THE UNIVERSITY OF WARWICK

Exploration of the intercellular heterogeneity of topotecan uptake into human breast cancer cells through compartmental modelling

S. Y. Amy Cheung[a], Neil D. Evans[b], Michael J. Chappell[b], Keith R. Godfrey[b], Paul J. Smith[c] and Rachel J. Errington[c]

[a]School of Pharmacy and Pharmaceutical Sciences, Coupland III Building, University of Manchester, M13 9PL, UK
[b]School of Engineering, University of Warwick, Coventry, CV4 7AL, UK
[c]Department of Medical Biochemistry and Immunology, School of Medicine, Cardiff University, CF 14 4XN, UK

Abstract

A mathematical multi-cell model for the *in vitro* kinetics of the anti-cancer agent topotecan (TPT) following administration into a culture medium containing a population of human breast cancer cells (MCF-7 cell line) is described. This nonlinear compartmental model is an extension of an earlier single-cell type model and has been validated using experimental data obtained using two-photon laser scanning microscopy (TPLSM).

A structural identifiability analysis is performed prior to parameter estimation to test whether the unknown parameters within the model are uniquely determined by the model outputs. The full model has 43 compartments, with 107 unknown parameters, and it was found that the structural identifiability result could not be established even when using the latest version of the symbolic computation software MATHEMATICA. However, by assuming that *a priori* knowledge is available for certain parameters, it was possible to reduce the number of parameters to 81, and it was found that this (Stage Two) model was globally (uniquely) structurally identifiable. The identifiability analysis demonstrated how valuable symbolic computation is in this context, as the analysis is far too lengthy and difficult to be performed by hand.

Keywords: Compartmental models, structural identifiability, topotecan, drug kinetics.

1. Introduction

In recent years, anti-cancer drugs with the ability to inhibit DNA topoisomerase I, have become one of the main foci in drug discovery. Topotecan (TPT, Hycamtin®), a cytotoxin alkaloid chemotherapy drug, is a semi-synthetic, water-soluble derivative of the drug camptothecin (CPT) [1]. CPT was originally isolated from a Chinese tree, *Campthotheca acuminata* which is widely found in South East Asia where it is known as Xi Shu. The main drawbacks of CPT are: (i) its poor water solubility; (ii) its lactone instability; (iii) the reversibility of the drug–target interaction; (iv) its drug resistance; and (v) its high toxicity [1, 2]. In contrast, the CPT analogue TPT has a more stable structure [3], and clinical evidence for the cytotoxic properties of TPT against lung, breast and ovarian cancers is well known [1].

Topoisomerase I, a monomeric protein, is an enzyme that acts to relax negative and positive super coils that accumulate in DNA during DNA replication [4] and transcription (S-phase), chromosome segregation and the efficient traverse of mitosis [5]. Topoisomerase I is necessary within the nucleus for cell proliferation and its level is constant throughout the cell cycle [6].

Acquired (external factors) or inherited gene mutations lead to the alteration of normal cell proliferation signals and hence a loss of cell cycle controls [7]. This leads to uncontrolled cell

proliferation and, as a consequence, tumours develop and become established. The cancer cells are characterised by a rapid proliferation rate which is paralleled by topoisomerase I activity. Since TPT activity is sensitive to proliferating cells and is S-phase specific [8], cancer cells are more likely to be targeted by the drug due to the increased levels of topoisomerase I.

TPT undergoes rapid, reversible, pH-sensitive, non-enzymatic hydrolysis [9] from a ring-closed lactone form (TPT_L) that predominates at low pH (<4) to an open-ringed hydroxyl form (TPT_H), which predominates at high pH (>10). The pharmacologically active lactone form, TPT_L, is a DNA topoisomerase I inhibitor which binds to the nuclear DNA. It disturbs the process of replication resulting in DNA breaks (S-phase) by trapping topoisomerase I and DNA in a covalently bound ternary complex. It also interferes with the catalytic function of the enzyme without trapping the covalent complex. As a result, cell division is prohibited, and cell arrest is caused in the target area [10].

Identification of dissimilitude in the uptake kinetics into human breast cancer cells (MCF-7 cell line) of the topoisomerase I inhibitor topotecan (Hycamtin®; SmithKline Beecham, Pharmaceutics) is essential for the development of therapeutic agents in oncology. The search for cell heterogeneity corresponds to a search for knowledge of specific factors involved in the differentiation of subpopulations in most neoplasms. Evidence of heterogeneous characteristics expressed in clonal cancer cells has been recorded in the literature [11, 12]. This heterogeneity is further supported in this paper by compartmental modelling of drug uptake.

In this work, a multi-cell compartmental model is derived to describe the *in vitro* kinetics of TPT following administration into a culture medium containing a population of human breast cancer cells (MCF-7 cell line) in suspension. The proposed model is an extension of that presented in [13], which did not differentiate between the behaviours of different cells. Thus, in the previous model, a heterogeneous population of cells was treated as a single averaged homogeneous type. A similar approach was used in [14, 15] for compartmental models that used high performance liquid chromatography (HPLC) data during parameter estimation.

Although the population of 10^5 cells modelled in this study initially originated from a single cloning stream, evidence of the inter-cellular heterogeneity of breast cancer cells has been found in *in vitro* experiments [16, 17].

The previous model [13] did not seek to address inter-cellular heterogeneity, but rather sought to represent the average population response. This established a firm foundation for the development of a more complex model that contains more compartments and parameters. Therefore, rather than treating the cells as a homogeneous population, the possible features of inter-cellular heterogeneity in TPT drug uptake is investigated through a model that includes different cell types. The source and effects of heterogeneity in cellular uptake kinetics is sought from estimates for the unknown parameters in the compartmental model. The experimental data used for the parameter estimation were collected from an *in vitro* study of the interaction of a culture of human breast cancer (MCF-7) cells and TPT using two-photon laser scanning microscopy (TPLSM).

Topotecan is a UV-excitable camptothecin and these auto-fluorescent properties have been exploited to evaluate drug resistance in differently derived cell lines using confocal microscopy [18]. The binding characteristics of the drug to DNA were observed in spectroscopy studies by using the high two-photon absorption cross-section properties demonstrated by TPT [19].

The experimental data were collected for the model using TPLSM; investigations [20] showed that TPT is UV light excitable, so that TPT will also be infrared two-photon excitable. The principle of two-photon excitation of fluorescence [21] is that two photons of longer wavelength light are simultaneously absorbed by a fluorochrome that would normally be excited by a single photon with a shorter wavelength. Hence, the fluorochrome excitation is limited to the point of focus by the non-linear optical absorption property of two-photon excitation.
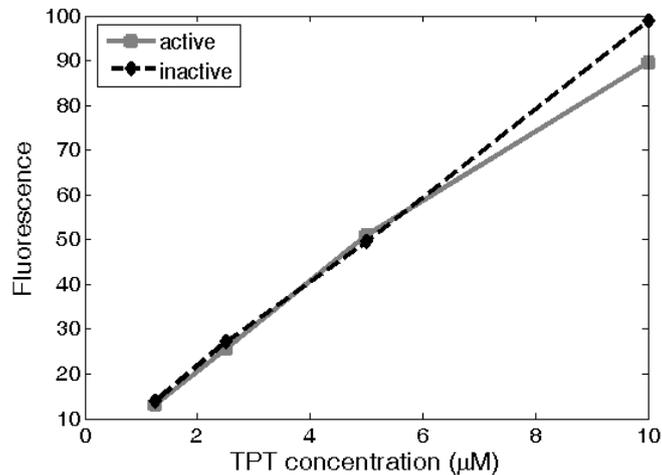

Fig. 1. The fluorescence intensity of topotecan in MCF-7 breast cancer cells, plotted against the concentration of the active and inactive form of the drug.

The advantages given by the features of non-linear excitation are that the images obtained have high contrast and are free of out-of-focus light. Also, this method is particularly beneficial to live cell imaging as the light is only restricted in the focal plane, thereby decreasing photo bleaching of the indicator and photo-damage to the cells. Although this method is not able to differentiate the active form (TPT_L) from the inactive form of drug (TPT_H), as shown in Figure 1, it has been considered to be a reliable method for the collection of high quality data. The experimental data only provide the total concentration of drug within the extracellular location, cytoplasm and nucleus (see Figure 2).
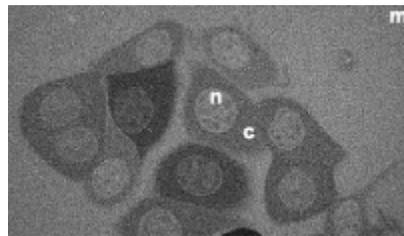
Fig. 2. A TPLSM image of topotecan localisation in MCF-7 breast cancer cells; where **n** denotes the nucleus, **c** the cytoplasm and **m** the medium.

The experiment performed to collect data for the purposes of parameter estimation imposes an output structure on the model. This output structure comprises the functions of the model variables that are directly measured in the proposed experiment (the output, or measurement, vector). Prior to the parameter estimation for the new multi-cell model, a structural identifiability analysis [22] is vital to test whether the output structure corresponding to the TPLSM experiment uniquely determines the unknown parameters. A globally identifiable model indicates that the estimated parameters are unique for a given model output vector.

Structural identifiability is independent of the experimental data and the analysis of it is a fundamental prerequisite to experiment design, system identification and parameter estimation.

The ultimate goal of this paper is the use of compartmental modelling techniques to implement a valid model to seek for the identification and verification of the importance of heterogeneity in the uptake kinetics of the drug in human cancer cells.

2. The multi-cell model

For a single tumour, not all the cells are homologous; rather there is a range of cells in a population expressing many different phenotypes. Cells might vary according to their structure (e.g., growth rate and morphology) and/or behaviour (e.g., invasion and metastasis) [23]. The verification and importance of heterogeneity of these cells is therefore explored using a new mathematical multi-cell compartmental model [24, 25] of the interaction between TPT and a population of MCF-7 breast cancer cells [8].

The experiment, using TPLSM, to track the delivery of TPT to the sub-cellular compartments of a population of MCF-7 cells, was performed on a cultural medium of 10^5 cells in suspension. Data for 13 individual cells were collected and are used to validate the multi-cell model proposed in this paper. The heterogeneous population of cells is assumed, for simplicity, to comprise 13 types corresponding to the cells for which data are available. The innate (as opposed to acquired) heterogeneity in the loading of drug within the population of MCF-7 cells was a striking feature of observations made during the experiment [8], which suggested the existence of inter-cellular heterogeneity.

2.1. Kinetics of the drug action

The principles of compartmental modelling were used to describe the behaviour of the system, that is, the uptake kinetics of the anti-cancer agent topotecan immediately after administration to a population of breast cancer cells.

Chemical reactions are assumed to occur, in accordance with the Law of Mass Action, at a rate that is directly proportional to the product of the concentrations of reactants [26]. The constant of proportionality is the corresponding rate constant. This principle is used in the proposed model to describe the reversible hydrolysis of TPT-L, the nuclear binding of TPT-L and the dissociation of bound drug. The transfers of drug between different physical locations are treated as first order processes. There is only one rate constant for the irreversible transfers corresponding to the dissociation of bound drug from the nucleus to the cytoplasm (as TPT_L or TPT_H); while the reversible transfers between locations have two constants (one for the forward direction and one for the reverse direction).

2.2. Complexity of the model

The basic model of one cell type proposed in [13] consists of 7 compartments and has 11 parameters. Using the same underlying set of experimental data for parameter fitting, the basic model is extended to one that is composed of 43 compartments (involving 109 parameters, 2 of which are known *a priori*), which describes the kinetics within a population of 13 distinct cell types. The extension of the model from 1 cell type to 13 cell types increases the complexity of the system with new pathways being added to account for the flux of the drug to each cell type. Both models were designed specially for parameter estimation using TPLSM data.

4

The increase in complexity of the system also increases the difficulty in performing a structural identifiability analysis, since a more complex system requires higher computational power to complete the analysis. Once the model is shown to be structurally identifiable, this complexity also affects parameter estimation. As the number of parameters increases, the time and computational power required for the process is increased as a consequence. The fitted proposed model is a stiff system [27] and so a specialist simulation and parameter estimation tool is used that is able to handle such a system (FACSIMILE, MCPA Software, UK). When the number of parameters is increased, there are an increased number of fluxes operating on different timescales, which increases the computational time within the simulation software used to handle the parameter estimation process.

2.3.   Assumptions of the compartmental model for 13 cell types

The new multi-cell compartmental model is a "model of process" [28]. The representation of this model involves equations describing the reactions and transfers within the system. These are written based on the basic laws of the processes that constitute the system. Therefore, the simulation results obtained in the virtual experiment on the model can then be compared with real experimental data. Assumptions are made when developing the model in order to fully describe the features of the real life situation. The validation of the assumptions is made once the structural identifiability and parameter estimation have been performed.

Assumption 1: The Law of Mass Action applies and is used to model all the chemical reactions within the system.

Assumption 2: The total concentration of drug in the experimental pool is divided into the medium, which is the environment in which the drug is added, and the extracellular region, which is the location in which the cells are situated [13]. Figures 3(a) and (b) show the response of one cell for which data were collected during the TPLSM experiment. The mixing that takes place in the medium, which necessitates this assumption, is evident in Figure 3(a).

(a)   (b)
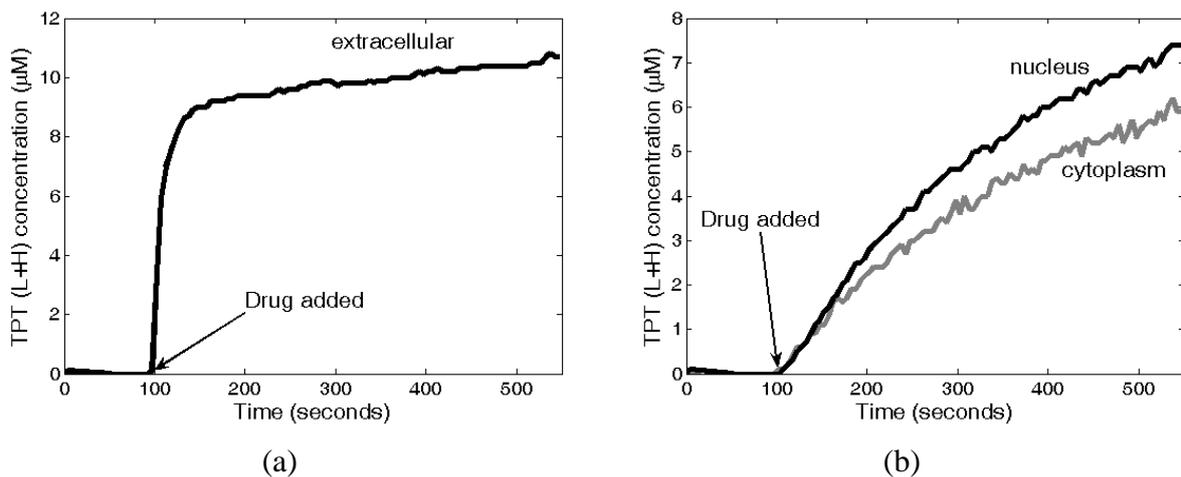
Fig. 3. The total concentration of TPT (TPT_L plus TPT_H) plotted against time.

