A simplified model of glycoprotein production within cell culture

Anna B Lambert, Frank T Smith, Ajoy Velayudhan

Complex biological products, such as those used to treat various forms of cancer, are typically produced by mammalian cells in bioreactors. The most important class of such biological medicines is proteins. These proteins typically bind to sugars (glycans) in a process known as glycosylation, creating glycoproteins, which are more stable and effective medicines. The glycans are large polymers that are formed by a long sequence of enzyme catalysed reactions. This sequence is not always completed, thus leading to a heterogeneous glycoprotein distribution. A better comprehension of this distribution could lead to more efficient production of high quality drugs.

To understand how the manufacturing process can affect the extent of glycosylation of protein, a nonlinear ODE model of glycoprotein production is developed which describes the bioreactor configuration as well as the protein production and glycosylation reactions within the cell. The entire system evolves eventually to a stable steady state. The earlier evolution itself is critical however, as the amount of product produced and its quality varies over time. The model is considered as two coupled systems: the bioreactor submodel and the glycosylation submodel. To investigate the early time evolution within the bioreactor submodel, analytical and numerical properties are derived using matched asymptotic expansions and a finite difference scheme for a range of initial conditions. This leads to qualitatively different regimes for aglycosylated protein production, which affect the glycosylation submodel. The discrete glycoprotein distribution is approximated as continuous and written as a first order PDE, with good agreement between the discrete and continuous models. The PDE is found to admit shocks, but the existence of these shocks is dependent on the early time evolution within the bioreactor submodel and leads to higher levels of glycosylation at early time. This suggests that changing the bioreactor configuration can lead to higher quality product at certain points in time.

1 Introduction

While historically the majority of medicines were small molecules, a rapidly increasing number of new drugs developed by the pharmaceutical industry are protein therapeutics [8]. Proteins are complex macromolecules which are involved in almost all processes within cells. Their complexity and specificity of action means that they are able to produce therapeutic effects that could not be replicated by small molecule drugs. Protein therapeutics have been used for a number of reasons: to aid drug delivery, as vaccines, and to treat diseases including cancer, genetic diseases and diseases of the immune system [17]. Additionally there are financial incentives for companies to develop protein therapeutics

since they typically have a shorter clinical development and approval time than small molecule drugs, and companies are able to acquire comprehensive patents [19].

The majority of these therapeutic proteins are produced using recombinant DNA technology. This entails the alteration of the genetic material of a host organism, (typically the bacterium *Escherichia coli*, yeast, or mammalian cells) to insert the gene coding for the protein of interest. The host organism then produces the recombinant protein alongside the native proteins necessary for usual function. These cells are produced outside of their natural environment in a bioreactor, where conditions such as temperature, pH and cell substrate levels are carefully monitored.

However, after the protein is synthesised by the host cell, the molecule is typically modified in a number of ways before being excreted by the cell. One of the most common of these post-translational modifications is glycosylation - the process by which carbohydrate molecules are attached to the protein. There are two kinds of glycosylation, the N-linking of sugars to asparagine residues, and O-linking of sugars to serine or threonine residues. The kind and extent of glycosylation seems to influence therapeutic efficacy as well as half-life in the body [10; 20]. It is therefore important to understand how to produce protein-based therapeutics with appropriate glycosylation [11]. Unfortunately, the biochemistry of glycosylation is extremely complex , involving many enzymes, various locations in the Golgi apparatus and endoplasmic reticulum, and a large number of product variants [3; 25].

Attempts have been made to model glycosylation as a complex network [14], leading to elaborate models. Engineering models have embedded glycosylation models into bioreactors [12; 16; 26], and have substantially improved our understanding of the behaviour of glycosylation in bioreactors. However, such models contain a large number of parameters, many of which cannot be measured directly. In this paper we take a different approach. Using a very simple kinetic model for the stepwise addition of one kind of sugar to a product, we analyse product formation in terms of quality and quantity in a chemostat. This gives rise to a set of coupled, nonlinear ordinary differential equations, which are analysed in detail to explore the influence of a wide range of initial conditions giving significant variation in quality and quantity distinct from that in steady state. The focus is on allowing a large number of such sugar additions; we also show that passing to the limit of a continuous mixture, a tractable partial differential equation (PDE) results. This PDE is hyperbolic, and admits of the formation of a shock. The condition for shock formation is established explicitly, and it is shown that selecting initial conditions that generate a shock results in improved product quality and quantity, relative to the product obtained at steady state. This unusual result suggests that the trade-off between quantity and quality is more subtle than previously believed. In the discussion, some suggestions are presented on how to extend the present analysis to more complex models of glycosylation.

2 Model Development

Our model considers the interaction between cell mass and substrate in a chemostat: where substrate is fed into the reactor at a constant rate, and a fraction of the bioreactor culture is washed out continuously. The reactor is stirred and therefore spatially homogeneous. All other important parameters, such as temperature or pH are held constant. The specific

growth rate of the cells increases with substrate concentration, following Monod kinetics. The rate of production of aglycosylated protein per cell again increases with substrate concentration according to Michaelis-Menten kinetics.

We then consider glycosylation as the enzyme catalysed stepwise addition of sugars to the protein. The concentration of other reactants necessary for glycosylation is assumed to not be a limiting factor and hence these reactions follow Michaelis-Menten kinetics. This process leads to differing glycoprotein variants, which are denoted p_j^*, where j is the number of sugars attached to the protein molecule and the asterisk denotes the dimensional quantity. The maximum number of sugars that can be attached is n, and so the final variant is p_n^*.

In the following the bioreactor parameters are the dilution rate D^* and the concentration of the substrate feed s_f^*. For the cell mass, r_x^* is the maximal specific growth rate, K_x^* is the half-saturation constant, k_d^* is the cell death rate and Y_x^* is the yield coefficient. The parameters concerning protein production are the maximal production rate r_0^*, the half saturation constant K_0^* and the yield coefficient Y_0^*. The reaction that leads to the formation of the glycoprotein variant p_j^*, where $0 \leq j \leq n$, has the following parameters: r_j^* is the maximal reaction rate, K_j^* the half-saturation constant and Y_j^* is the yield coefficient.

Thus the full dimensional model is as follows. It is first to solve for the cell mass x^*, the substrate s^* and the aglycosylated protein p_0 using

$$\frac{dx^*}{dt^*} = \frac{r_x^* s^* x^*}{K_x^* + s^*} - (D^* + k_d^*)x^*, \tag{2.1\,a}$$

$$\frac{ds^*}{dt^*} = D^*(s_f^* - s^*) - \frac{1}{Y_x^*}\left(\frac{r_x^* s^* x^*}{K_x^* + s^*}\right) - \frac{1}{Y_0^*}\left(\frac{r_0^* s^* x^*}{K_0^* + s^*}\right), \tag{2.1\,b}$$

$$\frac{dp_0^*}{dt^*} = \frac{r_0^* s^* x^*}{K_0^* + s^*} - \frac{1}{Y_1^*}\left(\frac{r_1^* p_0^*}{K_1^* + p_0^*}\right) - D^* p_0^*, \tag{2.1\,c}$$

and then for $1 \leq j \leq n$ to solve for the intermediate glycoprotein variants p_j^* as well as the final glycoprotein variant p_n^* using

$$\frac{dp_j^*}{dt^*} = \frac{r_j^* p_{j-1}^*}{K_j^* + p_{j-1}^*} - \frac{1}{Y_{j+1}^*}\left(\frac{r_{j+1}^* p_j^*}{K_{j+1}^* + p_j^*}\right) - D^* p_j^*, \tag{2.1\,d}$$

$$\frac{dp_n^*}{dt^*} = \frac{r_n^* p_{n-1}^*}{K_n^* + p_{n-1}^*} - D^* p_n^*. \tag{2.1\,e}$$

The initial conditions are

$$x^*(0) = x_0^*, \tag{2.2\,a}$$

$$s^*(0) = s_0^*, \tag{2.2\,b}$$

$$p_j^*(0) = 0 \text{ for all } j, \tag{2.2\,c}$$

where x_0^* and s_0^* are given constants from the operating conditions of the bioreactor. To reduce the considerable number of parameters in the system, non-dimensionalisation is now performed by setting

$$(t^*, x^*, s^*, p_j^*) = (\frac{t}{D^*}, K_x^* Y_x^* x, K_x^* s, K_{j+1}^* p_j). \tag{2.3}$$

While most of the values are intrinsic parameters of the system, we introduce K_{n+1}^* as a half-saturation constant for a theoretical reaction where the substrate is p_n^*, and K_{n+1}^* is used as a typical value for p_n^*. (It would be expected to be in line with K_n^*.) We also set

$$r_x = \frac{r_x^*}{D^*}, \tag{2.4 a}$$

$$k_d = \frac{k_d^*}{D^*}, \tag{2.4 b}$$

$$s_f = \frac{s_f^*}{K_x^*}, \tag{2.4 c}$$

$$r_0 = \frac{r_0^* K_x^* Y_x^*}{K_1 D^*}, \tag{2.4 d}$$

$$K_0 = \frac{K_0^*}{K_x^*}, \tag{2.4 e}$$

$$Y_0 = \frac{K_x^* Y_0^*}{K_1^*}, \tag{2.4 f}$$

$$r_j = \frac{r_j^*}{K_{j+1}^* D^*} \text{ for all } j >= 1, \tag{2.4 g}$$

$$Y_j = \frac{K_j^* Y_j^*}{K_{j+1}^*} \text{ for all } j >= 1. \tag{2.4 h}$$

The initial values $x(0) = x_0$ and $s(0) = s_0$ are thus as follows from (2.2),

$$s_0 = \frac{s_0^*}{K_x^*}, \tag{2.5 a}$$

$$x_0 = \frac{x_0^*}{K_x^* Y_x^*}. \tag{2.5 b}$$

From (2.1) we now have our non-dimensionalised model:

$$\frac{dx}{dt} = \frac{r_x s x}{1+s} - (1 + k_d)x, \tag{2.6 a}$$

$$\frac{ds}{dt} = s_f - s - \frac{r_x s x}{1+s} - \frac{1}{Y_0}\left(\frac{r_0 s x}{K_0 + s}\right), \tag{2.6 b}$$

$$\frac{dp_0}{dt} = \frac{r_0 s x}{K_0 + s} - \frac{1}{Y_1}\frac{r_1 p_0}{1 + p_0} - p_0, \tag{2.6 c}$$

and for $1 \le j \le n - 1$

$$\frac{dp_j}{dt} = \frac{r_j p_{j-1}}{1 + p_{j-1}} - \frac{1}{Y_{j+1}}\frac{r_{j+1} p_j}{1 + p_j} - p_j, \tag{2.6 d}$$

$$\frac{dp_n}{dt} = \frac{r_n p_{n-1}}{1 + p_{n-1}} - p_n. \tag{2.6 e}$$

The values of the parameters in our system will now be considered, with the values used throughout the paper detailed in Table 1. We have relative freedom for the initial conditions in our bioreactor s_0^* and x_0^*, and these will be varied to characterise different qualitative behaviour within the bioreactor. The experimenter also has choice when it comes to setting s_f^* and D^*. The cell growth and product production parameters r_x^*,

