# MODELING BIOLOGICAL SYSTEMS WITH CYBERNETIC CONTROL LAWS AND STEADY STATE FLUX DISTRIBUTIONS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Kritika Kashyap

December 2019

## ABSTRACT

Metabolism is a process by which organisms extract energy from its environment to power itself. The interplay between cellular environment, nutrients present and the chemical reactions occurring within the organisms dictate the execution of metabolism. Single-cell microbes, such as *E. coli*, *P. putida* and yeast, can be redesigned to be miniature chemical reactors that transform sugars and other carbon sources into biofuels. Enhancing bio-process conditions can stimulate the production of therapeutic proteins in mammalian cells. Hence, understanding the strategy of metabolism in organisms would help us build better technologies for controlling formation of end products like cell mass, biofuels, therapeutic proteins etc. Steady state flux analysis techniques have gained a lot of popularity in the field of systems biology due to successful qualitative analysis of large scale metabolic networks. This technique has been used to improve production of ethanol in *Saccharomyces cerevisiae* and lactic acid in *Escherichia coli (E. coli)*, to name a few examples. Despite this success, its major shortcoming remains in providing quantitative measurements of metabolite profiles that can be further validated with experiments. Previous cybernetic models have accurately validated experimental results and described molecular processes governing metabolism, but their application is limited to small scale metabolic networks. We have built a mathematical framework that combines solutions from steady state flux analysis with cybernetic principles to overcome the challenge of describing metabolic data for large scale metabolic networks. We have established computational models that can capture dynamic metabolic regulation in bacterial and mammalian cells. These models integrate chemical kinetics, mass

balances, enzyme catalysis and stoichiometric laws to predict biology that can help us optimize cellular machinery towards bioproduct formation.

## BIOGRAPHICAL SKETCH

Kritika Kashyap was born in Ajmer, India. Her earlier interest in science was fueled by her parents. She then went on to complete her Bachelor of Technology degree in Chemical Engineering from Vellore Institute of Technology (VIT) at Vellore in India. Here she worked on the modeling of Acetylcholine (ACh) neurotransmitter between two neuron cells to understand the role of ACh in neurodegenerative diseases. Upon graduating from VIT Vellore, she joined the group of Dr. Jeffrey D. Varner at Cornell University as a graduate student to understand the nuances of metabolism in eukarotic and prokaryotic organisms by using mathematical modeling and computing techniques.

To my parents, family and friends, who have always stood by me...

## ACKNOWLEDGEMENTS

I, first and foremost would like to thank my research advisor Dr. Jeffrey D. Varner for allowing me to work in the his lab, for his guidance, perspective and for imparting training to me in the field of metabolic engineering. My experiences in his lab have motivated my interest in learning computational techniques. I would also like to thank my committee member Dr. Matthew P. DeLisa for his wise words of advice. I would like to express my gratitude to Dr. Michael Vilkhovoy for brainstorming problem solving strategies and helping me build ideas that have been implemented and have contributed towards my thesis. I also want to thank all members of the Varner lab for their support, advice and ideas. A heartfealt thanks goes to my family. To my parents for their support, help, guidance and patience in my journey towards accomplishing my Masters degree and completing my thesis successfully. Thanks to my siblings Piyush and Swati for their never ending moral support through these years. I would like to extend my gratitude towards Jay Gandhi, Maithili Deshpande, Yash Joshi, Venktesh Katkar and all the wonderful people I have met at Cornell University for the fantastic experiences and friendships I have been fortunate enough to be a part of. I would like to extend my gratitude towards Avni Gupta, Kirti Krishnamurthy Iyer and Srishti Khanna for their friendship and for bolstering my willpower and dedication towards work through constant encouragement and heartful conversations. I am thankful to Pallavi Pradeep and Siddharth Gumber for fostering my interest in academic research and engaging me in intellectual conversations in the areas of modern science and technology. Lastly, I would like to thank the invisible superpower that gives power and capacity to the human brain to persist, persevere and explore the unknown.

## TABLE OF CONTENTS

## LIST OF FIGURES

CHAPTER 1

# INTRODUCTION

## 1.1 Modern Biology

Biology has been transformed from a descriptive science to a design science. This transformation has been guided by the discovery of DNA as the repository of genetic information, and of recombinant DNA as an effective way to alter this instruction set. Researchers seek to make genetic changes in microbes to alter their metabolism in order to produce specific products of interest. Genetic engineering and synthetic biology have been established as effective tools to engineer biological cells. Understanding regulatory mechanisms that navigate the flow of metabolites through a maze of reactive paths will help us build better technologies. These tools can enhance production of renewable biofuels and other bioproducts to industrialize biology and build a sustainable society.

Synthetic biology holds a promising future in improving our ability to engineer biological systems. CRISPR-enabled genetic editing tools have revolutionized our ability to modify the DNA in vivo. We can implement changes in cellular behavior by introducing a DNA alteration. Unfortunately, the ability to control the DNA changes that we make in the cell does not guarantee a perfect prediction of the outcome of cellular behavior. Our inability to accurately predict biological behavior after introducing perturbations to the genotype acts as a big hurdle in synthetic biology.

## 1.2 Systems Biology Models

Mathematical representation of biological networks helps in understanding regulatory mechanisms within an organism. Modeling helps in making predictions about complex biological processes. Such tools can help predict how nutrients are utilized for the production of essential products. Mathematical modeling provides a systematic manner to predict the behavior of engineered systems. There are various ways in which a model can be formulated and designed. Some theoretical principles demonstrate promising capabilities for addressing problems in systems biology. The best models capture regularity in biological data without excessive complexity. Concepts of optimization and control hold a huge role in building these models. An arbitrary data set can be represented in various mathematical forms. The utility of a model is derived from its ability to describe regularity in data without being excessively complex. Regularity, or coherence in a set of data, means that the data is generated as the result of some intelligent process.

## 1.3 Kinetic Models

Kinetic models explicitly take into account enzyme kinetics and are able to predict metabolite concentrations as a function of time. Purely kinetic models aim to describe metabolic regulation with high precision. This type of prediction is useful to metabolic engineers in order to design pathways that have the desired titers, rates, and yields.

Kinetic models are described by a system of ordinary differential equations

(ODEs) that track the changes in concentration of various chemical species. Reaction rates are defined by algebraic equations which are written as a function of metabolite concentration. Rate expressions dictate the speed at which chemical species are transformed. These rate equations also capture enzyme behavior via linear kinetics, Michaelis-Menten kinetics and Monod kinetics. Linear kinetic models are good for determining stability criterion around steady state. Michaelis-Menten kinetics can also approach linear behavior at low substrate concentrations.

Traditionally kinetic models have been used to predict dynamics of biological pathways. But building them is time consuming and not as resourceful. The biggest limitation of these models is determination of kinetic parameters that are not available in literature. They take a significant amount of time to be developed. A fair amount of intuition and expertise is required in building these models. One must predict which interactions are of primal importance and which ones can be lumped together. Ensemble modeling approaches can help the modeler determine which interactions are necessary to include but require significant computational input to implement. Adding to the difficulty of kinetic modeling is the fact that, in vivo, many parameters related to biological phenomena change from what is measured in vitro.

## 1.4 Constraint-Based Models

Computational biology is focusing on large-scale modeling of dynamical systems predicting phenotype from genotype. The most widely used and successful modeling technique in metabolic engineering involves analysis of metabolic

fluxes through stoichiometric models of metabolism. Constraints-Based Models (CBMs) help understand the mechanism of nutrient utilization for product formation in cells. CBMs are applied to genome-scale networks for quick analysis of metabolic states. CBMs offer quick solutions to determine metabolic fluxes with no prior knowledge of parameters. Metabolic flux values are constrained by stoichiometry, thermodynamic and evolutionary assumptions, or experimental data. Flux distribution within a metabolic network can be analysed via Flux Balance Analysis (FBA) and Markov Chain Monte Carlo (MCMC) sampling. FBA and MCMC sampling is used for predicting rates of chemical reactions in a metabolic network.

FBA is an effective tool for analyzing the efficiency and performance of complex metabolic networks. FBA assumes objectives and constraints the solution space of flux rates. FBA has been successfully used to improve production of ethanol in Saccharomyces cerevisiae and lactic acid, succinic acid, lycopene and L-valine in *Escherichia coli (E. coli)*. FBA is formulated as a linear programming problem designed to find the flux profile that would maximize an objective function. FBA optimizes the flux distribution in the metabolic pathway to force maximum carbon uptake from glucose towards secretion of the desired product which is described as the objective function. However the FBA framework of assuming an objective function is very questionable from a biological perspective.

MCMC helps in obtaining an unbiased sample of flux distribution for a defined solution space. The sample of flux distribution can be characterised to study the properties of a metabolic network. Incorporating the Monte Carlo approach became necessary to determine feasible flux solutions through an un-

biased analysis. The advantage of using flux distribution solutions obtained from MCMC sampling is that they are not biased towards a biological objective unlike FBA.

The data obtained from these models can help us analyse biological systems and to incorporate genetic interventions that bring cell metabolism closer to the desired phenotype. Flux analysis techniques have gained popularity because of their ability to enhance bioproduct formation via metabolic pathway manipulation. While this approach has provided significant success, it has also shown its limitations due to its simplicity. Stoichiometric models are limited for bioengineering purposes because they ignore enzyme kinetics and cannot accurately capture dynamic metabolic responses. CBMs do not describe the dynamics and regulation inherent to biological systems.

## 1.5 Review On Various Cybernetic Models

Another class of frameworks called cybernetic models can overcome the shortcoming of CBMs. Cybernetic models find varied applications in bioprocess engineering and have successfully predicted mutant behavior, steady state multiplicity and strain specific metabolism. Cybernetic models have been developed over the last three decades to predict various kinds of metabolic phenomena. Cybernetic approach has been validated for various metabolic phenomena including resource consumption in mixed-substrate and single substrate environments. A unique attribute of the cybernetic model, not shared by constraint based models, is its ability to predict both yield and productivity of the desired end product in the metabolic network. The ability to describe the time-

dependent regulation of cellular reactions in the form of dynamic ”metabolic goals” is the most attractive quality of cybernetic models.

Cybernetic models predict metabolic dynamics based on the theory that regulatory mechanisms are governed by the capability to take optimal decisions which are directed by a goal. Cybernetics is the most powerful language for describing a system that has goals. Hence we can view a cell as a cybernetic system to investigate their metabolism. A cell acts, senses, compares itself to its goals and self corrects itself until it successfully achieve its goals. The cybernetic approach is based on this art of steering a cellular system towards its goal. Every intelligent system has these properties in order to achieve a successful existence. Cells accomplish this navigation towards its goals via an intelligent molecular infrastructure.

Cybernetic models are built on the concept that chemical reactions inside the cell are controlled to achieve goals related to an organisms survival. Every organism has a purpose of survival which dictates its regulatory mechanism. This is achieved in the model by modulating enzyme activity and synthesis to maximize the rate of substrate uptake or the rate of growth. The advantage of using the cybernetic framework is that it combines multiple molecular phenomena into a combined regulatory principle that can guide all these individual actions. Every molecular incident is a step towards enhancing the cell's survival. Cybernetic goals such as maximizing growth or carbon uptake rate provide a basis for the regulation of individual chemical events.

Cybernetic models assume that every organism makes a frugal use of its resources for the synthesis of key enzymes so that the resulting growth rate is maximized. Every biological system has the capability to respond to its envi-

ronment with the aim of promoting its survival. Living systems have a sophisticated machinery that has been evolved over the ages to help them survive in accordance to their environment. Changing environments would lead to changing response in the organism in order to meet its goals and requirements. Hence understanding the execution of metabolism is complex since it would highly depend on the nutrients involved and the cell environment. The uptake pattern in the cybernetic model would ensure that the uptake pattern at any instant allows the synthesis of precursors ensuring maximum growth rate.

The incorporation of control variables in the cybernetic framework keeps a check on the modulation of cellular chemical reactions towards its goals. A biochemical pathway may have a survival goal. But the actual pathway interactions can be hinderance in achieving this goal. The control variable has the ability to gauge the amount of control each enzyme has on the overall flux of a pathway.

The mathematical framework of cybernetic models is derived from a range of topics including mechanistic, kinetic models and linear programming approach. The cybernetic approach outperforms kinetic models by making accurate predictions with lesser efforts for parameterization and model specification. The beauty of this approach is that it isn't highly specific for the biological system that it defines. This method can be extended to any biological system for the purposes of optimization and control.

Till now many mathematical frameworks that describe metabolic processes have been developed. The cybernetic framework is based on the concept that the cell has preprogrammed artificial intelligence that coordinates complex mechanisms within it. This behavior is a result of years of evolution and re-

finement of organism towards a seemingly optimal behavior. The Darwinian theory points out that organisms that orient their behavior towards the goal of survival are better equipped to pass on their genes to the next generation. Behaviors that demonstrate optimality or perfected survival strategies, will be the most competitive and outlasting. Various stages of work in the field of cybernetic modeling frameworks have been discussed in this section.