Assumption 3: The rates of hydrolysis between the active and inactive forms of the drug in the medium and extra-cellular location are the same for each particular direction, and the rate constants are denoted by k_{om} and k_{cm}, respectively.

Assumption 4: It is assumed that only the active form of TPT in the extracellular location is able to flow into and out of the cell, with the influx and efflux taken to be first order processes

with rate constants k_i and k_e, respectively. These assumptions are supported by previous experimental observations [14] and the physical properties of the hydroxy acid form with respect to the lipid bilayer traverse [29].

Assumption 5: The target site contains only the DNA complex, which is formed by the combination of DNA and topoisomerase I. Only the active form of drug, TPT_L, will bind to the DNA complex [19] and the entire drug in the nucleus is bound. Further experiments are essential to validate such a simplification of the real situation and it is important to investigate the possibility of modelling unbound drug located in the nucleus [13].

Assumption 6: In accordance with the Law of Mass Action, the product of the concentrations of available binding sites B_F, and cytosolic TPT_L is assumed to be proportional to the rate at which TPT_L binds to DNA. The concentration B_F can be rewritten as $(B_T - L_n)$ by applying suitable conservation laws [30], where B_T is the total concentration of sites available for TPT_L to bind to and L_n denotes the concentration of bound TPT_L in the nucleus.

Assumption 7: The dissociation of the drug from the nucleus can be to both active and inactive forms, i.e., to cytosolic TPT_L or TPT_H. The dissociation to either form occurs at a first order rate, $k_{dl}L_n$ and $k_{dh}L_n$, respectively.

2.4. A two-compartment model for the reversible hydrolysis of TPT

The drug TPT consists of two forms [14, 15], an active lactone form (TPT_L) and an inactive hydroxy acid form (TPT_H). Experiments indicate that for higher pH, the initial dose of TPT_L more rapidly undergoes hydrolysis to TPT_H and that the equilibrium concentration of TPT_L decreases [14]. A simple model with two compartments (see Figure 4) can be used to describe this reversible hydrolysis of TPT_L in solution; as was done in [14] for TPT in buffered solutions of different pH (in the range from 6.8 to 8.0). This simple model, where L_m and H_m denote the respective concentrations of TPT_L and TPT_H in the buffer (medium), forms the foundation for the overall multi-cell model. The pH in the medium during the experiments was 7.2 and so the values for the rate constants (for this pH) from [14] were used. Although the experiment only provides observations for the total concentration of TPT for estimation purposes, the model is able to predict individual concentrations of TPT_L and TPT_H.
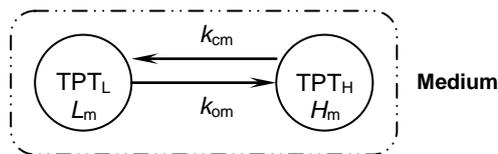
Fig. 4. Schematic of two compartmental model of the drug hydrolysis.

2.5. The multi-cell compartmental model

The model consists of two groups of pools; the first group consists of the medium (denoted by subscript m), in which the cells grow, and an extra-cellular location pool (denoted by subscript e) that contains all of the individual cells. Then, within each cell is another group that comprises two pools corresponding to the cytoplasm (denoted by subscript c) and the nucleus (denoted by subscript n). A two-compartment model (Figure 4) is used to model the reversible hydrolysis of the drug in the drug medium, extra-cellular location and the cytoplasm.

The model, shown in Figure 5, is made up of 43 compartments, each compartment representing a well-mixed homogeneous subsystem of the overall system. These subsystems

6

represent the concentration of the drug, in its active or inactive form, in a particular location with respect to time. The arrows represent the flows between compartments. These transfers can be as a result of the reversible hydrolysis process, the rate of flow of drug from one location to another location or the binding/association of the drug in the nucleus.
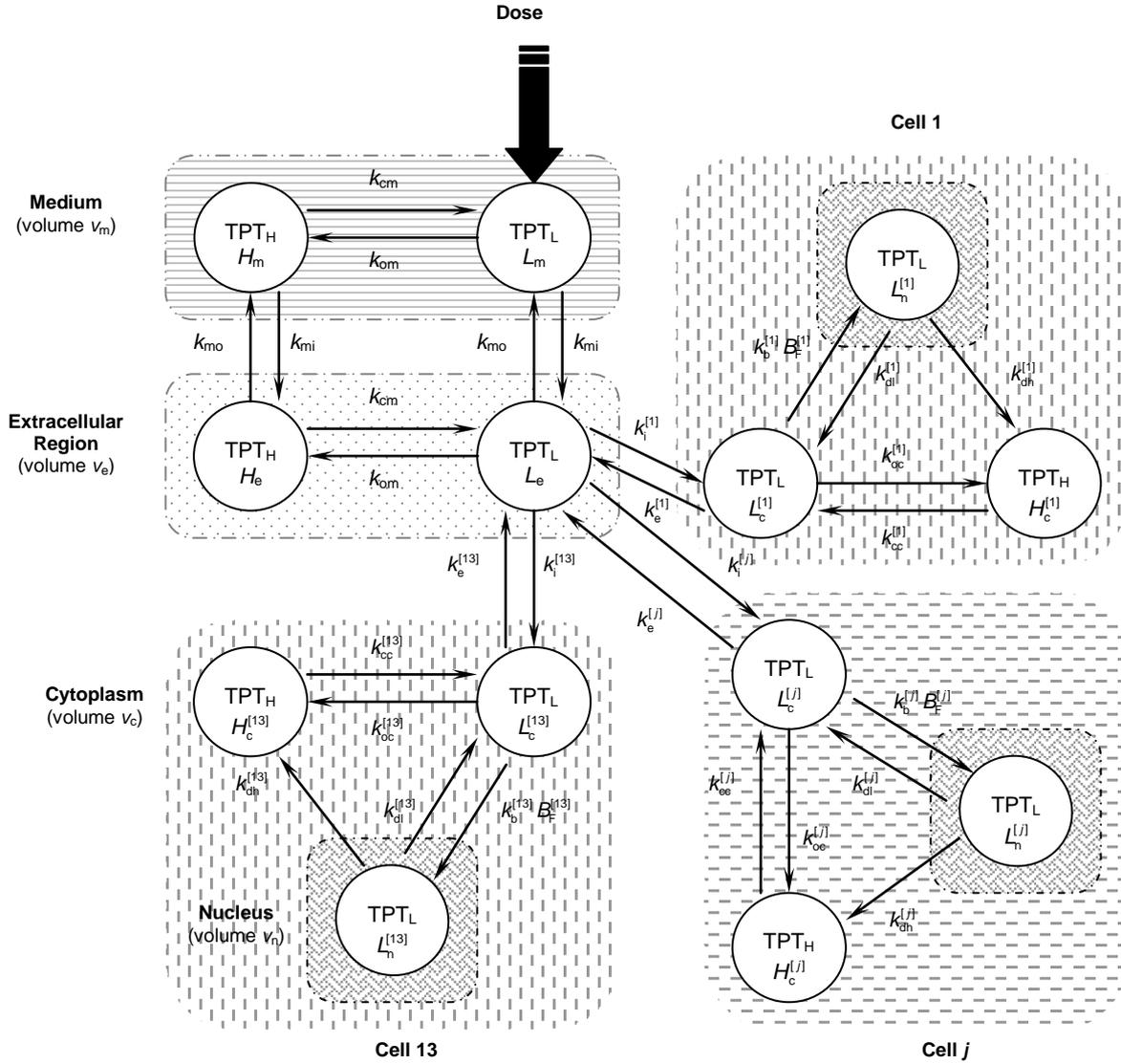
Fig. 5. Schematic of the proposed mathematical model for the *in vitro* kinetics of TPT, including its delivery to the DNA target in each cell nucleus. The concentration of free target sites (in cell type *j*) for TPT$_L$ is denoted by $B_F^{[j]}(t) = \left(B_T^{[j]} - L_n^{[j]}(t) \right)$.

In the experiment a dose of TPT$_L$ was injected into the medium in order to give an overall concentration of 10 μM (in the full 2mL physical medium). In the model this dose corresponds to time $t = 0$ and provides the initial concentration of the active form of drug (TPT$_L$) in the medium, $L_m(0)$. Mixing between the medium and extracellular locations then takes place with first order flows with rate constants k_{mi} and k_{mo}, respectively.

The full multi-cell compartment model (Figure 5) has 109 parameters in total divided between the medium (k_{mi}, k_{mo}, k_{om}, k_{cm} and $v_0 = v_e/v_m$) and the different cell types ($k_{oc}^{[j]}$, $k_{cc}^{[j]}$, $k_i^{[j]}$, $k_e^{[j]}$,

$k_{\mathrm{b}}^{[j]}$, $k_{\mathrm{dl}}^{[j]}$, $k_{\mathrm{dh}}^{[j]}$ and $B_{\mathrm{T}}^{[j]}$, $j = 1, \ldots, 13$). Since the pH in the medium during the experiment was 7.2, previously determined [14] values were used for the hydrolysis rate constants k_{om} and k_{cm}. In principle, if the pH of each cell were known then values could also be used for $k_{\mathrm{oc}}^{[j]}$ and $k_{\mathrm{cc}}^{[j]}$ ($j = 1, \ldots, 13$).

The model is described by the following nonlinear system of 43 differential equations:

$$\dot{L}_{\mathrm{m}}(t) = -\left(k_{\mathrm{om}} + k_{\mathrm{mi}}\right)L_{\mathrm{m}}(t) + k_{\mathrm{cm}}H_{\mathrm{m}}(t) + k_{\mathrm{mo}}v_0 L_{\mathrm{e}}(t) \tag{1}$$

$$\dot{H}_{\mathrm{m}}(t) = k_{\mathrm{om}}L_{\mathrm{m}}(t) - \left(k_{\mathrm{cm}} + k_{\mathrm{mi}}\right)H_{\mathrm{m}}(t) + k_{\mathrm{mo}}v_0 H_{\mathrm{e}}(t) \tag{2}$$

$$\dot{L}_{\mathrm{e}}(t) = \frac{k_{\mathrm{mi}}}{v_0}L_{\mathrm{m}}(t) - \left(k_{\mathrm{om}} + k_{\mathrm{mo}} + \sum_{j=1}^{13}k_{\mathrm{i}}^{[j]}\right)L_{\mathrm{e}}(t) + k_{\mathrm{cm}}H_{\mathrm{e}}(t) + \frac{1}{v_1}\sum_{j=1}^{13}k_{\mathrm{e}}^{[j]}L_{\mathrm{c}}^{[j]}(t) \tag{3}$$

$$\dot{H}_{\mathrm{e}}(t) = \frac{k_{\mathrm{mi}}}{v_0}H_{\mathrm{m}}(t) + k_{\mathrm{om}}L_{\mathrm{e}}(t) - \left(k_{\mathrm{cm}} + k_{\mathrm{mo}}\right)H_{\mathrm{e}}(t) \tag{4}$$

$$\dot{L}_{\mathrm{c}}^{[j]}(t) = k_{\mathrm{i}}^{[j]}v_1 L_{\mathrm{e}}(t) - \left(k_{\mathrm{e}}^{[j]} + k_{\mathrm{oc}}^{[j]}\right)L_{\mathrm{c}}^{[j]}(t) + k_{\mathrm{cc}}^{[j]}H_{\mathrm{c}}^{[j]}(t) + k_{\mathrm{dl}}^{[j]}v_2 L_{\mathrm{n}}^{[j]}(t) - k_{\mathrm{b}}^{[j]}\left(B_{\mathrm{T}}^{[j]} - L_{\mathrm{n}}^{[j]}(t)\right)L_{\mathrm{c}}^{[j]} \tag{5}$$

$$\dot{H}_{\mathrm{c}}^{[j]}(t) = k_{\mathrm{oc}}^{[j]}L_{\mathrm{c}}^{[j]}(t) - k_{\mathrm{cc}}^{[j]}H_{\mathrm{c}}^{[j]}(t) + k_{\mathrm{dh}}^{[j]}v_2 L_{\mathrm{n}}^{[j]}(t) \tag{6}$$

$$\dot{L}_{\mathrm{n}}^{[j]}(t) = \frac{k_{\mathrm{b}}^{[j]}}{v_2}\left(B_{\mathrm{T}}^{[j]} - L_{\mathrm{n}}^{[j]}(t)\right)L_{\mathrm{c}}^{[j]}(t) - \left(k_{\mathrm{dl}}^{[j]} + k_{\mathrm{dh}}^{[j]}\right)L_{\mathrm{n}}^{[j]}(t) \tag{7}$$

where the additional superscript ($j = 1, \ldots, 13$) is used to denote cell type. For example, $L_{\mathrm{c}}^{[j]}$ denotes the concentration of the active form of the drug in the cytoplasm of cell type j. The parameters $v_0 = v_{\mathrm{e}}/v_{\mathrm{m}}$, $v_1 = v_{\mathrm{e}}/v_{\mathrm{c}}$ and $v_2 = v_{\mathrm{n}}/v_{\mathrm{c}}$ are the ratios, respectively, of the volume of the extracellular location (v_{e}) to drug medium (v_{m}), extracellular location to cytoplasm (v_{c}) and nucleus (v_{n}) to cytoplasm. The cellular volumes correspond to the total volume for each cell type, where it is assumed, for simplicity, that these volumes are the same across cell types. The total volume, v_{T}, of the medium is equal to the sum of the volumes of the medium and extracellular location,

$$\text{i.e., } v_{\mathrm{T}} = v_{\mathrm{m}} + v_{\mathrm{e}} = \frac{v_{\mathrm{e}}}{v_0} + v_{\mathrm{e}} = \frac{v_{\mathrm{e}}\left(1 + v_0\right)}{v_0}. \tag{8}$$

Since $v_{\mathrm{T}} = 2$ ml (i.e., 2×10^{12} μm^3), Equation (8) may be rearranged to give:

$$v_{\mathrm{e}} = \left(2 \times 10^{12}\right)\left(\frac{v_0}{v_0 + 1}\right). \tag{9}$$

The volume of the cytoplasm, v_c, can be obtained by multiplying the mean volume of cytoplasm per cell (829 μm^3 [13]) by the total number of cells of a particular type:

$$v_{\mathrm{c}} = \left(829 \times 10^5\right)/13 \tag{10}$$

where it is assumed that the cell population is equally divided between the different types. Substituting the expressions from Equations (9) and (10) into $v_1 = v_{\mathrm{e}}/v_{\mathrm{c}}$ gives:

$$v_1 = \left[\left(2 \times 10^{12} \right) \left(\frac{v_0}{v_0 + 1} \right) \right] / v_c = 13 \times \frac{\left(2 \times 10^{12} \right)}{\left(829 \times 10^5 \right)} \frac{v_0}{v_0 + 1}$$

$$= 3.1363 \times 10^5 \times \left(\frac{v_0}{v_0 + 1} \right)$$

(11)

The value of v_1 is therefore calculated in terms of v_0.

