MODELING AND PARAMETER ESTIMATION IN BIOLOGICAL APPLICATIONS

MODELING AND PARAMETER ESTIMATION IN BIOLOGICAL APPLICATIONS

by

Brian Macdonald, B.Eng

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Master of Applied Science

McMaster University

MASTER OF APPLIED SCIENCE (2016) McMaster University

(Chemical Engineering) Hamilton, Ontario, Canada

TITLE:	Modeling and parameter estimation in biological applications
AUTHOR:	Brian Macdonald, B. Eng (McMaster University, Canada)
SUPERVISOR:	Dr. Prashant Mhaskar
NUMBER OF PAGES:	ix, 73

ABSTRACT

Biological systems, processes, and applications present modeling challenges in the form of system complexity, limited steady-state availability, and limited measurements. One primary issue is the lack of well-estimated parameters. This thesis presents two contributions in the area of modeling and parameter estimation for these kinds of biological processes. The primary contribution is the development of an adaptive parameter estimation process that includes parameter selection, evaluation, and estimation, applied along with modeling of cell growth in culture. The second contribution shows the importance of parameter estimation for evaluation of experiment and process design.

ACKNOWLEDGEMENTS

I would like to thank McMaster University, the Department of Chemical Engineering, and the McMaster Advanced Control Consortium for funding and support of this work. I would especially thank my supervisor, Dr. Prashant Mhaskar, for his tireless efforts and encouragements that let me push through when the goal seemed out of reach. I would like to thank my committee members, Dr. Ghosh and Dr. Adams, for taking the time to evaluate this work and attend my defence. I would like to thank Dr. Bramson and his students for providing experimental work, knowledge, and direction. I would like to thank all of my co-authors for their assistance.

I would like to thank my colleagues in graduate school for their own encouragement, assistance, and occasional necessary distractions. In particular, I would like to thank Giancarlo Dalle Ave for motivating me at the gym and helping me keep healthy, and Jake Nease for being a great friend, coach, and landlord.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Biological systems are often viewed as separate from traditional areas of chemical engineering research such as process system design, modeling, control, and optimization. However, all of these tools can be applied to biological systems.

Biological systems have several distinct features that present challenges that need to be met when it comes to modeling. Biological processes (whether it is cell culturing, drug production, or others) often take the form of batch processes, with no apparent steady-state. In a batch process, the process is started with some initial set of ingredients which then react and change over time. During the batch, additional ingredients may be added or removed, and changes to the batch conditions will often occur. Batch processes do not tend to reach any steady-state condition, and often involve highly non-linear reaction dynamics. These conditions make certain types of system identification and modeling more challenging.

Another challenge that we face with biological systems is defining the complexity of the system. Depending on the scale, a biological process can seem to be almost infinitely complex, with hundreds or thousands of simultaneous reactions. Each cell in a culture of millions is a tiny chemical factory producing many different products. Deciding at what level we describe a model of a biological system is an important element to these systems.

Biological systems are also often relatively poorly measured compared to traditional chemical processes. Whether it is due to risk of contamination, lack of sensitive instruments, or simply expense, biological systems often have relatively few available measurements, and even fewer that are available on-line. Many measurements need to be done after the batch has been completed, or require sampling of the batch to perform complex tests. Linking these few available measurements to the state of the batch as a whole may be a difficult task.

Finally, one of the biggest challenges in modeling biological systems is a lack of well-estimated parameters. Due to the potential complexity of the system, a model of a biological process can contain numerous parameters that are not available from literature or previous research. The difficulty with measurements, as described, can make independent measurement of a parameter difficult. Therefore, parameter estimation needs to be a priority when creating models of biological systems.

This work presents two cases of modeling and parameter estimation for a biological process or application. Chapter 2 presents work that has been done on a T cell culturing process. In this work, a model of T cells in culture was developed based on a population balance model, combined with a model of the media conditions. A parameter estimation process was implemented which including a parameter selection method, training estimation, and an adaptive estimation process to deal with biological variability between cultures.

Chapter 3 presents a work done on a PEGylation process in which a protein is modified to enhance its use for drug applications. The PEGylation process was modeled and then parameter estimation was performed to determine appropriate parameters of reaction kinetics. The model was then used to determine more optimal reactor conditions. The author of this thesis was not the primary author of that work, but contributed in the area of modeling, parameter estimation of the reaction kinetics, optimization, and assisted with collection of experimental data for validation of the estimated model.

Chapter 4 of this work presents the main conclusions and proposals for future areas of research in modeling and estimation of biological systems.

Chapter 2

Modeling and Adaptive Parameter Estimation of a T Cell Culture Process

Manuscript Overview

This chapter introduces the primary contribution of this work. In the manuscript presented below, a model T cell growth in batch cultures is developed. This model uses a population balance model method to describe the cell growth mechanisms, and links it to a 2-phase media model which describes the batch conditions the cells are grown in. As the model contains a number of uncertain parameters, we developed a parameter ranking, selection, and estimation procedure to provide a model with better predictive capabilities. This process is demonstrated on both simulated cases and on experimental data. One question raised by working with biological systems is the issue of biological variability. Variations between cells from different sources and origins, as would be seen in any personalized cell therapies, can be both significant and difficult to determine prior to the process. Therefore, we developed an adaptive estimation approach which uses data from partway into the growth process to update the parameter estimation and provide improved predictive capability for the remainder of the process.

Modeling and Adaptive Parameter Estimation of a T Cell Culture Process

Brian Macdonald, Xueya Feng, Seung Mi Yoo, Jonathan Bramson, Raja Ghosh, Prashant Mhaskar

Submitted: August, 2016, in *Journal of Biotechnology*

Abstract

This work considers the problem of control oriented modeling of a human cytotoxic T cell *in vitro* growth process that is utilized for cell-mediated therapy products. To this end, first, a population balance is used for describing the growth and division of cells. A structured two-phase media model is used to describe the external environment. Parameter estimation is first performed using simulated and experimental data. Finally, an adaptive parameter estimation procedure is implemented to address variations across samples, and demonstrate the viability of the proposed method for use in online control implementations.

2.1 Introduction

Cytotoxic T cells are an essential part of the immune system, and there are promising cellular therapies involving the use of engineered T cells to fight certain types of cancer [1, 2]. In order for a therapy to be effective, a sufficiently high number of T cells need to be cultured from a small initial batch. Viability of the process, and eventual automation, relies on the appropriate choice of the growth protocol, along with the ability to make adaptations to the protocol to account for biological variability across each new batch of cells.

Several different methods are currently used for T cell expansion. One method, and the one we will focus on in this paper, is characterized by the use of increasing culture vessel sizes as the number of cells expand [3]. Beginning with small culture volumes (such as 96 well plates), the cells are expanded into larger wells and subsequently into tissue culture flasks. Transferring the cells into larger vessels is intended to keep the cells at optimal cell density for growth, while also providing greater surface areas for oxygen transfer to occur. The change also refreshes the media, removing waste and adding nutrients [3]. Another method is to use larger culture volumes with some method of assisted oxygen transfer. This is seen particularly in the use of a gas-permeable rapid expansion vessel (G-Rex). Here, the cells are cultured in a large vessel with a gas-permeable membrane at the bottom of the vessel. This allows direct oxygen transfer to the site of the settled cells. There is no need to transfer cells, as new media is added to dilute waste products and renew concentration of nutrients. The membrane is intended to allows sufficient oxygen transfer [4].

Development of standard protocols provides an excellent base case for producing these cells. Successful implementation of these protocols however, requires the use of highly skilled technicians, and is not readily amenable for large scale production, which needs to adapt to biological variations and disturbances. This motivates the use of models to describe and predict cell growth, with the key ability to adapt and recognize biological variation. Developing a model with adaptive capability allows for adjustments to be made during a batch to reduce the impact of variability on that particular process and to provide a more consistent outcome.

Cell growth processes are inherently difficult to model for a number of reasons. Batch conditions can be difficult to measure without introducing potential contamination or disturbing cell growth, so limited measurements are available compared to traditional chemical processes. Secondly, cell growth and division are complicated processes. Limited or indirect measurements and potential noise associated with these measurements makes parameter estimation (for the model) difficult.

Despite these difficulties, models of cell growth have been developed for a number of different cell types under different conditions of growth. Some of the early efforts include development of a basis for mathematical modeling of bacterial populations, [5] [6] and mammalian cells [7]. One issue with these models is that one of the key characteristics used is cell age, which is not readily available for measurements. Similarly, others have looked at modeling division synchrony and other dynamics in bacterial populations. [8] [9] Models of eukaryotic cells have also been developed. One of the model cells for this type of model are hybridoma cells (hyrbridized antibody-producing cancer cells). Researchers have examined the modeling of hybridoma cultures in continuous suspension cultures. [10] [11] These models, however, do not translate directly to use with the T cell culture process in this paper, because both the cells and processing conditions (settled, un-mixed reactor and batch growth) are different.

More recent efforts of cell culture modeling have been focused on the promising tool of population balance modeling (see e.g., [12, 13, 14] for continuous cultures of yeast). Other works have examined the use and solution of population balance models from a more general perspective, and not for a specific growth process. [15, 16] This paper uses similar modeling techniques, although there are significant differences that separate the literature models and the model presented in this work. Most notably, a dynamic growth model for T Cell cultures simply does not exist. Secondly, existing models have typically been for continuous, homogeneous cultures as opposed to the present work where the cells grow in batch, non-homogeneous behavior with no steady-state available. In summary, there does not exist a model (with associated parameters) that can be potentially utilized for control and optimization of the T Cell growth process.

Motivated by these considerations, we present a modeling and adaptive parameter estimation approach for the T cell growth process. The remainder of this paper is organized as follows: In the second section, we introduce the necessary preliminaries for population balance models and parameter estimation. In the following section, we propose a model of cell growth which contains two main components. A population balance method is used to describe the growth of the cells, while a two-phase media model is used to model the balance of substrates and oxygen in the surrounding culture medium. These two components are linked into a single model of T cell growth in culture. In the next section, parameter ranking and estimation is first performed using simulated and experimental data. An adaptive parameter estimation procedure is next implemented. The use of an adaptive estimation process is particularly motivated by the issue of biological variation between different batches of cells (from different patients). This variation introduces difficulties for parameter estimation based on previously obtained measurements, as new batches may differ significantly from the older data. Simulation results demonstrate the viability of the proposed method for use in online control implementations. Finally, the last section summarizes the paper and presents the main conclusions.

2.2 Preliminaries

In this section, we provide an overview of T Cell growth mechanism, followed by a description of population balance models and parameter estimation concepts.

2.2.1 T Cell growth mechanism

On the individual cell level, cell growth and division are complex processes involving many biochemical reactions, signalling processes, and regulatory mechanics. At a simple level, the growth process can be viewed as containing the following main elements: the growth of cells through consumption of external substrates, and the division of cells to create new daughter cells, and cell death (see Figure 2.1 for a schematic).

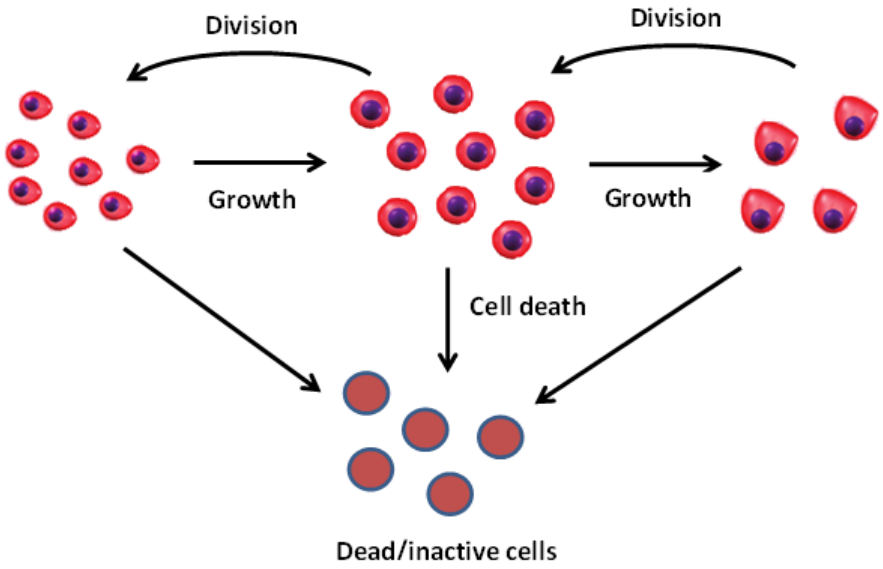


Figure 2.1: A schematic of generic cell growth processes

Some amount of cell growth must occur before a new daughter cell is ready to divide (otherwise a cell could divide indefinitely, creating smaller and smaller cells). Therefore, the division of cells is understood to be dependent on the size (or, in some models, age) of the cell, which in turn is dependent on the intake of substrates from the surrounding media. A natural means of modeling this behaviour is through population balance models, which are briefly introduced in the next section.

2.2.2 Population balance models

Population balance models (PBMs) are a useful tool that can be utilized for a variety of situations in which the system contains both a discrete particulate phase and a continuous phase in which the particles are dispersed. The mathematical basis of the method was proposed first as a way to model the growth of prokaryotic (bacteria) cells [6]. A similar work [5] used a slightly different approach for the same subject matter. In more recent work, PBMs have been proposed to model the growth of yeast in a continuous flow bioreactor. [13] [14]

The central idea behind a population balance model is the notion of modeling the rate of change of particular state variables which characterize the particles [17]. The choice of state variable depends on the system in question. Examples include size for a crystal nucleation and growth process, chemical concentration for the droplets in a liquid-liquid extraction process, or catalyst spatial density for a fluidized bed reactor [17]. The central part of a population balance model is a population balance equation (PBE) that maps how these characteristic variables change.

Although the characteristic variable may exist on a continuous basis, this cannot be readily captured in the models, as the resulting partial integro-differential equation is difficult to solve. Analytical solutions are possible, but only under certain restricted assumptions [18]. In order to solve the PBE numerically, it is common to use some method of discretization of the variable range into appropriate bins. Then the PBE maps the movement of particles from one bin to another. Finite difference methods are one common tool for this discretization. [12] [15] An alternate method, and the one used for this paper, is to use orthogonal collocation on finite elements for the discretization. In this way, the PBE is approximated by a set of nonlinear ordinary differential equations. The integral expressions in the PBE are approximated using Gaussian quadrature methods [19]. This method, although more complex, is more computationally efficient than finite differences, and potentially more accurate, depending on the collocation.

2.2.3 Parameter estimation

Parameter estimation in essence relies on solving an optimization problem. The starting point is a model of a system that often contains a number of parameters. These parameters may relate directly to elements of the real system, they may be simplifications of mechanics that occur in the system, or they may be mathematical constructs that are useful to help the model behaviour match the real system. In parameter estimation, we want to determine values of these parameters such that the difference between the model and the real system is minimized. In turn, this introduces the issue of measurements, as measurements provide the

comparison points between the model and system. In a system where limited measurements are available, we must realize that we likely cannot estimate all the parameters in our model. This is due to problems of identifiability and 'estimateability'. The influence of multiple parameters on a measured output means that the we cannot uniquely identify the impacts of each parameter on the model, nor can we then estimate all of those parameters. Instead, we can identify key parameters by the magnitude of their impact on the measured output of the model, and focus on estimating those key parameters.

A number of papers on parameter identifiability, 'estimateability', and selection have been published. Some of these techniques involve repeated parameter estimations to determine identifiability of each parameter, which can be very computationally intensive [20]. Other methods depend on a calculated sensitivity matrix to evaluate the overall effect of each parameter on the outputs [21]. This type of procedure is more prevalent when the outputs are steady-state, rather than dynamic measurements, as is the case with the present application.

