Chapter 4

Methodology of Mathematical Modelling

4.1 Introduction

Mathematical modelling involves the use of mathematical language to construct abstract models which describe the behaviour of a system. They are particularly useful in the natural sciences and engineering disciplines because they allow us to apply simple mathematical skills in order to gain insight into real problems (Edwards & Hamson, 1989). A model is a system of mathematical equations, which may be algebraic, differential, partial differential or integral or a combination thereof. There are three main reasons for developing mathematical models of biological systems, these being exploratory (the model must generate hypotheses), explanatory (it must give a detailed description of the underlying structure) and/or be predictive (it must be useful for making decisions). To be of value, a mathematical model must involve:

- The best possible understanding of biological phenomenon.
- The possibility of functioning (system optimization).

There are several types of models, though the two most common are:

- Knowledge models, where the mathematical equations are translations of the physical laws of the system.
- Simulation models giving a priori equations whose coefficients must be identified from available experimental data.

In practice, it is common to build intermediate models using both physical laws and experimental data. Both knowledge models and simulation models are equally useful, though it must be remembered that a model is an approximation of biological phenomenon, and each model has its own strengths and weaknesses as they are based on different approximations or simplifications (Cherruault, 1985).

4.1.1 Introduction to Differential Equations

Change is one of the most interesting aspects of most systems, and is the central importance across disciplines of differential equations. An ODE relates the derivative of a quantity of interest (i.e. the dependent variable) with respect to another variable (i.e. the independent variable) to a function of both the dependent and independent variables and describes how these variables change with each other. It is simply an equation, or a system of equations, written in terms of an unknown function and its derivatives with respect to a single independent variable, for example, time. A simple ODE is shown in equation (4.1.1). It can be solved by integration as shown in equation (4.1.2):

$$\frac{dy}{dt} = f(t) \rightarrow dy = f(t)dt \quad \textbf{(4.1.1)}$$

$$\rightarrow\ y(t_1) - y(t_0) = \int_{t_0}^{t_1} f(t\,)dt \quad \textbf{(4.1.2)}$$

In equation (4.1.1), the independent variable is time, t and the dependent variable is y. The term used to relate the two variables is $f(t)$, and $y\ (t_0)$ is the prescribed value of y at time, $t_0 < t_1$.

The order of a differential equation is the highest derivative that occurs and so the above example is a first-order differential equation. If the ODE contains terms which are linear in the dependent variable and its derivatives then it is said to be a linear equation and can be written in terms of a linear operator, $L_N(y)$ as shown in equation (4.1.3):

$$L_N \equiv \frac{d^N y}{dt^N} + A_1(t)\frac{d^{N-1}y}{dt^{N-1}} + \ldots A_{N-1}(t)\frac{dy}{dt} + A_N(t)y \quad \textbf{(4.1.3)}$$

Linear differential equations are important because they arise frequently when modelling physical systems, and also because it is relatively straightforward to solve them using standard techniques (Gershenfeld, 1999). The models that we have developed in this chapter represent the cell cycle of HCT116 cells and their response to RHPS4 and are formulated as a system of linear ODEs.

The data for the mathematical model that we present comes from a series of experiments involving incubation of the cells in media (control) and a range of concentrations of RHPS4 for up to 21 days. The cells were analysed via the SA-β-gal assay for senescence and the Nicoletti method for cell cycle analysis. The number of population doublings was also counted using a heamocytometer. Analysis of the cells occurred on days 4, 6, 10, 14 and 21.

According to MTT assays carried out up to 72 hours, the GI_{50} of RHPS4 on HCT116 cells was calculated as approximately 1 μM. Based on this analysis, the cells were incubated in media (control), 50 nM RHPS4 (0.05x GI_{50}), 100 nM RHPS4 (0.1x GI_{50}) and 1 μM RHPS4 (1x GI_{50}).

4.1.2 What Data can be derived to Apply to the Mathematical Model?

One desirable characteristic of our mathematical model is that it must be easily accessible to biologists with no prior knowledge or expectation of prior knowledge of mathematics. To fulfill this condition the data that is used incorporates as much information as is needed to describe the effects of RHPS4 and other anti-tumour agents. In the laboratory it is easy to derive information about the cell cycle, population doublings and senescence using the methods described in chapter 2. This data gives information regarding the proportion of cells in each phase of the cell cycle, the overall growth rate of the population of cells and the proportion of cells which have become senescent. The data is easily derived for cells under normal conditions and in response to anti-tumour agents and simple comparison allows information on the effects of each of these aspects of the cell cycle to be determined. The information that is not easily derived from these simple pieces of information is the kinetics or dynamic behaviour. The cell cycle data simply tells us what phase of the cell cycle the cells are in but not the rate at which they move between these compartments. The mathematical model, based on the derived data should enable us to determine these rates. We have five parameters that we are able to derive in the laboratory, those being the overall growth rate of the cells, the proportions of cells in senescence, G_1/G_0, S, and G_2/M. From this we can also calculate the proportions of dead cells and therefore we wanted to derive a mathematical model that includes no more than this data alone. Of course we can

incorporate the data into a model in more ways than those shown in sections 4.5 – 4.6, in particular senescence and cell death, however, to ensure the ODEs involved with the model are kept simple, the models shown in sections 4.5 – 4.6 are the most accurate way to do this.

4.2 Introduction to the Mathematical Models

As previously explained there is a variety of data describing the effects of RHPS4 on HCT116 cells. We have data showing:

- How the proportion of cells which are senescent changes in response to increasing concentrations of RHPS4 and doxorubicin.
- How the number of successful doublings that the cells undergo changes in response to increasing concentrations of RHPS4 and doxorubicin.
- The effects of RHPS4 and doxorubicin on the cell cycle. The data is available over a period of 21 days allowing the effects of both concentration and time to be derived.

Though conclusions and other insights regarding the effects of RHPS4 can be derived from the data (for example, over time the data shows there is an increase in the proportion of cells in G_1/G_0), what cannot easily be derived are conclusions regarding the drug's effects on the kinetics of the cell cycle. For example, while the flow cytometry data allows the proportions of cells in pre-G_1, G_1/G_0, S and G_2/M to be derived, it fails to reveal whether an increase in the proportion of cells in G_1/G_0, is due to an increase in the proportion of cells in G_1 or G_0 or whether there is a change in the movement of cells between the individual compartments or a reduction in the flow of cells into S phase. This information is difficult to appreciate without being able to measure the kinetics.

A basic cell cycle, which we use to derive our mathematical models, is depicted below in figure 4.1. This figure shows the progression of cells through the cell cycle from G_1 » S » G_2 » M at rates k_{g1s}, k_{sg2}, k_{g2m} and k_{mg1}, respectively. The basic model assumes that cells flow between G_0 and G_1 at rates k_{g0g1} and k_{g1g0} and that they enter Σ (become senescent) at rate $k_{g0\Sigma}$. We assume that cells which

enter compartment Σ are unable to return to the cell cycle. This conceptual description/representation of the cell cycle forms the basis of a series of mathematical models of increasing complexity which we use to analyse our experimental data and enables us to determine where in the cell cycle RHPS4 is having an effect. Our aim is to derive a mathematical model which utilises all the data available, and incorporates as many of the rate parameters shown in figure 4.1 as possible.

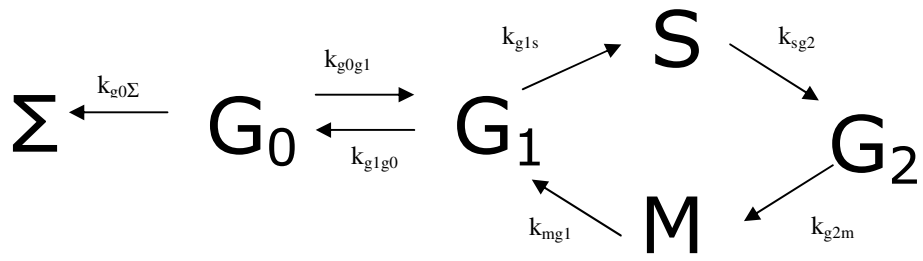


Figure 4.1 Schematic diagram of the cell cycle upon which the mathematical models are derived. The mathematical models derived in this chapter are based on the systematic diagram shown here and the rate of transition of the cells between the compartments are shown as $k_{g0\Sigma}$, kg_{0g1}, k_{g1g0}, k_{g1s}, k_{sg2}, k_{g2m} and k_{mg1}. The aim was to incorporate as much information from this diagram as possible. In the laboratory, we are able to derive information regarding the proportion of cells in pre-G_1, G_1/G_0, S, G_2/M, the number of population doublings and the proportion of cells in senescence. We are unable to derive the rate at which cells move between the individual compartments and the aim is derive mathematical models to allow the rate of transition between the compartments to be calculated and from this analyse the effects of anti-tumour agents on these rates.