Since the drug was administered into the medium at time $t = 0$ with dose D (equal to 10 µM) and was in active form only, the initial conditions for the model are given by:

$$L_m(0) = D(v_0 + 1)$$

(12)

and

$$H_m(0) = L_e(0) = H_e(0) = L_c(0) = H_c(0) = L_n(0) = 0.$$

(13)

As noted above, the full model has 109 parameters, but the values of k_{om} and k_{cm} were known experimentally, so leaving 107 parameters left to be estimated.

3.   Structural Identifiability Analysis

For a compartmental system, it is unlikely that all of the compartments (state variables) within the system model can be observed from a real experiment. Therefore, there is the question of whether the limited number of compartments being observed can provide enough information to uniquely determine all of the unknown model parameters. Depending upon the complexity of the model, and in particular how the parameters enter the system and output equations, a model in which all of the compartments are measured may prove to not uniquely determine all of the unknown parameters. For the structural identifiability approach [22] it can be assumed that an ideal experiment has been performed which provides flawless, smooth and noise free data.

Consider a general postulated parametric model that is expressed in a state variable representation of the form:

$$\dot{x}(t) = f\left(x(t, p), p \right), \ x(0, p) = x_0(p)$$

(14)

with

$$y(t, p) = C(p) x(t, p), \text{ where } C(p) = \begin{bmatrix} c_{11}(p) & \cdots & c_{1n}(p) \\ \vdots & \ddots & \vdots \\ c_{r1}(p) & \cdots & c_{rn}(p) \end{bmatrix}$$

(15)

and the unknown parameters of the model are constituted in a q-dimensional vector, $p = \left(p_1, ..., p_q \right)^T \in \Omega$ (Ω some open subset of $\mathbb{R}^q$). It is assumed that there exists a set $M \subseteq \mathbb{R}^n$ (parameter independent) that contains the trajectories of the n-dimensional state vector, $x(t, p) = \left(x_1(t, p), ..., x_n(t, p) \right)^T$, i.e., $x(t, p) \in M$ for all $t \geq 0$ (the model variables are the $x_i(t, p)$). The vector $y(t, p)$ represents the r-dimensional measurements in the experiment and $C(p)$ is an $r \times n$ matrix.

The following are the basic definitions used to distinguish the different types of parameter and system identifiability; they are consistent with the definitions originally proposed by Walter [31]. These definitions classify the uniqueness of the unknown parameters with respect to the observation function corresponding to the experiment used to collect data for parameter estimation. There are three cases: uniquely (globally) identifiable, locally identifiable and unidentifiable. It should be emphasised that uniqueness of a parameter in terms of identifiability does not guarantee that it will be well-determined from the actual experimental data.

For generic $p \in \Omega$, the parameter p_i (the ith component of p) is said to be uniquely (globally) identifiable if it is uniquely determined by the observation function, $y(t, p)$. That is, given the observed output of the compartmental model, $y(t, p)$, the parameter p_i is globally identifiable if for any $\tilde{p} \in \Omega$ such that

$$y(t, p) = y(t, \tilde{p}), \text{ for all } t \geq 0 \tag{16}$$

then $\tilde{p}_i = p_i$. If all of the parameters, p_i, in p are globally identifiable, then the model is said to be structurally globally identifiable.

For generic $p \in \Omega$, the parameter p_i (the ith component of p) is said to be locally identifiable if it can take on any value from a countable (usually finite) set in the parameter space. That is, there exists an $\varepsilon > 0$ such that for any $\tilde{p} \in \Omega$, if $\| \tilde{p} - p \| < \varepsilon$ and $y(t, p) = y(t, \tilde{p})$ for all $t \geq 0$ then $\tilde{p}_i = p_i$. If all of the parameters, p_i, in p are locally identifiable, and at least one is not globally identifiable, then the model is said to be structurally locally identifiable.

For generic $p \in \Omega$, the parameter p_i (the ith component of p) is said to be unidentifiable if it can take on an infinite (uncountable) number of possible values. That is, given any $\varepsilon > 0$ there exists a $\tilde{p} \in \Omega$, $\tilde{p}_i \neq p_i$, such that $\| \tilde{p} - p \| < \varepsilon$ and $y(t, p) = y(t, \tilde{p})$ for all $t \geq 0$. If any parameter, p_i, in p is unidentifiable, then the model is said to be unidentifiable.

If the model system is unidentifiable, any variation in an unidentifiable parameter could not be attributed specifically to heterogeneity. It is then important to consider the redesign of the experiment, for example by collecting additional data. An alternative approach is to redesign the model, postulating values for some of the unknown parameters in some other way. If neither of these is possible, then consideration should be given to a reparameterisation of the model [32]. Therefore, it is essential to perform an identifiability analysis before parameter estimation.

3.1 The Taylor series approach

For linear systems, there are several techniques available for structural identifiability analysis [33], but there are relatively few techniques available for the corresponding analysis for nonlinear systems – see, for example, [34, 35, 36]. In this paper, the Taylor series approach [37] is used, because it provided a tractable solution for the model with 81 unknown parameters (see Section 3.2).

In this approach, the Taylor series expansion of the observation function $y(t, p)$, that is the vector of model variables to be measured from the real experiment, is considered about some known time point (usually the initial condition $t = 0^+$). The successive derivatives of y, evaluated at the time point, then consist of information about the parameter vector p and are considered to be measurable and unique. This yields a (countable) series of simultaneous equations that are then solved for the parameters. Structural identifiability is then dependent

upon the uniqueness of solutions of this system of equations. Consider the Taylor series expansion of the observation $y_i(t, \boldsymbol{p})$ about the initial time point $t = 0$ with r linear combinations of the model variables measured:

$$y_i(t, \boldsymbol{p}) = y_i(0, \boldsymbol{p}) + y_i^{(1)}(0, \boldsymbol{p})t + y_i^{(2)}(0, \boldsymbol{p})\frac{t^2}{2!} + \cdots + y_i^{(\omega)}(0, \boldsymbol{p})\frac{t^\omega}{\omega!} + \cdots \quad i = 1, ..., r \quad (17)$$

where

$$y_i^{(\omega)}(t, \boldsymbol{p}) = \lim_{\tau \to 0} \frac{d^\omega y_i}{dt^\omega}(\tau, \boldsymbol{p}), \text{ for } i = 1, ..., r \text{ and } \omega = 1, 2, 3, ... \quad (18)$$

Uniqueness of the coefficients in the Taylor series with respect to the observations means that if $\tilde{\boldsymbol{p}} \in \Omega$ is such that $\boldsymbol{y}(t, \boldsymbol{p}) = \boldsymbol{y}(t, \tilde{\boldsymbol{p}})$ for all $t \geq 0$ then

$$y_i(0, \boldsymbol{p}) = y_i(0, \tilde{\boldsymbol{p}}) \text{ and } y_i^{(\omega)}(0, \boldsymbol{p}) = y_i^{(\omega)}(0, \tilde{\boldsymbol{p}}) \qquad i = 1, ..., r; \ \omega = 1, 2, 3, ... \quad (19)$$

It must be noted that Equation (19) provides an infinite list of equations that relate parameter vectors $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$ that give rise to identical outputs. If a finite number of these equations establishes that $\tilde{p}_i = p_i$ in any solution for $\tilde{\boldsymbol{p}}$ then this is sufficient to show that the parameter p_i is globally identifiable. However, for any finite number of equations, if it is found there are solutions for which $\tilde{p}_i \neq p_i$ then it is not necessarily true that the parameter p_i is not globally identifiable. It is often difficult, for nonlinear models, to establish an upper bound for ω in Equation (19) such that the resulting finite number of equations relating parameter vectors $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$ (with identical outputs) yields a necessary and sufficient condition for identifiability. Typically to demonstrate that such a bound, N say, has been found it must be shown that a solution for $\tilde{\boldsymbol{p}}$ of Equation (19) with $\omega \leq N$ is a solution of the full system of equations (i.e., $\omega = 1, 2, ..., \infty$). Margaria *et al.* [38] have provided an upper bound, using differential algebra methods, for systems of the form of (14) and (15) in which the right-hand side $\boldsymbol{f}$ is rational and $r = 1$.

3.2 Structural identifiability analysis of the model

The Taylor series approach is applied to test the structural identifiability of the model with 107 unknown parameters described above. The complex algebraic manipulations necessary when applying the Taylor series approach are handled within MATHEMATICA [39], a software package used to perform the symbolic computation involved in the structural identifiability analysis. The package is used to expand the Taylor series of the observations based upon the model definition, Equations (1)-(7), with the initial conditions given in Equations (12) and (13). It is then used to attempt to solve the resulting nonlinear simultaneous equations in (19) to test for uniqueness of the solution set. The structural identifiability result for this model could not be determined due to the high complexity of the computation involved in the analysis.

A second structural identifiability analysis was performed after reducing the number of unknown parameters to 81 by assuming that all the rate constants $k_{oc}^{[i]}$, $i = 1, ..., 13$ are known and equal; similarly for the rate constants $k_{cc}^{[i]}$, $i = 1, ..., 13$. Partial justification for this was provided by simulations with the cellular parameters at the pH values 7.0, 7.2 and 7.4 (within the feasible range, using values obtained in [14]), in which it was found that the model responses were not sensitive to these different values for the individual rate constants.

For the model with 81 unknown parameters, the Taylor series approach is performed on the model defined in Equations (1)-(7) with initial conditions given in Equations (12) and (13). The unknown parameter vector is given by:

$$\boldsymbol{p} = \left(k_{\text{mi}}, k_{\text{mo}}, k_{\text{i}}^{[1]}, \ldots, k_{\text{i}}^{[13]}, k_{\text{e}}^{[1]}, \ldots, k_{\text{e}}^{[13]}, k_{\text{b}}^{[1]}, \ldots, k_{\text{b}}^{[13]}, k_{\text{dl}}^{[1]}, \ldots, k_{\text{dl}}^{[13]}, k_{\text{dh}}^{[1]}, \ldots, k_{\text{dh}}^{[13]}, B_{\text{T}}^{[1]}, \ldots, B_{\text{T}}^{[13]}, v_0 \right)^{\text{T}}$$

where $\boldsymbol{p}$ is an 81-dimensional vector. The set of parameter vectors, Ω, consists of those vectors $(p_1, \ldots, p_{81})^{\text{T}}$ such that $p_i > 0$ $(1 \le i \le 81)$, given that each of the model parameters has positive value.

The state vector $\boldsymbol{x}(t, \boldsymbol{p})$ is given by

$$\boldsymbol{x}(t, \boldsymbol{p}) = \left(L_{\text{m}}(t, \boldsymbol{p}), H_{\text{m}}(t, \boldsymbol{p}), L_{\text{e}}(t, \boldsymbol{p}), H_{\text{e}}(t, \boldsymbol{p}), \right.$$
$$\left. L_{\text{c}}^{[1]}(t, \boldsymbol{p}), \ldots, L_{\text{c}}^{[13]}(t, \boldsymbol{p}), H_{\text{c}}^{[1]}(t, \boldsymbol{p}), \ldots, H_{\text{c}}^{[13]}(t, \boldsymbol{p}), L_{\text{n}}^{[1]}(t, \boldsymbol{p}), \ldots, L_{\text{n}}^{[13]}(t, \boldsymbol{p}) \right)^{\text{T}}$$

and the set M is taken to be $\mathbb{R}^{43}$. The output function is given by the following

$$\boldsymbol{y}(t, \boldsymbol{p}) = \left(L_{\text{e}} + H_{\text{e}}, L_{\text{c}}^{[1]} + H_{\text{c}}^{[1]}, \ldots, L_{\text{c}}^{[13]} + H_{\text{c}}^{[13]}, L_{\text{n}}^{[1]}, \ldots, L_{\text{n}}^{[13]} \right)^{\text{T}}$$
$$= \left(T_{\text{e}}, T_{\text{c}}^{[1]}, \ldots, T_{\text{c}}^{[13]}, L_{\text{n}}^{[1]}, \ldots, L_{\text{n}}^{[13]} \right)^{\text{T}} \tag{20}$$

so that the output matrix, $C(\boldsymbol{p})$, is given by

$$C(\boldsymbol{p}) = \begin{pmatrix} \boldsymbol{C}_0 & \boldsymbol{0}_{1\times3} & \boldsymbol{0}_{1\times3} & \cdots & \boldsymbol{0}_{1\times3} \\ \boldsymbol{0}_{2\times4} & \boldsymbol{C}_1 & \boldsymbol{0}_{2\times3} & \cdots & \boldsymbol{0}_{2\times3} \\ \boldsymbol{0}_{2\times4} & \boldsymbol{0}_{2\times3} & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \boldsymbol{0}_{2\times4} & \boldsymbol{0}_{2\times3} & \cdots & \cdots & \boldsymbol{C}_{13} \end{pmatrix} \qquad \text{where} \quad \begin{aligned} \boldsymbol{C}_0 &= \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix}, \\[1em] \boldsymbol{C}_j &= \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (j = 1, \ldots, 13) \end{aligned}$$

and $\boldsymbol{0}_{k \times m}$ is the $k \times m$ zero matrix. The initial state of the model is given by

$$\boldsymbol{x}_0(\boldsymbol{p}) = \left(D(1 + v_0), 0, \ldots, 0 \right)^{\text{T}}$$

where the dose, D is known. Let $\tilde{\boldsymbol{p}} \in \Omega$ be any parameter vector

$$\tilde{\boldsymbol{p}} = \left(\tilde{k}_{\text{mi}}, \tilde{k}_{\text{mo}}, \tilde{k}_{\text{i}}^{[1]}, \ldots, \tilde{k}_{\text{i}}^{[13]}, \tilde{k}_{\text{e}}^{[1]}, \ldots, \tilde{k}_{\text{e}}^{[13]}, \tilde{k}_{\text{b}}^{[1]}, \ldots, \tilde{k}_{\text{b}}^{[13]}, \tilde{k}_{\text{dl}}^{[1]}, \ldots, \tilde{k}_{\text{dl}}^{[13]}, \tilde{k}_{\text{dh}}^{[1]}, \ldots, \tilde{k}_{\text{dh}}^{[13]}, \tilde{B}_{\text{T}}^{[1]}, \ldots, \tilde{B}_{\text{T}}^{[13]}, \tilde{v}_0 \right)^{\text{T}}$$

such that $\boldsymbol{y}(t, \boldsymbol{p}) = \boldsymbol{y}(t, \tilde{\boldsymbol{p}})$ for all $t \ge 0$. The relationship between the co-ordinates of $\boldsymbol{p}$ and $\tilde{\boldsymbol{p}}$ is explored by equating successive coefficients of the Taylor series expansions of the components of $\boldsymbol{y}(t, \boldsymbol{p})$ and $\boldsymbol{y}(t, \tilde{\boldsymbol{p}})$. The symbolic computation software MATHEMATICA was employed due to the complexity of the coefficients obtained for the higher derivatives. The analysis of the model is shown below and the corresponding MATHEMATICA code used is found in Appendix A.

