DISSERTATION
submitted
to the
Combined Faculty for the Natural Sciences and for Mathematics
of the
Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Master of Science **Joanna Kawka**
Born in Chorzów, Poland

Oral examination:

Mathematical modeling of *SGK1* dynamics in medulloblastoma tumor cells

1. Advisor: Prof. Dr. Anna Marciniak-Czochra
2. Advisor:

Abstract

This work is devoted to mathematical modeling of deregulation of the Wnt/β-catenin signaling pathway in medulloblastoma resulting in abnormal dynamics of target genes.

Medulloblastoma is a brain tumor, mostly diagnosed in children. It is associated with several molecular genetic alterations. Specific aberrations of chromosome 6q, leading either to the chromosome copy-number loss (monosomy 6) or gain (trisomy 6), occur in two different subtypes of the tumor.

The model is a nine-dimensional system of ordinary differential equations and describes nonlinear dynamics of the key ingredients of the signaling process. The model is based on the law of mass action and accounts for a two-compartment architecture of a cell consisting of the nucleus and cytoplasm. The model helps to understand molecular differences between the two medulloblastoma mutation subtypes that are associated with different patient prognosis.

Our studies are based on a collaboration with the group of Prof. Dr. med. Stefan Pfister at the Division of Pediatric Neuro-oncology Research Group of the German Cancer Research Center (DKFZ). The model is used to evaluate data from the gene expression microarray data from the clinics in Heidelberg, Boston and Amsterdam.

Numerical simulations lead to new biological hypotheses related to a significant role of the regulatory loop *SGK1-GSK3*β*-MYC*, a part of the Wnt/β-catenin signaling pathway. Simulations indicate the advantage of using the pharmacological inhibitor of *SGK1* in patients with copy-number gain of chromosome 6q. Finally, the simulation results suggest a beneficial use of an adjuvant therapy in a trisomy 6 treatment.

Mathematical analysis of the ordinary differential equations system confirms the well-posedness of the model and provides basic properties of the solutions. Supported by numerical analysis, we conclude about global stability of a unique positive equilibrium corresponding to the homeostasis of the system. We also tackle the parameter estimation

problem using statistical assessment of the results and Gauss-Newton method. Sensitivity analysis provides insight into the role of model parameters. In particular, it confirms the sensitivity of the system to the parameter of *SGK1* degradation.

The model provides a powerful tool to study mechanistically the underlying process and to support the experiments.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Deregulation des Wnt/β-catenin Signalweges in Medulloblastomen, und der daraus resultierenden Fehlregulation seiner Zielgene.

Das Medulloblastom ist ein meist kindlicher Hirntumor, der mit charakteristischen genetischen Veränderungen einhergeht. Spezifische Aberrationen des Chromosoms 6q führen entweder zu einem Verlust (copy number loss, Monosomie 6) oder zu einem Zugewinn (copy number gain, Trisomie 6) von Genkopien und erlauben die Unterscheidung von zwei verschiedenen Medulloblastomsubtypen.

Das in dieser Arbeit entwickelte Modell besteht aus einem neundimensionalen System gewöhnlicher Differentialgleichungen, die die nichtlineare Dynamik der Schlüsselkomponenten des Wnt/β-catenin Signalweges beschreiben. Das Modell beruht auf dem Massenwirkungsgesetz und berücksichtigt die Kompartimentierung eukaryontischer Zellen in Zellkern (Nukleus) und Zytoplasma. Es dient dem Verständnis der molekularen Unterschiede zwischen den beiden erwähnten prognostisch unterschiedlichen Medulloblastomsubtypen.

Diese Arbeit basiert auf einer Kooperation mit der Abteilung Pädiatrische Neuro-Onkologie (Prof. Dr. med. Stefan Pfister) des Deutschen Krebsforschungszentrums (DKFZ). Das entwickelte Modell wird zur Evaluation von Microarray Genexpressionsdaten verschiedener Kliniken in Heidelberg, Boston und Amsterdam verwendet.

Numerische Simulationen führen zu neuen, biologisch relevanten Hypothesen im Hinblick auf die herausragende Rolle der Regulationsschleife *SGK1-GSK3β-MYC*, die Teil des Wnt/β-catenin Signalwegs ist. Die Simulationen legen den Einsatz pharmakologischer SGK1-Inhibitoren in Patienten mit copy number gain des Chromosoms 6 sowie positive Auswirkungen einer adjuvanten Therapie der Trisomie 6 nahe.

Die mathematische Untersuchung des Differentialgleichungssystems bestätigt die Wohlgestelltheit des Modells und liefert grundlegende Eigenschaften seiner Lösungen.

Gestützt durch numerische Methoden wird die globale Stabilität des eindeutigen positiven Gleichgewichtszustandes gezeigt, der der Homöostase des Systems entspricht. Mit Hilfe der Gauss-Newton-Methode wird das zugehörige Parameterschätzungsproblem betrachtet und die Ergebnisse werden statistisch untersucht. Eine Sensitivitätsanalyse bietet Einblick in die Auswirkungen verschiedener Modellparameter. Insbesondere bestätigt sie die Sensitivität des Systems bezüglich der Degradierungsrate von SGK1.

Das entwickelte Modell stellt ein leistungsfähiges Werkzeug zum mechanistischen Studium der dem Medulloblastom zugrundeliegenden Prozesse dar und liefert Erklärungsansätze für experimentelle Befunde.

Acknowledgements

Homo doctus in se semper divitias habet
"A learned person always has riches within himself"
Phaedrus, a Roman fabulist

To start is easier than to finish. I would like to give my great gratitude to everyone who helped me to accomplish my PhD thesis.
Especially, I would like to thank Prof. Dr. Anna Marciniak-Czochra who gave me invaluable guidance during whole scientific research and showed me mathematics in the everyday world.
Nevertheless, this project would not be possible without full support of Prof. Dr. med. Stefan Pfister who I am grateful for medical assistance and time, so costly in his profession. I cannot avoid mentioning his student, Dominik Sturm, who had patience to answer my biological questions.
I am also thankful to Dr. Thomas Carraro, who contributed to my PhD thesis while working on the parameter estimation problem, for his helpful tuition and clever advise.
Finally, I would like to thank my family, boyfriend and friends for encouragement and support coming from the heart. That, they never stopped believing in me and cheered me up with words and deeds.

Contents

CHAPTER 1

Introduction

Increased contribution of mathematical modeling is observed in many sections of modern life. It is involved in natural science, social science and engineering. Well defined models can describe complicated systems, such as signaling cell pathways, and can help in understanding the processes, as well as answering biological questions without time and money consuming experiments. One can monitor the process development and perceive its characteristic attributes. In the context of medulloblastoma, we model intracellular signaling processes belonging to the Wnt/β-catenin pathway, which are decisive for dynamics of the disease.

This work is based on systems biology approach to understand medulloblastoma dynamics, its clinical prognosis and to propose possible treatments. Medulloblastoma is a solid brain tumor linked to mutations related to chromosome 6q. The resulting aberration of the number of gene copies leads to abnormal dynamics of the Wnt/β-catenin pathway, which causes increased or decreased synthesis of several proteins influencing in turn cell proliferation, apoptosis and consequently the tumor growth. The aim is to develop a mathematical model which provides an insight into the dynamics of medulloblastoma. Our approach is based on the methods of mathematical modeling, analysis, simulation and parameter estimation.

Cancer development and growth of solid tumors is a multiscale process, where the macroscopically observed growth of tumor is a result of the abnormal cell turnover and movement, which is governed by intracellular signaling. In this work, we focus on a mathematical description of the intracellular processes.

The Wnt/β-catenin signaling pathway attracts a lot of interests in experimental biology [13], [31], [75]. Since it plays an important role in developmental processes [76] and

cancerogenesis [36], [45], the Wnt/β-catenin signaling pathway and related proteins are a broad area of investigations, also using the tools of mathematical modeling. The pioneers in modeling of the core elements of the pathway were Reinhart Heinrich, Roland Krüger and Ethan Lee in the context of colon cancer (for review, see [40], [42]). In this thesis, we do not focus on the system delineated by Heinrich's work. We develop a model that involves the ingredients of the pathway which are crucial for medulloblastoma tumor dynamics.

Medulloblastoma may be caused by several mutations. We consider monosomy and trisomy of chromosome 6q. The aberration of chromosome 6 in long arm refers to monosomy, when there is a loss of chromosome 6q. In opposite, the gain of chromosome 6q is ascribed to the trisomy. The striking fact is that the two types of medulloblastoma reveal extremely different prognosis.

The deregulation of the Wnt/β-catenin signaling pathway has a huge impact on the downstream target genes. The cascade of subsequent abnormal interactions between these genes results in tumorigenesis. In the case of monosomy 6 and trisomy 6, we observe the deregulation of mRNA levels. A question arises, how to investigate such complex problem in the context of modeling.

Based on the literature and collaboration with the group of Prof. Dr. med. Stefan Pfister at the Division of Pediatric Neuro-oncology Research Group of the German Cancer Research Center (DKFZ), we develop an ordinary differential system, describing dynamics of the key ingredients of the Wnt/β-catenin signaling system. To describe intracellular dynamics, we consider two-compartment approach taking into account transport between the cell nucleus and cytoplasm. To model the system, we reduce it to its main components. This enables obtaining a clear and comprehensive model, which still describes the complex system. We focus on the interplay between *MYC* and serum and glucocorticoidinducible kinase 1 (*SGK1*), which are the products of the Wnt/β-catenin signaling pathway, and glycogen synthase kinase (*GSK3*β). Numerical simulations of the model yield a better understanding of the process. In particular, the model indicates the importance of the *SGK1* gene in the development of medulloblastoma, where target genes of the Wnt/β-catenin signaling pathway show strong aberration. Finally, we employ the gene expression microarray data obtained from the clinics due to cooperation with Stefan Pfister and Dominik Sturm for two types of medulloblastoma, monosomy and trisomy of chromosome 6q.

We study dynamics of the system and investigate to what extent the prognosis is related to the deregulation of the *MYC* and *SGK1* mRNAs. After a series of numerical simulations, we formulate biological hypotheses on the significant role of loop *SGK1-GSK3*β*-MYC* (part of Wnt/β-catenin signaling pathway). We also propose a new therapy based on a pharmacological inhibitor as an adjuvant therapy to the one that is nowadays

in use.

The outline of this manuscript is the following. We introduce biological background in Chapter 2. We explain the differences between two types of medulloblastoma. Then, we bring closer molecular details of the Wnt/β-catenin signaling pathway and target genes of the pathway.

In Chapter 3 we elucidate dynamics of the system. We describe all underlying processes such as transcription, translation, phosphorylation, protein degradation and transport between the nucleus and cytoplasm. Consistently, we formulate mathematical equations based on the biological phenomena and the law of mass action.

Chapter 4 is devoted to the mathematical properties of the model. We discuss well-possedness in the sense of classical theory of ordinary differential equations. We prove global existence and uniqueness of the solutions and present numerical simulations of the system for several sets of initial conditions. Supported by numerical calculations, we show that the solution is globally, asymptotically stable with the equilibrium point classified as focus.

Numerical solutions of the model are presented in Chapter 5. Various aspects of particular protein kinetics are demonstrated. We present the solutions and compare them to reveal the biological meaning. The crucial role of *SGK1* in the system is explored. The suggested adjuvant therapy shows the influence of the pharmacological inhibitor of the SGK1 protein to the homeostasis in the investigated system.

To study the parameter estimation problem, we deliver the commonly used methods in Chapter 6. We perform the sensitivity analysis and run Gauss-Newton's algorithm to identify the parameters as well as we provide the statistical assessment of the solution. From the sensitivity analysis, we show the importance of the *SGK1* gene. At the end of the chapter the identified parameters are refined by the method of sequential optimal experimental design and hence we obtain a set of admissible estimates for the model.

The final Chapter 7 encompasses a summary and outline of the future research directions. This completes the content of the manuscript starting from the biology, through the modeling and finishing on the optimal experimental design.

CHAPTER 2

Biological background

A tumor, defined as abnormal mass of tissue, may be benign or malignant. In this chapter we present the nature of a specific brain tumor type *Medulloblastoma*, statistics and the biological issues that lie beyond the topic. We focus on the Wnt/β-catenin signaling pathway in medulloblastoma and in the end we derive a scheme of the modeled system, where we consider the fundamental biological processes.

2.1 Medulloblastoma

Medulloblastoma is a malignant brain tumor that mainly affects children. 85% of all medulloblastoma cases are below the age of eighteen [39]. The mortality during the first two years after diagnosis oscillates between 10% and 15% [38]. The rapidly-growing tumor is localized in the brain area which controls speech, balance, and posture. Depending on the age of the patient the symptoms can be headaches, vomiting, nausea, gait abnormality, eye squint, sensory neuropathy [2], [32], [61]. In the early stage of tumor development the symptoms are often not distinctive and for this reason the disease is not apparent. Fast recognition gives better chance for the children recovery, however symptoms are too general and often the disease is diagnosed in late stage. The diagnosis is confirmed after magnetic resonance imaging (MRI) and furthermore detailed

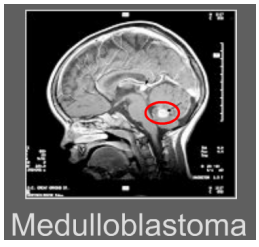

Figure 2.1: Computer tomography brain scan showing a medulloblastoma tumor (circled in red). Image thanks to S. Pfister.

tests are carried out (e.g., analysis of the cerebrospinal fluid). If the medulloblastoma is recognized, the treatment involves resection of the tumor and then, depending on the patient's stage and molecular aberration, radiotherapy, chemotherapy or both therapies are applied [58], [62].

Biological evidence 2.1 *Medulloblastoma is attributed to several mutations in the tumor genome, such as 17q gain, i(17q), MYC/MYCN amplification, 6q gain, 6q loss and Wnt pathway activation ([55], [70], Figure 2.2). Importantly, prognosis depends on the type of mutation causing the disease as well as on the stage of diagnosis (e.g., size and extent of the tumor, etc.). In general, tumors can be either non-invasive or metastasize through the cerebrospinal fluid [54].*

Nowadays, the treatment technics are improved, but still the side effects are a big concern. Although many children recover, they can have endocrinological, cognitive and neurological problems, such as speech difficulties.

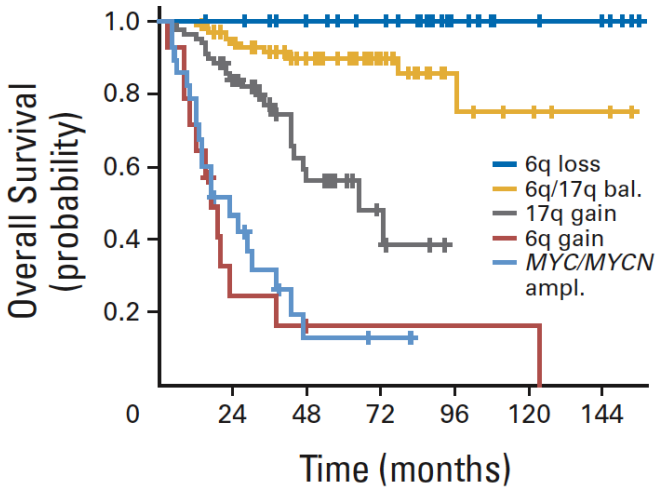

Figure 2.2: Overall survival probabilities for different chromosomal mutations in medulloblastoma. Adapted from Pfister, 2009.

2.1.1 Monosomy 6 and trisomy 6

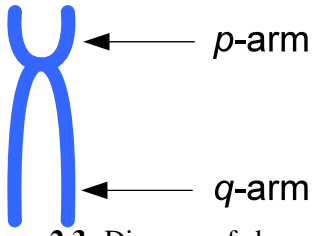

Figure 2.3: Diagram of chromosome structure pointing long and short arms. Copyright ©motifolio.com

In a healthy system, there are two (almost) identical copies of each chromosome. Each copy has a long and short arm (Figure 2.3). The short arm is described as *p*-arm, where *p* is taken from the french word *petit* - small. The long arm is called *q*-arm, where *q* is just the next letter from the Latin alphabet. Trisomy is a mutation in which there are three copies of chromosome 6 and in monosomy there is only one copy of chromosome 6. Our research is devoted to understanding the role of signaling pathways in two types of medulloblastoma, carrying either a monosomy or a trisomy of the long arm of chromosome 6q. Monosomy is loss of one chromosome copy (6q loss) and trisomy is gain of one chromosome copy (6q gain).

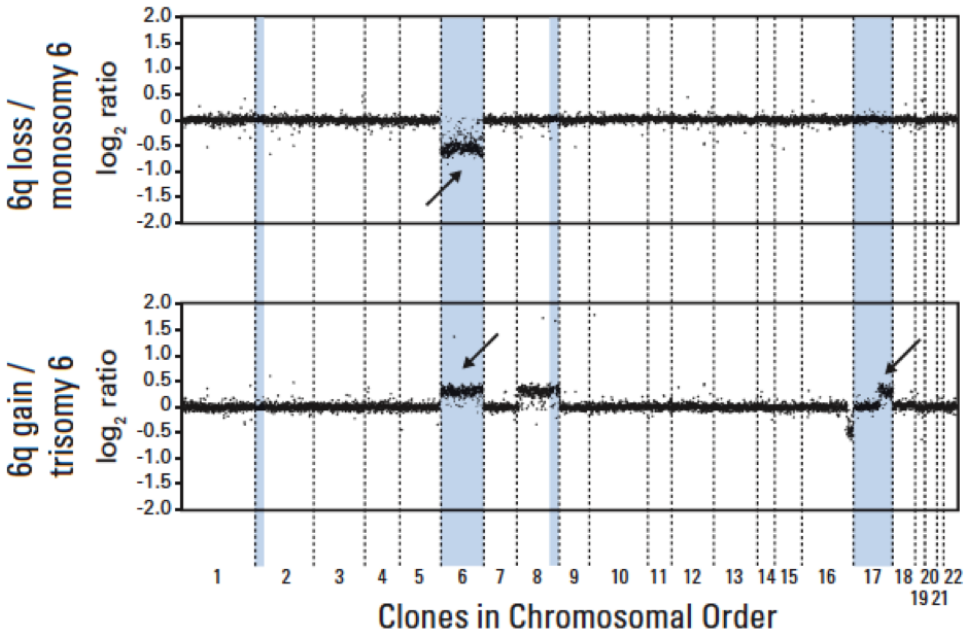

Figure 2.4: A genomic hybridization profiles of tumors with monosomy 6 and trisomy 6. Adapted from Pfister, 2009.

The DNA hybridization profiles from the two types of tumor show the discrepancy between chromosomal aberrations (Figure 2.4). Each of these mutations is associated with a strikingly different patient prognosis (Figure 2.5).

Biological evidence 2.2 *Medulloblastoma patients harboring a trisomy of chromosome 6q in the tumor genome are found to have a poor prognosis, while tumors characterized by a monosomy of chromosome 6q has a good prognosis following the medical treatment [55].*

Monosomy 6 is always found in combination with the β-catenin mutation, leading to constitutive Wnt signaling activation. It was discovered that mutation of β-catenin in medulloblastoma tumor cells was associated with a good prognosis in a pediatric patient [19].

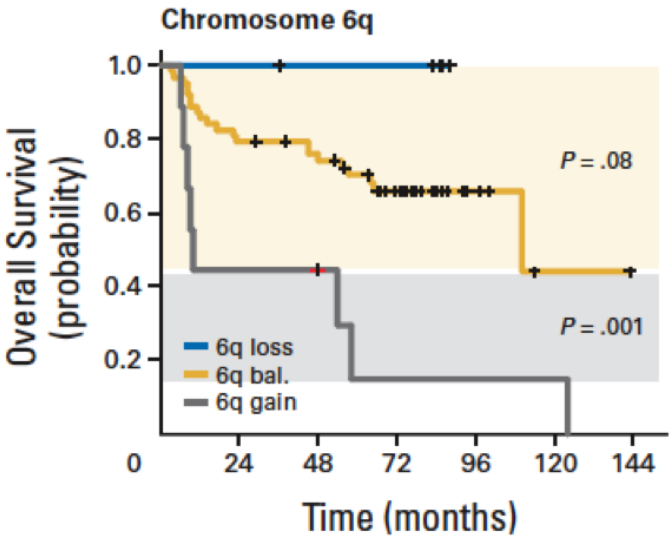

Figure 2.5: Kaplan Meier curves showing the estimated overall survival probabilities for medulloblastoma patients according to the copy-number status of chromosome arm 6q in the tumor. Adapted from Pfister, 2009.

The 6q loss and 6q gain appear to induce the deregulation of the expression of *MYC* and *SGK1* in the mutated cells (for biological notation, see Table 2.1). *MYC* is a transcription factor that regulates the expression of a number of genes and is involved in the biological processes such as cell growth and proliferation [14]. *SGK1* is responsible for the intracellular transport and cell survival [7], [68]. Both genes seem to play an important role in the cell's homeostasis and their deregulation can affect cellular processes finally resulting in carcinogenesis.

In the case of monosomy of chromosome 6q, we observe a downregulation of the mRNA level of *SGK1*, whereas its upregulation is detected in the case of trisomy 6 (Figure

Biological notation	
MYC	gene (italic)
MYC	mRNA (italic)
cMyc	protein (straight)

Table 2.1: Biological notation. Example provided for the *MYC* gene.

2.6). Interestingly, the *SGK1* gene is located on chromosome 6q, which suggests that a disruption in the chromosome balance alters the mRNA level of *SGK1*. However, the mechanism has not yet been explained. The concomitant upregulation of the mRNA level of *MYC* is found in both types of medulloblastoma. Since pediatric monosomy 6 of medulloblastoma is always related to the β-catenin mutation, the increase of the β-catenin translocation to the nucleus may explain upregulation of its target gene *MYC* [39]. The reason of *MYC* mRNA upregulation in trisomy 6 is not understood.

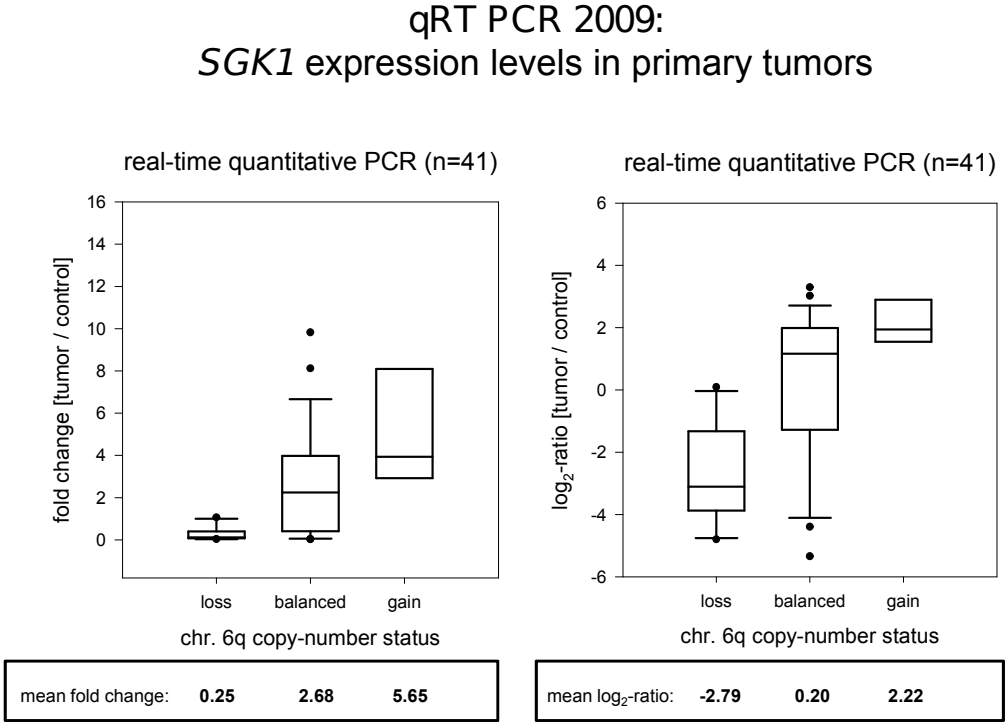

Figure 2.6: Fold change and $\log_2$-ratio of *SGK1* levels for 41 medulloblastoma samples. Comparison between different medulloblastoma subgroups: monosomy 6 (6q loss), trisomy 6 (6q gain) and "balanced", which stands for the medulloblastoma with no aberration in chromosome 6. Data thanks to S. Pfister and D. Sturm.

By definition *fold change* we understand the ratio of mRNA extracted from the tumor cells and control cells. Control cells are normal cells which are not affected by tumor.

2.2 Signaling pathway

Signaling pathway is a cascade of biochemical reactions taking place in a cell. It takes place as a response of the cells to signals they receive from environment or from other cells. It processes information from cell membrane to the genome. A membrane is equipped with receptors, which may be occupied by ligands. Those ligands are the molecules responsible for the signaling activation. If ligand binds to the receptor the cascade of protein reactions leads to an activation/deactivation of a particular set of genes (Figure 2.7).

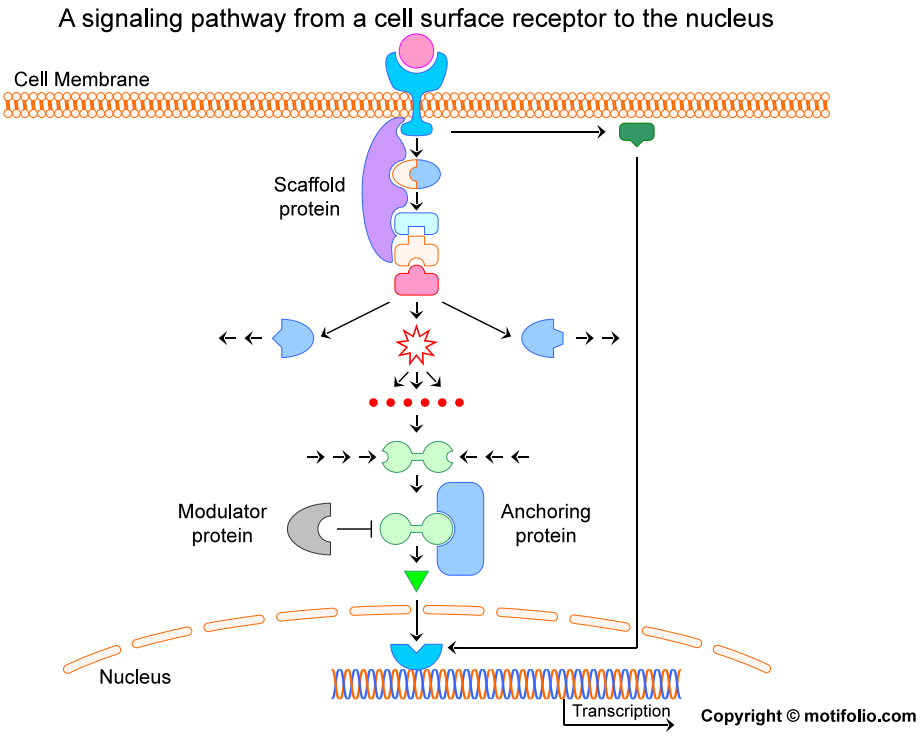

Figure 2.7: Scheme of the signaling pathway from a cell surface ligand-receptor to the nucleus. Copyright ©motifolio.com

Any changes on the molecular level that occur upon the activation of signal transmi-

tion is reflected in a cell fate, thus tissue fate (Figure 2.8). Further, different pathways often interact with each other creating a *signaling network* [24]. The proper work of the network is substantial to maintain the cellular homeostasis. Mutations taking place during cancerogenesis lead to perturbations in the corresponding signaling pathways and hence to abnormal gene expression. Here, we delineate the specific pathway which is called *Wnt/*β-catenin signaling pathway in its natural and mutated state.

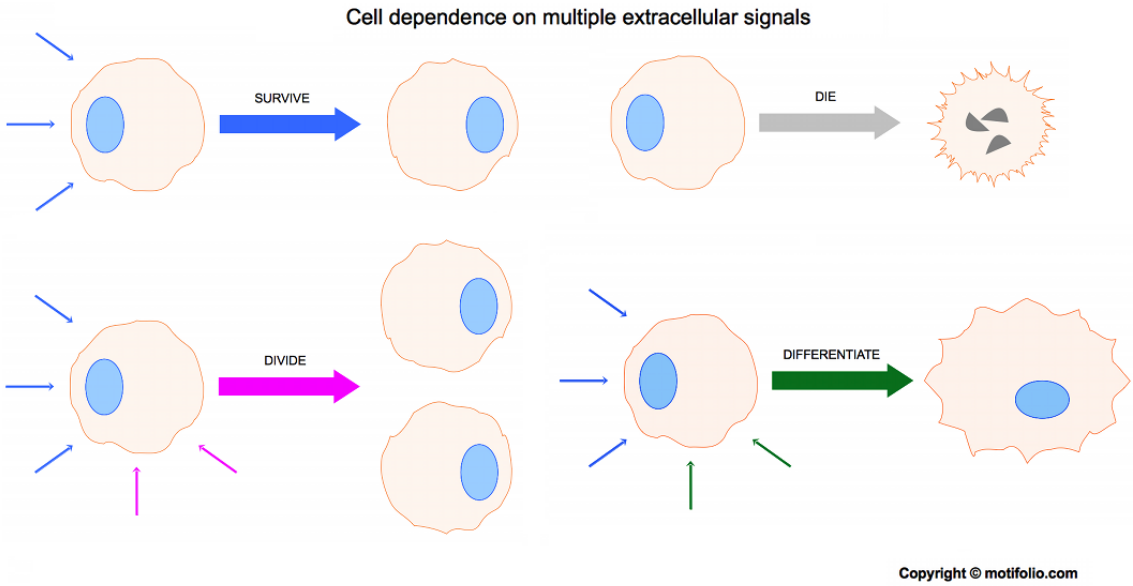

Figure 2.8: Scheme of the cell fate dependent on the multiple extracellular signals. Copyright ©motifolio.com

2.2.1 Regulation of the Wnt/β-catenin signaling pathway

The Wnt/β-catenin signaling pathway is one of the most important pathways in human cells. The Wnt gene was discovered already 30 years ago [35]. The knowledge about its function and role in the cell increased during the recent years [30]. The so called *canonical Wnt pathway* describes a cascade of reactions regulating embryonal development and adult tissue maintenance [76]. Substantially, destabilization of the Wnt pathway leads to tumorigenesis and was found in many cancers [36], [45].

Normal situation in the absence of Wnt ligands

Revealed in the former, cells can respond to the extracellular environment via ligands that bind to the receptors in the cell membrane. In the Wnt signaling pathway, if there are no Wnt ligands which could bind to low-density lipoprotein receptor-related proteins (LRP), no activation of cell surface receptor proteins (Frizzled) occurs. The multiprotein destruction complex captures then and phosphorylates β-catenin as the complex cannot be destroyed by inactive Frizzled [25], [71]. Such phosphorylated β-catenin is marked for degradation and in consequence there is no β-catenin translocation to the nucleus, so the transcription of genes, such as *SGK1* and *MYC*, cannot take place (Figure 2.9).

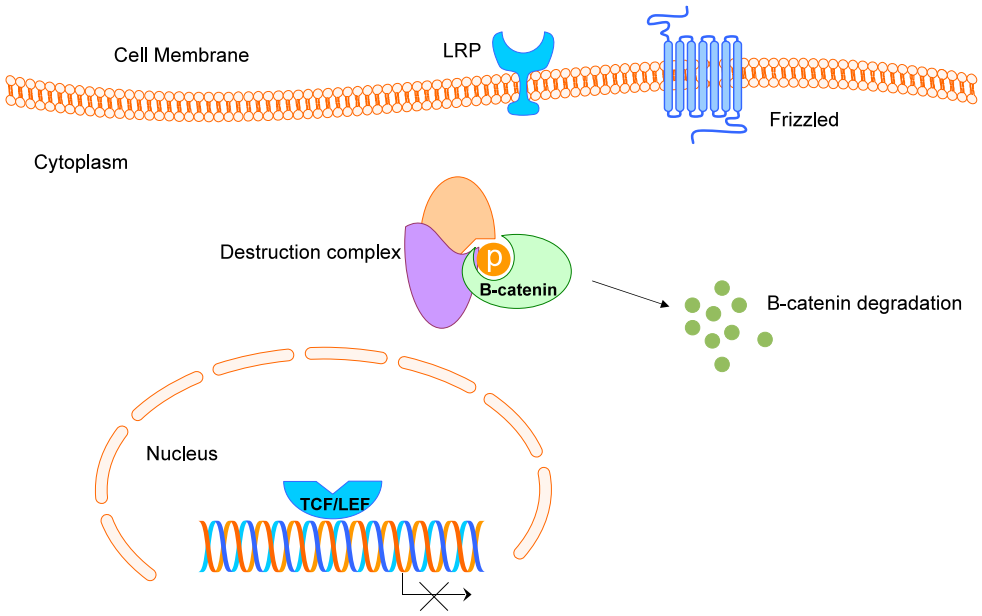

Figure 2.9: Schematic diagram of the Wnt signaling pathway in the absence of Wnt ligands. If there is deficiency of Wnt ligands, β-catenin is degraded via a destruction complex. Consistently, no transcription of its target genes is possible.

Normal situation in the presence of Wnt ligands

If on the cell surface Wnt ligands bind to LRP receptors, Frizzled is activated. The consequence of the latter is inhibition of the multiprotein destruction complex, leading to β-catenin accumulation in the cytoplasm. This protein abundance entails the translocation to the nucleus, where β-catenin binds to transcription factors (TCF/LEF) and initiates the transcription of particular genes [45], *MYC* and *SGK1*, among others (Figure 2.10).

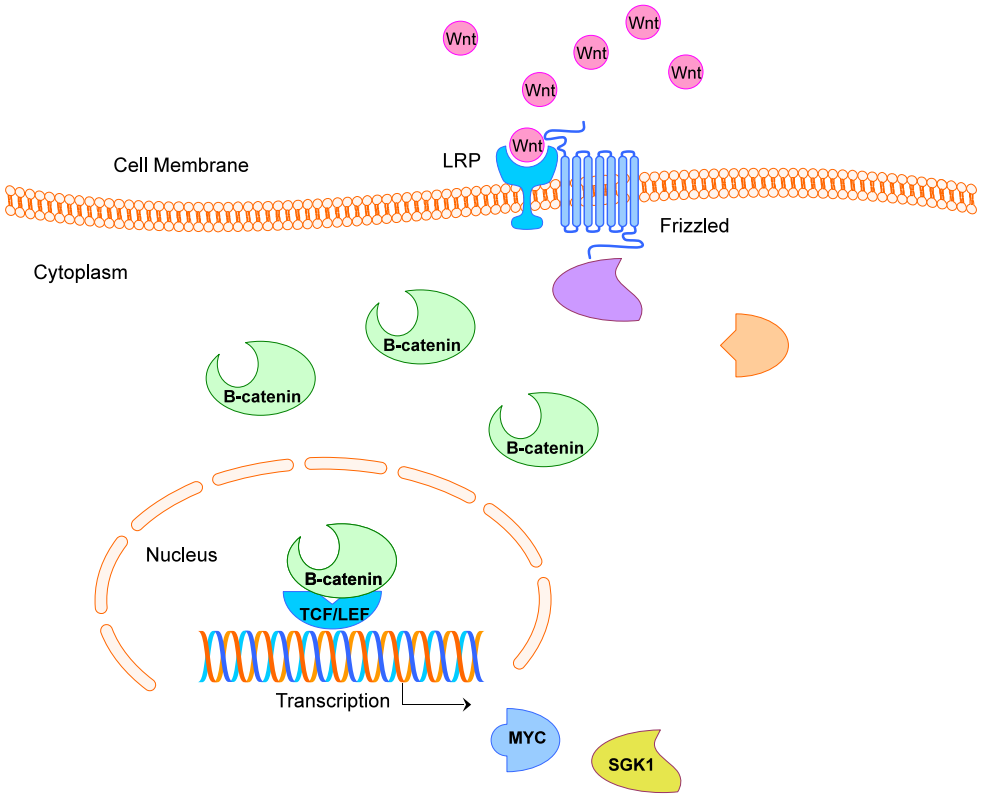

Figure 2.10: Schematic diagram of activated Wnt pathway. In the presence of Wnt ligands the destruction complex is destroyed by Frizzled and hence β-catenin is not phosphorylated, thus not degraded. The subsequent effect of β-catenin accumulation in the cytoplasm leads to its nuclear translocation. Then, the protein binds to TCF/LEF transcription family and initiates the transcription of several genes (e.g., *SGK1* and *MYC*).

Mutation in the Wnt pathway

The pathway is, however, not only modulated by the absence or presence of Wnt signaling. There are many mutations which give rise to an aberration (a mutation of any protein from the destruction complex), where the absence of Wnt ligands does not ensure the degradation of β-catenin. In medulloblastoma the mutation influences directly β-catenin, causing its resistance to degradation. The destruction complex cannot bind to the mutated β-catenin and therefore this protein is not degraded. Consecutive accumulation in the cytoplasm triggers β-catenin translocation to the nucleus, which results in abundant transcription of the downstream target genes of the pathway [22], [78], (Figure 2.11). This, together with β-catenin mutation, is always found in pediatric medulloblastoma carrying a monosomy 6. The corresponding medulloblastoma subtype is called WNT subgroup MBs.

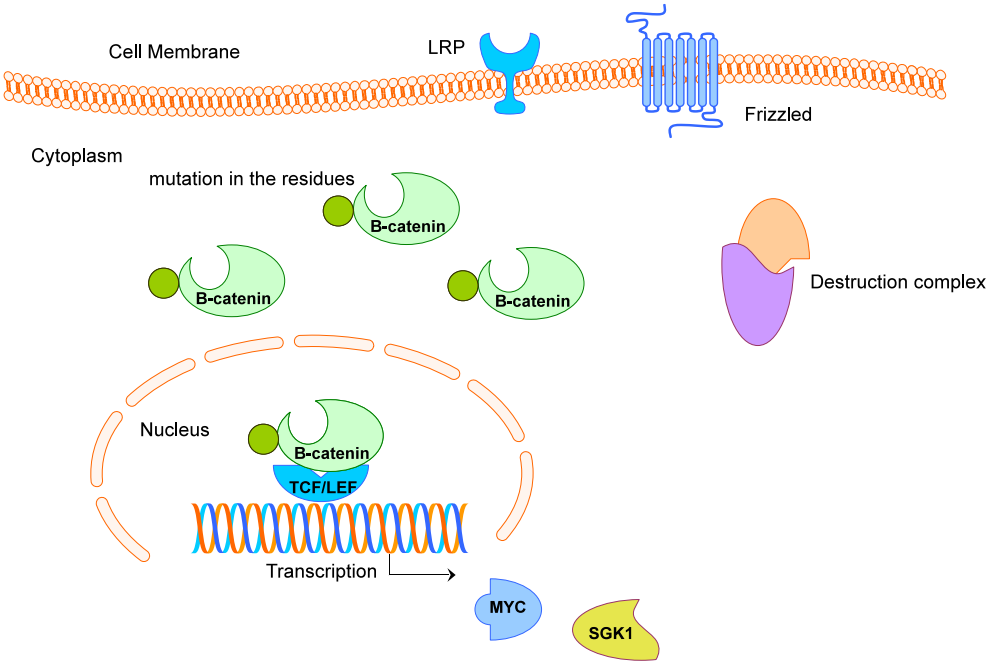

Figure 2.11: Schematic diagram of the Wnt pathway with mutated β-catenin. The absence of Wnt ligands should lead to β-catenin degradation due to activity of the destruction complex. However, a mutation on residues S33 and S37 (specific monomers of protein) of β-catenin blocks the destructive activity of the complex [57], [64]. β-catenin cannot be phosphorylated and for this reason is not marked for degradation. The increased stability of the protein implies the nuclear translocation and further transcription of *SGK1* and *MYC*.

2.2.2 Regulation of gene targets of the Wnt/β-catenin signaling pathway

In cells that are not deregulated, presence of β-catenin in the nucleus leads to the transcription of *MYC* [43] and *SGK1* [16] at normal level. Then the translation of their mRNA is observed in the cytoplasm. After translation each protein is transported to an appropriate cellular compartment depending on the role in the cell. The SGK1 protein remains in the cytoplasm if no additional signal is activated [21] and cMyc protein moves to the nucleus, where it functions as a transcription factor. SGK1 in the cytoplasm can bind to kinase GSK3β [18] and phosphorylate it, marking for degradation [3]. GSK3β is a protein that can shuttle between the cytoplasm and nucleus, and when it is in the cytoplasm, it can be degraded via an interaction with SGK1 [3], [77]. However, when GSK3β is shifted to the nucleus, it may bind to cMyc and phosphorylate this protein targeting it for degradation [26]. SGK1 and GSK3β return to the previous states after their phosphorylating activity, ready for new binding.

Since the transport between the cytoplasm and the nucleus is an essential process in the cell, it is necessary to take into consideration the transport rate as well as difference of the volumes of these two compartments. Additionally, some substrates may exist in different compartments of the cell and in the framework of ordinary differential equation modeling they have to be considered as separate variables. In Figure 2.12, we depict the modeled system.

Remark 2.1 *We use subscripts* t *and* p *to distinguish between the transcripts and the proteins, respectively. We do the same for the nuclear and cytoplasmic localization using subscripts* n *and* c*, respectively. When a protein binds to another protein, the process of phosphorylation takes place and then no other activity of these proteins is performed. For this reason, we label such state as* occupied*, denoting it by subscript* occ.

