Systematic Approaches to Identification of Dynamic Flux Balance Models

by

Ali Nikdel

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Prof. Michel Perrier Chemical Engineering, Polytechnique Montréal
Supervisor	Prof. Hector Budman Chemical Engineering, University of Waterloo
Internal Member	Prof. William Anderson Chemical Engineering, University of Waterloo
Internal Member	Prof. Raymond Legge Chemical Engineering, University of Waterloo
Internal-external Member	Prof. Brendan McConkey Biology, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 5 has been published as the paper below, author contributed as the first author in developing the idea, implementation of the method and writing the manuscript:

> Nikdel, A., & Budman, H. (2016). Identification of active constraints in dynamic flux balance analysis. *Biotechnol Prog*. doi: 10.1002/btpr.2388

Chapter 6 has been published as the paper below, author contributed as the first author in developing the idea, implementation of the method and writing the manuscript:

> Nikdel, A., Braatz, R. D., & Budman, H. M. (2018). A systematic approach for finding the objective function and active constraints for dynamic flux balance analysis. *Bioprocess Biosyst Eng*. doi: 10.1007/s00449-018-1899-y

For chapter 7 author contributed as the second author in the paper below, performed part of the experiments, contributed to the idea of the paper, writing and editing the paper and statistical analysis:

> Naderi, S., Nikdel, A., Meshram, M., McConkey, B., Ingalls, B., Budman, H., & Scharer, J. (2014). Modeling of cell culture damage and recovery leads to increased antibody and biomass productivity in CHO cell cultures. *Biotechnol J, 9*(9), 1152-1163. doi: 10.1002/biot.201300287

Abstract

Mathematical modelling of biological systems is an essential tool for better understanding and for optimizing biological processes. Simulating the experiments before performing them is a time-saving strategy when seeking for optimal operating conditions.

Dynamic flux balance analysis (DFBA) is a constraint-based dynamic modelling approach where the cell is assumed to act as an optimizing agent that allocates resources to maximize/minimize a suitable biological objective. In this modelling approach a linear programming (LP) problem is solved at each time interval involving the optimization of an objective function subject to constraints. This is a significantly different modelling approach than the one generally adopted for biochemical systems where dynamic mass balances are formulated for each metabolite. An important advantage of the dynamic metabolic flux models as compared to the more conventional models reported before is that DFBA models do not depend heavily on accurate kinetic information. The ultimate purpose of this research was to develop DFBA type models with a minimal number of parameters to fit and predict experimental data. Two main challenges had to be addressed: (i) - finding the biologically meaningful objective function that cells are trying to maximize/minimize and (ii) - finding a minimal set of limiting constraints. Limiting constraints includes any limit that potentially reduces the solution space such as kinetics, thermodynamics, gene expression, etc. The focus on this thesis was to develop algorithms to address these challenges in a systematic fashion. In the first published paper from this thesis (Chapter 5), I present two approaches for finding the kinetic constraints, an algorithm based on Lagrange multipliers and a parametric sensitivity algorithm. Both algorithms were applied using two case studies, one for *E.coli* that used a simplified metabolic network and a second for *B. Pertussis* with a more comprehensive metabolic network. Although both algorithms were capable of finding the constraints for the simple *E.coli* example, for the more complicated metabolic network used with *B.Pertussis*, the parametric sensitivity approach was found to be inefficient and

numerically challenging. Thus, it was concluded that the Lagrange multiplier approach is preferable when considering metabolic networks of large dimensions.

In the second paper published from this work (Chapter 6) I proposed a method for systematically and simultaneously finding the objective function and the limiting constraints necessary to describe data. To that aim, a set based approach that accounts for the errors in the experiments were proposed. Using this method both the active kinetic constraints and biological objective function were identified simultaneously using a bilinear optimization formulation.

In the final stage of this research, the developed algorithms were applied for building a DFBA model for CHO cell culture (Chapter 9). We systematically found an objective function for mammalian cells that involve the simultaneous maximization of growth and minimization of apoptosis. We argued in this part of the work that this objective function correlates well with reported phenomena for CHO cells including the occurrence of the Warburg effect that involves the preferable conversion of glucose to lactate.

In terms of experimental work we conducted an initial study to investigate the damaging effect of effect of CO_2 accumulation on CHO cell culture. This preliminary study motivated us to develop a perfusion process for high cell density culture of CHO cells based on an ATF filtration system. In this development work we identified different challenges for effective perfusion operation and we propose some practical measures to address these challenges.

Acknowledgements

I would like to express my sincere gratitude to Prof. Hector Budman for going above and beyond the role of supervisor – it has been reassuring to have you not only as my academic supervisor but as a mentor.

However, I was not fortunate to have Prof. Jeno Scharer as my supervisor for long, always remember him.

I would like to thank the other members of my committee, Professors William Anderson, Raymond Legge, Brendan McConkey, Michel Perrier for agreeing to be part of my PhD advising committee.

I am also very thankful to all the staff at Chemical Engineering department especially technical staff Ralph Dickhout and Bert Habicher.

NSERC is appreciated for funding of this research.

I would like also thank MilliporeSigma and especially Delia Lyons and Jeremiah Riesberg for their support and great discussions.

I also want to thank my friend and colleagues at Budman group: Rubin Hille, Mariana Carvalho, Vanessa Zavatti, Jasdeep Mandur, Hengameh Aghamohseni, Kaveh Ohadi.

In the end, it would not have been possible without the support and encouragement and sacrifices from my wife, Vida, my parents and my brothers.

Dedication

“Tomatoes can talk. “ (Darwin K.)

To all the adults who stay curious as kids.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Motivation

The modern biopharmaceutical industry that started with the discovery of Penicillin, the first antibiotic, received new momentum by the introduction of recombinant DNA technology. This technology emerged after Paul Berg presented his gene-splicing experiment and the idea progressed when Herbert Boyer and Stanley Cohen put forward the idea of inserting the recombinant DNA into bacteria. Genentech was the first biopharmaceutical company that used this technology for large-scale production of recombinant DNA. Genentech's first product was recombinant Insulin, the first human protein therapeutic using DNA technology. Five years later in 1982, USA Food and Drug Administration (FDA) approved the recombinant DNA based insulin and the product was further marketed by Eli Lilly. Since then many other therapeutic proteins including blood factors VII and VIII, interferons and many other life-saving drugs have been approved. The market of biopharmaceuticals has increased at an unprecedented rate. For example, in 2009, 200 biopharmaceuticals were reported in the market with more than US $99 billion sales, and in 2016 it reached to US $228 billion (Moorkens et al., 2017). Protein based biopharmaceuticals are in the top 5 the best-selling drugs (Craven, Whelan, & Glennon, 2014).

Monoclonal antibodies (mAbs) are a major group of therapeutic proteins that have a major role in today's biopharma industry. They are used for diagnostic purposes and research but their main role is treatment of various range of diseases from infections to inflammatory diseases and cancers. The success of monoclonal antibodies such as Infliximab (Remicade), Etanercept (Enbrel), Bevacizumab (Avastine), Rituximab (Rituxan), Adalimumab (Humira), Trustuzumab (Herceptin) has led to increased

demand of these products to more than 100kg/year and created a need for improving production yield through the improvement of cell lines and manufacturing processes.

To produce recombinant protein therapeutics proteins including mAbs, a host cell is required that is engineered to contain the gene of interest which is responsible for the production of the desired protein. The choice of host cell is very critical for producing the therapeutic proteins with the right efficacy and that can be safely used as a therapeutic agent.

The first host used for production of recombinant proteins was the bacteria *Escherichia coli* (*E.coli*) followed later by other hosts including mammalian cells, insect cells, fungal cells and plants. Nowadays the majority of recombinant therapeutic proteins are produced in mammalian cell lines such as Chinese hamster ovary (CHO), baby hamster kidney (BHK), human embryonic kidney (HEK-293), human retina derived (PERC6) and mouse myeloma (NS0). One of the main reasons for choosing mammalian cells as the host cell line for producing therapeutic proteins is their capability to provide post-translational modifications such as proper glycosylation of the protein of interest. Glycosylation is a process whereby different glycans are attached to proteins in particular configurations thus conferring recombinant glycoproteins with particular physical and functional properties (Butler & Spearman, 2014). More than 70% of recombinant therapeutic proteins are produced in CHO cells. Among the special features of CHO cells that has made them popular for large scale manufacturing are their ability to grow in a suspension culture, their inherent resistance to infection by viruses and their gene amplification qualities that makes them able to express larger quantities of proteins (Butler & Spearman, 2014).

In recent years, some of the ‘blockbuster’ mAbs have gone off patent so there is an opportunity for generic (biosimilar) manufacturing companies entering this market to further increasing the competition. Delays in market entry due to time spent on cell line and bioprocess development impacts

profitability and also increases the risk of loss of market share. This further motivates originator companies to optimize their manufacturing processes and to achieve better product quality.

Mathematical models are essential tools for improving the understanding of the cell metabolism and cell culture behavior and for their use in model based bioprocess optimization.

Biopharmaceutical process optimization is often more challenging as compared to other chemical manufacturing processes as biological organisms that are used for producing the desirable products exhibit significant variability in their responses to environmental conditions. Moreover, a particular combination of different environmental perturbations can heavily impact the intracellular level of these producing cells. Thus, to understand biological systems, it is not sufficient to just analyze each environmental factor individually as the interactions between those factors are of special importance. The discipline of System Biology studies a biological system as an integrated system of interacting parts (Tomar & De, 2013).

Since the introduction of mammalian cells for the production of biopharmaceuticals, different metabolic models have been used for the systematic understanding of cellular behavior at different conditions. These reported models have been generally used for describing a limited set of process conditions and also for some process optimization targets like finding the most appropriate feeding regime in fed-batch operation. Chinese hamster ovary (CHO) cell have been modelled using different approaches and, to a lesser degree, other models have been developed for the study of less utilized cell hosts such as hybridoma and HEK 293.

These metabolic models have been mostly developed for describing a static state of the biosystem which make it less useful for actual manufacturing processes that are of a dynamic nature. A few available dynamic models are available that only consider the metabolism loosely and instead they involve unstructured mass balances with many kinetic terms of the Monod form. These kinetic terms

generally involve a large number of parameters which makes these models highly sensitive to model error and measurement noise in the data used for model calibration, especially when data is scarce as it is often the case in biochemical operations (Ghorbaniaghdam, Henry, & Jolicoeur, 2013). Thus there is a genuine motivation to seek for structured models that correctly account for metabolic interactions while resulting in more parsimonious expressions with a smaller number of calibration parameters. Such models are expected to be more robust and better suited for model based prediction and optimization over a large range of operating parameters.

1.2 Objective

In this research, the object was to develop necessary algorithms required for building a Dynamic Flux Balance Analysis model to be used for mammalian cells. Unlike the conventional dynamic unstructured models which, as mentioned above, use many kinetic expressions involving a large set of tuning parameters, the currently proposed approach is based on using Dynamic Flux Balance Analysis (DFBA) algorithm (Mahadevan, Edwards, & Doyle, 2002). DFBA is based on the formulation of an optimization problem where a particular flux distribution in a metabolic network is sought such as a specific objective function is optimized subject to certain constraints. The rationale behind DFBA is that natural evolution had conditioned the cell to act as an agent that optimally allocates resources to satisfy a certain biological objectives. For bacterial and microbial organisms typical objectives that have been successfully used to describe experiments were maximization of growth and maintenance of a particular redox potential. However, optimization objectives have not been proposed as yet for mammalian cells due to their relative higher complexity as compared to bacteria. For example, mammalian cells exhibit particular regulatory mechanisms such as programmed cell death (apoptosis) or cell death by autophagy that have to be accounted for when defining an objective function for

optimization. Thus, a main goal of this proposed study is to find an objective function and specific metabolic constraints that can be combined together to formulate a suitable optimization problem which solution describes the experimental behavior of the mammalian cells.

Then, the resulting DFBA model can be employed in future process optimization approaches for genetic modification, media design or bioprocess operation.

Although the final goal was to formulate a dynamic metabolic flux model for CHO, I initially developed the model calibration approach for *E. coli* and *Bordetella Pertussis*. The reasons for developing the initial approach for these microorganisms were: (i)- a simple calibrated dynamic metabolic flux model of *E.coli* was available (Mahadevan et al., 2002) thus allowing us to test our identification approach on an a priori known model, (ii)- the metabolic network of reactions was *a priori* known for Pertussis with reasonable accuracy (Budman, Patel, Tamer, & Al-Gherwi, 2013) and was provided by a collaborating company, (iii)- considerable data were already available at the beginning of this project and iv- the assumption of growth rate as the objective function to be maximized within the model has been reported to work well for bacteria whereas the objective functions that are suitable for CHO have not been clearly identified as yet.

Using the data available for *Pertussis* or *in silico* data (simulated data) generated by the available *E. coli* model, we developed systematic approaches to identify the constraints and objective functions that are the basic elements of the dynamic metabolic flux models. Then, using the approaches for systematic identification of constraints and objective function, we developed a DFBA model for CHO. To calibrate this model two sets of data collected during two different batch operation were used. The use of batch mode was preferred over continuous operation since the data spans in the former a larger range of

operating conditions over the duration of the batch thus providing richer information for model calibration.

An additional part of the current project is a continuation of a study of CHO cell death under a mimicked perfusion cell culture. The perfusion operation was initially implemented by conducting intermittent replacements of part of the supernatant volume during a batch culture. Although the fed-batch operation is the most common type of operation in the biopharmaceutical industry due to a lower risk of contamination and the simplicity of operation, in recent years perfusion mode of operation has become very popular (Chotteau, 2015) especially by introduction of new perfusion systems. Perfusion processes are generally better able to maintain constant growth conditions over a larger range of cell densities. In the experiments it was realized that limitation in nutrients is not the only limiting factor and other factors such as higher concentration of carbon dioxide might play a critical role. Recognizing the importance of perfusion and following the results described above using an emulated perfusion system, the final part of the thesis involved the construction and experimentation of a continuous perfusion bioreactor system. Data collected from this perfusion system was used together with batch data to calibrate the proposed dynamic metabolic flux model described above.

Chapter 2

Literature review

2.1 Monoclonal antibodies

In the early twentieth century, Paul Ehrlich proposed the idea of "magic bullets" that referred to drug substances that will be able to specifically target an agent causing disease in the body and destroy it. The first monoclonal antibody was for use in patients with a kidney transplant to prevent the organ rejection by immune system. This product was developed in 1975 and approved in 1986 and the new era of developing monoclonal antibodies for different diseases including various type of cancers started (Liu, 2014). Among the ten best selling drugs of 2016, seven of them were monoclonal antibodies (Philippidis, 2017).

2.2 Chinese Hamster Ovary (CHO) cells

Chinese hamster was first used in 1919 for research on pneumococci. Chinese hamster was recognized as a suitable research host due to their low number of chromosomes. The Chinese hamster ovary (CHO) cell line was developed by Theodore T. Puck in 1957 (Marcus, Sato, Ham, & Patterson, 2006). Since their introduction CHO cells had many applications in biomedical research. They grow well in *in vitro* culture and can be grown in both adherent and suspension cultures. The first approved therapeutic protein obtained from recombinant mammalian cells, human tissue plasminogen activator, was produced using CHO cells. Since their approval in 1986, CHO cells have been recognized as the workhorse for producing recombinant therapeutic proteins including mAbs (Kim, Kim, & Lee, 2012). Other mammalian cells also have been used for this purpose but none of them are as popular as CHO cells. CHO cells are the host cell line for about 70% of all the recombinant therapeutic produced

(Jayapal, Wlaschin, Hu, & Yap, 2007). The reasons for their popularity are that CHO cells are well-studied and they are considered safe by drug agencies. They exhibit good gene amplification thus resulting in increased protein productivity, they support the necessary post translation modifications such as glycosylation and they adapt well to serum free media (Kim et al., 2012).

2.3 Vaccines

Vaccines are biological material enhances the immunity to a specific disease. Vaccines are one of the most important inventions of humankind. Edward Jenner introduced the first vaccine against smallpox in 1796 and since then vaccines have had a vital role in eradication of the disease (Nabel, 2013). Rabies vaccine was the next vaccine that is produced in the cultured cells (Plotkin, 2014). More than 70 vaccines are in the market to be used against about 30 types of microbes. Main groups of vaccines are live or attenuated organisms and killed (inactivated organisms) based vaccines. In some cases when the antigen responsible for immunity against the target disease is known purified or recombinant subunits of the vaccine could be used instead of the whole live or killed organism. Killed vaccines cannot replicate inside the body so they are usually administered in combination with a compound called adjuvant to increase their potency. Aluminum salts are among common adjuvants used in vaccine industry (Ulmer, Valley, & Rappuoli, 2006).

2.4 *Bordetella pertussis*

Bordetella pertussis is a gram negative bacteria. This bacteria is responsible for *whooping cough*, a potentially fatal bacterial disease. Two types of vaccines for *whooping cough* are available: whole-cell and acellular vaccines. Acellular vaccines are advantageous in terms of having fewer side effects thus they are normally the vaccines of choice. The acellular vaccine generally is composed of 1–4 protein

virulence factors: pertussis toxin (PT), pertactin (PRN), fimbriae (FIM) and filamentous hemagglutinin (FHA).

2.5 Cell metabolism

Mammalian and bacterial cells need energy for driving the cell functions. This energy is provided by several biochemical reactions that involve carbon and nitrogen sources (mainly carbohydrates and amino acids, lipids) that are normally consumed by the cells and are converted to energy carrier molecules (e.g. ATP and ADP) and other intermediate materials required for cell activities. The reactions occurring in the cell are generally classified into catabolic and anabolic. Catabolism is the process of breaking down the large molecules (e.g. glucose) into other useful molecules, energy and redox potential (e.g. NADH), and anabolism is the process of formation of new molecules that are required for cell functions using the biomolecules resulted from catabolism (Alberts, 2008). Figure (2-1) shows schematically the catabolic and anabolic pathways in the cell.

Enzymes, mostly made of proteins, are biochemical catalysts involved in regulating (e.g. accelerating) the biochemical reactions. Metabolic pathways generally referred to groups of biochemical reactions. For example, glycolysis, citric acid cycle (also known as TCA cycle or Krebs's cycle), oxidative phosphorylation, pentose phosphate pathway and urea cycle are among the main metabolic pathways.

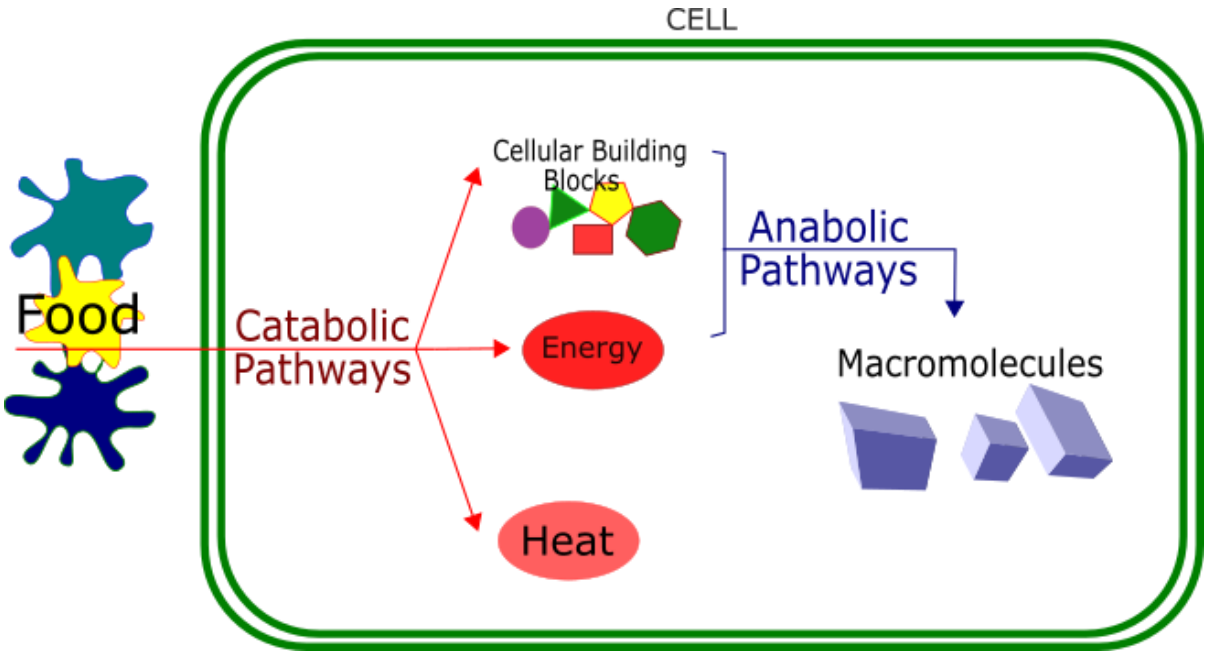


Figure 2-1 Catabolism and anabolism in cells. Catabolism is the process of breaking down the large molecules into other useful molecules, energy and redox potential, and anabolism is the process of formation of new molecules that are required for cell functions using the biomolecules resulted from catabolism.

2.5.1 Glycolysis

Glycolysis is the first and main pathway in glucose metabolism. In glycolysis, for each mole of glucose consumes, two moles of pyruvate and two moles of ATP are generated. Glycolysis takes place in the cytosol. The net reaction summarizing the glycolytic pathway, after canceling out intermediate reactions along the pathway, is:

$$\text{Glucose} + 2\ NAD^+ + 2\ Pi + 2\ ADP \rightarrow 2\ \text{Pyruvate} + 2\ NADH + 2\ ATP + 2\ H^+ + 2\ H_2O$$

2.5.2 Pentose phosphate

The pentose phosphate pathway (PPP) branches out from the glycolytic pathway and its main goal is to provide the NADPH needed for reductive biosynthesis such as synthesis of lipid and nucleic acid and also reduction of glutathione. PPP has two phases: an oxidative phase that generates NADPH (while oxidizing glucose-6-phosphate to ribose 5-phosphate) and a non-oxidative phase that interconverts sugars (Berg, Tymoczko, Stryer, & Stryer, 2002). The net reaction of the oxidative PPP phase can be written as:

$$\text{Glucose 6-phosphate} + 2\ NADP^+ + H_2O \rightarrow \text{ribulose 5-phosphate} + 2\ NADPH + 2\ H^+ + CO_2$$

The ratio of NADPH to NADP+ has been often suggested as the main factor for controlling the pentose phosphate pathway (Barcia-Vieitez & Ramos-Martinez, 2014). The pentose phosphate pathway occurs in the cytoplasm.

2.5.3 TCA cycle

Citric acid cycle, also known as TCA cycle, or Krebs's cycle is another important pathway in cell metabolism located in the mitochondria. The TCA cycle describes an aerobic process occurring in the cells which the main goal is the production of NADH needed for ATP production. In microorganisms that undergo glycolysis Pyruvate both converts to Acetyl- CoA and also feeds into TCA cycle. The net reaction is as follows:

$$\text{Pyruvate} + 2\ NAD + 2\ CoA \rightarrow 2\ \text{Acetyl CoA} + 2\ NADH + 2\ CO_2$$

Some microorganisms such as *Pertussis* undergo gluconeogenesis where the reactions occur in reverse order to glycolysis, i.e. from pyruvate towards the generation of glucose (Hu & Zhou, 2012).

Acetyl-CoA is also generated by oxidation of fatty acids. After formation, Acetyl-CoA breaks into two molecules of CO_2. As mentioned above, the TCA cycle provides reducing NADH that will be used in other reactions such as the oxidative phosphorylation pathway and it also has an important role in providing the precursors required for synthesis of some of the amino acids such as glutamine (Hu & Zhou, 2012). The TCA cycle net reaction can be written as follows:

$$\text{Acetyl CoA} + 3\ NAD^+ + FAD + ADP + Pi + H2O \rightarrow 2\ CO_2 + 3\ NADH + FADH_2 + ATP + CoA$$

2.5.4 Oxidative phosphorylation pathway

The oxidative phosphorylation pathway is an important pathway in producing energy for the cell. In this pathway, nutrients are oxidized to produce energy storage molecules of ATP. In fact, at this stage electrons that are released in glycolysis are consumed in the TCA cycle. It should be remembered that while oxidative phosphorylation has an important role in providing energy for the cell, reactive oxygen species (ROS) such as superoxide and hydrogen peroxide are also produced in this pathway resulting in an increase of free radicals that have detrimental effects on the cell including aging (cellular senescence) and also some diseases.

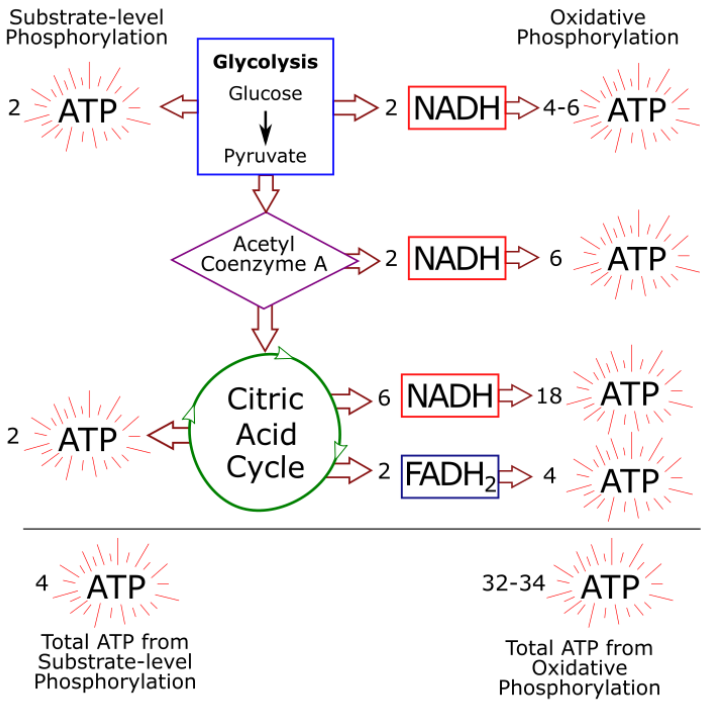


Figure 2-2 Summary of ATP production from one mole of glucose. This figure shows ATP production in each pathway of glycolysis and TCA cycle.

2.5.5 Metabolic regulation

Cells convert nutrients, mainly glucose, to energy for survival, proliferation and protein production. Glycolysis, the TCA cycle, and oxidative phosphorylation are the key pathways used by the cell to produce up to 36 moles ATP per mole of glucose consumed. Ammonia, lactate and CO_2 molecules are also key by-products of these pathways (Figure 2-3).

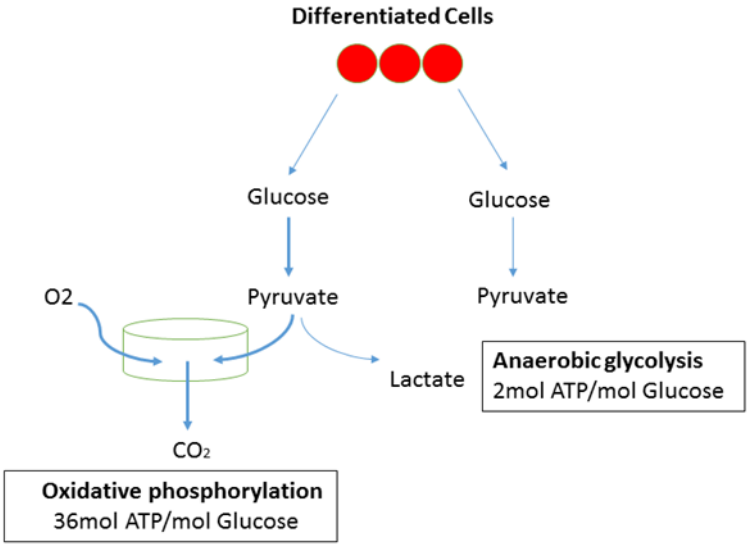


Figure 2-3 Schematic representing the differences between oxidative phosphorylation and anaerobic glycolysis

In the absence of enough oxygen, the carbon molecules follow from the pyruvate branch point an anaerobic metabolic pathway in which lactate production occurs instead of CO_2. In this situation, only two moles of ATPs are produced for every one mole of glucose being consumed (Figure 2-3). Surprisingly, proliferative mammalian cells such as tumor cells and most cell lines for producing recombinant proteins have been found to metabolize glucose to lactate even in the presence of sufficient amounts of available oxygen. This phenomenon has been referred to as the 'Warburg effect" or aerobic glycolysis. In fact the Warburg effect is viewed as an inefficient metabolic pathway since in the conversion of one mole of glucose to lactate only 4ATP are produced, as compared to oxidative phosphorylation where up to 36 moles of ATP are produced for each consumed mole of glucose (Figure 2-3) (Quek, Dietmair, Kromer, & Nielsen, 2010; Vander Heiden, Cantley, & Thompson, 2009). It has also been observed that the lactate resulting from glucose consumption is an important inhibition factor

for cell growth and product formation due mostly to increased acidity of the medium and increased osmolality (Ahn & Antoniewicz, 2012b).

On the other hand and despite the resulting inefficient ATP production the Warburg effect has some beneficial consequences since through a set of signaling and regulatory pathways it protects fast-growing cells by reducing mitochondrial activity, which is known to be partially responsible for apoptosis. Beyond ATP additional requirements for proliferating cells are nucleotides, amino acids, and lipids for growth. Amino acids need more equivalents of carbon and NADPH, as compared to ATP, to be synthesized inside the cell. Therefore, instead of converting all the glucose through the oxidative phosphorylation pathway to CO_2 in the mitochondria, cells allocate some glucose derived resources to macromolecular precursors such as acetyl-CoA.

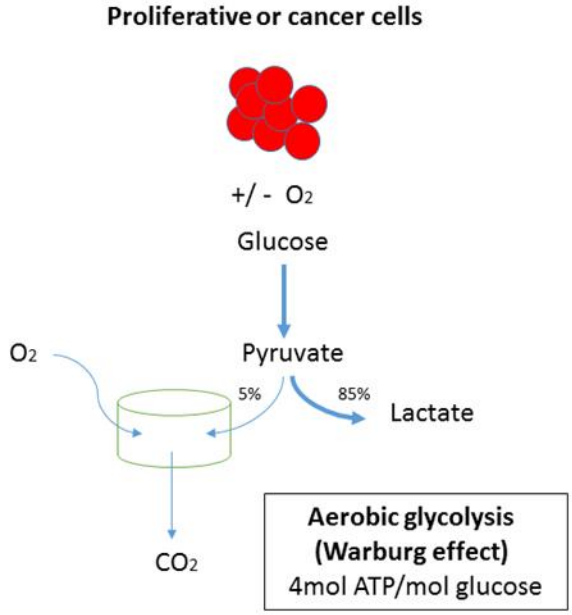


Figure 2-4 Schematic representing the aerobic glycolysis pathway in proliferative cells. These cells produce large amount of lactate and 4mol of ATP.

Glycolytic intermediates are required for anabolism of nonessential amino acids and ribose that is necessary for nucleotides' synthesis. Also glycolytic intermediates are needed for synthesis of fatty acids (Vander Heiden et al., 2009).

Another example of an energetically inefficient metabolic pathway in mammalian cells is partial oxidization by transamination of glutamine to ammonia and to non-essential amino acids such as alanine and aspartate (Quek et al., 2010). This phenomenon is actually associated with a truncated TCA cycle, where lower fluxes occur between pyruvate to citrate flux but a higher flux occurs from α-ketoglutarate to oxaloacetate (Lake-Ee & and Lars Keld, 2009).

As already mentioned above the by-products from incomplete metabolism of mammalian cells, mainly lactate and ammonia, create unfavorable living conditions for cells in the media. They have been generally found to be inhibitory for cell growth and protein production. Several strategies have been developed to cope with these challenges ranging from molecular biology methods for cell line modification to process optimization strategies (Quek et al., 2010), as described below.

2.6 Lactate reduction and control

Generally, in CHO cells producing proteins, it has been observed that lactate is produced during the growth phase when cells are growing faster as compared to the post-exponential phase of growth. It has been also observed that in some of the CHO cell lines there is a metabolic shift from lactate production to lactate consumption. There are different hypotheses on how this shift is regulated. For example, oxidative stress in the cells and/or lack of some key nutrient such as copper has been mentioned as possible factors (Luo et al., 2012).

Altamirano et al (2006), studied the lactate response by comparing two different culture conditions: 1- a combination of 20 mM glucose with 6mM glutamate and 2- a combination of 5mM glucose and 20mM galactose with 6 mM glutamate. They observed that for the second culture where part of the glucose used in the first culture was substituted with galactose, the formation of lactate was lower where glucose was initially used and then galactose was used. During the phase of galactose consumption they also observed consumption of endogenous lactate in the culture. To explain the results, they performed metabolic flux analysis (MFA) study and suggested that lactate is not consumed as a source of energy and hypothesized that other pathways exist involving the non-oxidative decarboxylation of pyruvate. These pathways have been also reported in cancer cells (Altamirano, Illanes, Becerra, Cairo, & Godia, 2006; Galleguillos et al., 2017). However in a separate modelling study based on flux balance analysis (FBA), Martinez et al. compared the fluxes during two phases, before and after shifting from lactate production to lactate consumption, and suggested that during the lactate production phase most of acetyl-CoA produced by pyruvate dehydrogenase (PDH) was used for the production of fatty acids and steroids while during the lactate consumption phase only 32% of the produced acetyl-CoA was later used for production of fatty acid and steroids while majority of acetyl-CoA went through the TCA cycle and was oxidized there. They suggested that the discrepancy between their findings with the study by Altamirano might be due to over-simplification of the metabolic network that was used in Altamirano's work (Martinez et al., 2013). Sengupta et al (2011) applied both MFA and isotopomer labeling for mapping the flux of CHO cell culture during the post-exponential phase of low growth. They observed that the majority of glucose was metabolized into NADPH. They suggested that the reason for this preferential pathway might be the higher demand of NADPH for reducing the oxidative stress. This research also indicated that lactate is not produced at this stage as most of the material flux originating from pyruvate was directed into TCA cycle (Sengupta, Rose, & Morgan, 2011). Templeton

et al. (2013) used MFA in combination with ^{13}C to study the different phases of cell growth and antibody production in CHO cell culture. They found that in the fast-growing phase, lactate was produced in higher amounts as compared to the post-exponential phase and the activity in the TCA cycle was minimal during the growth phase. Also, they observed that following a metabolism shift from higher growth to higher antibody production, the lactate response changed from production to consumption. Oxidative phosphorylation was the main source of energy production when antibody production was at its highest level but also the oxidative pentose phosphate pathway was observed to be highly active during this time period. This study also found that during the stationary phase while the growth rate declined, the TCA cycle and the oxidative pentose phosphate pathway remained active. They argue that during the time with the highest antibody production, cells were in their highest oxidative state (Templeton, Dean, Reddy, & Young, 2013). Wilkens et al performed a metabolic flux analysis for two different CHO culture conditions wherein one of them only glucose was fed to the culture and in the other glucose was fed together with galactose. In the culture fed with the combination of galactose and glucose after the initial consumption of glucose, galactose started to be consumed. This suggested that in the galactose consumption phase not enough pyruvate was available for the cells to use as energy source so the produced lactate was converted to pyruvate by changing the direction of the corresponding reaction regulated by lactate dehydrogenase. The resulting pyruvate was subsequently directed into the TCA cycle (Wilkens, Altamirano, & Gerdtzen, 2011). In a study by Xing et al in 2011, a CHO cell line expressing an antibody fusion protein (B1) metabolic flux analysis was used to optimize the amino acid composition in the cell culture medium. They increased the concentration of methionine, tryptophan, asparagine and serine that were considered as limiting and reduced the concentration of arginine, alanine and glycine with the goal of reducing the ammonia production. Then, as a result of the lower concentration of ammonia, both the cell density and protein

production were increased (Xing et al., 2011). Hefzi et al developed a genome-scale flux balance analysis model for CHO cells where they included 4455 metabolites, 6663 reactions and 1766 annotated genes in their model. This model connected the genotype and phenotype of CHO cells and showed the potential for protein production by engineered CHO cells that would utilize the resources optimally (Hefzi et al., 2016).

2.7 Genetic modification of mammalian cells

Knowledge about the cell metabolism has been often used to develop genetic modification strategies that would result in larger growth and productivity. For example, in a study on hybridoma cells producing monoclonal antibodies it was found that knocking down the lactate dehydrogenase (LDH) gene resulted in decreased lactate production and substantial improvements in cell viability and protein production (Chen, Liu, Xie, Sharp, & Wang, 2001).

In another investigation, the overexpression of yeast pyruvate carboxylase (PYC2) gene in BHK-21 cells transfected with Erythropoietin (EPO) gene was induced to enhance the activity of pyruvate carboxylase that promotes growth under glucose reduced conditions thus resulting in decrease lactate formation. Pyruvate carboxylase activity has not been observed in mammalian cells, however its gene does exist (Lake-Ee & and Lars Keld, 2009). The over-expression of the gene also resulted in about doubling of the efficiency of consumed glucose towards the production of EPO (Irani, Beccaria, & Wagner, 2002; Irani, Wirth, van Den Heuvel, & Wagner, 1999). Toussaint et al designed a CHO cell line producing a monoclonal antibody that resulted in a shift towards lactate consumption, extended growth phase and, higher cell concentration and productivity in fed-batch operation (Toussaint, Henry, & Durocher, 2016).

The glutamine pathway has also been considered as a target for cell line genetic engineering studies. The enzymatic reaction led by the glutamine synthetase (GS) enzyme is responsible for the synthesis of glutamine from glutamate and ammonia and it is the only reaction occurring in mammalian cells that synthesize glutamine. The GS enzyme becomes essential for mammalian cell survival when the medium is not externally supplied with glutamine. In some cell lines such as mouse myeloma, there is not sufficient expression of GS for survival thus glutamine has to be added. In CHO cells the GS expression is just enough for survival thus requiring the addition of glutamine. In these cells, Methionine sulphoximine (MSX) is added to inhibit endogenous GS activity and only cell lines transfected with the overexpressed GS survive in this medium. In summary, mammalian cells transfected using the GS system can be grown in a glutamine-free medium with just glutamate added as a substitute. Reduced glutamine usage results in the lower formation of toxic ammonia thus resulting in improved growth. This method has been commercialized by the Lonza Company and has gained wide acceptance in the industry (Bebbington et al., 1992; Fan et al., 2012; Quek et al., 2010).

2.8 Methods to increase cell density

Since productivity is often associated with cell mass, there is often a motivation to increase cell density. Also for experimental studies, it is important to obtain high density so as to increase the sensitivity of the experiment to changes in experimental conditions.

The mode of cell cultivation has a significant effect on viable cell density. Fed-batch and perfusion are the most common modes of operation in the industry for obtaining higher cell density. Fed-batch culture is nowadays the dominant approach in the biopharmaceutical industry for controlling the cell-specific growth rate and residual nutrient concentrations while leading to high cell density. A concentrated feed is typically used to provide enough nutrients and to maximize viable cell density and product titer. Cell

densities of ~10 million cells/ml can be attained in fed-batch by extending the productive state for long periods of time. A fed-batch culture consists of two phases: an initial growth phase followed by a cell density decline phase. The reason that perfusion culture is nowadays less common comparing to fed-batch is that perfusion operation is more complicated to run comparing to fed-batch. Also, the risk of contamination in perfusion is relatively higher and the right equipment for perfusion scale-up was not available until recently (Bonham-Carter, 2011).

Feeding is used as a strategy to increase cell density but cell growth can be hindered by the accumulation of inhibitory agents such as ammonia and lactate and also by elevated osmolality resulting from the addition of a base needed for pH control (Newland, Kamal, Greenfield, & Nielsen, 1994). Unlike microbial cells' cultures for which limiting the feed can be helpful in controlling the cultivation and limiting the by-products formation, in mammalian cells excessive starvation may induce apoptosis which results in a high cellular death rate (k_D) and decreased productivity. Based on these limitations some strategies have been developed for mammalian cells such as initial over-supply of nutrients followed by a cell arrest phase obtained by adding growth inhibitors or decreasing temperature. As mentioned above, perfusion is another mode of operation used to increase cell density. Because of its importance in the current research, it is described separately in the following chapter.

2.9 Mathematical metabolic modelling

In order to develop and optimize a process, a thorough understanding of the process is a critical step. Mathematical metabolic models are useful for describing and better understanding biological processes. A simple and well-developed model is often helpful in reducing the number of experiments required for developing and optimizing the process. Generally, models should be easy to build and use and the type of model to be developed may also depend on the specific expected application.

For example in order to find the best targets for genetic manipulation metabolic models could be used to rationally identify needed knock out/in/up/down of genes responsible for specific enzymes in the metabolic pathway to increase the production of desired metabolites. Without having a systematic view based on a model such a manipulation might need a higher number of trial and error based experiments. Instead, by forecasting the best targets based on mathematical modelling, recombinant DNA techniques can be implemented to change or delete the genetic function of a specific target (Zomorrodi, Suthers, Ranganathan, & Maranas, 2012). On the other hand, at the bioprocess level, models can also be used to design the feeding regime in a fed-batch operation or the perfusion rate schedule in a perfusion operation to maximize growth and/or productivity.

2.9.1 Mathematical model categories

Mathematical models for biochemical systems can be classified, according to their specifications and applications, into different groups: Structured or unstructured, segregated or unsegregated, deterministic or stochastic (Sidoli, Mantalaris, & Asprey, 2004). Structured models explicitly take into account the intracellular reactions reported for a particular organism unlike unstructured models than just describe mass balances of specific metabolites using empirical yields and Monod expressions but do not account for the different reactions involving a specific metabolite either as a reactant or as a product. Segregated cell models account for the heterogeneity of the cell culture thus assuming that different cells are at different metabolic stages at any given time. On the other hand un-segregated models assume that the cell population is homogeneous and thus describe an average status of the cell population within the culture. Unlike deterministic models where parameter and initial conditions values are fixed, stochastic models assumed these values to be random. The choice of the modelling

approach for a particular problem depends on the purpose of the model, the situation of the process and available data (Sidoli et al., 2004).

2.9.1.1 Kinetic Models

A common approach in biochemical modeling is to identify main reactions involved in the process and to describe each of the reaction rates by kinetic expressions of Monod or other types. These kinetic expressions are then combined together into a system of differential equations that represent dynamic mass balances of each of the metabolites participating in the identified reactions as substrates or products. These models are deterministic since they involve deterministic parameters and initial conditions. In some cases, simple unstructured empirical mass balances are used where the consumption of the metabolite towards biomass or towards another metabolite is determined by empirical Monod kinetic terms and yields. For example, a simple kinetic model that has been used extensively for glucose consumption for producing biomass by using Monod type equations is given as follows:

$$S \rightarrow X \tag{2-1}$$

$$\frac{ds}{dt} = -\frac{\mu_{max}}{k_s + s}\frac{X}{Y_{X/S}} \tag{2-2}$$

where $\frac{dS}{dt}$ is the consumption rate of glucose, $\mu_{\max}$ is the maximum specific growth rate, $\boldsymbol{S}$ and X are glucose and biomass concentration, k_s is the value of $\boldsymbol{S}$ when $\mu/\mu_{max} = 0.5$, $Y_{X/S}$ is the rate of biomass formation from substrate (glucose) consumption.

When considering many metabolites, these types of kinetic models are generally heavily parameterized. Then, in the presence of a large set of kinetic parameters, the downside of these models is their

sensitivity to model structure error and measurement noise especially when data is scarce. Also, the parameter estimation step is challenging since the kinetic equations are mostly expressed in non-linear forms thus necessitating nonlinear estimation techniques, e.g. when using Michaelis-Menten type equations (Sidoli et al., 2004). An additional problem is that parameters in the numerator and denominator of Michaelis-Menten expressions are highly correlated necessitating experiments at very low concentrations to determine the half-saturation constants while at low concentrations the data is generally inaccurate.

2.9.1.2 Metabolic flux analysis (MFA)

Metabolic flux analysis (MFA) is a modelling approach that identifies the distribution of flux rates at steady-state conditions. Based on stoichiometric relations among metabolites, dynamic mass balances can be written for the metabolites as follows:

$$\frac{d\mathbf{x}}{dt} = \boldsymbol{S}.\boldsymbol{v} \qquad (2\text{-}3)$$

where $\boldsymbol{S}$ is the stoichiometric matrix, **x** is the vector representing the concentration of the metabolites, $\boldsymbol{v}$ is the vector for metabolic fluxes and μ is the cell specific growth rate. In MFA it is usually assumed that the intracellular environment is at a quasi-steady state ($\frac{d\mathbf{x}}{dt} = 0$).

This assumption is based on the fact that changes in intracellular fluxes (within the cells) occur much faster comparing to extracellular ones but this may not be always accurate.

$$\boldsymbol{S}.\boldsymbol{v} = 0 \qquad (2\text{-}4)$$

Generally, input and output fluxes into and from the cell can be experimentally measured whereas intracellular fluxes are more difficult to measure. Following the assumption of quasi-steady state, it is possible to cancel out the reactions among intracellular intermediates and establish relations between

input and output fluxes but the resulting model is generally underdetermined given by equation 2-5 since the degrees of freedom are larger than zero:

$$d=n-k-m \quad (2\text{-}5)$$

where d is the degree of freedom, n is the number of fluxes, k is the number of constraints and m is the number of measurable fluxes.

There are two major approaches to address the under-determinacy problem or lack of measurements for calibrating stoichiometric models:

a) Using labelling data to experimentally find important/negligible intracellular fluxes

b) Fit extracellular data with a mathematical and biologically meaningful objective function subject to constraints where the latter are based on the stoichiometry and kinetic rate limits.

These two approaches are further discussed in the following sections.

Isotope-based metabolic flux analysis is a technique to measure intracellular metabolic fluxes. 13C-based metabolic flux analysis has been applied in different prokaryotic and eukaryotic systems (Yang, 2013). 13C-MFA technique involves the following steps: (1) selecting appropriate tracers for the system under study (2) implementing isotopic labeling experiments, (3) quantifying isotopic labeling distributions in metabolic products, (4) estimating metabolic fluxes using least-squares regression, and (5) evaluating the goodness of fit and computing confidence intervals for estimated fluxes (Antoniewicz, 2013).

