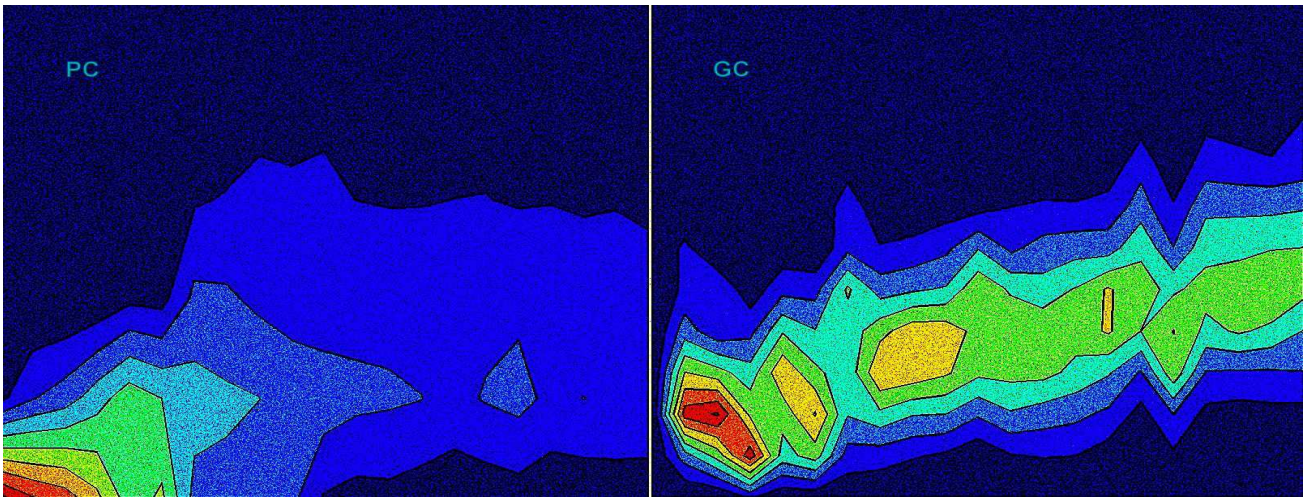
Mathematical Modelling in Systems Biology

Cell cycle regulation during leaf development in *Arabidopsis*

Leila Kheibarshekan



Promoters:

Prof. Dr. W. Govaerts

Prof. Dr. L. De Veylder

Doctor of Science in Mathematics

Academic year 2010–2011

Faculty of Science
Department of Applied Mathematics and
Computer Science, Ghent University

VIB
Department of Plant Systems Biology
Ghent University

UNIVERSITEIT GENT

FACULTY OF SCIENCES

There is a proverb that says ”There is an end point for any starting point”. This is the right place to thank all people who gave me good advice, who taught and educated me, who supported me financially and emotionally in my life and in particular during the completion of this PhD work. Therefore, I would like to thank Professor Willy Govaerts, my promoter whose generous help, guidance, encouragement and support gave me energy to understand and develop this project. He really made this thesis possible.

I wish to express my warm and sincere thanks to Professor Lieven De Veylder, my co-promoter for his kind support and guidance throughout my PhD project.

I am especially grateful to Professor Gerrit T.S. Beemster for his valuable input in this study.

I warmly thank Dr. Tim Lammens, Dr.Veŕonique Boudolf and Stijn Dhondt for their valuable advice and friendly help. Their extensive discussions on my work have been very helpful for this study.

I would like to show also my gratitude to the official referees, Professor Marnix Van Daele and Professor Peter Dawyndt from the Department of Applied Mathematics and Computer Science, Professor Gerrit T.S. Beemster from Plant System Biology (PSB) in VIB, Doctor Roeland Merks from the Netherlands Institute for Systems Biology (NISB) and Centrum Wiskunde and Informatica (CWI) and Professor Andreas Weiermann from the Department of Mathematics, the chairman of the assessment committee, for their detailed review, constructive criticism and excellent advice during the preparation of this thesis.

My deepest gratitude is also due to all my beloved family members, especially my unique, great and kind brother, Hossein, whose support and advice have been invaluable, my dedicated mother who is all the time besides me to support and encourage me and my father who I missed dearly,

God bless his soul. Also my kind-hearted sisters and sister-in-law who are close friends all the time. And most of all to my loving, supportive, encouraging, and patient husband Shahrouz whose faithful support during the final stages of my Ph.D. is so appreciated. Shahrouz, you are always giving me a feeling of warmth, hope and peace. It is so wonderful to have you beside me.

I am also indebted to all members of the Department of Applied Mathematics and computer Science, Ghent University and my colleagues at VIB as well. I especially like to thank my colleagues, Reza Khoshsiar, Neli Stoilova, Stéphanie Vanhove, Virginie De Witte, Patricia Victor, Charlotte Sonck and Elchin Jafarof for their help and consideration.

I am very happy and proud of making wonderful friends in Belgium. Especially I mention my best friends: Jacqueline and Nicole who made me feel Belgium as a second home. Furthermore I thank all Flemish people that I met, for their openness, friendship and hospitality.

Finally, I wish to deeply thank my best Iranian friends in Ghent (including lovely IranianGent group), for helping me get through the difficult times, and for all the emotional support, entertainment, and caring they provided. Someday I hope to give to you a fraction of all you have given me.

For those of you whose names I did not mention here, I offer my apologies but all those that have made this experience so special will never been forgotten. Thanks again to everybody for having been there for me and making my life in Belgium so memorable and pleasurable.

The financial support of the Fund for Scientific Research - Flanders (Belgium) (F.W.O.-Vlaanderen) under grant No. G.0065.07 is gratefully acknowledged.
Ghent, Belgium, February 2011
Leila Kheibarshekan Asl

Contents

Introduction

”Systems biology is the science of discovering, modelling, understanding and ultimately engineering at the molecular level the dynamic relationship between the biological molecules that define living organisms”[1].

In fact, systems biology is an academic field that seeks to integrate different levels of information to understand how biological systems function (Figure 1). System-level understanding, the approach advocated in systems biology [42], requires a shift in our notion of ”what to look for” in biology. A system-level understanding of a biological system can be derived from insight into four key properties: **system structures**, **system dynamics** (how a system behaves over time under various conditions), **the control method** (mechanisms that systematically control the state of the cell) and **the design method** (strategies to modify and construct biological systems having desired properties) [43]. Once we have attained an understanding of network structure, we will be able to investigate network dynamics. For dynamic analysis of a cellular system, we need to create a model. But first it is important to carefully consider the purpose of model building: Whether it is to obtain an in-depth understanding of system behaviour or to predict complex behaviours, we must first define the scope and abstraction level of the model [43].

Modelling in systems biology has three key steps consisting of several

[1] Leroy Hood, President ISB

components. The first step is construction, which includes construction of the model, scheme, constraints, and view components. The second step is analysis, which includes analysis of the model, context, engine, interpretation and input components. Note that the engine that executes the model determines the analytic process. The last step is the validation which includes validation of the model, observation, assumptions, and interpretation [27].

Model validation is possibly the most important step in the model building sequence. It is also one of the most overlooked. Often the validation of a model seems to consist of different statistical tests to see how well the model fits to the experimental data. Validation or verification is done to ensure that the model is programmed correctly and the algorithms have been implemented properly. Also to make sure that the model does not contain errors, oversights, or bugs. We note that validation does not ensure that the model meets a specified set of model requirements and correctly reflects the workings of a real world process. Moreover, no computational model will ever be fully verified, guaranteeing 100% error-free implementation. The question here is how the model can be validated if controlled experiments cannot be performed on the system? or if the real-world system being modelled does not exist? The ultimate goal of model validation is to make the model useful in the sense that the model addresses the right problem, provides accurate information about the system being modelled, and makes the model actually usable. Overall, a systems biology approach means integrating computational and theoretical methods with experimental efforts.

As shown in Figure 2 our research is based on a loop of collaborative processes in three steps (experiments, data handling and discussions, mathematical modelling) shared between mathematicians and biologists. We will discuss this in the following.

At first the biologists collect the experimental data and observations in

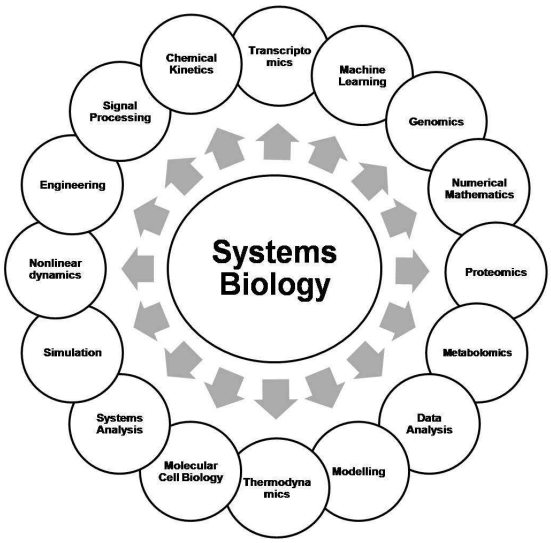


Figure 1: Different aspects of systems biology. To understand the function of biological systems, it is necessary to investigate the different aspects of the systems and study their interaction. This collection of information is called systems biology.

their research lab and discuss with mathematicians to generate a model. Mathematicians generate a model and analyse the experimental data and the results of the model (data handling) based on biological knowledge. In the next stage modellers give feedback to the biologists based on a comparison of their model results and experimental data. Biologists also renew their experiments to gather new lab results to obtain information for updating the mathematical model. This process ends in a satisfactory way if the model predictions correspond with experimental observations.

In this thesis, we studied the mathematical modelling approach of systems biology in plants. We have concentrated on two different issues related to the cell cycle and cell division (especially in the plant *Arabidopsis*) (Figure 3). The first issue is that of the epidermal cell population in the *Arabidopsis* leaf and the second issue deals with gene networks which

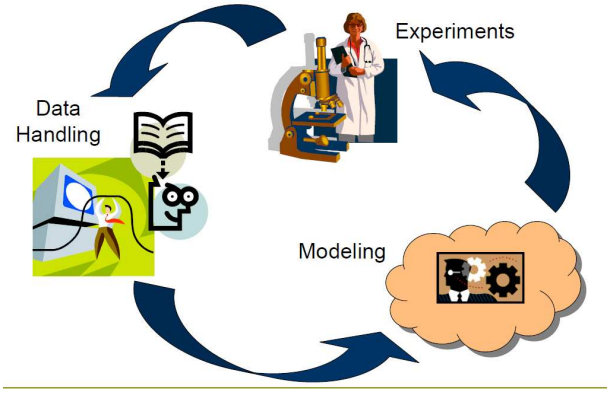


Figure 2: An iterative approach. Experimental biologists make observations and produce the experimental data about phenomena of biological interest. These data are handled and analysed. Afterwards they are used as input for the models. Models yield interpretations that prompt further experimental observations or, when compared with observations, question the validity of the assumptions. This process ends if the model predictions confirm the observations and vice versa. *Pic: Carlo Cosentino, Carnegie Mellon University, Pittsburgh, 2008.*

play an important role during the cell cycle. The chapters are grouped into four parts.

• **Part I**

Chapter 1 is an introduction to plant biology that can be skipped by readers familiar with the basics of this subject. We describe the cell cycle as the series of events that takes place in a cell leading to its division and duplication. One of the key words that will be explained in Chapter 1 is the endocycle process. This process occurs in a wide variety of cell types and is particularly prominent in dicotyledonous plants[2]. The normal cell cycle (mitotic cell cycle) progression and endocycle (endoreduplication) are linked events. Premature or delayed exit from the cell division program often results in an increased or decreased DNA ploidy, respectively. There-

[2]Having two seed leaves.

Figure 3: Arabidopsis. It is a small flowering plant native to Europe, Asia, and northwestern Africa. Arabidopsis is popular as a model organism in plant biology and genetics. Its genome was the first plant genome to be sequenced. Arabidopsis is a popular tool for understanding the molecular biology of many plant traits, including flower development and light sensing. *Pic:http://pested.ifas.ufl.edu/newsletters/september2007/hardtime.htm*

fore, the onset of endoreduplication must be controlled precisely. At the molecular level, endoreduplication is likely to be achieved through elimination of the components needed to progress through mitosis.

In Chapter 2, we state the general objective of the study. We address the various aspects of the problems and the key factors that are assumed to influence or cause the problems.

• **Part II**

We provide a comprehensive mathematical framework in part II (Chapter 3) to be used in the other chapters for the modelling, simulation and analysing purposes. This part can be skipped by mathematically-oriented

readers with a basic knowledge of systems biology. Here we introduce and study Michaelis-Menten kinetics (a model of enzyme kinetics) and the quasi-steady state assumption to reduce the complexity of the model. These concepts are necessary in the modelling approach of gene networks. We also introduce two basic mathematical models for the growth of cell size in plants. The first one considers growing of cells without any limitation, meaning that the cell can grow forever and does not stop growing. The second one puts a limitation on growing in a way that the cell can grow up to a certain size (threshold) and cannot become larger than that. Besides this, some optimization methods including gradient base methods and a derivative-free method are discussed. We will apply these optimization techniques in part III.

• **Part III**

Part III includes Chapter 4 and Chapter 5. In this part we consider two case studies related to the cell cycle and cell division in *Arabidopsis*.

The first case study in Part III (Chapter 4) is the temporal control of epidermal cell divisions in the *Arabidopsis* leaf. The growth of plant organs is the result of two processes acting on the cellular level, namely cell division and cell expansion. The precise nature of the interaction between these two processes is still largely unknown as it is experimentally challenging to disentangle them. The lower epidermal tissue layer of the *Arabidopsis* leaf is composed of two cell types, puzzle-shaped pavement cells and guard cells, which build the stomata that regulate gas exchange in the leaf. We determined the cell number and the individual cell areas separately for both cell types during development. To dissect the rules whereby different cell types divide and expand, the experimental data were fit into a computational model that describes all possible changes a cell can undergo from a given day to the next day. The model allows to calculate the probabilities for a precursor cell to become a guard or pavement cell, the maximum size at which it can divide into two pavement cells or two guard cells, the cell cycle duration and two different growth rates for two kinds of

cells (pavement and guard cells) in one population. Moreover, we utilized mathematical optimization methods to estimate the parameters, applying different optimization techniques.

The second case study in part III (Chapter 5) deals with the fact that atypical E2F activity restrains APC/C^{CCS52A2} function obligatory for endocycle onset. We have demonstrated that the atypical E2F transcription factor E2Fe/DEL1 controls the onset of the endocycle through a direct transcriptional control of APC/C activity. Because E2Fe/DEL1 represses the *CCS52A2* promoter, we hypothesize that its level must drop below a critical threshold to allow sufficient accumulation of *CCS52A2* during late S and G2 phase for cells to proceed from division to endoreduplication. We built a mathematical model to analyse the above hypothesize. Using a fitting approach, a non-linear curve approximation of E2Fe/DEL1 levels in a synchronized cell culture was obtained. An ODE model was built for *CCS52A2* expression in a synchronized cell culture, knowing that E2Fe/DEL1 inhibits the expression of *CCS52A2*. Applying that model to the *CCS52A2* expression in leaf development, the behaviour of *CCS52A2* in a mitotic cell can be written as a function of E2Fe/DEL1 expression during leaf development. The ODE model introduced in Chapter 5, is based on the model used for the binding of ligands to proteins with the help of Hill functions. This mathematical model helps to understand mechanistically how decreasing E2Fe/DEL1 levels can account for the division-to-endoreduplication transition. The model allowed an in-silico visualization of the cell cycle phase dependent relationship between E2Fe/DEL1 and *CCS52A2* in a developmental context.

• **Part IV**

Finally, we discuss some future work to extend the above research. Revealing other important factors that are involved in the systems, can help to improve the existing models. We suggest several ideas in this chapter to design new experiments and increase the value of the models.

The term ”we” is used throughout the text in the thesis to underline the

fact that every single result of the author's work as represented in this thesis was only possible because of the provision of experiments, equipment, materials and scientific input from others. This is also reflected in the long list of people mentioned in the acknowledgement and consistent with the fact that modern research relies on collaboration and teamwork. The results in this thesis have been published in or submitted for publication to specialized scientific journals or proceedings, see [32], [41], [46] and [56].

Chapter 1

Some concepts in plant biology

This chapter has been adapted from [38].

1.1 Cell cycle

The cell cycle (mitotic cycle), or cell-division cycle, is the series of events that takes place in a cell leading to its division and duplication (replication). The cell cycle is one of the most comprehensively studied biological processes, particularly given its importance for growth and development and in many human disorders. Studies on yeast, worms, flies, frogs, mammals and plants have contributed to a kind of universal picture on how the basic cell cycle machinery is regulated. The cell cycle consists of four distinct phases: G_1 phase, S phase (synthesis), G_2 phase and M phase (mitosis). M phase is itself composed of two tightly coupled processes: mitosis, in which the cell's chromosomes are divided between the two daughter cells, and cytokinesis, in which the cell's cytoplasm divides in half forming distinct cells. Activation of each phase is dependent on the proper progression and completion of the previous one.

In G_1 phase (Gap 1) the size of cells increases by synthesis of various enzymes that are required in S phase, mainly those needed for DNA replication. The G_1 checkpoint control mechanism ensures that everything is ready for DNA synthesis. In S phase DNA is replicated. When it is complete, all of the chromosomes have been duplicated. Thus, during this phase, the amount of DNA in the cell has effectively doubled.

During the gap between DNA synthesis and mitosis (G_2 phase), the cell continues to grow. The G_2 checkpoint controls mechanism ensures that everything is ready to enter the M (mitosis) phase and divide. Inhibition of protein synthesis during G_2 phase prevents the cell from undergoing mitosis. In M phase, during mitosis, the cell's chromosomes are divided between the two daughter cells, and during cytokinesis, the cell's cytoplasm divides in half forming distinct cells (Figure 1.1). After cell division, each of the daughter cells begins the interphase G_1 phase of a new cycle.

Polyploidy

Polyploidy occurs in cells and organisms when there are more than two paired sets of chromosomes in the nucleus. Most organisms are normally diploid, meaning they have two sets of chromosomes - one set inherited from each parent. Polyploidy may occur due to abnormal cell division in meiosis[1]. It is most commonly found in plants.

Endocycle

The endocycle represents an alternative cell cycle that is activated in various developmental processes. In endocycling cells, the mitotic cell cycle exit is followed by successive doublings of the DNA content, resulting in polyploidy. In plant endocycles are associated with the increase in cellular

[1]Meiosis is the type of cell division by which germ cells (eggs and sperm) are produced. Meiosis involves a reduction in the amount of genetic material.

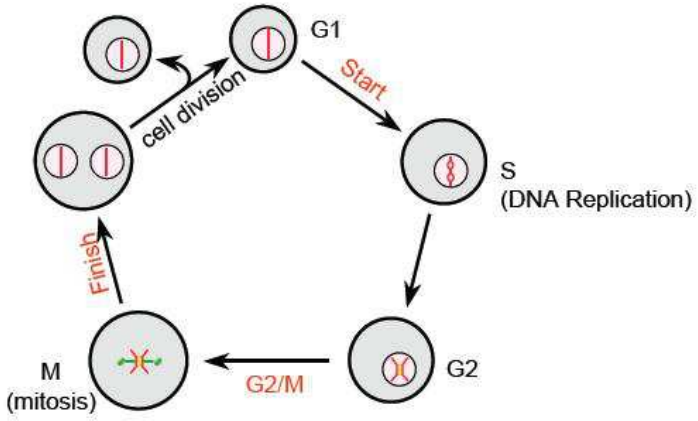


Figure 1.1: The cell division cycle. The cell cycle is divided into four phases. The cell grows continuously in interphase, which consists of three phases; G1, S, and G2. DNA is replicated in S phase. G1 is the gap between M phase and S phase, and G2 is the gap between S phase and M phase. During M phase, the nucleus divides first, in a process called mitosis; then the cytoplasm divides, in a process called cytokinesis. Checkpoints in the cellcycle control system ensure that key processes in the cycle occur in the proper sequence. The new cells called daughter cells can divide again by going through another cell cycle. *The figure is redrawn from [83].*

volume that drives most of plant organ growth. Hence, the increased or reduced growth may result from effects on mitotic or endocycles, or both.

1.2 Control of the cell cycle

Three principal checkpoints control the cell cycle. The G1 checkpoint makes the key decision as to whether the cell should divide, delay division or enter a resting stage. The G2 checkpoint assesses the success of DNA replication and triggers the start of mitosis (M) phase. If this checkpoint is passed, the cell initiates the beginning of mitosis. The accuracy of mitosis is assessed at the G2/M checkpoint. This checkpoint occurs dur-

ing metaphase, and triggers the exit from mitosis and cytokinesis and the beginning of G1. The progression of each checkpoint depends on multiple regulatory mechanisms, including reversible protein phosphorylation, the interactions of proteins and protein degradation. Phosphorylation of protein is one of the major mechanisms which control cell cycle progression. The role of different kinases, and particularly the family of cyclin-dependent protein kinases (CDKs), is important for phosphorylation.

At the G2 checkpoint, CDKs phosphorylate proteins that carry the cycle past the checkpoint into mitosis. During G2, the cell gradually accumulates the G2 cyclin. The cyclin binds to CDK to form a complex called MPF (mitosis promoting factor). When the level of MPF exceeds the threshold necessary to trigger mitosis, the G2 phase ends and mitosis begins. One of the functions of MPF is to activate proteins that destroy cyclin. As mitosis proceeds to the end of metaphase, CDK levels stay relatively constant, but the G2 cyclin is degraded, causing progressively less MPF to be available and initiating the events that end mitosis. After mitosis, the gradual accumulation of new cyclin starts the next turn of the cell cycle. The G1 checkpoint is thought to be regulated in a similar fashion. The level of G1 cyclin increases and associates with cyclin-depend kinase (CDK). Eventually, a threshold ratio that triggers the next round of DNA replication is reached. The cyclin is degraded and the cycle begins again.

1.3 Cyclins and CDKs

A cyclin-dependent protein kinase (CDK) is activated by association with a cyclin, forming a cyclin-dependent kinase complex to drive the cell cycle by phosphorylating key target proteins. It is required for cells to progress to the next phase of the cell cycle. CDKs are also involved in the regulation of transcription and mRNA processing. CDK proteins can be categorized into eight classes: CDKA to CDKG and the CDK like kinases (CKLs), as classified for Arabidopsis [57]. According to the amino acid sequence

similarities 152 CDKs from 41 plant species has been identified. The recent genome-wide transcript profiling of the core *Arabidopsis* cell cycle via Affymetrix[2] microarrays confirmed that most CDK-related kinase genes were relatively constantly expressed in synchronized cells.In contrast, the concentration of Cyclins, the partner of CDKs, varies in a cyclical fashion during the cell cycle. They are produced or degraded as needed in order to drive the cell through the different stages of the cell cycle. When concentrations in the cell are low, cyclins dissociate from CDK, thus they stop to activate. Cyclins themselves have no enzymatic activity. There are several different cyclins that are active in different parts of the cell cycle and that cause the CDK to phosphorylate different substrates. A cyclin-CDK complex can be regulated by several kinases and phosphatases, such as WEE1, and CDK-activating kinase (CAK). CAK adds an activating phosphate to the complex, while WEE1 adds an inhibitory phosphate. The plant CDK inhibitors such as the Kip-related proteins (ICK/KRPs) inhibit cell cycle progression through acting as pseudo-substrates[3] or by inducing conformational changes of the CDK/cyclin complex.

[2]Affymetrix is a manufacturer of DNA microarrays, based in Santa Clara, California, United States. The company was founded by Dr. Stephen Fodor in 1992.

[3]Any substance that mimics the substrate of an enzyme and thus inhibits its activity

Chapter 2

Objectives of the research

The cell cycle is the sequence of events by which a cell replicates its genome and distributes the copies evenly to two daughter cells. In most cells, the DNA replication-cell division cycle is coupled to the duplication of all other components of the cell, so that the interdivision time of the cell is identical to its mass doubling time; therefore the cell mass is often taken as a crucial parameter of the process [18, 61]. Usually mass growing is slower than DNA synthesis; therefore temporal gaps are inserted in the cell cycle between the phase of DNA synthesis and the phase of cell division (mitosis). The cell cycle has surveillance mechanisms that monitor progress through the cell cycle and stop the cell at crucial checkpoints, if necessary, so that DNA synthesis and cell division do not get out of order. These processes are regulated by a complex network of interacting genes and proteins. The cell-cycle regulation in plant has been studied experimentally for several years. It was proposed already in 1990, and since confirmed by experimental studies, that the cell division cycle is controlled by a common set of proteins that interact according to essentially the same rules [79]. Nevertheless, each species uses its own mix of proteins and interactions and generates its own type of growth and division. Recently, mathematical models are emerging in some of the best

understood cases which include fission yeast, budding yeast, frog eggs and mammalian cells. The generic features of the cell cycle control concern the common genes and proteins and the general dynamical principles of their interaction. However, the parts of the common machinery that function can be different in each type of cells. In addition, the function of cell division in the growth process of a multicellular organ(ism) is important. In this case, the role of individual cells is a controversial issue. From one point of view it considers cells as elemental building blocks of the organism and considers growth as a direct consequence of an increase in cell numbers. According to the organismal theory, cells are merely compartments of organismal space, the production of cellular compartments is imposed by the growth of the organ as a whole rather than the inverse [6, 82].

The aim of the research is to apply mathematical models to carry out detailed studies of the cell cycle control during the development of plant organs. Moreover, the growth of plant organs is the result of two processes acting at the cellular level, being cell division and cell expansion. Studying the precise nature of the interaction between these two processes is another challenge in this research. To address this issue we construct and apply a mathematical model to study a population of leaf epidermal cells consisting of a variety of cell types (Chapter 4). Each cell type has its own characteristics and different lineage to be created. Our goal is the study of cell division and cell growth of these cell types during leaf development of *Arabidopsis* in which different parameters have important roles. To study experimentally the behaviour of each cell type during leaf development, biologists must place them under a microscope and follow them every few hours to extract the necessary information from their images. Due to the lack of tools and techniques available so far, it is not possible to follow the behaviour of cells during a long period of time under the microscope. There are several reasons for this problem. One reason is that it is not possible to specify all cells, large and small, and study their behaviour under the microscope. Even in the rare cases, when a small area of leaf is taken

into account, and cells can be identified by markers, due to the growth of the cells, their place will change after a while and some of the cells will be out of the region of microscope view. Another reason is that fixing the leaf on a flat surface changes the growth conditions. Therefore, it is impossible to study experimentally the dynamics of each cell type over a long period of time. One goal of our research is to build a mathematical model with several parameters that are relevant to plant growth and development for the population of different kinds of cells. Some of the fundamental questions in this study are: does the cell size control cell cycle mechanisms? If so, do any thresholds exist for the cell size so that above the maximum threshold and below the minimum threshold the cell cannot divide? Is the average cell cycle duration fixed throughout leaf development? The mathematical model built for the above system helps to answer the above questions. Although at first glance it looks like a very complicated model, its basic structure is simple. In this model, all conditions of growth and cell division are considered. Finally, this is a general model which can be applied for all similar populations.

A part of the research on the cell cycle consists of studying networks of proteins and genes which have a great impact on the cell cycle process. A simple network is presented in Chapter 5 which deals with the onset of endoreduplication. Mitotic cell cycle progression and endoreduplication are linked events. Premature or delayed exit from the cell division program results in an increased or decreased DNA ploidy, respectively. Therefore, the onset of endoreduplication must be controlled precisely. At the molecular level, endoreduplication is likely to be achieved through elimination of the components needed to progress through mitosis. In [47], it became clear that the atypical E2F transcription factor, E2Fe/DEL1, accounts for the onset of the endocycle by regulating the temporal expression of *CCS52A2* gene during the cell cycle in a development-dependent manner.

Again, experimental studies of the behaviour of these two components during leaf development are nearly impossible because, although the cell

culture can be synchronized experimentally, endoreduplication cannot be triggered in *Arabidopsis* cell cultures. Inversely, cell division cannot be synchronized experimentally in a developing leaf. Therefore by combining molecular and computational techniques, it is possible to obtain results which help to explain the function of each component of the network in the onset of endoreduplication. The experimental observations in a synchronized cell culture display the inhibition of *CCS52A2* expression by E2Fe/DEL1. Therefore an Ordinary Differential Equation (ODE) model for *CCS52A2* gene expression was built, taking into account that E2Fe/DEL1 inhibits *CCS52A2* expression. Next to the above biological goals of this research, applying and testing various methods for modelling is a general goal of this research. Functions and parameters which can be used in any model are considered. The number of parameters is deliberately restricted to a minimum.

Chapter 3

Mathematical modelling tools

This chapter provides some mathematical modelling tools related to our work for the reader. This mathematical background is needed for understanding the rest of this thesis.

3.1 Hill functions

In the modelling of biological processes, sigmoid functions are often used. A sigmoid function is a monotonic, S-shaped function which raises from zero and approaches one. A Hill function is a particular case of a sigmoid function and is defined by

$$H^+(x) = \frac{x^n}{T^n + x^n} \quad (x \geq 0). \tag{3.1}$$

The parameter $T > 0$ is called a threshold, and the power $n > 0$ is called a Hill coefficient. Hill functions arise naturally in Michaelis-Menten kinetics, cf. §3.3 and in the binding of ligands to proteins, cf. §3.4. Several Hill functions are presented in Figure 3.1A. For $x < T$, the value of $H^+(x)$ is less than 0.5 and for small x it is c zero. For $x > T$ the Hill function is above 0.5 and for large x it is near 1. The function is one-to-one and

monotone and maps the interval $[0, \infty)$ onto $[0, 1)$. The first derivative of the Hill function is given by

$$\frac{d}{dx}H^+(x) = \frac{nx^{n-1}T^n}{(T^n + x^n)^2}, \tag{3.2}$$

which implies that

$$\frac{d}{dx}H^+(x) \mid_{x=T} = \frac{n}{4T}. \tag{3.3}$$

So the parameter n controls the slope of the Hill function at $x = T$. The larger n, the steeper the slope.

To find the inflection points of the Hill function, we consider the second derivative

$$\frac{d^2}{dx^2}H^+(x) = -\frac{nx^{n-2}T^n[(1-n)\,T^n + (n+1)\,x^n]}{(T^n + x^n)^3}. \tag{3.4}$$

$$\tag{3.5}$$

It follows that $\frac{d^2}{dx^2}H^+(x) = 0$, for

$$\begin{aligned} x &= \sqrt[n]{\tfrac{n-1}{n+1}}T,\ x = 0, && \text{for } n > 2, \\ x &= \sqrt[n]{\tfrac{n-1}{n+1}}T, && \text{for } 1 < n \leq 2. \end{aligned} \tag{3.6}$$

For $0 < n \leq 1$ the function has no inflection point. For $n = 1$, its profile is a branch of a hyperbola.
For $n > 1$, $x = T$ is not the inflection point. However for large n, the factor $\sqrt[n]{\frac{n-1}{n+1}}$ is near 1 and so $x = T$ approximately becomes the inflection point.