2.3 Model development

The model for the T cell growth process comprises of population and material balances. The population balance equation appropriate for cell cultures takes the following form:

$$\frac{\partial}{\partial t}W(m,t) + \frac{\partial}{\partial m}[K(S)W(m,t)] + [\Gamma(m,S)W(m,t+k_d)]W(m,t) = 2\int_m^\infty \Gamma(m')P(m,m')W(m',t)\mathrm{d}m \quad (2.1)$$

where m is the cell mass; $W(m,t)$ represents the cell mass distribution at time t; S is the concentration of glucose present in the cell; $K(S)$ is the cell growth rate; $\Gamma(m)$ is the division intensity function for a cell of mass m; and $p(m,m')$ is the partition function which models the probability distribution function of newborn cells. $\Gamma(m)$ represents the division intensity

function, the rate at which cells divide according to their mass and takes the following form:

$$\Gamma(m) = 0 \text{ if } m \leq m_t \tag{2.2}$$

$$\Gamma(m) = \gamma e^{-\epsilon(m-m_d)^2} \text{ if } m \in [m_t, m_d] \tag{2.3}$$

$$\Gamma(m) = \gamma \text{ if } m \geq m_d \tag{2.4}$$

As mass is conserved during cell division, it serves as the characteristic variable for the population balance equation. The formulation recognizes that below a minimum mass m_t, the cells are too small to begin dividing. Past m_t, the rate of division for a cell increases in a sigmoidal fashion up to a certain maximum rate γ at mass m_d. ϵ is a parameter that modifies the curve in the range between zero division and maximum division frequency.

The partition function describes how the mass of a dividing cell is split between the two daughter cells. Rather than assuming equal partitioning for all cell divisions, it is modeled as a symmetric beta distribution centred around the equal partition case. The partition function is governed by one parameter, q, which determines how narrow or broad the distribution function is.

$$p(m, m') = \frac{1}{B(q,q)} \frac{1}{m'} \Big(\frac{m}{m'}\Big)^{q-1} \Big(1 - \frac{m}{m'}\Big)^{q-1} \tag{2.5}$$

where $B(q,q)$ is the beta function.

The growth rate of the cells is modeled using Monod kinetics, considering only oxidation of glucose.

$$K(S', O) = \mu_{max} \frac{S'}{K_S + S'} \frac{O}{K_O + O} \tag{2.6}$$

where S' represents intracellular glucose concentrations and O represents oxygen concentration in the surrounding media (oxygen diffusion into the cells is assumed to be rapid enough that external concentration is identical to intracellular concentration). K_S represents the glucose saturation coefficient, and K_O represents the oxygen saturation coefficient. μ_{max} is the maximum specific growth rate.

The rate of formation of glucose (r_s) is modeled using a mass balance based on the growth

rate of the cells as well as the biomass yield ratio Y, representing the amount of cell mass gained per mass of glucose consumed.

$$r_S = -\int_0^\infty \frac{K(S', O)}{Y} W(m, t) \mathrm{d}m \tag{2.7}$$

The overall rate of change of glucose concentration in the media surrounding the cells is found through a mass balance between the consumption of glucose from Eq. 7 and the inflow of glucose through diffusion. Therefore:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = J_S \frac{A}{V_c} + \frac{r_S}{V_c} \tag{2.8}$$

where J_S is the mass flux of glucose into the surrounding media (see Eq. 12), A is the surface area, and V_c is the volume of the cell-rich layer, and S is the extracellular glucose concentration.

The rate of change of intracellular glucose depends on the glucose concentration in the surrounding media.

$$\frac{\mathrm{d}S'}{\mathrm{d}t} = \alpha(S - S') \tag{2.9}$$

where α is a constant parameter. The rate of formation of oxygen depends on the rate of formation of glucose, as the only source of oxygen depletion is through glucose oxidation. Therefore, oxygen consumption can be modeled on a stoichiometric balance with glucose consumption.

$$r_O = -\frac{192}{180} \int_0^\infty \frac{K(S', O)}{Y} W(m, t) \mathrm{d}m \tag{2.10}$$

Similarly to Eq. 8, the change in oxygen concentration is modeled through a mass balance combining the consumption of oxygen from Eq. 10 and the inflow through diffusion.

$$\frac{\mathrm{d}O}{\mathrm{d}t} = J_O \frac{A}{V_c} + \frac{r_O}{V_c} \tag{2.11}$$

where J_O is the flux of oxygen into the cell-rich layer (see Eq. 13).

2-phase media model

Although the population balance model is used to describe how cells grow, it is incomplete without considering the state of the surrounding cell culture media, as the PBM depends on the concentrations of glucose and oxygen in the media. To complete the model, this work develops a 2-phase media model to represent the settling behaviour of the cells, as well as the impact of oxygen and glucose diffusion through the media. Because of the size and slightly greater density of T cells, the cells tend to settle to the bottom of the culture vessels used. Although the settling is a slow process, the long length of the growth period allows us to assume the cells to be relatively settled for most of the growth process. Therefore, cell settling is not modeled as a process in this paper. Instead, it is assumed that the cells are settled into the bottom 1% of the vessel. This type of settling has been shown experimentally for murine hybridoma cells [22].

This model of settling results in a system with two distinct phases. First, there is a media-rich phase which composes the large bulk of the total media (99% of the height of the media). This phase is assumed to contain no cells. This phase is exposed to the external conditions (atmospheric oxygen), and serves as a diffusion layer through which both oxygen and glucose must diffuse to reach the cells. Second, there is a cell-rich phase below the media-rich phase. All cells are assumed to exist this layer. Due to the relatively small volume of this layer, the model assumes this phase to be well-mixed, with no local concentration differences of cells or any substrate (all cells exposed to the same external conditions).

The media rich phase acts as a reservoir of glucose. Diffusion through the media-rich phase is modeled as occurring along a linear gradient. As the model tracks both the concentration of glucose in the cell-rich phase as well as the total amount of glucose present, it is possible to calculate the gradient in concentration between the surface of the culture vessel and the boundary of the two phases. Subsequently, the flux can be found through Fick's 1st law:

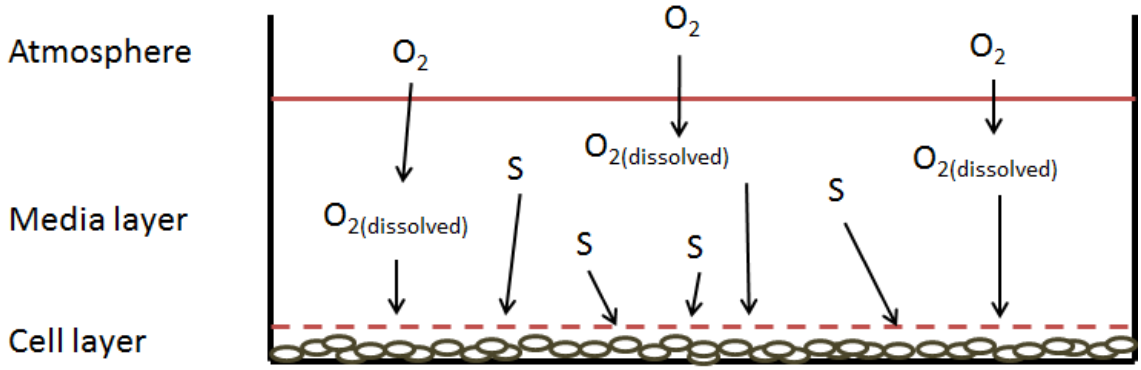


Figure 2.2: Diagram of the 2-phase media approximation used for settled cell growth

$$J_S = D_S \frac{(S^* - S)}{(h_l - h_{cells})} \tag{2.12}$$

where S^* represents the calculated concentration of glucose at the surface of the media:

$$S^* = 2\frac{S_{tot}}{V_{tot}} - S \tag{2.13}$$

where S_{tot} is total mass of glucose in the culture vessel at a particular point in time, V_{tot} is the total culture volume, and S is the concentration of glucose in the cell-rich phase.

Diffusion of oxygen also occurs through the media-rich phase. Concentration at the media surface is considered to be in equilibrium with atmospheric oxygen conditions. At the lower boundary, the concentration of oxygen is assumed equal to the concentration of oxygen in the cell-rich phase. Since no oxygen is consumed in the media-rich phase, a linear concentration gradient exists between the top and bottom of the media-rich phase. The oxygen flux is thus given by:

$$J_O = D_O \frac{(O^* - O)}{(h_l - h_{cells})} \tag{2.14}$$

where J_O is the flux of oxygen into the cell layer, D_O is the diffusivity (diffusion coefficient) of oxygen in cell culture media, O^* is the oxygen saturation concentration in cell culture

media, O is the oxygen concentration in the cell layer, h_l is the overall height of the media, and h_{cells} is the height of the cell layer.

It should be noted that this model of diffusion assumes that the concentration at the surface of the media layer remains at the oxygen saturation point. This assumption was tested by using an effective mass transfer coefficient to represent the oxygen transfer from the surrounding atmosphere into the media. For the simulations (and the initial choice of parameters), it was found that the transfer from the surrounding air into the media was not a limiting factor. Instead, the limiting rate is the transfer of oxygen through the media layer from the surface to the cell layer. Therefore, the assumption of saturated concentrations at the surface hold for the present model.

2.3.1 ODE approximation for numerical solution

From the approximation of the PBE shown in Eq. 1 using orthogonal collocation, we arrive at the following set of nonlinear ODEs:

$$\frac{\mathrm{d}W_j}{\mathrm{d}t} = -\frac{1}{h}K(S',O)\sum_{i=1}^{n} A_{j,i}W_i + h\sum_{i=1}^{n} 2w_i P_{j,i}\Gamma_i W_i - \Gamma_j W_j, \qquad j = 1,...,n \tag{2.15}$$

where, W_j represents the cell number density at collocation point j, n is the total number of collocation points. A is the collocation matrix, with w as a vector of the quadrature weights, and h scales each finite element to unity. K is the specific growth rate, depending on intracellular glucose and oxygen concentrations (see Eq. 6). $P_{j,i} = p(m_j, m_i)$ (see Eq. 5), and represents the partition function. $\Gamma_i = \Gamma(m_i)$ (see Eq. 2-4), and is the division intensity function. K, P, Γ, and W all vary with time. For this work, the number of collocation points, n, was 109, with 12 finite elements each with eight internal collocation points (any fewer points/elements lead to numerical errors).

The other ODEs used for the numerical solution represent the glucose and oxygen states in the media surrounding the cells, the calculated surface glucose concentration, and the

internal glucose concentrations, which are given as follows:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = J_S \frac{A}{V_c} - \frac{K(S', O)}{Y} h \sum_{i=1}^{n} w_i W_i \tag{2.16}$$

$$\frac{\mathrm{d}S'}{\mathrm{d}t} = \alpha(S - S') \tag{2.17}$$

$$\frac{\mathrm{d}O}{\mathrm{d}t} = J_O \frac{A}{V_c} - \frac{192}{180} \frac{K(S', O)}{Y} h \sum_{i=1}^{n} w_i W_i \tag{2.18}$$

The overall ODE system therefore consisted of 113 non-linear ODEs. The ODE system was written in C, and integrated into the *ode15s* solver in MATLAB using the MEX interface tool. The ODE system was used to integrate the system forward for each time-step (simulating one hour of process conditions). An outer loop in MATLAB was used to update the conditions given to the ODE system, record simulated measurements, and to implement the sub-batch changes as detailed in the method and meterials section later.

Remark 1 *Beyond the use for process control, development of a predictive model allows for exploratory simulations to be done in the place of physical experiments. Physical experiments for cell culture processes can be costly and time-consuming, while simulations are cheap and fast. A predictive model can therefore be implemented to explore promising experimental conditions or procedures first, and can be used to direct the course of physical experiments to validate those conditions.*

2.3.2 Model parameters

Prior to any rigorous parameter estimation, a manual fitting procedure was performed in order for the initial model to be close to representing the real system. This was done through the collection of a preliminary data set of experimental measurements, taken through the procedure outlined above. This manual fitting process resulted in our base model.

Parameter	Symbol	Units	Value	References
Oxygen diffusivity	J_O	cm^2/s	2.1 x 10^-5	[23]
Glucose diffusivity	J_S	cm^2/s	6.0 x 10^-6	[24]
Vessel surface area	A	cm^2	Varies along batch	Measured
Vessel volume	V	mL	Varies along batch	Measured
Cell layer height ratio	z	–	0.01	[22]

Table 2.1: List of fixed model parameters

The parameters in the model are divided into two groups. The first group consists of those parameters that are considered to be known or estimated to a high degree of certainty through literature. This includes values such as diffusivity of glucose and oxygen in cell culture media, as well as parameters that are dictated by the physical layout of the experiment (media surface area, media volume, glucose content of fresh media, and so on), and are presented in Table 2.1.

The second set of parameters are those for which no independent measurements exist, and need to be estimated for the process under consideration. These are listed in Table 2.2. Amonth these parameters, the biomass yield ratio (or growth yield) is known to be quite variable between cell cultures, different growth conditions, and available substrates. For glucose oxidation, ranges between 0.4 and 0.7 are found in literature. [25] [14] The cell mass values were estimated through available measurements of cell size during growth along with estimated correlations of cell volume to cell mass for similar types of cells [26]. For certain parameters (such as the division shape factor, the partition coefficient), initial values were chosen to create appropriate fittings. Initial estimates of glucose saturation and oxygen saturation coefficients, glucose uptake coefficient, and maximum growth rate were taken from work involving the continuous growth of *S. cerevisae* [14].

Remark 2 *The model structure chosen for this application, although based on certain first principles of cell growth and division, is not the only possible structure. The purpose of the model is to predict the behaviour of the system based on observable conditions, not to*

Parameter	Symbol	Units	Base Value	References
Biomass yield ratio	Y	g/g	0.6	[14] [25]
Minimum cell mass for division	m_t	g	3 x 10^{-12}	[26]
Cell mass of maximum division	m_d	g	8 x 10^{-12}	[26]
Maximum division frequency	γ	–	4	–
Division shape factor	ϵ	–	5	–
Partition coefficient	q	–	40	–
Maximum growth rate	μ_{max}	g/h	0.03x 10^{-12}	[14]
Glucose saturation coefficient	K_S	g/L	0.5	[14]
Oxygen saturation coefficient	K_O	g/L	5 x 10^{-4}	[14]
Glucose uptake constant	α	–	20	[14]

Table 2.2: List of uncertain model parameters

represent exactly what is happening in the biological system itself. We cannot know exactly what is happening beyond using experiments to validate a certain model structure over a sufficiently broad range of conditions. The model structure outlined in this paper is one such attempt.

2.4 Parameter ranking and adaptive estimation

In this section, we first rank the parameters in the model and then illustrate the ability to perform parameter estimation using simulated data. Next, we estimate the model parameters using experimental data. Finally, we present an adaptive parameter estimation approach and illustrate it using the simulation model.

2.4.1 Parameter impact and ranking

The model structure is informed by our current understanding/hypothesis of the various phenomenon that take place during the cell growth process. The description and inclusion of these phenomenon in the model necessitates the introduction of the model parameters. Such a model building exercise, however, does not guarantee that all the parameters can be uniquely identified from the measurements (even if all process variables were continuously measured). The first consideration in the parameter estimation procedure therefore is to evaluate the unique identifiability of the parameters in the model. The identifiability, however, is further influenced (and limited) by the availability of measurements. For the parameter estimation, we therefore perform a sensitivity analysis of the model with respect to the parameters.

For the cell culture growth process (and the data that is available to us) glucose and cell number measurements are available during the batch. In order to determine the sensitivity of the parameters with respect to these available measurements, each uncertain parameter was modified one at a time by $\pm 5\%$ (ensuring that no physical limits in the parameters were violated by this change). The updated model was used to generate new simulated measurements of glucose and cell numbers (reflecting the available physical measurements). The square of the relative difference between the base and updated models in each measurement at each time point was used as the metric for parameter impact.

$$RSSE = \sum_{i=1}^{N} \sum_{k=1}^{2} [\frac{(y_{(ik,base)} - y_{(ik,mod)})}{y_{(ijk,base)}}]^2 \tag{2.19}$$

where i denotes the time-step, N is the total number of time-steps, $y_{ik,base}$ is the measured variable k at time-step i using the base model parameters, $y_{ik,mod}$ is the measured variable k for time-step i given the current modified parameter. The parameter ranking is presented in Table 2.3.

