# Politecnico di Torino

Master Degree in Physics of Complex Systems



Master Degree's Thesis

# PID-APOLO locus in Arabidopsis thaliana: quantitative modeling for inferring operational principles


Internal supervisor:
Prof. Andrea PAGNANI

Candidate:
Roberto NETTI

Supervisors:
Dr. Silvia GRIGOLON
Dr. Olivier MARTIN


October 2022

**Abstract**

The development of living organisms is driven by the orchestrated expression of proteins and hormones which allow cell differentiation and thus organ formation. In plants, this role is mainly driven by the hormone auxin, which is produced at the single-cell level and actively transported across tissues, i.e., by the action of dedicated carriers. Among the carriers, PINs are known to mediate the outgoing flux of auxin and display polarization patterns across tissues. PIN polarity and therefore auxin transport is determined by the expression level of a protein kinase called Pinoid, encoded in the gene PID which is the neighboring gene of APOLO, a long non-coding RNA believed to be associated with changes in local epigenetic status.
A 2014 paper [1], pointed out how the transcription of APOLO drives the state of the PID-APOLO locus, switching it on or off therefore acting as a signed differentiator circuit, which turns on via auxin stimulus and then spontaneously re-closes in about 24 hours, even though the auxin stimulus is still there.
Although the past decades have seen extensive work on this subject, it is not known yet what ensures the loop closure, whether there is a role for basal transcription to maintain the closed loop or whether the observed basal level is ensured by a mixture of open and closed loops which undergo a specific dynamics of opening and closing.
The aim of this thesis is to use theoretical modelling based on dynamical systems and ordinary differential equations combined with inference methods to answer these questions. Based on molecular biology measurements in roots produced with APOLO overexpressors and RNAi knockdowns collected by Crespi's team at the Institut des Sciences du Végétal in Gif-sur-Yvette, mathematical models have been constructed and fine-tuned that attempt to provide a quantitative interpretation of the dynamics in the gene circuit. This approach made it possible to identify the most important interactions that shape the observed dynamics and to identify the role of the various actors in the system.

# Acknowledgements

First, I would like to give special thanks to my supervisors, without whom I could not have written this thesis. To Silvia and Olivier who taught me so much during these months, who always helped me with patience and dedication. Their guidance has been crucial, and discussing with them has always been inspiring. I am grateful for the support and very important advice they have given me.
A big thank you also goes to my family and friends, who have always comforted and encouraged me to move forward on my path.
Finally, thank you to all my classmates who always made me feel welcomed and eased this two-year journey.

# Contents

# List of Figures

# Chapter 1

# Introduction to biological sequences and eukaryotic gene expression

The aim of this first chapter is to introduce the reader to all the tools necessary to the understanding of the biological problem addressed in this thesis project. In this first chapter, we shall define the main biological sequences, their structure and how they are organized inside cells as well as the main mechanisms underlying gene expression, i.e., the production of proteins which allow all vital functions to be carried out, and gene regulation, in particular at the post-transcriptional level.

## 1.1 Biological sequences

The field of molecular biology studies macromolecules and the macromolecular mechanisms found in living organisms, such as the molecular nature of the gene and its mechanisms of gene replication, mutation, expression and regulation. These macromolecules are generally biopolymers, i.e., one-dimensional chains of building blocks called monomers, held together by chemical bonds and playing crucial roles in the development and aging of organisms. Biopolymers in nature can be of two different kinds: i) nucleic acids, such as *deoxyribonucleic acid* (DNA) and *ribonucleic acid* (RNA), or ii) proteins.
The DNA carries the "genetic material", that is the information required by an organism to function, and is transferred to the offspring. In all organisms owing a nucleus, such as eukaryotes, the DNA is located in the nucleus itself whereas proteins are made in the cytoplasm, that is outside the nucleus. The intermediate molecule transferring the genetic information out of the nucleus, i.e., between the DNA and the protein, is called the RNA. The

DNA normally occurs as a double-stranded helix: this means that it must be opened in order to be read and expressed. It thus serves as a template for the assembly of specific RNA, the messenger RNA (mRNA), which is read by dedicated proteins inside cells into groups of three nucleotides and then translated into proteins thanks to the action of other ribonucleic acids. These two processes are respectively known as *transcription* and *translation* (figure 1.1).

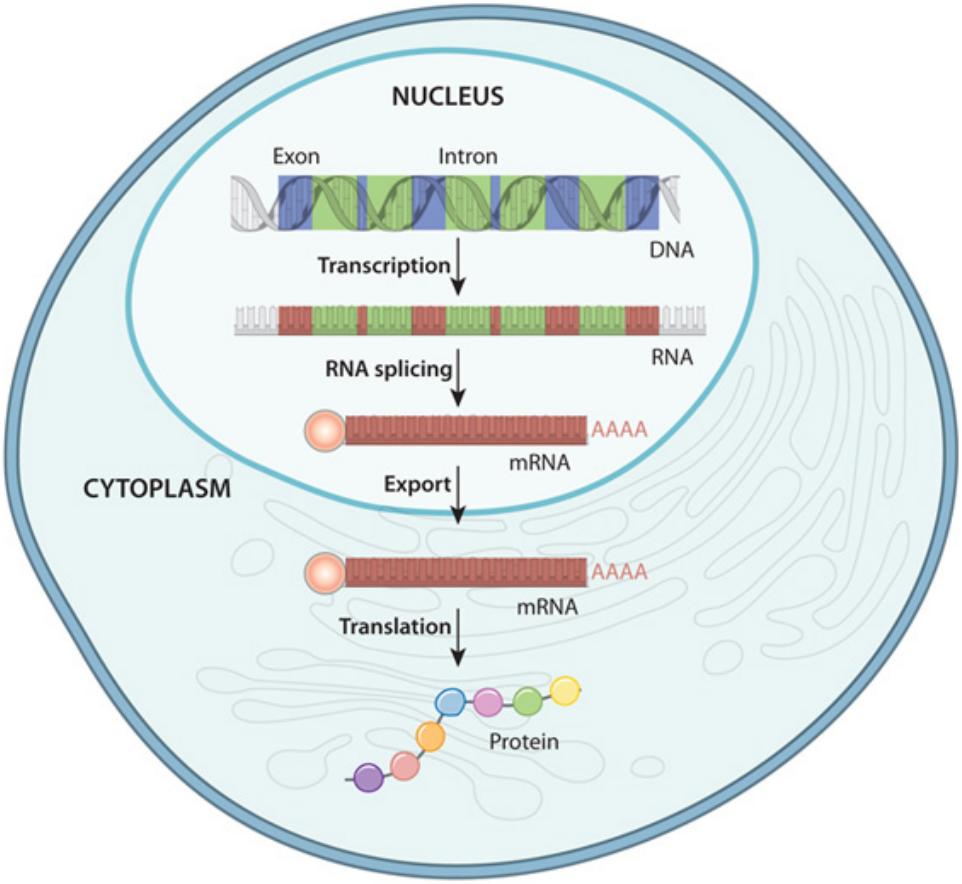



Figure 1.1: **An overview of the flow of information from DNA to protein in a eukaryote.** First, DNA is transcribed into mRNA, then the latter is prepared for export out of the nucleus. Once in the cytoplasm, the mRNA can be used to construct a protein [2].

This information flow in biology is summarized by the "central dogma", put forward by Francis Crick in 1958.

**Central Dogma of Molecular Biology.** *The central dogma states that once 'information' has passed into protein it cannot get out again. The transfer of information from nucleic acid to nucleic acid, or from nucleic*

*acid to protein, may be possible, but transfer from protein to protein, or from protein to nucleic acid, is impossible. Information means here the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein. [3]*

The set of processes characterizing the production of proteins is globally called **gene expression** and starts from the DNA. The overall idea is that one macromolecule can be used as a template to construct another and the information flow is unidirectional. Yet, in order to better understand how this takes place in cells, in the following we will introduce in detail the structure and organization of DNA, the different types of RNAs participating into the process of gene expression, the proteins as well as all the different mechanisms globally leading to gene expression.

### 1.1.1 Deoxyribonucleic acid (DNA) and Chromosomes

#### DNA:

The *deoxyribonucleic acid* is a biopolymer made of single repeated monomers called *nucleotides*.
Nucleotides are composed of three subunits: i) a phosphate group, ii) a sugar residue and iii) a nucleobase. The latter can be distinguished in turn into four types: Adenine (A), Cytosine (C), Guanine (G) and Thymine (T). A and G are called *purine*, C and T *pyramidine*. Thanks to the complementarity of their geometrical shape and hydrogen-bonding, each purine is coupled to one pyramidine according to the scheme A-T and C-G. This property allows the DNA to assume its characteristic double-helix structure, where a first strand is paired to a complementary one running in the opposite direction [4].

#### Chromosomes:

The main function of the DNA is to carry genes. A gene in molecular biology and genetics indicates the fundamental hereditary unit of living organisms. A gene is a portion of the nucleotide sequence of DNA that encodes the primary sequence of an end gene product, which can be either a structural or catalytic RNA, or a polypeptide. In eukaryotes, the DNA is enclosed in the cell nucleus in a highly compact state. For instance, if the double helices of human DNA could be laid end-to-end they would reach approximately two meters, yet the nucleus is only about 6 $\mu$m in diameter. This gives an idea of how complex is the task of DNA packaging in cells. More specifically, specialized proteins called *histones*, can bind to and fold the DNA, thus generating a series of coils and loops that provide increasingly higher levels of organization, preventing the DNA to become an

unmanageable tangle. The latter is divided among a series of chromosomes within the nucleus. Each of these chromosomes consists of a single long DNA molecule associated with proteins that package and fold it into a more compact structure.

**Chromatin:**

The whole set of DNA and proteins is called *chromatin*. In particular, proteins that bind the DNA are divided into two classes called *histones* and *non-histones chromosomal proteins*, respectively. Partially unfolded chromatin appears as a series of "beads on a string" if analyzed under electron microscope: the string is the DNA and each bead is called *nucleosome*, consisting of DNA wound around a core made of histones (figure 1.2). The nucleosome core is composed by an *histone octamer*, i.e., eight histones proteins, around which a double-stranded DNA of 147 nucleotide pairs is wound. This new structure allows the formation of a chromatin thread about one third of its initial length. Each of the core histones has an N-terminal amino acid tail, which extends out from the nucleosome core. These tails are involved in the control of critical aspects of chromatin structure and function. For instance, chromatin rarely adopts the extended "beads on a string" form, rather nucleosomes are packed together, generating regular arrays that make the DNA even more compact [5].
Specifically, chromatin can be found in two states of compaction: a highly condensed form, called *heterochromatin*, and a less condensed one called *euchromatin*. Actually, the term *heterochromatin* refers to several types of chromatin that share the common feature of being highly compacted.
The transition between these two types of chromatin is due to histone modifications. Histone modifications are reversible changes, created by one specific enzyme and removed by another. More specifically, the side chain of histones may be subject to covalent modifications, including acetylation, methylation of lysines, and phosphorylation of serines (occurring in the N-terminal histone tails). Acetylation of lysines, for example, tends to loosen the chromatin structure, reducing the affinity of the tails for adjacent nucleosomes. As we will see later, the main role of this switch in chromatin structure is that, by changing the type of packing, the production of RNA transcripts and proteins can be interrupted or increased.

[3]Meiosis is a special mode of cell division, which makes haploid cells from a diploid cell. It is essential for sexual reproduction in eukaryotes and diploid organisms and produces gametes [7].

[3]Mitosis is the process by which eukaryotic cells assure the equipartition of their chromosomes at cell division [8].

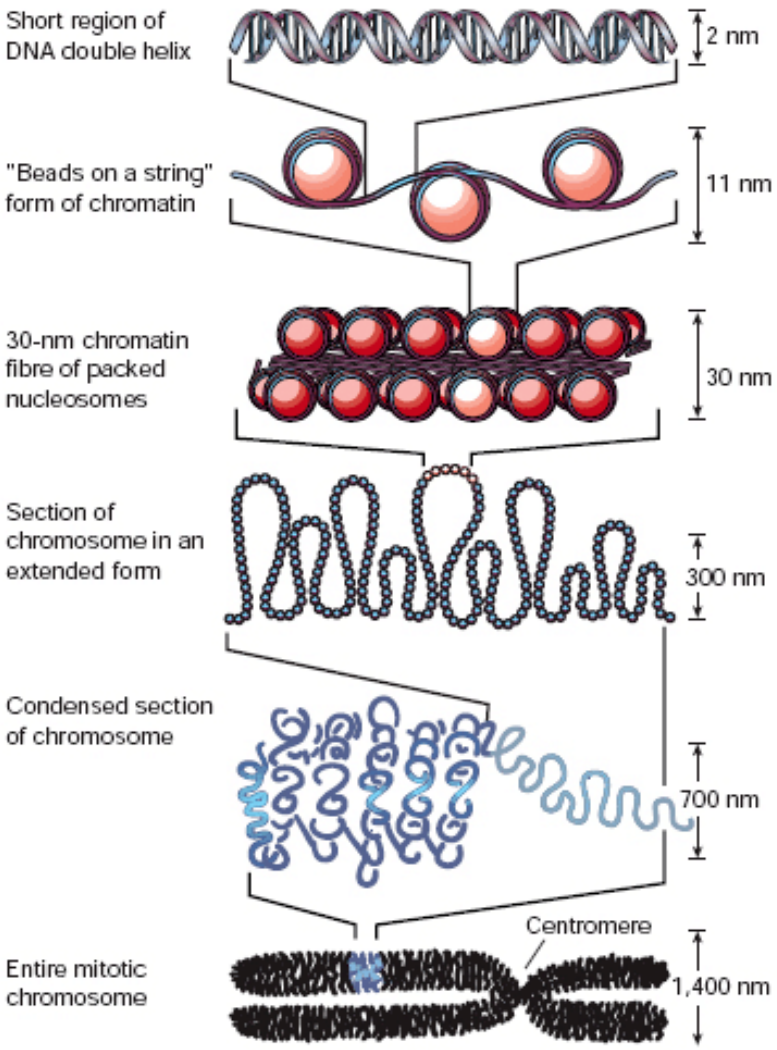



Figure 1.2: **Chromatin folding.** The first level of compaction is the association of the DNA with nucleosomes to create the 10 nm 'beads on a string structure'. This can condense into a 30 nm diameter fiber. Further compaction of chromosomes is observed in vivo, 60-300 nm diameter fibers are observed at interphase, these can be further compacted into fibers of around 700 nm diameter and finally fully condensed mitotic chromosomes occur during mitosis [1]and meiosis[23] [6].

### 1.1.2 Eukaryotic gene expression

Given the information so far, in this section we will go into the details of the templating procedure allowing genes to be expressed and proteins to be produced. Recall that by *gene* it is meant a portion of the DNA sequence of nucleotides encoding the information necessary for the production of either different RNAs or a protein.
In the following section, we will first discuss the structure of RNAs and proteins and then the processes leading to gene expression.

**Ribonucleic acid (RNA):**

The *ribonucleic acid* (RNA), like the DNA, is assembled as a chain of nucleotides. Despite this, there are three differences between the RNA and the DNA: i) the sugar is *ribose* instead of *2-deoxyribose*; ii) the nucleobase Uracile (U) replaces Thymine allowing the recognition of the RNA as foreign to the DNA and iii) the RNA appears as single-stranded.
RNAs can be divided into different categories depending on different criterions. Considering their length, it is possible to distinguish between:

- small RNA: usually shorter than 200 nt;
- long RNA: larger than 200 nt;

while considering their role:

- messenger RNA (mRNA), which is a copy of the DNA that carries information used for translation into proteins;
- non-coding RNA (ncRNA), which does not carry any genetic information to be expressed into protein.

Non-coding RNAs like transfer RNAs (tRNAs) have many fundamental roles in *translation* or in RNA processing. Other types of ncRNAs, for example, long-non-coding RNAs (lncRNAs) and small interfering RNAs (siRNAs) are involved in gene regulation, which will be discussed in the following paragraphs as they will play a central role in this thesis project.

**Proteins:**

Proteins are the final product of gene expression. A protein is a linear chain of building blocks called amino acids forming a *polypeptide chain*. Amino acids are kept together thanks to peptide bonds[4] and are mainly composed

[4]Peptide bond is a covalent bond that is established between two molecules when the carboxylic group of one reacts with the amine group of the other through a condensation reaction (or dehydration, leading to that is, the elimination of a molecule of water).

by an *amine group* $-NH_2$ and a *carboxyl group* $-COOH$. To each molecule a *side chain* R is bound, which allows the distinction of 20 different types of amino acids.
The polypeptide chain folds onto itself forming the protein structure which is described by four levels:

- primary structure, which refers to the amino acid sequence held together by peptide bonds, determined from the gene;
- secondary structure which refers to the formation of highly regular substructures ($\alpha - helices$ and $\beta - sheets$), due to hydrogen bonds;
- tertiary structure which is the 3D structure of the fully folded peptide chain that can include one or several domains;
- quaternary structure which is a 3D structure composed by different polypeptide chains that form a larger protein.

Thanks to their structure, proteins are able to carry out the majority of vital functions, e.g., catalyzing metabolic reactions, responding to stimuli and are involved in DNA replication and structural functions amongst all.

**Gene expression:**

Once the main players have been introduced, we can move on to describing the entire process of gene expression, allowing the information flow from the DNA to the final product, proteins. The two main mechanisms by which cells read out the genetic instructions in their genes are called *transcription* and *translation*. The synthesis of most protein molecules takes between 20 seconds and several minutes [9], but the fact that many identical copies of RNA can be produced by the same gene and each RNA can itself produce many identical proteins allows the cell to make a large quantity of molecules in a short time. Moreover, each gene can be transcribed and translated with different efficiency, each cell can regulate the transcription of its genes, making possible the production of large quantities of some molecules and tiny quantities of others. The steps allowing protein production from the DNA sequence are the following:

- transcription: the DNA is transcribed into a chain of pre-mRNA;
- pre-mRNA processing: modifications of the pre-mRNA that lead to the mature mRNA formation;
- transport of the mRNA to the ribosomes;

- translation: the mRNA is translated into the polypeptide chain, i.e., the protein.

As already stated in the *Central Dogma*, information is transferred from the DNA to the proteins and not backwards.

**Transcription (from DNA to RNA):**

During this first step, the cell reads out the part of interest of the genetic instructions and the corresponding DNA sequence is copied into an RNA one. The information in the RNA is still written in the form of nucleotides, hence the name *transcription*. Transcription begins with the opening of the double helix, which exposes the bases on the two strands of the DNA, the latter referred to as the *template* strand and *coding* strand, respectively. Then, starting from the template strand, complementary base pairing enables RNA synthesis. The role of reading the DNA before copying it is carried out by some RNA-like enzymes called *RNA-polymerases* which bind to the selected gene's *promoter* locus[5]. Having identified the starting site, the polymerase moves stepwise along the DNA, unwinding the helix to expose the next region for complementary base-pairing. The process ends, when a "stop" sequence is reached. During the procedure, the RNA *transcript* is elongated one nucleotide at a time and the final sequence is exactly complementary to the DNA template. In the end, since hydrogen bonds do not act between RNA and DNA, the first one is immediately released as single stranded.
The product of this phase is the pre-mRNA. It undergoes the so called *splicing* that subsequently produces the mature mRNA.

**Translation (from RNA to proteins):**

Once the mature mRNA is ready, the next step is the production of the corresponding protein. How is the information in a sequence of nucleotides translated in a sequence of chemically different units, i.e., the protein amino acids? We have seen how the transfer of information between DNA and RNA is possible thanks to the complementarity of their units, their coding language does not entirely change and the symbols used are closely related. The last step between RNA and protein is more complicated though. In this case, the two sequences are written in two different languages with different symbols: thus a *translation* procedure is needed. Which are the rules of this process? At the base of translation is the fact that nucleotides

[5]In genetics, a promoter is a sequence of DNA to which proteins bind to initiate transcription of a single RNA transcript from the DNA downstream of the promoter. Promoters are located near the transcription start sites of genes, upstream on the DNA.

are read in consecutive groups of three, each group being called *codon*. The combination of three nucleotides gives rise to $4^3 = 64$ different codons, each codon is associated with one of the 20 amino acids, but one amino acid can be associated to more than one codon, and the whole process is carried out in a specific macromolecular machinery called *ribosomes*. It is important to specify that each codon is not able itself to recognize its amino acid: it is in fact necessary the intervention of transfer RNAs that bind together both codons and amino acids. The primary structure of the protein will be then formed thanks to the bonds between amino acids, that will subsequently fold into the protein functional form.

## 1.2 Regulation of gene expression

Multicellular organisms contain many cell types, each of them differing dramatically from one another in both composition and function. It is known experimentally that this differentiation is due to changes in gene expression rather than in the cell genome. Yet, a central question is: how do cells differentiate? Moreover, all these types of cells must behave differently and do so in the correct way. For a cell to function properly, the necessary proteins must be synthesized at the right time and at the right amount, and it must also respond quickly to changes in the environment. What are the mechanisms by which the cell responds? The answer is gene regulation.

Thanks to gene regulation, each cell type has different sets of active genes causing the presence of different concentrations of proteins. Some proteins are abundant in specialized cells in which they have a specific function and are completely absent in others. One example is given by hemoglobin which is present in red blood cells but undetectable in all the others. Another fundamental factor in gene regulation is the response of the cell to external factors: some cells respond to the presence of a signal, for instance a hormone, changing their gene expression pattern, enhancing or reducing the production of some specific proteins. Clearly, different cell types do not respond to an external signal in the same way.
We have seen there are many stages to pass from the DNA to proteins. At each stage, gene expression can be regulated by:

- transcriptional control, controlling directly the transcription of a gene;
- RNA processing control, acting on the splicing and the processing of RNA transcripts ;
- RNA transport and localization control, which is determinant in the spatio-temporal articulation of gene expression;

- translational control, i.e., the selection of mRNA which is translated into proteins;
- mRNA degradation control, i.e., the destabilization of certain mRNA types;
- protein activity control, acting at the protein level through the intervention of enzymes that lower the activation energy of protein reactions.

Despite this large amount of stages, transcriptional control is the most common and will be discussed in detail in the following paragraphs.

**Gene regulatory proteins:**

The transcription of each gene is controlled by a *regulatory region*, i.e., a sequence of nucleotides in the DNA close to the respective gene. These regions act as a switch for the initiation of the DNA transcription in response to an external signal. These switches are made of two components:

- short stretches of DNA which act as binding sites;
- gene regulatory proteins that bind to these DNA stretches.

The role of the *gene regulatory proteins* is to recognize specific nucleotide sequences, complementary in shape, in the DNA double helix and turn a specific set of genes on or off.

**Transcriptional control:**

How do these components work to regulate the gene transcription?
Let us introduce this topic starting with a clear example of a *Transcriptional repressor*. The simplest case is indeed that of an on-off switch that respond to a single signal. As explained in the previous section, the promoter is the locus in which the polymerases open the DNA helix and begin synthesizing the RNA molecule. Within the promoter, there is a region recognized by the transcriptional repressor. When the repressor is bound to the DNA it blocks the access to the promoter by the polymerase, preventing the transcription. For obvious reasons, this mode of gene regulation is called *negative control*. In contrast to the negative control, let us bring an example of the so called *positive control*. Many gene promoters are actually marginally functional on their own, in other words the probability for the polymerase to bind and begin the transcription is really small. In these cases, the role of the *transcriptional activators* is to bind the promoter and dramatically increase the affinity between the former and the polymerase. The result is a larger

probability for the polymerase to open the double helix and initiate the transcription.
Another feature of gene regulation is so-called *DNA looping*. A first case is due to the ability of the repressor to bind two different DNA sites, creating a loop, which leads to even greater levels of repression of transcription [10], but the same can occur with the result of activating transcription of the gene. This process has been discovered to be fundamental in eukaryotic gene regulation.

Usually in eukaryotes the processes leading to the gene regulation are more complex than the previous simple cases. Typically, gene regulatory proteins bind far away from the promoter and the DNA sequences controlling the expression are spread over long stretches of DNA. The whole region, including the promoter, involved in the transcriptional control is referred as *gene control region*. In the case of gene activators, their main role is to attract, position and modify the general transcription factors and the polymerase, so the transcription can begin. In addition they can also modify the chromatin structure to promote transcription. In practice, gene activators trough histone modifications are able to make chromatin more accessible, facilitating the assembly of transcription factors. However, even before gene regulatory proteins have a role, many genes are *"marked"* to become rapidly activated. When considering eukaryotic repressors, it is important to note that they have many more ways to act on transcription, most of which do not directly compete with polymerase to access the DNA, but they rather use a variety of other mechanisms.