The first coefficient obtained when setting $\omega = 0$ in Equation (19), for $i = 1, \ldots, 27$, is equal to zero indicating the absence of information. For the first derivative, that is $\omega = 1$ in Equation (19) ($i = 1, \ldots, 27$), it is seen that

$$\tilde{k}_{\text{mi}} = \frac{k_{\text{mi}}(1 + v_0)\tilde{v}_0}{(1 + \tilde{v}_0)v_0}. \tag{21}$$

Using the relation in Equation (21) with the coefficients obtained in the next derivative ($\omega = 2$ in Equation 19) yields the following relations

$$\tilde{k}_{\mathrm{mo}} = k_{\mathrm{mo}} + \frac{k_{\mathrm{mi}}\left(v_0 - \tilde{v}_0\right)}{\left(1+\tilde{v}_0\right)v_0} + \frac{k_{\mathrm{i}}^{[j]}\left(\tilde{v}_0 - v_0\right)}{\left(1+v_0\right)\tilde{v}_0} \quad \text{and} \quad \tilde{k}_{\mathrm{i}}^{[j]} = \frac{k_{\mathrm{i}}^{[j]}\left(1+\tilde{v}_0\right)v_0}{\left(1+v_0\right)\tilde{v}_0} \quad \text{for } j = 1,\ldots,13. \quad (22)$$

By combining the set of coefficients in the Taylor series expansion ($\omega = 3$ in Equation 19) and the relations given in Equations (21) and (22), yields the following relations

$$\tilde{k}_{\mathrm{b}}^{[j]} = \frac{B_{\mathrm{T}}^{[j]}k_{\mathrm{b}}^{[j]}}{\tilde{B}_{\mathrm{T}}^{[j]}}, \ \tilde{k}_{\mathrm{e}}^{[j]} = k_{\mathrm{e}}^{[j]} \ (\text{for } j = 1,\ldots,13) \text{ and } \tilde{v}_0 = v_0 \quad (23)$$

Therefore, in order for Equation (19) to be satisfied for each $j = 1,\ldots,27$ and $\omega = 1,\ldots,3$, the following relationships must hold true when taking into account all of the relations between the parameters in p and $\tilde{p}$ given in Equations (21-23):

$$\tilde{k}_{\mathrm{mi}} = k_{\mathrm{mi}}, \ \tilde{k}_{\mathrm{mo}} = k_{\mathrm{mo}}, \ \tilde{k}_{\mathrm{i}}^{[j]} = k_{\mathrm{i}}^{[j]}, \ \tilde{k}_{\mathrm{e}}^{[j]} = k_{\mathrm{e}}^{[j]}, \ \tilde{k}_{\mathrm{b}}^{[j]} = \frac{B_{\mathrm{T}}^{[j]}k_{\mathrm{b}}^{[j]}}{\tilde{B}_{\mathrm{T}}^{[j]}} \ (j = 1,\ldots,13) \text{ and } \tilde{v}_0 = v_0. \ (24)$$

Using the relations in Equations (24), together with the fourth derivative terms ($\omega = 4$ in Equation 19), it follows that

$$\tilde{k}_{\mathrm{dl}}^{[j]} = k_{\mathrm{dl}}^{[j]} \text{ and } \tilde{k}_{\mathrm{dh}}^{[j]} = k_{\mathrm{dh}}^{[j]} \ (j = 1,\ldots,13). \quad (25)$$

Equating the relations in Equation (25) with the set of coefficients in the Taylor series expansion for the fifth derivative terms ($\omega = 5$ in Equation 19) and the relations obtained in Equations (24) and (25), the set of coefficients returned is equal to zero. Determining the Taylor series coefficients corresponding to the sixth derivative ($\omega = 6$ in Equation 19) and using the relations in Equations (24) and (25) gives

$$\tilde{B}_{\mathrm{T}}^{[j]} = B_{\mathrm{T}}^{[j]} \ (j = 1,\ldots,13). \quad (26)$$

Therefore, for Equation (19) to be satisfied, it is necessary that

$$\tilde{k}_{\mathrm{i}}^{[j]} = k_{\mathrm{i}}^{[j]}, \ \tilde{k}_{\mathrm{e}}^{[j]} = k_{\mathrm{e}}^{[j]}, \ \tilde{k}_{\mathrm{b}}^{[j]} = k_{\mathrm{b}}^{[j]}, \ \tilde{k}_{\mathrm{dl}}^{[j]} = k_{\mathrm{dl}}^{[j]}, \ \tilde{k}_{\mathrm{dh}}^{[j]} = k_{\mathrm{dh}}^{[j]}, \ \tilde{B}_{\mathrm{T}}^{[j]} = B_{\mathrm{T}}^{[j]} \ (j = 1,\ldots,13),$$
$$\tilde{k}_{\mathrm{mi}} = k_{\mathrm{mi}}, \quad \tilde{k}_{\mathrm{mo}} = k_{\mathrm{mo}}, \quad \text{and} \quad \tilde{v}_0 = v_0, \quad (27)$$

that is, $\tilde{p} = p$. Since this is true for generic $p \in \Omega$, the model with 81 unknown parameters is structurally globally identifiable, which implies that all of the unknown parameters can be uniquely determined by the input-output behaviour corresponding to the experiment performed to collect the data used for parameter estimation.

4.   Parameter estimation

Experimental data used for parameter estimation were collected using TPLSM, as described in [5, 13]. The duration of the experiment was 7.5 minutes with measurements taken every 5 seconds, resulting in a dataset of 27 different observations of 90 time points. The data include the total concentration level of topotecan uptake in the extracellular location (T_{e}), the cytoplasm (T_{c}) and the nucleus (L_{n}) of 13 cells (see Equation 20).

Parameter estimation is performed using the commercial software package, FACSIMILE (MCPA Software, UK), which employs the VA05 routine from the Harwell Subroutine Library (a hybrid method combining features from the Newton-Raphson, Steepest descent and

Marquardt methods). This computer-modelling tool is designed to numerically solve differential equations (using a backward-difference predictor-corrector method), with a particular focus on modelling the kinetics of physical and chemical systems [40]. A particular advantage of this package is the robust numerical integrator that is able to handle stiff systems, that is, systems with widely varying rate constants. The numerical integrator can solve all the ordinary differential equations of the model simultaneously and uses the parameter-fitting option available to fit the simulated output of the model to the experimental data. The result of this optimisation process is the estimation of the model parameters.

Goodness of fit is measured by the total error between the model and the experimental data, and the estimated confidence intervals on the parameter values. The residual sum of squares (RSS), that is, the sum of the square of the error at each time point, is used for the error measure, making the optimisation a least squares problem. To account for different magnitudes and accuracies of measurements, the error is weighted using the average magnitude for each time series:

$$ RSS = \sum_{i=1}^{r} \sum_{j=1}^{n} \left(\frac{y_{\text{obs},i}(j) - y_{\text{sim},i}(t_j)}{\sigma_i} \right)^2 \tag{28} $$

where $y_{\text{sim},i}(t_j)$ is the ith model output at the jth sampling time (t_j); $y_{\text{obs},i}(j)$ is the corresponding experimental data point; and $\sigma_i = e \cdot R_i$ is an estimate for the standard error for the ith output, in which $e = 0.015$ is the estimated overall accuracy of the data and R_i is the range for $y_{\text{sim},i}(t_j)$ [40]. This approach is also referred to as a weighted least squares method as the random error at each sampling time t_j is multiplied by a constant weight, $1/\sigma_i^2$, resulting in normalisation of the residuals if σ_i can be chosen to be the standard deviation for each of the random errors.

Confidence levels give a statistical measure of how well the parameters are defined by the model and the data. Sometimes the data will not determine a parameter value within tight enough bounds and so will be defined as 'not well-determined' (NWD) by the data in a statistical analysis that is performed during the fitting procedure. FACSIMILE works in terms of internal parameters that are the natural logarithms of the given model parameters. Information is also returned on the estimated correlation between the well-determined parameters and the standard deviation of the natural logarithm (SDLN) of each of the well-determined parameters, p^0, which is estimated from the variance-covariance matrix of $p\text{-}p^0$ [40].

4.1. Alternating least squares method

The number of unknown parameters that need to be estimated for the model is relatively large (81 in total). To reach maximum accuracy and minimum numerical error in the process of parameter estimation, a general approach, the alternating least squares method (ALSM), was applied. This approach was first introduced in 1976 by Young, de Leeuw and Takane [41, 42] and since then, has been used widely by different researchers (for example [43]) to obtain optimal parameter estimates.

The alternating least squares method involves estimating a limited number of parameters while the remaining parameters are assumed known and hence kept fixed. For subsequent iterations in the fitting procedure, previously estimated parameters are fixed to allow any remaining parameters to be estimated. This process is repeated until all of the parameters are estimated and a "small" residual sum of squares is obtained. The advantage of the ALSM

approach is that at each stage of the parameter estimation, the optimisation algorithm only requires adjustment of a proper subset of the free parameters. This reduces the computational complexity at each optimisation stage and hence increases the likelihood of convergence and well-determined parameter estimates. The original ALSM, described in [41], involved subdivided subsets of parameters for the alternating estimation process that are not suitable to illustrate the interaction of the cells. Therefore, the procedure of subdividing parameters into different subsets was replaced by the two approaches detailed in the next subsection.

4.2. Cell pairing approaches

All of the unknown parameters within each cell are inter-related so that separation of the unknown parameters into subsets with individual estimation is not appropriate. The natural classification of the unknown parameters in the model into logical subsets corresponding to each cell forms the primary step in the subdivision of parameters. This subdivision comprises 13 subsets of parameters associated with each of the cells (consisting of six parameters each) and one subset of parameters corresponding to the medium (comprising three parameters).

Although parameter estimation using ALSM for these 14 sets of unknown parameters is feasible, the connectivity between the cells and the medium would be neglected. This connectivity results from the medium acting as a reservoir of drug, and also as a bridge for drug flow between cells. To counter this issue, two cell pairing approaches are introduced to retain this inter-connection. Applying the ALSM the parameters corresponding to each pair of cells (or a single cell and the medium) are fitted separately while the remaining unknown parameters are fixed. For the first approach the cells are paired at random. In the second approach cells are paired such that the combined maximum concentration of drug in each pair, observed in the experimental data, is approximately the same for each pair (~24 μM). Both approaches are compared with respect to the RSS after the parameter estimation procedures. After several initial parameter estimation runs were conducted using both approaches independently, the random pairing of cells approach was found to result in smaller RSS value. This approach was then adopted as the main one for the full parameter estimation.

The initial input values for the unknown parameters used in the parameter estimation routine were important. These values allow the determination of the local region of parameter space for the optimisation routines involved in the parameter estimation. These initial input parameter values (given in Table 2 in the next subsection) represent the average behaviour of the model and were estimated in [13] for the one cell model using the averaged experimental data and assuming different cellular pH values of 7.0, 7.2 and 7.4.

4.3. Results

Using FACSIMILE, sets of simulation data for the multi-cell population model are generated. Each set of simulation outputs is compared with the corresponding observed data and the parameter values are iteratively altered until the RSS (Equation 28) is minimised. The estimates obtained for the unknown parameters in the model are shown in Table 1, and these were generated for a cellular pH of 7.0 (giving $k_{oc} = 1.2913 \times 10^{-4}$ s^{-1} and $k_{cc} = 3.1578 \times 10^{-4}$ s^{-1} [13]). Since the package FACSIMILE works in terms of internal parameters that are natural logarithms of the model parameters [40], the SDLN values given in Table 1 correspond to the estimated standard deviation for the natural logarithm of each parameter. The SDLN values are automatically generated within FACSIMILE, and correspond to the fitting of the parameters for the appropriate pair of cells, rather than the complete fit across all cells. However, for each parameter grouping FACSIMILE determines the sensitivity matrix (via finite-differences) with respect to all observation functions, which thus provides estimates for the corresponding columns of the full sensitivity matrix. Since it might be expected that the observation

functions for each cell would be insensitive to parameters from other cells, and each grouping consists of all of the parameters for a pair of (randomly selected) cells, the SDLN values provide an approximate accuracy of each parameter estimate. Those SDLN values corresponding to parameters related to the linking between the cellular subsystems and the medium might be estimated with lower accuracy because of the use of the ALSM.

Cell	k_i (SDLN) (s^{-1})	k_e (SDLN) (s^{-1})	k_b (SDLN) $(s^{-1} \mu M^{-1})$	k_{dl} (SDLN) (s^{-1})	k_{dh} (SDLN) (s^{-1})	B_T (SDLN) (μM)
1	9.6793×10^{-4} (0.01)	6.2390×10^{-3} (0.02)	1.0363×10^{-3} (0.27)	1.6352×10^{-1} (0.39)	3.4443×10^{-7} (NWD)	$8.3271 \times 10^{+1}$ (NWD)
2	1.7991×10^{-3} (0.01)	6.3897×10^{-3} (0.02)	6.9339×10^{-4} (0.27)	1.0082×10^{-1} (0.27)	5.3976×10^{-7} (NWD)	$8.9980 \times 10^{+1}$ (NWD)
3	1.7790×10^{-3} (0.01)	8.8861×10^{-3} (0.02)	5.4223×10^{-4} (0.25)	7.3321×10^{-2} (0.26)	2.3474×10^{-9} (NWD)	$8.8863 \times 10^{+1}$ (NWD)
4	2.9240×10^{-3} (0.02)	2.9843×10^{-2} (0.02)	9.6381×10^{-5} (0.02)	1.4977×10^{-2} (0.03)	2.9111×10^{-5} (NWD)	$9.0000 \times 10^{+1}$ (NWD)
5	1.7511×10^{-3} (0.01)	1.0744×10^{-2} (0.02)	3.8854×10^{-4} (0.12)	5.1612×10^{-2} (0.12)	4.8398×10^{-8} (NWD)	$8.6378 \times 10^{+1}$ (NWD)
6	1.8373×10^{-3} (0.01)	5.3843×10^{-3} (0.03)	2.8155×10^{-4} (0.10)	3.7808×10^{-2} (0.10)	1.3193×10^{-5} (NWD)	$8.9492 \times 10^{+1}$ (NWD)
7	2.0484×10^{-3} (0.02)	6.6008×10^{-3} (0.02)	2.5786×10^{-4} (0.07)	3.4608×10^{-2} (0.08)	1.0840×10^{-8} (NWD)	$8.9068 \times 10^{+1}$ (NWD)
8	1.9419×10^{-3} (0.01)	7.1130×10^{-3} (0.02)	5.0752×10^{-4} (0.14)	6.7302×10^{-2} (0.14)	3.2643×10^{-8} (NWD)	$8.6353 \times 10^{+1}$ (NWD)
9	1.2397×10^{-3} (0.01)	9.9466×10^{-3} (0.02)	8.2157×10^{-4} (0.25)	8.6975×10^{-2} (0.19)	4.5910×10^{-6} (NWD)	$6.6101 \times 10^{+1}$ (NWD)
10	1.3153×10^{-3} (0.01)	7.2792×10^{-3} (0.02)	1.1271×10^{-3} (0.36)	1.4692×10^{-1} (0.33)	5.2075×10^{-7} (NWD)	$8.3445 \times 10^{+1}$ (NWD)
11	1.7472×10^{-3} (0.01)	7.2440×10^{-3} (0.02)	4.5567×10^{-4} (0.13)	6.7669×10^{-2} (0.13)	2.7069×10^{-9} (NWD)	$8.8169 \times 10^{+1}$ (NWD)
12	1.5702×10^{-3} (0.01)	5.0707×10^{-3} (0.03)	3.6026×10^{-4} (0.13)	5.1449×10^{-2} (0.14)	4.5926×10^{-8} (NWD)	$8.9683 \times 10^{+1}$ (NWD)
13	1.8230×10^{-3} (0.01)	6.5954×10^{-3} (0.02)	6.9745×10^{-4} (0.22)	1.0707×10^{-1} (0.23)	4.1422×10^{-9} (NWD)	$8.9687 \times 10^{+1}$ (NWD)

Table 1. Estimates for cell-specific parameters with the estimated standard deviation of the natural logarithm of the parameter (SDLN) for individual run.