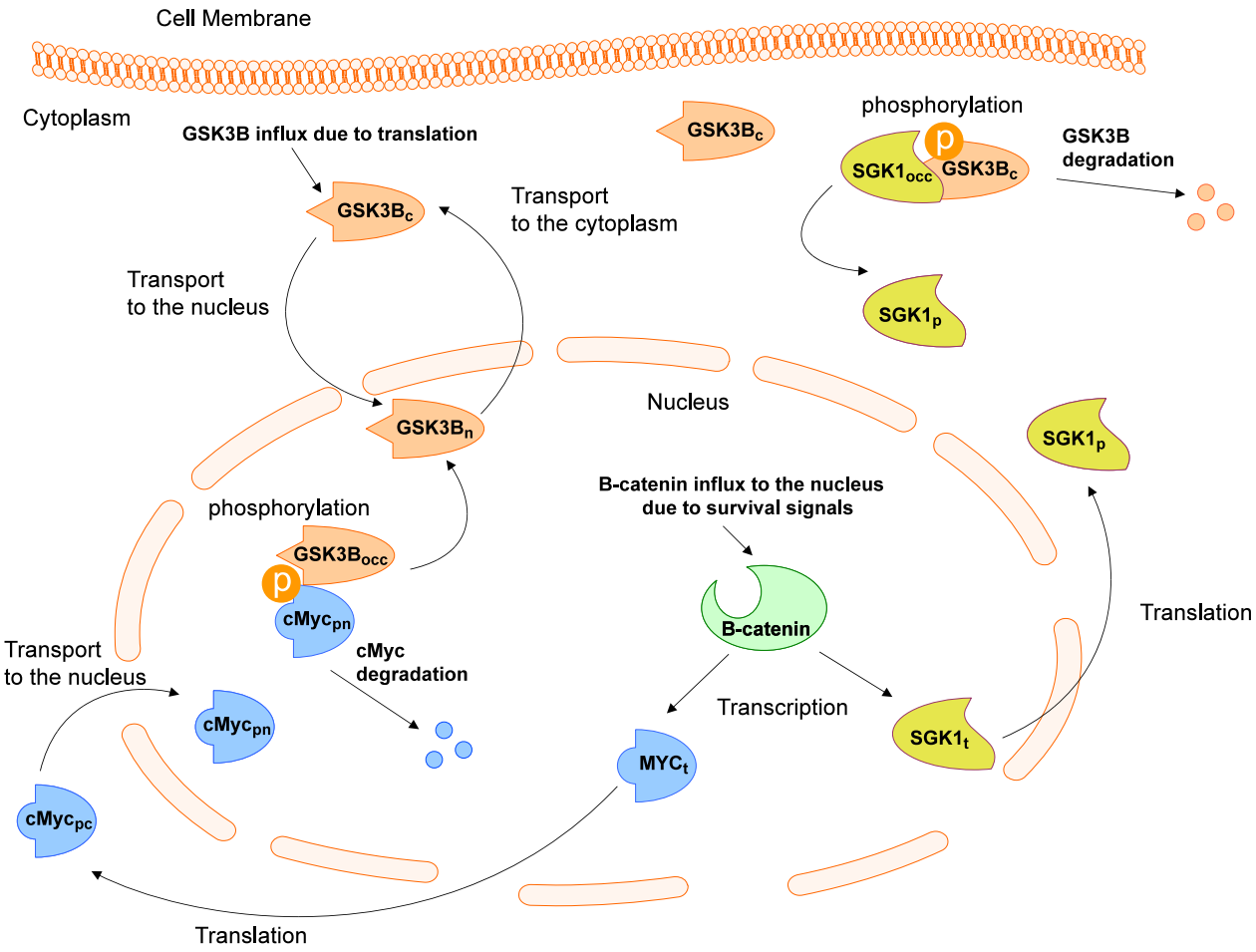

Figure 2.12: Schematic diagram of the modeled biological dynamics. Abbreviations:
$SGK1_t$ - concentration of the *SGK1* mRNA (transcript),
$SGK1_p$ - cytoplasmic concentration of the SGK1 protein,
MYC_t - concentration of the *MYC* mRNA (transcript),
$cMyc_{pc}$ - cytoplasmic concentration of the cMyc protein,
$cMyc_{pn}$ - nuclear concentration of the cMyc protein,
$GSK3\beta_c$ - cytoplasmic concentration of the GSK3β protein,
$GSK3\beta_n$ - nuclear concentration of the GSK3β protein,
$GSK3\beta_{occ}$ - phosphorylating GSK3β in a complex with cMyc,
$SGK1_{occ}$ - phosphorylating SGK1 in a complex with GSK3β.

Summary of Chapter 2

In this chapter we outlined the biological background that helps in understanding the basic processes in medulloblastoma. We presented medulloblastoma as a brain tumor, mostly found in children and diagnosed with different prognosis. To describe the phenomena found in medulloblastoma on the cellular level, we introduced a molecular overview of the signaling pathway, particularly the Wnt/β-catenin signaling pathway. The chosen key players compose an interesting system, which we want to model to investigate monosomy 6 and trisomy 6 subgroups of the medulloblastoma.

CHAPTER 3

Systems biology approach to understand the dynamics of medulloblastoma

In this chapter we develop a mathematical model of medulloblastoma. Our model is focused on a part of the Wnt signaling pathway, whose perturbations are crucial for the disease development and dynamics. We discuss biological phenomena involved in the modeled system. A systematic description of the underlying processes delivers a base to formulate autonomous ordinary differential equations (ODEs) of the interactions between particular molecules. We propose a model having the form of nine nonlinear ODEs. Our model reflects chemical kinetics of the reactants' concentrations in the nucleus-cytoplasm environment highlighting the discrepancy in prognosis of two types of medulloblastoma associated with a monosomy and trisomy of chromosome 6q.

3.1 Processes described by the model

We propose a model of the time dynamics of four major biochemical species involved in the system. We consider β-catenin, *SGK1*, *MYC* and *GSK3*β, accounting for the different stages in their synthesis (mRNA, protein), different intracellular location (nucleus, cytoplasm) and different biological processes, see Figure 2.12.

3.1.1 Transcription and translation

Production of *MYC* and *SGK1* is mainly regulated by β-catenin and members of TCF/LEF protein family. TCF/LEF, however, is present in abundance in comparison to other components of the system, therefore for sake of simplicity in our model we neglect its impact.

The main role for the intensity of the transcription of *MYC* and *SGK1* is credited to the amount of the concentration of β-catenin. We assume that both transcription rates are equal to t_1 and t_3, respectively, whereas translation is assigned by translation rates s_2 and s_4 for the *SGK1* mRNA and for the *MYC* mRNA, respectively

β-catenin $\xrightarrow{t_1}$ *SGK1* mRNA,
β-catenin $\xrightarrow{t_3}$ *MYC* mRNA,
SGK1 mRNA $\xrightarrow{s_2}$ SGK1 protein,
MYC mRNA $\xrightarrow{s_4}$ cMyc protein.

3.1.2 Spontaneous degradation

The spontaneous degradation of the *SGK1* mRNA and *MYC* mRNA is described by coefficients d_1 and d_3, respectively. Their proteins are degraded spontaneously at rates d_2 for SGK1 and d_4 for cMyc. The degradation of GSK3β is activated by various proteins (e.g., Dishevelled, PKB and SGK1 [77]). Neglecting details of the dynamics of Dishevelled and PKB, we model a spontaneous degradation of GSK3β protein as a linear process at a degradation rate d_o. Thus, we have

SGK1 mRNA $\xrightarrow{d_1}$ degraded *SGK1* mRNA,
MYC mRNA $\xrightarrow{d_3}$ degraded *MYC* mRNA,
SGK1 protein $\xrightarrow{d_2}$ degraded SGK1 protein,
cMyc protein $\xrightarrow{d_4}$ degraded cMyc protein,
GSK3β protein $\xrightarrow{d_o}$ degraded GSK3β protein.

3.1.3 Degradation through the interaction with proteins

We consider two degradation processes mediated by proteins in the investigated system. In particular, one of the cytoplasmic isoforms of SGK1 binds to GSK3β with binding rate d_6 and promotes its degradation via phosphorylation. Subsequent dissociation of the [SGK1·GSK3β] complex occurs at a rate p_9, and results in releasing SGK1 protein molecules that are active again. We do not go into details of the process of phosphorylated GSK3β degradation, so we assume that this process is immediate, taking place simultaneously with SGK1 releasing from the complex [3],

SGK1 + GSK3β $\xrightarrow{d_6}$ [SGK1·GSK3β],
[SGK1·GSK3β] $\xrightarrow{p_9}$ SGK1.

A scheme of these processes is presented in Figure 3.1.

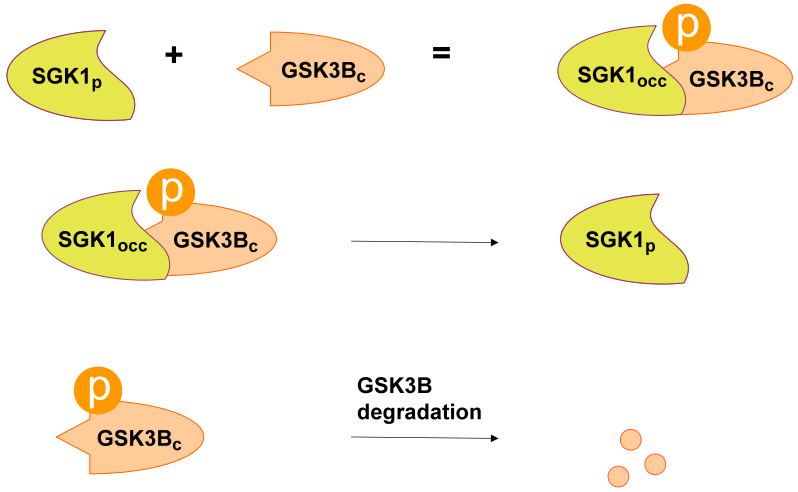

Figure 3.1: Steps in the process of GSK3β phosphorylation by SGK1.

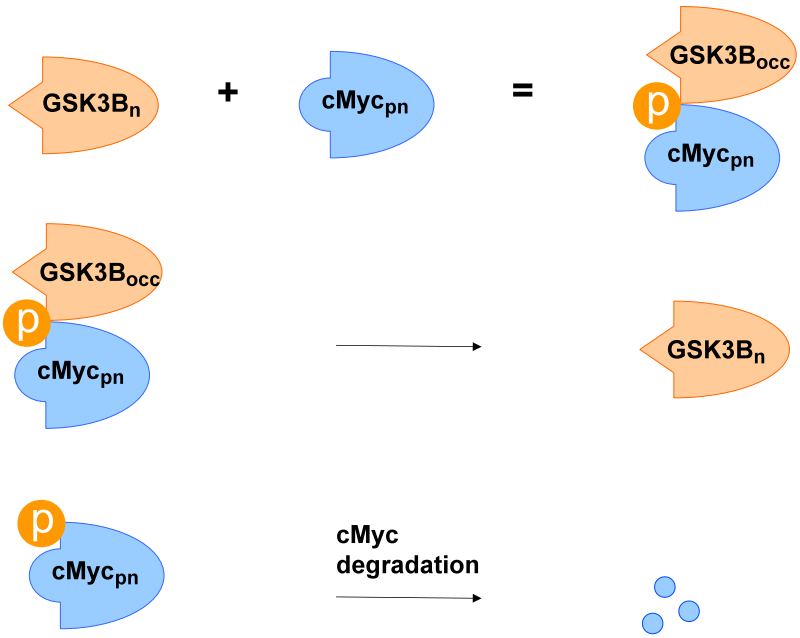

Figure 3.2: Steps in the process of cMyc phosphorylation by GSK3β.

Additionaly, in the nucleus GSK3β binds to cMyc, phosphorylates and leads to cMyc degradation, see Figure 3.2. It is important to note that after dissociation of the complex, GSK3β is ready for new binding and cMyc undergoes a degradation [23],

GSK3β + cMyc $\xrightarrow{d_5}$ [GSK3β·cMyc],
[GSK3β·cMyc] $\xrightarrow{p_8}$ GSK3β.

3.1.4 Transport between the nucleus and cytoplasm

When describing signaling pathway the architecture of the intracellular space can be treated as a two compartment structure consisting of the cell nucleus and the cell cytoplasm. We model transport between two compartments using scaling factor k_v which reflects the difference in volume between the nucleus and cytoplasm [41], [44].

Two compartment model of transport processes in cell

A volume of the substrate in the nucleus and cytoplasm is assigned by s_n and s_c, respectively. By the law of mass action the total amount of substrate $s_n + s_c$ is constant. Additionally, the speed of transport is proportional to the concentration of the transported substrate. Therefore, equations delineating the transport of the substrate concentrations in the nucleus and cytoplasm can be formulated in the form

$$\frac{\mathrm{d}s_c}{\mathrm{d}t} = K_n S_n - K_c S_c \qquad / : V_c, \tag{3.1}$$

$$\frac{\mathrm{d}s_n}{\mathrm{d}t} = K_c S_c - K_n S_n \qquad / : V_n, \tag{3.2}$$

where S_n and S_c are the nuclear and cytoplasmic substrate concentrations. K_n and K_c are coefficients of the cytoplasmic and nuclear transport, respectively. Dividing equation (3.1) by volume of the cytoplasm V_c and equation (3.2) by volume of the nucleus V_n results in

$$\frac{\mathrm{d}S_c}{\mathrm{d}t} = k_n S_n - k_c S_c, \tag{3.3}$$

$$\frac{\mathrm{d}S_n}{\mathrm{d}t} = k_v k_c S_c - k_v k_n S_n, \tag{3.4}$$

where $k_c = K_c/V_c$ and $k_n = K_n/V_c$ are scaled transport coefficients, and $k_v = V_c/V_n$ is a scaling factor.

k_v factor is used in the following equations to scale the transport rate between the nucleus and cytoplasm. It reflects differences between the size of the nucleus and cyto-

plasm. Since cancer cells are characterized by abnormally large nuclei the coefficient k_v changes in cells during the neoplastic transformation. Furthermore, transport coefficients are different for different proteins as well as for different movement directions. GSK3β import to the nucleus is assigned by the rate c_6 and its export by the rate c_7. cMyc import rate to the nucleus is ascribed by c_4,

cytoplasmic cMyc $\xrightarrow{k_v c_4}$ nuclear cMyc,
cytoplasmic GSK3β $\xrightarrow{k_v c_6}$ nuclear GSK3β,
nuclear GSK3β $\xrightarrow{c_7}$ cytoplasmic GSK3β.

3.2 Formulation of the mathematical model

Mathematical approach to describe biochemical phenomena gives a great opportunity to analyze biological system without laborious and cost consuming experiments.

Based on the law of mass action, introduced by Guldberg and Waage in 1864, we formulate a system of ODEs describing the biochemical interactions. The law states that the rate of any given chemical reaction at which substance reacts is proportional to the product of molar concentrations of the reactants, which are called *active mass* [46]. Additionally, velocity of chemical reactions is proportional to the product of reacting substances and to a constant which characterizes each reaction. Taking into account the biochemical kinetics described in Section 3.1, we obtain a system, which describe evolution in time of the investigated substances.

3.2.1 From biology to equations

The ODEs represent biological processes, which appear on the intracellular level. To keep the biological meaning of the particular species in this section, we define each model variable using its biological nomenclature. Each equation represents the time change of a variable, which is the concentration of one biochemical species. All variables are functions of time.

The first equation (gene activation equation) models the dynamics of *SGK1* transcript,

$$\frac{\mathrm{d}SGK1_t(t)}{\mathrm{d}t} = t_1\beta cat - d_1 SGK1_t(t). \tag{3.5}$$

The production of *SGK1* is described by the first term of the right hand side (r.h.s.) in

(3.5). The spontaneous degradation of the *SGK1* transcript is modeled by the second term on the r.h.s. We assume β-catenin to be constant, because in a system without mutation we observe its constant influx on the normal level [76]. Modeling of β-catenin mutation is described in Chapter 5.

The dynamics of the SGK1 protein in the cytoplasm is described by four terms,

$$\frac{\mathsf{d}SGK1_p(t)}{\mathsf{d}t} = s_2 SGK1_t(t) - d_2 SGK1_p(t) - d_6 SGK1_p(t) \cdot GSK3\beta_c(t) + p_9 SGK1_{occ}(t). \tag{3.6}$$

The first term of the r.h.s. of (3.6) stands for the translation (protein synthesis). We assume only a spontaneous degradation of SGK1 that is expressed by the next term. Binding of SGK1 to GSK3β is modeled using a bilinear term $d_6 SGK1_p(t) \cdot GSK3\beta_c(t)$ and it promotes GSK3β degradation via phosphorylation. Protein SGK1 from this complex is occupied and cannot bind to another protein. The last term regards dissociation of the [SGK1·GSK3β] complex and results in releasing the SGK1 molecules after phosphorylating GSK3β. Released SGK1 is freed from the complex and it becomes active again.

We model the *MYC* mRNA in the same way as the *SGK1* mRNA. Therefore, we have transcription of *MYC* and spontaneous degradation terms,

$$\frac{\mathsf{d}MYC_t(t)}{\mathsf{d}t} = t_3 \beta cat - d_3 MYC_t(t). \tag{3.7}$$

The dynamics of the cMyc protein in the cytoplasm is described by

$$\frac{\mathsf{d}cMyc_{pc}(t)}{\mathsf{d}t} = s_4 MYC_t(t) - d_4 cMyc_{pc}(t) - c_4 cMyc_{pc}(t). \tag{3.8}$$

The first term corresponds to the cMyc synthesis and the second one to spontaneous degradation. The expression $c_4 cMyc_{pc}(t)$ stands for the process of the transport between the nucleus and cytoplasm.

Dynamics of the cMyc protein in the nucleus consists of two terms. We model here transport from the cytoplasm using the scaling coefficient k_v, and then the degradation of cMyc, which occurs when GSK3β binds to this molecule,

$$\frac{\mathsf{d}cMyc_{pn}(t)}{\mathsf{d}t} = k_v c_4 cMyc_{pc}(t) - d_5 cMyc_{pn}(t) \cdot GSK3\beta_n(t). \tag{3.9}$$

Now, we write the equation for the cytoplasmic GSK3β as

$$\begin{aligned}\frac{\mathsf{d}GSK3\beta_c(t)}{\mathsf{d}t} &= P_{GSK} + c_7 GSK3\beta_n(t) - c_6 GSK3\beta_c(t) - d_6 SGK1_p(t)\cdot GSK3\beta_c(t)\\ &- d_o GSK3\beta_c(t). \end{aligned} \tag{3.10}$$

The first term of the r.h.s. of (3.10), P_{GSK}, corresponds to the influx of GSK3β protein due to translation. The second and third term describe import from the nucleus, $c_7 GSK3\beta_n(t)$, and transport to the nucleus, $c_6 GSK3\beta_c(t)$, respectively. The last two terms stand for the degradation of GSK3β. Here, SGK1 binds to GSK3β and promotes its degradation in the phosphorylation process, whereas the spontaneous degradation proceeds at the rate d_o.

Next, we model the dynamics of the nuclear GSK3β as

$$\begin{aligned}\frac{\mathsf{d}GSK3\beta_n(t)}{\mathsf{d}t} &= k_v c_6 GSK3\beta_c(t) - k_v c_7 GSK3\beta_n(t) - d_5 GSK3\beta_n(t)\cdot cMyc_{pn}(t)\\ &+ p_8 GSK3\beta_{occ}(t). \end{aligned} \tag{3.11}$$

The first two terms determine the import to the nucleus and the export to the cytoplasm. The third term stands for the loss of GSK3β that binds to cMyc and is involved in the process of phosphorylation. The last term is responsible for GSK3β, which is freed after phosphorylation. The process of phosphorylation in the complex $[GSK3\beta_n \cdot cMyc_{pn}]$ is modeled analogously to the phosphorylation in the complex $[SGK1_p \cdot GSK3\beta_c]$.

Considering the process of phosphorylation as time consuming, we introduce an additional variable to model it. To distinguish the states in which the proteins are ready for new binding and from those which are still in the complex, we use an additional variable for the latter, i.e., *occupied* GSK3β_{occ}, which is described by an ODE with two terms in the r.h.s.

$$\frac{\mathsf{d}GSK3\beta_{occ}(t)}{\mathsf{d}t} = d_5 GSK3\beta_n(t)\cdot cMyc_{pn}(t) - p_8 GSK3\beta_{occ}(t). \tag{3.12}$$

The first term corresponds to the phosphorylation process, where GSK3β phosphorylates cMyc, and thus initiates its degradation. The second term indicates the loss of bound GSK3β that is already dissociated from cMyc with its phoshorylating activity restored.

We also consider occupied SGK1 as

$$\frac{\mathrm{d}SGK1_{occ}(t)}{\mathrm{d}t} = d_6 SGK1_p(t) \cdot GSK3\beta_c(t) - p_9 SGK1_{occ}(t). \tag{3.13}$$

The first term stands for the complex arising, where SGK1 phosphorylates GSK3β and the second term is the loss of SGK1 that becomes active again.

3.2.2 System of ordinary differential equations

Rewriting equations (3.5) - (3.13) in the terms of new variables $u = [u_1, ..., u_9]^T$ describing each biological species and taking $k_1 = t_1\beta cat, k_3 = t_3\beta cat, P = P_{GSK}$ lead to the following system of equations.

Ordinary differential equations model of the medulloblastoma signaling

$$\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}t} &= k_1 - d_1 u_1, && (3.14)\\
\frac{\mathrm{d}u_2}{\mathrm{d}t} &= s_2 u_1 - d_2 u_2 - d_6 u_2 u_6 + p_9 u_9, && (3.15)\\
\frac{\mathrm{d}u_3}{\mathrm{d}t} &= k_3 - d_3 u_3, && (3.16)\\
\frac{\mathrm{d}u_4}{\mathrm{d}t} &= s_4 u_3 - d_4 u_4 - c_4 u_4, && (3.17)\\
\frac{\mathrm{d}u_5}{\mathrm{d}t} &= k_v c_4 u_4 - d_5 u_5 u_7, && (3.18)\\
\frac{\mathrm{d}u_6}{\mathrm{d}t} &= P + c_7 u_7 - c_6 u_6 - d_6 u_2 u_6 - d_o u_6, && (3.19)\\
\frac{\mathrm{d}u_7}{\mathrm{d}t} &= k_v c_6 u_6 - k_v c_7 u_7 - d_5 u_5 u_7 + p_8 u_8, && (3.20)\\
\frac{\mathrm{d}u_8}{\mathrm{d}t} &= d_5 u_5 u_7 - p_8 u_8, && (3.21)\\
\frac{\mathrm{d}u_9}{\mathrm{d}t} &= d_6 u_2 u_6 - p_9 u_9, && (3.22)
\end{aligned}$$

with initial condition $u_i(0) = u_{io}$, where $u_{io} \geq 0 \;\; i = 1, ..., 9$.
We assume positivity of parameters: $c_i, d_i, k_i, p_i, s_i, k_v, P > 0 \;\; \forall\, i$.

Variable notations and their biological meanings are listed in Table 3.1.

Key players	Related model variables	Biological meaning
$SGK1_t$	u_1	*SGK1* transcript in the nucleus
$SGK1_p$	u_2	SGK1 protein in the cytoplasm
MYC_t	u_3	*MYC* transcript in the nucleus
$cMyc_{pc}$	u_4	cMyc protein in the cytoplasm
$cMyc_{pn}$	u_5	cMyc protein in the nucleus
$GSK3\beta_c$	u_6	GSK3β protein in the cytoplasm
$GSK3\beta_n$	u_7	GSK3β protein in the nucleus
$GSK3\beta_{occ}$	u_8	phosphorylating GSK3β protein in the nucleus
$SGK1_{occ}$	u_9	phosphorylating SGK1 protein in the cytoplasm

Table 3.1: Description of model 3.14 - 3.22 variables. Biological nomenclature and their equivalent mathematical notation.

Summary of Chapter 3

In this chapter we formulated the ODE system describing the dynamics of medulloblastoma signaling found in the literature and discussed with biologists. To indicate the ratio of the compartments volume of the nucleus and cytoplasm, we introduced scaling coefficient k_v, an important factor while formulating the two-compartment model.

CHAPTER 4

Mathematical analysis of the medulloblastoma model

In this chapter we present an analysis of the ODEs model of the dynamics of medulloblastoma signaling. We show that the model is well-posed.

Model well-posedness

1. Global existence: there exists a solution of $u(t)$ for $t \in [0, \infty)$.
2. Uniqueness: the solution of the considered problem is unique.
3. Nonnegativity: the solution is nonnegative for nonnegative initial data.

The proposed model (3.14) - (3.22) is a system of autonomous ODEs

$$\frac{\mathrm{d}u}{\mathrm{d}t} = F(u), \tag{4.1}$$

with initial condition $u(0) = u_o$. $u := [u_1, ..., u_n]^T$, $F := [f_1, ..., f_n]^T$, $F : G \longrightarrow \mathbb{R}^n$ is Lipschitz continuous and G is an open subset of $\mathbb{R}^n$, $n = 9$. f_i stands for the r.h.s. of the i-th equation of (3.14) - (3.22). The solutions of this Cauchy problem describe dynamics of each reactant concentration.

4.1 Existence and uniqueness of global solution

Theorem 4.1 *System* (3.14) - (3.22) *with nonnegative initial data has a global unique solution* $u(t), t \geq 0$.

Proof 4.1

- *Local existence and uniqueness*

  Since the right hand side of the system F(u) is a C^1 *function, it is also a locally Lipschitz-continuous function. Then, we apply Picard-Lindelöf Theorem [29] and obtain local existence and uniqueness of the solution. The classical theory of ODEs provides also smoothness of the solution and its continuous dependence on the initial data and model parameters [29].*

- *Nonnegativity*

  To show nonnegativity of the solution, we check that

$$\frac{\mathrm{d}u_i}{\mathrm{d}t}|_{u_i=0} \geq 0 \qquad \forall\, i, i = 1, ..., 9 \tag{4.2}$$

  holds. Then, the set $\mathbb{R}^9_+ \cup \{0\}$ *is positive invariant with respect to the generated flow of the ODE system. Indeed, since the model parameters are positive (see Section 5.1.2), it holds:*

$$\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}t}|_{u_1=0} &= k_1 \geq 0, && (4.3)\\
\frac{\mathrm{d}u_2}{\mathrm{d}t}|_{u_2=0} &= s_2u_1 + p_9u_9 \geq 0, && (4.4)\\
\frac{\mathrm{d}u_3}{\mathrm{d}t}|_{u_3=0} &= k_3 \geq 0, && (4.5)\\
\frac{\mathrm{d}u_4}{\mathrm{d}t}|_{u_4=0} &= s_4u_3 \geq 0, && (4.6)\\
\frac{\mathrm{d}u_5}{\mathrm{d}t}|_{u_5=0} &= k_vc_4u_4 \geq 0, && (4.7)\\
\frac{\mathrm{d}u_6}{\mathrm{d}t}|_{u_6=0} &= P + c_7u_7 \geq 0, && (4.8)\\
\frac{\mathrm{d}u_7}{\mathrm{d}t}|_{u_7=0} &= k_vc_6u_6 + p_8u_8 \geq 0, && (4.9)\\
\frac{\mathrm{d}u_8}{\mathrm{d}t}|_{u_8=0} &= d_5u_5u_7 \geq 0, && (4.10)\\
\frac{\mathrm{d}u_9}{\mathrm{d}t}|_{u_9=0} &= d_6u_2u_6 \geq 0. && (4.11)
\end{aligned}$$

Using continuity of the solution and nonnegativity of the initial data, i.e., $u_o \geq 0$*, we conclude that solutions cannot leave* $\mathbb{R}^9_+ \cup \{0\}$.

- *Global existence*

Multiplying equations by properly chosen positive constants and adding them leads to

$$\frac{\mathsf{d}}{\mathsf{d}t}\left(\frac{s_2}{d_1}u_1+u_2+\frac{s_4}{d_3}u_3+u_4+\frac{u_5}{k_v}+u_6+\frac{u_7}{k_v}+\frac{u_8}{k_v}+u_9\right) = P+\frac{k_1 s_2}{d_1}+\frac{k_3 s_4}{d_3}-d_4u_4-d_2u_2 - d_0u_6 - \frac{d_5}{k_v}u_5u_7 - d_6u_2u_6. \quad (4.12)$$

Due to the nonnegativity of the solution, we obtain

$$\frac{\mathsf{d}}{\mathsf{d}t}\left(\frac{s_2}{d_1}u_1 + u_2 + \frac{s_4}{d_3}u_3 + u_4 + \frac{u_5}{k_v} + u_6 + \frac{u_7}{k_v} + \frac{u_8}{k_v} + u_9\right) \leq K, \quad (4.13)$$

where K is a positive constant. Integrating both sides of (4.13)*, we obtain the estimate*

$$\frac{s_2}{d_1}u_1 + u_2 + \frac{s_4}{d_3}u_3 + u_4 + \frac{u_5}{k_v} + u_6 + \frac{u_7}{k_v} + \frac{u_8}{k_v} + u_9 \leq Kt + C, C \in \mathbb{R}. \quad (4.14)$$

Consequently, we obtain that linear combination of the model variables is bounded for all finite times. Since the solutions are nonnegative, we conclude that they are bounded. Hence, the solutions of the model can be extended globally in time for $t \in [0, \infty)$. ■

Observation 4.1 *Numerical simulations suggest that the solutions are also uniformly bounded (for simulations, see next chapter).*

4.2 Asymptotic behavior of model solutions

In this section we determine the stability of the steady states of the ODE system applying linearization [10]. We apply the Hartman-Grobman Theorem [10], which justifies the use of linearization in the study of stability of nonlinear dynamical system.

4.2.1 Steady States

Definition 4.1 *A solution* $\bar{u}$ *with* $F(\bar{u}) = 0$ *is called a steady state solution or an equilibrium of* (4.1).

To study the stability of equilibria of the model, we calculate steady state solution $\bar{u} = [\bar{u}_1, ..., \bar{u}_9]$ of (4.1). This corresponds to the homeostasis in the biological system. Thus, we set

$$F(\bar{u}) = 0. \tag{4.15}$$

Theorem 4.2 *The model* (3.14) - (3.22) *has a unique positive steady state.*

Proof 4.2 *By solving the corresponding algebraic equations of* (4.15)*, we obtain a unique positive equilibrium given by*

Steady state of the model

$$\bar{u}_1 = \frac{k_1}{d_1}, \tag{4.16}$$

$$\bar{u}_2 = \frac{k_1 s_2}{d_1 d_2}, \tag{4.17}$$

$$\bar{u}_3 = \frac{k_3}{d_3}, \tag{4.18}$$

$$\bar{u}_4 = \frac{s_4}{d_4 + c_4}\frac{k_3}{d_3}, \tag{4.19}$$

$$\bar{u}_5 = \frac{s_4}{d_4 + c_4}\frac{k_3}{d_3}\frac{k_v c_4 c_7}{d_3 d_5 c_6 P}\left(\frac{k_1 s_2 d_6}{d_1 d_2} + d_o\right), \tag{4.20}$$

$$\bar{u}_6 = \frac{P d_1 d_2}{k_1 s_2 d_6 + d_o d_1 d_2}, \tag{4.21}$$

$$\bar{u}_7 = \frac{c_6}{c_7}\frac{P d_1 d_2}{k_1 s_2 d_6 + d_o d_1 d_2}, \tag{4.22}$$

$$\bar{u}_8 = \frac{k_v c_4}{p_8}\frac{s_4}{d_4 + c_4}\frac{k_3}{d_3}, \tag{4.23}$$

$$\bar{u}_9 = \frac{d_6}{p_9}\frac{k_1 s_2}{d_1 d_2}\frac{P d_1 d_2}{k_1 s_2 d_6 + d_o d_1 d_2}. \tag{4.24}$$

■

4.2.2 Stability of ordinary differential system

Definition 4.2 *An equilibrium $\bar{u}$ of* (4.1) *is said to be stable if for each $\epsilon > 0$, there exists a $\delta > 0$ such that*

$$|u(t) - \bar{u}| < \epsilon \qquad for\ all\ \ t \geq 0, \qquad \textit{whenever}\ |u(0) - \bar{u}| < \delta. \tag{4.25}$$

Definition 4.3 *An equilibrium $\bar{u}$ of* (4.1) *is said to be asymptotically stable if it is stable and there exists a > 0 such that*

$$\lim_{t\to\infty} |u(t) - \bar{u}| = 0, \qquad \textit{whenever}\ |u(0) - \bar{u}| < a. \tag{4.26}$$

Intuitively, Definition 4.3 means that an equilibrium $\bar{u}$ of (4.1) is asymptotically stable if the trajectories starting near $\bar{u}$ approach $\bar{u}$ for $t \to \infty$. If $\bar{u}$ is not $stable$, it is said to be *unstable*.

To determine stability of $\bar{u}$, we use the method of linearization.

4.2.3 Linearization of the medulloblastoma model

Definition 4.4 *The linearization of system* (3.14) - (3.22) *at the equilibrium $\bar{u}$ is given by $\frac{\mathrm{d}u}{\mathrm{d}t} = \mathbf{J}(u - \bar{u})$, where $\mathbf{J} = DF(\bar{u})$ is the Jacobian matrix of function F at $\bar{u}$.*

The Jacobian $\mathbf{J}$ of system (3.14) - (3.22) is of the form:

$$\mathbf{J} = \begin{pmatrix} -d_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ s_2 & -d_2 - d_6\bar{u}_6 & 0 & 0 & 0 & -d_6\bar{u}_2 & 0 & 0 & p_9 \\ 0 & 0 & -d_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & s_4 & -d_4 - c_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k_v c_4 & -d_5\bar{u}_7 & 0 & -d_5\bar{u}_5 & 0 & 0 \\ 0 & -d_6\bar{u}_6 & 0 & 0 & 0 & -c_6 - d_6\bar{u}_2 - d_0 & c_7 & 0 & 0 \\ 0 & 0 & 0 & 0 & -d_5\bar{u}_7 & k_v c_6 & -k_v c_7 - d_5\bar{u}_5 & p_8 & 0 \\ 0 & 0 & 0 & 0 & d_5\bar{u}_7 & 0 & d_5\bar{u}_5 & -p_8 & 0 \\ 0 & d_6\bar{u}_6 & 0 & 0 & 0 & d_6\bar{u}_2 & 0 & 0 & -p_9 \end{pmatrix} \tag{4.27}$$

To study the behavior of the linearized system in the neighborhood of $\bar{u}$, we need to assure that the behavior of the linearized system at sufficiently small neighborhood of $\bar{u}$ is topologically equivalent to the dynamics of nonlinear system. We apply classical theory given by

Theorem 4.3 *(Grobman-Hartman [10]) If $\bar{u}$ is a hyperbolic equilibrium of* (4.1)*, then there exists a homeomorphism H such that the orbits of* (4.1) *in the region U of $\bar{u}$ are mapped by H to orbits of the linearized system, i.e.,*

$$\frac{\mathsf{d}u}{\mathsf{d}t} = DF(\bar{u})(u - \bar{u}) \tag{4.28}$$

for $u \in U$.

Definition 4.5 *An equilibrium $\bar{u}$ is called a hyperbolic equilibrium of $\frac{\mathsf{d}u}{\mathsf{d}t} = F(u)$ if all eigenvalues $\lambda_{i,i=1,\ldots,9}$ of the Jacobian matrix $\mathbf{J} = DF(\bar{u})$ have nonzero real part.*

Theorem 4.4 *If $\bar{u}$ is a hyperbolic equilibrium of $\frac{\mathsf{d}u}{\mathsf{d}t} = F(u)$ and if all eigenvalues of the linear transformation $DF(\bar{u})$ have negative real parts, then $\bar{u}$ is asymptotically stable [10].*

Since the system of nine equations is not treatable analytically, we use numerical calculations of the eigenvalues of **J** for different set of parameters.

We study (3.14) - (3.22) for a special choice of parameters obtained based on the available data (see Chapter 5 for details). We obtain the following steady states for a healthy and trisomy 6 sample:

Steady states	
Healthy sample [μM]	**Trisomy 6 sample [μM]**
$\bar{u}_1$ = 0.0001143,	$\bar{u}_1$ = 0.00071,
$\bar{u}_2$ = 0.057,	$\bar{u}_2$ = 0.35,
$\bar{u}_3$ = 0.000266,	$\bar{u}_3$ = 0.009,
$\bar{u}_4$ = 0.049,	$\bar{u}_4$ = 1.55,
$\bar{u}_5$ = 0.1,	$\bar{u}_5$ = 4.92,
$\bar{u}_6$ = 0.0012,	$\bar{u}_6$ = 0.0002,
$\bar{u}_7$ = 0.0039,	$\bar{u}_7$ = 0.00063,
$\bar{u}_8$ = 0.000489,	$\bar{u}_8$ = 0.0039,
$\bar{u}_9$ = 0.0000332,	$\bar{u}_9$ = 0.00003,

where in healthy sample - *SGK1* mRNA: 1, *MYC* mRNA: 1, k_v: 2 and in the trisomy 6 sample - *SGK1* mRNA: 6.19, *MYC* mRNA: 31.86, k_v: 0.5 (for details, see Subsection 5.1.1). The eigenvalues of **J** (4.27) for the healthy and trisomy 6 samples are:

Eigenvalues	
Healthy sample [s^{-1}]	**Trisomy 6 sample [s^{-1}]**
λ_1 = -0.4504,	λ_1 = -2.8606,
λ_2 = -0.6004,	λ_2 = -0.6001,
λ_3 = -0.0204,	λ_3 = -0.1092,
λ_4 = -0.0015 + 0.0005i,	λ_4 = -5e-05 + 0.0001i,
λ_5 = -0.0015 - 0.0005i,	λ_5 = -5e-05 - 0.0001i,
λ_6 = -0.0004,	λ_6 = -0.0004,
λ_7 = -0.0007,	λ_7 = -0.0007,
λ_8 = -0.0022,	λ_8 = -0.0022,
λ_9 = -0.0006,	λ_9 = -0.0006.

We remark that all eigenvalues $\lambda_{i,i=1,\dots,9}$ have negative real parts. Next, we perform simulations for a large range of parameters and we obtain that calculated eigenvalues have negative real parts too. From Theorem 4.3 and 4.4, we conclude that the nontrivial equilibrium of (3.14) - (3.22) is asymptotically stable. Additionally, we classify $\bar{u}$ as a focus as there exist λ_i with $\mathfrak{Im}(\lambda_i) \neq 0$, [72].

For a good representation of solution dynamics we plot solutions u_5 and u_8, where we can easily see equilibrium to be a focus for a trisomy 6 sample. Figure 4.1 presents numerical simulations of the solutions u_5 and u_8 for a healthy sample and Figure 4.2 for a trisomy 6 sample.

Observation 4.2 *Numerical simulations suggest that the system is also globally asymptotically stable as we start the solution from various initial conditions and we finish at steady states (see Figure 4.3, Figure 4.4 and Figure 4.5).*

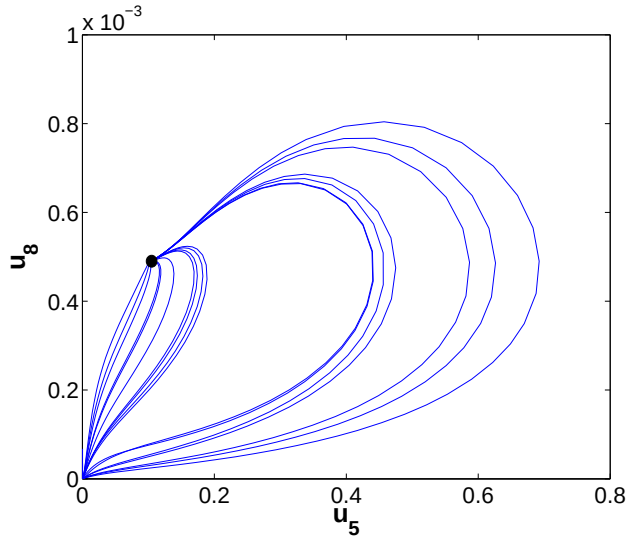

Figure 4.1: Phase portrait with marked equilibrium (black dot) for the state variables u_5 and u_8 of several different sets of initial conditions (healthy sample - *SGK1* mRNA: 1, *MYC* mRNA: 1, k_v: 2, see Chapter 5). Due to eigenvalue analysis, the critical point is classified as focus.

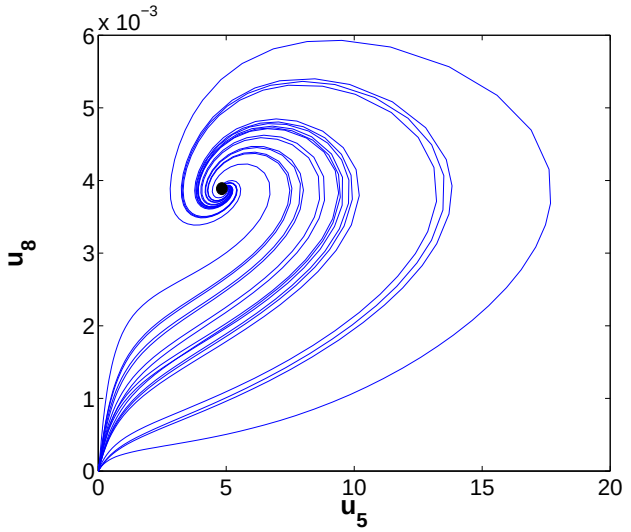

Figure 4.2: Phase portrait with marked equilibrium (black dot) for the state variables u_5 and u_8 of several different sets of initial conditions (trisomy 6 sample - *SGK1* mRNA: 6.19, *MYC* mRNA: 31.86, k_v: 0.5, see Chapter 5). Due to eigenvalue analysis, the critical point is classified as focus.

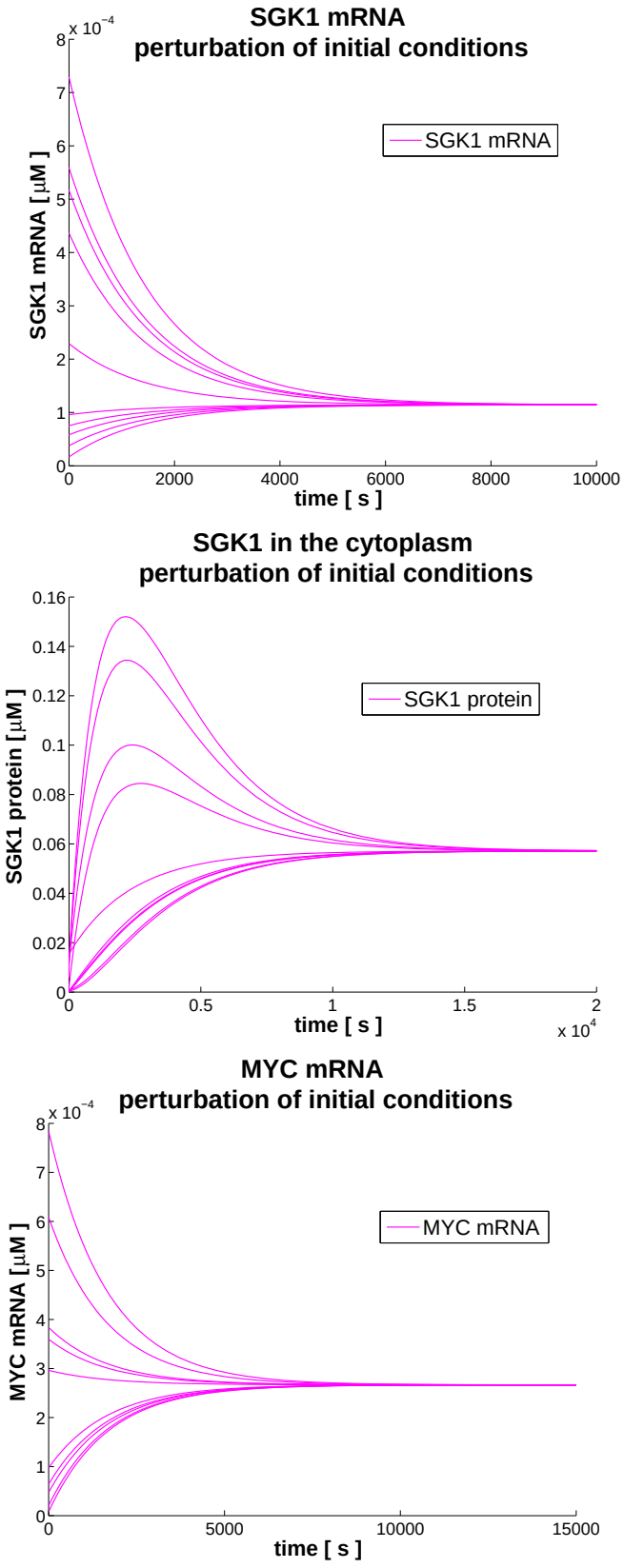

Figure 4.3: Perturbation of initial conditions for *SGK1* mRNA, cytoplasmic SGK1 and *MYC* mRNA for the healthy sample. Simulations show convergence of solutions to the equilibrium.

