COMBINED COMPUTATIONAL AND EXPERIMENTAL ANALYSIS TO CHARACTERIZE IMPACT OF BIOPROCESS CONDITIONS ON CHINESE HAMSTER OVARY CELL METABOLISM AND SITE-SPECIFIC N-LINKED GLYCOSYLATION OF AN IgG WITH DISTINCT Fab AND Fc GLYCANS

by

Jayanth Venkatarama Reddy

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical and Biomolecular Engineering

Fall 2024

COMBINED COMPUTATIONAL AND EXPERIMENTAL ANALYSIS TO CHARACTERIZE IMPACT OF BIOPROCESS CONDITIONS ON CHINESE HAMSTER OVARY CELL METABOLISM AND SITE-SPECIFIC N-LINKED GLYCOSYLATION OF AN IgG WITH DISTINCT Fab AND Fc GLYCANS

by

Jayanth Venkatarama Reddy

Approved: _____

Millicent O. Sullivan, Ph.D.
Chair of the Department of Chemical and Biomolecular Engineering

Approved: _____

Jamie D. Phillips, Ph.D.
Interim Dean of the College of Engineering

Approved: _____

Louis F. Rossi, Ph.D.
Vice Provost for Graduate and Professional Education and
Dean of the Graduate College

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Marianthi G. Ierapetritou, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Eleftherios T. Papoutsakis, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Abraham M. Lenhoff, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Kelvin H. Lee, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Michael J. Betenbaugh, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to extend my warmest thanks to my advisors, Prof. Marianthi Ierapetritou and Prof. Terry Papoutsakis, for making this work possible. The combination of complex modeling and rigorous experiments to study biological systems would not have been achievable without your support, mentorship, and encouragement. Thank you for the many sacrifices that contributed to creating such a wonderful lab environment; I truly felt like a part of both labs.

I am extremely grateful to Samik Das, Jessica Belliveau, Nikola Malinov, Katie Raudenbush, Will Thompson, and Oluwadare Badejo for their technical discussion, assistance with experiments, mentorship, and friendship. You all made my decision to join the lab an easy one.

I also owe a debt of gratitude to Prof. Kelvin Lee, Prof. Sumit Singh, Dr. Lauren Cordova, and Thomas Leibiger. Many parts of this dissertation would not have been possible without your help. Thank you for your support.

My passion for cell culture and metabolic engineering began in Prof. Michael Betenbaugh's lab. I would like to thank Mike, Harnish, and Swetha for their friendship, mentorship and for providing a strong foundation in cell culture engineering.

It is often said that graduate school can be a lonely experience, but I can gladly say that I disagree. My time in grad school was truly enjoyable, thanks to support of my dear friends. Special thanks to Sujay, Priyanka, Akanksha, Anantha, JohnD, Siddharth, Saketh, Archana, Soumitra, Esun, Aalap, and many others. I am grateful for your friendship.

Finally, I would like to express my gratitude for the love and support of my family —my parents, Venkatarama and Annapurna, and my brother Dileep. Thank you for providing me with a wonderful upbringing. There were numerous instances in graduate school that were easier to handle because of the social and technical skills you taught me over the decades. Any success I achieve is a testament to your immense support, and I appreciate your encouragement in my decision to pursue graduate school.

Additionally:

- I thank Prof. Kelvin Lee, Prof. Sumit Singh and Thomas Leibiger for analyzing the N-linked glycosylation samples that led to the results shown in Chapter 2 and Chapter 4. I would also like to thank Dr. Lauren Cordova for measuring the ammonia concentrations in my samples.
- This thesis is based on work funded by the US Food and Drug Administration under the grant numbers DHHS-FDA-R01FD006588 and U01FD007695A.
- The computational work in this thesis would not be possible without the support of the Caviness computing cluster.
- I would also like to thank Nikola Malinov for providing me with perfusion bioreactor experimental data used in Chapter 4.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Monoclonal antibodies (mAbs) have revolutionized the field of medicine over the last two decades. MAbs have demonstrated extraordinary ability to treat diseases such as cancer, autoimmune diseases, and viral infections. The 100th mAb was approved by the US FDA in 2021, with many more in clinical development. The increase in demand for mAbs, increase in number of mAbs in development, and the development of novel biotherapeutic modalities requires innovation in biomanufacturing for supply to keep up with demand. Biotherapeutic proteins are produced via biological processes that involve growing cells in bioreactors. Chinese Hamster Ovary (CHO) cells have emerged as the leading platform for producing glycosylated therapeutics. Development of these processes has traditionally relied on experimentation to determine optimal operating conditions. Over the last several years, mathematical models have demonstrated capabilities to aid with biotherapeutic process development. However, many of these mathematical models do not consider the effect of bioreactor parameters such as pH, dissolved oxygen, and temperature. The major focus of this thesis is to utilize mathematical models to elucidate the effect of these process parameters on CHO cell metabolism and N-linked glycosylation. This mechanistic understanding is further used to develop predictive mathematical models that have various applications to aid with process intensification.

Chapter 2 involves cultivating CHO cells at various bioreactor pH conditions and measuring the concentrations of viable cells, glucose, lactate, amino acids, mAbs, and ammonia. These measurements were utilized to perform flux balance analysis. The

resulting fluxes calculated from flux balance analysis provided mechanistic insights into the effect of bioreactor pH on CHO cell metabolism. The mAb used in this study contains glycosylation sites in the Fab and Fc regions. The effect of any process parameter on site specific N-linked glycosylation has not been reported. Hence, we decided to utilize glycopeptide mapping to elucidate the effect of bioreactor pH on Fab and Fc glycosylation. Bioreactor pH had a complex effect on glycosylation. Certain glycan fractions were impacted at one site but not the other.

Several mathematical models for glycosylation have been developed over the last two decades. However, none of these models can model site-specific N-linked glycosylation differences that can fundamentally arise from the glycosylation process. The work performed in chapter 3, led to the development of a site-specific N-linked glycosylation model. This model was used to elucidate the effect of bioreactor pH on Fab and Fc glycosylation. The model was also subsequently used to predict the effect of galactose and $MnCl_2$ supplementation on Fab and Fc glycosylation. These predictions were validated by performing fed-batch experiments in shake flasks.

Chapter 4 involved developing a mathematical model that can predict concentrations of viable cells, glucose, lactate, ammonia, amino acids, and mAb if provided with the bioreactor pH, seeding cell density, media composition, and feed addition time. A complex dynamic metabolic flux analysis model was constructed to achieve this aim. Chapter 4 also involves demonstrating the application of this model to predict intensified fed-batch and perfusion bioreactor performance by using a model trained only on fed-batch data. These predictions were successfully validated. This thesis involves utilizing computational and experimental approaches to develop tools for bioprocess development.

Chapter 1

INTRODUCTION

1.1 Overview of biopharmaceutical market

The last thirty some years have seen an explosive growth in the development of protein biotherapeutics, and notably of monoclonal antibodies (mAbs), to treat various diseases including a multitude of cancers and autoimmune diseases. Therapeutic antibodies have revolutionized the practice of medicine [1]. In 2021, a milestone was reached as the 100th therapeutic was approved by the FDA, with nearly 870 more in clinical development [2]. According to a market analysis on biopharmaceuticals, sales of biopharmaceuticals are expected to grow from 330.7 billion USD in 2021 to 478 billion USD in 2026 [3].

Despite the evident growth in the mAb market, these biologic products are expensive. The mAb drug price per quality adjusted life year (QALY) gained by a patient has been shown to increase with drug approval data [4]. To ensure cost effectiveness of new therapeutics, a "willingness to pay" threshold per QALY can be assessed to address value-based health decisions. For example, a typical threshold of $100,000/QALY was applied in a study comparing two PD-L1 inhibitors for the treatment of non-small cell lung cancer. Pembrolizumab was shown to be cost effective at $98,421/QALY, but this was not the case for a competitor, nivolumab, which had a ratio of $117,857/QALY despite a larger increase in QALY compared to pembrolizumab [5]. For both the benefit of humanity and the ability to succeed in a market which is increasingly moving towards value-based health decisions,

development of biosimilars may drive down market costs of these biotherapeutics. When novel drugs come off patent, more competitive pricing becomes available as biosimilar products can be offered at significantly lower prices (up to 87% lower) as there are far less research and development costs to recoup [6]. After loss of patent protection, market demand may shift to the most affordable biosimilar. Cutting production costs enables affordability in the competitive market of biosimilars. The growth of the mAb market, the large cost of biologics and the development of biosimilars has led to investigation of biomanufacturing methods that can manufacture the product at a lower cost.

1.2 Motivation to utilize mathematical models for bioprocess development

Optimization of the biomanufacturing process requires efforts in cell line development, media development, process monitoring and control, and determining optimal bioreactor operating conditions while meeting critical quality attributes (CQAs) such as those pertaining to N-linked glycosylation, mAb aggregation, and mAb fragmentation [7]. Quality by Design (QbD) guidelines set by the FDA emphasize the need to establish the relationship between process parameters and CQAs when designing a biomanufacturing process [8]. The classical method for implementing quality by design is through extensive experimental testing in a statistical design of experiments (DoE). These statistical DoEs can be very time consuming and expensive, scaling exponentially with the number of process parameters. The use of mechanistic models can reduce the total size of the design of experiments by maximizing the information produced in a set of experiments, iteratively improving the process understanding and mechanistic model as the DoE proceeds [9, 10]. For this improvement to QbD practices, the impact of operating conditions on cellular

metabolism and quality attributes must be captured in the mechanistic models. This chapter explores the latest advances in the field of modeling mammalian cellular metabolism and N-linked glycosylation to improve biomanufacturing, thus providing a summary of existing mathematical models for bioprocesses, their applications, advantages, and drawbacks.

Chinese hamster ovary (CHO) cells are the primary platform used for production of monoclonal antibodies. Nearly two-thirds of the approved therapeutic monoclonal antibodies are produced using CHO cells. Mammalian cells such as human embryonic kidney (HEK293), baby hamster kidney (BHK), and murine myeloma cells (NS0 and SP2/ 0) are also used for manufacturing of biologics [11]. As CHO cells are the most widely used cells in biomanufacturing, most of the modeling literature has focused on modeling CHO cell-based processes.

The majority of published models do not consider the impact of bioreactor parameters on modeling cell metabolism or N-linked glycosylation. Yet, parameters such as pH, dissolved oxygen (DO), and bioreactor heterogeneities have a significant impact on the metabolism and glycosylation of mammalian cells [12]. Developing and utilizing models that provide mechanistic insights into the effect of such bioreactor parameters constitute an essential need. We thus review the limited relevant literature in order to identify and detail future directions. The capabilities of different types of models to optimize various aspects of biomanufacturing are dissected and assessed.

1.3 Models for cellular metabolism

Models for cellular metabolism can be grouped in several categories using different key attributes. A first logical categorization is in structured versus unstructured models [13]. Structured models include biological processes or reactions that occur

within a cell, which enables more detailed and nuanced descriptions of metabolism. Such models can be difficult to validate and use, as not all intracellular reactions can be possibly accounted for, and thus, judicious simplifications must be made in order to facilitate their use. Also, for validation purposes it is quite tedious and difficult to experimentally measure a comprehensive set of intracellular reaction rates, particularly in different compartments inside the cell. On the other hand, unstructured models describe cell metabolism without the consideration of intracellular reactions. Instead, the cell is treated as a black box, with its inputs and outputs defined by extracellular macroscopic reactions. Another way to categorize models for cell metabolism is whether they are segregated or unsegregated [13]. Instead of considering all cells in terms of their average behavior like unsegregated models, segregated models describe cell behavior by accounting for the presence of a heterogeneous population of cells based on the distinct cellular states observed. The overall cell culture kinetics are more accurately captured by relating the kinetics to distinct parts of the cell culture population. However, they require proportionally more experimentation for the number of segregated phases captured. Thus, most models for cellular metabolism are unsegregated and unstructured.

Models for cellular metabolism can also be classified into kinetic, stoichiometric, and integrated. Kinetic models consist of semi-empirical parametric differential equations that are formulated based on knowledge of biological systems. Stoichiometric models are non-parametric models that consist of a reaction network of intracellular reactions. A schematic of the core reaction network for mAb producing mammalian cell metabolism is shown in Figure 1.1. Integrated models utilize components from both kinetic and stoichiometric models to utilize the advantages of

both types of models. The following sections discuss the development of each of these types of models to describe metabolism of mammalian cells. The application of these models for process simulation and optimization is discussed in Section 1.7.

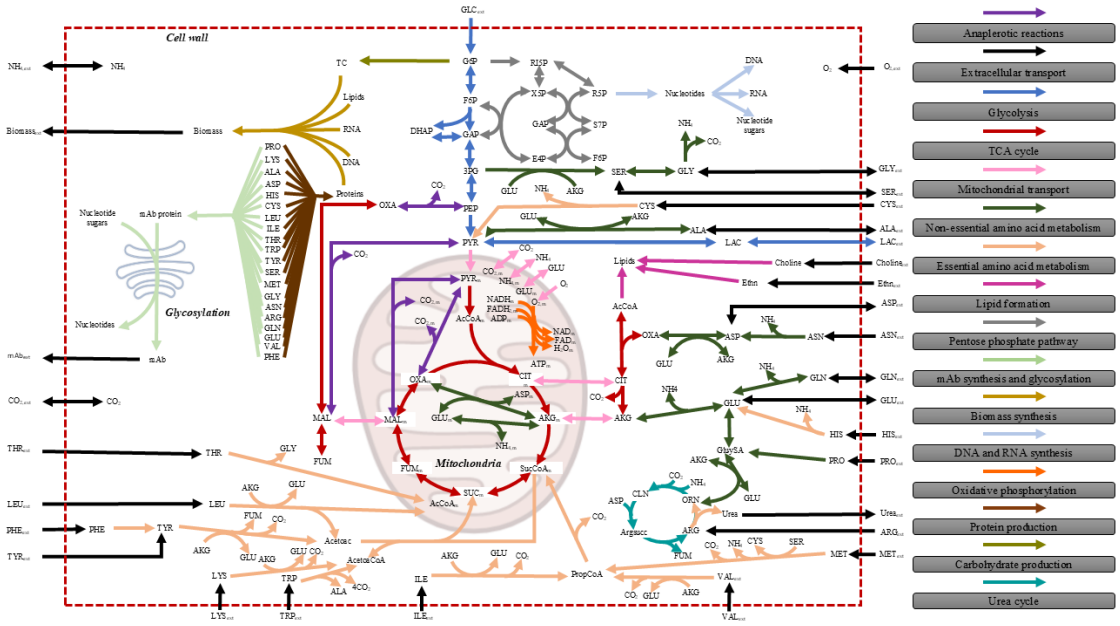

Figure 1.1: Schematic for core cellular metabolism of mammalian cells used to produce mAbs. The reaction network contains amino acid metabolism, TCA cycle, oxidative phosphorylation, urea cycle, anaplerotic reactions, biomass synthesis, antibody synthesis, antibody glycosylation and glycolysis. Detailed reaction networks can be accessed via genome scale models discussed in Section 1.3.2. Created with BioRender.com.

1.3.1 Kinetic models

Kinetic models of biological systems can be used to describe the dynamics of the biological process by typically using a formalism based on differential equations. These differential equations model the rate of change of concentration of metabolites as a function of the cell density, concentrations of substrates and concentrations of

inhibitors. These differential equations are usually formulated based on expert reasoning, statistical methods, and metabolic network analysis. The expressions of these differential equations contain parameters that are determined by regressing the model to a large number of experimental data [14]. These types of models provide a platform for describing concentrations of metabolites as a function of time. The inputs, model schematic and model output are shown in Figure 1.2. These types of models have existed for many decades and there have been various review papers and textbooks over the decades that have described how to develop these kinetic models for microbial and mammalian systems [13-15]. There are many publications and textbooks that derive the rate expression for various reaction mechanisms [16-18]. These publications can be used as a guide to formulate the rate expression for models, provided the reaction mechanism is known. The models are quite complex and various fitting strategies are used to determine an optimal set of parameters. These strategies are discussed in Section 1.8.

The size of a kinetic model can vary drastically depending on the application of the model. There are various formulations of kinetic models in the literature. For example, an unstructured and unsegregated kinetic model of CHO cells [19] used Monod type equation to describe the effect of glucose on growth rate, yield coefficients to determine the glucose uptake rate and Luedeking-Piret type equation to describe mAb production rate was formulated to correlate viable cell density, glucose concentration and mAb production [19]. This model was relatively simple and used 3 differential equations and 6 parameters. The use of Luedeking-Piret type equation allowed for relating the mAb production rate to the growth rate as well as the number of cells [19].

To make these models more applicable for the industrial bioprocess the kinetic modeling framework has been developed over the years to consider more metabolites, some even capturing intracellular details [20]. For example, an unstructured and unsegregated kinetic model [20] on NS0 (murine myeloma) cells containing 9 differential equations and 18 parameters was used to describe the dynamic profiles of cell density, viability, glucose concentration, lactate concentration, ammonia concentration, glutamic acid concentration, glutamine concentration and mAb titers. The ability to describe the dynamic profiles of more metabolites helped the authors of this study to use this model to optimize the feed amount and feed time of cell culture media in the fed-batch process and increase mAb titers by 30% [20].

In another study [21], a structured kinetic model containing 34 kinetic expression, and 139 parameters was used to describe the rate of change of concentrations of amino acids, antibody, glucose, lactate, ammonia and a few intracellular metabolites in CHO cells. This kinetic model was used to compare the effects of two different culture media on batch and fed-batch cultures [21]. Literature studies have also made the distinction between normal growing, normal resting and apoptotic CHO cells to develop a segregated and structured kinetic model [22]. The authors of this study used fluorescence microscopy to distinguish between normal viable, apoptotic and necrotic CHO cells. Relating the mAb production to both normal cells and apoptotic cells provided a better fit to mAb production data as age related variations in cell population occurred [22].

Kinetic models have also been shown to predict intracellular nucleotide sugar concentrations, such as UDP-Gal with respect to the concentration of extracellular metabolites, such as galactose concentration. Nucleotide sugars are important

precursors to the glycosylation process. Low intracellular levels of UDP-Gal have been shown to reduce the galactosylation index of mAbs [23]. These kinetic models can be used to determine the optimal galactose concentrations required to produce more intracellular UDP-Gal to improve the galactosylation index. These kinetic models can make predictions on the concentration of metabolites, products, and cell density as a function of time and concentration of metabolites. These predictions can help determine feeding strategies for fed-batch processes as addition of feed (spike in concentration of nutrients) can have a positive or negative effect on the cells depending on time of addition. The application of these models to optimize bioprocesses and the application of this framework within other models is described in Section 1.7.

Kinetic models do have limitations. As kinetic models capture more metabolites and reactions, they contain more parameters which require more experimentation to determine. There are thousands of reactions in a CHO cell, so a complete kinetic model of CHO cells is infeasible. Also, it is often difficult to determine the relationship between input metabolites and products. Capturing all will lead to complex differential equations with more parameters. Alternatively, linking the consumption of a single metabolite to cell density or to the production of another metabolite leads to a simplified model that cannot be used to predict the concentration of the metabolite under different media conditions. For example, amino acid consumption can lead to ammonia synthesis. Unbalanced amino acid concentrations in cell culture media can lead to high ammonia production. Linking ammonia synthesis to only cell density without considering the effect of the other amino acids leads to incorrect predictions under different media conditions.

A comprehensive review paper on the various types of kinetic expressions used in kinetic models along with a detailed description on formulation of kinetic models has been published in the literature [24]. The major challenges of developing and utilizing kinetic models are scarcity in broad data sets, lack of mechanistic understanding of complex regulatory events, large number of parameters and high dimensionality. In a review paper, Saa and Nielsen also discuss these limitations and the future outlook of kinetic models for bioprocess development [24].

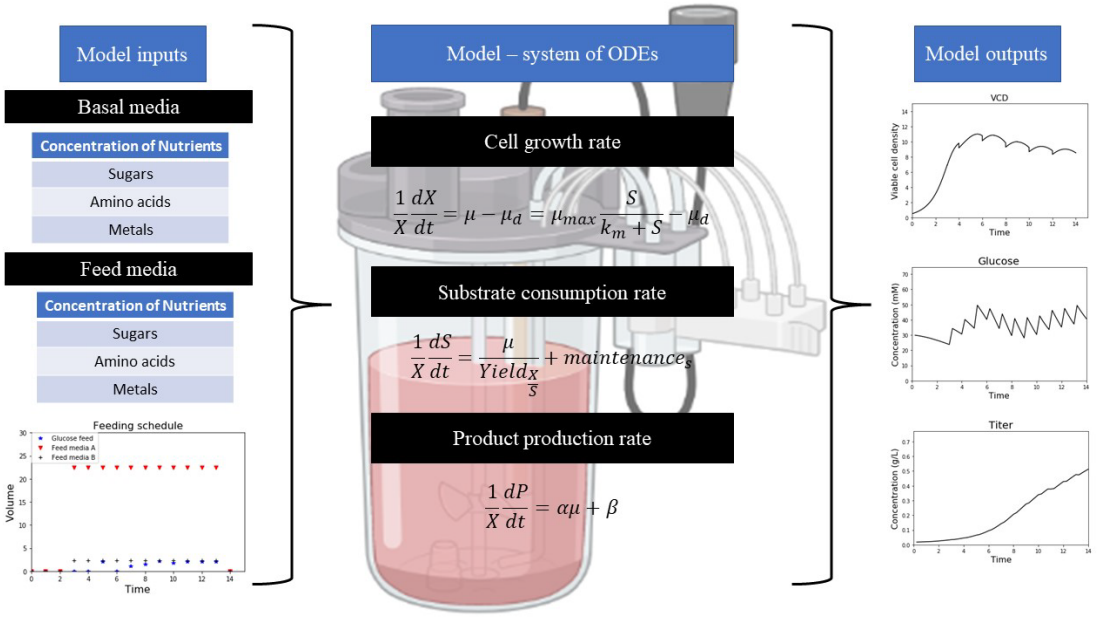

Figure 1.2: Schematic for kinetic model representation of cellular metabolism. Kinetic models utilize media nutrient concentrations and feed addition times to predict dynamic metabolic profiles by solving a system of ODEs that relate the cell density (X), product concentration (P) and substrate concentration (S). The cell growth rate and death rates are represented by μ and μ_d respectively. The yield of cells from substrate is represented by $Y_{\frac{X}{S}}$. The specific product production rate is modeled using Luedeking-Piret equation. Luedeking-Piret equation relates the specific product production rate to the growth rate by the constant α and also has a non-growth associated term β. Created with BioRender.com.

1.3.2 Stoichiometric models

Macroscopic reactions are used to develop kinetic models. These macroscopic reactions are determined by linking an input metabolite to an output metabolite by looking at intracellular reactions and collapsing them. On the other hand, stoichiometric models are built around a set of intracellular reactions that form a metabolic network. The number of intracellular reactions in these models vary across different models. This network is usually reconstructed based on metabolic pathway data that is available from the KEGG database or the BioCyc database [25, 26]. Stoichiometric models rely on a mass balance of metabolites in the reaction network. The rate of change of intracellular metabolite concentrations is equal to the net rate of reactions involving the metabolite of interest. According to the pseudo-steady state assumption, the rate of change of intracellular metabolite concentrations is negligible compared to the values of the reaction fluxes. An order of magnitude analysis was performed in the literature to show the validity of the pseudo-steady state assumption in mammalian cells by comparing the values of fluxes to the rate of change of intracellular metabolite concentrations [27]. The process of development of stoichiometric models for microbial and mammalian systems has also been widely reviewed in the literature. Depending on the size of the reaction network and the number of measurements, the stoichiometric model can be classified into an underdetermined system or an overdetermined system [28]. Underdetermined systems have more unknown fluxes than can be determined by the measured fluxes. Underdetermined systems can have infinite sets of solutions. Solutions to these types of systems can be determined by using optimization techniques. A specific reaction or a set of reactions have to be chosen as an objective function. The solution to all the fluxes is given based on minimizing or maximizing the objective function. This technique of solving the underdetermined system by maximizing or minimizing the objective

function is called flux balance analysis (FBA). A summary of many different objective functions can be found in the literature [29]. Overdetermined systems have more measured variables than unknown variables. These systems are solved by using metabolic flux analysis (MFA). MFA performs a least square regression to determine the solution that best describes all the measurements. The differences between FBA and MFA have also been reviewed in the literature [30].

Stoichiometric models are applicable due to the existence of metabolic steady states. Metabolic steady state is an approximation that the intracellular fluxes do not change with time. The existence of metabolic steady state is usually approximated for distinct cell culture phases or continuous cultures with fixed perfusion rates and cell densities [31]. A simple method to characterize a metabolic steady state in batch culture with an exponential growth rate without using intracellular measurements has been described in the literature [32]. Constant yield coefficients indicate the presence of a metabolic steady state. This method checks for a linear relationship between extracellular concentrations of various metabolites at different time points to identify constant yield coefficients and verify the presence of a metabolic steady state. [32]. The application of metabolic steady state to understand cellular metabolism is reviewed in the following subsections. A schematic of model inputs, model representation, solution method, and model outputs is shown in Figure 1.3.