During any fitting procedure, FACSIMILE performs a statistical analysis to detect parameters that are not well determined (NWD) by the data. The values for these parameters are then fixed before continuing with the parameter fitting, and treated as unknown in subsequent statistical tests [40]. Of the fitted parameters, k_{dh} and B_T in each cell are NWD by the experimental data, which means that small variations in the estimates do not significantly alter the closeness of the fit. A similar situation was also observed in the parameter estimation for the single-cell model [13], providing further support for the lack of sensitivity of the model output to variation in these parameters.
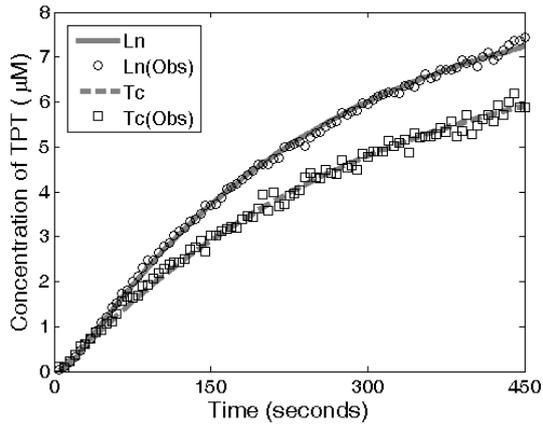
(a) Cell 1
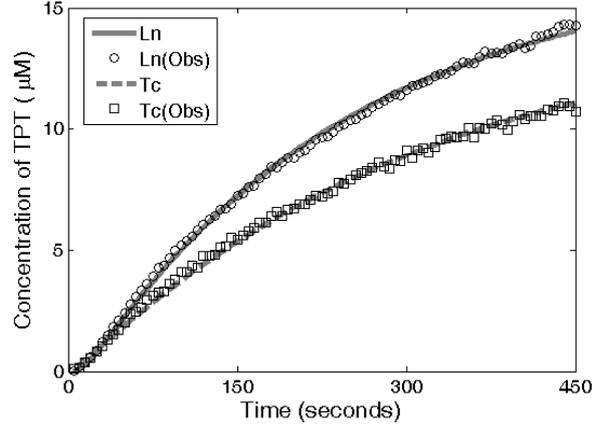
(b) Cell 2
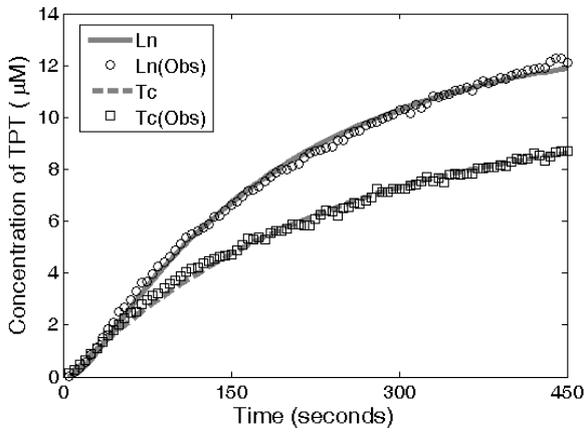
(c) Cell 3
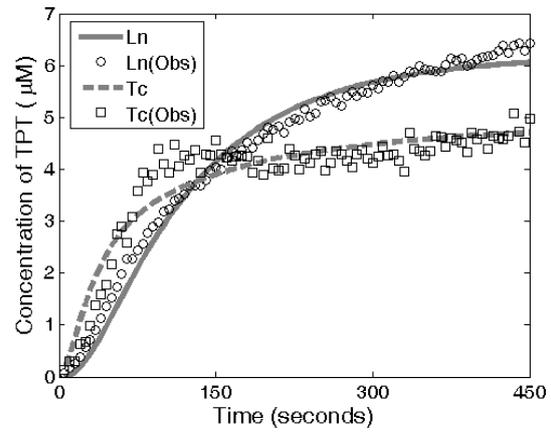
(d) Cell 4
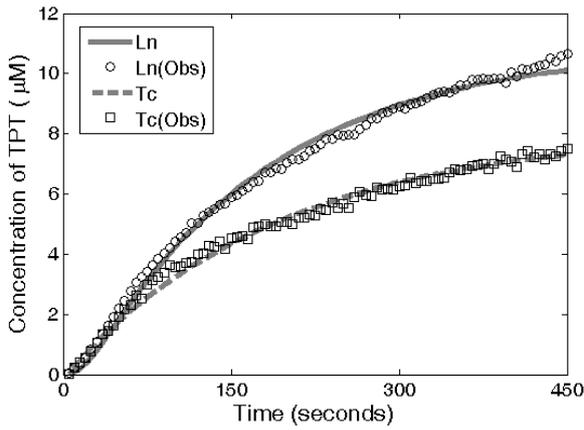
(e) Cell 5
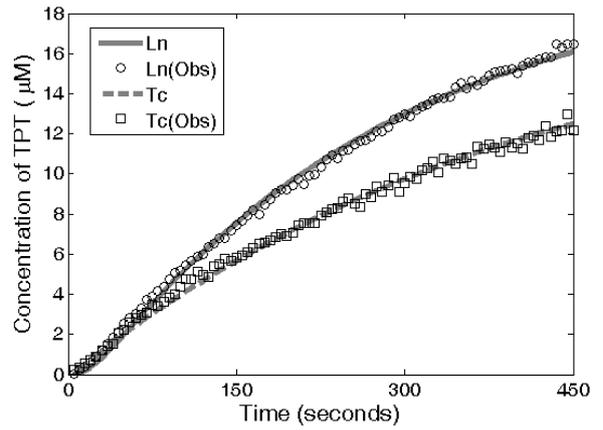
(f) Cell 6

Fig. 6. Comparison of model simulations with experimental data (denoted by 'Obs'). The total concentration of drug in the cytoplasm is denoted by T_c and in the nucleus by L_n. (Cells 1 to 6.)

17
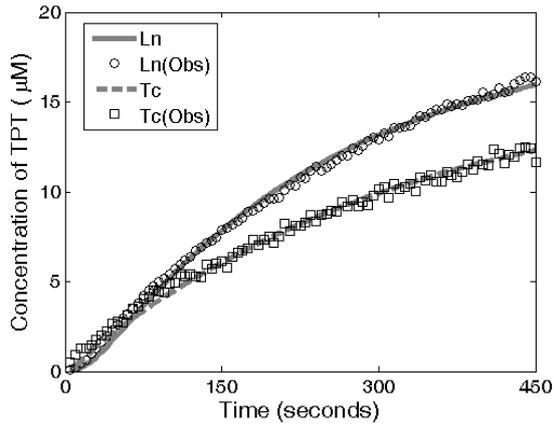
(a) Cell 7
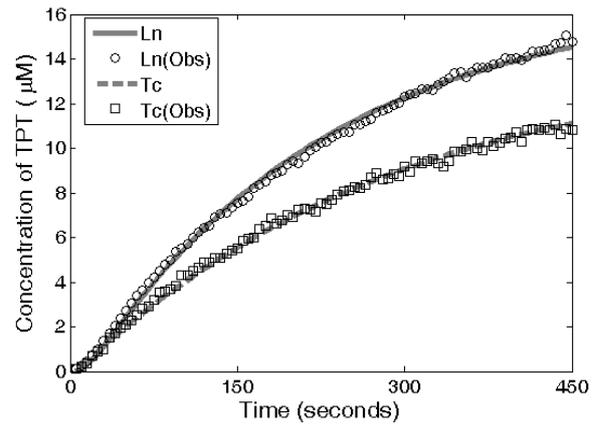
(b) Cell 8
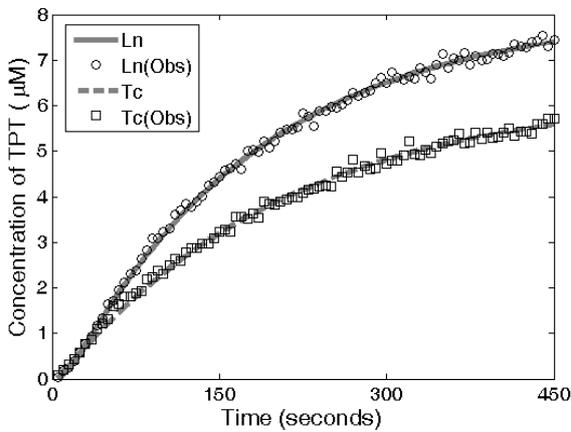
(c) Cell 9
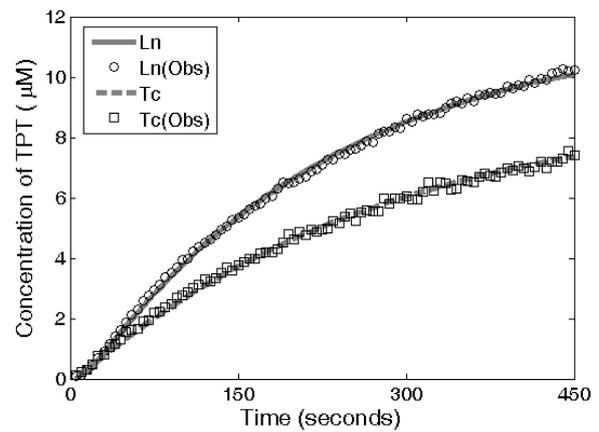
(d) Cell 10
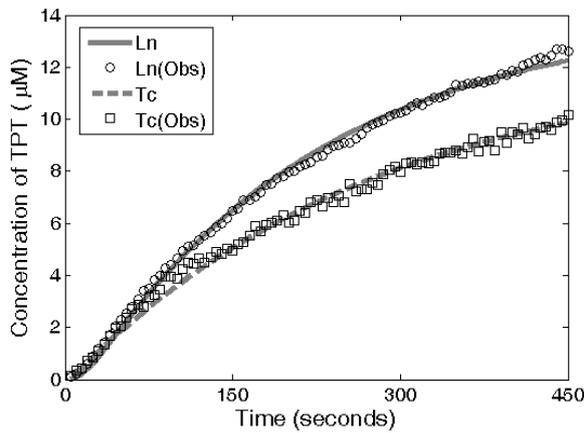
(e) Cell 11
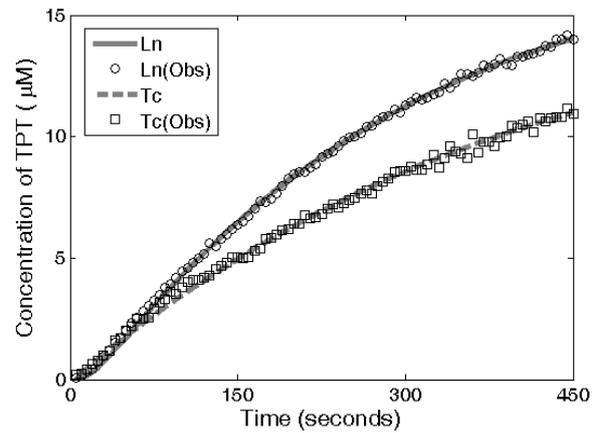
(f) Cell 12

Fig. 7. Comparison of model simulations with experimental data (denoted by 'Obs'). The total concentration of drug in the cytoplasm is denoted by T_c and in the nucleus by L_n. (Cells 7 to 12.)

(a) Cell 13   (b) Extracellular region
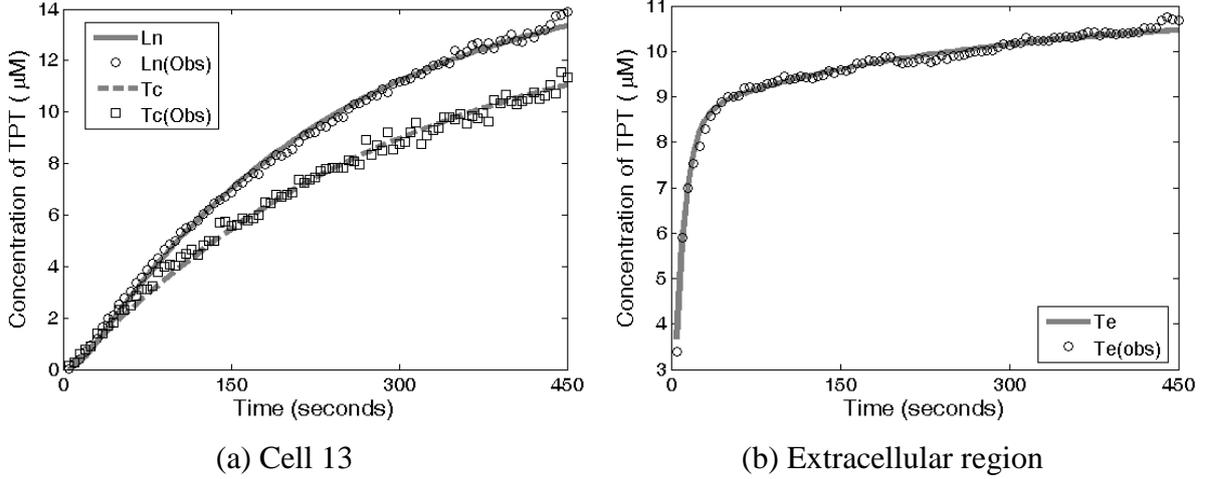
Fig. 8. Comparison of model simulations with experimental data (denoted by 'Obs'). The total concentration of drug in the extracellular region is denoted by T_e, in the cytoplasm by T_c and in the nucleus by L_n. (Cell 13 and Extracellular Region.)

