UPTEC STS07 036
Examensarbete 20 p
November 2007

Battle Cancer Chemotherapeutic Drug Resistance

Using Cell Cycle Phase Models

Mikael Lindahl

UPPSALA
UNIVERSITET

Teknisk- naturvetenskaplig fakultet
UTH-enheten

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
http://www.teknat.uu.se/student

Abstract

Battle Cancer Chemotherapeutic Drug Resistance

Mikael Lindahl

Cancer patients develop chemotherapeutic drug resistance during repetitive treatments. Patients who developed drug resistance stop responding to treatment and their chances to survive drastically decrease. Alternative drugs that reverse the resistance and methods that can be used to find such drugs are needed.

In this thesis work a novel computational method for finding alternative drugs for treatment of resistant cancer cells has been developed. Mathematical models of the dynamic cell cycle have been developed and used to characterize dynamics of sensitive and resistant cell lines with as well as without drug treatments. Using Bayesian inference, a procedure for assigning probabilities to different candidate models given an observed cell cycle time series has been developed. The assigned probabilities were used to determine the drugs with the highest probabilities of reversing the drug resistance among a set of substances.

The method has been evaluated on in silico created experimental data of cell cycle progression. The result is promising, from a database containing cell cycle models for varies drugs the method successfully singled out the ones with ability to reverence the resistance.

Handledare: Mats Gustafsson
Ämnesgranskare: Rolf Larson
Examinator: Elisabet Andrésdóttir
ISSN: 1650-8319, UPTEC STS07 036

Populärvetenskaplig beskrivning

Cancer är ett samlingsnamn för sjukdomar där celler i kroppen gått in i ett irreversibelt tillstånd med konstant tillväxt. Det är en sjukdom som berör oss på många sätt. De flesta känner antagligen till någon som har eller har haft cancer och var tredje person i Sverige drabbas någon gång under sin livstid av cancer.

Cancer kan behandlas på tre sätt; genom strålning, kirurgi och cellgiftbehandling. Behandlingarna kan kombineras för att öka effektiviteten. I ett fåtal men farliga cancer sorter (bl.a. olika sorters blodcancer även kallat leukemi) är cellgift huvudbehandlingen. Utvecklande av resistens vid behandling av dessa cancrar är förödande och i värsta fallet är patienten dödsdömd efter att cancer cellerna utvecklat resistens. Vid resistens har cancercellerna utvecklat en mekanism som gör att cellgiftet fungerar avsevärt sämre på dem. En resistensmekanism kan till exempel vara att det skett en effektivisering av pumpproteiner i cell membranet som har till uppgift att pumpa ut cellgiftet. Det leder till att cellgiftnivån i cellerna inte når den nödvändiga koncentrationen för att stoppa celltillväxten. I ett sådant fall vore det önskvärt att identifiera en substans som kan hämma pumparna så att mindre cellgift pumpas ut och koncentrationen av cellgift når önskvärda nivåer.

I detta arbete har en metod utvecklas som kan användas till att identifiera läkemedel för behandling av resistenta cancerceller. Till det används modeller som beskriver cellcykeltillväxten över tiden. I metoden identifieras skillnaden mellan cellcykelmodeller av resistenta och känsliga cancerceller. Skillnaden beskriver hur cellcykelsvaret bör förändras för att resistenta celler ska övergå till att uppföra sig som känsliga. Vidare går det att för läkemedelskandidater identifiera hur de förändrar modellen för känsliga celler. Dessa modellförändringar kan sparas i en databas i vilken läkemedel kan sökas efter som genererar en liknande modellförändring som den mellan resistenta och känsliga celler.

Metoden testades på *in silico* skapad experimentellt cellcykeldata med en cell linje och 20 substanser. Den lyckades med att utskilja kandidat substanserna för behandling av resistenta cancerceller. Metoden är helt ny och i detta arbete har metoden utvecklats från scratch samt testat i ett pilotförsök för att undersöka dess potential. Nästa steg i utvecklingen är att använda metoden på experimentellt material från *in vitro* studier.

Table of contents

1 Introduction 5

1.1 Background 5
1.1.1 Chemotherapy 5
1.1.1.1 Cancer Chemotherapy 5
1.1.1.2 The relevance of the cell cycle in chemotherapy 6
1.1.2 Chemotherapy drugs resistance 6
1.1.2.1 Relevance in the medical field 6
1.2 Battling chemotherapy drug resistance 6
1.2.1 Conventional methods 6
1.2.2 Novel developments 7
1.2.2.1 Connectivity map 7
1.2.2.2 The work of Panetta et al 7
1.2.2.2.1 Relevance of cell cycle modeling 9
1.2.2.3 Limitations 12
1.2.2.3.1 C-Map 12
1.2.2.3.2 Cell cycle modeling 12
1.3 A novel dynamic approach to battle drug resistance 13
1.4 The aim of the thesis 13

2 Theory 14

2.1 The cell cycle phase model structure 14
2.2 Simulation of noisy trajectories 15
2.3 Parameter estimation 16
2.3.1 Method 16
2.3.2 Problems with parameter estimation 18
2.3.2.1 Identifiability 18
2.3.2.2 Estimability 18
2.4 Finding alternative drugs for treatment of resistant cancer cells 20
2.4.1 A general method finding alternative drugs 20
2.4.2 Finding alternative drugs using parameter intervals 22
2.4.2.1 Problems with intervals 23
2.4.3 Finding alternative drugs using Bayesian probability theory 24
2.4.3.1 Probability assumption 26
2.4.3.2 Estimating sub-model probabilities 26
2.4.3.3 Probability matrices 28
2.4.3.4 Discrete subinterval change 29
2.4.3.5 Probability of alternative drugs 29
2.4.3.6 Alternative drug search – summing up 30

3 In silico experimental setup 31

3.1 Experiments 31
3.2 In silico sensitive cell line and substances 31
3.2.1 Cell line 31
3.2.2 Substances 32

4 Results 35

4.1 Setup 35
4.1.1 Subintervals 35
4.2 Example 1 36
4.3 Example 2 39

4.4 Summary of the results 41

5 Discussion 42

5.1 Model relevance 42
5.1.1 Delay of maximum drug effect 42
5.1.2 Error model 44
5.2 In vitro applicability 45
5.3 Suggestion for future work 45
5.3.1 In silico substance library 45
5.3.2 Covering different alternative drugs search 45
5.3.3 Subintervals 46
5.3.4 Drug classification 47

6 Conclusions 48

6.1 Results 48
6.1.1 Model generalization and simulation 48
6.1.2 Cell lines and cells in silico 48
6.1.3 Alternative drug search method 49
6.1.4 Summary 49
6.2 Limitations 50
6.3 Future work 50
6.3.1 Further computational investigation 50
6.3.2 In vivo experiments 50

7 Bibliography 51

8 Acknowledgements 53

Glossary

Arrayscan™: is an instrument which optically scans plates and count the number of cells with specific biomarkers.
Flow cytometry: it is a technique for counting, examining and sorting cell suspended in a stream of fluid. It allows simultaneous multiparametric analysis of the physical and/or chemical characteristics of single cells flowing through an optical and/or electronic detection apparatus.
Metastasis: the spread of cancer from one part of the body to another.
Mitosis: the process in cell division (mitotic cell – a cell that undergoes cell division)
In vitro: experiment in an artificial environment, such as a test tube.
In silico: experiment within a computer-simulated environment.
Proliferation: growth (often used when talking about cell growth)
Reversed chemical control: finding alternative drugs which can make resistant cells sensitive again

Contents overview

Chapter 1: Introduction, introduction and background to the subject and goal of this work.

Chapter 2: Theory, description of the mathematical cell cycle phase model, of the model parameter estimation and of the alternative drug search method.

Chapter 3: In silico experimental setup, a description of the setup and assumption of the in silicon experiments used to evaluate alternative drug search method. a method which can be used to find alternative drugs for treatment of resistant cancer cells.

Chapter 4: Results, the results of the performance of the alternative drug search method when used upon the in silico created experimental data.

Chapter 5: Discussion, discussion of the result, theoretical assumptions, and future problems to be solved.

Chapter 6: Conclusions.

Chapter 7: Bibliography.

1 Introduction

This Master of Science thesis work has been conducted at the Unit of Clinical Pharmacology at the Department of Medical Sciences at Uppsala University. The goal has been to create a method which uses mathematical cell cycle phase models of treated and untreated cancer cells to identify drugs and substances for refined treatment of resistant cancer cells.

1.1 Background

1.1.1 Chemotherapy

Chemotherapy was coined by Ehrlich at the beginning of the century and was meant to describe the use of chemical compound in the destruction of infective agents (Pang et al 1999). The definition of chemotherapy today has broadened and now also describes the use of antibiotics – substances produced by microorganisms that inhibit the growth of other microorganisms. The term chemotherapy is now also used to describe synthetically or organically produced chemical compounds which are used to inhibit the growth of malignant or cancerous cells within the body. (Pang et al 1999)

1.1.1.1 Cancer Chemotherapy

Cancer chemotherapy is one out of three main approaches for treating established cancer. The other two are surgical excision and irradiation. The approach to be used depends upon the type of cancer and the stage of its development. Chemotherapy is only the main treatment method for a few types of cancers[1] but it is often used in combination with surgery or irradiation. It has been proven difficult to identify cancer cell specific properties in comparison to normal cells which chemotherapy substances can target. There exist four main characteristics that in varying degree distinguish them from normal cells; uncontrolled cell proliferation, dedifferentiation, invasiveness and metastasis.(Pang et al 1999) In uncontrolled proliferation the normal processes that regulate cell division are disabled. Dedifferentiated cells have loss their ability to differentiate from stem cells into mature cells such as muscle or liver cells. Cells that are invasive have the ability to function outside there tissue origin, for example liver cell that appears in the bladder. Metastases occur when a secondary tumor is developed out of cells from the primary tumor at a new location in the body. These characteristic vary between cancers which means that drugs targeting one of these characteristic can have varying effects upon different cancers.

[1] For example: Hodgkin's disease, Non-Hodgkin's lymphoma, Chronic granulocytic leukemia, Acute lymphocytic leukemia, Hairy cell leukemia, Germ cell cancer (festis, ovary), Choriocarcinoma and Prostate cancer.

1.1.1.2 The relevance of the cell cycle in chemotherapy

The mitotic cell cycle can be considered to consist of four different phases G1 (gearing up before DNA replication), S (DNA replication), G2 (Cell division preparation) and M (mitosis or cell division) (See figure 2). A property of cancer cells is that they are constantly dividing and therefore always is in the mitotic cell cycle. By contrast normal cells are often placed in the passive G0 phase. Many chemotherapy cancer drugs acts upon the cell cycle forcing the cancer cells to go into apoptosis programmed cell death. Most of the current available chemotherapeutic drugs such as cytarbine, hydroxyuera, flouracil, methotrexate and merceptopurin act in S phase but some also act in M phase such as the vinca alkaloids. Some of these compounds also act upon G1 phase. Moreover, there are also a number of drugs like alcylating agents, dactinomyecin, doxorubicin and cisplatin which have no cycling specific inhibitor effect. (Pang et al 1999) The fact that many chemotherapeutic drugs act selectively (but also non-selectively) upon the cell cycle is a reason to why the cell cycle as model can help to characterize chemotherapeutic drug effects.

1.1.2 Chemotherapy drugs resistance

The resistance of cancer cells against cytotoxic drugs can either be present when the drug is first given or acquired during treatment. Resistance can be acquired through adaptation with the emergence of cells which are less effected by the drug and therefore has an selective advantage over the sensitive cells. (Pang et al 1999)

1.1.2.1 Relevance in the medical field

The development of chemotherapeutic drug resistance in cancer cells is a very serious problem. Studies have shown that cancer cell which becomes resistant towards a specific drug during chemotherapeutic treatment also may become cross-resistant towards other drugs with different drug mechanisms. Moreover an alarming fact is that drug resistance is thought to cause treatment failure in over 90 percent of patients with metastatic cancer. (Longly and Johnston 2005) Thus if drug resistance would be overcome then the treatment survival rate would be significantly increased. Refined methods which can combat drug resistance are clearly needed.

1.2 Battling chemotherapy drug resistance

1.2.1 Conventional methods

A natural approach to combat resistance would be to identify alternative drugs which could suppress the resistance mechanisms making the resistant cancer cells sensitive again. However, how to find such drugs is a highly nontrivial task. The standard approaches today are high throughput chemical screening for new compounds and target identification via gene expression microarrays comparing sensitive and resistant cells. Unfortunately, microarrays are expensive to produce and high throughput chemical screening is both expensive and difficult to accomplish considering the large number of possible molecules in chemical space.

1.2.2 Novel developments

1.2.2.1 Connectivity map

Fortunately a novel strategy which we here denote *reversed chemical control* was recently invented and published in Science by Lamb et al (Lamb et al 2006). Lamb et al identified 164 drug induced mRNA signatures. The MCF7 cancer cell line was treated with each of the 164 drugs and the mRNA levels were measured using microarrays at 0 hours and 6 hours. Then for each drug the difference between the mRNA expression levels obtained at 0 and 6 hours were calculated and stored in a database. Further analyzes with pattern-matching tools which can identify similarities among the signatures could then be performed. The resource they developed was referred to as the "Connectivity Map" (C-Map) due to its prospective in revealing "connections" among drugs, genes and diseases.

Microarrays can be used to identify how the gene activity differs between treated resistant and treated sensitive cancer cells. The mRNA activity induced in a cell by a drug is measured by incubating the cancer cells together with the drug for 6 hours. At the end of the time period mRNA are extracted and the mRNA activity are measured using microarrays. The differences in gene levels between treated sensitive and treated resistant cancer cells are then identified. Finally the resistance mechanism is then reversed using drugs that induce the opposite gene activity. Such drugs are search for in the C-Map database were the changes in mRNA activity induced by different drugs in sensitive MCF7 cells are stored. Treating the resistant cancer cells with a drug that induces the opposite change in gene activity will hopefully then make the resistant cells sensitive again through the reversal of mRNA activity of the resistant cells. The resistant cells should then respond as treated sensitive cells to a combination treatment consisting of this alternative drug found and the original drug resistance has been developed against. The method presented in this thesis may be regarded as an attempt to generalize the C-Map approach to accomplish reversed chemical control but instead of using static mRNA signatures, dynamic cell cycle models signatures are used.

1.2.2.2 The work of Panetta et al

The background to the cell cycle model used in this paper can be found in the two papers "*A mathematical model of in vitro cancer cell growth and treatment with the antimitotic agent curacin A"* (Kozusko et al 2000) and "*Mechanistic mathematical modeling of mercaptopurine effects on cell cycle of human acute lymphoblastic leukaemia cells*" (Panetta et al 2006) The first paper focus on modeling drug dosing effect and the second paper focus on revealing mechanisms behind drug resistance through modeling. In both articles a cell cycle model is developed to describe the dynamics of cancer growth.

Cell cycle model of the first paper has in the second paper "*Mechanistic mathematical modeling of mercaptopurine effects on cell cycle of human acute lymphoblastic leukaemia cells* "been expanded. Instead a two compartments model representing cell cycling now a three compartment model is employed. To give a background to the origin the cell cycle model used in this work the second paper briefly will be presented and discussed.

The second paper deals with the problem of cancer drug resistance. A mathematical model supposed to capture the dynamics of lymphoblastic leukemia cells treated with mercaptopurine (MP) an antimetabolic agent in one sensitive - Molt-4 sensitive - and two resistant – Molt-4 resistant and P12 – cell lines has been developed. The results from the mathematical modeling showed that the MP sensitive cells lines had a significantly higher rate of entering apoptosis (2.7 fold) compared to the resistant cell lines. In addition the model revealed that when treated with MP, the Molt-4 sensitive cell lines showed a significant increase in the rate at which cells entered apoptosis (2.4 fold) compared to its control. Also the model suggested that resistant cell lines had a higher rate of antimetabolite incorporation (1.4 fold) into the DNA of viable cells. Finally, in contrast to the other two cell lines the model showed how the Molt-4 resistant cell line continued to cycle after incorporation of MP into their DNA though at a slower rate then its control rather then entering apoptosis. This led to a large S-phase block in the Molt-4 but not a higher rate of cell death.