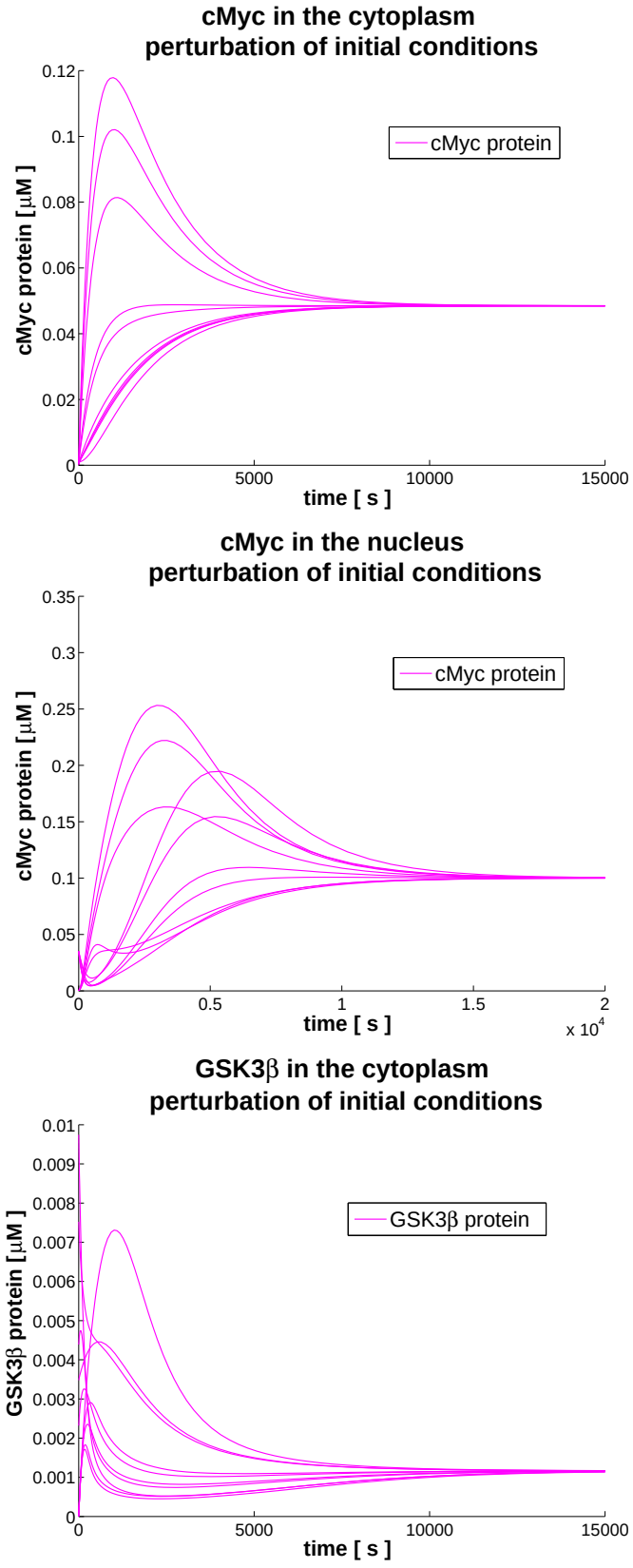

Figure 4.4: Perturbation of initial conditions for cytoplasmic cMyc, nuclear cMyc and cytoplasmic GSK3β for the healthy sample. Simulations show convergence of solutions to the equilibrium.

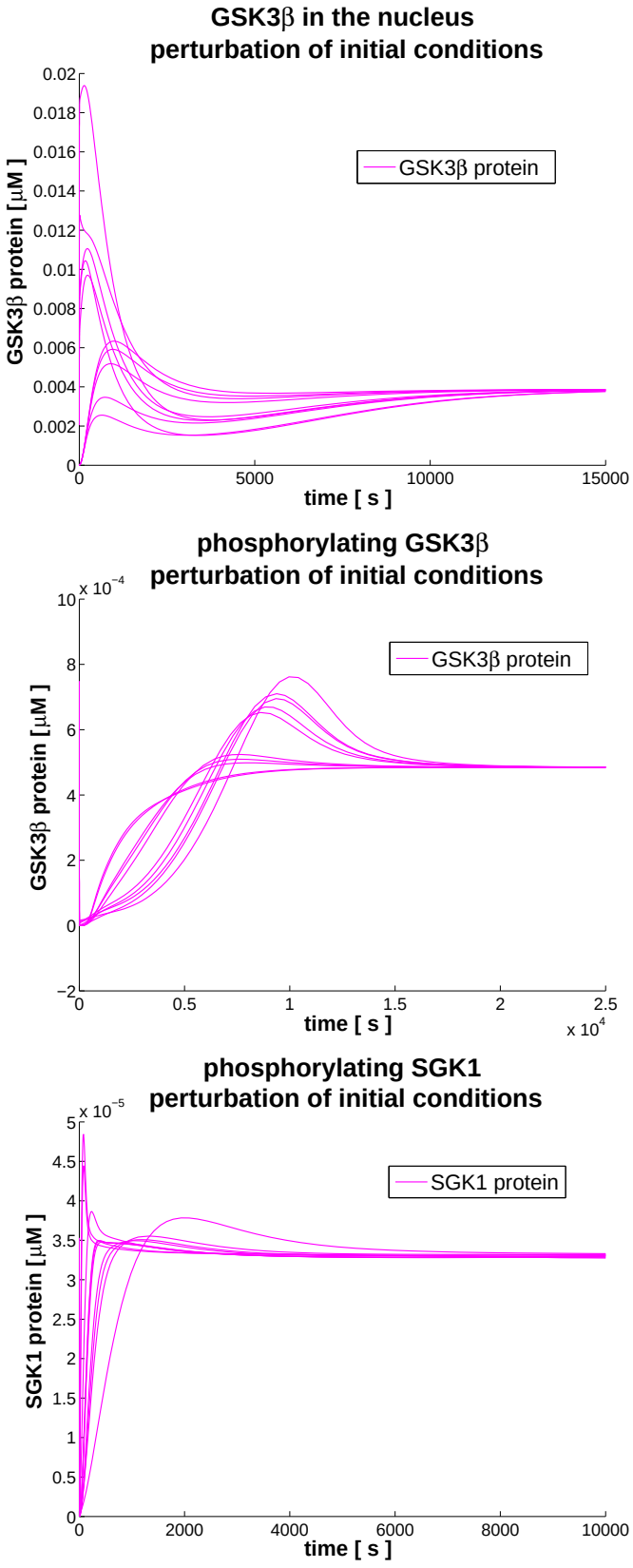

Figure 4.5: Perturbation of initial conditions for nuclear GSK3β, phosphorylating GSK3β and phosphorylating SGK1 for the healthy sample. Simulations show convergence of solutions to the equilibrium.

Summary of Chapter 4

In this chapter we proved existence, uniqueness and nonnegativity of the solutions of the system (3.14) - (3.22). We provided also estimates assuring the global existence. To analyze the stability of the system (3.14) - (3.22), we linearized it and performed stability analysis supported by numerical calculations. We found the model to be locally asymptotically stable and the equilibrium to be a focus. Finally, we conclude that the system has a global unique positive equilibrium which is asymptotically stable.

CHAPTER 5

Numerical simulations

Our model (3.14) - (3.22) is applied to simulate the behavior of two types of medulloblastoma, characterized by monosomy 6 and trisomy 6, respectively. To validate the model, we use patient data on the *SGK1* and *MYC* mRNAs obtained in the clinics in Heidelberg (Division of Molecular Genetics, German Cancer Research Center (DKFZ) Heidelberg, Germany [37]), Boston (Children's Hospital Boston, Boston, MA 02115, USA [12]) and Amsterdam (Department of Human Genetics, Academic Medical Center, Amsterdam, the Netherlands [60]) thanks to Stefan Pfister and Dominik Sturm.

In this chapter we describe the patient data from the clinics and how we introduce them into our model. We present parameters chosen for the model and the numerical tool used. We outline the dynamics of each population for monosomy 6, trisomy 6 and normal cells. We investigate the system dynamics and we make comparison between the patient data and the results obtained from the simulations. Finally, we highlight crucial features of the particular medulloblastoma subgroups, which lead us to formulate biological hypotheses. Numerical simulations of the mathematical model reveal the importance of *SGK1* expression for the prognosis in medulloblastoma.

5.1 Data, parameters and numerical tools

In this section we deal with the patient data. We delineate origin of the parameter values and a graphical user interface that we apply to investigate the dynamics of the model.

5.1.1 Patient data

Each patient sample is profiled by the set of features: gender, age, metastasis progression (mstg), death status and others. We do not apply these data to our model but compare them to the results obtained from the simulations.

The gene expression of *SGK1* and *MYC* (that is the *SGK1* and *MYC* mRNAs) is measured using DNA microarray analysis and it is provided by experimentalists. These data are fed to the model.

To model monosomy 6 and trisomy 6 associated medulloblastoma, we use different parameterizations of the *SGK1* transcript synthesis and *MYC* transcript synthesis. The parameterization corresponds to the experimentally observed fold change (see Section 2.1.1) of the level of the *SGK1* and *MYC* mRNAs in medulloblastoma tumor cells and in the normal cells. The *MYC* mRNA fold change in both cases is higher than 1. The *SGK1* mRNA fold change is smaller than 1 in monosomy 6 and larger than 1 in trisomy 6. We obtain that fold change of *SGK1*, which is synthesized on chromosome 6q, in 6q loss and 6q gain is not proportional to the copy-number of chromosome (Table 5.1). Intuitively, in monosomy (one copy of chromosome) the fold change should be equal to about 0.5 and in trisomy (three copies of chromosome) to 1.5.

Biological evidence 5.1 *Recent experiments show that the nonlinear relation between the number of gene copy and resulting gene expression of SGK1 is linked to the pattern of DNA methylation [56].*

To simulate dynamics of both medulloblastoma types, we consequently adjust the *SGK1* transcription rate such that the levels of mRNA agree with the patient data. We do the same procedure for the *MYC* mRNA. The mRNA experimental data which we follow in simulations, are collected from individual patients resected tumors. Exemplary fold changes of monosomy 6 and trisomy 6 samples are given in Table 5.1.

Biological evidence 5.2 *Monosomy 6 samples are characterized by the SGK1 mRNA downregulation compared to a healthy cells. In trisomy 6 sample we can observe overexpression compared to the healthy cells. MYC mRNA is overexpressed in both medulloblastoma subgroups and in the monosomy 6 case it is even higher. Such dependencies are found in a majority of the collected patient samples.*

monosomy	***SGK1* mRNA**	***MYC* mRNA**
fold change tumor/control	0.74	16.27
trisomy	***SGK1* mRNA**	***MYC* mRNA**
fold change tumor/control	7.11	6.45

Table 5.1: Examples of microarray data values from the clinics for two types of medulloblastoma.

To adjust in the simulations the *SGK1* and *MYC* transcription rates such that levels of mRNAs agree with the patient data, first we determine the basic set of parameters (Table 5.3) for the model of the control case (see Section 2.1.1). Then, to indicate the increase or decrease of the particular *SGK1* or *MYC* mRNA in the tumor cells, we extend the model to account for the mutation introducing coefficients L_1 and L_3 (parameterization) in two equations,

$$\frac{\mathsf{d}SGK1_t(t)}{\mathsf{d}t} = L_1 t_1 \beta cat - d_1 SGK1_t(t), \tag{5.1}$$

$$\frac{\mathsf{d}MYC_t(t)}{\mathsf{d}t} = L_3 t_3 \beta cat - d_3 MYC_t(t), \tag{5.2}$$

where L_1 is responsible for the *SGK1* mRNA increase/decrease and L_3 is responsible for the *MYC* mRNA increase. L_1 and L_3 play the role of control variables, i.e., independent variables that can be manipulated or controlled in an experimental design to understand how they affect the dependent variables of the model. They correspond to the fold change of selected transcripts. Originally, L_1 and L_3 are equal 1.

Furthermore, an important feature observed in medulloblastoma is a relatively small volume of the cytoplasm and quite large volume of the nucleus, which is typical for other malignant cells [50], [65], [51]. In our model, the ratio of two compartments is considered, hence we adjust the scaling factor k_v (see Section 3.1.4), accordingly. For normal cells, we assume $k_v = 2$, i.e., the cytoplasm is twice larger than the nucleus by volume. For malignant cells, we assume $k_v = 0.5$, i.e., the volume of the nucleus is twice larger than the cytoplasm. The scaling factor may be different for different cell types.

5.1.2 Model calibration

Our model consists of nine nonlinear ODEs and involves nineteen parameters. The parameters describe rates of the basic processes, e.g., translation, transport, association, etc. Orders of the magnitude of parameters are chosen following the Reference [44]. Thus, our

parameter magnitudes (places after decimal point) correspond to the analogous processes from [44]. The translation rate is of the order of 10^{-1}, spontaneous protein degradation is of the order of 10^{-4}, etc.

In the strategy of chosing parameters, we consider the protein size and the length of the transcript (see Table 5.2). More detailed, the size of the cMyc protein is 439 amino

Gene	**Transcript Length**	**Protein Size**
MYC	6.001 bases	439 amino acids
SGK1	148.867 bases	421-526 amino acids depending on isoform
GSK3β	273.095 bases	420 amino acids

Table 5.2: Gene description.

acids and size of GSK3β is 420 amino acids. Smaller molecules should be transported faster to the nucleus, hence the transport coefficient of GSK3β to the nucleus is equal to c_6 = 0.003 s^{-1} and of cMyc it is c_4 = 0.002 s^{-1}. The transport coefficient of GSK3β to the cytoplasm is equal to c_7 = 0.0009 s^{-1}, consequently it is smaller than the transport coefficient of GSK3β to the nucleus. We assume that GSK3β has crucial phosphorylating activity in the complex with the nuclear cMyc, so there is a bigger demand of GSK3β in the nucleus. Therefore, we assume that GSK3β shuttles faster to the nucleus ($c_6 > c_7$). Other factors can also influence the transport, but for sake of simplicity in our model we stick to the size rule. Further, the length of the *MYC* mRNA transcript is 6001 bases and the length of the *SGK1* mRNA transcript is 148867 bases. Taking into account the difference in both transcript lengths, we assume that the *MYC* transcript is produced faster than the *SGK1* transcript. As a result, we assume higher transcription rate for *MYC*. The size of the SGK1 protein isoforms is in the range 421-526 amino acids. The difference in the protein size of SGK1 and cMyc is not large, but the difference in transcripts length is great, hence the translation of *MYC* should be also faster than the translation of *SGK1* (translation of the nucleotide sequence takes time). The bigger difference between the transcript and the protein size, the longer translation process and the smaller the total amount of the new synthesized protein in the same period of time. Thus, we take s_4 = 0.4 s^{-1} as translation rate of *MYC* and s_2 = 0.2 s^{-1} as translation rate of *SGK1*. Next, in our model we have a spontaneous degradation for both transcripts and proteins. We want to preserve the assumption that level of the *MYC* gene, as well transcript as protein, is higher than level of the *SGK1* gene (transcript, protein). For this reason, we apply higher degradation rate of the *SGK1* gene than *MYC* for both the transcript and protein. From the molecular point of view transcripts are less stable than proteins, what is reflected in differences of degradation rates. The transcript degradation rate is larger than the pro-

tein degradation rate for the same gene. Considering degradation through the interaction with other proteins, we stick to the rule that the probability of meeting of two proteins is higher in the nucleus than in the cytoplasm. The reason is that the nucleus is smaller than cytoplasm in the normal cells and binding process may occur much faster in this compartment. In our system the phosphorylation of the SGK1 protein is indicated as a fast process. Thus, the dissociation of the complex [SGK1·GSK3β] is assumed to be faster than dissociation of the complex [GSK3β·cMyc]. Following the above methodology, we choose the parameters of the model.

Symbol	Value	Description
P_{GSK}	0.00002 μMs^{-1}	constant influx of the GSK3β protein to the cytoplasm
βcat	0.4 μM	transcription factor for the *MYC* and *SGK1* mRNA synthesis
t_1	2×10^{-7} s^{-1}	transcription rate of *SGK1*
t_3	4×10^{-7} s^{-1}	transcription rate of *MYC*
s_2	0.2 s^{-1}	translation rate of SGK1
s_4	0.4 s^{-1}	translation rate of cMyc
d_o	0.00005 s^{-1}	spontaneous degradation rate of the GSK3β protein in the cytoplasm
d_1	0.0007 s^{-1}	spontaneous degradation rate of the *SGK1* mRNA
d_2	0.0004 s^{-1}	spontaneous degradation rate of the SGK1 protein
d_3	0.0006 s^{-1}	spontaneous degradation rate of the *MYC* mRNA
d_4	0.0002 s^{-1}	spontaneous degradation rate of the cMyc protein in the cytoplasm
d_5	0.5 μM$^{-1}s^{-1}$	degradation rate of the cMyc protein in the nucleus by GSK3β
d_6	0.3 μM$^{-1}s^{-1}$	degradation rate of the GSK3β protein in the cytoplasm by SGK1
k_v	2	scaling coefficient, cytoplasm to the nucleus ratio
c_4	0.002 s^{-1}	transport of cMyc to the nucleus coefficient
c_6	0.003 s^{-1}	transport of GSK3β to the nucleus coefficient
c_7	0.0009 s^{-1}	transport of GSK3β to the cytoplasm coefficient
p_8	0.4 s^{-1}	dissociation coefficients of the [GSK3β·cMyc] complex
p_9	0.6 s^{-1}	dissociation coefficients of the [SGK1·GSK3β] complex

Table 5.3: A summary of model parameters for healthy sample (cf. [44]).

Statement 5.1 *Numerical simulations show robustness of the model behavior with respect to the parameters. The results are qualitatively conserved for a large range of parameter values.*

5.1.3 Graphical user interface

We perform numerical simulations with MATLAB®ODE solver ode23s [67], using crude error tolerances to resolve stiff systems [52]. Our system is stiff since the stiff coefficient for basic set of parameters

$$s = \frac{max_{i=1,2...m}|\lambda_i|}{min_{i=1,2...m}|\lambda_i|} \tag{5.3}$$

is equal to s = 1500 for healthy sample parameterization and s = 7000 for chosen trisomy 6 sample parameterization (see eigenvalues λ_i in Section 4.2.3). For sake of simplicity, we created graphical user interface (GUI) to make simulations and apply different tests (see Figure 5.1 and 5.2). The advantage of such application is that there is no need to change the source code of MATLAB® m-file every time when the parameters or other factors of the simulation are changed. The GUI consists of three types of elements: values that user can adjust, values that user gets after performing simulations (not available for changes of the user) and the graph of the solutions. Parameters of the mathematical model, initial values, time settings and number of visualized solutions can be adjusted by the user. A pop-up menu gives additional opportunity to either display the time dependence or the phase portrait. The important data which we get due to simulations visible on GUI are: steady state, eigenvalue and fold change. We present GUI as a convenient tool, where different scenarios of parameters can be easily tested.

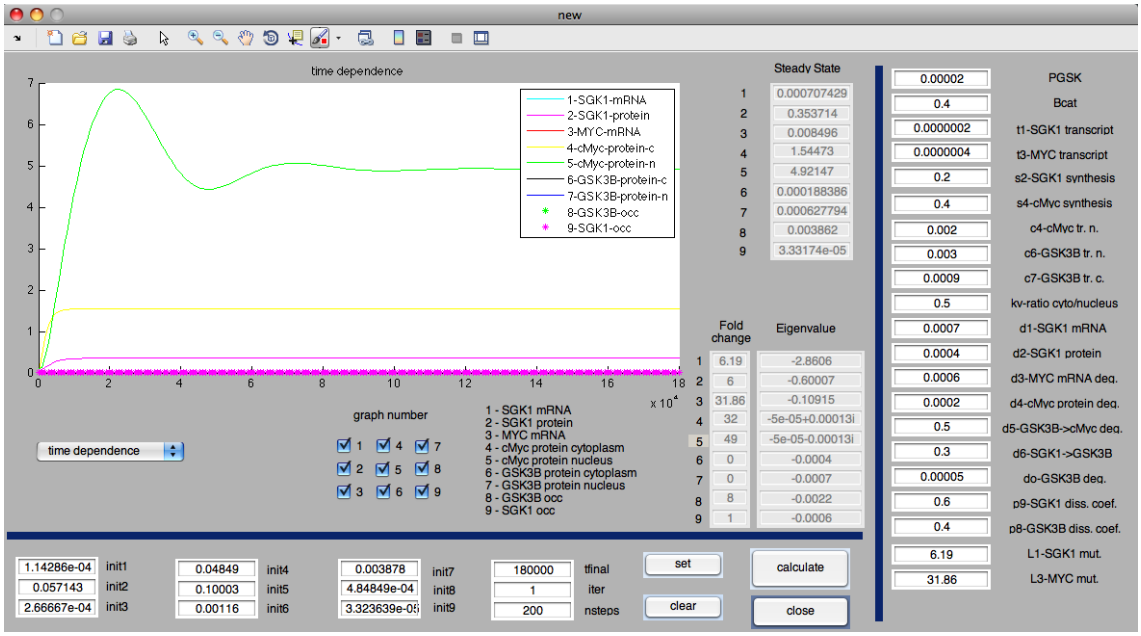

Figure 5.1: Graphical user interface - numerical solution of trisomy 6 sample - time dependence graph.

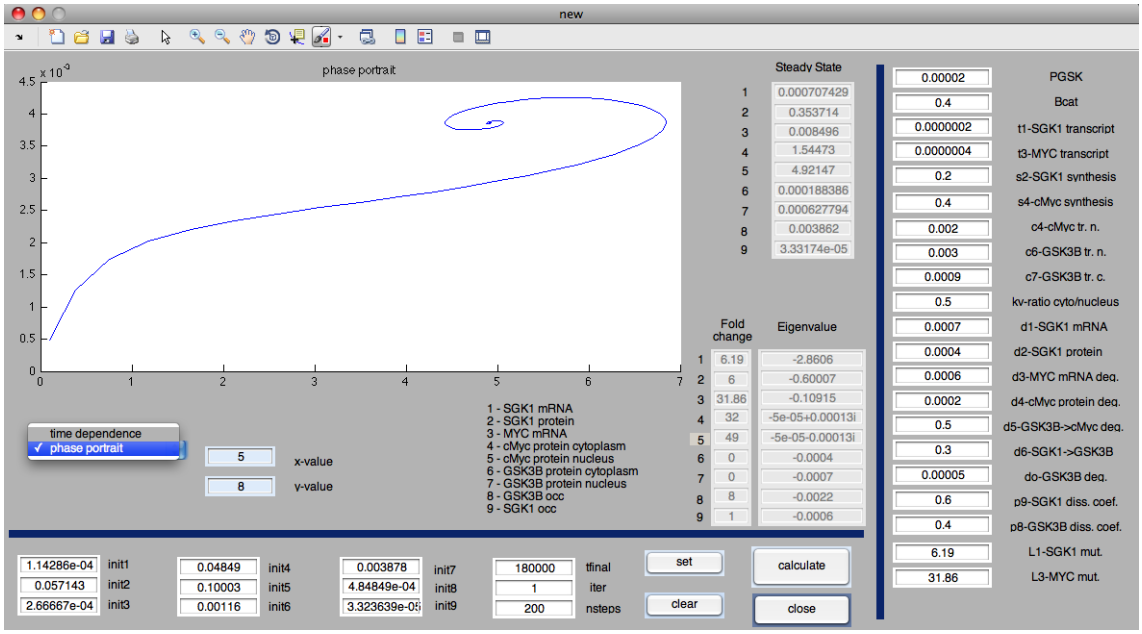

Figure 5.2: Graphical user interface - numerical solution of trisomy 6 sample - phase portrait.

5.2 Simulation results

In this section we present numerical simulations based on 49 patient samples (20 trisomy samples and 29 monosomy samples). The simulations are done for the initial conditions corresponding to the stationary state values, see (4.16) - (4.24), of the "healthy" system (signaling in "normal" - non-malignant cells). We compare the dynamics between monosomy 6, trisomy 6 and normal cells. We also investigate the correlation between patient data and simulations. Proceeding the simulation studies, we formulate new hypotheses concerning the difference in dynamics of the two types of medulloblastoma based on the regulatory loop *SGK1-GSK3β-MYC*.

5.2.1 Monosomy 6, trisomy 6 and control case

We perform numerical simulations for 6q loss, 6q gain and the control case for chosen patient samples (see Table 5.4) to better understand the dynamics of the medulloblastoma subgroups and the control. We depict these three cases on Figures 5.3, 5.4 and 5.5 for all model variables.

	monosomy	**trisomy**	**control**
fold change *SGK1* mRNA	L_1 = 0.74	L_1 = 7.11	L_1 = 1
fold change *MYC* mRNA	L_3 = 16.27	L_3 = 6.45	L_3 = 1
scaling factor	k_v = 0.5	k_v = 0.5	k_v = 2

Table 5.4: Examples of microarray data values from the clinics for two types of medulloblastoma and control.

• Figure 5.3

[A] Dynamics of the *SGK1* mRNA:

In the trisomy 6 cells we obtain the exponential increase of the *SGK1* mRNA and simultaneous exponential decrease in monosomy 6.

[B] Dynamics of the SGK1 protein in the cytoplasm:

SGK1 protein in the cytoplasm follows its mRNA values, in both monosomy 6 and trisomy 6.

[C] Dynamics of the *MYC* mRNA:

We observe the exponential *MYC* mRNA increase in both types of medulloblastoma.

• Figure 5.4

[A] Dynamics of the cMyc protein in the cytoplasm:

The behavior of the cytoplasmic cMyc is based on the *MYC* mRNA level; most often the cytoplasmic cMyc is lower in the trisomy 6 case than in the monosomy 6 case.

[B] Dynamics of the cMyc protein in the nucleus:

There is a peak of cMyc and then decrease in both types of medulloblastoma. Most often the nuclear cMyc in trisomy 6 is higher than in monosomy 6, opposite to the cytoplasmic cMyc.

[C] Dynamics of the GSK3β protein in the cytoplasm:

GSK3β in the cytoplasm is increased in 6q loss and decreased in 6q gain.

• Figure 5.5

[A] Dynamics of the GSK3β protein in the nucleus:

In both types of medulloblastoma the level of GSK3β in the nucleus is higher than for the cytoplasmic GSK3β.

[B] Dynamics of the phosphorylating GSK3β protein in the nucleus:

GSK3β increase is observed in monosomy 6 and trisomy 6. The increase is larger in the case of monosomy 6, however in both types of cancer the nuclear cMyc is high. Thus, we have increased phosphorylation activity of GSK3β.

[C] Dynamics of the phosphorylating SGK1 protein in the cytoplasm:

In 6q gain, there is a peak of the phoshorylating SGK1 and then a decrease. Inversely, in 6q loss we have at first a decrease and then we observe that the SGK1 level increases.

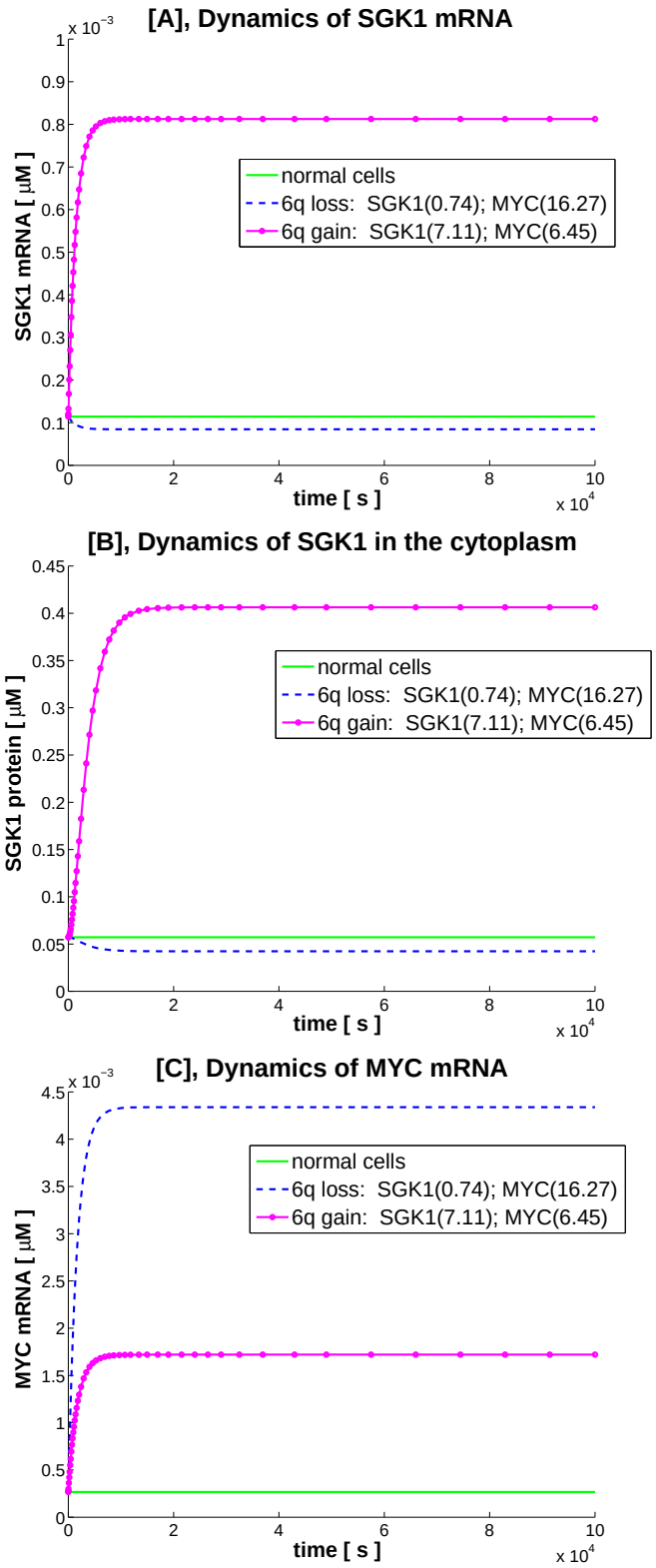

Figure 5.3: Numerical simulations based on exemplary microarray data values from the clinics. Population dynamics of *SGK1* mRNA, cytoplasmic SGK1 and *MYC* mRNA. Results for 6q gain, 6q loss and normal cells.

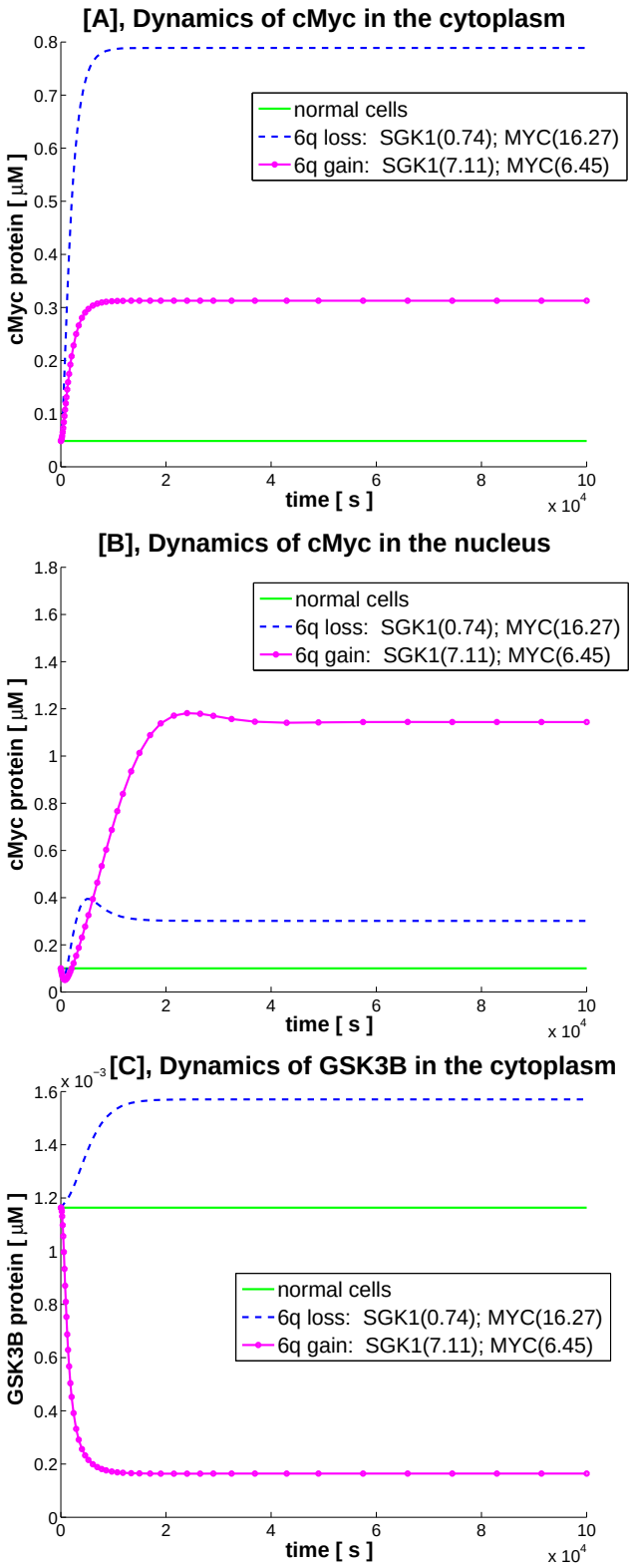

Figure 5.4: Numerical simulations based on exemplary microarray data values from the clinics. Population dynamics of the cytoplasmic and nuclear cMyc protein and GSK3β in the cytoplasm. Results for 6q gain, 6q loss and normal cells.

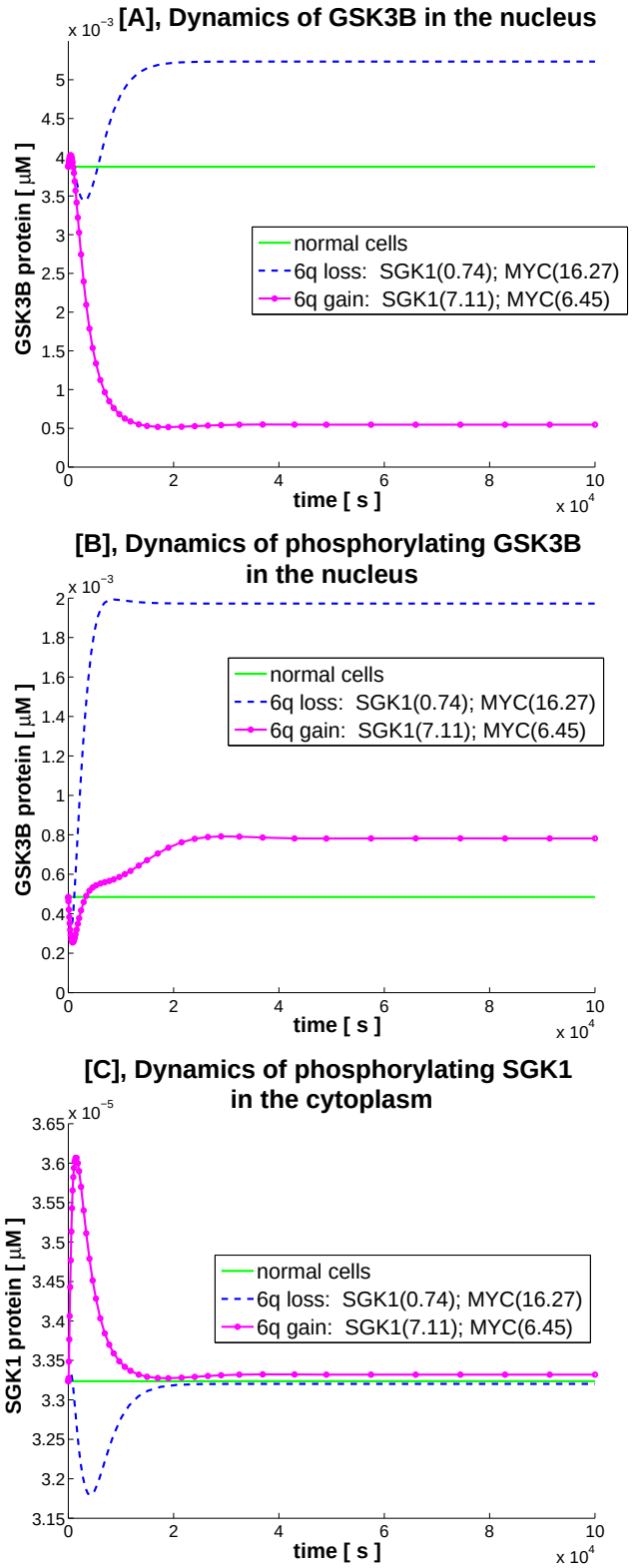

Figure 5.5: Numerical simulations based on exemplary microarray data values from the clinics. Population dynamics of protein GSK3β in the nucleus, phosphorylating GSK3β in the nucleus and phosphorylating SGK1 in the cytoplasm. Results for 6q gain, 6q loss and normal cells.

Now, we focus on the long time behavior of particular protein concentration. In both types of cancer the cMyc protein is always higher than control. The cytoplasmic cMyc demonstrates even higher fold change in monosomy 6 than in the case of trisomy 6. On the opposite, the nuclear cMyc level indicates much lower fold change in the monosomy 6 case than in the trisomy 6 case. As a result, we observe significant differences between the two cellular compartments with respect to the type of medulloblastoma tumor cells. Further, the nuclear GSK3β is lower in trisomy 6 than in monosomy 6, even much lower than in the normal cells. Taking into account the cytoplasmic SGK1 level, we have a lower level (even below control) in monosomy 6 compared to trisomy 6. In general, based on the results from the simulation, we can notice the following dependencies that are crucial for the model understanding and may explain discrepancy in medulloblastoma prognosis.

Observation 5.1 *The results of simulations show following qualitative relations:*

$$\textit{cytoplasmic SGK1} \uparrow \xrightarrow{cytoplasmic\ GSK3\beta\downarrow} \textit{nuclear GSK3}\beta \downarrow \ \rightarrow \ \textit{nuclear cMyc} \uparrow$$
and
$$\textit{cytoplasmic SGK1} \downarrow \xrightarrow{cytoplasmic\ GSK3\beta\uparrow} \textit{nuclear GSK3}\beta \uparrow \ \rightarrow \ \textit{nuclear cMyc} \downarrow.$$

5.2.2 Comparison between the patient data and simulations based on the microarray data

In previous section we investigated concentrations of each species in the case of monosomy 6, trisomy 6 and normal cells. We want to compare the results from the simulations with the patient data. We link the death status with the cMyc protein and find a strong correlation between the death status and high nuclear cMyc level.

Trisomy 6

In our analysis we compare samples of high nuclear cMyc (from simulations) with patients assigned to the group of positive death status. We found a positive correlation. It seems that high level of the cMyc protein in the nucleus favors patients death. The comparison indicates also that the critical high nuclear cMyc differs for different clinics. The explanation may lie in short time diagnosis or environmental features promoting longer survival time of the patient (Figure 5.6).

	patient no	SGK1 mRNA	MYC mRNA	cMyc nucleus	age	gender	mstg
gain							
Heidelberg							
	1M19	7.82	103.24	201	4	0	M0
	1M21	4.16	29.78	31	8	1	M0
	1M20	7.92	12.02	24	4	0	M3
	3M2	1.12	13.59	4	10	0	M3
	2M7	1.64	0.92	0	4	1	M1
Amsterdam							
	455	6.15	431.67	662	2	F	M0
	259	4.09	10.96	11	15	M	M0
	264	7.11	6.45	11	7	F	M0
	458	5.03	7.81	10	3	M	M1
	365	5.7	5.22	7	10	M	M0
	268	7.86	3.05	6	5	F	M0
	337	1.05	7.35	2	16	M	>M2
	326	3.17	2.57	2	5	M	M0
Boston							
	79	7.05	155.39	273	4	M	M0
	143	6.19	31.86	49	14	M	M3
	103	12.6	4.12	13	6	M	M0
	264	1.7	9.84	4	4	M	M0
	106	3.38	3.5	3	2	M	M0
	48	6	1.95	3	2	M	M0
	47	3.42	3.27	3	6	F	M0

Figure 5.6: Trisomy 6: presented data are composed of the patient part and results from the simulations. We notice that for high nuclear cMyc from the simulations (bright grey) there are samples where death status (dark grey) is assigned. Fold change of the cMyc protein equal to zero is due to the rounding a point decimal number and in fact is below 0.5 value.

Monosomy 6

In the case of 6q loss we have lower correlation between level of the nuclear cMyc and death status. Nevertheless, we still observe a tendency (Figure 5.7). Generally, the prognosis based on the patient data is much better in the case where level of the nuclear cMyc is substantially lower.

General observations

Our analysis yields the conclusion that the level of cMyc in the nucleus lends itself as a predictive factor. System dynamics reveal interesting features of monosomy 6 and trisomy 6. We claim that the observations indicate divergence in prognosis for the two types of medulloblastoma. Additionally, we remark that even if the nuclear cMyc in monosomy 6 is higher than in the normal cells (most cases), it is not too high and, hence, the treatment still provides positive effects. In the case of trisomy 6 the nuclear cMyc is too high and treatment cannot overcome the disease.

	patient no	SGK1 mRNA	MYC mRNA	cMyc nucleus	age	gender	mstg
loss							
Heidelberg							
	5M19	0.47	61.91	7	29	1	M1
	1M16	0.54	28.59	4	11	1	M0
	1M14	0.74	16.27	3	10	1	M0
	4M55	0.65	11.76	2	12	0	M0
	5M4	4.79	1.91	2	25	1	M0
	4M28	0.44	10.06	1	16	1	M0
	1M4	0.22	18.16	1	12	0	M1
	1M15	0.19	20.5	1	13	1	M0
	5M15	0.19	17.18	1	27	1	M0
	5M7	0.17	20.83	1	18	0	M0
	1M13	0.12	19.15	1	8	0	M0
	5M5	0.12	14.97	0	19	0	M0
	5M13	0.11	0.95	0	44	0	M0
	1M22	0.07	22.45	0	12	1	M0
Amsterdam							
	255	0.67	59.25	10	19	M	ND
	371	0.56	48.2	7	11	F	M0
	369	0.43	63.74	7	10	M	M0
	260	0.53	33.41	4	8	M	M0
	379	0.27	38.11	3	10	M	ND
	340	0.27	34.04	2	7	F	M0
	425	0.27	31.06	2	6	F	M0
	372	0.15	60	2	12	M	M0
	270	0.32	12.33	1	20	F	M0
Boston							
	120	1.06	31.78	8	10	F	M0
	244	0.82	24.54	5	4	F	na
	144	0.79	12.99	3	7	F	M0
	240	0.47	17.2	2	10	M	M0
	15	0.19	13.99	1	10	F	M0
	112	0.15	8.58	0	7	F	M0

Figure 5.7: Monosomy 6: presented data are composed of the patient part and results from the simulations. We notice that for high nuclear cMyc from the simulations (bright grey) there are samples where the death status (dark grey) is assigned. Fold change of the cMyc protein equal to zero is due to the rounding a point decimal number and in fact is below 0.5 value.