1.3.2.1 Subgenomic structured models and ^{13}C-based flux analysis

^{13}C-MFA utilizes isotope labeling data and metabolic exchange rates to provide constraints to an underdetermined stoichiometric model. Using only extracellular flux measurements to determine intracellular metabolism has drawbacks. Intracellular reactions have many branches and cyclic pathways. Fluxes in cyclic pathways such as

the pentose phosphate pathways cannot be estimated using only extracellular rates. Isotopic tracers can be used to provide additional constraints based on the labeling pattern of intracellular metabolites to determine these fluxes [33]. ^{13}C-MFA involves feeding cells carbon-13 labeled isotope substrates and measuring the rates of uptake of nutrients and isotope labeling of intracellular metabolites. The model quantifies the intracellular fluxes that can be used to understand intracellular metabolism under different growth conditions and stages. A detailed guide to performing ^{13}C-MFA in mammalian cells has been published in the literature [34].

^{13}C-MFA methods can be divided into stationary and non-stationary ^{13}C-MFA. Stationary ^{13}C-MFA requires isotopic steady state along with the metabolic steady state approximation. Isotopic steady state is achieved when the intracellular isotopic labeling fractions do not change with time. To perform stationary ^{13}C-MFA a single sample can be taken after isotopic steady state has been reached [31]. Stationary ^{13}C-MFA has been applied to study the intracellular metabolism of CHO cells in various studies. The isotopic labeling of metabolites in glycolysis reached steady state in 1.5 hours when [1,2-^{13}C]glucose was used as a substrate. However, the metabolites of the TCA cycle did not reach isotopic steady state in the same time frame with the substrate [1,2-^{13}C] glucose. TCA cycle metabolites reached isotopic steady state in 3 hours when [U-^{13}C]glutamine was fed as a substrate. Hence, parallel labeling studies using ^{13}C isotope labelled glucose and glutamine were performed on CHO cells to study differences in intracellular fluxes during the exponential growth phase and stationary phase. The quantification of intracellular fluxes in this study revealed that moving from the growth phase to the stationary phase resulted in a slight reduction in TCA cycle fluxes and increase in flux through the oxidative pentose phosphate pathway [35].

Parallel labeling experiments were carried out in another study to determine the contribution of glucose, glutamine and asparagine towards TCA cycle fluxes in CHO cells during early and late exponential growth phase at low and high glutamine concentrations. This study performed parallel labeling experiments using ^{13}C labelled glucose, glutamine and asparagine to show that glucose was the primary source for TCA cycle fluxes during both early and late exponential phase. Glutamine contribution towards TCA cycle fluxes was higher during early exponential phase and high glutamine concentration and asparagine contribution towards TCA cycle fluxes was higher at the early exponential phase with low glutamine concentrations. Aspartate was preferred over glutamine and asparagine during late exponential phase [36].

Non-stationary ^{13}C-MFA does not require isotopic steady state and can be performed at shorter time scales than stationary ^{13}C-MFA. Intracellular isotope labeling fractions at different time points must be measured to obtain dynamic ^{13}C labeling data that is used to perform non-stationary ^{13}C-MFA [31]. Non-stationary ^{13}C-MFA has been performed on CHO cell cultures using a single tracer [U-^{13}C] glucose and labeling patterns of only extracellular metabolites. In addition to quantifying the intracellular metabolic fluxes, the reversibility of transport reactions for many metabolites (pyruvate, lactate, and non-essential amino acids) was also studied. Non-essential amino acid metabolism such as production of intracellular serine from glycolysis and its subsequent conversion to pyruvate, glycine or biomass was quantified.

The capability to determine intracellular metabolic fluxes in a compartmentalized network via measurements of only extracellular isotope labeling presents a method that does not require extraction of intracellular metabolites [37]. The ^{13}C-MFA studies reviewed in this section used compartmentalized networks to study

pyruvate transport into mitochondria, anaplerotic reaction fluxes and amino acid metabolism. This enabled the study of different compartmental metabolic pools of various metabolites [35, 37]. Application of these models provides insights into intracellular metabolism. The application of ^{13}C-MFA to improve understanding of metabolism of mAb producing CHO cells, study the effect of media conditions on CHO cell metabolism and identify bottlenecks in pathways that can help engineer better cell clones has been reviewed in the literature [38], hence improving fundamental understanding of cellular metabolism. There are numerous studies that use ^{13}C-MFA to study the metabolism of mammalian cells. However, there is a significant lack of literature on utilizing these tools to provide insights into the effect of critical bioreactor parameters such as temperature, pH and dissolved oxygen on cellular metabolism. This could be attributed to the fact that ^{13}C tracers are expensive and the tracer experiments are most often performed in small scale shake flasks rather than bioreactors to avoid large costs. Hopefully the widespread utilization of miniaturized high throughput bioreactor systems will provide a platform that can help perform these experiments on bioreactors to study the effect of parameters relevant to bioreactor operation. Software such as METRAN, OpenFlux, and INCA provide platforms for ^{13}C-MFA model development and flux estimation [39-41].

1.3.2.2 Genome scale models (GSMs)

Technological developments in the last 10-15 years have dramatically enhanced the ability to collect and process genomic data. Identification of the genes and reactions in a microorganism has allowed for the development of large stoichiometric models that contain thousands of reactions. These models are called genome scale models. The genome scale models for different CHO cell lines have been published [42]. Due to the

large number of reactions in these models they are typical examples of underdetermined systems. The macromolecules involved in CHO cell biomass formation include proteins, lipids, DNA, RNA, and carbohydrates. The composition of these macromolecules, the fractions of different amino acids in the proteins and the types of lipids present vary for different cell lines, which presents a challenge in utilizing an accurate biomass formation reaction in the CHO genome scale model for a specific cell line. This challenge was overcome by experimentally measuring the macromolecular composition of the cells, amino acid composition in proteins and the lipid composition for 13 different parental and producer CHO cell lines under various experimental conditions. This was included in the CHO genome scale model biomass reaction to improve the accuracy of predictions. However, the study reported that the model predictions were more significantly influenced by the quantification of essential amino acid content in biomass and the exchange rates than changes in biomass composition [43].

The effect of errors in experimental measurements that are used as constraints for genome scale models has been studied in the literature. Small errors in measurements of essential amino acid fluxes led to a large difference in growth rate predictions. Hence, it is important to take accurate measurements of essential amino acids with low uptake rates at a high sampling frequency when using genome scale models [44, 45]. Another study cultured 10 CHO clones with various metabolic phenotypes and used transcriptomic data to curate a cell line specific metabolic model. This model was used to study the effect of model parameters such as biomass weight, biomass composition, cell death rate, protein secretion costs, biomass turnover rate, amino acid catabolism byproduct synthesis, and different objective functions on model predictions. The study

showed that the most influential parameter was the choice of objective function [46, 47]. The predictions from the genome scale model for CHO cell metabolism were improved by using an uptake rate based objective function in addition to the constraining the growth rate to the experimentally observed value. This approach independently minimizes the uptake rate of each individual non-essential nutrient to provide better predictions of the amino acid uptake rates rather than using the conventional biomass objective function [48]. This approach has further been improved to aid with CHO cell fed-batch culture. The measurement of growth rate was used to determine the concentration of amino acids in cell culture media at any point of a fed-batch process via model-based predictions. Hence, this model provided a platform for monitoring of amino acid concentration in a fed-batch culture without regular measurements. This was used to design feeding of essential amino acids in a fed-batch process [49].

The TCA cycle fluxes are most often under predicted due to the absence of accounting for ATP demand for cellular processes such as protein folding, macromolecular regeneration and redox balance maintenance. Estimation and addition of the maintenance energy requirements have led to improvements in flux predictions of the CHO genome scale model. These calculations were carried out using the genome scale model with the growth rate as the objective function and a constraint on demand reaction for ATP at different flux values. The constraints of the demand reaction for ATP were adjusted so that the TCA fluxes in the genome scale model FBA solution agrees with experimentally measured ^{13}C MFA data from the literature [44, 45].

The addition of enzyme related constraints for flux balance analysis can constrain the underdetermined system to yield biologically meaningful solutions. This

has been done by adding enzyme turnover numbers to improve the accuracy of intracellular flux predictions. This method led to an improvement in the prediction of TCA cycle fluxes compared to FBA without these additional constraints. The constraints on the enzyme turnover numbers also led to improved predictions of lactate secretion and growth rate as a function of increasing glucose uptake rates [50].

FBA was performed on a reaction network that contains 154 reactions and 120 metabolites to study the effect of cell culture temperature on nucleotide sugar production rates in CHO cells. The study revealed that reduction in culture temperature led to a decrease in nucleotide sugar production fluxes and hence affected the N-linked glycosylation [51]. Stoichiometric models are very good at modeling the uptake and production rates of a large number of metabolites. These models that are non-parametric provide mechanistic insights into relationships between a number of metabolites. Hence, there is no need for a large amount of experimental data to use these models. The disadvantage of using these models is that it is not easy to carefully curate these models. Curation of genome scale metabolic models requires constructing a draft model based on the genome annotation of the organism and subsequent manual revision of all gene and reaction components. The model is then analyzed for inconsistencies and gap filling [52].

Genome scale models have to be modified to include the latest results from multi-omics studies that reveal new metabolites, genes, and reactions which can be added to the reaction network. Analysis of inconsistencies and gaps in the CHO genome scale model showed the presence of many blocked reactions and dead-end metabolites. The application of computational gap filling methods such as GapFind/GapFill [53] and GAUGE [54] led to identification of a minimal subset of reactions that can resolve the

inconsistencies. The remaining blocked reactions were manually assessed based on data from the Human Metabolome Database, BiGG database and KEGG [25, 55, 56]. This effort led to the addition of 773 new reactions and 335 new genes to the model. Considering these new reactions and genes in the genome scale model resulted in a reduction of blocked reactions from 21.6 % to 11.3% and the percentage of dead-end metabolites dropped from 21.4 % to 6.6 % [57].

A drawback of using stoichiometric models is that these models do not directly provide information about dynamic metabolic profiles. They can only be used to provide information at a given snapshot of the metabolism of the cell. It is desirable to use the minimum number of reactions to describe the phenotype of the cell to minimize the complexity of the model. Algorithms such as lumpGEM and redGEM have been developed to reduce genome scale models into relatively small-scale models that are much easier to use for the development of models for MFA or process analytics [58-60]. Elementary flux modes (EFMs) can be used to determine a set of macroscopic reactions from stoichiometric models to describe the phenotypes observed. The theory behind utilizing EFMs to determine the macroscopic reactions along with two case studies to highlight the complexity of EFMs with increasing metabolic network size has been published [61]. Reaction networks of different sizes can be used based on the needs of the model applications. A list of models of CHO cells of various reaction network sizes has been tabulated in Table 1.1.

Table 1.1: Stoichiometric models for mammalian cell metabolism.

Number	Model details	Method used	Application	Reference
1	34 reactions and 24 metabolites	MFA	This network is part of a dynamic metabolic flux analysis model that predicts transient metabolic profiles for CHO cells.	[62]
2	272 reactions and 228 metabolites	MFA	Modifications were made to the 272 reactions to perform MFA on steady-state data obtained for hybridoma and CHO cell cultures.	[63]
3	79 reactions and 52 metabolites	^{13}C-MFA	Parallel labeling experiments were used to study CHO intracellular metabolism in exponential growth phase and early stationary phase.	[35]
4	68 reactions and 58 metabolites	^{13}C-MFA	^{13}C-MFA was used to study intracellular fluxes in mAb producing and non-producing CHO cells.	[64]
5	34 reactions and 30 metabolites	MFA	CHO cell metabolism was studied in growth and non-growth phases of cell culture using MFA.	[22]
6	6663 reactions and 2341 metabolites	FBA	Genome scale models for different CHO cell lines (CHO-K1, CHO-S and CHO-DG44) was developed.	[42]
7	154 reactions and 120 metabolites	FBA	FBA was used to study the effect of temperature and phase of CHO cell cultures on intracellular nucleotide sugar production rates.	[51]
8	85 reactions and 49 metabolites	^{13}C-MFA	^{13}C-MFA was used to study the metabolism of glucose, glutamine and asparagine in CHO cell cultures at various culture phases.	[36]
9	60 reactions and 25 metabolites	^{13}C-MFA	Non-stationary ^{13}C-MFA was performed on CHO cell cultures by using a single tracer ([U-^{13}C]glucose) with only extracellular measurements.	[37]
10	7336 reactions and 4527 metabolites	FBA	The gaps and dead end metabolites have been reduced in the CHO genome scale model.	[57]
11	100 reactions and 72 metabolites	EFM	Macroscopic reactions linking extracellular substrates to products of CHO cell metabolism were determined by computing a minimal set of EFMs.	[65]
12	103 reactions and 95 metabolites	EFM	Kinetic model was developed by determining a set of macroscopic reactions (EFMs) based on CHO cell culture data at different culture phases.	[66]
13	144 reactions and 101 metabolites	FBA	FBA was performed on the reaction network to determine nucleotide sugar production fluxes. These fluxes were used as inputs for a glycosylation model.	[67]
14	66 reactions and 42 metabolites	^{13}C-MFA	^{13}C-MFA was used to study metabolic changes in 1G3 cell line after infection with a human adenovirus.	[68]
15	354 reactions and 335 metabolites	FBA	Glucose and lactate metabolism was studied in HEK293 cells by using FBA.	[69]
16	76 reactions and 47 metabolites	MFA	MFA was used to determine the theoretical metabolite demand for production of glycoprotein in AGE1.HN cells.	[70]

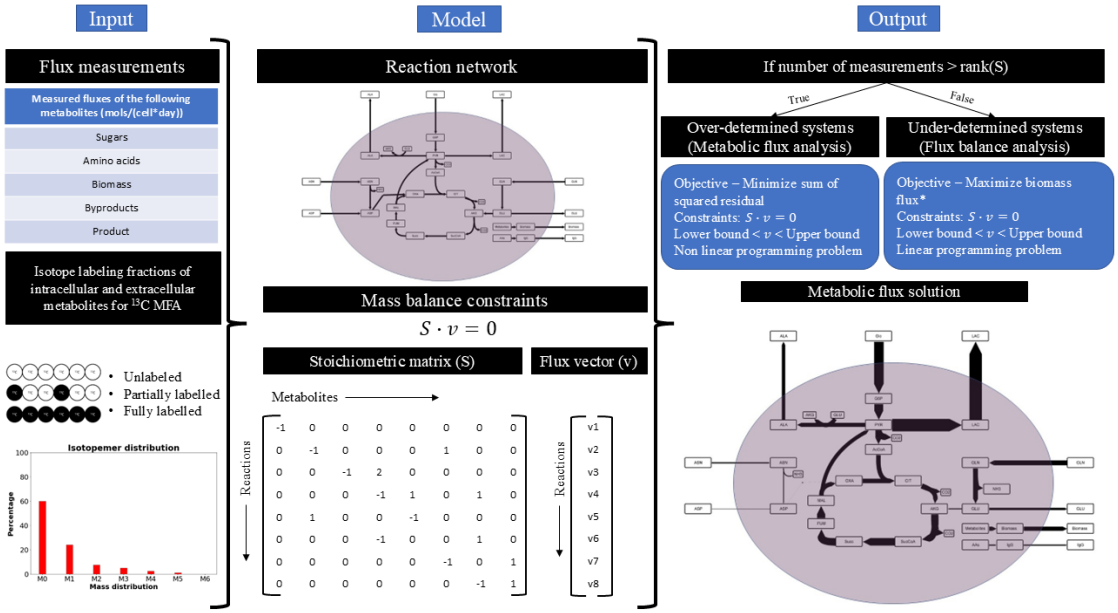

Figure 1.3: Schematic for stoichiometric model representation of cellular metabolism. Stoichiometric models are represented as a mass balance problem. Inputs to the model, model formulation and the methods to solve stoichiometric models are depicted here. The stoichiometric matrix is represented by S and the vector of fluxes is represented by v. Each column of the stoichiometric matrix represents the metabolites, and the rows represent the reactions. The entries in this matrix are the stoichiometric coefficients of the corresponding metabolite in their respective reaction. The fluxes in the vector of fluxes are ordered in accordance with the reactions in the stoichiometric matrix. The product of the stoichiometric matrix with the vector of fluxes represents the mass balance of each metabolite in the system. In mammalian or microbial systems, the rate of change of the concentration of intracellular metabolites is much smaller than the changes in external fluxes. Hence, the pseudo-steady state assumption can be applied. Application of the pseudo-steady state assumption results in the product of the stoichiometric matrix and the vector of fluxes to have a value of zero.

1.3.3 Integrated models

The two previous sections described how kinetic models can be used to provide information on transient metabolic concentrations and the utilization of stoichiometric

models to determine intracellular fluxes at a particular time point. The disadvantages of using these models alone can be overcome by combining the above two models to develop integrated models. This section describes different types of integrated models and how the disadvantages of using only a kinetic or stoichiometric model can be overcome. An integrated modeling framework for CHO cells containing 34 reactions, 24 metabolites and 12 kinetic expressions has been developed [62]. The 34 reactions describe glycolysis, TCA cycle, amino acid metabolism, biomass formation, and antibody formation. Kinetic expressions were used to link the rate of intracellular cytosolic reactions to extracellular metabolite concentrations. The kinetic expressions provided fluxes that can directly be transferred to the stoichiometric model. MFA was performed on the reaction network with 34 reactions by using the fluxes determined from kinetic expressions as constraints. The MFA solution was assumed to be a constant metabolic steady state for 0.25 days and this was used to determine the extracellular concentrations at the end of 0.25 days. This process was repeated until harvest day. This model predicts the concentration profiles for the antibody, glucose, lactate, alanine, asparagine, aspartate, cystine, glutamine, glycine, serine, ammonia, glutamate and biomass. The dependence on the reaction network reduces the number of parameters needed to be regressed compared to a kinetic model. It is often difficult to use a kinetic model to predict the ammonia and lactate formation from CHO cells due to the dependence on many metabolites. In this case, the amount of ammonia coming out of the system is not described through any kinetic expression. The ammonia concentration is determined by performing MFA on the reaction network. [62]. This model could capture the trends for the production and consumption of ammonia and lactate as a

function of concentrations of other metabolites. This is often a challenge for most models.

Reaction networks and kinetic modeling have also been integrated by using EFMs to determine a reduced set of macroscopic reactions from a larger reaction network and subsequently develop kinetic expressions for these macroscopic reactions. A reaction network of 100 reactions describing CHO cell metabolism was used to build a kinetic model by using EFMs to describe a CHO cell-based batch process. The relationship between input metabolites and product changes with culture duration. For example, high glycolytic fluxes during exponential growth phase results in lactate production and this lactate is consumed during subsequent culture phases. The authors of this study divided the cell culture duration into three phases. The three phases are exponential growth phase, transition phase and cell death phase. The measurements from each of these phases along with the reaction network was used to determine EFMs for each of the three phases. A separate macroscopic kinetic model was built for each of the phases and linear switching functions were used to piece together the overall model [65]. In another study, a column generation-based optimization algorithm was used to efficiently generate the EFMs [71]. Improvements in formulating reversible kinetic expressions along with macroscopic reactions generated by EFMs led to development of a single model capable of modeling CHO cell metabolism under all phases of cell culture [66, 72]. EFMs provide a platform to determine the essential macroscopic reactions that describe the given data.

Integrated stoichiometric and kinetic models provide a platform for dynamic modeling of cellular metabolism while integrating a large number of metabolites. Hence, these types of models have the potential to be used to optimize multiple process

parameters that using only kinetic or only stoichiometric models fail to do. The advantages and drawbacks of each model type are described in Table 1.2.

Table 1.2: Comparison of advantages and disadvantages of different types of metabolic models.

Model type	Description	Advantages	Disadvantages
Kinetic model	Semi empirical models that link reaction rates with concentrations of metabolites and bioreactor parameters such as pH or temperature.	• Provides information on transient metabolic profiles. • Models are most often simple, easy to build and understand.	• Extensive experimental data are required to accurately estimate model parameters. • Parameter estimation is complex. • Formulation of semi empirical equations is challenging when process knowledge is low. • Scaling to many metabolites increases the number of parameters.
Stoichiometric model	Reaction networks used to calculate reaction fluxes using metabolic flux analysis (MFA) or flux balance analysis (FBA).	• Cellular behavior can be determined without performing experiments to regress parameters. • Large number of metabolites can be included as models are non-parametric.	• Cannot directly be used to provide transient metabolic profiles. • Model inputs do not depend on concentration of metabolites but depend on fluxes.
Integrated model	Reaction networks along with semi empirical differential equations are used to model cellular metabolism.	• Inclusion of a stoichiometric model improves mechanistic capabilities of the model while reducing the number of parameters required. • Transient metabolic profiles can be predicted. • Large number of metabolites can be included in the model.	• The parameter estimation problem is still challenging due to large computational requirements to solve the model.

1.3.4 Data-driven models

The development and application of the models discussed in the above sections rely on mechanistic understanding of the bioprocess and require skilled engineers and scientists. However, such in-depth mechanistic understanding is not always needed. Relationships between process conditions (temperature, pH, dissolved oxygen, nutrient concentrations, etc.) and process performance (titer, cell density and product quality) can be identified and quantified using statistical methods as well. Although data-driven models can be applied to conditions where mechanistic knowledge is absent, they suffer from poor predictivity outside the design space that they have been trained on. One study developed a mechanistic model for cellular metabolism and N-linked glycosylation from data generated using a two-factorial design of experiments to study galactose addition, manganese addition and viable cell density in perfusion bioreactors and compared it with a response surface model built using the same data. Although the response surface model performed marginally better than the mechanistic model at the edges of the design of experiments, it performed poorly and yielded infeasible solutions when the predictions were extrapolated outside the design space. On the other hand, the mechanistic model provided valid predictions both inside and outside the design space. The mechanistic model also outperformed the statistical model at the center points of the factorial design. The mechanistic model provided insights into the kinetics of the process at different predictions that can be used to perform future experiments [73].

One study on mAb producing hybridoma cells used principal component analysis (PCA) to identify correlations between amino acid uptake/production rates and used partial least square (PLS) regression to quantify the effect of each amino acid and glucose uptake/production rates on uptake/production rates of lactate, ammonia, mAb, and growth. The results of this study showed that key amino acids (glycine, tyrosine,

phenylalanine, methionine, histidine, lysine, valine, isoleucine, and leucine) are positively correlated to mAb production. Although it is expected for essential amino acids to be directly correlated to the growth rate, this study showed that many non-essential amino acids are also directly correlated to cell growth rate. The study also confirmed the strong correlation between glutamine-ammonia and glucose-lactate [74].

PCA has also been used to identify the phase of culture (early exponential, late exponential, stationary, and the phase of decline) of CHO cells in 10 L, 100 L, and 1000 L bioreactors when provided the nutrient uptake rates of amino acids and sugars. The study combined K-means clustering with PCA to cluster the uptake rates into 4 distinct clusters for each of the phases [75]. One study measured the metabolic fluxes of extracellular metabolites and growth rates for the duration of the culture. The metabolic fluxes were linearly linked to the growth rate. However, this linear model cannot hold for the entire duration of the culture as there exist multiple metabolic steady states throughout the culture duration. Hence, the number of metabolic steady states was determined by studying the rate of change of exchange fluxes with respect to growth rate. Hierarchical clustering was used to determine the number of metabolic phases and a separate linear model was determined for each phase. The linear models for each phase were combined together to get a segmented linear model to model the complete culture duration. This model was developed from data on a 2 L bioreactor, and it was applied to predict metabolic uptake rates of large scale 2000 L bioreactor. The model successfully predicted dynamic metabolic concentration profiles at the 2000 L scale. Hence, it showed the potential to be used for in-silico optimization of feeding profiles [76].

PCA and OPLS (orthogonal partial least squares) have been used in another study to identify amino acids that are strongly correlated to growth rate and titer production rate. Cell culture media has been reformulated based on the results. The reformulated cell culture media resulted in a 55% increase in total cell growth and 80 % increase in total mAb productivity [77]. PLS-DA (PLS differential analysis) and Bayesian inference were used to study differences in metabolic and glycosylation process outcomes between different bioreactor scales (250 mL, 5 L, and 2000 L). Bayesian marginal probabilities showed that lactate flux and final titer were the key differences between the scales. PLS-DA further showed that the lactate flux was positively correlated to peak lactate, final titer, average pH, and percentage aggregate. This study also indicated that 2000 L bioreactor cultures exhibited different glycolysis steady states than the cultures at 5 L and 250 mL. It was also suggested that this method could further be improved by including metabolic fluxes in the analysis [78].

Process disruptions and instrument disruptions can occur while performing experiments. Issues with cell counters while performing an experiment could lead to missing data or inaccurate data in a culture run. One study trained a PLS model on bioprocess data (temperature, pH, lactate, glucose, osmolarity, titer, pO_2, pCO_2, and bioreactor impeller rotations per minute) from day 0 to day 6 and day 10 to day 14. This PLS model was used to predict viable cell density from day 6 to day 10 of the culture. The predictions were experimentally validated. This study showed the potential to use statistical methods to fill missing data [79].

Another study used a genetic algorithm for the optimization of a process. Key process parameters such as feed amount, glutamine concentration, glucose concentration, temperature shift, and pH were bounded by a lower and upper bound.