The model they used is shown below in figure 1. Untreated cells are assumed to behave as a system with 5 compartments G0/G1, S, G2/M, A and N. When treated with MP the cell lines are assumed to behave as a system with 3 additional compartments G0/G1*I*, S*I* and G2/M*I*. Furthermore, when treated with MP, *f* in the model represent the fraction of cells which continues through cell cycle of untreated cells at least one more time before entering apoptosis and conversely 1-*f* represents the fraction of cells which goes into the treated cell cycle.

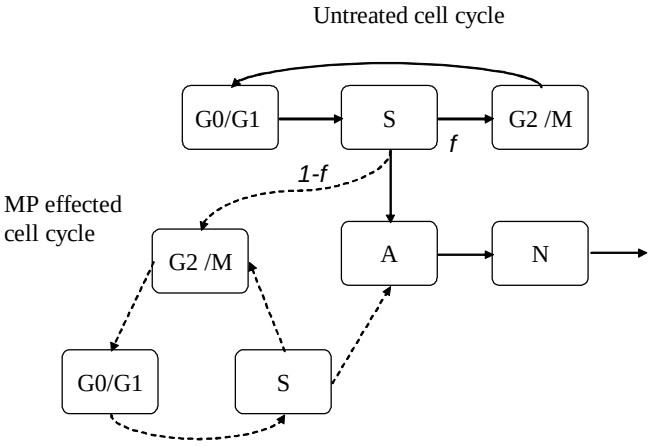


Figure 1 - Panetta's compartment model over the cell cycle and cell death process of untreated and treated cancer cells. The solid arrows connect the compartments used to model untreated cell cycle. The solid arrows together with the dashed arrows connect the compartments used to model MP treated cells. The compartments extension which the dashed arrows connect describes the cell cycle of MP incorporated cancer cell. The parameter f represents the fraction of cells which continues through the non-treatment effected cell cycle and conversely 1-f represents the fraction of cells which goes into the MP incorporated cell cycle.

The authors measured five quantities; total cell count, cell cycle distribution, percent viable, percent apoptotic and percent death of three cell lines P12, Molt-4 sensitive and Molt-4 resistant. They conducted experiments measuring the five cell quantities every 12, 24, 48 and 72 hour. Then the model parameters were fitted to the experimental values by using the program ADAPT II[2]. ADAPT II is a tool for analyzing pharmacokinetic and pharmacodynamic systems developed by Dr. David Z. D'Argenio at the University of Southern California. Finally models well fitted to the experimental material

1.2.2.2.1 Relevance of cell cycle modeling

The value of mathematical cell cycle modeling can be questioned. Could the conclusion drawn by Panetta be obtained without cell cycle modeling? The main conclusions reached are listed below:

I. MP sensitive cell lines had a significant higher rate of entering apoptosis (2.7 fold) compared to resistant cell lines.
II. MP treated sensitive cell lines showed a significant increase in the rate at which cells entered apoptosis compared to control (2.4 fold).
III. The resistant cell lines had a higher rate of MP incorporation into there DNA. (1.4 fold).
IV. Molt-4 continued through cell cycle at a lower rate than its controls after incorporation of MP into the DNA witch led to an S-phase block instead of entering apoptosis.

The results from the experiment on Molt-4 sensitive cell line presented in figures 3 shows that that the fraction of dead cells increases rapidly and the fraction of viable decreases rapidly. Between these two stages the apoptotic stage is settling after approximately 12 hours at a constant level. Furthermore no S-phase halt can be observed. To be able to explain the rapid decrease in viable cells and the rapid increase in non-viable cells one may be compelled to draw the conclusion that the rate of which cells entering apoptosis has increased. Why use a mathematical model to draw conclusions I if it can be reached by directly interpret the experimental data? The answer depends on whether the knowledge about the magnitude of the decrease is important or not and if it could be useful in further studies. It could be important if the goal is to reverse the process, that is force the resistant cells to respond as the sensitive cells to treatment. If you assume that it is possible to produce a drug which manipulates a certain rate then by knowing the exact change of rate between resistant cells and sensitive cells you could introduce a drug compensating for the rate difference forcing a resistant cell to become sensitive. Conclusion II is in the same way possible to reach based exclusively upon the experimental data as conclusion I. The experimental data shows that more Molt-4 sensitive cells die when MP is added. This implies that the rate cells entering apoptosis has increased. As for conclusion I the non-obvious information from models are the information about the magnitudes of the relative increases in apoptotic rate.

[2] American Type Culture Collection, Rockville, MD, USA

Conclusion III and IV are more interesting than I and II because they could not been drawn without the information from the mathematical model. The amount of cells incorporated with MP can not be directly observed. Thus no conclusion about differences in rates of MP incorporation between cell lines can be drawn by direct observation. Moreover, it is not possible to say anything from the experimental data about whether the cell cycle continues after MP incorporation. Conclusion III and IV are therefore non-trivial conclusions suggested by the mathematical modeling. Furthermore, the information about the rate change of MP incorporation is useful if the goal is to compensate rate change in resistant cells by adding new drugs.

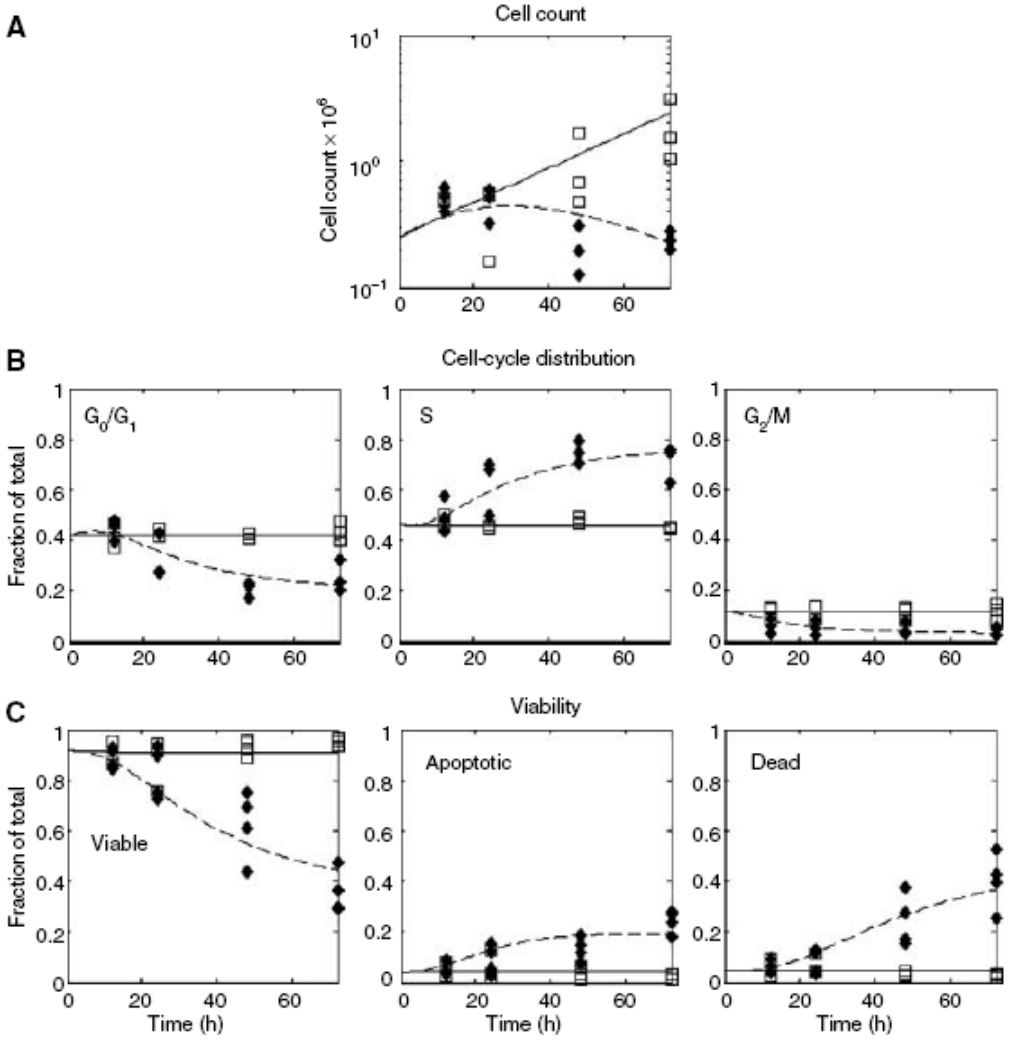


Figure 2 - P12 cell line: - □ -Control data. - ♦ -MP treated data. The solid line represents the model fit to the control data and the dashed line represents the model fit to MP-treated data. (A) Total cell counts verses time in hours. (B) Cell cycle distribution (that is fraction G0/G1, S, G2/M, respectively). (C) Cell viability distribution (that is fraction viable, fraction apoptotic and fraction dead respectively. (Panetta et al 2006)

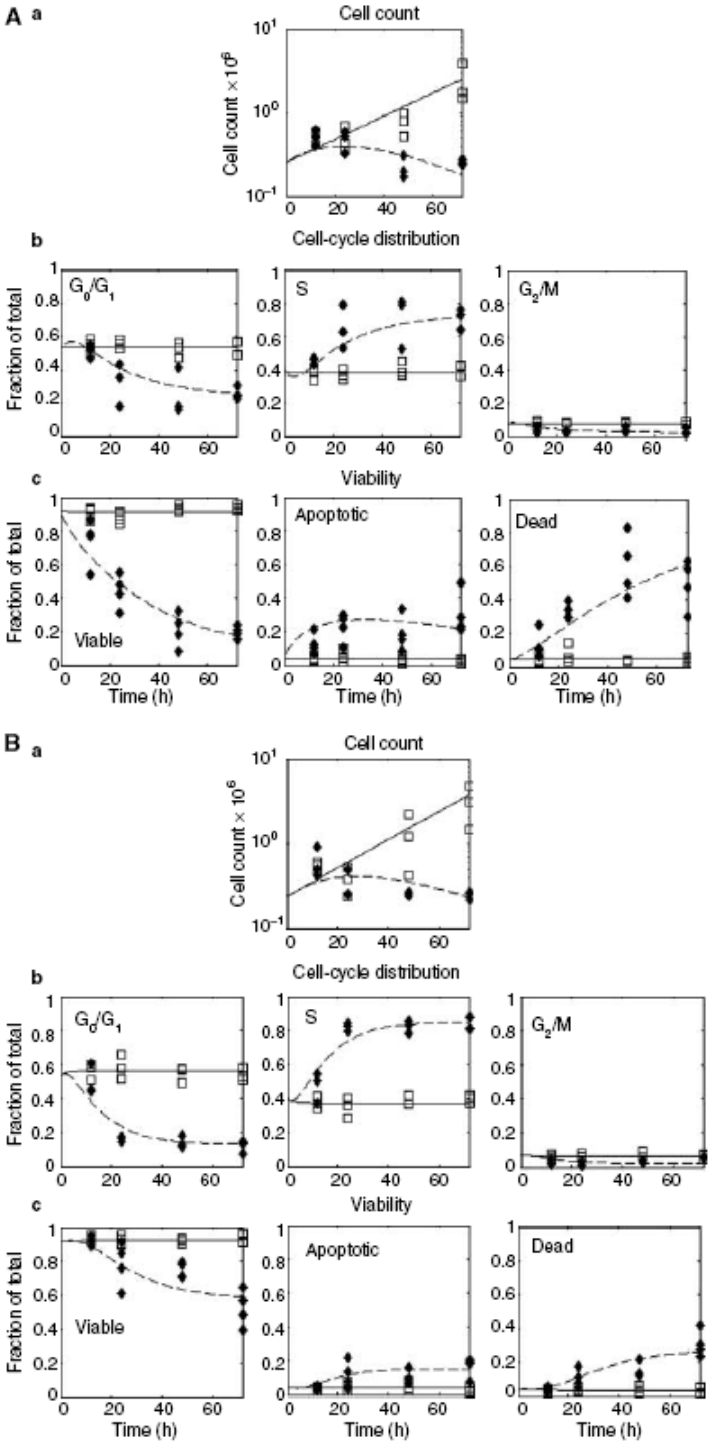


Figure 3 - Molt cell line. (A) Molt-4 sensitive cell line and (B) Molt-4 resistant cell line: - □ - Control data. - ♦ -MP treated data. The solid line represents the model fit to the control data and the dashed line represents the model fit to MP-treated data. Panel a: Total count versus time in hours. Panel b: Cell cycle distribution (fraction G0/G1, S G2/M, respectively). Panel c: Cell viability (that is fraction of viable, fraction of apoptotic, and fraction of dead respectively). (Panetta et al 2006)

The facts about how the cell lines react when incorporated with MP are non trivial information obtained from the mathematical modeling are. From the models one finds that Molt-4 sensitive and P12 cells continue through one cell cycle but no more after MP incorporation and that Molt-4 resistant cells continue through even more cycles. Important to remember is that the relevance of such non-trivial conclusions are dependent on the reliability of the mathematical model employed. Fore example no information of the MP incorporated cell cycle had been obtained if not the occurrence cell cycling after incorporation in all cell lines had been assumed. The non trivial results depend on the validity of the mathematical model. The validity of the model could be tested using drugs targeting certain rates. For example experiments could be performed with a drug inhibiting the incorporation of MP. The rate of MP incorporation is then expected to have decreased in the model. If it is not true then the model validity could be questioned.

Summary, it is important to think about the usefulness of the information that mathematical models can give because otherwise the modeling could become a waste of time. Moreover, one should always ask whether the model accurately describes what actually happens. Finally the model that Panetta uses differs in some aspects from the model used in this work. The presentation of Panetta's model is intended to illustrate a state-of-the-art application of cell cycle modeling as a background to current work in the field of cell cycle modeling.

1.2.2.3 Limitations

1.2.2.3.1 C-Map

The mRNA microarray data that are stored in the C-Map data base only contain static information from a single difference between two time points (the difference in mRNA levels between 0 and 6 hours). They do not give any dynamic information about how drug induced cells are effected over time. Moreover the microarrays used are expensive which limiting to C-Map research.

1.2.2.3.2 Cell cycle modeling

The experiment using the commercial method flow-cytometry to measure the fraction of cells in each cell cycle phase is less expensive than microarrays. The cell cycle modeling performed by Panetta et al is scientifically important for understanding of the mechanism behind MP-resistance but to a person concerned with the large scale discovery of drugs which can overcome resistance it is of limited help. Panetta does not claim that the cell cycle model developed can be used to study the effect of other chemotherapeutic drugs other than mercaptopurine. The model is created based upon pre-knowledge of how mercaptopurine works which limits the applicability of the model. For example the model can not be used to identify if apoptosis flows other than S-phase is triggered since only the S-phase flow is allowed in the model.

1.3 A novel dynamic approach to battle drug resistance

As already mentioned, the C-Map approach is promising but has the limitations of being non dynamic and expensive. It would be beneficial to have a similar approach which is dynamic and less expensive. The cell cycle model used by Panetta is dynamic but its applicability is limited. What if the applicability of the cell cycle models could be extended so that the cell cycle model could be used to characterize all known and candidate chemotherapeutic drugs? These models could then be used accordingly to the idea of reversed chemical control. For example, cell cycle models could be built which characterized selected chemotherapeutic drugs and then stored in a database. The cell cycle change between resistant cells and sensitive cells could then be estimated. Finally a search could be performed where drugs reversing the identified cell cycle change are found. In order to realize the above described the following two points need to be accomplished.

I. Build a cell cycle model which can be used to characterize known and unknown cell cycle specific chemotherapeutic drugs.
II. Invent a method which uses the general cell cycle models accordingly to the C-Map approach of inverse chemical control in the search of drugs for treatment of resistant cancer cells.