On some occasions, we need to define the Hill function as a decreasing function, which maps the interval $[0, \infty)$ onto $[1, 0)$. It is then defined as

$$H^-(x) = 1 - H^+(x) = \frac{T^n}{T^n + x^n}. \tag{3.7}$$

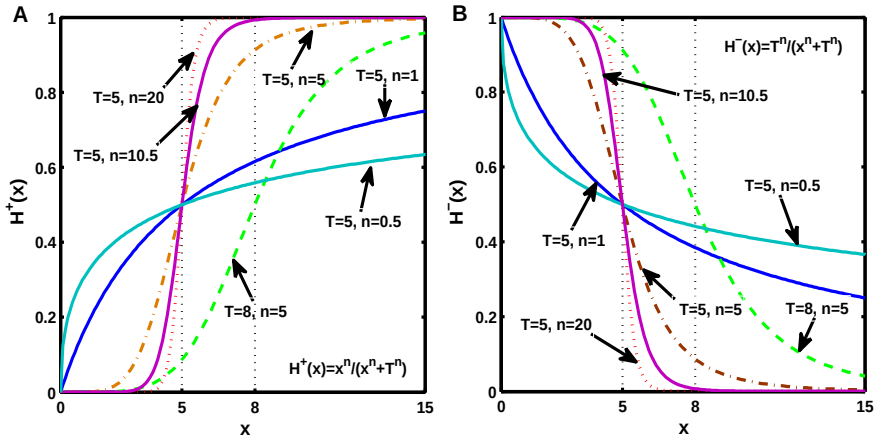


Figure 3.1: (A,B) Hill functions. T is the threshold and n is the Hill coefficient. Increasing n increases the steepness of the curve.

In this case, for x near zero the value of the function is near 1 and for large x it is near zero (Figure3.1B). We note that, to reduce the number of operations and the possibility of overflow and underflow, one can use other forms of formulas (3.1) and (3.7):

$$H^+(x) = \frac{1}{1+\left(\frac{T}{x}\right)^n}, \qquad H^-(x) = \frac{1}{1+\left(\frac{x}{T}\right)^n}. \tag{3.8}$$

In general, Hill functions are useful for detailed modelling. The essence of these functions is the transition between low and high values, with a characteristic threshold T.
Sometimes, it is useful to use other functions that capture the essential behaviour of these functions. The functions following

$$F^+(x) = \frac{1+\tanh(m(x-T))}{2}, \tag{3.9}$$

$$F^-(x) = \frac{1-\tanh(m(x-T))}{2}. \tag{3.10}$$

have a behaviour similar to those in (3.1) and (3.7). The main advantage is that (3.9) and (3.10) are defined for $x \in (-\infty, \infty)$ while (3.1) and (3.7) are defined only for $x \in [0, \infty)$ (Figure 3.2).
Here the parameter m controls the steepness and the parameter T is the threshold. The graphs of these functions are already very steep for $m = 1$ (Figure 3.3).

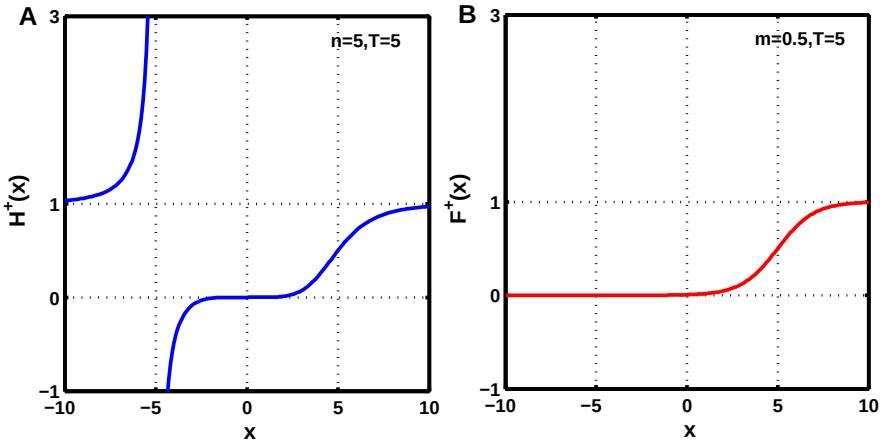


Figure 3.2: (A,B) The Hill function and Tangent hyperbolic function . For negative values of x the Hill function is not applicable while the $tanh$ function can still be used.

In general, the graph of $F^+(x)$ is very close to that of $H^+(x)$ for carefully chosen values of n for which $n \gg m$. This is illustrated in Figure 3.4 for $T = 5$ and the conclusions $(m, n) = (1, 10), (2, 20), (3, 30) and (4, 40)$. We conclude that, for $n > 10$, replacing m by $n/10$ in the $\tanh$ function gives approximately the same graph as the Hill function (Figure 3.3). We note that the graphs of these two functions for small n are less similar (Figure 3.5).

However, the execution time (in elapsed CPU seconds) is quite dif-

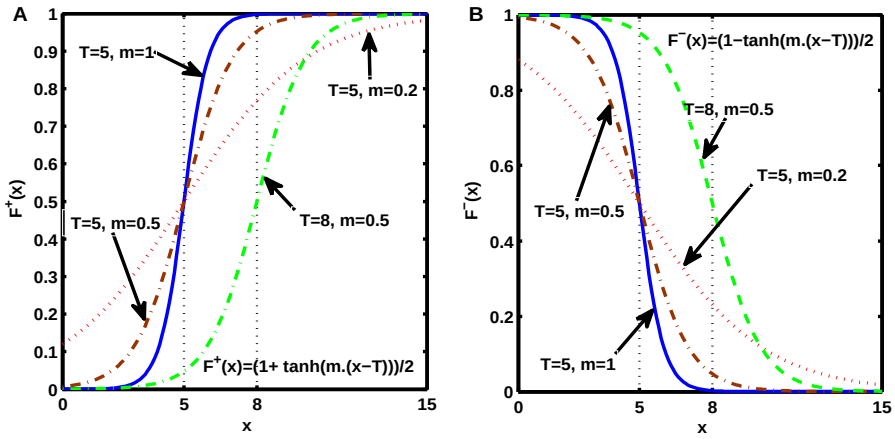


Figure 3.3: (A,B)Tangent hyperbolic functions. T is the threshold and m controls the steepness of the curve.

ferent. For a vector $x = (1, 1.1, 1.2, 1.3, \ldots, 500)$ where $m = 2, n = 20, T = 5$, the computation of the Hill function H^+ in (3.1) and (3.8) takes 0.0045 and 0.0027 seconds, respectively while for $\tanh$, it takes only $3.7128e-004$ seconds which is almost 12 times less than the Hill function in (3.1) and 7 times less than the Hill function in (3.8). A similar result can be observed for $m = 2.05, n = 20.5, T = 5$. The computation time for the Hill function in (3.1), (3.8) and $\tanh$, are $0.0036, 0.0024$ and $2.2210e-004$, respectively. Therefore, for large n, the $\tanh$ function is time efficient. For small n, e.g. $n = 2$ and $m = 0.2$ the time of computation is $1.5225e-004$ for the Hill function in (3.1) and $4.8610e-005$ for $\tanh$ which differ by only a factor 3. These computations were executed under Windows XP using Matlab R2008b.

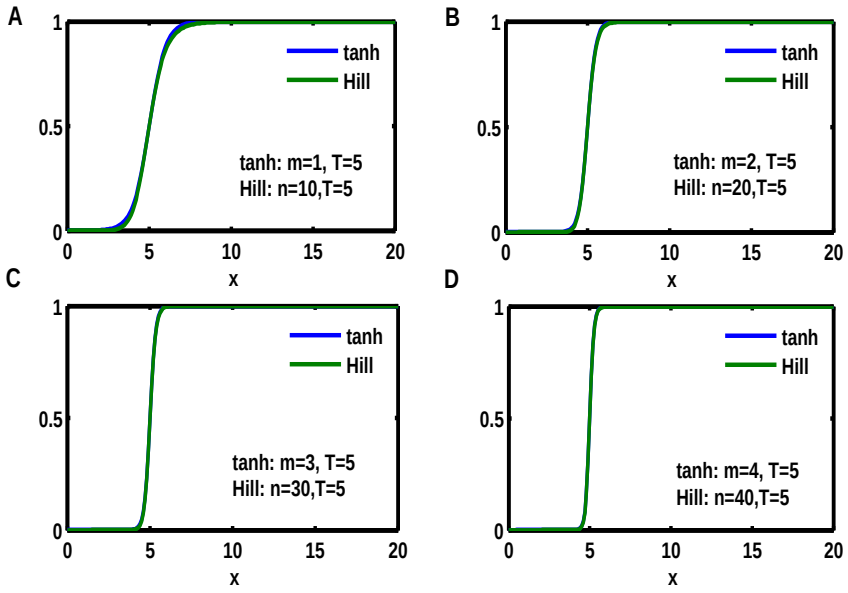


Figure 3.4: Hill function vs. $\tanh$ function for values $\mathbf{n \gg m}$.

3.2 The law of mass action

The mass action law, which is a key assumption in biochemical kinetics, was introduced by Guldberg and Waage (1864 - 1879). It states that the reaction rate (i.e., the change of concentration of reactant per time t) is proportional to the product of the concentrations of the participating molecules (reactants). For instance, in a simple reaction,

$$S_1 + S_2 \xrightarrow{k} P, \tag{3.11}$$

S_1 and S_2 are two substances which take part in a reaction and produce the product P. According to the law of mass action, the rate of reaction (v) is given by

$$\frac{d[P]}{dt} = v = k[S_1][S_2], \tag{3.12}$$

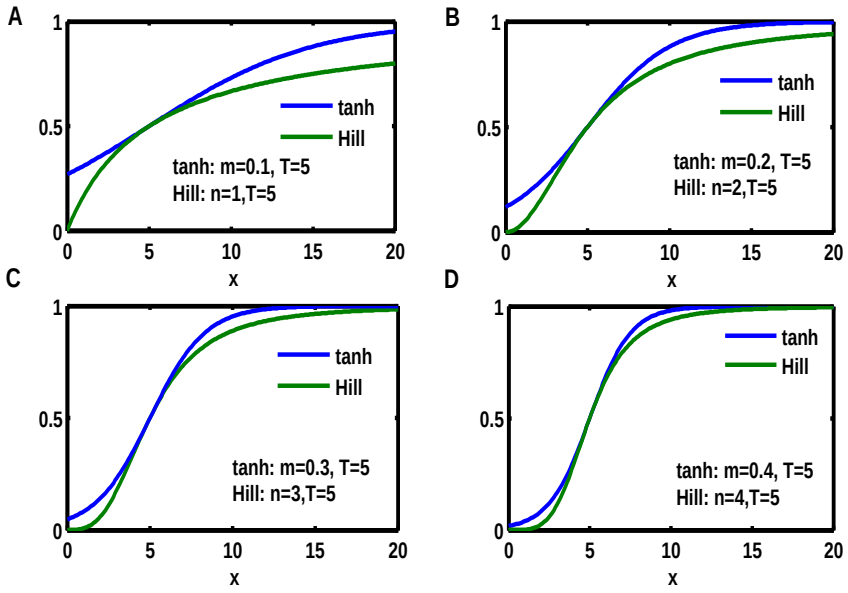


Figure 3.5: Hill function vs. $\tanh$ **function for small** $\mathbf{n}$**.**

$$\frac{d[S_1]}{dt} = \frac{d[S_2]}{dt} = -v = -k[S_1][S_2], \tag{3.13}$$

where $[S_1], [S_2]$ and $[P]$ are the concentrations of S_1, S_2 and P, respectively. The right-hand side of (3.12) is the reaction rate for the product and is positive, since the concentration of the product increases after each reaction. The right-hand side of (3.13) is the reaction rate of the reactant and is negative due to the decrease of the reactants when the reaction takes place. k is the proportionality factor, the so-called rate constant. If we measure the concentration in moles per litre ($molL^{-1}$ or M) and the time in seconds (s), then the unit of reaction rate is $M.s^{-1}$ and the unit of the rate constant is $\frac{1}{Ms}$.

In a two-sided reaction, the two substances take part in the forward reaction (association) and produce the product; the product also takes part in the backward reaction (dissociation) and increases the reactant concentra-

tion. In the following reaction

$$S_1 + S_2 \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} 2P, \tag{3.14}$$

the reaction rate is calculated by

$$v = v_1 - v_{-1} = k_1[S_1][S_2] - k_{-1}[P]^2. \tag{3.15}$$

v is the net rate, v_1 is the rate of forward reaction and v_{-1} is the rate of backward reaction. k_1 and k_{-1} are the rate constants for forward and backward reactions, respectively [44].

3.3 Michaelis-Menten kinetics

Brown [1](1902) proposed the first enzymatic mechanism for the following reaction:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \xrightarrow{k_2} E + P. \tag{3.16}$$

Enzyme E binds to the substrate S and forms the complex ES, which is subject to an irreversible reaction that releases the product P from the enzyme. k_1, k_2 and k_{-1} are the rate constants for forward and backward reactions in (3.16), respectively. We note that the total enzyme concentration (which contains free enzyme and the intermediate complex ES) does not change over time, so $[E_{total}] = [E] + [ES] = constant$. Enzyme is needed in the reaction but is not consumed in the end-product. According to the mass action law, we have

$$\frac{d[S]}{dt} = -k_1[E][S] + k_{-1}[ES], \tag{3.17}$$

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_2[ES], \tag{3.18}$$

[1]Adrian John Brown (1852-1919) was a British Professor at the University of Birmingham and a pioneer in the study of enzyme kinetics.

$$\frac{d[E]}{dt} = -k_1[E][S] + k_{-1}[ES] + k_2[ES], \tag{3.19}$$

$$\frac{d[P]}{dt} = k_2[ES]. \tag{3.20}$$

For simplicity, Michaelis and Menten (1913) assumed that the composition of E and S and decomposition of ES into E, S are much faster than the decomposition of ES into E and P ($k_2 \ll k_{-1}$). This assumption is called a *quasi-equilibrium* assumption. This means that $k_2[ES]$ is negligible and the concentration of ES is in the equilibrium. In a chemical process, the chemical equilibrium is the state in which the chemical activities or concentrations of the reactants and products have no change over time. Usually, this would be the state that results when the forward chemical processes proceed at the same rate as the reverse reaction. The reaction rates of the forward and backward reactions are generally not zero but, being equal, there are no changes in any of the reactant or product concentrations. It implies that

$$k_1[E][S] = k_{-1}[ES] \Rightarrow \frac{d[ES]}{dt} = 0 \Longrightarrow [ES] = \frac{k_1[E][S]}{k_{-1}}, \tag{3.21}$$

$$[ES] = [E_{total}] - [E],$$

$$\Longrightarrow [ES] = \frac{[E_{total}][S]}{[S] + \frac{k_{-1}}{k_1}}. \tag{3.22}$$

Therefore the reaction rate $V = \frac{d[P]}{dt}$ will be

$$V = \frac{V_{max}[S]}{[S] + K_m}, \tag{3.23}$$

where the parameters V_{max} and K_m are given by

$$V_{max} = k_2[E_{total}], \tag{3.24}$$

$$K_m = \frac{k_{-1}}{k_1}. \tag{3.25}$$

Briggs and Haldane (1925) made another assumption for the enzyme kinetic. They assume that, in the beginning of the reaction when the concentration of substrate S is much larger than the concentration of enzyme E, the reaction reaches a state where the concentration of the complex ES remains constant. This assumption is called the *quasi-steady state* assumption. Using (3.18) and (3.19) we have,

$$\frac{d[ES]}{dt} + \frac{d[E]}{dt} = 0. \tag{3.26}$$

This implies that $[ES] + [E] = c_0 + e_0$ where c_0 and e_0 are the concentrations of ES and E, respectively, at the starting point. In the beginning $c_0 = 0$, therefore $[E] = e_0 - [ES]$. We can now substitute $e_0 - [ES]$ for $[E]$ into the equations (3.17) and (3.18), giving us a system of two equations:

$$\frac{d[S]}{dt} = k_{-1}[ES] - k_1(e_0 - [ES])[S], \tag{3.27}$$

$$\frac{d[ES]}{dt} = k_1(e_0 - [ES])[S] - (k_{-1} + k_2)[ES]. \tag{3.28}$$

We can simplify those equations by a simple assumption. The key property of enzymes is that they are effective at very low concentration. So we can assume that e_0 is small, thus $[ES]$ is small and also $[ES] \leq e_0$. By this assumption we can define a new variable v,

$$v = \frac{[ES]}{e_0} \Rightarrow [ES] = e_0 v. \tag{3.29}$$

Then, according to (3.28) we have

$$\begin{aligned} \frac{dv}{dt} = \frac{1}{e_0}\frac{d[ES]}{dt} &= k_1(1 - \frac{[ES]}{e_0})[S] - (k_{-1} + k_2)\frac{[ES]}{e_0} \\ &= k_1(1 - v)[S] - (k_{-1} + k_2)v. \end{aligned} \tag{3.30}$$

The other equation (3.27) will be

$$\frac{d[S]}{dt} = k_{-1}e_0 v - k_1(e_0 - e_0 v)[S]$$

$$= \quad e_0(k_{-1}v - k_1(1-v)[S]). \tag{3.31}$$

Since e_0 is small, (3.31) indicates that $[S]$ changes much more slowly. So in (3.30) we can proceed as if $[S]$ were constant. It implies that

$$\frac{dv}{dt} = k_1[S] - (k_{-1} + k_2 + k_1[S])v, \;\; v(0) = 0 \tag{3.32}$$

$$\Rightarrow v = \frac{k_1[S]}{k_{-1} + k_2 + k_1[S]} + \frac{e^{-(k_{-1}+k_2+k_1[S])t}k_1[S]}{-(k_{-1} + k_2 + k_1[S])}. \tag{3.33}$$

We easily see that v approaches $\frac{[S]}{[S]+K_m}$ when t tends to infinity, where $K_m = \frac{k_{-1}+k_2}{k_1}$. By (3.29) $[ES]$ tends towards $\frac{e_0[S]}{[S]+K_m}$ which is called a *stable equilibrium point*.

Now, we go back to equation (3.20) which shows that

$$\frac{d[P]}{dt} = k_2[ES]. \tag{3.34}$$

Then, according to the above results, we find that the reaction rate $V = \frac{d[P]}{dt}$ as,

$$V = \frac{V_{max}[S]}{[S] + K_m}, \tag{3.35}$$

where the parameter,

$$V_{max} = k_2 e_0 = k_2[E_{total}] \tag{3.36}$$

is called *the maximal velocity*, which is the maximal rate of reaction (when the enzyme is completely saturated with substrate). The parameter K_m is called the *Michaelis constant* and given by $K_m = \frac{k_{-1}}{k_1}$ in the *quasi-equilibrium* assumption and $K_m = \frac{k_{-1}+k_2}{k_1}$ in the *quasi-steady state* assumption.
(3.23) and (3.35) are Hill functions with $n = 1$. Figure 3.6 shows that increasing the substrate does not always increase the reaction rate in the

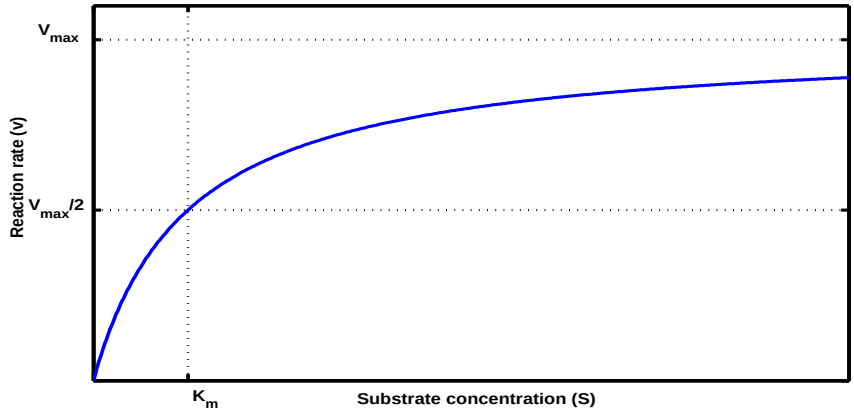


Figure 3.6: Michaelis-Menten kinetics. V_{max} is the maximal reaction rate which is approximated when the substrate concentration is large. When the substrate concentration equals K_m, half this maximal rate is obtained.

same way. When the concentration of the substrate exceeds the threshold K_m, the reaction rate increases slowly.

Normally, the requirements for applying the steady state approximation are laxer than for the equilibrium assumption. The concentration of the intermediate is only needed to be low and more or less constant (as seen, this has to do only with the rates at which it appears and disappears) but it is not needed to be at equilibrium, which is usually difficult to prove and involves heavier assumptions. This approximation can be used to yield similar results as the *quasi-equilibrium* approximation.

A more illustrative version of the *Michaelis-Menten* equation is the *Lineweaver-Burk*[2] equation,

$$\frac{1}{V} = \frac{K_m}{V_{max}} \frac{1}{[S]} + \frac{1}{V_{max}}. \tag{3.37}$$

[2]Hans Lineweaver and Dean Burk (1934)

The *Lineweaver-Burk* equation provides a line with a slope $\frac{K_m}{V_{max}}$ and y-intercept $\frac{1}{V_{max}}$. The x-intercept, a theoretical point since $\frac{1}{[S]}$ cannot be negative, is $\frac{-1}{K_m}$. The *Lineweaver-Burk* plot (Figure 3.7) was widely used to determine important terms in enzyme kinetics, such as K_m and V_{max}, before the wide availability of powerful computers and non-linear regression software. It also gives a quick, visual impression of the different forms of enzyme inhibition. The *Lineweaver-Burk* plot is classically used in older

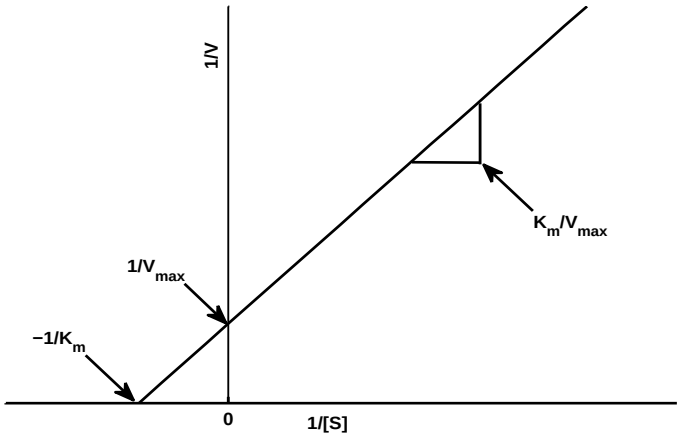


Figure 3.7: Lineweaver-Burk plot.

texts, but is prone to error, as the y-axis takes the reciprocal of the rate of reaction increasing any small errors in measurement. Also, most points on the plot are found far to the right of the y-axis (due to limiting solubility not allowing for large values of $[S]$ and hence no small values for $\frac{1}{[S]}$), calling for a large extrapolation back to obtain x and y intercepts.

3.4 Binding of ligands to proteins and cooperativity

Hill functions also appear when considering the equilibrium binding of the ligands to the binding sites of a protein. The first derivation of the Hill function was based on experimental results concerning the binding of oxygen to hemoglobin (Hb) (Hill 1910, 1913). In 1904, Bohr and co-workers found that the plot of the fractional saturation of Hb with oxygen against the oxygen partial pressure had a sigmoid shape [10]. Hill explained this with interactions between the binding sites located at the hemoglobin subunits. Hill assumed complete cooperativity and predicted an experimental Hill coefficient [36].

A ligand is a signal triggering molecule that binds to a binding site on a target protein to serve a biological purpose. A binding site is a region on a protein, DNA, or RNA to which ligands, or more specifically, protein ligands form a chemical bond. A protein can have more than one binding site and different types of ligands can bind to the binding sites. The term saturation refers to the fraction of total binding sites that are occupied at any given time. When more than one type of ligand binds to the binding sites, competition ensues.

In this section we consider binding of the ligands to monomer[3] and oligomer[4] proteins. Let S be a ligand and E be a protein with one binding site,

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \tag{3.38}$$

where k_1 and k_{-1} are the constants of binding and dissociation, respectively. By the law of mass action we have,

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES]. \tag{3.39}$$

[3] a small molecule that may chemically bind to other monomers to form a polymer.
[4] polymers with a relatively low number of units.

An equilibrium can exist between unbound ligands and bound ligands meaning that concentrations of the unbound ligands and bound ligands do not change over time. We then have $\frac{d[ES]}{dt} = 0$, and

$$\frac{[ES]}{[E_{total}]} = \frac{[S]}{K_m + [S]}, \tag{3.40}$$

where (3.40) is the fractional of saturation or probability of binding and

$$[E_{total}] = [E] + [ES], \tag{3.41}$$

$$K_m = \frac{k_{-1}}{k_1}. \tag{3.42}$$

Now consider a dimeric protein with two identical binding sites. Assume that the binding to the first ligand promotes the binding to the second ligand:

$$E_2 + S \xrightarrow{slow} E_2S \tag{3.43}$$

$$E_2S + S \xrightarrow{fast} E_2S_2, \tag{3.44}$$

where E_2 is a dimer. The fractional saturation is given by

$$f = \frac{\text{Total bound sites}}{\text{Total sites}} = \frac{[E_2S] + 2[E_2S_2]}{2[E_{2,total}]}. \tag{3.45}$$

$[E_{2,total}]$ is given by

$$[E_{2,total}] = [E_2] + [E_2S] + [E_2S_2]. \tag{3.46}$$

If the affinity to the second ligand is strongly increased by binding to the first ligand, then E_2S will react with S as soon as it is formed and the concentration of E_2S can be neglected. We assume that the ligands bind simultaneously (unrealistic), thus (3.43) and (3.44) are reduced to

$$E_2 + 2S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} E_2S_2. \tag{3.47}$$

Again, the existence of an equilibrium between bound and unbound ligands implies that

$$k_1[E_2][S]^2 = k_{-1}[E_2S_2]. \tag{3.48}$$

Now, the fractional saturation is given by

$$f = \frac{2[E_2S_2]}{2[E_{2,total}]}, \tag{3.49}$$

where $[E_{2,total}] = [E_2]+[E_2S_2]$. We use (3.48) to simplify (3.49) as follows

$$f = \frac{[E_2S_2]}{[E_2] + [E_2S_2]} = \frac{[S]^2}{K_m^2 + [S]^2}. \tag{3.50}$$

Here, K_m^2 equals $\frac{k_{-1}}{k_1}$.
Generally, for a protein with n binding sites it holds that

$$E_n + nS \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} E_nS_n. \tag{3.51}$$

Then, the collision rate is $k_1[E_n][S]^n$ and

$$\frac{d[E_nS_n]}{dt} = k_1[E_n][S]^n - k_{-1}[E_nS_n]. \tag{3.52}$$

By the *quasi- equilibrium assumption* we have

$$k_1[E_n][S]^n = k_{-1}[E_nS_n] \Longrightarrow \quad \frac{[E_nS_n]}{[E_{n,total}]} = \frac{[S]^n}{K_m^n + [S]^n} \tag{3.53}$$

where, $[E_{n,total}] = [E_nS_n] + [E_n]$ and we define the constant K_m by

$$K_m^n = \frac{k_{-1}}{k_1}. \tag{3.54}$$

(3.53) is the general form of the Hill function. Plotting the fractional of the saturation $\frac{[E_nS_n]}{[E_{n,total}]}$ versus the concentration of ligands $[S]$ yields a sigmoid

curve with the threshold point at K_m [44]. When $n = 1$, we obtain the Michaelis-Menten equation.

The Hill coefficient, n, describe the degree of cooperativity. Reactions with Hill coefficients $n > 1$ are often called reactions with positive cooperativity. Reactions with Hill coefficients $n < 1$ are called reactions with negative cooperativity. If $n = 1$, the equation reverts to a simple hyperbolic response which has no cooperativity.
Moreover, the probability of unbinding is given by:

$$1 - \frac{[E_n S_n]}{[E_{n,total}]} = \frac{K_m^n}{K_m^n + [S]^n}. \tag{3.55}$$

Some researchers feel that the above model (3.53) is so unrealistic that the Hill equation should be considered as an empirical result. Instead, they derive the system of ODEs from

$$E_2 + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} E_2 S, \tag{3.56}$$

$$E_2 S + S \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} E_2 S_2, \tag{3.57}$$

for the dynamics of the reactions to find an alternative model for the fractional saturation. The system is given by

$$\begin{aligned}
\frac{d[E_2]}{dt} &= -k_1[E_2][S] + k_{-1}[E_2 S], && (3.58)\\
\frac{d[S]}{dt} &= -k_1[E_2][S] + k_{-1}[E_2 S] - k_2[E_2 S][S] \\
&\quad + k_{-2}[E_2 S_2], && (3.59)\\
\frac{d[E_2 S]}{dt} &= k_1[E_2][S] - k_{-1}[E_2 S] - k_2[E_2 S][S] \\
&\quad + k_{-2}[E_2 S_2], && (3.60)
\end{aligned}$$

$$\frac{d[E_2S_2]}{dt} = k_2[E_2S][S] - k_{-2}[E_2S_2]. \tag{3.61}$$

Since (3.44) is a fast reaction, according to the *quasi-equilibrium* assumption, we have

$$\frac{d[E_2S]}{dt} = 0, \tag{3.62}$$

$$\frac{d[E_2S_2]}{dt} = 0. \tag{3.63}$$

These equations imply that

$$[E_2S] = K_1[E_2][S], \tag{3.64}$$

$$[E_2S_2] = K_2[E_2S][S] = K_1K_2[E_2][S]^2, \tag{3.65}$$

where $K_1 = \frac{k_1}{k_{-1}}$ and $K_2 = \frac{k_2}{k_{-2}}$. According to (3.45) and (3.46), the fractional saturation is then given by

$$f = \frac{[E_2S] + 2[E_2S_2]}{2([E_2] + [E_2S] + [E_2S_2])} = \frac{K_1[S] + 2K_1K_2[S]^2}{2(1 + K_1[S] + K_1K_2[S]^2)}, \tag{3.66}$$

which is more realistic. (3.66) is called the Adair model [1]. Figure 3.8 depicts the curves of fractional saturation for protein with two binding sites which are computed by (3.50) and (3.66), respectively.