### 1.2.1 Post-transcriptional control

As discussed before, gene regulation can act at every stage of the process from the DNA to the proteins, indeed many genes are regulated by multiple mechanisms. In the previous section we went through the transcriptional control, but other controls can act after the RNA polymerase has bound the gene promoter. This type of controls are globally are called *post-transcriptional controls*. These regulatory mechanisms include the attenuation of the RNA transcript by premature termination, alternative RNA splicing, RNA editing, control of transport and many others.
It is important to mention the *small non-coding RNAs*. Non-coding RNAs, which have been already discussed previously, have many roles in cells, especially in gene regulation. Among the *small non-coding RNAs*, *microRNA (miRNA)* play a crucial role inside cells. Once made and assembled with a set of proteins, miRNAs bind to a specific complementary RNA sequence producing several outcomes. Post-transcriptional control operated by miRNAs can occur through either translational repression or mRNA degrada-

tion. It has been shown that an individual miRNA is able to control the expression of more than one target mRNAs and that each mRNA may be regulated by multiple miRNAs [11]. In addition to the fine-tuning of a target protein level it has been proved the role of miRNA in efficient noise controlling of gene expression, thus conferring precision and stability to the overall process [12, 13]. Finally, because miRNAs can interact with multiple mRNAs, the noise control effect can be extended to multiple mRNAs that can can influence each other's fluctuations [14].

**RNA-interference:** One of the roles of small RNAs is to lead to the degradation of some double-stranded RNAs [15]. This mechanism, called *RNA-interference*, is structured as follows: a double-stranded RNA triggers the *RNAi* process by attracting a protein complex that is able to cleave the double-stranded RNA into small pieces: *small interfering RNAs (siRNA).* The single-stranded siRNAs, by binding other components, are able to exactly match the RNAs and lead to their destruction.
The RNA-interference mechanism is also able to interact with RNA transcripts and modify nearby histones directing the spread of heterochromatin that fatherly prevent gene expression.
Apart from regulatory and defense roles, RNA-interference has brought a powerful experimental tool for scientists, allowing almost any gene to be intentionally inactivated by evoking RNAi response [16].

**RdDM pathway:** Another epigenetic mechanism for regulating gene expression at the transcription and post-translation levels in eukaryotes is represented by DNA methylation[6]. The principle of RNA-mediated epigenetic pathway is the *RNA Directed DNA Methylation (RdDM)* pathway, which mediates de novo DNA methylation of cytosine bases, using small interfering RNAs (siRNAs). An important element of the RdDM pathway is a particular protein complex in which AGO4 (argonaute 4)[7] binds to small RNAs, including siRNAs or long non-coding RNAs (lncRNAs), and cleaves the target RNA transcripts. Moreover, AGO4 catalytic activity is important for the creation of secondary siRNAs that strengthen its repressive effects [17]. The protein complex guides DNA methylation at homologous loci through binding of small RNAs to AGO4 [18]. Studies suggest that the RdDM pathway largely participates in various physiological activities. In

[6]DNA methylation is a biological process by which methyl groups are added to the DNA molecule. Methylation can change the activity of a DNA segment without changing the sequence. When located in a gene promoter, DNA methylation typically acts to repress gene transcription.

[7]Argonaute proteins are the catalytic component of the RNA-induced silencing complex, the riboprotein complex responsible for the gene expression silencing process known as RNA interference (or RNAi).

plants for example: seed development, dormancy, fruit maturation, sexual reproduction and coloration [19, 20, 21, 22, 23, 24].

**Active DNA methylation in eukaryotes:** We have seen that DNA cytosine methylation is an important epigenetic mark and plays a central role in the regulation of gene expression. The DNA methylation mark, besides being added only through DNA methylation pathways, can also be removed through DNA demethylation pathways. Compared to the DNA methylation, the active DNA demethylation is less well understood in both mammals and plants. DNA methylation can be passively lost or actively removed, and these processes are referred to as passive DNA demethylation and active DNA demethylation, respectively. The entire methylcytosine base is actively removed from the DNA backbone by the action of specific enzyme families, in the case of plants these are: ROS1/DME. The latter are recruited to the target loci where they initiate demethylation [25]. Fundamental to the thesis will be the case where the action on DNA is mediated by ROS1, DML2, and DML3 enzymes (*RDD pathway*), which are shown to be triggered by auxin, a key plant hormone.

# Chapter 2

# Auxin and its role in plant gene expression: the regulation of the PID-APOLO locus

This thesis project focuses on the modelling, the analysis and the understanding of the mechanisms regulating the expression of a specific gene locus in plants. In this chapter, we will introduce the reader to the world of plant morphogenesis, defining the main biological components involved in the system as well as the experimental results which motivated this research project.

## 2.1 Biological components

This first section is devoted to introducing the key players in the system under consideration. Starting with auxin and its role in morphogenesis, we will introduce the PID gene, a regulator of auxin transport, and its neighboring gene APOLO, which is co-regulated along with PID.

### 2.1.1 Plant morphogenesis and the role of auxin in plant development

The word *morphogenesis* represents the set of processes that characterize the formation of organs from embryonic tissues of *stem cells*. These latter are undifferentiated cells, that need to specialize for the organism to take its shape and carry out its vital functions.
How does cell differentiation take place though? The appearance of new tissues and organs at the right place and moment during a plant life cycle must be guided in some way. This is the role of *morphogens* which, thanks to different levels of their concentrations and gradient in the tissues, encode

the different fundamental signals for cell differentiation. Those signals are, in many cases, received from surrounding cells and the environment.
In the particular case of plants, that we are analyzing in this thesis, the role of cell differentiation is fulfilled by a complex network of interactions between various substances, one of the most important being *auxin*.
Deriving from the Greek *auxein* = to grow, it is difficult to classify auxin in any category because of its multiplicity of roles played in plant morphogenesis. It was first named as an *hormone* and later referred as a *phytohormone*, a substance produced not only in plants but also in algae. It is involved in both development and regulatory processes, it is often seen as a signal that triggers various reactions and can be expressed in different tissues, it can be transported over the whole plant and influences numerous processes. For instance, in the case of the plant *A. thaliana*, which will be introduced in more detail later, auxin is involved in the development of different parts of the plant and in different processes such as *gravitropism* in roots, i.e., the ability of the plant to respond to gravity. Auxin is transported throughout plants and this transport is critical for plant development. Its transport, differently than other morphogens in other organisms, does not rely on simple diffusion: in fact auxin is actively transported across tissues, i.e., membrane proteins mediate auxin transport inside and outside cells [26]. It has been found that a family of these proteins, the PINs, relocalize on cell membranes thus displaying polarized patterns across tissues which provide directional and positional information essential for developmental processes. Several recent studies have investigated the complex mechanisms by which PIN-dependent and tissue-specific auxin transport creates localized gradients that are essential for plant organs or meristems[8] formation, patterning, and maintenance [26, 27].

### 2.1.2 PID

PIN-mediated auxin transport also appears to be regulated by the PINOID (PID) gene. The latter encodes a regulatory kinase[9] that determines the polar localization of the PIN-FORMED auxin transporter in root cells and is required for root development [28].

### 2.1.3 APOLO

As we will show in the next section, PID is co-regulated in an intricate network of epigenetic and transcriptional factors with another gene, which

[8]The meristem is a type of tissue found in plants. It consists of undifferentiated cells (meristematic cells) capable of cell division. Cells in the meristem can develop into all the other tissues and organs that occur in plants.

[9]A kinase is an enzyme that adds phosphate groups to other molecules.

is called APOLO. Long non-coding RNAs (lncRNAs) are becoming increasingly important regulators of gene expression: *APOLO (AUXIN REGULATED PROMOTER LOOP RNA)* is one example of lncRNA with multiple roles in gene regulation and root development, with both positive and negative regulatory capacity [29].

### 2.1.4 Arabidopsis thaliana

We conclude this section by introducing the plant taken under analysis in this work, i.e., the model organism for plant biology: *Arabidopsis thaliana*. Its genome was completely sequenced in 2000, and this is one of the reasons why it has become so popular in genetic experiments. Other reasons are that Arabidopsis is able to produce a large amount of seeds, has a short intergeneration time, is quite small in size, and has a very short genome. Arabidopsis is structured in three parts:

- apical part which includes leaves and reproductive organs;
- middle part, connecting apical and basal parts;
- basal part, situated below the surface and thus in the soil and involved in seeking nutrients.

## 2.2 Noncoding Transcription by Alternative RNA Polymerases Dynamically Regulates an Auxin-Driven Chromatin Loop

In this second section, we will introduce the experimental results that motivated this research project. In 2014, Ariel and collaborators revealed the molecular mechanism underlying the co-regulation of PID and APOLO [1]. Before illustrating and explaining their results, let us take up some concepts that will prove useful. As explained previously, chromatin structure can play a regulatory role by influencing DNA accessibility. Some types of chromatin folding can put distant gene elements in close proximity with both local and long-distance loops. In other cases, loops may be formed by lncRNAs that participate in Polycomb-dependent repression[10]. lncRNAs may also contribute to gene repression through histone modification and DNA methylation. Finally, lncRNAs, mediated by Pol IV and Pol V, can lead to the production of 24-nt siRNAs involved in transcriptional silencing

[10]The Polycomb repressive system comprises two central protein complexes, Polycomb repressive complex 1 (PRC1) and PRC2, which are essential for normal gene regulation and development.

of genes by directing DNA methylation of the locus of origin.
Based on this, Ariel and coworkers demonstrated how dual transcription by polymerases II and IV regulates the formation of chromatin loops spanning the PID promoter. Expression of APOLO alters chromatin loop formation, while RNA-dependent DNA methylation, active DNA demethylation and Polycomb complexes control loop dynamics. The result is the regulation of PID expression patterns.

### PID expression depends on its neighbor lincRNA APOLO

PID and APOLO are two neighboring genes involved in divergent transcription. APOLO was found to be transcribed by both Pol II and Pol V. External auxin treatment led to increased expression (RNA levels) of both genes, indicating that PID and APOLO are co-regulated, although the transcript levels of PID were consistently higher than its neighbor. Confirming the hypothesis that both genes are co-regulated, RNAi-APOLO plants showed repression of PID expression.

### Plant PRC1 and PRC2 modulate the epigenetic landscape of PID-APOLO locus

**APOLO alters the epigenetic landscape:** Repressive H3K27me3 and active H3K9Ac marks[11] were detected during the analysis, and both modifications were modulated during auxin treatment. Notably, the former is highly repressed in RNAi-APOLO plants, supporting the hypothesis that APOLO regulates the epigenetic landscape.

**LHP1 protein binds PID-APOLO locus, influencing their expression:** It is known that *Polycomb-repressive complex 2* (PRC2) deposits the repressive mark H3K27me3 while PRC1[12] stabilizes the marks through the chromodomain-containing protein LIKE HETEROCHROMATIC PROTEIN 1 (LHP1). Analysis of LHP1 confirmed the result that it binds to the APOLO locus and the PID promoter. Consistent with the removal of H3K27me3 in response to auxin, the binding of LHP1 decreased rapidly up to 6 hr. Finally, *lhp1* mutants showed a late decay of APOLO transcript in response to auxin, resulting in an increased PID response.

[11]Repressive and active marks are histone modifications that represent a hallmark of the chromatin state that suppresses or activates gene transcription, respectively.
[12]Polycomb-repressive complex 1.

### Chromatin loop encompassing PID promoter is regulated directly by APOLO and PRC1 component LHP1

**LHP1 affects loop formation:** The modulation of H3K27me3 and the dynamics of LHP1 following auxin treatment led to hypothesize the presence of an LHP1-mediated chromatin loop. Indeed, the authors were able to detect the loop linking PID promoter and APOLO locus, in addition *lhp1* mutant showed a basal loop formation decreased of 50%.

**LHP1 binds APOLO RNA with an effect on chromatin state:** Other considerations led the authors to think that the presence of the chromatin loop depends partly on PRC1 and PRC2 but also on APOLO transcripts.
Using RNA immunoprecipitation, it was evaluated whether LHP1 can bind APOLO RNA directly. Indeed it did: LHP1 binds APOLO transcripts but not PID transcripts. Finally, the delayed loop formation in the *lhp1* mutant, in conjunction with the delayed decay of APOLO levels and the enhanced PID response, led to the hypothesis that chromatin loop formation depends on the PRC1 component of LHP1 through binding to APOLO RNA.

### RNA-mediated DNA methylation and Active DNA demethylation influence loop formation

As mentioned above, the APOLO locus is highly methylated and, recognized by Pol V, produces 24 nt siRNAs that act in Transcriptional Gene Silencing (TGS). At the same time, auxin-induced opening of the chromatin loop allows transcription of APOLO by Pol II. The question at this point is whether TGS participates in loop formation. Experiments with mutants of the TGS pathway showed no loop formation, demonstrating the role of TGS in loop closure. In addition, an analysis of chromatin compaction showed that the APOLO locus is more condensed in RNAi plants, while the intergenic region appeared more relaxed, a feature already seen at other loci in the literature. Further analysis of DNA methylation after loop opening suggested that APOLO Pol II transcripts play a role in loop reconformation that precludes Pol II from accessing the PID promoter. Finally, experiments with *rdd* mutants showed strongly delayed APOLO and enhanced loop formation, clearly pointing to RDD action in loop opening.

### Noncoding transcription influences local chromatin topology

As explained above, RNAi silencing of APOLO led to amplified methylation of the APOLO locus, but also to a less compacted intergenic region. Two interpretations have been proposed to explain this phenomenon. Histone depletion facilitates the formation of chromatin loops, encouraging the

hypothesis that the DNA methylation state modulated by APOLO transcription shapes the nucleosome density of the intergenic region. The second interpretation concerns RNA-dependent DNA Methylation (RdDM): RNAi-mediated silencing of APOLO triggers RdDM, which spreads heterochromatin along the APOLO gene.

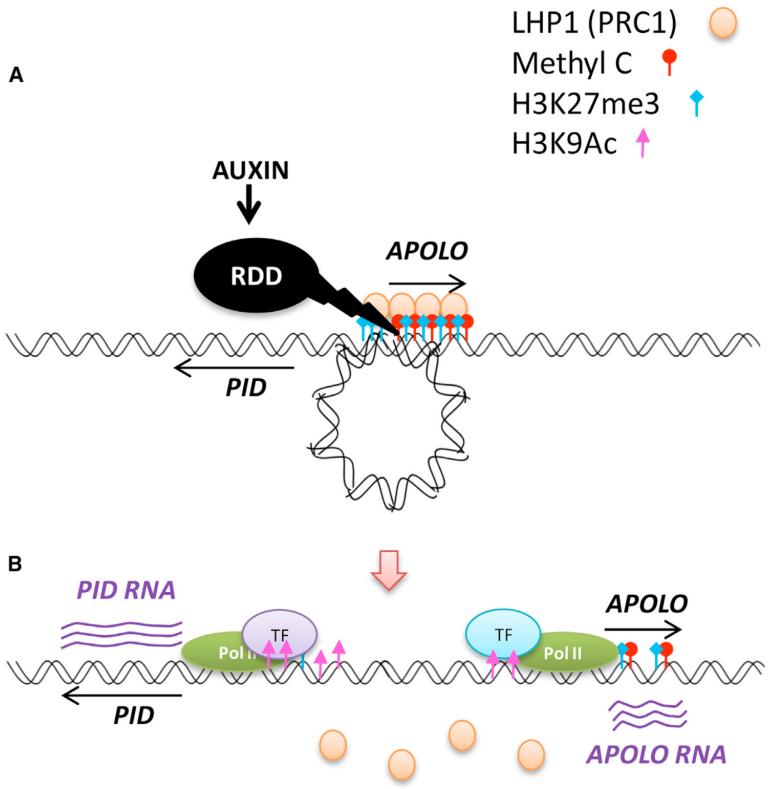



Figure 2.1: **APOLO Transcription by Alternative Polymerases Controls Chromatin Loop Dynamics to Fine-Tune PID Promoter Activity (first half)**
(A) Auxin activates RDD-mediated APOLO DNA demethylation and opens the loop encompassing the PID promoter region. (B) The H3K27me3 mark (pale blue symbols) decreases, whereas H3K9Ac levels (pink symbols) increase. Pol II divergent transcription leads to increased accumulation of PID and APOLO transcripts [1].

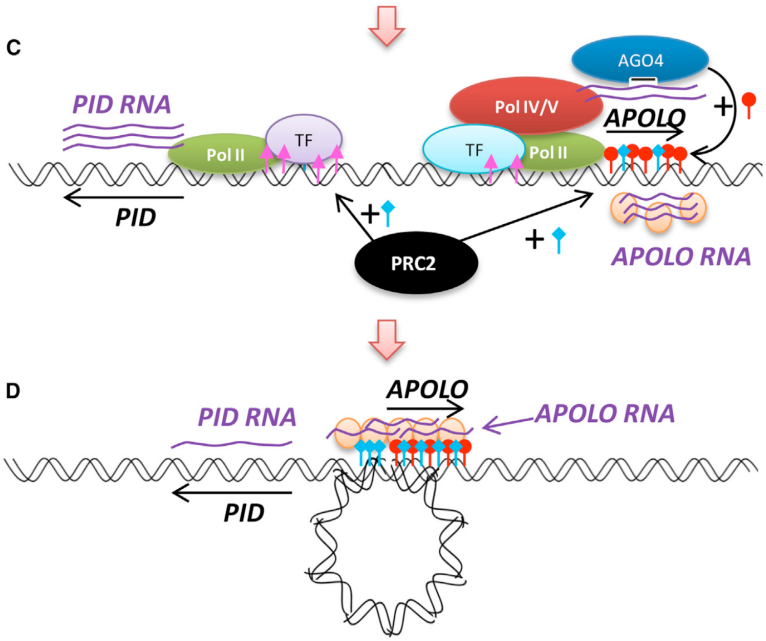
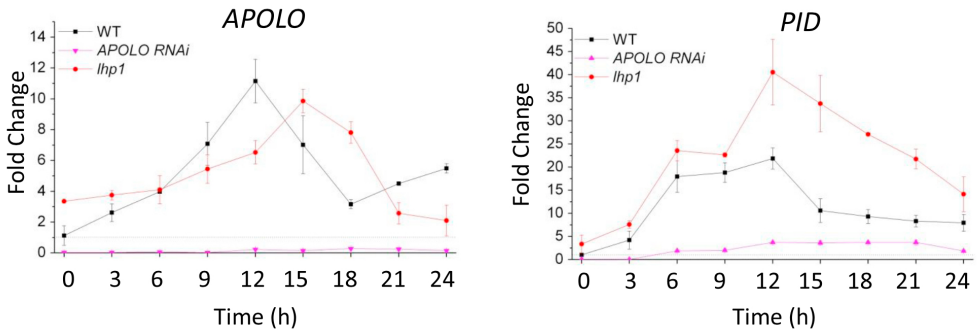



Figure 2.2: **APOLO Transcription by Alternative Polymerases Controls Chromatin Loop Dynamics to Fine-Tune PID Promoter Activity (second half)**
(C) Pol II APOLO transcripts gradually recruit LHP1 to mediate loop formation, whereas Pol IV/V transcription triggers DNA methylation (red symbols), and PRC2 likely redeposits repressive marks in the region. (D) Pol II APOLO-LHP1 mediated loop is finally conformed and maintained by Pol IV/V-dependent DNA methylation to down regulate levels of PID transcripts [1].

## Conclusions

Summarizing the results, auxin treatment was shown to activate the RDD demethylation pathway on the APOLO locus leading to loop opening, triggering divergent Pol II transcription of the two genes and recruiting Pol V on the APOLO locus. Subsequently accumulation of APOLO transcripts triggers siRNA-mediated DNA methylation and at the same time LHP1 is recruited by Pol II transcripts for efficient loop formation, ensuring down regulation of both PID and APOLO (figure 2.3). Finally, the H3K27me3 repressive mark and LHP1-chromatin interactions (which may be mediated by APOLO RNA) on loci suggest that PRC1 and PRC2 complexes contribute to loop closure.

Yet, once the biological problem is presented, some questions naturally arise. What are the properties of this genetic circuit and how such a dy-




(a) (b)

Figure 2.3: APOLO (fig. 2.3a) and PID (fig. 2.3b) transcript levels in wild-type (WT), an RNAi-APOLO line, and *lhp1* mutant seedlings in response to a 24 hr auxin time course. Error bars represent the standard deviation of three biological replicates. [1]

namics is related to the genetic circuit structure in particular with respect to auxin perturbations? What is the role of basal APOLO and PID transcriptions, if any? The goal of this thesis is to answer these questions by the aid of theoretical modelling and mathematical analysis and validate it by the aid and comparison to the experimental results illustrated in this chapter.

# Chapter 3

# Mathematical modeling

Given the biological introduction so far, we will now introduce the aim of this thesis project. The goal is to build a mathematical model based on the observations collected by Ariel et al. as well as the current biological knowledge to understand quantitatively PID-APOLO gene regulation. In the first section, we will introduce the theory of gene regulatory networks which gives all tools used to translate the biological observations and hypotheses into mathematical equations and thus to concretely formulate the model. The second and third sections will explain how dynamical systems theory allows us to better understand the model and extract information from it.

## 3.1 Gene Regulatory Networks and modeling

In the 1960s, genetic and biochemical experiments demonstrated the presence of regulatory sequences in close proximity to genes. It was also shown that there are some specific proteins that can bind these regulatory regions and control, by activating or repressing, the expression of the respective genes.
However, these proteins are themselves encoded by genes, thus creating a complex network of regulatory interactions. As explained in the first chapter, there are many different pathways of gene regulation that also act at different stages of gene expression thus further complicating the picture.
In recent years, the increased availability of gene expression data, accompanied by the flourishing of network theory (which allows the analysis of complex systems that made of many non-trivially interacting elements), has stimulated the use of mathematical and computational tools to understand and model the complicated regulatory processes in cells. For these reasons, *Gene Regulatory Networks* have attracted a lot of interest and many methods have been introduced for their statistical inference from gene expression

data.
Based on these premises, the following is an interesting definition that puts us in a perspective where data is the focus:

**Definition.** *We call a network that has been inferred from gene expression data a "gene regulatory network," briefly denoted as GRN [30].*

The survival of an organism requires that its cells continuously monitor the environment and calculate the amount of protein needed. They are able to sense many different signals from the environment, for example temperature and pressure, but also nutrients or harmful chemicals, and the production of the right proteins is the cell's internal response. Putting ourselves in this other perspective, however, a GRN can be defined as the set of all the interactions between molecular species that govern gene expression patterns by regulating the levels of mRNAs and proteins that are critical for the proper functioning of the mentioned cells.
As anticipated, these interactions in the network can be of various types: gene-gene, protein-gene, protein-protein (e.g., in a complex) and many others. They also occur at different levels of the gene expression process.
All this variety, multiple molecular interactions, and multiple levels of interaction lead to the fact that, in general, the gene regulatory networks within cells are extremely large and function in nontrivial ways.
To build a model that represents them, it is important to ask some questions. Is it really necessary to take all this complexity into account? An important point is therefore to identify which interactions are of real interest for the case studied. Moreover, as it turns out, gene regulation occurs on many different scales. Identifying the scales to work on can lead to a great simplification of the system under consideration, turning it into something more manageable. In addition, in the large GRNs inside the cells not every component has to be considered as fundamental for contributing to a cell's specific function.
Usually the role is given to some subgraphs, which contain a small part of the network components and their interactions. These, in a more or less complicated way, can be modeled as circuits and are called *"gene regulatory circuits"*. An example of a simple circuit is represented in figure 3.1.