Using figure 4.1, we can construct a system of time-dependent ODEs which describe how the number of cells at each stage of the cell cycle changes over time, t. We denote by $G_1(t)$, $G_2(t)$, S(t), M(t), $G_0(t)$ and Σ(t), the number of cells at time, t in G_1 phase, G_2 phase, S phase, M phase, G_0 phase and Σ phase, respectively.

The system of ODEs that correspond to figure 4.1 are as follows:

$$\frac{dG_0}{dt} = k_{g1g0}G_1 - (k_{g0g1} + k_{g0\Sigma}G_0 \quad \textbf{(4.2.1)}$$

$$\frac{d\Sigma}{dt} = k_{g0\Sigma}G_0 \quad \textbf{(4.2.2)}$$

$$\frac{dG_1}{dt} = 2k_{mg1}M + k_{g0g1}G_0 - (k_{g1g0} + k_{g1s})G_1 \quad \textbf{(4.2.3)}$$

$$\frac{dS}{dt} = k_{g1s}G_1 - k_{sg2}S \quad \textbf{(4.2.4)}$$

$$\frac{dG_2}{dt} = k_{sg2}S - k_{g2m}G_2 \quad \textbf{(4.2.5)}$$

$$\frac{dM}{dt} = k_{g2m}G_2 - k_{mg1}M \quad \textbf{(4.2.6)}$$

In equation (4.2.1), we assume that the processes which control the number of cells in G_0 are the rate at which cells enter G_0 from G_1 (i.e. the cells leaving the cell cycle) and the rates at which they leave G_0 to enter G_1 (i.e. resume cell cycling) or Σ (i.e. become senescent, in which case they will not re-enter the cell cycle). Combining these effects, we deduce that equation (4.2.1) describes how the number of cells in G_0 changes over time.

Another important feature of the model depicted in figure 4.1 is revealed by equations (4.2.3) and (4.2.6) which state that twice as many cells enter G_1 as leave M-phase. This is because when a single parent cells completes M-phase (at rate k_{mg1}), it will produce two daughter cells which we assume enter G_1 phase. We note further that, with the exception of the term $k_{mg1}M$, all terms in equations (4.2.1) – (4.2.6) appear twice. They appear once with a positive coefficient and once with a negative coefficient because as cells leave one compartment of the cell cycle (hence a negative coefficient) they enter a new compartment (hence a positive coefficient). The exception to this rule is the terms incorporating mitosis in equations (4.2.3) and (4.2.6) as described previously.

Before analysing equations (4.2.1) - (4.2.6), we pause to consider several simpler cell cycle models and introduce the techniques we use to study them. It is also useful to note that there is no cell death.

4.3 One-Compartment Model

The simplest mathematical model treats all cells within a population as identical, regardless of their position in the cell cycle. This case is shown in figure 4.2. Compartment Y represents all cells within the population and the model describes how the number of cells changes with respect to time.

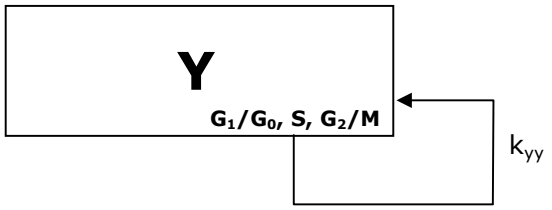


Figure 4.2 One-compartment mathematical model. The mathematical model presented here assumes that all cells within the population are in the same compartment, Y and growth of this compartment is via the rate parameter k_{yy}. The model utilises only the population doublings data as there is no consideration of the phase of the cell cycle in which the cells are residing.

The dynamics of the cell cycle are condensed into a single parameter,k_{yy}, which models the net rate at which the cells proliferate. In this case, the governing ODE can be written as follows:

$$\frac{dY}{dt} = k_{yy}Y \qquad \textbf{(4.3.1)}$$

Equation (4.3.1) predicts there is exponential growth of the population of cells. If at time, t = 0, there are $Y(t = 0) = Y_0$ cells, then we deduce that at time, t, $Y(t) = Y_o e^{k_{yy}t}$.

While this model is the simplest description of the cell cycle, it is of limited value because it utilises only the doublings data. We could fit data to this model, however, we would not be able to draw any new conclusions that we cannot already derive from the raw data. Additionally much of the available data would not be used.

4.4 Two-Compartment Model

The next level of complexity involves decomposing the cell population into two compartments, with $X(t)$ representing the number of cells in the G_0, G_1 and S phases of the cell cycle and $Y(t)$ representing the number in the G_2 and M phases at time, t. In this way we effectively decompose the cell cycle into two phases: a synthesising stage (X) and a replicating stage (Y) (see figure 4.3).

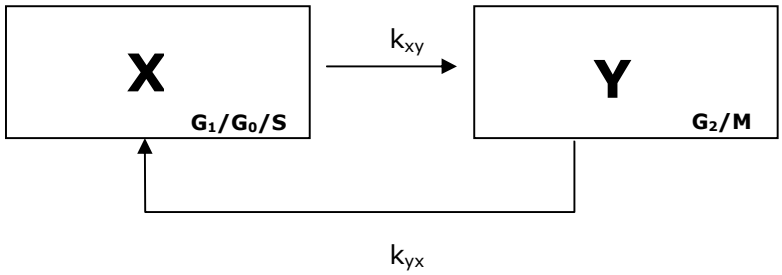


Figure 4.3 Two-compartment mathematical model. The mathematical model shown here depicts the cell cycle to be in two compartments, X (incorporating G_1/G_0 and S) and Y (incorporating G_2/M) and transition between the compartments is represented by k_{xy} and k_{yx}, respectively. The two-compartment model utilises the population doublings data in addition to the cell cycle data, however, does not incorporate the pre-G_1 or senescence data. The model could also be described as 'closed' as there is no allowance for exit of cells from the model.

In the case of the two-compartment model, the governing ODEs are given as:

$$\frac{dX}{dt} = 2k_{yx}Y - k_{xy}X \quad \textbf{(4.4.1)}$$

$$\frac{dY}{dt} = k_{xy}X - k_{yx}Y \quad \textbf{(4.4.2)}$$

with $X(t=0) = X_0$ and $Y(t=0) = Y_0$.

As before (see equations (4.2.3) - (4.2.6)), the factor of two premultiplying the coefficient k_{yx} in the ODE for X (equation (4.4.1)) is required because when a cell undergoes mitosis, it produces two daughter cells.

Since the governing ODEs are linear, we seek trial solutions in the form of exponentials $X(t) = \hat{X}e^{\alpha t}$ and $Y(t) = \hat{Y}e^{\alpha t}$. For a non-trivial solution (i.e. $\hat{X}, \hat{Y} \neq 0$), it is possible to show that the exponent α must satisfy the following quadratic equation:

$$0 = \alpha^2 + (k_{xy} + k_{yx})\alpha - k_{xy}k_{yx} \quad \textbf{(4.4.3)}$$

Since the expression for α is a quadratic, there are two possible solutions to equations (4.4.1) - (4.4.2) and hence the general solution is of the form:

$$X(t) = X_1e^{\alpha_1 t} + X_2e^{\alpha_2 t} \quad \textbf{(4.4.4)}$$

$$Y(t) = Y_1e^{\alpha_1 t} + Y_2e^{\alpha_2 t} \quad \textbf{(4.4.5)}$$

Where α_1 and α_2 are the roots of the quadratic (4.4.3) and X_1, X_2, Y_1 and Y_2 are constants of integration which are determined below.

