

# Mathematical Models of Cell Cycle Progression: Applications to Breast Cancer Cell Lines

Kate T. Simms

*Thesis submitted for the degree of  
Doctor of Philosophy  
in  
Applied Mathematics  
at  
The University of Adelaide*

Discipline of Applied Mathematics, School of Mathematical Sciences,  
Faculty of Engineering, Computer and Mathematical Sciences



June 16, 2012

# Contents

|                                                            |           |
|------------------------------------------------------------|-----------|
| Abstract . . . . .                                         | 13        |
| <b>1 Introduction</b>                                      | <b>15</b> |
| 1.1 Thesis summary . . . . .                               | 16        |
| <b>2 Biological background</b>                             | <b>21</b> |
| 2.1 The cell structure . . . . .                           | 21        |
| 2.2 Cancer is the result of mutations in DNA . . . . .     | 25        |
| Causes of mutations . . . . .                              | 25        |
| 2.3 Cancers come in different stages . . . . .             | 26        |
| 2.4 The typical hallmarks of cancer . . . . .              | 29        |
| Hallmark 1: Self-sufficiency of growth signals . . . . .   | 29        |
| Hallmark 2: Insensitivity to anti-growth signals . . . . . | 30        |
| Hallmark 3: Evasion of cell death signals . . . . .        | 30        |
| Hallmark 4: Unlimited replicative potential . . . . .      | 31        |
| Hallmark 5: Sustained angiogenesis . . . . .               | 32        |
| Mathematical models of pre-angiogenic growth . . . . .     | 34        |

|                                                                       |           |
|-----------------------------------------------------------------------|-----------|
| Mathematical models of capillary growth during angiogenesis . . . . . | 36        |
| Hallmark 6: Invasion and metastasis . . . . .                         | 37        |
| An enabling hallmark: Increased rates of mutation . . . . .           | 38        |
| 2.5 Breast cancer . . . . .                                           | 42        |
| Structure of the breast . . . . .                                     | 43        |
| The role of estrogen and progesterone in the breast . . . . .         | 43        |
| Hormone replacement therapies and the development of cancer . . . . . | 45        |
| 2.6 Breast cancer cell lines - MCF-7 and T47D . . . . .               | 46        |
| 2.7 Cell proliferation and the cell cycle . . . . .                   | 48        |
| 2.8 Proteins which encourage proliferation . . . . .                  | 50        |
| 2.9 Commitment to cell division and the the R point . . . . .         | 52        |
| How the R point is regulated . . . . .                                | 53        |
| Inhibitors of cell cycle progression . . . . .                        | 54        |
| <b>3 Mathematical modelling of cell cycle progression</b>             | <b>56</b> |
| 3.1 Flow Cytometry: detecting cell cycle phases . . . . .             | 56        |
| Flow cytometry and fluorescent DNA detection . . . . .                | 57        |
| Flow cytometry and detecting cell death . . . . .                     | 60        |
| 3.2 Previous mathematical models of cell cycle progression . . . . .  | 62        |
| Protein models of cell cycle progression . . . . .                    | 63        |
| Age structured models . . . . .                                       | 65        |
| 3.3 Our model of cell cycle progression . . . . .                     | 68        |
| The inclusion of the storage phases . . . . .                         | 70        |

|                                                                                                                          |            |
|--------------------------------------------------------------------------------------------------------------------------|------------|
| Steady-state: growth in unchanging environmental conditions. . . . .                                                     | 75         |
| <b>4 The MCF-7 breast cancer cell line</b>                                                                               | <b>83</b>  |
| 4.1 Determining the duration of the storage phases . . . . .                                                             | 83         |
| Mitotic selection - an experimental procedure . . . . .                                                                  | 83         |
| A mitotic selection experiment using MCF-7 breast cancer cells . . . . .                                                 | 84         |
| Approximating the duration of the storage phases using the mitotic selection<br>experiment . . . . .                     | 87         |
| 4.2 Determining the remaining environment independent parameters . . . . .                                               | 90         |
| 4.3 Model analysis . . . . .                                                                                             | 91         |
| Average cell cycle duration and population doubling times . . . . .                                                      | 93         |
| Sensitivity analysis: How the experiment independent variables influence the model                                       | 95         |
| 4.4 The doubling time and the average cell cycle duration . . . . .                                                      | 99         |
| Other calculations of the average cell cycle duration from the literature . . . . .                                      | 102        |
| How the average duration in each of the phases is calculated in [28] . . . . .                                           | 103        |
| Discrete Simulation . . . . .                                                                                            | 105        |
| Proof that by counting only some of the average residence times, we arrive at the<br>formula presented in [28] . . . . . | 109        |
| 4.5 A novel representation of the model in equations (3.7) . . . . .                                                     | 111        |
| Summary of our analysis of equations (3.7) after parameterisation to the MCF-7<br>cell line . . . . .                    | 117        |
| <b>5 Applying the parameterised model to experiments involving the MCF-7 breast<br/>cancer cell line</b>                 | <b>119</b> |

|          |                                                                                                                      |            |
|----------|----------------------------------------------------------------------------------------------------------------------|------------|
| 5.1      | Applying our model to mitotically selected cells . . . . .                                                           | 119        |
|          | Modelling the pure cell population . . . . .                                                                         | 121        |
|          | Modelling the impure cell population . . . . .                                                                       | 121        |
|          | Results of optimisation . . . . .                                                                                    | 122        |
| 5.2      | ‘Slowly cycling population’? . . . . .                                                                               | 127        |
|          | The experiment-independent variables, and their influence over model behaviour<br>when not in steady-state . . . . . | 131        |
| 5.3      | Changing Environmental Conditions . . . . .                                                                          | 133        |
| 5.4      | Biological literature and the $\gamma$ function . . . . .                                                            | 135        |
| 5.5      | Fitting the parameters from equation (5.2) . . . . .                                                                 | 137        |
|          | Using experimentally determined cyclin D and myc . . . . .                                                           | 139        |
| 5.6      | Applying the cell cycle model to other experiments involving growth factors . . .                                    | 145        |
| 5.7      | Modelling the effects of anti-estrogens . . . . .                                                                    | 150        |
|          | Sensitivity analysis: solving the inverse problem for myc . . . . .                                                  | 155        |
| 5.8      | Application of the model to tamoxifen exposed MCF-7 breast cancer cells . . . .                                      | 158        |
| 5.9      | Tamoxifen exposure to mitotically selected cells . . . . .                                                           | 163        |
|          | The cytotoxic effects of tamoxifen . . . . .                                                                         | 168        |
|          | Incorporation of cell death . . . . .                                                                                | 169        |
|          | The form of $\beta_D(t)$ after continuous exposure to $12\mu\text{M}$ of tamoxifen . . . . .                         | 173        |
|          | Summary of our results of modelling cell cycle progression in MCF-7 breast cancer<br>cells. . . . .                  | 177        |
| <b>6</b> | <b>The T47D breast cancer cell line</b>                                                                              | <b>179</b> |

|          |                                                                                                                  |            |
|----------|------------------------------------------------------------------------------------------------------------------|------------|
| 6.1      | A discussion of different experiments . . . . .                                                                  | 180        |
| 6.2      | A summary of experiments . . . . .                                                                               | 181        |
|          | Results from [137] . . . . .                                                                                     | 181        |
|          | Results from [19] . . . . .                                                                                      | 182        |
|          | Results from [102] . . . . .                                                                                     | 182        |
|          | Results from [111] . . . . .                                                                                     | 182        |
|          | Results from [80] . . . . .                                                                                      | 182        |
|          | Summary of these experiments . . . . .                                                                           | 183        |
|          | Correlation between experimental results . . . . .                                                               | 183        |
| 6.3      | The storage phase durations for the T47D cell line . . . . .                                                     | 185        |
| 6.4      | The $\gamma$ function for the T47D cell line . . . . .                                                           | 190        |
|          | Generating the cyclin D profile from data in [81] . . . . .                                                      | 191        |
| 6.5      | Applying the model to another cyclin D stimulation experiment . . . . .                                          | 195        |
|          | Summary . . . . .                                                                                                | 198        |
| <b>7</b> | <b>Progesterin effects on cell cycle progression</b>                                                             | <b>201</b> |
| 7.1      | Quantitative observations of protein changes in T47D breast cancer cells after exposure to progesterin . . . . . | 203        |
|          | Changes in the S phase proportion and protein concentrations after progesterin as observed in [84] . . . . .     | 203        |
|          | Attempting to simulate S phase changes using known changes in cyclin D . . . . .                                 | 207        |
|          | Changes in cyclin-CDK activity after progesterin exposure . . . . .                                              | 209        |
| 7.2      | Determining the molecular make-up of cyclin-CDK complexes using gel filtration chromatography . . . . .          | 211        |

|          |                                                                                                          |            |
|----------|----------------------------------------------------------------------------------------------------------|------------|
|          | Gel filtration chromatography and the elution profile . . . . .                                          | 212        |
|          | Downsides of gel filtration chromatography . . . . .                                                     | 214        |
| 7.3      | The elution profiles from [84] determined using gel filtration chromatography . .                        | 215        |
|          | Interpreting the elution profiles from [84] . . . . .                                                    | 217        |
|          | The cyclin E1 profile . . . . .                                                                          | 219        |
|          | The cyclin D profile . . . . .                                                                           | 220        |
|          | The p27 profile . . . . .                                                                                | 221        |
|          | The p21 profile . . . . .                                                                                | 221        |
| 7.4      | Quantitative analysis of the elution profiles from [84] . . . . .                                        | 222        |
|          | The modelled elution profile for p27 . . . . .                                                           | 222        |
|          | The modelled elution profile for cyclin E1 . . . . .                                                     | 226        |
|          | The modelled elution profile for p21 . . . . .                                                           | 229        |
|          | The modelled elution profile for cyclin D . . . . .                                                      | 231        |
|          | Summary . . . . .                                                                                        | 234        |
| <b>8</b> | <b>A model of progestin effects on cell cycle proteins</b>                                               | <b>235</b> |
|          | Previous mathematical models of protein concentration changes during cell cycle<br>progression . . . . . | 237        |
| 8.1      | A simple model of cyclin E-CDK2 and p27 . . . . .                                                        | 238        |
|          | Steady-states . . . . .                                                                                  | 241        |
|          | Stability analysis . . . . .                                                                             | 241        |
| 8.2      | An interesting relationship between cyclin E-CDK2 and p27 . . . . .                                      | 244        |
|          | Phosphorylation at T187 . . . . .                                                                        | 245        |

|     |                                                                                                        |     |
|-----|--------------------------------------------------------------------------------------------------------|-----|
|     | Mechanism proposed by [106] in 1997 . . . . .                                                          | 245 |
|     | Mechanism proposed by [74] in 1999 . . . . .                                                           | 245 |
|     | Mechanism proposed by [140] in 1999 . . . . .                                                          | 246 |
|     | Previous models of p27 and cyclin E in the literature . . . . .                                        | 246 |
| 8.3 | Including the mechanism of T187 phosphorylation in our model of p27 and cyclin E247                    |     |
|     | Nondimensionalisation . . . . .                                                                        | 248 |
|     | Steady-state for the model in equation (8.6) . . . . .                                                 | 249 |
|     | Determining the rate parameters and starting conditions of our model . . . . .                         | 250 |
|     | Stability analysis . . . . .                                                                           | 252 |
|     | Determining the drivers of our system, $D_{act}(t)$ , $G_c(t)$ and $S(t)$ . . . . .                    | 255 |
| 8.4 | Modelling the experiment from [84] . . . . .                                                           | 257 |
|     | Model results and final parameterisation of $\beta_{pe}$ and $\beta_{ph}$ . . . . .                    | 258 |
|     | Steady-state analysis of our model . . . . .                                                           | 261 |
|     | The steady-state value of $P + P_E$ in terms of $E$ . . . . .                                          | 262 |
|     | The steady-state value of $E + P_E$ in terms of $E$ . . . . .                                          | 263 |
| 8.5 | A discussion of experimental results from [37] . . . . .                                               | 267 |
| 8.6 | Expanding our model - introducing cyclin D . . . . .                                                   | 270 |
|     | Introducing the mechanism of reduction in cyclin D-CDK4 activity after progestin<br>exposure . . . . . | 272 |
|     | Combining the model of cyclin D and p27 with the model of cyclin E-CDK2 and<br>p27 . . . . .           | 274 |
|     | Determining the rate variables for the model in equation (8.18) . . . . .                              | 275 |

|          |                                                                                                            |            |
|----------|------------------------------------------------------------------------------------------------------------|------------|
| 8.7      | Combining the protein model from equations (8.18) with the cell cycle model from equations (3.7) . . . . . | 278        |
|          | Results of our combined model . . . . .                                                                    | 279        |
|          | Summary of our model results . . . . .                                                                     | 284        |
| 8.8      | The effects of a second dose of progestin, administered at 40 hours after the first dose . . . . .         | 285        |
|          | Simulating the second dose of progestin after 40 hours . . . . .                                           | 286        |
|          | Summary . . . . .                                                                                          | 292        |
| <b>9</b> | <b>Conclusion</b>                                                                                          | <b>295</b> |
| 9.1      | Thesis summary . . . . .                                                                                   | 295        |
| 9.2      | Current avenues being explored for future work . . . . .                                                   | 297        |
| 9.3      | Longer term views for future work . . . . .                                                                | 298        |
|          | Understanding the role of progesterone in the normal human breast . . . . .                                | 298        |
|          | Understanding breast cell proliferation during the normal menstrual cycle . . . .                          | 299        |
|          | Gaining a better understanding of how therapies influence breast cancer risk . .                           | 299        |
| 9.4      | Final words . . . . .                                                                                      | 300        |

# Declaration

I hereby declare this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Kate Simms and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968. I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

*Signature*..... *Date*.....



# Acknowledgments

My time as a PhD student has been one of the most rewarding periods of my life. I have had great supervisors and have met wonderful people throughout my PhD, whom I will never forget.

I would like to give special thanks to my supervisors Nigel Bean and Adrian Koerber for being patient and providing me with direction throughout my thesis. In particular, I am very grateful to my primary supervisor, Nigel, who always had time for me throughout my thesis and gave me encouragement and guidance when it was needed. I have said to many people that you are the best supervisor anyone can ever get!

I would also like to thank the team at the Hanson Institute, including my supervisor Wayne Tilley and Theresa Hickey, for being so welcoming and providing me with opportunities to communicate my results with a biological audience, and for answering my biological questions. In addition, I would also like to thank Professor Liz Musgrove from the Garvan Institute, for providing me with additional data and also for answering my biological questions.

Special thanks also goes to my fellow PhD students - I will thoroughly miss our long daily coffees and our long philosophical discussions! Also thanks to mum for listening to all my complaints so patiently. Lastly, a big thanks to Xiang, for being so encouraging and reading my work when I needed a second pair of eyes, and for just being there.



# Abstract

The aim of this thesis is to develop mathematical models of cell cycle progression which can be used in conjunction with biological experiments. The thesis focusses on modelling processes which have biological relevance, and uses mathematics to investigate biological hypotheses about mechanisms which drive experimental results.

In this thesis, we introduce a mathematical model of cell cycle progression and apply it to the MCF-7 breast cancer cell line. The model considers the three typical cell cycle phases, which we further break up into model phases in order to capture certain features such as cells remaining in phases for a minimum amount of time. This results in a unique system of delay differential equations which are solved numerically using MATLAB. The model is also able to capture a uniquely important part of the cell cycle, during which time cells are responsive to their environment. The model parameters are carefully chosen using data from various sources in the biological literature. The model is then validated against a variety of experiments, and the excellent fit with experimental results allows for insight into the mechanisms that influence observed biological phenomena. In particular, the model is used to question the common assumption that a ‘slow cycling population’ is necessary to explain some results. A model analysis is also performed, and used to discuss misconceptions in the literature regarding the average length of the cell cycle. An extension is developed, where cell death is included in order to accurately model the effects of tamoxifen, a common first line anticancer drug in breast cancer patients. We conclude that the model has strong potential to be used as an aid in future experiments to gain further insight into cell cycle progression and cell death.

The model is then applied to the T47D cell line, which has significantly different cell cycle kinetics to the MCF-7 cell line. The aim of modelling this cell line, which is naturally receptive to the effects of progestins, is to model the effects of progestins on cell cycle progression. It is important to understand the effects of this substance, as it has been used in hormone replacement therapies, and its effects on cell cycle progression are still not understood.

In order to understand how progestins influence cell cycle progression, a more detailed protein model is developed to get a better understanding of how progestin influences protein concentrations within a cell. We find that progestin effects on cell cycle progression

are complex, and that progestin can be considered to be both a proliferative hormone and an anti-proliferative hormone, depending on the cell's previous history of progestin exposure, and on the length of time the cells have been exposed to progestin. The fact that the timing of progestin exposure can have different effects on cell behaviour has profound implications for treatments that contain progestins, such as combined hormone replacement therapies.

In summary, this thesis develops mathematical models representing different aspects of the cell cycle, and uses a variety of sources in the literature to parameterise the models. The model results are used to give insight into mechanisms that play a role in cell cycle progression under different experimental conditions. The models have the potential to be used alongside experiments, giving further insight into the mechanisms that influence events, such as cell cycle progression in the presence of hormones, as well as cell death.

# Chapter 1

## Introduction

Cancer biology has its beginnings in the late 1800's, when scientists began to query the origins of cancer. During this time, many common diseases were known to be caused by external agents (such as viruses) which invaded the body. It was assumed that cancer, like other diseases, was also caused by an invasion of foreign agents. However, research along these lines proved to be less fruitful than originally hoped, and the causes of most human cancers remained a mystery. It wasn't until the 1950's that cancer research really boomed. This was due to the release of two independent epidemiological reports which both indicated that smokers were at least ten times more likely to develop lung cancer in their lifetime when compared to non-smokers. Tobacco companies subsequently became motivated to fund cancer research in an attempt to find contrary evidence to these reports that could potentially put them out of business. Not only did they fund research which investigated the causes of lung cancer, but they took the opportunity to appear philanthropic, giving money to any form of cancer research. Even some early mathematical models of cancer were funded by the tobacco research council [138]. Since this time, cancer research in both the biological sciences and mathematical sciences has grown greatly.

In the 60 years since this surge in cancer research, cellular and molecular biology has provided significant insight into the causes of cancer. Understanding how molecular changes in DNA result in changes in cell behaviour has provided a better understanding of the drivers of cancer, and treatments have been developed to directly counter these drivers.

In the mathematical sciences, statistics has proved to be of great use when deciphering which genes are involved in cancer via the use of microarray experiments. Mathematical modelling has also been used to better understand the complexity of cellular processes. As more experimental data becomes available, mathematical and statistical tools are likely to become more useful in interpreting this data. However, even today, mathematicians and cancer biologists seem to perform investigations somewhat independently of each other.

## 1.1 Thesis summary

In this thesis, we will combine the tools of mathematical modelling with biological experimentation in order to gain a better understanding of breast cancer cell division. Cell division, or **proliferation**, is a complex process whereby a cell produces a second cell (almost) identical to itself. It is important to understand the process of cell proliferation, because cancer is caused by cells that proliferate uncontrollably. It is the continuing proliferation of cancer cells, even in the presence of anti-proliferative signals, that causes damage to the organs in which the cells are growing.

When cells proliferate, they progress through four distinct phases (known as the *phases of the cell cycle*) that eventually culminate in cell division. In this thesis we aim to understand how cells progress through the cell cycle phases, and how these cell cycle phases are altered in cancer cells. In particular, we will focus on understanding cell cycle progression of breast cancer cells growing in a controlled laboratory environment, known as growth **in vitro**. It is important to first understand how cancer cells behave in a controlled environment. This is because cells growing in a more complex environment (such as the human breast) have many factors influencing their behaviour, and it is difficult to separate the influence of the environment from the intrinsic behaviour of the individual cells, if you don't already understand that intrinsic behaviour. We begin by developing a model of cell cycle progression, which allows us to investigate cell proliferation after exposure to substances such as estrogen, insulin and chemotherapy drugs. We later develop a second model which investigates the changes in internal cell protein concentrations after exposure to progestin, which is an artificial form of the natural hormone progesterone. The models we develop are done so with careful recognition of the current biological understandings

of cell proliferation. By comparing the model results with those from experiments, we are able to gain significant insight into the behaviour of breast cancer cells.

In Chapter 2, we give an overview of the biology of cell behaviour and the biology of cancer development. We begin this chapter by describing the structure of a cell, and also the phases a cell progresses through as it undergoes normal cell division. In this chapter, we also describe how cancer progresses from a mass of normal cells to a dangerous malignancy, and we describe some mathematical models that have been developed to describe cells at these different stages of cancer malignancy. We also discuss how cancer forms in the human female breast, and the role of the hormones estrogen and progesterone in breast cancer development.

In Chapter 3, we describe the experimental techniques used to determine the changes in the cell cycle phases after exposure to various substances, and we also describe some previous mathematical models that have been used to investigate the process of progression through the phases of the cell cycle. We finish this chapter by introducing our mathematical model of cell cycle progression.

In Chapter 4, we discuss how we can apply our mathematical model to a particular population of cells known as MCF-7 cells (we will describe these cells in more detail in Chapter 2). We begin this chapter by showing how different experiments from the literature allow us to accurately determine the parameters of our model. We then use the model to get a better understanding of how different parameters influence model behaviour, and we also address some misconceptions in the literature regarding the average time a cell spends in the different phases of the cell cycle.

In Chapter 5, the parameterised mathematical model is then used to simulate cell cycle progression in the presence of estrogen, insulin and different types of anti-estrogens, with excellent comparison with experimental data. In particular, we will use the model to question a common assumption that a ‘slowly cycling population’ of cells must make up a proportion of otherwise fast cycling cells. We finish this chapter by proposing an extension to our model to include the effects of cell death in the presence of high concentrations of tamoxifen, with some interesting insight into the phases that this drug exerts its killing effects.

In Chapter 6, we re-parameterise our model so that we can simulate cell cycle pro-

gression using T47D cells. We find that T47D cells behave very differently to MCF-7 cells, even when exposed to similar substances, and we use our model to get a better understanding of the differences between these two cell lines. In this chapter, we consider a range of experiments that use T47D cells, and we use our cell cycle model to make conclusions about some experiments which appear to be outliers. The main aim of extending the model to T47D cells is so that we can get a better understanding of the effects of progestins (which are artificial forms of progesterone that are commonly used commercially in drugs) on cell cycle progression. T47D cells (unlike MCF-7 cells) are responsive to the hormone progesterone, and therefore these cells are used in experiments which investigate the role of this hormone (and its artificial variants) in cell proliferation, and this is why we extend our model to this cell line.

In Chapter 7, we investigate the effects of progestins on cell cycle progression in significantly more detail. In this chapter, we perform a thorough investigation of some qualitative experimental data on the changes in internal protein concentrations after progestin exposure. We are able to develop a quantitative understanding of how protein concentrations change after exposure to progestin by virtue of our mathematical investigations.

In Chapter 8, we build a more detailed protein model to get a better understanding of how progestin influences protein concentrations within a cell, and how these protein concentrations relate to progression of cells through the cell cycle. We find that progestin effects on cell cycle progression are complex, and that progestin can be considered to be both a proliferative hormone and an anti-proliferative hormone, depending on the cell's previous history of progestin exposure, and on the length of time the cells have been exposed to progestin. The fact that the timing of progestin exposure can have different effects on cell behaviour has profound implications for treatments that contain progestins, such as combined hormone replacement therapies.

The mathematical model of cell division we develop in this thesis is applied to a variety of experimental results from the literature, and to our knowledge, this is the first time a mathematical investigation has been performed on these experimental results. In addition, the mathematical model we develop is based on the principal that cells that move into a new cell cycle phase are 'stored' in this phase for a fixed amount of time,

which results in a unique set of delay differential equations. The idea behind the model development, and the resultant set of delay differential equations is an original way to deal with certain features of cell cycle progression. This unique form of our equations allows us to connect changes in internal cell cycle proteins to changes in the rate of cell cycle progression. We have written a journal article which has been accepted by the *Bulletin of Mathematical Biology* [108]. The journal paper focuses on results from our model of cell cycle progression.

The second model we develop in Chapters 7 and 8 is, to our knowledge, a first attempt to use mathematical modelling to understand the effects of progestins on internal cell cycle protein concentrations. The protein model developed in these chapters provides insight into the role of progestins in cell proliferation, and makes suggestions to biologists to perform experiments which may confirm certain hypotheses about the complex actions of progestins.



## Chapter 2

# Biological background

Cancer is a disease which affects individual cells. Thus to understand how cancer develops, we must have a good understanding of how a normal cell is supposed to behave, and how aberrations in its behaviour emerge. We will limit our discussion to cells found in animal bodies (known as **eukaryotic** cells as opposed to prokaryotic cells, which are bacterial).

### 2.1 The cell structure

All animal bodies are composed of millions of different cells. Each cell is specialised to perform its required function, and different types of cells can have vastly different shapes and sizes. Regardless, there are certain internal cell structures (known generally as **organelles**) that are present in each cell across all animal species, some of which are shown in Figure 2.1. In this figure we can see the **mitochondria**. These organelles are the powerhouse of the cell because they are the main source of chemical energy (known as ATP) for the cell. We can also see **lysosomes**, which are the cells' waste disposal system. The lysosomes break down unwanted proteins and other waste products that the cell produces.

Within each cell is also a **nucleus**, which contains **chromosomes**. The nucleus surrounds these chromosomes to protect them from harmful products that are present in the cell. Chromosomes are highly organised structures made up of molecules known as deoxyribonucleic acid or **DNA**. The DNA is the instruction manual of the cell, and con-

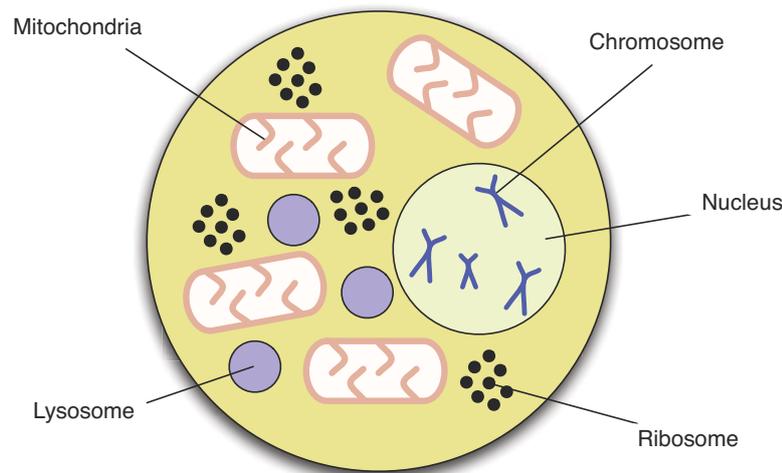


Figure 2.1: *The basic structure of an animal cell.*

tains all of the information required for a cell to perform its intended function. Anytime a cell is required to produce a protein, the DNA is consulted first for the code outlining the proteins' construction. To produce this protein, the information encoded in the DNA is copied onto a template, known as messenger RNA, or mRNA. The mRNA is exported from the nucleus and sent to the **ribosomes** (see Figure 2.1), which then translate mRNA into the final protein product. Once the protein has performed its intended action, it is often broken down in the lysosomes.

Let us investigate the structure of the chromosomes in more detail. Figure 2.2 shows how the DNA is tightly coiled together to make up the structure of a chromosome. If we were to unravel all of the DNA from the 46 chromosomes within a human cell, the total length of DNA would span two metres [2]! Figure 2.2 also shows that the DNA itself also has a structure. The exact structure of DNA was not known until 1953, when Watson and Crick discovered that DNA is made up of two chains twisted around each other in the shape of a helix [132]. Subsequent investigations by Crick showed that each of these chains, or backbones, are connected by molecules called **nucleotides** or **bases**. These nucleotides can be one of only four possible molecules: adenine, cytosine, guanine and thymine. Each of these bases is always paired with its complementary base on the opposing strand, as shown in the more detailed diagram of DNA in Figure 2.3 (thymine always pairs with adenine and cytosine always pairs with guanine). Thus a base and its

NOTE:  
This figure is included on page 23 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 2.2: *A typical cell. Every animal cell contains a nucleus, which houses multiple chromosomes. The chromosomes are made up of DNA. DNA can be thought of a distinct sets of genes, which code for specific proteins. Figure taken from [90].*

complementary pair are known as **base pairs**. The fact that there are only four types of bases that string together to make up a strand of DNA led to the realisation that traits are indeed stored in a code, and that this code could be broken simply by deciphering the sequences of base pairs that make up DNA.

It is now known that strings of these bases code for the proteins essential for cell function. A sequence of bases that code for any one protein is called a **gene**. The concept of a gene was introduced by Gregor Mendel in 1860 (before DNA was discovered), when he noticed that traits of an organism were inherited from the parents independently of other traits. He suggested that each trait was stored as an independent packet of information. Although the gene is still thought of as an independent packet of information, it is now known that most traits in an organism are not coded by a single gene, and rely on many genes to properly develop. An example of a gene on a strand of DNA is shown in Figure 2.2.

## NOTE:

This figure is included on page 24 of the print copy of the thesis held in the University of Adelaide Library.

Figure 2.3: *The structure of DNA contains two backbones, made up of sequential nucleotides. The backbones wind around each other in the shape of a double helix. The nucleotides in each strand (shown in detail at the bottom of the figure) contain a segment which makes up part of the backbone (the P and S symbol) and a segment called a base (represented here by the C symbol). There are four types of bases - adenine, cytosine, guanine and thymine. The two backbones in the double helix are connected by the bases. Figure taken from [90].*

## 2.2 Cancer is the result of mutations in DNA

It is now known that most human cancers arise because of changes to the DNA structure (or **genotypic** changes) within a cell [46]. This results in changes in the behaviour (or **phenotypic** changes) of the cell. An unintended change to the structure of DNA is referred to as a **mutation**.

There are many ways in which the DNA within a cell can be altered. The most common cause of mutations is through exposure to agents which change the chemical composition of DNA [43]. These agents can cause mutations in one of two ways [135, pg. 38]. One way is to directly change the sequence of DNA (for example, by changing a base from cytosine for thymine), known as genetic mutation. Another is through epigenetic means, whereby the structure of DNA remains unchanged, but the **expression** of certain genes becomes altered [33] (for example a gene can be **silenced**, or switched off, by the attachment of molecules to the DNA). The expression of a gene is its ability to create the protein that it codes for, and so an increase in expression would result in more of that protein, while a decrease in gene expression would result in a reduction in that protein.

### Causes of mutations

Some mutations are caused by environmental factors. Common environmental factors known to cause cancer include tobacco, radiation and viruses such as the Human Papillomavirus (or HPV) and the sarcoma virus [135, pg. 65]. Tobacco is highly efficient at causing damage to cells in the lung, increasing the lifetime chance of cancer in a smoker by a factor of ten [59]. High frequency radiation is known to directly alter the genetic make-up of DNA, causing mutations that eventually lead to cancer [135, pg. 46]. This includes UV radiation from the sun, the main cause of skin cancers [29]. Viruses often cause cancer through epigenetic means, attaching themselves to DNA and changing the expression of particular genes, and this is how the virus HPV causes cancers such as cancer of the cervix [51]. Some viruses, such as HIV, can increase the risk of cancer indirectly by weakening the body's immune system and hence the immune response to the presence of cancer [135, pg. 677].

Mutations can also occur during cellular replication, when the DNA from the parent

cell must be copied, and the copy passed down to the offspring. Errors in copying are thought to be quite common, and it is thought that many cancers begin because of this unavoidable and unlucky circumstance [42].

It is known that distinct abilities are progressively gained by cells as they obtain cancerous mutations. It is not that a cell directly moves towards these traits - the site of a mutation is random, and many mutations that occur within a cell do not provide a survival benefit to the genes within that cell (indeed many mutations may not be compatible with cell survival, and that cell may be removed). However, it is possible that after some time, a mutation may occur which gives the mutated cell a new phenotype that happens to provide the cell a survival advantage over its neighbours. Such a phenotype is more likely to 'spread' its genetic code, resulting in more and more of that genotype in the population. With more cells in the population that contain the advantageous phenotype, the chance that a second mutation will occur alongside the first is increased, and this new population will eventually outnumber both the original population and the population with only one mutation, resulting in a mass of cells with two beneficial mutations. This process can be repeated, resulting in a cluster of cells all with similar behaviour, which will be different to their normal cell counterparts due to the several successive changes in cell phenotype. This is essentially the idea of evolution by natural selection applied to the emergence of cancer.

### **2.3 Cancers come in different stages**

By the time that cancer is found in a human body, the cells with the most beneficial mutations have become most prevalent in the region, causing a mass of cells with similar genotype to one another, but significantly different to the cells that would normally exist in that region. Thus, cancers that are discovered early tend to resemble the normal cell neighbours more than cancers that are discovered late. Over the years, cancers of different degrees of aggressiveness have been discovered. It was these different degrees of aggressiveness, with the least aggressive cancers resembling the normal cells in that region, that indicated that cancers originated from inside the body rather than from external sources.

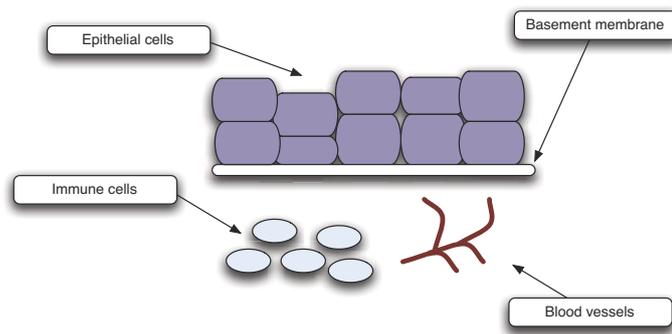


Figure 2.4: *Epithelial cells (dark purple) are tightly packed together and are separated from the rest of the body by the basement membrane. Underneath the basement membrane is the stroma, where the remaining tissue of the organ resides. Shown here are immune cells and blood vessels making up part of the stroma.*

Most human cancers arise from normal cell types known as **epithelial** cells [125], and such cancers are referred to as **carcinomas**. Epithelial cells are found lining organs, such as skin, breast ducts and the gastro-intestinal tract. Thus, they are often directly exposed to the external environment of the organ, and this is thought to be the reason as to why so many cancers originate in epithelial cells. Epithelial cells are tightly packed together and are separated from the other cells in the organ by the **basement membrane** (see Figure 2.4). The other side of the basement membrane is referred to as the **stroma** and contains blood vessels, immune cells and molecules which make up the structure of the tissue (such as collagen).

The least aggressive type of cancer growth is referred to as **hyperplastic** growth, which is simply an increase in the number of epithelial cells [135]. The epithelial cells still retain their connections with each other and remain above the basement membrane. An example of hyperplasia is shown in Figure 2.5. Such growths are not considered harmful if discovered in a body, and are referred to as **benign** growths. A cancer is not considered to be harmful or **malignant** unless the epithelial cells have been observed to break through the basement membrane [135]. This is shown in Figure 2.6. The act of local invasion of cancer cells is not in itself harmful. However, the cells now have direct access to the blood stream and have the option of progressing to distant sites in the body. The most dangerous cancers are generally ones that have spread to other parts of the body through the blood supply. Cancers that have spread to other sites are called **metastatic** cancers, and are generally considered to be incurable.

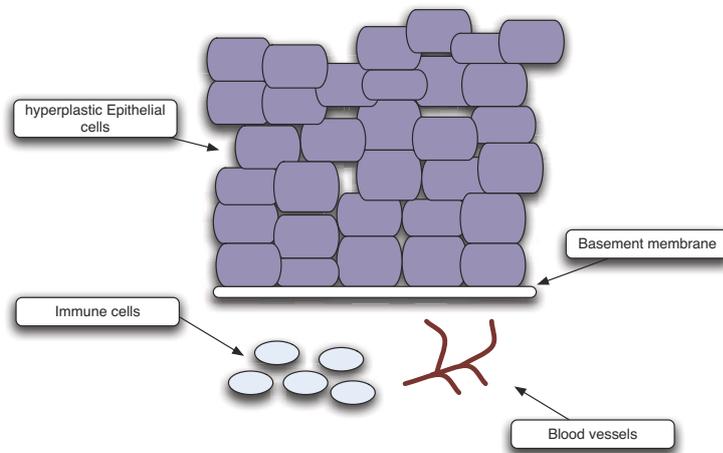


Figure 2.5: *Hyperplasia of the epithelial cell layer is shown here. The epithelial cells retain their connections between each other and remain above the basement membrane.*

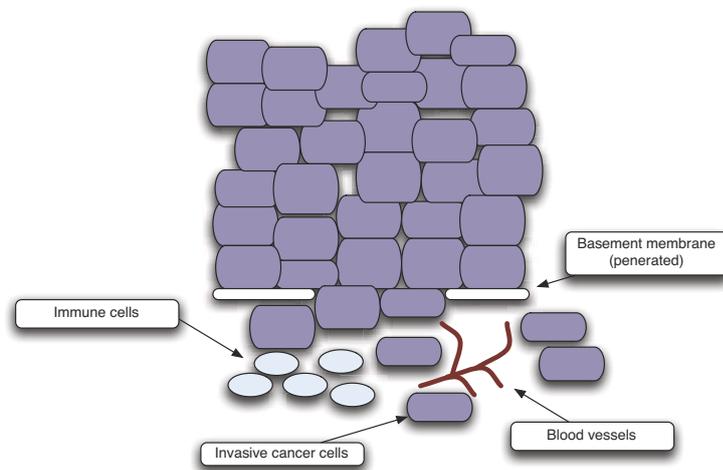


Figure 2.6: *The epithelial cells have broken through the basement membrane and invaded the stromal tissue. Such cancers are now considered to be malignant and represent a serious threat to the body.*

## 2.4 The typical hallmarks of cancer

The different stages of cancer growth just discussed are features which we can recognise when we observe a cancerous mass forming, and these features are used clinically to judge how advanced the cancer has become. However, the specific changes a cell must undergo at the level of DNA as it moves from being normal to cancerous is more difficult to determine, and has been intensively studied. In [39], it was proposed that very specific mutations give rise to most cancer types cancers. The authors proposed that six such independent changes in cell behaviour are required for a cell to move from being normal to cancerous. This article, published in 2000, had a significant impact on the cancer research community, because the complexity of cancer was summarised and classified into six simple, distinct events. By May 2011, the article became the most cited article of all time in the Journal *Cell*, with 10,000 citations. We will discuss these six hallmarks in this section.

### Hallmark 1: Self-sufficiency of growth signals

Cell division, or **proliferation**, is a process that is usually tightly regulated within the human body. Cells rely on signals from their neighbouring environment in order to divide. Proteins secreted by cells in the neighbouring environment confer messages that eventually control a cell's decision to divide and even the decision to survive. For instance, if a cell is surrounded by many other cells, signals discouraging cell growth, or even encouraging cell death, may be secreted by nearby cells. A cell is constantly receiving different types of signals from its environment. The cell constantly weighs the effect of these signals, and as long as the anti-proliferative signals remain stronger within a cell, the cell will not commit to cell division. However, once the proliferative signals sufficiently outweigh the anti-proliferative signals, the cell makes a (usually irreversible) commitment to undergo cell division.

Due to this tight regulation by neighbouring cells, it is difficult for a cell to take the decision to proliferate into its own hands. Densely packed cells are often surrounded by a large amount of anti-growth signals and very few growth encouraging signals. Thus, one of the first hurdles that a normal cell must overcome is the ability to produce its own

growth signals. This is often performed by over-expressing genes which are known to cause proliferation. Proliferative genes are known as **oncogenes**. A cell which over-expresses oncogenes has the effect of drowning out the anti-growth signals from its neighbours, and keeping the balance in favour of proliferation.

### **Hallmark 2: Insensitivity to anti-growth signals**

When the environment discourages cell division in neighbouring cells (which happens when, for example, it becomes too crowded), a cell responds by increasing the expression of genes which discourage proliferation. Such genes are called **tumour suppressor genes**.

If a small clump of cells which have the ability to secrete their own growth factors grow too large in number (due to the over-expression of oncogenes), then signals which discourage growth become more intense as the mass of cells becomes larger, and these signals can halt the growth of the cancer mass. Thus, another essential step towards cancer development is the ability for the cells to become insensitive to the anti-growth signals, often through switching off the tumour suppressor genes.

### **Hallmark 3: Evasion of cell death signals**

Cell survival is also controlled by the surrounding environment. If the body deems it more beneficial to have a cell removed for the greater benefit of the organism, certain signals (driven by tumour-suppressor genes) will converge on the target cell, requesting it to activate its internal cell suicide process, known as **apoptosis**. A cell undergoing apoptosis will begin to break down its internal structures, while releasing certain signals, calling upon the body's clean-up population of cells, known as **macrophages**. These macrophages will converge on the dying cell, engulfing it before it bursts and spills its otherwise toxic contents. Cell death by apoptosis leaves minimal debris in its wake that may otherwise compromise the survivability of neighbouring cells.

Cell death signals are also released when too many cells exist in one region, making cells which have high rates of proliferation more likely to receive signals to activate apoptotic programs. Avoidance of apoptosis is essential in the development of cancer.

**Hallmark 4: Unlimited replicative potential**

In the 1960's, it was demonstrated that cells would eventually stop growing in culture despite the presence of growth factors [18, 73]. This is because mammalian cells do not have the ability to divide indefinitely, and so eventually cells will be unable to create daughter cells.

The reason cells are unable to divide after a certain number of divisions is because the DNA strands become unstable. Experiments have shown that the ends of single DNA strands fuse together, creating instability [68]. In mammalian cells, the ends of DNA strands have long repeating sequences of **telomeres** which act to prevent the fusion of the ends of DNA strands [92]. Each time the DNA is replicated in preparation of cell division, some of the ends of the telomeres are cut, and hence subsequent copies of DNA have shorter telomere length. As more divisions take place, the telomeres become progressively shorter until eventually they are too short to effectively protect the ends of chromosomes. Fusion of the ends of chromosomes occurs once the telomeres are too short, creating chromosomal instability or **crisis**, and often causes the cell to undergo apoptosis [30, 4].

In many cases, cells will recognise the short length of their telomeres before chromosomal instability induces apoptosis. In this case, cells activate programs which prevent any further cell division, a process called **senescence** or **differentiation** [30]. Senescent (or differentiated) cells no longer divide, however they remain healthy within the body [135, page 480].

Thus, a population of cells in the same vicinity are not really identical biologically, as each of these cells will have a different level of differentiation. If we assume that every cell division reduces the telomere length of both daughter cells, then we will eventually reach a point when no cells in our body will be able to divide - they will all have become senescent or entered crisis! Thankfully, this is not what happens, as it is now known that there exists a population of cells called **stem cells** in each tissue. These cells show no signs of a limit to their replicative potential and are thus called **immortal**. Stem cells, however, rarely divide. On occasion, they will divide and produce what is known as a **transit-amplifying cell**. These transit-amplifying cells do not have unlimited replicative

potential, however they have high rates of division, and divide often enough to create the bulk of tissue we observe. Only one transit-amplifying cell is required to produce a large amount of offspring, and so stem cells rarely need to divide.

The concept of stem cells existing in tissues in the body makes the comparison between natural selection of animal species and natural selection at the cellular level more complicated (animal offspring cannot ‘run out of’ the ability to procreate). It is clear that unlimited replicative potential must be obtained in a cancer mass, otherwise the population of cells will run out of telomeres and never amount to cancer. It is possible that some cancer cells originate as stem cells, and therefore were never required to ‘overcome’ this hallmark. In many cancer types, it is not known for certain whether the cancer populations start with a cancerous stem cell or a transit amplifying cell that developed unlimited replicative potential.

### **Hallmark 5: Sustained angiogenesis**

We note that a mass that has acquired any, or all, of the first four hallmark requirements (a difficult task) will likely display hyperplasia (as shown in Figure 2.5). As we discussed earlier, hyperplastic growths are considered benign, and so although the mass has already overcome four hurdles, the remaining hurdles are often much more difficult to overcome.

One of the difficult traits for the cells of a cancer mass to gain is the ability to grow very large. This is due to the fact that there is a limited nutrient supply, which imposes a limit to the number of cells that can form in an area. This was demonstrated by Judah Folkman in the 1970’s. He injected a sample of tumour cells into an eye surface (which has only limited blood supply), and observed that within 2 weeks, the mass had grown to about 0.5mm in diameter, and thereafter, ceased to grow [122]. This tumour dormancy had been observed in patients before, and it was thought to be due to cell cycle arrest (possibly because of reductions in telomere length) or to intervention of the body’s immune system. However, Judah Folkman showed that when the sample was injected into the iris (which is dense in capillary tissue), the mass grew to 16,000 times its original size within three weeks. This provided clear evidence that the dormant tumours found in these patient went dormant because of a lack of blood supply, and not because the cells had entered senescence.

Further investigations have shown that if capillaries are not too far away, they appear to grow towards the tumour, eventually infiltrating the tumour and supplying it with its own direct blood supply! We now know that it is not just the cancer cells themselves that send out signals to encourage the growth of capillaries, but the nearby immune cells also send out signals, and help to encourage growth of the blood vessels.

Why would normal cells, such as immune cells and cells that make up the capillary walls, provide benefits to the threatening mass of abnormal cells growing from the nearby epithelium? The answer is that the cancer cells trick the body into thinking that the cancerous mass is a wound. Cancer cells eventually gain the ability (through mutation events) to secrete signals such as platelet derived growth factor (PDGF) and matrix metalloproteinase (MMP). Both of these signals are secreted by epithelial cells when the surface of the epithelium has been wounded. PDGF stimulates nearby immune cells, which will actively secrete growth factors to help stimulate the apparently injured epithelium. These stimulated immune cells are also ready to clean up any dead cells or other debris/bacteria that may cause further harm to the cells in that region. The MMP's released by the cancer cells cleave the surrounding tissue and release vascular growth factors, subsequently encouraging capillary growth into the 'wounded' site through the proliferation of cells that line the capillary walls, called **endothelial cells**. The influx of capillaries into the cancer mass is referred to as **angiogenesis**.

The presence of non-cancerous cells within a tumour mass is critical to the survivability of the tumour, and dispels the previously thought notions that everything about a cancerous mass can be understood by looking at the genetic make-up of the cancer cells. In fact, in most carcinoma masses, 90% of the cells present in the mass are non-cancerous stromal cells, such as immune cells and endothelial cells [135, pg. 528].

It has been observed that a mass can remain dormant in its pre-angiogenic state for many years. However, once a cell in the cancerous mass learns to behave like a wound, the process of angiogenesis is initiated, and the influx of vessels causes a sudden and dramatic increase in the growth of the tumour. This sudden change is referred to as the **angiogenic switch**.

The discovery of the role of angiogenesis in tumour growth produced a lot of optimism. Before this time, many drugs aimed at directly killing the cancer cells (e.g. chemothera-

peudic agents). However, due to the naturally high rate at which cancer cells can adapt to new environments, the success of such drugs was often temporary, and the cancer mass would be observed to grow again, however this time it would be resistant to the chemotherapy used. In addition, it is often hard to get the drug to reach all parts of the tumour mass due to the density of cells in this region. The possibility of introducing an anti-angiogenic treatment is attractive because such a treatment would be aimed at killing the normal blood vessel cells, which would be unlikely to ‘mutate’ and become resistant to the therapy. In addition, there would be no trouble in making sure the drug reached the site, because all it would need to do is find its way into the blood supply and it would eventually be transported to the capillaries at the cancerous site. The drug could be specifically tailored to target capillaries within the tumour, as it is known that capillaries that grow in response to wound sites are quite different to normal blood vessels, thus other parts of the body could be spared the attack from the drug (unlike chemotherapy, which kills any growing cell, causing many side effects to patients taking the drug). This avenue is still being heavily investigated as an option for cancer treatment.

### **Mathematical models of pre-angiogenic growth**

Many mathematical models have been developed of growth prior to angiogenesis (termed **avascular** growth), and in these models it is assumed that the first 4 hallmarks are present. In [14], the growth of a cancerous mass was investigated mathematically to determine whether a mass would become limited in size due to lack of nutrients, as observed by Judah Folkman. The models assumed that the mass was made up of a **clonal population** of cells (which is a population of cells with identical genotype and phenotype) that had gained the first 4 hallmarks, and it was also assumed that the mass was spherical. Differential equations were used to describe the growth of the radius of the tumour mass, where growth depended upon the radius of the tumour mass and the amount of nutrient the tumour was being supplied. The nutrient supply to the tumour was presumed to come from the edges of the tumour (and hence the larger the radius, the more nutrient could be supplied to the tumour). The diffusion equation was used to model the diffusion of nutrients towards the centre of the tumour, and cells which were too far away from the nutrient source were removed from the system. Solving the equations

NOTE:  
This figure is included on page 35 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 2.7: *The results of a cellular automata simulation of avascular tumour growth, obtained from [76].*

showed that once the tumour reached a certain radius, the rate at which cells are removed from the centre of the mass due to the limited nutrient supply equalled the rate of growth of the tumour mass, and so overall the growth of the tumour mass became stagnant. Such studies provided mathematical evidence for Judah Folkman's observations that the growth of a tumour can be limited by simply considering the availability of nutrient sources. These mathematical models have been further modified to include additional features such as the effect of the debris left behind by the dead cells on the nearby live ones [131].

Other models of avascular tumour growth have included cellular automata, which are useful because the assumption that the tumour mass has a uniform radius is not required in such models. Cellular automata models tend to produce results which have more jagged edges, which is more comparable to experimental observations. Cellular automata models are made up of some tiling space, often a square grid. Each grid space would follow certain probabilistic rules which determines what that grid space would be occupied by. An example of a simulation of avascular tumour growth using a square grid is shown in Figure 2.7. In this representation, grid spaces either contain a tumour cell that is able to divide, a tumour cell that receives enough nutrient to remain alive but not enough to

divide (called **quiescent**) or a tumour cell that has died from lack of nutrients (called a **necrotic** cell). In these simulations, the edges of the tumour appear jagged, and resemble better the tumour masses observed experimentally when compared to simulations from differential equation models.

Note that in these simulations, the tumour cells are assumed to have ample space to move into, reminiscent of growth in a petri dish. However, in most early epithelial cancer growths, the epithelial cells are unable to penetrate the basement membrane until angiogenesis begins. In addition, the nutrient source is not available from anywhere on the outside of the developing mass - the vessels reside below the basement membrane, and hence the supply of nutrient will only come from diffusion across the basement membrane. In the case of skin cancers for instance, the cells can only get nutrient from below the skin, and so assuming that nutrients are available from all angles is unrealistic in these circumstances. Including the effects of the basement membrane could produce simulated masses that better resemble skin cancer growths, or polyps that are often found in the colon and other areas of the gastro-intestinal tract.

### **Mathematical models of capillary growth during angiogenesis**

Many mathematical models have been developed to investigate the growth of capillaries when cancer masses activate the angiogenic switch. The release of vascular growth factors by the cancer mass, diffusing through the intermediate tissue and subsequently stimulating the growth of endothelial cells has motivated various random walk models and continuous diffusion models of the process. Various simulations of blood vessel growth patterns have emerged from these studies, one example is shown in Figure 2.8. In this figure, we can see that the vessels start to branch out in the direction of the source of the the vascular growth factors, and this branching pattern compares well with those observed in experiments.

We note that few mathematical models have taken into account the presence of other stromal cells in the stimulation of capillary growth, which is important since a more advanced cancer mass is often made up of 90% stromal cells and only 10% cancer cells. In fact, as we saw earlier, the presence of immune cells in a cancer mass has been deemed an additional critical hallmark of cancer in the updated *Hallmarks of Cancer* article [38], as these cells contribute greatly to the ability of cancer cells to become invasive and

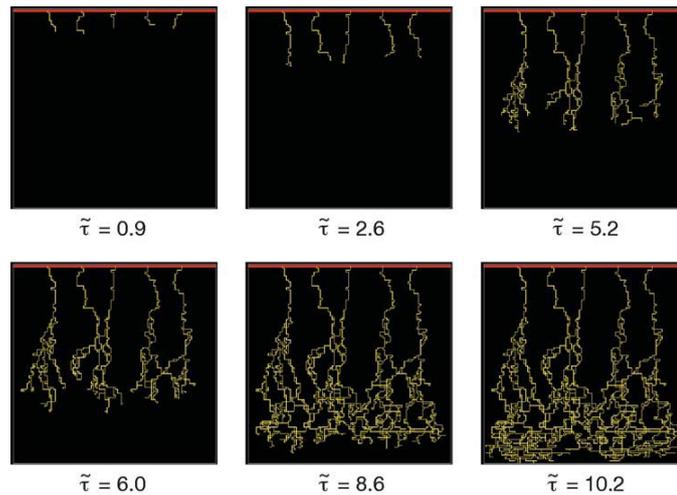


Figure 2.8: *Stochastic simulation of angiogenesis performed in [71]. Capillary cells grow in the direction of the vascular growth factor (bottom of the simulation screen), forming the pattern shown over time.*

metastatic. A model which can appropriately account for the influence of other cell types, and their interactions with the surrounding cancer cells, could provide insight into the mechanisms of cancer cell invasion and metastasis.

### Hallmark 6: Invasion and metastasis

The final hurdle that a cancerous mass must overcome is the ability to invade the local tissue and metastasise to distant sites. Invasion (breaking through the basement membrane) occurs at a similar time to the angiogenic switch. This is because breaking through the basement membrane relies upon the presence of MMP's. In initiating the angiogenic switch, the MMP's cleave the surrounding tissue releasing vascular growth factors. The MMP's can also cleave the basement membrane, breaking it down and providing the cancer cells access to the stroma outside of the basement membrane. There is, however, an additional feature of the invasion process, and this is the ability of the cancer cells to survive without being attached to each other. As discussed earlier, epithelial cells are tightly connected to each other. The ability for epithelial cells to survive without this attachment comes through the activation of the wound healing pathway. When there is a wound, the epithelial cells near the edge of the wound move across the wound to cover it from further external sources of harmful molecules or bacteria. To do this, epithelial

cells undergo a change in phenotype from their normal epithelial nature to a phenotype that resembles more of a mobile stromal cell called a **mesenchymal** cell, a process called epithelial-mesenchymal-transition (EMT). It has been observed that cancer cells near the edge of a cancer mass can undergo this transition and invade the nearby stromal tissue.

Once cells have gained the ability to invade the surrounding tissue, they have ample access to the blood stream and hence have the ability to undergo metastasis. The process of metastasis is not well understood. It is thought that once a cell undergoes EMT and invades the local stromal environment, it can then enter the blood stream through the capillary system. However, the blood stream is hostile, and the hydrodynamic forces within a blood vessel are likely to kill most cancer cells that make it this far [135]. Experiments show that cancer cells can recruit large amounts of platelets (possibly attracted to the cancer cell via the signals they release). These platelets can surround the cancer cell, protecting it from the harshness of the blood stream and allowing it to move safely to distant sites [135].

Once at the distant site, the cancer cell must penetrate the capillary walls at this site, and then successfully grow in the foreign environment. This last part of the metastasis process, known as **colonisation**, is not well understood. However, as metastasis is the process responsible for almost all cancer related deaths, many cancer researchers have been focussing on the changes that must occur in order for a cell to undergo this final hurdle [135].

### **An enabling hallmark: Increased rates of mutation**

The six hallmarks described above are each gained by cancer cells through changes in their DNA. However, there are many DNA monitoring proteins present within cells, constantly protecting the DNA from any mutations, or recognising when changes do occur and quickly delegating other proteins to repair the changes. Thus, increasing the rate of DNA mutations within a cell is an essential characteristic. It was not considered one of the six hallmarks of cancer in [39], however it was listed as an essential ability that must be gained in order for the other hallmarks to occur within a normal human lifetime.

There are many levels of protection a cell employs in order to avoid unwanted changes

NOTE:  
This figure is included on page 39 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 2.9: *During DNA replication, the double-helix unwinds. The protein DNA polymerase (not shown) attaches to each strand and creates a complementary strand. Obtained from [86].*

to a cells genome. For example, the very structure of DNA is naturally protective to the base-pairs that make up the DNA code: The helix-shaped DNA is essentially ‘squished’ so that the parts of DNA that are directly exposed to the outside are predominantly the DNA backbone, while the base-pair codes are hidden inside [135]. This sort of packaging provides protection from mutagens that may be present in the cell, and also from the effects of radiation. The only time the DNA is not in this protective form is when the double-helix structure unwinds during DNA replication, and the base-pairs become more exposed. In addition, the process of DNA replication itself is wrought with copy-error, which can result in unwanted mutations. Thus, cell division is a dangerous process, and significantly increases the chance that cancerous traits will emerge, and this is why the first two hallmarks discussed previously focus on increasing a cells overall rate of cell division.

To minimise copy-error, the cell has a variety of error-checking proteins to ensure that DNA is copied faithfully to the daughter cell. During DNA replication, the double-helix structure unwinds, and the protein **DNA polymerase** creates a complementary strand to each original strand (shown in Figure 2.9). DNA polymerase has a proof-checking mechanism. Each time a complementary base is created, DNA polymerase checks that

the base created is correct [135, page 477]. If the base is not correct (for instance if adenine, instead of cytosine, was created to pair up with guanine) then the polymerase unit would backtrack to re-create the copied base. If the DNA polymerase fails to pick up an error made in the replication process, another family of proteins called *mismatch repair* enzymes [135, page 477] follow close behind, and will attempt to pick up any mistakes which the DNA polymerases missed through their proof-reading mechanisms.

Together these checks result in a very low rate of copy errors in DNA replication. DNA polymerases will make only one mistake in  $10^5$  nucleotides. The proof-reading mechanism employed by the DNA polymerases will miss 1 in  $10^2$  errors that have been made. The final level of protection, the mismatch repair enzymes, will only miss 1 in 100 mistakes that were not previously picked up. This results in an overall error rate of 1 in  $10^9$  nucleotides [135, page 479]. As each cell contains about  $6 \times 10^9$  nucleotides [135, pg. 258], this results in about 6 errors that go unchecked per cell each time it divides. This may seem high, but about 97% of human DNA is not known to code for anything [135, page 479]. Such DNA is known as ‘junk’ DNA. Thus, most mutations that occur within a cell are likely to occur on junk DNA segments and therefore will have no impact on the cell’s function. From Figure 2.9, we can see that once the DNA has successfully been replicated, the two new copies of DNA will contain one original strand (gray) and one new copied strand (brown). Thus, both daughter cells will have one copy of the original DNA strand and one copy of the new strand, making cell division truly symmetric, in that neither cell will receive both copied strands (which would make it more likely to have inherited a mistake).

Mutations that manage to bypass the repair enzymes are not always permanent, and this is due to the presence of the p53 protein. The p53 protein, otherwise known as the ‘master guardian’ of the cell, can detect whether changes have been made to the DNA [46]. If the protein detects a mutation, it will immediately pause the cell division process, dispatching other proteins to fix the damage that was done [15, 46]. If the damage is deemed too severe to be effectively reversed, then p53 activates the cells’ internal apoptotic program, and the cell itself is discarded [32, 72]. The p53 protein can also detect mutations that occur through other means, such as radiation damage, or the presence of carcinogens [66], and thus this protein is very effective at preventing mutations to the DNA that may increase the risk of cancer. The p53 gene is itself found mutated (and therefore less effective) in about half of all human cancer types [54], implying that

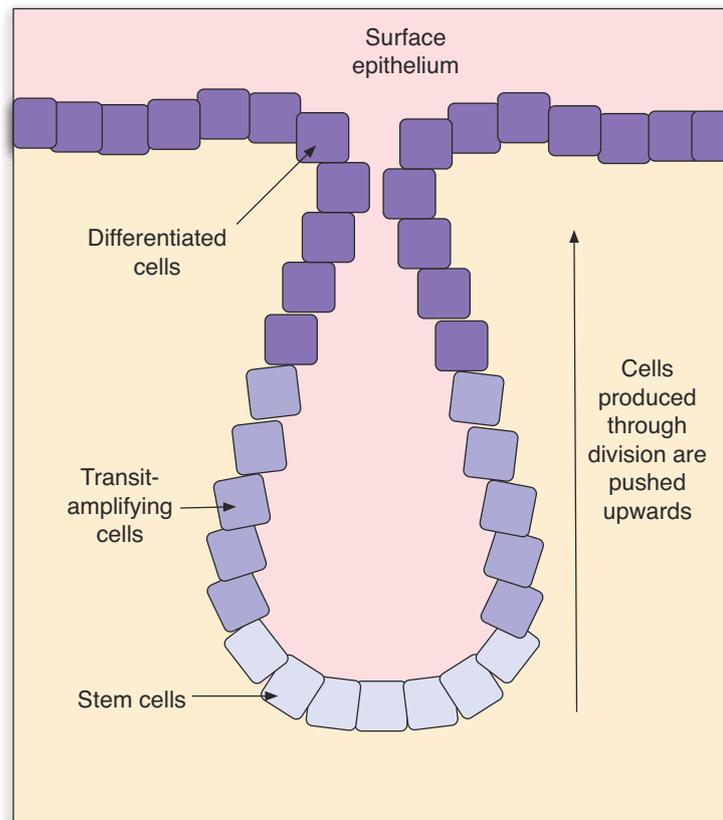


Figure 2.10: *The structure of the colon epithelium. Stem cells are at the bottom of a niche (light blue coloured cells). When they divide to produce a transit-amplifying cell (darker blue), this cell is pushed further towards the outside of the epithelium. Progressive cell divisions produce daughter cells closer to the outside, as indicated by the arrow pointing upwards. Thus cells that exist at the surface of the epithelium are differentiated cells (purple).*

without this gene, cancerous traits are much more likely to emerge.

### **The structure of the epithelium as a protection against cancer**

Most cancers occur in epithelial cell populations and, as mentioned earlier, this is thought to be due to the fact that these cells are exposed to the external environment of the organ. In many organs, the structure of the epithelium is designed to protect the stem cell population from the harsh outside. The structure of the colon epithelium is shown in Figure 2.10. This figure shows how the stem cell population, placed at the bottom of a niche, is protected from the harshness of the outside of the organ [75, 22]. The differentiated cells may be exposed to mutagens, however changes in their DNA are not

carried down to successive generations because these cells run out of telomeres (and lose the ability to divide). Thus the expendable cells, which have only a finite amount of replicative potential, make up the outer (more dangerous) layer of an organ while the stem cells are protected in safer compartments, away from the harmful exposure of external environments.

Stem cells divide very infrequently [96, 22], and this small rate of division of stem cells means that they are less likely to develop mutations in their DNA due to copy errors. Although transit-amplifying cells divide very frequently, they eventually become differentiated. On the rare occasion that stem cells do divide to produce a transit-amplifying cell, they perform asymmetric cell division, whereby both of the copied DNA strands are passed to the transit amplifying cell, and both original DNA strands are kept with the stem cell [77, 57]. Asymmetric cell division is another layer of protection to the DNA within a stem cell population.

In summary, we can see that there are many levels of surveillance and protection against unwanted changes to a cell's DNA, and if a cell cannot bypass some, or all, of these levels of protection, then it will be unable to gain the six hallmarks discussed earlier within a normal human lifetime.

So far, we have given an outline of how cancer develops. We have discussed the specific traits that gradually emerge in cancer cells, and how these traits relate to the degrees of aggressiveness of a cancerous mass. We now wish to discuss, specifically, how breast cancer develops.

## 2.5 Breast cancer

Breast cancer is the most common cancer in women [52], and so we wish to investigate how cancer originates, and how it progresses, in the breast. Hormones such as estrogen and progesterone play a critical role in the development of breast cancer. Treatment plans for patients with breast cancer rely heavily on the cell's sensitivity to these hormones [62, 105]. Many studies have focussed on deciphering the role of estrogen and progesterone in normal growth and development [94, 13, 55], however the role of these hormones in the development and progression of cancer is still not well understood [23, 61].

NOTE:  
This figure is included on page 43 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 2.11: *The structure of a typical female breast is shown in (a). The breast consists mainly of connective tissue and ducts. The image in (b) is of a cross-section of a breast duct, showing two populations of epithelial cells, and the basement membrane. Images taken from unpublished articles by Wayne Tilley's research team.*

### **Structure of the breast**

The breast is made up predominantly of connective tissue (which contains stromal cells and fat tissue) and breast ducts as shown in Figure 2.11(a). The vast majority of breast cancers originate from the epithelial cells that line the breast ducts (referred to as milk ducts in Figure 2.11(a)). A cross-sectional view of a breast duct is shown in Figure 2.11(b). We can see that there are two populations of epithelial cells that line the breast duct here, surrounded by an outer layer of basement membrane. Most cancer cells originate in either one of the epithelial cell populations, and the cancer becomes invasive when it breaks through the basement membrane of the duct to invade the breast tissue.

### **The role of estrogen and progesterone in the breast**

Estrogen and progesterone play a critical role, not just in breast cancer, but also in normal breast development. There are three types of estrogen - estrone (E1), estradiol (E2) and estriol (E3). E2 is the most common estrogen in the body, and is the estrogen primarily responsible for breast development, and the one most implicated in breast cancer. From this point on, estrogen will refer to E2, and we will not discuss the role of the other estrogens.

There is only one type of naturally occurring progesterone. However, this hormone has a very short half life (2-4 hrs). Thus, when administered in drugs, an artificial progesterone

is used. There are several types of artificial progesterone, which are collectively referred to as progestins.

Estrogen is only recognised by cells which express the estrogen receptor (ER) (known as **ER positive** cells) and similarly, progesterone is only recognised by progesterone receptor (PR) positive cells. Not all cells express ER or PR. In fact, only about 10% of breast cells will express a hormone receptor [24]. Of those cells that are hormone receptor positive, 96% of them will express both ER and PR [24]. Thus, about 90% of the cells in the breast are not responsive to either estrogen or progesterone, and of the remaining 10%, almost all are sensitive to both estrogen and progesterone.

It is known that estrogen increases the proliferation rate of cells within the breast. Interestingly, when an ER positive cell becomes stimulated by estrogen, it does not proliferate itself but releases signals which stimulate the proliferation of nearby ER negative cells. Thus, the ER positive cells cannot proliferate themselves in response to estrogen [24, 104]. However, two thirds of breast cancer masses found in the human female breast contain a high proportion of actively dividing cells that are ER positive [24]. In contrast to the normal breast, these cells actually divide in response to estrogen themselves. It is not known whether breast cancer starts out with the ER negative cells becoming cancerous and gaining the ability to express ER, or whether the ER positive cells become cancerous, gaining the ability to proliferate themselves.

### **Diagnosis of breast cancer**

Almost all breast cancers are detected initially by the presence of an unusual lump in the breast, which is denser tissue made up of cancer cells. Many women have regular screens for breast cancer, resulting in breast lumps being detected early on in disease progress. Many of these lumps are benign, and may never develop into breast cancers [67], however due to a general fear of breast cancer in the community, surgical removal is often chosen [67], even in cases where the lumps appear harmless. There has been a move in research to understand what causes some breast lumps to become cancerous and some to remain harmless for a woman's lifetime, and so diagnostic tools have begun to move towards cell based techniques. Mathematical and statistical tools have been used to determine whether certain cell markers are indicative of aggressive tumours [53, 56].

Information about a woman's previous history of pregnancy, obesity, family history of breast cancer and age at which menopause began, are all diagnostic tools that are used in conjunction with mammogram scans. If cancer is diagnosed in a patient, then often the treatment plans are based again upon the cellular markers found in the cancerous lumps [94]. In particular, the presence or absence of estrogen receptors on the cancer cells is a big indicator of prognosis, and heavily influences treatment choices.

Cancer development in the breast is thought to progress through the classic hallmarks of cancer described previously. Cancers originating in the human breast are significantly more common than cancers in any other part of the human body [49]. It is thought that this is because estrogen increases breast cell proliferation and therefore increases the risk of DNA mutations. Unlike cancers such as lung cancer, breast cancers are generally not influenced by external mitogens and are therefore difficult to prevent through lifestyle changes. However, it is now known that some external mitogens can increase the risk of breast cancer, in particular, some hormone replacement therapies. We will discuss these therapies in the next section.

### **Hormone replacement therapies and the development of cancer**

During menopause, estrogen levels in the body drop. This is because menstruation stops, and so the production of estrogen from ovaries ceases [25]. This drop in estrogen can cause osteoporosis, arteriosclerosis, night sweats and hot flashes [126]. To counteract these effects, scientists developed a hormone replacement treatment consisting of estrogen, which was effective in reducing these symptoms and in reducing the risk of associated disease [87]. However, it was soon discovered that this treatment alone resulted in an increase in the incidence of endometrial cancer [87] (endometrial cells are cells that line the uterus). This led to a new hormone therapy, known as combined hormone replacement therapy (cHRT), which included estrogen as well as a synthetic progesterone, (known as progestin), to counteract the increase in incidence of endometrial cancer [44]. Progestins were included because progesterone is known to decrease the proliferation rates of endometrial cells. Synthetic progestins were used because they have a longer half-life and high bioavailability when compared to natural progesterone. This therapy was successful in most cases at reducing the side effects of menopause without the increased incidence of

endometrial cancer. However, in 2002 a large study was conducted by the Women's Health Initiative to determine whether cHRT could increase the risk of developing breast cancer. The study was stopped early, due to overwhelming evidence that it does indeed increase the risk of developing breast cancer [65]. This was publicised, and resulted in a massive decrease in the number of women using cHRT. Within 2 years, there was a significant decrease in breast cancer incidence [100]. It is now known that when cHRT was developed, two incorrect assumptions about progestins were made - that progesterone induces an anti-proliferative effect in all parts of the body (just like it does in the uterus), and that synthetic progesterone mimics the actions of natural progesterone in all parts of the body [103]. It is clear now that these two assumptions are incorrect, and that progestins can be dangerous. For some time now, the role of both estrogens and progestins in normal and breast cancer cell proliferation have been investigated. Understanding the role of estrogen and progestin in breast cell proliferation is part of the large task of understanding why cHRT increases the risk of breast cancer.

Many laboratories investigate the role of hormones such as estrogen and progestin using clonal populations of cells from the breast cancer cell lines known as MCF-7 and T47D. These cell lines are used because much is known about the behaviour of cells in these cell lines, and so it is easier to separate the effects of the hormones from the intrinsic behaviour of the cells. We will discuss these two cell lines in the next section.

## 2.6 Breast cancer cell lines - MCF-7 and T47D

Cancer cell lines (which are cell populations derived from a common ancestor cell) are often used in experiments to investigate the role of hormones in the breast. If cells from different sources are used in different experiments, then it is difficult to compare results from such experiments, as the two cells may have very different behaviours. Thus, cell lines are attractive because results from experiments involving the same cell lines can be compared with reasonable confidence.

The cell lines MCF-7 and T47D are often used in cancer research. The MCF-7 breast cancer cell line originated in 1970 from a 69 year old woman with breast cancer [41]. Cells from the MCF-7 breast cancer cell line are widely used in biological experiments

because they are responsive to estrogens [117] and anti-estrogens [134], however they are less responsive to progesterone than other breast cancer cell lines [80]. There are many experiments published in the literature which investigate cell cycle changes in MCF-7 breast cancer cells in response to various growth factors [81, 134, 98, 136, 97].

The T47D cell line was originally derived from a 54 year old female cancer patient, and cells from this cell line express the progesterone receptor at all times, and are therefore responsive to hormones such as progesterone and progestins. It is particularly attractive because T47D cells do not require estrogen to express the progesterone receptor (unlike most breast cancer cells, and normal breast cells). This enables investigations into the role of progesterone in breast cell behaviour without the need to pre-expose the cells to estrogen. The T47D cell line has been used in many experiments [37, 84, 61].

Cell lines, such as the MCF-7 and T47D cell lines, are often derived from tumour samples through culturing. This process involves planting a human tumour sample obtained through surgery and allowing it to grow. After a few days, many of the cells in the culture die, but if left to grow for longer, a stable population of cells emerge. It is thought that this stable population represents the small amount of stem cells that were present in the original sample. The stable population that emerges is called a cell line.

In [28], mathematical models were used to simulate the emergence of a stable cell line population from the original culture of cells in the tumour sample. The emergence of this population from the original culture was simulated by considering cell cycle progression in two distinct populations - the slow growing population, which is the stem cell population, and the fast growing population, which makes up the bulk of the tumour mass. The emergence of the cell line was simulated, and was shown to have a different population doubling time to the original sample, which is in agreement with experimental data. This model has potential to investigate the theory that some cancer masses are driven by a small population of tumour stem cells, while the bulk of abnormal cells in the tissue are comprised of abnormal cells which have limited replicative potential. A model that can simulate this hypothesis, and compare results to the more standard theory that each of the tumour cells in a tumour mass have unlimited replicative potential, will give insight into the main differences between the two possible tumour cell populations.

Both MCF-7 and T47D cells are cells that have already gained the six hallmarks of

cancer. This means that any experimental results involving these cells can only aid in understanding how fully developed cancer cells behave, and cannot necessarily be extended to understanding how normal cells behave under the same conditions. In this thesis, we will focus on modelling experiments involving cells from these two cell lines, to get a better understanding of how different substances influence the behaviour of these breast cancer cells. Extending our results to normal cell populations, or even less developed cancer cell populations, is a larger task. We are currently considering extending our mathematical model to consider ZR75 cells, as suggested by the researchers at the Hanson Institute for Cancer Research. These cells behave much more like normal breast cells, and so results involving cells of this cell line can be more easily extended to normal cell populations.

No matter what cell lines we consider, the influence of different substances on cell behaviour is often investigated by determining how these substances influence the rate of proliferation of the cells. Thus, we wish to discuss the process of cell proliferation in detail, and we do this in the following section.

## 2.7 Cell proliferation and the cell cycle

As we have seen, cell division significantly increases the chance of developing a dangerous mutation, and hence increases the chance of cancer. This is one of the reasons that most cancers originate in the epithelium of organs - because epithelial cells proliferate at high rates in order to replenish cells lost from exposure to the harsh external environment. It has also been shown that cancers which contain cells that have high proliferation rates have a poor prognosis [121]. Thus, understanding how certain drugs or treatments influence the proliferation rates of cells can help with developing better treatments for people with cancer. In order to do this, we must understand how proliferation occurs in cells.

Once a cell commits to cell division, it goes through a series of phases (shown schematically in Figure 2.12) which eventually leads to mitosis: the splitting of a cell into two cells. The G1 phase (or Gap 1 phase) of the cell cycle is the only phase in which the cell is receptive to its surroundings. It receives constant signals from the environment and will remain in this phase until it receives enough encouragement to divide. The G1 phase is the most intensively studied phase of the cell cycle, as it is often the deregulation of

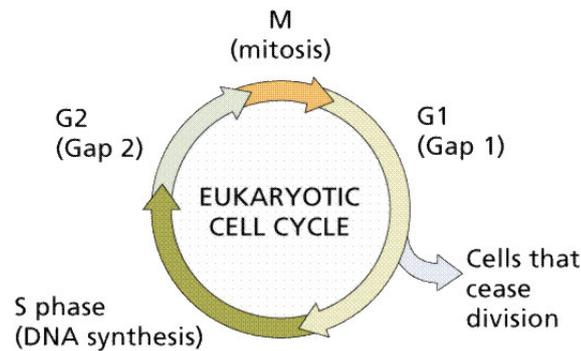


Figure 2.12: *The Cell Cycle.*

proteins produced during this phase that seems to lead to the development of cancer.

As can be seen in Figure 2.12, some cells exit the cell cycle part way through the G1 phase (‘cease division’) and become differentiated. We discussed earlier how cells become differentiated when their telomeres become too short to undergo the process.

If a cell does not become differentiated and also receives signals from its environment to divide, then it progresses irreversibly into the S phase (synthesis phase). Once the cell has decided to progress from G1 into S phase, it is no longer receptive to external signals, and has fully committed to cell division. In the S phase, a cell prepares for division by synthesising new DNA. Due to the massive amount of DNA present (about  $6 \times 10^9$  base pairs of DNA) it commonly takes 6-9 hours [135, pg. 258] to complete the S phase.

After the cell has completed the S phase, it moves into the G2 phase (Gap 2 phase), where the cell prepares itself (in currently poorly understood ways) for entry into the M phase (or mitotic phase). The G2 phase usually takes 3-5 hours to complete [135, pg. 258].

The mitotic phase is where cell division, or *mitosis*, takes place, and takes about 1 hour [135, pg. 258] to complete. All internal cell structures and DNA components will have been replicated by this time, and during this phase, the chromosomes and other cell structures will be distributed evenly between the two halves of the cell. The splitting of the cell into two daughter cells marks the completion of the mitotic phase.

As implied by the title of the process, the cell cycle is cyclic and so once a cell has completed the mitotic phase, both daughter cells return to the G1 phase of the cell cycle,

where the cycle starts again. We will now discuss how proteins are used to transition a cell from one phase to the following phase.

## 2.8 Proteins which encourage proliferation

The progression of a cell through the cell cycle phases is regulated at the level of protein expression. The proteins that regulate cell cycle progression are found within an individual cell, and almost always in the nucleus of the cell. Thus in this section, we will be discussing the changes in relevant proteins related to cell cycle progression that are found within the nucleus of any one cell.

The main proteins involved in cell cycle progression are called cyclin-dependent kinases (**CDKs**). CDKs on their own exhibit very little **activity** (i.e. they cannot interact with other proteins or cell structures to directly influence the cell), and are therefore not sufficient to encourage a cell through the cell cycle. To become active, they must bind with members of another family of proteins called **cyclins** (hence the name cyclin-dependent kinases) [135, page 263]. Each CDK will bind with one or more members of the cyclin family forming an active cyclin-CDK complex.

To regulate the progression of cells through the cell cycle, different CDKs become active as the cell progresses through different cell cycle phases. The changes in activity of the various CDKs are not due to changing concentrations of these proteins, as it has been observed that the concentration of CDKs do not change significantly throughout the cell cycle [135, pg. 263]. CDK activity is instead controlled by the changing concentration of cyclins. The concentration of CDKs is usually well in excess of the concentration of cyclins, and so the activity of a particular CDK is usually highly correlated with the concentration of the cyclins which can bind and activate it. As the concentration of cyclins changes throughout the cell cycle, different cyclin-CDK complexes become more or less active. The changes in cyclin-CDK activity as a cell progresses through the cell cycle is described below, and summarised again in Figure 2.13.

The concentration of D-type cyclins (of which there are three types, D1, D2 and D3, with D1 being the most commonly regulated and best studied of the three types) depends on the presence of external signals [98, 135]. The D-type cyclins are important proteins

in cell proliferation, and they are over-expressed in many cancers [135]. The concentration of D-type cyclins (which we will collectively refer to as cyclin D) increases if the surrounding environment contains a high concentration of growth promoting substances. Cyclin D binds to CDK4, or CDK6, and so an increase in cyclin D concentration results in an increase in the activity of cyclin D-CDK4 or cyclin D-CDK6 complexes (collectively referred to as cyclin D-CDK4/6 complexes). As cyclin D-CDK4/6 complexes increase in concentration, they encourage the cell to progress past what is known as the **Restriction point** (or R point). After this point, the cell will no longer consult its environment and will progress through the remaining phases of the cell cycle in a programmed fashion. Cells which have passed the R point are considered to be actively dividing cells. As the R point occurs about 3 hours before the cell enters the S phase [135, page 262], the actively dividing phase of the cell cycle contains not only the S, G2 and M phases but also the final hours of the G1 phase, often referred to as the ‘late G1’ phase.

Once a cell has passed through the R point, the concentration of E-type cyclins (of which there are two, E1 and E2, although we will refer to these cyclins collectively as cyclin E) increases and associates with CDK2, increasing the activity of cyclin E-CDK2 complexes. The activity of this complex allows the cell to progress from the G1 phase into the S phase.

Entry into the S phase results in a decrease in the concentration of cyclin E and almost mirroring this decrease is an increase in the concentration of A-type cyclins (of which there are two, A1 and A2, however again we will refer to these collectively as cyclin A). Cyclin A, like cyclin E, binds with CDK2. Later in the S phase, cyclin A begins to form complexes with CDK1 instead of CDK2 allowing the cell to progress from the S into the G2 phase of the cell cycle.

Towards the end of the G2 phase of the cell cycle, cyclin A is degraded and B-type cyclins (B1 and B2) increase forming active cyclin B-CDK1 complexes. These complexes trigger mitosis, after which both daughter cells begin again in the G1 phase. Upon entry into the G1 phase, the degradation of cyclin B is triggered and the cells again rely on the concentration of cyclin D to determine their progression through the remainder of the cell cycle.

As is implied by their name, cyclins are regulated in a cyclic manner, and the con-

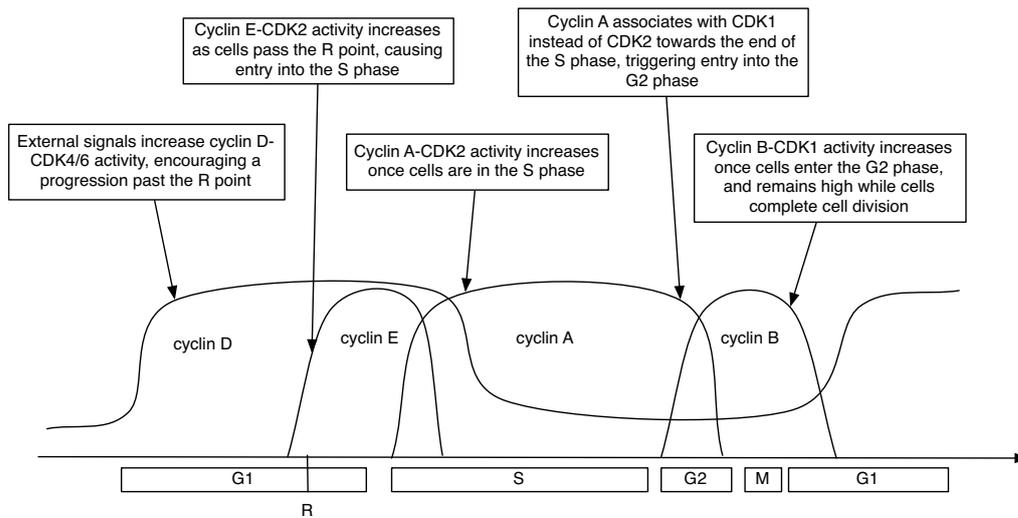


Figure 2.13: *Changes in cyclin-CDK activity as a cell progresses through the cell cycle. The cyclin which is responsible for the increase in activity during the different phases is listed below the curves.*

centration of most cyclins is determined by the phase of the cell cycle. Cyclin D concentrations, however, are regulated primarily by the presence or absence of external signals and not by the cell cycle phases, and thus these particular cyclins represent the communication link between the outside and inside of a cell. Other proteins, such as Myc, are also known to usher cells past the R point [3]. The protein Myc, although not a cyclin, is employed by many growth promoting substances (such as estrogen) to push cells past the R point. Once cyclin D-CDK4/6 or Myc molecules have ushered a cell past the R point, the changes in concentration of cyclin E, A and B change in an autonomous manner which is dependent upon the phase the cell exists in.

## 2.9 Commitment to cell division and the the R point

Cancer cells are thought to progress past the R point even in unfavourable conditions, and so understanding what kind of mechanisms regulate the R point, as well as how cancer cells alter these mechanisms, is crucial to understanding cancer cell growth.

## How the R point is regulated

The retinoblastoma protein (referred to as pRb) is a critical protein whose role is to block progression through the R point. In its active form, it has no phosphate groups attached to it (i.e. it is **unphosphorylated**, see glossary for a full description of this term), and binds to another protein, a member of the E2F family of proteins, which are inactive when bound to pRb. As cyclin D-CDK4/6 activity increases in a cell, it phosphorylates pRb, but not on all possible phosphorylation sites that the pRb protein has. This partially phosphorylated state is referred to as **hypophosphorylation**, and is not sufficient to allow cells to pass the R point. Instead, cyclin E-CDK2 activity begins to rise at this point, and it is these molecules which are able to phosphorylate pRb on the remaining unphosphorylated sites, resulting in **hyperphosphorylated** pRb. The Myc protein also increases transcription of cyclin E-CDK2, encouraging cells through the R point via the phosphorylation of the pRb protein.