### 3.1.1 Modeling gene circuits

Since the 1960s, methods drawn from mathematics and physics have been used to more rigorously describe and simulate small gene networks.
Various theoretical models have been formulated to explain how gene expression occurs and to try to understand numerous hidden aspects in this regard. There is always a tension between the generality and the level of detail of a model: depending on the scale involved and the nature of the

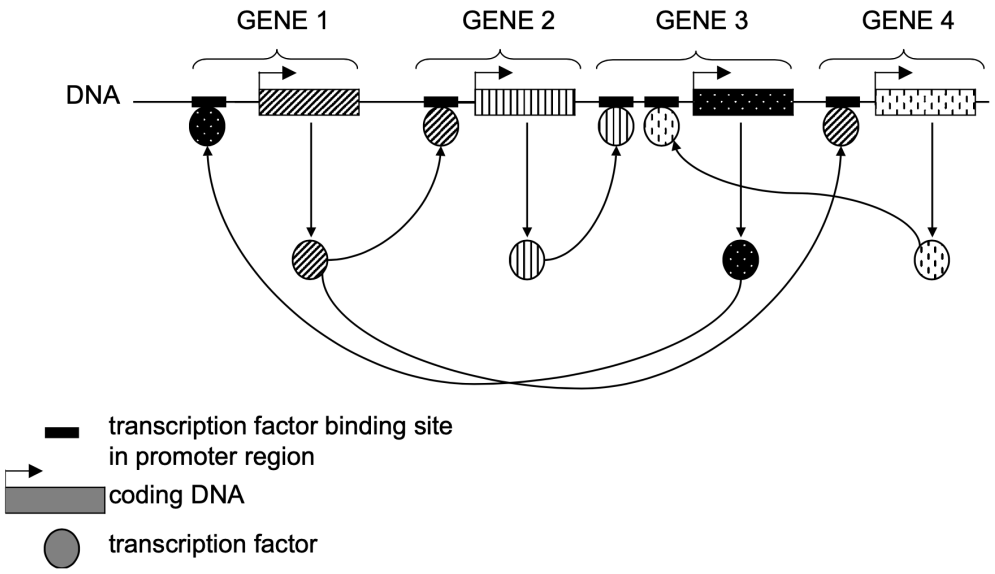



Figure 3.1: **Representation of a simple, fictional transcription factor network.** All genes shown encode transcription factors that control the activity of genes encoding transcription factors [31].

available information, an appropriate mathematical framework must be chosen. In addition, situations are usually more complicated than expected and may also involve a wide range of interactions and scales, so it may even be necessary to use a combination of different approaches.
One way to classify models can be done according to increasing level of detail. According to [32], it can be distinguished between: i) *parts lists*, a collection, description and systematization of network elements in a biological system (e.g., transcription factors, promoters, and transcription factor binding sites); ii) *topology models*, a description of the connections between the parts (this can be viewed as diagrams where directed or undirected connections between genes represent different types of interactions); and going towards a higher level of detail: iii) *dynamic models*, the simulation of the real-time behavior of the network and the prediction of its response to environmental changes, external, or internal stimuli. The following will explain how to build a detailed model step by step, paying more attention to topology and dynamic models.

### Topology models:

Once we know all the interacting molecular species (i.e., the *parts list* of the system), we can increase the level of detail and describe the regulatory network/circuit by graphs with nodes and edges. In this kind of models, the nodes correspond to genes, messengers or proteins and the edges represent

regulatory interactions (activations, inhibitions or formation of complexes) among the network components.
For the purpose of this thesis, it is important to introduce the following types of interactions: i) activation, ii) repression, iii) degradation, iv) basal transcription and v) complex formation.
The simplest case of activation, as well as repression, is that of transcription factors (TFs) that bind to the gene promoter by regulating its expression. The effect of TFs can be to increase the rate of transcription (in this case we speak of *activator*) or to decrease it (*repressor*). Taking this as an example, consider an edge connecting nodes $A$ to $B$ (genes in this case): it means that the product of gene $A$ can bind the promoter of gene $B$ by influencing its transcription. The interactions between $A$ and $B$ are not only represented by an arrow, but a sign is also required. In the case of activation, the interaction is assigned the sign $(+)$ and is represented by a regular arrow $A \longrightarrow B$, while in the other case the sign is $(-)$ and the interaction is represented by a blunt-tipped arrow $A \dashv\!\!\!\rightarrow B$. Usually a TF acts primarily in one mode for its target gene, as an activator or repressor, whereas a gene can be activated by several TFs and repressed by others.
Figure 3.2 shows a simple example of a two-gene topology model with both positive and negative regulatory interactions. That being said, we can now consider the fact that molecular species are usually subject to degradation, which may be, for example, related to an average lifetime of the molecule. This event is usually considered with a common arrow that starts from the species but points to the empty set: $A \longrightarrow \emptyset$ , which means that $A$ is subjected to degradation.
Also, important to consider for the purpose of this thesis is the basal production of certain RNAs, proteins, etc. This process can be due to multiple factors, from basal transcription of a gene, to the constant external injection of a molecular species into the cell. This phenomenon is usually represented in models with an arrow indicating the species that has basal activity: $\longrightarrow A$. Finally, let us take into account the interaction between two molecular species that leads to the creation of a complex. This is represented with an undirected edge between the species: $A - B$.

**Dynamic models:**

The knowledge of the *parts list* of a network and its topology are necessary requirements in order to expand the model to capture dynamical changes during time.
This further step in the modeling process aims at describing and often simulating the dynamic changes in the state of the system and predicting the network's response to various environmental changes and stimuli. Recalling the biological questions posed in the previous chapter, it finally becomes

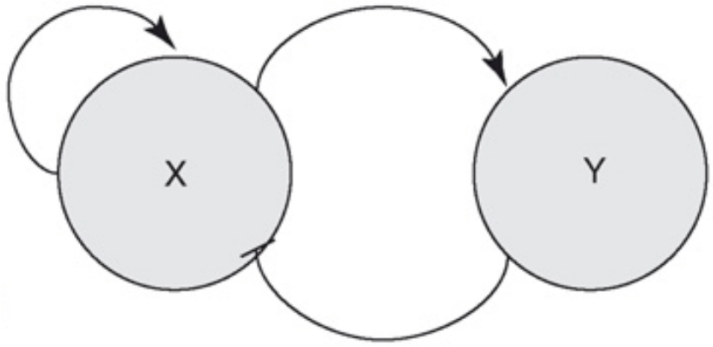



Figure 3.2: **Example of a simple two-gene GRN dynamic model.** Two ($X$ and $Y$) element GRN with positive (arrowhead edges) and negative (flat end edges) regulatory interactions [31].

clear why these kinds of models are necessary for our goal.
Again, the type of problem being considered, the available data, and many other factors can influence the choice between the different approaches that go by the name of dynamic modeling: quantitative versus qualitative, deterministic (e.g., differential equations) versus probabilistic (e.g., Markov theory), continuous versus discrete, and many others.
For the purpose of this thesis and the analysis of the gene regulatory circuit considered in the second chapter, we will consider continuous and deterministic models based on ordinary differential equations (ODEs). These kinds of models allow more detailed descriptions of network dynamics, by explicitly modelling the concentration changes of molecules over time. The result will be a system of ODEs of the following type:

$$\begin{cases} \frac{d}{dt}X_1(t) = f_1(X_1, ..., X_n, \vec{\theta}) \\ \frac{d}{dt}X_2(t) = f_2(X_1, ..., X_n, \vec{\theta}) \\ ... \\ \frac{d}{dt}X_n(t) = f_n(X_1, ..., X_n, \vec{\theta}) \end{cases} . \tag{3.1}$$

In this system, each equation contains in the LHS the time derivative of the concentration of a molecular species $X_i(t)$, while in the RHS there is a function $f_i\left(X_1, ..., X_n, \vec{\theta}\right)$ of the concentrations of the other molecular species and a vector of parameters $\vec{\theta}$. In fact, differential models depend on numerical parameters that are often difficult to measure experimentally but can be estimated by inference methods. The $f_i(...)$ function, which is often very complicated, represents the set of interactions between molecular species that lead to the variation in time of the corresponding component $X_i$ of the network. The question that arises in this context is: how should this

function be defined? How can the interactions be expressed in mathematical terms? The next section is entirely devoted to answering these questions.

## 3.2 From the dynamics of GRNs to ODEs

So far, we have briefly discussed how gene circuits and, more generally, GRNs are schematized thanks to the use of graphs. The next step is the translation into mathematical language of all the processes involved in gene regulation. This will allow the methods of mathematics and in particular dynamical systems theory to be used to explore in depth and in a more quantitative way the phenomenon under consideration. In particular, the general process of formulating *ordinary differential equations* (ODEs) representing the dynamics of the gene regulatory network will be made explicit in this section.

### 3.2.1 Input functions

Before starting with the actual construction of differential equations, it is important to spend some time on the introduction of *input functions*. Indeed, input functions are a key step in understanding how to mimic the dynamics of a gene circuit with mathematical tools. The complexity of the topic dictates that an entire section of this chapter be devoted to it.
We already considered the case in which two components of a gene circuit, $A$ and $B$, influence each other production: the strength of this effect can be quantified using an input function $f(A)$. The interaction $A \longrightarrow B$ is modeled through the regulatory function representing the concentration of products of $B$ per unit time as a function of the concentration of $A$ (from now on it will be implied in the formulas that molecular species are given in terms of concentration). Intuitively, $f(A)$ must be a monotonically increasing function if $A$ is an activator and decreasing in case $A$ is a repressor. Summarizing, we can write:

$$\textit{rate of production of}\, B = f(A).$$

To better grasp the mathematical meaning of these objects, let us start with a simple example: *logical input functions*.

**Logical input functions:**

Often used as an approximation for the input functions, the essence of these functions is that the gene is either OFF $f(A) = 0$ or maximally ON

$f(A) = \beta$ where $\beta$ is the *maximum promoter activity*. There is an activation threshold $K$ such that for activators the input function is:

$$f(A) = \beta \cdot \Theta(A > K),$$

while for repressors:

$$f(A) = \beta \cdot \Theta(A < K)$$

[33]. It is clear that the parameter $\beta$ is a rate of change and is expressed in concentration over time.

**Hill input functions:**

Another notable example of an input function, often used to describe many cases, is the Hill function. Although beyond the scope of this chapter, the *Hill function* can be derived by considering the equilibrium binding of the transcription factor to its site on the promoter. In the case of an *activator* it is characterized by three parameters:

- $K$: *activation coefficient* that defines the concentration of $A$ required to significantly activate the expression. More precisely, the half-maximal expression rate is reached by the condition $A = K$;
- $\beta$: *maximal promoter activity* reached when $A \gg K$;
- $n$: *Hill's coefficient* representing the steepness of the input function; the higher $n$, the steeper the Hill function is.

The functional form is as follows:

$$f(A) = \beta \frac{A^n}{K^n + A^n} \text{ Hill function for an activator.} \tag{3.2}$$

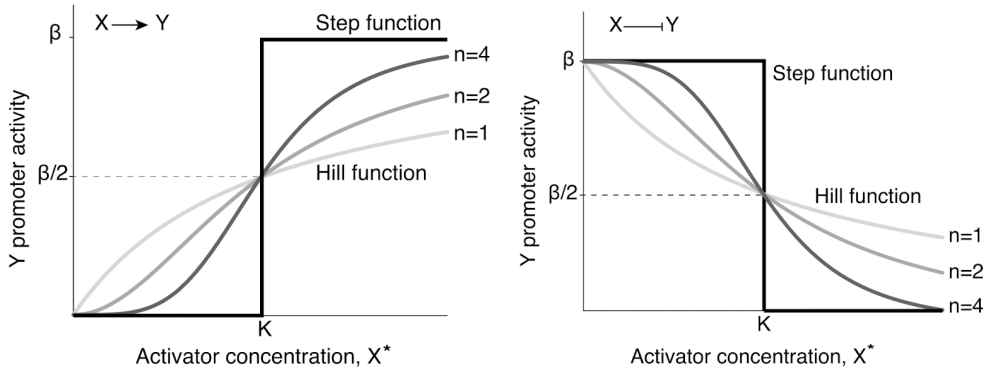
For a repressor, as already explained, the input function must be monotonically decreasing. The formulation of the Hill function in this case is as follows:

$$f(A) = \beta \frac{K^n}{K^n + A^n} \text{ Hill function for a repressor.} \tag{3.3}$$

It is important to note that in the last formula, the *maximal promoter activity* is reached for $A = 0$ while $f(A) = 0$ only if $A$ is really large. Again, half of the maximal expression rate is reached for $A = K$.
Figure 3.3 shows an example of increasing and decreasing Hill function as $n$ varies. Drawing conclusions, three parameters can be assigned to each arrow of a transcription network: $\beta$, $K$ and $n$. $K$, for example, can be modulated by changes in the DNA sequence of the binding site or if its position is changed, while $\beta$ may depend on mutations in the RNAp[13] binding site or

[13]RNA polymerase (RNAp) is an enzyme that synthesizes RNA from a DNA template.

other various factors [33].
Note that while $n$ is an integer and $K$ is expressed in concentration, $\beta$ in this case must be expressed in the unit such that the right-hand side represents a rate (concentration per unit time).




(a) (b)

Figure 3.3: Increasing (3.3a) and decreasing (3.3b) Hill function as the Hill's coefficient changes [33].

### Basal expression and multivariate input functions:

Usually the input function ranges from a zero transcription rate to a maximum basal expression, but sometimes it can happen that the gene has a nonzero minimum expression called *basal expression level*. Basal expression can be considered in the mathematical representation by adding a constant term $\beta_0$.
It is also important to underline that a gene can be regulated by many transcription factors, in other words, a network node can have more than one input arrow. Therefore, the input function must be a multivariate function. Three examples are presented below, using the logical approximation discussed earlier. We first consider the case in which two activators must bind the promoter to initiate transcription, much like an AND gate:

$$f(A_1, A_2) = \Theta(A_1 > K_1) \cdot \Theta(A_2 > K_2) \sim A_1 \text{ AND } A_2. \tag{3.4}$$

The second case is where binding of one of the two activators (similar to an OR gate) is sufficient:

$$f(A_1, A_2) = \Theta(A_1 > K_1 \text{ OR } A_2 > K_2) \sim A_1 \text{ OR } A_2. \tag{3.5}$$

Finally, when both activators can act cooperatively to activate the gene, an

SUM input function is often used in this case:

$$f(A_1, A_2) = \beta_1 A_1 + \beta_2 A_2 \tag{3.6}$$

[33].Note that in this example the unit of $\beta_1, \beta_2$ is the inverse of time.
Clearly other kind of function are also possible.

### 3.2.2 Building ordinary differential equations

By constructing, step by step, an exemplary differential equation that mimics the dynamics of the system, this section aims to explain the basic rules needed to transform biological hypotheses into mathematical ODEs.

To facilitate the reader's understanding, we will refer to a simple example: $A \longrightarrow B$ (i.e., TF $A$ activates gene $B$ in response to a signal from the environment). When the external signal is absent, $A$ is inactive and $B$ is not produced. As soon as the signal appears, $A$ is activated and binds the promoter of $B$. The latter begins to be transcribed by RNAp, its mRNA is produced and subsequently translated into the $B$ protein. From the previous section, we know that the whole process from mRNA transcription to protein translation is summarized by the input function $f(A)$, which gives the rate of production of protein $B$.
We can then write that the change per unit time of the concentration of protein $B$ is proportional to its production rate, i.e., the input function, as follows:

$$\frac{d}{dt}B = f(A). \tag{3.7}$$

This is not the only process involved in the dynamics of $B$ concentration, in fact, it is also important to consider degradation. The *degradation term* represents the loss in concentration of the $B$ protein that can be due to a variety of factors. Some examples may be the natural lifetime of a molecule in the cell or the process of RNAi silencing. For simplicity, we consider only the lifetime of the protein within the cell and define the generic *degradation rate* $\delta$ as:

$$\delta = \frac{1}{\tau}, \tag{3.8}$$

where $\tau$ is the *average lifetime of the molecule.*
It is now possible to write the full version of the equation 3.8, subtracting

the degradation term given by the product of the degradation rate and the concentration of B:

$$\frac{d}{dt}B = f(A) - \delta \cdot B. \tag{3.9}$$

This is the basic equation that forms the building block of ODE models, but the case of basal transcription and complex formation still remains to be modeled.
As already discussed in the section devoted to input functions, the case of basal transcription can be taken into account adding to equation 3.9 a constant rate (concentration over time) $\beta_0$ as follows:

$$\frac{d}{dt}B = \beta_0 + f(A) - \delta \cdot B.$$

If we start instead concerning the formation of complexes, the situation becomes more complicated. Let us consider equation 3.9 and let us add the possibility that the products of $A$ and $B$ form a complex $C$ following the chemical equation:

$$A + B \underset{K^{(\text{off})}}{\overset{K^{(\text{on})}}{\rightleftharpoons}} C, \tag{3.10}$$

where $K^{(\text{on})}$ and $K^{(\text{off})}$ are the coefficient related to the creation and separation of the complex $C$. These represent the probability of the reaction occurring per unit time and unit particle.
If the *mass action law* is assumed valid, these constants do not depend on the number of particles. Based on this law and assuming that $B$ and $C$ are subjected to the generic fluxes of concentration (incoming $\nu_b^+$,$\nu_c^+$ and outgoing $\nu_b^-$,$\nu_c^-$), the equations governing the change in concentrations of $A$,$B$ and $C$ can be written as:

$$\begin{cases} \frac{d}{dt}A(t) = f(A) \ - \delta \cdot B \ - K^{(\text{on})}A \cdot B + K^{(\text{off})}C \\ \frac{d}{dt}B(t) = \nu_b^+ \ - \nu_b^- \ - K^{(\text{on})}A \cdot B \ + K^{(\text{off})}C \\ \frac{d}{dt}C(t) = \nu_c^+ \ - \nu_c^- \ + K^{(on)}A \cdot B \ - K^{(\text{off})}C \end{cases} . \tag{3.11}$$

In the previous equations the terms: $K^{(\text{on})} \cdot AB$ and $K^{(\text{off})} \cdot C$ are the fluxes of the chemical reactions respectively direct and inverse.
Finally, at the conclusion of this section, to make things more rigorous, some remarks need to be made. In what follows, we will assume that the system is normally in its *stationary state* in which the concentration of its components is constant and stable. When an external signal reaches the cell, the system responds, and the steady state is reached again only after

some time after the signal has disappeared. Consider a simple example in which only basal expression is present:

$$\frac{d}{dt}B = \beta_0 - \delta \cdot B. \tag{3.12}$$

In the stationary state, the constant concentration level is represented by

$$\frac{d}{dt}B_{ss} = 0 \iff \beta_0 - \delta \cdot B_{ss} = 0 \iff B_{ss} = \frac{\beta_0}{\delta}. \tag{3.13}$$

The external signal leads the system to leave the steady state, but how is it reached again? Suppose the system is out of equilibrium and that the signal ceases at time $t = 0$, then:

$$B(t) = B_{ss} - C_0 \frac{e^{-\delta \cdot t}}{\delta}. \tag{3.14}$$

The steady state is reached again with an exponential decay $e^{-\delta \cdot t}$. $C_0$ is a coefficient depending on the initial conditions.

## 3.3 Theory of dynamical systems

As anticipated, this section will explain how dynamical systems theory can be used to study those systems of equations (e.g., equation 3.1) that represent gene regulatory networks. We will first briefly introduce the definition of a *ordinary differential equation.*

**Definition.** *an* ***ordinary differential equation*** *of order n is an equation*

$$\frac{d^n}{dt^n}y = F\left(t; y, \frac{d}{dt}y, ..., \frac{d^{n-1}}{dt^{n-1}}y\right),$$

*where F is a differentiable function defined in a domain U of a space of dimension $n + 1$ [34].*

In this project we will be dealing with many coupled equations, the so-called system of ODEs.

**Definition.** *By a system of ordinary differential equations we shall mean a system of equations in n unknown functions:*

$$\frac{d^{n_i}}{dt^{n_i}}y_i = F_i\left(t; y, ...\right), \quad i = 1, ...n,$$

*where the arguments of each function $F_i$ are the independent variable t, the dependent varibles $y_j$, and their derivatives of orders less than $n_j$ ($j = 1, ..., n$) respectively [34].*

For the purposes of this thesis and the modeling of regulatory circuits, it is sufficient to consider systems of first-order ordinary differential equations.

**Solutions of ODE systems:** Given the initial point, the solution of these systems is not always simple. An ODE can have no solution, a unique solution, or even an infinite number of solutions. Under given assumptions there are many theorems that prove the existence and uniqueness of the solution, which can be found either locally or globally. Two examples may be the *Peano's existence theorem* or the *Cauchy-Lipschitz theorem*, which prove the uniqueness and existence of the solution [34].
Very often, these systems become too complicated to solve analytically; instead, some numerical methods can be used to integrate the equations numerically. This will be the case in this thesis.

### 3.3.1 Stationary points

Let us return to biology for a moment: it has been said that the transcription network at rest is assumed to be in a stationary state. This state is the one achieved by the system after a very long time and to which it tends to return after an external signal has perturbed the cell.
The stationary condition in biology can be translated into mathematical terms by imposing that the time derivative of the concentrations in the system is zero:

$$\frac{d}{dt}\mathbf{y} = 0. \tag{3.15}$$

This is exactly what is called the *stationary point* of a differential equation.

**Definition.** *The point* $\bar{\mathbf{y}} \in \mathbb{R}^n$ *is an stationary point for the differential equation*

$$\frac{d}{dt}\mathbf{y} = \mathbf{F}(t, \mathbf{y})$$

*if* $\mathbf{F}(t, \bar{\mathbf{y}}) = 0$ *[34].*

**Physical conditions of the stationary states:** As explained in the definition, the mathematical stationary can take any value on the real axis. Obviously, the same cannot apply to the physical values of the concentrations in the system. Only an stationary point belonging to the positive real semi-axis should be considered as a biological steady state. This condition is of fundamental importance for the results to be interpretable.

### 3.3.2 Analysis of the stability

Given the respective ODEs and a set of initial conditions, the system begins to follow a deterministic orbit in phase space, completely dependent on the above conditions. This is exactly how the mathematical model mimics the behavior of the real system, that is, how the gene circuit reacts to the external signal. For these reasons, the study of the orbit provides a lot of information and especially a quantitative analysis of the biological problem under consideration. Many types of orbits can be found in the theory of dynamical systems: from orbits that converge to a stationary point to orbits that escape from it, to periodic orbits and many others. Several criteria have been developed to tell whether an orbit belongs to one type or another. In particular, to understand whether an orbit is *stable* or *unstable*. In the interest of this thesis, we will mainly discuss what is known as *local stability*, with a small introduction to *global stability* for completeness.
Let us consider an autonomous system[14] $\dot{\mathbf{y}} = \mathbf{F}(\mathbf{y}(t))$ with $\mathbf{F}(\bar{\mathbf{y}}) = 0$ ($\bar{\mathbf{y}}$ stationary point):

**Definition.** *An stationary point $\bar{\mathbf{y}}$ is **orbitally stable** if for every neighborhood $\mathcal{O}$ of $\bar{\mathbf{y}}$ in $\mathbb{R}^n$ there is a neighborhood $\mathcal{O}_1$ of $\bar{\mathbf{y}}$ in $\mathcal{O}$ such that every solution $\mathbf{y}(t)$ with $\mathbf{y}(0) = \mathbf{y}_0 \in \mathcal{O}_1$ is defined and remains in $\mathcal{O}$ for all $t > 0$ [35].*

**Definition.** *If $\mathcal{O}_1$ can be chosen above so that, in addition to the properties for stability, we have $\lim_{t\to\infty} \mathbf{y}(t) = \bar{\mathbf{y}}$, then we say that $\bar{\mathbf{y}}$ is **asymptotically stable** [35].*

In other terms, a stationary point is *orbitally stable* if any orbit will stay indefinitely close to it, is *asymptotically stable* if any orbit converges to the point, but can also be *unstable*.

**Definition.** *An stationary point $\bar{\mathbf{y}}$ that is not stable is called **unstable**. This means there is a neighborhood $\mathcal{O}$ of $\bar{\mathbf{y}}$ such that for every neighborhood $\mathcal{O}_1$ of $\mathbf{y}$ with $\mathcal{O}_1 \subset \mathcal{O}$, there is at least one solution $\mathbf{y}(t)$ starting at $\mathbf{y}(0) \in \mathcal{O}_1$ that does not lie entirely in $\mathcal{O}$ for all $t > 0$ [35].*

Figure 3.4 shows the phase portrait of a 2D stable and an unstable system, in the generic variables $y_i$ and $y_j$. In the case considered in this project, the steady state of the biological system coincides with an asymptotically stable point, assuming it is in the positive real semi-axis.

[14] In mathematics, an autonomous system or autonomous differential equation is a system of ordinary differential equations which does not explicitly depend on the independent variable. When the variable is time, they are also called time-invariant systems. (In this thesis the biological systems studied do not depend explicitly on time, we will then use autonomous ODEs).

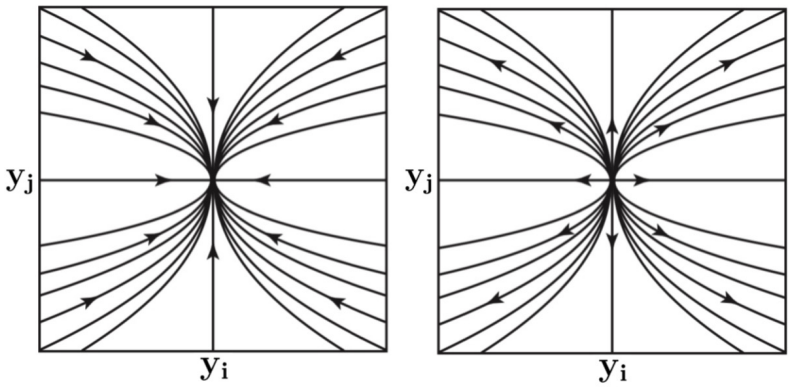



Figure 3.4: From left to right, respectively, a phase portrait of a 2D system with one orbitally stable and one unstable singular point [35].