### 1.5.1 Early Cybernetic Models

Early cybernetic formulation adapt a framework of ODEs that describe the time-based evolution of metabolite and enzyme concentrations. A complex biological machinery governs the changes in concentrations in a cell. The most important feature of cybernetic models are control variables that interpret the sum of regulatory processes inside a cell as actions that seek a goal related to the cells survival. These variables can accurately describe metabolism of a variety of biological systems without the knowledge of full details of mechanisms related to biological control. Monod's work on diauxie growth was the first system described by the cybernetic approach. *Bacillus subtilis* and *Escherichia coli (B. coli)* cells were exposed to the environment with multiple carbon substrates upon which the cells could grow. Changes in substrate and cell concentrations were tracked over time for this system. The results indicated that the cells preferred some substrates and would not simultaneously consume carbon sources.

### 1.5.2 Lumped Cybernetic Models

The goal of cybernetic modeling was to capture microbial growth on multiple substrates. Control variable u helps achieve this goal by describing resource allocation for enzyme generation for each substrates digestion. Kompala expanded upon some of the earlier work by adding another control variable v that describes enzyme activity for the digestion of different substrates. Following this, constitutive enzyme synthesis was incorporated into the framework. In other work by the same authors, provisions were added to describe maintenance, non-growth associated processes, within the cell. These models were able to explain the behavior of cell cultures at low growth rates on single and multiple substrates. However, it was incapable of predicting transients in continuous culture. To remedy this, Baloo and Ramkrishna included specifications for cellular resources that become limited during the lag period in step-up experiments, namely transcriptional resources for enzyme production. Kompala's model was unable to describe the simultaneous uptake of substrates. This was due to the assumption that growth rate on a mixture is never greater than the maximum growth rate of individual substrates. Kompalas model regarded the formation of biomass from substrate as a single step. Ramakrishna et al. were able to describe simultaneous uptake patterns in mixtures of substrates by including more than one biomass precursors. The inclusion of biomass precursors was a major development in the cybernetic approach. This marked the beginning of applications of cybernetic approach in metabolic engineering and systems biology because it describes the formation of biomass in multiple steps.

### 1.5.3 Structured Cybernetic Models

The next class of cybernetic models decomposed lumped pathways into functional units describing individual parts of the metabolic networks. The first effort in this direction sought to apply cybernetic principles related to resource competition among generic features present in metabolic networks including linear, branched and cyclic pathways. In principle, these features could be modeled with local objective functions and then integrated together to develop models for larger networks. Straight (1991) contributed towards developing cybernetic models for addressing metabolic networks by decomposing them into basic units and forming strategies for uptake of substrates. Large metabolic networks oscillate between coordinating regulation of intracellular reactions through local objectives versus global objectives that dictate the uptake of external nutrients. This approach is not good because the set of local objectives do not relate to the global metabolic goals. Varner and Ramkrishna expanded upon this scope of work. They postulated that individual enzymes are controlled by the product of both local and global objective functions. The local objectives determine resource allocation within pathways while global objectives determine regulation between different pathways. The local and global control policies complement one another towards the cells ultimate goal of survival. This model was able to accurately predict most flux ratios in the central carbon metabolism of *E. coli* Namjoshi et al. used structured cybernetic models to predict steady-state multiplicity for hybridoma cells in chemostat reactors using a highly simplified reaction network. Young et al refined the notion of global control variables. The global control variables maximize the rate of formation of various growth precursors which results in the dynamic maximization of growth. The key element of this work is how it divides the metabolic

network into related components which are similarly controlled.

### 1.5.4 Pathway Cybernetic Models

The formulation of structured cybernetic models was very cumbersome. It depended upon the modelers intuition when dividing the reaction network into various modules. Young et al developed a new way to incorporate reaction networks into the cybernetic framework by using elementary flux modes (EFMs). EFMs are minimal sequence of reactions beginning with the uptake of a substrate from outside the cell followed by a sequence of intracellular reactions and terminating in an extracellular product. If any reaction in the sequence is excluded, no flux can occur through the sequence. EFMs are non-decomposible pathways through a metabolic network from beginning to end. EFMs are not biased towards a local goal and provide a comprehensive breakdown of the network into its basic functional units. Hence, EFMs provide a consistent and general approach to modeling metabolic networks. In this method, a lumped biomass reaction which represents the conversion of precursors, ATP and NADH into a unit of biomass is used.

The cybernetic approach to a metabolic network aims to combine EFMs in such a way that the defined metabolic goal such as maximizing growth rate or uptake of substrate is accomplished. The metabolic goal of the organism is its global goal. Local goals control individual reactions in the EFM and will be employed towards enforcing the global goal. Identification of the set of EFMs is required for the implementation of such models. EFMs have the capability to capture the combination of all possible metabolic behaviors.

Pathway cybernetic models provide a mathematical framework that can help us unravel cellular mechanisms. EFMs are calculated and incorporated for solving these models. EFM calculation for genome scale metabolic networks is very cumbersome. This gave birth to a new breed of cybernetic models which Kim et al referred to as Hybrid Cybernetic Models (HCMs).

### 1.5.5 Hybrid Cybernetic Models

Kim et al expanded upon Pathway Cybernetic Models by building a modeling technique called Hybrid Cybernetic Models (HCMs). Kim et al applied the concepts of cybernetic modeling to large reaction networks. Cybernetic models use dynamic objective functions that optimize the system to achieve goals at each time through the inclusion of control variables that regulate enzyme synthesis and activity. Instead of exhaustively describing the kinetics of each reaction as the kinetic model does, hybrid cybernetic models (HCMs) decompose the reaction network into a set of pathways or macroscopic reactions termed EFMs that are expressed at varying levels over time.

HCMs uses makes a pseudo steady state assumption for the intracellular metabolites that regulate the effects of external fluxes into different EFMs. This eliminates the need for kinetic parameters for intermediate reactions. This is a huge advantage and helps in the application of HCMs to more complex reaction networks.

HCMs aim to maximize growth rate of an organism over a given period of time. A fixed amount of resources are utilized for the synthesis of enzymes that metabolize the carbon sources in the organism. HCMs choose the correct com-

bination of biochemical flux modes to achieve a desired biological goal. Each biochemical mode is catalyzed by a pseudoenzyme. The synthesis and activity of the pseudoenzyme is controlled by an optimal decision. HCMs have the capability of describing biological networks.

## 1.6 Combining Cybernetic and Constraint-Based Models

Generation of EFMs can be challenging and infeasible. Using flux analysis techniques instead of EFMs can reduce computational burden along with providing good results. In this work, a fusion of hybrid cybernetic modeling has been has been applied with steady state flux analysis techniques. These models can overcome the challenge of describing metabolic data for large scale metabolic networks. Hybrid cybernetic models can be integrated with steady state flux solutions to predict metabolism. This technique is an extension of the work done by Kim et al. HCMs can use flux solutions from FBA by ensuring maximal substrate uptake as its objective function. Our research proposes a detailed modeling framework that combines MCMC based flux solutions with cybernetic principles (HCM-MCMC). We have implemented HCM-FBA and HCM-MCMC focusing on the central carbon metabolism in *E. coli* and genome scale metabolic network of *Chinese Hamster Ovary cells (CHO-K1)*. The simulations predict real time concentration data for major metabolites. The output from these models validates experimental results. In silico modeling can improve our understanding of cell design by manipulating cellular pathways through media, feeding strategies, and other process parameters. Modeling tools such as these can help us optimize cellular machinery towards bioproduct formation and improve industrial yields.

HCM-FBA and HCM-MCMC successfully describe the formation of biomass in *E. coli* with glucose and acetate as substrates. This modeling framework has the capability to describe the sequential uptake of substrate in bacterial cells as well as simultaneous uptake of substrates in mammalian cells. HCM-FBA and HCM-MCMC are promising tools for building dynamic metabolic models with reduced computational burden and with higher accuracy.

The objective functions used in FBA are not universally descriptive nor applicable in all substrate consumption scenarios. Our efforts have been concentrated towards establishing that HCM-MCMC has no performance penalty compared to HCM-FBA. HCM-MCMC in fact offers a better solution strategy by eliminating the need to define an objective function that might not be biologically relevant.

It is the goal of this work to establish that return on investment can be broadened to describe objective functions in mammalian cells. The fundamental assumption of cybernetic models is that cells control metabolism towards goals related to their survival and propagation. This can be further extended to maximizing production of a certain protein of interest along with cell survival.

CHAPTER 2

# MATERIALS AND METHODS

## 2.1 Mathematical Formulation

The HCM-FBA and HCM-MCMC approach are modified form of HCM where FBA and MCMC solutions replace elementary mode solutions. The extracellular metabolites dictate the dynamics of the system while the intracellular metabolites are at pseudo steady state.

The reaction rates of each chemical transformation can be tracked. The regulated rate expressions for chemical transformation take a form similar to Michaelis-Menten kinetics and can be represented as follows

$$r_{M,l} = e_l \frac{k_{M,l} s_l}{K_{M,l} + s_l} \tag{2.1}$$

where $k_{M,l}$ is the maximum reaction rate and $K_{M,l}$ is the saturation constant of substrate $s_l$ in mode $l$. $s_l$ is the substrate that is being consumed for product formation. These are similar to the parameters from the kinetic model with one main distinction. They describe the rate of uptake into a pathway or set of reactions instead of the rate of a single reaction.

Pseudoenzymes that catalyze each pathway is formed in accordance to the following expression

$$r_{E,l} = \frac{k_{E,l} s_l}{K_{E,l} + s_l} \tag{2.2}$$

where $k_{E,l}$ is the rate constant for enzyme synthesis and $K_{E,l}$ is the saturation constant of substrate $s_l$ in mode $l$. $s_l$ is the substrate that is being consumed for product formation in mode $l$

In cybernetic modeling, there is a dichotomy between so-called regulated rates, $r_{M,l}$ , and unregulated rates, $r_{E,l}$. The unregulated rates are the enzyme-specific kinetics that determine the maximal rate at which a reaction or pathway can be expressed. The regulated rates are a function of the un-regulated rates and determine the time dependent changes in metabolites, $\frac{dx_i}{dt}$

The specific growth rate can be given by

$$r_{G,l} = z_{biomass,l} r_{M,l} \tag{2.3}$$

where $z_{biomass,l}$ denotes the flux via biomass reaction in mode $l$ and $r_{M,l}$ is the reaction rate for mode $l$. The framework being followed in this work normalizes flux through each mode by the specified objective flux. In our case, the objective flux is through the biomass reaction. Hence $z_{biomass,l}$ should be equal to unity after normalization.

Cybernetic control variables $u_l$ and $v_l$ guide the induced synthesis of enzymes and the allosteric regulation of enzyme activity respectively. Induced enzyme formation is expressed as some function of each pathways return on investment (ROI). ROIs typically are defined as each pathways rate of substrate uptake. This implies that cells prioritize their limited capacity for enzyme synthesis towards pathways with higher payoff. The set of $u_l$ control variables sum to 1. In this control variable, it is evident that there will be a higher investment of resources for a pathway if there is a higher return on investment or rate of growth.

To calculate the control of enzyme formation, ROIs are compared for each pathway in

$$u_l = \frac{z_{sl} r_{M,l}}{\sum_{l=1}^{L} z_{sl} r_{M,l}} \tag{2.4}$$

where $z_{sl}$ denotes the uptake flux of substrate $s$ through mode $l$. The denominator represents the sum of ROIs for all pathways. This means that the fraction of a finite resource pool devoted to the production of enzymes for one pathway is proportional to the ROI for that pathway. In addition to the efficient allocation of resources, pathways are also regulated by mechanisms that adjust their activity. To model this, pathways that have lower rates of growth than the maximal pathway are turned down via the proportional law. The activity of the different metabolic pathways is controlled by the $v_l$ variable which takes the form

$$v_l = \frac{z_{sl} r_{M,l}}{\max_{l=1,..L} z_{sl} r_{M,l}} \tag{2.5}$$

The pathway with the highest ROI will be fully expressed. All other pathways with lower ROIs will be down-regulated proportionally. For example, in the case of glucose and acetate pathways , one or the other should be turned down in order to reduce futile cycling. Therefore, the proportional cybernetic variables turn down pathways that do not represent the highest return on investment and fully express the pathway that does.

The total growth rate is given by

$$\mu = \sum_{l=1}^{L} r_{G,l} v_l \tag{2.6}$$

where L denotes the number of flux modes

The abundance of extracellular species $i(x_i)$, the pseudoenzyme $e_l$ (which catalyzes the flux through mode l), and cell mass is governed by

$$\frac{dx_i}{dt} = \sum_{j=1}^{R} \sum_{l=1}^{L} \sigma_{ij} z_{ij} r_{M,l} v_l c \qquad i = 1, ..., M \tag{2.7}$$

$$\frac{de_l}{dt} = \alpha_l + r_{E,l} u_l - (\beta_l + \mu) e_l \qquad l = 1, ..., L \tag{2.8}$$

$$\frac{dc}{dt} = \mu c \tag{2.9}$$

where R and M denote the number of reactions and extracellular species in the model, and L denotes the number of flux modes. The quantity $\sigma_{ij}$ denotes the stoichiometric coefficient for species $i$ in reaction $j$, and $z_{jl}$ denotes the normalised flux for reaction $j$ in mode $l$. If $\sigma ij > 0$, species $i$ is produced by reaction $j$; if $\sigma_{ij} < 0$, species $i$ is consumed by reaction $j$; if $\sigma_{ij} = 0$, species $i$ is not connected with reaction $j$. Extracellular species balances were subject to initial conditions $x(t_o) = x_o$ determined from the experimental data. The term $r_{M,l} v_l$ denotes the specific uptake/secretion rate for mode $l$, where e denotes the pseudoenzyme and c denotes the cell mass. Flux through each mode was catalyzed by a pseudoenzyme $e_l$, synthesized at the regulated specific rate $r_{E,l} u_l$, and constitutively at the rate $\alpha_l$. The term $u_l$ denotes the cybernetic variable

controlling synthesis of enzyme $l$. The term $\beta_l$ denotes the rate constant governing nonspecific enzyme degradation, and $\mu$ denotes the specific growth rate through all modes. Here $c$ represents biomass concentration. Concentration of the amount of pathway specific enzyme, $e_l$ , present in the system governs the conversion of substrate into biomass. The formation of cell mass is determined by the summation of the activity controlled cell mass formation reactions. The model equations were implemented in Julia (1.1.1) and solved using DifferentialEquations.jl.

