dimers (d): free dimer (rr), inhibited dimer (rrI).
  - Activated dimers (Arr), inhibited activated dimer (ArrI).
- Inhibitors: PatS (s), HetN (n) and diffusible pentapeptide (i).
- Activators: PatA and HetF that are presented in several forms:
  - Free: HetF (f) and PatA (a).
  - Forming activation complexes: frr and farr.
- Fixed Nitrogen: G

MODELED REACTIONS:

- Occupation of the promoters:
  We only consider the promoter dynamics of the genes that are regulated by HetR: *hetR* itself, *patS* and *patA* (Herrero et al., 2013).

$$H_0^r + Arr \underset{K_r}{\rightleftarrows} H_A^r \tag{B.1}$$

$$H_0^s + Arr \underset{K_s}{\rightleftarrows} H_A^s \tag{B.2}$$

$$H_0^a + Arr \underset{K_a}{\rightleftarrows} H_A^a \tag{B.3}$$

  The first approximation that we realize is to consider that the activated dimer has the same affinity for all the promoters of the several genes regulated by HetR $K_r = K_s = K_a$.

- Basal production of HetR (Herrero et al., 2013) and HetF (Risser and Sean M. Callahan, 2008):

$$H_0^r \underset{\beta_r}{\longrightarrow} H_0^r + r \tag{B.4}$$

$$H_0^f \underset{\beta_f}{\longrightarrow} H_0^f + f \tag{B.5}$$

- Regulated production of HetR, PatS and PatA (Herrero et al., 2013) in vegetative cells:

$$H_A^r \underset{\beta_r+\rho_r}{\longrightarrow} H_A^r + r \tag{B.6}$$

$$H_A^s \underset{\rho_s}{\longrightarrow} H_A^s + s \tag{B.7}$$

$$H_A^a \underset{\rho_a}{\longrightarrow} H_A^a + a \tag{B.8}$$

- Regulated production of HetN (S. M. Callahan and Buikema, 2001) and fixed Nitrogen (Fay et al., 1968) in heterocysts:

$$\text{Heterocyst} \underset{\rho_n}{\longrightarrow} \text{Heterocyst} + n \tag{B.9}$$

$$\text{Heterocyst} \underset{\rho_G}{\longrightarrow} \text{Heterocyst} + G \tag{B.10}$$

- Conversion of PatS and HetN into the inhibitory form (Rivers et al., 2018):

$$s \xrightarrow[c_s]{\text{diffusion}} i \tag{B.11}$$

$$n \xrightarrow[c_n]{\text{diffusion}} i \tag{B.12}$$

We assume that both PatS and HetN inhibit the HetR regulation in a similar mechanistic manner (concretely in the form of a diffusible hexapeptide (Corrales-Guerrero et al., 2013). This hypothesis is supported by the evidence that both genes require the same hexapeptide (ERGSGR) for being able to inhibit heterocyst formation (Rivers et al., 2018).

- Degradation and dilution of proteins and fixed Nitrogen:

$$r \xrightarrow[\alpha_r]{} \varnothing \tag{B.13a}$$

$$rr \xrightarrow[\alpha_d]{} \varnothing \tag{B.13b}$$

$$rrI \xrightarrow[\alpha_{Id}]{} \varnothing \tag{B.13c}$$

$$Arr \xrightarrow[\alpha_{Ad}]{} \varnothing \tag{B.13d}$$

$$ArrI \xrightarrow[\alpha_M]{} \varnothing \tag{B.13e}$$

$$s \xrightarrow[\alpha_s]{} \varnothing \tag{B.13f}$$

$$n \xrightarrow[\alpha_n]{} \varnothing \tag{B.13g}$$

$$i \xrightarrow[\alpha_i]{} \varnothing \tag{B.13h}$$

$$f \xrightarrow[\alpha_f]{} \varnothing \tag{B.13i}$$

$$a \xrightarrow[\alpha_a]{} \varnothing \tag{B.13j}$$

$$G \xrightarrow[\alpha_G]{} \varnothing \tag{B.13k}$$

The degradation terms in fact combine both degradation and dilution, since terms for both this effect would have the same form, they can be combined. There are several degradation rates for the HetR dimers due to the fact that experimental results indicate that both inhibited (Risser and Sean M. Callahan, 2009) and activated (Risser and Sean M. Callahan, 2008) dimers degrade faster than the unaltered ones. We consider that the modifications to the degradation rate are not cumulative, so the inhibited activated dimer(ArrI) will have $\alpha_M = \max(\alpha_{Ad}, \alpha_{Id})$

- HetR dimer formation:

$$r + r \underset{K_b}{\rightleftarrows} rr \tag{B.14}$$

- Activation of the HetR dimers[§] (Orozco et al., 2006; Risser and Sean M. Callahan, 2008):

$$rr + f \underset{k_{-f}}{\overset{k_f}{\rightleftarrows}} frr \underset{\tau_f}{\longrightarrow} Arr + f \tag{B.15}$$

$$rr + f + a \underset{k_{-a}}{\overset{k_a}{\rightleftarrows}} farr \underset{\tau_a}{\longrightarrow} Arr + f + a \tag{B.16}$$

- Inhibition of the HetR dimers (Rivers et al., 2018):

$$rr + i + i \underset{K_d^2}{\rightleftarrows} rrI \tag{B.17}$$

$$Arr + i + i \underset{K_d^2}{\rightleftarrows} ArrI \tag{B.18}$$

This equation supposes that the inhibition of the dimers is cooperative and that the affinity of the second inhibitor is much bigger than the affinity of the first one. We also define K_d with the quadratic form to maintain the units of concentration and then allow the direct comparison between $[i]$ and ${K_d}^2$. Given that there is no evidence that the inhibitor targets preferentially the activated dimers and the fact that the experimental stoichiometry of HetR and inhibitors is 1:1 (Feldmann et al., 2011, 2012) we consider that Arr and rr are indistinguishable for the inhibitor.

[§] *We initially considered an even bigger equivalent system where the reaction* $rr + f + a \underset{k_{-a}}{\overset{k_a}{\rightleftarrows}} farr$ *was substituted by both* $rr + e \underset{k_{-e}}{\overset{k_e}{\rightleftarrows}} err$ *and* $f + a \underset{K_A}{\rightleftarrows} e$. *But, as the PatA and HetF interactions are not known, this only introduced an additional unknown parameter to adjust.*

ORIGINAL DIFFERENTIAL EQUATIONS:

As we already did in the previous appendix A, we can use the Mass Action Law over the previous reactions to obtain the time evolution of each of our variables. Again, this proposition assumes that the system is in thermodynamic equilibrium. Then we suppose that a mean field approach to the cell interior without any intracellular heterogeneity or transport. In those homogeneous conditions, the probability or rate at which a reaction occurs is proportional to the concentration of the reactants in that reaction.

Promoters dynamics:

$$\frac{d[H_0^r]}{dt} = -k_r[H_0^r][Arr] + k_{-r}[H_A^r], \tag{B.19a}$$
$$\frac{d[H_A^r]}{dt} = k_r[H_0^r][Arr] - k_{-r}[H_A^r], \tag{B.19b}$$
$$\frac{d[H_0^s]}{dt} = -k_s[H_0^s][Arr] + k_{-s}[H_A^s], \tag{B.19c}$$
$$\frac{d[H_A^s]}{dt} = k_s[H_0^s][Arr] - k_{-s}[H_A^s], \tag{B.19d}$$
$$\frac{d[H_0^a]}{dt} = -k_a[H_0^a][Arr] + k_{-a}[H_A^a], \tag{B.19e}$$
$$\frac{d[H_A^a]}{dt} = k_a[H_0^a][Arr] - k_{-a}[H_A^a]. \tag{B.19f}$$

HetR dynamics:

$$\frac{d[r]}{dt} = \beta_r[H_0^r] + \big(\beta_r + \rho_r\big)[H_A^r] + 2k_{-b}[rr] - 2k_b[r]^2 - \alpha_r[r], \tag{B.20a}$$
$$\begin{aligned}\frac{d[rr]}{dt} &= k_b[r]^2 - k_{-b}[rr] - k_f[rr][f] + k_{-f}[frr] - k_a[rr][f][a] + \\ &\quad + k_{-a}[farr] - k_d[rr][i]^2 + k_{-d}[rrI] - \alpha_d[rr],\end{aligned} \tag{B.20b}$$
$$\frac{d[rrI]}{dt} = k_d[rr][i]^2 - k_{-d}[rrI] - \alpha_{Id}[rrI], \tag{B.20c}$$
$$\begin{aligned}\frac{d[Arr]}{dt} &= \tau_f[frr] + \tau_a[farr] - k_d[Arr][i]^2 + k_{-d}[ArrI] - \\ &\quad - k_r[H_0^r][Arr] + k_{-r}[H_A^r] - k_s[H_0^s][Arr] + k_{-s}[H_A^s] - \\ &\quad - k_a[H_0^a][Arr] + k_{-a}[H_A^a] - \alpha_{Ad}[Arr],\end{aligned} \tag{B.20d}$$
$$\frac{d[ArrI]}{dt} = k_d[Arr][i]^2 - k_{-d}[ArrI] - \alpha_M[ArrI]. \tag{B.20e}$$

PatS dynamics:

$$\frac{d[s]}{dt} = \big(1 - \delta_{Hcist}\big)\rho_s[H_A^s] - 2c_s[s] - \alpha_s[s]. \tag{B.21}$$

HetN dynamics:

$$\frac{d[n]}{dt} = \delta_{Hcist}\rho_n - 2c_n[n] - \alpha_n[n]. \tag{B.22}$$

Inhibitor dynamics (of a cell j):

$$\begin{aligned}\frac{d[i]_j}{dt} &= c_s\big([s]_{j+1} + [s]_{j-1}\big) + c_n\big([n]_{j+1} + [n]_{j-1}\big) - 2d_i[i]_j - \alpha_i[i]_j - \\ &\quad - 2k_d[rr]_j[i]_j^2 + 2k_{-d}[rrI]_j - 2k_d[Arr]_j[i]_j^2 + 2k_{-d}[ArrI]_j.\end{aligned} \tag{B.23}$$

PatA dynamics:

$$\frac{d[a]}{dt} = \rho_a[H^a_A] - k_a[rr][f][a] + k_{-a}[farr] + \tau_a[farr] - \alpha_a[a]. \tag{B.24}$$

HetF dynamics:

$$\begin{aligned}\frac{d[f]}{dt} = \beta_f - k_f[rr][f] + k_{-f}[frr] + \tau_f[frr] - \\ - k_a[rr][f][a] + k_{-a}[farr] + \tau_a[farr] - \alpha_f[f].\end{aligned} \tag{B.25}$$

Activation complexes dynamics:

$$\frac{d[frr]}{dt} = k_f[rr][f] - k_{-f}[frr] - \tau_f[frr], \tag{B.26}$$

$$\frac{d[farr]}{dt} = k_a[rr][f][a] - k_{-a}[farr] - \tau_a[farr]. \tag{B.27}$$

Fixed Nitrogen dynamics (of a cell j):

$$\frac{d[G]_j}{dt} = \delta_{Hcist}\rho_G + d_G\left([G]_{j+1} - [G]_j + [G]_{j-1}\right) - \alpha_G[G]_j. \tag{B.28}$$

MODEL SIMPLIFICATIONS:

Off all the presented variables of our model to reduce the computational efforts we are only going to simulate a few of them, the ones with the slowest dynamics (that is the one that involves protein production through transcription and translation). Those are the total amount of the main regulator HetR (r), the inhibitor producers PatS (s), the produced fixed Nitrogen (G), the inhibiting hexapeptide (i), and the two activating proteins PatA(a) and HetF(f).

First, we will simplify the promotor occupation dynamics (eqs. (B.19))*. These reactions are much faster than the *hetR* transcription and translation necessary to obtain a HetR monomer (the reactions (B.4) and (B.6)). Due to this, we can consider that both eqs. (B.19a) and (B.19b) are in equilibrium $\left(\frac{d[H^r_0]}{dt} = \frac{d[H^r_A]}{dt} = 0\right)$.