### Local stability

Local stability concerns what happens to an orbit in the immediate vicinity of an stationary point.
Let us consider the equation $\dot{\mathbf{y}} = \mathbf{F}(\mathbf{y})$ and suppose $\mathbf{F}(\bar{\mathbf{y}}) = 0$ ($\bar{\mathbf{y}}$ stationary point). Let us assume to be close enough to $\bar{\mathbf{y}}$ such that it is possible to expand the function $\mathbf{F}$ around $\bar{\mathbf{y}}$, up to the first order. This allows us to reduce the system to its linear part: $\dot{\mathbf{y}}(t) \simeq \mathbf{J}|_{\bar{y}} \cdot (\mathbf{y} - \bar{\mathbf{y}})$. The latter with a simple change of coordinates $\mathbf{y} = \bar{\mathbf{y}} + \mathbf{z}$ can be re-written as: $\dot{\mathbf{z}}(t) \simeq \mathbf{J}|_{\bar{z}} \cdot \mathbf{z}$. In the last expression $\mathbf{J}|_{\bar{y}}$ is the Jacobian matrix[15] of the function $\mathbf{F}$ evaluated in the stationary point.

Thanks to this procedure, the analysis of system stability has been reduced to the study of the stability of a much simpler *linear* system. Local stability is now completely determined by the eigenvalues $\lambda$ of the Jacobian matrix:

- if $Re(\lambda) < 0 \ \ \forall\lambda$ then the system is asymptotically stable;
- if $\exists\lambda$ such that $Re(\lambda) > 0$ then the system results to be unstable for the local analysis.

It is important to pay attention to the fact that the same properties are also true for the complete system[16], but only near the stationary point [34, 35].

[15]The Jacobian matrix of a vector-valued function of several variables is the matrix of all its first-order partial derivatives.

[16]Complete system here means the system before the linearization procedure.

**Imaginary part of the eigenvalues:** The real part of the eigenvalues of the Jacobian is not the only thing that matters. It is critical to understand the linear stability of the system, but it says nothing about how the orbit converges (or diverges) to the stationary point. It may be the case that the orbit simply converges to a stable point, but in some cases the orbit may begin to oscillate around the equilibrium and reach it after a while. This is the case when the eigenvalues of the Jacobian matrix have a nonzero imaginary part. Depending on the stability of the stationary point, the system may form the so-called *spiral sink* or *source*.

#### Global stability

Local stability is often used for its simplicity, but it has the inherent limitation of being reliable only in the immediate vicinity of the stationary point. What happens far from it? Again, there are many methods for analyzing the global stability of a system, which under certain assumptions can tell whether or not a point is stable for global dynamics. An important example is the use of the *Lyapunov function*, which is a continuous, differentiable function of points in phase space whose analysis provides global stability. Global stability is usually complicated to study analytically, especially when dealing with complicated systems. For this reason, we will rely on local stability analysis in the thesis.

### 3.3.3 Linear response analysis

Linearization of the system not only provides information about stability near the stationary point, but can be incredibly informative about the response to infinitesimal perturbations of the system. This is particularly useful for our purpose because the system under consideration has been perturbed through auxin treatment.
In addition, linear response analysis provides important information about the *resilience* of a system. *Biological resilience* in fact refers to the ability of a system to return to its initial state after being deformed or perturbed. A quantitative measure of resilience in a model can be given by the time it takes the system to return to its initial state after the disappearance of a perturbation. This concept can be further explored through response analysis.
Let us suppose that for a given auxin level $A_0$ the system reaches its (unique) steady state. Let us call $\Delta\mathbf{C}$ the vector containing the deviations of the concentration in the system from the steady-state values. The linearization already seen for local stability leads to the following differential equation for $\Delta\mathbf{y}$:

$$\frac{d}{dt}\Delta\mathbf{y} = \mathbf{J}|_{\bar{y}} \cdot \Delta\mathbf{y}. \tag{3.16}$$

Now we introduce a perturbation with respect to the previous auxin level $A(t) = A_0 + \Delta A(t)$. The latter will lead to the following modification of 3.16:

$$\frac{d}{dt}\Delta\mathbf{y} = \Delta\mathbf{S} + \mathbf{J}|_{\bar{y}} \cdot \Delta\mathbf{y}. \tag{3.17}$$

Where $\Delta\mathbf{S}$ is the vector containing the perturbation of auxin concentration, taking into account the equations of the ODE system considered [17]. Now, assuming that the system is at steady state before the application of the perturbation, we can write:

$$\Delta\mathbf{y} = \int_{-\infty}^{t} \chi(t - t') \cdot \Delta\mathbf{S}\,(t')\,, \tag{3.18}$$

where $\chi$ is the matrix:

$$\chi(t - t') = e^{J|_{\bar{y}} \cdot (t - t')}, \tag{3.19}$$

$\chi$ is known as *linear response function* and is assumed to be $\chi\,(t) = 0$ in the range $t < t'$.

How can be the function $\chi$ interpreted? Let us consider the example case in which the auxin perturbation appears only in one equation of the ODE system and is represented by a Dirac delta pulse at time $t_{pulse}$ as follows:

$$\Delta\mathbf{S} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \delta(t_{pulse}) \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.20}$$

.

Injecting the perturbation into equation 3.18 gives:

[17] When considering a perturbation of a system, it is not easy to determine which equations are actually affected and which are not. It is possible that all are perturbed or, as in the case of the system analyzed in this thesis, that only one is.

$$\Delta \mathbf{y}(t) = \chi(t - t_{pulse}) = e^{J|_{\bar{y}} \cdot (t - t_{pulse})}. \tag{3.21}$$

In other words the linear response function represents the value of $\Delta \mathbf{y}(t)$, i.e., the variation of $\mathbf{y}$ at time $t$ after applying a delta function pulse as perturbation [36].

# Chapter 4

# Models formulation and analysis

In this last section we will address the formulation and analysis of some models that attempt to capture the nature of the phenomenon under consideration. We will begin by considering only the minimum ingredients to build a model as simple as possible; in the second section, however, the simple formulations will be expanded to better capture the dynamics.

## 4.1 Minimal models

Returning to the biological problem, many components participate in the entire dynamics of the system. In this section we will try to reduce the complexity of the real system as much as possible by considering only the main actors in the transcription network.

### 4.1.1 Simple model

The first model formulated is an attempt to see if the patterns of regulation shown in the experiments can be summarized with a few simple principles and assumptions.

**Ingredients:**

The two key ingredients without which comparison with experimental data would be difficult are clearly the two RNA transcripts: PID and APOLO. As Ariel and collaborators demonstrated in the 2014 article [1], they are co-regulated in response to auxin and lead to loop opening. In modeling, this intermediate process between auxin treatment and PID-APOLO expression, was simplified by considering what we called *"euchromatin state"* $\epsilon$, which represents how compact chromatin is. The value of $\epsilon$, which varies in the range $[0, 1]$ (where $\epsilon = 0$ indicates a fully heterochromatic DNA and $\epsilon = 1$ a fully euchromatic one), is affected initially by auxin flux into the system.

**Interactions:**

As mentioned earlier, the RDD demethylation pathway plays an important role in opening the loop in response to auxin concentration. It was hypothesized that its action simply increases euchromatin state in a manner proportional to a parameter $\beta_\epsilon$. To account for the time required for RDD to actually change the chromatin state, $\beta_\epsilon$ was multiplied by the auxin concentration but delayed in time by a constant $\tau$: $A_\tau(t) = A(t-\tau)$.
Auxin concentration in time was assumed to be a decreasing exponential function:

$$A_\tau(t) = A_0 + C_A \cdot e^{-\frac{t-\tau}{\tau_A}},$$

where $C_A$ is a coefficient while $\tau_A$ is one over the mean lifetime of auxin in the system.
To ensure that the euchromatin state extends into the range $[0, 1]$, the increase of $\epsilon$ was set to be proportional to $(1-\epsilon)$ (which clearly measures how heterochromatic the DNA is). Proportional to $\epsilon$, through an input function $F(\epsilon)$, but with different maximal promoter activity ($\beta_a$ and $\beta_p$) APOLO and PID are transcribed. In the analysis conducted in the following sections, however, the input function was varied between linear function and Hill function with different parameters to study the different behaviours.
The natural lifetime of PID and APOLO was considered through the degradation terms: $\delta_a$ and $\delta_p$ and the basal transcription of APOLO through $\beta_0$. Finally, the action of the RdDM pathway was assumed to act directly on the state of euchromatin $\epsilon$ and proportionally to the concentration of APOLO. This assumption clearly simplifies the model, skipping the role of siRNA, which will be considered later. The model can be written as follows:

$$\begin{cases} \frac{d}{dt}[APOLO](t) = \beta_0 + \beta_a F(\epsilon) - \delta_a[APOLO] \\ \frac{d}{dt}[PID](t) = \beta_p F(\epsilon) - \delta_p[PID] \\ \frac{d}{dt}\epsilon(t) = \beta_\epsilon A_\tau(t)\,(1-\epsilon) - \delta_{RdDM}[APOLO]\epsilon \end{cases} . \tag{4.1}$$

**Analysis:**

Now that the model has been formulated, the next step is to take advantage of what has already been explained in chapter three to better understand its capabilities.

**Parameter reduction:** Although the system of ODEs considered is simple, it can be further simplified. In fact, by rescaling some quantities, some parameters can be eliminated, thus facilitating the study of the model. In this case, we considered the new quantities: $[\widetilde{APOLO}] = [APOLO]/\beta_a$ and

$[\widetilde{PID}] = [PID]/\beta_p$. Finally, by rescaling the parameters $\beta_0$ and $\delta_{RdDM}$ accordingly, it is possible to write:

$$\begin{cases} \frac{d}{dt}[\widetilde{APOLO}](t) = \tilde{\beta}_0 + F(\epsilon) - \delta_a[\widetilde{APOLO}] \\ \frac{d}{dt}[\widetilde{PID}](t) = F(\epsilon) - \delta_p[\widetilde{PID}] \\ \frac{d}{dt}\epsilon(t) = \beta_\epsilon A_\tau(t)\,(1-\epsilon) - \tilde{\delta}_{RdDM}[\widetilde{APOLO}]\epsilon \end{cases} . \tag{4.2}$$

Thanks to this easy trick it was possible to eliminate two parameters ($\beta_a$ and $\beta_p$) from the system.

**Stationary states:** The analysis of the model continues with the study of stationary states and their stability. First, it is necessary to find the points that satisfy the steady-state condition:

$$\begin{cases} \frac{d}{dt}[\widetilde{APOLO}](t) = 0 \\ \frac{d}{dt}[\widetilde{PID}](t) = 0 \\ \frac{d}{dt}\epsilon(t) = 0 \end{cases} . \tag{4.3}$$

Solving the system of equations, it can be seen that not all three equations are significant. By first solving the steady states of PID ($[\widetilde{PID}]_{ss}$) and APOLO ($[\widetilde{APOLO}]_{ss}$) and then substituting the last equation, it is possible to reduce the system to the solution of a *"self-consistency"* equation for $\epsilon_{ss}$:

$$\epsilon_{ss} = \frac{\beta_\epsilon A}{\beta_\epsilon A + \delta_\epsilon \frac{F(\epsilon_{ss})+\tilde{\beta}_0}{\delta_a}}. \tag{4.4}$$

It is worth noting that equation 4.4 must be accompanied by the condition that all concentrations are greater than zero. The solution will provide a value for $\epsilon_{ss}$ by which $[\widetilde{\text{PID}}]_{ss}$ and $[\widetilde{\text{APOLO}}]_{ss}$ can also be calculated. By inverting the rescaling it is possible, in the end, to obtain $[\text{PID}]_{ss}$ and $[\text{APOLO}]_{ss}$.
By fixing the degradation parameters (we will take care to set reasonable values for the parameters in the calibration of the model) and varying the parameters $\beta_\epsilon A$ and $\tilde{\beta}_0$, it was possible to solve the system 4.3 by the procedure just explained. The solution found was unique and always acceptable[18]. As expected, the steady-state value of euchromatin $\epsilon_{ss}$ decreases with $\beta_0$ and increases proportionally with $\beta_\epsilon A$. Higher basal transcription of APOLO resulted in higher levels of APOLO transcripts in the steady state, directly affecting chromatin state. In contrast, the increase $\beta_\epsilon A$[19] leads to more

[18]For the condition to be acceptable, the stationary state must be positive or equal to zero.

[19]$\beta_\epsilon A$ can be seen as a *"perceived auxin flux"*.

open loops and a higher value of $\epsilon_{ss}$. Finally, the stationary value of PID follows the behavior of euchromatin. Indeed, it is obvious that the more euchromatic the DNA is, the higher the production of PID and vice versa if $\epsilon_{ss}$ decreases.

**Analysis of the stability:** Regarding stability, local dynamics were analyzed. As explained in the corresponding section, in this case it is essential to study the eigenvalues of the Jacobian matrix evaluated at the stationary point. Continuing to work with the reduced-parameter model, it is possible to write the entire Jacobian matrix as a function solely of $\epsilon_{ss}$:

$$J(\epsilon_{ss}) = \begin{pmatrix} -\delta_a & 0 & \frac{d}{d\epsilon}F(\epsilon_{ss}) \\ 0 & -\delta_p & \frac{d}{d\epsilon}F(\epsilon_{ss}) \\ -\tilde{\delta}_\epsilon \epsilon_{ss} & 0 & \beta_\epsilon A - \tilde{\delta}_\epsilon \frac{F(\epsilon_{ss})+\tilde{\beta}_0}{\delta_a} \end{pmatrix}. \tag{4.5}$$

As is evident from the ODEs, the dynamics of the PID is completely dependent on the other two. This is reflected in this matrix, which clearly has an always negative (stable) eigenvalue. Taking advantage of this observation, the Jacobian matrix can be simplified into a $2 \times 2$ matrix:

$$J_{\text{reduced}}(\epsilon_{ss}) = \begin{pmatrix} -\delta_a & \frac{d}{d\epsilon}F(\epsilon_{ss}) \\ -\tilde{\delta}_\epsilon \epsilon_{ss} & \beta_\epsilon A - \tilde{\delta}_\epsilon \frac{F(\epsilon_{ss})+\tilde{\beta}_0}{\delta_a} \end{pmatrix}. \tag{4.6}$$

As before, the behavior was studied by fixing the degradation parameters and varying $\beta_\epsilon A$ and $\tilde{\beta}_0$. The two eigenvalues, always complex conjugate, have negative real part for each choice of the two varied parameters, confirming the stable nature of the steady state. Anticipating future analysis, this will be a feature preserved in all the models we have studied, which might suggest that it is intrinsic to this type of models. The imaginary part of the eigenvalues, almost zero everywhere, showed a nonzero value only for small values of $\beta_\epsilon A$ and $\tilde{\beta}_0$, indicative of oscillatory behaviour. In the end, different input functions (linear and Hill with $n = 1, 2, 3$) resulted only in different forms of the previously mentioned region.

**Linear response analysis:** In addition to the stability, we are interested in the linear response of the system in the steady-state to a perturbation of auxin concentration over time: $A(t) = A_0 + \Delta A(t)$. Where $A_0$ is the value of the auxin concentration (on which the steady state was built) and $\Delta A(t)$ is the perturbation. In other words, we assume that the system is in the steady state when the perturbation appears. Let us start with the linearized equations: $\frac{d}{dt}\Delta\mathbf{y}(t) = J|_{ss} \cdot \Delta\mathbf{y}(t)$ where

$$\frac{d}{dt}\Delta\mathbf{y}(t) = J|_{ss} \cdot \Delta\mathbf{y}(t) \text{ where } \Delta\mathbf{y}(t) = \begin{bmatrix} \Delta[APOLO](t) \\ \Delta[PID](t) \\ \Delta\epsilon(t) \end{bmatrix}. \tag{4.7}$$

In our specific case, the perturbation $\Delta\mathbf{S}$ in the equation 3.18 acts only in the equation for the euchromatin state, leading to:

$$\Delta\mathbf{S}(t) = \begin{bmatrix} 0 \\ 0 \\ \beta_\epsilon\left(1 - \epsilon(t)\right) \cdot \Delta A(t) \end{bmatrix}. \tag{4.8}$$

Once the perturbation is written, the linear response can be calculated as explained above:

$$\Delta\mathbf{y}(t) = \int_{-\infty}^{t} dt' \chi(t - t') \cdot \begin{bmatrix} 0 \\ 0 \\ \beta_\epsilon\left(1 - \epsilon(t')\right) \cdot \Delta A(t') \end{bmatrix}, \tag{4.9}$$

where $\chi(t - t') = \exp\left[J(\epsilon_{ss}) \cdot (t - t')\right]$.
The reader will recall that the auxin perturbation is delayed in time by an amount $\tau$, this is necessary to account for the effect of the RDD demethylation pathway. This delay causes no problem in defining the linear response; the only effect is to shift the response in time by $\tau$. Based on this reasoning, the response analysis can be carried out regardless of the delay: the reader should interpret the relative graphs as shifted in time by $\tau$.
It was already mentioned above that the analysis was repeated with different activation functions: linear activation and Hill activation with $n = 1, 2$; and the results are explained in the following.
As already evident from the system of equations, PID and APOLO share the same behavior and their response is completely superposed. The Hill function leads the response to have a larger amplitude than the linear function in the $n = 1$ case. The Hill input function $n = 2$ makes things more subtle if $A_0 = 0$. In the graphs of figure 4.1, it can be seen that the linear response of PID and APOLO is zero in this case: this is because the second derivative of the Hill function with $n = 2$ has $\epsilon$ at the numerator, but when calculated in the steady state with $A_0 = 0$, where $\epsilon_{ss} = 0$, the corresponding term goes to zero and leads to the observed behaviour.

**Code:** All the study explained so far has been done using the programming language Wolfram Mathematica. Integration of the equation was done

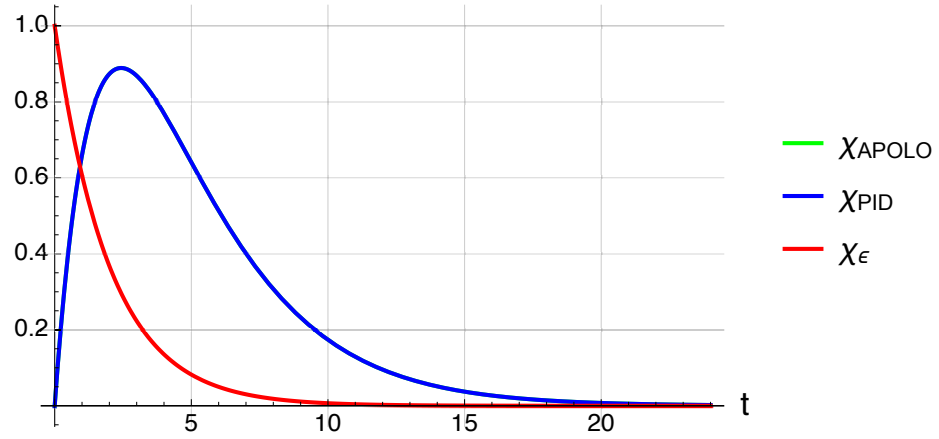



Figure 4.1: **Linear response of the minimal model.** PID and APOLO share the same dynamics and their linear responses are completely superposed. (Parameters: $\delta_a = \delta_\epsilon = 1/3$ and $\beta_\epsilon = \beta_0 = 1$ and $A_0 = 0$).

numerically by means of the function **NDSolve[ ]** placing the system at time zero at the stationary point. The latter was found by solving the system of equations 4.3 with **NSolve[ ]** with the conditions $[\widetilde{\text{APOLO}}], [\widetilde{\text{PID}}], \epsilon \geq 0$. Specifically, the values of $\tilde{\beta}_0$ and $\beta_\epsilon A$ were made to vary and, for each combination of the two parameters, the system was solved with **NSolve[ ]**. The same procedure was used for the Jacobian matrix and eigenvalues computed with the integrated function **Eigenvalues[ ]** [37].

### 4.1.2 Adding siRNA

The toy model presented so far seems to be going in the right direction to reproduce the experiments. Despite its simplicity, it is able to capture the main features of PID and APOLO regulation in A. thaliana. The opening of the loop is well reproduced by the rapid increase of $\epsilon$ and its subsequent decay. Loop opening allows transcription of PID and APOLO, which are subsequently degraded, generating the oscillating gene expression measured by Ariel and coworkers [1]. Despite inaccuracies as in the PID-APOLO responses that superpose, the results obtained with the toy model were encouraging and prompted us to continue in this direction. Once the basic structure of the model is defined, the next step is to add more details to make it more realistic. In the case considered so far, an important approximation was that of APOLO directly influencing the euchromatin state. As explained in the third chapter, the actual role is played by the small interfering RNA produced by APOLO. The introduction of siRNA into the model should make it as simple as possible, however, a new equation must be added to our ODE system to account for it.

$$\begin{cases} \frac{d}{dt}[APOLO](t) = \beta_0 + F(\epsilon) - \delta_a[APOLO] \\ \frac{d}{dt}[siRNA](t) = \beta_s[APOLO] - \delta_s[siRNA] \\ \frac{d}{dt}[PID](t) = F(\epsilon) - \delta_p[PID] \\ \frac{d}{dt}\epsilon(t) = \beta_\epsilon A_\tau(t)\,(1-\epsilon) - \delta_{RdDM}[siRNA]\epsilon. \end{cases} \tag{4.10}$$

siRNA is produced proportionally to [APOLO], through a constant parameter $\beta_s$ and is simply degraded with $\delta_s$. The main difference from the previous version is that now $\epsilon$ is no longer degraded proportionally to [APOLO] but to [siRNA]. Intuitively one would expect that since siRNA is produced only after APOLO transcription, this additional step in the process would lead to delayed or reduced loop closure.

**Comparison:** The analysis already done in the previous section was also repeated here and the results were compared. Again it is possible to reduce the number of parameters and simplify the ODE system. The single stationary point turns out to be always stable, with oscillating behavior in some regions of the $\beta_\epsilon A$-$\beta_0$ plane. The overall behavior of both models seems to be very similar, the main differences appear clear in the analysis of the linear response (figure 4.2). First, contrary to expectation, the decay of the euchromatin state from 1 to 0 after auxin perturbation is faster in the presence of siRNA. This is probably because the effective steady-state level reached by siRNA is higher than that of APOLO, contributing to faster loop closure. At the same time, the greatest effect occurs in the expression of PID and APOLO. Although they are still overlapping, their linear response is reduced by a factor of 4 due to the addition of the small interfering RNA. The latter result compared with the experimental data does not give any further information, it is clear that to improve the model it is necessary to differentiate the dynamics of PID and APOLO.