Eight levels were defined in between these bounds for each parameter. A fitness function of total mAb titer was used in the genetic algorithm. Experiments were performed in batches of 22. The fitness function was evaluated after each batch and the genetic algorithm was used to determine the levels of parameters for the next 22 experiments based on the results of the fitness function. This was carried out for 4 batches. This resulted in a 2.7 fold increase in mAb titer [80].

Statistical methods have clearly shown the ability to relate process parameters to process outcomes, often enabling substantial improvements to the process design. These methods can further be improved by combining them with the kinetic and stoichiometric models discussed above to develop hybrid models of cellular metabolism. PLS and genome-scale metabolic models were combined to identify metabolic bottlenecks caused by suboptimal media formulation in human alpha-1 antitrypsin producing CHO-DG44 cell cultures [81]. Another study used PLS to identify the contributions of metabolites to the cell density, growth rate, titer and specific titer production rate based on fed-batch data from CHO cells grown using different feed media. A CHO genome scale model was subsequently used to evaluate the significance of the media effects on intracellular metabolism [82, 83].

Genome scale models have been combined with machine learning models trained on data from ten CHO cell clones grown in two fed-batch processes. This hybrid model resulted in better predictions of uptake/production rates of different metabolites [47, 84]. PCA has been used to add constraints to a CHO-K1 genome scale model in order to reduce the solution space and improve the predictions of growth rate via FBA compared to using only FBA without the constraints added by PCA. Measured fluxes for 27 metabolites from 21 different experiments were used to perform PCA to

determine 6 principal components that can explain 90% of the variance in the flux data. Addition of these principal components to the objective function and addition of new constraints on the principal components resulted in a modified genome scale model that uses a non-parametric approach to model cellular metabolism in cases where mechanistic knowledge to add constraints is not available. This hybrid approach resulted in much better predictions of growth rate compared to FBA alone [85].

Although the development of hybrid models of genome scale models and machine learning in mammalian cell systems is sparse, the development of these models in microbial cells has been reviewed in the literature [86]. Mechanistic models rely on prior process understanding to reduce the number of experiments required to develop the model compared to statistical models. However, in cases of limited process understanding, development of statistical models is necessary. Hybrid models can combine mechanistic models with statistical models to overcome drawbacks arising from partial knowledge of the process. These hybrid models require less experimental data than stand-alone statistical models. One study in the literature has compared dynamic predictions of cell density, titer concentration and metabolite concentrations via using a mechanistic kinetic model to predictions from a hybrid model developed by combining an artificial neural network (ANN) with a mechanistic kinetic model. According to this comparison, the advantage of mechanistic models is that they can provide fair predictions of the process after being trained on limited experimental data but the predictions do not drastically improve when new experimental data are used to train the model. Hybrid models on the other hand outperform mechanistic models when optimal experimental data are available. They conclude that the choice of model

depends on available process knowledge, quantity, and quality of experimental data [87].

A comparison of different types (mechanistic, statistical, and hybrid) of upstream modeling strategies has been discussed in detail in a recent review paper. The comparison has been made on the basis of experimental data required, over parameterization, model development, predictive capability, quality of data, and availability of process knowledge. [88]. The development of high throughput bioreactor systems will aid in gathering experimental data required for the development of hybrid mathematical models.

1.4 Models for N-linked glycosylation

The N-linked glycosylation of biotherapeutics is a critical quality attribute (CQA) because reproducibility in defining N-linked glycosylation characteristics is necessary to maintain consistent quality parameters such as solubility, thermal stability, protease resistance, aggregation, serum half-life, immunogenicity, and efficacy [89]. Maintaining consistent N-linked glycosylation is challenging because the process of producing these glycans is non-template driven. The information for glycan building is encoded in the genome, which is translated by the Golgi apparatus into a branched polymer of sugar molecules assembled onto a protein. Competitive enzymatic reactions and the absence of a template and quality control mechanism leads to a stochastic output [90]. This leads to heterogeneity in the structure of glycans on the mAb. Quality by design (QbD) guidelines set by the FDA emphasize the need to understand the relationship between critical process parameters (CPP) and critical material attributes (CMA) on CQA [91]. Modeling the effect of CPPs such as bioreactor temperature, pH, and dissolved oxygen, and CMAs such as cell culture media on the N-linked

glycosylation provides a platform for process design that is in line with the QbD guidelines. Approval of biosimilars requires comparison of the N-linked glycosylation profile of the biosimilar to that of the reference product. The difficulty in controlling the N-linked glycan profile makes it the most challenging task in biosimilar development [92]. During the past decade, there have been several efforts in developing models that can predict the structure of glycans produced by CHO cells under different conditions. These efforts are discussed in the following subsections.

1.4.1 Kinetic models of N-linked glycosylation

To develop a model to predict the heterogeneity in glycoprotein structures, the first step is to generate the reaction network that produces the glycans. The process of N-linked glycosylation starts in the endoplasmic reticulum but most of the process occurs in the Golgi apparatus. The process involves assembly of sugar motifs on an N-linked glycosylation site of a protein. The process of N-linked glycosylation along with details of the enzymes and nucleotide sugars has been well reviewed in the literature [89]. The process of N-linked glycosylation contains a few enzymes that are capable of catalyzing many reactions. Hence, a rule-based reaction network generation strategy has been adopted in the literature. One of the earliest models defined eight rules for eight enzymes. An example of an enzyme-based rule is that Golgi alpha-mannosidase I acts on glycans with more than 5 mannose molecules to produce a glycan that has one fewer mannose molecule. Similar rules were defined for the seven other enzymes and this resulted in the generation of a reaction network that consists of 33 reactions and 33 distinct glycan structures for a system of eight enzymes and eight rules [93]. This system was improved by increasing the number of enzymes to eleven and using twenty rules. This led to the generation of a reaction network of 22871 reactions that involved 7565

distinct glycans [94]. This massive number of reactions and structures make this process difficult to model. However, certain application-oriented assumptions can help reduce the complexity of the reaction network. For example, one study generated a glycosylation reaction network to describe the glycosylation on the Fc region of a mAb. The authors restricted the glycosylation reaction network to bi-antennary structures and justified this assumption by citing literature that explains the difficulty of producing tri-antennary and tetra-antennary glycan structures on the Fc region of a mAb. This led to the generation of a reaction network with 95 reactions and 75 distinct glycans [95]. The different modeling frameworks that utilize these reaction networks are described in the following paragraphs.

One of the first models for N-linked glycosylation modeled the Golgi apparatus as a series of continuous stirred tank reactors (CSTRs). The CSTRs represent the cis-, medial-, and trans-Golgi cisternae, and the trans-Golgi network. Kinetic rate expressions that depend on enzyme concentrations and glycoprotein concentrations were developed for each reaction. This model was applied to the reaction network that contains 33 reactions and 33 distinct glycan structures. The model is based on a mass balance across the CSTRs. This leads to a system of 132 nonlinear equations that needs to be solved to obtain the steady-state glycan distribution for a secreted glycoprotein.

The effect of overexpressing different enzymes on the distribution of glycans was captured by the model as a function of the enzyme concentrations in the compartments [93]. This framework was expanded to include 22871 reactions with 7565 distinct glycans. The expanded framework included the concentration of nucleotide sugars in the kinetic expressions [94]. There have been efforts to model the Golgi apparatus as a plug flow reactor (PFR) with recycle of Golgi resident proteins

such as the enzymes and transport proteins. The kinetic rate expressions were modified to include sequential order kinetics. This type of model formulation leads to solving a series of partial differential equations to obtain the steady-state glycan distribution for a secreted glycoprotein [95].

The concept of EFM has also been applied to develop an N-linked glycosylation model called glycan residues balance analysis (GReBA). GReBA correlates the final glycan fractions to the extracellular sugar concentration via Michaelis-Menten kinetic expressions without using a CSTR- or PFR-based approach. The Michaelis-Menten kinetic expressions are developed for simplified macro reactions relating extracellular sugars to the glycan fractions. Perfusion cultures were performed at varying galactose and mannose concentrations in the culture to gather data for model parameterization and subsequent validation of model predictions [96].

The process of N-linked glycosylation is very complex. Small perturbations of the process can lead to a different glycan profile on the product. These types of models provide information on the reactions of the glycosylation pathway. Hence, these models are capable of providing invaluable insight into the effect of process parameters and media additives on the glycosylation pathway. The application of these models to study these effects is discussed in Section 1.7. The differences between the CSTR-based approach and the PFR-based approach are highlighted in Figure 1.4.

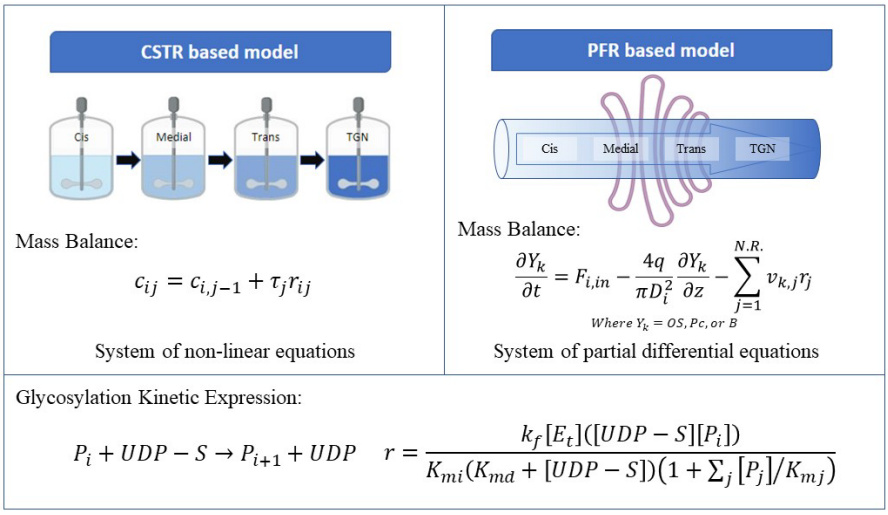

Figure 1.4: Schematic for kinetic models for N-linked glycosylation. Models that approximate the Golgi apparatus as a series of CSTRs solve a system of nonlinear equations to predict the glycan fractions. The PFR-based models use a system of PDEs to do the same. The reactant glycan P_i reacts with nucleotide sugar UDP-S to produce the product glycan P_{i+1} and the nucleotide UDP. The mannosidase reactions result in trimming of the mannose sugar from the glycans. These reactions do not involve nucleotide sugars. The rate equation shown in this figure is Michaelis-Menten kinetics with competitive inhibition. Although this is used in a few literature publications, other publications have expanded upon this to include more complex kinetics that are more representative of the enzymatic reactions such as sequential order Bi-Bi kinetics and random-order Bi-Bi kinetics. The mass balance for the CSTR is represented by the concentration of glycan i leaving tank j (C_{ij}) is equal to the amount of glycan i entering tank j ($C_{i,j-1}$) added to the product of the residence time (τ_j) and the net amount of glycan i reacted in tank j (r_{ij}). This mass balance results in a system of non-linear equations that can be solved to obtain the glycan structure fractions. The PFR-based model also relies on a mass balance of glycans, nucleotide sugars and nucleotides; the concentrations of these species are represented by Y_k. In a PFR, the concentration of the reactant varies with respect to the time (t) and axial distance (z) in the PFR. The net rate of reaction of the glycan j is represented by the sum of product of stoichiometric coefficient of glycan j in reaction k ($v_{k,j}$) and the rate of reaction (r_j). Hence, this mass balance is represented in the form of a PDE with respect to time and distance. Created with BioRender.com.

1.4.2 Non-parametric approaches to model N-linked glycosylation

The explosion in the number of kinetic rate parameters due to the large number of reactions has led researchers to look for non-parametric approaches to model this process. For example, glycosylation has been modeled as a Markov chain [97]. A similar rule-based approach was developed to generate the reaction network. Each glycan was considered as a state in a stochastic network that transitions from one state (reactant glycan) to another (product glycan) with certain transition probabilities. Since the glycosylation reactions only depend on the current structure of the glycan and not on its past states, these transition probabilities can be described as a time-discrete Markov chain with initial distribution concentrated on the initial glycan species that enter the Golgi and a transition matrix that comprises the probabilities of transitions between glycans. In order to model glycan secretion from the trans-Golgi, artificial absorbing states are introduced to allow the chain to virtually stop in a particular glycan, thereby mimicking its secretion [97].

Another non-parametric approach to model N-linked glycosylation is the use of ANN. The authors of this study generated data on the intracellular nucleotide sugar concentration and N-linked glycosylation profile under different cell culture conditions. The intracellular nucleotide sugar concentrations were used as an input to the ANN to predict the output glycan profile. The ANN was trained on these data. The predictions of the ANN were compared to those of a PFR-based kinetic model. The ANN-based model performed better than the PFR-based kinetic model both in terms of the error of the predictions as well as the computational requirements [98].

Data from a PFR-based kinetic model have been used to develop a dynamic kriging-based surrogate model that predicts the glycan profile based on the specific titer production rate, cell density, glucose concentration, bioreactor pH, and temperature

[99], providing a model that is computationally less intensive. Another model called Glyco-Mapper allows users to fix nutrient composition and glycosylation enzymes to determine the appearance or disappearance of one or more glycan structures in the final product. This system uses fuzzy parameters to relate the nutrient concentrations to central carbon metabolism and uses the outputs from the central carbon metabolism along with the glycosylation genes to determine the presence or absence of a glycan on the final product [100].

Network-based approaches have also been used to gain mechanistic understanding on N-linked glycosylation of mAbs. Metabolic flux analysis and flux balance analysis are powerful tools that can be used to gain insight into the fluxes of each reaction in a reaction network. The rate of production of glycans can be determined from the specific titer production rate and the percentage of different glycans present on the mAb. The experimentally determined rates of production of glycans can be used as constraints to perform metabolic flux analysis or flux balance analysis on the N-linked glycosylation reaction network. This concept has been used to develop a tool called glycosylation flux analysis (GFA). GFA has been applied to study the effect of culture duration and addition of manganese on the N-linked glycosylation of a mAb. A study in the literature showed that the enzymatic activity of galactose transferase dropped over time potentially due to change of culture conditions and that the addition of manganese increased the enzymatic activity of galactose transferase in the early part of the culture but did not significantly affect it in the late stage of the culture potentially because the effect of culture conditions on the enzyme activity were greater than that due to the addition of manganese in the later part of the culture. This study developed a tool that can be applied to study the effect of many different process parameters on glycosylation

[101]. GFA has also been used to study N-linked glycosylation in perfusion cultures. GFA showed that changes in IgG glycan composition in the studied perfusion cultures could be explained by changes in galactosylation flux activity. Random forest regression analysis was used to link the changes in galactosylation with specific IgG production rate and extracellular ammonia concentration [102].

The different types of non-parametric approaches used to model glycosylation along with their respective inputs and model outputs are represented in Figure 1.5. The choice of an N-linked glycosylation model depends on the amount of data available, the objective of the model and the knowledge of the process. A review of the development and comparison of mechanistic and data driven models for N-linked glycosylation has been published [103].

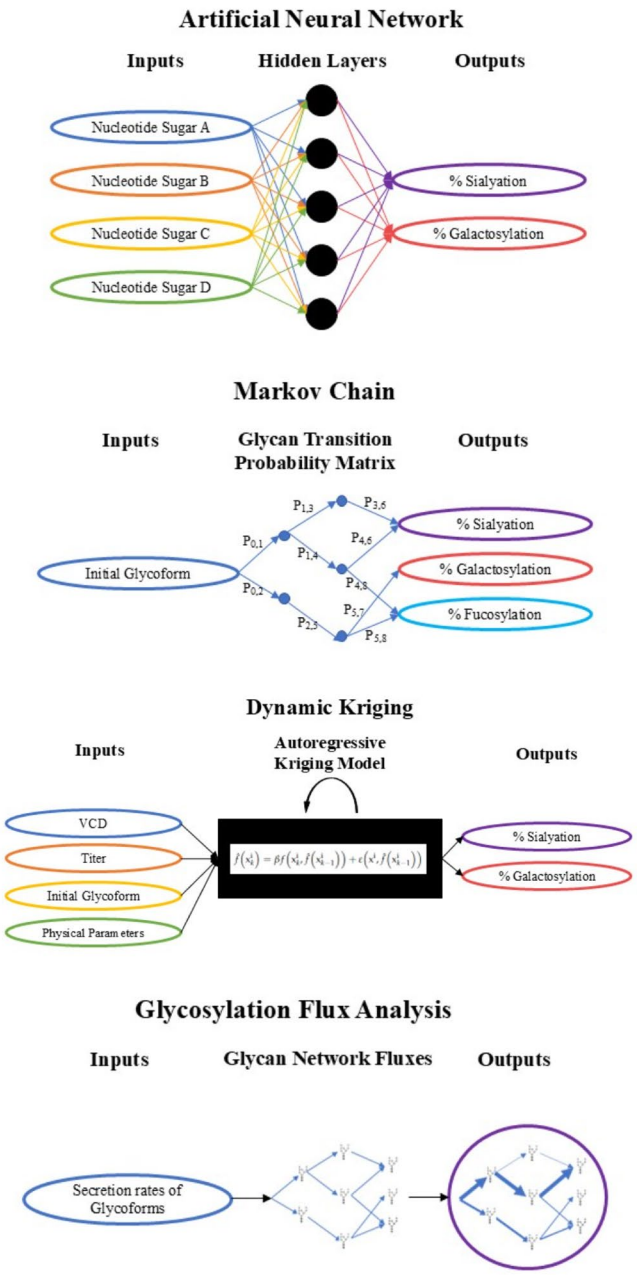

Figure 1.5: Schematics of surrogate and reaction network-based models for glycosylation

1.4.3 Recommendations for modeling N-linked glycosylation

Mechanistic kinetic compartment models for N-linked glycosylation such as the CSTR- and PFR-based models discussed in Section 1.4.1 are computationally intensive [104]. To determine the parameters of these models, these models need to be solved thousands of times. Hence, parameter regression is a computationally intensive process that takes a long time. It is important to develop parallel computing packages to be able to utilize computer clusters to reduce the time required to determine the parameters of these models. Data required for development of these models include accurate measurements of glycan fractions and sometimes intracellular measurements of nucleotide sugars. The protocol for measuring data relevant to glycosylation models, the impact of experimental data precision, glycosylation model design, evaluation of model parameters and evaluating the accuracy of predictions has been reviewed in the literature [105]. About 15 to 20 percent of human IgG molecules have N-linked glycans on the Fab region of the antibody [106]. Fab glycans are produced due to somatic hyper mutation. The development of various analytical techniques has led to characterization of the glycans present on the Fab and Fc regions of mAbs. It is often observed that the glycans on the Fab regions are far more developed than the glycans on the Fc region. The importance of Fab glycosylation is still understudied. However, it has been shown that the Fab glycosylation has a significant effect on antibody half-life, antibody aggregation, immune modulation, antigen binding and stability of the mAb [107-110].

The effect of cell culture parameters on the Fab glycan formation needs to be studied by using experiments and computational models to gain mechanistic insights. The ANN-based N-linked glycosylation model has been modified to predict the glycans on multiple N-linked glycosylation sites of two fusion proteins by including the intracellular glycosyltransferase concentrations and intracellular nucleotide sugar

concentrations as inputs [98]. Such studies should be expanded to study site specific Fc and Fab glycosylation on mAbs. The process of glycosylation is known to be influenced by various external parameters such as pH, temperature, dissolved oxygen, ammonia, manganese, sodium butyrate, dimethyl sulfoxide (DMSO), etc. The effects of these factors vary across cell lines and protein structure [111]. Variations in N-linked glycosylation were seen when the same mAb was expressed in CHO-K1, CHO-S, and CHO-DG44 cell lines [112]. Hence, it is difficult to use various literature data that are produced on different cell lines and protein products to draw general trends on the effect of process parameters on N-linked glycosylation. A brief comparison of the different types of glycosylation models and their applications has been described in Table 1.3.

Table 1.3: Comparison and applications of different types of models for N-linked glycosylation.

Model type	Advantages	Limitations	Reference
Golgi apparatus approximated as a series of CSTRs	Each parameter of this model has a physical representation in the glycosylation process. It can provide mechanistic insights into the process. Depending on the size of the system of nonlinear equations, this could be easier to solve than the PFR model.	Solving the system of nonlinear equations is still a time-consuming step. Parameter regression is complex. The PFR-based model has been used more in the literature. Hence, the nucleotide sugar synthesis has not yet been connected to this model.	[93, 94, 113]
Golgi apparatus approximated as a PFR	Each parameter of this model has a physical representation in the glycosylation process. It can provide mechanistic insights into the process. This model has been used to gain mechanistic insight and modulate glycan structures to a larger extent than the CSTR-based model. This model has been developed over many iterations by the original authors of the model.	This model is known to be very computationally demanding. The computational demand of this model makes the parameter regression challenging.	[18, 95, 114-116]
Glycosylation flux analysis	No parameter regression needed as this model does not have parameters. This model can be applied to study differences in glycosylation fluxes in different experiments to better understand which enzymes are affected by changes in culture conditions.	The inputs to this model are fit to logistic equations. Differentiating these logistic equations to get fluxes that are used as constraints for the flux analysis is not ideal as it drastically depends on the regressed parameters of the logistic equations. The model cannot be used to generate dynamic glycan profiles.	[101, 102]
Artificial neural network	The model is developed as a less computationally intensive surrogate model to the PFR-based model. This can be applied to design controllers and perform model-based optimization.	The model does not provide any mechanistic insights. Development of this model requires regression of the PFR-based model to experimental data.	[67, 98, 104]

1.5 Integrating models for cellular metabolism with models for N-linked glycosylation

Bioreactors are most often operated in either fed-batch or perfusion mode. Dynamic changes of nutrient concentrations, product concentrations, cell density and N-linked glycosylation profiles are intrinsic to the process. Supplementing sugars such as galactose, fucose, and mannose to the cell culture medium at regular intervals is commonly done to modulate N-linked glycosylation profiles. These sugars can lead to production of different nucleotide sugars in the cell. Mechanistic linkage between sugar supplementation and glycan fractions has been reviewed [117]. Changes in intracellular nucleotide sugar concentrations, enzyme expression levels, specific titer production rates, and extracellular ammonia concentration with respect to time can influence the N-linked glycosylation process [18, 73]. Hence, it is important to modify the previously described models to describe the dynamic profiles of N-linked glycosylation. This has been done in a few different ways in the literature.

A kinetic model for metabolism has been used to determine the dynamic profiles of concentrations of glucose, lactate, ammonia, asparagine, lactate, glutamine, mAb, and cell density. The outputs from this kinetic model for metabolism have been used as inputs to another kinetic model that determines the intracellular concentrations of nucleotide sugars as a function of culture time. The specific titer production rate and the intracellular nucleotide sugar concentrations have been used as inputs for a PFR-based glycosylation model [114]. A similar approach has also been shown with the ANN-based glycosylation model [104]. Hence, a platform is provided for describing dynamic glycan profiles. This approach was used to model the effect of ammonia, galactose, manganese concentrations, and cell density on N-linked glycosylation of a mAb produced in a perfusion bioreactor [73]. The kinetic model used to determine the

nucleotide sugar rates utilizes the glucose concentration and glutamine concentration to determine the intracellular concentration of nucleotide sugar. Nucleotide sugar metabolism is complex and it is difficult to model it with limited inputs. Hence, a stoichiometric model has been used to link more nutrient uptake rates to nucleotide sugar production rates calculated via flux balance analysis. The nucleotide sugar production fluxes have been used as inputs to the ANN-based model [67]. This leads to a more comprehensive way to link extracellular nutrients to the glycan profile. However, a dynamic model of metabolism that describes the dynamic profiles of many nutrients has not been developed. The addition of such a model to this framework will provide a comprehensive model to describe dynamic N-linked glycan profiles. This approach can adapt either PFR- or CSTR-based glycosylation models. The method used to combine the models for metabolism and N-linked glycosylation depends on the available measurements. Different methods that exist for this integration are shown in Figure 1.6. The application of these models to optimize cell culture CQAs are discussed in Section 1.7.