Remark 3 *Parameter selection for non-linear dynamic systems with no steady-state comparisons can be a difficult task. Recent methods for parameter ranking and selection have been*

Parameter	Ranking	RSSE
Biomass yield ratio	1	47.39
Cell mass of maximum division	2	16.59
Maximum specific growth rate	3	12.59
Oxygen saturation coefficient	4	1.75
Glucose saturation coefficient	5	0.518
Maximum division frequency	6	0.038
Division intensity coefficient	7	0.0059
Partition function coefficient	8	1.67 x $10^{-}3$
Glucose uptake coefficient	9	1.82 x $10^{-}4$
Minimum cell mass for division	10	0

Table 2.3: Parameter impact results

proposed [27], but were chosen not to be used for this work due to computational complexity reasons.

2.4.2 Parameter estimation using simulated data

As can be seen from Table 2.3, the impact of the parameters ranked fourth and below drops significantly, suggesting that these parameters cannot be uniquely estimated. Thus the biomass yield ratio, the maximum specific growth rate, and the cell mass required for maximum division rates were chosen for estimation.

For training data, a set of 40 simulated batches was generated, each representing cells from a new patient. The biomass yield ratio (the one that has the highest impact on the observations) was varied across the batches. This parameter was varied following a normal distribution with a mean determined by the base value, and a standard deviation chosen to give a wide variation without violating physical limits (such as avoiding a biomass yield ratio greater than 1). All of the other parameters were constant between each batch, but

Parameter	Base Value	Patient Value
Biomass yield ratio	0.6	0.681
Minimum cell mass for division	3 x 10^{-12}	2.16 x 10^{-12}
Cell mass of maximum division	8 x 10^{-12}	11.28 x 10^{-12}
Maximum division frequency	4	2.98
Division shape factor	5	6.03
Partition coefficient	40	35.64
Maximum specific growth rate	0.3 x 10^{-13}	0.23 x 10^{-13}
Glucose saturation coefficient	0.5	0.456
Oxygen saturation coefficient	5 x 10^{-4}	4.27 x 10^{-4}
Glucose uptake constant	20	25.24

Table 2.4: List of new parameter values for a selected patient

different from the base model and varied in a random fashion, as shown in Table 2.4.

Parameter estimation was performed using this training data through an optimization framework. The decision variables for the optimization were the three parameters chosen in the parameter impact evaluation. The objective function was the minimization of the sum of squares of the relative difference between the model prediction and all 40 training batches in both glucose and cell numbers. Thus, the first set of results demonstrate an attempt to describe the patient behaviours (as best as possible) using a single set of parameters. The optimization formulation is given as follows:

$$\begin{aligned} \underset{\theta_1,\theta_2,\theta_3}{\text{minimize}} \quad & \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{2}\left[\frac{\left(y_{(ijk,base)} - y_{(ik,estim)}\right)}{y_{(ijk,base)}}\right]^2 \\ \text{s.t.} \quad & \text{Eq. 13 - 16} \\ & \text{lb} \leq \theta \leq \text{ub} \end{aligned}$$

where i denotes the time-step, N is the total number of time-steps, M is the total number of

Parameter	Base value	Estimated value
Biomass yield ratio	0.6	0.844
Maximum cell size	8	10.9
Maximum specific growth rate	0.3	0.244

Table 2.5: Parameter estimation results - Simulated training

simulated batches, θ_1 is the biomass yield ratio, θ_2 is the maximum specific growth rate, θ_3 is the size of maximum division frequency, $y_{ijk,base}$ is the measured variable k from simulated batch j at time-step i, $y_{ik,estim}$ is the predicted variable k for time-step i given the current set of parameters. The results of the optimization are shown in Table 2.5.

Figure 2.3 illustrates the measurements from a select set of training data, and from the model after the estimation process. As expected, the model parameters give a 'best fit' value that passes through the observed variables in the training data set. In principle the fit could be improved by allowing the optimizer to pick a different growth ration for each of the patients (thus having forty two decision variables instead of just the three). While such a model would certainly result in a better fit with respect to each patient in the training set, it would not be very useful for a new patient, for which there is no way of forecasting what that particular parameter value is before measurements start being recorded. In contrast, the proposed approach allows for starting with a nominally 'good' model for a new patient with the possibility of improving the model as data is collected.

Remark 4 *The choice of estimating only a subset of the parameters, rather than attempting to estimate all uncertain parameters, revolves around the ability to uniquely identify the parameters. Because of the non-linear and complex nature of the model, the parameter estimation problem is likely non-convex. Similarly, certain parameters may interact or be indistinguishable from others. Therefore, attempting to estimate all parameters is likely to achieve poor, non unique estimates, and thus was not attempted.*

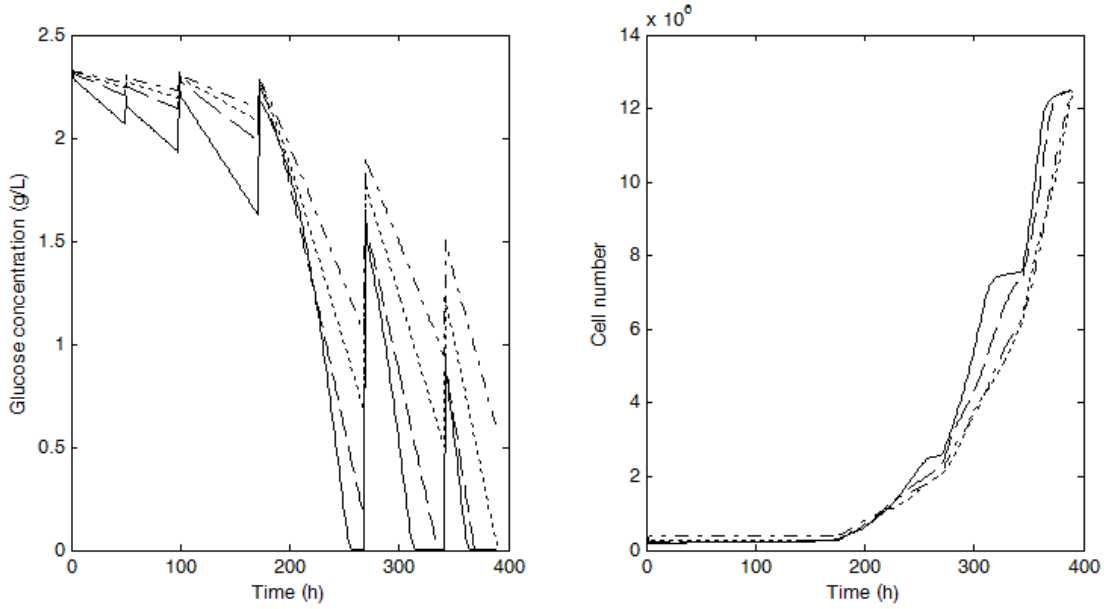


Figure 2.3: Illustrated training data from three of the patients (dashed, dotted, and dot-dash lines) and estimated model (solid black line)

2.4.3 Parameter estimation using experimental data

The parameter estimation approach was next implemented on experimental data. Four separate experimental measurement sets were used to construct the overall data set. These experiments were performed under similar, but not exact conditions, as described next.

Method and materials

Cell preparation

All studies using human subjects received prior approval by McMaster Immunology Research Centre (MIRC). After informed consent, peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using LeucosepTM (Greiner Bio-One International, Monroe, NC) which is prefilled with Ficoll-Paque density centrifugation media. The isolated PBMCs were cryopreserved with 10% DMSO and stored in liquid nitrogen tank for multiple experiments.

The donors were varied in age and race with designated coding (i.e. MAC 001). On day 1, PBMCs were plated onto 96-well plate (100,000 cells) and cultured in T-cell culture media (Life Technology, Grand Island, NY), consisting of RPMI media 1640, 10% FBS, HEPES (1 M), L-Glutamine (0.2 M), Penicillin (100 U/mL)/Streptomycin (100 μg/mL), NEAA, Sodium Pyruvate (100 mM), 2-Mercaptoethanol (55 mM). Cytokine hIL2 and hIL7 (PeproTech, Rocky Hill, NJ) was added right before media exchange in 1:1000 ratio. 100,000 cells were seeded in each 96 multi-well plate with Gibco Dynabeads human T-activator CD3/CD28 (Life technology, Burlington, ON) in a bead-to-cell ratio of 0.8:1. On day 2, the lentivirus for engineering T-cells was added in MOI of 2. On day 3, 100 μL of cell culture media was added to 96 well plate. On day 5, cells were scaled up to 24 well plate with 1 mL of fresh media. On day 8, cells were combined from 24 well and transferred to T-25 with 4 mL of fresh media. On day 10, engineered T-cells were scaled up to T-75 with 10 mL of fresh media. On day 13, the cells were scaled up to a T-150 flask with 10 mL of fresh media.

Glucose and cell measurements

HydrionTM pH papers (Micro Essential Laboratory, Inc, Brooklyn, NY) and Contour® Next EZ glucose meter (Bayer AG, Cleveland, TN) was utilized for measuring glucose concentration and pH in the T-cell culture media. Sample was taken 10 μL (3 μL for glucose strips and 7 μL for pH meter) directly from cell suspensions. Cells were counted using a Cellometer Auto 2000 Cell Viability counter (Nexelcom Bioscience).

Parameter Estimation Results

Note that two of the experiments only had cell number measurements available, while the other two had slightly different concentrations of glucose in the fresh media (measured at 2.34 g/L in the fresh media for the first set, and 2.52 g/L for the second set). One further difference was that the first set had viral transduction performed at 24 hours, while the second experimental set-up did not. This meant that the first experiment had 100 μL of media replaced with fresh media at 24 hours in to the growth process.

The parameter impact ranking was retained from the simulated results. However, during estimation using the parameters shown in Table 2.5, the biomass yield ratio reached the upper bound provided to the optimization problem. As this upper bound was set due to physical limitations, this parameter was instead set to that maximum value and removed from the set of estimated parameters, and the fourth ranked parameter was added to the estimation set. During the estimation process, it was found that equal weighting on both measurements did not provide satisfactory model fitting. Therefore different weightings were performed between the two available measurements in order to improve the model fit. This weighting also reflected the increased confidence in the cell number measurements as compared to the glucose measurements (due to the differences in the experimental methods used). The resultant optimization problem takes the following form:

$$\begin{aligned} \underset{\theta_1,\theta_2,\theta_3}{\text{minimize}} \quad & \sum_{j=1}^{2}\sum_{i=1}^{N_j}\sum_{k=1}^{2} w_k\left[\frac{\left(y_{(ijk,meas)} - y_{(ijk,estim)}\right)}{y_{(ijk,meas)}}\right]^2 \\ \text{s.t.} \quad & \text{Eq. 13 - 16} \\ & \text{lb} \leq \theta \leq \text{ub} \end{aligned}$$

As two types of experimental data were used (as discussed above), two separate constructed models were used in the optimization formula. Here, $j = 1$ and $j = 2$ represent models capturing the different batch/experimental conditions. N_j is the number of available measurements for model j. $y_{(ijk,meas)}$ is the measurement i of measured variable k under model condition j, while $y_{(ijk,estim)}$ is the predicted value of measurement i of measured variable k using model j. w_k represents the weighting factor for each measured variable (1 for glucose measurements, 10 for cell number measurements). The same set of estimated parameters (θ_1,θ_2,θ_3) were used in both models. Figures 2.4 and 2.5 show the results of parameter estimation using an increased weighting on the cell number measurements (weighted by a factor of 10:1).

Remark 5 *The focus of this work is not to develop an entirely predictive and accurate model. Thus it is not meant to describe and capture all the intricacies of cell behaviour. Instead, it is intended to predict the characteristics that are important. The characteristics*

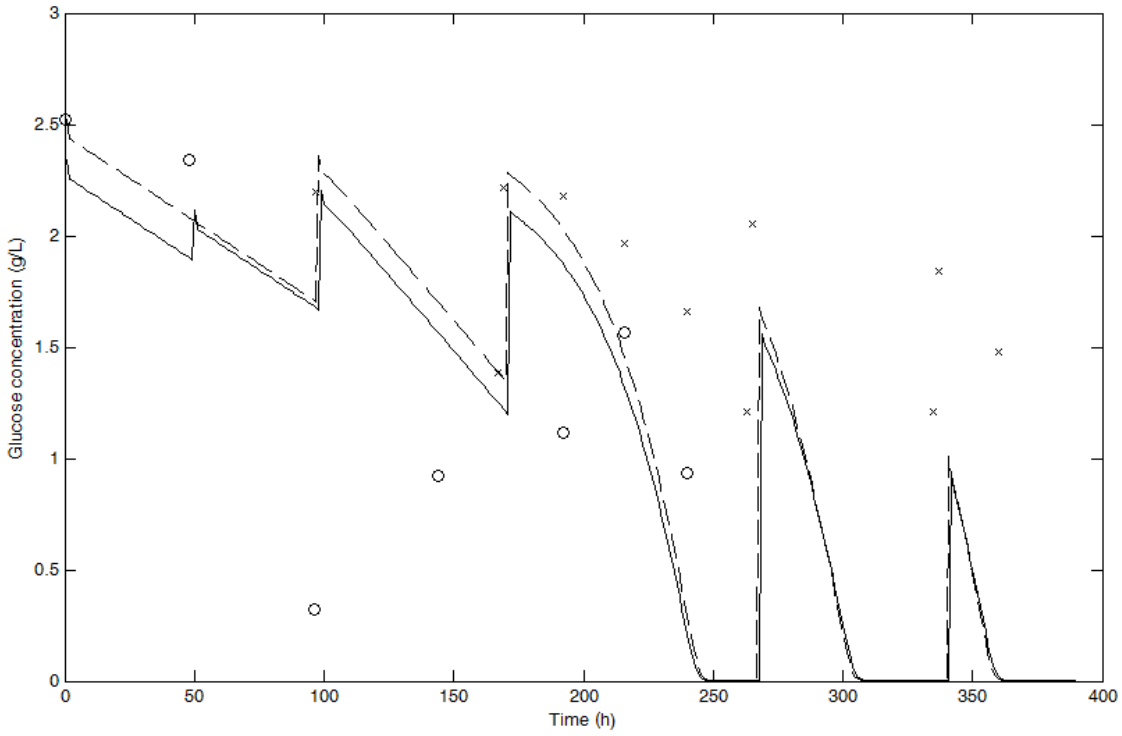


Figure 2.4: Experimental glucose measurements and model predictions. Model 1 prediction (solid line) shown with measurements taken under model 1 conditions (x). Model 2 prediction (dashed line) shown with measurements taken under model 2 conditions (o).