* *This first simplification is analogous to the one presented in appendix A. You can jump directly to its final expression eq. (B.34) if you want.*

With that we can obtain the following expression for K_r:

$$K_r = \frac{[H^r_0][Arr]}{[H^r_A]} \quad \text{where} \quad K_r = \frac{k_{-r}}{k_r}. \tag{B.29}$$

Then using that the total amount of promoter sites $[H^r]$ will be constant

$$[H^r] = [H^r_0] + [H^r_A], \tag{B.30}$$

we can transform eq. (B.29) using eq. (B.30)

$$[H^r_A] = K_r^{-1}[Arr][H^r_0] = K_r^{-1}[Arr]\Big([H^r] - [H^r_A]\Big) \Rightarrow \frac{K_r}{[Arr]} = \frac{[H^r]}{[H^r_A]} - 1 \Rightarrow$$

$$\Rightarrow [H^r_A] = [H^r]\frac{1}{1 + \frac{K_r}{[Arr]}},$$

to obtain an expression of both $[H^r_A]$ and $[H^r_0]$ that only depends on the variables $[H^r]$ and $[Arr]$:

$$[H^r_A] = [H^r]\frac{\frac{[Arr]}{K_r}}{1 + \frac{[Arr]}{K_r}}. \tag{B.31}$$

$$[H^r_0] = [H^r]\frac{1}{1 + \frac{[Arr]}{K_r}}. \tag{B.32}$$

After that if we introduce eqs. (B.30) and (B.31) into eq. (B.20a) we obtain

$$\frac{d[r]}{dt} = \beta_r[H^r] + \rho_r[H^r]\frac{\frac{[Arr]}{K_r}}{1 + \frac{[Arr]}{K_r}} + 2k_{-b}[rr] - 2k_b[r]^2 - \alpha_r[r]$$

As we suppose that the number of total promoters of hetR is constant we can absorb the $[H^r]$ into the production constants β_r and ρ_r to reduce the number of parameters. The resultant expression of the evolution of the concentration of monomers is:

$$\frac{d[r]}{dt} = \beta_r + \rho_r g\big([Arr]\big) + 2k_{-b}[rr] - 2k_b[r]^2 - \alpha_r[r], \tag{B.33}$$

where we define:

$$g([Arr]) \equiv \frac{\frac{[Arr]}{K_r}}{1 + \frac{[Arr]}{K_r}}. \tag{B.34}$$

Analogously, using the approximation that $K_r = K_a = K_s$, we can obtain the following expressions for eqs. (B.21), (B.24) and (B.20d):

$$\frac{d[s]}{dt} = (1 - \delta_{Hcist})\rho_s g([Arr]) - 2c_s[s] - \alpha_s[s], \tag{B.35}$$

$$\frac{d[a]}{dt} = \rho_a g\big([Arr]\big) - k_a[rr][f][a] + k_{-a}[farr] + \tau_a[farr] - \alpha_a[a], \tag{B.36}$$

$$\begin{aligned}\frac{d[Arr]}{dt} = {} & \tau_f[frr] + \tau_a[farr] - k_d[Arr][i]^2 + k_{-d}[ArrI] - \\ & - [H^r]\left[+\cancel{k_r\Big(1 - g\big([Arr]\big)\Big)[Arr]} - \cancel{k_{-r}g\big([Arr]\big)}\right] - \\ & - [H^s]\left[+\cancel{k_r\Big(1 - g\big([Arr]\big)\Big)[Arr]} - \cancel{k_{-r}g\big([Arr]\big)}\right] - \\ & - [H^a]\left[+\cancel{k_r\Big(1 - g\big([Arr]\big)\Big)[Arr]} - \cancel{k_{-r}g\big([Arr]\big)}\right] - \alpha_{Ad}[Arr],\end{aligned}$$

$$\frac{d[Arr]}{dt} = \tau_f[frr] + \tau_a[farr] - k_d[Arr][i]^2 + k_{-d}[ArrI] - \alpha_{Ad}[Arr]. \quad (B.37)$$

Secondly, we also know that HetR activation (the reactions (B.15) and (B.16)), which we suppose is an enzymatic reaction, is much faster than HetR production. As a result of that, we can consider that both the activation complexes are in equilibrium. Imposing $\frac{d[frr]}{dt} = \frac{d[farr]}{dt} = 0$ we obtain:

$$[frr] = K_F[rr][f] \quad \text{where} \quad K_F = \frac{k_f}{k_{-f} + \tau_f} \quad (B.38a)$$

$$[farr] = K_A[rr][f][a] \quad \text{where} \quad K_A = \frac{k_a}{k_{-a} + \tau_a}. \quad (B.38b)$$

We also know that the total amount of proteins a and f[††] is conserved for this equation, then:

$$[F] \equiv [f] + [frr] + [farr] \quad (B.39a)$$

$$[A] \equiv [a] + [farr]. \quad (B.39b)$$

This conservation can be proven by writing the dynamics of those variables considering only the activation complex formation:

$$\begin{aligned}
\frac{d[f]}{dt} &= -k_f[rr][f] + k_{-f}[frr] + \tau_f[frr] - k_a[rr][f][a] + k_{-a}[farr] + \tau_a[farr], \\
\frac{d[frr]}{dt} &= k_f[rr][f] - k_{-f}[frr] - \tau_f[frr], \\
\frac{d[a]}{dt} &= -k_a[rr][f][a] + k_{-a}[farr] + \tau_a[farr], \\
\frac{d[farr]}{dt} &= k_a[rr][f][a] - k_{-a}[farr] - \tau_a[farr],
\end{aligned}$$

$$\frac{d[F]}{dt} = \frac{d[f]}{dt} + \frac{d[frr]}{dt} + \frac{d[farr]}{dt} = 0.$$

$$\frac{d[A]}{dt} = \frac{d[a]}{dt} + \frac{d[farr]}{dt} = 0.$$

†† On the more simplified model that we will use in chapter 4 we will avoid modeling the full HetF and PatA dynamics. In order to do so, we will consider that HetF is only expressed basally at a lower level. Due to this, the total amount of [F] in the system will be roughly constant and low. Then its concentration will be much lower than the one of both its substrate, HetR, and assistant, PatA (which are up regulated during nitrogen deprivation). Due to this, the error associated with considering that all the HetF is active all the time is low. If all, our simplified model in chapter 4 will overestimate the efficiency of HetF in cells with low HetR and PatA concentrations. This occurs typically near a heterocyst where the activation is less relevant anyway due to heavy inhibition.

Using eqs. (B.38) and eq. (B.39) we can obtain:

$$[A] = [a] + K_A[rr][f][a] \implies [a] = \frac{[A]}{1 + K_A[rr][f]}$$

$$[F] = [f] + K_F[rr][f] + K_A[rr][f][a] \implies [F] = [f] + K_F[rr][f] + \frac{K_A[rr][f][A]}{1 + K_A[rr][f]} \implies$$

$$\implies [F] + K_A[rr][f][F] = [f] + K_A[rr][f]^2 + K_F[rr][f] + K_F K_A[rr]^2[f]^2 + K_A[rr][f][A] \implies$$

$$\implies \Big(K_A[rr]\Big(1 + K_F[rr]\Big)\Big)[f]^2 + \Big(1 + [rr]\Big(K_F + K_A\Big([A] - [F]\Big)\Big)\Big)[f] - [F] = 0$$

The solutions for this quadratic equation are:

$$[f] = \frac{-1 - [rr]\Big(K_F + K_A\big([A] - [F]\big)\Big) \pm \sqrt{\Big[1 + [rr]\Big(K_F + K_A\big([A] - [F]\big)\Big)\Big]^2 + 4K_A[rr]\Big(1 + K_F[rr]\Big)[F]}}{2K_A[rr]\Big(1 + K_F[rr]\Big)} \tag{B.40}$$

and therefore for $[a]$[††] in equilibrium:

$$[a] = \frac{[A]}{1 + K_A[rr][f]} \tag{B.41}$$

[††] *This same reason (much lower concentration of HetF than PatA) allows us to simplify PatA dynamics. We can consider that* $[a] >> [farr]$ *and therefore* $[A] \approx [a]$*. Then we can directly simulate the concentration of free PatA without distorting much of the dynamics.*

We can obtain the dynamics of $[F]$ and $[A]$ using eqs. (B.25), (B.36) and (B.39) and eqs. (B.38):

$$\frac{d[A]}{dt} = \frac{d[a]}{dt} + \frac{d[farr]}{dt} = \rho_a g\big([Arr]\big) - \alpha_a[a] \tag{B.42}$$

$$\frac{d[F]}{dt} = \frac{d[f]}{dt} + \frac{d[frr]}{dt} + \frac{d[farr]}{dt} = \beta_f - \alpha_f[f]. \tag{B.43}$$

And using eqs. (B.38) on eqs. (B.37) and (B.20b) we obtain:

$$\frac{d[rr]}{dt} = k_b[r]^2 - k_{-b}[rr] - \tau_F[rr][f] - \tau_A[rr][f][a] - k_d[rr][i]^2 + k_{-d}[rrI] - \alpha_d[rr], \tag{B.44}$$

$$\frac{d[Arr]}{dt} = \tau_F[rr][f] + \tau_A[rr][f][a] - k_d[Arr][i]^2 + k_{-d}[ArrI] - \alpha_{Ad}[Arr]. \tag{B.45}$$

where

$$\tau_F = \tau_f K_F = \frac{\tau_f k_f}{k_{-f} + \tau_f}$$
$$\tau_A = \tau_a K_A = \frac{\tau_a k_a}{k_{-a} + \tau_a} \tag{B.46}$$

and the expressions for $\left[f\left([rr],[F],[A]\right)\right]$ and $\left[a\left([rr],[F],[A]\right)\right]$ in equilibrium are defined by eqs. (B.40) and (B.41).

Lastly, we are going to consider that both HetR dimer formation (binding and unbinding of two monomers) and HetR inhibition (which consist of the binding/unbinding between a pair of inhibitors and a HetR dimer) are much faster than the protein synthesis. As a result of this we can consider that the reactions (B.14),(B.17), an (B.18) are in equilibrium:

$$[rr] = \frac{[r]^2}{K_b} \quad \text{where} \quad K_b = \frac{k_{-b}}{k_b}, \tag{B.47}$$

$$[rrI] = \frac{[rr][i]^2}{K_d^2} \quad \text{where} \quad K_d^2 = \frac{k_{-d}}{k_d}, \tag{B.48}$$

$$[ArrI] = \frac{[Arr][i]^2}{K_d^2} \quad \text{where} \quad K_d^2 = \frac{k_{-d}}{k_d}, \tag{B.49}$$

First, we will consider the conservation of the total amount of inactivated HetR and activated dimers for these two interactions:

$$[R] \equiv [r] + 2[rr] + 2[rrI] \tag{B.50a}$$

$$[AR] \equiv [Arr] + [ArrI] \tag{B.50b}$$

This conservation can be proven by writing the dynamics of those variables considering only the dimerization and the inhibition:

$$\left.\begin{aligned} \tfrac{d[r]}{dt} &= -2k_b[r]^2 + 2k_{-b}[rr] \\ \tfrac{d[rr]}{dt} &= k_b[r]^2 - k_{-b}[rr] - k_d[rr][i]^2 + k_{-d}[rrI] \\ \tfrac{d[rrI]}{dt} &= k_d[rr][i]^2 - k_{-d}[rrI] \end{aligned}\right\} \frac{d[R]}{dt} = \frac{d[r]}{dt} + 2\frac{d[rr]}{dt} + 2\frac{d[rrI]}{dt} = 0$$

$$\left.\begin{aligned} \tfrac{d[Arr]}{dt} &= -k_d[Arr][i]^2 + k_{-d}[ArrI] \\ \tfrac{d[ArrI]}{dt} &= k_d[Arr][i]^2 - k_{-d}[ArrI] \end{aligned}\right\} \frac{d[AR]}{dt} = \frac{d[Arr]}{dt} + \frac{d[ArrI]}{dt} = 0.$$

Therefore,

$$[R] = [r] + 2\frac{[r]^2}{K_b} + 2\frac{[r]^2}{K_b}\frac{[i]^2}{K_d^2} \Rightarrow \frac{1}{K_b}\left(1 + \frac{[i]^2}{K_d^2}\right)[r]^2 + [r] - [R] = 0 \Rightarrow$$

$$\Rightarrow [r] = \frac{-K_b K_d^2 \pm \sqrt{K_b^2 K_d^4 + 8K_b K_d^2\left(K_d^2 + [i]^2\right)[R]}}{4\left(K_d^2 + [i]^2\right)} = \frac{-1 \pm \sqrt{1 + \frac{4[R]}{K_b}\left(1 + \frac{[i]^2}{K_d^2}\right)}}{\frac{2}{K_b}\left(1 + \frac{[i]^2}{K_d^2}\right)}, \tag{B.51}$$

$$[AR] = [Arr] + [Arr]\frac{[i]^2}{K_d^2} \Rightarrow [Arr] = \frac{[AR]}{1 + \frac{[i]^2}{K_d^2}}. \tag{B.52}$$

Additionally, one should also consider the conservation of the inhibitors: $[I] = [i] + [rrI] + [ArrI]$:

$$[I] = [i] + [rr]\frac{[i]^2}{K_d^2} + [Arr]\frac{[i]^2}{K_d^2} = [i] + \frac{[r]^2}{K_b}\frac{[i]^2}{K_d^2} + \frac{[AR]}{1+\frac{[i]^2}{K_d^2}}\frac{[i]^2}{K_d^2} \Rightarrow$$

$$\Rightarrow \Big([I]-[i]\Big)\Big(K_d^2+[i]^2\Big) = \frac{[r]^2}{K_b}[i]^2 + [AR][i]^2 \Rightarrow$$

$$\Rightarrow [i]^3 + \Big(\frac{[r]^2}{K_b} + [AR] - [I]\Big)[i]^2 + K_d^2[i] - K_d^2[I] = 0. \tag{B.53}$$

Then if we solve this cubic equation, we will obtain the expression of free inhibitor ($[i]$)‡‡.