Figures 6–8 show the comparisons between the simulated model outputs (with the fitted parameter values) and the observed data. In the graphs plotted in Figures 6–8 the total concentration of TPT in the extracellular location (T_e), cytoplasm (T_c) and nucleus (L_n) in each of the cell types is shown against time. The starred lines denote the experimental data and the solid lines the simulation results. The figures show very good visual fits for all locations in all cells, except those for the Cell 4 nucleus data. This is possibly as a result of experimental error in the data or Cell 4 being a particularly drug resistant cell. The root-mean-square error (RMSE) provides a quantitative measure of the closeness of fit and is calculated according to the following formula:

$$RMSE = \sqrt{\frac{RSS}{N-q}} = 1.5136 \quad (\mu M) \tag{29}$$

where N is the number of data points and q is the number of fitted parameters such that $N - q$ is the number of degrees of freedom. This implies that the difference between the model prediction and the actual data for the overall time series is approximately 2.3% of the range of the data set:

$$\left| y_{observed,i}(j) - y_{simulated,i}(t_j) \right| = 1.5136 \times \sigma_i = 0.0227 \times R_i. \tag{30}$$

5. Analysis of results

Table 3 shows the parameter estimates obtained in [13] for the single cell-type version of the model, assuming a cellular pH value of 7.0, which were used as the initial estimates in the parameter fitting for the multi-cell type model presented in this paper. Also shown are the mean values for the estimates across the cells in the multi-cell version of the model. In comparing the means for the multi-cell estimates to the single-cell estimates, there is good correspondence for all parameters except the influx rate constant, k_i. There is a significant decrease in value for the mean across the multi-cell estimates. This decrease affects the ratio of the efflux/influx rate constants (E/I ratio), an indication of the drug resistance of a cell; in the single cell model the ratio is 0.36 and for the multi-cell model (see Table 4) the mean E/I

ratio is 5.05 (standard deviation is 2.18) which is significantly higher than the value obtained for the single cell model. However, the single cell model represents all cells, rather than a fraction (1/13) of them as in the multi-cell model. Assuming that the single cell model is representative of an average cell comparable to those constituting the multi-cell model, this would lead to an E/I ratio of 4.64 (13×0.357).

Comparing the E/I ratios for each of the cells, shown in Table 4, it appears that the approach of [13] in considering the population via a single average type, results in a slight underestimation of drug resistance. However, the highest ratio is for Cell 4, for which the poorest visual fit is seen, see Figure 6(d). The relatively poor fit for Cell 4 might account for an artificially high ratio, which in turn would produce an overestimate for the mean. In this case the 'true' E/I ratio might be closer to the single cell value. Although the visual fit is not as good for Cell 4, the estimates obtained for the influx and efflux rate constants do have very low SDLN values suggesting that they are well-determined by the data.

Parameter	Single Cell	Mean for Multi-cell
k_{mi} (s^{-1})	1.3974×10^{-6}	1.4941×10^{-6}
k_{mo} (s^{-1})	8.5551×10^{-2}	9.2632×10^{-2}
v_0	1.5045×10^{-5}	2.2324×10^{-5}
k_i (s^{-1})	2.2110×10^{-2}	1.7869×10^{-3}
k_e (s^{-1})	7.8915×10^{-3}	9.0034×10^{-3}
k_b (s^{-1} μM^{-1})	3.8085×10^{-4}	5.5039×10^{-4}
k_{dl} (s^{-1})	5.5522×10^{-2}	8.0489×10^{-2}
k_{dh} (s^{-1})	1.5639×10^{-7}	4.7793×10^{-7}
B_T (μM)	$8.9972\times10^{+1}$	$8.9440\times10^{+1}$
RMSE	1.0238	1.5136

Table 2. Comparison of the parameter estimates for the single cell model [13] (with cellular pH of 7.0) and the mean values of the individual estimates for the 13 cells.

Cell	1	2	3	4	5	6	7	8	9	10	11	12	13
E/I ratio	6.45	3.55	4.99	10.20	6.14	2.93	3.22	3.66	8.02	5.53	4.15	3.23	3.62

Table 3. The E/I ratios for all cells, calculated using the parameter estimates for the full multi-cell model. The mean value is 5.05, with a standard deviation of 2.18. The corresponding ratio for the single-cell model [13] is 4.64.

A statistical analysis is performed on the estimated parameter values, in order to rank their importance with respect to heterogeneity. The coefficient of variation [44] of each parameter is used to measure the significance of the spread of the estimates, and hence as a measure for the heterogeneity of cancer cells. The coefficient of variation (C_v), of each of the parameters is calculated according to the formula [44]:

$$C_v = \sigma / \overline{x}$$

where σ is the standard deviation and $\overline{x}$ is the mean or expected value of the estimated parameter. The value of C_v provides a relative measure of data dispersion with respect to the mean. It is a dimensionless quantity that can be represented as a percentage or as a simple

decimal value, and is directly proportional to the size of the data scatter; if the data scatter is small then C_v is small, and vice-versa.

Since there is some discrepancy between the model simulations for Cell 4 and the data, particularly for the cytoplasm, the C_v is calculated for each parameter both with and without Cell 4 taken into account. The values of the mean, standard deviation and C_v of the parameters in the model are presented Table 5. The table also shows the rank of the parameters from highest C_v to lowest C_v. The larger the C_v, the higher the rank and the more likely that the parameter is to be a factor of heterogeneity.

In Table 5, although the C_v of each of the model parameters decreases when Cell 4 is excluded from the analysis, there are no major differences in the order of C_v observed. From the ranking of the two sets of C_v values (with or without Cell 4) it can be seen that the rate constant of dissociation of bound TPT as the inactive form, k_{dh}, has the largest C_v, 316.5% or 137.4%; these values represent the highest variation in the parameter estimates across the different cell types. However, the large C_v for the estimates for k_{dh} cannot be accepted as indication of heterogeneity since this parameter is not well determined (NWD) during the parameter estimation in FACSIMILE.

In terms of well-determined parameters, the largest decrease in C_v obtained when neglecting Cell 4 is observed for the cellular efflux rate constant, k_e (from 71.8% to 23.2%). When neglecting the parameter estimate for Cell 4 the coefficient of variation drops from the second largest to the fourth largest. The former rank suggests that efflux is a factor of heterogeneity, while the second does not. If the relatively poor fit for cytoplasmic TPT and the high E/I ratio for Cell 4 are indicators of drug resistance, the higher rank for C_v suggests that this resistance is manifested through the efflux pathway. Since the C_v for the rate constants related to TPT binding, k_b and k_{dl}, are over 40% in both cases (with or without Cell 4), they are thought to be factors of heterogeneity.

Although the parameter relating to the total concentration of potential binding sites, B_T, is also not well determined during parameter fitting, the coefficient of variation is still small (less than 1.5%). This very low value for C_v suggests that the abundance of potential binding sites for nuclear TPT-L is not likely to be a source of heterogeneity. It should also be noted that B_T is constrained to be in the interval 0–90 μM and that this is implemented in FACSIMILE by defining B_T to be a bounded function of a positive dummy parameter that is estimated rather than B_T [13]. This bounded function is quite insensitive to variation in the dummy parameter resulting in B_T being not well determined. All estimates for this parameter were towards the higher end of the constraint interval (mean value of approximately 89 μM).

	All estimates			Estimates without Cell 4		
	Mean ($\bar{x}$)	StDev (σ)	C_v	Mean ($\bar{x}$)	StDev (σ)	C_v
k_i	1.7869×10^{-3}	4.7542×10^{-4}	26.61% (5)	1.6873×10^{-3}	3.2540×10^{-4}	19.28% (5)
k_e	9.0034×10^{-3}	6.4610×10^{-3}	71.76% (2)	7.2680×10^{-3}	1.6830×10^{-3}	23.16% (4)
k_b	5.5039×10^{-4}	2.6568×10^{-4}	48.27% (4)	5.8809×10^{-4}	2.3843×10^{-4}	40.54% (3)
k_{dl}	8.0489×10^{-2}	4.1840×10^{-2}	51.98% (3)	8.5931×10^{-2}	3.8596×10^{-2}	44.92% (2)
k_{dh}	4.7793×10^{-7}	1.5127×10^{-6}	316.52% (1)	5.8924×10^{-8}	8.0947×10^{-8}	137.37% (1)
B_T	$8.9440 \times 10^{+1}$	1.1955	1.34% (6)	$8.9397 \times 10^{+1}$	1.2377	1.38% (6)

Table 4. The mean, standard deviation and C_v for each of the estimated parameters in the model. The rank of the C_v, from highest to lowest, is also indicated in brackets.

Examining the differential equation (7) it can be seen that the affinity of binding [45] of TPT to DNA within a particular cell nucleus is given by

$$K_A^{[j]} = \frac{k_b^{[j]}/v_2}{k_{dl}^{[j]} + k_{dh}^{[j]}}.$$

(31)

The affinity of binding for each of the cells is given in Table 5.

Cell	1	2	3	4	5	6	7	8
K_A	1.55×10^{-2}	1.75×10^{-2}	1.86×10^{-2}	1.64×10^{-2}	1.90×10^{-2}	1.85×10^{-2}	1.88×10^{-2}	1.91×10^{-2}
Cell	**9**	**10**	**11**	**12**	**13**	**Mean**	**StDev**	C_v
K_A	1.68×10^{-2}	1.80×10^{-2}	1.69×10^{-2}	1.78×10^{-2}	1.65×10^{-2}	1.76×10^{-2}	1.14×10^{-3}	6.46%

Table 5. The affinity of binding, K_A (μM^{-1}), calculated using equation (31) for each of the cells. Neglecting Cell 4 the mean K_A is 1.77×10^{-2}, with a standard deviation of 1.13×10^{-3}, giving a coefficient of variation of 6.35%.

An analysis of the propagation of errors in equation (31) is as follows:

$$\left(\frac{\Delta K_A^{[j]}}{K_A^{[j]}} \right)^2 \approx \left(\frac{\Delta k_b^{[j]}}{k_b^{[j]}} \right)^2 + \left(\frac{\left(\Delta k_{dl}^{[j]}\right)^2 + \left(\Delta k_{dh}^{[j]}\right)^2}{\left(k_{dl}^{[j]} + k_{dh}^{[j]}\right)^2} \right)$$

For the means and standard deviations reported in Table 5 it is seen that

$$\left(\frac{\Delta K_A^{[j]}}{K_A^{[j]}} \right)^2 \approx 0.4827^2 + \left(\frac{\left(4.184 \times 10^{-2}\right)^2 + \left(1.5127 \times 10^{-6}\right)^2}{\left(8.0489 \times 10^{-2} + 4.7793 \times 10^{-7}\right)^2} \right) \approx 0.50.$$

This analysis predicts that if the variation in the parameter values is random, the calculated binding affinity, K_A, should have a coefficient of variation of 69%. However, this is not the case, with a consistent estimate of the affinity being obtained across the cells. The affinity of binding is a measure of the interaction of TPT with its binding sites and with such a relatively low coefficient of variation it would not be expected to account for inter-cellular heterogeneity. In contrast, the individual pathways involved in binding and dissociation (parameters k_b, k_{dl} and k_{dh}) are candidates for heterogeneity due to high coefficients of variation being obtained.

In Figure 9 a further investigation is carried out for the identified heterogeneous parameters k_b and k_{dl} by plotting the normalised (with respect to the mean) parameter values against the cell number. A high correlation is observed between these parameters across the cells. To verify this observation, the blocks of the parameter correlation matrix pertaining to parameters k_b and k_{dl} were examined after a fit for Cells 1 to 4 was performed. FACSIMILE reports the correlation matrix after parameter estimation and the relevant blocks are shown in Table 6 for Cells 1-4. Similar such matrices for the other cells are omitted due to space constraints and the fact that the pattern of correlation is very similar. It can be seen that there exists very strong correlation between k_b and k_{dl} for a given cell but there is no significant correlation between these parameters for the different cells shown.
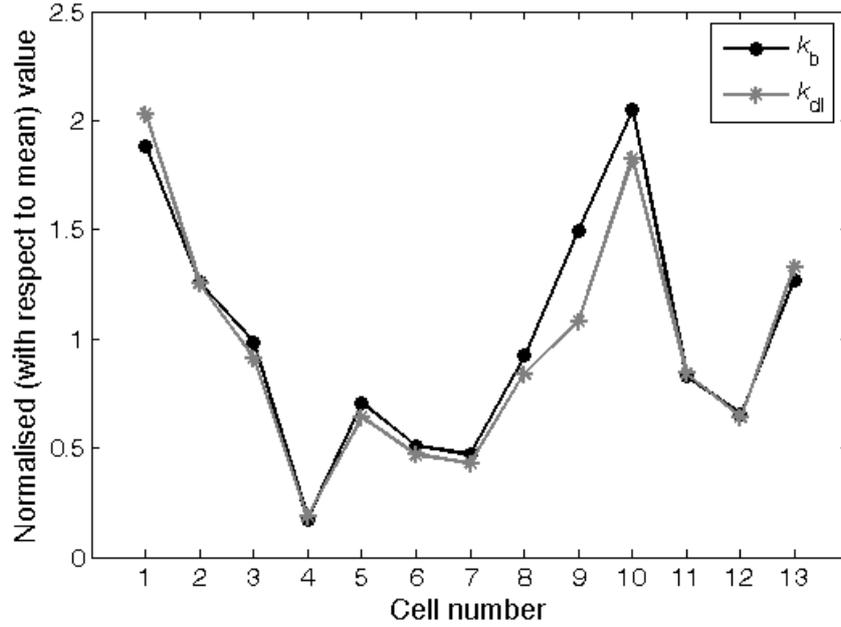
Fig. 9. Correlation between binding and (active) dissociation rate constants.

	$k_{\text{b}}^{[1]}$	$k_{\text{dl}}^{[1]}$	$k_{\text{b}}^{[2]}$	$k_{\text{dl}}^{[2]}$	$k_{\text{b}}^{[3]}$	$k_{\text{dl}}^{[3]}$	$k_{\text{b}}^{[4]}$	$k_{\text{dl}}^{[4]}$
$k_{\text{b}}^{[1]}$	1.000	1.000	0.002	0.002	0.002	0.002	0.002	0.002
$k_{\text{dl}}^{[1]}$	1.000	1.000	0.002	0.002	0.002	0.002	0.002	0.002
$k_{\text{b}}^{[2]}$	0.002	0.002	1.000	1.000	0.003	0.003	0.003	0.003
$k_{\text{dl}}^{[2]}$	0.002	0.002	1.000	1.000	0.003	0.003	0.003	0.003
$k_{\text{b}}^{[3]}$	0.002	0.002	0.003	0.003	1.000	0.999	0.003	0.003
$k_{\text{dl}}^{[3]}$	0.002	0.002	0.003	0.003	0.999	1.000	0.003	0.003
$k_{\text{b}}^{[4]}$	0.002	0.002	0.003	0.003	0.003	0.003	1.000	0.987
$k_{\text{dl}}^{[4]}$	0.002	0.002	0.003	0.003	0.003	0.003	0.987	1.000

Table 6. Estimated correlation matrix for the estimated parameters k_{b} and k_{dl} for Cells 1–4 (the other cells being similar).