Statement 5.2 *There exists a positive correlation between high nuclear cMyc obtained from the simulations and patients death status taken from the patient data (see Figure 5.6 and Figure 5.7).*

5.2.3 Discrepancy in dynamics of genes *MYC* and *SGK1* in trisomy 6 and monosomy 6

In this section we investigate differences in the nuclear cMyc. To do so, we simulate particular mRNA of *MYC* and *SGK1*. As dynamics of the *MYC* and *SGK1* mRNAs strongly depend on the type of medulloblastoma, we see how they influence the patient prognosis

(i.e., the amount of the nuclear cMyc). We simulate different patient samples and compare them for different scenarios.

COMPARISON OF THE SAMPLES WITH SIMILAR VALUES OF THE *MYC* MRNA IN BOTH TYPES OF MEDULLOBLASTOMA

Since *MYC* regulates many genes, we perform simulations of the monosomy 6 and trisomy 6 model to investigate how *SGK1* impacts the *MYC* production. Numerical simulations show that the change in the *MYC* transcription rate may result in markedly different levels of the cMyc protein depending on the magnitude of perturbations in the *SGK1* dynamics, which may be the reason for the different prognosis (Table 5.5 and Table 5.6). Based on the fold change of the *SGK1* and *MYC* mRNA, we present the dynamics of the cytoplasmic and the nuclear cMyc in Figure 5.8 and Figure 5.9 to hightlight the difference between the two cellular compartments. The cytoplasmic level follows the dynamics of the *MYC* mRNA and the nuclear cMyc follows the system dynamics.

monosomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	0.54	28.59	4
trisomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	4.16	29.78	31

Table 5.5: Microarray data values from the clinics for two types of medulloblastoma of similar *MYC* mRNA and different *SGK1* mRNA production.

monosomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	0.32	12.33	1
trisomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	7.92	12.02	24

Table 5.6: Microarray data values from the clinics for two types of medulloblastoma of similar *MYC* mRNA and different *SGK1* mRNA production.

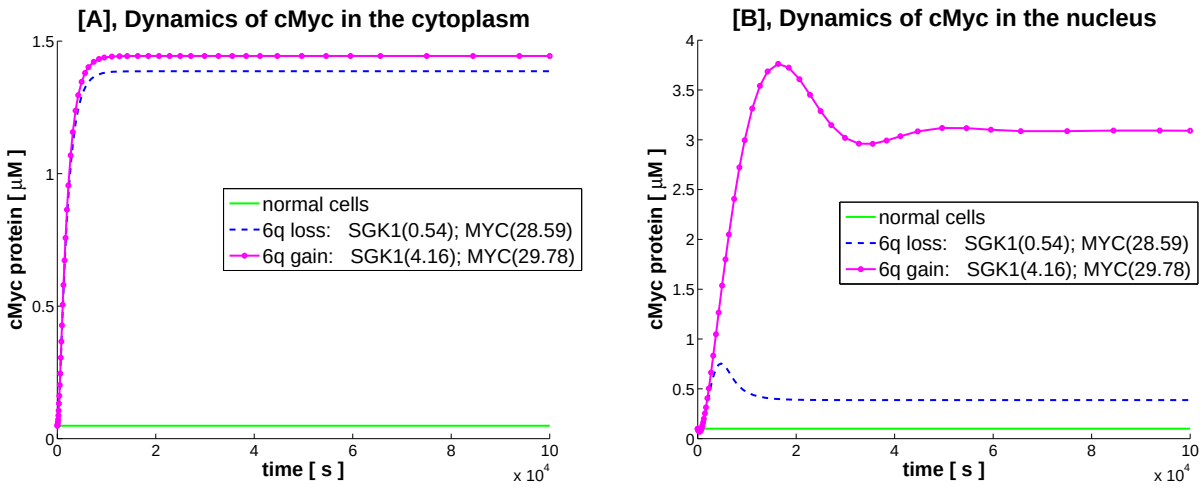

Figure 5.8: Numerical simulations based on exemplary microarray data values from the clinics. Dynamics of [A] cMyc protein in the cytoplasm and [B] cMyc protein in the nucleus in the trisomy 6 case (pink-solid dotted line), monosomy 6 case (blue-dashed line) and healthy tissue (green-solid line), each corresponding to different production rate of *SGK1* and similar production rate of *MYC* for two types of medulloblastoma.

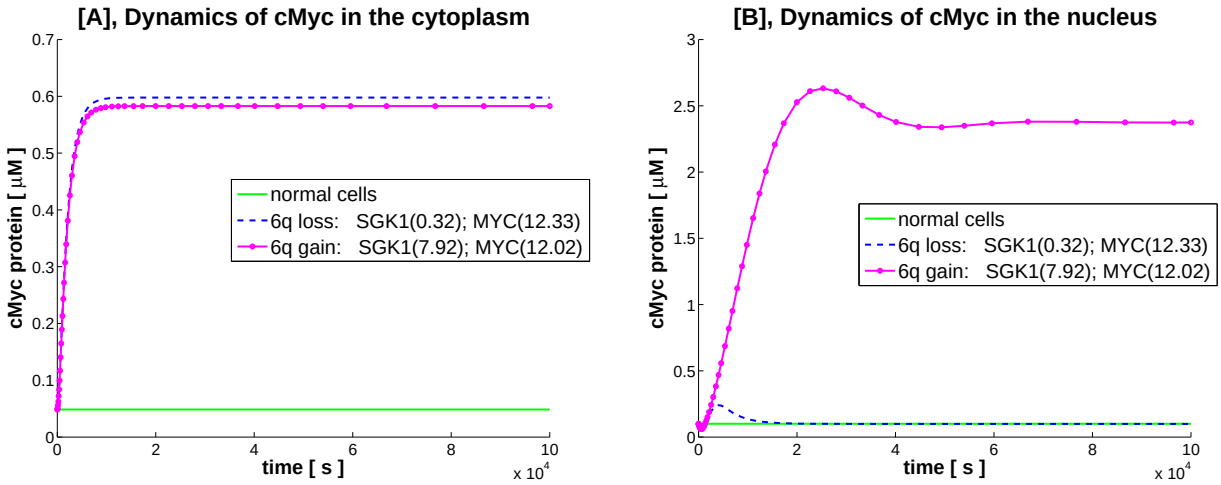

Figure 5.9: Numerical simulations based on exemplary microarray data values from the clinics. Dynamics of [A] cMyc protein in the cytoplasm and [B] cMyc protein in the nucleus in the trisomy 6 case (pink-solid dotted line), monosomy 6 case (blue-dashed line) and healthy tissue (green-solid line), each corresponding to different production rate of *SGK1* and similar production rate of *MYC* for two types of medulloblastoma.

COMPARISON OF THE SAMPLES WITH SIMILAR VALUES OF THE *MYC* MRNA WITHIN EACH MEDULLOBLASTOMA TYPE

Due to simulations, samples with similar mRNA of *MYC* and varying levels of the *SGK1* expression within one type of medulloblastoma show different nuclear cMyc levels. Again, we notice that the *SGK1* mRNA is responsible for the difference (Table 5.7 and Table 5.8). Additionally, we observe strong influence of higher *SGK1* mRNA on the levels of the nuclear cMyc in the case of trisomy 6. We present the dynamics of the cytoplasmic and the nuclear cMyc in Figure 5.10 and Figure 5.11.

monosomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	0.19	17.18	1
monosomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	0.47	17.20	2

Table 5.7: Microarray data values from the clinics for monosomy 6 of similar *MYC* mRNA and different *SGK1* mRNA production.

trisomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	7.92	12.02	24
trisomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	1.12	13.59	4

Table 5.8: Microarray data values from the clinics for trisomy 6 of similar *MYC* mRNA and different *SGK1* mRNA production.

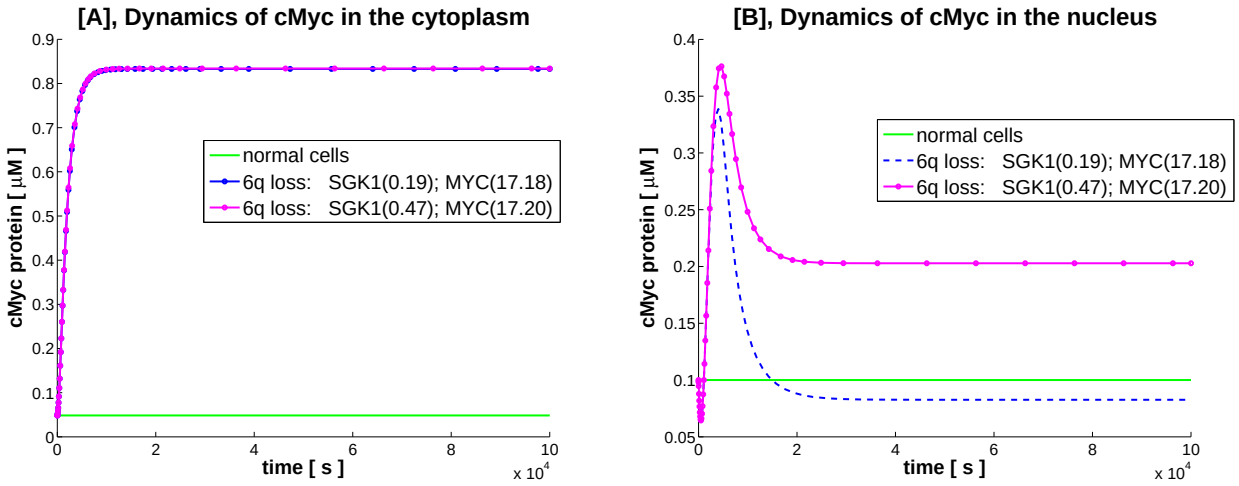

Figure 5.10: Numerical simulations based on exemplary microarray data values from the clinics. Dynamics of [A] cMyc protein in the cytoplasm and [B] cMyc protein in the nucleus in the first monosomy 6 case (pink-solid dotted line), second monosomy 6 case (blue-dashed line) and healthy tissue (green-solid line), each corresponding to different production rate of *SGK1* and similar production rate of *MYC*.

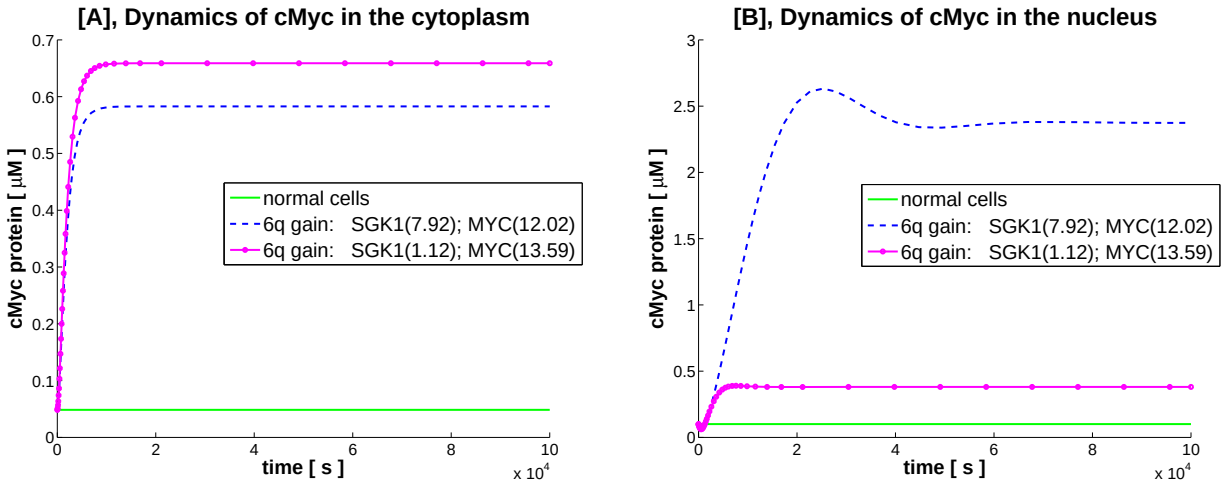

Figure 5.11: Numerical simulations based on exemplary microarray data values from the clinics. Dynamics of [A] cMyc protein in the cytoplasm and [B] cMyc protein in the nucleus in the first trisomy 6 case (pink-solid dotted line), second trisomy 6 case (blue-dashed line) and healthy tissue (green-solid line), each corresponding to different production rate of *SGK1* and similar production rate of *MYC*.

COMPARISON OF THE SAMPLES WITH THE *MYC* MRNA MUCH HIGHER IN MONOSOMY 6 THAN IN TRISOMY 6

We observe that for the patient samples with much higher *MYC* mRNA production in monosomy 6 than in trisomy 6 we still have nuclear cMyc lower in monosomy 6 than in trisomy 6 (Table 5.9 and Table 5.10). This indicates much higher *SGK1* mRNA production in trisomy 6 than in monosomy 6. Intuitively, we would claim that the nuclear cMyc concentration depends mostly on the *MYC* mRNA production. However, in this case we confirm a strong impact of *SGK1* on the nuclear cMyc. The dynamics of cytoplasmic and the nuclear cMyc are presented in Figure 5.12 and Figure 5.13.

monosomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	0.54	28.59	4
trisomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	7.92	12.02	24

Table 5.9: Microarray data values from the clinics for two types of medulloblastoma of *MYC* mRNA production higher in monosomy 6 than in trisomy 6.

monosomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	0.65	11.76	2
trisomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	5.7	5.22	7

Table 5.10: Microarray data values from the clinics for two types of medulloblastoma of *MYC* mRNA production higher in monosomy 6 than in trisomy 6.

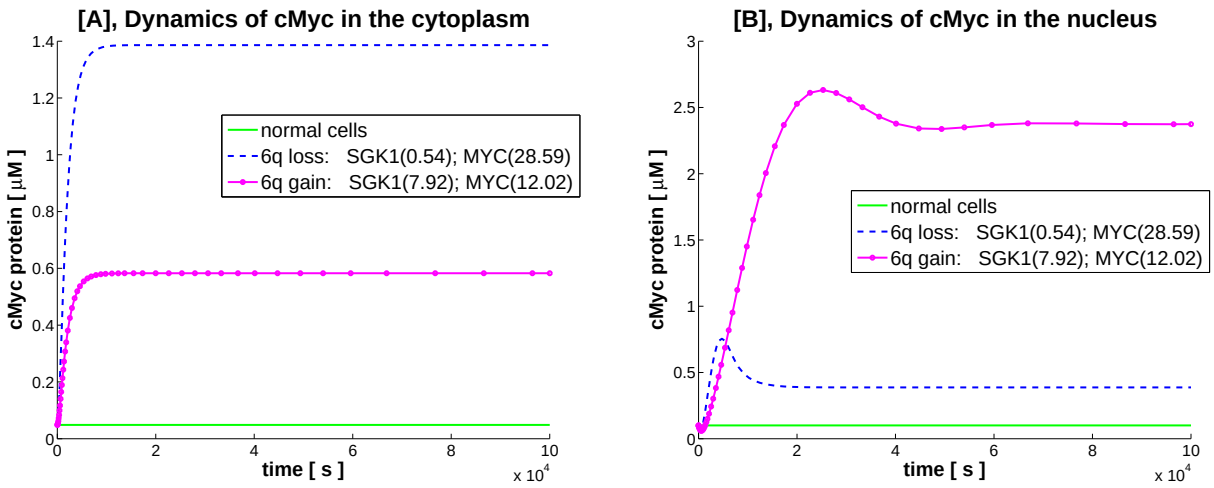

Figure 5.12: Numerical simulations based on exemplary microarray data values from the clinics. Dynamics of [A] cMyc protein in the cytoplasm and [B] cMyc protein in the nucleus in the trisomy 6 case (pink-solid dotted line), monosomy 6 case (blue-dashed line) and healthy tissue (green-solid line), each corresponding to production rate of *MYC* higher in monosomy 6 than in trisomy 6.

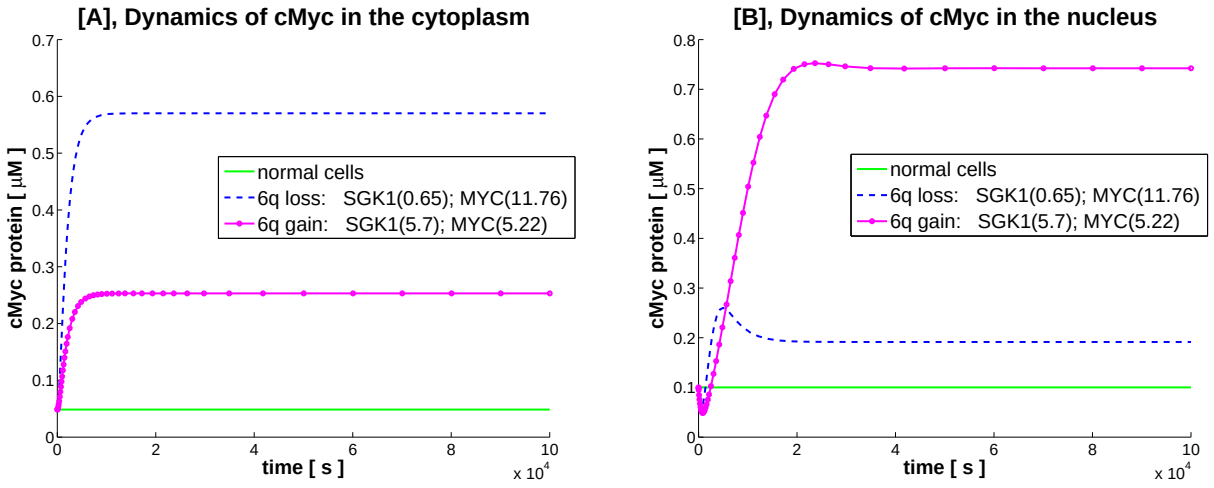

Figure 5.13: Numerical simulations based on exemplary microarray data values from the clinics. Dynamics of [A] cMyc protein in the cytoplasm and [B] cMyc protein in the nucleus in the trisomy 6 case (pink-solid dotted line), monosomy 6 case (blue-dashed line) and healthy tissue (green-solid line), each corresponding to production rate of *MYC* higher in monosomy 6 than in trisomy 6.

COMPARISON OF THE SAMPLES WITH SIMILAR SGK1 MRNA WITHIN MONOSOMY 6 AND TRISOMY 6

We perceive the important role of the *SGK1* mRNA in the patient samples of medulloblastoma. However, we cannot disregard the *MYC* mRNA which itself also has influence on the nuclear cMyc protein. Consistently, we perform simulations for similar *SGK1* mRNA production and different *MYC* mRNA production for samples in monosomy 6 (Table 5.11). Then, we repeat this scenario for the trisomy 6 samples (Table 5.12). Our intuition is confirmed by the tests and suggests that the *MYC* mRNA production has also an impact on the nuclear cMyc.

monosomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	0.27	38.11	3
monosomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	0.27	31.06	2

Table 5.11: Microarray data values from the clinics for monosomy 6 of similar *SGK1* mRNA and different *MYC* mRNA production.

trisomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	7.92	12.02	24
trisomy	***SGK1* mRNA**	***MYC* mRNA**	**cMyc in the nucleus**
fold change tumor/control	7.86	3.05	6

Table 5.12: Microarray data values from the clinics for trisomy 6 of similar *SGK1* mRNA and different *MYC* mRNA production.

The dynamics of the cytoplasmic and nuclear cMyc are presented in Figure 5.14 and Figure 5.15. In the situation of monosomy 6, we notice that even if the *MYC* mRNA production is high the resultant amount of the cMyc protein in the nucleus is much lower than in the trisomy 6. Therefore, we may explain why prognosis in 6q loss is better than in 6q gain. This shows that high *MYC* mRNA in monosomy 6, even much higher than in trisomy 6, is not fatal for patients. Numerical simulations confirm the medical observations that good prognosis is correlated with monosomy 6 and poor prognosis is correlated with trisomy 6.

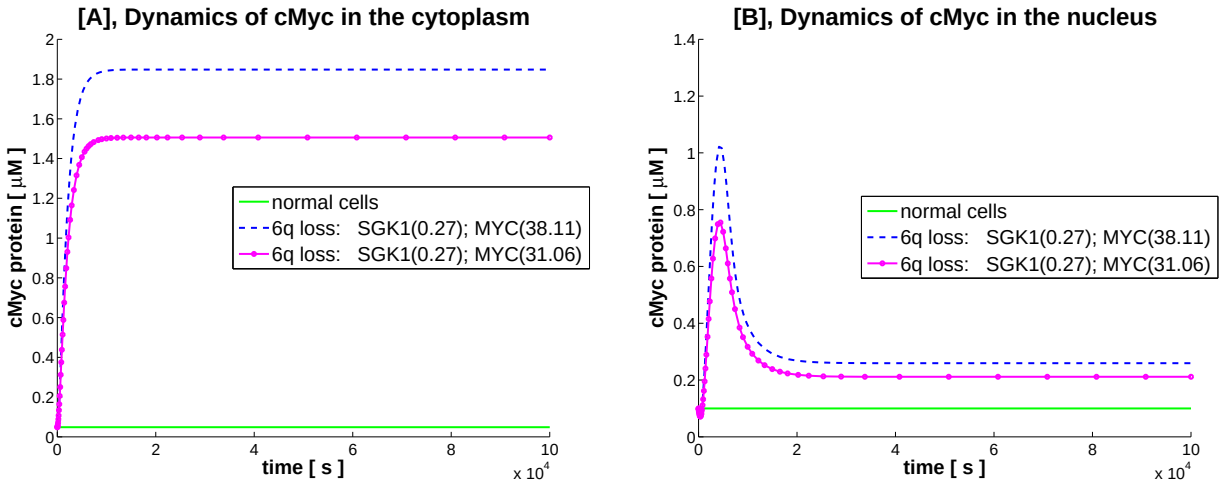

Figure 5.14: Numerical simulations based on exemplary microarray data values from the clinics. Dynamics of [A] cMyc protein in the cytoplasm and [B] cMyc protein in the nucleus in the first monosomy 6 case (pink-solid dotted line), second monosomy 6 case (blue-dashed line) and healthy tissue (green-solid line), each corresponding to different production rate of *MYC* and similar production rate of *SGK1*.

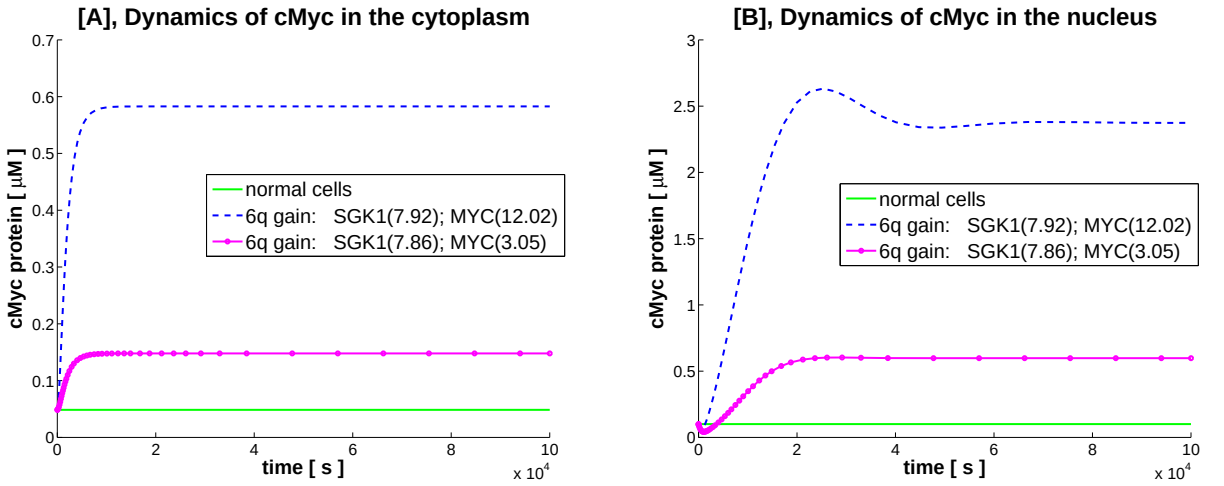

Figure 5.15: Numerical simulations based on exemplary microarray data values from the clinics. Dynamics of [A] cMyc protein in the cytoplasm and [B] cMyc protein in the nucleus in the first trisomy 6 case (pink-solid dotted line), second trisomy 6 case (blue-dashed line) and healthy tissue (green-solid line), each corresponding to different production rate of *MYC* and similar production rate of *SGK1*.

COMPARISON OF THE SAMPLES WITH GSK3β DYNAMICS IN MONOSOMY 6 AND TRISOMY 6

The graphs presented before show how the nuclear cMyc changes with respect to the *SGK1* and *MYC* mRNA production. However, there is no direct interaction between these two proteins.

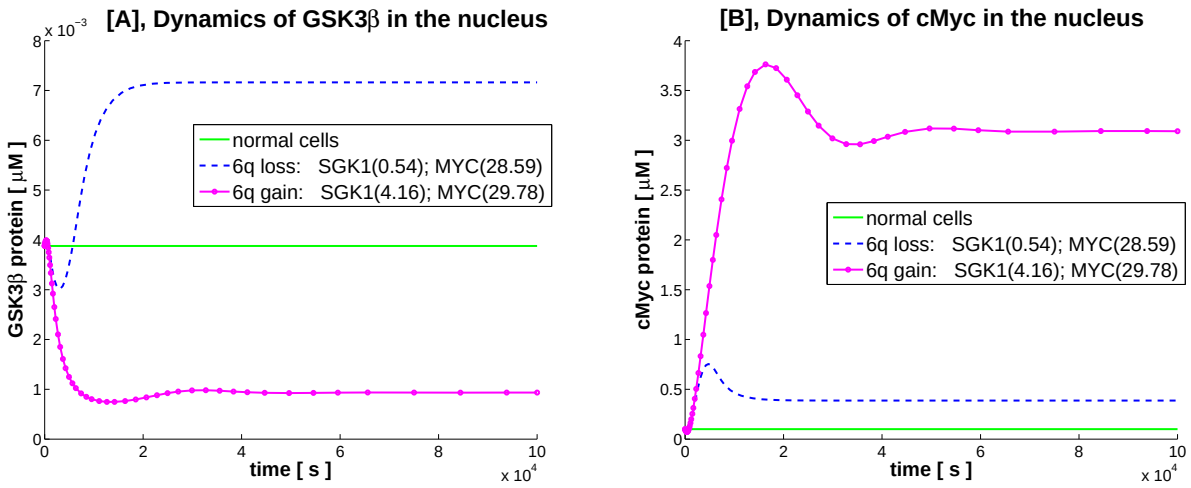

Figure 5.16: Numerical simulations based on exemplary microarray data values from the clinics. [A] GSK3β protein in the nucleus and [B] cMyc protein in the nucleus in the trisomy 6 case (pink-solid dotted line), monosomy 6 case (blue-dashed line) and healthy tissue (green-solid line). High GSK3β protein leads to low nuclear cMyc protein and low GSK3β protein leads to high nuclear cMyc protein.

GSK3β protein is a link between SGK1 and cMyc as it interacts with both of the proteins. We perceive that in the case of monosomy 6 the GSK3β protein is higher than in the control case. It is caused by weaker interaction with SGK1, which is downregulated in monosomy 6. On the other side, in the case of trisomy 6 the GSK3β protein decreases under the control level due to the SGK1 upregulation (Figure 5.16). The following changes in GSK3β influence further the nuclear cMyc. The regulatory loop *SGK1-GSK3β-MYC* demonstrates totally diverse behavior applying to the two types of medulloblastoma. This underlines the crucial role of these proteins in trisomy 6 and monosomy 6.

SUMMARY OF THE OBSERVATIONS FROM THE SIMULATIONS

- Similar *MYC* mRNA production in trisomy 6 and monosomy 6 results in different nuclear cMyc according to the *SGK1* mRNA production (Figure 5.8, Figure 5.9).

- Similar *MYC* mRNA production in each medulloblastoma subgroup underlies the influence of varying *SGK1* mRNA production on the nuclear cMyc (Figure 5.10, Figure 5.11).

- Higher *MYC* mRNA production in monosomy 6 than in trisomy 6 may still result in much lower nuclear cMyc (Figure 5.12, Figure 5.13).

- High *MYC* mRNA in monosomy 6 may still bring positive prognosis (Figure 5.14).

- Low *MYC* mRNA in trisomy 6 may lead to high nuclear cMyc due to influence of *SGK1* (Figure 5.15).

- Degradation of the nuclear cMyc by GSK3β is higher in the case of monosomy 6 than in the case of trisomy 6 (Figure 5.16).

- Dynamics of the regulatory loop *GSK3β-SGK1-MYC* may explain discrepancy in the nuclear cMyc.

Statement 5.3 *We identify the crucial role of SGK1 as a driving factor which strongly influences the* cMyc *level in the nucleus. MYC is assigned to be a proto-oncogene and is found upregulated also in other cancers [15], [53], [63]. Increased MYC leads to upregulated expression of many genes, some of which are involved in the cell proliferation. These changes may result in cancer. High increase in the* cMyc *protein follows the increase of the* SGK1 *protein. In monosomy 6 the production of SGK1 is lower than 1, thus the* cMyc *protein is not increased due to the SGK1 activity. Indeed, we observe that prognosis is much better in this case. Taking into account both types of medulloblastoma, we claim that SGK1, which is correlated with MYC, may play a main role in the survival differential.*

In Figure 5.17, we present relationship between the cMyc protein in the nucleus, the *MYC* and *SGK1* mRNAs obtained through the numerical simulations of the dynamics of the model with different parameter sets corresponding to the different virtual patients.

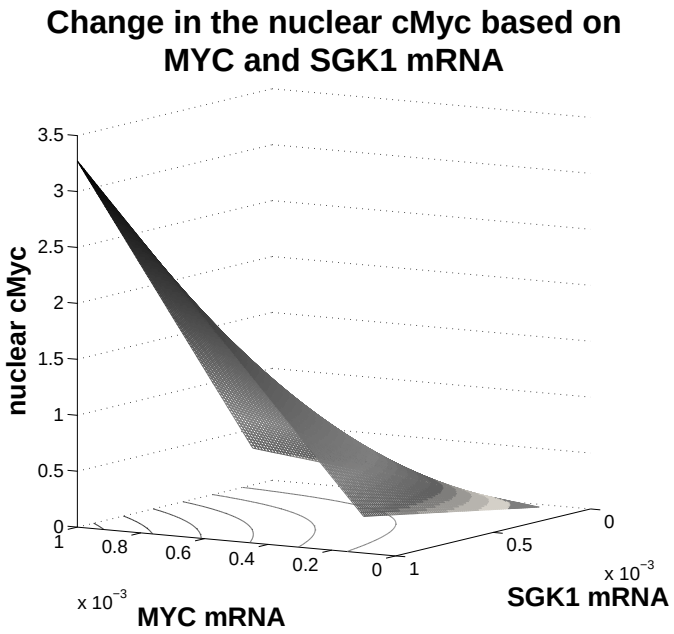

Figure 5.17: Relationship between the cMyc protein in the nucleus, the *MYC* and *SGK1* mRNAs obtained through the numerical simulations of the dynamics of the model with different parameter sets corresponding to the different virtual patients. We note the almost linear relationship between the *SGK1* mRNA and the nuclear cMyc when the *MYC* mRNA is close to 0.

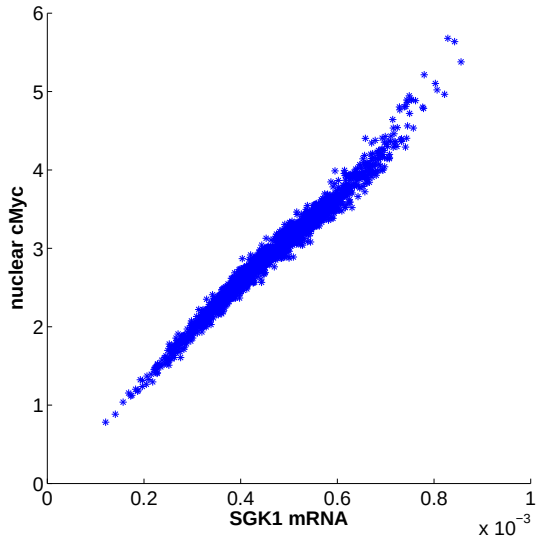

Figure 5.18: Correlation between the level of the cMyc protein in the nucleus and *SGK1* mRNA obtained through the numerical simulations for varying values of parameters t_1 and t_3. We obtain positive correlation between the *SGK1* mRNA expression and levels of the nuclear cMyc protein.

5.2.4 Correlation between the *SGK1* mRNA production and cMyc protein level in the nucleus

To estimate the correlation between the *SGK1* mRNA production and the nuclear cMyc level in the medulloblastoma samples, we perform the following numerical experiment. We fix the point in time for which we test the correlation. We vary the values of the rates, t_1 and t_3, by adding to them the pseudorandom values drawn from the standard normal distribution. We plot 500 points which relate to the number of vector entries of the randomly varied parameters. We obtain positive correlation. For higher values of the *SGK1* mRNA production we have higher values of the nuclear cMyc. The inference is that the *SGK1* production increases cMyc protein level in the nucleus (see Figure 5.18).

5.2.5 Role of inhibition in the SGK1 protein

In the case of critically high cMyc in the nucleus, we use the model to study the effect of the reduction of the protein amount. There are several options to obtain such effect. From the practical point of view, the best medical option is to apply a pharmacological inhibitor. The effect of inhibition (decreased activity of the protein) is in our system modeled by an increased degradation of the protein. Numerical simulations predict that increased degradation of the SGK1 protein (Figure 5.19[A]) leads to a significant decrease of cMyc levels in the nucleus. Moreover, using the model we can compare the effects of different pharmacological strategies. Therefore, we also simulate system (3.14) - (3.22) with the cMyc level reduction (Figure 5.19[B]) by increased degradation. Then, we show that indeed inhibition of the SGK1 protein is a more efficient way to decrease the level of cMyc than increased degradation of the cMyc protein. Importantly, the impact of SGK1 on stabilization of the cMyc protein is maintained through the degradation of GSK3β. SGK1 phosphorylates GSK3β resulting in its degradation, hence inhibition of SGK1 leads to increase in the GSK3β concentration. The consequence of the latter is higher degradation of cMyc by GSK3β. The degradation of cMyc itself is not sufficient enough as the upregulated SGK1 still stimulates the cMyc increase in the nucleus. Numerical simulations of the model reveal that the SGK1 inhibition is more efficient in decreasing the cMyc concentration.

We obtain that ten times increased SGK1 degradation yields a very good outcome, where level of the nuclear cMyc is almost equal to the control level (Table 5.13). The application of the SGK1 inhibition can be treated as an adjuvant treatment to the traditional one.

However, we should not undertake the same actions in monosomy 6, where we already have a low level of SGK1. Further decreasing of SGK1 could lead to unexpected mutations in the cell because of such low SGK1 level (Figure 5.20 and Table 5.14).

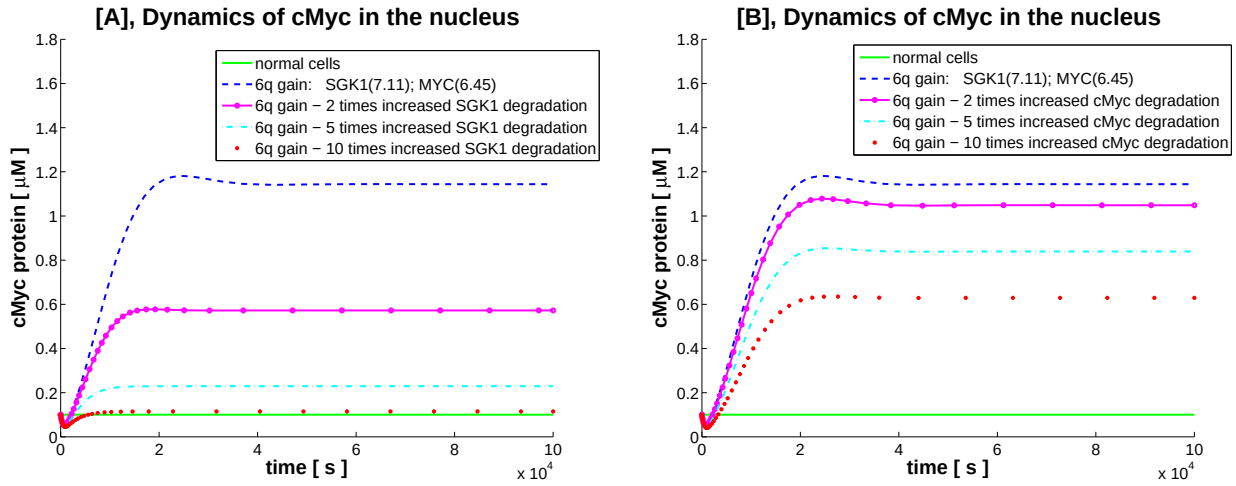

Figure 5.19: Numerical simulations based on exemplary microarray data values from the clinics. Effect of increased degradation [A] of SGK1 (removal of the SGK1 protein) and [B] of cMyc (removal of the cMyc protein); trisomy 6 sample (*SGK1* mRNA fold change is equal to 7.11, *MYC* mRNA fold change is equal to 6.45).

Fold change of the nuclear cMyc after SGK1 inhibition	
SGK1(7.11); MYC(6.45)	11
2 times inhibition	6
5 times inhibition	2
10 times inhibition	1

Fold change of the nuclear cMyc after cMyc inhibition	
SGK1(7.11); MYC(6.45)	11
2 times inhibition	10
5 times inhibition	8
10 times inhibition	6

Table 5.13: Fold change of the nuclear cMyc after SGK1 inhibition and cMyc inhibition. Trisomy 6 case.

Fold change of the nuclear cMyc (no inhibition)	
SGK1(0.74); MYC(16.27)	3
SGK1(0.19); MYC(20.5)	1
SGK1(0.44); MYC(10.02)	1

Fold change of the nuclear cMyc after 2 times SGK1 inhibition	
SGK1(0.74); MYC(16.27)	1.5
SGK1(0.19); MYC(20.5)	0.5
SGK1(0.44); MYC(10.02)	0.56

Table 5.14: Fold change of the nuclear cMyc before and after SGK1 inhibition. Monosomy 6 case.

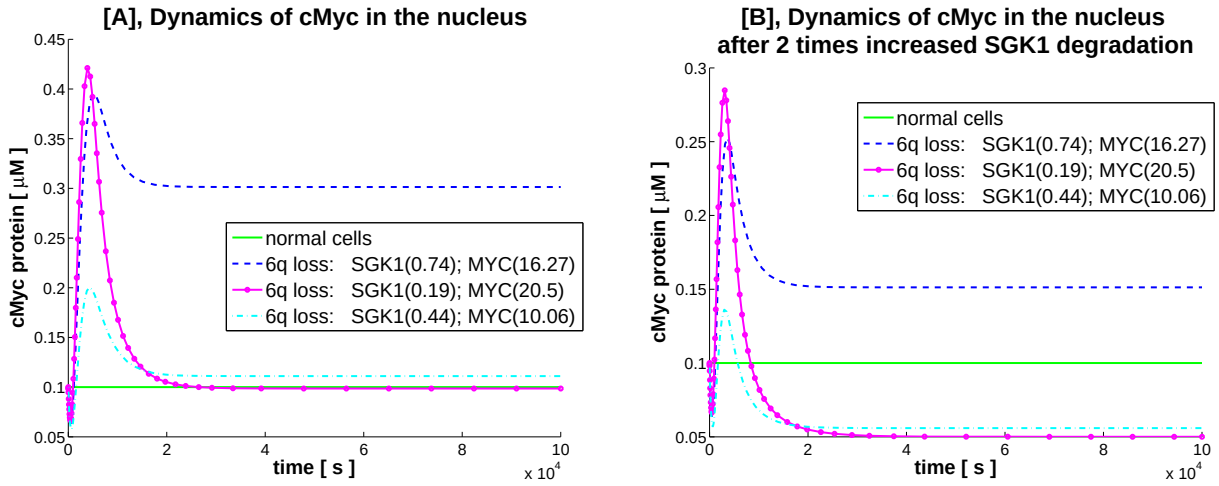

Figure 5.20: Numerical simulations based on exemplary microarray data values from the clinics. Comparison of the nuclear cMyc [A] before and [B] after SGK1 inhibition in the monosomy 6 different cases. SGK1 degradation increased two times.

Observation 5.2 *SGK1 inhibition is more efficient than cMyc inhibition to obtain low nuclear cMyc. Treatment in trisomy could be extended to inhibition of SGK1.*

5.2.6 Effect of the GSK3β protein stabilization

In parallel, we study the dynamics of GSK3β in the case of the cMyc and SGK1 inhibition. Simulations indicate that inhibition of cMyc does not influence the dynamics of GSK3β (Figure 5.21) and that GSK3β stays downregulated compared to the normal cells. It explains why the level of cMyc in the nucleus remains high in spite of inhibition. Simulations of the SGK1 inhibition reveal a positive influence on the levels of GSK3β (Figure 5.22 and Table 5.15), which become elevated. This in turn generates a decrease in cMyc level in the nucleus. The coupling loop *SGK1-GSK3β-MYC* seems to play an important role in both types of medulloblastoma, showing different properties between monosomy 6 and trisomy 6 with respect to aberrations in the *SGK1* and *MYC* mRNAs.