As a consequence of the Bolzano theorem we can affirm that every cubic equation with real coefficients has at least one real solution, the other two solutions may be two distinct real roots, a multiple real root or two complex conjugated roots. To know which is our case we can use the cubic discriminant:

$$\Delta_3 = b^2c^2 - 4ac^3 - ab^3d - 27a^2d^2 + 18abcd, \tag{B.54}$$

which is positive ($\Delta_3 > 0$) if the equation has three distinct real roots, zero ($\Delta_3 = 0$) if the equation has a multiple root and another distinct real root and negative ($\Delta_3 < 0$) if the equation has two complex conjugated roots and a real root.

Firstly, we will obtain the discriminant for our particular case (eq. (B.53)):

$$\Delta_3 = \left(\frac{[r]^2}{K_b} + [AR] - [I]\right)^2 K_d^4 -$$
$$-4K_d^6 + \left(\frac{[r]^2}{K_b} + [AR] - [I]\right)^3 [I]K_d^2 -$$
$$-27[I]^2K_d^4 - 18\left(\frac{[r]^2}{K_b} + [AR] - [I]\right)K_d^4[I],$$

‡‡ This cubic equation appears as a result of considering that a significant amount of inhibitor is attached to HetR. Due to this, the inhibitory effect would be less relevant in cells with more concentration of HetR because the sequestration of the inhibitor would create a diffusive bottleneck. If one considers instead that the amount of inhibitor in the system is much higher than the total amount of HetR [R] (given the lower expression of *hetR* (Di Patti et al., 2018)), then the error associated to simulate directly the dynamics of the free inhibitor would be much lower. It simply there won't be enough inactivated dimmers [rr] to effectively protect the active ones [Arr] through inhibitor sequestration.

If we rearrange

$$\Delta_3 = K_d^4 \left[\left(\frac{[r]^2}{K_b} + [AR] - [I] \right) \left(\frac{[r]^2}{K_b} + [AR] + \frac{(\frac{[r]^2}{K_b} + [AR] - [I])^2 [I]}{K_d^2} - 19[I] \right) - \right.$$
$$\left. - 4K_d^2 - 27[I]^2 \right]$$

And since we only need to consider if $\left\{ \begin{array}{l} \Delta_3 < 0 \\ \Delta_3 = 0 \\ \Delta_3 > 0 \end{array} \right\}$ we can ignore the K_d^4 factor as $K_d^4 \geqslant 0$:

$$\Delta_3 = \left(\frac{[r]^2}{K_b} + [AR] - [I] \right) \left(\frac{[r]^2}{K_b} + [AR] + \frac{(\frac{[r]^2}{K_b} + [AR] - [I])^2 [I]}{K_d^2} - 19[I] \right) -$$
$$- 4K_d^2 - 27[I]^2. \tag{B.55}$$

Using this discriminant we can decide, for each particular case, if we need to compute all the solutions or only the one that is always real (eq. (B.56)).

One can obtain the solutions using the Cardano's formula which estates that given a cubic Polynomial $P : ax^3 + bx^2 + cx + d = 0$ the solutions are:

$$x_1 = S + T - \frac{b}{3a} \tag{B.56}$$
$$x_2 = -\frac{S+T}{2} - \frac{b}{3a} + \frac{i\sqrt{3}}{2}(S - T) \tag{B.57}$$
$$x_2 = -\frac{S+T}{2} - \frac{b}{3a} - \frac{i\sqrt{3}}{2}(S - T), \tag{B.58}$$

where $S = \sqrt[3]{R + \sqrt{Q^3 + R^3}}$, $T = \sqrt[3]{R - \sqrt{Q^3 + R^3}}$, and in turn $Q = \frac{3ac - b^2}{9a^2}$, $R = \frac{9abc - 27a^2d - 2b^3}{54a^3}$.

And after computing the solutions using eqs. (B.56) to (B.58) with our particular coefficients from eq. (B.53), we will accept the one with values between 0 and [I]. This solution will be represented in this document as [i].

Then the time evolution of [R] and [AR] and [I] will be:

$$\begin{aligned}\frac{d[R]}{dt} &= \frac{d[r]}{dt} + 2\frac{d[rr]}{dt} + 2\frac{d[rrI]}{dt} = \\ &= \beta_r + \rho_r g\left([AR],[i]\right) - \alpha_r[r] - 2\tau_F[rr][f] - 2\tau_A[rr][f][a] - 2\alpha_d[rr] - 2\alpha_{Id}[rrI] = \\ &= \beta_r + \rho_r g\left([AR],[i]\right) - 2\tau_F\frac{[r]^2}{K_b}[f] - 2\tau_A\frac{[r]^2}{K_b}[f][a] - \alpha_r[r] - 2\alpha_d\frac{[r]^2}{K_b} - 2\alpha_{Id}\frac{[r]^2[i]^2}{K_b K_d^2}, \end{aligned} \tag{B.59}$$

$$\begin{aligned}\frac{d[AR]}{dt} &= \frac{d[Arr]}{dt} + \frac{d[ArrI]}{dt} = \tau_F[rr][f] + \tau_A[rr][f][a] - \alpha_{Ad}[Arr] - \alpha_M[ArrI], \\ &= \tau_F\frac{[r]^2}{K_b}[f] + \tau_A\frac{[r]^2}{K_b}[f][a] - \alpha_{Ad}\frac{[AR]}{1+\frac{[i]^2}{K_d^2}} - \alpha_M\frac{[AR]\frac{[i]^2}{K_d^2}}{1+\frac{[i]^2}{K_d^2}}, \end{aligned} \tag{B.60}$$

$$\frac{d[I]_j}{dt} = \frac{d[i]}{dt} + 2\frac{d[rrI]}{dt} + 2\frac{d[ArrI]}{dt} = d_s\left([s]_{j+1} + [s]_{j-1}\right) + d_n\left([n]_{j+1} + [n]_{j-1}\right) - 2d_i[i]_j - \alpha_i[i]_j. \tag{B.61}$$

And we can obtain the expression for $g([AR],[i])$ introducing eq. (B.52) on eq. (B.34) and adding a lineal inhibition from the fixed nitrogen products ([G])¶:

$$g([AR],[i],[NF]) = \frac{\frac{[AR]}{K_r}}{1+\frac{[i]^2}{K_d^2}+\frac{[AR]}{K_r}+\frac{[G]}{K_d}} \tag{B.62}$$

¶ As we already did in the previous model appendix A we add, phenomenology, a lineal inhibition from the fixed nitrogen products (G) to our model. Given tot we are not considering the nitrogen sensing system (which is mediated through *ntcA*) we have to add this term to reproduce the nitrogen effect over the heterocyst differentiation

COMPLETE MODEL:

With all the simplifications presented before, we present here the final set of differential equations that can be used use to reproduce the behavior of the system with a more identifiable version of our model.

$$\frac{dR_j}{dt} = \beta_r + \rho_r g\left(AR_j, i_j\right) - 2\tau_F \frac{r_j^2}{K_b} f_j - 2\tau_A \frac{r_j^2}{K_b} f_j a_j - \alpha_r r_j - 2\alpha_d \frac{r_j^2}{K_b} - 2\alpha_{Id} \frac{r_j^2 i_j^2}{K_b K_d^2} \tag{B.63}$$

$$\frac{dAR_j}{dt} = \tau_F \frac{r_j^2}{K_b} f_j + \tau_A \frac{r_j^2}{K_b} f_j a_j - \alpha_{Ad} \frac{AR_j}{1 + \frac{i_j^2}{K_d^2}} - \alpha_M \frac{AR_j \frac{i_j^2}{K_d^2}}{1 + \frac{i_j^2}{K_d^2}} \tag{B.64}$$

$$\frac{dA_j}{dt} = \rho_a g\left(AR_j, i_j\right) - \alpha_a a_j \tag{B.65}$$

$$\frac{dF_j}{dt} = b_f - \alpha_f f_j \tag{B.66}$$

$$\frac{ds_j}{dt} = (1 - \delta_{hc,j})\rho_s g\left(AR_j, i_j\right) - 2c_s s_j - \alpha_s s_j \tag{B.67}$$

$$\frac{dn_j}{dt} = \delta_{hc,j}\rho_n - 2c_n n_j - \alpha_n n_j \tag{B.68}$$

$$\frac{dI_j}{dt} = d_p(p_{j-1} + p_{j+1}) + c_n(n_{j-1} + n_{j+1}) + d_i(i_{j-1} - 2i_j + i_{j+1}) - \alpha_i i_j \tag{B.69}$$

$$\frac{dG_j}{dt} = \delta_{hc,j}\rho_G + d_G(G_{j-1} - 2G_j + G_{j+1}) - \alpha_G G_j \tag{B.70}$$

with the boundary condition that if $j+1$ or $j-1$ is outside the filament then $X_{j+1} = 0$ or $X_{j-1} = 0$ and where

$$\tau_F = \tau_f K_F = \frac{\tau_f k_f}{k_{-f} + \tau_f}, \qquad \tau_A = \tau_a K_A = \frac{\tau_a k_a}{k_{-a} + \tau_a},$$

$$g(AR_j, i_j, NF_j) = \frac{\frac{AR_j}{K_r}}{1 + \frac{i_j^2}{K_d^2} + \frac{AR_j}{K_r} + \frac{G_j}{K_d}}, \tag{B.71}$$

$$r_j(R_j, i_j) = \frac{-K_b K_d^2 + \sqrt{K_b^2 K_d^4 + 8K_b K_d^2 R_j\left(K_d^2 + i^2\right)}}{4\left(K_d^2 + i^2\right)}, \tag{B.72}$$

$$f_j(R_j, i_j, F_j, A_j) = \frac{-1 - \frac{r_j^2}{K_b}\left(K_F + K_A\left(A_j - F_j\right)\right) + \sqrt{\left[1 + \frac{r_j^2}{K_b}\left(K_F + K_A\left(A_j - F_j\right)\right)\right]^2 + 4F_j K_A \frac{r_j^2}{K_b}\left(1 + K_F \frac{r_j^2}{K_b}\right)}}{2K_A \frac{r_j^2}{K_b}\left(1 + K_F \frac{r_j^2}{K_b}\right)}, \tag{B.73}$$

$$a_j(R_j, i_j, F_j, A_j) = \frac{A_j}{1 + K_A \frac{r_j^2}{K_b} f_j}, \tag{B.74}$$

with i_j as the solution of this cubic equation (B.53):

$$i_j^3 + \left(\frac{r_j^2}{K_b} + AR_j - I_j\right) i_j^2 + K_d^2 i_j - I_j K_d^2 = 0. \tag{B.75}$$

BIBLIOGRAPHY

Aarts, E. H. L. and Jan Korst (1989). *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Chichester [England] ; New York: Wiley. ISBN: 978-0-471-92146-2.