The hyperphosphorylated pRb (which remains hyperphosphorylated until mitosis has occurred) dissociates from the E2F molecules, which then become active. These active E2F molecules increase the rate of production of cyclin E. The increase in concentration of cyclin E results in an increase in concentration of active cyclin E-CDK2 complexes, which phosphorylate pRb, releasing even more E2F. This sort of process, whereby ‘protein A increases protein B which increases protein A...’ is known as a **positive feedback loop**.

The increase in cyclin E-CDK2 activity (as well as the increase in active E2F molecules) results in a cascade of other proteins designed to shuttle the cell through the R point and, consequently, through the remainder of cell cycle. There are other checkpoints at later stages of the cell cycle which may halt cell cycle progression via the up-regulation of the p53 protein, due to DNA damage either from DNA replication or from carcinogens. However, the vast majority of cells will progress through the later checkpoints without halting the cell division process. This means that cells that have passed the R point have almost certainly committed themselves to undergoing cell division. The decision to progress past the R point is found to be deregulated in most, if not all, human cancers [135, page 262].

Since the later phases (S, G2 and M phases) occur similarly in both normal and cancer

cells, many mathematical models have focussed on the dynamics which allow a cell to pass the R point [99, 88, 114, 141]. In these models, regulation of a cell's decision to progress past the R point was modelled as a complex set of interactions between different proteins. Other models have also investigated how different proteins interact to help cells progress through the later autonomous cell cycle phases [119, 120, 35].

### Inhibitors of cell cycle progression

With knowledge of how the various proteins change in concentration, it might seem that we can determine what phase a cell is in by investigating the concentration of various proteins present within that cell. However, inhibitors within a cell can de-activate cyclin-CDK complexes. So the total concentration of a particular cyclin is not always correlated with the activity of that cyclin-CDK complex.

There are different families of inhibitors which inhibit cell cycle progression by directly targeting cyclins, CDKs or cyclin-CDK complexes. However we will discuss just two inhibitors: p21<sup>Cip1</sup> and p27<sup>Kip1</sup>. The proteins p21<sup>Cip1</sup> and p27<sup>Kip1</sup>, which we will refer to simply as p21 and p27, exert their inhibitory effects on proliferation by binding to otherwise active cyclin-CDK complexes. When p21 or p27 are bound to the cyclin-CDK complexes, the cyclin-CDK complexes are no longer active and cannot exert their proliferative effects on the cell. The exception to this is cyclin D-CDK4/6 complexes. When cyclin D-CDK4/6 are bound to either p21 or p27, the complex remains active, and is still able to exert its proliferative effects on the cell. This also means that total cyclin D concentration is usually a good indicator of overall CDK4/6 activity.

The differences between p21 and p27 include the rate at which they bind to different cyclin-CDK complexes. For instance, p21 preferentially binds to cyclin D-CDK4/6 complexes over cyclin E-CDK2 complexes (i.e. p21 has a higher **affinity** for cyclin D-CDK4/6 than for cyclin E-CDK2 ) whereas p27 preferentially binds to cyclin E-CDK2 over cyclin D-CDK4/6 complexes [40]. The proteins p21 and p27 are also regulated in a different manner: p27 levels in a cell are altered by changing the rate of degradation of the protein [74, 36], whereas p21 concentrations tend to be regulated by changing the rate of protein production, or protein **transcription** [12]. The mechanism of p27 degradation will be discussed in detail later in Section 8.2.

While p21 and p27 are thought of as potent inhibitors of cyclin-CDK activity in general, their inability to inhibit cyclin D-CDK complexes actually results in a paradoxical effect. Normally, cyclin D-CDK4/6 complexes have a high rate of decay, with a half-life of 30 minutes [135]. However, when bound to either p21 or p27, the complex becomes more stable, exhibiting a longer half-life within the cell.

This mechanism of cyclin D-CDK4/6 complexes binding to p21 and p27 sheds light on yet another mechanism by which cyclin D-CDK4/6 complexes can encourage cell cycle progression. As discussed earlier, cyclin D-CDK4/6 is able to directly aid in a cell's progression through the R point by phosphorylating pRb. However, an increasing presence of cyclin D-CDK4/6 will also result in an increasing proportion of the p21 and p27 inhibitors bound to cyclin D-CDK4/6 and so less of these inhibitors are bound to other cyclin-CDK complexes such as cyclin E-CDK2. Thus increasing concentrations of cyclin D-CDK4/6 will not only increase the rate of hypophosphorylation of pRb, but will increase the amount of active cyclin E-CDK2 within the cells, accelerating the rate of pRb hyperphosphorylation. The complex interaction between the p21/p27 inhibitors, the cyclin D-CDK4/6 molecules and other cyclin-CDK molecules will be investigated in more detail in Chapters 7 and 8.

In this thesis, we will combine the tools of mathematical modelling with the biological understandings discussed in this chapter in order to gain an even better understanding of cell cycle progression in breast cancer cells. As we are mainly focussing on cells that make up breast cancer cell lines (such as MCF-7 or T47D cells), we will assume that our cells have already overcome the hallmarks of cancer described in this chapter. In particular, we do not need to keep track of the number of cell divisions that a cell has undergone, because each cell has unlimited replicative potential. The models we develop are done so with careful recognition of the current biological understandings of cell proliferation, including the influence of the changing concentrations of cyclins and the mechanics of progression past the the R point. In addition, depending on the cell lines we are considering, we keep in mind the sensitivities of the cells to various substances (for instance, MCF-7 cells will respond to estrogen and not progesterone, whereas T47D cells are significantly more sensitive to progesterone). In the next chapter, we discuss how such a model can be developed.

## Chapter 3

# Mathematical modelling of cell cycle progression

Different mathematical modelling approaches have been taken in order to understand the process of cell cycle progression. Most models take into account the strengths and limitations of experimental biology techniques that are often used to detect cell cycle progression. Thus, before we discuss any mathematical models in detail, we must first understand how cell behaviour is investigated in the laboratory.

### 3.1 Flow Cytometry: detecting cell cycle phases

In order to understand how certain drugs or hormones influence the proliferation rates of breast cancer cell lines, biologists could simply determine how the total number of cells changes in the sample over time. However, there is a significant lag time between when a cell makes the commitment to cell division, and when it actually undergoes division. A more useful method, which is commonly used in laboratories, involves investigating the amount of DNA within a cell. Recall that G<sub>0</sub>/G<sub>1</sub> phase cells will have only one copy of DNA. S phase cells are in the process of DNA replication, and so the amount of DNA content in S phase cells will vary from one to two copies of DNA. The G<sub>2</sub> and M phase cells will each have two copies of DNA. By investigating the amount of DNA within a cell, biologists can accurately determine the phase that each cell is in (however, one

cannot distinguish between G2 and M phase cells). Often in experiments which contain many cells, results are reported as an overall proportion of cells that exist in each of the above phases. This information is determined experimentally by a process called **flow cytometry**.

In biological experiments, a flow cytometer is commonly used to determine many different characteristics of cells including cell size, survival status (i.e. whether a cell has begun to break down its internal constituents for preparation of cell death) and DNA content. These characteristics are determined by passing a light beam through the cells in a sample. The light is refracted at all angles from the cell, and by investigating the intensity of light at different angles, much information can be inferred.

### **Flow cytometry and fluorescent DNA detection**

To determine the amount of internal DNA content of cells, the cells are first exposed to a solution containing a fluorescent antibody. These antibodies attach to DNA strands, and so the higher the DNA content within a cell, the higher the concentration of DNA-bound antibodies. As the cells pass through the flow cytometer, a light beam is passed through each cell. This light beam causes the antibodies that are attached to the DNA strand to become excited and emit light at a particular wavelength. Thus, cells with higher DNA content will emit a higher intensity of light at that wavelength, and so light intensity is used to determine the internal DNA content. This process is called **fluorescence detection**.

For flow cytometry to be effective, it is critical that cells pass through the laser beam one at a time (otherwise the intensity of the emitted light will be artificially high if more than one cell passes through the beam at any time). Thus, a sample of cells must somehow be separated to allow for individual analysis. This is achieved by injecting the cell sample into a fast moving stream of fluid (see Figure 3.1). Injecting the cell sample into the high speed fluid stream creates a laminar flow, resulting in a single file progression of cells through the narrowest part of the funnel.

The cells, now moving in single file, pass through a light beam. When the light beam interacts with each cell, the fluorescent antibodies that are DNA-bound become

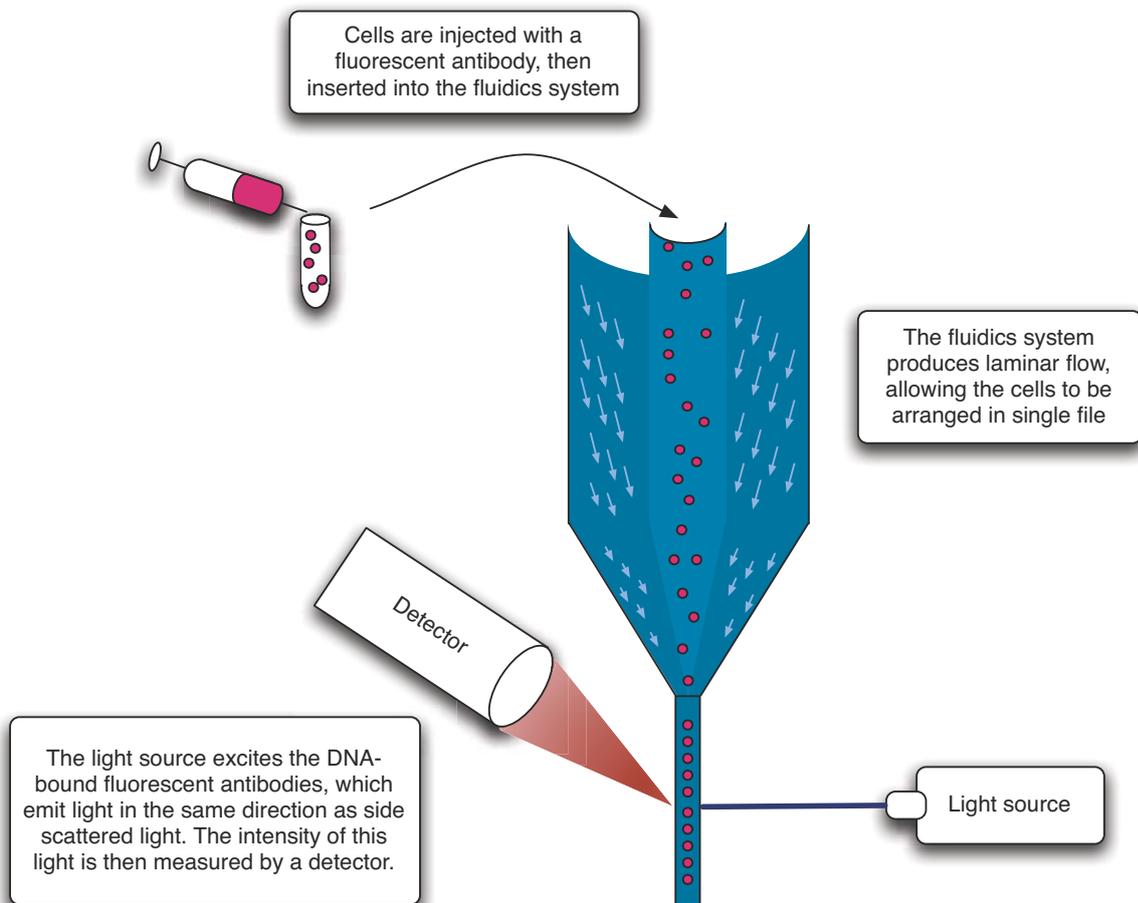


Figure 3.1: *Fluorescence detection using flow cytometry. The cell sample is injected into a funnel-like device containing fast flowing liquid, resulting in laminar flow of the sample. The cells then pass through the light source in single file, and the light source excites the DNA-bound antibodies. The excited antibodies emit light at a particular wavelength, and the intensity of the wavelength of light emitted is then measured using a detector.*

NOTE:  
This figure is included on page 59 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 3.2: *DNA histogram representing the number of cells (y-axis) containing various amount of DNA content (x-axis, called ‘channel number’ here). The smaller channel numbers (50 - 75) represents G0/G1 cells. Channel numbers 75 - 115 represent S phase cells. Higher channel numbers (115 - 130) represent G2/M phase cells. This figure was obtained from [78], where it was part of a larger series of flow diagrams, and so some peripheral features have been removed from the original figure.*

excited and emit light. The intensity of the specific wavelength of light emitted can then be measured using a detector, and analysed using computer software. A typical representation of DNA content from a flow cytometer, called a **DNA histogram**, is shown in Figure 3.2. The smaller channel numbers (50 - 75) represent cells containing one copy of DNA, and hence these cells must be G0/G1 cells. Higher channel numbers (115 - 130) represent cells containing two copies of DNA, and hence represent G2/M phase cells. Cells which are in the process of DNA replication (S phase cells) will contain an intermediate amount of DNA content, and are shown here in channels 75 - 115.

In theory, we would expect all G0/G1 and G2/M phase cells to be observed with exactly the same DNA content. However from Figure 3.2, we observe that there is a range of channel numbers that we denote as G0/G1 phase cells, and another range denoted G2/M phase cells. The variability in the channel numbers represents noise due to imperfect experimental procedures and data gathering techniques.

The representation in Figure 3.2 must be converted into a total number of cells in each of the phases. In Figure 3.2, it appears that there is a clear distinction between where the G0/G1 phase curve ends and where the S phase curve begins. Similarly, there

is a clear distinction between the G2/M phase data and the S phase data. However, in many cases there is significant overlap, and it is not easy to discern an accurate range of channel numbers that represent specific phases. This is overcome by assuming that cells with the same DNA content will be normally distributed about a given channel number. Thus, the amount of cells in the G phase can be determined by fitting Gaussian (normal) distributions to the data. Similarly, we assume that a Gaussian distribution is a good representation of the number of G2/M phase cells. A flat line is often fitted to the S phase data, or sometimes a polynomial curve (the S phase does not always appear as flat as that shown in Figure 3.2, and a polynomial representation may produce more accurate results). The resultant curves are then fitted to the data and used to quantitate the amount of cells in the three phases.

Figure 3.2 shows the distribution of cells across the three phases for only one point in time. In biological experiments, the distribution of cells across the phases is often investigated at regular time intervals after exposure to some substance. Thus, data like that in Figure 3.2 is determined many times using the flow cytometer during the course of an experiment.

The accuracy of this technique for determining the cell cycle phase distribution is generally quite good. This can be seen by considering results where the proportion of cells in each of the cell cycle phases is constant. Later we will consider a range of experiments which show flow cytometry results during instances when the proportion of cells in each of the cell cycle phases is constant. We find by observing this experimental data, that the error on flow cytometry is less than  $\pm 4\%$ .

### **Flow cytometry and detecting cell death**

Some substances can cause apoptosis in cells which are exposed to them [115], and it is important that flow cytometry techniques can successfully identify which cells are apoptotic. There are several techniques which can be used to detect the presence of apoptosis in a cell population. However as we will see, some techniques cannot accurately detect dead cells. We discuss some of these techniques here. Later in Section 5.9, we encounter an experiment which counts dead cells as if they were live cells, and in that section we find how miscounting cells can distort the results from flow cytometry experiments.

An apoptotic cell displays certain morphological features. For instance, the cell loses its ability to remain bound to its neighbouring environment, DNA strands begin to get broken down, and the cell expresses surface signals which advertise the process of apoptosis. Macrophages and other immune cells recognise these cell-surface antigens, and react by engulfing the apoptotic cell in a process called **phagocytosis**. In this way, dying cells are removed so that they don't compromise the health of nearby undamaged cells. Cells that die in experiments often float to the top of the petri dish once they lose the ability to remain connected to the medium they are growing in, and can be easily washed off. However, losing the ability to remain bound to the medium of growth is something that occurs later in the death process, thus it is possible that cells are being counted as live cells while they retain the ability to remain bound to the surrounding medium. For most of the experiments we consider in this thesis, no special attempt is made by experimentalists to determine whether cells in their sample are apoptotic or not, except to regularly wash away the dead layer of cells at the surface.

We wish to investigate whether simple flow cytometry techniques, which detect changes in DNA content, will be accurate in detecting dead cells in a sample. As DNA strands are broken down (and sometimes excreted from the cell) during apoptosis, apoptotic cells may appear to have less DNA than their live counterparts. If apoptosis occurs primarily during the G phase of the cell cycle, DNA flow cytometry may detect another peak further to the left of the G<sub>0</sub>/G<sub>1</sub> peak, which can represent the presence of cells whose DNA is being broken down due to apoptosis. If apoptosis occurs in other phases, such as the S or G<sub>2</sub>/M phases, apoptotic cells may appear simply as cells in an earlier phase. Thus, measurement of DNA content alone may produce unreliable results if apoptosis occurs in the later phases. However in [27], it was found that experimental techniques could be used to cause more DNA excretion in apoptotic cells. These treated cells will have a peak on the DNA histogram even further to the left, and therefore have minimal overlap with non apoptotic cells. An example of a DNA histogram of a population containing apoptotic cells is shown in Figure 3.3. Unfortunately, experiments have shown that it takes at least 12 hours [93] - 24 hours [48] to observe changes in DNA content due to apoptosis, meaning that many dead cells will appear as live cells when flow cytometry is performed.

Therefore, other techniques must be used to detect changes in cell morphology ear-

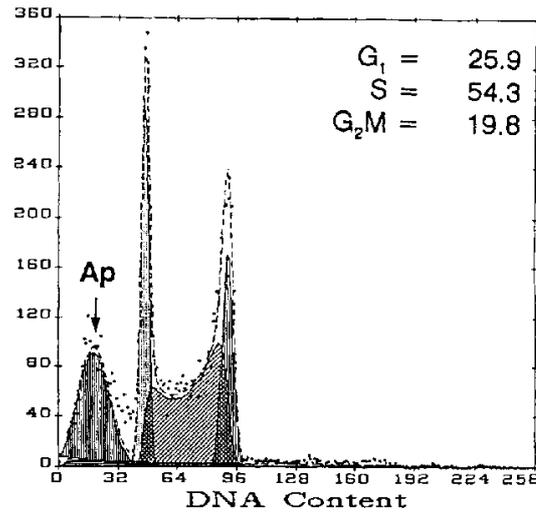


Figure 3.3: DNA histogram of a cell population containing apoptotic cells. The apoptotic cell population (labelled *Ap* here) have less DNA content than cells in other phases of the cell cycle. Figure obtained from [27].

lier. One of these early morphological changes is the movement of a protein called phosphatidylserine (a macrophage signalling protein) towards the outside of the cell membrane, where it can be recognised by phagocytes. Investigations have shown that this physical change in the cell can be detected easily in the laboratory. In addition, the externalisation of phosphatidylserine occurs early on in apoptosis, before any changes in DNA structure, and therefore this method may more accurately detect the initial stages of apoptosis which would otherwise be overlooked in DNA flow cytometry [58]. Unfortunately, this method has not been considered in the experiments we deal with in this thesis.

We will now conclude our discussion on apoptotic cells and consider some mathematical approaches from the literature which investigate the process of cell cycle progression.

### 3.2 Previous mathematical models of cell cycle progression

Various mathematical approaches have been used to understand the progression of cells through the cell cycle phases. Some models have considered the change in the amount of DNA as cells progress through the cell cycle so that their results can be directly compared to experimental data on DNA histograms [6, 7, 8]. Others have focussed on modelling the change in internal cell cycle proteins (such as cyclins E and A) as cells progress from

one phase to another [89, 88, 141, 11], while others have considered simpler approaches which bypass the effects of internal proteins and internal DNA concentrations altogether [28]. We will discuss some of the different approaches here.

### Protein models of cell cycle progression

As the progression past the R point is deregulated in cancer cells, many mathematical models investigate the interactions between proteins that control progression past this point [89, 88, 141, 11]. The interaction between various proteins, such as cyclin E and cyclin D are considered in many of these models. In [141], a mathematical model was built that considered the effects of cyclins D, E, myc and E2F on a cells' decision to pass the R point. For certain parameter values, they found that if cyclin D and myc reached a particular threshold, E2F would increase significantly in concentration, causing the cell to progress through the remaining cell cycle phases. They also found that if cyclin D and myc concentrations decreased again, E2F would remain high, suggesting two possible stable states in the concentration of this protein (i.e. stable states of both high and low E2F concentrations existed). This result provides insight into the possible mechanisms a cell uses to allow cells to progress through the later cell cycle phases, even when concentrations of cyclin D and myc drop. In [88], a large protein model was developed that included the effects of inhibitor proteins such as p21 and p27 on cyclin-CDK activity. They also considered cell cycle proteins in the other phases (such as cyclins A and B, which rise during the S and G2 phases). This results in a large model of 22 differential equations, and an even larger set of parameters that must be determined in order to model this process. Novak et. al. [88] were able to model the changes in cyclin concentrations, and their results compared well with experiments.

Mathematical models of cell cycle proteins can be tricky because, as we saw earlier, activity of a cyclin-CDK unit depends on whether they are bound to the inhibitors p21 and p27 or not, and determining this experimentally is difficult. In addition, the location of proteins within a cell is critical, as these proteins only exert their effects when in the nucleus of the cell, and this again is information that cannot be obtained easily using traditional experimental techniques. Later in Chapter 8, we consider a model of protein-protein interactions. We will see that even with a simplified model and a good set of data

on the concentration of proteins and their corresponding association with the p21 and p27 inhibitors, the choice of parameter values can significantly alter the behaviour of such a model.

### **A comment on simple models vs large complex models**

Mathematical models can be useful when trying to understand underlying mechanisms. When modelling processes which have many variables influencing its behaviour, such as natural biological processes, we realise that there are many more factors at play than can be realistically implemented into a mathematical model. Every modeller must make a decision about which mechanisms to include in the model and which to exclude.

The attraction to building large, detailed mathematical models is that such models are more realistic, as many relevant mechanisms known to influence the process are considered. The mathematical model presented in [88] involves considering the changes in concentration of many different cell cycle proteins, such as cyclins D, E, A and B, and many other cell-cycle related proteins, resulting in 22 differential equations with many more unknown rate parameters. These large, detailed models are perhaps easier to ‘sell’ to a biological audience because of this level of realism. However, the level of complexity inherent in the large size of the system poses many problems. Determining accurate rate parameters is not possible, given the large number of parameters used in such a system, and a study of parameter sensitivity when there are so many parameters is also a very difficult task. Parameter values are often chosen carefully so that the model exhibits realistic behaviour. This detracts from the purpose of the model, which is to investigate the causes of the mechanisms that drive the process, not to just simply replicate them. In addition, such models are not malleable, and any further modifications will require re-parameterisation just so the model produces realistic results.

It is often the case that simple models, which contain just the essential mechanisms of the process of interest, can often provide surprising insight into the process under investigation. With simple models, it is easier to see what mechanisms are driving model behaviour, and so the models can feed back directly into biological understandings of the process. Simple models are also highly malleable. For example, a simple cell cycle model which uses differential equations to represent the progression of cells through the four cell

cycle phases has been used in the literature many times, often with small modifications so that the authors can focus on a particular problem e.g. cells becoming senescent (by including a G0 phase) or cell death due to chemotherapeutic drugs. However, large complex models such as that in [88] cannot be easily used by other researchers, and even those who constructed the model may find it difficult to deconstruct and modify if other proteins of interest are to be implemented. We suggest that simple models, including only the key mechanisms, are overall better to employ, even though we recognise this may be counter-intuitive.

### Age structured models

Age structured models such as those described in [107] and [9] model the progression of cells through the cell cycle phases using differential equations, however such models do not explicitly consider the influence of proteins in the rates of progression. Consequently, such models involve less variables than the protein model counterparts described previously. Age structured models keep track of the time a cell has spent in any one phase as well as the time since the beginning of the experiment, resulting in a set of partial differential equations in two dimensions. In these models of the cell cycle, the number of cells that have been in phase  $P$  for  $\tau_P$  hours at time  $t$  hours after the beginning of the experiment is defined as  $\mathbf{n}_P(\tau_P, t)$ . The rates at which cells progress from one phase to the next can depend upon how long they have spent in that phase (i.e. depend upon the value of  $\tau_P$ ). For instance, the rate at which cells leave the S phase cells is represented in [9] as

$$\frac{\partial \mathbf{n}_S(\tau_S, t)}{\partial t} + \frac{\partial \mathbf{n}_S(\tau_S, t)}{\partial \tau_S} = -\alpha(\tau_S, t) \mathbf{n}_S(\tau_S, t).$$

In this representation, the rate at which cells leave the S phase depends upon both time  $t$  and the age of cells in the S phase,  $\tau_S$ . The dependence on the variable  $t$  is required since the rate of progression between phases can change after the addition of a chemotherapy drug. It is useful to be able to keep track of the time a cell has spent in the S phase, because the time a cell has spent in this phase will correlate with the amount of DNA it has been able to replicate, and so results from these models can be compared directly with DNA histograms. In [6], [7] and [8], the authors also included a diffusion term to represent the error in the accuracy of flow cytometry measurement of S phase DNA content. In

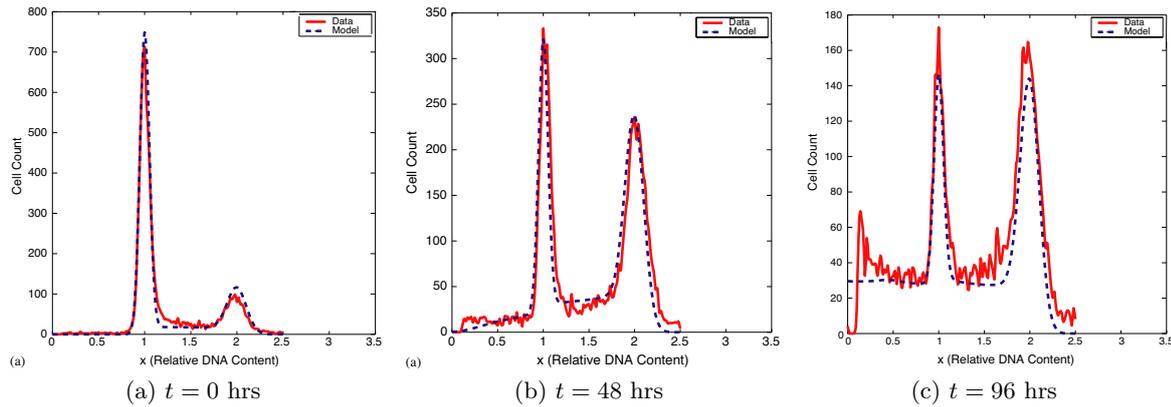


Figure 3.4: Model results after paclitaxel exposure. Figure obtained from [8].

[8], they used their model to simulate the effects of the anti-cancer drug paclitaxel, which causes arrest in the M phase of the cell cycle, and also causes cells to undergo apoptosis. Their model results are reproduced here in Figure 3.4 (model results are the blue dotted curves). These representations show that their model results compare very well with experiments. In Figure 3.4(b) and (c), we can see that they have been able to capture the presence of apoptotic cells, which have DNA content less than one.

In [107], an age-structured approach was also used to model cell cycle progression. The model considers cells becoming senescent, moving from the G1 into the G0 phase (see Figure 2.12). As we saw earlier, one of the critical hallmarks of cancer is cells gaining unlimited replicative potential, and hence in most cancers, cells do not enter quiescence. However, we also mentioned that in some cancers, there is a small population of stem cells which have unlimited replicative potential, and when they divide, their progeny, which may make up the bulk of observed tumour cells, could actually have limited potential. This is known to be true of leukaemias, which are driven by a population of stem cells in the bone marrow, and the abnormal cells circulating the blood have limited replicative potential. The model in [107] was applied to the growth of leukaemia cells in the blood. Cell death due to the presence of chemotherapy drugs was also considered.

The theory of the presence of a stem cell population was also considered in [28], and mathematical models were again used to simulate progression through the cell cycle phases. The equations used in [28] did not consider how the rates of transition between various cell cycle phases depends upon the age of cells in each of the phases. For instance,

in [28], the rate of change in the number of S phase cells was represented as

$$\frac{dN_S(t)}{dt} = r_{G1 \rightarrow S} N_{G1}(t) - r_{S \rightarrow G2} N_S(t),$$

where  $r_{G1 \rightarrow S}$  and  $r_{S \rightarrow G2}$  represent the rate of transition from the G1 into the S phase and the S phase into the G2 phase respectively, and  $N_{G1}(t)$  and  $N_S(t)$  represent the number of G1 and S phase cells respectively. In this representation, we note that the rate at which cells leave the S phase is constant, and so the rate at which cells leave their current phase is independent of how long they have already spent there. However, we find that this is not very realistic, and that age structured models are biologically more accurate because the rate at which cells leave a phase is dependent upon how long a cell has spent in that phase. In particular, authors of [107] and [10] found that cells have a higher rate of leaving a phase if they have spent a long time in that phase, and have a significantly lower rate if they have recently entered it. We also find that MCF-7 and T47D cells have a higher rate of leaving a phase the longer they have spent in that phase, however we will use different techniques to deal with this feature of cell cycle progression. We will discuss our model of cell cycle progression in the next section.

The age-structured models that are developed in these papers are very general, and the rate of transition between any two phases can be any function of time  $t$  and the age of cells in a phase,  $\tau_P$ . Thus, such models can be applied with great flexibility to many different conditions. In [10], analysis of the existence and uniqueness of solutions to cell cycle phase models using age-structured approaches was performed. It was shown that if the rates of transition between phases depends only upon how long cells spend in that phase (i.e. the rates of transition depend upon  $\tau_P$  and not on  $t$ ) then there exists a unique solution to the set of equations describing the transition rates. Although these models are very general, they can also be more difficult to analyse, particularly if the rate at which cells leave a phase is a complicated function depending upon the age distribution of cells in that phase. In the next section, we will introduce an new model of cell cycle progression.

### 3.3 Our model of cell cycle progression

We begin by developing a simple model, which describes the progression of a clonal population of cells through the phases of the cell cycle. We represent the number of cells in the G1 and S phases as  $N_G$  and  $N_S$  respectively. The number of cells in the G2 and mitotic phases are often reported together in biological experiments ([134, 81, 98, 136, 97]). Mathematically, we could model the G2 and mitotic phases separately and then add them together for comparison with biological results. However, once the cell has exited the G1 phase of the cell cycle, it will continue through the remainder of the cell cycle via a programmed series of steps which are independent of environmental signals [135, pg. 262]. This allows us to group the G2 and mitotic phases together as one term in our mathematical model without changing the dynamics of the model. Thus for simplicity we define  $N_M$  to represent the number of cells in the G2 and mitotic phases (From now on, the G2 and mitotic phase will collectively be referred to as the M phase). by the same argument, we could lump the S phase in with this term without changing any model dynamics, however we will consider the S phase cells as a separate population as biological experiments report these cells separately.

The influence of the external environment on the progression of the cells through the cell cycle is represented by the function  $\gamma(t)$ . It is assumed that environmental signals can only influence cells in the G1 phase of the cell cycle (for simplicity, we will refer to the G1 phase from now on as the G phase). In Section 5.3 we will describe how we convert the influence of environmental signals, such as the presence of growth factors or hormones, into the mathematical expression  $\gamma(t)$ .

The number of cells in a cell cycle phase at any given time must be an integer. However, when investigating the population after some change to its environment, it is not just the number of cells in a phase at some later time point that is of interest, but the number of cells in that phase *relative* to the starting number of cells in the sample. Mathematically, this would involve considering the change in  $\frac{N_S(t)}{N_T(0)}$  instead of the change in  $N_S(t)$ , where  $N_T(0)$  represents the total number of cells in the sample at  $t = 0$ . If we do this, then  $\frac{N_S(t)}{N_T(0)}$  is not necessarily an integer, but is a rational number. However, if  $N_T(0)$  is large, then  $\frac{N_S(t)}{N_T(0)}$  can be approximated reasonably accurately by a continuum. Thus, we will

assume that the number of cells in any phase is sufficiently large so we can represent the rate of change in  $\frac{N_S(t)}{N_T(0)}$  using continuous differential equations.

Consider the rate of change in the number of cells in the S phase of the cell cycle, which is represented in equation (3.1) by  $\frac{dN_S(t)}{dt}$ , where  $\frac{d}{dt}$  is the rate of change with respect to time. The first term on the RHS of equation (3.1) represents the rate at which the G phase cells enter the S phase due to the influence of environmental signals,  $\gamma(t)$ . The second term on the RHS of (3.1) represents cells leaving the S phase once they have completed DNA synthesis, which we assume occurs at a rate  $\alpha$ . Since the rate at which cells leave the S phase is generally considered to be the same for cells of the same cell line, and is independent of environmental signals ([112, 78, 135]),  $\alpha$  is assumed to be constant. Thus,

$$\frac{dN_S(t)}{dt} = \gamma(t)N_G(t) - \alpha N_S(t). \quad (3.1)$$

Now consider the rate of change in the number of cells in the M phase of the cell cycle, represented in equation (3.2). As cells leave the S phase, they enter the M phase, and so the term representing the loss of the S phase cells in equation (3.1) appears on the RHS of (3.2) as the growth of the number of cells in the M phase. The second term on the RHS of (3.2) represents cells leaving the M phase once they have completed mitosis, and occurs at the rate  $\beta$ . It is generally accepted that the rate of leaving the M phase for a cell of the same cell line is constant and independent of environmental signals ([112, 78, 135]), and so  $\beta$  is constant. Thus,

$$\frac{dN_M(t)}{dt} = \alpha N_S(t) - \beta N_M(t). \quad (3.2)$$

The rate of change in the number of G cells is represented in equation (3.3). As cells leave the M phase in equation (3.2), they split into two G phase cells, and hence the term representing the growth of G cells in equation (3.1) is double the term representing the rate of cells leaving the M phase in equation (3.3). Again, the term representing the loss of the G phase cells in (3.3) appears again in (3.1) as the growth of S phase cells, and so we have

$$\frac{dN_G(t)}{dt} = 2\beta N_M(t) - \gamma(t)N_G(t). \quad (3.3)$$

We do not consider a constant rate of cell death in our mathematical model. Biological experiments which investigate cell cycle progression often expose their cells to ample growth factors and nutrients, and as a result, the rate of cell death is assumed to be insignificant, although cell death in the presence of a killing agent will be investigated later in Section 5.9.

The system of equations presented in (3.1)-(3.3) is quite general, and can be used to model the progression of any population of cells through the phases of the cell cycle, provided the functional form of  $\gamma(t)$  is known for the population of cells being studied. For a clonal population of cells, the rates of progression from one phase to the next phase should be the same, and any environmental conditions (such as the presence of growth factors) can be represented in the function  $\gamma(t)$ , as these external factors only influence cells that are in the G phase of the cell cycle.

### The inclusion of the storage phases

The mathematical model presented in equations (3.1) - (3.3) assumes that once a cell enters a phase, it is immediately available to transition into the following phase at a fixed rate. Biological experiments, however, have shown that this is not the case. For one thing, when a cell makes the decision to enter the S phase, it is then forced to spend a few hours in ‘late G1’ whereby the cell up-regulates certain proteins which will help it to start the process of DNA replication. Similarly, once a cell enters the S phase, it must spend a minimum amount of time synthesising the large amount of DNA present in the nucleus, and for that time is forbidden to progress onto the M phase.

There are two more instances where cells are not immediately available to transition into the following phase. We assume that once a cell enters the M phase, it must spend a mandatory minimum amount of time there as the internal cell constituents align properly in preparation for mitosis. In addition, we will also assume that cells which have completed mitosis and entered the G phase are not immediately receptive to environmental signals, and must spend some time recuperating and re-attaching to their surroundings after cell

division before they can process the signals in their environment.

It has also been noted in other mathematical models of cell cycle progression that the rate of progression from one phase to the subsequent phase is not accurately captured using a constant rate, as shown in equations (3.1)-(3.3). Stochastic models were developed in [64] to quantify the variance of the S phase duration in MCF-7 cells and in [63] to estimate the variance of the G2 phase in MCF-7 cells, and in these papers, it was proposed that cells entering the S phase must spend a fixed minimum amount of time in this phase before being able to progress into subsequent phases. It was similarly noted in all of [6, 7, 8] that the rate of cells leaving the S phase should increase the longer they have been in this phase.

We choose to deal with the observation that cells are not immediately available to transition into subsequent cell cycle phases by breaking up the G, S and M phases further into multiple *model phases*. The model phases are characterised by one of two types of phases - one in which cells must spend a mandatory amount of time, called a *storage phase*, and another during which cells leave at some rate, called a *non-storage phase*, which is included to capture variability in cell cycle progression times. We will assume that the G phase can be represented by three model phases. Let  $N_{G_a}$  represent the number of G phase cells which are recuperating after having just completed mitosis (we will call this phase the  $G_a$  model phase). We assume that a cell will spend a mandatory  $\tau_{G_a}$  hours in this phase after mitosis before the cell immediately moves onto the next phase, in which the cell is receptive to environmental signals. We will call this receptive phase the  $G_b$  model phase and the number of cells in this model phase will be represented by  $N_{G_b}$ . Cells in this phase are immediately available to move onto the next phase at a rate  $\gamma(t)$ . Once a cell has decided to progress further through the cell cycle, it will move into the following phase, which we will call the  $G_c$  model phase. The number of cells in the  $G_c$  phase is represented by  $N_{G_c}$ , and it is assumed that a cell will spend  $\tau_{G_c}$  hours here until it immediately transitions into the S phase. Thus the G phase is split into two storage phases -  $G_a$  and  $G_c$ , and the non-storage phase,  $G_b$ . The  $G_b$  model phase is the only model phase in which cells are receptive to external signals.

Once a cell has completed the  $G_c$  model phase it will move into the S phase. We will assume that once a cell enters the S phase, it must spend a minimum amount of time here

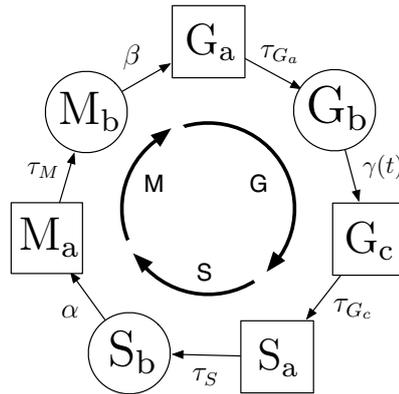


Figure 3.5: *Cell cycle progression through the seven model phases. Each model phase is associated with an actual cell cycle phase (inner cyclic arrows). Model phases in square boxes represent storage phases and model phases in circles represent non-storage phases. The parameters on arrows from storage phases represent the fixed time that a cell spends in that model phase, while the parameters on arrows from a non-storage phase represent the rate of transition from that phase.*

synthesising new DNA, and so we start the S phase with a storage stage which we call the  $S_a$  model phase. Let the number of cells in this model phase be  $N_{S_a}$  and let the duration of this storage phase be  $\tau_S$ . Once a cell has spent  $\tau_S$  hours here it immediately moves into the following non-storage phase which we call the  $S_b$  model phase. The number of cells in this phase will be represented by  $N_{S_b}$  and it is assumed that cells in this phase are available to transition into the M phase at a constant rate  $\alpha$ .

After leaving the S phase, the cell will enter the M phase. As with the S phase, the cell must spend a minimum amount of time here preparing for cell division, and so we begin the M phase with a storage phase which we call the  $M_a$  model phase. Let the number of cells in this phase be represented by  $N_{M_a}$  and let the duration of this storage phase be  $\tau_M$ . Once the cell has spent  $\tau_M$  hours here it immediately moves into the  $M_b$  model phase, which is a non-storage phase. The number of cells in this phase is represented by  $N_{M_b}$  and cells in this phase will complete mitosis and re-enter the  $G_a$  phase at a rate  $\beta$ . Thus, the G, S and M phases are all broken into storage and non-storage phases. We note that for the S phase, the storage phase concept captures the fact that DNA replication takes a long time, however, the storage phases in the G and M phases are due to other factors. The progression of cells through the seven model phases is shown schematically in Figure 3.5.

Note that although we are using the same symbols as we did in equations (3.1) - (3.3) for our rate coefficients of the transition between our cell cycle phases ( $\gamma(t)$ ,  $\alpha$  and  $\beta$ ), the values will be different from the ones presented in equations (3.1) - (3.3) due to the inclusion of the storage phases. Different algebraic symbols could have been introduced, however we choose to re-use the rate symbols for simplicity.

Let us consider the rate of change in the number of  $G_a$  phase cells, which can be expressed as

$$\frac{dN_{G_a}(t)}{dt} = 2\beta N_{M_b}(t) - 2\beta N_{M_b}(t - \tau_{G_a}). \quad (3.4)$$

The first term on the RHS of equation (3.4) represents the  $M_b$  phase cells completing the  $M_b$  phase, splitting in two, and entering the  $G_a$  phase. Cells leave the  $M_b$  model phase at rate  $\beta N_{M_b}(t)$  and so the rate at which cells enter the  $G_a$  model phase is  $2\beta N_{M_b}(t)$ . This growth term is similar to the first term on the RHS of equation (3.3) except now we only allow  $M_b$  phase cells to transition into the  $G_a$  phase, rather than allowing any cell in the  $M$  phase to transition into the  $G_a$  phase.

The  $G_a$  phase is a storage phase. This means that cells do not leave this phase at a constant rate; rather, once they enter this model phase, they remain here for  $\tau_{G_a}$  hours and then immediately progress onto the next model phase, the  $G_b$  phase. Thus, the rate at which cells leave the  $G_a$  phase at time  $t$  is equivalent to the rate at which cells entered this phase  $\tau_{G_a}$  hours earlier (i.e. at time  $t - \tau_{G_a}$ ), which is  $2\beta N_{M_b}(t - \tau_{G_a})$ . The flow chart below summarises this procedure.

|                                                           |   |                                                        |   |                                                                                                                                                                         |
|-----------------------------------------------------------|---|--------------------------------------------------------|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cells enter the $G_a$ phase at a rate $2\beta N_{M_b}(t)$ | → | Cells remain in the $G_a$ phase for $\tau_{G_a}$ hours | → | The number of cells leaving the $G_a$ phase at time $t$ is equivalent to the number that entered at time $t - \tau_{G_a}$ , which is $2\beta N_{M_b}(t - \tau_{G_a})$ . |
|-----------------------------------------------------------|---|--------------------------------------------------------|---|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Now let us consider the rate of change in the number of the  $G_b$  phase cells which is represented in equation (3.5) below.

$$\frac{dN_{G_b}(t)}{dt} = 2\beta N_{M_b}(t - \tau_{G_a}) - \gamma(t)N_{G_b}(t). \quad (3.5)$$

As cells leave the  $G_a$  phase, they enter the  $G_b$  phase, and so the term representing the loss of  $G_a$  phase cells in (3.4) appears in equation (3.5) as the growth term of  $G_b$  phase cells. The  $G_b$  model phase is a non-storage phase, and we assume that cells leave this phase at a rate  $\gamma(t)$ .

The rate of change in the number of  $G_c$  model phase cells can be represented by

$$\frac{dN_{G_c}(t)}{dt} = \gamma(t)N_{G_b}(t) - \gamma(t - \tau_{G_c})N_{G_b}(t - \tau_{G_c}). \quad (3.6)$$

As cells leave the  $G_b$  model phase, they enter the  $G_c$  model phase, and so we have the loss of  $G_b$  cells appearing as the growth of  $G_c$  cells in equation (3.6). As  $G_c$  is a storage phase, The rate at which cells leave the  $G_c$  phase at time  $t$  is equivalent to the rate at which cells entered at time  $t - \tau_{G_c}$ , which is  $\gamma(t - \tau_{G_c})N_{G_b}(t - \tau_{G_c})$ . We note that when considering cells which left the  $G_b$  phase at some previous time, the rate at which they left is dependent upon the value of  $\gamma$  at that earlier time, and not on the current value of  $\gamma$ .

We can derive similar equations for the rates of change in the remaining model phases shown in full in the system of equations (3.7) below (equations (3.4) - (3.6) are re-stated as (3.7a)-(3.7c) below). We note that in equations (3.7d) and (3.7e) there are terms which contain two storage duration variables -  $\tau_{G_c}$  and  $\tau_S$ . This is because the  $G_c$  and  $S_a$  phases, which are adjacent to each other, are both storage phases, and so the later storage phase will be considering the values of terms  $\tau_{G_c} + \tau_S$  hours ago.

$$\frac{dN_{G_a}(t)}{dt} = 2\beta N_{M_b}(t) - 2\beta N_{M_b}(t - \tau_{G_a}), \quad (3.7a)$$

$$\frac{dN_{G_b}(t)}{dt} = -\gamma(t)N_{G_b}(t) + 2\beta N_{M_b}(t - \tau_{G_a}), \quad (3.7b)$$

$$\frac{dN_{G_c}(t)}{dt} = \gamma(t)N_{G_b}(t) - \gamma(t - \tau_{G_c})N_{G_b}(t - \tau_{G_c}), \quad (3.7c)$$

$$\frac{dN_{S_a}(t)}{dt} = \gamma(t - \tau_{G_c})N_{G_b}(t - \tau_{G_c}) - \gamma(t - \tau_{G_c} - \tau_S)N_{G_b}(t - \tau_{G_c} - \tau_S), \quad (3.7d)$$

$$\frac{dN_{S_b}(t)}{dt} = \gamma(t - \tau_{G_c} - \tau_S)N_{G_b}(t - \tau_{G_c} - \tau_S) - \alpha N_{S_b}(t), \quad (3.7e)$$

$$\frac{dN_{M_a}(t)}{dt} = \alpha N_{S_b}(t) - \alpha N_{S_b}(t - \tau_M), \quad (3.7f)$$

$$\frac{dN_{M_b}(t)}{dt} = \alpha N_{S_b}(t - \tau_M) - \beta N_{M_b}(t). \quad (3.7g)$$

The set of equations (3.7) represent our mathematical model for the cell cycle process. In the next section, we will discuss how to determine parameters for the model presented in equations (3.7).

### Steady-state: growth in unchanging environmental conditions.

Any change in the environmental conditions is represented by a change in the variable  $\gamma(t)$ , which subsequently changes the rate of progression through the  $G_b$  phase and hence the number of cells in the remaining phases. However, other variables ( $\alpha$ ,  $\beta$ ,  $\tau_{G_a}$ ,  $\tau_{G_c}$ ,  $\tau_S$  and  $\tau_M$ ) are independent of environmental signals. Thus, we can classify variables as either environment-dependent or environment-independent variables. This is useful as once we have determined the values of the environment-independent variables, we can assume that they will take those values under any experimental conditions.

Note that this is only true of cells of the same cell line. Cells of different cell lines may have different rates of progression through certain phases, and therefore the environment independent variables above are not cell line independent variables. For instance, cells of the T47D cell line have an average S phase duration of 6 hours ([78], [81]), whereas cells of the MCF-7 cell line have an average S phase duration closer to 9 hours ([112], [115]). This means that the value of  $\alpha$  and  $\tau_S$  will be different for these two cell lines.

In this section, we will show that investigating steady-state conditions of equations (3.7) can produce additional relationships between the variables. We will then use these relationships, along with experimental results during steady-state growth of cells, to determine the values of our environment independent variables.

From equations (3.7) we see that  $\gamma(t)$  is the only rate coefficient that can change with time. Therefore if  $\gamma$  is constant with respect to time, then our system of equations will give rise to a steady-state in the sense that the proportion of cells across the model phases (and consequently the cell cycle phases) will be constant. A system which has a constant proportion of cells across the model phases will be referred to as being in *phase steady state*. A consequence of this relationship is that the growth of the population as a whole will become exponential (this is shown below in equations (3.13) - (3.15)).

Thus, with the aim of investigating this steady-state behaviour during growth in unchanging environmental conditions, we will consider the *proportion* of cells across the cell cycle phases instead of the *number* of cells in each of the phases. Once we have equations representing the rate of change in the proportion of cells we will then consider the case where the proportion of cells in each phase is constant. Let  $N_T(t)$  represent the total number of cells in our system at time  $t$ . We define  $G_a = \frac{N_{G_a}(t)}{N_T(t)}$ ,  $G_b = \frac{N_{G_b}(t)}{N_T(t)}$ , ... and so on, so that  $G_a(t)$ ,  $G_b(t)$ ,  $G_c(t)$ ,  $S_a(t)$ ,  $S_b(t)$ ,  $M_a(t)$  and  $M_b(t)$  represent the proportion of cells in their respective phases at time  $t$ . To determine the rate of change in the proportion of cells across the cell cycle phases, we use the product rule for differentiation. For instance, the rate of change in the proportion of cells in the  $G_a$  phase can be expressed as

$$\frac{dG_a(t)}{dt} = \frac{d\frac{N_{G_a}(t)}{N_T(t)}}{dt} = \frac{1}{N_T(t)} \frac{dN_{G_a}(t)}{dt} - \frac{N_{G_a}(t)}{N_T(t)^2} \frac{dN_T(t)}{dt}. \quad (3.8)$$

To determine an expression for  $\frac{dN_T(t)}{dt}$  we sum equations (3.7a) - (3.7g) to give

$$\frac{dN_T(t)}{dt} = \beta N_{M_b}(t),$$

which can be written as

$$\frac{dN_T(t)}{dt} = \beta M_b(t) N_T(t), \quad (3.9)$$

as  $\frac{N_{M_b}(t)}{N_T(t)} = M_b(t)$  from the definition of  $M_b(t)$ . Using equations (3.7a) and (3.9) we now rewrite (3.8) as

$$\begin{aligned}\frac{dG_a(t)}{dt} &= \frac{1}{N_T(t)}(2\beta N_{M_b}(t) - 2\beta N_{M_b}(t - \tau_{G_a})) - \frac{N_{G_a}(t)}{N_T(t)^2}\beta M_b(t)N_T(t), \\ &= 2\beta M_b(t) - 2\beta \frac{N_{M_b}(t - \tau_{G_a})}{N_T(t)} - \beta G_a(t)M_b(t).\end{aligned}\quad (3.10)$$

The same argument can be applied to the remaining 7 model phases, giving us the following set of equations (with equation (3.10) restated as equation (3.11a) below)

$$\frac{dG_a(t)}{dt} = 2\beta M_b(t) - 2\beta \frac{N_{M_b}(t - \tau_{G_a})}{N_T(t)} - \beta M_b(t)G_a(t), \quad (3.11a)$$

$$\frac{dG_b(t)}{dt} = 2\beta \frac{N_{M_b}(t - \tau_{G_a})}{N_T(t)} - \gamma(t)G_b(t) - \beta M_b(t)G_b(t), \quad (3.11b)$$

$$\frac{dG_c(t)}{dt} = \gamma(t)G_b(t) - \gamma(t - \tau_{G_c})\frac{N_{G_b}(t - \tau_{G_c})}{N_T(t)} - \beta M_b(t)G_c(t), \quad (3.11c)$$

$$\begin{aligned}\frac{dS_a(t)}{dt} &= \gamma(t - \tau_{G_c})\frac{N_{G_b}(t - \tau_{G_c})}{N_T(t)} - \gamma(t - (\tau_{G_c} + \tau_S))\frac{N_{G_b}(t - (\tau_{G_c} + \tau_S))}{N_T(t)} \\ &\quad - \beta M_b(t)S_a(t),\end{aligned}\quad (3.11d)$$

$$\frac{dS_b(t)}{dt} = \gamma(t - (\tau_{G_c} + \tau_S))\frac{N_{G_b}(t - (\tau_{G_c} + \tau_S))}{N_T(t)} - \alpha S_b(t) - \beta M_b(t)S_b(t), \quad (3.11e)$$

$$\frac{dM_a(t)}{dt} = \alpha S_b(t) - \alpha S_b(t - \tau_M) - \beta M_b(t)M_a(t), \quad (3.11f)$$

$$\frac{dM_b(t)}{dt} = \alpha S_b(t - \tau_M) - \beta M_b(t) - \beta M_b(t)M_b(t). \quad (3.11g)$$

Now let us assume that our cells are growing in unchanging environmental conditions so that we have reached the steady-state condition where the proportion of cells across the cell cycle phases remains constant with respect to time. Dropping the dependence on  $t$ , we define  $G_a, G_b, G_c, S_a, S_b, M_a$  and  $M_b$  to be the unchanging proportion of cells in their respective phases during phase steady-state. If we focus again on the  $G_a$  model phase, steady-state conditions imply that  $\frac{dG_a(t)}{dt} = 0$  and so we write (3.11a) as

$$0 = 2\beta M_b - 2\beta \frac{N_{M_b}(t - \tau_{G_a})}{N_T(t)} - \beta M_b G_a, \quad (3.12)$$

where we drop the time dependence on the terms  $G_a$  and  $M_b$  because the cells are in phase

steady-state. We can simplify the second term in equation (3.12) by evaluating  $N_{M_b}(t)$  and  $N_T(t)$  explicitly. To evaluate  $N_T(t)$ , we note that in unchanging environmental conditions, equation (3.9) becomes

$$\frac{dN_T(t)}{dt} = \beta M_b N_T(t). \quad (3.13)$$

Thus, the cell population is growing exponentially with constant growth rate  $\beta M_b$ . For simplicity, we will define the constant  $\rho$  as

$$\rho = \beta M_b. \quad (3.14)$$

We can determine  $N_T(t)$  explicitly from (3.13) as

$$N_T(t) = N_0 e^{\rho t}, \quad (3.15)$$

where  $N_0$  is the total number of cells in our system at time 0. To get an expression for  $N_{M_b}(t)$  we use the fact that  $M_b = \frac{N_{M_b}(t)}{N_T(t)}$  so that

$$\begin{aligned} N_{M_b}(t) &= M_b N_T(t) \\ &= M_b N_0 e^{\rho t}. \end{aligned} \quad (3.16)$$

and so using equations (3.15) and (3.16), we can write the second term in equation (3.12) as

$$\begin{aligned} 2\beta \frac{N_{M_b}(t - \tau_{G_a})}{N_T(t)} &= 2\beta \frac{M_b N_0 e^{\rho(t - \tau_{G_a})}}{N_0 e^{\rho t}} \\ &= 2\rho e^{-\rho \tau_{G_a}}. \end{aligned} \quad (3.17)$$

Hence equation (3.12) simplifies to

$$0 = 2\rho - 2\rho e^{-\rho \tau_{G_a}} - \rho G_a. \quad (3.18)$$

Likewise, expressions for the remaining 6 model phases can be determined. Letting  $\gamma$  represent the constant value of  $\gamma(t)$  for our unchanging environmental conditions, we can generate the following set of equations (equation (3.18) is repeated as equation (3.19a)),

where all the variables are now constants

$$0 = 2\rho - 2\rho e^{-\rho\tau_{G_a}} - \rho G_a, \quad (3.19a)$$

$$0 = 2\rho e^{-\rho\tau_{G_a}} - \gamma G_b - \rho G_b, \quad (3.19b)$$

$$0 = \gamma G_b - \gamma G_b e^{-\rho\tau_{G_c}} - \rho G_c, \quad (3.19c)$$

$$0 = \gamma G_b e^{-\rho\tau_{G_c}} - \gamma G_b e^{-\rho(\tau_S + \tau_{G_c})} - \rho S_a, \quad (3.19d)$$

$$0 = \gamma G_b e^{-\rho(\tau_S + \tau_{G_c})} - \alpha S_b - \rho S_b, \quad (3.19e)$$

$$0 = \alpha S_b - \alpha S_b e^{-\rho\tau_M} - \rho M_a, \quad (3.19f)$$

$$0 = \alpha S_b e^{-\rho\tau_M} - \rho - \rho M_b. \quad (3.19g)$$

As we mentioned earlier, the values of  $\alpha$ ,  $\beta$ ,  $\tau_{G_a}$ ,  $\tau_{G_c}$ ,  $\tau_S$  and  $\tau_M$  are considered environment independent. This means that even under different environmental conditions, these variables take the same values, and so once their values have been determined for one experiment, they will not change even across different experiments and different environmental conditions.

Biological experiments will often leave cells in an unchanging environment for 1-2 days to allow the cells to reach this steady-state condition before they change the cells' environment ([134, 81, 98, 136, 97]). Let us assume for now that during this initial period of unchanging environmental conditions we can determine the value of  $\rho$  and the steady state proportion of cells across the three phases -  $G$ ,  $S$  and the  $M$ . We will also assume that we have determined the values for  $\tau_{G_a}$ ,  $\tau_{G_c}$ ,  $\tau_S$  and  $\tau_M$  (we discuss the estimation of these variables in Section 4.1). This leaves 10 unknown variables which are listed in Table 3.1. The set of equations (3.19) are not linear in our unknown variables - specifically due to the presence of the terms  $\gamma G_b$  and  $\alpha S_b$  throughout the equations in (3.19). However, we can rearrange (3.19b) and (3.19g) to find expressions for the non-linear terms as follows

$$\gamma G_b = 2\rho e^{-\rho\tau_{G_a}} - \rho G_b, \quad (3.20)$$

$$\alpha S_b = e^{\rho\tau_M}(\rho + \rho M_b). \quad (3.21)$$

Replacing the terms  $\gamma G_b$  and  $\alpha S_b$  in the remaining equations in (3.19) with the ex-

Table 3.1: *Table of variables, classified as assumed known or assumed unknown and environment dependent or environment independent.*

|                                | Assumed known | Assumed unknown |
|--------------------------------|---------------|-----------------|
| <b>Environment independent</b> | $\tau_{G_a}$  | $\alpha$        |
|                                | $\tau_{G_c}$  | $\beta$         |
|                                | $\tau_S$      |                 |
|                                | $\tau_M$      |                 |
| <b>Environment dependent</b>   | $G$           | $G_a$ $\gamma$  |
|                                | $S$           | $G_b$ $S_b$     |
|                                | $M$           | $G_c$ $M_a$     |
|                                | $\rho$        | $S_a$ $M_b$     |

pressions in equations (3.20) and (3.21) gives us the following set of linear equations:

$$0 = 2 - 2e^{-\rho\tau_{G_a}} - G_a, \quad (3.22a)$$

$$0 = (2e^{-\rho\tau_{G_a}} - G_b)(1 - e^{-\rho\tau_{G_c}}) - G_c, \quad (3.22b)$$

$$0 = e^{-\rho\tau_{G_c}}(2e^{-\rho\tau_{G_a}} - G_b)(1 - e^{-\rho\tau_S}) - S_a, \quad (3.22c)$$

$$0 = (2e^{-\rho\tau_{G_a}} - G_b)e^{-\rho(\tau_S + \tau_{G_c})} - e^{\rho\tau_M}(1 + M_b) - S_b, \quad (3.22d)$$

$$0 = e^{\rho\tau_M}(1 + M_b)(1 - e^{-\rho\tau_M}) - M_a. \quad (3.22e)$$

Equations (3.19) contained 7 non-linear equations in 9 unknowns. Equations (3.22) now consists of 5 linear equations in 7 unknowns (the terms  $\gamma$  and  $\alpha$  no longer appear in (3.22)). In addition to the above 5 equations, we also have the identities

$$G = G_a + G_b + G_c, \quad (3.23)$$

$$S = S_a + S_b, \quad (3.24)$$

$$M = M_a + M_b. \quad (3.25)$$

(Note that we do not need to include the equation  $G + S + M = 1$  since we have assumed that  $G, S$  and  $M$  are already known.)

Thus we have 8 linear equations in 7 unknowns. Clearly they cannot be linearly

independent since we have more equations than unknowns. We can rewrite the 8 equations in (3.22) and (3.23) - (3.25) in the following matrix form:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 - e^{-\rho\tau G_c} & 1 & 0 & 0 & 0 & 0 \\ 0 & e^{-\rho\tau G_c}(1 - e^{-\rho\tau S}) & 0 & 1 & 0 & 0 & 0 \\ 0 & e^{-\rho(\tau G_c + \tau S)} & 0 & 0 & 1 & 0 & e^{\rho\tau M} \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 - e^{\rho\tau M} \\ -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} G_a \\ G_b \\ G_c \\ S_a \\ S_b \\ M_a \\ M_b \end{bmatrix} = \begin{bmatrix} 2 - 2e^{-\rho\tau G_a} \\ 2e^{-\rho\tau G_a}(1 - e^{-\rho\tau G_c}) \\ 2e^{-\rho(\tau G_a + \tau G_c)}(1 - e^{-\rho\tau S}) \\ 2e^{-\rho\tau G_a}e^{-\rho(\tau S + \tau G_c)} - e^{\rho\tau M} \\ e^{\rho\tau M} - 1 \\ -G \\ -S \\ -M \end{bmatrix} \quad (3.26)$$

The equations are not linearly independent however they are consistent (the final equation  $M_a + M_b = M$  is dependent upon the others since all column sums are zero) and have a unique solution given by

$$G_a = 2 - 2e^{-\rho\tau G_a}, \quad (3.27a)$$

$$G_b = 2e^{-\rho\tau G_a} - (2 - G)e^{\rho\tau G_c}, \quad (3.27b)$$

$$G_c = (2 - G)e^{\rho\tau G_c} - (2 - G), \quad (3.27c)$$

$$S_a = 2 - G - (2 - G)e^{-\rho\tau S}, \quad (3.27d)$$

$$S_b = (2 - G)e^{-\rho\tau S} - (1 + M), \quad (3.27e)$$

$$M_a = 1 + M - (1 + M)e^{-\rho\tau M}, \quad (3.27f)$$

$$M_b = (1 + M)e^{-\rho\tau M} - 1. \quad (3.27g)$$

We note that the RHS of equations (3.27) contain only variables which we assume can be determined.

We can now substitute results from equations (3.27) into equation (3.21) to evaluate  $\alpha$ , into (3.14) to evaluate  $\beta$  and into (3.20) to evaluate  $\gamma$ , giving us the following expressions:

$$\alpha = \frac{\rho(1 + M)}{(1 + S + M)e^{-\rho\tau_S} - (1 + M)}, \quad (3.28)$$

$$\beta = \frac{\rho}{(1 + M)e^{-\rho\tau_M} - 1}, \quad (3.29)$$

$$\gamma = \frac{\rho(2 - G)}{2e^{-\rho(\tau_{G_a} + \tau_{G_c})} - (2 - G)}. \quad (3.30)$$

Therefore, if we know the steady-state values of  $\rho, G, S$  and  $M$  for an experiment, as well as the values for  $\tau_{G_a}, \tau_{G_c}, \tau_S$  and  $\tau_M$ , then we can uniquely determine the remaining 10 unknown variables (see Table 3.1) for that experiment and in particular, we can determine the values of  $\alpha$  and  $\beta$ . Fortunately, the values of  $\rho, G, S$  and  $M$  are often provided in (or at least they can be determined from) experimental results in the literature. Unfortunately, the storage phase durations are not provided explicitly in experimental results. In the next chapter, we will discuss how we can estimate the values of  $\tau_{G_a}, \tau_{G_c}, \tau_S$  and  $\tau_M$  in the MCF-7 breast cancer cell line. Once we have a fully parameterised model for the MCF-7 cell line, we will analyse the model to get a better understanding of how the parameters influence model behaviour. We will also use our model to discuss some misconceptions in the literature regarding the average duration of the cell cycle.

## Chapter 4

# The MCF-7 breast cancer cell line

In this chapter, we will parameterise the cell cycle model from equations (3.7) to the MCF-7 breast cancer cell line, and we will then analyse the equations and get a better understanding of what parameters influence model behaviour. We begin by discussing how we can determine the values of the experiment-independent variables.

### 4.1 Determining the duration of the storage phases

The storage phases were introduced to model the fact that cells must spend a minimum amount of time in each phase. However the time taken to complete these compulsory tasks for each phase cannot be measured directly in biological experiments, and thus the duration of the storage phases has not been explicitly determined biologically. However we can still use the results of biological experiments to estimate the duration of these phases. We begin this section by describing a particularly useful biological experiment in detail, and later show how the results of this experiment can be used to estimate the values of  $\tau_{G_a}$ ,  $\tau_{G_c}$ ,  $\tau_S$  and  $\tau_M$

#### Mitotic selection - an experimental procedure

In the human body, breast cells are attached to the lining of breast ducts and in biological experiments, petri dishes contain substances to which these cells can attach. During

mitosis however, cells detach from the substratum as they split into two. Thus if a petri dish is shaken, the cells that fall off from the dish are cells that are in the mitotic phase of the cell cycle. Cells chosen in this way are called mitotically selected cells. As cells complete mitosis very quickly (within 30 minutes [78]), mitotically selected cells are highly synchronised, and subsequent observations of cell cycle progression allow us to observe the length of the storage phases.

Cells which are shaken off the petri dish during mitotic selection are placed into a new petri dish which contains the same growth factors as the original dish. Thus we assume that cells which have been mitotically selected are not exposed to any changes in environmental conditions<sup>1</sup>.

Unfortunately, the process of mitotic selection does not always produce cells exclusively in the mitotic phase, and cells in other phases may fall off the petri dish when shaken. Selected cells which are not in the mitotic phase will be referred to as impure cells, whereas those which have been successfully selected from the mitotic phase will be referred to as pure mitotically selected cells.

### **A mitotic selection experiment using MCF-7 breast cancer cells**

An experiment which investigates cell cycle progression in mitotically selected MCF-7 breast cancer cells is outlined in [78]. Cells were mitotically selected and the selected cells were re-plated onto a fresh medium for growth and attachment. After three hours, and at intervals thereafter, the proportion of cells across the three cell cycle phases was determined. Figure 4.1 shows how the cell cycle phase distribution changes over a 34.5 hour period.

It would be expected that after mitotic selection, the majority of cells would exist in the M phase. However at the first point of investigation, 3 hours after mitotic selection, only about 3% of cells exist in the M phase whereas 90% of cells are in the G phase (see Figure 4.1). This is because the pure mitotically selected cells have re-entered the G

---

<sup>1</sup>This is not entirely true. When cells are selected by mitotic selection, they are replanted onto a new petri dish, which contains the same growth conditions as the original petri dish. Cancer cells however are noted for their ability to secrete growth factors. This means that the original petri dish may have higher concentrations of secreted growth factors and therefore be more encouraging for growth. However [78] has shown that the difference in environments due to these secreted growth factors has minimal impact on the cell cycle distribution.

## NOTE:

This figure is included on page 85 of the print copy of the thesis held in the University of Adelaide Library.

Figure 4.1: *Cell cycle distribution of mitotically selected cells, reproduced from [78]. Cells were mitotically selected at time 0. By the first point of investigation, 3 hours after selection, the mitotically selected cells had completed mitosis and moved into the G phase. Points are the average of at least two independent experiments. The curve through the data points is the result of an unknown curve fitting process.*

phase by this time (it was noted in [78] that almost all of the pure mitotically selected cells entered the G phase within 30 minutes of mitotic selection). The 10% of cells which were not in the G phase at this time represent the impure population of selected cells.

Once the mitotically selected cells have moved into the G phase, the S phase proportion does not increase significantly until 7.5 hours into the experiment. Although this is the first noticeable increase in the S phase proportion, it is still possible that the G phase cells began to progress into the S phase at any time between 4.5 and 7.5 hours. This is because at the 4.5 hour mark, the G phase proportion was at a maximum and thereafter began to decrease. The S phase proportion increases from about 7% to 25% by 7.5 hours, and continues to increase, reaching a maximum of about 85% of cells in the S phase at 13.5 hours.

The proportion of cells in the M phase remains approximately constant until 16.5 hours after selection. The previous point of investigation, which indicated no increase in the M phase proportion, was 1.5 hours earlier at 15 hours, and so the first point of increase could have occurred at time points between 15 and 16.5 hours. The proportion of cells in the M phase increases from about 3% to 11% by 16.5 hours, and continues to increase with the M phase cells reaching a maximum proportion of 39% at 19.5 hours.

As the mitotically selected cells move through the S phase and the M phase, the proportion of cells in the G phase remains approximately constant. At about 21 hours into the experiment, a noticeable increase in the G phase proportion is observed. The increase in the G phase proportion could have occurred at anytime between 19.5 and 21 hours, as 19.5 hours was the last point of investigation when the G phase proportion had not significantly changed. This increase in G phase cells represents the mitotically selected cells completing mitosis (for the second time since the start of the experiment) and moving back into the G phase.

Once the mitotically selected cells have progressed through the G phase again, they re-enter the S phase between 28.5 and 31.5 hours, as evidenced by the increased proportion of S phase cells and the drop in the G phase proportion during these times. Note that the rise in the S phase percentage is not as high as when the mitotically selected cells entered the S phase the first time after selection, indicating that the cells are not as synchronised during this second round of the cell cycle. Cells will continue to lose synchrony until they

eventually reach a steady-state proportion across the phases.

### **Approximating the duration of the storage phases using the mitotic selection experiment**

We can use Figure 4.1 to estimate the duration of our storage phases. Observe that once the S phase proportion begins to increase at 4.5-7.5 hours, the M phase proportion does not change until sometime between 15 and 16.5 hours. This implies that once cells enter the S phase, they must remain in this phase for a minimum amount of time, that minimum amount of time being in the range 7.5-12 hours. Thus, we would choose  $\tau_S$  to be in this range.

Similarly, once the cells first enter the M phase sometime between 15 and 16.5 hours, the G phase proportion does not increase again until 19.5 - 21 hours, and so once the cells enter the M phase they do not begin to progress into the G phase until sometime between 3 and 6 hours later. Thus we assume  $\tau_M$  to be in this range.

As the S phase duration has not increased significantly until 4.5 - 7.5 hours we estimate the *total* storage duration of the G phase to be somewhere between 4 and 7 hours. Thus  $\tau_{G_a} + \tau_{G_c}$  takes a value between 4.5 and 7.5.

Fortunately, we can improve on the estimated ranges for the storage durations. We will use the results of another similar experiment to first improve on the ranges chosen for the value of  $\tau_{G_a} + \tau_{G_c}$ . Cell cycle progression in mitotically selected MCF-7 cells was also investigated in [115]. In their investigations (not shown), mitotic selection resulted in a similar number of cells in the G phase within 30 minutes of selection. Thereafter, the cell cycle phase distributions were similar to the results found in [78].

One difference between the experiments is the reduced amount of noise present for the first 6 hours after selection. The observed G phase proportion remained high for at least 6 hours, and by 7.5 hours the cells had begun to leave the G phase. Thus,  $\tau_{G_a} + \tau_{G_c}$  must be in the range 6 - 7.5 hrs. With no additional information, we take the midpoint and set  $\tau_{G_a} + \tau_{G_c} = 6.75$  hours. This refined range for the value of  $\tau_{G_a} + \tau_{G_c}$  also allows refinement of the range for  $\tau_S$ . As we now know that the G phase proportion decreases sometime between 6 and 7.5 hours, then this is also when the S phase starts to increase,

and so  $\tau_S$  must be between  $15 - 7.5 = 7.5$  and  $16.5 - 6 = 10.5$ . Taking the midpoint gives  $\tau_S = 9$  hours.

We can also improve on the range we have found for the value of  $\tau_M$ , which is currently 3 - 6 hrs. We use our model equations to improve this range (instead of using results from [115], which we did when calculating  $\tau_{G_a} + \tau_{G_c}$  and  $\tau_S$ ). If we reconsider the steady-state equation for  $M_b$  in (3.27) from Section 3.3 we see that if we take  $\tau_M$  to be very large,  $M_b$  becomes small and if we let  $\tau_M$  be too large,  $M_b$  may become negative which is not biologically possible. In the next section, we encounter an experiment which finds that during steady state growth of MCF-7 cells, the population of cells grows with  $\rho = 0.029\text{hr}^{-1}$  and  $M = 0.117$ . Substituting these values into the last equation in (3.27) gives the following expression

$$M_b = (1 + 0.117)e^{-0.029\tau_M} - 1.$$

As we must have  $M_b \geq 0$  we can write

$$\begin{aligned} (1 + 0.117)e^{-0.029\tau_M} - 1 &\geq 0 \\ \Rightarrow \tau_M &\leq 3.8. \end{aligned}$$

Thus we must take  $\tau_M$  to be in the range 3 - 3.8 hours instead of 3 - 6 hours, and taking the midpoint gives  $\tau_M = 3.4$  hours.

We are now left to determine the values of  $\tau_{G_a}$  and  $\tau_{G_c}$ . Note that we cannot differentiate between the  $\tau_{G_a}$  and  $\tau_{G_c}$  phase when using mitotic selection. In order to differentiate between the two G phase storage durations, we use the results from another experiment. Musgrove et al. [85] attempt to identify at which points in time during the G phase that cells were responsive to external signals (i.e. they tried to identify when cells entered the  $G_b$  model phase once they entered the G phase). They did this by mitotically selecting cells, and then exposing the mitotically selected cells to growth arresting substances at intervals thereafter. If a substance is added before the cells enter the  $G_b$  model phase and then maintained in the medium, then all of the mitotically selected cells will be responsive to this change in the environment, and we would expect to notice an alteration in the cell cycle phase profiles for the remainder of the experiment. However, if the substance is added too late - for instance, when the mitotically selected cells have progressed into the

Table 4.1: Values of the durations of the storage phases

| parameter    | range          | midpoint |
|--------------|----------------|----------|
| $\tau_{G_a}$ | 4 - 7.5 hrs    | 5.75 hrs |
| $\tau_{G_c}$ | 0 - 2 hrs      | 1 hr     |
| $\tau_S$     | 7.5 - 10.5 hrs | 9 hrs    |
| $\tau_M$     | 3 - 3.8 hrs    | 3.4 hrs  |

$G_c$  model phase - then the cells will not be responsive to the environmental change and the cell cycle phase distributions will be relatively unchanged when compared to that of control cells. The following flow chart represents this concept.

|                         |                         |                     |
|-------------------------|-------------------------|---------------------|
| $G_a$                   | $G_b$                   | $G_c$               |
| ↑                       | ↑                       | ↑                   |
| Substances added        | Substances added        | Substances added    |
| during this period will | during this period will | during this period  |
| have maximal effect     | have reduced effects    | will have no effect |

Musgrove et al. [85] used these techniques to identify the period in the G phase when cells are responsive to environmental signals. Hydroxycyclophene was added to mitotically selected cells at different times after mitotic selection, and then changes in cell cycle phase distributions were determined (not shown). Results of the experiment indicate that cells that are exposed to hydroxycyclophene for times  $t < 2$  hours do not display different cell cycle phase kinetics when compared to cell populations that are exposed at time 0. If hydroxycyclophene is added at times later than 2 hours after mitotic selection, the authors note changes in the cell cycle phase kinetics when compared to populations where exposure occurs at the beginning of the experiment. It may be tempting at this point to take  $\tau_{G_a} = 2$  hours, however we must consider that the addition of a substance will take some time to produce internal proteins that will in turn, take time to influence the progression past the Rb point, and thus all we can really say is  $\tau_{G_a} \leq 2$ . Thus, in the absence of additional information, we take the midpoint and set  $\tau_{G_a} = 1$ . This leaves  $\tau_{G_c}$  in the range 6-7.5 hrs, and we also take the midpoint of this range and set  $\tau_{G_c} = 5.75$  hrs. Table 4.1 shows the values of the storage phase durations that we have determined in this section.

## 4.2 Determining the remaining environment independent parameters

In Section 4.1, we determined the values of the storage durations for each of the storage phases in the MCF-7 breast cancer cell line. As we discussed earlier, if we know these four values, as well as the steady-state values  $G, S, M$  and  $\rho$ , then we can determine all other remaining variables, including the steady-state model phase proportions and the values of  $\alpha$  and  $\beta$ . In this section, we use the results from an experiment (which we call the *base experiment*) to determine what the values of  $\alpha$  and  $\beta$  are. As  $\alpha$  and  $\beta$  are environment independent, we only need to determine the values of these variables once, and they will be set for any other experiment involving MCF-7 cells. We choose to use the experiments detailed in [112] to determine our choices for  $\alpha$  and  $\beta$ . Sutherland et al. [112] perform multiple experiments to determine the average steady-state proportions of cells across the 3 cell cycle phases during exponential growth, and report these values along with a standard error. In particular, they find that growth was exponential for the first 7 days, and during this time the steady-state proportion of  $G, S$  and  $M$  was measured as  $48.9 \pm 0.6\%$ ,  $39.4 \pm 0.6\%$  and  $11.6 \pm 0.3\%$  respectively. They also report that the doubling time for the cell population during this exponential growth period is 24 hours. Although a margin of error was not reported for this observation, the experimental results were performed in multiple flasks (somewhere between 8 and 28 flasks, exact number not provided) and the experiment was then repeated 2 - 7 times. It would be expected that the observed cell population doubling time, as with the observed cell cycle phase distributions, would have a small margin of error due to the large number of observations.

We can now substitute the experimentally observed values of  $G, S, M$  and  $\rho$ , along with the storage phase durations calculated in Section 4.1, into equations (3.28) and (3.29). Doing this gives  $\alpha = 0.670\text{hr}^{-1}$  and  $\beta = 2.31\text{hr}^{-1}$  to 3 decimal places respectively, and as we said earlier, these values are experiment-independent and will not change as we change environmental conditions across experiments. Thus, values of the environment independent variables for the MCF-7 breast cancer cell line have now been fully determined, and are shown in Table 4.2.

Now that we know the values of all experiment-independent variables, we can now

Table 4.2: *Summary of the values of our experiment-independent variables, shown to 3 decimal places.*

| Known experiment-independent variables | Experiment dependent variables |          |
|----------------------------------------|--------------------------------|----------|
| $\tau_{G_a} = 1$ hr                    | $G$                            | $G_a$    |
| $\tau_{G_c} = 5.75$ hrs                | $G_b$                          | $G_c$    |
| $\tau_S = 9$ hrs                       | $S$                            | $S_a$    |
| $\tau_M = 3.4$ hrs                     | $S_b$                          | $M$      |
| $\alpha = 0.670$ h <sup>-1</sup>       | $M_a$                          | $M_b$    |
| $\beta = 2.31$ h <sup>-1</sup>         | $\rho$                         | $\gamma$ |

simulate cell cycle progression in MCF-7 breast cancer cells. We will eventually use our parameterised model to simulate cell cycle progression under different experimental conditions, however before we do this, we wish to analyse our model a bit further to get a better understanding of how the model behaves.

### 4.3 Model analysis

When we first derived the steady-state equations in (3.19), we considered the variables  $\alpha$  and  $\beta$  to be unknown variables. Now that we know the values of these experiment-independent variables for the MCF-7 breast cancer cell line, we can re-assess our steady-state system with a different perspective on what is known and unknown to us.

As discussed in Section 3.3, knowledge of the steady-state values for  $G, S, M$  and  $\rho$  are sufficient to determine the remaining experiment-dependent variables in Table 4.2. However the values of these variables during the steady-state period of growth are not always provided in full for each biological experiment. Thus one reason for re-assessing our steady-state system is to determine which of the variables -  $G, S, M$  and  $\rho$  - need to be given in order to determine the remaining unknowns of our system.

It is clear that once 2 of the 3 variables  $G, S$  and  $M$  are known, the third can be determined using the equation  $G + S + M = 1$ . Thus we could claim that only 3 of the variables needs to be given - 2 of the steady-state proportions across the G, S and M phases, and the value for  $\rho$ . However we can do better than this. Consider equation (3.29) from Section 3.3. The equation is an expression for  $\beta$  but since  $\beta$  is now known,

we will rearrange to make  $M$  the subject:

$$M = \left(\frac{\rho}{\beta} + 1\right)e^{\rho\tau_M} - 1. \quad (4.1)$$

Now let us substitute this value for  $M$  into equation (3.28) from Section 3.3 and rearrange to make  $S$  the subject:

$$S = \left(\left(\frac{\rho}{\alpha} + 1\right)\left(\frac{\rho}{\beta} + 1\right)e^{\rho\tau_S} - \left(\frac{\rho}{\beta} + 1\right)\right)e^{\rho\tau_M}. \quad (4.2)$$

Substituting equations (4.1) and (4.2) into the equation  $G + S + M = 1$  gives an expression for  $G$ :

$$G = 2 - \left(\frac{\rho}{\alpha} + 1\right)\left(\frac{\rho}{\beta} + 1\right)e^{\rho(\tau_S + \tau_M)}. \quad (4.3)$$

From equations (4.1) - (4.3), we can see that if  $\rho$  is known, then the values of  $G$ ,  $S$  and  $M$  can be determined (all other variables that appear on the RHS of equations (4.1) - (4.3) are known, see Table 4.2).

We cannot determine an explicit expression for  $\rho$  using equations (4.1) - (4.3) and so it is not immediately obvious whether knowledge of the cell cycle proportions will allow us to determine a unique value for  $\rho$ . However it can be shown that equations (4.1) - (4.3) are one-to-one for  $\rho \geq 0$  (as the first derivative w.r.t.  $\rho$  is strictly positive for equations (4.1) and (4.2) and strictly negative for equation (4.3)). This means that if we know  $G$ ,  $S$  or  $M$ , then the value of  $\rho$  can be uniquely determined numerically and in turn, so can the remaining 2 cell cycle phase proportions.

Thus of the 4 variables -  $G$ ,  $S$ ,  $M$  or  $\rho$  - only *one* of these needs to be provided in order to determine the remaining 3, and in turn, these 4 variables can be used to determine the remaining experiment-dependent variables in Table 4.2 for growth in unchanging environmental conditions (see Section 3.3). This will turn out to be particularly useful, as many biological experiments will only report the steady-state values of  $G$ ,  $S$ ,  $M$  or  $\rho$ , and so the earlier analysis gives us confidence that this is enough information to determine the remaining unknown variables.

### Average cell cycle duration and population doubling times

As we have just discussed, equation (4.3) is monotonically decreasing. If we consider it further, we notice that if we take the limit as  $\rho \rightarrow \infty$ ,  $G \rightarrow -\infty$ . Thus  $G$  will eventually take on negative values for large enough  $\rho$ . We calculate that for  $\rho \geq 0.473\text{hr}^{-1}$ ,  $G$  becomes negative. Since this is not biologically possible, we must impose an upper bound on  $\rho$  in order to produce biologically relevant values for  $G$ ,  $S$  and  $M$ . This upper bound for  $\rho$  is equivalent to a lower bound for the doubling time of 14.6 hrs. We choose to convert the proliferation rate into a value for the doubling time (i.e. we chose to utilise  $\frac{\ln(2)}{\rho}$  instead of  $\rho$ ) as doubling times have a more intuitive physical interpretation - ‘the time taken for the population to double in size’.

Observe that the sum of the duration of the storage phases -  $\tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M$  - is 19.15 hrs (Using results from Table 4.2), which is greater than our calculated minimum doubling time of 14.6 hrs. A cell population in phase steady-state cannot double in size faster than the time taken for any one cell to complete a full round of cell division, and so we must re-think our calculation for the minimum doubling time. The sum of the storage phase durations (which we just calculated to be 19.15 hrs) is also not the value of the minimum doubling time, because the rate of progression through the non-storage phases  $S_b$  and  $M_b$  is not instantaneous, and so the minimum doubling time must be larger than this value.

To correctly calculate the biologically minimum doubling time, recall that the value of  $\rho$  can only be changed by changing the rate at which cells progress through the  $G_b$  phase, as the rates of progression through any of the other model phases is constant. Thus, the smallest biologically reasonable value of the doubling time occurs when progression through the  $G_b$  model phase is instantaneous, or when  $\gamma \rightarrow \infty$ . To determine the value for the minimum doubling time, we derive an equation relating  $\rho$  and  $\gamma$  and then solve for  $\rho$  as  $\gamma \rightarrow \infty$ . By substituting the expression for  $G$  from equation (4.3) into equation (3.30), we get the following expression for  $\gamma$  in terms of  $\rho$

$$\gamma = \frac{\rho(\frac{\rho}{\alpha} + 1)(\frac{\rho}{\beta} + 1)}{2e^{-\rho(\tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M)} - (\frac{\rho}{\alpha} + 1)(\frac{\rho}{\beta} + 1)}. \quad (4.4)$$

We find the value for  $\rho$  as  $\gamma \rightarrow \infty$  by identifying when the denominator of equation (4.4)

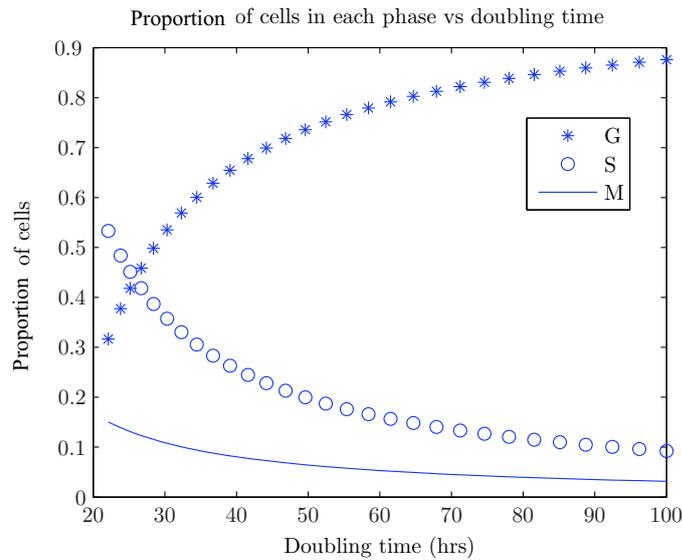


Figure 4.2: *Proportion of cells across the cell cycle phases versus the doubling time of the population. Equations (4.1) - (4.3) were solved numerically for doubling times from 21.04-100 hours, where 21.04 hours is the smallest possible doubling time.*

is zero to find  $\rho = 0.033\text{hr}^{-1}$  (we note the numerator of equation (4.4) is positive for this value of  $\rho$ ). This is equivalent to a doubling time of 21.04 hours, and thus we take this to be the biologically minimum doubling time for the MCF-7 cell population growing in phase steady-state.