Substituting with (4.4.4) – (4.4.5) into the initial conditions yields:

$X(t = 0) = X_1 + X_2 = X_0$ **(4.4.6)**

$Y(t = 0) = Y_1 + Y_2 = Y_0$ **(4.4.7)**

Thus the initial conditions supply two equations for our four unknowns. We may use equations (4.4.6) - (4.4.7) to eliminate two variables:

$X_1 = X_0 - X_2$ or $X_2 = X_0 - X_1$ **(4.4.8)**

$Y_1 = Y_0 - Y_2$ or $Y_2 = Y_0 - Y_1$ **(4.4.9)**

The original ODEs supply us with the following additional information:

$\frac{dX}{dt} = \alpha_1 X_1 e^{\alpha_1 t} + \alpha_2 X_2 e^{\alpha_2 t}$.

$= 2k_{yx}\,(Y_1 e^{\alpha_1 t} + Y_2 e^{\alpha_2 t}) - k_{xy}(X_1 e^{\alpha_1 t} + X_2 e^{\alpha_2 t})$, **(4.4.10)**

and so

$0 = \left(\alpha_1 X_1 - 2k_{yx}Y_1 + k_{xy}X_1\right)e^{\alpha_1 t} + \left(\alpha_2 X_2 - 2k_{yx}Y_2 + k_{xy}X_2\right)e^{\alpha_2 t}$. **(4.4.11)**

Equation (4.4.11) must hold true for all time, $t > 0$ and is only possible if the coefficients of $e^{\alpha_1 t}$ and $e^{\alpha_2 t}$ vanish simultaneously, that is if the following identities are satisfied (i.e. make equation (4.4.11) equal to zero):

$0 = \alpha_1 X_1 - 2k_{yx}Y_1 + k_{xy}X_1$, **(4.4.12)**

$0 = \alpha_2 X_2 - 2k_{yx}Y_2 + k_{xy}X_2$. **(4.4.13)**

Equations (4.4.12) - (4.4.13) constitute two additional equations which can be combined with equations (4.4.9) and (4.4.10) to determine the four unknown parameters X_1, X_2, Y_1, and Y_2. In more detail, rearranging (4.4.12) – (4.4.13) yields:

$Y_1 = \frac{(\alpha_1 + k_{xy})}{2k_{yx}} X_1$ and $Y_2 = \frac{(\alpha_2 + k_{xy})}{2k_{yx}} X_2$. **(4.4.14)**

From equation (4.4.8), we are able to simplify equation (4.4.14) to derive the following:

$$Y_2 = \frac{(\alpha_2 + k_{xy})}{2k_{yx}} X_2 = \frac{(\alpha_2 + k_{xy})}{2k_{yx}} (X_0 - X_1). \quad \textbf{(4.4.15)}$$

By combining (4.4.14) and (4.4.15), Y_1 can be eliminated and the following expression for X_1 can be derived:

$$2k_{yx}Y_0 = (\alpha_2 + k_{xy})X_0 + (\alpha_1 - \alpha_2)X_1. \quad \textbf{(4.4.16)}$$

Equation (4.4.16) can be rearranged to give:

$$(\alpha_1 - \alpha_2)X_1 = 2k_{yx}Y_0 - (\alpha_2 + k_{xy})X_0. \quad \textbf{(4.4.17)}$$

The general solution in equation (4.4.4) - (4.4.5) contains the four unknown constants X_1, X_2, Y_1 and Y_2.

The coefficients, α_1 and α_2, that appear in equation (4.4.17) are known in terms of the model parameters k_{xy} and k_{yx} (see equation 4.4.3).

We needed four equations to determine X_1, X_2, Y_1 and Y_2. The initial conditions in equation (4.4.8) - (4.4.9) provide two equations, whilst the ODEs shown in equation (4.4.15) provide another two. Combining them we are able to show the following:

$$X_1 = \frac{2k_{yx}}{(\alpha_1 - \alpha_2)} Y_0 - \frac{(\alpha_2 + k_{xy})}{(\alpha_1 - \alpha_2)} X_0 \quad \textbf{(4.4.18)}$$

$$Y_1 = \frac{(\alpha_1 + k_{xy})}{2k_{yx}} X_1 \quad \textbf{(4.4.19)}$$

$$Y_2 = Y_0 - Y_1 \quad \textbf{(4.4.20)}$$

$$X_2 = X_0 - X_1 \quad \textbf{(4.4.21)}$$

A key assumption in our analysis is that the quadratic for α possesses a dominant root which regulates the system's behaviour at long times. The results presented in figure 4.4 confirm that the behaviour of one component of the solution (that associated with the larger root) dominates. In order to illustrate this using MATLAB, we plotted the changes with respect to time in $X_1 e^{\alpha_1 t} + X_2 e^{\alpha_2 t}$. Both α_1 and α_2 have a value and as time, t, gets larger, it is α_1 that dominates the growth

of the system. Separate simulations were carried out to show the growth of both α_1 and α_2 over time separately and when added together. The results show that α_1 grows much more quickly than α_2 and reflects the growth of $X_1e^{\alpha_1 t} + X_2e^{\alpha_2 t}$ much more closely, thereby validating our assumption that the quadratic for α contains a dominant root which dominates the behaviour of the system. To carry out the simulation to illustrate the changes in $X_1e^{\alpha_1 t} + X_2e^{\alpha_2 t}$ with respect to time and determine that it is the α_1 that is dominant, the following conditions were imposed:

1. The fixed values of k_{xy} and k_{yx} were 5 and 2, respectively.
2. The values of X_0 and Y_0 were 0 and 2, respectively
3. The values of X_1, Y_1, X_2 and Y_2 were calculated in MATLAB 7 using equations (4.4.18) – (4.4.21).
4. The value of X(t) = $X_1e^{\alpha_1 t} + X_2e^{\alpha_2 t}$ and Y(t) = $Y_1e^{\alpha_1 t} + Y_2e^{\alpha_2 t}$.
5. The total numbers of cells was N(t) = X(t) + Y(t)

Suppose Real (α_1) > Real (α_2). Then, at long times, we have

$X(t) \sim X_1e^{\alpha_1 t}$. **(4.4.22)**

$Y(t) \sim Y_1e^{\alpha_1 t}$. **(4.4.23)**

i.e. the long-time behaviour is dominated by the solution with the larger real part. In what follows, we assume that the long-time behaviour is dominant, and exploit this to derive equations relating the rate parameters k_{xy} and k_{yx} to the experimental data. A simulation using MATLAB 7 was carried out to show the concept of equations (4.4.22) – (4.4.23) and the results are shown in figure 4.4.

The above analysis gives us information regarding the dynamics of the model. The quadratic for α ($0 = \alpha^2 + (k_{xy} + k_{yx})\alpha - k_{xy}k_{yx}$) must also be solved.

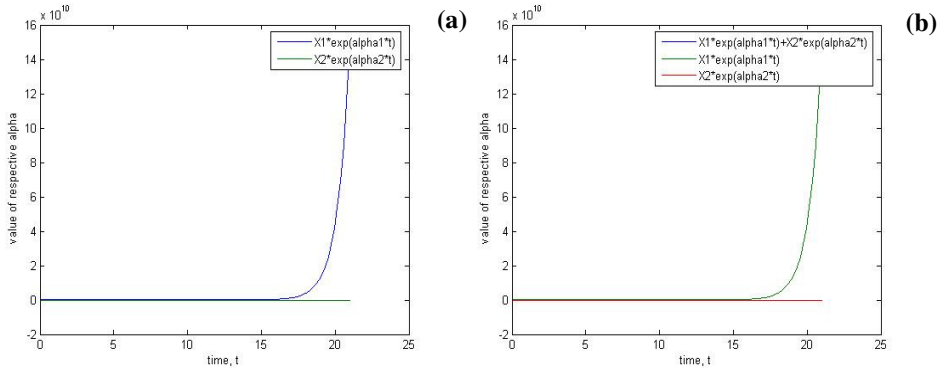


Figure 4.4 Determining that α_1 is dominant over α_2. The model solutions are sums of exponentials and so we have $X(t) = X_1 e^{\alpha_1 t} + X_2 e^{\alpha_2 t}$, where X_1 and Y_1 are the constant of integration determined by the initial conditions and α_1 and α_2 are constants. If $\alpha_1 > \alpha_2$ then at long times $X_1 e^{\alpha_1 t}$ dominates the behaviour of the growth of the system. In (a) we show how the components $X_1 e^{\alpha_1 t}$ and $X_2 e^{\alpha_2 t}$ evolve over time. In (b) we compare these components with X(t) = $X_1 e^{\alpha_1 t} + X_2 e^{\alpha_2 t}$ and find that at long times, $X(t) \sim X_1 e^{\alpha_1 t}$. The graph clearly shows that the growth of X_1 is more reflective of the growth of $X_1 + X_2$ proving that α_1 is dominant.