Abed, R. M. M. et al. (Jan. 2009). "Applications of Cyanobacteria in Biotechnology." In: *J. Appl. Microbiol.* 106.1, pp. 1–12. ISSN: 13645072. DOI: 10.1111/j.1365-2672.2008.03918.x. pmid: 19191979.

Alberts, Bruce (2015). *Molecular Biology of the Cell*. Sixth edition. New York, NY: Garland Science, Taylor and Francis Group. ISBN: 978-0-8153-4432-2.

Allard, Jun F. et al. (Dec. 2007). "Heterocyst Patterns without Patterning Proteins in Cyanobacterial Filaments." In: *Dev. Biol.* 312.1, pp. 427–434. ISSN: 00121606. DOI: 10.1016/j.ydbio.2007.09.045. pmid: 17976569.

Álvarez, Consolación et al. (Aug. 2020). "Endophytic Colonization of Rice (*Oryza Sativa* L.) by the Symbiotic Strain *Nostoc Punctiforme* PCC 73102." In: *MPMI* 33.8, pp. 1040–1045. ISSN: 0894-0282, 1943-7706. DOI: 10.1094/MPMI-01-20-0015-SC.

Archibald, John M. (Oct. 2015). "Endosymbiosis and Eukaryotic Cell Evolution." In: *Current Biology* 25.19, R911–R921. ISSN: 09609822. DOI: 10.1016/j.cub.2015.07.055.

Arite, Tomotsugu et al. (2009). "D14, a Strigolactone-Insensitive Mutant of Rice, Shows an Accelerated Outgrowth of Tillers." In: *Plant Cell Physiol.* 50.8, pp. 1416–1424. DOI: 10.1093/pcp/pcp091.

Asai, Hironori et al. (Oct. 2009). "Cyanobacterial Cell Lineage Analysis of the Spatiotemporal *hetR* Expression Profile during Heterocyst Pattern Formation in *Anabaena* Sp. *PCC 7120*." In: *PLoS ONE* 4.10. Ed. by Bruce Riley, e7371. ISSN: 19326203. DOI: 10.1371/journal.pone.0007371. pmid: 19823574.

Ashikari, Motoyuki et al. (2005). "Cytokinin Oxidase Regulates Rice Grain Production." In: *Science* 309.5735, pp. 741–745. DOI: 10.1126/science.1113373.

Baker, Rodger W. and Gabor T. Herman (1972). "Simulation of Organisms Using a Developmental Model Part 2: The Heterocyst Formation Problem in Blue-Green Algae." In: *Int. J. Biomed. Comput.* 3.4, pp. 251–267. ISSN: 0020-7101. DOI: 10.1016/0020-7101(72)90030-X.

Barrass, Iain et al. (2006). "Mode Transitions in a Model Reaction–Diffusion System Driven by Domain Growth and Noise." In: *Bull. Math. Biol.* 68.5, pp. 981–995. DOI: 10.1007/s11538-006-9106-8.

Bastet, Laurène et al. (Oct. 2010). "NtcA Regulates *patA* Expression in *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 192.19, pp. 5257–5259. ISSN: 10985530. DOI: 10.1128/JB.00640-10. pmid: 20639317.

Bauer, C.C et al. (June 1997). "Suppression of Heterocyst Differentiation in *Anabaena PCC* 7120 by a Cosmid Carrying Wild-Type Genes Encoding Enzymes for Fatty Acid Synthesis." In: *FEMS Microbiol. Lett.* 151.1, pp. 23–30. ISSN: 0378-1097. DOI: 10.1111/j.1574-6968.1997.tb10390.x.

Black, T. A. and C. P. Wolk (1994). "Analysis of a Het- Mutation in *Anabaena* Sp. Strain *PCC 7120* Implicates a Secondary Metabolite in the Regulation of Heterocyst Spacing." In: *J. Bacteriol.* 176.8, pp. 2282–2292. ISSN: 00219193. DOI: 10.1128/jb.176.8.2282-2292.1994. pmid: 8157596.

Black, Todd A. et al. (July 1993). "Spatial Expression and Autoregulation of *hetR*, a Gene Involved in the Control of Heterocyst Development in *Anabaena*." In: *Mol. Microbiol.* 9.1, pp. 77–84. ISSN: 13652958. DOI: 10.1111/j.1365-2958.1993.tb01670.x. pmid: 8412673.

Borthakur, Pritty B. et al. (July 2005). "Inactivation of *patS* and *hetN* Causes Lethal Levels of Heterocyst Differentiation in the Filamentous Cyanobacterium *Anabaena* Sp. *PCC 7120*." In: *Mol. Microbiol.* 57.1, pp. 111–123. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2005.04678.x. pmid: 15948953.

Brinkmann, Felix et al. (2018). "Post-Turing Tissue Pattern Formation: Advent of Mechanochemistry." In: *PLoS Comput. Biol.* 14.7, e1006259. DOI: 10.1371/journal.pcbi.1006259.

Brown, Aidan I. and Andrew D. Rutenberg (Aug. 2012a). "Heterocyst Placement Strategies to Maximize the Growth of Cyanobacterial Filaments." In: *Phys. Biol.* 9.4, p. 46002. ISSN: 14783967. DOI: 10.1088/1478-3975/9/4/046002. pmid: 22733109.

— (Feb. 2012b). "Reconciling Cyanobacterial Fixed-Nitrogen Distributions and Transport Experiments with Quantitative Modelling." In: *Phys. Biol.* 9.1, p. 16007. ISSN: 14783967. DOI: 10.1088/1478-3975/9/1/016007. pmid: 22313598.

— (Jan. 2014). "A Storage-Based Model of Heterocyst Commitment and Patterning in Cyanobacteria." In: *Phys. Biol.* 11.1, p. 16001. ISSN: 14783967. DOI: 10.1088/1478-3975/11/1/016001. pmid: 24384886.

Buikema, W. J. and R. Haselkorn (1991). "Isolation and Complementation of Nitrogen Fixation Mutants of the Cyanobacterium *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 173.6, pp. 1879–1885. ISSN: 00219193. DOI: 10.1128/jb.173.6.1879-1885.1991. pmid: 1900504.

— (Feb. 2001). "Expression of the *Anabaena hetR* Gene from a Copper-Regulated Promoter Leads to Heterocyst Differentiation under Repressing Conditions." In: *Proc. Natl. Acad. Sci. U. S. A.* 98.5, pp. 2729–2734. ISSN: 00278424. DOI: 10.1073/pnas.051624898. pmid: 11226308.

Burnat, Mireia et al. (Dec. 2014). "Cell Envelope Components Influencing Filament Length in the Heterocyst-Forming Cyanobacterium *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 196.23, pp. 4026–4035. ISSN: 10985530. DOI: 10.1128/JB.02128-14. pmid: 25201945.

Callahan, S. M. and W. J. Buikema (May 2001). "The Role of *hetN* in Maintenance of the Heterocyst Pattern in *Anabaena* Sp. *PCC 7120*." In: *Mol. Microbiol.* 40.4, pp. 941–950. ISSN: 0950382X. DOI: 10.1046/j.1365-2958.2001.02437.x. pmid: 11401701.

Casanova-Ferrer, Pau et al. (Aug. 2022). "Terminal Heterocyst Differentiation in the Anabaena patA Mutant as a Result of Post-Transcriptional Modifications and Molecular Leakage." In: *PLoS Comput Biol* 18.8. Ed. by Tobias Bollenbach, e1010359. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1010359.

Chan, Yuki et al. (Sept. 2012). "Hypolithic Microbial Communities: Between a Rock and a Hard Place: Hypolithic Microbial Communities." In: *Environ. Microbiol.* 14.9, pp. 2272–2282. ISSN: 14622912. DOI: 10.1111/j.1462-2920.2012.02821.x.

Corrales-Guerrero, Laura et al. (June 2013). "Functional Dissection and Evidence for Intercellular Transfer of the Heterocyst-Differentiation *patS* Morphogen." In: *Mol. Microbiol.* 88.6, pp. 1093–1105. ISSN: 0950382X. DOI: 10.1111/mmi.12244. pmid: 23663167.

Corrales-Guerrero, Laura et al. (Aug. 2014a). "Relationships between the ABC-exporter *hetC* and Peptides That Regulate the Spatiotemporal Pattern of Heterocyst Distribution in *Anabaena*." In: *PLoS ONE* 9.8. Ed. by Dirk-Jan Scheffers, e104571. ISSN: 19326203. DOI: 10.1371/journal.pone.0104571. pmid: 25121608.

Corrales-Guerrero, Laura et al. (Oct. 2014b). "Subcellular Localization and Clues for the Function of the *hetN* Factor Influencing Heterocyst Distribution in *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 196.19, pp. 3452–3460. ISSN: 10985530. DOI: 10.1128/JB.01922-14. pmid: 25049089.

Corrales-Guerrero, Laura et al. (Apr. 2015). "Spatial Fluctuations in Expression of the Heterocyst Differentiation Regulatory Gene *hetR* in *Anabaena* Filaments." In: *PLoS Genet.* 11.4. Ed. by Lotte Søgaard-Andersen, e1005031. ISSN: 15537404. DOI: 10.1371/journal.pgen.1005031. pmid: 25830300.

Crampin, Edmund J et al. (1999). "Reaction and Diffusion on Growing Domains: Scenarios for Robust Pattern Formation." In: *Bull. Math. Biol.* 61.6, pp. 1093–1120. DOI: 10.1006/bulm.1999.0131.

De Koster, Chris G. and Aristid Lindenmayer (Dec. 1987). "Discrete and Continuous Models for Heterocyst Differentiation in Growing Filaments of Blue-Green Bacteria." In: *Acta Biotheor.* 36.4, pp. 249–273. ISSN: 1572-8358. DOI: 10.1007/BF02329786.

de Tezanos Pinto, Paula et al. (Jan. 2016). "Morphological Traits in Nitrogen Fixing Heterocytous Cyanobacteria: Possible Links between Morphology and Eco-Physiology." In: *Hydrobiologia* 764.1, pp. 271–281. ISSN: 0018-8158, 1573-5117. DOI: 10.1007/s10750-015-2516-6.

Di Patti, Francesca et al. (May 2018). "Robust Stochastic Turing Patterns in the Development of a One-Dimensional Cyanobacterial Organism." In: *PLoS Biol.* 16.5. Ed. by Jeff Gore, e2004877. ISSN: 15457885. DOI: 10.1371/journal.pbio.2004877. pmid: 29727442.

Du, Yaru et al. (2020). "Expression from DIF1-motif Promoters of *hetR* and *patS* Is Dependent on HetZ and Modulated by PatU3 during Heterocyst Differentiation." In: *PLoS ONE* 15.7, e0232383. DOI: 10.1371/journal.pone.0232383.

Du, Ye et al. (May 2012). "Identification of the *hetR* Recognition Sequence Upstream of *hetZ* in *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 194.9, pp. 2297–2306. ISSN: 00219193. DOI: 10.1128/JB.00119-12. pmid: 22389489.

Dvořák, Petr et al. (Apr. 2015). "Species Concepts and Speciation Factors in Cyanobacteria, with Connection to the Problems of Diversity and Classification." In: *Biodivers Conserv* 24.4, pp. 739–757. ISSN: 0960-3115, 1572-9710. DOI: 10.1007/s10531-015-0888-6.

Ehira, Shigeki and Masayuki Ohmori (Dec. 2006). "NrrA Directly Regulates Expression of *hetR* during Heterocyst Differentiation in the Cyanobacterium *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 188.24, pp. 8520–8525. ISSN: 00219193. DOI: 10.1128/JB.01314-06. pmid: 17041048.

Ehira, Shigeki et al. (Feb. 2018). "Spatial Separation of Photosynthesis and Ethanol Production by Cell Type-Specific Metabolic Engineering of Filamentous Cyanobacteria." In: *Appl. Microbiol. Biotechnol.* 102.3, pp. 1523–1531. ISSN: 14320614. DOI: 10.1007/s00253-017-8620-y. pmid: 29143082.

Elhai, Jeff and Ivan Khudyakov (2018). "Ancient Association of Cyanobacterial Multicellularity with the Regulator *hetR* and an RGSGR Pentapeptide-Containing Protein (PatX)." In: *Mol. Microbiol.* 110.6, pp. 931–954. ISSN: 13652958. DOI: 10.1111/mmi.14003. pmid: 29885033.