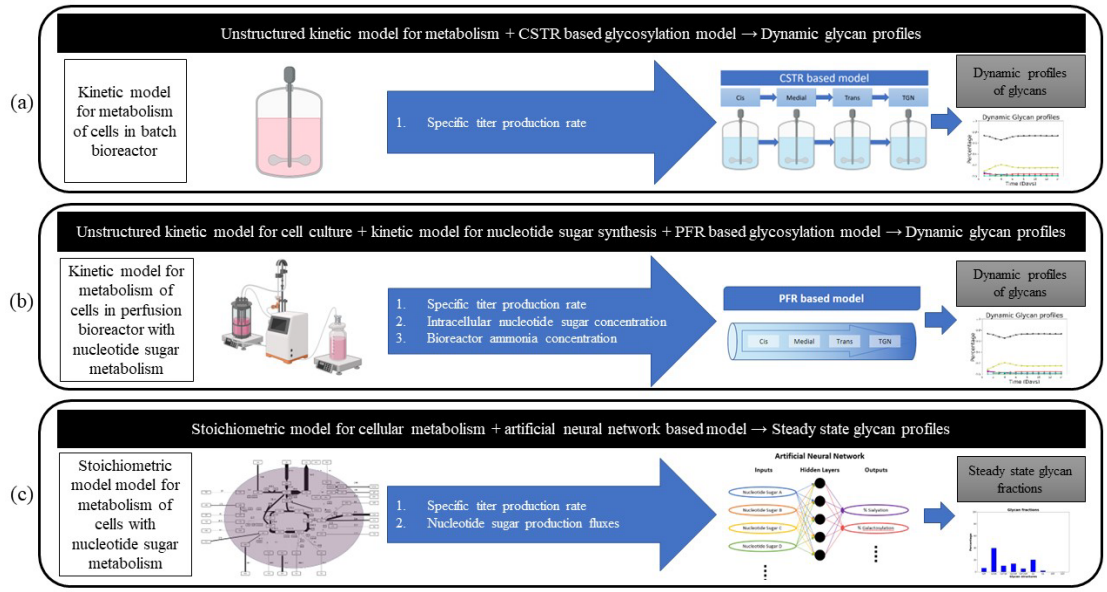

Figure 1.6: The different methods used to integrate models for metabolism with models for N-linked glycosylation are shown here. (a) A kinetic model was used to determine specific titer production rate until harvest day. This specific titer production rate was fed to a CSTR-based model for N-linked glycosylation to determine dynamic glycan fractions. (b) A kinetic model that includes nucleotide sugar metabolism was used to determine ammonia concentration, specific titer production rates, and intracellular nucleotide sugar concentrations in a perfusion bioreactor. This was fed to a PFR-based model for N-linked glycosylation to determine dynamic glycan fractions. (c) A stoichiometric model for metabolism including nucleotide sugar synthesis reactions was used to determine specific titer production rate, and nucleotide sugar synthesis rates. This was fed to an ANN-based model for N-linked glycosylation to determine steady state glycan fractions. Created with BioRender.com.

1.6 Modeling physiochemical parameters in bioreactors

Considerable work has been done to develop kinetic and stoichiometric models which predict bioreactor outcomes based on chemical inputs and can be used to optimize bioreactor media, feeding methods, bioreactor design, and bioreactor operating conditions. However, many of the models in the literature have been built using

metabolomics data from shake flasks which cannot sufficiently control important physical parameters including pH, dissolved oxygen (DO), and partial pressure of carbon dioxide (pCO_2) that can have a strong influence over biological responses and must be controlled properly in large scale production bioreactors. For example, changes in cell culture pH leads to changes in the uptake rate of amino acids [118]. The optimal media and feeding schedule will change due to the changes in uptake rates. Thus, the models described in the previous section that were used to optimize the feeding schedule or media will fail to capture these changes and cannot be used for optimization. Table 1.4 details the known effects of parameter changes on uptake rates, production rates and glycosylation profiles. These effects are described in further detail in the subsections below. Importantly, very little work has been performed to capture the effect of these physical parameters in cell culture models. Incorporating physical parameters into metabolic and N-linked glycosylation modeling can improve the predictive power of modeling and provide mechanistic insights on the effects of these physical parameters on cells. This section reviews efforts that describe the effect of these bioreactor parameters on metabolism and glycosylation and modeling the effects of these parameters.

1.6.1 Modeling temperature effects

Industrial bioreactors are often seeded at 37 °C as this is the optimal temperature for CHO cell growth. However, it is common practice to reduce the operating temperature when the cells reach a stationary phase. This reduced temperature helps to extend the stationary phase, which is the phase at which the cells are most productive in producing antibodies [119]. However, low temperatures hinder maturation of glycans and affect quality of final product. Flux balance analysis on a CHO cell reaction network

has shown that reducing the culture temperature leads to lower nucleotide sugar production fluxes [51]. Multiple experimental studies in the literature have found that lowering the culture temperature leads to a reduced growth rate but an increase in specific titer production rate. These studies have also shown that the specific amino acid uptake or production rates also vary with temperature [118, 120, 121].

Few models in the literature consider the effect of temperature on metabolism. Of the models that do, most are discrete with respect to temperature; data are collected at two temperature values and unique sets of parameters are fitted to each. These models can be used to optimize the day or time that the temperature switch should occur to maximize titer. The authors of one study simulated the effect of the day of temperature shift on the culture and determined the optimal temperature shift day for maximizing production. The optimal value was confirmed experimentally [122].

However, since these models fit distinct parameter sets for two discrete temperatures, they cannot provide insight into a possible optimal temperature outside of the two temperature values tested. Discrete models also cannot be used to determine the effect of non-instantaneous change of temperature, such as ramping down temperature over several hours. Rather than using a discrete model, the use of one continuous model with respect to temperature to describe metabolism will make it possible to find an optimal temperature and shift schedule. Dynamic metabolic flux analysis (DMFA) was used to study the effect of lowering culture temperature on intracellular metabolism. In the temperature shift condition, reduction in cell specific growth rate simultaneously reduced the biomass precursor requirements and most intracellular fluxes as expected. In the condition without temperature shift, the cell specific titer production rate increased slightly initially but as culture duration progressed there was a decline in cell

specific titer production rates, possibly due to increased accumulation of lactate and ammonia in a culture without temperature shift. By day 8 of the culture, antibody concentration in the shift condition had surpassed that of the constant condition. The authors showed that reduction of growth rates and most intracellular fluxes may have redirected resources towards mAb production in the temperature shift condition, leading to an increase in final antibody titer by 42% [123].

The PFR-based N-linked glycosylation model has been used to model the effect of temperature shift on N-linked glycosylation of the mAb. The parameters of the model implied that the GalT enzyme concentrations had to be reduced for the model to predict the N-linked glycan fractions at the lower temperature. Experimental measurement of GalT expression levels showed that the GalT expression was reduced at lower culture temperature [115].

There is limited scientific literature on modeling the effect of temperature on glycosylation. Temperature is a bioreactor parameter that can significantly improve mAb production while decreasing product quality [124]. Hence, there is need for more studies focusing on modeling the effect of temperature on both metabolism and glycosylation to better understand these opposing effects.

1.6.2 Modeling the profound effect of pH on growth and product formation

Optimizing the pH of a CHO cell culture is challenging for many reasons. Differences in specific nutrient uptake and secretion rates, cell growth, specific titer production rate, and glycosylation profiles have been observed while operating the bioreactor at different pH and temperature values [118, 125-130]. Figure 1.7 presents kinetic data from experiments with antibody producing CHO cells grown in bioreactors held at three pH values using the same feeding pattern. Cell growth and titer production

are significantly affected by these pH differences, but certain amino acids experience different accumulation and depletion patterns contributing to the observed process performance. Papers published in the literature also report a change in uptake/production rates of amino acids when CHO cells are grown at different pH values [118]. This highlights the need for determining unique optimal feeding patterns and media nutrient concentrations for pH values individually. However, a design of experiments approach to do this requires extensive experimental efforts. The development of models which can incorporate the effect of bioreactor parameters such as pH is warranted to determine optimal feeding patterns across all pH values with fewer experimental requisites.

The pH of CHO cell cultures also has a significant effect on glycosylation. An increase in pH has shown to reduce terminal galactosylation and a reduction of pH has been shown to increase sialylation of mAbs [111]. However, this effect usually changes with different cell lines and products, making it difficult to choose an optimal pH across CHO cell cultures.

There are very few models that provide information on the variable effects of pH on cell lines in the literature. A study has been published in the literature on model-based optimization of pH shift of cell culture. The authors used a kinetic model that consists of five differential equations incorporating 6 system components. Model parameters were determined based on experimental data at pH values of 6.6, 6.8, 7, and 7.2. This model was used to determine an optimal pH shift schedule. Experimental verification of this pH shift schedule led to an increase in product titers of 40% compared to the conditions of constant pH [131]. However, the publication did not have a model for N-linked glycosylation associated with it. Utilizing a model for N-linked

glycosylation that incorporates the effect of pH along with the kinetic model for metabolism would have helped place additional constraints while optimizing the pH shift schedule.

Another study in the literature developed a mechanistic model for metabolism and glycosylation that includes the effect of pH and temperature. This model was subsequently used to develop a computationally inexpensive surrogate model by using dynamic kriging. The surrogate model was applied to determine the feasible operating region for temperature and pH [99]. This model was not regressed to experimental data, thus limiting its use.

Transcriptomics, proteomics, metabolomics, and glycomics have also been used to study the effect of pH on CHO cell cultures. A recent study revealed that differences in intracellular pathways such as vesicular trafficking, apoptosis, and cell cycle affect glycosylation, growth rate and specific titer production rates [132]. Such studies can be combined with genome scale models to provide mechanistic insights into the effects of pH. Models have shown potential to improve product titers by optimizing the pH shift, but there is a need for further modeling efforts to investigate the effect of pH.

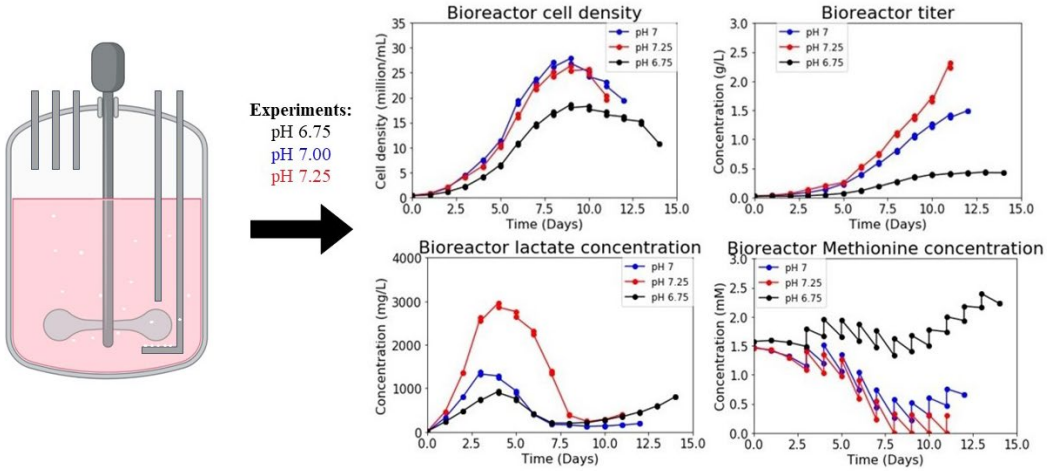

Figure 1.7: Bioreactor pH affects culture performance. A bioreactor operated at three different pH values using the same feeding schedule resulted in changes to CHO cell growth rate, monoclonal antibody production, ammonia concentration, and methionine concentration in the bioreactor. Created with BioRender.com.

1.6.3 Modeling the effect of dissolved oxygen

Sufficient supply of oxygen to cells through dissolved oxygen (DO) in the media is vital for respiration and other cellular processes and is necessary for viability, growth, productivity, and product quality. The DO level in a bioreactor is typically held at 40%-50% of the full saturation of oxygen from air at standard temperature and pressure. Sufficient DO is difficult to maintain because of its relatively fast consumption rate and low solubility. Though shaken flasks can sustain cells at low volumes, sparging is required in even lab-scale bioreactors to supply oxygen to the suspended cells [133]. Dissolved oxygen and pH are both controlled through sparging gases, which creates a dependence between these terms [134]. DO levels also depend on local gas transfer from bubbles to the fluid, from which convection is required to distribute oxygen throughout the bioreactor. As high cell density bioreactors are scaled up, the mixing time increases

while volumetric oxygen uptake rate remains constant, which could result in oxygen deprived zones. Thus, dissolved oxygen is highly susceptible to heterogeneous distributions in large scale bioreactors [135].

Dissolved oxygen has a significant effect on CHO cell growth and productivity. Experiments performed at DO levels 5%, 40%, 60%, and 90% of air saturation found that cell growth is relatively constant at levels 40% or higher but is four times lower at 5%. Reasonably, the productivity of cells held at 5% DO was also four times lower. Productivity was best at 40% saturation, but higher DO levels resulted in approximately half as much product as the standard 40% case [136]. Another study investigated DO levels of 10%, 25%, and 40% and found that within this range, increasing oxygen levels result in higher productivity and average and maximum cell growth [137, 138]. Another CHO cell line producing human recombinant erythropoietin (EPO) was studied at DO levels varying from 3% to 200%; the specific growth rate was found to be constant at all conditions except for 200% DO, at which it was significantly lower. The specific EPO production rate was constant between 10% and 100% DO, and dropped substantially for 3% or 200% DO [139].

CHO cells experience a metabolism shift at limited oxygen levels, which results in a higher production of lactate [140]. Lactate subsequently affects the pH of the bioreactor and this can affect glycosylation. However, the effect of DO on glycosylation is highly cell line and protein specific. Some findings have shown that DO ranging from 10-100% typically allows for consistent glycosylation [141]. Others have shown significant changes, including reduced fucosylation [139], reduced galactosylation [142] and higher sialylation [143] at low DO levels, and higher triantennary and sialylated glycans when exposed to oscillating DO levels [111, 144].

Stoichiometric models can inherently capture the oxygen uptake rate by cells if oxygen is added as a metabolite to the stoichiometric model and all the major reactions that include oxygen as a product or reactant are included in the reaction network. For example, the reaction network used in the integrated model described in Section 1.2.3 can determine oxygen demand [62]. An MFA study of process parameter impacts on CHO cell metabolism in the late stationary phase was performed, and effects of changing media, carbon dioxide and coupled effects of DO and temperature were compared. High DO was shown to cause oxidative stress, indicated by the positive impact towards the oxidative pentose phosphate pathway (PPP). Dissolved carbon dioxide had an opposite effect on fluxes towards the oxidative PPP, as well as on alanine and glutamic acid consumption and formation [145]. The model was not functionalized to predict fluxes outside of the tested DO levels, though the findings shed light on pathways affected by environmental DO levels. A kinetic model was also adapted for hybridoma cells to capture the metabolite uptake and production rates at various DO concentrations using correction factors determined experimentally. This kinetic model for mAb producing hybridoma cells was integrated with computational fluid dynamic (CFD) modeling to capture oxygen transfer to the fluid and alongside uptake by cells, enabling predictions in large, heterogeneous tanks [146].

In another study, a kinetic model was developed to describe CHO cells' oscillatory and dampening effects to transition from an oxidative metabolism to anaerobic metabolism after introducing the cells to a hypoxic environment (from 40% to 10% saturation with air) at 48 hours into cultivation. Kinetic expressions for fluxes of metabolites are described by Michaelis-Menten dynamics while oxygen consumption rate is defined by a Monod-type relationship to extra-cellular oxygen concentration.

Oscillations were captured mathematically, and parameters were fit by setting initial parameters from literature and maximizing ATP-to-ADP oscillatory behavior while minimizing simulation error for an ideal smooth response behavior. Thus, this model would greatly benefit from validation against experimental data [147].

Another study investigated a drop in viable cell density and viability upon scale-up of CHO cell culture using metabolomic and proteomic approaches and found the cause to be high oxidative stresses, likely due to periodic contact to lower oxygen regions [148]. Low oxygen and oscillating oxygen levels have been shown to cause increased reactive oxygen species (ROS) [149, 150]. The metabolomic and proteomic approach enabled partial recovery of cell density and viability through trace copper addition, which downregulated apoptotic regulators and reduced ROS production [148].

Oxygen effects on CHO cell metabolism and glycosylation are complex, and further work is needed to elucidate the effects of oscillating DO levels present in large scale bioreactors.

1.6.4 Modeling the effect of the partial pressure of CO_2

The pH in CHO cell bioreactors is frequently controlled through CO_2 sparging to bring the pH down to a level that cells prefer [151]. Sparging in CO_2 works to lower the pH without increasing the osmolality of the solution. The partial pressure of CO_2 (pCO_2) increases in large bioreactors due to CO_2 accumulation from cell metabolism and insufficient mixing [152-156]. However, CO_2 affects CHO and generally animal-cell metabolism and glycosylation, which must be considered in process design and operation [152-156] and thus modeled. Mixing, sparging, and agitation for oxygen supply are also used to prevent CO_2 accumulation (and thus high pCO_2 levels) whether CO_2 is externally supplied or derived from cell metabolism. pCO_2 and pH levels are

inextricably linked in terms of control but have differential impacts on cell metabolism and glycosylation. High levels of carbon dioxide, which may accumulate in large scale bioreactors, can reduce or completely eliminate the metabolic shift to lactate consumption. This reduces cell viability and product titer.

Metabolic flux analysis has been performed to examine the effects of elevated dissolved carbon dioxide on the metabolic switch to lactate consumption. The findings support complex effects of pCO_2 on cell metabolism, which may be hard to capture in a comprehensive model [157]. The effects of pH and pCO_2 on CHO cell metabolism and mAb glycosylation was also investigated, determining that these parameters have interaction effects on average cell growth, production, amino acid uptake rates, mAb acidic charge variants, and mAb sialyation [137, 138]. Thus, it is important to consider both pH and pCO_2 to capture their individual and interaction effects.

Table 1.4: Impacts of physiochemical parameters on cellular metabolism and N-linked glycosylation.

Physiochemical parameter	Effect on metabolism and N-linked glycosylation
Temperature	• Humanized mAb producing CHO (DUXB) cells grown at a temperature of 33 °C experienced lower growth rate compared to cells grown at 37 °C. The specific mAb productivity increased at lower temperatures [120]. • Temperature shift from 36.5 to 32 °C in IgG producing CHO-T cell line cultures led to higher viability on harvest day, 25% increase in specific mAb productivity, decrease in intracellular amounts of certain nucleotide sugars (UDP-Glc, UDP-Gal and UDP-GlcNAc), reduction in nutrient consumption rates and increase in the under processed glycan structures in the mAb [51]. • The specific uptake rate of glucose and lactate reduced with decrease in culture temperature. The specific production rate of ammonia and lactate also reduced with decrease in culture temperature. Reduction in temperature led to decrease in growth rate. The reduction in temperature only resulted in higher specific titer production rates at a pH of 7, 7.2, and 7.4 but was similar to higher temperature conditions at other pH values. This study was performed on EPO producing CHO (DUKX-B11) cell line [118].

	• Reduction in temperature led to higher specific erythropoietin (EPO) production rates. The N-linked glycosylation of the EPO was not drastically affected from a temperature of 37 to 32 °C. A reduction of tetra-antennary structure and tetra-sialylated glycans was observed below 32 °C. This study was performed in perfusion cultures using CHO (DUX-KX-B11) cell line [124]. • Effect of temperature reduction was studied using EPO-Fc producing CHO (DUKX-B11) cell line. Reduction of temperature led to supressed growth rates. Lowering culture temperature led to increase in specific titer production rate. Reduction in temperature led to reduction of EPO-Fc sialylation [12].
pH	• The maximum growth rate was achieved at a pH of 7.2. Culture longevity increased with decrease in pH. Increase in pH led to increase in specific uptake rate of glucose and glutamine. Increasing pH also led to decrease in specific production rates of ammonia and lactate. This study was performed on EPO producing CHO (DUKX-B11) cell line [118]. • Reduction of pH in mAb producing CHO (DUXB) cell line cultures led to increase of galactosylation index and sialylation index. However, mAb productivity also dropped [158]. • Specific mAb production rates were measured in mouse placenta lactogen producing CHO-K1 cultures from pH of 6.1 to pH of 8.7. The maximum value was observed at a pH of 7.8 [126]. • The maximum growth rate was achieved at a pH of 7.2. Increase in pH led to decrease in specific mAb production rates. Increase in pH resulted in reduction of terminal galactosylation and sialylation. This study was performed on IgG producing murine hybridoma cell line HFN 7.1 [159].

Table 1.4: Continued

Dissolved Oxygen	• Low DO causes reduction in growth rate and specific titer production rates. Reduction in specific titer rates was also observed at high DO. The effects of DO were studied on fusion protein producing CHO-K1 cells [136]. • A DO level of 5% caused reduction in cell growth rate and specific titer production rates, and no effects to these terms were observed for DO levels of 40%, 60%, and 90% in mAb producing CHO cell cultures [137]. • Specific lactate production rates increase with a decrease in dissolved oxygen. Viable cells drop with low DO, as does mAb titer. Cultures in this study used mAb producing CHO cells [140]. • DO ranging from 10% to 100% did not impact N-linked glycosylation [141]. • Decreasing DO from 100% to 50% to 10% led to reduced galactosylation and sialylation. A murine B-lymphocyte hybridoma cell line was used in these studies [142, 143].
Partial Pressure Carbon Dioxide	• High pCO_2 led to reduction in metabolic shift from lactate production to lactate consumption. High pCO_2 also led to reduction in cell viability and reduced product titer in mAb producing CHO cell cultures [157]. • The combined effect of elevated pCO_2 and osmolality resulted in a lower growth rate and specific tPA production by a CHO cell line MT2-1-8 [153].

1.7 Applications of CHO cell models

There are various methods to model cellular metabolism. These methods have been developed to perform different tasks. Kinetic models have been developed to capture dynamic changes in the bioreactor while stoichiometric models can be used to study the consumption of specific metabolites within the cell by determining the intracellular fluxes. The differences in these models make them ideal to model different key aspects of CHO cell-based processes. This section looks at the applications of these models in process simulation, process optimization, and in developing a deeper mechanistic understanding of CHO cell metabolism and glycosylation. A brief summary of the different types of applications is listed in Table 1.5.

1.7.1 Bioprocess simulation

Developing a model that can simulate a large number of process or genetic conditions provides a platform to study and develop the process under different

operating conditions. The QbD guidelines provided by the FDA highlight the importance of relating process conditions to the product yield and quality to help develop robust processes that produce superior quality product consistently [91]. The ability to simulate the effect of process conditions on product quality and yield provides a framework to study the sensitivity of process conditions. Evaluation of the design space by these models can be subsequently used to determine optimal process conditions, develop control systems or perform techno economic analysis. The dynamic MFA-based model discussed in Section 1.3.3 has been used to simulate 9200 bioreactor runs by choosing various combinations of process variables (seed density, harvest day, temperature shift value and temperature shift day) and knockdowns of 8 different metabolic enzymes. The different simulated process conditions led to variations in titer values. Some conditions resulted in a reduction in titer while some cases led to a greater than 2-fold improvement in titers [160]. Hence, dynamic MFA modeling provided a platform to study the effect of different process related conditions on the final titers. The application of CHO genome scale models to simulate the consumption rates of essential amino acids based on the growth rates of the cells has been described in Section 1.3.2. This approach has been developed to overcome the limitation of underestimating the consumption rates of many amino acids by combining the genome scale model with a statistical model that was trained on bioreactor data from 10 different clones from two different fed-batch processes. This approach led to significant improvements in prediction of consumption rates of most amino acids except alanine and glycine, providing a model that can simulate the consumption rates of amino acids based on growth rates of the cells [47, 84].

A kinetic model of the effect of bioreactor pH on CHO viable cell density, viability, apoptotic cell density, and titer production rate was developed in the literature. This model was used to simulate different static and dynamic pH shift conditions to study the tradeoffs between productivity, cell death, and culture volume. High rates of cell death can lead to issues with separation and product quality and is undesirable in operation. Large culture volumes due to excessive addition of base are also undesirable. This work simulated different pH shift conditions to evaluate a large number of process operations that would not be possible to perform via experiments alone [161]. Genome scale models provide a platform for process simulation that does not require a large amount of experimental data as they are non-parametric. One study used pathway enrichment analysis after using the CHO genome scale model on data from two fed-batch experiments that used different feed media. Fed-batch process A used a feed medium containing lower amino acid concentrations than that of fed-batch process B. Process B resulted in higher titers. The pathway enrichment analysis revealed increased fluxes in N-glycan biosynthesis, glycolysis/gluconeogenesis, valine, leucine and isoleucine degradation, TCA cycle, alanine metabolism, and amino sugar and nucleotide sugar metabolism in the fed-batch process B. Genome scale model solutions are not always reliable as the models are under determined. Hence, transcriptomics was used to confirm the results from pathway enrichment analysis. Feed B contained increased amounts of serine, lysine, threonine and tyrosine. The metabolic uptake rates from fed-batch process A were coupled with increased uptake rates of serine, lysine, threonine and tyrosine to perform FBA. The results of the FBA showed an increase in titer. This has already been confirmed by fed-batch process B. The pathway enrichment analysis revealed valine, leucine and isoleucine metabolism as targets for further media

optimization. This was tested in-silico by increasing the uptake rates of leucine, isoleucine and valine. The results from FBA showed a further increase in specific titer production rate of 19%. Since isoleucine was already overfed in the cultures, experimental validation was performed by increasing the concentration of leucine and valine in the feed medium. Experimental validation showed an increase in titer of 11.8% [162].

This demonstrates that in-silico model prediction based on pathway enrichment analysis and transcriptomic data has the potential to improve media formulations to increase process titers. The development of more detailed models for CHO cell metabolism and glycosylation that includes kinetics and the effect of bioreactor related parameters (pH, temperature, and DO) will provide a platform with more robust process simulation capabilities.