From the ODEs shown in equations (4.4.1) - (4.4.2), we are able to derive information regarding the kinetics of the cell cycle under normal conditions, and in response to RHPS4. The laboratory data gives us information regarding the cell cycle and the number of doublings. From the cell cycle data, we are able to determine the proportion of cells in G_1/G_0, S and G_2/M; however, no information regarding the flow of cells through these phases is derived. By using ODEs (4.4.1) - (4.4.2), we aim to derive equations for the unknown rate parameters k_{xy} and k_{yx} in terms of the experimental data and use them to gain information about how RHPS4 affects cell cycle progression.

As explained above the general solutions of equations (4.4.1) – (4.4.2) are stated in (4.4.4) - (4.4.5), where the exponents α_1 and α_2 are the roots of the quadratic (4.4.11). Moreover, at long times the root with the larger real part is dominant. If we assume that Real (α_1) > Real (α_2), then the general solutions (4.4.4) – (4.4.5) simplify at long times to give:

$$X(t) = X_1 e^{\alpha_1 t} \quad \textbf{(4.4.24)}$$

$$Y(t) = Y_1 e^{\alpha_1 t} \quad \textbf{(4.4.25)}$$

Experimentally, we have data regarding the following:

$\alpha = \frac{1}{doubling\ time}$ (i.e. overall growth rate = α_1) **(4.4.26)**

$\frac{X}{X+Y} = \frac{X}{N}$ (i.e. proportion of cells in X) **(4.4.27)**

$\frac{Y}{X+Y} = \frac{Y}{N}$ (i.e. proportion of cells in Y) **(4.4.28)**

Where N = total number of cells. We now explain how we may use this data to estimate the model parameters k_{xy} and k_{yx}.

Substitution of equations (4.4.24) - (4.4.25) into (4.4.1) – (4.4.2) yields:

$\alpha X = 2k_{yx}Y - k_{xy}X.$ **(4.4.29)**

$\alpha Y = k_{xy}X - k_{yx}Y.$ **(4.4.30)**

Addition of equations (4.4.29) and (4.4.30) allows us to eliminate k_{xy} and therefore determine k_{yx} as follows:

$\alpha\,(X+Y) \equiv \alpha N = k_{yx}Y$ **(4.4.31)**

Which, on re-arrangement, supplies:

$\boldsymbol{k_{yx} = \alpha \frac{N}{Y} = \frac{\alpha}{(Y/N)}}.$ **(4.4.32)**

Using equation (4.4.30) we are able to determine k_{xy} in terms of α and $\frac{Y}{N}$:

$(\alpha + k_{xy})X = 2k_{yx}Y = 2\alpha N$ **(4.4.33)**

Or, equivalently,

$(\alpha + k_{xy})\frac{X}{N} = 2\alpha$ **(4.4.34)**

So that now we have,

$k_{xy} = \frac{2\alpha}{X/N} - \alpha$ **(4.4.35)**

Where $\frac{X}{N} \equiv 1 - \frac{Y}{N}$ and so that:

$\boldsymbol{k_{xy} = \left(\frac{2\alpha}{(1-Y/N)}\right) - \alpha = \alpha \frac{(1+Y/N)}{(1-Y/N)}}$ **(4.4.36)**

Equations (4.4.32) and (4.4.36) depict that both k_{xy} and k_{yx} depend on the proportion of cells in Y, however, in differing manners. Using these expressions and noting that $0 \leq \frac{Y}{N} < 1$, it is clear that the following inequalities hold true:

$$k_{xy} > \alpha \quad \textbf{(4.4.37)}$$

$$k_{yx} = \frac{\alpha}{(Y/N)} > \alpha \quad \textbf{(4.4.38)}$$

Figure 4.5 shows the effect of varying $\frac{Y}{N}$ and α on the value of the rate parameters k_{xy} and k_{yx}. The value of k_{xy} increases as the proportion of cells in Y increases. However, as α decreases, the rate at which k_{xy} increases is reduced. The opposite is true for the rate parameter k_{yx}, which decreases as the proportion of cells in Y increases. Again, as α decreases, the rate at which k_{yx} decreases is reduced. Both k_{yx} and k_{xy} depend on the overall growth rate of the model in the same manner: they are both proportional to α.

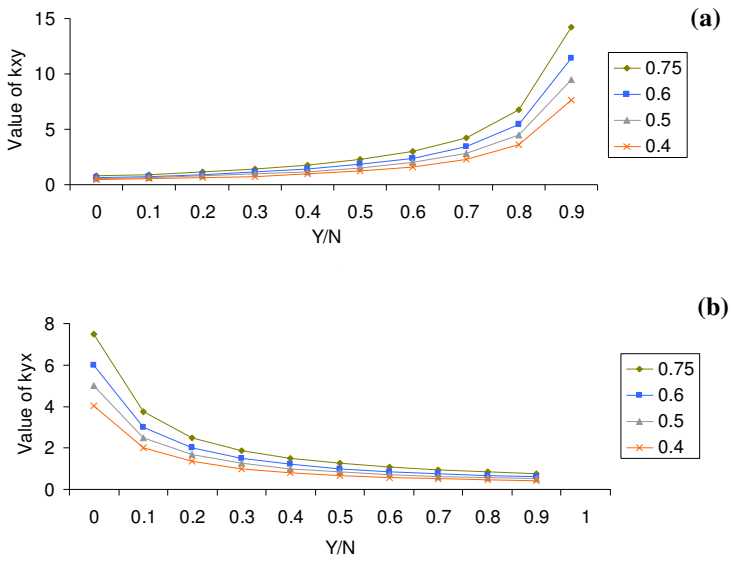


Figure 4.5 Simulation of the effects of the proportion of cells in Y and alpha on the value of the rate parameters (a) k_{xy} and (b) k_{yx}. The graph shown in (a) depicts that as the value of the proportion of cells in Y and alpha increases, the value of k_{xy} increases. It is also evident from (b) that as the value of α decreases, and the proportion of cells increases, there is a reduction in the value of k_{yx}.

Using MATLAB 7 to solve the ODEs, numerical simulations of the two-compartment model were performed and used to investigate how the model parameters k_{xy} and k_{yx} are affected by the proportion of cells in X and Y.

Figure 4.6 shows how, at long times, the proportion of cells in compartments X and Y change as k_{xy} varies. As the value of k_{xy} increases, the proportion of cells in X decrease and the proportion of cells in Y increase. These results are to be expected when we recall that k_{xy} is the rate at which cells move from compartment X to compartment Y and so increasing this rate should increase the proportion of cells in Y.

