Mathematical Modelling of Cell Cycle and Telomere Dynamics

Bartholomäus Hirt

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

March 2013

Abstract

The eukaryotic cell cycle primarily consists of five phases, namely a resting state, G_0, and four cycling phases G_1, S, G_2 and M phase, with cells progressing in this order before dividing into two cells back in phase G_1. Understanding how a drug affects the cell cycle can give insight into the drug's mechanism of action and may assist research into potential treatment strategies. The pentacyclic acridinium salt RHPS4 (3,11-difluoro-6,8,13-trimethyl-8H-quino$[4,3,2-kl]$ acridinium methosulfate) is an attractive agent because it is potentially cell-cycle specific and inhibits the activity of telomerase, an enzyme known for its role in cellular immortalisation in human cancer. The precise mechanism of action of the drug on the cell cycle dynamics, however, remains unclear.

We have devised experiments, collected experimental data and formulated a mathematical model describing the cell cycle dynamics of cancer cells and their time- and dose-dependent modulation by RHPS4 to investigate how the compound affects cells in each stage of the cell cycle. In addition to a control case, in which no drug was used, we treated colorectal cancer cells with three different concentrations of the drug and fitted simulations from our models to experimental observations. We have shown that the model is "identifiable", meaning that, at least in principle, the parameter values can be determined from observable quantities. Our fitting procedure also generates information on the sensitivity of parameters in the model. We found that RHPS4 caused a marked concentration-dependent cell death in treated cells, which is well modelled by allowing the rate parameters corresponding to cell death to be sigmoidal functions of time. Since the drug uptake into the nucleus is rapid (saturation within 5 hours), the observed delay effect of 5 days of the compound is unexpected and is a novel finding of our research into

this compound. Our results show that, at low concentrations, RHPS4 primarily affects the cells in the G_2/M phase, and that the delay decreases at larger doses. We propose that secondary effects lead to the induction of observed cell death and that changes in the molecular structure of the non-coding DNA sequences at chromosome ends, called telomeres, might be a precursor of delayed cell death.

We therefore investigated the dynamics of telomere length in different conformational states, that is, t-loops, G-quadruplex structures and those being elongated by telomerase. By formulating differential equation models we studied the effects of various levels of telomerase and RHPS4 concentrations on the distribution of telomere lengths and analysed how these effects evolve over large numbers of cell generations. As well as calculating numerical solutions, we use quasicontinuum methods to approximate the behaviour of the system over time, and predict the shape of the telomere length distribution. We showed that telomere length maintenance is tightly regulated: too high levels of telomerase lead to continuous telomere lengthening, and large concentrations of RHPS4 lead to progressive telomere erosion. Our results suggest different effects of RHPS4 dependent on the drug concentration used: low concentrations reduce telomere length, but do not impair the equilibrium of the system, and high concentrations destabilise the system leading to chromosome degradation and senescence and/or cell death. Moreover, our models predict a positively skewed distribution of telomere lengths at equilibrium, and our model predictions are in good agreement with experimental data.

Published papers

A paper named "The effects of a telomere destabilizing agent on cancer cell-cycle dynamics – Integrated modelling and experiments", authored by Bartholomäus V. Hirt, Jonathan A.D. Wattis, Simon P. Preston and Charles A. Laughton, containing the results from Chapters 4 to 6, was accepted for publication in October 2011 and has been published in the Journal of Theoretical Biology in January 2012.

Bartholomäus V. Hirt, Jonathan A.D. Wattis and Simon P. Preston also submitted another paper entitled "Modelling the regulation of telomere length: the effects of telomerase and G-quadruplex stabilising drugs" to the Journal of Mathematical Biology, which presents the results from Chapter 7 and is currently under review.

Acknowledgements

It is an honour to thank my supervisors, Drs. Jonathan Wattis, Simon Preston and Charles Laughton. In particular, I would like to thank Jonathan and Simon for weekly intensive and fruitful discussions of mathematical and statistical problems, their invaluable continuous assistance, guidance and their meticulous reading of my early drafts of this thesis. I also would like to thank Charlie for generously funding my tissue culture experiments and for many stimulating and insightful comments and suggestions during the course of this work. I have very much enjoyed the discussions with all of you and appreciate that you always took time for my questions. I am also thankful to the School of Mathematical Sciences, University of Nottingham, for providing funding for this work.

I would like to thank all members of the Pharmacology Group at the Centre for Biomolecular Science, University of Nottingham, for their much-appreciated introduction, comments and support with the experiments in the tissue culture lab. In particular, I am indebted to Tracey Bradshaw and Lyn Warner for the amount of time they spent teaching me all relevant cell culture techniques. I am also grateful for the great support regarding biological questions, interpretation of experimental data and many interesting discussions with Charles Matthews, and for all the help and support I received from Lucy Johnson, Jessica Chu, Ashraf Abdalla, Liu Xiangrui and many more in the lab.

My extended thanks go to Helen Cunliffe for always being there when needed and doing a great job as a research secretary, and to Hilary Lonsdale for all the great help about organising my trips to various conferences.

Finally, I would like to thank Kasia, my loving wife, for her continuous and amazing involvement in my project, for making me so happy and taking such good care of me.

Without you this work would not have been possible. I am also indebted to my wonderful family, in particular to my mum, my mother-in-law and my grandmother (who passed away before I could finish my degree), for their mental and motivational support, their love and care, and always believing in me. Particular thanks go to my friends, Gem and Chris, for being so loyal and lovable.

Contents

I Preliminaries **1**

1 Introduction and literature review **2**

1.1 Introduction 2

1.2 Cell cycle dynamics, senescence and apoptosis 4

1.2.1 Biological background 4

1.2.2 Mathematical modelling of cell-cycle dynamics 9

1.3 Telomeres and the end-replication problem 14

1.4 DNA damage repair and telomere capping 18

1.5 Telomerase and telomere length regulation in cancer cells 24

1.6 Quantitative models of telomere length dynamics 27

1.6.1 Modelling telomere loss 28

1.6.2 Modelling telomere length regulation by telomerase 34

1.6.3 Modelling telomerase-independent telomere maintenance 38

1.7 Anti-cancer effects of RHPS4 38

1.8 Outline of thesis 40

2 Compartmental ODE, PDE and DDE models of the cell cycle **42**

2.1 A three-compartment ODE model 43

2.2 A five-compartment ODE model 45

2.3 A seven-compartment ODE model 49
2.4 Development of PDE and DDE cell cycle models 51
2.4.1 High order ODE model 52
2.4.2 Deduction of PDE model 55
2.4.3 Deduction of a three-compartment DDE model 58
2.5 Analysis of the three-compartment DDE model 61
2.6 Conclusions 68

3 Identifiability analysis **70**
3.1 Introduction and definitions 70
3.2 Taylor series method and application to the five-compartment ODE model 73
3.3 Transfer function method 74
3.4 Practical identifiability analysis 76

II RHPS4 and Cell Cycle Dynamics **80**

4 Experimental work **81**
4.1 Experimental protocol, cell cycle analysis and normalisation of data 81
4.1.1 Suppliers of reagents 81
4.1.2 Drug stock and cell line 82
4.1.3 Cell stocks and general cell culture 83
4.1.4 Seeding of cells 84
4.1.5 Flow cytometric cell cycle analysis 84
4.1.6 Data collection and normalisation 86
4.2 Experimental results 88
4.3 Interpretation of results and discussion 93

5 Mathematical model of cell cycle dynamics and the effects of RHPS4 96

5.1 Mass balance equations and analysis 97

5.2 Markov-chain model and deduction of ODE model 104

5.3 Statistical model of experimental data 107

5.4 Structural identifiability analysis – application of transfer function method 109

5.5 Parameter estimation techniques 110

5.6 Model selection and evaluation 112

6 Fitting model to data 115

6.1 Results of best fit 116

6.2 Parameter fitting - model results 119

6.3 Sensitivity analysis 126

6.4 Accuracy of fit 128

6.5 Biological implications 131

6.6 Discussion and conclusions 132

III Telomere length dynamics 135

7 Models of telomere length dynamics in telomerase-positive cells 136

7.1 Estimation of kinetic model parameters 139

7.2 Model (i): A closed model for telomere length dynamics in the S/G_2 phase 141

7.3 Model (ii): An open model of length dynamics engaging a negative feedback mechanism in S/G_2 147

7.4 Model (iii): Model of telomere length dynamics over large numbers of cell divisions 156

7.4.1 Model formulation 156

7.4.2 Solution by quasi-continuum approximation 158

7.4.3 Interpretation of results 160

7.5 Summary of results and discussion 166

8 Discussion **171**

8.1 Conclusions 171

8.2 Future work 176

References **180**

A Example Codes for Identifiability **197**

A.1 Taylor series method 197

A.2 Transfer function method 198

B MATLAB code for parameter estimation **200**

C Computational notes and data tables for practical identifiability analysis **205**

C.1 Numerical computation of $\mathbf{J_F}(\widehat{\mathbf{p}})$ 205

C.2 Correlation coefficients for rate parameters of model M_2^{**} 206

C.3 Eigenvalues and eigenvectors for the confidence region of $\widehat{\mathbf{p}}$ in model M_2^{**} 207

D Coefficients of solution (7.3.11) **208**

List of Figures

1.1 The basic phases of the eukaryotic cell cycle. 5

1.2 A histogram of DNA content analysis using a DNA binding dye and flow cytometry. 6

1.3 Cyclical changes of Cdk activity, which are dependent on cyclin protein levels, indicated together with their time point of activity controlling certain events in the cell-cycle. 7

1.4 An illustration of the morphological changes of apoptotic cells. 9

1.5 A two-compartment model of cycling (X_1) and resting (X_2) cells including cell loss of rate η from the X_1 compartment and of rate γ from the X_2 compartment. 9

1.6 A three-compartment model describing the transitions of cells from the ith compartment ($i = 1, 2, 3$) as age-dependent rate functions $\Gamma_i(\tau)$. 11

1.7 A cell cycle diagram of an unperturbed cell population consisting of four phases (G_1, S, G_2, M), where the S phase, which is defined by DNA synthesis, is represented by an age-structured compartment in this model. . . 12

1.8 A six-compartment model of a cell population, which comprises either dead cells or cells in the different phases of the cell cycle, G_0, G_1, S, G_2, and M. 13

1.9 Nucleotide pairing and the structure of the DNA double helix. 15

1.10 Initiation of DNA replication at replication origins, leading to the formation of replication bubbles. 15

1.11 The asymmetric progression of DNA replication. 16

1.12 The mechanisms of the lagging strand synthesis. 17

1.13 Telomere replication and resection of the 5′ end by a nuclease. 19

1.14 Three different types of homologous recombination: (a) T-loop HR, (b) T-SCE and (c) recombination with interstitial sites. 22

1.15 Shelterin complex and t-loop structure at the telomeric end. 23

1.16 G-quadruplex folding topology with three G-quartets, stacked and stabilised by a positively charged metal ion with antiparallel orientation of the DNA strands. 24

1.17 Telomere elongation by the enzyme telomerase. 25

1.18 Telomere length dynamics caused by the end-replication problem leading to a deletion of length μ on telomeric DNA strands. 29

1.19 Frequencies of human diploid fibroblasts with varying population doubling potential. 29

1.20 An illustration of telomere length dynamics due to the end-replication problem and, additionally, (A) single-strand breaks at 5′ ends or (B) C-strand processing at telomeric ends. 31

1.21 Incomplete DNA replication causes telomere deletion of varying size. . . . 32

1.22 Increasingly skewed telomere length distributions of fibroblast cells at four different cumulative population doublings. 33

1.23 Schematics of telomere shortening, a graph illustrating the dependence of the equilibrium telomere length on the number of telomerase molecules, and the probability p that any site in the telomere is bound to TRF2, and model simulations fitting telomere length histogram data. 36

1.24 Telomere length distributions, $K(l,t)$, simulated using (1.6.8) for different generations ($t = 1, 5, 10, 20, 30$) in a telomerase-positive cell, where telomere shortening dominates telomere elongation. 38

1.25 Chemical structure of the pentacyclic acridinium salt RHPS4. 39

2.1 A three-compartment model containing compartments of resting, cycling and permanently arrested cells (senescence/apoptosis). 43

2.2 A model with five compartments and transitions between the compartments and to apoptosis from the X, Y, Z and Σ compartments describes the basic dynamics of the cell cycle. 45

2.3 Numerical simulations of the five-compartment model (2.2.1)-(2.2.5) with $k_{YZ} = 1$, $k_{ZX} = 2$ and $k_{\Sigma A} = 0.25$, and k_{XY} and $k_{X\Sigma}$ as indicated above each plot, where we assumed $k_{XA} = k_{YA} = k_{ZA} = 0$. 50

2.4 An ODE model with seven compartments describing all phases of the cell cycle and including transitions to apoptosis from the senescent phase Σ, the S phase and the G_2 phase. 50

2.5 A two-compartment model, with compartments A and B, and constant transition rate k_{AB}. 53

2.6 Cumulative distribution functions $F(\tau;\lambda) = 1 - \mathrm{e}^{-\lambda\tau}$, $\tau \geq 0$, and probability density functions $f(\tau;\lambda) = \lambda\,\mathrm{e}^{-\lambda\tau}$, $\tau \geq 0$, of the exponential distribution with rates $\lambda = 0.5, 1, 1.5$. 53

2.7 A model with two compartments A and B, where the A compartment is split into N subcompartments A_i. 54

2.8 Probability density functions $f(\tau;N,\lambda) = \tau^{N-1}\,\lambda^N\,\mathrm{e}^{-\tau\lambda}/\Gamma(N)$, $\tau > 0$, of the gamma distribution $\Gamma(N, 1/\lambda)$ with scales $N = 1, 2, 3, 5$ and shape $1/\lambda > 0$. 54

2.9 An ODE model with the Y compartment of cycling cells divided into N subcompartments, describing different stages of the S+G_2+M phase. . . . 55

2.10 A diagram of a three-compartment ODE-PDE model with age-structured Y compartment of length T_Y (residence time of cells in the S+G_2+M phase) and age-dependent loss terms $k_{YA}(\tau)$ from the Y compartment to apoptosis. 58

2.11 A diagram of a three-compartment model with a delay in the Y compartment of constant length T_Y. 60

2.12 Plot of the function $f(y)$ and the straight line $g(y)$ against y for $\tau = 1$, $\alpha = -1$, $\beta = 2$. 64

2.13 Plot of the function $h(\tau)$ against τ for $h(1/k_{YA}) = 2\,k_{XY}/(e\,k_{YA}) > \pi$. . . . 66

2.14 Three plots showing the different dynamic behaviour of the solution trajectories of model (2.4.27)-(2.4.29) for $k_{XY} = 0.5$, $k_{XA} = 0$ and $k_{YA} = 0.4$. . 67

2.15 Three plots showing the different dynamic behaviour of the solution trajectories of model (2.4.27)-(2.4.29) for $k_{XY} = 1$, $k_{XA} = 0.4$ and $k_{YA} = 0$. . . 67

3.1 A simple nonidentifiable two-compartment model with only one observable state X. 71

4.1 Gating of singlets and 'clean' histogram of FL3 of gated cells using the software package WinMDI. 86

4.2 Cell densities $N_i(t)$, at time point t, for each experiment used to analyse cells at day i, where $N_i(0)$ are the respective seeding densities per 2 ml of medium. 88

4.3 Total numbers $N(t_i)$ of cells and the numbers of dead cells (measured by trypan blue dye-exclusion) for no drug and each drug concentration (50 nM, 100 nM, 1 μM) of RHPS4 shown on a log scale at each day t_i, $i = 1, \ldots, 10$ of analysis. 90

4.4 Proportion of cells in each phase of the cell cycle at each day of analysis. Data were collected for control cells and three concentrations (50 nM, 100 nM, 1 μM) of RHPS4. 91

4.5 Distributions of the DNA content for control cells and cells treated with 1 μM of RHPS4, each analysed at day 4 and day 10, with manually set gates for the estimation of the corresponding cell cycle distributions. 92

5.1 A cell cycle model of viable cells including death from each of the three phases. 97

5.2 A seven-compartment model with compartments X, Y, Z, $\bar{X}$, $\bar{Y}$ $\bar{Z}$, A, arising from the data collected in Chapter 4. 98

5.3 A schematic illustration of (i) rate model M_0, (ii) rate model M_1, and (iii) rate model M_2, governing transition from viable cells to dead cells in the seven-compartment model in Figure 5.2. 100

5.4 Simulations for two sets of parameter values in system (5.1.1)-(5.1.7), displaying exponential growth and exponential decay. 102

5.5 Simulations for two sets of parameter values in system (5.1.1)-(5.1.7), displaying asymptotically constant behaviour and asymptotically linear growth of the state variable $\bar{Z}(t)$. 103

5.6 A simple two-compartment model of cells cycling in the V compartment with doubling rate μ and cells exiting the cell cycle to the $\bar{V}$ compartment with rate δ. 104

6.1 The objective function values $f(\mathbf{p})$ plotted against generation number G for several repeats of the optimisation routine with a logarithmic scale for the ordinate. 117

6.2 Results of fitting model M_2^{**} to experimental data from treatment with 50 nM of RHPS4. Optimal parameter values from the global (SRES) and local optimisation routine (LM) for 30 runs are plotted with their optimal function value for a choice of 8 parameters. 118

6.3 Solutions of the model fitting procedure to experimental data. Simulations stem from the 'best' models chosen in Table 6.1. 120

6.4 Cell cycle dynamics of model (5.1.1)-(5.1.7) for the behaviour of control cells over 9 days and for the behaviour of cells treated with 50 nM of RHPS4 over 10 days. 121

6.5 Cell cycle dynamics of model (5.1.1)-(5.1.7) for the behaviour of cells treated with 100 nM and 1 μM of RHPS4. 121

6.6 Selected sensitivity plots for model M_2^{**} (where $t_0 \approx 5$) and results of fitting data from treatment with 50 nM RHPS4. 127

6.7 Residuals $\mathbf{e}_{i,j}$ for time points t_i, $i = 1, \ldots, 10$, and model categories $j = 1, \ldots, 6$, plotted on the time scale for model M_2^{**} and experimental data from treatment with 50 nM of RHPS4. 129

6.8 Parameter estimates for control cells and treatment of cells with 50 nM, 100 nM and 1000 nM RHPS4 from fitting model M_0^* (control), model M_2^* and model M_2^{**} (treated cells). 130

6.9 Projections of the ellipsoidal regions for the estimated parameter $\hat{\mathbf{p}}$ onto the k_{ZX}-$k_{Z\bar{Z}0}$-plane and the k_{ZX}-$\Delta k_{Z\bar{Z}}$-plane shown for 50 nM of RHPS4. . 131

7.1 A HT Q-FISH histogram of the telomere length distribution of HeLa cells, where $n = 495$ nuclei were analysed. 138

7.2 A Gaussian probability density function, $p(x)$, for the telomere length in HeLa cells, with mean L_0 and standard deviation σ, and the rate of t-loop formation, $k_c(x)$, modelled by a sigmoidal function of telomere length with shape parameters α, β and δ. 141

7.3 Model of telomeric states U, B, G, C. Kinetics for each reaction are described by their rate constants k. Free telomerase (T) and RHPS4 (R) in the nucleus bind open forms (U) and G4 structures (G), respectively. Telomerase elongation occurs at rate ρ. 142

7.4 Simulation of the steady state numbers of telomeres B – bound by telomerase, U – in the open form, and G – in G4 form, dependent on telomerase molecule numbers T_0 in a drug-free assay ($R_0 = 0$). 145

7.5 Simulation of the average length of telomeres after telomere shortening due to the end-replication problem and subsequent telomere elongation by telomerase for a period of $t_r = 4.5, 6, 7.8$ h as a function of telomerase molecules T_0, where we assume the system is in "quasi-steady" state. . . . 146

7.6 Simulation of the average length of telomeres after telomere shortening due to the end-replication problem and subsequent telomere elongation by telomerase for a period of $t_r = 7.8$ h with $T_0 = 2500, 5000, 10000$ telomerase molecules: each curve plotted as a function of R_0. 146

7.7 Model of telomeric states U, B, G, C. Telomeres enter the system at rate k_e and exit the system due to t-loop formation at rate $k_c(x)$ and due to G4-stabilisation by RHPS4 at rate k_r. 147

7.8 Simulations of telomere length distributions $p(x)$ and $k_c(x)\,U(x) + k_r\,C(x)$ of telomeres entering and telomeres leaving the open system per unit of time, respectively, for $\delta = 5 \times 10^{-5}$ s^{-1} and different numbers of telomerase molecules, $T_0 = 500, 1000, 1700, 5000$ (no drug, $R_0 = 0$). 150

7.9 Simulations of steady state solutions $U(x)$, $B(x)$, $G(x)$ for $\delta = 5 \times 10^{-5}$ s^{-1} and different numbers of telomerase molecules, $T_0 = 1000, 1700, 5000, 10000$ (no drug, $R_0 = 0$). 150

7.10 Simulations of telomere length distributions $p(x)$ and $k_c(x)\,U(x) + k_r\,C(x)$ of telomeres entering and telomeres leaving the open system per unit of time, respectively, for $T_0 = 5000$ and different concentrations of RHPS4, $R_0 = 0, 50, 100, 1000$ nM. 150

7.11 Simulations of steady state solutions $U(x)$, $B(x)$, $G(x)$, $C(x)$ for $T_0 = 5000$ and different concentrations of RHPS4, $R_0 = 0, 50, 100, 1000$ nM. 150

7.12 Simulations of telomere numbers $k_c(x)\,U(x)$ and $k_r\,C(x)$ leaving the system from U and C, respectively, per unit time for $T_0 = 5000$ and different concentrations of RHPS4, $R_0 = 0, 50, 100, 1000$ nM. 151

7.13 The mean telomere length $\widehat{L}_{R_0=0}$ of telomeres leaving the system, $k_c(x)\,U(x) + k_r\,C(x)$ plotted against the parameter δ ($\sigma = 800$ bp) and σ ($\delta = 5 \times 10^{-5}$ s^{-1}), controlling the rate of t-loop formation and the standard deviation of the initial telomere length distribution $p(x)$, respectively. 154

7.14 The mean telomere length $\widehat{L}$ of telomeres leaving the system, $k_c(x)\,U(x) + k_r\,C(x)$, plotted against the concentration R_0 of RHPS4 for three different numbers of telomerase molecules, $T_0 = 419, 1000, 1700$, and $\widehat{L}_{R_0\to\infty}$ plotted as a function of T_0. 154

7.15 Simulations of telomere length distributions $p_0(x)$ and $p_i(x)$ of telomeres entering the open system at generations 0 and $i = 25$, respectively, in units of s^{-1} for $T_0 = 500$, $\delta = 5 \times 10^{-5}$ s^{-1} and different concentrations of RHPS4 (0nM, 50nM and 1000nM). 155

7.16 Simulations of telomere length distributions $p_0(x)$ and $p_i(x)$ of telomeres entering the open system at generations 0 and $i = 10$, respectively, in units of s^{-1} for $T_0 = 5000$ (super-telomerase cells), $\delta = 5 \times 10^{-5}$ s^{-1} and different concentrations (0nM, 50nM and 1000nM) of RHPS4. 155

7.17 A closed model of telomeric states U, B, G, C, with telomeres of length x losing μ basepairs when they exit the system by t-loop formation (rate $k_c(x)$) or after forming a complex with RHPS4 (rate k_r); telomeres re-enter the system in the open (U) form. 156

7.18 T_0-R_0-regions where physical steady state solutions U exist, plotted for $\delta = 5 \times 10^{-3}$ s^{-1} ($\delta < k_f$) and $\delta = 2.5 \times 10^{-2}$ s^{-1} ($\delta > k_f$). 163

7.19 The mean telomere length $\widehat{L}$ of telomeres $k_c(x)\,U(x) + k_r\,C(x)$ exiting the state U with rate $k_c(x)$ or the state C with rate k_r at steady state per unit time, plotted as a function of the concentration of RHPS4, R_0, and for four different numbers of telomerase molecules, $T_0 = 1000, 2000, 3000, 4000$. . . 164

7.20 Plots of steady state curves $U(x)$, $B(x)$ and $G(x)$ against telomere length for two different values of ρ and varying numbers of telomerase molecules ($T_0 = 406, 415, 420$ and $T_0 = 25, 27, 29$), where we assume that telomeres shorten by $\mu = 45$ nt between each round of replication and $R_0 = 0$. 165

7.21 Plots of telomere length distributions $p(x)$ and $k_c(x)\,U(x) + k_r\,C(x)$ at steady state for two different values $\rho = 4.5 \times 10^{-3}$ nt·s^{-1}, $\rho = 6.287 \times 10^{-2}$ nt·s^{-1} and varying numbers of T_0. 165

7.22 Contour plots of the mean telomere length $\widehat{L}$ in (T_0, R_0) space for $\rho = 4.5 \times 10^{-3}$ nt·s^{-1} and $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}, where $\delta = 5 \times 10^{-5}$ s^{-1} in both cases. 166

7.23 A plot of the mean telomere length $\widehat{L}$ of telomeres $k_c(x)\,U(x) + k_r\,C(x)$ exiting the system per unit time as a function of the number T_0 of telomerase molecules for the case of no drug ($R_0 = 0$) in the system ($\rho = 4.5 \times 10^{-3}$ nt·s^{-1} and $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}). 166

7.24 A plot of the mean telomere length $\widehat{L}$ of telomeres $k_c(x)\,U(x) + k_r\,C(x)$ exiting the system per unit time as a function of the concentration R_0 of RHPS4 for three different numbers of telomerase molecules, $T_0 = 406, 500, 750$ ($\rho = 4.5 \times 10^{-3}$ nt·s^{-1}) and $T_0 = 25, 30, 35$ ($\rho = 6.287 \times 10^{-2}$ nt·s^{-1}). 167

7.25 Plots of telomere length distributions $p(x)$ and $k_c(x)\,U(x) + k_r\,C(x)$ at steady state of the system (7.4.1)-(7.4.4) per unit time for varying concentrations of RHPS4 ($R_0 = 0, 50, 100, 250$ nM). 167

7.26 Plots of telomere length distributions $k_c(x)\,U(x)$ and $k_r\,C(x)$, leaving the system at steady state per unit time, for varying concentrations of RHPS4 ($R_0 = 0, 50, 100, 250$ nM). 167

8.1 An extension of model (i) from Section 7.2, containing the unbound, bound, G-quadruplex and RHPS4-G-quadruplex complex telomeric states U, B, G, C, respectively, showing the effects of the shelterin protein POT1 on telomere dynamics during the S/G$_2$ phases of the cell cycle. 178

List of Tables

4.1 Seeding densities (cell numbers per 2 ml of RPMI medium) used for cell cycle analysis of control cells and cells treated with 3 different concentrations of RHPS4, where cells were incubated between 1 and 10 days. 87

4.2 The Table shows the average values and respective standard deviations of the total number of doublings and the proportions (in %) of HCT116 cells that are apoptotic (trypan-blue stained), in the pre-G_1, G_0/G_1, S and G_2/M phases of the cell cycle. 89

6.1 The optimal objective function values $f_{\min} = f(\hat{\mathbf{p}})$, variance estimates $\hat{\sigma}^2$ of the model residuals and the values μ_{cAIC} of the corrected Akaike criterion given for models M_0, M_1 and M_2 and each concentration of RHPS4. 118

6.2 Results of parameter estimation. The transition rates k are displayed for all concentrations of the drug RHPS4, together with respective residence times T_X, T_Y, T_Z of cells in the G_0/G_1, S, G_2/M phase, and doubling times T_d (before the time of marked cell death) of viable cells. 123

6.3 Percentages of live cells in each model compartment. 125

6.4 Percentages of which phases of the cell cycle cells die from at each dose of RHPS4 at times $t > t_0$. 125

C.1 Upper triangular part of the correlation matrix $\mathbf{Corr}(\widehat{\mathbf{p}})$ with entries $r_{ij}\,(\widehat{\mathbf{p}})$ computed as in (3.4.7) for model M_2^{**} and the data fit from treatment with 50 nM RHPS4. 206

C.2 Eigenvalues μ_k and radii l_k of the principal axes for the ellipsoidal 95% confidence region of the $L = 9$-dimensional parameter vector $\widehat{\mathbf{p}}$ for model M_2^{**} and 50 nM RHPS4. 207

C.3 Eigenvectors $\mathbf{u}_k$ describing the directions of the principal axes of the ellipsoidal confidence region of $\widehat{\mathbf{p}}$ for model M_2^{**} and 50 nM RHPS4. 207

Part I

Preliminaries

CHAPTER 1

Introduction and literature review

1.1 Introduction

Among human diseases, cancer is the leading cause of death worldwide according to the World Health Organisation [149]. Incidence data derived from population-based cancer registries estimate that around 7.6 million people died of cancer worldwide in 2008, which makes about 0.1% of the world's total population and accounts for around 13% of all deaths and nearly 60% of all diagnosed cancer [63]. Cancer is characterised by rapid and uncontrollable cell division, which can lead to the invasion of surrounding tissue, but also to metastasis at distant sites by spreading through the bloodstream or lymphatic system. The most common cancer types are female breast, lung, prostate, cervix, stomach and colorectal cancer, to mention just a few, with the latter comprising 1.2 million new cases and 608,700 deaths, estimated in 2008. Viral infections account for the development of a large proportion of liver cancer (by the hepatitis B virus) and cervical cancer (by the human papillomavirus), where viruses transform normal cells into tumour cells by the activation of so-called oncogenes (genes that lead to cell transformation when mutated), with vaccinations now being available to prevent these infections. The causes of most of the many other cancer types, however, are not well understood, and genetic defects, environmental and lifestyle factors, such as environmental pollutants, radiation, diet and cigarette smoking, are considered to be the potential risk factors of cancer. The effectiveness of cancer treatment is therefore often limited, with surgery, radiation and chemotherapy being the typical treatment strategies, but these can cause considerable

side effects and are often only able to prolong a patient's lifespan instead of eradicating all cancer cells.

Chemical signals control cell proliferation in normal cells and when cells become overcrowded some of them are "told" to die. There is a strictly coordinated series of processes that facilitate cell replication, known as the cell division cycle, which includes the precise replication of the DNA and division of cells into two daughter cells. Most cells cycle and divide a limited number of times, a discovery first made by growing normal human fibroblasts in culture (60-80 population doublings) about 50 years ago [55]. The limited number of divisions is referred to as the Hayflick limit [56]. On reaching the Hayflick limit, cells cease proliferation permanently and enter a state called replicative senescence in which they are still alive but do not grow. Irreversible functional changes at protective structures located at the ends of chromosomes, known as telomeres, are responsible for the limited lifespan of a cell. Telomeres are eroded during each replication, marking the ageing of cells and eventually triggering irreversible cell cycle exit. Cancer cells, however, are able to avoid this natural limit, displaying reduced sensitivity to signals that control the cell cycle.

In 2009, Elizabeth H. Blackburn, Carol W. Greider and Jack W. Szostak were awarded the Nobel Prize in Physiology or Medicine for their discovery of the role of telomeres, the previously unknown enzyme telomerase (catalysing the elongation of chromosomal telomeres), and its role in chromosome protection. Telomerase was found to build the end of chromosomes, protecting them from degradation and thus delay cellular senescence. Stem cells and most cancer cells express telomerase and have the potential to divide indefinitely without any significant erosion of telomeres. This discovery has stimulated the development of new strategies for cancer therapy, with telomeric structures and telomerase being new targets for drug design. The DNA of cells is relatively exposed during cell division, making cells that actively progress through the cell cycle good targets for drugs. There is increasing research in this area and various compounds are being investigated with the aim of understanding how they interact with telomeres and the enzyme telomerase in the cell.

Modern experimental techniques generate a vast amount of data that have to be inter-

preted for useful systematic information to be extracted. Mathematical models can be a useful means for integrating different types of experimental data to predict the mechanism of action of compounds [148]. In this thesis I aim to enhance the understanding of how a potential anti-cancer drug affects cell cycle dynamics and the processes of telomere length maintenance in cancer cells.

1.2 Cell cycle dynamics, senescence and apoptosis

This Chapter provides some basic concepts of the biological cell cycle model that can be found in the standard literature, for example the book by Alberts [4], and some concepts of the function of telomeres in cell replication. It also includes a literature review of mathematical models of the cell cycle and telomere dynamics.

1.2.1 Biological background

The cell cycle is the essential series of events by which all living beings reproduce and primarily replicate and segregate their genome into two identical daughter cells. Unicellular organisms, such as bacteria and yeast, create new organisms each time cells divide, whereas multicellular organisms require a long and complex series of cell divisions to become a coherent system of tissues and different cell types. We distinguish between prokaryotes (bacteria and archaea), which are organisms that lack a distinct nucleus or any other organelles (membrane-bound specialised subunits of a cell), and eukaryotes, such as yeast, plant and animal cells, which possess nuclei containing the cell's genetic material, and on which we henceforth focus.

The eukaryotic cell cycle is the mechanism that ensures accurate duplication of DNA, organelles, and other cellular material, and the precise division of these copies into two genetically identical daughter cells. These processes define the two major phases of the cell cycle, the *synthesis (S) phase* and the *mitotic (M) phase*. The typical cell cycle time is about 24 hours. Chromosome duplication occurs during the S phase, which takes about 10-12 hours in a typical mammalian cell. After the S phase, cell division, which is subdivided into mitosis (chromosome segregation and nuclear division) and cytokine-

sis (cytoplasmic division), occurs during the M phase, and requires typically less than one hour. Cells have extra *gap phases*, a G_1 *phase* between the M phase and the S phase, and a G_2 *phase* between the S phase and mitosis, where they double their number of proteins and organelles and grow in size. The G_1, S, G_2 phase together are called *interphase*, which occupy about 23 hours of a 24-hour cycle. The length of the G_1 phase is variable as its main purpose is to monitor the internal and external environment of the cell and to check whether it is suitable for the cell to commit to DNA synthesis. If conditions are unfavourable for replication, cells may enter a specialised resting phase (*quiescence*), also known as the G_0 *phase*, which can last for several days up to even several months when signals for cells to resume proliferation are missing. Most somatic cells are usually found in the G_0 phase, and they return to G_1 when cellular conditions are favourable or else they die. A schematic of the cell cycle is shown in Figure 1.1.

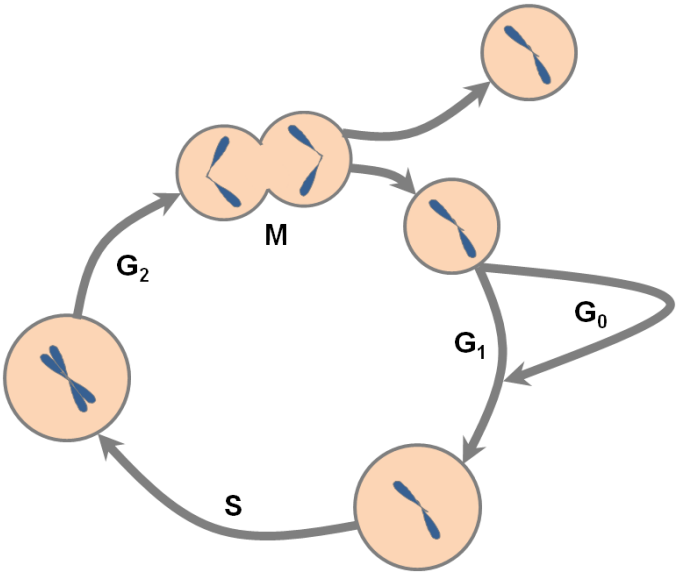

Figure 1.1: The basic phases of the eukaryotic cell cycle.

To determine the phase of a cell within the cell cycle, we can measure its DNA content (which doubles during the S phase), by using fluorescent binding dyes and a flow cytometer, which allows for a rapid and automatic analysis of a large number of cells. An example of a histogram plot of DNA content is shown in Figure 1.2. A quantitative model is used to determine the number of cells in each phase of the cell cycle, which assumes that the population of cells is separated into compartments consisting of cells in either the G_0/G_1 phase or the G_2/M phase, or at a certain stage of DNA replication within the

S phase. The fluorescence intensity of cells in each compartment is assumed to be distributed normally with the mean intensity of those in the G_2/M phase twice as high as those in the G_0/G_1 phase. Cells in the S phase have several mean fluorescence intensities that are located between the mean intensities of the G_0/G_1 and G_2/M compartment. It is also possible to use flow cytometry to determine the length of the G_0/G_1, S, and G_2/M phases by synchronising the initial cell population at some point in the cell cycle and, after release from synchrony, following the progression of the population through the cycle. Similarly, S phase cells can be visualised by adding the artificial thymidine analog bromo-deoxyuridine (BrdU) to cell culture, which is incorporated by the cell into newly synthesised DNA, and by subsequent staining of these molecules with anti-BrdU antibodies.

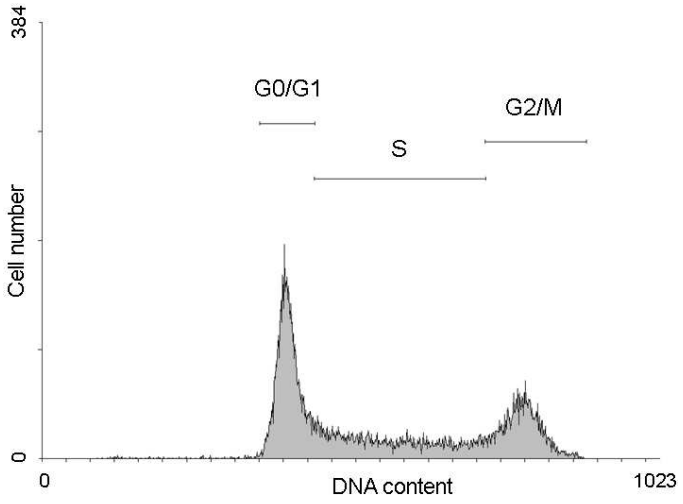

Figure 1.2: DNA content analysed by flow cytometry using a DNA binding dye. The histogram shows the fluorescence intensity of a typical cell population and indicates the associated cell cycle phases.

The order of cell-cycle events is controlled by a regulatory network of proteins, known as the cell-cycle control system, which consists of biochemical switches initiating cell cycle events and once an event is launched, its progression is irreversible. Three checkpoints regulate key cell-cycle transitions. The first of which is known as *Start* (in yeast), or the *restriction point* (in mammalian cells), when the cell commits itself to cell-cycle entry and chromosome replication. The two subsequent checkpoints are the *G_2/M checkpoint*, where the cell triggers the early mitotic events, and the *metaphase-to-anaphase transition* or *Finish*, where the control system triggers the separation of sister chromosomes and completion of mitosis and cytokinesis. In case of problems that affect the cell cycle, the checkpoints

can block further progression through the cycle until the problem has been resolved.

Core controllers of the cell-cycle control system are cyclin-dependent kinases (Cdks), which trigger major cell cycle events once they are activated. Cdk activity is cyclic and is regulated by binding of proteins known as cyclins to form cyclin-Cdk complexes. Cyclins undergo a cycle of synthesis and degradation and bind different Cdks at different stages of the cell cycle, each of the cyclin classes stimulating certain events in the cell cycle. The cyclin levels and cyclin-Cdk complexes associated with their specific time point of activity in the cell cycle are illustrated in Figure 1.3. More information on regulatory networks of cell cycle control can be found in the reviews from Aguda [2] and Schafer [120], and more quantitative discussions are presented by Novak and Tyson [93] and Tyson and Novak [138].

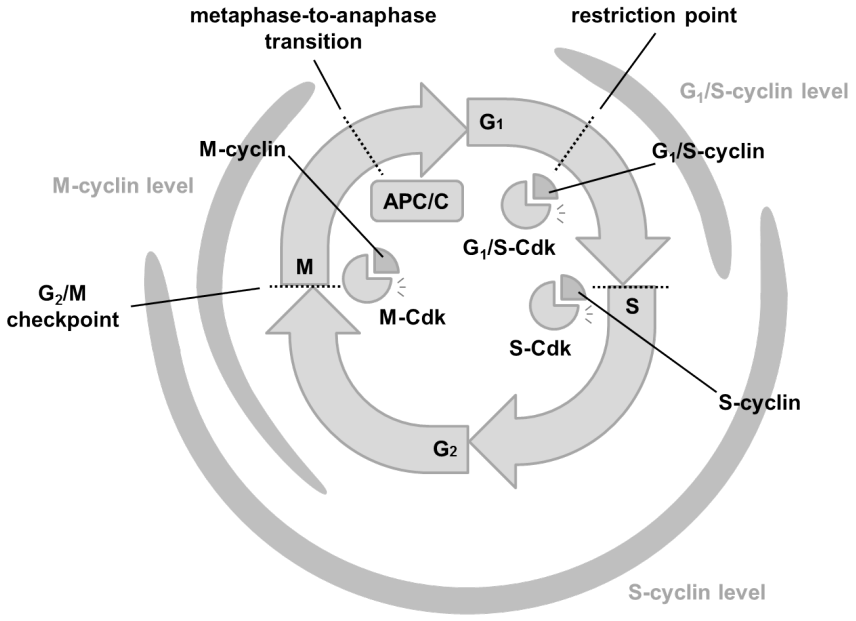

Figure 1.3: Cyclical changes of Cdk activity, which are dependent on cyclin protein levels, are indicated together with their time point of activity controlling certain events in the cell-cycle. The metaphase-to-anaphase transition is initiated by a separate regulatory protein called APC/C.

The replication capacity of somatic cells is usually limited and cells undergo a permanent cell-cycle arrest after several divisions. This state is called *replicative senescence* and is primarily triggered by the activation of the gene p53. Activation of p53 then triggers the gene expression of p21, which is a Cdk inhibitor, and thus induces cell cycle arrest. The expression of p53 and p21 is only transient and other Cdk inhibitors, such as the G_1 cell-cycle inhibitor p16, are upregulated later, which causes cell proliferation to halt

with cells being in a G_0/G_1 state. Cell viability is not affected at this stage and cells are still metabolically active and can survive for a long period of time. Dimri et al. [42] discovered that the enzyme β-galactosidase displays abnormal behaviour associated with senescent cells, a feature that they used to detect senescent cells and distinguish them from cells in other phases of the cell cycle. The behaviour is termed senescence-associated β-galactosidase (SA β-gal) activity, where the enzyme β-galactosidase is normally active at pH 4 in human, but in senescent cells is often found to be active at pH 6.

Cells die through several different mechanisms, but the one most relevant to the work in subsequent Chapters is controlled cell death called *apoptosis* (Figure 1.4). This form of cell death can occur at any point during the cell cycle and is characterised by morphological changes such as cell shrinking, cytoplasmic and nuclear condensation, chemically altered cell surface, cleavage of the nuclear chromatin and splitting of the cell into membrane-enclosed debris called apoptotic bodies. Cells dying by apoptosis do so by phosphorylation of p53 (but not p21) and do not elicit inflammatory responses, but are engulfed and digested by phagocytes, specialised white blood cells that protect the body by ingesting dead cells in living organisms. A discussion of different forms of eukaryotic cell death and their terminology has been given by Fink and Cookson [45].

Apoptotic cell death can be measured by the characteristic features these cells display. For example, nuclear and cytoplasmic condensation and cellular fragmentation can be observed by light and electron microscopy. Furthermore, DNA fragmentation, which is considered a biochemical characteristic of apoptosis, can be detected as a ladder pattern of chromosomal DNA fragments by the use of gel electrophoresis. Another technique is labelling the ends of DNA strand breaks in the so-called TUNEL (Terminal deoxynucleotidyl transferase-mediated dUTP Nick-End Labelling) technique, which are then visualised by standard microscopy techniques. The change of the plasma membrane is detected experimentally by labelling cells with Annexin V, and another dye, trypan blue (which we use in our tissue culture experiments in Chapter 4), which does not colour viable cells, but is taken up by dead cells whose membrane has become permeable.

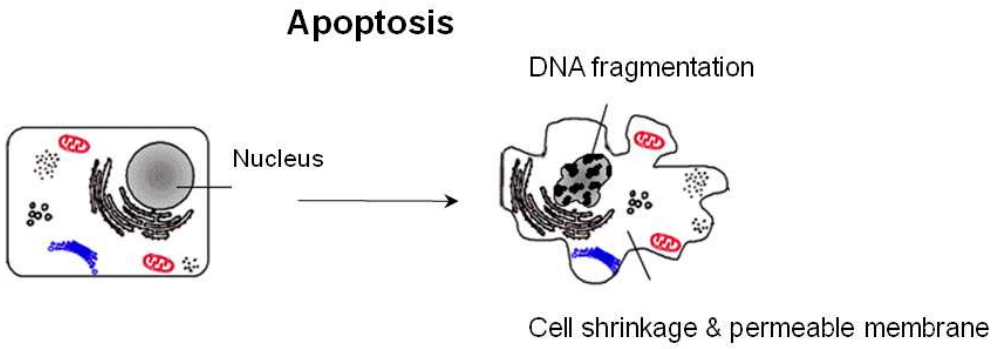

Figure 1.4: An illustration of the morphological changes of apoptotic cells, taken from the free encyclopaedia http://en.wikipedia.org/wiki/Apoptosis.

1.2.2 Mathematical modelling of cell-cycle dynamics

Several mathematical models have been developed to describe the different phases of the cell-cycle and their transition dynamics, including cell death. In particular, some models have been established describing the dynamics of tumour cell populations that are treated by certain anti-cancer drugs. A selection of these models is given below.

Panetta and Adam [102] designed a two-compartment ordinary differential equation (ODE) model of cycling and resting cells, which incorporates chemotherapeutic effects of a cell cycle-specific drug and includes some constraint equations for the effect of the drug on other non-cancerous cells. The simulations obtained from the model identify an appropriate period and dose of the drug to balance cell kill of cancer cells and normal cycling tissue cells. The general model is shown in Figure 1.5.

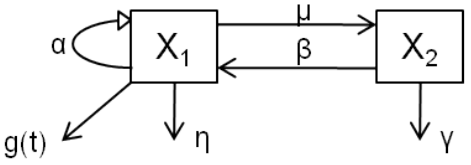

Figure 1.5: A two-compartment model of cycling (X_1) and resting (X_2) cells including cell loss of rate η from the X_1 compartment and of rate γ from the X_2 compartment. Cells double with rate α and cells transfer from X_1 to X_2 with rates μ and β, respectively. Chemotherapeutic effects are described by additional cell loss of rate $g(t)$, being a function of time t, from the compartment of cycling cells.

The system of ODEs describing the effect of a cell cycle-specific drug on cancerous cells

is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \alpha - \mu - \eta - g(t) & \beta \\ \mu & -\beta - \gamma \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}. \tag{1.2.1}$$

Here, X_1 is the mass of cycling tumour cells with growth rate α and X_2 is the mass of resting tumour cells; μ and β are the transition rates between cycling and non-cycling cells; natural decay is given by the rates η and γ for cycling and noncycling cells, respectively. All parameters are assumed to be constant and positive. The chemotherapeutic effects together with the decay of the drug can be modelled, for example, by a piecewise continuous function of the form $g(t) = h\,\mathrm{e}^{-\gamma(t-n\tau_D)}$ for $n\,\tau_D \leq t < (n+1)\,\tau_D$ with $n = 0, 1, 2, \ldots$, where h is a cell kill parameter, γ the decay of the drug and τ_D the period between administrations of the drug.

In contrast to the deterministic model of Panetta and Adam [102], Smith and Martin [126] considered slowly proliferating cells, where the population of quiescent cells introduces a large variation in the duration of the G_0/G_1 phase. The probabilistic model partitions the intermitotic phase into a phase of a fixed length of time containing $S/G_2/M$ cells, and a state in which cells stay for a random time and transfer to $S/G_2/M$ with a certain probability.

Sherer et al. [124] used a partial differential equation (PDE) model that incorporates age-dependent transition rates between the phases of the cell cycle and describes the number density of cells in each phase as a function of time t and age τ (the time since the last transition from a preceding cell-cycle phase). The model shown in Figure 1.6 is derived from three compartments, representing the G_0+G_1 phase, the S phase and the G_2+M phase. DNA histograms of the total cell population provide data to estimate the corresponding transition rate functions. Analysis of the age density is facilitated by labelling a proportion of S-phase cells with BrdU, which is incorporated into newly-synthesised DNA at time point $t = 0$. One subsequently observes both labelled and unlabelled subpopulations of cells in the S phase over time. The underlying assumption of this method is that the total cell population grows in a balanced fashion, that is, the population densities of the single compartments maintain a constant ratio to one another. The three-

compartment model is described by the following PDEs

$$\frac{\partial n_i(\tau,t)}{\partial t} + \frac{\partial n_i(\tau,t)}{\partial \tau} + \Gamma_i\, n_i = 0, \qquad i = 1,\ldots,3, \tag{1.2.2}$$

where $n_i(\tau,t)$ denote the number densities of cells within the three cell compartments, and $\Gamma_i(\tau)$ is the transition rate of cells that have spent a period of time τ in the ith compartment since entering it; cell loss is neglected in this model. Data of human leukaemia cells were fitted by the model, where the best fit suggested a bimodal transition age probability distribution for cells in the G_0/G_1 phase.

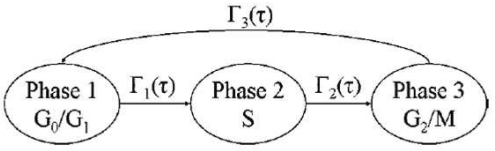

Figure 1.6: A three-compartment model describes the transitions of cells from the ith compartment ($i = 1,2,3$) as age-dependent rate functions $\Gamma_i(\tau)$. The illustration has been taken from Sherer et al. [124]. Reprinted with permission from John Wiley and Sons.

Basse et al. [13], by contrast, assume constant model parameters but include loss terms for cell death from each compartment of a four-compartment model describing the growth of human tumour cells and their response to the anticancer drug paclitaxel which arrests cells in the mitotic phase. The four compartments shown in Figure 1.7 correspond to the phases of constantly proliferating cells, that is the G_1, S, G_2 and M phase. The independent variables of the associated system of ODEs with one PDE are time t, age τ_S of cells in the S phase, and the relative DNA content x of cells. This set-up of model equations accounts for the process of DNA synthesis and enables the simulation of steady-state distributions of the DNA content in cell populations, which can be compared to experimental measurements from flow cytometry DNA analysis.

In a later work, Basse et al. [15] considered a similar model as follows. We will use the model formulation involving temporal ordinary and partial differential equations in

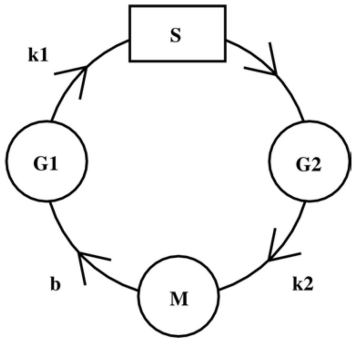

Figure 1.7: A cell cycle diagram of an unperturbed cell population consisting of four phases (G_1, S, G_2, M) with transition rates k_1, k_2 and b. The S phase, which is defined by DNA synthesis, is represented by an age-structured compartment (rectangular box) in this model. Loss terms have been neglected. The diagram is taken from Basse et al. [15]. Reprinted with kind permission from Springer Science and Business Media.

Chapter 2.4, which is given by the system

$$\frac{\mathrm{d}\, n_1(t)}{\mathrm{d}t} = 2\, k_{M1}\, n_M(t) - k_{1S}\, n_1(t)\,, \tag{1.2.3}$$

$$\frac{\partial\, n_S(t;\tau_S)}{\partial\, \tau_S} + \frac{\partial\, n_S(t;\tau_S)}{\partial\, t} = 0\,, \tag{1.2.4}$$

$$\frac{\mathrm{d}\, n_2(t)}{\mathrm{d}t} = n_S(t;\tau_S)\Big|_{\tau_S=T_S} - k_{2M}\, n_2(t)\,, \tag{1.2.5}$$

$$\frac{\mathrm{d}\, n_M(t)}{\mathrm{d}t} = k_{2M}\, n_2(t) - k_{M1}\, n_M(t)\,, \tag{1.2.6}$$

with the initial condition for the S phase being

$$n_S(t;\tau_S)\Big|_{\tau_S=0} = k_{1S}\, n_1(t)\,. \tag{1.2.7}$$

Here, $n_1(t)$, $n_2(t)$, $n_M(t)$ are the numbers of cells in the G_1, G_2 and M phase at time t, respectively; k_{1S}, k_{2M}, k_{M1} are the respective transition rates between the single phases; $n_S(t;\tau_S)$ is the number density of cells having spent τ_S hours in the S-phase at time t; and T_S is the time after which cells transfer from the S phase to the G_2 phase. Death rates are neglected in this model. Experimental data from 11 different human tumour cell lines unperturbed by cancer therapy were compared *in vitro* to gain insight into why some human cancer patients respond differently to cancer therapy while having similar initial profiles.

A more general model of Basse and Ubezio [12] considers the exposure of human tu-

mour cell populations to cancer therapy and can be used to examine the effects of different therapies, such as radiation therapy, the administration of the drug paclitaxel and the S-phase specific compound camptothecin, and combination treatments. The model allows for a user-defined number of phases and time-dependent transition rate functions, with a PDE system, similar to that of Sherer et al. [124], which describes the cell cycle. A further extension, which incorporates interactions between drugs and tumour microenvironments can be found in Venkatasubramanian et al. [139]. Cell-cycle progression, drug and nutrient diffusion, intracellular metabolism, cellular drug effects and drug pharmacokinetics are combined in a multi-compartment model, where the cell cycle is essentially described by six compartments (Figure 1.8) and modelled by coupled nonlinear PDEs, ODEs and algebraic equations consisting of functions of time and populations of cells in three-dimensions. Also, cell cycle transition rates are functions of the cellular energy production, which in general increase with increasing cellular energy production, whereas transition rates to quiescence and cell death increase with decreasing cellular energy production.

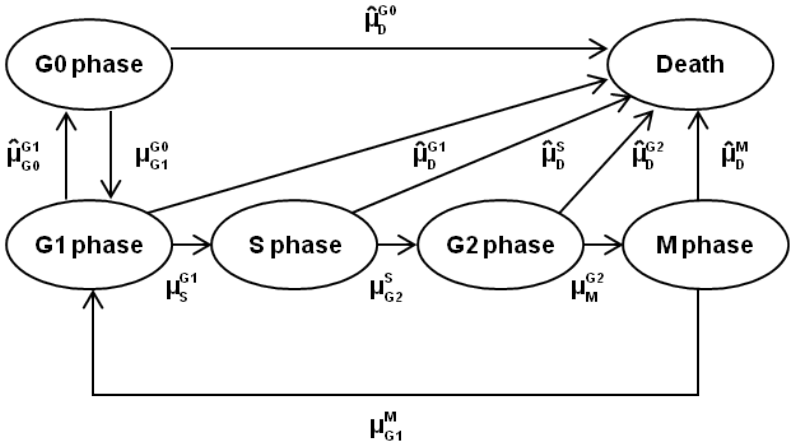

Figure 1.8: A six-compartment model of a cell population is shown, which comprises either dead cells or cells in the different phases of the cell cycle, G_0, G_1, S, G_2, and M. Each line signifies a transition from one cell phase to another and transition rates between the cell cycle phases are denoted by $\mu_{\text{to}}^{\text{from}}$. Cell cycle transition rates are dependent on the intracellular energy metabolism, where transitions μ to the phases of cycling cells increase with increasing cellular energy production, and transitions $\hat{\mu}$ to quiescence (G_0) and cell death increase with decreasing cellular energy production. The diagram is redrawn and adapted from Venkatasubramanian et al. [139] with permission from Elsevier.

Additionally, mathematical models on the cell-cycle control system can be found in Tyson [137], Novak and Tyson [94] and, in relation to molecular pathways involved in cell growth, in Yang et al. [153]. Modelling of cell cycle regulation together with general

principles of mathematical modelling are reviewed in Fuss et al. [46]. More cell-cycle models can be found in Basse et al. [14], who describe the impact of DNA degradation of apoptotic cells arrested in mitosis (by the drug paclitaxel) on flow-cytometric DNA distribution profiles in melanoma cells (skin cancer), and in Sherer et al. [123] and Basse et al. [16], who deal with the effects of chemotherapy and radiation on cancer cell-cycle dynamics, respectively.

1.3 Telomeres and the end-replication problem

A cell's DNA molecules are packed inside the nucleus into long polynucleotide chains, where a pair of tightly connected DNA chains forms a chromosome. Chromosomes are complemented with other proteins to build up a compact structure called chromatin. During the S phase of the cell cycle, the cell replicates its chromosomes by separating the paired DNA strands and using each as a template for polymerisation to make a new DNA strand, called the primer strand, containing a complementary sequence of nucleotides. Nucleotides form the basic subunits of the DNA, and a nucleotide consists of a sugar-phosphate molecule with a nitrogen-containing side group, or base, attached to it. The sugar-phosphate molecules make up the backbone of the DNA structure, holding the polynucleotide chain together in the shape of a double helix. The nucleobases are of four types, adenine (A), guanine (G), cytosine (C) and thymine (T). The bases of the existing strand bind successively to bases of the strand being synthesised by weak hydrogen bonds and according to a strict rule defined by the complementary structures of the bases: T pairs only with A, and G only with C. Figure 1.9 illustrates the nucleotide-pairing mechanism and the structure of the DNA helix.

DNA replication is initiated by special *initiator proteins* that pry the two DNA strands apart at positions, called replication origins, at which the DNA is first synthesised. DNA helicases (special enzymes) open up the quite stable DNA double helix ahead of the localised region of replication, referred to as a replication fork, which moves progressively along the parental DNA double helix. The replication forks are formed in pairs and create a replication bubble as they move in opposite directions away from the common point

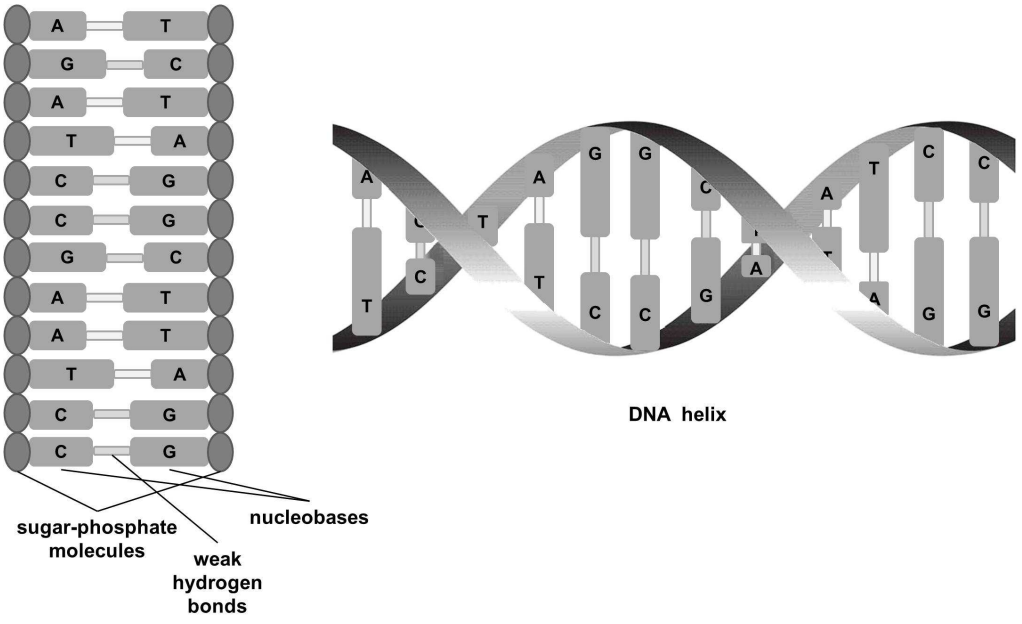

Figure 1.9: Nucleotide pairing and the structure of the DNA double helix.

of origin (Figure 1.10), stopping only when they collide head-on with a replication fork moving in the opposite direction or when they reach a chromosome end. There are many replication forks in an average-size human chromosome, containing about 150 million nucleotide pairs, moving simultaneously on each chromosome, but their replication origins are never at the end of DNA strands due to the specialised structure of chromosomes (see below). A multienzyme complex containing primase and DNA polymerases synthesises the DNA of newly formed daughter strands at the replication forks. First, the enzyme DNA primase synthesises a short RNA primer on the DNA strand, which is about 10 nucleotides long, and DNA polymerases then extend the primer by polymerisation of free nucleotides into DNA. The chemical structure of RNA is very similar to DNA, and a strand of RNA can form base pairs with a strand of DNA as long as the nucleotide sequences are complementary.

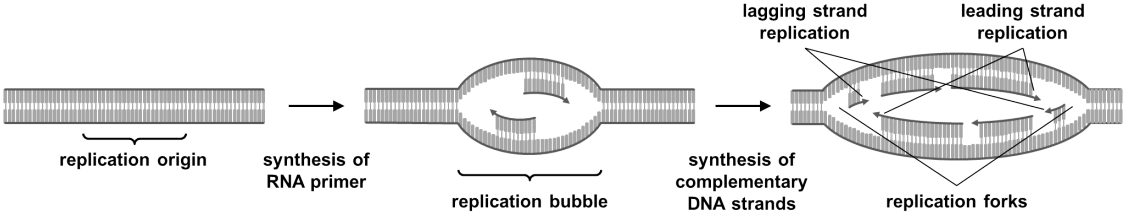

Figure 1.10: Initiation of DNA replication at replication origins, leading to the formation of replication bubbles.

Due to the the general structure of nucleotides and their corresponding binding mechanism, DNA polymerases can synthesise only in 5′-to-3′ direction of a DNA strand - the notation of the direction of a DNA strand derived from the numbering of carbon atoms in the sugar of a nucleotide. The two DNA strands in a DNA helix are, however, oriented in an antiparallel way, hence the growth of both strands cannot proceed in the same way. Indeed, one of the two DNA strands grows piecewise in 5′-to-3′ direction (the one that grows opposite to the overall direction of DNA chain growth), the pieces of DNA known as Okazaki fragments, which are 100-200 nucleotides long in eukaryotes. A replication fork therefore has an asymmetric structure, consisting of one DNA daughter strand that is synthesised continuously, known as the *leading strand*, and the other daughter strand that is synthesised discontinuously, known as the *lagging strand* (see Figures 1.10 and 1.11). The synthesis of the leading strand slightly precedes the synthesis of the lagging strand.

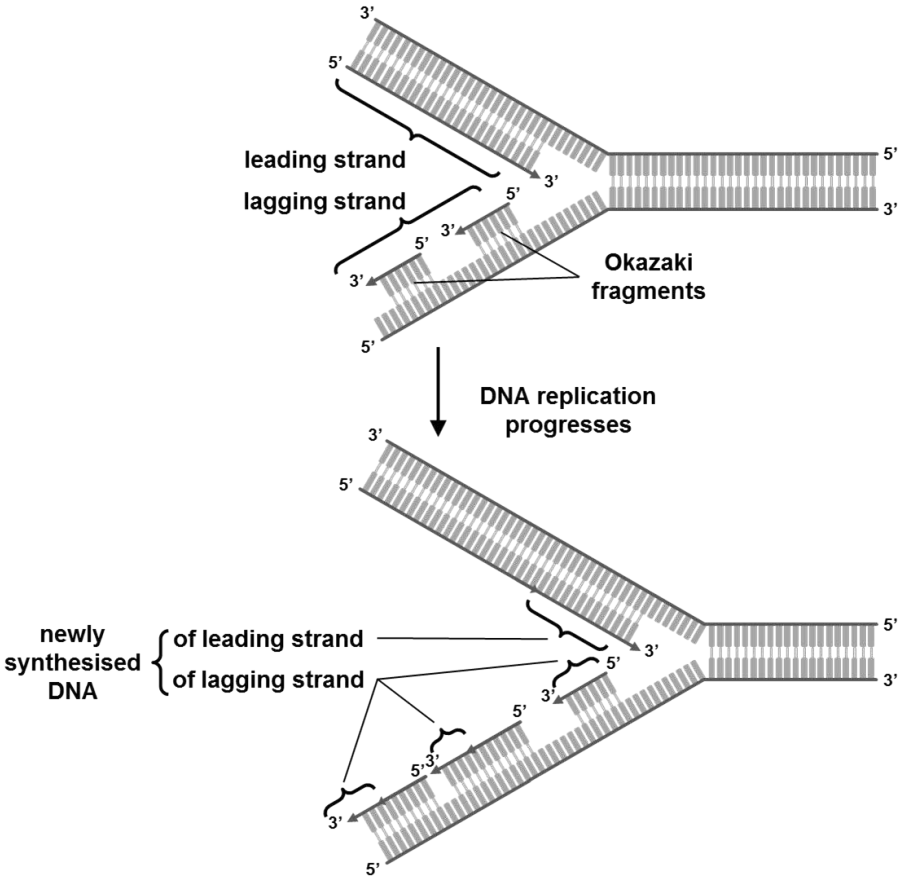

Figure 1.11: The asymmetric progression of DNA replication.

DNA polymerase enzymes require a previously formed base-paired 3′ end of a primer strand to synthesise new DNA, therefore a special mechanism is needed at the lagging strand to synthesise a completely new fragment at a site further along the template strand. RNA primers are produced by DNA primase at intervals of 100-200 nucleotides in eukaryotes, which can then be elongated by DNA polymerases at the 3′ end to begin Okazaki fragments. To produce a continuous DNA chain from the many DNA fragments made on the lagging strand, nucleases participate in the removal of the old RNA primer and a DNA polymerase replaces the RNA primer eventually with DNA, whenever an Okazaki fragment has been completed. The 3′ end of the new DNA fragment is joined to the 5′ end of the previous one by an enzyme called DNA ligase (see Figure 1.12). The newly replicated DNA copies are then attached to one another by a centromere, a specialised DNA sequence holding together the identical sister chromatids (consisting each of a template strand and a primer strand), facilitating the controlled separation and distribution of daughter chromosomes into daughter cells in the M phase.

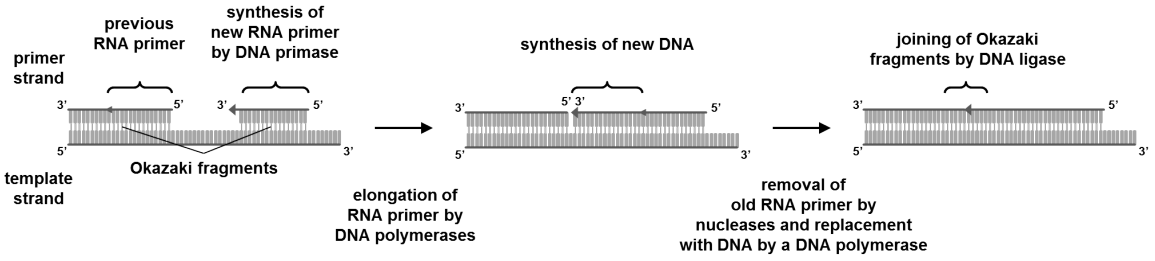

Figure 1.12: The mechanisms of the lagging strand synthesis.

During each replication, the cell encounters a special problem at the 3′ end of the linear chromosome, called the *end-replication problem*, as there is no place for an RNA primer to be attached at the very end of the DNA strand to start an Okazaki fragment and, therefore, the terminal RNA primer cannot be replaced by DNA after its removal. In bacteria, chromosomes are of circular shape, which solves the problem. Eukaryotes, in contrast, solve it by having specialised nucleotide sequences at the end of linear chromosomes, that are protected by specialised protein complexes. These structures are called *telomeres* and contain short tandem repeats, which are similar in various organisms. The sequence of the repeat unit is GGGTTA in mammalian cells and the majority of eukaryotes, and human telomeres contain around 10-15 kilobasepairs at birth. The telomeric strand that con-

stitutes the 3′-end is generally rich in guanosine and devoid of cytosine, the two strands of the telomeric DNA are hence referred to as the G- and C-strands. The single-stranded 3′-protrusion of the G-strand is known as the *3′ overhang*, which varies between 50-500 nucleotides in mammalian cells and is considerably shorter in most other eukaryotes. About 3 base pairs are lost at the 5′ end of the lagging strand due to the end-replication problem on each round of cell division, because the last RNA primer is degraded by enzymes and cannot be changed to DNA. However, human and mouse telomeres shorten by about 50-200 base pairs during each replication and the average telomere length in human cells decreases by roughly 2-4 kilobases during their lifetime. Furthermore, 3′ overhangs have been found to be present at both strands of chromosome ends, with overhangs of the leading strand having about half the length of the overhangs of the lagging strand. A more likely explanation (see Palm and de Lange [101], for example) of the intensive and double-sided telomere shortening is postreplicative processing of the 5′ ends by a nuclease (Figure 1.13). C-strand resection also explains an observed correlation [60] between the rate of telomere shortening and the length of the 3′ overhangs in human cell lines.

This progressive telomere erosion has been designated the reason why a normal mammalian somatic cell can divide only a finite number of times *in vitro* (50-80 divisions), with the maximum number often being referred to as the *Hayflick limit* [56]. The cell then triggers a permanent growth arrest and enters the senescent phase when it can no longer undergo the process of cell division. It is, however, not yet clear whether it is the average telomere length [85] or the length of the shortest telomere [58] that is critical for the onset of cell cycle arrest in a cell.

1.4 DNA damage repair and telomere capping

Ageing of cells is hallmarked by the accumulation of damage and insufficient repair that occur continually in the DNA caused by stress. These stresses are caused by not only environmental influences, such as chemical substances and radiation of various sorts, but also reactive oxygen species, which are produced by mitochondria and increased by mutations in mitochondrial DNA. Oxidative stress, containing free radicals, causes

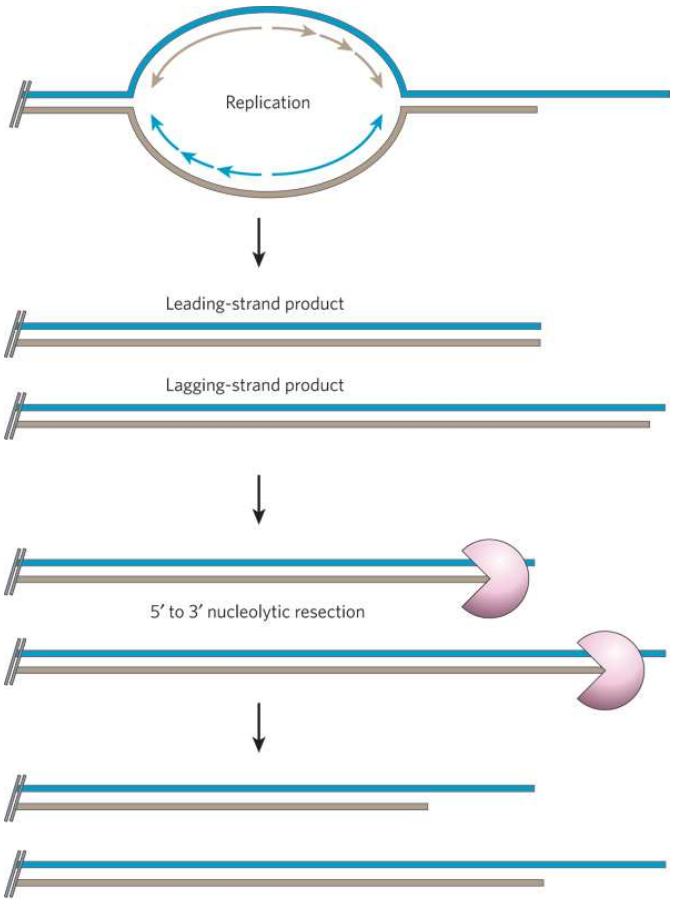

Figure 1.13: Telomere replication and resection of the 5′ end by a nuclease. The illustration is adapted from Verdun and Karlseder [140] with permission from Macmillan Publishers.

accidental lesion in the DNA, which interferes with transcription and DNA replication of the cell leading to mutations, cell-cycle arrest and apoptosis (reviewed by von Zglinicki et al. [142]).