1.7.2 Bioprocess optimization

Media development is the process of optimizing the constituents of cell culture media to assist cell growth and production of the antibody. Cell culture media is expensive and consists of many components. It can be considered as the raw material for the process as a portion of its constituents eventually get converted to the monoclonal antibodies and the rest of the compounds are necessary for optimal growth of cells. There is a great deal of interest in the optimization of all nutrient concentrations in cell culture media which satisfy the nutrient requirements to sustain the growth of the cell line and maximize the production of monoclonal antibodies. Cell culture media costs are often the most significant contributor to cost associated with the upstream production process. The media cost decreases with increasing productivity [163]. Optimization of media composition leads to an increase in growth rate and titers. It has

often been observed that cell culture media that has been optimized for a certain cell line shows sub-optimal performance when used for a different cell line. This is because different cell lines consume nutrients at different rates. This has been experimentally shown in a study that used three different CHO cell lines (CHO-K1, CHO-DG44, and CHO-S). The same monoclonal antibody was expressed in all three CHO cell lines to avoid product-related effects on the host cell line. The cells were grown in ActiCHO P media and CD CHO media. It was observed that the final mAb titer in the fed-batch process in ActiCHO P media was highest for CHO-DG44 followed by CHO-K1 and CHO-S. However, the same trend was not observed when the cells were grown in fed-batch mode by using the CD CHO media. In CD CHO fed-batch cultures the highest final mAb titer was observed for CHO-K1 followed by CHO-S and the lowest final titer was observed for CHO-DG44. Thus, the specific nutrient demands of the cell lines need to be considered while designing cell culture media [112].

Genome scale models of CHO-K1, CHO-DG44, and CHO-S have been used to highlight the difference in amino acid uptake rates of the three cell lines [48]. Genome scale models have also been used to understand the metabolic costs of secretory pathways and limitations of protein secretion [164].

Utilizing mammalian cells to produce products can inherently lead to release of proteins from the cell via lysis or secretion. These proteins are called host cell proteins (HCPs). Thousands of HCPs have been identified in culture supernatant from CHO cell cultures. HCPs such as glycosidase and protease can affect product quality [165]. The genome scale model used to understand the metabolic costs of protein secretory pathway [164] was used to determine the effect of knocking out genes responsible for production of HCPs on protein production. The total HCP content in the product of various knock

out clones was analyzed and a significant reduction of 40% to 70% was observed for different cell lines. Knockout of 11 HCPs resulted in a significant reduction in the titer. However, the knockout of only 6 HCPs resulted in a significant reduction in total HCP concentration as well as an increase in titer [166].

The composition of nutrients in cell culture media has traditionally been optimized using statistical methods. Statistical methods such as the one-factor-at-a-time (OFAT) approach are very labor intensive and require a large number of experiments to optimize cell culture media. Other design of experiments-based methods such as Plackett-Burman, Central Composite and Box-Behnken have been shown to reduce the number of experiments required to identify major effects. However, there will always be a need to gather large amounts of data to use these statistical methods. The potential of using metabolic flux analysis (MFA) and flux balance analysis (FBA) to determine the optimal media for a particular process is thus being explored. MFA and FBA provide mechanistic insight into how the fluxes from intake of nutrients from the media is directed into pathways that are utilized for production of biomass and product [167]. The application of these models to develop highly specialized cell culture media that is tailored to a specific cell line has been explored in the literature [168-170]. ^{13}C-MFA has been used to identify modifications that need to be made to cell culture media to improve growth and titer of the specific cell line. It was identified that the concentrations of alanine, arginine, glutamine and glycine should be decreased, and the concentrations of methionine, tryptophan, asparagine and serine should be increased. These modifications improved the peak cell density by 55 % and increased the titer by 27 % [171].

Genome scale models have been used to identify nutrients that need to be added to cell culture media to boost titers. The Flux Variability Scanning Based on Enforced Objective Flux (FVSEOF) algorithm was used to predict strategies for increasing the titers [172]. It was identified that the presence of 15 metabolites influenced the antibody production. A Plackett-Burman-based design of experiments was performed to identify the most significant effects from the 15 metabolites and determine the optimal metabolite concentrations. It was identified that increases in arachidonate and threonine had the most significant effect on the titers. A two-fold increase in total antibody production was achieved using these metabolites as culture supplements in feed solutions [168].

The above examples look at the optimization of amino acid concentrations and sugar concentrations. However, a large portion of media consists of salts and trace metals [167]. It should also be mentioned that these models did not include the effect of bioreactor parameters that can significantly affect process outcomes. The ability of stoichiometric models to include a large number of metabolites without the need to regress parameters makes it ideal to study the effect of a large number of metabolites on cellular metabolism. This is critical for media formulation. However, it does not provide information on the effect of the concentration of metabolites in media on cellular metabolism. In a fed-batch process the optimization of feeding schedule and feed concentrations has been shown to have a significant effect on cell growth and titers. It is an industrial practice to optimize these processes by running experiments under different feeding schedules and feed concentrations to choose strategies that have the best performance. Models that can provide insight into the effect of metabolite

concentrations on cell growth and titers can be used to optimize these processes and reduce experimental efforts.

There are various examples of the utilization of these models to predict cell growth and titers based on concentration inputs to the model. The authors of one study developed a simple kinetic model that describes the effect of the concentration of glucose, lactate, ammonia, glutamine and glutamate on cell density and titer production. The model was used to maximize the titer production for a given amount of feed media and the minimal interval between feed additions was set to 4 hours. This model-based optimization problem was solved to determine the optimal feeding schedule for the fed-batch process. This was experimentally validated, and the product titers were compared with the fed-batch process used to get the parameters for the model. This optimization resulted in a 30% increase in mAb titers [20].

Another study determined the optimal concentrations of tryptophan and cysteine needed in the feed media to maximize titer production via model-based simulations and validated the results via experiments. This study utilized Monod kinetics to determine the growth rates as a function of concentrations of key metabolites. The specific uptake/secretion rates of other metabolites were expressed as a linear function of growth rate as described in another publication [76]. The model was trained on experimental data at various concentrations of cysteine and tryptophan. Model predictions on variations of cysteine and tryptophan led to development of contour plots of the mAb titer as a function of cysteine and tryptophan concentrations. The model predictions were validated by comparing the contour plot with experimental data that was not used to determine model parameters. The optimal concentrations of cysteine and tryptophan

were determined by using a generalized reduced gradient method combined with multi-start function to determine the global optimum [173].

1.7.3 Mechanistic understanding of process conditions

The QbD framework highlights the need to mechanistically understand the relationship between CMAs and CQAs. One of the most important materials used in biomanufacturing is cell culture media. The trace metals of the media (zinc, copper, manganese, etc.) are CMAs that affect cellular metabolism [174]. In order to improve mechanistic understanding of trace metals on cellular metabolism, one study performed two bioreactor experiments at low copper concentration in one bioreactor and 20X the copper concentration in the second bioreactor. The difference in energy metabolism was studied by using ^{13}C-MFA during the stationary phase of the culture. The metabolic network consisted of 106 reactions and 38 metabolites. This study shows that deficiency of Cu leads to significant redistribution of fluxes in the pentose phosphate pathway and the TCA cycle due to lower oxidative phosphorylation leading to increased lactate production [169]. Another study concluded that there is an increased need to study the effects of salts and trace metals due to the adverse impacts they can have on N-linked glycosylation [170].

A combined experimental and modeling approach has been used in the literature to increase the percentage fraction of bigalactosylated glycans on a mAb from 11.9% to 61.4%. The parameters of the PFR-based glycosylation model described in Section 1.4.1 were regressed to the glycosylation profile of a CHO cell line to get a baseline. The parameters of this model were modified to mimic an increase in nucleotide sugar production, nucleotide sugar transport to the Golgi and overexpression of enzymes in the glycosylation reactions. The authors subsequently built 11 single gene constructs

and transfected the CHO cells to overexpress each of these 11 genes individually. The only experiment that led to an increase in galactose fractions was overexpression of one of the galactosyltransferases β4GalT1. This led to a bigalactosylated fraction of 47.3%. Overexpression of the other galactosyltransferase β4GalT2 did not significantly change the glycan composition. The authors mentioned that this is most likely because this enzyme has been reported to have a higher K_m value. Hence, it might be less efficient under certain conditions compared to β4GalT1. Overexpression of the genes responsible for transport and synthesis of UDP-Gal did not significantly affect the glycan fractions. The model predicted that overexpressing galactosyltransferase led to the most significant increase in galactose fractions. The model also predicted that increasing the UDP-Gal synthesis or transport alone did not significantly affect the glycan fraction and that the nucleotide sugar synthesis or production rate is not the rate limiting step. Both predictions have been confirmed by experiments.

The authors also performed multiparameter perturbations on the model to see how overexpressing more than one gene at a time can help increase the bigalactosylated glycan fractions. The model revealed that overexpression of galactosyltransferase to a certain value led to depletion of UDP-Gal and UDP-Gal transport or synthesis is now the rate limiting step. Based on the modeling results, the authors developed a multigene construct to overexpress the galactosyltransferase β4GalT1, UDP-galactose transporter SLC35A2, galactose kinase GALK1 and UDP-sugar transporter SLC35D1. This led to an increase in bigalactosylated glycan fractions to 61.4% [116]. This publication shows how the integration of mechanistic understanding of the glycosylation process into the model can be used to design experiments to improve product quality.

Impurities such as CHO cell derived HCPs and DNA can adversely affect product quality and results in increased downstream operation costs as additional chromatography steps are required to remove these impurities. To understand and quantify where these impurities come from, a CHO cell-based population balance model that accounts for viable cell density, dead cells, lysed cells, mAb production rate, osmolarity, and cell volume has been used. This model consists of a term that quantifies the rate of production of HCP or DNA as a function of lysis of cells and secretion by viable cells. The parameters of this term will provide insights into how much HCP or DNA is secreted by viable cells vs lysed cells. The model quantified that more than half of the amount of HCP in the media is secreted by viable cells [175]. Unbalanced amino acid composition in cell culture media can result in increased production of ammonia. Ammonia is known to inhibit cell growth, decrease cell viability, affect glycosylation and reduce specific titer production rates. Addition of valine to the cultures reduced ammonia accumulation. A genome scale model was used to understand this affect. The model showed that there was an increased TCA cycle flux due to valine feeding that resulted in production of intracellular glutamate from AKG and ammonia that subsequently reduced cell specific ammonia production rates [176].

Human cell line AGE1.HN has been used to produce Alpha 1-antitrypsin (A1AT) as the A1AT derived from this cell line has similar anti-inflammatory activity compared to A1AT derived from plasma. Differences between the producer cell line and the host cell lines have been studied. The producer cell line has shown higher levels of total cellular RNA, lipids and phosphatidylcholine. A reaction network approach has been used to study the metabolic burden on the producer cell line compared to the host cell line. This network-based approach considers the indirect metabolic needs for

production and synthesis of glycoproteins, such as energy required for synthesis of mRNA, glycosylation, protein assembly, and transport reactions for precursors. The simulation revealed that increased mRNA production leads to increased secretion of glutamic acid and glycine in the producer cell line. This difference was experimentally confirmed. A comparison of MFA results from the producer and host cell lines revealed changes in intracellular metabolism of glycine, glutamic acid and lipid synthesis. These mechanistic insights can be used to adjust concentrations of amino acids in the media and incorporated feeding of lipids and lipid precursors into the cell culture media [70].

A CHO reaction network consisting of 319 proteins catalyzing 183 reactions and 188 metabolites paired with transcriptomic data has been used to identify targets for genome editing to improve glycolysis and amino acid catabolism and develop a more efficient mAb producing cell line. The CHO genome scale model was used to determine the lethality of disrupting the target genes to make sure that gene targets do not result in cell death or reduction in growth rates. Disruptions of genes related to catabolism of tyrosine, phenylalanine and glutamate resulted in an increase in growth rate and a reduction in specific production rates of lactate and ammonia [177]. Genome scale models have been widely used for strain engineering in microbial systems. However, there is limited literature on their application for cell line development of mAb manufacturing. The models in this section provided insights into the process that are not easy to determine by looking at the experimental data alone, providing valuable mechanistic insights into the process.

Table 1.5: Applications of different types of models to simulate, optimize or provide mechanistic insights into the process.

Application	Model type	Specific application details	Reference
Bioprocess simulation	Integrated model (Dynamic MFA)	An integrated model was used to simulate 9200 bioreactor runs by varying seed density, harvest day, temperature shift value and temperature shift day.	[160]
	Stoichiometric model (Genome scale)	A genome scale model along with statistical model was capable of accurately simulating amino acid uptake rates for 10 different CHO cell lines if provided a growth rate.	[84]
	Kinetic model	A kinetic model was used to simulate different static and dynamic pH shifts to study the tradeoffs between productivity, cell death and culture volume.	[161]
	Stoichiometric model (Genome scale)	Pathway enrichment analysis along with transcriptomics of two fed batch processes guided the reformulation of feed media in CHO fed-batch cultures.	[162]
	Statistical model	Uptake fluxes of nutrients at various metabolic steady states of culture were regressed to a linear function of growth rate. This was successfully used to make predictions of dynamic nutrient concentrations at 2 L and 2000 L scale.	[76]
	Statistical model (Markov chain)	A Markov chain-based N-linked glycosylation model has been used to simulate product quality by either overexpressing or knocking-down various N-linked glycosylation enzymes.	[97]
Bioprocess optimization	Stoichiometric model (Genome scale)	FVSEOF was used to identify the metabolites that impacted mAb productivity. Increase in concentration of some of these metabolites resulted in two-fold increase in titers.	[168]
	Stoichiometric model (^{13}C-MFA)	^{13}C-MFA has been used to identify two lists of metabolites. Metabolites that need to have their concentration decreased and the metabolites that need to have their concentrations increased. This led to increase in peak cell density by 55% and increase in titer by 27%.	[171]
	Kinetic model	Kinetic model for cellular metabolism was used to determine optimal feeding schedule in a fed-batch process. This results in 30% increase in mAb titers.	[20]
	Hybrid model (Statistical + kinetic model)	Statistical correlation between metabolic uptake rates as a linear function of growth rate was combined with Monod kinetics to determine optimal concentrations of cysteine and tryptophan in feed media.	[173]
	Statistical model (Genetic algorithm)	A genetic algorithm was used to guide iterations of experiments based on varying process parameters to maximize titer production. A total of four iterations with 22 experiments per iteration resulted in a 2.7 fold increase in titer.	[80]
	Kinetic model	Kinetic model that includes the effect of pH on metabolism was used to determine optimal pH shift time in CHO fed-batch cultures. The optimal pH shift strategy resulted in 40% increase in titers.	[131]
	Stoichiometric model (FBA)	CHO reaction network along with transcriptomics data has been used to determine genome editing targets to produce an efficient mAb producing cell line. Disruptions of genes related to catabolism of tyrosine, phenylalanine, and glutamate resulted in an increase in growth rate.	[177]

Table 1.5: Continued

Improved metabolic understanding	Stoichiometric model (^{13}C-MFA)	^{13}C-MFA was used to show that deficiency of Cu leads to significant redistribution of fluxes in the Pentose Phosphate Pathway and the TCA cycle due to lower oxidative phosphorylation leading to increased lactate production.	[169]
	N-linked glycosylation model (PFR-based model)	PFR-based N-linked glycosylation model was used to determine the bottle necks in the N-linked glycosylation production process. The model identified gene targets that need to be overexpressed. Overexpression of these targets led to increase in bigalactosylated fraction from 11.9% to 61.4%.	[116]
	Kinetic model	Kinetic model was used to quantify the amount of HCP or DNA secreted by viable cells and lysed cells. The model showed that more than half of the amount of HCP in the media is secreted by viable cells.	[175]
	Stoichiometric model (Genome scale)	Reduction in specific ammonia production rates was observed when valine was fed to CHO cell cultures. A genome scale model was used to explain the observed phenomena.	[176]
	Stoichiometric model (MFA)	Comparison of MFA results from producer and host AGE1.HN cell line showed differences in intracellular metabolism of glycine, glutamic acid and lipid synthesis.	[70]
	Stoichiometric model (GFA)	GFA was used to show that supplementation of manganese increased the enzymatic activity of galactose transferase in the early part of the culture but did not significantly affect it in the late stage of the culture.	[101]
	Stoichiometric model (FBA)	FBA has been used to study the effect of culture phase and bioreactor temperature on intracellular nucleotide sugar rates.	[51]

1.7.4 Recommendations to improve applicability of models

To improve the industrial applicability of models, more developments are needed in the field of biomanufacturing modeling. A few recommendations are listed below. The models discussed in Section 1.3 can be further improved by including all amino acids and the effect of trace metals on cellular metabolism. Creating a model with sugars, amino acids and trace metals as model inputs will help develop a framework that can be used to optimize nearly all media components. N-linked glycosylation is only one of the many CQAs of the mAb manufacturing process. The process conditions that affect other CQAs such as host cell protein content, mAb fragmentation, charge variants and mAb aggregation need to be better studied [178]. Improvements in mechanistic understanding of the relationship between process conditions and these CQAs will aid

with development of mathematical models that can predict these CQAs with varying process conditions. Including the effect of bioreactor related parameters such as pH, temperature, dissolved oxygen, sparging rate and scale up related parameters into mathematical models of cellular metabolism and N-linked glycosylation will help make these models more relevant to industrial scale processes. The inter-linked nature of media formulation, bioreactor operation and fed-batch scheduling make the development of a multiscale model crucial as these models can relate process parameters across different scales.

Stoichiometric models (^{13}C-MFA and genome scale models) have been used to improve mechanistic understanding and improve media formulations. However, they have not yet been used to understand the effect of pH and temperature on metabolism. Changes in bioreactor temperature or pH can result in changes in nutrient uptake rates as described in Section 1.6. Mechanistic understanding of cellular metabolism via ^{13}C-MFA studies can aid in reformulation of media to account for these changes in nutrient uptake rates. Addition of transcriptomic data can help tailor the model to specific cell lines and provide additional constraints to the genome scale models. Although transcriptomic data are challenging to collect for every experimental condition at different time points, the benefits of using such data cannot be refuted.

The majority of publications related to CHO genome scale model application for media development or cell line development have been published in the last few years. Future studies on integration of transcriptomics data and genome scale models for cell line development will aid the development of efficient CHO clones that have increased growth and titer production rates.

Downstream chromatography operations can be modeled using an open source modeling and simulation framework called Chromatography Analysis and Design Toolkit (CADET) [179]. This package makes it easy for academic and industry users to perform simulations on downstream operations. Upstream operation simulations can benefit from the development of a similar open-source software package.

Models for cellular metabolism and N-linked glycosylation can be computationally intensive, making it challenging to use these models for optimization of a process. This can be overcome by building surrogate models which use statistical models in place of the computationally intensive mechanistic models. The use of dynamic kriging and neural networks to build surrogate models for computationally expensive mechanistic models [99, 104] has been reviewed in Section 1.4. The application of surrogate models to evaluate the design space of continuous downstream chromatography operations using models that are not as computationally intensive when compared to the mechanistic models to evaluate the effect of changes in process conditions on process outcome has shown promise in the literature [180].

There is a similar need to study the effect of upstream operating conditions such as media nutrient concentrations, media feed schedule, bioreactor agitation rate, bioreactor pH, temperature etc. on the final titers. Flowsheet modeling utilizes mass balances across all of the unit operations involved in the process to identify bottlenecks related to environmental and economic impacts. These approaches have been used to compare manufacturing of mAbs in CHO cells in continuous and batch operating modes, highlight process conditions that could lead to large costs and help determine the optimal process operating conditions [181]. This approach has also been used to compare manufacturing of viral vectors in continuous and batch processes [182]. The

use of flow sheet modeling to study the economic and ecological impacts of continuous biomanufacturing has also been performed in the literature [183]. The main limitation of using flowsheet-based models for biopharmaceutical applications is the need for assumptions on product titer, media consumption rate, cell density and other process dependent outcomes. This is due to the lack of well-defined models for each unit operation. The models reviewed in this paper can overcome this limitation and lead to a more realistic techno economic analysis framework. Developments in the above-mentioned points can facilitate the creation of industrially applicable workflows for model assisted process optimization of a mAb product.

1.8 Kinetic model parameter estimation

Kinetic models for cellular metabolism and N-linked glycosylation are semi-empirical models. These models contain parameters that need to be determined from experimental data. Kinetic models can be represented by the following generalized Equation (1.1).

$$y = f(\theta, x) \tag{1.1}$$

where y represents the model output and x represents the model input. These outputs and inputs can represent the concentration of different metabolites, cell density, titer or bioreactor parameters such as pH or temperature. θ is the set of parameters that need to be estimated to make sure that the model fits the experimental data. To obtain the θ values, least square minimization is most often used to minimize the sum of squared difference between the model predicted results y_{model} and the experimental results y_{exp}. The objective function that needs to be minimized can be written as follows

$$F(\theta) = \sum_{i}^{n} \sum_{j}^{m} \sum_{t}^{T} \left(y_{exp_{i,j,t}} - y_{model_{i,j,t}} \right)^2 \tag{1.2}$$

where $y_{exp_{i,j,t}}$ and $y_{model_{i,j,t}}$ are the *j*th specific uptake/production rate of the *i*th experiment at time *t* in a set of *m* specific uptake/production rates, *n* experiments and *T* time points.

The magnitude of each value of y varies drastically. For example, the specific uptake rate of glucose in much larger than the specific uptake rate of methionine. Parameter estimation without normalizing these values can lead to a set of parameters that favors the set of *y* values with large magnitudes. To overcome this, normalization is proposed as follows.

$$y_{exp_{i,j,t}} = \frac{y_{exp_{i,j,t,original}} - \overline{y_{exp_{i,J,t,original}}}}{\sigma_{exp_{i,original}}} \tag{1.3}$$

$$y_{model_{i,j,t}} = \frac{y_{model_{i,j,t,original}} - \overline{y_{exp_{i,J,t,original}}}}{\sigma_{exp_{i,original}}} \tag{1.4}$$

where $y_{model_{i,j,t,original}}$represents the outputs that are determined from the kinetic expressions and $y_{exp_{i,j,t,original}}$represents the experimental data, and $\overline{y_{exp_{i,J,t,original}}}$ represents the average value of the data across all time points from all the experiments. The standard deviation of the original data across all time points from all experiments is $\sigma_{exp_{i,original}}$. In this way, all the data are transformed to a data format with a mean of zero and standard deviation of 1 without changing the shape of the data and the trends and the contributions of each component are weighed equally.

The transformation of the experimental and model predicted data is problem specific. Parameter estimation is complex because the objective function $F(\theta)$ can exhibit multiple local minimum values across the vast parameter search space due to the nonlinear nature of the model and the objective function. Utilizing mechanistic knowledge of biochemical processes to constrain the parameter values to experimentally reasonable values and selecting a reasonable set of initial parameters is

crucial for parameter estimation. The presence of local minimum values requires the use of global optimization algorithms along with local optimization to determine the global minimum representing the best possible set of parameters. Many different optimization methods can be used to determine the best possible set of parameters. A list of methods and models used in the literature is tabulated in Table 1.6. Global sensitivity analysis can provide important insights into the relationships between model parameters and the model outputs. It allows for parameters to be ranked based on their effects on the model output. This can be used to reduce the number of parameters that need to be estimated. In the literature global sensitivity analysis has been extensively used for parameter estimation problems [18, 21, 184]. The ideal steps that should be followed to determine the parameters of large-scale kinetic models are represented as a flow chart in Figure 1.8.

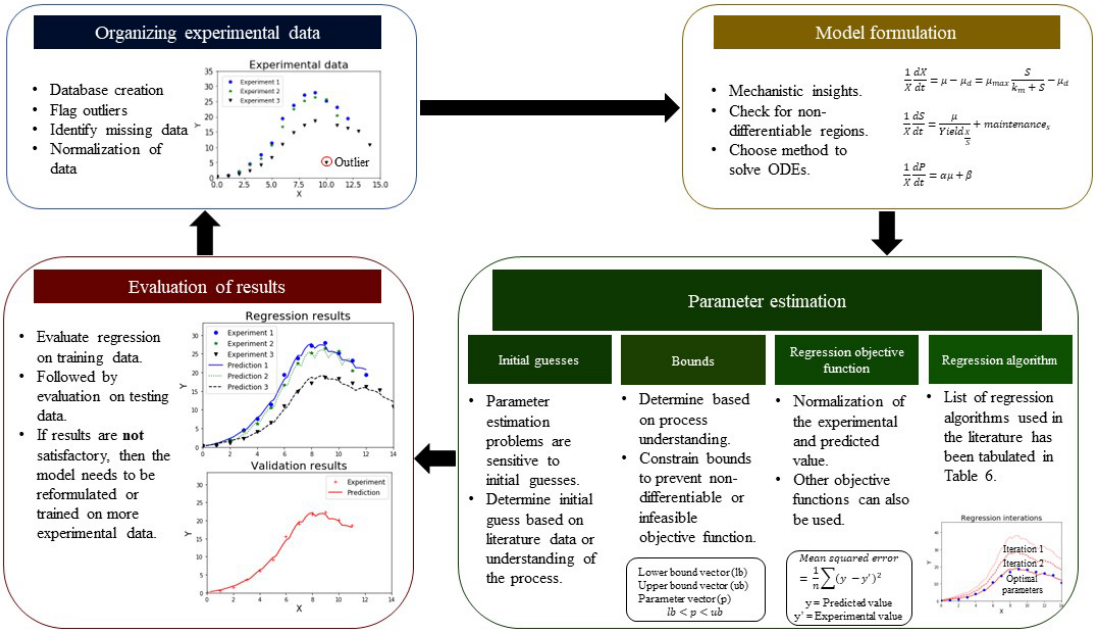

Figure 1.8: Flow chart for steps to be followed for kinetic model development and parameter estimation.