1.4 The aim of the thesis

The aim of this thesis work was to investigate the potential pursuing and implementing the novel dynamic approach described in 1.3. More specifically the goals may be summarized as:

I. Generalization and simulation of the cell cycle model developed by Panetta et al such that it can be used to characterize all known and candidate cell cycle effective chemotherapeutic drugs.
II. Development of artificial (in silico) cancer cell lines defined by parameters in the cell cycle model (chapter 3).
III. Development of artificial drugs that reflects different perturbation of the cell cycle parameters (chapter 3).
IV. Development of “mutated” cell cycle models in which the model parameters have been perturbed in such a way that the mutated cell lines become resistant to the original treatment (chapter 4).
V. Evaluate the possibility to perform estimation of the model parameters from time-series data to define unique cell cycle model “fingerprints” or “signatures” similar to those in C-Map database discussed earlier (chapter 2).
VI. Develop an alternative approach to parameter estimation based on Bayesian model selection for defining useful “fingerprints” (chapter 2).
VII. Perform in silico evaluation of the Bayesian fingerprint approach for combating drug resistant cancer cells (chapter 4).

2 Theory

2.1 The cell cycle phase model structure

The mitotic cell cycle is usually represented by five states G0, G1, S, G2 and M (see figure 5) (Weinberg 2007). In the model used here the cell cycle is described by three states G0/G1, S and M (see figure 4, 5 and equation (1)). The resting phase G0 and separation phase for DNA duplication G1 has been combined into one state G0/G1. The majority of the cells in G0/G1 will most likely be G1 phase cells since most cancer cells constantly are dividing and therefore never are in the resting phase G0. Furthermore, the preparation phase for cell division G2 and the cell division phase M are combined into one state G2/M. An important aspect of this representation is that S and M phase are in separate states. Most of the known chemotherapeutic drugs (Pang et al 1999) act in either S or M phase which makes it important to have a model which can classify between S and M specific drugs.

It has been shown that the protein machinery needed for a cell to enter apoptosis is present in all phases (Alenzi et al 2004). Thus it is reasonably to assume that apoptosis can be triggered in any phase of the cell cycle. The cell cycle model (figure 1) therefore has apoptosis flows from all cell cycle states.

The cell cycle model (figure 4, (1)) has four compartments or state-variables G0/G1-phase (G), S-phase (S), G2/M-phase (M) and apoptosis-phase (A). Furthermore, there are seven state-flows, three describing the flows between cell cycle phases p_{GS} (G0/G1→S), p_{SM} (S→G2/M) and p_{MG} (G2/M→G0/G1) three describing the apoptosis flows one from each cell cycle phase, p_{GA} (G0/G1→A), p_{SA} (S→A) and p_{MA} (G2/M→G0/G) and one describing how rapidly apoptosis cells become nonviable p_{AN} A→N.

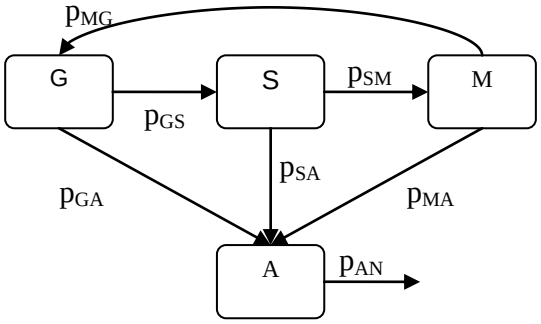


Figure 4 - The cell cycle model has four state-variables, G0/G-phase (G), S-phase (S), G2/M-phase (M) and apoptosis (A) and seven state-flows p_{GS} (G0/G1→S), p_{SM} (S→G2/M), p_{MG} (G2/M→G0/G1), p_{GA} (G0/G1→A), p_{SA} (S→A), p_{MA} (G2/M→G0/G) and p_{AN} (A→N).

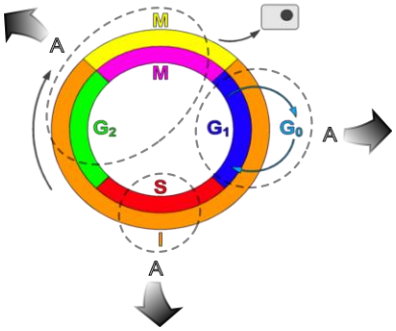


Figure 5 - The figure shows how the general 5 phase model representation of the cell cycle has been reduced into the 3 phase model representations used in this work. The resting phase G0 is combined with the cell duplication preparation phase G1 and cell division preparation phase G2 is combined whit the division phase M. The arrows represent the apoptosis flows.

Four differential equations (1) governs the change of the four state-variables G0/G1-phase (G), S-phase (S), G2/M-phase (M) and apoptosis (A) over time.

$$\begin{aligned}
\frac{dG}{dt} &= 2p_{MG}M - p_{GS}G - p_{GA}G \\
\frac{dS}{dt} &= p_{GS}G - p_{SM}S - p_{SA}S \\
\frac{dM}{dt} &= p_{SM}S - p_{MG}M - p_{MA}M \\
\frac{dA}{dt} &= p_{GA}G + p_{SA}S + p_{MA}M - p_{AN}A
\end{aligned} \qquad (1)$$

2.2 Simulation of noisy trajectories

Given that the parameters in equation (1) are known then can one perform model simulations and generate time series of the cell cycle phase progression of G, S, M and A. From this data samples G_t, S_t, M_t and A_t can be obtained at different time points $t=1\ldots T$. Vectors of noisy trajectories G^*, S^*, M^* and A^* can then be created by adding for example normal distributed errors $errG_t$, $errS_t$, $errM_t$ and $errA_t$ to each sampled data point as described in equation (2).

$$\begin{aligned}
G_t^* &= G_t + errG_t \\
S_t^* &= S_t + errS_t \\
M_t^* &= M_t + errM_t \\
A_t^* &= A_t + errA_t
\end{aligned} \qquad (2)$$

2.3 Parameter estimation

2.3.1 Method

To construct a cell cycle model describing the behavior of drug treated or untreated cancer cells we have to estimate the parameters p_{GS}, p_{SM}, p_{MG}, p_{GA}, p_{SA}, p_{MA} and p_{AN} of the system of differential equations in (1). The process of parameter estimation procedure is illustrated in figure 6. First experiments are conducted (A) in which the number of cells in each compartment is sampled at specific time points (B). The experimental data is then stored in a computer (C). A computer program is then used to find the model parameters which produce the best model fit to the collected data. (D). The best fit will have the minimal error between model simulated values and observed values. Finally the best result is displayed (E).

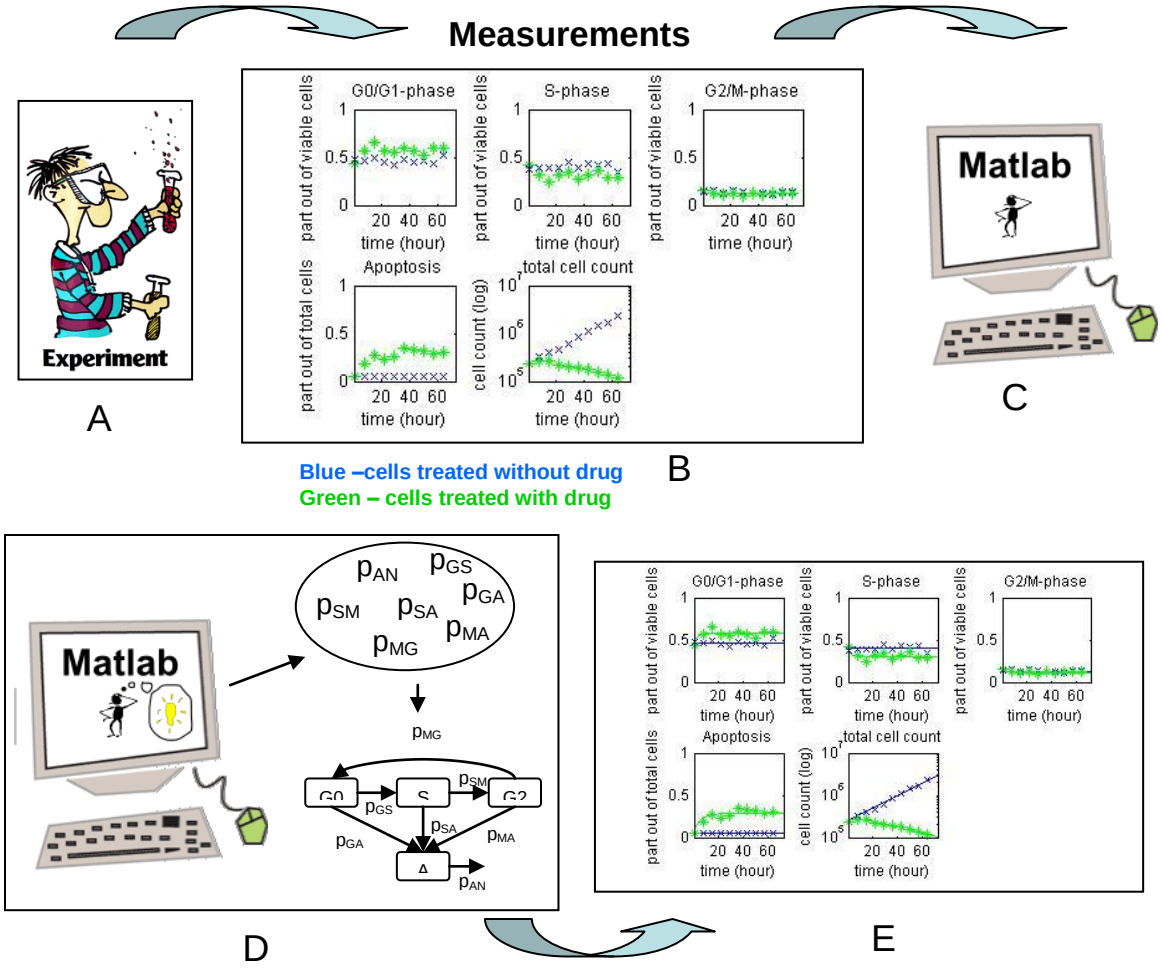


Figure 6 - An overview of the process of parameter estimation. From experiments (A) is time series retrieved (B) which are put into the computer (C). A computer program estimates the model parameters (D) and for the parameters resulting in the best model fit, the results time profiles are displayed (E).

Parameter estimation is commonly done by minimizing an objective function having the parameters as variables. The objective function describes the deviation between model simulated values and the actual experimental time profiles. (Ljung 1987) The objective function used in this work is the total error *e(p)* presented in equation (3). *Experimental* time series of the state-variables $G_t^{experimenal}$, $S_t^{experimenal}$, $M_t^{experimenal}$ and $A_t^{experimenal}$ $t=,\ldots,T$ where T is the number of sample time points are directly measured in vivo experiments. Furthermore *simulated* time series of the state-variables $G_t^{simulated}$, $S_t^{simulated}$, $M_t^{simulated}$ and $A_t^{simulated}$ $t,\ldots,T$ are calculated given a set of parameters *p*. For each parameter vector *p* the total error *e* in equation (3) is calculated from the experimental and the simulated data.

$$
\begin{aligned}
e(\bar{p}) = \sum_{t}^{T} \Big[& \left(G_t^{experimental} - G_t^{simulated}(\bar{p})\right)^2 + cdcs\left(S_t^{experimental} - S_t^{simulated}(\bar{p})\right)^2 \\
& + \left(M_t^{experimental} - M_t^{simulated}(\bar{p})\right)^2 + \left(A_t^{experimental} - A_t^{simulated}(\bar{p})\right)^2 \Big] \\
\bar{p} = & \left[p_{GS} \quad p_{SM} \quad p_{MG} \quad p_{GA} \quad p_{SA} \quad p_{MA} \quad p_{AN}\right]
\end{aligned}
\tag{3}
$$

The objective function can be minimized using the multidimensional nonlinear minimization method *fminsearch* in Matlab (Matlab 2007). Fminsearch is based upon Nelder-Meads[3] nonlinear optimization algorithm also known as the simplex method. The method approximately finds a locally minimal solution to a problem with *N* variables when the objective function varies smoothly.

The problem with simplex algorithms like fminsearch is that it often gets stuck in local optimal solutions instead of proceeding to the global optimal solution. This can happen if the initial parameter guess used is to far from the global optimal solution. There is no absolute solution to this problem. One way to increase the probability of finding the global optimum is for example to solve the minimization problem for a large number of initial parameter guesses and then choose the best solution.

Another way to minimize (3) instead of using the simplex method would be to use a genetic algorithm (for example see Goldberg 1989). Genetic algorithms are stochastic iterative processes which evolves a population of candidate solutions that is replaced in each iteration (generation) by a novel population created by mechanisms inspired by genetic inheritance such as chromosomal cross-over and mutations. A genetic algorithm was never used since the simplex method worked well on the problem considered in this work.

[3] See Nelder-Mead http://en.wikipedia.org/wiki/Nelder-Mead_method (2007-03-14).

2.3.2 Problems with parameter estimation

For a given application there are two potential problems associated with the parameter estimation step:

I. *Identifiability*, given that we have perfect measurements is there some parameters in the model which *a priori* can not be identified?
II. *Estimability*, although all parameters are identifiable can they practically be estimated when we have short time series and noisy measurements?

2.3.2.1 Identifiability

The following definition of identifiability is given by John Jacquez and Peter Greif (Jacquez and Grief 1985); given a model of a system and specific input-output experiments, with error free data, are all the model parameters uniquely determined? Thus, identifiability analysis is concerned with the problem of whether the parameters are possible to estimate given an experimental setup. Consequently, if some parameters are not identifiable then are they impossible to estimate using the current model and experimental setup. Rither more additional quantities have to be measured or the number of parameters in the model has to be reduced in order to make all parameters identifiable.

Two kinds of identifiability analysis can be performed; global and a local. Assume that a model that can explain a noise free observation y as $y=f(p)$. A global identifiability analysis tells us whether p is uniquely determined by the equation $f(p)=y$, if p is a set of solutions p_n $n=1,\ldots,N$ or undetermined having infinitely many solutions. A local identifiability analysis only tells us that p either is finitely determined (possible uniquely) or undetermined by $f(p)=y$.(Audoly 1998) There exist methods (Audoly et al 1998, Audoly et al 2001) to test for global identifiability in both the linear and the non-linear case which involve non-trivial algebraic mathematics. These methods have not been implemented due to lack of time. Global identifiability falls outside the scope of this thesis.

The method used in this thesis work for local identifiability analysis of linear and nonlinear compartments models was developed by John Jacquez and Timothy Perry presented first in *Endocrinology and metabolism* (Jacquez and Perry 1990). The result from the identifiability analysis using Jacquez and Perry's method showed that all of the parameters in equation (1) were identifiable. The identifiability analysis was conducted in order to confirm whether the cell cycle phase model would prove to be useful or not in the applications of interest.

2.3.2.2 Estimability

It can be difficult to estimate the parameters even if all the parameters are locally identifiable. In reality we often have few data sampling points and experimental measurement disturbances. Because of this even the correct parameter solution will produce an error e (equation 2). For example if a parameter has bad estimability then we

will be able to change that parameter significantly and still be within an acceptable error range.

Parameter estimation studies of the cell cycle model (1) showed that the estimability of the apoptosis model parameters p_{GA}, p_{SA} and p_{MA} are poor. Figure 7, which show six good model fits to data with 10 percent disturbance, illustrates the poor estimability of the three apoptosis parameters. Figure 8 shows the variety of the apoptosis parameters that apparently yield similar time profiles close to the observations. The consequence is that the apoptosis parameters can not be determined with any reasonable accuracy.