### 4.1.3 RNAi silencing APOLO

So far we have discussed many ways in which genes and transcripts are regulated, some of which were highlighted in Ariel's article [1] to act in A. thaliana. With the aim of adding more and more features to the model to make it more realistic, siRNA was introduced in the previous section. The role of siRNA is not only to act in methylation, but also to lead to the degradation of some RNAs. In this section, RNA-interference leading to APOLO degradation will be added to the model.
Intuitively, the decay of APOLO due to siRNA should depend on the concentrations of both APOLO and siRNA; to keep things as simple as possible,

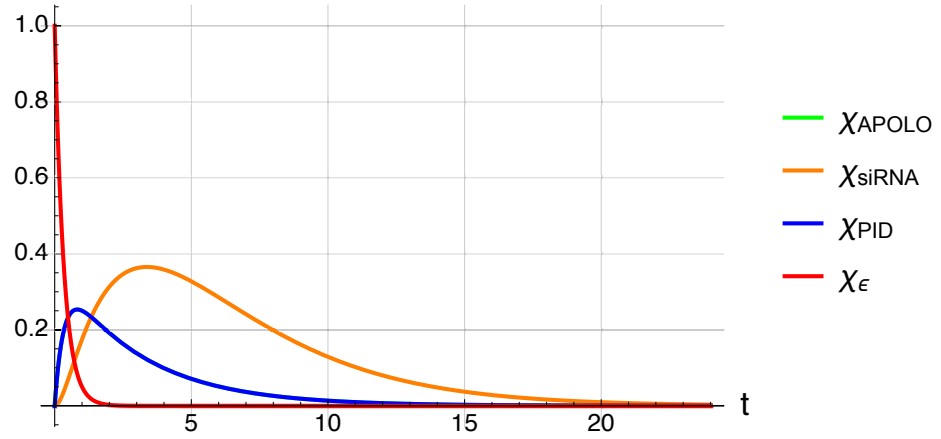



Figure 4.2: **Linear response of the toy model with siRNA.** (Parameters: $\delta_a = \delta_s = \delta_\epsilon = 1/3$ and $\beta_\epsilon = \beta_0 = 1$ and $A_0 = 0$).

we will consider it proportional to the product of $[APOLO] \cdot [siRNA]$. The term $\delta_{s,a} \cdot [APOLO] \cdot [siRNA]$ can be added in the ODEs as follows:

$$\begin{cases} \frac{d}{dt}[APOLO](t) = \beta_0 + F(\epsilon) - \delta_a \cdot [APOLO] - \delta_{s,a} \cdot [APOLO] \cdot [siRNA] \\ \frac{d}{dt}[siRNA](t) = \beta_s \cdot [APOLO] - \delta_s \cdot [siRNA] \\ \frac{d}{dt}[PID](t) = F(\epsilon) - \delta_p \cdot [PID] \\ \frac{d}{dt}\epsilon(t) = \beta_\epsilon \cdot A_\tau(t) \cdot (1 - \epsilon) - \delta_{RdDM} \cdot [siRNA] \cdot \epsilon \end{cases} . \tag{4.11}$$

**Analysis and comparison:** As usual, the model parameters were reduced as much as possible. The steady state found is always stable, and all dynamic characteristics are similar to previous models. Studying the linear response reveals some important differences. Due to the RNAi term, the dynamics of the two neighboring genes are differentiated:

- linear input function: The response of APOLO remains quite similar, while that of PID is considerably larger in amplitude. Moreover, APOLO reaches its peak earlier than PID;
- Hill function ($n = 1$): PID response is still much higher than that of APOLO, but in this case both are of much smaller amplitude than in the linear case;
- Hill function ($n = 2$): because of the greater complexity of this model than the previous version, the case of the Hill function $n = 2$ does not have a null answer. In fact, it resembles the answer of $n = 1$.

This small change in the equations pushes the system toward a more realistic behavior, similar to that of the experiments. This difference between the expression of PID and APOLO is actually seen in the measurements: as can be seen in figure 4.3, the fold change of APOLO peaks lower than that of PID. Clearly, RNA-interference reduces the expression of APOLO, which is less able to induce loop closure. The result is that PID transcripts are more concentrated and peak with some delay compared with APOLO.

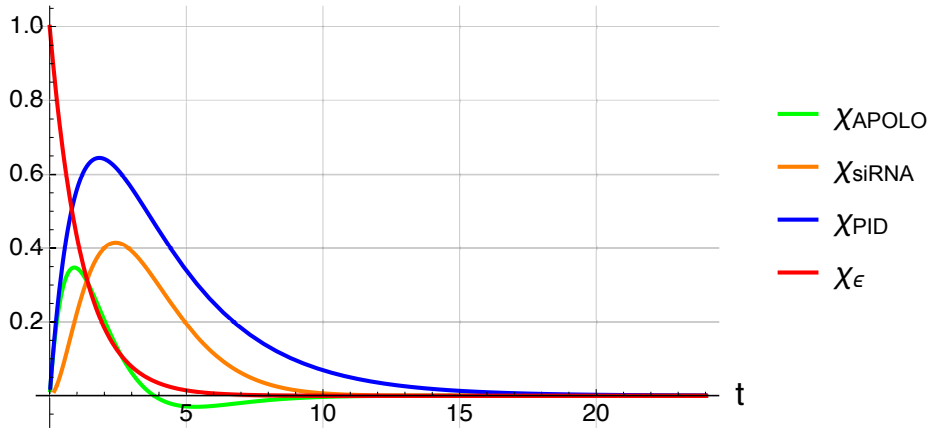



Figure 4.3: **Linear response of the toy model with RNA-interference.** (Parameters: $\delta_a = \delta_s = \delta_\epsilon = \delta_{sa} = 1/3$ and $\beta_\epsilon = \beta_0 = 1$ and $A_0 = 0$).

To clarify this difference, we studied the phase portrait of the dynamical system and compared it with the previous one. The system is placed in the steady state at $A_0 = 0$ and then treated, for a finite time interval, with a higher concentration of auxin: $A_0 + \Delta A$. The system begins to move toward the *transient state* (red circle) corresponding to the new level of auxin concentration and, once the perturbation time $\tau_s$ has elapsed, returns to the original point, which we have shown to be always stable.
Comparing the phase portrait in the $\epsilon$-$[APOLO]$ plane, it is clear how regardless of the $\Delta A$ perturbation the dynamics on the APOLO axis is incredibly reduced in the RNAi case. The same behavior is observed for the dynamics of [siRNA] in the $\epsilon$-$[siRNA]$ plane.

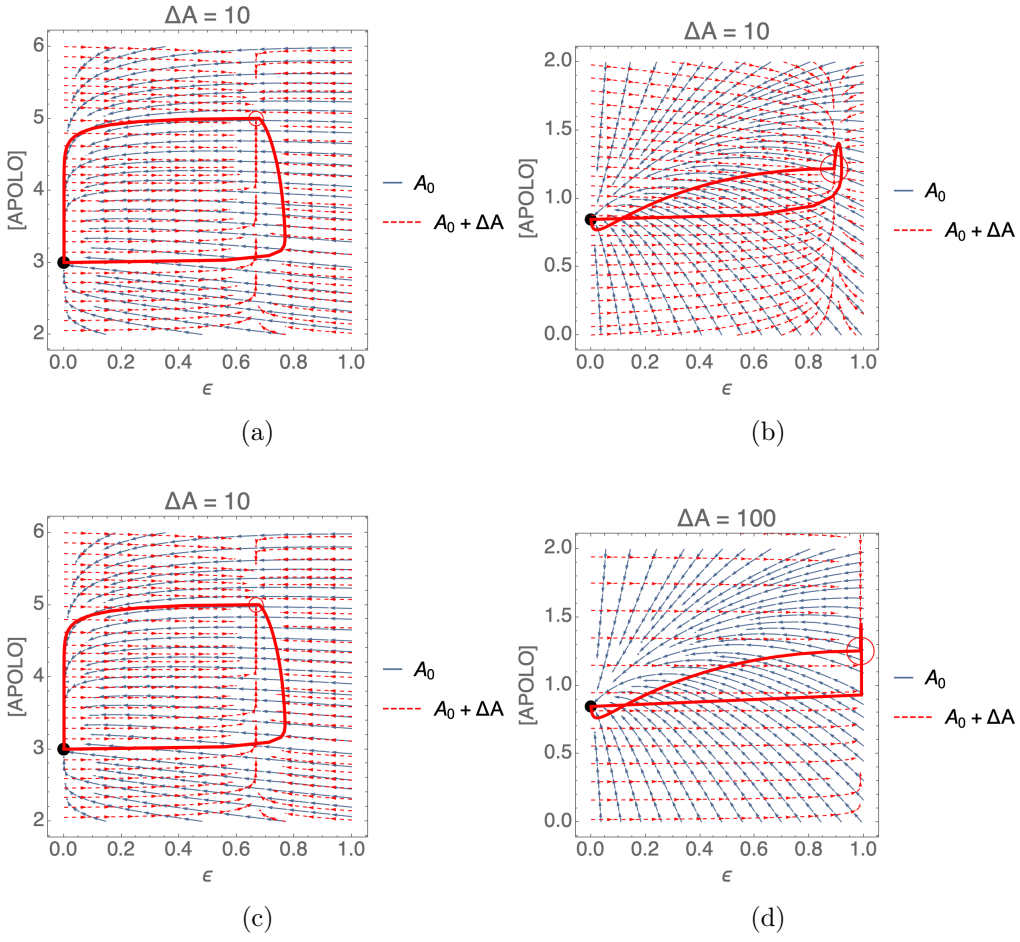



Figure 4.4: **Phase portrait comparison.** Phase portrait in the plane $[APOLO]$-$\epsilon$ rispectively: i) in the first formulation (addition of siRNA), with $\Delta A = 10$ (4.4a) and $\Delta A = 100$ (4.4c) and ii) in the second formulation (addition of RNA-interference), with $\Delta A = 10$ (4.4b) and $\Delta A = 100$ (4.4d). Note that the orbits cross in the graphs on the right since the phase portraits are projected onto the plane in which $[siRNA] = [siRNA]_{ss}$.

## 4.2 Full models

The toy models studied so far encourage us to think that the path taken may be the right one. The next step is then to expand the model, trying to include all the biological features that contribute to make it more realistic and able to reproduce experimental measurements. In this section we will begin by presenting the first formulation of the complete model and then, step by step, all the modifications and corrections needed to improve it.

### 4.2.1 First formulation

The first step from the toy model to the more complete one was made by keeping as much of the old structure as possible. Clearly [APOLO], [PID], and [siRNA] were retained, but also the general concept of *euchromatin state*, which seems to work well in representing the DNA state that triggers gene transcription.

**LHP1:** The fundamental addition to the model is the PRC[20] component LHP1, which has not been considered so far but has multiple roles like binding DNA as well as APOLO transcripts. Let us construct step by step the equation for $\frac{d}{dt}[LHP1]$ starting with its fundamental component and continuing with the terms corresponding to the various interactions involving [LHP1].
Its concentration was considered to grow over time with a constant basal transcription $\beta_l$ and degrade with the term $\delta_l \cdot [LHP1]$, defining a first equation of the type:

$$\frac{d}{dt}[LHP1](t) = \beta_l - \delta_l \cdot [LHP1]. \tag{4.12}$$

**LHP1 complexes:** As pointed out in the 2014 article [1], LHP1 is recruited for efficient loop formation and, together with APOLO transcripts, plays a role in precluding Pol II from accessing the PID promoter. Given that we are maintaining the general idea of a euchromatin state ($\epsilon$), the simplest way to model this process is to imagine that LHP1 and APOLO form a complex with a generic species representing chromatin. The fraction of chromatin bound to APOLO and LHP1 influences the state of euchromatin, reducing the $\epsilon$ value. In other words, the LHP1-APOLO-Chromatin complex ($LAC$) leads the system toward more heterochromatic DNA, encouraging loop formation.
What is the dynamics of this molecular complex formation? To turn this idea into equations it is necessary to make some assumptions. We will

[20]Polycomb Repressive Complex.

assume that first APOLO RNA binds chromatin and that only after this event can the APOLO-Chromatin complex ($AC$) be bound by LHP1. We assume that the formation of the two types of complexes is regulated by the following chemical reactions:

$$[APOLO] + [\text{free chromatin}] \underset{K^{\text{off}}_{AC}}{\overset{K^{\text{on}}_{AC}}{\rightleftharpoons}} [AC], \tag{4.13}$$

$$[AC] + [LHP1] \underset{K^{\text{off}}_{LAC}}{\overset{K^{\text{on}}_{LAC}}{\rightleftharpoons}} [LAC], \tag{4.14}$$

where $K^{\text{on}}_{AC}$, $K^{\text{off}}_{AC}$ and $K^{\text{on}}_{LAC}$, $K^{\text{off}}_{LAC}$ are the coefficients related to the creation and degradation of $AC$ and $LAC$, respectively.
How does $LAC$ affect the state of euchromatin in the equation? As already assumed for the RdDM pathway, the $LAC$ complex makes DNA more heterochromatic (reducing $\epsilon$), proportionally to a parameter $\delta_{LAC}$. Finally, as pointed out in the article [1], LHP1 is able to bind APOLO transcripts but not PID transcripts. In this first comprehensive model, this interaction was also considered on the basis of the following chemical equation:

$$[APOLO] + [LHP1] \underset{K^{\text{off}}_{LA}}{\overset{K^{\text{on}}_{LA}}{\rightleftharpoons}} [APOLO\text{-}LHP1], \tag{4.15}$$

with $K^{\text{on}}_{LA}$, $K^{\text{off}}_{LA}$ as the creation/degradation coefficients.
Before introducing the equations, it is necessary to emphasize an important concept. The total chromatin concentration in the region where APOLO can bind can be assumed constant, let us call it $[C_{tot}]$. As assumed in the previous lines, this total amount of chromatin can be in three different conditions: free, bound to APOLO ($AC$) or bound to both APOLO and LHP1 ($LAC$). By simply calling $C_{free}$, $AC$ and $LAC$ the fractions of chromatin in the three respective states, we can write:

$$C_{free} + AC + LAC = 1.$$

This means that we can write the concentration of free chromatin as: $[C_{free}] = (1 - AC - LAC)\,[C_{tot}]$ and do the same reasoning for $[AC] = AC[C_{tot}]$ and $[LAC] = LAC[C_{tot}]$. In order to simplify the equations and refer only to chromatin fractions ($C_{free}$, $AC$, $LAC$), we will take advantage of the fact that $[C_{tot}]$ is constant and absorb it in the coefficients: $\hat{K}^{\text{on}}_{AC} = K^{\text{on}}_{AC}[C_{tot}]$, $\hat{K}^{\text{off}}_{AC} = K^{\text{off}}_{AC}[C_{tot}]$, $\hat{K}^{\text{on}}_{LAC} = K^{\text{on}}_{LAC}[C_{tot}]$, $\hat{K}^{\text{off}}_{LAC} = K^{\text{off}}_{LAC}[C_{tot}]$.

Once all the appropriate assumptions have been made, the complete model equations can be written:

$$
\begin{cases}
\dfrac{d}{dt}[APOLO] = \beta_0 + \beta_a F(\epsilon) - \delta_a[APOLO] - \delta_{s,a}[APOLO][siRNA] + \\
\qquad\qquad -K^{\text{on}}_{LA}[APOLO][LHP1] - \hat{K}^{\text{on}}_{AC}[APOLO]\,(1 - AC - LAC) + \\
\qquad\qquad +\hat{K}^{\text{off}}_{LA}[LHP1\text{-}APOLO] + \hat{K}^{\text{off}}_{AC} AC \\
\\
\dfrac{d}{dt}[siRNA] = \beta_s[APOLO] - \delta_s[siRNA] + \beta_{RNAi} \\
\\
\dfrac{d}{dt}[PID] = \beta_p F(\epsilon) - \delta_p[PID] + \beta_{p,0} \\
\\
\dfrac{d}{dt}[LHP1] = \beta_l - \delta_l[LHP1] - K^{\text{on}}_{LA}[LHP1][APOLO] + \\
\qquad\qquad +K^{\text{off}}_{LA}[LHP1\text{-}APOLO] - \hat{K}^{\text{on}}_{LAC} AC[LHP1] + \hat{K}^{\text{off}}_{LAC} LAC \\
\\
\dfrac{d}{dt}[LHP1\text{-}APOLO] = K^{\text{on}}_{LA}[LHP1][APOLO] - K^{\text{off}}_{LA}[LHP1\text{-}APOLO] \\
\\
\dfrac{d}{dt} AC = K^{\text{on}}_{AC}[APOLO]\,(1 - AC - LAC) - K^{\text{off}}_{AC} AC + \\
\qquad -K^{\text{on}}_{LAC}[LHP1] AC + K^{\text{off}}_{LAC} LAC \\
\\
\dfrac{d}{dt} LAC = K^{\text{on}}_{LAC}[LHP1] AC - K^{\text{off}}_{LAC} LAC \\
\\
\dfrac{d}{dt}\epsilon = \beta_\epsilon A_\tau(t)\,(1 - \epsilon) - \delta_{RdDM}[siRNA] \cdot \epsilon - \delta_{LAC} LAC \cdot \epsilon
\end{cases}
\tag{4.16}
$$

Note that i) while in the equations for [APOLO] and [LHP1], $[C_{tot}]$ is absorbed in the chemical coefficients, in those for $AC$ and $LAC$ it can be completely simplified; and ii) that PID's basal transcription was introduced with the parameter $\beta_{p,0}$.

**Mutant plants:** As shown in the second chapter, many of the experimental results were obtained with mutants. With this comprehensive model, some cases of mutant plants can also be taken into account. By setting some parameters in certain values, corresponding to the effect of a mutant plant type, we are able to see how the system behaves in those cases and compare it with the experiments. The following are some examples:

- RNAi mutant: by introducing the parameter $\beta_{RNAi}$ as a kind of basal

siRNA transcription it is possible to reproduce WT by setting $\beta_{RNAi} = 0$, while RNA-interference is enhanced by setting $\beta_{RNAi} > 0$;

- *lhp1* mutant: the case where LHP1 is missing can be simulated by setting $\beta_l = 0$;
- *rdd* mutant: the RDD pathway mutant is reproduced thanks to $\beta_\epsilon$. By setting it to zero, the model should reproduce the behaviour of the mutant plant;
- mutant of the RdDM pathway: parameter $\delta_{RdDM}$ was introduced in the simple model to consider the RdDM pathway. It is evident that by eliminating $\delta_{RdDM}$ or, equivalently, by fixing its value to zero, the mutant of the RdDM pathway can be considered.

**Parameters reduction:** The process of reducing the number of parameters was also repeated for the full models. By rescaling some variables it was possible to reduce the number of parameters, although the greater complexity of the ODE system did not allow for much reduction. the result was a decrease from 20 parameters to 17 new parameters.

**Stationary states:** Next step is to find the stationary points of the ODE system. Despite the increased complexity, again it was possible to arrive analytically at a "consistency equation", in particular with a series of substitutions an equation for the unique [APOLO] can be written.
This equation is rather complicated (for this reason is not shown in the text). Therefore, it was solved numerically by assigning fixed values to the parameters. As in the case of the simple models, we let $\beta_\epsilon A$ and $\beta_0$ vary while keeping the other parameters fixed and studied the results. Surprisingly, the complete model also retains the same main feature as the others: there is always an acceptable stationary point. Analysis of the values of the stationary points as a function of the two mentioned parameters showed that:

- as expected $[APOLO]_{ss}$ and $[siRNA]_{ss}$ grow as $\beta_0$ increases;
- as before, $[PID]_{ss}$ is directly proportional to the *"effective auxin flux"* $\beta_\epsilon A$, but it decreases when basal transcription of APOLO increases, probably because of the resulting greater loop closure;
- not surprisingly, the [LHP1-APOLO] complex has a higher steady-state level when the basal transcription of APOLO is increased, whereas the level of [LHP1] idoes not depend on the two parameters: $[LHP1]_{ss} = \frac{\beta_l}{\delta_l}$;

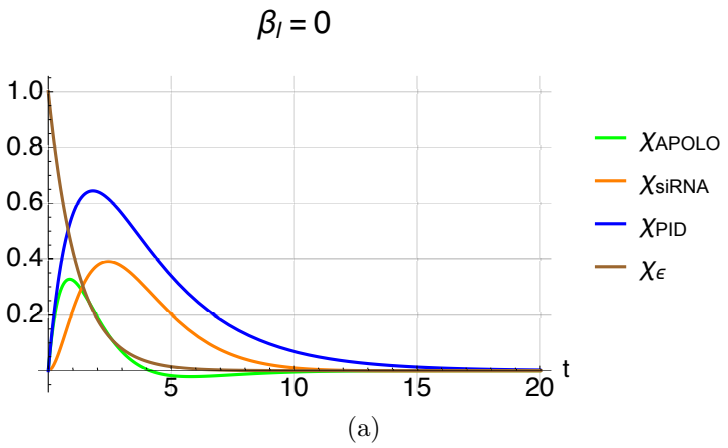
- the levels of the two complexes $AC$ and $LAC$ increase with both $\beta_0$ and $\beta_\epsilon A$.

**Analysis of the stability:** Looking at the local dynamics of this complete system, it is easy to see that here too, the dynamics of PID can be neglected because it is totally dependent on the other molecular species. Consequently, the Jacobian matrix can be written in its reduced form. The matrix is not shown in this section so as not to burden the reading.
Excluding PID, the remaining seven eigenvalues always have a negative real part, reflecting the stability of the stationary point in the linear regime. The imaginary parts of the eigenvalues, often coupled together as complex conjugates, depend on the values of $\beta_\epsilon A$ and $\beta_0$. Some eigenvalues have an imaginary part that is always zero, while others assume a nonzero imaginary part in specific regions of the $\beta_\epsilon A$-$\beta_0$ plane, leading to oscillatory (but always stable) behavior.

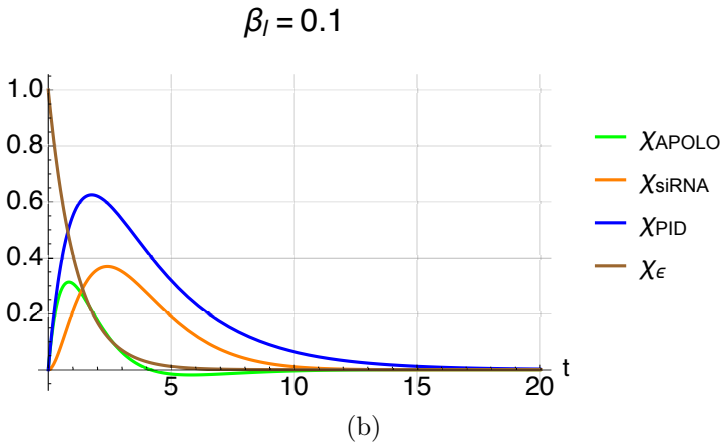
**Linear response analysis:** Regarding the linear response, the model still behaves as in the simplest cases when considering the wild type: the euchromatin state decays rapidly while the PID response is greater than that of APOLO, as shown by experiments. When considering mutants, things are different.
Let us first analyze the RNAi mutant: as explained in the corresponding paragraph, setting $\beta_{RNAi} > 0$ simulates the behavior of the mutant. In fact, linear response analysis shows that by increasing the value of the parameter, the response of PID and APOLO strongly decreases. This is probably because higher levels of siRNA in the cell rapidly degrade APOLO transcripts and lead to loop closure.
In *lhp1* plants the results were also interesting but less encouraging. Although, as expected, the APOLO and PID responses decrease with increasing $\beta_l$, in the mutant case ($\beta_l = 0$,) both APOLO shows no obvious delay in the peak and the PID response remains about the same amplitude, contrary to what was shown in the article.

### 4.2.2 Explicit DNA methylation

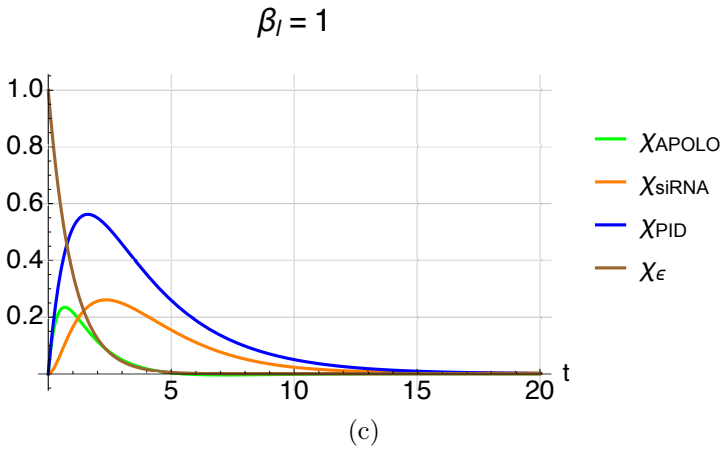
The ODE system analyzed so far has been the first attempt to build a comprehensive model that tries to take into account all the features of the biological system, but it is still an initial version that needs to be improved. One of the crudest approximations that needs to be refined is the representation of DNA and chromatin with the generic *euchromatin state*. In this section we will introduce a new version of the full model that more accurately accounts for DNA dynamics.




(a)




(b)




(c)

Figure 4.5: **Linear response comparison in *lhp1* mutant**. 4.5a, 4.5b and 4.5c represents the cases in which $\beta_l$ assumes respectively the values: 0.0, 0.1 and 1.0. $\beta_l$ is a concentration change rate, i.e., it is defined in units of concentraiton/time.

The most important distinction to be made is between histone binding molecular species (LHP1, APOLO, etc.) and methylated/unmethylated DNA under the effect of the RdDM pathway. In more detail, while before we considered the complexes formed by APOLO and LHP1 with a general concept of chromatin, now it is important to specify these complexes more precisely: Histone-APOLO ($HA$) and Histone-APOLO-LHP1 ($HAL$). Again, they represent the fraction of histones bound to other molecular species, in fact the following equation is true:

$$1 = H_{free} + HA + HAL, \tag{4.17}$$

where $H_{free}$ denotes the fraction of free histones. The same reasoning about chromatin fraction and concentration is valid here for the histones.
As for methylation, on the other hand, DNA can be divided into two fractions: methylated DNA ($D_m$) and unmethylated DNA ($D_{nm}$) satisfying:

$$1 = D_m + D_{nm}. \tag{4.18}$$

The equation 4.18 allows us to equivalently write an equation for methylated or unmethylated DNA. For simplicity, we will use an equation for unmethylated DNA.