Although promising the “Application of MFA technique using 13C” poses several challenges as follows:

a) Dilution of isotopic labels of intracellular metabolites

b) High flow rate of extracellular metabolites such as lactate may result in lower intracellular labeling due to the high exchange rate between intracellular and extracellular lactate.

c) Compartmentalization of metabolism since some metabolites like pyruvate, citrate, malate, and amino acids are involved in different metabolic pathways whereas current techniques cannot label them separately. Hence, just the sum of the individual contributions can be quantified (Ahn & Antoniewicz, 2012b).

The second approach is called flux balance analysis that is discussed in the next section.

2.9.1.3 Flux Balance Analysis (FBA) modelling

Flux Balance Analysis (FBA) is one of the modelling methods that has been extensively used for studying metabolic activities in a wide range of systems ranging from microbes to human cells. FBA is categorized among the type of models referred to as "constraint-based" models. By imposing various meaningful constraints on the system, the solution space of an underdetermined system of equations can be reduced to reach a unique solution. The main advantage of FBA type models is their capability of predicting the behaviour of metabolites in metabolic networks without a need for excessive kinetic information about that network. FBA mathematically describes the system under study by developing a matrix of stoichiometric coefficients. This matrix is used for formulating mass balances in terms of the fluxes associated with the different reactions involved in the metabolic network. The initial FBA approach was proposed for describing a steady state situation according to which the production rates of produced metabolites are equal to the consumption rates of consumed nutrients as already presented in equations (2-3) and (2-4). As explained in the former section, the resulting system of stoichiometric equations is usually underdetermined (i.e. more unknowns than equations). To partially address this problem it is considered that cells are trying to maximize/minimize an objective function subject to stoichiometric or other constraints and the resulting optimization problem will force a solution at particular constraints. Accordingly, the FBA can be formulated as an optimization problem as follows:

$$\max \ \boldsymbol{C^T v} \tag{2-6}$$

Subject to $\boldsymbol{S.v} = 0$

$$0 \leq \boldsymbol{v} \leq \boldsymbol{v_{max}}$$

where $\boldsymbol{C^T v}$ is an objective function which has to be maximized, $\boldsymbol{C^T}$ is a vector of weights for each flux in the objective function that reflects the relative contributions of amino acids or other metabolites to biomass and $\boldsymbol{v}$ is the set of reaction fluxes that are subject to constraints. Constraints like the availability of substrates are usually the most common ones used in previously reported work. Upper and lower bounds for each reaction flux are associated to reaction rate limitations. As for all FBA problems setting a biological meaningful objective function is very important. For example, maximization of biomass is a common optimization objective that has been used to predict the system behavior in some microorganisms especially for bacteria. It should be noticed that FBA can only solve for the fluxes between reactants and products for each reaction but it cannot provide information about the individual metabolites' concentrations (Mahadevan et al., 2002).

There are some extension models to FBA that will be briefly discussed in the following subsections.

2.9.1.4 Flux variability analysis

Due to a lack of a sufficient number of constraints FBA may lead to multiple solutions. This situation might have a negligible impact on the prediction of the model or it could be an important one depending on the proximity of the solutions to each other and to the importance of the fluxes that are not determined uniquely.

2.9.2 Dynamic Flux Balance Analysis

FBA is only capable of modelling the flux distribution at steady state. Hence, it is not able to determine the evolution with time of the metabolites' concentrations. To capture the dynamics of the system, a dynamic extension of FBA has been proposed that is referred to as Dynamic flux balance analysis (DFBA). In the original DFBA formulation proposed by Palsson group (Savinell & Palsson, 1992) it was assumed that the fluxes are piecewise constant, i.e. they remain constant during short time intervals. Hence, the resulting fluxes calculated at each time interval T can be used to calculate the concentrations in a recursive fashion using the Euler discretization of the corresponding dynamic balance as follows:

$$\Psi_{i+1} = \Psi_i + (\mathrm{S}.v_i)\ X_i.\Delta t \qquad \text{(T= time interval)} \qquad (2\text{-}7)$$

For example, a dynamic model for the diauxic growth of *E.coli* on glucose and acetate (Mahadevan et al., 2002) was formulated by using DFBA as follows:

@ each sampling time *i* (2-8)

max $c^T v$

Subject to $v_i > 0$

$$\frac{1}{\bar{X}}\frac{d\Psi}{dt_{higher}} \leq \frac{1}{\Delta t.X_i}(\Psi_{i+1} - \Psi_i) = \mathrm{S}.v_i \leq \frac{1}{\bar{X}}\frac{d\Psi}{dt_{lower}}$$

$$\Psi_{i+1} = \Psi_i + (\mathrm{S}.v_i)\ X_i\ \mathrm{T} \qquad \text{(T= time interval)}$$

where here $\bar{X}$ is an average biomass concentration over a given period of time (e.g. duration of exponential phase or post-exponential phase), X_i is the biomass at time i, $(d\Psi/dt)_{higher}$ and $(d\Psi/dt)_{lower}$ are higher and lower bounds on consumption or production rates. These limiting metabolic rates can be defined as a function of concentration/s of metabolites participating in the corresponding reaction.

These rate constraints are determined from experiments. T is a time interval value used for discrete numerical integration (Budman et al., 2013).

The key advantage of DFBA, and the reason that is chosen for the current research is that it often needs a limited number of constraints to describe the evolution of metabolites thus often requiring a small number of parameters to describe the system (Mahadevan et al., 2002). This may be a significant advantage as compared to kinetic models that were reviewed in the previous sections and that require many parameters thus making them sensitive to noise and also to regular FBA models that do not model the process dynamics. These advantages make DFBA a promising modeling method for optimization purposes but its application to experimental systems poses several challenges. First, the models that have been proposed in the literature such as the *E.coli* model given above, generally involve a small number of metabolites. For this case, the constraint and objective function are generally determined by trial and error. On the other hand when the number of metabolites that need to be modeled is much larger as compared to the *E.coli* model given above, finding the biological function and the constraints necessary to explain the experimental data is much more challenging. Finding systematically the minimal number of constraints of the DFBA model is a main focus of the current work.

2.10 Challenges regarding the DFBA modelling of mammalian cells

Modeling of mammalian cells is generally more challenging as compared to other microorganisms such as bacteria. Mammalian cells are subject to death mechanisms such as apoptosis (programmed cell death) that significantly limit their long term viability. Also, due to their slow growth as compared to bacteria, experiments are slow with the resulting increased chances of contamination while viability limitations impose serious constraints on the ability to collect data for model calibration.

Bacterial cells are generally grown in relatively simple chemically defined media. Most flux constraints that have been used within metabolic models of bacteria were generally related to the uptake rate of a major ingredient such as glucose. On the other hand, in mammalian cells, media is much more complex. This complexity combined with the fact that manufacturers of medium for mammalian cells do not reveal the media composition poses challenges for the formulation and calibration of mathematical models.

2.10.1 The choice of the objective function

FBA type models have generally adopted the optimality assumption of evolutionary biology whereby cells are using resources optimally to maximize or minimize a specific objective function in order to survive. A common objective function used especially for bacterial systems is the maximization of growth or final biomass at the end of a batch (Orth, Thiele, & Palsson, 2010). However, several other objective functions also have been investigated such as minimization of the production rate of redox potential, minimization of ATP production rate, maximization of ATP production rate, maximization, and minimization of nutrient uptake rate, maximization of biomass yield per unit flux) maximization of ATP yield (maximal energy efficiency), minimization of the overall intracellular flux, maximization of ATP yield per flux unit (maximizing ATP yield while minimizing enzyme usage), maximization of biomass yield per flux unit (maximizing biomass yield while minimizing enzyme usage), minimization of glucose consumption (more efficient usage of substrate) and minimization of reaction steps (minimization of the number of reaction steps for cell growth) (Schuetz, Kuepfer, & Sauer, 2007). The FBA modeling studies involving different objective functions have been mostly done based on the central metabolic network or genome-scale network of *E.coli*. Defining the right objective function has been reported to greatly impact the model prediction accuracy. In certain studies the minimization of

the production rate of redox potential under nutrient limitations, the linear maximization of the overall ATP or biomass yields have been found to more accurately predict the system behavior as compared to the objective of maximizing growth or final biomass (Pramanik & Keasling, 1997; Schuetz et al., 2007)

.

Mammalian cells might accomplish different functionalities during their time evolution comparing to bacteria. A previous study for finding a suitable objective function on hybridoma cell central metabolism studied the following possibilities: 1) minimization of ATP production, 2) minimizing total nutrient uptake and 3) minimizing redox metabolism through minimizing NADH production. That study suggested that while not a single objective could solely rule the cell behavior, minimizing the NADH production can better describe the typical characteristic behavior of hybridoma cells such as their inefficient use of nutrients. Such inefficient use of resources translates into higher rates of consumption of glutamine and production of alanine (Savinell & Palsson, 1992) .

Burgard et al. proposed a systematic approach using a bi-level optimization principle in which while at the upper level minimized the sum of square errors between the experimentally measured fluxes and the predicted fluxes, FBA model equations were used in the lower level (Burgard & Maranas, 2003). Where v_j represent the flux of reaction j (obtained from optimization problem), v_j^{exp} experimentally determined flux of j,c_j are weights determining the contribution of each flux to the objective function which are adjusted by the outer optimization problem such as to minimize the sum of the squared error between the experimental and optimization flux rates and S_{ij} is the stoichiometric coefficient of metabolite i in reaction j. The work was based on experimental data from the central metabolic network of *E.coli* using isotopomer analysis. The cost of the lower level optimization problem in equation (2-7) is a linear combination of fluxes multiplied by different weights where the latter are also decision

variables. For example, if the correct objective function following optimization would be the maximization of growth, those fluxes which contribute to biomass production would be multiplied by the weights corresponding to the relative contributions of fluxes to growth while all other weights would be zero. This approach can be potentially used to both identify the hypothesized objective function and to examine how the objective function may change by changing environmental conditions (Burgard & Maranas, 2003). It should be noted that this latter work was done for continuous operation where the system is at steady state.

To the knowledge of the authors, systematic approaches to find the most suitable objective function for dynamic cultures have not been proposed. The bi-level optimization formulation given above has been used in the current thesis for comparison with the new methodologies proposed in the current work. Furthermore, Dynamic Metabolic Flux Models have not been formulated as yet for mammalian cells and it remains an open area of research that will be addressed in the current work.

Chapter 3

Methodology

3.1 Cell line

Two cell lines were used: dihydrofolate reductase-deficient CHO cell line (dhfr- CHO) able to produce anti-RhD MAb that was originally provided by Cangene Corporation (Mississauga, ON) and a CHOZN ZFN-modified GS-/- CHO cell producing anti-rabies IgG provided by SAFC/MilliporeSigma (St Louis, MO). The latter has been reported to provide much higher titer as compared to the former.

3.2 Cell culture media

Both cell lines were cultivated in serum-free media. The medium used for (dhfr- CHO) was HyClone SFX-CHO by Thermo Fisher Scientific (Waltham, MA), a liquid serum-free medium designed for CHO cells. For CHOZN ZFN, the medium used was EX-CELL-CD CHO Fusion by SAFC/MilliporeSigma (St Louis, MO). Since (dhfr- CHO) cannot to synthesize glutamine, and cells need glutamine for growth, 1mM glutamine (Thermo Fisher Scientific, Waltham, MA) was added to the medium before its use in the cell culture experiments. CHOZN ZFN is a GS cell line which can to synthesize glutamine so addition to the medium for growth is not required.

3.3 Seed culture

To prepare the seed culture, a cryogenic vial of cells (containing 1ml of cells at a density of 5-10 $\times 10^6$ cells/ml) was rapidly thawed in a water bath at 37 °C under a laminar flow hood and quickly diluted in a 125 ml shaker flask containing 30 ml of pre-warmed serum-free media.

3.4 Cell count

To determine the cell density and viable cell density of the culture a Trypan blue exclusion method was used. 100 µl of Trypan blue (0.4%) was mixed with 100 µl of the cells (in their medium) and a 20 µl of this mixture was loaded into a hemocytometer. The total number of cells and the blue stained (viable) cells and the blue stained (viable) cells were counted under a microscope.

3.5 Sub-culturing

Based on growth rate, the cells should be passaged 2 to 3 times a week. The passaging step involves removing some part of the old medium (containing cells) and adding a new medium to the flask to lower the cell density to approximately 0.25×10^6 cells/ml at any single passaging operation. This procedure maintains the cells within their exponential growth phase. During the exponential growth phase, the cell viability was above 95%.

3.6 Weaning of the cells (adaptation to low serum media)

(dhfr- CHO) used in this study were an adherent cell line and required 10% serum. To adapt it to the suspension condition in spinner flask, it was necessary to gradually decrease the amount of the serum in the medium. For this purpose, a sequential process was used. First cells were cultured in was monolayer in a T-flask containing cell culture medium to which 10% fetal bovine serum (Invitrogen, CA, USA) was added. Following two sub-cultur steps under these conditions and once the cells were identified to be in their growth phase with more than 90% viability, they were sequentially sub-cultured gradually and adapted to 7.5%, 5%, 2.5% and finally 1% serum. For each of these subcultures, the cells were dislodged by trypsin-EDTA solution (Sigma, MI, USA), centrifuged at 100 X g for 4 min and the trypsin-EDTA were aspirated from the flask. After adapting to 1% serum in T-flask, cells in their mid-

growth phase with more than 90% viability were transferred to a spinner flask with a cell density of around 0.3 million cells/mL and the experiment was started (Radford, Niloperbowo, Reid, & Greenfield, 1991).

3.7 Metabolites analysis

Lactate, ammonia, glucose and glutamine were analyzed using a BioProfile 400 (Nova Biomedical, MA, USA).

3.8 Quantification of mAb

Quantification of monoclonal antibody was done in MilliporeSigma (MI, USA) by using a ForteBio (Pall ForteBio LLC, CA, USA).

3.9 Amino acid analysis

To remove the cell debris from the cell culture medium, samples from the cell culture medium were centrifuged at 4500rpm for 4 minutes. The amino acid analysis was performed with a pre-column derivatization method using 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) Waters AccQFluor Reagent (Waters, MA.USA). Sample preparation was done by mixing 70 µL of AccQFluor borate buffer in a 10 µL volume of sample. After vortexing the sample for approximately 10 seconds, 20 µL of AccQFluor reagent was added to the mixture and further vortexed, for 10 seconds the mixture was incubated for 1 minute at room temperature and later it was placed in a water bath at 55^{o} C for 10 min. This step was done to hydrolyze the excess AccQFluor reagent to 6-Aminoquinolone (AMQ), N-hydroxysuccinimide (NHS) and carbon dioxide. A five-point standard calibration curve was prepared by successive dilution and analysis of a mixture of standard amino acids that contains a 2.5 mM of each of the hydrolysate amino acids and 1.25 mM of cysteine (Sigma Aldrich, MI, USA).

For amino acid analysis, a Waters 1525 chromatography system (Waters Corporation) with a fluorescence detector (W2475; Waters 1525; Waters Corporation) was used. An excitation wavelength of 250 nm and an emission wavelength of 400 nm was applied. Also, the gain value was set to 1.0 and a filter value of 0.5. The volume of each sample was 5 µl that was provided by placing an insert inside the sample bottles. A Waters AccQ.Tag Amino Acid Analysis Column C18 (4 µm; 3.9 mm × 150 mm), specifically designed for this amino acid analysis method was used. As this analysis involves separation of 19-22 different amino acids and compounds, the procedure requires attention to the following steps:

1) The column is kept at conditions suggested by the manufacturer,
2) The water used for buffer preparation must be Millipore grade water with a resistivity of 18.2 MΩ.cm (at 25°C)
3) All the glassware including the buffer glasses is washed carefully,
4) All the buffers should be filtered and degassed properly before usage. Buffers being used are Buffer A: an acetate-phosphate buffer which is a mix of 100 mL AccQTag A concentrate with 1-liter Millipore water) and Buffer B: a mix of acetonitrile and water (60:40%).
5) The column temperature was set to 37 ºC and the column was conditioned with buffer B for 10 minutes at a flow rate of 1 mL/min. Before every run, the column was equilibrated for 10 minutes at a 1 mL/min flow with 100 % of Buffer A (Carrillo-Cocom et al., 2015; Cohen, 2000). A Gradient table for the HPLC runs is presented in Table 3-1.

Table 3-1 Gradient table for the HPLC runs, buffers AccQ.Tag buffer A and B

Time (min)	**Flow rate** (mL/min)	**A%**	**B%**
Initial	1.0	100	0
0.5	1.0	98	2
15	1.0	93	7
19	1.0	90	10
32	1.0	67	33
33	1.0	67	33
34	1.0	0	100
37	1.0	0	100
38	1.0	100	0

Chapter 4

Perfusion process development

4.1 The basics of perfusion operation

The perfusion mode of operation unlike the batch and fed-batch operating modes is considered a continuous operation. In perfusion operation the spent medium is continuously removed from the bioreactor and replaced by fresh medium, while cells are kept inside the bioreactor using a cell retention device. During medium replacement, the nutrients consumed as a result of cell metabolism are usually replenished by feeding of the new medium. A key advantage of perfusion processing is that lactate and ammonia, the most common by-products of cell metabolism that usually have a toxic or inhibitory effect on cell growth and protein production, are regularly removed. Thus perfusion operation generally promotes higher cell growth and lower cells death thus leading to much higher cell densities. Since the product titer is generally correlated with cell mass, higher cell density in the bioreactor generally translates to higher productivity per volume. The perfusion operation is generally implemented in two phases. During the first phase, the goal is to achieve a certain level of cell density while in the subsequent second phase the goal is to keep the cell density at a steady state (plateau) for a longer period of protein production. The level of cell density during this second phase is usually adjusted by regulating the perfusion rate and harvest rate and by bleeding out a certain amount of cells, including the dead cells. The overall equation for the perfusion process is shown in equation (4-1).

$$\frac{\mathrm{dX}}{\mathrm{dt}} = \mu \mathrm{X} - k_D \mathrm{X} - \mathrm{D} X_H - \frac{CBR}{V} X \qquad (4\text{-}1)$$

where μ is the specific growth rate, k_D is the cell specific death rate, X is the concentration of biomass, D is the dilution (perfusion) rate, $\mathrm{X_H}$ is the biomass concentration of the harvested stream. If the cell

retention device has 100% efficiency in retaining the cells this value should be equal to zero (Konstantinov et al., 2006).

In the early days of the modern biopharmaceutical industry, i.e. during the 1980s and 1990s, perfusion systems were initially introduced and they attracted significant attention as on those days the product yields of the available industrial cell lines were only a few milligrams to few hundreds milligram. Hence, perfusion was considered as an attractive option for producing a larger amount of products. The most common cell retention devices used at that time were spin filters, which will be introduced in the next section. However, those devices were not really efficient, especially due to scale-up challenges. When product yield was significantly improved as a result of cell engineering advancements, perfusion systems became less popular compared to fed-batch operations due to the relative operating complexity, additional hardware requirements and potential contamination of perfusion systems. For some products that were not stable, such as recombinant Antihemophilic Factor VIII, perfusion remained almost the only viable process option because the medium containing the expressed protein must be taken out of the bioreactor and sent directly for purification before it denatures inside the bioreactor.

Beyond these special applications, perfusion is gaining attention once again mainly because the development of new cell retention devices that are scalable. Other factors that have motivated the use of perfusion are the need for increasing amounts of therapeutic proteins, the availability of novel single-use bioprocess systems that make it possible to use only one manufacturing line for the production of multiple products and advancement in continuous purification systems to match the continuous production from perfusion (Bonham-Carter, 2011; Jacquemart et al., 2016).

Beyond their use in the large-scale production of proteins, in recent years perfusion processes have been found to be very useful in development of personal cancer treatment especially in immunotherapy using chimeric antigen receptor-modified T cell (CAR-T cell) in which it is critical to produce sufficient

amount of patients' modified T cells in a short amount time (Levine, Miskin, Wonnacott, & Keir, 2017; Wang & Riviere, 2016).

4.2 Cell retention devices

The lack of effective scalable cell retention device has been the main obstacle for using perfusion mode in bio-manufacturing operations. Cell retention devices can be based on different physical principles such as filtration, gravity, and acceleration. A cell retention system should be able to separate close to 100% of the cells from the effluent stream and it has to keep this efficiency even at high-density cell conditions (Woodside, Bowen, & Piret, 1998). Several factors should be considered regarding the choice of cell retention systems such as cell retention and cell density capacity, the residence time of cells in the bioreactor, mechanical shear stress and the retention of the protein of interest inside the bioreactor. Some of the cell retention devices have a re-circulation loop in which the medium containing cells is circulated inside this loop and only part of that stream will be extracted out for the protein harvesting. However, this recirculation process combined with the use of peristaltic pumps may result in detrimental mechanical shear stress on cells. The cell retention time during cell separation is another factor that may have negative effects on cell culture conditions as during this time cells may be exposed to stressful culture conditions. For example, oxygen shortage might occur during the time the cells are within the cell retention devices (Chotteau, 2015). Some of the cell retention systems available are reviewed below.

4.2.1 Gravitation/Acceleration based devices

These devices are working based on acceleration or gravity where the media is usually pumped through a peristaltic pump to these settlers. These devices are normally used for anchorage-dependent cells

attached on micro- carriers due to the higher solid density in these cases as compared to suspension cell systems without micro-carriers that cannot be effectively separated with gravitational devices.

4.2.1.1 Inclined settler

Inclined settlers consisted of inclined plates on which the cell sedimentation occurs. The cell culture medium containing cells is pumped into the lower part of the settler. Then, sedimented cells are sent back to the bioreactor and the cell-free media is pumped out from the top part of the settler to the harvest tank. The main advantage of inclined settlers is that cell bleeding, i.e. elimination of dead cells from the bioreactor, is possible in this type of settling device. Also, inclined settlers contribute negligible shear stress to the cells. On the other hand, cells in an inclined settler usually adhere to the plates resulting in an increase of the residence time within the settler in the range of 30 minutes to few hours. To mitigate this problem, vibration can be applied to the system and also the cell metabolism can be manipulated to inhibit the cell adhesion to the settler plates. Manipulation of the metabolism might be accomplished by using a heat exchanger to lower the temperature of the media that goes to the settler. Despite these possible modifications, the cell perfusion capacity within inclined settlers is relatively low and thus scale-up is considered a challenge (Chotteau, 2015; Woodside et al., 1998).

4.2.1.2 Acoustic Settler

In this type of settlers, acoustic waves are applied to a chamber where the cell broth has been previously pumped in. The cells are concentrated to a particular density so they experience resonance at the frequencies of the imposed acoustic waves. When the waves are interrupted, the aggregated cells settle rapidly, falling back into the bioreactor. This settling device needs significant power to generate the

waves and the transfer of this vibrational energy to the cells may damage them. The use of this device has been reported in pilot scale applications only (Chotteau, 2015).

4.2.2 Filtration based devices

Filters have been developed and used for separation of mammalian cells from culture medium but fouling remains a major challenge for long-term operation. Different modifications have been proposed to address fouling.

4.2.2.1 Spin-filters

Spin filters consist of a cylindrical membrane rotating within the bioreactor around the same centerline axis as the impeller. The spent culture medium is drawn from the culture volume into the cylinder and then out of the reactor while the cells are retained by the membrane. Spin-filters differ from cross-flow filters in two major aspects: the fluid flow relative to the filtration surface is produced by the rotation of the cylinder and the filtration surface is typically inside the reactor. Hence, the trans-membrane pressure drop can be manipulated independently from the cross-flow velocity, the retentate cell concentration is more uniform over the whole filter surface and the cells remain within the controlled environment (Woodside et al., 1998).

4.2.2.2 Cross-flow filters

Cross-flow filtration or tangential flow filtration is a common filtration methodology that has been used in many applications. Unlike normal dead-end filtration where the flow passes through the filter, in cross-flow filtration the feed flows tangentially to the filter surface. Hence, in tangential flow filtration, fouling is reduced as compared to processes involving dead-end filters and thus the process can be operated continuously since intermittent washing is not required.

ATF perfusion system (ATF: Alternating tangential flow) is a particular perfusion system manufactured by Repligen that has been used in the current research. Membrane fouling is significantly reduced in this system due to the use of a special diaphragm that produces an alternating tangential flow (Kelly et al., 2014) .Hollow-fiber cartridges are typical choices of cross-flow filters for cell retention in mammalian cell perfusion applications. In these systems, the suspension is pumped to an external cartridge and it is concentrated as it flows across a membrane. The concentrated material containing the cells is recycled back to the reactor, while the cell-free permeate exits the system. The pressure drop between the inlet and recycle streams drives the cross-flow across the membrane and determines the surface shear rate. The resulting trans-membrane pressure difference is the highest near the inlet of membrane and it may even be reversed at the filter outlet thus generating a non-uniform permeate flux. The permeate flux has been reported to be the most important parameter in determining the filter's fouling.

Among all the aforementioned cell retention systems, ATF perfusion system is currently viewed as especially advantageous as it is usually easier to use and control and it is amenable to scale- up. Thus, the ATF perfusion system has been chosen for the perfusion experimentation conducted in the current thesis.

4.3 Challenges in the operation of a perfusion bioreactor:

4.3.1 Level control:

As mentioned earlier, in perfusion operation the media is continuously substituted by fresh medium. In order to keep the level in the bioreactor constant during this exchange of contents the flow of fresh medium entering the bioreactor should match the flow of spent medium. Although the use of the same

peristaltic pump and the same tubing would be expected to result in a balance of the flows, due to several factors such as fatigue in the tubing (Chotteau, 2015) or differences between the pressure in the bioreactor and the harvest tank this balance is not always maintained. The resulting deviations from the balance between fresh medium and spent medium flowrates might cause during long term operation to overflow or to a critical reduction of the medium level inside the bioreactor. To address this problem two solutions are available. A first option is to control the level by using by adjusting the feeding pump based on a level sensor's measurements. This method is only partially effective because due to foaming at the top of the bioreactor, especially at higher cell density, the readings from the level sensor are highly inaccurate. A second solution used in the current work is to place the bioreactor on top of a scale and use the weight as an indicator for the level inside the bioreactor. Based on the weight a level controlling system was designed using a PID controller that based on the weight measurement sends corrective actions to a feeding peristaltic pump.

4.3.2 Oxygen and Carbon dioxide transfer

Providing enough oxygen for the cells in a perfusion operation is a key challenge, especially in a smaller scale bioreactor, since cell density is substantially higher than a batch or fed-batch operation (up to hundred times). As a result of this high density, the overall oxygen consumption rate (OUR) is generally high as per the balance equation for oxygen (equation 4-2). Also, according to the dynamic balance of oxygen within the bioreactor (equation 4-3), it is evident that a higher oxygen transfer rate (OTR) is required to keep the oxygen level above the critical point of dissolved oxygen (C in equation 4-2) which is normally 20- 50 % of saturation for CHO cells.

$$\text{OUR}=q_{O_2}X \qquad (4\text{-}2)$$

$$\frac{dC}{dt} = OTR - OUR \quad (4\text{-}3)$$

The OTR is often empirically calculated based on the volumetric transfer coefficient ($k_L a$) as given in equation 4-4. According to this equation, it is possible to increase OTR by increasing $k_L a$. Some of the factors affecting $k_L a$ are the rotational speed of the impeller, the bioreactor geometry, temperature and pressure and the amount of surfactants used to mitigate foaming.

$$OTR = k_L a(C_s - C) \quad (4\text{-}4)$$

Although increasing the impeller speed (rpm) is an easy way to increase OTR it should be noted that mammalian cells are more susceptible to damage resulting from shear stress. Thus, there is a limit on the speed of impeller that usually depends on the type of cells being cultured. Also decreasing the bubble size by manipulating the sparger operation results in more surface for oxygen transfer with a resulting increase of $k_L a$ but it also produces foam that has a detrimental effect on cells. Manipulating the amount of surfactants is also a common practice to increase $k_L a$ but it also has drawbacks. First surfactants may be toxic to the cells and also its addition to filter-based perfusion systems might cause clogging of the filter.

In addition to oxygen control, regulating the amount of CO_2 in the bioreactor is essential for effective operation of mammalian cell cultures. As a result of the cell metabolism carbon dioxide usually accumulates inside the bioreactor and lowers the pH. For example, Gary et al suggested that when pCo2 was in the range of 30-76 mmHg it better supported the protein production in CHO cells growing in perfusion culture and higher level of pCO2 (> 105 mm Hg) had an inhibitive and destructive effect on the cell growth (Gray, Chen, Howarth, Inlow, & Maiorella, 1996). An initial project of the current thesis addressed this issue and is reported later in Chapter 7. It was found that the accumulation of carbonic

acid caused the reduction of pH to values that were found to be inhibitory to cell growth in spinner flask cultures of CHO cells.

To tackle OTR limitations and accumulation of CO_2 we have proposed in the current work to use two different type of impeller with different agitation profile, one Rushton at the bottom and one Marine type closer to the surface of the liquid in the bioreactor. Following this design, the oxygen had a higher chance of being dissolved in the medium.

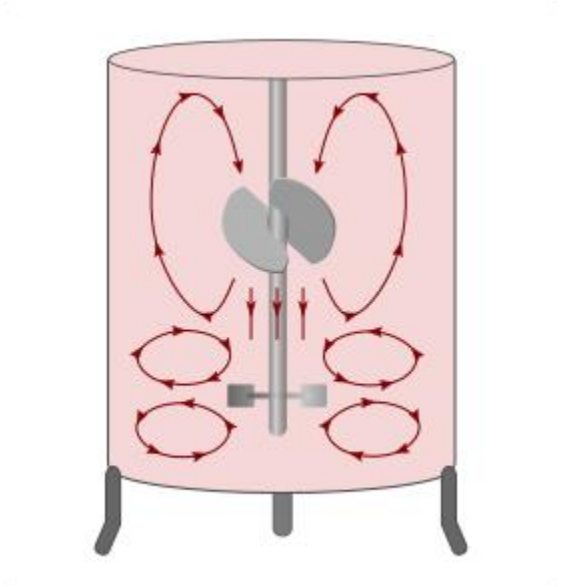


Figure 4-1 Idealized mixing profile in the bioreactor with two different impellers

Also, we used a ring sparger that by producing larger bubble size helped to expel the CO_2 from the bioreactor and in addition, a microsparger with smaller pore size was used to provide enough oxygen to the culture. Thus, the effect of the addition of the microsparger was to increase the $k_L a$ as compared to the values obtained with the ring sparger alone.

At the earlier batch phase occurring before the perfusion phase, the ring sparger was sufficient. This sparger is connected to the air supply and as mentioned above promotes the stripping of CO_2 from the

culture while providing enough dissolved oxygen (DO). During the second phase of operation when perfusion starts pure oxygen was provided through the microsparger while the ring pipe sparger was still providing regular air. The reason for switching the flow from air to pure oxygen was to satisfy the higher oxygen demand at this stage thus maintaining a suitable air/oxygen ratio that ensures the right level of DO while avoiding excessive foam formation inside the bioreactor.

4.3.3 Antifoam addition

Excess foam formation is detrimental to cell growth and protein. Bursting of the bubbles within the foam can damage the cells and decrease the cell viability (Routledge, Hewitt, Bora, & Bill, 2011). Inside the bioreactor, foaming can block the filters and increase the pressure inside. Addition of antifoam is a common practice in bioprocess engineering to decrease the foam formation. However, there are some points that should be considered regarding antifoam usage. Antifoam addition will decrease the oxygen transfer inside the bioreactor, and addition of antifoam could be an issue for the downstream and purification process because antifoams are known to foul the membranes. Also depending on the cell line, some antifoams might have toxic effects on the cells. Thus, it is generally recommended to perform a toxicity test for the antifoam to be used in the bioreactor by adding a different concentration of the antifoam to the cell line in the flask and assessing the resulting toxicity.

4.3.4 Effect of peristaltic pumps and tubing

In the perfusion operation used in the current work, the cells were extracted from the bioreactor by using a peristaltic pump and then they were filtered using an ATF filter. Both the pumping and filtering operation result in shear stress that is potentially damaging to the cells. One way to reduce this shear stress was to use tubes with larger diameters, however it could result in an increase of the loop volume.

A useful strategy utilized to mitigate the shear effects was to use tubes with a larger diameter but with a shorter length that while reducing the shear stress they are also able to minimize the residence time in the circulation loop. It should be noted that generally it is suggested to keep the residence time less than 2 minutes (Chotteau, 2015).

4.4 Materials and methods

4.4.1 Cell preparation for inoculum

For this study, a CHOZN ZFN-modified GS-/- CHO cell line (SAFC/MilliporeSigma, MO, USA) was used. Cells were thawed and suspended in a 125 ml flask containing 30 mL of medium, shaker condition of 36.9^{o}C, 5% CO_2 and rpm of 100 were used during the cell expansion. After two days, cells were dispensed to two 125 ml flasks with the cell density of 0.3 million cells/ml and were fed with fresh medium. Cell expansion continued for almost two weeks so as enough cells were available for inoculation of the bioreactor and that the cells were in their growth phase with more than 95% viability. Cell count and cell viability were performed using a hemocytometer under the microscope and the Trypan blue method (Strober, 2001).

4.4.2 Preparation of the bioreactor and ATF perfusion system

The bioreactor's vessel was connected to different sections of tubing required for feed, inoculum, harvest, perfusion, cell bleed, gases, base addition and antifoam. Each of these lines was connected to a disk filter with an easy connector (Figure 4-2).

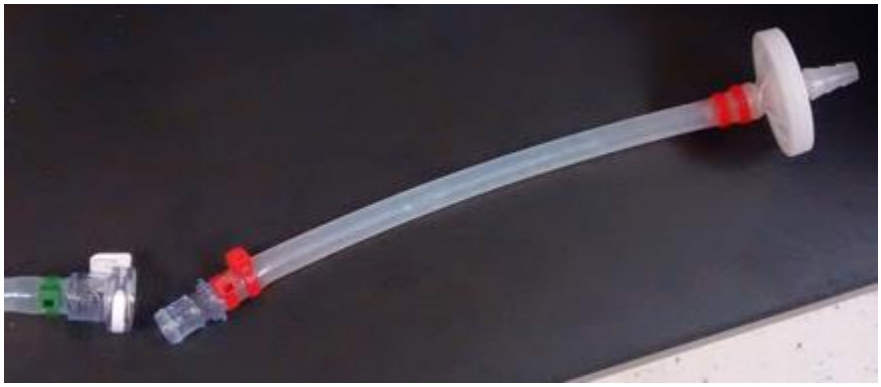


Figure 4-2 End tube connection to disk filter.

For each run the ATF perfusion's filtration set-up was equipped with a new diaphragm and connected to the bioreactor vessel using a dip tube went to the bottom of the vessel that was connected to the top of the filter. The port at the bottom of the filter was equipped with a tube with an easy connector to allow for connectivity to the harvest tank. The bioreactor vessel and the ATF filter were autoclaved while connected together. The bioreactor's vessel was connected to the controller and the probes were calibrated. The pipeline considered for the cell bleed was designed to be closer to the surface of the cell culture media where usually the cell debris was found to accumulate (Figure 4-3).

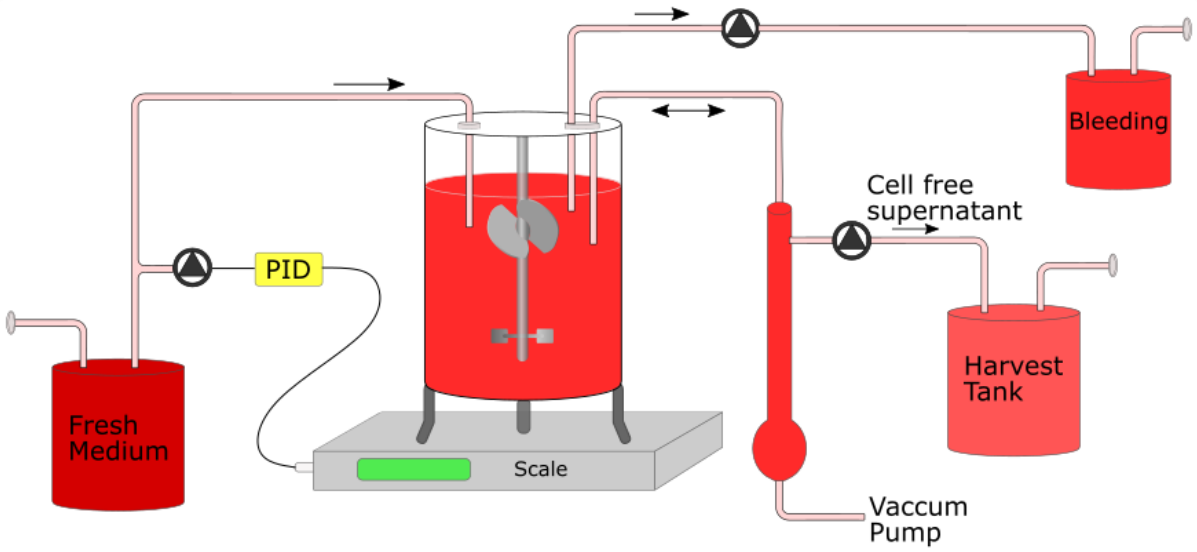


Figure 4-3 Schematic of perfusion set-up.

4.4.3 Bioreactor operation set points

The bioreactor was inoculated to a cell density of about 1 million cells/ml, the temperature was set to 36.9^{o}C , DO to 40%, pH to 6.9 and the impeller agitation speed was set at 150 rpm.

4.4.4 pH control by base addition

Following the above discussion about the need for reducing the CO_2 level in section 4-3-2 with its resulting detrimental effect on pH and cell viability, it was required to add a base to keep the pH at its desired level. For the cell line used in this study, a pH of 6.9-7.0 was recommended by the supplier. Sodium carbonate (Na_2CO_3) and Sodium hydroxide (NaOH) were used as recommended for mammalian cell cultures. Sodium carbonate is milder compared to sodium hydroxide but it is known to increase the osmolality, that could be detrimental to a high cell density culture. Hence, 1M sodium hydroxide was used. As sodium hydroxide could be detrimental to the cell culture at the feeding location

the base addition line was positioned immediately above the impeller, i.e. at a region that is considered to be well mixed inside the bioreactor.

4.5 Perfusion rate

The perfusion rate is an important factor in running perfusion processes. This rate has a great impact on cell growth and protein production and quality since it determines whether enough nutrients are available for the cells and whether toxic by-products are removed fast enough from the bioreactor. From an economic perspective consuming less cell culture medium is highly desired but this consideration should not come with the price of sacrificing cell growth, protein production, and protein quality. Hence, finding the optimum perfusion rate is critical for a successful operation. There are a few common criteria for adjusting the perfusion rate that is discussed here as follows:

1) Based on the level of key nutrients such as glucose
2) Based on trial and error and former experience; this method usually needs extensive experimentation so it was not pursued in the current work
3) Based on a desired cell-specific perfusion rate (CSPR). This is the most common method in the industry where perfusion rate is adjusted based on the cell density and by using the parameter CSPR(pL/Cell/Day) that relates the perfusion rate to the cell density as per the following formula:

$$\text{CSPR} = \frac{D}{C_v} \qquad (4\text{-}5)$$

In equation (4-5), D is the perfusion rate in units of (Reactor volume/Day, RV/Day), and C_v is viable cell density (millions of viable cells/ml, MVC/mL). Generally, the CSPR is kept constant by changing

the perfusion rate (D) as a function of the measured cell density. As mentioned earlier it is desired to use less medium while supporting cell growth and protein production and quality. Thus, keeping the CSPR at the lowest possible rate is an important goal. An acceptable CSPR for the industry is around 40-100 pL/cell/day. Konstantinov and colleagues introduced a method called "push-to-low" that is acceptable for products for which a longer residence time will not adversely affect the protein quality. In this method, CSPR is decreased in an iterative-stepwise fashion to find the minimum CSPR that is also acceptable in terms of cell growth and protein production quantity and quality (Konstantinov et al., 2006).

To find the perfusion rate in this research, the following approach was taken. The two different phases of operation, i.e. the initial growth in a batch operation and the second phase involving perfusion while maintaining a constant cell density were approached differently.

1) For the growth phase, the bioreactor was sampled each day for cell counting. Based on the counted cells and a prediction of the viable cells for the following day was estimated from equation (4-6).

$$C_2 = C_1 e^{\mu t} \qquad (4\text{-}6)$$

2) The prediction provided an estimate of the time that the cell density will reach the desired level at which the second phase of perfusion should be started. For the perfusion operation where the cell density was kept approximately constant, perfusion was started and bleeding of the cells was implemented to remove the dead cells and to keep the cell density at steady-state. The perfusion rate was found based on the desired CSPR. In the current study, we were aiming at a CSPR of 80 pL/day/cell. The harvest rate was calculated from the mass balance based on the perfusion and the bleeding rates.

4.6 Results, discussion, and suggestions for a perfusion run

During a two weeks perfusion experiment, the highest cell density reached at day 12 was 51 million cells per mL (Figure 4-4) starting from about one million cell per mL. The cell viability at this point was about 80%. Subsequently, the viable cell density dropped to around 35 million cells/ml on the last day of this experiment (day 14) with a cell viability of 75%. The perfusion operation started at day 3, when the viable cell density was around 4.7 million cells/ml with around 98% cell viability and the glucose concentration as the main carbohydrate source decreased to approximately 3.2 g/l (~16 mM) from its initial level of 5 g/l (Figure 4-5). As mentioned above, the desired CSPR in this study was 80pL/cell/day and thus based on the measured the cell density the perfusion rate that started at 0.67 RV/Day was increased to around 3 RV/Day. The cell bleed rate was started at day 9 when the cell viability dropped to about 80%. The bleed rate started at 0.07 RV/Day and it was gradually increased to 0.18 RV/Day. Up to day 4, the ring sparger was enough to keep the DO at the set point but after that, the microsparger with pure oxygen was required and at this point, foam formation started to become a challenge. The addition of antifoam (Antifoam C from MilliporeSigma) was done based on the observed layer of foam that formed at the top of the working volume in the bioreactor. The observed thickness of the foam layer was 1 cm at day 5, 3 cm of foam at day 8 and approximately 6-8 cm of foam at day 10. Antifoam addition (0.1 % Antifoam C (MilliporeSigma, MI, USA)) was performed using a peristaltic pump and its flow rate was adjusted based on the relative amount of foam that was observed at the top of the culture.

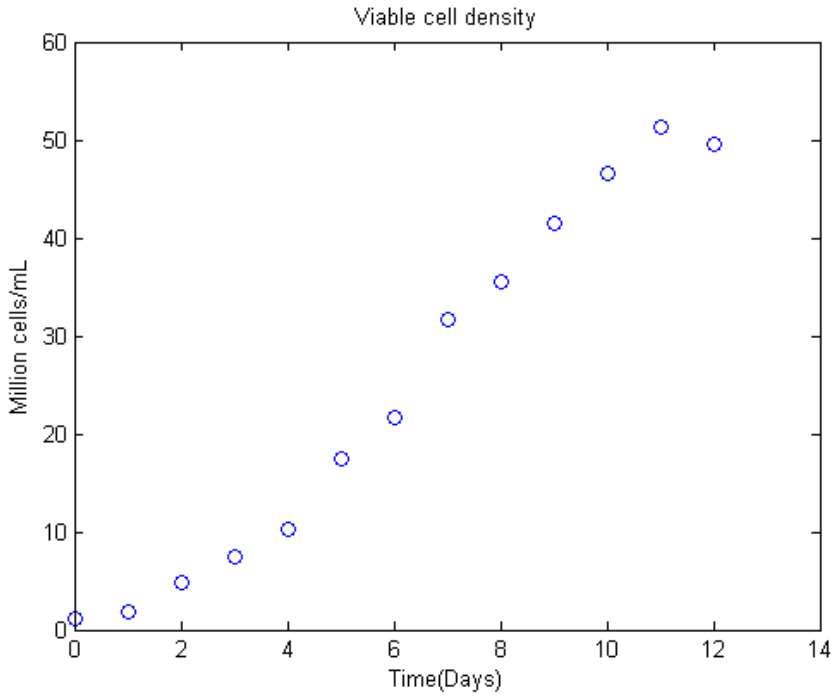


Figure 4-4. Viable cell density vs. time. Cells reach to their maximum cell density of around 50 million cells per mL.

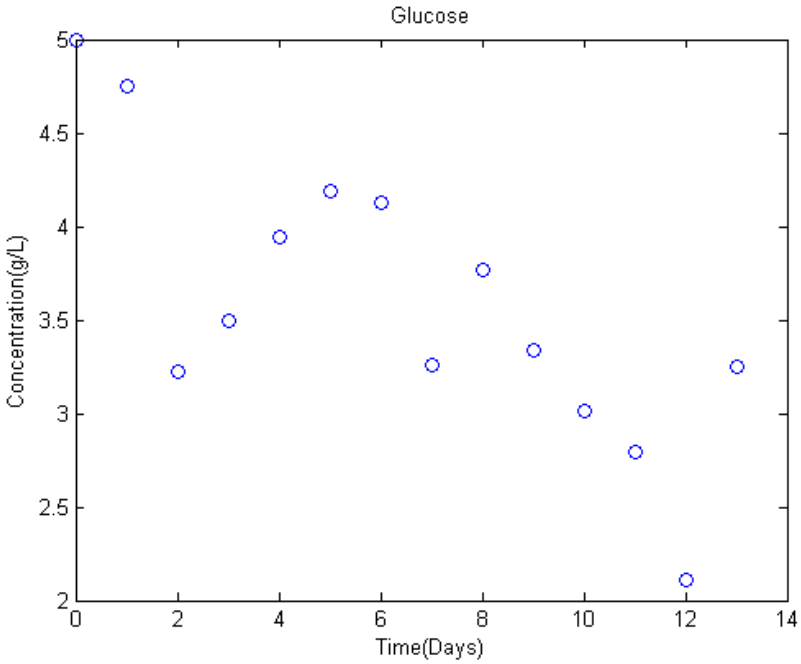


Figure 4-5 Glucose profile in perfusion culture. Glucose level is kept above critical level for cell growth.