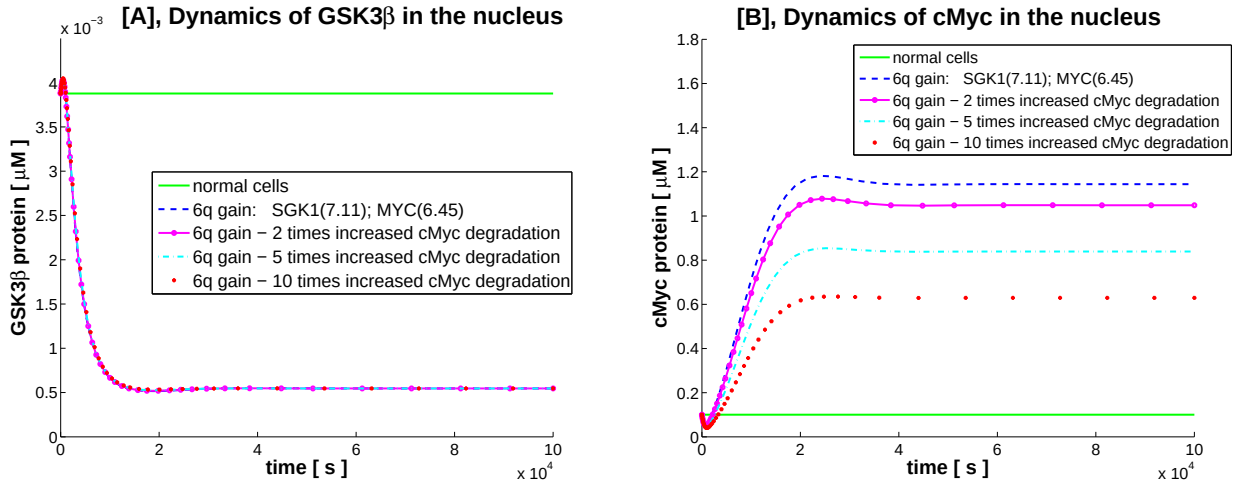

Figure 5.21: Numerical simulations based on exemplary microarray data values from the clinics. Influence of cMyc inhibition [A] on the GSK3β level, [B] on the nuclear cMyc.

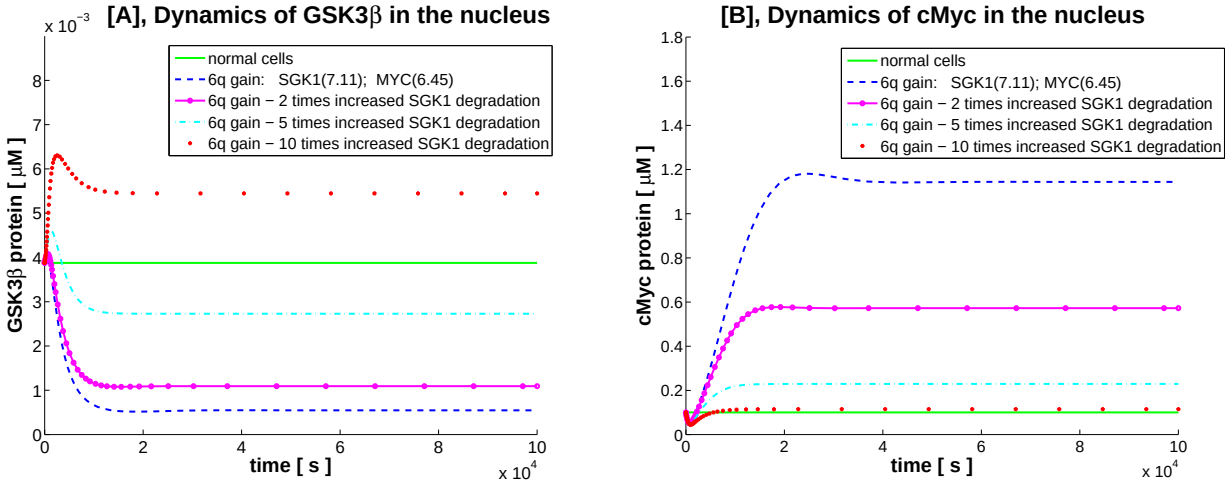

Figure 5.22: Numerical simulations based on exemplary microarray data values from the clinics. Influence of SGK1 inhibition [A] on the GSK3β level, [B] on the nuclear cMyc.

Observation 5.3 *Inhibition of SGK1 contributes to the balance of a wider range of proteins, e.g., GSK3β and cMyc.*

Fold change of the nuclear GSK3β after SGK1 inhibition	
SGK1(7.11); MYC(6.45)	0.14
2 times inhibition	0.28
5 times inhibition	0.7
10 times inhibition	1.4

Fold change of the nuclear cMyc after SGK1 inhibition	
SGK1(7.11); MYC(6.45)	11
2 times inhibition	6
5 times inhibition	2
10 times inhibition	1

Table 5.15: Fold change of GSK3β and cMyc in the case of SGK1 inhibition.

Summary of Chapter 5

In this chapter we presented numerical simulations of the ODE model discussed in previous chapters. We performed several tests to investigate the proteins behavior and interactions that occur on the intracellular level. Our study strongly emphasize the crucial role of *SGK1* in the process of biological homeostasis in the cell. We suggest that *SGK1* is the key gene indicating the discrepancy of prognosis in medulloblastoma and has big influence on the amount of the nuclear cMyc. Inhibition of the SGK1 protein is much more efficient procedure than inhibition of cMyc to decrease the nuclear cMyc. Additionally, reducing the SGK1 protein restores also the balance of the GSK3β protein in the system. All facts assert that high *SGK1* leads to negative prognosis and its inhibition brings positive results. We hope that inhibition of the SGK1 protein can be used as an adjuvant therapy for the patients with trisomy 6 diagnosis.

CHAPTER 6

Parameter estimation and optimal experimental design

This chapter is devoted to the identification of model parameters and application of optimal experimental design. Complexity of the biological model can strongly influence the robustness of the optimal parameter estimation. The goal is to have a predictive model to investigate the biological environment according to different biological signals. To perform good system predictions the parameters should be identified with the smallest possible uncertainties taking into consideration the model dynamics, which reflect the biological behavior.

Our goal is to estimate parameters by fitting the model to the experimental data to obtain the minimal costs of experimentation and enhance the quality of estimation.

Nowadays, processing techniques have improved a lot and more methods in data collection are found. However, large amount of data do not assure the best experimental set-up to infer estimates that are the closest to the real parameter values. Very often, in non-optimal designs there is less useful information and much more experimental runs must be performed. In consequence of latter, estimation of parameters is a laborious process and sometimes it can happen that parameters are obtained with low precision. With optimal experimental design the scheme is optimized in a way that taken measurements give the best estimate with the lowest cost.

6.1 Parameter estimation problem

In this section we approach the parameter estimation problem and the Gauss-Newton method, which provides a numerical solution of the stated problem. We introduce the solution operator, the observation operator as well as the cost functional, which quantifies

the difference between experimental and simulated data. To define the Gauss-Newton method, we derive the gradient and the approximation of the Hessian of the least squares functional and finally we present a calculation procedure in subsequent steps of the algorithm.

6.1.1 General formulation

We consider model (3.14) - (3.22) of biochemical interactions in the form of

$$\dot{u} = F(u, q) \tag{6.1}$$

in medulloblastoma tumor cells (see Figure 2.12), where q is a finite number of unknown model parameters q := (q_1,...,q_m) $\in Q \subset \mathbb{R}^m$ with m = 19. The concentration of each species is represented by the solution u, $F := [f_1, ..., f_n]^T$, $F : G \to \mathbb{R}^n$, where $G := U \times Q$ and $F \in C^1(U \times Q)$. Model solutions depend on time and model parameters $U : [0, \infty) \times Q \to \mathbb{R}^n$ with n = 9 . Theorem 4.1 yields $U \in C^1([0, \infty) \times Q)$. Table 6.1 shows the respective parameters used. The experimental measurements are gathered into a vector $\bar{C}$, which is in the observation space $D \subset \mathbb{R}^l$. It is crucial to satisfy $l \geq m$, [66], to be able to determine the model parameters. Further, we introduce the solution operator $S(q) : Q \to U$, which maps model parameters to the solution space. The observation operator $C(u) : U \to D$ maps the solution to the observation space. Here we use the point measurement $C(u) := u(t_k)$, where t_k is the time point in which measurement is taken, $k \in \mathbb{N}$. From practical point of view, we are aware that each of the measured values have a random error which we call perturbation $p \in P \subset D$. Consequently, the measurement values are $\bar{C} = \bar{C}_{true} + p$, i.e., the sum of real value and measurement error.

To evaluate the deviation between experimental measurements $\bar{C}$ and state values $C(u)$ dependent on parameters q, we apply the cost functional $J(u, q) : U \times Q \to \mathbb{R}$. Finally, we can formulate the perturbed parameter estimation problem in the following form.

Problem 6.1 *To find the optimal solution for model parameters the minimization of the cost functional under constraints is carried out. That reads*

$$\begin{cases} \min\limits_{(u,q)\in(U\times Q)} J(u, q), \\ s.t. \quad \dot{u} = F(u, q). \end{cases} \tag{6.2}$$

From Problem (6.2) it follows that we compare experimental data with data obtained from the model. Introducing the weighted least squares functional and using the reduced formulation [6], i.e., $\hat{C}(q) := C(S(q))$, where $u := S(q)$, we can rewrite the statement (6.2) into

$$\min_{q\in Q} \tfrac{1}{2} \left\| \hat{J}(q) \right\|_D^2. \tag{6.3}$$

Notation for estimates	**Notation of parameters used in model**	**Biological meaning**
q_1	t_1	transcription rate of *SGK1*
q_2	β_{cat}	transcription factor for the *MYC* and *SGK1* mRNA synthesis
q_3	d_1	spontaneous degradation rate of the *SGK1* mRNA
q_4	s_2	translation rate of SGK1
q_5	d_2	spontaneous degradation rate of the SGK1 protein
q_6	d_6	degradation rate of the GSK3β protein in the cytoplasm by SGK1
q_7	p_9	dissociation coefficients of the [SGK1·GSK3] complex
q_8	t_3	transcription rate of *MYC*
q_9	d_3	spontaneous degradation rate of the *MYC* mRNA
q_{10}	s_4	translation rate of cMyc
q_{11}	c_4	transport of cMyc to the nucleus coefficient
q_{12}	k_v	scaling coefficient, cytoplasm to the nucleus ratio
q_{13}	d_5	degradation rate of the cMyc protein in the nucleus by GSK3β
q_{14}	P_{GSK}	constant influx of GSK3β to the cytoplasm
q_{15}	c_7	transport of GSK3β to the cytoplasm coefficient
q_{16}	c_6	transport of GSK3β to the nucleus coefficient
q_{17}	d_o	spontaneous degradation rate of GSK3β in the cytoplasm
q_{18}	p_8	dissociation coefficients of the [GSK3· cMyc] complex
q_{19}	d_4	spontaneous degradation rate of cMyc in the cytoplasm

Table 6.1: Description of model variables. From left to right: notation used to define parameters to estimate, equivalent variables in ODE model (3.14) - (3.22) and biological meaning.

$\hat{J}(q) := \hat{C}(q) - \bar{C}$ is a reduced cost functional, where $\hat{C}(q) - \bar{C}$ is called a residual vector and $\|\cdot\|_D^2 = (C_D^{-1}\cdot, \cdot)$ is a weighted norm with weighting matrix C_D^{-1} (see Section 6.2.2), which corresponds to the covariance matrix of the measurements [17], [73]. Additionally, we assume that measurement errors are independent and normally distributed, which justifies the use of the least squares functional [9].

6.1.2 Solving the nonlinear problem

Problem 6.2 *Our goal is to solve the nonlinear least squares (LS) problem in reduced formulation*

$$\min_{q \in Q} \tfrac{1}{2} \big\| \hat{J}(q) \big\|_D^2. \tag{6.4}$$

We solve the optimization Problem 6.4 applying the Gauss-Newton method iteratively [34]. To find a local minimum of reduced formulation of $\hat{J}(q)$, we choose an initial guess of q^0. Our problem can be stated in the following way.

Algorithm 6.1 *Calculate q iteratively starting with q^0 and make update by setting $q^{i+1} = q^i + \alpha^i \Delta q^i$, where α^i is the stepsize and Δq^i is the search direction defined by the gradient and approximation of the Hessian in each iteration i, where $i = 0, 1, 2...$*

Convergence is assured for a small residual problem and obtained updating q by Δq, which is expressed by

$$\Delta q := -\big(\nabla^2 \hat{J}(q)\big)^{-1} \nabla \hat{J}(q), \tag{6.5}$$

where $\nabla \hat{J}(q)$ is the gradient of the cost functional $\hat{J}(q)$ in the form

$$\nabla \hat{J}(q) := \hat{G}(q)^T C_D^{-1} \big(\hat{C}(q) - \bar{C}\big). \tag{6.6}$$

$\hat{G}(q)_{ij} := \frac{\partial \hat{C}_i(q)}{\partial q_j}$ is the Jacobian (see Section 6.3.1).

The term $\nabla^2 \hat{J}(q)$ is the approximation of the Hessian of the cost functional $\hat{J}(q)$, where the second order derivatives are neglected [8]

$$\nabla^2 \hat{J}(q) := \hat{G}(q)^T C_D^{-1} \hat{G}(q). \tag{6.7}$$

We write the update of the Gauss-Newton problem by inserting (6.6) and (6.7) in (6.5).

Definition 6.1 *The Gauss-Newton update*

$$\begin{cases} q^{i+1} &= q^i + \alpha^i \Delta q^i, \\ \Delta q^i &= -\big(\hat{G}(q)^T C_D^{-1} \hat{G}(q)\big)^{-1} \hat{G}(q)^T C_D^{-1} \big(\hat{C}(q) - \bar{C}\big). \end{cases} \tag{6.8}$$

In each iteration the condition that the new calculated value of the LS functional is lower than the one calculated in previous iteration is imposed [27]

$$LS(q^{i+1}) < LS(q^i). \tag{6.9}$$

If condition (6.9) is not fulfilled, the backtracking line search strategy is used, i.e., in the $i\text{-}th$ iteration α^i is updated as follows. We choose an initial stepsize $\alpha_0^i = 1$ and a $\varrho \in (0, 1)$. For iterations $j = 0, 1, ...$ we update $\alpha_j^i = \alpha_0^i \varrho^j$ (ϱ^j is ϱ to the power of j) until the condition (6.9) is fulfilled. If fulfilled, q^{i+1} is calculated for the new α_j^i. In our application we used ϱ = 0.5.

We proceed updates defined in (6.8) till the main breaking condition is satisfied, which is when Δq is lower than assumed accuracy ϵ_1. The convergence is obtained and next steps would bring only quite low variation in parameter value, see [48].

Remark 6.1 *The Gauss-Newton method is a local optimization algorithm. If there are several local minima, several q^0 should be chosen to assure the global solution.*

The scheme of the Gauss-Newton method algorithm is presented in Table 6.2. We use the Gauss-Newton method in the next sections to find the model parameter estimates systematically, highlighting different aspects of parameter estimation problem.

The Gauss-Newton nonlinear least squares algorithm

Initialize:
  initial guess of q^0, initial residual - calculate LS value at q^0, $\varrho = 0.5$
Global loop:
  calculate $\hat{G}(q)$ for q^i
  calculate gradient based on $\hat{G}(q)$
  calculate approximation of the Hessian based on $\hat{G}(q)$
  calculate Δq
  set $\alpha_0^i = 1$
  update q^{i+1}
  evaluate residuals - calculate LS value at q^{i+1}
  Local loop:
    break loop if LS(q^{i+1}) < LS(q^i)
    update α_j^i
    update q^{i+1}
    evaluate residuals - calculate LS value at updated q^{i+1}
    increment local iteration counter j
    error exit if maximal number of iterations exceeded
  End local loop
  increment global iteration counter i
  error exit if maximal number of iterations exceeded
  normal exit if ($|\Delta q| < \epsilon_1$)
End global loop

Table 6.2: Pseudocode of the Gauss-Newton algorithm.

6.2 Confidence region in the form of ellipses

The goal of this section is to feature the probabilistic character of the parameter estimation problem and subsequently to show its graphical interpretation. We introduce the formula of *covariance matrix* that allows us to draw the *confidence region*. The confidence region is a region of parameter uncertainties around the estimated parameters q_i, q_j $(i \neq j)$, i.e., the confidence region is evaluated for two different parameters. This graphical representation is in the form of an *ellipse*, also denominated here *confidence ellipse*. The confidence region is a great tool to visualize the reliability of an estimated solution.

6.2.1 Definition of covariance matrix

Taking into account the probabilistic representation of perturbations p, we have the Gaussian nature of measurement error $e_M := p \sim \mathcal{N}(0, \sigma^2)$, that stands for the normal distribution with expected value of the measurement error equal zero and variance is given by the weighting matrix C_D with constant uncorellated variance $var(e_M) = \sigma^2$. We assume that there are no outliers [20].

Definition 6.2 *Measurement error follows an accuracy of which measurement is extracted by the experimentalist.*

To see, how the estimates depend on the measurement error we perform the covariance analysis. In the following [74], the covariance matrix is assumed to be

$$\mathrm{Cov} := \left(\hat{G}^T C_D^{-1} \hat{G}\right)^{-1}, \tag{6.10}$$

where $\hat{G}$, introduced in Subsection 6.1.2, represents the Jacobian with calculated sensitivities.

6.2.2 Geometrical interpretation of confidence region

Based on Cov we can interpret the parameter estimation problem geometrically, taking into account the probabilistic character of the parameter estimation problem [1], [4], [74]. Geometrical representation which approximates the statistical distribution of the estimated parameters is in the form of an ellipse [20], [59].

The ellipse is centered at searched value q_i, q_j and the uncertainties of estimated parameters are denoted by axes. We represent the ellipses in a coordinate system where lengths of principal axes correspond to the square root of eigenvalues $\sqrt{\lambda}$ of the covariance matrix Cov for particular parameters. $\sqrt{\lambda_{max}}$ and $\sqrt{\lambda_{min}}$ stand for the longer and shorter ellipse axis, respectively. The ellipses are derived from the concept of a 95% confidence interval for a normally distributed random value [17]. We present the statistical

results on parameter error estimation in the form of ellipses in Figure 6.1, where we set $i = 1, j = 2$.

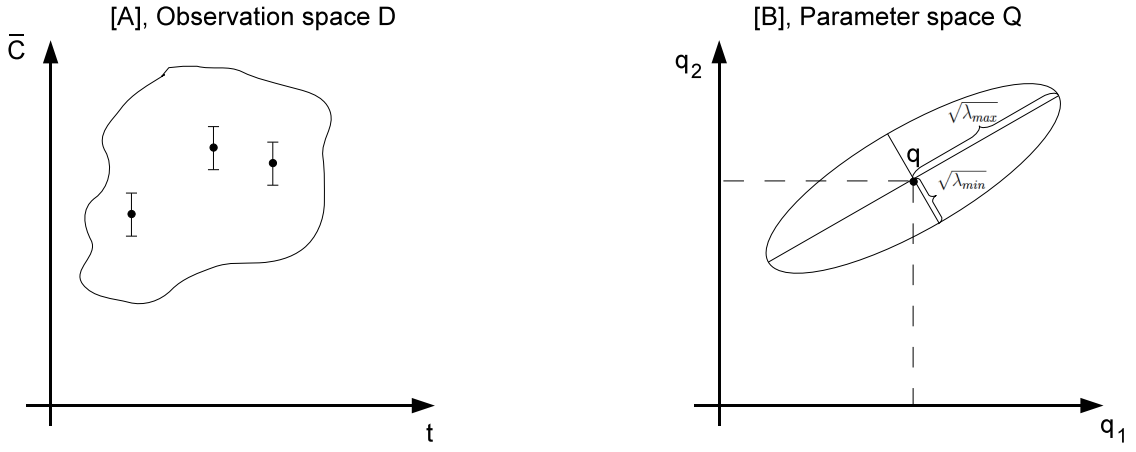

Figure 6.1: Graphical representation of the parameter estimation problem. [A] Measurements sampled in particular time point projected into [B] parameter space Q. The uncertainties of estimated parameters depend on the measurement error and sampling time.

Confidence ellipses visualize how the variance of measurement errors influences the error in parameter estimation. The variance of the measured data is expressed in (6.10) by C_D, whereas $\hat{G}$ is the Jacobian (see Section 6.3.1). Measurements are taken for different time probes t_k and therefore the Jacobian depends also on the time points of measurements. If we find good measurement points, the uncertainties of estimated parameters are small. If the uncertainties are below an expected threshold, which depends on the application, then we are satisfied with the estimates.

If measurements of two species are independent and have the same error, the data covariance matrix is $C_D = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The second option is that, measurements are independent but errors differ, thus $C_D = \sigma^2 \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ with a,b > 0 and a $\neq$ b. If measurements are correlated and sampled with different error the covariance matrix is not diagonal and reads $C_D = \sigma^2 \begin{pmatrix} a & c \\ d & b \end{pmatrix}$ with a,b,c,d > 0 and a $\neq$ b $\neq$ c $\neq$ d. The experimentalist may assume or predict the measurement error and may know if data are correlated or not. In our work we assume that measurements are independent and have the same error.

Interestingly, depending on the measurement points the ellipses are more or less slant-

ing in the coordinate system. The shape of ellipses, so alignment of the principal axes with respect to the parameter axes, describe the correlation between parameters. More slant-

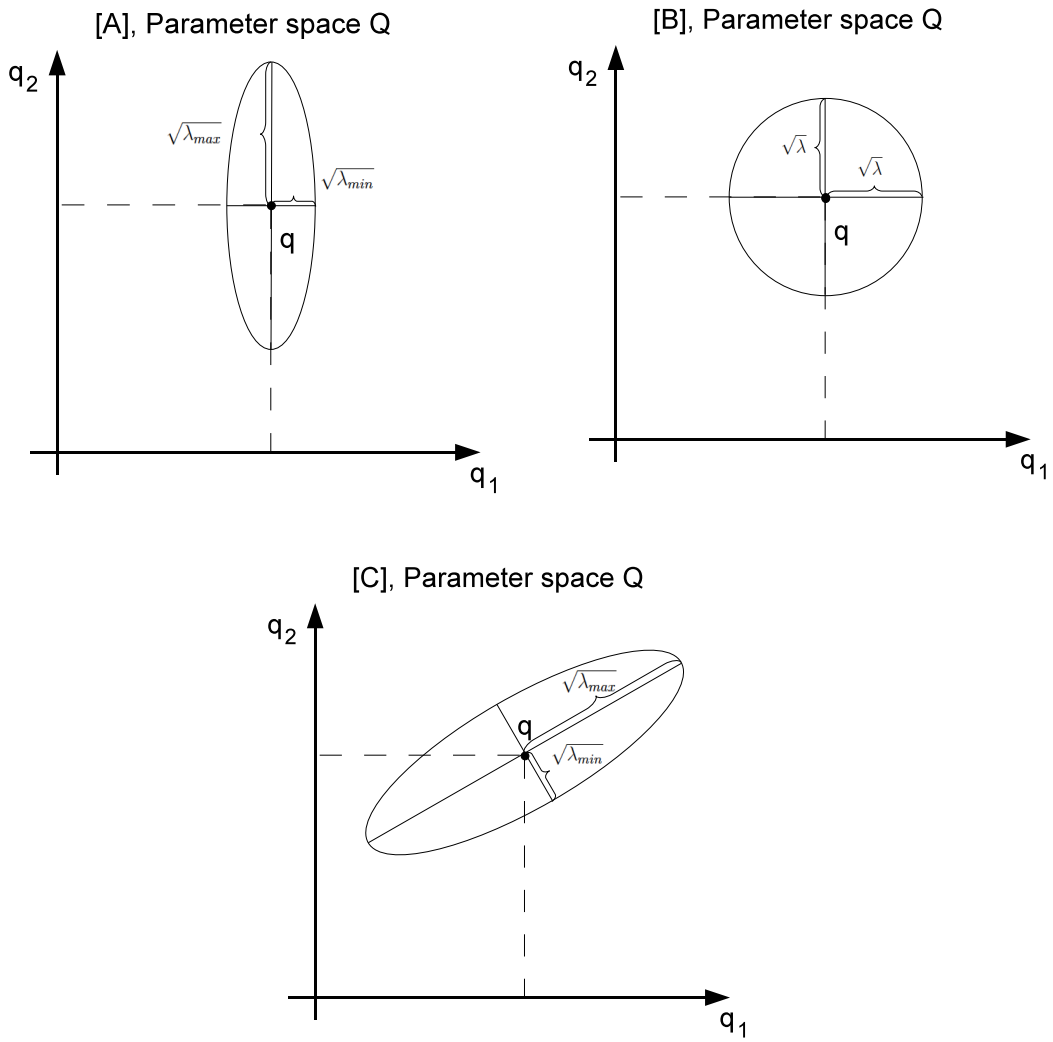

Figure 6.2: 95% confidence ellipses. [A] The parameters are not correlated, the estimation of parameter q_1 is better than estimation of parameter q_2, [B] parameters are not correlated, the error in estimation is the same for both parameters, [C] parameters are correlated and parameter q_1 is worse in estimation than parameter q_2.

ing ellipses refer to the parameters that are more correlated due to chosen measurement points, what shows that the error in one parameter influences the error in the correlated parameter. Following that $\mathrm{Cov} = \left(\hat{G}^T C_D^{-1} \hat{G}\right)^{-1}$ describes coupling between measurement error and probes depending on sampling time, we plot types of the confidence ellipses corresponding to different parameter correlations (Figure 6.2).

The graphical method of ellipses is the linearized approximation of the problem, thus to check its reliability, we also perform Monte Carlo simulations using the Gauss-Newton method described in Section 6.1 and compare both approaches in one plot.

Definition 6.3 *The Monte Carlo method is a numerical method based on random sampling of the data [47].*

6.3 Parameter estimation in medulloblastoma model

In this section we investigate different aspects of the parameter estimation problem. The medulloblastoma model consists of nonlinear ODEs and depends on unknown paramaters, which can be estimated if we take advantage of experimental data. We show the behavior and difficulty in estimating the parameters of the medulloblastoma model (3.14) - (3.22).

We perform our analysis for two species, i.e., the concentration of the nuclear cMyc protein and the concentration of occupied GSK3β in the nucleus, respectively. We take measurements to identify the parameters of the model. We do not have real experimental data, therefore we use randomly perturbed simulations as surrogate data.

In Subsection 6.3.1 we approach the parameter sensitivity analysis problem. Then, in Subsection 6.3.2 we deal with parameter nonlinearities, coupling between parameters and the influence of measurement error on estimation. In Subsection 6.3.3 we raise the problem of impact on estimation of variation in parameter values, which account for the parameters found in the literature, often presented by range in values.

At the end of each subsection we make a table with reliable estimates.

6.3.1 Parameter sensitivity analysis

Parameter sensitivity analysis [11], [33], is one of the steps to examine the parameter estimation problem. The measurements collected using some specific experimental protocol may not allow for the reliable parameter estimation that can be highly uncertain. We investigate if parameters are *weak* or *strong* sensitive to the measurements accounting for a chosen sampling protocol. If we have a good sampling protocol our estimation bring positive results and the behavior of the model for given set of parameters can explain the measurements. For future goals the model with well established parameters can predict dynamics under different inputs.

Definition 6.4 *Sensitivity analysis in a parameter estimation problem is the research of the sampling protocol, i.e., the time points for which model shows the highest sensitivity. This corresponds to the time point in which the model contributes the highest information to the estimation process.*

Problem 6.3 *Find model species $\hat{C}_i$ for which the sensitivity is the highest. Find time points for which particular parameters have the highest sensitivity.*

Definition 6.5 *We define a Jacobian sensitivity matrix $\hat{G}(q)_{ij}$ of the state $\hat{C}_i(q)$ with respect to parameter q by a change in q_j, where i = 1,...,n and j = 1,...,m.*

$$\hat{G}(q)_{ij} := \frac{\partial \hat{C}_i(q)}{\partial q_j} \tag{6.11}$$

The definition can be found in [74]. Differentiation of $\hat{G}(q)_{ij}$ gives the *system of sensitivities*

$$\frac{d}{dt}\left(\frac{\partial \hat{C}_i(q)}{\partial q_j}\right) = \sum_{i=1}^{n} \frac{\partial f_i(\hat{C}, q, t)}{\partial \hat{C}_i} \frac{\partial \hat{C}_i}{\partial q_j} + \frac{\partial f_i(\hat{C}, q, t)}{\partial q_j}. \tag{6.12}$$

Remark 6.2 *The condition number of an $m \times m$ matrix A is $cond(A) := \|A\|\|A\|^{-1}$ and is a measure of how close a matrix is to being singular. If $cond(A)$ is large depending on the operating system, e.g., greater than 1e15, then numerically matrix A is singular.*

To improve the condition number, we introduce a scaling $q = \bar{q} \cdot q_{ref}$ (also applied in model equations (3.14) - (3.22)) and reformulate (6.12) into

$$\frac{d}{dt}\left(\frac{\partial \hat{C}_i(\bar{q} \cdot q_{ref})}{\partial (\bar{q} \cdot q_{ref})_j}\right) = \sum_{i=1}^{n} \frac{\partial f_i(\hat{C}, \bar{q} \cdot q_{ref}, t)}{\partial \hat{C}_i} \cdot \frac{\partial \hat{C}_i}{\partial (\bar{q} \cdot q_{ref})_j} + \frac{\partial f_i(\hat{C}, \bar{q} \cdot q_{ref}, t)}{\partial (\bar{q} \cdot q_{ref})_j} \cdot q_{ref_j}, \tag{6.13}$$

where $\bar{q} = 1$ is an initial guess, $q_{ref} = const.$ is a reference parameter and q is a rescaled parameter.

Remark 6.3 *For future application when we write about model species, we stick to the notation C_i for the simplicity and to emphasize that measurements are mapped to the observation space, $C(u_i) = u(t_k)$.*

Numerical simulations

To solve the sensitivity problem, we calculate the sensitivities by numerical evaluation of (6.1) and (6.13). To check if our sensitivities are well calculated we use a finite difference method [69] and we compare both numerical solutions in one plot. In our model, where n = 9 and m = 19, we obtain $180 = (19 \cdot 9)_{sensitivity\_eqs} + 9_{model\_eqs}$ equations to be simulated in MATLAB®. We check for which species and for which parameters the system reveals the highest sensitivity. To find parameters with the highest absolute amplitude of sensitivity, we perform simulations for species C_5 (the concentration of the nuclear cMyc protein) found to have the highest sensitivity.

In our model q_5 = d_6 (degradation rate of the SGK1 protein) is the most sensitive parameter with respect to species C_5. The parameters $q_7 = p_9$ (dissociation coefficients of [SGK1·GSK3β]) and $q_{17} = d_o$ (spontaneous degradation rate of the GSK3β protein in the cytoplasm) show sensitivity close to zero. This suggests that we are not able to estimate all parameters. The absolute amplitude of sensitivity is denoted by $\mathcal{AS}$. The amplitudes of sensitivity of C_5 to the parameters are given in Table 6.3.

Parameter	$\mathcal{AS}$	**Parameter**	$\mathcal{AS}$
q_1	0.32	q_{11}	0.31
q_2	2.67	q_{12}	2.41
q_3	3.45	q_{13}	1.26
q_4	1.67	q_{14}	1.69
q_5	8.75	q_{15}	1.02
q_6	1.67	q_{16}	1.67
q_7	$\sim$0	q_{17}	$\sim$0
q_8	2.64	q_{18}	0.48
q_9	2.38	q_{19}	0.23
q_{10}	2.64		

Table 6.3: The absolute amplitude of sensitivity $\mathcal{AS}$ for model parameters regarding species C_5. Parameter q_5 shows the highest sensitivity.

Now, we investigate how the chosen measurement time points influence the accuracy of estimates. We start our analysis with only three parameters. We fix other parameters, by deleting the corresponding row and column from the species Jacobian matrix to calculate the proper Cov to plot ellipses. We choose $m_c = 3 < m$ parameters, i.e., parameter q_5 with the highest sensitivity, arbitrary parameter $q_{13} = d_5$ (degradation rate of the cMyc protein in the nucleus by GSK3β) and arbitrary parameter $q_{14} = P_{GSK}$ (constant influx of the GSK3β protein to the cytoplasm). For m_c = 3 we choose $l = 3$ time points of the

measurements to fulfill $l \geq m_c$ and perform numerical comparison based on the method of ellipses. We assume a measurement error e_M = 10% represented by the covariance matrix C_D. From practical point of view too little error is not realistic and too large does not bring reliable results. We choose the following time points measured in seconds:
t_1 = 12000, t_2 = 19000, t_3 = 20000, t_4 = 21000, t_5 = 23000,
where each of time point represents different amplitudes of sensitivity.

Definition 6.6 *We call the design parameter the parameter, which describes the sampling protocol. Here, it is the time $t_{k,k=1,...,l}$, where t_l corresponds to the last time point of the experiment and we define the design parameter as $\xi \in \Pi$, where Π is a design space and $\xi = \{t_1, ..., t_l\}$.*

We prepare the *experiment design* using the design parameter $\xi = \{t_v, t_2, t_5\}$, where t_2 and t_5 are fixed. We vary $t_{v,v=\{1,3,4\}}$ to investigate different sensitivity scenarios for t_1, t_3 and t_4, where the inequality $\mathcal{AS}(t_1) < \mathcal{AS}(t_3) < \mathcal{AS}(t_4)$ holds (see Figure 6.3 to Figure 6.5), i.e., amplitude of sensitivity in point t_1 is lower than amplitude of sensitivity in point t_3, and the amplitude of sensitivity in point t_3 is lower than amplitude of sensitivity in point t_4.

We describe the design sets:

L: $t_v = t_1$ chosen for low amplitude $\mathcal{AS}$ of sensitivity,

M: $t_v = t_3$ chosen for middle amplitude $\mathcal{AS}$ of sensitivity,

H: $t_v = t_4$ chosen for high amplitude $\mathcal{AS}$ of sensitivity,

for our model

L: $\xi = \{12000, 19000, 23000\}$,

M: $\xi = \{20000, 19000, 23000\}$,

H: $\xi = \{21000, 19000, 23000\}$.

The sensitivities with different measurement points are plotted in Figure 6.3 to Figure 6.5 and corresponding ellipses for parameter estimation problem are in Figure 6.6 to Figure 6.8. For chosen three parameters we extract data only from species C_5.

Remark 6.4 *We call set Q_{ad} be a set of admissible, in sense possible, parameters q that we can estimate, where $q_{min} < q < q_{max}, q_{min} = 0, q_{max} = 2q$.*

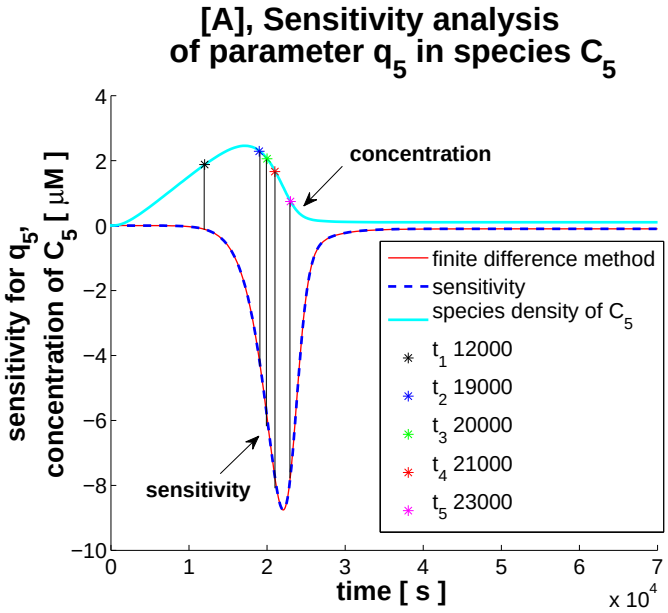

Figure 6.3: Sensitivity with time points for species C_5 in analysis of [A] parameter q_5.

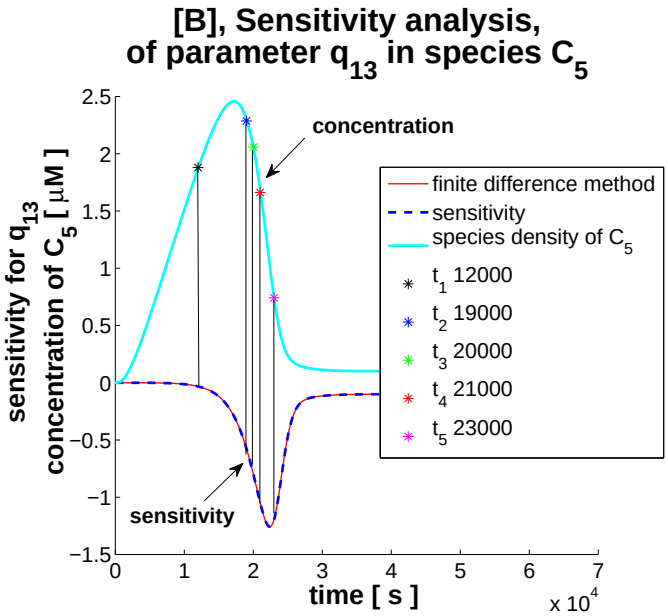

Figure 6.4: Sensitivity with time points for species C_5 in analysis of [B] parameter q_{13}.

On the first glance, we perceive that estimation for three chosen parameters is possible, as the error is smaller than 100%, and hence, we do not obtain negative values for the estimates. For different design sets we obtain better or worse identifications. Therefore,

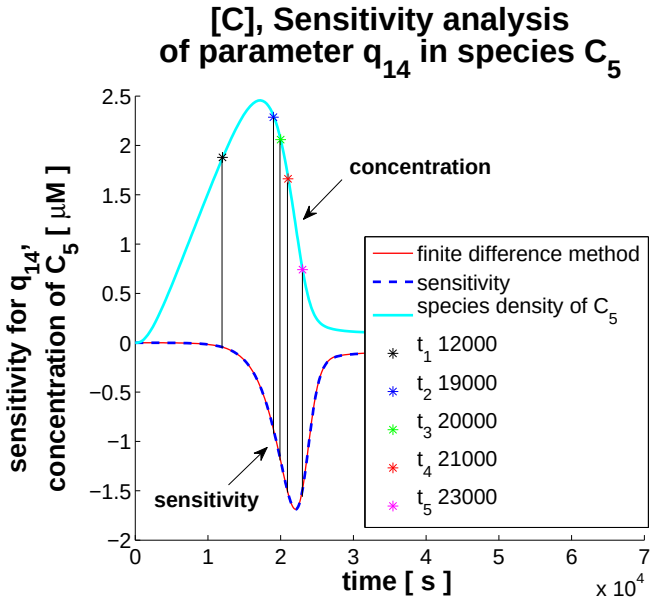

Figure 6.5: Sensitivity with time points for species C_5 in analysis of [C] parameter q_{14}.

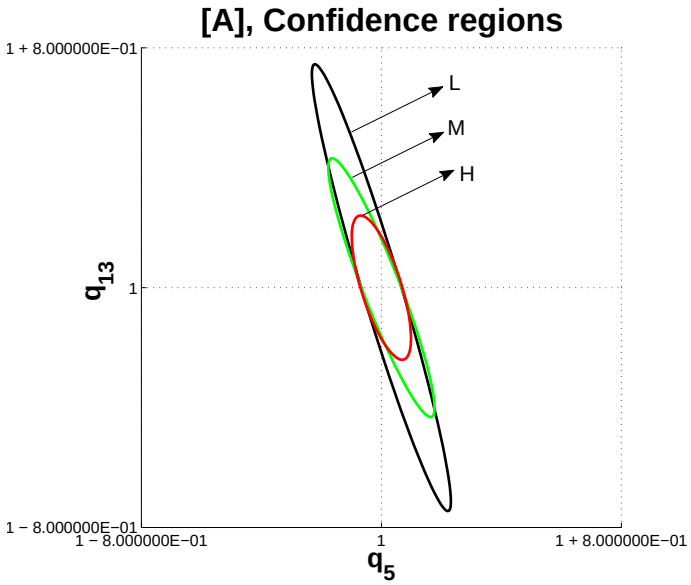

Figure 6.6: Comparison of 95% confidence regions: case H, M, L for C_5 for design parameters H: $\xi = \{19000, 23000, 21000\}$, M: $\xi = \{19000, 23000, 20000\}$, L: $\xi = \{19000, 23000, 12000\}$. [A] parameters q_5 and q_{13}.

we see how choice of the sampling time points is important for the estimation problem. For species C_5 the set H with sampling point of high $\mathcal{AS}$, so $t_v = t_3 = 20000$, depicts the

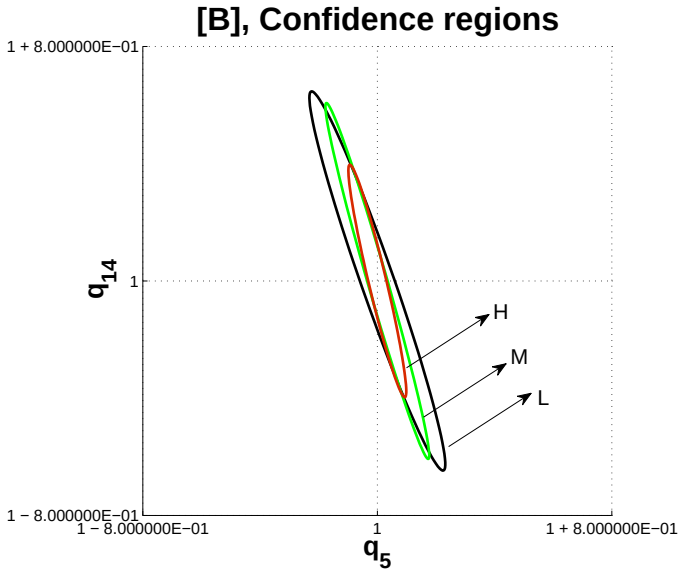

Figure 6.7: Comparison of 95% confidence regions: case H, M, L for C_5 for design parameters H: $\xi = \{19000, 23000, 21000\}$, M: $\xi = \{19000, 23000, 20000\}$, L: $\xi = \{19000, 23000, 12000\}$. [B] parameters q_5 and q_{14}.

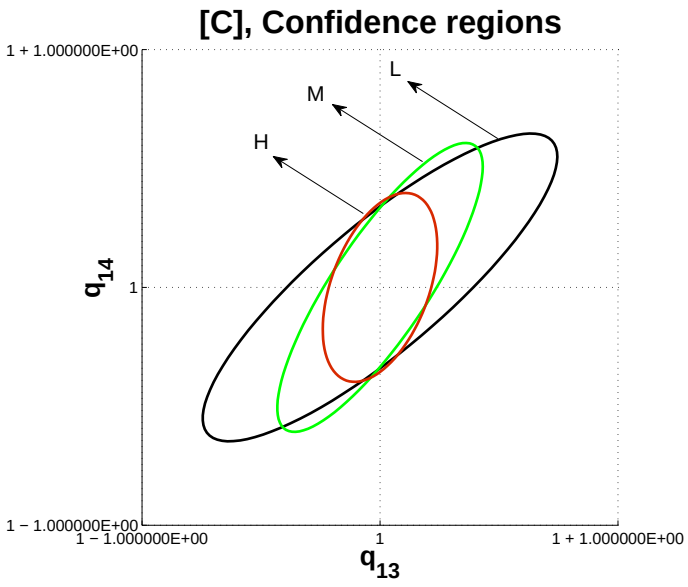

Figure 6.8: Comparison of 95% confidence regions: case H, M, L for C_5 for design parameters H: $\xi = \{19000, 23000, 21000\}$, M: $\xi = \{19000, 23000, 20000\}$, L: $\xi = \{19000, 23000, 12000\}$. [C] parameters q_{13} and q_{14}.

smallest ellipses, where parameters are estimated with the smallest uncertainty. Following this argumentation, the biggest ellipses are for set L with sampling time point of low $\mathcal{AS}$, i.e., $t_v = t_1 = 12000$ is taken for low sensitivity. The estimation is the worst in this case. We notice clear difference in the ellipses size when adapting t_v for different sensitivity scenarios.