Elowitz, Michael B. and Stanislas Leibler (Jan. 2000). "A Synthetic Oscillatory Network of Transcriptional Regulators." In: *Nature* 403.6767, pp. 335–338. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/35002125.

Erisman, Jan Willem et al. (Oct. 2008). "How a Century of Ammonia Synthesis Changed the World." In: *Nature Geoscience* 1.10, pp. 636–639. ISSN: 1752-0908. DOI: 10.1038/ngeo325.

Ernst, A. et al. (Oct. 1992). "Synthesis of Nitrogenase in Mutants of the Cyanobacterium *Anabaena* Sp. Strain *PCC 7120* Affected in Heterocyst Development or Metabolism." In: *J. Bacteriol.* 174.19, pp. 6025–6032. ISSN: 00219193. DOI: 10.1128/jb.174.19.6025-6032.1992. pmid: 1328150.

Evenson, Robert E and Douglas Gollin (2003). "Assessing the Impact of the Green Revolution, 1960 to 2000." In: *science*. DOI: 10.1126/science.1078710.

Fan, Chuchuan et al. (2006). "GS3, a Major QTL for Grain Length and Weight and Minor QTL for Grain Width and Thickness in Rice, Encodes a Putative Transmembrane Protein." In: *Theor. Appl. Genet.* 112.6, pp. 1164–1171. DOI: 10.1007/s00122-006-0218-1.

Fay, P. et al. (Nov. 1968). "Is the Heterocyst the Site of Nitrogen Fixation in Blue-Green Algae?" In: *Nature* 220.5169, pp. 810–812. ISSN: 00280836. DOI: 10.1038/220810b0. pmid: 4880705.

Feldmann, Erik A. et al. (Nov. 2011). "Evidence for Direct Binding between *hetR* from *Anabaena* Sp. *PCC 7120* and *patS*-5." In: *Biochemistry* 50.43, pp. 9212–9224. ISSN: 00062960. DOI: 10.1021/bi201226e. pmid: 21942265.

Feldmann, Erik A. et al. (Mar. 2012). "Differential Binding between *patS* C-terminal Peptide Fragments and *hetR* from *Anabaena* Sp. *PCC 7120*." In: *Biochemistry* 51.12, pp. 2436–2442. ISSN: 00062960. DOI: 10.1021/bi300228n. pmid: 22397695.

Flores, Enrique and Antonia Herrero (Jan. 2010). "Compartmentalized Function through Cell Differentiation in Filamentous Cyanobacteria." In: *Nat. Rev. Microbiol.* 8.1, pp. 39–50. ISSN: 17401526. DOI: 10.1038/nrmicro2242. pmid: 19966815.

Fogg, G. E. (July 1949). "Growth and Heterocyst Production in *Anabaena* Cylindrica Lemm: II. In Relation to Carbon and Nitrogen Metabolism." In: *Ann. Bot.* 13.3, pp. 241–259. ISSN: 03057364. DOI: 10.1093/oxfordjournals.aob.a083217.

Fowler, David et al. (July 2013). "The Global Nitrogen Cycle in the Twenty-First Century." In: *Phil. Trans. R. Soc. B* 368.1621, p. 20130164. ISSN: 0962-8436, 1471-2970. DOI: 10.1098/rstb.2013.0164.

Frías, José E. et al. (Nov. 1994). "Requirement of the Regulatory Protein NtcA for the Expression of Nitrogen Assimilation and Heterocyst Development Genes in the Cyanobacterium *Anabaena* Sp. *PCC 7120*." In: *Mol. Microbiol.* 14.4, pp. 823–832. ISSN: 13652958. DOI: 10.1111/j.1365-2958.1994.tb01318.x. pmid: 7534371.

Frigola, David et al. (Feb. 2012). "Asymmetric Stochastic Switching Driven by Intrinsic Molecular Noise." In: *PLoS ONE* 7.2. Ed. by Matjaz Perc, e31407. ISSN: 19326203. DOI: 10.1371/journal.pone.0031407. pmid: 22363638.

Gardiner, C. W. et al. (1983). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Vol. 42. Springer Berlin Heidelberg. ISBN: 978-3-540-61634-4. DOI: 10.2307/2531274.

Gardner, Timothy S. et al. (Jan. 2000). "Construction of a Genetic Toggle Switch in Escherichia Coli." In: *Nature* 403.6767, pp. 339–342. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/35002131.

Gerdtzen, Ziomara P. et al. (June 2009). "Modeling Heterocyst Pattern Formation in Cyanobacteria." In: *BMC Bioinformatics* 10.SUPPL. 6, S16. ISSN: 14712105. DOI: 10.1186/1471-2105-10-S6-S16. pmid: 19534741.

Gierer, A. and H. Meinhardt (Dec. 1972). "A Theory of Biological Pattern Formation." In: *Kybernetik* 12.1, pp. 30–39. ISSN: 03401200. DOI: 10.1007/BF00289234. pmid: 4663624.

Gillespie, Daniel T. (1976). "A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions." In: *J. Comput. Phys.* 22.4, pp. 403–434. ISSN: 0021-9991. DOI: 10.1016/0021-9991(76)90041-3.

— (July 2000). "Chemical Langevin Equation." In: *J. Chem. Phys.* 113.1, pp. 297–306. ISSN: 00219606. DOI: 10.1063/1.481811.

Golden, James W. and Ho Sung Yoon (Dec. 2003). "Heterocyst Development in *Anabaena*." In: *Curr. Opin. Microbiol.* 6.6, pp. 557–563. ISSN: 13695274. DOI: 10.1016/j.mib.2003.10.004. pmid: 14662350.

Griffiths, Anthony J. F. et al. (2020). *Introduction to Genetic Analysis*. Twelfth edition. New York, NY: W.H. Freeman & Company/Macmillan Learning. ISBN: 978-1-319-11478-7.

Grover, James P. et al. (2019). "Dynamics of Nitrogen-Fixing Cyanobacteria with Heterocysts: A Stoichiometric Model." In: *Mar. Freshw. Res.* 71.5, pp. 644–658. ISSN: 13231650. DOI: 10.1071/MF18361.

Grzebisz, Witold et al. (Oct. 2022). "Soil Fertility Clock—Crop Rotation as a Paradigm in Nitrogen Fertilizer Productivity Control." In: *Plants* 11.21, p. 2841. ISSN: 2223-7747. DOI: 10.3390/plants11212841.

Harish and Kunal Seth (June 2020). "Molecular Circuit of Heterocyst Differentiation in Cyanobacteria." In: *J Basic Microbiol*. ISSN: 0233111X. DOI: 10.1002/jobm.202000266.

Hedden, Peter (2003). "The Genes of the Green Revolution." In: *TRENDS Genet.* 19.1, pp. 5–9. DOI: 10.1016/s0168-9525(02)00009-4.

Hense, Inga and Aike Beckmann (2006). "Towards a Model of Cyanobacteria Life Cycle—Effects of Growing and Resting Stages on Bloom Formation of N2-fixing Species." In: *Ecol. Model.* 195.3, pp. 205–218. ISSN: 0304-3800. DOI: 10.1016/j.ecolmodel.2005.11.018.

Herrero, Antonia and Enrique Flores (Jan. 2019). "Genetic Responses to Carbon and Nitrogen Availability in *Anabaena*." In: *Environ. Microbiol.* 21.1, pp. 1–17. ISSN: 14622920. DOI: 10.1111/1462-2920.14370. pmid: 30066380.

Herrero, Antonia et al. (Oct. 2004). "Cellular Differentiation and the NtcA Transcription Factor in Filamentous Cyanobacteria." In: *FEMS Microbiol. Rev.* 28.4, pp. 469–487. ISSN: 01686445. DOI: 10.1016/j.femsre.2004.04.003. pmid: 15374662.

Herrero, Antonia et al. (2013). "Gene Expression during Heterocyst Differentiation." In: *Genomics of Cyanobacteria*. Ed. by Franck Chauvat and Corinne Cassier-Chauvat. Vol. 65. Advances in Botanical Research. Academic Press, pp. 281–329. DOI: 10.1016/B978-0-12-394313-2.00008-1.

Herrero, Antonia et al. (Nov. 2016). "The Multicellular Nature of Filamentous Heterocyst-Forming Cyanobacteria." In: *FEMS Microbiol. Rev.* 40.6. Ed. by William Margolin, pp. 831–854. ISSN: 15746976. DOI: 10.1093/femsre/fuw029. pmid: 28204529.

Higa, Kelly C. and Sean M. Callahan (June 2010). "Ectopic Expression of *hetP* Can Partially Bypass the Need for *hetR* in Heterocyst Differentiation by *Anabaena* Sp. Strain *PCC 7120*." In: *Mol. Microbiol.* 77.3, pp. 562–574. ISSN: 13652958. DOI: 10.1111/j.1365-2958.2010.07257.x. pmid: 20545862.

Higa, Kelly C. et al. (Feb. 2012). "The RGSGR Amino Acid Motif of the Intercellular Signalling Protein, *hetN*, Is Required for Patterning of Heterocysts in *Anabaena* Sp. Strain *PCC 7120*." In: *Mol. Microbiol.* 83.4, pp. 682–693. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2011.07949.x. pmid: 22220907.

Hill, Archibald Vivian (Oct. 1913). "The Combinations of Haemoglobin with Oxygen and with Carbon Monoxide. I." In: *Biochem. J.* 7.5, pp. 471–480. ISSN: 0306-3283. DOI: 10.1042/bj0070471.

Hill, D. J. (1975). "The Pattern of Development of Anabaena in the Azolla-Anabaena Symbiosis." In: *Planta* 122.2, pp. 179–184. ISSN: 0032-0935, 1432-2048. DOI: 10.1007/BF00388657.

Hou, Shengwei et al. (Jan. 2015). "The *hetR*-Binding Site That Activates Expression of *patA* in Vegetative Cells Is Required for Normal Heterocyst Patterning in *Anabaena* Sp. *PCC 7120*." In: *Sci. Bull.* 60.2, pp. 192–201. ISSN: 20959281. DOI: 10.1007/s11434-014-0724-5.

Hu, Hai Xi et al. (2015). "Structural Insights into *hetR'patS* Interaction Involved in Cyanobacterial Pattern Formation." In: *Sci. Rep.* 5, p. 16470. ISSN: 20452322. DOI: 10.1038/srep16470. pmid: 26576507.

Huang, Xianzhong et al. (2009). "Natural Variation at the DEP1 Locus Enhances Grain Yield in Rice." In: *Nat. Genet.* 41.4, pp. 494–497. DOI: 10.1038/ng.352.

Huang, Xu et al. (Apr. 2004). "*hetR* Homodimer Is a DNA-binding Protein Required for Heterocyst Differentiation, and the DNA-binding Activity Is Inhibited by *patS*." In: *Proc. Natl. Acad. Sci. U. S. A.* 101.14, pp. 4848–4853. ISSN: 00278424. DOI: 10.1073/pnas.0400429101. pmid: 15051891.

Ideker, Trey et al. (2001). "A NEW APPROACH TO DECODING LIFE: Systems Biology." In: *Annu. Rev. Genomics Hum. Genet.* 2.1, pp. 343–372. DOI: 10.1146/annurev.genom.2.1.343.

Ikeda-Kawakatsu, Kyoko et al. (2009). "Expression Level of ABERRANT PANICLE ORGANIZATION1 Determines Rice Inflorescence Form through Control of Cell Proliferation in the Meristem." In: *Plant Physiol.* 150.2, pp. 736–747. DOI: 10.1104/pp.109.136739.

Ingalls, Brian P. (2013). *Mathematical Modeling in Systems Biology: An Introduction*. Cambridge, Massachusetts: MIT Press. ISBN: 978-0-262-01888-3.

Ishihara, Jun-ichi et al. (Apr. 2015). "Mathematical Study of Pattern Formation Accompanied by Heterocyst Differentiation in Multicellular Cyanobacterium." In: *J. Theor. Biol.* 371, pp. 9–23. ISSN: 10958541. DOI: 10.1016/j.jtbi.2015.01.034. pmid: 25665721.

Ishikawa, Shinji et al. (2005). "Suppression of Tiller Bud Activity in Tillering Dwarf Mutants of Rice." In: *Plant Cell Physiol.* 46.1, pp. 79–86. DOI: 10.1093/pcp/pci022.