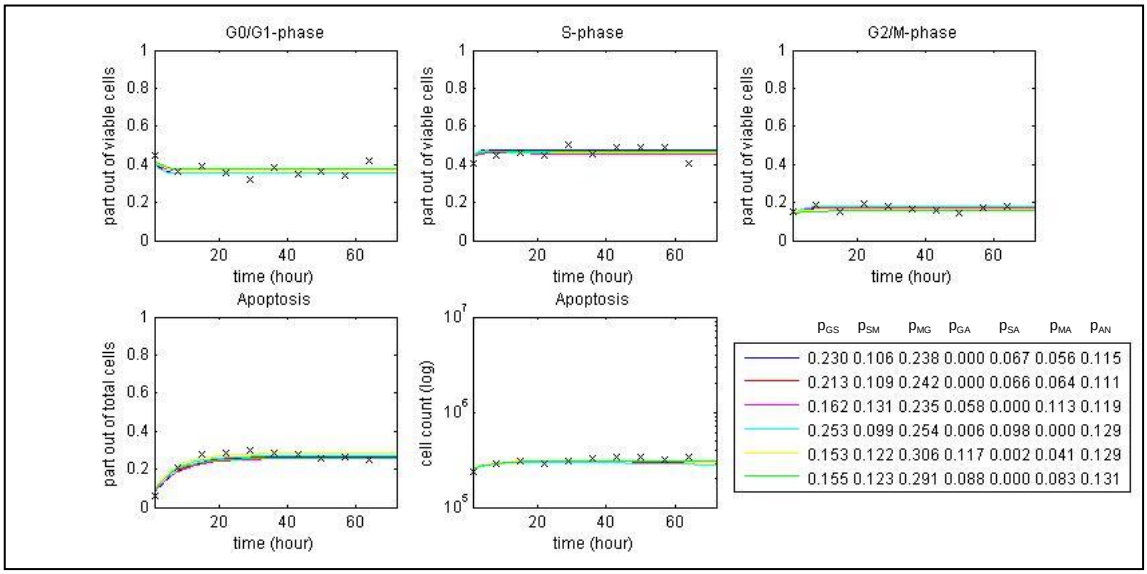


Figure 7 - The six different lines in the graphs represent six different model fits to experimental data (the black spots) where each model has its model parameters listed in the bottom left table. The experimental disturbance is 10 percent.

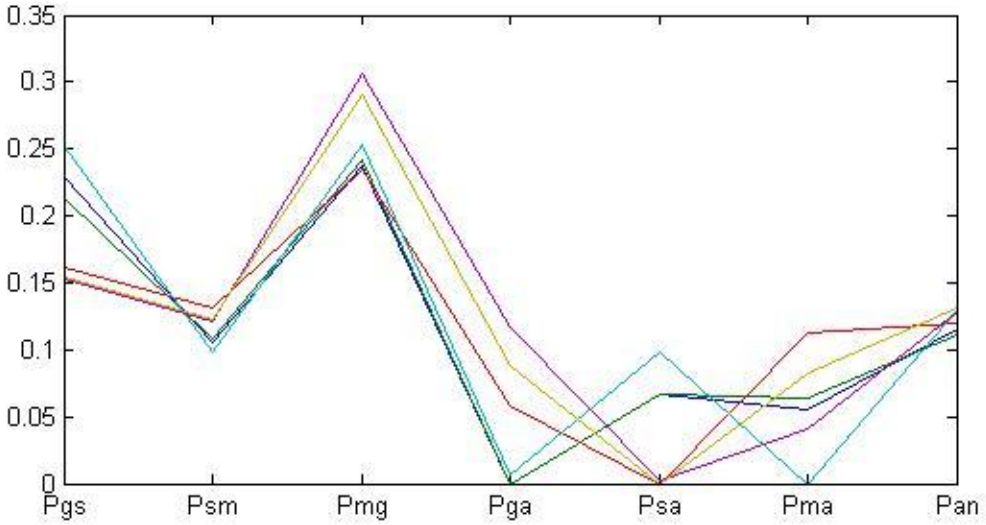


Figure 8 - The figure shows how the 6 good parameter estimates from figure 7 may vary. Parameter p_{ga}, p_{sa}, and p_{ma} shows great relative variance in between estimates which is a sign of bad estimability of these parameters.

2.4 Finding alternative drugs for treatment of resistant cancer cells

2.4.1 A general method finding alternative drugs

This part describes the general idea of how the cell cycle model can be used to find alternative drugs for treatment of resistant cancer cells.

Step 0: (figure 9) Estimate parameters for untreated cell line

Step 1: (figure 9) Estimate parameters for the cell line treated with each of the drugs separately. For each drug calculate the induced parameter change by subtracting the parameters of untreated cells from the parameters of drug treated cells. This will generate a fingerprint consisting of the parameter changes that each drug induces.

Step 2: (figure 10) Store the fingerprints in a database.

Step 3: Estimate model parameters for treated[4] resistance cancer cells. Then estimate the model parameters for treated sensitive cancer cells and calculate the parameter changes induced by the resistance mechanism by subtracting the parameters of treated resistant cells from the parameters of treated sensitive cells. Figure 11 shows an example where the difference between untreated resistant and treated sensitive cancer cells is calculated.

[4] Treated with the drug which resistance has been developed against.

The changes in parameters represent the effect an ideal drug should have on the resistant cancer cells in order to make them respond as sensitive treated cancer cells.

Step 4: Finally search the data base for an alternative drug which induces the desired changes calculated in step 3. Hopefully the database contains a drug which can generate the desired change (se figure 12).

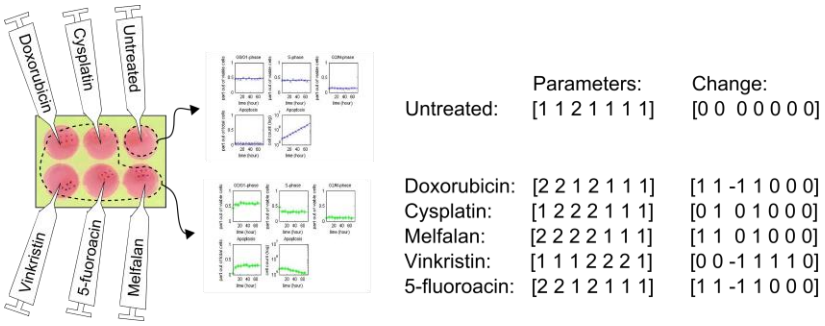


Figure 9 - (Step 1) Time series of the cell cycle phase progression of untreated and treated cancer cells are measured. For each drug the changes between the model parameters during treatment and the parameters of untreated cells are calculated.

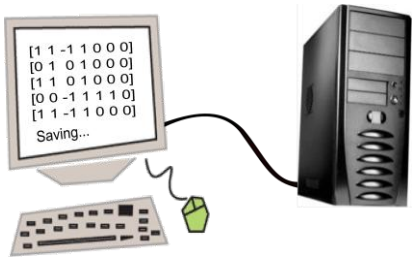


Figure 10 - (Step 2) The parameter changes calculated in step 1 is stored in a database.

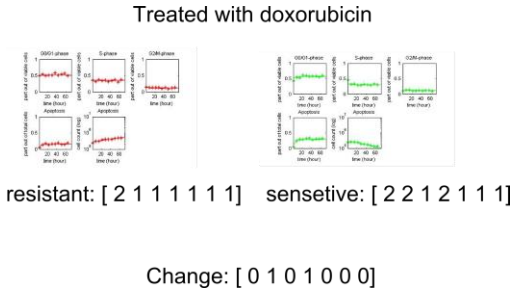


Figure 11 - (Step 3) Parameters are estimated based on cell cycle phase progression data obtained from treated resistant cancer cells and treated sensitive cancer cells respectively. Finally the parameter changes between resistance and treated sensitive cancer

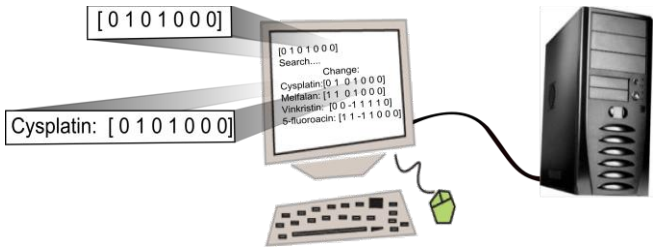


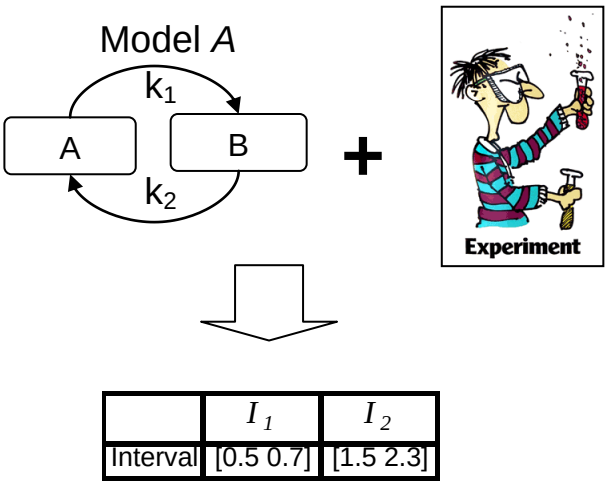
Figure 12 - (Step 4) The database is searched for an alternative drug which induces the change calculated in step 3 (see figure 8).

2.4.2 Finding alternative drugs using parameter intervals

Unfortunately, the basic idea described in 2.3.1 can not be directly implemented. This is due to the poor estimability of the parameters which the example in 2.2.2.2 illustrated. It is not possible to get the unique parameter estimates which are a requirement for success of the method. Therefore another way to represent the models and changes has to be found.

The models could be represented by parameters intervals where parameters pick within the intervals should generate acceptable time profiles that are similar to experimental observations. For example assume that the simplified model *A* in figure 13. Let $y(k_1,k_2)_t^{sim}$ be the simulated value at time point *t* with parameters k_1 and k_2 and let y_t^{obs} be the observed value at time point *t*. The intervals I_1 and I_2 are defined in such a way that equation (4) holds when choosing parameters k_1 and k_2 from within the intervals.

$$\sum_{n=1}^{N}\left(y(k_1,k_2)_t^{sim}-y_t^{obs}\right)^2 \leq cutoff \qquad (4)$$



	I_1	I_2
Interval	[0.5 0.7]	[1.5 2.3]

Figure 13 - From experimental measurements of treated or untreated resistant and sensitive cancer cells we can determine intervals of k_1 and k_2 such that equation (4) holds.

Change between parameter intervals (instead of change between unique parameters as in 2.4.1) could be calculated by subtracting the corresponding endpoints of two intervals. This will be the *interval fingerprint* representing the cell cycle model change induced by a drug. It is illustrated in figure 14 how intervals then can be used according to the basic idea described in 2.4.1. First is the change between untreated and treated cancer cells calculated by subtracting their corresponding interval endpoints (A). Secondly the preferred interval change of the resistant cancer cells is calculated (B) and then finally alternative drugs are searched for that induces the same interval change (C).

A

				Interval change	
Cell type	Treatment	k1	k2	k1	k2
Sensitive	none	[0.5 0.7]	[1.5 2.3]	[0]	[0]
Sensitive	Drug 1	[0.1 0.5]	[1 2]	[-0.4 -0.2]	[-0.5 -0.3]
Sensitive	Drug 2	[0.2 0.6]	[0.5 1]	[-0.2 -0.1]	[-1 -1.3]
Sensitive	Drug 3	[1 1.6]	[0.1 0.7]	[0.5 0.9]	[-1.4 -1.6]

Drug 2 induces a similar change.

B

C

				Interval change	
Cell type	Treatment	k1	k2	k1	k2
X-Resistant	Drug X	[0.4 0.6]	[1.6 2.4]	[0]	[0]
Sensitive	Drug X	[0.1 0.5]	[0.6 0.9]	[-0.3 -0.1]	[-1 -1.5]

Figure 14 - Reversed chemical control using parameter intervals. (A) First estimate the change between treated and untreated cancer cells by subtracting their corresponding end points. (B) Estimates the desired change of the resistant cancer cells. (C) Search for alternative drugs which induce the same change.

2.4.2.1 Problems with intervals

The interval approach has important limitations. The main problem is that the intervals are hard to define and compare. Intervals can be calculated by generating sets of parameters consistent with experimental data using for example fminsearch several times. The interval of each parameter could then be set to lowest and highest values in each calculated parameter set. The problem with this process is that it is time consuming but perhaps manageable at least for small models like the one considered here. Another fundamental problem is that correlation between parameters is ignored. It is not unlikely that only certain combinations of parameters from within the intervals are consistent with data. For example, suppose we have the two parameters k_1 and k_2 and their corresponding intervals I_1 and I_2 as defined in figure 13. Let us assume that the parameter combinations k_1=0.52 and k_2=1.73 and k_1=0.67 and k_2=2.25 produce time profiles consistent with observed data. Further more assume that the parameter combination k_1=0.52 and k_2=2.24 does not produce time profiles consistent with observed data. The reason to this is parameter correlation. It could actually be the case that there exist two clusters c_1 and c_2 as defined in table 1. Only if we draw k_1 and k_2 from intervals with in the same clusters do we get simulations consistent with data. The problem then becomes to perform reversed chemical control using clusters and not only intervals. The cluster approach is something not considered any further in this work but it could be fruitful to investigate it further.

	c_1	c_2
k_1	[0.5 0.55]	[0.65 0.7]
k_2	[1.7 1.8]	[2.2 .23]

Table 1 - Two different intervals c_1 and c_2 both from which parameters consistent with data can be drawn.

Moreover, it is difficult to compare interval fingerprints. For example; how the query interval in figure 15 with intervals I_1 and I_2 be compared? Maybe the query interval is more similar to I_2 since it is more in the center of that interval or maybe should the query interval be considered to be equal similar to both of them? The point is that it is tricky to set up criteria for how to compare intervals.

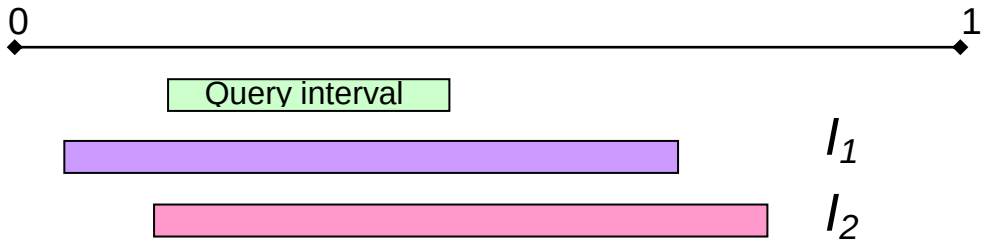


Figure 15 - It is difficult to compare intervals. Which one of I_1 or I_2 is the query interval most similar to?

It is also problematic when the parameters differ in estimability. For example imagine that we would have the case described by table 2 were the estimated interval of parameter k_1 is centralized around the true values and the estimated interval for parameter k_2 is relatively independent of the true values. Interval k_2 could then be considered as less informative since it does not give any hint to k_2's real value. The interval of parameter k_1, which seems more informative, would be the interval to rely upon when comparing different fingerprints. Thus it would be hard in a real situation to determine how much one should rely upon different parameters and therefore it would become a very subjective decision.

k_1	k_1 true	k_2	k_2 true
[0.5 0.7]	0.6	[0 2]	0.01
[0.01 0.05]	0.02	[0 2.1]	0.1
[2 3]	2.3	[0 1.8]	1
[1 1.6]	1.3	[0 1.9]	1.3

Table 2 - The k_1 interval is more informative than the k_2 interval. How should it be account for?

In the thesis work reported here the difficulties presented above were not studied any further. Instead of pursuing the interval approach an alternative strategy to obtain fingerprints based on Bayesian inference was investigated.

2.4.3 Finding alternative drugs using Bayesian probability theory

The idea of Bayesian probability fingerprints is here illustrated with the following example. Assume that we have the simplified cell cycle phase model *M* presented in

figure 16. The model has only two cell cycle phases A=G0/G1/S and B=G2/M and two flows k_1 and k_2 and no apoptosis flows.

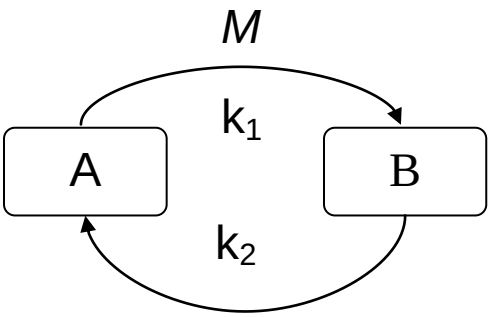


Figure 16 - The compartment model M with two state-variables A=G0/G1/S and B=G2/M and two state-flows k_1 and k_2.