3.5 Growth model of cell size

Cell size growth is defined as the increase in volume which, in elongating axes, is nearly proportional to the length. Growth in such axes is traditionally discussed as being generated by two processes: cell division and cell elongation. These processes are considered to be responsible for the formation of the cell pattern which, in the one-dimensional case, reduces to a

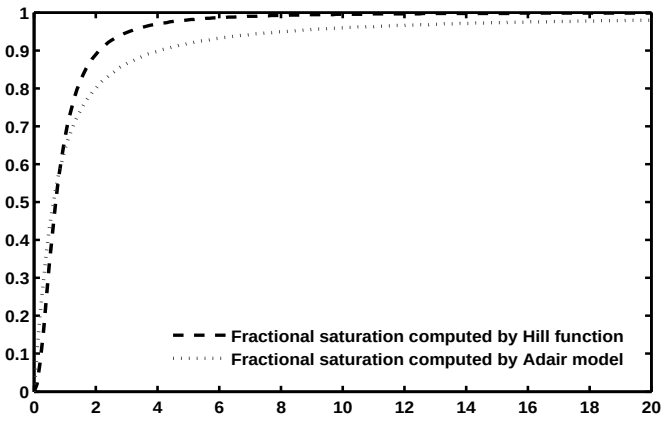


Figure 3.8: Fractional saturation. The fractional saturation or probability of binding given by two models. The first is computed by the Hill function and the second according to the Adair model. The constants of binding and dissociation in (3.43) and (3.44) are $k_1 = 0.2, k_{-1} = 0.1, k_2 = 10, k_{-2} = 8$. Although the two curves are similar, the Adair model is more realistic.

characteristic sequence of mean cell lengths down an axis. When the developmental processes and the cell pattern are consistent through time, the growth of the axis is in a steady state. In this study we used two different models for the growth of the cell area.

3.5.1 Malthusian growth model

The Malthusian[5] growth model, sometimes called the simple exponential growth model. Exponential growth occurs when the growth rate of a mathematical function is proportional to the function's initial value. This model is widely regarded in the field of population ecology as the first principle of population dynamics. We use this model for the growth of the cell area in

[5]The Reverend Thomas Robert Malthus (1766 - 1834), was a British scholar.

Arabidopsis. The form of the exponential growth model is the following:

$$\frac{ds}{dt} = gs, \tag{3.67}$$

where $g \geq 0$. In this model $s(t)$ represents the area of the cell at time t and the constant of proportionality g represents the rate constant. If at time $t = 0$ the cell has area s_0, then the solution at any time t of (3.67) is

$$s(t) = e^{gt} s_0 \tag{3.68}$$

where $s_0 = s(0)$. A graph of the solution (3.68) for $g > 0$ and $g = 0$ is shown in Figure 3.9. We observe that the area grows exponentially and is unbounded for $g > 0$; while for $g = 0$ the area maintains the constant value s_0 [71].

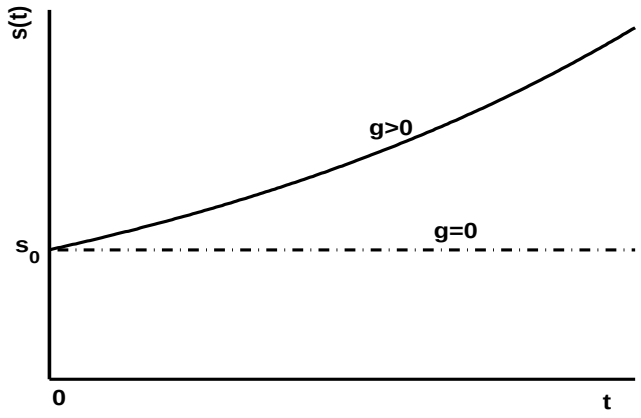


Figure 3.9: Exponential growth.

For $g > 0$ the Malthusian growth model implies that there is no limit to the area of cells and a cell can grow up to infinity. This is unrealistic for some cells such as guard cells in the leaves of *Arabidopsis* plant, cf. §4.2.3. We then use the logistic growth model described below.

3.5.2 Logistic growth model

At first, the logistic growth model was developed by Pierre-François Verhulst[6] for the growth of a population. The initial stage of growth is approximately exponential; then, as saturation begins, the growth slows, and at maturity, growth stops.
The model is defined by the following simple differential equation

$$\frac{ds}{dt} = gs - \epsilon s^2, \tag{3.69}$$

where g and ϵ are positive constants and are called the vital coefficients. ϵ is small relative to g. When s is small, the term ϵs^2 is very small compared to gs so s will grow at nearly an exponential rate. When s becomes large, the rate of growth will approach zero. The solution of the simple non-linear differential equation (3.69) with initial value $s(0) = s_0$ is given by

$$s(t) = \frac{gs_0}{\epsilon s_0 + (g - \epsilon s_0)e^{-gt}}. \tag{3.70}$$

When $t \to \infty$ then $s(t) \to \frac{g}{\epsilon}$. So regardless of the initial value, $s(t)$ ultimately approaches the limiting value $\frac{g}{\epsilon}$, which is called the carrying capacity [71].
The graph of (3.70) has an elongated S-shape and is called the logistic curve (Figure 3.10).

We propose the following logistic differential equation for the area growth of guard cells:

$$\frac{ds}{dt} = gs(1 - \frac{s}{M}), \tag{3.71}$$

where $s(t)$ is the area of guard cells at time t and M is the maximum area of guard cells. The reason for using formula (3.71) is that guard cells cannot

[6]Pierre François Verhulst (1804 - 1849) was a Belgian mathematician who got a PhD from Ghent University in 1825 for a thesis on number theory. He collaborated with the famous statistician Adolphe Quetelet.

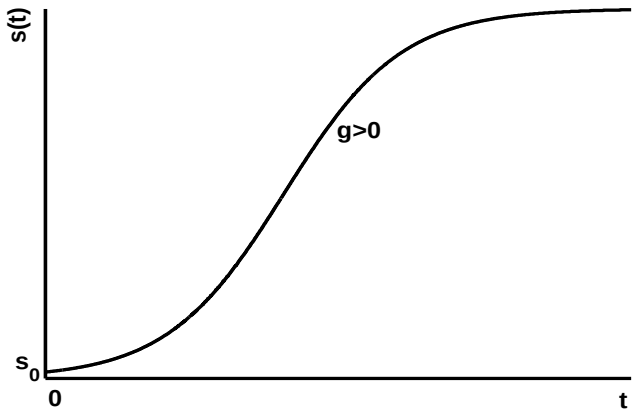


Figure 3.10: Logistic curve.

grow unboundedly and there is a limit M to which they grow.
The solution of the above differential equation is

$$s(t) = \frac{M}{1 + (\frac{M}{s_0} - 1)e^{-gt}}, \tag{3.72}$$

$$g = \frac{\ln[s(t)(\frac{M}{s_0} - 1)] - \ln(M - s(t))}{t}. \tag{3.73}$$

We will use these equations in Chapter 4 for modelling the cell size growth.

3.6 Unconstrained optimization algorithms

In mathematics, optimization, or mathematical programming, refers to finding the best element among the available alternatives. In particular, we use optimization techniques to minimize or maximize a real function which is called an objective function.

An objective function $f(X)$ has a local (or relative) maximum at the point X^* if $f(X^*) \geq f(X)$ for all X nearby X^*. In the case of a global maximum, $f(X^*) \geq f(X)$ for all X in the domain of f. Similarly, an objective function has a local minimum at the point X^* if $f(X^*) \leq f(X)$ for all X nearby X^*. In the case of a global minimum, $f(X^*) \leq f(X)$ for all X in the domain of f. By the above definition, a global maximum (minimum) is also a local maximum (minimum).

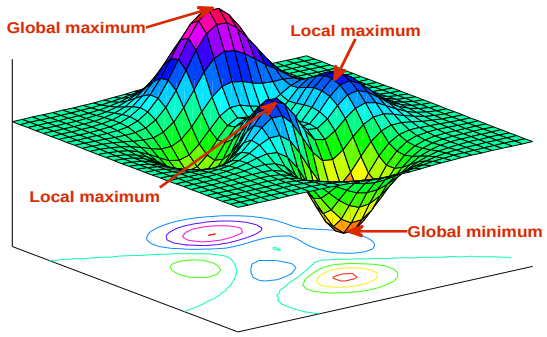


Figure 3.11: Local and global maximum of the objective function.

In this section we consider algorithms for solving the unconstrained non-linear programming problem

$$\min f(X), \tag{3.74}$$

where $X \in \Re^N$ is the vector of decision variables and $f : \Re^N \rightarrow \Re$ is the objective function. We assume that the function f has at least a local minimum in its range of definition and is twice continuously differentiable, even if in many techniques only one time continuous differentiability is needed. According to a theorem of Fermat[7], the optimal points of unconstrained problems are found at stationary points, where the first derivative or the gradient of the objective function is zero (first-order condition).

[7]Pierre de Fermat (1601 - 1665)

More generally, they may be found at critical points, where the first derivative or gradient of the objective function is zero or is undefined, or on the boundary of the choice set.

While the first derivative identifies the stationary points, the second derivatives of the objective function are used to distinguish whether the stationary point is a maximum, minimum or a saddle point (second-order conditions). The matrix of second derivatives is called the Hessian matrix. The algorithms discussed in this section generate a sequence X_n, starting from X_0, by the following iteration

$$X_{n+1} = X_n + \alpha_n d_n, \tag{3.75}$$

where $d_n \in \Re^N$ is a search direction and $\alpha_n \in \Re$ is a stepsize along d_n. Methods differ in the way the direction and the step-size are chosen. We note that different choices of d_n and α_n yield different convergence properties.

Roughly speaking, the search direction d_n affects the local behaviour of an algorithm and its rate of convergence, whereas the step-size α_n has an effect on the global convergence. Most methods determine α_n by a line search technique[8] (for more details see [77, 65]). In this thesis we will briefly describe the Gradient descent method, Newton's method, the quasi-Newton method, non-linear least squares methods and derivative free methods.

3.6.1 Trust region method

The trust region method is an algorithm which essentially approximates only a certain region (the so-called trust region) around the current search point, where a quadratic model of the objective function for local minimization is trusted to be correct and steps are chosen to stay within this

[8] In optimization, a line search technique is a method to find the minimum value of a function in one dimension.

region. The size of the region is modified during the search, based on how well the model agrees with actual function evaluations. The evaluation method is to observe the ratio of expected improvement from the quadratic approximation with the actual improvement observed in the objective function. Simple thresholding of the ratio is used as the criteria for expansion and contraction. Trust region methods are in some sense dual to line search methods: trust region methods first choose a step size (the size of the trust region) and then a step direction while line search methods first choose a step direction and then a step size (see [65]).

3.6.2 Gradient descent method

Gradient descent or steepest descent is a first-order optimization algorithm. To find a local minimum of a function using Gradient descent, we use the first-order of Taylor series for $f(X)$ around the point X_n:

$$f(X_n + \delta X) \approx f(X_n) + g_n^T \delta X, \tag{3.76}$$

where

$$g_n = \nabla f(X_n) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N} \right]^T. \tag{3.77}$$

$f(X_n + \delta X) - f(X_n) \leq 0$ implies that $g_n^T \delta X \leq 0$, thus the gradient of the function $f(X)$ gives the direction of steepest increase. Thus a natural minimization algorithm is to go in the direction opposite the gradient a certain amount. For any $\alpha \in \mathbb{R}$, the Gradient descent algorithm (for minimizing a function f) with learning rate α, chooses the following iteration:

$$X_{n+1} = X_n - \alpha \nabla f(X_n), n \geq 0. \tag{3.78}$$

The sequence (3.78) can converge to the desired local minimum. Note that the value of the step size can change at every iteration [65, 77].
Although Gradient descent works in spaces of any number of dimensions,

even in infinite-dimensional ones, it can take many iterations to converge towards a local minimum, if the curvature in different directions is very different. Moreover, finding the optimal α per step can be time-consuming. Using a fixed α can yield poor results. Newton's method is often a better alternative.

3.6.3 Newton's method

Newton's method is a well-known algorithm for finding roots of equations in one or more dimensions. It can also be used to find the stationary points of a function, as such points are the roots of the derivative function. To find the stationary points of (3.74), we expand $f(X)$ by its Taylor series around the point X_n:

$$f(X_n + \delta X) \approx f(X_n) + g_n^T \delta X + \frac{1}{2} \delta X^T H_n \delta X \tag{3.79}$$

and the Hessian is the symmetric matrix

$$H_n = H(X_n) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \vdots & \ddots & \\ \frac{\partial^2 f}{\partial x_1 \partial x_N} & & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}. \tag{3.80}$$

For a minimum we require that $\nabla f(X) = 0$. Again using Taylor series for $\nabla f(x)$, we obtain

$$g_n + H_n \delta X = 0 \Longrightarrow \delta X = -H_n^{-1} g_n, \tag{3.81}$$

from which we get the following iteration

$$X_{n+1} = X_n - H_n^{-1} g_n \tag{3.82}$$

yielding

$$f(X_n + \delta X) \approx f(X_n) - \frac{1}{2} g_n^T H_n^{-1} g_n. \tag{3.83}$$

From this expression it is clear that as long as $g_n^T H_n^{-1} g_n$ is positive, $f(X_n)$ is in a downhill direction. A fundamental theorem due to Jacobi shows that $g_n^T H_n^{-1} g_n$ is described by the quadratic form $\sum_{i=1}^{\mathbb{N}} \lambda_i x_i^2$, where λ_i are the eigenvalues of H_n^{-1} and x_i is the component of g_n along the eigenvector that corresponds to λ_i. Hence, if the eigenvalues of the Hessian at X_n are all positive, then X_n is a local minimum.

Although Newton's method converges fast towards a local maximum or minimum, finding the inverse of the Hessian in high dimensions can be an expensive operation [65, 67]. In such cases, there exist various quasi-Newton methods, where an approximation for the Hessian (or its inverse) is built up from changes in the gradient.

3.6.4 Quasi-Newton methods

If the problem size is large and the Hessian matrix is dense then it may be infeasible or inconvenient to compute it directly. Quasi-Newton methods avoid this problem by keeping a rolling estimate of $H(x)$, updated at each iteration using new gradient information.

The most common quasi-Newton algorithms are the *DFP*[9], and the *BFGS*[10]. The *BFGS* is generally regarded as the best performing method [67, 65]:

$$H_{n+1} = H_n + \frac{q_n q_n^T}{q_n^T s_n} - \frac{H_n^T s_n^T s_n H_n}{s_n^T H_n s_n}, \tag{3.84}$$

where

$$s_n = x_{n+1} - x_n, \tag{3.85}$$

$$q_n = g_{n+1} - g_n, \tag{3.86}$$

[9] for Davidon, 1959, and Fletcher and Powell, 1963

[10] for Broyden, 1969, Fletcher, 1970, Goldfarb, 1970, and Shanno, 1970

$$H_0 \;=\; I. \tag{3.87}$$

The estimate H_n is used to form a local quadratic approximation as before. In the Quasi-Newton method H itself is not computed anywhere in the iterations but rather represented compactly by a few stored vectors (for a proof see [65]). For $f : \mathbb{R} \longrightarrow \mathbb{R}$, BFGS with unit step size gives the secant method:

$$x_{n+1} = x_n - \frac{f'(x_n)}{H_n}, \quad H_n = \frac{f'(x_n) - f'(x_{n-1})}{x_n - x_{n-1}}. \tag{3.88}$$

3.6.5 Non-linear least squares

Non-linear least squares[11] is the form of least squares analysis which is used to fit a set of m observations with a model that is non-linear in n unknown parameters $(m > n)$. The basis of the method is to approximate the model by a linear one and to optimize the parameters by successive iterations. Let us consider a set of M experimental data points, and a model function $Z = Q(Y, X)$ where $Z = (z_1, z_2, \ldots, z_M)$ depends on the variable $Y = (y_1, y_2, \ldots, y_M)$ and n parameters, $X = (x_1, x_2, \ldots, x_N)$ with $M \geq N$. The aim is to find the vector of parameters such that the curve fits best the given data in the least squares sense, that is, the sum of squares

$$f(X) = \sum_{i=1}^{M} {r_i}^2, \tag{3.89}$$

is minimized, where the residuals (errors) r_i are given by

$$r_i = z_i - Q(y_i, X). \tag{3.90}$$

In general, the above non-linear problem is very hard to solve exactly. The Gauss-Newton algorithm is a method to solve it. It can be seen as

[11] The least squares solution is the overall solution that minimizes the sum of the squares of the errors made in solving every single equation.

a modification of Newton's method for finding a minimum of a function. Unlike Newton's method, the Gauss-Newton algorithm can only be used to minimize a sum of squared function values, but it has the advantage that second derivatives, which can be challenging to compute, are not required.

Gauss-Newton method

The $M \times N$ Jacobian of the vector of residuals r is defined as

$$J(X) = \begin{pmatrix} \frac{\partial r_1}{\partial x_1} & \cdots & \frac{\partial r_1}{\partial x_N} \\ \vdots & \ddots & \\ \frac{\partial r_M}{\partial x_1} & \cdots & \frac{\partial r_M}{\partial x_N} \end{pmatrix}. \tag{3.91}$$

The minimum value of (3.89) occurs when the gradient is zero. Since the model contains N parameters, there are N gradient equations. Consider

$$\frac{\partial}{\partial x_k} \sum_i {r_i}^2 = \sum_i 2 r_i \frac{\partial r_i}{\partial x_k}. \tag{3.92}$$

Hence

$$\nabla f(X) = 2 J^T r. \tag{3.93}$$

For the Hessian we require

$$\begin{aligned} \frac{\partial^2}{\partial x_l \partial x_k} \sum_i {r_i}^2 &= 2 \frac{\partial}{\partial x_l} \sum_i r_i \frac{\partial r_i}{\partial x_k} \\ &= 2 \sum_i \frac{\partial r_i}{\partial x_k} \frac{\partial r_i}{\partial x_l} + 2 \sum_i r_i \frac{\partial^2 r_i}{\partial x_k \partial x_l}. \end{aligned} \tag{3.94}$$

Hence

$$H(X) = 2 J^T J + 2 \sum_i^M r_i R_i. \tag{3.95}$$

Note that the second-order term in the Hessian $H(X)$ is multiplied by the residuals r_i and in most problems, the residuals will typically be small.

Therefore, the second-order term is often ignored, giving the Gauss-Newton approximation to the Hessian:

$$H(X) = 2J^T J. \tag{3.96}$$

Hence, explicit computation of the full Hessian can again be avoided. So the iteration to find the minimum of $f(x)$ in (3.89) is

$$X_{n+1} = X_n - \alpha_n {H_n}^{-1} g_n, \quad H_n(X) = 2J_n^T J_n. \tag{3.97}$$

Levenberg-Marquardt algorithm

The Levenberg-Marquardt algorithm (LMA) is a numerical technique to solve the minimizing problem of a function, generally non-linear, over a space of parameters of the function. These minimization problems arise especially in least squares curve fitting.
The LMA interpolates between the Gauss-Newton algorithm (GNA) and the method of Gradient descent. The LMA is more robust than the GNA, which means that in many cases it finds a solution even if it starts very far from the final minimum. On the other hand, for well-behaved functions and reasonable starting parameters, the LMA tends to be a bit slower than the GNA [50].

Like all non-linear optimization methods, LMA is iterative. Initiated at the starting point X_0, it produces a series of vectors $X_1, X_2, \ldots$ that converge towards a local minimum for (3.89). Hence, at each iteration, it is required to find the step δ that minimizes (3.89). It can be seen that simple gradient descent and Gauss-Newton iteration are complementary in the advantages they provide. Levenberg proposed an algorithm based on this observation, whose update rule is a blend of the above mentioned algorithms and is given as

$$X_{n+1} = X_n - (H + \lambda I)^{-1} \nabla f(X_n), \tag{3.98}$$

where H is the Hessian matrix evaluated at X_n, λ is the damping parameter and f is defined in (3.89). This update rule is used as follows. If the error goes down following an update, it implies that our quadratic assumption on $f(X)$ is working and we reduce λ (usually by a factor of 10) to reduce the influence of gradient descent. On the other hand, if the error goes up, we would like to follow the gradient more and so λ is increased by the same factor.

The above algorithm has the disadvantage that if the value of λ is large, the calculated Hessian matrix is not used at all and then the algorithm behaves like

$$X_{n+1} = X_n - \frac{1}{\lambda}\nabla f(X_n), \tag{3.99}$$

which we recognize to be the Method of Gradient Descent, with $\alpha = \frac{1}{\lambda}$ in (3.78). Conversely, if $\lambda \to 0$, then it should be clear that LMA performs similarly to Newton's method, cf. §3.6.3. Marquardt then derived some advantage out of the second derivative. He replaced the identity matrix in (3.98) with the diagonal of the Hessian resulting in the Levenberg-Marquardt update rule:

$$X_{n+1} = X_n - \left(H + \lambda diag[H]\right)^{-1} \nabla f(X_n). \tag{3.100}$$

Since the Hessian is proportional to the curvature of f, (3.100) implies a large step in the direction with low curvature (i.e., an almost flat terrain) and a small step in the direction with high curvature (i.e, a steep incline). It is to be noted that while LMA is in no way optimal but is just a heuristic, it works extremely well in practice. The only flaw is its need for matrix inversion as part of the update (for more details see [50]).

3.7 Derivative-free optimization

The standard mathematical characterization of a local minimum of differentiable functions, requires that the first-order derivatives are zero. However, for some reasons there are many examples where (at least some)

derivatives are unavailable or unreliable. Also, when the function evaluations are expensive tasks, applying finite-difference derivative approximations is also infeasible. Thus a class of non-linear optimization techniques called derivative-free optimization methods is needed. Note that the near-optimal solution obtained by a derivative-free method is often less accurate than that obtained by a derivative-based method, assuming derivative information is available. But it is still efficient when the function evaluations are noisy or expensive. One of the limitations of the derivative-free optimization is that it converges rather slowly compared to those methods that use approximations to the Hessian matrix (such as quasi-Newton methods). We will later refer to the Tables 4.4 and 4.5 and Figure 4.18 as an example.

In this thesis we consider derivative-free methods that sample the objective function at a finite number of points at each iteration and decide which actions to take next solely based on those function values and without any explicit or implicit derivative approximation or model building. A classical example of a derivative-free direct-search algorithm is the Nelder-Mead method which is described in the following section.

3.7.1 The downhill simplex method

The downhill simplex method or Nelder-Mead method is a non-linear optimization technique, due to J. A. Nelder and R. Mead (1965) [60, 55, 17]. It minimizes an objective function in a multidimensional space. It approximately finds a locally optimal solution to a problem with N variables when the objective function varies smoothly.

It uses the concept of a simplex, which is characterized by the $N + 1$ distinct vectors (its vertices) in the N-dimensional space. In the two-space, a simplex is a triangle; in the three-space, it is a pyramid. At each step of the search, a new point in or near the current simplex is generated by three operations: reflection, expansion, and contraction. Figure 3.12 illustrates

these operations in a three-dimensional variable space. **Reflection** involves moving the worst point (vertex) of the simplex (where the value of the objective function is the highest) to a point reflected through the straight line connecting the gravity center of all points and the worst point. If this point is better than the best point, then the method attempts to expand the simplex along this line. This operation is called **expansion**. On the other hand, if the new point is not much better than the previous point, then the simplex is **contracted** along the same straight line mentioned above. This procedure is called contraction. Moreover, if the new point is worse than the previous points, the simplex is contracted along all dimensions toward the best point and steps down the valley which is called **shrink**. By repeating this series of operations, the method finds the optimal solution (see Figure 3.12).

In all cases, the function value at the new point is compared with the function's values at the vertices of the simplex and, usually, one of the vertices is replaced by the new point, giving a new simplex. This step is repeated until the diameter of the simplex is less than the specified tolerance. The downhill simplex method must be started not just with a single point, but with $N + 1$ points, defining an initial simplex. If only one initial point x_1 is given, then the other N points are chosen by $x_i = x_1 + a_i e_{i-1}, i = 2, \ldots, N + 1$ where e_i are N unit vectors, and a_i are constants that characterize the length scale for each vector direction.

The downhill simplex algorithm

The algorithm generates a new test position by extrapolating the behaviour of the objective function measured at each test point of the simplex. The algorithm then chooses to replace one of these test points with the new test point and so the technique progresses. x_r, x_e and x_c are the new points generated by three operations, namely reflection, expansion, and contraction, respectively. If the new point generated by neither of those operations is efficient, the simplex shrinks along all dimensions toward the best point. By

repeating this series of operations, the method finds the optimal solution.

1. First order according to the values at the vertices:

$$f(x_1) \leq f(x_2) \leq \cdots \leq f(x_{N+1}).$$

2. Compute a **reflection**: $x_r = x_o + \alpha(x_o - x_{N+1})$.
   x_o is the center of gravity of all points except x_{N+1} and $\alpha > 0$. If $f(x_1) \leq f(x_r) < f(x_N)$, then we compute a new simplex with x_r by rejecting x_{N+1}. Go to step 1.

3. **Expansion**: If $f(x_r) < f(x_1)$, then compute $x_e = x_o + \gamma(x_o - x_{N+1})$ where $\gamma > 0$.
   If $f(x_e) < f(x_r)$ compute a new simplex with x_e by rejecting x_{N+1} and go to step 1. Else, compute a new simplex with x_r by rejecting x_{N+1} and go to step 1.

4. **Contraction**: If $f(x_r) \geq f(x_N)$, let $x_c = x_{N+1} + \rho(x_o - x_{N+1})$ where $0 < \rho < 1$.
   If $f(x_c) \leq f(x_{N+1})$, compute a new simplex with x_c by rejecting x_{N+1}. Go to step 1. Else go to step 5.

5. **Shrink** step: Compute the n vertices evaluations:
   $x_i = x_1 + \sigma(x_i - x_1)$ for all $i \in \{2, \ldots, N+1\}$ where $0 < \sigma < 1$.
   Go to step 1.

α, ρ, γ and σ are respectively the reflection, the expansion, the contraction and the shrink coefficient. Standard value are $\alpha = 1, \gamma = 2, \rho = \frac{1}{2}$ and $\sigma = \frac{1}{2}$. The initial simplex is also important, since a too small initial simplex can lead to a local search, consequently the method can get more easily stuck. So this simplex should depend on the nature of the problem.

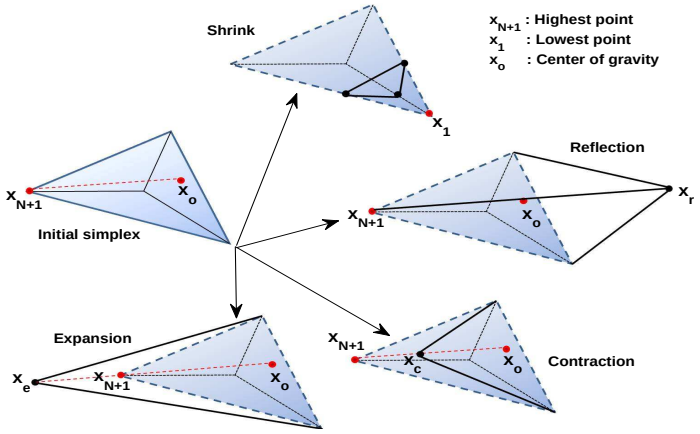


Figure 3.12: Four basic operations in the downhill simplex method. x_r, x_e and x_c are the new points generated by three operations, namely reflection, expansion, and contraction, respectively.

3.8 Optimization algorithms in Matlab

Matlab has two toolboxes that contain optimization algorithms, Optimization Toolbox and Global Optimization Toolbox. Optimization Toolbox provides widely used algorithms for standard and large-scale optimization. These algorithms solve constrained and unconstrained continuous and discrete problems. The toolbox includes functions for linear programming, quadratic programming, binary integer programming, non-linear optimization, non-linear least squares, systems of non-linear equations, and multi-objective optimization. Global Optimization Toolbox provides methods that search for global solutions to problems that contain multiple maxima or minima. It includes global search, multi-start, pattern search, genetic algorithm, and simulated annealing solvers.

Here we consider two functions, *fminunc* and *fminsearch* for solving unconstrained non-linear optimization programs and the function *lsqnonlin* for least squares. All are available in the Optimization Toolbox. Function

fminunc is based on a gradient-based algorithm with two modes, large-scale and medium-scale. Large-scale mode is a subspace trust-region[12] method, cf. §3.6.1. It can take a user-supplied Hessian or approximate it using finite differences (Newton's method). Medium-scale mode is a cubic line-search method. It uses quasi-Newton updates of the Hessian. In Chapter 4, we will use both the Newton and the quasi-Newton algorithms for our unconstrained non-linear program. Function *fminsearch* uses a derivative-free method based on Nelder-Mead simplex which we described above in detail. This is the method used for most optimizations discussed in this thesis. The last function is *lsqnonlin* which uses one of the following three methods:

• **trust-region-reflective method**,

By default, lsqnonlin chooses the trust-region-reflective algorithm. This algorithm is a subspace trust-region method and is based on the interior-reflective Newton method described in [15, 16]. Each iteration involves the approximate solution of a large linear system using the method of preconditioned conjugate gradients (PCG).

[12]The term, "trust-region" was first coined by Celis, Dennis and Tapia at Rice University.

Chapter 4

Model-based analysis of cell divisions in *Arabidopsis* leaf

This chapter introduces an investigative study that I conducted as part of my doctoral research. The overall aim of this study was to investigate the precise nature of the interaction between cell division and cell expansion during leaf development. The growth of plant organs results from these two processes that act at the cellular level, but the interaction between these process is still largely unknown because of the experimental challenge to disentangle them. The abaxial epidermal layer of the first *Arabidopsis thaliana* leaf pair consists of two major cell types, puzzle-shaped pavement cells and guard cells. Following a kinematic analysis of leaf growth, we determined cell area distributions for both cell types during complete leaf development. To dissect the rules by which different cell types divide and expand, the experimental data were fit into a mathematical model, describing all possible changes a cell can undergo from one day to the next. This work has been part of a team-work, and the word ”we” refer to the team members of this study including S. Dhondt, V. Boudolf, G.T.S. Beemster, T. Beeckman, D. Inzé, W. Govaerts and L. De Veylder. The author as a member of the team was responsible for the parts of modelling, simulations, optimization, data analysis and all mathematical issues related to this study.

4.1 Introduction

In most plant species the above-ground plant body is dominated by the leaves. These organs are specialized for photosynthesis, a process that converts carbon dioxide into organic components, using the energy from sunlight. This makes leaves the site for energy production in the plant and the engine for growth. To maximize its light capturing capacity a leaf is typically flat and thin. This characteristic shape is established during the process of leaf development. Leaves first arise on the shoot apical meristem at sites of auxin accumulation [7]. At this position, a number of cells start to bulge out from the meristem and eventually, will form the basis of the leaf primordium when cell division proceeds [64, 70]. Dorsiventrality (Flattened and having distinct upper and lower surfaces) is specified early in primordium development[1] and defines the adaxial and abaxial side of the leaf [12]. Divisions at the margin of the primordium drive leaf blade inception. Further expansion of the leaf blade is controlled by a strong preference for anticlinal divisions, leading to the primarily lateral outgrowth of different tissue layers of which the epidermis is the main layer driving leaf growth [24, 73].