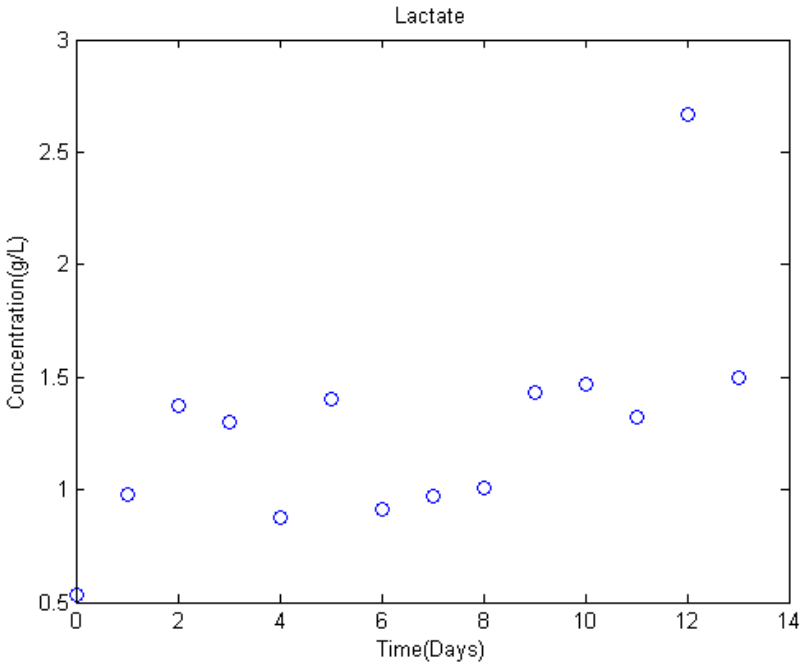


Figure 4-6 Lactate profile in perfusion culture. It was tried to keep the lactate level lower than harmful level.

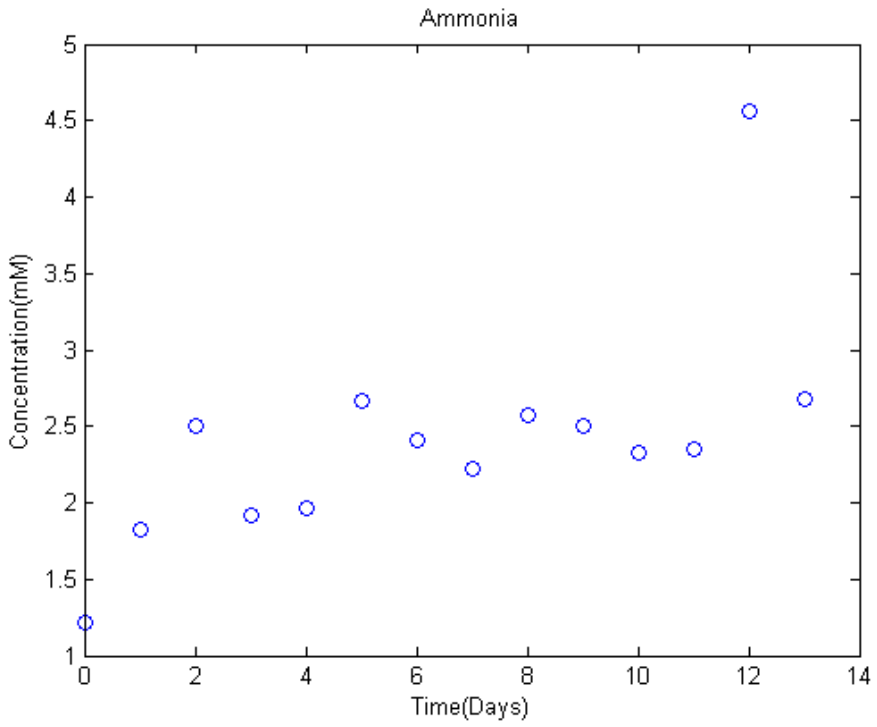
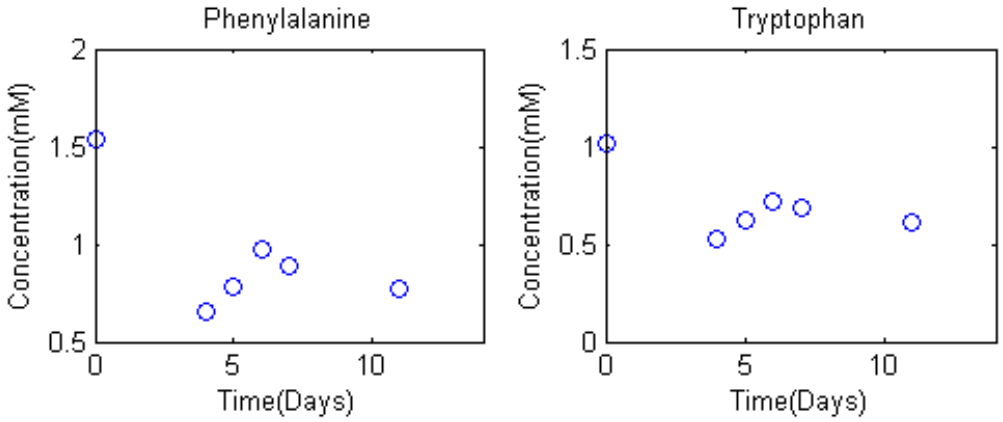
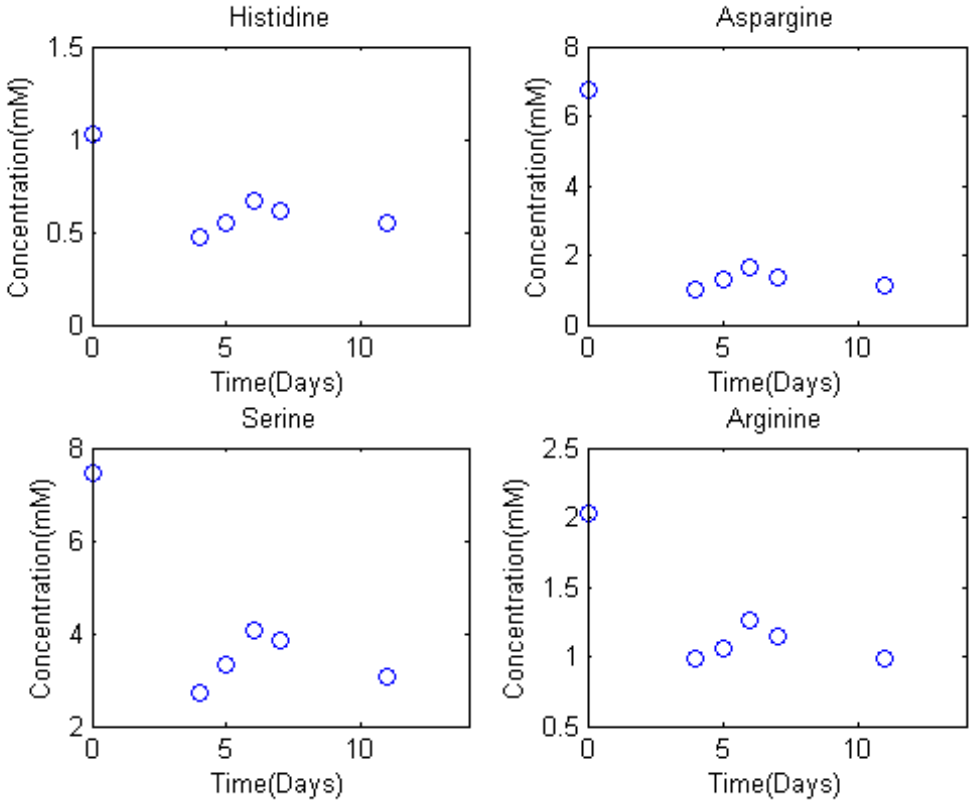
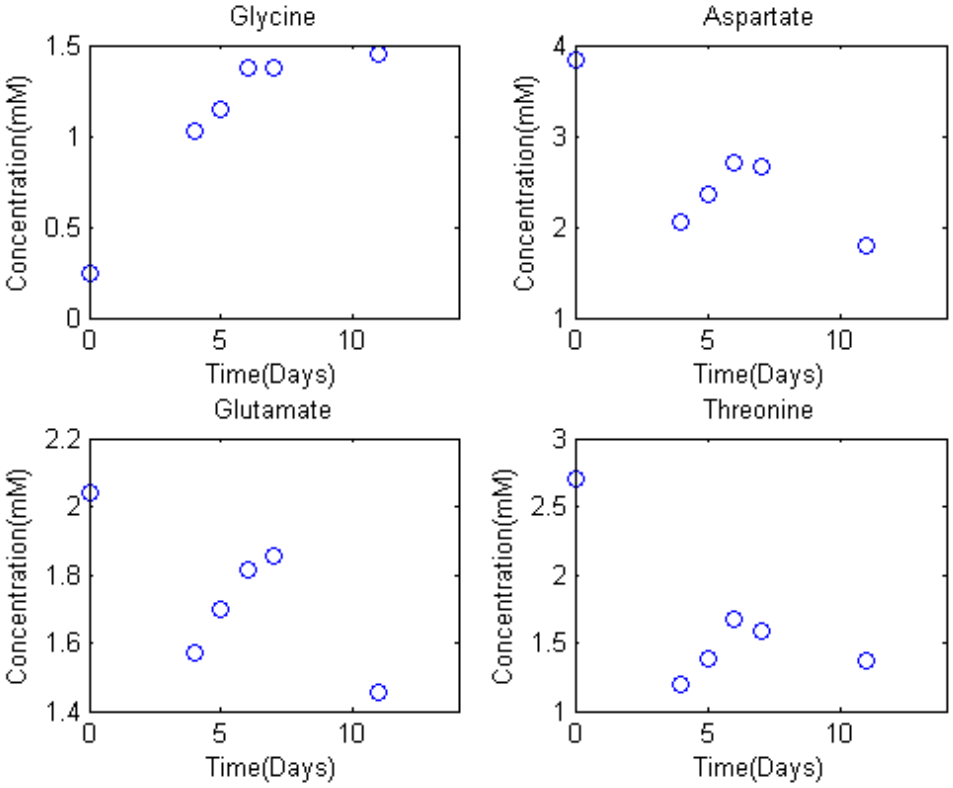


Figure 4-7 Ammonia profile in perfusion culture. It was tried to keep the ammonia level lower than harmful level.




Histidine
Concentration(mM)
Time(Days)
Aspargine
Concentration(mM)
Time(Days)
Serine
Concentration(mM)
Time(Days)
Arginine
Concentration(mM)
Time(Days)


Glycine
Concentration(mM)
Time(Days)
Aspartate
Concentration(mM)
Time(Days)
Glutamate
Concentration(mM)
Time(Days)
Threonine
Concentration(mM)
Time(Days)

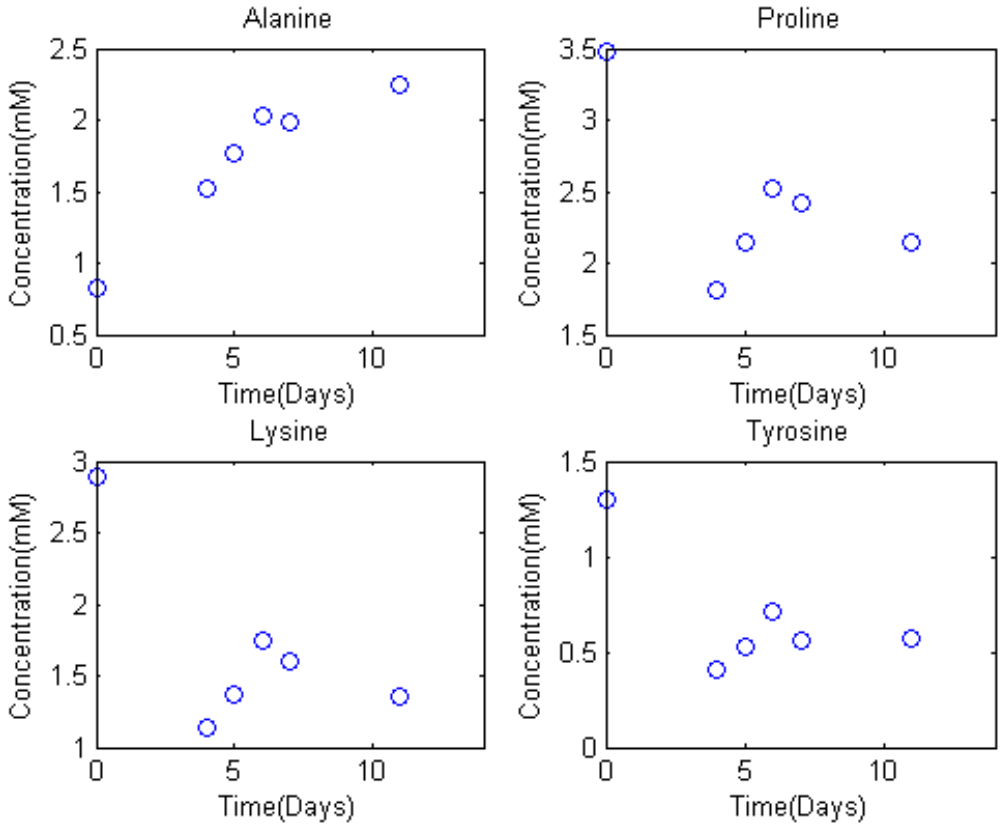


Figure 4-8 Amino acid concentration profile in perfusion culture.

During this perfusion run glucose concentration for most of them were kept above 3 g/l (16.7 mM), which is an acceptable range for glucose concentration. Lactate concentration was mostly between 1-1.5 g/l (11-16 mM) (Figure 4-6) and ammonia concentration was kept mainly under 2.4 mM (Figure 4-7). Asparagine was consumed very quickly during the batch phase from 6.4 mM to less than 2mM (Figure 4-8). Similar results were observed previously by Duarte et al. and the reason might be that asparagine and serine are the major nitrogen sources for the GS-CHO cells (Duarte et al., 2014). Compared to other reported perfusion experiments with CHO cells (Clincke et al., 2013) which reached to about 30 million cells/mL in a two week experiment, in this study we reached a higher cell density

(50 million cells/mL). Among the reasons for this improvement in cell density might be that in this experiment by having the lactate and ammonia as the toxic by-products of the culture concentrations were kept lower, which favored increased cell growth. It has been discussed by several groups (Aghamohseni et al., 2014; Andersen & Goochee, 1995; Butler, 2006) that accumulation of ammonia could impact the protein quality in the culture especially by affecting the glycosylation. It has been shown higher ammonia will decrease the terminal sialylation so in this experiment by controlling the level of production of ammonia the quality of the produced protein has potentially been improved.

Chapter 5

Identification of Active Constraints in Dynamic Flux Balance Analysis

[1]This study deals with the calibration of dynamic metabolic flux models that are formulated as the maximization of an objective, subject to constraints. Two approaches were applied for identifying the constraints from data. In the first approach a minimal active number of limiting constraints is found based on data that are assumed to be bounded within sets, whereas in the second approach, the limiting constraints are found based on parametric sensitivity analysis.

The ability of these approaches to finding the active limiting constraints was verified through their application to two case studies: an *in-silico* (simulated) study describing the growth of *E.coli* and an experimental study for *Bordetella pertussis (B.pertussis).*

5.1 Introduction

Living cells can be described as complex biological systems involving a combination of different molecular elements such as proteins, lipids, sugars, and nucleic acids. These species interact with each other through biochemical reactions involving different metabolites according to pre-defined metabolic networks (Toya, Kono, Arakawa, & Tomita, 2011). Cells use these metabolites for growth and function. The fluxes associated with biochemical reactions involving these metabolites are highly dependent on

[1] This chapter has been published as the paper: Nikdel, A., & Budman, H. (2016). **Identification of active constraints in dynamic flux balance analysis**. *Biotechnol Prog*. doi: 10.1002/btpr.2388

environmental conditions. The ability to mathematically describe the changes in the fluxes due to different environmental perturbations would allow for systematic understanding and optimization of the biological system under study.

Stoichiometric-based metabolic models are an important group of structured metabolic modeling consisting of systems of mass balance equations that are based on stoichiometric relations among metabolites. Stoichiometric models generally assume that the intracellular environment is at a quasi-steady state (Baroukh, Muñoz-Tamayo, Steyer, & Bernard, 2014). Following this assumption, it is possible to establish relations between input and output fluxes, but the resulting model is generally underdetermined since the degrees of freedom are larger than zero (more fluxes than balanced metabolites existed in the metabolic network):

$$d = n-k \tag{5-1}$$

where d is the degrees of freedom, n is the number of fluxes and k is the number of constraints. To partially address the under-determinacy it has been proposed to formulate the problem as a maximization or minimization of a biologically meaningful objective function subject to constraints. This method is referred to as flux balance analysis (Orth et al., 2010; Raman & Chandra, 2009; Sidoli et al., 2004). Because most biotechnological processes are transient in nature, dynamic models are sought to describe them (Ahn & Antoniewicz, 2012a). Most reported dynamic metabolic models are based on dynamic mass balances of dominant metabolites involving kinetic expressions of the biochemical reactions involved in the metabolic network. The identification of this type of models is normally posed as an optimization problem where the unknown kinetic parameters are sought by minimizing the least squared differences between experimental and simulated data (Borchers et al., 2013; Nolan & Lee, 2011). Since the resulting optimization problems are often non-linear and non-

convex, finding the global solutions is challenging. Also, with increasing number of kinetic parameters, these models are often difficult to calibrate and more sensitive to sensor noise (Moles, Mendes, & Banga, 2003).

An alternative to these extensive dynamic kinetic models is to use the dynamic extension of the FBA modeling approach, referred to as dynamic flux balance (DFB), which is the focus of the current study and it can be used to calculate the dynamic evolution of the metabolic fluxes. In contrast to the aforementioned dynamic models that are based on dynamic mass balances of dominant metabolites, DFB does not require extensive kinetic information (Mahadevan et al., 2002; Schuetz et al., 2007). The key idea in the DFB approach is to solve the FBA model at discrete sampling intervals where at each interval we seek to optimize a biological meaningful objective function subject to certain constraints (Budman et al., 2013; Mahadevan et al., 2002). A typical DFB formulation is as follows:

@ each time interval k (k=1,…, T) (5-2)

$$\max_{v_k} \boldsymbol{c}^{\boldsymbol{T}}\mathbf{v_k} \qquad (\mathbf{v_k}=(v_1, v_2, \dots, v_n)_k)$$

Subject to:

$$g(\psi_k) \le |\mathbf{S}.\mathbf{v_k}| \le f(\psi_k)$$

$$\mathbf{v}_k \ge 0$$

Where **S** is the matrix of stoichiometric coefficients, $\mathbf{v}_k$ is the vector of fluxes at time k, $\boldsymbol{c}^T\mathbf{v}_k$ is an expression for a biologically motivated objective function, e.g. growth, ATP consumption etc. that have to be maximized or minimized (objective function), ψ_k represents a concentrations of extracellular metabolites and *f* and *g* are vector functions of the metabolites concentrations at each time interval that could be of Michaelis-Menten or other type.

Hence, it is not necessary with this modeling approach to calibrating kinetic expressions for all the metabolites of interest but only for a few limiting ones participating as arguments in the functions *f* and *g* in (5-2) while the evolutions of the other metabolites and the fluxes are determined by stoichiometric relations.

However, the main challenge for calibrating such models is to identify the limiting constraints that will result in a good fit of experimental data. A study by Mahadevan et al (Mahadevan & Schilling, 2003b) proposed a method to find upper and lower limits on fluxes at steady state but the application of flux models to dynamic situations requires the identification of kinetic constraints that are dependent on metabolites' concentrations. Kinetic and metabolic concentration related constraints are of particular importance for non-genome scale based models where gene regulatory limitations are not explicitly considered in the model (Bordbar, Monk, King, & Palsson, 2014; Mahadevan & Schilling, 2003a; Reed, 2012). A simple DFB model reported for *E.coli* (Mahadevan et al., 2002) showed that few constraints such as rate limits on glucose or oxygen consumption can be sufficient to describe the processes under study. Since the metabolic networks considered in the previous studies were relatively small, the constraints given by functions *f* and *g* in (5-2) were mostly identified by trial and error.

For systems represented by more complex metabolic networks, identifying the limiting constraints by trial and error is challenging. For example, a DFB model of a *Bordetella pertussis* microorganism as used in one of the case studies of the current work is described by over 40 reactions. In previous work by one of the authors (Budman et al., 2013), a DFB model for *Bordetella Pertussis* was formulated where the necessary limiting constraints to explain the experimental data were found by a trial and error based search. This procedure was computationally expensive and prone to errors since it was found that the data could be explained by many combinations involving redundant constraints. Since each of

the constraints has to be expressed by the functions f and g in equation (5-2) and these functions involved kinetic parameters that must be identified from data, including redundant constraints in the DFB model will result in a larger than necessary number of kinetic expressions and parameters thus leading to higher sensitivity to measurement noise. In this work, we are presenting and comparing two data-based approaches for finding the limiting constraints required for formulating DFB models: one based on convex sets and the other is based on parametric sensitivity analysis.

5.2 Materials and methods:

Two different approaches are presented and compared: i- a convex set-based approach and ii- a parametric sensitivity analysis based approach.

i. Sets based approach

The key idea for finding the limiting constraints in the DFB model by this method is to represent the data by convex sets. To this purpose, it is assumed that because of measurement error or unmeasured disturbances, the metabolite concentrations are bounded by upper and lower limits at each time interval for which data is collected. For example, typical set constraints for glucose concentrations for the E-coli diauxic metabolism with 10% bounds are depicted in Figure 5-1.

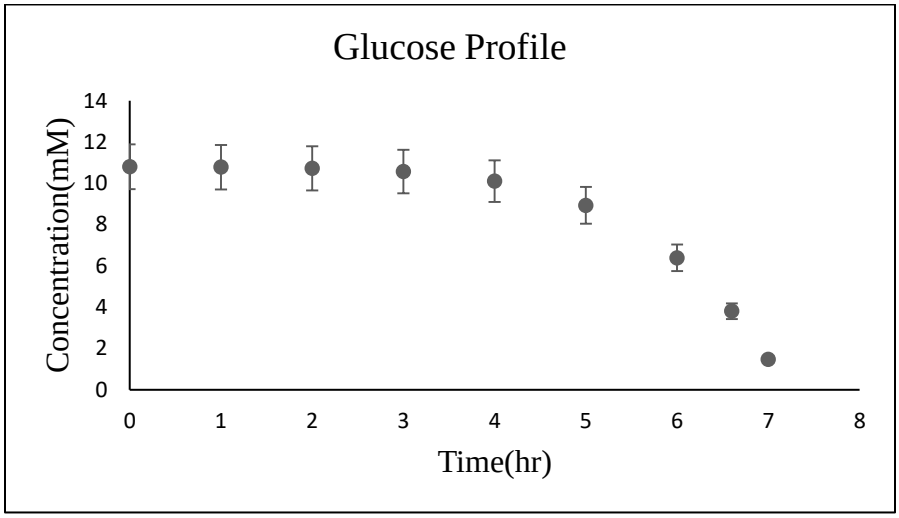


Figure 5-1. Set constraints for glucose concentrations for the *E-coli* diauxic metabolism (Mahadevan et al., 2002) with 10% wide bound.

By this representation, the actual value of metabolite concentrations at each time interval is assumed to be always contained within these constraints and all the possible data points, bounded by them, are referred heretofore as sets. Equivalently the maximal uptake or consumption rates of metabolites can be calculated from the upper and lower bounds on metabolite concentrations and can then be added as linear constraints in the optimization problem formulated for the DFB model.

Using this set-based approach the following 3-step procedure is proposed to find a DFB model to describe the experimental data:

Step 1: Find a flux distribution that maximizes the objective function to be chosen for the process, e.g. maximization of growth subject to the set based constraints. These constraints can be given in terms of upper or lower bounds on concentrations as explained above or as upper and lower limits of uptake rate or production rate of external metabolites. Mathematically, the problem is expressed based on (5-2) as follows:

@ all T intervals (k=1,…,T) (5-3)

$$\max_{\mathbf{v_k}} \mathbf{c^T}.\mathbf{v_k}$$

Subject to:

$$\frac{1}{\Delta t.X_k}(\psi_k - \psi_{k+1}^L) \leq \mathbf{S}.\mathbf{v_k} \leq \frac{1}{\Delta t.X_k}(\psi_k - \psi_{k+1}^U)$$

Where matrix S is the stoichiometric coefficients, $\mathbf{v_k} = (\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n})_\mathbf{k}$ is the vector of fluxes of n metabolites at the time interval of k. $\boldsymbol{c^T}.\mathbf{v_k}$: is the cost to be maximized or minimized (objective function), ψ_k and X_k are representing measured concentrations of extracellular metabolites and biomass at the time of k. The inequality constraint in (5-3) is derived from the discretization of the general mass balance equation for metabolites ($\frac{d\boldsymbol{\psi}}{dt} = \boldsymbol{S}\boldsymbol{v}X$). The inequality in equation (5-3) is for a metabolite being consumed whereas for a metabolite that is being produced, the inequality operators have to be inverted. It should be noticed that in both (5-2) and (5-3) the same objective is optimized but while in (5-2) the constraints are given as a function of the concentrations through the vector functions *f* and *g*, in (5-3) these functions are not known a priori and they have to be identified from measurements. For that purpose, in (5-3) the constraints are set to measured upper and lower bounds on specific rates of consumption/production of metabolites.

Step 2: Find the key metabolites such that by only imposing kinetic constraints on those metabolites the experimental data, represented by convex sets as explained above, can be properly simulated by the DFB model.

For finding the uptake or production rates which are limiting the solution space of the optimization problem, at the optimal point, a Lagrange multiplier approach (Bertsimas & Tsitsiklis, 1997) is used. The optimization (minimization or maximization) of an objective function as in here f ($\mathbf{v_k}$) = $\mathbf{c^T v_k}$ can be formulated as follows:

$$\mathbf{v_k}=\underset{\mathbf{v_k}}{\operatorname{argmin}}\ \mathbf{c^T v_k} \quad (5\text{-}4a)$$

Subject to:

$$G_{LI}(\mathbf{v_k}) \leq 0 \quad (5\text{-}4b)$$

$G_{LI}(\mathbf{v_k}) \leq 0$ are the inequality constraints(lower and upper bounds of the production or the uptake rate of measured metabolites . The Lagrange multipliers (λ) and Lagrange function (Lagrangian) are defined as per the following equation:

$$L(\mathbf{v_k}, \lambda) = \mathbf{c^T v_k} - \lambda\, G_{LI}(\mathbf{v_k})$$

Which is equivalent to the following set of conditions:

$$\lambda \leq 0 \quad (5\text{-}4c)$$

$$G_{LI}(\mathbf{v_k}) \leq 0 \quad (5\text{-}4d)$$

$$\lambda \times G_{LI}(\mathbf{v_k}) = 0 \quad (5\text{-}4e)$$

$$\nabla_{\mathbf{v_k}} L(\mathbf{v_k}, \lambda) = 0 \quad (5\text{-}4f)$$

Where $G_{LI}(\mathbf{v_k})$ is an inequality constraint, $\mathbf{v_k}$ is the vector of decision variables, $\mathbf{v_k}$ is the optimal solution of the fluxes. By solving equations (4c-4f) the value of the Lagrange multipliers (λ) for each of the constraints can be calculated. Equation (5-4e) represents the complementarity condition that forces either the Lagrange multiplier or the corresponding inequality to be zero. The value of the Lagrange multiplier at the optimum is equal to the rate of change in the value of the objective function as the constraint is relaxed. A value of zero for λ implies that by relaxing the constraint the value of the function which is intended to be maximized does not change and thus, this constraint is not active. In contrast, a higher value of the Lagrange multiplier is an indicator that the constraint is active since the objective function changes as a result of deviations from the value of that constraint.

By identifying the active constraints, it is possible to assess which metabolite's consumption or production rates are limiting constraints at each interval of the solution of the DFB model.

Step 3- The goal of the final step of this method is finding kinetic expressions as a function of metabolites that mathematically describe the active constraints found in step 2. To this purpose, at each interval, uptake or production rates of the active constraints are formulated as Michaelis-Menten-type or another type of kinetic model as a function of the corresponding concentration of metabolite associated to the active constraint:

$$\frac{\mathrm{d}\psi_{\mathrm{m}}}{\mathrm{dt}} = \frac{\mathrm{V_{m,max}}.\psi_m}{\mathrm{K_m}+\psi_m} X \qquad (5\text{-}5)$$

Where here $\frac{d\psi_m}{dt}$ is the uptake or production rates of a metabolite (active constraints) identified in step 2 as a function of its concentrations ψ_m at each corresponding time interval and $\mathrm{V_{m,max}}, \mathrm{K_m}$ are the kinetic parameter values to be identified. The consumption rates or production rates of metabolites are assumed in this work to be bounded by a function of the extracellular metabolite concentration (please see the last equation in the set of equation 5-3). This is strictly a model assumption that has been previously done in other reported metabolic flux models e.g. (Mahadevan et al., 2002). The validity of this assumption is assessed by the ability of the model to capture the data. The Lineweaver-Burk and Eadie-Hofstee plots are graphical methods that can be used for identifying the constants in (5-5) (Shuler & Kargi, 1992) if the kinetics is of Michaelis-Menten type. Using the identified kinetic parameters of active metabolic constraints (i.e. uptake or production rate of active constraints), it is possible to generate a predictive dynamic metabolic flux model for given initial conditions of the concentrations of all metabolites. For the case that Michaelis-Menten kinetics is found unsuitable for representing the dependency of the metabolite rate of change with respect to concentration, other nonlinear functions are fitted using the *lsqcurvefit* function in MATLAB.

The key numerical advantages of this set-based approach, are: i- step 1 and step 2 involve the solutions of a linear programming problems and therefore are simple to compute and ii-step 3 involves separated fitting of each metabolite profile as compared to other modelling approaches that require simultaneous fitting of all the metabolites by simultaneous calibration of a larger number of model parameters.

For each of the microorganisms under study, we created the sets of upper and lower bounds of uptake and production rates based on measurements of extracellular concentrations of metabolites at two consecutive time intervals.

For the *E. coli* example *in silico* data generated with a reported model (Mahadevan & Schilling, 2003b) was used for model calibration. The reason for using *in silico* instead of actual experimental data was to test the ability of the proposed methodology to correctly identify the constraints and parameters of the model. To generate the *in silico* data a 10 percent Gaussian random noise was introduced to simulate noisy data and used in lieu of experimental data.

For *B. pertussis*, two sets of experimental data of metabolite concentrations at regular time intervals were available. The available data included amino acids and biomass concentrations at different time points. However, we found that the sampling frequency of the experimental data was insufficient for generating smooth estimates of the uptake rates bounds. Therefore, we applied a non-linear polynomial regression function in MATLAB to interpolate the data and generate smoother estimates of uptake rates.

One of the limitations of this proposed approach is that due to stoichiometric correlations the activation of constraints is sensitive to measurement noise. For instance, in the presence of noise, the concentration of several metabolites may reach constraint values at the same time. To provide robustness on noise related errors we performed the identification of the active constraints as per the 3-step procedure described above for progressively smaller sets of values contained within the original convex sets. For

example, if the maximum noise levels in the data were $\pm 10\%$ with respect to an average value between the upper and lower bound of concentration at each time interval, we performed the identification of the limiting constraints for sets with magnitudes of $\pm 10\%$, $\pm 7\%$, $\pm 5\%$, and $\pm 2\%$ of the measurements. In case that at a particular time interval different limiting constraints were obtained for different assumed noise levels, we selected as limiting constraints the ones that occur most times.

Finally, for both *E.coli* and *B.Pertussis*, as commonly assumed for bacteria (Feist & Palsson, 2010a, 2010b; Meadows, Karnik, Lam, Forestell, & Snedecor, 2010) we considered the maximization of the growth rate as the objective function for the DFB model. The linear programs defined in this section were solved using the *cplexlp* solver ('interior-point' algorithm) of IBM ILOG CPLEX for MATLAB Toolbox.

ii- Parametric sensitivity analysis based approach

We adopted the idea of bi-level optimization that previously was used by Burgard et al. (Burgard & Maranas, 2003) for finding the objective function in FBA models and Raghunathan et al. (Arvind U. Raghunathan, PÉRez-Correa, Agosin, & Biegler, 2006; A. U. Raghunathan, Perez-Correa, & Biegler, 2003) for parameter estimation in FBA for a parametric sensitivity analysis based approach as a comparison to our proposed set-based approach.

The idea behind this approach is to calculate the sensitivity of a function describing the level of fit between data and model predictions on different possible kinetic constraints. Then, the degree by which the constraint affects the quality of fit function is used as an indicator for whether the specific constraint should be considered or ignored in the DFB model. To assess the sensitivity, $\pm 10\ \%$ changes were introduced in the uptake or production rates of external metabolites with respect to a set of nominal values calculated in a bi-level optimization problem as explained below (Model of (Mahadevan et al., 2002). Then, the effects of the deviations from these nominal parameter values on the sum of squared

errors (SSE) of predicted metabolites concentrations were calculated and compared. All possible combinations of constraints on the different metabolites were tested for parametric sensitivity.

The uptake and production rates were initially assumed to be described by Michaelis-Menten type kinetics as a function of the concentration of the metabolite (m) being produced or consumed as in equation (5-5).

The sensitivity of the SSE was calculated with respect to deviations of the maximum uptake or production rate of a metabolite from its nominal value, i.e. changes in the coefficient of the numerators of the Michaelis-Menten expressions in equation (5-5) on their nominal values calculated in an outer optimization level.

A key challenge for the application of this method is to find the nominal values of the parameters to be used in the Michaelis-Menten expressions describing the uptake and consumption rates.

The nominal parameters' values for calculating the sensitivities were obtained from the following bi-level optimization problem:

$$\underset{V_{m_1,max},\dots,V_{m_N,max},K_{m_1,\dots,K_{m_N}}}{\text{Min}} \text{SSE} = \sum_{k=1}^{T} \left(\left(\frac{\psi_{m_{1k}} - \psi^{e}_{m_{1k}}}{\overline{\psi^{e}_{m_{1k}}}}\right)^2 + \cdots + \left(\frac{\psi_{m_{Nk}} - \psi^{e}_{m_{Nk}}}{\overline{\psi^{e}_{m_{Nk}}}}\right)^2 \right) \qquad (5\text{-}6)$$

s.t:

@ each T interval

$$\max_{v_k} \mathbf{c^T}.\mathbf{v_k}$$

s.t:

$$\left| \mathbf{S}^{\mathbf{m_i}}.\mathbf{v_k} \right| \leq \frac{V_{m_i,max}.\psi_{m_i,k}}{K_{m_i}+\psi_{m_i,k}} \ \frac{\text{mmol}}{\text{gdw.hr}} \qquad (i=1, 2,\dots, N) \qquad N: \text{total number of metabolites}$$

Where, $\overline{\psi^{e}_{m_{N_k}}}$ is the mean value of the concentration of metabolite *i* at time interval k. The division by the average was done for the purpose of normalization of each metabolite with respect to the others. In

this bi-level or nested optimization problem, the inner optimization problem solves the DFB model by maximizing (or minimizing) the objective function subject to kinetic constraints. The parameters defining the kinetic constraints (Michaelis-Menten expressions) are calculated from the outer optimization problem in which the sum of square error (SSE) between model predictions and experimental data of metabolites are minimized. This bi-level optimization problem is solved in MATLAB. A linear programming algorithm was used to solve the inner level using an 'interior–point' algorithm implemented by the MATLAB function *linprog* function. The nonlinear upper level optimization problem was solved using the *fmincon* function. Due to the nonlinearity of the optimization problem involving a quadratic objective and nonlinear constraints, we found that the results are highly dependent on the initial guesses. Thus, to avoid local optimal solutions, we conducted a global search over a large range of initial guesses using the *globalsearch* function in the MATLAB Global Optimization Toolbox.

An additional obvious difficulty with the parametric sensitivity approach based on the solution of the bi-level optimization (5-6) is that initially the consumption rates and production rates of all metabolites must be assumed to be limiting thus requiring the calibration of many nominal parameters' values. Although some of these constraints were found ultimately to be redundant and eliminated following the parametric sensitivity analysis, the required initial search of nominal values partially neutralizes the key advantage of DFB models, which aim to describe the system with few constraints. Also, all combinations of constraints have to be tested to eliminate redundant constraints. Accordingly, the parametric sensitivity approach is impractical for problems involving many metabolites. For that reason, it is applied in this work only for the *E.coli* case study that involves four reactions but it is not pursued in the *B. pertussis* system that involves over 40 metabolites.

5.3 Results and discussion

To verify the two approaches presented above, we consider two case studies for two different microorganisms: *E. coli* and *B. pertussis* (Budman et al., 2013; Mahadevan et al., 2002). The set-based approach was applied to both, whereas the parametric sensitivity approach was only applied for *E.coli* for the purpose of comparison with the set-based technique.

For the *E.coli* model, we considered three extracellular metabolites present in the network: glucose, acetate and oxygen as potential limiting constraints and for *B.Pertussis* we considered 16 external metabolites as possible active constraints: alanine, arginine, aspartate, glutamate, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tyrosine, and valine.

The *E.coli* case study was an easy tool to verify the correctness of the methods since the solution was known a priori. The model was used to generate *in silico* data that was artificially corrupted by noise to simulate actual data. Then, we check whether the original DFB model of *E-coli* proposed in (Mahadevan et al., 2002) could be recovered based on the simulated data with noise. For the *B. pertussis* case study, actual experimental data and a preliminary DFB model were available (Budman et al., 2013). However, the latter model was previously calibrated by trial and error as reported in (Budman et al., 2013) and thus it was not known a priori whether the assumed limiting constraints for that preliminary model were the best ones for fitting the available data. Accordingly, the set-based methodology was used to produce a better model that resulted in improved fitting between data and predictions as compared to the previously reported model.

5.3.1 Case study 1: *E.coli*

5.3.1.1 Set based approach

The model assumes four species: glucose, acetate, oxygen, and biomass. The stoichiometric matrix for the problem is given in Appendix A. Glucose and oxygen are consumed throughout the batch culture while acetate is first produced and it is only consumed after glucose is depleted. Initially, a total of three candidate constraints were considered for glucose, acetate, and oxygen.

In order to provide robustness to noise, as explained in the previous section, the identification of the active constraints was done for different levels of noise in the measurements, i.e. for ±2%, ±5%, ±7% and ±10% with respect to the average value of the upper and lower bound at each point. By comparing the Lagrange multipliers for all the possible constraints, i.e. glucose, acetate and oxygen, during the first two hours of the culture only oxygen transfer was found to be a limiting constraint until time=4.2 hours at which time glucose became the dominant constraint and, towards the end of the batch, oxygen again became an active constraint. By the end of the culture, only for the ±10% noise level, acetate was also identified as an active constraint for a few time intervals but for all other levels of noise, it was identified that only glucose and oxygen remain as the active constraints while the acetate constraint remains inactive.

Since the dependency of the identified limiting rate in glucose with respect to the corresponding glucose concentrations appear to be of Michaelis-Menten type the kinetic expressions were obtained using the Lineweaver-Burk and Eadie-Hofstee plots. Lineweaver-Burk, a plot of $\frac{1}{V}$ versus $\frac{1}{\psi}$ (Figure 5-2), was used for finding V_{max}. To find K_m we used the Eadie-Hofstee plot, a plot of V versus $\frac{V}{[\psi]}$ (Figure 5-3) in which K_m is the slope of this plot. Confidence intervals for the parameters were calculated for the

bounds on data corresponding to the maximal level of noise of $\pm 10\%$ error in metabolic data and were as follows: V_{max}= [10.6, 10.7] and K_m= [0.0123, 0.0195]. The constraint on oxygen rate was found to be constant and equal to OUR=13.5

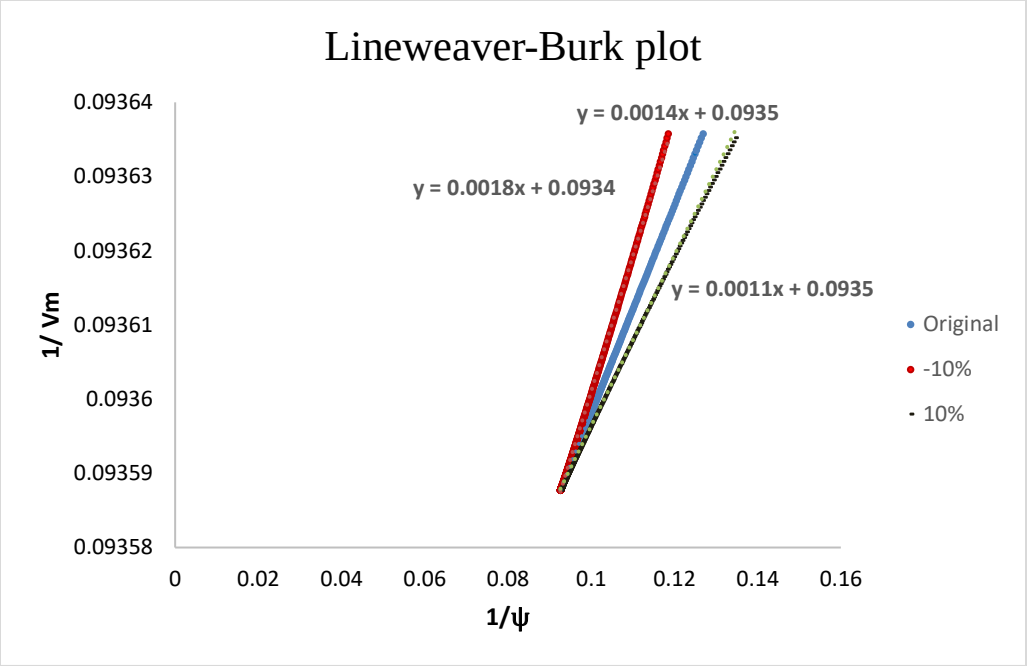


Figure 5-2 Lineweaver-Burk plot for determining V_m. Blue line is plot without noise in the data, red and black lines are the plot with ± 10 noise in the data

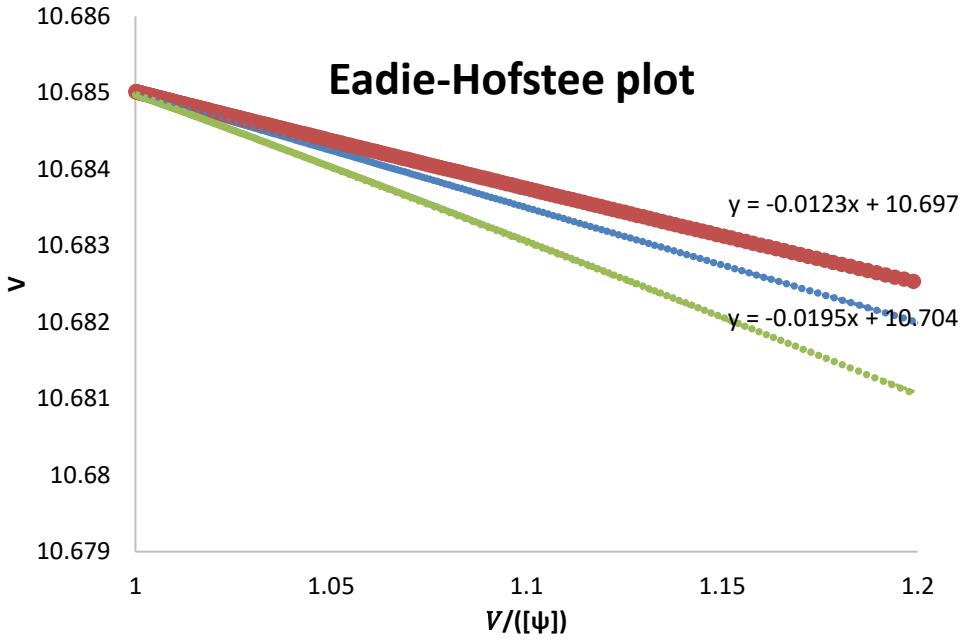


Figure 5-3 Eadie -Hofstee plot for determining K_m. Blue line is plot without noise in the data, red and green lines are the plot with $\pm$10 noise in the data

The two constraints in glucose and oxygen were found to be sufficient for building a DFB model that was able to explain the experimental data. This coincides with the constraints that were considered in the original model used to generate the *in silico* data thus demonstrating that the method identified the correct constraints. The identified parameters are also very similar to the ones used in the original model (V_m= 10 and K_m=0.015). The model with the identified kinetics was used to simulate a batch as shown in Figure 4, which is in agreement with the model results shown in (Mahadevan et al., 2002). It should be noticed that in this process, glucose is preferentially used until it is completely depleted at 7.5 hours. Subsequently acetate is used as nutrient until it is completely depleted at time=10 hours. As illustrated in Figure 5-4, the consumption of oxygen during the glucose consumption period is considerably higher

comparing to the period during which acetate is consumed which is expected from the stoichiometric relations among the metabolites.