## 2.2 Markov Chain Monte Carlo

For a given metabolic network, the number of MCMC modes can be quite high. Therefore, yield analysis is used to reduce the modes down to a minimal set that spans a given yield space. This makes the generation of the model for the subsequent results facile and reduces the number of parameters for the models specification. In steady state scenarios, HCM will only express one of these numerous modes that embodies some extreme edge of the yield space as determined using yield analysis. The flux modes were calculated using Artificial Hit and Run (ACHR) method of sampling implemented via COBRA toolbox on MATLAB R2017a.

## 2.3 Flux Balance Analysis

FBA modes are defined as the flux distribution through the network that connects substrate uptake to formation of extracellular products. FBA makes two

major assumptions to do so. One is that intracellular metabolites are at some pseudo-steady state. The other is that the cells organize their metabolic fluxes to optimize some objective function. FBA is written as an optimization problem as

$$\begin{aligned} \max_{w} \quad & (w_{obj} = \theta^T w) \\ s.t. \quad & S w = 0 \\ & \alpha_i \leq w_i \leq \beta_i \quad i = 1, 2, ..., R \end{aligned} \tag{2.10}$$

where S denotes the stoichiometric matrix, w denotes the unknown flux vector, $\theta$ denotes the objective function vector, and $\alpha_i$ and $\beta_i$ denote the lower and upper bounds on flux $w_i$, respectively. The lower and upper bounds are determined by experimental evidence as well as thermodynamic constraints on the reactions (i.e. some reactions only work in the forward direction). The experimental evidence is typically used to constrain the uptake rates of substrates consumption and product formation in the model. Other intracellular constraints can be used, but these quantities can be difficult to measure. The FBA problem was solved using GLPK on MATLAB R2017a, which is a software package intended for solving large-scale linear programming. For each FBA mode, the objective $w_{obj}$ was to maximize either the specific growth rate or the specific rate of byproduct formation. Multiple FBA modes were calculated for each objective by allowing the oxygen and nitrate uptake rates to vary. Each FBA mode was normalized by the specified objective flux.

## 2.4 Estimating Model Parameters

Model parameters can be estimated by minimizing the difference between simulations and experimental results (squared residual).

$$\min_{k} \sum_{\tau=1}^{T} \sum_{i=1}^{M} \left( \bar{x}_i(\tau) - x_i(\tau, k) \right)^2 \tag{2.11}$$

where $\bar{x}_i(\tau)$ denotes the measured value of species i at time $\tau$, $x_j(\tau, k)$ denotes the simulated value for species i at time $\tau$. The outer summation is with respect to time and the inner summation is with respect to state. The model residual was minimized using Neldermead implemented in the Julia programming language.

HCMs can accommodate large networks of intracellular reactions. HCMs require nonlinear optimization to estimate the kinetics of uptake into different modes. Identifying model parameters in this fashion can be computationally cumbersome and minimization searches do not guarantee that the search will conclude at a global minima. In the process of identification of parameters, the search function may have multiple local minima. Validation of parameters relies on predicting metabolic states on which the model was not trained.

CHAPTER 3

# *ESCHERICHIA COLI* SYSTEM

## 3.1 Abstract

Metabolic engineering can be applied to derive bioproducts from microbes with higher efficiency. The central carbon metabolism plays a crucial role in directing carbon fluxes towards formation of desired products. A mathematical model that describes the real time based variation of metabolites within *Escherichia coli (E. coli)* would help us enhance production processes. This chapter proposes a mathematical model that can describe dynamics of metabolic reactions within *E. coli*.

## 3.2 Introduction

Metabolic engineering aims to design biological systems based on quantitative predictions. Mathematical models can help aid these efforts. The lack of knowledge of mechanism of enzyme catalysis and kinetic parameters driving metabolism poses a huge challenge towards mathematical model formulation in the field. Kinetic modeling for metabolic networks is cumbersome due to complexity of the problem. Production of desired metabolites can be increased for industrial manufacturing by manipulating the central metabolic pathways towards amplifying the flux via objective reactions. Identifying target reactions that can be manoeuvred by metabolic pathways can help us build strategic models.

There are a number of objective functions that adequately describe a metabolic process. Bacterial cells are commonly modeled using objective functions that maximize growth yield. Maximization of carbon uptake is a popular objective function in the cybernetic framework. The modeling framework presented in this work can incorporate these biological objectives. FBA provides the flux distribution in a metabolic network which optimizes the objective function defined by the modeler. MCMC sampling can help achieve a sample of flux distribution in the region of maximum biomass yield or any other objective. HCM aims to club multiple regulatory mechanisms of biological control into an overarching principle of regulation.

The biological justification for using such objective functions while building mathematical models is still an open ended question. However such assumptions do have the capability to define the system dynamics. This is because these assumptions help define a basic structure for the model which is insensitive to the exact numerical value of kinetic parameters of the system. The given framework describes the central carbon metabolism in *E. coli* and supports the mentioned hypothesis. A rigid kinetic structure cannot capture the feedback and feedforward regulation of the metabolic network. Hence the mathematical framework must be designed towards an objective while incorporating the complexity of the system in a simple fashion. This chapter discuses the mathematical framework developed to describe the central carbon metabolism in *E. coli*

## 3.3 Methods

The simulated results have been validated against experimental results published by Varma et al.,1994 [3]. The initial metabolite concentration have been derived from aforementioned paper. The kinetic parameters for enzyme synthesis have been derived from Kim et al., 2008 [4]. The remaining kinetic parameters have been estimated.

A network of reactions describing the central carbon metabolism within the *E. coli* cell was used as a starting point. This network consisted of 118 reactions and 62 metabolites.

The modeling framework that we have implemented traces the time evolution of exchange fluxes with kinetic rate expressions. This helps in adapting flux analysis techniques to the cybernetic framework which helps in predicting dynamics of an unsteady system. This approach is based on the assumption that substrate uptake rates are major contributors towards determining the fate of the system.

Kinetic parameters were estimated by minimizing the difference between simulated data and metabolite time course data. The least squares minimization approach was followed for parameter estimation. This was formulated as a problem of non-linear optimization that was solved using Optim in Julia 1.1.1. The optimization solver used was Nelder-Mead. The objective function was to minimize the normalized sum squared difference between the simulated and measured concentrations, summed over all time points and extracellular metabolites. The mathematical formulation for parameter estimation is given in chapter 2 under section 2.4. It should be noted that a reasonable set of seed parame-

ters was critically important in the regression. Parameter estimation frequently failed to converge when random seeds were used. Therefore, seed parameters were first manually estimated based on experimental data and biochemical knowledge.

The reaction rates describing chemical transformations take a form similar to Michaelis-Menten kinetics. Twenty nine flux modes have been used to describe the dynamics of the system. The first eighteen flux modes utilize glucose for cell mass growth. The remaining eleven modes described cell mass growth via acetate consumption.

The mathematical formulation used in the work is as follows

$$r_{M,l} = e_l k_{M,l} \frac{s}{K_{s,l} + s} \tag{3.1}$$

where $k_{M,l}$ is the maximum reaction rate and $K_{s,l}$ is the saturation constant of substrate $s$ in mode $l$. $s$ represents glucose for the first 18 flux modes. $s$ represents acetate for the remaining 11 modes. The substrate is being consumed for product formation. This describes the rate of uptake into mode $l$ which consists of a set of reactions.

Pseudoenzymes that catalyze each pathway is formed in accordance to the following expression

$$r_{E,l} = k_{E,l} \frac{s}{K_{s,l} + s} \tag{3.2}$$

where $k_{E,l}$ is the rate constant for enzyme synthesis in mode $l$

The specific growth rate can be given by

$$r_{G,l} = z_{biomass,l} r_{M,l} \tag{3.3}$$

where $z_{biomass,l}$ denotes the flux via biomass reaction in mode $l$ and $r_{M,l}$ is the reaction rate for mode $l$.

Cybernetic control variables guide the induced synthesis of enzymes and the allosteric regulation of enzyme activity respectively. $u_l$ and $v_l$ can be written as follows

$$u_l = \frac{z_{s,l} r_{M,l}}{\sum_{l=1}^{L} z_{s,l} r_{M,l}} \tag{3.4}$$

$$v_l = \frac{z_{s,l} r_{M,l}}{\max_{l=1,..L} z_{s,l} r_{M,l}} \tag{3.5}$$

where $z_{s,l}$ denotes the uptake flux of substrate $s$ through the mode $l$. $s$ represents glucose for the first 18 flux modes. $s$ represents acetate for the remaining 11 modes. The total number of flux modes used in the given framework is 29, which means $L = 29$

The ordinary differential equations describing the time based variations in extracellular species, pseudoenzyme levels, and biomass concentration follows equation 2.7, 2.8 and 2.9 respectively from section 2.1 in chapter 2. The modeling equations describe batch fermentation of *E. coli*.

## 3.4 Results

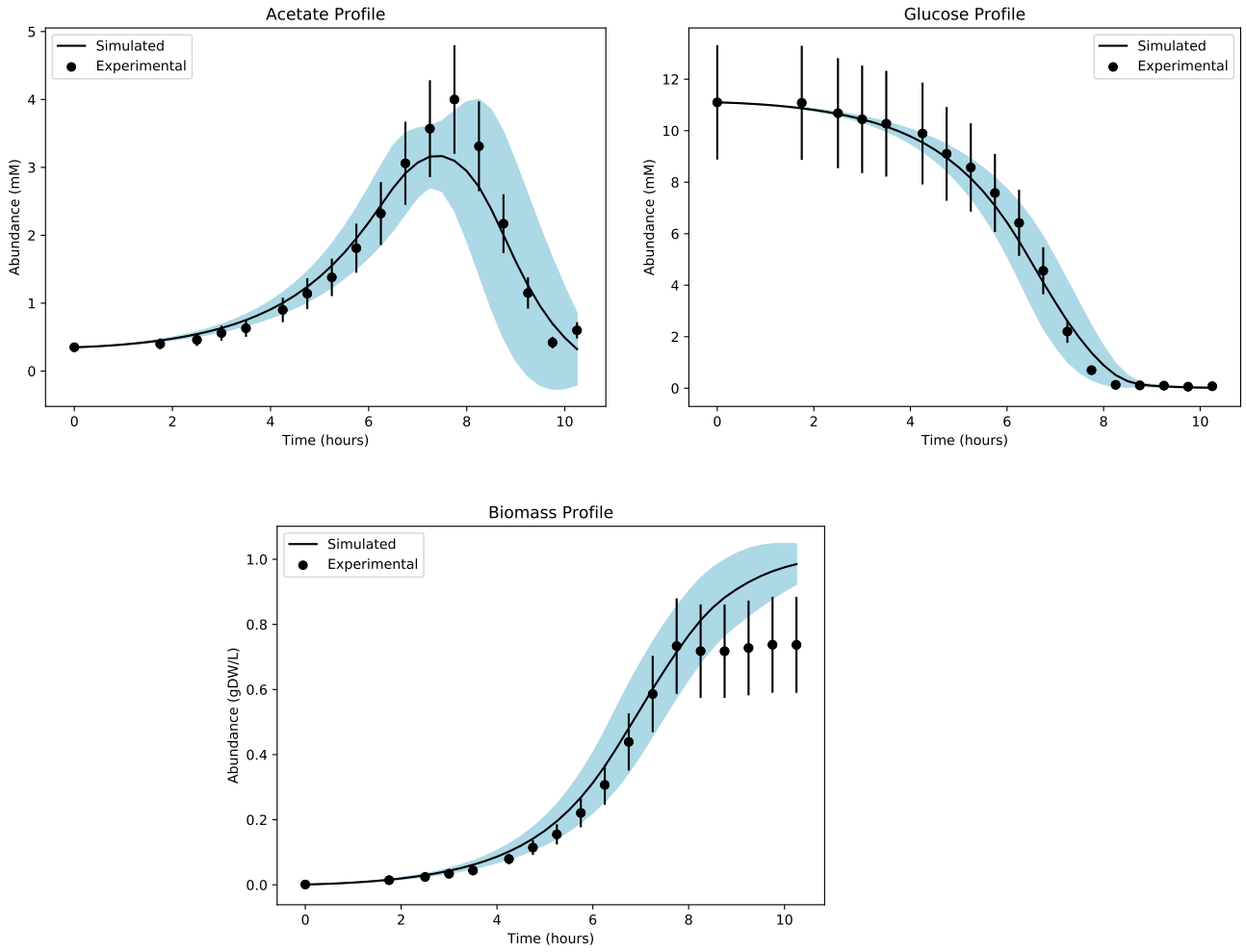


Figure 3.1: *E. coli* batch fermentation data versus HCM-MCMC (solid line) has been plotted. The shaded region represents 99% confidence interval. The error bars denote a 20% coefficient of variance.