During further leaf development in the model species *Arabidopsis thaliana*[2] (Figure 3) and other dicotyledonous species a cell proliferation phase, in which cells are actively dividing, is followed by a cell expansion phase where cells differentiate and expand. After expansion, cells mature and final leaf size is reached [5]. At the transition from proliferation to expansion phase cell division ceases along a longitudinal gradient from leaf tip to base [24]. In the epidermis, the onset of differentiation coincides with the formation of stomata [21]. A stoma comprises two guard cells

[1] A primordium (pl. primordia) is defined as an organ or tissue in its earliest recognizable stage of development. Primordial development in plants is critical to the proper positioning and development of plant organs. Different primordial types like the leaf and flower primordia arise from the shoot lateral meristem.

[2] *Arabidopsis* is a genus of small flowering plants and *thaliana* is one of the species.

that control the aperture of the stomatal pore between them. Stomata are epidermal valves that optimize water evaporation with the gas exchange needed for photosynthesis (Figure 4.1). Starting from a precursor meristemoid cell a series of subsequent asymmetric divisions produce a number of guard mother cells and daughter cells. The latter undergo cell fate specifications identical to the majority of cells produced during proliferation phase resulting in puzzle shaped pavement cells which start expanding and become the largest and most numerical epidermal cells. The two guard cells forming the stoma originate from a symmetric division of the guard mother cell, ending the stomatal lineage. Estimations suggest that most of the epidermal cells are generated by the stomatal lineage, making guard cells after the pavement cells the second most prominent cell type in the epidermis of mature leaves [31, 48].

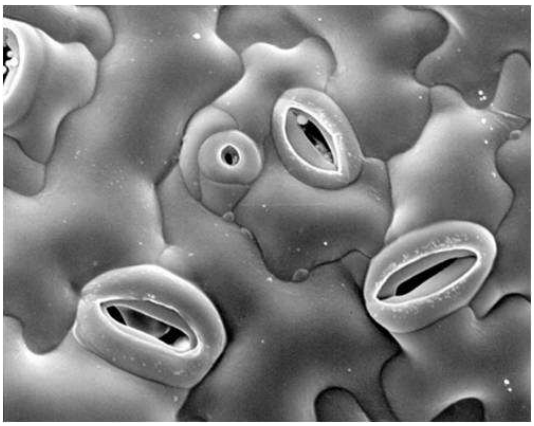


Figure 4.1: Stomata and pavement cells.

The final leaf size is determined by the total number of cells and the average cell size which are the result of cell division and cell expansion during leaf development, respectively. Kinematic analysis of leaf growth provides a powerful and rigorous framework to study the dynamics of these processes [28]. Knowledge on cell cycle duration, cell expansion and the

interaction of both on the level of individual cells is poorly understood, due to the fact that this is not easily experimentally determined and because inhibition of one often leads to the phenomenon of compensation whereby the other increases [81]. The data used in this thesis were obtained by Stijn Dhondt, Lieven De Veylder and Véronique Boudolf[3]. Stijn Dhondt measured individual cell sizes of pavement cells and guard cells separately throughout leaf development. Incorporation of these data in a mathematical model allowed the estimation of division and expansion parameters of pavement and guard cell populations within the growing leaf, allowing us to gain a better and more detailed insight in the processes that define leaf growth. Imaging of epidermal cells resulted into a good correlation between predicted and experimental cell growth data, supporting the model.

In fact, for both cell types, pavement and guard cells, we determine separately cell numbers and individual cell areas during development. To dissect the rules by which different cell types divide and expand, the experimental data were fit into the mathematical model that describes all possible changes a cell can undergo from one day to the next. The model allowed us to calculate the probabilities for a cell to become a guard or pavement cell, the maximum size at which it can divide, and its cell division and cell expansion rate at each point during the leaf development. Surprisingly, the cell cycle duration was found to be constant on the average throughout leaf development, whereas no evidence for a maximum cell size threshold for cell division of pavement cells was found. Furthermore, the model demonstrates that within the epidermis, neighbouring cells of different sizes expand at distinctly different rates, which might explain the complex shape of pavement cells.

[3]PSB/VIB - Ghent University

4.2 Results

4.2.1 Kinematic analysis of leaf growth

Recent developments in microscope and imaging technology suggest that cell tracking is the most suitable way to disentangle cell division and expansion in plants. This method allows to follow cells during 3–4 days and has been successfully applied to the root, the SAM and sepals [13-16]. However, to cover entire leaf development a much longer time frame is needed. Furthermore, such imaging techniques only cover very local changes, and thus do not report global growth characteristics. Therefore, a general kinematic analysis of leaf growth in *Arabidopsis thaliana* was used as a starting point for this research. In this approach, the first developing leaf pair (leaves 1 and 2) are harvested on a daily basis from 5 to 25 days after sowing (DAS). We used leaves 1 and 2 because they are nearly indistinguishable and probably are best synchronized among replicate plants. Microscopic drawings of abaxial epidermal cells are made at positions 25% and 75% of the distance from the base to the tip of the leaf (Figure 4.2), giving a precise estimate of the leaf average cell area. The average of the cell areas at these two positions is combined with the measured total leaf area to estimate the total cell number per leaf. When analyzed on a daily basis, this allows for the calculation of cell division and cell expansion rates.

Plotting of the leaf size evolution on a logarithmic scale revealed a linear increase until day 11, indicating exponential growth (Figure 4.4A). From day 12 onward, the leaf expansion rate decreased and the mature leaf size was reached approximately at 20 DAS. A similar evolution could be observed for the total cell number (Figure 4.4B), with cell division rates being high until day 10 (Figure 4.4C). Cell sizes remained relatively constant until day 10 (approximately $100\mu m^2$), indicating that cell division and expansion rates were balanced in the young leaf (Figure 4.4D). From day 10 onward, the average cell size increased to approximately 10-fold

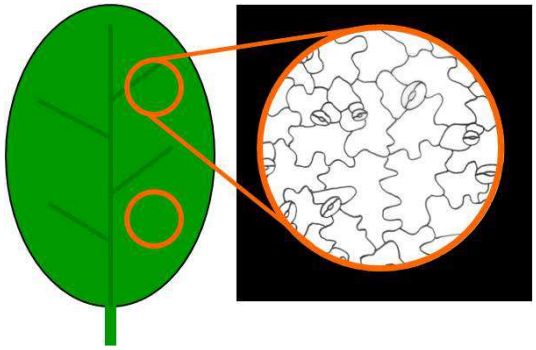


Figure 4.2: Measurements of the number of abaxial epidermal cells at two distinct positions, namely midway from the midvein and the leaf margin and at 25% and 75% of the distance from the base to the tip of the leaf, respectively.

by day 20, as the result from cell expansion in the absence of cell division (Figure 4.4D). Coinciding with the decrease in cell division rate, the stomatal index (fraction of guard cells among all cells) increased linearly, marking the leaf differentiation (Figure 4.4E). The relative leaf expansion rates were the highest during the high division rate period (Figure 4.4F). When considering pavement and stomatal cells separately, the total numbers of pavement cells increased gradually from day 5 to day 14 while the number of guard cells continued to increase until day 17, (Figure 4.7), indicating that the divisions giving rise to guard cells continued approximately 3 days longer than those to pavement cells.

4.2.2 Cell size distributions

Although the kinematic data give a characterization of the general growth processes during leaf development, the data consider leaves as homogenous cell populations, which is a simplification, because the epidermis consists of multiple cell types, each with distinct size characteristics at different time points during development. Furthermore, pavement cells

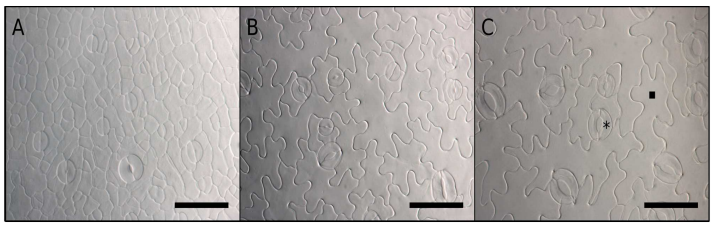


Figure 4.3: Abaxial *Arabidopsis* leaf epidermis during development. (A) Proliferating phase at 8 days after sowing (DAS). (B) Expansion phase at 14 DAS. (C) Mature abaxial epidermis at 20 DAS. $*$ guard cell; ■ puzzle-shaped pavement cell. Bar = 50 μm.

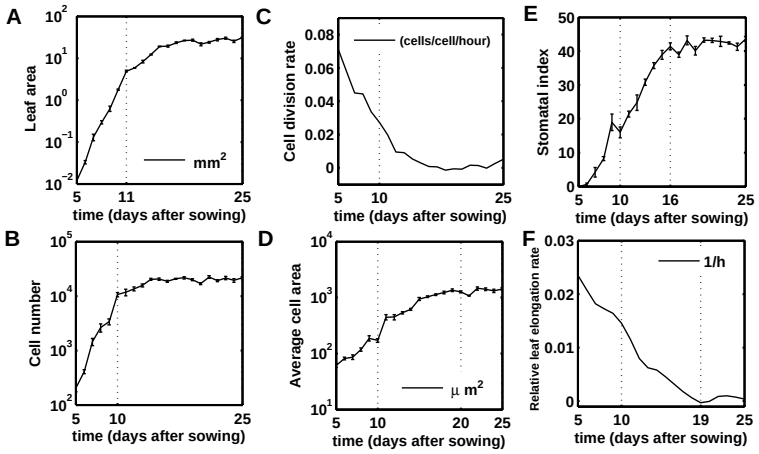


Figure 4.4: Kinematic data for the first true leaf pair of young *Arabidopsis* seedlings throughout their development. (A) Leaf area. (B) Total cell number. (C) cell division rate (CDR).(D) Mean cell area. (E) Stomatal index (SI). (F) Relative leaf elongation rate (RLER).

and guard cells are interdependent, because pavement cells are formed together with stomata[31]. Additionally, the size of pavement cells ranges from 50 up to 20,000μm^2, illustrating the heterogeneity of the population. To obtain a better insight into the cell area distribution during leaf devel-

opment, Stijn Dhondt and others[4] developed an image analysis algorithm that measures individual cell sizes from the microscopic drawings used in the kinematic analysis, cf. §4.5.

According to the experimental data, 97% of the pavement cells of 5-day-old leaves were smaller than $100\mu m^2$, whereas at day 8, only 62% of the cells had a size below this threshold, indicating that pavement cell sizes increase already during the cell proliferation phase. Early in the expansion phase, at 10 DAS, the maximum cell size detected was approximately $1600\mu m^2$ and 95% of the pavement cells were smaller than $500\mu m^2$. From day 11 onward, the pavement cell area distribution broadened and the pavement cell population was distributed over a large range of cell sizes (Figure 4.5A). Guard cell sizes ranged from 25 to $150\mu m^2$, with a mean area of approximately $75\mu m^2$. During leaf development, cell sizes increased, reaching a maximum size of approximately $300\mu m^2$ and an average area of $150\mu m^2$ at 25 DAS (Figure 4.5B).

To obtain quantitative information about the changes in cell size distributions during leaf development, we extrapolated the frequency distribution of the cell areas for the whole leaf on a given day *i* for pavement and guard cells, separately (for details, see Materials and Methods). This absolute representation of the data showed that from day 5 to 12, most pavement cells were very small (less than $300\mu m^2$) and that the frequency of these small cells increased significantly from day 6 to day 9 (Figure 4.5C), corresponding to the high cell division rate during these days. From day 9 until day 12, the peak of the distribution curves was less pronounced and was accompanied by a higher proportion of larger cells. The number of large cells also increased until day 17, after which the distribution of pavement cells remained relatively stable. The guard cell distribution was different. The graph had a symmetrical bell shape with a peak at the mean, revealing a roughly normal size distribution of guard cell sizes (Figure 4.5D). Prior to day 9, the number of guard cells was low, but afterwards

[4]PSB/VIB - Ghent University

increased significantly until day 17, indicating that most divisions of guard mother cells, leading to the formation of stomata, occurred relatively late during epidermal development. After day 17, the complete guard cell size distribution continued to shift to the right, hinting that cell divisions had ceased and cells just grew.

4.2.3 Mathematical model for leaf development

A general mathematical model was built to quantitatively study the crucial parameters for cell division and cell expansion during leaf development, cf.§4.5. The model takes only the pavement and guard cells into account, because in our experimental system it is impossible to distinguish pavement precursor cells from stomatal precursor cells. The model is based on the changes in size distribution of pavement and guard cells between successive days in function of cell expansion and on the changes in cell identity with each division event.

Therefore, in the model, all possible transitions were considered that a cell can undergo from one day to the next: a precursor cell might either expand and divide into two pavement cells or two guard cells, or expand in the absence of division, whereas guard cells do not divide, but can expand. The model included a maximum guard cell size (M_{GC}). As the final division in the stomatal lineage is symmetric, pavement cells larger than twice this size ($2M_{GC}$), cannot divide into two guard cells. Finally, a hypothetical threshold above which pavement cells can no longer divide (T_{PC}) was introduced into the model. Apart from these, specific parameters used in the model were the cell cycle duration time (T_c), cell growth rates for pavement (g_{PC}) and guard cells (g_{GC}), and the probabilities p_1 and p_2 for a precursor cell to divide into pavement or guard cells, respectively. p_3 (equal to $1 - p_1 - p_2$) represented the probability for a pavement cell not to be in division process. $T_{PC}, T_c, g_{PC}, g_{GC}, p_1$ and p_2 are allowed to vary from day to day while M_{GC}, as a maximum size, is kept the same

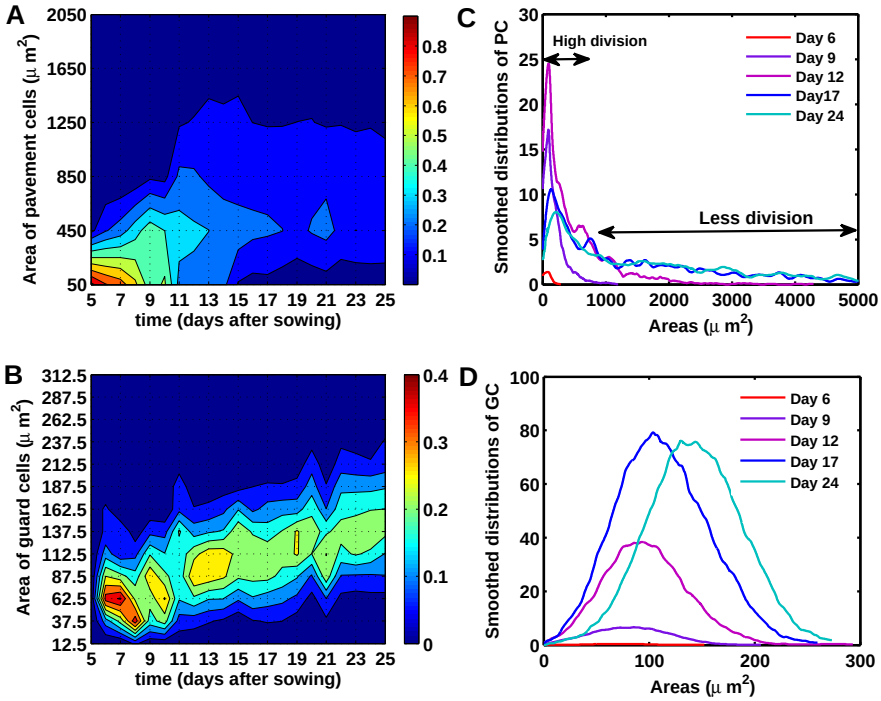


Figure 4.5: Cell size distributions during leaf development. (A and B) Relative size distribution of cells in the abaxial epidermis of leaf 1 and 2, represented in a contour plot. Color ranges indicate the fraction of the total number of cells present within a specific size bin. The vertical axis shows bins of cell areas (bin size is $400\mu m^2$ for pavement cells and $25\mu m^2$ for guard cells). Labels on the vertical axis mark the beginning of each bin. Red indicates a high percentage of cells and blue indicates a low percentage of cells. The colors in the ranges between two consecutive bins are obtained by interpolation. (A) Relative size distribution for pavement cells. (B) Relative size distribution for guard cell. (C and D) Distribution of the number of pavement (C) and guard (D) cells per leaf during its development. Until day 10, the peak of small pavement cells increases, indicating a high cell division rate (High division). After day 17, the peak decreases and the curves become wider, indicating that most cells enter into the expansion phase (Less division).

throughout development. Subsequently, we identified the different developmental categories for pavement cells at any given day (Figure 4.6A).

The first category contained cells with an area below T_{PC} and $2M_{GC}$ that could either divide into pavement or guard cells or stop dividing, contributing to fractions p_1 and p_2. The second category grouped those cells with an area below T_{PC} and above $2M_{GC}$ that could no longer become guard cells, because this would exceed the maximum threshold size. Therefore, a fraction p_1 of the cells in this category divided into pavement cells and the remainder expanded without division. The third category contained those cells with an area above T_{PC} that did not divide.

A precursor cell with an area a^* at day i, can transform into a cell with area a at day $i+1$ through several scenarios, called flows. These flows are defined by the types of cells at day i and $i+1$, and by the number of divisions in between. The number of divisions depends critically on T_c, in particular on the question whether $T_c < 24h$ or $T_c \geq 24h$. The average cell cycle duration could be estimated in a fully proliferating tissue, represented by days 5 to 8 in our analysis, and was calculated to be approximately $25.6 \pm 5.4h$ (see §4.5.3 and Table 4.3). Similarly, previously, the average cell cycle duration for a fully proliferating tissue was calculated to be $20.7h$ [21]. Thus, we could safely assume that (on average) $T_c \geq 18h$, implying that each pavement cell can undergo at most two divisions in one day. The a priori assumption that $Tc \geq 18h$ simplifies our modelling but T_c will be obtained more accurately as a result of our computations. As explained in detail below, eight possible flows (F_1 to F_8) for pavement cells and one flow (F_9) related to the guard cells were defined (Figure 4.6B). Flow F_1 was connected to the pavement cells that did not divide in 1 day, but only expanded. For dividing cells, possibilities differed based on whether the cell cycle duration time was shorter or longer than 24h. For flows F_2, F_3, and F_6, T_c was less than 24 h and pavement cells divided within 1 day either once (F_2) or twice (F_3) into pavement cells or guard cells (F_6). In flow F_3, for simplicity, we assumed that the newborn pavement cells are also in the process of dividing into pavement cells. Because $T_c \geq 18h$, not many cells take part in this flow and therefore the number of missed

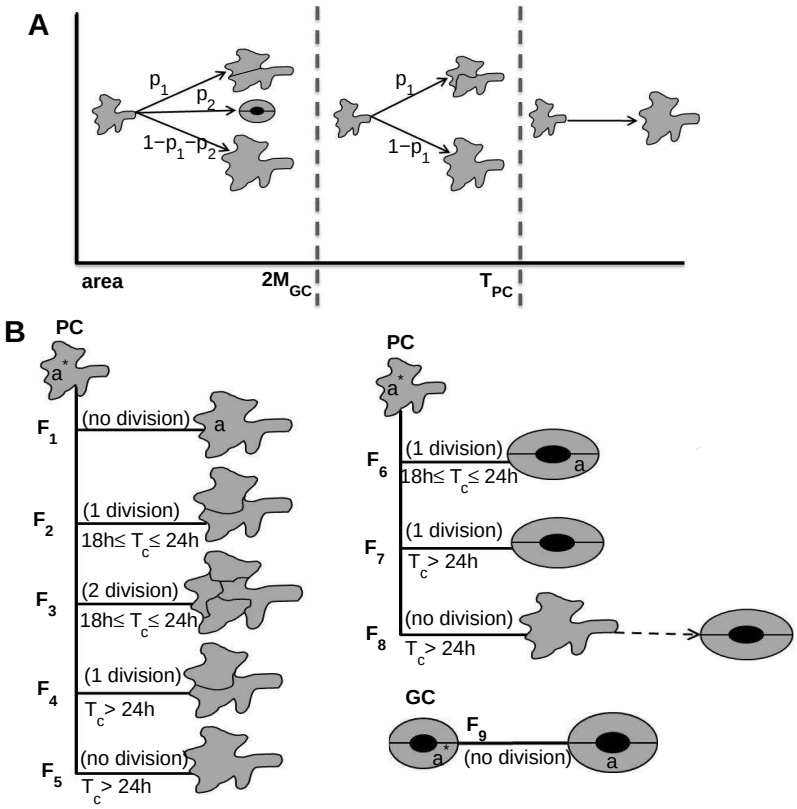


Figure 4.6: Different scenarios of division and expansion of a pavement and a guard cell in 1 day. (A) Categories of pavement cells. In category 1, pavement cells have an area smaller than $2M_{GC}$ and T_{PC}. For these cells, there are three possibilities: a fraction (p_1) divides into two new pavement cells, another fraction (p_2) divides into two new guard cells, and the remainder $(1 - p_1 - p_2)$ grows, but does not divide. In category 2, the area of pavement cells is between $2M_{GC}$ and T_{PC}; a fraction (p_1) divides into pavement cells and the remainder $(1 - p_1)$ remains as pavement cells. In category 3, the cells with area larger than T_{PC}, can only expand. (B) Global scheme of the flows for pavement and guard cells in 1 day. The flows are defined based on the fates of the cells and the duration of the cell cycle. PC and GC are pavement and guard cells, respectively.

events must be small. When $T_c > 24h$, pavement cells could divide at most once in 1 day. In flow F_4, pavement cells divided once, while in flow F_5,

they were in the process of dividing, but did not complete their division process within the first day. Flow F_7 was related to the pavement cells that divided into guard cells, and flow F_8 to those that were in the process of dividing into guard cells, but did not complete division in the first day. Finally, flow F_9 represented guard cells that only expanded. With these flows, we constructed a function for the transitions between the predicted size distributions of pavement and guard cells from one day to the next. This function was then used to optimize the values for the parameters to best fit the experimental data.

We emphasise that in our model we allow the nature of a cell to change from day to day. A cell which has the potential to divide on a given day (being in the fraction p_1) could well no longer have that potential on the next day, e.g. because it has entered the endocycle. If it is still in the process of dividing, the cell cycle duration could have changed because of the molecular interactions in the cell (which we do not model). Therefore we cannot take the information from the previous day into account. In fact, modelling individual cells and following them over time would be great but we know that this is experimentally hardly feasible, so it would not be possible to compare predictions with experiments.

4.2.4 Parameter estimation

To optimize the parameter values, it is useful to consider two phases in the development of the leaf: one that represents day 5 to 17, when cell divisions in the leaf occur at a regular basis, implying that both cell division and expansion might happen simultaneously for some cells, and one from day 18 to 25, when the final cell numbers are reached and all cells stop dividing (Figure 4.7). Absence of cell division in the second stage of development reduces the number of parameters because both probabilities p_1 and p_2 are 0. The maximum threshold for cell division of pavement cells, T_{PC} is also irrelevant and could be eliminated. Therefore, g_{PC}, g_{GC}, and

M_{GC} were the only parameters that affected the model in the second phase and had to be estimated.

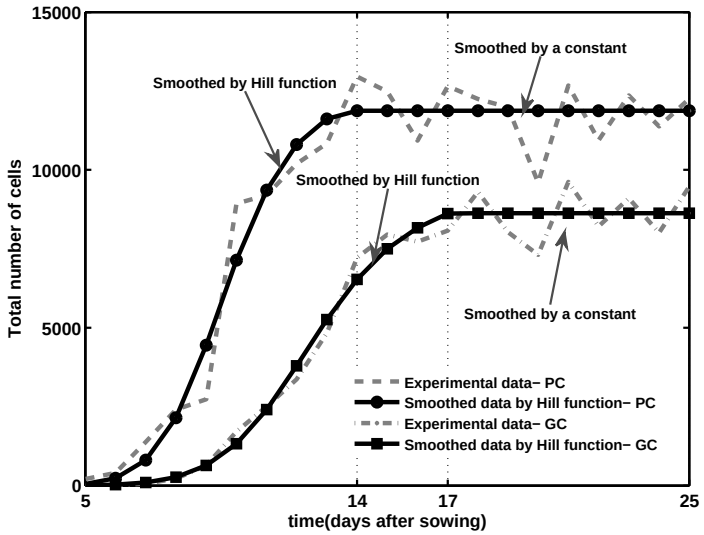


Figure 4.7: Experimental and smoothed total number of pavement and guard cells during leaf development. The total number of pavement cells increases until day 14; afterward, it becomes constant. To smooth the data, a Hill function was used to fit the data until day 14. For subsequent days, data were replaced by the average number of pavement cells between day 15 and day 25. The total number of guard cells increased until day 17, indicating that after day 14, still some pavement cells divided into guard cells. A Hill function was used to smooth the data until day 17. Afterwards, the data were replaced by the average number of guard cells between day 18 and day 25. PC and GC, mark pavement cells and guard cells, respectively.

In the first phase of development, some *a priori* restrictions to the model were imposed to reduce the range in which parameters needed to be optimized. As no cells can divide into guard cells and pavement cells simultaneously, we require $p_1 + p_2 \leq 1$. Division of all cells is supported by the uniform expression of cell cycle marker genes in the young leaf

[26]. Accordingly, in Table 4.3 that lists the number of cells during early leaf development, an approximately doubling of cells can be observed every $24h$, strongly indicating that nearly all cells divide thus, $p_1 + p_2 = 1$ at days 5, 6, and 7. Lastly, based on the size increase of guard cells over time (Figure 4.5B), the maximum threshold for precursor cells to divide into guard cells (M_{GC}) computed in the second phase, could also be applied to the first phase. With the model and the experimental distributions of the pavement and guard cells on a given day, we predicted the distributions for the next day. Although a best fit for the distribution of the guard cells at the earliest time points was difficult to find, because of the low number of guard cells at the early stage of leaf development, experimental and computed data fitted well from 9 DAS onward (Figure 4.8 and Figure 4.9).

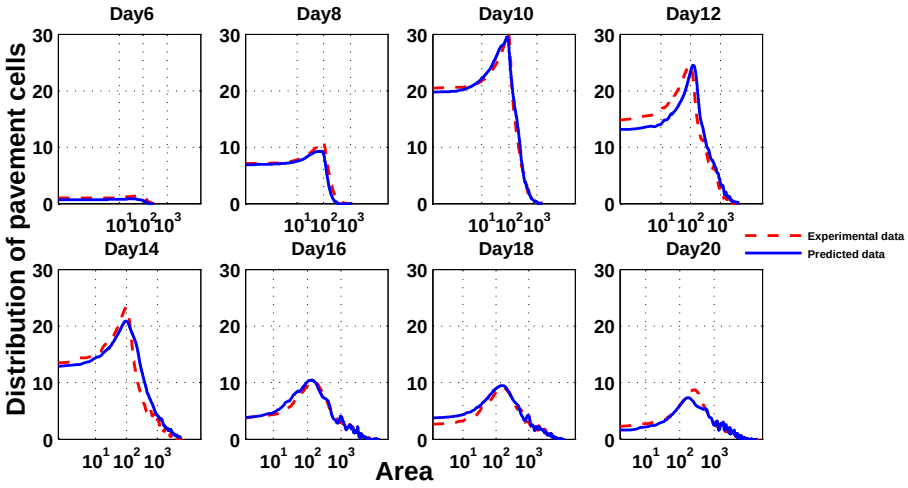


Figure 4.8: Comparison between experimental and model-predicted cell size distributions of pavement cells. A number of representative days are shown. In all graphs the correlation coefficient is nearly one.

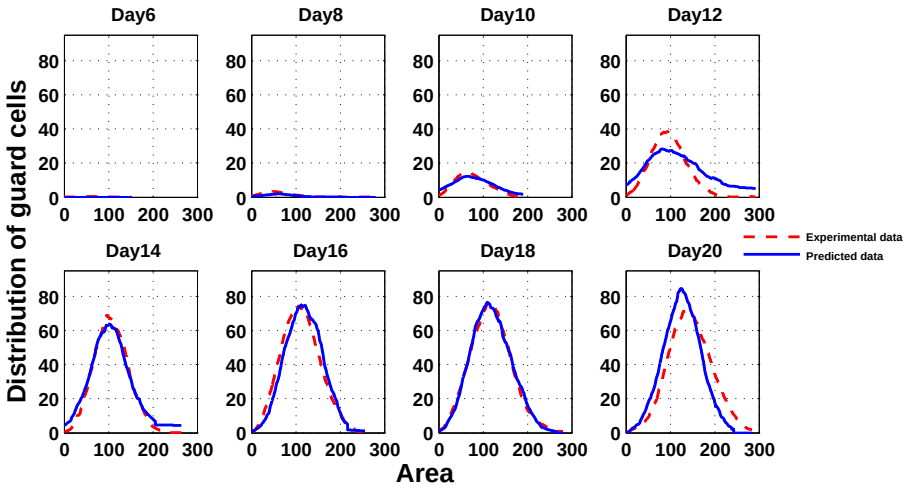


Figure 4.9: Comparison between experimental and model-predicted cell size distributions of guard cells. A number of representative days are shown. In all graphs the correlation coefficient is nearly one.

4.2.5 Constancy of the average cell cycle duration

Cell division rates measure how fast a cell progresses through the cell cycle. In the kinematic growth analysis, the cell division rate decreased progressively (Figure 4.4C). However, because the division rate was calculated based on the total number of cells in the leaf, the observed decrease might be due to a reduction in the proliferating fraction, an increase in the average cell cycle length, or a combination of both. To understand how the division rate is controlled at the cellular level, it is essential to quantify the fraction of cells dividing into pavement (p_1) and guard (p_2) cells.

As computed by the model, p_1 decreased gradually during leaf development and p_2 went up when p_1 went down (Figure 4.10A), indicating a shift from basal proliferation to cell division in the stomatal lineage. The probability p_2 had the largest values between days 10 to 14, meaning that most stomata were produced during these days. After day 14, both prob-

abilities dropped and decreased to 0, representing the exit of cell division during leaf development as indicated by an increase in probability p_3, in which more than 80% of the cells did not divide after day 16.

Because the model allowed us to split the total number of cells in a proliferative and an expanding population at the different stages of leaf development, a slow-down in the cell division rate and a reduction in the proliferative fraction could be discriminated. Cell cycle duration, which is a parameter in the mathematical model (T_c), could be derived from the cell division rate and was optimized for each day during leaf development. As discussed above, the impact of T_c in the model ended after day 18, when cell division stopped completely. Moreover, between days 16 and 18, the optimization results fluctuated a lot (Figure 4.11), implying that the proliferative fraction was too small to obtain relevant results. By contrast, for earlier time points, a stable output value was obtained. Remarkably, when different days were compared, the obtained T_c value was nearly constant (Figure 4.10B), indicating that the reduced cell division rate observed over the complete leaf during development was seemingly not caused by an increase in cell cycle duration, but solely by a decrease in the proliferative fraction within the leaf. We note that the cell cycle duration itself could differ from cell to cell, but in global perspective the average cell cycle duration was constant.