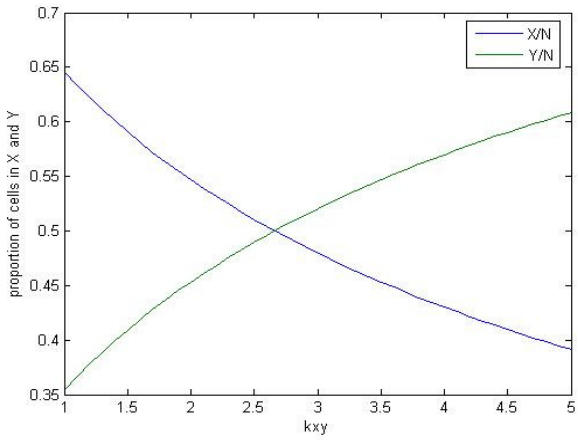


Figure 4.6 How the fraction of cells in X and Y change as k_{xy} varies. The simulation shows the changes in the proportion of cells in both X and Y with respect to changes in k_{xy}. The graph shows that as the value of k_{xy} increases there is an increase in the proportion of cells in Y and a decrease in the proportion of cells in X and shows similar information to figure 4.4. We note that for $k_{xy} < 3$, the proportion of cells in Y is less than the proportion of cells in X, however, at values of k_{xy} greater than 3 (considered to be the threshold value), the proportion of cells in Y exceeds that in X. In this simulation, k_{yx} was fixed at 2.

In addition to simulating how changes in k_{xy} influence the proportion of cells in X and Y, we also investigated how changes in both k_{xy} and k_{yx} affected the number of cells in compartments X and Y. In the first simulation we fixed $k_{xy} = 2$ and $k_{yx} = 5$ whilst in the second we fixed $k_{xy} = 5$ and $k_{yx} = 2$. The results are presented in figure 4.7 and show that, in both cases at long times, the growth of the system is exponential. The total number of cells at the end of the simulations appears to

depend on the choice of values of k_{xy} and k_{yx}. For example, a higher k_{yx} value increases the total number of cells at day 21 of the simulation. The growth of the system will depend on the initial conditions imposed, i.e. the values used for X(t=0) and Y(t=0). With respect to the proportion of cells in X and Y, when a lower k_{xy} value is applied, the proportion of cells in X initially increases, whilst the proportion of cells in Y decreases until both the proportions in each compartment reach steady state values which persist for the duration of the experiment. The opposite appears to be true when a higher value of k_{yx} is used. Initially there is a reduction in the proportion of cells in X whilst the proportion of cells in Y increases. Again, the proportions of cells in each compartment (i.e. $\frac{X}{N}$ and $\frac{Y}{N}$) reaches a plateau which persists for the duration of the simulation.

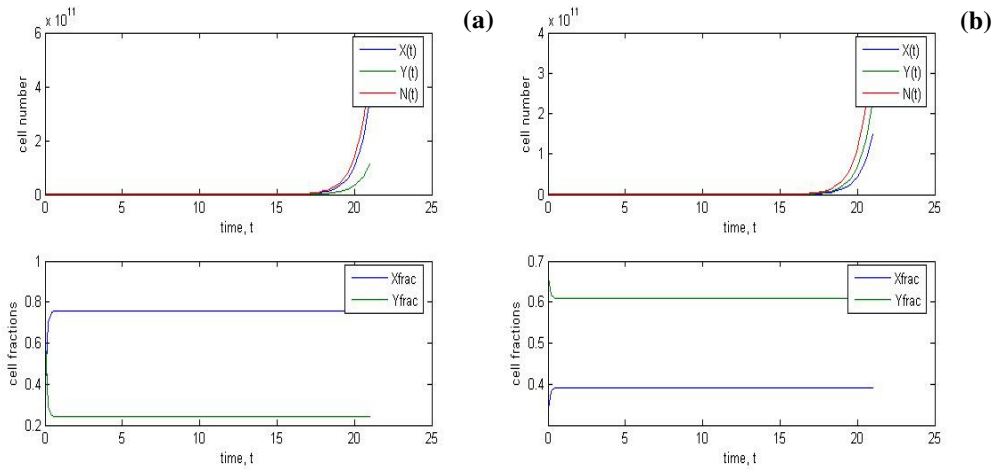


Figure 4.7 Simulating the effects of initial values of k_{xy} and k_{yx} on the proportion of cells in each compartment. The graphs shown in (a) are the results of simulations of the growth of the two-compartment model whereby the value of k_{xy} was fixed at 2 and the value of k_{yx} was fixed at 5. The graphs in (b) show the results of the same simulation, however, the value of k_{xy} was fixed at 5 and the value of k_{yx} was fixed at 2.

The two-compartment model is useful for introducing the techniques involved in mathematical modelling and also shows how information about the cell cycle can be obtained from data using only information about the proportion of cells in X and Y, and the overall growth rate. Although the two-compartment model provides some insight it should be used with caution as it does not distinguish

between the different phases of the cell cycle, assuming instead that cells transit go from a replicating stage (X) to a synthesizing stage (Y). In terms of our flow cytometry data we are essentially 'pooling' data on the proportion of cells in pre-G_1, G_1/G_0 and S into compartment X and the proportion of cells in G_2/M into compartment Y. In the next section, we explain how to extend the two-compartment model into a three-compartment model and, in doing so, how we may exploit the available data more fully.

4.5 Three-Compartment Model

While the two-compartment model provides a more realistic description of the cell cycle than the one-compartment model, it is overly simplistic given the data we are able to generate. We can distinguish G_0, G_1, S, G_2, M and pre-G_1. We also have data with respect to the number of doublings and the proportion of senescent cells. A natural way of extending the model is to decompose the cell cycle into three compartments, as shown in figure 4.8.

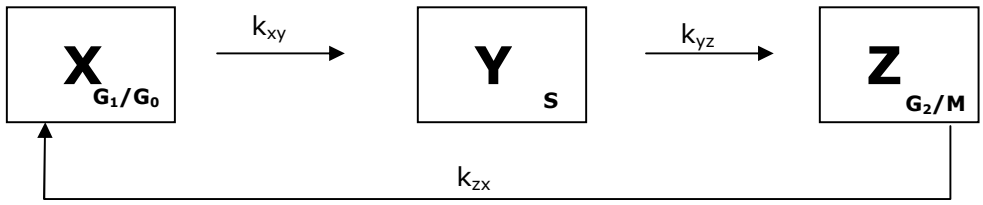


Figure 4.8 Three-compartment mathematical model. The model depicted here describes the cell cycle as a series of three-compartments X (incorporating G_1/G_0), Y (incorporating S) and Z (incorporating G_2/M). Transition of cells between the compartments is described by the rate parameters k_{xy}, k_{yz} and k_{zx}. The three-compartment model utilises the doublings data and incorporates the cell cycle data more accurately than the two-compartment model. Like the two-compartment model, the three-compartment model can also be described as a 'closed' model as there is no exit of cells from the model.

Like the two-compartment model, the three-compartment model, is 'closed' as no cells leave the system. Consequently senescence and cell death data are both neglected. We simply assume that cells begin cycling in X and enter compartment Y at rate k_{xy}. Following completion of the processes which take place in Y, the

cells enter compartment Z at rate k_{yz}. From Z, two cells enter X at rate k_{zx}, and commence their cell cycles.

The ODEs associated with the model depicted in figure 4.8 are:

$$\frac{dX}{dt} = 2k_{zx}Z - k_{xy}X \quad \textbf{(4.5.1)}$$

$$\frac{dY}{dt} = k_{xy}X - k_{yz}Y \quad \textbf{(4.5.2)}$$

$$\frac{dZ}{dt} = k_{yz}Y - k_{zx}Z \quad \textbf{(4.5.3)}$$

Since ODEs (4.5.1) - (4.5.3) are linear they admit solutions of the form:

$$X(t) = \hat{X}e^{\alpha t} \quad \textbf{(4.5.4)}$$

$$Y(t) = \hat{Y}e^{\alpha t} \quad \textbf{(4.5.5)}$$

$$Z(t) = \hat{Z}e^{\alpha t} \quad \textbf{(4.5.6)}$$

for the constants $\hat{X}$, $\hat{Y}$ and $\hat{Z}$.

As with the two-compartment model, α represents the overall growth rate of the model, and must be determined as part of the solution process. For the two-compartment model, α satisfied a quadratic equation as shown in section 4.4. For the three-compartment model, α satisfies a cubic which may be determined in the following way.