Itoh, Hironori et al. (2002). "The Gibberellin Signaling Pathway Is Regulated by the Appearance and Disappearance of SLENDER RICE1 in Nuclei." In: *Plant Cell* 14.1, pp. 57–70. DOI: 10.1105/tpc.010319.

Jiang, Liang et al. (2013). "DWARF 53 Acts as a Repressor of Strigolactone Signalling in Rice." In: *Nature* 504.7480, pp. 401–405. DOI: 10.1038/nature12870.

Jiao, Yongqing et al. (2010). "Regulation of OsSPL14 by OsmiR156 Defines Ideal Plant Architecture in Rice." In: *Nat. Genet.* 42.6, pp. 541–544. DOI: 10.1038/ng.591.

Kahru, Mati et al. (Dec. 2018). "Unexplained Interannual Oscillations of Cyanobacterial Blooms in the Baltic Sea." In: *Sci Rep* 8.1, p. 6365. ISSN: 2045-2322. DOI: 10.1038/s41598-018-24829-7.

Kaup, Maya et al. (Dec. 2021). "On the Move: Sloths and Their Epibionts as Model Mobile Ecosystems." In: *Biol Rev* 96.6, pp. 2638–2660. ISSN: 1464-7931, 1469-185X. DOI: 10.1111/brv.12773.

Khalil, Ahmad S. and James J. Collins (May 2010). "Synthetic Biology: Applications Come of Age." In: *Nat Rev Genet* 11.5, pp. 367–379. ISSN: 1471-0056, 1471-0064. DOI: 10.1038/nrg2775.

Khudyakov, Ivan et al. (2020). "Inactivation of Three Rg(S/t)Gr Pentapeptide-Containing Negative Regulators of *hetR* Results in Lethal Differenti-

ation of *Anabaena PCC 7120*." In: *Life* 10.12, pp. 1–16. ISSN: 20751729. DOI: 10.3390/life10120326.

Khudyakov, Ivan Y. and James W. Golden (Nov. 2004). "Different Functions of *hetR*, a Master Regulator of Heterocyst Differentiation in *Anabaena* Sp. *PCC 7120*, Can Be Separated by Mutation." In: *Proc. Natl. Acad. Sci. U. S. A.* 101.45, pp. 16040–16045. ISSN: 00278424. DOI: 10.1073/pnas.0405572101. pmid: 15520378.

Khush, Gurdev S (1999). "Green Revolution: Preparing for the 21st Century." In: *Genome* 42.4, pp. 646–655. DOI: 10.1139/g99-044.

Kim, Youngchang et al. (June 2011). "Structure of Transcription Factor *hetR* Required for Heterocyst Differentiation in Cyanobacteria." In: *Proc. Natl. Acad. Sci. U. S. A.* 108.25, pp. 10109–10114. ISSN: 00278424. DOI: 10.1073/pnas.1106840108. pmid: 21628585.

Kim, Youngchang et al. (2013). "Structures of Complexes Comprised of *Fischerella* Transcription Factor *hetR* with *Anabaena* DNA Targets." In: *Proc. Natl. Acad. Sci. U. S. A.* 110.19, E1716–E1723. ISSN: 00278424. DOI: 10.1073/pnas.1305971110. pmid: 23610410.

Kirkpatrick, S. et al. (May 1983). "Optimization by Simulated Annealing." In: *Science* 220.4598, pp. 671–680. ISSN: 00368075. DOI: 10.1126/science.220.4598.671. pmid: 17813860.

Kitano, Hiroaki (2002). "Systems Biology: A Brief Overview." In: *Science* 295.5560, pp. 1662–1664. DOI: 10.1126/science.1069492.

Kollmen, Jonas and Dorina Strieth (Jan. 2022). "The Beneficial Effects of Cyanobacterial Co-Culture on Plant Growth." In: *Life* 12.2, p. 223. ISSN: 2075-1729. DOI: 10.3390/life12020223.

Komárek, Jiří (July 2016). "A Polyphasic Approach for the Taxonomy of Cyanobacteria: Principles and Applications." In: *European Journal of Phycology* 51.3, pp. 346–353. ISSN: 0967-0262, 1469-4433. DOI: 10.1080/09670262.2016.1163738.

Komatsu, Keishi et al. (2003). "LAX and SPA: Major Regulators of Shoot Branching in Rice." In: *Proc. Natl. Acad. Sci.* 100.20, pp. 11765–11770. DOI: 10.1073/pnas.1932414100.

Kopp, Robert E. et al. (Aug. 2005). "The Paleoproterozoic Snowball Earth: A Climate Disaster Triggered by the Evolution of Oxygenic Photosynthesis." In: *Proc. Natl. Acad. Sci. U.S.A.* 102.32, pp. 11131–11136. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0504878102.

Laarhoven, Peter J. M. and Emile H. L. Aarts (1987). *Simulated Annealing: Theory and Applications*. Dordrecht: Springer Netherlands. ISBN: 978-94-015-7744-1.

Lacalli, Thurston C. (2022). "Patterning, from Conifers to Consciousness: Turing's Theory and Order from Fluctuations." In: *Front. Cell Dev. Biol.* 10, p. 871950. DOI: 10.3389/fcell.2022.871950.

Lee, Sang-Moo and Choong-Min Ryu (Feb. 2021). "Algae as New Kids in the Beneficial Plant Microbiome." In: *Front. Plant Sci.* 12, p. 599742. ISSN: 1664-462X. DOI: 10.3389/fpls.2021.599742.

Levitt, Michael and Arieh Warshel (1975). "Computer Simulation of Protein Folding." In: *Nature* 253.5494, pp. 694–698. DOI: 10.1038/253694a0.

Li, Bin et al. (Apr. 2002). "Expression of *hetN* during Heterocyst Differentiation and Its Inhibition of *hetR* Up-Regulation in the Cyanobacterium *Anabaena* Sp. *PCC 7120*." In: *FEBS Lett.* 517.1-3, pp. 87–91. ISSN: 00145793. DOI: 10.1016/S0014-5793(02)02582-6. pmid: 12062415.

Li, Xueyong et al. (2003). "Control of Tillering in Rice." In: *Nature* 422.6932, pp. 618–621. DOI: 10.1038/nature01518.

Liang, J. et al. (June 1992). "The *patA* Gene Product, Which Contains a Region Similar to CheY of *Escherichia coli*, Controls Heterocyst Pattern Formation in the Cyanobacterium *Anabaena 7120*." In: *Proc. Natl. Acad. Sci. U. S. A.* 89.12, pp. 5655–5659. ISSN: 00278424. DOI: 10.1073/pnas.89.12.5655. pmid: 1608976.

Liao, Zhigang et al. (2019). "SLR1 Inhibits MOC1 Degradation to Coordinate Tiller Number and Plant Height in Rice." In: *Nat. Commun.* 10.1, pp. 1–9. DOI: 10.1038/s41467-019-10667-2.

Liu, Duan and James W. Golden (Dec. 2002). "*hetL* Overexpression Stimulates Heterocyst Formation in *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 184.24, pp. 6873–6881. ISSN: 00219193. DOI: 10.1128/JB.184.24.6873-6881.2002. pmid: 12446638.

Luque, Ignacio et al. (June 1994). "Molecular Mechanism for the Operation of Nitrogen Control in Cyanobacteria." In: *EMBO J.* 13.12, pp. 2862–2869. ISSN: 02614189. DOI: 10.1002/j.1460-2075.1994.tb06580.x. pmid: 8026471.

Makarova, Kira S. et al. (June 2006). "Cyanobacterial Response Regulator *patA* Contains a Conserved N-terminal Domain (PATAN) with an Alpha-Helical Insertion." In: *Bioinformatics* 22.11, pp. 1297–1301. ISSN: 13674803. DOI: 10.1093/bioinformatics/btl096. pmid: 16543275.

Malatinszky, David et al. (Jan. 2017). "A Comprehensively Curated Genome-Scale Two-Cell Model for the Heterocystous Cyanobacterium *Anabaena* Sp. *PCC 7120*." In: *Plant Physiol.* 173.1, pp. 509–523. ISSN: 15322548. DOI: 10.1104/pp.16.01487. pmid: 27899536.

Marciniak-Czochra, Anna et al. (2017). "Instability of Turing Patterns in Reaction-Diffusion-ODE Systems." In: *J. Math. Biol.* 74.3, pp. 583–618. DOI: 10.1007/s00285-016-1035-z.

Marcon, Luciano et al. (2016). "High-Throughput Mathematical Analysis Identifies Turing Networks for Patterning with Equally Diffusing Signals." In: *Elife* 5, e14022. DOI: 10.7554/eLife.14022.

Massey Jr., Frank J. (1951). "The Kolmogorov-Smirnov Test for Goodness of Fit." In: *J. Am. Stat. Assoc.* 46.253, pp. 68–78. DOI: 10.2307/2280095.

Meeks, John C. and Jeff Elhai (Mar. 2002). "Regulation of Cellular Differentiation in Filamentous Cyanobacteria in Free-Living and Plant-Associated Symbiotic Growth States." In: *Microbiol. Mol. Biol. Rev.* 66.1, pp. 94–121. ISSN: 1092-2172. DOI: 10.1128/mmbr.66.1.94-121.2002. pmid: 11875129.

Meinhardt, Hans (2008). "Models of Biological Pattern Formation: From Elementary Steps to the Organization of Embryonic Axes." In: *Curr. Top. Dev. Biol.* 81, pp. 1–63. ISSN: 00702153. DOI: 10.1016/S0070-2153(07)81001-5. pmid: 18023723.

Metropolis, Nicholas et al. (June 1953). "Equation of State Calculations by Fast Computing Machines." In: *The Journal of Chemical Physics* 21.6, pp. 1087–1092. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1699114.

Michaelis, Leonor, Maud L Menten, et al. (1913). "Die Kinetik Der Invertinwirkung." In: *Biochem. z* 49.333-369, p. 352.

Middleton, Alistair M et al. (2012). "Mathematical Modeling Elucidates the Role of Transcriptional Feedback in Gibberellin Signaling." In: *Proc. Natl. Acad. Sci.* 109.19, pp. 7571–7576. DOI: 10.1073/pnas.1113666109.

Milo, Ron and Rob Phillips (2016). *Cell Biology by the Numbers*. New York, NY: Garland Science, Taylor & Francis Group. ISBN: 978-0-8153-4537-4.

Miura, Kotaro et al. (2010). "OsSPL14 Promotes Panicle Branching and Higher Grain Productivity in Rice." In: *Nat. Genet.* 42.6, pp. 545–549. DOI: 10.1038/ng.592.

Möllers, K. Benedikt et al. (2014). "Cyanobacterial Biomass as Carbohydrate and Nutrient Feedstock for Bioethanol Production by Yeast Fermentation." In: *Biotechnol. Biofuels* 7.1, pp. 1–11. ISSN: 17546834. DOI: 10.1186/1754-6834-7-64. pmid: 24739806.

Monna, Lisa et al. (2002). "Positional Cloning of Rice Semidwarfing Gene, Sd-1: Rice "Green Revolution Gene" Encodes a Mutant Enzyme Involved in Gibberellin Synthesis." In: *DNA Res.* 9.1, pp. 11–17. DOI: 10.1093/dnares/9.1.11.

Muñoz-García, Javier and Saúl Ares (May 2016). "Formation and Maintenance of Nitrogen-Fixing Cell Patterns in Filamentous Cyanobacteria." In: *Proc. Natl. Acad. Sci. U. S. A.* 113.22, pp. 6218–6223. ISSN: 10916490. DOI: 10.1073/pnas.1524383113. pmid: 27162328.

Murase, Kohji et al. (2008). "Gibberellin-Induced DELLA Recognition by the Gibberellin Receptor GID1." In: *Nature* 456.7221, pp. 459–463. DOI: 10.1038/nature07519.

Muro-Pastor, Alicia M. (May 2014). "The Heterocyst-Specific nsiR1 Small RNA Is an Early Marker of Cell Differentiation in Cyanobacterial Filaments." In: *mBio* 5.3. Ed. by Stephen J. Giovannoni, pp. 1–6. ISSN: 21507511. DOI: 10.1128/mBio.01079-14. pmid: 24825011.