Table 1. *The dimensionless parameter values used for all model runs, unless indicated otherwise.*

Parameter	Value
n	20
r_x	2
k_d	0.1
s_f	10
r_0	1
K_0	1
Y_0	0.2
r_j	1
Y_j	1

K_x^*, Y_x^*, r_0^*, K_0^* and Y_0^* vary widely between cell lines. Generally k_d^* in (2.4 b) is a small number, as cell death is slow in comparison to other cellular processes when the reactor is operating at ideal conditions [24].

The number of glycoprotein variants, here denoted as n, considered in previous models varies from 33 to 5685, depending on their level of detail [12; 16; 26]. Here we set $n = 20$, but the results are qualitatively similar for other values of n. The stoichiometry of each glycosylation reaction sets $Y_j^* = 1$. Determining the parameter values for the glycosylation reactions is a complex task [27]. Therefore for most of the paper we will assume that all glycosylation reactions are catalysed by the same enzyme, and thus r_j^* and K_j^* are constant for $j > 0$. Whilst there are multiple enzymes involved in the full reaction network, each enzyme catalyses many reactions [12], and thus this a reasonable starting point to understand this effect of each enzyme.

Since this model is generic and does not aim to produce results for a particular cell line, a set of representative dimensionless parameters each of order 1 is used, as illustrated in table 1. However, the advantage of the present analytical approach is that the effect of parameter changes can be deduced without running a large parameter sweep.

The whole model can be divided quite naturally into two coupled submodels - the cell-substrate interaction consisting of x, s and p_0, and the glycosylation system consisting of p_j with $0 \leq j \leq n$. The behaviour at finite time t turns out to be important throughout, as the extent of the glycosylation, and thus the quality of the product, varies in time.

3 Cell substrate interaction

In this section of the paper, we consider the submodel for the bioreactor, to determine the production of unglycosylated protein. This consists of the governing equations for the cell mass x (2.6 a), substrate s (2.6 b) and unglycosylated protein p_0 (2.6 c).

Part of the qualitative behaviour of the system will be examined, by determining the existence and properties of the steady states. The time evolution of the system is then examined first by solving the equations numerically. Second, by considering the relative initial conditions for cell mass and substrate to be asymptotically large or small, analytical solutions for whole time evolution are found.

3.1 Steady states and stability

There are two possible steady states for this system. The first is the trivial zero cell case, where $x = 0$, $s = s_f$ and $p_0 = 0$. If there are no cells, due to complete cell death or a zero initial condition, no protein can be produced and thus the system consists only of the substrate entering the reactor. If certain constraints on parameter values hold, there is a second steady state as follows:

$$s_e = \frac{1 + k_d}{r_x - (1 + k_d)}, \tag{3.1}$$

$$x_e = \frac{s_f - s_e}{r_x \frac{s_e}{1 + s_e} + \frac{r_0}{Y_0} \frac{s_e}{K_0 + s_e}}, \tag{3.2}$$

$$p_{0,e} = \frac{1}{2} \left[\frac{r_0 s_e x_e}{K_0 + s_e} - \frac{r_1}{Y_1} - 1 + \sqrt{\left(\frac{r_0 s_e x_e}{K_0 + s_e} - \frac{r_1}{Y_1} - 1 \right)^2 + 4 \frac{r_0 s_e x_e}{K_0 + s_e}} \right]. \tag{3.3}$$

Since all variables are non-negative, the other quadratic root for $p_{0,e}$ is neglected. For the same reason, this steady state will exist only if the conditions $r_x > 1 + k_d$, $s_f > s_e$ and $\frac{r_0 s_e x_e}{K_0 + s_e} > \frac{r_1}{Y_1} + 1$ hold. It is well established that this steady state is stable if it exists since the reactor is a chemostat [24]. The zero cell steady state is undesirable as no protein is produced, so from this point we assume that the conditions on the parameters hold and therefore the system has a unique nontrivial stable steady state to which it can evolve.

3.2 Time evolution

The equations were solved numerically using a finite difference scheme, to investigate the time evolution of the system. In addition to the equation parameters, the initial conditions for the cell mass, x_0 and substrate, s_0 have large effects on the early time behaviour of the system. This is illustrated in figure 1: the phase plane for the cell mass x and the substrate s on a log scale, with varying initial conditions for both x and s. Linear stability is visible for all trajectories close to the steady states, but there are clearly several regimes of qualitatively different behaviour, depending on the initial conditions. The three principal regimes are highlighted, and the full time-dependent numerical solutions for x, s and p_0 are given for regime A in figure 2(a), for regime B in figure 3 and for regime C in figure 4. In the next section these are explained and analytical solutions are found using matched asymptotic expansions. This asymptotic analysis allow us to clearly understand the most important processes within our system at each time scale, and thus clearly differentiate between each regime in figure 1. Similar methods have been used extensively when modelling biological systems [4; 9; 18]

We consider a typical initial-value parameter $\epsilon \ll 1$ and then explore the major details of the evolution where at least one of the initial conditions x_0 and s_0 is either $O(\epsilon)$ or $O(\frac{1}{\epsilon})$. Since these initial conditions affect the protein production at early time, and thus the extent of glycosylation, changing these is found to lead to differing quality levels thoroughout time.
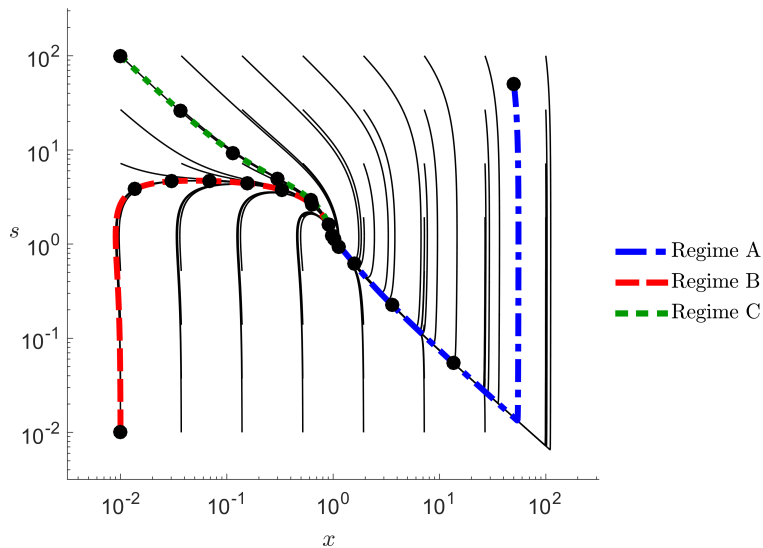
FIGURE 1. The phase plane for the cell mass x and the substrate s trajectories, displayed on a log scale. Three trajectories are highlighted as representative of three different regimes. The black dots are plotted at evenly spaced time intervals to illustrate the progress through time.

3.3 Main examples of evolution

The three principal regimes A, B and C are explored in the following subsections. Table 1 gives parameter values used throughout. These are supplemented by analysis which exposes interesting multi-scaling present for t of $O(1)$, cf [21–23] in other fields, and provides fresh insight and clarity as to how early time behaviour can affect quantity and quality.

3.3.1 *Regime A*

First, the behaviour of solutions is explored when the initial values s_0 and x_0 are order $\frac{1}{\epsilon}$. This is represented as regime A in the top right hand corner of figure 1, where the trajectories move downwards, initially curving to the right, before turning a sharp corner to move diagonally towards the steady state. The full numerical solutions for x, s and p_0 are illustrated in figure 2(a). While the solution for s seems to turn a sharp corner, the short time scales expanded upon in this section fully explain this behaviour. For these initial conditions, the denominators appearing in (2.6 a - 2.6 c) simplify and the effects of the s_f and r_1 contributions become secondary; so the governing equations of the model (2.6) become linearised. The main response arises first on the $O(1)$ time scale where $t = \bar{t}$

with the expansions

$$x = \frac{1}{\epsilon}\bar{x} + \cdots, \tag{3.4 a}$$

$$s = \frac{1}{\epsilon}\bar{s} + \cdots, \tag{3.4 b}$$

$$p_0 = \frac{1}{\epsilon}\bar{p}_0 + \cdots, \tag{3.4 c}$$

holding then. Here the size of p_0 is dictated by the protein production term involving sx over the $O(1)$ time scale; earlier p_0 is smaller of course due to the initial condition. At leading order, the governing equation for the cell mass x (2.6 a) thus becomes

$$\frac{d\bar{x}}{dt} = (r_x - 1 - k_d)\bar{x}. \tag{3.5}$$

Hence x is independent of s initially, as substrate is not limiting cell growth. As described in section 3.1, for the non-trivial steady state to exist, $r_x > 1 + k_d$. Hence $\bar{x}$ grows exponentially and has the form

$$\bar{x} = \bar{x}_0 e^{(r_x - 1 - k_d)t}. \tag{3.6}$$

The substrate $\bar{s}$ is governed by

$$\frac{d\bar{s}}{dt} = -\left(r_x + \frac{r_0}{Y_0}\right)\bar{x} - \bar{s}, \tag{3.7}$$

and hence the solution for $\bar{s}$ is a sum of exponential decay and negative exponential growth.

$$\bar{s} = (\bar{s}_0 + \gamma\bar{x}_0)e^{-t} - \gamma\bar{x}_0 e^{(r_x - 1 - k_d)t} \tag{3.8}$$

where $\gamma = \frac{r_x + \frac{r_0}{Y_0}}{r_x - k_d}$. The substrate is therefore rapidly depleted due to the large cell mass. The governing equation for the unglycosylated protein $\bar{p}_0$ is

$$\frac{d\bar{p}_0}{dt} = r_0\bar{x} - \bar{p}_0 \tag{3.9}$$

which can be solved to give

$$\bar{p}_0 = \frac{r_0}{r_x - k_d}\left(e^{(r_x - 1 - k_d)t} - e^{-t}\right). \tag{3.10}$$

Hence the amount of protein in the bioreactor increases rapidly initially because there is a large quantity of cell mass and the substrate does not limit protein production.