4.2.6 Growth rates of epidermal cells

The huge difference in maximum cell sizes found for pavement and guard cells suggests that both cell types might expand at distinct rates. Therefore, we used different average growth rates (per h) for pavement and guard cells. Note that in the model the two leaf growth-driving processes, cell division and cell expansion, were considered independently and that growth rates did only apply to expanding cells, but also to cell expansion during proliferation. In addition, the growth rates defined in the model allowed us

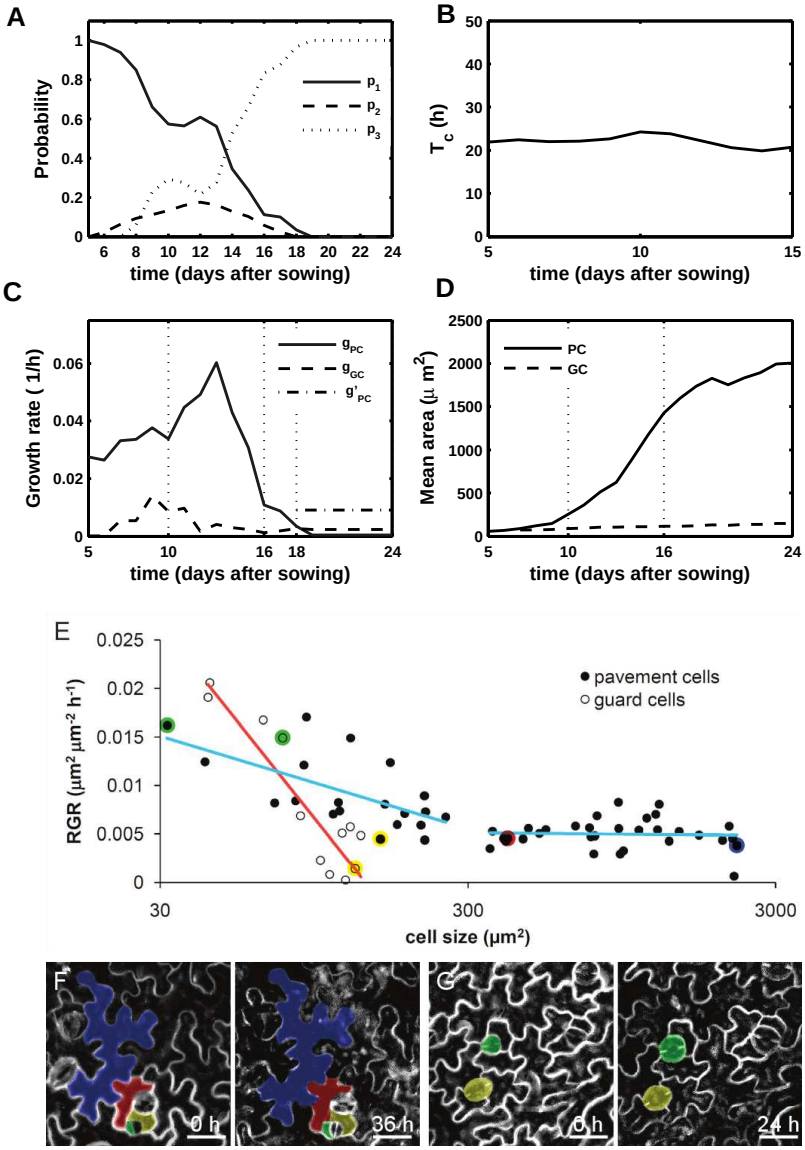


Figure 4.10: Predicted optimized parameters of the model and experimental validation. (A) Parameters of division. (B) Average cell cycle duration. (C) Average growth rates. g'_{PC} is the experimentally observed average growth rate of small pavement cells from days 18 to 25. (D) Experimental mean area. (E-G) Live imaging of epidermal cell growth in leaves. (E) Relative growth rate (RGR) measurements of pavement and guard cells on the abaxial epidermis of leaves 1 & 2 of 17 days-old-plants over a period of 36 h. Red and blue lines indicate a logistic regression for guard cells and for pavement cells smaller and larger than $300\mu m^2$. Color-coded data points represent high-lighted cells in F and G. (F) Epidermal cell pictures with 36 h interval highlighting pavement cells marked in E. (G) Epidermal cell pictures with 24 h interval highlighting stomata marked in E. Scale bars = $25\mu m$. PC,pavement cells. GC, guard cells.

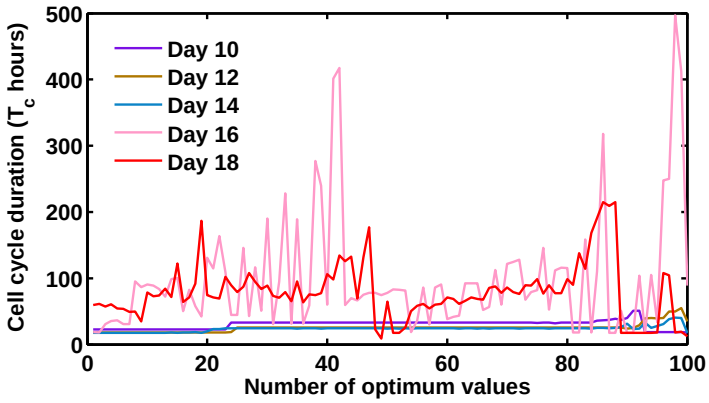


Figure 4.11: Result of 100 optimizations for the cell cycle duration. An optimization program for 100 different initial guesses for the parameters was implemented. The results were very similar until day 16. After day 16, the results for T_c became erratic because of too few dividing cells.

to calculate cell expansion rates for the different cell types studied (pavement cells and guard cells) in a mixed population and to compare them.

The model indicated that pavement cells had apparently a high and increasing growth rate from day 5 to day 13 (Figure 4.10C). From day 13 onward, the growth rate declined steeply, suggesting that pavement cells grew faster in the young proliferating leaf than in the older expanding leaf. Eventually, the growth rate reached a minimum value. These data are consistent with experimental data for the average cell area of pavement cells, because the mean area of pavement cells in the young leaves increased from $86\mu m^2$ to nearly $1500\mu m^2$ between days 7 and 16, but not significantly later on (Figure 4.10D). Because after day 18 pavement cells did not divide, the experimental size distribution graphs allowed to calculate the growth rate of small pavement cells ($< 300\mu m^2$) by measuring the shift in the distribution peak. The average growth rate for small cells obtained by the above experiment is $(0.009/h)$ whereas the average growth rate from day 18 until day 25 calculated by the model is $(0.0009/h)$ (Figure

4.10C). Therefore, by this analysis we found that during the last days of leaf development, small pavement cells grew 10-fold faster than the average population. Similar to the pavement cells, guard cells displayed higher growth rates in the young leaf compared to the old leaf. Initially the growth rate of pavement cells was higher than that of guard cells, but from day 18 onward, it was higher in stomata than in pavement cells, which means that guard cells stopped growing later than pavement cells (Figure 4.10C).

To confirm this cell growth data coming from the cell size distributions and predicted by the model, Stijn from PSB/VIB performed live imaging of epidermal cells in the leaf independently of the model. Confocal imaging allowed us to measure relative growth rates (RGR) of individual pavement and guard cells over a period of 36h. Cell tracking experiments on leaves at 17 DAS (it is the first day after which the majority of the cell stopped dividing.) confirmed that pavement cells smaller than $300\mu m^2$ grow faster than larger ones (p-value = $3.65e-5$, Student t-test) (Figure 4.10E,F and Table 4.1). The average RGR of small pavement cells decreased rapidly from about 0.015 to 0.005 when they reached a size of $300\mu m^2$. Afterwards, a relative constant RGR is maintained. Interestingly, we could also observe that small guard cells ($< 100\mu m^2$) grow faster than larger ones (p-value $= 0.0017$, Student t-test) (Figure 4.10E,G and Table 4.1). The guard cells do not have a biphasic growth pattern as was the case for pavement cells, but show a steady decrease in RGR with cell size. Strikingly, the average growth rate of pavement cells found by live imaging ($0.0065/h$) is close to those proposed by the model at the transition of day 17 to 18 ($0.0061/h$), independently confirming the model predictions (Figure 4.10C).

4.2.7 Non-existence of a size threshold for division

One of the parameters included in the model was the maximum guard cell size M_{GC}, which was designated a maximum threshold because pavement cells must have an area smaller than $2M_{GC}$ as a necessary condition for

Table 4.1: The area of color-coded cells in Figure 4.10. Pc and GC, pavement and guard cells.

time	0h	24h	36h	Relative growth rate
Small GC	1127	1919	-	1.70
Large GC	1872	1967	-	1.05
Small PC	306	-	453	1.48
Large PC	17486	-	19852	1.14

division into guard cells. As discussed above, the threshold for guard cells was calculated easily in the second stage of development (from days 18 to 25) when no cell divided. When we assumed that the threshold was constant, the parameter optimization procedure provided a value equal to $354\mu m^2$ for M_{GC}, which was only a little above the maximum guard cell size (M_{GC}) of $300\mu m^2$ found in the experimental data, illustrating the accuracy of the optimization methods. The value of M_{GC} indicated that the pavement cells with area larger than approximately $700\mu m^2$ had no ability to divide into guard cells.

In contrast to the robust value obtained for M_{GC} in nearly all simulations, the optimized parameter value for the threshold for pavement cells above which no divisions into pavement cells occur (T_{PC}), yielded erratic values. In cases with an optimum solution, a strong deviation between the experimental cell size distribution (used to build the model) and computed cell size distribution profiles was observed (Figure 4.12A,B) which indicates that the found optimum is a spurious one. To strengthen this observation, simulations were done with values for $T_{PC} = 300, 500, 1000,$ and $10{,}000\mu m^2$ at day 11 to 12, when the leaf consists of both dividing and expanding cells. Consistent with the results of the behaviour of the model with fitted T_{PC}, small fixed values of T_{PC} resulted in a sharp deviation from the experimental data (Figure 4.12C). As the T_{PC} value increased, the discrepancy between the experimental and computed data gradually disappeared, which might be explained by the fact that most cells were

smaller than the threshold and, thus, the T_{PC} value became less relevant. These observations demonstrate that the imposition of a threshold value for cell division in pavement cells is not compatible with the assumptions underlying the model. The non-existence of a threshold for pavement cell indicates that the mere size of a cell does not exclude it from dividing, i.e. size alone does not prevent division.

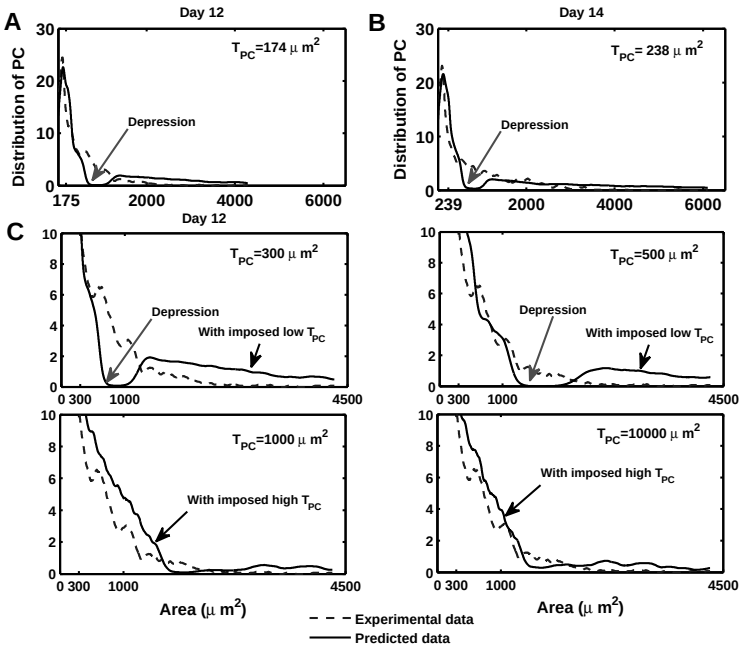


Figure 4.12: Distribution of pavement cells with optimized and imposed T_{PC} values. (A and B) Comparison between the experimental and predicted pavement cell size distribution with optimized T_{PC} for day 12 (A) and day 14 (B). (C) Comparison for different imposed values of T_{PC} at day 12. PC, pavement cells.

4.3 Discussion

Leaf development of *Arabidopsis* is driven by two main processes, cell division and cell expansion [34]. As both processes are intimately intertwined, their individual contribution to leaf growth is not easily studied experimentally. Moreover, the leaf epidermis in its simplest form consists at least of two different cell types, the pavement cells and stomatal guard cells, each with specific growth characteristics. We developed a mathematical model based on the fates of the two main different epidermal cells to fit experimentally determined cell size distributions during the development of the epidermis of a young dividing leaf into an adult organ. This model allowed us to disentangle cell division and cell expansion parameters for the individual cell types.

4.3.1 Differential cell expansion within the leaf epidermis

Individual cell size measurements yielded cell area distributions during leaf development. When comparing the growth rate of guard cells and pavement cells, our data show that the growth of these two cell types follows distinctly different dynamics (Figure 4.5A,B and Figure 4.10). The guard cells initially grow slower but continue to grow longer, whereas the larger pavement cells initially grow faster, but their growth stops around day 16. Differential growth rates between adjacent cells is surprising and have to our knowledge not been investigated in leaves.

In the root, this problem was studied by measuring cell sizes of different cell types [4]. Root growth is linear and its cell growth is symplastic, meaning that neighbouring cells grow uniformly, so that adjacent walls do not alter positions relative to each other. As a consequence, at a given distance from the tip, by definition all cells have the same relative elongation rate [33]. This constraint implies that differences in mature cell sizes between tissues reflect differences in cell proliferation [4]. In compari-

son, leaf growth is much more complex. As a leaf is a flat, exponentially growing structure with small and large cells dispersed over the epidermal surface, several parameters have to be taken into account in the equations representing leaf growth.

To validate the predictions of differential growth rates between cells we performed time lapse cell tracking experiments on 17 day old leaves. The results clearly indicate that small pavement and guard cells at the end of the cell expansion phase grow faster than larger cells of their representative cell type (Figure 4.10E,F-,G). Furthermore, modelling results indicated that pavement cell growth rate is strongly regulated during development, peaking during the early leaf expansion phase, followed by a rapid reduction until shut down at maturity (Figure 4.10C). these data indicate that cells exiting the division status grow faster than those being in full expansion phase. Taken together, these data reveal substantial cell growth differences within the epidermal layer. Because guard cells, and small and large pavement cells are dispersed throughout the leaf due to the patterning in the stomatal lineage (Figure 4.10F), our data suggests that adjacent guard and pavement cells of different sizes expand at different relative rates within the developing epidermis. Differential expansion of tissue layers in cylindrical organs, such as stems and roots, results in tissue tension, often motioning plants, as in the case of shoot phototropism and root gravitropic curvature [53, 80]. Epinasty of the leaves is also explained by differential growth of the abaxial and adaxial sides [40]. Within one single tissue layer of a flat tissue structure, such as the abaxial leaf epidermis, it is more difficult to translate tension into motion in order to release pressure. Breaking up symplastic growth in the root by uncoordinated differences in cell expansion between tissue layers cause cells to rupture and distortion of root growth [84, 88]. However in the leaf no cell ruptures are detected within the epidermis.

Possibly, the formation of lobes in the rigid cell wall of epidermal pavement cells represents a way to cope with the tissue tension. Remarkably,

the characterisitc jigjaw puzzle shape of pavement cells is established at the proliferation-to-expansion transition. At this time, the homogenously proliferating epidermis changes toward a tissue that contains guard cells and dividing and expanding pavement cells. Thus, the appearance of the puzzle shape fits with the moment at which differential cell expansion rates would first appear. Furthermore, it was shown that the emergence of lobes leading to the puzzle shape is preceded by the reorganization of cortical microtubules [45, 62], and that application of external mechanical stress to a tissue results in realignment of microtubules parallel to maximal stress directions [35]. These observations suggest that tissue tension within the epidermis layer could realign cortical microtubules, triggering puzzle-shape formation.

Interestingly, at maturity, the guard cells in Arabidopsis leaves are elevated above the surrounding pavement cells. Faster relative growth of the guard cells and surrounding smaller pavement cells compared to the larger pavement cells, which make up the bulk of the tissue area, could explain the development of this elevation.

4.3.2 Control of cell cycle duration

In the root, divisions in the meristematic zone ensure the constant cell production and indeterminate growth of the root. While cells mature, they are displaced from the root tip and enter the elongation zone [6]. Thus, in the root, the developmental stages are separated in space. Furthermore, in the root meristem, practically all cells divide actively [5, 8], implying that the cell division rate calculated for the meristem is a good overall average. In contrast, in leaves, the developmental stages are separated in time. Moreover, at the proliferation to expansion phase transition, the cell cycle arrest front that follows a longitudinal gradient [24] is accompanied by the differentiation of the first stomata [21], whereas cells at earlier stages in the stomatal lineage [39], dispersed through the leaf, keep on proliferating for

multiple days. Because of these constrains, the average cell division rate calculated on the basis of an increase in total cell number can only be measured experimentally during the early developmental stages, when all cells participate in division.

Our mathematical model yielded probabilities for a cell to divide into two pavement cells, into two guard cells, or to exit the mitotic cycle (Figure 4.10A), allowing the calculation of the proliferative fraction and the corresponding average cell cycle duration throughout complete leaf development (Figure 4.10B). Surprisingly, although cell cycle length can vary significantly between adjacent cells within the the shoot apical meristem (SAM)[5] and petals [69, 72], we found that the average cell cycle duration over the complete leaf epidermis remained constant during development. This observation is analogous to the situation in roots, where average cell cycle duration is considered to be constant between cell types and between different positions in the meristem [3, 34].

Moreover, the average cell cycle duration of approximately 20 h found for the leaf (Figure 4.10B) is comparable to that for the root [34]. These data indicate that the constancy of cell cycle duration is widespread and suggest that the basal cell division rate is strongly conserved on a cellular level [3].

4.3.3 Interaction between cell division and cell growth

For decades, cell biologists have been interested in the process that links cell size and cell division. The interaction between cell division and cell growth in plants is currently unclear. Some reports have argued that epidermal pavement cells of the Arabidopsis leaves divide only rarely at a cell size larger than $400\mu m^2$ [24, 31], suggesting a possible size threshold preventing cell division. By contrast, a study on the root meristem hinted at

[5]The shoot apical meristem (SAM) is a population of cells located at the tip of the shoot axis. It produces lateral organs, stem tissues and regenerates itself.

the lack of a maximum cell size threshold for cell division, because cells divide at very different sizes [6, 39]. Similarly, our mathematical model suggests that there is no threshold for cell division, because such a threshold would result in a deviation from the observed cell size distributions (Figure 4.12). Furthermore, our cell size distributions show that the average pavement cell size increases during the proliferation phase (Figure 4.5A). These observations suggest that in the leaf epidermis cells divide at different sizes and they provide extra evidence for the non-existence of a fixed threshold size for division in plants. Moreover, the increasing cell size during proliferation points toward a disequilibrium between cell growth and cell division. The constant cell cycle duration implies that this increase in cell size is due to an increase in cell growth rate, uncoupling cell size and cell division, and reflects also that the plasticity of the cell growth rate depends on the developmental context. In conclusion, in plants, cell expansion and cell division might occur independently and the cell cycle might act more as a timer mechanism than as a size-regulated machinery.

4.4 Materials and Methods

4.4.1 Plant material and growth conditions

We used wild-type *Arabidopsis thaliana* plants of the Columbia (Col-0) ecotype to perform a kinematic analysis of leaf growth. The plasma membrane marker 35S::GFP-PIP2a was conducted for live imaging of epidermal cell growth [19]. Seeds for in vitro analysis were sterilized in 3% bleach for 15 min and sown on medium containing 0.5x Murashige and Skoog medium (Duchefa) solidified with 0.9 gL plant tissue culture agar (Lab M) on round plates (1013; Becton-Dickinson). After a stratification period of 2 d, the plates were placed in a growth chamber under long-day conditions (16 h of light, 8 h of darkness) at $22°C$ with a light intensity of 80 to 100 $mE/m^2/s$ supplied by cool-white fluorescent tubes (Spectralux

Plus 36W/840; Radium).

4.4.2 Kinematic growth and image analysis

Kinematic analysis of leaf growth was performed as described earlier [21]. Briefly, from day 5 (when cotyledons started to expand) until day 21 after sowing (when leaves were fully expanded), leaves were harvested daily. Plants were placed in methanol overnight to remove chlorophyll, and subsequently they were cleared and stored in lactic acid for microscopy. The youngest plants were mounted whole on a slide and covered. Older primordia that had visible petioles (leafstalk) were dissected, whereas younger primordia were left on the plant. The primordia were observed with differential interference contrast optics on a microscope (DMLB; Leica, Wetzlar, Germany). The total leaf (blade) area of the oldest two primordia (leaves 1 and 2) of each seedling was determined from drawing tube images that were scanned with a flatbed scanner using the public domain image analysis program ImageJ (version 1.17y; http://rsb.info.nih.gov/ij/). At older stages, the primordia were digitized directly with a charge-coupled device camera mounted on a binocular (Stemi SV11; Zeiss, Jena, Germany), which was connected to a personal computer fitted with a frame-grabber board LG3 (Scion Corp., Frederick, MD) running the image analysis program Scion Im- age (version 3b for Windows NT). Cell density and stomatal index were determined from scanned drawing tube images of outlines of at least 100 cells of the abaxial epidermis located 25 and 75% from the distance between the tip and the base of the leaf primordium (or blade once the petiole was present), halfway between the midrib and the leaf margin. In the youngest primordia (up to day 6), a single group of cells was drawn. The microscopic drawings were scanned for digitization. An in-house developed image analysis algorithm was subsequently used to automatically extract detailed measurements, such as total area of the drawing, total number of cells and number of stomata, from the microscopic drawings. Kinematic growth characteristics were calculated and plotted

from these measurements. Extension of the image analysis algorithms allowed for discrimination between pavement and guard cells and individual cell area measurements for these two cell types separately. Guard cells were extracted as cells $< 500\mu m^2$ neighbouring the stomatal pores. The remaining non-guard cells were identified as pavement cells. These cell measurements were pooled by day and allowed the construction of cell area distributions. Therefore, cell areas of pavement and guard cells were classified into bins of $400\mu m^2$ and $25\mu m^2$, respectively. Relative and absolute frequency distributions were plotted using MATLAB. The image analysis algorithms were written in C++ scripting and make use of the SDC Morphology Toolbox for C++ (www.mmorph.com/cppmorph/) (done by Stijn Dhondt[6]).

4.4.3 Live imaging of epidermal cell growth in the leaf

At 16 DAS three complete plants of the plasma membrane marker 35S::GFP-PIP2a [19] were transfererd in aseptic conditions from in vitro plates to a round microscopic chamber (Warner instruments) with liquid growth medium containing 0.5x Murashige and Skoog (Duchefa). One leaf of the first leaf pair of each plant was flattened on the bottom of the chamber by overlaying it with a block of solidified agar. The abaxial epidermis was imaged by a confocal microscope with software package LSM510 (Zeiss) from 17 to 18 DAS. Z-stacks were recorded at five different positions with 12 h intervals over a total period of 36 h. Epidermal cells in the Z-stacks were projected onto a single reconstructed view by using the Extended Depth of Field plug-in [29] for ImageJ (http://rsbweb.nih.gov/ij). Cell lineages were manually tracked and measured at each time point using ImageJ. The relative growth rate (RGR) for each cell was calculated as the average RGR of this cell over the different time intervals. Growth rates of 53 pavement cells and 13 guard cells were tracked over time (done by Stijn

[6]PSB/VIB - Ghent University

Dhondt[7]).

4.5 Modelling

We built a general mathematical model that maps the distributions of pavement and guard cells on a given day of leaf development to the distributions on the next day. First, experimentally obtained kinematic data were used to obtain the distribution of pavement and guard cells for whole leaf during development. These distributions for a given day were used as an input for our model to predict the distributions of the next day. Then using the optimization method, we fitted the predicted distributions to the experimental distributions to estimate the parameters which included in the model (Figure 4.13).

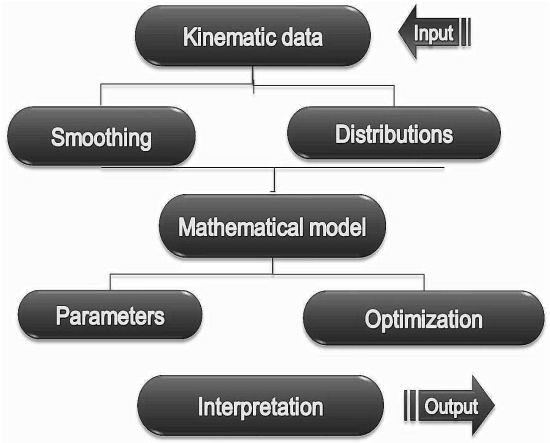


Figure 4.13: Scheme of the process for estimating the important parameters related to leaf development using kinematic data and a mathematical model.

[7]PSB/VIB - Ghent University

4.5.1 Measurement of total cell numbers

By using image analysis, the area and the number of cells in the two zones of a leaf (top and bottom) were measured. We put all these data together and measured the area of these zones. From then on, we call the small part of the leaf the union of the two parts mentioned above. The average cell area is also obtained from these data and by extrapolation, we approximately compute the total number of cells in the whole leaf. The following calculations show how we find the total number of pavement and guard cells separately in the whole leaf. For a given day i let $N^i(a_1)$ and $N^i(a_2)$ be the total number of cells including pavement and guard cells in a small part of a leaf with area a_1 and in the whole leaf with area a_2 respectively. $N^i(a_1)$ is experimentally obtained by image processing. So we have

$$\begin{aligned} s^i &= \frac{a_1}{N^i(a_1)} = \text{average cell area}, \\ N^i(a_2) &= \frac{a_2}{s^i}. \end{aligned} \tag{4.1}$$

Guard cells are also counted in the chosen parts to find the stomatal index SI^i at day i:

$$SI^i = \frac{N^i_{GC}(a_1)}{N^i(a_1)} \tag{4.2}$$

where $N^i_{GC}(a_1)$ is the number of guard cells in the area a_1 at day i. To compute the total number of pavement ($N^i_{PC}(a_2)$) and guard cells ($N^i_{GC}(a_2)$) in a typical leaf, we have

$$N^i_{GC}(a_2) = SI^i N^i(a_2), \tag{4.3}$$

$$N^i_{PC}(a_2) = N^i(a_2) - N^i_{GC}(a_2) = N^i(a_2)(1 - SI^i). \tag{4.4}$$

4.5.2 Smoothing total cell numbers

Smoothing the data set attempts to capture important patterns in the data, while leaving out noise or rapid phenomena. Many different algorithms

can be used in smoothing. For smoothing the total numbers of guard and pavement cells computed by (4.3) and (4.4) to reduce the biological variance, we divide them in two parts based on their bending points. The first part includes the data from day 5 to day 14 for pavement cells and from day 5 to day 17 for guard cells. A parametrized function is used to fit a smooth curve on the data of the first part. The Hill function, cf. §3.1, and a composition with an exponential function are two candidates for this fitting:

$$H_f(t, a_{max}, K_c, n) = a_{max}\frac{t^n}{K_c^n + t^n}, \tag{4.5}$$

$$E(t, a_{max}, K_c, K_d, n) = a_{max}e^{(\frac{t+K_c}{nt+K_d})^3}. \tag{4.6}$$

The least squares method was used to optimize the parameters of these functions i.e. a_{max}, K_c, n, K_d. The outcome is given in Table 4.2. These two techniques lead to nearly identical results. For simplicity we choose the Hill function as the curve to smooth the data (Figure 4.7).

Table 4.2: Parameters of the Hill function and the exponential function for pavement and guard cells. Pc and GC, pavement and guard cells.

Functions	$\mathbf{a_{max}}$	$\mathbf{K_c}$	**n**	$\mathbf{K_d}$
Hill - PC	12628.80	9.69	8.29	-
Hill - GC	9470.90	12.64	7.77	-
Exponential - PC	12656.56	-17.55	0.51	3.93
Exponential - GC	8780.05	-18.96	0.03	6.95

Since the total number of cells does not increase after the bending points, the second part of the data (data after the bending points) is smoothed by a constant which is the average value of the total number of pavement and guard cells separately for the second part of data. These smoothed data are used to compute the distributions. The correlation coefficients for

smoothed and experimental pavement and guard cell numbers were calculated to be 0.9760 and 0.9895 respectively.

4.5.3 Estimation of average cell cycle duration

We estimate the average cell cycle duration in the first few days of development (days 5, 6 and 7) using the assumption that nearly all pavement cells continuously divide.
Let i, T_c and N_i denote the day, time duration of the cell cycle and the number of continuously dividing cells at day i, respectively. Thus after T_c hours the number of cells become $2N_i$ and after $2T_c$ hours it rises to $2^2 N_i$ cells and so on. So we have,

$$N_{i+kT_c} = 2^k N_i, \tag{4.7}$$

where N_{i+kT_c} is the number of cells, kT_c hours later than day i. (4.7) holds also for non-integer values of k. Therefore $k = \log_2 \frac{N_{i+24h}}{N_i}$ and $T_c = \frac{24h}{k}$. Note that in this case, N_i is the total number of pavement cells at day i and N_{i+24h} is obtained by the total number of cells at day $i+1$ minus all number of guard cells produced before day $i+1$. The result of the estimation is reported at Table 4.3 which shows that the average cell cycle duration in the first few days of development is about $25.6 \pm 5.4h$. The average cell cycle duration is one the parameters involved in our mathematical model and for safety we keep $T_c > 18h$ and add this constrain in the model. It will be estimated more precisely using optimization.

4.5.4 Size distributions

The size distributions of pavement and guard cells in a discrete set of area values x on a given day i was defined by:

$$d^i_{PC}(x) \;=\; \frac{N^i_{PC}(x-\varepsilon_1, x+\varepsilon_1)}{N^i_{PC}(a_1)} \frac{N^i_{PC}(a_2)}{2\varepsilon_1}, \tag{4.8}$$

Table 4.3: Initial estimation of average cell cycle duration. Both experimental and smoothed (obtained by averaging method) cell numbers are given. For estimation of the cell cycle duration the smoothed data were used, with exception of the day 5 to day 6 transition, for which no smoothed data could be obtained. PC, pavement cells; GC, guard cells; T_C, cell cycle duration.

	Experimental data[a]		**Smoothed data**		
day	**PC**	**GC**	**PC**	**GC**	**T_c(hours)**
5	205	0	205	0	24.2
6	404	3	665	21	21.0
7	1385	61	1398	93	31.6
8	2404	216	2176	305	-
Average of T_c					$25.6 \pm 5.4h$

[a] Average of five independent measurements

$$d^i_{GC}(x) \;=\; \frac{N^i_{GC}(x-\varepsilon_2, x+\varepsilon_2)}{N^i_{GC}(a_1)} \frac{N^i_{GC}(a_2)}{2\varepsilon_2}, \tag{4.9}$$

where ε_1 and ε_2 are small positive values, x is the area of the cell and $N^i_{PC}(x-\varepsilon_1, x+\varepsilon_1)$ is the number of pavement cells with area in the range $(x-\varepsilon_1, x+\varepsilon_1)$ in the small part of the leaf, and similarly $N^i_{GC}(x-\varepsilon_2, x+\varepsilon_2)$ is the number of guard cells with area in the range $(x-\varepsilon_2, x+\varepsilon_2)$ in the small part of the leaf. $N^i_{PC}(a_2)$ and $N^i_{GC}(a_2)$ are the total number of pavement and guard cells, respectively; $N^i_{PC}(a_1)$ and $N^i_{GC}(a_1)$ are the total number of pavement and guard cells, respectively, in the small part of the leaf measured by experiment.