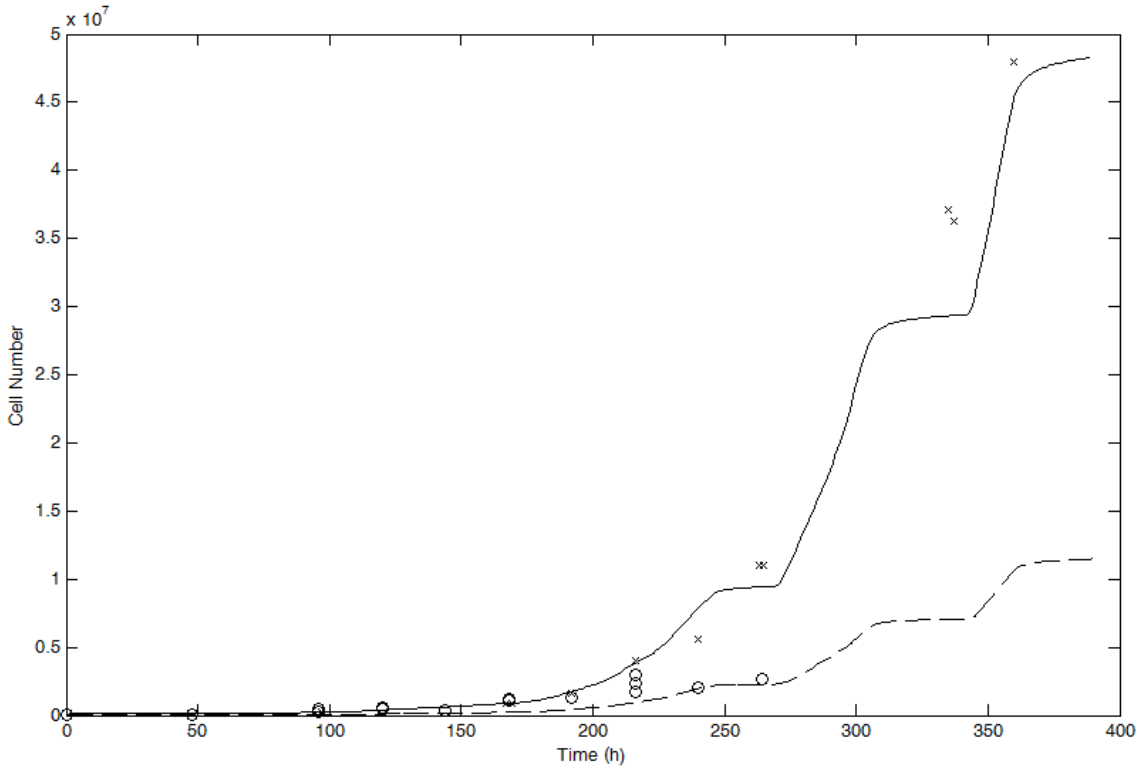


Figure 2.5: Experimental cell number measurements and model predictions. Model 1 prediction (solid line) shown with measurements taken under model 1 conditions (x). Model 2 prediction (dashed line) shown with measurements taken under model 2 conditions (o).

that are important are the ones that can be observed through the available measurements. This is not a limitation of the present approach, but a fundamental limitation of all models. Thus if a cell performs a certain function, and it has no impact on variables that can be currently observed, there exists no mechanism to capture that behaviour.

Remark 6 *It is important to note that the model structure used for this work is in many ways a simplification, but this simplification does not prevent the model from being useful. We could model more detailed elements such as regulatory pathways, glucose uptake limitations, additional substrates such as glutamine, and more. However, without additional measurements to allow us to uniquely identify and estimate the associated parameters, there is little that the model will gain from these additional details.*

Remark 7 *Although cells may seem to exhibit different behaviour at different times, this does not necessarily require a different model. Periods of seemingly low growth rates can be seen in any system that contains elements of exponential growth. A system of the form $\dot{x} = \alpha x$ will have low growth rate initially (when x is very small), and the growth rate will grow as the variable x grows larger. Thus, just the fact that the growth rate is different at different points in time does not imply that separate models are required to capture this behaviour. Additionally, while the cell behaviour could potentially be different under significantly different growth conditions, the focus in this work is to be able to capture the behaviour over a well defined set of conditions.*

Remark 8 *For the current approach, we use a structured model with determined parameters. Future work will focus on utilizing recent results that used subspace based model identification in which process measurements are used for model identification [28], and that have been illustrated for distributed parameter systems [30] and for handling the multi-rate nature of the process measurements [29].*

Remark 9 *The goal of the cell growth process highlighted in this work is to produce T cells that function for cellular therapies (as discussed previously). Therefore, there may be*

important variables surrounding cell functionality besides number and size that this work does not predict. Without more available measurements, however, we cannot predict those variables. More experimental work would need to be done to correlate cell functionality to the available measured variables.

2.5 Adaptive parameter estimation

The previously demonstrated method gives an estimate that essentially serves as the best fit for the average batch. However, biological variation in individual batches of cells from new patients means that this set of estimated parameters may be significantly different than those in a new batch. Therefore, we provide an adaptive estimation process to capture the biological variation and allow for better estimation.

In the simulated environment, 10 batches (simulating cell cultures taken from 10 new patients) were simulated by introducing variation in one parameter (the same parameter as in the training set). In the first implementation, no adaptive method was used, so the estimated parameter set for the model was the same as shown in Table 2.5. In the second implementation, a single adaptive update was performed 168 hours into the batch. Using only the data acquired by measurements up to that point in the batch, the highest impact parameter was re-estimated using the same optimization framework as before. The quality of fit for each process was calculated by the relative sum of squared errors (RSSE) over both measurements. A summary of the results for each patient is shown in Table 2.6. The improvement can be visualized with an example case in Figure 2.6. It can be seen that significant improvements were made in all 10 simulated cases, with the lowest reduction in RSSE being 91.5%. The proposed adaptive estimation step is deemed to be an effective and important addition to the parameter estimation process in order to address the potential variability in new batches.

Adaptive estimation was not performed on the experimental measurement to limited available measurements. Future work will focus on using newly available measurements to per-

Simulated case	RSSE - Non-adaptive	RSSE - Adaptive	Reduction in RSSE (%)
1	48.39	3.340	93.1
2	79.77	4.400	94.5
3	489.5	8.488	98.3
4	7042	17.36	99.8
5	891.9	9.452	98.9
6	1617	10.91	99.3
7	31.33	2.674	91.5
8	732.6	9.099	98.8
9	1178	10.06	99.2
10	1159	10.02	99.1

Table 2.6: Simulated patients - Comparison of non-adaptive to adaptive estimation

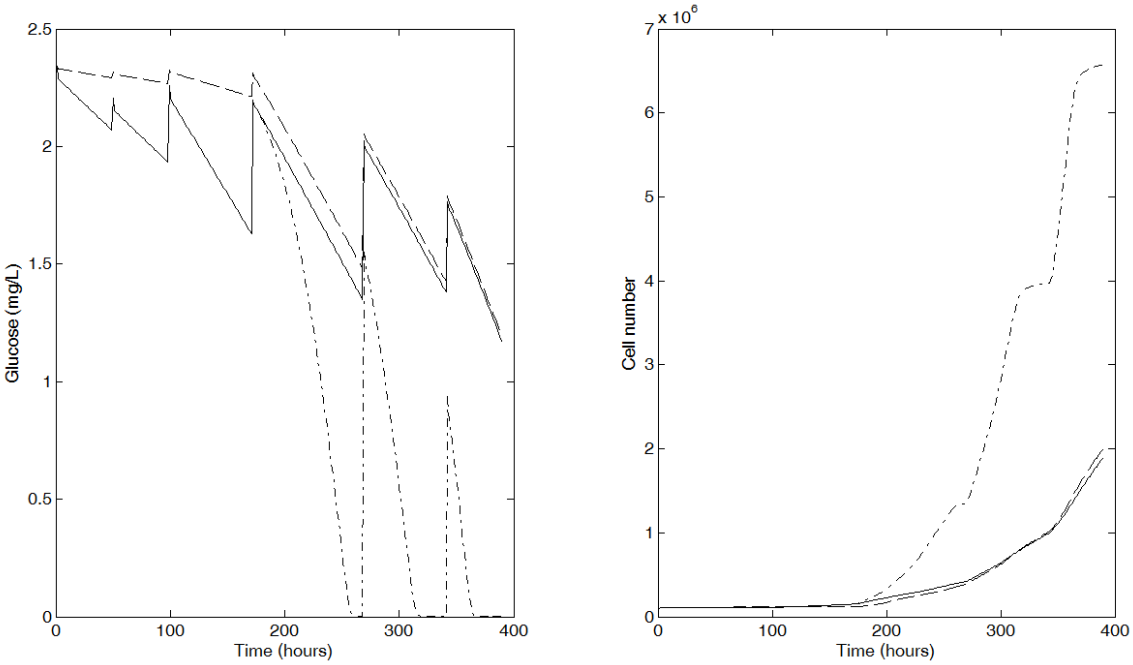


Figure 2.6: Example simulation result using adaptive estimation (Simulated patient (solid line), results using nominal non-adaptive model(dot-dash line), and results using adaptive estimation (dashed line))

form the adaptive estimation process and validate its performance on experimental data.

2.6 Conclusion

A first-principles model of the growth of T cells in culture was developed. The model consists of a population balance model to track the growth and division of cells paired with a 2-phase media model of the external environment. A parameter ranking and estimation procedure was implemented to select and estimate those parameters that can be estimated with the greatest confidence. The effectiveness of the parameter selection and estimation procedure was demonstrated using acquired experimental data. To address and minimize the effect of biological variation on model predictions, an adaptive estimation procedure was developed that added an on-line estimation during the batch process to improve model prediction later in the process. This adaptive estimation process was validated using a simulated set of data that had significant variations in a biological parameter. This illustrates the viability of the proposed parameter estimation process for the development of predictive models that can be subsequently used in process control implementation and design of experiments.

Acknowledgement

The authors thank McMaster Advanced Control Consortium (MACC) and the National Science and Engineering Research Council (NSERC) for financial support.

Bibliography

[1] A. R. McGray, R. Hallett, D. Bernard, S. L. Swift, Z. Zhu, F. Teoderascu, H. VanSeggelen, J. A. Hassell, A. A. Hurwitz, Y. Wan, *et al.*, "Immunotherapy-induced cd8+ t cells instigate immune suppression in the tumor," *Molecular Therapy*, vol. 22, no. 1, pp. 206–218, 2014.

[2] H. VanSeggelen, D. G. Tantalo, A. Afsahi, J. A. Hammill, and J. L. Bramson, "Chimeric antigen receptor–engineered t cells as oncolytic virus carriers," *Molecular therapy oncolytics*, vol. 2, p. 15014, 2015.

[3] A. Trickett and Y. L. Kwan, "T cell stimulation and expansion using anti-cd3/cd28 beads," *Journal of immunological methods*, vol. 275, no. 1, pp. 251–255, 2003.

[4] J. F. Vera, L. J. Brenner, U. Gerdemann, M. C. Ngo, U. Sili, H. Liu, J. Wilson, G. Dotti, H. E. Heslop, and A. M. Leen, "Accelerated production of antigen-specific t-cells for pre-clinical and clinical applications using gas-permeable rapid expansion cultureware (g-rex)," *Journal of immunotherapy (Hagerstown, Md.: 1997)*, vol. 33, no. 3, p. 305, 2010.

[5] P. Painter and A. Marr, "Mathematics of microbial populations," *Annual Reviews in Microbiology*, vol. 22, no. 1, pp. 519–548, 1968.

[6] A. Fredrickson, D. Ramkrishna, and H. Tsuchiya, "Statistics and dynamics of procaryotic cell populations," *Mathematical Biosciences*, vol. 1, no. 3, pp. 327–374, 1967.

[7] G. I. Bell and E. C. Anderson, "Cell growth and division: I. a mathematical model with applications to cell volume distributions in mammalian suspension cultures," *Biophysical journal*, vol. 7, no. 4, p. 329, 1967.

[8] T. B. Gage, F. Williams, and J. B. Horton, "Division synchrony and the dynamics of microbial populations: A size-specific model," *Theoretical population biology*, vol. 26, no. 3, pp. 296–314, 1984.

[9] M. Domach and M. Shuler, "A finite representation model for an asynchronous culture of e. coli," *Biotechnology and bioengineering*, vol. 26, no. 8, pp. 877–884, 1984.

[10] B. C. Batt and D. S. Kompala, "A structured kinetic modeling framework for the dynamics of hybridoma growth and monoclonal antibody production in continuous suspension cultures," *Biotechnology and bioengineering*, vol. 34, no. 4, pp. 515–531, 1989.

[11] R. Pörtner and T. Schäfer, "Modelling hybridoma cell growth and metabolismâĂŤa comparison of selected models and data," *Journal of biotechnology*, vol. 49, no. 1, pp. 119–135, 1996.

[12] M. J. Kurtz, G.-Y. Zhu, A. Zamamiri, M. A. Henson, and M. A. Hjortsø, "Control of oscillating microbial cultures described by population balance models," *Industrial & engineering chemistry research*, vol. 37, no. 10, pp. 4059–4070, 1998.

[13] G.-Y. Zhu, A. Zamamiri, M. A. Henson, and M. A. Hjortsø, "Model predictive control of continuous yeast bioreactors using cell population balance models," *Chemical Engineering Science*, vol. 55, no. 24, pp. 6155–6167, 2000.

[14] P. Mhaskar, M. A. Hjortsø, and M. A. Henson, "Cell population modeling and parameter estimation for continuous cultures of saccharomyces cerevisiae," *Biotechnology progress*, vol. 18, no. 5, pp. 1010–1026, 2002.

[15] N. V. Mantzaris, J.-J. Liou, P. Daoutidis, and F. Srienc, "Numerical solution of a mass structured cell population balance model in an environment of changing substrate concentration," *Journal of Biotechnology*, vol. 71, no. 1, pp. 157–174, 1999.

[16] F. R. Sidoli, S. P. Asprey, and A. Mantalaris, "A coupled single cell-population-balance model for mammalian cell cultures," *Industrial & engineering chemistry research*, vol. 45, no. 16, pp. 5801–5811, 2006.

[17] D. Ramkrishna, *Population balances: Theory and applications to particulate systems in engineering*. Academic press, 2000.

[18] M. Hjortso and J. Nielsen, "A conceptual model of autonomous oscillations in microbial cultures," *Chemical Engineering Science*, vol. 49, no. 8, pp. 1083–1095, 1994.

[19] B. Finlayson, *Nonlinear Analysis in Chemical Engineering*. McGraw-Hill, 1980.

[20] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer, "Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood," *Bioinformatics*, vol. 25, no. 15, pp. 1923–1929, 2009.

[21] R. Li, M. A. Henson, and M. J. Kurtz, "Selection of model parameters for off-line parameter estimation," *IEEE Transactions on control systems technology*, vol. 12, no. 3, pp. 402–412, 2004.

[22] L. Randers-Eichhorn, R. A. Bartlett, D. D. Frey, and G. Rao, "Noninvasive oxygen measurements and mass transfer considerations in tissue culture flasks," *Biotechnology and bioengineering*, vol. 51, no. 4, pp. 466–478, 1996.

[23] A. Richter, K. K. Sanford, and V. J. Evans, "Influence of oxygen and culture media on plating efficiency of some mammalian tissue cells," *Journal of the National Cancer Institute*, vol. 49, no. 6, pp. 1705–1712, 1972.

[24] C. Chung, C. Yang, and C. Chen, "Analysis of cell growth and diffusion in a scaffold for cartilage tissue engineering," *Biotechnology and bioengineering*, vol. 94, no. 6, pp. 1138–1146, 2006.

[25] W. Babel, R. H. Müller, and K. D. Markuske, "Improvement of growth yield of yeast on glucose to the maximum by using an additional energy source," *Archives of microbiology*, vol. 136, no. 3, pp. 203–208, 1983.

[26] W. H. Grover, A. K. Bryan, M. Diez-Silva, S. Suresh, J. M. Higgins, and S. R. Manalis, "Measuring single-cell density," *Proceedings of the National Academy of Sciences*, vol. 108, no. 27, pp. 10992–10996, 2011.

[27] K. A. McLean, S. Wu, and K. B. McAuley, "Mean-squared-error methods for selecting optimal parameter subsets for estimation," *Industrial & Engineering Chemistry Research*, vol. 51, no. 17, pp. 6105–6115, 2012.

[28] B. Corbett and P. Mhaskar, "Subspace identification for data-driven modeling and quality control of batch processes," *AIChE Journal*, 2016.

[29] M. Rashid, P. Mhaskar, and C. Swartz, "Handling multi-rate and missing data in variable duration economic model predictive control of batch processes with application to electric arc furnace operation," *AIChE Journal*, submitted.