DNA repair is a set of processes which ensures that only a few damages accumulate as mutations in the DNA sequence. DNA damage often occurs in the form of DNA strand breaks, and we distinguish between single- and double-strand breaks. The repair machinery is elicited by a signalling pathway known as DNA damage response with the central protein kinases ATM (ataxia telangiectasia mutated) and ATR (ataxia telangiectasia and Rad3-related) kinases. The ATM and ATR kinases associate with the site of damage, where ATM primarily senses double-strand breaks and ATR responses to single-strand breaks, and their activation causes phosphorylation of various target proteins that eventually initiate cell-cycle arrest.

The cell's DNA repair machinery can replace the damaged nucleotides of one strand by using the other strand as a template for DNA synthesis if only one DNA strand is damaged. DNA double-strand breaks, in contrast, require more sophisticated repair mechanisms and can introduce genome rearrangements and mutations. Typical mechanisms for the repair of double-strand breaks are direct joining of the two double-strand ends (nonhomologous end joining), and the use of a homologous chromosome (homologous recombination), or a sister-chromatid in the late-S/G_2 phase, as a template for the fill-in of the gap caused by the DNA breakage.

A specific protein complex named shelterin (or telosome) caps and protects chromosome ends in order to avoid inappropriate DNA damage repair at telomeres which are likely to be recognised by the cell as double-strand breaks. The shelterin complex in mammalian cells consists of the six proteins TRF1, TRF2 (Telomeric Repeat binding Factor 1 and 2, respectively), POT1 (Protection Of Telomeres 1), TIN2 (TRF1- and TRF2-Interacting Nuclear protein 2), TPP1 (TIN2 and POT1 interacting Protein 1) and Rap1 (Repressor/Activator Protein 1), which are exclusively associated with telomeric DNA, and prevent telomeres from being recognised as DNA damage throughout the life-span of a cell. The proteins TRF1 and TRF2 bind directly to double-stranded telomeric DNA and also recruit TIN2 and Rap1, which lacks DNA binding activity, while POT1 binds to the single-stranded DNA. POT1 can form a complex with the protein TPP1, which is assumed to be crucial for the recruitment of POT1 to telomeres, where TIN2 bridges TPP1 and TRF2, joining the shelterin components of the single-stranded and the double-stranded telomeric DNA. DNA repair at telomeres can result in chromosome aberrations, such as chromosome fusions or rings, due to inadequate repair. Inhibiting or deleting different shelterin proteins has revealed how telomeres protect themselves from DNA repair mechanisms. Depletion of TRF2 or TPP1 results in activation of the ATM kinase pathway, while removal of POT1 proteins from telomeres in mouse cells gives rise to DNA damage response mediated by the ATR kinase independent of ATM [54].

Insufficiently protected telomeres are detected by the DNA repair machinery, form telomere dysfunction-induced foci (TIFs), which are the origins of the telomeric DNA damage signal, and are processed by the cell. Telomeres that are insufficiently protected are recog-

nised by the cell as DNA double-strand breaks and are predominantly repaired by end-to-end joining of chromosomes, mainly established by nonhomologous end joining (NHEJ), or by homologous recombination (HR), which is also known as homology-directed repair. The first mechanism involves the loss of G-overhangs due to chromosome end-processing of the chromosome ends that are to be joined and subsequent ligation of the processed chromosome ends. The fusion of several chromosomes can create long trains of chromosomes, which leads to cell senescence or the initiation of apoptosis pathways in the cell. The second mechanism involves three types of HR that have been observed at telomeres. T-loop HR starts from a loop-like structure and abscises the circular part of the telomere by recombining terminal with internal telomeric DNA, leaving a telomeric circle and a shortened telomere (see Figure 1.14a). The mechanism of Telomere Sister Chromatid Exchanges (T-SCE) is similar to HR of double-strand breaks in coding DNA sequences. T-SCE occurs at sister telomeres after chromosome replication, where the duplicated chromosome ends intersect, recombine and segregate again, which typically leaves one shortened and one elongated telomere (Figure 1.14b). The junction formed between the four DNA strands of two homologous chromosomes is referred to as Holiday junction (HJ). Telomere length changes depend on the location of the cross-over at either telomere and may affect the proliferative capacity of the cell population. Finally, recombination with interstitial sites is a less common type of HR in human and mouse, but more frequent in other animals that possess larger amounts of telomeric repeat sequences at chromosomal internal sites. This type of HR, which produces extrachromosomal DNA in the form of telomeric rings that may even contain the centromere, is similar to T-loop HR, but with the strand invasion occurring at interstitial telomeric sequences (Figure 1.14c). The different types of recombination mentioned above are reviewed in greater detail in Palm and de Lange [101].

The shelterin component TRF2 has been shown not only to prevent the activation of ATM kinases, but also to be involved in a second mechanism for telomere protection, namely the formation of a loop structure hiding the single-stranded terminal DNA from being detected as DNA damage. TRF2 is believed to facilitate coiling of the duplex structure such that the protruding single-stranded overhang can tuck its end into the telomeric

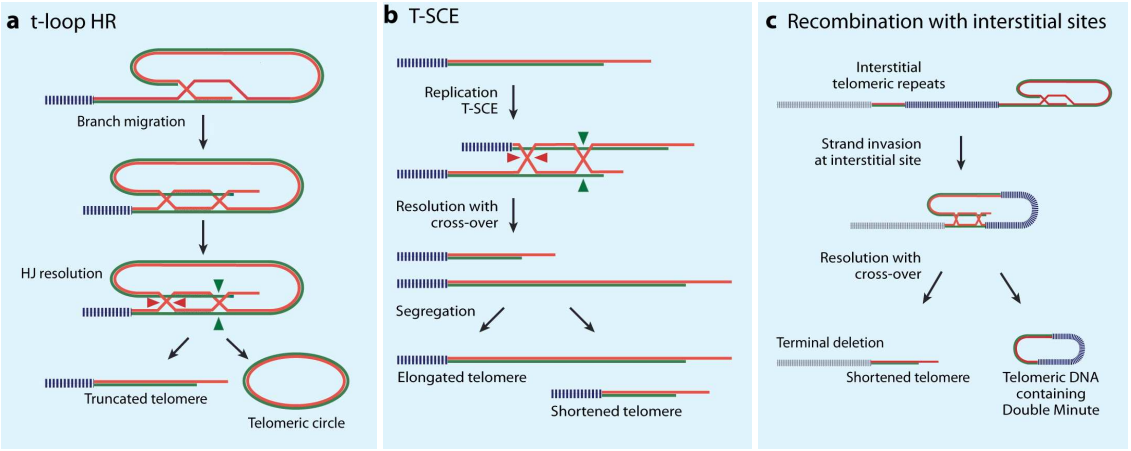

Figure 1.14: Three different types of homologous recombination: (a) T-loop HR, (b) T-SCE and (c) recombination with interstitial sites (see text for more information). The illustrations are adapted from Palm and de Lange [101]. Reprinted with permission from Annual Reviews.

duplex DNA and form a so-called *t-loop* [52]. The G-overhang then forms base pairs with the C-strand of the invaded double-stranded DNA and a smaller displacement D-loop is established by the displaced, single-stranded G-strand at the site of invasion (Figure 1.15). It has been found that TRF2 preferentially binds to the junction between single- and double-stranded DNA at the 3'-end of the telomere, and Stansel et al. [130] suggest a model in which t-loop formation is mediated by interaction of the TRF2 complex bound to that junction with another TRF2 complex bound to an internal site of the telomeric duplex DNA. The circular part of t-loops, however, can vary in size between different organisms (0.3-50 kilobases in eukaryotes) and also within an individual cell and the size does not seem relevant for its function.

The strand invasion of the G-overhang provides the chromosome ends with a structure which distinguishes them from the ends of broken DNA molecules. The t-loop is held together by shelterin proteins, where POT1 specifically binds to the single-stranded telomeric DNA in the D-loop, which is necessary for maintenance of the telomeric cap. Inhibition of either POT1 or TRF2 causes reduction or loss of 3' overhangs, chromosome fusions, senescence and apoptosis [154]. The replication fork might dissolve the t-loop structure during DNA replication, however it is not yet known whether t-loops switch into an open state during the S phase or persist throughout the cell cycle. Reviews of the function of shelterin in telomere protection, including the formation and function of t-loops, are given by de Lange [39], Palm and de Lange [101] and de Lange [40].

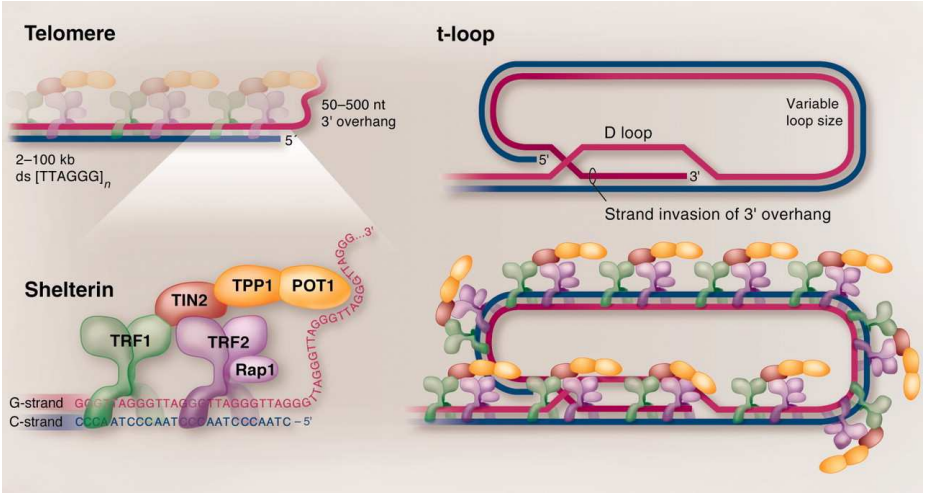

Figure 1.15: Shelterin complex and t-loop structure at the telomeric end. From de Lange [40]. Reprinted with permission from AAAS.

Another functional structure capping chromosome ends, known as the *G-quadruplex*, has been discovered by Henderson et al. [59]. A single-stranded, telomeric G-rich DNA sequence can spontaneously fold into a 4-stranded structure which involves repeating patterns of three G-quartet formations that are planar, hydrogen bonded arrangements of four guanine bases (Hoogsteen base-paired guanines) as illustrated in Figure 1.16. The G-quartets stack to produce the G-quadruplex with a negatively charged channel running through the centre of the planes of quartets. This channel stabilises the structure via positively charged metal ions such as sodium (Na^+) and potassium (K^+) which are co-ordinated in it. The twelve guanines, making up the G-quadruplex, form four backbones which can be connected repeatedly in various ways dependent on the telomeric sequence and choice of metal ion. The human telomeric sequence forms an intramolecular G-quadruplex *in vitro* which can appear with parallel and anti-parallel orientation of the DNA strands involved, and an example with anti-parallel orientation is shown in Figure 1.16. The three stacked G-quartets are connected by two lateral loops and a central diagonal loop, together comprising four backbones forming the parallel edges of this loop-like configuration.

Schaffitzel et al. [121] employed single-chain antibody fragment probes that are specific for G-quadruplex DNA to show that G-quadruplexes form *in vivo*. In their experiments with ciliated protozoa *stylonychia* (a species of eukaryotic microorganisms containing two

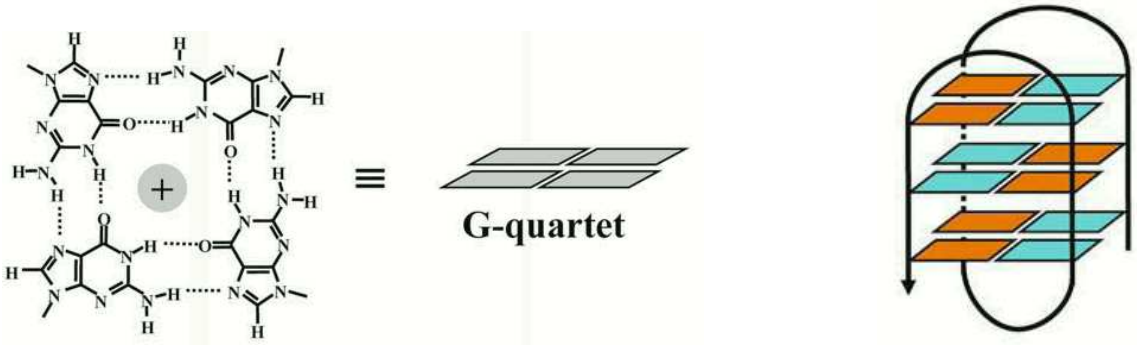

Figure 1.16: G-quadruplex folding topology with three G-quartets, stacked and stabilised by a positively charged metal ion with antiparallel orientation of the DNA strands. The illustrations stem from Moon and Jarstfer [91]. Reprinted with permission from Frontiers in Bioscience.

nuclei, a macronucleus and a micronucleus) G-quadruplex staining occurred everywhere in the macronucleus, but not in the macronuclear replication band, the area where chromosomes replicate, indicating that G-quadruplexes unfold for telomere synthesis to allow for DNA base-pairing. In some human cell lines, POT1 might play a role in the destabilisation of G-quadruplexes at telomere ends [156]. For reviews of G-quadruplex structures *in vitro* and *in vivo*, see Lipps and Rhodes [81] and König et al. [72]. More general reviews on telomere structures and their function in chromosome-end protection can be found in Oganesian and Karlseder [96] and Xu [152].

Cells that have become critically short, however, and thus unprotected, are recognised by the DNA damage response machinery of the cell, typically leading to irreversible cell cycle arrest or apoptosis. Those cells which pass this point in cell replication by inactivation of p53 continue dividing and lose all their protective telomeric DNA until they enter a state called crisis, causing enormous genomic instability, carcinogenesis and eventually cell death: see Greenberg [51] for a review.

1.5 Telomerase and telomere length regulation in cancer cells

The enzyme telomerase can counteract the continuous erosion of telomeres due to the end-replication problem, thus inhibiting telomere uncapping which occurs when telomeres become too short. Telomerase is a specialised RNA-protein complex that mainly consists of TERC (Telomerase RNA Component) with a template region for copying telom-

eric repeat sequences, and the catalytic protein TERT (Telomerase Reverse Transcriptase), which catalyses the 3'-end extension of linear chromosomes. The enzyme telomerase requires the telomeric G-overhang for telomere elongation. TERC recognises and anneals to the overhang and serves as a substrate for the synthesis of telomeric repeat sequences which are added to the telomeric overhang in 5'-to-3' direction (see Figure 1.17). The complementary, lagging strand is then filled in by DNA polymerases using the telomeric extensions as a template, as is conventional in DNA replication. TERC is generally, and independently of telomerase activity, highly expressed in all cells, whereas the concentration of TERT is estimated at less than 50 copies per cell. In normal somatic cells the catalytic subunit TERT is repressed, but it is upregulated in immortal cells, suggesting that TERT is the major determinant for telomerase activity. It was found that telomeres uncap in gastrointestinal progenitor cells (typically telomerase-positive) of mice lacking TERT *in vivo*, and that these telomeres induce apoptosis in late S/G_2 phase, indicating a connection between repression of telomerase, DNA replication and telomere uncapping [109].

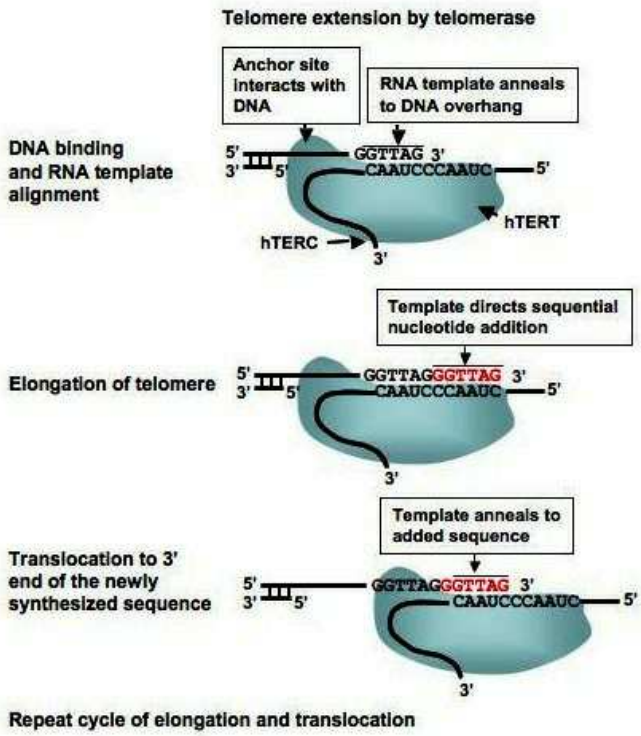

Figure 1.17: Telomere elongation by the enzyme telomerase. The illustration is taken from the free on-line resource at http://gallus.reactome.org.

Telomerase is a reverse transcriptase that replenishes the telomeric sequences each time a cell divides. The telomeric cap not only stabilises telomeres, but also functions as a regulator of telomere length. TRF1 and other shelterin proteins increase with the number of telomere repeats and have been found to inhibit telomere elongation in mammalian telomerase-positive cells, where removal of TRF1 from telomeres leads to telomere elongation. The amount of TRF1 may thus be used to estimate the length of telomeres. Also, direct competition between POT1 and telomerase for the single-stranded telomere terminus has been observed, where reduced loading of POT1 at the 3′ end leads to telomerase-dependent telomere elongation, which is reviewed in Smogorzewska and de Lange [128]. Also TPP1 was found to have inhibitory effects on telomerase. However, some findings are ambiguous, such as the role of POT1 in telomere length regulation, as binding of the POT1-TPP1 complex to internal sites of telomeres *in vitro* resulted in telomere elongation [145], making the shelterin component a telomerase attractor and inhibitor at the same time.

Additionally, Zahler et al. [155] tested oligonucleotides of the hypotrichous ciliate *oxytricha nova*, possessing $n \geq 4$ telomeric repeat sequences, in telomerase assays in presence and absence of the cation K^+, and found that telomerase is less effective in telomere elongation as the concentration of K^+ increases. However, telomerase elongation was not affected when the number of telomeric repeat sequences was reduced to less then 4 sequences per oligonucleotide. Hence G-quadruplex formation seems to affect telomerase activity and is therefore believed to be a telomerase inhibitor. Despite the length variation of individual telomeres within a cell or an organism, average telomere length is maintained within a narrow range that is specific for each species, indicating that telomerase-positive cells establish an equilibrium between telomere attrition and elongation.

Most normal human somatic (body) cells, however, do not contain active telomerase and have a limited replicative capacity. In a few normal human cell types, telomerase activity is strictly regulated during development, mostly lost during embryonic differentiation and remains active only in renewing tissues such as germ cells and stem cell populations, usually displaying a high proliferative potential during their life-span. The majority of cancer cells, in contrast, activate telomerase continually, they possess altered telomeres

and have the potential for unlimited replication. Telomerase was found to be present in 85-90% of cancerous cells and it is believed that its specific role is to immortalise these cells [68].

Most of the remaining 10-15% of cancer cells can maintain their telomeres, and thus avoid the induction of senescence or escape crisis, by a telomerase-independent pathway called alternative lengthening of telomeres (ALT). This mechanism has so far only been observed in anomalous cells (cancerous and genetically modified cells) and is associated with extrachromosomal telomeric DNA, recombination-based mechanisms and large variation or rapid changes in the length of telomeres. ALT cells are believed to elongate their telomeres either by unequal T-SCEs resulting in telomeres with elongated and telomeres with shortened ends, or by HR-mediated DNA synthesis, using an adjacent telomeric sequence as a template for DNA replication. The template for the latter mechanism can be the sister-chromatid telomere, a telomere of another chromosome, (linear or circular) extrachromosomal telomeric DNA or the same telomere through t-loop formation (see Section 1.4). The initiation of ALT is likely to involve intermediate levels of shelterin proteins at telomeres, which are sufficiently low to elicit a DNA damage response, but still high enough to prevent end-to-end fusions. Furthermore, ALT cells typically lose p53 function, which leads to more recombination events and continuing cell divisions (referred to as 'survivorship'). The full mechanisms of how cells activate or prevent ALT has, however, not yet been established. A recent review of ALT has been given by Cesare and Reddel [28].

1.6 Quantitative models of telomere length dynamics

There have been several approaches based on mathematical modelling to understand telomere length dynamics of somatic and cancerous cells and how they contribute to chromosome stability and the initiation of senescence or apoptosis.

1.6.1 Modelling telomere loss

The first papers on quantitative modelling of telomere dynamics describe the process of telomere shortening by simple deterministic [79] and probabilistic [6, 99] models, which only take account of losses due to the end-replication problem. The probabilistic models account for the variability in the lifetimes of cells, assuming that intermitotic times are exponentially distributed, and use Markov branching processes [6, 99]. Cell cycle arrest is assumed to set in as soon as one [6] or more [79] telomeres reach a critical length (zero), or cells die with a probability that differs for cells containing a single critically short telomere [99]. The dynamics of telomere loss is described according to the following principle (see also Figure 1.18): if l_i and r_i denote the length of telomeres at the left and right end of a chromosome, respectively, with $i = 1$ denoting the leading strand (3′ end) and $i = 2$ the lagging strand (5′ end) of each telomere, we have, with the vector notation $(l_1, l_2; r_1, r_2)$, two types of transition rules,

$$(n, n-\mu; m, m) \begin{cases} \rightarrow (n, n-\mu; m, m) \\ \rightarrow (n-\mu, n-\mu; m, m-\mu)\ , \end{cases} \tag{1.6.1}$$

and

$$(n, n; m, m-\mu) \begin{cases} \rightarrow (n, n; m, m-\mu) \\ \rightarrow (n, n-\mu; m-\mu, m-\mu)\ , \end{cases} \tag{1.6.2}$$

where μ denotes the deletion due to the end replication problem. Hence, there are, depending on the choice of daughter chromosome, two deletions or no deletions during one chromosome replication, and telomere length is assumed not to change in chromosomes that are of type $(0, 0; m, m-\mu)$ and $(n, n-\mu; 0, 0)$.

The *in vitro* life span has been found [127] to vary widely between different cell cultures of normal human fibroblasts and also between individual clones derived from the same culture. The proliferative potential of cell clones has been shown to follow a bimodal distribution (see Figure 1.19), meaning that a subset of cells is subject to a much earlier, sudden onset of senescence when growing in culture.

Rubelj and Vondracek [115] extended the previously established models of telomere short-

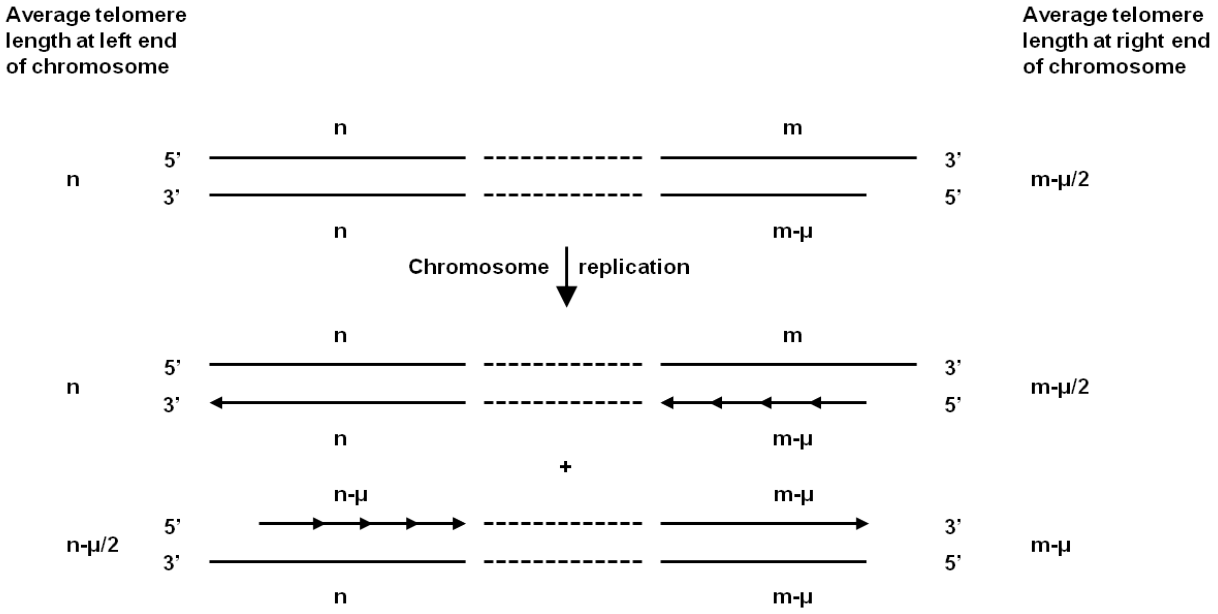

Figure 1.18: Telomere length dynamics caused by the end-replication problem leading to a deletion of length μ on telomeric DNA strands. The illustration is redrawn and adapted from Proctor and Kirkwood [106] with permission from Elsevier.

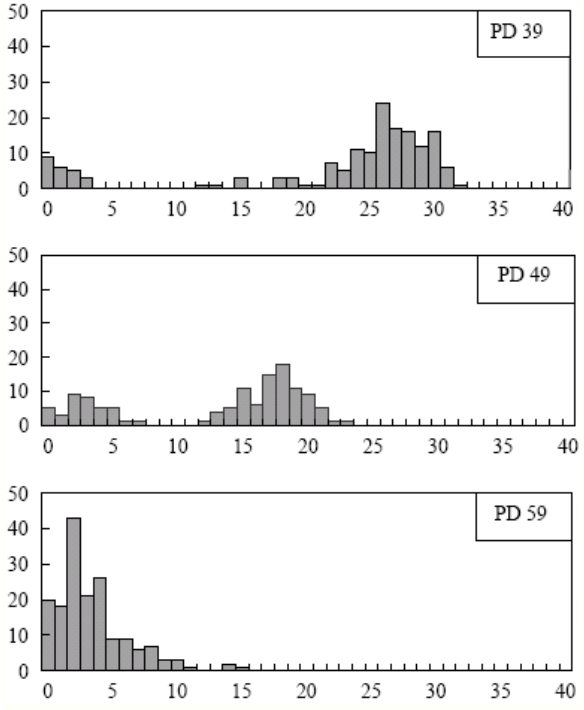

Figure 1.19: Frequencies of human diploid fibroblasts with varying population doubling (PD) potential, taken from Sozou and Kirkwood [129], who redrew the experimental data from Smith and Whitney [127]. Reprinted with permission from Elsevier.

ening ('gradual telomere shortening', see above) by the introduction of the possibility of 'abrupt telomere shortening' caused by DNA repair mechanisms due to accumulation of

DNA damage. Recombination between two telomeres, self-recombination or digestion of the C-strand by a nuclease produce sudden, stochastic changes in telomere length, which become more frequent as telomere shortening advances, and leading to a faster occurrence of senescence. Sozou and Kirkwood [129] included environment-dependent components affecting telomere shortening, where gradual telomere reduction is still the primary cause of senescence, but oxidative stress in the form of endogenous reactive oxygen species produced by mutant mitochondria, whose occurrence increases with age in various tissues, is assumed to be the cause of substantial telomere loss. The allocation of mitochondria and daughter chromosomes to daughter cells, the mutation of mitochondria and the amount of telomere loss are all random processes in this model. Cells also suffer somatic mutations in nuclear DNA whose accumulation is independent of telomere length and affects the expression of housekeeping genes. The lack of expression of these genes allows cells to undergo only a limited number of replications before they can no longer support cell division and become senescent. Similarly, Proctor and Kirkwood [106] modelled the onset of senescence by considering telomere shortening due to the end-replication problem and single-strand breaks, where they included that exonucleolytic C-strand processing can produce longer G-overhangs on either end of human chromosomes (see Figure 1.20). Single-strand breaks turned out to play a significant role in telomere shortening only at high levels of radicals in the nucleus. Another approach explaining the bimodal distribution of the doubling potential of individual cells in a clone is presented by Tan [133], who considered variations in 3'-overhang lengths observed in human fibroblasts (see Figure 1.21), leading to stochastic rates of telomere shortening. The stochasticity of the overhang length is produced by the sum of (i) the sequences that were occupied by the last RNA primer at the end of the DNA template, which is removed later, and (ii) the sequences that were missed by the primer at the end of the telomere.

Telomere length has been measured using different techniques, among which telomere restriction fragment (TRF) analysis using Southern blotting [69], and quantitative fluorescence *in situ* hybridization (Q-FISH) [104] have been frequently used. The former method uses restriction enzymes that digest genomic DNA, but do not cut telomeric sequences. The telomeric DNA fragments are then separated by gel electrophoresis and

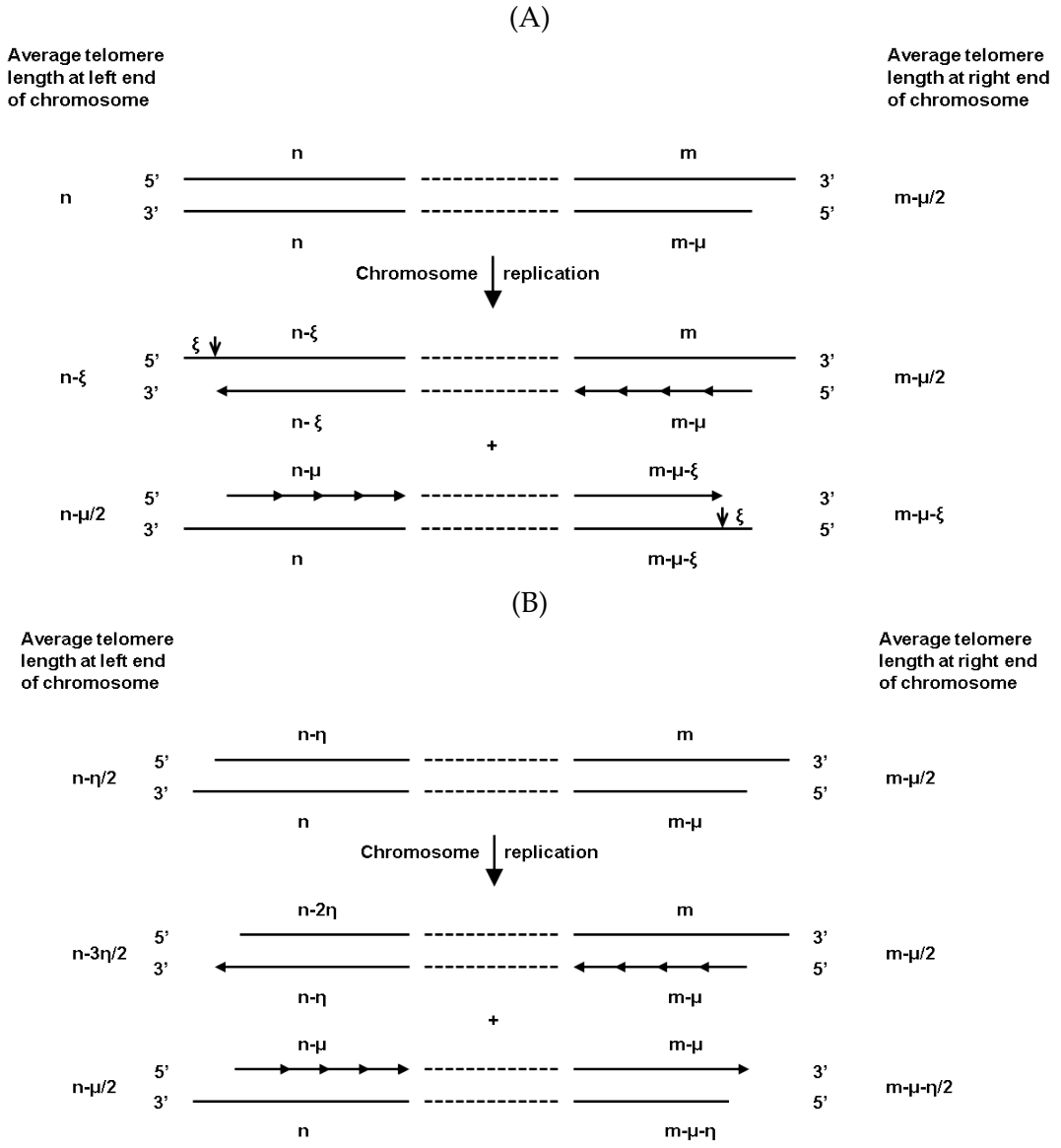

Figure 1.20: An illustration of telomere length dynamics due to the end-replication problem (deletion of length μ, see also Figure 1.18) and, additionally, (A) single-strand breaks at 5′ ends leading to a deletion of length ξ or (B) C-strand (5′ end) processing leading to a deletion of length η at telomeric ends. The position of single-strand breaks in (A) is indicated by down arrows (↓). Redrawn and adapted from Proctor and Kirkwood [106] with permission from Elsevier.

hybridized to labelled probes specific for telomeric repeats, which subsequently reveal the average length of telomeric DNA in the form of a smear in gel. The latter method, in contrast, can only be applied to metaphase chromosomes, but provides measurements for the telomeric length of individual chromosomes, where staining of chromosomal regions with a combination of telomere-specific and chromosome-specific probes allows for a comparison of the resulting fluorescence signals of individual telomere spots. The length of telomeres with a known subtelomeric DNA sequence can be detected by Single

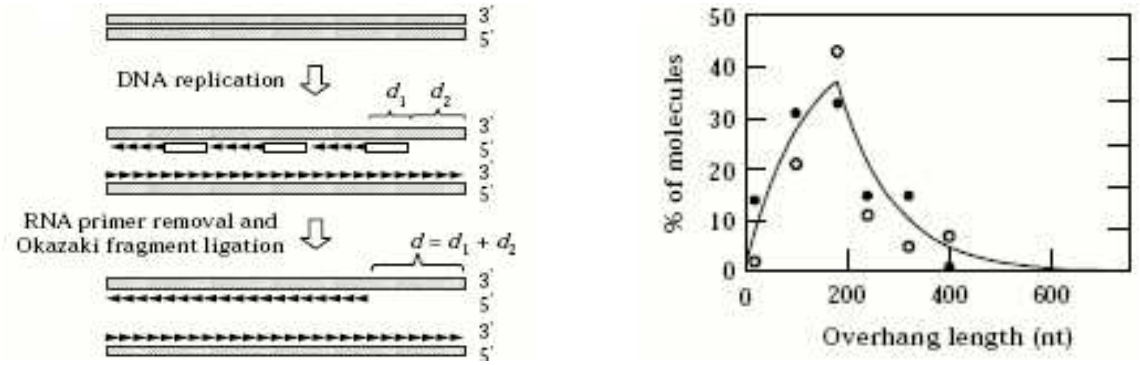

Figure 1.21: Incomplete DNA replication causes telomere deletion of varying size. Taken from Tan [133]. Reprinted with permission from Elsevier.

Telomere Length Analysis (STELA), which is also appropriate for measuring the length of very short telomeres [21]. The size of single-stranded telomeric DNA (G-overhangs) can be measured by the Telomere-Oligonucleotide Ligation Assay (T-OLA) and the overhang protection assay (OPA). Reviews of the techniques of telomere length measurement can be found in the literature, including focus on factors such as telomeric proteins that influence telomere length [117], on techniques for cloning, sequencing, visualization and analysis of the three-dimensional structure (G-quartets or telomeric loops) of telomeres [43], or on factors contributing to variations in average and individual telomere length of human cells [119].

Investigation of Martens et al. [85] into human diploid fibroblasts having a limited lifespan showed that short telomeres increasingly accumulate in cells and the length distribution of telomeres becomes positively skewed close to senescence (Figure 1.22). Here, the *mean* telomere length, and not the length of the shortest telomere, seemed to be correlated with the onset of replicative senescence.

Proctor and Kirkwood [107] considered the uncapping of telomeres by the opening of t-loops as a trigger for replicative senescence, which they incorporated into their previous model [106] to account for the experimental results found by Martens et al. [85]. Not just critically short telomeres, but the uncapping of *any* telomere is here the trigger of permanent cell cycle arrest. The shorter the telomeres, the fewer binding sites for TRF2 are present, and, in turn, the less likely t-loop formation becomes. Hence the probability of uncapping is modelled as a decreasing function of telomere length. Golubev et al. [48] investigated different possible causes for the observed positive skewness

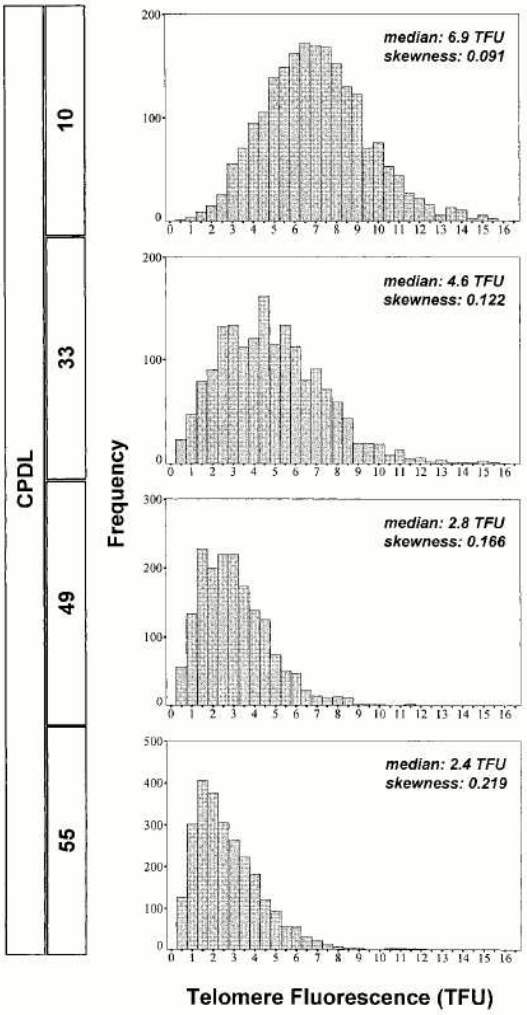

Figure 1.22: Increasingly skewed telomere length distributions of fibroblast cells at four different cumulative population doublings (CPDL). Taken from Martens et al. [85]. Reprinted with permission from Elsevier.

of telomere length distribution by mathematical modelling. Computer-assisted simulations, using virtual cells that contain only one telomere each, show that the Hayflick limit is markedly influenced by DNA damage caused by free radicals, and not by telomere length itself. The frequency of committing events, that is, processes such as changes in gene expression and cell differentiation, committing a cell to divide only a certain number of times before cell-cycle arrest, also increases with the increase of free radicals. Grasman et al. [50] characterise the dynamics of telomere shortening by the property that longer telomeres are more vulnerable to oxidative stress, as they are larger targets. Hence, DNA strand breaks, causing extensive telomere shortening, occur more frequently in longer telomeres.

Note that different numbers of chromosomes with critically short telomeres, varying from only one to twenty, or the average telomere length, are assumed to be necessary for the signalling of senescence in the models above, indicating that the issue of how many chromosomes contribute to the permanent arrest of the cell cycle has not yet been resolved.

Models on population growth dynamics that are determined by telomere length dynamics have been established by Portugal et al. [105] and Olofsson [97], who explain the occurrence of sigmoidal growth behaviour observed in mammalian and yeast cells, respectively. Portugal et al. [105] refers to experimental findings showing that telomere shortening correlates linearly with the probability of cell division, that is, cells of a certain cell line with shorter telomeres divide less frequently than cells with longer telomeres. Olofsson [97] investigates the reproduction of budding yeast cells, where the mother cell is distinct from the daughter cell and is able to reproduce several times. Furthermore, Itzkovitz et al. [61] develop a population mixture model with a re-populating pool of stem cells of constant telomere length and a derived pool which experiences constant decrease in telomere length, where one daughter cell of the repopulating pool stays and one transfers to the derived pool after cell replication (by differentiation). The model considers constant cell losses and predicts exponential decrease in telomere length over time.

1.6.2 Modelling telomere length regulation by telomerase

The enzyme telomerase is active at a low level in some somatic cells [86], such as in normal human fibroblasts, preserving the G-overhang and telomeric structure in these cells.

Telomerase-dependent shortening, leading to positively skewed telomere length distributions (see Martens et al. [85]), has been explained by op den Buijs et al. [100] through telomere-length-dependent negative feedback regulation. The model describes telomere loss per population doubling as

$$\Delta T = -\left(\left(1 - f_s\right) T + \Delta T_0\right) , \tag{1.6.3}$$

where T is the current telomere length, ΔT_0 the constant loss due to the end-replication

problem and f_s is a shortening factor, probably dependent on the amount of C-strand processing, oxidative damage and telomerase processivity. For positive $f_s < 1$ and when a Weibull distribution is assumed to represent the initial telomere length distribution, the resulting distributions fit experimental data well, suggesting that telomere shortening is a length-dependent mechanism in telomerase-positive cells.

However, telomere length is maintained in most telomerase-positive cells and mechanisms contributing to telomere length equilibrium have been considered in a number of publications. Blagoev [24] proposed a model in which telomere extension by telomerase occurs more frequently at short telomeres than at long telomeres. A logistic function of telomere length describes the probability of the occurrence of an extendible state of telomeres, opposed to a capped state, which was inspired by data from experiments in Teixeira et al. [136] on telomere elongation in yeast cells. Furthermore, Kowald [74] develop an ODE model involving the concentration of a capping protein, which can bind the 3′ overhang once it is sufficiently long, inhibits telomerase and facilitates DNA replication by maintenance of the single-stranded overhang it is bound to. The capping protein is released after telomere replication when telomerase acts on both chromosome ends, synthesising 3′ overhangs until free capping proteins block the 3′-end. In addition to telomere length maintenance, Kowald [74] explained the increase of telomere length levels when oligonucleotides are added to cell culture medium, as experimentally observed by Wright et al. [150], by considering extracellular and intracellular oligonucleotide concentrations in interaction with telomerase and capping protein levels. On the other hand, Sidorov et al. [125] investigated the impact of telomerase inhibition on the growth of tumours possessing either homogeneous or heterogeneous telomere length distributions, and determined the relationship between initial telomere length distributions, replicative capacity of cells and the time it takes for a tumour to reduce in volume to a certain threshold value.

The telomeric protein TRF2 binds to telomeric repeat sequences and controls telomere length dynamics by the inhibition of telomerase. To account for the assumption that the *state* of the telomere rather than its length determines the fate of a cell, Arkus [7] considered the fraction of viable cells using Michaelis-Menten equations for the binding and dis-

sociation of TRF2 dimers to telomeric repeat sequences, assuming that TRF2 caps telomeres, inhibits telomerase and that lack of TRF2 on telomeric ends causes senescence and apoptosis (thus the cell fate is only indirectly dependent on telomere length). Rodriguez-Brenes and Peskin [113] proposed another approach of modelling telomere length maintenance processes based on the biophysics of t-loop formation, which determine the state of a telomere and also decide the cell's fate. The shelterin component TRF2 can facilitate the conversion of telomeres into t-loops, which form the capped telomere states that cannot be elongated by telomerase. An ODE model and a stochastic model involve shortening by the end-replication problem, C-strand processing and telomerase-induced telomere elongation, where the senescence onset point is determined by the amount of time that telomeres remain in an uncapped state rather than by telomere length. Telomere equilibrium length has been simulated as a function of telomerase molecule number and as a function of TRF2 binding frequency. Simulated telomere length distributions fitted well to data of somatic and immortal cell lines in that work (see Figure 1.23) .

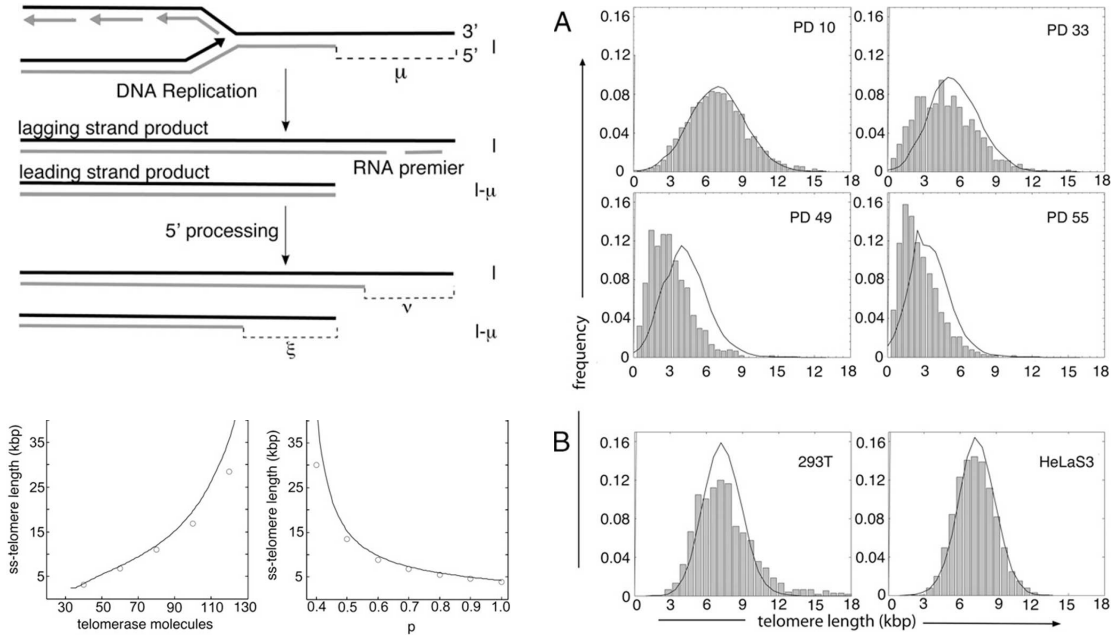

Figure 1.23: Schematics of telomere shortening (top left), a graph illustrating the dependence of the equilibrium telomere length on the number of telomerase molecules, and the probability p that any site in the telomere is bound to TRF2 (bottom left). Model simulations fit telomere length histogram data (right) collected from somatic cells (A, Martens et al. [85]) and immortal cell lines (B, Canela et al. [26]) well. Taken from Rodriguez-Brenes and Peskin [113]. Reprinted with permission from PNAS.

Other work on the dynamics of telomere length by Qi [108] compares the effects of normal ageing, accelerated ageing of patients with Werner's syndrome and the unlimited

lifespan of telomerase-positive cells. The typical way the process of telomere replication is modelled for normal ageing is

$$K_l^g \longrightarrow K_{l-\mu+T}^{g+1} + K_{l+T}^{g+1} \tag{1.6.4}$$

where K_l^g is the number of telomeres of length l at generation g, where it is assumed that telomere loss of amount μ occurs on only one daughter telomere, that is, one telomere of length l produces one telomere of length l and one of length $l-\mu$ in generation $g+1$. Telomerase elongates daughter telomeres with rate T per cell division, possibly with $T = 0$. The process (1.6.4) can be mathematically modelled by the algebraic equation

$$K_l^{g+1} = K_{l+\mu-T}^g + K_{l-T}^g \,. \tag{1.6.5}$$

Equation (1.6.5) permits separable solutions of the form $K_l^g = e^{\gamma g + \chi l}$, where the growth rate γ depends on χ, the rate of change of the telomere length distribution with l. Substituting this solution into (1.6.5), assuming that χ is small, gives

$$e^{\gamma} \approx 2 + A\chi + B\chi^2 \,, \tag{1.6.6}$$

with appropriate constants A and B. Taking the logarithm of (1.6.6) and expanding for small χ yields the dispersion relation

$$\gamma = \ln 2 + \ln(1 + A\,\chi + B\chi^2) \approx \ln 2 + \frac{A}{2}\,\chi + \frac{4\,B - A^2}{8}\,\chi^2 \,. \tag{1.6.7}$$

Considering a high number of replication events and treating the generation number g as a continuous time variable, t, as well as the telomere length l and the number of telomeres $K_l^t = K(l,t)$ as continuous, one can formulate a continuum analogue of equation (1.6.5), namely

$$\frac{\partial K}{\partial t} = K \ln 2 + \frac{A}{2}\frac{\partial K}{\partial l} + \frac{4\,B - A^2}{8}\frac{\partial^2 K}{\partial l^2} \,, \tag{1.6.8}$$

being the simplest partial differential equation with the dispersion relation (1.6.7). Under appropriate boundary conditions, solutions of (1.6.8) are Gaussian distributions, hence (1.6.8) can be used to simulate the time evolution of telomere length distributions (see

Figure 1.24 for an example). Additionally, Qi [108] considered models involving telomere length-dependent telomere loss and/or gain as well as telomere length-dependent probabilities for cell division.

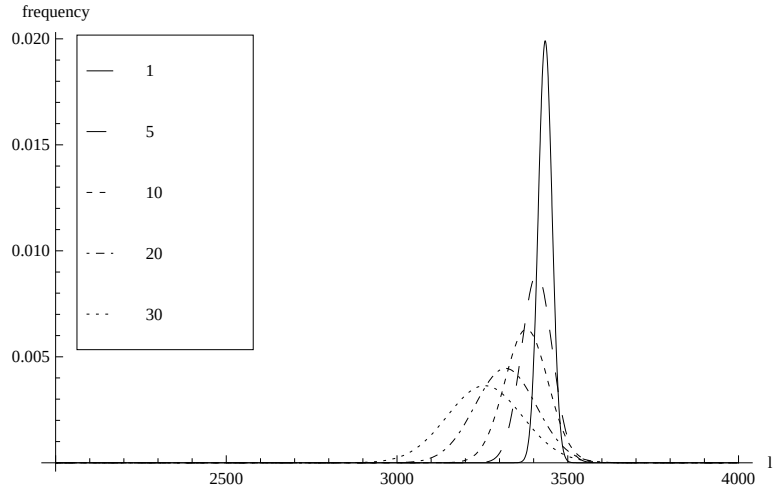

Figure 1.24: Telomere length distributions, $K(l,t)$, simulated using (1.6.8) for different generations ($t = 1, 5, 10, 20, 30$) in a telomerase-positive cell, where telomere shortening dominates telomere elongation.

1.6.3 Modelling telomerase-independent telomere maintenance

Telomerase-independent pathways of telomere length maintenance are considered in Olofsson and Bertuch [98] who used general branching processes to capture the mechanisms of survivorship of individual budding yeast cells, by which they bypass the crisis point and develop a recombination-based telomere-maintenance mechanism. Antal et al. [5] also presented an approach to understand the mechanism of alternative lengthening of telomeres (ALT) by modelling the effects of telomere sister-chromatid exchange through superimposing stochastic telomere length variations upon the systematic decrease in telomere length, which leads to wide distributions of telomere length, as indeed has been observed in cultures of ALT cells.

1.7 Anti-cancer effects of RHPS4

We aim to briefly review the effects of an anticancer drug which we consider later in greater detail, focusing on the dynamics of the cell cycle and telomeres in cancer cells.

We mentioned in Sections 1.4 and 1.5 that increasing the concentration of the metal ions associated with G-quadruplexes leads to augmented G-quadruplex formation. Stabilisation of G-quadruplexes by specific ligands can limit telomerase activity and alter telomere function in cancer cells. Anti-cancer researchers are now trying to design G-quadruplex ligands that will mimic the effect of the metal ions and inhibit telomerase, with the aim of achieving antitumour activity through effective stabilisation of G-quadruplexes [89].

The pentacyclic acridinium salt RHPS4 (3,11-difluoro-6,8,13-trimethyl-8H-quino $[4,3,2\text{-}kl]$ acridinium methosulphate, see Figure 1.25) is such a potent telomerase inhibitor [49], and induces a marked decrease of cell growth in human cell lines such as the 21NT breast cancer cells and A431 vulval carcinoma cells *in vitro* after 15 days and for concentrations lower than the level of acute cytotoxity. Furthermore, Cookson et al. [34] showed a notable reduction in telomere length of MCF-7 breast cancer cells when treated with subtoxic doses of RHPS4. RHPS4 also rapidly induces telomere dysfunction by telomere uncapping, which leads to short-term cell death through usage of higher doses and is dependent on the ATR damage response pathway [118] (primarily sensing single-strand breaks, see Section 1.4).

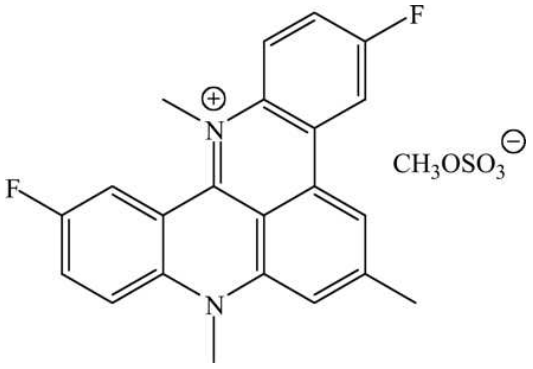

Figure 1.25: Chemical structure of the pentacyclic acridinium salt RHPS4, taken from Cookson et al. [34]. Reprinted with permission from ASPET.

The G-quadruplex ligand RHPS4 together with the 3,6,9-trisubstituted acridine compound BRACO-19 [53] and telomestatin [135], are promising compounds among the cancer inhibitor agents and have come close to clinical testing [22]. The compounds all inhibit telomerase activity, limiting long term proliferation of cancer cells, and directly target components of the protective cap of telomeres, leading to immediate effects on cancer cell proliferation [92]. Both critically short telomeres and direct disruption of the shelterin

structure can initiate telomere dysfunction and thus block tumorigenesis in most tissues by the activation of either an apoptosis or a senescence pathway [41]. RHPS4 has been found to reduce the growth of human tumours xenografted[1] in mice *in vivo* and did not show any toxic effects in mice. Also, RHPS4 did not induce telomere damage in normal human fibroblasts and did not show any signs of toxicity in nonmalignant human cells, suggesting that RHPS4 preferentially targets tumour cells over normal cells [118]. Moreover, the therapeutic efficacy of RHPS4 appeared to be superior to conventional drugs that are administered in human cancer therapy, particularly when the compound was used in combination with certain camptothecins (a group of antineoplastic agents), in mice that bear human tumours of different cell lines [78].

Only a few experiments to understand the effects of RHPS4 on cell cycle dynamics have been performed so far. RHPS4 treatment of human melanoma lines possessing relatively long telomeres resulted in a dose-dependent decrease in cell replication and accumulation of cells in the S-G_2/M phase of the cell cycle for short term assays of 10 days [77]. Furthermore, Rizzo et al. [112] showed that chronic exposure of transformed fibroblasts expressing hTERT to RHPS4 activates a DNA damage response when cells enter the S phase of the cell cycle, which is followed by a transient delay in the S-G_2 transition at days 4 and 5. The precise cell-cycle specific behaviour of RHPS4 and its mechanism of action in cancer cells, however, are still to be elucidated.

1.8 Outline of thesis

The effects of the potential anti-cancer drug RHPS4 are not yet well understood and the aim of this thesis is to investigate how RHPS4 affects the cell cycle dynamics of cancerous cells. We use differential equation models to describe the dynamics of the cell cycle and fit our models to the experimental data we collected in the tissue culture lab. Telomere length is a key factor determining the proliferative potential of a cell; therefore, a second objective of this work is to consider mechanisms regulating telomere length to understand how RHPS4 affects telomere dynamics at different drug concentrations and

[1] A xenograft is a graft of tissue taken from a donor of one species (here: a human) and grafted into a recipient of another species.

different time scales. The thesis is divided into three parts, altogether containing eight chapters, which are outlined below.

The first part contains two further preliminary chapters. In Chapter 2, we consider various compartmental differential equation models of cell cycle dynamics and analyse their behaviour. In Chapter 3, we introduce the problem of structural and practical model identifiability, which is important when fitting a model to data, and present two methods to analyse linear models for structural identifiability.

The second part consists of three chapters investigating the effects of RHPS4 on cell cycle dynamics. In Chapter 4, we describe the experimental tissue-culture work, where we analyse and interpret the behaviour of human cancer cells exposed to different concentrations of the drug RHPS4 with respect to their DNA content, growth and intensity of cell death. In Chapter 5, we develop a compartmental model of the cell cycle and introduce a statistical description of the experimental data based on the deterministic model. We also check our model for structural identifiability and introduce parameter estimation and model selection techniques for fitting the model to data. In Chapter 6, we apply the techniques of model fitting introduced in Chapter 5 and discuss the results in terms of accuracy of fit, parameter sensitivity and biological implications.

In the third part, Chapter 7, we develop and analyse several differential equation models of the effects of RHPS4 on telomere length dynamics based on experimental findings in the literature, where we consider different time-scales, from one cell cycle to a large number of cell replications.

Finally, in Chapter 8, we review the results of our work and discuss directions for future work.

CHAPTER 2

Compartmental ODE, PDE and DDE models of the cell cycle

The overall aim of the project is to use mathematical modelling to investigate how RHPS4 affects the cell cycle. It is not yet known for sure whether the drug is cell-cycle specific, although preliminary work of Johnson [64] suggests that it is. We aim to investigate which phase of the cell cycle is primarily affected by the drug, and whether, at low concentration, the drug causes cessation of cycle and, at higher concentration, the drug causes cell death. Answers to these questions could also provide information about the optimal dose of RHPS4 to inhibit telomerase effectively.

The following models embody different assumptions about how cells progress through the cell cycle, in particular through the S and G_2 phases, which we think is a potentially important part of the cell-cycle dynamics, in view of the drug treatment with RHPS4. We analyse the according models with respect to their typical dynamic behaviour.

We introduce a simple three-compartment model in Section 2.1, which we extend to a five-compartment model in Section 2.2, where we analyse and simulate the solution behaviour of this model. A further extension of this model is given in Section 2.3, comprising the typical cell cycle phases in seven compartments. Based on the three-compartment ODE model of Section 2.1, in Section 2.4 we develop a PDE model by splitting the compartment of cycling cells into a large number of subcompartments. We also consider the waiting time distributions and the notion of the residence time of cells within a compart-

ment. Furthermore, we show how the PDE model can be reduced to a delay differential equation (DDE) model, which we analyse in Section 2.5 finding oscillatory behaviour. We also simulate the dynamic behaviour for different values of the delay term occurring in the DDE model. Most of work in this Chapter is derived from existing models; however, the analysis of Section 2.5 is original. Section 2.6 concludes the chapter.

2.1 A three-compartment ODE model

We first model the phases of the cell cycle using a reduced number of compartments. Three compartments describe the G_0+G_1 phase (X compartment), the S+G_2+M phase (Y compartment) and the senescent/apoptotic phase (A compartment), as shown in Figure 2.1. We derive this model from the two-compartment model (shown in Figure 1.5) developed by Panetta and Adam [102], where we add a compartment (A) that collects the outflow from the compartments of resting and cycling cells. The model presented in this Section is a basis for the design and demonstration of the more complex models presented in the following Sections.

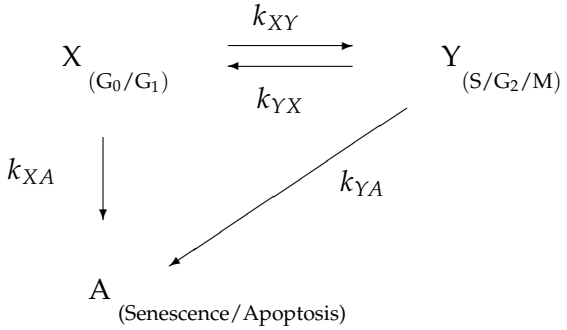

Figure 2.1: A three-compartment model containing compartments of resting, cycling and permanently arrested cells (senescence/apoptosis).

Cells in the Y compartment have committed to the cell cycle and double at the end of the M phase, giving rise to two cells in the X compartment at transfer from Y to X. Cells in the X compartment, on the other hand, may be quiescent (G_0 cells) or in the G_1 phase. We combine G_0 and G_1 cells into one compartment as it is difficult to distinguish between these phases experimentally. We assume that cells can exit the cell cycle permanently from the X and the Y compartment by undergoing senescence or apoptosis. The mass

action equations

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 2\,k_{YX}\,Y - (k_{XY} + k_{XA})\,X\,, \tag{2.1.1}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = k_{XY}\,X - (k_{YX} + k_{YA})\,Y\,, \tag{2.1.2}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = k_{XA}\,X + k_{YA}\,Y\,, \tag{2.1.3}$$

describe the dynamic behaviour of the numbers X, Y and A of cells in the G_0+G_1 phase, the S+G_2+M phase and the senescent/apoptotic phase, respectively. Equation (2.1.3) decouples from (2.1.1) and (2.1.2), it hence suffices to consider the equations for X and Y alone in order to find the solutions to this model. The characteristic polynomial of the corresponding coefficient matrix

$$\mathbf{M} = \begin{pmatrix} -k_{XY} - k_{XA} & 2\,k_{YX} \\ k_{XY} & -k_{YX} - k_{YA} \end{pmatrix}\,, \tag{2.1.4}$$

is a quadratic in the eigenvalue λ, hence it is possible to solve the system of ODEs of this model explicitly. The solution for the state variable A can then be easily derived by integration of (2.1.3) with respect to t. The eigenvalues of $\mathbf{M}$ are

$$\lambda_{1,2} = \frac{1}{2}\left(-b \pm \sqrt{b^2 - 4\,c}\right)\,, \tag{2.1.5}$$

where $b = k_{XY} + k_{YX} + k_{XA} + k_{YA}$ and $c = k_{XY}\,(k_{YA} - k_{YX}) + k_{XA}\,(k_{YA} + k_{YX})$, that is $b^2 \geq 4\,c$. Since all rate parameters are positive, $b^2 > 4\,c$, and we have one eigenvalue $\lambda_1 < 0$ and for the other eigenvalue $\mathrm{sgn}\,(\lambda_2) = -\mathrm{sgn}\,(c)$. Exponential growth occurs thus only if $\lambda_2 > 0$ ($c < 0$), that is, $k_{XA}/k_{XY} < (k_{YX} - k_{YA})\,/\,(k_{YX} + k_{YA})$, and a non-trivial steady state solution of (2.1.1)-(2.1.2) exists if and only if $c = 0$, that is, $k_{XA}/k_{XY} = (k_{YX} - k_{YA})\,/\,(k_{YX} + k_{YA})$, when we find asymptotically linear growth in the A compartment. A steady state of the full system (2.1.1)-(2.1.3) requires $X = Y = 0$ and A to be a constant that is determined by the initial condition on $A(0)$.

The three-compartment model is an oversimplified description of the actual cell-cycle dynamics and we do not expect it to capture biological behaviour very well. An extension

of the three-compartment ODE model together with a more detailed analysis is given in the following Section.

2.2 A five-compartment ODE model

We now model all phases of the cell cycle that can be directly observed by DNA content measurements and senescence indicators to provide a basic model which can be fit to available data sets and includes all relevant transition rates to determine how the drug RHPS4 affects the cell-cycle dynamics. Johnson et al. [65] have previously developed a five-compartment model of cycling and non-cycling cells that includes transitions of cells from the G_0/G_1, S, G_2/M and the senescent phase, which assumes that cells only leave the cell cycle from the G_0/G_1 phase when they become senescent and then eventually undergo cell death. We extend that model by including possible transitions from all other phases of the cell cycle to the apoptotic phase to account for potential cell death due to the influence of higher drug doses.

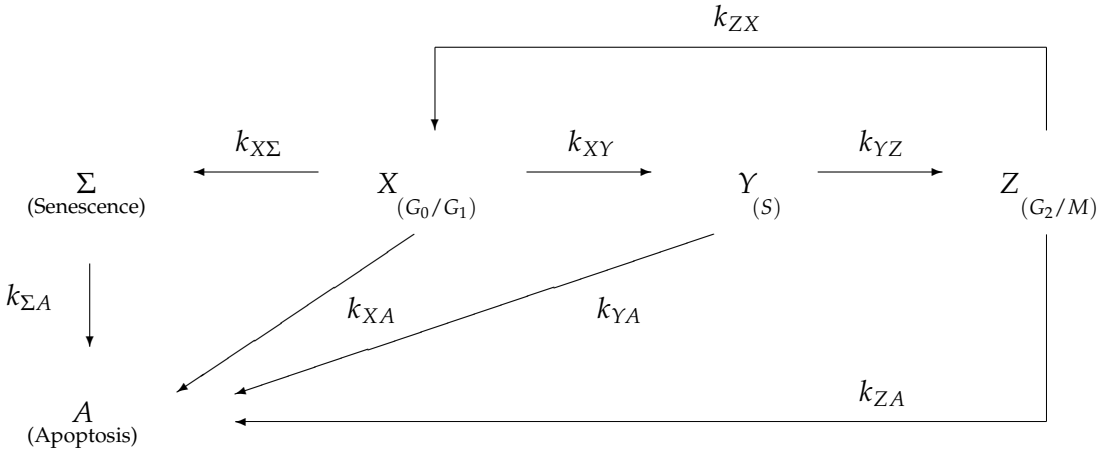

Figure 2.2: A model with five compartments and transitions between the compartments and to apoptosis from the X, Y, Z and Σ compartments describes the basic dynamics of the cell cycle. The phases which correspond to the single compartments are given in brackets.

The five-compartment model is depicted in Figure 2.2 and is represented by the system

of ODEs

$$\begin{aligned}
\frac{\mathrm{d}X}{\mathrm{d}t} &= 2\,k_{ZX}\,Z - (k_{XY} + k_{X\Sigma} + k_{XA})\,X\,, && (2.2.1)\\
\frac{\mathrm{d}Y}{\mathrm{d}t} &= k_{XY}\,X - (k_{YZ} + k_{YA})\,Y\,, && (2.2.2)\\
\frac{\mathrm{d}Z}{\mathrm{d}t} &= k_{YZ}\,Y - (k_{ZX} + k_{ZA})\,Z\,, && (2.2.3)\\
\frac{\mathrm{d}\Sigma}{\mathrm{d}t} &= k_{X\Sigma}\,X - k_{\Sigma A}\,\Sigma\,, && (2.2.4)\\
\frac{\mathrm{d}A}{\mathrm{d}t} &= k_{\Sigma A}\,\Sigma + k_{XA}\,X + k_{YA}\,Y + k_{ZA}\,Z\,, && (2.2.5)
\end{aligned}$$

which describe the rate of change of the number of cells within the compartments X, Y, Z, Σ and A corresponding to G_0/G_1, S, G_2/M, senescent and apoptotic cells, respectively. The quantities k_* denote the rates of transition between the compartments. Cells from the Z compartment divide, yielding two cells in the X compartment. Many cells divide only a limited number of times before they undergo a permanent cell-cycle arrest (replicative senscence, Σ). From the senescent phase, cells usually undergo cell death (apoptosis, A).

Equations (2.2.1)-(2.2.3) for X, Y and Z are mathematically independent of Σ and A. Therefore, it is possible to consider separately the 3×3 system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} -k_{XY} - k_{X\Sigma} - k_{XA} & 0 & 2\,k_{ZX} \\ k_{XY} & -k_{YZ} - k_{YA} & 0 \\ 0 & k_{YZ} & -k_{ZX} - k_{ZA} \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}. \quad (2.2.6)$$

Let $\mathbf{M}$ denote the coefficient matrix of (2.2.6) and let $\mathbf{v} = (X, Y, Z)^T$ be the vector of cycling cells X, Y, Z. The system (2.2.6) can accordingly be written as $\dot{\mathbf{v}} = \mathbf{M}\,\mathbf{v}$.

Now let $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ be the eigenvectors of $\mathbf{M}$ corresponding to eigenvalues $\lambda_1, \lambda_2, \lambda_3$, respectively. We assume that all rate constants k_* are positive and, for simplicity, that all eigenvalues are distinct. Solutions of (2.2.6) are then of the form

$$\mathbf{v}\,(t) = a_1\,e^{\lambda_1 t}\,\mathbf{w}_1 + a_2\,e^{\lambda_2 t}\,\mathbf{w}_2 + a_3\,e^{\lambda_3 t}\,\mathbf{w}_3\,, \quad (2.2.7)$$

where a_1, a_2, a_3 are real constants which are determined by the initial conditions.

Equation (2.2.4) is a first order linear differential equation and can be solved by the integrating factor method. We reformulate (2.2.4) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{k_{\Sigma A}\,t}\Sigma\right) = k_{X\Sigma}\,e^{k_{\Sigma A}\,t}X\,, \tag{2.2.8}$$

which yields

$$\Sigma\,(t) = a_4\,e^{-k_{\Sigma A}\,t} + k_{X\Sigma}\,e^{-k_{\Sigma A}\,t}\int e^{k_{\Sigma A}\,t}\,X\,(t)\,\mathrm{d}t\,, \tag{2.2.9}$$

with a_4 a real constant which is determined by the initial condition on Σ. The general solution for Σ is hence

$$\Sigma\,(t) = a_4\,e^{-k_{\Sigma A}\,t} + k_{X\Sigma}\sum_{i=1}^{3}\frac{a_i\,(\mathbf{w}_i)_1}{k_{\Sigma A}+\lambda_i}\,e^{\lambda_i\,t}\,, \tag{2.2.10}$$

$(\mathbf{w}_i)_1$ being the first component of $\mathbf{w}_i$.

Equation (2.2.5) can now be solved by integration with respect to t and is of the form

$$A\,(t) = b_5 + b_4\,e^{-k_{\Sigma A}\,t} + \sum_{i=1}^{3} b_i\,e^{\lambda_i t}\,, \tag{2.2.11}$$

where $b_1,\ldots,b_5$ are (possibly complex) constants. We observe exponential decay of X, Y, Z and Σ (A increases to the steady state solution $A = b_5$), with an asymptotically constant solution of system (2.2.1)-(2.2.5) if all of the real parts of the eigenvalues are negative. Otherwise, the system grows exponentially for large values of $t \geq 0$ (with oscillations in the case of complex eigenvalues).

To find a steady solution of system (2.2.6) for (X, Y, Z), one eigenvalue, say λ_1, must be zero. To simplify our analysis, we assume $k_{XA} = k_{YA} = k_{ZA} = 0$. The characteristic polynomial (for eigenvalues λ) of the system is then

$$\begin{aligned} p\,(\lambda) &= 2\,k_{XY}\,k_{YZ}\,k_{ZX} - (k_{X\Sigma}+k_{XY}+\lambda)\,(k_{YZ}+\lambda)\,(k_{ZX}+\lambda) \\ &= \sum_{i=0}^{3} c_i\,\lambda^i\,, \end{aligned} \tag{2.2.12}$$

where $c_i \in \mathbb{R}$ denote the coefficients of the characteristic polynomial. The constant coef-

ficient of the polynomial p is

$$c_0 = (k_{XY} - k_{X\Sigma})\, k_{YZ}\, k_{ZX}\,, \tag{2.2.13}$$

and since we need $c_0 = 0$ and $\lambda_1 = 0$ for a steady state solution, this is satisfied if and only if $k_{XY} = k_{X\Sigma}$ (remember that rate constants are assumed to be positive). Hence, the steady state solutions of system (2.2.6) are of the form $(d/k_{XY}, d/k_{YZ}, d/k_{ZX})^T$ with an arbitrary real (and positive) constant d. The remaining eigenvalues are then

$$\begin{aligned}\lambda_{2,3} &= \frac{1}{2}\Big(-2\,k_{XY} - k_{YZ} - k_{ZX} \pm \sqrt{4\,k_{XY}^2 + (k_{YZ} - k_{ZX})^2 - 4\,k_{XY}\,(k_{YZ} + k_{ZX})}\Big) \\ &= \frac{1}{2}\Big(-2\,k_{XY} - k_{YZ} - k_{ZX} \pm (2\,k_{XY} - k_{YZ} - k_{ZX})\Big)\,,\end{aligned} \tag{2.2.14}$$

that is, $\lambda_2 = -2\,(k_{YZ} + k_{ZX}) < 0$ and $\lambda_3 = -4\,k_{XY} < 0$. A steady state $(X^s, Y^s, Z^s)^T$ of system (2.2.6) hence implies $\Sigma(t) \to k_{X\Sigma}\, X^s / k_{\Sigma A}$ as $t \to \infty$. A steady state solution of (2.2.1)-(2.2.4) is therefore described by $(d/k_{XY}, d/k_{YZ}, d/k_{ZX}, d\,k_{X\Sigma}/(k_{XY}\,k_{\Sigma A}))^T$, in which case $A(t)$ and the total cell population $N(t) = X(t) + Y(t) + Z(t) + \Sigma(t) + A(t)$ both grow linearly for large values of t. Note that a steady state solution of the full system (2.2.1)-(2.2.5) implies $d = 0$, that is, $X = Y = Z = \Sigma = 0$ and $A = b_5$, due to $\lambda_{1,2,3} < 0$.