**Interactions:** In this paragraph we will explain how the interactions between molecular species were reviewed.

- RdDM pathway: the role of siRNAs should now be explicit in the equation for DNA methylation. The interaction between $D_{nm}$ and [siRNA] through the parameter $\delta_{RdDM}$ increases DNA methylation (in particular, it reduces $D_{nm}$);
- $HAL$ in DNA methylation: the increase of the $HAL$ complex affects DNA methylation. In particular, it reduces unmethylated DNA with a term proportional to both $HAL$ and $D_{nm}$: $-\delta_{HAL} HAL \cdot D_{nm}$;
- $D_{nm}$ influences $HA$ and $HAL$: the greater the level of DNA methylation, the greater the chance that $HA$ and $HAL$ complexes will be formed. As explained earlier, we will use an equation for $D_{nm}$, which will act equivalently with a negative term in the equation for $HA$ and $HAL$ proportional to a new parameter $\beta_m$.

Finally, it is crucial to replace the *euchromatin state* in the transcriptional imput function of PID and APOLO. As pointed out in the 2014 work [1], their transcription is largely influenced by chromatin loop, more specifically

open loops enable transcripts production. In this new model, a suitable indicator for loop opening can be the fraction of free histones:

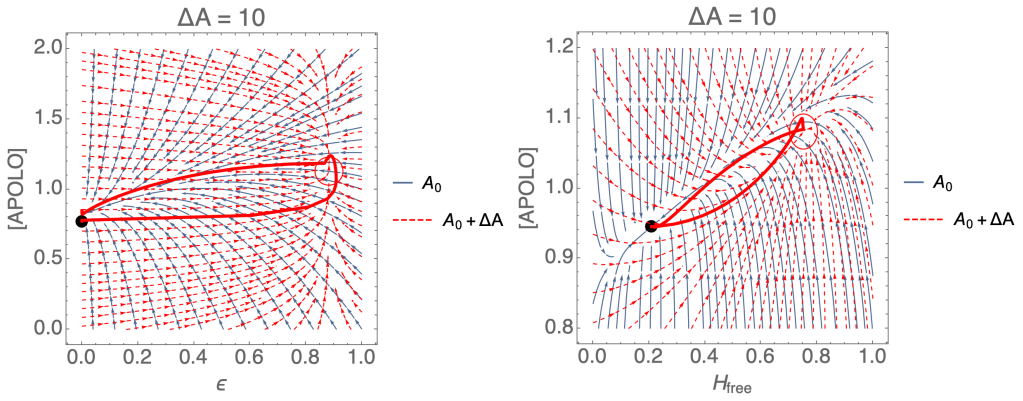
$$H_{free} = 1 - HA - HAL,$$

which will be the choice for the argument of the input function: $F\,(1 - HA - HAL)$. Once the new interactions have been introduced, the ODE system corresponding to this second formulation of the complete model can be written:

$$\begin{cases} \dfrac{d}{dt}[APOLO] = \beta_0 + \beta_a F(1 - HA - HAL) - \delta_a[APOLO] - \delta_{s,a}[APOLO][siRNA] + \\ \qquad -K_{LA}^{on}[APOLO][LHP1] - \hat{K}_{HA}^{\text{on}}[APOLO]\,(1 - HA - HAL) + \\ \qquad +K_{LA}^{\text{off}}[LHP1\text{-}APOLO] + \hat{K}_{HA}^{\text{off}} HA + \beta_m D_{nm}\,(HA + HAL) \\ \\ \dfrac{d}{dt}[siRNA] = \beta_s[APOLO] - \delta_s[siRNA] + \beta_{RNAi} \\ \\ \dfrac{d}{dt}[PID] = \beta_p F(1 - HA - HAL) - \delta_p[PID] + \beta_{p,0} \\ \\ \dfrac{d}{dt}[LHP1] = \beta_l - \delta_l[LHP1] - K_{LA}^{\text{on}}[LHP1][APOLO] + \\ \qquad +K_{LA}^{\text{off}}[LHP1\text{-}APOLO] - \hat{K}_{HAL}^{\text{on}} HA[LHP1] + \hat{K}_{HAL}^{\text{off}} HAL \\ \\ \dfrac{d}{dt}[LHP1\text{-}APOLO] = K_{LA}^{\text{on}}[LHP1][APOLO] - K_{LA}^{\text{off}}[LHP1\text{-}APOLO] \\ \\ \dfrac{d}{dt}HA = K_{HA}^{\text{on}}[APOLO]\,(1 - HA - HAL) - K_{HA}^{\text{off}} HA + \\ \qquad -K_{HAL}^{\text{on}}[LHP1]HA + K_{HAL}^{\text{off}} HAL - \beta_m D_{nm} HA \\ \\ \dfrac{d}{dt}HAL = K_{HAL}^{\text{on}}[LHP1]HA - K_{HAL}^{\text{off}} HAL - \beta_m D_{nm} HAL \\ \\ \dfrac{d}{dt}D_{nm} = \beta_D A_\tau(t)\,(1 - D_{nm}) - \delta_{RdDM}[siRNA] \cdot D_{nm} - \delta_{HAL} HAL \cdot D_{nm} \end{cases} . \tag{4.19}$$

**Analysis and comparison:** Again, parameter reduction is possible, although not as efficient: their number can be reduced by three. The reduced ODE system can then be solved to find the stationary states. The increased complexity of the new version does not allow a single consistency equation to be written for the stationary states. The only possible simplification is

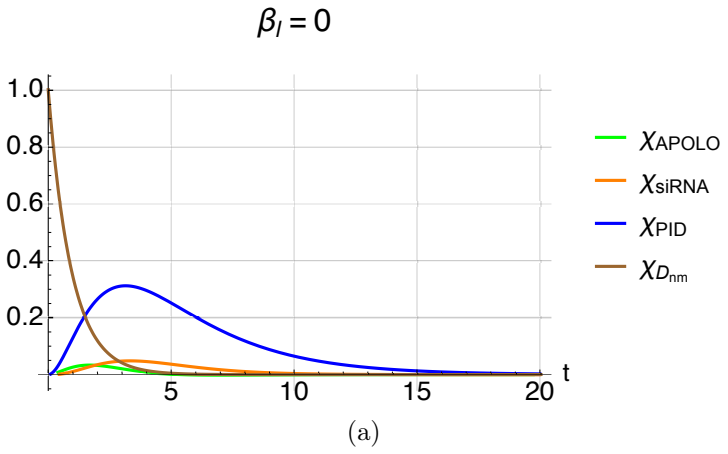
to reorganize the equations so that the problem is reduced to the solution of three equations for [APOLO], [LHP1] and $HAL$.
As usual, these equations were solved numerically, fixing all parameters and varying $\beta_D A$ and $\beta_0$. The stationary point is still unique, and the general results resemble the previous cases. The seven nontrivial eigenvalues indicate always stable dynamics, which takes on oscillatory behavior in some regions. It is worth noting that, unlike the last model, in this case the stationary value of LHP1 concentration is no longer constant as $\beta_D A$ and $\beta_0$ vary.
Regarding the linear response, it is evident that in this new formulation the peaks of [siRNA] and [APOLO] are lowered and delayed in time, while that of PID does not seem to be affected much in its amplitude (figures 4.5c,4.7c). The same observation is confirmed when comparing the phase portraits of the two models (figure 4.6).



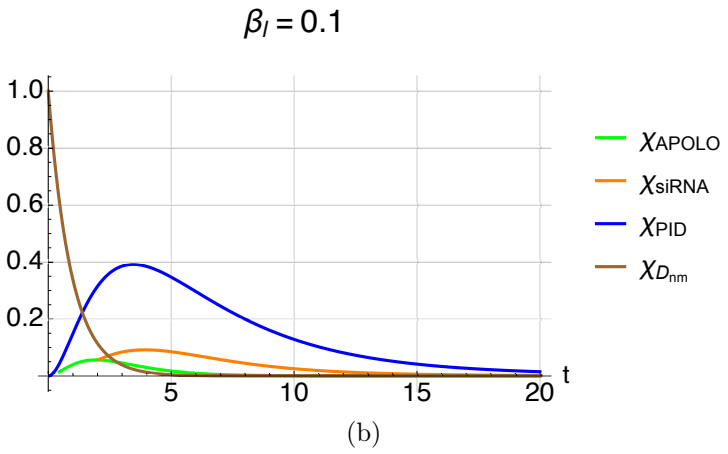

(a) (b)

Figure 4.6: **Phase portrait comparison between first and second formulation**. 4.6a and 4.6b show (respectively for the first and second model) the phase portrait of [APOLO] with respect to $\epsilon$ and $H_{free}$. Notice that the role of the *euchromatin state* in APOLO transcription is assumed by the fraction of free chromatin in the second formulation.

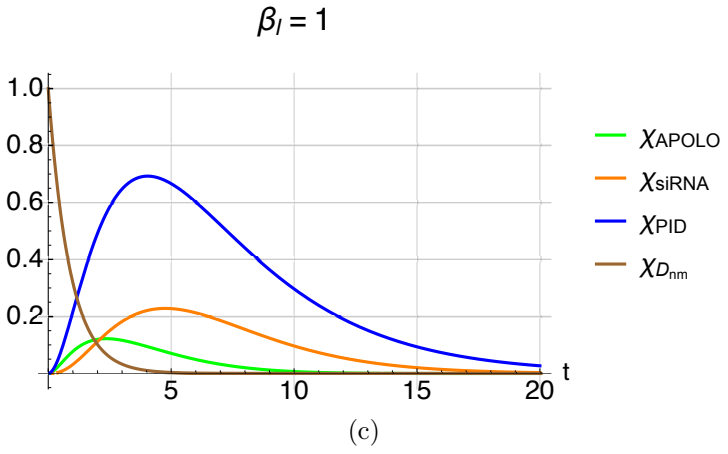
Turning to the comparison of mutants, we started to analyze the results in RNAi plants. In both versions of the complete model, the RNAi mutation strongly affects gene transcription. Both [APOLO] and [siRNA] responses are disrupted, while that of [PID] is strongly reduced, as the experiments show a delayed peak in the latter case (figure 4.8).
As for *lhp1* plants, the responses of APOLO and PID are strongly reduced contrary to the experiments and no peak delay is evident (figure 4.7).




(a)

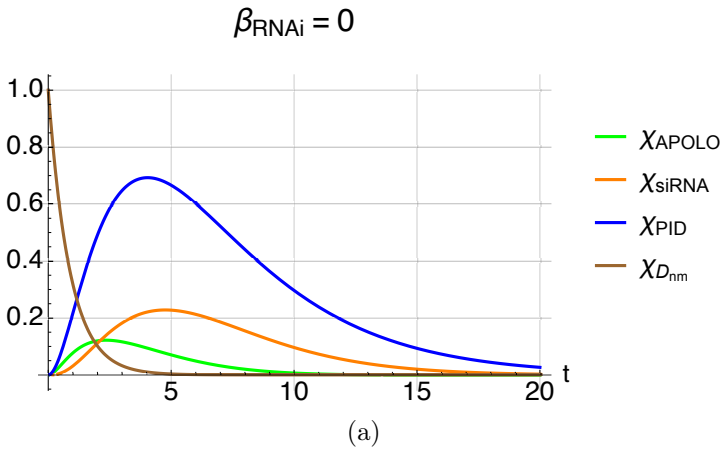



(b)

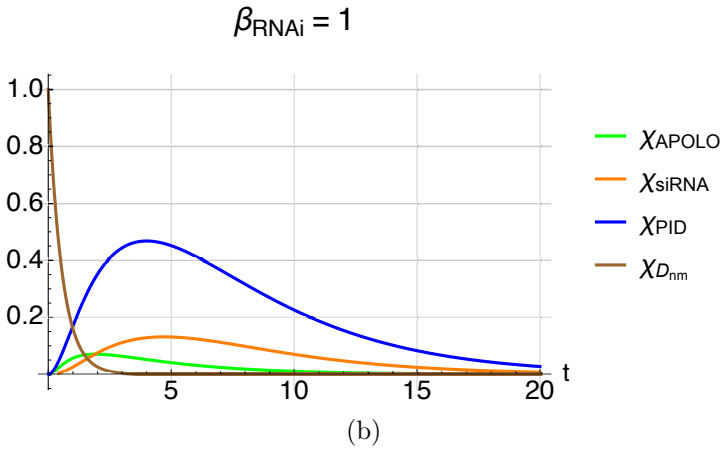



(c)

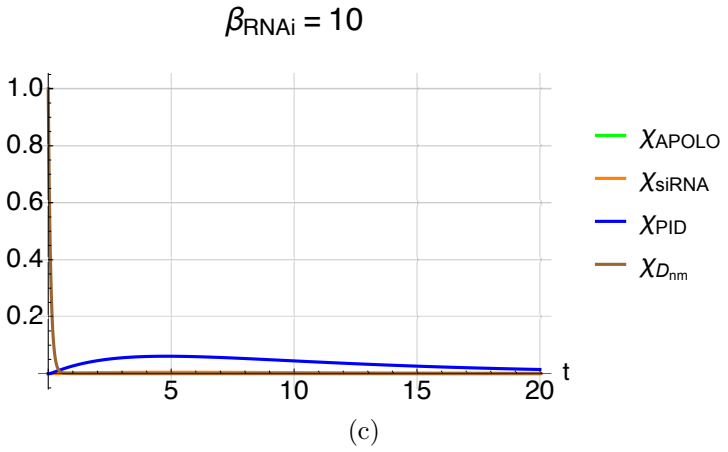
Figure 4.7: **Second formulation: comparison of the linear responses by varying** $\beta_l$. 4.7a, 4.7b and 4.7c represents the cases in which $\beta_l$ assumes respectively the values: 0.0 (*lhp1* mutant), 0.1 and 1.0. $\beta_l$ is a concentration change rate, i.e., it is defined in units of concentraiton/time.




(a)




(b)




(c)

Figure 4.8: **Second formulation: comparison of the linear responses by varying** $\beta_{RNAi}$. 4.8a, 4.8b and 4.8c represents the cases in which $\beta_{RNAi}$ assumes respectively the values: 0, 1 and 10. As previously mentioned, a value of $\beta_{RNAi} > 0$ represents the RNAi mutation: the higher $\beta_{RNAi}$ the stronger is RNA-interference., while $\beta_{RNAi} = 0$ represents WT. $\beta_{RNAi}$ is a concentration change rate , i.e., it is defined in units of concentration/time.

### 4.2.3 Separate genic region

The need to improve the dynamics of the model led us to make more details explicit in the formulation. The third test in modeling the complete system was to definitively divide the three gene regions: APOLO, PID and intergenic region. Another important difference concerns the complexes formed with histones. While in the previous case we assumed that first APOLO binds histones and only afterwards LHP1 can bind to the $HA$ complex, a discussion with biologists pointed out an error in this interpretation. In fact, it is more likely that histones are bound first by LHP1, forming $HL$, and only then does APOLO interact with it in the $HAL$ complex. This will lead to slightly different chemical equations:

$$[Histones] + [LHP1] \underset{K_{HL}^{\text{off}}}{\overset{K_{HL}^{\text{on}}}{\rightleftharpoons}} [HL]; \tag{4.20}$$

$$[HL] + [APOLO] \underset{K_{HAL}^{\text{off}}}{\overset{K_{HAL}^{\text{on}}}{\rightleftharpoons}} [HAL]. \tag{4.21}$$

**Three genic regions:** There are three DNA regions of interest in the analyzed biological system. So far we have considered the DNA to be unique, which is a reasonable approximation. The results of the last model made us think about going beyond this assumption and actually considering the three regions as separate. The consequence in mathematical terms is the need to introduce new equations: $HL^{(a)}$, $HAL^{(a)}$ for the APOLO locus, $HL^{(p)}$, $HAL^{(p)}$ for the PID's, and $HL^{(IR)}$, $HAL^{(IR)}$ for the intergenic region, but still satisfying the conditions:

$$\begin{cases} 1 = H_{free}^{(a)} + HL^{(a)} + HAL^{(a)} \\ 1 = H_{free}^{(p)} + HL^{(p)} + HAL^{(p)} \\ 1 = H_{free}^{(IR)} + HL^{(IR)} + HAL^{(IR)} \end{cases}. \tag{4.22}$$

Accordingly, it is appropriate to define three equations for $D_{nm}^{(a)}$, $D_{nm}^{(p)}$, $D_{nm}^{(IR)}$. As shown in the experiments, only the APOLO locus is subject to DNA methylation, and therefore both $D_{nm}^{(p)}$ and $D_{nm}^{(IR)}$ are fixed at one. Consequently, only one equation corresponding to $D_{nm}^{(a)}$ needs to be added while the equations for $HL^{(p)}$, $HAL^{(p)}$, $HL^{(IR)}$ and $HAL^{(IR)}$ are completely independent of DNA methylation.

$$
\begin{cases}
\underline{\text{APOLO region}}: \\
\\
\dfrac{d}{dt}D_{nm}^{(a)} = \beta_D A_\tau(t)\left(1 - D_{nm}^{(a)}\right) - \delta_{RdDM}[siRNA]\cdot D_{nm}^{(a)} \\
\\
\dfrac{d}{dt}HL^{(a)} = K_{HL}^{\text{on}}[LHP1]\left(1 - HL^{(a)} - HAL^{(a)}\right) - K_{HL}^{\text{off}}HL^{(a)} + \\
\qquad\qquad -K_{HAL}^{\text{on}}[APOLO]HL^{(a)} + K_{HAL}^{\text{off}}HAL^{(a)} - \beta_m D_{nm}^{(a)}HL^{(a)} \\
\\
\dfrac{d}{dt}HAL^{(a)} = K_{HAL}^{\text{on}}[APOLO]HL^{(a)} - K_{HAL}^{\text{off}}HAL^{(a)} - \beta_m D_{nm}^{(a)}HAL^{(a)} \\
\\
\\
\underline{\text{PID region}}: \\
D_{nm}^{(p)} = 1 \\
\\
\dfrac{d}{dt}HL^{(p)} = K_{HL}^{\text{on}}[LHP1]\left(1 - HL^{(p)} - HAL^{(p)}\right) - K_{HL}^{\text{off}}HL^{(p)} \\
\qquad\qquad -K_{HAL}^{\text{on}}[APOLO]HL^{(p)} + K_{HAL}^{\text{off}}HAL^{(p)} \\
\\
\dfrac{d}{dt}HAL^{(p)} = K_{HAL}^{\text{on}}[APOLO]HL^{(p)} - K_{HAL}^{\text{off}}HAL^{(p)} \\
\\
\\
\underline{\text{Intergenic region}}: \\
D_{nm}^{(IR)} = 1 \\
\\
\dfrac{d}{dt}HL^{(IR)} = K_{HL}^{\text{on}}[LHP1]\left(1 - HL^{(IR)} - HAL^{(IR)}\right) - K_{HL}^{\text{off}}HL^{(IR)} + \\
\qquad\qquad -K_{HAL}^{\text{on}}[APOLO]HL^{(IR)} + K_{HAL}^{\text{off}}HAL^{(IR)} \\
\\
\dfrac{d}{dt}HAL^{(IR)} = K_{HAL}^{\text{on}}[APOLO]HL^{(IR)} - K_{HAL}^{\text{off}}HAL^{(IR)}
\end{cases}
\tag{4.23}
$$

**PID/APOLO transcription:** Next step is to understand how transcription of the two genes depends on these quantities. As explained in the second chapter, the transcription of genes in response to auxin is due to two factors: DNA compaction and loop opening. For these reasons, we will make

PID-APOLO transcription proportional to the product of two quantities: a measure of chromatin compaction and a probability that the loop is open. How can chromatin compaction be measured? As already done for the minimal models, we will introduce a quantity $\epsilon$ that will take value 1 if the chromatin is not compact and 0 vice versa. Clearly $\epsilon$ will be differentiated for the three regions and proportional to the fraction of unmethylated DNA $D_{nm}$ and free histons $H_{free}$ (where $H_{free} = 1 - HL - HAL$).

$$\begin{cases} \epsilon_a(t) = H^{(a)}_{free}(t) \cdot D^{(a)}_{nm}(t) \\ \epsilon_p(t) = H^{(p)}_{free}(t) \;\text{ since }\; D^{(p)}_{nm} = 1 \;\text{ always} \\ \epsilon_{IR}(t) = H^{(IR)}_{free}(t) \;\text{ since }\; D^{(IR)}_{nm} = 1 \;\text{ always} \end{cases} . \tag{4.24}$$

Concerning the probability of the loop being open, it can be clearly defined as:

$$P_{\text{loop open}} = 1 - P_{\text{loop closed}}. \tag{4.25}$$

In the experimental results, it was evident that the effective loop closure was due to the increased compaction of the PID and APOLO loci, while the intergenic region remained "relaxed." It is easy to define the probability of loop closure as the product of the compaction measures of the PID and APOLO loci:

$$P_{\text{loop closed}} = (1 - \epsilon_a)\,(1 - \epsilon_p)\,. \tag{4.26}$$

Now that all the appropriate assumptions have been made, the equation for the remaining molecular species can be written.

$$
\begin{cases}
\dfrac{d}{dt}[\text{APOLO}] = \beta_0 + \beta_a \epsilon_a P_{\text{loop open}} - \delta_a [APOLO] + \\
\qquad\qquad -\delta_{s,a}[APOLO][siRNA] - K_{LA}^{\text{on}}[APOLO][LHP1] + \\
\qquad\qquad -\hat{K}_{HAL}^{\text{on}}[APOLO]\left(HL^{(a)} + HL^{(p)} + HL^{(IR)}\right) + \\
\qquad\qquad +\hat{K}_{HAL}^{\text{off}}\left(HAL^{(a)} + HAL^{(p)} + HAL^{(IR)}\right) + \\
\qquad\qquad +K_{LA}^{\text{off}}[LHP1\text{-}APOLO] + \beta_m D_{nm} HAL \\
\\
\dfrac{d}{dt}[siRNA] = \beta_s [APOLO] - \delta_s [siRNA] + \beta_{RNAi} \\
\\
\dfrac{d}{dt}[PID] = \beta_p \epsilon_p P_{\text{loop open}} - \delta_p [PID] + \beta_{p,0} \\
\\
\dfrac{d}{dt}[LHP1] = \beta_l - \delta_l [LHP1] - K_{LA}^{\text{on}}[LHP1][APOLO] + K_{LA}^{\text{off}}[LHP1\text{-}APOLO] + \\
\qquad\qquad -\hat{K}_{HL}^{\text{on}} HL[LHP1]\left(H_{free}^{(a)} + H_{free}^{(p)} + H_{free}^{(IR)}\right) + \\
\qquad\qquad +\hat{K}_{HL}^{\text{off}}\left(HL^{(a)} + HL^{(p)} + HL^{(IR)}\right) + \beta_m \left(HL^{(a)} + HAL^{(a)}\right) \\
\\
\dfrac{d}{dt}[LHP1\text{-}APOLO] = K_{LA}^{\text{on}}[LHP1][APOLO] - K_{LA}^{\text{off}}[LHP1\text{-}APOLO]
\end{cases}
\tag{4.27}
$$

For simplicity, we consider only a linear input function in this formulation, since we are primarily interested in the interaction between the components. Functional dependence can be studied in a future analysis.

**Stationary states and stability:** The more complicated the model, the more difficult it is to reduce the number of equations to solve. In the latter model, the solution was found by solving four equations for: [APOLO], [LHP1], $HL^{(a)}$ and $HAL^{(a)}$. The result was used to calculate, from the stationary values of these quantities, those of all the others. As usual, it is not possible to do this analytically, but only by substituting the numerical values of the parameters. The steady state of this new system was always found to be stable under varying the parameters $\beta_D A$ and $\beta_0$, and the dependence of the molecular species on these two parameters closely resembled the behavior of the old models.
Next, the Jacobian matrix was evaluated at the stationary point and its eigenvalues were analyzed for local stability. As mentioned previously, a common feature of all these models is that their steady state is always stable, and this case was no different. All eigenvalues have, in fact, a negative

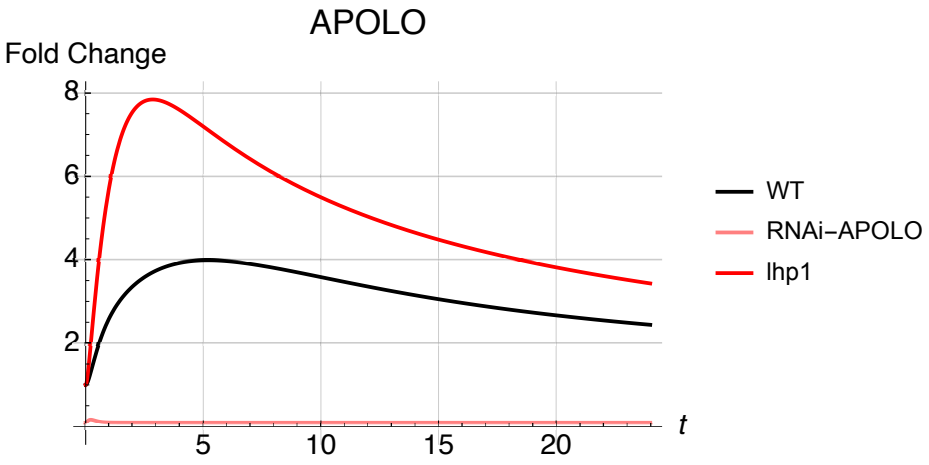
real part, while the imaginary one was non-zero in some regions of the $\beta_D A$-$\beta_0$ plane.