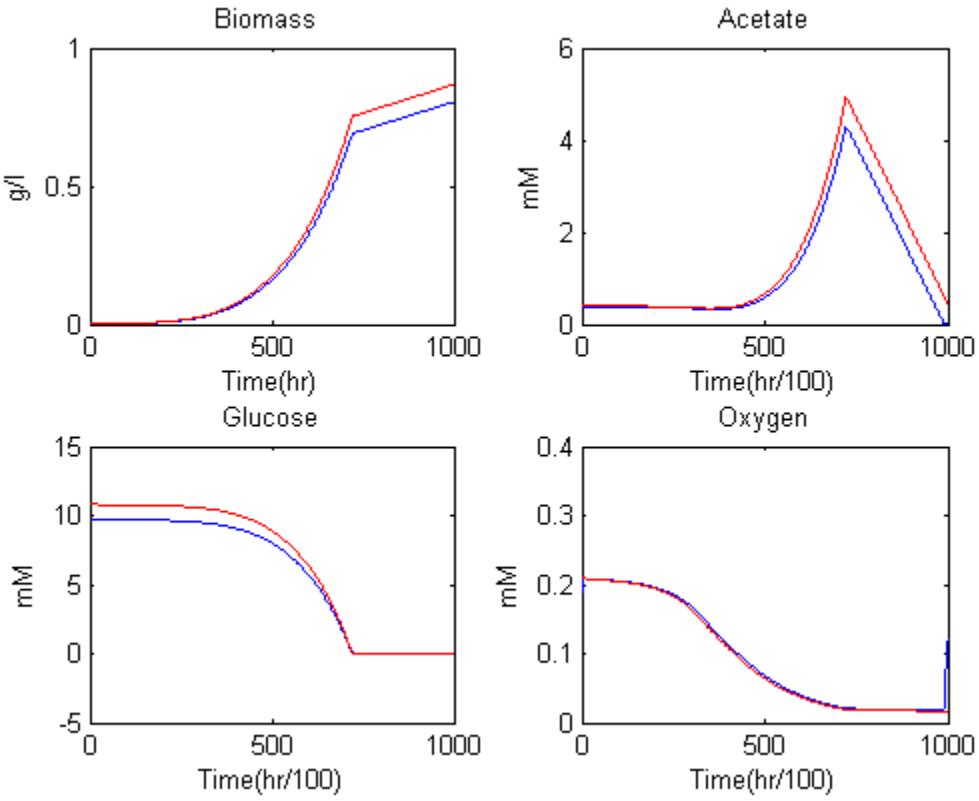


Figure 5-4 Metabolic concentration profiles resulted from the DFB model for *E.coli*; red is the original model and blue is the one that resulted by introducing the noise

5.3.1.2 Parametric sensitivity analysis approach for *E.coli*

Following the general formulation given in (5-6) the nominal parameters values were solved from the following bi-level optimization problem:

$$\min_{V_{Glu,max},K_{Glu},V_{Ace,max},K_{Ace}} \mathrm{SSE} = \sum_{k=1}^{T}\left(\left(\frac{\psi_{Ace_k}-\psi^{e}_{Ace_k}}{\overline{\psi^{e}_{Ace_k}}}\right)^2+\left(\frac{\psi_{Glc_k}-\psi^{e}_{Glc_k}}{\overline{\psi^{e}_{Glc_k}}}\right)^2+\left(\frac{\psi_{X_k}-\psi^{e}_{X_k}}{\overline{\psi^{e}_{X_k}}}\right)^2+\left(\frac{\psi_{O_{2k}}-\psi^{e}_{O_{2k}}}{\overline{\psi^{e}_{O_{2k}}}}\right)^2\right) \quad (5\text{-}7)$$

Subject to:

@ All T intervals

$$\max_{\mathrm{v}} \mathrm{c}^{\mathrm{T}}.\mathrm{v}$$

$$\left|\mathbf{S^{Glc}}.\mathbf{v_k}\right| \le \frac{\mathrm{V_{Glc,max.}}\psi_{Glc,k}}{\mathrm{K_{Glc}}+\psi_{Glc,k}}\frac{mmol}{gdw.hr}$$

$$\left|\mathbf{S^{Ac}}.\mathbf{v_k}\right| \le \frac{\mathrm{V_{Ac,max.}}\psi_{\mathrm{Ac,k}}}{\mathrm{K_{Ac}}+\psi_{\mathrm{Ac,k}}}\frac{mmol}{gdw.hr}$$

$$\left|\mathbf{S}^{\boldsymbol{O_2}}.\mathbf{v_k}\right| \le OUR\ \frac{mmol}{gdw.hr}$$

The nominal parameters' values are listed in Table 5-1 and resulted in a minimal sum of square errors of SSE=94.63. These parameters are identical to the original parameter values used in Mahadevan et al. (Mahadevan et al., 2002). Although the acetate related constraint is not present in the original mathematical model, it was initially assumed to be non-zero with the intention to test whether the parametric sensitivity analysis proposed here shows it to be a redundant constraint.

Table 5-1 Estimated kinetic parameters for *E.coli* model

Parameter	Value
$V_{Glc,max}$	10
K_{Glc}	0.02
OUR	15
$V_{Ac,max}$	0.1
K_{Ac}	2.2

As depicted in Figure 5-1, for glucose, either 10% increases or decreases in the glucose uptake constraint caused a sharp increase in the value of SSE around the minimal value i.e. $V_{Glc,max}$=10. This indicates that the glucose uptake rate is an active constraint around that nominal value of the kinetic parameter. Then, it was found that a 10 % decrease in the oxygen uptake rate (OUR) around the value of OUR=15, increased the SSE from 94.6 to 121.94 (Figure 5-6), but increasing the oxygen uptake rate did not affect the SSE. Similarly, it was observed that a 10% decrease in $V_{Ac,max}$ from a nominal value of 2.2, resulted in an increase in the SSE from 94.6 to 101.86 (Figure 5-7). On the other hand a 10% increase in $V_{Ac,max}$ did not affect the SSE indicating that beyond this value its impact on the SSE is null. When only a glucose constraint was used the SSE increased to 206.13. On the other hand, when constraints on both glucose and acetate were used without a constraint on oxygen the SSE was 95 slightly higher than the value of SSE obtained when using glucose and oxygen together which resulted in an SSE = 94.69. Therefore, it was concluded that the acetate constraint is redundant because it cannot be used to further reduce the SSE as compared to the use of a glucose and oxygen constraints.

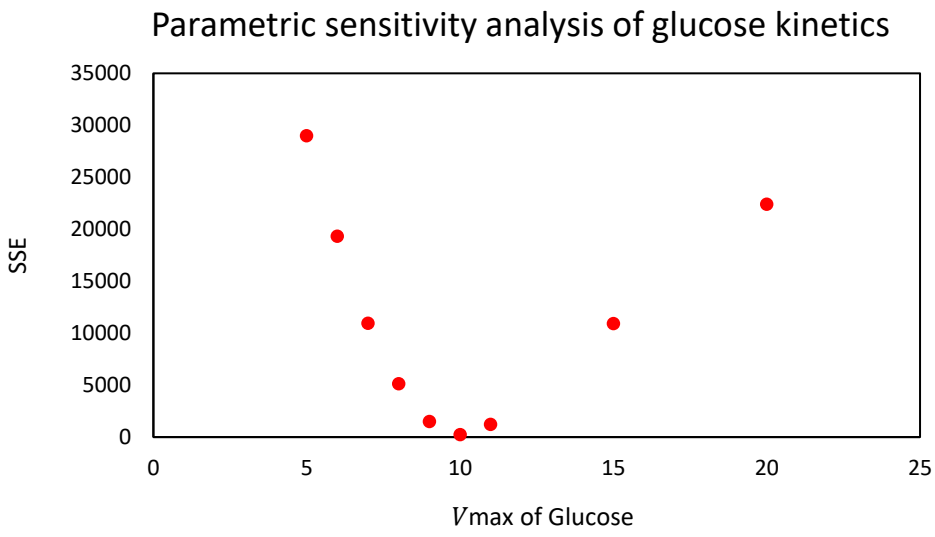


Figure 5-5 parametric sensitivity analysis of glucose kinetics. With perturbation in V_{max} of glucose around the estimated value, SSE increases sharply.

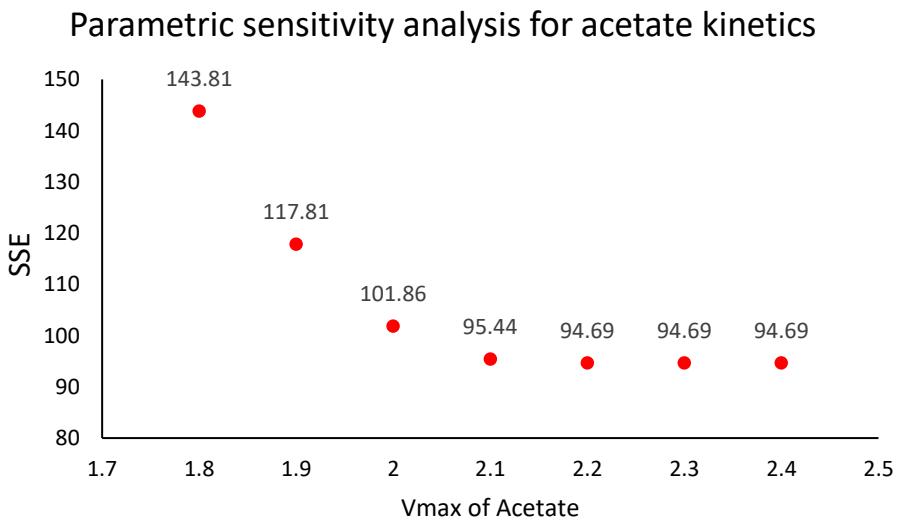


Figure 5-6. Parametric sensitivity analysis of acetate kinetics. Perturbation in V_{max} of acetate around the estimated value does not have a big impact on the SSE value.

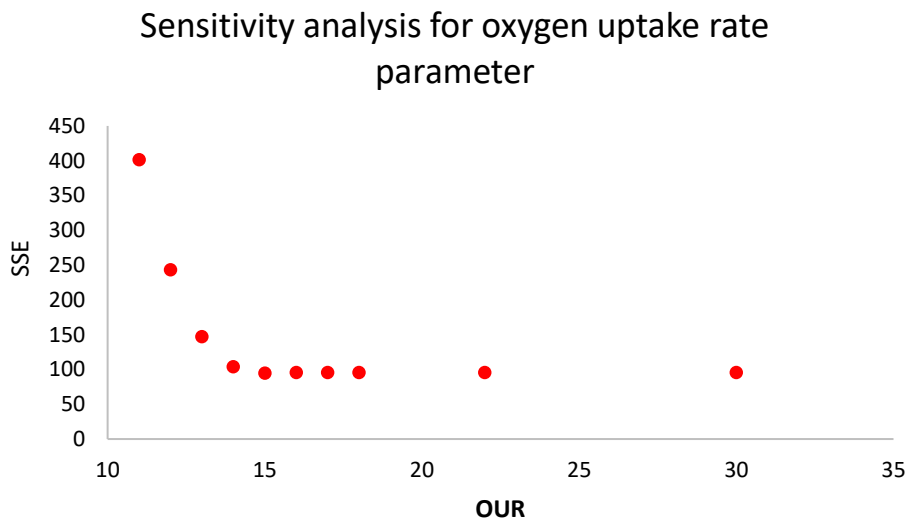


Figure 5-7 parametric sensitivity analysis for oxygen uptake rate.

In summary, it is evident that both the set-based and parametric sensitivity approaches are successful in identifying the limiting constraints for this relatively simple model of *E.coli* comprising only four reactions. However, this case study clearly illustrated that the parametric sensitivity approach is numerically more complex for two main reasons: i- It requires a priori calculation of nominal values for the parameters from a nonlinear bi-level optimization problem and ii- It requires testing different combinations of constraints and their effect on the value of the SSE. This numerical complexity is of particular importance when trying to solve problems of much higher dimensions, i.e. a large number of metabolites, than the *E.coli* case such as the second case study presented below. Furthermore, for simplicity, the parametric sensitivity analysis that was used in this work was local in nature thus correlations among parameters were ignored. Global sensitivity analysis is possible (Rand, 2008) but it will further increase the numerical complexity of the parametric sensitivity approach.

An additional significant advantage of the set-based method is that the kinetic parameters involved in each constraint are identified separately using data related to the metabolite involved in the specific constraint whereas in the parametric sensitivity method the parameters of all the constraints must be identified together by solving a bi-level optimization.

5.3.2 Case study 2: *Bordetella pertussis*

Bordetella Pertussis is a gram-negative bacteria causing the *whooping cough* disease (Thalen et al., 2006). A metabolic network of *Pertussis* has been previously proposed by one of the authors (Budman et al., 2013). The model involves 47 metabolites with 49 reactions as listed in Appendix B. The system under study was operated in batch mode during the exponential growth phase followed by fed-batch operation after the main nutrient, glutamate, was depleted. The feeding of glutamate during the fed-batch phase occurred at a rate 4.3 g/h. Replicated data for 16 amino acids was available at the times of 0, 6, 24, 29, 31.1, 32, 34, 48.3, 51.9 and 55.6 hours.

The application of the parametric sensitivity based approach for this case involving 49 reactions was found to be very challenging since it required the solution of a bi-level non-convex optimization problem with multiple parameters if the uptake/consumption rates of all the measured metabolites would be initially assumed as potential limiting constraints. Therefore, only the set based method was investigated for this case study.

In the set-based approach applied for *B. pertussis*, first, all of the 16 measured external metabolites were considered as potential active constraints. Following the application of the three-step set based approach, the Lagrange multipliers had higher values for phenylalanine uptake rate, compared to the multipliers associated with the other metabolites. However, through the end of culture (greater than 45 hrs), some other Lagrange multipliers related to alanine, serine, leucine, lysine, isoleucine threonine,

proline, valine, aspartate at their higher bound of uptake rate (which correlates to their production rate) are becoming important. Hence, we limited the production rate of these metabolites by imposing a zero production rate constraint.

Based on the dependency of uptake rates of phenylalanine and histidine as a function of its concentration identified in step 2 of the method and shown in Figure 5-8 and Figure 5-9, it appeared that the consumption of phenylalanine could not be described by a Michaelis-Menten type kinetics. Instead, the uptake rate of this metabolite seems to follow an exponential type of dependency between the uptake rate and the corresponding concentration. It should be remembered that the kinetic constraint on the consumption/production rates of metabolites may describe a combination of different reactions thus explaining a kinetic model that is not of Michaelis-Menten type. It seems that saturation will occur at some higher concentration of the metabolites. We used MATLAB *lsqcurvefit* function to fit a Hill function as follows. We have shown a possible saturation behavior of the curve with a dashed line in Figure 5-8.

$$\mathrm{r_{Phe}} = \frac{0.0066{\psi_{Phe}}^{3.2}}{0.022^{3.2}+{\psi_{Phe}}^{3.2}} \qquad (5\text{-}8)$$

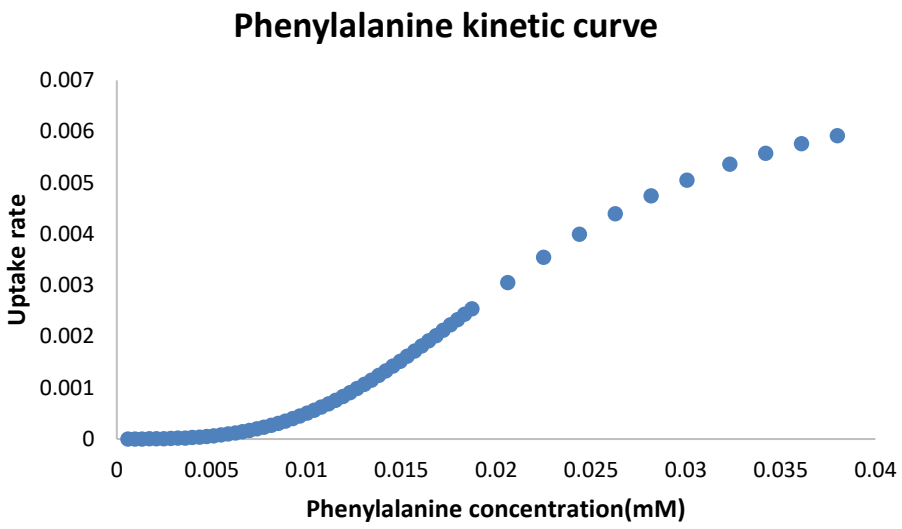


Figure 5-8 Phenylalanine kinetic curve. Data points is based on the nonlinear regression on the experimental data

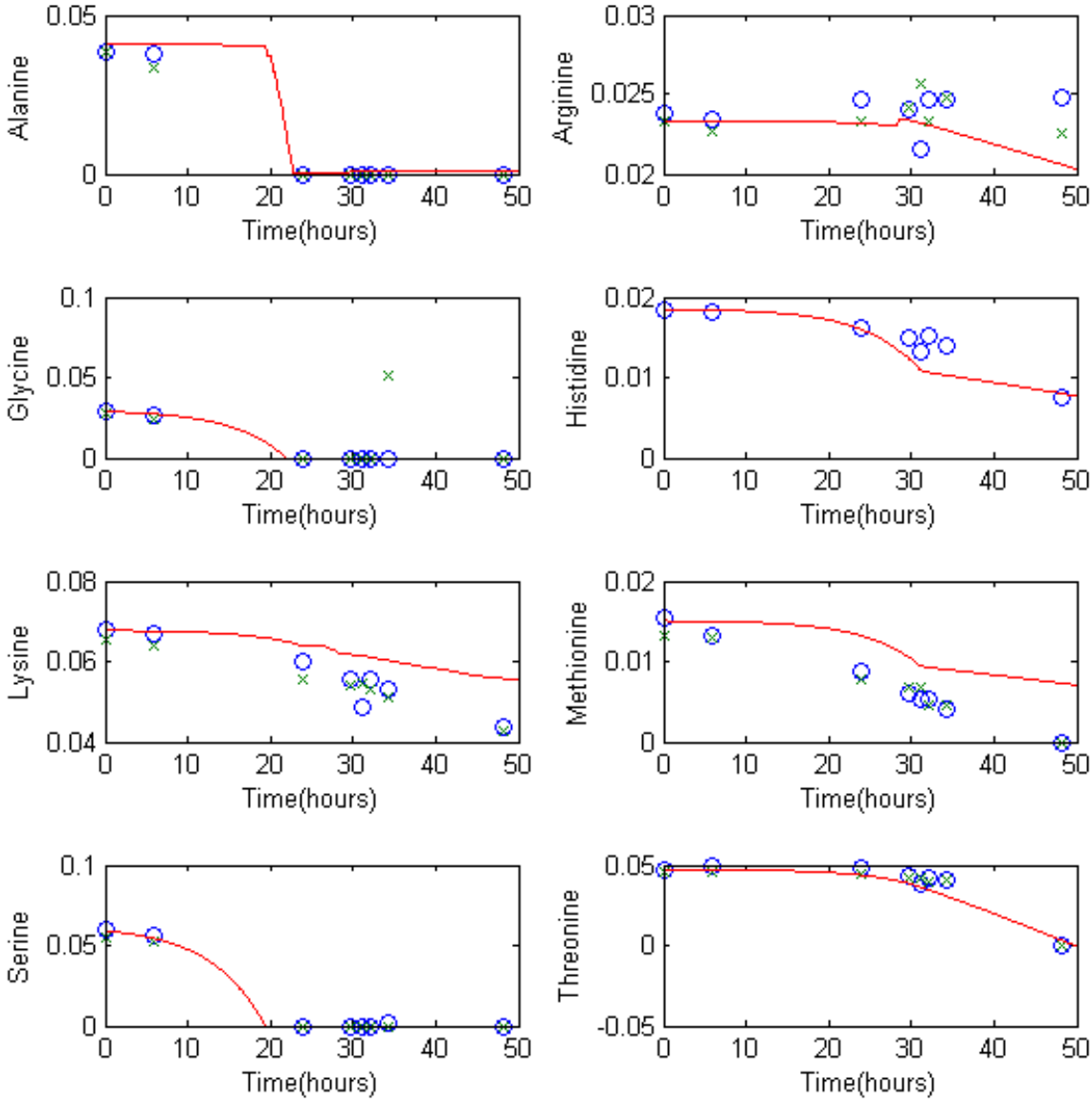
To verify the need for all the constraints, identified by the Lagrange multipliers solution, we investigated two models: i- a model that uses only a constraint on the uptake rate of phenylalanine. ii- a model that uses phenylalanine constraint together with maximal zero production rates of alanine, serine, leucine, lysine, isoleucine threonine, proline, valine, aspartate. When only phenylalanine and was considered as the only active constraint although the resulting model was able to describe most of the experimental data, alanine, serine, leucine, lysine, isoleucine, threonine, proline, valine, aspartate were slightly overestimated. On the other hand, the second model with the mentioned zero production rates' constraints on metabolites above could describe the experimental data with very good fitting accuracy (SSE=8.84) between the model predictions and the experimental data. Chapman et al. (Chapman, Paget, Johnson, & Schwartz, 2016) also have mentioned the possible necessity of such constraints. The metabolic profiles resulted from this new DFB model is shown in Figure 5-9. For

confidential purposes, the data shown in this figure is normalized based on the glutamate concentration at time=0 of the batch. It is noteworthy that this model showed better fitting of the experimental data (SSE=8.84) as compared to the SSE=15.42 obtained with the earlier model reported by Budman et al. (Budman et al., 2013) where the necessary rate limiting constraints were obtained by trial and error. In order to gain an understanding for the reason that a kinetic constraint based on phenylalanine is needed to improve the fit between data and predictions, we conducted the model simulation with all the maximum consumption/production rates but without the kinetic constraint (phenylalanine). The result of this simulation is that most amino acids were depleted much earlier, at about 25 hours from the beginning of the batch instead of 31 hours as obtained with the phenylalanine constraint and as observed in the data. For example, the glutamate evolution without the kinetic constraints is shown in Figure 5-10 and compared to data. In fact, the discrepancies observed between the data and model without the phenylalanine constraint were solely observed at lower concentrations of amino acids but the difference in the sum of square errors is very significant, SSE=252 without phenylalanine constraint versus SSE=8.84 with phenylalanine constraint. The need for the kinetic constraints to describe the evolution of metabolites at lower concentrations values is related to the fact that amino acids are strongly correlated with the biomass formation equation. This correlation is because all the amino acids contribute to biomass and from the fact that the mass balances of metabolites are calculated from specific consumption/production rates multiplied by the biomass concentration. Phenylalanine, cannot be biosynthesized according to the metabolic network and thus the depletion of phenylalanine determines the stoppage of biomass growth thus affecting all other amino acids at low concentration values. Only a few amino acids that contribute to biomass cannot be produced: phenylalanine, histidine and methionine and phenylalanine have the largest share in the biomass composition among these three. It is not surprising that other amino acids that can be synthesized, such as glutamate or proline, are not

limiting constraints since their depletion does not necessarily result in stoppage of growth as it is the case with phenylalanine.

Furthermore, to show the correctness of the model regarding biomass formation, biomass as a function of time obtained with the model is given in Figure 5-11 showing good agreement with data (circles in the figure). It should be noted that biomass was not used for calibration of the model.

Sources of remaining errors between data and model predictions are HPLC measurement error, possible inaccuracy of the metabolic network and inaccuracy of the biomass equation (see (Budman et al., 2013), for how the biomass equation was formulated based on protein/lipids/nucleotides estimated split and elemental analysis of biomass). Also, it should be remembered that the biomass composition may change along with time (Dikicioglu, Kirdar, & Oliver, 2015).


Alanine
0.05
0
0 10 20 30 40 50
Time(hours)
Arginine
0.03
0.025
0.02
0 10 20 30 40 50
Time(hours)
Glycine
0.1
0.05
0
0 10 20 30 40 50
Time(hours)
Histidine
0.02
0.01
0
0 10 20 30 40 50
Time(hours)
Lysine
0.08
0.06
0.04
0 10 20 30 40 50
Time(hours)
Methionine
0.02
0.01
0
0 10 20 30 40 50
Time(hours)
Serine
0.1
0.05
0
0 10 20 30 40 50
Time(hours)
Threonine
0.05
0
-0.05
0 10 20 30 40 50
Time(hours)

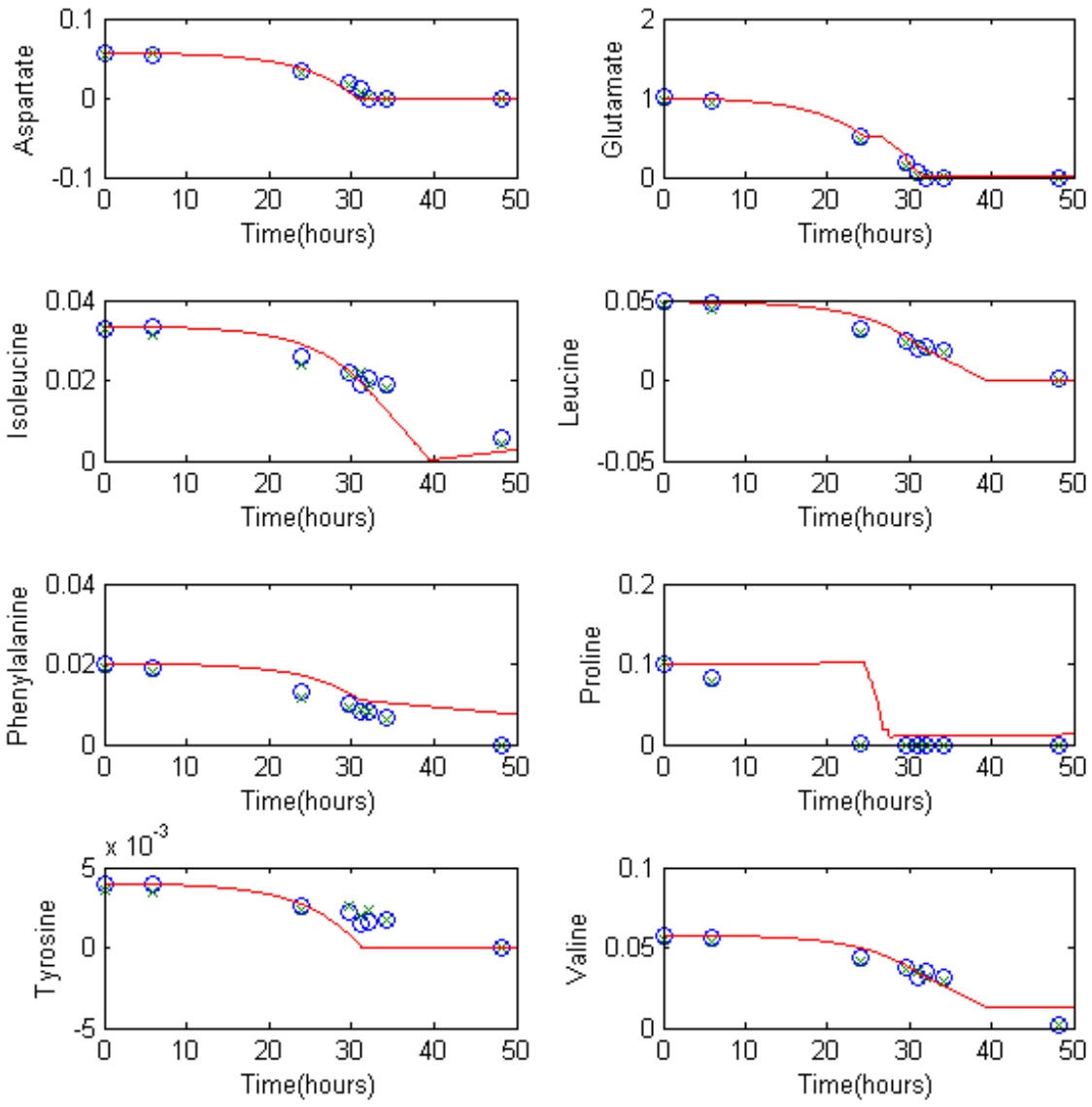


Figure 5-9 DFB model of *B. pertussis* in Fed-Batch mode. Blue circles are the experimental data and red line is the model simulation

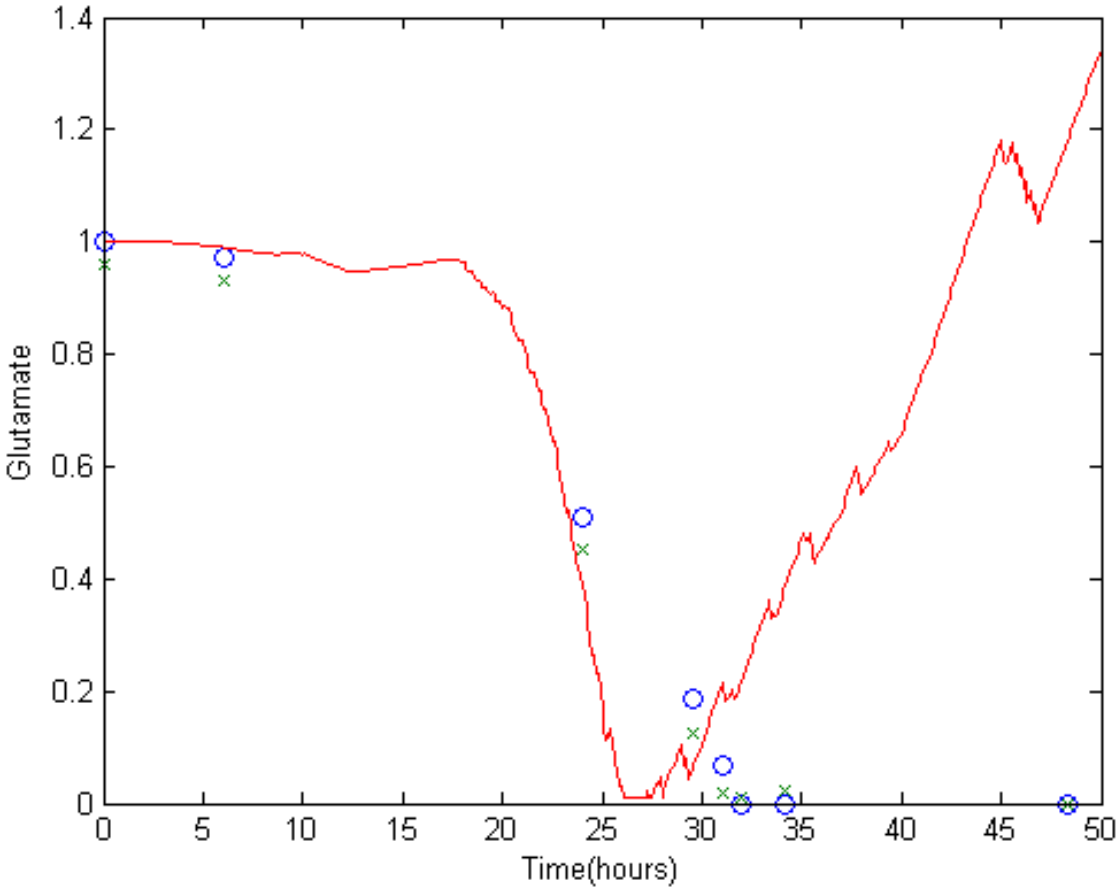


Figure 5-10 Glutamate concentration evolution (without the use of kinetic constraints). As it shows without the use of kinetic constraints, the model is not able to predict the glutamate properly especially after the first day of the experiment

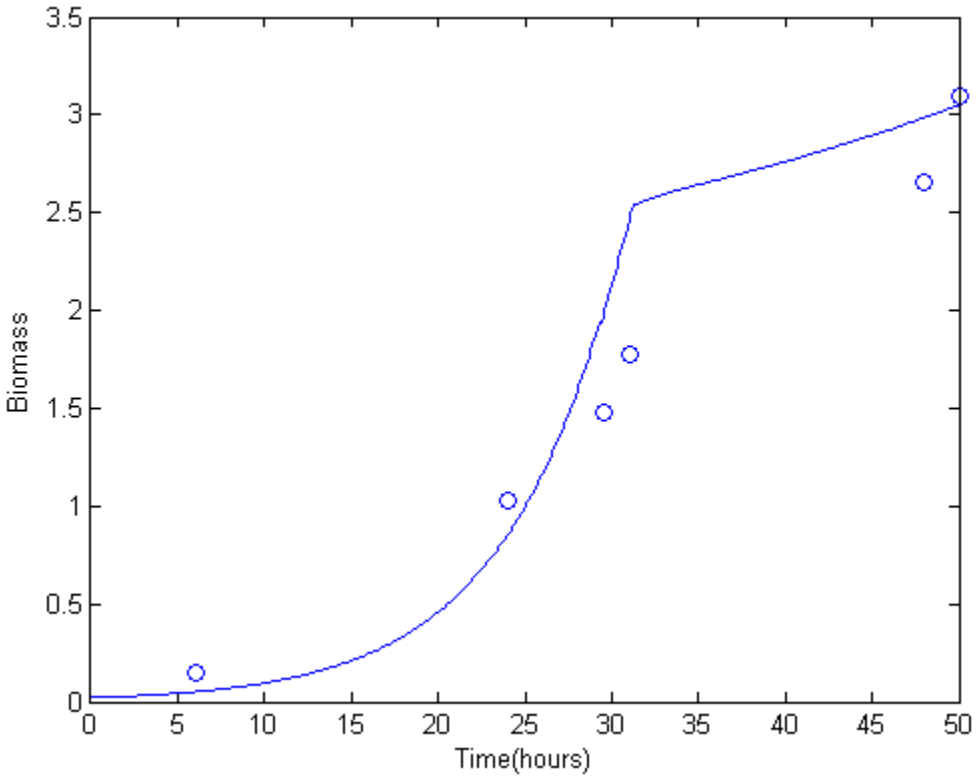


Figure 5-11 Evolution of biomass by time. Circles are experimental data and line is model prediction.

5.4 Conclusions

We presented two approaches for identifying limiting constraints required for the formulation of DFB models. Both methods were successful in identifying the same set of limiting constraints to describe the simulated and experimental data. However, the parametric sensitivity analysis based approach was found computationally challenging for identifying limiting constraints for the *B. pertussis* culture, which involved a large number of reactions.

The parametric sensitivity analysis exhibit two main computational challenges: i- it required the initial calculation of nominal parameters values from a non-convex bi-linear optimization and ii-requires testing the effect of all combinations of possible constraints on the value of the sum of square errors between measurements and predictions. Therefore it was concluded that the approach is not practical for large metabolic networks such as the *B. pertussis* system considered in the second case study. The calculations in the set-based approach are significantly simplified by assuming the data to lie within sets. This method is numerically advantageous since it involves the solution of LP problems and also finding of the kinetic expressions describing the constraints, can be done one at a time. Beyond its use for the formulation of the DFB models, the knowledge about the limiting constraints could also be used in the future for targeted genetic manipulations of the organisms under study.

5.5 Acknowledgment

The authors would like to thank Natural Science and Engineering Research Council (NSERC) for funding this research.

Chapter 6

A Systematic Approach for Finding the Objective Function and active constraints for Dynamic Flux Balance Analysis

[2]Dynamic flux balance analysis (DFBA) has become an instrumental modeling tool for describing the dynamic behavior of bioprocesses. DFBA involves the maximization of a biologically meaningful objective subject to kinetic constraints on the rate of consumption/production of metabolites. In this chapter, we propose a systematic data-based approach for finding both the biological objective function and a minimum set of active constraints necessary for matching the model predictions to the experimental data. The proposed algorithm accounts for the errors in the experiments and eliminates the need for ad-hoc choices of objective function and constraints as done in previous studies. The method is illustrated for two cases: (i) for *in silico* (simulated) data generated by a mathematical model for *E. coli* and (ii) for actual experimental data collected from the batch fermentation of *Bordetella pertussis* (*whooping cough*).

6.1 Introduction

Cell metabolism is generally described by a complex network of biochemical reactions involving different metabolites interacting with each other (Schilling, Edwards, Letscher, & Palsson, 2000; Schilling, Letscher, & Palsson, 2000). Different levels of cellular control orchestrate the evolution of

[2] This chapter has been published as the paper: Nikdel, A., Braatz, R. D., & Budman, H. M. (2018). A systematic approach for finding the objective function and active constraints for dynamic flux balance analysis. *Bioprocess Biosyst Eng*. doi: 10.1007/s00449-018-1899-y

these reactions within kinetic or thermodynamic constraints. Systems biology elucidates the combinatorial role of different level of interaction between the reactions inside the cell as a system and its response to the changes in the surrounding environment of the cell (Jaqaman & Danuser, 2006). This systematic understanding is instrumental for optimizing the growth and productivity of existing cell lines or for designing new cell lines.

Various metabolic models have been presented for an understanding of biological systems, including constraint-based flux balance analysis (FBA) as an important category of metabolic models. FBA models are used to describe the fluxes at steady state based on the optimality assumption of evolutionary biology whereby cells are using resources optimally to maximize or minimize a specific objective function in order to survive (Orth et al., 2010).

Since most of the biological processes are dynamic in their nature, e.g. batch operations, models that represent these processes during transients are sought. Dynamic flux balance analysis (DFBA) is the dynamic extension of flux balance analysis (FBA) models to have the ability to estimate the metabolites' concentrations over time (Foguet et al., 2016; Llaneras, Sala, & Picó, 2012).

Models of dynamic biological systems have traditionally consisted of systems of differential equations, each describing a dynamic balance of a main nutrient or important byproduct that is being consumed or produced generally following Michaelis-Menten reaction kinetics. These models have been often found lacking since they involve a large number of kinetic parameters that need to be calibrated to fit the experimental data. Being based on an optimization with few limiting constraints, DFBA models are advantageous over these traditional dynamic models because they typically required a smaller number of calibration parameters (Mahadevan et al., 2002; Varma & Palsson, 1995).

Typically, DFBA models involve the solution at each time interval *k* of the optimization:

$$\max_{\boldsymbol{v_k}} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v_k}, \tag{6-1}$$

Subject to:

$$\boldsymbol{g}(\psi_{\boldsymbol{k}}) \leq |\boldsymbol{S}\boldsymbol{v_k}| \leq \boldsymbol{f}(\psi_{\boldsymbol{k}})$$

$$\boldsymbol{v_k} \geq \boldsymbol{0}$$

$$\psi_{\boldsymbol{k}} \geq 0$$

where $\boldsymbol{S}$ is the stoichiometric matrix, $\boldsymbol{v_k} = (v_1, \ldots, v_n)_k$ is a vector of fluxes (moles/hr·mole of biomass) at time instance k, n is the number of reactions, and the biological objective function is expressed as a specific linear combination of fluxes, $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{v_k}$. This objective function is generally related to a biological variable that is representative of the overall evolution of the culture such as growth rate, ATP production, etc. The constraints given by functions $\boldsymbol{f}$ and $\boldsymbol{g}$ that are dependent on specific metabolites' concentrations at time *k* ψ_k are describing kinetic limitations associated to the consumption or production of these metabolites.

Thus, the formulation of a DFBA model involves the choice of two main elements:

1) A biological meaningful objective function
2) A set of limiting constraints

The stoichiometric matrix ($\boldsymbol{S}$), generally available from public sources such as the KEGG (Kyoto Encyclopedia of Genes and Genomes), describes the metabolic reactions among species for any microorganism of interest. The idea behind the use of the constraints in this model is that only a few key metabolites are limiting in terms of their rate of consumption or production whereas all other

metabolites will follow the dynamic behavior of the limiting metabolites based on static stoichiometric relations among all metabolites.

The formulation of the model as a constrained optimization is based on the assumption that cells have evolved through time to act as optimizers that allocate their resources through maximizing/minimizing an objective function subject to some kinetic constraints.

The choice of a biological meaningful objective function (Sanchez & Saez, 2014) is of key importance for formulating an accurate DFBA that will not require a large number of associated constraints and calibrating parameters. The choice of the objective function has been reported to greatly impact the model prediction accuracy (Sanchez & Saez, 2014).

Common objective functions used in previous studies, mainly for bacterial systems, are the maximization of growth rate or final biomass at the end of a batch (Orth et al., 2010). However, several other objective functions were found, by trial and error, to more accurately predict the system behavior such as minimization of the production rate of redox potential, minimization of ATP production rate, maximization of ATP production rate, maximization and minimization of nutrient uptake rate, maximization of biomass yield per unit flux, maximization of ATP yield (maximal energy efficiency), minimization of the overall intracellular flux, maximization of ATP yield per flux unit (maximizing ATP yield while minimizing enzyme usage), maximization of biomass yield per flux unit (maximizing biomass yield while minimizing enzyme usage), minimization of glucose consumption (more efficient usage of substrate), and minimization of reaction steps (minimization of the number of reaction steps for cell growth) (Knorr, Jain, & Srivastava, 2007; Pramanik & Keasling, 1997; Sanchez & Saez, 2014; Schuetz et al., 2007). Furthermore, it has been suggested that the objective function that rules the cell behavior may be a nonlinear combination of different specific objectives such as redox minimization,

growth maximization, or ATP production (Burgard & Maranas, 2003) rather than a linear combination as used in earlier DFBA models.

In terms of the choice of a suitable objective function, mammalian cells might accomplish different functionalities during the course of the fermentation comparing to bacteria. For example, a previous study on hybridoma cell central metabolism (Burgard & Maranas, 2003) studied the three choices of objective function: 1) minimizing ATP production, 2) minimizing total nutrient uptake, and 3) minimizing redox metabolism through minimizing NADH production. That study concluded that, while no single objective could solely rule the cell behavior, minimizing the NADH production can better describe the typical characteristic behavior of hybridoma cells such as their inefficient use of nutrients. This inefficient use of resources translates into higher rates of consumption of glutamine and production of alanine (Savinell & Palsson, 1992).

Some methods have been proposed for systematically finding the biological objective function. For example, an approach referred to as ObjFind has been reported (Knorr et al., 2007; Schuetz et al., 2007) that is based on the minimization of the sum of squared errors between identified fluxes from experimental and simulated data. A drawback of this method is that it involves a non-convex constrained optimization for which local optima are possible (Burgard & Maranas, 2003; Knorr et al., 2007). Furthermore, the location of these multiple optima will be sensitive to the choice of limiting constraints.

Knorr et al. (Knorr et al., 2007) developed a Bayesian-based method for selecting the most suitable objective function for *E. coli* growing on succinate and producing acetate. Among 5 possible objective functions, the minimization of the production rate of redox potential resulted in a better model in terms of its ability to fit the data, and was the only objective function capable of prediction of the acetate production by the cells.

The Biological Objective Solution Search (BOSS) method is an additional reported bi-level optimization algorithm for inference of the objective function that does not require assumptions of candidate objectives (Gianchandani, Oberhardt, Burgard, Maranas, & Papin, 2008; Mahadevan et al., 2002). However, this method was found to lead to overfitting of experimental data and to the finding of objective functions that do not have a particular biological meaning but are rather a particular combination of certain metabolic fluxes. Furthermore, the resulting bi-level optimization is non-convex and computationally expensive (Gianchandani et al., 2008; Sanchez & Saez, 2014).

FBA models are often referred as being "constraint-based" since they rely on biological constraints that represent the limited ability of the cell to consume certain nutrients or produce certain byproducts (Llaneras & Picó, 2008). These constraints can be categorized into two groups: non-adjustable and adjustable. Non-adjustable constraints such as stoichiometric constraints, enzyme, and transporter capacities are time independent and describe intrinsic characteristics of the cells under consideration. On the other hand, kinetic and regulatory constraints may change as a function of environmental conditions and therefore are reasonable to adjust in order to match the model predictions to the experimental observations (Llaneras & Picó, 2008).

In an earlier work by the authors (Nikdel & Budman, 2016), the limiting constraints were found by inspecting the values of the Lagrange multipliers and assessing from them whether particular constraints are active or inactive at the solution. The disadvantage of that approach is that we had to decide on threshold values for the Lagrange multipliers as these values could not be systematically related to actual uncertainty/noise in the measurements.

An additional challenge in our earlier study was related to multiplicity of solutions that is also common in FBA models due to a large number of reactions that are considered versus the limited number of measured variables that are used to constrain the problem. In the literature, the addition of both ad-hoc

chosen capacity constraints and/or thermodynamic constraints have been proposed to limit the solution space (Maranas, Zomorrodi, & Wiley Online Library, 2016). In our previous work, it was necessary to choose by trial and error a set of loose constraints to limit the solution space to address multiplicity.

In this work, we propose a novel optimization-based algorithm to systematically and simultaneously identify both the limiting adjustable constraints (that is, an additional set of constraints to limit the solution space) and the biological objective function. For the limiting constraints, the objective is to identify the metabolites for which a change in their uptake or production rates significantly affects the value of the objective function and as a result on the estimated flux values. Constraints that are used to limit the solution space are also systematically identified from the proposed optimization. The significance of all the constraints is directly related in this work to measurement error. For the objective function, we investigate various probable objective functions by considering them simultaneously within the proposed optimization to find the solution of this problem that results in the best fit for describing the data. Then, after identifying the necessary constraints and biological objective function, a predictive model is formulated where the identified constraints are related to the corresponding metabolite concentrations and the future metabolite concentrations can be predicted using the fluxes resulting from the optimization into dynamic mass balances of all metabolites.

This chapter is organized as follows. Section 6.2.1 presents the mathematical algorithm for identifying the limiting constraints and objective function, and Section 6.2.2 formulates a predictive model of metabolites based on the identified objective function and constraints. Section 6.3 presents experimental methods used to collect data for the *B. pertussis* case study and how to generate the *in silico* data used for the *E. coli* study. Section 6.4 presents the Results followed by Conclusions in Section 6.5.

6.2 Theoretical and Experimental Methods

The development of the DFBA model in this study involves two steps:

1) An identification algorithm of limiting constraints and the biological objective function

2) The development of a predictive DFBA model based on the identified constraints and biological objective function.

These two steps are discussed separately below.

6.2.1 Identification algorithm

The proposed identification algorithm was formulated to meet 3 main goals:

1) Formulate the identification problem as a single-level optimization.

2) Find limiting constraints such as the level of fit that can be directly related to the measurement error.

3) Integrate into the algorithm a search for a biologically motivated objective function that results in best fit between data and model predictions.

Regarding Goal 1, as mentioned in the Introduction, some approaches that were proposed for identifying DFBA models are based on bi-level optimization formulations which can be numerically challenging due to the existence of local minima. This problem is especially important for the objective of simultaneously identifying the objective function and the set of limiting constraints. Bi-level formulations derive from the need to minimize some measure of the quality of fit of model predictions and experiments while maximizing a biological objective as required in DFBA formulations as shown in equation (6-1). The measure of the fit, e.g. the sum of squared errors, is calculated with the input

data which generally consist of the concentrations of extracellular metabolites at different time intervals. To avoid a bi-level formulation, a key idea for the approach proposed in the current study is that the data (concentrations of extracellular metabolites) are represented by convex sets whereby, at each time interval, data points are limited by upper and lower bounds. For example, Figure 6-1 shows the evolution of glucose concentration during a batch where at each time interval the glucose level lies between an upper and lower limit.