These analytical solutions are compared to the numerical solutions in figure 2(b) and show good agreement initially, with the solutions diverging as the initial assumptions (3.4) fail. The solution for $\bar{s}$ is the leading factor for the expansions ceasing to hold, as it decreases to zero at a finite time $t = t_0$ (contrasting with $\bar{x}$ and $\bar{p}_0$ which remain $O(1)$ then). This leads to a new expansion, beginning around the time t_0 when $\bar{s}$ in effect decreases to $O(\epsilon)$, so s becomes $O(1)$ and the influence of the denominators in (2.6 a - 2.6 c) comes into play as the substrate begins to limit cell growth and protein production.
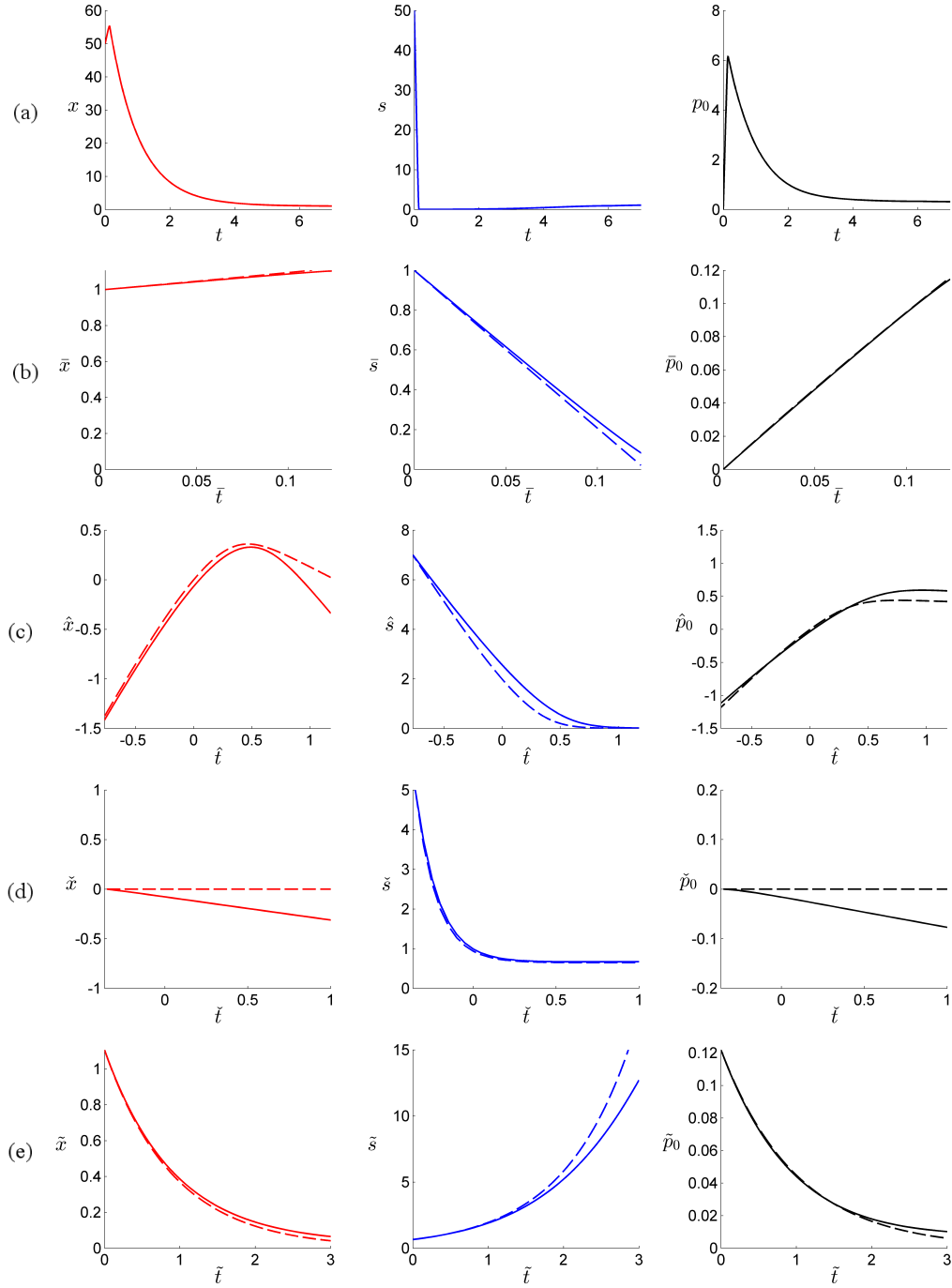
FIGURE 2. Numerical and analytical solutions for cell mass x, substrate s and protein p_0 in regime A, where $x_0 = 50$ and $s_0 = 50$. We take $\epsilon = 0.02$. The solid curves represent numerical solutions and the dotted curves analytical solutions.

(a) illustrates the full time evolution to steady state.

(b) illustrates the first timescale from the expansions (3.4) with the solutions (3.6), (3.8) and (3.10). Here $\bar{x}_0 = \bar{s}_0 = 1$.

(c) illustrates the second timescale from the expansions (3.11) with the solutions (3.13), (3.16) and (3.19). Here $t_0 = 0.13$, $\hat{s}_0 = 2$, $\hat{x}_0 = 1.1$, $\hat{p}_{0,0} = 0.117$.

(d) illustrates the third timescale from the expansions (3.21) with the solution (3.23). Here $\check{t}_0 = 1$, $\check{s}_0 = 1$, $\check{x}_0 = 1.094$, $\check{p}_{0,0} = 0.123$.

(e) illustrates the fourth timescale, with the expansions (3.25) and the solution (3.28). Here $\tilde{x}_0 = 1.094$, $\tilde{s}_0 = 1.31$, $\tilde{p}_0 = 0.123$.

Thus:

$$t = t_0 + \epsilon \hat{t}, \tag{3.11 a}$$

$$x = \frac{1}{\epsilon}(\hat{x}_0 + \epsilon \hat{x}) + \cdots, \tag{3.11 b}$$

$$s = \hat{s} + \cdots, \tag{3.11 c}$$

$$p_0 = \frac{1}{\epsilon}(\hat{p}_{0,0} + \epsilon \hat{p}_0) + \cdots, \tag{3.11 d}$$

with the hatted variables to be found. The governing equations then become

$$\frac{d\hat{s}}{d\hat{t}} = -r_x \frac{\hat{x}_0 \hat{s}}{1+\hat{s}} - \frac{r_0}{Y_0}\frac{\hat{x}_0 \hat{s}}{K_0 + \hat{s}}, \tag{3.12 a}$$

$$\frac{d\hat{x}}{d\hat{t}} = \frac{r_x \hat{x}_0 \hat{s}}{1+\hat{s}} - (1+k_d)\hat{x}_0, \tag{3.12 b}$$

$$\frac{d\hat{p}_0}{d\hat{t}} = \frac{r_0 \hat{x}_0 \hat{s}}{1+\hat{s}} - \hat{p}_{0,0}. \tag{3.12 c}$$

For simplicity we set $K_0 = 1$ here, although this assumption could be relaxed if required. We then set $\beta = r_x + \frac{r_0}{Y_0}$ and integrate to obtain an implicit relation between $\hat{t}$ and $\hat{s}$,

$$-\beta \hat{x}_0 \hat{t} = \hat{s} - \hat{s}_0 + \log\left(\frac{\hat{s}}{\hat{s}_0}\right). \tag{3.13}$$

Explicit features are found for limiting values of $\hat{t}$. For large negative $\hat{t}$ (matching back to the previous stage) we have

$$\hat{s} \sim \hat{s}_0 - \frac{\beta \hat{x}_0 \hat{s}_0}{1+\hat{s}_0}\hat{t}, \tag{3.14}$$

and for large positive $\hat{t}$,

$$\hat{s} \sim \hat{s}_0 e^{-\beta \hat{x}_0 \hat{t}}; \tag{3.15}$$

so $\hat{s} \to 0$ as $\hat{t} \to \infty$. Meanwhile the governing equation for $\hat{x}$ gives the relation

$$\hat{x} = -\hat{x}_0 \left[\frac{r_x}{\beta}(\hat{s} - \hat{s}_0) + (1+k_d)\hat{t}\right]. \tag{3.16}$$

At large negative $\hat{t}$,

$$\hat{x} \sim \hat{x}_0 \left[\frac{r_x \hat{x}_0 \hat{s}_0}{1+\hat{s}_0} - (1+k_d)\right]\hat{t}, \tag{3.17}$$

which is increasing if we ensure this phase is begun early enough. At large positive $\hat{t}$,

$$\hat{x} \sim \frac{r_x \hat{x}_0 \hat{s}_0}{\beta} - (1+k_d)\hat{t}, \tag{3.18}$$

and so $\hat{x}$ has a maximum in this time range and then begins to decrease due to the small values of $\hat{s}$ at this point. This can be clearly seen in figure 1, as the trajectories turn significantly in the bottom right hand corner. Similarly we can integrate our equation for $\hat{p}_{0,0}$ to obtain

$$\hat{p}_0 = -\frac{r_0 \hat{x}_0}{\beta}(\hat{s} - \hat{s}_0) - \hat{p}_{0,0}\hat{t} \tag{3.19}$$

which mimics the behaviour of $\hat{x}$. These analytical solutions are again compared with numerical solutions in figure 2(c).