Equation (2.2.12) shows that the coefficients c_1, c_2 and c_3 of the characteristic polynomial are negative. Hence, $c_0 < 0$ implies that all real solutions of (2.2.12) are negative, and a positive solution exists only for $c_0 > 0$. Therefore, we expect to find exponential decay of X, Y, Z for $k_{XY} < k_{X\Sigma}$ and exponential growth for $k_{XY} > k_{X\Sigma}$.

We now show that if oscillations occur, they do not govern the behaviour of the system. For complex eigenvalues $\lambda = v + i\,w$ with $v, w \in \mathbb{R}$, we obtain respectively

$$\mathrm{Re}\,(p(\lambda)) = w^2\,(3\,v + k_2) - \left(v^3 + k_2\,v^2 + k_1\,v + k_0\right)\,, \tag{2.2.15}$$

$$\mathrm{Im}\,(p(\lambda)) = w\left(w^2 - 3\,v^2 - 2\,k_2\,v - k_1\right)\,, \tag{2.2.16}$$

with $k_0 = k_{YZ}\,k_{ZX}\,(k_{X\Sigma} - k_{XY})$, $k_1 = k_{YZ}\,(k_{XY} + k_{X\Sigma}) + k_{ZX}\,(k_{XY} + k_{YZ} + k_{X\Sigma}) > 0$ and $k_2 = k_{XY} + k_{YZ} + k_{ZX} + k_{Z\Sigma} > 0$ for the real and imaginary part of the characteristic polynomial in (2.2.12).

Function (2.2.16) is a cubic in w with roots $w_1 = 0$ and

$$w_{2,3} = \pm\sqrt{3\,v^2 + 2\,k_2\,v + k_1}\,, \tag{2.2.17}$$

that is, one eigenvalue, say λ_1, is real and $w_{2,3}$ are the imaginary parts of the two remaining eigenvalues, λ_2 and λ_3. Expression (2.2.15) is a quadratic in w with roots

$$\widetilde{w}_{2,3} = \pm\sqrt{(v^3 + k_2\,v^2 + k_1\,v + k_0)\,/\,(3\,v + k_2)}\,. \tag{2.2.18}$$

As both equations (2.2.17) and (2.2.18) must be fulfilled for complex eigenvalues, that is, the equation

$$8\,v^3 + 8\,k_2\,v^2 + 2\,(k_1 + k_2^2)\,v + k_1\,k_2 - k_0 = 0\,, \tag{2.2.19}$$

must be fulfilled for the eigenvalues λ_2 and λ_3, we obtain information on their imaginary parts by solving a cubic in v. As the constant coefficient, $k_1\,k_2 - k_0$, of the cubic (2.2.19) is positive, all solutions of equation (2.2.19) must be negative. Therefore, oscillations decay exponentially, and thus, when the overall solution is exponentially increasing, oscillatory behaviour vanishes with increasing time t.

The analysis of more complex forms of ODEs – using positive death rates (k_{XA}, k_{YA}, $k_{ZA} > 0$) or including more compartments – has been conducted using the computer algebra package MATHEMATICA, but yielded equations too complicated to be solved algebraically. Numerical simulations for the five-compartment model are shown in Figure 2.3, illustrating exponential growth and exponential decay behaviour of the cell-cycle system, and showing a system tending towards a steady state for the variables X, Y, Z and Σ.

2.3 A seven-compartment ODE model

Cells can leave the cell cycle and enter a resting state G_0 for a certain time, when the environment is unfavourable for cell division, and return to the cell cycle (i.e. $G_1 \rightarrow S \rightarrow G_2 \rightarrow M \rightarrow 2\,G_1$) when signals favouring replication are present. We propose a seven-compartment model derived from the six-compartment model developed by Venkatasub-

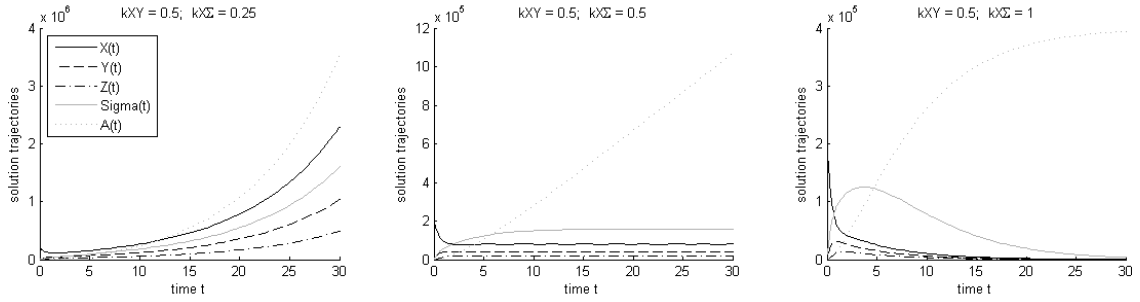

Figure 2.3: Numerical simulations of the five-compartment model (2.2.1)-(2.2.5) with $k_{YZ} = 1$, $k_{ZX} = 2$ and $k_{\Sigma A} = 0.25$, and k_{XY} and $k_{X\Sigma}$ as indicated above each plot, where we assumed $k_{XA} = k_{YA} = k_{ZA} = 0$. We observe exponential growth (decay) for $k_{XY} > k_{X\Sigma}$ ($k_{XY} < k_{X\Sigma}$) and cells approaching an equilibrium state in the compartment of living cells X, Y, Z, Σ, for $k_{XY} = k_{X\Sigma}$. Apoptotic cells (compartment A) grow exponentially, asymptotically linear or approach an equilibrium for $k_{XY} > k_{X\Sigma}$, $k_{XY} = k_{X\Sigma}$ or $k_{XY} < k_{X\Sigma}$, respectively.

ramanian et al. [139] (see Figure 1.8), which reflects the dynamics of the transitions between the G_0 and G_1 phase, and also distinguishes between the G_2 and M phase. In addition, we distinguish between cells in the G_0 and the senescent phase (Σ), where senescent cells, contrary to G_0 cells, do not reenter the cell cycle and eventually undergo apoptosis (A).

Dynamics in the G_2 phase may be particularly important regarding transitions to apoptosis for treatment of cells with a telomere-interactive drug. Rajaraman et al. [109] found that telomeres uncap and induce apoptosis directly in late S phase or in the G_2 phase when their telomeres become critically short. It may therefore suffice to include transitions to apoptosis from these phases (and the senescent phase) only. The seven-compartment cell-cycle model is shown in Figure 2.4.

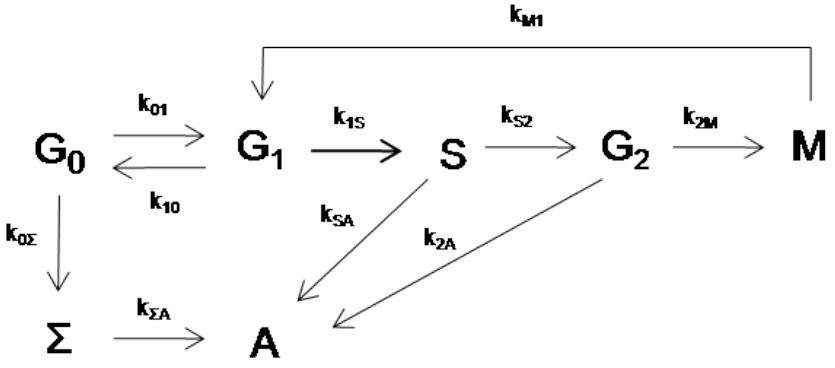

Figure 2.4: An ODE model with seven compartments describing all phases of the cell cycle and including transitions to apoptosis from the senescent phase Σ, the S phase and the G_2 phase.

The model equations for the seven-compartment model are given by

$$\frac{dG_0}{dt} = k_{10}\,G_1 - (k_{01} + k_{0\Sigma})\,G_0\,, \tag{2.3.1}$$
$$\frac{dG_1}{dt} = k_{01}\,G_0 + 2\,k_{M1}\,M - (k_{10} + k_{1S})\,G_1\,, \tag{2.3.2}$$
$$\frac{dS}{dt} = k_{1S}\,G_1 - (k_{S2} + k_{SA})\,S\,, \tag{2.3.3}$$
$$\frac{dG_2}{dt} = k_{S2}\,S - (k_{2M} + k_{2A})\,G_2\,, \tag{2.3.4}$$
$$\frac{dM}{dt} = k_{2M}\,G_2 - k_{M1}\,M\,, \tag{2.3.5}$$
$$\frac{d\Sigma}{dt} = k_{0\Sigma}\,G_0 - k_{\Sigma A}\,\Sigma\,, \tag{2.3.6}$$
$$\frac{dA}{dt} = k_{\Sigma A}\,\Sigma + k_{SA}\,S + k_{2A}\,G_2\,, \tag{2.3.7}$$

with transition rates k_* and state variables representing cell numbers in cell cycle phases as indicated by their denotation, with Σ and A denoting senescent and apoptotic cell numbers, respectively.

The seven-compartment model also accounts for the difference of cells in the G_0 and G_1 phase with respect to transitions to the senescence phase, which only occurs from the phase of quiescent cells (G_0) that have not yet committed to cell-cycle entry. However, it may be difficult to obtain data on all of the cell-cycle phases contained in the model, therefore the model may be too complex to infer parameter estimates from experimental data.

2.4 Development of PDE and DDE cell cycle models

The Y=S/G_2/M compartment of model (2.1.1)-(2.1.3) is of particular interest as it contains the phase of DNA synthesis which might be especially affected by the drug RHPS4. We therefore use different mathematical descriptions to account for the potential variability of the behaviour of cells in this compartment. The basic 3-compartment ODE model in Section 2.1 is a simplification of the five-compartment ODE model in Section 2.2 and enables us to derive solution trajectories explicitly; the age-distribution of cells in the Y-compartment in the three-compartment ODE model can be modelled by splitting the

compartment into several or infinitely many subcompartments resulting in a $(2+N)$-compartment ODE or a three-compartment partial differential equation model, respectively, the latter of which can be reduced to a 3-compartment differential-delay equation that describes the processes in the Y compartment by a delay term instead of a density function.

2.4.1 High order ODE model

The three-compartment ODE model in Section 2.1 describes the loss of cells in the Y-compartment by a constant rate. The effect of this is that residence times of cells in this compartment are exponentially distributed as shown below. However, cell material might not be homogeneous in the Y=S/G_2/M compartment, that is residence times may not be exponentially distributed. We can estimate the age of a particle at the time of transition from one compartment to another by probabilistic means.

We consider a simple system of two compartments A and B with constant transition rate k_{AB} at a fixed time point t (see Figure 2.5). We define $A_t(\tau)$ as the amount of material in the A-compartment, at time t, with age larger than or equal to τ, that is, material that has been in the A-compartment for a time of at least τ. The probability, at a time point t, for the age T_{AB} of cells at the time of transition from A to B being greater than a certain age τ is given by $\mathrm{P}(T_{AB} > \tau | \, t) = A_t(\tau)/A_t(0)$. We estimate $A_t(\tau)$ by approximating the loss of $A_t(\cdot)$ during an infinitesimal time interval $\Delta\tau$ as

$$A_t(\tau + \Delta\tau) - A_t(\tau) \approx -k_{AB}\,\Delta\tau\,A_t(\tau)\,. \tag{2.4.1}$$

As $\Delta\tau \longrightarrow \infty$, we obtain

$$\frac{\mathrm{d}A_t(\tau)}{\mathrm{d}\tau} = -k_{AB}\,A_t(\tau)\,, \tag{2.4.2}$$

which is solved by $A_t(\tau) = A_t(0)\,\mathrm{e}^{-k_{AB}\tau}$. We use this expression to show that the distribution of T_{AB} is time-independent, that is $\mathrm{P}(T_{AB} > \tau | \, t) = \mathrm{e}^{-k_{AB}\tau} = \mathrm{P}(T_{AB} > \tau)$. T_{AB} is thus exponentially distributed with rate $\lambda = k_{AB}$, that is, $T_{AB} \sim \mathrm{Exp}(1/k_{AB})$, where $X \sim \mathrm{Exp}(1/\lambda)$ means X is a random variable with distribution $\mathrm{Exp}(1/\lambda)$.

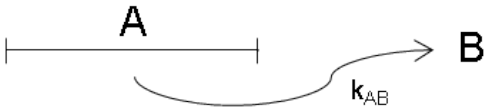

Figure 2.5: A two-compartment model, with compartments A and B, and constant transition rate k_{AB}.

The residence time of cells in a particular cell-cycle phase is the amount of time a cell spends in that phase. For a compartment, say A, with losses from the A compartment only of rate k_{AB}, the mean value of the probabilistic waiting time, T_{AB}, to the next transition from A to B is $1/\lambda = 1/k_{AB}$. An example for exponential distributions with different values of λ is given in Figure 2.6. However, it may be rather unlikely that cells leave a compartment immediately after entry into that compartment as it is suggested by the exponential distribution of the waiting time. If we split the compartment A into N subcompartments $A_1, \ldots, A_N$ with identical transition rates λ (as shown in Figure 2.7), the age $\widehat{T}_N$ at the time of transition from the compartment A to B is the sum over all transition ages T_i of each single compartment A_i; and as $T_i \sim \mathrm{Exp}(1/\lambda)$, which is equivalent to $T_i \sim \Gamma(1, 1/\lambda)$, that is, T_i is gamma distributed with scale 1 and shape $1/\lambda$, we find $\widehat{T}_N = \sum_{i=1}^{N} T_i \sim \Gamma(N, 1/\lambda)$ with mean N/λ. Examples for gamma distributions with different values of N and either constant $\lambda = 0.5$, or varying $\lambda = N/6$ are given in Figure 2.8.

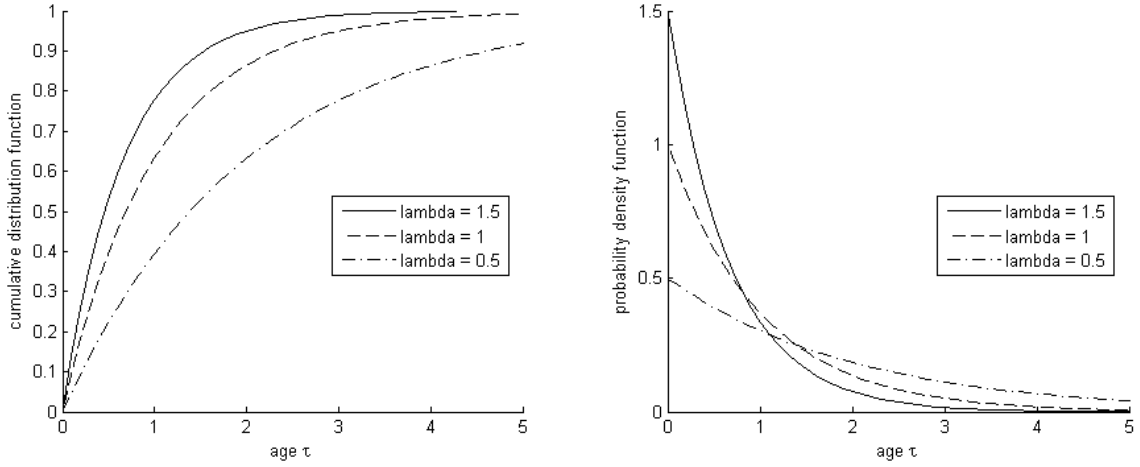

Figure 2.6: Cumulative distribution functions $F(\tau;\lambda) = 1 - \mathrm{e}^{-\lambda\tau}$, $\tau \geq 0$, (left) and probability density functions $f(\tau;\lambda) = \lambda\,\mathrm{e}^{-\lambda\tau}$, $\tau \geq 0$, (right) of the exponential distribution $\mathrm{Exp}(1/\lambda)$ with rates $\lambda = 0.5, 1, 1.5$.

The gamma distribution may be a better description of the real dynamics with respect to

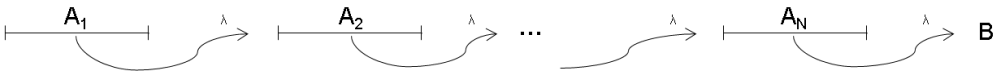

Figure 2.7: A model with two compartments A and B, where the A compartment is split into N subcompartments A_i. All transition rates are equal to a constant λ.

transition times, as the mode of its density function is shifted towards larger times as N tends to infinity; hence, for larger N, transitions are rather more likely to occur at a later time point than directly after entry into the compartment.

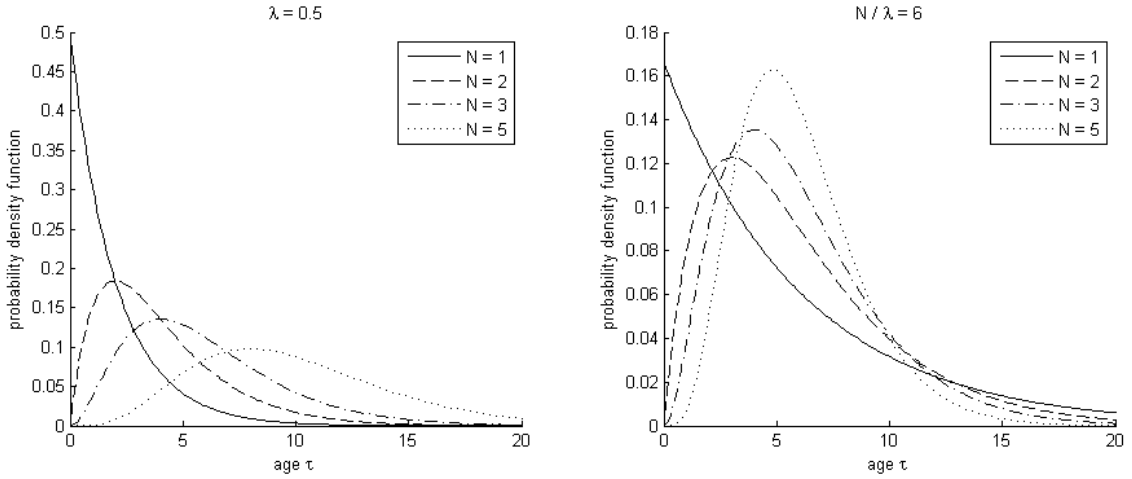

Figure 2.8: Probability density functions $f(\tau; N, \lambda) = \tau^{N-1}\,\lambda^N\,\mathrm{e}^{-\tau\lambda}/\Gamma(N)$, $\tau > 0$, of the gamma distribution $\Gamma(N, 1/\lambda)$ with scales $N = 1, 2, 3, 5$ and shape $1/\lambda > 0$. The graph on the left-hand-side shows f for the fixed rate $\lambda = 0.5$ and different values of N, whereas the graph on the right-hand-side shows f for varying λ with $N/\lambda = 6$.

We represent the Y compartment in the ODE model (2.1.1)-(2.1.3) by several compartments Y_i, $i = 1, \ldots, N$, as depicted in Figure 2.9, and attribute a rate parameter k_Y to the transition from one subcompartment to the next, and a different rate parameter, k_{YiA}, for the transition to the A compartment from each subcompartment.

For purposes of simplification, we assume that all rate parameters between the single Y_i compartments have same value k_Y, but describe loss terms from the Y_i compartments by

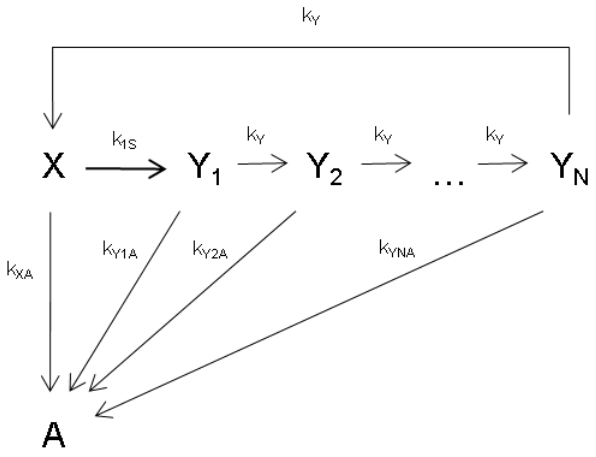

Figure 2.9: An ODE model with the Y compartment of cycling cells divided into N subcompartments, describing different stages of the S+G_2+M phase. The X compartment contains cells in the G_0/G_1 phases. The model contains transitions to apoptosis, A, at rate k_{YiA} from each Y_i compartment. Transitions between the Y_i compartments are all assumed to be equal.

arbitrary rate parameters k_{YiA}. The system of ODEs is given by

$$\frac{dX}{dt} = 2\,k_Y\,Y_N - (k_{XY} + k_{XA})\,X\,, \tag{2.4.3}$$

$$\frac{dY_1}{dt} = k_{XY}\,X - (k_Y + k_{Y1A})\,Y_1\,, \tag{2.4.4}$$

$$\frac{dY_i}{dt} = k_Y\,Y_{i-1} - (k_Y + k_{YiA})\,Y_i\,, \qquad i = 2,\ldots,N\,, \tag{2.4.5}$$

$$\frac{dA}{dt} = k_{XA}\,X + \sum_{i=1}^{N} k_{YiA}\,Y_i\,. \tag{2.4.6}$$

The model accounts for the processes in the S phase and the G_2/M phase by splitting the Y compartment into N subcompartments all of which contain different transition rates to the apoptotic phase.

2.4.2 Deduction of PDE model

The ODE-PDE model (1.2.3)-(1.2.7) in Chapter 1.2.2 (see Figure 1.7) of Basse et al. [15] describes the age-dependent evolution processes in the S phase by adding a second independent variable, τ_S, to the model, which denotes the age of a cell according to its entry into the S phase. We aim to deduce a similar model for the S+G_2+M phase and approximate the age-distribution of cells in the Y=S/G_2/M compartment by splitting the compartment into infinitely many subcompartments, that is by describing the Y compart-

ment by the number density $y(t,\tau)$ of the cells that, at time t, have been in the S+G_2+M phase for a time τ. We denote the maximum residence time for cells in this phase by T_Y. We assume cells undergo apoptosis with rate $k_{YA}(\tau)$ which may depend on the time τ spent in the S+G_2+M phase. Age-dependent transition rates have also been incorporated into cell-cycle modelling by Sherer et al. [124] (see Figure 1.6 for the corresponding three-compartment model), for example. We regard equations (2.4.3)-(2.4.6) as a piecewise continuous approximation of such an age-structured model and define the density function of the overall compartment $Y_1 + Y_2 + \ldots + Y_N$ by $\widehat{y}_N(t,\tau) = y(t, i\,\Delta\tau_N)$, and age-dependent rates to apoptosis by $\widehat{k}_{YA}^{(N)}(\tau) = k_{YA}(i\,\Delta\tau_N)$, for all $\tau \in ((i-1)\,\Delta\tau_N, i\,\Delta\tau_N]$ with $i = 1,\ldots,N$, and $\Delta\tau_N = T_Y/N$. We express Y_i in terms of $y(t, i\,\Delta\tau_N)$, namely

$$Y_i(t) = \int_{(i-1)\Delta\tau_N}^{i\,\Delta\tau_N} \widehat{y}_N(t,\tau)\,\mathrm{d}\tau = \Delta\tau_N\, y\big(t, i\,\Delta\tau_N\big)\,, \tag{2.4.7}$$

and similarly, $k_{YiA} = \widehat{k}_{YA}^{(N)}(\tau) = k_{YA}(i\,\Delta\tau_N)$ for $\tau \in ((i-1)\,\Delta\tau_N, i\,\Delta\tau_N]$. Equations (2.4.3)-(2.4.6) can be rewritten as

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 2\,k_Y\,\Delta\tau_N\,y(t,T_Y) - (k_{XY}+k_{XA})\,X\,, \tag{2.4.8}$$

$$\Delta\tau_N\,\frac{\partial y}{\partial t}\big(t,\Delta\tau_N\big) = k_{XY}\,X - \Big(k_Y + k_{YA}\big(\Delta\tau_N\big)\Big)\,\Delta\tau_N\,y\big(t,\Delta\tau_N\big)\,, \tag{2.4.9}$$

$$\frac{\partial y}{\partial t}\big(t,i\,\Delta\tau_N\big) = k_Y\,y\big(t,(i-1)\Delta\tau_N\big) - \Big(k_Y + k_{YA}\big(i\,\Delta\tau_N\big)\Big)\,y\big(t,i\,\Delta\tau_N\big)\,,\ (2\le i\le N)\,, \tag{2.4.10}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = k_{XA}\,X + \Delta\tau_N\sum_{i=1}^{N} k_{YA}\big(i\,\Delta\tau_N\big)\,y\big(t,i\,\Delta\tau_N\big)\,, \tag{2.4.11}$$

respectively.

For an arbitrary $\tau \in (0,T_Y]$ and every $N \in \mathbb{N}$, there is an $i_N \in \{1,\ldots,N\}$ such that $\widehat{y}_N(t,\tau) = y(t, i_N\,\Delta\tau_N)$ and $\widehat{k}_{YA}^{(N)}(\tau) = k_{YA}(i\,\Delta\tau_N)$, and it can be shown that respectively $\widehat{y}_N(t,\tau) \to y(t,\tau)$ and $\widehat{k}_{YA}^{(N)}(\tau) \to k_{YA}(\tau)$ as $N \to \infty$. If we assume that $k_Y = 1/\Delta\tau_N = N/T_Y$, that is, increasing transition rates with N between the Y_i compartments, we have

for $\tau \in ((i_N - 1)\,\Delta\tau_N, i_N\,\Delta\tau_N]$, and for each $N \in \mathbb{N}$, from (2.4.10)

$$\begin{aligned}\frac{\partial y}{\partial t}\big(t, i_N\,\Delta\tau_N\big) &= k_Y\left(y\big(t,(i_N-1)\,\Delta\tau_N\big) - y\big(t, i_N\,\Delta\tau_N\big)\right) - k_{YA}\big(i_N\,\Delta\tau_N\big)\,y\big(t, i_N\,\Delta\tau_N\big)\\ &= \frac{\widehat{y}_N\,(t,\tau-\Delta\tau_N) - \widehat{y}_N\,(t,\tau)}{\Delta\tau_N} - \widehat{k}_{YA}^{(N)}\,(\tau)\;\widehat{y}_N\,(t,\tau)\;,\end{aligned} \tag{2.4.12}$$

and taking $\Delta\tau_N \to 0$, $N \to \infty$ with τ fixed,

$$\frac{\partial y}{\partial t}(t,\tau) = -\frac{\partial y\,(t,\tau)}{\partial \tau} - k_{YA}\,(\tau)\;y\,(t,\tau)\;. \tag{2.4.13}$$

Also, taking the limit of (2.4.9) as $\Delta\tau_N \to 0$, $N \to \infty$ yields

$$0 = k_{XY}\,X\,(t) - y\,(t,0)\;. \tag{2.4.14}$$

Furthermore, the terms

$$\frac{T_Y}{N}\sum_{i=1}^{N} k_{YA}\big(i\,\frac{T_Y}{N}\big)\,y\big(t, i\,\frac{T_Y}{N}\big)\,, \tag{2.4.15}$$

from the R.H.S. of (2.4.11) are the Riemann sums of $\int_0^{T_Y} k_{YA}\,(\tau)\,y\,(t,\tau)\,\mathrm{d}\tau$. It follows that the limit system of (2.4.3)-(2.4.6) as $N \to \infty$ (under the above assumptions) can be formulated as the system of ordinary and partial differential equations

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 2\,y\,(t,T_Y) - (k_{XY} + k_{XA})\,X\,, \tag{2.4.16}$$

$$\frac{\partial y\,(t,\tau)}{\partial t} = -\frac{\partial y\,(t,\tau)}{\partial \tau} - k_{YA}\,(\tau)\;y\,(t,\tau)\;, \qquad (0 < \tau < T_Y)\;, \tag{2.4.17}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = k_{XA}\,X + \int_0^{T_Y} k_{YA}\,(\tau)\,y\,(t,\tau)\,\mathrm{d}\tau\,, \tag{2.4.18}$$

with boundary condition

$$y\,(t,0) = k_{XY}\,X\,(t)\;. \tag{2.4.19}$$

In particular, equations (2.4.16)-(2.4.18) and (2.4.19) result from (2.4.8),(2.4.10)-(2.4.11) and (2.4.9), respectively (by taking the limit as $N \to \infty$). Figure 2.10 shows a diagram of the age-structured model.

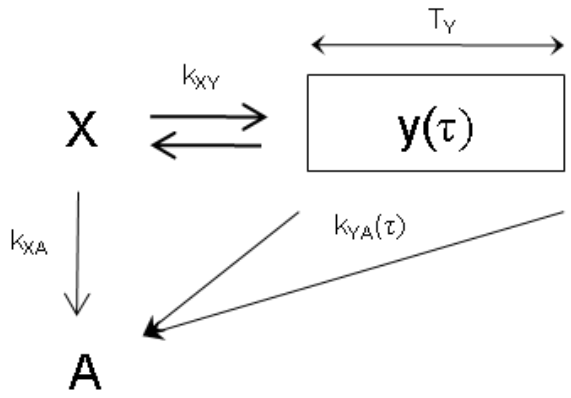

Figure 2.10: A diagram of a three-compartment ODE-PDE model with age-structured Y compartment of length T_Y (residence time of cells in the S+G_2+M phase) and age-dependent loss terms $k_{YA}(\tau)$ from the Y compartment to apoptosis (A compartment). The X compartment corresponds to the G_0/G_1 phase of the cell cycle.

We solve the first order PDE (2.4.17) by using the method of characteristics. The substitution $u = t - \tau$, $v = t$ gives

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial u} + \frac{\partial}{\partial v}, \qquad \frac{\partial}{\partial \tau} = -\frac{\partial}{\partial u}, \tag{2.4.20}$$

and transforms (2.4.17) into the ODE

$$\frac{\partial y}{\partial v} = -k_{YA}\, y\,. \tag{2.4.21}$$

We now assume $k_{YA}(\tau) = k_{YA} =$ constant throughout the following analysis of model (2.4.16)-(2.4.18). The solution can then be formulated as

$$y(t,\tau) = C(u)\, \mathrm{e}^{-k_{YA}\, v} = C(t-\tau)\, \mathrm{e}^{-k_{YA}\, t}\,, \tag{2.4.22}$$

where C is determined by the boundary condition (2.4.19).

2.4.3 Deduction of a three-compartment DDE model

We finally construct a mathematical model which contains only one independent variable, t, for the time evolution of the state variables, but incorporates a parameter for the residence time of cells in the S+G_2+M phase. This parameter describes a delay in the cell

cycle, which stems from the time cells have spent in the phase, and therefore accounts for possible processes during cell cycle progression, which halt the cycle in, for example, the S or G_2 phase if the drug RHPS4 is added.

If we assume that the rates of transition between N subcompartments, A_i, are all equal to λ, as depicted in Figure 2.7, and λ grows proportional to N when we let $N \longrightarrow \infty$, we have $N/\lambda \longrightarrow \tau^*$ for some constant τ^*. We showed in Section 2.4.1 that the transition age of the joint compartment $Y_1+\ldots+Y_N$ is then gamma-distributed with "scale" N and "shape" $1/\lambda$, that is, $Y_1+\ldots+Y_N \sim \Gamma(N, 1/\lambda)$ with probability density function $f(\tau; N, \lambda) = \tau^{N-1}\,\lambda^N\,\mathrm{e}^{-\tau\lambda}/\Gamma(N)$, $\tau > 0$. As, for N sufficiently large, it is possible to estimate $\Gamma(N, 1/\lambda)$ by a normally distributed random variable with mean N/λ and standard deviation N/λ^2, we find that $\Gamma(N, 1/\lambda)$ tends to the normal distribution with mean τ^* and vanishing standard deviation as $N \longrightarrow \infty$. Such a distribution is also called a Dirac delta distribution. All material in the joint compartment $Y_1+\ldots+Y_N$ leaves the compartment with approximately the same age τ^*, when N, the number of subcompartments, is large. We aim to find the corresponding three-compartment model in only one independent variable, time t.

We derive a differential-delay equation (DDE) model by reformulating the three-compartment PDE models (2.4.16)-(2.4.18) from Section 2.4.2. The left-hand-side boundary condition at $\tau = 0$ (2.4.19) of these models can be written as

$$y\,(t, 0) \quad = \quad C\,(t)\,\mathrm{e}^{-k_{YA}\,t}\,, \tag{2.4.23}$$

by using the general solution (2.4.22) of the PDE (2.4.17). By using equations (2.4.22), (2.4.23) and the boundary condition (2.4.19), we express the solution $y\,(t, \tau)$ of the PDE in terms of the state variable $X(t - \tau)$, namely

$$\begin{aligned} y\,(t, \tau) &= C\,(t - \tau)\,\mathrm{e}^{-k_{YA}\,t} \\ &= C\,(t - \tau)\,\mathrm{e}^{-k_{YA}(t-\tau)}\,\mathrm{e}^{-k_{YA}\tau} \\ &= y\,(t - \tau, 0)\,\mathrm{e}^{-k_{YA}\tau} \\ &= k_{XY}\,X\,(t - \tau)\,\mathrm{e}^{-k_{YA}\tau}\,. \end{aligned} \tag{2.4.24}$$

We now transform the PDE into a differential-delay equation by defining

$$Y(t) = \int_0^{T_Y} y(t,\tau)\,\mathrm{d}\tau\,, \tag{2.4.25}$$

and assuming k_{YA} to be constant. It follows from equations (2.4.25), (2.4.17), (2.4.19) and (2.4.24) that

$$\begin{aligned}\frac{\mathrm{d}Y}{\mathrm{d}t} &= \int_0^{T_Y} \frac{\partial y(t,\tau)}{\partial t}\,\mathrm{d}\tau \\ &= -\int_0^{T_Y} \frac{\partial y(t,\tau)}{\partial \tau}\,\mathrm{d}\tau - k_{YA}\int_0^{T_Y} y(t,\tau)\,\mathrm{d}\tau \\ &= y(t,0) - y(t,T_Y) - k_{YA}\,Y(t) \\ &= k_{XY}\,X(t) - k_{XY}\,X(t-T_Y)\,\mathrm{e}^{-k_{YA}\,T_Y} - k_{YA}\,Y(t)\,.\end{aligned} \tag{2.4.26}$$

We can thus rewrite the system (2.4.16)-(2.4.18) as an equivalent system of DDEs, which is given by

$$\frac{\mathrm{d}X}{\mathrm{d}t} = 2\,k_{XY}\,X(t-T_Y)\,\mathrm{e}^{-k_{YA}\,T_Y} - (k_{XY}+k_{XA})\,X(t)\,, \tag{2.4.27}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = k_{XY}\,X(t) - k_{XY}\,X(t-T_Y)\,\mathrm{e}^{-k_{YA}\,T_Y} - k_{YA}\,Y(t)\,, \tag{2.4.28}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = k_{XA}\,X + k_{YA}\,Y\,. \tag{2.4.29}$$

Figure 2.11 shows a diagram of the three-compartment DDE model.

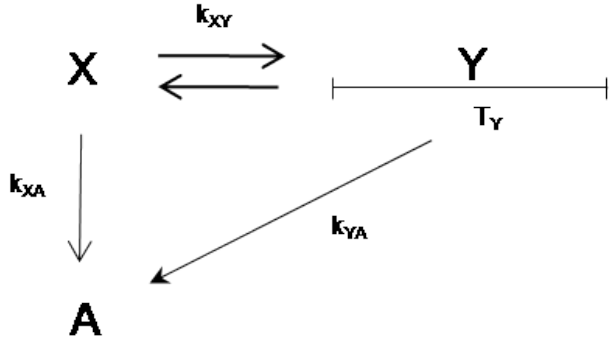

Figure 2.11: A diagram of a three-compartment model with a delay in the Y compartment of constant length T_Y. The model describes senescence/apoptosis (A compartment) from the G_0+G_1 phase (X compartment) and the S+G_2+M phase (Y compartment). All transition rates between the compartments X,Y and A are positive and constant.

As equation (2.4.27) is mathematically independent of Y and A, it decouples from equations (2.4.28) and (2.4.29), and we can analyse equation (2.4.27) first and derive the behaviour of solutions for Y and then A by integrating equation (2.4.28) first, and later (2.4.29), using the integrating factor method for equation (2.4.28).

2.5 Analysis of the three-compartment DDE model

We now want to investigate the solution behaviour of the DDE model (2.4.27)-(2.4.29), in particular, we wish to determine whether oscillations could occur and how the dynamics of the DDE model differ from the ODE models in Section 2.1 and 2.2.

We therefore consider solutions of (2.4.27) of the form $X_\lambda(t) = X_0 \mathrm{e}^{\lambda t}$. Inserting $X_\lambda(t)$ into equation (2.4.27) yields

$$\lambda = \beta \mathrm{e}^{-\tau\lambda} + \alpha\,, \tag{2.5.1}$$

a transcendental equation for $\lambda \in \mathbb{C}$, with $\tau = T_Y > 0$, $\alpha = -k_{XY} - k_{XA} < 0$ and $\beta = 2\,k_{XY}\,\mathrm{e}^{-T_Y k_{YA}} > 0$. We can split the transcendental equation (2.5.1) into two equations corresponding to the real and imaginary parts, with $\lambda = \mu + i\,\omega$, $\mu, \omega \in \mathbb{R}$, being any complex eigenvalue of (2.5.1), that is

$$\mu = \beta \mathrm{e}^{-\tau\mu} \cos(\omega\,\tau) + \alpha\,, \tag{2.5.2}$$

$$\omega = -\beta \mathrm{e}^{-\tau\mu} \sin(\omega\tau)\,. \tag{2.5.3}$$

It is clear from equation (2.5.2) that there is exactly one real eigenvalue $\lambda = \mu$, ($\omega = 0$) for every $\alpha \in \mathbb{R}$, $\beta > 0$ and $\tau > 0$. If $k_{XY} + k_{XA} > 2\,k_{XY} \exp(-T_Y k_{YA})$, then $\lambda = \mu < 0$ and monotone decay is possible. If $k_{XY} + k_{XA} < 2\,k_{XY} \exp(-T_Y k_{YA})$, then $\lambda = \mu > 0$ and it is possible to find monotone growth. (This will be more rigorously established in proposition 2 below.) We want to compare the purely real eigenvalue with the real parts of complex eigenvalues of (2.5.1) in order to find the relationship between exponentially growing and oscillating solutions. Let λ for this purpose denote the purely real eigenvector of (2.5.1) only, and let μ denote the real part of possibly complex eigenvalues $\mu + i\,\omega$ with $\omega > 0$. Note that $(\mu, -\omega)$ is solution of (2.5.2) and (2.5.3) whenever (μ, ω)

is a solution, it hence suffices to consider solutions of (2.5.2) and (2.5.3) with positive ω only.

For $\omega > 0$, equation (2.5.3) yields

$$\frac{\sin(\omega\tau)}{\omega} = -\mathrm{e}^{\tau\mu}\,\beta^{-1} < 0. \tag{2.5.4}$$

Hence, we find that for solutions (μ, ω) of (2.5.2), (2.5.3) we require $\omega\tau \in (2k\pi - \pi, 2k\pi)$, $k \in \mathbb{N}$.

Proposition 1 *An exponentially growing (decaying) solution corresponding to a purely real eigenvalue $\lambda > 0$ ($\lambda < 0$) dominates all oscillatory solutions for large $t > 0$, that is we find $\lambda > \mu$.*

Proof We distinguish between three cases to show that $\lambda > \mu$:

(i) $\omega\tau \in \left(2k\pi - \pi, 2k\pi - \frac{1}{2}\pi\right)$, $k \in \mathbb{N}$:

Dividing (2.5.2) by (2.5.3) yields

$$\frac{\mu - \alpha}{\omega} = -\frac{1}{\tan(\omega\tau)}, \tag{2.5.5}$$

and we have, using equation (2.5.1) and (2.5.5),

$$\begin{aligned} \lambda - \mu &= \beta\,\mathrm{e}^{-\tau\lambda} + \alpha - \mu & (2.5.6)\\ &= \beta\,\mathrm{e}^{-\tau\lambda} + \frac{\omega}{\tan(\omega\tau)}, & (2.5.7) \end{aligned}$$

that is, $\lambda > \mu$, as $\tan(\omega\tau) > 0$ for $\omega\tau \in \left(2k\pi - \pi, 2k\pi - \frac{1}{2}\pi\right)$ for all $k \in \mathbb{N}$.

(ii) $\omega\tau = 2k\pi - \frac{1}{2}\pi$, $k \in \mathbb{N}$:

Equation (2.5.2) yields $\mu = \alpha$. It follows from equation (2.5.1) that

$$\lambda - \mu = \beta\,\mathrm{e}^{-\tau\lambda}. \tag{2.5.8}$$

Hence, $\lambda > \mu$ and equation (2.5.3) shows that we have at most one positive solution for ω, namely $\omega = \beta\,\mathrm{e}^{-\tau\alpha}$ if $k = \beta\tau\mathrm{e}^{-\tau\alpha}/2\pi + 1/4 \in \mathbb{N}$.

(iii) $\omega\tau \in \left(2k\pi - \frac{1}{2}\pi, 2k\pi\right)$, $k \in \mathbb{N}$:

Employing the fact that in this case $0 < \cos(\omega\tau) < 1$, we aim to show that $\lambda > \mu$ by a brief indirect proof. Let us assume that $\lambda < \mu$. From equations (2.5.1) and (2.5.2) we derive

$$\lambda - \alpha = \beta\, \mathrm{e}^{-\tau\lambda} > \beta\, \mathrm{e}^{-\tau\mu} = \frac{\mu - \alpha}{\cos(\omega\tau)} > \frac{\lambda - \alpha}{\cos(\omega\tau)} > \lambda - \alpha\,. \tag{2.5.9}$$

This is a contradiction and hence $\lambda \geq \mu$. Let us now assume that $\lambda = \mu$. We again use equations (2.5.1) and (2.5.2) to derive

$$\beta\, \mathrm{e}^{-\tau\lambda} + \alpha = \lambda = \mu = \beta\, \mathrm{e}^{-\tau\mu} \cos(\omega\tau) + \alpha\,, \tag{2.5.10}$$

and find $\cos(\omega\tau) = 1$. This contradicts $\omega\tau \in \left(2k\pi - \frac{1}{2}\pi, 2k\pi\right)$. □

Consequently, we expect solutions of (2.4.27) in general not to oscillate around 0 for large enough t. In the following, we analyse the general solution behaviour of the three-compartment model further.

Proposition 2 *The sign of $\alpha + \beta$ determines the general growth behaviour of solutions $X(t)$, that is we obtain exponential growth for $\alpha + \beta > 0$ and exponential decay for $\alpha + \beta < 0$.*

Proof Consider

$$z(\lambda) = \alpha - \lambda + \beta\, \mathrm{e}^{-\tau\lambda}\,, \tag{2.5.11}$$

then

$$z'(\lambda) = -1 - \beta\, \tau\, \mathrm{e}^{-\tau\lambda} < -1\,. \tag{2.5.12}$$

Hence $z(\lambda)$ is a strictly monotone decreasing function of λ with $z \to +\infty$ as $\lambda \to -\infty$, $z \to -\infty$ as $\lambda \to +\infty$ and $z(0) = \alpha + \beta$. So if $\alpha + \beta > 0$, the root $z(\lambda) = 0$ satisfies $\lambda > 0$, and if $\alpha + \beta < 0$, then root $z(\lambda) = 0$ corresponds to $\lambda < 0$. □

Note that for positive λ, equation (2.5.1) yields $\lambda < \alpha + \beta$, providing an upper bound on the growth rate of exponentially growing solutions. Furthermore, a steady state solution for $X(t)$ exists if and only if $\alpha + \beta = 0$.

Proposition 3 *There are an infinite number of solutions* (μ, ω) *for equations* (2.5.2) *and* (2.5.3).

Proof Solving equation (2.5.3) for μ and inserting in equation (2.5.2) yields (after taking exponentials on both sides and defining $y = \omega\tau$)

$$\frac{e^{\alpha\tau}}{\beta\tau}\, y = -\sin(y) \exp\left(\frac{y}{\tan(y)}\right), \tag{2.5.13}$$

an equation which is independent of μ. We show that equation (2.5.13) has exactly one solution $y \in (2k\pi - \pi, 2k\pi)$ for each $k \in \mathbb{N}$. It suffices for this purpose to show that the function

$$f(y) = -\sin(y) \exp\left(\frac{y}{\tan(y)}\right), \tag{2.5.14}$$

on the right-hand side of equation (2.5.13) is strictly monotonic decreasing in each interval $(2k\pi - \pi, 2k\pi)$, $k \in \mathbb{N}$, with $f(y) \to +\infty$ as $y \searrow 2k\pi - \pi$ and $f(y) \to 0$ as $y \nearrow 2k\pi$, as the left-hand side of (2.5.13) is a positive, strictly monotonic increasing function, $g(y) = e^{\alpha\tau}\, y/(\beta\,\tau)$, in y for $y > 0$ (see Figure 2.12).

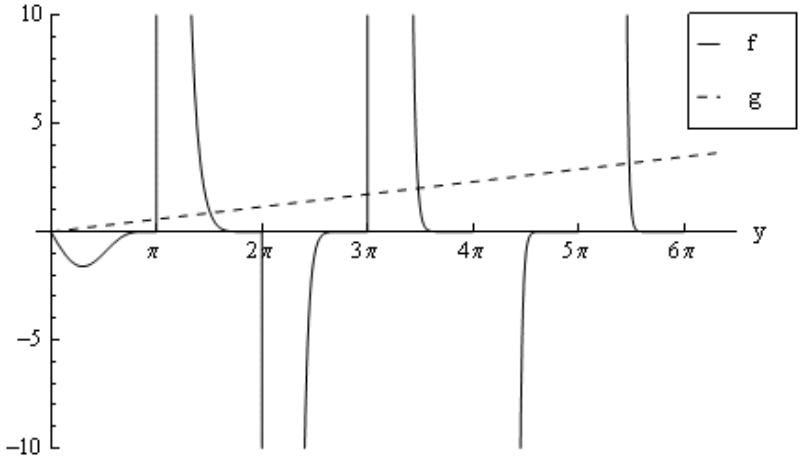

Figure 2.12: Plot of the function $f(y)$ and the straight line $g(y)$ against y for $\tau = 1$, $\alpha = -1$, $\beta = 2$. There is exactly one intersection between f and g in each interval $y \in (2k\pi - \pi, 2k\pi)$, $k \in \mathbb{N}$.

The derivative of f is

$$f'(y) = \left(\frac{y}{\sin(y)} - 2\cos(y)\right) \exp\left(\frac{y}{\tan(y)}\right). \tag{2.5.15}$$

For $y \in \left(2k\pi - \pi, 2k\pi - \frac{1}{2}\pi\right]$, we find $-1 \leq \sin(y) < 0$ and $-1 < \cos(y) \leq 0$, hence

$$\frac{y}{\sin(y)} - 2\cos(y) \leq -y + 2 < -\pi + 2 < 0\,, \tag{2.5.16}$$

and thus $f'(y) < 0$. For $y \in \left(2k\pi - \frac{1}{2}\pi, 2k\pi\right)$, we find $\sin(y) < 0$ and $\cos(y) > 0$, and $f'(y) < 0$ follows immediately. Moreover, $f(y)$ can be analytically continued to $y_k = 2k\pi - \frac{\pi}{2}$ with $f(y_k) = 1$. Thus, f is a strictly monotonic decreasing function in $(2k\pi - \pi, 2k\pi)$ for all $k \in \mathbb{N}$.

We have shown that there is exactly one solution of (2.5.13) for $y \in (2k\pi - \pi, 2k\pi)$ for each $k \in \mathbb{N}$. Consequently, there are infinitely many solutions (μ, ω) for (2.5.2) and (2.5.3), with $\omega = y/\tau$ and $\mu = \ln(-(\beta\,\tau\sin y)/y)/\tau$. □

It is trivial to show, however, that there are only a finite number of exponentially growing oscillatory solutions of the DDE (2.4.27). We infer from (2.5.3) that $\mu > 0$ is equivalent to $y < -\beta\,\tau\sin y$, hence exponentially growing oscillations require y to be sufficiently small, in particular $y = \omega\,\tau < \beta\,\tau$. As there is only one solution in each interval $(2k\pi - \pi, 2k\pi)$, $k \in \mathbb{N}$, the number of solutions (μ, ω) for (2.5.2) and (2.5.3) with $\mu > 0$ is bounded above.

An upper bound on the number N_μ^+ of complex eigenvalues with positive real part μ is thus determined by the upper bound $\beta\,\tau$ on $y = \omega\,\tau$. We have at most $\lceil(\beta\,\tau - \pi)/2\,\pi\rceil$ such complex eigenvalues with $\omega > 0$, that is, $0 \leq N_\mu^+ \leq \lceil(\beta\,\tau - \pi)/\pi\rceil$, where $\lceil x\rceil$ is defined to be the smallest integer not less than x. If $\beta\,\tau \leq \pi$, the solution $X(t)$ does not display any exponentially growing oscillatory behaviour. The function $h(\tau = T_Y) = \beta\,T_Y = 2\,k_{XY}\,T_Y\exp(-k_{YA}\,T_Y)$, illustrated in Figure 2.13, has a unique extrema (maximum) in $\widehat{T}_Y = 1/k_{YA}$ for $k_{YA} > 0$ with $h(\widehat{T}_Y) = 2\,k_{XY}/(e\,k_{YA}) > 0$, where $h(T_Y) \to -\infty$ as $T_Y \to -\infty$ and $h(T_Y) \to 0$ as $T_Y \to +\infty$. Exponentially growing oscillations hence occur only for T_Y in an interval around $\widehat{T}_Y$ where $h(T_Y) > \pi$. Thus, if $2\,k_{XY}/(e\,k_{YA}) < \pi$, exponentially growing oscillations do not occur whatever choice of T_Y.

The period of oscillation of a solution $X(t) = X_0\,\mathrm{e}^{(\mu + i\,\omega)t}$ is $P_X = \frac{2\pi}{\omega}$, the corresponding amplitude is $A_X(t) = X_0\,\mathrm{e}^{\mu t}$. We can give an upper bound for P_X by using the lower bound π/τ on $|\omega|$, that is

$$0 \leq P_X < 2\,\tau\,. \tag{2.5.17}$$

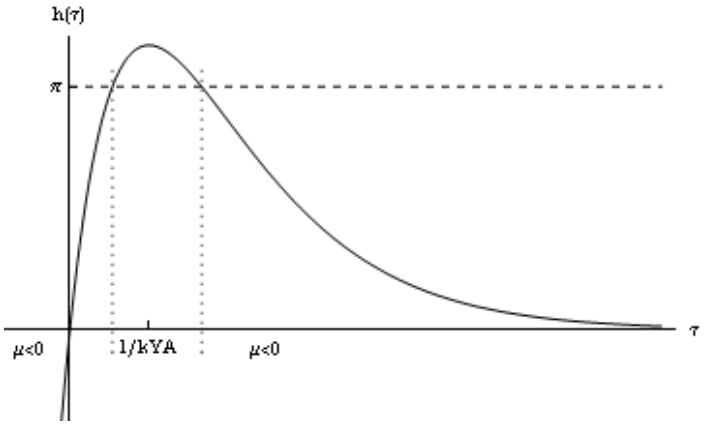

Figure 2.13: Plot of the function $h(\tau)$ against τ for $h(1/k_{YA}) = 2k_{XY}/(e\,k_{YA}) > \pi$. For the delay term T_Y with $h(T_Y) < \pi$, all complex eigenvalues of (2.5.1) have negative real parts.

Equation (2.5.17) shows that $P_X \to 0$ for small delays $\tau \ll 1$. Hence, the period P_X of the oscillation is likely to become smaller with decreasing τ.

Altogether, we have shown that the DDE model has exactly one purely real eigenvalue, which is larger than the real parts of any complex eigenvalues, that is, oscillatory solutions do not in general dominate the growth behaviour of the system and we do not expect negative values in the solutions. The general growth behaviour of the system is determined by the sign of $\alpha + \beta = k_{XY}\,(2\exp(-T_Y\,k_{YA}) - 1) - k_{XA}$, yielding exponential growth if $\alpha + \beta > 0$, decay if $\alpha + \beta < 0$ or a steady state if $\alpha + \beta = 0$, for large values of time t. Furthermore, we have shown that there are an infinite number of eigenvalues, only a finite number of which have a positive real part, and derived an upper bound for this number, N_μ^+. Thus, one can construct solutions that do not show persistent oscillations by choosing appropriate parameter values and initial conditions. This can be achieved in particular by varying the delay term T_Y, which also determines the period of underlying oscillatory modes, where we find decreasing values of P_X with smaller values of T_Y.

Figures 2.14 and 2.15 show different dynamics of the DDE model (2.4.27)-(2.4.29) with varying T_Y, where we use the DDE solver "dde23" in Matlab for the numerical computations and assume that the initial conditions are constant, that is, $(X(t), Y(t), A(t)) = (X_0, Y_0, Z_0)$ for $t \in [-T_Y, 0]$ with $X_0 = 200,000$ and $Y_0 = A_0 = 0$. The solution behaviour of $Y(t)$ and $A(t)$ can be derived as in Section 2.2 for the five-compartment ODE model,

and we find that $Y(t)$ is asymptotic to a constant and $A(t)$ asymptotes to linear growth in time for steady states of $X(t)$. If $X(t)$ grows exponentially, $Y(t)$ and $A(t)$ grow exponentially as well, whereas, if $X(t)$ decays exponentially, $Y(t)$ tends to zero and $A(t)$ tends to a constant for large t. Figure 2.14 illustrates the three different types of growth behaviour according to the sign of $\alpha + \beta$, where T_Y is chosen such that $\beta\tau < \pi$ ($N_\mu^+ = 0$) for the case of exponential growth, hence no visible oscillations appear. Figure 2.15, in contrast, shows exponential growth with $\beta\tau > \pi$, that is, where persistent oscillatory behaviour occurs ($N_\mu^+ \geq 1$), and the period of oscillation varies with the choice of T_Y.

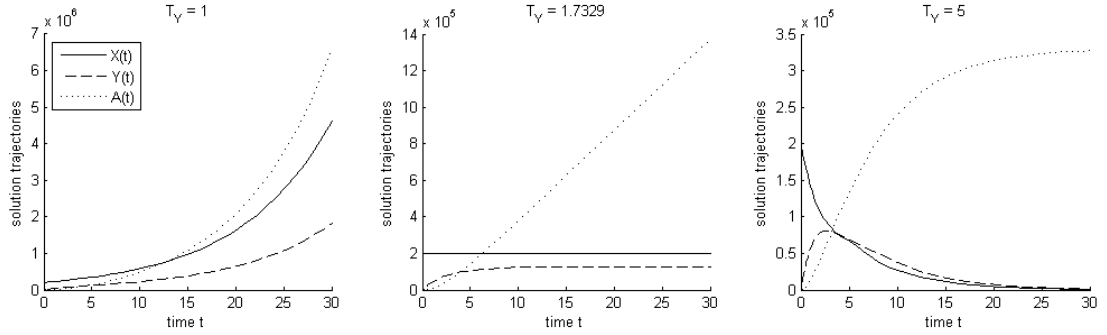

Figure 2.14: The three plots show the different dynamic behaviour of the solution trajectories of model (2.4.27)-(2.4.29) for $k_{XY} = 0.5$, $k_{XA} = 0$ and $k_{YA} = 0.4$. The delay term T_Y takes the values 1, $2.5 \ln 2$ and 5, that is $\alpha + \beta > 0$, $\alpha + \beta = 0$ and $\alpha + \beta < 0$, respectively. As $N_\mu^+ = 0$ for $T_Y = 1$, there are no visible oscillations in the exponentially growing curves.

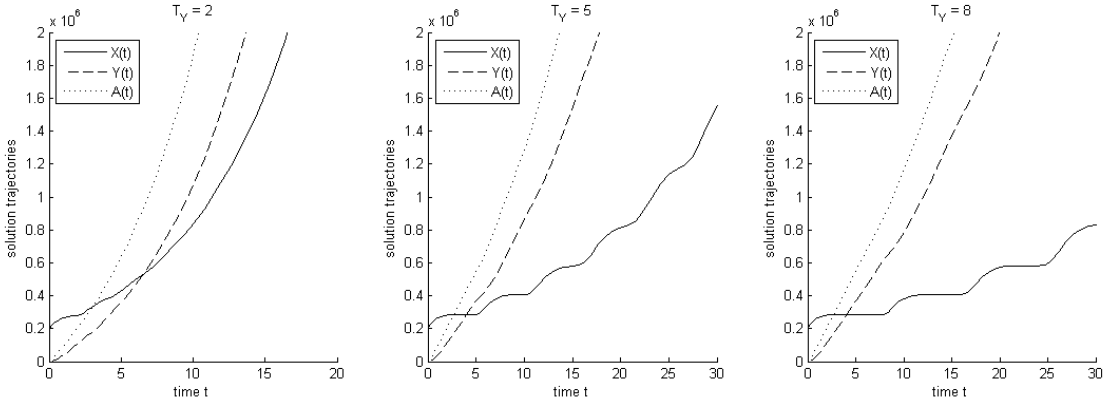

Figure 2.15: The three plots show the different dynamic behaviour of the solution trajectories of model (2.4.27)-(2.4.29) for $k_{XY} = 1$, $k_{XA} = 0.4$ and $k_{YA} = 0$. The delay term T_Y takes the values 2, 5 and 8, and we have exponential growth with visible oscillatory behaviour, where $P_X \leq 4$, $P_X \leq 10$ and $P_X \leq 16$, respectively. Oscillations decay exponentially for $T_Y = 2$ ($N_\mu^+ = 0$), but grow exponentially for $T_Y = 5$ and $T_Y = 8$ ($N_\mu^+ \geq 1$).

2.6 Conclusions

We have developed and analysed various differential equation models describing the dynamics of the cell cycle including the phases of senescence and apoptosis. The developed three-compartment ODE model is simple and analytic solutions can be derived explicitly, but it does not distinguish between the dynamics in the S, G_2 and M phases, which are potentially important when exposing the culture to a cell-cycle specific drug. We therefore developed a five-compartment ODE model that distinguishes between the phases of the cell cycle that can be detected by DNA content analysis and senescence markers, and which allows for apoptosis from each phase of the cell cycle and is still simple enough to analyse its dynamic behaviour. A refined seven-compartment ODE model distinguishes between the cell-cycle phases G_0, G_1, S, G_2 and M, and includes the dynamics of quiescent cells (G_0 phase), which may return to the cell cycle after a longer period of growth arrest. However, it may be difficult to extract valuable information from this model as one typically observes only joint compartments of the model and the number of observations might be too small to infer the rate parameters of the system.

We also developed a PDE and a DDE model from the three-compartment ODE model presented earlier. The PDE model requires age-structured data for model comparison to data, which are difficult to obtain experimentally. The DDE model, in contrast, accounts for a variety of dynamic behaviour including oscillatory behaviour for exponentially growing solutions, requires only knowledge of the time-course behaviour of cells and contains a small number of parameters to estimate. The three-compartment DDE model therefore seems to be more powerful with respect to the description of the real dynamics. The only additional knowledge which is required for numerical simulations of the model is an estimation of the initial evolution of the state variables for a time of length T_Y before the initial time point of observation. The initial conditions, however, do not affect the general growth behaviour and their effect on the solution dynamics generally vanishes over time.

The developed differential equation models are potentially good descriptions of cell cycle dynamics and provide information on the specific cell-cycle behaviour when compared

to experimental data. One should, however, choose a model which is not only sufficiently detailed to describe the dynamic behaviour of the cell cycle processes observed, but also sufficiently simple in its structure to allow for a reliable and sufficiently accurate parameter inference, that is, to ensure that parameters are identifiable from the observations obtained. We discuss the problem of parameter and model identifiability in the following Chapter 3.

CHAPTER 3

Identifiability analysis

3.1 Introduction and definitions

Mathematical models can be constructed using compartmental systems, which are made up of a finite number of subunits. These compartments are assumed to consist of homogeneous and well-mixed material and are linked by material flowing between them.

Models contain parameters whose identification from experimental data is of particular interest. In general, however, it is possible that more than one parameter set can fit the experimental data equally well, even under the ideal assumption that for all observables error-free and continuous data are available. Before we define the concept of structural identifiability (first introduced by Bellman and Astrom [19]) that deals with this problem, we introduce some notation for compartmental models that we use for our definitions.

Let $\mathbf{v}$ be the vector function of state variables with dimension m, whose components represent the variations with time t of the individual compartments. Furthermore, let $\mathbf{v}_0$ be the respective vector of initial conditions and $\mathbf{A} = \mathbf{A}(t, \mathbf{p})$ the $m \times m$ matrix of transitions rates, which depend on the L-dimensional vector $\mathbf{p}$ of model parameters and possibly t. The state variables $\mathbf{v}$ depend on the model parameters $\mathbf{p}$, and when components of the vector of initial conditions $\mathbf{v}_0$ are not known, $\mathbf{v}_0$ also depends on $\mathbf{p}$.

A linear compartmental model can then be written as

$$\frac{\mathrm{d}\mathbf{v}(t, \mathbf{p})}{\mathrm{d}t} = \mathbf{A}(t, \mathbf{p})\ \mathbf{v}(t, \mathbf{p})\ , \qquad \mathbf{v}(0, \mathbf{p}) = \mathbf{v}_0(\mathbf{p})\,. \tag{3.1.1}$$

The vector function $\mathbf{y}$ of observations with dimension M is given by

$$\mathbf{y}\left(t, \mathbf{p}\right) = \mathbf{B}\,\mathbf{v}\left(t, \mathbf{p}\right) , \tag{3.1.2}$$

where $\mathbf{B}$ is the constant $M \times m$ observation matrix of rank M. The model (3.1.1) is referred to as a time-invariant linear compartmental model for $\mathbf{A}\left(t, \mathbf{p}\right) = \mathbf{A}(\mathbf{p})$ being independent of time t.

The problem of structural (*a priori*) identifiability consists of determining for which of the unknown parameters $\mathbf{p}$ of system (3.1.1)-(3.1.2) there is a unique solution (*globally or uniquely identifiable*), a finite number (≥ 2) of solutions (*locally or nonuniquely identifiable*), or an infinite number of solutions (*nonidentifiable*), given that the observation function $\mathbf{y}$ is known. The study of structural model identifiability aims to determine whether the unknown parameters of the model can be estimated from the experimental setting. The model is uniquely identifiable if all its parameters $\mathbf{p}$ are uniquely identifiable, it is locally identifiable if all its parameters are identifiable (but at least one parameter nonuniquely), and nonidentifiable if at least one of its parameters is nonidentifiable. Reviews of identifiability analysis can be found in Cobelli and Distefano [30] and Walter and Pronzato [144].

Not all models are suitable for parameter estimation, in particular, there are models which are not uniquely (or locally) identifiable. We give an example of a simple, nonidentifiable two-compartment ODE model below (see Figure 3.1), that is, a model which has the same dynamical behaviour for an infinite number of parameter values.

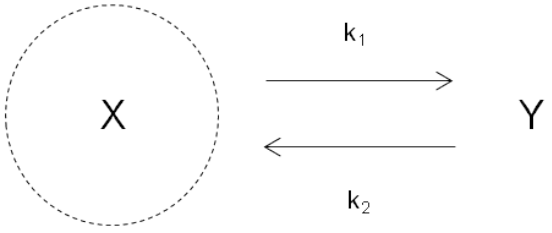

Figure 3.1: A simple nonidentifiable two-compartment model with only one observable state X. The compartments X and Y are connected by two constant flows with rates k_1 and k_2.

The rate of change of the according state variables X and Y can be described by a system

of ODEs, namely

$$\frac{\mathrm{d}X}{\mathrm{d}t} = -k_1\,X + k_2\,Y\,, \qquad \frac{\mathrm{d}Y}{\mathrm{d}t} = k_1\,X - k_2\,Y\,, \tag{3.1.3}$$

with the initial conditions $X(0) = x_0$, $Y(0) = y_0$ and rate parameters k_1, $k_2 > 0$. We assume that rate parameters as well as initial conditions are unknown parameters, and that we observe $X(t)$ only. We solve this system of ODEs explicitly by computing the eigenvalues of the corresponding coefficient matrix. The characteristic polynomial is $p(\lambda) = \lambda\,(\lambda + k_1 + k_2)$, the eigenvalues are hence $\lambda_1 = 0$ and $\lambda_2 = -k_1 - k_2$. We derive the general solution from the eigenvalues and according eigenvectors, $v_1 = (k_2, k_1)$, $v_2 = (1, -1)$, respectively, that is

$$X(t) = \frac{1}{c_0}\left(c_1\,\mathrm{e}^{-c_0 t} + c_2\right)\,, \qquad Y(t) = \frac{1}{c_0}\left(-c_1\,\mathrm{e}^{-c_0 t} + c_3\right)\,, \tag{3.1.4}$$

with $c_0 = k_1 + k_2$, $c_1 = k_1\,x_0 - k_2\,y_0$, $c_2 = k_2\,(x_0 + y_0)$ and $c_3 = k_1\,(x_0 + y_0)$. Observing X gives c_0, c_1, c_2, but not c_3. The question is, can k_1, k_2, x_0, y_0 be obtained from c_0, c_1, c_2? Since the number of equations is smaller than the number of unknowns, the problem is underdetermined and the solution is not unique. It follows that $x_0 = (c_1 + c_2)/c_0$, and $k_1 = c_0 - k_2$ and $y_0 = (c_0\,c_2 - k_2\,(c_1 + c_2))/(k_2\,c_0)$ are functions of the parameter k_2 that can be chosen arbitrarily. The unknown parameters are hence not identifiable from the observations, that is the model is non-identifiable.

Similarly, observing $X + Y$ only gives one piece of information, $X + Y = (c_2 + c_3)/c_0$, so we cannot determine four data k_1, k_2, x_0, y_0. However, observing both states, $X(t)$ and $Y(t)$ yields an additional condition on the parameter values and fixes the free parameter with $k_2 = k_1\,c_2/c_3$, resulting in a model which is uniquely identifiable.

Two methods for the identifiability analysis of time-invariant linear compartmental models are given in Section 3.2 and 3.3, which we apply to the five-compartment model from Section 2.2 and later, in Section 5.4, to a seven-compartment model (presented in Section 5.1), respectively. The concept of practical identifiability analysis is introduced in Section 3.4, and can be employed to evaluate the accuracy of parameter estimation results, when fitting models to data that are noisy and contain only a finite number of measurements.

3.2 Taylor series method and application to the five-compartment ODE model

It is possible for models to contain parameters which cannot be identified from the observations independent of the noise in the data. It is useful to check models for *a priori* identifiability, and we now aim to show that the five-compartment model of Section 2.2 is globally identifiable, that is, two different parameter sets cannot describe the same observation function of the model.

We write model (2.2.1)-(2.2.5) in the form (3.1.1)-(3.1.2), that is, as a linear system of ODEs with state vector $\mathbf{v} = (X, Y, Z, \Sigma, A)^T \in \mathbb{R}^5$, parameter vector

$$\mathbf{p} = (k_{XY}, k_{YZ}, k_{ZX}, k_{X\Sigma}, k_{\Sigma A}, k_{XA}, k_{YA}, k_{ZA}) \in \mathbb{R}^8\,, \tag{3.2.1}$$

transition matrix

$$\mathbf{A}(\mathbf{p}) = \begin{pmatrix} -k_{XY} - k_{X\Sigma} - k_{XA} & 0 & 2\,k_{ZX} & 0 & 0 \\ k_{XY} & -k_{YZ} - k_{YA} & 0 & 0 & 0 \\ 0 & k_{YZ} & -k_{ZX} - k_{ZA} & 0 & 0 \\ k_{X\Sigma} & 0 & 0 & -k_{\Sigma A} & 0 \\ k_{XA} & k_{YA} & k_{ZA} & k_{\Sigma A} & 0 \end{pmatrix} \in \mathbb{R}^{5\times 5}\,, \tag{3.2.2}$$

and initial conditions $\mathbf{v}_0 = (X_0, Y_0, Z_0, \Sigma_0, A_0)$, which are arbitrary, possibly unknown data with $\mathbf{v}_0 \geq 0$. We observe all states, hence we have $\mathbf{B} = \mathbf{id}_5$ and the observation function is $\mathbf{y} \equiv \mathbf{v}$.

We aim to show that the equality $\mathbf{v}(t, \mathbf{p}) = \mathbf{v}(t, \bar{\mathbf{p}})\ \forall\, t \in [0, T]$ for two rate parameter vectors $\mathbf{p}$ and $\bar{\mathbf{p}}$ implies $\mathbf{p} = \bar{\mathbf{p}}$ irrespective our choice of $\mathbf{p}$, $\bar{\mathbf{p}}$. We can use the Taylor series approach [103] assuming $\mathbf{v}$ is analytic, that is, $\mathbf{v}$ has infinitely many derivatives with respect to t. The function $\mathbf{y}(t, \mathbf{p})$ is then uniquely described by the coefficients $\mathbf{y}^{(k)}(0, \mathbf{p})$,

$k = 0, \ldots, \infty$, of its power series expansion around $t = 0$, with

$$\mathbf{y}^{(0)}(0, \mathbf{p}) = \mathbf{B}(\mathbf{p})\,\mathbf{v}_0\,, \tag{3.2.3}$$

$$\mathbf{y}^{(k)}(0, \mathbf{p}) = \mathbf{B}(\mathbf{p})\,(\mathbf{A}(\mathbf{p}))^k\,\mathbf{v}_0\,, \qquad k \geq 1\,, \tag{3.2.4}$$

where $\mathbf{y}^{(k)}(0, \mathbf{p})$ denotes the kth derivative with respect to time. The model is uniquely identifiable at $\mathbf{p}$ if the implication

$$\mathbf{y}^{(k)}(0, \mathbf{p}) = \mathbf{y}^{(k)}(0, \bar{\mathbf{p}}) \quad \forall\, k \geq 0 \quad \Longrightarrow\ \mathbf{p} = \bar{\mathbf{p}}\,, \tag{3.2.5}$$

holds, and the model is structurally globally identifiable if it is uniquely identifiable at $\mathbf{p}$, for generic $\mathbf{p}$.

Considering the initial states $\mathbf{v}_0$ as unknown parameter values, we can immediately see from equation (3.2.3) that $\mathbf{v}_0$ is uniquely identifiable from the observation function $\mathbf{y}$ in the case of $\mathbf{B} = \mathrm{id}_5$.

In order to show that (3.2.5) is satisfied, we solve the respective equations for $\mathbf{p}$ iteratively using formulas (3.2.3) and (3.2.4) for $k = 0, 1, 2, 3, \ldots$, and employing the previous results for every subsequent equation. A MATHEMATICA code can be found in Appendix A.1 and shows each step in the algorithm, yielding $\mathbf{p} = \bar{\mathbf{p}}$ already after comparison of the first and second derivatives of $\mathbf{y}$ only.

The Taylor series method is a powerful method, which enabled us to show that the five-compartment model is globally identifiable. Assuming the model describes the biological behaviour sufficiently well, we thus know that every observed dynamic behaviour (continuous and noise-free) can be uniquely allocated to one parameter vector.

3.3 Transfer function method

Another method to check a time-invariant linear compartmental model for identifiability can be found in Cobelli and Distefano [30] and Jacquez [62], and will be briefly outlined in this Section. An example of how this method can be applied is given later in Section

5.4, where we show a seven-compartment ODE model to be structurally identifiable.