The use of convex sets to represent biological experimental data has been recently reported in several studies dealing with identification of biological models (Borchers et al., 2013; Nikdel & Budman, 2016). In this approach, the experimental data within the convex set is generally assigned the same probability to occur. Experimental data in biological experiments are highly variable due to measurement noise or lack of exact repeatability due to unmeasured disturbances such as variability in initial conditions. Then, if the number of experimental repeats is not large, it is reasonable to ascribe the same probability to the values contained between the bounds at each time interval as shown in Figure 6-1.

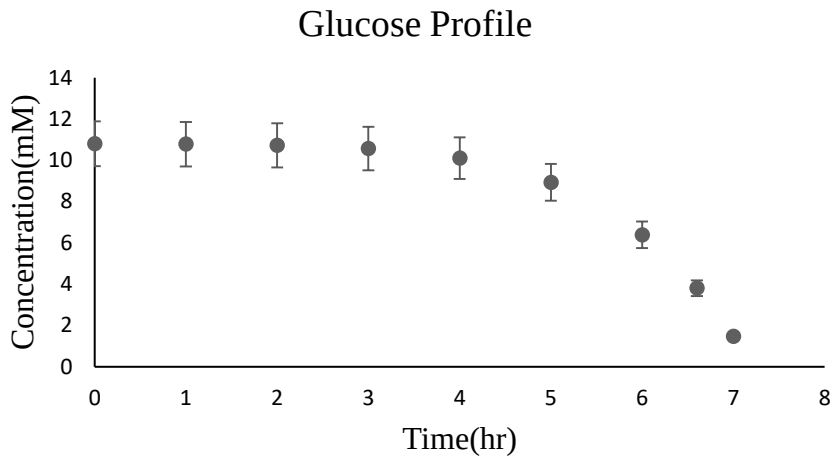


Figure 6-1 Convex set constraints for glucose concentrations of E. coli diauxic metabolism with 10% error bound

Since the convex sets can be formulated as inequality constraints within an optimization, there is no need in our proposed approach to use a dedicated optimization level for minimizing the errors in fitting between data and model predictions. Hence, by using set constraints, a bi-level optimization formulation is avoided.

The key idea for finding the limiting constraints in the DFB model by this method is to represent the data by convex sets. To this purpose, it is assumed that because of measurement error or unmeasured disturbances, the metabolite concentrate-actions are bounded by upper and lower limits at each time interval for which data is collected. For example, typical set regarding Goal 2, to identify a model that is robust with respect to uncertainty in measurements, the key idea is to perform an optimization with respect to two different sets of fluxes: one set of fluxes ($\boldsymbol{J_T}$) that satisfies the convex sets' inequalities without measurement uncertainty and a second set of fluxes ($\boldsymbol{J_R}$) that satisfies these inequalities with an accuracy related to the measurement error. This means that the original convex sets (used to solve $\boldsymbol{J_T}$)

represent lack of repeatability of batches due to unmeasured disturbances, e.g. unmeasured changes in inoculum amount and population content, unmeasured changes in minor species, etc.

In addition to these unmeasured disturbances there is measurement noise of magnitude epsilon. The flux vector $\boldsymbol{J_R}$ is solved by satisfying the original convex sets with the added noise.

Then, we seek a minimal set of constraints by adding to the optimization objective one term that it is related to the error between $\boldsymbol{J_T}$ and $\boldsymbol{J_R}$ and additional terms that penalize the activation of the limiting constraint. By minimizing the resulting objective function, we are minimizing the number of constraints that will be active at the solution while at the same time minimizing the error between the ideal set of fluxes $\boldsymbol{J_T}$ and the uncertain set of fluxes $\boldsymbol{J_R}$.

Regarding Goal 3, the search for the most suitable biologically meaningful objective function for solving the DFBA is conducted by adding different candidate functions, e.g. growth rate, redox potential etc., into the overall objective. Then, to force the solution towards one specific candidate, the cross-products of the candidates' biological functions are also added to the objective function such as, at the solution of the minimization, these cross-products will be driven to zero and only one particular candidate will be left.

Following the above considerations, the resulting optimization is

$$\min_{\boldsymbol{J_R J_T, w^L, w^U, w_c}} (\boldsymbol{J_T} - \boldsymbol{J_R})^{\mathrm{T}}\mathbf{H}(\boldsymbol{J_T} - \boldsymbol{J_R}) - \sum_{i=1}^{N_C}(w_{c_i})\,\boldsymbol{c_i}^{\mathrm{T}}\boldsymbol{J_T} - \sum_{i=1}^{N_C}(w_{c_i})\boldsymbol{c_i}^{\mathrm{T}}\boldsymbol{J_R} + \frac{1}{2}\sum_{\substack{i=1\\ j=1\\ i\neq j}}^{N_C}(w_{c_i})(w_{c_j}) + \left(\sum_{i=1}^{N_m} w_i^L - \sum_{i=1}^{N_m} w_i^U\right)\frac{n_g}{\varepsilon N_m} \quad (6\text{-}2a)$$

subject to:

$$\left.\frac{d\psi_i}{dt}\right|^l w_i^l \leq \boldsymbol{S_i}\,\boldsymbol{J_R} X_k \leq \left.\frac{d\psi_i}{dt}\right|^u w_i^u \quad (6\text{-}2b)$$

$$\left.\frac{d\psi_i}{dt}\right|^{l} \leq \boldsymbol{S_i J_T} X_k \leq \left.\frac{d\psi_i}{dt}\right|^{u} \quad \text{(6-2c)}$$

$$1 \leq w_i^u \leq 1 + \varepsilon \quad \text{(6-2d)}$$

$$1 - \varepsilon \leq w_i^l \leq 1 \quad \text{(6-2e)}$$

$$\boldsymbol{J_R} \geq \mathbf{0} \quad \text{(6-2f)}$$

$$\boldsymbol{J_T} \geq \mathbf{0} \quad \text{(6-2g)}$$

$$\psi_{\boldsymbol{k}} \geq \mathbf{0} \quad \text{(6-2h)}$$

$$0 \leq w_{c_i} \leq 1 \quad \text{(6-2i)}$$

$$\sum_{i=1}^{N_c} w_{c_i} = 1 \quad \text{(6-2j)}$$

We defined $\left.\frac{d\psi_i}{dt}\right|^{l} = \frac{\psi_i^{l,k+1} - \psi_i^k}{\Delta T}$ and $\left.\frac{d\psi_i}{dt}\right|^{u} = \frac{\psi_i^{u,k+1} - \psi_i^k}{\Delta T}$. The first term in the objective function (6-2a) is a quadratic term that minimizes the sum of square errors between the assumed two sets of fluxes J_T and J_R, $\boldsymbol{J_T}$ represents the fluxes that satisfy tight constraints on metabolites' concentrations or consumption and production rates of these metabolites as observed from the limited data available, and $\boldsymbol{J_R}$ are the fluxes that satisfy relaxed constraints for which it is assumed that the data are uncertain. The $\boldsymbol{S_i}$ is the row of the stoichiometric matrix representing metabolite i, and X_k is the biomass value at time k**.** The Hessian is

$$\mathbf{H} = u_w \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \quad \text{(6-2k)}$$

the sum of square errors between the elements of the vector $\boldsymbol{J_T}$ and the elements of the vector $\boldsymbol{J_R}$ weighted by the scalar $u_w$that is selected sufficiently large so as for at the solution the first term in (6-2a) is smaller than a tolerance (10^{-4}) that is of the order of the noise $\boldsymbol{n_g}$ in the biological objective (e.g. the measurement noise of the growth rate if the latter is used as the biological objective).

To account for the uncertainty in consumption/production rates, the bounds on the relaxed constraints are obtained by multiplying the bounds on the tight constraints by weights w_i^u and w_i^l in Eq. 6-2a. These weights change between 1 to 1+ε or 1–ε to 1 for the upper or lower constraints respectively (equations (6-2d) and (6-2e)) where ε is chosen to reflect the expected uncertainty in the measurements of consumption/production rates of metabolites.

The second and third term in the cost function (6-2a) describe the biological objective functions to be maximized in a DFBA model computed as a function of either the $\boldsymbol{J_T}$ or the $\boldsymbol{J_R}$ vectors of fluxes that satisfy either the tight and relaxed constraints respectively. N_C is the number of the objective functions' candidates for the problem and the w_{c_i} are the weight coefficients corresponding to the candidates of the objective function.

The fourth term in Eq. (6-2a) are cross-products of the corresponding weights of any two candidate biological objective functions. The minimization in (6-2a) is expected to force these cross-products close to zero values thus only leaving one nonzero value of the weights w_{c_i}. This non-zero value of a w_{c_i}will indicate which one of the candidate biological objectives is the one to be chosen for the problem, i.e. the goal is that at each time interval only one of the weight coefficients which has the highest impact will be driven to one and the rest will tend to zero. Mathematically there may be a situation that the resulting bilinear optimization will give fractional values, especially in the case that the sum of the cross-products of the weights is not significant as compared to the other terms in the objective function of the problem. If this happen it could be resolved by weighting the cross-products with a higher scalar weight.

The fifth term in Eq. (6-2a) is to drive as many constraints as possible to non-active status, i.e. the redundant ones so as to identify only the necessary ones. It should be noticed that $1 \leq w_i^u \leq 1 + \varepsilon$ and

$1 - \varepsilon \leq w_i^l \leq 1$ and these inequalities imply that the extreme values of the weights are $w_i^u = 1 + \varepsilon$ and the smallest $w_i^l = 1 - \varepsilon$. Thus, the minimization of the 5th term ($\sum_{i=1}^{N_m} w_i^L - \sum_{i=1}^{N_m} w_i^U$) will drive as many as possible w_i^l to $1 - \varepsilon$ or w_i^u to $1 + \varepsilon$ respectively and thus, when they are active at these extreme values, the corresponding constraints are not necessary (not active) for the given noise. Hence if I can expand the bounds on the consumption/production rates up to the limit of the noise without affecting the error $\boldsymbol{J_T - J_R}$ in a significant way, this means that the corresponding constraints are not necessary to explain the data. Since the fifth term is of the order of εN_m (the ones are cancelling out with each other), we must introduce the factor, $\frac{n_g}{\varepsilon N_m}$, that ultimately render this term of the order of the assumed noise in the growth rate of magnitude n_g.

In fact the first and the fifth terms are introduced in the objective function of the optimization (Eq. (6-1)) to achieve a tradeoff between the error $(\boldsymbol{J_T - J_R})$ and the limiting constraints active at the solution of $\boldsymbol{J_T}$ and $\boldsymbol{J_R}$.

Constraints (6-2b) and (6-2c) are the key elements for the development of a predictive DFBA model presented in the next subsection. The idea is to find for which amino acid constraints (6-2b) and (6-2c) become active and to eventually replace these constraints by their corresponding kinetics functions to be able to predict future concentrations. More details for this process has been explained in Section 2.2. The objective function in (6-2a) is nonlinear involving a bilinear term and a quadratic term, and the constraints are all linear. Since the bilinear term involves products of weights that are each bounded between 0 and 1, the term is convex. Thus, overall this formulation results in a nonlinear convex optimization that can be solved by common optimization solvers such as *fmincon* in *MATLAB*. Accordingly, the optimization search must be conducted for different initial guesses.

For the special case that the biological objective function is known or assumed *a priori* and only the active constraints need to be identified, the nonlinear optimization of Eq. (6-2) that results from the presence of bilinear terms (terms 2 and 3 in Eq. (6-2a)) converts into the simple quadratic optimization:

$$\min_{\boldsymbol{J_R},\boldsymbol{J_T},\boldsymbol{w^L},\boldsymbol{w^U}} (\boldsymbol{J_T}-\boldsymbol{J_R})^{\mathrm{T}}\mathbf{H}(\boldsymbol{J_T}-\boldsymbol{J_R}) - \boldsymbol{c}^{\mathrm{T}}(\boldsymbol{J_T}+\boldsymbol{J_R}) + (\textstyle\sum_{i=1}^{N_m} w_i^L - \sum_{i=1}^{N_m} w_i^U)\frac{n_g}{\varepsilon N_m} \quad (6\text{-}3a)$$

subject to:

$$\left.\frac{d\psi_i}{dt}\right|^l w_i^l \le \boldsymbol{S_i}\,\boldsymbol{J_R}X_k \le \left.\frac{d\psi_i}{dt}\right|^u w_i^u \quad (6\text{-}3b)$$

$$\left.\frac{d\psi_i}{dt}\right|^l \le \boldsymbol{S_i}\,\boldsymbol{J_T}X_k \le \left.\frac{d\psi_i}{dt}\right|^u \quad (6\text{-}3c)$$

$$1 \le w_i^u \le 1+\varepsilon \quad (6\text{-}3d)$$

$$1-\varepsilon \le w_i^l \le 1 \quad (6\text{-}3e)$$

$$\boldsymbol{J_R} \ge \mathbf{0} \quad (6\text{-}3f)$$

$$\boldsymbol{J_T} \ge \mathbf{0} \quad (6\text{-}3g)$$

$$\psi_{\boldsymbol{k}} \ge \mathbf{0} \quad (6\text{-}3h)$$

This quadratic optimization can be solved, for example, using *cplexqp* (**IBM CPLEX**) in a MATLAB-compatible code. If the fluxes $\boldsymbol{J_R} = \boldsymbol{J_T}$ at the solution of the QP then the quadratic term in (6-3a) vanishes and the solution is identical to the solution of an LP that corresponds to the original formulation of a DFBA model equation (6-1) augmented by the second term that represents added uncertainty to the biological objective as explained before in this section.

A main negative consequence of the multiplicity of solutions is that, for some problems such as the *B. pertussis* case study in this work, some of the rate constraints in equation (6-2b) and (6-2c) become active only on isolated time intervals along a batch fermentation and this sporadic activation of constraints depends highly on the initial guesses assumed for the optimization. On the other hand, it

should be remembered that one of the expected benefits of the DFBA model is to predict concentrations using only a few kinetic constraints so a small number of kinetic parameters would be needed. The sporadic activation of many constraints defeats this purpose since it would require constraining many amino acids and calibration of many corresponding kinetic expressions.

A way to address this situation is by limiting the solution space for the optimization through the imposition of upper bounds on consumption or production rates of metabolites. Sometimes, these bounds are available from the literature. In the case that these bounds are not available, we propose to find the bounds from a modification of the optimization in (6-2a):

$$\min_{\boldsymbol{J_R J_T, w^L, w^U, w_c, w^{sc}}} (\boldsymbol{J_T} - \boldsymbol{J_R})^{\mathrm{T}} \mathbf{H} (\boldsymbol{J_T} - \boldsymbol{J_R}) - \sum_{i=1}^{N_C} (w_{c_i})\, \boldsymbol{c_i}^{\mathrm{T}} \boldsymbol{J_T} - \sum_{i=1}^{N_C} (w_{c_i}) \boldsymbol{c_i}^{\mathrm{T}} \boldsymbol{J_R} + \frac{1}{2} \sum_{\substack{i=1 \\ i \neq j}}^{N_C} (w_{c_i})(w_{c_j}) +$$

$$\left(\sum_{i=1}^{N_m} w_i^L - \sum_{i=1}^{N_m} w_i^U\right) \frac{n_g}{\varepsilon N_m} - \left(\sum_{i=1}^{N_{sc}} w_i^{sc}\right) \frac{n_g}{W^U N_{sc}} \tag{6-4a}$$

subject to:

$$\left.\frac{d\psi_i}{dt}\right|^l w_i^l \leq \boldsymbol{S_i J_R} X_k \leq \left.\frac{d\psi_i}{dt}\right|^u w_i^u \tag{6-4b}$$

$$\left.\frac{d\psi_i}{dt}\right|^l \leq \boldsymbol{S_i J_T} X_k \leq \left.\frac{d\psi_i}{dt}\right|^u \tag{6-4c}$$

$$1 \leq w_i^u \leq 1 + \varepsilon \tag{6-4d}$$

$$1 - \varepsilon \leq w_i^l \leq 1 \tag{6-4e}$$

$$\boldsymbol{J_R} \geq \mathbf{0} \tag{6-4f}$$

$$\boldsymbol{J_T} \geq \mathbf{0} \tag{6-4g}$$

$$\psi_{\boldsymbol{k}} \geq \mathbf{0} \tag{6-4h}$$

$$0 \leq w_{c_i} \leq 1 \tag{6-4i}$$

$$\sum_{i=1}^{N_c} w_{c_i} = 1 \tag{6-4j}$$

$$|\boldsymbol{S_i J_T} X_k| \leq w_i^{sc} \quad (6\text{-}4k)$$

$$0 \leq w_i^{sc} \leq W^u \quad (6\text{-}4l)$$

The optimization (6-4a), in contrast with (6-2a), includes a sixth term in the objective function whose purpose is to maximize at each instance the absolute consumption/production rate of each metabolite (defined in 6-4k) towards an upper bound W^u as defined by (6-4l). It can be easily verified that the additional term in (6-4a) is also of the order of the noise,n_g since the term is normalized by the maximum allowable value W^u (Eq. 6-4l) and by the total number of metabolites N_{SC}. To constrain the solution space, it has been often proposed in the FBA literature to bound all the fluxes by a value of the order of the largest known consumption rate of the limiting nutrients (Maranas et al., 2016). In this work, we have used a similar criterion to choose the value of W^u. In principle, this value can be iteratively adjusted by executing the optimization in (6-4a)–(6-4e) for different values of W^u until a satisfactory fitting between the vectors of fluxes J_T and J_R is obtained.

The fundamental difference between constraints (6-4b)–(6-4c) with respect to the constraints (6-4k)–(6-4l) is that the former are tight constraints of the order of the noise (ε) on the metabolite consumption/production rates that are frequently active along a batch whereas the latter are loose rate constraints related to the noise n_g in the biological objective that are only important in isolated time intervals along the batch. In principle, the additional term in (6-4a) together with constraints (6-4i)–(6-4l) result in time-varying values of the weights w_i^{sc} for each metabolite i. For the purpose of limiting the solution space of a predictive model, one time-independent upper bound is found for each metabolite i from

$$w_i^{sc,max} = \max\,[w_i^{sc}(k)], \quad (6\text{-}5)$$

where k ranges from 0 to T. Selecting the maximum bound is justified since, for the worst case that all $w_i^{sc} = W^u$, the effect of the sixth term in (6-4a) is guaranteed to be no larger than n_g, i.e. the noise in the biological objective function. Then, the bound $w_i^{sc,max}$ for each metabolite, i is used as an upper bound for the rate of that metabolite in the predictive model that is presented in the next section.

6.2.2 Predictive DFBA model

The predictive DFBA model is given by

At each time instance $\boldsymbol{k}$ ($\boldsymbol{k = 1, \dots, T}$)

$$\max_{\boldsymbol{v_k}} \boldsymbol{c}^{\mathrm{T}} \boldsymbol{v_k}, \tag{6-6a}$$

subject to:

$$|\boldsymbol{S^{m_i}} \boldsymbol{v_k}| \leq \frac{V_{m_i,max} \cdot \psi_{m_i,k}}{K_{m_i} + \psi_{m_i,k}} \frac{\text{mmol}}{\text{g·dw·hr}} \quad (\text{i} = 1, 2, \dots, \text{N}_\text{m};\ \text{N}_\text{m} \text{ is the number of metabolites}) \tag{6-6b}$$

$$\boldsymbol{v_k} \geq \boldsymbol{0} \tag{6-6c}$$

$$\left|\frac{d\varphi_i}{dt}\right| \leq w_i^{sc,max} \tag{6-6d}$$

$$\psi_{\boldsymbol{k}} \geq \boldsymbol{0} \tag{6-6e}$$

$$\psi_k = \psi_{k-1} + S v_k X_k \Delta T \geq 0 \tag{6-6f}$$

where $\boldsymbol{S}$ is the stoichiometric matrix and $\boldsymbol{S^{m_i}}$ is the row of the stoichiometric matrix associated with the metabolite i. Equations (6-6a)–(6-6f) are a modified version of the optimization (6-4a)–(6-4k) with respect to four elements:

The objective function (6-6a) includes only the biological objective (second term in 6-4a). This objective is based on the assumption that, at the optimal solution of (6-4a), the first term vanishes since, at $\boldsymbol{J_T \approx J_R}$, the third term becomes identical to the second term, the fourth and sixth terms are neglected

since they are of the order of the noise $\boldsymbol{n_g}$, and the fifth term also vanishes since only one $\boldsymbol{w_{c_1} = 1}$ and the rest are zero.

The limiting constraints identified, as explained above, the upper and lower bounds in (6-4b) and (6-4c), are replaced by explicit functions (6-6b) of the corresponding metabolites concentrations at time $\boldsymbol{k}$. Constraints (6-4i) and (6-4l) are replaced by upper bound constraints (6-6d) using the value$\boldsymbol{s\ w_i^{sc,max}}$ calculated from (6-5).

Equations to predict the concentrations of metabolites over time (Eq. 6-6f) is formulated based on mass balances. That is, Equation (6-6f) are time-discretized mass balances for all metabolites that are used to integrate the concentrations over time.

The optimization (6-6a)–(6-6f) is an LP that is solved with the CPLEX optimization software. A fundamental advantage of the proposed model (6-6a)–(6-6f) is that the parameters in the kinetic expressions corresponding to the limiting constraints (6-6b) can be separately identified for each metabolite. This is particularly advantageous as compared to previously reported algorithms where the parameters corresponding to all the kinetic expressions involved in the constraints of the DFBA problem must be simultaneously identified from bi-level optimization formulations.

The ability to identify the kinetic expression separately is due to the fact that, at each time instance, the solution of the optimization (6-4a) provides which constraint i is active; it is possible to find the corresponding values of the gradients, either $\frac{d\psi_i}{dt}\Big|^u$ or $\frac{d\psi_i}{dt}\Big|^l$, that are active at the constraint. Then, Michaelis-Menten expressions

$$\frac{d\psi_i}{dt} = \frac{V_{i,max} \cdot \psi_i}{K_i + \psi_i} X \qquad (6\text{-}7a)$$

or alternative types of kinetic expressions can be found to correlate the bounds on the uptakes or production rates ($\frac{d\psi_i}{dt}$) of the metabolite as a function of its concentration (ψ_i) at every time interval, where $V_{i,max}, K_i$ are the kinetic parameters that should be identified. For example, if it is found that glucose is a limiting constraint, then a kinetic expression can be used to represent the relation:

$$\frac{dGLC}{dt} = \frac{V_{GLC,max}.GLC}{K_{GLC}+GLC} X \qquad (6\text{-}7b)$$

Then, the two parameters $V_{GLC,max}$ and K_{GLC} can be found from time data of glucose alone since the gradient of glucose on the left-hand side of Eq. (6-5) and the corresponding concentration of glucose at each time interval are available

If Michaelis-Menten kinetics are assumed, graphical methods such as the Lineweaver-Burk are used to identify the parameters in the kinetic expressions (6-7a) or (6-7b) (Shuler & Kargi, 1992). For other types of kinetic expressions, nonlinear regression is done using the Curve Fitting Toolbox in the MATLAB environment.

6.3 Experimental and *in silico* data Used in the Case Studies

The proposed method is applied in two case studies for two model microorganisms. The first study is the batch fermentation of *E. coli* system for which a DFBA model has been reported by Mahadevan et al. (Mahadevan et al., 2002).The model was used to generate *in silico* data consisting of simulated data with superimposed white noise. Since the DFBA model with the required biological objective function and constraints was available a priori, this case study serves to verify whether our methodology is able to identify the correct objective function and limiting constraints.

For the second case study, we studied the fermentation of *Bordetella pertussis* (*B. pertussis*) used to produce the antigens of the whooping cough vaccine. The metabolic network for this organism involves

49 reactions. This process was assumed to operate initially in batch and, after depletion of the main nutrient, in fed-batch mode. Glutamate was the main nutrient for this microorganism. For this example, we use actual measurements of sixteen amino acids concentrations measured by HPLC.

6.4 Results

6.4.1 Case Study 1: *E. coli*

The DFBA model describing the diauxic growth of an *E. coli* system involves four species: glucose, acetate, oxygen, and the biomass in its network. All species except the biomass are considered as potentially active constraints.

The model proposed by Mahadevan et al. (Mahadevan et al., 2002) is based on a simplified metabolic network of four reactions involving the four aforementioned species shown in Appendix A. Mahadevan's model assumes that the growth rate is maximized at each time interval subject to two kinetic limiting constraints on glucose and oxygen consumption:

At all-time instances *k*:

$$\max_{\boldsymbol{v}} \boldsymbol{c}^{\mathrm{T}}\boldsymbol{v} \tag{6-8}$$

$$\left|\mathbf{S^{Glc}}\boldsymbol{v}\right| \leq \frac{V_{Glc,max.}\psi_{Glc,k}}{K_{Glc}+\psi_{Glc,k}}\frac{\text{mmol}}{\text{gdw}\cdot\text{hr}}$$

$$\left|\mathbf{S^{O_2}}\boldsymbol{v}\right| \leq OUR\ \frac{\text{mmol}}{\text{gdw}\cdot\text{hr}}$$

This model was simulated for a total of 10 hr to simulate experimental conditions. Gaussian noise was added to the metabolites responses at a level of 5% of the total range of change of each metabolite during the batch. These simulated values were used as *in silico* data to test the proposed identification methodology as given by equations (4a)–(4e). The sampling interval used for discretization was 0.1 hr

and the oxygen uptake rate was OUR=10.Initial condition for metabolites and biomass at time zero was glucose (10.4mM), acetate (0.4mM), oxygen (0.21mM) and biomass (0.001g). Based on this level of noise, $\boldsymbol{\varepsilon}$ was set equal to 0.1 and the upper bound in constraint (4k), W^u, was set to 10 mM/hr·gdw. The weight u_w defining the Hessian as per equation (2j) was set to 0.01. As mentioned in the methodology section, this weight is selected large enough such as the first term in (4a) at the optimum is of the order of $\boldsymbol{n_g}$ (estimated noise in the growth rate). We used *interior-point* algorithms with cplexqp and cplexlp function in CPLEX for MATLAB toolbox.

In order to test our methodology, we assume that any of the nutrients (glucose, acetate, or oxygen) can be potentially limiting. In addition, we assume two possible candidate biological objectives for maximization: (1) growth rate as given by equation (8), i.e. $\boldsymbol{c_1}^{\mathrm{T}} = [1\ 1\ 1\ 1]$ (see equation 4a) and (2) maximal consumption rate of acetate, which is represented by $\boldsymbol{c_2}^{\mathrm{T}} = [39.43\ \ 0\ \ -1.24\ \ -12.12]$ in eq. (4a).

The solution over time of the weights w_i^u and w_i^l for all the species is identified. These values indicate which rate constraints are considered at each time interval according to equations (4b) and (4c). Following the solution of problem (4a), it was found that the upper bound of oxygen depletion rate is active from $t = 0$ until the end of the batch whereas the upper bound in glucose becomes active after $t = 4.2$ hr and thereafter. These results coincide with Mahadevan's model where the limiting rate constraints are glucose and oxygen while acetate is not limiting, thus corroborating the ability of the method to identify the constraints.

The weights involved in constraints (4k) and (4l) which are introduced to limit the solution space were found to be equal to their upper bound, i.e. $w_i^{sc} = W^u$, for all the 4 species and for the entire duration of the batch. The weights w_{c_1} and w_{c_2} that correspond to the growth rate and to the acetate objectives

in equation (4a) are 1 and 0 for the entire duration of the batch, thus correctly identifying the growth rate as the objective to be maximized. This again verifies the ability of the proposed approach to identify the correct biological objective function. For the purpose of formulating the predictive model according to equations (6a)–(6f), it is necessary to identify kinetic expressions of the two identified limiting rate constraints as a function of the corresponding metabolite concentrations, i.e. the rate constraints in glucose and oxygen are expressed as a function of the corresponding glucose and oxygen concentrations respectively. Figure 6-2 shows the consumption rate of glucose obtained at each time interval as a function of the corresponding glucose concentration. For this plot, a Michaelis-Menten kinetic expression was identified using the Lineweaver-Burk graphical method. The resulting kinetic parameters for the glucose are as in the following table 6-1. It should be noted that these results are almost identical to the non-linear parametric estimation.

Table 6-1 Estimated and original values of the parameters

Parameter	Original Value	Estimated Value
$V_{Glc,max}$	10	11.11
k_{Glc}	0.015	0.02

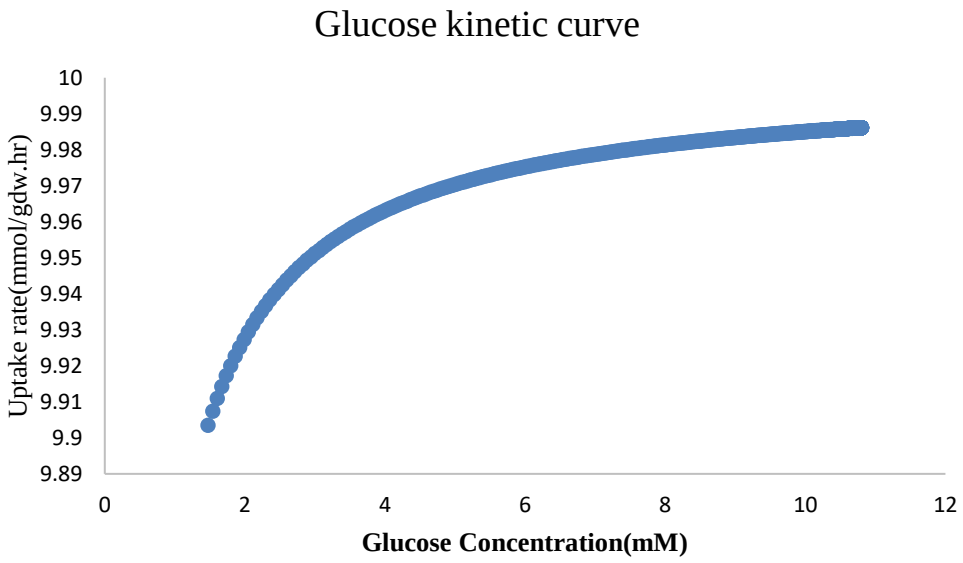


Figure 6-2 Kinetic curve of glucose (glucose uptake rate vs glucose concentration) in *E.coli* example

The procedure identifies, in accordance with Mahadevan's model used to generate the in silico data, Michaelis-Menten kinetics for the glucose consumption rate bound whereas a constant upper bound for the consumption of oxygen is identified. The numerical values of the kinetic constants are slightly different from Mahadevan's model because they were identified by the procedure in (6-4a)–(6-4l) from simulations of the original model with superimposed noise.

Figure 6-3 shows the comparison between the time evolutions of the 4 species as a function of batch time obtained from the original model without the noise (red line) and the identified model (blue line). The simulated results are slightly different between the original and identified models due to the differences in the kinetic rate expressions for the limiting constraints that resulted from the assumed noise.

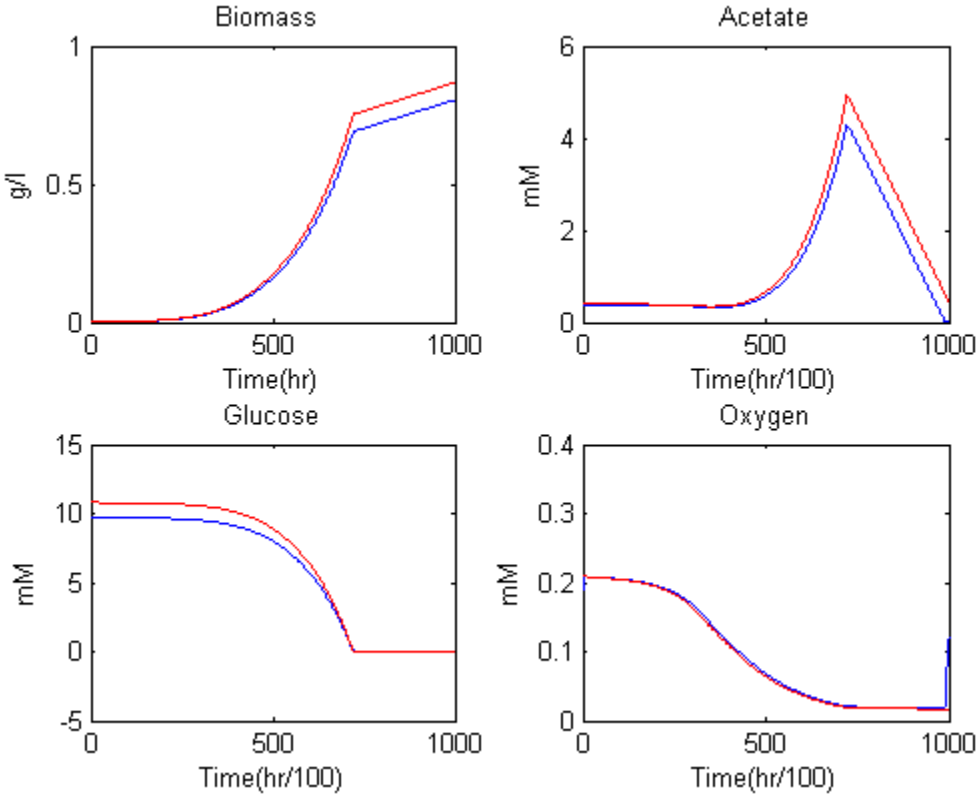


Figure 6-3 Metabolite concentration profiles of *E. coli* case study red line is the simulated original model and the blue line is our identified model

6.4.2 Case Study 2: *B. Pertussis*

Bordetella pertussis is a gram-negative, aerobic, and pathogenic bacteria causing Pertussis or *whooping cough* (Thalen et al., 2006). In this case, our algorithm is used for systematic identification from experimental data of limiting constraints and a suitable biological objective function of a *B. pertussis* fed-batch culture. Glutamate is the main carbon source used for the feeding of this culture, which is started at time 0 in batch operation until depletion of glutamate occurs and then the fed-batch operation is started with a constant glutamate feed rate of 4.3 g/h. A metabolic network involving 49 metabolites

and 50 reactions was available from a previous study by Budman et al.(Budman et al., 2013). The reactions that were considered are listed in Appendix B. HPLC measurements of 16 amino acids were available for a fifty hours fermentation process at 8 time instances: 0, 6, 24, 29, 31.1, 32, 34, 48.3 hr. The data is shown by circles in Figure 6-4. Due to confidentially reasons, all the presented data in this figure have been normalized based on the initial concentration of glutamate which is the main nutrient in this fermentation process.

A nonlinear polynomial interpolation function was applied to interpolate the data between the available measured values. The sampling interval used for the identification procedure given by equations (6-4a)–(4l) was 0.1 hr. The level of noise in the HPLC measurements was estimated as 10% of the full range of change of each amino acid and this value was used to estimate the ε = 0.1 used in the rate constraints (6-4b) and (6-4c). The upper bound W^u in constraint (6-4l) was set to 10 mM/hr·mM as in the previous case study. In principle, this value can be iteratively further adjusted as explained in Section 2. The weight u_w defining the Hessian in equation (6-2j) was set to 0.01.

Two candidates were considered as possible biological objectives for maximization: maximization of the growth rate and maximization of the lactate production rate. The growth rate is defined by the stoichiometric relations derived from the reactions listed in Appendix B.

The identification procedure given by equations (6-4a)–(6-4l), was applied to the data to identify the objective function and limiting constraints. Among constraints (6-4bc), only one constraint is active for the entire duration of the batch corresponding to the upper bound in the consumption rate of phenylalanine. The biological reason for phenylalanine being the limiting active constraint might be due to the fact that phenylalanine is not produced by *B. pertussis* and should be externally supplied in the medium.

A key difference between the current case study and the previous case study is that here only a fraction of the species involved in the network is measured whereas in the previous study all species were measured. Thus, the current case study is highly under-determined and consequently, constraints (6-4k)–(6-4l) that are used to limit the solution space become important. These constraints were found to be important for limiting the solution space for both consuming and producing species that are not being measured.

The weights w_i^{sc} (constraint 6-4k) were identified at each time interval and accordingly $w_i^{sc,max}$ was calculated using equation (6-5).

Six of the amino acids can be produced according to the metabolic reactions listed in Appendix B: Alanine, Aspartate, Glutamate, Isoleucine, Arginine, and Proline. However, it was found from the data that the concentration of these metabolites did not exhibit any noticeable increases in time. The reason is that these metabolites start to be synthesized only after they are completely depleted so as to further generate biomass, i.e., the production of these amino-acids is exactly balanced by their consumption towards biomass, especially during the fed-batch phase. In principle, an upper bound of 0 could have been assumed for the production rates but it was found that due to Euler discretization that it was difficult to satisfy the positivity constraint (6-6e) when the produced amino acids concentrations are very close to zero. For these six amino acids, having a constant upper bound with the value of 10^{-4} was necessary. The preferable objective function was found to be the growth rate. This coincides with the objective function that has been prevalent if previous bacterial models and also corresponds to the objective function used in a *B. Pertussis* model reported in Budman et al. (Budman et al., 2013).

Using the growth rate as the biological objective, it is possible to formulate the optimization problem as a QP as explained in equation (6-3a). Solving the problem as a QP serves to verify the nonlinear

optimization used to solve the problem in (6-3a)–(6-3h). The *cplexqp* solver in MATLAB was used to solve this quadratic optimization. The results were very similar to the results of the nonlinear optimization except for small differences in the values of the weights w_i^{sc}.

To formulate a predictive model as given by equations (6a)–(6f), a kinetic expression was sought relating the maximum consumption rate in phenylalanine as a function of its concentration. From the relation between the measured consumption rates and the corresponding concentrations, it was concluded that a Michaelis-Menten relation is not suitable. Instead, as assumed in our earlier work (Nikdel & Budman, 2016), a Hill-type function (Weiss, 1997) was used to describe the maximal uptake rate of the phenylalanine by

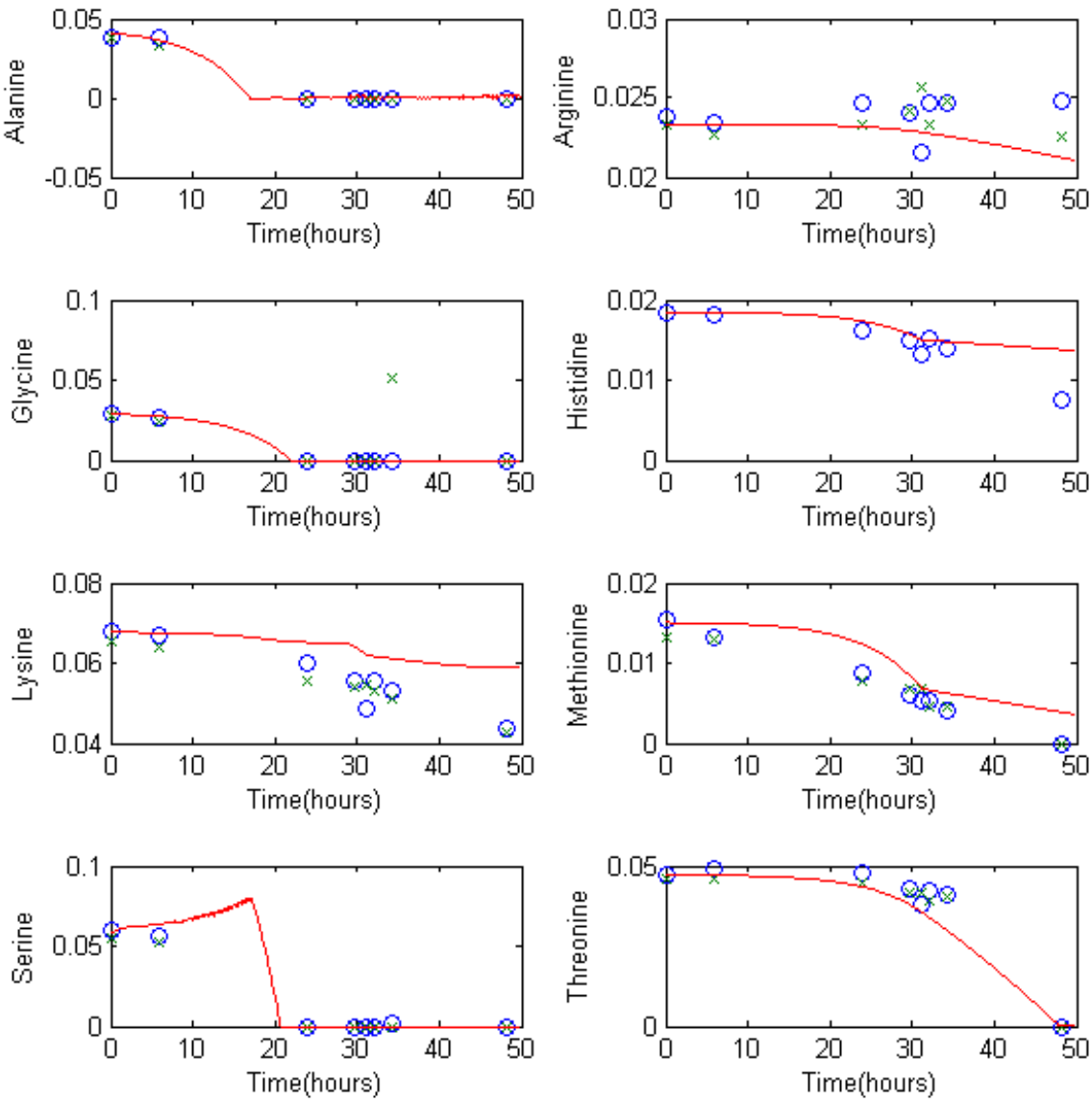
$$r_{\mathrm{Phe}} = \frac{0.0066\ \psi_{\mathrm{Phe}}{}^{3.2}}{0.022^{3.2}\ \psi_{\mathrm{Phe}}{}^{3.2}} \qquad (6\text{-}9)$$

The values of the parameters in this expression were found using the *lsqcurvefit* of MATLAB.

The bounds $w_i^{sc,max}$ used in constraints (6-6d) are obtained from the calculated values of the corresponding w_i^{sc} according to equation (6-5).

The simulated results for the 16 measured amino acids using equations (6a–f) using the phenylalanine expression in equation (9) and the constraints (6b) are shown in Figure 4. The sum of squared errors between simulations and measured values is SSE = 6.25. This error is significantly lower as compared to our earlier model of SSE = 15.42 (Budman et al., 2013) where the limiting constraints in developing the DFBA model was based on trial and error in contrast with the systematic approach proposed in this work. The reason that our model is not able to give a perfectly prediction for some of the metabolites might be due to the inaccuracies in the stoichiometric matrix and in the calculation of biomass amino acids' composition.

In order to test further the predictive ability of the model, no constraints were included for biomass. The goal was to test whether the model without any biomass-related constraints could correctly predict the biomass evolution over time. This test is significant since all amino acids contribute to the composition of biomass and thus predicting correctly the evolution of biomass implies also a certain level of fidelity in the prediction of the amino acids. Biomass measurements were available for two different feeding rates of glutamate 4.3 g/hr and 4.7 g/ hr. The model predictions of biomass for both feeding rates show very good agreement between measured and predicted values (Figure 6-5).


Alanine
0.05
0
-0.05
0 10 20 30 40 50
Time(hours)
Arginine
0.03
0.025
0.02
0 10 20 30 40 50
Time(hours)
Glycine
0.1
0.05
0
0 10 20 30 40 50
Time(hours)
Histidine
0.02
0.01
0
0 10 20 30 40 50
Time(hours)
Lysine
0.08
0.06
0.04
0 10 20 30 40 50
Time(hours)
Methionine
0.02
0.01
0
0 10 20 30 40 50
Time(hours)
Serine
0.1
0.05
0
0 10 20 30 40 50
Time(hours)
Threonine
0.05
0
0 10 20 30 40 50
Time(hours)

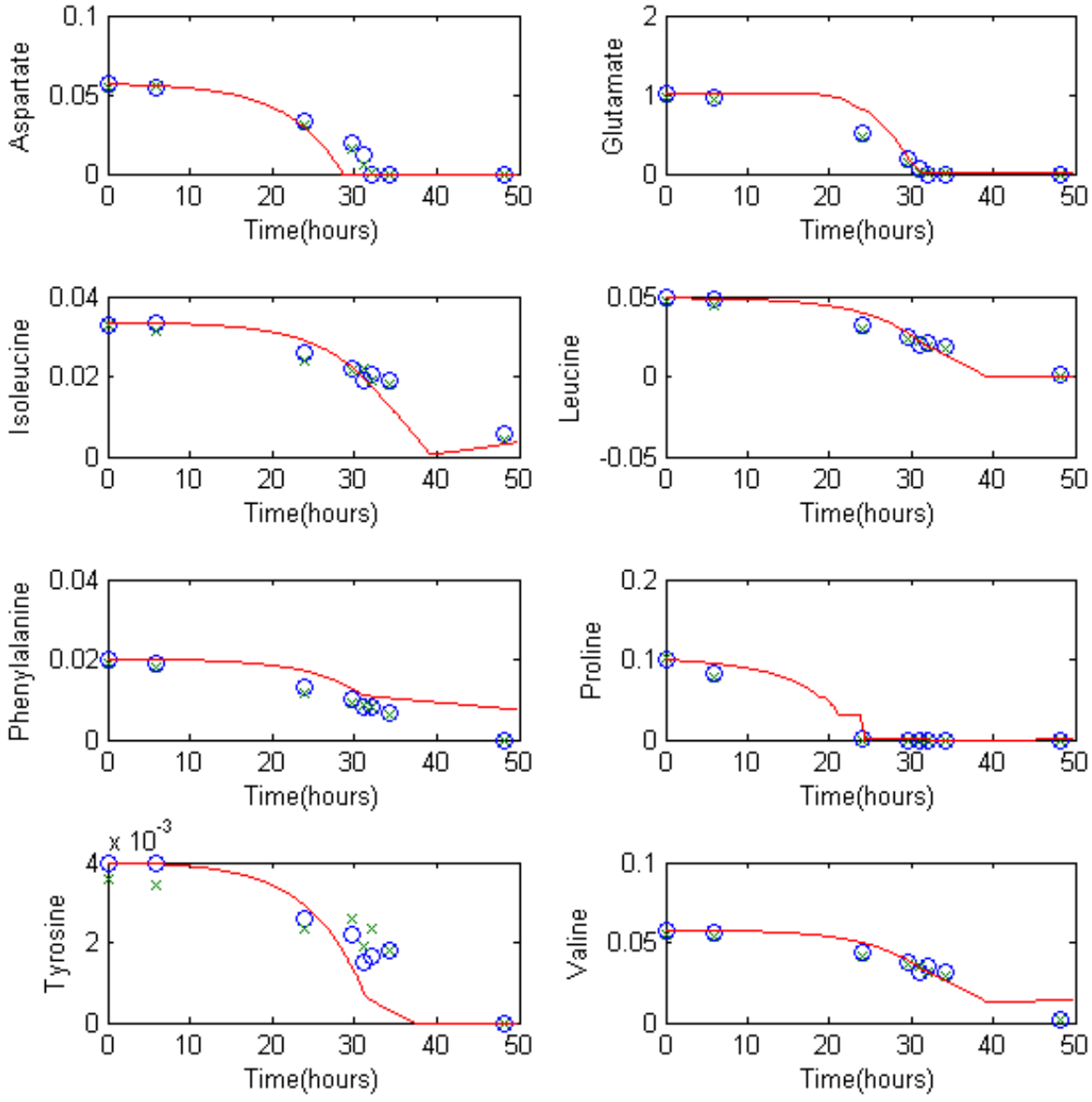


Figure 6-5 Metabolite concentration profiles of *B-Pertussis* case study resulted from the DFBA.