We note that guard cells are much smaller than pavement cells. Therefore $\varepsilon_2 < \varepsilon_1$. $\varepsilon_1 = 80\mu m^2$ and $\varepsilon_2 = 30\mu m^2$ were chosen after some trial and error (for very small ε the values of d_{PC} and d_{GC} are erratic; for too large values of ε the distributions are flattened too much). The distribution values were stored in a discrete set of area values. We use 10000 areas for pavement cells, uniformly distributed between 0 and the maximum size of pavement cells and 5000 area values for guard cells, again uniformly distributed between 0 and the maximum size of guard cells. To reduce the

errors, we smoothed the distributions by an averaging method. We took average values of 50 data after and 50 data before, for pavement cells and 30 data after and 30 data before, for guard cells. The first and last 50 data of pavement cells and the first and last 30 data of guard cells were not smoothed (Figures 4.14 and 4.15). We note that the correlation coefficients for the smoothed and original distributions of pavement and guard cells in each day are larger than 0.99.

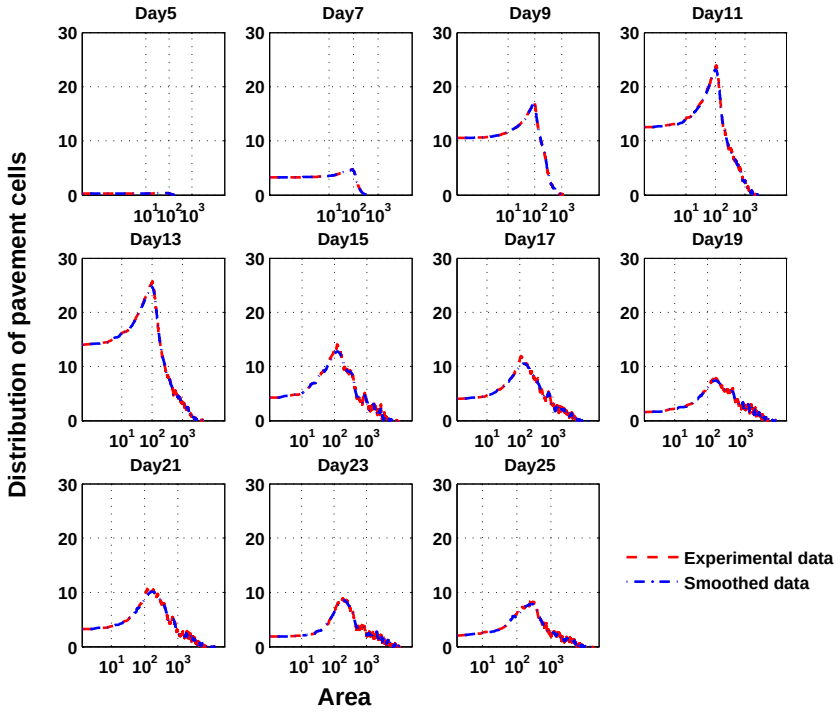


Figure 4.14: Logarithmic plot of experimental and smoothed distribution of pavement cells during leaf development.

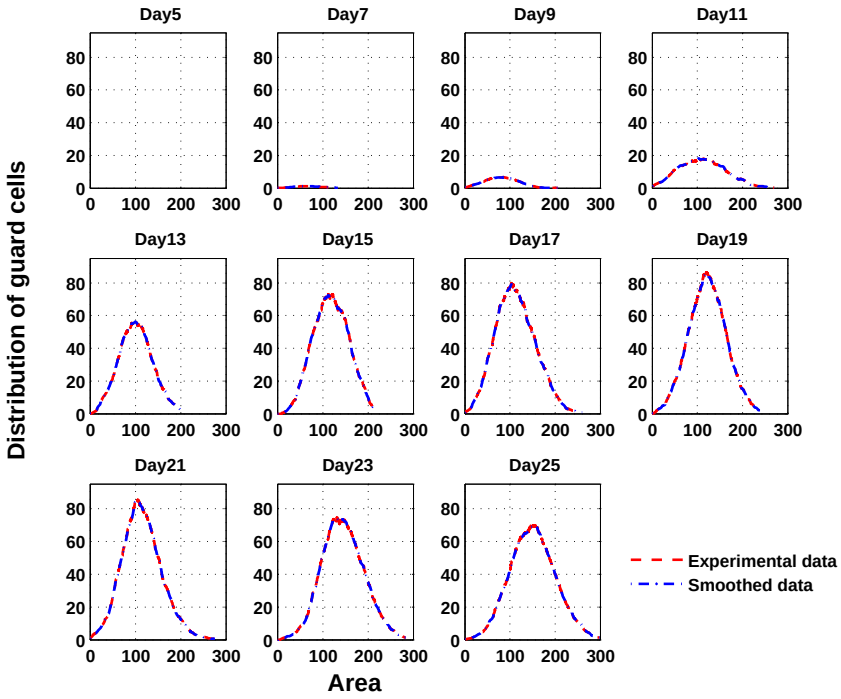


Figure 4.15: Experimental and smoothed distribution of guard cells during leaf development.

4.5.5 Exponential growth

For simplicity we model the growth of pavement cells by the equation $\frac{ds}{dt} = g_{PC}s$ if no division occurs. Here s is the average area of the pavement cells and g_{PC} is the growth factor at a given time point, so it can change from day to day. During a short time span (day 24 hours) we then have $s(t) = e^{g_{PC}t}s(0)$, where $s(0)$ is the average area of pavement cells at $t = 0$, cf. §3.5. According to above formula, the area of pavement cells can grow without restriction. Note that if a symmetric division occurs, the area of the cell become half of the area in case no division occurs.

For the guard cells, since their area cannot grow to more than M_{GC}, we propose the formula $\frac{ds}{dt} = g_{GC}s(1 - \frac{s}{M_{GC}})$ for their growth, cf. §3.5.2. We note that g_{GC} also varies from day to day. Similarly, s is the average area of the guard cells at a given time t.

4.5.6 Map of distributions

We constructed a map from (d^i_{PC}, d^i_{GC}) to $(D^{i+1}_{PC}, D^{i+1}_{GC})$ to build a link between the distribution of cells on a given day i to predicted distributions of the cells at the next day $i + 1$. Each cell with area a at day $i + 1$ has a precursor cell with area a^* at day i, but there are several scenarios by which this can happen. We call each scenario a flow and each flow can be described mathematically by a partial map from (d^i_{PC}, d^i_{GC}) into $(D^{i+1}_{PC}, D^{i+1}_{GC})$. The sum of these maps defines the global flow from (d^i_{PC}, d^i_{GC}) to $(D^{i+1}_{PC}, D^{i+1}_{GC})$. Although all flows are different, they have a basic structure in common.

Let $F^{i+1}_k(a)$ (respectively, $G^i_k(a^*)$) be the distribution of cells (pavement cells or guard cells) with average area a at day $i + 1$ (respectively, a^* at day i) related to the kth flow and let $f_{2,k}$ be the function defined by the requirement that $f_{2,k}(a) = a^*$. Let n_k be the number of cells at day $i + 1$ that originate from one cell at day i through the flow F_k and let ε be a small number. Suppose that

$$f_{2,k}(a - \varepsilon) = a^* - \varepsilon^*, \tag{4.10}$$

$$f_{2,k}(a + \varepsilon) = a^* + \varepsilon^*. \tag{4.11}$$

If $N_k(\varepsilon)$ is the number of cells at day $i+1$ in $[a - \varepsilon, a + \varepsilon]$ that is involved in the k-th flow then this corresponds to $\frac{N_k(\varepsilon)}{n_k}$ cells in $[a^* - \varepsilon^*, a^* + \varepsilon^*]$ at day i (Figure 4.16).
The distribution at day $i + 1$ is then given by

$$F^{i+1}_k(a) = \lim_{\epsilon\to 0} \frac{N_k(\epsilon)}{2\epsilon} = \lim_{\epsilon\to 0} \frac{\frac{N_k(\epsilon)}{n_k}}{2\epsilon^*} n_k \lim_{\epsilon\to 0} \frac{\epsilon^*}{\epsilon}$$

$$= \quad G_k^i(a^*)n_k \lim_{\epsilon \to 0} \frac{\epsilon^*}{\epsilon}. \tag{4.12}$$

Now by (4.10) and (4.11) we have,

$$\frac{f_{2,k}(a+\varepsilon) - f_{2,k}(a-\varepsilon)}{2\varepsilon} = \frac{\varepsilon^*}{\varepsilon}. \tag{4.13}$$

Hence $\lim_{\varepsilon \to 0} \frac{\varepsilon^*}{\varepsilon} = \frac{df_{2,k}(a)}{da}$. We conclude that

$$F_k^{i+1}(a) = G_k^i(f_{2,k}(a)).f_{1,k}(a), \tag{4.14}$$

where $f_{1,k}(a) = n_k \frac{df_{2,k}(a)}{da}$.

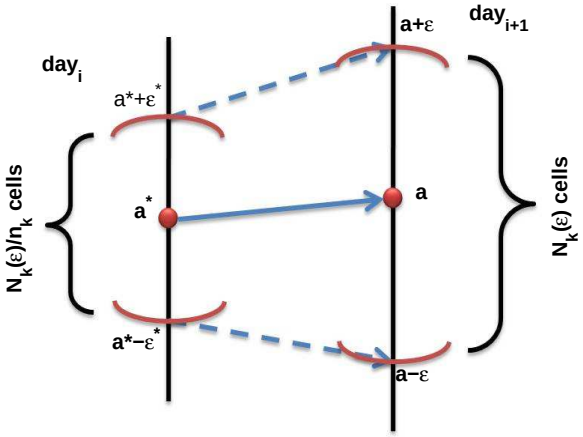


Figure 4.16: Changing the number of cells and the cell size in the k-th flow from day i to day $i+1$.

4.5.7 Flows

All the possible flows for pavement cells and one flow related to guard cells are defined here (Figure 4.6).

Flow F_1 from PC to PC without division

This flow deals with pavement cells which are not in the process of dividing (Figure 4.6B F_1). This includes firstly all pavement cells with area larger than T_{PC}, secondly a fraction $(1-p_1)$ of those with area less than T_{PC} and larger than $2M_{GC}$ and thirdly a fraction $(1-p_1-p_2)$ of those with area less than T_{PC} and $2M_{GC}$ (Figure 4.6A). The area of a pavement cell at day $i+1$ (a) is given by:

$$a = a^* e^{g_{PC}(24h)} \Longrightarrow a^* = a e^{-g_{PC}(24h)} = f_{2,1}(a).$$

Therefore, it holds that

$$\begin{aligned} G_1^i(a^*) &= d_{PC}^i(a^*)[H(n,a^*,T_{PC}) + (1-p_1-p_2)(1-H(n,a^*,T_{PC})) \\ &\quad (1-H(n,a^*,2M_{GC})) + (1-p_1)(1-H(n,a^*,T_{PC})) \\ &\quad (H(n,a^*,2M_{GC}))], \\ F_1^{i+1}(a) &= G_1^i(a^*).f_{1,1}(a) = G_1^i(f_{2,1}(a)).f_{1,1}(a), \qquad (4.15) \\ f_{1,1}(a) &= \frac{df_{2,1}(a)}{da} = e^{-g_{PC}(24h)}. \end{aligned}$$

From this moment onwards, $H(n,a^*,x) = \frac{a^{*n}}{a^{*n}+x^n}$, is a Hill function, cf. §3.1 and x is one of the thresholds T_{PC} or $2M_{GC}$. In the all computations a^* is a variable and $n = 20$ is fixed (to get a proper slope). Therefore, $H(n,a^*,T_{PC})$ is nearly 1, for $a^* > T_{PC}$ and it is nearly 0, for $a^* < T_{PC}$. Similarly, $H(n,a^*,2M_{GC})$ is nearly 1, for $a^* > 2M_{GC}$ and it is nearly 0, for $a^* < 2M_{GC}$. Note that in this flow, the factor $n_k = 1$ because there is no division.

Flow F_2 from PC to PC with one division for $18h \leq T_c \leq 24h$

If $18h \leq T_c \leq 24h$ then a fraction p_1 of pavement cells with area less than T_{PC} divides at least once and at most twice during one day (Figure

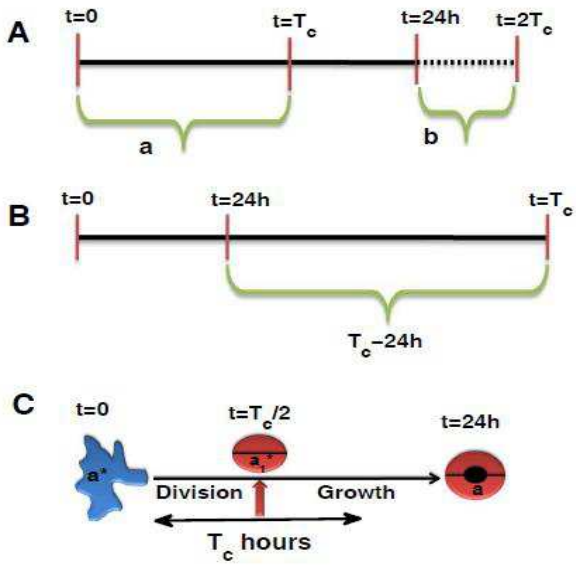


Figure 4.17: Division parts related to the flows. (A) Flows $F2$ and $F3$. (B) Flows $F4$ and $F5$. (C) Division process of flow $F6$.

4.6B F_2). Figure 4.17A shows the division moments on the time axis. For simplicity we assume that on a given day all cells in the same category are uniformly distributed over the time axis. By this hypothesis, a fraction $\frac{2T_c-24h}{T_c}$ of the corresponding pavement cells have the second division in part B (they already divide in part A). The rest fraction $\frac{24h-T_c}{T_c}$ of the corresponding pavement cells divide twice during one day.
Here we consider the first part with one division. Then the area of pavement cell at day $i+1$ (a) is given by:

$$a = \frac{1}{2}a^* e^{g_{PC}(24h)} \Longrightarrow a^* = 2ae^{-g_{PC}(24h)} = f_{2,2}(a).$$

Therefore, it holds that

$$G_2^i(a^*) \;=\; \frac{2T_c - 24h}{T_c} d_{PC}^i(a^*) p_1 [1 - H(n, a^*, T_{PC})]$$

$$
\begin{aligned}
& [1 - H(n, T_c, 24h)]H(n, T_c, 18h), \\
F_2^{i+1}(a) &= G_2^i(a^*) f_{1,2}(a) = G_2^i(f_{2,2}(a)) f_{1,2}(a), \qquad (4.16) \\
f_{1,2}(a) &= 2\frac{df_{2,2}(a)}{da} = 4e^{-g_{PC}(24h)}.
\end{aligned}
$$

Similarly, $H(n, T_c, 24h)$, is a Hill function, cf. §3.1. It is nearly 1, for $T_C > 24h$ and nearly 0, for $T_C < 24h$. The same goes for $H(n, T_c, 18h)$. Note that in this flow, the factor $n_k = 2$ because there is only one division.

Flow F_3 from PC to PC with two divisions for $18h \leq T_c \leq 24h$

For simplicity, we assume that when a cell is actively dividing, all cells which originate from it are also actively dividing during the same day. In flow F_3 both daughter cells born in the first division will divide once again in the same day. So the case in which one daughter cell divides and the other does not is not considered. We note that, since $T_c > 18h$ not many cells involve in such a case.

According to Figures 4.6B F_3 and 4.17A, we obtain

$$a = \frac{1}{4} a^* e^{g_{PC}(24h)} \Longrightarrow a^* = 4ae^{-g_{PC}(24h)} = f_{2,3}(a).$$

Therefore, it holds that

$$
\begin{aligned}
G_3^i(a^*) &= \frac{24h - T_c}{T_c} d_{PC}^i(a^*) p_1 [1 - H(n, a^*, T_{PC})] \\
& \quad [1 - H(n, T_c, 24h)]H(n, T_c, 18h), \\
F_3^{i+1}(a) &= G_3^i(a^*) f_{1,3}(a) = G_3^i(f_{2,3}(a)) f_{1,3}(a), \qquad (4.17) \\
f_{1,3}(a) &= 4\frac{df_{2,3}(a)}{da} = 16e^{-g_{PC}(24h)}.
\end{aligned}
$$

Note that in this flow, the factor $n_k = 4$ because there are two divisions.

Flow F_4 from PC to PC with one division for $T_c > 24$

If $T_c > 24h$ then the fraction p_1 of the pavement cells with area less than T_{PC} can divide at most once during one day (Figure 4.6B F_4). Figure 4.17B shows that a fraction $\frac{24h}{T_c}$ of those divides once and a fraction $\frac{T_c-24h}{T_c}$ does not divide during a given day and they divide later.
Here we consider the first part with one division. Then, we have,

$$a = \frac{1}{2}a^* e^{g_{PC}(24h)} \Longrightarrow a^* = 2ae^{-g_{PC}(24h)} = f_{2,4}(a).$$

Therefore, it holds that

$$\begin{aligned} G_4^i(a^*) &= \frac{24h}{T_c} d_{PC}^i(a^*) p_1 [1 - H(n, a^*, T_{PC})][H(n, T_c, 24h)], \\ F_4^{i+1}(a) &= G_4^i(a^*) f_{1,4}(a) = G_4^i(f_{2,4}(a)) f_{1,4}(a), \\ f_{1,4}(a) &= 2\frac{df_{2,4}(a)}{da} = 4e^{-g_{PC}(24h)}. \end{aligned} \tag{4.18}$$

Note that in this flow, the factor $n_k = 2$ because there is one division.

Flow F_5 from PC to PC with no division for $T_c > 24h$

Here we consider the case $T_c > 24h$, where a fraction $\frac{T_c-24h}{T_c}$ of those cells below T_{PC} does not divide during a given day (Figure 4.6B F_5 and 4.17B). Then we have

$$a = a^* e^{g_{PC}(24h)} \Longrightarrow a^* = ae^{-g_{PC}(24h)} = f_{2,5}(a).$$

Therefore, it holds that

$$\begin{aligned} G_5^i(a^*) &= \frac{T_c - 24h}{T_c} d_{PC}^i(a^*) p_1 [1 - H(n, a^*, T_{PC})][H(n, T_c, 24h)], \\ F_5^{i+1}(a) &= G_5^i(a^*) f_{1,5}(a) = G_5^i(f_{2,5}(a)) f_{1,5}(a), \\ f_{1,5}(a) &= \frac{df_{2,5}(a)}{da} = e^{-g_{PC}(24h)}. \end{aligned} \tag{4.19}$$

Note that in this flow, the factor $n_k = 1$ because there is no division.

Flow F_6 from PC to GC for $18h \leq T_c \leq 24h$

If $18h \leq T_c \leq 24h$, then a fraction p_2 of pavement cells with area less than T_{PC} and $2M_{GC}$ divides once during one day to produce two new guard cells (Figure 4.6B F_6). By our assumptions

$$\begin{aligned} \frac{ds}{dt} &= g_{GC}s(1-\frac{s}{M_{GC}}) \Longrightarrow \\ s(t) &= \frac{cM_{GC}}{c+e^{-g_{GC}t}M_{GC}-ce^{-g_{GC}t}}, \\ \text{where } c &= s(0). \end{aligned} \tag{4.20}$$

On the other hand, the area of a pavement cell in this part can be at most $2.M_{GC}$ just before dividing. Therefore we propose the formula for the growth rate

$$\begin{aligned} \frac{ds}{dt} &= g_{PC}s(1-\frac{s}{2M_{GC}}) \Longrightarrow \\ s(t) &= \frac{2cM_{GC}}{c+2e^{-g_{PC}t}M_{GC}-ce^{-g_{PC}t}}, \\ \text{where } c &= s(0). \end{aligned} \tag{4.21}$$

Let a^* be the area of the pavement cell at time 0 of day i. It will divide somewhere between time points 0 and T_c to produce new guard cells. We assume that the pavement cell approximately divides after $\frac{T_c}{2}$ hours. Let ${a_1}^*$ be the area of the new-born guard cell (Figure 4.17C), therefore using (4.21) we have

$$\begin{aligned} a_1^* &= \frac{2a^*M_{GC}}{a^*+2e^{-g_{PC}(\frac{T_c}{2})}M_{GC}-e^{-g_{PC}(\frac{T_c}{2})}a^*} \\ &= \frac{2M_{GC}}{1+2e^{-g_{PC}(\frac{T_c}{2})}\frac{M_{GC}}{a^*}-e^{-g_{PC}(\frac{T_c}{2})}}. \end{aligned} \tag{4.22}$$

Moreover, using (4.20) we have

$$\begin{aligned} a &= \frac{a_1^* M_{GC}}{a_1^* + e^{-g_{GC}(24h-\frac{T_c}{2})}M_{GC} - e^{-g_{GC}(24h-\frac{T_c}{2})}a_1^*} \\ &= \frac{M_{GC}}{1 + e^{-g_{GC}(24h-\frac{T_c}{2})}\frac{M_{GC}}{a_1^*} - e^{-g_{GC}(24h-\frac{T_c}{2})}}. \end{aligned} \tag{4.23}$$

Combining (4.22) and (4.23) we obtain,

$$a^* = 2\frac{a e^{0.5 g_{GC}(-48h+T_c)} e^{-0.5 g_{PC} T_c} M_{GC}}{M_{GC} - a + a e^{0.5 g_{GC}(-48h+T_c)} e^{-0.5 g_{PC} T_c}} = f_{2,6}(a). \tag{4.24}$$

To summarize, we have

$$\begin{aligned} G_6^i(a^*) &= d_{PC}^i(a^*) p_2 [1 - H(n, a^*, T_{PC})] \\ &\quad [1 - H(n, a^*, 2M_{GC})][1 - H(n, T_c, 24h)]H(n, T_c, 18h), \end{aligned} \tag{4.25}$$

$$F_6^{i+1}(a) = G_6^i(a^*) f_{1,6}(a) = G_6^i(f_{2,6}(a)) f_{1,6}(a), \tag{4.26}$$

$$\begin{aligned} f_{1,6}(a) &= 2\frac{df_{2,6}(a)}{da} \\ &= 4\,\frac{e^{0.5 g_{GC}(-48h+T_c)} e^{-0.5 g_{PC} T_c} {M_{GC}}^2}{\left(M_{GC} - a + a e^{0.5 g_{GC}(-48h+T_c)} e^{-0.5 g_{PC} T_c}\right)^2}. \end{aligned} \tag{4.27}$$

Note that in this flow, the factor $n_k = 2$ because there is only one division.

Flow F_7 from PC to GC with one division for $T_c > 24h$

If $T_c > 24h$, then a pavement cell with area less than T_{PC} and $2M_{GC}$ with probability p_2 is in the stage where it can divide into guard cells (Figure 4.6B F_7). Similar to flows F_4 and F_5 a fraction $\frac{24h}{T_c}$ of these cells will divide once and a fraction $\frac{T_c - 24h}{T_c}$ will have no division during a given day. Here we consider the first part with one division, so the cell can divide

somewhere in the time interval $[0, 24h]$. Similar to flow F_6 we assume it divides at time $12h$. Then, we have

$$a^* = 2\frac{ae^{-12g_{GC}}e^{-12g_{PC}}M_{GC}}{M_{GC} - a + ae^{-12g_{GC}}e^{-12g_{PC}}} = f_{2,7}(a), \tag{4.28}$$

Therefore, it holds

$$\begin{aligned} G_7^i(a^*) &= \frac{24h}{T_c}d_{PC}^i(a^*)p_2[1 - H(n, a^*, T_{PC})] \\ &\quad [1 - H(n, a^*, 2M_{GC})]H(n, T_c, 24h), \\ F_7^{i+1}(a) &= G_7^i(a^*)f_{1,7}(a) = G_7^i(f_{2,7}(a))f_{1,7}(a), \\ f_{1,7}(a) &= 2\frac{df_{2,7}(a)}{da} = 4\frac{e^{-12hg_{GC}}e^{-12hg_{PC}}M^2{}_{GC}}{(M_{GC} - a + ae^{-12hg_{GC}}e^{-12hg_{PC}})^2}. \end{aligned} \tag{4.29}$$

Note that in this flow, the factor $n_k = 2$ because there is only one division.

Flow F_8 from PC to GC with no division for $T_c > 24h$

In this case, a fraction $\frac{T_c - 24h}{T_c}$ of the corresponding pavement cells will have no division during a given day, so they just grow (Figure 4.6B F_8). Then, using (4.21) we have,

$$\begin{aligned} a &= \frac{2a^*M_{GC}}{a^* + 2e^{-g_{PC}(24h)}M_{GC} - e^{-g_{PC}(24h)}a^*}, \\ a^* &= 2\,\frac{ae^{-24h\,g_{PC}}M_{GC}}{2\,M_{GC} - a + ae^{-24h\,g_{PC}}} = f_{2,8}(a). \end{aligned} \tag{4.30}$$

Therefore, it holds that

$$\begin{aligned} G_8^i(a^*) &= \frac{T_c - 24h}{T_c}d_{PC}^i(a^*)p_2[1 - H(n, a^*, T_{PC})] \\ &\quad [1 - H(n, a^*, 2M_{GC})]H(n, T_c, 24h), \end{aligned}$$

$$F_8^{i+1}(a) = G_8^i(a^*) f_{1,8}(a) = G_8^i(f_{2,8}(a)) f_{1,8}(a), \quad (4.31)$$

$$f_{1,8}(a) = \frac{df_{2,8}(a)}{da} = 4\,\frac{e^{-24h\,g_{PC}}\,{M_{GC}}^2}{\left(2\,M_{GC} - a + ae^{-24h\,g_{PC}}\right)^2}. \quad (4.32)$$

Note that in this flow, the factor $n_k = 1$ because there is no division.

Flow F_9 from GC to GC with no division

Since guard cells do not divide, there is no division in this case. So the cells just grow up to the threshold (Figure 4.6B F_9). Using (4.20), we have

$$a = \frac{a^* M_{GC}}{a^* + e^{-g_{GC}(24h)} M_{GC} - e^{-g_{GC}(24h)} a^*},$$

$$a^* = \frac{a e^{-g_{GC}(24h)} M_{GC}}{M_{GC} - a + a e^{(-24h) g_{GC}}} = f_{2,9}(a). \quad (4.33)$$

Therefore, it holds that

$$G_9^i(a^*) = d_{GC}^i(a^*),$$

$$F_9^{i+1}(a) = G_9^i(a^*) f_{1,9}(a) = G_9^i(f_{2,9}(a)) f_{1,9}(a), \quad (4.34)$$

$$f_{1,9}(a) = \frac{df_{2,9}(a)}{da} = \frac{e^{-(24h)\,g_{GC}}\,{M_{GC}}^2}{\left(M_{GC} - a + ae^{-24h\,g_{GC}}\right)^2}.$$

Note that in this flow, the factor $n_k = 1$ because there is no division.

4.5.8 Predicted distributions

Using the functions $F_1, F_2, ...F_9$ we compute the distribution of cells with area a at day $i+1$ as follows,

$$D_{PC}^{i+1}(a) = \sum_{k=1}^{5} F_k^{i+1}(a) + F_8^{i+1}(a), \quad (4.35)$$

$$D_{GC}^{i+1}(a) = \sum_{k=6}^{7} F_k^{i+1}(a) + F_9^{i+1}(a). \tag{4.36}$$

4.5.9 The optimization strategy

Using (4.35) and (4.36), we predicted distributions for day $i+1$ by using the smoothed data for the distributions on a given day i and the parameters on day i. If d_{PC}^i and d_{GC}^i are the smoothed experimental distributions at day i then we define

$$D_{PC}^{i+1} = MapPC(d_{PC}^i, d_{GC}^i, parameters), \tag{4.37}$$

where $MapPC$ is obtained from (4.35).The same goes for guard cells, so we have,

$$D_{GC}^{i+1} = MapGC(d_{PC}^i, d_{GC}^i, parameters), \tag{4.38}$$

where $MapGC$ is obtained from (4.36).
Then we use a least squares method to optimize the parameter values. The function which we minimize is the objective function

$$\begin{aligned} F_{min} &= \sum_k \left[D_{PC}^{i+1}(a_k) - d_{PC}^{i+1}(a_k)\right]^2 \\ &+ \sum_k \left[D_{GC}^{i+1}(b_k) - d_{GC}^{i+1}(b_k)\right]^2. \end{aligned} \tag{4.39}$$

The a_k and b_k are selected from the discrete set of areas mentioned in §4.5.4. We select those a_k from the list of areas with a sufficiently high frequency. The areas with very small frequency are not considered since cells with such areas are rare outliers.

To optimize the parameters, we minimize the objective function (4.39) by the downhill simplex method or Nelder-Mead method, cf. §3.7.1. To stabilize the result, we considered the transition from day 18 to day 25 as a whole to find the average values for growth rates during these days.

We used those growth rates as the initial guesses for g_{PC} and g_{GC} in the optimization program for the other days. In each step of the optimization and evaluation of parameters, these guesses are modified to be more compatible with the computed values at adjacent days. To obtain reliable results, we used the optimization program for 100 different initial values to find sets of parameters for which F_{min} has a local minimum (up to convergence thresholds) ; then we selected the one with minimal F_{min} among the biologically acceptable ones. To generate the initial values for day i, we randomly selected several sets of parameters from intervals centred at the optimal value for day $i-1$ and wide enough to cover all reasonable situations. Figures 4.8 and 4.9 show the comparison between experimental and model-predicted cell size distributions.

The optimization algorithm was applied for maximum 6000 iterations and function evaluations up to $1e-9$ tolerance on the function and parameter values. These experiments were performed in the "twigrid" cluster of the Department of Applied Mathematics and Computer Science (Ghent University), under Linux 2.6.18-6-xen-amd64 using Matlab R2008a. Note that, although in the downhill simplex method convergence is typically rather slow, in the same conditions we achieved more accurate results than in three other methods, Newton's method, quasi-Newton's method, Levenberg-Marquardt algorithm, cf. §3.6. We selected the above four methods to run on our program for day 18. First, the quasi-Newton's method and Nelder-Mead method were taken to optimize three parameters. The results are reported in Figure 4.18 and Table 4.4.