We wish to further explore the relationship between  $G$ ,  $S$ ,  $M$  and  $\rho$ . We can plot equations (4.1) - (4.3) to get a visual representation of how the cell cycle phase proportions vary with the doubling time. Figure 4.2 illustrates the relationship between the 3 cell cycle phases and the doubling time of the population during periods of exponential growth. Doubling times in Figure 4.2 range from 21.04 hrs, the minimum biologically relevant value, up to 100 hours. As the doubling time increases, the proportion of cells in the  $G$  phase increases while the proportion of cells in the  $S$  and  $M$  phases decrease. For larger values of the doubling time (not shown),  $G$  tends to 1 and  $S$  and  $M$  tend to 0. This is what we would expect, as large doubling times correspond to cells spending larger amounts of time in the  $G$  phase.

As we mentioned earlier, knowledge of just one of the variables  $G$ ,  $S$ ,  $M$  or  $\rho$  allows us to determine the remaining experiment dependent variables. However, if we consider

Figure 4.2, we can see that this may not always be a reliable technique. For example, small variations in G, S or M phase data for slow growing populations (high doubling times) produce a large range of possible values of the doubling time of the population. Experimental data will never provide exact values for the steady-state parameters, and so we have to be careful about their sensitivity when determining the values of other parameters. This figure also shows that for cell populations with a high growth rate (low doubling time), G and S phase proportions are good candidates for making conclusions about the corresponding doubling times. However it appears that steady-state M phase proportions are never good candidates for making solid conclusions about the corresponding doubling time of the population.

### **Sensitivity analysis: How the experiment independent variables influence the model**

Figure 4.2 provides a lot of information about the behaviour of the cell line under steady-state conditions. We wish to investigate how such a representation may change if we change the values of the experiment-independent variables. From now on, a representation like Figure 4.2 will be referred to as *phase diagram*.

First, we wish to summarise the steps required to produce such a diagram. First we determine the storage phase durations. We then use an experiment with reliable data, called the base experiment, to determine values of G, S, M and  $\rho$  in phase steady-state (we call  $G, S, M$  and  $\rho$  from this experiment the base experiment variables, as they have been determined from our base experiment). We then use these variables, along with the storage phase durations, to determine the remaining variables  $\alpha$  and  $\beta$ . With fully determined experiment-independent variables, we can then uniquely determine the values of G, S and M for any given value of the doubling time, and hence can produce a phase diagram similar to Figure 4.2.

What if the chosen storage phase values were different? Let's begin by assuming that instead of  $\tau_S = 9$ , we chose  $\tau_S = 1$ , a significant change in this variable. All other storage phase durations are unchanged, and we still use the base experiment discussed in Section 4.2 to determine  $\alpha$  and  $\beta$ . With this change in  $\tau_S$  we find that  $\alpha = 0.09$ , compared with the value of 0.67 determined when  $\tau_S = 9$ . The value of  $\beta$  is unchanged,

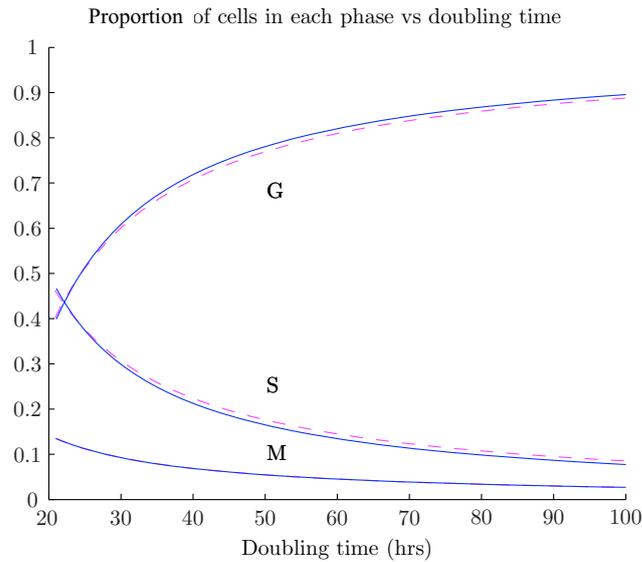


Figure 4.3: *Proportion of cells across the cell cycle phases for doubling times ranging from 21.04 hours to 100 hours. The solid blue curve represents the phase diagram for  $\tau_S = 9$  and the dashed magenta curve represent that for  $\tau_S = 1$ .*

as it is independent of  $\tau_S$ . The corresponding phase diagram with this change in  $\tau_S$  is shown in Figure 4.3. As we can see, this change in  $\tau_S$  has produced little change in the phase diagram. We find that doing a similar investigation by changing  $\tau_M$  also produces little change in this phase diagram. Thus, our phase diagram appears to be relatively independent of these experiment-independent variables.

Why is this the case? And what, then, drives the shape of the curves in the phase diagram? We suggest it is the choice of the base experiment variables (i.e. the values of  $G$ ,  $S$ ,  $M$  and  $\rho$  from the base experiment). The base variables really control measurements such as the average cell cycle phase durations. The choice of variables such as the storage phase durations only changes the distribution with which cells spend in a phase. Specifically, variables such as  $\tau_S$  and  $\alpha$  determine the relative amount of time a cell spends in the  $S_a$  and  $S_b$  phase respectively, however the overall time spent in the S phase is determined quite strongly by the base variables. This is evidenced in Figure 4.4, where we see that different values of  $\tau_S$  produce significantly different phase diagram for  $S_a$  and  $S_b$ , but that in each, the overall S phase proportion is not significantly different (compare black and magenta symbols respectively). We find that altering the value of  $\tau_M$  produces similar results to when we alter the value of  $\tau_S$  - i.e. we only alter the relative amount of time

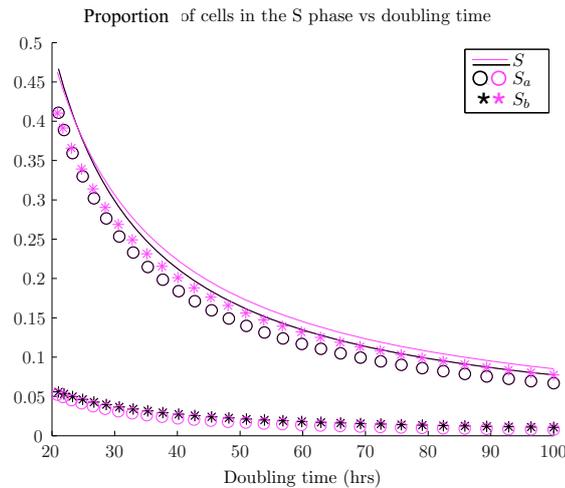


Figure 4.4: *Proportion of cells in the S phase for doubling times ranging from 21.04 hours to 100 hours are shown as the black solid curve when we choose  $\tau_S = 9$  and the magenta solid curve when  $\tau_S = 1$ . The corresponding proportion of  $S_a$  phase cells is also shown as the hollow circles (black for  $\tau_S = 9$  and magenta for  $\tau_S = 1$ ) and the corresponding proportion of  $S_b$  phase cells is shown as the asterisks (black for  $\tau_S = 9$  and magenta for  $\tau_S = 1$ ) respectively.*

spent in the  $M_a$  and  $M_b$  phases, and the phase diagram is not significantly altered. The fact that the values of  $\tau_S$  and  $\tau_M$  do not significantly influence steady-state relationships between the doubling time and the corresponding cell cycle phase proportion is in line with the original definition of the storage phases. The storage phases were constructed simply to control the distribution with which cells left a phase, but not measurements such as the average duration of a phase. However, the storage phase durations are still an important aspect of our model, and their effects on model behaviour will become more apparent when we consider changes in the cell cycle phases in the next chapter.

Let's investigate how the phase diagram changes when we change the base variables. What if the base doubling time was, say, 30 hours instead of 24 hours? The corresponding phase diagram is shown in Figure 4.5. It is clear that this alteration in the base doubling time has significant effects on the phase diagram. Similarly, choosing different G S or M phase values for the base variables produces significantly different phase diagrams.

Although the phase diagram is not sensitive to changes in the  $\tau_S$  and  $\tau_M$  variables, we find that altering the values of  $\tau_{G_a}$  or  $\tau_{G_c}$  does not leave the phase diagram unchanged. We find that changing the value of these variables has a significant impact on the value of the

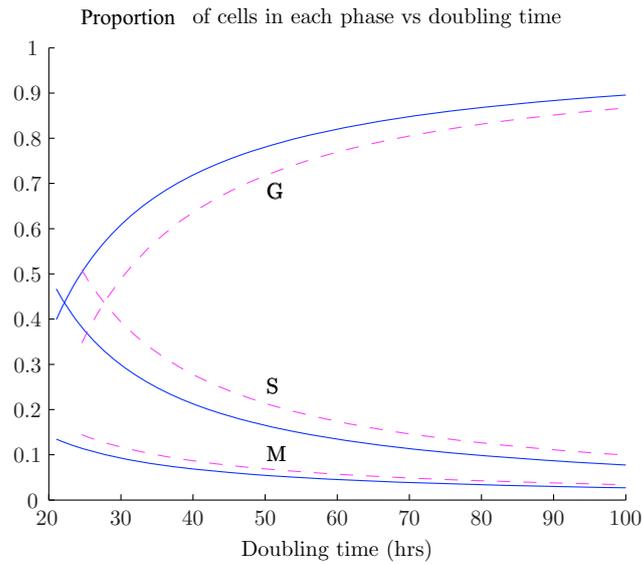


Figure 4.5: *Percent of cells across the cell cycle phases versus the doubling time of the population. The solid blue curve represents the phase diagram for when the base doubling time is 24 hours. The dashed magenta curve represents that when the base doubling time is 30 hours.*

*minimum doubling time* of the population. Figure 4.6 shows the phase diagram for varying values of  $\tau_{G_a} + \tau_{G_c}$ . As we can see, different values for  $\tau_{G_a} + \tau_{G_c}$  have a strong influence on the value of the minimum doubling time, and hence how far the phase diagram extends, but have little influence on the remainder of the phase diagram, in the same way that  $\tau_S$  and  $\tau_M$  do not significantly influence the shape of the phase diagram. The reason  $\tau_{G_a}$  and  $\tau_{G_c}$  extend the limits of the phase diagram is because these variables influence the value of the minimum doubling time. This can be seen by considering equation (3.30), which shows that  $\gamma$  (which we take to infinity when we calculate the minimum doubling time) depends on the storage phase durations  $\tau_{G_a}$  and  $\tau_{G_c}$ , but not  $\tau_S$  or  $\tau_M$ .

This investigation has provided us with further insight into how the variables in our system influence model behaviour. This investigation will prove even more useful later in Section 6.3, when we must produce a phase diagram for the T47D cell line to investigate the validity of certain experiments. We will come back to phase diagrams again in Section 6.3.

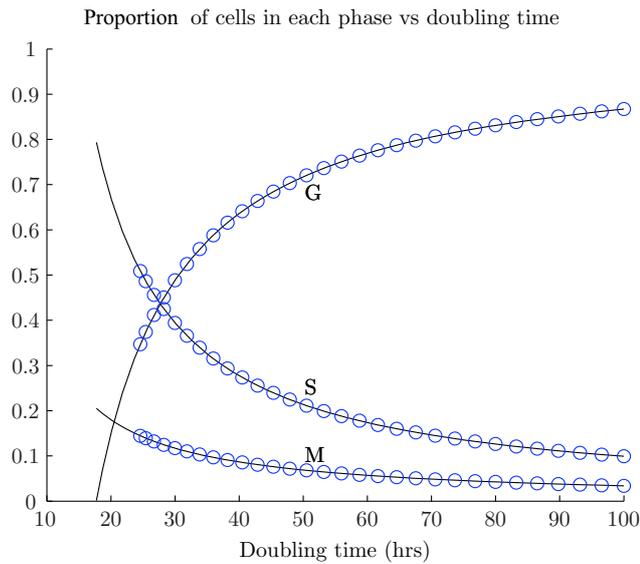


Figure 4.6: Phase diagram for  $\tau_{G_a} + \tau_{G_c} = 0$  (black line) and  $\tau_{G_a} + \tau_{G_c} = 6.75$  (blue circles).

#### 4.4 The doubling time and the average cell cycle duration

We now use equations (4.1) - (4.3) to discuss some misconceptions in the literature regarding the relationship between the doubling time of a cell population and the average cell cycle duration. Often, the phase steady-state values of  $G, S, M$  and  $\rho$  are used to make conclusions about the average duration of a cell cycle phase. For instance, in [115], cells growing exponentially were said to have a doubling time of 24 hours and a steady-state S phase percentage of 39.4%. It was then concluded that the average duration of the S phase in MCF-7 cells is  $0.394 \times 24 = 9.5$  hours. Similar conclusions have been made in [80], [83] and in [5].

The commonly used formula for the S phase duration can be stated as  $S \times \text{doubling time} = \text{average S phase duration}$ , or as  $S \times \frac{\ln 2}{\rho} = \text{average S phase duration}$ . As the average time a cell spends in the S phase is an experiment-independent value, then the relationship just described implies that  $S$  is proportional to  $\rho$ , a clear disagreement with equation (4.2) where we can see the relationship between  $S$  and  $\rho$  is more complex. In this section, we will discuss why this simple formula is incorrect.

In addition to the simple formula for the S phase duration (stated in the previous paragraph as  $S \times \frac{\ln 2}{\rho}$ ), we can derive similar formulae for the average G and M phase

durations which can be expressed as  $G \times \frac{\ln 2}{\rho}$  and  $M \times \frac{\ln 2}{\rho}$  respectively. By adding the formulae for the G, S and M phase durations together, and noting that the sum of the average duration of the cell cycle phases is equivalent to the average cell cycle phase duration, we have

$$\begin{aligned} & \text{the average cell cycle phase duration} = \\ & G \times \frac{\ln 2}{\rho} + S \times \frac{\ln 2}{\rho} + M \times \frac{\ln 2}{\rho} = (G + S + M) \times \frac{\ln 2}{\rho} = \frac{\ln 2}{\rho}. \end{aligned}$$

Thus, the commonly assumed formula implies that the doubling time of the population is equivalent to the average cell cycle duration. However this is not the case, as we now show.

We do this by using our model equations to calculate the average duration of the G, S and M phases for the MCF-7 cell line. The average duration of the S phase using the model results is calculated by summing the average duration of the  $S_a$  phase with that of the  $S_b$  phase. Cells will always spend exactly  $\tau_S$  hours in the  $S_a$  phase, and so all that remains is to calculate the average duration of the  $S_b$  phase. From equation (3.7e) we have

$$\frac{dN_{S_b}(t)}{dt} = \gamma(t - \tau_{G_c} - \tau_S)N_{G_b}(t - \tau_{G_c} - \tau_S) - \alpha N_{S_b}(t).$$

When considering the average amount of time a cell spends in a phase, we are only concerned with the rate at which they leave the phase, and so we are not concerned with the rate of entry into the phase. Hence we consider the differential equation

$$\frac{dN_{S_b}(t)}{dt} = -\alpha N_{S_b}(t), \quad (4.5)$$

which has solution

$$N_{S_b}(t) = N_{S_b}(0)e^{-\alpha t}.$$

If we divide both sides by  $N_{S_b}(0)$  then we have

$$\frac{N_{S_b}(t)}{N_{S_b}(0)} = e^{-\alpha t},$$

where the RHS is the complementary distribution function of the well known exponential

distribution. We can calculate the mean lifetime of the exponential distribution using

$$\begin{aligned} E(t) &= \int_0^{\infty} te^{-\alpha t} \\ &= \frac{1}{\alpha}. \end{aligned}$$

Thus, the average duration of the  $S_b$  phase is  $\frac{1}{\alpha}$ , and so we calculate the average duration of the S phase as  $\tau_S + \frac{1}{\alpha} = 10.5$  hours. Similarly, the average duration of the M phase is  $\tau_M + \frac{1}{\beta} = 3.8$  hours and the average duration of the G phase is  $\tau_{G_a} + \frac{1}{\gamma} + \tau_{G_c}$  which depends upon the value of  $\gamma$ , as would be expected. We also note that our expressions for the average duration of the S and M phases only depend upon experiment-independent parameters, and hence will not change value when environmental signals change.

We can use these expressions to determine the average duration of the entire cell cycle as simply the sum of the average time spent in each of the cell cycle phases, and can be expressed as

$$\tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M + \frac{1}{\gamma} + \frac{1}{\alpha} + \frac{1}{\beta}. \quad (4.6)$$

We wish to compare this expression with the doubling time. If we re-arrange equation (4.4), we can write the following expression

$$\begin{aligned} 2e^{-\rho(\tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M)} &= \left(\frac{\rho}{\alpha} + 1\right)\left(\frac{\rho}{\beta} + 1\right) + \frac{\rho}{\gamma}\left(\frac{\rho}{\alpha} + 1\right)\left(\frac{\rho}{\beta} + 1\right) \\ &= \left(\frac{\rho}{\gamma} + 1\right)\left(\frac{\rho}{\alpha} + 1\right)\left(\frac{\rho}{\beta} + 1\right), \end{aligned} \quad (4.7)$$

which, after some more manipulation, can be written as

$$\begin{aligned} \frac{\ln 2}{\rho} &= \tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M + \frac{\ln\left(\left(\frac{\rho}{\alpha} + 1\right)\left(\frac{\rho}{\beta} + 1\right)\left(\frac{\rho}{\gamma} + 1\right)\right)}{\rho} \\ &= \tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M + \frac{\ln\left(\frac{\rho}{\alpha} + 1\right)}{\rho} + \frac{\ln\left(\frac{\rho}{\beta} + 1\right)}{\rho} + \frac{\ln\left(\frac{\rho}{\gamma} + 1\right)}{\rho}. \end{aligned} \quad (4.8)$$

On the LHS of equation (4.8), we have the doubling time, and if we compare the RHS of this equation to the expression in (4.6), we note that the expressions are different, and hence the average cell cycle duration is not the same as the doubling time. We can further show that the average cell cycle duration is greater than or equal to the doubling time of

the population.

We begin by showing that  $\ln(x + 1) \leq x$  for  $x \geq 0$ . When  $x = 0$ ,  $\ln(x + 1) = x$ . For  $x > 0$ ,  $\ln(x + 1)$  has a smaller gradient than  $x$  (which can be confirmed by finding the derivative of both), and so  $\ln(x + 1) < x$  when  $x > 0$ . Substituting  $x = \frac{\rho}{\gamma}$  gives  $\ln(\frac{\rho}{\gamma} + 1) \leq \frac{\rho}{\gamma}$  which can be re-arranged to give

$$\frac{\ln(\frac{\rho}{\gamma} + 1)}{\rho} \leq \frac{1}{\gamma},$$

with equality only when  $\frac{1}{\gamma} = 0$ . Similar inequalities can be derived for the remaining terms in equation (4.8), allowing us to write

$$\begin{aligned} \frac{\ln 2}{\rho} &= \tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M + \frac{\ln(\frac{\rho}{\alpha} + 1)}{\rho} + \frac{\ln(\frac{\rho}{\beta} + 1)}{\rho} + \frac{\ln(\frac{\rho}{\gamma} + 1)}{\rho} \\ &\leq \tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M + \frac{1}{\alpha} + \frac{1}{\beta} + \frac{1}{\gamma}. \end{aligned}$$

Thus, the average cell cycle duration is always greater than or equal to the doubling time. Equality occurs only when progression through the non-storage phases is instantaneous, and in this case, the cell cycle duration is deterministic.

### Other calculations of the average cell cycle duration from the literature

Our result, that the average cell cycle duration is always greater than or equal to the doubling time, is not in agreement with results from [28], which conclude that the average cell cycle duration is *less* than the doubling time. In [28], a mathematical model was developed to describe cell cycle progression, and they derived their expression for the average cell cycle duration using their model equations. The equations used to describe cell cycle progression are not the same as our model equations, implying that the difference could come from different equations being used to represent cell cycle progression. However, we show that even with the model equations used in [28], the average cell cycle duration is greater than the doubling time. The equations shown in (4.9) are essentially reproduced

from [28] (in the absence of cell differentiation)

$$\begin{aligned}\frac{dN_G(t)}{dt} &= 2\beta N_M(t) - \gamma N_G(t), \\ \frac{dN_S(t)}{dt} &= \gamma N_G(t) - \alpha N_S(t), \\ \frac{dN_M(t)}{dt} &= \alpha N_S(t) - \beta N_M(t).\end{aligned}\tag{4.9}$$

In [28], it was noted that the average time a cell spends in the G, S and M phases depends upon the overall rate of cell growth (which we refer to as  $\lambda$ , as this is the notation used in [28]) *as well as* the rate at which cells leave their respective phases. Specifically, their calculated average G, S and M phase proportions are expressed as  $\frac{1}{\gamma+\lambda}$ ,  $\frac{1}{\alpha+\lambda}$  and  $\frac{1}{\beta+\lambda}$ , respectively. We will show that it is not possible for the average cell cycle phase durations to depend both on their rate of leaving and the overall proliferation rate,  $\lambda$ .

Consider two separate cell populations, call them pop1 and pop2, each growing in phase steady-state (phase steady-state is referred to as balanced exponential growth (BEG) in [28]). Let's also assume that in pop1, the value of  $\gamma$  is greater than that in pop2, while  $\alpha$  and  $\beta$  are the same in both populations. The overall rate of growth will be higher in pop1, as cells transition through the G phase faster due to the higher value of  $\gamma$ , and hence the overall growth rate (i.e. the value of  $\lambda$ ) will be different in these two populations. If  $\lambda$  is different, then according to the expression for the average S and M phase durations in [28], the average duration of both the S and M phases will be different in the two different populations of cells, because their formulas depend upon  $\lambda$ .

However, as the values of  $\alpha$  and  $\beta$  are the same in both populations, we do not see why cells in the S and M phases should behave differently just because the rate of transition through the G phase is different in the two populations. Cells in the S and M phases should transition through their respective phases at rates only dependent upon  $\alpha$  and  $\beta$  respectively. We proceed to provide further evidence of this claim.

### **How the average duration in each of the phases is calculated in [28]**

In [28], the derivation of the average residence time was not provided (although the proposed value of the average residence times were reported in equations (15)-(17) on

page 567 of their paper, and listed near the bottom of page 103 of this thesis), and the reader was directed to consider the age-structured model of cell cycle progression presented in [9]. There was no derivation of the average residence times in this paper, however we were able to derive the residence times from the expressions presented in this paper.

Near the top of page 1680 in [9], the authors derived an expression for the age distribution of cells in each of the phases. If we apply this expression to the model presented here in equations (4.9), then the number of cells in the G phase at time  $t$  with age  $\tau$  can be expressed as

$$n_G(t, \tau) = N_G(\lambda + \gamma)e^{\lambda t}e^{-\tau(\lambda+\gamma)}.$$

We can determine the average age of cells in this phase at time  $t$  by first converting  $n_G(t, \tau)$  into a probability distribution and then calculating the mean of this distribution. To convert it into a probability distribution, we must determine the total number of cells in the G phase at each point in time  $t$ . The total number of cells in the G phase can be calculated by evaluating

$$\int_0^\infty n_G(t, \tau)d\tau = N_Ge^{\lambda t}.$$

Thus  $\frac{n_G(t, \tau)}{N_Ge^{\lambda t}} = (\lambda + \gamma)e^{-\tau(\lambda+\gamma)}$  is a probability density function. This is clearly the density of the exponential distribution with parameter  $\lambda + \gamma$ , and thus has mean  $\frac{1}{\lambda+\gamma}$ . What we have just calculated is the *current average age* of cells in the G phase. As this formula is identical to the one proposed in [28] for the average G phase duration, we suspect that this is how the average G phase duration was derived. However, we claim that the mean *age* of cells in a phase at time  $t$  is not equivalent to the average time a cell spends in this phase. This is because the G phase population is growing (when in phase steady-state, this population is growing exponentially), and so at any point in time, there are more cells entering than leaving. Thus, the average duration of a cell cycle phase will be greater than the average age of cells in that phase.

We believe that it is difficult to consider values such as the average time a cell spends in a phase using continuous differential equations, and so we endeavoured to create a discrete simulation whose limiting model can be represented by equations (4.9). Using the simulation, we are able to further show that the average time a cell spends in each phase is indeed the reciprocal of the rate at which it leaves the phase, and is independent

of the proliferation rate,  $\lambda$ . We will also show that if we run this simulation under some seemingly intuitive (but incorrect) assumptions, we get biased estimates of the average cell cycle phase durations that compare well with the formulae presented in [28]. Thus, it is possible that such a simulation was performed in [28], providing them with the same answer as the method of calculating the average age of cells in each phase as discussed in the previous paragraph. However, removing these incorrect assumptions provides estimates of the average cell cycle phase durations that are simply the inverse of the rate at which cells leave their phase (in agreement with the expression provided in this thesis, and also in agreement with [9] on page 1678). The simulation is discussed in detail in the following section.

### Discrete Simulation

In [28], an explicit formula for the average duration of the G, S and M phases was produced. In agreement with the notation used in equations (4.9), the average G, S and M phase durations proposed in [28] are  $\frac{1}{\gamma+\lambda}$ ,  $\frac{1}{\alpha+\lambda}$  and  $\frac{1}{\beta+\lambda}$  respectively, where  $\lambda$  is the overall proliferation rate. However, on page 101 of this thesis, we concluded that the average G, S and M phase durations can be expressed as  $\frac{1}{\gamma}$ ,  $\frac{1}{\alpha}$  and  $\frac{1}{\beta}$  respectively. We will perform a discrete simulation of events and calculate what the average duration is, and compare with these two hypotheses.

We start with cells in each of the three phases, and the number of cells in the phases is chosen so that the population is in phase steady-state. As each cell is experiencing a constant rate of leaving its current phase, we use the exponential probability distribution to determine at what times each of the cells will be leaving.

For each phase we assign two vectors. One is called the birth vector, which contains the entry times for each of the cells currently residing in that phase, and the other is called the leaving vector, which contains the times at which the cells are calculated to leave. For instance, if we start with 10 cells in the G phase, then we can create a birth vector of length 10 which denotes the time which these cells entered this phase. At the start of the simulation, we assume that all 10 cells were born at time  $t = 0$  hours, and so we start with a vector of zeros as the birth vector. We then generate the leaving vector, whose elements are generated based on a cell's rate of leaving the G phase. As we can see

from equations (4.9), cells leave the G phase at a constant rate  $\gamma$ , and so G phase cells have exponentially distributed residence times. We calculate the time each cell leaves the G phase by evaluating  $t_{leaving} = -\frac{\log(rand)}{\gamma}$ , where  $rand$  is a uniformly distributed random number between 0 and 1. This provides a random vector of leaving times representing the times at which each cell in the G phase moves into the S phase. The same is done for the S and M phases (i.e. a birth vector and a leaving vector is generated for cells in the S and M phases as well). We note that initially, choosing a birth vector for each of the cells in the phases containing all zeroes is unrealistic. However, if we let the simulation run for long enough, the initial bias in choosing these birth times will become insignificant.

The simulation then proceeds as follows. The smallest time over all leaving vectors and over all cell cycle phases is chosen, and the cell that corresponds to that leaving time is chosen to move into the following phase. Let's assume that the cell with the earliest leaving time comes from the S phase. The residence time of that cell is calculated by subtracting its birth time from its leaving time, and is recorded in a vector representing the residence times of S phase cells. The birth and leaving time for that cell is deleted from the birth and leaving vectors in the S phase, and a new birth time is created in the M phase (which is equal to the cell's old leaving time). A leaving time is then generated for this new M phase cell which is equal to  $-\frac{\log(rand)}{\beta}$ . We note that if the cell chosen to leave was an M phase cell, then two cells would have been created in the G phase, each with the same birth time, and with leaving times according to  $-\frac{\log(rand)}{\gamma}$  (the leaving times may be different for each daughter cell because the value of  $rand$  is chosen randomly).

Once the vectors have been updated, the simulation is repeated - the earliest leaving time is determined, and the cell with that leaving time is chosen to move into the subsequent phase. All the vectors are updated, and the residence time for that cell is recorded. The duration of each cell cycle phase can be calculated by finding the mean of the residence vectors for each phase, and the average cell cycle duration is simply the sum of average G, S and M phase durations.

It may seem intuitive that to determine the average residence time of each of the three phases (and consequently the average cell cycle duration), we can run the simulation described above for a sufficiently long period of time, and then stop the simulation and consider the residence times we have thus far collected. However, these residence times

will be values from cells that have *completed* their times in any phase. This calculation will likely be biased, as there are many cells which have not had the chance to complete their time in a phase, and these cells will likely have longer durations in their corresponding phase (because they have not yet left, whereas those recorded up until now have left their phase). We claim that to get a true picture of the average time each cell spends in the phases, we should designate a stop time for which cells can be considered, and then proceed with the simulation to allow those remaining cells to exit the phases they are currently residing in.

Let's investigate what happens when we run the above simulation, and then halt it after a sufficiently long period of time. We will consider two cases: one case when we consider only residence times we have collected by the stop time, and the case when we continue to allow remaining cells to exit their current phases. We allow the simulation to run long enough so that about 2000 events occur in each phase (that is, about 2000 cells have completed the G phase, 2000 completed the S phase and 2000 completed the M phase, and thus we have collected 2000 completed residence times for each phase, sufficient to accumulate a reasonably accurate average residence time). When we stop the simulation, we find that there are just over 2000 cells remaining, distributed across the phases. The average residence times collected for the G phase (when we ignore the cells still remaining in the G phase) are shown in Figure 4.7 as the 'biased average G phase duration'. The overall cell cycle duration (when we ignore cells remaining in the system after the stop time) is also shown in Figure 4.8 as the 'biased average cell cycle duration'. We note that in these simulation results, we consider varying values of  $\gamma$  only, however we get similar results if we vary the values of  $\alpha$  and  $\beta$  as well. We also note that we obtain similar results for the calculated S and M phase durations (not shown).

These representations show that the formulae for the average cell cycle durations from [28] compare well with the calculated average phase durations. Thus, such a simulation could have been performed by Daukste et al. [28] and could be the method they used to determine their expressions for the average cell cycle phase durations. However, these simulation results are biased because we have ignored cells that have not yet been given the full opportunity to complete their phase. In fact, we can calculate the mean residence times of the cells that are still remaining in the G phase after we stop the simulation. The average residence times of *just* cells that are still in the G phase at the end of the

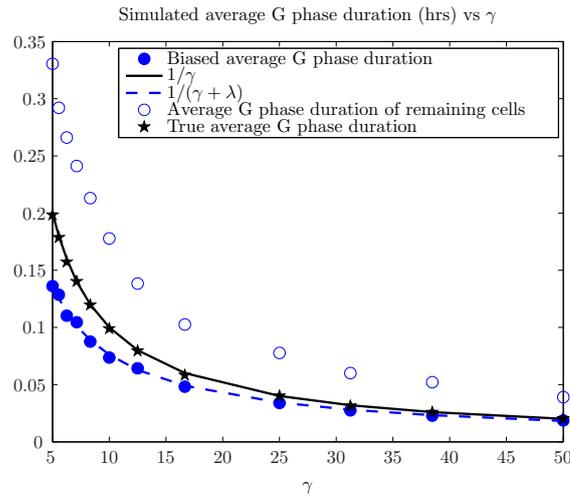


Figure 4.7: The simulated biased average G phase duration for varying values of  $\gamma$  is shown as the filled blue dots, when we only consider cells that have already completed the G phase by the simulation stop time. The hollow dots represent the average G phase residence times for cells that were remaining in the simulation after the stop time. The black stars represent the true average G phase duration. Solid black and dashed blue curves represent our formula ( $\frac{1}{\gamma}$ ) and the formula from [28] ( $\frac{1}{\gamma+\lambda}$ ) respectively. In this simulation, we have fixed  $\alpha = 10$  and  $\beta = 20$ .

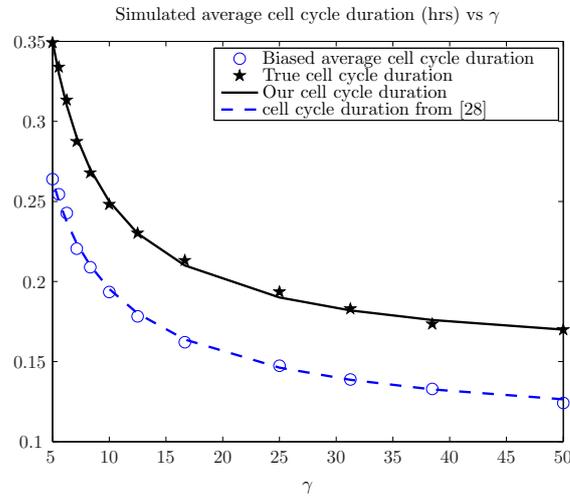


Figure 4.8: The biased average cell cycle duration is calculated by summing the biased average duration of the G, S and M phases and is shown here as the hollow blue dots. The true average cell cycle duration is calculated by summing the true G, S and M phase durations and is shown here as the black stars. The solid black and dashed blue curves represent our formula ( $\frac{1}{\gamma} + \frac{1}{\alpha} + \frac{1}{\beta}$ ) and the formula from [28] ( $\frac{1}{\gamma+\lambda} + \frac{1}{\alpha+\lambda} + \frac{1}{\beta+\lambda}$ ) respectively. In this simulation, we have fixed  $\alpha = 10$  and  $\beta = 20$ .

simulation is also shown in Figure 4.7 as the hollow blue circles. We can see that the average residence times of the remaining cells is significantly larger than the average residence times of cells that have already completed the G phase by the stop time. It may be tempting to think that if we chose a later stop time (i.e. if we allow the simulation to continue for a longer period of time), we may eliminate the bias. However this is not the case - the bias remains no matter how long the simulation is run for! This is because the longer the simulation runs for, the more cells are generated and the more cells will be remaining once we stop the simulation. This large number of cells continues to contribute to the bias.

Taking the true mean of the residence times by taking into account the residence times of G phase cells that are still in the G phase after the simulation stop time is shown in Figure 4.7 as the black asterisks. The true average cell cycle duration when we take into account these remaining cells is also shown in Figure 4.8 as the black asterisks. We can see that the true residence times compare very well with the formulae we presented in this thesis. We conclude that ignoring the cells that are still remaining in the phases introduces a bias in the calculated average residence times, because cells that exited their phases before the simulation stop time are counted, but cells that have a longer residence time are not counted. This means that those cells with relatively short residence times have a higher chance of being included in the calculation of the average residence time, while those with long residence times have a smaller chance of being included in this calculation, making the average appear shorter than its true value. The true average residence times of the G, S and M phases can be expressed as  $\frac{1}{\gamma}$ ,  $\frac{1}{\alpha}$  and  $\frac{1}{\beta}$  respectively, and the average cell cycle duration is the sum of these times, which is greater than the doubling time of the population, as discussed in Section 4.3.

**Proof that by counting only some of the average residence times, we arrive at the formula presented in [28]**

We can prove that by only considering the residence times of cells that have left by some time,  $T$ , that the formula for the average residence times in a phase for those cells is equivalent to the formula presented in [28]. This is useful, because it shows that the calculated residence times can be skewed if we censor the residence times of certain cell

populations. We will use the equations representing cell cycle progression from [28] (shown in this thesis in equations (4.9)) to show this.

We consider the residence times of cells in one phase only, the S phase. We note however that the proof can be applied to calculate the residence times of cells in the G and M phases similarly.

From equation (4.9), we note that cells leave the S phase at rate  $\alpha N_S(t) = \alpha S_0 e^{\lambda t}$ , where  $S_0$  is the number of S phase cells at time  $t = 0$ . Therefore, the probability that a cell which entered at time  $t$  leaves before time  $T$  is  $1 - e^{-\alpha(T-t)}$ . Thus, the density of the single cell leaving at time  $t + x$ , and hence having a residence time of  $x$  can be expressed as  $\alpha e^{-\alpha x}$ . Given that we have this expression, and recalling from equation (4.9) that the rate at which cells enter the S phase is  $\gamma N_G(t) = \gamma G_0 e^{\lambda t}$ , we determine the *total* average residence time of cells that have left the S phase by time  $T$ , denoted  $Life_S$ , as

$$\begin{aligned}
Life_S &= \int_{t=0}^T \gamma G_0 e^{\lambda t} \int_{x=0}^{T-t} x \alpha e^{-\alpha x} dx dt \\
&= \gamma G_0 \int_{t=0}^T e^{\lambda t} \left( - \int_{x=0}^{T-t} x \frac{d}{dx} e^{-\alpha x} dx \right) dt \\
&= \gamma G_0 \int_{t=0}^T e^{\lambda t} \left( (t-T)e^{\alpha(t-T)} - \frac{1}{\alpha} e^{\alpha(t-T)} + \frac{1}{\alpha} \right) dt \quad (\text{integration by parts}) \\
&= \gamma G_0 \int_{t=0}^T e^{-\alpha T} t e^{(\lambda+\alpha)t} - \left(T + \frac{1}{\alpha}\right) e^{-\alpha T} e^{(\lambda+\alpha)t} + \frac{1}{\alpha} e^{\lambda t} dt \\
&\quad \vdots \\
&= \gamma G_0 \left( \frac{1}{\lambda \alpha} (e^{\lambda T} - 1) - \frac{1}{\lambda + \alpha} \left( \frac{1}{\alpha} + \frac{1}{\lambda + \alpha} \right) (e^{\lambda T} - e^{-\alpha T}) \right).
\end{aligned}$$

As this is the *total* average residence time of S phase cells that have entered and left between times  $t = 0$  and  $t = T$ , we can calculate the expected residence time of any one cell by dividing the above expression by the total number of cells that have entered and left the S phase between times  $t = 0$  and  $t = T$ , which we denote by  $ns(T)$ . We write

$$\begin{aligned}
ns(T) &= \int_{t=0}^T \gamma G_0 e^{\lambda t} (1 - e^{-\alpha(T-t)}) dt \\
&= \gamma G_0 \left( \int_{t=0}^T e^{\lambda t} - e^{-\alpha T} \int_{t=0}^T e^{(\lambda+\alpha)t} dt \right) \\
&= \gamma G_0 \left( \frac{1}{\lambda} (e^{\lambda T} - 1) - \frac{1}{\lambda + \alpha} (e^{\lambda T} - e^{-\alpha T}) \right).
\end{aligned}$$

We can thus calculate the average S phase duration by evaluating  $\frac{Life_S}{ns(T)}$ . If we consider this for very large  $T$

$$\begin{aligned}
\frac{Life_S}{ns(T)} &= \frac{\gamma G_0 \left( \frac{1}{\lambda \alpha} (e^{\lambda T} - 1) - \frac{1}{\lambda + \alpha} \left( \frac{1}{\alpha} + \frac{1}{\lambda + \alpha} \right) (e^{\lambda T} - e^{-\alpha T}) \right)}{\gamma G_0 \left( \frac{1}{\lambda} (e^{\lambda T} - 1) - \frac{1}{\lambda + \alpha} (e^{\lambda T} - e^{-\alpha T}) \right)} \\
&\approx \frac{\frac{1}{\lambda \alpha} e^{\lambda T} - \frac{1}{\lambda + \alpha} \left( \frac{1}{\alpha} + \frac{1}{\lambda + \alpha} \right) e^{\lambda T}}{\frac{1}{\lambda} e^{\lambda T} - \frac{1}{\lambda(\lambda + \alpha)} e^{\lambda T}} \quad (\text{for very large } T) \\
&\approx \frac{\frac{1}{\lambda \alpha} - \frac{1}{\alpha} \frac{1}{\lambda + \alpha} \frac{\lambda + 2\alpha}{\lambda + \alpha}}{\frac{\alpha}{\lambda(\lambda + \alpha)}} \\
&\approx \frac{\frac{(\lambda + \alpha)^2 - \lambda(\lambda + 2\alpha)}{\lambda \alpha (\lambda + \alpha)^2}}{\frac{\alpha}{\lambda(\lambda + \alpha)}} \\
&\approx \frac{1}{\lambda + \alpha}.
\end{aligned}$$

Thus, if we restrict the S phase cells that we are considering to cells that have left this phase by some time  $T$ , then we arrive at the expression for the average S phase duration presented in [28]. This exercise shows that it is important not to censor certain cell populations when calculating the average cell cycle phase durations.

## 4.5 A novel representation of the model in equations (3.7)

Graeme Wake [128], a reviewer of the paper we submitted [108], suggested a novel representation of our cell cycle model from equations (3.7). We will discuss this representation in this section.

We begin by defining the notation

$$[\tau] = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \tau_6 \\ \tau_7 \end{bmatrix} \text{ and } \mathbf{x}(t - [\tau]) = \begin{bmatrix} x_1(t - \tau_1) \\ x_2(t - \tau_2) \\ x_3(t - \tau_3) \\ x_4(t - \tau_4) \\ x_5(t - \tau_5) \\ x_6(t - \tau_6) \\ x_7(t - \tau_7) \end{bmatrix}.$$

The model in equations (3.7) can be written in the following form.

$$\frac{d\mathbf{N}(t)}{dt} = A\mathbf{N}(t) + B\mathbf{N}(t - [\delta]) + C\mathbf{N}(t - [\omega]), \quad (4.10)$$

where

$$\mathbf{N}(t) = \begin{bmatrix} N_{G_a}(t) \\ N_{G_b}(t) \\ N_{G_c}(t) \\ N_{S_a}(t) \\ N_{S_b}(t) \\ N_{M_a}(t) \\ N_{M_b}(t) \end{bmatrix}, \quad [\delta] = \begin{bmatrix} 0 \\ \tau_{G_c} \\ 0 \\ 0 \\ \tau_M \\ 0 \\ \tau_{G_a} \end{bmatrix}, \quad [\omega] = \begin{bmatrix} 0 \\ \tau_{G_a} + \tau_S \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 2\beta \\ 0 & -\gamma & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\beta \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -2\beta \\ 0 & 0 & 0 & 0 & 0 & 0 & 2\beta \\ 0 & -\gamma(t - \tau_{G_a}) & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma(t - \tau_{G_a}) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha & 0 & 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\gamma(t - \tau_{G_a} - \tau_S) & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma(t - \tau_{G_a} - \tau_S) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

It can be confirmed by expanding the expression in equation (4.10) that this representation produces equations (3.7).

We will use Wake's set-up [128] described above to show that when  $\gamma$  is constant in time (i.e. when cells are growing in phase steady-state), the rate of exponential growth (which we have called  $\rho$  throughout this thesis) as well as the steady state proportion of cells across the phases is unique, and equivalent to previous calculations we have performed throughout this thesis.

Let's assume that our cells are growing in phase steady-state. In this case,  $\gamma$  becomes constant, and we can drop the time dependence of  $\gamma$  in matrices  $B$  and  $C$ . We then look for solutions of equation (4.10) in phase steady-state of the form  $\mathbf{N}(t) = \mathbf{c}e^{\lambda t}$ , where  $\mathbf{c}$  is a vector containing elements independent of time. We let  $\mathbf{c} = [c_1, c_2, \dots, c_7]$  such that  $c_1 + c_2 + \dots + c_7 = 1$ .

We wish to determine what possible values of  $\lambda$  will work in our set of equations so that we can confirm that there is a unique solution. We have previously defined  $\mathbf{N}(t) = \mathbf{b}e^{\rho t}$  to be the only solution to equations (3.7), where  $\mathbf{b} = [G_a, G_a, \dots, M_b]^t$ . However, we never proved that  $\rho$  was unique. Using Wake's approach [128], we show that the only possible positive  $\lambda$  that satisfies  $\mathbf{N}(t) = \mathbf{c}e^{\lambda t}$  corresponds to the value of  $\rho$  we have discussed earlier.

Substituting  $\mathbf{N}(t) = \mathbf{c}e^{\lambda t}$  into equation (4.10) gives

$$\lambda \mathbf{c}e^{\lambda t} = \mathbf{A} \mathbf{c}e^{\lambda t} + \mathbf{B} \mathbf{c}e^{\lambda(t - \lceil \delta \rceil)} + \mathbf{C} \mathbf{c}e^{\lambda(t - \lceil \omega \rceil)}, \quad (4.11)$$

where

$$[[\tau]] = \begin{bmatrix} [\tau]' \\ [\tau]' \\ [\tau]' \\ [\tau]' \\ [\tau]' \\ [\tau]' \\ [\tau]' \end{bmatrix}.$$

For example,

$$[[\omega]] = \begin{bmatrix} 0 & \tau_{G_a} + \tau_S & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_{G_a} + \tau_S & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_{G_a} + \tau_S & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_{G_a} + \tau_S & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_{G_a} + \tau_S & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_{G_a} + \tau_S & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_{G_a} + \tau_S & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Equation (4.11) can be simplified to

$$\lambda I - A - B e^{-\lambda[[\tau_1]]} - C e^{-\lambda[[\tau_2]]} = 0.$$

We proceed to find the characteristic equation by taking the determinant of this expression:

$$\begin{vmatrix} \lambda & 0 & 0 & 0 & 0 & 2\beta e^{-\lambda\tau_{G_a}} - 2\beta \\ 0 & \gamma + \lambda & 0 & 0 & 0 & -2\beta e^{-\lambda\tau_{G_a}} \\ 0 & \gamma e^{-\lambda\tau_{G_c}} - \gamma & \lambda & 0 & 0 & 0 \\ 0 & \gamma e^{-\lambda(\tau_{G_c} + \tau_S)} - \gamma e^{-\lambda\tau_{G_c}} & 0 & \lambda & 0 & 0 \\ 0 & -\gamma e^{-\lambda(\tau_{G_c} + \tau_S)} & 0 & 0 & \alpha + \lambda & 0 \\ 0 & 0 & 0 & 0 & \alpha e^{\lambda\tau_M} - \alpha & \lambda \\ 0 & 0 & 0 & 0 & -\alpha e^{\lambda\tau_M} & 0 & \beta + \lambda \end{vmatrix} = 0 \quad (4.12)$$

Expanding this expression using MATLAB gives the following characteristic equation,

which we think of, in an algebraic sense, as a nonlinear eigenvalue problem

$$\lambda^4((\lambda + \alpha)(\lambda + \beta)(\lambda + \gamma) - 2\alpha\beta\gamma e^{-\lambda(\tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M)}) = 0. \quad (4.13)$$

Clearly  $\lambda = 0$  is a solution to the nonlinear eigenvalue problem (henceforth, in a slight abuse of notation, referred to as an eigenvalue). If we ignore the  $\lambda^4$  term, then the remaining eigenvalues are solutions to the expression

$$(\lambda + \alpha)(\lambda + \beta)(\lambda + \gamma) - 2\alpha\beta\gamma e^{-\lambda(\tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M)} = 0. \quad (4.14)$$

In the previous section, we derived an equation relating  $\gamma$  and  $\rho$ , shown in equation (4.7). By comparing equation (4.14) with equation (4.7), we see that they are identical, except that in equation (4.7),  $\lambda$  is called  $\rho$ . Thus,  $\rho$  is an eigenvalue, as expected. However we do not know whether there are other solutions to the expression in equation (4.14), and this is something we did not address earlier when we had the same equation involving  $\rho$ .

Wake [128] suggested plotting the functions  $(\lambda + \alpha)(\lambda + \beta)(\lambda + \gamma)$  and  $2\alpha\beta\gamma e^{-\lambda(\tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M)}$ . The intersection of these two functions represents the real solutions to equation (4.14), and hence the remaining real eigenvalues of our characteristic function. Figure 4.9 shows these two functions. It is clear that these two functions intersect exactly once for some  $\lambda_i > 0$ . Thus the only real eigenvalues are  $\lambda = 0$  and  $\lambda = \rho$ . We conclude that either  $\mathbf{N}(t)$  is constant ( $\lambda = 0$ ) or  $\mathbf{N}(t)$  grows exponentially with growth rate  $\rho$ . Thus, this very different approach to dealing with equations (3.7), suggested by Wake, provides us with the same conclusions we made earlier regarding the value of  $\rho$  and its relationship with  $\gamma$ , and can be thought of as a verification to the approaches thus far used.

We have just determined that  $\lambda = \rho$  is the only eigenvalue of interest for our analysis. We can substitute this eigenvalue into equation (4.11) to determine the corresponding eigenvector. Thus we solve

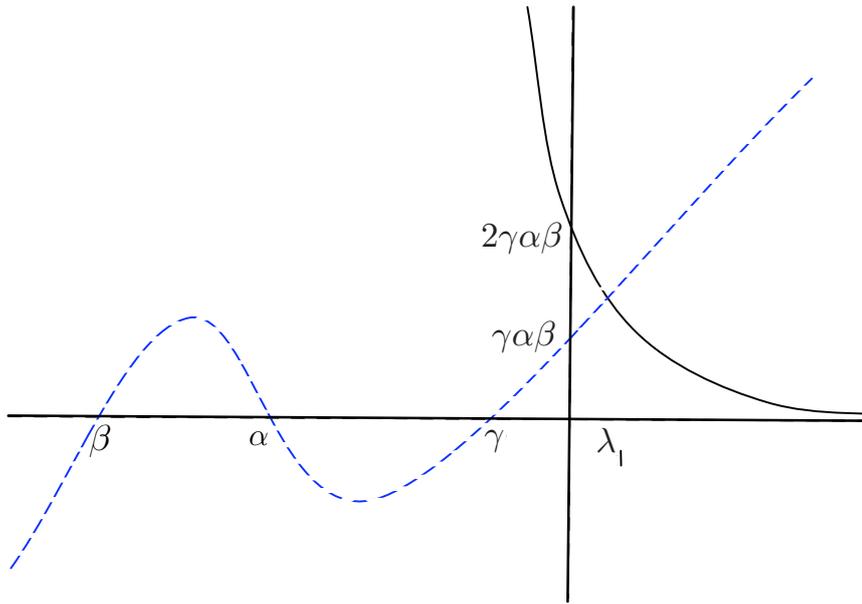


Figure 4.9: The function  $(\lambda + \alpha)(\lambda + \beta)(\lambda + \gamma)$  is shown as the dashed blue curve. The function  $2\alpha\beta\gamma e^{-\lambda(\tau G_a + \tau G_c + \tau S + \tau M)}$  is shown as the solid black curve. The two curves intersect exactly once. The  $x$ -value of the intersection point is labelled  $\lambda_1$  here.

$$\begin{bmatrix} -\rho & 0 & 0 & 0 & 0 & 0 & 2\beta - 2\beta e^{-\rho\tau G_a} \\ 0 & -\gamma - \rho & 0 & 0 & 0 & 0 & 2\beta e^{-\rho\tau G_a} \\ 0 & \gamma - \gamma e^{-\rho\tau G_c} & -\rho & 0 & 0 & 0 & 0 \\ 0 & \gamma e^{-\rho\tau G_c} - \gamma e^{-\rho(\tau G_c + \tau S)} & 0 & -\rho & 0 & 0 & 0 \\ 0 & \gamma e^{-\rho(\tau G_c + \tau S)} & 0 & 0 & -\alpha - \rho & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha - \alpha e^{\rho\tau M} & -\rho & 0 \\ 0 & 0 & 0 & 0 & \alpha e^{\rho\tau M} & 0 & -\beta - \rho \end{bmatrix} \mathbf{c} = 0.$$

If we sum all the rows of the above expression together, we get the following

$$\begin{aligned} \rho(c_1 + c_2 + c_3 + c_4 + c_5 + c_6 + c_7) &= \beta c_7 \\ \Rightarrow \rho &= \beta c_7. \end{aligned}$$

Using this relationship, and row-reducing the above matrix equation using MATLAB gives

$$\mathbf{c} = \begin{bmatrix} 2 - 2e^{-\rho\tau G_a} \\ 2e^{-\rho\tau G_a} - (2 - (c_1 + c_2 + c_3))e^{\rho\tau G_c} \\ (2 - (c_1 + c_2 + c_3))e^{\rho\tau G_c} - (2 - (c_1 + c_2 + c_3)) \\ 2 - (c_1 + c_2 + c_3) - (2 - (c_1 + c_2 + c_3))e^{-\rho\tau S} \\ (2 - (c_1 + c_2 + c_3))e^{-\rho\tau S} - (1 + (c_6 + c_7)) \\ 1 + (c_6 + c_7) - (1 + (c_6 + c_7))e^{-\rho\tau M} \\ (1 + (c_6 + c_7))e^{-\rho\tau M} - 1 \end{bmatrix}. \quad (4.15)$$

The expression in equation (4.15) may look messy, however if we let  $\mathbf{c} = [G_a, G_b, \dots, M_b]^t$  then since  $c_1 + c_2 + c_3 = G$  and  $c_6 + c_7 = M$  we find that this expression is equivalent to the set of equations representing the steady-state proportion of cells in the cell cycle phases determined earlier in Section 3.3 and shown in equations (3.27). Thus, this approach of finding solutions to our cell cycle model in phase steady-state has produced the same set of solutions as the approach we took in Section 3.3.

### Summary of our analysis of equations (3.7) after parameterisation to the MCF-7 cell line

In this chapter, we have developed a model of cell cycle progression. The model uses the concept of storage phases to account for parts of the cell cycle during which cells must always spend a mandatory amount of time, resulting in a set of ordinary delay differential equations which only takes seconds to solve numerically. A unique part of the model is the ability to capture the period when cells are responsive to environmental signals - what we have called the  $G_b$  phase - which is known to occur during some part of the G phase. However, the exact position within this phase has been difficult to define experimentally. Being able to identify when this responsive phase occurs allows for direct connection between the concentrations of internal cell cycle related proteins, which we will discuss in more detail in the following chapter. In this chapter, we also parameterised the model to the MCF-7 cell line. The parameters are chosen after careful consideration of the biological literature, using results from mitotic selection experiments [78] and phase steady-state experiments [112].

The parameterised model is also used to uncover common misconceptions in the literature. In particular, we address the correctness of methods used to calculate the average duration of the cell cycle phases. The formulae commonly used to determine the average duration of cell cycle phases assume that the doubling time of a population is equivalent to the average cell cycle duration. Another paper proposed a different formula for the average duration of the cell cycle, which results in the conclusion that the average duration of the cell cycle is less than the doubling time. However, we use our model equations to show that the average duration of the cell cycle is greater than the doubling time, and we are able to provide values for the average S and M phase durations using our model parameters.

In the next chapter, we show how we can apply our cell cycle model to a range of experiments. We re-visit the experiment on mitotic selection discussed in this chapter and investigate whether our model can predict the cell cycle phase proportions after mitotic selection. We consider other experiments that do not involve changing environmental conditions, and use our model to address the common assumption that a ‘slowly cycling population of cells’ is required to explain their experimental results. Later in the chapter we discuss how we convert changing environmental conditions into a change in the value of  $\gamma(t)$ , and we apply our model to a range of experiments that consider changing environmental conditions. We finish the chapter by extending the model to include the effects of cell death in the presence of chemotherapy drugs.

## Chapter 5

# Applying the parameterised model to experiments involving the MCF-7 breast cancer cell line

In this chapter, we apply the cell cycle model from equations (3.7) to a range of experiments involving the MCF-7 breast cancer cell line. At this point, we still have not discussed how to convert environmental signals into an expression for  $\gamma(t)$  (we do this later in Section 5.3), and so we cannot yet model cell cycle progression in response to changes in the environment. However, we can still apply the model to a variety of interesting situations which do not involve changes in environmental conditions. We begin by re-visiting the mitotic selection experiment from the previous chapter, which involves synchronised cells growing in unchanging environmental conditions.

### 5.1 Applying our model to mitotically selected cells

We wish to simulate cell cycle progression in the mitotic cell experiment discussed in the previous chapter. The results of this experiment are shown in Figure 4.1 on page 85, and they are also reproduced here for ease of reference.

To model mitotic selection, we model the pure population and the impure cell population separately (see Chapter 4 for the definitions of these terms). In both instances,

NOTE:  
 This figure is included on page 120 of the print copy of  
 the thesis held in the University of Adelaide Library.

Figure 5.1: *Cell cycle distribution of mitotically selected cells, reproduced from [78]. Cells were mitotically selected at time 0. Points are the average of at least two independent experiments. The curve through the data points is the result of an unknown curve fitting process.*

we numerically solve the delay differential equations represented in equations (3.7) in Section 3.3. When solving ordinary differential equations, the starting values of all the unknown variables (which includes the 7 model phase values as well as  $\gamma$ ) must be known. However, as we are dealing with *delay* differential equations, we must not only know the values of the variables at time 0, but we must know the values these variables take at times prior to the beginning of the experiment. This is because, once the model begins at time 0, it will be calling on the value of variables which contain delay terms (and hence their values prior to time 0 will be required). As the largest storage phase duration is 9 hours (the  $S_a$  phase duration), then we must be able to determine the values of our variables for 9 hours prior to the beginning of the experiment, and these values will be stored in a history function. Thus, before we can solve these equations, we must first determine the 7 phase proportions (and the value of  $\gamma$ ) at time  $t = 0$  and we must also

develop a history function representing the values of the phase proportions and the value of  $\gamma$  for times  $-9 < t < 0$  as well.

### Modelling the pure cell population

To model the pure cell population, we must determine which part of the M phase to place the cells in at  $t = 0$ . It is known that mitotic selection produces a large number of cells in the biological M phase of the cell cycle and by definition, all of our pure cells must start in the M phase. As the model equations (3.7) do not differentiate between the G2 and M phases of the cell cycle, it is not obvious in which model phase the newly selected cells must be placed. We can determine this by recalling that the majority of mitotically selected cells complete the M phase and enter the G phase within 30 minutes. If we consider the rate at which cells leave the  $M_b$  phase from equation (3.7g), we have

$$\begin{aligned}\frac{dN_{M_b}(t)}{dt} &= -\beta N_{M_b}(t) \\ \Rightarrow N_{M_b}(t) &= N_{M_b}(0)e^{-\beta t},\end{aligned}\tag{5.1}$$

Thus the rate at which cells leave the  $M_b$  phase is exponential. The value of  $\beta$  for the MCF-7 cell line is 2.31, which allows us to calculate that  $e^{-2.31 \times 0.5} = 32\%$  of cells will remain in the  $M_b$  phase after 30 minutes, and 68% will have moved into the G phase within this time. In addition, those cells that have moved into the G phase will have doubled in number (as cell division occurs when cells move from the  $M_b$  phase into the  $G_a$  phase), so that the overall proportion of cells in the G phase after 30 minutes will be  $\frac{2 \times 68}{2 \times 68 + 32} = 81\%$ . This value compares well with the observation that most mitotically selected cells will enter into the G phase within 30 minutes, and thus, we assume that the mitotically selected cells begin in the  $M_b$  phase at the beginning of the experiment. The only unknown rate variable when modelling this population is the value of  $\gamma$ .

### Modelling the impure cell population

To model the impure cell population we assume, for simplicity, that this population is in phase steady-state. Thus, as we discussed in Section 4.3, we only require knowledge of one of  $G, S, M$  or  $\rho$  in order to determine the remaining unknowns in the model system.

Knowledge of any of these four variables will also allow determination of  $\gamma$ , and as this population is growing in the same environment at the mitotically selected population, the value of  $\gamma$  will be the same as the value of  $\gamma$  which we require in modelling the impure cell population. Thus, we only need one of  $G, S, M$  or  $\rho$  (we choose to determine  $\rho$ ) to model *both* the pure and the impure cell population. In addition to this, there is one more variable that we must determine, and that is the proportion of the starting population that are pure cells.

Thus, to model mitotic selection, we need to determine two variables - the value of  $\rho$  and the proportion of the starting population that are pure cells. To determine these values, we use optimisation. We do this by considering different values of  $\rho$  and the pure cell proportion of the starting population. For each pair of values we try, we solve the delay differential equations numerically using MATLAB's **dde23** function. We use MATLAB's **lsqnonlin** function to determine how well the modelled cell cycle phase distributions compare with the experimentally determined values from Figure 5.1, and the parameter values which produce the best fit to the cell cycle phase proportions in a least squares-sense are the parameters we choose. We do not find the best fit to all phases of the cell cycle data - we choose to find the best fit to the S phase data, as the remaining phases will simply be a result of how the S phase changes.

## Results of optimisation

Performing the optimisation process provides an optimal value of  $\rho$  as  $0.03 \text{ hr}^{-1}$  (equivalent to a doubling time of 23.3 hours) and the proportion of the pure cell population at time 0 is 0.59. The pure cell population making up 59% of the total population at time 0 may seem low. However, when the mitotically selected cell population completes mitosis (most of them doing so within 30 minutes), they have doubled in number and so they make up a much larger proportion of the total population. We calculate that after the pure population has doubled, they make up  $\frac{2 \times 0.59}{2 \times 0.59 + 0.41} = 74\%$  of the entire cell population<sup>1</sup>. From Figure 5.1, we observe that to G phase proportion at this time is close to 85%, 74% of which is made up of the pure cell population, and the remaining 11% is

---

<sup>1</sup>This calculation is not entirely correct, as the impure cell population is also growing exponentially and so it will actually constitute a higher proportion of the total cell population. However, the increase in the overall population due to impure cell growth is very minor and does not significantly alter the final result.

Table 5.1: *Steady-state values of variables for MCF-7 breast cancer cells during exponential growth when the doubling time of the population is 23.3 hours. The calculated variables are shown here to 2 significant figures.*

| variable | value        |
|----------|--------------|
| $G_a$    | 5.8%         |
| $G_b$    | 14%          |
| $G_c$    | 28%          |
| $G$      | 48%          |
| $S_a$    | 36%          |
| $S_b$    | 4.9 %        |
| $S$      | 40%          |
| $M_a$    | 10%          |
| $M_b$    | 1.3%         |
| $M$      | 12%          |
| $\rho$   | $0.03h^{-1}$ |
| $\gamma$ | $0.39h^{-1}$ |

made up of impure cells that are still in the G phase at this time.

The value  $\rho$  allows us to determine the steady-state values of the model phase proportions and the value of  $\gamma$ , shown in Table 5.1. These model phase proportions are not the phase proportions we observe when modelling mitotic selection. These values represent the phase proportions that cells would have been in prior to mitotic selection based on the environmental conditions in which the cells are growing. These model phase proportions are therefore the proportions which the impure cell population is in prior to mitotic selection, and hence for the entire experiment. If we let the simulation run for long enough, it is also the phase proportions that our model will eventually settle into. The actual observed model phase proportions at time 0 are shown in Table 5.2. These model phase proportions are calculated by assuming that at time 0, 41% of the population has model phase proportions as in Table 5.1 (i.e. multiply each of the model phase proportion in Table 5.1 by 0.41), and then the remaining 59% of the population is added to the  $M_b$  phase proportion.

The delay differential equations in (3.7) were evaluated for the duration of the experiment performed in [78], i.e. 34.5 hours with the values of the parameters just discussed, and the changes in the cell cycle phases with time are shown in Figure 5.2.

Table 5.2: *The model phase proportions at time 0.*

| variable | value  |
|----------|--------|
| $G_a$    | 2.4%   |
| $G_b$    | 5.7%   |
| $G_c$    | 11.5%  |
| $G$      | 19.7%  |
| $S_a$    | 14.76% |
| $S_b$    | 2.0%   |
| $S$      | 16.4%  |
| $M_a$    | 4.1%   |
| $M_b$    | 59.5 % |
| $M$      | 63.6%  |

As we can see, the simulated G, S and M phase proportions take the values of 18, 15 and 67% at time 0 respectively, which are in agreement with the values from Table 5.2. Within the first 30 minutes, the pure cell population doubles, and the G phase proportion increases with a corresponding decrease in the M phase proportion. The S phase proportion also decreases slightly within the first 30 minutes. This is because the total cell population has increased during this time due to the doubling of the pure cell population, and all of the new cells are concentrated in the G phase. Thus, even though there has been no change in the number of S phase cells during this time, they make up a smaller proportion of the population as a whole.

For  $0.5 \leq t \leq 6$  hours, the G, S and M phase proportions are about 88, 10 and 3% respectively. Note that during this time, the phase proportions are not exactly constant, and appear to be changing slowly - the G phase is slowly decreasing while the S and M phase proportions are slowly increasing. During this time, the pure cells are all in the G phase, however the minor changes in each of the cell cycle phases is due to the presence of the impure cell population. The impure cell population is in phase steady-state, with G, S and M phase proportions at 40, 48 and 12% respectively. Thus, as this population of cells increases exponentially over time, they make up a larger and larger proportion of the population as a whole, while the pure cell population is stagnant in the G phase.

At exactly  $\tau_{G_a} + \tau_{G_c} = 6.75$  hours (although this exact number is hard to read off from the graph shown), the pure cell population begins to leave the G phase and enter

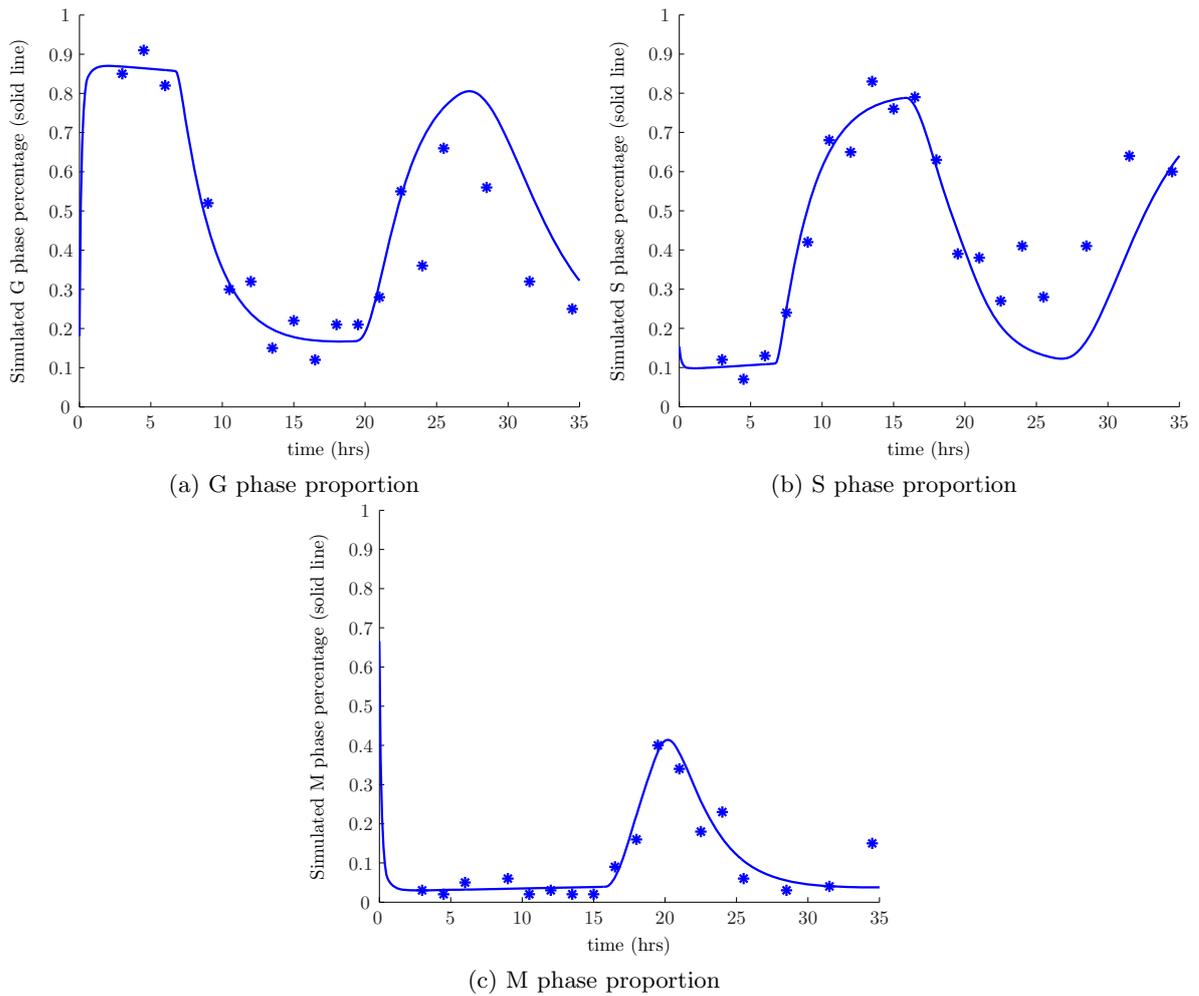


Figure 5.2: *The modelled proportion of cells in (a) the G phase, (b) the S phase and (c) the M phase for 35 hours after mitotic selection. The modelled phase proportions are represented by the solid curve in each figure and the experimentally determined phase proportions from [78] are shown as the asterisks.*

the S phase, as evidenced by the decrease in the G phase proportion and increase in the S phase proportion after this time. The S phase proportion increases from about 11% until  $t = \tau_{G_a} + \tau_{G_c} + \tau_S = 15.75$  hours, where it reaches a maximum of about 78%. After this time, S phase cells are available to transition into the M phase, and this results in a decrease in the S phase proportion with a corresponding increase in the M phase at this time. Even though the S phase cells have become available to transition into the M phase at this time, some G phase cells will still be entering the S phase, and so it is possible that we may not observe a noticeable decrease in the S phase proportion as soon as the cells become available to transition into the G phase. However, in this case, the rate of

leaving the S phase is higher than the rate at which cells are entering the S phase at this time, resulting in a noticeable decrease in the S phase proportion.

The M phase increases from about 4% at 15.75 hours to 41% at  $t = 20.25$  hours, and thereafter decreases. We may have expected the M phase proportion to reach a maximum at time  $t = \tau_{G_a} + \tau_{G_c} + \tau_S + \tau_M = 19.15$  hours, as at this time some cells will be able to exit the M phase. Although it is true that some cells may be available to exit the M phase at this time, the rate at which cells are still entering the M phase from the S phase is higher than the rate of exit, resulting in a net increase in the M phase proportion. Thus the M phase proportion does not begin to noticeably decrease until 20.25 hours, when the rate at which cells leave the M phase (and enter the G phase) exceeds the rate at which cells enter. Once cells re-enter the G phase after this time, they appear to move at a more constant rate from the G into the S phase after about 27 hours. This is because cells are losing their synchrony, and are not necessarily transitioning between the phases at similar points in time.

The modelled G and S phase proportions compare well with the experimentally determined G and S phase proportions (asterisks) for the first 23 hours of the simulation (i.e. for the first round of cell division). For later times, the modelled G phase proportion is significantly higher (with a correspondingly lower S phase proportion) than the experimentally determined percentage. One possible reason for this is that the environmental conditions had changed sometime during the first round of cell division. Consequently, when the pure cells completed cell division and re-entered the G phase at about 23 hours, the environmental conditions were more favourable for cell division, resulting in a higher rate of transition from the G phase into the S phase. It is possible that the change in environmental conditions within the first 23 hours of the experiment could be due to cancer cells secreting growth factors which aid in cell division.

The modelled M phase proportion compares well with the experimentally determined M phase proportion up until the final point at 34.5 hours, where the modelled M phase proportion is not as high as the experimentally determined M phase proportion at this time. The increase in the M phase proportion at this time is due to the S phase cells entering the M phase for a second time since the beginning of the experiment. If we continue the simulation (not shown), we find that the S phase cells enter the M phase a

second time at about  $t = 38.3$  hours. The increase in the observed M phase proportion therefore occurs about 4 hours earlier when compared to the simulated increase in the M phase proportion. It may be tempting to conclude, as we did above, that secreted growth factors could be causing this earlier increase in the M phase proportion than what we would have expected. However, no matter how fast the cells transition from the  $G_b$  phase, they must still spend the required amount of time in the storage phases. Thus, it will be impossible for the simulated M phase proportion to increase earlier than  $t = 2 \times 19.15 = 38.3$  hours due to the presence of the storage phases. As only one observation at 34 hours after selection shows an increase in the M phase proportion, it is possible that this could be due to some observational error. Indeed, we note that in another mitotic selection experiment performed in [115], which investigated cell cycle proportions up until 34 hours after selection, the M phase proportion was not observed to increase at 34 hours (not shown).

## 5.2 ‘Slowly cycling population’?

In this section, we use our model to discuss a theory proposed in the literature that the MCF-7 cell line is composed of a small population of slowly cycling cells.

Another mitotic selection experiment was performed in [115], which has similar results on the observed changes in the cell cycle phase proportions after mitotic selection. Taylor et al. [115] noted that the G phase proportion never reached zero after mitotic selection, and appeared to reach a minimum of only 15% between 10-18 hours (this can also be seen in results from [78], where the G phase proportion does not go below about 12%). They were surprised, as they thought that the G phase proportion would get much closer to zero while the mitotically selected cells were in other phases of the cell cycle. Taylor et al. [115] concluded that there must exist a population of cells, which they called ‘slowly cycling cells’, which remained in the G phase of the cell cycle while most of the cells continued through the other phases after selection. These slowly cycling cells were suggested to have different cell cycle kinetics when compared to other cells in the population because they have a much slower rate of exit from the G phase. This interpretation of the data was likely given because it was becoming apparent that cancers could be driven by a small

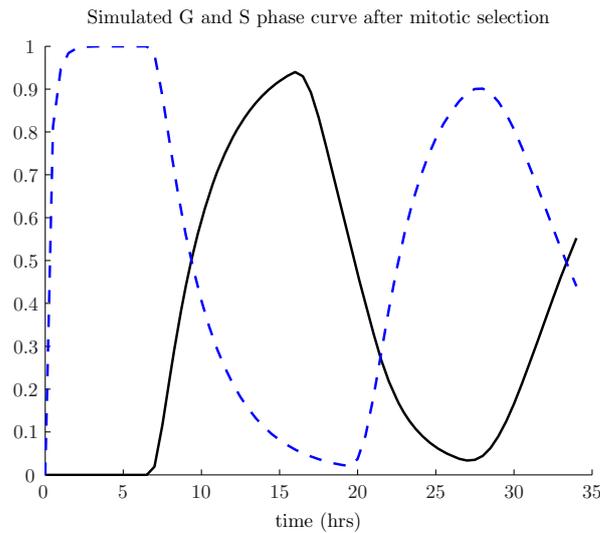


Figure 5.3: *The modelled G (dashed curve) and S (solid curve) phase changes after mitotic selection for a 100% pure cell population.*

population of stem cells. These stem cells may divide infrequently, however the transit amplifying population they give rise to can populate the bulk of the tumour, meaning that cancers could be made up of a small population of slowly dividing stem cells which pass on dangerous traits to the frequently dividing transit amplifying population. However, we note that our model also shows that the G phase proportion also never reaches zero after mitotic selection, without having to explicitly include the presence of this ‘slowly cycling’ population. The reason that the modelled G phase proportion does not drop below about 16% in Figure 5.2 is because of the presence of the impure cell population. If we simulate mitotic selection without any impure cells (i.e. if we start with 100% pure cells), then the G phase proportion does get very close to zero. The results of this simulation are shown in Figure 5.3, which shows that the G phase proportion gets very close to zero at about 19 hours after selection. We suggest that introducing a separate slowly cycling population with different rate parameters is not necessary to explain the changes in cell cycle phase distributions observed, and that the presence of the impure cell population is sufficient to account for the experimentally observed features of mitotic selection.

In [112], the presence of a slowly cycling population was again used to explain a different set of experimental results. MCF-7 cells growing in phase steady-state were exposed to the drug ICRF 159, which prevents cells from completing mitosis and hence from re-entering the G phase. We would expect that the G phase proportion would

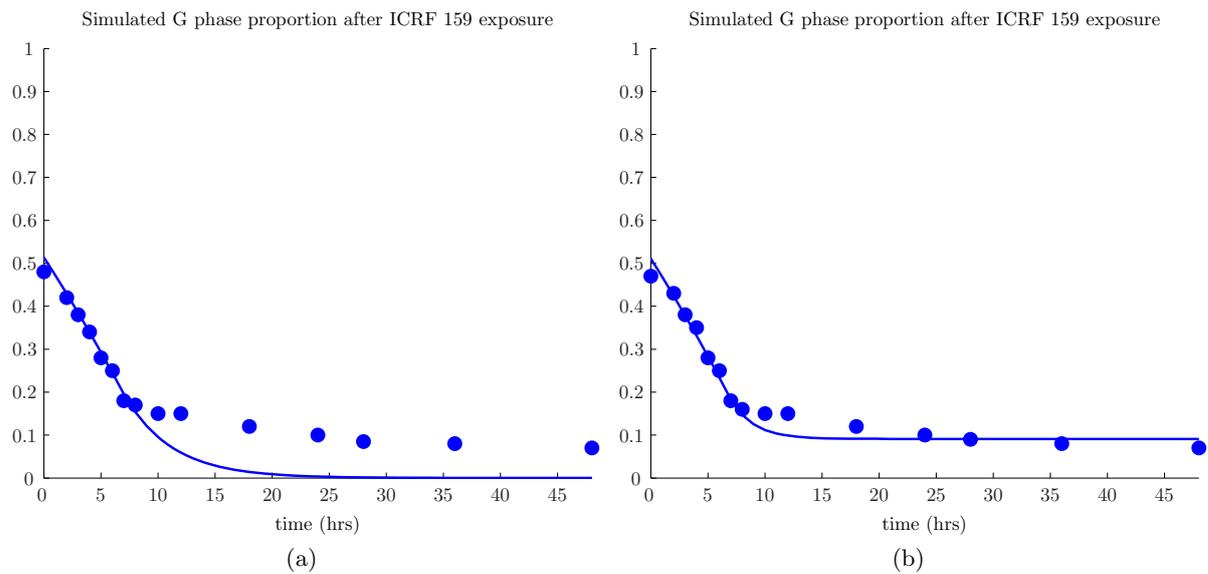


Figure 5.4: *The simulated G phase proportion after ICRF when (a) we assume that ICRF is 100% effective at preventing completion of mitosis and (b) we assume that ICRF is 100% effective at preventing completion of mitosis and that there is a slowly cycling population consisting of 10% of the entire cell population. Experimental data from [112] is shown as circles which represent the mean of 2-6 separate experiments.*

eventually reach zero as cells continue to leave this phase at a constant rate, under the assumption that the presence of ICRF does not alter the value of  $\gamma$ . However this was not observed in the experimental results.

We can simulate this experiment. We assume that cells start in phase steady-state. At  $t = 0$ , when the cells are exposed to ICRF 159, cells are prevented from completing the M phase. Figure 5.4(a) shows the experimental and modelled G phase proportion after ICRF 159 exposure.

As we can see, assuming that cells are completely arrested in the M phase results in a very fast decline in the G phase proportion. The model results compare very well with experiments for the first 9 hours, but thereafter the model produces results that are significantly below that of the experiment. The presence of a slowly cycling population of cells was again used to explain the fact that the observed G phase proportion remained above zero for the duration of the experiment. We can include the presence of this population in the model and simulate this hypothesis. We assume that there exists a population of cells that remains in the G phase for the entire experiment. We do not need to solve equations (3.7) for this population, we simply add a constant number of cells

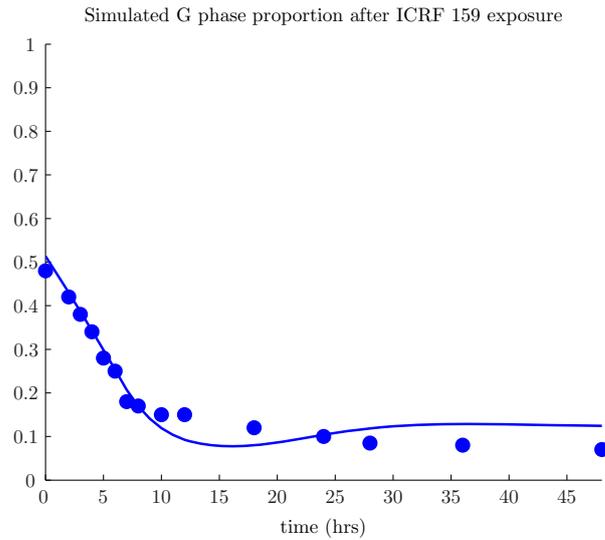


Figure 5.5: *The G phase proportion when we assume that ICRF 159 affects 99.96% of cells in the M phase rather than 100% of the cell population. Experimental data from [112] is shown as circles which represent the mean of 2-6 separate experiments.*

to the G phase number once we have simulated cell cycle progression after ICRF 159. If we assume that the slowly cycling population makes up 10% of the total population (which remains at 10% for the full duration of the experiment, as the entire cell population does not grow due to the presence of ICRF), then we can produce Figure 5.4(b), which compares very well with the experimental data for the entire duration of the experiment.

The quality of the fit to the data in Figure 5.4(b) would suggest that the presence of a background G phase population is a good hypothesis. However there is an alternative hypothesis, and that is to assume that ICRF is not 100% effective at preventing cells from exiting the M phase. We can model this by assuming that ICRF 159 prevents  $x\%$  of cells completing mitosis at any time. We determine the value of  $x$  by finding the best fit to the data in a least-squares sense. Doing this gives  $x = 99.96\%$ , and the resultant changes in the G phase proportion are shown in Figure 5.5. The modelled G phase profile compares much better with experimental data for times  $t > 9$  hours when compared to Figure 5.4(a). It is surprising that altering the value of  $x$  from 100% to 99.96% has such a drastic effect on the modelled G phase proportion. However, it is biologically reasonable to assume that the drug is not 100% effective at preventing completion of mitosis, and perhaps even unrealistic to assume that the drug is 100% effective at preventing cells from exiting the M phase.

The assumption that there exists a slowly cycling population of cells provides an excellent fit to the data. However, there seems to be little biological evidence from other sources for such a population. Cells from the MCF-7 cell line are a clonal population of cells, meaning that each cell should be a replica of all other cells in the population, and there is therefore no reason to suspect that there would exist two distinct cell populations. It is possible that subsequent divisions may cause mutations leading to slightly different cell behaviours, however it is unlikely that such a large number of cells (10% of the population) would have developed such similar behaviour. If this were to happen, subsequent cell divisions performed by the actively dividing population of cells would mean that the slowly cycling population would make up a smaller and smaller proportion of the overall population, resulting in vastly different experimental results if the same experiment was performed just days later. In addition, this slowly cycling population may raise problems with convergence to phase steady-state. If we instead assume that ICRF 159 is not 100% effective at preventing M phase completion, then the model also produces a reasonable fit to the G phase data. We believe it is more realistic to assume that ICRF 159 is not 100% effective at preventing completion of the M phase, rather than to assume that there is a heterogeneous population of cells amongst this clonal cell line.

### **The experiment-independent variables, and their influence over model behaviour when not in steady-state**

As discussed in Section 4.3, the choice of the experiment-independent variables do not significantly influence the phase diagram (representing the behaviour of the cell cycle phase proportions in steady-state), which depends more heavily upon the choice of the base experiment variables. However, the storage phase durations significantly influence the G, S and M phase proportions when the cell population is not in steady-state. Let's consider the results of the mitotic selection experiments in Section 5.1.

The high values of the storage phase durations for the MCF-7 cell line means that it takes a long time for the cell population to reach steady-state. If we allow our mitotic selection simulation to run until time  $t = 1000$  hours, then we find it is not until about 400 hours after selection (about 17 days!) that steady-state is reached (details not shown).