It also should be noted that the accuracy of the model could be increased by including more measurements in the identification step.

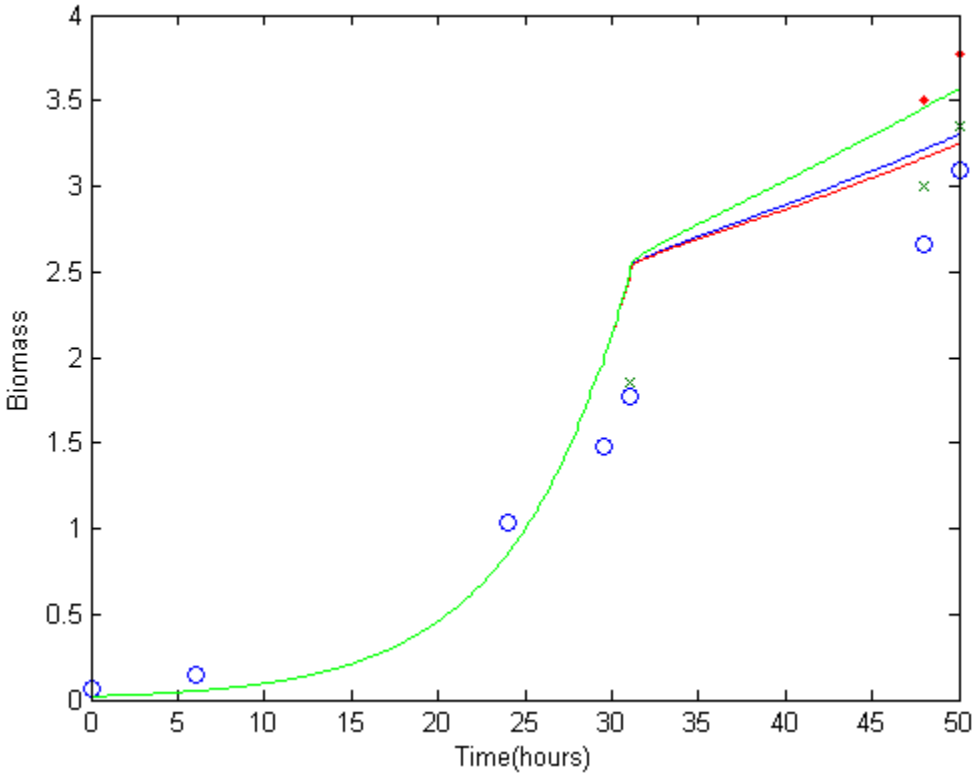


Figure 6-6 Biomass profile of *B. Pertussis* in fed-batch mode: red, blue, and green lines are model predictions with the fed-batch rate of 4.3, 6, and 7.4 g/h glutamate; circle, ×, and pink dots mark experimental data points for the feed rate of 4.3, 6, and 7.4 g/h glutamate.

6.5 Conclusion

The formulation of a dynamic metabolic flux balance model requires the identification of a suitable biological objective function to be maximized or minimized and active limiting constraints. In this contribution, we propose a systematic approach to identify both of these attributes for developing DFBA models. A key advantage of the proposed approach is that its formulation as a one-level optimization that simultaneously maximizes the biological objective together with terms that are related to the limiting constraints on consumption and production rates of amino acids. Proper weighting of these additional terms allows the constraint to be related to the relative errors incurred in measuring both amino acid concentrations as well as the biological objective variable. The use of set-based constraints ensures fitting between the data and the model predictions while avoiding the need for two-level optimization formulations. Two case studies were investigated. In the first case study for E. coli for which a priori known model was used to generate *in silico* data, the proposed approach successfully recognized the correct objective function and constraint of the model. In the second case study that involved experimental data of *B. pertussis* fermentation, the scarcity of data and the complexity of the metabolic network rendered the problem underdetermined. Additional constraints were required in order to only allow realistic solutions. In this case, our approach systematically identified a single kinetic rate bound on one amino acid (phenylalanine) and a set of constant upper bounds on consumption and production rates where the latter served to limit the solution space.

The resulting model was able to predict the evolution of biomass concentration with time whereas biomass values were not used for model calibration.

6.6 Acknowledgments

The authors would like to thank Natural Science and Engineering Research Council (NSERC).

Chapter 7

Inhibitory pH effect in culturing CHO cells

[3]This chapter is the contribution of the author of this thesis as the second author of the paper titled: ”Modeling of cell culture damage and recovery leads to increased antibody and biomass productivity in CHO cell cultures” (Naderi et al., 2014) so only an introduction and my main contribution to the paper has brought here. In this paper, a mathematical model was proposed for the investigation of cell damage of a Chinese hamster ovary (CHO) cell culture secreting recombinant anti-RhD monoclonal antibody (mAb). Reduction of pH was found to be a responsible factor for irreversible cell damage. A Tessier-based model was used to describe this irreversible damage to cellular function. In this model, the actively growing fraction of cells is dependent on an intracellular metabolic product acting as a growth inhibitor. To further verify the model, an offline model-based optimization of mAb production in the cell culture was carried out, with the goal of minimizing cell damage and thereby enhancing productivity through intermittent refreshment of the culture medium. An experimental implementation of this model-based strategy resulted in doubling of the yield as compared to the batch operation and the resulting biomass and productivity profiles agreed with the model predictions.

[3] This chapter is based on the paper: Naderi, S., Nikdel, A., Meshram, M., McConkey, B., Ingalls, B., Budman, H., & Scharer, J. (2014). Modeling of cell culture damage and recovery leads to increased antibody and biomass productivity in CHO cell cultures. *Biotechnol J*, *9*(9), 1152-1163. doi: 10.1002/biot.201300287

7.1 Introduction

Cell death following exponential growth in culture is still a major obstacle for efficient and economical production on an industrial scale. Cell death is primarily caused by apoptosis, a form of programmed cell death that can be triggered by environmental causes, such as depletion of essential nutrients, accumulation of metabolic by-products (ammonia, lactate and carbon dioxide), changes in pH, oxygen limitation, and hyperosmolality (Han, Kim, Kim, & Lee, 2010; Laken & Leonard, 2001). Autophagy, another death process, has been attributed to nutrient exhaustion (Hwang & Lee, 2008). Lysozymes released upon cell death affect both the titer and glycosylation, thus impacting both the quantity and quality of mAbs (Chee Furng Wong, Tin Kam Wong, Tang Goh, Kiat Heng, & Gek Sim Yap, 2005). An understanding of the mechanism and dynamics of these cell death processes will be crucial for the development of robust strategies for optimization of culture yield.

Fed-batch and perfusion operations are two efficient process strategies that are often applied to mitigate the stressful conditions that occur during batch culture. In both culture techniques, controlled nutrient feeding, either continuously or at regular intervals, is performed; spent medium (either in part or in whole) is removed in the perfusion mode (Rodrigues, Costa, Henriques, Azeredo, & Oliveira, 2010). These strategies result in longer cell viability and higher recombinant protein production (Burky et al., 2007; Zhang, Shen, & Zhang, 2004). However, formulating an optimal operating strategy (i.e. frequency and volume of feeding and medium removal) requires a detailed understanding of cell metabolism and physiology. Due to the complexity of cell metabolism and the many unknown intracellular factors involved in regulating cellular mechanisms, most strategies applied to date have been based on trial-and-error (Altamirano, Paredes, Illanes, Cairo, & Godia, 2004).

In another work the effect of key nutrients—glucose, glutamine, and asparagine on cell growth, maximum population density, cell death and longevity of the culture; given adequate initial reserves, these effects were found to be insignificant (Naderi et al., 2011a). Subsequently, the response of cultures to periodic feeding strategies was addressed in an attempt to improve culture conditions by compensating for the possible exhaustion of nutrients. Due to uncertainty about the identity of limiting nutrient(s), supplementation with complete medium was used. These earlier studies revealed that supplementation does not have a significant impact on mAb productivity. To model cell growth, a logistic model was used, because a limiting nutrient could not be identified.

Excursions in pH levels are common and very significant in large scale bioreactors, even when operated under closed loop pH control, due to spatial gradients within the vessel (Nienow, 2006). In this present work, we provide evidence suggesting that the accumulation of a pH reducing compound results in cell damage. Towards these objectives, a set of experiments were carried out involving partial or complete replacement of the spent medium to mitigate the effects of the observed inhibition. Periodic complete replacement of the spent medium—a method often referred to as intermittent harvesting (Ozturk & Hu, 2006)- can mitigate cell damage and may increase cell longevity and productivity. Moreover, it appears that some cells may undergo replicative senescence rather than direct transition to apoptosis. As described below, once stationary phase was reached, most cells were committed to growth arrest, but there was a small fraction of cells that reverted to growth after each period of intermittent harvesting. This effect may be related to oncogene-induced growth arrest or replicative senescence under unfavorable environmental conditions, which has been reported previously (Ruiz et al., 2008).

7.2 Materials and Methods

7.2.1 Cell Line and Culture Process

A CHO (DHFR-) cell line, containing a fully human IgG-Variant gene for anti-RhD mAb, was provided by our industrial partner, Cangene Corporation, Winnipeg, MB. This recombinant mAb (IgG1-r9B8) has shown a relatively high affinity for the human RhD antigen, making it suitable for therapeutic application. The CHO cells were cultured in SFX-CHO (Hyclone) medium, supplemented with 1% serum (Invitrogen) and L-glutamine (Sigma) to a final concentration of 1 mM. No animals or human subjects were used in this research.

7.2.2 Analytical Methods

Viability assays were carried out with Trypan Blue staining followed by cell counting on a hemocytometer. The glucose content of cell culture samples was measured enzymatically using a glucose test kit (Megazyme Glucose Test Kit). Lactate was assayed with a Lactate Plus meter from Nova Biomedical. The dissolved ammonia concentration was measured using a pH/ISE meter equipped with an Ammonia Ion-Selective electrode (VWR, model 710A). Dissolved oxygen (DO) was measured using VWR Symphony Dissolved Oxygen meters and corresponding probes. Electrode calibration was performed in water-saturated air in a special calibration chamber. The pH of the broth was measured by VWR SympHony SP21 meter.

7.2.3 Experimental Design

To corroborate and model the adverse effect of pH on growth for the cell line under consideration, experiments were conducted in T-flasks at different initial pH values using 1N HCl for pH adjustment. Cell growth and pH were measured routinely in each flask. The culture was adherent in T-flasks.

Cell culture experiments were conducted in 500 ml spinner flasks (Bellco, USA) in a humidified CO2 incubator (Sanyo IR Sensor, 37°C, 5% CO_2) at 100 rpm. The inoculum density was approximately 0.2 $\times 10^6$ cells/mL. In this series of experiments, pH was left intentionally uncontrolled. Samples were taken daily for immediate and offline analysis of cells and metabolites.

Inhibitory effects of pH on growth of mammalian cells have been reported previously (Link et al., 2004; Lipscomb, Palomares, Hernandez, Ramirez, & Kompala, 2005). To verify the effect of pH levels on cell growth in the system under study, 3 sets of culture at different initial pH were prepared. Initially, the pH of 3 shaker flasks containing fresh medium (SFX-CHO) was adjusted at 7.6, 7.1 and 6.8, respectively, with 1N HCl. The initial cell density of CHO cell seeded in each shaker flask was $0.2x10^6$ cells/mL with 1% serum and 1mM glutamine.

The culture from each shaker flask was split into 5 T-flasks 20 mL of cell culture in each. All 15 T-flasks were incubated at 37°C and 5% CO_2. During the 5 days incubation period, cell concentration and pH of one T-flask in each set were measured. The results in Fig.7-1A and Fig.7-1B show that a reduction in pH results in significant reduction in cell growth. The differences were statistically corroborated by an ANOVA test ($p=0.05$) for the 3 pH levels tested for the viable cell densities measured at 63 hrs and 85 hrs which correspond to the maximal densities. On the other hand, the differences among the cultures at the beginning or end of the T-flask experiments are not statistically significant.

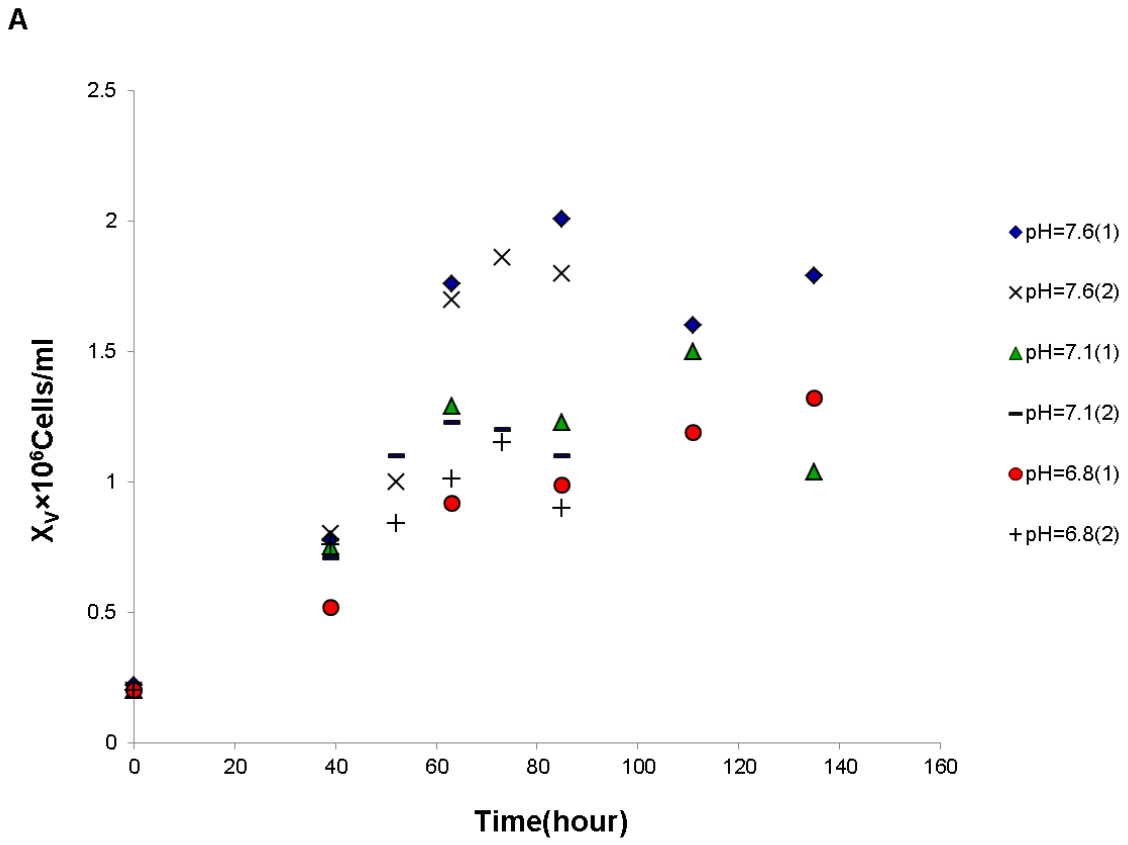


Figure 7-1 A Total viable cell concentration in T-flask. 1 ml of each flask was used for cell counting by Trypan Blue staining. Two repeats were done for each of a total of three pH levels.

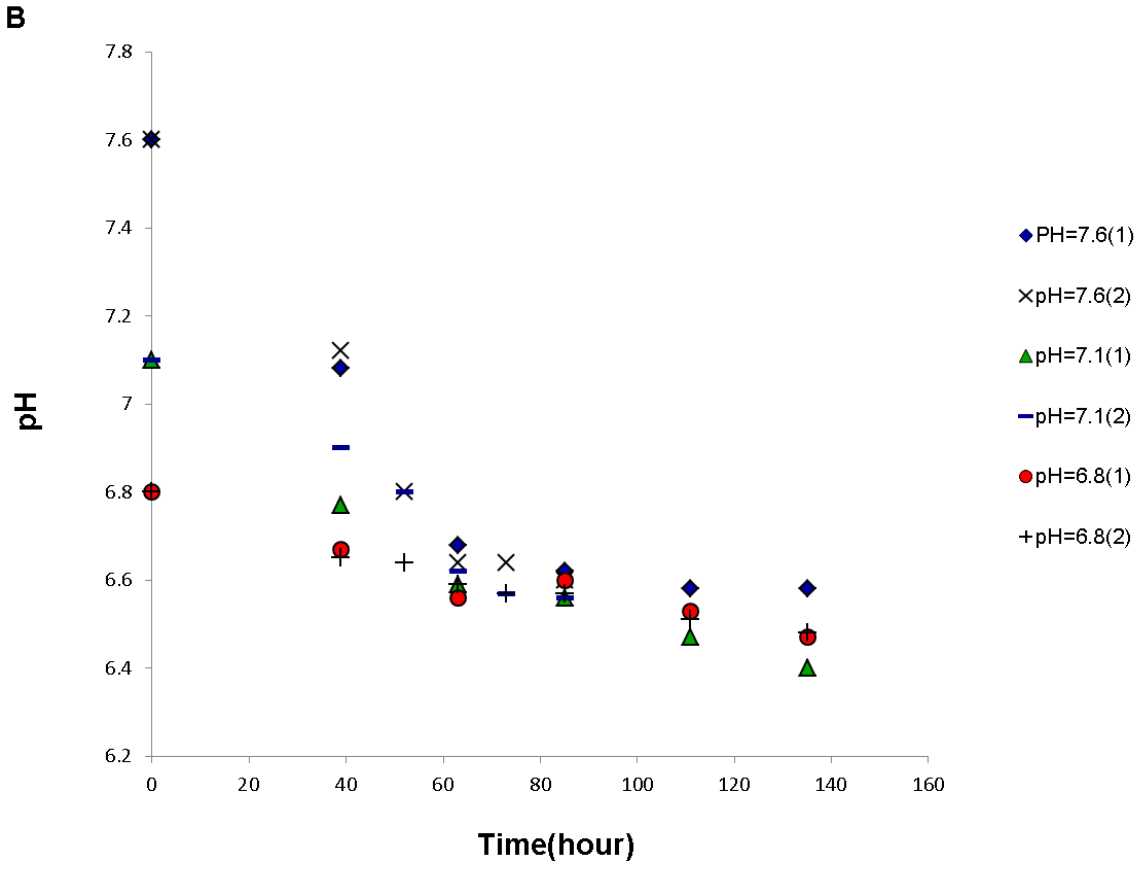


Figure 7-1B pH of the culture in T-flask. 2 ml samples from each flask were used for pH measurement. Two repeats were done for each of a total of three pH levels. After measurements of cell count and pH the corresponding flask was discarded.

To assess the effect of gaseous (oxygen/carbon dioxide) mass transfer effects on the pH, a series of experiments were performed in spinners at differing surface area to volume ratios (250 mL and 500 mL) at identical agitation speeds (100 rpm) and at initial pH=7.6 adjusted by base addition. It was hypothesized that, for different culture volumes, mass transfer conditions may lead to different rates of accumulation of acidic metabolites. The use of different culture volumes at identical stirring speeds

provided a simple and convenient means to evaluate the potential effect of gas/liquid mass transfer on basal metabolism. The expected differences in pH levels for spinners with different culture volumes were fully corroborated as shown in Table 7-1 and in Figure 7-2, which indicates that for the 500 mL spinner consistently lower level of pH were attained as compared with the 250 mL spinner. The differences in pH between the two volumes were corroborated by a t-test (p=0.05) for all pH data obtained after time=48 hours (F=5.5 and $F_{critical}$=3.2). Periodically a portion of the spent culture (50mL in the spinner containing 250 mL and 100 mL in the spinner containing 500 mL) was withdrawn and substituted with an equivalent amount of fresh medium. Before each withdrawal, we allow the cells to settle down (~15 min.) so as to minimize cell loss during withdrawal. It was expected that, if an inhibitory compound were accumulating in the supernatant, each withdrawal would help to renew growth by reducing the level of such compound.

Table 7-1. Dissolved oxygen concentration (DO %) and pH of culture for two different volumes in 500 mL capacity spinner flasks.

	V=250mL		**V=500mL**	
Time (h)	DO%	pH	DO%	pH
145	33.2	6.62	38.6	6.46
196	27.0	6.50	40.1	6.32
244	49.1	6.43	55.3	6.37
293	43.9	6.44	46.5	6.36
338	45.6	6.47	49.9	6.37

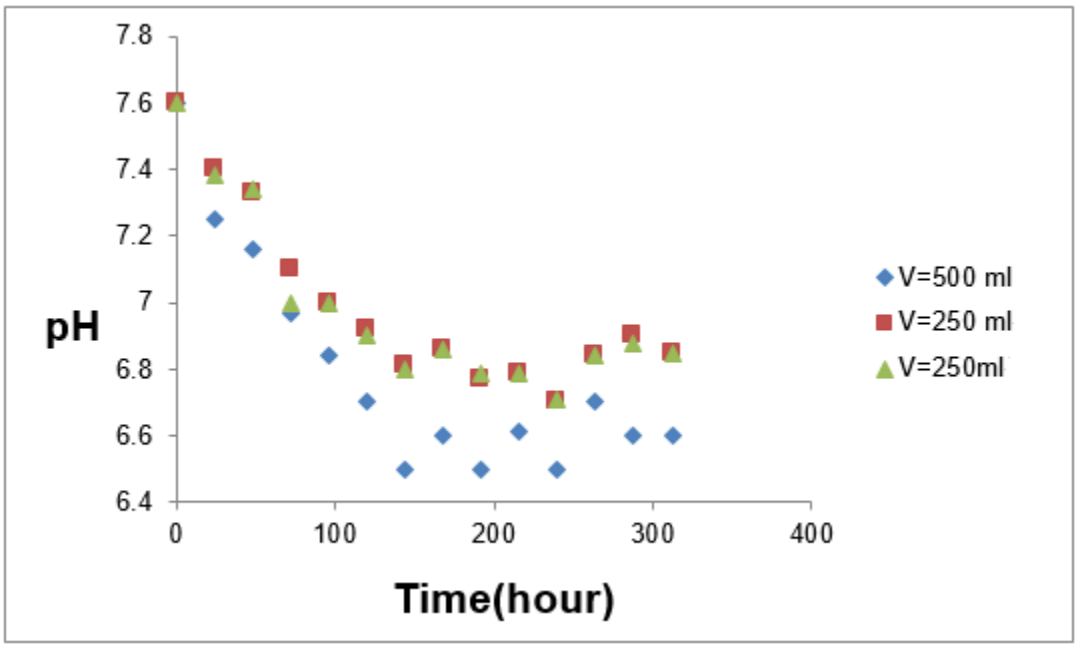


Figure 7-2 pH profile during partial intermittent feeding culture at volumes of 250 ml and 500 ml. Feeding volume of 50ml and 100 ml at the times of 145, 196, 244 and 293 (hours)

The total viable cell density assays for these two spinners are presented in Fig.7-3A. The differences in cell density between the two volumes were corroborated by a t-test (p=0.05) for densities collected after t=49 hrs (F=5.31 and $F_{critical}$=2.53). Following the first withdrawal, the cultures recovered and reached cell density maxima and cell viability close to those achieved before dilution (almost equal, for the spinner containing 250 mL). This behavior is consistent with the hypothesis that the accumulation of an acidic compound results in pH reduction, leading to growth inhibition. However, no improvements in viability were observed in the subsequent withdrawal steps. The time course of viability in the 250 mL and 500 mL cultures is shown in Figure7-3B. A progressive decline in viability was observed in spite of intermittent partial refreshment of the medium. Dissolved oxygen concentration measurements given in Table 7-1 for both spinners confirmed that dissolved oxygen was not in the limiting range,

even at peak cell density. The levels ranged between 30%-90% of air saturation at which DO has no significant effect on cell metabolism (Trummer et al., 2006). On the other hand the pH levels, also shown in Table 7-1, were found to be below the range of 6.80-7.60, which has been reported as optimal for other CHO cell lines (Link et al., 2004; Yoon, Choi, Song, & Lee, 2005).

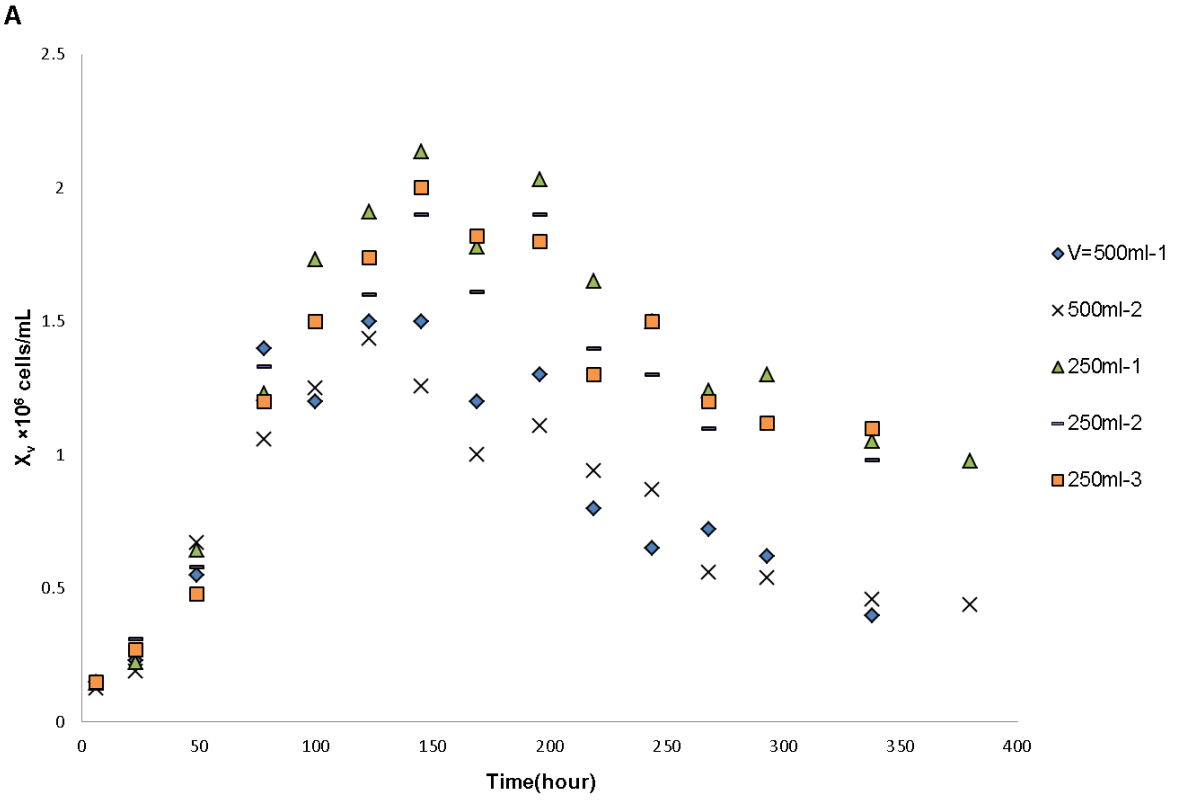


Figure 7-3A Harvesting operation at two different volumes. V=250 mL, ΔV=50 mL, V=500mL, ΔV=100 mL. Harvesting occurred at the times of 145,196,244,293 hours. Three repeats were done for 250 ml and two repeats for 500 ml.

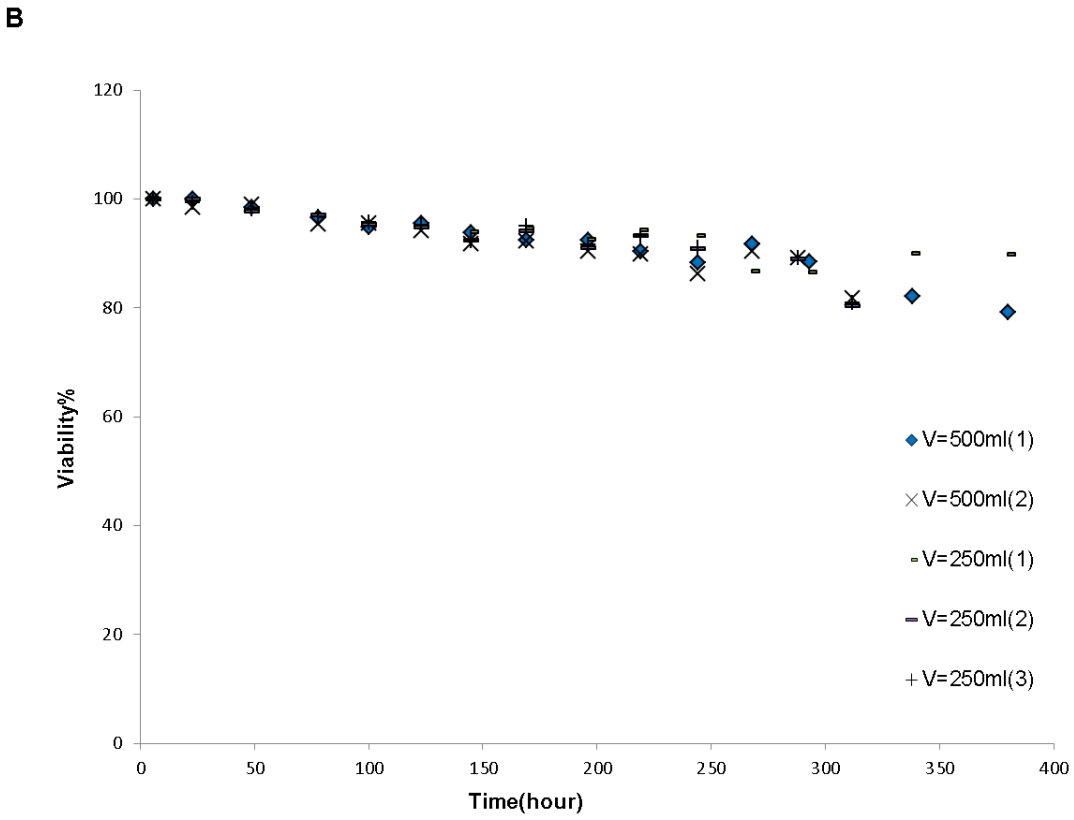


Figure 7-3B Viability graph. Harvesting operation at two different volumes. V=250 mL, ΔV=50 mL, V=500mL, ΔV=100 mL. Harvesting Occurred at the times of 145,196,244,293 hours. Three repeats were done for 250 ml and two repeats for 500 ml.

Figures 7-3C and 7-4A-B show the metabolic condition observed in the spinners operating with 250ml and 500ml working volume, respectively. As shown, the higher cell density in the low-volume spinner led to a higher relative consumption rate of glucose in the post-exponential phase. They also show lactate and ammonia concentrations are nearly identical and below inhibitory level in both cultures.

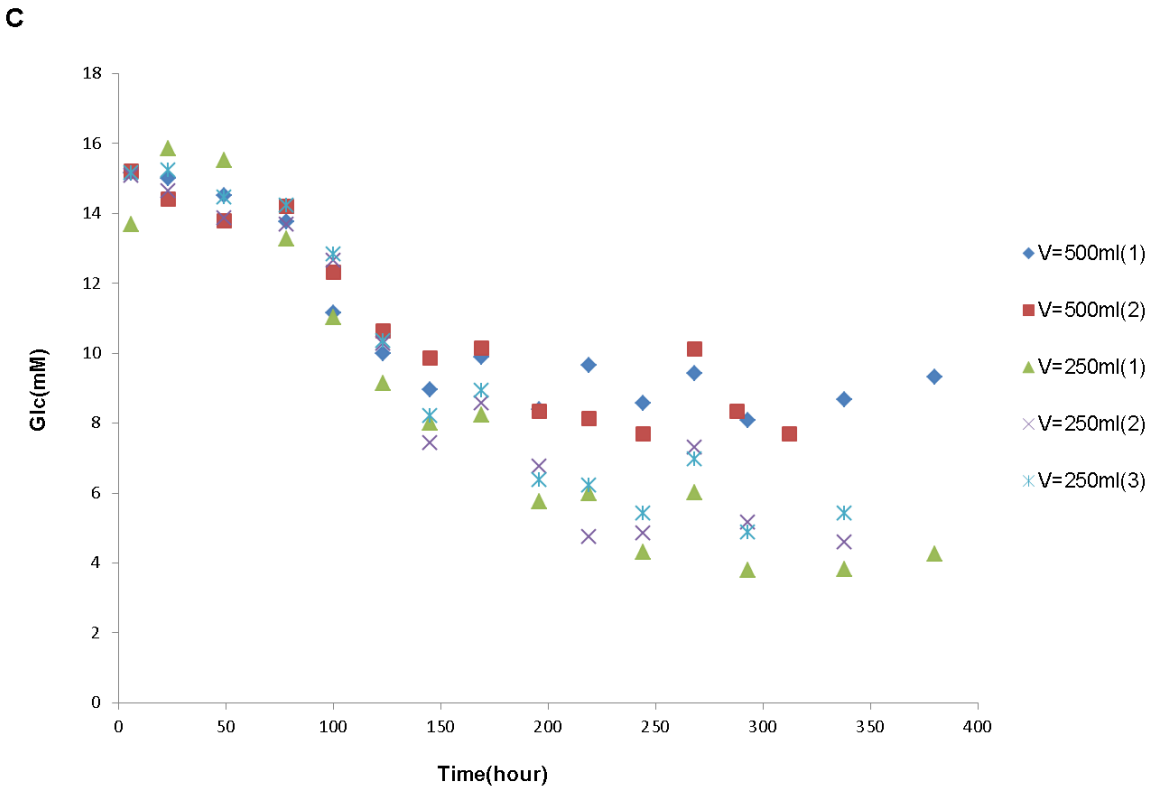


Figure 7-3C Glucose profile for 250 and 500 mL flasks. Harvesting operation at two different volumes. V=250 mL, ΔV=50 mL, V=500mL, ΔV=100 mL. Harvesting Occurred at the times of 145,196,244,293 hours. Three repeats were done for 250 mL and two repeats for 500 mL.

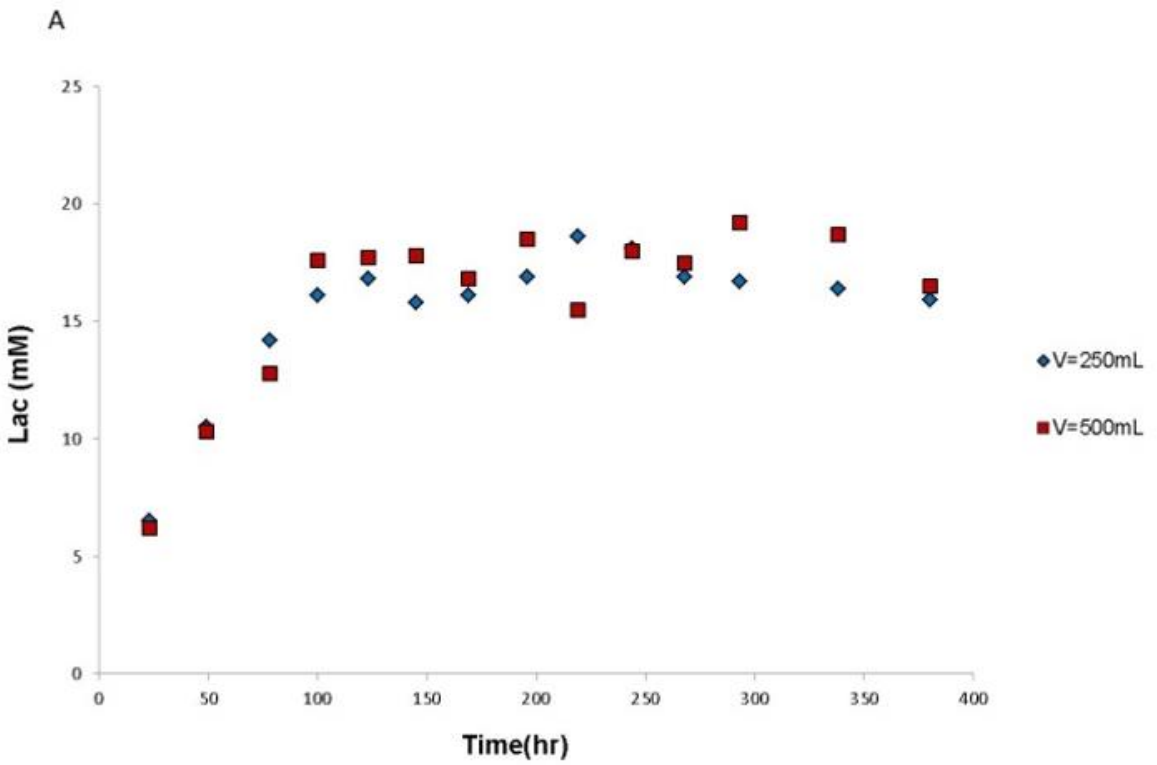


Figure 7-4A Lactate profile for 250 and 500 mL flasks. Harvesting operation at two different volumes. V=250 mL, ΔV=50 mL, V=500mL, ΔV=100 mL. Harvesting Occurred at the times of 145,196,244,293 hours.

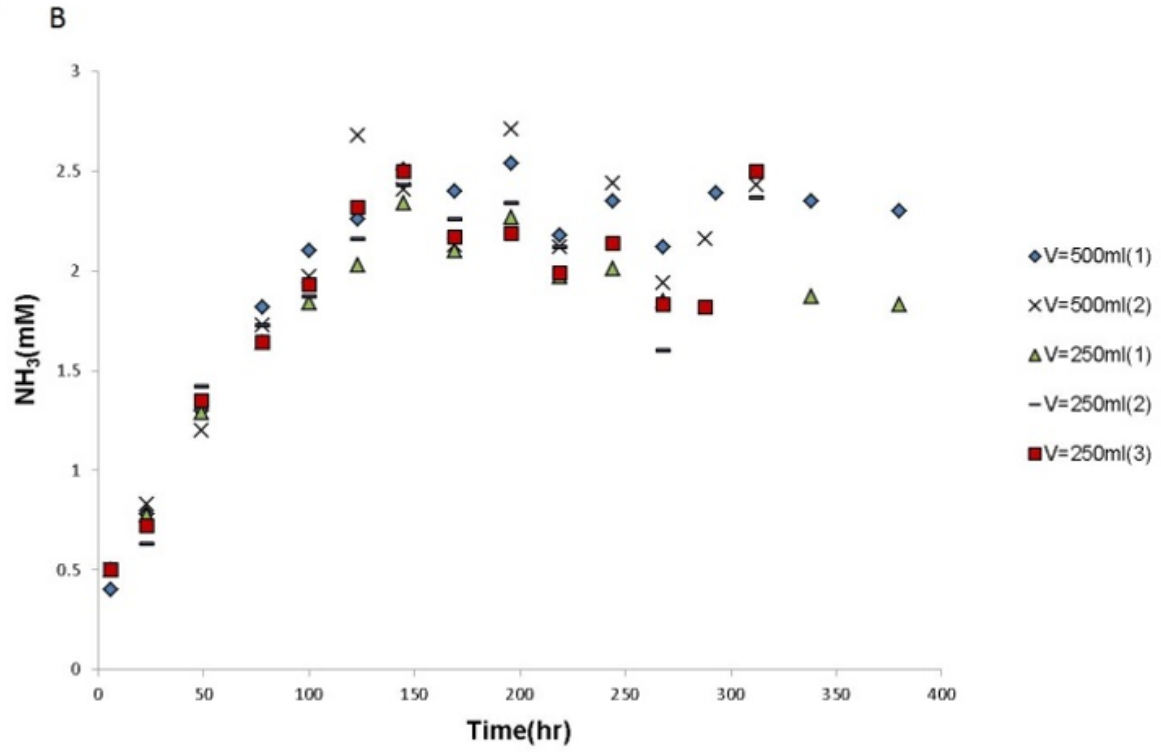


Figure 7-4B Ammonia profile for 250 and 500 mL flasks. Harvesting operation at two different volumes. V=250 mL, ΔV=50 mL, V=500mL, ΔV=100 mL. Harvesting Occurred at the times of 145,196,244,293 hours. Three repeats were done for 250 mL and two repeats for 500 mL

The observed decline in viable cell concentrations during intermittent harvesting operation (Fig.7-3A) suggests that an irreversible growth arrest occurred during long-term periodic operations. This steady decline in viability rate coincided with a decrease of pH to growth-inhibiting level as shown in Table 7-1 (pH< 6.6). Oncogene-induced growth arrest or replicative senescence has been reported to occur under harsh environmental conditions and it is often related to p53 regulation (Ruiz et al., 2008) and the effect of low pH on the regulation of p53 and cell damage has been clearly established (Xiao, Li, Yang, & Liu, 2003) .

To further assess whether the reduction in growth is due to depletion of a substrate or whether there is accumulation of an inhibitory compound in the medium two additional experiments were conducted as follows: (i) a fed-batch culture supplemented with 5× concentrated feed; and (ii) a full harvesting operation. The full harvesting culture was prepared with 500 mL of 2× concentrate SFX-CHO medium; the fed-batch culture employed an initial volume of 300 mL normal (1×) SFX-CHO medium. Once the cell density had reached an approximate maximum (80 h), the culture conditions were manipulated. In the full harvesting spinner, complete renewal of the culture supernatant was performed (by centrifuging of the total culture, discarding the supernatant and returning the cells in a same volume of fresh medium). In the fed-batch spinner, four feedings of 50 mL concentrated 5× SFX-CHO were added every 48 h, starting at 80 h. The results are shown in Fig. S1A–D (See Appendix D). Figure S1A, which shows viable cell concentration illustrates the growth-inhibition effect of pH at different working culture volumes. Until the peak in cell concentration is reached (~80 h) these experiments are equivalent to the experiments described in Fig. 7-3 and Fig 7-4. While both spinners started with equal cell density, a higher peak cell level was achieved in the fed-batch spinner, which started with a volume of 300 mL compared with the 500 mL culture used for intermittent harvesting. However, in the fully harvested spinner, not only was cell growth renewed after the harvesting event (80 h), but the cells maintained higher density for a significant time (~100 h). Conversely, in the fed-batch run, no improvement in viability was observed after each feeding, and the cells gradually entered a decline phase during which viable cell density and total viability decreased. Metabolite levels for these runs are shown in Fig. S1B–D. The glucose level (Fig. S1-C) never falls below the limiting concentration, due to refreshing of the culture. The glutamine consumption rate in the full harvesting culture was low (Figure S1-D), while glutamine was depleted in the fed-batch culture due to consumption and dilution. Figure S1-B shows lactate production, revealing a high value of lactate at the early stage of the culture (which may have

been caused by the high initial cell density). In the full harvest culture, lactate was produced quickly after the medium replacement event: the concentration increased from near zero to over 10 mM in 24 h. This rapid production of lactate demonstrates how quickly culture conditions can be changed by the metabolic activity of a high viability culture. In a separate experiment, an attempt was made to improve growth in a fed-batch mode by controlling pH at 7.1–7.2 by manual addition of NaOH (Data not shown). Although the growth in this pH-controlled culture was slightly higher than for cultures without pH regulation, a metabolic shift toward glycolysis gave rise to a very large specific consumption rate of glucose, almost 6 times larger than the consumption rates for cultures without pH regulation. Glucose was completely consumed after 3 days at which time the growth stopped and the lactate levels reached 45 mM. In principle, we could have added more glucose after its depletion but this would have resulted in further lactate accumulation and so it could not lead to improvements in yield. Thus, for improvement of yield, we concluded that an intermittent harvesting operation is a better approach.

7.3 Conclusion

This chapter describes the development of a comprehensive dynamic model for Chinese hamster ovary (CHO) cells producing anti-RhD monoclonal antibody (mAb) that accounts for metabolic behaviour over time. Development of the model required identification of the underlying parameters that play key roles in cell viability and monoclonal antibody productivity.

It was found that a decrease in pH, attributed to the accumulation of an acidic compound, significantly inhibited cell growth. Although this suppression effect can be temporarily mitigated by partial or full replacement of culture with fresh medium, the culture rapidly returns to unfavorable conditions (due to high cell density and subsequently high metabolic activity). As a consequence, cells gradually—and irreversibly—lose their capacity for regeneration.

We extended our previously published model to account for the cumulative reduction in pH and consequent growth inhibition. The model accounts for the effects of several significant factors on cell growth and metabolic activity, and model simulations successfully reproduce the time profile of cell populations and major external metabolites. It should be noticed that the model does not account for effects that have not been explicitly tested in the experiments such as glucose depletion. To test the accuracy of the mathematical model it was used as the basis for an offline optimization of an intermittent harvesting schedule. Our experimental results showed that the predicted intermittent refreshment of the medium at optimal instances minimized cell damage and significantly improved productivity as compared to a batch operation. The novel description of accumulated cell damage presented here can be used to incorporate the dynamics of this crucial mechanism into models of CHO cell culture, thus improving the accuracy of model-based predictions of culture yield.