Now, assume that it is only biological reasonable to have values of k_1 and k_2 within the intervals presented in table 4. Then it is reasonable to divide the each of the intervals into a number of subintervals. In table 4, they are divided in *two* subintervals *low* and *high*.

	Interval
k_1	**[0.01 1]**
k_2	**[0.1 10]**

Table 3 The biological reasonable intervals of k_1 and k_2.

	Subinterval	
	Low	**High**
k_1	**[0.01 0.1]**	**[0.1 1]**
k_2	**[0.1 1]**	**[1 10]**

Table 4 The intervals k1 and k_2 divided into subintervals *low* and *high*.

Four sub-models *M* - M_1, M_2, M_3 and M_4 (see figure 17) can now be defined based on the subinterval. Each model can be said to represent all observations where k_1 and k_2 are limited to a particular set of subintervals.

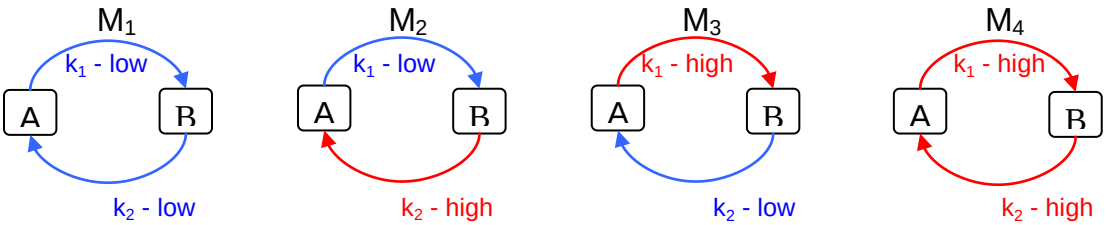


Figure 17 - Four sub-models defined from the subintervals of k_1 and k_2.

Suppose now that we can calculate for each of the four sub-models the probability that it explains experimental data y_{data}. The set of probabilities, $P(M_1),...,P(M_4)$ i=1,…,4 will then represent the sub-model *probability fingerprint* of the experimental data.

2.4.3.1 Probability assumption

In this work, it is assumed that the errors *errG*, *errS*, *errM* and *errA* in equation (2) are a normally distributed stochastic variable with a user defined variance σ^2.

It is important to point out that the noise assumption is a result of a compromise between the time available to investigate the matter and the practical usefulness of the assumption in application. It would be much more time consuming and complicated to calculate the probability fingerprints without this normality assumption. However if the fingerprints calculated using the assumption can prove useful when searching for alternative drugs it would be an important step forward.

2.4.3.2 Estimating sub-model probabilities

The probability $P(M_i|y)$ of each sub-model $M_1,...,M_I$ given an experimental time profile y from a cell population has to be determined in order to calculate the probability fingerprints. But how are these sub-model probabilities $P(M_i|y)$ i=,1…,N calculated? According to Bayes theorem the conditional probability $P(M_i|y)$ can be rewritten as;

$$P(M_i|y) = \frac{p(y|M_i)P(M_i)}{p(y)} \tag{5}$$

Equation (4) and the probability density function values $p(y)$ and $p(y|M_i)$ tell us that the probability $p(M_i|y)$ can indirectly be calculated through the quantities $P(M_i)$, $p(y)$ and $p(y|M_i)$. Since there is no prior knowledge of which model is more probable we have $P(M_1)=,\ldots,=P(M_N)=1/N$ where N is the number of models. Thus, the probabilities $P(M_i)$ $i=1,..N$ split the probability space considered into equally probable parts. The probability distribution function $p(y)$ according to the *law of total probability* (Blom 2001) is given by;

$$p(y) = \sum_{i=1}^{N} p(y|M_i)P(M_i) = \frac{1}{N}\sum_{i=1}^{N} p(y|M_i) \tag{6}$$

Equation (5) shows that the probability of receiving a measurement y is proportional to a sum of the probabilities $p(y|M_i)$ $i=1,\ldots,N$ weighted with $P(M_i)$. $P(y|M_i)$ is the likelihood function for the observation when assuming sub-model M_i. This is the challenging quantity to calculate in equation (4). It is the essentially the probability that the experimental data y has been generated by sub-model M_i. The function $p(y|M_i)$ can be estimated numerically given the probability assumption of the errors *errG*, *errS*, *errM* and *errA* in equation (2) from part 2.4.3.1. Under the assumption, $p(y|M_i)$can be calculated through expression (7) below which follows from basic probability theory.

$$p(y|M_i) = \int_{R_i} p(y|\theta, M_i)p(\theta|M_i)d\theta = E_\theta\left\{p(y|\theta, M_i)\ |M_i\right\} \tag{7}$$

Equation (7) tells us that $p(y|M_i)$ is equal to the integral of the product of $p(y|\theta,M_i)$ where we integrate over the parameter region R_i of sub-model M_i that defines all possible values of the parameter in model M_i. The parameter region R_i is defined by the parameter subintervals of sub-model *i*. Put differently equation (7) tells us that $p(y|M_i)$ is equal to the expected average value of the probability $p(y|\theta,M_i)$. We need an expression for $p(y|\theta,M_i)$ in order to evaluate expression (7). Under the assumption given in 2.4.3.1 $p(y|\theta,M_i)$ is defined as;

$$p(y|\theta_n) = \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\frac{1}{2\sigma^2}\|y-y(\theta_n)\|^2} \tag{8}$$

Here y represents the measured cell cycle phase experimental data and $y(\theta)$ denotes the model simulated cell cycle phase data. N is the number of sampling time points. In order to solve the integral for a sub-model M_i we thus need to evaluate;

$$\int_{R_i} \frac{1}{(2\pi)^{N/2}\sigma^N} e^{-\frac{1}{2\sigma^2}\|y-\hat{y}(\theta)\|^2} p(\theta|M_i)d\theta \tag{9}$$

Evaluation of expression (9) is challenging to solve mathematically due to the many parameters the integration is performed across. An approximation can be obtained by estimating the expected value $E_\theta\{p(y|\theta,M_i)|M_i\}$, equation (7). The expected value can be estimated by using the following Monte Carlo[5] algorithm;

1. Let n=1.
2. Randomly select a parameter vector θ_n from the sub-model parameter region R_i of M_i. A uniform distribution in the sub-model parameter region R_i of the parameters is assumed.
3. Use the cell cycle model simulator to calculate a trajectory $\hat{y}(\theta_n)$.
4. Calculate $p(y|\theta_n)$ using formula (8) where the standard deviation σ is a user defined variable that should reflect the variability in the expected measurement and model errors.
5. Let n=n+1.
6. Repeat 2-5 until the sum $S_{M_i} = \frac{1}{N}\sum_{n=1}^{N} p(y|\theta_n) \approx E_\theta\{p(y|\theta,M_i)|M_i\}$ converges.

Remember from above that we had $P(M_1)=,\ldots,=P(M_N)=1/N$ where N is the number of models. This fact together with equation (6) makes it possible to rewrite equation (5) as;

[5] For more information about the Monte Carlo method go to http://en.wikipedia.org/wiki/Monte_Carlo_method

$$P(M_i|y) = \frac{p(y|M_i)P(M_i)}{\sum_{i=1}^{I} p(y|M_i)P(M_i)} = \frac{p(y|M_i)}{\sum_{i=1}^{I} p(y|M_i)} \approx \frac{S_{M_i}}{\sum_{i=1}^{I} S_{M_i}} \quad (10)$$

The probability p*(Mi|y)* can now be calculated through expression (10). Once again, consider the example the model in figure 16. Using (10) we can numerically approximate *p(M1|y), p(M2|y), p(M3|y)* and *p(M4|y)*. In the next part it is shown how the calculated probabilities can be used to find alternative drugs for treatment of resistance cancer cells.

2.4.3.3 Probability matrices

Let y_1 be experimental data from untreated sensitive cancer cells and let y_2 be experimental data from the same sensitive cancer cells treated with drug *d*. We can then calculate the probabilities $p(M_1|y_1)$, $p(M_2|y_1)$, $p(M_3|y_1)$, $p(M_4|y_1)$ andd $p(M_1|y_2)$, $p(M_2|y_2)$, $p(M_3|y_2)$, $p(M_4|y_2)$ through formula (10). The effect of the drug will then be characterized by a matrix containing all possible sub-model probability changes between untreated and treaded cancer cells. For example, the probability that sub-model M_i explained the untreated data y_1 and that sub-model M_j explained data y_2 can be calculated using equation (11).

$$P(M_i \to M_j|y_1 \to y_2) \equiv P(M_i, M_j|y_1, y_2) = p(M_i|M_j, y_1, y_2)P(M_j|y_1, y_2) = \\ = P(M_i|y_1)P(M_j|y_2) \quad (11)$$

The last evaluation in equation (11) is possible since model M_i is independent of M_j and y_2 and model M_j is independent of y_1. Equation (11) says that the probability for the event of going from sub-model M_i to sub-model M_j, given that the experimental data has change from untreated experimental data y_1 to drug treated experimental data y_2, is equal to the product of probability of the separate events $p(M_i|y_1)$ and $p(M_j|y_2)$. Thus it is assumed that the event having model M_i when observing y_1 and the event having model M_j when observing y_2 are independent of each other. The probability changes calculated from (11) for M_i $i=1,\ldots,N$ and M_j $j=1,\ldots,N$ where N is the number of models will be defined as the *probability fingerprint* which is the probabilities of all possible sub-model changes that a drug induces.

As an example, imagine that we have got the probabilities presented in table 5.

Cell type	Treatment	P(M1|y)	P(M2|y)	P(M3|y)	P(M4|y)
Sensitive	**none**	**0.10**	**0.30**	**0.10**	**0.50**
Sensitive	**Drug 1**	**0.23**	**0.40**	**0.10**	**0.17**

Table 5 - The probability results calculated from experimental data of untreated sensitive cells and sensitive cells treated with drug 1.

A matrix *P* with the probability jumps can be calculated where for example p_{12}= $P(M_1,M_2|y_{untreated},y_{drug1})$=0.10*0.40=0.04 represents the probability that experimental data

first were explained by model M_1 and then after drug treatment were explained by model M_2. This generates the *probability matrix* of model transitions and is illustrated in table 6.

p_{11}	p_{21}	p_{31}	p_{41}
p_{12}	p_{22}	p_{32}	p_{42}
p_{13}	p_{23}	p_{33}	p_{43}
p_{14}	p_{24}	p_{34}	p_{44}

$p_{12}=P(M_1M_2|y_{untreated}y_{drug1})$

Table 6 The probability matrix P where p_{ij} describes the probability $p(M_i \rightarrow M_j | y_{untreated} \rightarrow y_{drug1})$.

2.4.3.4 Discrete subinterval change

Assume that the two compartment model from section 2.4.3 has three subintervals low, medium and high for each parameter. Let model M_i represent the subinterval combination [Medium Low], model M_j [Medium Medium], model M_k [High Medium] and model M_l [High High]. Going from M_i to M_j can be represented by the vector [0 +1] meaning that parameter 2 has changed plus one subinterval from low to medium. Similar going from M_k to M_l can also be represented as the change [0 +1] were parameter 2 has changed from medium to high. By adding the probabilities $P(M_i \rightarrow M_j)$ and $P(M_k \rightarrow M_l)$ the probability of the change [0 +1] can be calculated. Conversely going from M_j to M_i and from M_k to M_l will be interpreted as the change [0 -1] and by adding the probabilities $P(M_j \rightarrow M_i)$ and $P(M_k \rightarrow M_l)$ the probability for the change [0 -1] is calculated. Generally let c_k be a vector $1xN$ where N is the number of parameters in the model defining a particular subinterval change k and assume that there are K possible unique changes. For example if the parameters can take 3 values there are 5 possible changes; -2, -1, 0, 1, and 2. If there are two parameters then there are $5^2=25$ unique changes c_k. The probability $P(c_k | y_1 \rightarrow y_2)$ of a parameter subinterval change c_k can be calculated in the following way.

$$P(c_k | y_1 \rightarrow y_2) = \sum_{i,j \in S} P(M_i \rightarrow M_j | y_1 \rightarrow y_2) \qquad (12)$$

Where S is the set with index pairs i,j representing model changes $M_i \rightarrow M_j$ which stand for same discrete subinterval change. $P(c_k | y_1 \rightarrow y_2)$ can then be calculated for all $k=1,\ldots,K$. This will results in a k-dimensional vector ρ with probabilities for all the possible subinterval changes.

2.4.3.5 Probability of alternative drugs

Assume that there are only three alternative drugs a_1, a_2 and a_3. A sensitive cell line is treated with each drug separately and a time series is collected from each treatment. For each drug a_i, a probability vector ρ^{ai} can be calculated which contains the probabilities $P(c_k | y_0 \rightarrow y_i)$ $k=1,\ldots,K$ were y_0 is trajectory from untreated sensitive cells and y_i is trajectory from drug a_i treated sensitive cells. Moreover assume that a resistant cell line has evolved from the sensitive cell. Assume that the cell line is resistant to a drug x. It is desirable to identify drugs which have the highest probability to reverse the resistance.

This can be calculated by first collecting time series from sensitive and resistant cell line treated with drug *x*. Assume that these trajectories can be denoted y_r and y_s respectively. the collected trajectories of sensitive cells can then be obtained. Then the corresponding probability vector ρ^* for $y_r = y_0$ and $y_s = y_i$ can be calculated. The probability that alternative drug a_i reverses the resistance can then be calculated in the following way.

P(ai orsakar önskad förändring yr –ys| observerad förän

$$P(a_i | y_r \rightarrow y_s) = \sum_{k=1}^{K} P(a_i | c_k) P(c_k | y_r \rightarrow y_s) \quad (13)$$

The probability $P(c_k | y_r \rightarrow y_s)$ equals ρ^*_k. $P(a_i|c_k)$ can be rewritten using Bayes theorem;.

$$P(a_i | c_k) = \frac{P(c_k | a_i) P(a_i)}{\sum_{i=1}^{I} P(c_k | a_i) P(a_i)} \quad (14)$$

The drug probabilities $P(a_i)$ are assumed to be equal for all drugs. This means that $P(a_i)$ will be canceled out in equation (14). The probability $P(c_k|a_i)$ equals $\rho^{ai}{}_k$. It is then straightforward to calculated $P(a_i|c_k)$.

2.4.3.6 Alternative drug search – summing up

Below is a stepwise description of the method used to find alternative drugs for treatment of resistance cancer cells.

Step 1: Calculate the probability matrix for each drug *i* which represents the drug induced change in sensitive cancer cells. Then calculate the probability vectors ρ^{ai} $i=1,\ldots,I$ with probabilities for the discrete subinterval changes using (12).

Step 2: Store the results in a database.

Step3: Calculate the probability matrix for the change between treated resistant cells and treated sensitive cells which represents the change an ideal drug induce in the resistant cancer cells. Then calculate the probability vector ρ^* with probabilities for discrete subinterval changes.

Step 4: Rank the alternative drugs accordingly to their probability $P(a_i | y_r \rightarrow y_s)$ of reversing the resistance. The one or those with the highest probability will then be considered as the best alternative drug.

3 In silico experimental setup

3.1 Experiments

In the simulated experiments performed the state-variables G0/G1, M, G2/S and A are measured at 10 uniformly spread time points over a period of 72 hours. The cell cycle time is measured if it turns out to be shorter than 72 hours. Cell cycle trajectories were generated by sampling from simulation and then adding 10 percent measurement disturbance in order to reflect a realistic case. Notably 10 percent measurement disturbance could be an underestimation. However the results from Panetta et al indicated that 10 percent measurement disturbance is a realistic assumption (Panetta et al 2006).

In a real wet experiment the cell cycle progression over time can be measured for example by Arrayscan™ or flow cytometry. Flow cytometry was the method used by Panetta (Panetta et al 2006) to collect data.