An *E. coli* network of 118 reactions and 62 metabolites was constructed. 29 flux modes were generated using FBA and MCMC sampling. The initial concentration of metabolites is derived from the experimental data by Varma et al., 1994 [3]. The kinetic parameters for enzyme synthesis have been derived from Kim et al., 2008 [4]. Model parameters were estimated by minimizing the sum of squared errors between simulated and experimental results. The given math-

ematical model could capture the shift from glucose to acetate consumption and describe experimental results from Varma et al., 1994 [3]. *E. coli* first consumes extracellular glucose while generating acetate as a byproduct. Once the glucose has been exhausted, the culture then shifts its metabolic state to consume the acetate product.

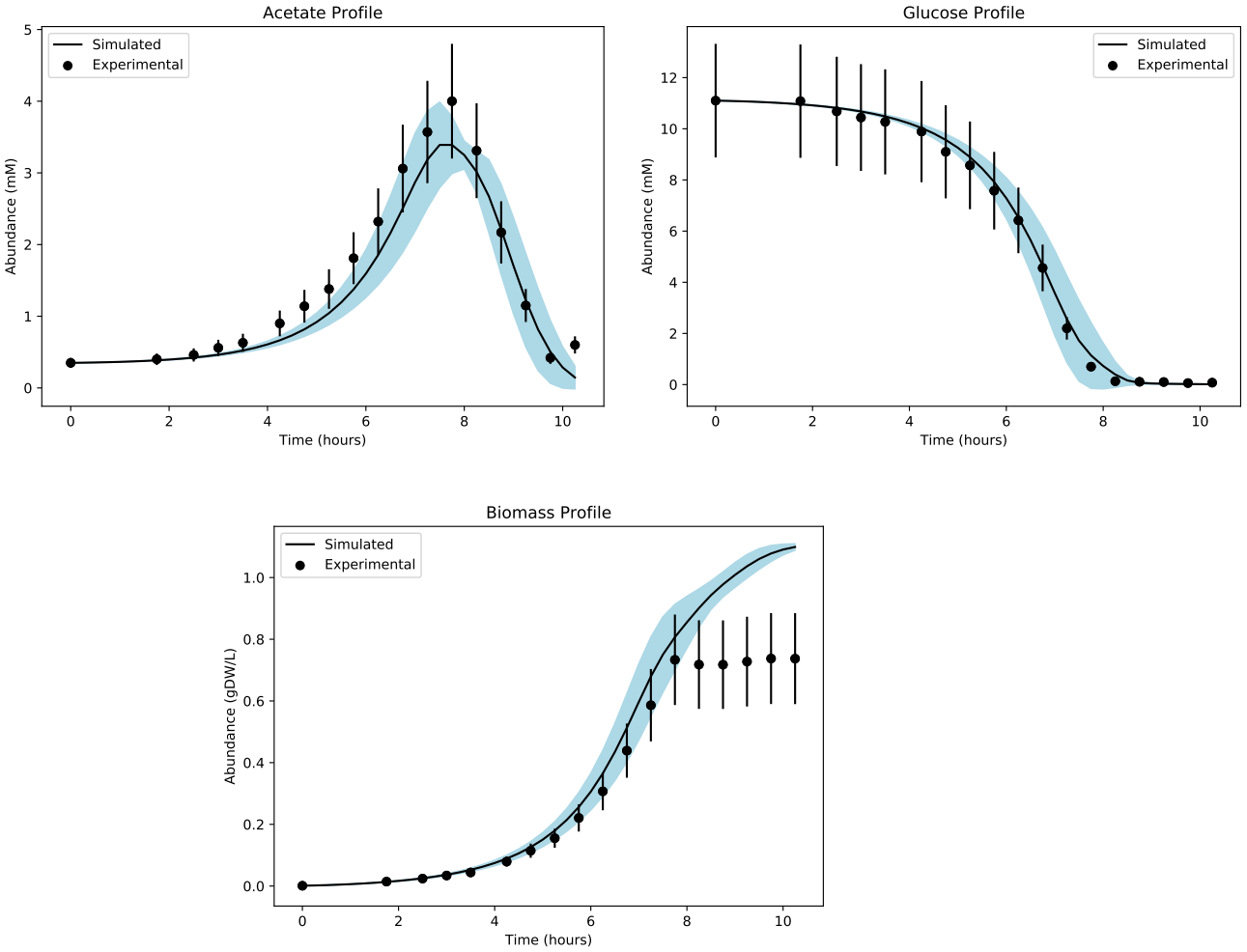


Figure 3.2: *E. coli* batch fermentation data versus HCM-FBA (solid line) has been plotted. The shaded region represents 99% confidence interval. The error bars denote a 20% coefficient of variance.

## 3.5 Discussions

The mathematical model accurately describes growth of biomass. The model system captured the initial growth phase on glucose and the second growth phase on acetate. The model can also capture the lag phase which represents transitioning from the preferred substrate to the secondary one. *E. coli* growing in a mixture of glucose and acetate, first consumes glucose and switches to acetate, a carbon source upon which it has a lower growth rate. The values of the control variables for each pathway changes dynamically. As the glucose level depletes, the $u$ variable for the glucose pathway decreases from one towards zero. This transition is made during the lag phase of the diauxic growth. At the same time, the $u$ variable for acetate pathways increase from a value close to zero towards one.

This work discusses the use of cybernetic modeling in combination with flux analysis techniques to simulate the dynamics of metabolism. These come under the umbrella of hybrid cybernetic models. HCM-FBA refers to a combination of cybernetics with FBA analysis. HCM-MCMC refers to a combination of cybernetic with MCMC flux sampling.

The complexity of internal relulatory mechanism in *E. coli* drives diauxic growth. HCMs are built on the hypothesis that synthesis and activity of the enzymatic machinery is regulated to maximize a return on investment. The HCM framework equates microorgamisms to optimal control systems. The metabolism in these systems is modulated via control variables for enzyme synthesis and enzyme activity. Another assumption this model make is that microorganisms possess limited internal resources that can be used for achieving

their goals. This modeling framework helped predict distinct growth and lag phases of diauxie in *E. coli*.

HCM-FBA and HCM-MCMC described dynamics of glucose, acetate and cell mass in *E. coli* by analysing its central carbon metabolism. A metabolic network containing 118 reactions has been studied. EFM analysis was infeasible for the given network. HCM-FBA and HCM-MCMC described cell-mass growth and the shift from glucose to acetate consumption with only 29 modes. Calculation of flux modes by FBA and MCMC sampling is very trivial even for a genome scale metabolic network. Hence HCM-FBA and HCM-MCMC opens the possibility for dynamic genome scale modeling of bacterial and mammalian cell metabolism.

The regulatory mechanism within a cell keeps a check on substrate uptake for meeting nutritional demand. A host of actions are coordinated when *E. coli* switches from growing on glucose to growing on acetate. Acetate uptake is achieved via Acetyl-coenzyme A synthetase (ACS) pathway where it is converted to other growth precursors via increased flux through the tricarboxylic acid (TCA) cycle. Upon the exhaustion of the preferred substrate, the production of lumped enzyme for the secondary substrate commences. Concentration of the amount of pathway specific enzyme, $e_l$ , present in the system governs the conversion of substrate into biomass. The rate of reaction is dependent on enzyme specific growth rate $r_G$ , substrate concentration, $s_l$ , and the total concentration of cells $c$.

HCM employs a combination of flux vectors (derived from FBA and MCMC) over time to describe the changes in fluxes. The cybernetic policy is selective towards producing the enzyme for one pathway. Once a steady state is reached,

there is only one pathway that will have a maximum unregulated rate for the substrate concentration present at a steady state. The inclusion of multiple flux solutions for a single substrate becomes an advantage for HCM as it can describe the span of an organisms metabolic concentration space. HCM incorporates multiple flux solution vectors (obtained from FBA and MCMC). This helps to capture the dynamic states predicted by the kinetic model best. The fact that goal signals can validate dynamics of metabolic networks makes a strong case for using them as objective functions in the modeling framework. This approach has potential applications ranging from bioprocessing to pharmacology.

CHAPTER 4

# CHINESE HAMSTER OVARY CELL SYSTEM

## 4.1 Abstract

There is a dearth of quantitative models which can predict the dynamics of metabolism in mammalian cells. This hampers the ability to build process optimization techniques for mammalian cell cultures. This work provides a framework for being able to simulate the dynamics of metabolism within *Chinese Hamster Ovary (CHO)* cells and understand the integration of biosynthetic pathways that contribute towards cell growth in fed-batch cultures. *CHO* cell lines are popularly used for the production of different biopharmaceutical drugs. Their high rate of consumption and a fast duplication cycle makes them an ideal biological clone. A careful balance of medium, bioreactor operational parameters and bioprocess is required to ensure optimal biopharmaceutical production. Constraint based MCMC sampling and FBA have been used to calculate flux distributions of intracellular and extracellular metabolites in the genome scale metabolic network of *CHO-K1*. The combination of steady state flux analysis and HCM provides a mathematical formulation that can help us derive solutions that describe the system dynamics. The framework discussed in this work is designed to define growth patterns in the controllable environmental conditions of a vessel.

## 4.2 Introduction

A biologic drug, also referred to as biologics, is a product that is produced from living organisms or contain components of living organisms. Biologic drugs include products derived from human, animal, or microorganisms by using biotechnology. Biologic products may contain proteins that control the action of other proteins and cellular processes, genes that control production of vital proteins, modified human hormones, or cells that produce substances that suppress or activate components of the immune system. Biologic drugs are sometimes referred to as biologic response modifiers because they change the manner of operation of natural biologic cellular actions.

Biologics have provided successful treatments for numerous diseases, including diabetes, arthritis, multiple sclerosis, cancer, anemia, and HIV. There is a high demand of biologics for building therapeutic interventions. Eukaryotic cells are popular hosts for producing recombinant therapeutics proteins since they enable post-translational modifications. *CHO* cells have become a standard industrial host since it has well established technologies for gene transfection, amplification and clone selection. Therefore, *CHO* cells are popularly used for pharmaceutical protein production. Fed-batch cultures are commonly used for growing *CHO* cells due to ease of operation, scale-up and higher volumetric productivity as compared to batch re-feed and perfusion cultures.

Fed-batch bioprocesses have been used since a long time for growing cells. However, its knowledge is not sufficient to describe cellular metabolism which is unique to every organism. Process perturbations and fluctuation in nutrient concentrations occur throughout the duration of fed-batch reactions. Lack of

knowledge of cellular metabolism would lead to poor formulation of the media and sub-optimal concentrations of the nutrients due to their utilization in unaccounted metabolic pathways. Thus there is requirement to enhance bio-process conditions that stimulate the development of mammalian cell lines for pharmaceutical interventions. Statistical design of experiments and high throughput screening (HTS) methods provide limited insights into the mechanistic details of a cell and does not give enough information for process optimization.

Mathematical models of metabolism can answer this problem and can help us optimize the production efficiency of a cellular process. Engineering *CHO* cells for increasing protein production is highly dependent upon time consuming and labour intensive empirical optimization. Mathematical modeling and simulation can aid these efforts by directing the utilization of necessary resources towards a chosen objective. Systematic study of time dependent metabolomics data can help in better understanding of molecular networks and cellular behaviour. The timing of highest protein production and changes in different cell growth phases may be ignored without time series data analysis. This knowledge will lead to better cellular design of *CHO* cells such that there is maximum production of the product of interest. In silico modeling can improve our understanding of cell design by manipulating cellular pathways through media, feeding strategies, and other process parameters. This contributes towards gaining knowledge for building better bioprocess optimization techniques.

Flux analysis techniques like FBA and MCMC sampling are used to elucidate cellular metabolism. These techniques can be used to calculate the redistribution of fluxes inside a cell when exposed to various environmental conditions. Flux rates for the biochemical reactions within *CHO-K1* metabolic network are

estimated based on experimental data. The flux rates of extracellular substrates dictate metabolism of the system. Extracellular substrates like alanine, ammonia, aparagine, asparate, glucose, glutamine, glutamate, glycine, lactate and serine have been constrained based on experimental data. Flux analysis can help us gain mechanistic insights into the metabolic events happening inside the cell. Ongoing work in understanding *CHO* cell metabolism has helped in establishing databases that trace fluxes into biomass and byproducts. This will eventually help in building better bioreactor models for *CHO* cells. The knowledge of flux distribution within a metabolic pathway can guide researchers towards building better mathematical models. The mathematical framework used in this work combines steady state flux analysis and kinetic rate equations. These are referred to as Hybrid Cybernetic Models. Hybrid Cybernetic Models define uptake and output reactions using kinetic rate equations. This takes into account the time and concentration dependency. All intracellular reactions are assumed to be at steady state. This simplifies the calculation by a significant amount and reduces time of solving by eliminating the need to calculate kinetic parameters for intracellular reactions. The simulation describes the behavior of the culture at various time steps, assuming the existence of a pseudo-steady state at each step and the extracellular metabolites dictating the metabolism of the system. Due to steady state assumption, these models can only predict the dynamic cellular responses to external medium changes and other perturbations.