We wish to investigate what would happen if our storage phase durations were not so

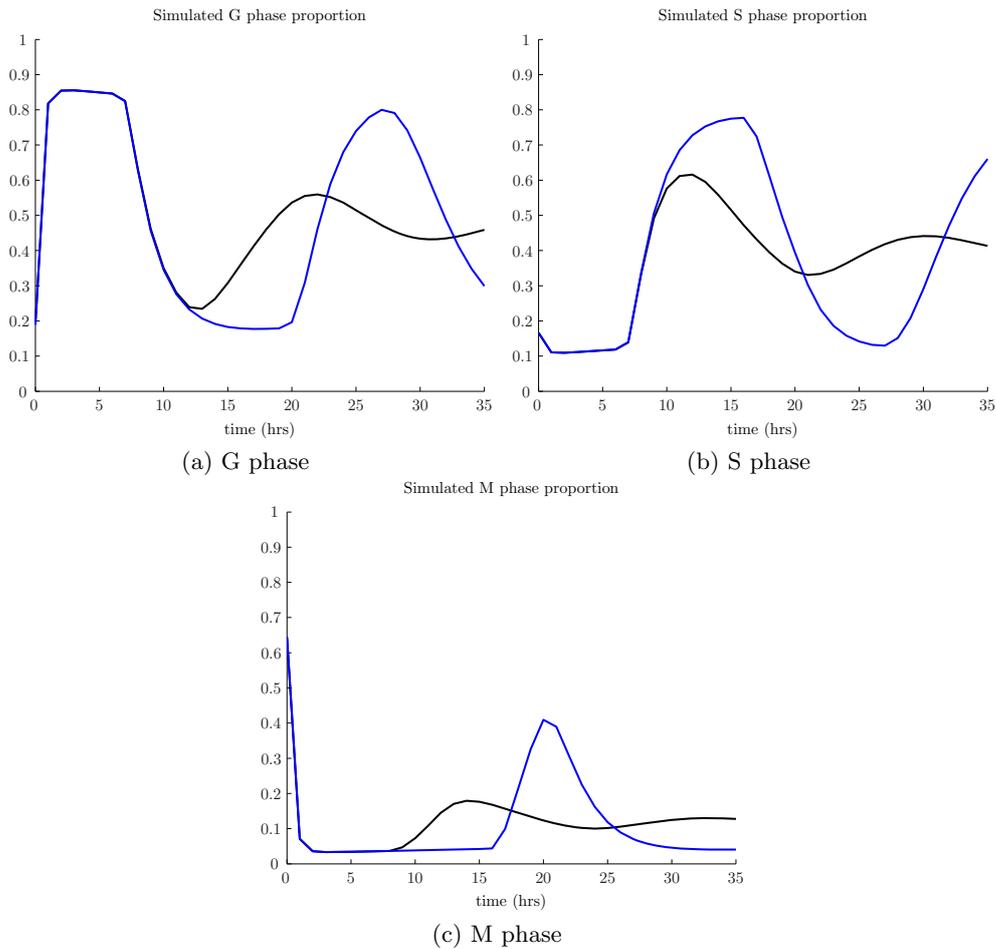


Figure 5.6: Cell cycle progression after mitotic selection when  $\tau_S = 9$  (blue lines) and when  $\tau_S = 1$  (black lines).

high. We choose  $\tau_S = 1$ , which through the base variables, causes a change in the value of  $\alpha$  from 0.67 to 0.09. Figure 5.6 shows the results of mitotic selection in this case. We can see that the significantly smaller value of  $\tau_S$  has produced a significantly different cell cycle phase profile after mitotic selection. For the first 12 hours, the cell cycle phase proportions are no different to when  $\tau_S = 9$  hours. However, after this time, the fact that cells are not forced to remain in the S phase for a full 9 hours means that a significant number of cells move into the M phase after this time. The shape of the cell cycle phase profiles do not change as dramatically over time when the storage phase durations are not so high, and the steady-state G, S and M phase proportions are reached within 60 hours after selection (not shown), which is significantly faster than when  $\tau_S = 9$  hours. Similar results can be seen if we reduce  $\tau_{G_a}$ ,  $\tau_{G_c}$  and/or  $\tau_M$ . The investigations in this

section shed further light on the role that the experiment-independent variables play in model behaviour, and provide strong evidence for the need for storage phase durations when modelling cell cycle progression in the MCF-7 breast cancer cell line.

### 5.3 Changing Environmental Conditions

So far, we have only dealt with cell populations exposed to unchanging environmental conditions. This means that our rate coefficient,  $\gamma(t)$ , which represents the environmental influence on the cell cycle, has remained constant with respect to time. If we wish to consider situations where the environmental conditions change with time, such as a change in growth factor concentration, or the addition of a hormone, then we must be able to translate this environmental change into a change in the value of  $\gamma(t)$ .

As discussed in Chapter 3, changes in environmental conditions only affect cells that are in the  $G_b$  model phase. During this period, a cell will review the different signals it receives from the environment and will make a decision about whether to progress through the remainder of the cell cycle, or remain in the G phase. Once a cell decides to progress past the G phase, it will complete cell division without further consultation of the environmental conditions ([135], pg. 262). Thus  $\gamma(t)$  represents the only point of influence the environment has on cell division.

Different environmental signals will have different effects on the cell cycle. For instance, estrogen causes an increase in the growth rate of a cell population [98, 135]. However progesterone has a biphasic effect on the cell cycle, where growth is initially stimulated, and then later growth is inhibited ([37, 84, 61]). Our  $\gamma(t)$  function must be able to incorporate the effects of different environmental signals if we wish to apply our model to biologically interesting situations.

Most environmental signals which are added to a cell population, including hormones and growth factors, influence the cell cycle through communication with internal proteins. In particular, there are two internal proteins which strongly affect the rate at which cells progress through the cell cycle - these are cyclin D and myc. Most experiments which investigate cell cycle progression also report changes in the concentration of these two proteins, and it is thought that the concentration of these two proteins correlates well

with cell cycle progression. The effect these proteins have on the cell cycle has been discussed in detail in Section 2.9.

We assume that each environmental signal can be translated into an effect on the concentration of cyclin D and myc (which we refer to as  $d(t)$  and  $m(t)$  respectively), which in turn will affect the value of  $\gamma(t)$ . Doing this provides a ‘universal language’ by which the environment can communicate with internal cell cycle machinery. Although it is known that there are many other proteins which can have an effect on cell cycle progression, we find that including the effect of just these two proteins is good enough for our purposes. Including the effects of other proteins may add unnecessary complexity to the model, and may blur insight into the mechanisms involved. When modelling any real world phenomena, it is best to keep the model simple, only adding new levels of complexity if the previous levels are not enough to accurately represent the system.

Thus we assume  $\gamma(t)$ , the environment’s influence over cell cycle progression, can be represented as a function of cyclin D and myc concentrations. Because  $\gamma(t)$  is not explicitly dependent on time, but rather on the time-dependent concentrations of cyclin D and myc, it may be more correct to write it as  $\gamma(d(t), m(t))$ . However, for simplicity, we will continue to refer to it as  $\gamma(t)$ , and we will also refer to it as the  $\gamma$  function.

We assume that  $\gamma(t)$  increases at a rate proportional to the concentrations of cyclin D and myc. We will also assume that cyclin D and myc are not equally effective at increasing the value of  $\gamma(t)$ , and so the rate at which  $\gamma(t)$  increases in response to cyclin D and the rate at which  $\gamma(t)$  increases in response to myc may be different. We may express this growth term as  $\omega_D d(t) + \omega_m m(t)$ , where  $\omega_D$  and  $\omega_m$  represent the rate at which  $\gamma(t)$  increases in response to cyclin D and myc respectively. We choose, however, to write this term as

$$\omega(\alpha^D d(t) + (1 - \alpha^D)m(t)),$$

where  $\alpha^D$  represents the relative influence on  $\gamma(t)$  of cyclin D when compared to myc, and  $\omega$  represents the effect that a linear combination of cyclin D and myc have on the production of  $\gamma(t)$ .

We have observed that cells react to growth factors faster when they start in favourable steady-state conditions (i.e. high values of  $\gamma(t)$ ) compared to cell populations that began

in starved steady-state conditions (i.e. low values of  $\gamma(t)$ ). To capture this behaviour, we will assume that the interaction between cyclin D and  $\gamma(t)$ , and the interaction between myc and  $\gamma(t)$ , affect the rate of growth of  $\gamma(t)$ . Thus, we include an additional growth term which can be written as  $\gamma(t)(\omega_D^\gamma d(t) + \omega_m^\gamma m(t))$ . However as we did earlier, we will write it as

$$\omega_\gamma \gamma(t)(\alpha_\gamma^D d(t) + (1 - \alpha_\gamma^D)m(t)),$$

where  $\alpha_\gamma^D$  represents the relative influence of the interaction between cyclin D and  $\gamma(t)$  when compared to the interaction between myc and  $\gamma(t)$ . The rate  $\omega_\gamma$  represents the rate at which the interaction between  $\gamma$ , and the linear combination of cyclin D and myc, influences the production of  $\gamma(t)$ .

We also include a natural rate of decay of the variable  $\gamma(t)$ , which we assume decays at a constant rate  $\beta_\gamma$ . Thus we can now write the full model for the rate of change in  $\gamma(t)$  as

$$\frac{d\gamma(t)}{dt} = \omega(\alpha^D d(t) + (1 - \alpha^D)m(t)) + \omega_\gamma \gamma(t)(\alpha_\gamma^D d(t) + (1 - \alpha_\gamma^D)m(t)) - \beta_\gamma \gamma(t). \quad (5.2)$$

Thus, the value of  $\gamma(t)$  over time can be determined from knowledge of cyclin D and myc over time, and then substituted as the rate coefficient into equations (3.7), and changes in the cell cycle phase proportions can then be determined.

## 5.4 Biological literature and the $\gamma$ function

Many experiments have been carried out which investigate the change in cell cycle distribution due to changes in cyclin D and myc concentration ([134, 133, 98, 97, 20, 69]). We use the experimental results from [134, 133, 98, 20] in this section and the following section to parameterise our  $\gamma$  function. In this section, we briefly review these experiments, and highlight those that will aid us in determining the unknown parameters from equation (5.2).

Anti-estrogens down-regulate cyclin D and/or myc and in turn decrease the rate of

proliferation, resulting in a high proportion of cells in the G phase. Having a high steady-state proportion of G phase cells means that many cells exist in the  $G_b$  phase. Having a high  $G_b$  phase proportion is desirable because any subsequent proliferative agents which are added to the cell population will result in a highly synchronous progression through the remainder of the cell cycle. Thus, many experiments which aim to investigate the effects of proliferative agents on the cell cycle will pre-expose their cell population to anti-estrogens. This means that understanding the effects of anti-estrogens on the cell cycle is critical.

There are several different types of anti-estrogens. However, for most experiments which investigate the effects of proliferative agents on MCF-7 breast cancer cells, the anti-estrogen ICI 182780 is used to decrease the rate of proliferation. This anti-estrogen is a member of the class of *pure* anti-estrogens. Some anti-estrogens, such as tamoxifen, will sometimes mimic the effects of estrogen [70, 21], resulting in a partially proliferative effect. Pure anti-estrogens, on the other hand, exert a strictly anti-proliferative effect on breast cells [127] and are therefore more effective at decreasing the overall proliferation rate of a cell population than other anti-estrogens. In this section, we use the biological literature to determine the quantitative effect of ICI 182780 on changes in cyclin D and myc concentrations.

Experiments performed in [133] investigate changes in cyclin D protein concentration in response to the anti-estrogen ICI 182780 in MCF-7 breast cancer cells. It was found that cyclin D protein levels are reduced to 40% of the starting concentration after ICI 182780 exposure. The authors comment that myc is likely to be reduced in response to this anti-estrogen as well, however they do not quantitatively investigate the changes in concentration of this protein.

Investigations in [20] show that in response to ICI 182780, myc concentrations are reduced to 12% of the initial concentration. In both this experiment and the one described in [133], the starting steady-state S phase proportions prior to anti-estrogen exposure were similar, implying that cells were growing in similar environmental conditions. Thus we will assume that ICI 182780 reduces cyclin D concentrations to 40% and myc to 12% of the starting concentration. We note that under different environmental conditions, ICI 182780 will not always reduce cyclin D and myc concentrations to 40 and 12% respectively.

However, we will assume that this anti-estrogen has the same relative effect on cyclin D and myc: that is, the effect of ICI 182780 on myc will be about 4 times stronger than the effect on cyclin D.

## 5.5 Fitting the parameters from equation (5.2)

As discussed in Section 5.3,  $\gamma(t)$  can be determined directly from knowledge of cyclin D and myc by solving the differential equation expressed in equation (5.2). The resultant  $\gamma$  function can then be substituted as the value for  $\gamma(t)$  into equations (3.7) and the cell cycle phase distributions can subsequently be determined. However, at this point, we do not know the values of the rate parameters in equation (5.2) and so we cannot use knowledge of cyclin D and myc to determine the phase proportions. In this section, we will describe how we can use knowledge of cyclin D and myc, along with the cell cycle phase proportions over time, to determine the optimal values of the parameters in equation (5.2). These parameters are environment independent and so we only need to determine them once and they will take the same values across all experiments with the same cell line.

To find the values of the parameters in equation (5.2), we will require knowledge of cyclin D and myc over time, however we do not require knowledge of all three cell cycle phase proportions over time. As we saw in Section 4.3, when we assumed that the environment was unchanging, knowledge of just one of the three steady state variables,  $G$ ,  $S$  or  $M$ , was all that was required to determine the remaining experiment dependent variables from Table 4.2. When the model system is not in steady-state (i.e. when  $\gamma(t)$  is not constant), then knowledge of just one phase proportion is still enough to determine the remaining unknown variables in (3.7). However, we must have knowledge of this phase proportion for a long enough *period* of time and not just at one point in time. In steady-state, we only need a single phase proportion at a single point in time because we know that the environmental conditions are constant. However in changing environmental conditions, knowledge of a phase proportion (the S phase, for instance) at one point in time is not enough, as we do not know what kind of environmental conditions are driving the changes in the S phase proportion - whether this S phase proportion is in steady-

state, or whether the S phase proportion is currently increasing due to the addition of a growth factor. Fortunately, experimental results provide us with a rich enough history of cell cycle phase proportions so that we need only use the history for one of the cell cycle phases. We will use the S phase profiles provided in our experiments, as some experiments will only provide the S phase profiles (and will not provide changes in the G or M phase proportions).

We choose to use the results of an experiment which investigates the growth-promoting factor, estrogen (detailed in [98]), to determine the parameters in equation (5.2). We choose to use this experiment as it provides the most data on the S phase, cyclin D and myc changes when compared to other experiments that investigate cell cycle progression after exposure to an environmental agent.

Experiments performed in [98] investigate the effects of estrogen on cell cycle progression. MCF-7 cells growing in growth-factor supplemented medium were exposed to ICI 182780 and left to reach phase steady-state. Cells were then exposed to estrogen and the changes in cell cycle phase proportions were determined at time points thereafter. Relative changes in cyclin D and myc protein concentrations were also determined at intervals after estrogen exposure. Figure 5.7 shows the relative changes in the concentration of (a) cyclin D, (b) myc and the changes in the proportion of cells in the (c) S phase after estrogen exposure (figures extracted from [98]). Figure 5.7(a) (black dots) shows that, after the addition of estrogen, Cyclin D protein levels are unchanged for the first 2 hours, increase to about 5.5 fold above control levels by 8 hours, and thereafter decrease rapidly, but remain at levels at least 2-fold higher than control. Figure 5.7(b) (also black dots) shows that concentrations of myc protein increase 8 fold within the first 2 hours, remain this high until about 4 hours after estrogen exposure, and then decline, levelling at 2.5-4 fold above control levels. Concentrations of cyclin D and myc are therefore regulated to a similar extent following estrogen exposure, although myc increases in concentration earlier, and reaches a higher magnitude than cyclin D. The S phase profile is shown in Figure 5.7c. Changes in the S phase proportion occur several hours after the changes in cyclin D and myc. The S phase proportion increases noticeably by 12 hours after estrogen exposure from about 9% to 15%, and reaches a maximum of 60% by 21 hours. After this point, the S phase proportion begins to decrease. The decrease in the S phase proportion after 21 hours is not due to the decrease in cyclin D and myc, but rather to the cells having

NOTE:  
This figure is included on page 139 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 5.7: *The relative changes in (a) cyclin D and (b) myc concentration after estrogen exposure are shown as the black dots. the white dots in (a) and (b) are not relevant for our purposes. The white circles in (c) represent changes in the S phase proportion after exposure to estrogen, and the black circles in (c) represent the S phase proportion in the absence of estrogen addition. Figures reproduced from [98].*

completed the S phase and progressing into the M phase (not shown). The increase in S phase proportion at about 36 hours after exposure is due to cells entering a second round of the cell cycle, though with less synchrony.

### Using experimentally determined cyclin D and myc

We can use the S phase profile in this experiment to determine the parameters in equation (5.2) as long as we are given the concentration of cyclin D and myc over time. However, we are only provided with changes in cyclin D and myc relative to their starting values.

Fortunately, we can find the absolute concentrations of these two proteins by determining the starting values of cyclin D and myc.

At time 0, the cell cycle phase proportions are in steady-state. Knowledge of the starting proportion of S phase cells (given in Figure 5.7) allows us to determine the value of  $\gamma$  (see Section 4.3). As  $\gamma$  is in steady-state at time 0, we can write equation (5.2) as

$$0 = \omega(\alpha^D d(0) + (1 - \alpha^D)m(0)) + \omega_\gamma(\alpha_\gamma^D d(0) + (1 - \alpha_\gamma^D)m(0))\gamma - \beta_\gamma\gamma.$$

In addition, we use the fact that after exposure to the anti-estrogen ICI 182780, myc is reduced to levels about 4 times lower than that of cyclin D after the same change in environment (see Section 5.4). Thus  $m(0) = \frac{d(0)}{4}$  and we have

$$0 = \omega(\alpha^D d(0) + (1 - \alpha^D)\frac{d(0)}{4}) + \omega_\gamma(\alpha_\gamma^D d(0) + (1 - \alpha_\gamma^D)\frac{d(0)}{4})\gamma - \beta_\gamma\gamma$$

which, after re-arranging for  $d(0)$ , gives:

$$d(0) = \frac{\beta_\gamma\gamma}{\omega(\alpha^D + (1 - \alpha^D)\frac{1}{4}) + \omega_\gamma(\alpha_\gamma^D + (1 - \alpha_\gamma^D)\frac{1}{4})\gamma},$$

which we note only depends upon the experiment-independent parameters from equation (5.2) and the starting value of  $\gamma$ , which we can determine from the starting steady-state phase proportions. Thus, as we go through the optimisation process and try different values for the parameters in equation (5.2), we can determine what the starting values of cyclin D and myc are for each set of parameter values, and hence we know the absolute concentrations of cyclin D and myc over time.

Note also that we are only provided information about the cyclin D and myc concentrations for the first 24 hours after estrogen exposure, and therefore we will not be able to use the given cyclin D and myc concentrations to model the S phase changes for the entire duration of the experiment. The rate of change in the S phase proportion can be determined by adding equations (3.7d) and (3.7e) to get

$$\frac{dS(t)}{dt} = \gamma(t - \tau_{G_c})N_{G_b}(t - \tau_{G_c}) - \alpha N_{S_b}(t).$$

Thus,  $\gamma$  affects the S phase proportion, but with a delay, so that any changes in  $\gamma$  at time  $t$  will affect the S phase proportion at time  $t + \tau_{G_c}$ . This means that we can use the first 24 hours of known cyclin D and myc concentrations to model the S phase profile for up to  $24 + \tau_{G_c} = 29.75$  hours. Thus, to solve for the parameters in equation (5.2), we will only use the data provided on the S phase proportion for the first 29.75 hours of the experiment.

Let's reconsider the cyclin D and myc profiles shown in Figures 5.7a and 5.7b. We are only provided cyclin D and myc concentrations at certain time-points throughout the experiment. To make full use of the cyclin D and myc profiles, we wish to interpolate the data points. One way to do this is to assume that interpolated points lie on a line between known data points. This may be a feasible option for generating a profile of cyclin D. However, the myc data appears quite noisy, and so we do not wish for our curve to be forced to pass through all known data points (thereby capturing the unwanted noise).

Thus to generate an myc profile for times less than 24 hours, we will assume that myc can be represented by a continuous function which best fits the data points. To determine what this function might be, we will consider some important features that appear in the myc profile from Figure 5.7b. At time zero, myc takes the value of 1, and so the functional representation of myc must also take this value. If we assume that prior to time 0, myc is constant, then this would imply that the functional representation of myc at this time will have zero gradient. However also note that within 2 hours of estrogen exposure, myc has increased 7 fold. If we force the function to have zero gradient at time 0, then it is hard for the function to capture this initial quick incline. Instead, we choose the function to simply take the value of 1 at time zero, and we choose not to enforce a zero gradient at this time. As our main aim is to simply generate a function which best represents an interpolation of the known data points, it is not necessary to have this function model the actual biological process of cyclin D and myc up-regulation.

After estrogen exposure, myc increases, reaching a maximum of about 8-fold above control by 4 hours after exposure. Thereafter it decreases, reaching a minimum of about

4-fold above control 12 hours after estrogen exposure. At later times, myc is relatively unchanging. Thus within 12 hours, myc experiences 2 turning points: a maximum at some unknown early time point (close to 4 hours after estrogen exposure) and a minimum at time  $t = 12$ . To capture this behaviour, we represent myc by a cubic polynomial within the first 12 hours. We will force the polynomial to take the value of 1 at time zero, and also enforce a zero gradient at time  $t = 12$ . Although there is another turning point when myc reaches its maximum value, we will allow the fitting program to decide exactly when this occurs. After  $t = 12$  hours, we will assume that myc can be represented by a straight line, with the condition that the function remains continuous at time  $t = 12$  hours.

Thus we wish myc to satisfy

$$\frac{dm(t)}{dt} = \begin{cases} a(t-12)(t-b) = at^2 - a(12+b)t + 12ab & \text{for } t \leq 12 \\ e & \text{for } t > 12. \end{cases} \quad (5.3)$$

Integrating this expression, and imposing both continuity of the function at  $t = 12$  as well as the condition that  $m(t) = 1$  at  $t = 0$  gives

$$m(t) = \begin{cases} a\frac{t^3}{3} - a(b+12)\frac{t^2}{2} + 12abt + 1 & \text{for } t \leq 12 \\ et + f & \text{for } t > 12, \end{cases} \quad (5.4)$$

where we enforce  $f = a \times \frac{12^3}{3} - a(12+b) \times \frac{12^2}{2} + 12ab \times 12 + 1 - e \times 12$  so that we have continuity at time  $t = 12$ . We can now fit the unknown coefficients  $a, b$  and  $e$ .

Performing a least squares fit between the known values of myc and the functional form of myc, we find  $a = 0.059, b = 4.6$  and  $e = 0.072$ , to 2 significant figures. The functional representation of myc is shown in Figure 5.8.

To find a functional form representing relative changes in cyclin D, we will simply form a cubic interpolation between the data points, as there is minimal noise in the data for the changes in this protein. The resulting curve representing cyclin D changes is shown in Figure 5.9. These functional representations of cyclin D and myc allow us to generate values for cyclin D and myc for all times  $0 \leq t \leq 24$  and not just at the times where the concentrations were determined experimentally.

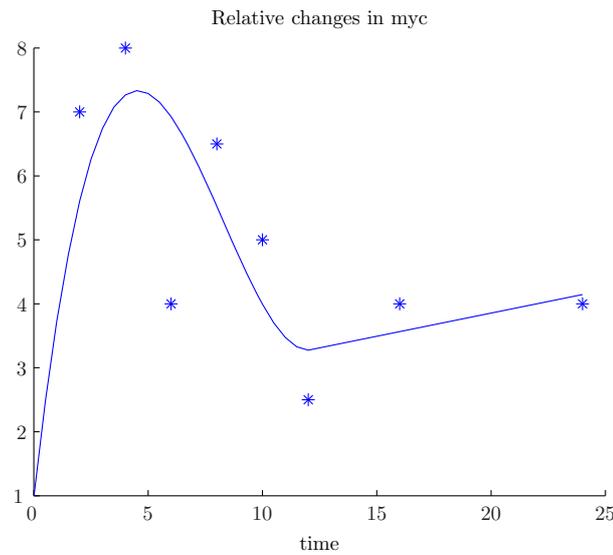


Figure 5.8: The functional representation of *myc* is represented here by the curved line. The asterisks are the experimentally determined relative changes in *myc* from [98].

Table 5.3: The value of the rate parameters from equation (5.2) (to 3 significant figures) after performing a least squares best fit between the modelled and experimentally determined S phase over the first 29.75 hours.

| variable          | value |
|-------------------|-------|
| $\omega$          | 2.00  |
| $\omega_\gamma$   | 123   |
| $\alpha^D$        | 0.594 |
| $\alpha_\gamma^D$ | 0.208 |
| $\beta_\gamma$    | 0.261 |

We substitute these profiles into equation (5.2) to determine the value of the parameters in equation (5.2) which produce the best fit to the S phase proportion (i.e. we use the cyclin D and *myc* profiles to parameterise the  $\gamma$  function). Doing this gives the parameter values shown in Table 5.3. We can check that the fitted parameter values from Table 5.3 do give good results for the change in the S phase over time. Equation (5.2) was solved using the parameter values shown in Table 5.3, and the interpolated cyclin D and *myc* profiles from Figures 5.8 and 5.9 were also substituted into this equation to drive the changes in  $\gamma(t)$ . The resultant  $\gamma$  function was substituted as the rate coefficient in equations (3.7) and the model was run for 29.75 hours. The modelled G, S and M phase proportions are shown in Figure 5.10. The modelled G and S phase curves are shown for

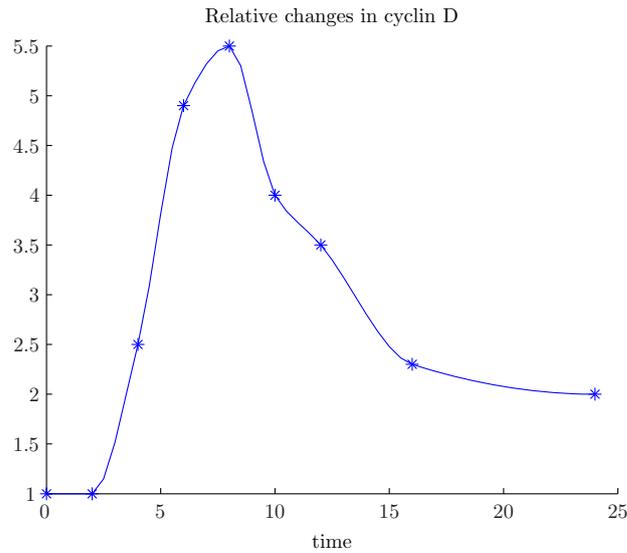


Figure 5.9: *The cubic interpolation of cyclin D is represented here by the curved line. The asterisks are the experimentally determined relative changes in cyclin D from [98].*

29.75 hours, however we are able to model the M phase proportion for the full 36 hours after estrogen exposure. This is because the value of  $\gamma(t)$  does not affect the proportion of cells in the M phase until times  $t + \tau_{G_c} + \tau_S = t + 14$  (see equations (3.7f) and (3.7g)) and so knowledge of  $\gamma$  for the first 24 hours allows us to model the M phase proportion for the full duration of the experiment. The representation in Figure 5.10 shows that our model produces an excellent fit to the data points. The fact that model parameters could be found so that our model compares so well with experiments gives strong evidence that the model captures the important cell kinetic features of the MCF-7 cell line.

At this point we would like to note that the cyclin D and myc profiles observed in this experiment are very similar, and therefore it is possible that the values of  $\alpha^D$  and  $\alpha_\gamma^D$  have not been accurately resolved. However, in the next section we model cell cycle progression in cell populations which are stimulated to express very different cyclin D and myc concentrations and, as we will see, the chosen values for  $\alpha^D$  and  $\alpha_\gamma^D$  produce accurate S phase profiles. Thus, we find there is no need to do further fitting to modify the current values of  $\alpha^D$  and  $\alpha_\gamma^D$ .

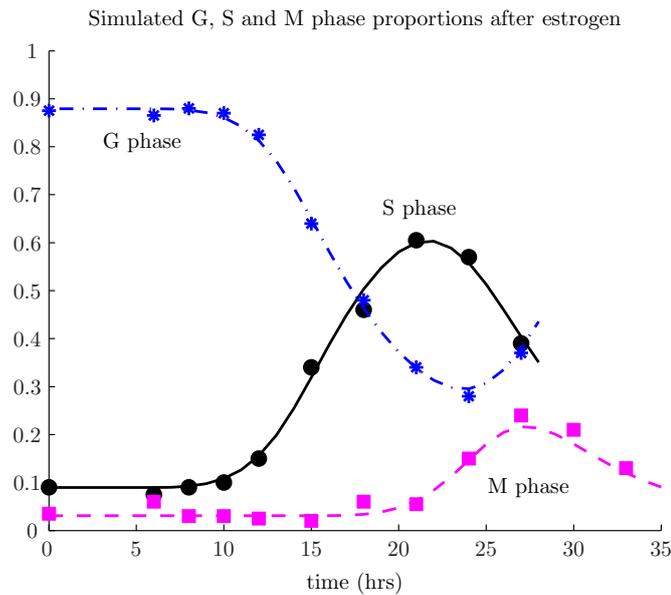


Figure 5.10: Cell cycle phase changes using equations (3.7) are shown as the blue dot-dashed curve (*G phase*), solid black curve (*S phase*) and the dashed magenta curve (*M phase*). The asterisks, circles and squares represent the experimentally observed *G*, *S* and *M* phase changes after estrogen exposure from [98].

## 5.6 Applying the cell cycle model to other experiments involving growth factors

Now that the model represented in equation (5.2) has been fully parameterised, we can simulate cell cycle progression after exposure to a range of substances, so long as we are provided the corresponding changes in the cyclin D and myc concentrations.

There are several experiments which have investigated the effects of growth factors and anti-growth factors on S phase profiles in MCF-7 breast cancer cells. In this section, we consider two experiments which investigate growth factor effects on S phase progression. These two experiments were performed in [60] and [69]. In [69], it was noted that the degree to which cyclin D and myc were induced correlated to the degree of changes in the S phase proportion. However such observations only provide minor evidence that cyclin D and myc are the sole determinants of cell cycle progression, and does not rule out the possibility that other internal proteins are influencing cell cycle progression in undetermined ways. The mathematical model presented in this thesis offers a way to quantify the effects of cyclin D and myc across different experiments, providing insight

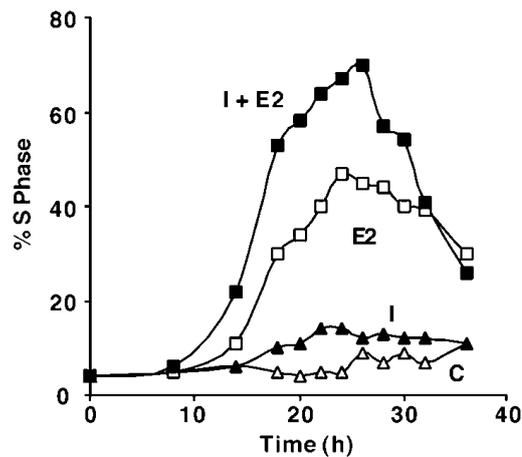


Figure 5.11: The S phase profile was reproduced from [60]. Cells were exposed to ICI 182780 and left to reach phase steady-state. At time 0, cells were exposed to insulin (I, black triangles), estrogen (E2, white squares) or estrogen + insulin (I + E2, black squares). White triangles represent the S phase in the absence of growth factor addition. The curve through each set of data is the result of an unknown interpolation.

into whether cyclin D and/or myc are sufficient to produce the observed changes.

Lai et al. [60] investigate the effects of estrogen, insulin and combinations of estrogen and insulin on S phase progression in MCF-7 breast cancer cells. Cells growing in serum-free medium were exposed to ICI 182780 and left to reach phase steady-state. Cells were then exposed to either estrogen, insulin or estrogen + insulin and the S phase percentage was determined at time intervals for 36 hours following exposure. The experimentally determined S phase profiles are shown in Figure 5.11. After insulin exposure (black triangles), there is a minor increase in the S phase proportion over the next 36 hours, with the S phase percentage increasing from about 5 to 12% after 22 hours. Estrogen (white squares) is more effective at increasing the S phase proportion. After estrogen exposure, the S phase proportion begins to increase by 13 hours, reaches a maximum of about 48% after about 24 hours of exposure, and thereafter decreases slowly, reaching 30% by 36 hours after exposure. Combinations of insulin and estrogen (black squares) have the most potent effect on the S phase percentage, with the S phase percentage increasing by 13 hours to 22%, reaching a maximum of 70% by 26 hours, and decreasing after this point to about 26% by 36 hours.

Cyclin D concentration changes were also observed in [60]. Cyclin D changes were measured every 5 hours for the first 25 hours of exposure to insulin (I), estrogen (E2) and

combined insulin and estrogen (I + E2). The experimentally observed changes, along with a piecewise polynomial which we fitted through the data points, are shown in Figure 5.12a. We will not give details of how the piecewise polynomial through these data points was determined (see discussion around Figure 5.8 on how we choose to perform curve fitting through such data points). We can see that cyclin D reaches a maximum of about 2.5 fold after insulin exposure, 5 fold after estrogen exposure and about 8 fold after estrogen + insulin.

The changes in *myc* concentration were not investigated in this experiment. However in [69], cells growing in the same culture conditions were exposed to the same combinations of growth factors, and the changes in the concentration of *myc* was determined 4 hours after exposure, which was reported to be when *myc* reached its highest concentration. We choose to use the *myc* changes reported in [69] along with the form of the cyclin D changes shown in Figure 5.12(a) to model cell cycle progression after insulin, estrogen and insulin + estrogen exposure. Cyclin D changes were also investigated in [69], but it was only measured once at 5 hours after mitogen exposure. We note that the measured cyclin D concentration at this point in time was similar to that observed in [60].

Results from [69] only show the *myc* concentration at 4 hours after exposure to the mitogens. To determine a continuous *myc* profile after growth factor exposure, we assume that the *myc* profile is qualitatively similar to the cyclin D profile when exposed to the same mitogen, with the only difference being *myc* reaches its highest concentration 4 hours after exposure, with a magnitude equal to that observed in [69]. The resulting *myc* profiles are shown in Figure 5.12b. We can see that *myc* concentration reaches a maximum of about 2 fold after insulin exposure, 8 fold after estrogen exposure and about 10 fold after estrogen + insulin.

As we did when considering the experiment in [98], we assume that *myc* is reduced to levels 4 times lower than cyclin D due to the presence of ICI 182780. Substituting the relative changes in cyclin D and *myc* into equation (5.2), and then substituting the value for  $\gamma$  into equations (3.7), we can generate the S phase changes for up to  $25 + \tau_{G_c} = 30.75$  hours after exposure to insulin, estrogen and insulin + estrogen, shown in Figure 5.13a. The modelled changes in the S phase proportion compare remarkably well with the experimentally observed changes after I, E2 and I + E2 for the first 25 hours.

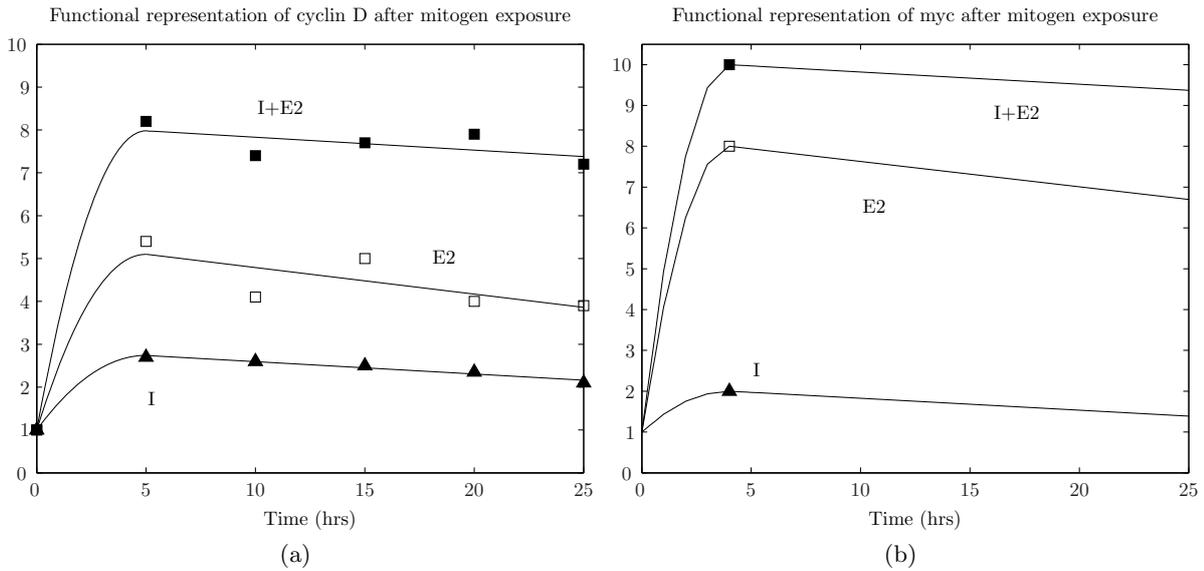


Figure 5.12: *Piecewise polynomial fits to the (a) cyclin D data points after exposure to insulin (I, black triangles), estrogen (E2, white squares) and insulin + estrogen (I+E2, black squares). The myc function (b) uses the qualitative features from the cyclin D profile in (a) but takes the polynomial through the known values of myc at points 4 hours after mitogen exposure.*

The results of this model provide very strong evidence that changes in cyclin D and myc are sufficient to produce the observed changes in the S phase proportion after exposure to insulin, estrogen or the combination of the two, as was suggested by [69].

Although the fits to the experimental data are generally good, we note that after 25 hours of I + E2, the modelled S phase proportion is clearly lower than that determined experimentally. This could be because at later times, we have not chosen accurate values for our input variables  $d(t)$  and  $myc(t)$ , and thus have not got a high enough value for  $\gamma$ . Our modelled S phase proportion decreases after 25 hours because the cells which entered the S phase hours earlier have completed the S phase and are moving onto the M phase (not shown). For the S phase to remain high beyond  $t = 25$  hours, we need the rate at which cells enter the S phase to be similar to the rate of exit. If we consider equation (3.7d), the rate of influx into the S phase depends upon the number of cells in the  $G_b$  phase. If there are very few cells in the  $G_b$  phase, then no matter how large the value of  $\gamma$  is, the cell population will be very unresponsive to the environmental changes. Figure 5.13(b) shows the modelled  $G_b$  phase proportion after exposure to insulin + estrogen. After  $t = 20$  hours the  $G_b$  phase proportion is so low that any subsequent changes to the value of  $\gamma(t)$  will not produce significant changes in the modelled S phase proportion. Thus the

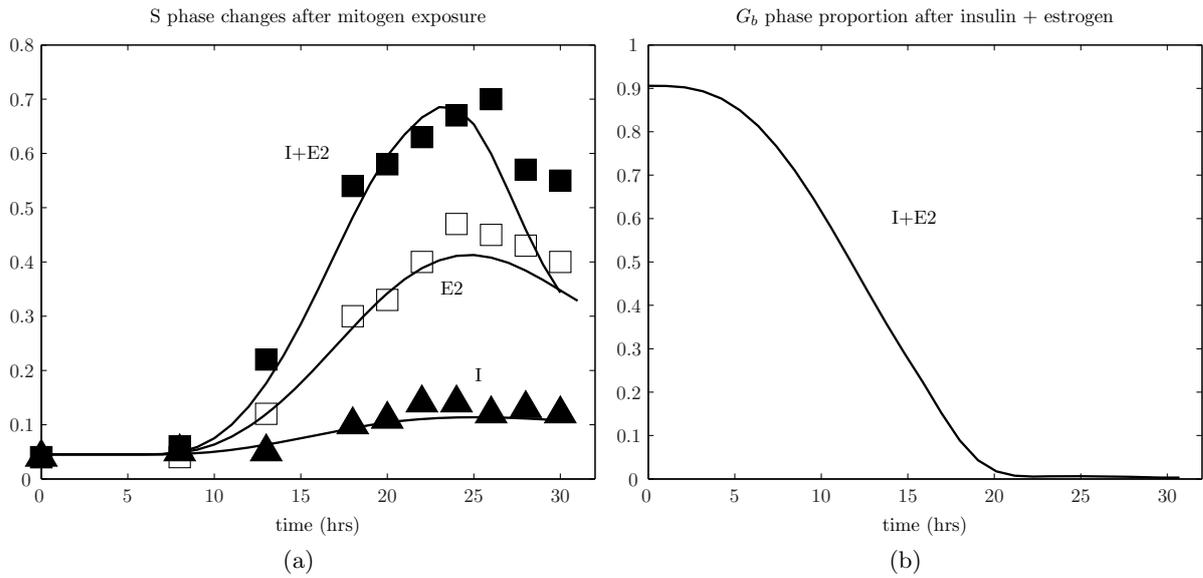


Figure 5.13: The modelled S phase changes and experimentally observed S phase changes are shown in (a) after insulin (I, black triangles), estrogen (E2, white squares) and insulin + estrogen (I+E2, black squares). The data points are representative of at least 3 independent experiments, however error bars are not provided. The proportion of cells in the G<sub>b</sub> model phase after exposure to I+E2 is shown in (b).

discrepancy between modelled and experimentally observed S phase changes at these later times is unlikely to be due to the concentrations of cyclin D and myc used to generate the S phase profile. Another possibility is that the values of the storage phase durations chosen, particularly the value of  $\tau_S$ , does not allow cells to remain for long enough in the S phase. However, trying different values of  $\tau_S$  within the bounds discussed earlier in Section 4.1 does not produce a significantly different S phase profile (not shown). We note that as there are no error bars provided on the experimentally observed S phase changes after mitogen exposure, it is difficult to determine whether the modelled S phase profile after I+E2 for these later times is still a reasonable representation of the observed S phase changes.

In this section, we have used the mathematical model to provide strong quantitative evidence to the claim that cyclin D and myc are sufficient drivers when observing cell cycle phase changes after exposure to various growth factors. We were able to simulate cell cycle progression in MCF-7 cells after exposure to a range of mitogens, such as insulin, IGF and estrogen, and our model produces S phase changes that compare very well with experimental data. The model was also able to quantify the number of cells which would

be responsive to these growth factors at any time (i.e. the number of  $G_b$  phase cells), a critical feature previous models have not been able to capture. It is useful to be able to know how many cells are in this unique responsive phase at any point in time, as the potency of a drug or growth factor depends highly upon how many cells are in this phase.

In the next section, we model the effects of tamoxifen on cell cycle progression, and gain insight into the phases during which this drug exerts its effects.

## 5.7 Modelling the effects of anti-estrogens

In the previous section we used information provided about the changes in cyclin D and myc concentrations to verify certain claims about the causes of S phase increases after exposure to mitogens. In addition to using cyclin D and myc to solve for the S phase proportion after substance exposure, the model also has the potential to use the S phase profile to solve the inverse problem for cyclin D or myc (if one of these is also known). In this section, we use the model to determine how myc concentrations changes in response to various anti-estrogens.

We consider an anti-estrogen experiment from [134], in which cell cycle phase changes were investigated after exposure to the anti-estrogen ICI 164384. This anti-estrogen, like ICI 182780, is a pure anti-estrogen, however it is known to be less effective than ICI 182780 at reducing proliferation [129]. The experimental results from [134] inform us of the S phase changes and cyclin D concentration changes after this anti-estrogen, however myc changes were not investigated. We use our model in this section to determine how myc concentrations change in response to this anti-estrogen. To the best of our knowledge, the changes in myc concentration after ICI 164384 have never been quantified experimentally.

The experimental results from [134] do not explicitly provide changes in cyclin D concentrations after anti-estrogen. Instead, cyclin D mRNA changes are provided. Messenger RNA (mRNA), is the template for protein synthesis read off from DNA, and can be thought of as an intermediate step between reading the code of the protein from DNA and creating the protein. Figure 5.14 shows the experimental results regarding the changes in cyclin D mRNA concentration after exposure to ICI 164384. We will assume that the changes in cyclin D protein follows the same curve as the mRNA, as it has been observed

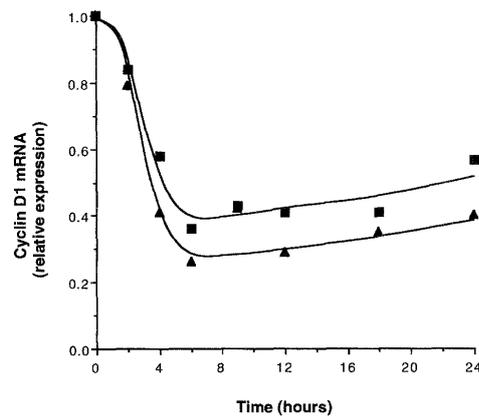


Figure 5.14: *The cyclin D1 mRNA profile was reproduced from [134]. Black squares represent the relative changes in cyclin D1 over time after exposure to the anti-estrogen ICI 164384. The curve through the data points is the result of an unknown curve fit through the data points. Black triangles, and the corresponding curve through them, represent tamoxifen-induced changes in cyclin D1 concentration, and will be discussed in more detail in Section 5.8.*

that cyclin D protein and mRNA concentration parallel each other [123]. It is not always the case that cyclin D protein and mRNA are equivalent, as has been observed in [137, 98] (the relationship between cyclin D mRNA and cyclin D protein from [98] after estrogen can be seen in Figure 5.7(a) of this thesis). In [98], the reason for the disparity is suggested to be the presence of estrogen, which may alter the rate of decay of cyclin D protein without altering the corresponding rate of mRNA decay. In [137], it was suggested that the reason cyclin D mRNA and cyclin D protein levels were not equivalent is because of disparity between the two different methods used to detect protein and mRNA. However, for the experiment in [134] (which we consider in this section), we assume that cyclin D protein and mRNA will parallel each other as is generally expected.

In addition to the changes in cyclin D concentration shown in Figure 5.14, the authors also present data on the changes in the S phase proportion after ICI 164384 exposure, shown in Figure 5.15. In response to ICI 164384, the S phase proportion begins to decrease from about 35%, reaching a minimum of about 13% 25 hours after ICI 164384 exposure. The S phase proportion remains relatively unchanged after 25 hours, settling at 15-18%.

As we discussed earlier, a rich enough source of data on the S phase progression after exposure to an environment change is enough to solve the inverse problem for  $\gamma$ . As we also know the cyclin D profile from Figure 5.14, the only unknown from equation

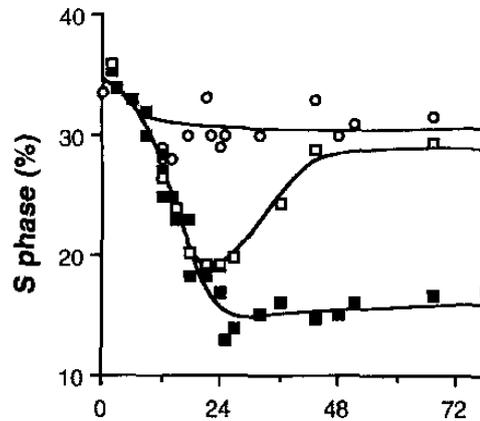


Figure 5.15: *The S phase profile reproduced from [134]. Black squares represent the changes in the S phase proportion over time after exposure to the anti-estrogen ICI 168384. The curve through the data points is the result of an unknown curve fitting process. White circles represent S phase proportions in the absence of anti-estrogen addition. White squares represent a different experiment, which is not relevant to the discussion here.*

(5.2) is the myc concentration over time. Thus, we can use the cyclin D and S phase profiles to solve the inverse problem for myc. As we know very little about the shape of the myc curve, we could represent myc by a high degree polynomial, and find the coefficients of the polynomial which produces the best fit to the S phase curve. High degree polynomials allow for a wide range of curves which we can fit to, however fitting to high degree polynomials result in long computation times, as all of the coefficients need to be determined. In addition, high degree polynomials have a lot of flexibility and will often attempt to fit every single data point, consequently picking up unwanted noise from the data. We could guess a functional form for myc to simplify the polynomial representation, however without knowledge of  $\gamma$ , it is difficult to make any guesses about the shape of the myc function. Thus, instead of using the S phase profile to fit directly to myc, we will use the S phase profile to first solve the inverse problem for  $\gamma$ . Once we have a  $\gamma$  function, we can solve the inverse problem for myc from equation (5.2). Although it appears that we have to do twice the amount of work - solve two inverse problems instead of one - solving the inverse problem for myc from the  $\gamma$  function involves solving a linear inverse problem, as all other variables in equations (5.2) are known for this experiment, and we can simply re-arrange to make myc the subject.

To solve the inverse problem for  $\gamma$ , we will assume that it can be represented by a

polynomial. Unlike the myc curve, we can make an educated guess on what the  $\gamma$  curve must look like in order to generate the S phase profile shown. The fact that the S phase decreases quickly within the first 24 hours of ICI 164384 exposure, and then remains low for most of the experiment, suggests that  $\gamma$  must also decrease quickly and then remain essentially unchanged after a certain time. The rate of change in the number of S phase cells, which can be written as

$$\frac{dS(t)}{dt} = \gamma(t - \tau_{G_c})N_{G_b}(t - \tau_{G_c}) - \alpha N_{S_b}(t),$$

shows that  $\gamma$  affects the S phase proportion with a delay of  $\tau_{G_c} = 5.75$  hours. It might be tempting to assume that since the S phase proportion reaches a minimum at about 25 hours,  $\gamma$  would reach its minimum at about  $25 - \tau_{G_c} = 19.75$  hours. However, we note that the rate at which cells leave the S phase depends upon the number of  $S_b$  cells. When cells first enter the S phase, they enter the  $S_a$  phase, and after a delay of  $\tau_S = 9$  hours, cells enter the  $S_b$  phase, where the rate of leaving is quite high (see equations (3.7)). Thus, the S phase may reach its minimum not because the rate at which cells enter the S phase has reached a minimum (due to  $\gamma$  having reached a minimum), but because the rate at which cells leave the S phase has increased (due to an increase in the number of  $S_b$  phase cells). This, too, could cause a turning point in the proportion of S phase cells. We will let  $t_c$  denote the point at which the  $\gamma$  function takes its minimum, and use optimisation to determine what this value should be.

Thus, we will assume that  $\gamma$  decays according to a quadratic, reaching a minimum at  $t = t_c$  hours, and thereafter, we will assume that  $\gamma$  remains unchanged in value. Omitting the details, we determine the optimum quadratic representation for  $\gamma$ , and the resultant  $\gamma$  curve is shown in Figure 5.16(a). The  $\gamma$  curve shown, when substituted into equations (3.7) produces a G and S phase profile as shown in Figure 5.16(b). We can see that the chosen  $\gamma$  curve produces G and S phase changes that compare very well with experimental data (shown as the asterisks in Figure 5.16(b)).

Now that we have a  $\gamma$  profile, we can use the cyclin D profile from Figure 5.14 to determine an myc profile. We still need to develop a continuous representation for cyclin D, as so far we only have data at distinct time points after ICI 164384 exposure. As there is not much noise in this data, we will generate a continuous representation of cyclin

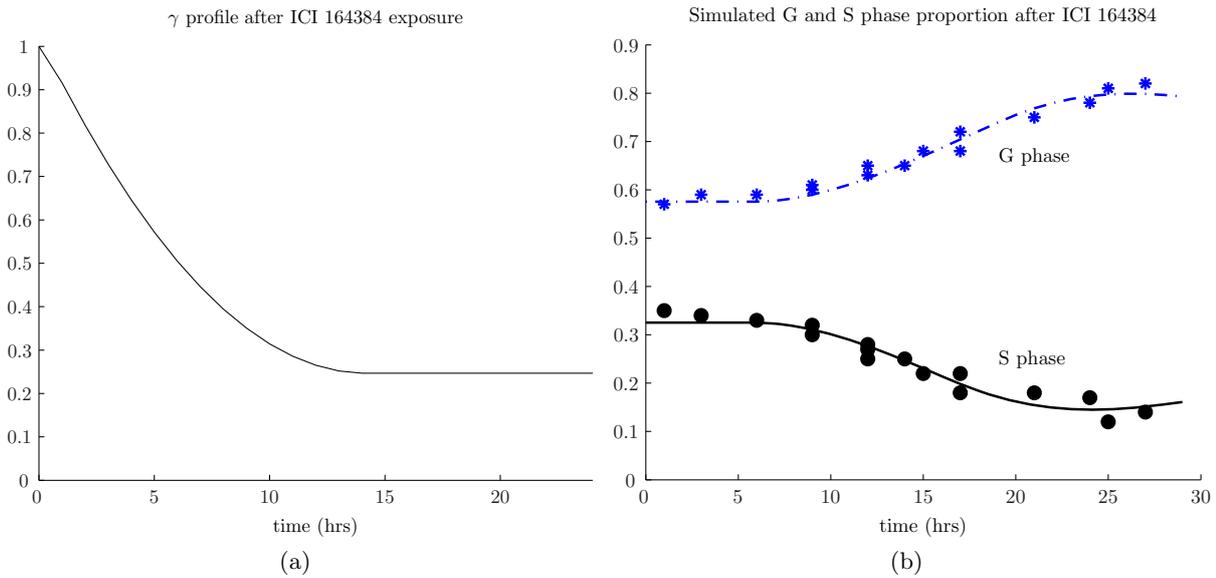


Figure 5.16: The *S* phase profile from [134] was used to solve the inverse problem for  $\gamma$ , shown in (a). Substituting the  $\gamma$  function shown in (a) into equations (3.7) gives the corresponding corresponding *G* and *S* phase profiles as shown in (b). The known experimental changes in the *G* and *S* phase proportions after ICI 164384 are shown as the black dots and blue asterisks respectively.

D concentration by simply interpolating the data points. We must deal with one last complication before we can determine the myc profile for this experiment. We are only provided relative changes in cyclin D concentration after anti-estrogen exposure, and not absolute values of cyclin D concentration. To remedy this, we only need to determine the concentration of cyclin D at  $t = 0$ . Recall that we know the value of  $\gamma(0)$ . If we make the assumption that  $d(0) = m(0)$ , then we can determine the absolute concentration of cyclin D (see Section 5.5 for more details). We can now simply substitute the  $\gamma$  and cyclin D profile into equation (5.2) and solve for myc, which is shown in Figure 5.17. We can see that the myc curve decreases within one hour after anti-estrogen exposure, reaching a minimum of about 45% of its initial value. Cyclin D concentrations also decrease to reach about 40% of control, however this occurs on a much slower timescale, and is not observed to reach this value until about 6 hours. The concentration of myc has been observed to decrease earlier than cyclin D in other antiestrogen experiments [20] and, as we saw in Figure 5.7b Section 5.5, myc also changes in concentration earlier than cyclin D when cells are exposed to estrogen. Thus the observation here that myc changes in concentration significantly earlier than cyclin D is in agreement with the general behaviour of myc.

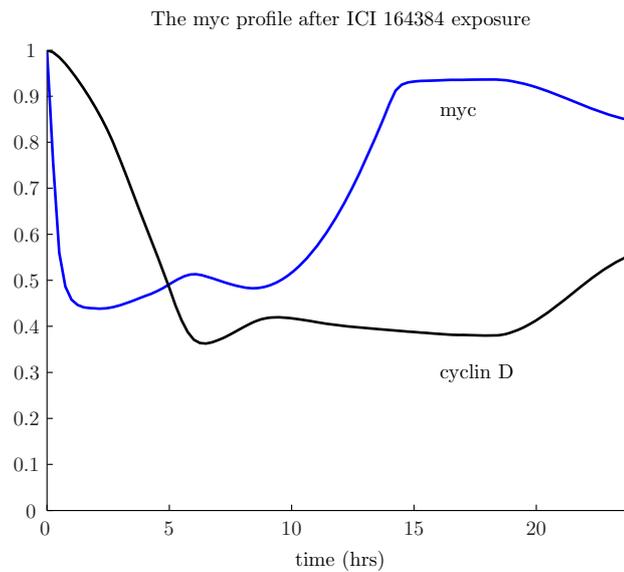


Figure 5.17: The  $\gamma$  and cyclin D profiles were used to solve for myc from equation (5.2). The resultant relative change in myc is shown as the blue curve, and can be compared with the known cyclin D profile in the black curve.

The myc concentration does not change significantly from its minimum until about  $t = 10$  hours when it begins to increase, eventually reaching control levels at 14 hours after exposure, and thereafter, does not change significantly. The increase seen at 10 hours is not something that has been observed in previous experimental observations of myc changes after antiestrogen. In [20], myc concentrations remained low after ICI 182780 for the full 24 hours of observation. Perhaps this is why it is thought that ICI 182780 is more potent than ICI 164384 - because it keeps myc levels lower for longer.

### Sensitivity analysis: solving the inverse problem for myc

We wish to investigate how sensitive the model is to inputs like cyclin D and myc. In particular, we wish to know what features of our myc profile shown in Figure 5.17 are critical in producing an accurate S phase profile. Let's start by considering how changes in myc at early times influences cell cycle progression.

What happens to the cell cycle phase proportions if myc remains unchanged for the first 2.5 hours instead of decreasing rapidly we found it does when solving the inverse problem for myc? This 'test' myc profile shown in Figure 5.18(a) produces a G and S phase profile shown in 5.18(b). We can see that the G and S phase profiles no longer

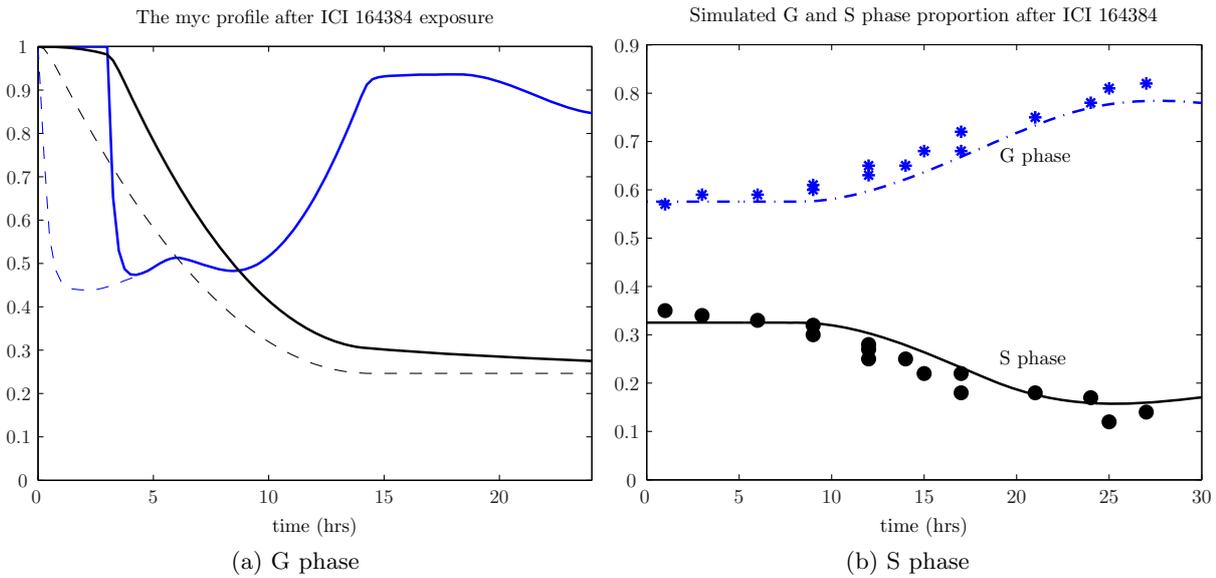


Figure 5.18: *The original myc and corresponding  $\gamma$  functions are shown in (a) as the dotted blue and black curves respectively. The delayed myc curve and corresponding delayed  $\gamma$  functions are also shown in (a) as the solid blue and black curves respectively. The modified  $\gamma$  curve was substituted into equations (3.7) and the G and S phase proportion were evaluated and shown here in (b) as the blue and black curved respectively. Experimental data are also shown as the asterisks and dot.*

compare as well with the experimental data, and this is likely to be due to the fact that an early change in the myc profile produces a change in  $\gamma$  that persists for the entire experiment, as can be seen in Figure 5.18(a). The  $\gamma$  curve is now persistently higher than the original  $\gamma$  curve, producing an S phase profile that is consistently higher than observed values. We conclude that influencing the myc profile at early times can significantly alter the cell cycle phase profiles. It is likely that after ICI 164384 exposure, myc must decrease quickly in order to produce the observed changes in the G and S phase proportions. As mentioned previously, the fast reaction of myc to the anti-estrogen has been observed in other experiments.

We would also like to investigate whether the behaviour of myc at later times also significantly alters the G and S phase profiles. If we assume that after the initial decrease in the first 10 hours of exposure, myc remains unchanged for the remainder of the experiment, we produce an myc curve as shown in Figure 5.19(a). The corresponding G and S phase profile is shown in Figure 5.19(b). This representation shows that alterations in myc at later times does not significantly alter the G and S phase profiles. This is likely to be

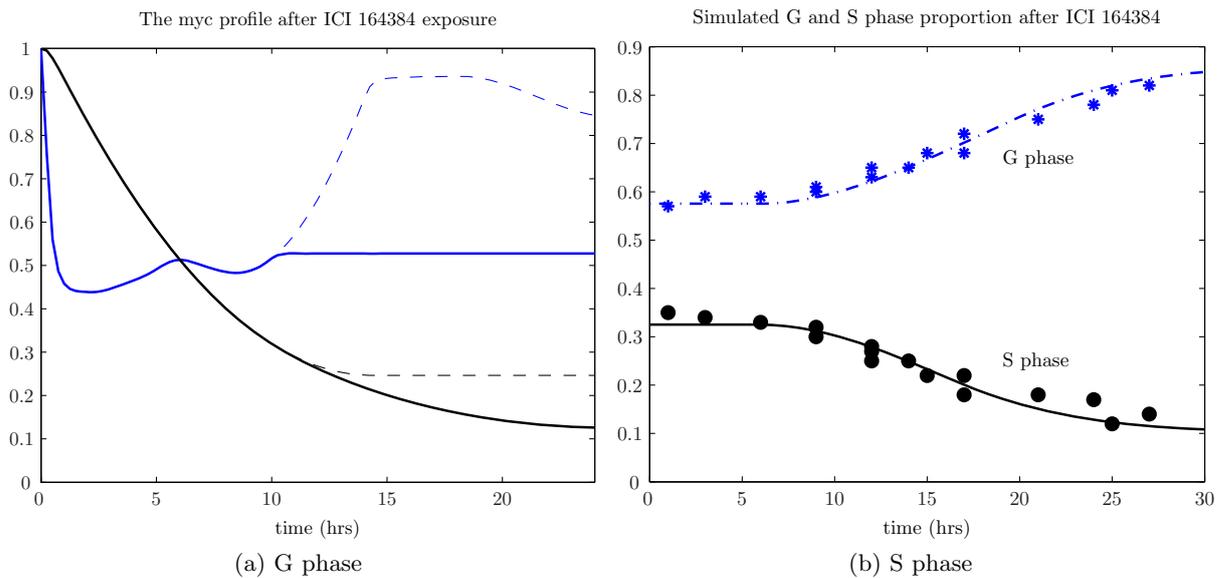


Figure 5.19: *The modified myc shown in (a) produces the G and S phase profiles in (b).*

due to the fact that the modified  $\gamma$  curve does not differ from the original  $\gamma$  curve until after about 14 hours. As  $\gamma$  changes do not influence the S phase profile until after a delay of  $\tau_{G_c} = 5.75$  hours, we would expect to see changes in the cell cycle phase proportions only after about 19.75 hours. Indeed, the G and S phase proportions compare well with experiments until after 20 hours, when the S phase profile is slightly lower than what is observed experimentally. However, the difference is not great, and we conclude that this is still a reasonable profile for the G and S phase. Thus, this modified myc profile produces cell cycle phase changes that compare well with the data. As mentioned previously, myc concentrations after antiestrogen exposure have been observed to decrease quickly initially, but have not been observed to increase again at later times. It is likely that after ICI 164384, myc does not increase as we observed in Figure 5.17, and that the representation shown here in Figure 5.19 is a more biologically realistic profile for myc after ICI 164384 exposure.

In this section, we have solved the inverse problem for myc after exposure to ICI 164384, which has not been investigated in experiments before (to our knowledge). We have also seen that using S phase data to solve the inverse problem for concentrations of myc produces accurate results for the first few hours of myc changes, however for later times, this method of inverse solving is not very reliable. Inverse solving for an myc profile (and, presumably for a cyclin D profile) using the model will not produce an accurate

profile for these proteins without more qualitative information regarding the behaviour of the proteins after exposure at later times.

We now apply the model to tamoxifen-exposed cells, and extend the model eventually to include the effects of tamoxifen-induced cell death.

## 5.8 Application of the model to tamoxifen exposed MCF-7 breast cancer cells

The antiestrogen tamoxifen is known to decrease the proliferation rate of many breast cancer cell types, and has been used as a first line treatment for post-menopausal female breast cancer patients. However, unlike ICI 164384 and ICI 182780, tamoxifen is not a pure antiestrogen, as it has been shown to increase the expression of some growth promoting genes [45]. Investigations have been performed to determine the proteins that are altered in response to tamoxifen in MCF-7 breast cancer cells. Tamoxifen has been shown to decrease expression of cyclin D1 mRNA [134]. In [118] it was found that tamoxifen does not affect the expression of the *myc* gene. However investigations in [45] showed that tamoxifen transiently increased expression of the *myc* gene, which is in disagreement with the results from [118]. To our knowledge, there has not been any further studies which have attempted to reconcile the conflicting results of these two investigations.

In the previous section, we saw that Watts et al. [134] investigated cell cycle progression after cells were exposed to ICI 164384. They also investigated the effects of tamoxifen on cell cycle progression. MCF-7 cells were exposed to 100nM tamoxifen and the concentrations in cyclin D mRNA, as well as the proportion of cells in the G and S phases were determined at intervals after tamoxifen exposure. In this section, we use the experimental data to solve the inverse problem for *myc*, keeping in mind that using our model to inverse-solve for *myc* may not produce accurate results for the *myc* profile at later times. The results we get from determining the inverse profile for *myc* will also be used in the following section, when investigating the effects of tamoxifen-induced death.

The changes in cyclin D concentration after tamoxifen exposure are shown in Figure 5.14 on page 151 (triangles represent data). After tamoxifen exposure, cyclin D1

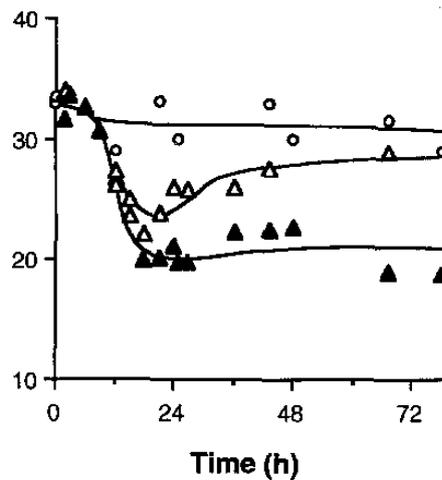


Figure 5.20: *The S phase profile reproduced from [134]. Black triangles represent the changes in the S phase proportion over time after exposure to the anti-estrogen tamoxifen. The curve through the data points is the result of an unknown curve fitting process. White circles represent S phase proportions in the absence of anti-estrogen addition. White triangles represent a different experiment, which is not relevant to the discussion here.*

mRNA decreases, reaching a minimum of 0.25 of control levels 5 hours after tamoxifen exposure. After this time, cyclin D1 mRNA levels increase slowly, reaching about 0.35 of control levels 24 hours after tamoxifen exposure.

The changes in the S phase proportion were also investigated after tamoxifen exposure, and are shown in Figure 5.20 as the black triangles. In response to tamoxifen, the S phase proportion begins to decrease from about 32% in control cells to 20% 18 hours after tamoxifen exposure, remaining this low until at least 25 hours after tamoxifen exposure. A slight increase in the S phase percentage is observed at the next time of observation 26 hours after exposure, where the S phase increased marginally to about 23%, remaining at about this level until 48 hours after exposure. By 66 hours, the S phase proportion decreased again, remaining at about 18% for the remainder of the experiment. This seeming increase at 48 hours and decrease at 66 hours could well be due to noise and experimental error. In fact the amount of noise in this experiment can be seen when we consider the hollow circles in Figure 5.20, which represent the observed S phase proportion in phase steady-state. We would expect the S phase proportion in phase steady-state to be constant for the duration of the experiment. However, we observe that the hollow dots fluctuate between about 28% and 34% throughout the experiment, and so the change in the S phase proportion after tamoxifen from 23% at 48 hours to 18% at 66 hours could

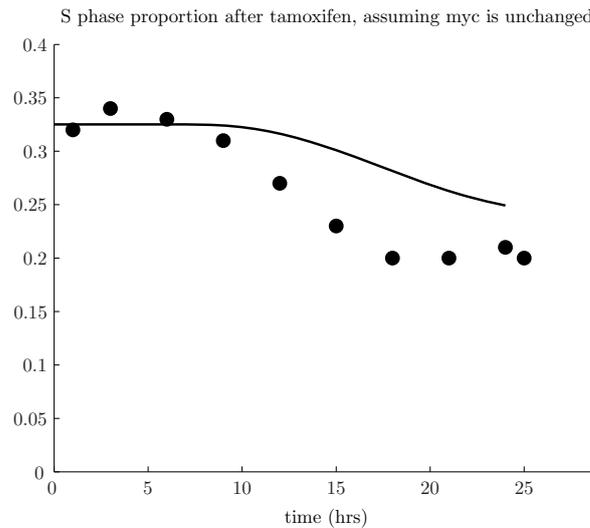


Figure 5.21: *The S phase changes after tamoxifen exposure when the cyclin D profile from Figure 5.14 is used to generate  $\gamma$ , and myc is assumed unchanged in the presence of tamoxifen. The dots represent observed changes in the S phase proportion after tamoxifen exposure from Figure 5.20.*

be due to noise.

If we compare the changes in cyclin D1 mRNA after tamoxifen with those after ICI 164384 (black circles, same figure), we note that tamoxifen consistently reduced cyclin D1 mRNA to lower levels than did ICI 164384. However, if we compare with the observed changes in the S phase percentage after ICI 164384 exposure in Figure 5.15, we note that the S phase percentage after tamoxifen exposure is consistently higher than the S phase percentage after ICI 164384 exposure. This is likely to be due to the fact that myc is regulated differently after tamoxifen, as discussed earlier.

We now use our model and the experimental data from [134] to investigate how myc concentrations change after tamoxifen exposure. As discussed earlier, the experimental results from [118] indicated that myc concentrations do not change after exposure to tamoxifen. We use the observed changes in cyclin D concentration from [134] and, assuming that myc does not change in concentration, we can determine the S phase changes using our model. As we did in Section 5.7, we assume that prior to tamoxifen exposure  $d(0) = m(0)$ . Substituting the cyclin D profile from Figure 5.14 into the  $\gamma$  equation from (5.2) and assuming that myc concentrations do not change after tamoxifen exposure produces an S phase profile shown in Figure 5.21. It is clear from this figure that our

modelled S phase profile does not compare well with experimental data. In order to accurately replicate the observed S phase changes, we must assume that *myc* concentrations *decrease* after tamoxifen exposure. This is in disagreement with [118], which states that *myc* is unaffected by tamoxifen and [45], which states that *myc* *increases* transiently after tamoxifen. We wish to further investigate how much *myc* concentrations must decrease by in order to produce S phase changes that compare well with experimental results. To do this, we solve the inverse problem for *myc*.

We use the S phase profile from Figure 5.20 and the cyclin D profile in Figure 5.14 to solve the inverse problem for *myc*. As we did in Section 5.7, we interpolate the cyclin D and S phase data points so that we have continuous representations of these variables. We then solve the inverse problem for  $\gamma$  and then solve the linear inverse problem for *myc* using equation (5.2). To do this, we must determine a functional form for  $\gamma(t)$  after tamoxifen exposure. As the S phase profile after tamoxifen exposure is qualitatively similar to the S phase profile after ICI 164384 exposure from Figure 5.15, we assume that  $\gamma$  has the same functional form as when solving the inverse problem for *myc* after ICI 164384 exposure. Once we determine the  $\gamma(t)$  profile using this inverse-solving technique (not shown), the *myc* profile is then determined by rearranging equation (5.2). Doing this gives us the *myc* profile as shown in Figure 5.22(a), and the corresponding G and S phase profiles are shown in 5.22(b). After tamoxifen exposure, *myc* decreases in concentration quickly, reaching a minimum of about 45% of control after about 1 hour of exposure. After this time, *myc* increases, reaching control levels by 10 hours after exposure. The concentration of *myc* remains high, slightly exceeding control levels for the remainder of the experiment. As discussed earlier, experimental observations of changes in *myc* after anti-estrogen exposure indicate that *myc* changes in concentration faster than cyclin D does, and so the fact that we have achieved a similar feature here is in agreement with these observations. However, the later increase in *myc* concentration is not a feature that has been described previously. As we know that inverse-solving for *myc* at later times is not accurate, we wish to investigate whether we can still produce reasonable results if we assume that *myc* does not increase so significantly after it reaches its minimum.

However, we cannot produce accurate results if we assume that *myc* remains low for the entire duration of the experiment. In fact, we find that *myc* has to increase significantly after it reaches its minimum in order for the modelled cell cycle phase proportions to

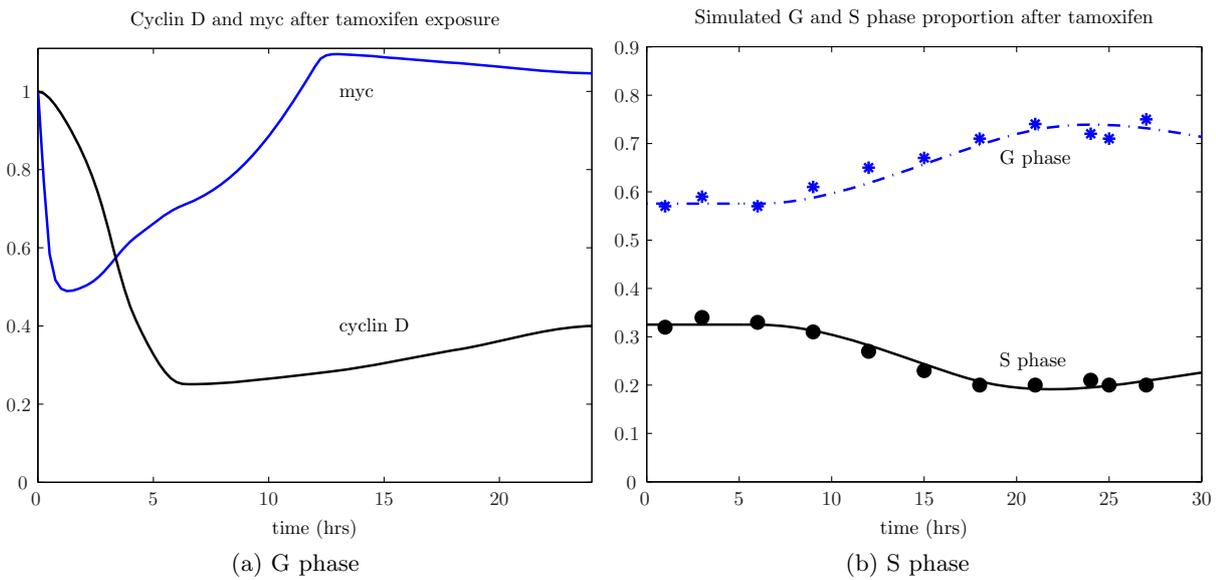


Figure 5.22: The *myc* profile determined using our model is shown in (a) as the blue line, and the continuous cyclin D profile is also shown here as the black line. The cyclin D and *myc* profiles are used to generate the G and S phase curves shown in (b) as the blue and black curves respectively. The asterisks and black dots represent G and S phase data respectively.

compare well with experiments. For instance, if we assume that *myc* increases to about 75% of controls and does not increase further, we produce Figure 5.23. It is clear that for later times, the G and S phase profiles do not compare well with experimental data because too few cells are leaving the G phase and entering the S phase at these times. Thus, we must have *myc* at higher levels at these times. We assume that the original *myc* profile from Figure 5.22 is the more accurate profile, and we use it in the following section where we further investigate the effects of tamoxifen.

We noted earlier that other experimental observations of changes in *myc* after tamoxifen exposure indicate that *myc* does not decrease after exposure to tamoxifen. We do not know why our modelled results indicate that *myc* decreases significantly after tamoxifen exposure, and why this protein then increases to control levels again by 10 hours after exposure.

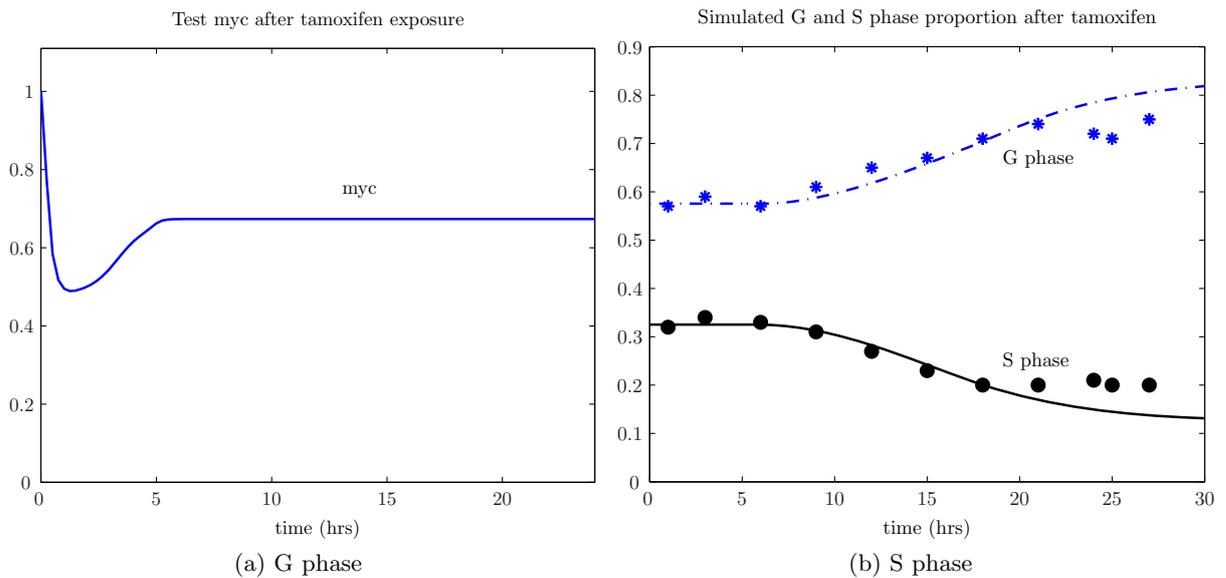


Figure 5.23: The *myc* profile determined using our model is shown in (a) as the blue line, and the continuous cyclin *D* profile is also shown here as the black line. The cyclin *D* and *myc* profiles are used to generate the *G* and *S* phase curves shown in (b) as the blue and black curves respectively. The asterisks and black dots represent *G* and *S* phase data respectively.

## 5.9 Tamoxifen exposure to mitotically selected cells

We wish to explore the effects of tamoxifen in MCF-7 cells further by considering other experiments which investigate cell cycle progression after tamoxifen exposure. In this section, we consider experiments performed in [115], which investigated the changes in cell cycle phase distributions after exposure to various concentrations of tamoxifen. Cells were mitotically selected and re-plated in a medium which contained differing amounts of tamoxifen. The *G*, *S* and *M* phases were subsequently determined at intervals after exposure. Figure 5.24 shows the changes in the *G*, *S* and *M* phase proportions after mitotically selected cells were re-plated in medium without tamoxifen (black circles) and with  $7.5\mu\text{M}$  of tamoxifen. Cells that were mitotically selected and replated in medium without tamoxifen displayed similar changes in the cell cycle phases as was observed in the mitotic selection experiment described in Figure 5.1 (see page 120). By the first point of investigation (about 1 hour into the experiment), the mitotically selected cells had entered the *G* phase, as evidenced by the high *G* phase proportion at this time. Between 6 and 10 hours, the tamoxifen exposed cells began to progress into the *S* phase, with an observed increase in the *S* phase proportion from about 5% before 6 hours to about 15%

NOTE:  
This figure is included on page 164 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 5.24: *Cells were mitotically selected and replated in a medium without tamoxifen (black circles) and in a medium with 7.5 $\mu$ M of tamoxifen (white circles). Points are the average of at least two independent experiments, and points with error bars represent the mean of at least 4 independent experiments. The curve through the data points is the result of an unknown curve fitting process. Results obtained from [115].*

10 hours after selection. The S phase proportion reached a maximum of 50% by about 11-12 hours after selection, remained high until 16 hours after selection, and thereafter began to decrease, remaining low for the remainder of the experiment. It is clear that the tamoxifen-exposed cell populations have a reduced rate of entry into the S phase, as evidenced by the smaller proportion of cells in the S phase after selection. Mitotically selected cells that are exposed to tamoxifen do not enter a second round of cell division after 24 hours of selection, in contrast to cells that are not exposed to tamoxifen.

We wish to simulate changes in the G, S and M phase proportions after exposure to tamoxifen in mitotically selected cells, and then compare our simulated results with the experimental data presented here. To do this, we must be able to simulate cell cycle

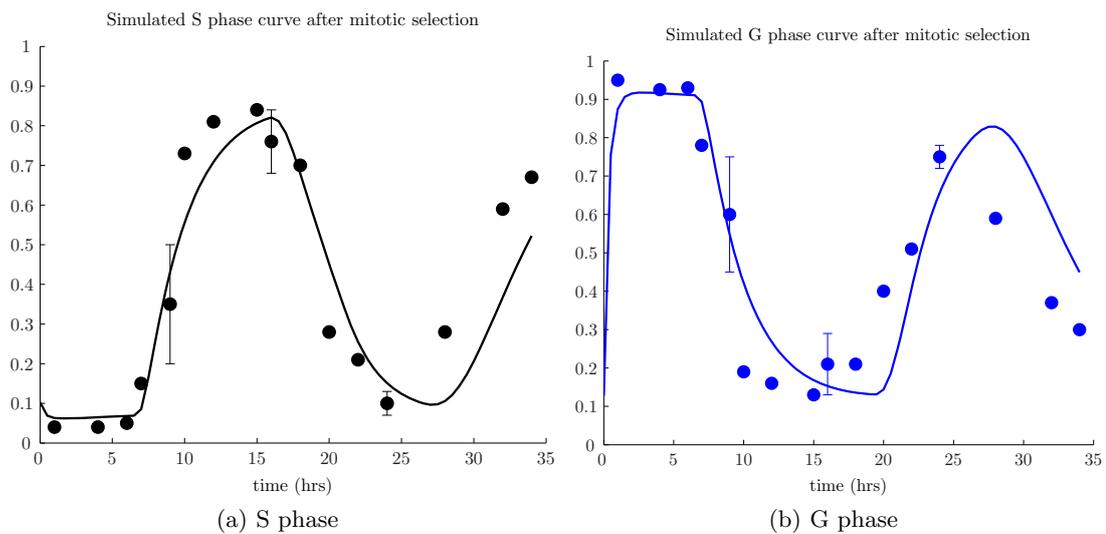


Figure 5.25: The proportion of cells in the (a) S phase and (b) G phase after mitotic selection. Experimental data from [115] (see Figure 5.24) on the G and S phase changes after mitotic selection is shown as circles (error bars also shown, where available).

progression in the absence of tamoxifen (as this is the background growth conditions of the experiment), and then we can use our known changes in cyclin D and myc after tamoxifen exposure from the previous section to investigate how mitotically selected cells behave after tamoxifen exposure. Thus, we begin by simulating cell cycle progression of mitotically selected cells in the absence of tamoxifen.