[30] G. Abhinav and P. Mhaskar, "Subspace model based predictive control of particulate batch processes," *J. Proc. Cont.*, submitted.

Chapter 3

Modeling and Optimization of Protein PEGylation

Manuscript Overview

In the previous chapter, modeling and parameter estimation for a cell growth process was considered. This work provided a focus on the linking of model development to a parameter estimation procedure. In this chapter, a process of the PEGylation of lysozyme was considered as a test case for protein PEGylation. This work also built a model based on first-principle understanding of the process involved. In this case, the relative simple reaction kinetics model with fewer parameters allowed the estimation of the full set of kinetic parameters for each pH condition tested. A quadratic fit was developed to relate the kinetic parameters to pH. In this way, the developed model could be used for both design of experiments (to suggest future conditions of testing) and for process optimization (to suggest optimal operating conditions to achieve some goal, whether it was reaction time required or selectivity for the desired product). The author of this thesis was not the primary author of the following manuscript, but contributed in the area of modeling, parameter estimation of the reaction kinetics, optimization, and assisted with collection of experimental data for validation of the estimated model.

Modeling and Optimization of Protein PEGylation

Shang Xiaojiao, Brandon Corbett, Brian Macdonald, Prashant Mhaskar, Raja Ghosh

Submitted: August 2016, in *Journal of Industrial and Engineering Chemistry*

Abstract

A PEGylated protein is prepared by conjugating polyethylene glycol (or PEG) with the protein, a process known as PEGylation. Most PEGylation processes lead to synthesis of different PEGylated forms of the protein, amongst which only one form is typically of interest. In this work we propose a modeling and optimization based approach to determining optimal operating conditions for protein PEGylation. To this end, a first principles models is proposed and targeted experiments carried out to estimate the model parameters. A simulation based optimization is then carried out to suggest the best operating conditions for the experiment. Specifically, results suggest that to maximize the concentration of mono-PEGylated product, the reaction should be carried out at high pH and with a high ratio of PEG to protein. Subsequent experiments are conducted to confirm the validity of the modeling and optimization approach.

3.1 Introduction

Protein PEGylation, which involves the covalent attachment of one or more poly (ethylene glycol, or PEG) chains, has become a well-established technology for the covalent modification of biopharmaceutical drugs [1, 2]. The primary advantage of PEGylation is the prolonged in vivo half-life of drug molecules made possible by the increase in hydrodynamic radius. As a result, PEGylated drugs need to be administered less frequently [3, 4, 5]. Other advantages arise from the shielding-effect provided by PEG including reduced enzymatic degradation, lower immunogenicity, and lower protein aggregation [3, 4, 6, 7]. Several PEGylated protein drugs such as AdagenÂő (PEG-adenosine Deaminase, Enzon Pharmaceuticals, approved 1990) [8]; NeulastaÂő (PEG-G-CSF, Amgen, approved 2002) [8] have been approved by the FDA.

The amino acid side chains of a protein are the usual target sites for PEG attachment [10, 11]. Protein PEGylation is usually carried out as a homogeneous liquid phase batch reaction. Due to the nature of the reaction, various PEGylated forms could potentially be synthesized. These forms are differentiated by the number of PEG chains attached to the protein molecule as well as in terms of attachment sites [12]. The mono-PEGylated form, is usually desirable because it retains bioactivity of the protein to the maximum extent while having the desired pharmacodynamic effect [11, 13]. In higher-PEGylated forms it is more likely for the PEG chain to shield the active site of the protein, resulting in deactivation. A high PEG:protein molar ratio is typically used to maximize protein PEGylation. Unfortunately, carrying out the reaction in the presence of excess PEG promotes the formation of higher-PEGylated proteins [14]. The separation of different PEGylated forms of a protein is technically challenging due to the similarities in their physicochemical properties. These separations require extensive procedures and are demanding in time and cost. The selective synthesis of mono-PEGylated form is therefore highly desirable. While the problem of synthesizing mono-PEGylated protein has not been approached from an explicit model-based optimization standpoint, several researchers have tried and maximize the selectivity of synthesis of the mono-PEGylated form through adoption of alternate PEGylation chemistries

and experimental process optimization.

Along one such direction, site-specific protein PEGylation has been extensively pursued, see [15, 16, 1, 17, 18, 19, 20] for specific examples and excellent reviews. This includes N-terminal PEGylation, involving the conjugation of PEG at N-terminal α-amino group [15, 1]; cysteine-specific PEGylation, conjugation at thiol group in either a naturally present or a genetically introduced cysteine residue [15, 1, 16, 17] and other site-specific conjugations at histidine affinity tag or amide group of glutamine [15, 17]. Note however, that even when a site-specific PEGylation (e.g. N-terminal PEGylation) is carried out, significant amount of by-products such as di- and tri-PEGylated protein may be synthesized [18, 19, 20]. Some researchers have also demonstrated physical manipulations for specificity improvement, such as carrying out PEGylation reactions using size exclusion chromatography column [10, 21], packed bed chromatography column [22], membrane stack reactor [23], hollow fiber membrane reactor [24], and microfluidic device [25].

Some of the efforts to study the effect of operating conditions on the PEGylation process include those where the effect of pH of the reaction medium is investigated [26, 27, 28, 29, 30]. Other parameters that have been studied include the initial molar ratio of PEG to protein. Specifically, it has been reported that excess of PEG results in the formation of diverse PEGylated proteins [14, 28, 30]. There also exist results focused on quantifying the kinetics of PEGylation reaction. Some examples include Nojima et al. [26] that quantifies the kinetics of the reaction between branched mPEG-NHS and bovine lactoferrin, focusing on pH dependency. Moosmann et al. [29] evaluated the effect of temperature, reaction time, and protein concentration on conversion and selectivity of PEGylation reactions between two mPEG-aldehyde of different sizes and two proteins (lysozyme and single-chain variable fragment or scFv). Puchkov et al. [30] investigated the effect of pH, buffer system, initial concentrations of protein, PEG and sodium cyanoborohydride (used as a reducing agent) on the yield of conjugation using recombinant granulocyte colony-stimulating factor (Filgrastim) and mPEG-aldehyde.

In summary, while there exist a number of efforts focusing on synthetic chemistry-based

approaches, or relatively simple explorations of the parameters effect on the production of mono-PEGylated protein, the production process stands to benefit from a rigorous modeling and optimization approach to the process that allows a more thorough exploration of the parameter space. Motivated by these considerations, in this work we propose a model based optimization approach for the protein PEGylation process. The rest of the paper is structured as follows: First, a mathematical model structure is derived based on an assumed reaction mechanism and a few key assumptions to reduce the model complexity. In the next section the experimental procedure for PEGylation of the model protein lysozyme (MW, 14,100, pI 11) by methoxy-PEG-$(CH_2)_5$COO-NHS (5kDa PEG equivalent) is described. Note that the choice of the PEG is consistent with what is used in a typical PEGylation reaction on a molecular weight equivalence basis a PEG molecule is significantly bulkier than a protein, so attaching a larger PEG to Lysozyme will completely block out any activity. Targeted experiments are carried out to determine the parameters in the proposed model. In Section 3.4, the parameters of the proposed model structure are identified from the experimental concentration profiles. Subsequently, in Section 3.5, the resulting model is used to determine the optimal operating conditions that maximize the production of the mono-PEGylated form. Additional experiments, presented in Section 3.6, demonstrate the efficacy of the proposed modeling and optimization approach. We finally present some conclusions and directions for future work.

3.2 Model Construction

In this section, we propose a model structure for the dynamic evolution of the PEGylation process. A key decision in proposing a process model is determining which model structure to use. One option is to propose entirely data driven models, where the model structure is predefined, and the model parameters are determined using statistical techniques [31, 32]. Given that there exists some level of understanding of the mechanisms involved in the process, it was decided to propose a first principles model for the process, and determine the parameters from the experiments. In particular, it is understood that the production of

mono-PEG follows a pathway of the following form:

$$L + P \rightarrow M \tag{3.1}$$

$$M + P \rightarrow H_2 \tag{3.2}$$

$$H_2 + P \rightarrow H_3 \tag{3.3}$$

$$\vdots$$

Where L represents lysozyme, P represents mPEG-NHS, M represents mono-PEGylated protein, and H_i represents protein that has been PEGylated i times. With this recognized, the next key decision was the form of the rate law. The parameters of any given rate law can only be determined uniquely if they are observable from the measured outputs. Furthermore, the choice of the rate law dictates the number of parameters in the model. With these factors in mind we opted to begin our study with a relatively simple first order rate law. This rate law may be revised, as necessary, in subsequent work. Additionally, to minimize inclusion of parameters that need to be estimated, we decided to carry out the experiments in constant volume batches.

For the batch configuration (with constant liquid volume) of the process with the chosen rate law, a dynamic moles balance on the individual species yields a model of the following form:

$$\frac{dC_L}{dt} = -k_0 C_L C_P \tag{3.4}$$

$$\frac{dC_P}{dt} = -k_0 C_L C_P - k_1 C_M C_P - \cdots \tag{3.5}$$

$$\frac{dC_M}{dt} = k_0 C_L C_P - k_1 C_M C_P \tag{3.6}$$

$$\frac{dC_{H_2}}{dt} = k_1 C_M C_P - k_2 C_{H_2} C_P \tag{3.7}$$

$$\vdots$$

where k_i is the rate constant for the reaction of protein that has been PEGylated i times; and C_L, C_P, C_M, and C_{H_i} are the concentrations of lysozyme, mPEG-NHS, mono-PEGylated protein, and protein PEGylated i times respectively (where $i \geq 2$). Note that rate constants were taken to be functions of pH with the explicit functional dependence (ie $k_i = k_i\,(\text{pH})$). A

more rigorous model structure might take pH as an additional state (ie define pH dynamics). Experimentally, however, these reactions are usually carried out in a buffered solution. Under buffered conditions, the information needed to determine pH dynamics is unavailable.

We next simplify the model by defining C_H as the concentration of all protein PEGylated more than once. Mathematically this can be represented as:

$$C_H = C_{H_2} + C_{H_3} + C_{H_4} + \cdots \tag{3.8}$$

Given this definition of C_H, we can describe $\frac{dC_H}{dt}$ as:

$$\frac{dC_H}{dt} = \frac{dC_{H_2}}{dt} + \frac{dC_{H_3}}{dt} + \cdots \tag{3.9}$$

By substituting in our original expressions for $\frac{dC_{H_i}}{dt}$ we obtain:

$$\frac{dC_H}{dt} = (k_1 C_M C_P - k_2 C_{H_2} C_P) + (k_2 C_{H_2} C_P - k_3 C_{H_3} C_P) + \cdots \tag{3.10}$$

Which, by term cancellation and without loss of accuracy, simplifies to

$$\frac{dC_H}{dt} = k_1 C_M C_P \tag{3.11}$$

In experimental practice, it is only practicable to quantitatively distinguish between the mono-PEGylated and higher order product. Therefore, we make the assumption that the reaction rate for protein PEGylated more than one time is approximately the same. This simplification is made to retain only the parameters in the model that can be identified with reasonable confidence. Having made this assumption, the equation describing the evolution of mPEG-NHS can be rewritten as:

$$\frac{dC_P}{dt} = -k_0 C_L C_P - k_1 C_M C_P - k_h C_H C_P \tag{3.12}$$

where k_h denotes the lumped reaction rate.

The assumption about the equality of the reaction rates was adopted in this work because it allows the development of a model structure which describes experimental data. As more data is obtained, the model structure could be further refined as appropriate.

Equations 3.4, 3.12, 3.6, and 3.11 represent the simplified dynamic models structure used throughout the remainder of this work. These equations are reproduced below for convenience:

$$\frac{dC_L}{dt} = -k_0 C_L C_P \tag{3.13}$$

$$\frac{dC_P}{dt} = -k_0 C_L C_P - k_1 C_M C_P - k_h C_H C_P \tag{3.14}$$

$$\frac{dC_M}{dt} = k_0 C_L C_P - k_1 C_M C_P \tag{3.15}$$

$$\frac{dC_H}{dt} = k_1 C_M C_P \tag{3.16}$$

This simplified description of the system depends on only three parameters. These are the reaction rate constants: k_0, k_1 and k_h. This model structure describes the evolution of the concentrations of lysozyme, mPEG-NHS, mono-PEGylated protein, and the total concentration of protein PEGylated more than one time.

Note that this model structure does not explicitly account for temperature. It is well known that temperature plays a significant role in the evolution of this process. However, the inclusion of temperature effects in the model would significantly increase the complexity of the identification problem. Furthermore, experimental setup required to capture dynamic temperature behavior is significantly more complex. Moreover, biological conjugation reactions such as protein PEGylation are typically carried out at room temperature. Therefore, in this work the decision was made to assume constant temperature.

Remark 1: The dynamic model presented in equations 3.14 to 3.16 is valid for the batch configuration of the process. As a result, optimization of this model is limited to offline determination of the optimal initial recipe. However, given the first-principals basis of the model, it can be easily extended, without the need for additional parameter estimation, to describe fed-batch or other configurations. These configurations present the problem of determining optimal input trajectories (see, e.g., [33]). Furthermore, for these configurations implementation of online measurement would allow the model to be used in online monitoring and control (see, e.g., [34, 35, 36]).

3.3 Experimental Design and Procedure

In this section, we describe the methodology of the set of experiments carried out to obtain data for modeling the PEGylation process. The first two sections, sections 3.3.1 and 3.3.2, describe the materials and methods used for the PEGylation reaction itself. The subsequent section, section 3.3.3, describes two measurement techniques which were used to qualitatively and quantitatively analyze the resulting PEGylated protein. To collect the data to estimate the model parameters, the following set of experiments were planned (see Table 3.1).

Table 3.1: Summary of all experiments

Buffered pH	Ratio of PEG to protein	Repetitions	Number of sample times
7.0	4:1	1	7
7.5	4:1	3	7
8.0	4:1	3	7
8.5	4:1	1	8
7.0	2:1	1	1
7.5	2:1	2	7
8.0	2:1	1	1
7.0	8:1	1	1
7.5	8:1	1	1
8.0	8:1	1	1

3.3.1 Materials

Ammonium persulfate (A3678), 30% acylamide solution (A3699), bromophenol blue (B0126), Brilliant Blue R concentrate (B8647), glycerol (G2025), 25% glutaraldehyde solution (G6257), glycine (G8898), lysozyme (L6876), sodium dodecyl sulfate (L3771), sodium phosphate monobasic (S0751), sodium phosphate dibasic (S0876), sodium hydroxide (S5881), sodium chloride (S7653), Trizma base (T1503), Trizma-hydrochloride (T3253), N,N,N',N'-tetramethylethylenediamine (T9281), DL-dithiothreitol (43817), 70% perchloric acid (77227),

barium chloride (202738), hydrochloric acid (258148), and iodine (326143) were purchased from Sigma-Aldrich, St. Louis, MO, USA. Acetic acid (1000-1) and methanol (6700-1) were purchased from Caledon Laboratories LTD., Georgetown, ON, Canada. Potassium iodide (74210-140) was purchased from Anachemia, Montreal, QC, Canada. Methoxy-PEG-(CH2)5COO-NHS (5 kDa, catalog number SUNBRIGHT ME-050HS) was purchased from NOF Corporation, Tokyo, Japan. Purified water (18.2 M cm) was obtained from a SIMPLICITY 185 water purification unit (Millipore, Molsheim, France) for preparation of all test and buffer solutions. AmiconÂő Ultra-4 centrifugal filters (3 kDa MWCO, UFC800324) purchased from EMD Millipore Co., Billerica, MA, USA were used for concentrating and desalting samples. Hydrophilized PVDF membrane (0.22 μm, GVWP) purchased from Millipore (Billerica, MA, USA) were used for hydrophobic interaction membrane chromatography of liquid phase PEGylation mixtures.