Once more, our assumptions break down due to the exponential decay of $\hat{s}$ at large time, and another expansion begins at the time $\check{t}_0$ when

$$\epsilon \check{s}_0 \sim \hat{s}_0 e^{-\beta \hat{x}_0 \hat{t}}. \tag{3.20}$$

With this logarithmic time shift the time scale and expansions are:

$$t = t_0 + \epsilon(F\check{t}_0 + \check{t}), \tag{3.21 a}$$

$$x = \frac{1}{\epsilon}(\check{x}_0 + \epsilon\check{x} + \cdots), \tag{3.21 b}$$

$$s = \frac{1}{\epsilon}(\epsilon^2 \check{s} + \cdots), \tag{3.21 c}$$

$$p_0 = \frac{1}{\epsilon}(\check{p}_{0,0} + \epsilon\check{p}_0 + \cdots), \tag{3.21 d}$$

where $F = -\frac{1}{\beta \hat{x}_0} \log(\epsilon \frac{\check{s}_0}{\hat{s}_0})$ is a large positive constant. To leading order, our governing equations become

$$\frac{d\check{s}}{d\check{t}} = s_f - \beta \hat{x}_0 \check{s}, \tag{3.22 a}$$

$$\frac{d\check{x}}{d\check{t}} = 0, \tag{3.22 b}$$

$$\frac{d\check{p}_0}{d\check{t}} = 0, \tag{3.22 c}$$

and so $\check{x}$ and $\check{p}_0$ remain at their initial values $\check{x}_0$ and $\check{p}_{0,0}$ respectively, while

$$\check{s} = \frac{1}{\beta \hat{x}_0} \left[s_f - (s_f - \beta \hat{x}_0 \check{s}_0) e^{\beta \hat{x}_0 \check{t}} \right], \tag{3.23}$$

all of which are illustrated and compared to the numerical solutions in figure 2(d).

As $\check{t} \to \infty$, the exponential term in $\check{s}$ becomes small, and therefore we define a new time scale which begins when

$$\frac{s_f - \beta \hat{x}_0 \check{s}_0}{\beta \hat{x}_0} e^{\beta \hat{x}_0 \check{t}} \sim \epsilon. \tag{3.24}$$

Here we let $\check{t} = G\tilde{t}_0$, where $G = -\frac{1}{\beta \hat{x}_0 \check{s}_0} \log(\frac{\epsilon \beta \hat{x}_0}{s_f - \beta \hat{x}_0})$ and $\tilde{t}_0$ is an arbitrary constant of order 1. Thus the final time scale and expansions are as follows:

$$t = t_0 + \epsilon(F\check{t}_0 + G\tilde{t}_0) + \tilde{t}, \tag{3.25 a}$$

$$x = \frac{1}{\epsilon}\tilde{x} + \cdots, \tag{3.25 b}$$

$$s = \epsilon\tilde{s} + \cdots, \tag{3.25 c}$$

$$p_0 = \frac{1}{\epsilon}\tilde{p}_0 + \cdots, \tag{3.25 d}$$

with

$$\tilde{s}(0) = \frac{s_f}{\beta \hat{x}_0}, \qquad (3.26\,a)$$

$$\tilde{x}(0) = \check{x}_0, \qquad (3.26\,b)$$

$$\tilde{p}_0(0) = \check{p}_{0,0}. \qquad (3.26\,c)$$

To leading order the governing equations become

$$0 = s_f - \beta \tilde{x} \tilde{s}, \qquad (3.27\,a)$$

$$\frac{d\tilde{x}}{d\tilde{t}} = -(1 + k_d)\tilde{x}, \qquad (3.27\,b)$$

$$\frac{d\tilde{p}_0}{d\tilde{t}} = -\tilde{p}_0, \qquad (3.27\,c)$$

and so we have the following pure exponential solutions:

$$\tilde{s} = \frac{s_f}{\beta \tilde{x}_0} e^{(1+k_d)\tilde{t}}, \qquad (3.28\,a)$$

$$\tilde{x} = \check{x}_0 e^{-(1+k_d)\tilde{t}}, \qquad (3.28\,b)$$

$$\tilde{p}_0 = \check{p}_{0,0} e^{-\tilde{t}}. \qquad (3.28\,c)$$

This regime is again clearly visible as the straight diagonal line in figure 1. Additionally, these analytical solutions are compared with the numerical ones in figure 2(e). In logarithmic time our assumptions fail, but the qualitative behaviour does not markedly change as the solutions display linear stability.

While explicit solutions have only been given for the case where $x_0 \sim \frac{1}{\epsilon}$, $s_0 \sim \frac{1}{\epsilon}$, similar or related qualitative behaviour is seen in figure 1 for other initial conditions. This can be explained with reference to regime A. First, if $x_0 \sim 1$ and $s_0 \sim \frac{1}{\epsilon}$, the solution is very similar to regime A illustrated before. However, the first stage represented by the barred variables in (3.4) is slightly different, with (3.7) modified to be

$$\frac{ds}{dt} = -s \qquad (3.29)$$

and hence s falls more slowly. On the other hand, since x is growing exponentially in (3.6), the x term re-enters (3.29) and hence the solution continues as previously described, but the exponential growth lasts for longer, as is visible in figure 1. The solution for p_0 also remains unmodified.

Similarly, if $x_0 \sim \frac{1}{\epsilon}$ and $s_0 \sim 1$, the solution is identical in form to regime A, except that we begin with the hatted variables and (3.11) and (3.12 c). Again, if we have $x_0 \sim \frac{1}{\epsilon}$ and $s_0 \sim \epsilon$, the solutions for cell mass and substrate are the same as for regime A, beginning with the checked variables in (3.21) and (3.22), moving on to the tilde variables (3.25) and (3.27). However, the solution for p_0 in the tilde timescale will clearly be different, as (3.25 d) will not hold due to the zero initial condition. Instead we introduce

$$p_0 = \tilde{\tilde{p}}_0 \qquad (3.30)$$

and hence (3.28 c) is replaced by

$$\frac{d\tilde{\tilde{p}}_0}{d\tilde{t}} = r_0 \tilde{x}\tilde{s} - \frac{r_1}{Y_1} \frac{\tilde{\tilde{p}}_0}{1 + \tilde{\tilde{p}}_0} - \tilde{\tilde{p}}_0, \qquad (3.31)$$

which can be solved implicitly.

3.3.2 *Regime B*

We now describe the behaviour shown by regime B in figure 1, which occurs when the initial values $x_0 \sim \epsilon$ and $s_0 \sim \epsilon$ are both small. The full time evolution is shown in figure 3. These initial conditions and the forcing due to s_f in (2.6 b) imply that initially we have an $O(1)$ time scale and the following expansions:

$$t = T, \tag{3.32 a}$$

$$s = S + \cdots, \tag{3.32 b}$$

$$x = \epsilon X + \cdots, \tag{3.32 c}$$

$$p_0 = \epsilon P_0 + \cdots, \tag{3.32 d}$$

so that the governing equation (2.6 a), (2.6 b) and (2.6 c) become

$$\frac{dS}{dT} = s_f - S, \tag{3.33 a}$$

$$\frac{dX}{dT} = \frac{r_x S X}{1 + S} - (1 + k_d) X, \tag{3.33 b}$$

$$\frac{dP_0}{dT} = \frac{r_0 S X}{K_0 + S} - (\frac{r_1}{Y_1} + 1) P_0. \tag{3.33 c}$$

Thus the expressions

$$S = s_f + (S_0 - s_f) e^{-T}, \tag{3.34 a}$$

$$X = X_0 e^{(r_x - 1 - k_d) T} \left[\frac{(1 + s_f) e^T + (S_0 - sf)}{1 + S_0} \right]^{\frac{-r_x}{1 + s_f}}, \tag{3.34 b}$$

$$P_0 = \frac{\int \frac{r_0 X S}{1 + S} e^{(\frac{r_1}{Y_1} + 1) T} \, dT}{e^{(\frac{r_1}{Y_1} + 1) T}}, \tag{3.34 c}$$

describe the solutions at this stage. Since $r_x - 1 - k_d$ is always positive, exponential growth dominates the expression for x. The numerical and analytical solutions are compared in figure 3, where it is clear that the expansions break down in logarithmic time, and the cell growth term $\frac{r_x x s}{1 + s}$ re-enters the s equation. This is where the trajectory begins moving downwards towards linear stability in the phase plane in figure 1.