Although the quasi-Newton's method quickly reaches the local solution, the result for one of the parameters (M_{GC}) is more accurate in the Nelder-Mead method. Therefore for safety we keep Nelder-Mead method as a main optimization method for the programs with more than three parameters. Another experiment for day 18 has been done with only one parameter (Table 4.5). Again Table 4.5 shows that the two methods, quasi-Newton and Nelder-Mead, are time efficient among the aforementioned

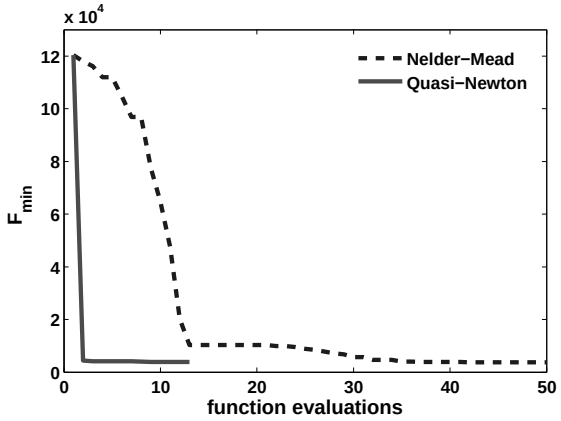


Figure 4.18: Results of Nelder-Mead and quasi-Newton's methods for the function F_{min} with three parameters at day 18.

Table 4.4: Results of Nelder-Mead and quasi-Newton's methods for the function F_{min} at day 18. time–eval is the evaluation time of the objective function and func–Count reports the total number of objective function evaluations.

Algorithms	**Nelder-Mead**	**Quasi-Newton**
Initial values		
g_{PC}	0.030	0.030
g_{GC}	0.003	0.003
M_{GC}	300	300
F_{min}	120365	120365
Optimum values		
g_{PC}	0.003	0.003
g_{GC}	0.016	0.018
M_{GC}	357	300
F_{min}	2404	3934
Output arguments		
time-eval (sec)	4892	56
iteration	300	12
func-count	561	104

four methods. The evaluation of the objective function also shows that they are sufficiently precise. Afterwards, we tested the above two efficient methods (quasi-Newton and Nelder-Mead) for the program with two parameters (Table 4.6). Again, the Nelder-Mead method is more accurate.

Table 4.5: Results of four methods for the function F_{min} with one parameter at day 18. time–eval is the evaluation time of the objective function and func–count reports the total number of objective function evaluations. Maximum number of iterations and function evaluations is 3000. The exitflag is an integer code for the reason the solver halted its iterations: 1 means the function converged to a solution x; 2 means change in x was less than the specified tolerance; 4 means magnitude of search direction was smaller than the specified tolerance Tolx; 5 means predicted decrease in the objective function was less than the TolFun tolerance. Tolerance of function (TolFun) and variables (Tolx), both are 1.00E-12.

Algorithms	**Nelder-Mead**	**Quasi-Newton**	**Levenberg-Marquardt**	**Newton**
Initial values				
g_{PC}	0.03	0.03	0.03	0.03
F_{min}	238	238	238	238
Optimum values				
g_{PC}	0.0035	0.0035	0.0035	0.0035
F_{min}	20	20	238	238
Output arguments				
time-eval (sec)	20	17	38	35
iteration	44	9	58	69
func-count	88	78	153	140
exitflag	1	5	4	2

The relevant optimization methods all converge to local minima which may be identical or not. To reduce the objective function value as far as possible, in all optimization process we use the Nelder-Mead method to increase the chance to achieve the global minimum. We must be very aware that different initial values and tolerances also lead to different results. In

general we choose the most relevant parameters with the smallest objective function.

Table 4.6: Results of Nelder-Mead and quasi-Newton's methods for the function F_{min} for two parameters at day 18. time–eval, func-count and exitflag were described in Table 4.5.

Algorithms	**Nelder-Mead**	**Quasi-Newton**
Initial values		
g_{GC}	0.003	0.003
M_{GC}	300	300
F_{min}	120127	120127
Optimum values		
g_{GC}	0.0159	0.0177
M_{GC}	357	300
F_{min}	2384	3914
Output arguments		
time-eval (sec)	93	23
iteration	150	6
func-count	305	72
exitflag	1	5

Chapter 5

Modelling the control of the APC/C^{CCS52A2}

This chapter is a part of the studies which have been conducted by T. Lammens, V. Boudolf, L. Kheibarshekan, L. P. Zalmas, T. Gaamouche, S. Maes, M. Vanstraelen, E. Kondorosi, N. B. La Thangue, W. Govaerts, D. Inzé and L. De Veylder. Here, I only highlight my contribution which deals with the modelling and simulation issues.

5.1 Introduction

During the mitotic cell cycle, DNA which is duplicated during the S phase, is divided at the M phase, so that each daughter cell produced has a genomic DNA content equal to that of its parents. In contrast, during the endoreduplication cycle, no cytokinesis occurs between rounds of DNA replication, resulting in successive doubling of the DNA ploidy level. This process occurs in a wide variety of cell types in arthropods and mammals and is particularly prominent in dicotyledonous plants[1] [25], especially in species with a small genome and a short life cycle, in which repetitive DNA replication might support growth under conditions that require rapid

[1] a group of flowering plants whose seed typically has two embryonic leaves.

development [2, 58]. The endocycle is activated in various developmental processes, including placental formation, *Drosophila* oogenesis[2], and leaf development. The timing of endocycle onset is crucial for correct development, because polyploidization is linked with cessation of cell division and initiation of terminal differentiation, cf.§1.1. Mitotic cell cycle progression and endoreduplication are linked events. Premature or delayed exit from the cell division program results in an increased or decreased DNA ploidy, respectively [11, 14, 22, 75, 85, 87, 89]. Therefore, the onset of endoreduplication must be controlled precisely. At the molecular level, endoreduplication is likely achieved through elimination of the components needed to progress through mitosis [52]. Predominant roles in this process are played by the anaphase-promoting complex/cyclosome (APC/C) activator genes, such as *CDH1*, *FZR*, and *CCS52A*, which have been found to promote endocycle onset and progression in human, *Drosophila melanogaster*[3], and *Medicago truncatula*[4] cells, respectively [9, 13, 49, 59, 74, 76]. However, the mechanisms controlling the transcriptional activity of these genes remain unclear.

Over the years, it has become clear that the E2F transcriptional network acts as a key regulator in the balanced expression of many essential genes involved in proliferation and differentiation [23]. Recently, a class of novel atypical E2F proteins was identified in *Arabidopsis thaliana* (E2Fd/DEL2, E2Fe/DEL1 and E2Ff/DEL3) and mammals (E2F7 and E2F8) that operate as transcriptional repressors [23, 37]. Similar to the typical E2F proteins, the atypical E2F proteins bind the consensus E2F recognition sequence, but they have two DNA-binding domains and do not require a DP partner to bind DNA. In contrast to the classical E2F proteins, the physiological relevance of the novel E2Fs is less clear. E2F7 and E2F8 have been demonstrated to play a role in controlling E2F1-dependent apoptosis [51, 90],

[2]Oogenesis is the creation of an egg cell.
[3]known as the common fruit fly.
[4]is a small plant native to the Mediterranean region that is used in genomic research.

whereas in plants, atypical E2Fs operate as inhibitors of postmitotic events; mutants of E2Fe/DEL1 display increased endoreduplication levels [86], whereas E2Ff/DEL3-deficient plants are prone to rapid cell expansion [68]. K. Vlieghe and others[86] report that the enhanced endoreduplication levels observed in E2Fe/DEL1 knock-out plants arise from a premature onset of the endocycle. Through microarray analysis and chromatin immunoprecipitation[5] (ChIP), T. Lammens and others [46] identified the APC/C activator gene *CCS52A2* as a direct E2Fe/DEL1 target. E2Fe/DEL1 misregulation results in untimely *CCS52A2* transcription, affecting the timing of endocycle onset. Correspondingly, ectopic *CCS52A2* expression drives cells into the endocycle prematurely. Dynamic simulation in this chapter, illustrate that E2Fe/DEL1 accounts for the onset of the endocycle by regulating the temporal expression of *CCS52A2* during the cell cycle in a development-dependent manner. Remarkably, an association of E2F7 to the *CDH1* promoter in mammalian cells was observed, suggesting that the transcriptional control of the APC/C activator genes through atypical E2Fs might be conserved across species.

5.2 Modelling

An Ordinary Differential Equation (ODE) model for *CCS52A2* gene expression was built, taking into account that E2Fe/DEL1 inhibits *CCS52A2* expression. We applied mathematical optimization approach to fit the parameters which were evaluated using the relative 2-norm error and computed the correlation coefficient between approximate and experimental data. In the first step we built a model for *CCS52A2* and E2Fe/DEL1 expressions in the synchronized cell culture.

[5]This process can be used to isolate and concentrate a particular protein from a sample containing many thousands of different proteins.

5.2.1 E2Fe/DEL1 and *CCS52A2* in a synchronized cell culture

Table 5.1 presents the time evolution of levels of E2Fe/DEL1 and *CCS52A2* during 18 hours in a synchronized cell culture. Cells are synchronized at the G1/S border, but lose synchrony at the M/G1 transition. Cell cycle is about 24h and the S phase takes about 5h (time points 0-6). Mitosis occurs between 12-18h. So G2 can be taken from 6-12, and G1 18-24. Figure 5.1B and Table 5.1 were obtained from the experimental data. Each column in Table 5.1 is scaled independently in such a way that the concentration at time 0 is 1.

Table 5.1: E2Fe/DEL1 and *CCS52A2* transcript levels during the cell cycle as quantified by qRT-PCR. Values are scaled in such a way that the concentrations at time 0 are 1.

time(h)	E2Fe/DEL1	*CCS52A2*
0	1	1
2	1.125	0.6859
4	1.1935	0.5545
6	1.2364	0.4509
8	1.3914	0.2191
10	2.1031	0.3173
12	1.6431	0.1975
14	1.4724	0.2389
16	1.1644	0.4045
18	1.1284	0.5171

Approximation of E2Fe/DEL1 levels

A non-linear fitting through the least squares method was undertaken to find a curve for E2Fe/DEL1 expression in a synchronized cell culture based

on experimental data (Figure 5.1B). The non-linear function,

$$DEL1(t) = c_4 + \frac{c_1}{1 + c_2(t - c_3)^2}, \tag{5.1}$$

was proposed for this to mimic the essential features of the experimental data. This model was fitted to the data by minimizing the following square error

$$\sum_{i=0}^{9} (DEL1(2i) - DEL1_{2i})^2, \tag{5.2}$$

where $DEL1(2i)$ is given by (5.1) and $DEL1_{2i}$ is the level of E2Fe/DEL1 at time $2i$, scaled so that $DEL1_0 = 1$. Using the *minimize* command in Maple, which implies a quasi-Newton method, cf. §3.6.4, for optimization problems, we found $c_1 = 1.07194, \mathrm{c}_2 = 0.2964, \mathrm{c}_3 = 10.4737, \mathrm{c}_4 = 1.0727$. The associated relative 2-norm error is 0.051 and the correlation coefficient 0.97 (Figure 5.1C).

A dynamic model for *CCS52A2* transcription

Knowing that E2Fe/DEL1 inhibits the expression of *CCS52A2*, an ODE model was built for *CCS52A2* gene expression in a synchronized cell culture. We propose the following formula:

$$\frac{dCCS52A2}{dt} = v\frac{k^n}{k^n + [DEL1(t)]^n} - gCCS52A2, \tag{5.3}$$

in which v is the maximal production rate of *CCS52A2*, g the degradation rate of *CCS52A2* and k a threshold (whose dimension is that of a concentration). Exponent n is a determinant of the strength of E2Fe/DEL1 binding to the promoter of *CCS52A2*, which biologically could be represented as the number of available binding sites, cf. §3.1 and §3.4. The transcriptional activity of the *CCS52A2* promoter, as a function of the E2Fe/DEL1 availability in time, can be expressed as

$$\frac{k^n}{k^n + [DEL1(t)]^n}. \tag{5.4}$$

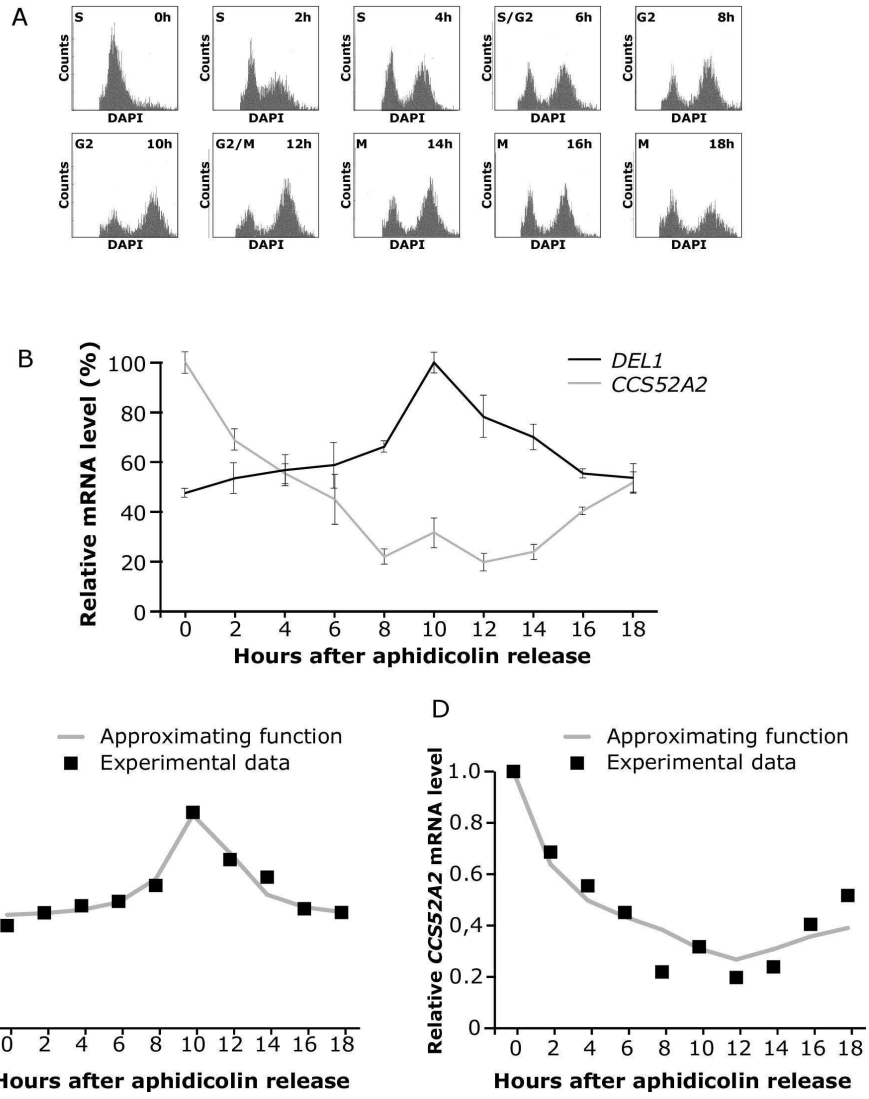


Figure 5.1: Kinetics of E2Fe/DEL1 and *CCS52A2* mRNA during the cell cycle. (A) Flow cytometric analysis of *Arabidopsis* suspension cells after release of an aphidicolin block in late G1/early S, showing the coherent cell population progressing synchronously through the cell cycle. (B) E2Fe/DEL1 and *CCS52A2* transcript levels during the cell cycle as quantified by qRT-PCR. Values are mean $\pm SD$ obtained from two independent synchronization experiments. (C) Fitting of a non-linear model function for E2Fe/DEL1 expression in cell cultures, approximating the experimental data. (D) Curve fitting of *CCS52A2* expression during the cell cycle, allowing determination of parameters in the model describing the dynamic relation between E2Fe/DEL1 and *CCS52A2*.

Multiplying (5.4) with the maximum production rate v, gives the E2Fe/DEL1-dependent term in the *CCS52A2* expression. By analytically solving (5.3), we obtain

$$CCS52A2(t) = e^{-gt}\left(C + v\int_0^t \frac{k^n}{k^n + [DEL1(\theta)]^n} e^{g\theta} d\theta\right), \qquad (5.5)$$

where $C = CCS52A2(0)$. Parameters were estimated with the experimentally obtained data for *CCS52A2* expression in a synchronized cell culture (Figure 5.1B). A least-squares fit, cf. §3.6.5, computing local minima of the objective function, was used for parameter optimization. As a first step, we assumed that $C = \textit{CCS52A2}(0) = 1$. In this case, the objective function became

$$\begin{aligned} G \;=\; \sum_{i=0}^{i=9} \Big[CCS52A2_{2i} - e^{-2ig}\,(C+ \\ v\int_0^{2i} \frac{k^n}{k^n + [DEL1(\theta)]^n} e^{g\theta} d\theta\Big)\Big]^2, \end{aligned} \qquad (5.6)$$

in which $CCS52A2_{2i}$ represents the experimental data for *CCS52A2* expression at time $(h) = 2i$. Assuming $n = 1$, the number of parameters to be estimated is reduced to 3. Solving the below nonlinear system of equations (5.7), parameters could be estimated making use of Newton iterations, cf. §3.6.3.

$$\frac{\partial G}{\partial v} = 0, \frac{\partial G}{\partial k} = 0, \frac{\partial G}{\partial g} = 0. \qquad (5.7)$$

Optimal parameters were $k = 0.0138$, $v = 17.0210$ and $g = 0.4940$. The associated relative 2-norm error was 0.15 and the correlation coefficient 0.95 (Figure 5.1D). We note that the minimization problem turned out to be ill-conditioned so that many near-minima could be found. We chose the one with the highest v, which is consistent with other information on the production rate of *CCS52A2* (Table 5.2.1).

Table 5.2: The result of optimization for several initial values of the parameters in ODE model.

Results	k	v	g	n	Relative 2-norm error	Correlation
1	0.0668	4.0003	0.5237	1	0.1534	0.9457
2	0.0447	5.2415	0.4885	1	0.1529	0.9443
3	0.0370	6.3654	0.4901	1	0.1524	0.9448
4	0.01420	16.5748	0.4939	1	0.1514	0.9457
5	0.0138	17.0210	0.4940	1	0.1514	0.9457

Scaling of *CCS52A2* expression levels

The absolute levels of E2Fe/DEL1 and *CCS52A2* in a synchronized cell culture at $time = 0(h)$ were not equal (as indicated by their difference in $\triangle C_t$ values, being $C_{t,GENE} - C_{t,ACTIN2}$, obtained by Real-Time qRT-PCR). To take this value into account in equation (5.3), let c_i and d_i be the time-normalized Real-Time qRT-PCR values of *CCS52A2* and E2Fe/DEL1 expression, obtained by calculating $2^{-(C_{t,GENE}-C_{t,ACTIN2})}$ where $i = time(h)$. Let $c_{real,i}$ be the computed data for *CCS52A2* expression in cell culture at $time = i(h)$. In this view, we can state,

$$\frac{c_i}{c_{real,i}} = C^{te} = \frac{c_0}{c_{real,0}} \Longrightarrow c_{real,i} = c_i \frac{c_{real,0}}{c_0}. \tag{5.8}$$

Experimentally, we found $\frac{d_{real,0}}{c_{real,0}} = 18.0673$. If we take $d_{real,0} = d_0 = 1$, then $c_{real,0} = 0.05534$ and (5.8) implies $c_{real,i} = 0.05534 c_i$. So, the speed at which *CCS52A2* mRNA is formed, which is a function of E2Fe/DEL1, needs to be corrected by a factor 0.05534. That establishes v in formula (5.3) as 0.94208, where all concentrations are measured with the concentration of E2Fe/DEL1 at $time = 0(h)$ as unit.

5.2.2 E2Fe/DEL1 and *CCS52A2* during leaf development

Subsequently, we built a model for the expression of *CCS52A2* and E2Fe/DEL1 during leaf development. Because cell division cannot be synchronized experimentally in a developing leaf and endoreduplication cannot be triggered in *Arabidopsis* cell cultures, we combined leaf and cell culture expression data mathematically.

E2Fe/DEL1 expression

Figure 5.2A shows the experimentally observed profile of E2Fe/DEL1 expression during leaf development. In a developing leaf, no synchrony can be obtained; consequently, the corresponding expression values at each day could be seen as an average over the whole cell cycle that took place during that day. In the modelling, we consider days 8 to 10 only, because beyond day 10 more and more cells undergo endoreduplication, possibly affecting the cell cycle-dependent expression profile of the E2Fe/DEL1 and *CCS52A2* genes. Leaves at day 8 can be assumed to be only consisting of cells going through cell division. This assumption, together with the previous statement, implies that the cell cycle expression profile of E2Fe/DEL1 at days 8, 9 and 10 can be seen as a scaling of the profile obtained by synchronisation. This makes

$$DEL1_{day8}(t) = \lambda \left(c_4 + \frac{c_1}{1 + c_2(t - c_3)^2} \right), \tag{5.9}$$

where λ is a factor to be estimated by using further experimental data (see estimating λ). Similarly, we find the approximate functions for E2Fe/DEL1 in each day,

$$DEL1_{dayi}(t) = \alpha_i DEL1_{day8}(\mathrm{t}), \tag{5.10}$$

where α_i is the normalized expression value at day i as plotted in Figure 5.2A.

Estimating λ

We let d_i be the normalized E2Fe/DEL1 expression value in the cell culture at time $i = 0, 2, 4...18$ (Table 5.1) and ld_j the result of $2^{-(C_{t,GENE}-C_{t,ACTIN2})}$ at day $j = 8, 9, 10$ during leaf development (Table 5.2.2). Assuming s as

Table 5.3: Scaled level of E2Fe/DEL1 during the leaf development.

days	DEL1
8d	1
9d	0.3747
10d	0.2608
11d	0.1859
12d	0.1279
13d	0.0610
14d	0.0520
15d	0.0898
16d	0.1284
17d	0.0709
18d	0.1289
19d	0.0545
20d	0.0233
21d	0.0293

the experimentally measured proportion of E2Fe/DEL1 expression at time $i = 0$ in the cell culture to the expression value of E2Fe/DEL1 during leaf development at day 8, $ld_8 = \frac{d_0}{s}$. According to (5.9), $ld_8 = \lambda.\overline{d}$, where $\overline{d}$ is the averaged normalized expression value of E2Fe/DEL1 in cell culture. So $\lambda = \frac{d_0}{\overline{d}s} = 0.2391$ (Table 5.2.2).

Table 5.4: Scalar factor of E2Fe/DEL1 in the leaf development at day 8.

s	$\bar{\mathbf{d}}$	λ
3.1080	1.3458	0.2391

***CCS52A2* expression**

Applying the model of *CCS52A2* expression in a synchronized cell culture, the behaviour of *CCS52A2* in a mitotic cell can be written as a function of E2Fe/DEL1 expression during leaf development, meaning that *CCS52A2* expression during leaf development can be described through combination of functions (5.5) and (5.10),

$$\begin{aligned} &CCS52A2_{dayi}(t) = e^{-gt}\left(C + v\int_0^t \frac{k^n}{k^n+[DEL1_{dayi}(\theta)]^n} e^{g\theta} d\theta\right), \\ &C = CCS52A2_{dayi}(0) = \mu, \quad i = 8, \ldots, 15 \end{aligned} \tag{5.11}$$

in which we assume μ_8 as the average value of *CCS52A2* in the cell culture i.e. 0.025381. μ_9 becomes $1.781\mu_8$ (proportion of *CCS52A2* expression at 9 days vs. 8 days) and μ_{10} becomes $2.434\mu_8$ (proportion of *CCS52A2* expression at 10 days vs. 8 days).

This mathematical modelling permitted an in silico visualization of the cell cycle phasedependent relationship between E2Fe/DEL1 and *CCS52A2* in a developmental context. The simulation showed that decreasing E2Fe/DEL1 levels during leaf maturation triggered a preferential increase in *CCS52A2* transcripts during the late S-G2 and M phases (Figure 5.2C). These data suggest that E2Fe/DEL1 controls the cell cycle phasedependent *CCS52A2* transcription profile in a developmentally dependent manner.

5.3 Results

Our data suggest that E2Fe/DEL1 levels determine the timing of cell cycle exit and onset of endoreduplication by controlling the transcription of *CCS52A2*. To understand mechanistically how decreasing E2Fe/DEL1 levels can account for the division-to-endoreduplication transition, we mathematically modelled the cell cycle phasedependent expression pattern of

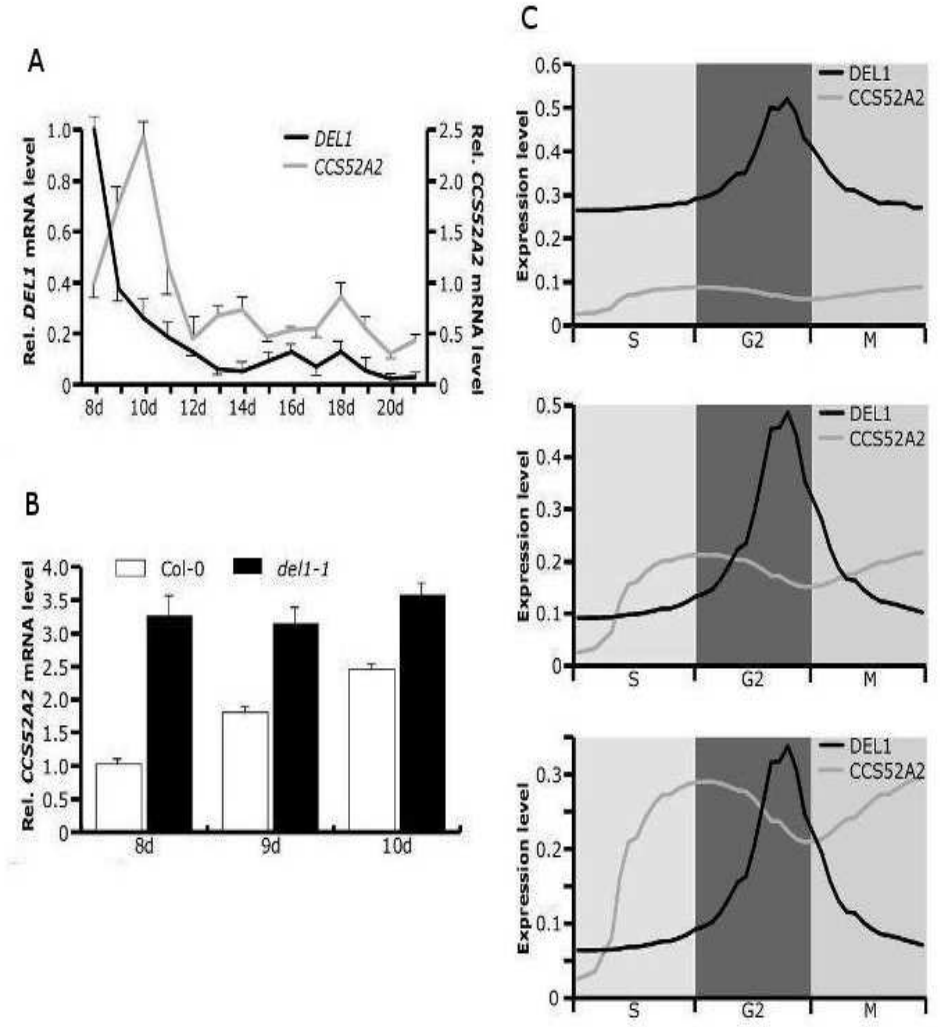


Figure 5.2: Control of development-dependent expression of *CCS52A2* by E2Fe/DEL1. (A) Kinetics of E2Fe/DEL1 and *CCS52A2* transcription during leaf development. Transcript levels were measured by real-time PCR. All values were normalized to the ACT2 housekeeping gene. The $\triangle Ct$ method was used for relative quantification of transcripts. Values are means $\pm SD(n = 3)$. Note that transcription of *CCS52A2* peaked at day 10, marking the endocycle onset. (B) *CCS52A2* mRNA level during leaf development in wild-type (Col-0) and *del1-1* mutants, respectively. Data are mean $\pm SD(n = 3)$. (C) Simulation of *CCS52A2* accumulation during leaf development showing a progressive increase in *CCS52A2* transcript levels during the S and G2 phases (days 8, 9 and 10) in response to decreasing levels of E2Fe/DEL1.

CCS52A2 during leaf development. E2Fe/DEL1 mRNA levels were abundant mainly in the early stages of leaf development. To do so, we measured E2Fe/DEL1 and *CCS52A2* transcription levels during leaf development and during cell cycle progression of a synchronized cell cultture.

In contrast, *CCS52A2* transcripts accumulated and reached a maximum at day 10, the time of cell cycle exit and onset of endoreduplication (Figure 5.2A). Because of its low expression level before day 10, *CCS52A2* might be repressed by E2Fe/DEL1 during the dividing phase of leaf development. This hypothesis was confirmed in *del1-1* plants, in which *CCS52A2* expression levels were clearly higher during the early leaf growth stages than those of control plants (Figure 5.2B). In a synchronized cell culture, E2Fe/DEL1 and *CCS52A2* display complementary transcription profiles, with a predominance of *CCS52A2* expression during the G1 and S phases (Figure 5.1B) [30]. The *CCS52A2* expression profile corresponded with the anticipated function of its gene product in preventing premature accumulation of mitotic cyclins in interphase cells but allowing their accumulation during the late S and G2 phases, allowing the M phase to proceed [63]. E2Fe/DEL1 expression levels peaked during G2, similar to what has been observed for its mammalian counterparts E2F7 and E2F8 [20, 54].

5.4 Concluding remarks

We have found that the atypical E2F transcription factor E2Fe/DEL1 controls the onset of the endocycle through a direct transcriptional control of APC/C activity. Because E2Fe/DEL1 represses the *CCS52A2* promoter, we hypothesize that its level must drop below a critical threshold to allow sufficient accumulation of *CCS52A2* during late S and G2 for cells to proceed from division to endoreduplication, a model suggested by the dynamic simulation of the cell cycle phasedependent expression level of *CCS52A2* during leaf development. The steady increase in *CCS52A2* during late S and G2 likely counteracts the mitotic cyclin-dependent kinase

(CDK) activity that builds up during these cell cycle phases [78, 66], eventually blocking the G2 to M transition and thereby triggering endoreduplication.

However, this might be an over-simplification since, no clear endoreduplication phenotype was observed in 8-day old E2Fe/DEL1 knockout plants. At this earliest developmental stage examined, the leaves were still mitotically active, corresponding to high cyclin transcription rates [5]. We propose that at this stage, the cyclin production rates are so high that the cyclin abundance is insensitive to the counteracting action of the E2Fe/DEL1 controlled APC/C^{CCS52A2} activity. In contrast, when the leaf matures, the cyclin production rates decrease, and the effects of increasing *CCS52A2* levels may become apparent. The combination of decreased cyclin production rates and increased control at the protein stability level may ensure a unidirectional onset of the endoreduplication program.

Chapter 6

Conclusions and future work

6.1 Conclusions

In this thesis we developed two mathematical models and innovative simulation techniques for systems biology relevant to plant growth and development. Here we discuss the progress represented by these models and the conclusions that we could draw from this work.