Table 1.6: Methods used in the literature to determine model parameters for different types of models with various number of parameters

Number	Model details	Method used	Reference
1	Integrated model for metabolism that consisted of 47 parameters	Simulated Annealing	[62]
2	Kinetic model for glycosylation that consisted of 24 parameters	Levenberg Marquardt	[94]
3	Kinetic model for glycosylation that consisted of 16 kinetic parameters	Maximum likelihood estimation	[95]
4	Kinetic model for metabolism that consisted of 24 kinetic parameters	Sequential Quadratic Programming	[184]
5	Kinetic model for metabolism coupled with a PFR-based kinetic model for glycosylation. The model consisted of 14 parameters	Enhanced scatter search method and fmincon solver	[73]
6	Kinetic model for metabolism and glycosylation that consisted of 41 parameters	Nelder-Mead method	[185]
7	Kinetic model for metabolism that consisted of 18 parameters	Maximum Likelihood estimation	[20]
8	Kinetic model for metabolism that consisted of 19 parameters	Nelder Mead method	[186]
9	Kinetic model for metabolism that consisted of 17 parameters	Differential evolution technique	[187]
10	Kinetic model for metabolism that consisted of 41 parameters	Genetic algorithm	[131]

1.9 Conclusions

Optimization of bioprocess development is challenging due to the interlinked nature of the subprocesses such as media formulation, bioreactor operation, bioreactor design, process control, and cell line development. There are various types of models that can be applied for optimization of a particular subprocess. Using existing models for process intensification is challenging as models designed to optimize a subprocess need to be connected together to optimize the overall process. Hence, there is a need to combine these models together to develop multi-scale models that encapsulate the effect of more factors that are needed to optimize the whole process. However, to develop these complex models, mechanistic understanding of each process parameter is important to accomplish this task. One of the challenges of building and applying these models is parameter estimation. One section was dedicated to explaining the challenges encountered while estimating the model parameters and the different techniques that can be used to overcome these challenges. Another challenge is the lack of literature on incorporating the effect of bioreactor-based parameters such as pH, temperature, and dissolved oxygen levels, likely due to the high cost of bioreactor experiments required to generate data needed to develop these models. Improvements and widespread availability of miniaturized high throughput bioreactor systems will improve capabilities to generate large amounts of data that can be used to train and develop better models for metabolism and glycosylation. Generation of large datasets through these high throughput systems will pave the way to use statistical data-driven models along with mechanistic models.

We discussed different types of models that can be used to describe cellular metabolism and N-linked glycosylation. A list of models that have contributed significantly to the field of modeling cellular metabolism and N-linked glycosylation of

mammalian cells are listed in Table 1.7. The use of these models to gain mechanistic insights into processes has been shown. The potential of these models to optimize processes for mAb production while reducing experimental efforts has been highlighted using different examples. The effect of each of these bioreactor parameters on cellular metabolism and N-linked glycosylation has been discussed along with the efforts to model these bioreactor parameters.

These modeling techniques are not restricted to mAb production processes. Viral vector, cell therapy and vaccine manufacturing can also benefit from improved modeling approaches and model-based optimization. Incorporation of bioreactor parameters into these models will make them more applicable to industry by providing process level insights. In order to consistently improve CAR-T cell culture performance, it is important to carry out more detailed optimization of cell culture media and bioreactor operating conditions with special emphasis on DO, pH, and biomechanical forces, three important parameters affecting T cell expansion, receptor expression and activation [188-192]. As such, predictive modeling of these effects is an unmet need. CHO cell-based modeling efforts (especially those including process parameters such as DO, pH, and biomechanical forces) can prove useful in developing optimal culture conditions for CAR-T therapies to help increase cell proliferation and reduce bioreactor run times. Improving our understanding of T cell metabolism via ^{13}C-MFA studies can also provide important insights for improving the CAR-T culture performance.

Table 1.7: Timeline of important publications on modeling of biomanufacturing.

Year	Description	Reference
1995	Intracellular flux analysis of hybridoma cells using ^{13}C measurements from Nuclear Magnetic Resonance (NMR)	[27]
1997	First model for N-linked glycosylation. The model approximated the Golgi apparatus as a series of CSTR.	[93]
2005	The CSTR-based model was expanded to 7565 glycan structures and 22871 reactions by including more enzymes.	[94]
2011	The Golgi apparatus was modeled as a PFR. The nucleotide sugar transport into the Golgi was also modeled.	[95]
2011	Dynamic metabolic flux analysis-based model for CHO cell culture.	[62]
2013	Parallel labeling strategy to perform ^{13}C-MFA on CHO cells.	[35]
2015	Application of kinetic model to optimize feed addition time in mammalian cell culture to increase mAb titers.	[20]
2016	Genome scale model for CHO cells was published.	[42]
2017	Application of kinetic model for metabolism and PFR-based model for N-linked glycosylation to a perfusion bioreactor system while accounting for the addition of galactose and manganese to control N-linked glycosylation.	[73]
2019	Application of N-linked glycosylation model to improve the glycosylation fractions.	[116]
2019,2020	Application of genome scale model to improve CHO cell fed-batch process by providing insights into media formulation.	[193]
2021	Artificial neural network-based model for N-linked glycosylation.	[67]

The potential of mathematical modeling to improve bioprocess development has been reviewed in Chapter 1. A detailed survey of the literature shows that the major limitation of current mathematical models is the inclusion of bioreactor related parameters such as pH, DO, pCO_2, and temperature while predicting a large number of key bioreactor performance metrics, such as cell density, mAb titer, glucose concentration, lactate concentration, ammonia concentration, amino acid concentration, and the N-linked glycosylation fractions. The literature survey also showed the advantages and limitations of current modeling technologies, thus providing us with the knowledge to develop and choose the appropriate modeling tools to build a predictive mechanistic mathematical model that includes the effect of bioreactor pH on CHO cell metabolism and N-linked glycosylation.

Development of mathematical models requires experimental data. Hence, Chapter 2 focuses on characterizing the effect of bioreactor pH on CHO cell metabolism and N-linked glycosylation via rigorous experimentation. CHO cells were grown in bioreactors at different bioreactor pH set points. Measurements of concentrations of glucose, lactate, amino acids, ammonia, mAb titers, and viable cells led to the development of an experimental dataset. The uptake and secretion rates from this data set were utilized to constrain a stoichiometric model and perform flux balance analysis to understand the impacts of bioreactor pH on CHO cell metabolism. A survey of the literature also showed that there exist very limited studies on the impact of process conditions on site-specific N-linked glycosylation of glycoproteins with multiple N-linked glycosylation sites. The mAb used in this dissertation contains N-linked glycosylation sites in the Fc and Fab region. Hence, glycopeptide mapping was used to

quantify the effect of bioreactor pH on the distinct Fab and Fc glycan fractions of this mAb.

There exist several N-linked glycosylation models as reviewed in Section 1.4. However, none of these models account for site-specific N-linked glycosylation that can fundamentally arise from N-linked glycan processing in the Golgi. Thus, Chapter 3 was dedicated to the construction of a site-specific N-linked glycosylation model. Applications of this model were demonstrated by utilizing it to study the complex impacts of bioreactor pH on Fab and Fc glycosylation and the impacts of feeding glycosylation modulators (galactose and $MnCl_2$) on Fab and Fc glycosylation.

Mammalian cell culture performance can be affected by the concentrations of many nutrients. A review of the literature (Section 1.6.2) showed that the uptake rates of these nutrients can be impacted by bioreactor pH. However, literature mathematical models do not consider these effects. Hence, Chapter 4 was dedicated to the development of a DMFA model that is capable of predicting the concentrations of viable cells, glucose, lactate, ammonia, mAb and amino acid as a function of bioreactor pH, seeding cell density, media composition, and feed media supplementation schedule. This model was trained on fed-batch data generated in Chapter 2. Flowsheet models that are used to compare fed-batch and perfusion bioreactors process economics suffer from the drawback of utilizing mass balances to determine the amount of raw material required and mAb titer concentrations. This is due to the lack of reliable mathematical models to describe the performance of fed-batch and prefusion bioreactors. Hence, we decided to utilize the DMFA mathematical model (trained on fed-batch data) to make predictions of performance of intensified fed-batch and perfusion cultures. The successful experimental validation of these predictions demonstrates that the DMFA

model built in Chapter 4 is capable of predicting cell metabolism under different conditions of bioreactor pH, seeding cell density, media composition, and feed-media composition.

The end of the dissertation (Chapter 5) focuses on providing recommendations to improve the mathematical models built in this dissertation along with potential applications as digital twins that can be used to monitor, derisk, or optimize biotherapeutic process development.

Chapter 2

FLUX BALANCE ANALYSIS AND PEPTIDE MAPPING ELUCIDATE THE IMPACT OF BIOREACTOR pH ON CHINESE HAMSTER OVARY (CHO) CELL METABOLISM AND N-LINKED GLYCOSYLATION IN THE Fab AND Fc REGIONS OF THE PRODUCED IgG

2.1 Introduction

Process intensification involves determining the best bioreactor operating conditions to maximize mAb titers while maintaining product quality. It is a widespread practice in the industry to shift the temperature of the culture to maximize production of the mAbs [119-122]. However, literature studies have also used pH shifts to maximize the titer and control product quality [12, 131]. Scaling up the production to large scale bioreactors shows that gradients in pH within the bioreactor can significantly affect the metabolism and N-linked glycosylation of the cells [137, 138]. Bioreactor pH is known to affect titer production rates, growth rates and nutrient uptake rates [118, 126]. These differences in uptake rates of nutrients can lead to suboptimal media performance at certain pH values. Studies have shown that culture pH can affect N-linked glycosylation of biologics [126]. More specifically it has been shown that bioreactor pH can affect the galactosylation of glycoproteins [158, 159]. However, the effect of critical process parameters (pH, temperature, and DO) on site specific N-linked glycosylation of mAbs has not been reported in the literature. This is important because about 20% of the IgG in the human serum contain glycans in the Fab region and the Fc region [108, 109].

The effect of Fc glycosylation on mAb affinity, half-life, aggregation, and thermal stability have been well studied over the past few decades [194]. Recent studies also show that Fab glycans can influence mAb affinity, activity, half-life and aggregation [108]. One study introduced an N-linked glycosylation site in the Fab region of a mAb by modifying the amino acid sequence. It was shown that mAbs with glycans in the Fab region had increased half-life in mice and showed similar efficacy to the original molecule [195]. Another study looked at thermal stability of mAbs with and without Fab glycans and showed that the presence of Fab glycans improves thermal stability of antibodies [107]. Hence there is the need to control Fab N-linked glycan profiles and treat them with the same importance that we treat the Fc glycan fractions. From a Quality by Design (QbD) perspective, it is important to understand the impact of process parameters on site specific N-linked glycan fractions. Although this study focuses on Fab and Fc glycans, the importance of studying the effect of critical process parameters on site specific N-linked glycosylation can extend to any glycoprotein that has several N-linked glycosylation sites.

To study the effect of bioreactor pH on CHO cell metabolism and site-specific N-linked glycosylation, CHO VRC01 cells producing a broadly neutralizing anti-HIV antibody were cultivated in bioreactors at three different pH conditions. Measurements of glucose, lactate, amino acids, titers, ammonia, and site-specific N-linked glycan fractions at different timepoints allowed us to study the effect of bioreactor pH and culture duration on CHO cell metabolism and product quality. A survey of the literature showed that there are limited number of studies that utilize stoichiometric models to study the effect of bioreactor critical process parameters (pH and DO) on cellular metabolism [196]. In the literature, parsimonious flux balance analysis (pFBA) along

with constraining the maintenance energy has been used to study the metabolism of CHO cells [44]. Another study has used FBA to study the effect of bioreactor temperature shifts on CHO cell metabolism [51]. Measurements of glucose, lactate, amino acids, titers, and ammonia have allowed us to use a similar approach to explain the differences in metabolic uptake/production rates observed at the various pH values and provide insights into intracellular metabolism. However, using an underdetermined system of reactions to make conclusions on the effect of pH on cell metabolism can be tricky, due to the existence of multiple possible solutions [197].

To overcome this challenge, flux variability analysis (FVA) has been used to study the range of possible fluxes. This study utilized experiments and stoichiometric models to study the differences in cell metabolism at different pH conditions and culture durations. The results presented in this study reveal the drastic differences between glycans observed in the Fab and Fc regions of the antibody. It was found that the changes in overall glycosylation indices as a function of bioreactor pH were significant. Changes in glycosylation indices in one site do not always indicate a similar change in the same glycosylation index on the other site, hence showing that the N-linked glycan fractions across the two sites are not uniformly affected by changes in bioreactor pH. This is an unexpected and novel finding. Improving our knowledge of this can help control the site-specific N-linked glycan fractions and aid with implementing quality by design for glycoproteins with multiple N-linked glycosylation sites.

2.2 Materials and methods

2.2.1 Cell culture

CHO-K1 VRC01 cell line (Clone A11 from the Vaccine Research Center at the National Institute of Health) stably expressing a broadly neutralizing anti-HIV monoclonal antibody with Fab and Fc glycosylation was used in this study. CHO cells were grown in 125 mL Erlenmeyer flasks at 30 mL culture volume for passage and inoculum preparation. HyClone Actipro media (Cytiva) supplemented with 6 mM L-glutamine was used as the basal media. HyClone Cell Boost 7a supplement and HyClone Cell Boost 7b supplement were used as feed media. CHO cells were grown in 1 L Eppendorf BioFlo 120 bioreactors at an initial culture volume of 750 mL. The bioreactor was operated at a temperature of 37 °C, dissolved oxygen setpoint of 40 % and agitation set to 90 RPM using a pitched blade impeller. Three different bioreactor pH values (6.75, 7, and 7.25) were studied by running biological triplicates for each condition. The bioreactor pH was controlled by sparging with CO_2 and pumping 6% sodium bicarbonate solution. A significant amount of base was required only for the pH 7.25 condition due to higher lactate accumulation. The feed 7b medium has a basic pH to improve solubility of tyrosine. Thus, sodium bicarbonate addition was not necessary for the pH 7 condition. The dissolved oxygen was controlled by sparging with oxygen and air.

Starting on day 3 of the culture, feed media was added daily: 22.5 mL of HyClone Cell Boost 7a and 2.25 mL of HyClone Cell Boost 7b. Starting on day 5, glucose was added to the bioreactor every day to bring the glucose concentration to 9 g/L. Antifoam C was added when foaming was observed.

Samples were collected daily for measurements of cell counts, glucose concentration, lactate concentration, ammonia concentration, amino acid concentration and titer. Samples were collected every 2 days starting from day 4 for site specific glycan analysis. Cultures were terminated when viability dropped below 80%. The viability dropped below 80% on day 11 or 12 for the bioreactor runs at pH 7 and pH 7.25. Thus, site specific glycan analysis has been compared only from day 4 to day 10 for these two conditions. The titers on day 4 were very low for bioreactor run at pH 6.75. Thus, site specific glycan analysis was performed from day 6 to day 14 for this condition. Cell counts and viability were measured by using the trypan blue assay on the DeNovix automatic cell counter.

2.2.2 Measurements of substrate, metabolite, and IgG concentrations

Measurements of glucose and lactate were performed using the YSI 2900 bioanalyser. Ammonia concentrations and osmolarity were measured by using the NOVA bioanalyser. Amino acid concentrations were measured using an Agilent HPLC 1260 instrument. Standards and a HPLC column (AdvanceBio Amino Acid Analysis column) were purchased from Agilent. The amino acids were derivatized with OPA for primary amino acids and FMOC for secondary amino acids as per the protocol provided by Agilent. Derivatization was performed on the autosampler. Antibody titers were measured on the Agilent HPLC 1260 using protein A chromatography using a POROS A HPLC column (Catalog number, 1502226; ThermoFisher). The method was performed in accordance with the manufacturer's protocol. Mobile phase A consisted of 50 mM phosphate and 150 mM NaCl at pH 7.0. Mobile phase B consisted of 12 mM hydrochloric acid at pH 1.9. The gradient of mobile phases is in Supplementary Table A2. A UV detector (280 nm) was used to detect the mAbs. Additional method details

are provided in the supplementary material. HPLC grade IgG standard (Catalog number MFCD00163923) was purchased from Sigma-Aldrich.

2.2.3 Site-specific glycan analysis

Starting day 4, samples (8 mL) were taken every 2 days for site-specific glycan analysis. Due to the laborious and expensive nature of performing site specific glycan analysis, it was performed for biological duplicates. Protein A purification was performed on these samples using an AKTA Pure and protein A HiTrap mAb select column (catalog number 29497628) purchased from Cytiva. These samples were further concentrated using 10 kDa centrifugal filters. 1 mg of mAb was extracted by removing the appropriate volume and vacuum evaporation. 1 mg of mAb was not available on day 4 for duplicate bioreactors run at pH 6.75 because of low titer values. Hence, samples starting on day 6 were used for pH 6.75 bioreactor condition. Duplicate samples were also not available for bioreactors run at pH 7.00 for day 4 and day 8 of the cultures. Hence single samples were used only for these two days. Duplicate samples were used for all the other conditions and timepoints.

Trypsin digestion and Glu-C digestion were performed on the mAb. LC-MS was performed using a Waters BioAccord with ACQUITY Premier system equipped with a Waters BEH C18 column (300Å, 1.7 μm, 2.1 mm ×150 mm, catalog number 186003687). Trypsin and Glu-C (ammonium bicarbonate, pH 7.8) were selected as digest reagents in the analysis method. An absolute retention time mass tolerance of 0.1 min, and target identification mass tolerance of 10 PPM (20 PPM for fragment match tolerance) were specified. Identification of at least five primary fragment ions was required for confirmation. The glycan modification library used for data processing is given in Supplemental Table A3 and A4. Additional LC-MS method parameters are

listed in Supplemental Table A5. UNIFI software and the modification library were used to determine the glycan structures at both sites. The glycosylation indices were calculated separately for the Fc region data and the Fab region data based on equations from the literature [198]. The fucosylation index is calculated by dividing the number of fucosylated glycans (F1) by the sum of the fucosylated glycans (F1) and the afucosylated glycans (F0). This is shown in Equation (2.1). Only biantennary glycans were observed on this mAb. Hence, the galactosylation index (Equation (2.2)) was calculated by dividing the sum of galactose in G2 (galactose is present in both the branches) and G1 (galactose is present on only one branch) species by the total number of potential galactose fractions that can be present in a biantennary glycan (two galactose molecules per branch). G0 represents glycans without galactose. Similarly, S2 represents sialic acid molecules on both the branches, S1 represents sialic acid present on only one branch and S0 represents the absence of sialic acid. The sialylation index (Equation (2.3)) is defined as the total molecules of sialic acid present divided by the total possible number of sialic acid residues. Biantennary glycans contain two branches and can contain a sialic acid molecule on each branch.

$$Fucosylation\ index\ \% = \ 100 \times \frac{F1}{F0 + F1} \tag{2.1}$$

$$Galactosylation\ index\ \% = \ 100 \times \frac{2 \times G2 + G1}{2\ \times G2 + 2\ \times G1 + 2\ \times G0} \tag{2.2}$$

$$Sialylation\ Index\ \% = \ 100 \times \frac{2\ \times S2 + S1}{2\ \times S2 + 2\ \times S1 + 2\ \times S0} \tag{2.3}$$

2.2.4 Parsimonious Flux Balance Analysis (pFBA)

Parsimonious FBA was performed by using a literature network consisting of 144 reactions and 101 metabolites [199]. This model has been used to study intracellular

metabolism at different culture phases. It provides an alternative to genome scale models but contains relevant metabolic information to study CHO cell metabolism. Growth rates, uptake and production rates of amino acids, glucose, cells, titer, lactate, and ammonia have been calculated by plotting the Integral Viable Cell Density (IVCD) vs the cumulative amount consumed or produced [51]. Glutamine degradation rate has been reported for the Actipro media in the literature [43]. The uptake rate of glutamine by the cells is equal to the experimentally observed rate of change in glutamine concentrations divided by the number of cells subtracted by the glutamine degradation rate reported in the literature. Degradation of glutamine results in production of ammonia and pyrrolidonecarboxylic acid [200].

The ammonia production rate was calculated as the experimentally observed change in ammonia concentration divided by the number of cells added to the rate of glutamine degradation. These calculations were made for three different phases of the fed-batch process at the three different pH conditions. The culture durations were split into early growth phase (day 0 to 3), late growth phase (day 4 to 7) and stationary phase (day 8 to 11). The data from biological triplicates are tabulated in supplementary Table A1. A two-tailed T-test was performed to determine statistical significance between the uptake and secretion rates across the different pH conditions. The following metric is used to depict statistical significance in plots of uptake rates. P values below 0.001 are represented as ***, p values below 0.01 are represented as ** and p values below 0.1 are represented as *.

A survey of the literature showed that biomass maximization is the most used objective function to perform flux balance analysis during the growth phase of cultured cells [43, 44, 51, 199, 201]. In fed-batch conditions in which the growth rates, nutrient

uptake rates, byproduct secretion rates, and titer production rates are constrained to the experimentally measured value, biomass maximization objective function has been used in the growth phase of the cultures and maximization of specific titer production rates has been used in the stationary phases of culture [51, 193, 202]. Hence, biomass maximization was used as the objective function for the early growth phase and late growth phase.

Maximization of specific titer production rates was used as the objective function for the stationary phase. The lower bounds and upper bounds were determined based on the standard deviation of the experimental measurements for all the measured fluxes. The bounds for all the reversible and irreversible reactions were set based on the information from the literature [199]. The bounds of fluxes for metabolites that were not measured were set at ±30% of the values reported in the literature [203]. Flux Variability Analysis was also performed to study the range of possible fluxes in each condition with optimality factor set to 0.95 and pFBA factor set to 1.1. pFBA and FVA were performed using cobrapy [204]. Solutions to underdetermined systems are determined using pFBA. However, there are infinite solutions that are possible for a range of objective functions. Hence, FVA was performed for all the different objective functions used.

2.3 Results and discussion

2.3.1 Bioreactor pH impacts cell growth rates and peak cell densities

The effect of bioreactor pH on viable cell density (VCD) and growth rates are shown in Figure 2.1A & 2.1B, respectively. The highest peak VCD of 28.8 million cells/mL was achieved at pH 7, followed by 27.1 million cells/mL at pH 7.25 and 18.2

million cells/mL at pH of 6.75. The statistical test performed on the data from biological triplicates at the three pH conditions show that the peak VCD at pH 7 was not significantly higher than the peak VCD at pH 7.25 (p value > 0.05). However, the peak cell densities at pH 7 and pH 7.25 were significantly higher than the value observed at pH 6.75 (p value < 0.001).

The fed-batch culture has been divided into three phases. According to the fed-batch feeding protocol in the literature for this cell line, the cells were cultivated in Actipro media (basal media) with 6 mM glutamine, followed by feeding Cell Boost 7a and Cell Boost 7b. Glutamine was not added with the feed media. [205]. High growth rates were observed in the first few days (0 to 3; early growth phase) of the culture. This can be attributed to uptake of glutamine. Figure 2.1D shows that glutamine was depleted on day 3 of the cultures. Depletion of glutamine leads to the second phase of the culture. For all three pH values, during the second late-growth phase of the culture, nutrients (except glutamine) were not limiting. The carbon source in the late growth phase is glucose but isotope labeling experiments from the literature have also shown that asparagine can drive TCA cycle fluxes in the absence of glutamine [36]. For the pH 7 and pH 7.25 conditions, the VCD started dropping when asparagine was depleted. Asparagine was added in the feed media, but the amount added could not sustain the growth of cells at such large cell densities. Asparagine (Figure 2.1G) was depleted before the measurement at the next 24-hour time point. Depletion of both asparagine and glutamine has been shown to reduce growth rates by 90% [198].

For pH 7 and 7.25, depletion of nutrients led to the third stationary-culture phase. Figure 2.1J shows the plot of osmolarity for all the conditions. In the case of pH 6.75, cell death can be attributed to the increase in osmolarity of the culture. Increase in

osmolarity from 320 mOsm per kg to 500 mOsm per kg in CHO cell cultures has been shown to reduce growth rates drastically [206]. The reasons for the increase in osmolarity in the pH 6.75 condition are discussed later. In the early growth phase, the highest growth rate (Figure 2.1B) was observed for the pH 7 condition, followed by pH 7.25 and pH 6.75.

The growth rates in the late growth phase were similar for all pH conditions. It was surprising to see that growth rates were only affected during the early growth phase. This could be due to the time taken for rewiring of metabolism to adjust for the difference in pH conditions. The peak VCD was observed on different days for all three pH conditions due to differences in growth rates in the early growth phase. Cultures were terminated when cell viability reached 80%. This occurred on different days for the different pH conditions. The differences in viability profiles seen in Figure 2.1C can be attributed to nutrient depletion, osmolarity increases, and the potential production of growth inhibitory metabolites. The viability dropped below 80% between day 11 and 12 for pH 7 and pH 7.25. For pH 6.75 viability dropped below 80% on day 14.

As stated, the feeding media do not contain glutamine. We tested the impact of adding glutamine to the feeding media in shake flask experiments. We found that it resulted in a significant increase in ammonia accumulation and drop in cell viability.