First, we note that:

$$\frac{dX}{dt} = \alpha\hat{X}e^{\alpha t} = \alpha X \quad \textbf{(4.5.7)}$$

$$\frac{dY}{dt} = \alpha\hat{Y}e^{\alpha t} = \alpha Y \quad \textbf{(4.5.8)}$$

$$\frac{dZ}{dt} = \alpha\hat{Z}e^{\alpha t} = \alpha Z \quad \textbf{(4.5.9)}$$

Substituting from (4.5.7) – (4.5.9) into (4.5.1) – (4.5.3), we obtain the following system of simultaneous equations for $\hat{X}, \hat{Y}$ and $\hat{Z}$ in terms of the exponent, α, and the system parameters, k_{xy}, k_{yz} and k_{zx}.

$$\alpha\hat{X} = -k_{xy}\hat{X} + 2k_{zx}\hat{Z} = 2k_{zx}\hat{Z} - k_{xy}\hat{X} \quad \textbf{(4.5.10)}$$

$$\alpha\hat{Y} = k_{xy}\hat{X} - k_{yz}\hat{Y} \quad \textbf{(4.5.11)}$$

$$\alpha\hat{Z} = k_{yz}\hat{Z} - k_{zx}\hat{Z} \quad \textbf{(4.5.12)}$$

We have experimental measurements of proportions of cells in X, Y and Z, which we denote as $\frac{X}{N}, \frac{Y}{N}$ and $\frac{Z}{N}$, respectively, where N= X+Y+Z. We also know growth rate, α. By combining this data with our model it is possible to derive three algebraic equations for the three unknown rate parameters k_{xy}, k_{yz} and k_{zx} to the data. Re-arranging equations (4.5.10) – (4.5.12) we obtain:

$$(\alpha + k_{xy})X = 2k_{zx}Z \quad \textbf{(4.5.13)}$$

$$(\alpha + k_{yz})Y = k_{xy}X \quad \textbf{(4.5.14)}$$

$$(\alpha + k_{zx})Z = k_{yz}Y \quad \textbf{(4.5.15)}$$

We now show how (4.5.13) – (4.5.15) can be used to obtain expressions for the rate parameters k_{xy}, k_{yz} and k_{zx}. First we note from (4.5.13) – (4.5.15) that:

$$\frac{X}{Z} = \frac{2k_{zx}}{(\alpha+k_{xy})} \quad \textbf{(4.5.16)}$$

$$\frac{X}{Y} = \frac{(\alpha+k_{yz})}{k_{xy}} \quad \textbf{(4.5.17)}$$

$$\frac{Y}{Z} = \frac{(\alpha+k_{zx})}{k_{yz}} \quad \textbf{(4.5.18)}$$

The cubic for α can be derived by noting the following:

$$\left(\frac{X}{Z}\right) = \left(\frac{X}{Y} \times \frac{Y}{Z}\right)$$

This identity enables us to deduce the following relationship between k_{xy}, k_{yz} and k_{zx}:

$$\frac{2k_{zx}}{(\alpha+k_{xy})} = \frac{(\alpha+k_{yz})}{k_{xy}} \times \frac{(\alpha+k_{zx})}{k_{yz}}$$

Or, equivalently,

$$0 = (\alpha + k_{xy})(\alpha + k_{yz})(\alpha + k_{zx}) - 2k_{zx}k_{xy}k_{yz} \quad \textbf{(4.5.19)}$$

Expanding (4.5.19), we deduce:

$$0 = \alpha^3 + (k_{zx}k_{xy}k_{yz})\alpha^2 + (k_{zx}k_{xy} + k_{zx}k_{yz} + k_{xy}k_{yz})\alpha - k_{xy}k_{zx}k_{yz} \quad \textbf{(4.5.20)}$$

This cubic equation for α allows the dominant/long time growth rate of the model to be calculated. Equation (4.5.20) has 3 roots and we know α from the data and we assume that the data corresponds to the root of (4.5.20) which grows most

rapidly and we now utilise (4.5.20), or equivalently (4.5.19), to supply a relationship between the data and the system parameters.

We now explain how we can estimate k_{xy}, k_{yz} and k_{zx} from the experimental data on the doubling rate and the proportions of cells in the different compartments $\frac{X}{N}, \frac{Y}{N}$ and $\frac{Z}{N} = 1 - \frac{X}{N} - \frac{Y}{N}$.

Addition of equations (4.5.10), (4.5.11) and (4.5.12) supplies:

$\alpha(X + Y + Z) = k_{zx}Z$

Or equivalently,

$\frac{\alpha}{k_{zx}} = \frac{Z}{N} = 1 - \frac{(X+Y)}{N}$

And so we have the following:

$$\boldsymbol{k_{zx} = \frac{\alpha}{(1-\frac{(X+Y)}{N})} = \frac{\alpha}{Z/N}} \quad \textbf{(4.5.21)}$$

By manipulating equation (4.5.10), we deduce:

$$\left(\alpha + k_{xy}\right)\frac{X}{N} = 2k_{zx}\frac{Z}{N} \quad \textbf{(4.5.22)}$$

By combining (4.5.21) and (4.5.22) we deduce:

$$\left(\alpha + k_{xy}\right)\frac{X}{N} = 2\alpha \quad \textbf{(4.5.23)}$$

By re-arranging (4.5.23) we derive the following expression for k_{xy} in terms of α and $\frac{X}{N}$:

$$k_{xy} = \frac{2\alpha}{\left(\frac{X}{N}\right)} - \alpha = \frac{\alpha}{\left(\frac{X}{N}\right)}\left(2 - \frac{X}{N}\right) \quad \textbf{(4.5.24)}$$

$$\boldsymbol{k_{xy} = \frac{\alpha\,(2 - X/N)}{X/N}} \quad \textbf{(4.5.25)}$$

It remains to derive an expression for k_{yz}. Manipulation of equation (4.5.12) yields:

$$k_{yz}Y = (\alpha + k_{zx})Z \quad \text{or } k_{yz}\left(\frac{Y}{N}\right) = (\alpha + k_{zx})\frac{Z}{N} \tag{4.5.26}$$

Where, from above, $k_{zx}\left(\frac{Z}{N}\right) = \alpha$

$$\boldsymbol{k_{yz} = \frac{\alpha(1+Z/N)}{Y/N}} \tag{4.5.27}$$

To summarise, by re-arranging and/or adding equations (4.5.10) - (4.5.12) we have derived a series of expressions for the rate parameters k_{xy}, k_{yz} and k_{zx}. These equations show us how data on the proportion of cells in each compartment can be used to estimate the rate parameters. Alternatively, or equivalently, the expressions provide information about how the rate parameters affect the proportion of cells in each compartment as follows:

$$\frac{X}{N} = \frac{X}{(X+Y+Z)} = \frac{2\alpha}{(\alpha+k_{xy})} \tag{4.5.28}$$

$$\frac{Y}{N} = \frac{Y}{(X+Y+Z)} = \frac{k_{xy}}{(\alpha+k_{xy})}\left(\frac{2\alpha}{(\alpha+k_{yz})}\right) \tag{4.5.29}$$

Where $(\alpha + k_{xy})(\alpha + k_{yz}) = \frac{2k_{zx}k_{xy}k_{yz}}{\alpha+k_{zx}}$

And so $\frac{Y}{N} = \frac{\alpha}{k_{zx}k_{yz}}(\alpha + k_{zx})$ **(4.5.30)**

$$\frac{Z}{(X+Y+Z)} = \frac{\alpha}{k_{zx}} \tag{4.5.31}$$

One of the main weaknesses with the two and three-compartment model is that they are 'closed' models as no cells leave the system: cells simply transfer between the compartments. In particular, processes by which cells leave the cell cycle (due to senescence or cell death) are not considered. Data about the proportion of senescent and dead cells is available, however, only proliferating cells are accounted for.

4.6 Five-Compartment Model

The 'closed' three-compartment model utilizes the data available about the number of successful population doublings and the cell cycle. As stated previously, we have data regarding the number of successful population doublings, cell cycle data

and senescence data. A natural way of extending the three-compartment model and to utilise the senescence data is to decompose the three compartments into five compartments (see figure 4.9) and assume that cells can only become senescent from compartment X. When telomeres eventually shorten to a certain point, the cell initiates a senescence in G_1 (equivalent to compartment X) and this arrests the cells and so seems the most realistic way of incorporating senescence into the mathematical model (Bailey & Murnane, 2006). It has also been reported that senescent cells are arrested in G_1 based on evidence that indicated that senescent cells have a G_1 DNA content (Afshari *et al*, 1993).