**APOLO and LHP1 conservation:** Unlike the other cases, the more complex ODE system in this section may suggest that some interactions can contain inaccuracies. To check that everything has been well considered, a good index is the conservation principle.
For a molecular species to be conserved over time during the integration of the equations, all terms representing a sink or a source must be eliminated. Let us first consider the case of [APOLO]. From the first equation of the system 4.27 it is easy to distinguish sink/source terms from others. While $+\beta_0$, $+\beta_a \epsilon_a P_{\text{loop open}}$ are sources and $-\delta_a[APOLO]$, $-\delta_{s,a}[APOLO][siRNA]$ are the sinks, the others represent the binding or detachment of APOLO from other complexes. By setting $\beta_0 = \beta_a = \delta_{s,a} = 0$ the concentration of APOLO must remain unchanged over time. By integrating the equations with these conditions, it was possible to show that this is indeed the case. The same discussion can be repeated with the concentration of LHP1 by eliminating the source/sink terms by fixing $\beta_l, \delta_l = 0$. Even in this second case, [LHP1] turns out to be conserved over time.

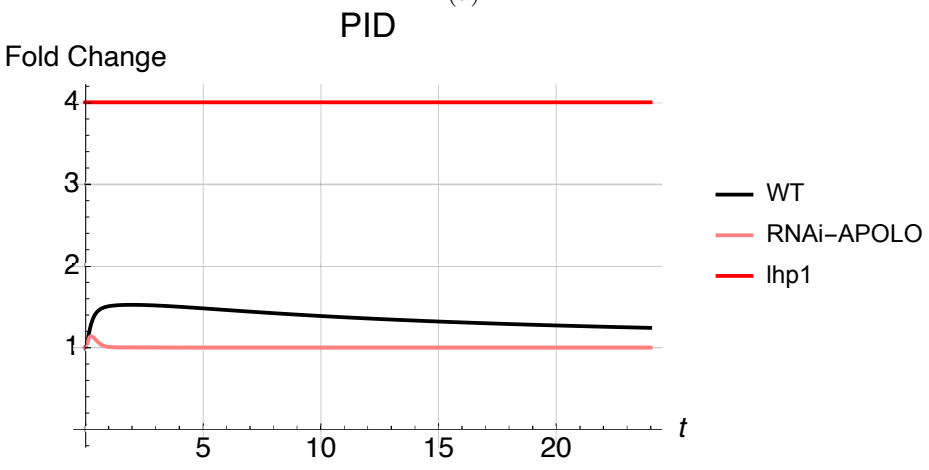
**Fold change and issue with *lhp1* mutant:** So far we have analysed the numerical integration in time of the ODE systems and their linear response to obtain a qualitative representation of the experiements. In the paper [1], data are collected in the form of fold change[21] with respect to the steady state of the wild type. To improve the model's ability to replicate real experiments, we turned to the study of fold change. After initializing the model parameters, we calculated the steady-state values of the molecular species and then integrated the ODE system. The obtained trajectory was normalized by dividing it by the steady-state values, thus obtaining a fold change prediction. Clearly, the qualitative behavior was still comparable with the experiments, but the large number of parameters made it difficult to approach the experimental curves quantitatively. In these cases it may be meaningful to attempt a different approach, to calculate the best values of the parameters that reproduce the experiments, in other words, inference may be necessary.
As for APOLO, the results seemed interesting: compared with the wild type, the *lhp1* mutant showed a higher peak, while the RNAi mutant was completely turnded off, exactly as in the experiments (figure 4.9a).
Regarding PID, however, while the fold change of the RNAi mutant was

[21] Fold change is a measure describing how much a quantity changes between an original and a subsequent measurement. It is defined as the ratio between the two quantities; for quantities A and B the fold change of B with respect to A is B/A.

consistent with the measurements, the *lhp1* plant showed a peculiar behavior. Upon integrating the equation, PID concentration remained constant and equal to zero regardless of auxin perturbation, leading to zero fold change (figure 4.9b). This result highlighted a problem in modeling the genetic circuit. In fact, in the *lhp1* mutant, in which $\beta_l = 0$, there is no production of LHP1 and this causes the absence of the $[LHP1\text{-}APOLO]$, $HL^{(a)}$, $HL^{(p)}$, $HL^{(IR)}$ and $HAL^{(a)}$, $HAL^{(p)}$, $HAL^{(IR)}$ complexes at steady state. Moreover, while $D_{nm}^{(a)}$ is subject to the influence of auxin, $D_{nm}^{(p)} = 1$ always. Consequently, APOLO responds to auxin action with an increased amplitude, caused by the absence of $[LHP1\text{-}APOLO]$, but in the equation for PID the source term $\beta_p \epsilon_p P_{\text{loop open}} = \beta_p = constant$ always. The result is that PID is completely indifferent to auxin perturbation and its concentration is always fixed at the steady-state value, which is effectively $[PID]_{ss} = \frac{\beta_p}{\delta_p}$.




(a)




(b)

Figure 4.9: **Separate genic regions: fold change**. 4.9a and 4.9b show the fold change obtained in the simulation in the three cases: wild type (WT), *lhp1* and RNAi mutant.

### 4.2.4 Considering histone marks

The separation of the three genic regions led us to acceptable results in fold change, but at the same time highlighted a problem in modeling the *lhp1* mutant. In this last section, starting from the discrepancy with the experiments found, the model will be further modified and extended.
One way to resolve the issue of the *lhp1* mutant can be found in the introduction of the hitherto neglected *histone marks*. Briefly summarized, the deposition of histone marks can produce both acetylation and methylation of DNA, which are the most common modifications of histone tails. In a very general way, it can be stated that while acetylation decreases histone-DNA interaction consequently activating transcription, methylation does not act on histone-DNA interaction resulting in both repression and activation of gene transcription [38].

**Histone mark dynamics:** As already done for histones in general, the presence of histone marks can be represented with a variable $H_{mark}$ that takes a value of zero if no marks are deposited and vice versa equals one. Clearly, all the different gene regions must be taken into account here as well, so three equations must be introduced for: $H^{(a)}_{mark}$, $H^{(p)}_{mark}$ and $H^{(IR)}_{mark}$.
Let us first consider the processes that lead to the growth of $H_{mark}$ in the system. Histone marks are deposited and removed in response to specific enzymes called *writers* and *erasers*; the appropriate localisation and activity of these enzymes on chromatin is, in part, regulated by chromatin *readers*. The balance between the effect of writing and deletion leads to a basal level of markers that increases even when markers are already present. This led us to introduce a constant basal term $\beta_{m,0}$.
Secondly, as pointed out by Ariel and coworkers, in the case of Arabidopsis thaliana, the $Histone\text{-}LHP1$ and $Histone\text{-}APOLO\text{-}LHP1$ complexes have a positive effect in histone mark deposition, the effect of which can be taken into account by letting the mark production be proportional to the $HL$ and $HAL$ quantities. Third, histone modifications are not only subject to the effect of writers/erasers, but undergo a phenomenon called spreading. It is easy to observe that spreading can be modeled with an additional source term, proportional to $H_{mark}$ itself.
It is important to remember that the condition $H^{(i)}_{mark} \in [0, 1]$ (where $i = a, p, IR$) must be satisfied: therefore all three terms must be multiplied by $(1 - H^{(i)}_{mark})$. Finally we make the entire positive term proportional to a constant parameter $\gamma_m$. Regarding the degradation of the quantity $H^{(i)}_{mark}$, two components need to be considered. A basal level of deletion can be modeled with the usual $-\delta_m H^{(i)}_{mark}$, while the active demethylation pathway must be introduced not only in proportion to $H^{(i)}_{mark}$ but also to the auxin

concentration: $-\delta_{RDD} A_\tau(t)\, H^{(i)}_{mark}$. The new three equations can be written explicitly:

$$\frac{d}{dt}H^{(i)}_{mark} = \gamma_m \left(\beta_{m,0} + H^{(i)}_{mark} + HL^{(i)} + HAL^{(i)}\right)\left(1 - H^{(i)}_{mark}\right) + \\ - \left(\delta_m + \delta_{RDD} A_\tau(t)\right) H^{(i)}_{mark}. \quad (4.28)$$

**DNA methylation and histone marks:** Both DNA methylation and histone modification are involved in shaping patterns of gene repression during development. It has recently become apparent that the DNA methylation and histone modification pathways may depend on each other: DNA methylation can activate histone modifiers writing, while heterochromatic histone marks are able to activate the DNA methylation pathway [39]. Due to the introduction of histone marks into the equations, the influence of methylated DNA in mark deposition, through the effect of the $HL$ and $HAL$ complexes, becomes apparent: DNA methylation will increase the formation of complexes in the system, which, in turn, will promote mark deposition. This new interaction replaces the role of the $HAL$ complex in the equation for $D^{(a)}_{nm}$. The equations for the three intergenic regions become:

$$\begin{cases} \underline{\text{APOLO region}}: \\ \\ \dfrac{d}{dt}D^{(a)}_{nm} = \beta_D A_\tau(t)\left(1 - D^{(a)}_{nm}\right) - \delta_{RdDM}[siRNA]\cdot D^{(a)}_{nm} \\ \\ \dfrac{d}{dt}HL^{(a)} = K^{\text{on}}_{HL}[LHP1]\left(1 - HL^{(a)} - HAL^{(a)}\right) - K^{\text{off}}_{HL}HL^{(a)} + \\ \qquad -K^{\text{on}}_{HAL}[APOLO]HL^{(a)} + K^{\text{off}}_{HAL}HAL^{(a)} + \\ \qquad -\beta_m D^{(a)}_{nm}HL^{(a)} \\ \\ \dfrac{d}{dt}HAL^{(a)} = K^{\text{on}}_{HAL}[APOLO]HL^{(a)} - K^{\text{off}}_{HAL}HAL^{(a)} - \beta_m D^{(a)}_{nm}HAL^{(a)} \\ \\ \dfrac{d}{dt}H^{(a)}_{mark} = \gamma_m \left(\beta_{m,0} + H^{(a)}_{mark} + HL^{(a)} + HAL^{(a)}\right)\left(1 - H^{(a)}_{mark}\right) + \\ \qquad - \left(\delta_m + \delta_{RDD} A_\tau(t)\right) H^{(a)}_{mark} \end{cases} ; \quad (4.29)$$

$$\begin{cases} \underline{\text{PID region}}: \\ \\ D_{nm}^{(p)} = 1 \\ \\ \dfrac{d}{dt}HL^{(p)} = K_{HL}^{\text{on}}[LHP1]\left(1 - HL^{(p)} - HAL^{(p)}\right) - K_{HL}^{\text{off}}HL^{(p)} + \\ \qquad\qquad -K_{HAL}^{\text{on}}[APOLO]HL^{(p)} + K_{HAL}^{\text{off}}HAL^{(p)} \\ \\ \dfrac{d}{dt}HAL^{(p)} = K_{HAL}^{\text{on}}[APOLO]HL^{(p)} - K_{HAL}^{\text{off}}HAL^{(p)} \\ \\ \frac{d}{dt}H_{mark}^{(p)} = \gamma_m\left(\beta_{m,0} + H_{mark}^{(p)} + HL^{(p)} + HAL^{(p)}\right)\left(1 - H_{mark}^{(p)}\right) + \\ \qquad\qquad -\left(\delta_m + \delta_{RDD}A_\tau\left(t\right)\right)H_{mark}^{(p)} \end{cases} ; \tag{4.30}$$

$$\begin{cases} \underline{\text{Intergenic region}}: \\ \\ D_{nm}^{(IR)} = 1 \\ \\ \dfrac{d}{dt}HL^{(IR)} = K_{HL}^{\text{on}}[LHP1]\left(1 - HL^{(IR)} - HAL^{(IR)}\right) - K_{HL}^{\text{off}}HL^{(IR)} + \\ \qquad\qquad -K_{HAL}^{\text{on}}[APOLO]HL^{(IR)} + K_{HAL}^{\text{off}}HAL^{(IR)} \\ \\ \dfrac{d}{dt}HAL^{(IR)} = K_{HAL}^{\text{on}}[APOLO]HL^{(IR)} - K_{HAL}^{\text{off}}HAL^{(IR)} \\ \\ \frac{d}{dt}H_{mark}^{(IR)} = \gamma_m\left(\beta_{m,0} + H_{mark}^{(IR)} + HL^{(IR)} + HAL^{(IR)}\right)\left(1 - H_{mark}^{(IR)}\right) + \\ \qquad\qquad -\left(\delta_m + \delta_{RDD}A_\tau\left(t\right)\right)H_{mark}^{(IR)} \end{cases} . \tag{4.31}$$

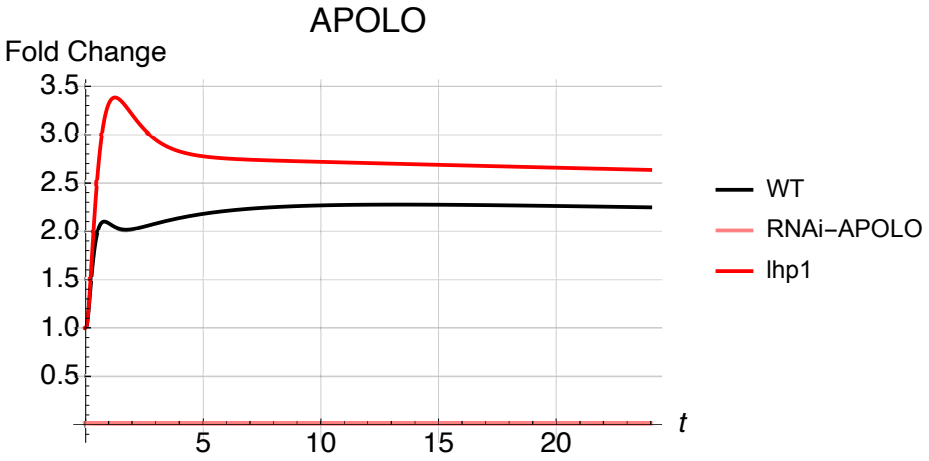
**PID/APOLO transcription:** The changes added to the system do not directly affect the equations of the other molecular species in the model, so reference can still be made to the previous ODE system 4.27. The only two terms that need to be modified indirectly are those referring to the PID and APOLO transcription. In the previous formulation, it was assumed that transcripts were produced in proportion to the product between a variable representing the *"euchromatin state"* and the probability that the chromatin loop was open. Since it is known that the two signals of heterochromatin

are represented by DNA methylation and histone marks, the only update to be made is to consider the euchromatin state inversely proportional to the deposition of the marks, multiplying it by $(1 - H^{(i)}_{mark})$.

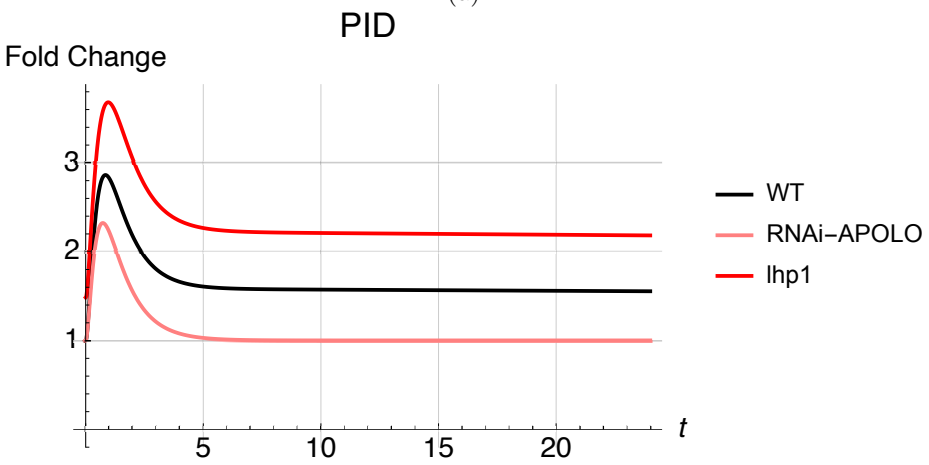
$$\begin{cases} \epsilon_a(t) = H^{(a)}_{free}(t) \cdot \left(1 - H^{(a)}_{mark}\right) \cdot D^{(a)}_{nm}(t) \\ \epsilon_p(t) = H^{(p)}_{free}(t) \cdot \left(1 - H^{(p)}_{mark}\right) \quad \text{since} \quad D^{(p)}_{nm} = 1 \quad \text{always} \\ \epsilon_{IR}(t) = H^{(IR)}_{free}(t) \cdot \left(1 - H^{(IR)}_{mark}\right) \quad \text{since} \quad D^{(IR)}_{nm} = 1 \quad \text{always} \end{cases} . \tag{4.32}$$

**Stationary states and stability:** The large number of equations and the increased complexity make parameter reduction not very useful. The first step in the analysis is therefore the solution of the steady-state equations. They can be reduced to the solution of five equations for: [APOLO], $HL^{(a)}$, $H^{(a)}_{mark}$, $H^{(p)}_{mark}$ and $H^{(IR)}_{mark}$, from which the steady-state values of the other molecular species can be calculated. As usual, the steady state found is unique and stable. There is no notable difference in the behavior of the steady-state concentrations from the previous model, but it is interesting to note that $[LHP1]_{ss}$ is constant regardless of the parameters $\beta_D A$ and $\beta_0$.

**Fixing issue in *lhp1* mutant:** The second step is to produce and study the fold change of the new model. In the last formulation, the lack of histone marks led to the null expression of PID in the *lhp1* mutant. Notably, the fact that the DNA is always unmethylated in the PID region makes the "euchromatin state" indifferent to auxin perturbations.
The introduction of histone marks, another index of heterochromatin along with DNA methylation, solves the problem. In fact, histone modifications are strongly affected by auxin treatments through the RDD pathway, causing the increase in euchromatin state, which in turn produces PID transcripts.
The fold change graphs (figure 4.10) show that the qualitative behavior is now fully accounted for: the expression of PID in the mutant *lhp1* is indeed reduced compared to WT, but not null. Ultimately, the large number of parameters makes it really difficult to try to obtain even quantitative similarity: an inference process is required.




(a)




(b)

Figure 4.10: **Considering histone marks: fold change**. 4.10a and 4.10b show the fold change obtained in the simulation in the three cases: wild type (WT), *lhp1* and RNAi mutant.

# Conclusions and future developments

In this thesis, an attempt was made to obtain a quantitative modeling of the PID-APOLO gene regulatory network with the aim of giving a qualitative and quantitative explanation of the mechanisms regulating this network.
We began by presenting the biological background: from the basics of biological sequences to the mechanisms of gene regulation. The first chapter aimed to provide the necessary tools for a clear understanding of the biological question.
The latter was introduced later in the second chapter. First by presenting the main players in the system and second by showing the work and results obtained by Ariel and his collaborators. Briefly summarizing: PID and APOLO genes are co-regulated in response to a perturbation of the system by auxin. Auxin triggers the opening of the chromatin loop allowing transcription of genes; when the concentration of transcripts reaches certain levels, APOLO RNA itself leads to loop closure, halting the entire transcription process.
Quantitative modeling of systems is not possible without mathematical tools: the goal of the third chapter is to expose them. Starting from the basics of transcriptional network theory, the assumptions under which biological interactions can be transformed into mathematical equations were presented. By making use of the theory of dynamical systems, the last part of the chapter aimed to explain the means by which mathematical equations can be studied to extract information from them.
The fourth chapter was entirely devoted to the modeling procedure and analysis of the system. We started by first identifying the main biological actors in the network by studying extremely simplified models.
A toy model was constructed, adding, step by step, all the most important components and interactions of the system. This resulted in a simple ODE system that, despite its coarseness, qualitatively reproduces the experimental measurements. The concentrations of APOLO, PID and siRNA, as well as an index representing DNA euchromatin, proved sufficient to encapsulate the main features of the dynamics in the wild type. Certainly fundamen-

tal to the dynamics are the transcription terms (which depend directly on how euchromatic the system is) and the degradation terms, which represent the natural lifetimes of the molecular species. However, greater interest is captured by the active demethylation and RNA-mediated methylation pathways, which in turn allow the chromatin loop to open in response to auxin and ultimately to close again when siRNA transcripts increase. These interactions are thus primarily responsible for the oscillation in measured gene expression. Of particular relevance is the RNA-interaction of siRNA acting on APOLO transcripts: without this term, PID and APOLO would be identical in their expression. The result of RNA-interference is a significantly lower response of APOLO; this led us to hypothesize that RNAi may play a key role in explaining why APOLO expression is repressed in amplitude in experiments compared with PID.
Statics analysis showed the obvious effect of basal transcription of $\beta_0$ in increasing the steady-state concentration of the respective APOLO gene. The proportionality of the steady-state concentrations of both genes to the *"effective auxin flux"* also emerged. Less evident is that basal transcription of APOLO does not have a strong influence on either euchromatin state or [PID], at least for low values of $\beta_{epsilon}A$. Moreover, an increase in basal transcription of APOLO would lead to a higher level of $[APOLO]_{ss}$, as explained above, but at the same time would reduce the effect of auxin perturbation and consequently reduce the peak in fold change of APOLO itself. In summary, it can be said that basal transcription plays a minor role in the dynamics and statics of the system. In conclusion, the minimal models seemed to point in the right direction, but they are still too coarse to describe many details. For example, the behavior of mutant plants has not been considered so far.
The second section of the fourth chapter is devoted to the formulation and analysis of more complex models. In the first formulation we introduced the role of LHP1 and the complexes formed with APOLO and chromatin. Despite its complexity, the new model showed the same static and dynamic characteristics as the previous ones. In addition, however, it allows the consideration of mutants. Simulations of the RNAi mutant gave interesting results from the beginning, while the *lhp1* mutant did not deviate much from WT behavior (contrary to the experiments).
The first attempt to improve the model was to challenge the assumption that DNA dynamics can be represented by the *chromatin state* alone. Therefore, the latter was first replaced by a variable representing DNA methylation, and then the model was further extended to account for both DNA methylation and chromatin state in the system. Finally, the three gene regions of the system (PID, APOLO and intergenic) were split in the equations to also account for the differences in methylation (at the different loci) measured in the experiments. The resulting model was shown to retain all the

dynamic features of the previous ones, as well as the qualitative behavior of the experiments. Despite the good results obtained, a problem emerged during the analysis with the *lhp1* mutant that produced constant PID gene expression in the simulations. For PID to be correctly expressed, it is not sufficient to consider only the DNA methylation of its locus. Indeed, experiments showed that the PID locus was always and consistently unmethylated, causing insensitivity to auxin stimulus in the *lhp1* mutant. In this case it was necessary to introduce histone modifications, which are another signal of heterochromatin in the DNA. The latter addition, besides making the system more realistic, allowed PID expression to be activated in the *lhp1* mutant.