Muro-Pastor, Alicia M. and Wolfgang R. Hess (2012). "Heterocyst Differentiation: From Single Mutants to Global Approaches." In: *Trends Microbiol.* 20.11, pp. 548–557. ISSN: 0966842X. DOI: 10.1016/j.tim.2012.07.005. pmid: 22898147.

Muro-Pastor, Alicia M. et al. (May 2002). "Mutual Dependence of the Expression of the Cell Differentiation Regulatory Protein *hetR* and the Global Nitrogen Regulator NtcA during Heterocyst Development." In: *Mol. Microbiol.* 44.5, pp. 1377–1385. ISSN: 0950382X. DOI: 10.1046/j.1365-2958.2002.02970.x. pmid: 12068814.

Muro-Pastor, M. Isabel et al. (Oct. 2001). "Cyanobacteria Perceive Nitrogen Status by Sensing Intracellular 2-Oxoglutarate Levels." In: *J. Biol. Chem.* 276.41, pp. 38320–38328. ISSN: 00219258. DOI: 10.1074/jbc.m105297200. pmid: 11479309.

Murray, James Dickson (2003). *Mathematical Biology, II: Spatial Models and Biomedical Applications*. Springer.

Nakamasu, Akiko et al. (2009). "Interactions between Zebrafish Pigment Cells Responsible for the Generation of Turing Patterns." In: *Proc. Natl. Acad. Sci. U. S. A.* 106.21, pp. 8429–8434. DOI: 10.1073/pnas.0808622106.

Newman, M. E. J. and G. T. Barkema (1999). *Monte Carlo Methods in Statistical Physics*. Oxford : New York: Clarendon Press ; Oxford University Press. ISBN: 978-0-19-851796-2.

Oikawa, Tetsuo and Junko Kyozuka (2009). "Two-Step Regulation of LAX PANICLE1 Protein Accumulation in Axillary Meristem Formation in Rice." In: *Plant Cell* 21.4, pp. 1095–1108. DOI: 10.1105/tpc.108.065425.

Olmedo-Verd, Elvira et al. (Mar. 2005). "*hetR*-Dependent and Independent Expression of Heterocyst-Related Genes in an *Anabaena* Strain Overproducing the NtcA Transcription Factor." In: *J. Bacteriol.* 187.6, pp. 1985–1991. ISSN: 00219193. DOI: 10.1128/JB.187.6.1985-1991.2005. pmid: 15743946.

Orozco, Christine C. et al. (Mar. 2006). "Epistasis Analysis of Four Genes from *Anabaena* Sp. Strain *PCC 7120* Suggests a Connection between *patA* and *patS* in Heterocyst Pattern Formation." In: *J. Bacteriol.* 188.5, pp. 1808–1816. ISSN: 00219193. DOI: 10.1128/JB.188.5.1808-1816.2006. pmid: 16484191.

Pereira, Ana L. (May 2018). "The Unique Symbiotic System between a Fern and a Cyanobacterium, Azolla-Anabaena Azollae: Their Potential as Biofertilizer, Feed, and Remediation." In: *Symbiosis*. Ed. by Everlon Cid Rigobelo. InTech. ISBN: 978-1-78923-224-0. DOI: 10.5772/intechopen.70466.

Peters, G A and J C Meeks (June 1989). "The Azolla-Anabaena Symbiosis: Basic Biology." In: *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* 40.1,

pp. 193–210. ISSN: 1040-2519. DOI: 10.1146/annurev.pp.40.060189.001205.

Pingali, Prabhu L (2012). "Green Revolution: Impacts, Limits, and the Path Ahead." In: *Proc. Natl. Acad. Sci.* 109.31, pp. 12302–12308. DOI: 10.1073/pnas.0912953109.

Pinzon, Neissa M. and Lu Kwang Ju (Dec. 2006). "Modeling Culture Profiles of the Heterocystous N2-fixing Cyanobacterium *Anabaena flos-aquae*." In: *Biotechnol. Prog.* 22.6, pp. 1532–1540. ISSN: 87567938. DOI: 10.1021/bp060163c. pmid: 17137298.

Rajagopalan, Ramya and Sean M. Callahan (Feb. 2010). "Temporal and Spatial Regulation of the Four Transcription Start Sites of *hetR* from *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 192.4, pp. 1088–1096. ISSN: 00219193. DOI: 10.1128/JB.01297-09. pmid: 20008074.

Raspopovic, Jelena et al. (2014). "Digit Patterning Is Controlled by a Bmp-Sox9-Wnt Turing Network Modulated by Morphogen Gradients." In: *Science* 345.6196, pp. 566–570. DOI: 10.1126/science.1252960.

Rieu, Ivo et al. (2008). "The Gibberellin Biosynthetic Genes AtGA20ox1 and AtGA20ox2 Act, Partially Redundantly, to Promote Growth and Development throughout the Arabidopsis Life Cycle." In: *Plant J.* 53.3, pp. 488–504. DOI: 10.1111/j.1365-313X.2007.03356.x.

Rippka, Rosmarie et al. (Mar. 1979). "Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria." In: *Microbiology* 111.1, pp. 1–61. ISSN: 1350-0872, 1465-2080. DOI: 10.1099/00221287-111-1-1.

Risser, Douglas D. and Sean M. Callahan (Mar. 2007). "Mutagenesis of *hetR* Reveals Amino Acids Necessary for *hetR* Function in the Heterocystous Cyanobacterium *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 189.6, pp. 2460–2467. ISSN: 00219193. DOI: 10.1128/JB.01241-06. pmid: 17220221.

— (Dec. 2008). "*hetF* and *patA* Control Levels of *hetR* in *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 190.23, pp. 7645–7654. ISSN: 00219193. DOI: 10.1128/JB.01110-08. pmid: 18835986.

— (Nov. 2009). "Genetic and Cytological Evidence That Heterocyst Patterning Is Regulated by Inhibitor Gradients That Promote Activator Decay." In: *Proc. Natl. Acad. Sci. U. S. A.* 106.47, pp. 19884–19888. ISSN: 00278424. DOI: 10.1073/pnas.0909152106. pmid: 19897721.

Risser, Douglas D. et al. (2012). "Biased Inheritance of the Protein PatN Frees Vegetative Cells to Initiate Patterned Heterocyst Differentiation." In: *Proc. Natl. Acad. Sci. U. S. A.* 109.38, pp. 15342–15347. ISSN: 00278424. DOI: 10.1073/pnas.1207530109. pmid: 22949631.

Rivers, Orion S. et al. (Dec. 2014). "Mutation of *sepJ* Reduces the Intercellular Signal Range of a *hetN*-Dependent Paracrine Signal, but Not of a *patS*-Dependent Signal, in the Filamentous Cyanobacterium *An-*

abaena Sp. Strain *PCC 7120*." In: *Mol. Microbiol.* 94.6, pp. 1260–1271. ISSN: 13652958. DOI: 10.1111/mmi.12836. pmid: 25336355.

Rivers, Orion S. et al. (May 2018). "Phenotypic Assessment Suggests Multiple Start Codons for *hetN*, an Inhibitor of Heterocyst Differentiation, in *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 200.16. Ed. by Yves V. Brun, pp. 1–14. ISSN: 10985530. DOI: 10.1128/JB.00220-18. pmid: 29784882.

Roumezi, Baptiste et al. (Jan. 2020). "The Pkn22 Kinase of *Nostoc PCC 7120* Is Required for Cell Differentiation via the Phosphorylation of *hetR* on a Residue Highly Conserved in Genomes of Heterocyst-Forming Cyanobacteria." In: *Front. Microbiol.* 10.January, pp. 1–18. ISSN: 1664302X. DOI: 10.3389/fmicb.2019.03140.

Ryu, Min-Hyung et al. (Feb. 2020). "Control of Nitrogen Fixation in Bacteria That Associate with Cereals." In: *Nat Microbiol* 5.2, pp. 314–330. ISSN: 2058-5276. DOI: 10.1038/s41564-019-0631-2.

Sasaki, A et al. (2002). "A Mutant Gibberellin-Synthesis Gene in Rice." In: *Nature* 416.6882, pp. 701–702. DOI: 10.1038/416701a.

Saunders, Marissa G. and Gregory A. Voth (May 2013). "Coarse-Graining Methods for Computational Biology." In: *Annu. Rev. Biophys.* 42.1, pp. 73–93. ISSN: 1936-122X, 1936-1238. DOI: 10.1146/annurev-biophys-083012-130348.

El-Shehawy, Rehab M. and Diethelm Kleiner (2003). "The Mystique of Irreversibility in Cyanobacterial Heterocyst Formation: Parallels to Differentiation and Senescence in Eukaryotic Cells." In: *Physiol. Plant.* 119.1, pp. 49–55. ISSN: 00319317. DOI: 10.1034/j.1399-3054.2003.00096.x.

Shomura, Ayahiko et al. (2008). "Deletion in a Gene Associated with Grain Size Increased Yields during Rice Domestication." In: *Nat. Genet.* 40.8, pp. 1023–1028. DOI: 10.1038/ng.169.

Smith, Stephen and Neil Dalchau (2018). "Beyond Activator-Inhibitor Networks: The Generalised Turing Mechanism." In: *ArXiv Prepr. ArXiv180307886*. arXiv: 1803.07886.

Song, Xian-Jun et al. (2007). "A QTL for Rice Grain Width and Weight Encodes a Previously Unknown RING-type E3 Ubiquitin Ligase." In: *Nat. Genet.* 39.5, pp. 623–630. DOI: 10.1038/ng2014.

Spielmeyer, Wolfgang et al. (2002). "Semidwarf (Sd-1),"Green Revolution" Rice, Contains a Defective Gibberellin 20-Oxidase Gene." In: *Proc. Natl. Acad. Sci.* 99.13, pp. 9043–9048. DOI: 10.1073/pnas.132266399.

Strogatz, Steven H. (2015). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Second edition. Boulder, CO: Westview Press, a member of the Perseus Books Group. ISBN: 978-0-8133-4910-7.

Suutari, Milla et al. (Dec. 2010). "Molecular Evidence for a Diverse Green Algal Community Growing in the Hair of Sloths and a Specific Association with Trichophilus Welckeri(Chlorophyta, Ulvophyceae)." In: *BMC Evol Biol* 10.1, p. 86. ISSN: 1471-2148. DOI: 10.1186/1471-2148-10-86.

Takeda, Taito et al. (2003). "The OsTB1 Gene Negatively Regulates Lateral Branching in Rice." In: *Plant J.* 33.3, pp. 513–520. DOI: 10.1046/j.1365-313x.2003.01648.x..

Tanigawa, Ryohei et al. (2002). "Transcriptional Activation of NtcA-dependent Promoters of *Synechococcus* Sp. *PCC* 7942 by 2-Oxoglutarate in Vitro." In: *Proc. Natl. Acad. Sci. U. S. A.* 99.7, pp. 4251–4255. ISSN: 00278424. DOI: 10.1073/pnas.072587199. pmid: 11917135.

Torres-Sánchez, Alejandro et al. (Mar. 2015). "An Integrative Approach for Modeling and Simulation of Heterocyst Pattern Formation in Cyanobacteria Filaments." In: *PLoS Comput. Biol.* 11.3. Ed. by Stanislav Shvartsman, e1004129. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1004129. pmid: 25816286.

Triveni, Sodimalla et al. (June 2015). "Evaluating the Promise of *Trichoderma* and *Anabaena* Based Biofilms as Multifunctional Agents in *Macrophomina Phaseolina*- Infected Cotton Crop." In: *Biocontrol Science and Technology* 25.6, pp. 656–670. ISSN: 0958-3157, 1360-0478. DOI: 10.1080/09583157.2015.1006171.

Turing, Alan Mathison (Aug. 1952). "The Chemical Basis of Morphogenesis." In: *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 237.641, pp. 37–72. ISSN: 2054-0280. DOI: 10.1098/rstb.1952.0012.

Valladares, Ana et al. (Sept. 2008). "Transcription Activation by NtcA and 2-Oxoglutarate of Three Genes Involved in Heterocyst Differentiation in the Cyanobacterium *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 190.18, pp. 6126–6133. ISSN: 00219193. DOI: 10.1128/JB.00787-08. pmid: 18658268.