6. Conclusions

The implementation of a mathematical multi-cells model which describes the *in vitro* kinetics of the anti-cancer agent topotecan (TPT) following administration into a culture medium containing a population of human breast cancer cells (MCF-7 cell line) has been presented in this paper. This model was then used to consider inter-cellular heterogeneity via the analysis of parameter estimates obtained from experimental data.

The new model is an extension of the homogeneous cell model proposed in [13] using the same underlying set of experimental data obtained from two-photon laser scanning microscopy (TPLSM) for parameter fitting. The previous model does not seek to address inter-cellular differentiation, but rather represents the population response by one average cell type. It also provides a firm fundamental structure for the new highly complex model presented here that consists of 43 compartments and has 109 parameters (2 known and 107

unknown *a priori*). This greater level of complexity is necessary to permit investigation of the heterogeneity in the uptake kinetics of topotecan into the cells.

A formal structural identifiability analysis using the Taylor series approach [37] was performed as a prerequisite to parameter estimation to ensure that the unknown parameters in the model could be uniquely identified by the corresponding input-output behaviour. The model imposed with different numbers of unknown parameters (107 and 81) underwent structural identifiability analyses using the Taylor series approach. The structural identifiability analyses for both models were analogous, though a result for the model with 107 (unknown) parameters could not be obtained. The complex algebraic manipulations necessary in the Taylor series approach were handled by MATHEMATICA, a software package used to perform symbolic computation.

The identifiability analysis for the model with 107 unknown parameters proved intractable due to the high complexity in the symbolic computation involved. Therefore a reduced model (with 81 unknown parameters) was chosen for subsequent use and determined to be structurally globally (uniquely) identifiable. The increased complexity of the system being investigated increased the difficulties for the structural identifiability analysis for both models. This was especially true for the model with 107 unknown parameters, which requires higher computational power to complete the analysis or to perform a reparameterisation [32].

The parameter estimation of the model was carried out using the alternating least squares method (ALSM) in the computer software FACSIMILE. In order to apply the ALSM, the unknown parameters were grouped into subsets that retained the connectivity between cells whilst maintaining a reasonable number of free parameters to estimate in a single FACSIMILE run. This approach permitted a large number of parameters to be estimated, though some information is lost concerning the correlation between parameters in different subsets. A good visual fit was obtained for all cells except Cell 4, which showed some marked deviation from the cytoplasm data (see Figure 6(d)). The ratio of the efflux and influx rate constants (E/I ratio), a measure of drug resistance, is highest for Cell 4. In addition, the estimates for these rate constants have low SDLN values indicating that they are well determined by the data. These points suggest that Cell 4 is a drug resistant cell, and this might account for the relatively poor visual fit. From Table 5 it is seen that the only significant change to the mean estimated values that was observed when neglecting Cell 4 is for the efflux rate constant. Similarly, there is a significant reduction in the coefficient of variation for this parameter when Cell 4 is neglected. These factors suggest that drug resistance is manifest in the cellular efflux pathway.

Through analysis of the coefficient of variation, C_v, for each of the estimated parameters, it was found that the heterogeneity of the cell population is manifested (separately) in the binding and dissociation pathways at the nuclear DNA target. Although there is relatively little variation in the affinity of binding, K_A, across cells ($C_v = 6.46\%$; or 6.35% neglecting Cell 4), there is significant variation for the individual parameters for binding, k_b, and dissociation (as active form), k_{dl}. An analysis of the propagation of errors in estimates for the affinity of binding of TPT to nuclear DNA showed that the variation in K_A is small enough to suggest that a consistent estimate has been obtained across cells. This observation gives further indication that the mathematical model embodies the behaviour of a physical cell system. Finally, a high correlation was observed in the estimates for the two heterogeneous factors, name k_b and k_{dl}, across the cells. This observation may indicate varying drug resistance in the cells.

The model described in this paper is currently in use at the School of Medicine in Cardiff in the following applications:

- Determining the impact of TPT (a highly cell cycle selective agent) on cells, as part of studies on how bolus delivery can induce DNA damage (in discrete time frames) using the hyperphosphoryllation of a histone variant to detect early events in genomic damage stress induction. For this, it is essential to clarify the pharmacokinetic behaviour of a drug at its first stage.

- Providing an assessment of the ability of cells to modulate the delivery of active drug to nuclear targets as part of a rapid method of profiling the function in single cells of multidrug resistance modulators.

- Assessing the 'plasticity' of expression of cellular features, such as cell size, efflux and nuclear-to-cytoplasmic ratio, that can affect drug responses. This forms part of ongoing research on the origins of resistance in small cell lung cancer.

In the future, it is envisaged that the modelling work will provide an important interface in research into new methods of drug tracking, in particular in flow cytometry and cell imaging, and will allow users to extract parameters of interest from the acquired data.

References

[1]   C. Bailly, Topoisomerase I poisons and suppressors as anticancer drugs, Current Medicinal Chemistry 7 (2000) p. 39-58.

[2] F. Zunino, G. Pratesi, Camptothecins in clinical development, Expert Opinion on Investigational Drugs 13 (2004) p. 269–284.

[3] J. O'Leary, F.M. Muggia, Camptothecins: a review of their development and schedules of administration, European Journal of Cancer 34 (1998) p. 1500-1508.

[4] J.C. Wang, DNA topoisomerases, Annual Review of Biochemistry 65 (1996) p. 653-692.

[5] R.J. Errington, S.M. Ammer-Beg, B. Vojnovic, L.H. Patterson, M. Zloh, P.J. Smith, Advanced microscopy solutions for monitoring the kinetics and dynamics of drug-DNA targeting in living cells, Advanced Drug Delivery Reviews 57 (2005) p. 153-167.

[6] D. Yang, J.T. Strode, H.P. Spielmann, A.H.-J. Wang, T.G. Burke, DNA interactions of two clinical camptothecin drugs stabilize their active lactone forms, Journal of Americal Chemical Society 120 (1998) p. 2979-2980.

[7] R.C. Robert Jr., D.W. Kufe, R.E. Pollock, R.R. Weichselbaum, J.F. Holland, III E. Frei, T.S. Gansler, editors, Cancer Medicine, Hamilton (Canada): BC Decker Inc, Fifth edition, 2000.

[8] G.P. Feeney, R.J. Errington, M. Wiltshire, N. Marquez, S.C. Chappell, P.J. Smith, Tracking the cell origins for escape from topotecan action by breast cancer cells, British Journal of Cancer 88 (2003) p. 1310-1317.

[9] F. Zunino, S. Dallavalle, D. Laccabue, G. Beretta, L. Merlini, G. Pratesi, Current status and perspectives in the development of camptothecins, Current Pharmaceutical Design 8 (27) (2002) p. 2505-2520.

[10] M. Taron, C. Plasencia, A. Abad, C. Martin, M. Guillot, Cytotoxic effects of topotecan combined with various active G_2/M-phase anticancer drugs in human tumour-derived cell lines, Investigation of New Drug 18 (2000) p. 139-147.

[11] R. Alexandrova, Tumour heterogeneity, Experimental Pathology and Parasitology 4/6 (2001) p. 57-67.

[12] P. Duesberg, Genetic instability of cancer cells is proportional to their degree of aneuploidy, Cell Biology 95:23 (1998) p. 13692-13697.

[13] N.D. Evans, R.J. Errington, M.J. Chapman, P.J. Smith, M.J. Chappell, K.R. Godfrey, Compartmental modelling of the uptake kinetics of the anti-cancer agent topotecan in human breast cells, International Journal of Adaptive Control and Signal Processing 19 (2005), p. 395- 417.

[14] N.D. Evans, R.J. Errington, M. Shelley, G.P. Feeney, M.J. Chapman, K.R. Godfrey, P.J. Smith, M.J. Chappell, A mathematical model for the *in vitro* kinetics of the anti-cancer agent topotecan, Mathematical Biosciences 189 (2004) p. 185-217.

[15] N.D. Evans, R.J. Errington, M. Shelley, G.P. Feeney, M.J. Chapman, K.R. Godfrey, P.J. Smith, M.J. Chappell, Compartmental modeling to assess stability of topotecan bound to DNA, In: Proceedings of the 5th IFAC Symposium on Modelling and Control in Biomedical Systems, Melbourne, Australia, 21-23 August, 2003, (D. D. Feng and E. R. Carson, Eds) Pergamon, Oxford, p. 403-408.

[16] A. Paradiso, A. Mangia, A. Barletta, S. Fusilli, F. Marzullo, F. Schittulli, M. De Lena, Heterogeneity of intratumour proliferative activity in primary breast cancer: biological and clinical aspects, European Journal of Cancer, 31A (6) (1995) p. 911-916.

[17] M. Nugoli, P. Chuchana, J. Vendrell, B. Orsetti, L. Ursule, C. Nguyen, D. Birnbaum, E.J. Douzery, P. Cohen, C. Theillet, Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications, BMC Cancer, (24 April 2003) 3:13.

[18] T. Litman, M. Brangi, E. Hudson, P. Fetsch, A. Abati, D.D. Ross, K. Miyake, J.H. Resau, S.E. Bates, The multidrug-resistance phenotype associated with over expression of the new ABC half-transporter, MXR (ABCCG2), Journal of Cell Science 113 (Pt 11) (2000) p. 2011-2021.

[19] I. Gryczynski, Z. Gryczynski, J.R. Lakowicz, D.Z. Yang, T.G. Burke, Fluorescence spectral properties of the anticancer drug topotecan by steady-state and frequency domain fluorometry with one-photon and multi-photon excitation, Photochemistry and Photobiology 69 (1999) p. 421-428.

[20] LOCI, Laboratory for Optical and Computational Instrumentation website, multi-photon excitation fluorescence microscopy page, 2004.

[21] Gonda, Neuroscience and genetic research centre, UCLA, Two photon or multiphoton laser-scanning microscopy website, 1997.

[22] R. Bellman, K.J. Åström, On structural identifiability, Mathematical Biosciences 7 (1970) p. 329-339.

[23] L.M. Franks, M.N. Teich, Introduction to the Cellular and Molecular Biology of Cancer, Oxford University Press, Oxford, New York, Tokyo, Third edition, 1997.

[24] K.R. Godfrey, Compartmental Models and their Application, Academic Press, London, New York, 1983.

[25] J.A. Jacquez, Compartmental Analysis in Biology and Medicine, BioMedware, Ann Arbor, Third edition, 1996.

[26] H.P. Rang, M.M. Dale, J.M. Ritter, Pharmacology, Churchill Livingston, Edinburgh, London, New York, Fourth edition, 1999.

[27] K. G. Guderley, C.-C. Hsu, A predictor-corrector method for a certain class of stiff differential equation, Mathematics of Computation 26 (1972) p. 51-69.

[28] J. A. Jacquez, Modelling with Compartments, BioMedware, Ann Arbor, 1999.

[29] I. Chourpa, J. M. Millot, G. D. Sockalingum, J. F. Riou, M. Manfait, Kinetics of lactone hydrolysis in antitumor drugs of camptothecin series as studied by fluorescence spectroscopy, Biochimica et Biophysica Acta 1379 (1998) p. 3353-3366.

[30] M. J. Chappell, G. D. Thomas, K. R. Godfrey, A. R. Bradwell, Optimal tumor targeting by antibodies: Development of a mathematical model, Journal of Pharmacokinetics and Biopharmaceutics 19 (1991) p. 227-260.

[31] E. Walter, Identifiability of State Space Models, Springer Verlag, 1982.

[32] N. D. Evans, M. J. Chappell, Extensions to a procedure for generating locally identifiable reparameterisations of unidentifiable systems, Mathematical Biosciences 168 (2000) p. 137-159.

[33] K. R. Godfrey, J. J. DiStefano III, Identifiability of model parameters. In E. Walter, editor, Identifiability of Parametric Models, Pergamon Press, Oxford, Chapter 1 (1987) p. 1-20.

[34] M. J. Chappell, K. R. Godfrey, S. Vajda, Global identifiability of the parameters of nonlinear systems with specified inputs: A comparison of methods, Mathematical Biosciences 102 (1990) p.41-73.

[35] L. Ljung and T. Glad, On global identifiability for arbitrary parametrizations, Automatica 30 (1994) p. 265-276.

[36] M. P. Saccomani, S. Audoly and L. D'Angio, Parameter identifiability of nonlinear systems: the role of initial conditions, Automatica 39 (2003) p. 619-632.

[37] H. Pohjanpalo, System identifiability based on the power series expansion of the solution, Mathematical Biosciences 41 (1978) p. 21-33.

[38] G. Margaria, E. Riccomagno, M.J. Chappell, H.P. Wynn, Differential algebra methods for the study of the structural identifiability of rational polynomial state-space models in the biosciences, Mathematical Biosciences 174 (2001) p. 1–26.

[39] S. Wolfram, The Mathematica Book (Mathematica version 5.0), Wolfram Media/Cambridge University Press, 2005.

[40] AEA Technology, Facsimile (Unix version) User Guide. Harwell Laboratory: Didcot, Oxfordshire, UK, 1994.

[41] F. W. Young, J. de Leeuw, T. Takane, Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features, Psychometrika 41 (1976) p. 505-529.

[42] Y. Takane, F. W. Young, J. de Leeuw, Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features, Psychometrika 42 (1977) p. 7-65.

[43] M. L. Ochoa, P. B. Harrington, Detection of methamphetamine in the presence of nicotine using in situ chemical derivatization and ion mobility spectrometry, Analytical chemistry 76 (2004) p. 985-991.

[44] B. Iglewicz, Some Properties of the Coefficient of Variation, Virginia Polytechnic Institute, PhD. Thesis (1967).

[45] P. Leff, J. Prentice, H. Giles, G.R. Martin, J. Wood, Estimation of Agonist Affinity and Efficacy by Direct, Operational Model-Fitting, Journal of Pharmacological Methods 23 (1990) p. 225-237.

Appendix A: Structural identifiability analysis within the symbol computation software package MATHEMATICA

The sample code used for the structural identifiability analysis of the model is provided in the following section. It also shows how invaluable the package is when performing a structural identifiability analysis on such a model.