We modelled mitotic selection previously for a different experiment in Section 5.1. As we found in that section, simulating mitotic selection when cells are re-plated in the same medium they have been growing in only requires us to determine the value of two variables -  $\rho$ , and the proportion of the impure cell population at time  $t = 0$ . We use optimisation to determine these values, and we find that  $\rho = 0.03$ , and that the impure cell population makes up 27% of the total population at time  $t = 0$ . The resultant changes in G, S and M phase proportions after mitotic selection in these conditions are shown in Figure 5.25. These figures show that the modelled G and S phase proportions compare very well with experimental data for the entire experiment, and further demonstrates the validity of our model.

To model the changes in mitotically selected cells after re-plating in a medium containing tamoxifen, we use the cyclin D and myc profile determined earlier in Section 5.8 to generate the S phase changes after tamoxifen exposure. However, the concentration of ta-

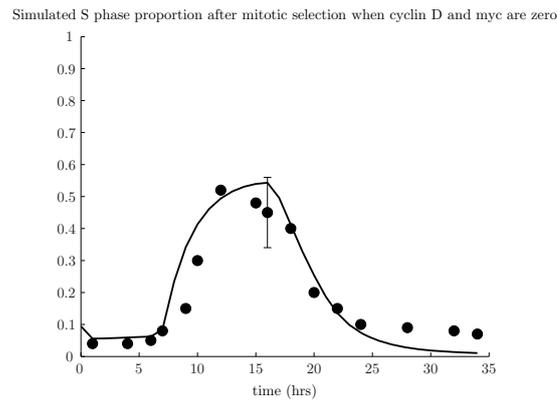


Figure 5.26: *The S phase changes after mitotically selected cells are re-plated into a medium which causes the concentrations of cyclin D and myc to be instantly reduced to zero are shown. Experimental data from [115] (see Figure 5.24) on the S phase changes after mitotically selected cells were re-plated into a medium containing  $7.5\mu\text{M}$  of tamoxifen is shown as black dots (error bars also shown).*

moxifen used in the experiment from [134] was  $100\text{nM}$ . In the mitotic selection experiment shown in Figure 5.24,  $7.5\mu\text{M}$  of tamoxifen was used - a significantly higher concentration. Thus, the cyclin D and myc profiles determined in Section 5.8 may not produce as strong a reduction in the proportion of cells entering the S phase.

In fact, if we simply assume that cyclin D and myc are instantly reduced to zero after  $7.5\mu\text{M}$  of tamoxifen exposure, the most dramatic effect the model can elicit, we produce changes in the S phase proportion after mitotic selection as shown in Figure 5.26 (note that cyclin D and myc having zero concentration within a cell is biologically unlikely, however they could be reduced to such low concentrations that their affect on the value of  $\gamma$  is negligible). In this figure, we have also shown the experimental data representing the observed changes in the S phase proportion after  $7.5\mu\text{M}$  of tamoxifen (copied from Figure 5.24). The modelled S phase proportion after tamoxifen exposure compares well with experimental data for times less that 24 hours after tamoxifen exposure. Thereafter, the artificial stipulation that cyclin D and myc concentrations remain at 0 results in cells being unable to enter the S phase after the first round of cell division.

The reason that some cells are still able to enter the S phase even after the concentration of cyclin D and myc have been reduced to zero is because the  $\gamma$  function does not instantly reduce to zero even though cyclin D and myc do. If we re-consider the  $\gamma$  function from equation (5.2) in the complete absence of cyclin D and myc, then we see

NOTE:  
This figure is included on page 167 of the print copy of the thesis held in the University of Adelaide Library.

Figure 5.27: *MCF-7 cells were mitotically selected and the selected cells were re-plated in medium containing varying concentrations of tamoxifen. The S phase percentage after 9 (first column), 16 (second column) and 24 (third column) hours of tamoxifen exposure is shown. The concentration of tamoxifen in the medium is shown in  $\mu\text{M}$  on the x-axis. Results obtained from [115].*

that  $\gamma$  decays exponentially with rate parameter  $\beta_g$ .

The model results imply that  $7.5\mu\text{M}$  of tamoxifen results in the most dramatic effect on cyclin D and myc - instantly reducing their concentration to 0. We would therefore expect that even higher concentrations of tamoxifen would not produce more drastic reductions in the S phase proportion. However this is not the case. In [115], further investigations were performed, exposing the mitotically selected cells to varying concentrations of tamoxifen, investigating the effects of  $5\mu\text{M}$ ,  $10\mu\text{M}$  and  $12.5\mu\text{M}$  of tamoxifen on cell cycle progression. The authors investigate the S phase proportion after 9, 16 and 24 hours after exposure to these different concentrations of tamoxifen. Their results are reproduced in Figure 5.27. These investigations show that after exposure to  $10\mu\text{M}$  and  $12.5\mu\text{M}$  of tamoxifen, the S phase is significantly more reduced than after exposure to  $7.5\mu\text{M}$  of tamoxifen. Thus, the mathematical model of cell cycle progression as it stands, cannot realistically replicate the observed changes in the G and S phase proportions observed experimentally after higher concentrations of tamoxifen.

A possible reason for the discrepancy between modelled and observed cell cycle phase changes after tamoxifen is that higher concentrations of tamoxifen may cause a significant amount of cell death, and this may affect the distribution of cells across the cell cycle phases. In particular, if tamoxifen-induced cell death occurs primarily in one phase, or a group of phases, this could skew the observed proportion of cells across the different phases of the cell cycle. In [112], it was observed that concentrations of tamoxifen above

10 $\mu$ M not only inhibited cell division (i.e. had a *cytostatic* effect of cells) but also killed cells (i.e. had a *cytotoxic* effect). We introduce cell death and attempt to model the cell cycle phase changes in mitotically selected cell populations exposed to high concentrations of tamoxifen.

### The cytotoxic effects of tamoxifen

To introduce the cytotoxic effects of tamoxifen, we must first determine in which phases this cytotoxic effect takes place. When determining the effects of other drugs on cell cycle progression, we have always assumed that their growth effects occur in the G<sub>b</sub> phase of the cell cycle, however this is not the case with the cytotoxic effects of tamoxifen. In [115], the period during which tamoxifen exerted its cytotoxic effects was investigated. This was done by exposing mitotically selected cells to tamoxifen at intervals after mitotic selection. Mitotically selected cells were exposed to tamoxifen for two-hourly bursts, starting at different times after mitotic selection. After the burst of tamoxifen exposure, cells were washed and re-plated into a drug-free medium. As cell death can take several hours to become apparent, cells were left to grow for 16 hours, and then the proportion of cells in the dish that were still alive was determined. Clonogenic assays were then used to determine the surviving fraction of cells 16 hours after mitotic selection. Clonogenic assays are an accurate but tedious method for determining whether a cell is alive or dead [34].

Tamoxifen bursts during the period 0-2 hours or 2-4 hours after mitotic selection reduced the proportion of cells which entered the S phase when compared to control cells. However, the surviving fraction of cells was 1 at 16 hours after selection (i.e. 100% of cells are alive 16 hours after selection). Cells exposed to tamoxifen during the interval 4-6 hours after selection resulted in a surviving fraction of 0.5 at 16 hours after selection. Cells exposed to tamoxifen in the interval 6-8 hours after selection have a surviving fraction of 0.8. Thus, tamoxifen exposure in the interval 4-6 hours after selection had the strongest cytotoxic effect on cells.

The fact that 100% of cells survived to 16 hours when tamoxifen was exposed in the 2-hourly periods 0-2 hours and 2-4 hours after selection [115], the same period where growth arrest was maximal, implies that cell death must not occur in the same part of

the cell cycle during which cell cycle progression is influenced (i.e. cell death does not occur in the  $G_b$  phase). In a more recent paper, similar investigations were made to determine the phases when cell death due to tamoxifen occurs [26], and similar results were obtained - that tamoxifen did not cause cell death in G1 arrested cells (i.e. cells arrested in the G1 phase due to an anti-growth agent, which would have arrested them in the  $G_b$  model phase). Further investigations by [26] showed that cells arrested in the S phase did experience cell death after tamoxifen exposure, which led the authors to conclude that tamoxifen-induced cell death occurred in the S phase. However, in [115], cells exposed to tamoxifen 6-8 hours after mitotic selection (the 2-hour period with the highest proportion of S phase cells) did not result in as much cell death compared to the 4-6 hour period of exposure (surviving fraction of 0.8 compared with 0.5). These results indicate that the highest rate of cell death occurs in the late G phase, or the  $G_c$  model phase. This is not in disagreement with [26] who only showed that cell death does not occur in cell cycle arrested cells (which we assume are arrested in the  $G_b$  phase, as this is when the vast majority of drugs arrest cell cycle progression). So the experiment performed in [26] did not rule out the possibility that death occurred in the later  $G_c$  phase. Although [26] showed that cell death did occur in the S phase when exposed to tamoxifen, implying that cell death due to tamoxifen can occur in both the  $G_c$  and the S phase. We find that including cell death in both the  $G_c$  and S phase does not produce significantly different results from including cell death in just the  $G_c$  phase. Thus, for simplicity, we only discuss the results from including cell death in the  $G_c$  phase.

### Incorporation of cell death

To include the effects of cell death in the mathematical model, we assume that death occurs at a rate of  $\beta_D(t)$  in the  $G_c$  phase, where the function  $\beta_D(t)$  is yet to be explicitly defined. Upon exposure to tamoxifen, we assume that the concentrations of cyclin D and myc are reduced to zero, as we did earlier when considering the effects of  $7.5\mu\text{M}$  of tamoxifen. We include the effects of cell death in the  $G_c$  phase by adding the term  $-\beta_D(t)N_{G_c}(t)$  into equations (3.7c). However, we must further modify Eq. (3.7c) when considering death in the  $G_c$  model phase. Currently, the rate at which cells exit the  $G_c$  phase and enter the  $S_a$  phase, shown as the decay term in Eq. (3.7c), only depends

upon the number of cells that left the  $G_b$  phase  $\tau_{G_c}$  hours earlier. This means that this transition term will always move exactly the number of cells that exited the  $G_b$  phase  $\tau_{G_c}$  hours earlier into the  $S_a$  phase, and will not consider the reduced number of cells that will be able to make this transition due to the effects of cell death during the  $G_c$  phase.

To update the rate at which cells move from the  $G_c$  phase into the  $S_a$  phase, we note that the number of cells that would normally make this transition at time  $t$  have spent exactly  $\tau_{G_c}$  hours in the  $G_c$  phase, and thus they have experienced a rate of death for the last  $\tau_{G_c}$  hours. This implies that the proportion of cells that have survived the last  $\tau_{G_c}$  hours in the  $G_c$  phase at time  $t$  is

$$\Phi_{G_c}(t) = e^{-\int_{t-\tau_{G_c}}^t \beta_D(s) ds}.$$

We therefore modify the rate of transition from the  $G_c$  phase into the  $S_a$  phase by multiplying by this surviving proportion. We include the cytotoxic effects of tamoxifen into the model as shown below:

$$\begin{aligned} \frac{dN_{G_c}(t)}{dt} &= \gamma(t)N_{G_b}(t) - \gamma(t - \tau_{G_c})N_{G_b}(t - \tau_{G_c})\Phi_{G_c}(t) - \beta_D(t)N_{G_c}(t), \\ \frac{dN_{S_a}(t)}{dt} &= \gamma(t - \tau_{G_c})N_{G_b}(t - \tau_{G_c})\Phi_{G_c}(t) - \gamma(t - \tau_{G_c} - \tau_S)N_{G_b}(t - \tau_{G_c} - \tau_S)\Phi_{G_c}(t - \tau_S), \\ \frac{dN_{S_b}(t)}{dt} &= \gamma(t - \tau_{G_c} - \tau_S)N_{G_b}(t - \tau_{G_c} - \tau_S)\Phi_{G_c}(t - \tau_S) - \alpha N_{S_b}(t), \\ \frac{dN_D(t)}{dt} &= \beta_D(t)N_{G_c}(t), \end{aligned}$$

where  $N_D(t)$  represents the number of dead cells at time  $t$ . The updated rate of transition between the  $G_c$  and the  $S_a$  phase is included as the decay of  $G_c$  phase cells and as the growth of  $S_a$  phase cells. The  $S_a$  phase is a storage phase, and so the rate of exit from this phase is equivalent to the rate at which cells entered this phase  $\tau_S$  hours earlier. Thus we modify the rate at which cells exit the  $S_a$  phase (and hence enter the  $S_b$  phase) to include the effects of cell death. The rates of change in all other model phases are the same as in equations (3.7), and are not shown again here.

We must also consider how the experimental methods used to detect dead cells may alter the results. As discussed in in Section 3.1, there will be a delay in the time taken for dead cells to be detected by flow cytometry. Other investigations have investigated

NOTE:  
This figure is included on page 171 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 5.28: *DNA histogram of mitotically selected MCF-7 cells 16 hours after selection. The top histogram represents the mitotic cell population re-plated in control medium, and the bottom histogram represents the cell population re-plated in medium containing  $12\mu\text{M}$  of tamoxifen. The left-most peak in both histograms corresponds to chicken erythrocytes, which acted as a standard in each experiment. SF refers to the surviving fraction of the population. Results obtained from [115].*

this further, and found that when cells die from tamoxifen exposure, it takes at least 12 hours [93] - 24 hours [48] to observe changes in DNA content. After this time, the dead cells appear to have significantly less DNA than their live counterparts and can therefore be observed using flow cytometry. Therefore, for several hours after apoptosis, these dead cells will appear as live cells when flow cytometry is performed. This implies that the experimental data used in this section and shown in Figure 5.24 may be reporting an artificially high proportion of cells in the G phase as cells that are dying in this phase are still being counted as live G phase cells when flow cytometry is performed.

DNA histograms were also reported in [115], from which we may be able to see if there are any cells that could represent the dead population. The DNA histograms after  $12\mu\text{M}$  of tamoxifen are reproduced in Figure 5.28. These DNA histograms do not indicate the presence of an early peak of dead cells. It would appear that even after 16 hours of exposure, the dead cells do not have noticeably less DNA than their live counterparts, and cannot be spotted on a histogram. This is in agreement with the timescale suggested earlier, that dead cells are detected in flow cytometry after 12 - 24 hours.

Other experts suggest that dead cells lose their attachment to the petri dish they are growing in once they die [Wayne Tilley, personal communication]. These dead cells ‘float’ to the top of the growing layer, and are easily washed away. If this is the case, then there may not be any significant lag time between cell death and their removal from the system, and these cells would also not be detectable on flow cytometry because they have been washed away. However, we believe it is unlikely that dead cells will lose attachment and float to the top as soon as they become apoptotic. We therefore attempt to model the effects of tamoxifen by considering two different cases, which we refer to as Case 1 and Case 2. In Case 1, we assume that the dead cells are being efficiently removed as soon as they die (because they ‘float’ to the top and get washed away), and that flow cytometry is only considering live cells. In Case 2, we assume that the dead cells are not being efficiently removed because they remain attached for some time after they have begun the apoptotic process, and are therefore being artificially included as live cells.

We need to discuss how we model the scenario in Case 2. We would normally calculate, say, the G phase proportion as  $\frac{N_G(t)}{N_T(t)}$ , which is how we calculate this proportion in Case 1. However if the dead cells are artificially being counted as live cells, then since the dead cells are dying in the G phase, the observed G phase proportion would be

$$\frac{N_G(t) + N_D(t)}{N_T(t) + N_D(t)}.$$

The observed S and M phase proportions would be calculated as  $\frac{N_S(t)}{N_T(t) + N_D(t)}$  and  $\frac{N_M(t)}{N_T(t) + N_D(t)}$  respectively. As we mentioned in the previous paragraph, it takes between 12 - 24 hours before dead cells will have a noticeable change in the amount of DNA content within their cells. Thus, we assume that it takes 18 hours until the dead cells are successfully identified, though we note that choosing any other value of time between 12 and 24 hours makes little difference to the results of our model.

Thus, the observed G phase proportion will be calculated as

$$G(t) = \frac{N_G(t) + N_D(t) - N_D(t - 18)}{N_T(t) + N_D(t) - N_D(t - 18)}.$$

Similarly, we calculate the S and M phase proportions as

$$S(t) = \frac{N_S(t)}{N_T(t) + N_D(t) - N_D(t-18)}$$

and

$$M(t) = \frac{N_M(t)}{N_T(t) + N_D(t) - N_D(t-18)}$$

respectively.

### The form of $\beta_D(t)$ after continuous exposure to $12\mu\text{M}$ of tamoxifen

For both Case 1 and Case 2, we assume that after exposure to  $12\mu\text{M}$  of tamoxifen, the concentrations of cyclin D and myc are instantly reduced to zero. We also assume that for times  $t \geq 0$ ,  $\beta_D(t)$  takes on a constant, non-zero value called  $\beta_D$ . Thus, to model the effects of high concentrations of tamoxifen, the only unknown variable is the value of  $\beta_D$ . We choose the value of  $\beta_D$  in both Case 1 and Case 2 such that at 16 hours after tamoxifen exposure, the modelled S phase percentage is 16%, as observed in [115]. Doing this for Case 1 gives  $\beta_D = 0.33$  and the resulting cell cycle phase proportions, as well as the surviving fraction after tamoxifen exposure (calculated by evaluating  $\frac{N_T(t)}{N_T(t) + N_D(t)}$ ), are shown in Figure 5.29.

The inclusion of a constant death rate in the  $G_c$  phase significantly reduces the S phase proportion when compared to just altering concentrations of cyclin D and myc (compare Figure 5.29(a) with Figure 5.26). The value of  $\beta_D$  chosen results in a surviving fraction of 0.5 at 16 hours after tamoxifen exposure, which compares very well with the experimentally determined value of 0.45 shown as the hollow circle in Figure 5.29. In addition, the G and S phase proportions at 9 hours after tamoxifen exposure compare well with the experimentally determined values at this time. However, the modelled G and S phase proportions 24 hours after exposure do not compare well with the experimentally observed proportions at this time. This may be due to the fact that we have assumed that  $\beta_D$  is constant and that cyclin D and myc remain at zero concentration for the full duration of the experiment, whereas in reality these variables may change.

It is interesting to note that continual exposure to high concentrations of tamoxifen does not result in all cells eventually dying, and the surviving proportion appears to

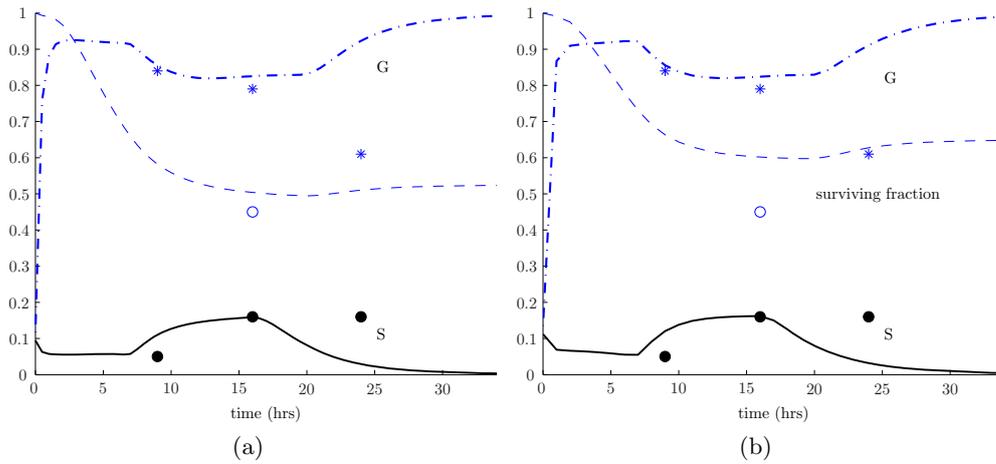


Figure 5.29: The proportion of cells in the G phase (dot-dash curve) and S phase (solid curve) after continuous tamoxifen exposure for (a) Case 1 and (b) Case 2. The surviving fraction is also shown in each figure as the dotted curve. The experimentally observed G and S phase proportions are shown as the asterisks and filled circles respectively and the surviving fraction 16 hours after tamoxifen exposure is shown as the hollow circle.

asymptote to about 0.5 in Figure 5.29(a). This is because in addition to the cytotoxic effects of tamoxifen, tamoxifen also dramatically reduces the concentrations of cyclin D and myc, which arrests cells in the  $G_b$  phase. This prevents cells from progressing into the  $G_c$  phase and hence from experiencing the cytotoxic effects of tamoxifen. This is in agreement with results from [112], which showed that the surviving fraction of cells exposed to  $12\mu\text{M}$  of tamoxifen 24 hours after exposure was 0.47, similar to the value of 0.45 observed after 16 hours of exposure in [115]. As [26] concluded, it may be beneficial to stimulate breast cancer cells to progress past the  $G_b$  model phase using growth promoting agents and then expose them to tamoxifen so as to cause the largest amount of cell death.

The value of  $\beta_D$  for Case 2 is 0.20, and the resulting cell cycle phase proportions after tamoxifen exposure is shown in Figure 5.29(b). As we can see, Case 2 produces very similar cell cycle phase proportions to Case 1, however the surviving fraction is significantly higher at 16 hours after tamoxifen exposure than what is observed experimentally (60% in the model results compared with 45% in experiments), implying that Case 1 is a better representation of the experiment. However, there are no error bars provided for this single data point, and so this higher value could well be reasonable.

As we think that Case 2 is a more realistic representation of the measure of the cell cycle phase proportions after tamoxifen exposure, we wish to consider it further. We determine

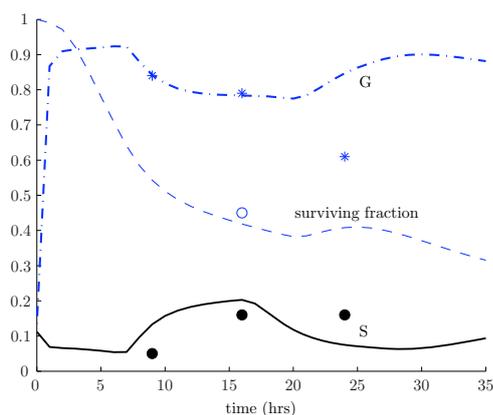


Figure 5.30: The proportion of cells in the G phase (dotted-dashed curve) and S phase (solid curve) after continuous tamoxifen exposure for Case 2 when we use the cyclin D and myc profiles from Figure 5.22. The surviving fraction is also shown in each figure as the dotted curve. The experimentally observed G and S phase proportions are shown as the asterisks and filled circles respectively and the surviving fraction 16 hours after tamoxifen exposure is shown as the hollow circle.

whether Case 2 will produce more realistic results if we assume that cyclin D and myc are not instantly reduced to zero after tamoxifen exposure. We instead use the cyclin D and myc profiles from Figure 5.22 (determined after exposure to 100nM of tamoxifen). We find that using the cyclin D and myc profiles and choosing  $\beta_D = 0.22$  produces the cell cycle phase changes shown in Figure 5.30. In this case, the cell cycle phase proportions are not significantly different from the attempts shown in Figure 5.29. However, for later times, the S phase proportion is slightly higher, better matching the observed S phase proportion at 24 hours than any other previous attempt, but still significantly different from that observed experimentally. The surviving fraction at 16 hours is now about 42%, which is similar to that observed experimentally, which is 45%.

We now wish to reconsider the experimental results after  $7.5\mu\text{M}$  of tamoxifen, which we attempted to model in Figure 5.26 by assuming that cyclin D and myc are instantly reduced to zero. In [112], results show that after 24 hours of  $7.5\mu\text{M}$  of tamoxifen, the surviving fraction is about 87%. We wish to see if we can model the effects of  $7.5\mu\text{M}$  of tamoxifen using the cyclin D and myc profiles from Figure 5.22. Although we have already determined that the concentrations of cyclin D and myc must be instantly reduced to zero to model the cell cycle phase changes after exposure to this concentration of tamoxifen, we wish to see if we can avoid making such an assumption by including the effects of cell

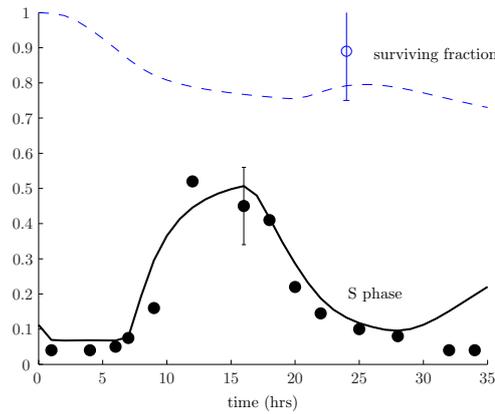


Figure 5.31: The proportion of cells in the S phase (solid curve) after continuous tamoxifen exposure for Case 2 when we use the cyclin D and myc profiles from Figure 5.22. The surviving fraction is also shown as the dotted curve. The experimentally observed S phase proportions are shown as the filled circles and the surviving fraction 24 hours after tamoxifen exposure is shown as the hollow circle, and is the mean of 8 observations. The error bar for the surviving fraction at this time is also shown.

death.

As we did earlier, we use the information we have on the S phase proportion after tamoxifen exposure to determine the value of  $\beta_D$ . However, we have more data on the S phase proportions after  $7.5\mu\text{M}$  of tamoxifen when compared to that after  $12.5\mu\text{M}$  of tamoxifen. Thus, we determine the value of  $\beta_D$  by finding the best fit to the S phase data shown in Figure 5.24. Doing this, we find that  $\beta_D = 0.076$ , and the resulting cell cycle phase proportions are shown in Figure 5.31. In this case, the model compares well to the data (black circles) except for the last 2 data points. The surviving proportion at 24 hours is lower than that observed experimentally, though it is still within the standard error determined in [112]. It is possible that the cyclin D and myc profiles from Figure 5.22 are not a realistic representation of how these proteins change after exposure to the high concentrations of tamoxifen considered in [115]. It is likely that cyclin D and myc are reduced to lower levels than after 100nM of exposure. This would mean that we would not need to choose such a high value for  $\beta_D$  to produce an accurate S phase proportion at 16 hours after selection, and we would consequently have a higher surviving fraction for times after tamoxifen exposure. In addition, less cells would enter the S phase after 24 hours, and our model would not produce such a high S phase proportion at these later times.

In this section, we introduced cell death into the model, and we were able to accurately capture cell cycle phase changes after high concentrations of tamoxifen without making unrealistic assumptions, such as cyclin D and myc being instantly reduced to 0. By including the effects of cell death in the  $G_c$  phase, we were able to use the cyclin D and myc profiles from Figure 5.22 to produce good results when compared to experimental data. Given the recent claims in [26] that tamoxifen kills cells in the S phase and not in G1-arrested cells, and the older results from [115] that suggest that tamoxifen kills cells in the late G phase, it would seem that further investigation is required to determine how tamoxifen induces cell death, and in what phases this effect is most potent. This model has potential to investigate the phases during which cell death does occur, however more data is required before we can make strong conclusions about how this drug behaves.

### **Summary of our results of modelling cell cycle progression in MCF-7 breast cancer cells.**

In this chapter, we used our parameterised cell cycle model to simulate cell cycle progression in MCF-7 cells after exposure to a range of substances, including estrogen, IGF, insulin, ICRF, ICI 164384, ICI 182780 and tamoxifen. In each of these instances, we were able to produce results that compared very well with experimental data, without the need to change the values of fixed cell-line-specific parameters. A unique part of the model is the ability to capture the period when cells are responsive to environmental signals - what we have called the  $G_b$  phase. Being able to identify when this responsive phase occurs allows for direct connection between the concentrations of internal cell cycle related proteins, such as cyclin D and myc, and cell cycle progression. In addition, the ability to quantify the number of cells in any of the model phases at all times means that cell cycle specific drugs can be administered at optimal times for maximal cytotoxic effects.

We also consider the common assumption that the presence of a ‘slowly cycling population’ of cells is necessary to interpret experimental data. In particular, we show that this population is not necessary to explain the results of mitotic selection experiments. We also show that this population is not necessary to explain experimental results when cells are exposed to the drug ICRF (which arrests cells in the M phase of the cell cycle) if we assume that this drug is 99.96% effective at preventing M phase completion instead

of 100% effective.

The model is applied to a range of different experiments, and at many instances is used to give insight into the underlying biological mechanisms at play. We also used the model to determine how the protein myc must change in concentration after exposure to various anti-estrogens. The model was later extended to include the effects of tamoxifen, which both inhibits cell cycle progression and causes cell death, and we were able to determine in what phases cell death occurs after tamoxifen exposure. Overall, we conclude that the model has great potential to be used in parallel with biological experiments.

We now wish to apply our model to the T47D breast cancer cell line, which is commonly used to investigate the effects of progesterone and progestin on cell cycle progression. We will need to re-parameterise our model in order to apply it to this cell line. In the next chapter, we discuss how we can do this, and we show how our model compares to some other experimental results from the literature that involve T47D cells, concluding that one particular experiment from the literature appears to have results that are inconsistent with other experiments. We also use the model to discuss the different behaviours of the popular MCF-7 and T47D cell lines.

## Chapter 6

# The T47D breast cancer cell line

The T47D breast cancer cell line is another widely used cell line for investigations into breast cancer growth. One of the benefits of using the T47D cell line is that the cells are responsive to progesterone and progestin (although recall that they are not sensitive to estrogen), whereas the MCF-7 cells are not responsive to either progesterone or progestin [80]. It is important to understand how progesterone, and particularly progestin, influences cell cycle progression, as progestin has been implicated as a possible contributing factor to breast cancer [17].

We will use the model from equations (3.7) to model cell cycle progression in T47D breast cancer cells. However, we cannot assume that the values of the experiment-independent variables are the same in the T47D cell line as the MCF-7 cell line. As we did with the MCF-7 cell line, we use published experimental results to estimate the values of the experiment-independent variables.

Recall from Chapter 4 that the first step in determining the experiment-independent variables for the MCF-7 cell line involved estimating the values of the storage phase durations. We did this using results from mitotic selection experiments from [78]. Once we had the storage phase durations, we used results from a single set of experiments (which we called the *base experiment*) on the steady-state values of  $G$ ,  $S$ ,  $M$  and  $\rho$  (which we called the *base variables*, because we obtained them from the base experiment), which allowed us to determine the values of  $\alpha$  and  $\beta$ .

Unfortunately, there is not as much data available in the literature for the T47D cell

line, and so we are required to use a different approach. In this chapter, we discuss how we determine the experiment-independent variables for this cell line by pooling results from a range of experiments.

## 6.1 A discussion of different experiments

For the MCF-7 breast cancer cell line, we used data from [112] to determine the experiment-independent variables because the observed data had a very small standard error. There exists a similar experiment for the T47D cell line [111], performed in the same laboratory and around the same point in time. Experimental results from [111] were made from repeated observations, making it desirable as the base experiment for the T47D cell line. In addition, the reported values of  $G$ ,  $S$ ,  $M$ , and the doubling time in the text, can be confirmed by observing the published experimental results. It is important to be able to confirm the doubling time from their data, because the terms ‘average cell cycle duration’ and ‘population doubling time’ are often used interchangeably. However, as we saw in Section 4.3, they are not the same.

Results of [111] report that in the presence of growth factors, the steady-state cell cycle phase proportions were observed as  $G = 62.5\%$ ,  $S = 26\%$ ,  $M = 11.5\%$  and the doubling time was measured as 22 hours (the reported value compares well with measurements of changes in total cell number). Unfortunately, we find that these values are not consistent with other reported experimental results. In particular, [102] reports that under certain environmental conditions, T47D cells grew in phase steady-state with  $G = 63 \pm 0.4\%$ ,  $S = 27 \pm 0.4\%$  and  $M = 9.5 \pm 0.4\%$  respectively. They also reported a doubling time of about 34 hours, which is significantly different to the 22 hour doubling time reported in [111] even though the steady-state cell cycle phase distributions are quite similar. Although [102] was published well after [111], there was no mention of the discrepancies between the two experiments.

These differences in the reported doubling times - 22 hours and 34 hours - are too large to simply account for experimental error. To have such vastly different doubling times while having very similar cell cycle phase proportions, the rates of progression through all three phases must be different. The cell population which doubled in 22 hours must

spend a smaller amount of time in *all* of the G, S and M phases in order to generate a cell population with very similar phase proportions but a much shorter doubling time. However, it is well accepted that the S and M phase durations in the same cell line are experiment-independent, and should therefore not change value when experimental conditions change.

The properties of the T47D cell line have been shown to vary across laboratories, and it has been suggested that T47D cells are genetically unstable, giving rise to many variants across laboratories [101]. This feature of the T47D cell line could be the reason for the discrepancies described above. We wish to investigate the cell kinetics of T47D cells across a variety of experiments to decide whether there is much variation across many experiments, or whether most experiments have similar kinetics, but some are outliers. To do this, we consider the results from a variety of experiments, and determine whether there exists a set of experiment-independent parameters which can be used to describe the majority of experimental results.

## 6.2 A summary of experiments

In this section, we list experiments in the literature that we could find which report features of T47D cell growth under phase steady-state conditions, and then compare these experimental results later. For an experiment to be considered for comparison, the doubling time and at least one cell cycle phase proportion in phase steady-state needs to be reported (without this information, we cannot compare the experiments).

### Results from [137]

In [137] the effects of retinoids on T47D cell cycle progression were investigated. Before exposure to retinoids, cells grew in phase steady-state, and cell numbers over a 5 day period were determined. By investigating this data, we note that the population doubling time is about 33 hours. It is noted that the S phase proportion under these conditions is 25%, however the G and M phase proportions were not reported.

**Results from [19]**

In [19], the effects of the drug Genistein on T47D cell cycle progression was investigated. In control cells, it was noted that 4 days corresponded to two population doublings, and so we conclude that the doubling time is 2 days, or 48 hours. The G, S and M phase proportions were reported to be 70, 22 and 8% respectively in the absence of Genistein.

**Results from [102]**

In [102], glucose transport kinetics in differentiated (or growth-arrested) T47D cells are investigated. It was observed that before cells were growth arrested, the T47D cells had a doubling time of 34 hours, as reported in the text, and the G, S and M phase proportions were reported as 63, 27 and 9.5% respectively. After exposure to retinoic acid, which prevents progression past the R point, the doubling time was reported to be 50 hours and the G,S and M phase proportions were reported as 75, 19 and 6% respectively.

**Results from [111]**

The effects of the drug Medroxyprogesterone Acetate on T47D cells was investigated in [111]. The cells were maintained in constant exponential growth with a doubling time of 22 hours. The corresponding G, S and M phase proportions were reported to be 62.5, 26 and 11.5% respectively.

**Results from [80]**

The effects of the progestin ORG 5020 on cell cycle kinetics was investigated in the T47D cell line. Cells were reported to grow with a doubling time of 32-48 hours, and taking the midpoint we estimate cells grew with a doubling time of 40 hours. The corresponding phase proportions were observed to be  $G = 68$ ,  $S = 21$  and  $M = 11\%$ .

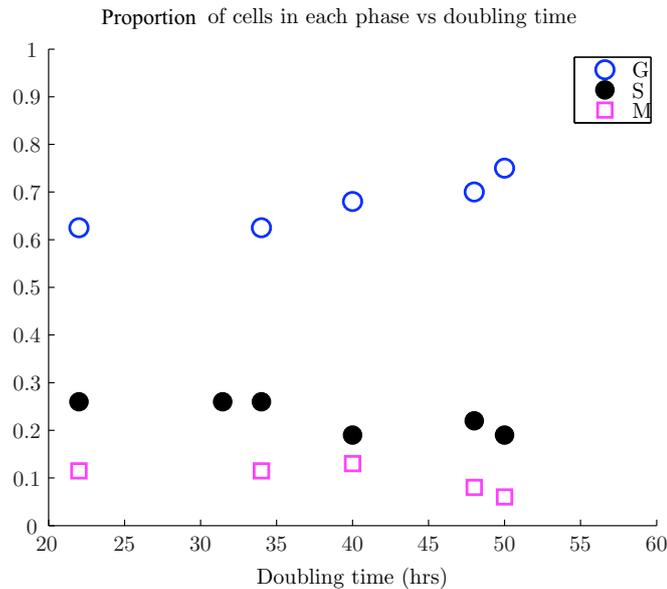


Figure 6.1: A summary of the experiments discussed in this section. The reported  $G$ ,  $S$  and  $M$  phase proportion for the corresponding doubling time is shown as the hollow blue circle, the solid black circle and magenta square, respectively. The experiment from [137] which only reports the  $S$  phase proportion is also shown.

### Summary of these experiments

We can summarise the experimental results of this section. In total we have phase steady-state information from six observations (recall that in [102] steady-state values under two different growth conditions were reported), with [137] only reporting the doubling time and corresponding  $S$  phase proportion ( $G$  and  $M$  phase proportions were not reported). Figure 6.1 summarises the reported data from these six observations by plotting the observed phase steady-state cell cycle proportions against the observed doubling time.

### Correlation between experimental results

We already know that two of the above mentioned experiments do not agree with each other. To decide whether the remaining experiments are generally in agreement, we use our model. In particular, we will see if we can generate a phase diagram similar to that produced earlier in Section 4.3 (see Figure 4.2, page 94). We wish to determine whether there is a phase diagram that is consistent with the data points produced in Figure 6.1.

In Section 4.3, we investigated the sensitivity of the phase diagram to varying values of our experiment-independent parameters. We found that the phase diagram in Figure 4.2 was highly sensitive to the choice of the base variables, but was insensitive to the choice of the experiment-independent variables (such as the storage phase durations and the values of the parameters  $\alpha$ ,  $\beta$  and  $\gamma$ ).

Thus to determine what kinds of phase diagrams will go through our data points, we only need to consider different values of the base variables  $G$ ,  $S$ ,  $M$  and  $\rho$ . We can determine the optimal choice of these base variables that produces the phase diagram that fits these data points as well as possible. Even though it doesn't matter what values of the storage phase durations we choose, we still need to know what they are, and we choose to use the storage phase durations we determine later, shown in Table 6.1. We also note that we have tried varying the values of the storage phase durations for the T47D cell line, and as with the MCF-7 cell line, varying these experiment-independent variables does not significantly change the phase diagram.

We use an optimisation process to determine the phase diagram that best goes through all the data points. To do this, we choose a doubling time and use optimisation to determine what the corresponding values of  $G$ ,  $S$  and  $M$  are (it doesn't matter what doubling time we choose. The only restriction is that the doubling time is larger than the minimum possible doubling time for this cell line). For a doubling time of 34 hours, we find  $G = 0.71$ ,  $S = 0.2$ ,  $M = 0.09$ . Figure 6.2(a) shows the phase diagram in this case. We can see that the fit is not very good because the experiment from [111] (the experiment with a doubling time of 22 hours) is an outlier. We wish to determine what kind of fit we can produce if we do not consider this experiment. For a doubling time of 34 hours, we find that the corresponding optimal values of  $G$ ,  $S$  and  $M$  are  $G = 0.6$ ,  $S = 0.28$ ,  $M = 0.12$ . Figure 6.2(b) shows the  $G$ ,  $S$  and  $M$  phase proportions for varying values of the doubling time using these values for the base variables. The phase proportions compare much better with all experiments except for the one from [111], implying that this experiment is the only outlier. Thus, we choose not to consider the experiment from [111] in our investigations of this cell line. Other than the explanation that T47D cells are genetically unstable [101], we have no other explanation as to why these results are so different from those in the other experiments.

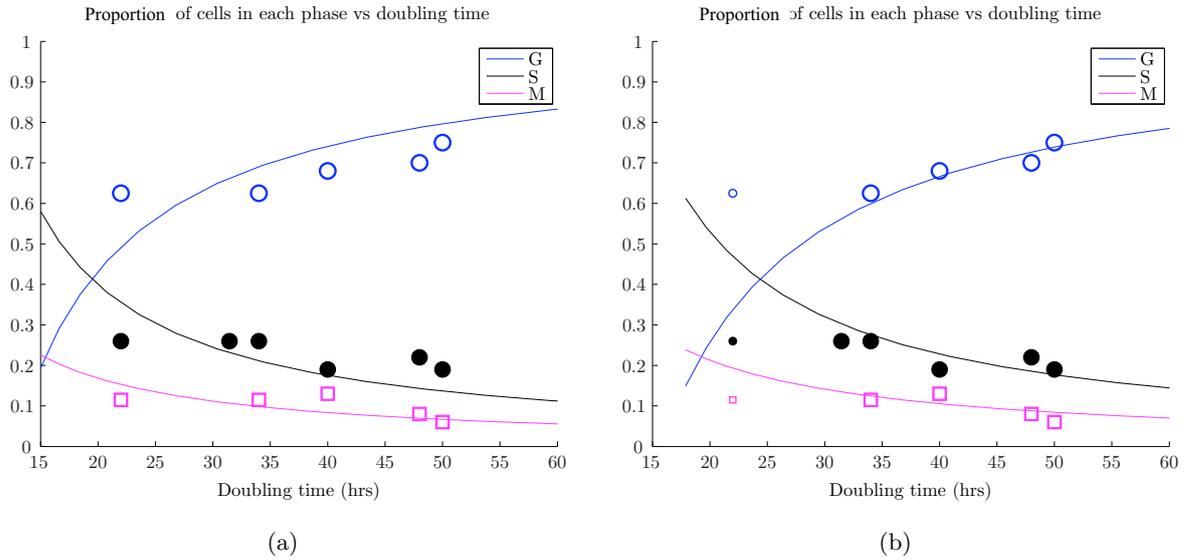


Figure 6.2: *Steady-state G (blue dotted-dashed curve), S (black solid curve) and M (pink dashed curve) phase proportions for varying values of the doubling time. Optimisation was used to determine the optimal base variables. In (b) all experiments except for the one performed in [111] were considered when finding the optimal values of the base variables, and the G, S and M phase proportions from this experiment are shown in (b) as the smaller sized hollow blue circle, solid black circle and magenta square, respectively.*

### 6.3 The storage phase durations for the T47D cell line

In this section, we discuss how to determine the storage phase durations for this cell line. We use the values of the base variables we have just determined. Recall that the base variables give information about how long a cell spends in each of the three cell cycle phases,  $G$ ,  $S$  and  $M$ . Variables such as the storage phase durations determine how much time a cell spends in the corresponding storage phase for that cell cycle phase. Thus the base variables can be used to determine an upper bound on the storage phase durations. We do this by considering equations (3.27b), (3.27e) and (3.27g) from Section 3.3 (see page 81). Substituting the base variables we chose in the previous section into these equations gives:

$$G_b = 2e^{-0.02\tau_{G_a}} - (2 - 0.63)e^{0.02\tau_{G_c}}, \quad (6.1)$$

$$S_b = (2 - 0.63)e^{-0.02\tau_S} - (1 + 0.095), \quad (6.2)$$

$$M_b = (1 + 0.095)e^{-0.02\tau_M} - 1. \quad (6.3)$$

As we must have  $G_b, S_b$  and  $M_b$  all  $\geq 0$ , we can rearrange to find upper bounds on the storage phase durations, which must also be non-negative:

$$\begin{aligned} 0 &\leq \tau_{G_a} + \tau_{G_c} \leq 21.3, \\ 0 &\leq \tau_S \leq 8.14, \\ 0 &\leq \tau_M \leq 5.50. \end{aligned} \tag{6.4}$$

If we simply take the midpoint of the ranges for the storage phase durations, we get  $\tau_{G_a} + \tau_{G_c} = 10.65$ ,  $\tau_S = 4.07$  and  $\tau_M = 2.75$  hours, however we note that this is unlikely to give great results because the ranges are quite large.

We find, however, that the  $\tau_S = 4$  is a reasonable choice, as we will now show. We consider the S phase changes in [136], shown later in Figure 6.10, page 196. We see that the S phase profile is very sharp - the S phase proportion increases quickly from about 9 hours, reaching a maximum at 15 hours and subsequently decreasing quickly. If  $\tau_S$  is large, then cells will be forced to remain in the S phase for long periods of time, and would not be able to enter the S phase and leave quickly. The sharpness of the S phase profile in Figure 6.10 strongly implies a short S phase duration. This sharpness of the S phase is also observed in [80]. We find that choosing  $\tau_S = 4$  hours, equivalent to the midpoint of the ranges discussed above, allows us to produce reasonable fits to the sharp S phase profiles we observe.

We find that  $\tau_M = 2.5$  is not a good choice, and in fact,  $\tau_M$  must be quite small. We can see this by considering the experimental results from [81]. In this experiment, T47D cells are injected with a plasmid that continuously expresses cyclin D protein over time. Cell cycle progression is then investigated after cyclin D stimulation, and the changes in cell cycle phase distributions are shown in Figure 6.3. We note that this experiment was not considered in our earlier set of experiments because the corresponding doubling time was not reported for the steady-state conditions, and so we could not have used it as one of our experiments used to generate the phase diagram in the previous section. The shallow, flat features of the M phase profile shown in Figure 6.3 indicates that  $\tau_M = 2.5$  is not a good choice, because a large value of  $\tau_M$  would result in a steep increase in the M phase proportion, followed by a sharp decline (see Section 4.3 for a discussion on how the storage phase durations influence cell cycle phase proportions). A shallow and flat

NOTE:  
This figure is included on page 187 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 6.3: *Cell cycle distribution of T47D cells 9 - 36 hours after cyclin D exposure, reproduced from [81]. The white circles represent control populations. The black circles represent changes in cell cycle phase proportions after cyclin D stimulation.*

M phase proportion can also be seen in [80] (not shown). For this reason, we choose a smaller value for  $\tau_M$  at 0.5 hours.

We can also use the experiment performed in [81] to discuss the G phase storage durations. We note that our calculated range of the possible G phase storage durations is large, and simply taking the midpoint of this wide range is unlikely to produce good results. To determine better ranges for the G phase storage durations, we will use our model to try and simulate the experimental results from [81]. We attempt to model these results by finding a  $\gamma$  function that can produce the observed cell cycle phase changes for different values of  $\tau_{G_a} + \tau_{G_c}$ . We start by taking  $\tau_{G_a} + \tau_{G_c} = 10.65$ , the midpoint of the possible range of storage phase durations, and show that we cannot simulate the changes in cell cycle progression observed in Figure 6.3. We note we have not explicitly specified the values of  $\tau_{G_a}$  and  $\tau_{G_c}$ . However, distributing 10.65 between the variables  $\tau_{G_a}$  or  $\tau_{G_c}$  merely ‘offsets’ when cells enter the S phase after stimulation. For example, if we choose  $\tau_{G_a} = 10$ , then  $\tau_{G_c} = 0.65$ , then within an hour of stimulation, cells will begin to enter the S phase. However, if we were to choose  $\tau_{G_a} = 0.65$  and  $\tau_{G_c} = 10$ , then there would be a 10 hour delay before cells that are stimulated from the  $G_b$  phase to enter the S phase

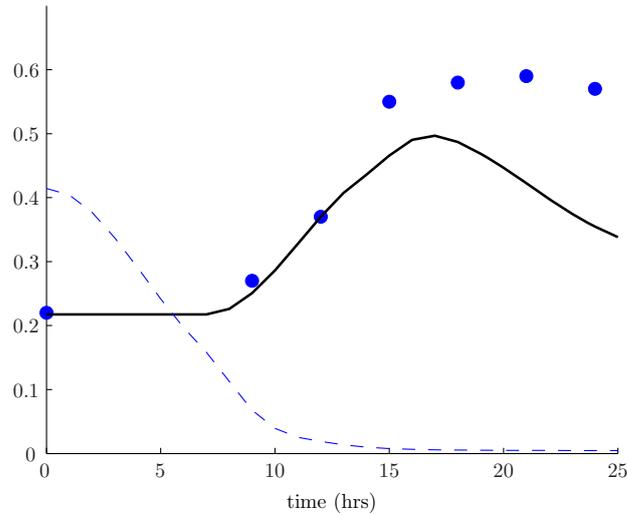


Figure 6.4: Simulated S phase proportions over time is shown as the solid line. The  $G_b$  phase proportions are shown as the dashed line. The dots represent the experimentally determined S phase proportion after cyclin D stimulation, from Figure 6.3.

(due to their long duration in the  $G_c$  phase). Thus, we choose a value for  $\tau_{G_c}$  so that cells enter the S phase at the appropriate time.

Results of our attempt to simulate the cell cycle phase changes with  $\tau_{G_a} + \tau_{G_c} = 10.65$  are shown in Figure 6.4. The figure shows that the modelled S phase proportion (solid black curve) does not compare well with the experimental data for times later than 15 hours. The modelled  $G_b$  phase proportion, also shown in this figure as the blue dashed curve, quickly reduces to zero by about 10 hours, and remains at about zero for the remainder of the experiment. This means that even if we increased the value of  $\gamma$  after this time, we would not increase the S phase proportion any further. This means that there is no  $\gamma$  curve which can replicate the S phase changes from [81] if we assume that  $\tau_{G_a} + \tau_{G_c} = 10.65$ . Choosing a larger value of  $\tau_S$  would help the S phase profile here, however as discussed earlier, we are limited as to how large we can choose  $\tau_S$  because of the sharpness observed in some S phase experimental data.

Another reason for why we cannot produce a response in the S phase similar to that shown in Figure 6.3 is that there are not enough  $G_b$  phase cells to react to any environmental change (shown as the blue dashed line Figure 6.3). If we decrease the value of  $\tau_{G_a} + \tau_{G_c}$  then we increase the starting number of cells in the  $G_b$  phase. This will provide us with more cells that we can shuffle through to the S phase, and hence may

enable us to more accurately replicate results.

Figure 6.5 shows our attempts to replicate the S phase changes for smaller values of  $\tau_{G_a} + \tau_{G_c}$ . Figure 6.5 shows that we must choose small values for  $\tau_{G_a} + \tau_{G_c}$  to be able

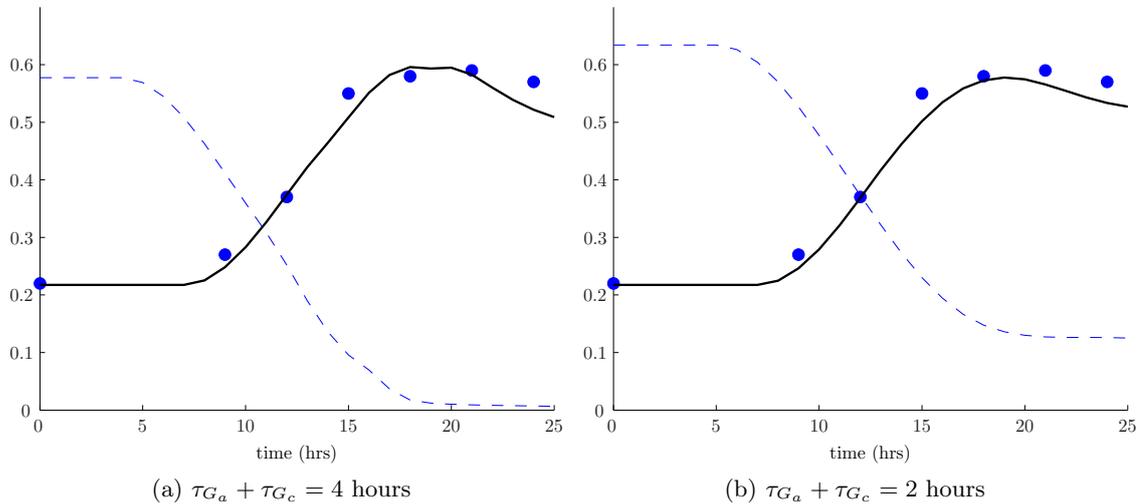


Figure 6.5: Simulated S phase proportions over time is shown as the solid line. The value of  $\tau_{G_a} + \tau_{G_c}$  was assumed to be (a) 4 hours and (b) 2 hours. The  $G_b$  phase proportions are shown as the dashed line. The dots represent experimentally determined S phase proportion after cyclin D stimulation, from [81].

to replicate the changes in the S phase proportion over time. In Figure 6.5(a), we chose  $\tau_{G_a} + \tau_{G_c} = 4$  hours (compare with Figure 6.4, where we chose  $\tau_{G_a} + \tau_{G_c} = 10.65$ ). We were able to find a  $\gamma$  function which produces a reasonable S phase curve. However, we needed to choose very high values of  $\gamma$  just to get the S phase up to this level for this amount of time - this is evidenced by the fact that the  $G_b$  phase is essentially zero for later times. This figure shows that  $\tau_{G_a} + \tau_{G_c} = 4$  is really an upper bound on these storage phase duration values. Thus we have a refined range for the G phase storage durations ( $0 \leq \tau_{G_a} + \tau_{G_c} \leq 4$  hours), so without any additional information, we take the midpoint and set  $\tau_{G_a} + \tau_{G_c} = 2$  hours. Figure 6.5(b) shows an example of a  $\gamma$  curve that simulates the S phase changes when  $\tau_{G_a} + \tau_{G_c} = 2$  hours. In this case, the chosen  $\gamma$  produces a good S phase curve, as compared to the experimental data. As we can see, the  $G_b$  phase is not very close to zero at any time, and so in this case we did not have to resort to choosing high values of  $\gamma$  to find a good S phase curve, meaning we have more flexibility in our choice of  $\gamma$  and this will make it easier when we later try to find a relationship between known cyclin D changes and  $\gamma$  changes. Table 6.1 shows the chosen values of the storage

Table 6.1: *The experiment-independent variables for the T47D cell line*

| Parameter                 | value                  |
|---------------------------|------------------------|
| $\tau_{G_a} + \tau_{G_c}$ | 2 hrs                  |
| $\tau_S$                  | 4 hrs                  |
| $\tau_M$                  | 0.5 hrs                |
| $\alpha$                  | 0.154 hr <sup>-1</sup> |
| $\beta$                   | 0.174 hr <sup>-1</sup> |

phase durations for the T47D cell line, as well as the corresponding values of  $\alpha$  and  $\beta$ , calculated using the base variables determined in the previous section. The chosen values of the storage phase durations are significantly lower overall when compared to the storage phase duration values determined for the MCF-7 breast cancer cell lines. As we saw in Section 4.3, smaller storage phase durations mean that steady-state is reached faster upon a change in environmental conditions. We also found in Section 4.3 that having shorter storage phase durations (with correspondingly longer non-storage phase durations) means that the doubling time and average cell cycle duration are no longer similar.

## 6.4 The $\gamma$ function for the T47D cell line

As we did in Section 5.3, we now develop a function which converts the changes in internal protein concentrations into a value of  $\gamma$ . In Section 5.3, we found that the proteins myc and cyclin D were the most influential players in causing cell cycle progression in MCF-7 cells. For the T47D cell line, there is very little data concerning internal myc concentration changes after exposure to different substances, and most experiments only monitor changes in cyclin D concentration. In [79] it was noted that cyclin D appears to be the main regulator of cell cycle progression in T47D cells after progestin exposure, and that myc plays less of a role when compared to cyclin D. Thus, we will create a  $\gamma(t)$  function which depends only upon cyclin D. In particular, we choose the form

$$\frac{d\gamma(t)}{dt} = f_0 + \omega^T d(t) + \omega_\gamma^T d(t)\gamma(t) - \beta_\gamma^T \gamma(t). \quad (6.5)$$

The term  $\omega^T$  represents the influence of cyclin D on the production rate of  $\gamma$ . The term  $\omega_\gamma^T$  represents the influence of the interaction between cyclin D and  $\gamma$  on the production rate

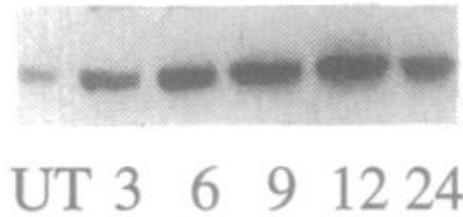


Figure 6.6: Immunoblot of cyclin D concentration over time after zinc stimulation from [81]. UT represents the concentration of cyclin D in untreated cells. The numbers represent hours after zinc stimulation.

of  $\gamma$  and the term  $\beta_\gamma^T$  is the rate at which  $\gamma$  decays. The terms  $\omega^T$ ,  $\omega_\gamma^T$  and  $\beta_\gamma^T$  are similar to the terms  $\omega$ ,  $\omega_\gamma$  and  $\beta_\gamma$  used in equation (5.2) from Section 5.3. In equation 6.5,  $f_0$  represents the influence of other internal proteins (for instance myc), which we assume do not change significantly after exposure to the various substances we will be investigating, and therefore are represented here as a constant ‘background’ effect on  $\gamma(t)$ .

### Generating the cyclin D profile from data in [81]

As we did for the MCF-7 breast cancer cell line, we will use experimental results regarding internal protein changes to find the best fit for the growth parameters in the  $\gamma$  function presented here in equation (6.5). The experiment from [81], which we have already used to help determine the value of  $\tau_{G_a}$  and  $\tau_{G_c}$ , provides changes in cyclin D concentration over time. Unfortunately, [81] has only provided immunoblots representing the concentration of cyclin D over time, and it is difficult to read off quantitative values of cyclin D from this representation. However the apparent intensity of the blots can give an idea of what type of functional form we should choose for cyclin D concentration over time. Figure 6.6 shows the immunoblot of cyclin D protein concentration after zinc stimulation from [81]. This representation shows that cyclin D protein concentration increases by 3 hours after stimulation and reaches an observed maximum concentration at about 12 hours after stimulation and by 24 hours, cyclin D concentration is still higher than untreated control cells, although the concentration has noticeably decreased from what it was at 12 hours after cyclin D stimulation. It was noted that the highest observed concentration of cyclin D protein was about 5 fold higher than controls, which we assume was at about 12 hours after exposure, given that the highest observed concentration was at that time.

Let's reconsider the S phase profile shown in Figure 6.3. The S phase proportion remains high for at least 21 hours, and by 24 hours, the S phase proportion has begun to drop. Recall that if the S phase begins to decline after reaching a maximum, then this could be due to a reduction in the rate at which cells enter the S phase at this time, or an increase in the rate of leaving the S phase. As  $\gamma$  affects the rate of entering the S phase after a delay of  $\tau_{G_e} = 2$  hours, and influences the rate at which cells leave the S phase after a delay of  $\tau_S = 4$  hours, we would expect  $\gamma$  to have reached its maximum somewhere between  $21-4 = 17$  and  $21-2 = 19$  hours.

If we assume that cyclin D influences the value of  $\gamma$  quickly, then if cyclin D reaches a maximum concentration at  $t = 12$  hours, we will only maintain a high S phase proportion up to  $(12+2 = 14) - (12+4 = 16)$  hours, and not up to the observed 21 hours. To keep the S phase proportion high until  $t = 21$  hours, we need cyclin D to be high until sometime between 17-19 hours and for simplicity, we take the midpoint and assume that cyclin D is high until 18 hours. Thus, we do not assume that cyclin D reaches a maximum at  $t = 12$  hours, but rather at about 18 hours after exposure. We assume that if the maximum at 18 hours was significantly higher than that at 12 hours, then it would have been more likely to have been noticed during the experiment. Thus, we choose the maximum at 18 hours to be about 6 fold higher than control (compared to five-fold higher at 12 hours).

After 18 hours, we assume that cyclin D begins to decrease, however we do not know what level cyclin D decreases to by 24 hours after exposure. Let's consider the cell cycle phase changes for times  $t > 30$  hours. We observe that for these times, the phase proportions are unchanging, implying that phase steady-state has been reached. Thus, we can calculate the final steady-state value of  $\gamma$ , as we know the phase proportions for these later times. The value of  $\gamma$  can be substituted into equation (6.5) to give a final concentration of cyclin D in terms of the variables we are trying to fit from equation (6.5). Once  $\gamma$  reaches this constant value, it will take some time before the cell cycle phases settle into their constant phase proportions. We choose  $\gamma$  to reach its final unchanging value at about 24 hours after exposure so that by 30 hours, the cell cycle phase proportions have settled into their steady-state values (recall that steady-state is reached quite quickly in this cell line because the storage phase durations are small). We assume that after 18 hours, cyclin D concentration decreases to reach this final value (call it  $D_f$ ) by 24 hours, and thereafter, cyclin D concentrations remain unchanged. We will see later that our

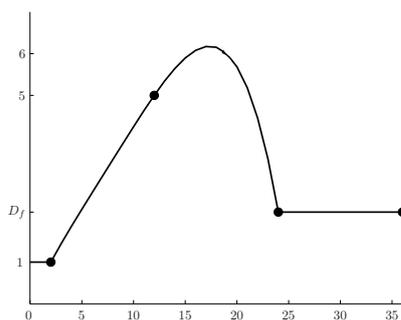


Figure 6.7: The dots represent the concentrations of cyclin D known at the respective points in time, as discussed in the text. The curve is a hypothetical curve which may be fitted to the cyclin D data, and which depends upon the value of  $D_f$ .

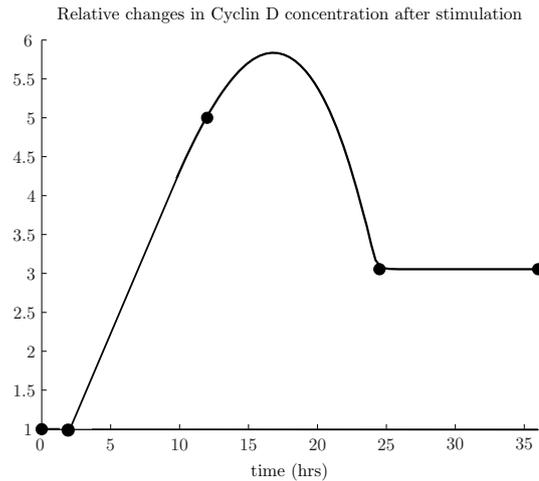
fitted concentration of cyclin D at 24 hours is higher than that at 6 hours, but less than the 9 hour concentration, which is in agreement with the immunoblots shown in Figure 6.6

We must still determine how fast cyclin D concentrations change immediately after stimulation. In [82], which also increased expression of cyclin D by plasmid stimulation, it was observed that cyclin D concentrations did not change noticeably until three hours after stimulation. We assume that in this case, cyclin D concentrations have begun to increase earlier so as to allow for it to become a noticeable change by three hours into the experiment. We assume that in [81] cyclin D begins to increase by two hours after stimulation. Cyclin D concentration increases from  $t = 2$  hours, continues to increase reaching about five-fold above controls at 12 hours (as observed by [81]), reaches a maximum concentration of six-fold above control by 18 hours and thereafter decreases, reaching  $D_f$  by 24 hours, after which time it does not change concentration. This information allows us to generate a cyclin D profile. We assume that for times  $t < 2$  hours, cyclin D is unchanging at control levels.

Figure 6.7 summarises the features of cyclin D changes after stimulation (this is NOT our cyclin D profile, as we have yet to determine the values of our unknown parameter  $D_f$ ). We can now use this continuous representation to solve for the unknown parameters in equation (6.5), by finding the value of these parameters which produce the best fit to the S phase given the cyclin D profile shown in Figure 6.6. Performing a least squares best fit gives the parameters from equation (6.5) shown in Table 6.2. The corresponding cyclin D curve is shown in Figure 6.8. We note that the final cyclin D concentration is about 3

Table 6.2: *The parameters of the  $\gamma$  function for the T47D cell line.*

| Parameter       | value |
|-----------------|-------|
| $f_0$           | 0.01  |
| $\omega$        | 0.91  |
| $\omega_\gamma$ | 5.30  |
| $\beta_\gamma$  | 1.96  |

Figure 6.8: *Cyclin D profile determined using results from [81] and the form of the cyclin D profile from Figure 6.7.*

fold higher than the starting concentration. This is a reasonable final concentration, as the immunoblots shown in Figure 6.6 indicate that the cyclin D concentration at time  $t = 24$  is higher than the starting concentration, but lower than the concentration at  $t = 12$  hours. The simulated S phase changes after cyclin D stimulation are shown in Figure 6.9(a). The S phase curve shown corresponds well with the experimentally determined S phase changes, which are shown as asterisks. The corresponding changes in the G and M phase proportions are shown in Figure 6.9(b). The simulated G and M phase proportions compare reasonably well with the observed changes. The M phase proportion appears to overshoot the observed values between 21 and 30 hours, however this is only by a small amount if we consider the absolute values of  $M$ . As mentioned earlier, it is important we keep the value of  $\tau_S$  low because increasing it further makes the M phase profile even sharper.

The values of these parameters from equation (6.5) are significantly different from the

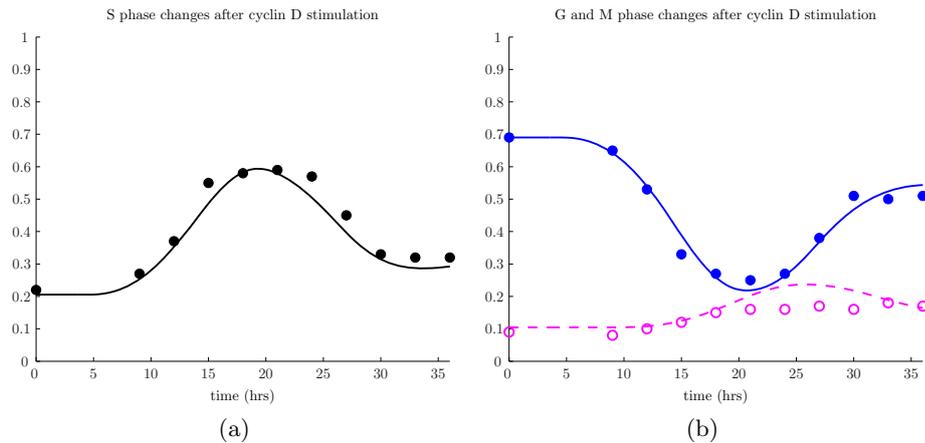


Figure 6.9: The cyclin D profile from Figure 6.8 was used to generate cell cycle phase proportions after T47D cells were stimulated to express cyclin D. The simulated S phase proportions are shown in (a) as the solid black curve and the simulated G and M phase proportions are shown in (b) as the solid blue and dashed magenta curve respectively. The corresponding dots represent the data from [81] in each figure.

ones in the MCF-7 cell line. For instance, the value of  $\beta_\gamma$ , which represents the rate at which the  $\gamma$  function decays, is much larger in this cell line. We also note that the value of  $f_0$  is very small. This is likely to be due to the fact that other proteins are not at all influential in this experiment, and that changes in cyclin D concentration are the main drivers of cell behaviour in this experiment.

## 6.5 Applying the model to another cyclin D stimulation experiment

We wish to verify the results of the previous section by investigating whether our chosen rate parameters for equation (6.5) will produce reasonable results when applied to other experiments. To test this, we consider the results from an experiment performed in [136]. In [136], T47D cells were exposed to the antiestrogen ICI 164384. After phase steady-state was reached, cells were stimulated to express cyclin D and subsequent changes in cell cycle phase distributions were investigated. The resultant S phase changes are shown in Figure 6.10. This experiment and the one discussed in Section 6.4 both encouraged cell cycle progression by stimulating T47D cells to express cyclin D. The main difference between

NOTE:  
This figure is included on page 196 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 6.10: *S phase changes in antiestrogen arrested T47D cells after cyclin D stimulation. The squares are the means of four to six experiments and the bars represent the standard error. The curve through the data points is the result of an unknown curve fitting process. Results from [136].*

the two experiments is that [136] pre-exposed cells to the antiestrogen ICI 164384, whereas in [79], cells were serum starved and then stimulated to express cyclin D. The presence of antiestrogen in [136] is likely to be the reason for the lower starting S phase proportion of about 13% when compared to the starting S phase proportion in [79] which is about 22%. We also note that the S phase changes in [136] are very different to the S phase changes observed in [81]. In [136], the S phase increases at a similar point in time as in [81] - about 9 hours after stimulation. However in [136], the S phase proportion reaches a maximum by about 16 hours and thereafter declines to reach control levels by 24 hours after exposure. Thus the S phase in [136] has a much sharper profile, decreasing rapidly after it reaches its maximum.

We wish to model the S phase changes observed after cyclin D stimulation. We are only told that the cyclin D protein concentration increased about 2-4 fold 6-9 hours after cyclin D stimulation. This is not sufficient information to generate a cyclin D profile. Instead, we will use the S phase profile to solve the inverse problem for cyclin D. We would not be able to solve the inverse problem for just cyclin D in the MCF-7 cell line, as the  $\gamma$  function was a function of both cyclin D and myc, and we would need to have information about myc as well as the S phase changes in order to infer cyclin D concentration changes over time. For the T47D cell line, we can use S phase changes to directly solve for cyclin D as we are not considering the effects of myc in this cell line.

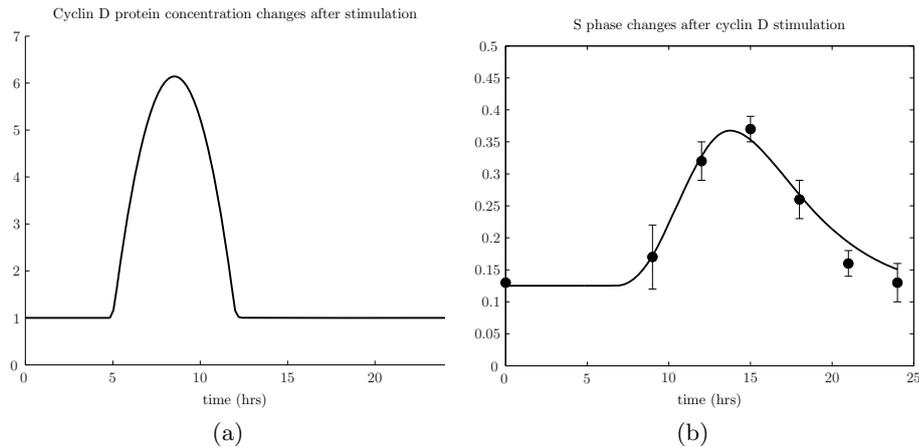


Figure 6.11: (a) The cyclin D protein changes determined by solving the inverse problem for cyclin D. (b) The S phase changes using the cyclin D profile shown in (a). Dots represent experimental data from Figure 6.10.

Previously when solving the inverse problem for *myc*, we first solved the inverse problem for  $\gamma$  and then for *myc*. However in this section, we will directly solve for cyclin D. We assume that cyclin D concentrations remain unchanged for  $t_1$  hours. At  $t = t_1$  we assume that cyclin D increases, reaching a maximum and then decreasing again to reach control levels at time  $t_2$  hours, after which the concentration of cyclin D remains unchanged. The S phase profile shown in Figure 6.10 is reasonably symmetric about the observed maximum S phase percentage at about 15 hours after cyclin D stimulation. Thus, we also assume that the cyclin D profile is symmetric about its maximum concentration. Thus, we assume the cyclin D profile takes the form of a quadratic in between times  $t_1$  and  $t_2$ . We write

$$d(t) = \begin{cases} d_0 & \text{for } t \leq t_1, \\ a(t - t_1)(t - t_2) + d_0 & \text{for } t_1 < t \leq t_2, \\ d_0 & \text{for } t > t_2, \end{cases} \quad (6.6)$$

where  $d_0$  is the starting cyclin D concentration, and  $a$  is some unknown parameter. We use the S phase data to find the best fit to the values of  $a$ ,  $t_1$  and  $t_2$ . Doing this gives  $a = -0.28$ ,  $t_1 = 5.3$  and  $t_2 = 13$  to two significant figures, and the resultant cyclin D profile is shown in Figure 6.11(a). Cyclin D protein concentration remains at controls until 5.3 hours, when it starts to increase, reaching a maximum concentration of about 5 fold higher than controls at about 9 hours after exposure, and decreasing to reach control

levels again by 13 hours after exposure.

We note that our profile for cyclin D increases from about 3.5 to 6 fold above controls 6-9 hours after cyclin D stimulation. This compares well with the observations made in [136] that cyclin D concentration increased 2-4 fold at 6-9 hours after cyclin D stimulation. We suggest that a reason that our model required a higher increase in cyclin D concentration in order to fit the S phase curve is due to the fact that the cells were pre-exposed to antiestrogen. As we saw earlier when we investigated the MCF-7 cell line, *myc* is strongly reduced after antiestrogen exposure. Perhaps the reductions in *myc* concentration are the reason for the reduced starting S phase proportion in [136] when compared to [81]. If *myc* is a cause of the reduced S phase proportion in control cells, then the corresponding cyclin D concentration at time 0 in our model would not be so low. If cyclin D did not start at such low levels as we have assumed in this experiment, then we would not require such high increases in the relative concentration of cyclin D 6-9 hours after stimulation, and we may have been able to produce a good S phase profile by increasing cyclin D only 2-4 fold above controls, as observed in [136]. However, as mentioned earlier, there is simply not enough data on *myc* changes to include their effects in modelling cell cycle phase changes in this cell line. Nonetheless, this analysis suggests that further investigations into the effect of *myc* on T47D cell cycle progression would be worthwhile.

## Summary

In this chapter, we have re-parameterised our cell cycle model to the T47D cell line. We used this model to compare a variety of experimental results, concluding that one of these experiments is probably an outlier. As we mentioned in Chapter 2, the T47D cell line is used in experiments because it is responsive to progesterone, and doesn't require the cells to be pre-exposed to estrogen. We wish to investigate the effects of artificial progestinones (known collectively as progestins) on cell cycle progression using this cell line. We choose to investigate progestins, and not the natural hormone progesterone, because progestins are used in hormone replacement therapies which are known to increase the risk of breast cancer, and also because progesterone is rarely used in experimental conditions due to the short half-life of the hormone. We discuss the effects of progestins on cell proliferation in the next chapter, and we also use the cell cycle model developed in this chapter to

investigate the effects of progestins on cell cycle progression.



## Chapter 7

# Progestin effects on cell cycle progression

We saw in Chapter 2 that the presence of progestins (which are artificial variants of progesterone) in combined hormone replacement therapies (cHRT's) are thought to be the cause of the increased cancer incidence amongst women who take cHRT's during menopause. Since this was discovered in the Million Women Study in 2002, many other studies have been performed to try and understand how progestins influences breast cell proliferation. However, there are still some disagreements about the role of progestins in breast cell proliferation [24, 61]. The role of progesterone and progestins in breast cell proliferation is unclear because various studies have shown different results. For instance, it is known that breast cell proliferation varies during the menstrual cycle, and reaches its highest rate when progesterone is at the highest concentration in the breast (and NOT when estrogen is at its highest concentration) [109]. However, cell experiments indicate that progesterone is anti-proliferative [24, 1] in both normal and breast cancer cells.

Another complication comes from the fact that for cells to be responsive to progesterone, they need to be first exposed to estrogen. This is because estrogen causes cells to express the progesterone receptor (PR), without which cells would not recognise the presence of progesterone (this is presumably also true for progestins). Thus to investigate the effects of progestins on breast cells, one must first expose the cells to estrogen, a strong mitogen (as we saw previously), and this makes it very difficult to investigate the

role of progestins in breast cancer cell proliferation.

To circumvent this issue, artificial breast cancer cell lines have been developed. One such cell line, the T47D line, is able to express PR without estrogen priming, making it very useful for research into the effects of progestins on breast cancer cell proliferation. In the years 1990-2000, several studies were performed investigating the role of progestins on T47D breast cancer cell proliferation. The results showed that progestin has more complex effects on cell cycle progression than those of estrogen. Initially, progestin up-regulates growth promoting proteins, such as cyclin D and myc [80, 37, 84], resulting in an increase in the S phase proportion. However, at later hours, inhibitory proteins such as p27 and p18 are up-regulated, exerting inhibitory effects on the activity of important cell cycle proteins such as cyclin D and cyclin E [37, 84, 113]. This results in a decrease in the S phase proportion to levels below that of control, and it has been observed that the S phase proportion remains below control levels for at least 40 hours after initial progestin exposure [80, 37, 84].

Thus, progestin is said to exert a biphasic response within a cell, initially stimulating cells to enter the S phase, and later discouraging S phase entry. Interestingly, Groshong *et al.* [37] showed that if the cells are exposed to a second dose of progestin again during the later period of growth arrest, the initial increase in the proliferation rate (which was observed after the first dose of progestin) does not occur after the second dose, and the S phase proportion remains below control levels.

Prior to the Million Women Study in 2002, some investigations were performed to get a better understanding of the effects of progestins on breast cancer cell lines [37, 84, 113]. However, the results of the million women study resulted in a shift in focus, with much research focussing on the role of progestins in early cancer development (as opposed to their effects on fully developed cancer cells, like the T47D cell line). In this chapter, we re-visit these older studies on the effects of progestins on T47D cells to get a better understanding of the underlying mechanisms behind the effects of progestins on breast cell proliferation.

## 7.1 Quantitative observations of protein changes in T47D breast cancer cells after exposure to progestin

To understand which proteins are crucial in progestin-induced cell cycle changes, it is important to investigate the changes in protein concentrations and changes in the activity of protein complexes after progestin. Recall that monitoring the *activity* of a protein is crucial, because even though a certain protein may be high in concentration, if the protein exists mainly in its inactive form, then its effects in the cell will be small. Such an experiment was performed in [84]. T47D cells in phase steady-state were exposed to progestin at time zero and the changes in the S phase proportion, cyclin concentration, cyclin-CDK activity and p21/p27 inhibitor concentration were all investigated. We will discuss the results from this investigation in detail, as these results give us good insight into the causes of progestin-induced cell cycle progression.

### Changes in the S phase proportion and protein concentrations after progestin as observed in [84]

The T47D cells used in the experiment were pre-exposed to ample growth factors, resulting in a high steady-state S phase proportion of about 27%, as indicated in Figure 7.1 (white circles). The cells were exposed to these growth factors because authors wanted to focus on the growth-arrest phase of progestin (high growth conditions mean that the proliferation rate will not increase much after further stimulation, and the effects of anti-growth signals will be more obvious).

At time  $t = 0$ , non-control cells were exposed to progestin (see Figure 7.1). After progestin exposure, the S phase proportion increased slightly so that by 6 hours after exposure, the S phase proportion was just above that of control cells at 31% and reached an observed maximum at 12 hours of about 34%. Thereafter, the S phase proportion decreased. By 18 hours after exposure, the S phase proportion was just below that of control cells at about 23% and continued to decline reaching a minimum of about 10% by 24 hours. The S phase proportion remained this low for the remainder of the experiment, i.e. until 30 hours after exposure. As we can see, the S phase proportion does not increase by very much after exposure to progestin, but the S phase proportion decreases

NOTE:  
This figure is included on page 204 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 7.1: *Changes in the S phase proportion after progestin exposure were determined at regular 6 hourly intervals and shown here as black circles. The white circles represent the S phase proportion without exposure to progestin (i.e. control cells). Curves through the data points were produced using an unknown curve fitting program. Results obtained from [84].*

significantly after progestin, implying that progestin is mainly inhibitory. However, as mentioned previously, this is simply because cells started in high growth conditions, and it is therefore harder to increase any further the rate of progression past the  $G_b$  phase. In [80], the effects of progestins was investigated in T47D cells growing in a more starved environment. Cells were then exposed to ORG 2058, and the subsequent changes in cell cycle phase proportions were investigated. Their results are re-printed in Figure 7.2. We can see that in this experiment, the initial stimulatory effect of progestins is more obvious, and this experiment shows the clear biphasic nature of progestins. We will not discuss results of this particular experiment any further, because no other information was provided on changes in internal cell cycle proteins, and so we go back to discussions from [84], keeping in mind that progestin can cause a strong proliferative effect initially followed by a strong anti-proliferative effect.

Changes in concentration of cyclin D1, cyclin D3 and cyclin E1 were investigated in [84] in addition to the changes in the S phase proportion. The changes in *total* cyclin D1, cyclin D3 and cyclin E1 concentrations are shown in Figure 7.3. (In this figure, it is noted that cyclin E was measured, and not cyclin E1. However, Musgrove et al. [84] have confirmed that they only measured cyclin E1. They did not measure cyclin E2 because they believe that this cyclin is not significant in T47D cells. However, more recent evidence has shown that cyclin E2 is an important regulator of cell cycle progression in T47D cells [16]. Thus, we wish to be explicit about the differences between the two proteins. The

NOTE:  
This figure is included on page 205 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 7.2: *The S phase changes in T47D cells after exposure to ORG 2058 are shown as the black squares. The hollow squares are the cell cycle phase percentages in control cells which have not been exposed to ORG 2058. The points represent the means of six independent experiments and the bars represent the standard error. Results from [80].*

role of cyclin E2 will be discussed later in Section 7.4.) This representation shows that within the first 6 hours of progestin exposure, cyclin D1 concentration increases to about 1.5 times control levels. Between 6 and 18 hours, cyclin D1 concentration decreases to about 70% of control levels, and by 24 hours, cyclin D1 remained low at about 60% of control levels. Cyclin D3 levels are regulated similarly, with a slight increase within the first 12 hours of exposure, eventually decreasing to about 50% of controls by 24 hours after exposure.

Cyclin E1 concentrations do not change significantly within the first 6 hours of progestin treatment. Between 6 and 24 hours after progestin treatment, cyclin E1 levels decrease at an almost constant rate, reaching about 30% of control levels by 24 hours after progestin treatment. Cyclin E1 experiences a slight increase in concentration between 24 and 30 hours after treatment, even though there is no sign of any increase in the S phase proportion after this time (see Figure 7.1). Other experiments have shown that cyclin E1 concentration begins to increase at about this time [37] and continues to increase eventually exceeding control levels. These experiments also show that the S phase proportion remains low, even in the face of these high cyclin E1 levels. We will discuss the reasons for the increase in cyclin E1 in Section 8.4.

The changes in concentration of the inhibitors p21 and p27 were investigated in [84] after cells were exposed to progestin. These changes are shown in Figure 7.4. Their results show that within the first 12 hours, there was no significant change in the concentration

## NOTE:

This figure is included on page 206 of the print copy of the thesis held in the University of Adelaide Library.

Figure 7.3: *Relative changes in cyclin D1 (black squares), cyclin E1 (white circles, called cyclin E) and cyclin D3 (white squares) concentration after progestin exposure. Curves through the data points were produced using an unknown curve fitting program. Results obtained from [84].*

## NOTE:

This figure is included on page 206 of the print copy of the thesis held in the University of Adelaide Library.

Figure 7.4: *Relative changes in p21 (white circles) and p27 (black circles) protein concentration after progestin exposure. Curves through the data points were produced using an unknown curve fitting program. Results obtained from [84].*

of either p21 or p27. By 18 hours after progestin exposure, p21 protein concentration had increased to levels 2-fold above control levels, and remained this high until 24 hours. By 30 hours, p21 had decreased again to control levels. The concentration of p27 did not increase noticeably until 18 hours, when a minor increase of about 1.25 of control levels is observed. By 24 hours, p27 was about 1.4 times control and at 30 hours, p27 had increased to 2 times control.

### **Attempting to simulate S phase changes using known changes in cyclin D**

In Chapter 6, we discussed how our mathematical model of cell cycle progression can be applied to the T47D cell line. We concluded that cyclin D concentration is the main driver of cell cycle progression in the T47D cell line, and we saw that our model produces good results across different experiments. Let's investigate the changes in the S phase proportion if we substitute the known changes in cyclin D from Figure 7.3 into our cell cycle model. In order to produce a cyclin D profile for our model, we would normally consider the relative change in both cyclin D1 and cyclin D3, and so we would consider a cyclin D profile for our simulation that is somewhere 'in between' the cyclin D1 and cyclin D3 curves from Figure 7.3. However, in this instance we will just use the cyclin D1 profile from Figure 7.3 rather than considering both cyclin D1 and cyclin D3. This is because (as we will see when we produce our simulation results) we require our cyclin D profile to decrease after 12 hours quite quickly, which is more in line with the known changes in cyclin D1. Biologically, this could be interpreted as assuming that cyclin D1 is more potent at encouraging cell cycle progression than cyclin D3. Producing a continuous representation of cyclin D1 from Figure 7.3 and substituting this profile our model gives the S phase changes shown in Figure 7.5. The solid curve in Figure 7.5(a) shows the continuous representation of cyclin D concentration changes after exposure, and the solid curve in Figure 7.5(b) represents the corresponding S phase changes using our model. As we can see, the modelled S phase profile does not compare well with experimental data.