**Model calibration:**

As anticipated earlier, dynamical models such as those formulated contain many parameters. Some of them can be extracted from the literature, but many others cannot be measured experimentally. For this reason, the next step in the study of formulated models is calibration and validation. Calibration of a model consists of looking for the unknown parameters that allow the model prediction to fit the experimental data.
We built these models step by step by first trying to follow the qualitative behavior of the PID-APOLO circuit. This helped us to understand the role of different interactions in the whole system, but now we will turn to a more quantitative analysis.
The data available are from measurements produced with APOLO overexpressors and RNAi Knockdowns and were collected by Crespi's team at the Institut de Sciences du Végétal in Gif-sur Yvette; one example was shown in figure 2.3.
Returning to the different models analyzed, as mentioned in the corresponding section, when the division into three gene regions was added to the equations, the models began to be too complex even to extract from them a qualitative behavior closer to reality. Below, we will show the calibration of the first model taking into account the three gene regions on the available fold change data.
The procedure used to fit the model to the data is based on the minimization of an objective function $f_{obj}$ that measures the distance between experiments and model predictions.
The fitting routine was implemented in Matlab. First, an array of parameters was generated according to the random distribution $\Gamma(k, \theta)$[22] where $\theta$ was set equal to the value of each parameter obtained from the best man-

[22]Gamma distribution is usually considered in order to avoid negative initial parameters.

ual fit and $k = 1$. Once the initial conditions for the parameters were set, the values of the species were initialized at time zero at the corresponding steady state and the equations were simulated using the *ode15_s()* algorithm[23] [40].
From the integration of the ODE system, the concentrations of PID and APOLO at time instants $t_i$ $(i = 1, ...9)$, which are the time instants corresponding to the experimental measurements, were extracted. To obtain a fold change, once collected, these concentrations were divided by the steady-state concentrations calculated at the beginning. We call the experimental and theoretical fold changes respectively $\psi^{th}_{APOLO}(t_i)$, $\psi^{th}_{\mathrm{PID}}(t_i)$ and $\psi^{exp}_{\mathrm{APOLO}}(t_i)$, $\psi^{exp}_{\mathrm{PID}}(t_i)$, now the objective function can be defined as follows:

$$f_{obj} = \sum_{i=1}^{9} \left( \frac{\psi^{th}_{APOLO}(t_i) - \psi^{exp}_{\mathrm{APOLO}}(t_i)}{\langle \psi^{exp}_{\mathrm{APOLO}}(t_j) \rangle_{t_j}} \right)^2 + \sum_{i=1}^{9} \left( \frac{\psi^{th}_{PID}(t_i) - \psi^{exp}_{\mathrm{PID}}(t_i)}{\langle \psi^{exp}_{\mathrm{PID}}(t_j) \rangle_{t_j}} \right)^2, \tag{4.33}$$

where $\langle \psi^{exp}_{\mathrm{APOLO}}(t_j) \rangle_{t_j}$ and $\langle \psi^{exp}_{\mathrm{PID}}(t_j) \rangle_{t_j}$ are the average values in the time interval $[t_1, t_9]$. As can be seen from the equation 4.33, the objective function is represented by the squared Euclidean distance so as to penalize discrepancies between experimental and theoretical data. After computing the objective function, the search for minima with respect to the parameters was carried out using the built-in function *fminsearch()* in Matlab, which applies the Nelder-Mead algorithm [40, 41].
The routine was repeated 100 times until the algorithm found a minimum. Whenever it did not, the procedure was repeated by initializing the parameters to the last output of the minimization algorithm.
As shown in Figure 4.11, the range of values of the minima of $f_{obj}$ is quite wide. Clearly, the algorithm often got stuck in a local minimum around $f_{obj} \simeq 0.35$, but then it was able to reach a lower distribution of minima centred around $f_{obj} \simeq 0.1$.

In order to consider only the region where $f_{obj}$ is lowest and discard the local minima, we cut out the whole interval $f_{obj} \geq 0.1$. For each estimated parameter, we examined the *Coefficient of Variation* (CV), which is the ratio of the standard deviation to the mean value across runs of the algorithm. Figure 4.12 shows that even considering only the lower distribution of minima, the coefficient of variation is never less than 1. One reason could be that we are letting all the parameters vary, and the CV can probably be made smaller by fixing some of them according to experimental estimation.

[23] ode15_s() algorithm is a variable step, variable order method based on Backward Difference Formulas (BDF) for stiff systems.

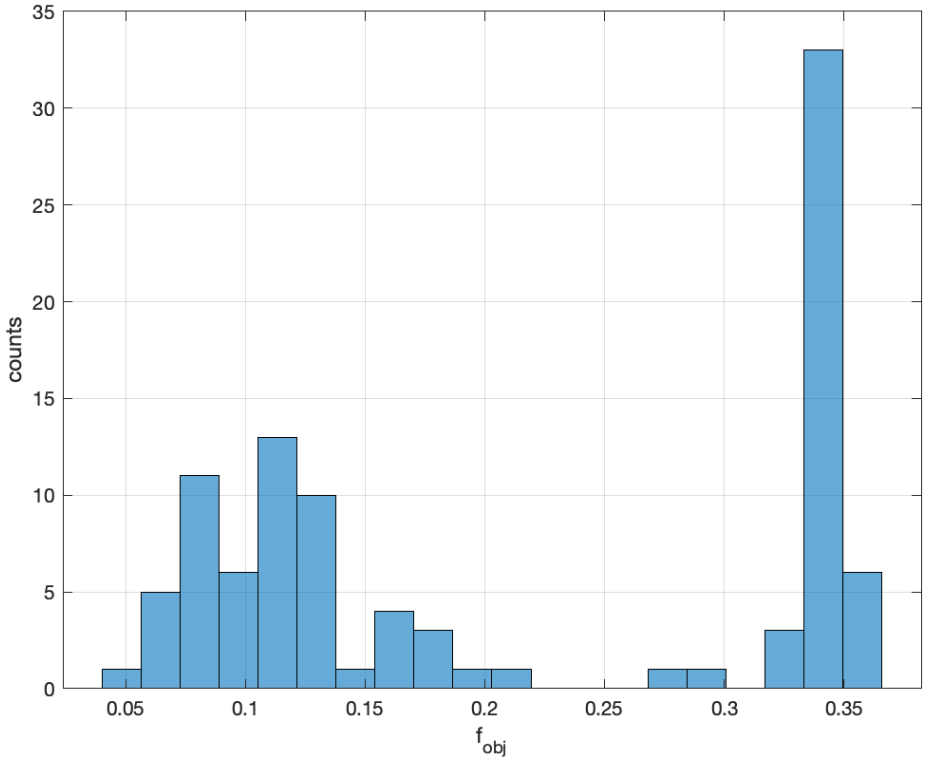



Figure 4.11: Histogram of the results of calibration. One local minimum is found around $f_{obj} \simeq 0.35$ and a distribution of lower minima at $f_{obj} \simeq 0.1$.

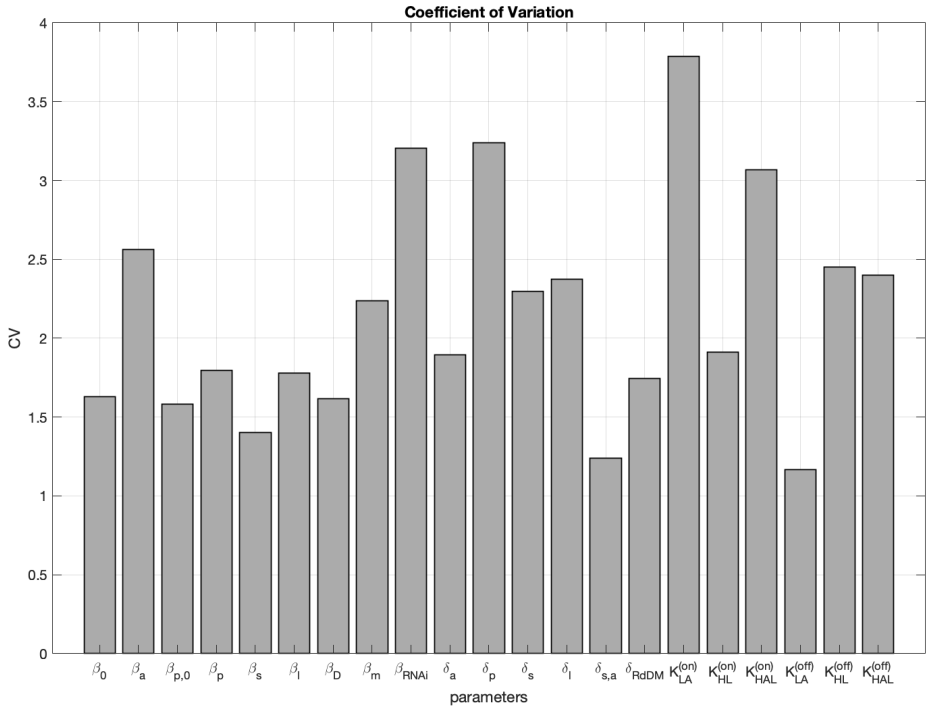



Figure 4.12: Coefficient of variation computed for each parameter considering only the region $f_{obj} \leq 0.1$

In the end, the best result achieved an objective function value of $f_{obj} = 0.049172$, and as shown in figure 4.13 is quite good at reproducing the experimental data.

The parameters corresponding to the best fit are shown in Table 4.1.

| $\beta_0$ | $\beta_a$ | $\beta_{p,0}$ | $\beta_p$ | $\beta_s$ | $\beta_l$ | $\beta_D$ | $\beta_m$ | $\beta_{RNAi}$ |
|---|---|---|---|---|---|---|---|---|
| 0.0006 | 107.7514 | 0.0002 | 0.6839 | 0.0007 | 2.3474 | 24.2990 | 10.4319 | 0.0 |

| $\delta_a$ | $\delta_p$ | $\delta_s$ | $\delta_l$ | $\delta_{s,a}$ | $\delta_{RdDM}$ |
|---|---|---|---|---|---|
| 0.0361 | 2.4121 | 0.2948 | 0.0199 | 2.7968 | 0.0762 |

| $K_{LA}^{\text{on}}$ | $K_{HL}^{\text{on}}$ | $K_{HAL}^{\text{on}}$ | $K_{LA}^{\text{off}}$ | $K_{HL}^{\text{off}}$ | $K_{HAL}^{\text{off}}$ |
|---|---|---|---|---|---|
| 0.0537 | 2.2346 | 0.0029 | 0.0217 | 0.2411 | 0.1344 |

Table 4.1: **Parameters estimates of the best fit.** Respectively divided in positive terms, negative terms and coefficients for complex formation.
$\beta_0, \beta_a, \beta_p, \beta_l, \beta_m, \beta_{RNAi}, K_{HL}^{\text{off}}, K_{HAL}^{\text{off}}$ are expressed in terms of concentration/time;
$\beta_s, \beta_s, \delta_a, \delta_s, \delta_p, \delta_l, K_{HL}^{\text{on}}, K_{HAL}^{\text{on}}, K_{LA}^{\text{off}}$ in terms of $1/\textit{concentration}$;
$\beta_D, \delta_{s,a}, \delta_{RdDM}, K_{LA}^{\text{on}}$ in terms of $1/(\text{concentration} \cdot \text{time})$.

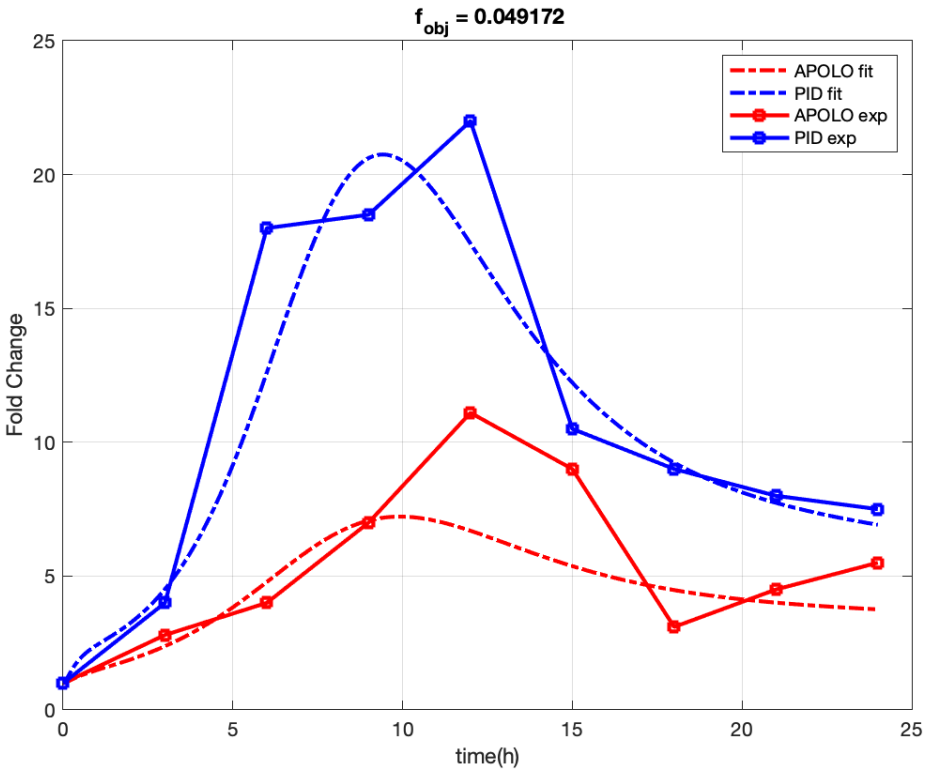



Figure 4.13: **Comparison between the best theoretical prevision and experimental measures.**

### Results and future developments:

Summarizing the results, our analysis first showed that the fundamental structure, which these models have in common, makes them all monostable, thus ensuring the existence of a single steady state that is achieved regardless of the initial perturbation.
The active demethylation (RDD) and RNA-mediated methylation pathways can be considered to underlie the oscillation of gene expression and thus responsible for initial growth and subsequent decay. The role of basal transcription seems to be scaled down from expectations: while the basal transcription of APOLO has little influence in determining DNA methylation and loop closure, that of PID seems irrelevant. Furthermore, it was found that RNA-interference is critical in the context of the minimal models to differentiate PID and APOLO expression and reproduce experimental results. This may not be true when considering complete models in which interactions are more complex and the division of gene regions differentiates the two responses at base.
To better understand these issues, new and more specific measurements may be needed. The detailed values of $[\text{APOLO}]_{ss}$ and $[\text{PID}]_{ss}$ can indeed

be compared with the theoretical values provided by model predictions. This can provide more information about the basal transcription of PID and APOLO and their influence in the system. Another aspect that can be further investigated is the involvement of RNA-interference in differentiating the response of genes, even in complete models. Here again, specific measurements may be useful.

Turning to mutants, we can say that the modeling of RNAi mutants was satisfactory. The basal term in the siRNA equation rapidly increases its concentration by disrupting APOLO transcription and loop opening at the beginning. More complicated, however, was the modeling of the *lhp1* mutant, in which histone modification had to be introduced. It turned out that heterochromatin marks on histones are crucial to explain the behavior of the *lhp1* mutant. Indeed, regulation involving DNA methylation of the PID locus does not explain the reduced but not null expression of PID in the mutant. The lack of LHP1 in the cell affects the system at a deeper level than the RNAi mutation. In fact, it leads to the lack of the LA, HL and HAL complexes. While a lower concentration of LA simply increases the expression of APOLO, the lack of HL and HAL completely blocks the PID locus on the DNA in its initial state, shutting down at base any possibility of changing the rate of expression. It is clear now that, based on the constructed model, the experimental results can only be interpreted with the fact that it is the histone modifications that bring the effect of auxin and the RDD pathway to the PID locus and initiate transcription in the *lhp1* mutant.

Clearly, the previous model cannot be used to fit the available mutant data, except for the RNAi case. It is therefore necessary to switch to calibration of the model in the case with the addition of histone markers. This would give more information about the model's ability to reproduce experiments quantitatively. In addition, the model formulation can be expanded to consider the other mutant cases studied in [1], on which the calibration can be repeated.

In the end parameters calibration showed that the model can be able to fit the experimental measurements in Wild Type. Even though this is still at an initial stage, the results clearly underlined that the roles of both PID and APOLO basal transcriptions are negligible.

In the case of inference, many improvements can be achieved. First, so far we have used only a portion of the available data. To completely validate and calibrate the model, it will be necessary to make use of the data from mutant experiments, as mentioned earlier. In addition, we have imposed the model to fit the experimental mean values and thus neglected the standard deviations associated to the data, which can be used to weight the data points during calibration. Taking into account all these different aspects could definitely help to obtain more robust results. Finally, an important

step in the whole calibration procedure is to evaluate the robustness of the fit, to see how reliable the results obtained are. One hypothesis could be the use of the bootstrapping procedure: by using the best fit and reshuffling the deviations between the model and the experiments, it is indeed possible to generate new artificial data to validate the model itself.
As mentioned earlier, the inference part is still at an early stage, but we are still working on it with the goal of completing the analysis as explained.

In conclusion of this thesis, it can be said that despite the satisfactory results, there are still some things to be understood. Starting with a more precise test of the mutants to see if the formulated model can predict their behavior, one could then move on to a more in-depth analysis of the system's response to external stimuli. In particular, it might be interesting to delve into the biological resilience characteristics of the system and its ability to return to its initial stage after different types of perturbation. The study of these circuits can lead to incredible advances in biology. First, because of the growing interest in both i) the multiple tasks performed by noncoding RNAs and ii) the role of DNA loop dynamics and more generally the structural topology of the genome in regulating gene expression in organisms. Second, as anticipated, understanding the expression of PID and how its tight control is implemented through this complicated process is critical with regard to root development.

# Bibliography


[1] F. Ariel, T. Jegu, D. Latrasse, N. Romero-Barrios, A. Christ, M. Benhamed, and M. Crespi, "Noncoding transcription by alternative rna polymerases dynamically regulates an auxin-driven chromatin loop," *Molecular cell*, vol. 55, no. 3, pp. 383–396, 2014.

[2] J. Kuang, X. Yan, A. J. Genders, C. Granata, and D. J. Bishop, "An overview of technical considerations when using quantitative real-time pcr analysis of gene expression in human exercise research," *PloS one*, vol. 13, no. 5, p. e0196438, 2018.

[3] F. H. Crick, "On protein synthesis," in *Symp Soc Exp Biol*, vol. 12, p. 8, 1958.

[4] A. Klug, "The discovery of the dna double helix," *Journal of Molecular Biology*, vol. 335, no. 1, pp. 3–26, 2004.

[5] B. Alberts, *Molecular biology of the cell*. WW Norton & Company, 2017.

[6] C. De Bustos, *Genetic and Epigenetic Variation in the Human Genome: Analysis of Phenotypically Normal Individuals and Patients Affected with Brain Tumors*. PhD thesis, Acta Universitatis Upsaliensis, 2006.

[7] H. Ohkura, "Meiosis: an overview of key differences from mitosis," *Cold Spring Harbor perspectives in biology*, vol. 7, no. 5, p. a015859, 2015.

[8] J. R. McIntosh, "Mitosis," *Cold Spring Harbor perspectives in biology*, vol. 8, no. 9, p. a023218, 2016.

[9] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, "From rna to protein," in *Molecular Biology of the Cell. 4th edition*, Garland Science, 2002.

[10] K. Matthews, "Dna looping," *Microbiological reviews*, vol. 56, no. 1, pp. 123–136, 1992.

[11] Y. Cai, X. Yu, S. Hu, and J. Yu, "A brief review on the mechanisms of mirna regulation," *Genomics, proteomics & bioinformatics*, vol. 7, no. 4, pp. 147–154, 2009.

[12] M. Osella, C. Bosia, D. Corá, and M. Caselle, "The role of incoherent microrna-mediated feedforward loops in noise buffering," *PLoS computational biology*, vol. 7, no. 3, p. e1001101, 2011.

[13] S. Grigolon, F. Di Patti, A. De Martino, and E. Marinari, "Noise processing by microrna-mediated circuits: the incoherent feed-forward loop, revisited," *Heliyon*, vol. 2, no. 4, p. e00095, 2016.

[14] C. Bosia, F. Sgrò, L. Conti, C. Baldassi, D. Brusa, F. Cavallo, F. D. Cunto, E. Turco, A. Pagnani, and R. Zecchina, "Rnas competing for micrornas mutually influence their fluctuations in a highly non-linear microrna-dependent manner in single cells," *Genome biology*, vol. 18, no. 1, pp. 1–14, 2017.

[15] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello, "Potent and specific genetic interference by double-stranded rna in caenorhabditis elegans," *nature*, vol. 391, no. 6669, pp. 806–811, 1998.

[16] J. Downward, "Rna interference," *Bmj*, vol. 328, no. 7450, pp. 1245–1248, 2004.

[17] Y. Qi, X. He, X.-J. Wang, O. Kohany, J. Jurka, and G. J. Hannon, "Distinct catalytic and non-catalytic roles of argonaute4 in rna-directed dna methylation," *Nature*, vol. 443, no. 7114, pp. 1008–1012, 2006.

[18] G. Meister, "Argonaute proteins: functional insights and emerging roles," *Nature Reviews Genetics*, vol. 14, no. 7, pp. 447–459, 2013.

[19] P.-H. Wang, K. T. Wittmeyer, T.-f. Lee, B. C. Meyers, and S. Chopra, "Overlapping rddm and non-rddm mechanisms work together to maintain somatic repression of a paramutagenic epiallele of maize pericarp color1," *PloS one*, vol. 12, no. 11, p. e0187157, 2017.

[20] J. Cheng, Q. Niu, B. Zhang, K. Chen, R. Yang, J.-K. Zhu, Y. Zhang, and Z. Lang, "Downregulation of rddm during strawberry fruit ripening," *Genome biology*, vol. 19, no. 1, pp. 1–14, 2018.

[21] H. Vaucheret, "Plant argonautes," *Trends in plant science*, vol. 13, no. 7, pp. 350–358, 2008.

[22] M. Iwasaki, L. Hyvärinen, U. Piskurewicz, and L. Lopez-Molina, "Non-canonical rna-directed dna methylation participates in maternal and environmental control of seed dormancy," *Elife*, vol. 8, p. e37434, 2019.

[23] R. C. Kirkbride, J. Lu, C. Zhang, R. A. Mosher, D. C. Baulcombe, and Z. J. Chen, "Erratum: Maternal small rnas mediate spatial-temporal regulation of gene expression, imprinting, and seed development in arabidopsis (proceedings of the national academy of sciences of the united states of america (2019) 116 (2761–2766," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, no. 17, p. 8633, 2019.

[24] H. T. Chow, T. Chakraborty, and R. A. Mosher, "Rna-directed dna methylation and sexual reproduction: expanding beyond the seed," *Current opinion in plant biology*, vol. 54, pp. 11–17, 2020.

[25] R. Liu and Z. Lang, "The mechanism and function of active dna demethylation in plants," *Journal of integrative plant biology*, vol. 62, no. 1, pp. 148–159, 2020.

[26] J. J. Blakeslee, W. A. Peer, and A. S. Murphy, "Auxin transport," *Current opinion in plant biology*, vol. 8, no. 5, pp. 494–500, 2005.

[27] R. Benjamins, A. Quint, D. Weijers, P. Hooykaas, and R. Offringa, "The pinoid protein kinase regulates organ development in arabidopsis by enhancing polar auxin transport," 2001.

[28] J. Friml, X. Yang, M. Michniewicz, D. Weijers, A. Quint, O. Tietz, R. Benjamins, P. B. Ouwerkerk, K. Ljung, G. Sandberg, *et al.*, "A pinoid-dependent binary switch in apical-basal pin polar targeting directs auxin efflux," *Science*, vol. 306, no. 5697, pp. 862–865, 2004.

[29] B. B. Amor, S. Wirth, F. Merchan, P. Laporte, Y. d'Aubenton Carafa, J. Hirsch, A. Maizel, A. Mallory, A. Lucas, J. M. Deragon, *et al.*, "Novel long non-protein coding rnas involved in arabidopsis differentiation and stress responses," *Genome research*, vol. 19, no. 1, pp. 57–69, 2009.

[30] F. Emmert-Streib, M. Dehmer, and B. Haibe-Kains, "Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks," *Frontiers in cell and developmental biology*, vol. 2, p. 38, 2014.

[31] E. R. Alvarez-Buylla, M. Benitéz, E. B. Dávila, A. Chaos, C. Espinosa-Soto, and P. Padilla-Longoria, "Gene regulatory network models for plant development," *Current Opinion in Plant Biology*, vol. 10, no. 1, pp. 83–91, 2007.

[32] T. Schlitt and A. Brazma, "Current approaches to gene regulatory network modelling," *BMC bioinformatics*, vol. 8, no. 6, pp. 1–22, 2007.

[33] U. Alon, *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, 2006.

[34] V. I. Arnold, *Ordinary differential equations*. Springer Science & Business Media, 1992.

[35] M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential equations, dynamical systems, and an introduction to chaos*. Academic press, 2012.

[36] S. Grigolon, B. Bravi, and O. C. Martin, "Responses to auxin signals: an operating principle for dynamical sensitivity yet high resilience," *Royal Society open science*, vol. 5, no. 1, p. 172098, 2018.

[37] W. R. Inc., "Mathematica, Version 13.1." Champaign, IL, 2022.

[38] A. K. H. K. Cloos PA, Christensen J, "Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease.," *Genes Dev*, 2008.

[39] B. Y. L. Cedar, H., "Dna methylation and histone modification: patterns and paradigms," *Nat Rev Genet*, no. 10, p. 295–304, 2009.

[40] MATLAB, *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.

[41] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder–mead simplex method in low dimensions," *SIAM Journal on optimization*, vol. 9, no. 1, pp. 112–147, 1998.