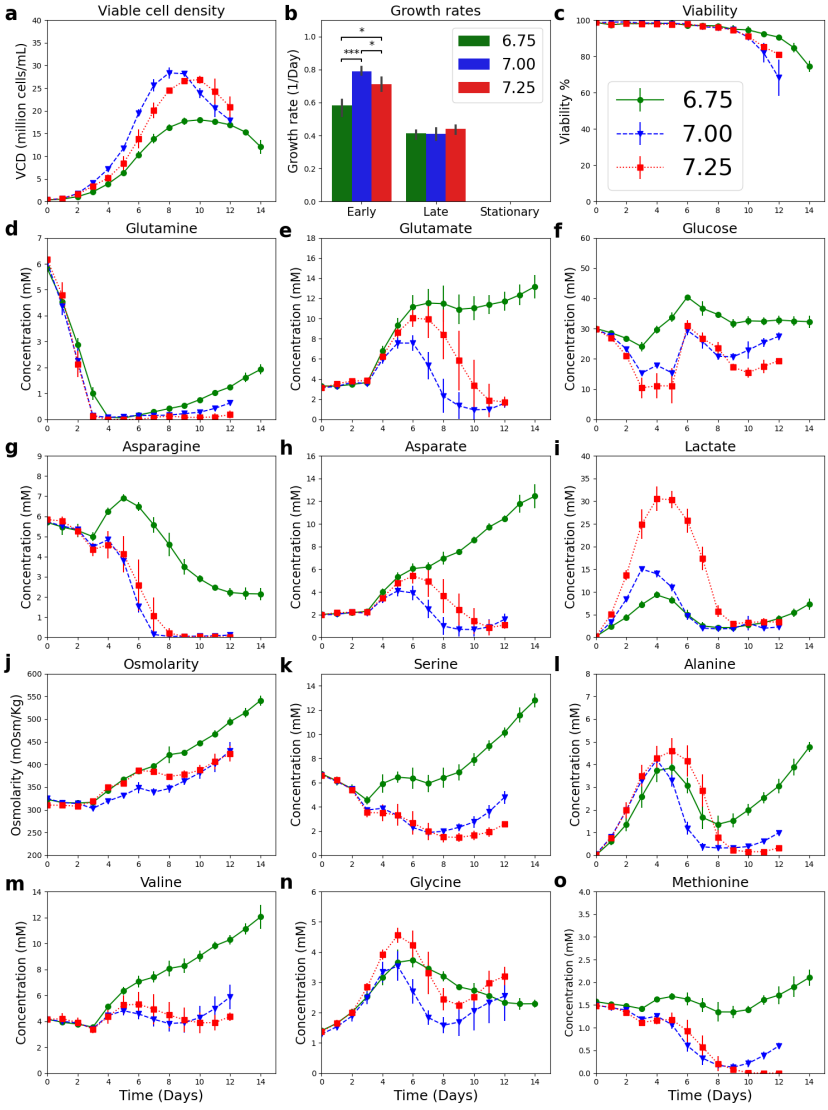

Figure 2.1: Bioreactor pH affects nutrient depletion and osmolarity thus impacting growth rates, peak viable cell density, time to peak viable cell density, harvest day, cell viability, amino acid concentrations, glucose concentrations and lactate concentrations. Data shown for biological triplicates. A) Viable cell density, B) Growth rate during the early growth phase (day 0 to 3), late growth phase (day 4 to 7), and stationary phase (day 8 to 11), C) Viability, D) Concentration of glutamine, E) Concentration of glutamate, F) Concentration of glucose, G) Concentration of asparagine, H) Concentration of aspartate, I) Concentration of lactate, J) Osmolality, K) Concentration of serine, L) Concentration of alanine, M) Concentration of valine, N) Concentration of glycine, and O) Concentration of methionine.

2.3.2 Higher bioreactor pH leads to higher rates of glucose uptake and lactate production

Figure 2.1F and Figure A2 show that higher bioreactor pH results in a statistically significant increase in cell specific glucose uptake rates. The bioreactor at pH 7.25 showed this trend throughout the culture duration. This trend has been observed by other literature studies [126, 207]. Increased glycolysis flux can be explained as a function of glycolytic enzyme activities and intracellular cytosolic pH. Glycolytic enzymes have been shown to have increased activity at alkaline pH [208]. Accumulation of intracellular pyruvate coming from glycolysis can lead to production of alanine and lactate as observed in Figures 2.1I and 2.1L. Figure 2.1I and supplementary Figure A2B show that increasing bioreactor pH leads to an increase in cell specific lactate production rates. A switch in lactate metabolism, from production to uptake, has also been reported to be caused by depletion of glutamine [209].

As glutamine enters cell metabolism through the TCA cycle, its depletion can result in reduced TCA cycle fluxes [209]. For pH 7.00 and pH 6.75, the switch in lactate metabolism occurred exactly when glutamine was depleted. Lactate is also known to reduce the pH of the media. Significant accumulation of lactate will require addition of base to control the bioreactor pH. The bioreactor operated at pH 6.75 did not require addition of base as small amounts of lactate were produced. The bioreactor operated at pH 7.00 produced a significant amount of lactate. However, addition of the feed media resulted in an increase in pH. Feed 7b has a high pH value as it is required to dissolve tyrosine. No additional base was required for bioreactor operated at pH 7.00. Bioreactor operated at pH of 7.25 required addition of a significant amount of base, the consequence of which is an increase in osmolarity of the culture as can be observed in

Figure 2.1J. If it is desired to operate at high pH conditions, methods to control glycolysis fluxes would result in better culture performance.

2.3.3 Bioreactor pH affects amino acid metabolism and mAb production

Bioreactor pH had a significant effect on uptake and production rates of different amino acids. Glutamine was rapidly consumed by the cells and had the highest uptake rate among all the amino acids. Glutamine is known to be an alternative energy source for rapidly dividing cells [210]. This rapid consumption of glutamine could lead to increased growth rates in the early phase of the culture (Figure 2.1). Glutamine supplementation in CHO cells has also been shown to increase the alanine and lactate fluxes [36]. Since glutamine can produce alanine, lactate, aspartate and glutamate, its depletion could have resulted in a switch in metabolism of these metabolites. During the glutamine consumption phase (Figure 2.1), lactate, aspartate, glutamate, and alanine were produced. After depletion of glutamine, these metabolites were consumed. An exception to this is the pH 7.25 condition that had increased glycolysis rates that led to a delay in switch of metabolism of lactate and alanine. After glutamine is depleted, glucose is the main driver of TCA cycle fluxes. Parsimonious flux balance analysis has been used to understand this metabolism in Section 2.3.4.

Bioreactor pH also significantly affected the metabolism of other non-essential amino acids (Figure 2.1 and A2). The metabolism of alanine, glycine and serine is strongly linked to glycolytic fluxes. Glycine, which is not added in the feed media, was produced and consumed by the cells throughout the culture. Glycine is known to be produced from serine in CHO cells [203]. This serine could come from glycolysis or due to serine uptake from the media. Hence, the reasons for switch in glycine metabolism could be attributed to reduction in glucose and serine uptake rates in the

different phases of culture. Accumulation of intracellular pyruvate can also result in production of alanine. When comparing bioreactor pH 7.25 to the other pH conditions, unlike lactate, we did not observe a significant increase in alanine production rates early in the culture. The reasons for lack of increased alanine production rates with increased bioreactor pH are explored by using pFBA in Section 2.3.4. We noticed a delayed switch in alanine metabolism as a function of increased glycolytic fluxes at higher pH and delayed glutamine depletion at pH 6.75. Serine uptake rates did not vary drastically among the three pH conditions, but they did vary significantly with culture duration. Literature studies have reported serine uptake rates in the range 1.9 fmol/cell/hour to 17 fmol/cell/hour [35, 207, 211]. The serine uptake rate here was 31 fmol/cell/hour during the early phase of the culture. High serine fluxes are examined in detail in Section 2.3.4.

Essential amino acid metabolism was strongly correlated to cell growth for all pH conditions (Figures 2.1, A1, and A2). Differences in growth rates resulted in different nutritional needs of the cells. Changing the feed amounts could have resulted in improved performance across the different pH conditions. For example, delaying feed addition time and reducing the amount of feed added for the pH 6.75 condition could have led to improved osmolarity of the culture. This could have extended and improved its growth rates. Bioreactors operated at pH 7.25 showed increased amino acid consumption rates towards the end of the culture (Figure A2), and this led to depletion of methionine (Figure 2.1). The data show that optimal feeding patterns for one pH condition are not optimal for another condition. Thus, it is important to recalibrate the fed-batch process and media with changes in bioreactor pH as the uptake rates are affected by pH.

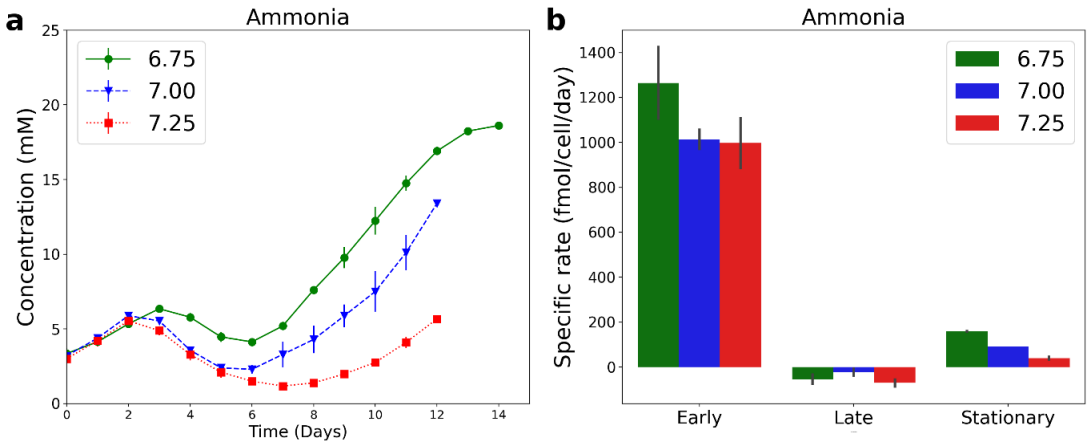

Figure 2.2: Reduction of bioreactor pH leads to accumulation of ammonia. Ammonia metabolism undergoes multiple shifts in production and consumption rates during fed-batch cultures. Data shown for biological duplicates. A) Concentration of ammonia and B) Cell specific uptake/production rates of ammonia.

Ammonia is a byproduct of CHO cell metabolism due to amino acid catabolism, and notably that of glutamine and asparagine. Ammonia accumulation inhibits cell growth [212, 213] and affects the cell metabolism and the N-linked glycosylation of proteins [127, 212]. The effect of ammonia on N-linked glycosylation can be attributed, in part at least, to its effect on the pH in the Golgi [214]. As each N-linked glycosylation enzyme has an optimal pH [73]. Hence, it is desirable to design the process aiming to minimize ammonia production. Reduction in glutamine and asparagine concentrations in the media has been shown to reduce the ammonia accumulation in the literature [36, 215]. Cell specific ammonia production rates were very high for the first few days of the culture (Figure 2.2B). The second culture phase led to depletion of glutamine, and consumption of glutamate and aspartate. Depletion of glutamine could have resulted in reduced ammonia production. Intracellular production of glutamine and asparagine

from glutamate and aspartate requires one ammonia each. This could result in consumption of extracellular ammonia as observed in the experimental data during the second phase of the culture. It is interesting that bioreactor pH significantly affected ammonia production rates in the third phase of the culture. Lower pH leads to an increase in specific ammonia production rates and increase in extracellular ammonia. When ammonia concentrations increase drastically, it has been reported that alanine production is observed [216]. Alanine was produced in the early phase of the culture and consumed from day 4 onwards. However, metabolism of alanine went through another switch at pH 6.75. On day 8 of the culture, when ammonia concentration was 10 mM, the cells began producing alanine. In the bioreactor run at pH 7.00, on day 10, when ammonia concentration reached 10 mM there was a switch in alanine metabolism and cells began producing alanine. Ammonia concentrations never exceeded 10 mM at pH 7.25, and thus a shift in alanine metabolism was not observed. The bioreactor run at pH 7.25 had significantly higher amino acid consumption rates but surprisingly lower ammonia production rates. This suggests that the amino acids were used for protein synthesis, thus preventing ammonia production.

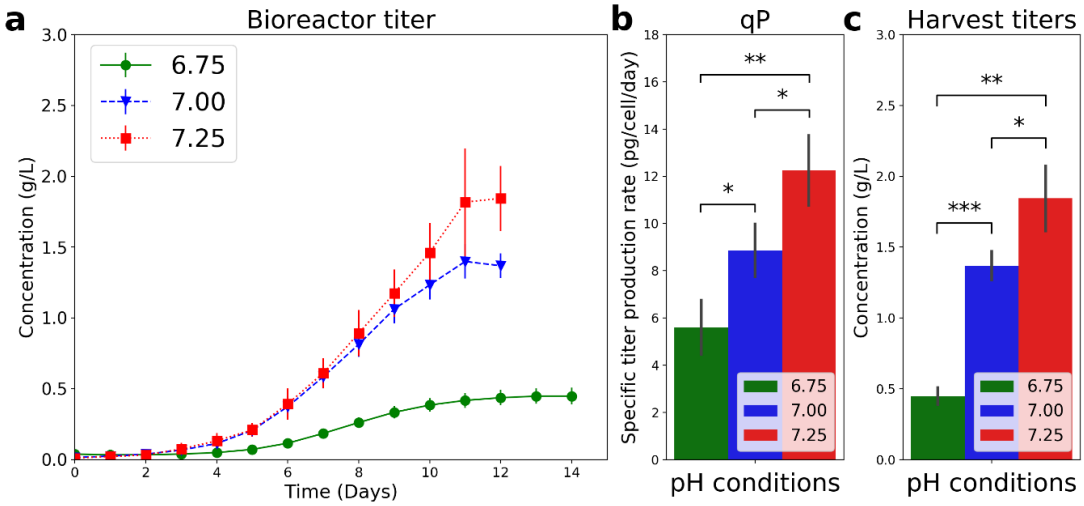

Figure 2.3: Bioreactor pH affects cell specific antibody production rates. Changes in cell-specific antibody production rates and the viable cell density led to differences in mAb titers. Data shown for biological triplicates. A) Concentration of mAb, B) Cell specific titer production rates (qP) and C) mAb titers at harvest.

Figure 2.3 shows that increase in bioreactor pH led to increased mAb concentrations and increased specific titer production rates. Using transcriptomic and proteomic tools, it was shown that, in the range of 6.9 to 7.1, higher pH leads to upregulation of secretory pathways [132]. This upregulation can be used to explain the increase in mAb production rates. It can be observed that the best cell specific mAb production rate (qP) is not necessarily observed at the optimal growth rate. Thus pH-shift strategies can be explored for optimizing cell growth and mAb production.

2.3.4 Elucidating the effect of bioreactor pH and culture duration on cellular metabolism using parsimonious flux balance analysis (pFBA)

Stoichiometric models that are constructed from knowledge of metabolic pathways can be used to study the intracellular metabolism of the cell. These models

have been used to study the effect of sparging stress, culture duration, temperature, and media composition on CHO cell metabolism [22, 63, 201, 217]. However, there are limited studies in the literature that use these models to study the effect of bioreactor parameters such as pH or dissolved oxygen on CHO cell metabolism. As demonstrated above (Figure 2.1-2.3), the effect of bioreactor pH on CHO cell metabolism is complicated and can result in switch in metabolism of metabolites such as ammonia, lactate, aspartate, glutamate, glycine, and alanine. To gain more insights into the effect of pH on CHO cell metabolism, a combination of pFBA, and FVA were used.

pFBA analysis (Figure 2.4) confirmed that during early culture phases, higher pH increased the flux from glucose to pyruvate. The highest glycolysis fluxes throughout the culture duration were observed at pH 7.25. As expected, increased glycolytic fluxes resulted in increased lactate-production fluxes, but this same trend was not observed for alanine production. This is most likely due to the requirement of glutamate to provide nitrogen to convert pyruvate to alanine and bioreactor pH did not significantly affect glutamine (the main source of glutamate) consumption rates. Glutaminolysis is the conversion of glutamine to glutamate that is incorporated into the TCA cycle via AKG and leaves the TCA cycle at malate to produce pyruvate and may eventually end up as lactate [218]. This explains why the upper half of the TCA cycle fluxes were much lower than the lower half of the TCA cycle fluxes (Figure 2.4A, 2.4D, 2.4G and A5). In Section 2.3.2, we noted that the switch from lactate production to lactate uptake coincided with glutamine depletion, indicating that glutamine was converted to lactate. The results from pFBA add further insights into this possibility, indicating that glutamine was responsible for pyruvate production that resulted in lactate production. In addition to affecting lactate metabolism, reduction of the pyruvate

production fluxes, and intracellular glutamate concentrations could have also resulted in a switch in metabolism of alanine.

During the initial phase of the culture, aspartate and glutamate were produced (Figures A2G and A2I), the latter from glutamine conversion. Aspartate production can be linked to two sources, consumption of asparagine and production from oxaloacetate produced from glutaminolysis. All subfigures in Figure 2.4 show that there is a constant production of aspartate from asparagine, but there is a drastic reduction in production of aspartate from oxaloacetate after depletion of glutamine, indicating that the reduction of glutaminolysis led to reduction in intracellular production of aspartate and this led to consumption of aspartate from the media. Thus, the switch in aspartate metabolism can also be linked to depletion of glutamine.

During the early phase of the culture, glucose and glutamine were the two major sources of TCA cycle fluxes. Upon depletion of glutamine, the glycolytic flux feeds the TCA cycle fluxes (Figure 2.4B, 2.4C, 2.4E, 2.4F, 2.4H, 2.4I and A5) and as a result the fluxes in the upper half are similar to the lower half fluxes of the TCA cycle. Analysis of the data of Figure 2.1 and A2 in Section 2.3.3 shows that this CHO cell line exhibits unusually high cell specific rates of serine consumption. pFBA analysis showed that intracellular serine was produced from serine uptake from the media and from reactions involving glycolytic intermediates and glutamate (Figure 2.4 and Figure A3). pFBA analysis also showed that the intracellular serine is converted to pyruvate, ammonia and glycine (Figure 2.4).

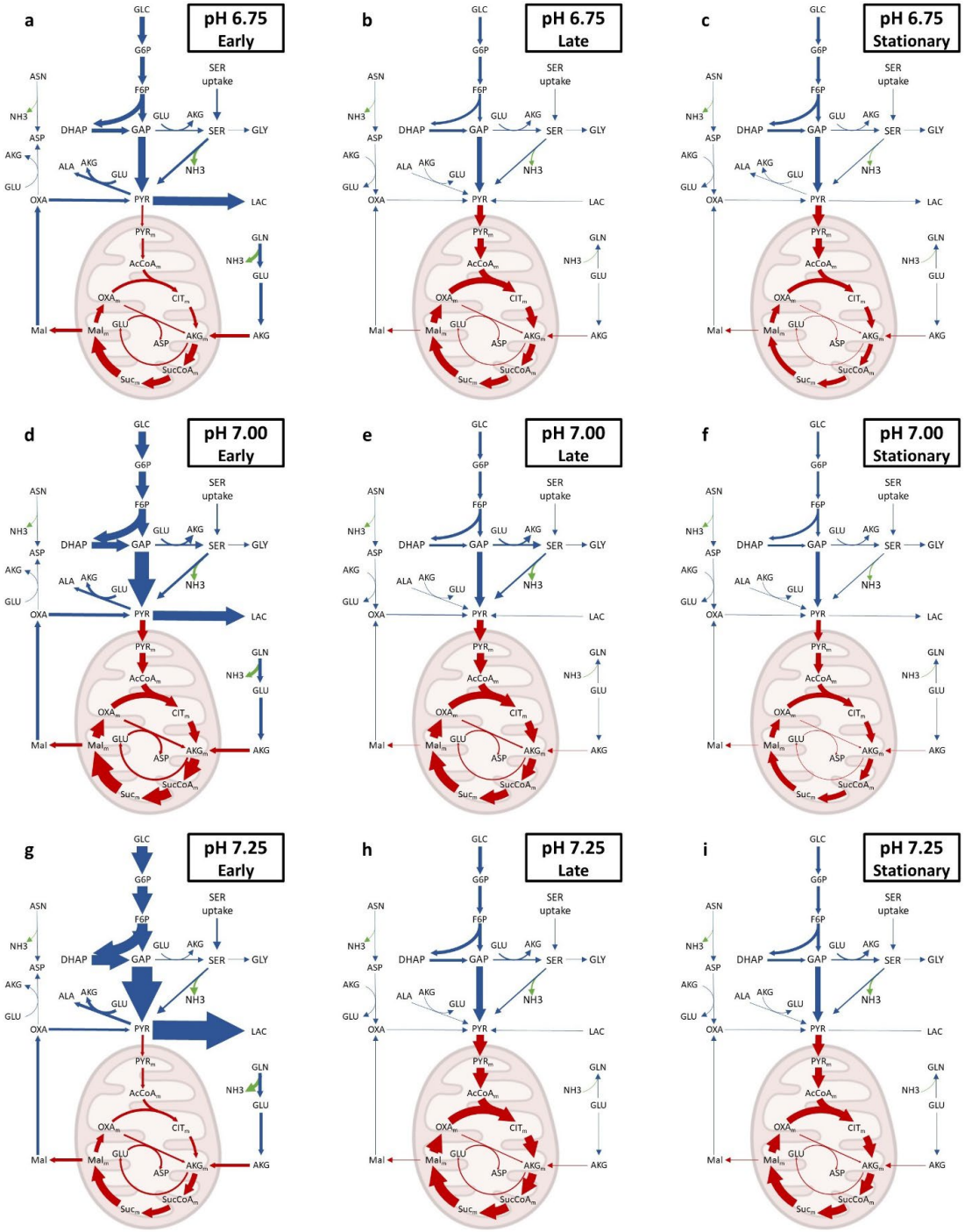

Figure 2.4: Flux balance analysis reveals the major sources of ammonia production and differences in intracellular fluxes at various culture phases for different bioreactor pH conditions.

The switch in ammonia metabolism (Figure 2.2) is complicated as it results from the metabolism of several amino acids. Ammonia production reactions (Figures 2.4 and A4) indicate that in the early phase of the culture, most of the ammonia is produced from glutamine conversion to glutamate. Surprisingly, the second largest source of ammonia production is not from asparagine but from serine producing pyruvate. pFBA shows that the reaction converting serine to pyruvate led to the largest ammonia production flux in the second and third phase of the cultures. Hence, reduction of serine concentrations in the media could have resulted in better culture performances for product quality and increased titers.

2.3.5 Bioreactor pH impacts Fab and Fc glycosylation

Glycopeptide mapping revealed major differences in Fab and Fc glycans on the antibody produced by the CHO VRC01 cell line (Figure 2.5A). It is not surprising that the Fc region has very high fucosylation and the most abundant glycans were the G0F and G1F. Low amount of high mannose, G0, G2F, and G0F-GlcNAc glycans were also detected in the Fc region. The lack of G2F and sialylation indicates that the galactosyltransferases and sialyltransferase enzymes cannot easily access the glycosylation site in the Fc region. A survey of the glycosylation of over 150 mAbs approved by the FDA by May 2023 has been published in the literature [219]. Out of the 150 approved mAbs, it has been reported that 57 IgG1 molecules with glycans only in the Fc region had very similar N-linked glycosylation profiles. G0F glycans were the most abundant followed by G1F and G2F glycans. All 57 IgG1s had very high fucosylation [219].