Modeling can be a challenge even after implementing the pseudo steady state assumption. It can be difficult to define a model for all extracellular reactions which have concentration dependent rate equations. This calls for the need of more model simplifications which can be achieved by reducing the network size, using macroscopic reactions, using cellular objectives like maximiz-

ing biomass, recombinant protein production or energy utilization, adding constraints, such as pre-defined reaction rates or empirical equations and dividing the process into distinct phases, with defined functions to transition between phases.

The aforementioned assumptions might be applicable to certain biological systems depending on the goal. An ideal model would be flexible enough to describe dynamics of multiple biological system scenarios. The process engineer building these formulations must ensure that the model only uses initial metabolite concentrations, cell density and parameter set as the input for the model. It must be ensured that the mechanism describing the system dynamics can be traced back to the biochemical pathways describing the system. The modeling framework must be capable of describing how manipulating the process variables can impact the system dynamics on a large time scale (scaling range includes minutes, hours, days and weeks).

In this work, a novel simulation framework is presented which builds on the HCM framework to predict the metabolic dynamics of *CHO* cells in fed-batch culture. The model accounts for coordinated regulation of the various pathways by incorporating control variables.

## 4.3 Methods

The simulated results have been validated against experimental results published by Nolan et al., 2011 [1]. The initial metabolite concentration and saturation constants have been derived from aforementioned paper. The authors have measured and described dry cell weight as 350 pg/cell. The experiments

carried by the authors have determined that $10^6$ cells/mL is equivalent to 2.31 mM biomass. The conversion factors and units in this work have been based on the calculations carried by Nolan et al., 2011 [1].

A network of reactions describing the metabolism within the *CHO-K1* cell lines was taken from Hefzi et al., 2016 [2] and was used as a starting point. This network consisted of 4723 reactions and 2773 metabolites, comprising the metabolic pathways of glycolysis, pentose phosphate pathway (PPP), citrate cycle, amino acid metabolism, oxidative phosphorylation, and biomass synthesis. An addition was made to the reaction network provided the Hefzi et al., 2016 [2]. Two more reactions describing antibody formation were added to the reaction network described by Hefzi et al., 2016 [2]. The reactions that were added have been taken from the reaction network described by Nolan et al., 2011 [1]. Due to this addition, the final reaction network consisted of 4725 reactions and 2775 metabolites.

The modeling framework that we have implemented traces the time evolution of exchange fluxes with kinetic rate expressions. This helps in adapting flux analysis techniques to the cybernetic framework which helps in predicting dynamics of an unsteady system. This approach is based on the assumption that substrate uptake rates are major contributors towards determining the fate of the system.

The initial metabolite concentration and saturation constants have been taken from the experimental data collected from Nolan at al.,2011 [1]. The remaining kinetic parameters were estimated by minimizing the difference between simulated data and metabolite time course data obtained from the aforementioned research paper. The least squares minimization approach was followed

for parameter estimation. This was formulated as a problem of non-linear optimization that was solved using Optim in Julia 1.1.1. The objective function was to minimize the normalized sum squared difference between the simulated and measured concentrations, summed over all time points and extracellular metabolites. The mathematical formulation for parameter estimation is given in chapter 2 under section 2.4. It should be noted that a reasonable set of seed parameters was critically important in the regression. Parameter estimation frequently failed to converge when random seeds were used. Therefore, seed parameters were first manually estimated based on experimental data and biochemical knowledge.

The reaction rates describing chemical transformations take a form similar to Michaelis-Menten kinetics. Three flux modes have been used to describe the dynamics of the system. Each flux mode corresponds to a particular time frame within the solution interval and represents multiple substrates being utilized in that time frame.

The rate equation for the first flux mode is formulated as follows

$$r_{M,1} = e_1 k_{M,1} \frac{s_1}{K_{s_1,1} + s_1} \frac{s_2}{K_{s_2,1} + s_2} \frac{s_3}{K_{s_3,1} + s_3} \frac{s_4}{K_{s_4,1} + s_4} \tag{4.1}$$

where $k_{M,1}$ is the maximum reaction rate and $K_{s_1,1}$, $K_{s_2,1}$, $K_{s_3,1}$ and $K_{s_4,1}$ are the saturation constant of $s_1$, $s_2$, $s_3$ and $s_4$ in mode 1. $s_1$ represents glucose, $s_2$ represents serine, $s_3$ represents asparagine and $s_4$ represents glutamine. These substrates are being consumed for product formation. This describes the rate of uptake into mode 1 which consists of a set of reactions.

Pseudoenzymes that catalyze the first pathway is formed in accordance to the

following expression

$$r_{E,1} = k_{E,1} \frac{s_1}{K_{s_1,1} + s_1} \frac{s_2}{K_{s_2,1} + s_2} \frac{s_3}{K_{s_3,1} + s_3} \frac{s_4}{K_{s_4,1} + s_4} \tag{4.2}$$

where $k_{E,1}$ is the rate constant for enzyme synthesis in mode 1

The specific growth rate can be given by

$$r_{G,1} = z_{biomass,1} r_{M,1} \tag{4.3}$$

where $z_{biomass,1}$ denotes the flux via biomass reaction in mode 1 and $r_{M,1}$ is the reaction rate for mode 1.

The rate equation for the second flux mode is formulated as follows

$$r_{M,2} = e_2 k_{M,2} \frac{s_1}{K_{s_1,2} + s_1} \frac{s_2}{K_{s_2,2} + s_2} \frac{s_3}{K_{s_3,2} + s_3} \frac{s_5}{K_{s_5,2} + s_5} \frac{s_6}{K_{s_6,2} + s_6} \frac{s_7}{K_{s_7,2} + s_7} \tag{4.4}$$

where $k_{M,2}$ is the maximum reaction rate and $K_{s_1,2}$, $K_{s_2,2}$, $K_{s_3,2}$, $K_{s_5,2}$,$K_{s_6,2}$ and $K_{s_7,2}$ are the saturation constant of $s_1$, $s_2$, $s_3$, $s_5$, $s_6$ and $s_7$ in mode 2. $s_1$ represents glucose, $s_2$ represents serine, $s_3$ represents asparagine, $s_5$ represents lactate, $s_6$ represents glycine and $s_7$ represents ammonia. These substrates are being consumed for product formation. This describes the rate of uptake into mode 2 which consists of a set of reactions.

Pseudoenzymes that catalyze the second pathway is formed in accordance to the following expression

$$r_{E,2} = k_{E,2} \frac{s_1}{K_{s_1,2} + s_1} \frac{s_2}{K_{s_2,2} + s_2} \frac{s_3}{K_{s_3,2} + s_3} \frac{s_5}{K_{s_5,2} + s_5} \frac{s_6}{K_{s_6,2} + s_6} \frac{s_7}{K_{s_7,2} + s_7} \tag{4.5}$$

where $k_{E,2}$ is the rate constant for enzyme synthesis in mode 2

The specific growth rate can be given by

$$r_{G,2} = z_{biomass,2} r_{M,2} \tag{4.6}$$

where $z_{biomass,2}$ denotes the flux via biomass reaction in mode 2 and $r_{M,2}$ is the reaction rate for mode 2.

The rate equation for the third flux mode is formulated as follows

$$r_{M,3} = e_3 k_{M,3} \frac{s_1}{K_{s_1,3} + s_1} \frac{s_2}{K_{s_2,3} + s_2} \frac{s_3}{K_{s_3,3} + s_3} \frac{s_5}{K_{s_5,3} + s_5} \frac{s_8}{K_{s_8,3} + s_8} \frac{s_9}{K_{s_9,3} + s_9} \tag{4.7}$$

where $k_{M,3}$ is the maximum reaction rate and $K_{s_1,3}$, $K_{s_2,3}$, $K_{s_3,3}$, $K_{s_5,3}$,$K_{s_8,3}$ and $K_{s_8,3}$ are the saturation constant of $s_1$, $s_2$, $s_3$, $s_5$, $s_8$ and $s_9$ in mode 3. $s_1$ represents glucose, $s_2$ represents serine, $s_3$ represents asparagine, $s_5$ represents lactate, $s_8$ represents alanine and $s_9$ represents asparate. These substrates are being consumed for product formation. This describes the rate of uptake into mode 3 which consists of a set of reactions.

Pseudoenzymes that catalyze the third pathway is formed in accordance to the following expression

$$r_{E,3} = k_{E,3} \frac{s_1}{K_{s_1,3} + s_1} \frac{s_2}{K_{s_2,3} + s_2} \frac{s_3}{K_{s_3,3} + s_3} \frac{s_5}{K_{s_5,3} + s_5} \frac{s_8}{K_{s_8,3} + s_8} \frac{s_9}{K_{s_9,3} + s_9} \tag{4.8}$$

where $k_{E,3}$ is the rate constant for enzyme synthesis in mode 3

The specific growth rate can be given by

$$r_{G,3} = z_{biomass,3} r_{M,3} \tag{4.9}$$

where $z_{biomass,3}$ denotes the flux via biomass reaction in mode 3 and $r_{M,3}$ is the reaction rate for mode 3.

The system of CHO-K1 cell growth is described by multisubstrate kinetics. Glutamine, asparagine and lactate play a very important role in substrate kinetics. To simplify understanding the description of control variables in this system, another variable called the cybernetic variable is described as follows

$$cyberneticvariable_l = z_{s,l} r_{M,l} \tag{4.10}$$

where $z_{s,l}$ describes flux of a particular substrate $s$ through a flux mode $l$ and $r_{M,l}$ describes the reaction rate of flux mode $l$. The two terms $z_{s,l}$ and $r_{M,l}$ define the cybernetic variable. The system of CHO-K1 cells has been defined by 3 flux modes. The first flux mode is dependent on the uptake of glutamine towards cell growth. Hence the cybernetic variable for the first flux mode is defined as follows

$$cyberneticvariable_1 = z_{s_4,1} r_{M,1} \tag{4.11}$$

where $z_{s_4,1}$ denotes the uptake flux of substrate $s_4$ i.e. glutamine through the first flux mode. The second flux mode is dependent on the uptake of asparagine

towards cell growth. Hence the cybernetic variable for the second flux mode is defined as follows

$$cyberneticvariable_2 = z_{s_3,2} r_{M,2} \tag{4.12}$$

where $z_{s_3,2}$ denotes the uptake flux of substrate $s_3$ i.e. asparagine through the second flux mode. The third flux mode is dependent on the uptake of asparagine and lactate towards cell growth. Hence the cybernetic variable for the third flux mode is defined as follows

$$cyberneticvariable_3 = (z_{s_3,3} + z_{s_5,3}) r_{M,3} \tag{4.13}$$

where $z_{s_3,3}$ denotes the uptake flux of substrate $s_3$ i.e. asparagine through the third flux mode and $z_{s_5,3}$ denotes the uptake flux of substrate $s_5$ i.e. lactate through the third flux mode.

Cybernetic control variables guide the induced synthesis of enzymes and the allosteric regulation of enzyme activity respectively. $u_l$ and $v_l$ can be written as follows

$$u_l = \frac{cyberneticvariable_l}{\sum_{l=1}^{L} cyberneticvariable_l} \tag{4.14}$$

$$v_l = \frac{cyberneticvariable_l}{\max_{l=1,..L} cyberneticvariable_l} \tag{4.15}$$

where total number of flux modes used in the given framework is 3, which means $L = 3$

The modeling equations for *CHO-K1* describe fed-batch operation. The experimental set up by Nolan et al., 2011 [1] describes a media flow feed stream that enters the reactor and contributes to the kinetics occuring within. This brings a factor of volumetric flow rate into the modeling equations. Starting on day 3 daily bolus feeds at 3 % v/v of a proprietary media feed were supplied to the reactor which would increase the volume of reaction mixture and contribute towards manipulating the reaction kinetics. The experimental design uses a 2-L Applikon bioreactor. Hence the initial volume of the reactor is assumed to be 1.2-L. The model equation for fed-batch operation of *CHO-K1* cell lines can be given as follows.

The volumetric flow rate ,F, at which the feed stream enters the reactor can be written as follows,

$$F = \frac{dV}{dt} \tag{4.16}$$

where V is the volume of reaction mixture

Adding a feed stream to the reactor would lead to entry of more substrate available for consumption. It also leads to dilution of the reaction mixture due to the volumetric changes that occur with it. Hence the rate of dilution, D, can be written as,

$$D = \frac{F}{V} \tag{4.17}$$

The abundance of extracellular species $i(x_i)$, the pseudoenzyme $e_l$ (which catalyzes the flux through mode l), and cell mass is governed by

$$\frac{dx_i}{dt} = D(x_f - x_i) + \sum_{j=1}^{R} \sum_{l=1}^{L} \sigma_{ij} z_{ij} r_{M,l} v_l c \qquad i = 1, ..., 10; j = 1, ..., 546 \qquad (4.18)$$

$$\frac{de_l}{dt} = \alpha_l + r_{E,l} u_l - (\beta_l + \mu) e_l \qquad l = 1, 2, 3 \qquad (4.19)$$

$$\frac{dc}{dt} = (\mu - D - k_d) c \qquad (4.20)$$

where $x_f$ is the concentration of extracellular species in the feed stream, $k_d$ is the death rate of cells and D is the rate of dilution.