3.2 In silico sensitive cell line and substances

3.2.1 Cell line

One sensitive in silico cell line was created. The cell line is represented by the model parameters shown in table 9. Panetta's (Panetta 2006) estimated cell lines parameter was used as a guidance to set realistic parameter values. Parameter p_{GS}, p_{SM} and p_{MG} are cell cycle parameters representing the cell transition rate between cell cycles. The rate between G (G0/G1) and S phase is of the same magnitude as the rate between S and M (G2/M) phase. The mitosis rate between M and G phase is 3 times faster. As a consequence the cancer cells generally stay shorter time in M phase than in G and S phase. Parameter p_{GA}, p_{SA} and p_{MA} are apoptosis parameter representing the apoptosis rate from respective G, S and M phase. The choice of apoptosis parameters should depend upon pre-known facts about the cell. Here no pre-known facts were used hence the apoptosis flow was considered to be of equal magnitude. Models where one or two apoptosis parameters play greater rolls could also be created. Parameter p_{AN} is the flow with which apoptosis cells becomes non-viable. The simulated cell cycle response of the cell line is presented in figure 18. The figure shows that part of cells in each phase is stabilized on a constant level and that the cell growth is exponential. The start values of cells in each phase are in the beginning in dynamic steady state proportions.

p_{GS}	p_{SM}	p_{MG}	p_{GA}	p_{SA}	p_{MA}	p_{AN}
0,15	0,12	0,33	0,011	0,009	0,012	0,15

Table 7 - The table shows the parameter values of the resistant cell lines.

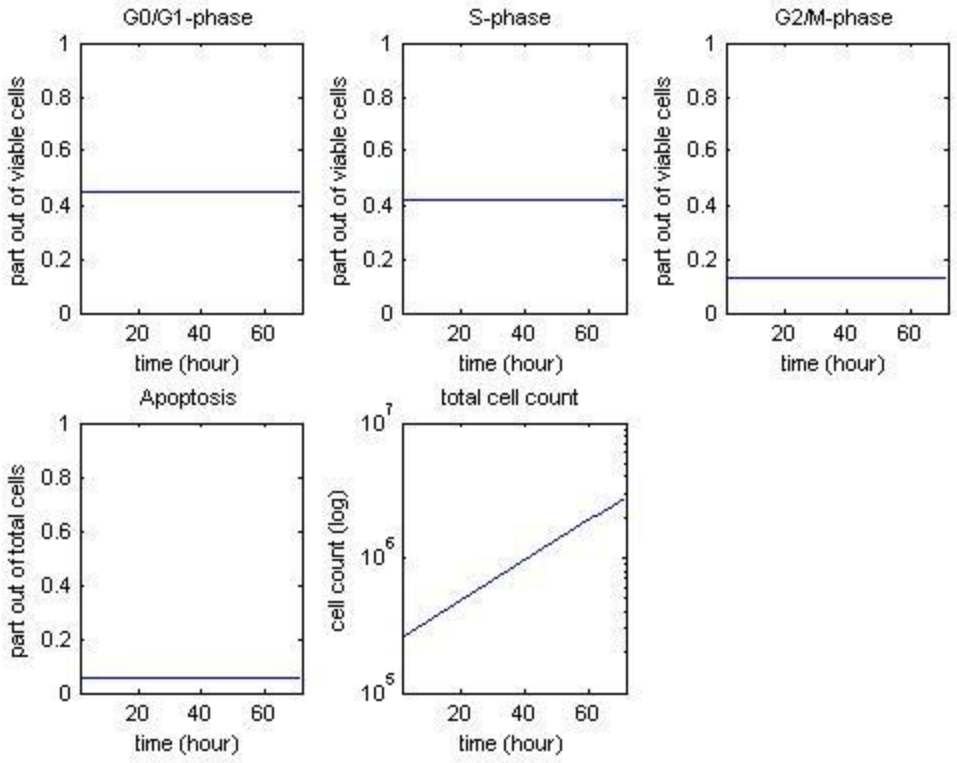


Figure 18 - The three top graphs show the relative distribution of cell line cells in G (G0/G1), S and M (G2/M) phases over time of. The bottom left graph show the fraction of cells in apoptosis over time out of total cell count of cell line 1 and the bottom right graph shows the total cell count.

3.2.2 Substances

20 different in silico substances were created. The effect of one substance was modeled as a multiplication of the substances parameters with the corresponding cell line parameters. The parameter changes for each of the substances are presented in table 8. The substances can be divided into three categories based upon their effect on cell growth. Substances 1-3, 7-8 and 13-15 decrease the cell growth in such magnitudes that there are fewer cells after 72 hours incubation than in the beginning (A the figure 19). Substances 4-6, 10-12 16-18 also decrease the cell growth of cell line 1 but there are still more cells after 72 hours incubation that in the beginning (B in figure 19). Finally substances 19-20 increase the cell growth (C in figure 19).

	p_{GS}	p_{SM}	p_{MG}	p_{GA}	p_{SA}	p_{MA}	p_{AN}
1	1	1	1	15	1	1	1
2	1	1	1	1	15	1	1
3	1	1	1	1	1	25	1
4	1	1	1	4	1	1	1
5	1	1	1	1	4	1	1
6	1	1	1	1	1	5	1
7	0,01	1	1	1	1	1	1
8	1	0,01	1	1	1	1	1
9	1	1	0,01	1	1	1	1
10	0,1	1	1	1	1	1	1
11	1	0,1	1	1	1	1	1
12	1	1	0,1	1	1	1	1
13	0,1	1	1	4	1	1	1
14	1	0,1	1	1	4	1	1
15	1	1	0,1	1	1	5	1
16	0,1	5	1	1	1	1	1
17	1	0,1	5	1	1	1	1
18	1	5	0,1	1	1	1	1
19	1	5	1	1	1	1	1
20	1	1	5	1	1	1	1

Table 8 - The table shows the parameter effect of each substances one the cell line parameters.

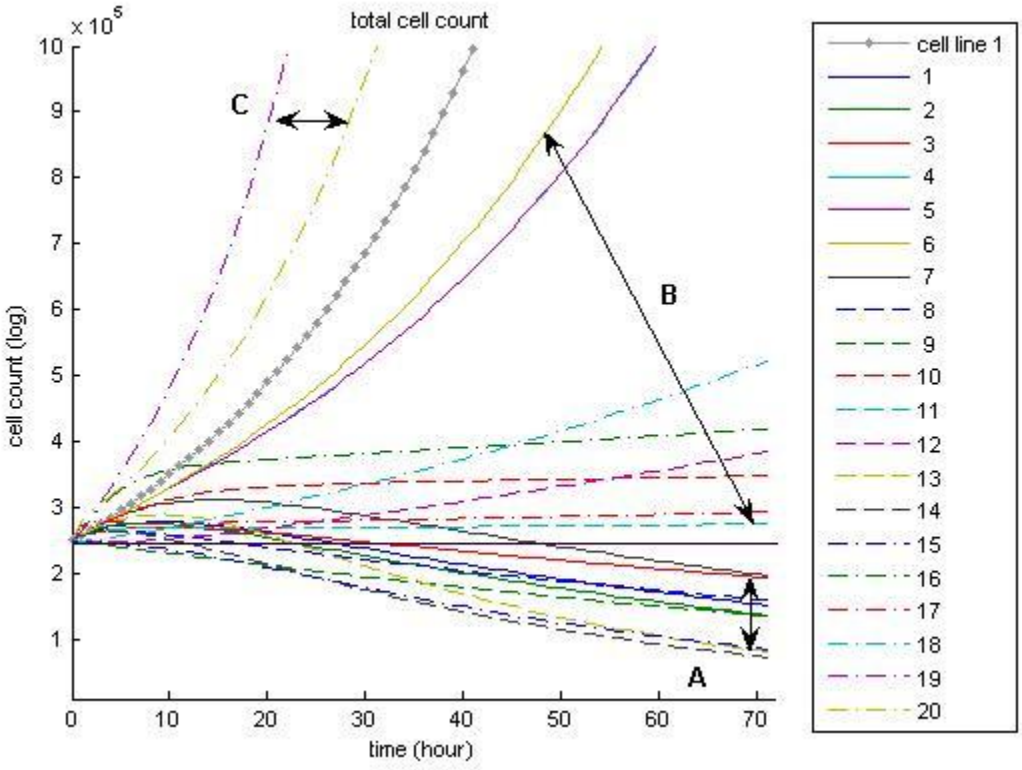


Figure 19 - Arrow (A) points to substances 1-3, 7-8 and 13-15 which decrease the cell growth in such magnitude that there are fewer cells after 72 hours incubation than in the beginning. Arrow (B) points to substances 4-6, 10-12 16-18 which also decrease the cell growth of cell line 1 (gray dotted line) but there are still more cells after 72 hours incubation that in the beginning. Finally arrow (C) points to substances 19-20 which increase the cell growth. Due to lack of resolution graph of substance 4 can not be spotted. It lies extremely close to the graph of substance 5.

4 Results

4.1 Setup

4.1.1 Subintervals

The parameters p_{GS}, p_{SM}, p_{MG}, p_{GA}, p_{SA}, p_{MA} and p_{AN} were divided into intervals (table 9). The endpoints of the intervals were thought to represent extreme scenarios. The low endpoints were considered to represent very slow flows between state-variables and the high endpoints were considered to represent very fast flows. Values outside these intervals were seen as biologically unrealistic values. It was assumed that the time for a cell to go from cell cycle phases G to S and S to M was at minimum 1 hour/cell and at maximum 20000 hours/cell. The minimum time for a cell to go from M to G was assumed to be 0.5 hour/cell and the maximum time was assumed to be 13333 hours/cell. The minimum time for a cell to go into apoptosis was assumed to be 20 hours/cell and the maximum time was assumed to be 20000 hours/cell. Finally the minimum time for a cell to go from A to N was assumed to be 5 hour/cell and the maximum time 10 hour/cell. From these assumptions were intervals defined for each of the parameters

The intervals of parameter p_{GS}, p_{SM}, p_{MG}, p_{GA}, p_{SA} and p_{MA} were further divided into three subintervals (table 10). The subintervals were created based knowledge of the estimability of the sensitive cell lines model parameters. Certain parameter combinations of p_{GS}, p_{MS} and p_{MG} drawn from within the mid subintervals together with certain parameter combinations of p_{GA}, p_{SA} and p_{MA} drawn from the low and mid interval generate models consistent with data. Figure 7 and 8 of part 2.3.2.2 shows this. In figure 8 six good model fits are plotted. It can be seen in figure 7 how the parameters varies between the fits. For each cell cycle parameter a single interval including both the highest and lowest value of each parameter is defined. The apoptosis parameters show a greater relative variability and therefore their values were split into two intervals. Here, another approach was implemented. It can be observed in figure 8 that the each parameter either are close to zero or in the interval [0.005 0.015]. For example p_{GA} and p_{MG} is in one fit high but p_{SM} is low. The low-subinterval was defined as going from 0.0005 to 0.005 and the mid-subinterval was defined as going from 0.005 to 0.02. The variance σ^2 in formula (8) was set to equal one when calculating the sub-model probabilities using formula (11).

The only parameter not divided into any subintervals was parameter p_{AN}. This eliminated the possibility to separate drugs based on how they parameter p_{AN}. The benefit of doing so is a three fold reduction in the number of sub-models to evaluate which leads to a three fold reduction of the computational time. The evaluation of the 3^6=729 sub-models of each substance took approximately three to four hours using four parallel working 2.9 gigahertz Intel Pentium processors. Consequently the complete calculation for all substances plus the cell line was 3*21=63 hours, approximately two and a half day.

	Interval
p_{GS}	[0.0005 1]
p_{SM}	[0.0005 1]
p_{MG}	[0.00075 2]
p_{GA}	[0.0005 0.5]
p_{SA}	[0.0005 0.5]
p_{MA}	[0.0005 0.5]
p_{AN}	[0.1 0.2]

Table 9 - The parameter intervals that are considered to be biological reasonable for each parameter.

	Subinterval		
	Low	Medium	High
p_{GS}	[0.0005 0.1]	[0.1 0.25]	[0.25 1]
p_{SM}	[0.0005 0.1]	[0.1 0.25]	[0.25 1]
p_{MG}	[0.0008 0.2]	[0.2 0.4]	[0.4 2]
p_{GA}	[0.0005 0.005]	[0.005 0.02]	[0.02 0.5]
p_{SA}	[0.0005 0.005]	[0.005 0.02]	[0.02 0.5]
p_{MA}	[0.0005 0.005]	[0.005 0.02]	[0.02 0.5]
p_{AN}	[0.1 0.2]		

Table 10 - The subinterval for the parameters of the cell cycle phase model.

4.2 Example 1

Substance 1 (marked by green in table 11) is considered to be the drug *x* which resistance has been developed against. The parameters of drug *x* treated sensitive cells therefore equal to those of the untreated sensitive cell line but with parameter p_{GA} increased 15 fold. The drug works by triggering the apoptosis when the cell is in G-phase. The cell cycle response of treated resistant cells is represented by substance 4 (marked by red in table 11). It is assumed that the resistant cell line has developed a mechanism which makes it less sensitive to the change drug *x* induced change of parameter p_{GA}. It could be case that the resistant cell has developed pumps specialized in pumping out drug *x* from the cell. Parameters p_{GA} is in the resistant cell line only increased 4 fold by the drug *x* which is compared to the 15 fold increase it induces in the sensitive cell line. All the substances of table 8 were included in the database. This means that substances 1 and 4 also play the role of being substances which could be identified as alternative drugs.

The result of ranking the substances accordingly to the probability calculated as in part 2.4.3.5 is presented in table 12. Substance 1 received the highest probability close followed by substance 4. Drug 6, 5, 2 and 3 then comes closely behind. They are all within 75 percent of the probability of the top ranked substance. Then there is a jump down to substance 20 which is within 57 percent of the top ranked substance, almost half as probable as the top ranked drug. Then is it a relative big drop in probability to the next substance which only is within 35 percent of the top ranked substance. Thus, the method has singled out substances 1, 4, 6, 2, 3 and possible 20 as more or less equally potential substances which could be used to reverse the resistance. They have an accumulated probability of 12+12+11+11+9+7=61 percent. Considering the coarse models and limited data involved these substances should be considered to be equally good candidates. The induced cell cycle response for each substance on the resistant cell line has been plotted in figure 20. Drug 1, 2, 3, 5 and 6 all decreased the cell growth of the resistant cell line.

Drug 1 and 4 does it by increasing the apoptosis rate from phase G where drug *x* triggered apoptosis. Drug 2, 3, 5 and 6 increases the apoptosis flow from a different phase than drug *x* does. It shows that it is possible to identify alternative drugs which targets the apoptosis mechanism in another cell cycle phase than the one drug *x* acts in. Substance 20 increases the cell growth and would therefore be useless as an alternative drug. As the figure show, the top ranked drugs induce cell cycle responses in the resistant drug *x* treated cells which are similar to the response of drug *x* treated sensitive cells. The drugs found could for example be triggering alternative pathways which makes the resistant cell line sensitive to drug *x* again.