3.3.2 Reaction Method

PEGylation reactions were carried out using 3 mL reaction mixture consisting of lysozyme (1 mg/mL) and PEG-NHS ester (PEG: lysozyme ratio as listed in Table 3.1), at pH listed in Table 3.1 at room temperature (22$\pm$1°C), with constant stirring. PEG-NHS ester is a commonly used reagent for protein PEGylation. As shown in Figure 3.1, the conjugation is based on stable amide linkages formation through acylation in the pH range of 7 to 9, with the NHS group being replaced by the protein [26, 27]. The N-terminal α-amino group (pKa 7.6-8) and the Îţ-amino group on lysine residues (pKa 9.3-9.5) are potential sites for PEG conjugation.

The reaction media contained 100 mM sodium phosphate and 150 mM sodium chloride. For each pH, seven reaction mixtures were prepared to carry out the reactions for seven different durations (5 min, 15 min, 30 min, 1 h, 2 h, 4 h, and 24 h). The reaction mixtures were quenched by adding hydroxylamide hydrochloride solution to make its final concentration of 50 mM. All quenched reaction mixtures were processed using centrifugal ultra-filters to remove low molecular weight species and to enhance concentration. The resultant re-

action mixtures were analyzed using SDS-PAGE and hydrophobic interaction membrane chromatography (HIMC) [19]. The experiments for pH 7.5 and 8 were repeated three times. Refer to Table 3.1 for a complete list of experimental conditions.

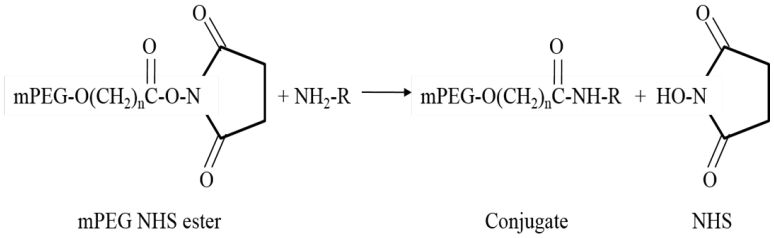


Figure 3.1: PEGylation reaction using PEG NHS esters.

Remark 2: In this study, methoxy-PEG-$(CH_2)_5$COO-NHS (5kDa PEG equivalent) was used to PEGylate the model protein - lysozyme (MW, 14,100, pI 11) in liquid phase batch reaction. NHS ester reactive group is known to be very susceptible to hydrolysis [27]. By increasing the distance between this group and the PEG group, its hydrolysis half-life could be improved dramatically [27]. For example, a PEG-NHS ester with three ether groups in between has a hydrolysis half-life of 23 h, while that of a PEG-NHS ester with only one ether group in between is only 0.75 h [37]. Therefore, the PEG-NHS ester used in this work is inferred to have a hydrolysis half-life substantially longer than 23 h.

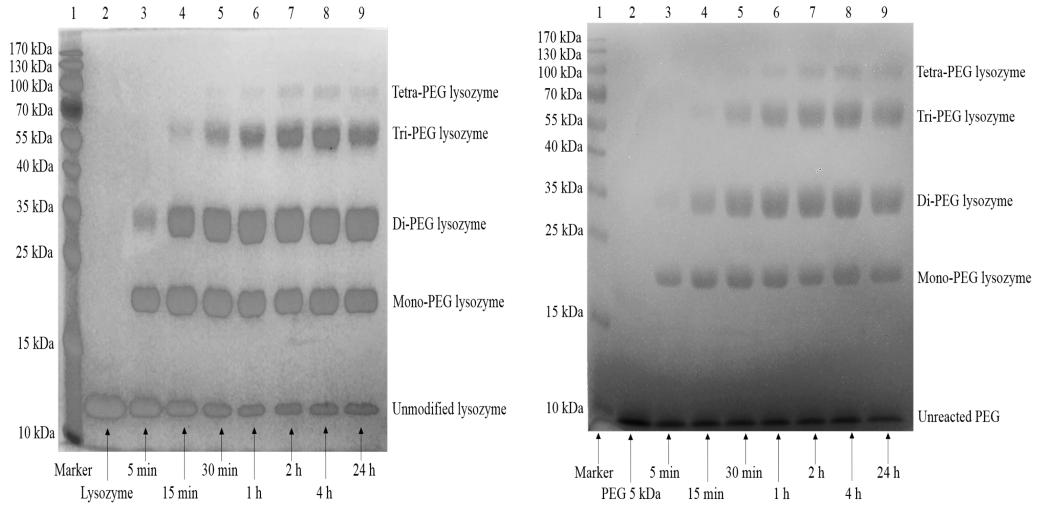
3.3.3 Analytical Techniques

In this section two independent measurement techniques are discussed. First, in section 3.3.3, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is described. SDS-PAGE is an electrophoresis method for qualitatively verifying the PEGylated forms resulting in the reaction mixture. In section 3.3.3, hydrophobic interaction membrane chromatography (HIMC) method is presented which provides quantitative analysis of the reacted protein. Particularly, this method provides partial concentration measurements.

SDS-PAGE

To verify the composition in the reaction mixtures, SDS-PAGE experiments [38] were carried out with an equal amount of total protein ($\sim$6 μg) loaded onto lanes of 12.5% non-reducing gels, using a miniVE vertical electrophoresis system (80-6418-77) purchased from GE Healthcare Life Sciences, Baie d'Urfe, QC, Canada. Coomassie Brilliant Blue R dye was used to stain the gel for visualization of protein-containing bands. The visualization of PEG-containing bands was made by the following protocol [39]: the gel was sequentially soaked in 50 mL of 5% glutaraldehyde solution, then 20 mL of 0.1 M perchloric acid, and finally a mixture of 5 mL 5% barium chloride and 2 mL 0.1 M iodine/potassium iodide. Each soak took 15 min.

Figures 3.2a and 3.2b show two example gels (generated using SDS-PAGE to identify the components) for the reaction mixture prepared at pH 8 and PEG: lysozyme molar ratio of 4:1. Both figures have protein molecular weight marker in lane 1 and reaction mixtures with different durations of 5 min, 15 min, 30 min, 1 h, 2 h, 4 h, and 24 h in lanes 3 to 9, respectively. Lane 2 consisted of pure lysozyme and pure PEG-NHS ester 5 kDa in figures 2a and 2b, respectively. Lane 3 (containing the 5-min reaction mixture) has mainly mono-PEG lysozyme (band between 15 and 25 kDa) and a very small amount of di-PEGylated lysozyme (band between 25 and 35 kDa) as shown in both figures. There are also unmodified lysozyme (band corresponding to the pure lysozyme band in lane 2) as shown in figure 2a and quenched PEG 5 kDa (band around 10 kDa) as shown in figure 2b. The amount of di-PEG form was increased greatly from the 5-min reaction (lane 3) to the 15-min reaction (lane 4), which can be seen from the intensity difference of the di-PEG form bands in both lanes. Tri-PEG form (band between 55 and 70 kDa) started to show up in lane 4 (15-min reaction). Lanes 5 to 9 contained faint bands of tetra-PEG lysozyme (band between 70 and 100 kDa) besides the bands of unmodified, mono-, di- and tri-PEG forms of lysozyme as well as quenched PEG. These two gels are clearly showing the components, the complexity of the components and their relative quantities in each reaction mixture. Generally speaking, longer duration results in more complex reaction mixtures with predominance of higher PEGylated forms of protein.



(a) Coomassie blue stained gel for liquid-phase PEGylation carried out at pH 8 and molar ratio of PEG: lysozyme of 4:1.

(b) PEG stained gel for liquid-phase PEGylation carried out at pH 8 and molar ratio of PEG: lysozyme of 4:1 at various durations.

Figure 3.2: SDS-PAGE for a representative reaction condition

Fractionation and analysis of PEGylation reaction mixtures using HIMC

After centrifugation, each PEGylation reaction mixture was fractionated using HIMC [23] with a stack of hydrophilized PVDF membranes. The working principle of HIMC is described in [19]. The membranes are environment-responsive, becoming hydrophobic when lyotropic salt is present and reverting back to hydrophilic when salt is removed. PEG is hydrophilic in nature, but becomes mildly hydrophobic in the presence of lyotropic salt as well. Therefore, a PEG-containing species has hydrophobic interaction with the environment-responsive membrane when a lyotropic salt is present. The fractionation of different species in a reaction mixture (i.e. unmodified protein, quenched PEG, mono-, di-, tri-, tetra- and/or high-PEGylated protein) depends on the difference in their apparent hydrophobicity in presence of salt. A greater number of PEG chains attached to a protein results in a higher apparent hydrophobicity of the protein, i.e. unmodified protein with no PEG attached is the least hydrophobic form, and followed by mono-, di-, tri-, tetra- and/or high-PEGylated forms. As described in the paper [19], during a HIMC experiment, unmodified lysozyme flow through

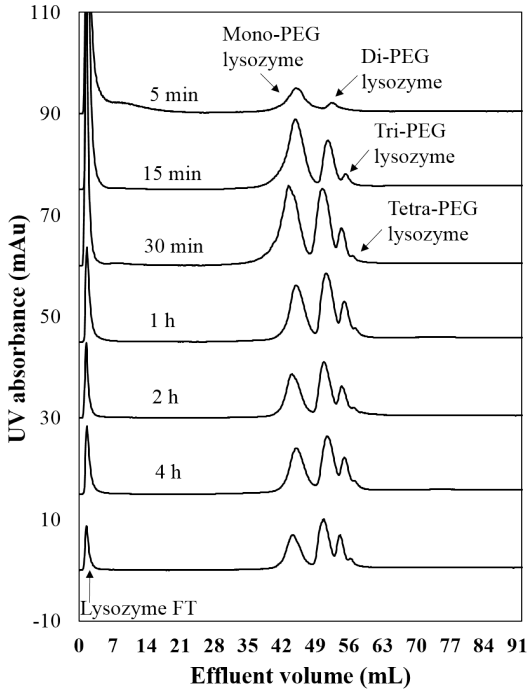


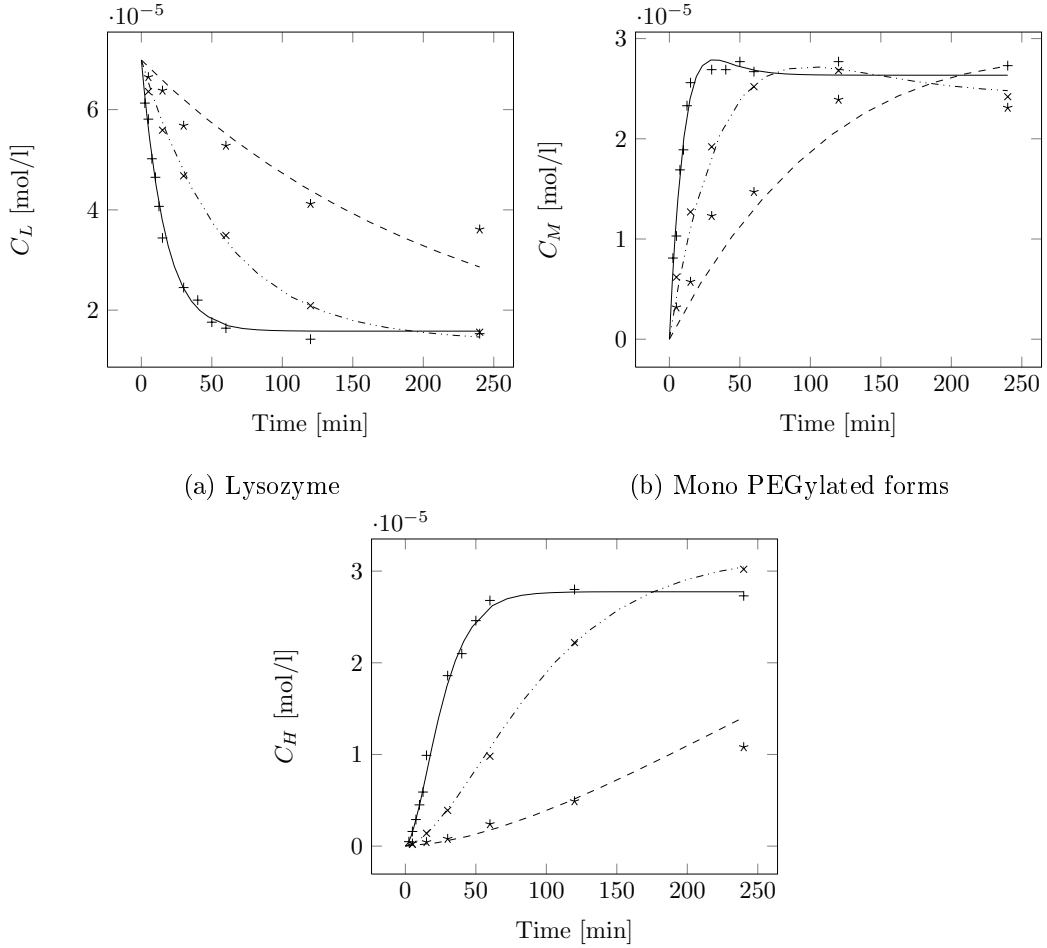
Figure 3.3: A sequence of HIMC chromatographs for a representative PEGylation batch

the membrane stack without any interaction under the operating condition, whereas all PEGylated forms are bound on the membranes. The bound species are eluted out in order of increasing apparent hydrophobicity by lowering the salt concentration in a gradient manner. Specifically, the mono-PEG form is eluted first, followed by di-, tri-, and tetra-PEG forms if present.

In this work, the eluting buffer contained 100 mM sodium phosphate at pH 7. The binding buffer with an ammonium sulfate concentration of 1.4 M was prepared in eluting buffer and adjusted to pH 7. A membrane module having 15 membrane discs with a diameter of 18 mm stacked inside was integrated with an AKTA Prime liquid chromatography system (GE Healthcare Bio-Sciences, QC, Canada). The separation was executed at 1 mL/min with 500 μL of feed sample containing about 250 μg of total protein loaded. A 50 mL gradient from 0% to 100% of eluting buffer was used for elution of bound species. Figure 3.3 shows a sequence of chromatograph results for one batch condition.

From the chromatographs generated by HIMC, the composition of each mixture was quanti-

fied by software, PrimeView, based on the integration of individual chromatographic peaks. Because each peak corresponds to one component present in the mixture, the relative mole percentage of each species can be calculated. Then, by mole balance, concentration for each species in each sample can be determined. By applying this for the batches quenched at different times, time-series concentration data was obtained for each lysozyme species. Figure 3.4 shows the resulting concentration trajectories.



(a) Lysozyme

(b) Mono PEGylated forms

(c) Higher order PEGylated forms

Figure 3.4: Modeled concentration trajectories (pH 7.0: dashed line, pH 7.5: dot dashed line, pH 8.0: solid line) compared to experimental data (pH 7.0: star, pH 7.5: x, pH 8.0: plus)

Remark 3: The results of HIMC can be verified to some extent by a comparative analysis with the SDS-page results. Particularly, the width of the bands in SDS-page corresponds in some sense to the relative concentration of species. For instance, unmodified lysozyme bands were reduced from 5 minute to 30 minute reactions, meaning that significant amount of unmodified lysozyme was being consumed from 5 minutes to 30 minutes. Such comparative analysis universally confirmed the results obtained by HIMC.