## 4.4 Results

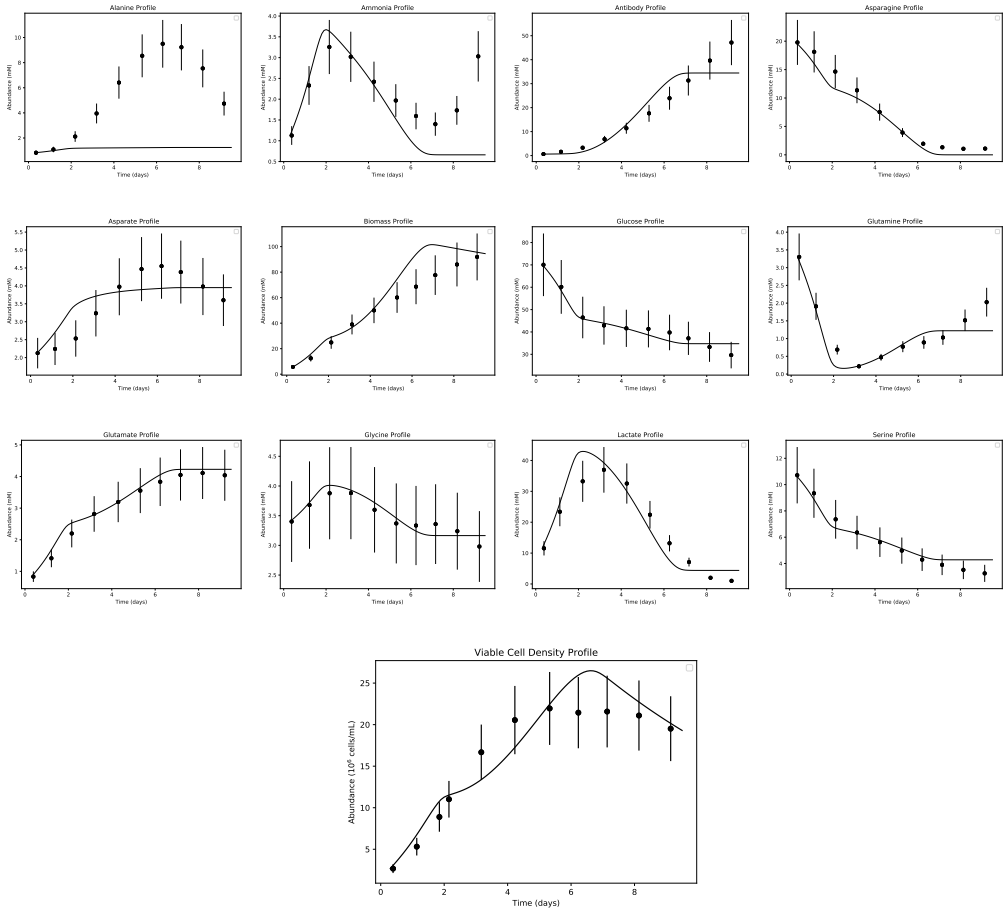


Figure 4.1: Experimental data for extracellular metabolite concentrations in *CHO-K1* cells versus HCM-MCMC (solid line) has been plotted. The error bars denote a 20% coefficient of variance.

The use of FBA and MCMC flux distribution assumes that the pseudo-steady state assumption is valid throughout the fed-batch. The model simulated profiles of the extracellular metabolite concentrations. The simulations were performed with an initial set of metabolite concentrations and parameter values as the only inputs.

Importantly, the simulations correctly predicted both the timing and magni-

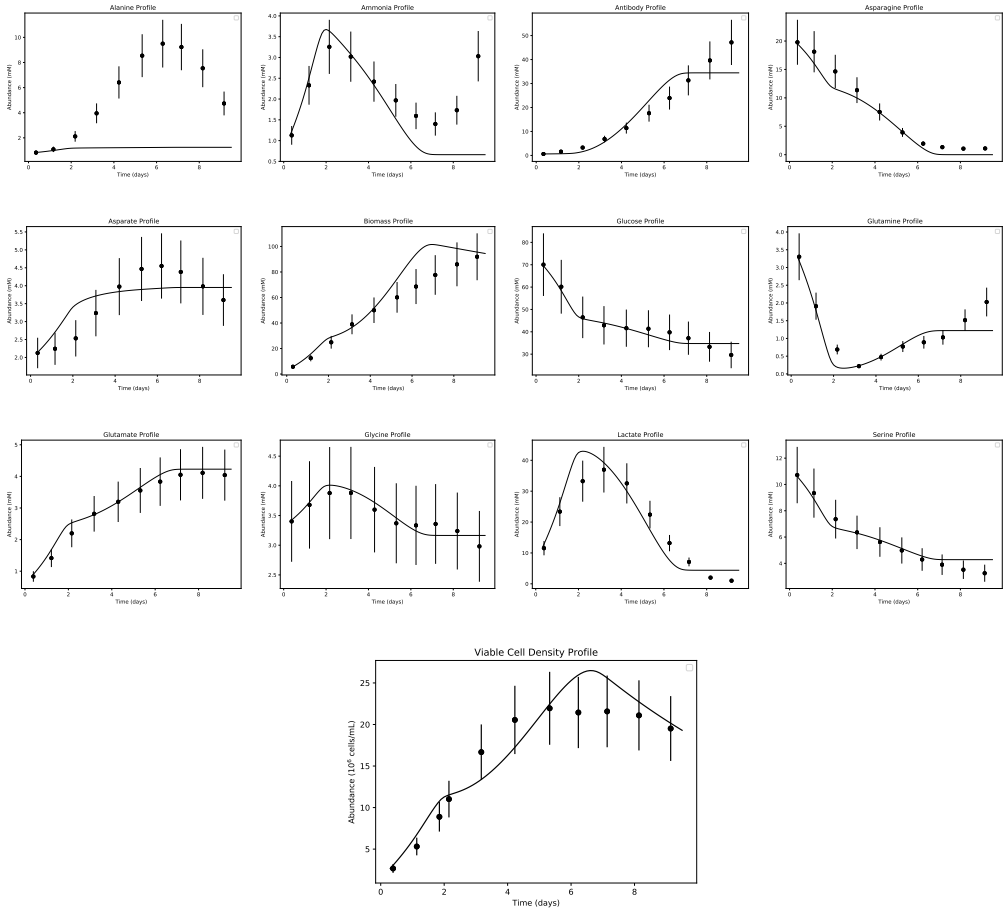


Figure 4.2: Experimental data for extracellular metabolite concentrations in *CHO-K1* cells versus HCM-FBA (solid line) has been plotted. The error bars denote a 20% coefficient of variance.

tude of metabolic shifts, which are characterized by transitions from production to consumption of external metabolites, such as lactate.

To validate the model, simulated concentration profiles of selected metabolites for representative process conditions were compared with experimentally determined trends. The simulated profiles closely tracked the measured profiles.

Time dependent metabolite profiles for extra-cellular alanine, ammonia, an-

tibody, aparagine, asparate, biomass, glucose, glutamine, glutamate, glycine, lactate,serine and viable cell density were measured for 9 days. Biomass abundance showed faster increase from day 0 to 2 after which rate of growth became slower.

The rate of glucose uptake slows down after 2 days. This impacts the rate of growth. The presence of lactate can compensate for the slow uptake rate of glucose and contribute towards cell proliferation. Lactate starts depleting at the same time as rate of glucose consumption lowers. This implies that the cell can grow in a multi-substrate environment and utilize different carbon sources simultaneously.

For 2 to 3 days cells exhibit very high consumption rates of glucose and glutamine. This can be referred to as the ideal growth phase. This phase also results in high secretion of lactate and glutamate. Amino acids like serine and asparagine are mostly depleted. At the end of 3 days, glucose consumption rate lowers and glutamine consumption stops. To compensate for this, the produced lactate is consumed as the carbon source along with glucose, serine and asparagine. After 3 days, glutamine is being produced in the system instead of being utilized as a carbon source. Amino acids like ammonia and glycine are not controlling factors towards the cellular objective and they oscillate between acting being produced and acting as a substrate. Asparate and glutamate are being produced by the system and do not act as nutrients for growth.

The model has not been able to capture the system dynamics of alanine. This might be a result of using only 3 flux modes in the solving strategy. To corrrectly represent biology of the system, more flux modes are required. The 3 flux modes that have been used might not be able to represent an alanine secreting route

towards the goal of the system. Hence incorporating these flux vectors into the ODE framework of Hybrid Cybernetic Modeling will not represent the true alanine dynamics of the system, since none of these flux vectors describes an alanine producing route. In order to improvise the model, more flux modes are needed to be calculated and incorporated in the solution strategy.

## 4.5 Discussions

This work demonstrates a mathematical formulation to describe *CHO* cell metabolism. The model simulates dynamics of metabolism and biosynthetic pathways in a fed-batch culture. The model calculates the concentration time profiles of extracellular metabolites in the reactor. The model successfully predicts the dynamics of metabolism and validates the experimental data derived from Nolan at al.,2011 [1]. The model represents rate expressions that are derived from enzyme kinetics. Michaelis-Menten saturation kinetics gives a form that takes care of reaction stoichiometries and enzyme regulation by pseudoenzymes.

The structure of the model equations is generic with respect to the type of mammalian cell , i.e. specificity for the *CHO* cell and fed-batch process results from the parameters, reaction network and stoichiometry used. In principle, application of the model to other parental *CHO* or mammalian cell lines, different culture processes, or extreme operating conditions would likely require re-estimation of at least some of the parameters.

The presented model accurately predicts the timing and magnitude of experimental metabolic profile. The model does rely on solutions obtained from

FBA and MCMC sampling. The pseudo steady state assumption supports experimental data but prevents detailed dynamic calculations within the cell. The current model does not account for fluctuations in the intracellular environment that is not driven by the extracellular metabolites. Hence this model will not predict the molecular mechanisms of pathway regulation involving the dynamics of signaling metabolites. The model presented here does a good job of predicting dynamics of the *CHO-K1* cell culture in a fed-batch reactor. Molecular regulations can be implemented by adding layers of sophistication to the model that account for the mechanism of signaling molecules.

The dynamic model described in this study does a good job at predicting the metabolic shift inside the fed-culture of *CHO* cells. Future work can take into account the mechanism of glycosylation for production of glycosylated proteins and monoclonal antibodies.

CHAPTER 5

# DRAWBACKS

Kinetic models rest on an explicit functional relationship connecting the rate of change of proteins and metabolites involved in the reaction. Michaelis-Menten kinetics is the most common choice when building a mathematical framework for these models (alternatives include generalized mass action, lin-log kinetics, or power-law models). But true mechanistic kinetic rate law for each specific reaction is unknown for most enzymes. There is a lack of reliable data for the enzyme activity and substrate affinity parameters used in these models. The effect of activators and inhibitors are typically unknown. Hence observations derived from in-vitro characterization may not hold true for in vivo conditions.

One potential drawback of the model is the reliance on flux balances. The assumption of pseudo-steady state, while supported experimentally, is still only an approximation, which ultimately prevents detailed calculations of dynamic changes within the cell. For example, the present model does not account for fluctuations in intracellular metabolite concentrations that are not reflected in the extracellular environment. Consequently, molecular mechanisms of pathway regulation, involving the accumulation or dissipation of signaling metabolites, cannot be directly explained or tested. In the present study, we found that this loss of detail did not compromise the ability of the model to simulate the long-term dynamics of a fed-batch culture. The inability to capture details of alanine dynamics in *CHO* metabolism is seen. It can be fixed by increasing the number of flux modes that are calculated by flux analysis techniques.

Bacterial cells compete fiercely to ensure the survival of their genes through numerous generations. Hence cybernetic objectives like the maximization of

growth rate or resource consumption hold validity for bacterial cultures. Mammalian cells do not have a clearly defined objective of survival like single-cellular organisms. Hence more thought can be given in this direction to derive a more suitable and biologically relevant system objective for mammalian cell cultures.

Variance based sensitivity analysis can be used to estimate and shortlist FBA modes that are crucial for model performance. Biomass yield on substrate can be used to define the performance function for sensitivity analysis.

# CHAPTER 6

## SUMMARY, CONCLUSIONS AND FUTURE WORK

This work represents the central carbon metabolism of *E. coli* and the genome scale metabolism of *CHO-K1*. *CHO-K1* was used as a subject of study due to special properties of mammalian cells to exhibit post translational modifications that can help derive protein therapeutics which can combat diseases like cancer in humans. This was a huge source of motivation towards pursuing the study of *CHO-K1* in the presented work. A similar genome scale study of metabolism can be extended for *E. coli* organism as well. The mathematical modeling techniques, principles and assumptions that have been applied for genome scale *CHO-K1* network can be extrapolated to a genome scale *E. coli* network as well. The modeling technique is flexible in nature and has general principles which is not specific to any organism. Specificity for modeling for a particular organism would result from the parameters, reaction network and stoichiometry used. The benefit of studying a genome scale *E. coli* network would be the vast availability of literature and experimental data that can help in deriving the required kinetic parameters for the cybernetic model and the various permutations and combinations of experimental designs that can be implemented and mathematically modeled for *E. coli*. Its also easier to control and monitor the growth environment and parameters manipulating a prokaryotic organism over a eukaryotic organism due to simpler biological machinery.