6.1.1 Conclusions related to the epidermal cell division model

The lower leaf epidermis of the plant *Arabidopsis thaliana* consists of two cell types, stomatal guard cells and pavement cells. Stomata are small pores on the surface of leaves whose aperture is controlled by two guard cells. In this study we concentrated on the development of epidermal leaf cells in *Arabidopsis* to investigate the role of cell size and cell age in cell division and cell differentiation. We also studied the evolution of the cell cycle duration during development. We considered the initiation and regulation of precursor cells that form guard cells and pavement cells; this study is restricted to the number and sizes of pavement and guard cells.

We built a computational framework to describe the development of the leaf epidermis. The model includes a number of parameters, among which the cell cycle duration and the growth rates of pavement and guard cells, as well as two thresholds for the area of pavement and guard cells. The model is based on a map from the size distribution of pavement and guard cells on a given day to the distributions on the next day.

From the biological point of view we obtained the two following important results. The first rather unexpected result is that the average cell cycle duration is nearly constant during leaf development. The second unexpected result is that there is no evidence for the existence of a size threshold, hence the cell size in itself does not determine whether the cell is in its division phase. Moreover, we found in the study of the model that small cells grow faster than large cells during leaf development. This realization convinced biologists (Lieven de Veylder, Stijn Dhondt and et al.)[1] to design new experiments. They performed live imaging of epidermal cells in the leaf independently of the model. Confocal imaging allowed them to measure relative growth rates of individual pavement and guard cells over a period of 36 hours. Cell tracking experiments on leaves confirmed that pavement cells smaller than $300\mu m^2$ grow faster than larger ones. Interestingly, they could also observe that small guard cells ($< 100\mu m^2$) grow faster than larger ones.

6.1.2 Conclusions related to E2Fe/DEL1 activity

The experimental studies of the expressions of *CCS52A2* and the level of atypical E2F transcription factor, E2Fe/DEL1 in a synchronized cell culture revealed that the E2Fe/DEL1 inhibits *CCS52A2* expression. The model proposed for the E2Fe/DEL1 activity during leaf development, permitted an in silico visualization of the cell cycle phase dependent relation-

[1] PSB/VIB, Ghent University

ship between E2Fe/DEL1 and *CCS52A2* in a developmental context. The results revealed that decreasing E2Fe/DEL1 levels during leaf maturation triggered a preferential increase in *CCS52A2* transcripts during the late S-G2 and M phases. These data suggest that E2Fe/DEL1 controls the cell cycle phase dependent *CCS52A2* transcription profile in a developmentally dependent manner.

6.2 Future work

In this thesis a number of important new conclusions were reached on the behaviour of cell division during leaf development and on the molecular vision of endocycle onset. However, simultaneously a new number of issues arose, that could be achieved in a follow-up of this thesis.

6.2.1 Future work related to the epidermal cell division model

- **Different growth model**
  One of the conclusions of the study of epidermal cell division is that the growth rate of both cell types, pavement and guard cells, are size dependent. Although we let the growth rate vary in the model during leaf development, meaning that it can depend on some internal factors, it would be more appropriate to replace the exponential growth model with formula that takes into account the size of the cell as an independent variable for the growth rate.

- **Applying the model to the analysis of mutants**
  Since the model is a general model which is valid for similar populations, it can be easily run for different mutants affected in cell division or/and cell expansion to compare with wild type. Moreover,

implementation of the model for different systems can help to check the model results and improve it in a simple way.

- **Applying the model to produce a visual development of the virtual leaf**
  Another notable work that we have intended to do it, is, applying the results of the model in a computer simulation to produce a virtual leaf and observe the development process virtually. This virtual leaf can be used to compare with the actual samples, and to review the results of the model. For this reason, it is necessary to produce the images and movies by using appropriate software such as L-studio[2].

- **Introduce some other cell types and extend the model**
  Stomatal development in *Arabidopsis* invariably requires three different precursor cells, the meristemoid mother cell (MMC), the meristemoid (M), and the guard mother cell (GMC). The pathway of stomata begins with the selection and asymmetric division of the MMC. Also the sister cell to meristemoid may differentiate into pavement cells, large jigsaw puzzle-shaped cells that undergo endoreduplication or become MMCs and start a new stomatal lineage or divide symmetrically and produce two cells that can independently follow one of these three fates.
  To our knowledge, no data are available for the number of meristemoid cells, sister cells and mother meristemoid cells. This thesis is restricted primarily to the number of pavement and guard cells.
  However some of the other cells which exist in the stomatal pathway can be used to extend the model. This can be a challenge for both biologists and modellers to see how they can distinguish the other

[2]L-studio is Windows software for creating simulation models and performing virtual experiments using L-systems (http://Algorithmicbotany.org/).

cells in the stomatal pathway and produce kinematic data for them to use in the extended model.

- **Experimental observation of the cell size threshold**
  As we saw in Chapter 4, the model does not point to the existence of a size threshold for cell division. For this purpose, we open a discussion with the biologists to study the possibility of empirical testing of this conclusion. For instance, one way to experimentally study the size threshold is to see whether very large cells can divide during leaf development. Another way to check this result is to consider cell division in a small part of leaf at nearly the end of development with large cells and build a simple model that contains the size threshold as a free parameter. The simulation of this model and comparison of the result of this simple model with the experimental observations could help to confirm or negate our conclusions.

6.2.2 Future work related to E2Fe/DEL1 activity

We demonstrated that E2Fe/DEL1 regulates the temporal expression of *CCS52A2* but not that of *CCS52A1*, implying that independent signalling pathways control the timing of endocycle onset and/or progression through the endoreduplication program. In *Arabidopsis*, endoreduplication is an integral part of the leaf development. The presence of multiple pathways protect plants from uncontrolled cell proliferation. A goal for future work is to recognize the other components active in the small network of E2Fe/DEL1 and *CCS52A2*. By adding the new factors and extending the network, we can better study the system which is responsible for endocycle onset. For this reason, we can either extend the network and add new unknown factors in the model and predict the behaviour of these unknown factors such as transcription factors or identify experimentally the other transcription

factors involved in the cell cycle controlling system. These both help to make the whole pathway more clear and to understand the whole system of controlling the cell cycle process during leaf development.

Glossary

Abaxial	On the side that is away from the axis or central line, usually on the underside, 47
Activator	An activator is a DNA-binding protein that regulates one or more genes by increasing the rate of transcription, 100
Adaxial	On the side that is towards the axis or central line, usually on the upper side, 47
Auxin	Any of a group of hormones that regulate plant growth, particularly by stimulating cell elongation in stems and inhibiting it in roots. For example, auxins influence the growth of stems toward light (phototropism) and against the force of gravity (geotropism). Auxins also play a role in cell division and differentiation, 47
Cell differentiation	Cell differentiation is a process in which a generic cell develops into a specific type of cell in response to specific triggers from the body or the cell itself, 48

Cell expansion
: Cell expansion is increase in cell volume. The mechanics of cell expansion is central to the regulation of cell shape and division, 49

Digitization
: Digitization is the representation of an object, image, sound, document or a signal (usually an analogue signal) by a discrete set of its points or samples. The result is called digital representation, 75

Enzyme
: Enzymes are proteins that catalyse (i.e., increase or decrease the rates of) chemical reactions. In enzymatic reactions, the molecules at the beginning of the process are called substrates, and the enzyme converts them into different molecules, called the products. Almost all processes in a biological cell need enzymes to occur at significant rates. Since enzymes are selective for their substrates and speed up only a few reactions from among many possibilities, the set of enzymes made in a cell determines which metabolic pathways occur in that cell, 1

Epinasty
: Increased growth of the upper surface of a plant part, such as a leaf, resulting in a downward bending of the part. The opposite of epinasty is, hyponasty, an increase in growth in a lower part of a plant causing it to bend upward, 70

Flow cytometry — Flow cytometry is a technology that simultaneously measures and then analyzes multiple physical characteristics of single particles, usually cells, as they flow in a fluid stream through a beam of light, 103

Gene — A gene is a unit of heredity in a living organism. It normally resides on a stretch of DNA that codes for a type of protein or for an RNA chain that has a function in the organism. All living things depend on genes, as they specify all proteins and functional RNA chains, xi

Least squares — The least squares solution of a system of equations minimizes the sum of the squares of the errors made in solving every single equation. Least squares problems have two categories: linear least squares and nonlinear least squares, depending on whether or not the residuals are linear in all parameters, 34

Mass action law — Mass action kinetics define chemical reaction rates as a product of a rate constant and the concentrations of the reactants. Both forward and reverse rates can be specified, 16

Michaelis-Menten kinetics	An approximation of mass action kinetics typically used for enzymesubstrate interactions when the concentration of the substrate is in excess of the enzyme. Conservation of mass and applying the equilibrium assumption on the intermediate complex reduces the number of equations, 11
Objective function	A function used to measure the goodness of a model. A common example is deviation from an ideal behaviour, which we attempt to minimize for parameter estimation, 32
Organismal theory	In plants, the organismal theory refers to the idea that during plant growth/evolution the original cell expanded into the whole organism. This is opposed to the cell theory that states that a cell multiplies into many cells with the original cell being equivalent to many, 7
Parameter estimation	The regression process by which parameters are estimated by comparing model output to experimental data, 32
Precursor cells	Precursor cells are cells capable of differentiating into one or two closely related final forms., 55
Prediction	An outcome obtained by executing a model and that has not yet been obtained experimentally or used in calibration, 23

Proliferating	Multiplying or increasing in number. In biology, cell proliferation occurs by a process known as cell division, 51
Protein	Proteins (also known as polypeptides) are organic compounds made of amino acids arranged in a linear chain and folded into a globular form. The sequence of amino acids in a protein is defined by the sequence of a gene, which is encoded in the genetic code, xv
Protein Kinase	A protein kinase is a kinase enzyme that modifies other proteins by chemically adding phosphate groups to them (phosphorylation). This class of proteins is further separated into subsets such as PKC alpha, PKC beta, and PKC gamma, each with specific functions. Phosphorylation usually results in a functional change of the target protein (substrate) by changing enzyme activity, cellular location, or association with other proteins. Up to 30 proteins may be modified by kinase activity, and kinases are known to regulate the majority of cellular pathways, especially those involved in signal transduction, the transmission of signals within the cell, 3
qRT-PCR	**q**uantitative **R**eal **T**ime **P**olymerase **C**hain **R**eaction (qRT-PCR) is a technique which is used to amplify and simultaneously quantify a targeted DNA molecule, 101

Repressor — A repressor is a DNA-binding protein that regulates the expression of one or more genes by binding to the operator and blocking the attachment of RNA polymerase to the promoter, thus preventing transcription of the genes. This blocking of expression is called repression, 100

Simulation — The process of using mathematical models to study the responses and properties of a system under differing conditions, usually different parameter values and occasionally different model structures, 11

Index

Bibliography

[1] G. S. Adair, A. V. Bock, and J. H. Field, *The oxygen dissociation curves of hemoglobin*, J. Biol. Chem. **63** (1925), 529–535.

[2] M. Barow and A. Meister, *Endopolyploidy in seed plants is differently correlated to systematics, organ, life strategy and genome size*, Plant Cell Environ **26** (2003), 571–584.

[3] T. I. Baskin, *On the constancy of cell division rate in the root meristem*, Plant Molecular Biology **43** (2000), 545–554.

[4] G. T. Beemster and T. I. Baskin, *Analysis of cell division and elongation underlying the developmental acceleration of root growth in arabidopsis thaliana*, Plant Physiol **116** (1998), no. 4, 1515–26.

[5] G. T. Beemster, L. De Veylder, S. Vercruysse, G. West, D. Rombaut, P. Van Hummelen, A. Galichet, W. Gruissem, D. Inzé, and M. Vuylsteke, *Genome-wide analysis of gene expression profiles associated with cell cycle mode transitions in growing organs of Arabidopsis thaliana*, Plant Physiol **138** (2005), 734–743.

[6] G. T. Beemster, F. Fiorani, and D. Inzé, *Cell cycle: the key to plant growth control?*, Trends Plant Sci **8** (2003), no. 4, 154–8.

[7] E. Benková, M. Michniewicz, M. Sauer, T. Teichmann, D. Seifertova, G. Jurgens, and J. Friml, *Local, efflux-dependent auxin gradients as*

a common module for plant organ formation, Cell **115** (2003), no. 5, 591–602.

[8] D. C. Bergmann and F. D. Sack, *Stomatal development*, Annu Rev Plant Biol **58** (2007), 163–81.

[9] U. K. Binné, M. K. Classon, F. A. Dick, W. Wei, M. Rape, W. G. Jr. Kaelin, A. M. Näär, and N. J. Dyson, *Retinoblastoma protein and anaphase promoting complex physically interact and functionally cooperate during cell cycle exit*, Nature cell biology **9** (2007), 225–232.

[10] C. Bohr, K. A. Hasselbalch, and A. Krogh, *Concerning a biologically important relationship - the influence of the carbon dioxide content of blood on its oxygen binding*, Skand. Arch. Physiol **16** (1904), 402–412.

[11] V. Boudolf, K. Vlieghe, G. T. Beemster, Z. Magyar, J. A. Torres-Acosta, S. Maes, E. Van Der Schueren, D. Inzé, and L. De Veylder, *The plant-specific cyclin-dependent kinase cdkb1;1 and transcription factor e2fa-dpa control the balance of mitotically dividing and endoreduplicating cells in Arabidopsis*, Plant Cell **16** (2004), 2683–2692.

[12] J. L. Bowman, *Axial patterning in leaves and other lateral organs*, Curr Opin Genet Dev **10** (2000), no. 4, 399–404.

[13] A. Cebolla, J. M. Vinardell, E. Kiss, B. Oláh, F. Roudier, A. Kondorosi, and E. Kondorosi, *The mitotic inhibitor ccs52 is required for endoreduplication and ploidy dependent cell enlargement in plants*, EMBO J. **18** (1999), 4476–4484.

[14] M. L. Churchman, M. L. Brown, N. Kato, V. Kirik, M. Hulskamp, D. Inzé, L. De Veylder, J. D. Walker, Z. Zheng, D. G. Oppenheimer,

T. Gwin, J. Churchman, and J. C. Larkin, *Siamese, a novel plant-specific cell cycle regulator, controls endoreplication onset in arabidopsis thaliana*, Plant Cell **18** (2006), 3145–3157.

[15] T. F. Coleman and Y. Li, *On the convergence of reflective newton methods for large-scale nonlinear minimization subject to bounds*, Mathematical Programming **67** (1994), 189–224.

[16] ———, *An interior, trust region approach for nonlinear minimization subject to bounds*, SIAM Journal on Optimization **6** (1996), 418–445.

[17] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization*, SIAM, 2009.

[18] A. Csikász-Nagy, D. Battogtokh, K. C. Chen, B. Novák, and J. J. Tyson, *Analysis of a generic model of eukaryotic cell-cycle regulation*, Biophysical J **90** (2006), 4361–4379.

[19] S. R. Cutler, D. W. Ehrhardt, J. S. Griffitts, and C. R. Somerville, *Random gfp::cdna fusions enable visualization of subcellular structures in cells of arabidopsis at a high frequency*, PNAS **97** (2000), 3718–3723.

[20] A. de Bruin, B. Maiti, L. Jakoi, C. Timmers, R. Buerki, and G. Leone, *Identification and characterization of e2f7, a novel mammalian e2f family member capable of blocking cellular proliferation*, J Biol Chem **278** (2003), 42041–42049.

[21] L. De Veylder, T. Beeckman, G. T. Beemster, L. Krols, F. Terras, I. Landrieu, E. Van Der Schueren, S. Maes, M. Naudts, and D. Inzé, *Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis*, Plant Cell **13** (2001), 1653–1667.

[22] W. Dewitte, C. Riou-Khamlichi, S. Scofield, J. M. Healy, A. Jacqmard, N. J. Kilby, and J. A. Murray, *Altered cell cycle distribution, hyperplasia, and inhibited differentiation in arabidopsis caused by the d type cyclin cycd3*, Plant Cell **15** (2003), 79–92.

[23] D. K. Dimova and N. J. Dyson, *The e2f transcriptional network: old acquaintances with new faces*, Oncogene **24** (2005), 2810–2826.

[24] P. M. Donnelly, D. Bonetta, H. Tsukaya, R. E. Dengler, and N. G. Dengler, *Cell cycling and cell enlargement in developing leaves of Arabidopsis*, Developmental Biology **215** (1999), 407–419.

[25] B. A. Edgar and T. L. Orr Weaver, *Endoreplication cell cycles: more for less*, Cell **105** (2001), 297–306.

[26] A. Engler Jde, L. De Veylder, R. De Groodt, S. Rombauts, V. Boudolf, B. De Meyer, A. Hemerly, P. Ferreira, T. Beeckman, M. Karimi, P. Hilson, D. Inzé, and G. Engler, *Systematic analysis of cell-cycle gene expression during arabidopsis development*, Plant J. **59** (2009), 645–660.

[27] A. Finkelstein, J. Hetherington, L. Li, O. Margoninski, P. Saffrey, R. Seymour, and A. Warner, *Computational challenges of systems biology*, Computer **37** (2004), 26–33.

[28] F. Fiorani and G. T. Beemster, *Quantitative analyses of cell division in plants*, Plant Molecular Biology **60** (2006), 963–979.

[29] B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser, *Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images*, Microsc Res Tech **65** (2004), 33–42.

[30] K. Fülöp, S. Tarayre, Z. Kelemen, G. Horváth, Z. Kevei, K. Nikovics, L. Bakó, S. Brown, A. Kondorosi, and E. Kondorosi, *Arabidopsis*

anaphase promoting complexes: multiple activators and wide range of substrates might keep apc perpetually busy, Cell Cycle **4** (2005), 1084–1092.

[31] M. Geisler, J. Nadeau, and F. D. Sack, *Oriented asymmetric divisions that generate the stomatal spacing pattern in Arabidopsis are disrupted by the too many mouths mutation*, The Plant Cell **12** (2000), 2075–2086.

[32] W. Govaerts, V. De Witte, and L. Kheibarshekan, *Using matcont in a two-parameter bifurcation study of models for cell cycle controls*, Proceedings of the ASME, International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (2009).

[33] P. B. Green, *Growth and cell pattern formation on an axis: Critique of concepts, terminology, and modes of study*, Botanical Gazette **137** (1976), no. 3, 187–202.

[34] P. B. Green and K. Bauer, *Analysing the changing cell cycle*, Journal of Theoretical Biology **68** (1977), 299–315.

[35] O. Hamant, M. G. Heisler, H. Jonsson, P. Krupinski, M. Uyttewaal, P. Bokov, F. Corson, P. Sahlin, A. Boudaoud, E. M. Meyerowitz, Y. Couder, and J. Traas, *Developmental patterning by mechanical signals in arabidopsis*, Science **322** (2008), 1650–1655.

[36] A. V. Hill, *The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves*, J. Physiol. **40** (1910), iv–vii.

[37] D. Inzé and L. De Veylder, *Cell cycle regulation in plant development*, Annu. Rev. Genet. **40** (2006).

[38] Dirk Inzé, *Cell cycle control and plant development*, Blackwell Publishing Ltd, 2007.

[39] V. B. Ivanov, A. E. Dobrochaev, and T. I. Baskin, *What the distribution of cell lengths in the root meristem does and does not reveal about cell division*, Journal of Plant Growth Regulation **21** (2002), 60–67.

[40] C. P. Keller and E. Van Volkenburgh, *Auxin-induced epinasty of tobacco leaf tissues (a nonethylene-mediated response)*, Plant Physiol **113** (1997), no. 2, 603–610.

[41] L. Kheibarshekan, S. Dhondt, V. Boudolf, G. T. S. Beemster, D. Inzé, W. Govaerts, and L. De Veylder, *Model-based analysis of the developmental control of epidermal cell growth and division in the Arabidopsis leaf*, submitted (2010).

[42] H. Kitano, *Foundations of systems biology*, MIT Press, Cambridge, MA, 2001.

[43] ———, *Systems biology: A brief overview*, Science **295** (2002), 1662–1664.

[44] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, *Systems biology in practice: Concepts, implementation and application*, Wiley-VCH, 141–157, 2005.

[45] A. Kotzer and G. Wasteneys, *Mechanisms behind the puzzle: microtubulemicrofilament cross-talk in pavement cell formation*, Canadian Journal of Botany **84** (2006), no. 4, 594–603.

[46] T. Lammens, V. Boudolf, L. Kheibarshekan, L. P. Zalmas, T. Gaamouche, S. Maes, M. Vanstraelen, E. Kondorosi, N. B. La Thangue, W. Govaerts, D. Inzé, and L. De Veylder, *Atypical e2f activity restrains apc/cccs52a2 function obligatory for endocycle onset*, Proc Natl Acad Sci USA **105** (2008), 14721–14726.

[47] T. Lammens, J. Li, G. Leone, and L. De Veylder, *Atypical e2fs: new players in the e2f transcription factor family*, Trends Cell Biol **19** (2009), 111–118.

[48] J. C. Larkin, M. D. Marks, J. Nadeau, and F. Sack, *Epidermal cell fate and patterning in leaves*, Plant Cell **9** (1997), no. 7, 1109–20.

[49] A. Lasorella, J. Stegmller, D. Guardavaccaro, G. Liu, M. S. Carro, G. Rothschild, L. de la Torre-Ubieta, M. Pagano, A. Bonni, and A. Iavarone, *Degradation of id2 by the anaphase promoting complex couples cell cycle exit and axonal growth*, Nature **442** (2006), 471–474.

[50] K. Levenberg, *A method for the solution of certain non-linear problems in least squares*, The Quarterly of Applied Mathematics **2** (1944), 164 168.

[51] J. Li, C. Ran, E. Li, F. Gordon, G. Comstock, H. Siddiqui, W. Cleghorn, H. Z. Chen, K. Kornacker, C. G. Liu, S. K. Pandit, M. Khanizadeh, M. Weinstein, G. Leone, and A. de Bruin, *Synergistic function of e2f7 and e2f8 is essential for cell survival and embryonic development*, Dev. Cell **14** (2008), 62–75.

[52] M. A. Lilly and R. J. Duronio, *New insights into cell cycle control from the drosophila endocycle*, Oncogene **24** (2005), 2765–2775.

[53] E. Liscum and E.L. Stowe-Evans, *Phototropism: a "simple" physiological response modulated by multiple interacting photosensory-response pathways*, Photochemistry and Photobiology **72** (2000), no. 3, 273–282.

[54] B. Maiti, J. Li, A. de Bruin, F. Gordon, C. Timmers, R. Opavsky, K. Patil, J. Tuttle, W. Cleghorn, and G. Leone, *Cloning and characterization of mouse e2f8, a novel mammalian e2f family member*

capable of blocking cellular proliferation, J Biol Chem **280** (2005), 18211–18220.

[55] J. H. Mathews and K. K. Fink, *Numerical methods using matlab, 4th edition*, Prentice-Hall Inc, Numerical Optimization, Chapter 8, 2004.

[56] K. Mehdi Khanlou, K. Vandepitte, L. Kheibarshekan, and E. Van Bockstaele, *Towards an optimal sampling strategy to assess the genetic variation within and among white clover (trifolium repens l.) cultivars using aflp*, submitted (2010).

[57] M. Menges, S.M. de Jager, W. Gruissem, and J.A.H. Murray, *Global analysis of the core cell cycle regulators of arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cycle control*, Plant J **41** (2005), 546–566.

[58] W. Nagl, *Dna endoreduplication and polyteny understood as evolutionary strategies*, Science **261** (1976), 614–615.

[59] K. Narbonne-Reveau, S. Senger, M. Pal, A. Herr, H. E. Richardson, M. Asano, P. Deak, and M. A. Lilly, *Apc/cfzr/cdh1 promotes cell cycle progression during the drosophila endocycle*, Development **135** (2008), 1451–1461.

[60] J. A. Nelder and R. Mead, *A simplex method for function minimization*, Comput J **7** (1965), 308–313.

[61] B. Novák and J. J. Tyson, *Quantitative analysis of a molecular model of mitotic control in fission yeast*, Theor. Biol **173** (1995), 283–305.

[62] E. Panteris and B. Galatis, *The morphogenesis of lobed plant cells in the mesophyll and epidermis: organization and distinct roles of cortical microtubules and actin filaments*, New Phytol **167** (2005), no. 3, 721–32.

[63] J. M. Peters, *The anaphase promoting complex/cyclosome: a machine designed to destroy*, Nat. Rev. Mol. Cell Biol. **7** (2006), 644–656.

[64] S. Pien, J. Wyrzykowska, S. McQueen-Mason, C. Smart, and A. Fleming, *Local expression of expansin induces the entire process of leaf development and modifies leaf shape*, Proc Natl Acad Sci U S A **98** (2001), no. 20, 11812–7.

[65] E. Polak, *Optimization: algorithms and consistent approximations*, Springer-Verlag, 1997.

[66] A. Porceddu, H. Stals, J. P. Reichheld, G. Segers, L. De Veylder, R. P. Barroco, P. Casteels, M. Van Montagu, D. Inzé, and V. Mironov, *A plant-specific cyclin-dependent kinase is involved in the control of g2/m progression in plants*, J Biol Chem **276** (2001), 36354–36360.

[67] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in c: The art of scientific computing*, Cabridge University Press, New York, 1992.

[68] E. Ramirez-Parra, M. A. López Matas, C. Frndt, and C. Gutierrez, *Role of an atypical e2f transcription factor in the control of arabidopsis cell growth and differentiation*, Plant Cell **16** (2004), 2350–2363.

[69] G. V. Reddy, M. G. Heisler, D. W. Ehrhardt, and E. M. Meyerowitz, *Real-time lineage analysis reveals oriented cell divisions associated with morphogenesis at the shoot apex of arabidopsis thaliana*, Development **131** (2004), 4225–4237.

[70] D. Reinhardt, T. Mandel, and C. Kuhlemeier, *Auxin regulates the initiation and radial position of plant lateral organs*, Plant Cell **12** (2000), no. 4, 507–18.

[71] C. Roberts, *Ordinary differential equations: Applications, models, and computing*, Chapman-Hall, 131–135, 2010.

[72] A. H. K. Roeder, V. Chickarmane, A. Cunha, B. Obara, B. S. Manjunath, and E. M. Meyerowitz, *Variability in the control of cell division underlies sepal epidermal patterning in arabidopsis thaliana*, PLoS Biol **8** (2010), e1000367.

[73] S. Savaldi-Goldstein, C. Peto, and J. Chory, *The epidermis both drives and restricts plant shoot growth*, Nature **446** (2007), 199–202.

[74] V. Schaeffer, C. Althauser, H. R. Sccherbata, W. M. Deng, and H. Ruohola Baker, *Notch dependent fizzy related/hec1/cdh1 expression is required for the mitotic to endocycle transition in drosophila follicle cells*, Curr. Biol. **14** (2004), 630–636.

[75] A. Schnittger, U. Schöbinger, D. Bouyer, C. Weinl, Y. D. Stierhof, and M. Hülskamp, *Ectopic d-type cyclin expression induces not only dna replication but also cell division in arabidopsis trichomes*, Proc Natl Acad Sci USA **99** (2002), 6410–6415.

[76] S. J. Sigrist and C. F. Lehner, *Drosophila fizzy related down regulates mitotic cyclins and is required for cell proliferation arrest and entry into endocycles*, Cell **90** (1997), 671–681.

[77] J. A. Snyman, *Practical mathematical optimization: An introduction to basic optimization theory and classical and new gradient-based algorithms*, Springer Publishing, 2005.

[78] D. A. Sorrell, M. Menges, J. M. Healy, Y. Deveaux, C. Amano, Y. Su, H. Nakagami, A. Shinmyo, J. H. Doonan, M. Sekine, and J. A. Murray, *Cell cycle regulation of cyclin-dependent kinases in tobacco cultivar bright yellow-2 cells*, Plant Physiol. **126** (2001), 1214–1223.

[79] H. Stals and D. Inzé, *When plant cells decide to divide*, Trends in Plant Science **6** (2001), 359–364.

[80] R. Swarup, E. M. Kramer, P. Perry, K. Knox, H. M. Leyser, J. Haseloff, G. T. Beemster, R. Bhalerao, and M. J. Bennett, *Root gravitropism requires lateral root cap and epidermal cells for transport and response to a mobile auxin signal*, Nat Cell Biol **7** (2005), no. 11, 1057–65.

[81] H. Tsukaya, *Interpretation of mutants in leaf morphology: genetic evidence for a compensatory system in leaf morphogenesis that provides a new link between cell and organismal theories*, Int Rev Cytol **217** (2002), 1–39.

[82] ———, *Controlling size in multicellular organs: focus on the leaf*, PLOS Biology **6** (2008), 1373–1376.

[83] J. J. Tyson and B. Novak, *Regulation of the eukaryotic cell cycle: Molecular antagonism, hysteresis, and irreversible transitions*, J. theor. Biol. **210** (2001), 249–263.

[84] S. Ubeda-Tomas, F. Federici, I. Casimiro, G. T. Beemster, R. Bhalerao, R. Swarup, P. Doerner, J. Haseloff, and M. J. Bennett, *Gibberellin signaling in the endodermis controls arabidopsis root meristem size*, Curr Biol **19** (2009), no. 14, 1194–9.

[85] A. Verkest, C. L. Manes, S. Vercruysse, S. Maes, E. Van Der Schueren, T. Beeckman, P. Genschik, M. Kuiper, D. Inzé, and L. De Veylder, *The cyclin-dependent kinase inhibitor krp2 controls the onset of the endoreduplication cycle during arabidopsis leaf development through inhibition of mitotic cdka;1 kinase complexes*, Plant Cell **17** (2005), no. 6, 1723–36.

[86] K. Vlieghe, V. Boudolf, G. T. Beemster, S. Maes, Z. Magyar, A. Atanassova, J. de Almeida Engler, R. De Groodt, D. Inzé, and

De Veylder L., *The dp e2f like del1 gene controls the endocycle in arabidopsis thaliana*, Curr. Biol. **15** (2005), 59–63.

[87] C. Weinl, S. Marquardt, S. J. Kuijt, M. K. Nowack, M. J. Jakoby, M. Hülskamp, and Schnittger A., *Novel functions of plant cyclin dependent kinase inhibitors, ick1/krp1, can act non cell autonomously and inhibit entry into mitosis*, Plant Cell **17** (2005), 1704–1722.

[88] Y. Yamamoto, Y. Kobayashi, and H. Matsumoto, *Lipid peroxidation is an early symptom triggered by aluminum, but not the primary cause of elongation inhibition in pea roots*, Plant Physiol **125** (2001), 199–208.

[89] Y. Yu, A. Steinmetz, D. Meyer, S. Brown, and W. H. Shen, *The tobacco a type cyclin, nicta;cyca3;2, at the nexus of cell division and differentiation*, Plant Cell **15** (2003), 2763–2777.

[90] L. P. Zalmas, X. Zhao, A. L. Graham, R. Fisher, C. Reilly, A. S. Coutts, and N. B. La Thangue, *Dna-damageresponse control of e2f7 and e2f8*, EMBORep **9** (2008), 252–259.