We can use the known S phase profile to solve the inverse problem for cyclin D (we did this in Chapter 6 when trying to determine the cyclin D profile using the S phase data from [136]). A continuous representation of the S phase profile is shown in Figure 7.5 (b) as the dashed curve, and the corresponding cyclin D profile required to generate the

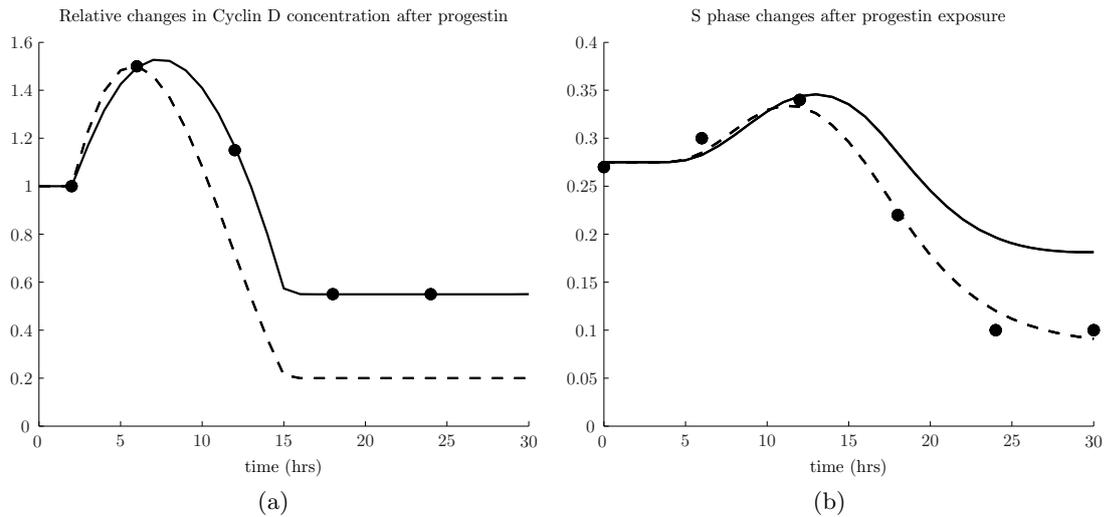


Figure 7.5: The cyclin D profile determined using experimental data (solid line, (a)) was used to generate the S phase profile using our model, shown as the solid line in (b). The S phase profile generated from observed S phase data ((b), dashed curve) was used to solve the inverse problem for cyclin D, shown in (a) as the dashed curve.

known S phase curve (determined by inverse-solving) is shown in Figure 7.5(a) also as the dashed curve. This investigation shows that to accurately model the low S phase proportion at later times, cyclin D concentration must be reduced to levels closer to 10% of control quite quickly after it has reached the maximum at about 6 hours after exposure. The discrepancy between our modelled S phase changes and the known S phase changes when using the cyclin D profile from Figure 7.3 cannot be due to the fact that we did not consider any influence from the cyclin D3 profile, because cyclin D3 decreases even slower than cyclin D1, and if we were to give any weighting to the cyclin D3 profile from Figure 7.3, then our modelled changes in the S phase profile would have resembled the known changes even less.

Musgrove et al. [84] suspected that the changes in cyclin D concentration observed could not account for the significant decline in the S phase proportion after progestin exposure, although they could not provide quantitative evidence for this. The model results presented here do provide quantitative evidence that cyclin D concentrations are not sufficient to explain the observed S phase changes for later times after progestin exposure, although they appear to explain the early behaviour quite well according to our model. In the next section, we consider some more data from [84] on observed changes in

NOTE:  
This figure is included on page 209 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 7.6: *Relative changes in CDK4 activity (black circles) and cyclin E-CDK2 activity (white circles, called cyclin E1). Curves through the data points were produced using an unknown curve fitting program. Results obtained from [84].*

cyclin-CDK activity, and we will see that cyclin D-CDK4 activity is observed to decrease to about 10% of the starting value, which is a significant drop when compared to the observed changes in cyclin D1 and cyclin D3 concentration. We will see that the observed changes in cyclin D-CDK4 activity compare very well with our cyclin D profile generated by inverse-solving the S phase profile from [84] (see Figure 7.5).

### **Changes in cyclin-CDK activity after progestin exposure**

As we discussed earlier, the total concentration of internal cell cycle proteins may not always be informative about the rate of cell cycle progression, as they are only active when bound to their partner CDK's. Even when cyclin E1 is bound to its CDK partner, CDK2, the cyclin E1-CDK2 complex is inactive if it is also bound to p27 or p21. Thus, a more accurate measure of the influence of these proteins on cell cycle progression is a measure of the activity of cyclin-CDK complexes, and in particular, we wish to investigate whether cyclin D-CDK4/6 activity may actually decrease significantly, providing some explanation for why the S phase proportion decreases after 10 hours of progestin exposure. This investigation was performed in [84], and the observed changes in cyclin activity are shown in Figure 7.6.

The activity of CDK4 was thought to represent the overall activity of cyclin D within the cells. Although cyclin D1 and cyclin D3 also bind to CDK6, the concentration of

CDK6 in the T47D cells used was low [84], and so measuring overall CDK4 was considered representative of cyclin D activity. The activity of cyclin E1-CDK2 are also shown (again note that cyclin E1 is referred to as cyclin E in this figure). Simply measuring the activity of CDK2 would not be representative of cyclin E1-CDK2 activity, as CDK2 also binds to cyclin A (and also to cyclin E2), and so it was necessary to measure the activity of the specific complex cyclin E1-CDK2.

As we can see from this representation, CDK4 activity did not change significantly within the first 12 hours of progestin treatment. By 18 hours after progestin exposure, CDK4 activity decreased to about 40% of control levels, and continued to decrease reaching 10% of control levels by 24 hours and remaining this low until 30 hours into the experiment. We note that CDK4 activity does not correlate well with observed changes in cyclin D1 or cyclin D3 concentration after progestin exposure (see Figure 7.3), implying that CDK4 activity is somehow being altered after progestin exposure, without correspondingly decreasing the concentration of cyclin D1/D3 molecules. In fact, the changes in CDK4 activity compare very well with the inverse-solved cyclin D changes we produced using our model (see Figure 7.5). This is true except for the data point at 6 hours: in our simulated profile, we predicted a 50% increase in the influence of cyclin D, whereas the data shown in Figure 7.6 indicates that CDK4 activity does not change significantly within the first 10 hours. We believe it is unlikely that CDK4 activity remains unchanged after progestin exposure because there is a noticeable increase in the rate of S phase entry after exposure, which can only occur if CDK4 activity has also increased. We will discuss how we deal with this discrepancy later in Section 8.4.

Figure 7.6 also shows that Cyclin E1-CDK2 activity increased slightly 6 hours after progestin treatment. By 12 hours, cyclin E1-CDK2 activity had declined, and was only 70% of control levels at this time. Cyclin E1-CDK2 activity continued to decline reaching a minimum of 23% by 24 hours. At 30 hours, cyclin E1-CDK2 activity had increased slightly and was at 40% of the control levels by this time. This slight increase in cyclin E1-CDK2 activity corresponds with the observed increase in cyclin E1-CDK2 concentration observed in Figure 7.3.

We now have an understanding of how certain protein concentrations change, as well as how the activity of critical proteins (such as cyclin D-CDK4) change after progestin expo-

sure. We would now like to get an understanding of the relationship between cyclin-CDK concentration, cyclin-CDK activity and the association between cyclin-CDK molecules and the inhibitors p21 and p27. As we discussed previously, cyclin E1-CDK2 activity can change if its association with the inhibitors p21 or p27 changes. The activity of cyclin E1-CDK2 molecules can also change if total cyclin E1-CDK2 concentrations change. We wish to investigate whether the main cause of cyclin E1-CDK2 activity changes are due to a decrease in total cyclin E1 concentrations, or if it is mainly due to increased association with the inhibitors p21 and p27. We would also like to get a better understanding of the cause of decreased cyclin D-CDK4 activity, which has decreased to less than 10% of control, while total cyclin D concentrations have only been reduced to 50% of control. This could be due to a decreased association between cyclin D and CDK4 after progestin exposure. However, we find that by considering some additional data (which will be discussed in the next section), that this is not the cause of the reduced activity of cyclin D-CDK4 molecules.

## 7.2 Determining the molecular make-up of cyclin-CDK complexes using gel filtration chromatography

In this section, we will investigate how the composition of each protein changes - that is, how the complexes that each protein is found in change in concentration over time. Such an investigation was carried out in [84]. This was determined by investigating the molecular mass of the complexes that the proteins appeared in. For instance, if cyclin E1 is found to be in low molecular mass complexes, then it is likely to be bound to very few other proteins. However, if cyclin E1 appears in high molecular mass complexes, then cyclin E1 is likely to be bound to many other proteins. Determining the type of complexes which contain cyclin E1 is performed using a process called **gel filtration chromatography**. Before we discuss their experimental results, we will describe how this technique works.

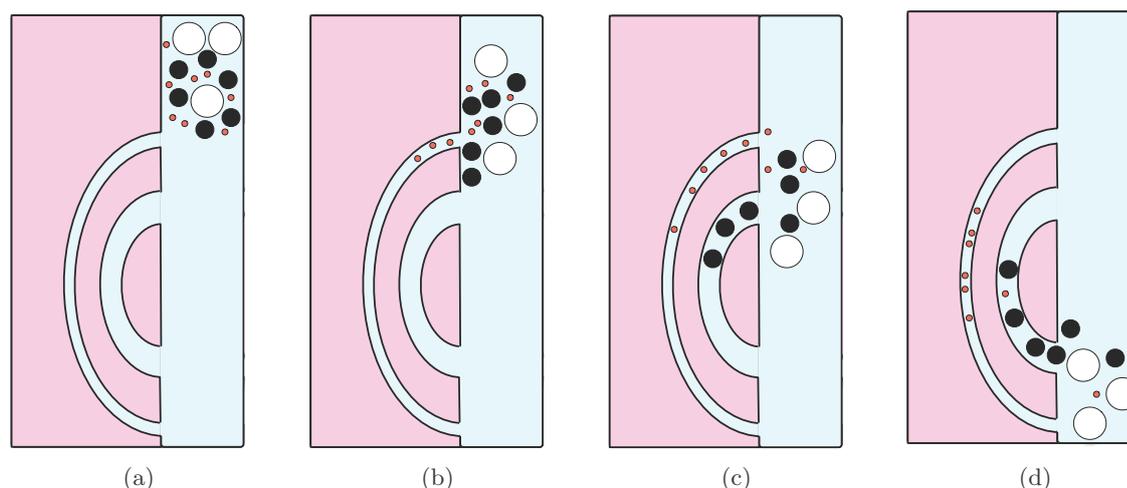


Figure 7.7: *Process of gel filtration chromatography, where a sample of molecules is separated by size. The sample is passed through a column, shown in (a), and the molecules proceed to move towards the exit of the column at the bottom. As the sample moves through the column, smaller molecules move into pores ((b) and (c)), while larger molecules that cannot fit continue to move through the main body of the column. Thus, the molecules in the original sample tend to leave the column in descending order of molecular size, as seen in (d). The actual layout of the column shown here is not based on the actual layout of columns used in gel filtration chromatography, and is to be interpreted purely as a conceptual tool for understanding the concept that pore detours result in smaller molecules taking longer to pass through the column than larger molecules.*

### Gel filtration chromatography and the elution profile

To determine the molecular mass distribution of a protein, gel filtration chromatography is used. The T47D cells that are under investigation are **lysed** (split open), and the internal cell fluid is extracted. The sample is injected into a vertical column at the top of the chromatographer (see Figure 7.7). As the sample moves down through the column towards the exit at the bottom, the molecules are able to ‘detour’ through some side passages that exist throughout the column. These side passages vary in diameter. This means that large molecules will not be able to fit into many of the pores in the column (they are simply too large), and will pass through the column with few detours. Small molecules, however, will be able to fit into many more pores, and may take many detours as they move through the column. Thus, larger molecules will exit the column earlier (or **elute** earlier) than smaller molecules. This process of separating a sample by molecule size is shown in Figure 7.7.

As the solution exits the column over time, the sample is collected in constant volumes

NOTE:  
This figure is included on page 213 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 7.8: *The percentage of total p27 protein eluting in ordered fractions labelled 1-12, modified from [84].*

called **fractions**. The first 2mL of solution that is observed to exit the column may be collected into the first fraction. The next 2mL of solution that exits the column will be collected into the second fraction, and so on, so that an ordered collection of constant volume fractions is collected. As larger molecules will exit the column first, the order of the fractions will correlate highly with the size of the molecules present in the fractions. For each fraction, the concentration of various proteins of interest are measured, and compared with the concentrations of those proteins present in other fractions. An example of an elution profile for the protein p27 is shown in Figure 7.8, which was taken and modified from [84]. In this experiment, T47D cells were lysed and internal cell samples were analysed. The representation in Figure 7.8 shows the percentage of total p27 eluting at various fraction numbers. This representation shows that little p27 elutes at low fraction numbers, and that most of the p27 protein elutes at higher fraction numbers (fractions 6-10), implying that p27 is mainly present in small molecule complexes.

We have already noted that increasing fraction numbers are correlated with decreasing molecular size. However, what is of more interest is molecular mass, from which the constituent protein composition can be more easily inferred. In [84], molecules of known molecular mass were passed through the column along with the sample, and the fractions which contained these known molecules were recorded. This is represented in Figure 7.8 by the numbers 440 kDa (molecular mass of ferritin) and 158 kDa (molecular mass of aldolase) at the top of fractions 1 and 7 respectively, and where kDa represents the unit of measurement (kilo daltons). Thus, the difference between fractions 1 and 6 corresponds to

a difference in molecular mass of about 300kDa, and so as a rough estimate, each fraction number in this experiment represents molecules about 50kDa apart.

### **Downsides of gel filtration chromatography**

There are some disadvantages to using gel filtration chromatography. As molecules pass through the column, the time they spend before eluting will have an element of randomness. Thus, molecules of the same molecular mass may not all elute at identical fractions, and may be present in a range of fraction numbers. In particular, small molecules may elute at a large range of fraction numbers, because there are so many pores into which they can pass through, or may instead bypass, and so some small molecules may elute very quickly, and some may take a long time to pass through.

Other disadvantages in this procedure come from choosing the width of the large column through which molecules are passed. For instance, choosing a wide column to pass samples through allows for even very large molecules to be analysed. However, there is an increased variance of the time taken to pass through the column, due to there being more space within the column to 'move around'. Thus, molecules will elute across a wider span of fraction numbers. This high variance is a problem, and often smaller columns are used. However, this means that molecules that are too large will not be able to be analysed. Thus, often when gel filtration is carried out, there is a small volume representing molecules that were too large to pass through the column. The volume of a sample containing these large molecules is known as the **void volume**.

When interpreting results from gel filtration experiments, it is often assumed molecular size is highly correlated with its mass. For roughly spherical molecules, this may be true. For asymmetrical molecules, however, this may not always be the case. When a molecule is not spherical, it may fit into certain pores when oriented one way, but may be too large when oriented a different way, resulting in a large variance in the fractions at which such a molecule will elute.

One last problem comes from the fact that each fraction number represents a range of molecular masses, meaning that molecules with similar molecular masses may be indistinguishable if the fraction volumes are too large. As we mentioned earlier, in the experiment

from [84] one fraction volume represents molecules that may differ in molecular mass of about 50kDa. This implies that molecules that differ in molecular mass by up to 50kDa may elute in the same fraction, and therefore such molecules are indistinguishable.

### 7.3 The elution profiles from [84] determined using gel filtration chromatography

Now that we understand how elution profiles are generated, and the possible problems with the experimental procedure, we will consider the elution profiles determined in [84] after T47D cells were exposed to progestin. Figure 7.9 shows the elution profiles of the proteins cyclin E1, p21 and p27 at times 12, 18 and 24 hours after progestin exposure as the black dots. The elution profiles of these proteins in control cells were also taken, and are shown in each profile at the white dots. The fraction number is shown on the x-axis, where the smaller fraction numbers were collected from the column first. Thus, high fraction numbers correspond to low molecular mass and lower fraction numbers correspond to high molecular masses. The fraction numbers represent 2 mL collection volumes, with fraction 1 representing volumes 57-58mL, fraction 2 representing volumes 59-60mL, ... , fraction 12 representing volumes 81-82mL. As fraction 1 starts at 57 mL, the void volume is 56 mL.

As we can see from the figure, the left most column represents the elution profile for cyclin E1 at 12, 18 and 24 hours after progestin exposure, and in each of these graphs, the elution profile for cyclin E1 in control cells is shown as the white dots for comparison. In control cells, almost all of the cyclin E1 elutes between fractions 4 and 11. We can see that at 12 hours after exposure, the cyclin E1 elution profile has not changed significantly from controls, however at 18 and 24 hours, significantly more cyclin E1 elutes in fractions 4-8, implying that progestin causes cyclin E1 molecules to associate in higher molecular mass complexes.

In the second column of Figure 7.9, we have the elution profile for p27 at 12, 18 and 24 hours after progestin exposure. In control cells, the vast majority of p27 elutes in fractions 7-9. However, after progestin exposure, more and more p27 is found eluting in fractions 3-6, so that by 24 hours after exposure, about 40% of total p27 elutes in these

## NOTE:

This figure is included on page 216 of the print copy of the thesis held in the University of Adelaide Library.

Figure 7.9: *Percent of total cyclin E1 (first column, referred to as cyclin E), p27 (second column) and p21 (third column) eluting at various fractions. The first row represent protein fractions at 12 hours after progestin, the second row at 18 hours and the third row at 24 hours after progestin treatment. In all graphs, the white circles represent the elution profiles of control cells. Results obtained from [84].*

fraction numbers, and the remaining 60% elutes in fractions 7-9, implying that p27 is found in larger molecular mass complexes after progestin exposure.

The third column of Figure 7.9 shows the elution profiles for p21 after progestin exposure. In control cells, we see that p21 mainly elutes in fractions 5-8, and after progestin exposure, a larger proportion of total p21 is found in fractions 3-5. Thus, like p27, p21 appears in larger molecular mass complexes after progestin exposure.

### **Interpreting the elution profiles from [84]**

We have just used the elution profiles to conclude, in a qualitative sense, how the molecular masses of the proteins cyclin E1, p27 and p21 change after progestin exposure. We now wish to determine in a quantitative sense, how the proteins change in association with each other after progestin exposure. We do this by assuming that certain molecular complexes elute in certain fraction numbers. For instance, it was noted in [84] that in the cyclin E1 elution profile, fractions 8-10 contained cyclin E1-CDK2 molecules, and fractions 4-8 contained cyclin E1-CDK2-p21 and cyclin E1-CDK2-p27 molecules, and so by investigating how the cyclin E1 elution profile changes after progestin exposure, we can quantify how the concentration of these specific complexes changes after progestin exposure.

As mentioned earlier, molecules of the same size may elute at a range of fractions, due to there being an element of randomness in the times taken to progress through the column. This means that molecules of the same type are likely to be present in a range of fraction numbers, and so there will not be a unique fraction number that a molecule of interest will exist in (as we have just discussed, specific complexes containing cyclin E1 were present in a range of fractions according to [84]). Thus, we will not assume that every molecule will elute at the fraction number corresponding to its actual mass 100% of the time. Instead, we will assume that the fraction numbers at which any one molecule elutes at follows a Gaussian (Normal) distribution. This means we will assume that molecules will most likely elute at the fraction number corresponding to their molecules mass, and with slightly reduced probability, they will elute at one of the neighbouring fraction numbers. With slightly reduced probability again, that same molecule may elute at an even further fraction number, and so on. Because we have

chosen a Gaussian distribution, we are assuming that a molecule is just as likely to elute at a higher fraction number as it is to elute at a lower fraction number. However, we note that the exact functional distribution is not likely to be Gaussian. For example, there is a minimum time with which a molecule can pass through the column (it cannot pass through instantaneously), however there is no upper bound on the time taken for a molecule to pass. Thus, we may expect a skewed distribution to be more realistic, in order to capture this effect. However, we find that a Gaussian distribution of fraction numbers is good enough for our purposes. We will assume that the distribution of fraction numbers that a molecule exists in takes the form of a Gaussian distribution. We will use the notation  $N(a, b)$  to describe a Gaussian distribution with mean  $a$  and variance  $b$ .

In gel filtration chromatography, some small molecules may move through the column without detouring through any side channels, and some small molecules may spend a lot of time in side channels. However, large molecules have much less options about the paths they can choose when moving through columns. Thus, small molecules will tend to have a larger variance in the fraction numbers they elute at, implying it is more realistic if we choose higher variances for larger elution fraction numbers. However, the presence of other factors (such as the presence of undetermined molecules in a fraction, or errors in measurement), also contribute to the variance of molecular distributions. Thus, we find that considering different variances for different mean elution fraction numbers is an unnecessary complication. For simplicity, we assume that the variance of the Gaussian distribution for each molecule of interest is 1, and we note that trying different values for the variance does not significantly change our results (not shown). To get an idea of what such a distribution may look like, the Gaussian distribution with mean 6 and variance 1 is shown in Figure 7.10. This representation shows that about 40% of the total complex will elute at the mean fraction number of 6, and about 90% of the complex will elute within 2 fraction numbers of the mean.

We can use this Gaussian representation to make conclusions about the protein complex make-up of the elution profiles. We will begin by investigating the protein complex make-up of cyclin E1 molecules.

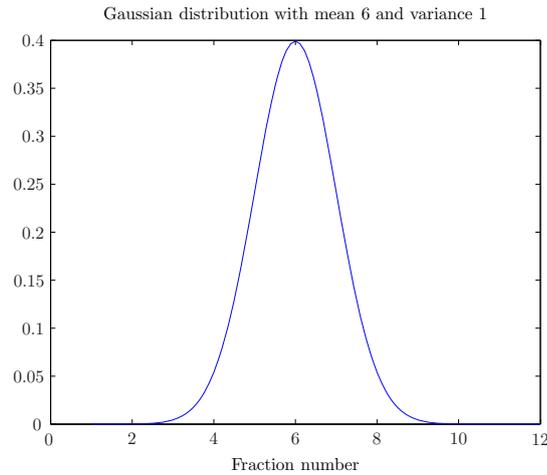


Figure 7.10: *Gaussian distribution of mean 6 and variance 1.*

### The cyclin E1 profile

The cyclin E1 elution profile for 12, 18 and 24 hours after progestin exposure can be seen in the first column in Figure 7.9. It was suggested in [84] that most cyclin E1 is found in cyclin E1-CDK2, cyclin E1-CDK2-p21 or cyclin E1-CDK2-p27 complexes. There were no observable free cyclin E1 molecules, because there was an abundance of CDK2, and so cyclin E1 was always bound to CDK2 due to the high affinity between the two proteins. Thus, we will assume that cyclin E1 exists in three forms: free cyclin E1-CDK2 cyclin E1-CDK2-p21 or cyclin E1-CDK2-p27. We will now determine which fraction numbers correspond to which of these forms.

It was noted in [84] that the cyclin E1 present in fractions 4-8 had very little kinase activity, however the cyclin E1 in fractions 8-10 exhibited kinase activity. It was then concluded that the cyclin E1 in fractions 4-8 must be bound to inhibitors such as p21 or p27 in addition to CDK2 (making them inactive cyclin E1-CDK2-p21 or cyclin E1-CDK2-p27 molecules), whereas cyclin E1 in fractions 8-10 must not be bound to an inhibitor (and therefore of the form cyclin E1-CDK2). As p21 and p27 only differ in molecular mass by 6kDa, the p21 and p27-bound cyclin E1-CDK2 complexes will elute at the same fraction numbers, i.e. in fractions 4-8. We take the midpoint of this range (which is 6) to determine the mean of our distribution, and we find that cyclin E1-CDK2-p21/p27 molecules elute with a distribution of  $N(6, 1)$ . Similarly, as free cyclin E1-CDK2 molecules elute at fractions 8-10, we will assume that these molecules can be represented by  $N(9, 1)$ .

NOTE:  
This figure is included on page 220 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 7.11: *The relative intensity at which cyclin D1 and D2 elute at various fractions both in control cells (top two rows) and 24 hours after progestin exposure (bottom two rows). Results obtained from [84].*

### The cyclin D profile

The elution profile for cyclin D1/D3 molecules was not shown in Figure 7.9, however it was investigated in [84], and the cyclin D1 and cyclin D3 elution profiles after progestin exposure are shown in Figure 7.11. It was observed that both before and after progestin treatment, almost all of the cyclin D1 and cyclin D3 complexes were bound to CDK4 and to p21 or p27. Thus, as the western blots have highest intensity at fractions 6-9 and 5-10 for cyclins D1 and D3 respectively after progestin treatment cyclin D1-CDK4-p21/p27 and cyclin D2-CDK4-p21/p27 complexes elute with a mean of 7.5. As cyclin D1 and D3 both have the same mean, are regulated similarly after progestin exposure (see Figure 7.3) and perform the same effects within a cell, we will, for simplicity, consider just cyclin D-CDK4-p21/p27 complexes, where cyclin D refers to the combination of cyclins D1 and D3. The small amount of cyclin D1 and cyclin D3 eluting in fractions 12-13 were suggested to be free cyclin D1/D3 molecules, as they exhibited no kinase activity [84], and we will assume that free cyclin D1 and cyclin D3 molecules elute with a mean of 13 (again, we will just refer to these two proteins collectively as cyclin D). We also observe that some cyclin D1 and cyclin D3 elute at fractions 11, and we will assume that this represents cyclin D-CDK4 molecules, as these molecules should have molecular mass between that of cyclin D and cyclin D-CDK4-p21/p27. It is clear from the western blots in Figure 7.11 that the concentration of cyclin D-CDK4 must be significantly lower than that of cyclin D-CDK4-p27 or free cyclin D molecules, and that after progestin exposure, the concentration of cyclin D-CDK4 molecules decreases significantly (the intensity of the blots in fractions 9-11 is low in control cells, and decreases to even lower levels after

progesterin exposure).

### The p27 profile

We have already determined that cyclin E1-CDK2-p27 elutes at fractions  $N(6, 1)$  and cyclin D-CDK4-p27 elutes at fractions  $N(7.5, 1)$ , and the presence of molecules at these fraction numbers can be seen in the p27 elution profile in Figure 7.9. However, we also observe that p27 elutes in smaller fraction numbers. In particular, after 24 hours of progesterin treatment, p27 is observed to elute at fractions 3-5. This cannot be explained by the presence of cyclin E1-CDK2-p27 molecules or cyclin D-CDK4-p27 molecules, which are represented by  $N(6, 1)$  and  $N(7.5, 1)$ .

In [16], it was discovered that cyclin E2 plays a major role in T47D-induced cell cycle progression. It was suggested that, although not investigated in [84], cyclin E2 plays a significant role in progesterin induced cell cycle progression [Musgrove, personal communication] and is likely to be present in p21 and p27 complexes. We will thus assume that cyclin E2-CDK2-p27 complexes are eluting at fractions with a mean of 4. Since cyclins E1 and E2 have the same molecular mass, we would expect them to elute at the same fractions. However, it was observed in [84] that the inhibitor-bound cyclin E1-CDK2 complexes elute at a lower fraction than anticipated, likely due to the presence of other proteins in the complex. It is not unreasonable to assume that inhibitor-bound cyclin E2-CDK2 complexes may be bound to other proteins, causing this complex to elute at an even lower fraction, and so we will assume that this is the case here, and choose cyclin E2-CDK2-p27 complexes to elute with a mean of 4.

### The p21 profile

We have already determined that cyclin E1-CDK2-p21 elutes in fraction  $N(6, 1)$  and cyclin D-CDK4-p21 elutes in fractions  $N(7.5, 1)$ . It was also determined that a significant amount of p21 elutes with cyclin A-CDK2 complexes, and so we must determine at what fraction number this complex will elute. However, the observed p21 elution profile appears to elute very near to fractions 6 and 7.5, and so there is no clear fraction number at which cyclin A-CDK2-p21 can elute at. We will determine the mean elution range for cyclin

Table 7.1: *Values of mean elution fractions*

| complex            | mean    |
|--------------------|---------|
| cyclin D-CDK4-p27  | 7.5     |
| cyclin D-CDK4-p21  | 7.5     |
| cyclin D           | 13      |
| cyclin E1-CDK2-p27 | 6       |
| cyclin E1-CDK2-p21 | 6       |
| cyclin E1-CDK2     | 9       |
| cyclin E2-CDK2-p27 | 4       |
| cyclin E2-CDK2     | unknown |

A-CDK2-p21 complexes later in Section 7.4.

We note that we have made the assumption that the total concentration of p27, p21, cyclin E1 and cyclin D is made up only of the complexes discussed above. It is likely that these proteins may be found in other protein-protein associations that we have not considered. However, to our knowledge, there are no other known complexes that we would expect to be of significance in progestin-induced cell cycle progression.

## 7.4 Quantitative analysis of the elution profiles from [84]

### The modelled elution profile for p27

Now that we know the protein complexes in which p27 can exist, we can determine how much of the total p27 protein exists in each of these complexes. For instance, because we are assuming that p27 exists in three possible complexes (cyclin D-CDK4-p27, cyclin E1-CDK2-p27 and cyclin E2-CDK2-p27), then the elution profile for p27 in Figure 7.9 can be expressed as a linear combination of the known molecular complexes that can contain p27. Using the known elution profile of p27 in control cells, we can write

$$[0 \ 0 \ 1 \ 3 \ 4 \ 6 \ 26 \ 37 \ 20 \ 5 \ 0 \ 0] = a \times N(4, 1) + b \times N(6, 1) + c \times N(7.5, 1), \quad (7.1)$$

where the numbers on the left were read off from the elution profile shown in Figure 7.9, and  $a$ ,  $b$  and  $c$  represent the relative contributions of cyclin E2-CDK2-p27, cyclin E1-CDK2-p27 and cyclin D-CDK4-p27 respectively. The values of  $a$ ,  $b$  and  $c$  should loosely

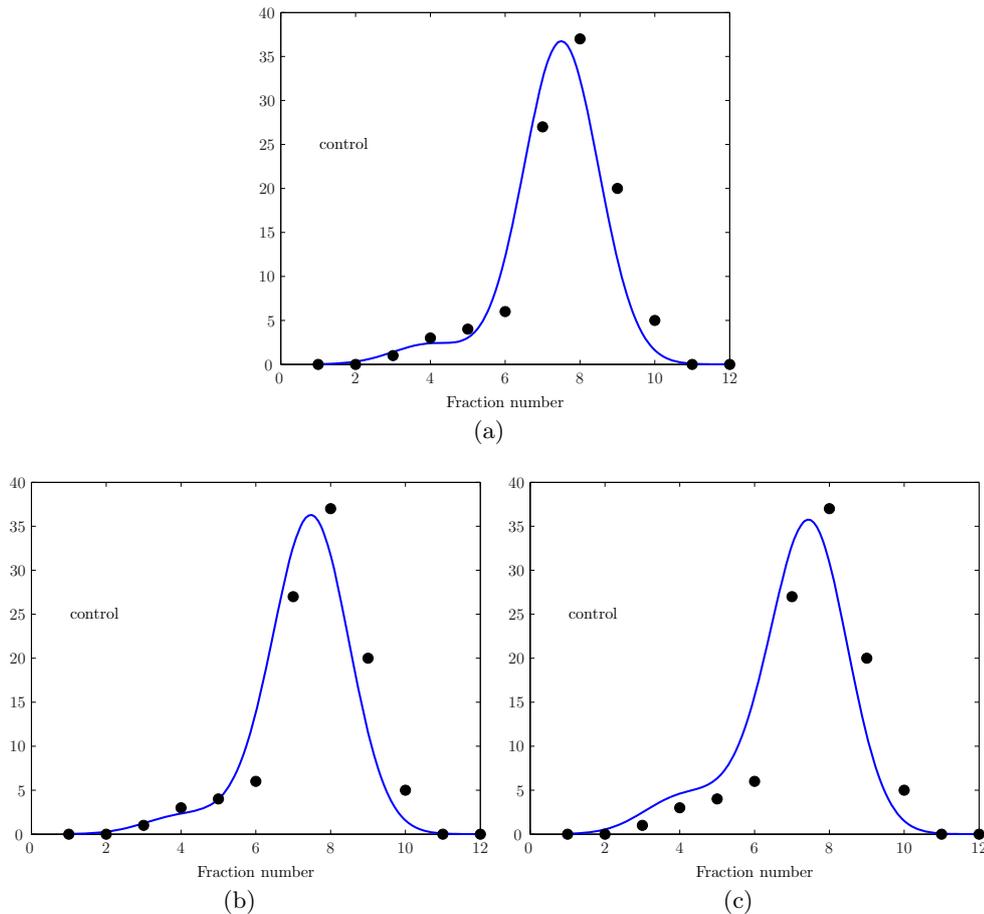


Figure 7.12: Modelled elution profiles for the p27 protein in control cells using (a)  $a = 5.8$ ,  $b = 0.0$ ,  $c = 92$  and (b)  $a = 5$ ,  $b = 5$ ,  $c = 90$  and (c)  $a = 10$ ,  $b = 10$ ,  $c = 80$  are shown as the blue curve in each figure. Experimental data from [84] is shown as the black dots.

represent the approximate percentage of total protein found in that protein complex. For instance,  $a = 50$  implies that about 50% of total p27 is bound to cyclin E2-CDK2 in control T47D cells.

We can determine the optimal values of  $a$ ,  $b$  and  $c$  by minimising the difference between the LHS and the RHS of equation (7.1) in a least squares sense. Doing this for control p27 cells gives  $a = 5.8$ ,  $b = 0.0$  and  $c = 92$ , to 2 sig. fig. respectively, and the resultant modelled p27 elution profile as shown in Figure 7.12(a). We can see that the modelled elution profile for p27 in control cells compares well with the observed elution profile at this time (shown as black dots). However, we may also get equally good fits to the modelled elution profile by considering other values of  $a$ ,  $b$  or  $c$ . The value of  $a$  is likely to be quite accurate, as we can see that the modelled curve fits well with the points around

fraction 4, and changing the values of  $b$  or  $c$  are not likely to alter the modelled curve around fraction 4. In addition, the value of  $c$  must remain quite high, as the known data points to the right of 7.5 are quite high, and so the contribution of cyclin D-CDK4-p27 must be significant. The value of  $b$  is less obvious because the amount at this fraction may be influenced partly by cyclin E2-CDK2-p27 and by cyclin D-CDK4-p27 molecules. In addition, we believe the solution  $b = 0$  is not realistic. If we enforce a minimum on the values of  $a, b$  and  $c$  of 5% of the total p27 protein in control cells, then performing the least-squares fit again produces  $a = 5$ ,  $b = 5$  and  $c = 90$ . The resultant modelled elution profile is shown in Figure 7.12(b). We can see the difference between the two representations in fractions 4-6, however we note that overall the fit is still very good. We can do a small investigation to determine at what point the fit starts to look unrealistic. If we enforce a minimum of 10% on each of the parameters, then our least squares fit gives  $a = 10$ ,  $b = 10$  and  $c = 80$ , and the resultant fit is shown in Figure 7.12(c). We can see that the fit is starting to look bad for fraction numbers 3-6. Thus, we will assume that  $a = 5$ ,  $b = 5$  and  $c = 90$ , and we have some confidence that this is accurate to  $\pm 5\%$ .

We can repeat the process for the p27 elution profile at 12, 18 and 24 hours after progestin exposure, and at each time point, we will assume that  $a, b$  and  $c \geq 5\%$ . Doing this gives Figures 7.13(a) - (d). We can see that the modelled elution profiles for p27 compare reasonably well with the data points at 12, 18 and 24 hours after progestin exposure, except for some isolated data points.

The elution profiles shown in this figure give us an idea of how the protein complexes change after progestin exposure. For instance, in (a) we have  $a = 5$  and in (c) we have  $a = 15$ , which implies that cyclin E2-CDK2-p27 complexes made up 5% of total p27 in control cells, and by 18 hours after exposure, cyclin E2-CDK2-p27 complexes make up 15% of total p27. We can combine these results from the modelled elution profile with the known changes in total p27 protein from Figure 7.4, to determine quantitatively how such protein-protein complexes change in concentration after progestin exposure. Doing this provides us with the full profile of the p27 protein, shown in Figure 7.13(e). This representation shows the relative change in concentration of cyclin E2-CDK2-p27 (dark shaded area), cyclin E1-CDK2-p27 (gray shaded area) and cyclin D-CDK4-p27 (light shaded area). The shaded areas are placed one on another so that the total shaded area represents the relative change in total p27 concentration after progestin exposure. In

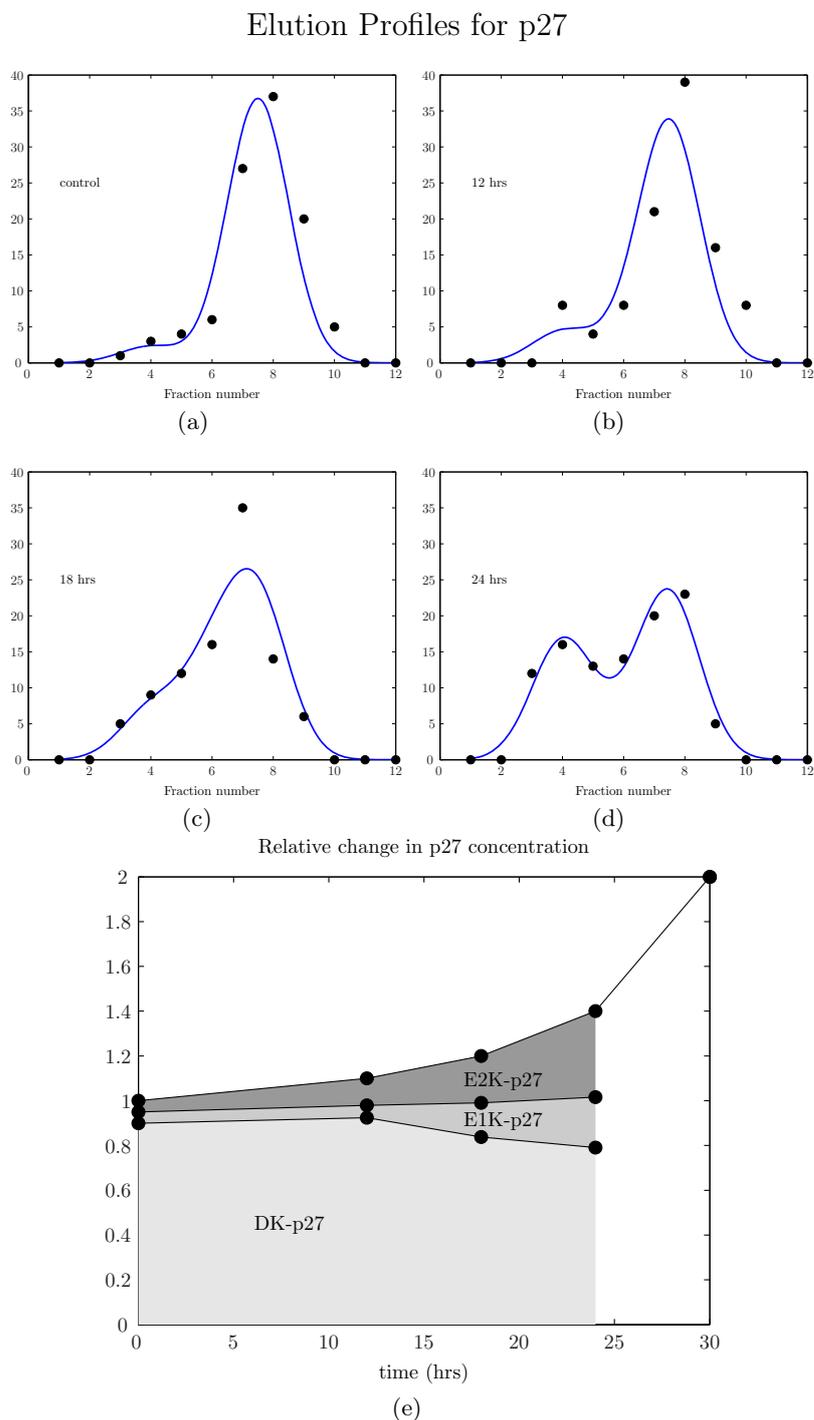


Figure 7.13: Modelled elution profile for the p27 protein at 0, 12, 18 and 24 hours after progestin exposure ((a)-(d)). The elution profiles and the total p27 concentration changes after progestin were used to produce (e) (see text for discussion of these figures).

Figure 7.13(e), we observe that the concentration of cyclin E2-CDK2-p27 has increased about 2-fold by 12 hours after exposure, but the concentrations of both cyclin E1-CDK2-p27 and cyclin D-CDK4-p27 have not noticeably changed in concentration by this time. By 18 hours, the concentration of cyclin E2-CDK2-p27 has increased to about 3 times control by 18 hours and then to 8 times control by 24 hours. The concentration of cyclin E1-CDK2-p27 increased significantly from undetectable levels in control cells to about 40% of total p27 by 18 hours, and then decreases to about 30% of the observed 18 hours level by 24 hours. The concentration of cyclin D-CDK4-p27 decreases to about 70% of control levels by 18 hours after progestin exposure, and then increases slightly to about 75% of control levels by 24 hours after exposure. In each of these figures, we note that these modelled elution profiles can be considered to be accurate to  $\pm 5\%$ , as we investigated earlier.

### The modelled elution profile for cyclin E1

We can perform a similar analysis for cyclin E1. We have already made the assumption that cyclin E1 can exist in three possible molecular complexes - cyclin E1-CDK2, cyclin E1-CDK2-p21 or cyclin E1-CDK2-p27. As both p21 and p27-bound cyclin E1 complexes elute at the same fractions in the cyclin E1 elution profile (mean of fraction 6), we have no way of distinguishing between them when considering just the cyclin E1 profile. Thus, we will consider inhibitor-bound cyclin E1 complexes (which includes both p21 and p27) and non inhibitor-bound complexes. As discussed earlier, inhibitor-bound complexes elute with a distribution  $N(6, 1)$  and cyclin E1-CDK2 complexes elute with a distribution of  $N(9, 1)$ . We thus determine the optimal values of  $a(t)$  and  $b(t)$  at times after progestin exposure in the expression

$$\text{elution profile cyclin E1}(t) = a(t) \times N(6, 1) + b(t) \times N(9, 1).$$

Performing a least squares fit between our modelled elution profile and the experimental data from [84] for times 0, 12, 18 and 24 hours after progestin allows us to produce Figures 7.14(a) - (d). The modelled elution profile compares well with the observed elution profile for fractions 4-10 after progestin exposure, however we have not been able to capture the presence of high molecular mass complexes eluting in fractions 1-3. Trying

## Elution Profiles for cyclin E1

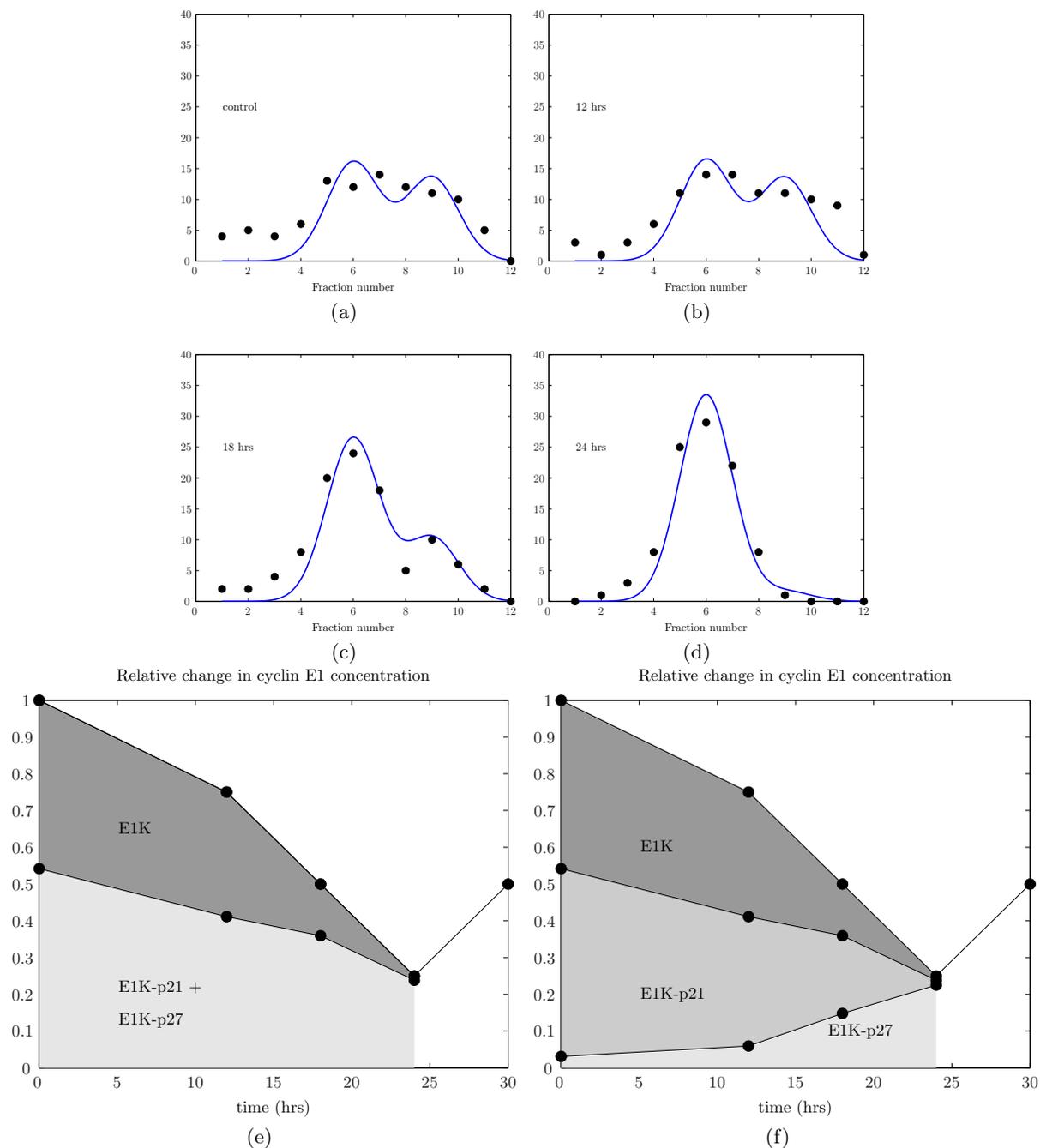


Figure 7.14: Modelled elution profile for the *E* protein at 0, 12, 18 and 24 hours after progestin exposure ((a)-(d)). The elution profiles and the total *E* concentration changes after progestin were used to produce (e) and the known changes in cyclin E1-CDK2-p27 from Figure 7.13(e) was used to produce (f).

different values of  $a(t)$  and  $b(t)$  will not improve our fits, as we have not included a population of cyclin E1-CDK2 with fraction numbers in the range of 1-3. We do not know what these complexes might be.

The two peaks we are trying to fit to have mean fraction values sufficiently far apart (means of 6 and 9). Thus we can assume that the chosen values of  $a$  and  $b$  at all times after progestin exposure are reasonably accurate in this case, since there will be little overlap of the two Gaussian curves. This is in contrast to the elution profile for p27, where we concluded there was an error of about  $\pm 5\%$  due to the amount of overlap in the Gaussians (see Figure 7.12 and the corresponding text discussing this figure).

As we did when modelling the changes in p27 after progestin, we combine the information on total cyclin E1 concentration changes with the changes in the percentage of total cyclin E1 known to be in certain complexes from our modelled elution profiles to produce Figure 7.14. This figure shows the changes in free cyclin E1-CDK2 and inhibitor-bound cyclin E1-CDK2. We can do better than this, because we also know the relative changes in concentration of cyclin E1-CDK2-p27 complexes from Figure 7.13(e), and so we can actually differentiate between cyclin E1-CDK2-p21 and cyclin E1-CDK2-p27. We will explain how we do this now. In [113], it was noted that after 24 hours of progestin exposure, the vast majority of cyclin E1 was bound to p27. Thus, if we make the assumption that at 24 hours after progestin exposure, 90% of total cyclin E1 is bound to p27, then we substitute the known relative changes in cyclin E1-CDK2-p27 from Figure 7.13(e) into Figure 7.14(e) to produce Figure 7.14(f).

In Figure 7.14(f), we can see that initially, a large amount of cyclin E1-CDK2 was bound to p21 molecules, and after progestin exposure, cyclin E1-CDK2-p21 complexes decrease in concentration while cyclin E1-CDK2-p27 complexes increase in concentration. We also observe that the concentration of free cyclin E1-CDK2 remains approximately constant until 18 hours, when it decreases to about 40% of control. By 24 hours, the amount of cyclin E1-CDK2 that is *not* attached to an inhibitor has decreased to virtually undetectable levels. This does not seem to be in agreement with the results from [84] shown earlier in Figure 7.6, where the amount of active cyclin E1-CDK2 complexes decreases to a minimum of 25% of controls by 24 hours after exposure to progestin. However, it was noted in [84] that cyclin E1-CDK2 activity at 24 hours was observed to occur

in low fraction numbers of cyclin E1 (fractions 1 - 3), and that activity in other fraction numbers was virtually undetectable, in agreement with our modelled elution profiles. It was also noted that the activity in fractions 1-3 at 24 hours was not consistently observed. Authors of [84] did not know why cyclin E1-CDK2 activity peaked in these fraction numbers on some occasions. Thus, as we do not know why activity was sometimes observed here, and also because this observation was not consistent, we will not consider this in our investigations, and assume that the large decrease in free cyclin E1-CDK2 concentration after progestin exposure, as shown in Figure 7.14, is accurate.

### The modelled elution profile for p21

We can also model the elution profile for p21. As mentioned in Section 7.3, p21 can be found in three different complexes - cyclin E1-CDK2-p21, cyclin A-CDK2-p21 and cyclin D-CDK4-p21. We also concluded in this section that cyclin E1-CDK2-p21 and cyclin D-CDK4-p21 would elute with a distribution of  $N(6, 1)$  and  $N(7.5, 1)$  respectively. However, we did not know at what fractions cyclin A-CDK2-p21 would elute. We can determine this by solving the equation

$$\text{p21 elution profile}(t) = a(t) \times N(6, 1) + b(t) \times N(7.5, 1) + c(t) \times N(d, 1), \quad (7.2)$$

for the parameter  $d$ , where  $a(t)$ ,  $b(t)$  and  $c(t)$  represent the influence of the complexes cyclin E1-CDK2-p21, cyclin D-CDK4-p21 and cyclin A-CDK2-p21 respectively, and  $d$  represents the mean fraction value for cyclin A-CDK2-p21. The value of  $d$  must be the same for all time points after progestin exposure.

We know the relative changes in the concentration of cyclin E1-CDK2-p21 from Figure 7.14, however we do not yet know how much the cyclin E1-CDK2-p21 complex makes up the overall concentration of p21. We only need to determine the value of  $b(t)$  for one time point after progestin, and the value of  $b(t)$  for all other time points can be determined.

We use a least-squares approach to find the optimal values of our parameters, and we find that  $b(0) = 20$  and  $d = 5$ . The results of our model are shown in Figure 7.15. The modelled p21 elution profile implies that the concentration of cyclin D-CDK4-p21 does not

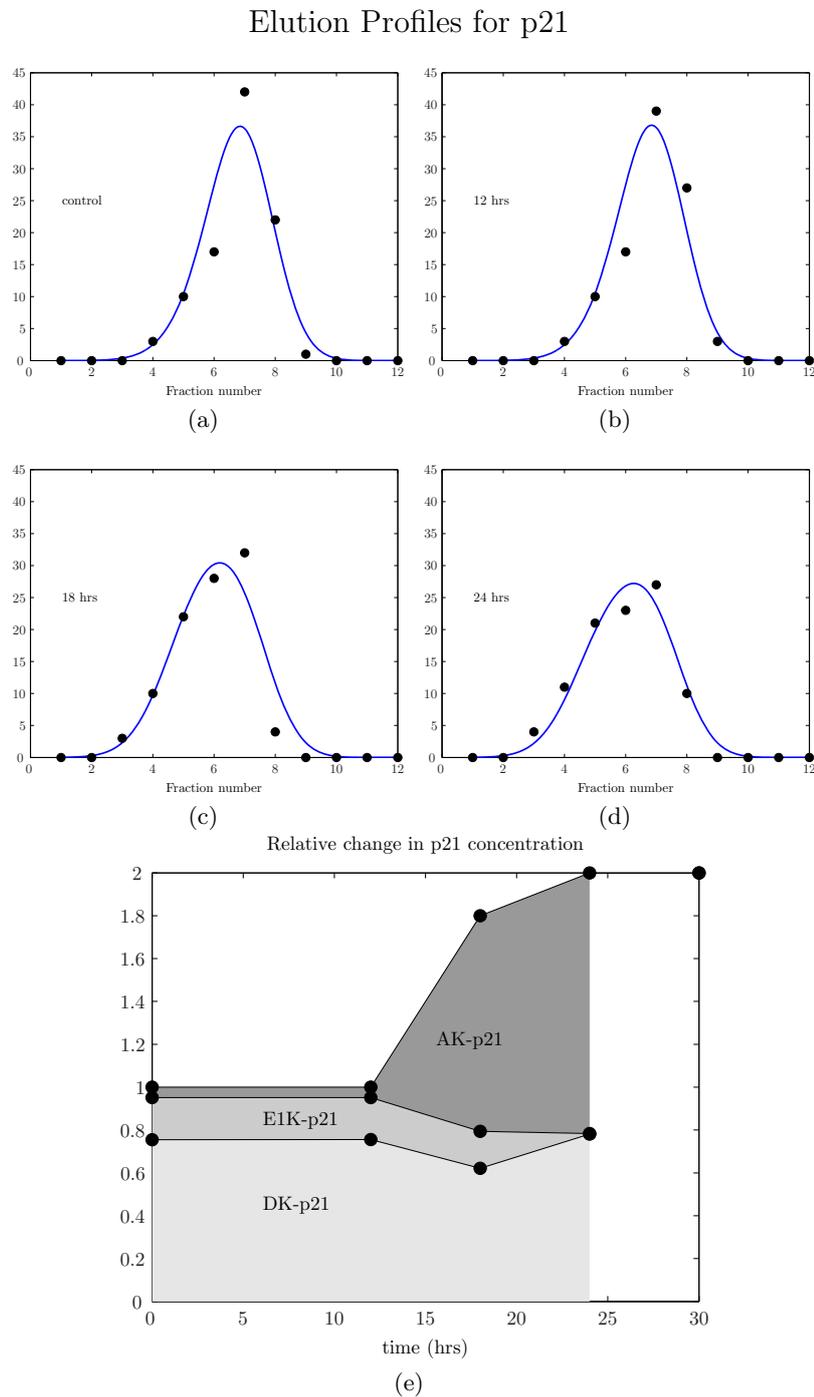


Figure 7.15: Modelled elution profile for the p21 protein at 0, 12, 18 and 24 hours after progestin exposure ((a)-(d)). The elution profiles and the total p21 concentration changes after progestin were used to produce (e).

change significantly, and that the concentration of cyclin A-CDK2-p21 increases drastically after progestin exposure. It was observed in [84] that cyclin A-CDK2-p21 complexes increased in concentration after progestin exposure, however there was no mention of the amount with which this occurred. In [37], it was observed that the overall concentration of cyclin A decreases to almost undetectable levels, however this observation implies that it is highly unlikely that cyclin A-CDK2-p21 complexes would increase in concentration, in disagreement with the observations in [84]. The observation that total cyclin A decreases drastically after progestin exposure from [37] also implies that our modelled results from Figure 7.15 are not accurate, as we observe a significant *increase* in cyclin A-CDK2-p21 molecules after exposure. Thus, we have a discrepancy here.

We could investigate other values of  $d$  and  $b(0)$  that produce reasonable fits to the data in an attempt to address the discrepancy discussed in the previous paragraph. However, we find that trying different values of these variables does not produce significantly different results. If we look at the elution profiles for p21, we note that for 18-24 hours after exposure, there is significantly more p21 in the low fraction number ranges. As this corresponds to the period when p21 doubles in concentration, we would expect most of the increase in p21 concentration to occur in these fraction ranges. We know from Figure 7.14(f) that cyclin E1-CDK2-p21 complexes do not increase in concentration (they decrease significantly) and the cyclin D-CDK4-p21 complexes exist in fraction numbers  $N(7.5, 1)$ , which is too far from the peak observed near  $N(4, 1)$ , meaning that cyclin A-CDK2-p21 complexes must be increasing at this time. It is possible that other proteins not discussed here are responsible for the observed increases in p21 protein concentration after progestin exposure, and this may warrant further investigation.

### The modelled elution profile for cyclin D

We can use the profiles already produced to determine the elution profile for cyclin D. We will ignore the presence of free cyclin D molecules, as, although these molecules make up a significant concentration of total cyclin D (see Figure 7.11), they do not contribute to cyclin D-CDK4 activity.

We wish to determine an elution profile for cyclin D. We will only consider cyclin D-CDK4, cyclin D-CDK4-p21 and cyclin D-CDK4-p27 complexes, and we will ignore free

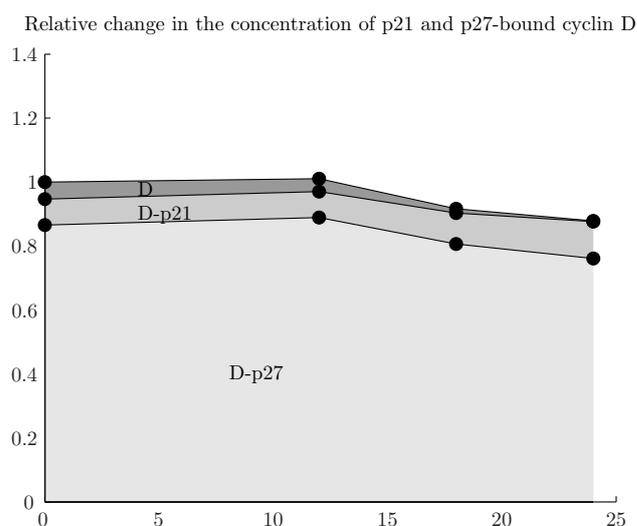


Figure 7.16: The relative changes in cyclin D-CDK<sub>4</sub> (darkest shaded region, labelled *D*), cyclin D-CDK<sub>4</sub>-p21 (dark shaded region, labelled *D-p21*) and cyclin D-CDK<sub>4</sub>-p27 (light shaded region, labelled *D-p27*) concentrations are shown. The shaded regions have been placed on top of each other so that the total shaded region represents the relative change in concentration of cyclin D-CDK<sub>4</sub> complexes, under the assumption that, in control cells, 85% of cyclin D-CDK<sub>4</sub> complexes are bound to p27 and 10% to p21 and 5% free.

cyclin D molecules. We can use the elution blots for cyclin D from Figure 7.11 to differentiate between cyclin D-CDK<sub>4</sub> and cyclin D-CDK<sub>4</sub>-p21/p27 molecules (as they elute at fractions 7.5 and 11 respectively), however differentiating between cyclin D-CDK<sub>4</sub>-p21 and cyclin D-CDK<sub>4</sub>-p27 complexes is more difficult. We can estimate the change in concentration of cyclin D-CDK<sub>4</sub>-p27 and cyclin D-CDK<sub>4</sub>-p21 complexes as follows. In [113], it was noted that in control cells, cyclin D was predominantly p27 bound, and not p21 bound. Thus we assume that in control cells, 85% of total cyclin D-CDK<sub>4</sub> complexes are p27 bound, 10% are p21 bound and 5% are in the form cyclin D-CDK<sub>4</sub> (we can see from the blots in Figure 7.11 that cyclin D-CDK<sub>4</sub> molecules make up a small proportion of total cyclin D in control cells). As we know the relative changes in concentration of p21 and p27-bound cyclin D-CDK<sub>4</sub> complexes, we can determine the change in concentration of inhibitor bound cyclin D-CDK<sub>4</sub> after progestin, allowing us to produce the elution profile for cyclin D-CDK<sub>4</sub>, shown in Figure 7.16. The representation in Figure 7.16 shows that cyclin D-CDK<sub>4</sub>-p27 complexes are predominant both before and after progestin treatment, in agreement with the observations in [113]. The concentration of total cyclin D-CDK<sub>4</sub> complexes do not change significantly after progestin exposure. This appears to be in disagreement with Figure 7.3, where cyclin D concentration was observed

to decrease by 50%. However in this section, we have chosen to ignore the presence of free cyclin D molecules. As the blots in Figure 7.11 indicate, these free molecules make up a significant portion of cyclin D concentration in control cells, but then decrease significantly in concentration after progestin exposure. These free molecules are likely accounting for the overall drop in cyclin D concentration of 50%.

As p27 and p21 are not thought to inhibit cyclin D-CDK4 activity, we would expect Figure 7.16 to also be a good representation of total cyclin D-CDK4 activity after progestin exposure. As there is not a significant change in the concentration of cyclin D-CDK4 complexes, we would conclude that there is not a significant change in the activity of cyclin D-CDK4 complexes. However, we can see from Figure 7.6 that CDK4 activity decreased by about 90% 24 hours after progestin exposure!

In [113], it was suggested that the decrease in CDK4 activity was due to the increase in p18 concentration. These p18 molecules bind CDK4, preventing them from binding cyclin D molecules. However, if the presence of p18 was significant in reducing CDK4 activity, we would expect to see a reduction in the concentration of cyclin D-CDK4 complexes, which is not observed in the representation shown in Figure 7.16. It is not just our representation in Figure 7.16 that implies that p18 is not responsible for the majority of the decrease in CDK4 activity, but observations from [113] which indicate that most cyclin D present in the cells is bound in cyclin D-CDK4-p27 complexes both before and *after* progestin exposure. We would need to observe more free cyclin D molecules after progestin exposure if p18 was responsible for the drop in cyclin D activity, not a sustained high concentration of cyclin D-CDK4-p27.

We conclude that to have a 90% reduction in cyclin D-CDK4 activity, p27 bound cyclin D-CDK4 complexes (which make up a majority of the overall concentration of cyclin D-CDK4) must move from being active to inactive complexes after progestin exposure. There is some recent evidence [50] suggesting that phosphorylation of p27 by tyrosine kinases (on Y88 or Y89) allows p27 to bind cyclin D-CDK4 complexes in a non-inhibitory manner (i.e. so that the cyclin D-CDK4-p27 complex is still active). If p27 is not phosphorylated in this way, then when bound to cyclin D-CDK4, it forms an inactive cyclin D-CDK4-p27 complex. Thus, a plausible mechanism of reduced cyclin D-CDK4 activity after progestin exposure is a reduction in phosphorylation of p27 (at the Y88 or Y89

sites), causing the p27 bound cyclin D-CDK4 complexes to become inactive instead of active. In [113], it was noticed that after T47D cells were exposed to progestin, there was a change in the phosphorylation state of p27, however it was not known (and still is not known) what the cause of the shift in p27 phosphorylation might be [Musgrove, personal communication].

We suggest that a very likely reason is the change in the phosphorylation of cyclin D-CDK4-p27 molecules, from non-inhibitory binding to inhibitory binding. We will discuss this possibility when modelling the changes in cyclin D concentration and activity later in Section 8.6.

## Summary

In this chapter, we delved further into the experimental results published in [84]. We determined quantitatively how the concentrations of the relevant protein-protein complexes changes after progestin exposure. In particular, we found that although cyclin D activity decreases significantly after progestin exposure, cyclin D-CDK4 complexes do not change significantly in concentration. The role of p18 was thought to be a significant contributor to reduced cyclin D activity [113], however as we discussed earlier, it cannot account for the majority of observed reduction in CDK4 activity. It was also suggested in [113] that p27 must play a significant role, however it was assumed to play a significant role in the sense that it inhibits cyclin E-CDK2 activity.

We suggest that in order to have a reduction in cyclin D activity while keeping cyclin D-CDK4 concentration high, cyclin D-CDK4-p27 complexes (which account for the majority of cyclin D-CDK4 complexes) must become inactive. In the next chapter, we will build a mathematical model which uses this theory, and comparing our model results to the data produced in [84], we get a deeper insight into the role of progestin in breast cell proliferation.

## Chapter 8

# A model of progestin effects on cell cycle proteins

In the previous chapter, we used data from [84] to determine changes in cyclin D, cyclin E1, p27 and p21 after progestin exposure. In this chapter, we will develop a model to simulate protein concentration changes after progestin exposure, with the aim of investigating whether the current understandings of protein interactions can explain the changes in protein concentrations observed after progestin exposure. The model introduced here provides insight into the key mechanisms that are responsible for the observed changes in protein concentrations after progestin exposure, and will later be used to simulate protein concentration changes under different conditions than those considered in [84].

If we were to include all of the proteins discussed in the previous chapter into our model, taking into account the different forms that proteins can appear in (e.g. free cyclin D, or cyclin D-CDK4 or cyclin D-CDK4-p27), we would have a large complicated model consisting of many variables, making it difficult to get real insight into the drivers of protein concentration changes after progestin exposure. We argue that some proteins are critical to understanding the role of progestin-induced cell cycle progression, while some proteins are not so significant, and we wish to simplify our model to include only those proteins that are important in progestin-induced cell cycle progression.

We begin by considering the role of p21 in progestin-induced cell cycle changes. It is thought that p21 only plays a minor role in cell cycle inhibition caused by progestin,

and that p27 is the main protein involved in cell cycle inhibition after progestin exposure [113]. As we saw in the previous chapter, p21 increased two-fold after progestin exposure, however we were unable to conclude which protein complexes p21 was found in. As a further complication, it is known that p21 is regulated mainly through its rate of transcription (it has a relatively constant rate of degradation), and we have no information about how the rate of p21 transcription changes after progestin exposure. In addition, we do not know what mechanisms drive the changes in cyclin E1-CDK2-p21 and cyclin D-CDK4-p21 complexes. Including the effects of p21 will involve many additional unknowns into our model system, making it more difficult to get an insight into the more relevant proteins involved in progestin-induced cell cycle inhibition. Thus we choose not to model the changes in p21 after progestin exposure.

In [16] it was suggested that cyclin E1 concentrations could be regulated in part by the proteins cyclin D and myc, and that cyclin E2 concentrations could not be regulated by myc, but could be regulated by cyclin D. This implies that there are differences in the way that these two proteins are regulated. However, in the T47D cell line, myc plays only a minor role after progestin exposure, and changes in myc concentration were not measured in the experiments performed in [84]. We have already seen that E type cyclins are predominantly regulated by E2F, which is in turn regulated by cyclin D. Thus, we will assume that both cyclin E1 and cyclin E2 are regulated in the same way after progestin exposure, and we will henceforth collectively refer to them as cyclin E.

Lastly, we choose not to consider the concentrations of free cyclin D and free cyclin E. This is because there are very high concentrations of CDK2 and CDK4 within the cells, so that almost all cyclin E and cyclin D molecules are bound to their respective CDK partners. Although there is a high concentration of free cyclin D in control cells, we cannot determine why this is the case, as there is a high concentration of CDK4 molecules for them to bind to. In addition, both free cyclin D and cyclin E are inactive, and therefore do not influence cell proliferation in any way. Lastly, we do not need to model the concentration of free CDK molecules, as these molecules tend to remain unchanged in concentration, and are always in abundance in the cell.

Thus, in summary, we will be considering the effects of progestins on the five proteins cyclin E-CDK2, cyclin D-CDK4, p27, cyclin E-CDK2-p27 and cyclin D-CDK4-p27. We

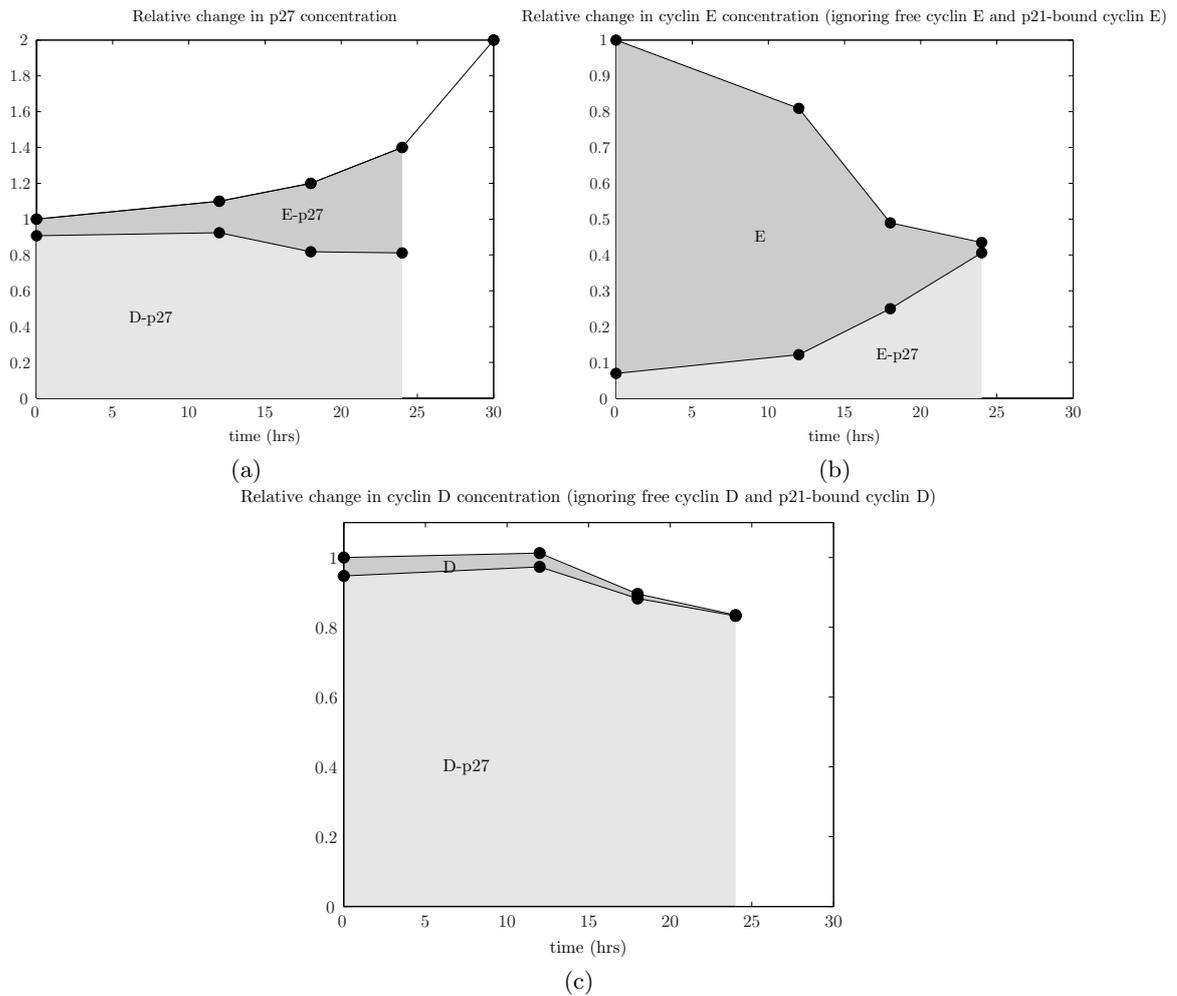


Figure 8.1: *Modified elution profiles for (a) p27, (b) cyclin E-CDK2 and (c) cyclin D-CDK4, constructed from those in Section 7.4*

can modify the cyclin E and cyclin D elution profiles from Section 7.4 by excluding the cyclin D and cyclin E1 complexes which contain p21, and by combining cyclin E1-CDK2 and cyclin E2-CDK2. Figure 8.1 shows the modified (and re-normalised) elution profiles for p27, cyclin E-CDK2 and cyclin D-CDK4 molecules.

### Previous mathematical models of protein concentration changes during cell cycle progression

Previous mathematical models have been developed in order to get a better understanding of how external substances influence internal cell cycle protein concentrations. Since the

later phases (S, G2 and M phases) occur similarly in both normal and cancer cells, many mathematical models have focussed on the dynamics which allow a cell to pass the R point [99, 88, 114, 141]. In [88], the proteins cyclin E-CDK2 and cyclin D-CDK4 were considered, as well as the inhibitors p21 and p27. In addition to these proteins, many other proteins were considered, resulting in a large model consisting of 22 differential equations. These 22 differential equations modelled not just progression past the R point, but also progression past the remaining cell cycle phases. In [141], cyclin D-CDK4, cyclin E-CDK2 and myc proteins were considered, however the influence of the inhibitors p21 and p27 were not considered. In [141], only progression past the R point was considered, and they only considered a set of 6 differential equations in their model.

Other models have also investigated how different proteins interact to help cells progress through the later autonomous cell cycle phases [119, 120, 35]. However, none of these models were able to use data that separated total cyclin concentrations with known concentrations of relevant molecular complexes (i.e. none were able to consider how much of each cyclin population was, say, bound to p27), and in particular, none of these models have been able to link changes in the concentration of these molecules to changes in the relevant CDK activities that we have available in this data. In addition, no previous models have considered the effects of progestin influences of cell cycle protein concentrations after progestin exposure. We will do this with our model of protein concentration changes that we develop over the next few sections.

## 8.1 A simple model of cyclin E-CDK2 and p27

In previous models that we have considered, we used variables that represented an average over many cells. For instance,  $S(t)$  represented the proportion of cells in a large system of cells that are in the S phase at time  $t$ . In this section, we will be considering a model of protein-protein interactions, and these interactions occur within an individual cell. However, we will be considering the concentration of these proteins averaged out over a large number of cells. This is because this is how the concentrations of these proteins are measured experimentally: cells are lysed (i.e. split open, allowing the internal contents to spill out) and then antibodies to a particular protein, say cyclin E, are injected into the

collection of lysed cells under investigation. The overall concentration of cyclin E relative to the overall number of cells in the sample is then measured, and the data recorded. Thus, the concentration of proteins recorded represents an average concentration per cell. Although it is possible that there may be a large variation in the concentration of a protein (for example, some cells may have had a very small amount of the protein of interest, while other cells may have had a very large concentration of that protein), we will assume that the reported average concentration of the protein is a good representation for the concentration of that protein in any one cell.

We will begin by developing a model representing the relationship between cyclin E-CDK2 and p27, which includes the molecules cyclin E-CDK2, p27 and cyclin E-CDK2-p27. This means that we are initially ignoring the effects of cyclin D-CDK4 and cyclin D-CDK4-p27. We do this for two reasons. One is that we wish to start with as simple a model as possible so that we can understand how the model behaves. Another reason is because we should be able to model the observed changes in cyclin E-CDK2, p27 and cyclin E-CDK2-p27 without needing to consider cyclin D-CDK4 and cyclin D-CDK4-p27 molecules. This is because the cyclin D-CDK4 and cyclin D-CDK4-p27 molecules do not interact with cyclin E-CDK2 molecules. Of course, total p27 concentration can be altered by changing the concentration of cyclin D-CDK4-p27. However, from Figure 8.1(b), we can see that cyclin D-CDK4-p27 concentration does not change significantly in concentration after progestin exposure. Thus, any changes in free p27, cyclin E-CDK2-p27 or cyclin E-CDK2 that take place in the experiment performed in [84] occur largely independently of cyclin D-CDK4 or cyclin D-CDK4-p27.

The concentrations of p27 and cyclin E-CDK2 at time  $t$  are denoted by  $P(t)$  and  $E(t)$  respectively. The cyclin E-CDK2 complex is not produced directly, but rather forms as the proteins cyclin E and CDK2 bind together. However, it is known that the concentration of cyclin E-CDK2 is highly correlated with the concentration of free cyclin E molecules (as discussed earlier, CDK2 molecules are well in abundance of cyclin E molecules, and so we would expect the concentration of cyclin E-CDK2 to be proportional to the concentration of free cyclin E molecules). Thus, instead of including multiple terms representing the production of cyclin E and the binding to CDK2, we assume that the rate of cyclin E-CDK2 formation is equivalent to the rate of cyclin E formation, as has been done in other mathematical models of cell cycle proteins [88].

We make the assumption that p27 is produced at the rate  $\alpha_p$ , where  $\alpha_p$  is a constant. As we saw in Chapter 2, cyclin E is regulated by E2F in cells that have passed the R point (and so it is produced only in  $G_c$  cells), which, in turn, is regulated by active cyclin D. We choose to bypass the intermediate step in our modelling, and we will assume that cyclin E is directly regulated by active cyclin D-CDK4 molecules, but only by cells that currently reside in the  $G_c$  phase. This may appear contrary to our claim that we are ignoring cyclin D-CDK4 molecules in this model, however we will later find that we can determine changes in cyclin D-CDK4 activity without needing to explicitly model changes in cyclin D-CDK4 concentration. Thus we assume that cyclin E-CDK2 is produced at the rate  $\alpha_e D_{act}(t) G_c(t)$ , where  $\alpha_e$  is constant,  $D_{act}(t)$  represents the relative concentration of active cyclin D molecules at time  $t$  and  $G_c(t)$  represents the  $G_c$  phase proportion at time  $t$ . We choose  $D_{act}(t)$  to represent the relative concentration, and not the absolute concentration, of active cyclin D-CDK4 molecules because (as we will see later,) we only have experimental data on the relative changes in the concentration of these molecules, and we will assume that the constant  $\alpha_e$  contains the relevant multiplication factor to convert  $D_{act}(t)$  into an absolute concentration. The p27 and cyclin E-CDK2 proteins have a natural decay rate of  $\beta_p$  and  $\beta_e S(t)$  respectively, where the term  $S(t)$ , which is the S phase proportion at time  $t$ , is included to capture the fact that degradation of cyclin E-CDK2 occurs mainly during the S phase. We also assume that p27 binds to cyclin E-CDK2 at a rate of  $\alpha_{pe}$ , creating the complex cyclin E-CDK2-p27, and the concentration of this complex at time  $t$  is denoted by  $P_E(t)$ . We do not include a natural decay rate of the complex cyclin E-CDK2-p27, as it is known that such inhibitor-CDK complexes are very stable [135, pg 273]. Instead, we assume that this complex disassociates to produce cyclin E-CDK2 and p27 at a rate of  $\beta_{pe}$ . These interactions can be expressed mathematically as

$$\frac{dP(t)}{dt} = \alpha_p - Ebind - \beta_p P(t), \quad (8.1a)$$

$$\frac{dE(t)}{dt} = \alpha_e D_{act}(t) G_c(t) - Ebind - \beta_e S(t) E(t), \quad (8.1b)$$

$$\frac{dP_E(t)}{dt} = Ebind, \quad (8.1c)$$

$$Ebind = \alpha_{pe} E(t) P(t) - \beta_{pe} P_E(t). \quad (8.1d)$$

As mentioned earlier, the values  $E(t)$ ,  $P(t)$  and  $P_E(t)$ , as well as  $S(t)$  and  $G_c(t)$ , represent average values over a large population of cells.

### Steady-states

The variables  $D_{act}(t)$ ,  $G_c(t)$  and  $S(t)$  are rate variables that depend upon the external environment, and their differing values over time directly drive the model system described here. The steady-state solution of the above system can be determined by setting each equation to zero, however there is no steady-state solution if the values of  $D_{act}(t)$ ,  $G_c(t)$  and  $S(t)$  are not constant. We denote  $D_{act}$ ,  $G_c$  and  $S$  to be the unchanging values of these variables. The steady-state solution can then be expressed as

$$P = \frac{\alpha_p}{\beta_p}, \quad (8.2a)$$

$$E = \frac{\alpha_e D_{act} G_c}{\beta_e S}, \quad (8.2b)$$

$$P_E = \frac{\alpha_{pe} E P}{\beta_{pe}}, \quad (8.2c)$$

where  $P$ ,  $E$  and  $P_E$  represent the steady-state values of p27, cyclin E and cyclin E-CDK2-p27 respectively. We can see that  $P$  is independent of any other variables, and will always be  $\frac{\alpha_p}{\beta_p}$ .  $E$  depends upon  $D_{act}$ ,  $G_c$  and  $S$  and so will depend on the environmental conditions. The value of  $P_E$  varies linearly with  $E$ , and hence is also dependent on the variables  $D_{act}$ ,  $G_c$  and  $S$ . We can also see that the steady-state is unique for any values of  $D_{act}$ ,  $G_c$  and  $S$ .

### Stability analysis

We determine whether this steady-state solution is stable or not by linearising the system; that is, by investigating whether small perturbations from the steady-state concentrations of p27, cyclin E-CDK2 and cyclin E-CDK2-p27 will continue to ‘move away’ from the steady-state solution (resulting in an unstable point), or whether small perturbations from the steady-state solution will tend to move back towards the steady-state solution (stable point). We do this by investigating the rate of change in concentration of each protein when we consider small perturbations from the steady-state solution. We define

the perturbations from the steady-state solution as

$$\begin{aligned}\bar{P}(t) &= P + \tilde{P}(t), \\ \bar{E}(t) &= E + \tilde{E}(t), \\ \bar{P}_E(t) &= P_E + \tilde{P}_E(t),\end{aligned}$$

where  $\tilde{P}(t)$ ,  $\tilde{E}(t)$  and  $\tilde{P}_E(t)$  are considered to be very small.  $\bar{P}(t)$ ,  $\bar{E}(t)$  and  $\bar{P}_E(t)$  are the values of p27, cyclin E-CDK2 and cyclin E-CDK2-p27 at time  $t$  when perturbed from the steady-state values of  $P$ ,  $E$  and  $P_E$  respectively. The rate of change in the concentration of  $\bar{P}(t)$  at this perturbed point can be expressed as

$$\begin{aligned}\frac{d\bar{P}(t)}{dt} &= \alpha_p - \alpha_{pe}(E + \tilde{E}(t))(P + \tilde{P}(t)) + \beta_{pe}(P_E + \tilde{P}_E(t)) - \beta_p(P + \tilde{P}(t)), \\ &\approx (\alpha_p - \alpha_{pe}EP + \beta_{pe}P_E - \beta_pP) - \alpha_{pe}(E\tilde{P}(t) + P\tilde{E}(t)) + \beta_{pe}\tilde{P}_E(t) - \beta_p\tilde{P}(t),\end{aligned}\tag{8.3}$$

where we ignore the second order term  $\tilde{E}(t)\tilde{P}(t)$ , as we assume its value is so small as to be negligible. The first term on the RHS of the equation (8.3) (i.e. the expression within the first set of brackets) must be zero due to the steady-state equations, and so we can write

$$\frac{d\bar{P}(t)}{dt} \approx -\alpha_{pe}(E\tilde{P}(t) + P\tilde{E}(t)) + \beta_{pe}\tilde{P}_E(t) - \beta_p\tilde{P}(t).$$

Similarly,

$$\begin{aligned}\frac{d\bar{E}(t)}{dt} &= \alpha_e D_{act} G_c - \alpha_{pe}(E + \tilde{E}(t))(P + \tilde{P}(t)) + \beta_{pe}(P_E + \tilde{P}_E(t)) - \beta_e S(E + \tilde{E}(t)), \\ &\approx (\alpha_e D_{act} G_c - \alpha_{pe}EP + \beta_{pe}P_E - \beta_e SE) - \alpha_{pe}(E\tilde{P}(t) + P\tilde{E}(t)) + \beta_{pe}\tilde{P}_E(t) - \beta_e S\tilde{E}(t), \\ &= -\alpha_{pe}(E\tilde{P}(t) + P\tilde{E}(t)) + \beta_{pe}\tilde{P}_E(t) - \beta_e S\tilde{E}(t), \\ \frac{d\bar{P}_E(t)}{dt} &= \alpha_{pe}(E + \tilde{E}(t))(P + \tilde{P}(t)) - \beta_{pe}(P_E + \tilde{P}_E(t)), \\ &\approx (\alpha_{pe}EP - \beta_{pe}P_E) + \alpha_{pe}(E\tilde{P}(t) + P\tilde{E}(t)) - \beta_{pe}\tilde{P}_E(t), \\ &= \alpha_{pe}(E\tilde{P}(t) + P\tilde{E}(t)) - \beta_{pe}\tilde{P}_E(t).\end{aligned}$$

We can express the above equations in matrix form:

$$\begin{bmatrix} \frac{d\bar{P}(t)}{dt} \\ \frac{d\bar{E}(t)}{dt} \\ \frac{d\bar{P}_E(t)}{dt} \end{bmatrix} \approx \begin{bmatrix} -\alpha_{pe}E - \beta_p & -\alpha_{pe}P & \beta_{pe} \\ -\alpha_{pe}E & -\alpha_{pe}P - \beta_eS & \beta_{pe} \\ \alpha_{pe}E & \alpha_{pe}P & -\beta_{pe} \end{bmatrix} \begin{bmatrix} \tilde{P}(t) \\ \tilde{E}(t) \\ \tilde{P}_E(t) \end{bmatrix}. \quad (8.4)$$

The steady state solution is stable if the eigenvalues of the above matrix have strictly negative real parts. Theoretically, the nature of the steady-state may depend upon the values of the rate coefficients we choose. However, we can show that no matter what rate coefficients we choose (so long as they are positive), the steady-state is stable. We determine the characteristic polynomial by expanding  $\det(\lambda I - A) = 0$  (where  $A$  is the coefficient matrix) as

$$\begin{aligned} \lambda^3 + (\beta_eS + \beta_p + \beta_{pe} + E\alpha_{pe} + P\alpha_{pe})\lambda^2 + (\beta_eS\beta_p + \beta_eS\beta_{pe} + \beta_p\beta_{pe} + E\alpha_{pe}\beta_eS + P\alpha_{pe}\beta_p)\lambda \\ + \beta_eS\beta_p\beta_{pe} = 0, \end{aligned} \quad (8.5)$$

where the solutions,  $\lambda$ , of the above polynomial represent the eigenvalues of the matrix. To determine whether the steady-state is stable, we only need to determine whether the eigenvalues all have negative real part. To do this, we will make use of the Routh Hurwitz stability theorem [47]. For a degree three polynomial

$$\lambda^3 + a\lambda^2 + b\lambda + c,$$

the Routh-Hurwitz stability criteria states that the roots all have negative real part iff

$$a > 0 \quad \text{and} \quad ab > c.$$

If we consider the characteristic polynomial in equation (8.5), then we note that the first criteria is satisfied, as each of the coefficients of our characteristic polynomial are positive. Consider the product of the coefficients of  $\lambda^2$  and  $\lambda$ , and note that the product of these coefficients contains the term  $\beta_eS\beta_p\beta_{pe}$  plus other positive terms, and thus will be greater than the constant term in equation (8.5). Since both of the stability criteria are satisfied, the system is stable for any choice of the rate coefficients.