Step One:

Define the model equations in the MATHEMATICA workspace by stating the functions for the first derivatives (with respect to t) of the model variable for the 43 compartments where $j = 1,\dots,13$:

```
Lm'[t_]:= -(kom+kmi) Lm[t]+kcm Hm[t]+kmo v0 Le[t]
Hm'[t_]:= kom Lm[t]-(kcm+kmi) Hm[t]+kmo v0 He[t]
```

$$\text{Le'[t\_]:= } \frac{\text{kmi}}{\text{v0}}\text{ Lm[t]}-\left(\text{kmo}+\text{kom}+\sum_{j=1}^{13}\text{ki[j]}\right)\text{Le[t]}+\text{kcm He[t]}$$

$$+\frac{1}{\text{v1}}\sum_{j=1}^{13}\text{ke[j] Lc[j, t]}$$

$$\text{He'[t\_]:= } \frac{\text{kmi}}{\text{v0}}\text{ Hm[t]}-\text{kmo Le[t]}-\text{(kcm+kmo) He[t]}$$

```
Derivative[0, 1][Lc][j_, t_]:= ki[j] v1 Le[t]+kcc[j] Hc[j, t]
    -(ke[j]+koc[j]) Lc[j, t]+kdl[j] v2 Ln[j, t]
      -kb[j](Bt[j]-Ln[j, t]) Lc[j, t]


Derivative[0, 1][Hc][j_, t_]:= koc[j] Lc[j, t]-kcc[j] Hc[j, t]
                                 +kdl[j] v2 Ln[j, t]
```

$$\text{Derivative[0, 1][Ln][j\_, t\_]:= } \frac{\text{kb[j]}}{\text{v2}}\text{(Bt[j]-Ln[j, t]) Lc[j, t]-}$$
$$\text{(kdl[j]+kdh[j]) Ln[j, t]}$$

The initial state of the system is given by:

```
Lm[0] = d (v0+1); Hm[0]= 0; Le[0]= 0; He[0]= 0;
Lc[j_, 0]:= 0
Hc[j_, 0]:= 0
Ln[j_, 0]:= 0
```

where **d** is the dose and is known. Using two-photon laser scanning microscopy to collect the experimental data results in the following output structure for the model:

```
Te[t_] = Le[t]+He[t]
Tc[j_, t]:= Lc[j, t]+Hc[j, t]
```

In order to perform the identifiability analysis, let $\tilde{p}$ be another parameter vector such that the output $y(t, p)$ is identical to $y(t, \tilde{p})$. The relationship between $\tilde{p}$ and p is explored through equating the coefficients of the Taylor series expansions of these outputs. When the

successive terms in the expansion of $y(t, \boldsymbol{p})$ are obtained, a set of substitution rules will then applied to obtain the corresponding terms in the expansion of $y(t, \tilde{\boldsymbol{p}})$:

```
subst = {kmi → kkmi, kmo → kkmo, kom → kkom, kcm → kkcm,
         koc[j] → kkoc[j], kcc[j] → kkcc[j], ki[j] → kki[j],
         ke[j] → kke[j], kb[j] → kkb[j], kdl[j] → kkdl[j],
         kdh[j] → kkdh[j], Bt[j] → BBt[j], v0 → vv0};
```

In the above expression, which will be used as the substitution rule with the MATHEMATICA replacement operator (`/.`), the parameter $k_b[j]$ in the vector $\boldsymbol{p}$ will be represented by `kb[j]` in MATHEMATICA and the corresponding parameter in the vector $\tilde{\boldsymbol{p}}$ will be denoted by `kkb[j]`.

It is assumed that the extracellular and intracellular hydrolysis rate constants are known *a priori* in the structural identifiability analysis for the reduced model (with 81 unknown parameters). This additional information can be included in the analysis as follows:

```
soln = {kom → kkom, kcm → kkcm, koc[j] → kkoc[j], kcc[j] → kkcc[j]};
```

where $j = 1, \ldots, 13$. In addition, the relation between the parameters v_0 and v_1 has to be taken into account (α is a known value and is equal to 3.1363×10^5):

$$
v1 = \frac{\alpha\, v0}{1 + v0};
$$

Corresponding to Equation (19) in the main text, the first coefficients in the Taylor series expansion of the components of the output ($\omega = 0$ in Equation 19) are zero and so yield no useful information.

The next coefficients are obtained by taking derivatives with respect to t and substituting $t = 0$. For $\omega = 1$ in Equation (19) it is seen that

```
obs = D[y[t], t];
y1 = Simplify[(obs /. t → 0)];
y1bar = Simplify[y1 /. subst];
newSoln = Simplify[Solve[y1[[1]] == y1bar[[1]], {*}]]
```

(where `*` inserts all of the unknown parameters) yields the following output

```
Solve::svar :
Equations may not give solutions for all "solve" variable. More...
```

$$
\left\{ \left\{ kkmi \rightarrow \frac{kmi\,(1+v0)\,vv0}{(1+vv0)\,vv0} \right\} \right\}
$$

The relationship between $\tilde{k}_{mi}$ (`kkmi`) and k_{mi} (`kmi`) can be extracted using the command `newSoln[[1]]`, and this new relationship can be added to the known *a priori* relations in `soln` where $j = 1, \ldots, 13$ using the `Union[]` function:

```
soln = Union[soln, newSoln[[1]]]
```

$$\left\{ \text{kom} \rightarrow \text{kkom, kcm} \rightarrow \text{kkcm, kkmi} \rightarrow \frac{\text{kmi (1+v0) vv0}}{\text{(1+vv0) v0}}, \text{koc[j]} \rightarrow \text{kkoc[j]}, \right.$$
$$\left. \text{kcc[j]} \rightarrow \text{kkcc[j]} \right\}$$

The next coefficient, for $\omega = 2$ in Equation (19), combined with the previous relations yields the following, where ***** is used to denote all of the unknown parameters except $\tilde{k}_{\text{mi}}$:

```
obs = D[obs, t];

y2 = Simplify[obs/. t → 0];

y2bar = Simplify[y2/.subst];

newSoln = Simplify[Solve[y2 == y2bar, {*}]]
```

```
Solve::svar :

  Equations may not give solutions for all "solve" variable. More...
```

$$\left\{ \left\{ \text{kkmo} \rightarrow \text{kmo} + \frac{\text{kmi (v0-vv0)}}{\text{(1+vv0) v0}} + \frac{\text{ki[j](vv0-v0)}}{\text{(1+v0) vv0}} \right\}, \left\{ \text{kki[j]} \rightarrow \frac{\text{ki[j](1+vv0) v0}}{\text{(1+v0) vv0}} \right\} \right\}$$

These new relations can then be used to update the previously obtained relations and can then be added to them using the following:

```
soln = Simplify[soln/. newSoln[[1]]];

soln = Union[soln, newSoln[[1]]];
```

For $\omega = 3$ in Equation (19), the third derivative can be obtained by using code similar that for the first and second derivatives, namely:

```
obs = D[obs, t];

y3 = Simplify[obs/. t → 0];

y3bar = Simplify[(y3/.subst) /. soln];
```

Since the coefficients obtained from this derivative are highly complicated, it is necessary to solve the relationship between $\tilde{p}$ and p for different unknown parameters in three parts. First, the part of the coefficients that contain similar relations is solved, where ***** denotes all the remaining unknown parameters that have not yet been solved:

```
eqns = {y3[[27]]- y3bar[[27]]};

For[i = 1, i < 13, i++,

      eqns = Append[eqns, y3[[27-i]]- y3bar[[27-i]]]

eqns = Map[# == 0 &, eqns];

eqns = Union[eqns, {BBt1 ≠ 0, BBt2 ≠ 0, BBt3 ≠ 0, BBt4 ≠ 0, BBt5 ≠ 0,

      BBt6 ≠ 0, BBt7 ≠ 0, BBt8 ≠ 0, BBt9 ≠ 0, BBt10 ≠ 0, BBt11 ≠ 0,

      BBt12 ≠ 0, BBt13 ≠ 0}];

newSoln = Simplify[Solve[eqns, {*}]]
```

The previous code gives the following new relations where ($j = 1, \ldots, 13$):

```
Solve::svar :
```

```
  Equations may not give solutions for all "solve" variable. More...
```

$$\left\{ \left\{ \text{kkb[j]} \rightarrow \frac{\text{Bt[j] kb[j]}}{\text{BBt[j]}} \right\} \right\}$$

This gives rise to an updated set of relations taking into account the previously obtained relations between $\tilde{p}$ and p :

```
soln = Simplify[soln/. newSoln[[1]]];
soln = Union[soln, newSoln[[1]]];
```

Second, another group of coefficients which will potentially give rise to the relations for $k_{\text{e}}^{[j]}$ is dealt with:

```
eqns = {y3[[2]]-y3bar[[2]]};
For[i = 3, i < 15, i++,
          eqns = Append[eqns, y3[[i]]-y3bar[[i]]]
eqns = Map[# == 0 &, Simplify[eqns /. soln]]];
eqns = Union[eqns, Map[# == 0 &, {*}]];
newSoln = Simplify[Solve[eqns,{*}]]
```

```
Solve::svar :
```

```
  Equations may not give solutions for all "solve" variable. More...
```

$$\left\{ \left\{ \text{kke[j]} \rightarrow \text{ke[j]} \right\} \right\}$$

Finally, the remaining group of terms is solved which gives rise to the following relation for $\tilde{v}_0$:

```
eqns = {y3[[1]]-y3bar[[1]]};
eqns = Map[# == 0 &, Simplify[eqns /. soln]]];
eqns = Union[eqns, Map[# == 0 &, {*}]];
newSoln = Simplify[Solve[eqns,{*}]]
```

```
Solve::svar :
```

```
  Equations may not give solutions for all "solve" variable. More...
```

$$\left\{ \left\{ \text{vv0} \rightarrow \text{v0} \right\} \right\}$$

Using all three new relations obtained for $k_{\text{b}}^{[j]}$, $k_{\text{e}}^{[j]}$ and $\tilde{v}_0$ to update the set of relations obtained for all coefficients examined yields the following:

```
soln = Simplify[soln/. newSoln[[1]]];
soln = Union[soln, newSoln[[1]]]
```

$$\left\{ \text{kom} \rightarrow \text{kkom}, \text{kcm} \rightarrow \text{kkcm}, \text{kkmi} \rightarrow \text{kmi}, \text{kkmo} \rightarrow \text{kmo}, \text{koc[j]} \rightarrow \text{kkoc[j]}, \right.$$

$$\text{kcc[j]} \rightarrow \text{kkcc[j]}, \text{kki[j]} \rightarrow \text{ki[j]}, \text{kke[j]} \rightarrow \text{ke[j]}, \text{kkb[j]} \rightarrow \frac{\text{Bt kb[j]}}{\text{BBt}}$$

$$\left. \text{vv0} \rightarrow \text{v0} \right\}$$

For $\omega = 4$ in Equation (19), the coefficients obtained for this derivative are solved in two parts. First, to obtain a potential relation for $\tilde{k}_{\mathrm{dl}}^{[j]}$ the following code is used:

```
obs = D[obs, t];
y4 = Simplify[obs /. t → 0];
y4bar = Simplify[(y4 /.subst) /. soln];
```

and

```
eqns = {y4[[27]]-y4bar[[27]]};
For[i = 1, i < 13, i++,
          eqns = Append[eqns, y4[[27-i]]-y4bar[[27-i]]]
eqns = Map[# == 0 &,eqns];
eqns = Union[eqns, Map[# == 0 &, {*}]];
newSoln = Simplify[Solve[eqns,{*}]]
```

yielding

```
Solve::svar :
  Equations may not give solutions for all "solve" variable. More...
```

$$\{\{\text{kkdl[j]} \rightarrow \text{-kkdl[j]+kdh[j]+kdl[j]}\}\}$$

These new relations are used to update the previously obtained ones and then added to them using the following:

```
soln = Simplify[soln /. newSoln[[1]]];
soln = Union[soln, newSoln[[1]]];
```

Next, to obtain potential relations for $\tilde{k}_{\mathrm{dh}}^{[j]}$ the following code is run:

```
eqns = {y4[[1]]-y4bar[[1]]};
For[i = 2, i < 15, i++,
                eqns = Append[eqns, y4[[i]]-y4bar[[i]]]
eqns = Map[# == 0 &, Simplify[eqns /. soln]]];
eqns = Union[eqns, Map[# == 0 &, {*}]];
newSoln = Simplify[Solve[eqns,{*}]]
```

which yields the following output

```
Solve::svar :
  Equations may not give solutions for all "solve" variable. More...
```

$$\{\{\text{kkdh[j]} \rightarrow \text{kdh[j]}\}\}$$

The following code uses all three new relations obtained for $\tilde{k}_{\mathrm{dl}}^{[j]}$ and $\tilde{k}_{\mathrm{dh}}^{[j]}$ to update the full set of all relations obtained so far:

```
soln = Simplify[soln /. newSoln[[1]]];
soln = Union[soln, newSoln[[1]]]
```

which yields the following output

$$\Big\{ \text{kom} \to \text{kkom}, \text{kcm} \to \text{kkcm}, \text{kkmi} \to \text{kmi}, \text{kkmo} \to \text{kmo}, \text{koc[j]} \to \text{kkoc[j]},$$

$$\text{kcc[j]} \to \text{kkcc[j]}, \text{kki[j]} \to \text{ki[j]}, \text{kke[j]} \to \text{ke[j]}, \text{kkb[j]} \to \frac{\text{Bt kb[j]}}{\text{BBt}},$$

$$\text{kkdl[j]} \to \text{kdl[j]}, \text{kkdh[j]} \to \text{kdh[j]}, \text{vv0} \to \text{v0} \Big\}$$

For $\omega = 5$ in Equation (19), initially an attempt was made to solve the coefficients but no solution was obtained. The reason for this, as seen in the output from the following code, is that the corresponding coefficients for the two outputs ($y(t, p)$ and $y(t, \tilde{p})$) are identical:

```
obs = D[obs, t];
y5 = Simplify[obs/. t → 0];
y5bar = Simplify[(y5 /.subst) /. soln];
Simplify[y5 - ybar]
```

$$\{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0\}$$

For $\omega = 6$ in Equation (19), the relations for $\tilde{B}_T^{[j]}$ are obtained using the following code:

```
obs = D[obs, t];
y6 = Simplify[obs/. t → 0];
y6bar = Simplify[(y6 /.subst) /. soln];
newSoln = Simplify[Solve[y6 == y6bar, {BBt[j]}]]
```

which yields the following output

$$\big\{\{\text{BBt[j]} \to \text{Bt[j]}\}\big\}$$

Using the new relations obtained for $\tilde{B}_T^{[j]}$ to updated the set of relations obtained so far, the following is obtained which indicates that $\tilde{p} = p$:

```
soln = Simplify[soln/. newSoln[[1]]];
soln = Union[soln, newSoln[[1]]]
```

$$\Big\{ \text{kom} \to \text{kkom}, \text{kcm} \to \text{kkcm}, \text{kkmi} \to \text{kmi}, \text{kkmo} \to \text{kmo}, \text{koc[j]} \to \text{kkoc[j]},$$

$$\text{kcc[j]} \to \text{kkcc[j]}, \text{kki[j]} \to \text{ki[j]}, \text{kke[j]} \to \text{ke[j]}, \text{kkb[j]} \to \text{kb[j]},$$

$$\text{kkdl[j]} \to \text{kdl[j]}, \text{kkdh[j]} \to \text{kdh[j]}, \text{BBt[j]} \to \text{Bt[j]}, \text{vv0} \to \text{v0} \Big\}$$

This final set shows that $\tilde{p} = p$ if $y(t, \tilde{p}) = y(t, p)$ for all $t \geq 0$ and hence the model is structurally globally (uniquely) identifiable since MATHEMATICA works in terms of generic parameters (for p) [39].