3.3.3 *Regime C*

We now consider the top left hand corner of the phase plane in figure 1, where $s(0)$ is large of order $\frac{1}{\epsilon}$ and $x(0)$ is small of order ϵ. The full evolution in time is shown in figure 4. Guided by the initial values and the influence of interactive terms we set the expansions

$$x = \epsilon \grave{x} + \cdots, \tag{3.35 a}$$

$$s = \frac{1}{\epsilon} \grave{s} + \cdots, \tag{3.35 b}$$

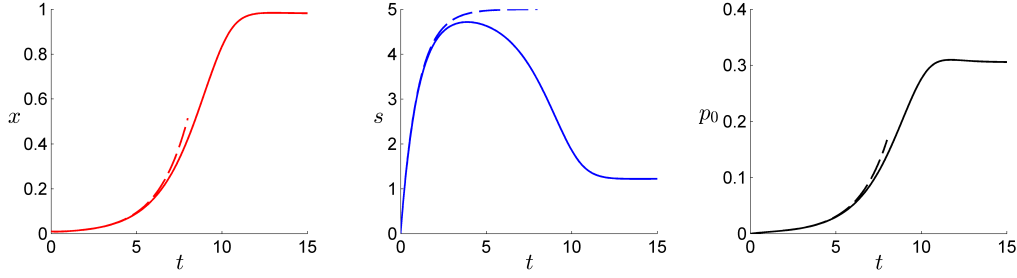$$p_0 = \epsilon \grave{p}_0 + \cdots, \tag{3.35 c}$$

FIGURE 3. Numerical and analytical solutions for cell mass x, substrate s and protein p_0 in Regime B, where $x_0 = 0.01$ and $s_0 = 01$. The solid curves represent the numerical solutions for the full time evolution. The dashed curves represents the analytical solution as given by the expansions in (3.32) and the solutions in (3.34). Here $\epsilon = 0.01$, $X_0 = 1$, $S_0 = 1$

for t of $O(1)$ and hence our governing equations become

$$\frac{d\grave{s}}{dt} = -\grave{s}, \tag{3.36 a}$$

$$\frac{d\grave{x}}{dt} = (r_x - 1 - k_d)\grave{x}, \tag{3.36 b}$$

$$\frac{d\grave{p}_0}{dt} = r_0\grave{x} - (\frac{r_1}{Y_1} + 1)\grave{p}_0. \tag{3.36 c}$$

The solutions are

$$\grave{s} = \grave{s}_0 e^{-t}, \tag{3.37 a}$$

$$\grave{x} = \grave{x}_0 e^{(r_x - 1 - k_d)t}, \tag{3.37 b}$$

$$\grave{p}_0 = \frac{r_0}{r_x - k_d + \frac{r_1}{Y_1}} \left(e^{(r_x - 1 - k_d)t} - e^{-(\frac{r_1}{Y_1} + 1)t} \right). \tag{3.37 c}$$

Hence the dominant process for substrate s is washout, leading to exponential decay. Cell mass x is growing at its maximum rate, but is tempered by washout and cell death, leading to exponential growth. The protein production mimics cell growth, although glycosylation and washout slow the growth. This behaviour is clearly visible in figure 1 in the top left hand corner. Again, a comparison with numerical solutions is shown in figure 4. The expansions cease to hold as s becomes order 1, and the cell growth and protein production terms are no longer linear. However the time evolution continues in a qualitatively similar way, before approaching linear stability.

By considering regimes A, B and C, the qualitative and quantitative behaviour of the entire phase plane can thus be explained. It can be seen that changing the initial conditions for cell mass x and substrate s (which can be done by the experimenter) has a large effect on the quantity of protein in the bioreactor at early time. In particular, the protein production in regime A, shown in figure 2(a), is markedly different from the protein production in regime B, shown in figure 3. In the latter half of the paper, we demonstrate how these two regimes have very different glycosylation patterns at early time, which has implications for the quality of the product.
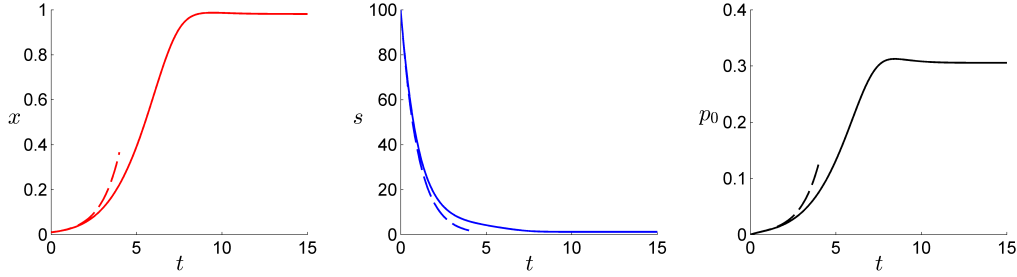
FIGURE 4. Numerical and analytical solutions for cell mass x, substrate s and protein p_0 in Regime C, where $x_0 = 0.01$ and $s_0 = 100$. The solid curves represent the numerical solutions for the full time evolution. The dashed curves represents the analytical solution as given by the expansions in (3.35) and the solutions in (3.37). Here $\epsilon = 0.01$, $\dot{x}_0 = 1$, $\dot{s}_0 = 1$.

4 Glycosylation submodel

We now turn our attention to the glycosylation submodel, in order to determine the shape of the dynamic glycosylation distribution $p_j(t)$. First the steady glycosylation distribution will be examined, before considering the time evolution. The governing equations of this subsystem consist of $(2.6\,c)$, $(2.6\,d)$ and $(2.6\,e)$ as stated earlier. Again, the number of glycoprotein variants, n, is assumed to be relatively large, and the reaction rates r_j and yield coefficients Y_j are assumed constant for all j.

4.1 Steady states

The steady states for each variant, where the time derivatives are equal to zero, depend only on the steady state of the previous variant, because each glycosylation reaction is stepwise. Since all quantities are nonnegative, we can see from $(2.6\,d)$ that each equation has one nonnegative root:

$$p_j = \frac{1}{2}\left(\frac{r_j p_{j-1}}{1 + p_{j-1}} - \frac{r_{j+1}}{Y_{j+1}} - 1 + \sqrt{\left(\frac{r_j p_{j-1}}{1 + p_{j-1}} - \frac{r_{j+1}}{Y_{j+1}} - 1 \right)^2 + 4\frac{r_j p_{j-1}}{1 + p_{j-1}}} \right). \quad (4.1)$$

From $(2.6\,e)$, the terminal glycoprotein has a steady state

$$p_n = \frac{r_n p_{n-1}}{1 + p_{n-1}}. \quad (4.2)$$

We repeat that the evolution in t however is of most concern, as by adjusting the configuration of the bioreactor via the initial conditions for cell mass x and substrate s, the glycosylation distribution can be manipulated. Thus the dynamic behaviour is examined in detail, to identify points at which the quality of the product is favourable.

5 A continuum glycosylation model

To further explore the dynamic glycoform distribution, we consider a partial differential equation model for the glycosylation system. The number of glycoform variants, n, is relatively large [16], and therefore we look to approximate the glycoform distribution

investigated in section 3.1 as continuous. We replace the discrete glycosylation index j with a continuous variable z that represents the extent of glycosylation, and obtain

$$p_j(t) = p(t, z). \tag{5.1}$$

Here the integer points of the continuous variable z correspond to the discrete indices j. However we now allow intermediate reactions where z takes non-integer values. Each glycosylation reaction is thus seen as increasing the extent of glycosylation by δz, i.e. each reaction is denoted by

$$p(t, z) \to p(t, z + \delta z). \tag{5.2}$$

The nondimensional parameters r_j and Y_j are also modified. Since r_j is the maximal reaction rate in the jth reaction, the equivalent rate for the reaction represented in (5.2) is $\frac{R(z)}{\delta z}$. The yield coefficient is defined from the non-dimensionalisation in equations 2.4

$$Y_j = \frac{K_j^* Y_j^*}{K_{j+1}^*}. \tag{5.3}$$

We have assumed perfect conversion and hence $Y_j^* = 1$. If the Monod constants K_j vary smoothly, then we model them by a smooth function $\kappa(z)$. Hence we take

$$Y_j = \frac{\kappa(z)}{\kappa(z + \delta z)}. \tag{5.4}$$

If δz is sufficiently small then we may write

$$\frac{1}{Y_j} = 1 - \frac{\kappa'(z)}{\kappa(z)} \delta z. \tag{5.5}$$

We also introduce the following function for our reaction terms,

$$f(p(t, z), R(z)) = \frac{R(z) p(t, z)}{1 + p(t, z)}. \tag{5.6}$$

The equation (2.6 d) for each intermediate glycoform variant becomes now

$$\frac{\partial p}{\partial t}(t, z) = \frac{1}{\delta z} f(p(t, z - \delta z), R(z)) - \frac{1}{\delta z} f(p(t, z), R(z + \delta z))(1 - \frac{\kappa'(z)}{\kappa(z)} \delta z) - p(t, z). \tag{5.7}$$

We then take the limit as $\delta z \to 0$ and obtain the first order PDE

$$\frac{\partial p}{\partial t}(t, z) = -\frac{df}{dz} + R(z) \frac{\kappa'(z)}{\kappa(z)} f(z) - p(t, z) \tag{5.8}$$

for the system. The function $f(z)$ allows this derivation to hold for many kinetic forms [1; 2]. However, for our case with Michaelis-Menten kinetics given by (5.6) we obtain

$$\frac{\partial p}{\partial t} + \frac{R(z)}{(1 + p)^2} \frac{\partial p}{\partial z} = (R(z) \frac{\kappa'(z)}{\kappa(z)} - R'(z)) \frac{p}{1 + p} - p, \tag{5.9}$$

with a zero initial condition at $t = 0$ and a dynamic boundary condition at $z = 0$ taken from the solution for p_0 from the bioreactor submodel; thus

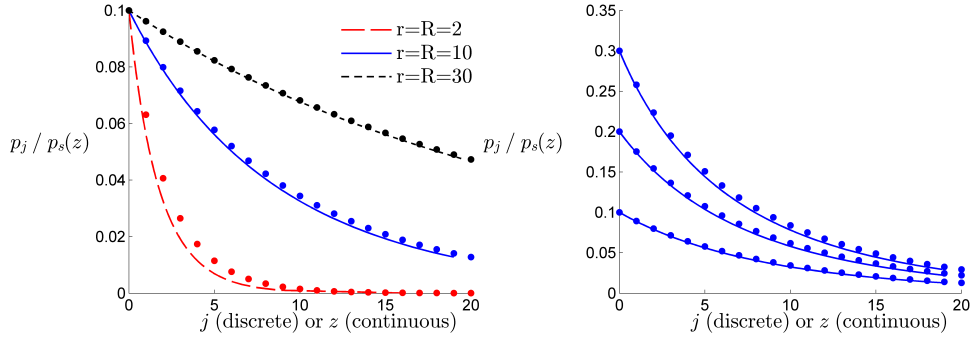$$p(z, 0) = 0, \tag{5.10}$$

$$p(0, t) = p_0(t). \tag{5.11}$$

FIGURE 5. A comparison of the steady-state glycoform distribution for the discrete and continuum models. The round markers are the discrete solutions for the distribution, the sequence p_j at steady state. The left plot shows solutions for different values of $r = R$ and the right plot shows the distribution where $p_0 = 0.1$, 0.2 and 0.3 for $r = R = 10$.