HCMs make the assumption that metabolic changes are driven by goals. Following this approach gives a concise way of modeling and understanding metabolism. HCMs capture biological process machinery, which means they represent an understanding of the generating process. The modeling assump-

tion of a goal directed metabolic system is not absurd since it can validate experimental results.

Mathematical models that capture dynamics of metabolism hold great applications in pharmacology and bioprocess engineering. HCMs provide simple kinetic framework to describe intricate regulatory mechanisms via specifying metabolic goal of the organism as objective functions. A good model can help understand the regulatory mechanisms. Perturbations to these metabolic systems can be induced to analyze the formation of metabolic products. The application area for this can be to study drug side-effects. Modeling the effects of drugs on cellular regulation will provide deeper insight into how drugs perturb the processes that control metabolism. This can help us gain insight of how drugs induce metabolic dysfunction. This allows in identification of better drug targets, treatment timing and dosage requirement. HCMs are formulated from the perspective that cells modulate metabolism to achieve goals. These goals are relevant to the organisms survival. When drugs perturb metabolic systems, they are interfering with the pre-programmed goal that drives metabolic regulation. There are number of interesting scenarios that can happen. The cellular regulation can be unaffected by drug perturbation. Cells can react to perturbation and still control metabolism towards the same goal. Some new metabolic objective function starts driving cellular regulation. Understanding cancer cells metabolism through model formulation can prevent the growth and evolution of the disease. The capacity of HCMs to revolutionize treatment in pharmacological applications has yet to be realized. The recommended research has the capability of impacting a range of important activities in the application of medicine. Systems biology has promising applications. Metabolic models can empower researchers to make sense of biological data and hypothesize connec-

tions between biomedical research and patient outcomes. The combination of metabolism and regulation are highly relevant to human health. HCMs offer an effective way to approximate the dynamics of both metabolites and regulation of metabolic reactions. New discoveries can be made using cybernetic models in the field of cancer. Modeling can help researchers trace time dependent changes in metabolites that affect disease progression.

The process of attachment of a glycan unit to a protein after translation is called glycosylation. Glycosylation is involved in malignant properties of cancer cells. It is commonly modeled separately from the rest of the cell metabolism. Peptide or protein can enter the Endoplasmic Reticulum (ER). Specific glycans are attached to the protein through amino residue of the asparagine inside ER. Additional glycosylation of the glycan part of the protein occurs in the ER. Two N-acetyl glucosamine and nine mannose residues are obtained at the end of this process. Next, the glycosylated protein enters the golgi apparatus (GA). Here, the glycan part of the protein is additionally modified by the help of specific enzymes. N-acetyl glucosamine, galactose, sialic acid, and fucose might be added to final glycan structures. The final products leave the GA after which the glycosylated proteins can be excreted from the cell, incorporated into membrane or transported to other places inside the cell. Cellular secretory capability plays an important role in the availability of glycosylation machinery. Modeling can be used to predict the distribution of glycoforms based on extracellular conditions and the form of glycan.

Protein-based drugs routinely display suboptimum therapeutic efficacies due to their poor physicochemical and pharmacological properties. Glycoengineering is a promising way to improve most parameters necessary for optimiza-

tion of protein drug in vivo efficacy (e.g., in vitro and in vivo molecular stability, pharmacodynamic responses, and pharmacokinetic profiles) while allowing for targeted drug delivery. In recent times, the global market for recombinant therapeutic proteins has gained popularity. Monoclonal antibodies (mAbs) offer novel therapy avenues for cancer, inflammatory diseases, infectious diseases, and autoimmune diseases and have been remarkably successful in the health market. Mammalian cells can produce therapeutic antibodies. *CHO* cells are extensively used for the production of recombinant antibodies as a result of their robust growth and the potential to produce non-immunogenic antibodies with glycosylation patterns similar to humans. The antibody production rate depends on factors such as recombinant gene expression level in the host cell. The gene expression level depends on the promoter strength and influences the rate of translation. The antibody production rate also depends on the availability of amino acid substrates, cofactors, inhibitor metabolites, and temperature. N-linked glycosylation plays a critical role in the biological properties of therapeutic immuno-globulin G (IgG). Understanding the interplay between cell growth, cell metabolism, IgG synthesis and glycosylation will contribute towards better bioprocess strategies and help in engineering cells towards maximum production of therapeutic proteins.

Glycosylation plays a crucial role in cancer mechanism and protein based therapeutics. Hence understanding the mechanism of glycosylation can enhance our knowledge of cancer and help build better protein-based therapeutics. The modeling framework can be designed as a continuous plug flow reactor (PFR) for representing maturation of glycoforms along the Golgi apparatus. Mass balance can be built around nucleotide sugar donors, byproducts, and transport proteins. A model based on glycosylation flux analysis (GFA) can be

proposed by applying constraint-based modeling of the glycosylation network by using a pseudo steady state assumption. GFA can be used to capture dynamical changes of glycoforms caused by media variations. A model combining metabolic pathways with glycosylation pathways can be valuable for estimating the quality of product titer.

Metabolism is a vastly complex process that is fundamental to life. Having a clearer picture of metabolic behavior enables a variety of societally important technologies ranging from pharmaceutical applications to biofuels. Our access to numerous omic technologies has increased our knowledge of biological machinery that drives metabolic processes. There is a need for models that give new insight into the volumes of biological data that are being generated by these technologies.

APPENDIX A

# APPENDIX I

The model parameters for batch fermentation of *E. coli* using HCM-MCMC are as follows

| Parameter | Value | Units | Source |
|---|---|---|---|
| $k_E$ | 0.41 | $hour^{-1}$ | Kim et al., 2008, [4] |
| $\alpha$ | 0.005 | $hour^{-1}$ | Kim et al., 2008, [4] |
| $\beta$ | 0.05 | $hour^{-1}$ | Kim et al., 2008, [4] |
| $K_{glucose}$ | 1 | mM | Estimated |
| $K_{acetate}$ | 1.5 | mM | Estimated |
| $k_{glucose,1}$ | 0.478 | $hour^{-1}$ | Estimated |
| $k_{glucose,2}$ | 0.382 | $hour^{-1}$ | Estimated |
| $k_{glucose,3}$ | 0.313 | $hour^{-1}$ | Estimated |
| $k_{glucose,4}$ | 0.257 | $hour^{-1}$ | Estimated |
| $k_{glucose,5}$ | 0.095 | $hour^{-1}$ | Estimated |
| $k_{glucose,6}$ | 0.093 | $hour^{-1}$ | Estimated |
| $k_{glucose,7}$ | 0.095 | $hour^{-1}$ | Estimated |
| $k_{glucose,8}$ | 0.087 | $hour^{-1}$ | Estimated |
| $k_{glucose,9}$ | 0.091 | $hour^{-1}$ | Estimated |
| $k_{glucose,10}$ | 0.086 | $hour^{-1}$ | Estimated |
| $k_{glucose,11}$ | 0.897 | $hour^{-1}$ | Estimated |
| $k_{glucose,12}$ | 0.090 | $hour^{-1}$ | Estimated |
| $k_{glucose,13}$ | 0.012 | $hour^{-1}$ | Estimated |
| $k_{glucose,14}$ | 0.011 | $hour^{-1}$ | Estimated |

| Parameter | Value | Units | Source |
|---|---|---|---|
| $k_{glucose,15}$ | 0.012 | $hour^{-1}$ | Estimated |
| $k_{glucose,16}$ | 0.011 | $hour^{-1}$ | Estimated |
| $k_{glucose,17}$ | 0.011 | $hour^{-1}$ | Estimated |
| $k_{glucose,18}$ | 0.011 | $hour^{-1}$ | Estimated |
| $k_{acetate,19}$ | 0.323 | $hour^{-1}$ | Estimated |
| $k_{acetate,20}$ | 0.346 | $hour^{-1}$ | Estimated |
| $k_{acetate,21}$ | 0.442 | $hour^{-1}$ | Estimated |
| $k_{acetate,22}$ | 0.007 | $hour^{-1}$ | Estimated |
| $k_{acetate,23}$ | 0.007 | $hour^{-1}$ | Estimated |
| $k_{acetate,24}$ | 0.008 | $hour^{-1}$ | Estimated |
| $k_{acetate,25}$ | 0.007 | $hour^{-1}$ | Estimated |
| $k_{acetate,26}$ | 0.006 | $hour^{-1}$ | Estimated |
| $k_{acetate,27}$ | 0.005 | $hour^{-1}$ | Estimated |
| $k_{acetate,28}$ | 0.005 | $hour^{-1}$ | Estimated |
| $k_{acetate,29}$ | 0.239 | $hour^{-1}$ | Estimated |

The model parameters for batch fermentation of *E. coli* using HCM-FBA are as follows

| Parameter | Value | Units | Source |
|---|---|---|---|
| $k_E$ | 0.41 | $hour^{-1}$ | Kim et al., 2008, [4] |
| $\alpha$ | 0.005 | $hour^{-1}$ | Kim et al., 2008, [4] |
| $\beta$ | 0.05 | $hour^{-1}$ | Kim et al., 2008, [4] |
| $K_{glucose}$ | 1 | mM | Estimated |
| $K_{acetate}$ | 1.5 | mM | Estimated |
| $k_{glucose,1}$ | 0.675 | $hour^{-1}$ | Estimated |
| $k_{glucose,2}$ | 0.599 | $hour^{-1}$ | Estimated |
| $k_{glucose,3}$ | 0.391 | $hour^{-1}$ | Estimated |
| $k_{glucose,4}$ | 0.286 | $hour^{-1}$ | Estimated |
| $k_{glucose,5}$ | 0.087 | $hour^{-1}$ | Estimated |
| $k_{glucose,6}$ | 0.088 | $hour^{-1}$ | Estimated |
| $k_{glucose,7}$ | 0.081 | $hour^{-1}$ | Estimated |
| $k_{glucose,8}$ | 0.083 | $hour^{-1}$ | Estimated |
| $k_{glucose,9}$ | 0.077 | $hour^{-1}$ | Estimated |
| $k_{glucose,10}$ | 0.085 | $hour^{-1}$ | Estimated |
| $k_{glucose,11}$ | 0.936 | $hour^{-1}$ | Estimated |
| $k_{glucose,12}$ | 0.081 | $hour^{-1}$ | Estimated |
| $k_{glucose,13}$ | 0.012 | $hour^{-1}$ | Estimated |
| $k_{glucose,14}$ | 0.012 | $hour^{-1}$ | Estimated |

| Parameter | Value | Units | Source |
|---|---|---|---|
| $k_{glucose,15}$ | 0.0101 | $hour^{-1}$ | Estimated |
| $k_{glucose,16}$ | 0.0161 | $hour^{-1}$ | Estimated |
| $k_{glucose,17}$ | 0.012 | $hour^{-1}$ | Estimated |
| $k_{glucose,18}$ | 0.011 | $hour^{-1}$ | Estimated |
| $k_{acetate,19}$ | 0.238 | $hour^{-1}$ | Estimated |
| $k_{acetate,20}$ | 0.296 | $hour^{-1}$ | Estimated |
| $k_{acetate,21}$ | 0.418 | $hour^{-1}$ | Estimated |
| $k_{acetate,22}$ | 0.002 | $hour^{-1}$ | Estimated |
| $k_{acetate,23}$ | 0.006 | $hour^{-1}$ | Estimated |
| $k_{acetate,24}$ | 0.008 | $hour^{-1}$ | Estimated |
| $k_{acetate,25}$ | 0.009 | $hour^{-1}$ | Estimated |
| $k_{acetate,26}$ | 0.007 | $hour^{-1}$ | Estimated |
| $k_{acetate,27}$ | 0.006 | $hour^{-1}$ | Estimated |
| $k_{acetate,28}$ | 0.007 | $hour^{-1}$ | Estimated |
| $k_{acetate,29}$ | 0.223 | $hour^{-1}$ | Estimated |

The model parameters for fed-batch fermentation of *CHO-K1* using HCM-FBA and HCM-MCMC are as follows

| Parameter | Value | Units | Source |
|---|---|---|---|
| $k_{E,1}$ | 1 | $day^{-1}$ | Estimated |
| $k_{E,2}$ | 10 | $day^{-1}$ | Estimated |
| $k_{E,3}$ | 10 | $day^{-1}$ | Estimated |
| $\beta$ | 0.25 | $day^{-1}$ | Calculated |
| $k_{M,1}$ | 0.528 | $day^{-1}$ | Estimated |
| $k_{M,2}$ | 0.288 | $day^{-1}$ | Estimated |

| Parameter | Value | Units | Source |
|---|---|---|---|
| $k_{M,3}$ | 0.72 | $day^{-1}$ | Estimated |
| $x_f$ | 9.5 | $mM$ | Estimated |
| $\alpha$ | 0.005 | $day^{-1}$ | Estimated |
| $K_{serine,1}$ | 1 | mM | Nolan et al., 2011, [1] |
| $K_{glucose,1}$ | 5 | mM | Nolan et al., 2011, [1] |
| $K_{asparagine,1}$ | 0.3 | mM | Nolan et al., 2011, [1] |
| $K_{glutamine,1}$ | 0.4 | mM | Nolan et al., 2011, [1] |
| $K_{serine,2}$ | 1 | mM | Nolan et al., 2011, [1] |
| $K_{glucose,2}$ | 20 | mM | Nolan et al., 2011, [1] |
| $K_{asparagine,2}$ | 0.3 | mM | Nolan et al., 2011, [1] |
| $K_{lactate,2}$ | 4 | mM | Nolan et al., 2011, [1] |
| $K_{glycine,2}$ | 3 | mM | Nolan et al., 2011, [1] |
| $K_{ammonia,2}$ | 0.6 | mM | Nolan et al., 2011, [1] |
| $K_{serine,3}$ | 1 | mM | Nolan et al., 2011, [1] |
| $K_{glucose,3}$ | 20 | mM | Nolan et al., 2011, [1] |
| $K_{asparagine,3}$ | 0.3 | mM | Nolan et al., 2011, [1] |
| $K_{lactate,3}$ | 4 | mM | Nolan et al., 2011, [1] |
| $K_{alanine,3}$ | 2.5 | mM | Nolan et al., 2011, [1] |
| $K_{asparate,3}$ | 1 | mM | Nolan et al., 2011, [1] |
| $K_{glycine,3}$ | 3 | mM | Nolan et al., 2011, [1] |
| $k_d$ | 0.03 | $day^{-1}$ | Estimated |

## BIBLIOGRAPHY

[1] Ryan P. Nolan and Kyongbum Lee. Dynamic model of *CHO* cell metabolism. *Metabolic Engineering*, Vol. 13, No. 1, pp. 108-124, 2011.