## 8.2 An interesting relationship between cyclin E-CDK2 and p27

As mentioned earlier, equation (8.2) shows that concentrations of cyclin E-CDK2-p27 are proportional to those of cyclin E-CDK2. However, if we consider Figure 8.1(b), we note that after progestin, cyclin E-CDK2-p27 concentration increases significantly as cyclin E-CDK2 concentrations decrease, a feature we are unlikely to capture with the model described in equations (8.1). It may seem possible that we could capture this feature when the model is moving from one steady-state to another (i.e. in transience), however even this is not possible with our model. This is because to have an increase in  $P_E(t)$  while  $E$  decreases would require a large increase in  $P(t)$ , however from equation (8.1a) that the only way for  $P(t)$  to increase is through an increase in  $E(t)$ , and so we cannot even generate this feature when the model is not in steady-state.

This feature is due to an additional important mechanism - where cyclin E-CDK2 phosphorylates and kills p27 molecules, which we will discuss in the next section. As mentioned earlier, the regulation of p27 concentration is mainly controlled by altering the rate at which the protein is degraded, rather than by altering the rate at which the protein is produced [135, pg. 273]. In this section, we will discuss some of the mechanisms involved in p27 degradation.

As with most proteins, degradation occurs by **phosphorylation** (which is when a phosphate group is added to the targeted protein) followed by **ubiquitination** (a common process for protein degradation). The initial step of phosphorylation allows the protein to be recognised by enzymes which start the process of ubiquitination, eventually resulting in the protein being broken down [135, pg. 242].

In the case of p27, there are several sites at which phosphorylation can cause degradation [124], and each of these sites are phosphorylated by different proteins. However by far the most studied and most influential in regulating p27 protein turnover is phosphorylation at the T187 site.

### Phosphorylation at T187

An experiment performed in [106] in 1997 showed that the very complex that p27 inhibits, cyclin E-CDK2, can phosphorylate p27, causing its degradation by ubiquitination. Results of [106] showed that cyclin E-CDK2 can phosphorylate p27 at a particular site called T187 (denoted with a 'T' because the site contains the amino acid called threonine), which causes its degradation.

This mechanism of p27 phosphorylation by the very protein it inhibits raised an important question - if p27 binds and inactivates cyclin E-CDK2 complexes, then how can an *inactive* cyclin E-CDK2 complex phosphorylate p27? There have been a few different hypotheses put forward to answer this, and we will examine each of these hypotheses in turn.

#### Mechanism proposed by [106] in 1997

The mechanism of p27 phosphorylation by cyclin E-CDK2 was proposed in [106] to occur as follows: p27 is able to bind with cyclin E-CDK2 in 2 ways, forming either a weak or a strong bond. When cyclin E-CDK2 and p27 first interact, they bind weakly. When bound weakly, the complex is not inhibited and the cyclin E-CDK2 component is able to phosphorylate the p27 molecule it is bound to (called *cis* phosphorylation). However, before phosphorylation occurs, it is possible for the cyclin E-CDK2-p27 molecule to transition into a stronger bound complex, which is inactive, and hence phosphorylation of p27 cannot occur once strong binding has taken place. The way the molecules transition into a stronger bond was not described.

#### Mechanism proposed by [74] in 1999

In [74], a different mechanism of p27 phosphorylation at the T187 site was proposed. Phosphorylation of p27 was observed to be the result of free cyclin E-CDK2 complexes. Experiments performed in [74] showed that free p27 could be phosphorylated by active cyclin E-CDK2. The authors also noted that phosphorylation of p27 was not unique to cyclin E-CDK2 complexes - active cyclin A-CDK2 and cyclin B-CDK1 complexes were

also able to phosphorylate p27. However for degradation to occur, the phosphorylated p27 had to be bound to a cyclin E-CDK2 complex. Thus degradation of p27 was proposed to be a two step process: first phosphorylation could occur by a free cyclin E-CDK2 molecule, but for the p27 molecule to be degraded, it must be bound in an inactive cyclin E-CDK2-p27 complex. The authors did not note any changes in binding strength between cyclin E-CDK2 and p27, as was proposed in [106].

### **Mechanism proposed by [140] in 1999**

A third mechanism for the phosphorylation and ubiquitination of p27 was proposed in [140]. In [140], it was shown that active cyclin E-CDK2 complexes (that are not bound to an inhibitor such as p27 or p21) can only phosphorylate p27 molecules that are already bound to cyclin E-CDK2 complexes (called trans-phosphorylation), and cannot phosphorylate free p27 molecules. This disagrees with the results from [74]. They did not investigate the effects of other cell cycle proteins such as cyclin A or cyclin B as was done in [74]. Results from [140] were published only two months after [74], and the results from [74] were not cited in this investigation.

Later studies conducted by the same authors of [74] have confirmed that the mechanism proposed in [140] is correct, and that the observation in [74] that free p27 can be phosphorylated on T187 was not correct, and neither is the hypothesis that cyclin E-CDK2-p27 can exist in two states [Pagano, personal communication]. Hence, we choose to adopt the mechanism of p27 degradation proposed in [140] when modelling p27 degradation.

### **Previous models of p27 and cyclin E in the literature**

In [116], a mathematical model was developed to investigate the mechanisms by which cyclin E caused p27 degradation. At the time of this paper (1999), it was only just becoming apparent that cyclin E phosphorylated (and hence caused the degradation of) p27 molecules, however the exact mechanism by which cyclin E performed this was still being investigated. Thus, different mechanisms of p27 degradation were considered in [116]. The mechanism originally proposed by the authors of [106] was considered (which we just discussed), where cyclin E binds with p27 weakly initially, and then progresses

to a stronger connection with the protein. In this model, cyclin E phosphorylates p27 molecules in the same complex when it is weakly bound, but does not phosphorylate the molecule when the bind is strong. A second mechanism was also investigated in [116]. This mechanism assumes that cyclin E-CDK2 not only causes degradation of free p27 molecules, but also degrades p27 when p27 is bound in a different p27-cyclin E-CDK2 complex. This is similar to the model which we will develop in the next section - the differences being that we do not allow cyclin E-CDK2 to influence the degradation of free p27 molecules, and we only allow cyclin E-CDK2 molecules to phosphorylate p27 molecules that are currently bound in a cyclin E-CDK2-p27 complex. As we discussed earlier, the current consensus is that free p27 molecules are not thought to be phosphorylated by active cyclin E-CDK2 molecules.

The primary purpose of the investigation in [116] was to identify conditions that would lead to an all-or-none release of cyclin E-CDK2 from p27, so that in steady-state, there were either very few active cyclin E-CDK2 molecules, or there were very many active cyclin E-CDK2 molecules. This is what appears to occur in cells: upon entry into the late G1 phase, cyclin E concentrations shoot up and most of these appear to be bound to CDK2 but not to p27 (i.e. most of these appear as active cyclin E-CDK2 molecules). However, in other phases, most cyclin E is found bound in p27 complexes, and hence there is very little free cyclin E-CDK2. We suggest that changes in the amount of free cyclin E-CDK2 concentration (and hence cyclin E activity) is much more related to a cell's entry into the late G1 phase (where cyclin E production rates increase) or its entry into the S phase (where the protein is degraded), rather than to the concentration of p27.

Now that we understand how cyclin E-CDK2 can influence the rate of p27 degradation, we wish to include this mechanism into our model, and investigate its effects.

### 8.3 Including the mechanism of T187 phosphorylation in our model of p27 and cyclin E

In addition to the mechanisms included in the model discussed in Section 8.1 and shown in equations (8.1), we assume that free cyclin E-CDK2 molecules will bind and phosphorylate only the p27 molecules present in a separate cyclin E-CDK2-p27 complex at the rate  $\beta_{ph}$ .

We assume that p27 molecules phosphorylated in this way are instantly removed from the system, and hence the cyclin E-CDK2 molecule that was bound in the cyclin E-CDK2-p27 complex becomes instantly freed upon phosphorylation. We therefore modify equations (8.1) to include this mechanism as follows:

$$\frac{dP(t)}{dt} = \alpha_p - Ebind - \beta_p P(t), \quad (8.6a)$$

$$\frac{dE(t)}{dt} = \alpha_e D_{act}(t) G_c(t) - Ebind - \beta_e S(t) E(t) + \beta_{ph} E(t) P_E(t), \quad (8.6b)$$

$$\frac{dP_E(t)}{dt} = Ebind - \beta_{ph} E(t) P_E(t), \quad (8.6c)$$

$$Ebind = \alpha_{pe} E(t) P(t) - \beta_{pe} P_E(t), \quad (8.6d)$$

where we note that the term  $\beta_{ph} E(t) P_E(t)$  in equation (8.6b) represents the cyclin E-CDK2 complex re-appearing after the p27 molecule was removed from the  $P_E$  complex upon phosphorylation. As the  $P_E$  complex has essentially been converted into a cyclin E-CDK2 complex upon phosphorylation, the term  $\beta_{ph} E(t) P_E(t)$  also appears as a loss in equation (8.6c). Recall that the variables  $D_{act}(t)$ ,  $G_c(t)$  and  $S(t)$  are variables which are not determined by our model system, rather their time-dependent values are inputs to the model. We will call  $D_{act}(t)$ ,  $G_c(t)$  and  $S(t)$  the drivers of the model, as their changing values will result in changes in our model behaviour.

It may seem that the analysis performed in the previous section for the model described in equations (8.1) was unnecessary. However, by comparing the model results that we will find using equations (8.6) with those from (8.1), we will be able to identify how the additional terms representing cyclin E-CDK2-induced death of p27 influences model behaviour. In addition, we will later find in Section 8.6 that equations (8.1) have the same functional form as the model we develop for cyclin D-CDK4 and p27, and so the earlier analysis of equations (8.1) will be relevant then as well.

### Nondimensionalisation

We wish to nondimensionalise equations (8.6) with respect to the starting concentration of cyclin E-CDK2,  $E(0)$ . We do this because we can determine the concentrations of other proteins relative to  $E(0)$  using the elution profiles from Section 7.4.

Letting  $E(t) = \bar{E}(t)E(0)$ ,  $P(t) = \bar{P}(t)E(0)$  and  $P_E(t) = \bar{P}_E(t)E(0)$ , we write

$$\begin{aligned}\frac{dP(t)}{dt} &= \frac{d\bar{P}(t)E(0)}{dt} = \alpha_p - Ebind - \beta_p\bar{P}(t)E(0), \\ \frac{dE(t)}{dt} &= \frac{d\bar{E}(t)E(0)}{dt} = \alpha_e D_{act}(t)G_c(t) - Ebind - \beta_e S(t)\bar{E}(t)E(0) + \beta_{ph}\bar{E}(t)E(0)\bar{P}_E(t)E(0), \\ \frac{dP_E(t)}{dt} &= \frac{d\bar{P}_E(t)E(0)}{dt} = Ebind - \beta_{ph}\bar{E}(t)E(0)\bar{P}_E(t)E(0), \\ Ebind &= \alpha_{pe}\bar{E}(t)E(0)\bar{P}(t)E(0) - \beta_{pe}\bar{P}_E(t)E(0),\end{aligned}$$

which we can simplify to

$$\begin{aligned}\frac{d\bar{P}(t)}{dt} &= \bar{\alpha}_p - \overline{Ebind} - \beta_p\bar{P}(t), \\ \frac{d\bar{E}(t)}{dt} &= \bar{\alpha}_e D_{act}(t)G_c(t) - \overline{Ebind} - \beta_e S(t)\bar{E}(t) + \bar{\beta}_{ph}\bar{E}(t)\bar{P}_E(t), \\ \frac{d\bar{P}_E(t)}{dt} &= \overline{Ebind} - \bar{\beta}_{ph}\bar{E}(t)\bar{P}_E(t), \\ \overline{Ebind} &= \bar{\alpha}_{pe}\bar{E}(t)\bar{P}(t) - \beta_{pe}\bar{P}_E(t),\end{aligned}$$

where  $\bar{\alpha}_p = \frac{\alpha_p}{E(0)}$ ,  $\bar{\alpha}_e = \frac{\alpha_e}{E(0)}$ ,  $\bar{\beta}_{ph} = \beta_{ph}E(0)$  and  $\bar{\alpha}_{pe} = \alpha_{pe}E(0)$ . This system now has the same form as the original system in equations (8.6), except that most of the rate coefficients take on different values. The rate coefficients that have not changed after nondimensionalisation are  $\beta_e$ ,  $\beta_p$  and  $\beta_{pe}$ . It is not important to us that the rate coefficients are non-dimensionalised, as we will be determining most of the values of the rate coefficients using the non-dimensionalised system anyway. The only rate coefficients which we determine directly from literature are  $\beta_e$  and  $\beta_p$ , which, as we can see, have not changed value after non-dimensionalising. Thus, for notational simplicity, we wish to drop the bars above the variables and rate coefficients. We consider the set of equations identical to that shown in equations (8.6), keeping in mind that we have nondimensionalised our system by expressing  $P(t)$ ,  $E(t)$  and  $P_E(t)$  relative to the value  $E(0)$ .

### Steady-state for the model in equation (8.6)

When  $D_{act}(t)$ ,  $G_c(t)$  and  $S(t)$  are unchanging (which we call  $D_{act}$ ,  $G_c$  and  $S$  respectively), then the system can reach steady-state, and the steady-state concentrations of each pro-

Table 8.1: *Unknown parameters from equation (8.6)*

| Parameter     |
|---------------|
| $\beta_e$     |
| $\beta_p$     |
| $\alpha_e$    |
| $\beta_{ph}$  |
| $\beta_{pe}$  |
| $\alpha_{pe}$ |
| $\alpha_p$    |
| $D_{act}(0)$  |
| $G_c(0)$      |
| $S(0)$        |
| $E(0)$        |
| $P(0)$        |
| $P_E(0)$      |

tein can be expressed as

$$P = \alpha_p \frac{\beta_{pe} + \beta_{ph}E}{\alpha_{pe}E^2\beta_{ph} + \beta_{pe}\beta_p + \beta_p\beta_{ph}E}, \quad (8.9a)$$

$$E = \frac{\alpha_e D_{act} G_c}{\beta_e S}, \quad (8.9b)$$

$$P_E = \frac{\alpha_{pe} EP}{\beta_{pe} + \beta_{ph}E}. \quad (8.9c)$$

This steady-state is unique for any given  $D_{act}$ ,  $G_c$  and  $S$ . This can be seen by noting that equation (8.9b) is an explicit equation for  $E$ . As  $P$  is expressed directly in terms of  $E$ ,  $P$  must also be unique. It also follows that  $P_E$  is unique, and hence there is only one steady-state for the system.

### Determining the rate parameters and starting conditions of our model

We aim to determine the values of the unknown parameters in our model, shown in Table 8.1. We can determine the values of  $\beta_e$  and  $\beta_p$  from the literature. In [130], it was noted that p27 had a half-life of about 6 hours in T47D cells. In [139] it was noted that for cells in the S phase, cyclin E-CDK2 had a half-life of 0.5 hours. These half life values can be converted to rates of decay, giving  $\beta_p = 1.1\text{hr}^{-1}$  and  $\beta_e = 1.4\text{hr}^{-1}$ .

We cannot determine the remaining rate variables from the literature. However, if

we assume that prior to progestin exposure, protein concentrations were in steady-state (this is reasonable, since we already know that the phase steady-state was reached before progestin was added), then we can use the starting concentrations of the proteins from [84] to determine more rate parameters.

We can determine the starting concentrations of proteins using data from Section 7.4. As we have non-dimensionalised with respect to  $E(0)$ , we set  $E(0) = 1$ . From Figure 7.14 in Section 7.4, we see that  $P_E(0)$  is about one fifth that of  $E(0)$ , and so we choose  $P_E(0) = 0.2$ . Authors of [84] note that  $P(0)$  is undetectable, however elution blot profiles from [113] show that  $P(0)$  is present in very small quantities, which we estimate to represent about 1% of total p27 in control cells, and so we choose  $P(0) = \frac{1}{100}$ .

We can determine the starting cell cycle phase proportions using the data from [84], and also using results from our cell cycle model from Chapter 6. We know that  $S(0) = 0.27$  from Figure 7.1. Using this value for  $S(0)$ , we use our cell cycle model from equations (3.7), as well as the parameters unique to the T47D cell line discussed in Chapter 6 to determine  $G_c(0) = 0.047$  (calculation details not shown). As mentioned in the previous section, we will assume that  $D_{act}(0) = 1$ .

Substituting the starting values of our variables into equations (8.9) gives

$$\frac{1}{100} = \alpha_p \frac{\beta_{pe} + \beta_{ph}}{\alpha_{pe}\beta_{ph} + 1.1\beta_{pe} + 1.1\beta_{ph}}, \quad (8.10a)$$

$$1 = \frac{\alpha_e \times 0.047}{1.4 \times 0.27}, \quad (8.10b)$$

$$0.2 = \frac{\alpha_{pe} \times \frac{1}{100}}{\beta_{pe} + \beta_{ph}}. \quad (8.10c)$$

Equation (8.10b) gives  $\alpha_e = 0.12$ . We are then left with two non-linear equations with 4 unknowns, and so we have at least two degrees of freedom. However, by considering equation (8.10c), we note that if we know  $\beta_{pe}$  and  $\beta_{ph}$  then we can rearrange to determine  $\alpha_{pe}$ . We will then be able to determine  $\alpha_p$  from equation (8.10a) as all other rate variables will have been determined. Thus, we only need to determine the values of  $\beta_{pe}$  and  $\beta_{ph}$ , and all the rate variables in our system can then be calculated. The values of the rate variables and starting values we have determined so far are summarised in Table 8.2.

In the next section we will investigate the stability of our model system, and use the

Table 8.2: *Values of parameters we have determined*

| Parameter     | value                                        |
|---------------|----------------------------------------------|
| $\beta_e$     | 1.4                                          |
| $\beta_p$     | 1.1                                          |
| $\alpha_e$    | 0.12                                         |
| $\beta_{ph}$  | unknown                                      |
| $\beta_{pe}$  | unknown                                      |
| $\alpha_{pe}$ | dependent upon $\beta_{ph}$ and $\beta_{pe}$ |
| $\alpha_p$    | dependent upon $\beta_{ph}$ and $\beta_{pe}$ |
| $D_{act}(0)$  | 1                                            |
| $G_c(0)$      | 0.047                                        |
| $S(0)$        | 0.27                                         |
| $E(0)$        | 1                                            |
| $P(0)$        | 0.01                                         |
| $P_E(0)$      | 0.2                                          |

results of the analysis to determine a viable range of values that the parameters  $\beta_{pe}$  and  $\beta_{ph}$  can take.

### Stability analysis

Now that we know the values of most of the rate variables, we wish to investigate the stability of the steady-state (shown in equations (8.9)). We do this by performing a linear stability analysis. We define

$$\begin{aligned}\bar{P}(t) &= P + \tilde{P}(t), \\ \bar{E}(t) &= E + \tilde{E}(t), \\ \bar{P}_E(t) &= P_E + \tilde{P}_E(t).\end{aligned}$$

After some algebra (not shown, but similar to the process used in Section 8.1), we can express the linear-stability system in matrix form as follows:

$$\begin{bmatrix} \frac{d\bar{P}(t)}{dt} \\ \frac{d\bar{E}(t)}{dt} \\ \frac{d\bar{P}_E(t)}{dt} \end{bmatrix} = \begin{bmatrix} -\alpha_{pe}E - \beta_p & -\alpha_{pe}P & \beta_{pe} \\ -\alpha_{pe}E & -\alpha_{pe}P + \beta_{ph}P_E - \beta_eS & \beta_{pe} + \beta_{ph}E \\ \alpha_{pe}E & \alpha_{pe}P - \beta_{ph}P_E & -\beta_{pe} - \beta_{ph}E \end{bmatrix} \begin{bmatrix} \tilde{P}(t) \\ \tilde{E}(t) \\ \tilde{P}_E(t) \end{bmatrix}. \quad (8.11)$$

Unlike in the previous section, we find that the stability depends upon the choice of the rate variables for the system. In the simple model we considered earlier, we only needed to know that all of the model variables were positive (which they were) and we could prove that the system was stable for any choice of rate parameters. However, the model presented in this section is not stable for all rate parameters, and we must ensure that the model is stable for a realistic range of environmental conditions. If we consider the matrix in equation (8.11), we note that to investigate the stability of our system, we must know the steady-state values  $S, P, E$  and  $P_E$ . As the steady-state values of  $P$  and  $P_E$  can be expressed in terms of  $E$ , then we really have two unknown variables -  $E$  and  $S$ . We can see from equation (8.9b) that  $E$  and  $S$  are also related, however there is an additional unknown in this equation -  $D_{act}$ . Thus, we need to know both the value of  $E$  and the corresponding value of  $S$  to investigate stability of this system.

Thus, we will consider stability for two steady-state point values that we do know. As we have just discussed, we are making the assumption that initially,  $E = 1, P = \frac{1}{100}, P_E = 0.2$  and  $S = 0.27$ , and so we can investigate stability for these initial conditions. Now after 30 hours of progestin exposure, the S phase proportion reaches a steady-state of about 0.11 (see Figure 7.1), and thus we know the value of  $S$  here. This allows us to determine the value for  $G_c(t)$ , which we calculate to be 0.017. If we also assume that  $D_{act}$  does not change from its value of 0.1 after 30 hours of progestin exposure (as justified by Figure 7.3), then we can calculate the eventual steady-state value of  $E$  using equation (8.9b). This gives  $E = 0.037$ . Substituting this value of  $E$  into equation (8.9a) and (8.9c) provides us the values for  $P$  and  $P_E$  in terms of the remaining undetermined rate variables, giving us another set of values with which we can investigate stability. Thus, we have two steady-state situations where we can investigate stability, one representing a highly proliferative system ( $E = 1$ ) and the other representing a growth arrested situation ( $E = 0.037$ ). We note that we cannot truly investigate the steady-state of these two states, because perturbations from these steady-states may also cause perturbations in the value of  $S$  (which we do not consider here), however we can still get a good indication of stability at these points.

We will investigate the stability of these two critical points for varying values of  $\beta_{pe}$  and  $\beta_{ph}$ . In particular, we will determine what values of  $\beta_{pe}$  and  $\beta_{ph}$  result in stable critical points. We do this by calculating the eigenvalues of the matrix shown in equation (8.11)

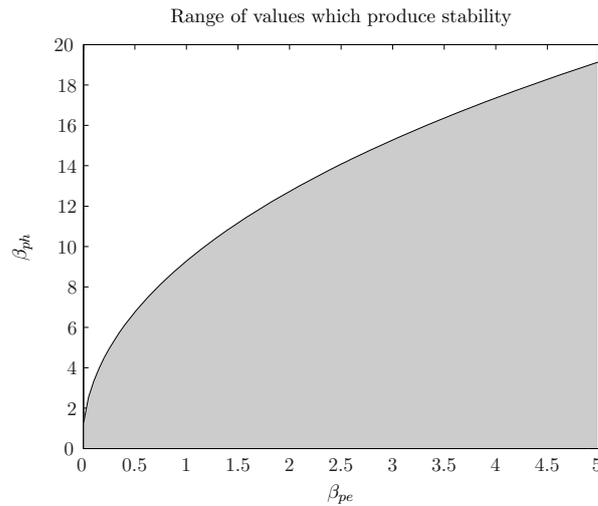


Figure 8.2: *The shaded area represents the values of  $\beta_{pe}$  and  $\beta_{ph}$  that result in a stable critical point when  $E = 0.037$ .*

for varying values of  $\beta_{pe}$  and  $\beta_{ph}$ . We find that in the case  $E = 1$ , the range of  $\beta_{pe}$  and  $\beta_{ph}$  values that produce stability (i.e. eigenvalues with negative real part) is very large. However, when  $E = 0.037$ , the range of  $\beta_{pe}$  and  $\beta_{ph}$  values that produce stability is much smaller. In addition, the range of  $\beta_{pe}$  and  $\beta_{ph}$  that produce stability when  $E = 1$  encompasses the values of  $\beta_{pe}$  and  $\beta_{ph}$  that produce stability when  $E = 0.037$ . Thus we will only consider results for the steady-state with  $E = 0.037$ .

The ranges of stability for varying values of  $\beta_{pe}$  and  $\beta_{ph}$  are shown in Figure 8.2. We can see from this representation that as  $\beta_{pe}$  increases, our range of possible values of  $\beta_{ph}$  increases. Although not obvious in this diagram, when  $\beta_{pe} = 0$ , our model system is stable for  $\beta_{ph} \leq 1.25$ . This representation gives us some restrictions on the values that  $\beta_{pe}$  and  $\beta_{ph}$  can take when simulating protein concentration changes after progestin exposure. By choosing different values of  $\beta_{pe}$  and  $\beta_{ph}$  within the ranges of stability shown in Figure 8.2, we will simulate our protein model and determine which values of  $\beta_{pe}$  and  $\beta_{ph}$  produce a good fit to the data on protein concentration changes after progestin exposure from [84]. As we discussed when we introduced the model, the addition of progestin will influence our model through the drivers  $D_{act}(t)$ ,  $G_C(t)$  and  $S(t)$ . Thus, before we can simulate our model to determine the values of  $\beta_{pe}$  and  $\beta_{ph}$ , we must first determine how progestin influences the drivers of our model, and we will do this in the next section.

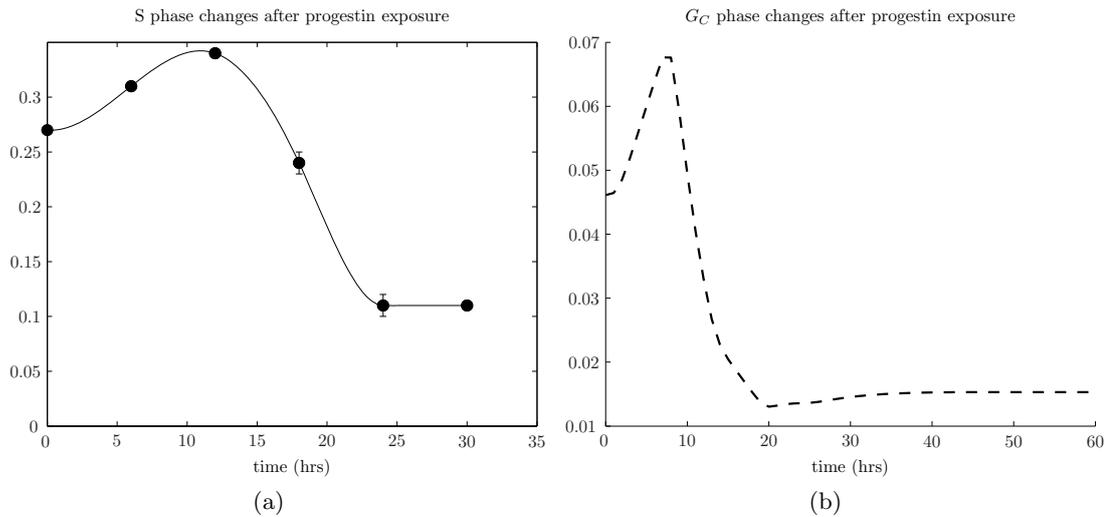


Figure 8.3: (a) The S phase changes after progestin exposure from Chapter 6. The experimental data from [84] on observed S phase changes is shown as the black dots. (b) The  $G_c$  phase proportion after T47D cells are exposed to progestin, calculated using our cell cycle model from equations (3.7).

### Determining the drivers of our system, $D_{act}(t)$ , $G_c(t)$ and $S(t)$

In order to simulate protein concentration changes after progestin exposure, we must also know how the drivers of our model in equations (8.6) change with time. Once we determine these values, we can simulate our model and compare our results with experimental data.

Data on the changes in  $S(t)$  after progestin exposure are provided for discrete time points in [84] (reproduced in Figure 7.1 of this thesis). In order to determine a continuous representation of  $S(t)$ , we can interpolate the data points representing experimentally observed changes in the S phase proportion after progestin exposure from [84], shown in Figure 8.3(a).

There is no data on the changes in  $G_c(t)$  after progestin exposure, however we can use the S phase profile from Figure 8.3(a) to generate the  $G_c(t)$  changes after progestin exposure using our cell cycle model from Chapter 6. Doing this produces the resultant  $G_c$  phase profile shown in Figure 8.3(b).

We now wish to determine a continuous profile for  $D_{act}(t)$  after progestin exposure. We have discrete points representing the change in cyclin D-CDK4 activity from [84], shown in Figure 7.6. However, as previously discussed, the second data point, which represents

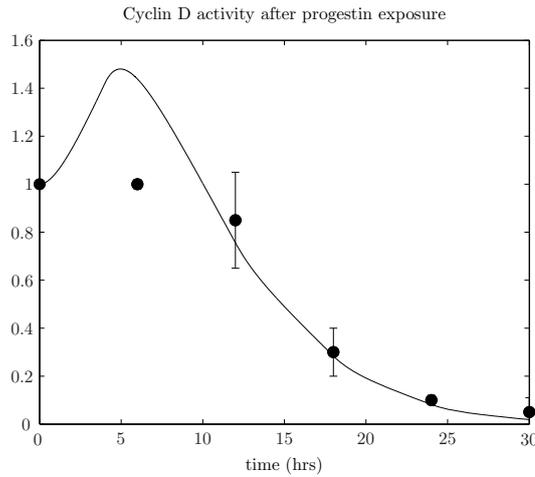


Figure 8.4: *Cyclin D activity after progestin exposure.* The dots represent experimentally observed values from [84]. The black curve represents our chosen profile for  $D_{act}(t)$ .

cyclin D-CDK4 activity at 6 hours after progestin, appears unchanged from control levels. Yet when we inverse-solved for cyclin D-CDK4 activity in Chapter 6, we found that we needed cyclin D-CDK4 activity to increase by 50% of control by 6 hours after exposure, otherwise we could not obtain the observed increase in the rate of S phase entry. For this reason, we will use the inverse-solved profile for cyclin D-CDK4 activity from Chapter 6, which we emphasise only differs from the data points presented in [84] at the point at 6 hours after progestin exposure (for which no error bars are provided). The continuous profile is shown Figure 8.4.

We now have a continuous representation of the changes in our drivers after progestin exposure. We can perform a quick check of these drivers, by investigating whether they will cause the decrease in concentration of free cyclin E-CDK2 molecules that are observed after progestin exposure. If we make the assumption that the steady-state value of  $E$  is reached fairly quickly relative to changes in  $D_{act}(t)$ ,  $G_c(t)$  and  $S(t)$  (i.e. we assume a separation of timescales on  $E$ ), then  $E(t)$  can be written as

$$E(t) = \frac{\alpha_e D_{act}(t) G_c(t)}{\beta_e S(t)}. \quad (8.12)$$

Since we are only interested in the *relative* change in  $E(t)$  after progestin exposure we do not need to know the values of  $\alpha_e$  and  $\beta_e$ . Substituting the continuous representations of  $D_{act}(t)$  and  $S(t)$  into equation (8.12) gives  $E(t)$  as shown in Figure 8.5. The known changes

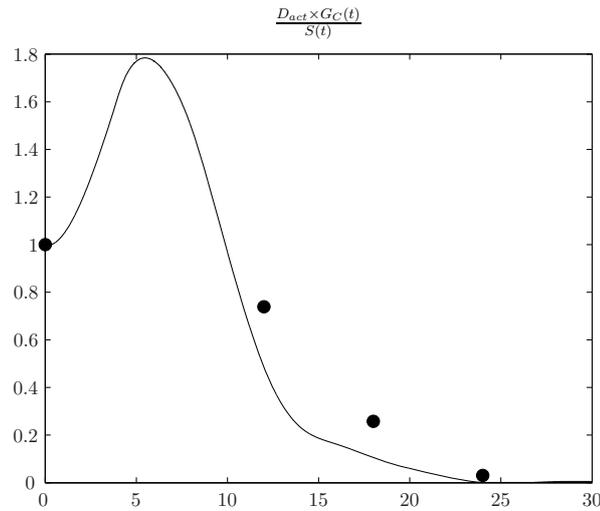


Figure 8.5: The relative concentration of  $E(t)$  from equation (8.12) using the drivers for  $G_c(t)$ ,  $S(t)$  and  $D_{act}(t)$ . The known changes in  $E(t)$  using the elution profiles from Figure 8.1 are shown here as the dots.

in  $E(t)$  using the elution profiles from Figure 8.1 are shown as dots for comparison. We can see that the modelled cyclin E-CDK2 concentration compares reasonably well with the limited experimental data. This is encouraging, as the cyclin E data was generated using the elution profiles from [84], however the modelled curve relies on both raw data from [84], and on results from our cell cycle model from equations (3.7).

Even though the cyclin E curve compares reasonably well with the data points, we note that the predicted cyclin E-CDK2 concentration changes occur on a faster timescale than what is observed experimentally. This is likely due to the fact that we have assumed that cyclin E-CDK2 concentration reaches steady-state quickly. However, as we will see later in Section 8.4, our modelled changes in cyclin E-CDK2 concentration will take a little longer to reach steady-state, and will compare even better with experimental data.

## 8.4 Modelling the experiment from [84]

In this section, we will use the drivers of our model to simulate changes in cyclin E-CDK2, cyclin E-CDK2-p27 and p27 concentration after progestin exposure. The experimental data on protein-protein concentration changes after progestin exposure was discussed in Section 7. Recall that we managed to summarise their results on protein-protein complex

changes, shown in Figure 8.1. We will compare our model results on protein concentration changes to the data presented in Figure 8.1.

### Model results and final parameterisation of $\beta_{pe}$ and $\beta_{ph}$

To simulate the model after exposure to progestin, we substitute the values of the drivers  $S(t)$  and  $D_{act}(t)$  as shown in Figure 8.3 and 8.4 into equations (8.6), and compare the changes in protein concentrations with those determined experimentally. We still have not determined the values of  $\beta_{pe}$  and  $\beta_{ph}$ , and so we will choose these values so that the model results compare with experimental data. We would normally do an automated best fit to data in order to determine the optimal values of unknown model parameters, however due to the limited data available, we find that this is not a reliable approach, and instead we will explore the model results for different values of  $\beta_{pe}$  and  $\beta_{ph}$  in order to determine accurate values of these variables.

Lets begin by considering small values of both  $\beta_{pe}$  and  $\beta_{ph}$ . If we choose  $\beta_{pe} = 1$  and  $\beta_{ph} = 1$  and substitute the known changes in  $D_{act}(t)$ ,  $G_C(t)$  and  $S(t)$  to drive equations (8.6), we can model the changes in cyclin E-CDK2 and p27 concentration, as shown in Figures 8.6(a) and (b). In Figure 8.6(a), the light shaded region represents changes in cyclin E-CDK2 concentration, and the darker shaded region represents changes in cyclin E-CDK2-p27 concentration after progestin exposure. The two shaded regions are placed one on another so that the entire shaded region represents changes in  $E(t) + P_E(t)$  after progestin exposure. Experimental data is also shown in this figure, with the hollow blue dots showing observed changes in free cyclin E-CDK2 after progestin, and the solid black circles show the experimental data on changes in  $E(t) + P_E(t)$  after progestin (the data we are using was shown in Figure 8.1). We can see here that total cyclin E-CDK2 concentration increases initially, due to the increase in cyclin D-CDK4 activity, and subsequently declines. The modelled changes in cyclin E-CDK2 and cyclin E-CDK2-p27 concentration do not compare well with experimental data after 12 hours of progestin exposure (compare hollow blue dots and solid black circles). In Figure 8.6(b), the modelled change in free p27 is shown as the light shaded region, and the modelled change in cyclin E-CDK2-p27 concentration is shown as the darker shaded region. Experimental data on changes in total p27 are shown as the solid black dots. As we can see, choosing  $\beta_{pe} = 1$  and  $\beta_{ph} = 1$  does

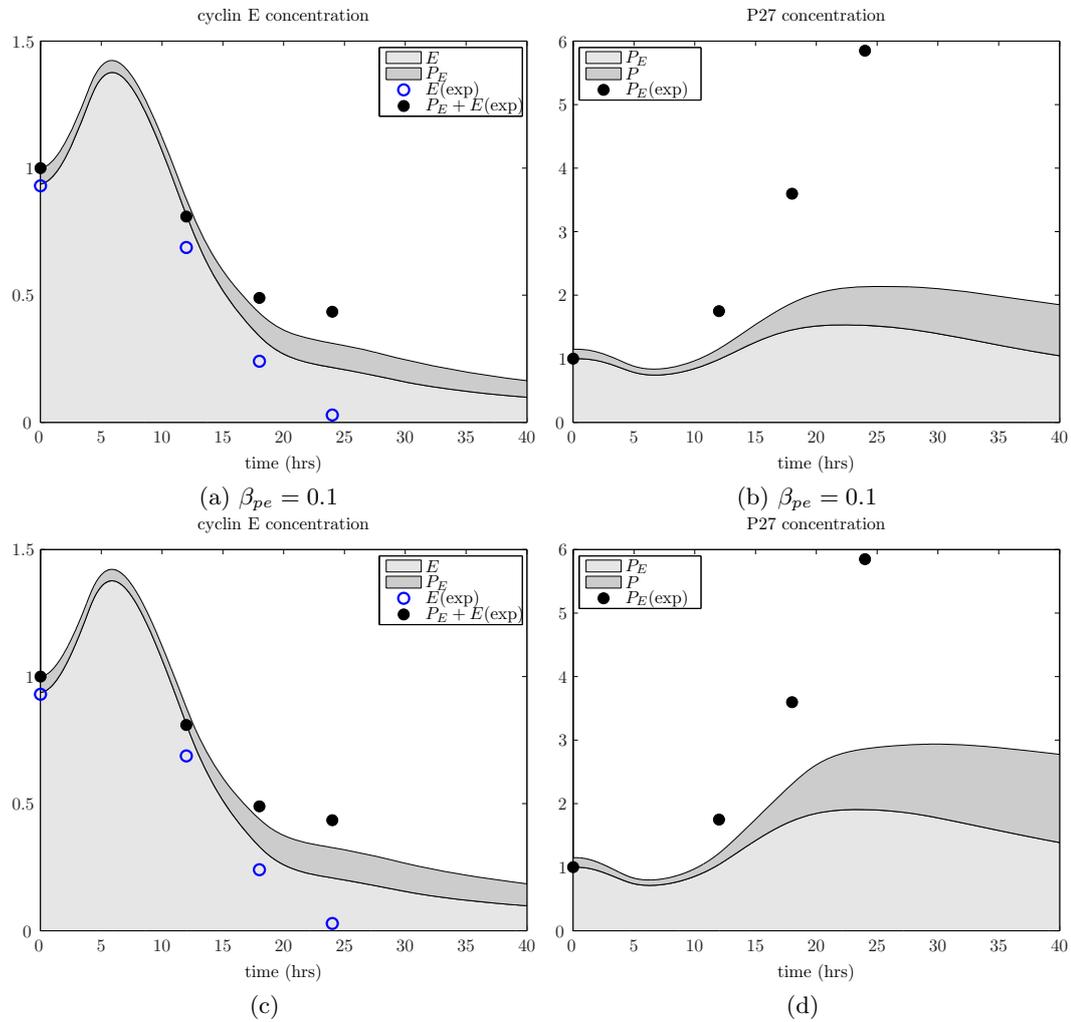


Figure 8.6: Figures (a) and (b) represent the changes in cyclin E-CDK2 and p27 concentration respectively when  $\beta_{pe} = 1$  and  $\beta_{ph} = 1$ . Figures (c) and (d) represent the changes in cyclin E-CDK2 and p27 concentration respectively when  $\beta_{pe} = 4$  and  $\beta_{pe} = 4$ . The hollow blue circles in (a) and (c) represent experimental data on  $E(t)$  concentration after progestin. The solid black circles in (a) and (c) represent experimental data on  $E(t) + P_E(t)$  after progestin exposure. In (b) and (d) the solid black circles represent  $P_E(t)$  after progestin exposure.

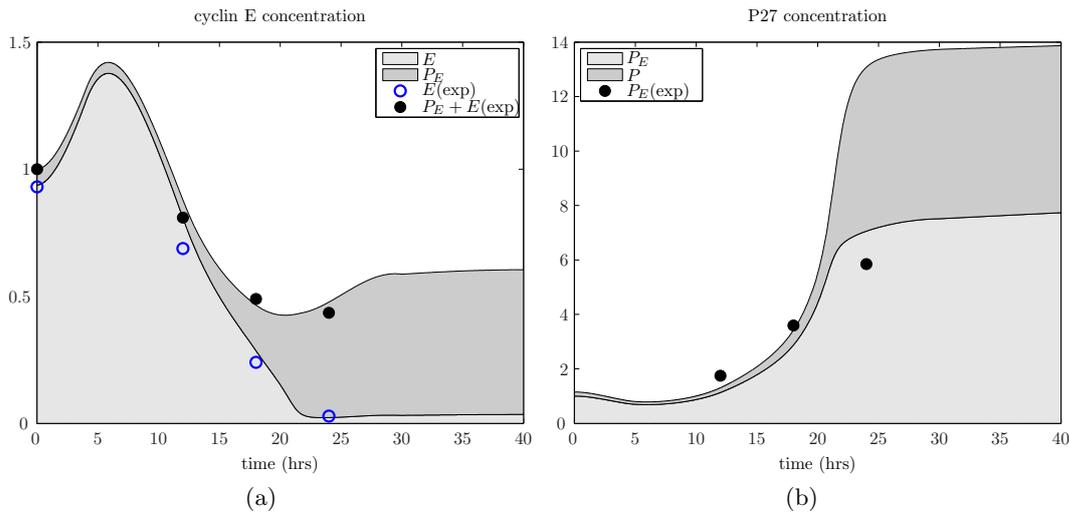


Figure 8.7: *The model results after progestin exposure when  $\beta_{pe} = 1$  and  $\beta_{ph} = 10$ .*

not produce changes in cyclin E-CDK2, cyclin E-CDK2-p27 or free p27 that compare well with experiments. If we also consider larger values of  $\beta_{pe}$  and  $\beta_{ph}$ , but still well within the ranges of stability (for instance, we choose  $\beta_{pe} = 4$  and  $\beta_{ph} = 4$ ), then our model produces results as shown in Figures 8.6(c) and (d). We can see that when we choose  $\beta_{pe} = 4$  and  $\beta_{ph} = 4$ , the model results also do not compare at all well with the experimental data. If we consider values of  $\beta_{pe}$  and  $\beta_{ph}$  closer to the edges of the stability region, for example we choose  $\beta_{pe} = 1$  and  $\beta_{ph} = 10$ , then our model produces results as shown in Figure 8.7. The simulation results shown when we choose  $\beta_{pe} = 1$  and  $\beta_{ph} = 10$  compare well with experimental observations. The changes in free cyclin E-CDK2 concentration compare well with experimental observations (hollow blue circles). Consequently, the changes in cyclin E-CDK2-p27 concentration also compare reasonably well with experiments.

We find that if we simulate our model for other values of  $\beta_{pe}$  and  $\beta_{ph}$  that are near the top edge of the stability region shown in Figure 8.2, our model produces results similar to those shown in Figure 8.7. That is, cyclin E concentrations decrease relatively quickly and p27 concentrations increase fast enough so that simulated results compare well with experimental data. However, the further away from this edge that we move, the more our model results look like those shown in Figure 8.6. Thus, we conclude that to produce realistic changes in protein concentrations after exposure to progestin, we must choose values of  $\beta_{pe}$  and  $\beta_{ph}$  that are near the edges of stability.

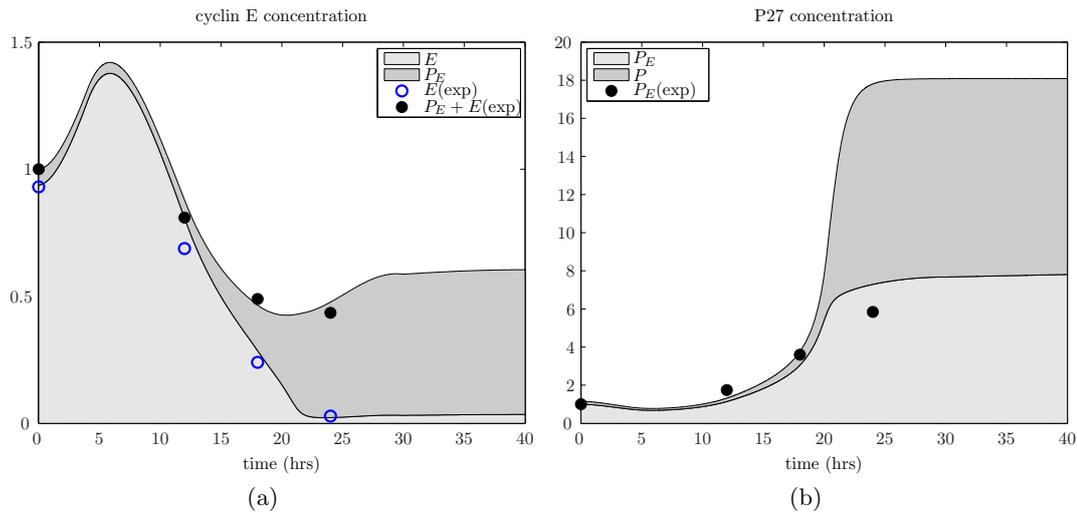


Figure 8.8: *The model results after progestin exposure when  $\beta_{pe} = 5$  and  $\beta_{ph} = 18$ .*

If we restrict our model parameters  $\beta_{pe}$  and  $\beta_{ph}$  to be values near the edge of stability, then we must still decide where on this edge will produce the best results. Figure 8.8 shows model results when we choose  $\beta_{pe} = 5$  and  $\beta_{ph} = 18$ , still near the edge of stability but with overall larger values. As we can see, the higher values of  $\beta_{pe}$  and  $\beta_{ph}$  result in a higher concentration of free p27 at times after 24 hours. We know that the concentration of total p27 must be about 2-fold higher than control levels by 30 hours, however this included cyclin D-CDK4-p27 molecules. From Figure 8.1, we note that cyclin D-CDK4-p27 concentration has not changed significantly in the first 24 hours after progestin exposure, and if we assume that cyclin D-CDK4-p27 concentration remains unchanged until 30 hours, then to get a 2-fold increase in total p27 by 30 hours would require a 14-fold increase in the concentration of  $P(t) + P_E(t)$ . From Figure 8.7, we note this occurs when we choose  $\beta_{pe} = 1$  and  $\beta_{ph} = 10$ . Therefore, from this point on we will choose these values for the rate variables. The full list of parameters is shown in Table 8.3.

### Steady-state analysis of our model

In this section, we will do some more steady-state analysis, as we find that it can provide further insight into this model system, and explain some features observed in experiments that are otherwise difficult to explain.

To investigate how the steady-state concentration of p27 and cyclin E-CDK2 molecules

Table 8.3: Values of the rate variables and starting concentrations

| Parameter     | value |
|---------------|-------|
| $\beta_e$     | 1.4   |
| $\beta_p$     | 1.1   |
| $\alpha_e$    | 0.12  |
| $\beta_{ph}$  | 10    |
| $\beta_{pe}$  | 1     |
| $\alpha_{pe}$ | 220   |
| $\alpha_p$    | 2.21  |
| $D_{act}(0)$  | 1     |
| $G_c(0)$      | 0.047 |
| $S(0)$        | 0.27  |
| $E(0)$        | 1     |
| $P(0)$        | 0.01  |
| $P_E(0)$      | 0.2   |

change with changing values of  $E$ , we will consider how the steady-state values of  $P + P_E$  and  $E + P_E$  change with changing value of  $E$ . To do this, we begin by determining the stationary points, if any, of  $P + P_E$  and  $E + P_E$  as functions of  $E$ .

### The steady-state value of $P + P_E$ in terms of $E$

From equations (8.9), we write

$$\begin{aligned}
 P + P_E &= \alpha_p \frac{\beta_{pe} + \beta_{ph}E}{\alpha_{pe}E^2\beta_{ph} + \beta_{pe}\beta_p + \beta_p\beta_{ph}E} \left(1 + \frac{\alpha_{pe}E}{\beta_{pe} + \beta_{ph}E}\right), \\
 &= \alpha_p \frac{\beta_{pe} + \beta_{ph}E + \alpha_{pe}E}{\alpha_{pe}E^2\beta_{ph} + \beta_{pe}\beta_p + \beta_p\beta_{ph}E}.
 \end{aligned} \tag{8.13}$$

The derivative of the above expression with respect to  $E$  is shown below.

$$\frac{d(P + P_E)}{dE} = -\alpha_p\alpha_{pe} \frac{(\alpha_{pe}\beta_{ph} + \beta_{ph}^2)E^2 + 2\beta_{pe}\beta_{ph}E - \beta_{pe}\beta_p}{\alpha_{pe}E^2\beta_{ph} + \beta_{pe}\beta_p + \beta_p\beta_{ph}E}. \tag{8.14}$$

The zeroes of this expression are the values of  $E$  which produce a zero value on the numerator. We would expect up to two stationary points because we have a quadratic on the numerator, however we can also see that we would have at most one positive stationary point, because the constant term is negative, implying that if there are two real roots, then one would be negative and one positive. The zeroes of this equation

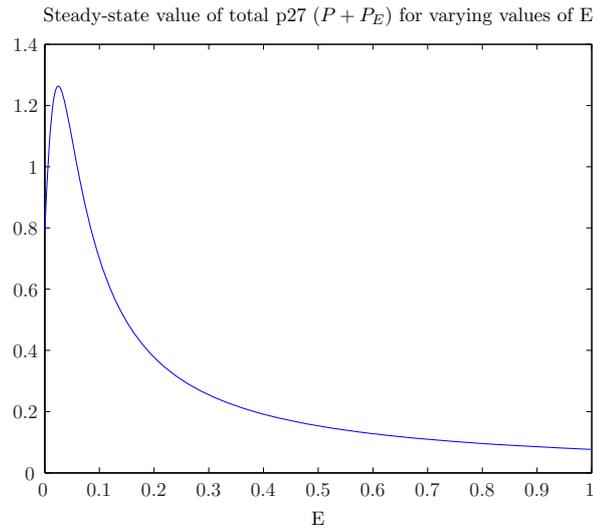


Figure 8.9: *The steady-state value of  $P + P_E$  (total p27) for varying values of  $E$  (free cyclin  $E$ ).*

depend heavily upon the choice of rate parameters, however for the values chosen earlier (see Table 8.3), we can calculate the changes in total p27 for varying values of  $E$  as shown in Figure 8.9. The single positive stationary point can be seen at  $E = 0.025$ . We note that for the values of  $E$  we consider in our simulations ( $E > 0.037$ ), total p27 decreases for increasing  $E$ . This is in contrast to the simple model of cyclin E and p27, where we found that increasing cyclin E-CDK2 concentration increased the overall concentration of p27. This is because the cyclin E-CDK2-p27 molecules are not free from degradation when  $\beta_{ph} \neq 0$ . Instead, they are subject to decay based on the concentration of active cyclin E-CDK2 molecules.

We can see this behaviour in the simulations of our protein model, shown in Figure 8.7. As cyclin E-CDK2 concentrations decreased after progestin exposure, cyclin E-CDK2-p27 concentrations subsequently increased, as did free p27, resulting in a significant increase in total p27 in parallel with the decrease in free cyclin E-CDK2 concentrations.

### **The steady-state value of $E + P_E$ in terms of $E$**

We can also determine how total cyclin E concentration varies for varying values of  $E$  (the concentration of free cyclin E-CDK2). The total steady-state concentration of cyclin

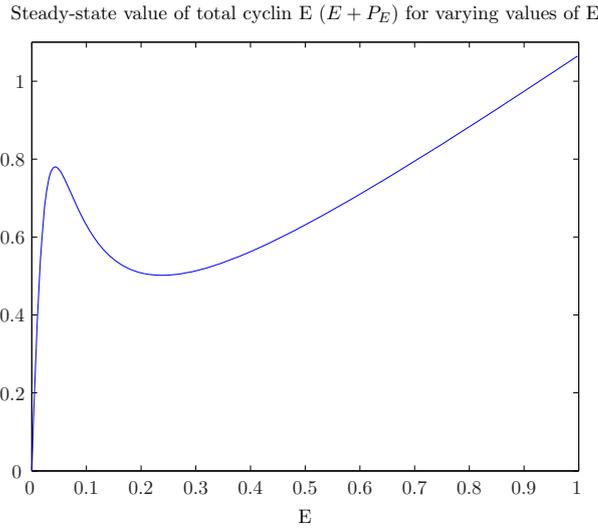


Figure 8.10: *The steady-state value of  $E + P_E$  (total cyclin  $E$ ) for varying values of  $E$  (free cyclin  $E$ ).*

$E$  can be expressed as

$$\begin{aligned}
 E + P_E &= E + \frac{\alpha_{pe}EP}{\beta_{pe} + \beta_{ph}E} \\
 &= E \frac{\alpha_{pe}(E)^2\beta_{ph} + \beta_p\beta_{ph}E + \beta_{pe}\beta_p + \alpha_{pe}\alpha_p}{\alpha_{pe}(E)^2\beta_{ph} + \beta_{pe}\beta_p + \beta_p\beta_{ph}E}. \quad (8.15)
 \end{aligned}$$

The derivative of this expression results in a quartic in terms of  $E$  on the numerator and the denominator is the square of the denominator from equation (8.15), which we know to be strictly positive. Thus, the solution to  $\frac{d(E+P_E)}{dE} = 0$  will be the solution to the quartic polynomial on the numerator.

The number of positive solutions to this quartic depends upon the choice of rate variables, and we will limit our investigations to looking at the stability of the system when the rate variables take on the values in Table 8.3. For these rate variables, there are two positive solutions to the quartic (details not shown). The relationship between  $E$  and  $E + P_E$  is shown in Figure 8.10. The two stationary points can be seen at  $E = 0.04$  (which is a local maximum) and  $E = 0.25$  (a local minimum). In Figure 8.10, we note that total cyclin  $E$  increases for increasing values of  $E$  when  $E < 0.04$  or  $E > 0.25$ . For  $E \in [0.04, 0.25]$ , increasing  $E$  actually decreases total cyclin  $E$ . This is because  $E$  in this range is killing off the  $P$  molecules that are in the  $P_E$  complex at the rate  $\beta_{ph}$ . This has the effect of freeing the  $E$  molecules that were originally bound to p27, and

hence subjecting these molecules to the effects of natural degradation (which they did not experience when in the complex  $P_E$ ). This is in contrast to the simple model discussed in Section 8.1, where we found that increasing cyclin E-CDK2 concentration always increased the overall concentration of cyclin E. Lets reconsider Figure 8.7. As free cyclin E-CDK2 concentrations decrease after progestin from a starting value of  $E(0) = 1$ , total cyclin E decreased. However, after 24 hours, total cyclin E concentrations increased slightly. This is because free cyclin E-CDK2 concentrations are in the range  $[0.04, 0.25]$  after 24 hours, and therefore total cyclin E-CDK2 concentrations increase during this time.

Results in Figure 8.10 imply that measurements of total cyclin E concentration within a cell are not indicative of the amount of active cyclin E-CDK2 within the cell, and hence not a good measure of the proliferation rate of that cell. This is in agreement with studies showing that total cyclin E concentration within a breast cell is not a reliable prognostic factor for breast cancer patients [110, 95]. It was also observed that high cyclin E concentrations did not always imply increased proliferation rates in [98]. In [110] it was noted that cyclin E concentrations often did not change significantly when cells become more or less proliferative. In [98], it was even observed that the overall concentration of cyclin E was 4 times higher in cells that were exposed to an anti-estrogen than in cells that were not exposed to the anti-estrogen. The authors of [98] did not know why this should be the case, given that cyclin E is known to aid in progression past the R point (and hence increase the proliferation rate of a cell). However, it is really cyclin E-CDK2 activity that decreases when cells become arrested, and the analysis performed in this section shows that low cyclin E-CDK2 activity (i.e. low values of  $E$ ) can result in high concentrations of total cyclin E-CDK2 (i.e. high values of  $E + P_E$ ), because  $P_E$  becomes large when  $E$  is small. This result provides a possible explanation for the experimental observations in [98].

So far, we have only discussed a model that contains cyclin E and p27, and have not included the effects of cyclin D. This is because we wished to investigate the relationship between cyclin E-CDK2 and p27, and did not want to complicate the analysis by including cyclin D-CDK4. Even though the model is simple, it requires several input variables -  $S(t)$ ,  $G_c(t)$  and  $D_{act}(t)$  in order to accurately model changes in cyclin E-CDK2 and p27 after progestin exposure. Requiring a large amount of input data in order to accurately model changes in these two proteins after progestin means that our model is not very

flexible, and is restricted in the number of situations we may consider. We wish to modify our model further so that we do not require as many input variables, and instead allow our extended model to calculate the input variables that we need. To do this, we must introduce the effects of cyclin D-CDK4 into our model so that we have a way of calculating  $D_{act}(t)$  after progestin exposure. To avoid using pre-calculated values for the changes in  $G_c(t)$  and  $S(t)$  after progestin exposure, we can use the modelled changes in cyclin D-CDK4 activity to drive cell cycle progression, by combining the protein model with the cell cycle model developed earlier in Chapter 6. We do this by using our modelled changes in cyclin D-CDK4 activity to drive changes in  $\gamma$  (using equation (6.5) on page 190), and subsequently, changes in cell cycle progression. Thus, we only need to calculate changes in cyclin D-CDK4 activity after progestin exposure, and our cell cycle model can provide the values of  $G_c(t)$  and  $S(t)$  after progestin exposure.

In summary, we will make two modifications to the model presented in equation (8.6). One is introducing the effects of cyclin D-CDK4 into the protein model, so that we can simulate changes in  $D_{act}(t)$  after progestin exposure. The second modification is combining the protein model with the cell cycle model from equation (3.7). When we introduce the effects of cyclin D and combine the protein models with the cell cycle model, we find that some features of our simulation are different to what is observed in [84]. As we have only discussed experimental results after progestin from one experiment, it is difficult to determine which features of protein concentration changes are essential to explain the effects of progestin, and which are perhaps not so critical. Thus, we wish to discuss the results of another experiment from [37], which investigates cell cycle phase changes, as well as protein concentration changes, after progestin exposure in T47D cells. By considering the results from two independent experiments, we can get a better idea of what features are common between the two experiments - this can tell us what features are critical to progestin exposure, and what features are not so essential in understanding progestin exposure. We will discuss this experiment now, and discuss how we extend our model in detail later in Section 8.7.

NOTE:  
This figure is included on page 267 of the print copy of  
the thesis held in the University of Adelaide Library.

Figure 8.11: *The black dots and the corresponding line connecting them line represents changes in (a) cyclin D1, (b) cyclin D3, (c) cyclin E1 and (d) p27 concentration after R5020 exposure. The x-axis represents the number of hours after progestin exposure. Other lines represent data from other investigations, which are discussed later in Section 8.7. Results obtained from [37].*

## 8.5 A discussion of experimental results from [37]

In [37], T47D cells were exposed to the progestin R5020, and protein concentration changes were observed until 80 hours after progestin exposure. The changes in total cyclin D1, cyclin E and p27 concentration were reported, and are reproduced here in Figure 8.11. These results show changes in total cyclin D1, cyclin D3, cyclin E (cyclin E1) and p27 concentration after progestin exposure as the black squares, and the black curve is a line through successive data points. The dotted line in the background of each figure represents the S + G2/M phase proportions after progestin exposure. The arrow in the figures at 40 hours, and the white dots that appear at times later than 40 hours, will be discussed in depth in later sections.

We can see that there are some differences between the results of [84] and [37], particularly in the changes in cyclin D1 concentrations after progestin exposure. Figure 8.11(a) shows a 3-fold increase in cyclin D1 concentration within the first 10 hours, whereas [84] only observed a 40% increase in cyclin D1 concentrations during this time (see Figure 7.3). Cyclin D1 concentrations were also not observed to drop below control levels in [37], however in [84], cyclin D1 concentrations fell to 50% of control within 24 hours. In Section 7.4, we concluded that the observed decrease in total cyclin D concentrations in [84] was partly due to a decrease in the concentration of free cyclin D1 molecules. It is possible that total cyclin D1 does not decrease to such low levels in [37] because the concentration of free cyclin D1 in control cells is not as high as was observed in [84]. We note that the cyclin D3 concentration changes appear very similar in both [84] and [37].

The changes in total cyclin E1 concentrations are similar in [84] and [37]. However in [37], cyclin E1 concentrations increased by 50% within the first 10 hours of exposure, in parallel with the increase in cyclin D1. This is likely to be due to the fact that the large increase in cyclin D observed in this experiment also results in an increase in cyclin D activity, causing an increase in the production rate of cyclin E1 molecules. In [84], cyclin E1 concentrations were not observed to increase by very much within the first 10 hours (Figure 7.3 shows an increase in cyclin E1 concentration of about 10-20%, and we note the standard error of this observations is about  $\pm 10\%$ ), and this is likely to be due to the fact that cyclin D-CDK4 was also not observed to increase in [84]. However, as we discussed previously, we have assumed in our model that cyclin D-CDK4 activity must increase within the first 10 hours in order to have an increase in the S phase proportion after progestin exposure. Thus, our model results from Figure 8.7 show an increase in total cyclin E-CDK2 concentrations during the first 10 hours of exposure due to an early increase in cyclin D-CDK4 activity, in agreement with [37]. Figure 8.11(c) shows cyclin E1 concentrations begin to decrease after 10 hours, reaching 30% of control by about 20 hours. This decrease was also observed in [84], and was captured in our model results from Figure 8.7. In both [84] and [37], cyclin E1 concentrations were observed to increase at about 30 hours, a feature also captured in our model results from Figure 8.7. As observations were made in [37] for a longer period of time, we observe cyclin E1 concentrations continuing to increase until 50 hours after exposure when protein concentrations levelled out to about 3-fold of control levels. Cyclin E1 concentrations were observed to remain high for the remaining hours of the experiment in [37].

The changes in p27 concentrations were also similar between [84] and [37], except in [37], p27 concentrations increased almost 2-fold within the first 10 hours of exposure and then decreases to control levels, in parallel with the transient increase in cyclin D concentrations during this time. This is likely to be due to the fact that cyclin D-CDK4 and p27 have a high rate of binding, and therefore the increase in cyclin D-CDK4 concentration within the first few hours will cause an increase in cyclin D-CDK4-p27 concentration. This increase in p27 was not observed in [84] because cyclin D concentrations did not increase by much after progestin in this experiment. The early increase in p27 concentration was also not observed in our model results shown in Figure 8.7. This is because we have not considered the effects of cyclin D in equations (8.6), and so we cannot capture an increase

in cyclin D-CDK4-p27. In both [84] and [37], p27 concentration increases to 2-fold above control at later times of progestin exposure (in [84], p27 concentration reached 2-fold at about 30 hours, and in [37] it occurred at about 40 hours after exposure). This later increase in p27 was captured in our model results, which we found occurred because of the decrease in free cyclin E-CDK2. In [37], p27 remained at about this level for the remainder of the experiment.

We wish to note that the features in the protein concentration profiles that are different between the two can be explained if we assume that in [37], cyclin D concentration and activity increases initially by a higher amount than in [84]. This is because a higher cyclin D-CDK4 activity will cause an increase in cyclin E production in the early hours (as we saw in our model results), and a higher cyclin D concentration will cause an increase in cyclin D-CDK4-p27, and therefore an increase in total p27 concentration shortly after progestin exposure. As we have chosen to use a cyclin D-CDK4 activity profile that increases within the first 6 hours of progestin exposure (see Figure 8.4), we might expect our modelled results to compare better with experimental data from [37] than from [84]. However, we will continue to use the experimental data from [84] when comparing our model results. This is because the model results on changes in total cyclin E-CDK2 and cyclin D-CDK4 concentrations do not include the presence of p21 molecules, and we have successfully managed to take these molecules out from the data in [84], however we cannot do this with the data from [37] because they do not provide the elution profiles for their proteins.

We note that the similarities in experimental results between [84] and [37] probably represent critical features of the protein concentration changes after progestin exposure. A short increase in cyclin D concentration within the first 10 hours, followed by a decline in concentration of cyclin D is likely to be critical to progestin exposure, however the level to which cyclin D increases initially, and the minimum it reaches, are not consistent between the two experiments. In both [84] and [37], cyclin E concentrations decline after 10 hours of progestin exposure, reaching about 25% - 30% of control by about 18 hours, and after about 30 hours, cyclin E concentrations increased. As this occurs in both experiments, we interpret this as a critical feature of progestin exposure. In both experiments, p27 concentrations increase to about 2-fold of control by 30-40 hours after exposure, and so again we interpret this as a critical feature. Our model of cyclin E-CDK2 and p27 from

equations (8.6) successfully captures these essential features in cyclin E-CDK2 and p27 concentration changes after progestin exposure.

We will now discuss how we can extend our model to include the effects of cyclin D-CDK4 activity.

## 8.6 Expanding our model - introducing cyclin D

We begin this section by describing a model of p27 and cyclin D interactions. We then combine this model with the one described in equations (8.6), and compare these model results with the experimental data from [84].

To develop a set of equations describing the interaction between cyclin D-CDK4 and p27, we begin by making the assumption that the rate of cyclin D-CDK4 production is equivalent to the rate of cyclin D production, and we call the concentration of cyclin D-CDK4 molecules at time  $t$ ,  $D(t)$  (note that the variable  $D(t)$  is to be contrasted with the variable  $d(t)$  used in Chapters 4 and 6, as  $D(t)$  here represents only cyclin D-CDK4 molecules, whereas  $d(t)$  previously represented all cyclin D molecules, including those bound to p27 molecules). We will make the assumption that cyclin D-CDK4 molecules are produced at the rate  $\alpha_d(t)$ , which depends upon time because cyclin D production depends highly upon the external environment. Cyclin D-CDK4 molecules decay at the rate  $\beta_d$ , where  $\beta_d$  is constant. The inhibitor p18 also causes the decay of cyclin D-CDK4 molecules, by causing the cyclin D and CDK4 molecules to unbind [113]. We could include this effect explicitly in our model, however, we do not know how p18 changes after progestin exposure. For simplicity, we therefore model the time-dependent influences on cyclin D-CDK4 concentration using only one term (namely,  $\alpha_d(t)$ ), and so  $\alpha_d(t)$  encapsulates both the change in production rate of cyclin D-CDK4 molecules, and the decay of cyclin D-CDK4 molecules due to the presence of p18. We assume that p27 binds to cyclin D-CDK4 at a rate of  $\alpha_{pd}$ , creating the complex cyclin D-CDK4-p27, and the concentration of this complex at time  $t$  will be referred to as  $P_D(t)$ . We do not include a natural decay rate of the complex cyclin D-CDK4-p27, as it is known that such inhibitor-CDK complexes are very stable [135, pg 273]. We assume that this complex disassociates to produce cyclin

D-CDK4 and p27 at a rate of  $\beta_{pd}$ . This system can be expressed as

$$\frac{dP(t)}{dt} = \alpha_p - Dbind - \beta_p P(t), \quad (8.16a)$$

$$\frac{dD(t)}{dt} = \alpha_d(t) - Dbind - \beta_d D(t), \quad (8.16b)$$

$$\frac{dP_D(t)}{dt} = Dbind, \quad (8.16c)$$

$$Dbind = \alpha_{pd} D(t) P(t) - \beta_{pd} P_D(t). \quad (8.16d)$$

We note that this set of equations is identical in form to those described in equations (8.1) (the simple model of cyclin E and p27). This means that cyclin D-CDK4 and p27 interact in the same way as cyclin E-CDK2 and p27, except that cyclin D-CDK4 does not cause the degradation of p27 molecules. As we intend to eventually combine this model with the model for cyclin E-CDK2 and p27 from equations (8.6), we also want to non-dimensionalise with respect to the starting concentration of cyclin E-CDK2,  $E(0)$ , as we did when we non-dimensionalised the model from equations (8.6). When we do this, we get a set of equations identical in form to the ones displayed here in (8.16), but with the new rate variables taking on different values to the original ones. The only rate variables that remain unchanged after non-dimensionalisation are  $\beta_d$ ,  $\beta_p$  and  $\beta_{pe}$ . We note that it is not significant that the rate variables are different, because we will be using the non-dimensionalised model to determine the values of the rate parameters anyway, and therefore the model displayed here in equations (8.16) can be considered non-dimensionalised with respect to  $E(0)$ .

When we combine the models from equations (8.16) with the model from equations (8.6), we wish to calculate the changes in cyclin D-CDK4 activity using equations (8.16), rather than use the pre-calculated form for cyclin D-CDK4 activity from Figure 8.4 (we must calculate cyclin D-CDK4 activity in order to determine the rate of production of cyclin E-CDK2 molecules). As we discussed previously, cyclin D-CDK4 activity is not equivalent to the total concentration of cyclin D-CDK4 molecules, and we suggested that for cyclin D-CDK4 activity to decrease by the amount observed in [84], cyclin D-CDK4-p27 molecules must move from being active to inactive. As we discussed in Section 7.4, other experiments have shown that cyclin D-CDK4-p27 molecules can become inactive by changing the phosphorylation status at the Y88/Y89 sites on p27,

and in our model, we will assume that progestin reduces cyclin D-CDK4 activity via this mechanism. In the next section, we will describe how we incorporate this mechanism into our model.

### **Introducing the mechanism of reduction in cyclin D-CDK4 activity after progestin exposure**

We will assume that in the absence of progestin, all p27 molecules are phosphorylated at the Y88/Y89 sites. This is a valid assumption, as total cyclin D concentration and cyclin D-CDK4 activity are highly correlated in T47D cells in the absence of progestin, and as most cyclin D-CDK4 molecules are found in cyclin D-CDK4-p27 complexes, these molecules must be active, and hence most of the p27 molecules must be phosphorylated at the Y88/Y89 sites. In [50] it was noted that the phosphorylation status of p27 is independent of what form it is in (i.e. independent of whether it is currently bound to other molecules such as cyclin D-CDK4). Given this information, we will assume that progestin directly reduces the phosphorylation status of all p27 molecules present in the cells.

It would be biologically accurate to model two different populations of p27 - one that is phosphorylated and one that is not. However this would also require us to monitor both phosphorylated and un-phosphorylated populations of free p27, cyclin E-CDK2-p27 and cyclin D-CDK4-p27. Including these additional variables makes our model much more complex, and unnecessarily so, because we are really only interested in whether cyclin D-CDK4-p27 complexes are Y88/Y89 phosphorylated (i.e. active) or not. Thus, for simplicity, we will assume that progestin directly influences the proportion of p27 molecules that are phosphorylated at the Y88/Y89 sites, and for simplicity, we only model the change in the proportion of cyclin D-CDK4-p27 molecules that are phosphorylated, and we will not monitor the populations of free p27 or cyclin E-CDK2-p27 that are phosphorylated.

In addition to the above, we will assume that the concentration of active cyclin D-CDK4-p27 molecules is equivalent to cyclin D-CDK4 activity. Although free cyclin D-CDK4 molecules are also active, we note that the concentration of these molecules at any time after progestin exposure is very small, and we choose not to consider this population

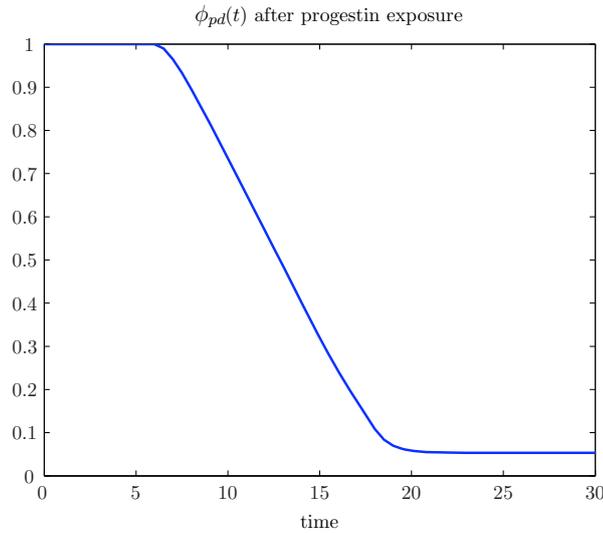


Figure 8.12: *The proportion of total cyclin D-CDK4-p27 that is active after progestin exposure.*

for simplicity.

Thus, we can express the concentration of active cyclin D-CDK4 molecules at time  $t$  as

$$D_{act}(t) = \phi_{pd}(t)P_D(t), \quad (8.17)$$

where  $\phi_{pd}$  represents the proportion of  $P_D(t)$  molecules that are active (or, equivalently, the proportion of total p27 that is Y88/Y89 phosphorylated) at time  $t$ . After exposure to substances other than progestins, we would expect  $\phi_{pd}(t)$  to remain at 1 (i.e. cyclin D-CDK4-p27 molecules are usually considered to be active at all times), however after progestin exposure, we would expect the value of  $\phi_{pd}(t)$  to decrease significantly.

We can determine the functional form of  $\phi_{pd}(t)$  at times after progestin as follows. We know the  $D_{act}(t)$  profile from the previous section (see Figure 8.4), and we know  $P_D(t)$  from our elution profile analysis performed in Section 7.4 (see Figure 8.1, page 237). This means we can re-arrange equation (8.17) to find an expression for  $\phi_{pd}(t)$ . Doing this gives the curve shown in Figure 8.12. We can see that cyclin D-CDK4-p27 molecules remain active for the first 6 hours and then become inactive at a fairly constant rate between 6 and 20 hours after progestin. After 20 hours, the proportion of active cyclin D-CDK4-p27 molecules remains unchanged at about 7% of the total pool of cyclin D-CDK4-

p27 molecules. If an experiment were to be performed to test our hypothesis regarding the change in active cyclin D-CDK4 molecules, then we would expect the proportion of phosphorylated p27 molecules would decrease according to Figure 8.12.

### Combining the model of cyclin D and p27 with the model of cyclin E-CDK2 and p27

We now wish to combine the model of p27 and cyclin D-CDK4 with the model for cyclin E-CDK2 and p27 from the previous section. The full model encompassing cyclins E, D and p27 is shown in equations (8.18).

$$\frac{dP(t)}{dt} = \alpha_p - Dbind - Ebind - \beta_p P(t), \quad (8.18a)$$

$$\frac{dE(t)}{dt} = \alpha_e G_C(t) D_{act}(t) - Ebind - \beta_e S(t) E(t) + \beta_{ph} E(t) P_E(t), \quad (8.18b)$$

$$\frac{dP_E(t)}{dt} = Ebind - \beta_{ph} E(t) P_E(t), \quad (8.18c)$$

$$\frac{dD(t)}{dt} = \alpha_d(t) - Dbind - \beta_d D(t), \quad (8.18d)$$

$$\frac{dP_D(t)}{dt} = Dbind, \quad (8.18e)$$

$$D_{act}(t) = \phi_{pd}(t) P_D(t), \quad (8.18f)$$

where

$$Ebind = \alpha_{pe} E(t) P(t) - \beta_{pe} P_E(t) \text{ and}$$

$$Dbind = \alpha_{pd} D(t) P(t) - \beta_{pd} P_D(t).$$

We note that by including the model for cyclin D-CDK4 and p27, we have two additional drivers in our combined model -  $\alpha_d(t)$ , which we have not yet determined, and  $\phi_{pd}(t)$ , which we calculated previously and which is shown in Figure 8.12. It may seem that by extending our model to include different proteins, we are increasing the number of input variables required to simulate our model - contrary to our claim that extending our model would require less prior information. However, later in Section 8.7 we will discuss how we can eliminate the need for pre-calculated input variables  $G_c(t)$  and  $S(t)$  by combining the protein model discussed here with the cell cycle model from Chapter 6.

**Determining the rate variables for the model in equation (8.18)**

We begin by determining the steady-state of equations (8.18), which will help us to determine many of the rate parameters. The steady-state can be expressed as

$$P = \alpha_p \frac{\beta_{pe} + \beta_{ph}E}{\alpha_{pe}E^2\beta_{ph} + \beta_{pe}\beta_p + \beta_p\beta_{ph}E}, \quad (8.19a)$$

$$E = \frac{\alpha_e G_C D_{act}}{\beta_e S}, \quad (8.19b)$$

$$P_E = \frac{\alpha_{pe}EP}{\beta_{pe} + \beta_{ph}E}, \quad (8.19c)$$

$$D = \frac{\alpha_d}{\beta_d}, \quad (8.19d)$$

$$P_D = \frac{\alpha_{pd}DP}{\beta_{pd}}, \quad (8.19e)$$

$$D_{act} = \phi_{pd} \times P_D. \quad (8.19f)$$

The steady-state values of  $P$ ,  $E$  and  $P_E$  are independent of the values of cyclin D-CDK4 complexes, and we will assume that  $P(0)$ ,  $E(0)$  and  $P_E(0)$ , as well as the rate parameters associated with these variables, take on the same values as they did in the model containing only free cyclin E-CDK2 p27 and cyclin E-CDK2-p27 (see Table 8.3). Therefore we need only to determine the rate variables and starting values of the variables associated with the addition of cyclin D.

To determine  $\beta_d$ , we note that cyclin D has a half-life of 20 minutes [31] which gives  $\beta_d = 2\text{hr}^{-1}$ . We can use the steady-state equations to determine the remaining variables. From Figure 7.13, we note that  $P_D(0) = 11 \times P_E(0) = 2.2$ . From Figure 8.1, we note that  $D(0) = 0.1 \times P_D(0) = 0.22$ . We will assume that all of the cyclin D-CDK4-p27 molecules are phosphorylated initially, i.e.  $\phi_{pd}(0) = 1$ . We note that cyclin D concentration must be able to change quickly in response to environmental changes. As we have assumed that cyclin D-CDK4-p27 complexes experience no decay in our model, we must ensure that these complexes have a fast rate of unbinding so that the cyclin D molecules can be degraded when necessary. Thus we will assume  $\beta_{pd} = 2$ , which we find is large enough for the protein to change in concentration in a reasonable time. Assuming that initial protein concentrations are in steady-state, we can use the steady-state equations (8.19)

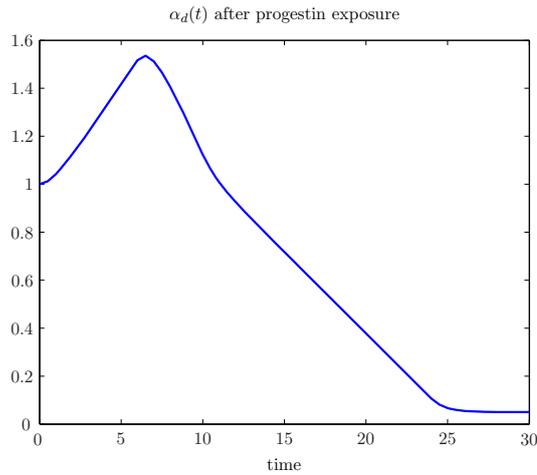


Figure 8.13: *The functional form for  $\alpha_d(t)$  after progestin exposure.*

and write  $\alpha_{pd} = \beta_{pd} \frac{P_D(0)}{D(0)P(0)} = 2000$ . The value of  $\alpha_{pd}$  is so large because in control cells both  $D(0)$  and  $P(0)$  are so small compared to  $P_D(0)$ .

We have now determined all the unknown rate parameters for our model, except for the functional form of  $\alpha_d(t)$  after progestin exposure. We can determine the value of this variable at  $t = 0$  by substituting the values of  $D(0)$  and  $\beta_d$  into the steady-state equations (8.19), rearranging  $D(0) = \frac{\alpha_d(0)}{\beta_d}$  to give  $\alpha_d(0) = 0.22 \times \beta_d = 0.44$ , and so we know  $\alpha_d(0)$ . If we assume that cyclin D-CDK4-p27 reaches steady-state quickly at all times (i.e.  $Dbind = 0$ ) and that  $D(t)$  reaches steady-state quickly at all times, then equation (8.16b) can be rearranged to give  $D(t) = \frac{\alpha_d(t)}{\beta_d}$ . Thus, we can determine  $\alpha_d(t)$  if we know  $D(t)$ . We can determine how  $D(t)$  changes after progestin exposure using our results from the previous section as follows. If we again assume that cyclin D-CDK4-p27 reaches steady-state quickly at all times, then solving  $Dbind = 0$  gives

$$D(t) = \frac{\beta_{pd} P_D(t)}{\alpha_{pd} P(t)}.$$

Now, we know from the model results in Section 8.4 how  $P(t)$  changes after progestin exposure (see Figure 8.7(b)). We also know the relative changes in  $P_D(t)$  from the p27 elution profile in Figure 8.1. Substituting these functions into the above expression allows us to determine how  $D(t)$  changes after progestin exposure, and is shown in Figure 8.13. Thus, as  $\alpha_d(t) = \beta_d D(t)$ , the functional representation in Figure 8.13 also represents the relative change in  $\alpha_d(t)$  after progestin exposure, and we will use this profile for  $\alpha_d(t)$ .

Table 8.4: *The parameter values and starting concentrations of proteins described in equations (8.18). Values that are not described here are displayed earlier in Table 8.3 on page 262.*

| <b>Parameter</b> | <b>value</b>    |
|------------------|-----------------|
| $\beta_d$        | 2               |
| $\alpha_d(t)$    | See Figure 8.13 |
| $\beta_{pd}$     | 2               |
| $\alpha_{pd}$    | 2000            |
| $\phi_{pd}(t)$   | See Figure 8.12 |
| $D(0)$           | 0.22            |
| $P_D(0)$         | 2.2             |

At this point we have determined each of the rate variables for our model and also the starting concentrations of each of the proteins. The values of the additional rate parameters and starting concentrations from the addition of cyclin D into the model are summarised in Table 8.4.

As mentioned in Section 8.1, the model in equations (8.1) represents the interaction between cyclin D-CDK4 and p27 molecules, and discussions in that section showed that the model is stable for any choice of rate variables. The model discussed in Section 8.3 (equations (8.6)) was also shown to be stable for the rate variables we discussed. The full model encompassing cyclins E, D and p27 discussed in this section can be thought of as merging of the two previous models from equations (8.1) and (8.6). However the behaviour of the model may not be identical around the steady-state, and so the region of stability may not be identical. We have investigated the stability of this extended model for the rate variables chosen in Section 8.4 (see Table 8.3), along with the additional rate variables determined in this section, and we find that both the non-proliferative ( $E = 0.037$ ) and the proliferative ( $E = 1$ ) steady-state values are stable for these rate variables (details not shown).