3.4 Parameter Estimation

Having obtained concentration evolution with time, least squares estimation was applied to determine the model coefficients from the model structure proposed in section 2. To estimate the parameters the following dynamic optimization problem was solved:

$$\min_{k_0,k_1,k_h} \sum_{i=0}^{n} \left(C_M(t_i) - \hat{C}_M(t_i)\right)^2 + \left(C_H(t_i) - \hat{C}_H(t_i)\right)^2 \tag{3.17}$$

$$s.t. \quad \frac{d\hat{C}_L}{dt} = -k_0 \hat{C}_L \hat{C}_P \tag{3.18}$$

$$\frac{d\hat{C}_P}{dt} = -k_0 \hat{C}_L \hat{C}_P - k_1 \hat{C}_M \hat{C}_P - k_h \hat{C}_H \hat{C}_P \tag{3.19}$$

$$\frac{d\hat{C}_M}{dt} = k_0 \hat{C}_L \hat{C}_P - k_1 \hat{C}_M \hat{C}_P \tag{3.20}$$

$$\frac{d\hat{C}_H}{dt} = k_1 \hat{C}_M \hat{C}_P \tag{3.21}$$

$$\hat{C}_P(t_0) = C_P(t_0) \tag{3.22}$$

$$\hat{C}_M(t_0) = C_M(t_0) \tag{3.23}$$

$$\hat{C}_L(t_0) = C_L(t_0) \tag{3.24}$$

$$\hat{C}_H(t_0) = C_H(t_0) \tag{3.25}$$

Where t_i denote the times at which measurements of C_M and C_H are available experimentally. In essence, the optimization problem computes the 'best-fit' values of the parameters (k_0, k_1, k_h) that enable matching the predicted values of C_M and C_H with the measured values. Note that the initial concentrations of all the species was known (based on the initial ration chosen for a particular run), and thus does not need to be estimated. The optimiza-

tion problem was solved using MATLAB's NLP solver `fmincon`. Note that the effect of the molar ratio on the model is accounted for explicitly by the initial concentrations $C_L(t_0)$ and $C_P(t_0)$. Furthermore, it is known that there are no PEGylated forms initially, which allows setting their initial concentrations to zero in the above optimization. However, pH is not explicitly accounted for. Therefore, the optimization problem was solved independently for each pH using only the experimental data collected at that pH (i.e. the problem was solved independently for data taken at pH 7.0, 7.5, and 8.0). By this process, different kinetic parameters were obtained for each pH value in the data set. A nonlinear function was then fit relating each kinetic parameter to pH. To this end, quadratic models were used. These quadratic models were fit to the values of the parameters at pH 7.0, 7.5, and 8.0. In this way the overall system behavior was related to the buffer pH.

Remark 4: The intent of the present contribution is not to develop 'the' perfect model for the PEGylation process, but demonstrate the possibility of using a dynamic model, estimating parameters, and utilizing the model subsequently to optimizing operating conditions. The data set used for the training, thus, does not contains repeats for all operating conditions (although it does for some of the conditions), and does not report parameter estimates with error bars. The key is that parameter estimation using this data set still yields useful information about the process dynamics, and suggests meaningful operating condition that yields an improvement. Future work will focus on using this approach in a variety of experimental set up to obtain a richer data set, and also perform optimization not just over operating conditions but also experimental configuration.

Remark 5: Note that the quadratic fit through three points used in this work exactly passes through each point. While this approach likely leads to over-fit, the three data points were not well explained by a linear relationship. The proposed approach, using quadratic relationships, provides an interpolation method between the experimental observations. With sufficient additional measurements a more rigorous approach to identifying the pH dependency (similar to that in [40]) would be possible.

Figure 3.4 shows the model fit compared to experimental data. This figure illustrates the

ability of the proposed model structure to explain the trends in the data. Note that the proposed model structure does not include a backward reaction. Even in the absence of backward reaction, the forward reaction rate, for instance for the lysozyme molecule, depends not only on the presence of lysozyme, but also the mPEG-NHS concentration. If there was abundance of mPEG-NHS (or atleast that mPEG-NHS was not the limiting reagent), then we expect the concentration of L to approach zero. However, if mPEG-NHS is the limiting reagent, the concentrations of neither L no M will approach zero. This is found to happen in the experiments, and is consistent with the proposed model structure. The model parameters obtained from least squares regression for each pH are shown in Table 3.2.

Table 3.2: Parameters for data collected at different pH values

pH	k_0	k_1	k_h
Units	L mol^{-1} s^{-1}	L mol^{-1} s^{-1}	L mol^{-1} s^{-1}
7.0	15.08	12.72	0.9958
7.5	49.87	44.91	432.4
8.0	186.0	194.3	3123

Remark 6: Insight about the effect of pH on the reaction can be drawn by comparing the rate coefficients, k_0, k_1, and k_h, for different pHs. The clear trend is that the coefficients increase for increasing pH. This indicates that the overall rate of the reaction increases as pH is increased. Note that k_0 and k_1 are similar in magnitude and increase with a similar trend. Recall that k_0 and k_1 correspond to the rate constants for the reaction to the mono and di-PEGylated forms respectively. The similar increasing trend and similar magnitudes of k_0 and k_1 indicate that the pH has minimal impact on the relative reaction rates to the mono and higher PEGylated forms.

Remark 7: It is believed that the nucleophilic attack involved in this type of PEGylation takes place when the operating pH is near or higher than the pKa value of the amino acid group [27]. As a result, PEGylation at the α-amino group should be favored at a pH value near or above the pKa of α-amino group and below the pKa of ϵ-amino group, and mainly

mono-PEGylated protein should be synthesized. A higher pH than the pKa of ϵ-amino group should in turn facilitate conjugation at the abundant ϵ-amino groups, resulting in the synthesis of the higher-PEGylated forms. Thus, in the proposed model, parameter values for k1 should increase much faster than k_0 with increasing pH. The best fit values of the parameters, however, do not follow this trend, neither do the experimental results. There could be two reasons for this mismatch. The first is because of the simplifying assumptions made in the model, or insufficient experimental data, due to which the parameters estimated from the model do not reflect the reaction mechanism in sufficient detail. The other reason would point toward a reevaluation of the current understanding of the reaction mechanism of protein PEGylation. Both these avenues will be pursued as future directions of research.

Remark 8: PEG reagents are usually designed to be mono-functional, i.e. one PEG molecule can be attached to only one protein molecule, whereas one protein molecule can react with several PEG chains at its various conjugation sites. Therefore, it could be presumed that an excess of protein would increase the specificity of mono-PEGylated form but at the cost of low conversion and wastage of expansive raw materials. However, the experimental data, and the proposed model suggest a different underlying physical phenomenon. Specifically, experimental results suggest that by increasing the concentration of PEG, the reaction rates for all reactions are proportionally increased. This is demonstrated in Figure 3.5 which shows two reactions carried out at pH 7.5 where one experiment is run with a ratio of 4:1 PEG to lysozyme and one with 2:1. Note that the 2:1 curve looks like a stretched version of the 4:1 curve. This observation is reflected in the modeling assumption that the reactions are first order. Mathematically this means that increasing the concentration of PEG would proportionally accelerate all reactions (i.e. to higher order PEGylated forms as well) regardless of the concentration of protein. Figure 3.5 also plots the model predictions. Note that both of the model prediction trajectories plotted in Figure 3.5 are based on a model built solely from the plotted 4:1 data. The predictive ability of this model for the 2:1 case strongly suggests that the first-order assumption is valid.

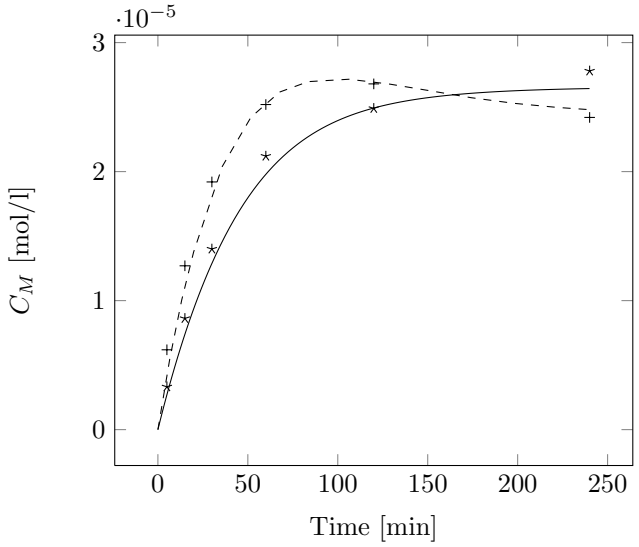


Figure 3.5: Comparison of reactions carried out with a 2:1 ratio of PEG to lysozyme (experimental data: stars, model predictions: solid line) and a 4:1 ratio of PEG to lysozyme (experimental data: dashed line, model prediction: plus)

Remark 9: Lysozyme is widely used as a model protein as its structure and properties have been extensively studied. It consists of a single chain, potential PEGylation sites are known, and it is relatively easy to analyse. It is therefore ideally suited for the present study which focuses on reaction engineering aspects of PEGylation. Note also that while it would be interesting to test for biological activity of mono-PEGylated lysozyme, such analysis would be better suited in a paper that focuses on clinical, biochemical and biophysical fundamentals regarding PEGylated proteins, and as such is outside the scope of the present work. Also, note that larger PEGs are attached to some therapeutic proteins. However, these molecules are also significantly bigger than lysozyme. The 1:3 peg:protein MW ratio was intentionally chosen to represent the typical ratio used in PEGylation.

3.5 Optimization

In this section we discuss the optimization of the operating conditions for batch PEGylation. In this work, the operating conditions were defined in terms of the buffer pH and the initial ratio of PEG to lysozyme. Correspondingly, these two parameters were the decision variables in our optimization problem. In this work optimization was performed using a grid search method. In this method, the decision variable space is discretized into a sufficiently fine mesh and the objective function is evaluated for each discrete point. The optimal is then arrived at by selecting the maximum (or minimum) corresponding objective value. The objective function was defined to maximize the concentration of the mono-PEGylated form. Note that it is much easier to separate reactants from mono-PEGylated protein compared to byproducts like di-PEGylated protein. Thus, terminating the batch at the peak of the mono-PEGylated protein concentration is beneficial from the downstream processing standpoint, and allows the possibility of recycling the unreacted reagents. This objective function was selected based on the assumption that the goal of the process is to generate the highest yield of mono-PEGylated product (which is synonymous with maximizing the concentration for a constant volume batch reactor where the initial concentration of protein is unchanged from batch to batch).

Figure 3.6 shows the objective function for the grid-search optimization carried out over a pH range of 7.0 to 9.0 and an initial ratio (PEG: lysozyme) range of 2:1 to 8:1. A greater than 1:1 ratio is essential to ensure adequate pegylation. 2:1 represents a low ratio and 8:1 represents a high ratio, thus the range includes all the relevant conditions. This figure clearly demonstrates the results of the optimization. In particular, it is apparent that maximum concentrations of mono-PEGylated lysozyme can be obtained by operating at high pH and a high ratio of PEG to lysozyme.

Remark 10: In addition to determining the optimal operating conditions, the optimization also predicts the optimal batch duration. Figure 3.7 shows the optimal batch durations for the range of operating conditions shown in figure 6. Note that the proposed operating conditions (high pH and high ratio PEG:lysozyme) also has the lowest optimal operating

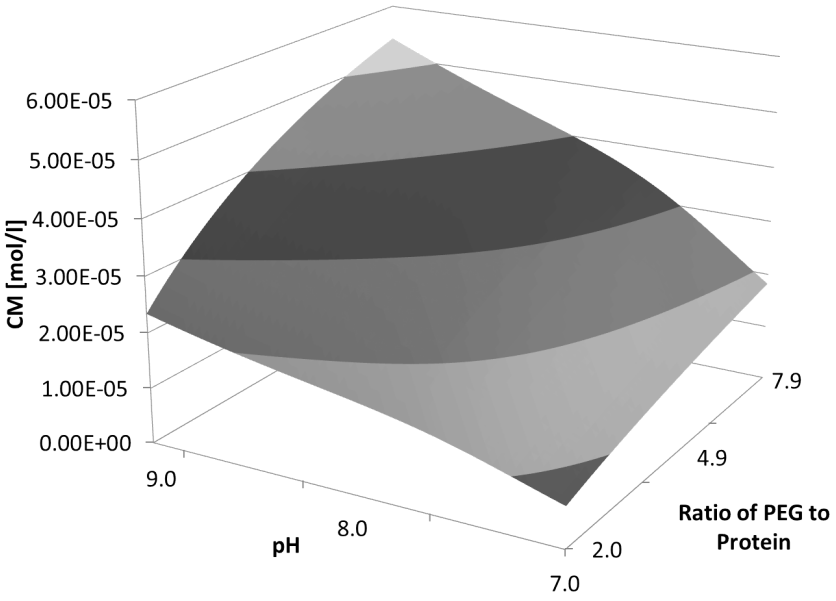


Figure 3.6: Model prediction of maximum concentration of mono-PEGylated product for a range of pH and ratio of PEG to lysozyme

duration. The yield reaches a maximum before it decreases. This is not a mathematical artifact. It is observed in the experiments (although not very pronounced; see Figure 3.4b and Figure 3.8b) and is consistent with the model structure. The mono-PEGylated protein will keep reacting and form higher PEGylated forms if left to react. A model (such as the one developed in this study) which captures this phenomenon can be utilized in determining not just optimum batch times, but also optimum operating conditions for other reactor configurations.

Remark 11: Note that after a model has been developed, the optimization criteria can be readily altered to investigate different requirements. Specifically, an alternate objective function could be to maximize the selectivity for mono-PEGylated product. The choice of this alternate objective function would be motivated by a desire to reduce downstream separation costs associated with separating mono-PEGylated from higher order PEGylated forms. However, an intuitive analysis of the process suggests that using the selectivity objective would complicate the optimization problem significantly. This is because it is

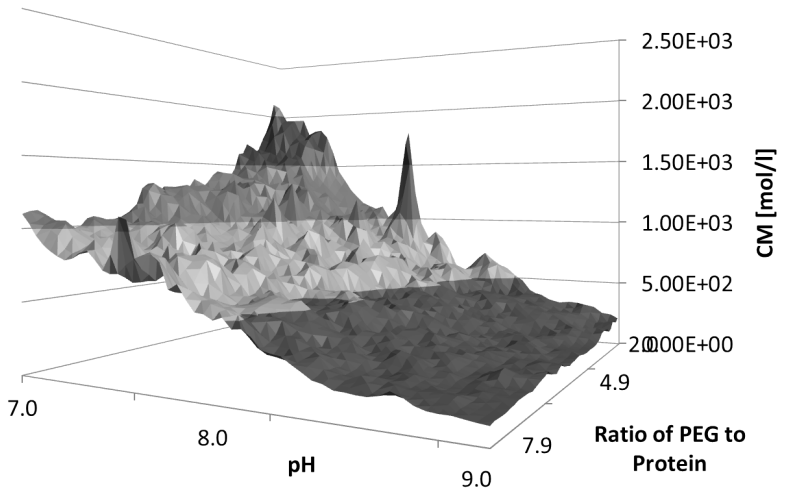


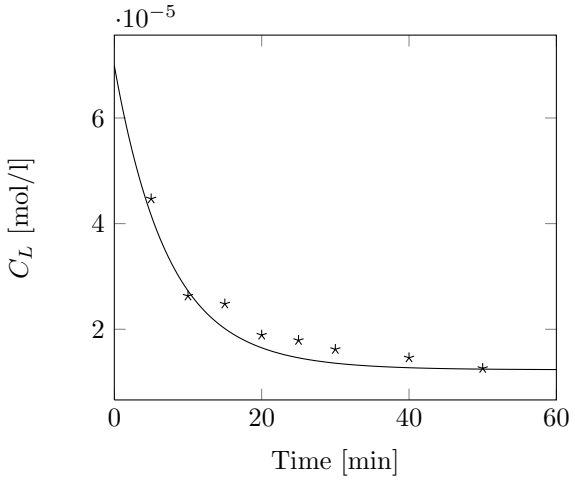
Figure 3.7: Model prediction of maximum concentration of mono-PEGylated product for a range of pH and ratio of PEG to lysozyme

possible (and in fact probable) that the highest selectivity would occur near the beginning of the batch before mono-PEGylated product reacts to higher order PEGylated products. In this case, the yield of mono-PEGylated product could be very low. In practice then, using selectivity as an objective function would require at least a hard constraint on minimum yield. Further research will focus on formulating a dual objective function considering both yield and selectivity, weighted by a factor determined by the costs of downstream operations.