Let F be the Laplace transform of f, that is

$$F(s) = \mathcal{L}[f(t)] = \int_0^\infty e^{-st} f(t)\, dt\,, \quad s \in \mathbb{C}\,, \qquad (t \geq 0)\,, \tag{3.3.1}$$

then $\mathcal{L}$ is linear and bijective and we have

$$\mathcal{L}\left[f'(t)\right] = s\,F(s) - f(0)\,. \tag{3.3.2}$$

Taking Laplace transforms of (3.1.1) and (3.1.2) yields $s\,\mathbf{V} - \mathbf{v}_0 = \mathbf{A}\mathbf{V}$ and $\mathbf{Y} = \mathbf{B}\mathbf{V}$, where $\mathbf{V}$ and $\mathbf{Y}$ are the Laplace transforms of $\mathbf{v}$ and $\mathbf{y}$, respectively. Thus we find

$$\mathbf{Y} = \mathbf{B}\,(s\mathbf{I} - \mathbf{A})^{-1}\,\mathbf{v}_0\,. \tag{3.3.3}$$

It can be shown that $\mathbf{Y}(s) = (\mathbf{Y}_1(s), \ldots, \mathbf{Y}_M(s))^T$ is a vector of M rational polynomials of the form

$$\mathbf{Y}_i(s) = \frac{\phi_{m+1,i}\, s^{m-1} + \phi_{m+2,i}\, s^{m-2} \ldots + \phi_{2m,i}}{s^m + \phi_{1,i}\, s^{m-1} + \ldots + \phi_{m,i}}\,, \tag{3.3.4}$$

with parameters $\phi_{h,i}$ which give sets of equations of the form

$$\begin{aligned}
&\phi_{1,i} = \sum_j \alpha_j^{(1,i)}\, \mathbf{p}_j\,, \qquad \phi_{2,i} = \sum_{j\leq k} \alpha_{j,k}^{(2,i)}\, \mathbf{p}_j\, \mathbf{p}_k\,, \qquad \phi_{3,i} = \sum_{j\leq k\leq l} \alpha_{j,k,l}^{(3,i)}\, \mathbf{p}_j\, \mathbf{p}_k\, \mathbf{p}_l\,, \ \cdots\,,\\
&\phi_{m+1,i} = \sum_{r=1}^{m} \alpha_r^{(m+1,i)}\, (\mathbf{v}_0)_r\,, \qquad \phi_{m+2,i} = \sum_j \sum_{r=1}^{m} \alpha_{r,j}^{(m+2,i)}\, (\mathbf{v}_0)_r\, \mathbf{p}_j\,,\\
&\phi_{m+3,i} = \sum_{j\leq k} \sum_{r=1}^{m} \alpha_{r,j,k}^{(m+3,i)}\, (\mathbf{v}_0)_r\, \mathbf{p}_j\, \mathbf{p}_k\,, \ \cdots\,,
\end{aligned} \tag{3.3.5}$$

where $\alpha_j^{(1,i)}$, $\alpha_{j,k}^{(2,i)}$, $\ldots$, and $\alpha_r^{(m+1,i)}$, $\alpha_{r,j}^{(m+2,i)}$, $\ldots$ are real constants and $(\mathbf{v}_0)_r$ denotes the rth component of the vector $\mathbf{v}_0$. The formulations (3.3.4) and (3.3.5) stem from (3.3.3) by the formula of the inverse of an $m \times m$ matrix $\mathbf{C}$, that is $\mathbf{C}^{-1} = \left((-1)^{i+j} \det\left(\mathbf{C}_{j,i}\right)\right)_{i,j=1}^{n} / \det \mathbf{C}$, where $\mathbf{C}_{j,i}$ denotes the matrix that results from $\mathbf{C}$ by removing the jth row and the ith column and $\det(\cdot)$ denotes the standard determinant of a matrix.

The quantities $\phi_{h,i}$ are called *observational parameters* and are uniquely determined by $\mathbf{Y}$ (cancelling common factors in the numerator and denominator if necessary). The observational parameters are therefore identifiable from the observations in the experiment. $\mathbf{Y}(s)$ is called a *transfer function* as it can be identified from the experimental results, and the set of equations (3.3.5) is called *exhaustive summary* of the model. The identifiability problem consists then of determining which of the unknown parameters $\mathbf{p}$ (and possibly $\mathbf{v}_0$) are uniquely determined by the exhaustive summary (3.3.5).

The seven-compartment ODE model introduced in Chapter 5, whose number of observable quantities is lower than its number of state variables, can be shown to be identifiable using the transfer function method, but using the Taylor series method yielded equations too complex to be analysed by MATHEMATICA. However, neither of the methods in Section 3.2 and in this Section enabled us to obtain results when checking the seven-compartment ODE model (2.3.1)-(2.3.7) in Section 2.3 for identifiability. Not every method guarantees results when we check a particular model for identifiability and an appropriate method needs to be chosen.

To address this problem, Audoly et al. [11] and Bellu et al. [20] have provided computer algebra tools to test linear and nonlinear compartmental models, respectively, for structural identifiability, determining whether a model is either uniquely, locally or nonidentifiable. The algorithms use the Gröbner basis and are implemented in the symbolic language REDUCE, a proprietary computer algebra program for which the University of Nottingham does not have a license. According to the authors Bellu et al. [20], however, an on-line version of the program translated in C++ may be available soon.

3.4 Practical identifiability analysis

When estimating unknown parameters, available data are often sparse and noisy. We therefore should take account of the uncertainty in the estimated parameter values $\hat{\mathbf{p}}$, which arise from the noise in the data. Practical or *a posteriori* identifiability is concerned with how well model parameters can be estimated considering the measurement noise, a summary of which can be found in Ashyraliyev et al. [9].

Let us denote experimental measurements by x_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, M$, when the occupancies of states $\mathbf{y}_j$ at time points t_i are observed. In the following, we assume that the model (3.1.1) is a valid representation of the biological system under study and $\mathbf{p}^*$ is the true parameter vector representing reality. The least-squares estimate, $f(\mathbf{p}) = \sum_{i=1}^{n} \sum_{j=1}^{M} \mathbf{e}_{i,j}^2(\mathbf{p})$, with $\mathbf{e}_{i,j}(\mathbf{p}) = \mathbf{y}_j(t_i) - x_{ij}$ is a measure (Euclidean norm) for the discrepancy between model values and experimental data. Under the assumption that the errors in the measurements x_{ij} are independent from each other and normally distributed with zero mean and variance σ^2, it is well-known that minimising the sum of squares $f(\mathbf{p})$ is equivalent to maximising the likelihood of our model given the data x_{ij}. Besides calculating $\hat{\mathbf{p}}$ by minimising $f(\mathbf{p})$ (see Section 5.5 for optimisation techniques), we characterise the accuracy of this estimate. The following steps summarize a more detailed description of the statistical analysis in Bates and Watts [17] and Aster et al. [10].

By an asymptotic result [122], the approximate sampling distribution of $\hat{\mathbf{p}}$ under the model, assuming the true parameter value $\mathbf{p}^*$, is

$$\hat{\mathbf{p}} \sim \mathcal{N}_L(\mathbf{p}^*, \hat{\mathbf{\Sigma}}), \tag{3.4.1}$$

where $\mathcal{N}_L(\cdot, \cdot)$ denotes the L-dimensional multivariate normal distribution and $\hat{\mathbf{\Sigma}} = \mathbf{Cov}(\hat{\mathbf{p}})$ denotes the covariance matrix of the least-squares parameter estimate. The covariance matrix of $\hat{\mathbf{p}}$ indicates whether parameters are practically identifiable, that is, whether we can accurately infer values for the parameters. If we assume that the model residuals $\mathbf{e}_{i,j}(\mathbf{p})$, $i = 1, \ldots, n$, $j = 1, \ldots, M$, are approximately linear in a small neighbourhood of $\hat{\mathbf{p}}$, then

$$\mathbf{e}_{i,j}(\mathbf{p}) = \mathbf{e}_{i,j}(\hat{\mathbf{p}}) + \nabla \mathbf{e}_{i,j}(\hat{\mathbf{p}})(\mathbf{p} - \hat{\mathbf{p}}), \tag{3.4.2}$$

where we can derive the covariance matrix $\hat{\mathbf{\Sigma}}$ of $\hat{\mathbf{p}}$ from linear regression analysis [8], that is,

$$\hat{\mathbf{\Sigma}} = \hat{\sigma}^2 \left(\mathbf{J}_{\mathbf{F}}(\hat{\mathbf{p}})^T \mathbf{J}_{\mathbf{F}}(\hat{\mathbf{p}}) \right)^{-1}, \tag{3.4.3}$$

where $\mathbf{F}$ is a vector function in $\mathbf{p}$ obtained by stacking the columns $\mathbf{e}_{*,j}$, $j = 1, \ldots, M$, of

the matrix $\left(\mathbf{e}_{i,j}\right)$ into a vector, and

$$\mathbf{J}_{\mathbf{F}}\left(\mathbf{p}\right) = \left(\frac{\partial \mathbf{F}_i\left(\mathbf{p}\right)}{\partial \mathbf{p}_k}\right), \tag{3.4.4}$$

is the Jacobian of $\mathbf{F}$ of size $N \times L$. Here, $N = M \times n$ is the number of experimental measurements and the quantity $\hat{\sigma}^2 = f\left(\hat{\mathbf{p}}\right)/(N-L)$ is an unbiased estimator of σ^2. The diagonal elements of $\hat{\mathbf{\Sigma}}$ are the marginal variances of the parameter estimates $\hat{\mathbf{p}}_k$, $k = 1, \ldots, L$.

The matrix $\mathbf{J}_{\mathbf{F}}\left(\hat{\mathbf{p}}\right)^T \mathbf{J}_{\mathbf{F}}\left(\hat{\mathbf{p}}\right)$ in equation (3.4.3) is called the Fisher information matrix and provides a measure of the available information content about the unknown parameters $\hat{\mathbf{p}}$ in the data x_{ij}. When the Fisher information matrix is singular, that is, the matrix is noninvertible, the model is practically nonidentifiable, corresponding to infinitely large confidence regions [30]. Confidence regions for parameter estimates tell us, where around the parameter estimate $\hat{\mathbf{p}}$ the 'true' values $\mathbf{p}^*$ lie, with a certain probability. We estimate an $(1-\alpha)$ confidence region (see Draper and Smith [44]) for the parameter vector $\hat{\mathbf{p}}$ by

$$\left(\hat{\mathbf{p}} - \mathbf{p}^*\right)^T \mathbf{J}_{\mathbf{F}}\left(\hat{\mathbf{p}}\right)^T \mathbf{J}_{\mathbf{F}}\left(\hat{\mathbf{p}}\right)\left(\hat{\mathbf{p}} - \mathbf{p}^*\right) \leq C(\alpha), \tag{3.4.5}$$

where $C(\alpha) = \mathrm{F}\left(L, N-L, 1-\alpha\right) L\, f(\hat{\mathbf{p}})/(N-L)$ with $\mathrm{F}(L, N-L, 1-\alpha)$ being the $(1-\alpha)$ point of the F-distribution with L and $N-L$ degrees of freedom.

We can investigate the ellipsoidal region described by (3.4.5), which is centred at $\hat{\mathbf{p}}$, by determining the eigenvalues μ_k and corresponding eigenvectors $\mathbf{u}_k$ of the Fisher information matrix $\mathbf{J}_{\mathbf{F}}\left(\hat{\mathbf{p}}\right)^T \mathbf{J}_{\mathbf{F}}\left(\hat{\mathbf{p}}\right)$. The vectors $\mathbf{u}_k$ describe the directions of the principal axes of the ellipsoid and the values

$$l_k = \sqrt{\frac{C(\alpha)}{\mu_k}}, \tag{3.4.6}$$

indicate the respective radii along these axes. Knowledge of the principal axes provides us with information about certain linear combinations along which parameter values cannot be identified with sufficient accuracy.

To investigate the correlation between parameter pairs, the correlation matrix

$$\mathbf{Corr}\left(\hat{\mathbf{p}}\right) = \left(\frac{\mathbf{Cov}_{ij}\left(\hat{\mathbf{p}}\right)}{\sqrt{\mathbf{Cov}_{ii}\left(\hat{\mathbf{p}}\right)\,\mathbf{Cov}_{jj}\left(\hat{\mathbf{p}}\right)}} \right) , \tag{3.4.7}$$

with entries, say $r_{ij}(\hat{\mathbf{p}})$, $i, j = 1, \ldots, L$, and the property $|r_{ij}\left(\hat{\mathbf{p}}\right)| \leq 1$, provides us with information on the degree of linear dependence between parameters. We have perfect positive (negative) linear relationship between the parameter estimates $\hat{\mathbf{p}}_i$ and $\hat{\mathbf{p}}_j$ for $r_{ij}\left(\hat{\mathbf{p}}\right) = 1$ ($r_{ij}\left(\hat{\mathbf{p}}\right) = -1$) and no correlation for $r_{ij}\left(\hat{\mathbf{p}}\right) = 0$.

The covariance matrix estimate $\hat{\mathbf{\Sigma}}$, the lengths l_k and directions $\mathbf{u}_k$ of the principal axes of the ellipsoidal confidence regions and the correlation matrix $\mathbf{Corr}\left(\hat{\mathbf{p}}\right)$ present measures of the accuracy of model-to-data fits and can help us identify parameters (or parameter combinations) that cannot be well defined due to low estimation accuracy. We employ the measures in Chapter 6, where we evaluate the fit of a compartmental cell cycle model that we develop in Chapter 5. The following Chapter 4 presents the experimental work that we have undertaken to collect data we use for the model-fitting procedure.

Part II

RHPS4 and Cell Cycle Dynamics

CHAPTER 4

Experimental work

This Chapter outlines the experimental work which has been performed in order to gain more information on the cell cycle dynamics of colorectal cancer cells that were treated with different concentrations of the anti-cancer drug RHPS4. The cell cycle proportions, according to the DNA content of cells, were measured as well as the viable fraction of treated cells and compared to control cells for each day of incubation over a period of 10 days. We varied the period of cell incubation and the drug concentration across the experiments. We give an overview of the basic experimental procedures and materials used to obtain quantitative data of the cell cycle process in Sections 4.1.1-4.1.5, where Sections 4.1.3-4.1.5 are standard procedures, and largely follow the formulation given by Johnson [64]. In Section 4.1.6, we explain how we manipulated the collected data in order to make the data suitable for comparison with models developed subsequently in Chapter 5. Section 4.2 provides a summary of the experimental results, which we interpret with respect to the experimental setting and drug specificity in Section 4.3.

4.1 Experimental protocol, cell cycle analysis and normalisation of data

4.1.1 Suppliers of reagents

Beckman Coulter (UK) Ltd., Buckinghamshire, UK

FlowCheck®beads.

Fisher Scientific UK Ltd., Leicestershire, UK

Industrial methylated spirits (IMS).

Pharminox Ltd., Nottingham, UK

3,11-difluoro-6,8,13-trimethyl-8H-quino [4,3,2-kl] acridinium methosulfate (RHPS4) (a gift).

Sigma-Aldrich Co. Ltd., Dorset, UK

Dimethylsulphoxide (DMSO), foetal bovine serum (FBS), Ribonuclease A from bovine pancreas (RNAse A), RPMI 1640 liquid medium (a basic cell culture medium containing 0.3 g/L L-glutamine and 2 g/L sodium bicarbonate), phosphate buffered saline (PBS) tablets, propidium iodide (PI) (HPLC grade), sodium citrate, Titron X-100, trypsin-EDTA $1\times$ solution, Trypan blue.

4.1.2 Drug stock and cell line

The pentacyclic acridinium salt RHPS4 is a water-soluble compound, which facilitates rapid uptake into cells (saturation within 6 h), appears to be localised in the nuclear membrane, intranuclear bodies and cytoplasm, and does not show evidence of degradation or metabolism [35, 57]. RHPS4 is a telomere-interactive molecule possessing antitumoral activity [49]. Cells were incubated in RHPS4 at different concentrations to analyse its mechanism of action for shorter and medium periods of up to 10 days.

The stock of RHPS4 was made up to a concentration of $c_{\text{stock}} = 10$ mM in DSMO and stored at 4°C protected from light. Stocks were used for a maximum of 15 weeks and prior to use were sterile filtered using a 0.2 μM Sartorius filter. The desired concentration c_{assay} used in assays of $V_{\text{assay}} = 2$ml was obtained by adding a volume of $V_{\text{stock}} = c_{\text{assay}}\, V_{\text{assay}} / (c_{\text{stock}} - c_{\text{assay}})$ from the stock. As V_{stock} is of smaller order than 10^{-3} ml for the concentrations of 50 nM, 100 nM and 1 μM of RHPS4 used in the assays, the change of volume in the assays is negligible.

The HCT116 cell line was used for all assays. It is one of three strains of human malignant cells isolated from a male with colonic carcinoma. HCT116 cells grow in a monolayer and have a relatively short doubling time of 20.5 hours [25]. The cell line has been chosen

because of its good sensitivity to RHPS4 and reliable growth properties which allow analysis of cell viability and cell cycle analysis [64]. Moreover, RHPS4 can manifest its effects more efficiently in HCT116 cells than in many other cell lines due to the quick doubling rate of the cell line.

4.1.3 Cell stocks and general cell culture

Cells stocks were kept in a cell bank frozen at $-180°$C. To set up new cultures from frozen cells, we thawed the cells in a water bath at 37°C. We subsequently added the cells to a single 75 cm^2 Costar tissue culture flask that contained RPMI tissue culture medium at a temperature of 37°C. The cells were sub-cultured (see below) with RPMI 1640 medium twice before use in assays. The medium contained additionally 10% heat-inactivated FBS, where FBS heat inactivation was achieved by heating FBS to 55-59°C for 1 hour, and cooling it before use in the RPMI tissue culture medium.

We carried out all cell culture techniques aseptically in a BioMat^2 MDH Class II microbiological safety cabinet equipped with a laminar flow system. Before each use of the safety cabinet, we wiped down the cabinet with 70% IMS in distilled water.

To sub-culture cells, we aspirated the RPMI tissue culture medium from the flask followed by a brief rinse with sterile PBS solution, and subsequently added trypsin-EDTA $1\times$ solution, which causes cells to detach from the flask. We then resuspended the cells in fresh RPMI tissue culture medium in a new flask at a ratio of 1:10 and incubated the flask in a LEEC incubator at 37°C, which contained a tray of dH_2O to provide a humidified environment and supplied the cells with CO_2 for a 5% atmosphere. In order to maintain the exponential growth of human colorectal HCT116 cells, we subcultured them in Costar tissue flasks (25 cm^2 and 75 cm^2) twice weekly, or when approximately 70% confluent. The confluence level of cells is the fraction of surface area in the well covered by cells, where too high confluence levels ($> 80\%$) diminish the proportion of cycling cells in the flask.

For maintenance of the cell stocks in the cell bank, we grew cells to approximately 70% confluence in a 75 cm^2 flask. We then aspirated the tissue culture medium, rinsed the

cells twice with sterile PBS and trypsinised them before we resuspended them in 3 ml freezing solution that consisted of 95% heat inactivated FBS and 5% DMSO. We split the cell suspension into 1 ml aliquots in cryovials (Nalgene, 1.2 ml) and directly stored them at -20°C for 24 hours. When the cells were frozen, we transferred them to -80°C for a further 24 hours. For long-term storage, cells were stored in liquid nitrogen at -180°C.

4.1.4 Seeding of cells

In order to seed cells at a certain density, we aspirated all RPMI tissue culture from the flasks and briefly rinsed them with sterile PBS, before we detached the cells from the flask with 1 ml trypsin-EDTA per 25 cm^2. We then resuspended the cells in 6 ml RPMI tissue culture medium and syringed the cells gently through a 23G needle to break cell clumps into single cells. We derived the total number of cells in the flask by taking two samples from the flask to count the number of cells within each sample using a haemocytometer and taking the average. We subsequently seeded the cells at the desired density by suspending the appropriate amount of cells in RPMI tissue culture medium.

To prepare cells for analysis, we grew HCT116 cells in 6-well plates in a total volume of 2 ml of RPMI tissue culture medium per 9.6 cm^2 well and kept them at 37°C in a 5% CO^2, humidified atmosphere. The initial cell densities varied between 1000 and 200 000 cells per well dependent on the length of time of incubation. Before treating cells with 50 nM, 100 nM or 1 μM of RHPS4, we incubated them for about five hours or until attached to the bottom of each well. We then re-incubated the treated cells in unmodified conditions. Cells grew at different rates according to agent concentration and duration for which they were exposed to the drug, and we did not passage cells before the day of analysis.

4.1.5 Flow cytometric cell cycle analysis

We adapted the protocol from Riccardi and Nicoletti [111] for the cell cycle analysis of the HCT116 cell line. To harvest the cells, we collected all RPMI tissue culture medium, possibly containing dead cells and debris, in fluorescence-activated cell sorting (FACS) tubes and detached the remaining cells from the flask with 350 μl trypsin-EDTA per well.

We resuspended the cells in the collected tissue culture medium and subsequently transferred the suspension back into the FACS tubes. To wash off the empty wells, we used about 1 ml of PBS solution, and transferred the resulting solution to according FACS tubes to ensure an as accurate cell count as possible. To pellet the cells, we centrifuged them at 1200 rpm for 5 min at 4°C and removed the supernatant. Depending on the number of cells, we resuspended the cell pellet in 100-500 ml of PBS and syringed the suspension gently through a 23G needle to disperse cell clumps. We mixed about 15 μl of the suspension with 15 μl of 0.4% trypan blue (staining non-viable cells blue) in microcentrifuge tubes and counted the number of viable and dead cells in a heamocytometer. We then centrifuged the remaining cells again at 1200 rpm for 5 minutes at 4°C, decanted the supernatant and resuspended the pellet in 400 μl hypotonic fluorochrome solution, which had been composed of 50 μg/ml of the fluorescent dye propidium iodide (PI), 0.1% sodium citrate, 0.1% Titron X-100 and 0.1 mg/ml RNAse A in distilled water.

Prior to cell cycle analysis, we kept the cells at 4°C for 24 hours in the dark. We measured the PI fluorescence of individual nuclei by use of a Coulter Epics XL-MCL^{TM} flow cytometer operated using Expo32^{TM} software. We carried out a quality control check using FlowCheck®beads according to the manufacturer's instructions to monitor instrument alignment. When the quality control check was satisfactory, we set the flow rate to low and analysed a minimum of 20 000 events. We collected list mode data for fluorescence emission in the FL3 channel (detecting fluorescent light emitted by PI) of particles and used a dot plot of AUX (channel for peak fluorescence signal) against FL3 of all detected events (see figure 4.1) to distinguish single cells from unwanted doublets as described in Nunez [95].

To estimate the proportion of cells in each phase of the cell cycle (pre-G_1, G_1/G_0, S and G_2/M), we manually set gates on the DNA content histograms that we obtained from FL3 fluorescence of single cells using the software package WinMDI (freeware designed by J. Trotter, allowing the removal of aggregates or "doublets" consisting of two cells via gating of list mode data). Here, the term 'pre-G_1 cells' denotes cells with DNA content less than that of G_0/G_1 cells. An example of a typical DNA histogram after the gating of singlets indicating gates for the separation of cell cycle phases is given in Figure 4.1.

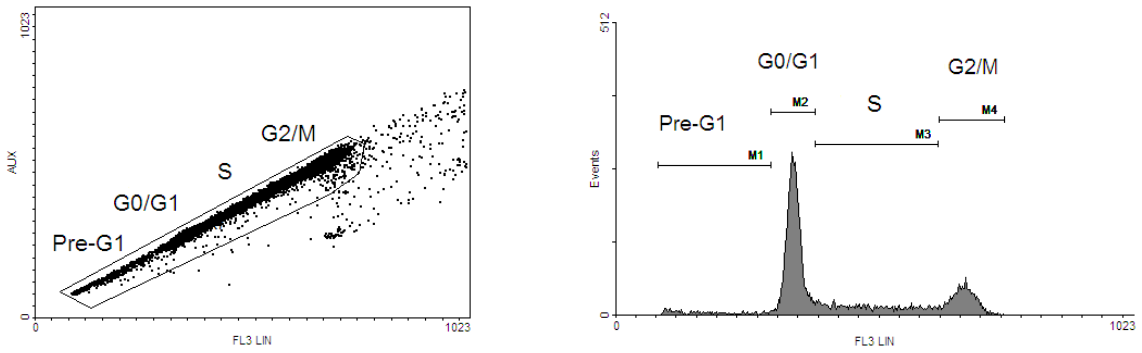

Figure 4.1: Gating of singlets and 'clean' histogram of FL3 of gated cells using the software package WinMDI.

4.1.6 Data collection and normalisation

The experimental procedure is inspired by work from Johnson [64], in which cell cycle dynamics for the HCT116 cell line were analysed over 21 days, where the time point of analysis coincided with cells approaching 70%-80% confluence (every 2-7 days), when they were subcultured at a density between 1 in 2 and 1 in 10 according to the anticipated doubling rate. In contrast, we did not split cells during culture, but adjusted seeding densities for each assay in order to minimise disruption of the natural cell growth, and analysed cells every day. The procedures, however, restricted the period of observation to 10 days due to the constraints of nutrient supply and extreme cell densities.

All cells for cell cycle analysis were taken from a HCT116 cell stock as described in Section 4.1.3. The cells were seeded at different densities in 6-well plates (see Section 4.1.4), where the seeding densities depended on the day of analysis in order to guarantee a confluence level of less than 80% in the cell population on the day of analysis (to maximise the proportion of cycling cells). Seeding densities were chosen carefully as, on the one hand, too dense cell-to-cell contacts can inhibit cell growth, but on the other hand, sufficient cell numbers are needed both for meaningful cell cycle analysis and to allow for a comparable cell environment up to the day of analysis and across the experiments. The seeding densities correspond to estimates of the population growth in each experiment. Higher initial seeding densities were used for treated cells to compensate for reduced population doublings. The seeding densities chosen for the individual experiments are shown in Table 4.1.

Every experiment was set up with six identical experiments (replicates) to assess the vari-

Day of analysis	Molar concentration of RHPS4			
	CONTROL	50 nM	100 nM	1000 nM
1	300000	300000	300000	300000
2	200000	200000	200000	200000
3	150000	150000	150000	200000
4	100000	100000	100000	200000
5	40000	50000	50000	100000
6	20000	50000	50000	100000
7	10000	20000	50000	100000
8	2000	12500	40000	100000
9	1000	12500	40000	100000
10	1000	12500	25000	100000

Table 4.1: Seeding densities (cell numbers per 2 ml of RPMI medium) used for cell cycle analysis of control cells and cells treated with 3 different concentrations of RHPS4, where cells were incubated between 1 and 10 days.

ability associated with the cell growth cycle: the cells were incubated and treated with RHPS4, and cell cycle analysis was conducted as in Section 4.1.5. Control cells (under normal conditions of no drug) and cells treated with 50 nM, 100 nM, 1 μM of the drug were used and analysed to a maximal period of 10 days.

The total number of doublings (TND), for each concentration of the drug, and for each day i of analysis, was computed for each experiment using the formula

$$\mathrm{TND}_i(t) = \frac{\ln N_i(t) - \ln N_i(0)}{\ln 2}\,, \tag{4.1.1}$$

where $\mathrm{TND}_i(t)$ and $N_i(t)$ are, at time point t, the total number of doublings and the total number of cells, respectively, from the experiment used to analyse cells at day i. We now assume that the total number of doublings is independent of the initial seeding density and write $\mathrm{TND}_i = \mathrm{TND}$ for all i.

Now we consider an experiment, where all initial seeding densities are identical, and normalise our data by setting $N_i(0) = 10^3$ for all i. The normalised data set can be derived from the original data by

$$N(t) = 10^3 \cdot 2^{\mathrm{TND}(t)} = 10^3\,\frac{N_i(t)}{N_i(0)}\,, \tag{4.1.2}$$

where $N(t)$ is the total number of cells that have grown from 10^3 cells in t days. The

number of cells in each phase of the cell cycle, including the number of viable and dead cells, can then be obtained by multiplying the respective proportions with $N(t)$. Table 4.2 contains the experimental data used for calculations in tabular form.

4.2 Experimental results

We are interested in the response of cells to 4 different levels of the drug and accordingly measured the cell growth, cell cycle distribution and cell viability in each assay throughout a period of 10 days as described in Section 4.1.

Day 1 is the first day of analysis when measurements have been taken. The total number $N_i(t)$ of cells at the day $t = i$ of analysis together with their respective cell seeding densities $N_i(0)$ are shown in Figure 4.2 for control cells and all three different concentrations of the drug.

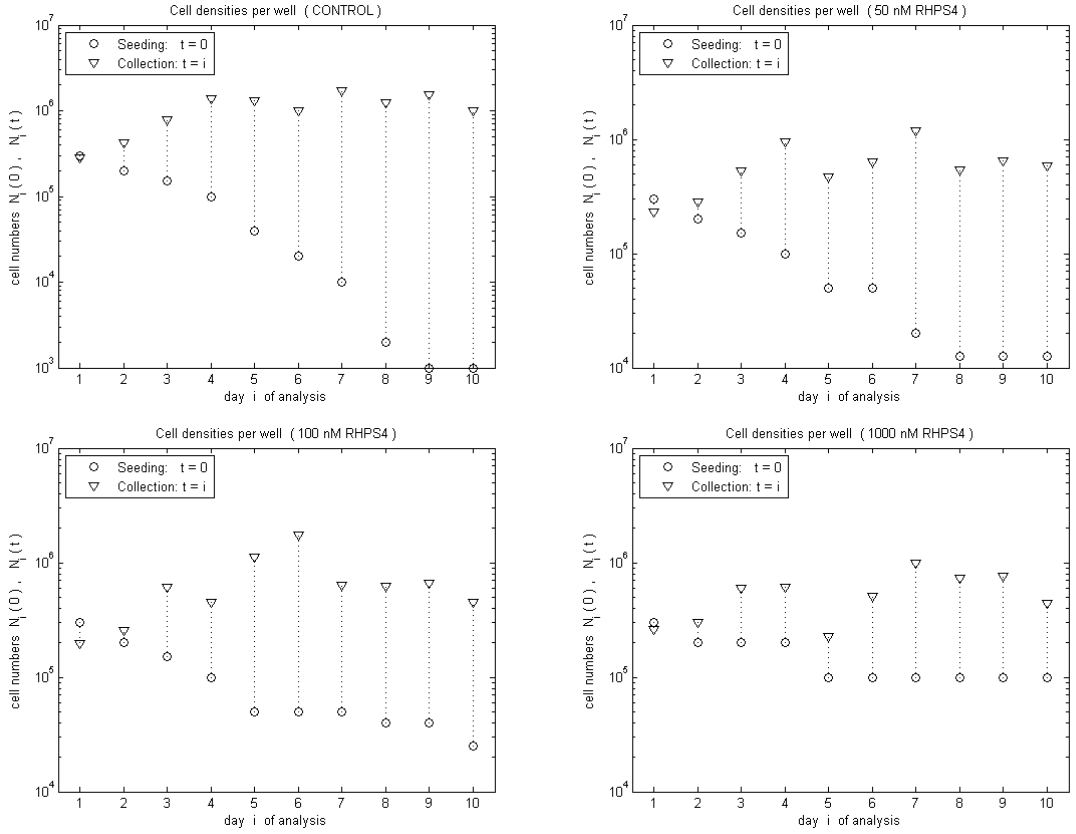

Figure 4.2: Cell densities $N_i(t)$, at time point t, for each experiment used to analyse cells at day i, indicated by inverted triangles, where $N_i(0)$ are the respective seeding densities per 2 ml of medium, denoted by circles. Data were collected for control cells and three concentrations (50 nM, 100 nM, 1 μM) of RHPS4, and cell numbers are shown on a log scale.

RHPS4 data

Day		Doublings		Apoptosis		Pre-G1 phase		G0/G1 phase		S phase		G2/M phase	
		Average	±SD	Average	±SD	Average	±SD	Average	±SD	Average	±SD	Average	±SD
1	**Control**	-0.1	0.3	6.9	1.7	1.7	0.9	48.0	1.9	20.9	1.0	29.8	2.3
1	**50nM**	-0.6	0.2	15.1	4.4	2.5	0.4	42.7	1.0	22.8	1.5	32.5	0.9
1	**100nM**	-0.4	0.1	13.0	3.4	2.1	0.7	42.9	2.2	21.3	0.8	34.2	2.2
1	**1μM**	-0.2	0.2	11.2	7.8	2.7	1.3	51.6	1.6	15.6	2.3	30.2	1.9
2	**Control**	1.1	0.2	7.5	4.7	1.5	0.9	46.5	0.4	28.0	0.9	24.4	1.3
2	**50nM**	0.5	0.2	7.9	7.1	1.9	0.7	44.0	3.2	28.1	2.9	26.4	1.9
2	**100nM**	0.3	0.3	14.0	8.7	2.9	1.7	46.5	2.1	26.8	2.7	23.6	2.2
2	**1μM**	0.5	0.4	12.3	3.2	1.5	0.6	54.2	1.4	22.8	1.8	21.9	0.9
3	**Control**	2.3	0.5	17.5	13.2	4.3	0.9	55.3	2.1	20.4	2.0	20.3	1.0
3	**50nM**	1.9	0.6	12.5	7.9	2.4	0.8	54.2	1.6	21.7	1.3	22.0	1.2
3	**100nM**	1.8	0.3	15.6	9.7	2.8	0.5	57.1	1.3	18.6	1.4	22.0	1.2
3	**1μM**	1.5	0.6	28.6	8.9	3.7	1.7	65.1	1.5	13.5	0.8	17.9	1.0
4	**Control**	3.8	0.3	20.9	3.4	1.3	1.0	47.2	3.5	23.0	1.7	28.5	3.7
4	**50nM**	3.2	0.5	23.5	11.5	3.9	1.9	71.4	2.0	7.8	0.8	17.1	3.2
4	**100nM**	2.1	0.4	39.9	12.0	2.8	0.5	58.6	0.9	12.4	0.6	26.4	1.0
4	**1μM**	1.6	0.4	76.7	3.2	4.0	0.6	69.9	1.0	5.0	0.9	21.2	0.7
5	**Control**	5.0	0.2	16.9	5.0	2.3	0.8	72.8	1.3	10.4	0.2	14.7	2.5
5	**50nM**	4.4	0.3	21.1	8.0	4.3	1.5	72.1	1.8	7.2	1.0	16.5	1.6
5	**100nM**	3.2	0.3	97.7	0.9	7.2	1.2	61.2	2.1	12.8	0.4	18.9	1.5
5	**1μM**	1.2	0.2	80.9	6.5	1.0	0.3	58.4	3.4	15.0	2.3	25.7	1.5
6	**Control**	5.6	0.4	18.0	6.7	3.5	1.3	56.6	0.6	20.7	0.6	19.4	1.0
6	**50nM**	5.1	0.3	15.5	4.0	2.4	0.6	63.9	2.3	11.6	1.2	22.1	3.1
6	**100nM**	3.6	0.2	96.7	1.3	15.0	7.4	51.1	7.5	14.7	2.0	19.5	1.7
6	**1μM**	2.3	0.1	97.6	2.7	2.7	1.6	60.1	4.3	12.0	2.4	25.6	2.5
7	**Control**	7.2	0.9	19.1	6.4	3.9	2.4	79.7	2.4	4.3	0.6	12.2	1.5
7	**50nM**	5.8	0.4	97.9	2.1	29.0	3.3	47.5	2.2	10.8	1.6	12.9	2.5
7	**100nM**	3.6	0.3	100.0	0.0	19.1	1.3	49.1	1.4	13.7	1.5	18.4	0.8
7	**1μM**	3.2	0.5	100.0	0.0	2.4	2.1	64.5	3.0	8.5	1.6	24.7	3.6
8	**Control**	9.2	0.2	8.3	2.9	1.5	0.8	62.6	3.0	17.9	1.7	18.3	1.2
8	**50nM**	5.4	0.3	96.8	4.1	2.9	0.4	67.6	2.0	8.6	1.9	20.9	0.4
8	**100nM**	3.9	0.1	98.9	1.0	4.8	1.2	65.7	2.8	9.3	0.8	20.3	2.5
8	**1μM**	2.9	0.0	99.8	0.3	4.3	0.9	68.4	2.3	7.4	0.6	20.0	1.0
9	**Control**	10.6	0.1	11.8	2.2	1.3	0.6	86.6	1.5	4.1	0.2	8.2	0.8
9	**50nM**	5.7	0.3	100.0	0.0	6.7	0.6	66.6	0.9	9.7	0.9	17.0	0.9
9	**100nM**	4.0	0.2	100.0	0.0	13.9	1.9	58.2	1.6	11.2	0.4	16.7	1.0
9	**1μM**	2.9	0.1	97.7	3.6	20.2	12.4	57.7	8.5	7.5	1.2	14.7	3.3
10	**Control**	9.9	0.4	14.3	5.4	1.1	0.1	79.7	1.4	5.1	0.9	14.2	0.9
10	**50nM**	5.5	0.1	100.0	0.0	10.5	1.4	52.1	2.7	17.3	1.6	20.6	1.0
10	**100nM**	4.2	0.1	100.0	0.0	9.3	0.6	55.0	1.7	15.1	2.0	21.2	0.5
10	**1μM**	2.1	0.3	99.8	0.4	19.3	1.4	52.4	1.9	10.6	3.7	18.0	2.4

Table 4.2: The Table shows the average values and respective standard deviations of the total number of doublings and the proportions (in %) of HCT116 cells that are apoptotic (trypan-blue stained), in the pre-G_1, G_0/G_1, S and G_2/M phases of the cell cycle. Cells were exposed to treatment with different concentrations of RHPS4 (control, 50 nM, 100 nM, 1 μM) and analysed throughout a period of 10 days.

We use $N_i(0)$ and $N_i(t)$ to derive normalised data sets according to (4.1.2). We observe that control cells grow exponentially between days 1 and 9 as represented by the approximately linear increase of total cell numbers $N(t_i)$, $i = 1, \ldots, 10$, on the log-scale in Figure 4.3. The data point of control cells at t_{10} is markedly lower than the data point at t_9, due to confluence and probably nutrient deficiency in the wells; we therefore disregard this data point in our analysis. Error bars result from the standard deviation over 6 replicates and are rather small (mostly less than 10%). It is important to note that they do not account for variability within different cell batches or repeated experiments incorporating intermediate breaks, as the replicates were obtained from cells of the same batch being seeded in 6 parallel wells. Further comments on the particular cell behaviour can be found in Section 4.3.

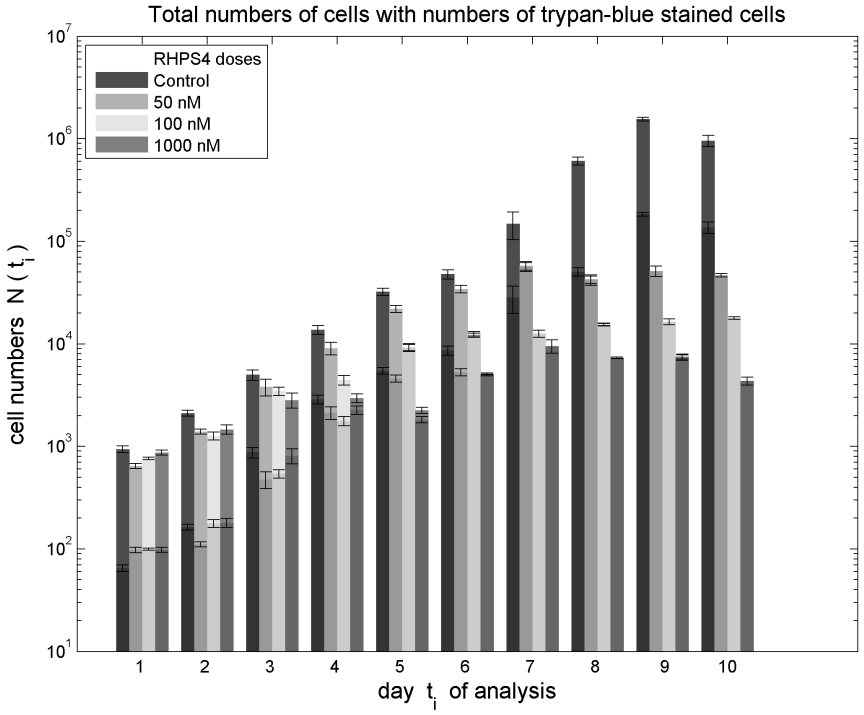

Figure 4.3: Total numbers $N(t_i)$ of cells and the numbers of dead cells (measured by trypan blue dye-exclusion, lower proportion of each bar) for no drug and each drug concentration (50 nM, 100 nM, 1 μM) of RHPS4 are shown on a log scale at each day t_i, $i = 1, \ldots, 10$ of analysis. Error bars represent the standard deviations of the corresponding values.

For treated cells, we observe growth inhibition and, in general, cell growth declines with increasing drug concentration. The total number of doublings at day 4 is 3.78 for control cells, 3.18 for 50 nM, 2.15 for 100 nM and 1.56 for 1 μM. Growth reduction occurs after a period of 6-7 days for 50 nM, 3-5 days for 100 nM and 3-4 days for 1 μM of RHPS4, with

the respective total cell numbers levelling off thereafter.

The average proportion of dead cells was around 14% for control cells, and low proportions of pre-G_1 cells (cells with fragmented DNA) at less than 5% were observed. Without treatment, the proportion of S phase and G_2/M phase cells ranged from 4% to 28% and from 8% to 30%, respectively, over the period of observation. These proportions tended to decrease gradually from higher to lower levels. Accordingly, the proportions of G_0/G_1 cells increased from 50% to 80%, that is, control cells progressively accumulated in the G_0/G_1 phase. The proportions of cells in each phase of the cell cycle (G_0/G_1, S, G_2/M and pre-G_1) are illustrated in Figure 4.4, where error bars represent standard deviations over 6 replicates.

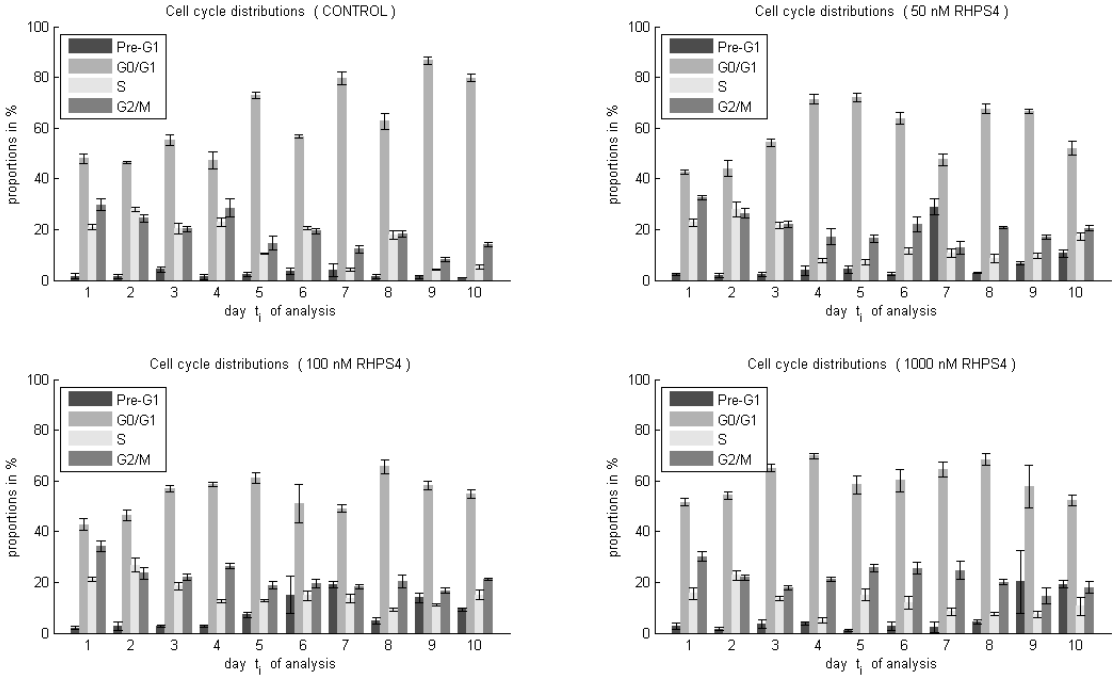

Figure 4.4: Proportion of cells in each phase of the cell cycle at each day of analysis. Data were collected for control cells and three concentrations (50 nM, 100 nM, 1 μM) of RHPS4. Error bars represent standard deviations of 6 replicates.

Interestingly, the proportion of treated cells in the S phase tended to drop from around 25% at the beginning of the experiment to about 8.5% on average by day 4 and returned to a slightly higher level of around 15% after this drop. The most dramatic effect was achieved with 1000 nM of drug when the S phase proportions dropped to about 5% at day 4. The proportions of G_2/M cells behaved in a similar way but with less remarkable trends, 30% at day 1 decreasing to 20% at day 4 and 18% on average during the subse-

quent days. The proportion of cells in the G_0/G_1 phase increased by 25% during the first 4 days from an average of 45% and thereafter decreased to around 50% for all treated cells.

Typical DNA distributions obtained from the flow cytometer at days 4 and 10 are shown for control cells and for cells treated with 1 μM RHPS4 in Figure 4.5. The DNA distribution of control cells contains a markedly higher peak in the range of G_0/G_1 DNA at day 10 compared to the DNA distribution at day 4. Analysis of cells treated with 1 μM RHPS4 results in a much smaller S phase DNA content and a higher pre-G_1 DNA content at day 4 and we observe a marked increase in the pre-G_1 DNA content at day 10.

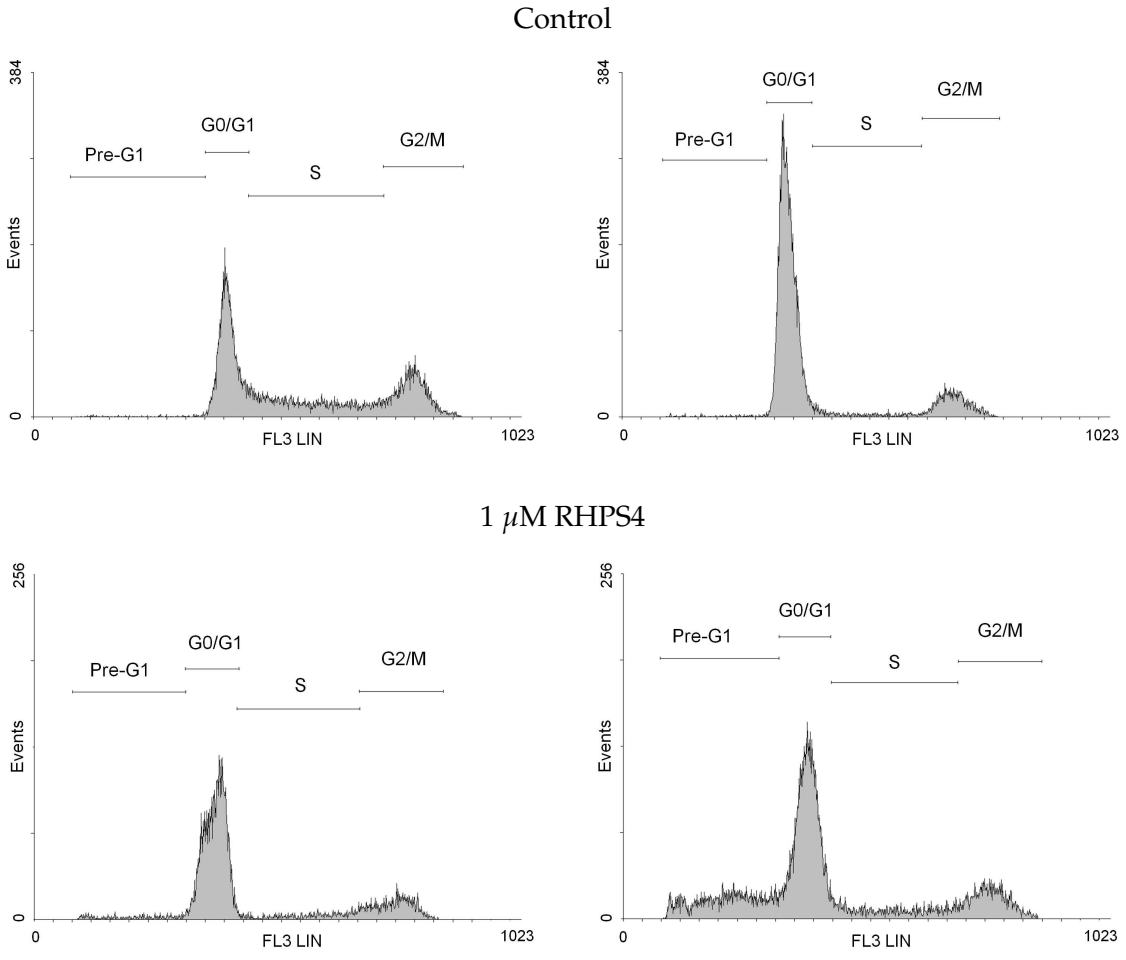

Figure 4.5: Distributions of the DNA content for control cells (top) and cells treated with 1 μM of RHPS4 (bottom), each analysed at day 4 (left) and day 10 (right), with manually set gates for the estimation of the corresponding cell cycle distributions.

Before the onset of considerable cell death, the proportions of pre-G_1 cells remained at a low level of about 1-4% for all drug concentrations, and then increased to around 25%,

where the increase occurred at the later time point of day 9 for the highest drug concentration. Strikingly, in the experiments with 50 nM and 100 nM of RHPS4, the proportions of pre-G_1 cells were observed to reduce again after day 7, to around 3-14%, possibly due to disintegration of dead cells in medium.

Marked cell death (trypan blue staining) was observed at day 7 for 50 nM, at day 5 for 100 nM and at day 4 for 1 μM of RHPS4, when nearly all cells died abruptly, from a prior level of around 15% at the beginning of the experiment. The intensive cell death set in more abruptly for lower drug concentrations than for higher doses. Altogether, the drug RHPS4 seems not only to affect entry into the cell cycle ($S/G_2/M$ phases) around day 4, but also to inhibit cell growth up to a complete cessation of the replication processes with increasing drug concentration.

4.3 Interpretation of results and discussion

Colorectal cancer cells were plated in medium with different concentrations of RHPS4 for up to 10 days and their growth properties and cell cycle distribution analysed. The decrease in the total numbers of control cells at day 10 is likely to result from environment-dependent growth inhibition, as the medium has not been replaced throughout the experiments in order not to interfere with the cells' natural environment. The proportions of G_0/G_1 cells displayed an upward trend over the period of 10 days, which may also represent a consequence of diminishing nutrient supply over time.

Also, the number of cells at day 1 was slightly lower than the value of the initial number of cells at seeding (day 0), indicating that cells did not replicate during the first day after seeding. This can also be partly due to not all cells attaching to the bottom of the well after seeding in medium. For simplicity of analysis, we consider day 1 as the start of the analysis of cell cycle modelling.

Estimates of the initial cell densities were good in terms of achieving around 70-80% confluence at the day of analysis for control cells and slightly lower confluence levels with increasing drug concentration. This was due to growth properties being harder to predict for treated cells. Also, cells seeded at densities lower than about 1000 cells did

not grow satisfactorily. There is a slight correlation between cell densities $N_i(t_i)$ at the day of analysis and respective proportions of cells in the G_0/G_1 phase, in particular for control cells (compare Figures 4.2 and 4.4). However, the varying confluence levels did not appear to affect cell growth.

As well as cell cycle analysis, we measured the proportion of nonviable cells by trypan blue dye-exclusion, as DNA degradation is not an early event in apoptosis and, therefore, should not be taken as the indicator of cell death [141]. Apoptotic cells and other dead cells split into apoptotic bodies and their DNA degradation takes place after the onset of cell death.

Propidium iodide is a fluorescent dye that binds stoichiometrically to DNA molecules. We analysed the amount of dye incorporated in stained cell material treated with different concentrations of RHPS4 to derive DNA histograms. Inaccuracies in measuring the fractions of pre-G_1 cells can arise from one or more of the following: firstly, on death, one cell can produce several fragments each with some DNA content. A correction for the number of apoptotic cells was suggested by Johnson [64] as $n_A = \sum_i n_i \, I_i / I_{G_0/G_1}$, where n_i is the number of detected pre-G_1 cell fragments with fluorescence intensity I_i and I_{G_0/G_1} is the fluorescence intensity of a G_0/G_1 cell measured by flow cytometry. Secondly, cells can undergo apoptosis from all phases of the cell cycle, hence debris from S or G_2/M cells may not be detected as 'pre-G_1' at all [67] as they may have the DNA content of normal cells (in G_0/G_1). The above formula for n_A, in this respect, is an overestimate of the number of apoptotic cells, as it assumes that all cells die with a G_0/G_1 DNA content. Since above reasons include arguments for under- and overestimation of the actual proportion of late-apoptotic cells, our method of deriving the pre-G_1 fraction serves as a good indication for the frequency of apoptotic cells.

There was a decisive reduction in the number of cells treated with 1 μM RHPS4 between days 7 and 10. In particular, the number of harvested cells incubated for more than 7 days decreased, although the respective seeding densities were the same in these experiments. Dead cells split into apoptotic bodies and disintegrate in the medium making it hard to detect such cells. Indeed, cell debris was observed while cells were counted, and a complete disintegration of apoptotic bodies in the medium may have been a reason for

the possible underestimate of the number of dead cells (and the number of pre-G_1 cells) towards the end of the observation period.

The fraction of cells in the S phase drops by 10-20% after 4 days suggesting that cells from the HCT116 cell line are affected by the drug in this phase. This rather temporary effect, which recovers in days 5-6 (see Figure 4.4), indicates that in many treated cells, the effect of the drug is to prevent cells passing the G_1-S check point where a cell commits to the DNA synthesis phase of cell division. This can be due to cells resolving certain chromosome replication defects introduced by the drug to make cellular conditions favourable for cell division.

In summary, we have shown that the compound RHPS4 has a concentration-dependent effect on the proliferation of HCT116 cells, where growth reduction occurs at a later time point for lower drug concentrations, but cell cycle progression is generally also affected at an early stage of incubation with the drug. Similar to interpretations of RHPS4-related effects in Rizzo et al. [112], who exposed HCT116 cells to the drug for 10 days, collected data at days 4,6,8,10, and found significant growth inhibition, we suggest that RHPS4 interferes with the replication fork during DNA synthesis, causing DNA damage and apoptosis. Cells may initially recover rapidly from the replication stress, explaining the dip in the S phase proportions at day 4 with a slight, subsequent increase, but eventually undergo cell death due to the inability to fix further defects.

The experimental results can be further analysed by means of mathematical models describing the cell cycle dynamics of control and treated cells, the development of which, and comparison to experimental data are presented in Chapters 5 and 6, respectively.

CHAPTER 5

Mathematical model of cell cycle dynamics and the effects of RHPS4

The aim of model development and analysis in this thesis is to find out how the drug RHPS4 affects cells of the HCT116 cell line through their cell cycle and growth. Here, we want to model the states and dynamics related to the experiments described in Chapter 4. Data of the proportions of viable and dead cells and data of proportions of cells in each phase, G_0/G_1, S, G_2/M, pre-G_1, of the cell cycle have been collected under control conditions and under treatment with different concentrations, 50 nM, 100 nM and 1 μM of the potential anti-cancer drug RHPS4 over 10 days.

The phases G_0/G_1, S, G_2/M can be distinguished by their DNA content in a cell, being onefold during the phases G_0 and G_1, and twofold during phases G_2 and M, and between one- and twofold during the synthesis phase S. Cells in one of these cell cycle phases are either viable or have been measured a short time after the onset of cell death. Pre-G_1 cells are dead, being detected some time after the onset of cell death and contain fragmented DNA with less DNA content than a G_0/G_1 cell. A simplified cell cycle model of viable cells is depicted in Figure 5.1. Cell death is indicated by loss of cell material from each of the three cell-cycle phases.

In this Chapter, we present a novel cell cycle model, describing the overall dynamics underlying our experimental data, and provide the theoretical basis for fitting these models to data. In Section 5.1, we present the mathematical cell-cycle model and analyse its

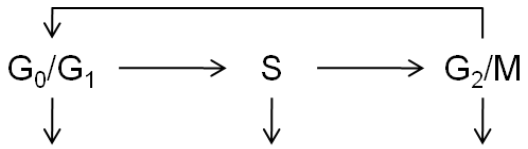

Figure 5.1: A cell cycle model of viable cells including death from each of the three phases.

behaviour with respect to the solution dynamics. We also introduce an asymptotic expression for the doubling time of a cell population. In Section 5.2, we show how our deterministic model can be derived from a continuous-time Markov process, where we introduce the probabilistically motivated expressions for the residence times of cells in a cell cycle phase. To describe our experimental measurements mathematically, in Section 5.3 we introduce a statistical model of the data. In Section 5.4, we show that our model is structurally identifiable, that is, each model solution corresponds to a unique set of parameter values. Numerical optimisation methods that we use to estimate cell cycle parameters from our experimental data are presented in Section 5.5, and techniques for model comparison and residual testing to evaluate model-data fits statistically are given in Section 5.6.

5.1 Mass balance equations and analysis

The nature of the cell states that can be detected by our experimental set-up suggests the assignment of seven compartments, namely X, Y, Z for viable cells being in G_0/G_1, S, G_2/M, respectively, $\bar{X}$, $\bar{Y}$, $\bar{Z}$ for cells dying recently in each of the cell-cycle phases, and A for all pre-G_1 cells. Note that the experimental measurements do not allow us to distinguish between viable and dead cells of the same DNA content. The observable states therefore differ from the classification, we can only observe cells in the combined compartments $X + \bar{X}$, $Y + \bar{Y}$, $Z + \bar{Z}$, $X + Y + Z$, $\bar{X} + \bar{Y} + \bar{Z} + A$ and A.

Model compartments consist of homogeneous and well-mixed material and are linked by material flows between them. We assume balanced cell growth, that is, cell proportions in each of the cell cycle phases maintain constant ratios to one another. Viable cells go around the cell cycle $X \to Y \to Z \to 2X \to \ldots$, where cells double at the transfer from Z =G_2/M to X =G_0/G_1. It is possible that cells die from all the phases, G_0/G_1,

S, G_2/M, of the cell cycle, that is, they are transferred as $X \to \bar{X}$, $Y \to \bar{Y}$ and $Z \to \bar{Z}$. Once cells have died, their DNA cannot be synthesised anymore and their nucleus is subject to DNA degradation, that is $\bar{X} \to A$, $\bar{Y} \to A$, $\bar{Z} \to A$. The seven-compartment model is a modification of model (2.2.1)-(2.2.5) in Section 2.2, refined by the addition of compartments $\bar{X}$, $\bar{Y}$, $\bar{Z}$ for early cell death in the G_0/G_1, S, G_2/M phase, respectively. In contrast to model (2.2.1)-(2.2.5), this model does not contain a compartment for senescent cells, which may be included in the X compartment. Figure 5.2 illustrates the seven-compartment model including the observable states. We use the principle of mass action

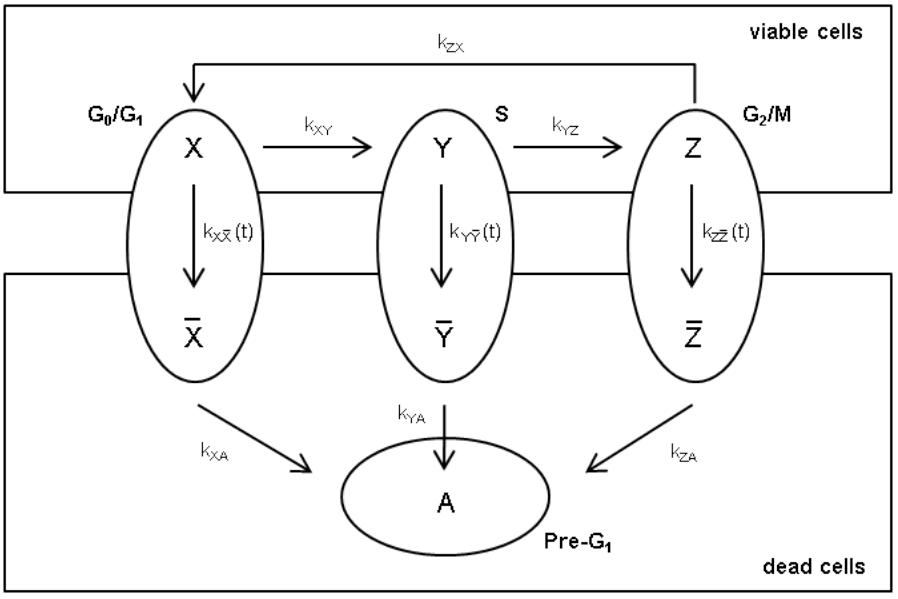

Figure 5.2: A seven-compartment model with compartments X, Y, Z, $\bar{X}$, $\bar{Y}$ $\bar{Z}$, A, arising from the data collected in Chapter 4. Transition rates k between compartments are assumed to be constant except for $k_{X\bar{X}}$, $k_{Y\bar{Y}}$, $k_{Z\bar{Z}}$ being possibly time-dependent functions. We group together the observed quantities, that is, the number of cells in each phase $G_0/G_1 = X + \bar{X}$, S$= Y + \bar{Y}$, $G_2/M = Z + \bar{Z}$, pre-$G_1 = A$ (oval areas), and viable $(X + Y + Z)$, dead $(\bar{X} + \bar{Y} + \bar{Z} + A)$ cells (rectangular areas).

to model the dynamics with the system of ODEs

$$\begin{aligned}
\frac{dX}{dt} &= 2\,k_{ZX}\,Z - (k_{XY} + k_{X\bar{X}})\,X\,, && (5.1.1)\\
\frac{dY}{dt} &= k_{XY}\,X - (k_{YZ} + k_{Y\bar{Y}})\,Y\,, && (5.1.2)\\
\frac{dZ}{dt} &= k_{YZ}\,Y - (k_{ZX} + k_{Z\bar{Z}})\,Z\,, && (5.1.3)\\
\frac{d\bar{X}}{dt} &= k_{X\bar{X}}X - k_{XA}\,\bar{X}\,, && (5.1.4)\\
\frac{d\bar{Y}}{dt} &= k_{Y\bar{Y}}Y - k_{YA}\,\bar{Y}\,, && (5.1.5)\\
\frac{d\bar{Z}}{dt} &= k_{Z\bar{Z}}Z - k_{ZA}\,\bar{Z}\,, && (5.1.6)\\
\frac{dA}{dt} &= k_{XA}\,\bar{X} + k_{YA}\,\bar{Y} + k_{ZA}\,\bar{Z}\,, && (5.1.7)
\end{aligned}$$

with initial values $(X, Y, Z, \bar{X}, \bar{Y}, \bar{Z}, A)^T|_{t=0} = (X_0, Y_0, Z_0, \bar{X}_0, \bar{Y}_0, \bar{Z}_0, A_0)^T \geq \mathbf{0}$ being non-negative. The rate of change of cell numbers in the system is assumed to be equal to the inflow minus the outflow of cell material. Cell material which is transferred from the Z compartment to the X compartment doubles during transition giving a factor of 2 in the equation for the rate of change in X. Transition rates between compartments are denoted by k subscripted with labels corresponding to relevant compartments (e.g. k_{XY} is the rate of transition from the X to the Y compartment) and are assumed to be non-negative and constant.

We aim to model the most significant cell cycle dynamics without and with treatment by RHPS4 and therefore aim to focus on capturing the dynamics related to the observed marked cell death of treated cells during the period of observation. We allow the rate coefficients governing transition from viable cells to dead cells to be time-dependent, that is we let $k_* = k_*(t)$ where (*) represents $X\bar{X}$, $Y\bar{Y}$, or $Z\bar{Z}$.

At first, we assume $k_*(t) = k_{*0}$ is constant (**model $\mathbf{M}_0$**), that is, transition rates from viable to dead cells do not change over time. This may be especially true for control cells as they probably grow in an unhindered manner. However, data corresponding to treated cells may not be reflected well by this model. Cell death has been observed to set in after a certain time lag which was dependent on the agent concentration. We want to consider two refined models, namely a model involving an ongoing rate increase (**model $\mathbf{M}_1$**) and

a model involving an abrupt rise (**model $\mathbf{M}_2$**) in the transition rates to cell death. For both models, we assume that these effects set in at a certain time point $t_0 > 0$, and that they remain after this time point. The basic rate functions describe, for each of $k_{X\bar{X}}$, $k_{Y\bar{Y}}$, $k_{Z\bar{Z}}$,

($\mathbf{M}_0$) constant behaviour: $k_*(t) = k_{*0}$,

($\mathbf{M}_1$) a linear increase: $k_*(t) = k_{*0} + m_*(t - t_0)\,\mathrm{H}(t - t_0)$,

($\mathbf{M}_2$) a sigmoidal increase: $k_*(t) = k_{*0} + \Delta k_*\, t^\beta / \left(t_0^\beta + t^\beta\right)$,

with k_{*0} being the initial rate of cell death from the compartment (*). The parameter m_* is the gradient of the linear increase from time point t_0 on and $\mathrm{H}(t)$ is the Heaviside function, which takes value 1 for $t \geq 0$ and 0 for $t < 0$. For the abrupt increase, we choose a sigmoidal function with the magnitude of the increase in $k_*(t)$ being Δk_* and β being a parameter controlling the shape of the transition of $k_*(t)$ at t_0. Note that the rate function of model M_0 is the special case of M_1 with $m_* = 0$ and of M_2 with $\Delta k_* = 0$. The rate models M_0, M_1, M_2 are illustrated in Figure 5.3. The choice of each of the parameters in the transition functions will later be evaluated and the best model chosen to find the most appropriate description of the cell cycle dynamics for each drug concentration.

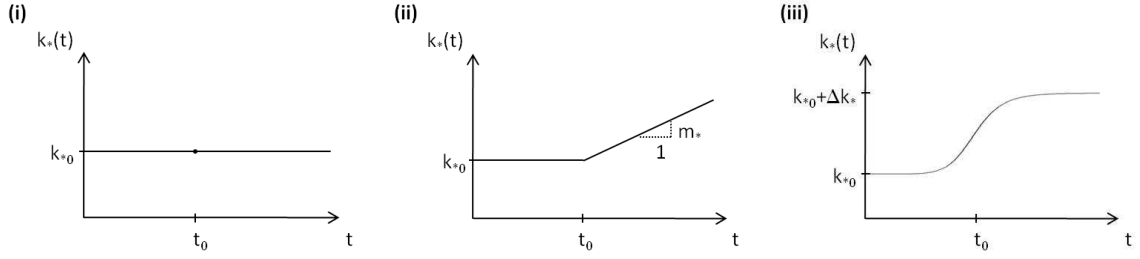

Figure 5.3: A schematic illustration of (i) rate model M_0, (ii) rate model M_1, and (iii) rate model M_2, governing transition from viable cells to dead cells in the seven-compartment model in Figure 5.2. The parameter k_{*0} denotes the initial rate of cell death, t_0 the time point of the onset of significant cell death, m_* the gradient of the rate increase of model M_1, Δk_* the magnitude of the increase in rate and β a shape parameter of model M_2.

Of course, other models (and rate changes) could be considered: but we know from other work [77] that the effect of the drug is to accelerate senescence and apoptosis. See Chapter 7 for more detailed analysis of how RHPS4 affects the stability of telomeres that is necessary to maintain the protective cap of chromosomes in cancer cells.

We now analyse the dynamic behaviour of the system (5.1.1)-(5.1.7), where the analysis of the model solutions is similar to that of the five-compartment model in Chapter 2.2.

Therefore we only state the results without derivation. The subsystem of equations (5.1.1)-(5.1.3) for X, Y and Z is mathematically independent of $\bar{X}$, $\bar{Y}$, $\bar{Z}$ and A. Hence, we consider separately the cell-cycle dynamics of viable cells, given by the 3×3 system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} -k_{XY} - k_{X\bar{X}} & 0 & 2\,k_{ZX} \\ k_{XY} & -k_{YZ} - k_{Y\bar{Y}} & 0 \\ 0 & k_{YZ} & -k_{ZX} - k_{Z\bar{Z}} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \tag{5.1.8}$$

where we consider only the case of constant rates to cell death (model M_0) with $k_*(t) = k_{*0}$. Let $\mathbf{c} = (X, Y, Z)^T$ be the vector of cycling cells and $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ be the eigenvectors of the coefficient matrix of (5.1.8) corresponding to eigenvalues ξ_1, ξ_2, ξ_3, respectively. We assume that all rate constants k_* are non-negative, and, for simplicity, that all eigenvalues are distinct. Solutions of (5.1.8) are then of the form

$$\mathbf{c}\,(t) = a_1\, e^{\xi_1 t}\,\mathbf{r}_1 + a_2\, e^{\xi_2 t}\,\mathbf{r}_2 + a_3\, e^{\xi_3 t}\,\mathbf{r}_3\,, \tag{5.1.9}$$

with a_1, a_2, a_3 being real constants, which are determined by the linear system

$$\mathbf{R}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix}, \tag{5.1.10}$$

where $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ is the fundamental matrix and $\mathbf{c}\,(0) = (X_0, Y_0, Z_0)^T$ are the initial conditions of (5.1.8) at $t = 0$.

Equations (5.1.4)-(5.1.6) are first order linear differential equations and can be solved by the integrating factor method. The general solutions for $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ respectively are

$$\bar{X}\,(t) = b_1\, e^{-k_{XA}\,t} + k_{X\bar{X}0} \sum_{i=1}^{3} \frac{a_i\,(\mathbf{r}_i)_1}{k_{XA} + \xi_i}\, e^{\xi_i\,t}\,, \tag{5.1.11}$$

$$\bar{Y}\,(t) = b_2\, e^{-k_{YA}\,t} + k_{Y\bar{Y}0} \sum_{i=1}^{3} \frac{a_i\,(\mathbf{r}_i)_2}{k_{YA} + \xi_i}\, e^{\xi_i\,t}\,, \tag{5.1.12}$$

$$\bar{Z}\,(t) = b_3\, e^{-k_{ZA}\,t} + k_{Z\bar{Z}0} \sum_{i=1}^{3} \frac{a_i\,(\mathbf{r}_i)_3}{k_{ZA} + \xi_i}\, e^{\xi_i\,t}\,, \tag{5.1.13}$$

with b_1, b_2, b_3 being real constants, which are determined by the initial conditions on $\bar{X}$, $\bar{Y}$, $\bar{Z}$, and $(\mathbf{r}_i)_j$ is the jth component of eigenvector $\mathbf{r}_i$.

Equation (5.1.7) can be solved by integration with respect to t, giving

$$A(t) = \sum_{i=1}^{3} d_i\, e^{\xi_i t} + d_4\, e^{-k_{XA}\, t} + d_5\, e^{-k_{YA}\, t} + d_6\, e^{-k_{ZA}\, t} + d_7\,, \tag{5.1.14}$$

with $d_1, \ldots, d_7$ being real constants.

If all of the real parts of the eigenvalues ξ_i are negative, we observe exponential decay of X, Y, Z, $\bar{X}$, $\bar{Y}$, $\bar{Z}$ for large t and A increases to the steady state solution d_7. If Re $\xi_i > 0$ for some i, the system grows exponentially for sufficiently large values of $t > 0$. Examples of exponential growth and exponential decay in the state variables of the system are given in Figure 5.4. Oscillations (in the case of complex eigenvalues) decline for increasing t. The complex eigenvalues in the solutions in both examples have relatively large negative real parts and hence oscillations vanish rapidly over time.