Chapter 8

Identification of a dynamic flux balance model of a CHO cell culture

Mathematical models are essential tools for improving the understanding of cell metabolism and cell culture behavior and for their use in model-based bioprocess optimization. Dynamic flux balance analysis (DFBA) type of models are based on the premise that biological cells have adapted through evolution to optimally allocate resources to achieve certain biological goal subject to some limitations. A key challenge in the formulation of a DFBA is to identify from experimental data the specific biological goal that the cell seeks to optimize (objective function) and a minimal set of kinetic bounds (constraints) so as to avoid overfitting of the data used for model calibration. This paper applies a systematic methodology for identifying an objective function and constraints necessary for describing a dynamic culture of mammalian CHO (Chinese Hamster Ovary) cells. Using this approach we are able to calibrate a model that describes the data up to an a priori known margin of noise and which provides rational explanations for biological phenomena such as the excessive production of lactate by anaerobic fermentation during the growth phase of the cell culture referred to as the "*Warburg effect"*. The model that best explained the data during the exponential phase of growth is based on the maximization of growth rate combined with the minimization of NADH production in the cytoplasm and minimization of NADPH consumption in mitochondria and requires constraints on the consumption/production rates of only 5 metabolites.

8.1 Introduction

The success of monoclonal antibodies has significantly increased their demands and created a need for improving production yield through the optimization of cell lines and manufacturing processes. Mammalian cells, in particular, Chinese Hamster Ovary cells, are the host of choice for monoclonal antibody production.

In order to optimally manipulate mammalian cell systems whether at the genetic or at the bioprocess level, it is imperative to understand the metabolic behavior of the biosystem under study and a mathematical model is a good tool to gain such understanding.

A very common approach in biochemical modeling is kinetic modeling of the main reactions involved in the process. These models are based on a system of differential equations that describe the evolution of specific metabolites with respect to time by empirical mass balances where the consumption of the metabolite towards biomass or towards another metabolite or the production rates or by-products are determined by empirical Monod kinetic terms and yields. Each one of these kinetic expressions involve maximal rate coefficients, half saturation constant and yield coefficients thus resulting in a large number of calibration parameters that need to be identified from available experimental data. Therefore, these types of models are generally heavy parameterized when many metabolites are considered thus rendering them very sensitive to noise and inaccurate especially when data is scarce. Furthermore, the numerical estimation of parameters for these models from the solution of nonlinear optimization problems is challenging due to the large number of parameters and the nonlinearity of the kinetic expressions that may lead to the occurrence of multiple minima.

An alternative modelling approach is based on the Dynamic Flux Balance Analysis (DFBA) algorithm (Mahadevan et al., 2002) DFBA is based on the formulation of an optimization problem where a

particular flux distribution in a metabolic network is sought such as a specific objective function is optimized subject to certain constraints. The rationale behind DFBA is that natural evolution had conditioned the cell to act as an agent that allocates resources such as a biological objective is optimized. Being based on an optimization with few limiting constraints, DFBA models have been reported to be advantageous over other aforementioned modelling approaches because for certain micro-organisms such as bacteria the have often required a smaller number of calibration parameters

A DFBA model can be formulated as follows:

@ each sampling time i

max $c^T v$

subject to $v_i > 0$ (8-1)

$f(\psi_i) \leq \frac{1}{TX_i}(\Psi_{i+1} - \Psi_i)$ =A. $v_i \leq g(\psi_i)$

$\Psi_{i+1} = \Psi_i$ + A. v_i X_i T (T= time interval)

Where $\bar{X}$ is the average biomass concentration, X_i is the biomass at time interval i, $f(\psi_i)$ and $g(\psi_i)_r$ are lower and upper consumption or production rates of constraining metabolites concentration. These metabolic rates can be defined as a function of the concentration of a metabolite participating in the corresponding reaction. These rate constraints must be determined from experiments. The sampling time interval, T is selected to discretize the mass balances for the purpose of numerical integration (Budman et al., 2013; Mahadevan et al., 2002).

Although in the past researchers have chosen the objective function and constraints in model (8-1) by trial and error, in our recent work (Nikdel & Budman, 2016) we have proposed a systematic approach for finding those from data and illustrated this method for a culture of microbial (non-mammalian) systems. For bacterial and microbial organisms typical objectives that can be successfully used to

describe experiments are maximization of growth and/or maintenance of a particular redox potential. However, suitable optimization objectives they have not been proposed as yet for dynamic metabolic models of mammalian cells due to their relative more complex behavior as compared to bacteria. For example, mammalian cells exhibit particularly regulatory mechanisms such as programmed cell death (apoptosis) that have to be taken into account when defining an objective function for optimization. Accordingly, systematic approaches for identifying DFBA models for mammalian cells have not been reported as yet.

A key challenge in the identification of dynamic metabolic flux models from data is that they require minimizing a sum of square errors between data and model predictions over the duration of the experiment while satisfying at each time interval the optimization in (8-1) thus resulting in an overall bi-level optimization problem. Since the resulting bi-level optimization is numerically challenging and prone to convergence to local minima, we proposed (Nikdel & Budman, 2016) instead a single level optimization algorithm, briefly summarized in Section 2, to identify DFBA models based on the idea of set based description of the experimental data. In this study we apply this methodology, which previously was used for a microbial systems, to mammalian cell data. The goal here is finding a suitable objective function and a set of kinetic constraints to describe the behavior of mammalian cells by a DFBA model. Another key objective of the current study is to assess whether a dynamic metabolic flux model of mammalian cells is indeed will require a significantly lower of tuning parameters as compared to previously reported dynamic models (Naderi et al., 2011a).

This chapter is organized as follows. In Section 2 the algorithm proposed by authors (Nikdel & Budman, 2016) for identifying the DFBA constraints, is summarized. Section 3 discusses the mammalian cells study including the metabolic network used and the experimental data. Section 4

presents the identification of the mammalian cell DFBA model and discusses the results followed by Section 5 with Conclusions.

8.2 Summary of the algorithm for identifying DFBA constraints

The method for finding the limiting constraints is based on representing the data by convex sets. It is assumed that because of sensor noise or unaccounted disturbances, the metabolites levels are bounded by upper and lower limits at each time interval for which data is collected. For example, set constraints for glucose concentrations during a batch culture that consumes glucose are illustrated in Figure 8-1.

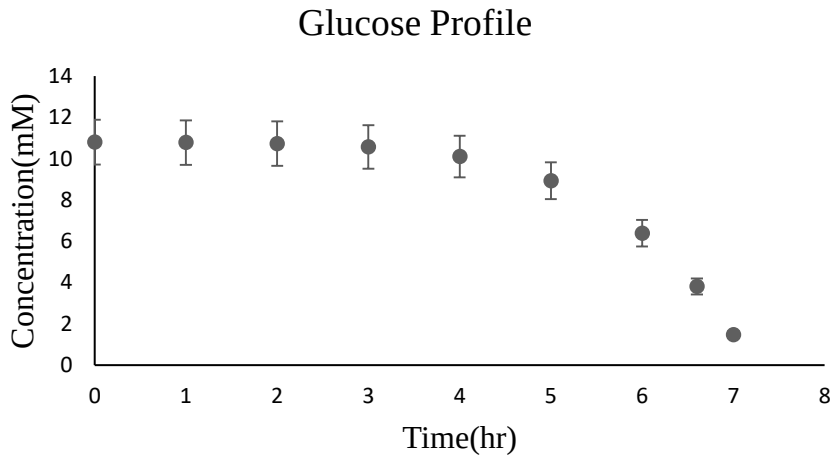


Figure 8-1 set constraints for glucose concentrations during a batch culture

Thus, the actual value of metabolites concentrations at each time interval are bounded within constraints that are referred heretofore as *sets*.

Using this set-based description of the data a 3-step procedure is used to find a DFBA model as follows:

Step 1: Calculate at each time interval a flux distribution from (8-1) that maximizes the objective function to be chosen for the process, e.g. maximization of growth subject to the set based constraints. Mathematically, the problem is expressed based on (8-2) as follows:

@ all T intervals (k=1,…,T) (8-2)

$$\max_{\mathbf{v_k}} \mathbf{c^T}.\mathbf{v_k}$$

Subject to:

$$\frac{-1}{\Delta t.X_k}(\psi_k - \psi_{k+1}^L) \leq \mathbf{S}.\mathbf{v_k} \leq \frac{-1}{\Delta t.X_k}(\psi_k - \psi_{k+1}^U)$$

Where matrix S is the stoichiometric coefficients, $\mathbf{v_k} = (\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n})_\mathbf{k}$ is the vector of fluxes of n metabolites at the time interval of k. $\boldsymbol{c^T}.\mathbf{v_k}$ is the cost to be maximized or minimized (objective function), ψ_k and X_k represent measured concentrations of extracellular metabolites and biomass at the time of k. The inequality constraints in (8-2) are derived from Euler discretization of the general mass balance equation for each metabolite ψ_m ($\frac{d\boldsymbol{\psi}}{dt} = \mathbf{S}\boldsymbol{v}X$) as done in (8-1). The inequality directions in equation 8-2 are applied for a metabolite that is being consumed whereas the directions are inverted for a metabolite that is being produced. It should be noticed that while in (8-1) the constraints are given as a function of the concentrations through the vector functions *f* and *g* to present kinetic limitations on the rate of reactions, in (8-2) the functions are unknown at this stage and they must be identified. Instead the production/uptake rates in (8-2) given by $\mathbf{S}.\mathbf{v_k}$ are constrained based on the upper and lower bounds identified directly from data, ψ_{k+1}^U and ψ_{k+1}^L respectively.

Step 2: Identify a minimal number of metabolites which uptake/production rates need to be constrained as in (8-2) in order to fit the experimental data described by convex sets.

To find the uptake/production rates which are limiting the solution space of equation (8-2), a Lagrange multiplier approach(Bertsimas & Tsitsiklis, 1997) is formulated as follows:

$$\mathbf{v_k}=\underset{\mathbf{v_k}}{\operatorname{argmin}}(-\mathbf{c^T v_k}) \quad (8\text{-}3)$$

S.t: $G_{LI}(\mathbf{v_k}) \leq 0$

$G_{LI}(\mathbf{v_k}) \leq 0$ are the inequality constraints lower and upper bounds of the production or the uptake rate of measured metabolites . If the Lagrangian is defined as follows:

$$L(\mathbf{v_k}, \lambda) = \mathbf{c^T v_k} - \lambda\, G_{LI}(\mathbf{v_k})$$

The Lagrange multipliers and the optimal solution $\mathbf{v_k}$ to (8-3) can be obtained from the following set of conditions:

$\lambda \geq 0$ (8-4a)

$G_{LI}(\mathbf{v_k}) \leq 0$ (8-4b)

$\lambda \times G_{LI}(\mathbf{v_k}) = 0$ (8-4c)

$\nabla_{\mathbf{v_k}} L(\mathbf{v_k}, \lambda)=0$ (8-4d)

Where $G_{LI}(\mathbf{v_k})$ is an inequality constraint, $\mathbf{v_k}$ is the optimal solution of the fluxes. By solving equations (4a-4d) the value of the Lagrange multipliers (λ) for each of the constraints can be calculated. Equation (4c) represents the complementarity condition that forces either the Lagrange multiplier or the corresponding inequality to be zero. Based on the interpretation of the Lagrange multipliers a $\lambda = 0$ implies that by relaxing the corresponding constraint the value of the function which is intended to be maximized does not change indicating that, this constraint is not active. In contrast, a large value of the Lagrange multiplier is an indicator that the constraint is active. From the active constraints, it is possible to assess which metabolite's consumption or production rates are limiting constraints at each interval of the solution of the DFBA.

Step 3- Find kinetic expressions as a function of metabolites' concentrations that can describe the time evolution of the active constraints found in step 2. To accomplish this, at each time interval, the active constraints are formulated with a particular type of kinetic expressions e.g. Michaelis-Menten or other as a function of the concentration of metabolite that participates in the reaction associated to the active constraint as follows:

$$\frac{\mathrm{d}\psi_{\mathrm{m}}}{\mathrm{dt}} = \frac{\mathrm{V_{m,max}}.\psi_m}{\mathrm{K_m}+\psi_m} X \qquad (8\text{-}5)$$

Where, $\frac{d\psi_m}{dt}$ is the uptake or production rates of a metabolite (active constraints) identified in step 2 as a function of its concentration ψ_m at each corresponding time interval and $\mathrm{V_{m,max}}, \mathrm{K_m}$ are the kinetic parameter values to be identified as described.

The identification of the parameters in (8-5) is done by regression of the values of the production/uptake rates $\frac{-1}{\Delta t.X_k}(\psi_k - \psi_{k+1}^U)$ and/or $\frac{-1}{\Delta t.X_k}(\psi_k - \psi_{k+1}^U)$ that were identified as active in step 2 with respect to the corresponding value of the metabolite concentration ψ_k. Using these parameters it is possible to generate a predictive dynamic metabolic flux model as shown in Equation (8-1) where expressions (8-5) are used as the functions *f* or *g* in (8-1).

The main numerical advantages of this three-step approach, are: i- step 1 and step 2 involve the solutions of a linear programming problems and therefore are simple to compute and ii-step 3 involves separated fitting of each metabolite profile as compared to other modelling approaches that require fitting of all the metabolites by simultaneous calibration of a larger number of model parameters.

The linear programs defined in this section (8-2)-(8-4) were solved using the *cplexlp* solver ('interior-point' algorithm) of IBM ILOG CPLEX for MATLAB Toolbox.

In an earlier work (Nikdel & Budman, 2016) showed that the above 3 step procedure is able to successfully identify an assumed dynamic metabolic flux model of *E.coli* from *in silico* data generated

by that model thus corroborating the ability of the algorithm in identifying the model used for generating the *in silico* data.

8.3 Mammalian cells study: metabolic network and experimental data

8.3.1 Metabolic network of reactions used to formulate the matrix S in equations (8-1) and (8-2).

The stoichiometric matrix ***S,*** in (8-1) and (8-2) is obtained from the main stoichiometric reactions for the mammalian cell under study. The metabolic network in this study it was taken by combining group of reactions used in different studies that were considered relevant to describe experiments with CHO cells (Naderi et al., 2011a; Zamorano, Wouwer, & Bastin, 2010) such as TCA cycle, glycolysis, amino acids' synthesis reactions and transamination reactions. The glutamine synthesis reaction was added since the cell line used in the current work is able to synthesize glutamine. This resulting network involves 38 reactions including one reaction describing the formation of biomass as a function of amino acid contributions. The reactions considered account for the main contributors to the carbon and nitrogen balances. For simplicity and since they were not measured, the balances for co-metabolites such as ATP/ADP and NADH/NAD+ are not explicitly considered at this stage.

Since the model calibration and comparisons have been limited in this study to the growth phase only, mechanisms of cell necrosis which are predominant during the post-exponential phase were not explicitly considered. However, we have considered in the model the occurrence of apoptosis (programmed cell death) which is expected to occur in different degrees from the beginning of the batch culture.

8.3.2 Experimental data used for model calibration/validation

The data was obtained with a CHO (Chinese Hamster Ovary) cell line modified GS producing anti-rabies IgG. This cell line is able to synthesize glutamine so the latter does not have to be added for growth.

Two different experimental media from (MilliporeSigma, St. Louis, MI, USA) - medium 1 and medium 2 were used for experimental purposes. Cells were in their exponential phase when inoculated for the actual experiment. Cells were inoculated at the cell density of about 1.1 million cells/mL for medium 1 and 2.4 million cells/mL for medium 2 in a 3 liter bioreactor with 1.5 liter working volume. Experiment one was continued for 50 hours and experiment two using medium 2 that had a richer medium was continued for 96 hours.

Cell density and viable cell density of the cultures were obtained from a Trypan blue exclusion method. Glucose, lactate, ammonia, and glutamine concentrations were measured with a Nova BioProfile analyzer. Amino acids were measured by HPLC. Due to the nature of the analytical methods data included amino acids and biomass concentrations were only available once per day. On the other hand, this data sampling frequency was insufficient for generating smooth estimates of the uptake rates bounds. Therefore, we used non-linear polynomial regression in MATLAB to interpolate the data and generate smoother estimates of metabolites' rates of production or consumption.

8.4 Identification of DFBA model for CHO cells

One of the main goals of the current work was to identify a DFBA model that can explain some of the particular behavior exhibited by most mammalian cell systems such as very high production of lactate during the growth phase and cell death due to apoptosis (programmed cell death).

Flux balance models have generally adopted the optimality assumption of evolutionary biology whereby cells are using resources optimally to maximizing or minimizing a specific objective function in order to survive. Maximization of growth or biomass (Orth et al., 2010) has been commonly used for most studies dealing with bacteria. The steady-state FBA modeling studies that have been reported involving different objective functions have been mostly done *E.coli*. Here we show that defining the right objective function can greatly impact the DFBA model prediction accuracy for mammalian systems.

The formulation of the DFBA model involves identifying a suitable objective function $c^T v$ equations 8-1 and 8-3, and rate limiting constraints corresponding to particular amino acids as per the method presented in Section 2.

The approach adopted in this work is to test the 3-step identification procedure with different objective functions $c^T v$ (equation 8-3) and check a posteriori the sum of square errors between model predictions and data in order to choose a best model, i.e. most suitable objective function and set of limiting constraints.

Mammalian cells must accomplish a much more diverse set of functionalities during their time evolution as compared to bacteria that seems to be targeted only to grow faster. For example, during their growth phase, mammalian cells make very inefficient use of resources in terms of ATP production since most of the glucose is initially converted to lactate instead of being consumed toward ATP production in the TCA cycle. This preferential consumption of glucose towards lactate is referred to as the *Warburg effect*. On the other hand fast nutrient depletion occurring during the growth phase in batch operation may induce apoptosis which results in high cellular death rate. It is known that apoptosis may

be enhanced through high TCA cycle activity occurring in the mitochondria (Fleury, Mignotte, & Vayssiere, 2002).

Following the above, to assess which objective function and constraints best fit the experiments and successfully capture some of the phenomena described above, e.g. *Warburg effect*, we identified DFBA models with the 3-step method shown in Section 2 for 3 candidate objective functions as follows:

O1-Maximization of growth rate at each time.

O2-Minimization of NADH production in the cytosol plus minimization of NAD(P)H consumption in mitochondria at each time interval.

O3-Combination of O1 and O2

O1 is the objective function typically used for bacteria for which it is assumed that natural evolution had driven the microorganisms to the ultimate goal of growing as much as possible despite environmental conditions. One can argue that O1 is also a rational choice for mammalian cells during the exponential phase of growth where cells exhibiting mostly a growth pattern with a minimal amount of dead cells.

Objective O2 arises from the observation mentioned above that self-programmed death apoptosis is largely correlated to the levels of cytochrome-C that is a key protein in the respiratory electron transport chain responsible for the synthesis of ATP. Then, since the electron transport reactions are coupled to the TCA cycle through NADH/FADH, the minimization of NADH production is tantamount to the cells striving to minimize the occurrence of apoptosis (programmed death). Although our stoichiometric matrix did not include a specific balance for NADH/FADH, its production could be inferred from the sum of the fluxes of the reactions producing NADH in cytosol

and reaction consuming NAD(P)H in mitochondria = v(R5) + v(R8) +v(R11) v (R13) + v(R34) + v(R40)+ v(R43).

Where the corresponding reaction numbers are given in the Appendix C.

Objective O3 combines both O1 and O2 to describe the possibility that cell attempt to simultaneously maximize its growth while reducing the occurrence of apoptosis.

To test which objective function is most suitable and to find the corresponding limiting constraints, we apply the 3-step identification procedure (Section 2) to the data thus obtaining 3 different models each corresponding to the 3 objective functions considered in the study. Then, the models were compared in their ability to fit the data for the duration of the batch runs based on the Sum of Squared Errors (SSE) between the models' predictions and the data.

Following the 3-step method outlined in section 2, step 1 was implemented to find set based constraints from data for each of the 3 objective functions. Secondly, using these set based constraints, the limiting constraints were found from the values of the Lagrange multipliers (see Section 2, Step-2) that for all the 3 models the active constraints corresponded to exactly the same 5 metabolites: alanine, glutamate, lactate, ammonia, and glycine.

It should be noticed that although the same metabolites must be constrained in the model, the observed dependency of the consumption/production rates as a function of the current concentration (found as per Step 3 in Section 2) of the corresponding metabolite were different among the 3 models.

Following the identification of the active constraints and in order to formulate a predictive model it is required to calibrate Monod or another type of kinetic expressions to the calculated values of consumption/production rates as a function of a corresponding metabolite concentration.

Different types of kinetic models, e.g. Monod, Hill and etc. were tried in order to fit the consumption production rates as a function of the metabolite concentrations. Although the model predictions using the identified kinetic models for the 5 constraints were acceptable, some of the predicted errors were found to be slightly higher than the measurement errors expected for the metabolites. To further improve the fitting we decided to use for the 5 identified constraints, look up tables that will provide the specific consumption/production rates as a function of concentration values. Since one of the objectives of formulating a predictive model was to optimize the process with respect to the feeding of nutrient concentrations we develop look-up tables as a function of the main nutrient for this process which is glucose. It should be noticed that two of the identified constraints are by-products: lactate and ammonia. Then, the proposed dependency of the specific production rates of these two by-products as a function of glucose can be justified metabolically by the fact that in reality, some intermediate metabolite of glycolysis is actually limiting but it is not measured and therefore, it cannot be used explicitly to formulate a predictive model.

In principle, the 5 constraints would suffice for defining a dynamic metabolic flux model as given in (8-1). However, it was found that with these constraints only the optimization problem in (8-1) resulted in multiple solutions thus necessitating the addition of coarse constraints to limit the solution space of the model. We found that these additional coarse constraints could be given as a function of constant upper bound values on the rate of consumption/production of all metabolites.

It should be noticed that the need for such additional coarse constraints to limit the solution space of dynamic models has been recognized and reported before by other researchers (Zakrzewski et al., 2012).

Using the kinetic constraints and the coarse constraints, 3 different models were obtained.

First, it was observed that all the 3 models based on the 3 objective functions and their corresponding constraints could be fitted reasonably well to the data. Table 8-1 shows the SSE for each of the models corresponding to each objective function.

Table 8-1: Sum of Squared Errors (SSE) for each model (number of parameters are all the same)

Objective Function	**SSE**
O1-max (growth)	1.49
O2-min (NADH in the cytosol)-min NAD(P)H consumption in mitochondria	8.74
O3-combination O1 and O2	1.3

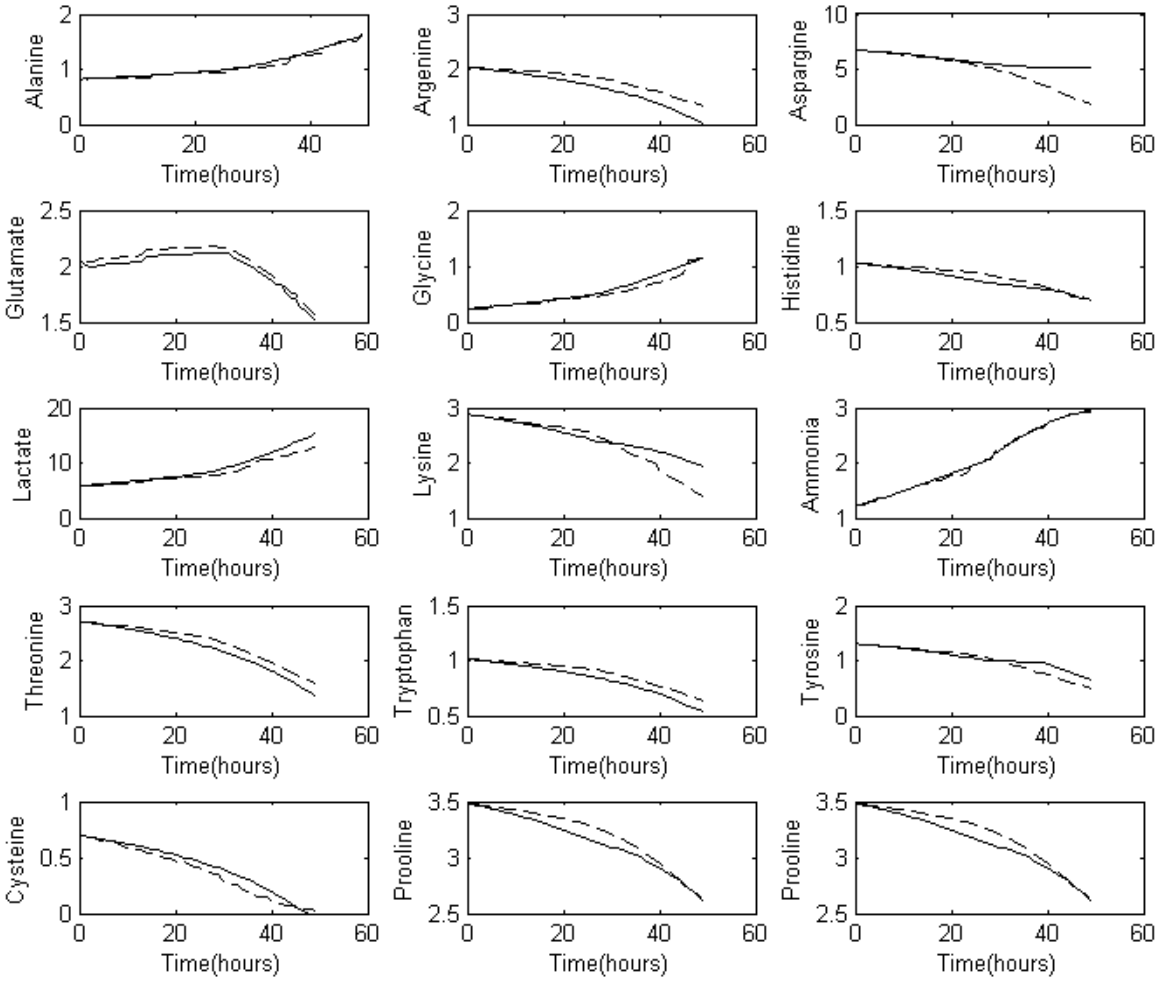
It is evident from Table 8-1 that O2 results in the worst fit. On the other hand, O1 and O3 provide a similar level of fit but O3 results in an improvement in the SSE as compared to O1.

Figure 8-2 shows a comparison between the data and the predictions obtained with objective O3, i.e. maximization of (growth-production of NADH in cytoplasm-NAD(P)H consumption in mitochondria). As shown in Figure 8-2 the agreement is very good and the maximal observed errors between the data and the model predictions were of the order of up to 10% of the maximal variation in each metabolite. These maximal errors are of the order of the expected measurement errors in HPLC (amino acids) or the Bioprofile (Glucose, Lactate).

It is of interest to inspect why the model based on objective function O3 resulted in the best fit to the data. To that purpose, it should be recalled that one of the salient characteristics of CHO cells used for cell culturing is their high production of lactate during the exponential phase of growth. This phenomena, referred to as the Warburg effect is particularly counterintuitive since the higher production

of lactate is occurring at the expense of reduced flow of glucose into the TCA cycle where the latter is very effective in the production of ATP/mol of glucose.

Then, to explain the higher suitability of objective function O3 as compared to the other objective functions we run the 3 models corresponding to the 3 objective functions with constraints on all metabolites but without imposing any constraint on lactate levels and we subsequently tested the amount of lactate produced by each of the models. Figure 8-3 shows the lactate as a function of time for the duration of the batch as predicted by the 3 models and as measured by the BioProfile. It is evident that objective 3 correctly predicts the large production of lactate which is a key feature in CHO cell cultures.


Alanine
Time(hours)
Argenine
Time(hours)
Aspargine
Time(hours)
Glutamate
Time(hours)
Glycine
Time(hours)
Histidine
Time(hours)
Lactate
Time(hours)
Lysine
Time(hours)
Ammonia
Time(hours)
Threonine
Time(hours)
Tryptophan
Time(hours)
Tyrosine
Time(hours)
Cysteine
Time(hours)
Prooline
Time(hours)
Prooline
Time(hours)

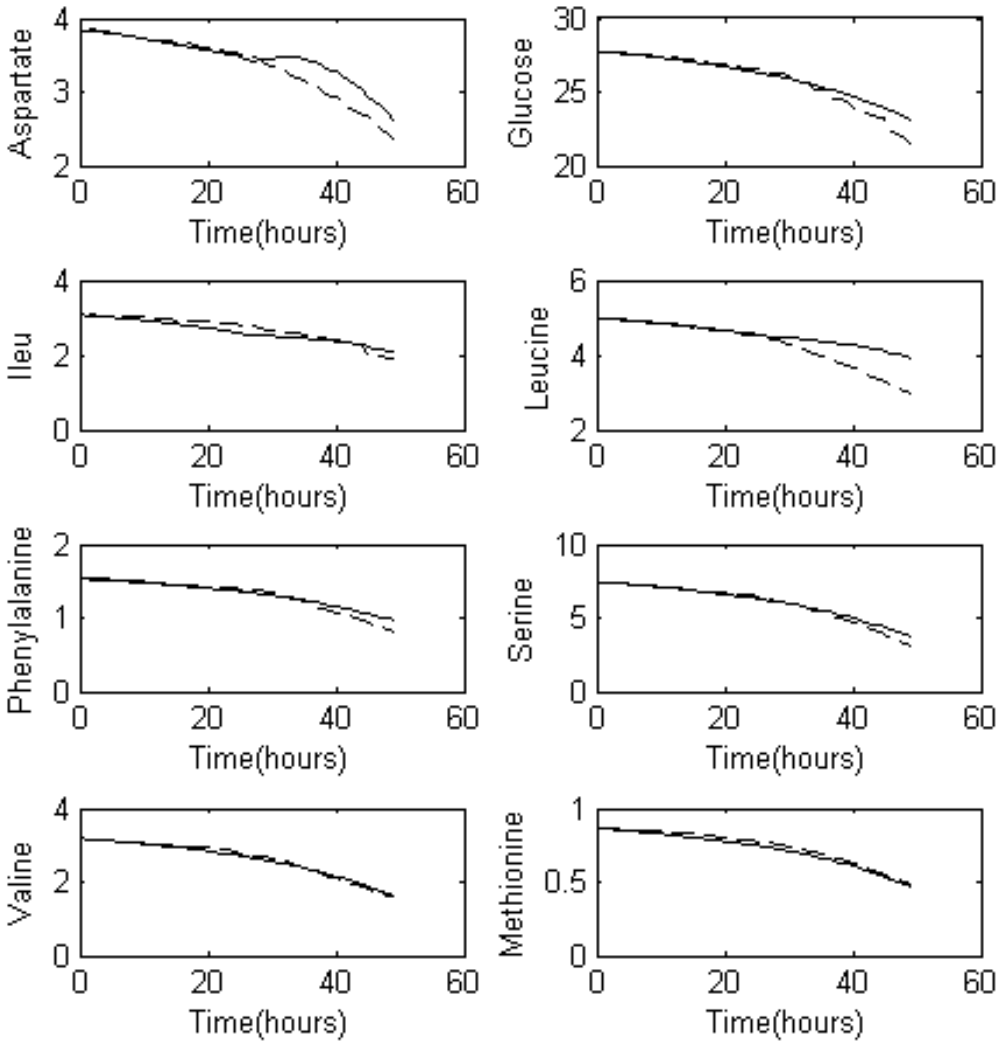


Figure 8-2 Dotted lines are the interpolation of experimental data of CHO cell culture number 1 and solid lines are the DFBA model predictions.

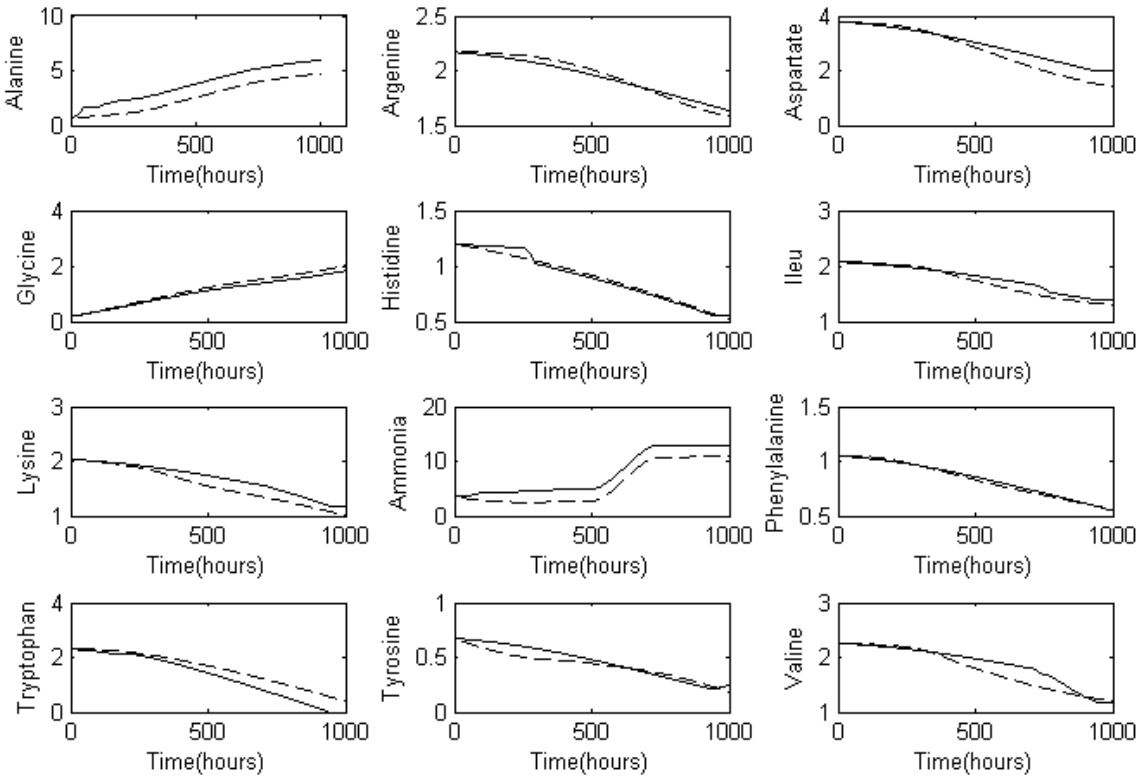


Figure 8-3-a Dotted lines are the experimental data of CHO cell culture number 2 and solid line are the model predictions

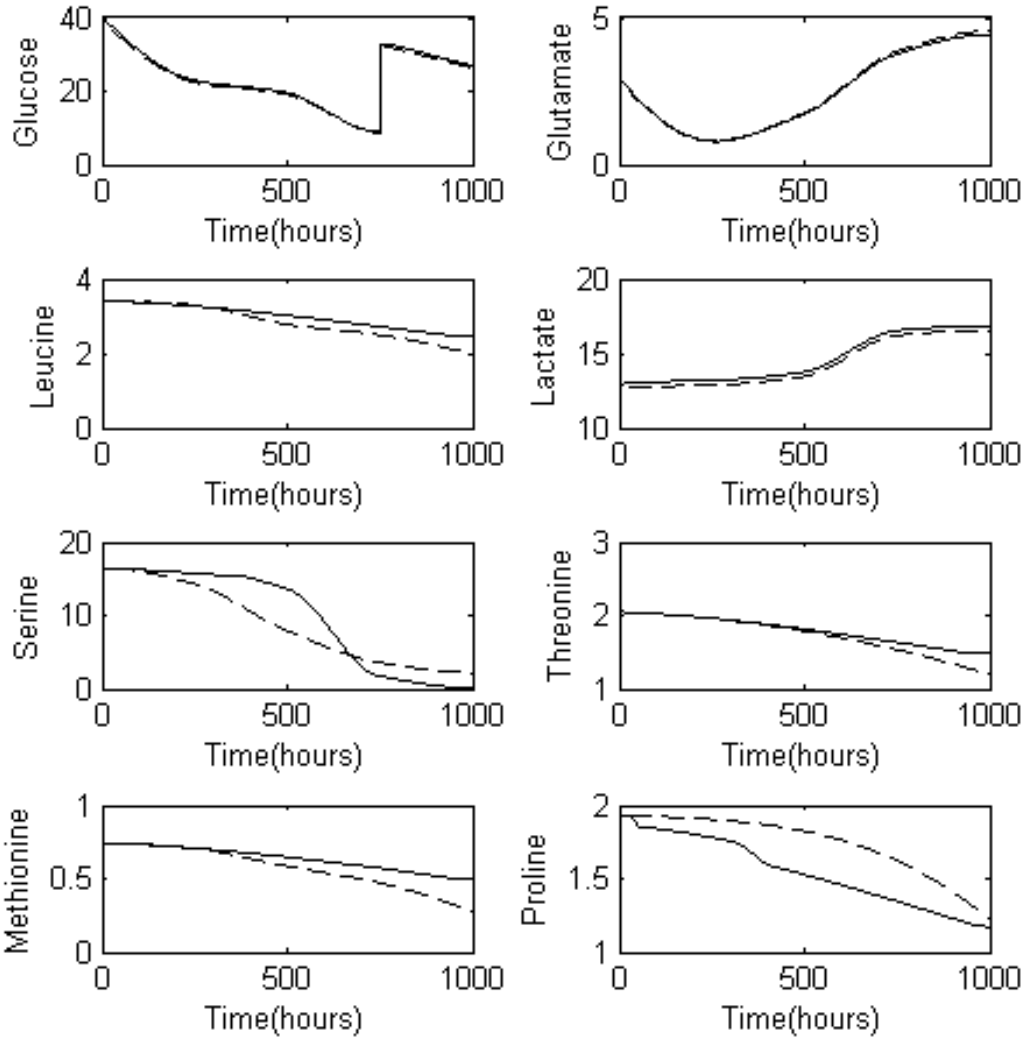


Figure 8-3-b Dotted lines are the interpolation of experimental data of CHO cell culture number 2 and solid line are the model predictions

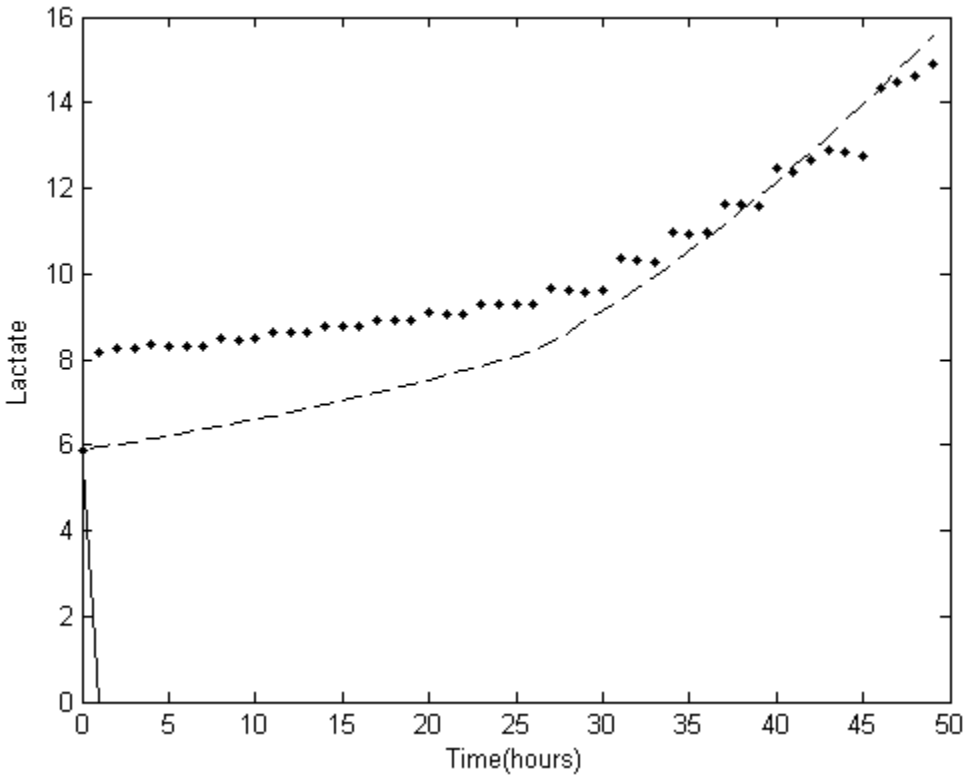


Figure 8-4 Prediction for lactate with two different objective functions: Dot line in the presence of objective function O3 and solid line when the objective function is considered O1, Dashed line is the experimental data

It should be noticed that the models obtained in this work are definitely advantageous in terms of the lower number of metabolites that have to be modelled in terms of kinetic behaviour as compared to other mammalian cell models. However, at this point not necessarily the required number of parameters is smaller as compared to other models in the literature since this number depends on the number of rows of the lookup tables formulated to describe the dependency of specific consumption/production

rates of 5 metabolites as a function of glucose concentrations. At this point, we have not optimized these tables so as to minimize the number of entries and this is left for future study.

Regarding the choice of objective function, although objective O3 is able to correctly predict the lactate production when constraints on all other metabolites are used and it also results in the lowest SSE, it may not be the best possible objective to be used for a CHO cell system. For instance, it should be noticed that objectives O1 and O2 were combined with an equal weight into objective O3. However, it is possible that an alternative objective that combines O1 and O2 with different relative weighting may result in a better model than the one obtained for objective O3. The search for such optimal relative weighting that could reflect ATP or NADH equivalents of growth rate or apoptosis is left for future investigation.

8.5 Conclusion

A dynamic metabolic flux model for mammalian cells was identified directly from data using a 3-step identification approach based on set based constraints. The model is able to correctly predict the measured metabolites while only 5 metabolites need to be described by kinetic expressions and used as constraints within the optimization problem describing the model. Three objective functions were investigated that represented the maximization of growth, the minimization of apoptosis and the combination of both objectives. It was found that the combination provides the best fitting to the data. A plausibility argument for the superiority of the objective function that simultaneously maximizes growth and minimizes apoptosis was that this constraint results in the superior prediction of the Warburg effect where lactate is produced in large amounts at the expense of ATP production by the electron transport chain.

Chapter 9

Conclusions and recommendations

The optimal operation of bioprocesses requires mathematical models to develop optimal operating recipes. Most mathematical models proposed in the literature that describe the time profiles of nutrients and by-products during cell cultures require a large number of parameters for fitting experimental data. Recognizing that the over-parameterization of these models make them very sensitive to noise and impact their prediction ability we investigated in this work the use of a particular type of models, dynamic metabolic flux models, that have been often reported as more compact in terms of the number of parameters that need to be identified. Thus, based on preliminary studies in the literature, our working hypothesis was that dynamic metabolic flux models would be able to describe the behavior in cell culture of different microorganisms with fewer parameters. To objectively assess this hypothesis we developed mathematical approaches that will be able to describe experiments with a minimal number of parameters.

The final goal of the thesis was to apply the proposed mathematical approaches to mammalian cell cultures. The focus on CHO was motivated by an earlier study that I conducted on cell death in CHO cell cultures as a continuation of the work of another member of our research group (S. Naderi). In a preliminary study of my thesis, I conducted experiments on CHO cell cultures at different pH for I identified carbonic acid as a limiting factor for growth and that intermittent media replacement was instrumental to extend the viability of the culture. Following these initial experiments and based on the fact that the models of Naderi (Naderi et al., 2011b) were highly parameterized we sought for more parsimonious models for CHO cell cultures. Also, the finding that intermittent media replacements were notably advantageous for operating CHO cell cultures motivated me to look further into

continuous perfusion as a logical continuation of intermittent perfusion operation. Following these goals, we were also able to establish a collaboration with Sigma-Aldrich in St Louis, Missouri that provided a high producing CHO cell line, growth media for culturing and maintained close communication with us on the directions of the project.

This thesis includes both theoretical and experimental part. In terms of theory the development of dynamic flux balance models requires finding a meaningful biological objective function to be maximized/minimized and active constraints. To address these challenges two main mathematical algorithms were suggested in this thesis which were published in separate papers. The algorithms were used initially to describe bacterial cultures, *E.coli* and *B. pertussis* but then were also applied CHO cell cultures. Following the recognition of the potential of perfusion for increasing cell growth, in the experimental section, I developed a bioreactor perfusion system for culturing CHO cells in a higher cell density.

The thesis includes 9 different Chapters involving the different theoretical and experimental aspects of the work. Chapter 2 reviewed the relevant literature and Chapter 3 presents the relevant experimental and theoretical methods and materials. In the first part of Chapter 4 I reviewed the perfusion mode of operation, cell retention systems and challenges that I faced when developing the perfusion bioreactor system with an ATF filtration system. Here I have clearly identified challenges that have not been properly addressed in the literature such as methods to provide enough DO and for stripping the carbon dioxide while controlling the foam formation. Following the identification of these challenges, I have formulated a series of practical solutions that help me to develop a high cell density CHO cell culture in a 3 liter bioreactor. I performed experiments with this perfusion system where a maximum cell density of around 50 million cells per mL culture was achieved and I also showed the ability of the system to control the level of potentially toxic by-products such as lactate and ammonia. Although

companies are currently conducting perfusion operations of CHO cultures the particular problems associated with these operations have not been thoroughly reported. Hence, a paper summarizing our findings regarding the perfusion practical challenges and solution is currently in preparation.