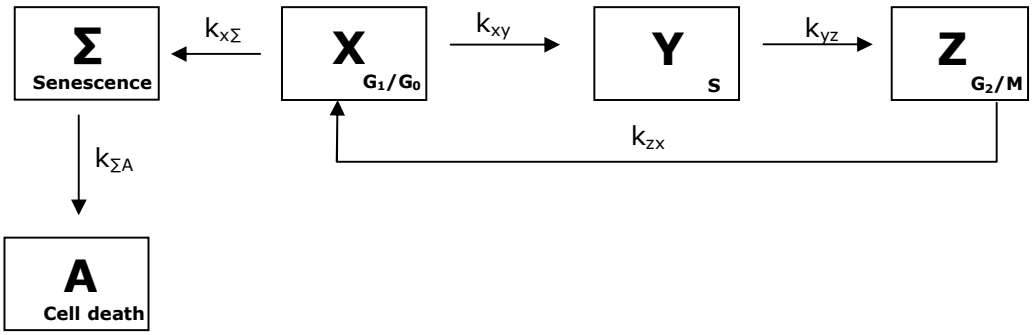


Figure 4.9 Five-compartment mathematical model. The cell cycle is split into five compartments, X (incorporating G_1/G_0), Y (incorporating S), Z (incorporating G_2/M), Σ (incorporating senescence) and A (incorporating pre-G_1). Transition of cells between the individual compartments is denoted by the rate parameters k_{xy}, k_{yz}, k_{zx}, $k_{x\Sigma}$ and $k_{\Sigma A}$. The five-compartment model is considered to be the most accurate of the mathematical models as all the available data is utilised and exit of cells from the model is taken into account.

Our five-compartment model accounts for cells leaving the system via cell death and/or senescence. As shown in figure 4.9, we assume that cycling cells flow between compartments X, Y or Z at rates k_{xy}, k_{yz} and k_{zx}. Thus compartments X, Y and Z are identical to those in our earlier three-compartment model. However, cells may exit the cell cycle from compartment X and enter the senescent compartment, Σ at rate $k_{x\Sigma}$. Cells in compartment Σ either remain there or undergo cell death when they enter compartment A. This process is assumed to occur at rate $k_{\Sigma A}$ (i.e. senescent cells can't return to the cell cycle). In reality, it is possible for cell death to occur at any stage of the cell cycle if conditions are unfavourable or if DNA damage has been sustained, however, to keep the mathematical models

as simple as possible in terms of the ODEs involved, we assume that the rate of cells entering cell death via rate $k_{\Sigma A}$ is much higher than the rate at which it is occurring from compartments X, Y and Z and so these rate parameters can be essentially 'ignored'.

The system of ODEs associated with the five-compartment model depicted in figure 4.9 is as follows:

$$\frac{dX}{dt} = 2k_{zx}Z - \left(k_{xy} + k_{x\Sigma}\right)X \quad \textbf{(4.6.1)}$$

$$\frac{dY}{dt} = k_{xy}X - k_{yz}y \quad \textbf{(4.6.2)}$$

$$\frac{dZ}{dt} = k_{yz}Y - k_{zx}Z \quad \textbf{(4.6.3)}$$

$$\frac{d\Sigma}{dt} = k_{x\Sigma}\mathrm{X} - \mathrm{k}_{\Sigma\mathrm{A}}\Sigma \quad \textbf{(4.6.4)}$$

$$\frac{dA}{dt} = k_{\Sigma A}\Sigma \quad \textbf{(4.6.5)}$$

As with the two- and three-compartment models, the ODEs are linear and so possess solutions of the form:

$$X(t) = \hat{X}e^{\alpha t} \quad \textbf{(4.6.6)}$$

$$Y(t) = \hat{Y}e^{\alpha t} \quad \textbf{(4.6.7)}$$

$$Z(t) = \hat{Z}e^{\alpha t} \quad \textbf{(4.6.8)}$$

$$\Sigma(t) = \hat{\Sigma}e^{\alpha t} \quad \textbf{(4.6.9)}$$

$$A(t) = \hat{A}e^{\alpha t} \quad \textbf{(4.6.10)}$$

for some constants $\hat{X}, \hat{Y}, \hat{Z}, \hat{\Sigma}$ and $\hat{A}$.

As with the two- and three-compartment models, the growth rate of the dependent variables is governed by α, which must also be determined as part of the solution process. For the five-compartment model, the compartments X, Y and Z determine the overall behaviour of the model and are independent of Σ and A (i.e. the ODEs for X, Y and Z are independent of Σ and A). As a result, α may be determined from the ODEs for X, Y and Z and so, as with the three-compartment model, α satisfies a cubic equation of the following form:

$$0 = \left(\alpha + k_{xy} + k_{x\Sigma}\right)\left(\alpha + k_{yz}\right)(\alpha + k_{zx}) - 2k_{zx}k_{xy}k_{yz} \quad \textbf{(4.6.11)}$$

In order to determine (4.611), first we note that:

$$\frac{dX}{dt} = \alpha X \quad \textbf{(4.6.12)}$$

$$\frac{dY}{dt} = \alpha Y \quad \textbf{(4.6.13)}$$

$$\frac{dZ}{dt} = \alpha Z \quad \textbf{(4.6.14)}$$

$$\frac{d\Sigma}{dt} = \alpha \Sigma \quad \textbf{(4.6.15)}$$

$$\frac{dA}{dt} = \alpha A \quad \textbf{(4.6.16)}$$

As with the two- and three-compartment model, when the general solutions (4.6.12) - (4.6.16) are substituted into the original ODEs (4.6.1) - (4.6.6) we generate the following algebraic equations for the constants $\hat{X}, \hat{Y}, \hat{Z}, \hat{\Sigma}$ and $\hat{A}$:

$$\alpha\hat{X} = 2k_{zx}\hat{Z} - \left(k_{xy} + k_{x\Sigma}\right)\hat{X} \quad \textbf{(4.6.17)}$$

$$\alpha\hat{Y} = k_{xy}\hat{X} - k_{yz}\hat{Y} \quad \textbf{(4.6.18)}$$

$$\alpha\hat{Z} = k_{yz}\hat{Y} - k_{zx}\hat{Z} \quad \textbf{(4.6.19)}$$

$$\alpha\hat{\Sigma} = k_{x\Sigma}\hat{X} - k_{\Sigma A}\hat{\Sigma} \quad \textbf{(4.6.20)}$$

$$\alpha\hat{A} = k_{\Sigma A}\hat{\Sigma} \quad \textbf{(4.6.21)}$$

Re-arranging equations (4.6.17) - (4.6.19) and noting that for a non-trivial solution we require $\hat{X}, \hat{Y}, \hat{Z} > 0$, we deduce that α must satisfy the following expression:

$$0 = \begin{vmatrix} \boldsymbol{X} & \boldsymbol{Y} & \boldsymbol{Z} \\ \alpha + k_{xy} + k_{x\Sigma} & 0 & -2k_{zx} \\ -k_{xy} & \alpha + k_{yz} & 0 \\ 0 & -k_{yz} & \alpha + k_{zx} \end{vmatrix} \begin{matrix} \boldsymbol{X} \\ \boldsymbol{Y} \\ \boldsymbol{Z} \end{matrix}$$

By expanding the above determinant we generate the cubic for α in terms of k_{xy}, k_{yz}, k_{zx} and $k_{x\Sigma}$ as shown in equation (4.6.11).