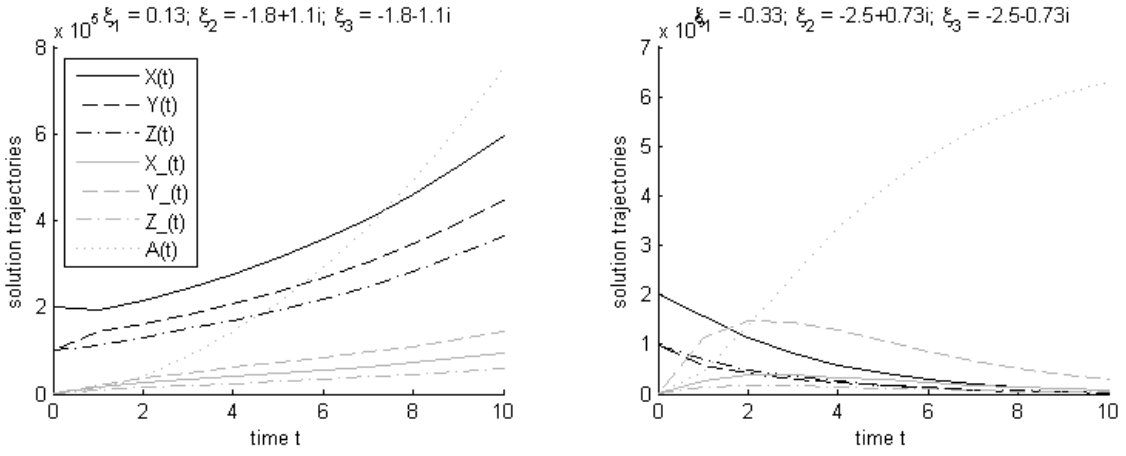

Figure 5.4: Simulations for two sets of parameter values in system (5.1.1)-(5.1.7), displaying exponential growth (left) and exponential decay (right). The corresponding parameter values are $k_{XY} = k_{YZ} = k_{ZX} = 1$, $k_{XA} = k_{YA} = k_{ZA} = 0.5$ and $X_0 = 2 \cdot 10^5$, $Y_0 = Z_0 = 10^5$, $\bar{X}_0 = \bar{Y}_0 = \bar{Z}_0 = A_0 = 0$ for both plots, and $k_{X\bar{X}0} = k_{Z\bar{Z}0} = 0.1$, $k_{Y\bar{Y}0} = 0.2$ for the left plot and $k_{X\bar{X}0} = k_{Z\bar{Z}0} = 0.2$, $k_{Y\bar{Y}0} = 2$ for the right plot.

It is possible that the system of cycling cells (X, Y, Z) tends to an equilibrium state, that is, with at least one eigenvalue being zero and the remaining eigenvalues having negative real parts. By inspecting (5.1.11)-(5.1.13), or solving (5.1.4)-(5.1.6), we then find that $\bar{X}$, $\bar{Y}$, $\bar{Z}$ tend either to a constant (for $k_{*A} > 0$, with (*) denoting respectively X, Y, Z), or to a linear function (for $k_{*A} = 0$), and A is asymptotic to a linear function. Examples for

both types of dynamic behaviour are given in Figure 5.5, where suitable parameter values have been chosen by assuring that the constant term in the characteristic polynomial for ξ_1, ξ_2, ξ_3 is zero.

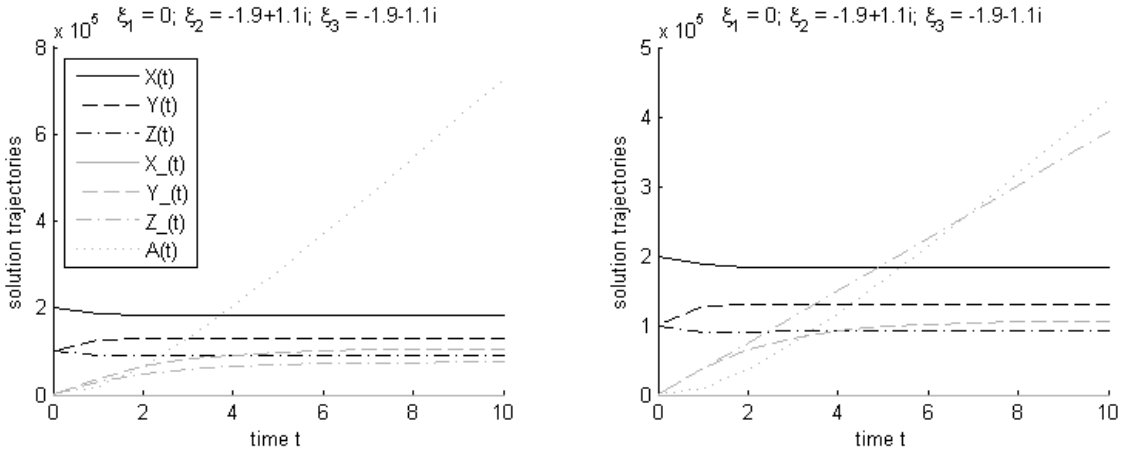

Figure 5.5: Simulations for two sets of parameter values in system (5.1.1)-(5.1.7), displaying asymptotically constant behaviour (left) and asymptotically linear growth (right) of the state variable $\bar{Z}(t)$. The corresponding parameter values are $k_{XY} = k_{YZ} = k_{ZX} = 1$, $k_{X\bar{X}0} = 0$, $k_{Y\bar{Y}0} = k_{Z\bar{Z}0} = -1 + \sqrt{2}$, $k_{XA} = 0$, $k_{YA} = 0.5$, $X_0 = 2 \cdot 10^5$, $Y_0 = Z_0 = 10^5$ and $\bar{X}_0 = \bar{Y}_0 = \bar{Z}_0 = A_0 = 0$ in both simulations. The parameter k_{ZA} varies in the two plots, where $k_{ZA} = 0.5$ in the left plot and $k_{ZA} = 0$ in the right plot.

In general, however, we expect cell division and growth of the population to dominate the dynamic behaviour of the system. A quantity of biological interest is the doubling time, T_d, of cells, which is the period of time required for the population of cells to double in number. It is a characteristic unit for the description of cell growth providing a more intuitive notion of the long-term impact of growth. In the following, therefore, we aim to determine the time T_d it takes for a population to double their number of viable cells, that is $V = X + Y + Z$.

To derive a mathematical expression for T_d, we firstly consider a simplified version (Figure 5.6) of the cell cycle model which can be described by the two equations

$$\frac{dV}{dt} = (\mu - \delta)\, V\,, \qquad \frac{d\bar{V}}{dt} = \delta V\,. \tag{5.1.15}$$

The expression $\mu - \delta$ in equation (5.1.15) is derived from cells producing a new cell in the V compartment with rate μ and cells leaving the V compartment with rate δ to the $\bar{V}$ compartment of dead cells. Initial conditions are $V(0) = V_0 > 0$ and $\bar{V}(0) = \bar{V}_0 \geq 0$. The corresponding solution for cycling cells is $V(t) = V_0\, e^{(\mu-\delta)t}$, providing us with

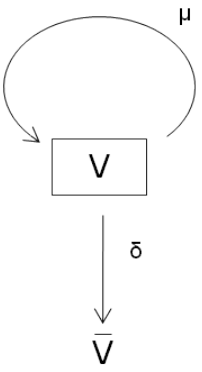

Figure 5.6: A simple two-compartment model of cells cycling in the V compartment with doubling rate μ and cells exiting the cell cycle to the $\bar{V}$ compartment with rate δ.

a formula for the doubling time, namely $T_d = \ln(2)/(\mu - \delta)$, by solving the equation $2\,V(t) = V(t + T_d)$.

We now approximate the doubling time T_d for the full model by considering the subsystem of equations (5.1.1)-(5.1.3) representing the cell-cycle dynamics of viable cells illustrated in Figure 5.1. The eigenvalues ξ_j of the coefficient matrix of (5.1.8) determine the behaviour of X, Y and Z, and if the largest eigenvalue, say ξ_s, is positive, we expect exponential growth in the total number of cycling cells. Through solving $2\,(X(t) + Y(t) + Z(t)) = X(t + T_d) + Y(t + T_d) + Z(t + T_d)$ for large values of t and assuming constant transition rates k, we approximate the doubling time by

$$T_d = \frac{\ln 2}{\xi_s}\,. \tag{5.1.16}$$

For exponential decay ($\xi_s < 0$), the absolute value of T_d gives the half-life of the population of cycling cells, though the total number of cells, including apoptotic material, A, will approach a finite positive constant.

5.2 Markov-chain model and deduction of ODE model

In this Section, we aim to investigate the relationship between a probabilistic model of the cell-cycle dynamics and the deterministic model in Section 5.1. When modelling the cell cycle we assume that it is a continuous-time Markov process with a countable state space

(see Wilkinson [147] for a review of stochastic processes), that is, a stochastic process $\{S(t) : t \geq 0\}$ with $S(t) \in \mathbb{N}_0^m$ for all $t \geq 0$, where the compartments of S represent the number of cells in each phase of the cell cycle, and where, given the *current* state of the system, the *past* behaviour of the system does not influence the time-evolution of the process. Thus, we have, in terms of probabilities,

$$\mathrm{P}\big(S(t+\Delta t) = s' \;\big|\; \{S(\tau) = s(\tau) \mid \tau \in [0,t]\}\big) = \mathrm{P}\big(S(t+\Delta t) = s' \;\big|\; S(t) = s(t)\big)\,, \quad (5.2.1)$$

for all $t \in [0,\infty)$ and $s' \in \mathbb{N}_0$, with Δt being an infinitesimally short time interval.

A Markov process is said to be time-homogeneous if the transition kernel

$$p(s,t,s',t') = \mathrm{P}\left(S(t+t') = s' \,|\, S(t) = s\right)\,, \quad (5.2.2)$$

is not dependent on t, that is $p(s,t,s',t') = p(s,s',t') \;\; \forall\, t$. Hence, the initial distribution for $S(0)$ together with the transition kernel $p(s,s',t')$ determine the probability distribution for the state at all future times. The derivatives

$$q(s,s') = \frac{\partial p(s,s',t')}{\partial t'}\Big|_{t'=0}\,, \quad (5.2.3)$$

give the rate of moving from state s to state s'. The quantity $p(s,s',\Delta t) = q(s,s')\,\Delta t$ is then the probability of a transition from state s to state s' in the infinitesimally short time interval $(t, t+\Delta t)$ conditional on the state being s at time t.

In our cell cycle model, we described the rates of a cell making a transition from one phase to another by constants k. We want to show that the probabilistic model can be taken as a basis for the mass balance equations (5.1.1)-(5.1.7) with constant transition rates, interpreting the state variables in the ODE model as the expected cell numbers of a Markov process. We consider only a sub-system $S = (X,Y,Z)$ of the cell cycle model M_0 and later on focus only on state changes in the X compartment of the model, as the equivalence between the probabilistic and the deterministic model can be shown in a similar way for the remaining state variables.

The probability of a state change in S in the next infinitesimal time interval Δt can be

described by the master equation (a brief summary of which is given in Gillespie [47]) as

$$\begin{aligned}
\mathrm{P}\left(X,Y,Z;\ t+\Delta t\right) = & \left[\mathrm{P}\left(X+1,Y-1,Z;t\right)k_{XY} + \mathrm{P}\left(X+1,Y,Z;t\right)k_{X\bar{X}0}\right]\ (X+1)\,\Delta t \\
& + \left[\mathrm{P}\left(X,Y+1,Z-1;t\right)k_{YZ} + \mathrm{P}\left(X,Y+1,Z;t\right)k_{Y\bar{Y}0}\right]\ (Y+1)\,\Delta t \\
& + \left[\mathrm{P}\left(X-2,Y,Z+1;t\right)k_{ZX} + \mathrm{P}\left(X,Y,Z+1;t\right)k_{Z\bar{Z}0}\right]\ (Z+1)\,\Delta t \\
& + \mathrm{P}\left(X,Y,Z;t\right)\left[1-\left(k_{XY}\,X + k_{X\bar{X}0}\,X + k_{YZ}\,Y + k_{Y\bar{Y}0}\,Y + k_{ZX}Z + k_{Z\bar{Z}0}\,Z\right)\Delta t\right] .
\end{aligned} \tag{5.2.4}$$

Each term in the sum on the right hand side of equation (5.2.4) corresponds to a probability that, at time t, the system (X, Y, Z) is in a certain state s and either remains in that state in $(t, t+\Delta t)$, or alters its state due to a transition occurring in $(t, t+\Delta t)$. Note that $k_{XY}\,\Delta t$, for instance, describes the probability of a transition for one particular cell from X to Y in the next time interval Δt, and it is multiplied by the number of cells, X, to give the probability for any such transition to occur. Summing over all possible values of Y and Z in (5.2.4) yields

$$\begin{aligned}
\mathrm{P}\left(X;t+\Delta t\right) \ = \ & \mathrm{P}\left(X;t\right)\left(1-\left(k_{XY}\,X + k_{X\bar{X}0}\,X + k_{ZX}\,\mathbb{E}(Z|X)\,(t)\right)\right)\Delta t \\
& + \left(\mathrm{P}\left(X+1;t\right)k_{XY} + \mathrm{P}\left(X+1;t\right)k_{X\bar{X}0}\right)\ (X+1)\,\Delta t \\
& + \mathrm{P}\left(X-2;t\right)k_{ZX}\,\mathbb{E}(Z|X-2)\,(t)\ \Delta t\ ,
\end{aligned} \tag{5.2.5}$$

using implications from basic probability theory such as $\sum_y y\,\mathrm{P}\left(X, Y=y\right) = \mathrm{P}\left(X\right)\mathbb{E}(Y|X)$ from the definition of the conditional expectation $\mathbb{E}(Y|X)$. Rearranging (5.2.5) and letting $\Delta t \to 0$, we obtain

$$\begin{aligned}
\frac{\partial}{\partial t}\mathrm{P}\left(X;t\right) \ = \ & -\mathrm{P}\left(X;t\right)\left(\left(k_{XY}+k_{X\bar{X}0}\right)X + k_{ZX}\,\mathbb{E}(Z|X)(t)\right) \\
& + \mathrm{P}\left(X+1;t\right)\left(k_{XY}+k_{X\bar{X}0}\right)(X+1) + \mathrm{P}\left(X-2;t\right)k_{ZX}\,\mathbb{E}(Z|X-2)(t)\ .
\end{aligned} \tag{5.2.6}$$

If we multiply both sides of equation (5.2.6) by X and sum over all values of X, it is easy

to show that

$$\frac{\partial}{\partial t}\,\mathbb{E}\,(X)\,(t) = -\,(k_{XY} + k_{X\bar{X}0})\,\mathbb{E}\,(X)\,(t) + 2\,k_{ZX}\,\mathbb{E}(Z)(t)\,, \tag{5.2.7}$$

demonstrating the relationship between the probabilistic and the deterministic model of the cell cycle dynamics.

Provided cell numbers are sufficiently large, we can therefore introduce the probabilistically-motivated notion of the residence time (see Chapter 2.4) as the expected value of the waiting time, in our deterministic model, and set

$$T_X = \frac{1}{k_{XY}}, \quad T_Y = \frac{1}{k_{YZ}}, \quad T_Z = \frac{1}{k_{ZX}}\,, \tag{5.2.8}$$

as the residence times of cells in the X, Y and Z compartments. For non-negligible loss rates k_{*0} from any of the three compartments to cell death, we have to correct our formulas for the residence time, and set

$$T_X = \frac{1}{k_{XY} + k_{X\bar{X}0}}, \quad T_X = \frac{1}{k_{YZ} + k_{Y\bar{Y}0}}, \quad T_Z = \frac{1}{k_{ZX} + k_{Z\bar{Z}0}}\,. \tag{5.2.8*}$$

5.3 Statistical model of experimental data

To fit the cell cycle model of Section 5.1 to the experimental data collected in Chapter 4, we need to describe the data mathematically, which we do using a statistical model. The cell-cycle dynamics are modelled by the system of ODEs (5.1.1)-(5.1.7), of the form

$$\frac{\mathrm{d}\mathbf{v}\,(t,\mathbf{p})}{\mathrm{d}t} = \mathbf{A}\,(t,\boldsymbol{\theta})\,\mathbf{v}\,(t,\mathbf{p})\,, \qquad 0 < t < t_e\,, \tag{5.3.1}$$

$$\mathbf{v}\,(0,\mathbf{p}) = \mathbf{v}_0\,, \tag{5.3.2}$$

where $\mathbf{v}$ is the $m = 7$-dimensional vector of state variables and t denotes the time in the interval $[0, t_e]$. The parameter vector $\mathbf{p} = (\boldsymbol{\theta}, \mathbf{v}_0)$ with domain $\boldsymbol{\Theta} \subseteq \mathbb{R}^L$ is an $L = 20$-dimensional vector of $l = 13$ unknown rate parameters $\boldsymbol{\theta}_k$, $k = 1, \ldots, l = 13$, and $m = 7$ initial conditions $(\mathbf{v}_0)_j$, $j = 1, \ldots, m = 7$, ($L = l + m = 20$) and $\mathbf{A}$ is an 7×7 matrix

with entries that depend on $\boldsymbol{\theta}$ and possibly t. These numbers are for the example system, model M_2 (see Figure 5.3), where we have $t_e = 10$, $\mathbf{v} = (X, Y, Z, \bar{X}, \bar{Y}, \bar{Z}, A)^T$ and

$$\boldsymbol{\theta} = (k_{XY}, k_{YZ}, k_{ZX}, k_{X\bar{X}0}, \Delta k_{X\bar{X}}, k_{Y\bar{Y}0}, \Delta k_{Y\bar{Y}}, k_{Z\bar{Z}0}, \Delta k_{Z\bar{Z}}, t_0, k_{XA}, k_{YA}, k_{ZA}) \,. \tag{5.3.3}$$

Whilst this model has $m = 7$ quantities $(X, Y, Z, \bar{X}, \bar{Y}, \bar{Z}, A)$, only $M = 5$ independent quantities are measured, respectively $w_1 = X + \bar{X}$, $w_2 = Y + \bar{Y}$, $w_3 = Z + \bar{Z}$, $w_4 = X + Y + Z$ and $w_5 = A$. Hence the vector $\mathbf{w}$ of measurements obtained is a linear combination of $\mathbf{v}$, this is, $\mathbf{w} = \mathbf{B}\,\mathbf{v}$ where

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} . \tag{5.3.4}$$

In our model the observables, w_j, are observed with noise. Therefore, we model w_j as random variables W_j. Our statistical model for $W_j(t_i)$, $j = 1, \ldots, M$, $i = 1, \ldots, n$, the jth observable at ith time point, is

$$\ln W_j(t_i) = \ln\left(\mathbf{B}\,\mathbf{v}(t_i, \mathbf{p})\right)_j + \varepsilon_{ij} \,, \tag{5.3.5}$$

where ε_{ij} are the measurement errors, which we assume are independently and identically distributed as $\varepsilon_{ij} \sim \mathrm{N}\left(0, \sigma^2\right)$ with variance σ^2. The probability distribution of $W_j(t_i)$ is hence log-normal and is described by the probability density function

$$g\left(w_{ij} | \boldsymbol{\alpha}\right) = \frac{1}{w_{ij}\,\sigma\,\sqrt{2\,\pi}} \exp\left\{ -\frac{\left(\ln\left(w_{ij}\right) - \ln\left(\mathbf{B}\,\mathbf{v}(t_i, \mathbf{p})\right)_j\right)^2}{2\,\sigma^2} \right\} , \tag{5.3.6}$$

which is a function of the variable w_{ij} with the $(L+1)$-dimensional parameter $\boldsymbol{\alpha} = (\mathbf{p}, \sigma)$ fixed, we require $w_{ij} > 0$ and $\left(\mathbf{B}\,\mathbf{v}(t_i, \mathbf{p})\right)_j > 0$ for (5.3.6) to be well-defined.

As the experimental data, $w_j(t_i)$, used for model fitting have positive values throughout

the experiments, and come from measurements which combine error sources by multiplication of errors rather than addition (see (4.1.2) for deriving normalised data), there is good reason to assume that the random variables $W_j(t_i)$ follow a log-normal distribution [73]. This choice of error distribution also seems appropriate since the data stem from an autocatalytic growth process, that is, a replication process, in which Gaussian variation in the growth rates will generate a log-normal error structure in the obtained cell numbers [80]. The assumption of a log-normal distribution is furthermore affirmed by the results of normality testing of the model residuals in Chapter 6.4.

5.4 Structural identifiability analysis – application of transfer function method

Before estimating parameters it is useful first to check whether the model (5.1.1)-(5.1.7) is identifiable. An identifiable model is one for which the unknown parameters of the model, our rates k, can be uniquely recovered from observed data under ideal conditions, that is, assuming we have error-free and continuous data for all observables available (see Chapter 3.1 for a definition of structural identifiability).

We apply the transfer function method from Chapter 3.3 to our model and use (3.3.3) with $\mathbf{B}$ as in (5.3.4),

$$\mathbf{A} = \begin{pmatrix} -k_{XY} - k_{X\bar{X}0} & 0 & 2k_{ZX} & 0 & 0 & 0 & 0 \\ k_{XY} & -k_{YZ} - k_{Y\bar{Y}0} & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{YZ} & -k_{ZX} - k_{Z\bar{Z}0} & 0 & 0 & 0 & 0 \\ k_{X\bar{X}0} & 0 & 0 & -k_{XA} & 0 & 0 & 0 \\ 0 & k_{Y\bar{Y}0} & 0 & 0 & -k_{YA} & 0 & 0 \\ 0 & 0 & k_{Z\bar{Z}0} & 0 & 0 & -k_{ZA} & 0 \\ 0 & 0 & 0 & k_{XA} & k_{YA} & k_{ZA} & 0 \end{pmatrix}, \tag{5.4.1}$$

and $\mathbf{v}_0 = (X_0, Y_0, Z_0, \bar{X}_0, \bar{Y}_0, \bar{Z}_0, A_0)^T$ to obtain the exhaustive summary (3.3.5) of the model, containing 43 equations in $\mathbf{p}$. For the calculations, we use the computational software program MATHEMATICA, the code of which is shown in Appendix A.2. In order

to determine which of the unknown parameters

$$\boldsymbol{\theta} = (k_{XY}, k_{YZ}, k_{ZX}, k_{X\bar{X}0}, k_{Y\bar{Y}0}, k_{Z\bar{Z}0}, k_{XA}, k_{YA}, k_{ZA}) \ , \tag{5.4.2}$$

are uniquely determined by (3.3.5), we pick nine of the 43 equations which are linear in the parameters $\boldsymbol{\theta}$ (that is the equations for the observable parameters $\Phi_{1,i}$ and $\Phi_{m+2,i}$) and formulate them as a system of equations $\mathbf{Q}\,\boldsymbol{\theta} = \boldsymbol{\phi}$, where $\mathbf{Q} \in \mathbb{R}^{9\times 9}$ is a matrix whose entries may contain initial conditions $\mathbf{v}_0$, and $\boldsymbol{\phi} \in \mathbb{R}^9$ is a vector of observational parameters. Matrix reduction of $\mathbf{Q}$ by elementary row operations yields that $k_{XY}, k_{YZ}, k_{ZX}, k_{XA}, k_{YA}, k_{ZA}$ can be uniquely determined by this system. Hence, we substitute the solutions for the identifiable parameters and consider the exhaustive summary now as a system of equations only in $(k_{X\bar{X}0}, k_{Y\bar{Y}0}, k_{Z\bar{Z}0})$. By again picking appropriate equations from (3.3.5), we obtain a system with coefficient matrix

$$\begin{pmatrix} k_{YZ}\,k_{ZX} & k_{XY}\,k_{ZX} & k_{XY}\,k_{YZ} \\ k_{YZ}+k_{ZX} & k_{XY}+k_{ZX} & k_{XY}+k_{YZ} \\ 1 & 1 & 1 \end{pmatrix} , \tag{5.4.3}$$

which has non-zero determinant if $(k_{XY}-k_{YZ})\,(k_{XY}-k_{ZX})\,(k_{YZ}-k_{ZX}) \neq 0$. Consequently, all model parameters can be uniquely identified and we have shown that the model is globally identifiable, that is, at least in principle, the parameter values can be determined from observable quantities.

5.5 Parameter estimation techniques

The parameters to be estimated in our model (5.3.5) are a maximum of $l = 13$ rate parameters $\boldsymbol{\theta}_k$ and $m = 7$ initial conditions $(\mathbf{v}_0)_j$ whose values are not known exactly. We use a least squares approach for determining optimal parameter values. The approach is equivalent to maximising the likelihood function under model (5.3.5), that is, maximising the probability that, for a given parameter vector $\boldsymbol{\alpha} = (\mathbf{p}, \sigma)$, certain realisations $w_j\,(t_i)$ of the random variables $W_j\,(t_i)$ occur, interpreted as a function of $\boldsymbol{\alpha}$. The sum of the squared

residuals is

$$f(\mathbf{p}) = \sum_{i=1}^{n} \sum_{j=1}^{M} \mathbf{e}_{i,j}^2(\mathbf{p}) \ , \tag{5.5.1}$$

where

$$\mathbf{e}_{i,j}(\mathbf{p}) = \ln \ (\mathbf{B}\,\mathbf{v}(t_i, \mathbf{p}))_j - \ln w_j(t_i) \ . \tag{5.5.2}$$

The least-squares parameter estimate is

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p} \in \boldsymbol{\Theta}} f(\mathbf{p}) \ , \tag{5.5.3}$$

which represents a nonlinear optimisation problem and is dealt with in the following paragraphs.

In order to minimize the cost function f in (5.5.1) we need an efficient optimisation technique that searches over all the parameter space $\boldsymbol{\Theta}$ and converges quickly to the global minimum. An ODE solver (ode45) in MATLAB is used for the numerical integration of (5.1.1)-(5.1.7) to obtain $\mathbf{v}(t, \mathbf{p})$. Local optimisation methods usually converge quickly but do so only if the initial guess is sufficiently close to the global minimum, otherwise, they are easily trapped in a local minimum. In contrast, global optimisation algorithms offer a more promising route to find the global minimum even if there is no proof of convergence in general and they require more time to find the optimum precisely.

Ashyraliyev et al. [8] therefore propose a hybrid method that uses the Stochastic Ranking Evolutionary Strategy (SRES) for a global search followed by the Levenberg-Marquardt (LM) method for a local search which starts with the results from the global optimisation SRES as an initial guess. For local optimisation in the case of least-square fits, the LM method [84] – a combination of a steepest-descent method with the fast-converging Gauss-Newton method – is recommended by Ashyraliyev et al. [9].

SRES [116] is a strategy inspired by biological evolution, which uses the idea of treating parameter sets as the 'genome' of an individual and a procedure of selection, recombination and mutation is repeated over G generations to find an optimal parameter set. Selected global optimisation methods were tested for the parameter estimation of a large ODE model by Moles et al. [88] and SRES presented the best convergence rates.

We also compared the SRES method to another standard global optimisation algorithm, Simulated annealing[1], and to MATLAB's built-in function fminbnd (which is a global optimisation technique that combines the golden-section search and parabolic interpolation techniques). We found that SRES performed far faster and more successfully than both of these other methods, when applied to the cell-cycle models.

We performed calculations in MATLAB using the implementation of Kleinstein et al. [71] with $G = 500$ generations for the SRES algorithm, which is run in a transformed space to avoid stagnation or misconvergence due to sampling from a uniform distribution, and using the function lsqnonlin for the LM optimisation, where we implement parameter constraints $\mathbf{p} \in \mathbf{\Theta}$ by adding penalties to the objective function ensuring the constraints are strictly obeyed in each step of the algorithm. An example code for fitting model M_2^{**} (a reduced form of model M_2, see Section 6.1 for details) to data from treatment with 50 nM of RHPS4 is presented in Appendix B.

In summary, the SRES + LM method implementation is an efficient approach for the parameter estimation of the cell-cycle model and is easy to handle once set up. As SRES is a stochastic optimisation algorithm, several runs should be evaluated to determine the global minimum, an example will be given in Chapter 6.1.

5.6 Model selection and evaluation

Having found best fits to experimental data for each of the models M_0, M_1 and M_2, the most appropriate model for each of the given drug concentrations is to be selected. A danger in choosing a model which is meant to describe the observations from certain experiments, is overfitting the experimental data by introducing many parameters. We prefer to choose the model with the smallest number of parameters, which still describes

[1]Simulated annealing is an optimisation method associated with the physical process of annealing [70]. Starting from a random initial state, a parameter set within the parameter constraints is randomly generated within a certain range of the previous state and accepted or rejected according to the Metropolis criterion: in the case of an improvement of the cost function $\Delta f \leq 0$, the move will be accepted, otherwise it will be accepted with probability $p(\Delta f) = e^{-\Delta f / T_k}$, with T_k being the current 'temperature' of the system at step k. At the beginning, T_k is set to be high implying that all states are equally likely to be accepted. The temperature T_k is slowly decreased with each new accepted state until the improvement of f within a chosen number of steps is smaller than a preset value ϵ. The slow cooling mimics the physical method of the controlled cooling of a melt with the aim of reaching its minimum energy state, which corresponds to the global minimum of the cost function.

the underlying dynamics sufficiently well. The Akaike information criterion [3] (AIC) is an information-theoretic criterion for model comparison, which incorporates not only the objective function $f(\hat{\mathbf{p}})$ but also a penalty based on the number of parameters in the model. Hence it characterises the trade-off between goodness of fit and model complexity. When comparing two models according to the Akaike criterion, they must be 'nested' in the sense that the parameter space of one model is a lower dimensional subspace of the other model. The value of the Akaike criterion is given by

$$\mu_{\mathrm{AIC}} = N \ln\left(f(\hat{\mathbf{p}})\right) + 2(L+1)\,, \tag{5.6.1}$$

where N is the number of experimental measurements, f is the sum of the squared residuals and L is the number of parameters which have been fitted to the data (L is the sum of the number of rate parameters, l, and the number of initial conditions, m) in each model. The correction term

$$\mu_{c\mathrm{AIC}} = \mu_{\mathrm{AIC}} + \frac{2\,(L+1)\,(L+2)}{N-L-2}\,, \tag{5.6.2}$$

[18] should be used for smaller sample sizes of $N \leq 40(L+1)$. The candidate model with the *smallest* value μ_{cAIC} is the selected model.

We can now evaluate how accurately our selected model reflects the properties of the collected data by employing statistical significance tests. In writing down our model (5.3.5), we have made several assumptions, namely that the model is a valid description of the underlying biological processes, that errors are normal on the log scale, that the errors are independent and have the same standard deviation σ. Under these assumptions, the residuals $\widehat{\mathbf{e}}_{i,j}\left(\mathbf{p}\right) = \mathbf{e}_{i,j}\left(\mathbf{p}\right)/\sigma$ are i.i.d. random variables each with a $\mathrm{N}\left(0,1\right)$ distribution.

To test the model residuals $\widehat{\mathbf{e}}_{i,j}$ for standard normality, that is, to test the null hypothesis H_0: $\widehat{\mathbf{e}}_{i,j} \sim \mathrm{N}\left(0,1\right)$, we use the Anderson-Darling Statistic

$$A^2 = -\frac{1}{N}\sum_{i=1}^{N}(2\,i-1)\left[\ln(z_i) + \ln(1-z_{N+1-i})\right] - N\,, \tag{5.6.3}$$

with $z_l = F\left(\widehat{e}_{(l)}\right)$, $l = 1,\ldots,N = n \times M$; this is one of the most powerful tests of H_0 [131]. Here, F is the cumulative standard normal distribution function and $\widehat{e}_{(1)} \leq \widehat{e}_{(2)} \leq$

$\ldots \leq \widehat{e}_{(N)}$ are the weighted residuals $\widehat{\mathbf{e}}_{i,j}$ in ascending order. We reject the null hypothesis at a chosen significance level α (with α being the probability of incorrectly rejecting H_0) if A^2 is larger than a critical value $\delta_{A^2}(\alpha)$. For $A^2 \leq \delta_{A^2}(\alpha)$, we have no evidence against the null hypothesis, which is supportive for the model choice, but in general a weaker argument than rejecting H_0. The value $\delta_{A^2}(\alpha)$ is derived from the distribution function of A^2 under H_0 ("null distribution"). Asymptotic results for the null distribution of A^2 can be found in Stephens [132].

We now test the residuals $\mathbf{e}_{i,j}$ for independence. Dependence between residuals is likely to be found as correlation between $\mathbf{e}_{:j}$ and $\mathbf{e}_{:k}$ for $j, k \in \{1, \ldots, M\}$, and as autocorrelation over the time points of observation, that is $\mathbf{e}_{(i+1),j}$ being correlated with $\mathbf{e}_{i,j}$. This could be checked by tests of the correlation coefficients for each pair $(\mathbf{e}_{:j}, \mathbf{e}_{:k})$ as in Kraemer [75] and by autocorrelation tests for each $\mathbf{e}_{:j}$ as in Cedersund and Roll [27]. For very small ($i \leq 10$) sample sizes one should additionally assess if the assumption of independence for the model errors holds by plotting the residuals.

In the following Chapter 6, the concepts and techniques for model inference and evaluation presented in this Chapter will be combined with our experimental results through fitting our cell cycle model to data. The best model-to-data fit will be selected and the model residuals will be statistically evaluated and analysed with respect to the accuracy of fit.

CHAPTER 6

Fitting model to data

To investigate the effects of RHPS4 on cell-cycle dynamics, we collected experimental data from colorectal cancer cells incubated in control (no drug) and three different concentrations of RHPS4 for periods of 10 days as described in Chapter 4. We tried to capture the underlying mechanisms using compartmental models in Chapter 5. We have observed marked cell death setting in after a time delay in treated cells, which we introduced into our models using transition rates from viable to dead cells which vary with time. To evaluate our model and infer more information about the effects of RHPS4 particularly with respect to its cell-cycle specificity and cell death processes, we now compare model predictions to our experimental data using the techniques of model inference and residual testing from Chapters 5.5 and 5.6. We also investigate the accuracy of our parameter fitting results by applying the techniques of practical parameter identifiability outlined in Chapter 3.4.

Results from fitting our models to experimental data are presented in Section 6.1, where the best model describing the data is determined and model reductions are performed to achieve a minimal realisation of the system. In Section 6.2, we present the results from parameter fitting and describe the model dynamics for each of the different concentrations of RHPS4. Information on parameter sensitivities generated by our fitting procedure is summarised in Section 6.3, and an analysis of the accuracy of model-to-data fit is given in Section 6.4. Section 6.5 investigates the biological implications of our results. Section 6.6 discusses the results and concludes this Chapter.

6.1 Results of best fit

We used three submodels, M_0, M_1 and M_2 (see Figure 5.3), each containing a maximum of 13 parameters, to describe the cell-cycle dynamics of control cells and cells treated with RHPS4. To find the best fit to experimental data, we inferred parameters using a combination of global and local optimisation techniques that we introduced in Chapter 5.5, and selected the best model according to the Akaike criterion [3].

As we do not observe the state variables of our model (5.3.1) directly, we do not have knowledge of the initial values of the system. Estimating the initial conditions $\mathbf{v}_0$ directly from experimental data did not notably influence the best fit and resulted in a higher accuracy for the estimated rate parameters (more than a 75% reduction in the standard error, see Chapter 3.4 for estimating standard errors). We chose estimates of initial values $\mathbf{v}_0$ according to experimental data at day 1, as cells take about a day to attach to the flask before resuming cell growth, and we assume that the proportions of viable cells in each of the cell-cycle phases G_0/G_1, S, G_2/M do not differ from the proportion of viable cells across the total cell population at day 1.

Applying the optimisation methods from Chapter 5.5, we found that the parameter estimates, $\hat{\mathbf{p}}$, of our models converged rapidly to the global minimum during the first $G = 100$ generations, and in every case of the 30 repeated runs of the SRES algorithm, based each time on a different random initial guess for $\hat{\mathbf{p}}$, the final objective function values were found to be close to the optimal objective function value, see Figure 6.1 for plots of residuals, f, against generation number. The parameter values obtained from the repeated runs of the SRES+LM routine shown in Figure 6.2 illustrate the case of 50 nM RHPS4 and demonstrate that the convergence of the method is stable and a unique minimum can be identified.

We then calculated the corrected Akaike criterion μ_{cAIC} (see Chapter 5.6) to select between models M_0, M_1 and M_2, for each drug concentration. Table 6.1 shows the values of μ_{cAIC}, as well as the values of the cost function f and the variance estimates $\hat{\sigma}^2$, suggesting that model M_0 describes best the behaviour of control cells and model M_2 the dynamical behaviour of treated cells. The parameter β controlling the shape of the transitions $k_*(t)$ to

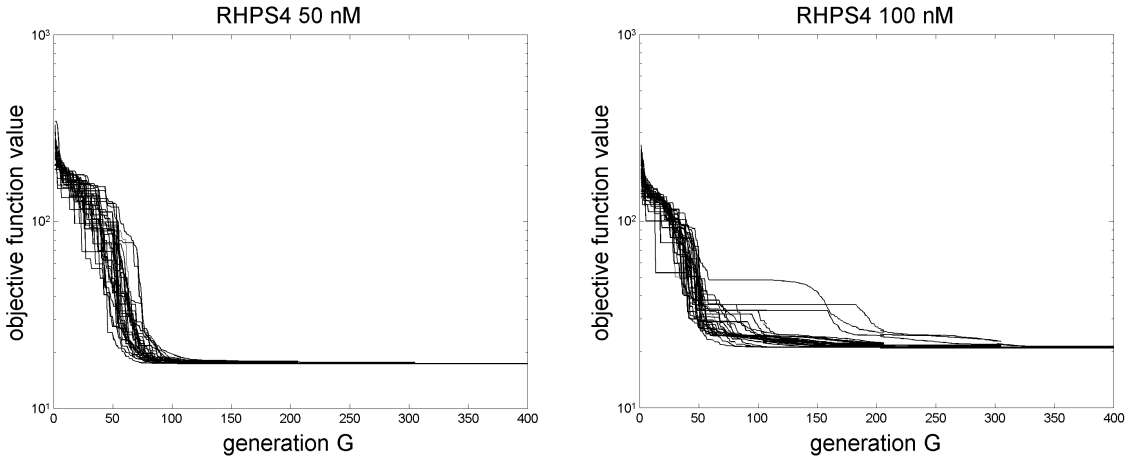

Figure 6.1: The objective function values $f(\mathbf{p})$ are plotted against generation number G for several repeats of the optimisation routine with a logarithmic scale for the ordinate. The results of fitting model M_2 to data from HCT116 cells treated with 50 nM and 100 nM of RHPS4 are shown. The algorithm converges especially rapidly during the first 100 generations, and 30 repeats have been run over 500 generations to identify the global minimum.

cell death for model M_2 (sigmoidal rate increase) was not part of the model fit as similar values of β resulted in equally good fits and additional fitting of β slowed down the optimisation process. A value of $\beta = 10$ was used for model fitting as smaller values of β resulted in worse fits. In particular, values of $\beta \leq 3$ did not fit the data showing that the drug introduces not gradual but abrupt changes in the number of viable cells. Note that, for control cells, the model with a higher cost function value but a smaller number of parameters has been chosen. For increasing drug concentration, the data variance estimators $\hat{\sigma}^2$ increase from about 0.2 (control) to 0.5 (1 μM RHPS4) indicating that more variability occurs in the data at higher doses of RHPS4. Model M_1 performs worse in terms of $f_{\min}$ than the selected models for all experimental data.

We found very low parameter values for $k_{Y\bar{Y}0}$ and $k_{Z\bar{Z}0}$ in control cells, which causes parameter identifiability problems for k_{YA} and k_{ZA}, as the choice of those parameter values, with $\bar{Y}$, $\bar{Z}$ being practically zero, did not influence the model dynamics significantly. This also resulted in large standard errors in the respective parameters. Therefore, we refitted model M_0 with only 5 parameters, setting $k_{Y\bar{Y}0} = k_{Z\bar{Z}0} = k_{YA} = k_{ZA} = 0$. The choice of the reduced model, denoted by M_0^*, resulted in similar parameter fits (see Table 6.2), with a smaller value of μ_{cAIC} (Table 6.1) compared to model M_0 due to the presence of fewer parameters in M_0^*, confirming the model reduction.

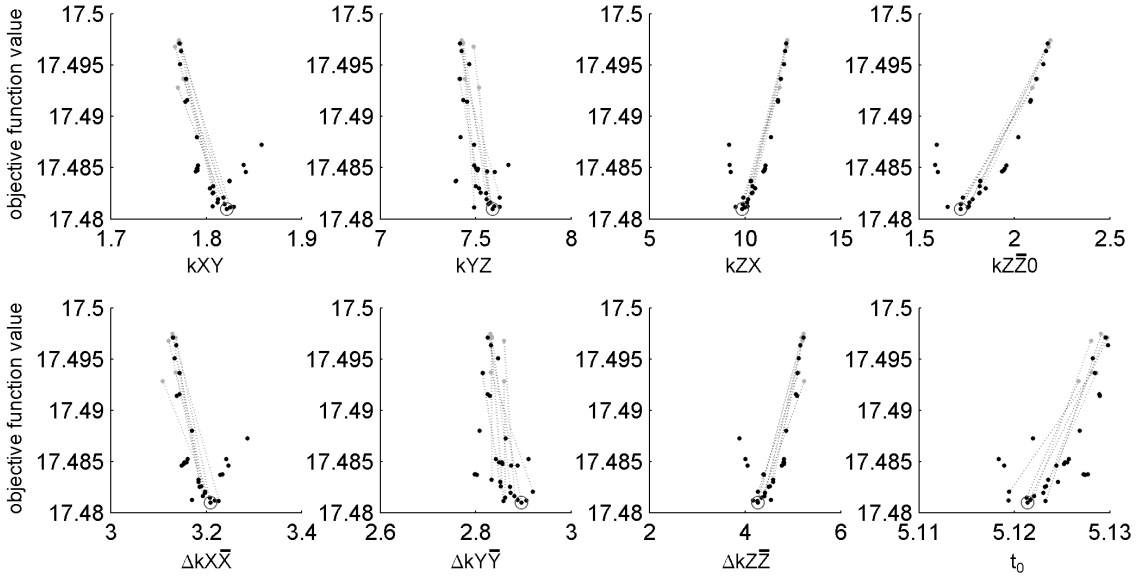

Figure 6.2: Results of fitting model M_2^{**} to experimental data from treatment with 50 nM of RHPS4 are shown with the parameter values being given in units of 1/day. Optimal parameter values from the global (SRES, grey dots) and local optimisation routine (LM, black dots) for 30 runs are plotted with their optimal function value for a choice of 8 parameters, each run being represented by a dotted line connecting the initial guess (optimal value from SRES) and the optimal value from the LM routine. The parameter set that resulted in the overall lowest cost function value has been chosen and is circled in this Figure. The optimal parameters of the 30 fits do not differ more than 0.5/day from the parameter value with the lowest cost function value for each of the different parameters, only the convergence of k_{ZX} and $\Delta k_{Z\bar{Z}}$ is less stable with a spectrum of about ± 2.5/day and ± 2/day, respectively.

RHPS4	**model type**	$f_{\min}$	$\hat{\sigma}^2$	μ_{cAIC}	model chosen
0 nM	M_0	8.1856	0.189	138.64	
	M_1	8.0979	0.198	151.72	
	M_2	7.9739	0.194	150.88	
	M_0^*	8.2152	0.168	127.51	✔
50 nM	M_0	118.29	2.319	310.88	
	M_1	20.240	0.436	217.79	
	M_2	17.481	0.372	209.00	
	M_2^{**}	17.481	0.343	196.16	✔
100 nM	M_0	50.604	0.992	259.93	
	M_1	28.033	0.596	237.34	
	M_2	20.985	0.446	219.96	
	M_2^*	20.985	0.411	207.12	✔
1 μM	M_0	36.423	0.714	240.20	
	M_1	31.121	0.662	243.61	
	M_2	26.038	0.554	232.91	
	M_2^*	26.038	0.511	220.06	✔

Table 6.1: The optimal objective function values $f_{\min} = f(\hat{\mathbf{p}})$ and variance estimates $\hat{\sigma}^2$ for the model residuals are given for models M_0, M_1 and M_2 and each concentration of the drug RHPS4, and the model fits are compared with respect to their values μ_{cAIC} of the corrected Akaike criterion. Data of model M_0^* is given for 0 nM, of model M_2^* for 100 nM and 1 μM, and of model M_2^{**} for 50 nM RHPS4. The 'best' model is chosen according to the lowest value of μ_{cAIC}.

Similarly, we reduced model M_2 by setting $k_{Y\bar{Y}0} = k_{Z\bar{Z}0} = k_{YA} = k_{ZA} = 0$ for cells treated with higher drug concentrations, producing model M_2^*; and setting $k_{X\bar{X}0} = k_{Y\bar{Y}0} = k_{XA} = k_{YA} = 0$ for treatment with 50 nM RHPS4 (model M_2^{**}). This is suggested since all these parameters k_{*0} satisfy $k < 10^{-6}$/day, with parameters k_{*0} that have not been removed taking values larger than 10^{-1}/day. Furthermore, lower μ_{cAIC} values for model M_2^* and model M_2^{**} than for model M_2 supported our choice of model reduction.

We find good agreement between model predictions of the selected models and experimental data, especially for control and 50 nM of RHPS4. The simulated and observed data are shown for each concentration and each observed quantity in Figure 6.3. For higher drug concentrations, the simulations overestimate the number of S phase cells at day 4. However, there is more noise in the data with increasing concentration of RHPS4, and the model still captures the major trend of the cell cycle dynamics well.

6.2 Parameter fitting - model results

To gain insight into the cell cycle dynamics predicted by the fitted models, we investigate the solution behaviour of the state variables of the selected models. We compare the estimated parameter values and effects on physical properties of the system, such as the doubling time T_d, defined by (5.1.16), and the residence times T_X, T_Y, T_Z of cells, defined by (5.2.8*), in the X, Y, Z compartments, respectively, for no drug and each of the three concentrations of RHPS4.

Simulations of the dynamical behaviour of viable cells confirm that control cells grow exponentially, with the number of G_0/G_1 cells being larger than the number of G_2/M cells and the G_2/M cell numbers being slightly larger than the number of S-phase cells across the observation period. The doubling time of $T_d = 19.8$ h simulated for control cells is within biological variability of the value $T_d = 20.5$ h quoted by Brattain et al. [25] for colorectal cancer HCT116 cells. Cells die from the G_0/G_1 phase, the number of non-viable S-phase and G_2/M-phase cells remain at a constant, very low level (see left panel of Figure 6.4). Note that t has been shifted one unit in the diagrams in accordance with the first measurements, which are our initial conditions, being on day 1 of the experiments,

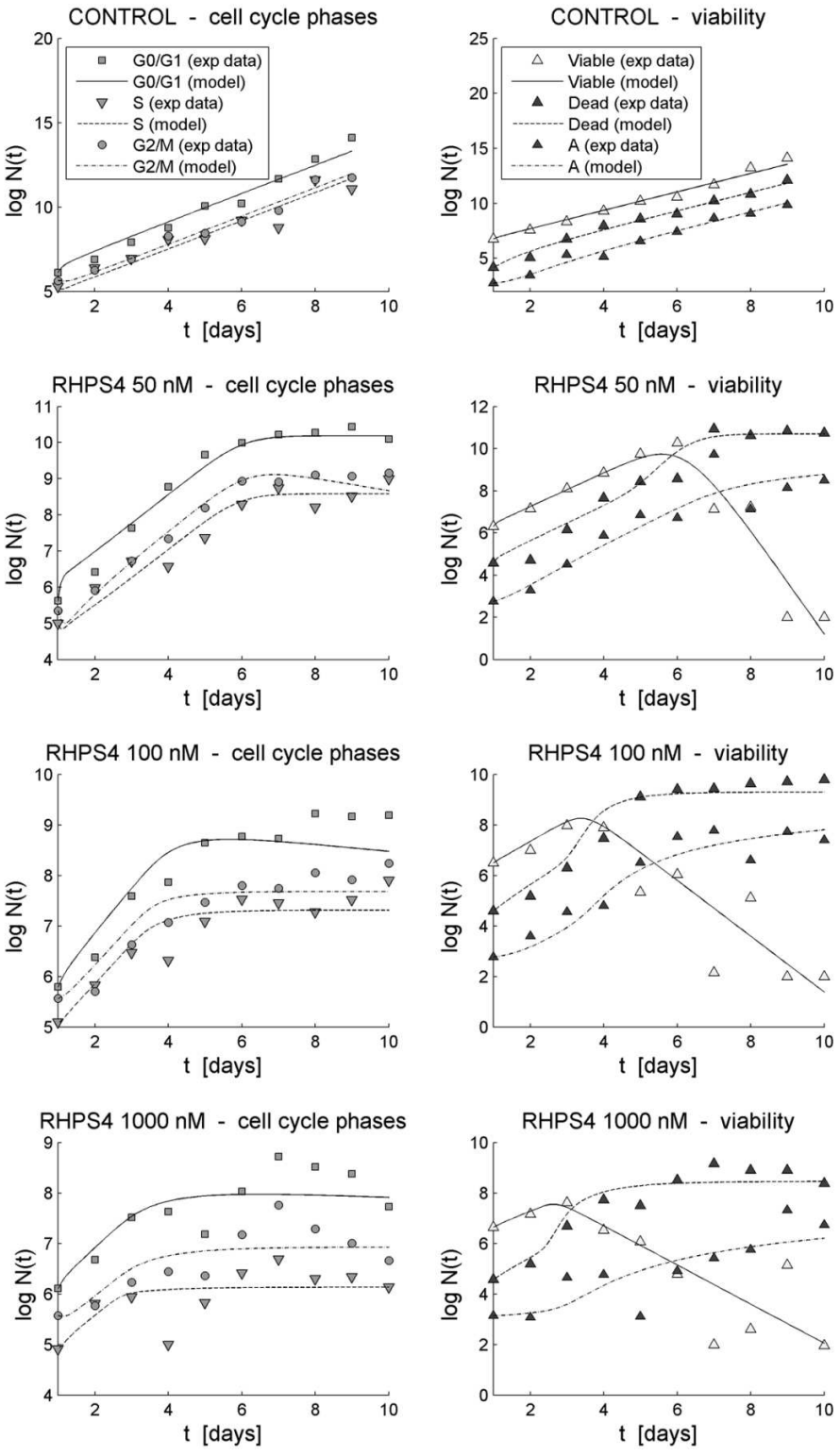

Figure 6.3: Solutions of the model fitting procedure to experimental data. Simulations stem from the 'best' models chosen in Table 6.1. Model curves are represented by lines and experimental data by different markers dependent on the states observed and logarithmic scales have been taken on the vertical axes. 'A cells' denotes pre-G1 cells and is a subset of nonviable cells. Standard deviations of the experimental data are not shown in order not to impair the visibility of the markers, since error bars would be largely smaller than the symbols.

as this is when we assume cell growth to start since cells take about a day to attach to the flask after seeding and to resume growth.

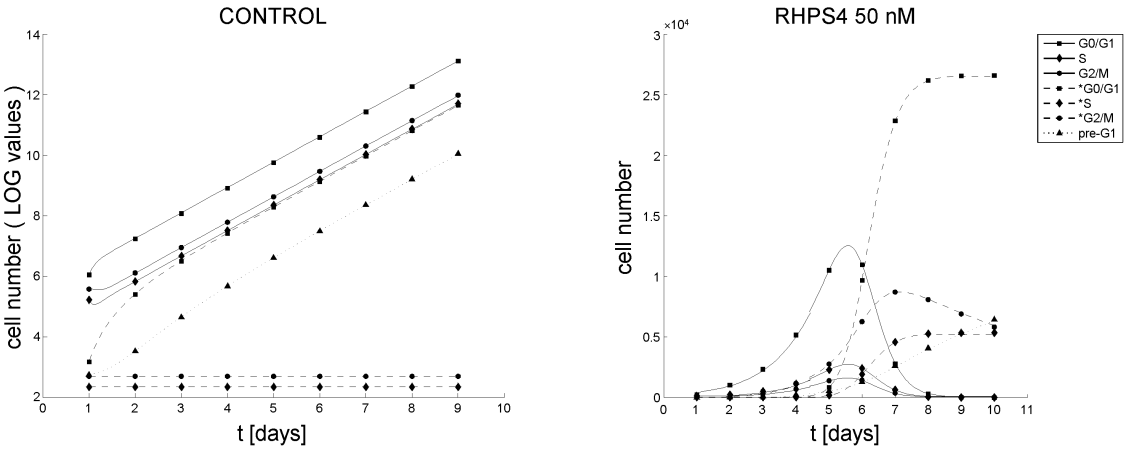

Figure 6.4: Cell cycle dynamics of model (5.1.1)-(5.1.7) for the behaviour of control cells over 9 days (left: log-plot for better visual distinction between the different phases of the cell cycle), and for the behaviour of cells treated with 50 nM of RHPS4 over 10 days (right: normal scale on vertical axis). The asterisk in front of phase names in the legend denotes dead cells.

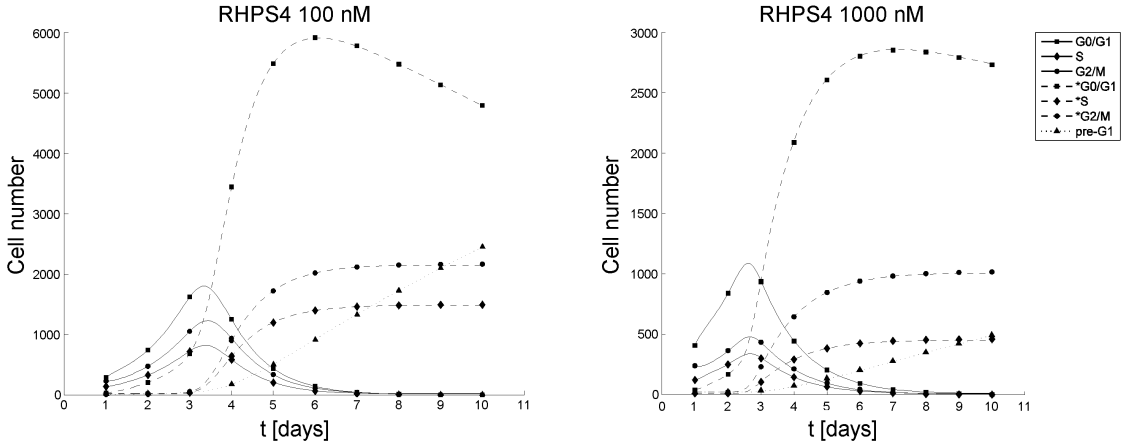

Figure 6.5: Cell cycle dynamics of model (5.1.1)-(5.1.7) for the behaviour of cells treated with 100 nM (1 μM) of RHPS4 are shown on the left (right) side of this Figure. Simulations stem from model fitting to experimental data. The asterisk in front of phase names in the legend denotes dead cells. Note different scales on vertical axes.

When the drug is added to the cells, the cell cycle dynamics change markedly. The number of viable cells in each phase of the cell cycle decays exponentially after an initially exponential increase and almost vanishes at the end of the observation period (see right panel of Figure 6.4 for 50 nM and Figure 6.5 for 100 nM and 1 μM RHPS4). Whereas the exponential growth rates, ξ_s, do not differ much across treatments for the exponential increase ($\xi_s \in [0.6/\text{day}, 0.9/\text{day}]$, setting $k_*(t) = k_{*0}$ in (5.1.8) with (*) representing

$X\bar{X}$, $Y\bar{Y}$, or $Z\bar{Z}$), these growth rates are negative and decline markedly with increasing drug concentration ($\xi_s = -2.2458$/day for 50 nM, $\xi_s = -0.98828$/day for 100 nM, $\xi_s = -0.68232$/day for 1 μM RHPS4, setting $k_*(t) = k_{*0} + \Delta k_*$ in (5.1.8)) for the exponential decrease. Model analysis revealed that we find oscillations in the state variables for each cell-cycle phase and each drug concentration around the beginning ($t = 1$) of the observation period (two eigenvalues are complex conjugates with negative real parts), hence their contributions to the cell numbers vanish rapidly as t increases. The peak of the number of viable cells is shifted towards earlier times at increasing drug concentration. The number of pre-G_1 cells grows exponentially for control cells and, due to the number of viable cells ceasing to proliferate, linearly in the second half of the observation period for treated cells. For the lowest drug concentration of 50 nM, however, the model predicts a higher number of viable S-phase cells than viable G_2/M-phase cells across the observation period contrary to respective trends for control cells, 100 nM and 1 μM treatments.

Table 6.2 displays all estimated rate parameter values and some derivations such as doubling times for each concentration of RHPS4. Control cells spend about 10.6 h in the G_0/G_1 phase, 3.2 h in the S phase and 5.0 h in the G_2/M phase according to the fitted parameters. The rates of transition between compartments of viable cells and the average amounts of time T_X, T_Y, T_Z a cell spends in each of the cell cycle phases G_0/G_1, S, G_2/M, do not display a trend with respect to changes in drug concentration. Notable changes predicted by the model, however, are in T_Z for 50 nM, being less than half the length of the corresponding estimates for control cells, and in T_Z for 100 nM and 1 μM, being about 1.5 times this length. There is also a slight increase of about 1.5 h in the predicted residence time of cells in the S phase for increasing drug concentration. The time point, t_0, around which the transition rates to cell death increase by Δk_*, decreases from 5.12 days to 1.68 days with increasing drug concentration, where the decrease is more significant for the lower drug concentrations.

There is no visible trend in the model of cells dying before the time point t_0. Control cells and cells treated with higher drug concentrations die with lower rates ($k_{X\bar{X}0} < 0.4$/day) from the G_0/G_1 phase; cells treated with 50 nM die markedly ($k_{Z\bar{Z}0} = 1.72$/day) from

RHPS4	(0 nM)	0 nM	50 nM	100 nM	1 μM
model	M_0	M_0^*	M_2^{**}	M_2^*	M_2^*
T_X	10.6	10.6	13.2	7.18	12.4
T_Y	3.21	3.22	3.16	4.07	4.70
T_Z	5.02	5.01	2.08	7.37	8.23
T_d	19.8	19.8	20.8	20.1	27.3
k_{XY}	2.03	2.03	1.82	2.98	1.72
k_{YZ}	7.47	7.45	7.59	5.90	5.10
k_{ZX}	4.78	4.79	9.84	3.26	2.92
t_0	-	-	5.12	2.46	1.68
$k_{X\bar{X}0}$	2.36×10^{-1}	2.35×10^{-1}	0	3.61×10^{-1}	2.11×10^{-1}
$k_{Y\bar{Y}0}$	4.87×10^{-17}	0	0	0	0
$k_{Z\bar{Z}0}$	3.28×10^{-16}	0	1.72	0	0
$k_{X\bar{X}0} + \Delta k_{X\bar{X}}$	2.36×10^{-1}	2.35×10^{-1}	3.21	3.13	1.86
$k_{Y\bar{Y}0} + \Delta k_{Y\bar{Y}}$	-	-	2.90	1.58	8.88×10^{-1}
$k_{Z\bar{Z}0} + \Delta k_{Z\bar{Z}}$	-	-	5.98	1.44	1.36
k_{XA}	1.63×10^{-1}	1.70×10^{-1}	0	7.05×10^{-2}	2.56×10^{-2}
k_{YA}	1.20×10^{-14}	0	0	0	0
k_{ZA}	4.54×10^{-1}	0	1.71×10^{-1}	0	0

Table 6.2: Results of parameter estimation. The transition rates k are displayed in units of 1/day for all concentrations of the drug RHPS4, together with respective residence times T_X, T_Y, T_Z of cells in the G_0/G_1, S, G_2/M phase, as defined by (5.2.8*), and doubling times T_d (before the time of marked cell death) of viable cells, as defined by (5.1.16), given in units of hours. Cells die with rates $k_{*0} + \Delta k_*$ after the time point t_0. The presented values stem from the 'best' models chosen in Table 6.1. Parameter values for model M_0 and model M_0^* are given for comparison.

the G_2/M phase. There is practically no death from the S phase before t_0. After the time point t_0, treated cells die with significantly higher rates from *all* phases of the cell cycle. In particular, cells treated with 50 nM RHPS4 undergo cell death from G_2/M with a markedly higher rate ($k_{Z\bar{Z}0} + \Delta k_{Z\bar{Z}} = 5.98/\text{day}$) than from the other cell cycle phases. DNA degradation in cells occurs largely from the phases of the cell cycle in which cells die before the time point t_0, that is, $\bar{X} \to A$ when $X \to \bar{X}$ before t_0 (control, 100 nM and 1 μM RHPS4), and $\bar{Z} \to A$ when $Z \to \bar{Z}$ before t_0 (50 nM RHPS4).

To analyse the data in Table 6.2 further, we approximate the percentages of live cells in each compartment X, Y, Z, and the percentages of which compartments X, Y, Z cells die from at times $t > t_0$ by considering the largest eigenvalue, ξ_s, and the corresponding eigenvector, $\mathbf{r}_s = (r_{s,1}, r_{s,2}, r_{s,3})^T$, of the coefficient matrix of (5.1.8) for each concentration of RHPS4. We use the scaled components $r_{s,i}/(r_{s,1} + r_{s,2} + r_{s,3})$, $i = 1, 2, 3$, of the eigenvector $\mathbf{r}_s$ to approximate the percentages of live cells in the X, Y, Z compartments, respectively. Furthermore, we approximate the percentages of which compartments X, Y, Z cells die from at each dose of RHPS4 respectively by $p_{\bar{X}}\, r_{s,1}/\bar{r}$, $p_{\bar{Y}}\, r_{s,2}/\bar{r}$, $p_{\bar{Z}}\, r_{s,3}/\bar{r}$, where

$$p_{\bar{X}} = \frac{k_{X\bar{X}0} + \Delta k_{X\bar{X}}}{k_{XY} + k_{X\bar{X}0} + \Delta k_{X\bar{X}}}\,, \quad p_{\bar{Y}} = \frac{k_{Y\bar{Y}0} + \Delta k_{Y\bar{Y}}}{k_{YZ} + k_{Y\bar{Y}0} + \Delta k_{Y\bar{Y}}}\,, \quad p_{\bar{Z}} = \frac{k_{Z\bar{Z}0} + \Delta k_{Z\bar{Z}}}{k_{ZX} + k_{Z\bar{Z}0} + \Delta k_{Z\bar{Z}}}\,, \tag{6.2.1}$$

are the death probabilities for cells in each compartment and

$$\bar{r} \;=\; p_{\bar{X}}\, r_{s,1} + p_{\bar{Y}}\, r_{s,2} + p_{\bar{Z}}\, r_{s,3}\,, \tag{6.2.2}$$

is a scaling factor.

When comparing cell cycle proportions of viable cells treated with RHPS4 to those of control cells, we find that, whereas S phase proportions are slightly higher (2-6%) when the drug is added to cell cultures, there is a drop of about 12% in the proportion of viable G_2/M phase cells for 50 nM of RHPS4, and an average increase of about 10% for higher concentrations of RHPS4. Correspondingly, G_0/G_1 phase proportions increase by around 10% for 50 nM and decrease by an average of 14% for 100 nM and 1 μM of RHPS4. Table 6.3 shows the percentages of live cells in each model compartment. In contrast to control

cells, which die primarily from the G_0/G_1 phases, cells treated with 50 nM of RHPS4 die at low percentages (7-8%) additionally from the S and G_2/M phases. For 100 nM and 1 μM of RHPS4, cell death from the S phase occurs at low proportions of about 7-12%, cell death from the G_2/M phases, however, increases to about 21-28%. The proportions of which phase of the cell cycle cells die from at different drug concentrations are given in Table 6.4.

Phase / Dose	G_0/G_1	S	G_2/M
	X	Y	Y
Control	64	16	21
50 nM	74	17	9
100 nM	45	21	34
1 μM	55	18	27

Table 6.3: Percentages of live cells in each model compartment.

Phase / Dose	G_0/G_1	S	G_2/M
	X	Y	Z
Control	100	0	0
50 nM	85	8	7
100 nM	60	12	28
1 μM	72	7	21

Table 6.4: Percentages of which phases of the cell cycle cells die from at each dose of RHPS4 at times $t > t_0$.

Altogether, our models predict exponential growth with a doubling time of a little less than a day for control cells, which is in agreement with data from the literature, and we find exponential growth is followed by exponential decay which occurs after a time delay for treated cells, where a larger delay corresponds to a lower drug concentration and *vice versa*. Cell death occurs primarily from the G_2/M phase for a dose of 50 nM RHPS4 before the time point t_0 of increased cell death, in contrast to higher drug concentrations, where cells are more likely to die from the G_0/G_1 phase. At times $t > t_0$, however, around 10% of treated cells die also from the S phase and cell death from the G_2/M phases switches on in particular for cells treated with higher concentrations of RHPS4, with more than 20% of cells dying from this phase. Hence, RHPS4 seems to particularly affect the G_2/M phase inducing cell death throughout the treatment period for 50 nM of RHPS4, and after a delay of about 2-3 days for the higher drug concentrations of 100 nM and 1 μM.

6.3 Sensitivity analysis

It is important to investigate how reliable and accurate model predictions are before drawing conclusions from the model output. We perform a detailed sensitivity analysis on all model parameters in this Section and consider how sensitive the model dynamics are to small changes in the estimated parameters. Thus, we identify which parameters are important in contributing to the prediction of the output variables.

Parameter sensitivity is a measure of how a small variation in each parameter around the optimal parameter values $\hat{\mathbf{p}}$ influence the model behaviour. To investigate the sensitivity of our model parameters $\mathbf{p}$, we define the time-dependent dimensionless sensitivity coefficients

$$S\left(\mathbf{v}_j, \mathbf{p}_k, t\right) = \frac{\mathbf{p}_k}{\mathbf{v}_j(t, \mathbf{p})} \frac{\partial \mathbf{v}_j(t, \mathbf{p})}{\partial \mathbf{p}_k} = \frac{\partial \log \mathbf{v}_j(t, \mathbf{p})}{\partial \log \mathbf{p}_k}, \tag{6.3.1}$$

for the kth parameter on the jth variable in our model, with $\mathbf{v}_j$ denoting the jth component of the vector of state variables $\mathbf{v} = (X, Y, Z, \bar{X}, \bar{Y}, \bar{Z}, A)$. The partial derivatives $\partial \mathbf{v}_j(t, \mathbf{p}) / \partial \mathbf{p}_k$ can be computed by solving the equations described in Appendix C.1. A zero value of S indicates that the respective model variable is insensitive to changes in the parameter value, and a value of 1 indicates that a change in $\mathbf{p}_k$ causes the same relative change in $\mathbf{v}_j$.

We have computed values of sensitivity coefficients according to (6.3.1) for the initial conditions $\mathbf{v}_0$ and the rate parameters k in our model. At the time point $t = 1$, each of the initial conditions $(\mathbf{v}_0)_i$ has sensitivity 1 on $\mathbf{v}_j$ with $j = i$ and sensitivity 0 on $\mathbf{v}_j$ for all $j \neq i$. In general, the sensitivities of the initial conditions are positive and low, with $S < 0.5$ for the sensitivities of each of the state variables on X_0, Y_0, Z_0 at times $t > 2$, vanishing sensitivities of $\bar{X}$, $\bar{Y}$, $\bar{Z}$ on $\bar{X}_0$, $\bar{Y}_0$, $\bar{Z}_0$ as well as A on $\bar{X}_0$, $\bar{Y}_0$, $\bar{Z}_0$, A_0 as $t \to \infty$ ($S < 0.1$ for $t > 5$, except for the sensitivities of A on A_0 at 1 μM RHPS4 where $S < 0.1$ for $t > 6.5$) and $S = 0$ for the sensitivities of X, Y, Z on $\bar{X}_0$, $\bar{Y}_0$, $\bar{Z}_0$, A_0, and $\bar{X}$, $\bar{Y}$, $\bar{Z}$ on A_0.

All rate parameters have sensitivity 0 at $t = 1$ as we assume initial conditions for day 1 to be independent of transition rate parameters. For control cells, the rate parameter k_{XY} has the greatest influence on each of the components of the system (data not shown); and t_0 has the greatest affect on model dynamics for cells with 50 nM RHPS4. We find

that changes in t_0 cause disproportionally larger changes in the state variables X, Y, Z for $t > 5$, furthermore an increase in t_0 evokes a marked decrease in $\bar{X}$ and $\bar{Y}$ at earlier times ($4 \leq t \leq 6$) and a significant increase at later times ($t \geq 7$). The effects on $\bar{Z}$ and A are of the same type but less marked; Figure 6.6 shows selected sensitivity plots. Determining t_0 to a good level of accuracy is therefore crucial in accurately simulating model behaviour for model M_2^{**}. For increasing drug concentrations, the effect of t_0 declines, but changes in t_0 cause changes in other variables at earlier times ($t = 3$ for 100 nM, $t = 2$ for 1 μM).

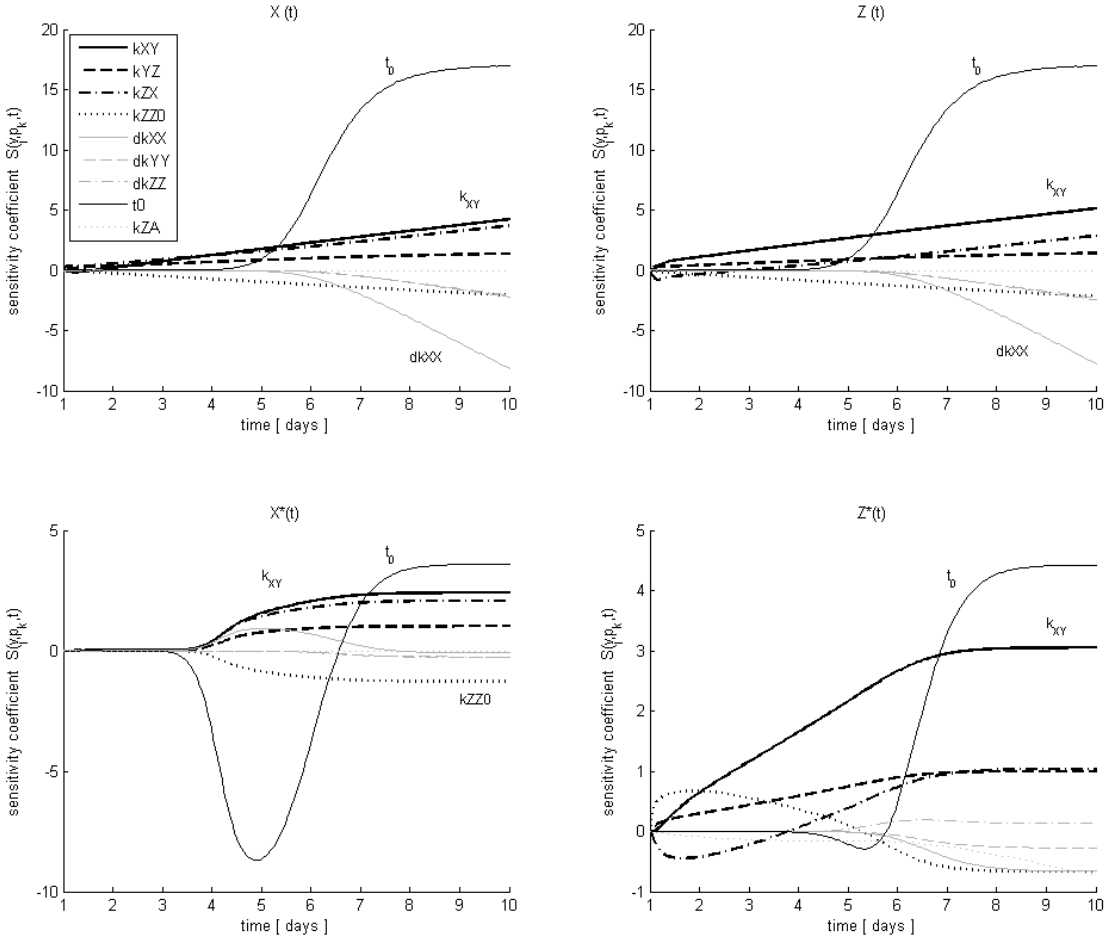

Figure 6.6: Selected sensitivity plots for model M_2^{**} (where $t_0 \approx 5$) and results of fitting data from treatment with 50 nM RHPS4. The plots show the sensitivity coefficients for all 9 model parameters and numbers of viable and dead cells in G_0/G_1 and G_2/M (in this Figure: X, $X^* = \bar{X}$, Z, $Z^* = \bar{Z}$, respectively). Overall, the parameters t_0, and then k_{XY}, have the greatest effect on cell cycle dynamics.