We have already confirmed that the changes in cyclin E-CDK2, p27 and cyclin E-CDK2-p27 are successfully captured in our model containing these variables from equations (8.6). By extending the model to contain cyclin D-CDK4 and cyclin D-CDK4-p27, we have chosen the drivers  $\alpha_d(t)$  and  $\phi_{pd}(t)$  so that cyclin D-CDK4 and cyclin D-CDK4-p27 change in concentration after progestin exposure according to the data provided in [84]. We note that we have simulated the model above using the drivers for  $\alpha_d(t)$  and

$\phi_{pd}(t)$  discussed in this section, and we can confirm that protein concentration changes compare well with those from [84] as expected (details not shown).

The last modification we wish to make to our model is to combine the protein model shown in equations (8.18) with the cell cycle model from Chapter 6, so that we do not have to use pre-calculated forms for  $G_c(t)$  and  $S(t)$ . Once we do this, we will simulate the full model and investigate the results, by comparing them with [84].

## 8.7 Combining the protein model from equations (8.18) with the cell cycle model from equations (3.7)

Now that we have a model that can describe changes in cyclin D activity after progestin exposure, we can use this model to simulate cell cycle progression in T47D cells, essentially combining the protein models discussed here in Section 8 with the cell cycle model discussed in Sections 3 to 6. We do this by substituting the value of  $D_{act}(t)$  as the value of  $D(t)$  in equation (6.5). The value of  $\gamma$  is then used to determine the rate at which cells progresses from the  $G_b$  into the  $G_c$  phase using the model proposed in equations (3.7), and will provide us the input variables  $G_c(t)$  and  $S(t)$  after progestin exposure. Thus the models of cell cycle progression and changes in cell cycle protein concentrations will become intimately linked, with values of cell cycle phase proportions influencing the protein model through changes in cyclin E production and degradation, and the protein model in turn influencing the cell cycle phase proportions through the changes in cyclin D activity. We will call this the combined model. Such a model will allow us to simulate cell cycle progression, and protein concentration changes, under different environmental conditions.

The combined model is shown, in full, below.

$$\begin{aligned}
\frac{dP(t)}{dt} &= \alpha_p - Dbind - Ebind - \beta_p P(t), \\
\frac{dE(t)}{dt} &= \alpha_e G_C(t) D_{act}(t) - Ebind - \beta_e S(t) E(t) + \beta_{ph} E(t) P_E(t), \\
\frac{dP_E(t)}{dt} &= Ebind - \beta_{ph} E(t) P_E(t), \\
\frac{dD(t)}{dt} &= \alpha_d(t) - Dbind - \beta_d D(t), \\
\frac{dP_D(t)}{dt} &= Dbind, \\
D_{act}(t) &= \phi_{pd}(t) P_D(t), \\
Ebind &= \alpha_{pe} E(t) P(t) - \beta_{pe} P_E(t), \\
Dbind &= \alpha_{pd} D(t) P(t) - \beta_{pd} P_D(t), \\
\frac{d\gamma(t)}{dt} &= f_0 + \omega D_{act}(t) + \omega_\gamma D_{act}(t) \gamma(t) - \beta_\gamma \gamma(t), \\
\frac{dN_{G_a}(t)}{dt} &= 2\beta N_{M_b}(t) - 2\beta N_{M_b}(t - \tau_{G_a}), \\
\frac{dN_{G_b}(t)}{dt} &= -\gamma(t) N_{G_b}(t) + 2\beta N_{M_b}(t - \tau_{G_a}), \\
\frac{dN_{G_c}(t)}{dt} &= \gamma(t) N_{G_b}(t) - \gamma(t - \tau_{G_c}) N_{G_b}(t - \tau_{G_c}), \\
\frac{dN_{S_a}(t)}{dt} &= \gamma(t - \tau_{G_c}) N_{G_b}(t - \tau_{G_c}) - \gamma(t - \tau_{G_c} - \tau_S) N_{G_b}(t - \tau_{G_c} - \tau_S), \\
\frac{dN_{S_b}(t)}{dt} &= \gamma(t - \tau_{G_c} - \tau_S) N_{G_b}(t - \tau_{G_c} - \tau_S) - \alpha N_{S_b}(t), \\
\frac{dN_{M_a}(t)}{dt} &= \alpha N_{S_b}(t) - \alpha N_{S_b}(t - \tau_M), \\
\frac{dN_{M_b}(t)}{dt} &= \alpha N_{S_b}(t - \tau_M) - \beta N_{M_b}(t).
\end{aligned} \tag{8.20}$$

### Results of our combined model

Using the combined model, we simulate cell cycle progression and protein concentration changes after T47D cells are exposed to progestin as was done in [84]. The resultant S phase changes, as well as the protein concentration changes, are shown in Figure 8.14.

In Figure 8.14(a), we have the modelled changes in cyclin D-CDK4-p27 as the light shaded region and the changes in free cyclin D-CDK4 as the dark shaded region. The

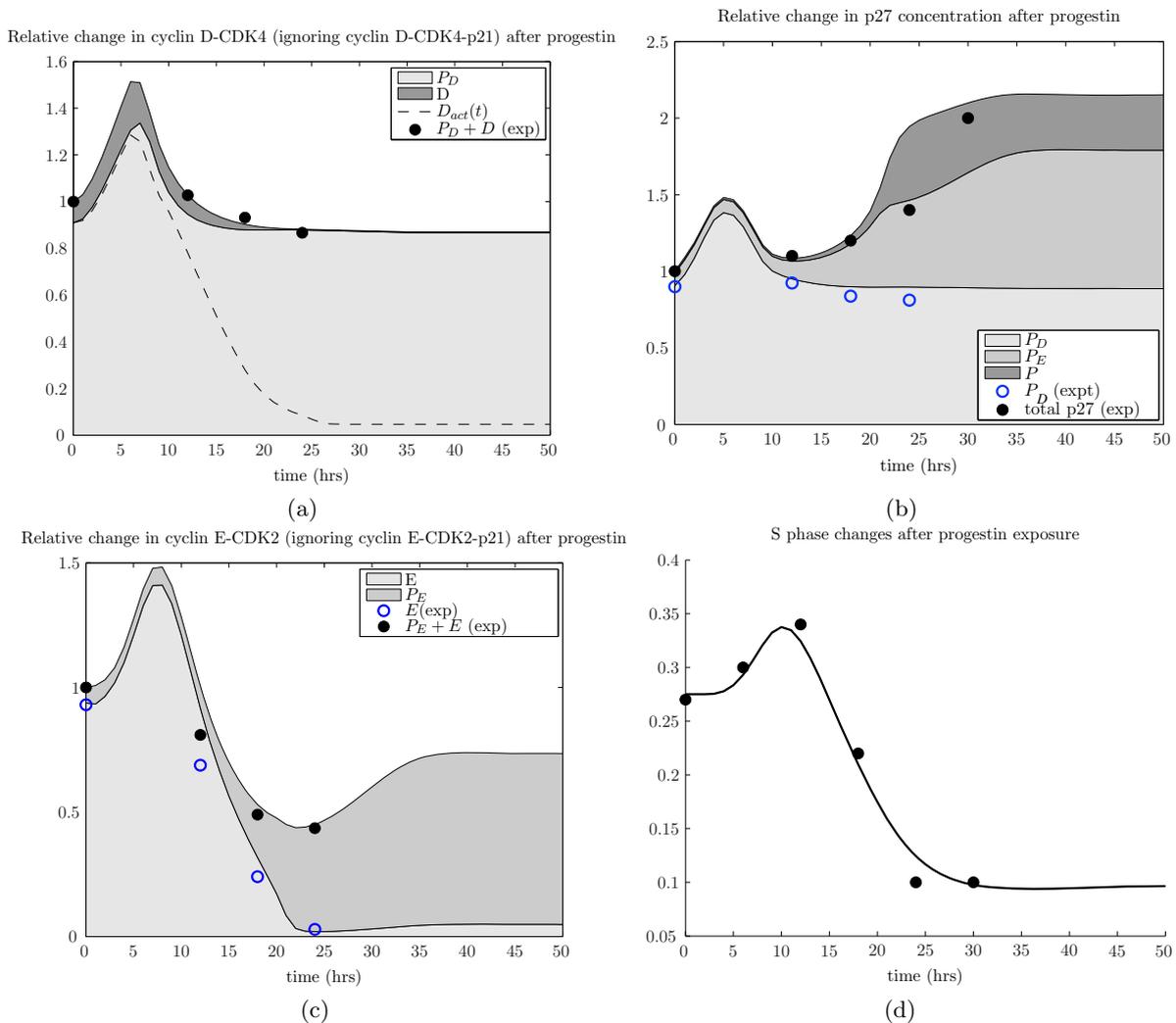


Figure 8.14: Results of the combined model after progestin exposure. Modelled changes in (a) cyclin D concentration, (b) p27 concentration, (c) cyclin E concentration, and (d) the S phase proportion, were simulated using the combined model in equations (8.20). The black dots in (a) represent experimental data on the changes in  $P_D(t) + D(t)$  after progestin exposure from [84] and the dashed curve in (a) represents the modelled changes in  $D_{act}(t)$  after progestin exposure. The dots in (c) represent experimental data on changes in the S phase proportion after progestin exposure. For a full description of the dots and the shaded areas in (b) and (c), see Figure 8.6 and the main body text which discussed this figure.

shaded regions are placed one on another so that the overall shaded region represents the relative changes in total concentration of cyclin D here (i.e. the overall shaded region represents  $D(t) + P_D(t)$  but does not include free cyclin D or cyclin D-CDK4-p21). The dashed line shows the changes  $D_{act}(t)$  after progestin exposure. The black solid dots are computed from experimental data from [84] as discussed in Section 7.4. We can see that  $D(t) + P_D(t)$  increases to about 1.5-fold above controls within the first 6 hours, and then decreases to about control levels by about 15 hours after exposure, remaining at this level for the rest of the experiment. The increase in cyclin D concentrations within the first few hours is due to the increase in  $\alpha_d(t)$  during this time. At later times,  $\alpha_d(t)$  decreases significantly, reaching a minimum of about 2% of control levels, and we can see that the concentration of free cyclin D-CDK4 decreases significantly in parallel with the changes in  $\alpha_d(t)$ . However, overall cyclin D levels do not drop significantly below the level of controls, because free p27 concentrations also increase in parallel with the decrease in  $\alpha_d(t)$ , resulting in essentially unchanging values of  $P_D(t)$  after progestin exposure, and hence we do not observe any significant change in the concentration of  $P_D(t)$  after progestin exposure. When we originally considered the experimental results from [84], we observed that total cyclin D concentrations decreased to about 50% of control after progestin exposure. However, after the analysis of this data in more detail in Section 7.4, we found that the decrease in total cyclin D concentration was likely due to decreases in the concentration of free cyclin D molecules (which we are not considering here) and that the concentration of total cyclin D-CDK4 concentration is likely to be essentially unchanged (see Section 7.4). We also note that the changes in cyclin D-CDK4 activity, shown as the dashed line, compares well with the cyclin D-CDK4 curve shown earlier in Figure 8.4, as we would expect. Allowing the simulation to continue shows that cyclin D-CDK4 and cyclin D-CDK4-p27 concentrations have reached steady-state. We do not observe the additional ‘bump’ in cyclin D concentrations that is observed in [37] between 30 and 40 hours after progestin exposure. Perhaps in [37], there is a small increase in  $\alpha_d$  at these later times that we have not included, although we do not know why this would be the case.

Let us now consider Figure 8.14(b). In this figure,  $P_D(t)$  is shown as the light shaded region,  $P_E(t)$  is shown as the darker shaded region, and  $P(t)$  is shown as the very dark shaded region. Again, the shaded regions are placed one on another so that the overall

shaded area represents the relative change in total p27 concentrations after progestin exposure. The experimentally observed changes in  $P_D(t)$  and total p27 are shown as the hollow blue dots and the solid black circles, respectively. We can see that total p27 concentration increases 1.5 fold within the first 6 hours because of an increase in cyclin D-CDK4-p27 complexes. These complexes increase during this time because of the increase in the production rate of cyclin D-CDK4 molecules, which bind to p27 to form the stable cyclin D-CDK4-p27 complex. This increase in total p27 concentration within the first 6 hours of exposure was not observed in [84], but it was observed in [37]. The concentration of p27 decreases again between 6 and 10 hours to reach control levels, and then increases again, reaching two-fold of control by 30 hours after exposure. This later increase is due to an increase in free p27 and cyclin E-CDK2-p27 concentrations. Both free p27 and cyclin E-CDK2-p27 concentrations increase because free cyclin E-CDK2 concentrations decrease during this time (see Figure (c)), and from Figure 8.9 on page 263, we can see that low cyclin E-CDK2 concentrations result in high p27 and high cyclin E-CDK2-p27 concentrations. It was also observed in [84] that the concentration of cyclin E-CDK2-p27 increased after progestin exposure, and that free p27 concentrations increased significantly by 30 hours after progestin exposure, in good agreement with our modelled results. The eventual two-fold increase in total p27 that our model produced was also observed in both [84] (solid black circles) and [37], implying that this two-fold increase is an important feature after progestin exposure. We also note that our model is able to differentiate between the causes of the increase observed in p27 within the first 6 hours, and the later increase observed at 30 hours, the first being due to an increase in cyclin D-CDK4-p27 concentration, and the latter being due to a decrease in cyclin E-CDK2 concentrations, causing an increase in both free p27 and cyclin E-CDK2-p27.

We turn our attention now to Figure 8.14(c). This figure shows changes in cyclin E-CDK2 concentration ( $E(t)$ ) as the light shaded region, and changes in cyclin E-CDK2-p27 ( $P_E(t)$ ) as the darker shaded region, and the total shaded region represents changes in  $E(t)+P_E(t)$  after progestin. The experimentally observed changes in  $E(t)$  and  $E(t)+P_E(t)$  are shown as the hollow blue circles and the solid black circles, respectively. In this representation, we observe an increase in  $E(t) + P_E(t)$  during the first 8 hours due to an increase in free cyclin E-CDK2 concentration during this time (cyclin E-CDK2-p27 concentration is essentially unchanged during this period). This is due to an increase in cyclin

D-CDK4 activity, causing an increase in the production rate of cyclin E-CDK2. As cyclin D-CDK4 activity declines, cyclin E-CDK2 concentration drops significantly between 12 and 24 hours after progestin exposure. We note that the concentration of cyclin E-CDK2-p27 increases as free cyclin E-CDK2 concentration drops. This is due to the fact that free cyclin E-CDK2 molecules kill p27 molecules that are bound to other cyclin E-CDK2 molecules, and so as cyclin E-CDK2 concentration drops, cyclin E-CDK2-p27 molecules experience less decay. Between 24 and 40 hours, cyclin E-CDK2 concentration remains low, causing an increase in cyclin E-CDK2-p27 concentration. Total cyclin E-CDK2 concentrations reach steady-state by 40 hours after progestin exposure. We note that our modelled changes in both  $E(t)$  and  $E(t) + P_E(t)$  compare very well with experimental observations from [84] (compare blue circles and solid black dots) for the first 24 hours after progestin exposure. The experiment performed in [37] also showed a decrease in total cyclin E within the first 25-30 hours, followed by an increase in concentration after this time. This implies that these features of cyclin E concentration changes are important, and likely to be universal features after progestin exposure. Thus, the fact that our model also captures these features means that our model has captured the important mechanisms regarding changes in cyclin E concentrations after progestin exposure.

In [37], cyclin E concentrations increase to reach 3-fold of control levels after 50 hours of progestin exposure. In our simulation,  $E(t) + P_E(t)$  concentration reaches a maximum of about 70% of control levels after progestin exposure, significantly lower than that observed in [37]. We are not sure why cyclin E concentrations have been observed to increase to such high levels in [37]. One reason could be because the progestin used was R5020, which is different to the one used in [84], which is ORG 2050, and could have slightly different effects on cyclin E regulation. Another reason could be the absence of cyclin E-CDK2-p21 complexes in our simulation, which could make up a significant proportion of total cyclin E at later hours after progestin exposure. However, as we saw earlier in Section 7.4, cyclin E-CDK2-p21 complexes decrease significantly within the first 24 hours of exposure, and it is therefore unlikely that they then increase significantly after this time. As long as we are only considering cyclin E-CDK2 and cyclin E-CDK2-p27 concentration changes, then it is impossible for our model to simulate a 3-fold increase in cyclin E concentrations after progestin exposure. This can be seen if we reconsider Figure 8.10 on page 264, (i.e. the results of our steady-state analysis) which shows that for any concentration of free cyclin

E-CDK2 that is less than the starting concentration, total cyclin E will be less than the starting concentration of total cyclin E as well.

Finally, we wish to consider Figure 8.14(d). In this Figure, we show the modelled change in the S phase proportion after progestin exposure. The modelled changes in the S phase proportion compare well with experimental observations (shown as the black circles).

### Summary of our model results

The model produces changes in cyclin E-CDK2, cyclin D-CDK4 and p27 concentrations that compare very well with experimental data. In particular, the model captures the crucial features, such as a 2-fold increase in p27 concentration after progestin exposure, with only a small decrease in cyclin D-CDK4 concentration after progestin exposure. The model is also able to show that total cyclin E-CDK2 concentration decreases, reaching a minimum after about 24 hours, and thereafter increases, although we were not able to capture the observed increase of 3-fold higher cyclin E concentration that was observed in [37]. The presence of free p27 at 30 hours after exposure was also captured in our simulation results, in agreement with observations from [84]. In addition to total protein concentration changes, we were also able to capture the changes in cyclin E-CDK2-p27 concentration, and show that this complex increases significantly after progestin exposure. In conclusion, our model in equations (8.18) is able to capture the changes in relevant protein concentration after progestin exposure.

In this model, it was required that we recognise the significant difference between total cyclin D-CDK4 concentration and the concentration of active cyclin D-CDK4 molecules. We suggest this is a feature unique to progestin, as it has not been observed to occur after the addition of other substances. In this section, we found that the proportion of cyclin D-CDK4-p27 molecules that were active decreased significantly after progestin exposure.

We suggest that an experiment be performed which investigates whether there is a change in the phosphorylation status of cyclin D-CDK4-p27 complexes after progestin, in particular, whether this change occurs at the Y88/Y89 sites. If this is indeed the case, then it would result in a significant shift in the current understanding of progestin effects

on cell cycle progression.

In the next section, we wish to use the model to investigate some observations in [37]. In particular, we wish to investigate how our model would behave if a second dose of progestin was administered 40 hours after the first dose. This investigation was performed in [37]. The authors had expected the second dose of progestin to cause cells to behave similarly to after the first dose, where cells are initially stimulated to proliferate and later, cells are growth arrested. However the second dose had unexpected effects on cells behaviour. We will discuss their results in the next section.

## 8.8 The effects of a second dose of progestin, administered at 40 hours after the first dose

Earlier in Section 8.5, we discussed experimental results from [37] which investigate cell behaviour after a dose of progestin. These results showed that the first dose of progestin initially stimulated cells to enter the S phase, and later arrested cells in the G phase, in agreement with the biphasic nature of progestin observed in [84]. In the same paper, at 40 hours after the first dose of progestin, a second dose of progestin was given to the T47D cells, and subsequent changes in protein concentrations, as well as cell cycle phase proportion was investigated. Surprisingly, authors of [37] did not observe an increase in the S phase proportion after the second dose of progestin and instead, the second dose of progestin continued to keep the S phase proportion low. It is particularly surprising that the second dose of progestin does not cause an initial increase in the S phase proportion because after the first dose has arrested cells, a large proportion of cells are in the  $G_b$  model phase, and therefore a large number of cells would be primed to react to any increase in the proliferation rate.

Protein concentration changes, and cell cycle phase proportions after the second dose of progestin were also investigated in [37]. The results can be seen in Figure 8.11. At 40 hours, there is an arrow in each figure, which represents the time at which the second dose was administered. The changes in the protein concentrations after this second dose are shown as the white squares after this time. If we consider Figure 8.11(a), we can see that the second dose of progestin doesn't appear to change cyclin D1 concentrations

significantly compared to cells which only experienced one dose (compare black and white squares). However by 50 hours, cyclin D1 concentrations increase to levels above that of cells that only experienced one dose. Figure 8.11(b) shows that cyclin D3 concentrations increase to levels higher than in cell which only experienced the first dose, implying that overall, the second dose of progestin increases cyclin D concentrations.

Figure 8.11(c) shows that the second dose of progestin decreases cyclin E1 concentrations to levels below that of cells that only experienced the first dose. Cyclin E1 concentrations remain below levels of cells that only experienced one dose for the rest of the experiment.

Figure 8.11(d) shows that p27 concentrations are significantly higher in cells that experienced two doses of progestin compared to cells that only experienced one dose. The concentration of p27 remains significantly higher for the remainder of the experiment.

The S + G2/M phase proportion after the second dose is shown as the thin solid line after this time in each figure. As we can see, the second dose of progestin does not cause the transient increase in proliferation that is seen within the first few hours of the first dose of progestin, and instead, the second dose just continues to arrest cells. It is relevant that we understand why this occurs, because there is some controversy in the literature around the role of progesterone and progestin in cell proliferation. As we saw earlier, progesterone levels in the breast correlate with high proliferation (for instance, during the menstrual cycle, breast proliferation is high when progesterone is high), however some cell experiments find that progesterone is anti-proliferative [1]. The significance of this observation, that progestin effects depend upon a cell's previous history of progestin exposure, could help to explain how these conflicting observations have come about.

### Simulating the second dose of progestin after 40 hours

We wish to simulate protein concentration changes, as well as cell cycle phase proportions, if we administer a second dose of progestin at 40 hours after the first, as was done in [37]. We must decide how the second dose of progestin influences our input drivers,  $\alpha_d(t)$  and  $\phi_{pd}(t)$ , and we will begin by discussing how the second dose affects  $\alpha_d(t)$ . We note that after the first dose of progestin,  $\alpha_d(t)$  increased from its starting value of 1 to 1.5, and so

we could conclude that progestin contributed about 0.5 units per hour to the production of cyclin D. However, if we allow the same increase in cyclin D-CDK4 concentration after the second dose of progestin (i.e. allowing  $\alpha_d(t)$  to increase by the same amount (0.5 units) after the second dose), we will observe a much larger spike in the overall concentration of cyclin D within the first 10 hours after the second dose. This is because the first dose of progestin has resulted in a very high concentration of free p27 molecules by 40 hours after exposure, and the same increase in cyclin D-CDK4 molecules will bind with a much larger pool of free p27 molecules. This will result in an increase in the stable cyclin D-CDK4-p27 complex, resulting in a large increase in total cyclin D concentration. If we consider the experimental observations regarding the concentrations of cyclin D1 and cyclin D3 after the second dose of progestin, we note that there is no significant increase in cyclin D1, and that there is some increase in cyclin D3 after progestin exposure, meaning that we should choose  $\alpha_d(t)$  to increase by only a small amount after the second dose. Instead of assuming that progestin contributes 0.5 units of production to cyclin D, we could, alternatively, note that progestin increased  $\alpha_d(t)$  by 50% after the first dose of progestin, and similarly assume that  $\alpha_d(t)$  increases by 50% of the value at 40 hours after the second dose of progestin exposure (which amounts to an increase of about 0.05 units, significantly lower than our originally proposed amount of 0.5 units). However, we find that this is too small an increase to make any significant difference in protein concentrations after 40 hours. Thus, we will choose  $\alpha_d(t)$  to increase by about 0.25 units, which is in between the two possibilities just discussed. The amount of increase in  $\alpha_d(t)$  after the second dose of progestin is shown in Figure 8.15. After the first dose of progestin, we noted that  $\alpha_d(t)$  reduced to levels of about 2% of the starting concentration of cyclin D. We will also assume that  $\alpha_d(t)$  later reduces to about 2% of the value it took at 40 hours, when the second dose is given. The profile we choose for  $\alpha_d(t)$  after two doses of progestin (one at  $t = 0$  and the second at  $t = 40$  hours) is shown in Figure 8.15.

We must also determine a profile for  $\phi_{pd}(t)$ . We will assume that after the second dose, the proportion of cyclin D-CDK4-p27 that are phosphorylated decreases by the same amount as after the first dose of progestin. The profile for  $\phi_{pd}(t)$  after both doses of progestin is shown in Figure 8.16. In this profile, we can see that the proportion of active cyclin D-CDK4 molecules is very small after the second dose, and it is possible that, biologically, it cannot reach such low levels because other factors that influence the

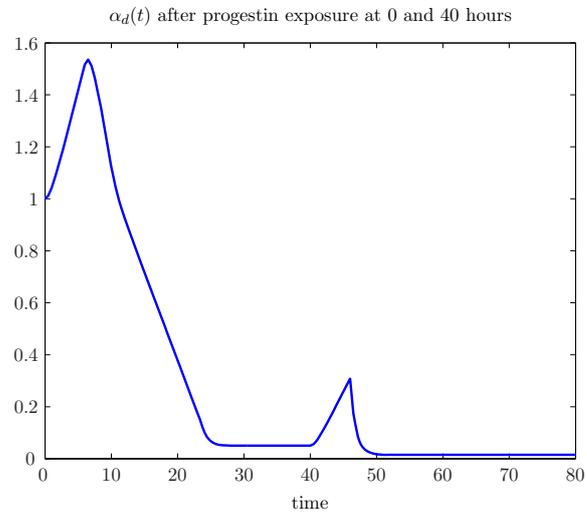


Figure 8.15: *The profile for  $\alpha_d(t)$  when progestin is administered at  $t = 0$  and  $t = 40$  hours.*

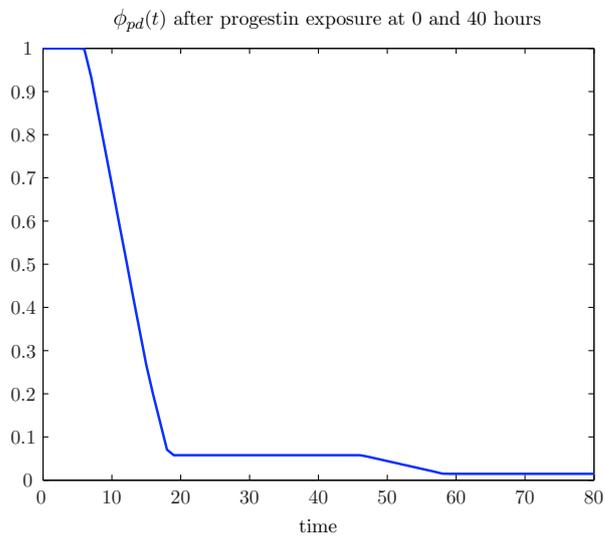


Figure 8.16: *The profile for  $\phi_{pd}(t)$  when progestin is administered at  $t = 0$  and  $t = 40$  hours.*

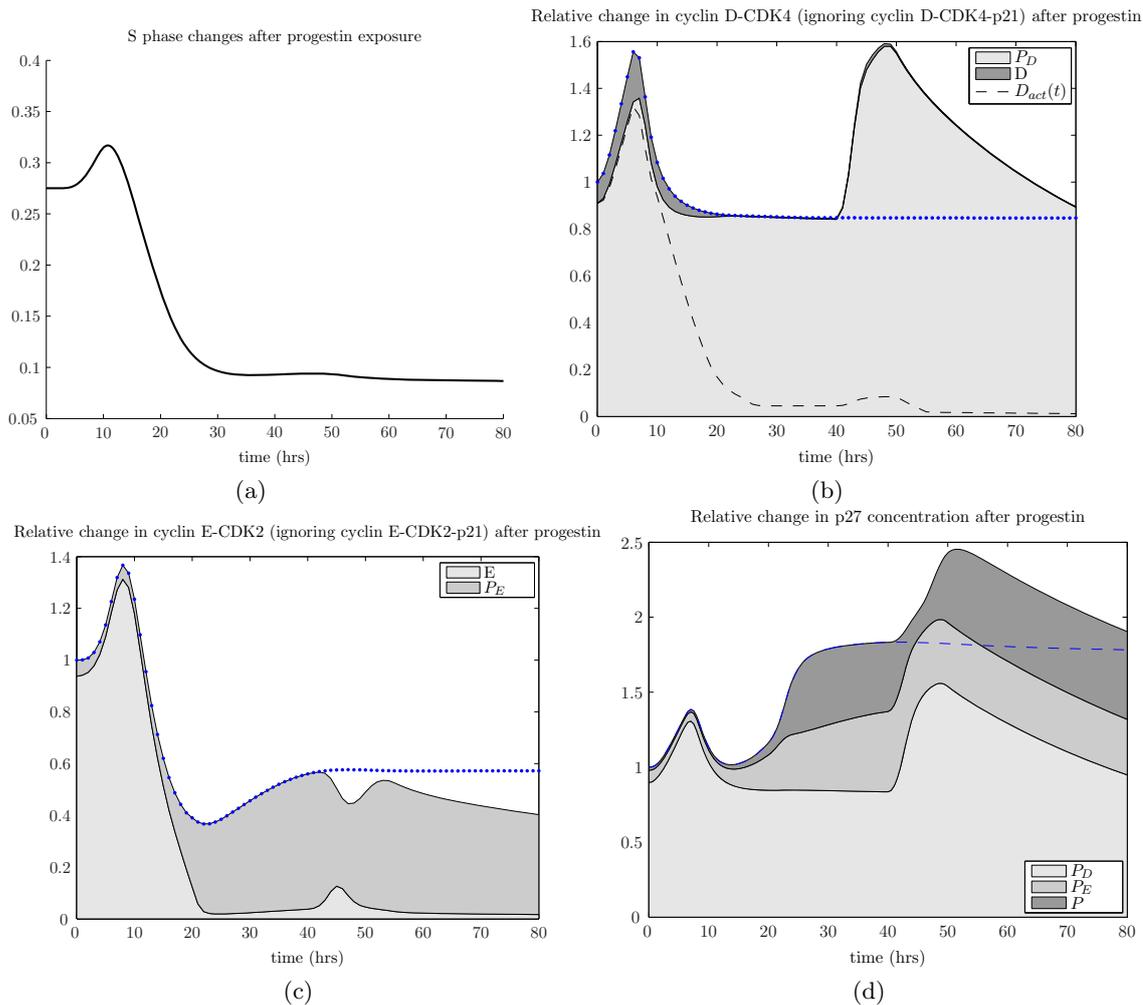


Figure 8.17: Modelled changes in (a) the S phase proportion, (b) cyclin E concentration, (c) cyclin D concentration and (d) p27 concentration after two doses of progestin - one at  $t = 0$  and the second at  $t = 40$  hours. The blue dotted curve represents changes in total protein concentration changes after only the first dose of progestin, and is shown for comparison.

cell may not allow the activity to get to such low levels. We will discuss other possible profiles for  $\phi_{pd}$  later when we discuss our model results.

Substituting these input drivers into equations (8.18) and allowing the simulation to run for 80 hours, gives the results shown in Figure 8.17. Figure 8.17(a) shows that the additional dose of progestin at 40 hours did not increase the S phase proportion significantly when compared to cells which only experienced one dose, in agreement with experimental observations from [37].

Figure 8.17(b) shows that cyclin D concentrations increased significantly after the

second dose of progestin, which is due to the increase in  $\alpha_d(t)$ . Although experimental results from [37] did not observe an increase in cyclin D1 after the second dose of progestin (from Figure 8.11), cyclin D3 concentrations did increase significantly, and so our modelled increase in cyclin D concentration is in reasonable agreement with experimental data from [37]. Our model results show that cyclin D concentration reached levels above that observed when the first dose was administered, even though  $\alpha_d(t)$  did not increase by as much as after the first dose. This is because there is more p27 in the system after 30 hours of the first dose of progestin, and so when we increase cyclin D-CDK4 after the second dose, the excess cyclin D-CDK4 molecules bind to more p27, creating more of the stable cyclin D-CDK4-p27 molecules, and so for some time after exposure, the concentration of total cyclin D increases significantly. The reason the increase in cyclin D did not cause an increase in the rate of S phase entry is because such a large amount of p27 is no longer Y88/Y89 phosphorylated after the first dose of progestin, and so the increase in cyclin D, which causes an increase in cyclin D-CDK4-p27 molecules, does not result in a significant increase in active cyclin D-CDK4-p27 molecules as evidenced by the black dashed curve shown in Figure 8.17(b).

Figure 8.17(c) shows that the concentration of total cyclin E molecules decreased after the second dose of progestin at 40 hours when compared to the concentration after only one dose (shown as the dashed blue curve in this figure), in good agreement with the experimental data from [37]. If we consider equation (8.18), then we note that the rate of production and degradation of cyclin E depends upon the proportion of cells in the  $G_c$  and S phase (neither of which have changed significantly after the second dose), however the concentration of active cyclin D-CDK4-p27 molecules has increased. This does result in an increase in the concentration of free cyclin E-CDK2 molecules, as evidenced by change in the light shaded region in Figure 8.17(c). The concentration of free cyclin E-CDK2 molecules reaches a maximum of about 10% of control levels within 10 hours after the second dose. We might expect this increase in free cyclin E-CDK2 molecules to result in an increase in the concentration of overall cyclin E. However, if we reconsider Figure 8.10, we note that  $E(t) = 0.1$  actually corresponds to a lower steady-state concentration of total cyclin E than when  $E(t) = 0.037$  (which is the value of  $E(t)$  after the first dose). Thus, we would expect total cyclin E concentrations to be *decreasing* during this time, even though free cyclin E-CDK2 concentrations are increasing, and this is what we are

seeing in our model results. After 45 hours, free cyclin E-CDK2 concentrations reduce again, going below levels observed at 40 hours after the first dose, due to the reduction in cyclin D-CDK4 activity after this time. This further decrease in free cyclin E-CDK2 concentration causes a decrease in total cyclin E. If we consider Figure 8.10, we note that  $E = 0.037$ , the value of  $E$  after the first dose of progestin exposure, is a local maximum in total cyclin E-CDK2 concentration, and so any further decrease in cyclin E-CDK2 concentration will result in a further decrease in total cyclin E-CDK2 concentrations. Thus, total cyclin E-CDK2 concentrations remain below levels observed after only one dose of progestin (see the blue curve in this figure), which is in agreement with results from [37].

As we just discussed, the initial decrease in cyclin E-CDK2 concentration after the second dose of progestin is due to an *increase* in free cyclin E-CDK2 concentration. However, the later decrease after 45 hours of exposure is due to a *decrease* in free cyclin E-CDK2 concentration because cyclin D-CDK4 activity is further reduced after this time. Recall that we experience a further reduction in cyclin D-CDK4 activity because we have chosen  $\phi_{pd}(t)$  to decrease below control levels at times after 40 hours of progestin exposure. However, if we assume that  $\phi_{pd}(t)$  has already reached its minimum at 40 hours after exposure (and does not decrease further after the second dose), then we do not observe this further decrease in total cyclin E-CDK2 concentrations after 45 hours. Total cyclin E-CDK2 concentration still decreases initially after the second dose due to the increase in cyclin D-CDK4 activity, however, as cyclin D-CDK4 activity approaches levels equivalent to that at 40 hours, total cyclin E-CDK2 concentrations increase again to reach levels observed just before the second dose. The fact that we require a further reduction on cyclin D-CDK4 activity to keep cyclin E-CDK2 concentrations below levels observed at 40 hours provides some evidence that the second dose of progestin further reduces cyclin D-CDK4 activity below levels observed at 40 hours because the changes in cyclin E-CDK2 concentration better match experimental observations.

Figure 8.17(d) shows that after the second dose of progestin at 40 hours, p27 concentration levels increased above those that experienced one dose only (compare total shaded region with blue dashed curve). This increase occurs because of the increase in concentration of cyclin D-CDK4 molecules which bind to p27, hence increasing the concentration of cyclin D-CDK4-p27 molecules. As we can see, the concentration of cyclin

D-CDK4-p27 molecules slowly declines after this time due to the reduction in  $\alpha_d(t)$ . It may seem surprising that cyclin D-CDK4-p27 concentration decreases so slowly, especially as we originally chose rate parameter values so that overall, cyclin D-CDK4 concentration would react quickly to changes in the environment. However, the rate of decay of cyclin D-CDK4 molecules depends on the current concentration of cyclin D-CDK4 molecules, which are very low after the first dose of progestin. Therefore, even though we have chosen a large rate parameter so that cyclin D-CDK4 concentration reaches steady-state quickly, the overall rate of decay of cyclin D-CDK4 molecules is slow because these molecules are at such a low concentration. It is perhaps biologically unlikely that such conditions would exist so as to make cyclin D-CDK4 concentrations this low. This is because there will always be some other cell processes that we are not considering that are likely to make some contribution to the concentration of cyclin D-CDK4 molecules.

### Summary

In the last two chapters, we have performed a thorough investigation of the effects of progestin on cell cycle progression. Initially, we considered experimental results from [84] in detail. We gave a more thorough analysis of their data on elution profiles, and managed to convert their data into a form that could be used for quantitative analysis. Doing this gave us some insight into the causes of progestin-induced cell cycle arrest. In particular, we observed that the concentration of cyclin D-CDK4 molecules does not decrease significantly after progestin exposure, even though it was observed that cyclin D-CDK4 activity did decrease after progestin exposure. This led us to conclude that cyclin D-CDK4-p27 molecules, which make up the bulk of the total pool of cyclin D-CDK4 molecules, must move from being active, to being inactive. Assuming that this is how progestin exerts its effects, we were able to develop a model of protein concentration changes within a cell after exposure to progestin, and connect this model with the model for cell cycle progression that we developed in earlier chapters. The model was able to accurately capture changes in internal cell protein concentrations, and to connect these internal cell protein concentrations to effects on cell cycle progression. In particular, we captured the biphasic effect of progestin, which initially stimulates cells to enter the cell cycle, but later inhibits them. In addition, we were able to model the effects of a second

dose of progestin on cell cycle progression, capturing the effect that the second dose does not result in an temporary increase in the rate of cell cycle progression, but continues to arrest cells. The fact that the effects of progestin on cell cycle progression depend upon the cell's history of progestin exposure has profound implications for progestin treatments, such as cHRT's.

The fact that our simulations compare so well with experimental data under different conditions indicates that we have captured the main mechanisms of progestin-induced cell cycle progression. To the authors knowledge, this is the first time that cell cycle progression has been modelled according to changes in internal cell cycle proteins such as cyclin D. Such a model has great potential to simulate cell cycle progression under conditions difficult to produce experimentally, and to compare results and test hypotheses, as was done in this section.

A similar model could be developed to investigate protein concentration changes in the MCF-7 cell line after exposure to estrogen, however we could not do this because, to our knowledge, there are no experiments which report elution profiles showing how protein concentrations change after exposure. If there were similar experiments in this cell line, then we could also make a protein model and combine that with our cell cycle model. If the data becomes available, we could model cell cycle progression after both estrogen and progesterone, which we could then use to investigate certain conditions where both proteins are present simultaneously. This could help answer many questions about how combined hormone replacement therapies influence cell division, and hence cancer development.



## Chapter 9

# Conclusion

This thesis uses mathematical modelling techniques to gain insight into biological processes. We have focussed on modelling the process of cell division in breast cancer cells, and have found that in a wide range of different experimental conditions, our models produce results which compare well with experimental data. We have also used our mathematical models to get a deeper understanding of the mechanisms that drive the experiments, providing an overall better understanding of breast cancer cell behaviour.

### 9.1 Thesis summary

In Chapter 3, we introduced a mathematical model of cell cycle progression through the G, S and M phases of the cell cycle. This model uses the concept of storage phases to account for parts of the cell cycle in which cells must always spend a mandatory amount of time, resulting in a set of continuous delay differential equations. The full model has many unknown parameters, however each model parameter is chosen carefully after a thorough investigation of the biological literature. The model is used to investigate the relationship between the average cell cycle duration and the population doubling time, and challenges the common view in the literature that the two expressions are one and the same. The model is also used to question the assumption that a ‘slowly-cycling population’ of cells is necessary to explain some experimental results.

We then extended our mathematical model to simulate cell cycle progression in the

presence of a range of substances. Instead of considering the influence of different substances on cell cycle progression independently, we devised a function (which we denoted  $\gamma$ ) which uses known changes in cyclin D and myc concentrations to drive progression past the R point. This means that different substances can be investigated in our cell cycle model if we understand how these substances alter the concentrations of these internal cell cycle proteins. Converting the effects of each substance into an effect on the proteins cyclin D and myc means that we can simulate a range of substances without having to re-parameterise our model to account for the different effects of each substance. The model was validated against a range of experiments which use a whole variety of substances to influence cell cycle progression, and the excellent fit to experimental data in these experiments allows for insight into the mechanisms involved in cell cycle progression. Substances we considered include estrogen, insulin, IGF and tamoxifen, and the model produces accurate results regarding the changes in the cell cycle phase proportions after exposure to these substances.

The model is not only used to generate cell cycle phase proportions using the inputs cyclin D and myc, but can use known phase proportions to solve the inverse problem for cyclin D and myc. The model is used in this way to determine how the concentration of myc changes after exposure to various anti-estrogens, something which, to our knowledge, has not been calculated previously in the literature. We later extend the model to include cell-cycle phase specific cell death due to the presence of tamoxifen. The investigations using this model extension allowed us to determine which phases cell death must occur after exposure to the common chemotherapy drug tamoxifen. We found, as others have, that cell death occurs after the  $G_b$  phase, and our model was able to further identify that cell death predominantly occurred in the  $G_c$  model phase. This result led to the conclusion that it may be beneficial to first stimulate the growth of cancer cells before exposing them to tamoxifen. This extension of the model has excellent potential for use in cancer treatments, as tamoxifen is a first-line cancer treatment for women with breast cancer.

We then use our model to investigate cell cycle progression in T47D cells in Chapter 6. We use a range of experiments to parameterise the model to this cell line, and follow with validation of the model against other experimental results, from which we were able to conclude that one of the experiments in the literature is clearly an outlier.

In Chapter 7, we further investigate the role of progestins in cell cycle progression. We use our cell cycle model to show that total cyclin D concentration changes are not a good predictor of cell cycle progression after exposure to progestins. We perform a thorough analysis of experimental results on the changes in protein concentrations and protein-protein complex associations after progestin exposure. It is also observed in a range of experiments that cyclin D activity does not compare at all well with cyclin D concentration after progestin exposure. We challenge some hypotheses about why cyclin D concentration and activity do not compare well, and suggest a new hypothesis based on the phosphorylation status of p27 molecules.

In Chapter 8, we built a mathematical model of the relevant proteins involved in progestin-induced changes in cell behaviour, and included mechanisms which allow us to test our hypothesis. The model parameter choices were chosen carefully, and mainly based on experimental data available in the literature. The model results compare well with experimental observations, and provided some insight into the mechanisms progestins use to cause cell cycle progression and also cell cycle arrest. We combined this new protein model with the cell cycle model developed in previous chapters, and were able to get a better understanding of how internal cell cycle proteins influence cell cycle progression after exposure to progestins. These investigations highlight the complex nature of progestins on cell cycle progression, and show that the observed effects of progestins depend on the current state of the cells, as well as whether a cell has recently been exposed to progestin.

## 9.2 Current avenues being explored for future work

We are currently investigating other avenues using our cell cycle model. We have been in further communication with cancer researchers from the Hanson Institute of Cancer Research, and we intend to extend the model to simulate cell cycle progression in another cell line called the ZR75 cell line. These cells represent normal breast cells more than MCF-7 and T47D cells. ZR75 cells divide much more slowly, and although this is one of the features which make them more like normal breast cells, it also makes them difficult to work with, as experiments need to be carried out over a longer period of time. Having an

accurate mathematical model of the cells of this cell line is useful because mathematical models can be simulated in a much shorter period of time (the cell cycle models we used in this thesis typically produced results within seconds).

Another project we are currently working on is addressing the question ‘is the MCF-7 breast cancer cell line driven by a small population of stem cells’? In Chapter 4 we find that the results of the experiments could be explained without assuming there was a population of slowly dividing cells (i.e. without the need to assume that a stem cell population existed), however these results did not rule out that such a population could be present in this cell line. We wish to use our cell cycle model to determine whether an experiment could be constructed that would give conclusive evidence to the presence of such a cell population in this cell line.

### **9.3 Longer term views for future work**

As discussed in Chapter 2, normal breast tissue can be classified into two main types of cells: cells which do not express the estrogen receptor (ER) or the progesterone receptor (PR) (making up about 90% of the total breast cell population), and cells which (mainly) express both ER and PR (making up about 10% of the cell population). In the MCF-7 breast cancer cell line, all the cells are ER positive, and all these cells proliferate in response to estrogen. In the normal breast, however, exposure to estrogen causes the small number of ER positive cells to release growth factors that stimulate the growth of nearby ER negative cells - a significantly different mechanism of action. It would be interesting to extend the model to include the mechanisms of behaviour that better resemble the normal breast. However, without more experimental data on certain parameters (such as how far the stimulated ER positive cells release their subsequent growth signals, or how strong such signals are at causing proliferation), such a task remains difficult.

#### **Understanding the role of progesterone in the normal human breast**

As we saw in Chapter 2, normal breast cells do not express the progesterone receptor (PR), but will express this receptor after exposure to estrogen. This means that cells will not be responsive to progesterone unless they have recently been exposed to estrogen, and so this

makes it difficult to separate the effects of progesterone from those of estrogen. Having an accurate mathematical model of the effects of progesterone on cell cycle progression is useful because it can be extended to resemble the situation in the normal human breast. In particular, understanding the effects of progesterone given the cell environment in normal breast tissue (only 10% of cells express the ER and the PR), the model could be applied to this situation, and a spatial model could be developed. We anticipate that cell cycle progression would appear very different in such a make-up of cells.

### **Understanding breast cell proliferation during the normal menstrual cycle**

How breast cell proliferation fluctuates during the menstrual cycle due to the changing concentrations of estrogen and progesterone is not something that is fully understood. It has been observed that breast cell proliferation is at its peak when progesterone is at its highest level in the cycle, and not when estrogen is at its highest level. This observation is not in agreement with the currently understood role of these two hormones, which is that estrogen is a strong driver of proliferation, and progesterone is a strong driver of cell differentiation. Building a model of the effects of these hormones in normal breast tissue, including possible delays of their effects, could provide a deeper insight into why breast proliferation appears highest during the period when the level of progesterone is highest.

### **Gaining a better understanding of how therapies influence breast cancer risk**

Understanding the role of progestins and progesterone in the normal human breast would greatly aid in understanding how drugs, such as those used in hormone replacement therapies, can increase the risk of developing breast cancer. It is known that the presence of progestins in hormone replacement therapies contribute to the observed increase in breast cancer incidence, as older hormone replacement therapies which did not include progestins did not increase the incidence of breast cancer. Combining a model of the effects of estrogens and progestins on normal breast cell proliferation would greatly benefit an understanding of how these hormones can influence cancer.

One of the research projects that the team at the Hanson Institute of Cancer Research is working on is determining the effects of androgens in breast cell proliferation. They

found that the androgen DHT acts to inhibit the proliferative actions of estrogen. It is also known that in women with breast cancer, those whose breast cancer cells are androgen-receptor positive, have a better clinical outcome than those who aren't [91]. Researchers at the Hanson Institute have also found that the progestin MPA (which is an active hormone present in combined hormone replacement therapies) can prevent DHT from binding to the androgen receptor, thus blocking the anti-estrogenic properties of DHT. It is possible that this hormone is also critical in the effects of combined hormone replacement therapies in breast cancer.

## 9.4 Final words

This thesis has advanced our understanding of the influence of a variety of growth substances and anti-growth substances on cell cycle progression. The application of mathematical models to the large amount of experimental data available in the literature has provided deep insight into the data generated through these experiments. With more interaction between experimentalists and mathematicians, we believe that a better understanding of the mechanisms behind cell cycle progression, and ultimately cancer and its treatment, can be attained.

# Glossary

**activity** Protein activity is the measure of a protein's ability to interact with other cell structures to cause direct changes in cell behaviour.

**affinity** A measure of the overall strength with which two proteins bind together.

**angiogenesis** Process by which new blood vessels are formed.

**apoptosis** Program of cellular self-destruction.

**basement membrane** A specialized layer that separates epithelial cells from stromal cells.

**benign** A mass of cells that have formed. The cells are not considered to be dangerous and in many cases, can never become cancerous.

**carcinoma** A cancer arising from epithelial cells.

**CDK's** Proteins called cyclin dependent kinases that are critical in the cell division process. To become active, they must bind to cyclins.

**chemokine** A chemical message that serves as an attractant for motile cells.

**chromosome** A very long DNA molecule that is condensed into a rod-like structure within the nucleus of a cell.

**colonisation** The process whereby a cancer cell manages to successfully grow in a part of the body far from the original site.

**crisis** A cell enters crisis when the telomeres are gone from the ends of its chromosomes due to shortening during cell division. The chromosome ends fuse together when the telomeres are gone.

**cyclins** Proteins critical in the cell division process. When they bind to CDK's, the CDK's become active.

**differentiated** A very specialised cell, which has often become senescent (e.g. nerve cells).

**DNA** A chain of deoxyribonucleotide units, which contain hereditary information about the organism.

**DNA histogram** A graphical representation of the average amount of DNA per cell in a sample, determined using flow cytometry.

**endothelial cells** Cell that form the lining of blood vessels or lymph ducts.

**eukaryotic cell** A cell which contains a distinct nucleus and cytoplasm. All living things, except for viruses and bacteria, are made up of eukaryotic cells.

**differentiation** Process whereby a cell acquires a specialised phenotype.

**extracellular matrix** Mesh of secreted proteins that surrounds most cells in a tissue. It creates structure in the intercellular space.

**epithelial cells** Cells that line an organ.

**fibroblast** Cell found in the connective tissue and stromal compartment of epithelial tissues.

**fluorescence detection** A process of detecting the intensity of light that is emitted after passing through a cell. This is used to determine information about a cell, such as the concentration of DNA content within a cell.

**flow cytometry** A process of examining internal cell constituents by passing them through a fluid, which streamlines each cell so that they pass through a detector one at a time.

**gene** A segment of DNA that codes for a single protein.

**genotype** Genetic constitution of an organism.

**hematopoietic cells** Cells which form the blood, including red blood cells, white blood cells and endothelial cells.

- haptotaxis** Directional migration of cells in response to an adhesion gradient.
- heterotypic** Referring to the interactions between two or more distinct cellular types.
- hyperplasia** Abnormal proliferation of cells, not yet considered to be dangerous.
- hypoxia** State of lower-than-normal oxygen tension.
- immortal** A cell is immortal if it has infinite replicative potential e.g. see stem cell.
- in vitro*** Occurring in a vessel (e.g. a petri dish) rather than in living tissue.
- in vivo*** Occurring in a living organism.
- invasion** Process whereby cancer cells move from a primary tumour into adjacent normal tissue. In the case of carcinomas, this involves breaching of the basement membrane.
- lysosome** A structure within a cell that breaks down waste and debris.
- macrophage** A type of white blood cell that ingests foreign material.
- malignant** A group of cells which are considered cancerous. In the case of carcinomas, this is because a mass of abnormal cells have invaded the basement membrane. Also refers to more advanced cancers.
- matrix metalloproteinase** A molecule designed to cleave the extracellular matrix.
- mesenchymal cells** Cells of the connective tissue as well as hematopoietic cells.
- metastasis** A malignant growth forming at a site in the body made up of cells from a malignancy located elsewhere.
- mitochondria** A structure within a cell that produces ATP, the cells chemical energy unit.
- mitogen** An agent that promotes cellular proliferation.
- mutation** Change in the genotype of a species.
- monocytes** Cells constituting functional muscles
- myocytes** The cells constituting functional muscles.

**myofibroblasts** A type of fibroblast normally involved in inflammation.

**necrosis** Cell death involving the breakdown of a cell through steps different from those in apoptosis.

**neuroectodermal cancer** Cancer of the nervous system.

**nucleotide** The molecules that join together to make up DNA.

**oncogene** A cancer causing gene.

**phagocytosis** The process whereby a macrophage cell engulfs an apoptotic cell to remove the dying cell from its environment.

**phenotype** An observable trait of an organism.

**phosphorylation** When a phosphate group is added to a protein. Phosphorylation either activates, de-activates or tags the protein for degradation.

**proliferation** The process of cell division.

**quiescent** A state where a cell has fully differentiated or has no plan to divide for a long period of time. Such cells still have the ability to divide if environmental conditions encourage it to do so.

**receptor** Protein found on the surface of a cell which emits signals to the inside of a cell in response to the external environment.

**restriction point** (or R point) A cell which passes the restriction point has made an irreversible commitment to cell division. It exists in the G phase.

**RNA** A copy of the code from DNA, usually produced as an intermediate step between reading the DNA and creating the protein.

**ribosome** A structure within a cell that creates proteins from mRNA templates.

**sarcoma** Tumour derived from mesenchymal cells.

**senescence** A state where the cell can no longer divide e.g. due to differentiation or crisis.

**stem cells** Cells which have infinite replicative potential. In particular, they regularly regenerate their telomeres so they never enter crisis.

**stroma** The mesenchymal components or blood tissue and epithelial tissue.

**telomere** The section of DNA at the end of the chromosome that exists to protect the chromosomal ends from fusing together.

**transcription** The process of making the RNA template for a protein by reading the code from the DNA.

**transit-amplifying cell** the progeny of stem cells which divide only a finite number of times. Almost all cells in an organism are transit-amplifying.

**transformation** Process of converting a normal cell into having some or many of the attributes of a cancer cell.

**tumour suppressor gene** A gene whose partial or complete inactivation leads to an increased likelihood of cancer development.

**ubiquitination** A process of protein degradation commonly carried on internal cell proteins.

# Bibliography

- [1] T.M. Ahola, T. Manninen, N. Alkio, and T. Ylikomi. G protein-coupled receptor 30 is critical for a progestin-induced growth inhibition in MCF-7 breast cancer cells. *Endocrinology*, 143(9):3376, 2002.
- [2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular biology of the cell*. Garland Science, 2002.
- [3] B. Amati, K. Alevizopoulos, J. Vlach, et al. Myc and the cell cycle. *Front Biosci*, 3:D250–D268, 1998.
- [4] S.E. Artandi and R.A. DePinho. Telomeres and telomerase in cancer. *Carcinogenesis*, 31(1):9, 2010.
- [5] JA Barnes, DJ Dix, BW Collins, C. Luft, and JW Allen. Expression of inducible Hsp70 enhances the proliferation of MCF-7 breast cancer cells and protects against the cytotoxic effects of hyperthermia. *Cell Stress & Chaperones*, 6(4):316–325, 2001.
- [6] B. Basse, B.C. Baguley, E.S. Marshall, W.R. Joseph, B. van Brunt, G. Wake, and D.J.N. Wall. A mathematical model for analysis of the cell cycle in cell lines derived from human tumors. *Journal of Mathematical Biology*, 47(4):295–312, 2003.
- [7] B. Basse, B.C. Baguley, E.S. Marshall, W.R. Joseph, B. van Brunt, G. Wake, and D.J.N. Wall. Modelling cell death in human tumour cell lines exposed to the anti-cancer drug paclitaxel. *Journal of Mathematical Biology*, 49(4):329–357, 2004.
- [8] B. Basse, B.C. Baguley, E.S. Marshall, G.C. Wake, and D.J.N. Wall. Modelling the flow of cytometric data obtained from unperturbed human tumour cell lines:

- parameter fitting and comparison. *Bulletin of Mathematical Biology*, 67(4):815–830, 2005.
- [9] B. Basse and P. Ubezio. A generalised age-and phase-structured model of human tumour cell populations both unperturbed and exposed to a range of cancer therapies. *Bulletin of Mathematical Biology*, 69(5):1673–1690, 2007.
- [10] R. Begg, D.J.N. Wall, and G.C. Wake. On a multicompartment age-distribution model of cell growth. *IMA Journal of Applied Mathematics*, 75(6):905, 2010.
- [11] F. Bekkal Brikci, J. Clairambault, B. Ribba, and B. Perthame. An age-and-cyclin-structured cell population model for healthy and tumoral tissues. *Journal of Mathematical Biology*, 57(1):91–110, 2008.
- [12] R.A. Blundell. The Biology of p21Waf1/Cip1-Review Paper. *American Journal of Biochemistry and Biotechnology*, 2(1):33–40, 2006.
- [13] C. Brisken and S. Duss. Stem cells and the stem cell niche in the breast: an integrated hormonal and developmental perspective. *Stem Cell Reviews and Reports*, 3(2):147–156, 2007.
- [14] N.F. Britton. *Essential mathematical biology*. Springer Verlag, 2003.
- [15] F. Bunz, A. Dutriaux, C. Lengauer, T. Waldman, S. Zhou, JP Brown, JM Sedivy, KW Kinzler, and B. Vogelstein. Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science*, 282(5393):1497, 1998.
- [16] C.E. Caldon, C.M. Sergio, J. Schutte, M.N. Boersma, R.L. Sutherland, J.S. Carroll, and E.A. Musgrove. Estrogen regulation of cyclin E2 requires cyclin D1 but not c-Myc. *Molecular and Cellular Biology*, 29(17):4623, 2009.
- [17] C. Campagnoli, F. Clavel-Chapelon, R. Kaaks, C. Peris, and F. Berrino. Progestins and progesterone in hormone replacement therapy and the risk of breast cancer. *Journal of Steroid Biochemistry and Molecular Biology*, 96(2):95–108, 2005.
- [18] J. Campisi. The biology of replicative senescence. *European Journal of Cancer*, 33(5):703–709, 1997.

- [19] V. Cappelletti, L. Fioravanti, P. Miodini, and G. Di Fronzo. Genistein blocks breast cancer cells in the G2M phase of the cell cycle. *Journal of Cellular Biochemistry*, 79(4):594–600, 2000.
- [20] J.S. Carroll, A. Swarbrick, E.A. Musgrove, and R.L. Sutherland. Mechanisms of growth arrest by c-myc antisense oligonucleotides in MCF-7 breast cancer cells implications for the antiproliferative effects of antiestrogens. *Cancer Research*, 62(11):3126–3131, 2002.
- [21] I. Chu, K. Blackwell, S. Chen, and J. Slingerland. The dual ErbB1/ErbB2 inhibitor, lapatinib (GW572016), cooperates with tamoxifen to inhibit both cell proliferation and estrogen-dependent gene expression in antiestrogen-resistant breast cancer. *Cancer Research*, 65(1):18–25, 2005.
- [22] M.F. Clarke and M. Fuller. Stem cells and cancer: two faces of eve. *Cell*, 124(6):1111–1115, 2006.
- [23] RB Clarke. Human breast cell proliferation and its relationship to steroid receptor expression. *Climacteric*, 7(2):129–137, 2004.
- [24] R.B. Clarke, A. Howell, C.S. Potten, and E. Anderson. Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Research*, 57(22):4987, 1997.
- [25] D. Curran. Menopause, 1994-2011.
- [26] M. Dalvai and K. Bystricky. Cell cycle and anti-estrogen effects synergize to regulate cell proliferation and ER target gene expression. *PloS ONE*, 5(6):e11011, 2010.
- [27] Z. Darzynkiewicz, S. Bruno, G. Del Bino, W. Gorczyca, MA Hotz, P. Lassota, and F. Traganos. Features of apoptotic cells measured by flow cytometry. *Cytometry Part A*, 13(8):795–808, 1992.
- [28] L. Daukste, B. Basse, B.C. Baguley, and D.J.N. Wall. Using a stem cell and progeny model to illustrate the relationship between cell cycle times of in vivo human tumour cell tissue populations, in vitro primary cultures and the cell lines derived from them. *Journal of Theoretical Biology*, 260(4):563–571, 2009.

- [29] FR De Gruijl. Skin cancer and solar UV radiation. *European Journal of Cancer*, 35(14):2003–2009, 1999.
- [30] Y. Deng and S. Chang. Role of telomeres and telomerase in genomic instability, senescence and cancer. *Laboratory Investigation*, 87(11):1071–1076, 2007.
- [31] J.A. Diehl, M. Cheng, M.F. Roussel, and C.J. Sherr. Glycogen synthase kinase-3 $\beta$  regulates cyclin D1 proteolysis and subcellular localization. *Genes & development*, 12(22):3499, 1998.
- [32] T. Enoch and C. Norbury. Cellular responses to DNA damage: cell-cycle checkpoints, apoptosis and the roles of p53 and ATM. *Trends in Biochemical Sciences*, 20(10):426–430, 1995.
- [33] M. Esteller. Epigenetics in cancer. *New England Journal of Medicine*, 358(11):1148–1159, 2008.
- [34] N.A.P. Franken, H.M. Rodermond, J. Stap, J. Haveman, and C. van Bree. Clonogenic assay of cells in vitro. *Nature Protocols*, 1(5):2315–2319, 2006.
- [35] H. Fuss, W. Dubitzky, C.S. Downes, and M.J. Kurth. Mathematical models of cell cycle regulation. *Briefings in Bioinformatics*, 6(2):163–177, 2005.
- [36] M. Grimmmler, Y. Wang, T. Mund, Z. Cilensek, E.M. Keidel, M.B. Waddell, H. Jäkel, M. Kullmann, R.W. Kriwacki, and L. Hengst. Cdk-inhibitory activity and stability of p27Kip1 are directly regulated by oncogenic tyrosine kinases. *Cell*, 128(2):269–280, 2007.
- [37] S.D. Groshong, G.I. Owen, B. Grimison, I.E. Schauer, M.C. Todd, T.A. Langan, R.A. Scrafani, C.A. Lange, and K.B. Horwitz. Biphasic Regulation of Breast Cancer Cell Growth by Progesterone: Role of the Cyclin-Dependent Kinase Inhibitors, p21 and p27Kip1. *Molecular Endocrinology*, 11(11):1593–1607, 1997.
- [38] D. Hanahan and R.A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, 2011.
- [39] D. Hanahan, R.A. Weinberg, et al. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.

- [40] JW Harper, SJ Elledge, K. Keyomarsi, B. Dynlacht, LH Tsai, P. Zhang, S. Dobrowolski, C. Bai, L. Connell-Crowley, E. Swindell, et al. Inhibition of cyclin-dependent kinases by p21. *Molecular Biology of the Cell*, 6(4):387–400, 1995.
- [41] Y. Hathout, M.L. Gehrman, A. Chertov, and C. Fenselau. Proteomic phenotyping: metastatic and invasive breast cancer. *Cancer Letters*, 210(2):245–253, 2004.
- [42] T. Helleday, E. Petermann, C. Lundin, B. Hodgson, and R.A. Sharma. DNA repair pathways as targets for cancer therapy. *Nature Reviews Cancer*, 8(3):193–204, 2008.
- [43] K. Hemminki. DNA adducts, mutations and cancer. *Carcinogenesis*, 14(10):2007, 1993.
- [44] D.A. Hill, N.S. Weiss, S.A.A. Beresford, L.F. Voigt, J.R. Daling, J.L. Stanford, and S. Self. Continuous combined hormone replacement therapy and risk of endometrial cancer. *American Journal of Obstetrics and Gynecology*, 183(6):1456–1461, 2000.
- [45] L.C. Hodges, J.D. Cook, E.K. Lobenhofer, L. Li, L. Bennett, P.R. Bushel, C.M. Aldaz, C.A. Afshari, and C.L. Walker. Tamoxifen functions as a molecular agonist inducing cell cycle-associated genes in breast cancer cells. *Molecular Cancer Research*, 1(4):300–311, 2003.
- [46] J.H.J. Hoeijmakers et al. Genome maintenance mechanisms for preventing cancer. *Nature*, 411(6835):366–374, 2001.
- [47] A. Hurwitz. On the conditions under which an equation has only roots with negative real parts. *Selected Papers on Mathematical Trends in Control Theory, Edited by R. Bellman and R. Kalaba, Dover, New York*, 1964.
- [48] A. Ichikawa, A. Jiro, and S. Koichi. G1 arrest and expression of cyclin-dependent kinase inhibitors in tamoxifen-treated MCF-7 human breast cancer cells. *Human Cell*, 21(2):28–37, 2008.
- [49] National Cancer Institute. Common cancer types, July 2011.
- [50] M.K. James, A. Ray, D. Leznova, and S.W. Blain. Differential modification of p27Kip1 controls its cyclin D-cdk4 inhibitory activity. *Molecular and Cellular Biology*, 28(1):498–510, 2008.

- [51] AM Jastreboff and T. Cymet. Role of the human papilloma virus in the development of cervical intraepithelial neoplasia and malignancy. *Postgraduate Medical Journal*, 78(918):225, 2002.
- [52] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M.J. Thun. Cancer statistics, 2009. *CA: A Cancer Journal for Clinicians*, 59(4):225–249, 2009.
- [53] V.C. Jordan, S. Gapstur, and M. Morrow. Selective estrogen receptor modulation and reduction in risk of breast cancer, osteoporosis, and coronary heart disease. *Journal of the National Cancer Institute*, 93(19):1449, 2001.
- [54] A. Kamb, N.A. Gruis, J. Weaver-Feldhaus, Q. Liu, K. Harshman, S.V. Tavtigian, E. Stockert, RS Day, B.E. Johnson, and M.H. Skolnick. A cell cycle regulator potentially involved in genesis of many tumor types. *Science*, 264(5157):436, 1994.
- [55] A. Kariagina, M.D. Aupperlee, and S.Z. Haslam. Progesterone receptor isoform functions in normal breast development and breast cancer. *Critical Reviews in Eukaryotic Gene Expression*, 18(1):11, 2008.
- [56] T. Kirkegaard, C.J. Witton, L.M. McGlynn, S.M. Tovey, B. Dunne, A. Lyon, and J. Bartlett. AKT activation predicts outcome in breast cancer patients treated with tamoxifen. *The Journal of pathology*, 207(2):139–146, 2005.
- [57] J.A. Knoblich. Mechanisms of asymmetric stem cell division. *Cell*, 132(4):583–597, 2008.
- [58] G. Koopman, CP Reutelingsperger, GA Kuijten, RM Keehnen, ST Pals, and MH Van Oers. Annexin V for flow cytometric detection of phosphatidylserine expression on B cells undergoing apoptosis. *Blood*, 84(5):1415–1420, 1994.
- [59] H. Kuper, P. Boffetta, and H.O. Adami. Tobacco use and cancer causation: association by tumour type. *Journal of internal medicine*, 252(3):206–224, 2002.
- [60] A. Lai, B. Sarcevic, O.W.J. Prall, and R.L. Sutherland. Insulin/insulin-like growth factor-I and estrogen cooperate to stimulate cyclin E-CDK2 activation and cell cycle progression in MCF-7 breast cancer cells through differential regulation of cyclin E and p21WAF1/CIP1. *Journal of Biological Chemistry*, 276(28):25823–25833, 2001.

- [61] C.A. Lange. Challenges to defining a role for progesterone in breast cancer. *Steroids*, 73(9-10):914–921, 2008.
- [62] R.G. Lapidus, S.J. Nass, and N.E. Davidson. The loss of estrogen and progesterone receptor gene expression in human breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, 3(1):85–94, 1998.
- [63] S. Larsson, T. Ryden, U. Holst, S. Oredsson, and M. Johansson. Estimating the distribution of the G2 phase duration from flow cytometric histograms. *Mathematical Biosciences*, 211(1):1–17, 2008.
- [64] S. Larsson, T. Ryden, U. Holst, S. Oredsson, and M. Johansson. Estimating the variation in S phase duration from flow cytometric histograms. *Mathematical Biosciences*, 213(1):40–49, 2008.
- [65] C.I. Li, K.E. Malone, P.L. Porter, N.S. Weiss, M.T.C. Tang, K.L. Cushing-Haugen, and J.R. Daling. Relationship between long durations and different regimens of hormone therapy and risk of breast cancer. *JAMA: the Journal of the American Medical Association*, 289(24):3254, 2003.
- [66] G. Li and VC Ho. p53-dependent DNA repair and apoptosis respond differently to high-and low-dose ultraviolet radiation. *British Journal of Dermatology*, 139(1):3–10, 1998.
- [67] M. Martinez-Alonso, E. Vilaprincho, R. Marcos-Gragera, and M. Rue. Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Research*, 2010, vol. 12, núm. 4, p. 58, 2010.
- [68] R.S. Maser and R.A. DePinho. Connecting chromosomes, crisis, and cancer. *Science*, 297(5581):565, 2002.
- [69] A. Mawson, A. Lai, J.S. Carroll, C.M. Sergio, C.J. Mitchell, and B. Sarcevic. Estrogen and insulin/IGF-1 cooperatively stimulate cell cycle progression in MCF-7 breast cancer cells through differential regulation of c-Myc and cyclin D1. *Molecular and Cellular Endocrinology*, 229(1-2):161–173, 2005.

- [70] FE May and BR Westley. Effects of tamoxifen and 4-hydroxytamoxifen on the pNR-1 and pNR-2 estrogen-regulated RNAs in human breast cancer cells. *Journal of Biological Chemistry*, 262(33):15894–15899, 1987.
- [71] S.R. McDougall, A.R.A. Anderson, and M.A.J. Chaplain. Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies. *Journal of Theoretical Biology*, 241(3):564–589, 2006.
- [72] D.W. Meek. Tumour suppression by p53: a role for the DNA damage response? *Nature Reviews Cancer*, 9(10):714–723, 2009.
- [73] M. Meyerson. Role of telomerase in normal and cancer cells. *Journal of Clinical Oncology*, 18(13):2626, 2000.
- [74] A. Montagnoli, F. Fiore, E. Eytan, A.C. Carrano, G.F. Draetta, A. Hershko, and M. Pagano. Ubiquitination of p27 is regulated by Cdk-dependent phosphorylation and trimeric complex formation. *Genes & Development*, 13(9):1181–1189, 1999.
- [75] K.A. Moore and I.R. Lemischka. Stem cells and their niches. *Science*, 311(5769):1880, 2006.
- [76] J. Moreira and A. Deutsch. Cellular Automation Models of Tumor Development: A Critical Review. *Advances in Complex Systems*, 5(2/3):247–268, 2002.
- [77] S.J. Morrison and J. Kimble. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature*, 441(7097):1068–1074, 2006.
- [78] E. A. Musgrove. *Hormonal control of breast cancer cell cycle progression*. PhD thesis, University of New South Wales, 1991.
- [79] EA Musgrove, JA Hamilton, CS Lee, KJ Sweeney, CK Watts, and RL Sutherland. Growth factor, steroid, and steroid antagonist regulation of cyclin gene expression associated with changes in T-47D human breast cancer cell cycle progression. *Molecular and Cellular Biology*, 13(6):3577–3587, 1993.
- [80] E.A. Musgrove, C.S. Lee, and R.L. Sutherland. Progestins both stimulate and inhibit breast cancer cell cycle progression while increasing expression of transforming growth factor alpha, epidermal growth factor receptor, c-fos, and c-myc genes. *Molecular and Cellular Biology*, 11(10):5032–5043, 1991.

- [81] E.A. Musgrove, C.S.L. Lee, M.F. Buckley, and R.L. Sutherland. Cyclin D1 induction in breast cancer cells shortens G1 and is sufficient for cells arrested in G1 to complete the cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 91(17):8022–8026, 1994.
- [82] E.A. Musgrove, B. Sarcevic, and R.L. Sutherland. Inducible expression of cyclin D1 in T-47D human breast cancer cells is sufficient for Cdk2 activation and pRB hyperphosphorylation. *Journal of Cellular Biochemistry*, 60(3):363–378, 1998.
- [83] E.A. Musgrove and R.L. Sutherland. Acute effects of growth factors on T-47D breast cancer cell cycle progression. *European Journal of Cancer*, 29(16):2273–2279, 1993.
- [84] E.A. Musgrove, A. Swarbrick, C.S.L. Lee, A.L. Cornish, and R.L. Sutherland. Mechanisms of cyclin-dependent kinase inactivation by progestins. *Molecular and Cellular Biology*, 18(4):1812–1825, 1998.
- [85] E.A. Musgrove, A.E. Wakeling, and R.L. Sutherland. Points of action of estrogen antagonists and a calmodulin antagonist within the MCF-7 human breast cancer cell cycle. *Cancer Research*, 49(9):2398–2404, 1989.
- [86] S. Muskopf. Biology Corner. Website, August 2011.
- [87] H.D. Nelson, L.L. Humphrey, P. Nygren, S.M. Teutsch, and J.D. Allan. Postmenopausal hormone replacement therapy. *JAMA: the Journal of the American Medical Association*, 288(7):872, 2002.
- [88] B. Novak and J.J. Tyson. A model for restriction point control of the mammalian cell cycle. *Journal of Theoretical Biology*, 230(4):563–579, 2004.
- [89] M.N. Obeyesekere, J.R. Herbert, and S.O. Zimmerman. A model of the G1 phase of the cell cycle incorporating cyclin E/CDK2 complex and retinoblastoma protein. *Oncogene*, 11(6):1199, 1995.
- [90] National Institute of General Medical Sciences. The New Genetics, July 2011.
- [91] Y. Ogawa, E. Hai, K. Matsumoto, K. Ikeda, S. Tokunaga, H. Nagahara, K. Sakurai, T. Inoue, and Y. Nishiguchi. Androgen receptor expression in breast cancer:

- relationship with clinicopathological factors and biomarkers. *International Journal of Clinical Oncology*, 13(5):431–435, 2008.
- [92] R.J. O’Sullivan and J. Karlseder. Telomeres: protecting chromosomes against genome instability. *Nature reviews. Molecular cell biology*, 11(3):171, 2010.
- [93] R.R. Perry, Y. Kang, and B. Greaves. Effects of tamoxifen on growth and apoptosis of estrogen-dependent and-independent human breast cancer cells. *Annals of Surgical Oncology*, 2(3):238–245, 1995.
- [94] M.C. Pike, D.V. Spicer, L. Dahmouch, and M.F. Press. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. *Epidemiologic Reviews*, 15(1):17, 1993.
- [95] P.L. Porter, W.E. Barlow, I. Yeh, et al. p27Kip1 and cyclin E expression and breast cancer survival after treatment with adjuvant chemotherapy. *Journal of the National Cancer Institute*, 98(23):1723, 2006.
- [96] C.S. Potten and M. Loeffler. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. *Development*, 110(4):1001, 1990.
- [97] O.W.J. Prall, E.M. Rogan, E.A. Musgrove, C.K.W. Watts, and R.L. Sutherland. c-Myc or cyclin D1 mimics estrogen effects on cyclin E-Cdk2 activation and cell cycle reentry. *Molecular and Cellular Biology*, 18(8):4499–4508, 1998.
- [98] O.W.J. Prall, B. Sarcevic, E.A. Musgrove, C.K.W. Watts, and R.L. Sutherland. Estrogen-induced activation of Cdk4 and Cdk2 during G1-S phase progression is accompanied by increased cyclin D1 expression and decreased cyclin-dependent kinase inhibitor association with cyclin E-Cdk2. *Journal of Biological Chemistry*, 272(16):10882–10894, 1997.
- [99] Z. Qu, W.R. MacLellan, and J.N. Weiss. Dynamics of the cell cycle: checkpoints, sizers, and timers. *Biophysical journal*, 85(6):3600–3611, 2003.
- [100] P.M. Ravdin, K.A. Cronin, N. Howlader, C.D. Berg, R.T. Chlebowski, E.J. Feuer, B.K. Edwards, and D.A. Berry. The decrease in breast-cancer incidence in 2003 in the United States. *New England Journal of Medicine*, 356(16):1670–1674, 2007.

- [101] R.R. Reddel, I.E. Alexander, M. Koga, J. Shine, and R.L. Sutherland. Genetic instability and the development of steroid hormone insensitivity in cultured T47D human breast cancer cells. *Cancer Research*, 48(15):4340–4347, 1988.
- [102] D. Rivenzon-Segal, E. Rushkin, S. Polak-Charcon, and H. Degani. Glucose transporters and transport kinetics in retinoic acid-differentiated T47D human breast cancer cells. *American Journal of Physiology- Endocrinology And Metabolism*, 279(3):E508–E519, 2000.
- [103] R.K. Ross, A. Paganini-Hill, P.C. Wan, and M.C. Pike. Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin. *Journal of the National Cancer Institute*, 92(4):328, 2000.
- [104] J Russo, C Grill, and I.H. Russo. Pattern distribution of cells positive for estrogen receptor alpha and progesterone receptor in relation to proliferating cells in the mammary gland. *Breast Cancer Research and Treatment*, 53:217–227, 1999.
- [105] V.W. Setiawan, K.R. Monroe, L.R. Wilkens, L.N. Kolonel, M.C. Pike, and B.E. Henderson. Breast cancer risk factors defined by estrogen and progesterone receptor status. *American Journal of Epidemiology*, 169(10):1251, 2009.
- [106] RJ Sheaff, M. Groudine, M. Gordon, JM Roberts, and BE Clurman. Cyclin E-CDK2 is a regulator of p27Kip1. *Genes & development*, 11(11):1464–1478, 1997.
- [107] E. Sherer, RE Hannemann, A. Rundell, and D. Ramkrishna. Analysis of resonance chemotherapy in leukemia treatment via multi-staged population balance models. *Journal of Theoretical Biology*, 240(4):648–661, 2006.
- [108] K. Simms, N. Bean, and A. Koerber. A mathematical model of cell cycle progression applied to the MCF-7 breast cancer cell line. *Bulletin of Mathematical Biology*, In press, 2011.
- [109] G. Soderqvist, E. Isaksson, B. von Schoultz, K. Carlstrom, E. Tani, and L. Skoog. Proliferation of breast epithelial cells in healthy women during the menstrual cycle. *American Journal of Obstetrics and Gynecology*, 176(1):123–128, 1997.
- [110] A. Ström, J. Hartman, J.S. Foster, S. Kietz, J. Wimalasena, and J.Å. Gustafsson. Estrogen receptor  $\beta$  inhibits  $17\beta$ -estradiol-stimulated proliferation of the breast can-

- cer cell line T47D. *Proceedings of the National Academy of Sciences of the United States of America*, 101(6):1566, 2004.
- [111] R.L. Sutherland, R.E. Hall, G.Y.N. Pang, E.A. Musgrove, and C.L. Clarke. Effect of medroxyprogesterone acetate on proliferation and cell cycle kinetics of human mammary carcinoma cells. *Cancer Research*, 48(18):5084–5091, 1988.
- [112] R.L. Sutherland, R.E. Hall, and I.W. Taylor. Cell proliferation kinetics of MCF-7 human mammary carcinoma cells in culture and effects of tamoxifen on exponentially growing and plateau-phase cells. *Cancer Research*, 43(9):3998–4006, 1983.
- [113] A. Swarbrick, C.S.L. Lee, R.L. Sutherland, and E.A. Musgrove. Cooperation of p27Kip1 and p18INK4c in progestin-mediated cell cycle arrest in T-47D breast cancer cells. *Molecular and Cellular Biology*, 20(7):2581–2591, 2000.
- [114] M. Swat, A. Kel, and H. Herzl. Bifurcation analysis of the regulatory modules of the mammalian G1/S transition. *Bioinformatics*, 20(10):1506, 2004.
- [115] I.W. Taylor, P.J. Hodson, M.D. Green, and R.L. Sutherland. Effects of tamoxifen on cell cycle progression of synchronous MCF-7 human mammary carcinoma cells. *Cancer Research*, 43(9):4007–4010, 1983.
- [116] CD Thron. Mathematical analysis of binary activation of a cell cycle kinase which down-regulates its own inhibitor. *Biophysical Chemistry*, 79(2):95–106, 1999.
- [117] C. Torricelli, V. Fortino, E. Capurro, G. Sacchi, P. Ponzio, A. Pacini, M. Muscettola, and E. Maioli. Role of PTHrp and PTHrp-engaged pathways in MCF-7 cells migration/invasion. *Matrix Biology*, 25(2):104–111, 2006.
- [118] L.C. Tsai, M.W. Hung, C.C. Yuan, P.L. Chao, S.Y. Jiang, G.G. Chang, and T.C. Chang. Effects of tamoxifen and retinoic acid on cell growth and c-myc gene expression in human breast and cervical cancer cells. *Anticancer Research*, 17(6D):4557–4562, 1997.
- [119] J.J. Tyson. Modeling the cell division cycle: cdc2 and cyclin interactions. *Proceedings of the National Academy of Sciences*, 88(16):7328, 1991.

- [120] J.J. Tyson. Models of cell cycle control in eukaryotes. *Journal of Biotechnology*, 71(1-3):239–244, 1999.
- [121] PJ Van Diest, E. Van Der Wall, and JPA Baak. Prognostic value of proliferation in invasive breast cancer: a review. *Journal of Clinical Pathology*, 57(7):675, 2004.
- [122] H.L. Van Epps. What tumors need. *The Journal of Experimental Medicine*, 201(7):1024, 2005.
- [123] H. Varma and S.E. Conrad. Antiestrogen ICI 182,780 decreases proliferation of insulin-like growth factor I (IGF-I)-treated MCF-7 cells without inhibiting IGF-I signaling. *Cancer Research*, 62(14):3985–3991, 2002.
- [124] J. Vervoorts and B. Lüscher. Post-translational regulation of the tumor suppressor p27 KIP1. *Cellular and Molecular Life Sciences*, 65(20):3255–3264, 2008.
- [125] B. Vogelstein and K.W. Kinzler. The multistep nature of cancer. *Trends in Genetics*, 9(4):138–141, 1993.
- [126] D.G. Von Mühlen, D. Kritz-Silverstein, and E. Barrett-Connor. A community-based study of menopause symptoms and estrogen replacement in older women. *Maturitas*, 22(2):71–78, 1995.
- [127] GN Wade, JD Blaustein, JM Gray, and JM Meredith. ICI 182,780: a pure antiestrogen that affects behaviors and energy balance in rats without acting in the brain. *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology*, 265(6):R1392–R1398, 1993.
- [128] Graeme Wake. Personal communication, February 2011.
- [129] A.E. Wakeling, M. Dukes, and J. Bowler. A potent specific pure antiestrogen with clinical potential. *Cancer Research*, 51(15):3867–3873, 1991.
- [130] Z. Wang, B.W. Yu, KM Rahman, F. Ahmad, and F.H. Sarkar. Induction of growth arrest and apoptosis in human breast cancer cells by 3, 3-diindolylmethane is associated with induction and nuclear localization of p27kip. *Molecular Cancer Therapeutics*, 7(2):341–349, 2008.

- [131] JP Ward and JR King. Mathematical modelling of avascular-tumour growth II: Modelling growth saturation. *Mathematical Medicine and Biology*, 16(2):171, 1999.
- [132] J.D. Watson and F.H.C. Crick. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738, 1953.
- [133] CK Watts, A. Brady, B. Sarcevic, A. DeFazio, EA Musgrove, and RL Sutherland. Antiestrogen inhibition of cell cycle progression in breast cancer cells is associated with inhibition of cyclin-dependent kinase activity and decreased retinoblastoma protein phosphorylation. *Molecular Endocrinology*, 9(12):1804–1813, 1995.
- [134] C.K.W. Watts, K.J.E. Sweeney, A. Warlters, E.A. Musgrove, and R.L. Sutherland. Antiestrogen regulation of cell cycle progression and cyclin D1 gene expression in MCF-7 human breast cancer cells. *Breast Cancer Research and Treatment*, 31(1):95–105, 1994.
- [135] R.A. Weinberg. *The Biology of Cancer*. Garland Science New York, 2007.
- [136] NR Wilcken, OW Prall, EA Musgrove, and RL Sutherland. Inducible overexpression of cyclin D1 in breast cancer cells reverses the growth-inhibitory effects of antiestrogens. *Clinical Cancer Research*, 3(6):849–854, 1997.
- [137] NR Wilcken, B. Sarcevic, EA Musgrove, and RL Sutherland. Differential effects of retinoids and antiestrogens on cell cycle progression and cell cycle regulatory genes in human breast cancer cells. *Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research*, 7(1):65–74, 1996.
- [138] T. Williams and R. Bjerknes. Stochastic model for abnormal clone spread through epithelial basal layer. *Nature*, 236, 1972.
- [139] K.A. Won and S.I. Reed. Activation of cyclin E/CDK2 is coupled to site-specific autophosphorylation and ubiquitin-dependent degradation of cyclin E. *The EMBO journal*, 15(16):4182, 1996.
- [140] X. Xu, T. Nakano, S. Wick, M. Dubay, and L. Brizuela. Mechanism of Cdk2/Cyclin E inhibition by p27 and p27 phosphorylation. *Biochemistry*, 38(27):8713–8722, 1999.

- [141] G. Yao, T.J. Lee, S. Mori, J.R. Nevins, and L. You. A bistable Rb–E2F switch underlies the restriction point. *Nature Cell Biology*, 10(4):476–482, 2008.