For our analysis, we focus on the case mentioned previously where the parameter values do not vary between reactions and hence $R(z)$ and $\kappa(z)$ are constant, giving

$$\frac{\partial p}{\partial t} + \frac{R}{(1+p)^2} \frac{\partial p}{\partial z} = -p. \tag{5.12}$$

The final glycoform variant $p_n(t)$ has an equation of a different form but can be found from the continuum model as

$$p_n(t) = \int_n^\infty p(t, z) \, dz \tag{5.13}$$

from all the glycoform variants.

5.1 Steady State analysis

We can obtain the steady state $p = p_s$, to compare with the discrete system, by taking $\frac{\partial p}{\partial t} = 0$ in (5.12) which yields

$$\frac{dp_s}{dz} = \frac{-p_s(p_s + 1)^2}{R}. \tag{5.14}$$

From this, certain properties of the glycoform distribution are immediately clear. First, the distribution is always monotonically decreasing. Second, higher values of R (and r for the discrete solutions) lead to a flatter distribution, and finally, higher values of p_0 lead to a steeper distribution. These attributes, as well as a comparison between the steady state results for the discrete and continuum models, are shown in figure 5. A close fit is obtained for values of r above 10, and the agreement is still reasonable at $r = 2$.

We now perform a linear stability analysis on this steady solution, by considering the evolution of a small perturbation to the steady state:

$$p(z, t) = p_s(z) + B(z)e^{-\alpha t} + \cdots \tag{5.15}$$

where $\alpha > 0$ and $B(z) \ll 1$. Substituting this solution into the original PDE (5.12) and

linearising, to first order the PDE becomes, on rearrangement,

$$\frac{dB}{B} = \left[(\alpha - 1)\frac{(p_s + 1)^2}{R} + \frac{2p_s'}{(p_s + 1)}\right] dz. \tag{5.16}$$

Here $'$ denotes differentiation with respect to z. However, from (5.14) we have

$$\frac{(p_s + 1)^2}{R} dz = \frac{dp_s}{-p_s} \tag{5.17}$$

since $p_s' dz = dp_s$ and so (5.16) gives

$$\frac{dB}{B} = \left[\frac{1 - \alpha}{p_s} + \frac{2}{p_s + 1}\right] dp_s. \tag{5.18}$$

Integrating this from $z = 0$, we obtain

$$\frac{B(z)}{B(0)} = \left(\frac{p_s(z)}{p_{0,s}}\right)^{1-\alpha} \left(\frac{p_s(z) + 1}{p_{0,z} + 1}\right)^2. \tag{5.19}$$

If we impose $\alpha > 1$, the perturbation decays in z and thus the steady state is stable as expected.

5.2 Dynamic Solution

Since the equation (5.12) is hyperbolic, it can be solved by the method of characteristics. Two sets of characteristics are found, one which transmits information from the zero initial condition and one that transmits information from the dynamic boundary condition, in the form

$$\frac{dt}{1} = \frac{dz}{\frac{R}{(1+p)^2}} = \frac{dp}{-p}. \tag{5.20}$$

We introduce ξ to parametrise the characteristics, as the point at which a characteristic crosses the t axis where $z = 0$. Along the characteristics p decays exponentially

$$p = p_0(\xi)e^{(\xi - t)}, \tag{5.21}$$

from (5.20). However, if the characteristic does not cross the positive t axis, then we apply the zero initial condition instead of the boundary condition, implying that $p = 0$ along such characteristics. Hence we obtain the "zero" characteristics:

$$z = R(t - \xi). \tag{5.22}$$

By integrating (5.20), the "non-zero" characteristics have the shape

$$z(t, \xi) = h(p_0(\xi)e^{(\xi - t)}) - h(p_0(\xi)) \tag{5.23}$$

where

$$h(p) = -R\left(log\left(\frac{p}{p + 1}\right) + \frac{1}{p + 1}\right). \tag{5.24}$$

The evolution of $p_0(t)$ therefore drives the shape of the characteristics. Thus the evolution of p_0, as discussed in section 3.2 is critical to understanding the dynamic glycoform distribution. Certain p_0 trajectories in fact lead to crossed characteristics. This in turn leads to a discontinuity in $p(z)$ which manifests itself as a shock. In figure 6(a), the initial
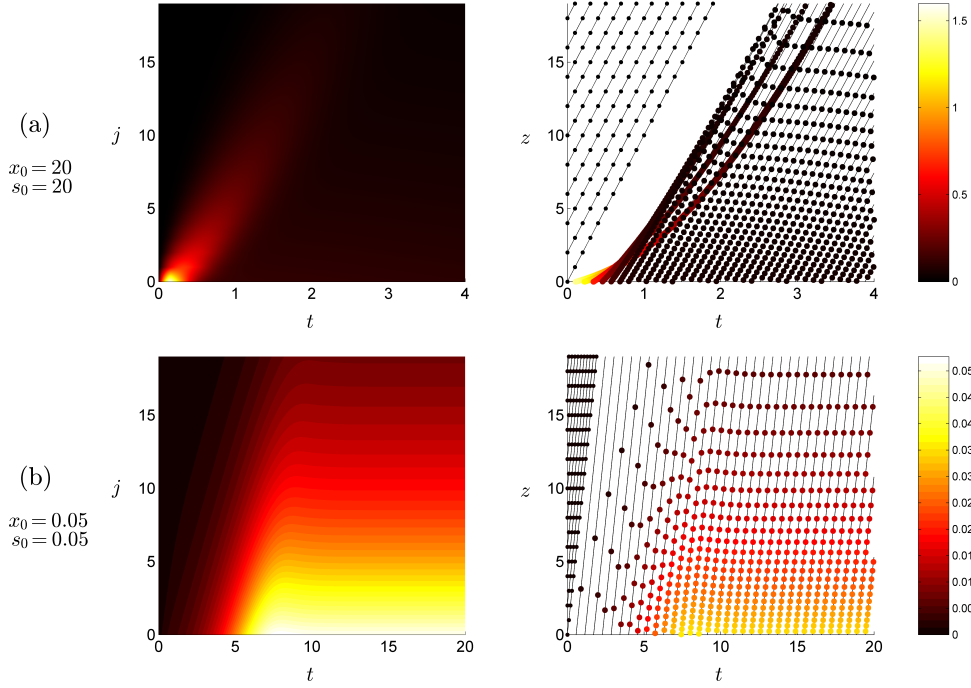
FIGURE 6. A comparison of the dynamical solutions for the discrete and continuous models. The left panels represent the discrete solution. The panels on the right show the characteristics, while the coloured dots give contours of constant p. In (a), the large initial conditions lead to a shock, whereas in (b), the small initial conditions do not. Here $r = R = 10$.

conditions correspond qualitatively to regime A in section 3.2 and a shock is present; whereas in figure 6(b), the initial conditions correspond qualitatively to regime B and a shock does not occur. In both cases there is also reasonably good agreement with the discrete case. The conditions determining the occurrence of a shock are investigated below.

5.3 Shock condition

For a shock to occur, two adjacent characteristics must meet. Hence for small $\delta\xi$,

$$z(t,\xi) = z(t,\xi + \delta\xi), \tag{5.25}$$

and so, from the limit $\delta\xi \to 0$, they will meet if

$$\frac{\partial}{\partial\xi}(z(t,\xi)) = 0, \tag{5.26}$$

and

$$(p_0' + p_0)e^{\xi-t}h'(p_0 e^{\xi-t}) + p_0'h'(p_0) = 0. \tag{5.27}$$

Thus the condition becomes

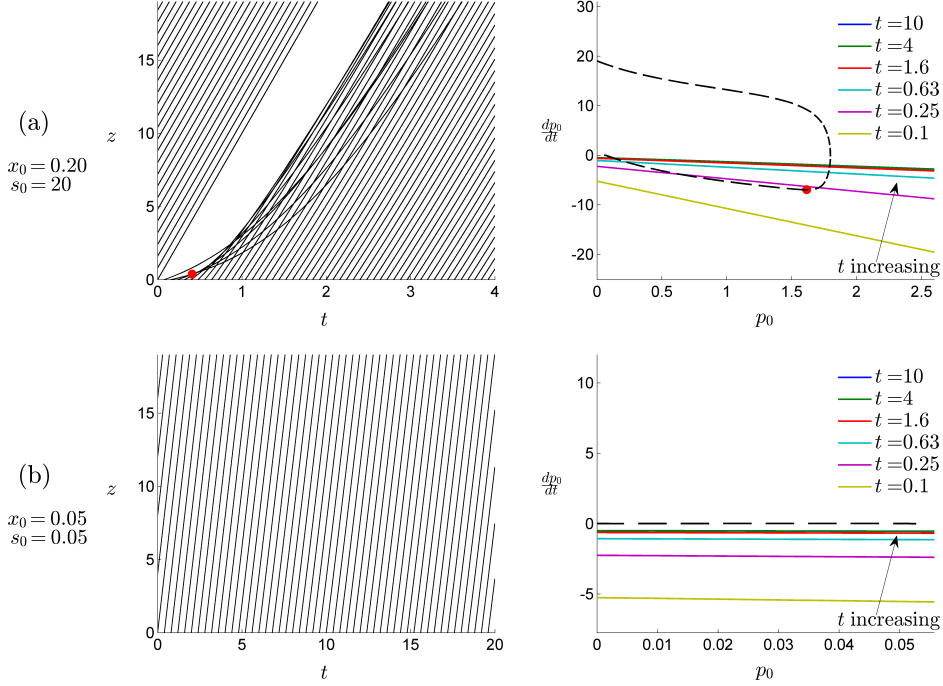$$p_0' = -\frac{(p_0 + 1)^2}{(1 - e^{\xi-t})(p_0(1 + e^{\xi-t}) + 2)}. \tag{5.28}$$

FIGURE 7. The right panels show the conditions for a shock to form. The solid curves represent the conditions for a shock to begin at varying values of t. From the bottom to the top, the values of t are increasing. The dashed curve is the p_0 trajectory corresponding to the initial conditions for x and s. In (a), the large initial conditions mean the trajectory satisfies the conditions for a shock to form and the circular marker shows the point at which the characteristics first cross. This marker in the left panel also shows where the shock begins in the (t, z) plane. In (b), the large initial conditions do not lead to a shock forming.