Further investigation of the relationship between particular parameters is appealing as it also corresponds to the absolute amplitude of the sensitivity $\mathcal{AS}$. Parameter q_5 in comparison to other parameters is identified with higher accuracy, where the amplitude $\mathcal{AS}$ is the highest (Figure 6.6, Figure 6.7). Comparing parameters q_{13} and q_{14} we see that q_{13} is estimated with lower precision (Figure 6.8). This reflects that absolute amplitude $\mathcal{AS}$ is lower for q_{13} than for q_{14}. Next, it can be noticed that for the measurement error of 10% the error in estimation of parameter q_5, parameter q_{13} and parameter q_{14} is less than 100% (see Remark 6.4) for a very little number of measurements $l = 3$. Estimation of the three chosen parameters for the model is acceptable and parameter estimation problem is not ill-posed, [28]. In next sections we observe that for more measurement points and for two species C_5 and C_8 (the concentration of occupied GSK3β in the nucleus) the error in parameter estimation is much lower.

Choosing measurements points based only on the sensitivity analysis disregarding their correlation does not necessarily bring positive results. The problem with optimal sampling time points is more intricate, as coupling between the parameters must be taken into account. The second issue is that state variables are nonlinear with respect to the parameters. To learn more about this problem we refer to [74].

Short Summary 6.1 *In this subsection we tackled the problem of sensitivity analysis. We investigated the influence of different time probes on the estimation process taking into account sensitivity amplitude. Corollary of our study are the following estimates that will be used in the sequential design (see Section 6.4.1) in the context of optimal experimental design:*

Estimates	**Investigated in**
q_5, q_{13}, q_{14}	Subsection 6.3.1

6.3.2 Parameter nonlinearities, coupling and measurement error

In order to maximize the amount of information extracted from the experiments, we have to deal with the following problems.

Problem 6.4 *Nonlinearities of the system can produce nonconvexity [5] of the least squares functional. In consequence, we may obtain multimodality.*

Multimodality means that we do not have one global minimum and multiple local minima can exist. We have to also deal with coupling between model parameters.

Problem 6.5 *If parameter, e.g., q_1 is afflicted with an error and is coupled with parameter, e.g., q_2, then coupling of these parameters produces an error in parameter q_2.*

Problem 6.6 *The strength of coupling and nonlinearities rises with increase of the measurement error.*

Remark 6.5 *In our model, the measurement error is set to the samples of species C_5 and C_8.*

Remark 6.6 *If for certain parameters the nonlinearities are too big, then we cannot apply the method of ellipses (Section 6.2) to estimate the range of parameters.*

Numerical simulations

To show the role of nonlinearities in parameter estimation, we perform numerical simulations for parameters $q_2 = t_1$ (transcription rate of *SGK1*), $q_5 = d_2$ (spontaneous degradation rate of the SGK1 protein), $q_8 = t_3$ (transcription rate of *MYC*), $q_{18} = p_8$ (dissociation coefficients of the [GSK3β·cMyc] complex) and l = 41 (number of measurements). The method of ellipses and Monte Carlo simulation sampling with 50 points and e_M = 10% of measurement error for arbitrary chosen 41 measurement points for species C_5 and C_8 is applied (Figure 6.9, Figure 6.10 and Figure 6.11).

In the Monte Carlo simulation we first choose the measurement time points, so in consequence we have the corresponding measurement for each time point. The next step is to perturb the measurements by adding an error to see how this error influences the parameter estimation error. By Gauss-Newton algorithm we visualize estimates for 50 samples.

For estimation in parameters q_2 and q_8 (see Figure 6.9) we observe non symmetric distribution of the points which means that our model indicates strong nonlinearity with respect to these parameters. In other cases all points are well distributed. For different measurement points the model may or may not indicate vulnerability to nonlinearities, respectively.

The error of measurement is e_M = 10% and the error in parameter estimation varies according to the particular estimates. Comparing the chosen parameters we can see that accuracy in parameter estimation for q_5 and q_{18} is the biggest, and in fact, much bigger than for any other parameters. To improve accuracy of estimates the number of measurement points can be increased or "not estimable" parameters can be fixed. Further investigation leads us to fixing the parameters q_2 and q_8, because they have the biggest influence on the estimate error increase comparing to other tested parameters. If we fix

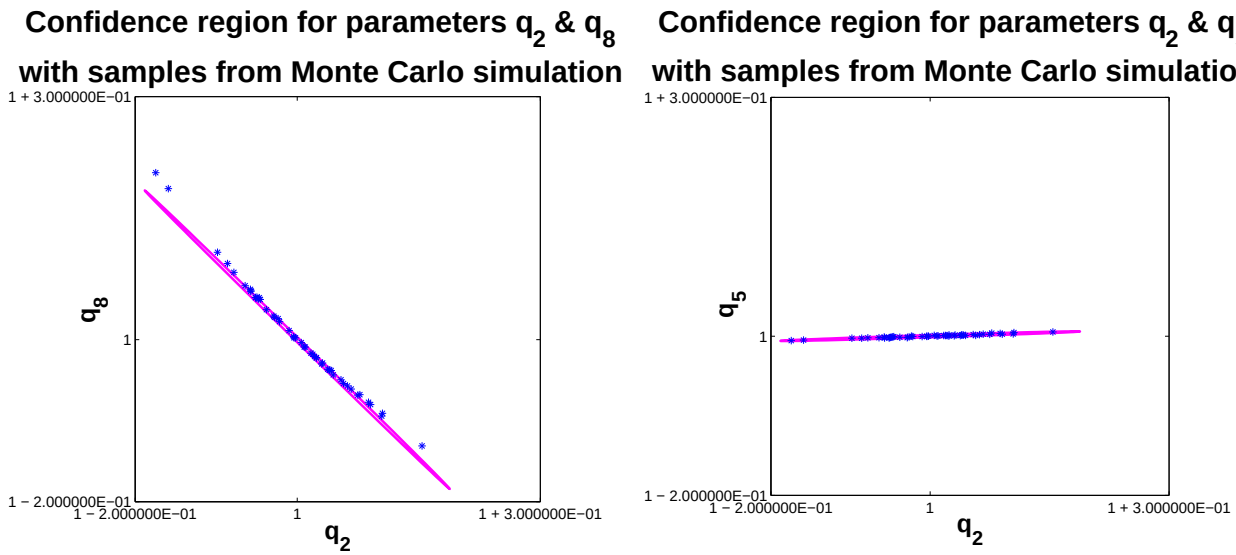

Figure 6.9: 95% confidence regions and Monte Carlo points, parameters q_2, q_5 and q_8, species C_5 and C_8. Detected nonlinearity for parameters q_2 and q_8.

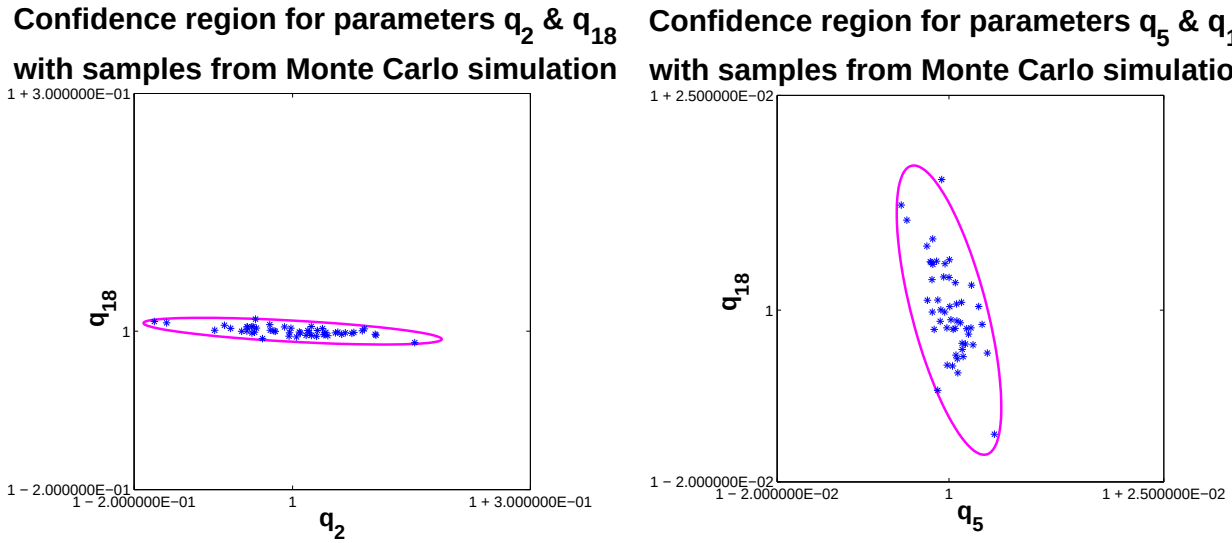

Figure 6.10: 95% confidence regions and Monte Carlo points, parameters q_2, q_5 and q_{18}, species C_5 and C_8.

them, the error in other estimates is smaller. We have to be careful during selection of the parameters, as not all parameters are estimable.

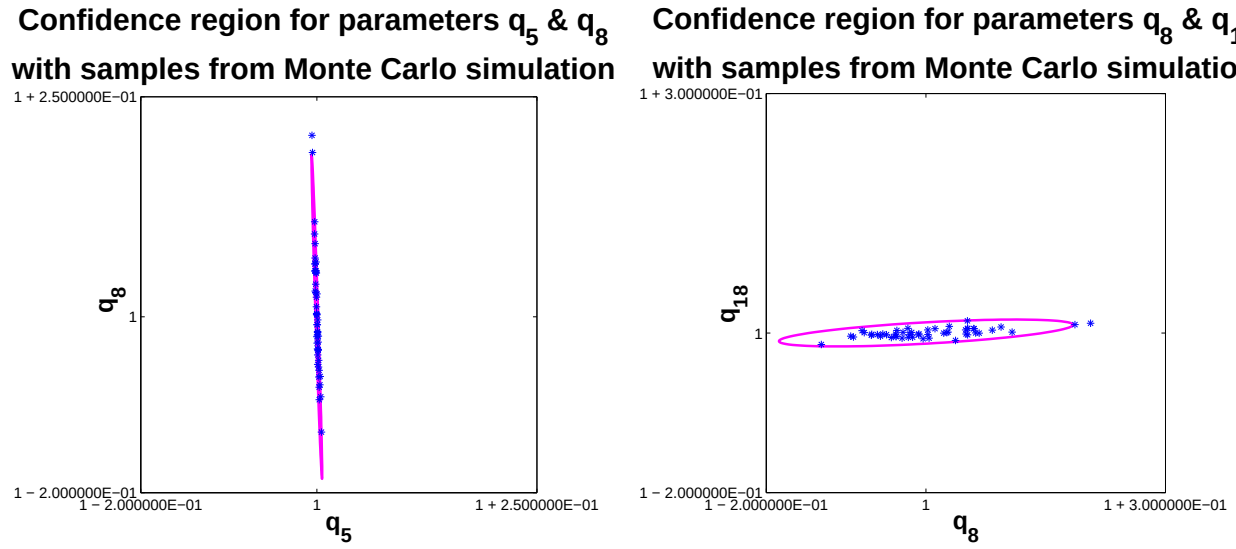

Figure 6.11: 95% confidence regions and Monte Carlo points, parameters q_5, q_8 and q_{18}, species C_5 and C_8.

Up to now, we performed simulations for a confidence region linearized about q(1,1), for each parameter. To further check the behavior of the ellipses with respect to nonlinearities, we introduce randomly small perturbations and linearize the confidence region for each parameter about q in the range {0.9-1.1}. We draw both ellipses (for q(1,1) and q(0.9-1.1,0.9-1.1)) in one plot with centers in point (1,1), and investigate the model nonlinearity regarding chosen parameters and measurements. If the difference is big the method of ellipses cannot be applied to estimate the range of searched parameters. We perform simulations for parameters q_5, $q_9 = d_3$ (spontaneous degradation rate of the *MYC* mRNA), $q_{12} = k_v$ (scaling coefficient, cytoplasm to the nucleus ratio), q_{13} and q_{14} for measurement error of 10%. The results (Figure 6.12 to Figure 6.16) show only little difference between shape in ellipses, therefore we state that we may estimate these parameters using the method of ellipses.

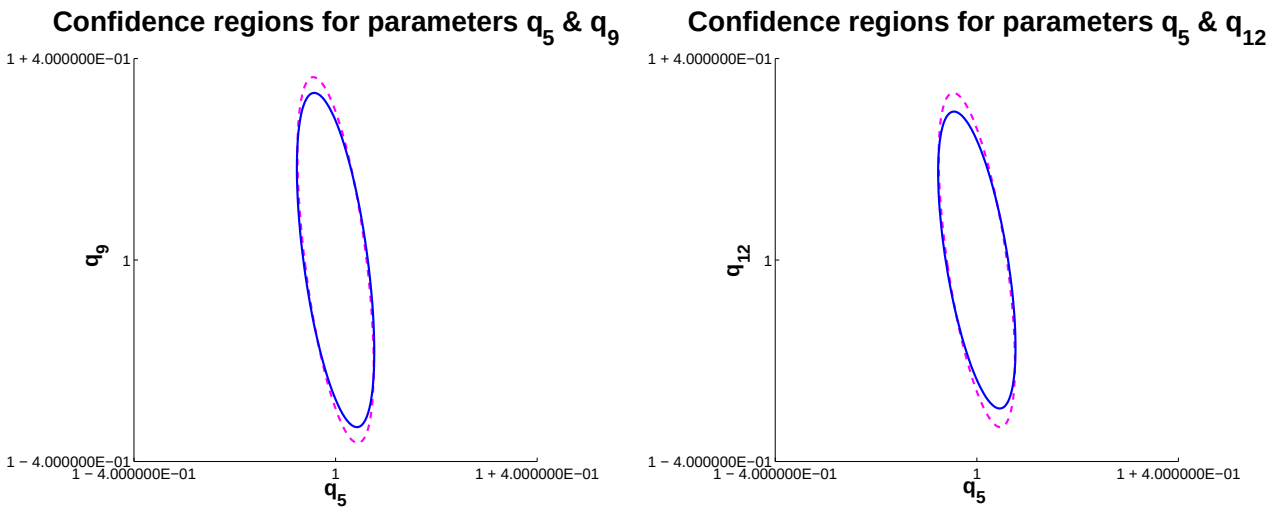

Figure 6.12: 95% confidence regions for different linearizations, parameters q_5, q_9 and q_{12}, species C_5 and C_8; pink-dashed: linearization about q(1,1), blue-solid: linearization about q(0.9-1.1,0.9-1.1).

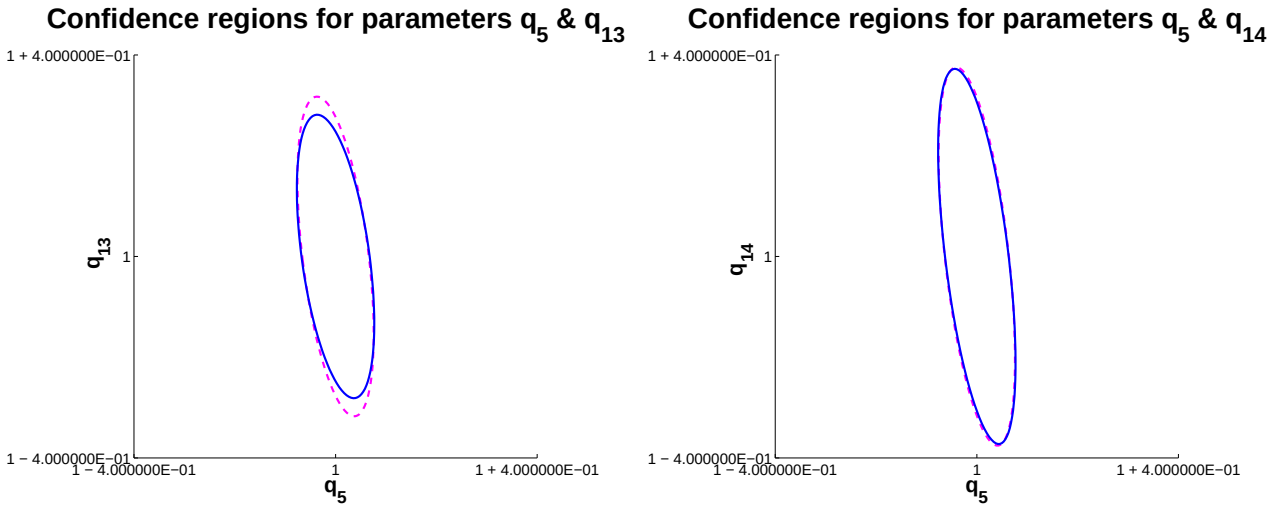

Figure 6.13: 95% confidence regions for different linearizations, parameters q_5, q_{13} and q_{14}, species C_5 and C_8; pink-dashed: linearization about q(1,1), blue-solid: linearization about q(0.9-1.1,0.9-1.1).

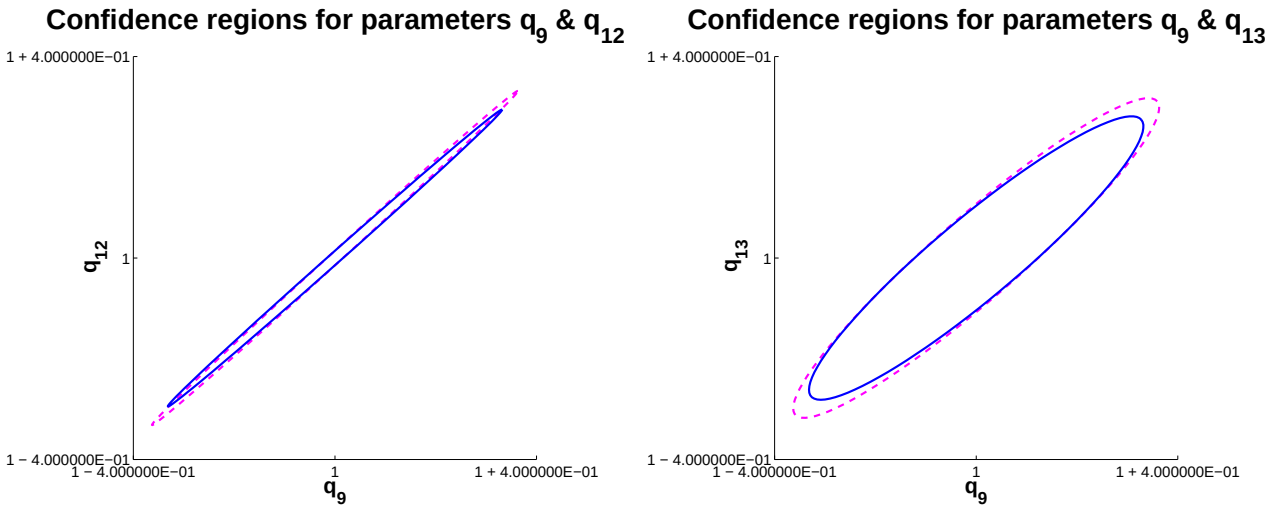

Figure 6.14: 95% confidence regions for different linearizations, parameters q_9, q_{12} and q_{13}, species C_5 and C_8; pink-dashed: linearization about q(1,1), blue-solid: linearization about q(0.9-1.1,0.9-1.1).

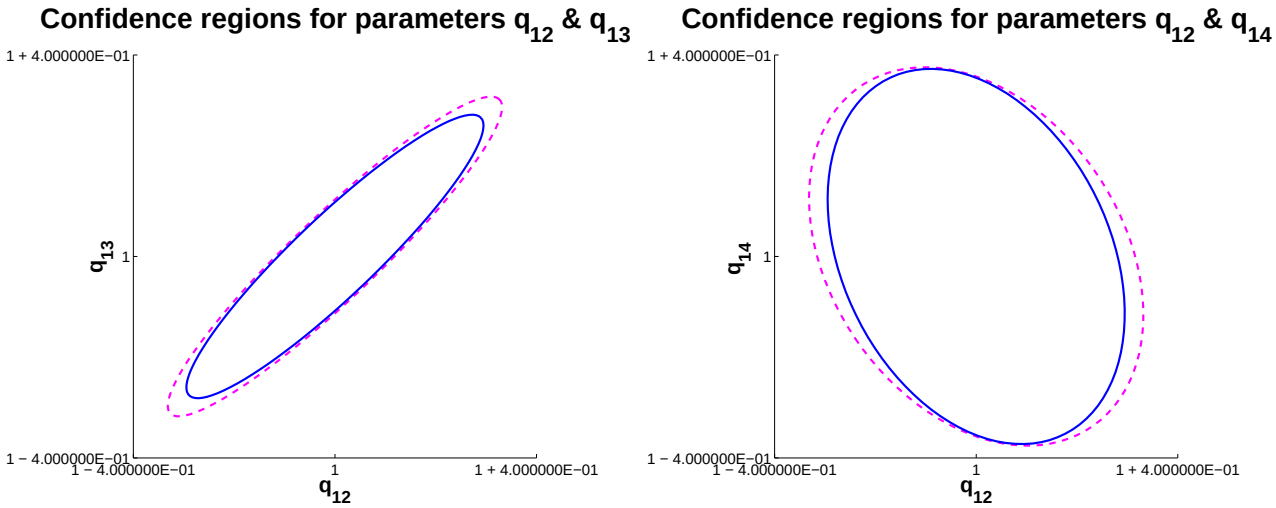

Figure 6.15: 95% confidence regions for different linearizations, parameters q_{12}, q_{13} and q_{14}, species C_5 and C_8; pink-dashed: linearization about q(1,1), blue-solid: linearization about q(0.9-1.1,0.9-1.1).

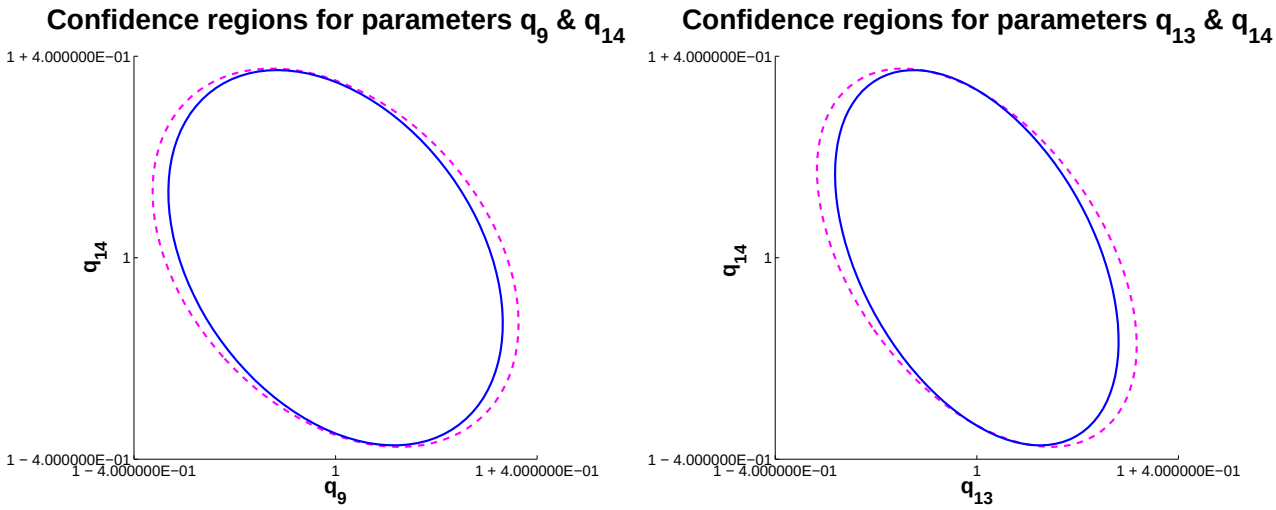

Figure 6.16: 95% confidence regions for different linearizations, parameters q_9, q_{13} and q_{14}, species C_5 and C_8; pink-dashed: linearization about q(1,1), blue-solid: linearization about q(0.9-1.1,0.9-1.1).

Next we check, how the variation in measurement error e_M reflects the parameter estimation. We present a numerical solution of both Monte Carlo 50 sampling points and ellipses for parameters q_5, q_9, q_{12}, q_{13}, q_{14}. We set l = 41 for $e_M = 1\%$ of measurement error (species C_5 and C_8) and also for $e_M = 10\%$ of measurement error (species C_5 and C_8). It is demonstrated that for chosen errors e_M and l = 41 the estimation error is below 100%. We find that even for $e_M = 10\%$ of measurement error we still have a reasonable solution. A measurement error of $e_M = 10\%$ results in parameter estimation error in all cases less than 40% and for $e_M = 1\%$ estimation error is less than 5% (Figure 6.17 to Figure 6.26).

Remark 6.7 *The Monte Carlo simulation for 50 sampling points is compared with the method of ellipses reflecting the 95% confidence regions. It is noticed that the method of ellipses represents a good linearized approximation of the model for different aspects of analysis.*

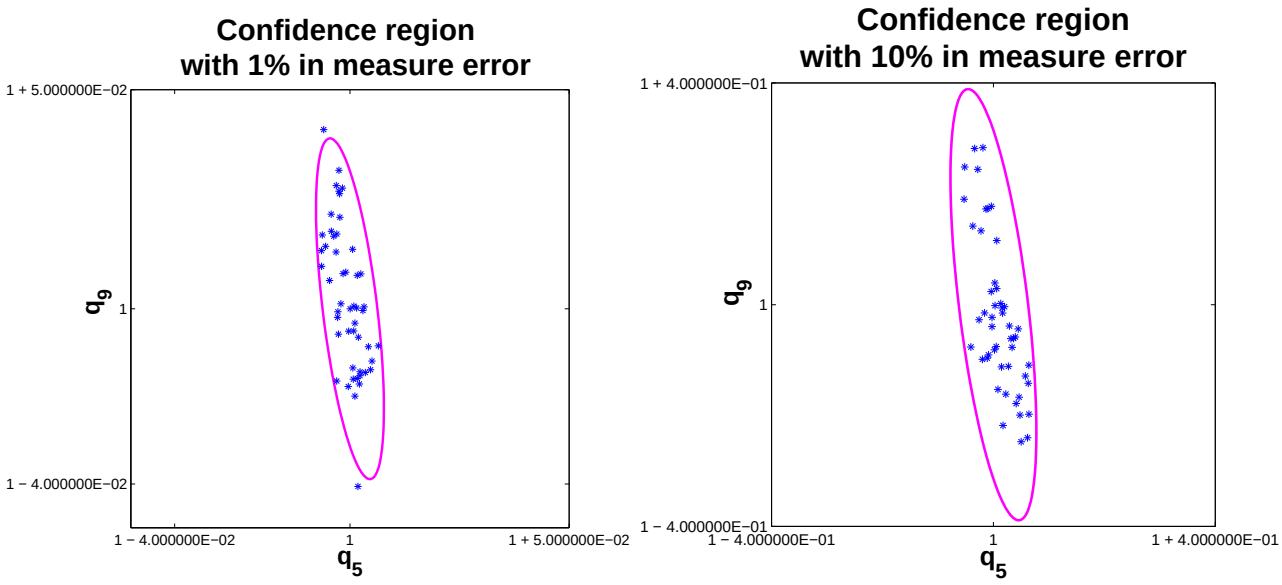

Figure 6.17: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_5 and q_9 for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

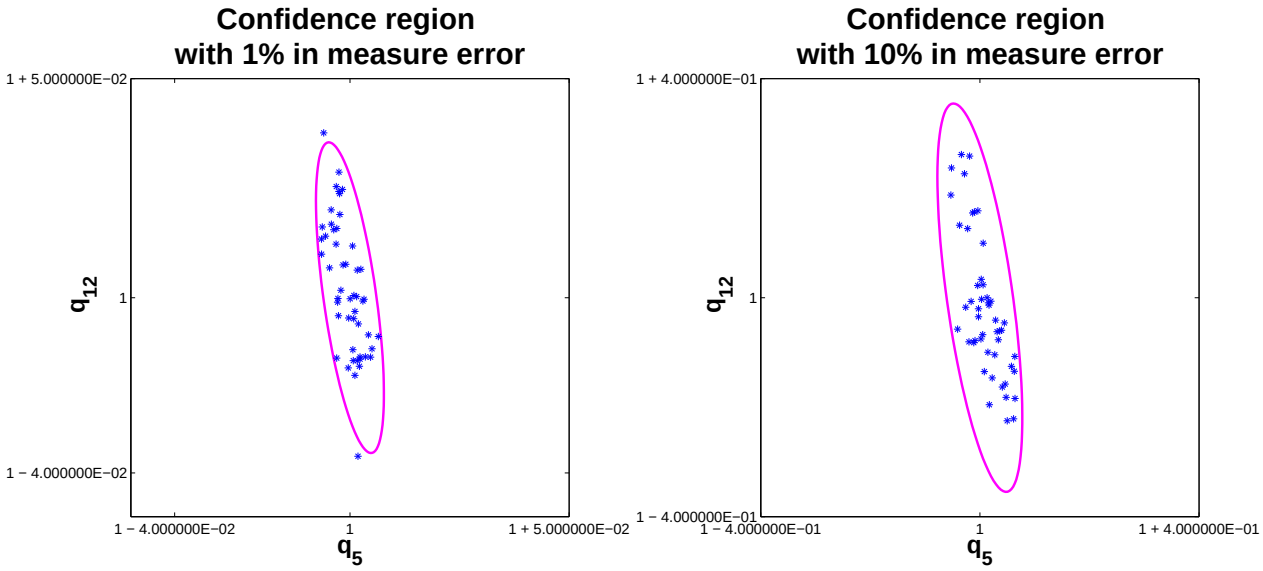

Figure 6.18: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_5 and q_{12} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

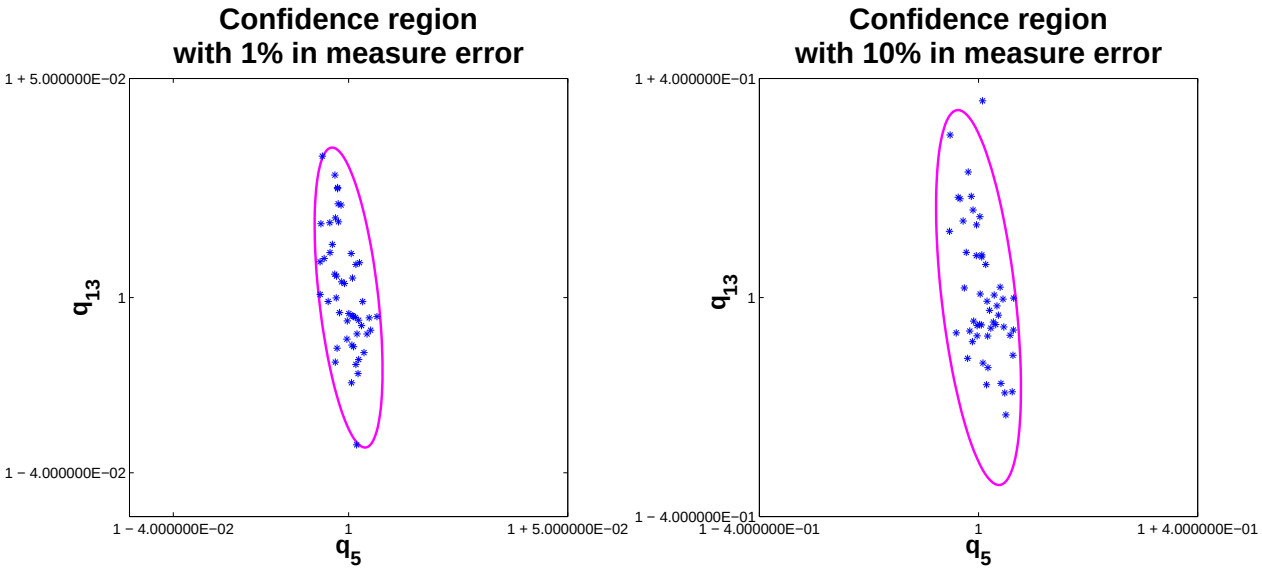

Figure 6.19: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_5 and q_{13} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

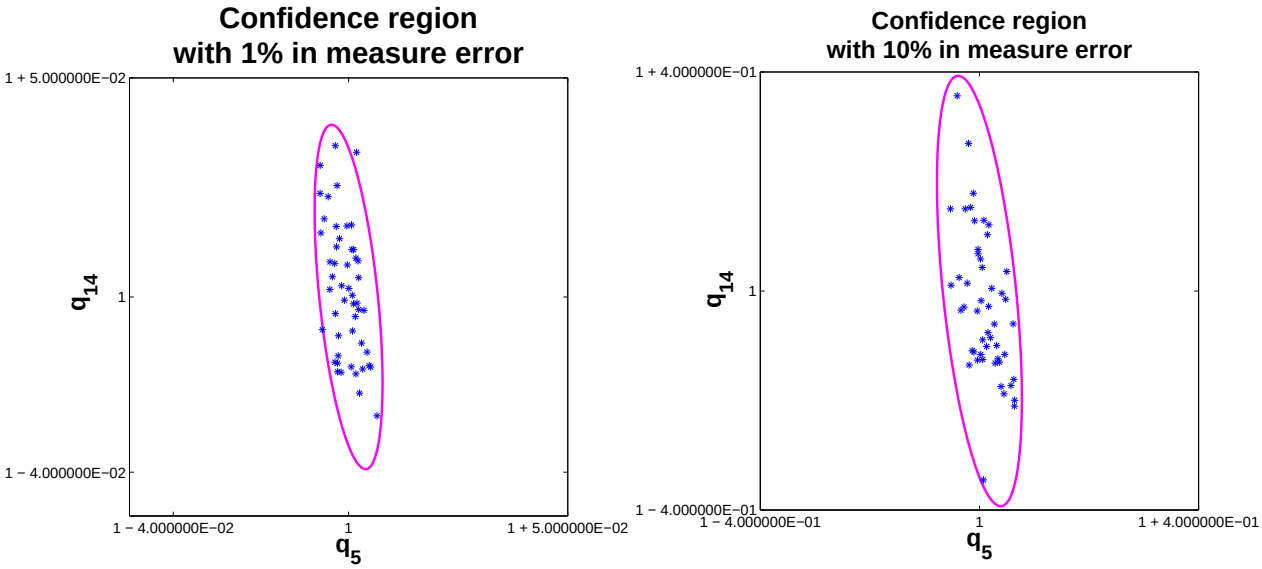

Figure 6.20: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_5 and q_{14} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

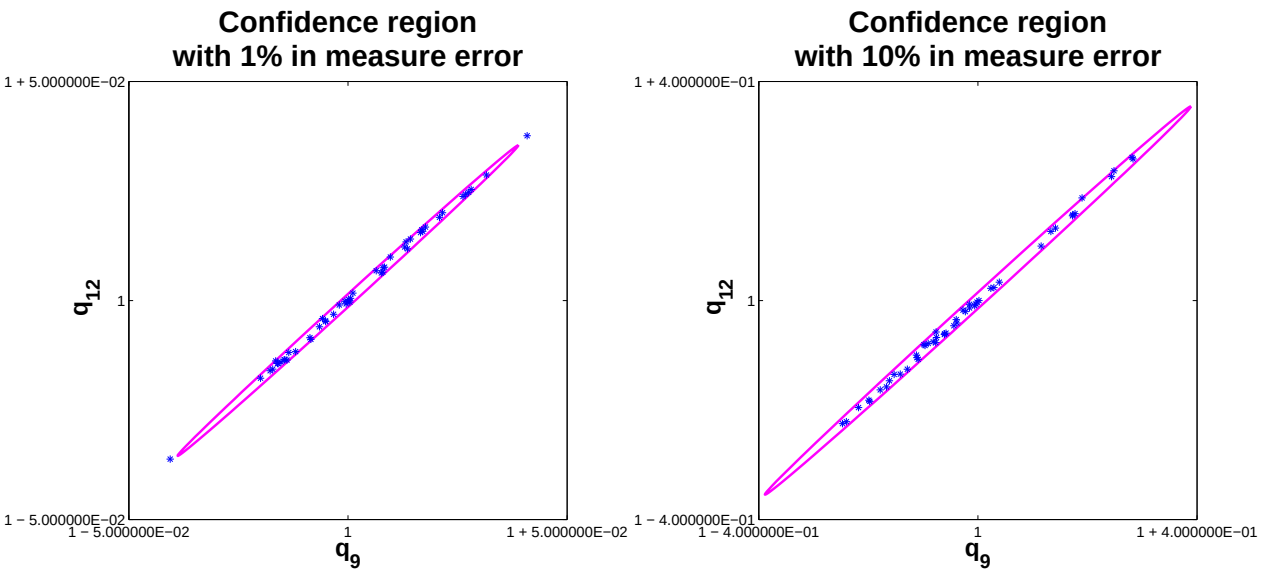

Figure 6.21: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_9 and q_{12} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

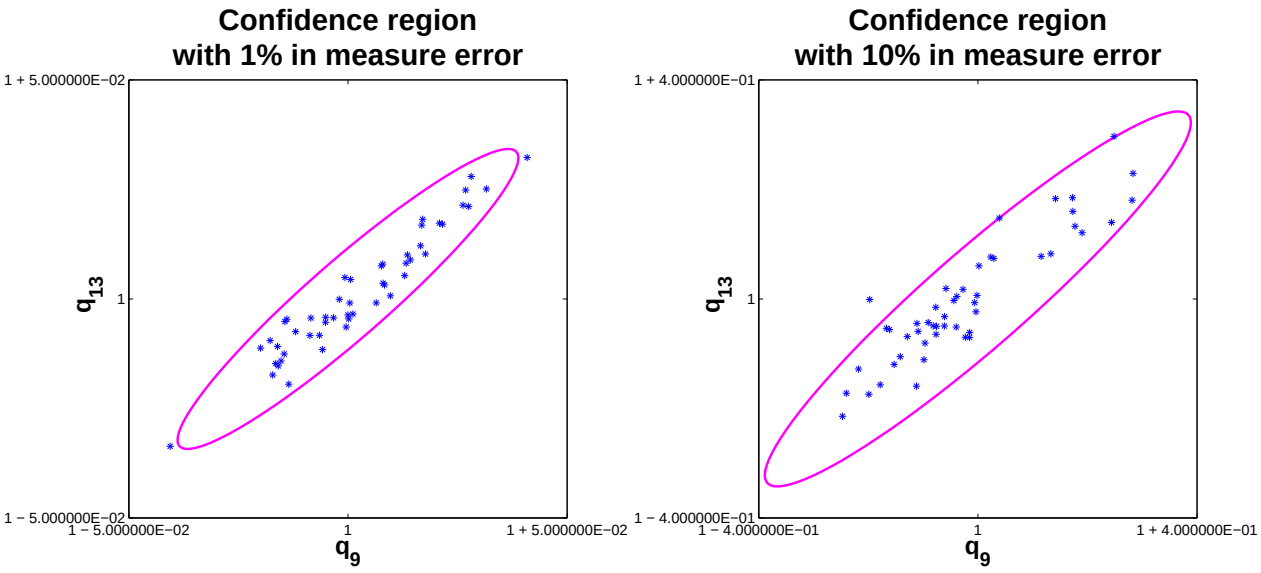

Figure 6.22: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_9 and q_{13} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

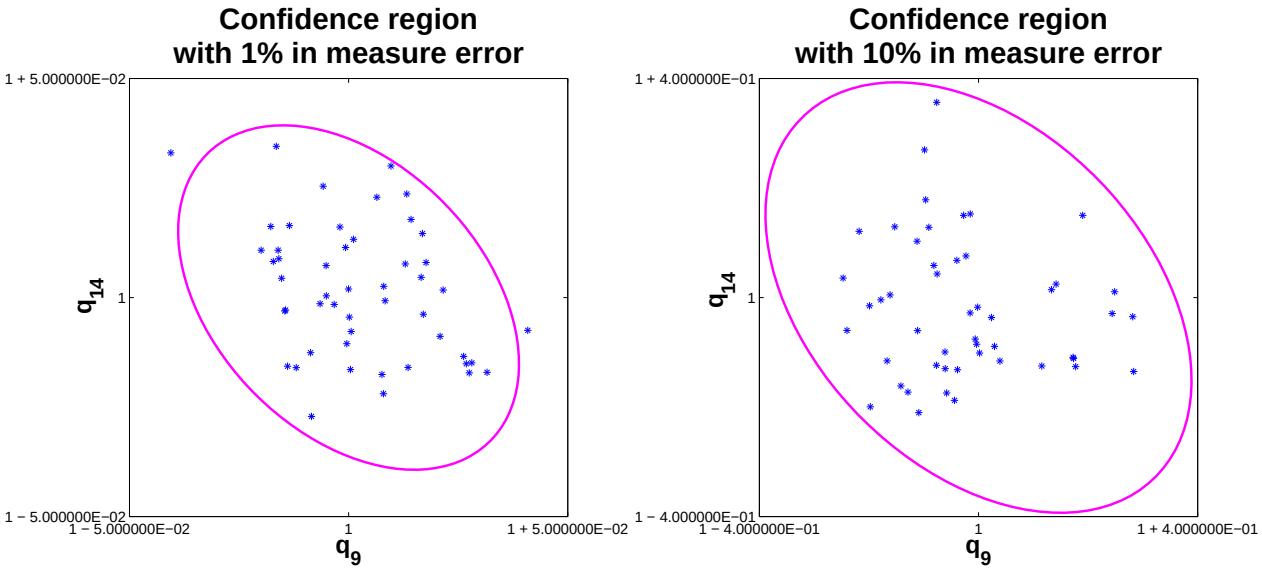

Figure 6.23: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_9 and q_{14} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

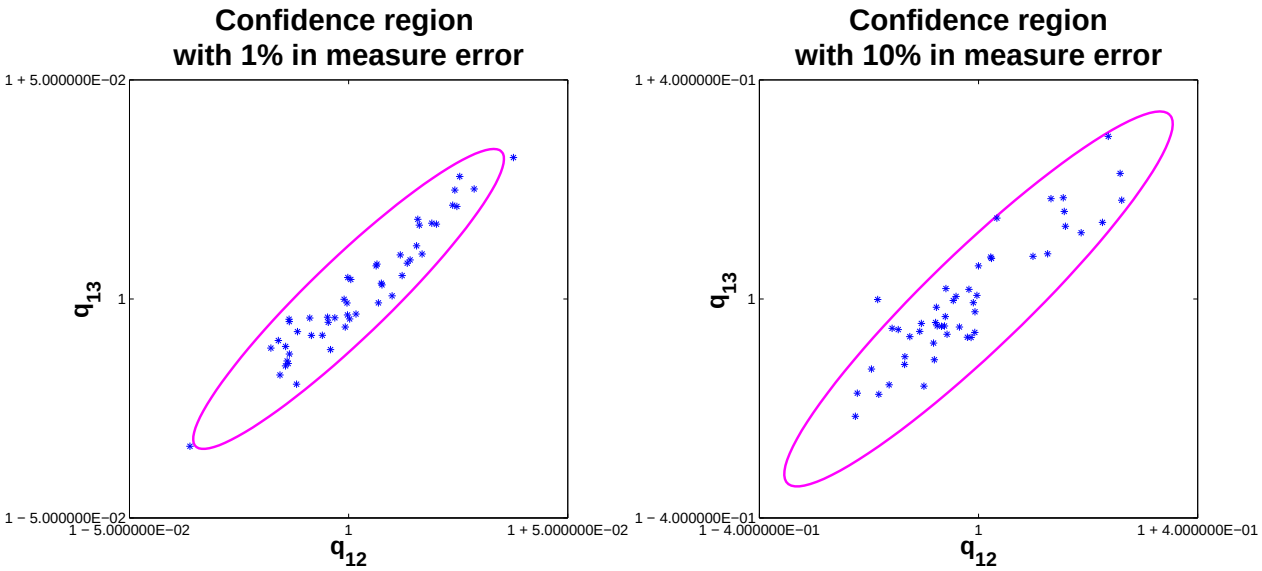

Figure 6.24: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_{12} and q_{13} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

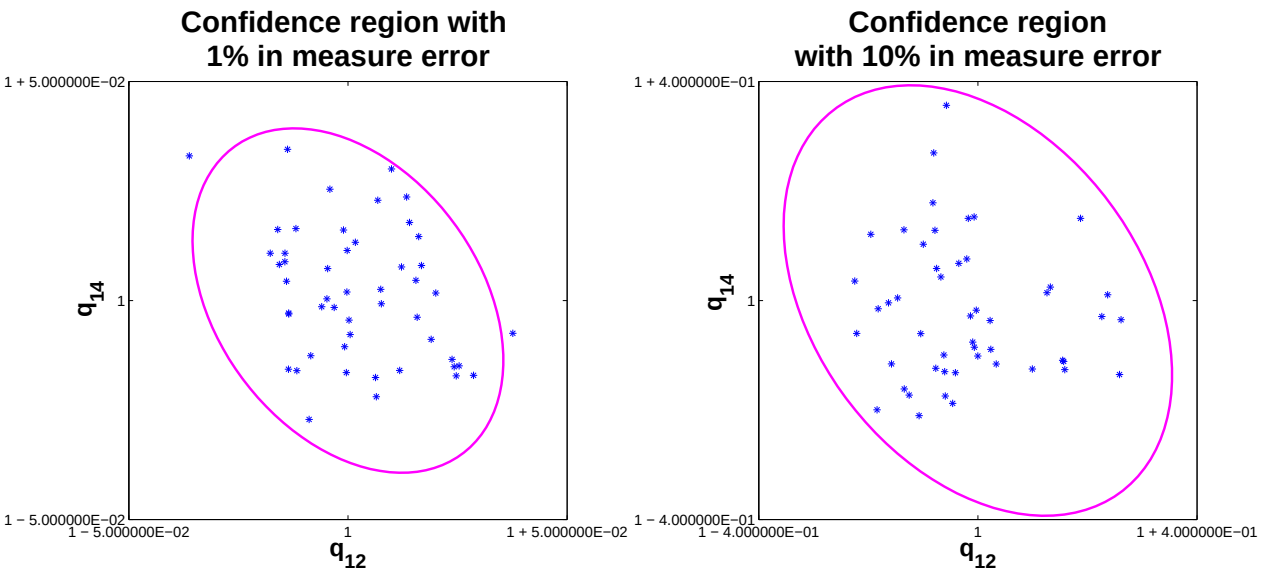

Figure 6.25: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_{12} and q_{14} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

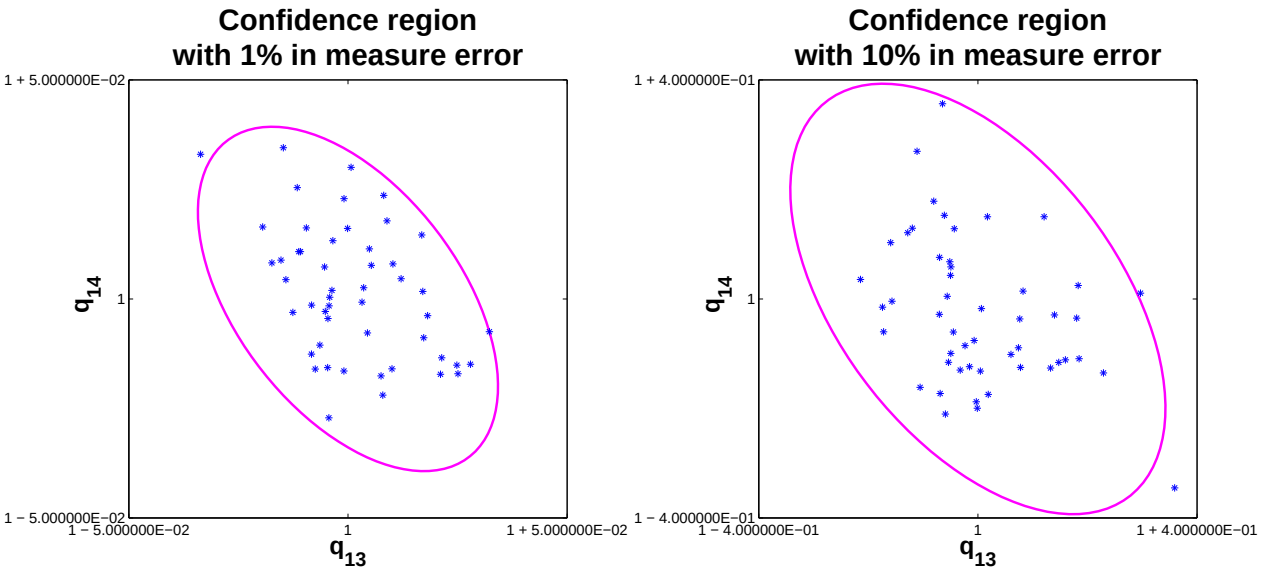

Figure 6.26: 95% confidence region and Monte Carlo numerical results for estimation of parameters q_{13} and q_{14} for C_5 and C_8 with [A] 1% and [B] 10% of measurement error.