In chapter 5, two algorithms for identifying active constraints for developing DFBA models were presented and compared. These algorithms are based on two different concepts: 1- sensitivity analysis and 2- identification of active constraints using Lagrange multipliers. These algorithms were initially applied for building DFBA models for two case studies of *E.coli* and *B. pertussis* in batch and fed-batch mode of operation. The *E.coli* case study used *in silico* (simulated) data and was purposely developed to check the ability of the algorithm to converge to a dynamic metabolic flux model that was a priori known. A simplified metabolic network of *E.coli* containing four reactions was used for this case study. The second case study was using *B.pertussis* producing four antigens required for manufacturing the whooping cough vaccine. Data for different batch and fed batch cultures were provided by a company. Both of these proposed methods were capable of identifying the active kinetic constraints for *E.coli* example but the one based on the sensitivity analysis was found to be not efficient for the *B.pertussis* case that had a larger metabolic network. The disadvantages of the sensitivity analysis method were as follows: i- the method involved the solution of a non-convex bi-level optimization problem that was very difficult to solve and require numerous initial guesses to converge to a best solution and ii- the sensitivity had to be tested around specific nominal values of the parameters that were not known a priori. Hence the method require an iterative procedure where initial nominal parameter values must be identified, then the sensitivity is tested around these values and constraints are identified and then with these constraints a new set of nominal values have to be calculated and so on.

While sensitivity analysis has been used before in the context of dynamic flux models, the second method that we proposed is novel and it is based on the idea of set-based constraints. The representation

of the data by this type of constraint greatly facilitates the identification procedure since it leads to a one level optimization problem. Moreover, we show that after identifying the constraints based on the values of Lagrange multipliers associated to each constraint, the fitting of a kinetic expression to describe the active constraint as a function of metabolite concentration can be done separately for each metabolite. This greatly simplifies the identification of the kinetic expressions necessary to formulate a predictive model. Although this algorithm based on Lagrange multiplier was shown to be very effective it was also found that it resulted in a multiplicity of solutions for the *B. Pertussis* case due to the large dimension of the metabolic network. To address this multiplicity time-independent upper bound constraints on the metabolites' consumption or production rates were required. An additional challenge for this algorithm is that the activity/inactivity of the constraint depends on a threshold value of the Lagrange multiplier. Hence the activity of the constraint had to be decided by testing the sensitivity of the Lagrange multipliers to different levels of noise. This chapter has been published in the Biotechnology Progress Journal.

Thus, two main challenges encountered with the algorithm presented in Chapter 5: 1- the algorithm in Chapter 5 find the constraints for an assumed biological objective that is maximized by the DFBA and 2- the algorithm in 5 required correct assessment of the effect of noise on the Lagrange multipliers and this was found to be a laborious task. In Chapter 6 a new set-based algorithm was suggested to address these problems for simultaneously finding the biological meaningful objective function together with the active constraints through one single optimization problem and noise was directly accounted for in the optimization. This method was advantageous compared to prior as it was capable of finding the active constraints, objective function, and the upper bound constraints simultaneously using a bilinear optimization formulation. This method was also verified using two case studies of *E.coli* and *B. pertussis*. After corroborating the accuracy of the algorithm with the simulated E.coli case study we

applied the algorithm to data from a *B. pertussis* case study (experimental data was provided by Sanofi Pasteur). The maximization of the biomass was identified as the best objective function and the uptake rate of phenylalanine was discovered as the single active constraint. Thus the algorithm was able to find a highly compact model with one single kinetic expression. This algorithm is particularly advantageous as compared to the algorithm in Chapter 5 since it formally addresses the level of noise and its effect on the measured consumption/production rates in the optimization problem. However, the problem of multiplicity of solutions and the need to limit the solution space of the optimization problem by adding time –independent upper bounds on the metabolites' consumption and production rates still remains as a challenge to be addressed in future research. The results in Chapter 6 were reported in a paper published in the journal Bioprocesses and Biosystems Engineering.

Chapter 7 reports my above mentioned contribution as a second author with a study dealing with cell damage due to the accumulation of carbon dioxide in cell culture flasks. By performing experiments at three different pH, I was able to identify that a reduction in pH significantly reduces the cell growth. Another experiment in shaker flask at two different volumes of 250 and 500mL was performed to study the effect of gaseous carbon dioxide. This experiment was performed in a semi-perfusion mode by substituting a portion of spent medium with fresh media at different times. The initial pH in the two spinner flask were the same but a higher pH drop was observed in the 500 mL flask as compared to the 250 mL flask that hinted at an accumulation of an acidic compound in the spinners affecting growth. Based on pH and pK calculations we were able to identify carbonic acid with the pka=6.4 as the most plausible compound affecting cell growth.

In chapter 8, the algorithms developed in chapter 6 for finding active constraints and biological objective function were used to develop a DFBA model for CHO cells in a batch experiment. At this point, we have only been able to fit the model for the exponential phase of growth. A key finding in

this chapter is that the biological function for mammalian cells that can be used to best describe the data is not as simple as for bacterial systems where only growth rate was maximized. In the case of mammalian cells, the simultaneous maximization of growth rate and minimization of NADH production in the cytoplasm and minimization of NADPH consumption in the mitochondria resulted in a better fit of the data. We have explained this finding by the fact that the production of NADH in the TCA cycle is correlated to apoptosis and thus the mathematical minimization of NADH production it describes the ability of the cell to optimize its chances for survival by minimizing apoptosis. Constraints on 5 metabolites were necessary to describe the data. Following fitting of the consumption/production rates of the limiting metabolites with standard kinetic expressions (e.g. Michaelis-Menten, Hill) it was found that the errors between data and model predictions were acceptable but not as accurate as expected considering the measurement error. To improve the fitting look-up tables were used that describe the consumption/production rates corresponding to the active constraints as a function of interpolated values among few glucose concentrations.

9.1 Future work

9.1.1 Genome-scale DFBA models

Following the publication of the genome sequence of CHO cells, the current models could be extended to include gene regulatory mechanisms. Following the multiplicity of solutions observed in some of the studies it would also be important to include constraints other than kinetic constraints considered in the current work e.g. thermodynamic constraints, protein expression constraints, gene expression related constraints etc.

9.1.2 In-silico Metabolic Engineering

One important application of the developed DFBA model would be to develop *in silico* metabolic engineering approaches for designing better CHO cell lines producing more desired proteins and fewer by-products such as lactate. For example, DFBA models could be used for investigating the gene-deletion targets in CHO cells based on identified constraints. By combining the DFBA model with the method OptKnock(Burgard, Pharkya, & Maranas, 2003) that has been proposed by for gene-deletion in bacterial strains one could dynamically optimize cell line productivity.

9.1.3 Cell culture design and process optimization

Finding the best combinations of nutrients in the cell culture media has always been an important goal in the biopharmaceutical industry. DFBA models can provide insights in terms of limiting external metabolites, intracellular metabolites and metabolites that are not easy to measure such as vitamins and trace elements. The model will be useful to ensure that sufficient amount of these elements are available in the media. Nowadays the most common technique for media optimization is black-box modelling based on techniques such as multivariate statistical tools. By combining the information from DFBA

model with multivariate data analysis based models into hybrid models (empirical+mechanistic) the growth medium could be optimized.

9.1.4 Optimizing the protein quality and Quality-by-Design approaches

Protein quality and lot-to-lot variability are important issues in the biopharmaceutical industry. One of the important protein quality attributes is glycosylation. Models have been developed that link the concentration of extracellular metabolites to glycosylation models but due to a higher number of parameters to be estimated developing such models has been very difficult. Using DFBA models with less parameters compared should be helpful in developing models that couple the cell metabolism with glycosylation processes.

References

Aghamohseni, H., Ohadi, K., Spearman, M., Krahn, N., Moo-Young, M., Scharer, J. M., . . . Budman, H. M. (2014). Effects of nutrient levels and average culture pH on the glycosylation pattern of camelid-humanized monoclonal antibody. *J Biotechnol, 186*, 98-109. doi: 10.1016/j.jbiotec.2014.05.024

Ahn, W. S., & Antoniewicz, M. R. (2012a). Towards dynamic metabolic flux analysis in CHO cell cultures. *Biotechnol J, 7*(1), 61-74. doi: 10.1002/biot.201100052

Alberts, B. (2008). *Molecular biology of the cell* (5th ed.). New York: Garland Science.

Altamirano, C., Illanes, A., Becerra, S., Cairo, J. J., & Godia, F. (2006). Considerations on the lactate consumption by CHO cells in the presence of galactose. *J Biotechnol, 125*(4), 547-556. doi: 10.1016/j.jbiotec.2006.03.023

Altamirano, C., Paredes, C., Illanes, A., Cairo, J. J., & Godia, F. (2004). Strategies for fed-batch cultivation of t-PA producing CHO cells: substitution of glucose and glutamine and rational design of culture medium. *J Biotechnol, 110*(2), 171-179. doi: 10.1016/j.jbiotec.2004.02.004

Andersen, D. C., & Goochee, C. F. (1995). The effect of ammonia on the O-linked glycosylation of granulocyte colony-stimulating factor produced by chinese hamster ovary cells. *Biotechnol Bioeng, 47*(1), 96-105. doi: 10.1002/bit.260470112

Antoniewicz, M. (2013). Using Multiple Tracers for 13C Metabolic Flux Analysis. In H. S. Alper (Ed.), *Syst Metab Eng,* (Vol. 985, pp. 353-365): Humana Press.

Barcia-Vieitez, R., & Ramos-Martinez, J. I. (2014). The regulation of the oxidative phase of the pentose phosphate pathway: new answers to old problems. *IUBMB Life, 66*(11), 775-779. doi: 10.1002/iub.1329

Baroukh, C., Muñoz-Tamayo, R., Steyer, J.-P., & Bernard, O. (2014). DRUM: A New Framework for Metabolic Modeling under Non-Balanced Growth. Application to the Carbon Metabolism of Unicellular Microalgae. *PLoS One, 9*(8), e104499. doi: 10.1371/journal.pone.0104499

Bebbington, C. R., Renner, G., Thomson, S., King, D., Abrams, D., & Yarranton, G. T. (1992). High-level expression of a recombinant antibody from myeloma cells using a glutamine synthetase gene as an amplifiable selectable marker. *Biotechnology (N Y), 10*(2), 169-175.

Berg, J. M., Tymoczko, J. L., Stryer, L., & Stryer, L. (2002). *Biochemistry*. New York: W.H. Freeman.

Bertsimas, D., & Tsitsiklis, J. (1997). *Introduction to Linear Optimization*: Athena Scientific.

Bonham-Carter, J. S., J. (2011, October). A Brief History of Perfusion Biomanufacturing. *BioProc Int*.

Borchers, S., Freund, S., Rath, A., Streif, S., Reichl, U., & Findeisen, R. (2013). Identification of Growth Phases and Influencing Factors in Cultivations with AGE1.HN Cells Using Set-Based Methods. *PLoS One, 8*(8). doi: 10.1371/journal.pone.0068124

Bordbar, A., Monk, J. M., King, Z. A., & Palsson, B. O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet, 15*(2), 107-120. doi: 10.1038/nrg3643

Budman, H., Patel, N., Tamer, M., & Al-Gherwi, W. (2013). A dynamic metabolic flux balance based model of fed-batch fermentation of Bordetella pertussis. *Biotechnol Prog, 29*(2), 520-531. doi: 10.1002/btpr.1675

Burgard, A. P., & Maranas, C. D. (2003). Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng, 82*(6), 670-677. doi: 10.1002/bit.10617

Burgard, A. P., Pharkya, P., & Maranas, C. D. (2003). OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng, 84*(6), 647-657. doi: 10.1002/bit.10803

Burky, J. E., Wesson, M. C., Young, A., Farnsworth, S., Dionne, B., Zhu, Y., . . . Sauer, P. W. (2007). Protein-free fed-batch culture of non-GS NS0 cell lines for production of recombinant antibodies. *Biotechnol Bioeng, 96*(2), 281-293. doi: 10.1002/bit.21060

Butler, M. (2006). Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by Mammalian cell systems. *Cytotechnology, 50*(1-3), 57-76. doi: 10.1007/s10616-005-4537-x

Butler, M., & Spearman, M. (2014). The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr Opin Biotechnol, 30*, 107-112. doi: 10.1016/j.copbio.2014.06.010

Carrillo-Cocom, L. M., Genel-Rey, T., Araiz-Hernandez, D., Lopez-Pacheco, F., Lopez-Meza, J., Rocha-Pizana, M. R., . . . Alvarez, M. M. (2015). Amino acid consumption in naive and recombinant CHO cell cultures: producers of a monoclonal antibody. *Cytotechnology, 67*(5), 809-820. doi: 10.1007/s10616-014-9720-5

Chapman, S. P., Paget, C. M., Johnson, G. N., & Schwartz, J. M. (2016). Corrigendum: Flux balance analysis reveals acetate metabolism modulates cyclic electron flow and alternative glycolytic pathways in Chlamydomonas reinhardtii (vol 6, 474, 2015). *Front Plant Sci, 7*. doi: 10.3389/Fpls.2016.00362

Chee Furng Wong, D., Tin Kam Wong, K., Tang Goh, L., Kiat Heng, C., & Gek Sim Yap, M. (2005). Impact of dynamic online fed-batch strategies on metabolism, productivity and N-glycosylation quality in CHO cell cultures. *Biotechnol Bioeng, 89*(2), 164-177. doi: 10.1002/bit.20317

Chen, K. Q., Liu, Q., Xie, L. Z., Sharp, P. A., & Wang, D. I. C. (2001). Engineering of a mammalian cell line for reduction of lactate formation and high monoclonal antibody production. *Biotechnol Bioeng, 72*(1), 55-61. doi: 10.1002/1097-0290(20010105)72:1

Chotteau, V. (2015). Perfusion Processes. In M. Al-Rubeai (Ed.), *Animal Cell Culture* (Vol. 9, pp. 407-443): Springer International Publishing.

Clincke, M. F., Molleryd, C., Zhang, Y., Lindskog, E., Walsh, K., & Chotteau, V. (2013). Very high density of CHO cells in perfusion by ATF or TFF in WAVE bioreactor. Part I. Effect of the cell density on the process. *Biotechnol Prog, 29*(3), 754-767. doi: 10.1002/btpr.1704

Cohen, S. A. (2000). Amino Acid Analysis Using Precolumn Derivatization with 6-Aminoquinolyl-N-Hydroxysuccinimidyl Carbamate. In C. Cooper, N. Packer & K. Williams (Eds.), *Amino Acid Analysis Protocols* (pp. 39-47). Totowa, NJ: Humana Press.

Craven, S., Whelan, J., & Glennon, B. (2014). Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. *J Process Contr, 24*(4), 344-357. doi: http://dx.doi.org/10.1016/j.jprocont.2014.02.007

Dikicioglu, D., Kirdar, B., & Oliver, S. G. (2015). Biomass composition: the "elephant in the room" of metabolic modelling. *Metabolomics, 11*(6), 1690-1701. doi: 10.1007/s11306-015-0819-2

Duarte, T. M., Carinhas, N., Barreiro, L. C., Carrondo, M. J. T., Alves, P. M., & Teixeira, A. P. (2014). Metabolic Responses of CHO Cells to Limitation of Key Amino Acids. *Biotechnol Bioeng, 111*(10), 2095-2106. doi: 10.1002/bit.25266

Fan, L., Kadura, I., Krebs, L. E., Hatfield, C. C., Shaw, M. M., & Frye, C. C. (2012). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnol Bioeng, 109*(4), 1007-1015. doi: 10.1002/bit.24365

Feist, A. M., & Palsson, B. O. (2010a). The biomass objective function. *Curr Opin Microbiol, 13*(3), 344-349. doi: 10.1016/j.mib.2010.03.003

Fleury, C., Mignotte, B., & Vayssiere, J. L. (2002). Mitochondrial reactive oxygen species in cell death signaling. *Biochimie, 84*(2-3), 131-141. doi: Pii S0300-9084(02)01369-X

Foguet, C., Marin, S., Selivanov, V. A., Fanchon, E., Lee, W. N. P., Guinovart, J. J., . . . Cascante, M. (2016). HepatoDyn: A Dynamic Model of Hepatocyte Metabolism That Integrates C-13 Isotopomer Data. *PLoS Comput Biol, 12*(4). doi: 10.1371/journal.pcbi.1004899

Galleguillos, S. N., Ruckerbauer, D., Gerstl, M. P., Borth, N., Hanscho, M., & Zanghellini, J. (2017). What can mathematical modelling say about CHO metabolism and protein glycosylation? *Comput Struct Biotechnol J, 15*, 212-221.

Ghorbaniaghdam, A., Henry, O., & Jolicoeur, M. (2013). A kinetic-metabolic model based on cell energetic state: study of CHO cell behavior under Na-butyrate stimulation. *Bioprocess Biosyst Eng, 36*(4), 469-487. doi: 10.1007/s00449-012-0804-3

Gianchandani, E. P., Oberhardt, M. A., Burgard, A. P., Maranas, C. D., & Papin, J. A. (2008). Predicting biological system objectives de novo from internal state measurements. *Bmc Bioinformatics, 9*. doi: Artn 43Doi 10.1186/1471-2105-9-43

Gray, D. R., Chen, S., Howarth, W., Inlow, D., & Maiorella, B. L. (1996). CO(2) in large-scale and high-density CHO cell perfusion culture. *Cytotechnology, 22*(1-3), 65-78. doi: 10.1007/BF00353925

Han, Y. K., Kim, Y. G., Kim, J. Y., & Lee, G. M. (2010). Hyperosmotic stress induces autophagy and apoptosis in recombinant Chinese hamster ovary cell culture. *Biotechnol Bioeng, 105*(6), 1187-1192. doi: 10.1002/bit.22643

Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., . . . Lewis, N. E. (2016). A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Syst, 3*(5), 434-443 e438. doi: 10.1016/j.cels.2016.10.020

Hu, W.-S., & Zhou, W. (2012). *Cell culture bioprocess engineering*. [Minnesota]: Wei-Shou Hu.

Hwang, S. O., & Lee, G. M. (2008). Nutrient deprivation induces autophagy as well as apoptosis in Chinese hamster ovary cell culture. *Biotechnol Bioeng, 99*(3), 678-685. doi: 10.1002/bit.21589

Irani, N., Beccaria, A. J., & Wagner, R. (2002). Expression of recombinant cytoplasmic yeast pyruvate carboxylase for the improvement of the production of human erythropoietin by recombinant BHK-21 cells. *J Biotechnol, 93*(3), 269-282.

Irani, N., Wirth, M., van Den Heuvel, J., & Wagner, R. (1999). Improvement of the primary metabolism of cell cultures by introducing a new cytoplasmic pyruvate carboxylase reaction. *Biotechnol Bioeng, 66*(4), 238-246.

Jacquemart, R., Vandersluis, M., Zhao, M., Sukhija, K., Sidhu, N., & Stout, J. (2016). A Single-use Strategy to Enable Manufacturing of Affordable Biologics. *Comput Struct Biotechnol J, 14*, 309-318. doi: 10.1016/j.csbj.2016.06.007

Jaqaman, K., & Danuser, G. (2006). Linking data to models: data regression. *Nat Rev Mol Cell Bio, 7*(11), 813-819. doi: 10.1038/nrm2030

Jayapal, K. R., Wlaschin, K. F., Hu, W. S., & Yap, M. G. S. (2007). Recombinant protein therapeutics from CHO cells - 20 years and counting. *Chem Eng Prog, 103*(10), 40-47.

Kelly, W., Scully, J., Zhang, D., Feng, G., Lavengood, M., Condon, J., . . . Bhatia, R. (2014). Understanding and modeling alternating tangential flow filtration for perfusion cell culture. *Biotechnol Prog*. doi: 10.1002/btpr.1953

Kim, J. Y., Kim, Y. G., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl Microbiol Biot, 93*(3), 917-930. doi: 10.1007/s00253-011-3758-5

Knorr, A. L., Jain, R., & Srivastava, R. (2007). Bayesian-based selection of metabolic objective functions. *Bioinformatics, 23*(3), 351-357. doi: 10.1093/bioinformatics/btl619

Konstantinov, K., Goudar, C., Ng, M., Meneses, R., Thrift, J., Chuppa, S., . . . Naveh, D. (2006). The "Push-to-Low" Approach for Optimization of High-Density Perfusion Cultures of Animal Cells. In W.-S. Hu (Ed.), *Cell Culture Eng* (Vol. 101, pp. 75-98): Springer Berlin Heidelberg.

Lake-Ee, Q., & and Lars Keld, N. (2009). Metabolic Engineering of Mammalian Cells *The Metabolic Pathway Engineering Handbook* (pp. 26-21-26-22): CRC Press.

Laken, H. A., & Leonard, M. W. (2001). Understanding and modulating apoptosis in industrial cell culture. *Curr Opin Biotechnol, 12*(2), 175-179.

Levine, B. L., Miskin, J., Wonnacott, K., & Keir, C. (2017). Global Manufacturing of CAR T Cell Therapy. *Mol Ther Methods Clin Dev, 4*, 92-101. doi: 10.1016/j.omtm.2016.12.006

Link, T., Backstrom, M., Graham, R., Essers, R., Zorner, K., Gatgens, J., . . . Noll, T. (2004). Bioprocess development for the production of a recombinant MUC1 fusion protein expressed by CHO-K1 cells in protein-free medium. *J Biotechnol, 110*(1), 51-62. doi: 10.1016/j.jbiotec.2003.12.008

Lipscomb, M. L., Palomares, L. A., Hernandez, V., Ramirez, O. T., & Kompala, D. S. (2005). Effect of production method and gene amplification on the glycosylation pattern of a secreted reporter protein in CHO cells. *Biotechnol Prog, 21*(1), 40-49. doi: Doi 10.1021/Bp049761m

Liu, J. K. (2014). The history of monoclonal antibody development - Progress, remaining challenges and future innovations. *Ann Med Surg (Lond), 3*(4), 113-116. doi: 10.1016/j.amsu.2014.09.001

Llaneras, F., & Picó, J. (2008). Stoichiometric modelling of cell metabolism. *J Biosci Bioeng, 105*(1), 1-11. doi: http://dx.doi.org/10.1263/jbb.105.1

Llaneras, F., Sala, A., & Picó, J. (2012). Dynamic estimations of metabolic fluxes with constraint-based models and possibility theory. *J Process Contr, 22*(10), 1946-1955. doi: https://doi.org/10.1016/j.jprocont.2012.09.001

Luo, J., Vijayasankaran, N., Autsen, J., Santuray, R., Hudson, T., Amanullah, A., & Li, F. (2012). Comparative metabolite analysis to understand lactate metabolism shift in Chinese hamster ovary cell culture process. *Biotechnol Bioeng, 109*(1), 146-156. doi: 10.1002/bit.23291

Mahadevan, R., Edwards, J. S., & Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophys J, 83*(3), 1331-1340.

Mahadevan, R., & Schilling, C. H. (2003a). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng, 5*(4), 264-276. doi: 10.1016/j.ymben.2003.09.002

Maranas, C. D., Zomorrodi, A. R., & Wiley Online Library. (2016). *Optimization methods in metabolic networks*. Hoboken, New Jersey: Wiley.

Marcus, P. I., Sato, G. H., Ham, R. G., & Patterson, D. (2006). A tribute to Dr. Theodore T. Puck (September 24, 1916-November 6, 2005). *In Vitro Cell Dev Biol Anim, 42*(8-9), 235-241. doi: 10.1290/0606039A.1

Martinez, V. S., Dietmair, S., Quek, L. E., Hodson, M. P., Gray, P., & Nielsen, L. K. (2013). Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. *Biotechnol Bioeng, 110*(2), 660-666. doi: Doi 10.1002/Bit.24728

Meadows, A. L., Karnik, R., Lam, H., Forestell, S., & Snedecor, B. (2010). Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation. *Metab Eng, 12*(2), 150-160. doi: DOI 10.1016/j.ymben.2009.07.006

Moles, C. G., Mendes, P., & Banga, J. R. (2003). Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res, 13*(11), 2467-2474. doi: 10.1101/gr.1262503

Moorkens, E., Meuwissen, N., Huys, I., Declerck, P., Vulto, A. G., & Simoens, S. (2017). The Market of Biopharmaceutical Medicines: A Snapshot of a Diverse Industrial Landscape. *Front Pharmacol, 8*. doi: Artn 314 10.3389/Fphar.2017.00314

Nabel, G. J. (2013). Designing tomorrow's vaccines. *N Engl J Med, 368*(6), 551-560. doi: 10.1056/NEJMra1204186

Naderi, S., Meshram, M., Wei, C., McConkey, B., Ingalls, B., Budman, H., & Scharer, J. (2011a). Development of a mathematical model for evaluating the dynamics of normal and apoptotic Chinese hamster ovary cells. *Biotechnol Prog, 27*(5), 1197-1205. doi: 10.1002/btpr.647

Naderi, S., Meshram, M., Wei, C., McConkey, B., Ingalls, B., Budman, H., & Scharer, J. (2011b). Development of a Mathematical Model for Evaluating the Dynamics of Normal and Apoptotic Chinese Hamster Ovary Cells. *Biotechnol Prog, 27*(5), 1197-1205. doi: 10.1002/btpr.647

Naderi, S., Nikdel, A., Meshram, M., McConkey, B., Ingalls, B., Budman, H., & Scharer, J. (2014). Modeling of cell culture damage and recovery leads to increased antibody and biomass productivity in CHO cell cultures. *Biotechnol J, 9*(9), 1152-1163. doi: 10.1002/biot.201300287

Newland, M., Kamal, M. N., Greenfield, P. F., & Nielsen, L. K. (1994). Ammonia inhibition of hybridomas propagated in batch, fed-batch, and continuous culture. *Biotechnol Bioeng, 43*(5), 434-438. doi: 10.1002/bit.260430512

Nienow, A. W. (2006). Reactor engineering in large scale animal cell culture. *Cytotechnology, 50*(1-3), 9-33. doi: 10.1007/s10616-006-9005-8

Nikdel, A., & Budman, H. (2016). Identification of active constraints in dynamic flux balance analysis. *Biotechnol Prog*. doi: 10.1002/btpr.2388

Nolan, R. P., & Lee, K. (2011). Dynamic model of CHO cell metabolism. *Metab Eng, 13*(1), 108-124. doi: 10.1016/j.ymben.2010.09.003

Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? *Nat Biotechnol, 28*(3), 245-248. doi: Doi 10.1038/Nbt.1614

Ozturk, S. S., & Hu, W.-S. (2006). *Cell culture technology for pharmaceutical and cell-based therapies*. New York, N.Y.; London: Taylor & Francis.

Philippidis, A. (2017). The Top 15 Best-Selling Drugs of 2016

Prospect of Price Curbs May Dent Future Results for Blockbusters. *GEN*.

Plotkin, S. (2014). History of vaccination. *Proc Natl Acad Sci U S A, 111*(34), 12283-12287. doi: 10.1073/pnas.1400472111

Pramanik, J., & Keasling, J. D. (1997). Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol Bioeng, 56*(4), 398-421. doi: Doi 10.1002/(Sici)1097-0290(19971120)56:4<398::Aid-Bit6>3.0.Co;2-J

Quek, L. E., Dietmair, S., Kromer, J. O., & Nielsen, L. K. (2010). Metabolic flux analysis in mammalian cell culture. *Metab Eng, 12*(2), 161-171. doi: 10.1016/j.ymben.2009.09.002

Radford, K., Niloperbowo, W., Reid, S., & Greenfield, P. F. (1991). Weaning of three hybridoma cell lines to serum free low protein medium. *Cytotechnology, 6*(1), 65-78.

Raghunathan, A. U., PÉRez-Correa, J. R., Agosin, E., & Biegler, L. T. (2006). Parameter estimation in metabolic flux balance models for batch fermentation—Formulation & Solution using Differential Variational Inequalities (DVIs). *Ann Oper Res, 148*(1), 251-270. doi: 10.1007/s10479-006-0086-8

Raghunathan, A. U., Perez-Correa, J. R., & Biegler, L. T. (2003). Data reconciliation and parameter estimation in flux-balance analysis. *Biotechnol Bioeng, 84*(6), 700-709. doi: 10.1002/bit.10823

Raman, K., & Chandra, N. (2009). Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform, 10*(4), 435-449. doi: 10.1093/bib/bbp011

Rand, D. A. (2008). Mapping global sensitivity of cellular network dynamics: sensitivity heat maps and a global summation law. *J R Soc Interface, 5 Suppl 1*, S59-69. doi: 10.1098/rsif.2008.0084.focus

Reed, J. L. (2012). Shrinking the Metabolic Solution Space Using Experimental Datasets. *PLoS Comput Biol, 8*(8). doi: 10.1371/journal.pcbi.1002662

Rodrigues, M. E., Costa, A. R., Henriques, M., Azeredo, J., & Oliveira, R. (2010). Technological progresses in monoclonal antibody production systems. *Biotechnol Prog, 26*(2), 332-351. doi: 10.1002/btpr.348

Routledge, S. J., Hewitt, C. J., Bora, N., & Bill, R. M. (2011). Antifoam addition to shake flask cultures of recombinant Pichia pastoris increases yield. *Microb Cell Fact, 10*, 17. doi: 10.1186/1475-2859-10-17

Ruiz, L., Traskine, M., Ferrer, I., Castro, E., Leal, J. F., Kaufman, M., & Carnero, A. (2008). Characterization of the p53 response to oncogene-induced senescence. *PLoS One, 3*(9), e3230. doi: 10.1371/journal.pone.0003230

Sanchez, C. E. G., & Saez, R. G. T. (2014). Comparison and Analysis of Objective Functions in Flux Balance Analysis. *Biotechnol Prog, 30*(5), 985-991. doi: 10.1002/btpr.1949

Savinell, J. M., & Palsson, B. O. (1992). Network Analysis of Intermediary Metabolism Using Linear Optimization .1. Development of Mathematical Formalism. *J Theor Biol, 154*(4), 421-454. doi: Doi 10.1016/S0022-5193(05)80161-4

Schilling, C. H., Edwards, J. S., Letscher, D., & Palsson, B. O. (2000). Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol Bioeng, 71*(4), 286-306. doi: Doi 10.1002/1097-0290(2000)71:4<286::Aid-Bit1018>3.0.Co;2-R

Schilling, C. H., Letscher, D., & Palsson, B. O. (2000). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol, 203*(3), 229-248. doi: 10.1006/jtbi.2000.1073

Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol Syst Biol, 3*, 119. doi: 10.1038/msb4100162

Sengupta, N., Rose, S. T., & Morgan, J. A. (2011). Metabolic Flux Analysis of CHO Cell Metabolism in the Late Non-Growth Phase. *Biotechnol Bioeng, 108*(1), 82-92. doi: 10.1002/bit.22890

Shuler, M. L., & Kargi, F. (1992). *Bioprocess engineering : Basic concepts*. Englewood Cliffs,N.J.: Prentice Hall.

Sidoli, F. R., Mantalaris, A., & Asprey, S. P. (2004). Modelling of Mammalian cells and cell culture processes. *Cytotechnology, 44*(1-2), 27-46. doi: 10.1023/B:CYTO.0000043397.94527.84

Strober, W. (2001). Trypan blue exclusion test of cell viability. *Curr Protoc Immunol, Appendix 3*, Appendix 3B. doi: 10.1002/0471142735.ima03bs21

Templeton, N., Dean, J., Reddy, P., & Young, J. D. (2013). Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. *Biotechnol Bioeng, 110*(7), 2013-2024.

Thalen, M., Venema, M., Dekker, A., Berwald, L., van den IJssel, J., Zomer, B., . . . Tramper, J. (2006). Fed-batch cultivation of Bordetella pertussis: Metabolism and Pertussis Toxin production. *Biologicals, 34*(4), 289-297. doi: 10.1016/j.biologicals.2005.12.001

Tomar, N., & De, R. K. (2013). Comparing methods for metabolic network analysis and an application to metabolic engineering. *Gene, 521*(1), 1-14. doi: 10.1016/j.gene.2013.03.017

Toussaint, C., Henry, O., & Durocher, Y. (2016). Metabolic engineering of CHO cells to alter lactate metabolism during fed-batch cultures. *J Biotechnol, 217*, 122-131. doi: 10.1016/j.jbiotec.2015.11.010

Toya, Y., Kono, N., Arakawa, K., & Tomita, M. (2011). Metabolic flux analysis and visualization. *J Proteome Res, 10*(8), 3313-3323. doi: 10.1021/pr2002885

Trummer, E., Fauland, K., Seidinger, S., Schriebl, K., Lattenmayer, C., Kunert, R., . . . Muller, D. (2006). Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors. *Biotechnol Bioeng, 94*(6), 1033-1044. doi: 10.1002/bit.21013

Ulmer, J. B., Valley, U., & Rappuoli, R. (2006). Vaccine manufacturing: challenges and solutions. *Nat Biotechnol, 24*(11), 1377-1383. doi: 10.1038/nbt1261

Vander Heiden, M. G., Cantley, L. C., & Thompson, C. B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science, 324*(5930), 1029-1033. doi: 10.1126/science.1160809

Varma, A., & Palsson, B. O. (1995). Parametric sensitivity of stoichiometric flux balance models applied to wild-type Escherichia coli metabolism. *Biotechnol Bioeng, 45*(1), 69-79. doi: 10.1002/bit.260450110

Wang, X., & Riviere, I. (2016). Clinical manufacturing of CAR T cells: foundation of a promising therapy. *Mol Ther Oncolytics, 3*, 16015. doi: 10.1038/mto.2016.15

Weiss, J. N. (1997). The Hill equation revisited: uses and misuses. *Faseb Journal, 11*(11), 835-841.

Wilkens, C. A., Altamirano, C., & Gerdtzen, Z. P. (2011). Comparative Metabolic Analysis of Lactate for CHO Cells in Glucose and Galactose. *Biotechnol Bioproc E, 16*(4), 714-724. doi: 10.1007/s12257-010-0409-0

Woodside, S., Bowen, B., & Piret, J. (1998). Mammalian cell retention devices for stirred perfusion bioreactors. *Cytotechnology, 28*(1-3), 163-175. doi: 10.1023/A:1008050202561

Xiao, H., Li, T.-K., Yang, J.-M., & Liu, L. F. (2003). Acidic pH induces topoisomerase II-mediated DNA damage. *P Natl Acad Sci, 100*(9), 5205-5210. doi: 10.1073/pnas.0935978100

Xing, Z., Kenty, B., Koyrakh, I., Borys, M., Pan, S.-H., & Li, Z. J. (2011). Optimizing amino acid composition of CHO cell culture media for a fusion protein production. *Process Biochem, 46*(7), 1423-1429. doi: https://doi.org/10.1016/j.procbio.2011.03.014

Yang, T. (2013). 13C-Based Metabolic Flux Analysis: Fundamentals and Practice. In H. S. Alper (Ed.), *Syst Metab Eng* (Vol. 985, pp. 297-334): Humana Press.

Yoon, S. K., Choi, S. L., Song, J. Y., & Lee, G. M. (2005). Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37.0 degrees C. *Biotechnol Bioeng, 89*(3), 345-356. doi: 10.1002/bit.20353

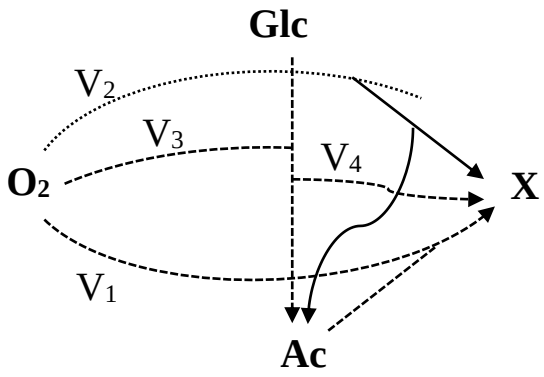
Zakrzewski, P., Medema, M. H., Gevorgyan, A., Kierzek, A. M., Breitling, R., & Takano, E. (2012). MultiMetEval: comparative and multi-objective analysis of genome-scale metabolic models. *PLoS One, 7*(12), e51511. doi: 10.1371/journal.pone.0051511

Zamorano, F., Wouwer, A. V., & Bastin, G. (2010). A detailed metabolic flux analysis of an underdetermined network of CHO cells. *J Biotechnol, 150*(4), 497-508. doi: 10.1016/j.jbiotec.2010.09.944

Zhang, L., Shen, H., & Zhang, Y. (2004). Fed-batch culture of hybridoma cells in serum-free medium using an optimized feeding strategy. *J Chem Technol Biot, 79*(2), 171-181. doi: 10.1002/jctb.940

Zomorrodi, A. R., Suthers, P. F., Ranganathan, S., & Maranas, C. D. (2012). Mathematical optimization applications in metabolic networks. *Metab Eng, 14*(6), 672-686. doi: 10.1016/j.ymben.2012.09.005

Appendix A



J_1 $39.43\ Ac + 35O_2 \rightarrow X$

J_2 $9.46Glc + 12.92O_2 \rightarrow X$

J_3 $9.84\ Glc + 12.73O_2 \rightarrow 1.24Ac + X$

J_4 $19.23Glc \rightarrow 12.12Ac + X$

Appendix B

Metabolic network of *B.Pertussis*

J1 Pyruvate → PEP

J2 Pyruvate + CoA → Acetyl-coA + CO2

J3 Acetyl-CoA + H2O +oxaloacetate → citrate + CoA

J4 citrate →α-ketoglutarate + CO2

J5 α-ketoglutarate + enz-N6 → succinyl transf. + CO2

J6 Succinyl transf. + CoA → succinyl-CoA + enz-N6

J7 Succinyl-CoA + phosphate → CoA + succinate

J8 Succinate + acceptor → fumarate + reduction acceptor

J9 Fumarate + H2O → malate

J10 malate → oxaloacetate

J11 glutamate + NH3 → phosphate + glutamine

J12 2 glutamate ← glutamine + α-ketoglutarate

J13 glutamate + H2O →α-ketoglutarate + NH3

J14 proline + 2H2O → glutamate

J15 oxaloacetate + glutamate → aspartate + α-ketoglutarate

J16 aspartate + NH3 ← aspargine

J17 PEP+HCO_3^- → phosphate + oxaloacetate*

J18 lactate → pyruvate

J19 acetate + CoA ← acetyl-CoA + phosphate

J20 2 acetyl-CoA → CoA + acetoacetyl-CoA

J21 acetoacetyl-CoA → PHB

J22 glucose-6-phosphate + 3 glyceraldehyde-3-phosphate → 3 ribose-5-phosphate

J23 acetyl-CoA + carrier-protein → acetoacetyl-carrier + CoA + CO_2

J24 Threonine + 3 pyruvate + 2 glutamate + acetyl-CoA + H_2O → NH_3 + $3CO_2$ + $2H_2O$ + isoleucine + α-ketoglutarate +valine + CoA + leucine

J25 serine + tetrahydrofolate ←→ 5,10-methylenetetrahydrofolate + glycine + H_2O

J26 serine ←→ pyruvate + NH_3

J27 threonine ←→ glycine + acetaldehyde

J28 aspartate→threonine + phosphate

J29 hydrogen sulfide + acetyl-CoA + serine → CoA + cysteine + acetate

J30 glutamate + pyruvate ←α-ketogultarate + alanine

J31 aspartate + pyruvate + glutamate + succynyl-CoA → phosphate + α-ketogultarate + succinate + lysine + CO_2 + CoA

J32 malate → pyruvate + CO_2

J33 amino acids → biomass

J34 amino acids → pertactin

J35 PEP → glyceraldehyde 3-P + phosphate

J36 2GAP + H_2O→glucose 6-P + phosphate

J37 J3

J38 amino acids → Pertussis toxin

J39 amino acids → fimbria

J40 amino acids → FHA

J41 Inverse of J14

J42 Inverse of J27†

J43 Inverse of J30

J44 2 glutamate + aspartate → fumarate + α-ketoglutarate + arginine

J45 Glucose + 6-P GAP + 2 PEP + glutamate → tyrosine + D-xylose + α-ketoglutarate

J46 Inverse of J26

J47 valine + α-ketoglutarate → glutamate + 4 methyl-2-oxopentanoate

J48 2 pyruvate +serine + aspartate + cysteine → CoA + 3 CO2 + glycine

J49 Inverse of J18

Stoichiometry of Purines

Adenine: pyruvate + 2 glutamine + 2 aspartate + glycine → adenine + 2 glutamate + fumarate

Guanine: pyruvate + 3 glutamine + aspartate +glycine → guanine + 3 glutamate + fumarate

Stoichiometry of Pyrimidines

UMP (Urydilic acid): glutamine + HCO_3^-+ aspartate + ribose-5-phosphate →UMP + CO2 + glutamate

CMP (Cytidilic acid): glutamine +HCO_3^- þ aspartate + ribose-5-phosphate + NH3 → CMP +CO2 + glutamate

TMP (Tymidilic acid): HCO_3^- + serine + glutamine + aspartate + ribose-5-phosphate → glycine + TMP + glutamate + CO2

Appendix C

Metabolic network of CHO

R1 Glc → G6P

R2 G6P → 2·3phosphaglycerate

R3 3phosphoglycerate →Pyr

R4 Pyr → Lac

R5 Pyr → AcCoA + CO2

R6 AcCoA + Oxal → Cit

R7 Cit → KG + CO_2

R8 KG →SucCoA + CO_2

R9 SucCoA → Suc

R10 Suc → Mal

R11 Mal → Oxal

R12 Mal → Pyr + CO_2

R13 Thr →Gly + AcCoA

R14 Trp → Ala + NH4 +2·AcCoA

R15 Lys → NH4 + KG

R16 Ile → Glu + AcCoA + SucCoA

R17 Leu → 2·AcCoA + 2·CO_2

R18 Tyr → Mal + Oxal + CO_2

R19 Ser + Met → Cys + NH4

R20 Val → SucCoA + KG

R21 Glu + Oxal → Asp + KG

R22 Glu → KG + NH4

R23 Glu + Pyr → Ala + KG

R24 Cys → Pyr

R25 Ser → NH4 + Pyr

R26 Gly → NH4 + CO2

R27 Ser + Thr → SucCoA

R28 Glu + 3phosphoglycerate → Ser + KG

R29 Ser → Gly

R30 Phe → Tyr

R31 Asn → Asp + NH4

R32 Gln → Glu + NH4

R33 Arg → Glu

R34 Glu → Pro

R35 His → Glu + NH4

R36 Gln + Asp → Glu + Asn

R37 0.0208·Glc + 0.0377·Gln + 0.0006·Glu + 0.007·Arg + 0.003·Hist + 0.0084·Ile + 0.0133·Leu + .0101·Lys + 0.0033·Met + 0.0055·Phe + 0.008·Thr + 0.004·Trp + 0.0096·Val + 0.0133·Ala + 0.026·Asp + 0.0004·Cys + 0.0165·Gly + 0.0081·Pro + 0.0099·Ser + 0.0077·Tyr →Biomass

R4, R21, R22, R23, R29, R30, R31, R32 and R36 are assumed reversible

Appendix D

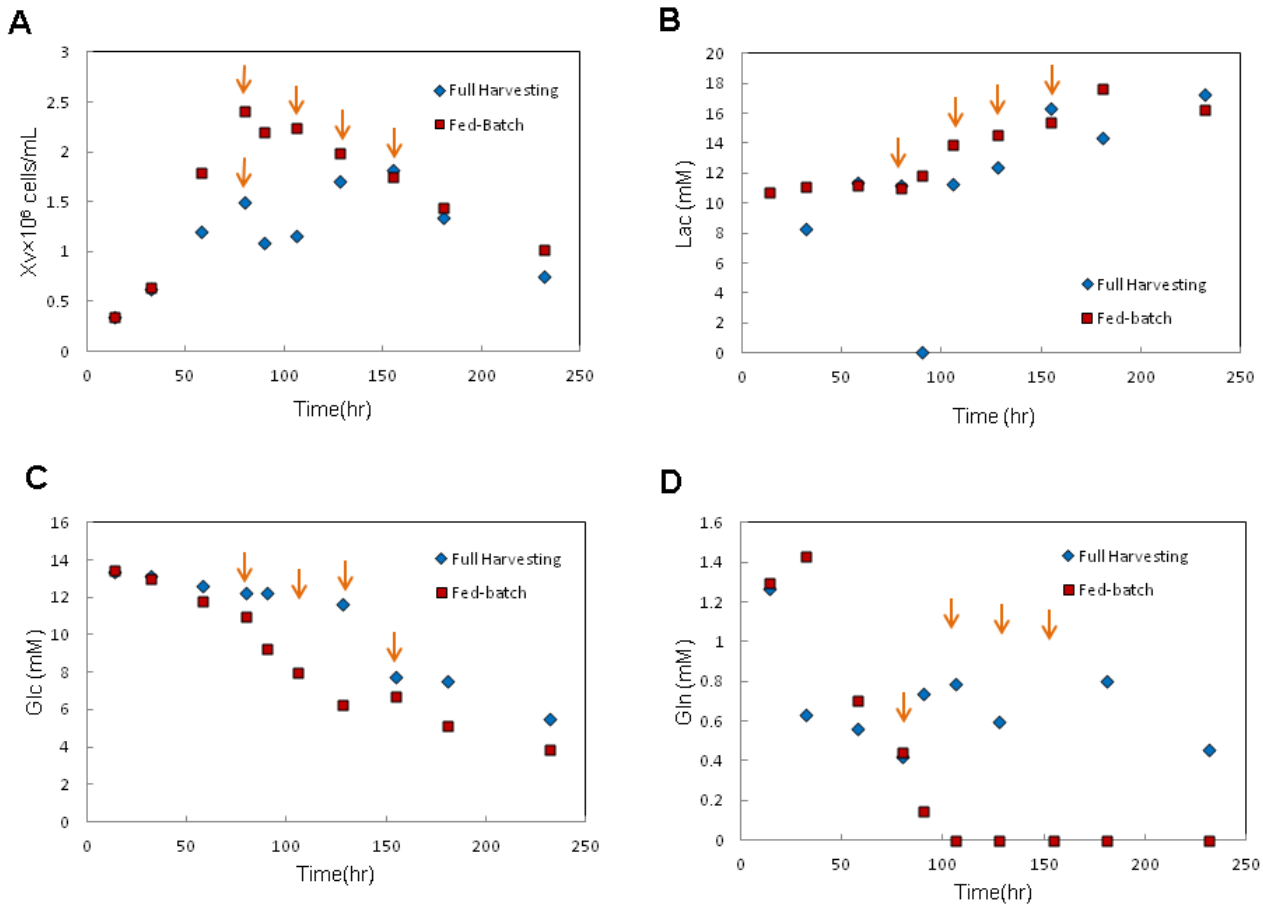


Figure S-1-A-D Single full harvesting and fed-batch culture