— (Feb. 2016). "The Heterocyst Differentiation Transcriptional Regulator *hetR* of the Filamentous Cyanobacterium *Anabaena* Forms Tetramers and Can Be Regulated by Phosphorylation." In: *Mol. Microbiol.* 99.4, pp. 808–819. ISSN: 13652958. DOI: 10.1111/mmi.13268. pmid: 26552991.

Valladares, Ana et al. (May 2020). "Interactions of *patA* with the Divisome during Heterocyst Differentiation in *Anabaena*." In: *mSphere* 5.3. Ed. by Grant R. Bowman, pp. 1–16. ISSN: 2379-5042. DOI: 10.1128/msphere.00188-20. pmid: 32434840.

Vázquez-Bermúdez, Maria Félix et al. (Apr. 2003). "Carbon Supply and 2-Oxoglutarate Effects on Expression of Nitrate Reductase and Nitrogen-Regulated Genes in *Synechococcus* Sp. Strain *PCC 7942*." In: *FEMS Microbiol. Lett.* 221.2, pp. 155–159. ISSN: 03781097. DOI: 10.1016/S0378-1097(03)00208-8. pmid: 12725920.

Vázquez-Bermúdez, María Félix et al. (2002). "2-Oxoglutarate Increases the Binding Affinity of the NtcA (Nitrogen Control) Transcription Factor for the *Synechococcus glnA* Promoter." In: *FEBS Lett.* 512.1-3, pp. 71–74. ISSN: 00145793. DOI: 10.1016/S0014-5793(02)02219-6. pmid: 11852054.

Vega-Palas, M. A. et al. (Feb. 1990). "Identification and Cloning of a Regulatory Gene for Nitrogen Assimilation in the Cyanobacterium *Synechococcus* Sp. Strain *PCC* 7942." In: *J. Bacteriol.* 172.2, pp. 643–647. ISSN: 00219193. DOI: 10.1128/jb.172.2.643-647.1990. pmid: 1967601.

Vega-Palas, M. A. et al. (July 1992). "NtcA, a Global Nitrogen Regulator from the Cyanobacterium *Synechococcus* That Belongs to the Crp Family of Bacterial Regulators." In: *Mol. Microbiol.* 6.13, pp. 1853–1859. ISSN: 13652958. DOI: 10.1111/j.1365-2958.1992.tb01357.x. pmid: 1630321.

Videau, Patrick et al. (Aug. 2015). "ABC Transporter Required for Intercellular Transfer of Developmental Signals in a Heterocystous Cyanobacterium." In: *J. Bacteriol.* 197.16. Ed. by P. J. Christie, pp. 2685–2693. ISSN: 10985530. DOI: 10.1128/JB.00304-15. pmid: 26055115.

Videau, Patrick et al. (Nov. 2016). "The Heterocyst Regulatory Protein HetP and Its Homologs Modulate Heterocyst Commitment in *Anabaena* Sp. Strain *PCC 7120*." In: *Proc. Natl. Acad. Sci. U. S. A.* 113.45, E6984–E6992. ISSN: 10916490. DOI: 10.1073/pnas.1610533113. pmid: 27791130.

Videau, Patrick et al. (July 2018). "The *hetZ* Gene Indirectly Regulates Heterocyst Development at the Level of Pattern Formation in *Anabaena* Sp. Strain *PCC 7120*." In: *Mol. Microbiol.* 109.1, pp. 91–104. ISSN: 13652958. DOI: 10.1111/mmi.13974.

Wang, Li et al. (Nov. 2019). "*patD*, a Gene Regulated by NtcA, Is Involved in the Optimization of Heterocyst Frequency in the Cyanobacterium *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 201.21. Ed. by Conrad W. Mullineaux. ISSN: 10985530. DOI: 10.1128/JB.00457-19. pmid: 31405917.

Wang, Shaokui et al. (2012). "Control of Grain Size, Shape and Quality by OsSPL16 in Rice." In: *Nat. Genet.* 44.8, pp. 950–954. DOI: 10.1038/ng.2327.

Wang, Yu and Xudong Xu (Dec. 2005). "Regulation by *hetC* of Genes Required for Heterocyst Differentiation and Cell Division in *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 187.24, pp. 8489–8493. ISSN: 00219193. DOI: 10.1128/JB.187.24.8489-8493.2005. pmid: 16321953.

Watanabe, I. (1982). "Azolla—Anabaena Symbiosis — Its Physiology and Use in Tropical Agriculture." In: *Microbiology of Tropical Soils and Plant Productivity*. Ed. by Y. R. Dommergues and H. G. Diem. Dordrecht:

Springer Netherlands, pp. 169–185. ISBN: 978-94-009-7531-6. DOI: 10.1007/978-94-009-7529-3_6.

Water, S. D. Van De and R. D. Simon (May 1982). "Induction and Differentiation of Heterocysts in the Filamentous Cyanobacterium *Cylindrospermum* Licheniforme." In: *J. Gen. Microbiol.* 128.5, pp. 917–925. ISSN: 00221287. DOI: 10.1099/00221287-128-5-917.

Wei, Shaobo et al. (July 2022). "A Transcriptional Regulator That Boosts Grain Yields and Shortens the Growth Duration of Rice." In: *Science* 377.6604, eabi8455. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abi8455.

Wei, T. F. et al. (1994). "*Anabaena* Sp. Strain *PCC 7120 ntcA* Gene Required for Growth on Nitrate and Heterocyst Development." In: *J. Bacteriol.* 176.15, pp. 4473–4482. ISSN: 00219193. DOI: 10.1128/jb.176.15.4473-4482.1994. pmid: 7913926.

Wolk, C. P. (1967). "Physiological Basis of the Pattern of Vegetative Growth of a Blue-Green Alga." In: *Proc. Natl. Acad. Sci. U. S. A.* 57.5, pp. 1246–1251. ISSN: 00278424. DOI: 10.1073/pnas.57.5.1246. pmid: 5231730.

Wolk, C. Peter (Oct. 1975). "Formation of One-Dimensional Patterns by Stochastic Processes and by Filamentous Blue-Green Algae." In: *Dev. Biol.* 46.2, pp. 370–382. ISSN: 00121606. DOI: 10.1016/0012-1606(75)90113-X. pmid: 810378.

Wong, F. C. Y. and J. C. Meeks (Apr. 2001). "The *hetF* Gene Product Is Essential to Heterocyst Differentiation and Affects *hetR* Function in the Cyanobacterium *Nostoc punctiforme*." In: *J. Bacteriol.* 183.8, pp. 2654–2661. ISSN: 00219193. DOI: 10.1128/JB.183.8.2654-2661.2001. pmid: 11274126.

Wu, Kun et al. (2020). "Enhanced Sustainable Green Revolution Yield via Nitrogen-Responsive Chromatin Modulation in Rice." In: *Science* 367.6478, eaaz2046. DOI: 10.1126/science.aaz2046.

Wu, Xiaoqiang et al. (Oct. 2004). "*patS* Minigenes Inhibit Heterocyst Development of *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 186.19, pp. 6422–6429. ISSN: 00219193. DOI: 10.1128/JB.186.19.6422-6429.2004. pmid: 15375122.

Xing, Wei-Yue et al. (Sept. 2022). "A Proteolytic Pathway Coordinates Cell Division and Heterocyst Differentiation in the Cyanobacterium *Anabaena* Sp. PCC 7120." In: *Proc. Natl. Acad. Sci. U.S.A.* 119.36, e2207963119. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2207963119.

Xing, Yongzhong and Qifa Zhang (Feb. 2010). "Genetic and Molecular Bases of Rice Yield." In: *Annu. Rev. Plant Biol.* 61, pp. 421–42. DOI: 10.1146/annurev-arplant-042809-112209.

Xu, Xiaomei et al. (Aug. 2020). "*hetL*, *hetR* and *patS* Form a Reaction-Diffusion System to Control Pattern Formation in the Cyanobacterium

Nostoc *PCC 7120*." In: *eLife* 9, pp. 1–48. ISSN: 2050084X. DOI: 10.7554/ELIFE.59190. pmid: 32762845.

Yao, Ruifeng et al. (2016). "DWARF14 Is a Non-Canonical Hormone Receptor for Strigolactone." In: *Nature* 536.7617, pp. 469–473. DOI: 10.1038/nature19073.

Yoon, H. S. and J. W. Golden (Apr. 2001). "*patS* and Products of Nitrogen Fixation Control Heterocyst Pattern." In: *J. Bacteriol.* 183.8, pp. 2605–2613. ISSN: 00219193. DOI: 10.1128/JB.183.8.2605-2613.2001. pmid: 11274121.

Yoon, Ho Sung and James W. Golden (Oct. 1998). "Heterocyst Pattern Formation Controlled by a Diffusible Peptide." In: *Science* 282.5390, pp. 935–938. ISSN: 00368075. DOI: 10.1126/science.282.5390.935. pmid: 9794762.

Young-Robbins, Shirley S. et al. (Sept. 2010). "Transcriptional Regulation of the Heterocyst Patterning Gene *patA* from *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 192.18, pp. 4732–4740. ISSN: 00219193. DOI: 10.1128/JB.00577-10. pmid: 20622060.

Yu, Haiyuan and Mark Gerstein (2006). "Genomic Analysis of the Hierarchical Structure of Regulatory Networks." In: *Proc. Natl. Acad. Sci.* 103.40, pp. 14724–14731. DOI: 10.1073/pnas.0508637103.

Zeng, Xiaoli and Cheng-Cai Zhang (Sept. 2022). "The Making of a Heterocyst in Cyanobacteria." In: *Annu Rev Microbiol* 76, pp. 597–618. ISSN: 1545-3251. DOI: 10.1146/annurev-micro-041320-093442.

Zhang, Cheng Cai et al. (Jan. 2006). "Heterocyst Differentiation and Pattern Formation in Cyanobacteria: A Chorus of Signals." In: *Mol. Microbiol.* 59.2, pp. 367–375. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2005.04979.x. pmid: 16390435.

Zhang, Cheng Cai et al. (2018). "Carbon/Nitrogen Metabolic Balance: Lessons from Cyanobacteria." In: *Trends Plant Sci.* 23.12, pp. 1116–1130. ISSN: 13601385. DOI: 10.1016/j.tplants.2018.09.008. pmid: 30292707.

Zhang, He et al. (Feb. 2018). "Functional Overlap of *hetP* and *hetZ* in Regulation of Heterocyst Differentiation in *Anabaena* Sp. Strain *PCC 7120*." In: *J. Bacteriol.* 200.9. Ed. by Victor J. DiRita, e00707–17. ISSN: 10985530. DOI: 10.1128/JB.00707-17. pmid: 29440250.

Zhang, Wei et al. (Nov. 2007). "A Gene Cluster That Regulates Both Heterocyst Differentiation and Pattern Formation in *Anabaena* Sp. Strain *PCC 7120*." In: *Mol. Microbiol.* 66.6, pp. 1429–1443. ISSN: 0950382X. DOI: 10.1111/j.1365-2958.2007.05997.x. pmid: 18045384.

Zhao, Jiaping et al. (Oct. 2017). "Genome-Wide Constitutively Expressed Gene Analysis and New Reference Gene Selection Based on Transcriptome Data: A Case Study from Poplar/Canker Disease Interaction."

In: *Front. Plant Sci.* 8, p. 1876. ISSN: 1664-462X. DOI: 10.3389/fpls.2017.01876.

Zhou, Ruanbao et al. (Apr. 1998a). "Characterization of *hetR* Protein Turnover in *Anabaena* Sp. *PCC 7120*." In: *Arch. Microbiol.* 169.5, pp. 417–426. ISSN: 03028933. DOI: 10.1007/s002030050592. pmid: 9560423.

Zhou, Ruanbao et al. (1998b). "Evidence That *hetR* Protein Is an Unusual Serine-Type Protease." In: *Proc. Natl. Acad. Sci. U. S. A.* 95.9, pp. 4959–4963. ISSN: 00278424. DOI: 10.1073/pnas.95.9.4959. pmid: 9560210.

Zhu, Mei et al. (Nov. 2010). "Maintenance of Heterocyst Patterning in a Filamentous Cyanobacterium." In: *J. Biol. Dyn.* 4.6, pp. 621–633. ISSN: 17513758. DOI: 10.1080/17513751003777507. pmid: 22881208.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

https://bitbucket.org/amiede/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of December 12, 2022 (`classicthesis` version 1.0).