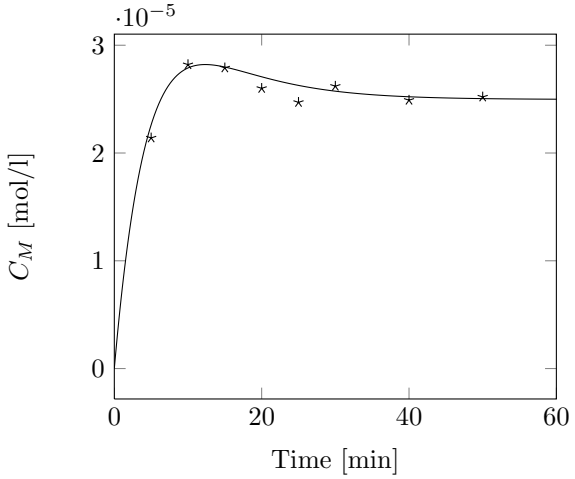
3.6 Validation

To validate these results (as well as the original model parameters and structure), a new experimental batch was run at pH 8.5 and molar ratio of 4:1. Recognizing that excessively high pH values would lead to deleterious effects such as degradation, we avoided the higher pH value suggested by the optimization solution. The concentration of mono PEGylated protein for this batch is plotted in Figure 3.8, as well as the model predictions (based on

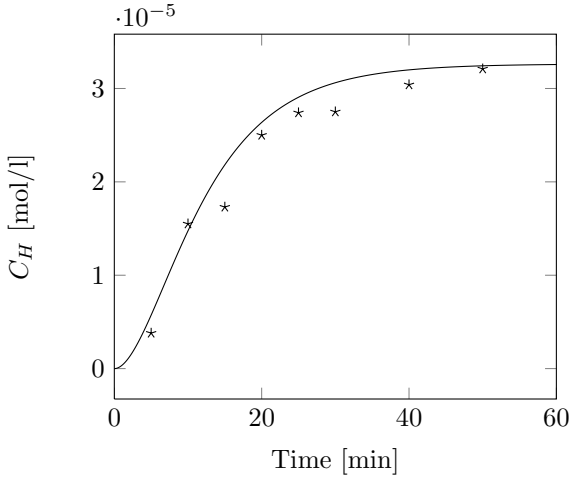
training data not including the new experiment). This figure demonstrates the predictive ability of the proposed model. Moreover, the experimental observation at 10 minutes (the projected optimal batch length) for the new batch demonstrates one of the highest concentrations of mono-PEGylated lysozyme over all the experiments conducted. Note that the model development is not aimed at developing a 'perfect' model, but rather a model that captures the qualitative behavior of the process, and can be used in an adaptive framework, to inform further experiments, and enable optimal operation of the process. A graphical representation of the prediction is therefore presented in the validation step instead of quantifying the error in the prediction.



(a) Lysozyme



(b) Mono-PEGylated product



(c) Protein PEGylated two or more times

Figure 3.8: Model predictions (solid line) and experimental measurements (stars) for a batch run at pH 8.5 and a ratio of 4:1, PEG to lysozyme.

3.7 Conclusions and Future Work

In this study, PEGylation reaction kinetics were investigated. First a model form was proposed based on first-principals and a few key assumptions. Model parameters were identified from experimental data. To obtain this data, batch PEGylation reactions were carried out for a range of pH and initial concentration conditions. Concentration measurements for the resulting batch products were obtained by hydrophobic interactive membrane chromatography. By repeating the reactions with differing durations, concentration trajectories were obtained. These trajectories were used to find the model parameters using the pre-defined model structure and least squares. The model was then applied in an optimization framework to determine the optimal operating conditions for the PEGylation process. Optimization results suggest that to maximize the concentration of mono-PEGylated product, the reaction should be carried out at high pH and with a high ratio of PEG to protein. Both the model and optimization results were validated by carrying out an additional experiment at pH 8.5. This experiment demonstrated both the ability of the model to predict the reaction outcome and the validity of the optimization result.

Optimization of the initial recipe and pH provide a starting point for the optimal PEGylation of protein. However, the inevitable reaction of mono-PEGylated product to undesired higher-PEGylated species provides motivation to investigate other reactor configurations. One of the benefits of this work is that the developed model is not restricted to application in the batch arrangement. Specifically, it is possible to use the existing model and parameters to make prediction for continuous and semi-continuous reactor arrangements with the hope that the yield of mono-PEGylated product can be improved.

Acknowledgement

We thank Natural Science and Engineering Research Council (NSERC) of Canada for funding this study. Paul Gatt for the Department of Chemical Engineering, McMaster University is thanked for fabricating the membrane modules used in this study. Raja Ghosh holds the

Canada Research Chair in Bioseparations Engineering. Prashant Mhaskar holds the Canada Research Chair in Nonlinear and Fault-Tolerant Control.

Bibliography

[1] Veronese F. M, Mero A. The impact of PEGylation on biological therapies *Biodrugs.* 2008;22:315-329.

[2] Ikeda Y, Nagasaki Y. PEGylation Technology in Nanomedicine in *Polymers in Nanomedicine* (Kunugi S, Yamaoka T. , eds.);247 of *Advances in Polymer Science*:115-140 2012.

[3] Abuchowski A, Mccoy J, Palczuk N, Vanes T, Davis F. Effect of Covalent Attachment of Polyethylene-glycol on Immunogenicity and Circulating Life of Bovine Liver catalase *Journal Of Biological Chemistry.* 1977;252:3582-3586.

[4] An Q, Lei Y, Jia N, et al. Effect of site-directed PEGylation of trichosanthin on its biological activity, immunogenicity, and pharmacokinetics *Biomolecular Engineering.* 2007;24:643-649.

[5] Bailon P, Fung W, Porter J. Strategies for the preparation and characterization of polyethylene glycol (PEG) conjugated pharmaceutical proteins. *Abstracts of Papers of the American Chemical Society.* 1997;213:183-POLY.

[6] Basu A, Yang K, Wang M, et al. Structure-function engineering of interferon-beta-1b for improving stability, solubility, potency, immunogenicity, and pharmacokinetic properties by site-selective mono-PEGylation *Bioconjugate Chemistry.* 2006;17:618-630.

[7] Hinds K, Koh J, Joss L, Liu F, Baudys M, Kim S. Synthesis and characterization of poly(ethylene glycol)-insulin conjugates *Bioconjugate Chemistry.* 2000;11:195-201.

[8] Alconcel S. N. S, Baas A. S, Maynard H. D. FDA-approved poly(ethylene glycol)-protein conjugate drugs *Polymer Chemistry.* 2011;2:1442-1448.

[9] Bennett C. L, Spiegel D. M, Macdougall I. C, et al. A Review of Safety, Efficacy, and Utilization of Erythropoietin, Darbepoetin, and Peginesatide for Patients with Cancer or Chronic Kidney Disease: A Report from the Southern Network on Adverse Reactions (SONAR) *Seminars in Thrombosis and Hemostasis.* 2012;38:783-796.

[10] Fee C, Van Alstine J. PEG-proteins: Reaction engineering and separation issues *Chemical Engineering Science.* 2006;61:924-939.

[11] Gaertner H, Offord R. Site-specific attachment of functionalized poly(ethylene glycol) to the amino terminus of proteins *Bioconjugate Chemistry.* 1996;7:38-44.

[12] Wong S. S, Jameson D. M. *Chemistry of protein and nucleic acid cross-linking and conjugation.* CRC Press 1991. Review of protein and nucleic acid chemistry.

[13] Harris J, Martin N, Modi M. Pegylation - A novel process for modifying pharmacokinetics *Clinical Pharmacokinetics.* 2001;40:539-551.

[14] Klenkler B. J, Sheardown H. Characterization of EGF coupling to aminated silicone rubber surfaces *Biotechnology and Bioengineering.* 2006;95:1158-1166.

[15] Jevsevar S, Kunstelj M, Porekar V. G. PEGylation of therapeutic proteins *Biotechnology Journal.* 2010;5:113-128.

[16] Veronese F, Pasut G. PEGylation, successful approach to drug delivery *Drug Discovery Today.* 2005;10:1451-1458.

[17] Cong Y, Pawlisz E, Bryant P, et al. Site-Specific PEGylation at Histidine Tags *Bioconjugate Chemistry.* 2012;23:248-263.

[18] Dou H, Zhang M, Zhang Y, Yin C. Synthesis and purification of mono-PEGylated insulin *Chemical Biology & Drug Design.* 2007;69:132-138.

[19] Yu D, Shang X, Ghosh R. Fractionation of different PEGylated forms of a protein by chromatography using environment-responsive membranes *Journal of Chromatography A.* 2010;1217:5595-5601.

[20] Zhou Z, Zhang J, Sun L, Ma G, Su Z. Comparison of site-specific PEGylations of the N-terminus of interferon β-1b: selectivity, efficiency, and in vivo/vitro activity *Bioconjugate Chem.* 2014;25:138-146.

[21] Fee C. J. Size-exclusion reaction chromatography (SERC): A new technique for protein PEGylation *Biotechnology and bioengineering.* 2003;82:200–206.

[22] Huang Z, Ye C, Liu Z, et al. Solid-phase N-terminus PEGylation of recombinant human fibroblast growth factor 2 on heparin-sepharose column *Bioconjugate chemistry.* 2012;23:740–750.

[23] Shang X, Yu D, Ghosh R. Integrated solid-phase synthesis and purification of PEGylated protein *Biomacromolecules.* 2011;12:2772–2779.

[24] Shang X, Ghosh R. Membrane reactor for continuous and selective protein mono-PEGylation *Journal of Membrane Science.* 2014;451:177–184.

[25] Ottow K. E, Lund-Olesen T, Maury T. L, Hansen M. F, Hobley T. J. A magnetic adsorbent-based process for semi-continuous PEGylation of proteins *Biotechnology journal.* 2011;6:396–409.

[26] Nojima Y, Iguchi K, Suzuki Y, Sato A. The pH-Dependent Formation of PEGylated Bovine Lactoferrin by Branched Polyethylene Glycol (PEG)-N-Hydroxysuccinimide (NHS) Active Esters *Biological & Pharmaceutical Bulletin.* 2009;32:523-526.

[27] Roberts M, Bentley M, Harris J. Chemistry for peptide and protein PEGylation *Advanced Drug Delivery Reviews.* 2002;54:459-476.

[28] Maiser B, Dismer F, Hubbuch J. Optimization of Random PEGylation Reactions by Means of High Throughput Screening *Biotechnology and Bioengineering.* 2014;111:104-114.

[29] Moosmann A, Blath J, Lindner R, Müller E, Böttinger H. Aldehyde PEGylation kinetics: a standard protein versus a pharmaceutically relevant single chain variable fragment *Bioconjugate chemistry.* 2011;22:1545–1558.

[30] Puchkov I, Kononova N, Bobruskin A, Bairamashvili D, MartÂŠyanov V, Shuster A. Recombinant granulocyte colony-stimulating factor (filgrastim): Optimization of conjugation conditions with polyethylene glycol *Russian Journal of Bioorganic Chemistry.* 2012;38:479–487.

[31] Corbett B, Macdonald B, Mhaskar P. Model Predictive Quality Control of Polymethyl Methacrylate *IEEE Transactions on Control Systems Technology.* 2015;23:687-692.

[32] Aumi S, Corbett B, Clarke-Pringle T, Mhaskar P. Data-driven model predictive quality control of batch processes *Aiche Journal.* 2013;59:2852-2861.

[33] Bonvin D, Srinivasan B, Hunkeler D. Control and optimization of batch processes *IEEE Control Systems Magazine.* 2006;26:34-45.

[34] Undey C, Tatara E, Cinar A. Intelligent real-time performance monitoring and quality prediction for batch/fed-batch cultivations *Journal of Biotechnology.* 2004;108:61-77.

[35] Birol G, Undey C, Cinar A. A modular simulation package for fed-batch fermentation: penicillin production *Computers & Chemical Engineering.* 2002;26:1553-1565.

[36] Nayhouse M, Kwon J. S.-I, Christofides P. D, Orkoulas G. Crystal shape modeling and control in protein crystal growth *Chemical Engineering Science.* 2013;87:216-223.

[37] Veronese F. M. Peptide and protein PEGylation: a review of problems and solutions *Biomaterials.* 2001;22:405–417.

[38] Laemmli U. K, others . Cleavage of structural proteins during the assembly of the head of bacteriophage T4 *nature.* 1970;227:680–685.

[39] Kurfürst M. M. Detection and molecular weight determination of polyethylene glycol-modified hirudin by staining after sodium dodecyl sulfate-polyacrylamide gel electrophoresis *Analytical biochemistry.* 1992;200:244–248.

[40] Li Y, Ogunnaike B. A, Roberts C. I. Multi-Variate Approach to Global Protein Aggregation Behavior and Kinetics: Effects of pH, NaCl, and Temperature for alpha-Chymotrypsinogen A *Journal of Pharmaceutical sciences.* 2010;99:645-662.

Chapter 4

Conclusions and Future Work

The contributions shown above develop the idea that modeling of biological systems and applications is assisted greatly by the use of some type of method for parameter estimation. It also shows that modeling and parameter estimation must be closely linked if we are to develop accurate and predictive models of these kinds of systems. The decision of complexity influences how we approach both the modeling and estimation parts of the problem. In the T cell work, the population balance model was based on first-principle notions of our understanding of cell growth, and it was combined with a model of media behaviour that fit our empirical understanding of the conditions in which the cells were grown. However, this resulted in a model with a number of unknown or uncertain parameters, with relatively few available measurements. In turn, this required the use of the parameter selection process shown, as estimation of the complete set of parameters would have been impossible. Finally, our understanding of the biological variability inherent in the cell culture process motivated the development of the adaptive process that was illustrated in that work.

On the other hand, the process in the PEGylation work was modeled as a simpler reaction mechanism, with relatively few parameters in the model. Therefore, the estimation technique used was able to estimate all of the parameters. Because this process did not have the kinds of biological variability that the cell culture process had, no adaptive estimation was developed.

There are many areas of potential for future investigations in this avenue of research. First, the parameter estimation procedures, especially in the case of the T cell culture process, had limited experimental data available. This limited the ability of the parameter estimation process (in terms of 'estimateable' parameters). Work should be directed in this area of linking the availability of measurements (both in number of measured variables, and in the frequency) to the utility of the parameter estimation process.

Another potential area of work is using the biological modeling shown above in the design of experiments. By developing these first-principle models and parameter estimation procedures, the goal is to build a predictive model for these biological systems. One issue with biological systems is the expense of experiments, in terms of both time and money. Predictive models should be used in combination with statistical tools in order to design experiments that get the maximum useful data at the lowest possible costs. A related area would be in design optimization, wherein the biological model created through the parameter estimation procedure is used to design a process to achieve some optimal goal or set target, and then validated experimentally.

One of the goals of process modeling is to be able to develop a model for use in control algorithms. This is of particular interest in the developing field of personalized cell therapies, as discussed in the T cell manuscript. The reliance on highly-skilled technicians for long-batch processes means very high production costs. Automated processes that can adapt to changing process conditions, biological variability, and other disturbances would be very useful to create better manufacturing processes. With the modeling and parameter estimation work shown, a next step could be the development of control algorithms for these batch-type cell culture processes.

In conclusion, biological systems and applications provide a rich field of potential research that demands the development of high quality models that can meet the needs of the application. Alongside of modeling, we must consider the question of parameter estimation, as these models will always contain some level of uncertain parameters, often with associated difficulties in process measurements due to the nature of the biological application in ques-

tion. This thesis presents two contributions to this field. Although these cases considered very different types of systems, both works were able to use first-principle modeling and parameter estimation techniques to provide predictive models for biological applications.