The condition is used to find the time at which the characteristics first cross, and therefore when the shock begins. This time is the minimum t for which the condition is satisfied, which occurs at the most negative value of p_0'. The value of ξ at which the minimum of p_0' occurs gives the leading characteristic into the shock. In figure 7, the shock condition (5.28) is illustrated with differing values of t for two different sets of initial conditions. In the right panel of figure 7(a) the trajectory intersects with the shock conditions and thus the characteristics begin to cross at $t = 0.24$, as shown by the red circular marker, which is also shown in the left panel as the point where the characteristics begin. In figure 7(b) by contrast, the p_0 trajectory does not intersect with the shock condition for any time, and thus the characteristics shown in the left panel do not cross. This shock condition indicates that a shock will only form if p_0 decreases sharply, as occurs in regime A in section 3.2, but not in regimes B or C.

5.4 Shock position

Since the shock represents a discontinuity, the model needs to be modified to give a physical solution. The simplest way of doing this is to add a small term including the

second z derivative, which is only felt close to the shock. This allows us to calculate the speed of the shock, using say

$$\frac{\partial p}{\partial t} + \frac{R}{(1+p)^2}\frac{\partial p}{\partial z} = -p + \epsilon\frac{\partial^2 p}{\partial z^2}. \tag{5.29}$$

The change of variables $z - h(t) \equiv \bar{z}$, where $z = h(t)$ is the position of the shock, yields

$$\frac{\partial p}{\partial t} - h'(t)\frac{\partial p}{\partial \bar{z}} + \frac{R}{(1+p)^2}\frac{\partial p}{\partial \bar{z}} = -p + \epsilon\frac{\partial^2 p}{\partial \bar{z}^2}, \tag{5.30}$$

and so in the shock layer, i.e. where $\bar{z} = \epsilon\hat{z}$, to leading order we have

$$-h'\frac{\partial p}{\partial \hat{z}} + \frac{R}{(1+p)^2}\frac{\partial p}{\partial \hat{z}} = \frac{\partial^2 p}{\partial \hat{z}^2}. \tag{5.31}$$

We also have conditions at the edge of the shock layer where the solution for p is well defined:

$$\text{As } \hat{z} \to +\infty, \quad p \to p_{+\infty} = p_0(\xi + \xi_{shock})e^{\xi + \xi_{shock} - t}; \tag{5.32}$$

$$\text{As } \hat{z} \to -\infty, \quad p \to p_{-\infty} = p_0(\xi - \xi_{shock})e^{\xi - \xi_{shock} - t}. \tag{5.33}$$

Additionally, as $\hat{z} \to \pm\infty$, $\frac{\partial p}{\partial \hat{z}} \to 0$. By integrating (5.31) between $+\infty$ and $-\infty$ we obtain

$$h'(t) = \frac{R}{(1 + p_{+\infty})(1 + p_{-\infty})} \tag{5.34}$$

which gives the slope of the shock. This is the geometric mean of the slope of the two characteristics entering the shock. The "height" of the shock is $p_{+\infty} - p_{-\infty}$, which has a factor of $e^{\xi-t}$ and therefore decays exponentially as $t \to \infty$, and hence the shock dissipates.

Thus when the model has been modified in this way, the discontinuous shock is replaced by a wave that moves forward in z and dissipates as time progresses. In the left panel of figure 8(a), this wave is visible in the results for the discrete model. In this case, a shock has occurred in the corresponding continuous model, due to the rapid decrease in p_0 visible in the right panel. Thus the glycosylation distribution is no longer always monotonically decreasing, which potentially has impact on the quality of the product. For example, at $t = 0.9$, $p_5 > p_0$, and hence there is more protein with a higher level of glycosylation. By contrast, in figure 8(b), a shock has not occurred in the corresponding continuous model and there is no wave. The glycosylation distribution is always monotonically decreasing, and thus there is no favourable point with higher levels of glycosylation.

6 Discussion

The glycosylation patterns of therapeutic proteins can be highly heterogeneous, and controlling this heterogeneity is an important element in maintaining product quality in the biopharmaceutical industry. Measuring the glycosylation distribution of therapeutic proteins experimentally is a difficult task, and thus mathematical modelling is a useful tool to further understanding. The complexity of detailed glycosylation models [12; 13; 16; 26] has made them difficult to use in process bioreactor models. In the rare cases where
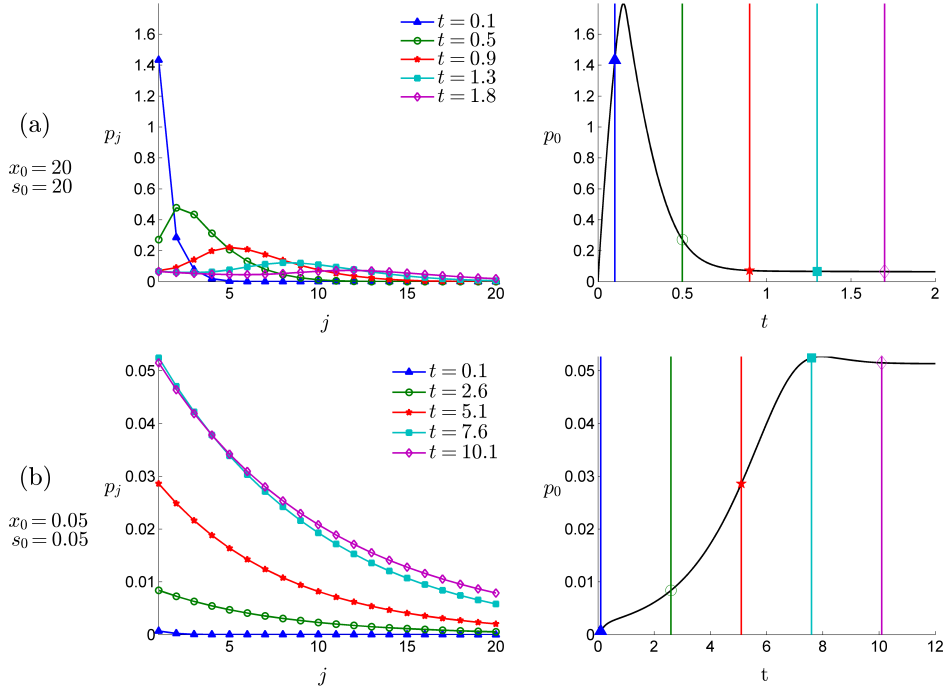
FIGURE 8. The left panels show solutions for the dynamic glycosylation distribution p_j at given time points. On the right the aglycosylated protein production $p_0(t)$ is shown, with the timepoints from the left panels also marked. In (a) the large inital conditions lead to a shock in the continuous model, and this manifests itself as a wave in the discrete model. In (b) the small initial conditions do not lead to a shock, and so there is no wavelike behaviour in the discrete model.

complex glycosylation models are used [15], the values of the parameters used are an important consideration for practical application.

Our focus has been on the detailed analysis of the consequences to the bioreactor of using a simple glycosylation model. The impact of the initial configuration of the bioreactor on the aglycosylated protein production is fully understood, as well as the impact this has on the glycosylation distribution. The simplicity of the model has allowed us to analyse completely the behaviour of any number of product variants, which led to the continuous-mixture limit [1; 7], in which a denumerably infinite number of product variants is considered. This limit has often produced insights that are hard to obtain from discrete models [2], and that trend has held true in this model. The occurrence of a shock in the hyperbolic PDE model, as shown in figure 6, predicted non-monotonic product distributions for certain cases, which also occurred in the discrete analogue with a finite number of product variants, but would have been much harder to discern.

Also on practical applications, it is not known yet if shock formation for example occurs in realistic configurations but the shock effect typifies the range where initial conditions lead to considerable unusual features; these and other features raise interesting possibilities for experimental study. The usefulness of the approximations involved also rests in their demonstration that the responses and subtle structures arising during evolution are quite

different from what is encountered in the steady state, a matter taken up below in terms of quality and quantity during product formation.

We now detail explicitly how the purposeful creation of a non-monotonic product distribution generates product of higher quality and quantity than is found at steady state. The total product quantity is denoted

$$p_{\text{total}}(t) = \sum_{j=0}^{n} p_j(t). \tag{6.1}$$

The quality of the product is defined in two ways: the mean level of glycosylation

$$q_{\text{mean}}(t) = \frac{\sum_{j=0}^{n} j p_j(t)}{p_{\text{total}}(t)}, \tag{6.2}$$

and the fraction of product that is glycosylated above a critical level c

$$q_{\text{thresh}}(t) = \frac{\sum_{j=c}^{n} p_j(t)}{p_{\text{total}}(t)}. \tag{6.3}$$

Changing this critical level does not change the qualitative behaviour. Thus, using the second quality metric, we also define the total amount of good quality product

$$p_{\text{qual}}(t) = p_{\text{total}}(t) q_{\text{thresh}}(t). \tag{6.4}$$

In figure 9, product quantity and quality are described as a function of time for two different initial conditions, $x_0 = s_0 = 20$, which is similar to regime A (dashed line) and $x_0 = s_0 = 0.05$, which is similar to regime C (solid line); these regimes were previously discussed in detail in section 3.2. The values shown are those in the bioreactor at the specific timepoint, but since the reactor is a chemostat these are proportional to those in the outflow which go on to be processed [24].