Overall, the effects of changes in parameter values become more visible the longer cells are kept in culture. An increase in rate parameters of cycling cells has, in general, a positive effect on cell cycle compartments. Depending on the model chosen, however, one of the rate parameters has positive and negative effects on those compartments of

dead cells that are most affected by the drug. For 50 nM RHPS4, for example, an increase in k_{ZX} causes a decrease of cell numbers in $\bar{Z}$ (and A) at times $t \leq 4$ and an increase that is nearly proportional to k_{ZX} at times $t \geq 6$. The final increase of non-viable cell numbers is caused by the source of viable cells growing with increasing k_{ZX}. The rate parameters Δk_* influences the states of viable cells rather strongly and the states of non-viable cells less strongly ($|S| < 2$). The effect of k_{ZA} is rather small except for the compartment of pre-G_1 cells (data not shown).

We find that the sensitivity curves of each of the state variables are approximately symmetric for k_{ZX} and $k_{Z\bar{Z}0}$ with respect to the time axis, hence the effects of both parameters on model dynamics nearly cancel, indicating the possibility of a strong correlation between the respective parameters for 50 nM RHPS4. The correlation between parameters will be studied in greater detail by investigating the correlation matrix in the next Section.

6.4 Accuracy of fit

We can evaluate how accurately our model predictions reflect the statistical assumptions we made about the experimental data by using the statistical hypothesis tests introduced in Chapter 5.6. To analyse the uncertainty in the estimated parameters $\hat{\mathbf{p}}$, we estimate the standard errors and investigate the correlation matrix of $\hat{\mathbf{p}}$ as well as parameter confidence regions as described in Chapter 3.4.

In writing down our model (5.3.5), we made several assumptions, namely that the model is a valid description of the underlying biological processes, and that errors are independent and log-normally distributed with the same standard deviation σ. Under the model (5.3.5) the model residuals $\mathbf{e}_{i,j}$ are independent with a $\mathrm{N}\left(0, \sigma^2\right)$ distribution (see Chapter 5.3). We employ statistical tests to check the model residuals for normality and independence. Using the Anderson-Darling Statistic [131, 132] we found no evidence at a 5% significance level against the hypothesis that $\mathbf{e}_{i,j} \sim \mathrm{N}\left(0, \sigma^2\right)$. Furthermore, using a t-statistic [75] we found no evidence against the assumption that residuals are independent. The level of correlation in residuals for control and at 50 nM RHPS4 is very low ($p > 0.1$ for most pairs of residuals), indicating that our models capture the dynamics

well. We also have not found any significant autocorrelation [27], that is, $\mathbf{e}_{i+1,j}$ is not correlated with $\mathbf{e}_{i,j}$, $i = 1, \ldots, n-1$, over the time points of observation in the residuals. Most of the p-values were larger than 0.5, none of them being smaller than 0.15. Figure 6.7 displays the residuals $\mathbf{e}_{i,j}$ for the case of 50 nM of RHPS4. For higher drug concentrations, we find statistically significant ($p < 0.05$) correlation between residuals from $X + \bar{X}$, $Y + \bar{Y}$, $Z + \bar{Z}$, $\bar{X} + \bar{Y} + \bar{Z} + A$, indicating that there is some systematic variability in the data that the model does not capture. The near-linear relationships between residuals may be caused by secondary dynamics in nonviable cells, which are not accounted for in our model.

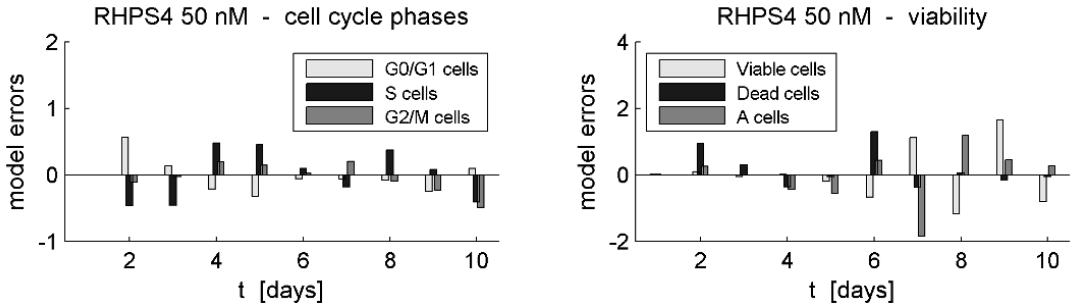

Figure 6.7: Residuals $\mathbf{e}_{i,j}$ for time points t_i, $i = 1, \ldots, 10$, and model categories $j = 1, \ldots, 6$, are plotted on the time scale for model M_2^{**} and experimental data from treatment with 50 nM of RHPS4. Using the Anderson-Darling Statistic, there is no evidence at a 5% level against the null hypothesis that the residuals are sampled from a normal distribution.

An illustration of the estimates of the rate parameters for the four RHPS4 assays including standard errors is given in Figure 6.8. Standard errors correspond to local analysis around $\hat{\mathbf{p}}$ and, therefore, large error bars indicate low accuracy in a neighbourhood of the optimal parameter. Consequently, error bars do not indicate the full range of possible parameter values for the model to be a good prediction of the cell cycle dynamics.

The degree of correlation between the estimates $\hat{\mathbf{p}}$ of the rate constants k was evaluated for each concentration of RHPS4 by computing the correlation matrix of $\hat{\mathbf{p}}$ (given for 50 nM of RHPS4 in Appendix C.2). For the control case and higher drug concentrations, we found only weak correlation between parameter estimates ($|r_{ij}(\hat{\mathbf{p}})| < 0.8$ and $|r_{ij}(\hat{\mathbf{p}})| < 0.9$ for all i, j, respectively). There is, however, a positive correlation between the estimates of k_{ZX} and $k_{Z\bar{Z}0}$ ($r_{34}(\hat{\mathbf{p}}) = 0.96$) for treated cells with 50 nM RHPS4, meaning that, with parameter values being at the best fit, changes in k_{ZX} and $k_{Z\bar{Z}0}$ proportional to each other do not significantly influence the model predictions in a small neighbour-

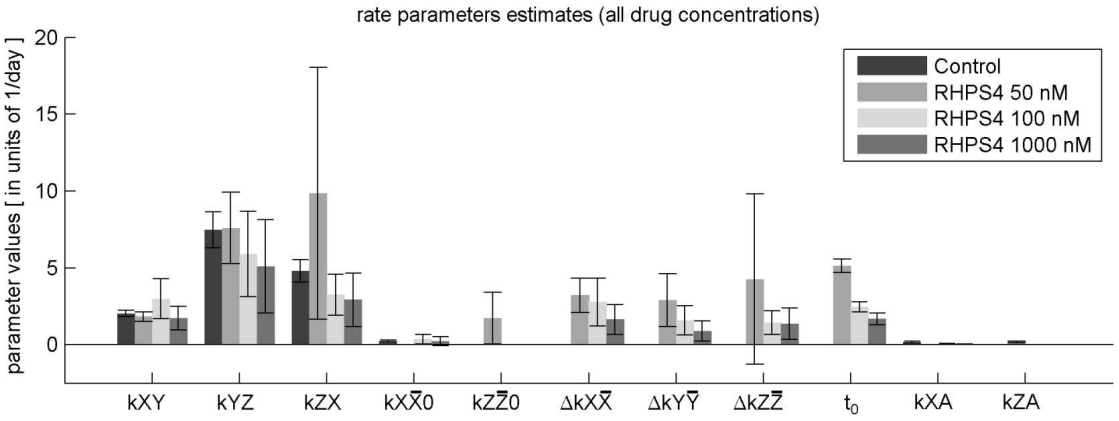

Figure 6.8: Parameter estimates for control cells and treatment of cells with 50 nM, 100 nM and 1000 nM RHPS4 from fitting model M_0^* (control), model M_2^* and model M_2^{**} (treated cells). Standard errors of parameter estimates result from model fit to experimental data and are shown by error bars.

hood of the estimated parameter values. Hence, good predictions for these parameters lie within a narrow ellipsoidal region in the k_{ZX}-$k_{Z\bar{Z}0}$-plane, indicated in Figure 6.9 (left plot), whose orientation and length of principal axes can be estimated from analysis of the eigenvectors and eigenvalues of the Fisher information matrix as described in Chapter 3.4. The parameter values can be slightly lower or higher but must approximately satisfy $(k_{ZX} - 9.84) \approx 5.05\,(k_{Z\bar{Z}0} - 1.72)$. The Tables of all eigenvalues and corresponding eigenvectors are given in the Appendix C.3. Lower accuracy and strong correlation between the estimates k_{ZX} and $k_{Z\bar{Z}0}$ therefore suggests that the effects of the drug on these parameters may be less dramatic at 50 nM RHPS4 than originally inferred from model fitting.

The error bar for $\Delta k_{Z\bar{Z}}$ is rather large for 50 nM RHPS4 corresponding to a relatively flat region in parameter space along $\Delta k_{Z\bar{Z}}$. Whereas the estimates of k_{ZX} and $k_{Z\bar{Z}0}$ are strongly correlated, the confidence region in the k_{ZX}-$\Delta k_{Z\bar{Z}}$-plane, for example, is a rather broad ellipsoid as shown in Figure 6.9, which corresponds to a weaker correlation ($r_{37}\,(\hat{\mathbf{p}}) = 0.66$).

In summary, for no drug and 50 nM of RHPS4, our model predictions capture the experimental data particularly well and reflect the statistical assumptions we made about the data. Apart from the parameters k_{ZX}, $k_{Z\bar{Z}0}$ and $\Delta k_{Z\bar{Z}}$ that were estimated with lower accuracy, most of the model parameters were estimated with good accuracy given the general biological variability existing in living systems.

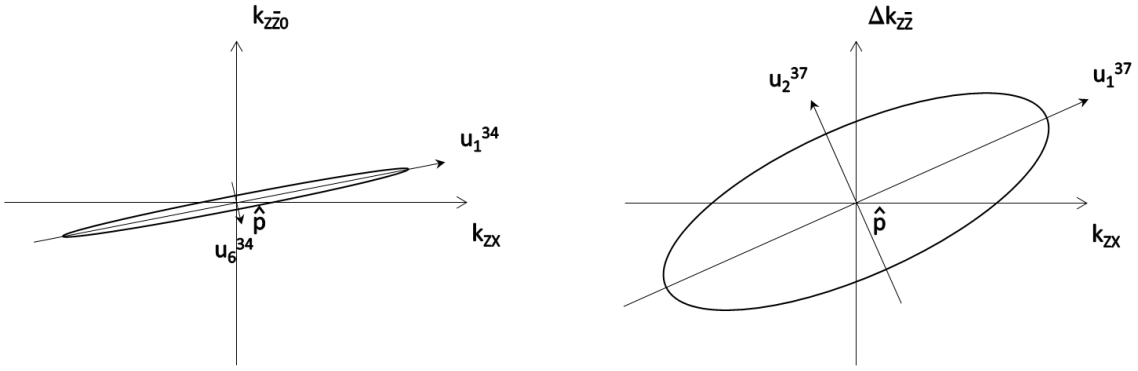

Figure 6.9: Projections of the ellipsoidal regions for the estimated parameter $\hat{\mathbf{p}}$ onto the k_{ZX}-$k_{Z\bar{Z}0}$-plane (left plot) and the k_{ZX}-$\Delta k_{Z\bar{Z}}$-plane (right plot) are shown for 50 nM of RHPS4. The ellipsoids are centred at the optimal parameter values $\hat{\mathbf{p}}$ and $\mathbf{u}_k^{ij} = (\mathbf{u}_{ki}, \mathbf{u}_{kj})$ are the projections onto the x_i-x_j-plane of the L-dimensional normalised eigenvectors $\mathbf{u}_k$ with components $\mathbf{u}_{k1}, \ldots, \mathbf{u}_{kL}$, which describe the directions of the principal axes in the L-dimensional ellipsoid. The lengths of the projected principal axes are $l_k \, \|\mathbf{u}_k^{ij}\|_2$, where l_k are the radii along the principal axes $\mathbf{u}_k$ of the L-dimensional ellipsoid and $\|\cdot\|$ denotes the Euclidean norm. The parameters k_{ZX} and $k_{Z\bar{Z}0}$ are strongly ($r_{34}(\hat{\mathbf{p}}) = 0.96$) correlated in contrast to k_{ZX} and $\Delta k_{Z\bar{Z}}$ ($r_{37}(\hat{\mathbf{p}}) = 0.66$).

6.5 Biological implications

We now interpret briefly the core biological findings from model fitting, thus relating our theoretical results to the biological processes of the mechanism of action of RHPS4 in the cell cycle.

We find that the drug affects the cell cycle phases differently at lower and higher concentrations. High concentrations do not have a pronounced cell-cycle specific effect on cells over the total observation period, whereas low concentrations seem to affect the G_2/M phases by increasing the rate of transition to the G_0/G_1 phases indicating fewer or faster processes occurring in the G_2/M phases, which simultaneously introduce marked cell death from G_2/M over the period of observation. One reason for the cell death in the G_2/M phases might be chromosome end-to-end joining, as described in Guo et al. [54], as chromosomes cannot be segregated during mitosis and cells do not pass the metaphase-anaphase checkpoint.

Furthermore, we observe a drug-dependent behaviour in terms of cells undergoing cell death around t_0 (the time point of significant cell death), where higher drug doses reduce the delay to the onset of marked cell death, which occurs in a largely cell-cycle independent manner.

The delay in the effects of the drug can be interpreted either as the time which the drug

requires to enforce its mechanism of action until the cell's repair machinery is depleted, or as the number of cell divisions required before the drug causes the occurrence of secondary effects leading to cell death.

A further effect of the drug, which we found by flow cytometry analysis, is the decrease of G_2/M and in particular S phase proportions at day 4 for all concentrations of RHPS4. We could, however, not capture these transient dynamics by our models.

6.6 Discussion and conclusions

In order to investigate the effects of the drug RHPS4 on cancer cell-cycle dynamics, we have collected experimental data from colon cancer cells as discussed in Chapter 4. We developed ODE models of the cell-cycle dynamics, allowing for different functions for the transition rates to cell death, and we analysed the model equations in Chapter 5 where we also discussed the behaviour of solutions. Expressions for the doubling time and residence times have also been derived by deterministic and probabilistic means, respectively. We have proposed a statistical model assuming log-normal errors for the experimental data and have tested the residuals of model-data fits for misspecification.

Model fitting has been performed in this Chapter by minimising a least squares cost function using global and local optimisation methods, and the Akaike criterion [3] was used to rank models of different complexity. The models 'best' describing the experimental data have been selected from the suggested candidate models. The best fit of data on the control cells was a model with constant transition rates; and the data for cells treated by 50 nM of RHPS4 was fitted best by including (sigmoidal) time-dependent transition rates. More complex effects influenced the cell cycle dynamics when the more potent treatment with 100 nM and 1 μM of the drug was applied, and less impressive agreement between model data and experimental data has been achieved for these cases. We also tried fitting one model to all four data sets (control and three doses) using transition rate functions k_* that are linear in the concentration of RHPS4, we could, however, not find a good fit to the data in this way. The estimated parameter values of the best-fitting models have been presented and we investigated the parameter sensitivities and the accuracy of our

estimated parameters. In each drug concentration, we found little correlation between the parameter estimates and generally reasonably good accuracy.

The error bars on our parameter estimates come from linearisation of the model around the optimal parameters and are thus symmetric. Symmetry of the standard errors about the parameter estimates is a rather unnatural assumption, in particular, when the parameters are purely positive and small. An approach based on the profile likelihood has recently been developed by Raue et al. [110] which does not require a linear approximation of the model. The method detects non-identifiabilities of model parameters and produces confidence intervals for parameter estimates by frequent re-fitting of the model, fixing one parameter each time. This technique might be an interesting alternative to estimating the accuracy of the parameter estimates in future work.

We report that the decrease in S (and consequently G_2/M) phase cells that RHPS4 causes around day 4 in HCT116 colorectal cancer cells is only a minor effect of the treatment, the cause of which we could not fully explain by our model. More significant is the delayed onset of rapid cell death within the population of treated cells, occurring largely from the G_2/M phase for 50 nM and from the G_0/G_1 phase for higher concentrations of RHPS4 (100 nM; 1 μM). In conclusion, we have demonstrated that the drug causes primarily abrupt cell death which is cell-cycle-specific and delayed.

The delay in the effect of the drug decreases with higher drug concentrations and the mechanisms for the delay remain to be discovered. RHPS4 stabilises G-quadruplexes, which has been found to inhibit telomerase [49], it thus shortens telomeres during replication. This could cause increased senescence of treated cells. The drug also causes telomere uncapping leading to apoptosis [118]. We suggest that RHPS4 affects telomeres of colorectal cancer cells in two ways: first, induced telomere shortening may decrease the fraction of telomeres in the capped state [113], and additional disruption of telomeric proteins may subsequently cause damage to the uncapped telomeres leading to activation of damage response pathways in cells traversing S phase [112]. These findings could serve as a potential explanation in particular for the late death of cells treated with 50 nM RHPS4, which we found occurs largely from the G_2/M phase. Secondly, higher concentrations of the drug may lock the telomeric end in G-quadruplex structures causing severe

replication stress by impairing fork progression in early S phase, which could explain the earlier onset of cell death for 100 nM and 1 μM of the drug. G-quadruplexes seem not to be compatible with chromosome replication: they have been observed *in vivo* throughout the cell cycle except for the phase of DNA replication [81]. Mathematical modelling of telomere replication processes with interference through G-quadruplex formation and stabilisation by RHPS4 can be a useful means to investigate the underlying mechanisms of the delayed cell death further, which is part of the following Chapter.

Part III

Telomere length dynamics

CHAPTER 7

Models of telomere length dynamics in telomerase-positive cells

Fitting our cell cycle model to experimental data we collected in the tissue culture lab suggested that the drug RHPS4 affects colorectal cancer cells markedly after a time delay which was concentration-dependent. We interpret the delay as the number of cell divisions the cell population has undergone after administering the drug and before cell death, and suggest that the observed cell death is associated to secondary effects of the drug at the terminating DNA sequences, called telomeres. In this Chapter we assume that, primarily, telomeres shorten due to inhibition of the enzyme telomerase by mechanisms that are induced by RHPS4. We aim to describe these mechanisms by mathematical models and investigate how the drug affects telomere length dynamics in cancer cells, particularly in the late S/G_2 phase of the cell cycle, that is, immediately after telomere duplication and before chromosome segregation and mitosis.

First, we briefly summarise the review of telomere biology from Part I that is relevant for the modelling presented in this Chapter. Telomeres of most somatic cells typically shorten in the S phase of the cell cycle due to the end-replication problem, that is, the inability to fully replicate the terminating DNA sequences [79]. Also C-strand resection by a nuclease may contribute to the cyclic shortening of telomeres [83]. A more detailed review of telomere replication can be found in Section 1.3.

The enzyme telomerase can antagonise telomere shortening by association with the telom-

eric end, where it progressively synthesises telomeric repeat sequences at the single-stranded overhang of the telomere (reviewed in Section 1.5). Human telomerase has been suggested to act rapidly on most ($\sim$ 70-100%) telomeres following replication, indiscriminately elongating leading and lagging DNA strands, where C-strand synthesis occurs after the synthesis of the G-overhang as a distinct event in the late S/G_2 phase [151, 158].

The unfolded, open form of telomeres is presumably the most natural form during telomere synthesis. Telomeres, however, can loop back and tuck their single-stranded end into the duplex DNA of telomeric sequences to form a t-loop (reviewed in Blackburn [23], de Lange [38]), protecting telomeric ends from DNA degradation and DNA damage responses (see Section 1.4 for greater detail). T-loops also function as telomerase inhibitors, as they hide the telomeric 3' end from access by telomerase, and structural rearrangements between t-loops and the open form of telomeres are likely to establish telomere length homeostasis.

In contrast, Cristofari and Lingner [36] found that HeLa telomeres, which were observed over 56 population doublings (PD), elongated at a constant rate of 415-635 bp/PD upon overexpression of the main functional subunits of the enzyme telomerase, the catalytic protein TERT and the telomerase RNA component (TERC). This massive telomerase activity is referred to as super-telomerase, and long telomeres did not change into a permanently non-extendible state in super-telomerase cells.

Alternatively, telomeric ends can spontaneously fold into guanine-rich structures called G-quadruplexes (G4), which are supported by monovalent cations such as potassium (K^+) in the nucleus (see Section 1.4). G4 structures form *in vivo* and probably unfold during telomere replication [121]. When G-quadruplexes are located at the very end of the telomeric G-overhang, which has been shown to be their preferred location [134], the enzyme telomerase is inhibited by the capping of the 3' end [155]. Optimal telomerase activity thus seems to require the non-folded single-stranded form of terminal telomere sequences.

Cheng et al. [29] compared relative quadruplex and duplex binding affinity constants of different quaternary polycyclic acridinium salts and found that quaternised quino$[4,3,2-kl]$-acridinium salts, such as RHPS4, selectively bind and stabilise quadruplex DNA.

Also, quadruplex DNA binding affinity correlated strongly with telomerase-inhibitory activity data for these G4 ligands. Thus, the quino$[4,3,2-kl]$acridinium methosulfate RHPS4 is an effective G4 ligand and potent telomerase inhibitor *in vitro* (see Section 1.7 for the effects of RHPS4 on cancer cells).

Standard techniques of telomere length measurements have been reviewed in Section 1.6.1. A high-throughput (HT) Q-FISH method has recently been developed [26], which can be used for large sample sets, generates telomere-length frequency histograms and allows for the analysis of interphase nuclei. Telomere length is maintained in telomerase-positive HeLa cells and has been measured at a mean value of 3.44 kb with standard deviation of 0.80 kb. A HT Q-FISH histogram of the telomere length distribution of HeLa cells is shown in Figure 7.1, which we represent by a normal distribution in Figure 7.2.

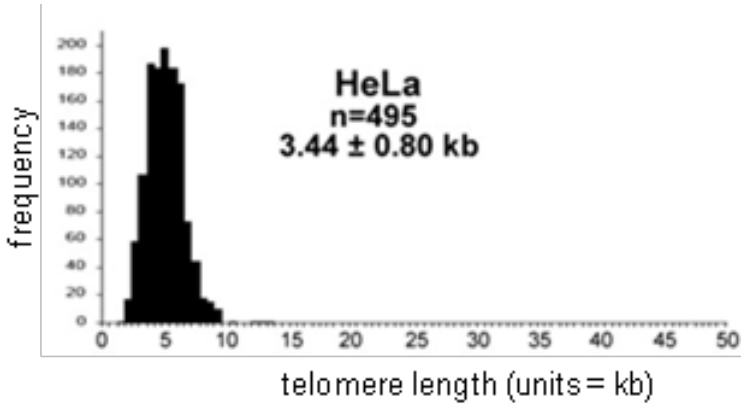

Figure 7.1: A HT Q-FISH histogram of the telomere length distribution of HeLa cells, where $n = 495$ nuclei were analysed (taken from Canela et al. [26]). Reprinted with permission from PNAS.

Our aim is to develop and analyse a variety of models of telomere length dynamics for telomerase-positive cancer cells, and to investigate how they respond to treatment with RHPS4. In Section 7.1, we gather data on the kinetics involved in the telomere-length regulatory processes described above, in order to obtain more accuracy in the predictions we make from simulating dynamic model behaviour. In Section 7.2, we consider a simple model describing the telomere length dynamics during one S/G_2 phase, where the model is closed in the sense that there is no influx or efflux of telomeres. In Section 7.3, we allow for a constant input of telomeres and include a negative feedback mechanism for telomere length regulation, where longer telomeres form t-loops and leave the system in this state. We also use this open model to describe telomere length dynamics not only for

one cell cycle, but also over a small number of cell generations. Finally, in Section 7.4, we feed the model output from the previous model back into the system and analyse under which conditions of telomerase and RHPS4 levels telomere length distributions stabilise over a large number of cell generations and predict the according steady-state length distributions. Section 7.5 summarises the results and contains a concluding discussion.

7.1 Estimation of kinetic model parameters

Some experimental results on the kinetic rate and equilibrium binding constants which we will use in simulation and analysis of model dynamics later (in Sections 7.2-7.4), can be found in the literature and will be described in this Section.

HeLa cells are telomerase-positive cells which were taken from a cervical cancer patient in 1951 and maintained in culture since then. The HeLa cell line is immortal, but when telomerase activity is inhibited in HeLa nuclei, telomeres shorten by $\sim$45 bp/cell division [158].

Binding kinetics of telomerase to single-stranded telomeric $(TTAGGG)_3$ sequences (no G4 folding possible) have been measured at 37° *in vitro* by pulse-labelling of primers in complex with telomerase (primer binding assay) and following the time-dependent dissappearance of labelled primers [143]. The dissociation rate of telomerase has been determined at $k_{\text{off}} = 0.013\,\text{min}^{-1}$, which is equivalent to a half-life of $t_{1/2} = \ln 2/k_{\text{off}} \approx 53$ min of the complex, and the equilibrium dissociation (Michaelis-Menten) constant has been measured at $K_m \approx 2$ nM. We consequently estimate the binding rate of telomerase as $k_{\text{on}} = k_{\text{off}}/K_m = 6.5 \times 10^{-3}\,\text{min}^{-1} \cdot \text{nM}^{-1}$.

Total overhang length in HeLa cells that were synchronised at the G_1/S transition gradually increased during 6-7.5 h after release [37, 158], indicating a phase of increased telomerase activity, and then decreased back to normal size during the G_2 phase, possibly due to fill-in of the C-strand. The exact mechanisms with respect to leading and lagging overhang generation, however, are not yet well understood.

The typical number ($R_{1/2}$) of telomeric repeats synthesised before half-life $t_{1/2}$ of the telomerase-telomere complex has been estimated to be between 0.66 and 4.1 [145], that

is between about 4 and 25 bases, and we derive the rate of telomere elongation as $\rho = R_{1/2}/t_{1/2}$. Hence we find $\rho \in \left[1.2 \times 10^{-3}, 7.7 \times 10^{-3}\right]$ in units of nt·s^{-1} (nucleotides per second) with mean value $\bar{\rho} = 4.5 \times 10^{-3}$ nt·s^{-1}. The variation in the measurements stems from different levels of the POT1-TPP1 complex used in assays, which seemed to enhance telomerase activity.

Zhao et al. [157] studied the formation of telomeric quadruplex structures by measuring the folding (k_f) and unfolding (k_u) rate constants of the human telomere sequence $(TTAGGG)_4$ at 150 mM of K^+ (typical intracellular concentration) and 37°C. The rate constants were determined by use of an optical biosensor as $k_f = 1.6 \times 10^{-2}$ s^{-1} and $k_u = 3.8 \times 10^{-3}$ s^{-1}, hence G4 structures and the unfolded, single-stranded form exist in a slow equilibrium *in vitro* with half lives of about 3 min and less than 1 min, respectively.

The equilibrium binding constant of RHPS4 with quadruplex forming human telomeric sequences $d[AG_3(TTAGGG)_3]$ has been estimated by fluorescence titration as $K = 2.70 \times 10^5$ M^{-1} and by surface plasmon resonance experiments (using an optical sensor) as $K = 110.0 \times 10^5$ M^{-1} [29]. We estimate K by choosing the mean value $K = 5.6 \times 10^6$ M^{-1}.

The dynamics of t-loop formation were described by a worm-like chain model in Rodriguez-Brenes and Peskin [113], who developed an algorithm to sample telomeric chromatin chains (modelled as semi-flexible polymer chains) at thermodynamic equilibrium. They assumed that the longer telomeres are, the more frequently telomere ends come into close proximity with internal positions of the telomere, and hence the more likely are invasions of double-stranded DNA by the G-overhang, which results in the formation of displacement loops together with t-loops. We approximate the rate of the formation of t-loops by a sigmoid function of telomere length (see Figure 7.2), inspired by a quantitative model [24] and data [136] for the telomere extension frequency in budding yeast cells (having shorter telomeres than human cells), which suggest that telomere length is regulated in a nonlinear fashion by a switch between telomerase-extendible and -nonextendible states. Using x to denote the number of basepairs (bp) of a telomere, we model the rate of t-loop formation (telomere capping) by

$$k_c(x) = \frac{\delta}{1 + \exp\left(\left(\alpha - x\right)/\beta\right)}, \tag{7.1.1}$$

with shape parameters $\alpha > 0$ bp, $\beta > 0$ bp and $\delta > 0$ s^{-1}. For small β, this has the form of a step function, with step at $x = \alpha$; $k_c(x) \approx 0$ for $x < \alpha$ and $k_c(x) \approx \delta$ for $x > \alpha$; and β describes the range of telomere lengths over which the transition occurs.

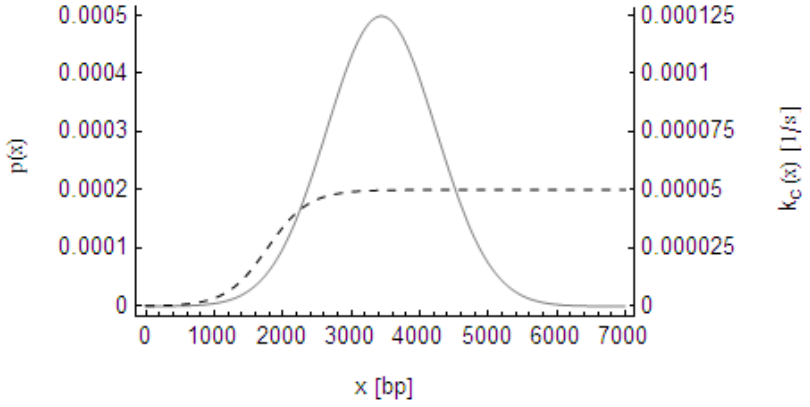

Figure 7.2: A Gaussian probability density function, $p(x)$, for the telomere length in HeLa cells, with mean $L_0 = 3440$ bp and standard deviation $\sigma = 800$ bp, is indicated by the solid gray line. The rate of t-loop formation (see formula (7.1.1)), $k_c(x)$, is modelled by a sigmoidal function of telomere length (dashed line) with shape parameters $\alpha = 1775$ bp, $\beta = 300$ bp and $\delta = 5 \times 10^{-5}$ s^{-1}. Shorter telomeres are more likely to be in an unlooped form than longer telomeres.

7.2 Model (i): A closed model for telomere length dynamics in the S/G$_2$ phase

The mechanisms described in Section 7.1 can be summarised in a simple model containing the states U, B, G and C respectively for the number of telomeres in the open (*U*ncapped) form, *B*ound to telomerase, in *G*-quadruplex formation and forming a *C*omplex with the drug RHPS4 (see Figure 7.3). After telomere duplication in the S phase, telomeres switch between the open and G4 forms (G4 folding rate k_f and G4 unfolding rate k_u), where telomeres in the open form bind free telomerase molecules T, with association rate k_{on} and dissociation rate k_{off}, which synthesise nucleotides with average rate ρ at the telomere end, thereby extending telomere length and compensating for incomplete replication of telomeres. Telomeric intramolecular G4 structures do not allow telomerase association with the 3′ end, and can be stabilised by free RHPS4 molecules R, where the association rate and dissociation rate of RHPS4 are k_s and k_d, respectively. We assume that *one* telomerase molecule binds *one* telomere to elongate the telomeric end, and *one* RHPS4 molecule is sufficient to stabilise a G4 form. Furthermore, all kinetic rates are

assumed to be constant and non-negative.

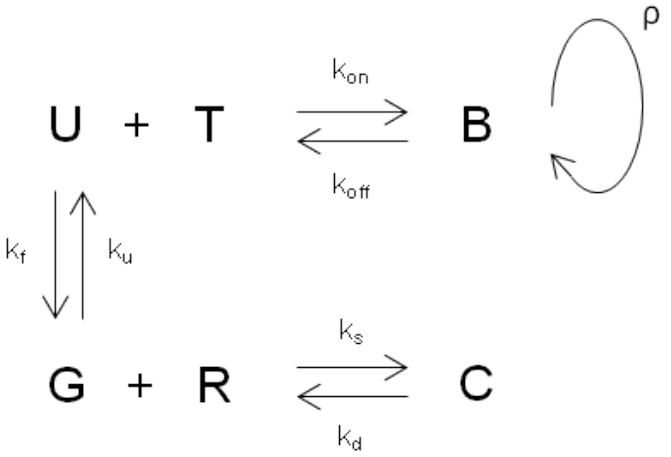

Figure 7.3: Model of telomeric states U, B, G, C. Kinetics for each reaction are described by their rate constants k. Free telomerase (T) and RHPS4 (R) in the nucleus bind open forms (U) and G4 structures (G), respectively. Telomerase elongation occurs at rate ρ.

We analyse the effect of the nuclear concentration of telomerase and RHPS4 on telomere-length dynamics in the time interval beginning at the end of chromosome duplication (when we assume that telomerase starts to act) and ending in the late G_2 phase (when telomerase stops acting and its activity is inhibited by other, yet unclear, mechanisms). Initially, all 184 telomeres within a cell (here, a HeLa cell) are assumed to be in the open form, that is, $U_0 = 184$, and we describe the mechanisms of our model by the following mass balance equations (ODEs),

$$\frac{\mathrm{d}}{\mathrm{d}t}U(t) = k_{\mathrm{off}}\,B(t) + k_u\,G(t) - \left(k_{\mathrm{on}}\,T(t) + k_f\right)U(t)\,, \tag{7.2.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}B(t) = k_{\mathrm{on}}\,T(t)U(t) - k_{\mathrm{off}}\,B(t)\,, \tag{7.2.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}G(t) = k_f\,U(t) + k_d\,C(t) - \left(k_u + k_s\,R(t)\right)G(t)\,, \tag{7.2.3}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}C(t) = k_s\,R(t)\,G(t) - k_d\,C(t)\,, \tag{7.2.4}$$

with $U(0) = U_0$ and $B(0) = G(0) = C(0) = 0$, where we treat the numbers of telomeres U, B, G, C, the numbers of telomerase molecules T and the numbers of RHPS4 molecules R as continuous variables. We assume that telomerase and RHPS4 molecule numbers are conserved quantities in the cell during the period of observation, that is, $T(t) + B(t) = T_0$ and $R(t) + C(t) = R_0$ at all time points $t \geq 0$. The system we consider is closed, the

number of telomeres is hence a conserved quantity as well, with

$$U(t) + B(t) + G(t) + C(t) \quad = \quad U_0\,. \tag{7.2.5}$$

According to the data given in Section 7.1, HeLa telomeres have an average telomere length of 3,440 bp and shorten due to the end replication problem and possibly C-strand resection by $\mu = 45$ bp per cell division, which makes an average length $l_0 = 3,395$ bp of telomeres after the duplication of telomeres and before telomerase has been activated in the cell. We assume that these are separate events for each telomere. We assume that the average length, $l(t)$, at time t, of telomeres in our system increases with rate ρ/U_0 when a telomerase molecule is associated with one telomere, hence

$$\frac{\mathrm{d}}{\mathrm{d}t}\, l(t) \quad = \quad \frac{\rho}{U_0}\, B(t)\,, \tag{7.2.6}$$

with $l(0) = l_0$.

Equation (7.2.5) can be used to express the quantity $C(t)$ in equation (7.2.3) in terms of $U(t)$, $B(t)$ and $G(t)$. Taking advantage of the fact that we can express the quantities $T(t)$ and $R(t)$ by other model variables, we reduce the system (7.2.1)-(7.2.4) to only three differential equations in $U(t)$, $B(t)$, $G(t)$, namely

$$\frac{\mathrm{d}}{\mathrm{d}t}\, U(t) \quad = \quad k_{\mathrm{off}}\, B(t) + k_u\, G(t) - \left(k_{\mathrm{on}}\, T_0 + k_f - k_{\mathrm{on}}\, B(t)\right) U(t)\,, \tag{7.2.7}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\, B(t) \quad = \quad k_{\mathrm{on}}\, (T_0 - B(t))\, U(t) - k_{\mathrm{off}}\, B(t)\,, \tag{7.2.8}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\, G(t) \quad &= \quad k_d\, U_0 + \left(k_f - k_d\right)\, U(t) - \left(k_u + k_s\, R_0 + k_d - k_s\, U_0\right) G(t) - k_d\, B(t) \\ &\quad - k_s\, G(t)\, (U(t) + B(t) + G(t))\,.\end{aligned} \tag{7.2.9}$$

We are interested in how the steady state of this system depends on the concentration of telomerase and how it is affected by RHPS4. Setting the left-hand side of each of the equations (7.2.7)-(7.2.9) to zero and solving for the variables U, B, G at steady state we

find

$$U = \frac{k_{\text{off}}\, B}{k_{\text{on}}\,(T_0 - B)}\,, \qquad\qquad G = \frac{k_{\text{off}}\, k_f\, B}{k_{\text{on}}\, k_u\,(T_0 - B)}\,, \tag{7.2.10}$$

and B is solution of the cubic

$$\begin{aligned} 0 \;\; = \;\; & {k_{\text{on}}}^2\, k_u^2\,(U_0 - B)\,(T_0 - B)^2 - {k_{\text{off}}}^2\, k_f\, k_s\, k_d^{-1}\,\left(k_f + k_u\right) B^2 \\ & - k_{\text{on}}\, k_{\text{off}}\, k_u\, B\,(T_0 - B)\left(k_u + k_f + k_f\, k_s\, k_d^{-1}\,(R_0 - U_0 + B)\right)\,, \end{aligned} \tag{7.2.11}$$

derived by inserting (7.2.10) into (7.2.9). To solve the cubic (7.2.11) we require estimates for the kinetic rate parameters, as it is difficult to solve (7.2.11) algebraically.

In order to adopt appropriate units for the parameters given in Section 7.1, we employ the fact that a concentration of 1 nM corresponds to approximately $10^{-9} N_A V_n \cdot \text{mol} \cdot \text{l}^{-1} \approx 415$ molecules per HeLa nucleus, where $N_A = 6.022 \times 10^{23}$ mol^{-1} is the Avogadro constant and $V_n = 6.9 \times 10^{-13}$ l is the nuclear volume of a HeLa cell. The BioNumbers data base of Milo et al. [87] provides us with an average volume for HeLa nuclei, which is taken from Monier et al. [90]. Using standard conversion factors, we obtain $k_{\text{on}} = 2.6 \times 10^{-7}$ s^{-1}, $k_{\text{off}} = 2.2 \times 10^{-4}$ s^{-1}, $\rho = 4.5 \times 10^{-3}$ nt·s^{-1}, $k_f = 1.6 \times 10^{-2}$ s^{-1}, $k_u = 3.8 \times 10^{-3}$ s^{-1} and $K = 1.4 \times 10^{-5}$ with $K = k_s\, k_d^{-1}$ for the equilibrium binding constant of RHPS4. The exact values for each of the rates of RHPS4 binding, k_s, and RHPS4 dissociation, k_d, are not relevant for the steady state of the system and we can set $k_s = 10^{-7}$ s^{-1} (according to the size of the other molecular binding rate, k_{on}), for example, then $k_d = k_s\, K^{-1} = 7.1 \times 10^{-3}$ s^{-1}.

We use MATHEMATICA to compute three numeric solutions of (7.2.11) in B, only one of which is physical, that is $0 \leq B < T_0$. Linear stability analysis over the ranges of $T_0 \in [0, 10000]$ and $R_0 \in [0\text{nM}, 1000\,\text{nM}]$ shows that the only physical solution is stable, that is, the real parts of the eigenvalues of the Jacobian matrix of system (7.2.7)-(7.2.9) evaluated at the steady state (U, B, G) are all negative. The steady state solutions in $U(t)$, $B(t)$, $G(t)$ are shown in Figure 7.4 as functions of the number of telomerase molecules for the case of no drug ($R_0 = 0$).

Assuming the system is in "quasi-steady" state, that is, the state variables operate close to

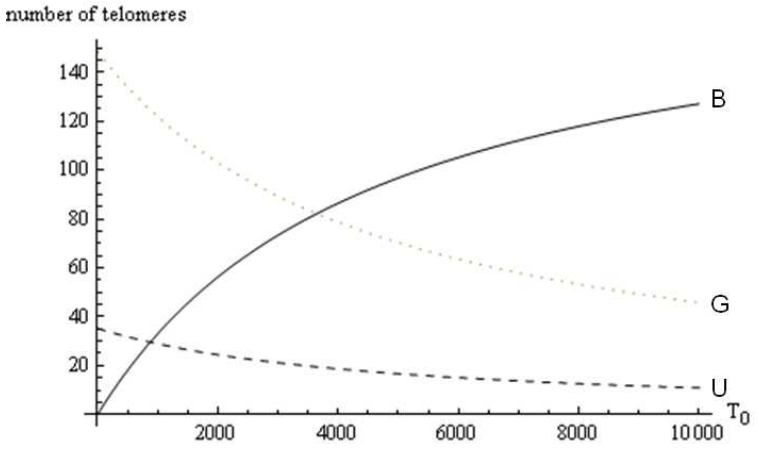

Figure 7.4: Simulation of the steady state numbers of telomeres B – bound by telomerase (solid line), U – in the open form (dashed line), and G – in G4 form (dotted line), dependent on telomerase molecule numbers T_0 in a drug-free assay ($R_0 = 0$).

the steady state and change slowly, the average telomere length at time t can be estimated by

$$l\,(t) = l_0 + \frac{\rho}{U_0}\,B\,t\,. \tag{7.2.12}$$

At an estimate of $T_0 = 2500$ telomerase molecules in the nucleus and no drug ($B \approx 65$), the time, t_r, it takes to replenish the number of nucleotides, $\mu = 45$ bp, lost due to the end replication problem is about 7.8 h. Figure 7.5 shows nucleotide addition of telomeres as a function of the number of telomerase molecules in the nucleus for different periods ($t_r = 4.5, 6, 7.8$ h) of telomerase activity. We also simulate telomere elongation by telomerase as a function of the concentration of RHPS4 for a fixed number of telomerase molecules T_0 in the nucleus, as shown in Figure 7.6 for $T_0 = 2500, 5000, 10000$ (the higher values of T_0 corresponding to super-telomerase cells).

It is possible to approximate the average telomere length $l(t)$ at $t = t_r$ as a function of the concentration of RHPS4, R_0, by rational polynomials of the form $1/(a\,R_0 + b) + c$ (with $a = 1.6 \times 10^{-7}$ nt^{-1}, $b = 1/\mu$ and $c = l_0$ for $T_0 = 2500$ and $t_r = 7.8$ h, for example, resulting in a maximum distance between curves of about 0.1 bp). One can find an approximation $l(t_r) - l_0$ for the number of nucleotides added during the time span t_R and replace the term T for telomerase elongation of telomeres in the model (1.6.4) in Section 1.6.2 by $l(t_r) - l_0$ to simulate changes in telomere length distributions over several chromosome replication events and for different drug concentrations R_0, as shown in

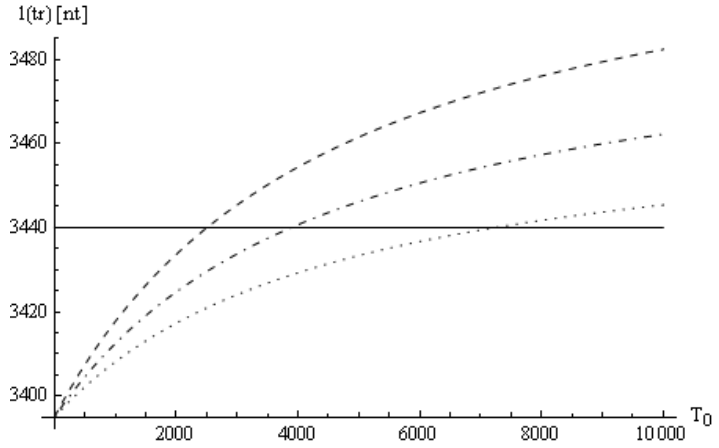

Figure 7.5: Simulation of the average length of telomeres after telomere shortening ($\mu = 45$ bp) due to the end-replication problem and subsequent telomere elongation by telomerase for a period of $t_r = 4.5$ h (dotted line), $t_r = 6$ h (dot-dashed line) and $t_r = 7.8$ h (dashed line) as a function of telomerase molecules T_0 and for the case of no drug ($R_0 = 0$), where we assume the system is in "quasi-steady" state. The average telomere length of HeLa cells is 3,440 bp and is indicated by the solid line.

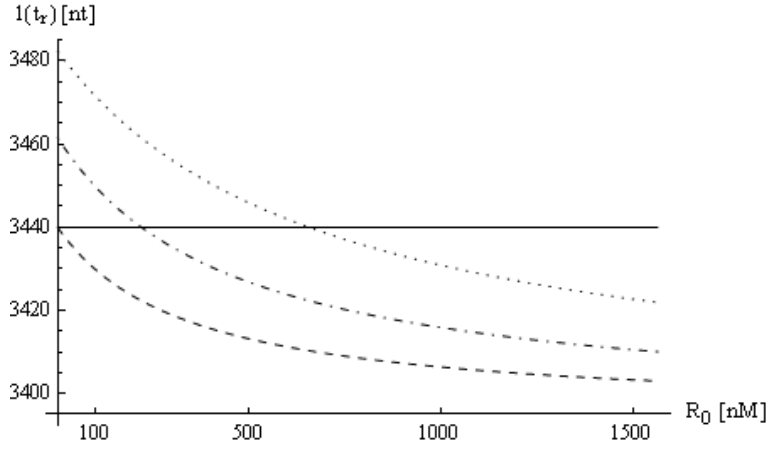

Figure 7.6: Simulation of the average length of telomeres after telomere shortening ($\mu = 45$ bp) due to the end-replication problem and subsequent telomere elongation by telomerase for a period of $t_r = 7.8$ h with $T_0 = 2500$ (dashed line), $T_0 = 5000$ (dot-dashed line) and $T_0 = 10000$ (dotted line) telomerase molecules: each curve plotted as a function of RHPS4 concentration R_0. We assume the system is in "quasi-steady" state. The average telomere length of 3,440 bp of HeLa cells is indicated by the solid line.

Figure 1.24 for the case of $l(t_r) - l_0 < \mu$. The model developed by Qi [108] distinguishes between leading- and lagging-strand replication, assuming telomere loss occurs on the leading telomere, but does not account for the effects of structural changes at the terminal sequences of telomeres. By contrast, we consider telomere length changes averaged over all telomeres in a cell, and aim to explore the impact of a feedback loop on telomere length dynamics over several replication events. Such a feedback loop may be introduced by the formation of t-loops of telomeric ends being in an open state.

7.3 Model (ii): An open model of length dynamics engaging a negative feedback mechanism in S/G_2

We want to simulate not only the dynamics of the average telomere length, but also of the telomere length distribution over time for control cells and cells treated with RHPS4. We refine our model (7.2.1)-(7.2.4) by including a variable x for the length of telomeres and allow for a constant influx of telomeres into the system, at rate k_e, and losses at rates k_c and k_r, being introduced by formation of t-loops and locking of G4 structures by RHPS4, respectively. We assume that the rate of t-loop formation is dependent on telomere length, that is k_c=$k_c(x)$ as introduced in Section 7.1, where shorter telomeres are more likely to form t-loops than longer telomeres and the relation is of a nonlinear (sigmoidal) nature. The extended model is illustrated in Figure 7.7.

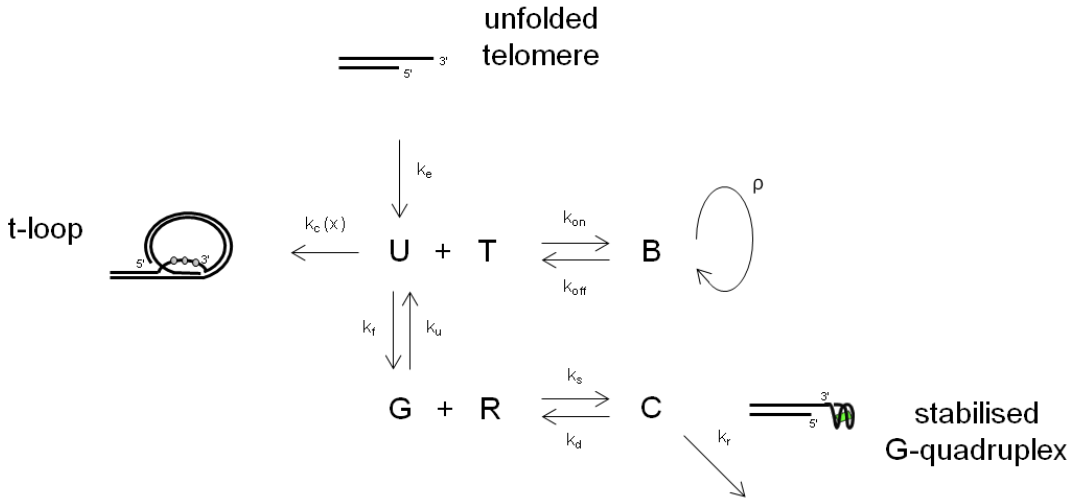

Figure 7.7: Model of telomeric states U, B, G, C. Kinetics for each reaction are described by their rate constants k. Free telomerase (T) and RHPS4 (R) in the nucleus bind open forms (U) and G4 structures (G), respectively. Telomerase elongation occurs at rate ρ. Telomeres enter the system at rate k_e and exit the system due to t-loop formation at rate $k_c(x)$ and due to G4-stabilisation by RHPS4 at rate k_r. Here, x is the length of a telomere.

Since the average telomere loss of about $\mu = 45$ bp during chromosome replication is much less than the initial telomere length of approximately 2k to 6k basepairs in HeLa cells, we treat telomere length, x, as a continuous variable. The dynamics of individual telomeres of length x at time t can be mathematically described by a partial differential equation (PDE) model of the number densities of telomeres in the states U, B, G, C, that

is,

$$\begin{aligned}
\frac{\partial}{\partial t} U(x,t) &= k_e\, p(x) + k_{\text{off}}\, B(x,t) + k_u\, G(x,t) \\
&\qquad - \left(k_c(x) + k_{\text{on}}\, T(t) + k_f\right) U(x,t)\,, \qquad (7.3.1)\\
\frac{\partial}{\partial t} B(x,t) &= k_{\text{on}}\, T(t)\, U(x,t) - k_{\text{off}}\, B(x,t) - \rho\, \frac{\partial}{\partial x} B(x,t)\,, \qquad (7.3.2)\\
\frac{\partial}{\partial t} G(x,t) &= k_f\, U(x,t) + k_d\, C(x,t) - \left(k_u + k_s\, R(t)\right) G(x,t)\,, \qquad (7.3.3)\\
\frac{\partial}{\partial t} C(x,t) &= k_s\, R(t)\, G(x,t) - \left(k_d + k_r\right) C(x,t)\,, \qquad (7.3.4)
\end{aligned}$$

where $p(x)$ is the probability density function of the length of telomeres entering the system at rate k_e, which we model as Gaussian, hence

$$p(x) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\, \exp\left(-\frac{(x-L)^2}{2\,\sigma^2}\right)\,, \qquad (7.3.5)$$

with average telomere length L and variance σ^2. This choice of $p(x)$ allows for small proportions of telomeres with negative length, which we will ignore in the following. One could also use a different distribution such as a log-normal distribution for $p(x)$ to avoid positive probabilities of telomeres with negative length; these, however, may reappear when we simulate telomere length distributions over several cell generations due to the shortening of telomeres that occurs at each cell division.

We adopt the functional form (7.1.1) for $k_c(x)$ as proposed in Section 7.1. Assuming that telomerase and RHPS4 are conserved quantities in the system, we have

$$T(t) + \int_0^\infty B(x,t)\,\mathrm{d}x = T_0\,, \qquad R(t) + \int_0^\infty C(x,t)\,\mathrm{d}x = R_0\,, \qquad (7.3.6)$$

for the numbers of free telomerase molecules, $T(t)$, and the numbers of free RHPS4 molecules, $R(t)$. The term $\rho\,\frac{\partial}{\partial x}\,B(x,t)$ is the only derivative term with respect to x in the model equations and accounts for the process of telomere elongation at rate ρ by telomerase.

We now assume that, for the system in steady state, the numbers of bound telomerase and bound RHPS4 molecules are small compared respectively to the numbers of free

telomerase and free RHPS4 molecules in the nucleus, that is, $T(t) \approx T_0$ and $R(t) \approx R_0$. Steady state telomere length distributions are then described by the equations

$$\begin{aligned} 0 &= k_e\, p(x) + k_{\text{off}}\, B(x) + k_u\, G(x) - \left(k_c(x) + k_{\text{on}}\, T_0 + k_f\right) U(x)\,, && (7.3.7)\\ \rho\, \frac{\partial}{\partial x}\, B(x) &= k_{\text{on}}\, T_0\, U(x) - k_{\text{off}}\, B(x)\,, && (7.3.8)\\ 0 &= k_f\, U(x) + k_d\, C(x) - (k_u + k_s\, R_0)\, G(x)\,, && (7.3.9)\\ 0 &= k_s\, R_0\, G(x) - (k_d + k_r)\, C(x)\,, && (7.3.10) \end{aligned}$$

for each of the four telomere states U, B, C, G. Using equations (7.3.9) and (7.3.10), we obtain $C \propto G \propto U$. We then express U as a function of B and p using equation (7.3.7), and subsequently rewrite (7.3.8) as an ODE for the variable $B(x)$. To avoid negativity in $B(x)$, a reasonable boundary condition to impose on (7.3.8) is $B(-\infty) = 0$ or $B(+\infty) = 0$ (imposing $B(0) = 0$ leads to a sign change of $B(x)$ in $x = 0$ due to $p(x) > 0$ for all $x \in \mathbb{R}$).

We estimate rate parameters as in Section 7.2 and we set $k_e = 9.2 \times 10^{-3}\ \text{s}^{-1}$ according to an approximate influx of 184 (normal number of chromosome ends in human cells) telomeres into the system per 5.5 h (the time during which telomere extension occurs [158]), and estimate, within biologically feasible ranges, the values $\alpha = 1775$, $\beta = 300$ and $\delta = 5 \times 10^{-5}\ \text{s}^{-1}$ for the sigmoid function $k_c(x)$ (compare Figure 7.2) and the loss rate $k_r = 5 \times 10^{-6}\ \text{s}^{-1}$. We simulate, by numerical integration of (7.3.8) using MATHEMATICA's built-in function NDSolve, steady state distributions for different molecule numbers T_0 of telomerase. Similarly, we simulate the telomere length distribution of telomeres leaving the system, $k_c(x)\, U(x) + k_r\, C(x)$, at steady state and compare it to the distribution $p(x)$ of telomeres entering the system as shown in Figure 7.8.

Our results show that for larger T_0, the distributions become increasingly positively skewed. Figure 7.9 shows the simulated steady-state telomere length distributions. The steady state curves $U(x)$ and $G(x)$ are of the same shape, but different amplitude.

We also simulate the telomere length distributions at steady state for $T_0 = 5000$, $\delta = 10^{-5}\ \text{s}^{-1}$ and varying molecule numbers R_0 of RHPS4 (see Figures 7.10 and 7.11). An increase in R_0 causes telomere length distributions to become less skewed. This is due to larger numbers of telomeres forming a complex, C, with RHPS4, which prevents telom-

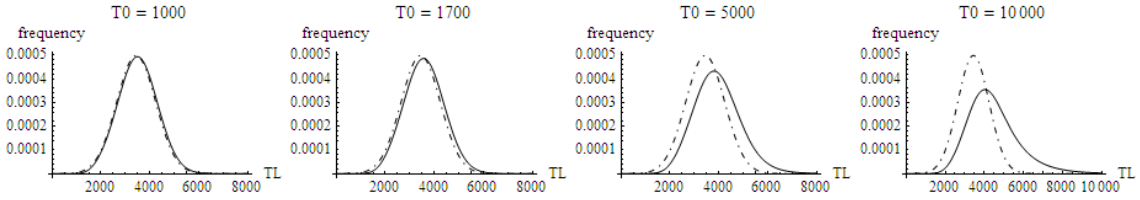

Figure 7.8: Simulations of telomere length distributions $p(x)$ (dot-dashed line) and $k_c(x)\,U(x) + k_r\,C(x)$ (solid line) of telomeres entering and telomeres leaving the open system per unit of time, respectively, for $\delta = 5 \times 10^{-5}$ s^{-1} and different numbers of telomerase molecules, $T_0 = 500, 1000, 1700, 5000$ (no drug, $R_0 = 0$). The x-axis represents telomere length (TL) in units of basepairs.

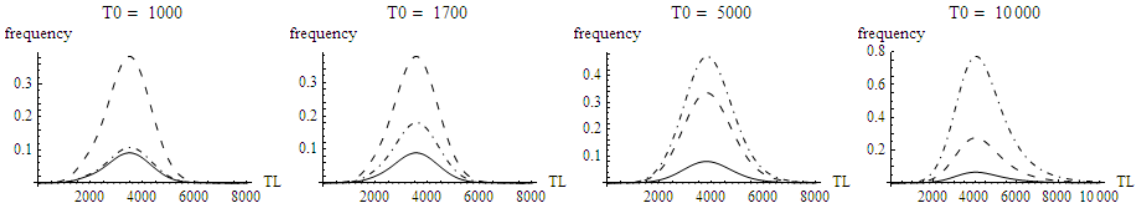

Figure 7.9: Simulations of steady state solutions $U(x)$, $B(x)$, $G(x)$ (represented by a solid, dot-dashed and dashed line, respectively) for $\delta = 5 \times 10^{-5}$ s^{-1} and different numbers of telomerase molecules, $T_0 = 1000, 1700, 5000, 10000$ (no drug, $R_0 = 0$). The x-axis represents telomere length (TL) in units of basepairs.

erase from adding telomeric sequences to telomere ends. Figure 7.12 presents a comparison of how many telomeres leave the system from U and C via t-loop formation and RHPS4 binding, respectively.

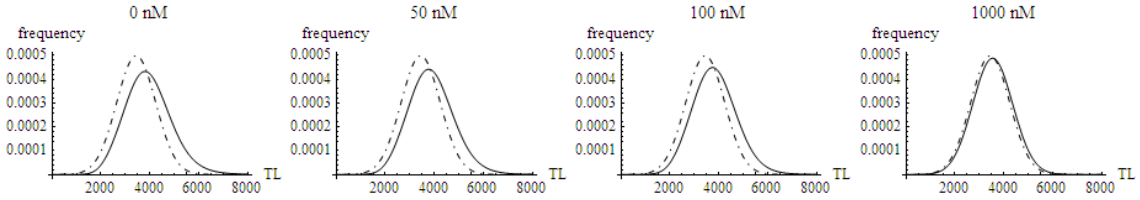

Figure 7.10: Simulations of telomere length distributions $p(x)$ (dot-dashed line) and $k_c(x)\,U(x) + k_r\,C(x)$ (solid line) of telomeres entering and telomeres leaving the open system per unit of time, respectively, for $\delta = 5 \times 10^{-5}$ s^{-1}, $T_0 = 5000$ and different concentrations of RHPS4, $R_0 = 0, 50, 100, 1000$ nM. The x-axis represents telomere length (TL) in units of basepairs.

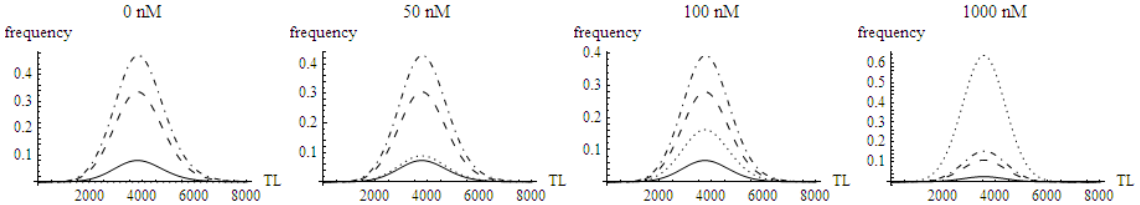

Figure 7.11: Simulations of steady state solutions $U(x)$, $B(x)$, $G(x)$, $C(x)$ (represented by a solid, dot-dashed, dashed and dotted line, respectively) for $\delta = 5 \times 10^{-5}$ s^{-1}, $T_0 = 5000$ and different concentrations of RHPS4, $R_0 = 0, 50, 100, 1000$ nM. The x-axis represents telomere length (TL) in units of basepairs.

In order to investigate which parameters control the shape of the telomere length dis-

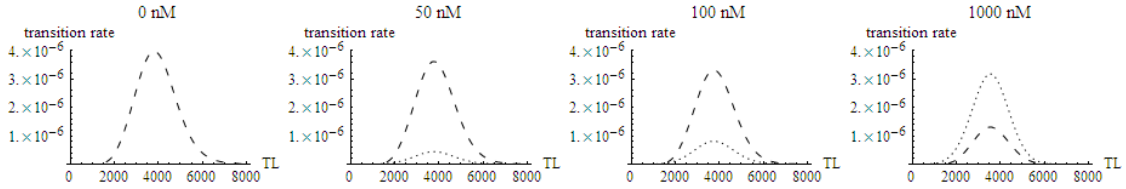

Figure 7.12: Simulations of telomere numbers $k_c(x)\,U(x)$ (dashed line) and $k_r\,C(x)$ (dotted line) leaving the system from U and C, respectively, per unit time for $\delta = 5\times10^{-5}\ \mathrm{s}^{-1}$, $T_0 = 5000$ and different concentrations of RHPS4, $R_0 = 0, 50, 100, 1000$ nM. The x-axis represents telomere length (TL) in units of basepairs.

tributions at steady state, we derive approximate analytical expressions for the distributions. We assume β is small and so approximate $k_c(x)$ in (7.1.1) by $\delta\,\mathrm{H}(x-\alpha)$, where $\mathrm{H}(x)$ denotes the Heaviside step function, to derive approximate analytical expressions at steady state for $U(x)$, $B(x)$, $G(x)$, $C(x)$, and the mean telomere length of telomeres leaving the system, $k_c(x)\,U(x) + k_r\,C(x)$. We use an integrating factor to solve the ODE (7.3.8) with the boundary conditions $B(-\infty) = 0$ and $B(+\infty) = 0$, to obtain

$$B(x) = \begin{cases} a_1\,e^{-a_2\,x}\left(1+\mathrm{erf}\left(\dfrac{x-L}{\sigma\sqrt{2}} - \dfrac{\sigma}{\sqrt{2}}\,a_2\right)\right), & x<\alpha\,, \\ b_1\,e^{-b_2\,x}\left(b_0+\mathrm{erf}\left(\dfrac{x-L}{\sigma\sqrt{2}} - \dfrac{\sigma}{\sqrt{2}}\,b_2\right)\right), & x>\alpha\,, \end{cases} \tag{7.3.11}$$

$$U(x) = \begin{cases} \dfrac{1}{c_1}\left(k_{\mathrm{off}}\,B(x) + k_e\,p(x)\right), & x<\alpha\,, \\ \dfrac{1}{\delta+c_1}\left(k_{\mathrm{off}}\,B(x) + k_e\,p(x)\right), & x>\alpha\,, \end{cases} \tag{7.3.12}$$

$$G(x) = \frac{k_f\,(k_d+k_r)}{k_u\,k_d + k_r\,(k_u + k_s\,R_0)}\,U(x)\,, \tag{7.3.13}$$

$$C(x) = \frac{k_s\,R_0}{k_d+k_r}\,G(x)\,, \tag{7.3.14}$$

where

$$a_2 = \frac{k_{\mathrm{off}}}{\rho}\left(1-\frac{k_{\mathrm{on}}\,T_0}{c_1}\right), \qquad b_2 = \frac{k_{\mathrm{off}}}{\rho}\left(1-\frac{k_{\mathrm{on}}\,T_0}{\delta+c_1}\right), \tag{7.3.15}$$

and

$$c_1 = k_{\mathrm{on}}\,T_0 + \frac{k_f\,k_s\,k_r\,R_0}{k_u\,k_d + k_r\,(k_u+k_s\,R_0)}\,, \tag{7.3.16}$$

and a_1, b_0, b_1 are lengthy explicit expressions involving kinetic parameters of the system, which are given in Appendix D. All constants satisfy $a_i, b_i > 0$. Here, the steady state

distribution of B is a product of a negative exponential function and an error function, where the error function dominates the exponential function for $x < \alpha$ and *vice versa* for $x > \alpha$. The parameter b_2 determines how rapidly the telomere length distribution approaches zero for increasing telomere length, x, with smaller values of b_2 increasing the positive skewness of the distribution. Hence, decreasing the rate of t-loop formation (by lowering δ) increases the positive skewness of $B(x)$, resulting in more longer telomeres, and so does increasing the number T_0 of telomerase molecules (or the rate ρ of telomere elongation), for example.

The parameters k_{off} and k_e function as scaling factors that determine the contribution of $B(x)$ and $p(x)$ to the telomere length distribution $U(x)$. For small k_{off}, when $B(x)$ is increasingly positively skewed, we expect $B(x)$ to have a larger tail than $U(x)$ due to the respectively decreasing and increasing contributions of $B(x)$ and $p(x)$ to $U(x)$. We note that $U(x)$ is independent of the rates k_f and k_u of G4 folding and unfolding, respectively, for control cells ($R_0 = 0$). The distributions of G and C are of the same shape as $U(x)$, where larger R_0 increases the ratio of telomere numbers C/G. The ratio of telomere numbers C/U increases with increasing R_0 in a nonlinear fashion, and tends to k_f/k_r for large R_0. On the other hand, the ratio of telomere numbers G/U decreases with increasing R_0 in an inverse fashion, and is equal to k_f/k_u for $R_0 = 0$. For larger T_0, the distributions become increasingly positively skewed.

By integrating the equations (7.3.7)-(7.3.10) over the interval $[-\infty, \infty)$, assuming $B(-\infty) = B(\infty) = 0$, and taking the sum of all these equations, we obtain the steady-state input-output balance

$$k_e = \int_{-\infty}^{\infty} k_c(x)\, U(x)\, \mathrm{d}x + k_r \int_{-\infty}^{\infty} C(x)\, \mathrm{d}x\,, \tag{7.3.17}$$

where $k_c(x)$ is given by (7.1.1). We have confirmed that this holds for the solutions plotted later.

An analytic expression for the mean telomere length, $\widehat{L} = \widehat{L}(T_0, R_0)$, of telomeres leaving the system at steady state, is

$$\widehat{L} = \frac{\int_{-\infty}^{\infty} x\, k_c(x)\, U(x) + x\, k_r\, C(x)\, \mathrm{d}x}{\int_{-\infty}^{\infty} k_c(x)\, U(x) + k_r\, C(x)\, \mathrm{d}x} = \frac{1}{k_e} \int_{-\infty}^{\infty} x\, k_c(x)\, U(x)\, \mathrm{d}x + \frac{k_r}{k_e} \int_{-\infty}^{\infty} x\, C(x)\, \mathrm{d}x\,, \tag{7.3.18}$$

and an approximate formula for (7.3.18), based on the approximation of $k_c(x)$ by $\delta\,\mathrm{H}(x-\alpha)$, has been derived using MATHEMATICA with a series of variable substitutions and simplifications, as the formulae involved in the computation are long and complex. We consider only the limiting cases $R_0 = 0$ (no drug), $R_0 \to \infty$ (high concentrations of RHPS4) and $\sigma \to 0$ (all telomeres are initially of the same length) for $\widehat{L}$, that is,

$$\widehat{L}_{R_0=0} = L + \frac{\rho}{2\,k_{\mathrm{off}}}\left(\mathrm{erfc}\left(\frac{L-\alpha}{\sigma\sqrt{2}}\right) + \frac{2\,k_{\mathrm{on}}\,T_0}{\delta}\right), \tag{7.3.19}$$

$$\begin{aligned}\widehat{L}_{R_0\to\infty} &= L + \frac{k_{\mathrm{on}}\,T_0\,\rho}{k_{\mathrm{off}}\,(\delta+k_f)} + \frac{\delta\,k_{\mathrm{on}}\,T_0\,\rho}{2\,k_{\mathrm{off}}\,k_f\,(\delta+k_f)}\,\mathrm{erfc}\left(\frac{L-\alpha}{\sigma\sqrt{2}}\right)\\ &\quad - \frac{\delta\,(k_{\mathrm{on}}\,T_0)^2\,\rho}{2\,k_{\mathrm{off}}\,k_f\,(\delta+k_f)\,(k_{\mathrm{on}}\,T_0+k_f)}\,\mathrm{erfc}\left(\frac{L-\alpha}{\sigma\sqrt{2}} + \frac{\sigma\,k_{\mathrm{off}}\,k_f}{\rho\sqrt{2}\,(k_{\mathrm{on}}\,T_0+k_f)}\right)\\ &\quad \times \exp\left(\frac{k_{\mathrm{off}}\,k_f}{\rho\,(k_{\mathrm{on}}\,T_0+k_f)}\left(L-\alpha+\frac{\sigma^2\,k_{\mathrm{off}}\,k_f}{2\,\rho\,(k_{\mathrm{on}}\,T_0+k_f)}\right)\right),\end{aligned} \tag{7.3.20}$$

$$\widehat{L}_{\sigma\to 0} = L + \frac{k_{\mathrm{on}}\,T_0\,\rho\,(k_u\,(k_d+k_r)+k_s\,k_r\,R_0)}{k_{\mathrm{off}}\,(\delta\,k_u\,(k_r+k_d)+k_s\,k_r\,R_0\,(\delta+k_f))}, \tag{7.3.21}$$

where $\mathrm{erfc}(x) = 1-\mathrm{erf}(x)$ is the complementary error function, and we assume positive concentrations of telomerase, $T_0 > 0$, and $L > \alpha$. The expression $\widehat{L}$ is independent of k_e, and $\widehat{L}_{R_0=0}$ and $\widehat{L}_{R_0\to\infty}$ are also independent of k_u, k_s, k_d and k_r. An increase in the parameter values T_0 (or ρ) leads to a linear increase in $\widehat{L}_{R_0=0}$, becoming nonlinear for positive values of R_0. An increase in σ leads to an increase in $\widehat{L}_{R_0=0}$. If we assume $\sigma_0 \to 0$, that is, all telomeres initially have the same length, we find $\widehat{L}_{\sigma\to 0}$ is independent of α and increases linearly with T_0 (or ρ).

We use (7.3.19) and (7.3.20), or (7.3.21) to derive expressions for the limits of $\widehat{L}_{\sigma\to 0}$ for extreme values of R_0, that is

$$\widehat{L}_{R_0=0,\sigma\to 0} = L + \frac{k_{\mathrm{on}}\,T_0\,\rho}{k_{\mathrm{off}}\,\delta}, \tag{7.3.22}$$

$$\widehat{L}_{R_0\to\infty,\sigma\to 0} = L + \frac{k_{\mathrm{on}}\,T_0\,\rho}{k_{\mathrm{off}}\,(\delta+k_f)}. \tag{7.3.23}$$

We use formula (7.3.19) to approximate $\widehat{L}_{R_0=0}$ for varying δ or σ, and different values of T_0 (see Figure 7.13). Here, a value of $T_0 = 500$ and $\delta = 5\times 10^{-5}\ \mathrm{s}^{-1}$ results in $\widehat{L}_{R_0=0} \approx L_0$ choosing $L = L_0 - \mu$ (initial mean telomere length of telomeres immediately after a telomere shortening event). Furthermore, we simulate $\widehat{L}$ for varying R_0 and different values of

T_0, and also plot $\widehat{L}_{R_0\to\infty}$ against T_0 (Figure 7.14), which results in an (approximately) straight line.