[2] Hooman Hefzi, Kok Siong Ang, Michael Hanscho, Aarash Bordbar, David Ruckerbauer, Meiyappan Lakshmanan, Camila A. Orellana, Deniz Baycin-Hizal, Yingxiang Huang, Daniel Ley, Veronica S. Martinez, Sarantos Kyriakopoulos, Natalia E. Jimnez, Daniel C. Zielinski, Lake-Ee Quek, Tune Wulff, Johnny Arnsdorf, Shangzhong Li, Jae Seong Lee, Giuseppe Paglia, Nicolas Loira, Philipp N. Spahn, Lasse E. Pedersen, Jahir M. Gutierrez, Zachary A. King, Anne Mathilde Lund, Harish Nagarajan, Alex Thomas, Alyaa M. Abdel-Haleem, Juergen Zanghellini, Helene F. Kildegaard, Bjrn G. Voldborg, Ziomara P. Gerdtzen, Michael J. Betenbaugh, Bernhard O. Palsson, Mikael R. Andersen, Lars K. Nielsen, Nicole Borth, Dong-Yup Lee and Nathan E. Lewis. A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems*, Vol. 3, No. 5, pp. 434-443, 2016.

[3] Amit Varma and Bernhard O. Palsson. Stoichiometric Flux Balance Models Quantitatively Predict Growth and Metabolic By-Product Secretion in Wild-Type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, Vol. 60, No. 10, pp. 3724-3731, 1994.

[4] Jin I. Kim, Jeffery D. Varner and Doraiswami Ramkrishna. A hybrid model of anaerobic *E. coli* GJT001: Combination of elementary flux modes and cybernetic variables. *Biotechnology Progress*, Vol. 24, No. 5, pp. 993-1006, 2008.

[5] Michael Vilkhovoy, Mason Minot and Jefferey D. Varner. Effective Dynamic Models of Metabolic Networks. *IEEE Life Sciences Letters*, Vol. 2, No. 4, pp. 2332-7685, 2016.

[6] Jefferey D. Varner and Doraiswami Ramkrishna. Metabolic Engineering from a Cybernetic Perspective. 1. Theoretical Preliminaries. *Biotechnology Progress*, Vol. 15, No. 3, pp. 407-425, 1999.

[7] A. Provost, G. Bastin, S. N. Agathos and Y.J. Schneider. Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. *Bioprocess Biosystem Engineering*, Vol. 29, No. 5-6, pp. 349-366, 2006.

[8] Jamey D. Young and Doraiswami Ramkrishna. On the Matching and Proportional Laws of Cybernetic Models. *Biotechnology Progress*, Vol. 23, No. 1, pp. 83-99, 2006.

[9] Jamey D. Young, Kristene L. Henne, John A. Morgan, Allan E. Konopka and Doraiswami Ramkrishna. Integrating Cybernetic Modeling With Pathway Analysis Provides a Dynamic, Systems-Level Description of Metabolic Control. *Biotechnology and Bioengineering*, Vol. 100, No. 3, pp. 542-559, 2008.

[10] Dhinakar S. Kompala, Doraiswami Ramkrishna, Norman B. Jansen and George T. Tsao. Investigation of Bacterial Growth on Mixed Substrates: Experimental Evaluation of Cybernetic Models. *Biotechnology and Bioengineering*, Vol. 28, No. 7, pp. 1044-1055, 1986.

[11] Cong T. Trinh, Aaron Wlaschin and Friedrich Srienc. Elementary Mode Analysis: A Useful Metabolic Pathway Analysis Tool for Characterizing Cellular Metabolism. *Applied Microbial Biotechnology*, Vol. 81, No. 5, pp. 813-826, 2009.

[12] Jurgen Zanghellini, David E. Ruckerbauer, Michael Hanscho and Christian Jungreuthmayer. Elementary flux modes in a nutshell: Properties, calculation and applications. *Biotechnology Journal*, Vol. 8, No. 9, pp. 1009-1016, 2013.

[13] Stefan Schuster and Claus Hilgetag. On Elementary Flux Modes In Biochemical Reaction Systems At Steady State. *Journal of Biological Systems*, Vol. 2, No. 2, pp. 165-182, 1994.

[14] Kuhn Ip, Caroline Colijn and Desmond S Lun. Analysis of complex metabolic behavior through pathway decomposition. *BMC Systems Biology*, Vol. 5, No. 91, pp. 1752-0509, 2011.

[15] Doraiswami Ramkrishna and Hyun-Seob Song. Dynamic Models of Metabolism: Review of the Cybernetic Approach. *American Institute of Chemical Engineers, Bioengineering, Food, And Natural Products*, Vol. 58, No. 4, pp. 986-997, 2012.

[16] Jamey D. Young. Learning from the Steersman: A Natural History of Cybernetic Models. *American Chemical Society, Industrial and Engineering Chemistry Research*, Vol. 54, No. 42, pp. 37235-1604, 2015.

[17] Ramprasad Ramakrishna and Doraiswami Ramkrishna. Cybernetic Modeling of Growth in Mixed, Substitutable Substrate Environments: Preferential and Simultaneous Utilization. *Biotechnology and Bioengineering*, Vol. 52, No. 1, pp. 141-151, 1996.

[18] Dhinakar S. Kompala, Doraiswami Ramkrishna and George T. Tsao. Cybernetic Modeling of Microbial Growth on Multiple Substrates. *Biotechnology and Bioengineering*, Vol. 26, No. 11, pp. 1272-1281, 1984.

[19] P. Dhurjati, D. Ramkrishna, M. C. Flickinger, and G. T. Tsao. A Cybernetic View of Microbial Growth: Modeling of Cells as Optimal Strategists. *Biotechnology and Bioengineering*, Vol. 27, No. 1, pp. 1-9, 1985.

[20] Doraiswami Ramkrishna. A Cybernetic Perspective of Microbial Growth. *American Chemical Society, Biochemical Engineering*, Vol. 83, pp. 0207-0161, 1983.

[21] A. Provost and G. Bastin. Dynamic metabolic modelling under the balanced growth condition. *Journal of Process Control*, Vol. 14, pp. 717-728, 2004.

[22] F. Zamoranoa, A.Vande Wouwera and G. Bastinb. A detailed metabolic flux analysis of an under-determined network of CHO cells. *Journal of Biotechnology*, Vol. 150, pp. 497-508, 2010.

[23] Philipp N.Spahn, Anders H.Hansen, Henning G.Hansen d, Johnny Arnsdorf, Helene F.Kildegaard and Nathan E.Lewis. A Markov chain model for N-linked protein glycosylation - towards a low-parameter tool for model-driven glycoengineering. *Metabolic Engineering*, Vol. 33, pp. 52-66, 2016.

[24] Julien Robitaille, Jingkui Chen and Mario Jolicoeur. A Single Dynamic Metabolic Model Can Describe mAb Producing CHO Cell Batch and Fed-Batch Cultures on Different Culture Media. *Plos One*, Vol. 10, No. 9, 2015.

[25] Han-Hsiu Hsu, Michihiro Araki, Masao Mochizuki, Yoshimi Hori, Masahiro Murata, Prihardi Kahar, Takanobu Yoshida, Tomohisa Hasunuma and Akihiko Kondo. A Systematic Approach to Time series Metabolite Profiling and RNA-seq Analysis of Chinese Hamster Ovary Cell Culture. *Nature, Scientific Reports*, Article No. 43518, 2017.

[26] Sofia Fernandes, Julien Robitaille, Georges Bastin, Mario Jolicoeur and Alain Vande Wouwer. Application of Dynamic Metabolic Flux Convex Analysis to CHO-DXB11 Cell Fed-batch Cultures. *International Federation of Automatic Control*, Vol. 49, No. 7, pp 466-471, 2016.

[27] Jan Schellenberger, Richard Que, Ronan M T Fleming, Ines Thiele, Jeffrey D Orth, Adam M Feist, Daniel C Zielinski, Aarash Bordbar, Nathan E Lewis, Sorena Rahmanian, Joseph Kang, Daniel R Hyduke and Bernhard Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature, Protocols*, Vol. 6, No. 9, 2011.

[28] Madhuresh Sumit, Sepideh Dolatshahi, An-Hsiang Adam Chu, Kaffa Cote, John J. Scarcelli, Jeffrey K. Marshall, Richard J. Cornell, Ron Weiss, Douglas A. Lauffenburger, Bhanu Chandra Mulukutla and Bruno Figueroa Jr. Dissecting N-Glycosylation Dynamics in Chinese Hamster Ovary Cells Fed-batch Cultures using Time Course Omics Analyses. *iScience, Cell Press*, Vol.12, pp 102-120, 2019.

[29] Roi Adadi, Benjamin Volkmer, Ron Milo, Matthias Heinemann and Tomer Shlomi. Prediction of Microbial Growth Rate versus Biomass Yield by a Metabolic Network with Kinetic Parameters. *Plos, Computational Biology*, Vol.8, No. 7, 2012.

[30] M. Yang and M. Butler. Effects of Ammonia on CHO Cell Growth, Erythropoietin Production, and Glycosylation. *Biotechnology and Bioengineering*, Vol.68, No. 4, 2000.

[31] Sjoerd Opdam, Anne Richelle, Benjamin Kellman, Shanzhong Li, Daniel C. Zielinski and Nathan E. Lewis. A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Systems*, Vol.4, pp 318-329, 2017.

[32] Nathan E Lewis, Xin Liu, Yuxiang Li, Harish Nagarajan, George Yerganian, Edward O'Brien, Aarash Bordbar1, Anne M Roth, Jeffrey Rosenbloom, Chao Bian, Min Xie, Wenbin Chen, Ning Li, Deniz Baycin-Hizal, Haythem Latif, Jochen Forster, Michael J Betenbaugh, Iman Famili, Xun Xu, Jun Wang and Bernhard O Palsson. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nature Biotechnology, Resource*, Vol.31, No. 8, 2013.

[33] Tiago M. Duarte, Nuno Carinhas, Laura C. Barreiro, Manuel J. T. Carrondo, Paula M. Alves, Ana P. Teixeira. Metabolic Responses of CHO Cells to Limitation of Key Amino Acids. *Biotechnology and Bioengineering*, Vol.111, No. 10, 2014.

[34] Mio-Sam Lao and Derek Toth. Effects of Ammonium and Lactate on Growth and Metabolism of a Recombinant Chinese Hamster Ovary Cell Culture. *American Chemical Society and American Institute of Chemical Engineers, Biotechnology Progess*, Vol.13, No. 5, 1997.

[35] Woo Suk Ahn and Maciek R.Antoniewicz. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. *Metabolic Engineering*, Vol.13, pp 598-609, 2011.

[36] Bhanu Chandra Mulukutla, Michael Gramer, Wei-Shou Hu. On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. *Metabolic Engineering*, Vol.14, pp 138-149, 2012.

[37] Woo Suk Ahn and Maciek R.Antoniewicz. On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. *Metabolic Engineering*, Vol.15, pp 34-47, 2013.

[38] Jin Hyoung Park, Jong Hwa Jin, Myung Sin Lim, Hyun Joo An, Jong Won Kim and Gyun Min Lee. Proteomic Analysis of Host Cell Protein Dynamics in the Culture Supernatants of Antibody-Producing CHO Cells. *Nature, Scientific Reports*, Vol.7, Article No. 44246, 2017.

[40] Sha Sha, Hemlata Bhatia, Seongkyu Yoon. An RNA-seq based transcriptomic investigation into the productivity and growth variants with Chinese hamster ovary cells. *Journal of Biotechnology*, Vol.271, pp 37-46, 2018.

[41] Woo Suk Ahn and Maciek R.Antoniewicz. Towards dynamic metabolic flux analysis in CHO cell cultures. *Biotechnology Journal*, Vol.7, pp 61-74, 2011.