The five-compartment model has five unknown rate parameters k_{xy}, k_{yz}, k_{zx}, $k_{x\Sigma}$ and $k_{\Sigma A}$. We aim to derive expressions for them in terms of the experimental data

(i.e. $\frac{X}{N}, \frac{Y}{N}, \frac{Z}{N}, \frac{\Sigma}{N}, \frac{A}{N}$ and α, where $\hat{N} = \hat{X} + \hat{Y} + \hat{Z} + \hat{\Sigma} + \hat{A}$. We now explain how equations (4.6.17) - (4.6.21) may be used to derive these relations. Addition of equations (4.6.17) - (4.6.21) yields the following equation for k_{zx}.

$$\boldsymbol{k_{zx} = \frac{\alpha}{Z/N}} \quad \textbf{(4.6.22)}$$

Using equation (4.6.19) we derive the following equation for the rate parameter k_{yz}.

$$(\alpha + k_{zx})\left(\frac{Z}{N}\right) = k_{yz}(\frac{Y}{N}) \quad \textbf{(4.6.23)}$$

Substituting from (4.6.22) with k_{zx} in equation (4.6.24) we deduce that

$$\boldsymbol{k_{yz} = \frac{\alpha(1+Z/N)}{Y/N}} \quad \textbf{(4.6.24)}$$

In a similar manner equation (4.6.18) supplies an equation for the rate parameter k_{xy}.

$$\boldsymbol{k_{xy} = \alpha \frac{\{1+(Y/N)+(Z/N)\}}{X/N}} \quad \textbf{(4.6.25)}$$

We now have expressions for k_{zx}, k_{xy} and k_{yz} in terms of the cell cycle data. It remains to derive equations for $k_{x\Sigma}$ and $k_{\Sigma A}$. From equation (4.6.17), we have:

$$\left(k_{x\Sigma} + k_{xy} + \alpha\right)\frac{X}{N} = 2k_{zx}\frac{Z}{N} \quad \textbf{(4.6.26)}$$

From this we are able to state the following:

$$k_{x\Sigma} = \frac{2k_{zx}Z/N}{(X/N)} - \alpha - k_{xy} \quad \textbf{(4.6.27)}$$

$$k_{x\Sigma} = \frac{2\alpha}{(X/N)} - \alpha - \frac{\alpha}{(X/N)}(1 + \frac{Y}{N} + \frac{Z}{N}) \quad \textbf{(4.6.28)}$$

$$\boldsymbol{k_{x\Sigma} = \frac{\alpha}{(X/N)}\ (1 - \frac{X}{N} - \frac{Y}{N} - \frac{Z}{N})} \quad \textbf{(4.6.29)}$$

The ODE for A can be simply re-arranged to give the following equation for $k_{\Sigma A}$:

$$\boldsymbol{k_{\Sigma A} = \frac{\alpha(A/N)}{\Sigma/N}} \quad \textbf{(4.6.30)}$$

We now have five equations for the five unknown rate parameters. These equations enable us to utilise fully the available data to generate conclusions regarding the effects of RHPS4, and other anti-tumour agents on the HCT116 cell line that would be difficult to determine otherwise.

4.7 Suitability of Cell Cycle Data Analysis

One of the major challenges of systems biology is that of data quality. Systems biology approaches rely heavily on information found in public databases which are often incomplete or are not properly annotated (Aderem, 2005). For a systems biology approach to be successful, the data needs to be of a high quality and the way in which it is derived needs to be rigorous to meet the demands of this approach. The manual method described in chapter 3 is sufficient for the analysis of trends and the effects of RHPS4 and other anti-tumour agents. For the short-term data presented in chapter 3, this approach was used and was sufficient, however, for a systems biology approach to the analysis of RHPS4 and the modelling of the cell cycle, this method is considered to not be sufficient nor rigorous enough. The data used for modelling, although derived in the same manner as the short term data presented in section 3.2.3 and 3.2.4, the long term data that will be used for modelling requires more accurate analysis since this data will be used in a mathematical model to probe the mechanism of action of RHPS4. The manual method, although appropriate for the analysis of G_1/G_0, S and G_2/M, is not appropriate or vigorous enough for the analysis of the pre-G_1 fraction. It appears that the analysis of the pre-G_1 fraction using both software packages winMDI and Cylchred would be the best way forward. With respect to the analysis of cell death and the pre-G_1 fraction, equation 3.1 should be used which utilizes the total DNA intensity of the population of cells and the average DNA intensity of a G_1 cell in order to define the proportion of cells in this fraction. In doing this, there will be a higher degree of confidence in the data and reduced noise and also lead to the feasibility of data pooling.

4.8 Conclusions

The mathematical models we have stated are simple enough to yield useful conclusions regarding RHPS4 and other anti-tumour agents that would be difficult to discern from the data alone. Our model enables us to draw conclusions about the effects on the cell cycle of RHPS4 and other anti-tumour agents. However, without the mathematical models, the rates of transfer from one phase to another are difficult to determine. For example, the flow cytometer provides information about the proportion of cells in each of the cell cycle phases under normal circumstances and in response to anti-tumour agents. However, it offers no information about the cell cycle kinetics and in particular the rate of flow of cells between the compartments. Such information provides us with further insight into the mode of action of different anti-tumour agents, in particular, RHPS4, and enables us to predict where in the cell cycle the drugs act. Armed with such knowledge, it should be possible to tailor treatment protocols to particular types of cancers, for example would we know whether a particular type of brain tumour would respond in the same way as another?

The data that is derived in the laboratory with respect to RHPS4 and other anti-tumour agents allows the emergent properties of these agents over a period of time to be gathered, however, the mathematical model aims to link these emergent properties to a possible protein/molecular target by turning the experimental data into a series of 'k' values which give us information with respect to the kinetics and how these change with respect to time and concentration of agent. How we get from the experimental data to the 'k' values is illustrated here:

- The data with regards to the proportion of cells in senescence is calculated as a proportion of 100 (as the proportions of cells in each phase must be equal to 100%) and so we have $\frac{\Sigma}{N}$.
- The data with regards to the proportions of cells in G_1/G_0, S, G_2/M and pre-G_1 is converted into proportions of cells in X, Y, Z and A, respectively and so we have the following: $\frac{X}{N}, \frac{Y}{N}, \frac{Z}{N}$ and $\frac{A}{N}$.

- We also assume that senescent cells are incorporated in the cell cycle data with G_0/G_1 cells when analysed on the flow cytometer since senescent cells have the same DNA content as G_1 cells and so we assume that $\frac{X}{N}$ is actually $\frac{X+\Sigma}{N}$ and so the 'true' proportion of cells in X is calculated as $\frac{X+\Sigma}{N} - \frac{X}{N}$ (Afshari *et al*, 1993).
- The total numbers of cells at each time point is calculated as $(1 + number\ of\ doublings) \times (2 \times 10^5)$. We know from the experiment that the cells are subbed as in section 2.2.9 at a density of 2 x 10^5 and will allow the overall growth rate of the cell population to be calculated.
- The doublings data is plotted on a graph and the gradient at each time point is derived and from this a value of Td is calculated. Td is the time taken to complete one population doubling and is calculated for each time point as $\frac{total\ number\ of cells}{gradient\ of\ graph\ at\ that\ point}$.
- The Td value is now converted into a value for α which is the overall growth rate of the population of cells at each time point and for each concentration and is calculated as $\frac{\ln 2}{Td}$.
- We now have the data required to convert the experimental data into the 'k' values, for example, the value of k_{zx} is calculated at each time point using the values for α and $\frac{Z}{N}$ and the values of k_{yz} is also calculated using the values for α and $\frac{Z}{N}$ in addition to the values for $\frac{Y}{N}$.
- The data is used to calculated 'k' values for each time point and concentration and the results are presented in chapter 5.

The mathematical models derived in this chapter are extremely simple models. In the next chapter they will be challenged with data from a well-characterised drug with a known mechanism of action and data from RHPS4, whose mechanism of action still remains to be elucidated. The mathematical models are by no means the end in determining the mechanism of action of anti-tumour agents; they simply allow the data to speak for itself. As such, they represent a starting point for the

development of more realistic models; however, still provide preliminary insights into the mode of action of drugs.

In the next chapter, we explain the experimental data derived for RHPS4 and doxorubicin with regards to cell cycle, senescence and population doublings. The data is applied to the mathematical models described in this chapter to demonstrate that conclusions can differ depending on the mathematical model used.