Summary, the method has singled out the useful drugs among a group of candidates. A useful drug has the quality to both reduce the cell growth of the resistant cells and change drug *x* treated resistant cells cell cycle pattern in such way that the pattern imitates that of treated drug *x* sensitive cells. A lesson learned is that substances triggering apoptosis within other phases than drug *x* could be used as alternative drugs. No credible ranking within the treatment effective drugs could be obtained. A reason could be the coarse nature of the cell cycle model, measurement noise and the limitation of available data on cell cycle time series.

	p_{GS}	p_{SM}	p_{MG}	p_{GA}	p_{SA}	p_{MA}	p_{AN}
1	1	1	1	15	1	1	1
2	1	1	1	1	15	1	1
3	1	1	1	1	1	25	1
4	1	1	1	4	1	1	1
5	1	1	1	1	4	1	1
6	1	1	1	1	1	5	1
7	0,01	1	1	1	1	1	1
8	1	0,01	1	1	1	1	1
9	1	1	0,01	1	1	1	1
10	0,1	1	1	1	1	1	1
11	1	0,1	1	1	1	1	1
12	1	1	0,1	1	1	1	1
13	0,1	1	1	4	1	1	1
14	1	0,1	1	1	4	1	1
15	1	1	0,1	1	1	5	1
16	0,1	5	1	1	1	1	1
17	1	0,1	5	1	1	1	1
18	1	5	0,1	1	1	1	1
19	1	5	1	1	1	1	1
20	1	1	5	1	1	1	1

Table 11 - The table shows the in silico created substances. The parameters perturbation of treated resistant cells marked by red and the parameters perturbation of treated sensitive marked by green.

prob	**drug id**
0.1221	1
0.1154	4
0.109	6
0.1074	5
0.0974	2
0.0949	3
0.0704	20
0.0429	12
0.0405	9
0.0355	15
0.0219	8
0.0213	11
0.0204	13
0.0192	14
0.0136	19
0.0129	17
0.0126	7
0.0115	10
0.0093	16
0.0079	18

Table 12 - The table shows the results from the alternative drugs search. The substances are ranked after the probability it has to change the resistant cell in the preferred way.

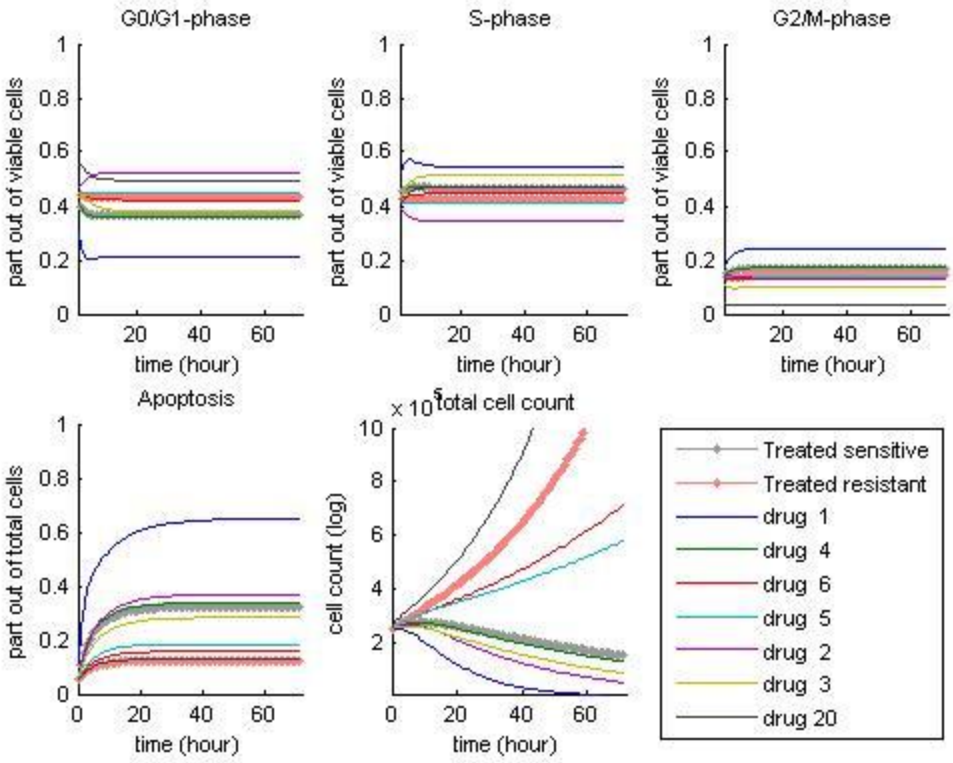


Figure 20 - The figure shows the plot of the substances listed in the legend in the right lower corner. It is the top ranked substances which the method has singled out.

4.3 Example 2

In this second example substance 13 (green in table 13) is considered to be the drug *xx* which resistance has been developed against. The parameters of treated sensitive cells are therefore equal to the cell line parameters of untreated sensitive cell line but with two changes where parameter p_{GS} has been decreased 10 fold and parameter p_{GA} has been increased 4 fold. Thus, in this care the drug of interest works by both lowering the transition rate between G and S phase and by triggering the apoptosis rate in G-phase. The cell cycle response of treated resistant cells is represented by substance 4 (red in table 13). It is assumed that the resistant cell line has developed a mechanism which makes it less sensitive to the change parameter p_{GS} induces (compare substance 4 and 13). It could be that the resistant cell has developed an inhibitor such that the drug no longer can lower the cell cycle rate between G and S. As in example 1 all the substances of table 8 are included into the database. This means that substances 13 and 4 also play the role of being possible drug candidates for treatment.

The result of ranking the substances is presented in table 14. Substance 13 received the highest probability. Substance 7 and substance 10 follows and are both within 70 percent of the top ranked substance. Then is it a relative big drop in probability to the next substance which is only within 31 percent of the top ranked substance. The method has singled out substances 13, 7 and 10 which have an accumulated probability of 26+19+18=63 percent. The induced cell cycle response of each of these three substances on the resistant cell line has been plotted in figure 21. The plot of cell growth in the figure shows that all three substances decrease the growth of the resistant cell line. Drug 13 decreases the transition rate between G and S phase and increases the apoptosis rate from G phase. The substance has the same effect as drug *xx*. It could work by effecting alternative pathway than drug *xx* which gives triggers the same cell cycle response as drug *xx* induces. Drug 7 and 10 both decreases the transition rate between G and S phase. Drug 7 does it 10 fold more than drug 10. Both drugs might work by triggering another pathway than drug *xx*. Commonly for all three substances is that they decreases the G to S transition rate. The result implies that such a property is crucial for an alternative drug to overcome drug *xx* resistance. As the figure show, the top ranked substances induces a cell cycle response of the resistant drug *xx* treated cells similar to the one of drug *xx* treated sensitive cells.

Summary, as in example 1 the method has singled out useful drugs among a group of candidates. From the example can be learned that is appears as a reduction of the transition rate between G and S phase is a crucial property for an alternative drug. No credible ranking within the treatment effective drugs could be obtained. A reason could be the coarse nature of the cell cycle model, measurement noise and the limitation of available data on cell cycle time series.

	p_{GS}	p_{SM}	p_{MG}	p_{GA}	p_{SA}	p_{MA}	p_{AN}
1	1	1	1	15	1	1	1
2	1	1	1	1	15	1	1
3	1	1	1	1	1	25	1
4	1	1	1	4	1	1	1
5	1	1	1	1	4	1	1
6	1	1	1	1	1	5	1
7	0,01	1	1	1	1	1	1
8	1	0,01	1	1	1	1	1
9	1	1	0,01	1	1	1	1
10	0,1	1	1	1	1	1	1
11	1	0,1	1	1	1	1	1
12	1	1	0,1	1	1	1	1
13	0,1	1	1	4	1	1	1
14	1	0,1	1	1	4	1	1
15	1	1	0,1	1	1	5	1
16	0,1	5	1	1	1	1	1
17	1	0,1	5	1	1	1	1
18	1	5	0,1	1	1	1	1
19	1	5	1	1	1	1	1
20	1	1	5	1	1	1	1

Table 13 The table shows the results from the alternative drugs search. In the right table are the parameters perturbation of treated resistant cells marked by red and the parameters perturbation of treated sensitive marked by green.

prob	**drug id**
0.2562	13
0.1888	7
0.1788	10
0.0822	17
0.0535	14
0.0515	20
0.0358	2
0.0259	3
0.025	16
0.0207	5
0.017	6
0.0153	4
0.0141	11
0.0131	1
0.0125	8
0.0044	19
0.0006	15
0	12
0	9
0	18

Table 14 - The table shows the results from the alternative drugs search. The substances are ranked the after probability to make the resistant cells sensitive again.

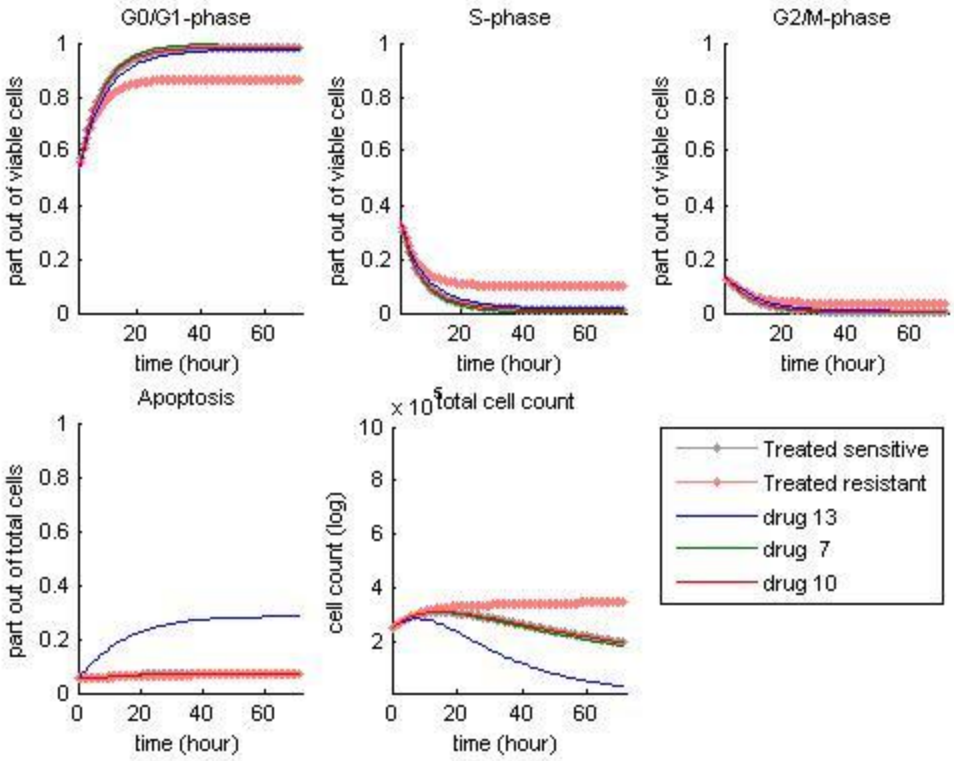


Figure 21 - The figure shows the plot of the substances listed in the legend in the right lower corner. It is the top ranked substances which the method has singled out.

4.4 Summary of the results

Example 1 and 2 illustrated how the method presented in part 2.3 works. The examples show that the method successfully can single out substance that has the potential to reverse resistance from a library with cell cycle phase data of different substances.

5 Discussion

5.1 Model relevance

The result presented in chapter 4 is promising. Given that the reality exhibits the same behavior as assumed in chapter 3 then is it know that the method using cell cycle model (1) performs well. But what if the model does not describe the reality sufficiently? This leads to bad performance of the method for alternative drug search. Here two important model assumptions are addressed.

5.1.1 Delay of maximum drug effect

The model used in this work does not capture the event of drug effect time delay. It is assumed that the drug starts to change the cell cycle instantly at a constant rate during treatment. This is probably not what happens in reality. There is most likely a period before the drug reaches its maximum effect, a period under which the cancer cells absorbs and distributes the drug. The maximum effect is not reached until there is a thermodynamic equilibrium of the drug concentration in the cell. To what extent the time delay undermines the explanatory power of the model should depend on the length of the time delay. In the following, one possible extension of model (1) in which models the time delay is considered.

Time delay of the drug effect can be incorporated in model (1) by changing the drug target parameter/parameters as in (13). At t=0 is $p_{XX}=p_{XXuntreated}$ and at $t>>p_{XXdelay}$ is $p_{XX}\approx p_{XXtreated}$. The time delay parameter $p_{XXdelay}$ determines how fast the influence of the $p_{XXuntreated}$ parameter is decreased and consequently how fast the influence of $p_{XXuntreated}$ is increased. The function gives rise to an hyperbole-formed curved. For example with $p_{XXdelay}=10$, $p_{XXuntreated}=0.05$, $p_{XXtreated}=0.1$ and t=[0 100] the following graph is obtained (figure 22).

$$p_{GS} = \frac{p_{GSdelay} p_{GSuntreated} + t p_{GStreated}}{p_{GSdelay} + t} \text{ where } p_{GStreated} = p_{GSuntreated} + p_{GSDrugchange}$$

$$p_{SM} = \frac{p_{SMdelay} p_{SMuntreated} + t p_{SMtreated}}{p_{SMdelay} + t} \text{ where } p_{SMtreated} = p_{SMuntreated} + p_{SMDrugchange}$$

$$p_{MG} = \frac{p_{MGdelay} p_{MGuntreated} + t p_{MGtreated}}{p_{MGdelay} + t} \text{ where } p_{MGtreated} = p_{MGuntreated} + p_{MGDrugchange}$$

$$p_{GA} = \frac{p_{GAdelay} p_{GAuntreated} + t p_{GAtreated}}{p_{GAdelay} + t} \text{ where } p_{GAtreated} = p_{GAuntreated} + p_{GADrugchange}$$

$$p_{SA} = \frac{p_{SAdelay} p_{SAuntreated} + t p_{SAtreated}}{p_{SAdelay} + t} \text{ where } p_{SAtreated} = p_{SAuntreated} + p_{SADrugchange}$$

$$p_{MA} = \frac{p_{MAdelay} p_{MAuntreated} + t p_{MAtreated}}{p_{MAdelay} + t} \text{ where } p_{MAtreated} = p_{MAuntreated} + p_{MADrugchange}$$

$$p_{AN} = p_{AN} \tag{15}$$

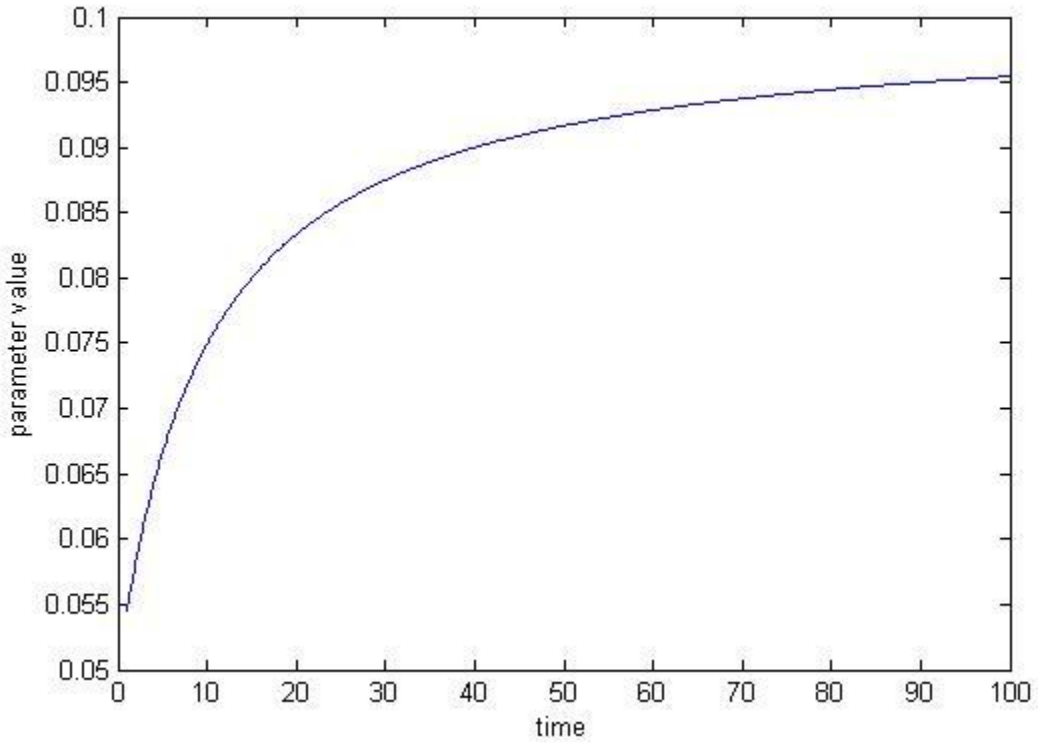


Figure 22 - The figure shows the function $p_{XX}(t)=(10*0.05+0.1*t)/(10+t)$. It illustrates how the function at the beginning assumes values 0.05 and then increases in 10 hours to 0.75. As the time goes to infinity the function equals 0.1.