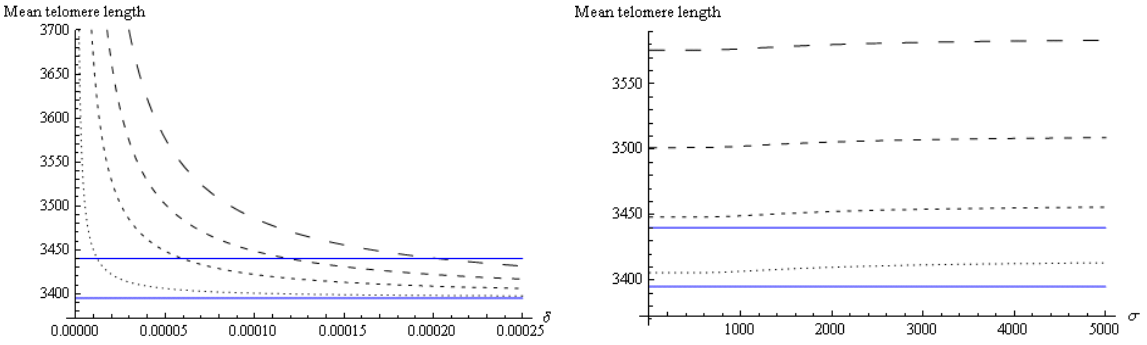

Figure 7.13: The mean telomere length $\widehat{L}_{R_0=0}$ of telomeres leaving the system, $k_c(x)\,U(x) + k_r\,C(x))$ is plotted against the parameter δ (left; $\sigma = 800$ bp) and σ (right; $\delta = 5\times10^{-5}$ s^{-1}), controlling the rate of t-loop formation and the standard deviation of the initial telomere length distribution $p(x)$, respectively. Simulations are shown for four different values $T_0 = 100, 500, 1000, 1700$ (from smaller to larger dashing of lines) and no drug. The solid lines indicate the threshold values $L = L_0 - \mu$ and $\widehat{L} = L_0$.

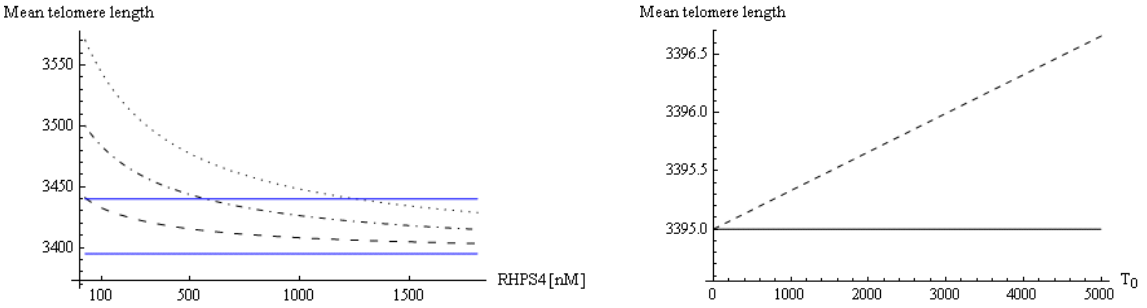

Figure 7.14: The mean telomere length $\widehat{L}$ of telomeres leaving the system, $k_c(x)\,U(x) + k_r\,C(x)$, plotted (left) against the concentration R_0 of RHPS4 for three different numbers of telomerase molecules, $T_0 = 419$ ($\widehat{L} \approx L_0$, dashed line), $T_0 = 1000$ (dot-dashed line), $T_0 = 1700$ (dotted line), where $\delta = 5\times10^{-5}$ s^{-1} and the solid lines indicate the threshold values $L = L_0 - \mu$ and $\widehat{L} = L_0$. The plot on the right-hand side of the Figure shows $\widehat{L}_{R_0\to\infty}$ (dashed line) as a function of T_0, where $L = L_0 - \mu$ is indicated by a solid line.

We aim to show numerical results of telomere length distributions over a few generations. To derive steady state distributions at the end of each replication, we initially assume the length distribution of telomeres before the first replication event to be Gaussian $p_0(x)$ as in (7.3.5) with $L = L_0$. Telomeres shorten at an average amount μ due to the end-replication problem and postreplicative processing, and consequently enter the system with length distribution $p(x) = p_0(x+\mu)$. By numerically integrating (7.3.8) and using (7.3.17) we simulate the telomere length probability density function of telomeres leaving the system, $p_1(x) = (k_c(x)\,U(x) + k_r\,C(x))/k_e$, at steady state and compare it to the distribution $p_0(x)$ of telomeres entering the system before telomere shortening takes place. Assuming telomere length does not change between the S/G$_2$ phases of subsequent cell

divisions, we treat the output telomere lengths of one cycle as the input telomere length for the next generation, $p = p_1(x + \mu)$. In particular, we use the steady state distribution of telomeres leaving the system at the end of the initial generation $i = 0$ as input (of rate k_e) into the system at the beginning of generation $i = 1$ and after telomere loss of amount μ occurred, that is $(k_c(x+\mu)\,U(x+\mu) + k_r\,C(x+\mu))/k_e = p_1(x+\mu)$ replacing $p(x)$ in equation (7.3.1), and derive steady state distributions $p_i(x)$ for higher generations i in this fashion.

Figures 7.15 and 7.16 show simulations of telomere length distributions, $p_i(x)$, of telomeres entering the system at generation i, for different concentrations of RHPS4 and $T_0 = 500$ (telomere length equilibrium, $i = 25$) and $T_0 = 5000$ (super-telomerase cells, $i = 10$), respectively. For increasing R_0 telomere length distributions are shifted towards 0 and their shape changes from positively to slightly negatively skewed.

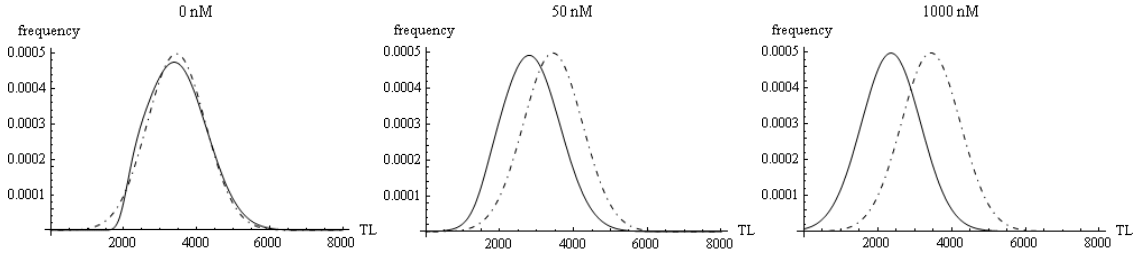

Figure 7.15: Simulations of telomere length distributions $p_0(x)$ (dot-dashed line) and $p_i(x)$ (solid line) of telomeres entering the open system at generations 0 and $i = 25$, respectively, in units of s^{-1} for $T_0 = 500$, $\delta = 5 \times 10^{-5}$ s^{-1} and different concentrations of RHPS4 (0nM, 50nM and 1000nM). Telomeres shorten by $\mu = 45$ nt between each round of replication. The x-axis represents telomere length (TL) in units of basepairs.

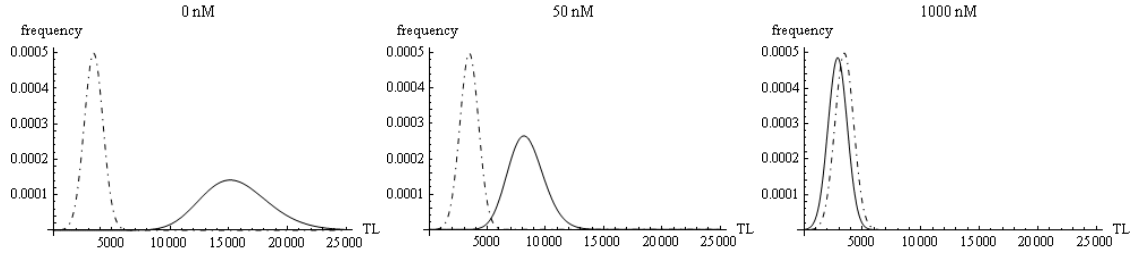

Figure 7.16: Simulations of telomere length distributions $p_0(x)$ (dot-dashed line) and $p_i(x)$ (solid line) of telomeres entering the open system at generations 0 and $i = 10$, respectively, in units of s^{-1} for $T_0 = 5000$ (super-telomerase cells), $\delta = 5 \times 10^{-5}$ s^{-1} and different concentrations (0nM, 50nM and 1000nM) of RHPS4. Telomeres shorten by $\mu = 45$ nt between each round of replication. The x-axis represents telomere length (TL) in units of basepairs.

7.4 Model (iii): Model of telomere length dynamics over large numbers of cell divisions

7.4.1 Model formulation

In order to investigate changes in the telomere length distributions over several cell divisions, we now modify the model further and feed the telomeres that exit the system at rates $k_c(x)$ and k_r back into the system. We assume that telomeres shorten by an amount μ due to the end replication problem (in the S phase) before they re-enter the system. The strategy we employed at the end of Section 7.3 to simulate telomere length distributions over several generations worked well for smaller generation numbers i, but was computationally too expensive for large values of i. Figure 7.17 illustrates an extended system incorporating model (7.3.1)-(7.3.4), but allowing for analysis of the telomere length distribution after large numbers of cell divisions, that is iterative S/G_2 phases, where we assume that telomere length does not change in the G_0/G_1 and the M phase of the cell cycle.

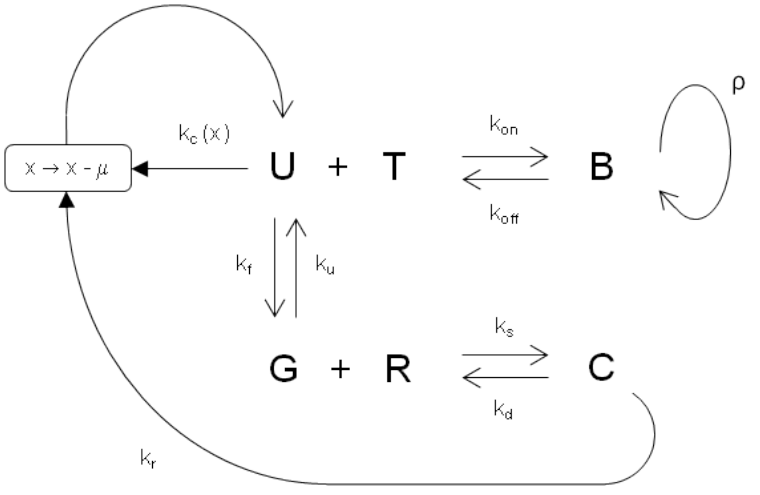

Figure 7.17: A closed model of telomeric states U, B, G, C, with telomeres of length x losing μ basepairs when they exit the system by t-loop formation (rate $k_c(x)$) or after forming a complex with RHPS4 (rate k_r); telomeres re-enter the system in the open (U) form. Kinetics for each reaction are described by their rate constants k. Free telomerase (T) and RHPS4 (R) in the nucleus bind open telomere forms and G4 structures (G), respectively. Telomerase elongation occurs at rate ρ.

We aim to analyse the model dynamics and compare the results of our analysis to experimental telomere length measurements and to the theoretical and experimental results we obtained in the work in Chapter 4 and 5. We replace $k_e\, p(x)$ in (7.3.1) (section 7.3) by

$k_c(x+\mu)\,U(x+\mu,t)+k_r\,C(x+\mu,t)$ and obtain the resulting, closed system of differential equations

$$\begin{aligned}
\frac{\partial}{\partial t}\,U\,(x,t) &= k_c\,(x+\mu)\;U\,(x+\mu,t)+k_r\,C\,(x+\mu,t)+k_{\text{off}}\,B\,(x,t)+k_u\,G\,(x,t)- \\
&\qquad -\left(k_c\,(x)+k_{\text{on}}\,T\,(t)+k_f\right)U\,(x,t)\;, && (7.4.1)\\
\frac{\partial}{\partial t}\,B\,(x,t) &= k_{\text{on}}\,T\,(t)\,U\,(x,t)-k_{\text{off}}\,B\,(x,t)-\rho\,\frac{\partial}{\partial x}\;B\,(x,t)\;, && (7.4.2)\\
\frac{\partial}{\partial t}\,G\,(x,t) &= k_f\,U\,(x,t)+k_d\,C\,(x,t)-\left(k_u+k_s\,R\,(t)\right)G\,(x,t)\;, && (7.4.3)\\
\frac{\partial}{\partial t}\,C\,(x,t) &= k_s\,R\,(t)\,G\,(x,t)-\left(k_d+k_r\right)C\,(x,t)\;, && (7.4.4)
\end{aligned}$$

where equations (7.4.2)-(7.4.4) are identical to equations (7.3.2)-(7.3.4) in Section 7.3, and (7.4.1) adds delay terms (in x) to equation (7.3.1).

We are interested in analysing the system when telomere length distributions are at a steady state to predict how telomerase and/or RHPS4 affect telomere length distributions over large numbers of cell divisions. It is a big assumption to assume steady state solutions, as there may not be enough telomerase to maintain a steady state, or there may be too much telomerase and telomere length tends to $+\infty$. Hence we expect a window of feasible solutions. As in Section 7.3, we assume that $T(t)\approx T_0$ and $R(t)\approx R_0$ hold at steady state and hence obtain the steady state equations

$$\begin{aligned}
0 &= k_c(x+\mu)\,U(x+\mu)+k_r\,C(x+\mu)+k_{\text{off}}\,B\,(x)+k_u\,G\,(x) \\
&\qquad -\left(k_c\,(x)+k_{\text{on}}\,T_0+k_f\right)U\,(x)\;, && (7.4.5)\\
\rho\,\frac{\partial}{\partial x}\;B\,(x) &= k_{\text{on}}\,T_0\,U\,(x)-k_{\text{off}}\,B\,(x)\;, && (7.4.6)\\
0 &= k_f\,U\,(x)+k_d\,C\,(x)-\left(k_u+k_s\,R_0\right)G\,(x)\;, && (7.4.7)\\
0 &= k_s\,R_0\,G\,(x)-\left(k_d+k_r\right)C\,(x)\;. && (7.4.8)
\end{aligned}$$

Solving (7.4.5), (7.4.7), (7.4.8) for B as a function of $U(x)$ and $U(x+\mu)$, we obtain

$$B(x)=\frac{1}{k_{\text{off}}}\left(\left(g(x)+k_{\text{on}}\,T_0\right)U(x)-g(x+\mu)\,U(x+\mu)\right)\;, \qquad (7.4.9)$$

with

$$g(x) = k_c(x) + c_R\,, \tag{7.4.10}$$

where

$$c_R = \frac{k_f\,k_s\,k_r\,R_0}{k_u\,k_d + k_r\,(k_u + k_s\,R_0)}\,. \tag{7.4.11}$$

Inserting (7.4.9) into the ODE (7.4.6) we reduce equations (7.4.1)-(7.4.4) to a single delay differential equation in $U(x)$, namely

$$\begin{aligned} 0 \;=\; & \frac{\rho}{k_{\rm off}}\,(g(x) + k_{\rm on}\,T_0)\,U'(x) - \frac{\rho}{k_{\rm off}}\,g(x+\mu)\,U'(x+\mu) + \left(g(x) + \frac{\rho}{k_{\rm off}}\,g'(x)\right)U(x) \\ & - \left(g(x+\mu) + \frac{\rho}{k_{\rm off}}\,g'(x+\mu)\right)U(x+\mu)\,. \end{aligned} \tag{7.4.12}$$

7.4.2 Solution by quasi-continuum approximation

Now we aim to find approximate solutions to equation (7.4.12) by using quasi-continuum approximations previously used for nonlinear waves in advance-delay equations [33, 114, 146].

By defining $y = x + \mu/2$ and using the relation $(g\,U)' = g'U + g\,U'$, we can re-write (7.4.12) as

$$\begin{aligned} 0 \;=\; & \frac{\rho}{k_{\rm off}}\,\left((g\,U)' + k_{\rm on}\,T_0\,U'\right)(y - \tfrac{1}{2}\mu) - \frac{\rho}{k_{\rm off}}\,(g\,U)'(y + \tfrac{1}{2}\mu) + (g\,U)(y - \tfrac{1}{2}\mu) \\ & -(g\,U)(y + \tfrac{1}{2}\mu)\,. \end{aligned} \tag{7.4.13}$$

It is useful to introduce the following notation. Let ∂_y denote the differential operator $\partial/\partial y$, and $\partial_y^{(n)}$ denote $\partial^n/\partial y^n$. For analytic functions and $\alpha \in \mathbb{R}$, we can express

$$f(y + \alpha) = \mathrm{e}^{\alpha\,\partial_y}\,f(y)\,, \tag{7.4.14}$$

where $\exp(\alpha\,\partial_y) = \sum_{n=0}^{\infty}\alpha^n\,\partial_y^{(n)}/n!$ is a differential operator which yields the Taylor series at y, that is,

$$f(y+\alpha) = \mathrm{e}^{\alpha\,\partial_y}\,f(y) = f(y) + \alpha\,f'(y) + \frac{\alpha^2}{2!}\,f''(y) + \frac{\alpha^3}{3!}\,f'''(y) + \ldots\,, \tag{7.4.15}$$

when applied to a function f. Assuming U is analytic, we use (7.4.14) with $\alpha = \mu/2$ and $\alpha = -\mu/2$ and re-formulate (7.4.13) as

$$\mathrm{e}^{-\frac{1}{2}\mu\,\partial_y}\left(\frac{k_{\mathrm{on}}\,T_0\,\rho}{k_{\mathrm{off}}}\,\partial_y\right)U + \left(\mathrm{e}^{-\frac{1}{2}\mu\,\partial_y} - \mathrm{e}^{\frac{1}{2}\mu\,\partial_y}\right)\left(\frac{\rho}{k_{\mathrm{off}}}\,\partial_y + 1\right)(g\,U) = 0\,. \tag{7.4.16}$$

Re-arranging (using commutativity of operator multiplication) and taking inverse yields

$$\begin{aligned} g\,U &= -\left(\mathrm{e}^{-\frac{1}{2}\mu\,\partial_y} - \mathrm{e}^{\frac{1}{2}\mu\,\partial_y}\right)^{-1}\left(\frac{\rho}{k_{\mathrm{off}}}\,\partial_y + 1\right)^{-1}\left(\mathrm{e}^{-\frac{1}{2}\mu\,\partial_y}\right)\left(\frac{k_{\mathrm{on}}\,T_0\,\rho}{k_{\mathrm{off}}}\,\partial_y\right)U \\ &= \mathfrak{A}\,U\,, \end{aligned} \tag{7.4.17}$$

where $\mathfrak{A}$ is an operator. Expanding the components of $\mathfrak{A}$ to third order in μ, we find

$$\mathrm{e}^{-\frac{1}{2}\mu\,\partial_y} = 1 - \tfrac{1}{2}\mu\,\partial_y + \frac{\mu^2}{8}\,\partial_y^2 - \frac{\mu^3}{16}\,\partial_y^3 + \mathcal{O}(\mu^4)\,, \tag{7.4.18}$$

and

$$\mathrm{e}^{-\frac{1}{2}\mu\,\partial_y} - \mathrm{e}^{\frac{1}{2}\mu\,\partial_y} = -\mu\,\partial_y - \frac{\mu^3}{24}\,\partial_y^3 + \mathcal{O}(\mu^5) = (-\mu\,\partial_y)\left(1 + \frac{\mu^2}{24}\,\partial_y^2 + \mathcal{O}(\mu^4)\right)\,, \tag{7.4.19}$$

and for the inverse operators

$$\left(1 + \frac{\mu^2}{24}\,\partial_y^2\right)^{-1} = 1 - \frac{\mu^2}{24}\,\partial_y^2 + \mathcal{O}(\mu^4)\,, \tag{7.4.20}$$

and

$$\left(\frac{\rho}{k_{\mathrm{off}}}\,\partial_y + 1\right)^{-1} = 1 - \frac{\rho}{k_{\mathrm{off}}}\,\partial_y + \frac{\rho^2}{{k_{\mathrm{off}}}^2}\,\partial_y^2 - \frac{\rho^3}{{k_{\mathrm{off}}}^3}\,\partial_y^3 + \mathcal{O}\Big(\frac{\rho^4}{{k_{\mathrm{off}}}^4}\Big)\,. \tag{7.4.21}$$

Hence, assuming the quantities μ and ρ/k_{off} are of the same order of magnitude (from Section 7.2, $\mu = 45$ nt, $\rho/k_{\mathrm{off}} \approx 18$ nt), we obtain

$$g\,U = \left(k_{\mathrm{on}}\,T_0\,\frac{\rho}{\mu\,k_{\mathrm{off}}} - k_{\mathrm{on}}\,T_0\,\frac{\rho}{\mu\,k_{\mathrm{off}}}\left(\tfrac{1}{2}\mu + \frac{\rho}{k_{\mathrm{off}}}\right)\partial_y + \mathcal{O}\left(\mu^2\right)\right)U\,. \tag{7.4.22}$$

Note that the integration constant that appears when one applies ∂_y^{-1} to U is equal to zero, as we assume $U^{(j)}(x) \to 0$ as $x \to \pm\infty$ for $j = 0, 1, 2, 3$. Thus we obtain the first-order

differential equation

$$\left(\tfrac{1}{2}\mu + \frac{\rho}{k_{\text{off}}}\right) U' = \left(1 - \frac{\mu\, k_{\text{off}}\, g}{k_{\text{on}}\, T_0\, \rho}\right) U\,, \tag{7.4.23}$$

which approximately describes the telomere length distribution $U(x)$ at steady state.

We rewrite (7.4.23) as

$$c_0\, c_T\, \frac{\mathrm{d}U}{\mathrm{d}x} + (g\,(x) - c_T)\, U = 0\,, \tag{7.4.24}$$

by defining

$$c_0 = \frac{1}{2}\,\mu + \frac{\rho}{k_{\text{off}}}\,, \qquad c_T = \frac{k_{\text{on}}\, T_0\, \rho}{\mu\, k_{\text{off}}}\,, \tag{7.4.25}$$

and use (7.4.24) to analyse how different numbers T_0 and R_0 affect the distribution of telomeres in the system at steady state. Equation (7.4.24) is a separable ODE, which we solve by integrating with respect to x and using (7.4.10) and (7.1.1) for $g(x)$ and $k_c(x)$, respectively, thus

$$\ln U(x) = -\frac{\beta\,\delta}{c_0\, c_T} \ln\left(\mathrm{e}^{x/\beta} + \mathrm{e}^{\alpha/\beta}\right) + \frac{1}{c_0}\left(1 - \frac{c_R}{c_T}\right) x + s_0\,, \tag{7.4.26}$$

where $s_0 \in \mathbb{R}$ is a constant that depends on the initial number of telomeres in the system. We simplify expression (7.4.26) by using the hyperbolic secant $\operatorname{sech}(x) = 2/(\exp(x) + \exp(-x))$ to obtain

$$U(x) = A\, e^{\lambda\, x} \operatorname{sech}^{\frac{\beta\,\delta}{c_0\, c_T}}\left(\frac{x - \alpha}{2\,\beta}\right)\,, \tag{7.4.27}$$

where $A = \exp(-\delta\,(2\,\beta \ln 2 + \alpha)/(2\, c_0\, c_T))\, \exp(s_0)$ and $\lambda = (2\, c_T - 2\, c_R - \delta)/(2\, c_0\, c_T)$ are constants. We choose s_0 such that $\int_{-\infty}^{\infty} U(x)\,\mathrm{d}x = 1$, that is, $U(x)$ is a probability density function.

7.4.3 Interpretation of results

Having constructed a steady-state approximation for the solution (7.4.27), and noted at the start of Section 7.4 that presuming the existence of a steady state is a significant assumption, we now analyse the conditions under which such a solution may be expected to be relevant. Necessary conditions for a steady solution, $U(x)$, of the continuum model (7.4.5)-(7.4.8) to be a distribution are that U must have a maximum at a point $\widehat{x}$ where

$U'(\widehat{x}) = 0$ and $U(x) \geq 0$ for all $x \in \mathbb{R}$. From (7.4.24) it follows that $g(\widehat{x}) = c_T$ must be satisfied for the relevant values of $T_0 \geq 0$ and $R_0 \geq 0$ in order for U to be physical. We note that solutions $U(x) > 0$ for $x < 0$, representing a positive probability of telomeres with a negative length, should be regarded as unphysical. Such solutions arise in the model from telomeres with length $x < \mu$ leaving the system via t-loop formation at rate $k_c(x) > 0$ or via G4-stabilisation by RHPS4 (for $R_0 > 0$) at rate k_r, subsequently re-entering the system at length $x - \mu < 0$ due to telomere shortening by an amount μ. To avoid the occurrence of negative telomere length, one could change the functional forms of the transition rates of telomeres exiting the system in order to prevent short telomeres becoming further reduced in length. This could be established by adjusting the rate functions such that $k_c(x)$ is strictly zero if $L < \alpha$, and $k_r(x) = 0$ if $L < \mu$, for example. The probability of telomeres with length $x < 0$, however, is usually very small in our simulations and we interpret the occurrence of larger proportions of telomeres with negative length to reflect the presence of telomeres that have lost all their telomeric sequences and are no longer functional. Such telomeres would typically cause a cell to become senescent or undergo apoptosis. In the following, we aim to find conditions on T_0 and R_0 that must be satisfied to yield solutions with $U \gneqq 0$, and we require $\widehat{x} > 0$ for such solutions to be physical.

It follows immediately from the definitions (7.4.10) and (7.1.1) that $g(x) < \delta + c_R$ for all $x \in \mathbb{R}$, and hence by (7.4.24) and (7.4.25) $c_T < g < \delta + c_R$ in $x > \widehat{x}$, providing us with an upper bound for the number T_0 of telomerase molecules, from (7.4.11) and (7.4.25)

$$T_0 < T_{\max}(R_0) = \frac{\mu\, k_{\text{off}}}{k_{\text{on}}\, \rho} \left(\delta + \frac{k_f\, k_s\, k_r\, R_0}{k_u\, k_d + k_r\, (k_u + k_s\, R_0)} \right) . \tag{7.4.28}$$

For high concentrations of RHPS4 namely as $R_0 \to \infty$ we find $T_{\max}(R_0) \to T_{\max}^{\infty} = \mu\, k_{\text{off}}\, (\delta + k_f) / (\rho\, k_{\text{on}})$.

By the same reasoning as for (7.4.28) and since, from (7.4.10), $g(x) > c_R$ for all $x \in \mathbb{R}$, we find $c_T > c_R$, which provides us with a lower bound on T_0, namely

$$T_0 > T_{\min}(R_0) = \frac{\mu\, k_{\text{off}}}{k_{\text{on}}\, \rho}\, \frac{k_f\, k_s\, k_r\, R_0}{k_u\, k_d + k_r\, (k_u + k_s\, R_0)} . \tag{7.4.29}$$

Thus, for high concentrations of RHPS4 ($R_0 \to \infty$) we find $T_{\min}(R_0) \to T_{\min}^{\infty} = \mu\, k_{\text{off}}\, k_f/(\rho\, k_{\text{on}})$. For physical solutions $U(x)$, $\widehat{x}$ needs to be positive, and since g is monotonic increasing in x, we need $g(\widehat{x}) > g(0) = c_R + \delta/(1+\mathrm{e}^{\alpha/\beta})$, which provides a larger lower bound on T_0 than (7.4.29), namely

$$T_0 > \tilde{T}_{\min}(R_0) = T_{\min}(R_0) + \frac{\mu\, k_{\text{off}}\, \delta}{k_{\text{on}}\, \rho\, (1+\mathrm{e}^{\alpha/\beta})} = \frac{\mu\, k_{\text{off}}}{k_{\text{on}}\, \rho} \left(\frac{k_f\, k_s\, k_r\, R_0}{k_u\, k_d + k_r\, (k_u + k_s\, R_0)} + \frac{\delta}{1+\mathrm{e}^{\alpha/\beta}} \right). \tag{7.4.30}$$

Note that in the limit $\beta \to 0$ $\tilde{T}_{\min} = T_{\min}$, however, in the figures below we use $\beta = 300$ bp.

For $\tilde{T}_{\min} < T_0 < T_{\max}$ we expect steady state solutions, for $T_0 < \tilde{T}_{\min}$ the amount of telomerase is insufficient and the telomere length decays causing the cell to become senescent or undergo apoptosis. For $T_0 > T_{\max}$ telomere length grows without limit.

Alternatively, we reformulate these inequalities to provide a lower and an upper bound on R_0, for given $T_0 > 0$, in a similar way to (7.4.28) and (7.4.29), that is,

$$R_0 > R_{\min}(T_0) = \frac{k_u\, (k_{\text{on}}\, \rho\, T_0 - \mu\, k_{\text{off}}\, \delta)\, (k_d + k_r)}{k_s\, k_r\, \left[\mu\, k_{\text{off}}\, (k_f + \delta) - k_{\text{on}}\, \rho\, T_0\right]}, \tag{7.4.31}$$

and

$$R_0 < R_{\max}(T_0) = \frac{k_{\text{on}}\, \rho\, T_0\, k_u\, (k_d + k_r)}{k_s\, k_r\, (\mu\, k_{\text{off}}\, k_f - k_{\text{on}}\, \rho\, T_0)}, \tag{7.4.32}$$

for $\mu\, k_{\text{off}}\, k_f > k_{\text{on}}\, \rho\, T_0$ ($T_0 < T_{\min}^{\infty}$), respectively. Note that $R_{\max}(T_0) \to \infty$ as $T_0 \to (T_{\min}^{\infty})^-$ and there is no upper bound on R_0 for $T_0 > T_{\min}^{\infty}$. Similarly, equation (7.4.31) is valid for $T_0 < T_{\max}^{\infty}$ only (where $R_{\min}(T_0) > 0$) with $R_{\min}(T_0) \to \infty$ as $T_0 \to (T_{\max}^{\infty})^-$, and no physical solutions $U(x)$ exist for $T_0 \geq T_{\max}^{\infty}$, as $T_{\max}^{\infty} > T_{\max}(R_0)$ (note $k_f < c_R$) for all $R_0 \geq 0$. If $\delta > k_f$, there is a range of telomerase concentrations, $T_{\min}^{\infty} < T_0 < T_{\max}(0) = \mu\, k_{\text{off}}\, \delta/(\rho\, k_{\text{on}})$, where no upper bound on R_0 exists for solutions $U(x)$ to be physical, and this is not true for $\delta < k_f$, because then $T_{\max}(0) < T_{\min}^{\infty}$.

The lower and upper bounds on R_0 or T_0 can be used to plot (T_0,R_0)-regions of parameter space where steady physical solutions $U(x)$ exist, we can then identify the effects of changes of telomerase and RHPS4 concentrations in the cell. Examples with different values of δ are given in Figure 7.18 to illustrate the cases $\delta < k_f$ and $\delta > k_f$, where we choose

$\rho = 2.5 \times 10^{-1}$ nt·s^{-1} to illustrate the shape of these regions (lower values of ρ result in unphysically large values of $T^{\infty}_{\min}$, for example $T^{\infty}_{\min} > 10^5$ for $\rho = 4.5 \times 10^{-3}$ nt·s^{-1}). For $\delta > k_f$, there exists a range of telomerase concentrations where a steady state solution exists no matter how large or how small the concentration of RHPS4 is. For $\delta < k_f$, the region of steady state solutions is much smaller, hence more care for the regulation of telomerase and/or RHPS4 is required.

Figure 7.19 illustrates the mean telomere length, $\widehat{L}$, of telomeres, $k_c(x)\,U(x) + k_r\,C(x)$, exiting the system at steady state as a function of R_0 for four different numbers of telomerase molecules, T_0, and $\rho = 2.5 \times 10^{-1}$ nt·s^{-1} and $\delta > k_f$ as in the right plot of Figure 7.18. The mean telomere length $\widehat{L}$ has been computed using the same formula as in (7.3.18) and tends to $-\infty$ for large values of R_0 and small values of T_0, and to $+\infty$ for small values of R_0 and large values of T_0.

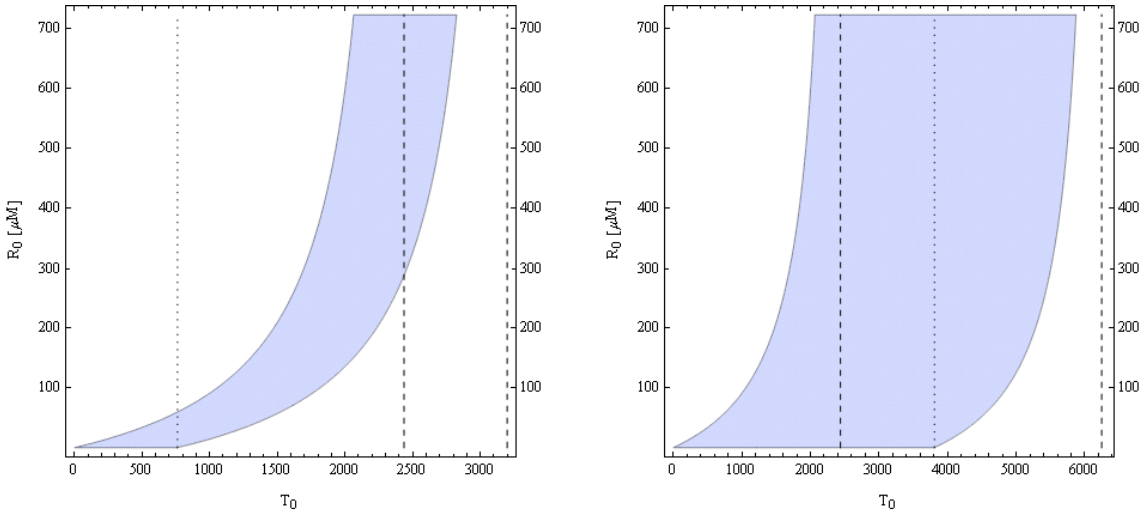

Figure 7.18: T_0-R_0-regions where physical steady state solutions U exist, plotted for $\delta = 5 \times 10^{-3}$ s^{-1} ($\delta < k_f$, left plot) and $\delta = 2.5 \times 10^{-2}$ s^{-1} ($\delta > k_f$, right plot). The rate of telomerase-induced telomere synthesis is $\rho = 2.5 \times 10^{-1}$ nt·s^{-1}. The lower ($T_{\min}(R_0)$) and upper ($T_{\max}(R_0)$) bounds on T_0 are defined by (7.4.29) and (7.4.28), respectively, and there is no visible difference between the lower bound $T_{\min}(R_0)$ and the larger lower bound $\tilde{T}_{\min}(R_0)$, defined by (7.4.30). The dotted line indicates the upper bound on telomerase, $T_{\max}(0)$, for the case of no drug and the two dashed lines in each plot indicate the lower ($T^{\infty}_{\min}$) and upper ($T^{\infty}_{\max}$) bounds on telomerase for large concentrations of RHPS4, where the values of $\tilde{T}^{\infty}_{\min}$ could not be distinguished from the values of $T^{\infty}_{\min}$ in these plots and are not shown.

We now plot the approximate telomere length distributions of telomeres in each of the four states and of telomeres leaving the system at steady state, where we chose $\rho = 4.5 \times 10^{-3}$ nt·s^{-1} (as in Section 7.3) or $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}, and vary the numbers T_0 of telomerase molecules (see Figures 7.20 and 7.21). All other model parameters are

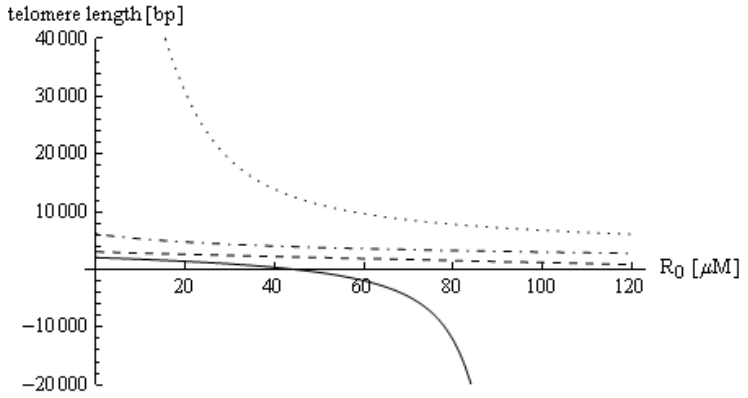

Figure 7.19: The mean telomere length $\widehat{L}$ of telomeres $k_c(x)\,U(x) + k_r\,C(x)$ exiting the state U with rate $k_c(x)$ or the state C with rate k_r at steady state per unit time, plotted as a function of the concentration of RHPS4, R_0, and for four different numbers of telomerase molecules, $T_0 = 1000$ (solid line), $T_0 = 2000$ (dashed line), $T_0 = 3000$ (dot-dashed line) and $T_0 = 4000$ (dotted line). The simulations correspond to the right plot in Figure 7.18, where $\delta = 2.5 \times 10^{-2}$ s^{-1} and $\rho = 2.5 \times 10^{-1}$ nt·s^{-1}. In all cases $T^{\infty}_{\min} = 2437$, $T_{\max}(0) = 3808$ and $T^{\infty}_{\max} = 6245$.

estimated as stated in Section 7.3 and 7.4. The bounds on T_0 for physical solutions $U(x)$ at $R_0 = 0$, that is

$$\tilde{T}_{\min}(0) = \frac{\mu\, k_{\text{off}}\, \delta}{k_{\text{on}}\, \rho\, (1 + \mathrm{e}^{\alpha/\beta})} < T_0 < T_{\max}(0)\,, \tag{7.4.33}$$

are $2 \leq T_0 \leq 423$ and $1 \leq T_0 \leq 30$ for $\rho = 4.5 \times 10^{-3}$ nt·s^{-1} and $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}, respectively. Note that the closer T_0 approaches $T_{\max}(0)$, the more positively skewed the distribution becomes. The mean telomere length $\widehat{L}$ of telomeres leaving the system at steady state is given in the caption of Figure 7.21 with its respective standard deviations, which are computed using the probability density function

$$\widehat{p}(x) = \frac{k_c(x)\,U(x) + k_r\,C(x)}{\int_{-\infty}^{\infty} k_c(x)\,U(x) + k_r\,C(x)\,\mathrm{d}x}\,. \tag{7.4.34}$$

The value $\rho = 6.287 \times 10^{-2}$ nt·s^{-1} has been chosen in the simulations, as the telomere length distribution for telomeres leaving the system for $T_0 = 25$ is in good agreement ($\widehat{L} = 3,440 \pm 1,516$ nt) with experimental data from HeLa cells (compare Figures 7.21 and 7.1). Figure 7.22 shows contour plots of $\widehat{L}$ as a function of T_0 and R_0, and Figures 7.23 and 7.24 show $\widehat{L}$ as a function of T_0 for $R_0 = 0$, and as a function of R_0 for chosen values of T_0, respectively.

Whereas an increase of T_0 leads to more positively skewed telomere length distributions, an increase in R_0 causes telomere length distributions becoming increasingly negatively

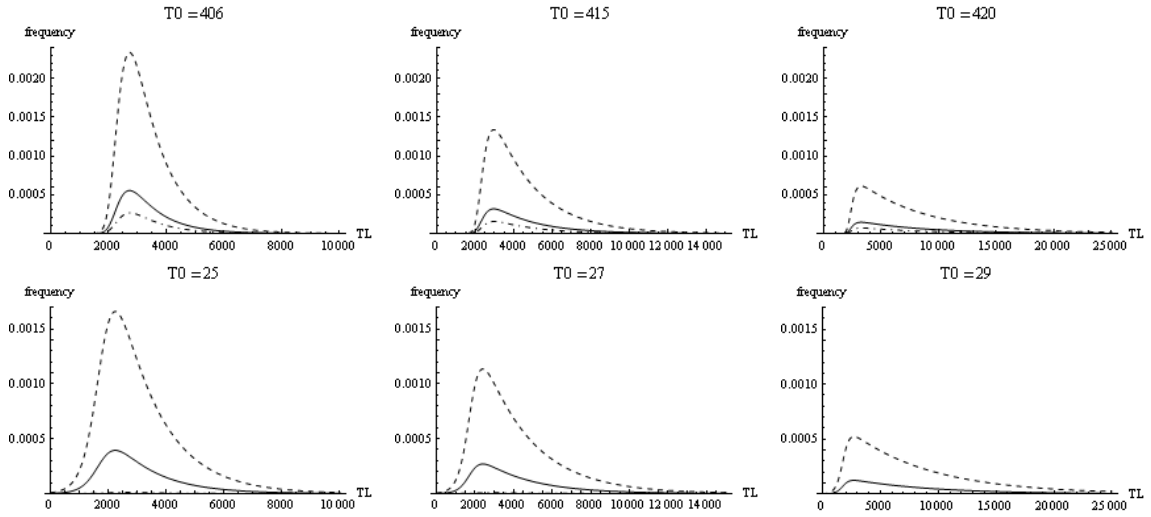

Figure 7.20: Plots of steady state curves $U(x)$, $B(x)$ and $G(x)$ against telomere length (represented by a solid, dot-dashed and dashed line, respectively) for two different values $\rho = 4.5 \times 10^{-3}$ nt·s^{-1} (top), $\rho = 6.287 \times 10^{-2}$ nt·s^{-1} (bottom) and varying numbers of telomerase molecules (top: $T_0 = 406, 415, 420$; bottom: $T_0 = 25, 27, 29$), where in each case we assume that $\delta = 5 \times 10^{-5}$ s^{-1}, that there is no RHPS4 ($R_0 = 0$) and telomeres shorten by $\mu = 45$ nt between each round of replication. The x-axis represents telomere length (TL) in units of basepairs. Note that there are different scales used for the x-axes in the plots.

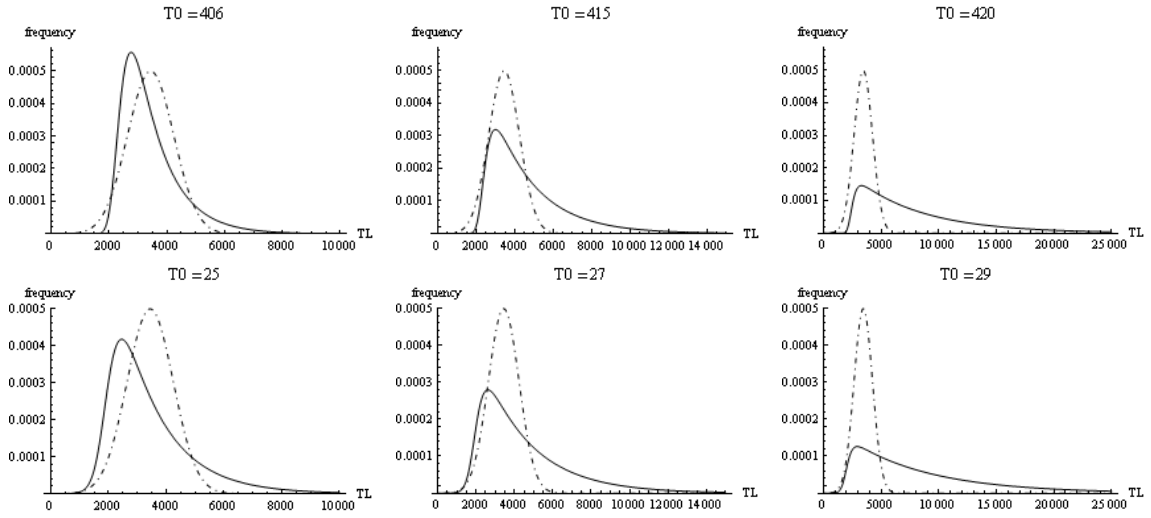

Figure 7.21: Plots of telomere length distributions $p\,(x)$ (normal distribution with $L_0 = 3440$ bp and $\sigma = 800$ bp, dot-dashed line) and $k_c(x)\,U(x) + k_r\,C(x)$ (telomeres exiting the state U with rate $k_c(x)$ and the state C with rate k_r, solid line) at steady state for two different values $\rho = 4.5 \times 10^{-3}$ nt·s^{-1} (top), $\rho = 6.287 \times 10^{-2}$ nt·s^{-1} (bottom) and varying numbers of T_0. The values of the mean telomere length $\widehat{L}$ are 3453 ± 1076 nt, 4679 ± 2233 nt, 7893 ± 4804 nt (top) and 3440 ± 1516 nt, 4549 ± 2560 nt, 8133 ± 5257 nt (bottom). Telomeres are assumed to shorten by $\mu = 45$ nt between each round of replication. The x-axis represents telomere length (TL) in units of basepairs in this plot.

skewed (see Figure 7.25 for a plot of telomere length distributions for different values of R_0). The negative skewness is predominantly caused by large numbers of telomeres forming a complex with RHPS4, whose length is overall shorter than the length of telomeres leaving the system when they are in the open form, as shown in Figure 7.26.

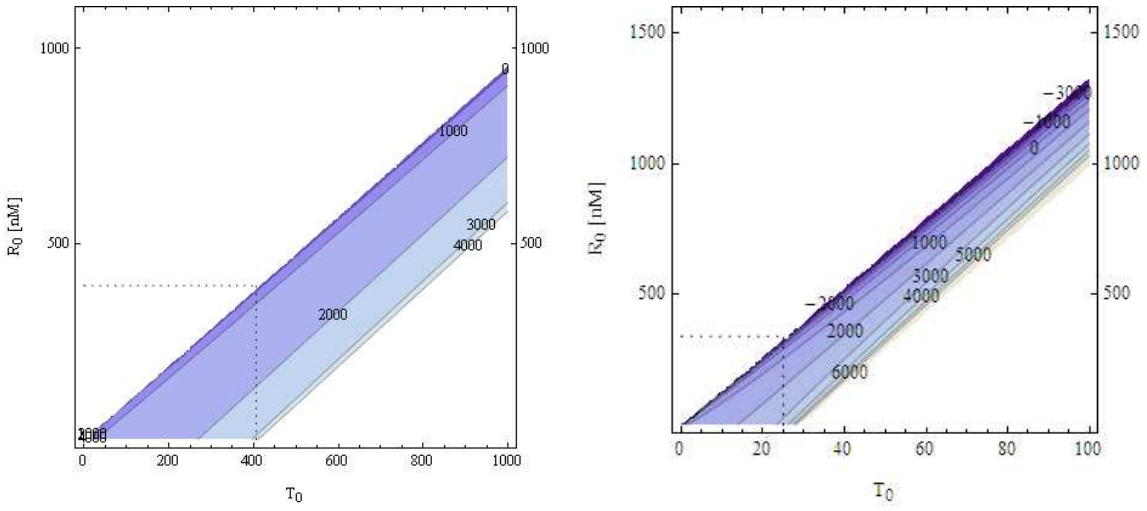

Figure 7.22: Contour plots of the mean telomere length $\widehat{L}$ in (T_0, R_0) space, left plot for $\rho = 4.5 \times 10^{-3}$ nt·s^{-1}, and right plot for $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}, where $\delta = 5 \times 10^{-5}$ s^{-1} in both cases. The dotted lines in each plot indicate the best approximation for T_0 such that $\widehat{L} = L_0$ ($T_0 = 406$ and $T_0 = 25$ for the left and the right plot, respectively), and the according upper limit $R_{\max}(T_0)$.

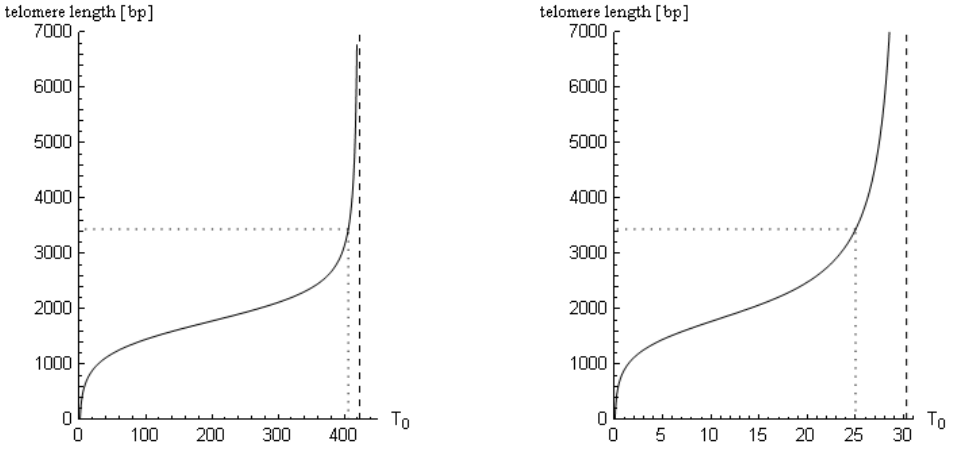

Figure 7.23: A plot of the mean telomere length $\widehat{L}$ of telomeres $k_c(x)\,U(x) + k_r\,C(x)$ exiting the system per unit time as a function of the number T_0 of telomerase molecules for the case of no drug ($R_0 = 0$) in the system (left plot: $\rho = 4.5 \times 10^{-3}$ nt·s^{-1}, right plot: $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}). The dashed lines in each plot represent the estimated bounds $T_{\max}(R_0)$ on T_0. The simulations correspond to the left and right plot in Figure 7.22, where $\delta = 5 \times 10^{-5}$ s^{-1} in both cases. The dotted lines in each plot indicate the value L_0 and according values T_0.

7.5 Summary of results and discussion

We have presented three simple models describing the telomere length dynamics in the S/G$_2$ phase where telomeres are exposed to a G4 binding agent (RHPS4). Model (i) accounts for the mechanism of telomeres forming a complex with RHPS4 in a folded G4 form, but not simultaneously a complex with the enzyme telomerase in an open form allowing for telomere elongation. Model (ii) is an extension of (i) which allows for t-

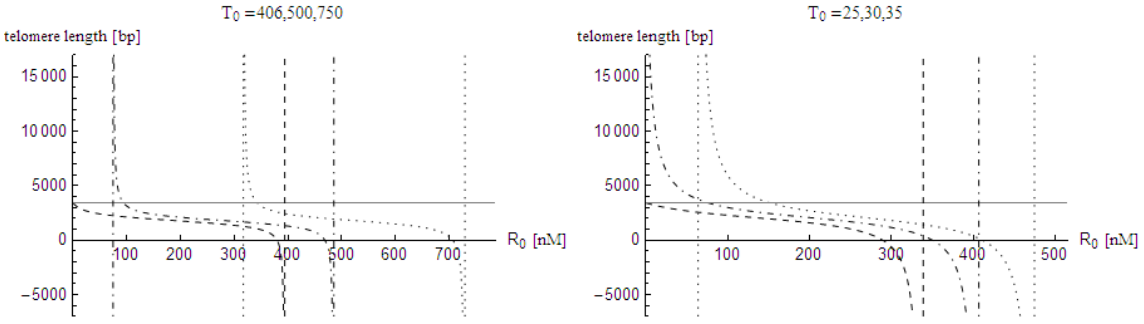

Figure 7.24: A plot of the mean telomere length $\widehat{L}$ of telomeres $k_c(x)\,U(x) + k_r\,C(x)$ exiting the system per unit time as a function of the concentration R_0 of RHPS4 for three different numbers of telomerase molecules, $T_0 = 406, 500, 750$ (left plot: $\rho = 4.5 \times 10^{-3}$ nt·s^{-1}) and $T_0 = 25, 30, 35$ (right plot: $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}), each represented by the dashed, dot-dashed and dotted line, respectively. Vertical lines show the corresponding asymptotes to these curves for both low or high concentrations of RHPS4. The simulations correspond to the left and right plot in Figure 7.22, where $\delta = 5 \times 10^{-5}$ s^{-1} in both cases. The solid, gray line in each plot indicates the value L_0.

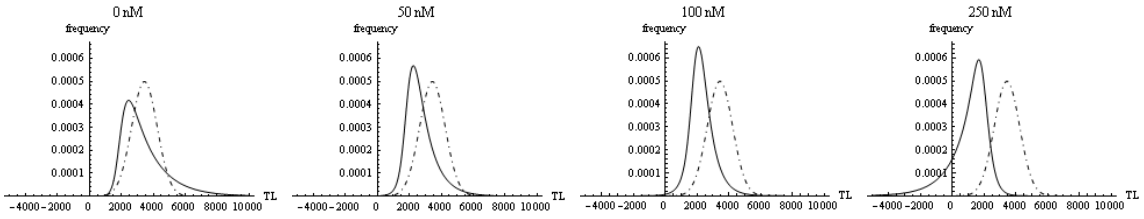

Figure 7.25: Plots of telomere length distributions $p\,(x)$ (dot-dashed line) and $k_c(x)\,U(x) + k_r\,C(x)$ at steady state of the system (7.4.1)-(7.4.4) (solid line) per unit time for varying concentrations of RHPS4 ($R_0 = 0, 50, 100, 250$ nM). In all cases $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}, $\delta = 5 \times 10^{-5}$ s^{-1} and $T_0 = 25$. The x-axis represents telomere length (TL) in units of basepairs.

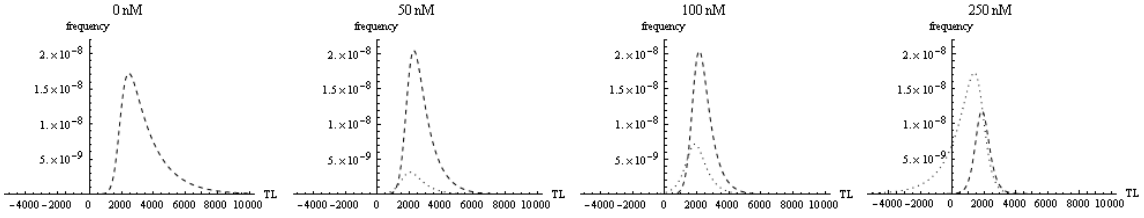

Figure 7.26: Plots of telomere length distributions $k_c(x)\,U(x)$ (dashed line) and $k_r\,C(x)$ (dotted line), leaving the system at steady state per unit time, for varying concentrations of RHPS4 ($R_0 = 0, 50, 100, 250$ nM). In all cases $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}, $\delta = 5 \times 10^{-5}$ s^{-1} and $T_0 = 25$. The x-axis represents telomere length (TL) in units of basepairs.

loop formation of open forms and the influx and efflux of telomeres into and out of the system, respectively. Model (ii) was used for the simulation of several telomere duplication events. Models (i) and (ii) describe a population of telomeres undergoing a single replication event. Model (iii) is a closed system, where telomeres shorten due to the end-replication problem, and possibly C-strand resection at the beginning of the next S phase each time they leave the cell cycle, and then re-enter the system at the S/G_2 level in the cell cycle being in an open form. Model (iii) can be used to predict steady state telomere

length distributions for a population of continuously cycling cells, as a function of RHPS4 and telomerase concentrations.

We assume in our models that t-loops and G4 structures are the key inhibitors of telomerase, and telomerase-induced telomere elongation. Constant telomere shortening due to the end-replication problem and C-strand resection, described in our model (iii) by a constant loss term, μ, complete the mechanisms that determine the shape of telomere length distributions. We modelled telomere elongation as a length-dependent process in that shorter telomeres are more likely to remain in an open state and be extendible by telomerase than longer telomeres, as longer telomeres have a higher tendency to coil up and form t-loops (described by a sigmoidal rate function of telomere length that increases with rising telomere length), which inhibits telomerase binding.

We further assume that telomeres only exit the cell cycle by t-loop formation, unless telomeric G4 structures are locked by RHPS4, when telomeres leave the G_2 phase of the cell cycle in this state and may unfold either at a later stage or at the beginning of the next cell cycle. We also suppose that the concentrations of telomerase and RHPS4 are much larger than respective concentrations of telomere-bound telomerase and RHPS4, and that these concentrations do not change over cell generations. Our results from simulations of different telomere states in the system are consistent with this latter assumption.

Our models (i)-(iii) describe the overall dynamics of telomere length in a cell, but do not account for the differences in telomere lengths at leading and lagging telomeric strands. Leading and lagging strands are both elongated by telomerase indiscriminately [158], but their overhang length evolves differently throughout the S/G_2 phase. The extent to which telomerase activity and C-strand resection contribute to the elongation of leading and lagging overhangs is not yet fully resolved [37].

We have estimated most of the kinetic parameters in each system by using experimental results from the literature, and estimated the remaining parameters in order to reproduce the experimental results of HeLa cells obtained by Canela et al. [26]. Our model predictions agree well with the experimental telomere length distribution of HeLa cells (compare Figures 7.1 and 7.21) and suggests a low concentration of about $T_0 = 25$ telomerase molecules per nucleus. In the literature the telomerase processivity parameter ρ is

given in the range $1.2 - 7.7 \times 10^{-3}$ nt·s^{-1}. We investigated the sensitivity of the results to ρ within this range. For small ρ we obtained narrower distributions of telomere lengths (and larger values of T_0) different to the experimental data. We therefore fitted ρ using a larger value, $\rho = 6.287 \times 10^{-2}$ nt·s^{-1}, to describe the experimental data. There is nearly no visible change in the telomere length distribution for smaller numbers (≤ 1000) of T_0 during one S/G$_2$ phase as illustrated by simulations of model (ii), shown in Figure 7.9, but visible changes occur after 25 generations, as shown in Figure 7.15 with $T_0 = 500$.

Note that we used a large value, $T_0 = 5000$, in model (ii) to simulate telomere length maintenance, where the choice of a larger value of ρ would yield much smaller values for the number of telomerase molecules, that is $T_0 \approx 120$ for $\rho = 6.287$ nt·s^{-1} at telomere length equilibrium. Furthermore, the T_0-R_0-regions of physical telomere length distributions shown in Figure 7.18 for model (ii) are for a value of ρ that is large ($\rho = 2.5 \times 10^{-1}$ nt·s^{-1}), this value has been chosen to illustrate the shape of these regions. For smaller (and more realistic) values of ρ, one is likely to find more elementary T_0-R_0-regions in the form of a band in the T_0-R_0-plane as illustrated in Figure 7.22.

Slightly smaller values of ρ with corresponding larger values of T_0 can produce telomere length distributions of similar shape. Having more knowledge of the values ρ, therefore, will help us determine the number of telomerase molecules in the nucleus more accurately. The small value of $T_0 = 25$ is, however, consistent with the average value of about 20-50 telomerase molecules per HEK-293 (human embryonic kidney) nuclei measured by Cohen et al. [31], which is the only quantitative measurement on telomerase levels in cells we are aware of in the literature.

Our simulations of conditions on T_0 and R_0 for physical solutions showed that the range of values $T_0 > 25$ reproduces well the experimental results in the literature, and telomeres grow unboundedly in length for values of T_0 larger than 30. In contrast, telomeres shorten beyond physical lengths (or indefinitely) if we use drug concentrations larger than $\sim$100 nM of RHPS4 (see Figure 7.25). Hence, the telomere equilibrium state is rather sensitive to the amount of telomerase and RHPS4 in the system: larger doses of RHPS4 lead to continuous telomere shortening, whilst telomerase overexpression, on the other hand, induces continuous telomere lengthening. The steady telomere length increase we

found for large T_0 is consistent with the findings of Cristofari and Lingner [36], who observed elongation of telomeres at a constant rate in super-telomerase HeLa cells for over 50 population doublings. Hence, telomere length homeostasis cannot be established with telomerase overexpression.

Our model suggests two different effects of the treatment with RHPS4 which are dependent on the drug concentration used: low concentrations reduce telomere length, but do not impair the equilibrium of the system, and high concentrations destabilise the system leading to chromosome degradation and senescence and/or cell death. Addition or overexpression of telomerase can counteract telomere degradation; however, telomerase addition should be carefully regulated to maintain the system in equilibrium and not trigger unlimited telomere elongation. Note that the upper limit for R_0, when telomere equilibrium can still be maintained, is probably lower than we predicted: The threshold value for telomere length as determined by the Hayflick limit [56], triggering senescense or apoptosis pathways in a cell terminating cell proliferation, is likely to be based on the shortest telomere in the cell, not the average length [1, 58]. As far as we are aware, telomere length frequency histograms for cells exposed to different drug concentrations of RHPS4 have not yet been experimentally determined, but would help to confirm the simulated effects of RHPS4 on telomere length distributions of cancer cells in our models.

CHAPTER 8

Discussion

8.1 Conclusions

The study of anti-cancer agents has not yet generated an efficient remedy for the most common cancer types. Much research is still necessary to understand the processes involved in cancer development as well as the mechanisms by which specific anti-cancer treatments work. The ability of new drugs to induce growth inhibition, changes in cell-cycle progression and cell death is commonly evaluated by tissue culture experiments, and comparing the results to observations from more detailed *in vitro* experiments on cell components can bring understanding into the actual mechanisms in living cells. We used a systems biology approach to investigate how the drug RHPS4 changes the cell cycle dynamics over short periods of time and at medium drug concentrations ($50 - 1000$ nM), and studied the potential effects of RHPS4 on telomere length using mathematical modelling. Our method involves not only the development and analysis of mathematical models, but also new experimental design, parameter estimation, and model evaluation techniques to gain detailed insight into the dynamics of cells in drug assays.

We have reviewed the current cell biology and mathematical modelling that is relevant to our work in Chapter 1, inclusive of the biological background and mathematical modelling of cell cycle dynamics, and the biology of telomeres including telomere replication, DNA damage repair, telomere capping mechanisms and the function of the enzyme telomerase in telomere length regulation. Furthermore, we have discussed existing math-

ematical models of telomere length dynamics and provided a brief overview of the effects of RHPS4 on cancer cells.

In Chapter 2, we developed and analysed compartmental ODE, PDE and DDE models of the cell cycle. A simple three-compartment ODE model of cycling, resting and dead cells was used as a basis for more refined models, such as a five- and a seven-compartment ODE models, which distinguish between the single phases of the cell cycle in greater detail than the three-compartment model. We additionally refined the three-compartment ODE model by splitting the compartment of cycling cells (containing the phases $S/G_2/M$) into several sub-compartments, as we considered this compartment more important with respect to the effects of RHPS4 than other compartments. We showed that the multi-compartmental ODE model turns into a PDE model as the number of sub-compartments becomes large. The PDE model, in turn, can be reduced to a DDE model containing only one variable for the time t and a delay term specifying the residence time in the $S/G_2/M$ phase. We focused on the analysis of two models, the five-compartment ODE model and the three-compartment DDE model and simulated typical solution behaviour. Whereas in the ODE model oscillations decay exponentially with time, in the DDE model oscillatory behaviour may occur in exponentially growing solutions depending on the values of the delay term and the model parameters. Additionally, we analysed a mathematical expression for the residence time of cells in a certain phase of the cell cycle which is equivalent to the average waiting time of cells within a model compartment.

In Chapter 3, we introduced the concept of model and parameter identifiability and reviewed two methods to check compartmental models for structural identifiability: the Taylor series method, which we applied to the five-compartment model in Chapter 2, and the transfer function method, which we used to show that the seven-compartment cell-cycle model in Chapter 5 is identifiable. Model parameters typically have to be estimated when fitting a model to experimental data. Thus, we presented the concept of practical identifiability, where we showed how one can obtain information on the accuracy of parameter estimates, by computing standard errors, confidence intervals and correlation coefficients.

We conducted our own experimental work, as described in detail in Chapter 4. We treated

colorectal cancer cells with three different concentrations of the drug RHPS4 over a period of 1 to 10 days with six replicates each. We analysed cells with respect to their growth behaviour, cell viability and cell cycle progress, we measured the DNA content of cells by use of flow-cytometry. We collected a larger number of data than is typical for standard biological experiments and analysed cells daily, as we aimed to obtain more information on the cell cycle dynamics and time-dependent behaviour of cells. In order to minimise disruption of natural growth we did not perform sub-culturing, but adjusted the seeding densities for each of the experiments, so that we achieved around 70-80% confluence at the day of analysis. This procedure limited the period of observation to a maximum of 10 days as extreme cell densities and low nutrient supply constrained cell growth.

We found that RHPS4 affects cells in a concentration-dependent manner, causing growth inhibition and cell death, finding that cell death is delayed, and the delay is larger for lower drug concentrations. Additionally, we observed a transient drop of S phase cells around day 4 for treated cells. Most of our results are consistent with work by Johnson et al. [65], who found growth inhibition when they exposed colorectal cancer cells to different concentrations of RHPS4 and incubated cells over a period of 21 days. However, their cells were sub-cultured about four times in their experiments, whenever they became near confluent, and their observations of inhibition of cell death at higher drug concentrations is contrary to our findings, that is, increasing cell death with time and drug concentration. This may be due to the different tissue culture techniques used, or the different measurement of apoptotic cells: whereas Johnson et al. [65] used the number of cells with pre-G_1 DNA content as an indicator of cell death, we measured apoptosis by trypan-blue dye exclusion and considered the proportions of pre-G_1 cells as a measure of late apoptosis only.

In Chapter 5, we developed a novel cell-cycle model distinguishing between viable and dead cells of the same DNA content, and distinguishing between early and late apoptosis. Our model additionally allows for different dynamics of cell death intensity, distinguishing between constant intensity, linear and sigmoidal increase in death rates setting in at a certain time point after drug treatment. We analysed our model for its solution behaviour and derived an asymptotic expression for the doubling time of cycling cells.

We then formulated a statistical model describing the experimental data we collected in the tissue culture lab. The observable quantities are a linear combination of the state variables in our model, and for this reason we confirmed *a priori* that the parameters were identifiable. To fit our model to data, we used a maximum likelihood approach. We introduced an information criterion [3], which is used for model selection, and gave examples of statistical tests that can be employed to check the model residuals against assumptions we made when writing down our statistical model, that is, identically and independently distributed errors that are log-normally distributed.

In Chapter 6, we fitted our models to experimental data, using the stochastic, global optimisation routine SRES [116], which converged rapidly and the convergence was relatively stable. We identified parameters for each of the models developed in Chapter 5 and selected the model best fitting the experimental data. We performed model reduction, where suggested parameter values were practically zero and thus caused identifiability problems. Our model fits captured the experimental data well, especially for control cells and the lower drug concentration of 50 nM. We performed a sensitivity analysis on our model and identified the influence of model parameters on the state variables of the system. We evaluated the accuracy of model fit using statistical hypothesis tests and found that there is only a little systematic variability in the data for higher drug concentrations that the model cannot capture. Furthermore, the degree of correlation between most of the model parameters is low, stronger correlated parameters had larger variances, meaning greater uncertainty in the estimates.

More detailed measurements could enhance our modelling results. It may be possible (by flowcytometric analysis) to perform experiments distinguishing between viable and trypan-blue stained cells of the same DNA content, which would allow us to obtain a higher accuracy for our parameter fits. Such experimental data would also be useful in confirming our model results. Measurements of senescent proportions for each day of analysis would enable us to extend our cell cycle model and include a compartment of non-cycling (senescent) cells into our model. Measurements that distinguish between G_0 and G_1 cells could additionally provide us with more accurate results when fitting cell-cycle models.

Our results of fitting the model to data showed that cells die rapidly after a concentration-dependent delay which decreases with higher doses when treated with RHPS4, and cell death largely occurs from the G_2/M phase for the lower drug concentration and from the G_0/G_1 phase for higher drug concentrations. We interpreted the phase difference as a reflection of differences in the cell cycle events the drug disrupts, blocking successful completion of mitotic events for lower drug concentrations and entry into the S phase for higher concentrations. In addition, we interpreted the time delay of cell death as the number of cell divisions that cells have undergone after the treatment with RHPS4 and before the onset of cell death, which we explained by the mechanisms of the drug impairing telomere stability and inducing apoptotic pathways.

In Chapter 7, we investigated the effects of the G-quadruplex interactive agent RHPS4 on telomere erosion, and developed mathematical models of telomere length dynamics. In particular, we considered telomere length dynamics during the S/G_2 phases, when telomerase replenishes telomeric sequences of open t-loops but not the G-quadruplex structure at each cell division. We determined steady-state length distributions, over small and large numbers of cell generations, with and without treatment with RHPS4. In particular, the simple formula for the telomere length distribution, (7.4.27) is a notable result of that chapter.

We derived approximate analytical expressions, and simulated numerically steady-states of length distributions of telomeres, at the time when a cell exits a cell cycle. We analysed the effects of different levels of telomerase and different concentrations of RHPS4 on telomere length during the S/G_2 phases and considered how these effects evolve over large numbers of cell generations. Our models predict positively skewed length distributions of telomeres and our model predictions are in good agreement with experimental observations in HeLa cells. Moreover, our predicted value of the number of telomerase molecules in the nucleus is consistent with experimental findings in cancer cells.

We found that too high concentrations of telomerase can lead to ongoing telomere lengthening, which is consistent with observations from experiments with telomerase-positive cells. We derived regions of different telomerase and RHPS4 levels that provide physically plausible solutions to our model of telomere length dynamics over large numbers

of cell generations and showed that telomerase expression must be strictly regulated for telomere length maintenance. Too high concentrations of RHPS4 can lead to progressive telomere erosion; we estimate that drug concentrations larger than ≈100 nM impair the equilibrium of the system leading to continuous telomere shortening and triggering senescence and apoptosis.

In summary, the main results of Chapter 7 show how telomerase acts as a regulator for telomere length, and how RHPS4 can disrupt this regulation as illustrated by Figures 7.23 and 7.24. At small concentrations, RHPS4 has little effect, but there is a critical concentration above which telomerase is unable to maintain a steady state, and rapid shortening occurs. The region of telomerase-RHPS4 parameter space where steady states exist has also been determined and is illustrated in Figure 7.18.

Altogether, our work presents novel mathematical models for cancer cells, describing cell-cycle dynamics and the length dynamics of telomeres. We investigated how the drug RHPS4 affected these dynamics with respect to its time- and dose-dependent modulation. Our synthesis of experimental work and mathematical modelling enhances the understanding of the mechanism of action of RHPS4 and may represent a valuable contribution to the research into anti-cancer drugs.

8.2 Future work

More direct extensions of this thesis are given by investigating our mathematical models of telomere length dynamics further. It would be interesting to derive an estimate of the average time it takes for a telomere to leave the system via t-loop or G-quadruplex formation, this could be modelled using Monte-Carlo simulations, following single telomeres in the system. Comparison of this average time with the residence times of control cells and cells treated with different drug concentrations would enable us to test whether the rate parameters k_r and δ of telomeres leaving the system respectively *via* G-quadruplex stabilisation and t-loop formation have been estimated accurately.

We approximated analytically the changes in the mean telomere length over one cell generation in Section 7.3. The case of large numbers of cell generations in Section 7.4 involves

complicated integrals and is left as interesting work for future investigations. We derived an approximate solution for the steady-state length distribution of telomeres over large numbers of generations using a first-order ODE approximation of the delay differential equation in Section 7.4.2. It would be interesting to investigate further the solution of a second-order approximation to obtain solutions of higher accuracy. Furthermore, measurements of telomere length distributions of telomerase-positive cells at different population doublings and for varying concentrations of telomerase and the drug RHPS4 would be useful in comparing our model predictions to experimental results.

Our work on the dynamics of telomere length in telomerase-positive cancer cells included the role of G-quadruplex and t-loop formation in telomere stability. There are other players, such as the shelterin protein POT1, which contribute to telomere maintenance and have not been considered in this thesis. POT1 is involved in several processes at the telomere end, which might be an interesting avenue to follow. Human POT1 binds single-stranded DNA at the D-loop of a telomere and at the telomeric overhang *in vitro*. The minimal binding site of POT1 comprises nine nucleotides, where POT1 binding is preferred close to the 3′ end of the telomeric overhang [82]. Two distinct functions of the protein have been identified, depending on the position of POT1 at the 3′-overhang: if POT1 is bound at the very end of the overhang (leaving less than 8 nt free at the 3′ end), telomerase cannot extend the telomere [76]. On the other hand, the formation of G-quadruplex structures requires that POT1 is not bound to the terminal four telomeric sequences involved in G-quadruplex formation. Since G-quadruplexes form spontaneously at the end of the telomere and are in dynamic equilibrium with unfolded or partially unfolded forms, POT1 binding of unfolding structures may trap telomeres in the open form [156].