Short Summary 6.2 *In this subsection we performed numerical simulations in order to deal with interesting aspects of parameter estimation, i.e., nonlinearities, coupling and measurement error. The successive estimates established after these simulations are added to our estimates table:*

Estimates	Investigated in
q_5, q_{13}, q_{14}	Subsection 6.3.1
q_9, q_{12}, q_{18}	Subsection 6.3.2

6.3.3 Variation in parameter values

In the following we consider the parameters q are unknown and that they are to be estimated. The parameters r are known and described with some range in values. Based on the preliminary simulations, we assume that ellipses approximate the behavior of Monte Carlo simulations and therefore we visualize only ellipses.

Definition 6.7 *We define the known parameters r described with some range as parameters for which laboratory experiments reveal slightly different values of the same parameter.*

Problem 6.7 *The aim of this subsection is to examine how the variation in parameter values combined with the error of measurement e_M influences the error in total parameter identification.*

In biological modeling the knowledge of exact parameter values, e.g., representing the rate constant, is fundamental. Sometimes, we can find in the scientific literature necessary data. Unfortunately, often data is not available or parameters are given in ranges. Evidently, the estimation process is better if we have more information available. Therefore, to use all possible information we can use the parameters with a range in values to make the estimation.

To analyze Problem 6.7, we fix some parameters r adding them the ranges of uncertainty and we test the influence of this variation on the accuracy of estimates. To do this, we plot ellipses based on the covariances derived from the Jacobian in the direction of estimated and fixed parameters.

Remark 6.8 *In this subsection, we redefine the already known covariance* Cov (6.10) *by* $\mathrm{Cov_q}$.

Definition 6.8 *The covariance matrix* $\mathrm{Cov_q}$ *is of the form*

$$\mathrm{Cov_q} := \left(\hat{G}_q^T C_D^{-1} \hat{G}_q\right)^{-1}, \tag{6.14}$$

where the covariance is derived in the direction of estimated parameters.

Definition 6.9 *The covariance matrix* $\mathrm{Cov_r}$ *is of the form*

$$\mathrm{Cov_r} := \left(\hat{G}_q^T C_D^{-1} \hat{G}_q\right)^{-1} \left(\hat{G}_q^T C_D^{-1} \hat{G}_r\right) C_r \left(\hat{G}_q^T C_D^{-1} \hat{G}_r\right)^T \left(\hat{G}_q^T C_D^{-1} \hat{G}_q\right)^{-1}, \tag{6.15}$$

where the covariance is derived in the direction of fixed parameters with ranges and matrix C_r *describes the range of the fixed parameters.*

Definition 6.10 *To plot the confidence ellipses, we apply the covariance of estimated and fixed parameters in the form*

$$\mathrm{Cov} := \mathrm{Cov_q} + \mathrm{Cov_r}\,. \tag{6.16}$$

Numerical simulations

We perform simulations to obtain confidence ellipses given by (6.16) for the following parameters q_5, q_9, q_{12}, q_{13} and q_{14}. We introduce variation (uncertainty) from 1% up to 10% with 1% step in the following two fixed parameters with ranges. The first parameter is $r_1 = \beta cat$ (transcription factor for the *MYC* and *SGK1* mRNA synthesis), and the second parameter is $r_3 = d_1$ (spontaneous degradation rate of the *SGK1* mRNA). The investigation containing variation in fixed parameters is tested (Figure 6.27 to Figure 6.36).

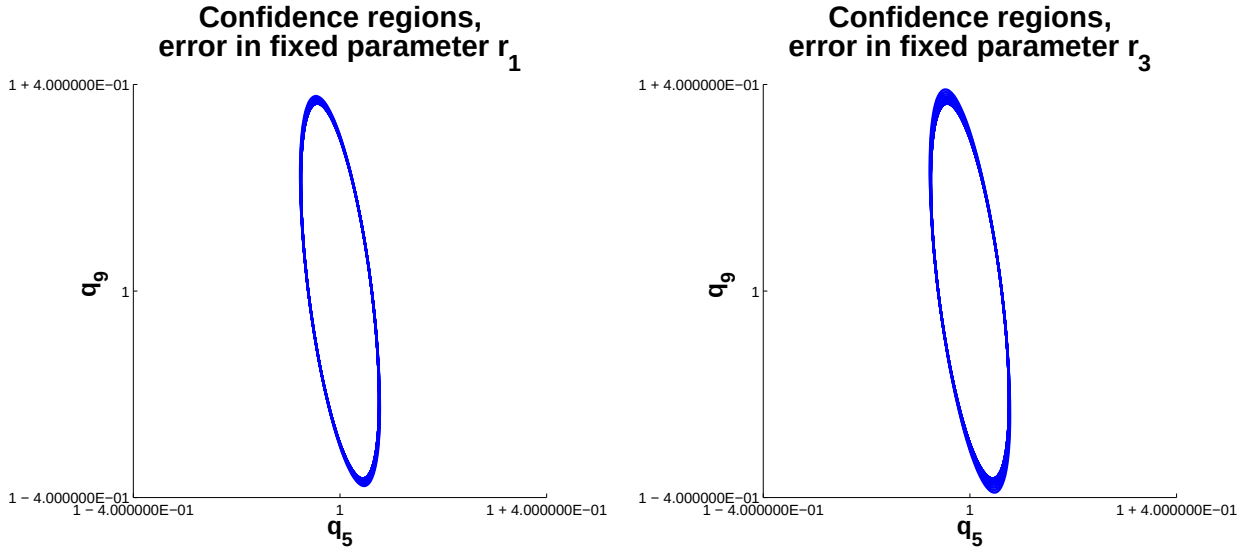

Figure 6.27: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_5, q_9, species C_5 and C_8 with 10% of measurement error.

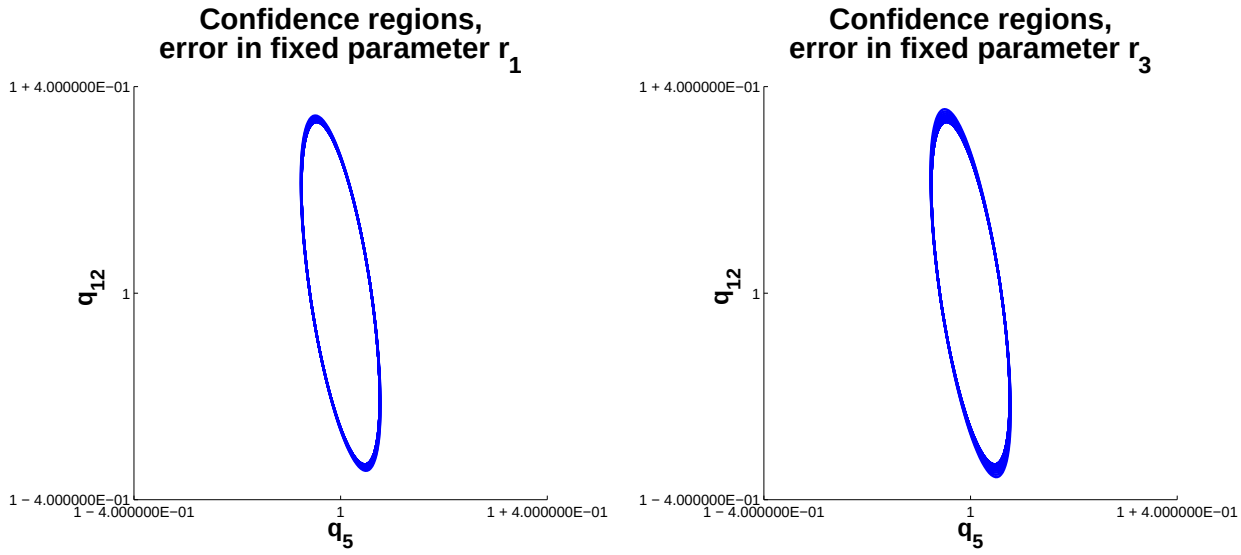

Figure 6.28: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_5, q_{12}, species C_5 and C_8 with 10% of measurement error.

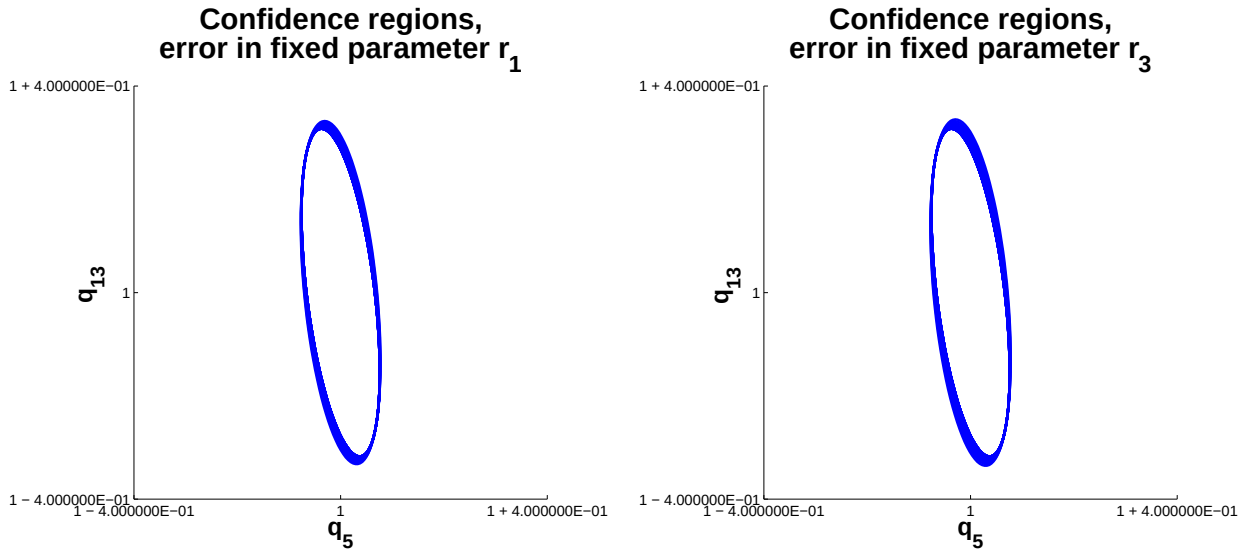

Figure 6.29: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_5, q_{13}, species C_5 and C_8 with 10% of measurement error.

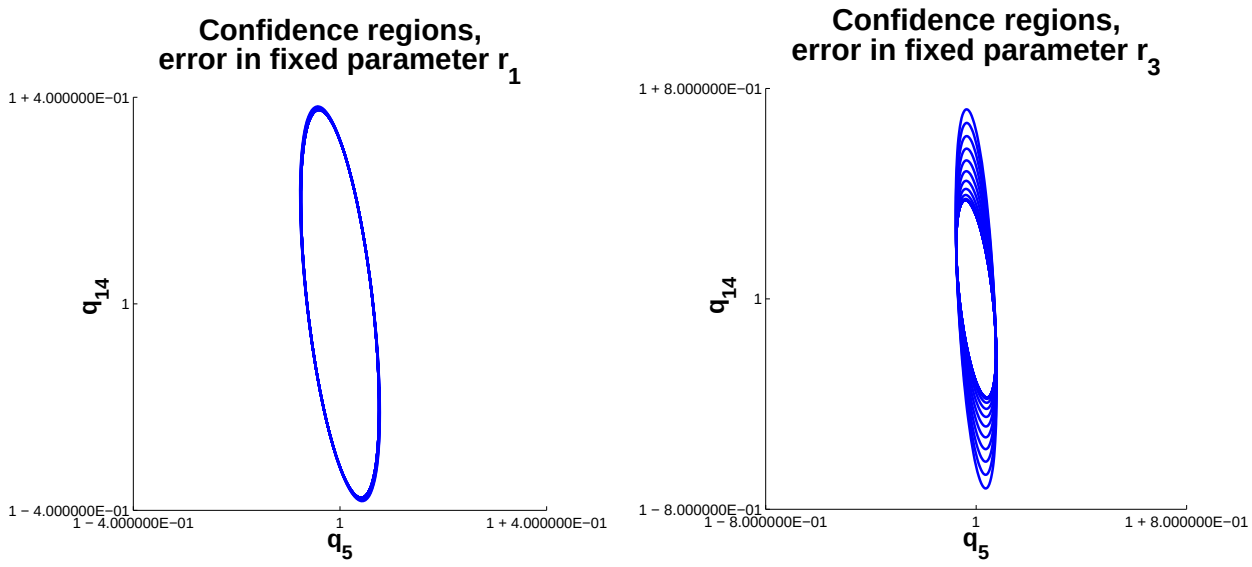

Figure 6.30: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_5, q_{14}, species C_5 and C_8 with 10% of measurement error.

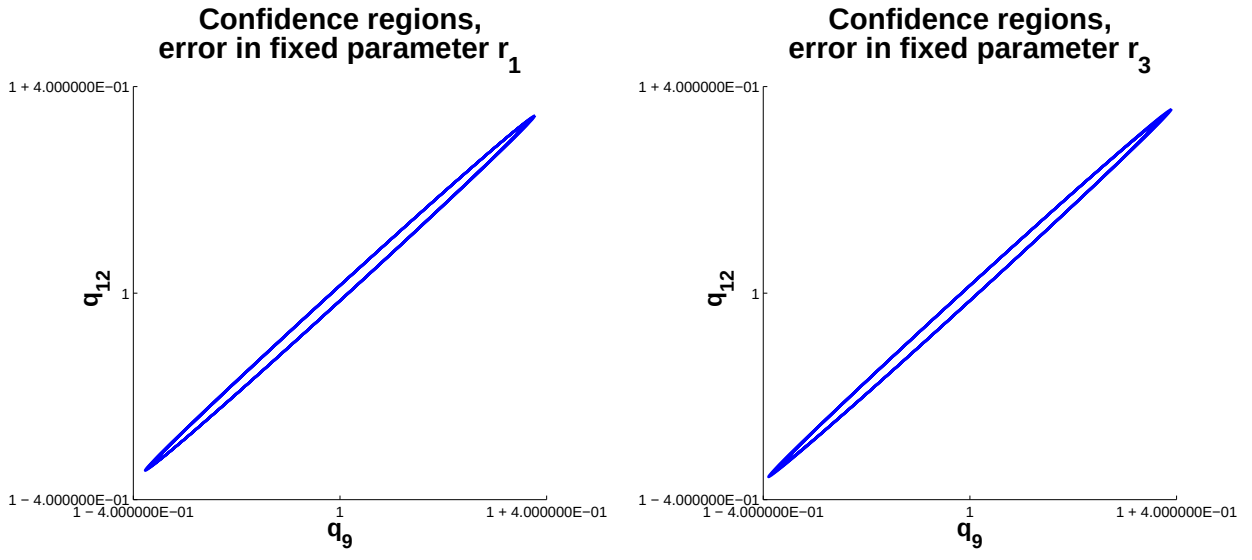

Figure 6.31: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_9, q_{12}, species C_5 and C_8 with 10% of measurement error.

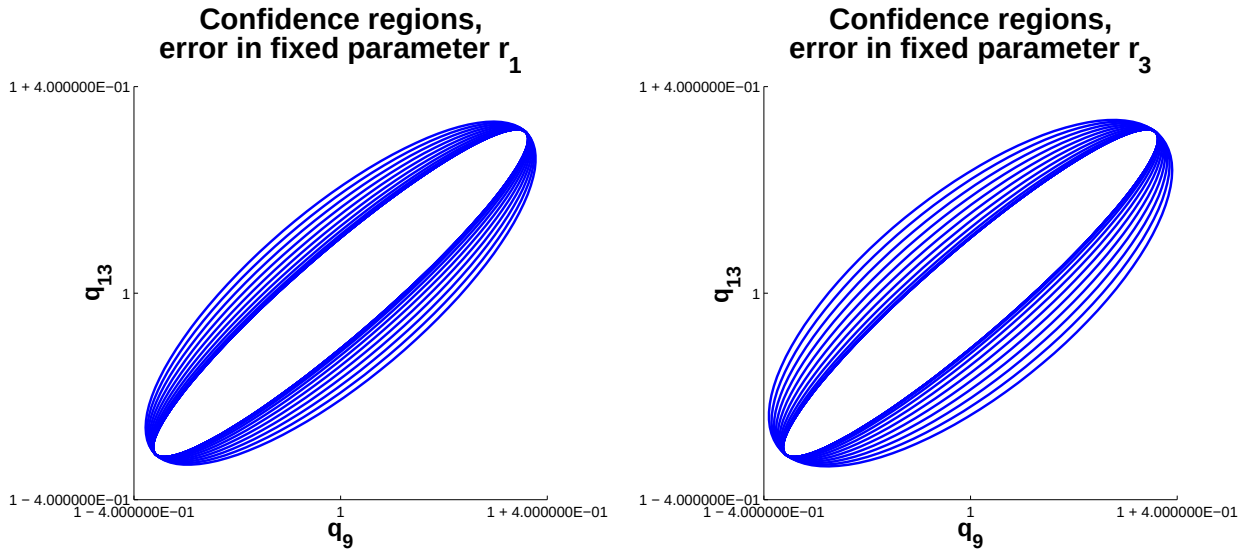

Figure 6.32: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_9, q_{13}, species C_5 and C_8 with 10% of measurement error.

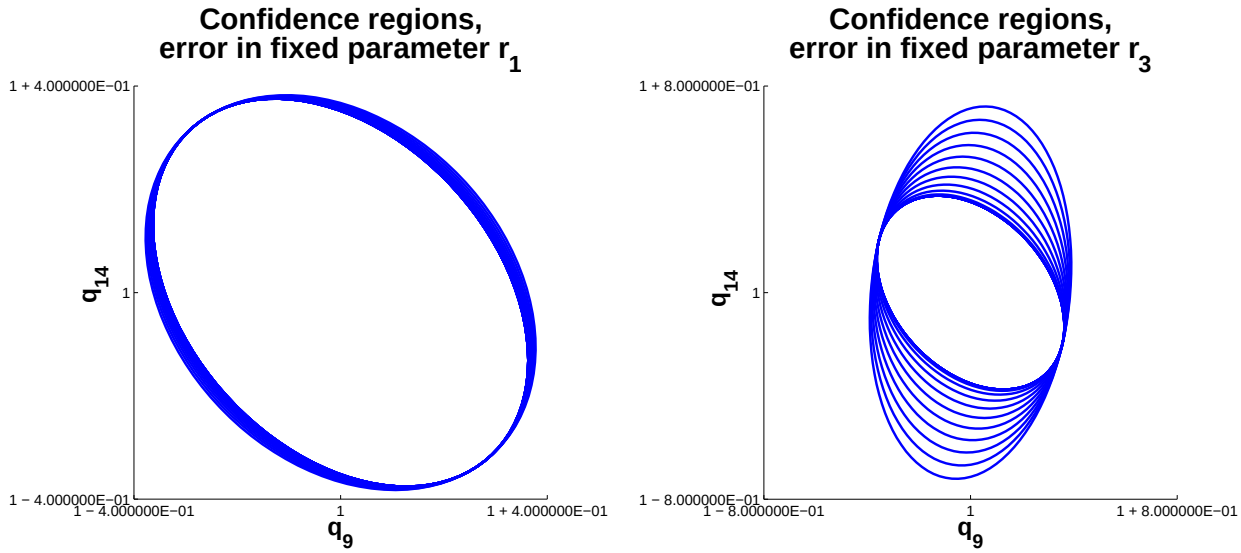

Figure 6.33: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_9, q_{14}, species C_5 and C_8 with 10% of measurement error.

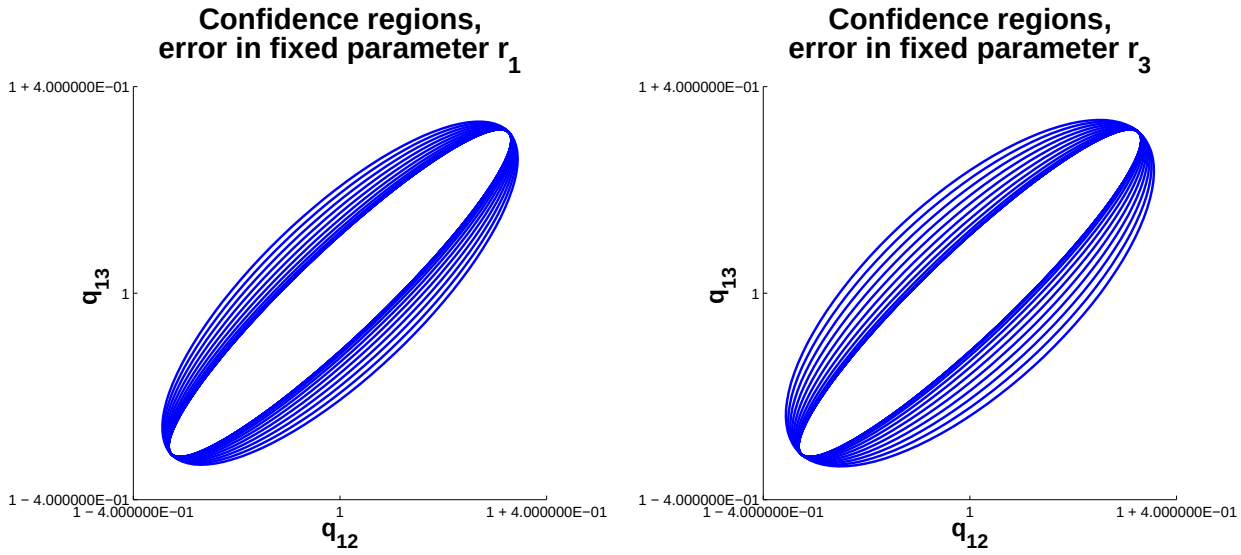

Figure 6.34: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_{12}, q_{13}, species C_5 and C_8 with 10% of measurement error.

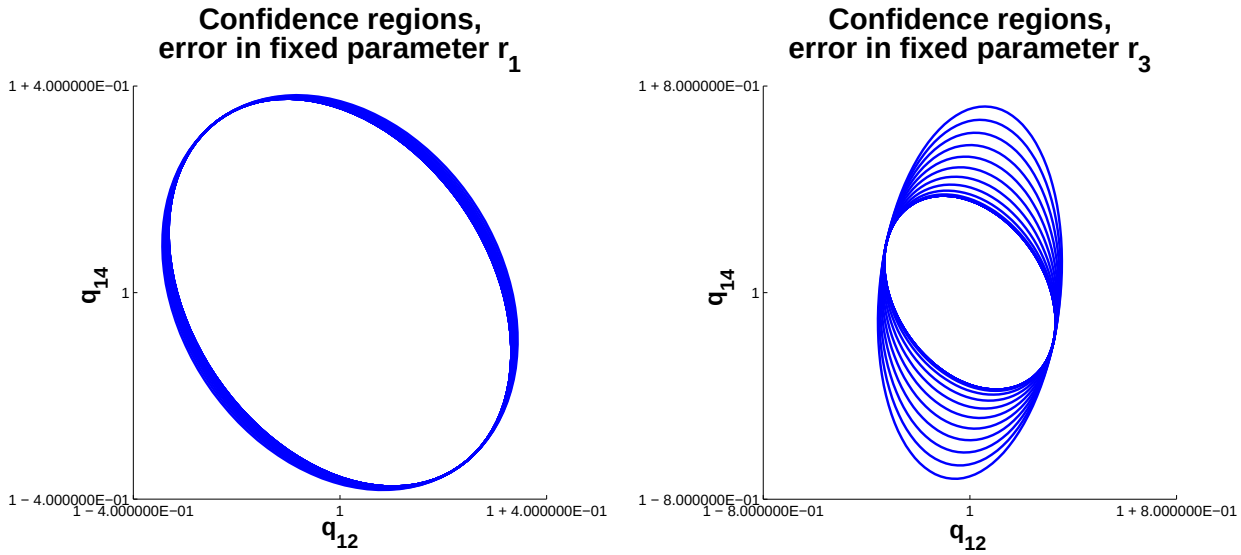

Figure 6.35: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_{12}, q_{14}, species C_5 and C_8 with 10% of measurement error.

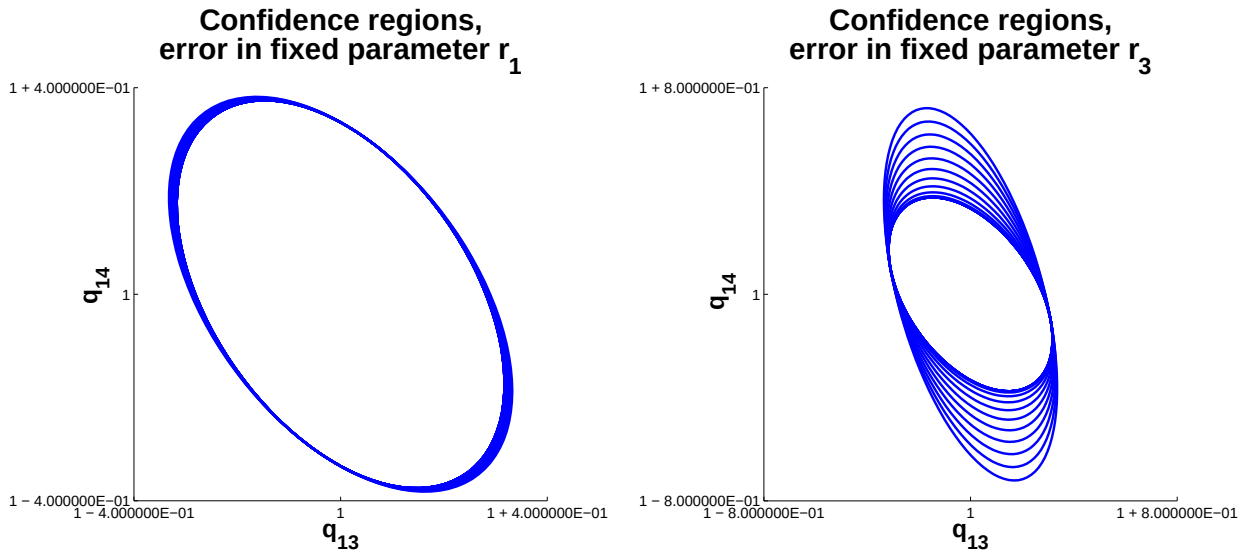

Figure 6.36: 95% confidence regions for variation in fixed parameters from 1% up to 10% for two parameters r_1 and r_3, estimation of parameters q_{13}, q_{14}, species C_5 and C_8 with 10% of measurement error.

In each Figure we depict ellipses starting from 1% of variation in fixed parameters up to 10% of variation with a step of 1%. The smallest ellipse corresponds to the lowest uncertainty in fixed parameters and the biggest ellipse to the largest uncertainty. The results illustrate that the ellipses of the estimated parameters vary more or less with respect to two fixed parameters.

Ellipses of parameters $\{q_5, q_9\}$, $\{q_5, q_{12}\}$, $\{q_5, q_{13}\}$, $\{q_9, q_{12}\}$, $\{q_9, q_{13}\}$, $\{q_{12}, q_{13}\}$ do not differ much for cases r_1 and r_3. Additionally, the variation in fixed parameters does not entail any large increase in estimation uncertainty. Therefore, we say that for particular estimations the variation in fixed parameters does not influence the solution. For other ellipses, i.e., $\{q_5, q_{14}\}$, $\{q_9, q_{14}\}$, $\{q_{12}, q_{14}\}$ and $\{q_{13}, q_{14}\}$ we observe big increase in parameter estimation uncertainty for 10% of variation in fixed parameter r_3. For parameter r_3 the accuracy in parameter estimation is much lower than for parameter r_1. Particularly, q_{14} is vulnerable to the variation in fixed parameter r_3. It follows that quality of estimation in parameter q_{14} is strongly affected by uncertainty in r_3. Altogether, we find that estimation is sensitive to the uncertainty in r_3. In consequence, it might be necessary to fix such parameter completely, if possible. However, the system is not sensitive to the uncertainty in r_1. Thus, when no information (or even any ranges) about such parameter is available, it can be considered as parameter to estimate.

Remark 6.9 *The size and shape of ellipses does not change much upon variation of r_1. Hence, we consider parameter r_1 to be estimable, since the sensitivity is high enough (Table 6.3).*

Therefore, we set $q_1 = r_1$ and we estimate this parameter too. We can observe that some of the fixed parameters with ranges cannot be taken into account in estimation, if we want to achieve precise estimates.

Short Summary 6.3 *In this subsection we introduced the covariance* Cov *composed of the covariance matrices derived in the direction of fixed and estimated parameters. We illustrated changes in the size and shape of the ellipses based on* Cov *with respect to applied measurement error. Our considerations are finalized by finding one more parameter that can be estimated for our model:*

Estimates	Investigated in
q_5, q_{13}, q_{14}	Subsection 6.3.1
q_9, q_{12}, q_{18}	Subsection 6.3.2
q_1	Subsection 6.3.3

The whole investigation ends up with establishing seven parameters, for which estimation error is below 100% (see Remark 6.4).

6.4 Optimal experimental design

The goal of this section is to minimize the variance of estimates according to the design parameter (experimental set-up) $\xi \in \Pi$.

Strictly speaking, we want to minimize the matrix $\mathrm{Cov} = \left(\hat{G}^T C_D^{-1} \hat{G}\right)^{-1}$ for sampling time, which in our model takes over the role of design parameter. Minimizing the variance corresponds to maximizing the certainty of the information.

Definition 6.11 *Optimal experimental design (OED) aims to the most significant identification of unknown parameters in the model.*

Goal 6.1 *Goal of OED is the reduction of the confidence region of the model parameters.*

We improve our estimation by choosing *optimal* sampling protocol (design parameter ξ) in order to reduce Cov. There are the following optimality criteria to find the optimal sampling protocol (see [17]), which are based on functional Ψ of the covariance Cov:

D-optimal design criterion $\Psi(\mathrm{Cov}) = \prod_i \lambda_i$,

E-optimal design criterion $\Psi(\mathrm{Cov}) = max(\lambda_i)$,

T-optimal design criterion $\Psi(\mathrm{Cov}) = \sum_i \lambda_i$,

C-optimal design criterion $\Psi(\mathrm{Cov}) = \frac{max(\lambda_i)}{min(\lambda_i)}$,

where $\lambda_{i,i=1,\dots,n}$ are eigenvalues of matrix Cov.

6.4.1 Sequential design

Remark 6.10 *Sequential experimental design leads to parameter refinement in an iterative process.*

Remark 6.11 *In sequential experimental design minimization of covariance matrix* Cov *is expected for each iteration.*

The equation for Cov contains Jacobian, which depends on the sensitivities of the state with respect to given parameter values. However, we want to estimate the parameter values which in fact are used in Jacobian. Therefore, we have to specify a prior guess q^0 to be used in Jacobian. We improve sequentially our estimates based on the Jacobian (initial guess) and the specific optimality criteria used.

The algorithm presented in Table 6.4 is as follows. We have initial parameter estimate q^0 for $\xi^0 = (t_1, t_2, ..., t_j)$ and measurements $\bar{C} = (\bar{C}_1, ..., \bar{C}_j)$. We minimize the functional of Cov over the new design and update ξ^{i+1} by new time point t_{j+1}.We perform the experiment to extract new sample measurement and calculate parameter estimates. We repeat the procedure untill we obtain satisfying parameter convergence given by ϵ_2 in sequential update of the parameter design.

	Sequential OED algorithm
1	**Initialize:**
2	initial design $\xi^o = (t_1, t_2, ..., t_j)$, measurements $\bar{C} = (\bar{C}_1, ..., \bar{C}_j)$, initial guess of parameter q^0
3	**Loop:**
4	$t_{j+1} = \arg\min_{\xi^i}(\Psi(\mathrm{Cov}(\xi^i)))$
5	$\xi^{i+1} = (\xi^i, t_{j+1})$
6	perform experiment to get measurement $\bar{C}_{j+1}$
7	calculate parameter estimate q^{i+1} based on Gauss-Newton algorithm
8	evaluate confidence region based on the ellipses
9	increment global iteration counters j and i
10	normal exit if ($\|q^{i+1} - q^i\| < \epsilon_2$)
11	**End loop**

Table 6.4: Pseudocode of sequential OED algorithm.

It happens that the initial guess q^0 of the parameters is far away from the solution q, what may result in a large confidence region. In such case, we would need a new strategy and we would have to prepare a new setup of initial guess of parameters and then follow the procedure of estimation process. For highly nonlinear models we can work with many

sets of initial parameter guesses in order to find the proper one and even then fail. For further details see [59].

Problem 6.8 *The goal is to derive parameter estimates with the highest possible accuracy based on the optimality criteria in the sequential design.*

We consider all the criteria described in the beginning of Section 6.4. Tests performed in previous sections demonstrate that parameter estimation is not possible for all model parameters, hence we have a subset Q_{ad} of admissible parameters. We apply the optimality criteria to that subset $Q_{ad} = \{q_1, q_5, q_9, q_{12}, q_{13}, q_{14}, q_{18}\}$ and find the minimum of the functional Ψ for a given criteria. For example, for the **T**-optimality design we search for the minimal value of the trace of covariance matrix Cov in each iteration, etc.

Heuristic design

To compare numerical solutions built on different design criteria in order to find optimal parameter, we start the first iteration with the following heuristic design preliminary chosen from sensitivity analysis:

$$C_5 : \xi^0 = \{t_1 = 18000, t_2 = 19500, t_3 = 20500, t_4 = 21000, \\ t_5 = 22000, t_6 = 23000, t_7 = 24000\};$$

$$C_8 : \xi^0 = \{t_1 = 20000, t_2 = 21000, t_3 = 22000, t_4 = 23500, \\ t_5 = 24200, t_6 = 24500, t_7 = 25000\}.$$

The measurement error is assumed to be $e_M = 1\%$. Depending on the sampled probes the error can be larger. In this case, however, we cannot afford to start with a high measurement error, as already for $e_M = 1\%$ the estimates exceed 60% of error (Table 6.5). Only parameters q_5 and q_{18} indicate high accuracy of estimation.

Parameters	q_1	q_5	q_9	q_{12}	q_{13}	q_{14}	q_{18}
Estimates error	58%	2.1%	67%	58%	15%	39%	2.8%

Table 6.5: Parameter estimation error for heuristic design; measurement error e_M = 1%.

Problem 6.9 *To refine the estimates using the idea of sequential design, we have to find the best sampling protocol.*

We apply the sequential design, and therefore, in the following iterations we minimize the covariance matrix Cov derived in the direction of estimated parameters. Numerical solutions show improvement in the estimates in the subsequent iterations evaluated by succesive optimization procedure for $Trace$ (T-optimality), Det (D-optimality), Eig (E-optimality) and $Ratio$ (C-optimality). Table 6.6 displays the sampling points found in the subsequent steps of optimization for the considered optimality criteria based on the covariance matrix Cov starting with heuristic design $\xi^0 = \{t_1, ..., t_7\}$.

	iteration 1 time t_8	iteration 2 time t_9	iteration 3 time t_{10}	iteration 4 time t_{11}	iteration 5 time t_{12}
Trace	2800	28400	13400	4200	32800
Det	6000	50000	13400	50000	5300
Eig	2100	27900	13400	3200	12300
Ratio	2100	27900	13400	3200	12300

Table 6.6: Sampling points in subsequent iterations for criteria: $Trace$, Det, Eig, $Ratio$; measurement error e_M = 1%.

We investigate the progressive refinement of the estimates for five iterations. We notice that Eig and $Ratio$ criteria consist of the same sampling design in every step and they differ from $Trace$ and Det except third iteration. $Trace$ and Det criteria also represent various sampling schemes. Such variety of possible sampling schemes leads to the situation where the experimentalist can provide measurements for his own preferred sampling scheme. For this reason the procedure of data collection can be less expensive, i.e., the experiment can be stopped earlier.

The Eig and $Ratio$ criteria meet the shortest time duration with $t_{last} = t_9$ = 27900. In this case we can benefit from Eig and $Ratio$ optimization and stop experiment on t_{last}. An interesting fact emerges in the case of Det, where during optimization process the same probe time is computed twice. This accounts for the repeated measurement for the same time point. The question is which protocol is the best for parameter identification considering costs and efficiency of the experimental procedure.

The enhancement of the parameter estimates is depicted in Table 6.7, 6.8 and 6.9.

Trace criterion	heuristic points	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5
q_1	58%	20%	11%	5.6%	5.4%	5%
q_5	2.1%	2%	1.9%	0.8%	0.8%	0.8%
q_9	67%	2.7%	2.6%	2.5%	1.5%	1.5%
q_{12}	58%	2.1%	2.1%	2.1%	1.2%	1.2%
q_{13}	15%	14%	7%	4.4%	4.3%	3.9%
q_{14}	39%	13%	9%	5%	4.3%	3.8%
q_{18}	2.8%	0.5%	0.5%	0.4%	0.4%	0.4%

Table 6.7: Parameter estimation error in subsequent optimization iterations for Trace criterion; measurement error e_M = 1%.