The model has at minimum 6($p_{GSuntreated},\ldots,p_{MAuntreated}$)+2(new)+1($p_{AN}$) parameters and at maximum 6+12+1 depending upon the number of parameters affected by the drug. Six of the parameters ($p_{GSuntreated}$, $p_{SMuntreated}$, $p_{MGuntreated}$, $p_{GAuntreated}$, $p_{SAuntreated}$ and $p_{MAuntreated}$) should be known from previous experiments on untreated cancer cells. For each of the drug affected model parameters p_{XX}, two new parameters $p_{XXdelay}$ and $p_{XXdrugchange}$ are introduced. This results in 3-13 unknown parameters. When estimating the parameters of the extended model all 13 parameters have to be estimated since no pre-knowledge of what parameters the drug targets are known. For the parameters p_{XX} that are left unaffected by the drug we should have $p_{XXdrugchange}=0$ so that $p_{XXtreated}=p_{XXuntreated}$.

Now, if the time delay $p_{XXdelay}$ is small then after a short period of time $p_{XXdelay}$ will be much smaller than t and $p_{XXuntreated}p_{XXdelay}$ will be much smaller that $tp_{XXtreated}$. Thus the influence of the untreated parameters will become insignificant such that p_{XX} equals the constant values of $p_{XXtreated}$. Thus if the time delay is short then the time delay model (15) can be approximated by model (1). But as the time delay grows lager the more significant the time delay becomes. It then would be preferable to used model (15) to accurately capture reality. However a major problem with model (15) is that it has 13 unknown parameters instead of the 7 of model (1). This will lead to even more severe estimability problems.

Another approach would be to directly estimate the time delay of each drug parameters $p_{XXdelay}$ based on for example drug absorption studies and/or studies of the penetrability of the drug active molecule/molecules. Such an approach would reduce the unknown parameters of the drug delay model (15) from 13 to the 7, the same number of parameters that model (1) has.

In summary; no model of the time delay is used in this work. A proposal for future studies would be to further investigate the potential of such an extension.

5.1.2 Error model

In the current approach presented in this thesis, a simple Gaussian error model is assumed (part 2.4.3.1). This error model was chosen mainly for practical reasons and should be realistic as long as the deviation between the observed and simulated time profiles is small. Additional experimental and theoretical work is needed to validate this assumption.

Another error model could be used. Instead of Gaussian error model a simpler model could be implemented as follows. First define a cutoff value of the total error *e* (equation 3). For each sub-model, draw parameters from within the subintervals of the sub-model. Perform model simulations with the parameters and then calculate the total error *e*. If the total error is below the cut off value then add one to the sub-model score parameter *s* formula below (16). The procedure is repeated for each sub-model until the score parameter *s* has converged. The score parameter *s* then equals the probability that sub-model M_i is *e* from data. The index *m* runs over the number of times the total error *e* is below the cut of value and the index *n* runs over the number of repetitions.

$$s_{m+1} = \frac{s_m + 1}{n} \tag{16}$$

A score value s^{Mi} is obtained for all sub-models which is the score value for model M_i. The probability for sub-model M_i then is calculated as follows.

$$p(M_i|y) = \frac{s^{M_i}}{\sum_{i=1}^{I} s^{M_i}} \tag{17}$$

5.2 In vitro applicability

The in vitro applicability of the method developed in this thesis work depends upon a number of things.

I. Descriptive power of the model; how well the cell cycle model describes reality
II. Experimental accuracy, how accurately and how many samples of the cell cycle progression that has to be measured.
III. Financial burden; is cell cycle experiments financially realistic?

All these three points has to be considered when deciding whether to go further with in vitro studies. The work of Panetta's suggested that a non-time delay model can successfully be used. They also prove that experimental accuracy is achievable. The financial aspect of cell cycle experiments is promising since cell cycle experiments is much less expensive and time consuming than micro-array experiments.

5.3 Suggestion for future work

5.3.1 In silico substance library

The in silico substance library could be extended. For example, the library could be generates by an algorithm which perturbed the model parameters in a randomized way. Each parameter perturbation is saved and stored as an alternative substance.

5.3.2 Covering different alternative drugs search

Moreover, the results presented in chapter 4 shows that drugs in the current library can be used to find alternative drugs which can change *treated* resistant cancer cells to behave as *treated* sensitive cancer cells. Other types of changes can be calculated depending upon whether you have measurements of treated or untreated resistant cells and treated or untreated sensitive cells. Using measurements of untreated and treated resistant cells and of untreated and treated sensitive cells is it possible to calculate four different types of changes;

I. Between untreated resistant and untreated sensitive cancer cells
II. Between untreated resistant and treated sensitive cancer cells
III. Between treated[6] resistant and untreated sensitive cancer cells
IV. Between treated resistant and treated sensitive cancer cells

It makes sense to calculate change between untreated resistant and untreated sensitive cancer cells (I), between untreated resistant and treated sensitive cancer cells (II) and between treated resistant and treated sensitive cancer cells (IV) but not to calculate the change between treated resistant and untreated sensitive cancer cells (III). Searching for a drug which turns untreated resistant into treated sensitive cancer cells (II) will directly kill the resistant cancer cells and such a drug is probably an analog to the drug which resistance has been developed against. A drug which changes untreated resistant to sensitive untreated cancer cells (I) preferable inverts the resistant mechanism and hopefully makes the resistant cancer cells sensitive to treatment again. Although, one can not be sure that such a drug will work in combination with the drug resistance has been developed against. It could be the case that the alternative drug simply changes the resistant cancer cells without blocking the resistance mechanism. Finally a drug which changes treated resistant to treated sensitive (IV) is a drug which makes the resistant cell sensitive again. In a combination treatment together with the drug resistance have been developed against should such a drug work. Alternative drugs found from searching in any of the three ways I, II or IV either directly kill the resistant cancer cell (I) or kill them in a combination treatment with the drug resistance has been developed against. A drug which changes treated resistant to untreated sensitive cancer cells (III) will be useless since such drugs neither has the potential to alone kill the resistant cancer cells or to do so in combination with the drug resistance has been developed against.

The method developed in this paper can be used to find drugs which change the resistant cancer cells in any of the three ways I, II and IV. In this work the type III change is searched for, the same type of change that Lamb et al looked to perform reversed chemical control using their C-Map of mRNA expression data.(Lamb et al 2006) One can never be certain in reality if you are going to find drugs which induce the specific change searched for. It might therefore be preferred in an in vitro study to search for alternative drugs which can change the resistant cancer cells in either one of the three ways I, II and IV.

5.3.3 Subintervals

The subintervals in part 4.1.1 were defined based on estimability information of the sensitive cell lines parameters. The purpose of defining the subintervals in this way was to make it is easier to observe drug induced change in sensitive cells. If the mid-subintervals for the cell cycle parameters together with the low or mid-subintervals of the apoptosis parameters are the most probable sub-model describing sensitive cell line then a drug induced change decreased or increased in a parameter should be easiest to observe.

[6] By saying treated resistant cancer cell is it refereed to treatment with the drug which the resistance has been developed against.

It is preferable that the mid-subintervals are as close as possible to the parameter set consistent with collected time series of the sensitive cell line in order for the slightest parameters change to be observed. If the subintervals are badly defined such that induced parameter changes does not show as sub-model change. Estimability study of the sensitive cell line is therefore important. Defining the subintervals in a correct way is essential to how well the alternative drug search method perform.

It is preferable to define subintervals specific to the treated resistant cell line. Remember that changes between treated resistant and treated sensitive cancer cells are used in the alternative drug search. The mid-subintervals should be tuned to treated resistant cells in such a way that the slightest parameter change of the resistant cells parameters can be observed. Such tuning will maximize the likelihood of observing the change between resistant and sensitive cancer cells parameters. Consequently, different subintervals compared to those used when estimating the induced change substances has on the sensitive cell line should be used when identifying the change between treated resistant and treated sensitive cancer cells.

5.3.4 Drug classification

The collected time series of drug induced cell cycle change in the sensitive cell line can be used to classify drugs. The classification could reveal novel cell cycle behavioral connections between drugs. For example if drug induced mRNA expression data is known then different pathways or different expression patterns can be connected according to cell cycle behavior.

6 Conclusions

6.1 Results

Several things have been accomplished in this thesis work. In part 1.4 the goals of the thesis are listed. In conclusion all the goals have been reached. A summary is presented below.

6.1.1 Model generalization and simulation

I. Generalization and simulation of the cell cycle model developed by Panetta et al in such a way that it can be used to characterizes all known and unknown cell cycle specific chemotherapeutic drugs.

Without generalization and simulation of the cell cycle model the result of this thesis work would not have been possible to achieve. The used Matlab code is not included in this thesis. An experience Matlab programmer should easily be able to reproduce the results presented in this thesis work.

6.1.2 Cell lines and cells in silico

II. Development of artificial (in silico) cancer cell lines defined by parameters in the cell cycle model (chapter 3).

One in silico cell line has been developed in this thesis work (part 3.2.1). The work of Panetta et al has been used as guidance. In future work, several more cell lines can be created by variation of the cell cycle parameters and investigated.

III. Development of artificial drugs that reflects different perturbation of the cell cycle parameters (Chapter 3).

20 substances have been developed in silico (part 3.2.2). The substances effect the cell cycle in different ways by perturbation of the cell line parameters. It is possible to add numerous of substance to the library. Such library could be used to further in silico study the performance of the alternative drug search method.

IV. Development of “mutated” cell cycle models in which the model parameters have been perturbed in such a way that the mutated cell lines become resistant to the original treatment (Chapter 4).

In each of the two examples in chapter 4 (part 4.2 and 4.3) is a resistant cell line defined. The resistant cell line is considered to have mutated in such way that it is unaffected by a drug induced parameter/parameters change. In the examples is the cell cycle respond of treated resistant cancer cells only used. No parameters for an untreated resistant cell line are therefore defined in this work.

6.1.3 Alternative drug search method

V. Evaluate the possibility to perform estimation of the model parameters from time-series data to define unique “fingerprints” or “signatures” similar to those in C-Map database discussed earlier (Chapter 2).

It has been shown that estimation of the model parameters is problematic. The estimability of the model parameters are pore. No exact values of the model parameter can be obtained. Such behavior causes problems implementing the idea of C-Map using cell cycle parameters instead of gene expression signatures. A way to overcome the problem of representation was needed. Interval representation was considered. The approach seemed promising but was abandon due to numerous issues concerning the practical in an alternative drug search method.

VI. Develop an alternative approach to parameter estimation based on Bayesian model selection for defining useful “fingerprints” (Chapter 2).

In part 2.4.3 is a novel Bayesian model selection method described. It is shown how the method can be used to search for candidate drugs. Subintervals have to be defined. The range of each subinterval is a user defined property.

VII. In silico evaluation of the Bayesian fingerprint approach for combating drug resistant cancer cells (Chapter 4).

The two examples described in part 4.2 and 4.3 show how the Bayesian model selection method successfully can be used to find candidate drugs for treatment of resistant cancer cells. The method has the ability to single out treatment effective substances from a library with a majority of non effective substances. The treatment effective substances have the ability to change the cell cycle response of the resistant cell line such that it imitates the cell cycle response of the sensitive cell line again.

6.1.4 Summary

The results of this work demonstrate that it is feasible to find alterative drug which can combat chemotherapeutic resistance in cancer cells using the method developed in part 2.4.3. In part 1.3 is the guide lines for the development of a novel method used to battled drug resistance drawn. The following to points had to be accomplished in order to realize such a novel method.

I. Build a cell cycle model which can be used to characterize known and candidate cell cycle specific chemotherapeutic drugs.

II. Invent a method which uses the general cell cycle models accordingly to the C-Map approach of inverse chemical control in the search of drugs for treatment of resistant cancer cells.

In conclusion both points have been fulfilled. A novel method for battling cancer resistance been has been presented in this thesis work. A next step would be to apply the method to in vitro experimental material to further test the methods integrity.

6.2 Limitations

The applicability in reality of the method used to find alternative drugs which can combat chemotherapeutic drug resistance presented in this work depends upon how plausible the experimental in silico conditions defined in chapter 3 are.

Moreover, in this work maximum drug effect time delay is not included in the model. This could pose a problem. If the time delay is short enough then would it be sufficient top use model (1) but as the time delay growths then model (13) should be used to capture reality. The draw back with the extended model (13) is the large number of parameters. A suggestion in order to reduce the number of parameters would be to obtain estimates of the drug effect time delay of the parameters from studies on drug absorption and molecule/molecules penetrability.

The measurement noise model is not validated. The noise model should be compared to measurement noise obtained from in vivo experiments.

6.3 Future work

6.3.1 Further computational investigation

It would be would be a good idea to test the method using several other in silico setups. Such work would strengthen the credibility of the method.

Furthermore, the consequences of time delay should be investigate further. One could analyze the performance of the method using the extended model (13) and/or investigate to what extent estimates of the time delay obtained from studies of drug absorption and molecule/molecules penetrability can be used.

Develop a method used to define the parameter subintervals. Such method should include parameter estimability.

6.3.2 In vivo experiments

Planning and execution of in vivo experiments to complete a database containing cell cycle time series of drugs and substances. Work should be dedicated to decide what kind of drugs and substances to be included. Novel substances with unknown mechanism and effect preferable could be included in the database.

The method should be validated using drugs with known cell cycle specific mechanism for which the outcome can be guessed.

7 Bibliography

Alenzi Faris, Q. B., (2004) "Links between apoptosis, proliferation and the cell cycle" *British Journal of Biomedical Science* Number 2, August.

Gunnar, Blom., (2001). "*Sannolikhetsteori med tillämpningar"* sp. 35, Studentlitteratur, Lund (In Swedish)

Feng, Xiao-jian., Rabitz, Herschel., (2004) "Optimal Identification of Biochemical Reaction Networks" *Biophysical Journal"* Volume 86 1270-1281 March.

Goldberg, D. E., (1989) "*Genetic algorithms in search, optimization, and machine learning"* Reading Mass, Addison-Wesley.

Jacques, J., Grief, P., (1985) "Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design". *Mathematical Biosciences"* 77:201-227.

Jacques, J., Perry, T., (1990) "Parameter estimation: local identifiability of parameters" American journal of physiology. *Endocrinology and metabolism* vol. 21, n°4, pp. E727-E736 (31 ref.)

Mathworks (2007) http://www.mathworks.com/products/matlab/?BB=1 2007-03-13

Lamb, Justin., Crawford, E. D., Peck, David., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J-P., Subramanian, A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., Golub, T. R., (2006) "The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease" *Science* vol. 313 29 September 2006.

Ljung, L., (1987) "*System identification"* Prentice Hall.

Longley, D.B., Johnston, P.G., (2005) "Molecular mechanisms of drug resistance" *Journal of Pathology* 205:275-292.

Kozusko, F., Chen, P-H., Grant, S. G., Day B. W., Panetta, J. C., (2000) "A mathematical model of in vitro cancer cell growth and treatments with antimitotic agent curacin A" *Mathematical Biosciences* 170 1-6.

Panetta, J.C., Evans, W.E., Cheok, M.H., (2006) "Mechanistic mathematical modelling of mercaptopurine effects on cell cycle of human acute lymphoblastic leukaemia cells" *British Journal of Cancer* 94(1):93-100.

Pang, H. P., Dale, M. M., Ritter, J. M., (1999) *"Pharmacology"* fourth edition, Churchill Livingstone.

Weinberg, R. A., (2007) *"The Biology of Cancer"* chapter 8, Garland Science Taylor and Francis Group LLC.

8 Acknowledgements

Unit of Clinical Pharmacology at the Department of Medical Sciences at Uppsala University

Professor Rolf Larsson at the Unit of Clinical Pharmacology

Associate professor Mats Gustafssonat Dept. of Medical Sciences & Dept. of Engineering Sciences.

Magnus Åberg