We therefore suggest further expansion of our models (i)-(iii) of telomere length dynamics from Chapter 7. Incorporating the mechanisms of POT1 in telomere replication, that is, telomerase and G-quadruplex inhibition, may help predict how different levels of POT1 affect telomere length dynamics during the S/G_2 phases of the cell cycle. We extend our model by introducing new telomere states, which represent open forms that do not allow for transitions to telomerase-bound and/or G-quadruplex states: POT1 bound at the very end of the 3′-overhang inhibits both the formation of G-quadruplexes and telom-

erase binding at telomeric ends. However, POT1 bound to one or more of the last four telomere sequences but not to the last one, inhibits G-quadruplex formation, but provides optimal conditions for telomere elongation by telomerase. A model diagram showing the extension of model (i) from Section 7.2 is presented in Figure 8.1.

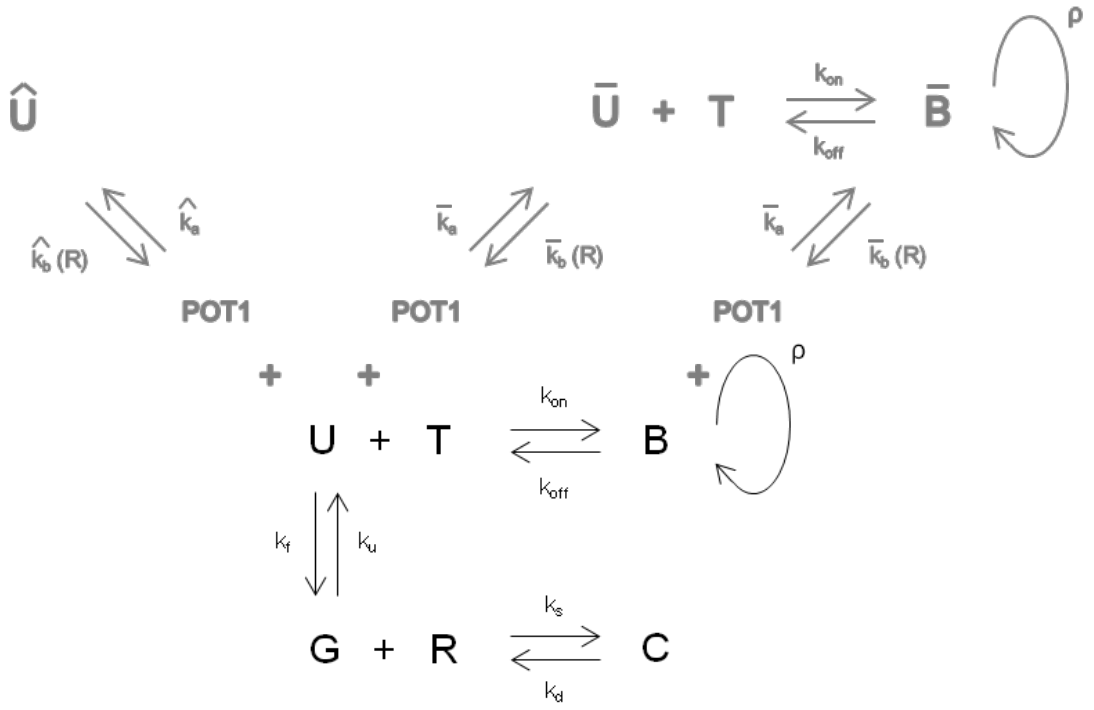

Figure 8.1: An extension of model (i) from Section 7.2, containing the unbound, bound, G-quadruplex and RHPS4-G-quadruplex complex telomeric states U, B, G, C, respectively, shows the effects of the shelterin protein POT1 on telomere dynamics during the S/G_2 phases of the cell cycle. POT1 binds single-stranded DNA of telomeric 3'-overhangs, where $\overline{U}$ denotes telomeres with POT1 bound to the terminal four telomeric sequences, but leaving 8 nt free at the 3' end, and $\widehat{U}$ denotes telomeres with POT1 bound to the end of the 3'-overhang (leaving less than 8 nt free at the 3' end). Free telomerase (T) binds both open forms U and $\overline{U}$, that is U+T$\rightarrow$B and $\overline{U} + T \rightarrow \overline{B}$, respectively. However, telomerase activity is inhibited by POT1-bound telomeres of the form $\widehat{U}$. Both of the POT1-bound forms, $\overline{U}$ and $\widehat{U}$, inhibit the formation of G-quadruplexes (G). Kinetics for each reaction in this model are described by their rate constants k and telomerase elongation occurs at rate ρ. The drug RHPS4 (R) stabilises G-quadruplexes (G+R$\rightarrow$C), but also causes POT1 delocalisation from telomeres, which is indicated in the Figure by the transition rates of POT1 dissociation ($\widehat{k}_b$ and $\overline{k}_b$) being functions of RHPS4 concentration, R.

Simulations of this extended model for different levels of POT1 binding might enable us to explain ambiguous findings in the literature. Colgin et al. [32] found increased telomere lengthening dependent on telomerase activity when POT1 was overexpressed in a human fibrosarcoma cell line, meaning that POT1 can act as an telomerase inhibitor and a telomerase attractor at the same time. When cells were chronically exposed to RHPS4 at a concentration of 1 μM, RHPS4 triggered removal of POT1 from telomeres [118]. However, the effect of the drug was suppressed by overexpression of POT1 in human trans-

formed fibroblasts and melanoma cells. Hence, it may be interesting to predict the effects of RHPS4 for varying POT1 levels, in particular POT1 being overexpressed or suppressed in telomerase-positive cells.

References

[1] P. Abdallah, P. Luciano, K. W. Runge, M. Lisby, V. Geli, E. Gilson, and M. T. Teixeira. A two-step model for senescence triggered by a single critically short telomere. *Nature Cell Biology*, 11:988–U174, 2009.

[2] B. D. Aguda. Kick-starting the cell cycle: From growth-factor stimulation to initiation of DNA replication. *Chaos*, 11:269–276, 2001.

[3] H. Akaike. New look at statistical-model identification. *IEEE Transactions on Automatic Control*, AC19:716–723, 1974.

[4] B. Alberts. *Molecular Biology of the Cell*. Garland Science, New York, 5th edition, 2008.

[5] T. Antal, K. B. Blagoev, S. A. Trugman, and S. Redner. Aging and immortality in a cell proliferation model. *Journal of Theoretical Biology*, 248:411–417, 2007.

[6] O. Arino, M. Kimmel, and G. F. Webb. Mathematical-modeling of the loss of telomere sequences. *Journal of Theoretical Biology*, 177:45–57, 1995.

[7] N. Arkus. A mathematical model of cellular apoptosis and senescence through the dynamics of telomere loss. *Journal of Theoretical Biology*, 235:13–32, 2005.

[8] M. Ashyraliyev, J. Jaeger, and J. G. Blom. Parameter estimation and determinability analysis applied to drosophila gap gene circuits. *BMC Systems Biology*, 2:83, 2008.

[9] M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, and J. G. Blom. Systems biology: parameter estimation for biochemical models. *FEBS Journal*, 276:886–902, 2009.

[10] R. C. Aster, C. H. Thurber, B. Borchers, and ScienceDirect. *Parameter estimation and inverse problems*. International geophysics series. Elsevier Academic Press, Amsterdam, 2005.

[11] S. Audoly, L. D'Angio, M. P. Saccomani, and C. Cobelli. Global identifiability of linear compartmental models - a computer algebra algorithm. *IEEE Transactions on Biomedical Engineering*, 45:36–47, 1998.

[12] 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*, 69:1673–1690, 2007.

[13] 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*, 47:295–312, 2003.

[14] 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 anticancer drug paclitaxel. *Journal of Mathematical Biology*, 49:329–357, 2004.

[15] B. Basse, B. C. Baguley, E. S. Marshall, G. C. Wake, and D. J. N. Wall. Modelling the flow cytometric data obtained from unperturbed human tumour cell lines: parameter fitting and comparison. *Bulletin of Mathematical Biology*, 67:1153–1153, 2005.

[16] B. Basse, W. R. Joseph, E. S. Marshall, and B. C. Baguley. Analysis of radiation-induced changes to human melanoma cultures using a mathematical model. *Cell Proliferation*, 43:139–146, 2010.

[17] D. M. Bates and D. G. Watts. *Nonlinear regression analysis and its applications*. Wiley series in probability and statistics. Wiley, New York; Chichester, 2007.

[18] E. J. Bedrick and C. L. Tsai. Model selection for multivariate regression in small samples. *Biometrics*, 50:226–231, 1994.

[19] R. Bellman and K. J. Astrom. On structural identifiability. *Mathematical Biosciences*, 7:329–339, 1970.

[20] G. Bellu, M. P. Saccomani, S. Audoly, and L. D'Angio. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine*, 88:52–61, 2007.

[21] L. Bendix, P. B. Horn, U. B. Jensen, I. Rubelj, and S. Kolvraa. The load of short telomeres, estimated by a new method, Universal STELA, correlates with number of senescent cells. *Aging Cell*, 9:383–397, 2010.

[22] A. E. Bilsland, C. J. Cairney, and W. N. Keith. Targeting the telomere and shelterin complex for cancer therapy: current views and future perspectives. *Journal of Cellular and Molecular Medicine*, 15:179–186, 2011.

[23] E. H. Blackburn. Switching and signaling at the telomere. *Cell*, 106:661–673, 2001.

[24] K. B. Blagoev. Cell proliferation in the presence of telomerase. *PloS One*, 4:e4622, 2009.

[25] M. G. Brattain, W. D. Fine, F. M. Khaled, J. Thompson, and D. E. Brattain. Heterogeneity of malignant-cells from a human colonic-carcinoma. *Cancer Research*, 41: 1751–1756, 1981.

[26] A. Canela, E. Vera, P. Klatt, and M. A. Blasco. High-throughput telomere length quantification by FISH and its application to human population studies. *Proceedings of the National Academy of Sciences of the United States of America*, 104:5300–5305, 2007.

[27] G. Cedersund and J. Roll. Systems biology: model based evaluation and comparison of potential explanations for given biological data. *FEBS Journal*, 276:903–922, 2009.

[28] A. J. Cesare and R. R. Reddel. Alternative lengthening of telomeres: models, mechanisms and implications. *Nature Reviews Genetics*, 11:319–330, 2010.

[29] M. K. Cheng, C. Modi, J. C. Cookson, I. Hutchinson, R. A. Heald, A. J. McCarroll, S. Missailidis, F. Tanious, W. D. Wilson, J. L. Mergny, C. A. Laughton, and M. F. G. Stevens. Antitumor polycyclic acridines. 20. Search for DNA quadruplex binding selectivity in a series of 8,13-dimethylquino 4,3,2-kl acridinium salts: Telomere-targeted agents. *Journal of Medicinal Chemistry*, 51:963–975, 2008.

[30] C. Cobelli and J. J. Distefano. Parameter and structural identifiability concepts and ambiguities - a critical-review and analysis. *American Journal of Physiology*, 239:R7–R24, 1980.

[31] S. B. Cohen, M. E. Graham, G. O. Lovrecz, N. Bache, P. J. Robinson, and R. R. Reddel. Protein composition of catalytically active human telomerase from immortal cells. *Science*, 315:1850–1853, 2007.

[32] L. M. Colgin, K. Baran, P. Baumann, T. R. Cech, and R. R. Reddel. Human POT1 facilitates telomere elongation by telomerase. *Current Biology*, 13:942–946, 2003.

[33] M. A. Collins. A quasicontinuum approximation for solitons in an atomic chain. *Chemical Physics Letters*, 77:342–347, 1981.

[34] J. C. Cookson, F. P. Dai, V. Smith, R. A. Heald, C. A. Laughton, M. F. G. Stevens, and A. M. Burger. Pharmacodynamics of the G-quadruplex-stabilizing telomerase inhibitor 3,11-difluoro-6,8,13-trimethyl-8H-quino 4,3,2-kl acridinium methosulfate (RHPS4) in vitro: activity in human tumor cells correlates with telomere length and can be enhanced, or antagonized, with cytotoxic agents. *Molecular Pharmacology*, 68: 1551–1558, 2005.

[35] J. C. Cookson, R. A. Heald, and M. F. G. Stevens. Antitumor polycyclic acridines. 17. Synthesis and pharmaceutical profiles of pentacyclic acridinium salts designed to destabilize telomeric integrity. *Journal of Medicinal Chemistry*, 48:7198–7207, 2005.

[36] G. Cristofari and J. Lingner. Telomere length homeostasis requires that telomerase levels are limiting. *EMBO Journal*, 25:565–574, 2006.

[37] X. Y. Dai, C. H. Huang, A. Bhusari, S. Sampathi, K. Schubert, and W. H. Chai. Molecular steps of G-overhang generation at human telomeres and its function in chromosome end protection. *EMBO Journal*, 29:2788–2801, 2010.

[38] T. de Lange. T-loops and the origin of telomeres. *Nature Reviews Molecular Cell Biology*, 5:323–329, 2004.

[39] T. de Lange. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & Development*, 19:2100–2110, 2005.

[40] T. de Lange. How telomeres solve the end-protection problem. *Science*, 326:948–952, 2009.

[41] Y. B. Deng, S. S. Chan, and S. Chang. Telomere dysfunction and tumour suppression: the senescence connection. *Nature Reviews Cancer*, 8:450–458, 2008.

[42] G. P. Dimri, X. H. Lee, G. Basile, M. Acosta, C. Scott, C. Roskelley, E. E. Medrano, M. Linskens, I. Rubelj, O. Pereirasmith, M. Peacocke, and J. Campisi. A biomarker that identifies senescent human-cells in culture and in aging skin in-vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 92:9363–9367, 1995.

[43] P. V. Dmitriev and Y. S. Vassetzky. Analysis of telomeric DNA: current approaches and methods. *Russian Journal of Developmental Biology*, 40:125–144, 2009.

[44] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley Series in Probability and Statistics. Wiley, New York; Chichester, 3rd edition, 1998.

[45] S. L. Fink and B. T. Cookson. Apoptosis, pyroptosis, and necrosis: mechanistic description of dead and dying eukaryotic cells. *Infection and Immunity*, 73:1907–1916, 2005.

[46] H. Fuss, W. Dubitzky, C. S. Downes, and M. J. Kurth. Mathematical models of cell cycle regulation. *Briefings in Bioinformatics*, 6:163–177, 2005.

[47] D. T. Gillespie. Exact stochastic simulation of coupled chemical-reactions. *Journal of Physical Chemistry*, 81:2340–2361, 1977.

[48] A. Golubev, S. Khrustalev, and A. Butov. An in silico investigation into the causes of telomere length heterogeneity and its implications for the Hayflick limit. *Journal of Theoretical Biology*, 225:153–170, 2003.

[49] S. M. Gowan, R. Heald, M. F. G. Stevens, and L. R. Kelland. Potent inhibition of telomerase by small-molecule pentacyclic acridines capable of interacting with G-quadruplexes. *Molecular Pharmacology*, 60:981–988, 2001.

[50] J. Grasman, H. M. Salomons, and S. Verhulst. Stochastic modeling of length-dependent telomere shortening in corvus monedula. *Journal of Theoretical Biology*, 282:1–6, 2011.

[51] R. A. Greenberg. Telomeres, crisis and cancer. *Current Molecular Medicine*, 5:213–218, 2005.

[52] J. D. Griffith, L. Comeau, S. Rosenfield, R. M. Stansel, A. Bianchi, H. Moss, and T. de Lange. Mammalian telomeres end in a large duplex loop. *Cell*, 97:503–514, 1999.

[53] M. Gunaratnam, O. Greciano, C. Martins, A. P. Reszka, C. M. Schultes, H. Morjani, J.-F. Riou, and S. Neidle. Mechanism of acridine-based telomerase inhibition and telomere shortening. *Biochemical Pharmacology*, 74:679–689, 2007.

[54] X. Guo, Y. Deng, Y. Lin, W. Cosme-Blanco, S. Chan, H. He, G. Yuan, E. J. Brown, and S. Chang. Dysfunctional telomeres activate an ATM-ATR-dependent DNA damage response to suppress tumorigenesis. *EMBO Journal*, 26:4709–4719, 2007.

[55] L. Hayflick. The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research*, 37:614–636, 1965.

[56] L. Hayflick. Cell biology of aging. *Journal of Investigative Dermatology*, 73:8–14, 1979.

[57] R. A. Heald, C. Modi, J. C. Cookson, I. Hutchinson, C. A. Laughton, S. M. Gowan, L. R. Kelland, and M. F. G. Stevens. Antitumor polycyclic acridines. 8. Synthesis and telomerase-inhibitory activity of methylated pentacyclic acridinium salts. *Journal of Medicinal Chemistry*, 45:590–597, 2002.

[58] M. T. Hemann, M. A. Strong, L. Y. Hao, and C. W. Greider. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell*, 107:67–77, 2001.

[59] E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, and E. H. Blackburn. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine guanine base pairs. *Cell*, 51:899–908, 1987.

[60] K. E. Huffman, S. D. Levene, V. M. Tesmer, J. W. Shay, and W. E. Wright. Telomere shortening is proportional to the size of the G-rich telomeric 3 ′-overhang. *Journal of Biological Chemistry*, 275:19719–19722, 2000.

[61] S. Itzkovitz, L. I. Shlush, D. Gluck, and K. Skorecki. Population mixture model for nonlinear telomere dynamics. *Physical Review E*, 78, 2008.

[62] J. A. Jacquez. *Compartmental analysis in biology and medicine*. BioMedware, Ann Arbor, MI, 3rd edition, 1996.

[63] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman. Global cancer statistics. *CA: a Cancer Journal for Clinicians*, 61:69–90, 2011.

[64] L. A. Johnson. *Integrative Biological Studies of Anti-tumour Agents*. PhD thesis, University of Nottingham, 2009. Available at http://etheses.nottingham.ac.uk.

[65] L. A. Johnson, H. M. Byrne, A. E. Willis, and C. A. Laughton. An integrative biological approach to the analysis of tissue culture data: application to the antitumour agent RHPS4. *Integrative Biology*, 3:843–849, 2011.

[66] L. W. Johnson and R. D. Riess. *Numerical analysis*. Addison-Wesley, Reading, Mass; London, 2nd edition, 1982.

[67] M. Kajstura, H. D. Halicka, J. Pryjma, and Z. Darzynkiewicz. Discontinuous fragmentation of nuclear DNA during apoptosis revealed by discrete "sub-G(1)" peaks on DNA content histograms. *Cytometry Part A*, 71A:125–131, 2007.

[68] N. W. Kim, M. A. Piatyszek, K. R. Prowse, C. B. Harley, M. D. West, P. L. C. Ho, G. M. Coviello, W. E. Wright, S. L. Weinrich, and J. W. Shay. Specific association of human telomerase activity with immortal cells and cancer. *Science*, 266:2011–2015, 1994.

[69] M. Kimura, R. C. Stone, S. C. Hunt, J. Skurnick, X. Lu, X. Cao, C. B. Harley, and A. Aviv. Measurement of telomere length by the southern blot analysis of terminal restriction fragment lengths. *Nature Protocols*, 5:1596–1607, 2010.

[70] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[71] S. H. Kleinstein, D. Bottino, G. S. Lett, A. Georgieva, and R. Sarangapani. Nonuniform sampling for global optimization of kinetic rate constants in biological pathways. *Proceedings of the 2006 Winter Simulation Conference, Vols 1-5*, pages 1611–1616, 2006.

[72] S. L. B. König, A. C. Evans, and J. L. Huppert. Seven essential questions on G-quadruplexes. *BioMolecular Concepts*, 1:197–213, 2010.

[73] A. L. Koch. Logarithm in biology. 1. Mechanisms generating log-normal distribution exactly. *Journal of Theoretical Biology*, 12:276–290, 1966.

[74] A. Kowald. Possible mechanisms for the regulation of telomere length. *Journal of Molecular Biology*, 273:814–825, 1997.

[75] H. C. Kraemer. Estimation and hypothesis testing problems for correlation-coefficients. *Psychometrika*, 40:473–485, 1975.

[76] M. Lei, A. J. Zaug, E. R. Podell, and T. R. Cech. Switching human telomerase on and off with hPOT1 protein in vitro. *Journal of Biological Chemistry*, 280:20449–20456, 2005.

[77] C. Leonetti, S. Amodei, C. D'Angelo, A. Rizzo, B. Benassi, A. Antonelli, R. Elli, M. F. G. Stevens, M. D'Incalci, G. Zupi, and A. Biroccio. Biological activity of the G-quadruplex ligand RHPS4 (3,11-difluoro-6,8,13-trimethyl-8H-quino 4,3,2-kl acridinium methosulfate) is associated with telomere capping alteration. *Molecular Pharmacology*, 66:1138–1146, 2004.

[78] C. Leonetti, M. Scarsella, G. Riggio, A. Rizzo, E. Salvati, M. D'Incalci, L. Staszewsky, R. Frapolli, M. F. Stevens, A. Stoppacciaro, M. Mottolese, B. Antoniani, E. Gilson, G. Zupi, and A. Biroccio. G-Quadruplex Ligand RHPS4 potentiates the antitumor activity of camptothecins in preclinical models of solid tumors. *Clinical Cancer Research*, 14:7284–7291, 2008.

[79] M. Z. Levy, R. C. Allsopp, A. B. Futcher, C. W. Greider, and C. B. Harley. Telomere end-replication problem and cell aging. *Journal of Molecular Biology*, 225:951–960, 1992.

[80] E. Limpert, W. A. Stahel, and M. Abbt. Log-normal distributions across the sciences: keys and clues. *Bioscience*, 51:341–352, 2001.

[81] H. J. Lipps and D. Rhodes. G-quadruplex structures: in vivo evidence and function. *Trends in Cell Biology*, 19:414–422, 2009.

[82] D. Loayza, H. Parsons, J. Donigian, K. Hoke, and T. de Lange. DNA binding features of human POT1 - A nonamer 5 '-TAGGGTTAG-3 ' minimal binding site, sequence specificity, and internal binding to multimeric sites. *Journal of Biological Chemistry*, 279:13241–13248, 2004.

[83] V. L. Makarov, Y. Hirose, and J. P. Langmore. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell*, 88:657–666, 1997.

[84] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11:431–441, 1963.

[85] U. M. Martens, E. A. Chavez, S. S. S. Poon, C. Schmoor, and P. M. Landsdorp. Accumulation of short telomeres in human fibroblasts prior to replicative senescence. *Experimental Cell Research*, 256:291–299, 2000.

[86] K. Masutomi, E. Y. Yu, S. Khurts, I. Ben-Porath, J. L. Currier, G. B. Metz, M. W. Brooks, S. Kaneko, S. Murakami, J. A. DeCaprio, R. A. Weinberg, S. A. Stewart, and W. C. Hahn. Telomerase maintains telomere structure in normal human cells. *Cell*, 114:241–253, 2003.

[87] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer. Bionumbers-the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38: D750–D753, 2010.

[88] C. G. Moles, P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Research*, 13:2467–2474, 2003.

[89] D. Monchaud and M.-P. Teulade-Fichou. A hitchhiker's guide to G-quadruplex ligands. *Organic & Biomolecular Chemistry*, 6:627–636, 2008.

[90] K. Monier, J. C. G. Armas, S. Etteldorf, P. Ghazal, and K. F. Sullivan. Annexation of the interchromosomal space during viral infection. *Nature Cell Biology*, 2:661–665, 2000.

[91] I. K. Moon and M. B. Jarstfer. The human telomere and its relationship to human disease, therapy, and tissue engineering. *Frontiers in Bioscience*, 12:4595–4620, 2007.

[92] S. Neidle. Human telomeric G-quadruplex: The current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS Journal*, 277:1118–1125, 2010.

[93] B. Novak and J. J. Tyson. Modelling the controls of the eukaryotic cell cycle. *Biochemical Society Transactions*, 31:1526–1529, 2003.

[94] B. Novak and J. J. Tyson. A model for restriction point control of the mammalian cell cycle. *Journal of Theoretical Biology*, 230:563–579, 2004.

[95] R. Nunez. DNA measurement and cell cycle analysis by flow cytometry. *Current Issues in Molecular Biology*, 3:67–70, 2001.

[96] L. Oganesian and J. Karlseder. Telomeric armor: the layers of end protection. *Journal of Cell Science*, 122:4013–4025, 2009.

[97] P. Olofsson. Can telomere shortening explain sigmoidal growth curves? *Journal of Biological Dynamics*, 4:527–538, 2010.

[98] P. Olofsson and A. A. Bertuch. Modeling growth and telomere dynamics in saccharomyces cerevisiae. *Journal of Theoretical Biology*, 263:353–359, 2010.

[99] P. Olofsson and M. Kimmel. Stochastic models of telomere shortening. *Mathematical Biosciences*, 158:75–92, 1999.

[100] J. op den Buijs, P. P. J. van den Bosch, M. W. J. M. Musters, and N. A. W. van Riel. Mathematical modeling confirms the length-dependency of telomere shortening. *Mechanisms of ageing and development*, 125:437–44, 2004.

[101] W. Palm and T. de Lange. How shelterin protects mammalian telomeres. *Annual Review of Genetics*, 42:301–334, 2008.

[102] J. C. Panetta and J. Adam. A mathematical-model of cycle-specific chemotherapy. *Mathematical and Computer Modelling*, 22:67–82, 1995.

[103] H. Pohjanpalo. System identifiability based on power-series expansion of solution. *Mathematical Biosciences*, 41:21–33, 1978.

[104] S. S. S. Poon, U. M. Martens, R. K. Ward, and P. M. Lansdorp. Telomere length measurements using digital fluorescence microscopy. *Cytometry*, 36:267–278, 1999.

[105] R. D. Portugal, M. G. P. Land, and B. F. Svaiter. A computational model for telomere-dependent cell-replicative aging. *Biosystems*, 91:262–267, 2008.

[106] C. J. Proctor and T. B. L. Kirkwood. Modelling telomere shortening and the role of oxidative stress. *Mechanisms of Ageing and Development*, 123:351–363, 2002.

[107] C. J. Proctor and T. B. L. Kirkwood. Modelling cellular senescence as a result of telomere state. *Aging Cell*, 2:151–157, 2003.

[108] Q. Qi. *Mathematical modelling of telomere dynamics*. PhD thesis, University of Nottingham, 2012. Available at http://etheses.nottingham.ac.uk.

[109] S. Rajaraman, J. Choi, P. Cheung, V. Beaudry, H. Moore, and S. E. Artandi. Telomere uncapping in progenitor cells with critical telomere shortening is coupled to S-phase progression in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 104:17747–17752, 2007.

[110] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmueller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25:1923–1929, 2009.

[111] C. Riccardi and I. Nicoletti. Analysis of apoptosis by propidium iodide staining and flow cytometry. *Nature Protocols*, 1:1458–1461, 2006.

[112] A. Rizzo, E. Salvati, M. Porru, C. D'Angelo, M. F. Stevens, M. D'Incalci, C. Leonetti, E. Gilson, G. Zupi, and A. Biroccio. Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. *Nucleic Acids Research*, 37:5353–5364, 2009.

[113] I. A. Rodriguez-Brenes and C. S. Peskin. Quantitative theory of telomere length regulation and cellular senescence. *Proceedings of the National Academy of Sciences of the United States of America*, 107:5387–5392, 2010.

[114] P. Rosenau. Dynamics of nonlinear mass-spring chains near the continuum-limit. *Physics Letters A*, 118:222–227, 1986.

[115] I. Rubelj and Z. Vondracek. Stochastic mechanism of cellular aging - abrupt telomere shortening as a model for stochastic nature of cellular aging. *Journal of Theoretical Biology*, 197:425–438, 1999.

[116] T. P. Runarsson and X. Yao. Stochastic ranking for constrained evolutionary optimization. *IEEE Transactions on Evolutionary Computation*, 4:284–294, 2000.

[117] S. N. Saldanha, L. G. Andrews, and T. O. Tollefsbol. Assessment of telomere length and factors that contribute to its stability. *European Journal of Biochemistry*, 270:389–403, 2003.

[118] E. Salvati, C. Leonetti, A. Rizzo, M. Scarsella, M. Mottolese, R. Galati, I. Sperduti, M. F. G. Stevens, M. D'Incalci, M. Blasco, G. Chiorino, S. Bauwens, B. Horard, E. Gilson, A. Stoppacciaro, G. Zupi, and A. Biroccio. Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *Journal of Clinical Investigation*, 117:3236–3247, 2007.

[119] O. Samassekou, M. Gadji, R. Drouin, and J. Yan. Sizing the ends. Normal length of human telomeres. *Annals of Anatomy-Anatomischer Anzeiger*, 192:284–291, 2010.

[120] K. A. Schafer. The cell cycle: a review. *Veterinary Pathology*, 35:461–478, 1998.

[121] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps, and A. Pluckthun. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proceedings of the National Academy of Sciences of the United States of America*, 98:8572–8577, 2001.

[122] G. A. F. Seber and C. J. Wild. *Nonlinear regression*. Wiley series in probability and mathematical statistics. Wiley, New York; Chichester, 1989.

[123] E. Sherer, R. E. Hannemann, A. Rundell, and D. Ramkrishna. Analysis of resonance chemotherapy in leukemia treatment via multi-staged population balance models. *Journal of Theoretical Biology*, 240:648–661, 2006.

[124] E. Sherer, E. Tocce, R. E. Hannemann, A. E. Rundell, and D. Ramkrishna. Identification of age-structured models: cell cycle phase transitions. *Biotechnology and Bioengineering*, 99:960–974, 2008.

[125] I. A. Sidorov, K. S. Hirsch, C. B. Harley, and D. S. Dimitrov. Cancer treatment by telomerase inhibitors: predictions by a kinetic model. *Mathematical Biosciences*, 181: 209–221, 2003.

[126] J. A. Smith and L. Martin. Do cells cycle? *Proceedings of the National Academy of Sciences of the United States of America*, 70:1263–1267, 1973.

[127] J. R. Smith and R. G. Whitney. Intraclonal variation in proliferative potential of human-diploid fibroblasts - stochastic mechanism for cellular aging. *Science*, 207: 82–84, 1980.

[128] A. Smogorzewska and T. de Lange. Regulation of telomerase by telomeric proteins. *Annual Review of Biochemistry*, 73:177–208, 2004.

[129] P. D. Sozou and T. B. L. Kirkwood. A stochastic model of cell replicative senescence based on telomere shortening, oxidative stress, and somatic mutations in nuclear and mitochondrial DNA. *Journal of Theoretical Biology*, 213:573–586, 2001.

[130] R. M. Stansel, T. de Lange, and J. D. Griffith. T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang. *EMBO Journal*, 20:5532–5540, 2001.

[131] M. A. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69:730–737, 1974.

[132] M. A. Stephens. Asymptotic results for goodness-of-fit statistics with unknown parameters. *Annals of Statistics*, 4:357–369, 1976.

[133] Z. Tan. Intramitotic and intraclonal variation in proliferative potential of human diploid cells: Explained by telomere shortening. *Journal of Theoretical Biology*, 198: 259–268, 1999.

[134] J. Tang, Z.-y. Kan, Y. Yao, Q. Wang, Y.-h. Hao, and Z. Tan. G-quadruplex preferentially forms at the very 3' end of vertebrate telomeric DNA. *Nucleic Acids Research*, 36:1200–1208, 2008.

[135] T. Tauchi, K. Shin-ya, G. Sashida, M. Sumi, S. Okabe, J. H. Ohyashiki, and K. Ohyashiki. Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia. *Oncogene*, 25:5719–5725, 2006.

[136] M. T. Teixeira, M. Arneric, P. Sperisen, and J. Lingner. Telomere length homeostasis is achieved via a switch between telomerase-extendible and -nonextendible states. *Cell*, 117:323–335, 2004.

[137] J. J. Tyson. Modeling the cell-division cycle - cdc2 and cyclin interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 88:7328–7332, 1991.

[138] J. J. Tyson and B. Novak. Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions. *Journal of Theoretical Biology*, 210: 249–263, 2001.

[139] R. Venkatasubramanian, M. A. Henson, and N. S. Forbes. Integrating cell-cycle progression, drug penetration and energy metabolism to identify improved cancer therapeutic strategies. *Journal of Theoretical Biology*, 253:98–117, 2008.

[140] R. E. Verdun and J. Karlseder. Replication and protection of telomeres. *Nature*, 447: 924–931, 2007.

[141] I. Vermes, C. Haanen, and C. Reutelingsperger. Flow cytometry of apoptotic cell death. *Journal of Immunological Methods*, 243:167–190, 2000.

[142] T. von Zglinicki, A. Burkle, and T. B. L. Kirkwood. Stress, DNA damage and ageing - an integrative approach. *Experimental Gerontology*, 36:1049–1062, 2001.

[143] G. Wallweber, S. Gryaznov, K. Pongracz, and R. Pruzan. Interaction of human telomerase with its primer substrate. *Biochemistry*, 42:589–600, 2003.

[144] E. Walter and L. Pronzato. On the identifiability and distinguishability of nonlinear parametric models. *Mathematics and Computers in Simulation*, 42:125–134, 1996.

[145] F. Wang, E. R. Podell, A. J. Zaug, Y. Yang, P. Baciu, T. R. Cech, and M. Lei. The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature*, 445:506–510, 2007.

[146] J. A. D. Wattis. Approximations to solitary waves on lattices. 3. The monatomic lattice with second-neighbour interactions. *Journal of Physics A: Mathematical and General*, 29:8139–8157, 1996.

[147] D. J. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC, London, 2006.

[148] O. Wolkenhauer, D. Fell, P. De Meyts, N. Bluthgen, H. Herzel, N. Le Novere, T. Hofer, K. Schurrle, and I. van Leeuwen. Sysbiomed report: Advancing systems biology for medical applications. *IET Systems Biology*, 3:131–136, 2009.

[149] World Health Organisation. Cancer, 2008. [Online] (Updated February 2012) Available at: `http://www.who.int/mediacentre/factsheets/fs297/en/index.html` [Accessed 06 May 2012].

[150] W. E. Wright, D. Brasiskyte, M. A. Piatyszek, and J. W. Shay. Experimental elongation of telomeres extends the lifespan of immortal x normal cell hybrids. *EMBO Journal*, 15:1734–1741, 1996.

[151] P. Wu and T. de Lange. Human telomerase caught in the act. *Cell*, 138:432–434, 2009.

[152] Y. Xu. Chemistry in human telomere biology: structure, function and targeting of telomere DNA/RNA. *Chemical Society Reviews*, 40:2719–2740, 2011.

[153] L. Yang, Z. G. Han, W. R. MacLellan, J. N. Weiss, and Z. L. Qu. Linking cell division to cell growth in a spatiotemporal model of the cell cycle. *Journal of Theoretical Biology*, 241:120–133, 2006.

[154] Q. Yang, Y. L. Zheng, and C. C. Harris. POT1 and TRF2 cooperate to maintain telomeric integrity. *Molecular and Cellular Biology*, 25:1070–1080, 2005.

[155] A. M. Zahler, J. R. Williamson, T. R. Cech, and D. M. Prescott. Inhibition of telomerase by G-quartet DNA structures. *Nature*, 350:718–720, 1991.

[156] A. J. Zaug, E. R. Podell, and T. R. Cech. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 102:10864–10869, 2005.

[157] Y. Zhao, Z. Y. Kan, Z. X. Zeng, Y. H. Hao, H. Chen, and Z. Tan. Determining the folding and unfolding rate constants of nucleic acids by biosensor. Application to telomere G-quadruplex. *Journal of the American Chemical Society*, 126:13255–13264, 2004.

[158] Y. Zhao, A. J. Sfeir, Y. Zou, C. M. Buseman, T. T. Chow, J. W. Shay, and W. E. Wright. Telomere extension occurs at most chromosome ends and is uncoupled from fill-in in human cancer cells. *Cell*, 138:463–475, 2009.

Appendices

APPENDIX A

Example Codes for Identifiability

A.1 Taylor series method

(* Taylor series method of the five-compartment ODE model of Chapter 3.2 *)

Clear[kXY, kYZ, kZX, kXS, kXA, kYA, kZA, kSA];
Clear[v0, x_0, y_0, z_0, s_0, a_0];
Clear[subst, yDer, eqn, Soln, newSoln];

(* dv/dt = $A\,v$, $v(0) = v0$, $y = B\,v$ *)

$$\mathbf{A} = \begin{pmatrix} -\mathrm{kXY} - \mathrm{kXS} - \mathrm{kXA} & 0 & 2*\mathrm{kZX} & 0 & 0 \\ \mathrm{kXY} & -\mathrm{kYZ} - \mathrm{kYA} & 0 & 0 & 0 \\ 0 & \mathrm{kYZ} & -\mathrm{kZX} - \mathrm{kZA} & 0 & 0 \\ \mathrm{kXS} & 0 & 0 & -\mathrm{kSA} & 0 \\ \mathrm{kXA} & \mathrm{kYA} & \mathrm{kZA} & \mathrm{kSA} & 0 \end{pmatrix}; \quad \mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix};$$

v0 = {x_0, y_0, z_0, s_0, a_0};

(* Comparison of power series coefficients for two different parameter vectors *)

subst = {kXY->kXYb, kYZ->kYZb, kZX->kZXb, kXS->kXSb, kXA → kXAb, kYA → kYAb, kZA → kZAb, kSA->kSAb}

{kXY → kXYb, kYZ → kYZb, kZX → kZXb, kXS → kXSb, kXA → kXAb, kYA → kYAb, kZA → kZAb, kSA → kSAb}

k = 1;

yDer = B.A.v0

$\{(-\mathrm{kXA} - \mathrm{kXS} - \mathrm{kXY})x_0 + 2\mathrm{kZX}z_0, \mathrm{kXY}x_0 + (-\mathrm{kYA} - \mathrm{kYZ})y_0, \mathrm{kYZ}y_0 + (-\mathrm{kZA} - \mathrm{kZX})z_0, -\mathrm{kSA}s_0 + \mathrm{kXS}x_0, \mathrm{kSA}s_0 + \mathrm{kXA}x_0 +$
$+\mathrm{kYA}y_0 + \mathrm{kZA}z_0\}$

eqn = yDer − (yDer/.subst)

$\{(-\mathrm{kXA} - \mathrm{kXS} - \mathrm{kXY})x_0 - (-\mathrm{kXAb} - \mathrm{kXSb} - \mathrm{kXYb})x_0 + 2\mathrm{kZX}z_0 - 2\mathrm{kZXb}z_0, \mathrm{kXY}x_0 - \mathrm{kXYb}x_0 + (-\mathrm{kYA} - \mathrm{kYZ})y_0 -$
$-(-\mathrm{kYAb} - \mathrm{kYZb})y_0, \mathrm{kYZ}y_0 - \mathrm{kYZb}y_0 + (-\mathrm{kZA} - \mathrm{kZX})z_0 - (-\mathrm{kZAb} - \mathrm{kZXb})z_0, -\mathrm{kSA}s_0 + \mathrm{kSAb}s_0 + \mathrm{kXS}x_0 - \mathrm{kXSb}x_0,$
$\mathrm{kSA}s_0 - \mathrm{kSAb}s_0 + \mathrm{kXA}x_0 - \mathrm{kXAb}x_0 + \mathrm{kYA}y_0 - \mathrm{kYAb}y_0 + \mathrm{kZA}z_0 - \mathrm{kZAb}z_0\}$

Soln = Simplify[Solve[eqn == 0, {kXY, kYZ, kZX, kXS, kXA, kYA, kZA, kSA}]][[1]]

Solve::svars : Equations may not give solutions for all "solve" variables. ⟩⟩

$\left\{\mathrm{kXA} \to \frac{(-\mathrm{kSA}+\mathrm{kSAb})s_0+\mathrm{kXAb}x_0-\mathrm{kYA}y_0+\mathrm{kYAb}y_0-\mathrm{kZA}z_0+\mathrm{kZAb}z_0}{x_0}, \mathrm{kXY} \to \frac{\mathrm{kXYb}x_0+(\mathrm{kYA}-\mathrm{kYAb})y_0+(\mathrm{kZA}-\mathrm{kZAb})z_0}{x_0},\right.$
$\left.\mathrm{kYZ} \to \frac{\mathrm{kYZb}y_0+(\mathrm{kZA}-\mathrm{kZAb})z_0}{y_0}, \mathrm{kZX} \to \mathrm{kZXb}, \mathrm{kXS} \to \frac{(\mathrm{kSA}-\mathrm{kSAb})s_0+\mathrm{kXSb}x_0}{x_0}\right\}$

k = 2;

yDer = B.MatrixPower[A, k].v0

$$\Big\{(-kXA - kXS - kXY)^2 x_0 + 2kYZkZXy_0 + (2(-kXA - kXS - kXY)kZX + 2(-kZA - kZX)kZX)z_0,$$
$$((-kXA - kXS - kXY)kXY + kXY(-kYA - kYZ))x_0 + (-kYA - kYZ)^2 y_0 + 2kXYkZXz_0, kXYkYZx_0 +$$
$$+((-kYA - kYZ)kYZ + kYZ(-kZA - kZX))y_0 + (-kZA - kZX)^2 z_0, kSA^2 s_0 + (-kSAkXS + kXS(-kXA - kXS - kXY))x_0 +$$
$$+2kXSkZXz_0, -kSA^2 s_0 + (kSAkXS + kXA(-kXA - kXS - kXY) + kXYkYA)x_0 + (kYA(-kYA - kYZ) + kYZkZA)y_0 +$$
$$+(kZA(-kZA - kZX) + 2kXAkZX)z_0\}$$

eqn = yDer − (yDer/.subst)

$$\Big\{(-kXA - kXS - kXY)^2 x_0 - (-kXAb - kXSb - kXYb)^2 x_0 + 2kYZkZXy_0 - 2kYZbkZXby_0 +$$
$$+(2(-kXA - kXS - kXY)kZX + 2(-kZA - kZX)kZX)z_0 - (2(-kXAb - kXSb - kXYb)kZXb + 2(-kZAb - kZXb)kZXb)z_0,$$
$$((-kXA - kXS - kXY)kXY + kXY(-kYA - kYZ))x_0 - ((-kXAb - kXSb - kXYb)kXYb + kXYb(-kYAb - kYZb))x_0 +$$
$$+(-kYA - kYZ)^2 y_0 - (-kYAb - kYZb)^2 y_0 + 2kXYkZXz_0 - 2kXYbkZXbz_0, kXYkYZx_0 - kXYbkYZbx_0 +$$
$$+((-kYA - kYZ)kYZ + kYZ(-kZA - kZX))y_0 - ((-kYAb - kYZb)kYZb + kYZb(-kZAb - kZXb))y_0 + (-kZA - kZX)^2 z_0 -$$
$$-(-kZAb - kZXb)^2 z_0, kSA^2 s_0 - kSAb^2 s_0 + (-kSAkXS + kXS(-kXA - kXS - kXY))x_0 -$$
$$-(-kSAbkXSb + kXSb(-kXAb - kXSb - kXYb))x_0 + 2kXSkZXz_0 - 2kXSbkZXbz_0, -kSA^2 s_0 + kSAb^2 s_0 +$$
$$+(kSAkXS + kXA(-kXA - kXS - kXY) + kXYkYA)x_0 - (kSAbkXSb + kXAb(-kXAb - kXSb - kXYb) + kXYbkYAb)x_0 +$$
$$+(kYA(-kYA - kYZ) + kYZkZA)y_0 - (kYAb(-kYAb - kYZb) + kYZbkZAb)y_0 + (kZA(-kZA - kZX) + 2kXAkZX)z_0 -$$
$$-(kZAb(-kZAb - kZXb) + 2kXAbkZXb)z_0\}$$

eqn = Simplify[eqn/.Soln] (* simplify eqn by use of Soln from k=1*)

$$\Big\{0, -\tfrac{1}{x_0 y_0}\left((kYA - kYAb)y_0 + (kZA - kZAb)z_0\right)\left(kXYbx_0^2 + (kXAb + kXSb + kXYb - kYAb - kYZb)x_0 y_0 - 2kZXby_0 z_0\right),$$
$$\tfrac{(kZA - kZAb)\left(-kYZby_0^2 + kXYbx_0 z_0 + (-kYAb - kYZb + kZAb + kZXb)y_0 z_0\right)}{y_0},$$
$$\tfrac{(kSA - kSAb)\left(-kXSbx_0^2 + s_0((kSAb - kXAb - kXSb - kXYb)x_0 + 2kZXbz_0)\right)}{x_0},$$
$$\tfrac{1}{x_0}\Big((kSAkXSb - kSAbkXSb + kXYb(kYA - kYAb))x_0^2 + 2kZXbz_0((-kYA + kYAb)y_0 + (-kZA + kZAb)z_0) +$$
$$+x_0\Big(\Big(kXYbkYA + kXAb(kYA - kYAb) + kXSb(kYA - kYAb) - kXYbkYAb - kYAkYAb + kYAb^2 - kYAkYZb +$$
$$+kYAbkYZb + kYZbkZA - kYZbkZAb\Big)y_0 + (kZA - kZAb)(kXAb + kXSb + kXYb - kZAb - kZXb)z_0) -$$
$$-(kSA - kSAb)s_0((kSAb - kXAb - kXSb - kXYb)x_0 + 2kZXbz_0))\Big\}$$

newSoln = Simplify[Solve[eqn == 0, {kXY, kYZ, kZX, kXS, kXA, kYA, kZA, kSA}]][[1]]

Solve::svars : Equations may not give solutions for all "solve" variables. 〉〉

$$\{kYA \to kYAb, kSA \to kSAb, kZA \to kZAb\}$$

Soln = Simplify[Soln/.newSoln] (* use results from k=2 for results from k=1 *)

$$\{kXA \to kXAb, kXY \to kXYb, kYZ \to kYZb, kZX \to kZXb, kXS \to kXSb\}$$

Soln = Union[Soln, newSoln] (* combine all results from k=1 and k=2 *)

$$\{kSA \to kSAb, kXA \to kXAb, kXS \to kXSb, kXY \to kXYb, kYA \to kYAb, kYZ \to kYZb, kZA \to kZAb, kZX \to kZXb\}$$

(* All rate parameters are uniquely identifiable *)

A.2 Transfer function method

(* Transfer function method of the seven-compartment ODE model of Chapter 5 *)

```
Clear[kXY, kYZ, kZX, kX0, kY0, kZ0, kXA, kYA, kZA];
Clear[v0, X0, Y0, Z0, XX0, YY0, ZZ0, A0];
Clear[s, H, HI, Y, YCoeffList];
```

(*dv/dt = $Av, v(0) = v0, y = Bv$*)

$$A = \begin{pmatrix} -kXY - kX0 & 0 & 2*kZX & 0 & 0 & 0 & 0 \\ kXY & -kYZ - kY0 & 0 & 0 & 0 & 0 & 0 \\ 0 & kYZ & -kZX - kZ0 & 0 & 0 & 0 & 0 \\ kX0 & 0 & 0 & -kXA & 0 & 0 & 0 \\ 0 & kY0 & 0 & 0 & -kYA & 0 & 0 \\ 0 & 0 & kZ0 & 0 & 0 & -kZA & 0 \\ 0 & 0 & 0 & kXA & kYA & kZA & 0 \end{pmatrix}; \quad B = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix};$$

v0 = {X0, Y0, Z0, XX0, YY0, ZZ0, A0};

(* Computation of the exhaustive summary *)

H = s * IdentityMatrix[7] − A; HI = Inverse[H];

Y = Together[B.HI.v0];

YCoeffList = Flatten[CoefficientList[Join[Numerator[Y], Denominator[Y]], s]];

MatrixForm[YCoeffList]; (* output not shown due to length *)

Length[YCoeffList]

49

YCoeffListLinear = Normal[CoefficientArrays[YCoeffList, {kXY, kYZ, kZX, kX0, kY0, kZ0, kXA, kYA, kZA}]][[2]]

(* coefficient list of equations that are linear in the parameters *)

{{0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {XX0, X0 + XX0, X0 + XX0 + 2Z0, X0 + XX0, X0 + XX0, X0 + XX0, X0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {X0 + Y0 + YY0, YY0, Y0 + YY0, Y0 + YY0, Y0 + YY0, Y0 + YY0, 0, Y0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {Z0 + ZZ0, Y0 + Z0 + ZZ0, ZZ0, Z0 + ZZ0, Z0 + ZZ0, Z0 + ZZ0, 0, 0, Z0}, {0, 0, 0, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0}, {X0 + Y0 + Z0, X0 + Y0 + Z0, X0 + Y0 + 2Z0, Y0 + Z0, X0 + Z0, X0 + Y0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0},
{A0, A0, A0, A0, A0, A0, A0 + XX0, A0 + YY0, A0 + ZZ0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0}, {1, 1, 1, 1, 1, 1, 1, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0},
{1, 1, 1, 1, 1, 1, 0, 1, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {1, 1, 1, 1, 1, 1, 0, 0, 1},
{0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {1, 1, 1, 1, 1, 1, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0},
{0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 0, 0, 0, 0, 0, 0, 0, 0}, {1, 1, 1, 1, 1, 1, 1, 1, 1},
{0, 0, 0, 0, 0, 0, 0, 0, 0}}

MatrixForm[Q = YCoeffListLinear[[{3, 7, 11, 14, 21, 26, 31, 36, 40}]]] (* pick 9 equations that are linear in the parameters *)

$$\begin{pmatrix}
XX0 & X0+XX0 & X0+XX0+2Z0 & X0+XX0 & X0+XX0 & X0+XX0 & X0 & 0 & 0 \\
X0+Y0+YY0 & YY0 & Y0+YY0 & Y0+YY0 & Y0+YY0 & Y0+YY0 & 0 & Y0 & 0 \\
Z0+ZZ0 & Y0+Z0+ZZ0 & ZZ0 & Z0+ZZ0 & Z0+ZZ0 & Z0+ZZ0 & 0 & 0 & Z0 \\
X0+Y0+Z0 & X0+Y0+Z0 & X0+Y0+2Z0 & Y0+Z0 & X0+Z0 & X0+Y0 & 0 & 0 & 0 \\
A0 & A0 & A0 & A0 & A0 & A0 & A0+XX0 & A0+YY0 & A0+ZZ0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0
\end{pmatrix}$$

MatrixForm[RowReduce[Q]]

$$\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & \frac{-Y0+Z0}{X0-Y0} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & \frac{X0-Z0}{X0-Y0} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

(* The parameters kXY, kYZ, kZX, kXA, kYA, kZA are uniquely identifiable *)

YCoeffListLinearReduced = Normal[CoefficientArrays[YCoeffList, {kX0, kY0, kZ0}]][[2]]; (* output not shown due to length *)

MatrixForm[R = YCoeffListLinearReduced[[{38, 39, 40}]]] (* pick 3 equations that are linear in the parameters *)

$$\begin{pmatrix}
kYZkZX & kXYkZX & kXYkYZ \\
kYZ+kZX & kXY+kZX & kXY+kYZ \\
1 & 1 & 1
\end{pmatrix}$$

Simplify[Det[R]]

$-(kXY-kYZ)(kXY-kZX)(kYZ-kZX)$

(* All rate parameters are uniquely identifiable *)

APPENDIX B

MATLAB code for parameter estimation

APPENDIX B: MATLAB CODE FOR PARAMETER ESTIMATION

```
function ParameterEstimationAlgorithm
% ------------------------------------------------------------------------
%  Parameter estimation for the seven-compartment ODE model M2**
%  with time-dependent (sigmoidal) transition rates to cell death
% ------------------------------------------------------------------------

global v0;

% number of parameters
ParNo = 9;

% number of runs for SRES
run_number = 30;

% maximum generation number
G = 500;

% lower & upper bounds for parameter estimates
% p = [ kXY kYZ kZX dkX dkY kZ0 dkZ t0 kZA ] for treatment with 50 nM RHPS4
lu = zeros(2,ParNo);
l0 = [  0.5 , 1,  1,  0,  0, 0,  0, 0 , 0 ];
u0 = [ 10   ,50, 50, 95, 95, 5, 95, 9 , 5 ];
lu(:,:) = [l0;u0];

% load experimental data from file (here: exp_data.mat)
% [t = time points (n-dim. vector), w_ij = measurements (n*m-dim. matrix),
%  m = number of state variables, n = numer of time points]
% w_ij(i,*) = (X+X_ , Y+Y_ , Z+Z_ , A , X+Y+Z , X_+Y_+Z_+A),  i=1,...,n
load exp_data t w_ij m n

% estimates of initial conditions v0
% v(1) = G0/G1(t) | v(2) = S(t) | v(3) = G2/M(t) |
% v(4) = G0/G1_dead(t) | v(5) = S_dead(t) |  v(6) = G2/M_dead(t) |
% v(7) = A(t)
p_viable = w_ij(1,5)/(w_ij(1,5)+w_ij(1,6)-w_ij(1,4));
v0 = [p_viable*w_ij(1,1) p_viable*w_ij(1,2) p_viable*w_ij(1,3) (1-p_viable)*w_ij(1,1)↙
(1-p_viable)*w_ij(1,2) (1-p_viable)*w_ij(1,3) w_ij(1,4)];

% initialise matrices / vectors to store optimisation results
f_min_global_vectors = [];
f_min_local_vectors = [];
p_min_global_matrices = [];
p_min_local_matrices = [];
f_min_best = inf;
p_min_best = zeros(1,ParNo);

% allocate cost function
cost = @(p) leastsquares(p);

for run = 1:run_number

    % compute global minimum using the global optimisation routine SRES
    [p_min_global Stats] = sres_nus( cost , 'min' , lu , 350 , G , 50 , 0.45 , 1 ,↙
{'log' 'log'} , '' );
    f_min_global = Stats(G,1);

    % Matlab code for the SRES method used here is available on the Web at
    % the following URL: <http://www.cs.princeton.edu/~stevenk/optimization>,
    % the first lines of which are shown below (parameter declaration).
```

```
    %
    %   function [xb,Statistics,performance_matrix,Gm] = sres_nus(fcn,mm,lu,
    %       lambda,G,mu,pf,varphi,nu_methods,logfile,varargin)
    %   SRES Evolution Strategy using Stochastic Ranking
    %   %
    %   %  fcn           : name of function to be optimized (string)
    %   %  mm            : 'max' or 'min' (for maximization or minimization)
    %   %  lu            : parameteric constraints (lower and upper bounds)
    %   %  lambda        : population size (number of offspring) (100 to 200)
    %   %  G             : maximum number of generations
    %   %  mu            : parent number (mu/lambda usually 1/7)
    %   %  pf            : pressure on fitness in [0 0.5] try around 0.45
    %   %  varphi        : expected rate of convergence (usually 1)
    %   %  nu_methods    : non-uniform methods, {} for uniform sampling,
    %   %                  otherwise {method for initial population sampling,
    %   %                  method used during iterations}
    %   %                  possible choices are 'log' and '' for uniform,
    %   %                  the default (e.g., {'log' 'log'})
    %   %  logfile       : write any intermediate output to the file
    %   %                  specified by the string logfile
    %   %                  No output is written if logfile == ''
    %   %  varargin      : problem-dependent parameters passed directly to
    %   %                  the function fcn
    %   %                  (e.g., feval(fcn,x,varargin{:}))
    %   %
    %   %  xb            : best feasible individual found
    %   %  Stats         : [min(f(x)) mean(f(x)) number_feasible(x)]
    %   %  Gm            : the generation number when "xb" was found

    %   %  Copyright (C) 1998-1999 Thomas Philip Runarsson
    %   %  (e-mail: tpr@verk.hi.is)
    %
    %   ...

    % compute local minimum using the local optimisation routine LM with
    % the global minimum as initial guess
    [p_min_local f_min_local residual exitflag output lambda jacobian] =↙
LM_ODE7_sigmoidal_rate ( p_min_global , f_min_global );

    % store optimisation results
    f_min_global_vector(run) = f_min_global;
    f_min_local_vector(run) = f_min_local;
    p_min_global_matrix(:,run) = p_min_global;
    p_min_local_matrix(:,run) = p_min_local;

    % update results with overall lowest cost function value and
    % corresponding parameter estimates
    if f_min_local < f_min_best
       f_min_best = f_min_local;
       p_min_best = p_min_local;
    end;

end

% display results
disp(' ')
disp('best value of cost function: ')
disp(f_min_best)
disp(' ')
```

```
disp('estimated optimal parameters: ')
disp(p_min_best(1:ParNo))
disp(' ')
disp('**********************************************************************')

end

% definition of cost function - sum of squared residuals
function f=leastsquares(p)

    global v0;

    % load experimental data from file (here: exp_data.mat)
    % w_ij(i,*) = (X+X_ , Y+Y_ , Z+Z_ , A , X+Y+Z , X_+Y_+Z_+A),  i=1,...,n
    load data_exp t w_ij m n

    % set parameter values to zero (model M2**)
    % pp(1)=kXY, pp(2)=kYZ, pp(3)=kZX,
    % pp(4)=kX0, pp(5)=dkX, pp(6)=kY0, pp(7)=dkY, pp(8)=kZ0, pp(9)=dkZ,
    % pp(10)=t0, pp(11)=kXA, pp(12)=kYA, pp(13)=kZA
    pp = [p(1:3) 0 p(4) 0 p(5:(end-1)) 0 0 p(end)];

    % numerical integration of ODEs using ode45 solver
    tspan = 0:1:t(end);
    [tt,v] = ode45( @modelODE7_sigmoidal_rate , tspan , v0 , [] , pp );

    % vector of observables
    w = [ sum(v(:,[1 4]),2) sum(v(:,[2 5]),2) sum(v(:,[3 6]),2) v(:,7) sum(v(:,[1 2 3]),2) sum(v(:,[4 5 6 7]),2) ];

    % sum of squared residuals
    f = 0;
    wij_log = log(w_ij);
    w_log = log(w);
    for i = 1:n
        for j = 1:m
            f = f + ( wij_log(i,j) - w_log(t(i)+1,j) )^2;
        end
    end

end

% definition of system of ODEs
function y=modelODE7_sigmoidal_rate(t,v,p)

    y = zeros(7,1);
    sigmoid_t0 = 1/(1+(t/p(10))^10);

    y(1) = 2*p(3)*v(3) - ( p(1)+p(4)+p(5)-p(5)*sigmoid_t0 )*v(1);
    y(2) = p(1)*v(1) - ( p(2)+p(6)+p(7)-p(7)*sigmoid_t0 )*v(2);
    y(3) = p(2)*v(2) - ( p(3)+p(8)+p(9)-p(9)*sigmoid_t0 )*v(3);
    y(4) = ( p(4)+p(5)-p(5)*sigmoid_t0 )*v(1) - p(11)*v(4);
    y(5) = ( p(6)+p(7)-p(7)*sigmoid_t0 )*v(2) - p(12)*v(5);
    y(6) = ( p(8)+p(9)-p(9)*sigmoid_t0 )*v(3) - p(13)*v(6);
    y(7) = p(11)*v(4) + p(12)*v(5) + p(13)*v(6);

end

% definition of local optimisation routine
```

```
function [p_min_local f_min_local residual exitflag output lambda jacobian] = LM_ODE7_sigmoidal_rate(p_min_global,f_min_global)

    global v0;

    % local optimisation method = Levenberg-Marquardt (LM)
    options = optimset('LargeScale ','off','LevenbergMarquardt ','on','MaxFunEvals', 10000,'MaxIter',100);
    [p_min_local f_min_local residual exitflag output lambda jacobian] = lsqnonlin( @model_residuals , p_min_global , [] , [] , options );

    % define vector of model residuals
    function e = model_residuals(p)

        % load experimental data from file (here: exp_data.mat)
        % w_ij(i,*) = (X+X_ , Y+Y_ , Z+Z_ , A , X+Y+Z , X_+Y_+Z_+A),  i=1,...,n
        load exp_data t w_ij m n

        % set parameter values to zero (model M2**)
        % pp(1)=kXY, pp(2)=kYZ, pp(3)=kZX,
        % pp(4)=kX0, pp(5)=dkX, pp(6)=kY0, pp(7)=dkY, pp(8)=kZ0, pp(9)=dkZ,
        % pp(10)=t0, pp(11)=kXA, pp(12)=kYA, pp(13)=kZA
        pp = [p(1:3) 0 p(4) 0 p(5:(end-1)) 0 0 p(end)];

        % numerical integration of ODEs using ode45 solver
        tspan = 0:1:t(end);
        [tt,v] = ode45( @modelODE7_sigmoidal_rate , tspan , v0 , [] , pp );

        % vector of observables
        w = [ sum(v(:,[1 4]),2) sum(v(:,[2 5]),2) sum(v(:,[3 6]),2) v(:,7) sum(v(:,[1 2 3]),2) sum(v(:,[4 5 6 7]),2)];

        % vector of model residuals
        e = zeros(1,m*n);
        wij_log = log(w_ij);
        w_log = log(w);
        for j = 1:m
            for i = 1:n
                e(1,(j-1)*m+i) = wij_log(i,j) - w_log(t(i)+1,j);
            end
        end

        % implement parameter constraints by adding penalties to cost
        % function
        for k = 1:size(p,2)
            if (p(k) < 0 || p(k) > 1000)
                e(1,:) = sqrt( f_min_global / (m*n-0.5) );
                % i.e. f_min_local > f_min_global
            end
        end

    end

end
```

APPENDIX C

Computational notes and data tables for practical identifiability analysis

In Chapter 3.4, we introduced the Fisher information matrix $\mathbf{J_F}\left(\widehat{\mathbf{p}}\right)^T \mathbf{J_F}\left(\widehat{\mathbf{p}}\right)$, which we employed for deriving confidence regions and correlation coefficients for the estimated parameters $\widehat{\mathbf{p}}$. Here, we describe a method for computing the entries of the Jacobian $\mathbf{J_F}\left(\widehat{\mathbf{p}}\right)$ (Appendix C.1), present data tables of the correlation coefficients $r_{ij}\left(\widehat{\mathbf{p}}\right)$ (Appendix C.2), and the eigenvalues μ_k and eigenvectors $\mathbf{u}_k$ of the Fisher information matrix to obtain respectively the radii and directions of the principal axes of the confidence regions of $\widehat{\mathbf{p}}$ (Appendix C.3). The data tables are shown for model M_2^{**} from Chapter 5.1 fitted to data from treatment with 50 nM of RHPS4.

C.1 Numerical computation of $\mathbf{J_F}\left(\widehat{\mathbf{p}}\right)$

For the computational side of the practical identifiability and sensitivity analysis, it is important to find a reliable method for the numerical calculation of $\mathbf{J_F}\left(\widehat{\mathbf{p}}\right)$, which involves derivatives of the state variables $\mathbf{v}_j$ with respect to the parameters $\mathbf{p}_k$. Using divided differences

$$\frac{\partial \mathbf{v}_j(t_i, \mathbf{p})}{\partial \mathbf{p}_k} = \frac{\mathbf{v}_j(t_i, \mathbf{p} + \epsilon\, \mathbf{s}_k) - \mathbf{v}_j(t_i, \mathbf{p})}{\epsilon} + \mathrm{O}\left(\epsilon\right) , \tag{C.1.1}$$

with $\mathbf{s}_k$ being the kth unit vector in $\mathbb{R}^L$, may lead to significant errors in the computation of the derivatives in $\mathbf{J_F}\left(\widehat{\mathbf{p}}\right)$ dependent on the choice of ϵ. Also, employing the Richardson

extrapolation (see Johnson and Riess [66]) using, for example,

$$\frac{\partial \mathbf{v}_j\left(t_i,\mathbf{p}\right)}{\partial \mathbf{p}_k} = \frac{1}{6\,\epsilon}\Big(\mathbf{v}_j\left(t_i,\mathbf{p}-\epsilon\,\mathbf{s}_k\right) - 8\,\mathbf{v}_j\left(t_i,\mathbf{p}-\frac{\epsilon}{2}\,\mathbf{s}_k\right) + 8\,\mathbf{v}_j\left(t_i,\mathbf{p}+\frac{\epsilon}{2}\,\mathbf{s}_k\right) - \mathbf{v}_j\left(t_i,\mathbf{p}+\epsilon\,\mathbf{s}_k\right)\Big) + \mathrm{O}\left(\epsilon^4\right), \tag{C.1.2}$$

to obtain approximations of higher order does not solve the problem of the right choice of ϵ. A better approach is to use the variational equations [8]

$$\frac{\partial}{\partial t}\frac{\partial \mathbf{v}\left(t,\mathbf{p}\right)}{\partial \mathbf{p}_k} = \frac{\partial \mathbf{A}\left(t,\boldsymbol{\theta}\right)}{\partial \mathbf{p}_k}\,\mathbf{v} + \mathbf{A}\left(t,\boldsymbol{\theta}\right)\frac{\partial \mathbf{v}\left(t,\mathbf{p}\right)}{\partial \mathbf{p}_k}, \qquad 0 < t < t_e, \tag{C.1.3}$$

$$\frac{\partial \mathbf{v}\left(0,\mathbf{p}\right)}{\partial \mathbf{p}_k} = \left\{\begin{array}{ll} \mathbf{s}_j & \text{if } \mathbf{p}_k = (\mathbf{v}_0)_j \text{ for any } j \\ \mathbf{0} & \text{else} \end{array}\right\}, \tag{C.1.4}$$

which are respectively the derivatives of equations (5.3.1) and (5.3.2) in Chapter 5.3 with respect to $\mathbf{p}_k$. We obtain $\partial\mathbf{v}(t,\mathbf{p})/\partial\mathbf{p}_k$ for every value $t \in [0,t_e]$ by numerical integration of (C.1.3) with respect to t. We then compute the $((j-1)n+i,k)^{\text{th}}$ entry, $i = 1,\ldots,n$, $j = 1,\ldots,M$, of the Jacobian $\mathbf{J_F}\left(\widehat{\mathbf{p}}\right)$ by

$$\frac{\partial \mathbf{F}_{(j-1)n+i}\left(\mathbf{p}\right)}{\partial \mathbf{p}_k} = \frac{\left[\mathbf{B}\left(\frac{\partial \mathbf{v}_1(t_i,\mathbf{p})}{\partial \mathbf{p}_k},\ldots,\frac{\partial \mathbf{v}_m(t_i,\mathbf{p})}{\partial \mathbf{p}_k}\right)^T\right]_j}{\left[\mathbf{B}\,\mathbf{v}\left(t_i,\mathbf{p}\right)\right]_j}, \tag{C.1.5}$$

with $\mathbf{B}$ as in Chapter 5.3.

C.2 Correlation coefficients for rate parameters of model $\mathbf{M}_2^{**}$

$\widehat{\mathbf{p}}_k$	k_{XY}	k_{YZ}	k_{ZX}	$k_{Z\bar{Z}0}$	$\Delta k_{X\bar{X}}$	$\Delta k_{Y\bar{Y}}$	$\Delta k_{Z\bar{Z}}$	t_0	k_{ZA}
k_{XY}	1	-0.3705	-0.7089	-0.5611	0.6719	-0.3118	-0.7014	-0.0160	0.1211
k_{YZ}		1	-0.1513	-0.1607	-0.3471	0.7008	-0.0754	-0.2058	-0.0167
k_{ZX}			1	0.9617	-0.4064	-0.0605	0.6632	0.1068	-0.4296
$k_{Z\bar{Z}0}$				1	-0.2750	-0.0385	0.5167	0.2015	-0.5210
$\Delta k_{X\bar{X}}$					1	-0.3917	-0.7752	0.3833	0.0714
$\Delta k_{Y\bar{Y}}$						1	-0.0711	0.0366	-0.0365
$\Delta k_{Z\bar{Z}}$							1	-0.0261	-0.1145
t_0								1	-0.0351
k_{ZA}									1

Table C.1: Upper triangular part of the correlation matrix $\mathbf{Corr}(\widehat{\mathbf{p}})$ with entries $r_{ij}\left(\widehat{\mathbf{p}}\right)$ computed as in (3.4.7) for model M_2^{**} (see Chapter 5.1) and the data fit from treatment with 50 nM RHPS4.

C.3 Eigenvalues and eigenvectors for the confidence region of $\widehat{\mathbf{p}}$ in model M_2^{**}

k	1	2	3	4	5	6	7	8	9
μ_k	0.0666	0.4881	0.8052	6.2742	26.753	39.139	352.33	2285.6	3388.0
l_k	39.148	14.466	11.263	4.0348	1.9540	1.6155	0.5384	0.2114	0.1736

Table C.2: Eigenvalues μ_k and radii l_k of the principal axes for the ellipsoidal 95% confidence region of the $L = 9$-dimensional parameter vector $\widehat{\mathbf{p}}$ for model M_2^{**} and 50 nM RHPS4 (see Chapter 5.1).

	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$	$\mathbf{u}_4$	$\mathbf{u}_5$	$\mathbf{u}_6$	$\mathbf{u}_7$	$\mathbf{u}_8$	$\mathbf{u}_9$
k_{XY}	-0.0268	-0.0252	0.0569	0.0185	0.0258	-0.2978	-0.5703	-0.7421	0.1725
k_{YZ}	-0.0403	0.1203	-0.8698	-0.4670	0.0756	-0.0209	-0.0263	-0.0468	0.0130
k_{ZX}	0.8976	-0.3670	-0.0840	-0.0304	-0.1041	0.1818	-0.0204	-0.0858	0.0096
$k_{Z\bar{Z}0}$	0.1778	-0.1597	-0.0264	0.0570	0.3741	-0.8396	0.0593	0.3008	-0.0081
$\Delta k_{X\bar{X}}$	-0.0534	-0.1911	0.1360	-0.1956	0.7236	0.3775	-0.4358	0.2182	-0.0185
$\Delta k_{Y\bar{Y}}$	-0.0112	0.0267	-0.4456	0.8572	0.1388	0.1359	-0.1573	0.0568	-0.0043
$\Delta k_{Z\bar{Z}}$	0.3968	0.8873	0.1230	-0.0053	0.1751	-0.0029	-0.0887	0.0397	-0.0024
t_0	0.0026	-0.0148	0.0141	0.0651	0.5189	0.0986	0.6659	-0.5210	0.0390
k_{ZA}	-0.0022	0.0044	0.0025	0.0012	-0.0073	0.0475	0.0655	0.1592	0.9839

Table C.3: Eigenvectors $\mathbf{u}_k$ describing the directions of the principal axes of the ellipsoidal confidence region of $\widehat{\mathbf{p}}$ for model M_2^{**} and 50 nM RHPS4.

APPENDIX D

Coefficients of solution (7.3.11)

The coefficients a_1, b_0, b_1 of (7.3.11) are given by the expressions

$$\begin{aligned} b_0 &= E(\widehat{h}) + \frac{\widehat{h}}{h}\,(1 - E(h)) \\ &\quad \times \exp\left[\frac{k_{\text{on}}\,k_{\text{off}}\,T_0}{\rho^2\,h^2\,\widehat{h}^2}\left(k_{\text{on}}\,k_{\text{off}}\,\sigma^2\,T_0\,\left(\delta\,h + h^2\right) + \delta\,h\,\widehat{h}\,\left(\alpha\,\rho - L_0\,\rho - k_{\text{off}}\,\sigma^2\right)\right)\right], \end{aligned} \tag{D.0.1}$$

$a_1 = F(h)$ and $b_1 = F(\widehat{h})$ with

$$h = k_{\text{on}}\,T_0 + \frac{k_f\,k_r\,k_s\,R_0}{k_r\,k_s\,R_0 + k_u\,(k_d + k_r)}\,, \qquad \widehat{h} = h + \delta\,, \tag{D.0.2}$$

where

$$E\,(Y) = \text{erf}\left[\frac{L_0 - \alpha}{\sigma\,\sqrt{2}} + \frac{\sigma\,k_{\text{off}}\,(Y - k_{\text{on}}\,T_0)}{\rho\,Y\,\sqrt{2}}\right], \tag{D.0.3}$$

$$F\,(Y) = \frac{k_e\,k_{\text{on}}\,T_0}{2\,\rho\,Y}\,\exp\left[\frac{k_{\text{off}}\,(Y - k_{\text{on}}\,T_0)\,\left(2\,\rho\,L_0\,Y + k_{\text{off}}\,\sigma^2\,Y - k_{\text{on}}\,k_{\text{off}}\,\sigma^2\,T_0\right)}{2\,\rho^2\,Y^2}\right]. \tag{D.0.4}$$