Det criterion	heuristic points	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5
q_1	58%	21%	12%	5.8%	5.4%	5.3%
q_5	2.1%	2%	1.9%	0.8%	0.8%	0.8%
q_9	67%	3.3%	3.1%	1.9%	1.8%	1.4%
q_{12}	58%	2.2%	2.1%	1.5%	1.5%	1.1%
q_{13}	15%	14%	6.5%	4.2%	3.9%	3.9%
q_{14}	39%	13%	9%	4.3%	4.1%	3.7%
q_{18}	2.8%	0.51%	0.5%	0.44%	0.4%	0.4%

Table 6.8: Parameter estimation error in subsequent optimization iterations for Det criterion; measurement error e_M = 1%.

Eig, Ratio criteria	heuristic points	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5
q_1	58%	20%	11%	6.1%	5.6%	5.2%
q_5	2.1%	2%	1.9%	0.9%	0.8%	0.7%
q_9	67%	3.7%	3.6%	3.6%	1.9%	1.9%
q_{12}	58%	3.1%	3%	3%	1.6%	1.6%
q_{13}	15%	14%	6.9%	4.5%	4.4%	4.3%
q_{14}	39%	13%	9.2%	6.1%	4.6%	4.4%
q_{18}	2.8%	0.53%	0.5%	0.4%	0.4%	0.4%

Table 6.9: Parameter estimation error in subsequent optimization iterations for Eig, Ratio criteria; measurement error e_M = 1%.

Already, after the first iteration the error of the estimates is strongly reduced for all criteria. The increase in accuracy of parameter estimates accompanies within subsequent steps. Numerical solutions indicate that the difference between third and fourth step is small. This implies that only three optimization cycles are essential to obtain reasonable estimates.

However, in further investigations we consider all five steps, as the result for five iterations is the best. The well observable fact is that for different parameters the optimization has different impact. For instance, examine parameter q_1 and q_{12}, where the heuristic design results in an estimation error of 58%. For parameter q_1 the reduction is only almost three times, whereas for parameter q_{12} the reduction is substantial of thirty times after first iteration for $Trace$ criteria. In the case of other parameters, we notice also the diversity in parameter estimates refinement. To clearly depict the process of identification according to the iterations and given criterion, we present results in Figure 6.37 to Figure 6.40. Each parameter is separately considered.

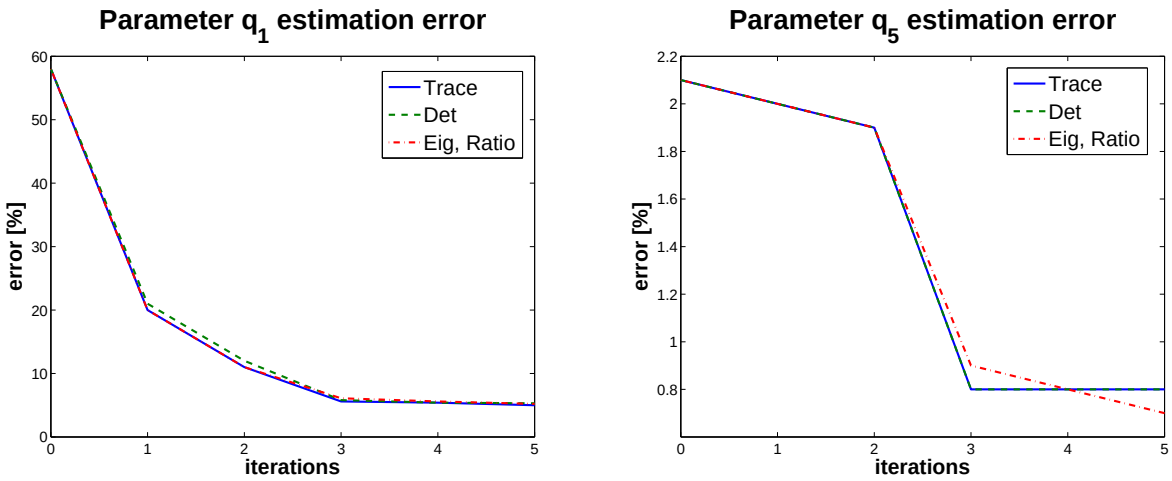

Figure 6.37: Comparison of optimality criteria of parameters q_1 and q_5 estimation error, five optimization iterations; measurement error e_M = 1%.

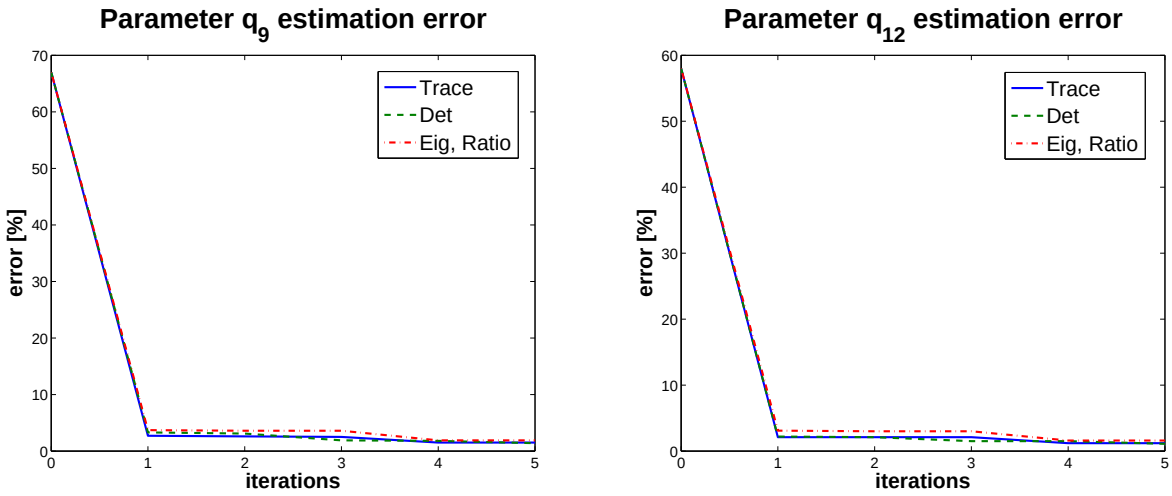

Figure 6.38: Comparison of optimality criteria of parameters q_9 and q_{12} estimation error, five optimization iterations; measurement error $e_M = 1\%$.

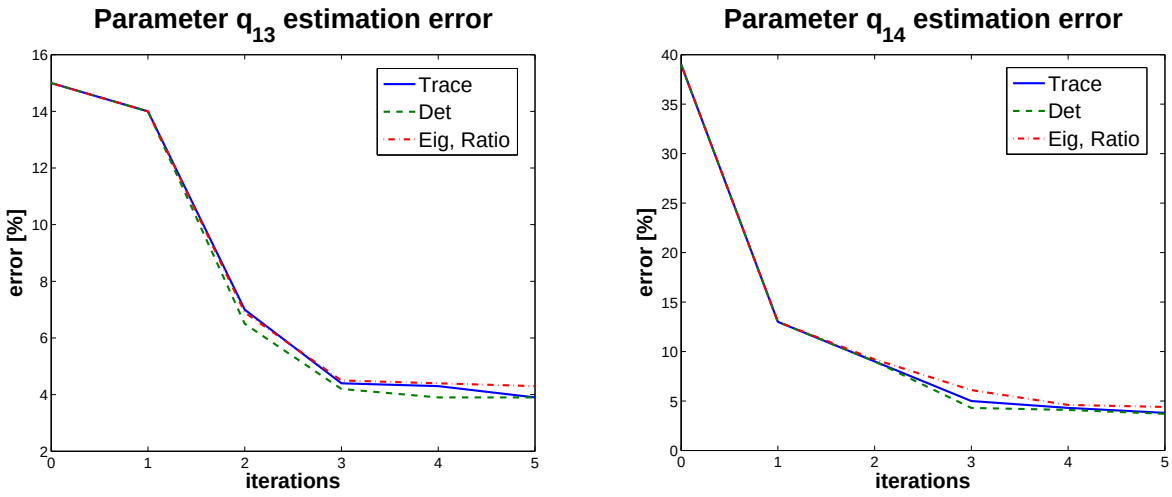

Figure 6.39: Comparison of optimality criteria of parameters q_{13} and q_{14} estimation error, five optimization iterations; measurement error $e_M = 1\%$.

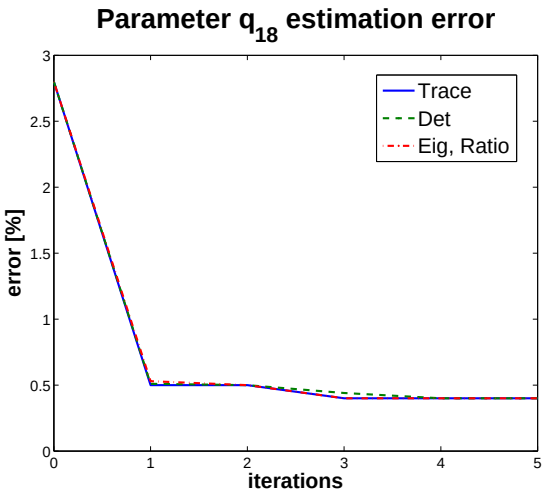

Figure 6.40: Comparison of optimality criteria of parameter q_{18} estimation error, five optimization iterations; measurement error e_M = 1%.

Comparison of optimality criteria for 5 steps of sequential design

Problem 6.10 *To obtain the best parameter estimates in the context of optimal experimental design using sequential design the best optimality criteria must be established.*

By analyzing the criteria in the context of error for parameter estimates, we can state that Eig and $Ratio$ are worse criteria for our model, (see Table 6.7, 6.8 and 6.9). In the case of Det and $Trace$ some estimations are better and some are worse. After first iteration the Det criterion seems to be worse, but in the fifth iteration it performs comparably to $Trace$. Essentially, using all criteria estimation finishes with reasonable results after five iterations.

To recall, for the heuristic design we make a sweep of the additional measurement time point from t_1 to $t_{l,l=50000}$ and we find the minimum to obtain the best time point in each iteration. In Figure 6.41 to Figure 6.44 we illustrate the behavior of particular property of covariance matrix Cov represented by the criteria $Trace$, Det, Eig and $Ratio$ in the time interval (t_1, t_{50000}).

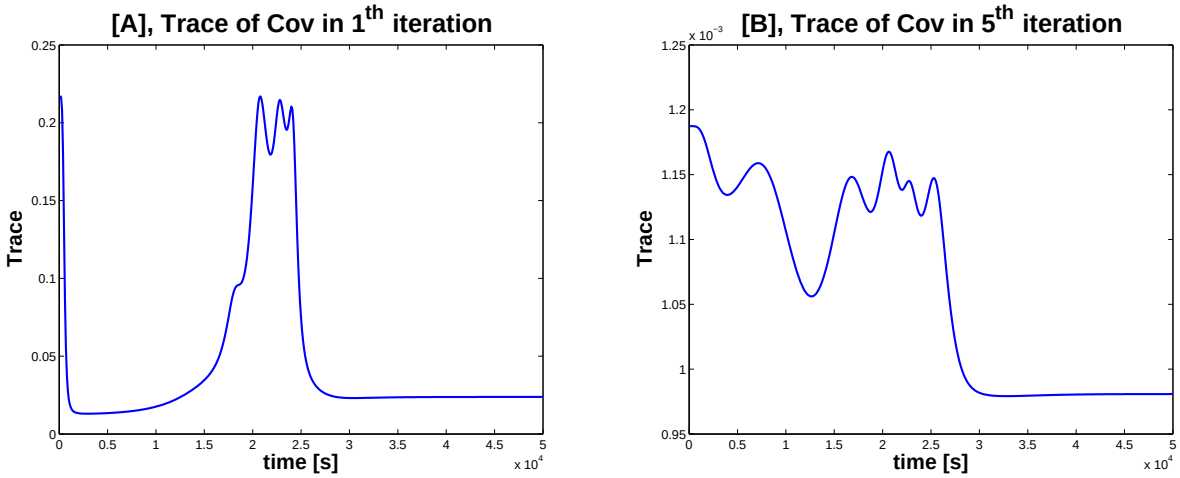

Figure 6.41: Trace of Cov in the process of sampling scheme optimization, first [A] and last [B] iteration.

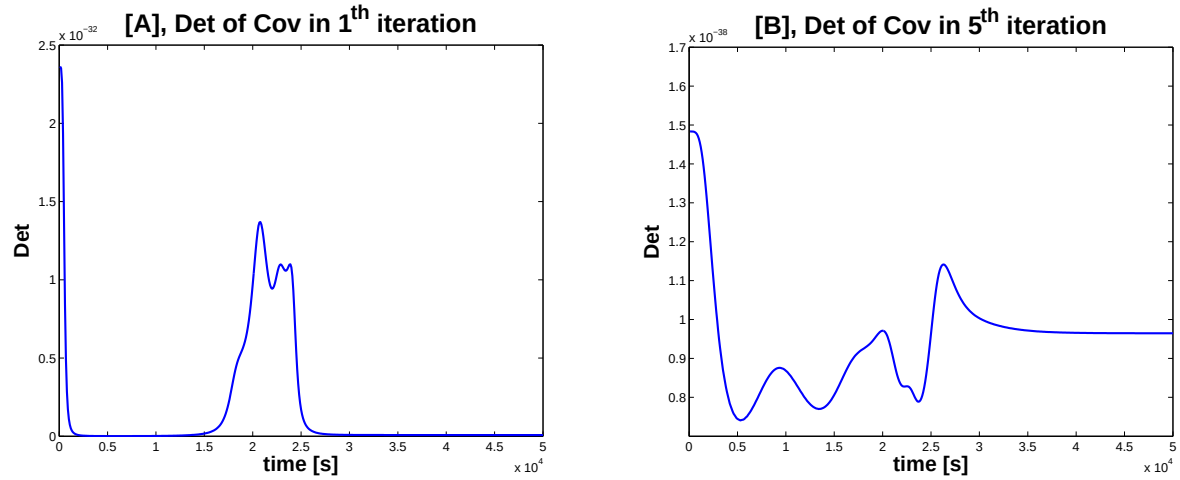

Figure 6.42: Determinant of Cov in the process of sampling scheme optimization, first [A] and last [B] iteration.

For each criterion we depict the first and last iteration. Values of particular property ($Trace$, Det, Eig, $Ratio$) are much higher in the first than the last iteration, for all criteria. It implies that reduction in values of $Trace$, Det, Eig and $Ratio$ entails reduction in error of parameter estimates for subsequent iterations.

We achieve a satisfying error not exceeding 5% for every parameter $q_{i,i=\{1,5,9,12,13,14,18\}}$ for measurement error $e_M = 1\%$. In the case of concentrations, the measurement error of the experimental probes can be (and in laboratorie's reality is found to be) bigger than $e_M = 1\%$. Hence, we carry forward our investigation. We introduce measurement errors $e_M = 5\%$ and $e_M = 10\%$ and give the results are in Table 6.10, 6.11 and 6.12 for all

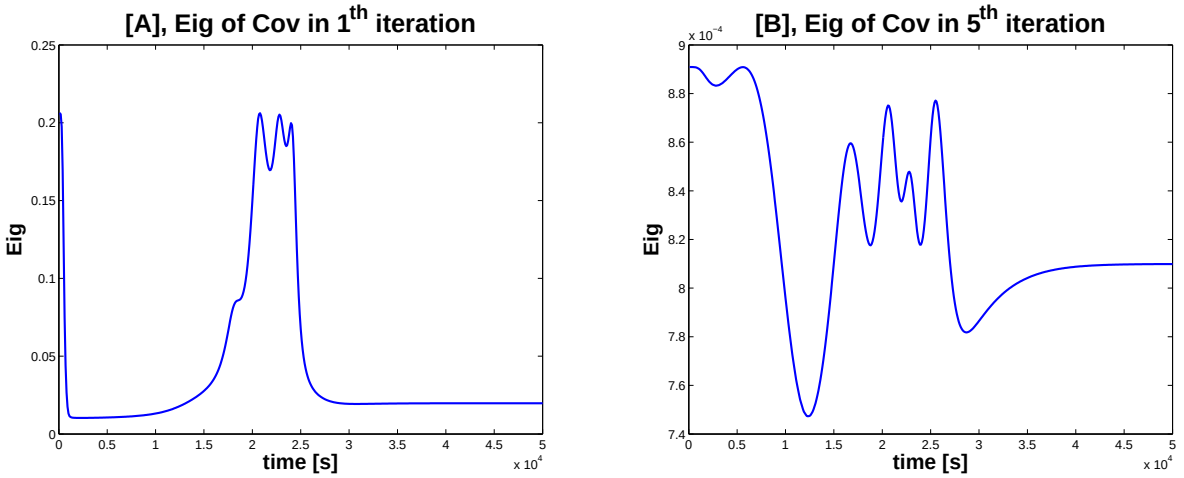

Figure 6.43: Maximal eigenvalue of Cov in the process of sampling scheme optimization, first [A] and last [B] iteration.

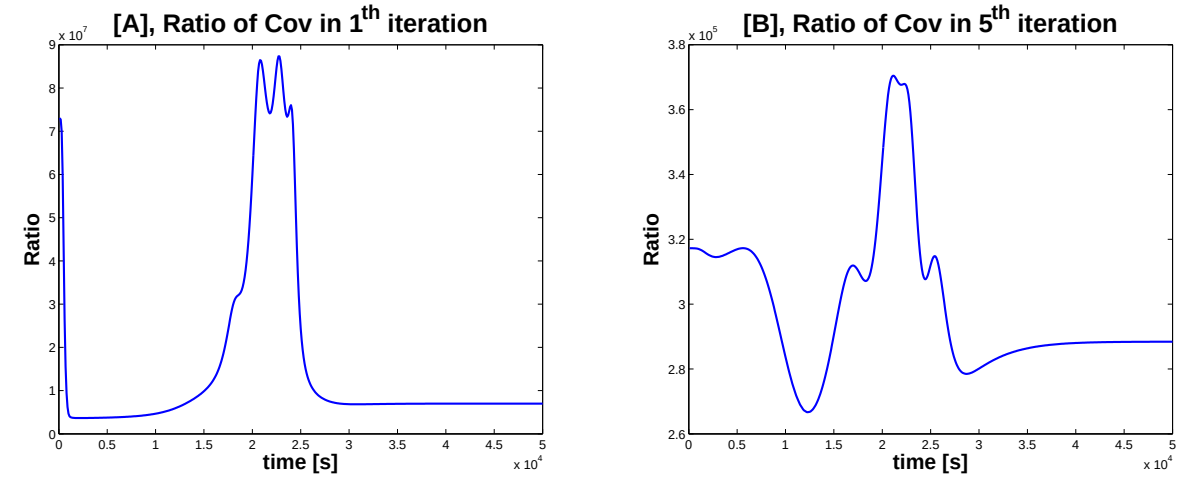

Figure 6.44: Ratio between the maximum eigenvalue and minimum eigenvalue of Cov in the process of sampling scheme optimization, first [A] and last [B] iteration.

optimality criteria.

Trace criterion	1%	5%	10%
q_1	5%	25%	50%
q_5	0.8%	3.8%	8%
q_9	1.5%	7.5%	15%
q_{12}	1.2%	6.1%	12%
q_{13}	3.9%	19%	39%
q_{14}	3.8%	19%	38%
q_{18}	0.4%	1.9%	3.8%

Table 6.10: Parameter estimation error for measurement error 1%, 5%, 10%. Trace criterion.

Det criterion	1%	5%	10%
q_1	5.3%	27%	53%
q_5	0.8%	3.9%	7.7%
q_9	1.4%	7%	14%
q_{12}	1.1%	5.4%	11%
q_{13}	3.9%	20%	39%
q_{14}	3.7%	19%	37%
q_{18}	0.4%	1.9%	3.8%

Table 6.11: Parameter estimation error for measurement error 1%, 5%, 10%. Det criterion.

Eig, Ratio criteria	1%	5%	10%
q_1	5.2%	27%	53%
q_5	0.7%	3.6%	7.3%
q_9	1.9%	9.3%	19%
q_{12}	1.6%	7.7%	15%
q_{13}	4.3%	19%	43%
q_{14}	4.4%	22%	44%
q_{18}	0.4%	1.9%	3.9%

Table 6.12: Parameter estimation error for measurement error 1%, 5%, 10%. Eig, Ratio criteria.

As for $e_M = 10\%$ the parameters are still estimable, the resultant error for some estimates is quite large and is even of 53% for Det, Eig and $Ratio$ criteria. However, four out of seven parameters, q_5, q_9, q_{12} and q_{18}, are obtained with a surprisingly good accuracy, below 20%. For $e_M = 5\%$ we have very good estimates for those parameters after five iterations for all optimality criteria, where the best is $Trace$ and Det.

Short Summary 6.4 *This subsection described strategies in the parameter estimation process regarding the optimal experimental design. We introduced the optimality criteria that support better estimation results. We faced the situation, where measurement error impacts strongly on the parameters accuracy. Finally, we showed that with $e_M = 5\%$ the estimates were achieved with good precision and that the best results were obtained with $Trace$ and Det optimality criteria. The drawback of deterministic estimation is that in our case the subset of admissible estimates Q_{ad} is small (see Remark 6.4).*

Summary of Chapter 6

The parameter estimation problem appears to be intricate under many aspects. The represented nonlinear Gauss-Newton method and linear approximation of confidence regions highlighted difficulties and details of parameter estimation. With the optimal experimental design, we refined estimates and showed which of the optimality criteria is the best and leads to the reduction in estimation error. In our model, however, we could find only seven from nineteen parameters, which are well estimable (Table 6.10, Table 6.11 and Table 6.12): $q_1, q_5, q_9, q_{12}, q_{13}, q_{14}, q_{18}$. Consequently, even in assistance of optimized measurement points, we could not estimate more than these.

CHAPTER 7

Summary

Understanding biological processes which are altered in medulloblastoma is important for building a relevant mathematical model. There are four distinct molecular variants of medulloblastoma [49]. These four variants are characterized by different chromosome aberrations, diverse even in a singular medulloblastoma subgroup [55]. Furthermore, prognosis radically varies depending on the molecular diagnosis and major differences are also found between the adult and pediatric case.

In this work we focus on the aberration of chromosome 6q in the pediatric case only. Monosomy 6 (6q loss) is linked to good prognosis and trisomy 6 (6q gain) to poor prognosis. Both types involve a distortion in the expression of the target genes of the Wnt/β-catenin signaling pathway. This may suggest changes at the embryonal or developmental level and may explain the larger incidence of medulloblastoma among children than among grown-ups. We observe the correlation between the *SGK1* deregulation and prognosis, i.e., increase of *SGK1* seems to favor a negative influence on the patient's prognosis. Conversely, the *SGK1* downregulation is more frequent in good prognosis.

We build a novel mathematical model in the form of a system of nonlinear ODEs to investigate the different prognosis in the two types of medulloblastoma. We model interactions between particular genes involving biological processes such as transcription, translation, phosphorylation, degradation and transport between the nucleus and cytoplasm. Because of the complexity of the biological system we choose genes which seem to play a crucial role in the system, hence we model dynamics of the loop *SGK1-GSK3*β*-MYC*. Numerical simulations indicate the importance of *SGK1*. We simulate different scenarios for the monosomy 6 and trisomy 6 patient samples based on the microarray data from the clinics. Consistently, we reveal that high *SGK1* mRNA production strongly

correlates with the patient poor prognosis. It is obtained due to the comparison between the patient data and nuclear cMyc level obtained in the simulations. We elucidate the discrepancy between prognosis of the two types of medulloblastoma due to the SGK1 protein concentration. Based on the patient data we formulate hypothesis that gene *SGK1* is an essential factor in medulloblastoma treatment. We propose a pharmacological inhibition of the SGK1 protein as a novel way of the patient treatment. We find that inhibition of the SGK1 protein yields the best effect as a pharmacological intervention for the modeled system and leads to decrease of the nuclear cMyc and the GSK3β stabilization. Our results are in line with recent experimental results obtained in the laboratory of the group of Prof. Dr. med. Stefan Pfister at the Division of Pediatric Neuro-oncology Research Group of the German Cancer Research Center (DKFZ) showing that *SGK1* is an important gene in the investigated tumor and its inhibition entails remission of malignancy [56]. We can "translate" our mathematical study to an application in medicine, what was the main goal of this work.

In the analytical part we show well-possedness of the proposed model in the sense of classical theory of ODEs. We show global existence and uniqueness of solutions.

At next, we perform parameter estimation process in the framework of optimal experimental design. Using sensitivity analysis and a tool of confidence ellipses combined with Monte Carlo simulations we are able to refine seven estimates for the model. This result shows difficulties in the process of deterministic estimation of the nonlinear problem. Importantly, we indicate also that the most sensitive model parameter, for the species of the nuclear cMyc concentration, is responsible for the degradation of the SGK1 protein. Consequently, we underline the importance of *SGK1* also in the sensitivity analysis.

The question that still arises is whether the effects of *SGK1* deregulation are executed just through the *MYC* dynamics or if there are other target genes that are significant for the disease development. It means that further studies could expose on which proteins *SGK1* has further negative impact. The processes in tumor development are complex and one should follow many tests in order to deeply comprehend the biological dynamics.

We have a well-posed model and a qualitative solution based on the input in the form of microarray data (mRNAs), but what about quantitative solutions on the protein level? Todays laboratory techniques do not assure the necessary data. However, we may estimate parameters using surrogate data. During investigation, we see that the deterministic methods used here failed to estimate all parameters for our nonlinear problem. Therefore, with the help of stochastic methods, which are more sophisticated, we could try to obtain estimates for more model parameters. The model with refined parameters may be a powerful tool for more accurate studies (e.g., individual patient dose) on the adjuvant therapy reagarding the *SGK1* inhibition in near future.

Bibliography

[1] Angelika E Altmann-Dieses, Johannes P Schlöder, Hans G Bock, and Otto Richter. Optimal experimental design for parameter estimation in column outflow experiments. *Water Resources Research*, 38(10):4–1, 2002.

[2] Celina Ang, David Hauerstock, Marie-Christine Guiot, Goulnar Kasymjanova, David Roberge, Petr Kavan, and Thierry Muanza. Characteristics and outcomes of medulloblastoma in adults. *Pediatric blood & cancer*, 51(5):603–607, 2008.

[3] Maria F Arteaga, Diego Alvarez de la Rosa, Jose A Alvarez, and Cecilia M Canessa. Multiple translational isoforms give functional specificity to serum-and glucocorticoid-induced kinase 1. *Molecular biology of the cell*, 18(6):2072–2080, 2007.

[4] Samuel Bandara, Johannes P Schlöder, Roland Eils, Hans G Bock, and Tobias Meyer. Optimal experimental design for parameter estimation of a cell signaling model. *PLoS computational biology*, 5(11):e1000558, 2009.

[5] Julio R Banga. Optimization in computational systems biology. *BMC systems biology*, 2(1):47, 2008.

[6] Roland Becker, Malte Braack, and Boris Vexler. Numerical parameter estimation for chemical models in multidimensional reactive flows. *Combustion Theory and Modelling*, 8(4):661–682, 2004.

[7] Larissa Belova, Sanjay Sharma, Deanna R Brickley, Jeremy R Nicolarsen, Cam Patterson, and Suzanne D Conzen. Ubiquitin–proteasome degradation of serum-and glucocorticoid-regulated kinase-1 (sgk-1) is mediated by the chaperone-dependent e3 ligase chip. *Biochemical Journal*, 400(Pt 2):235, 2006.

[8] Hans G Bock, Thomas Carraro, Willi Jäger, Stefan Körkel, Rolf Rannacher, and Johannes P Schlöder. *Model based parameter estimation. Theory and applications.* Springer, 2013.

[9] Hans Georg Bock, Ekaterina Kostina, and Johannes P Schlöder. Numerical methods for parameter estimation in nonlinear differential algebraic equations. *GAMM-Mitteilungen*, 30(2):376–408, 2007.

[10] Carmen Chicone. *Ordinary differential equations with applications*, volume 34. Springer, 2000.

[11] Kwang-Hyun Cho, Sung-Young Shin, Walter Kolch, and Olaf Wolkenhauer. Experimental design in systems biology, based on parameter sensitivity analysis using a monte carlo method: A case study for the tnfα-mediated nf-κ b signal transduction pathway. *Simulation*, 79(12):726–739, 2003.

[12] Y-J. Cho, A. Tsherniak, P. Tamayo, S. Santagata, A. Ligon, H. Greulich, R. Berhoukim, V. Amani, L. Goumnerova, C.G. Eberhart, et al. Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *Journal of Clinical Oncology*, 29(11):1424–1430, 2011.

[13] Hans Clevers. Wnt/β-catenin signaling in development and disease. *Cell*, 127(3):469–480, 2006.

[14] Chi V Dang. c-myc target genes involved in cell growth, apoptosis, and metabolism. *Molecular and cellular biology*, 19(1):1–11, 1999.

[15] Chi V Dang, Linda Resar, Eileen Emison, Sunkyu Kim, Qing Li, Julia E Prescott, Diane Wonsey, and Karen Zeller. Function of the c-myc oncogenic transcription factor. *Experimental cell research*, 253(1):63–77, 1999.

[16] Manuel Dehner, Michel Hadjihannas, Jörg Weiske, Otmar Huber, and Jürgen Behrens. Wnt signaling inhibits forkhead box o3a-induced transcription and apoptosis through up-regulation of serum-and glucocorticoid-inducible kinase 1. *Journal of Biological Chemistry*, 283(28):19201–19210, 2008.

[17] Jacques Delforge, Andre Syrota, and Bernard M Mazoyer. Experimental design optimisation: theory and application to estimation of receptor model parameters using dynamic positron emission tomography. *Physics in medicine and biology*, 34(4):419, 1989.

[18] Bradley W Doble and James R Woodgett. Gsk-3: tricks of the trade for a multitasking kinase. *Journal of cell science*, 116(7):1175–1186, 2003.

[19] David W Ellison, Olabisi E Onilude, Janet C Lindsey, Meryl E Lusher, Claire L Weston, Roger E Taylor, Andrew D Pearson, and Steven C Clifford. β-catenin status predicts a favorable outcome in childhood medulloblastoma: the united kingdom children's cancer study group brain tumour committee. *Journal of Clinical Oncology*, 23(31):7951–7957, 2005.

[20] D Faller, U Klingmüller, and J Timmer. Simulation methods for optimal experimental design in systems biology. *Simulation*, 79(12):717–725, 2003.

[21] Gary L Firestone, Jennifer R Giampaolo, and Bridget A OKeeffe. Stimulus-dependent regulation of serum and glucocorticoid inducible protein kinase (sgk) transcription, subcellular localization and enzymatic activity. *Cellular Physiology and Biochemistry*, 13(1):1–12, 2003.

[22] Riccardo Fodde and Thomas Brabletz. Wnt/β-catenin signaling in cancer stemness and malignant behavior. *Current opinion in cell biology*, 19(2):150–158, 2007.

[23] Sheelagh Frame and Philip Cohen. Gsk3 takes centre stage more than 20 years after its discovery. *Biochemical Journal*, 359(Pt 1):1, 2001.

[24] Scott F Gilbert. *Developmental biology: The anatomical tradition*. Sinauer Associates, 2000.

[25] Michael D Gordon and Roel Nusse. Wnt signaling: multiple pathways, multiple receptors, and multiple transcription factors. *Journal of Biological Chemistry*, 281(32):22429–22433, 2006.

[26] Mark A Gregory, Ying Qi, and Stephen R Hann. Phosphorylation by glycogen synthase kinase-3 controls c-myc proteolysis and subnuclear localization. *Journal of Biological Chemistry*, 278(51):51606–51612, 2003.

[27] Narendra Gupta and Raman Mehra. Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. *Automatic Control, IEEE Transactions on*, 19(6):774–783, 1974.

[28] Eldad Haber, Lior Horesh, and Luis Tenorio. Numerical methods for experimental design of large-scale linear ill-posed inverse problems. *Inverse Problems*, 24(5):055012, 2008.

[29] Philip Hartman. *Ordinary differential equations.* John Wiley & Sons, Inc. New York-London-Sydney, 1964.

[30] Stefan Hoppler and Claire L Kavanagh. Wnt signalling: variety at the core. *Journal of cell science*, 120(3):385–393, 2007.

[31] Hans A Kestler and Michael Kühl. From individual wnt pathways towards a wnt signalling network. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1495):1333–1347, 2008.

[32] Ehab M Khalil. Treatment results of adults and children with medulloblastoma nci, cairo university experience. *Journal of the Egyptian National Cancer Institute*, 20(2):175, 2008.

[33] David M King and Chris Perera. Sensitivity analysis for evaluating importance of variables used in an urban water supply planning model. In *Proceedings of the International Congress on Modelling and Simulation (MODSIM'07)*, pages 2768–2774, 2007.

[34] Peter K Kitanidis and Robert W Lane. Maximum likelihood parameter estimation of hydrologic spatial processes by the gauss-newton method. *Journal of Hydrology*, 79(1):53–71, 1985.

[35] Alexandra Klaus and Walter Birchmeier. Wnt signalling and its impact on development and cancer. *Nature Reviews Cancer*, 8(5):387–398, 2008.

[36] Frank T Kolligs, Guido Bommer, and Burkhard Göke. Wnt/beta-catenin/tcf signaling: a critical pathway in gastrointestinal tumorigenesis. *Digestion*, 66(3):131–144, 2002.

[37] M. Kool, J. Koster, J. Bunt, N.E. Hasselt, A. Lakeman, P. van Sluis, D. Troost, N. Schouten-van Meeteren, H.N. Caron, J. Cloos, et al. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PloS one*, 3(8):e3088, 2008.

[38] Andrey Korshunov, Axel Benner, Marc Remke, Peter Lichter, Andreas von Deimling, and Stefan Pfister. Accumulation of genomic aberrations during clinical progression of medulloblastoma. *Acta neuropathologica*, 116(4):383–390, 2008.

[39] Andrey Korshunov, Marc Remke, Wiebke Werft, Axel Benner, Marina Ryzhova, Hendrik Witt, Dominik Sturm, Andrea Wittmann, Anna Schöttler, Jörg Felsberg, et al. Adult and pediatric medulloblastomas are genetically distinct and require different algorithms for molecular risk stratification. *Journal of Clinical Oncology*, 28(18):3054–3060, 2010.

[40] Roland Kruger and Reinhart Heinrich. Model reduction and analysis of robustness for the wnt/beta-catenin signal transduction pathway. *Genome Informatics Series*, pages 138–148, 2004.

[41] Julia Kzhyshkowska, Anna Marciniak-Czochra, and Alexei Gratchev. Perspectives of mathematical modelling for understanding of intracellular signalling and vesicular trafficking in macrophages. *Immunobiology*, 212(9):813–825, 2008.

[42] Ethan Lee, Adrian Salic, Roland Krüger, Reinhart Heinrich, and Marc W Kirschner. The roles of apc and axin derived from experimental and theoretical analysis of the wnt pathway. *PLoS biology*, 1(1):e10, 2003.

[43] Han C Lee, Miran Kim, Jack R Wands, et al. Wnt/frizzled signaling in hepatocellular carcinoma. *Front Biosci*, 11(5):1901–1915, 2006.

[44] Tomasz Lipniacki, Pawel Paszek, Allan R Brasier, Bruce Luxon, and Marek Kimmel. Mathematical model of nf-κ b regulatory module. *Journal of theoretical biology*, 228(2):195–215, 2004.

[45] Catriona Y Logan and Roel Nusse. The wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.*, 20:781–810, 2004.

[46] Franklin C McLean. Application of the law of chemical equilibrium (law of mass action) to biological problems. *Physiol. Rev*, 18:495, 1938.

[47] Christopher Z Mooney. *Monte carlo simulation*, volume 116. SAGE Publications, Incorporated, 1997.

[48] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer Science+ Business Media, 2006.

[49] Paul A Northcott, Andrey Korshunov, Hendrik Witt, Thomas Hielscher, Charles G Eberhart, Stephen Mack, Eric Bouffet, Steven C Clifford, Cynthia E Hawkins, Pim French, et al. Medulloblastoma comprises four distinct molecular variants. *Journal of Clinical Oncology*, 29(11):1408–1414, 2011.

[50] Alex B Novikoff. A transplantable rat liver tumor induced by 4-dimethylaminoazobenzene. *Cancer Research*, 17(10):1010–1027, 1957.

[51] Fabio Orlandi, Enrico Saggiorato, Giovanni Pivano, Barbara Puligheddu, Angela Termine, Susanna Cappia, Paolo De Giuli, and Alberto Angeli. Galectin-3 is a presurgical marker of human thyroid carcinoma. *Cancer research*, 58(14):3015–3020, 1998.

[52] Andrzej Palczewski. *Równania różniczkowe zwyczajne: teoria i metody numeryczne z wykorzystaniem komputerowego systemu obliczeń symbolicznych*. Wydawnictwa Naukowo-Techniczne, 1999.

[53] Stella Pelengaris, Mike Khan, and Gerard Evan. c-myc: more than just a matter of life and death. *Nature Reviews Cancer*, 2(10):764–776, 2002.

[54] Stefan Pfister, Christian Hartmann, and Andrey Korshunov. Histology and molecular pathology of pediatric brain tumors. *Journal of child neurology*, 24(11):1375–1386, 2009.

[55] Stefan Pfister, Marc Remke, Axel Benner, Frank Mendrzyk, Grischa Toedt, Jörg Felsberg, Andrea Wittmann, Frauke Devens, Nicolas U Gerber, Stefan Joos, et al. Outcome prediction in pediatric medulloblastoma based on dna copy-number aberrations of chromosomes 6q and 17q and the myc and mycn loci. *Journal of Clinical Oncology*, 27(10):1627–1636, 2009.

[56] Sabrina V Pleier, Dominik Sturm, Stefan Pfister, Marcel Kool, Anna Marciniak-Czochra, Joanna Kawka, et al. Serum- and glucocorticoid-regulated kinase 1 is a druggable regulator of cell migration and apoptosis in high-risk medulloblastoma. *In Review*.

[57] Paul Polakis. Wnt signaling and cancer. *Genes & development*, 14(15):1837–1851, 2000.

[58] Scott L Pomeroy and Lisa M Sturla. Molecular biology of medulloblastoma therapy. *Pediatric neurosurgery*, 39(6):299–304, 2004.

[59] Luc Pronzato. Optimal experimental design and some related control problems. *Automatica*, 44(2):303–325, 2008.

[60] M. Remke, T. Hielscher, A. Korshunov, P.A. Northcott, S. Bender, M. Kool, F. Westermann, A. Benner, H. Cin, M. Ryzhova, et al. Fstl5 is a marker of poor prognosis in non-wnt/non-shh medulloblastoma. *Journal of Clinical Oncology*, 29(29):3852–3861, 2011.

[61] Laurent Riffaud, Stephan Saikali, Emmanuelle Leray, Abderrahmane Hamlat, Claire Haegelen, Elodie Vauleon, and Thierry Lesimple. Survival and prognostic factors in a series of adults with medulloblastomas: Clinical article. *Journal of neurosurgery*, 111(3):478–487, 2009.

[62] Alessandra Rossi, Valentina Caracciolo, Giuseppe Russo, Krzysztof Reiss, and Antonio Giordano. Medulloblastoma: from molecular pathology to therapy. *Clinical cancer research*, 14(4):971–976, 2008.

[63] Kevin M Ryan and George D Birnie. Myc oncogenes: the enigmatic family. *Biochemical Journal*, 314(Pt 3):713, 1996.

[64] Einat Sadot, Maralice Conacci-Sorrell, Jacob Zhurinsky, Dalia Shnizer, Zeev Lando, Dorit Zharhary, Zvi Kam, Avri Ben-Ze'ev, and Benjamin Geiger. Regulation of s33/s37 phosphorylated β-catenin in normal and transformed cells. *Journal of cell science*, 115(13):2771–2780, 2002.

[65] HR Schlesinger, JM Gerson, PS Moorhead, H Maguire, and K Hummeler. Establishment and characterization of human neuroblastoma cell lines. *Cancer research*, 36(9 Part 1):3094–3100, 1976.

[66] George AF Seber and Chris J Wild. *Nonlinear regression*. John Wiley $\&$ Sons, New York, New York, USA, 2003.

[67] Lawrence F Shampine and Mark W Reichelt. The matlab ode suite. *SIAM journal on scientific computing*, 18(1):1–22, 1997.

[68] Perikles Simon, Michaela Schneck, Tabea Hochstetter, Evgenia Koutsouki, Michel Mittelbronn, Axel S Merseburger, Cora Weigert, Andreas M Niess, and Florian Lang. Differential regulation of serum-and glucocorticoid-inducible kinase 1 (sgk1) splice variants based on alternative initiation of transcription. *Cellular Physiology and Biochemistry*, 20(6):715–728, 2007.

[69] Gordon D Smith. *Numerical solution of partial differential equations: finite difference methods*. Oxford University Press, 1985.

[70] Michael D Taylor, Paul A Northcott, Andrey Korshunov, Marc Remke, Yoon-Jae Cho, Steven C Clifford, Charles G Eberhart, D Williams Parsons, Stefan Rutkowski, Amar Gajjar, et al. Molecular subgroups of medulloblastoma: the current consensus. *Acta neuropathologica*, 123(4):465–472, 2012.

[71] Esther M Verheyen and Cara J Gottardi. Regulation of wnt/β-catenin signaling by protein kinases. *Developmental Dynamics*, 239(1):34–44, 2010.

[72] Ferdinand Verhulst. *Nonlinear differential equations and dynamical systems*. Springer verlag, 1996.

[73] Karina J Versyck, Kristel Bernaerts, Annemie H Geeraerd, Jan F Van Impe, et al. Introducing optimal experimental design in predictive modeling: a motivating example. *International Journal of Food Microbiology*, 51(1):39, 1999.

[74] Brian J Wagner and Judson W Harvey. Experimental design for estimating parameters of rate-limited mass transfer: Analysis of stream tracer studies. *Water Resources Research*, 33(7):1731–1741, 1997.

[75] Hong-Xing Wang, Francis R Tekpetey, and Gerald M Kidder. Identification of wnt/β-catenin signaling pathway components in human cumulus cells. *Molecular human reproduction*, 15(1):11–17, 2009.

[76] Jianbo Wang and Anthony Wynshaw-Boris. The canonical wnt pathway in early mammalian embryogenesis and stem cell maintenance/differentiation. *Current opinion in genetics & development*, 14(5):533–539, 2004.

[77] Kan Wang, Shuchen Gu, Omaima Nasir, Michael Föller, Teresa F Ackermann, Karin Klingel, Reinhard Kandolf, Dietmar Kuhl, Christos Stournaras, and Florian Lang. Sgk1-dependent intestinal tumor growth in apc-deficient mice. *Cellular Physiology and Biochemistry*, 25(2-3):271–278, 2010.

[78] Russell H Zurawel, Sharon A Chiappa, Cory Allen, and Corey Raffel. Sporadic medulloblastomas contain oncogenic β-catenin mutations. *Cancer research*, 58(5):896–899, 1998.

List of Figures