The top left panel shows the total product quantity, which evolves as predicted. For the small initial conditions, as seen in (3.35) $p_{\text{total}} \sim \epsilon$, and hence the growth to steady state is slow and monotonic. However, for the large initial conditions detailed in (3.4), there is initial fast growth, followed by a slower decline towards the steady state. Hence, the glycoprotein production is larger before the reactor reaches steady state.

However, this also has an impact on product quality in a non-obvious fashion. As seen in the top right and bottom left panels, with large initial conditions both quality measures rise quickly to a maximum before dropping to the steady state. Therefore, for a significant period of time the quality of the glycoprotein is higher. The maximum occurs because of the shock formation in the equivalent continuous model - the glycosylation distribution is then not monotonically decreasing, which therefore increases both quality measures. The bottom right panel shows the total amount of good quality protein, (the amount glycosylated above the threshold), which again shows a maximum for the large initial conditions. This quality increase due to difference in bioreactor configuration is not obvious, and merits further investigation.

One reason we were able to analyse this model fully is its relatively small number of parameters, which is further reduced by passing to the continuous limit of the PDE. The results were shown to be quite generally valid, which is very difficult to show for high-parameter systems, because the number of simulations needed to reveal all the
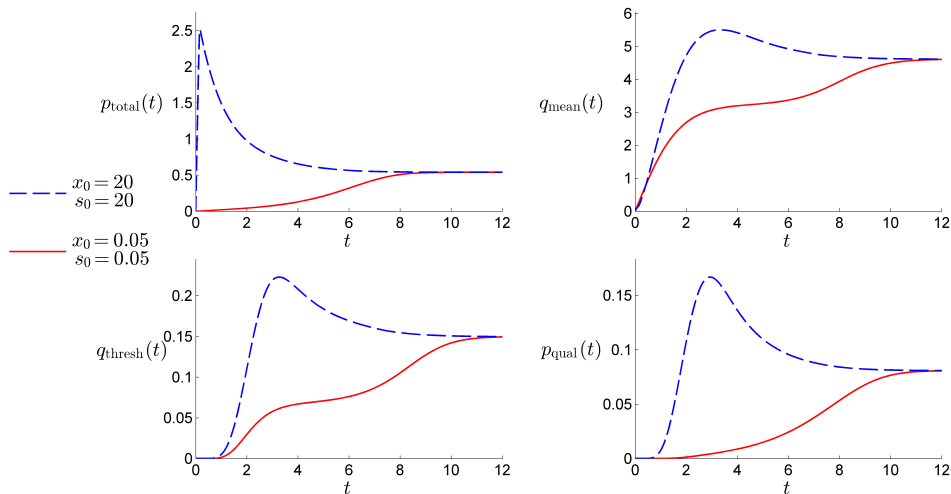
FIGURE 9. The quantity and quality of product formation with two differing initial conditions in the bioreactor, from the discrete model. Two quality metrics are shown, the mean level of glycosylation in the top right panel and the proportion above a threshold level of glycosylation in the bottom left panel. The total good quality glycoprotein in the bottom right panel shows the amount of protein that is glycosylated above that threshold. There are 20 glycoprotein variants and the glycosylation threshold is 10. Here $r_x = 10$.

interactions grows exponentially with the number of parameters. A related benefit is that the parameters of our model could be easily determined from experiment, since it is clearly not over-parametrised.

A natural extension is to apply the model to other reactor modes, especially the fed-batch mode that is widely used in therapeutic production. The simple model for cell growth could also be made more realistic, by accounting for cell heterogeneity. Production bioreactors, with volumes exceeding $10m^3$, inevitably are not well-mixed, and have regions of relatively low oxygen and substrate levels. The state of cells in these regions will also be different from those in other regions. These model adaptations could be investigated in future work.

The complex interactions among the product variants, even in this simple kinetic model, have led to unexpected operating conditions leading to effective productivities that are quite different from the corresponding steady-state values. Such results have been discussed in the literature for very different chemical systems [5; 6]. The possibility that such improved productivity could be attained for more realistic bioreactor modes is intriguing, and we are studying such generalisations.

Acknowledgements

References

[1] ARIS, R. (1989) Reactions in continuous mixtures. *AIChE Journal* **35**(4), 539–548.

[2] ARIS, R. & GAVALAS, G. (1966) On the theory of reactions in continuous mixtures. *Philosophical Transactions of the Royal Society A* **260**(1112), 351–393.

[3] ARNOLD, J. N., WORMALD, M. R., SIM, R. B., RUDD, P. M., & DWEK, R. A. (2007) The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annual Review of Immunology* **25**(1), 21–50.

[4] BAND, L. R. & KING, J. R. (2012) Multiscale modelling of auxin transport in the plant-root elongation zone. *Journal of Mathematical Biology* **65**(4), 743–85.

[5] BORESKOV, G. K. & MATROS, Y. S. (2006) Unsteady-State Performance of Heterogeneous Catalytic Reactions. *Catalysis Reviews* **25**(4), 551–590.

[6] BRUNS, D. & BAILEY, J. (1975) Process operation near an unstable steady state using nonlinear feedback control. *Chemical Engineering Science* **30**(7), 755–762.

[7] CHOU, M. Y. & HO, T. C. (1989) Lumping coupled nonlinear reactions in continuous mixtures. *AIChE Journal* **35**(4), 533–538.

[8] DIMITROV, D. S. (2012) *Therapeutic proteins.*, volume 899. Humana Press, New York.

[9] EDELSTEIN-KESHET, L. (1988) *Mathematical Models in Biology.* SIAM, Philadelphia, Pennsylvania.

[10] ELLIOTT, S., LORENZINI, T., ASHER, S., AOKI, K., BRANKOW, D., BUCK, L., BUSSE, L., CHANG, D., FULLER, J., GRANT, J., HERNDAY, N., HOKUM, M., HU, S., KNUDTEN, A., LEVIN, N., KOMOROWSKI, R., MARTIN, F., NAVARRO, R., OSSLUND, T., ROGERS, G., ROGERS, N., TRAIL, G., & EGRIE, J. (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nature Biotechnology* **21**(4), 414–21.

[11] HOSSLER, P., KHATTAK, S. F., & LI, Z. J. (2009) Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* **19**(9), 936–49.

[12] HOSSLER, P., MULUKUTLA, B. C., & HU, W.-S. (2007) Systems analysis of N-glycan processing in mammalian cells. *PLOS ONE* **2**(8), e713.

[13] JIMENEZ DEL VAL, I., NAGY, J. M., & KONTORAVDI, C. (2011) A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. *Biotechnology progress* **27**(6), 1730–43.

[14] KIM, P.-J., LEE, D.-Y., & JEONG, H. (2009) Centralized modularity of N-linked glycosylation pathways in mammalian cells. *PLOS ONE* **4**(10), e7317.

[15] KONTORAVDI, C., ASPREY, S. P., PISTIKOPOULOS, S., & MANTALARIS, A. (2005) *European Symposium on Computer-Aided Process Engineering-15, 38th European Symposium of the Working Party on Computer Aided Process Engineering*, volume 20 of *Computer Aided Chemical Engineering*. Elsevier, Amsterdam, Netherlands.

[16] KRAMBECK, F. J. & BETENBAUGH, M. J. (2005) A mathematical model of N-linked glycosylation. *Biotechnology and Bioengineering* **92**(6), 711–28.

[17] LEADER, B., BACA, Q. J., & GOLAN, D. E. (2008) Protein therapeutics: a summary and pharmacological classification. *Nature reviews. Drug discovery* **7**(1), 21–39.

[18] O'DEA, R. D. & KING, J. R. (2011) Multiscale analysis of pattern formation via

intercellular signalling. *Mathematical Biosciences* **231**(2), 172–85.

[19]REICHERT, J. M. (2003) Trends in development and approval times for new therapeutics in the United States. *Nature reviews. Drug discovery* **2**(9), 695–702.

[20]SINCLAIR, A. M. & ELLIOTT, S. (2005) Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. *Journal of Pharmaceutical Sciences* **94**(8), 1626–35.

[21]SMITH, F. T. (2010) Finite-time break-up can occur in any unsteady interacting boundary layer. *Mathematika* **35**(02), 256.

[22]SMITH, F. T. & BOWLES, R. I. (1992) Transition Theory and Experimental Comparisons on (I) Amplification into Streets and (II) a Strongly Nonlinear Break-up Criterion. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **439**(1905), 163–175.

[23]SMITH, F. T. & BURGGRAF, O. R. (1985) On the Development of Large-Sized Short-Scaled Disturbances in Boundary Layers. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **399**(1816), 25–55.

[24]SMITH, H. L. (1995) *The Theory of the Chemostat: Dynamics of Microbial Competition.* Cambridge University Press, England.

[25]SPEARMAN, M., DIONNE, B., & BUTLER, M. (2011) The role of glycosylation in therapeutic antibodies. In AL-RUBEAI, M., editor, *Antibody Expression and Production*, 251–292. Springer, Dordrecht, Netherlands.

[26]UMANA, P. & BAILEY, J. E. (1997) A Mathematical Model of N-Linked Glycoform Synthesis. *Biotechnology and Bioengineering* **55**(6), 890–908.

[27]VON DER LIETH, C. (2004) Bioinformatics for glycomics: status, methods, requirements and perspectives. *Briefings in bioinformatics* **5**(2), 164–178.