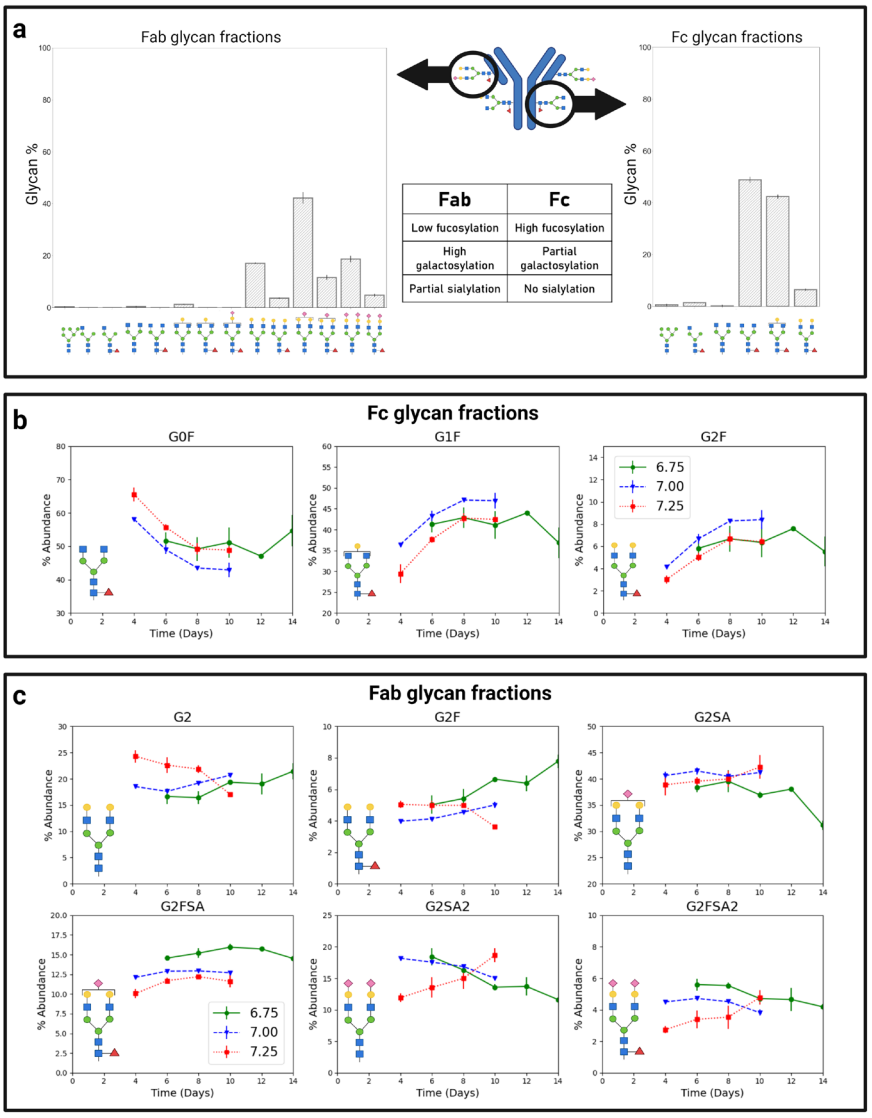

Figure 2.5: A) Glycopeptide mapping revealed significant differences in the glycans on the Fab and Fc mAb regions. Fab region glycans displayed large diversity, low fucosylation, high galactosylation and partial sialylation. Fc glycans displayed little diversity, high fucosylation, partial galactosylation and no sialylation. Data shown for biological duplicates. Data shown for harvest glycans from bioreactor operated at pH 7.25. B) Glycan fractions in the Fc region were tracked over time for each pH condition. C) Glycan fractions in the Fab region were tracked over time for each pH condition. Glycan fractions of glycans with low abundance are provided in the supplementary data.

More complex glycoforms were observed in the Fab region while compared to the Fc region (Figure 2.5A). The most abundant glycans were G2, G2F, G2S1. G2FS1, G2S2, and G2FS2. Glycans with high degrees of sialylation and galactosylation were observed in the Fab region, indicating increased accessibility of galactosyltransferase and sialyltransferase enzymes to the glycosylation site in the Fab region. The majority of glycans in the Fab region were afucosylated. This implies that for the mAb used in this study, fucosyltransferase enzyme had poor accessibility to this glycosylation site. 20% of IgGs in human serum have glycans in the Fab and Fc region. In the literature site-specific glycan analysis has been performed on IgGs in human serum to identify the differences in glycan fraction across the two sites. It has been reported that Fab glycans contained higher percentages of ternary glycan structures, galactosylation and sialylation but contained lower fucosylation while compared to the glycans present in the Fc region. [108]. Site specific N-linked glycosylation on IgG acquired from intravenous immunoglobulin from healthy donors has shown that the fucosylation in Fab region was close to 80% and greater than 95% in the Fc region [220]. The results in our study agree with the literature that the Fab region leads to increased galactosylation and sialylation. However, there were a few significant differences in findings in the literature to IgG from human serum vs. the mAb produced by the CHO VRC01 cells. The fucosylation of Fab glycans for the mAb (produced by CHO VRC01 cells) was less than 30%. This value is smaller than that reported in the literature (80%) for IgG acquired from intravenous immunoglobulin of healthy donors [220]. The high mannose types were also present in small quantities, indicating good accessibility of mannosidases and poor accessibility of *N*-acetylglucosaminyltransferase-IV (GnT4)

and *N*-acetylglucosaminyltransferase-V (GnT5) enzymes to the Fab N-linked glycosylation site.

Experimental determination of the effect of bioreactor pH and culture duration on Fab and Fc glycosylation fractions was an important goal of our study (Figure 2.5B and 2.5C). Bioreactor pH significantly impacted G0F, G1F, and G2F glycans in the Fc region (Figure 2.5B). The highest levels of G1F and G2F glycans were observed at pH 7.00. Increases in G1F and G2F glycans naturally result in lower G0F glycoforms. Increasing bioreactor pH led to reduction in G2FSA glycan fractions throughout the culture duration (Figure 2.5C). For the pH 7.25 condition, increase in culture time led to reduction in G2 and G2F fractions but increase in G2SA, G2FSA2, and G2SA2 fractions. The opposite trend was observed for these glycan fractions in the pH 6.75 and pH 7.00 conditions. To sum up, bioreactor pH can significantly affect the glycan fractions in the Fab and Fc region. The other glycan fractions present in lower abundance are shown in Figures A6 and A7.

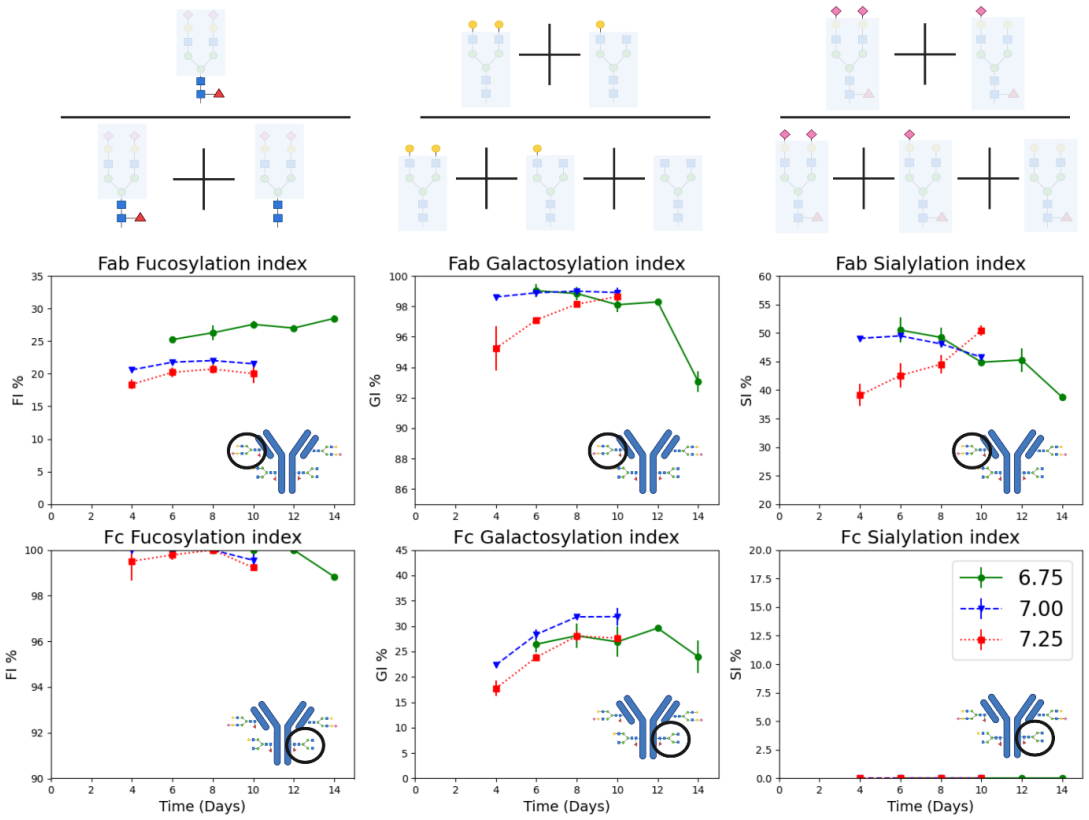

Figure 2.6: Fucosylation increased with decreasing bioreactor pH in the Fab region but did not change in the Fc region. Bioreactor pH had minimal effects on galactosylation in the Fab region but affected the galactosylation in the Fc region. Sialylation was absent in the Fc region but varied with bioreactor pH and culture duration in the Fab region. Data are presented using glycosylation indices described in Equations (2.1), (2.2) and (2.3). Plots of the effect of bioreactor pH on all the individual glycan fractions are provided in the supplementary data. Data shown for biological duplicates.

Glycosylation indices (as described in Equations 2.1-2.3) provide a succinct summary of the impact of culture conditions and duration on protein glycosylation. Bioreactor pH had a significant effect on the fucosylation in the Fab region but did not affect the fucosylation in the Fc region (Figure 2.6). Fucosylation in the Fc region was nearly 100% throughout the culture duration for all the different pH conditions. Culture

duration did not have a significant effect on the fucosylation index of the Fab region. However, an increase in bioreactor pH led to reduction in the fucosylation index. The galactosylation index of the Fc region was the highest at pH 7.00 throughout the culture duration. The galactosylation index in the Fc region increased from day 4 to day 8 for pH 7.00 and 7.25. The galactosylation index of the Fab region was very high throughout the culture duration. However, at pH 6.75 the galactosylation index in the Fab region dropped at the end of the culture (Figure 2.6). The sialylation index in the Fc region was zero as sialylation was not observed in the Fc region.

The effect of bioreactor pH on Fab sialylation was non-intuitive. For bioreactors run at pH 7.00 and pH 6.75, the sialylation index of the Fab region dropped as culture duration progressed. However, the opposite trend was observed for bioreactors operated at pH 7.25. On day 4, there was a drastic reduction in sialylation in the pH 7.25 condition while compared to the other bioreactor runs. The Fab sialylation index in this condition increased as culture progressed. Data in the literature suggest that sialyltransferase activity is most sensitive to changes in Golgi pH [73] and that the concentrations of ammonia significantly affect Golgi pH [214]. High lactate concentrations have also been shown to impact intracellular pH [214]. Figures 2.1 and 2.2 show that bioreactor pH had a significant effect of lactate and ammonia concentrations.

2.4 Conclusions

Our results indicate that bioreactor pH influences growth rates, glucose uptake rates, lactate production rates, amino acid uptake rates, ammonia secretion rates, and titer production rates, indicating that the bioreactor pH should not be neglected while designing a fed-batch process and the media (basal and feed) will not be optimal for all the different bioreactor pH conditions. Differences in growth rates and nutrient uptake

rates will lead to accumulation or depletion of nutrients at certain pH conditions. Taking this into account can improve product titers. pFBA showed that depletion of glutamine coincides with a metabolic shift from lactate production to lactate consumption. This analysis also showed us that bioreactor pH significantly affects glycolysis and intracellular pyruvate production rates but did not affect alanine production rates in the early culture phase. The effect of glutamine uptake rates on the shift in metabolism of aspartate and glutamate was also dissected by pFBA. We have shown that the switch in metabolism of aspartate is linked to glutamine uptake rates. Bioreactor pH and culture duration significantly affected ammonia metabolism. This CHO cell line also exhibited very high serine uptake rates. Analysis of our data suggests that serine produces pyruvate and leads to ammonia production.

This is the first study to report the effect of a critical process parameter (pH) on site specific N-linked glycosylation of a therapeutic glycoprotein. Our data demonstrated the diversity of glycans present in the Fab and Fc region of the mAb. Fc glycans contain high fucosylation, moderate galactosylation and no sialylation. Fab glycans contained many more species but contained moderate fucosylation, high galactosylation and moderate sialylation. Increased pH led to reduction in fucosylation in the Fab region but did not affect fucosylation in the Fc region. In the Fc region, the highest galactosylation was observed at pH 7. However, bioreactor pH did not seem to drastically affect Fab galactosylation. Sialylation in the Fab region seemed to be affected by metabolites that can impact the intracellular pH. Certain glycosylation attributes can be affected in one site but not the other site. Understanding the effect of bioreactor pH on CHO cell metabolism and site-specific N-linked glycosylation can improve bioprocess performance and implement quality by design.

Chapter 3

A NOVEL, SITE-SPECIFIC N-LINKED GLYCOSYLATION MODEL PROVIDES MECHANISTIC INSIGHTS INTO THE PROCESS CONDITION DEPENDENT DISTINCT Fab AND Fc GLYCOSYLATION OF AN IgG1 MONOCLONAL ANTIBODY PRODUCED BY CHO VRC01 CELLS

3.1 Introduction

Post translational modifications such as N-linked glycosylation are linked to the function of various glycoproteins such as therapeutic monoclonal antibodies (mAbs), fusion proteins, viruses, and other native mammalian proteins. N-linked glycosylation of therapeutic proteins is a closely monitored Critical Quality Attribute (CQA) as it can influence the half-life, immunogenicity, thermal stability, and efficacy [111]. To design robust processes for therapeutic manufacturing with consistent glycosylation fractions, regulatory authorities have set Quality by Design (QbD) guidelines to understand the impacts of process conditions on the CQAs [8]. Development of biosimilars requires matching the glycosylation profile of the originally approved therapeutic protein; this matching is one of the most challenging tasks in biosimilars development [92].

The production of glycoproteins is complex, stochastic, and non-template driven. Mathematical models have been developed in the past to understand and predict the types of glycans present on glycoproteins as a function of glycoprotein production rates, enzyme concentrations, and nucleotide sugar concentrations, among others [196]. These mathematical models have been applied to aid with reading mass spectrometry data, evaluating biomanufacturing design space for critical process parameters,

determining optimal media feeding schedules, and understanding the relationship between disease and protein glycosylation [99, 113-115, 221, 222].

About 20% of the IgG antibodies in the human body have glycosylation on the Fab region, whereby the glycosylation fractions in the Fab and Fc regions can vary significantly [108]. Recent advances in site-specific N-linked glycan analysis have allowed for better understanding of site-specific heterogeneity of glycoproteins. In addition to Fc glycosylation impacting product quality, it has been demonstrated that Fab glycans can also influence mAb half-life, efficacy, thermal stability, and aggregation [107, 108, 195]. Process related conditions such as media additives, pH, and temperature can affect the N-linked glycosylation of mAbs [111]. Mathematical models of glycosylation have been used in the literature to study the impact of culture conditions (temperature shift and manganese supplementation) on N-linked glycosylation [115, 223]. Another application of mathematical models is to predict glycan fractions after changes to media composition by addition of glycosylation modulators such as galactose or $MnCl_2$ [114].

The goal of this study is to develop a predictive site-specific glycosylation model for an IgG1 mAb containing glycans in both the Fab and Fc region. We have shown that bioreactor pH has significant but differential impact on both Fab and Fc glycosylation of the mAb produced by the Chinese hamster ovary (CHO) VRC01 cell line in Chapter 2 [224]. Inclusion of site-specific details in glycosylation models is essential for understanding the impact of process conditions on the glycans from all the glycosylation sites. To meet this need, here we report the development of a site-specific N-linked glycosylation model in order to understand the relationship between process conditions and the Fab and Fc glycans on the IgG1 produced from the CHO VRC01 cells. The site-

specific N-linked glycosylation model approximates the Golgi as a series of stirred tank reactors (STRs), [94] and includes site-specific details in the kinetics of glycosylation process in the Golgi. The kinetic model parameters explain the reasons for the major differences between experimentally observed Fab and Fc glycans. The model predictions were compared with measurements of glycosyltransferase mRNA expression levels and intracellular nucleotide sugar concentrations to understand the effect of bioreactor pH on site-specific N-linked glycosylation.

Another goal of this study was to show that the site-specific N-linked glycosylation model can predict the glycan fractions as a function of nutrient composition. To this effect, predictions of increased galactosyltransferase activity on the Fab and Fc glycan fractions were made and experimentally validated using data generated from cultures supplemented with galactose and $MnCl_2$. Although the model was used for the specific IgG1 in this study, it can be adapted to any protein molecule with multiple N-linked glycosylation sites.

3.2 Materials and methods

3.2.1 Cell culture

CHO-K1 VRC01 cell line (Clone A11 from the Vaccine Research Center at the National Institute of Health) expressing a broadly neutralizing anti-HIV mAb with Fab and Fc glycosylation was used in this study. Data from bioreactor experiments was adapted from Chapter 2 [224]. Additional shake flask experiments involved fed-batch cultures of CHO cells in 125 mL Erlenmeyer flasks at 30 mL culture volume. Cells were seeded at 0.4 million cells/mL and incubated at 37 °C in an incubator configured to 85% relative humidity, 20% O_2 and 5% CO_2. Cells were grown in HyClone Actipro media

(Cytiva) supplemented with 6 mM L-glutamine. HyClone Cell Boost 7a supplement and HyClone Cell Boost 7b supplement were used as feed media. Starting on day 3 of the culture, 3% of HyClone Cell Boost 7a and 0.3% of HyClone Cell Boost 7b were added daily. Starting on day 5, glucose was added to the flask to bring the glucose concentration to 9 g/L. Methods to culture the CHO VRC01 cell line have been described in the literature [205]. Galactose and $MnCl_2$ supplementation experiments involved adding galactose every day from day 4 onwards as described in Table 3.1. Experiments were performed in duplicate. D-Galactose (catalog number: G5388-100G) was obtained from Sigma-Aldrich. $MnCl_2$ (catalog number: M5005-100G) was obtained from Sigma-Aldrich. Cell density and viability were measured by using trypan blue assay with the DeNovix automatic cell counter.

Table 3.1: Description of shake flask experiments involving supplementation of galactose and MnCl2.

Flask number	Amount of galactose added	Additional details
1	No galactose added	Flask was extracted on day 4 to analyze glycans.
2	0.2 mL of 0 g/L galactose (only water) added daily from day 4.	Flask was harvested on day 9. No galactose was added to this flask.
3	0.2 mL of 37.5 g/L galactose solution added daily from day 4.	Flask was harvested on day 9. This flask will be referred to as '1X galactose added'.
4	0.2 mL of 150 g/L galactose solution added daily from day 4.	Flask was harvested on day 9. This flask will be referred to as '4X galactose added'.
5	0.2 mL of 150 g/L galactose solution added daily from day 4. $MnCl_2$ was added on day 4 to bring the concentration of Mn to 1 μM.	Flask was harvested on day 9. This flask will be referred to as '4X galactose + Mn added'.

3.2.2 Measurements of substrate, metabolite, and IgG concentrations

Measurements of glucose, lactate, galactose, and ammonia were performed on a YSI 2700 bioanalyser. Antibody titers were measured using Protein A HPLC with an Agilent 1260 system and a POROS A column (Catalog number, 1502226; ThermoFisher). The method details involving the mobile phase composition, mobile phase flow rates, and injection volume have been described in Chapter 2 [224]. The methods for calculation of specific uptake and production rates have also been described in Chapter 2 [224].

3.2.3 Measurements of intracellular nucleotide sugar concentrations

CHO cells (10 million cells) were extracted from bioreactor experiments every two days and from shake flask experiments during harvest by centrifuging and removing the media. The cells were subsequently washed with 0.9% w/v ice cold sodium chloride solution and centrifuged immediately to remove cell culture media residues. The washing procedure was performed twice followed by flash freezing the cells for storage at -80 °C for further analysis. Extraction of nucleotide sugars from the cell pellets involved resuspending the cell pellets in ice-cold 50% aqueous acetonitrile, vortex, and incubating on ice for 10 minutes [225]. The solution was centrifuged for extraction of nucleotide sugar containing supernatant. This solution was dried on a SpeedVac and resuspended in 200 µL water. The samples were filtered using a 0.22 µm syringe filter and immediately analyzed on the HPLC. The method for extraction of nucleotide sugars described above was adapted from the literature [114, 226, 227].

Measurement of nucleotide sugar concentrations was performed on an Agilent HPLC 1260 with a UV detector at 262 nm using a Dionex CarboPac PA1 analytical column (2 mm diameter and 250 mm length; ThermoFisher (catalog number 057178)

together with a guard column (2mm diameter and 50 mm length, ThermoFisher, catalog number 057179) [114, 117, 201, 226, 228]. Mobile phase A was 3 mM NaOH and mobile phase B was 1.5 M NaOAc in 3 mM NaOH. Quantification of UDP-Gal and CMP-Neu5Ac was performed with the gradient provided by the manufacturer: t_0 = 80% A, t_2 = 80% A, t_{10} = 50% A, t_{12} = 50% A, t_{13} = 40% A, t_{22} = 40% A, t_{27} = 80% A, and t_{34} = 80% A. Measurement of GDP-Fuc required modifying the gradient: t_0 = 80% A, t_2 = 80% A, t_{10} = 70% A, t_{37} = 70% A, t_{50} = 60% A, t_{54} = 40% A, t_{64} = 40% A, t_{65} = 80% A, and t_{80} = 80% A. The flow rate for both methods was set to 0.3 mL/min. The sample injection volume was set to 10 µL to measure GDP-Fuc and 2.5 µL for measurement of UDP-Gal and CMP-Neu5Ac. Column temperature was set to 30 °C. Nucleotide sugars were measured at the same cell culture time points that were used to perform glycan analysis. Nucleotide sugar samples were measured on day 4, 6, 8, and 10 for bioreactor samples from bioreactors operated at pH 7.00 and 7.25. The bioreactor operated at pH 6.75 did not have enough antibody titer to measure glycans on day 4. Hence, both glycans and nucleotide sugars were measured on day 6, 8, and 10. For shake flask samples, nucleotide sugars were measured on days 4 and 9.

3.2.4 Measurement of glycosylation-enzyme mRNA levels

Extraction of mRNA from cell pellets (10 million cells) was performed using the RNeasy Mini Kit (Qiagen, Catalog number 74104). mRNA extracted from these samples was converted to cDNA by performing reverse transcription using a QuantiTect Reverse Transcription Kit purchased (Qiagen, Catalog number 205311). mRNA levels were measured by performing quantitative Real-time polymerase chain reaction (qRT-PCR) with PowerTrack SYBR Green Master Mix (ThermoFisher, Catalog number A46110). qRT-PCR was performed by using three technical replicates and two

biological replicates on the CFX96 Optical Reaction Module (Bio-Rad). mRNA levels were quantified using the comparative CT method ($2^{-\Delta\Delta Ct}$) [229]. β-actin was used as the housekeeping gene. For each target, the mRNA level of the target in the bioreactor sample at pH 7.00 day 4 was used as the reference sample and mRNA expression results were calculated as a fold change of the reference sample. mRNA expression analysis was performed for N-acetylglucosaminyltransferase I (GnT1), β-1,4-galactosyltransferase 1 (β4GalT1), β-1,4-galactosyltransferase 2 (B4GalT2), β-1,4-galactosyltransferase 3 (B4GalT3), UDP-Gal transporter (SLC35A2), ST3 β-galactoside α-2,3-sialyltransferase 3 (ST3Gal3), ST3 β-galactoside α-2,3-sialyltransferase 4 (ST3Gal4), fucosyltransferase 8 (FUT8), GDP-Fuc transporter (SLC35C1), UDP-GalNAc transporter (SLC35D1), and CMP-Neu5Ac transporter (SLC35A1). The primers used are listed in supplementary Table B1. mRNA levels were measured for the same timepoints that nucleotide sugars and glycans were measured.

3.2.5 Site-specific glycan analysis

The mAb used in this study contains glycans in the Fc and Fab region. For bioreactor operated at pH 7.00 and pH 7.25, samples were collected on day 4, 6, 8, and 10 for site-specific glycan analysis. Bioreactors operated at pH 6.75 did not have enough antibody titer on day 4 for glycan analysis. Hence, site-specific glycan analysis was performed on day 6, 8, and 10. The site-specific N-linked glycosylation data for bioreactor samples was adapted from Chapter 2 [224]. For the shake flask cultures described in Table 3.1, glycans were analyzed on day 4 for the first condition and on day 9 for the rest. Site-specific glycan analysis was performed for a total of 16 different samples (32 samples including biological duplicates). The site-specific glycan analysis method has been described in Section 2.2.3 [224].

3.3 Model development

3.3.1 N-linked glycosylation model formulation

Models for N-linked glycosylation have approximated the Golgi as a series of stirred-tank reactors (STRs) [93, 94] or a plug flow reactor (PFR) [95]. These models have shown applications in understanding the effect of process conditions on mammalian cell glycosylation [115] and simulating glycosylation profiles to aid with glycoengineering strategies for process or cell line development [73, 100, 114, 116]. None of the glycosylation kinetic models can be used to model site-specific N-linked glycosylation. To address this need we have developed a kinetic model for site-specific N-linked glycosylation.

The first step to modeling N-linked glycosylation is to generate the reaction network containing all the precursor glycans required to produce the experimentally observed glycans. Site-specific N-linked glycan analysis led to identification of 15 glycans in the Fab region and 7 glycans in the Fc region (Figure 3.1). Rules for generation of glycosylation reaction networks have been published [113, 230]. According to these rules, α-mannosidase class 1 (Man1), α-mannosidase class 2 (Man2), N-acetylglucosaminyltransferase I (GnT1), N-acetylglucosaminyltransferase II (GnT2), fucosyltransferase (FucT), galactosyltransferase (GalT), and sialyltransferases (SiaT) are required to produce the experimentally observed glycans. Based on these 7 enzymes and the reaction network generated from enzyme-reaction rules [94], 51 reactions and 41 glycans are identified as summarized in Figure 3.1. Four reactions are catalyzed by Man1, four reactions are catalyzed by Man2, one reaction is catalyzed by GnT1, two reactions are catalyzed by GnT2, four reactions are catalyzed by FucT, eighteen reactions are catalyzed by GalT, and eighteen reactions are catalyzed by SiaT.

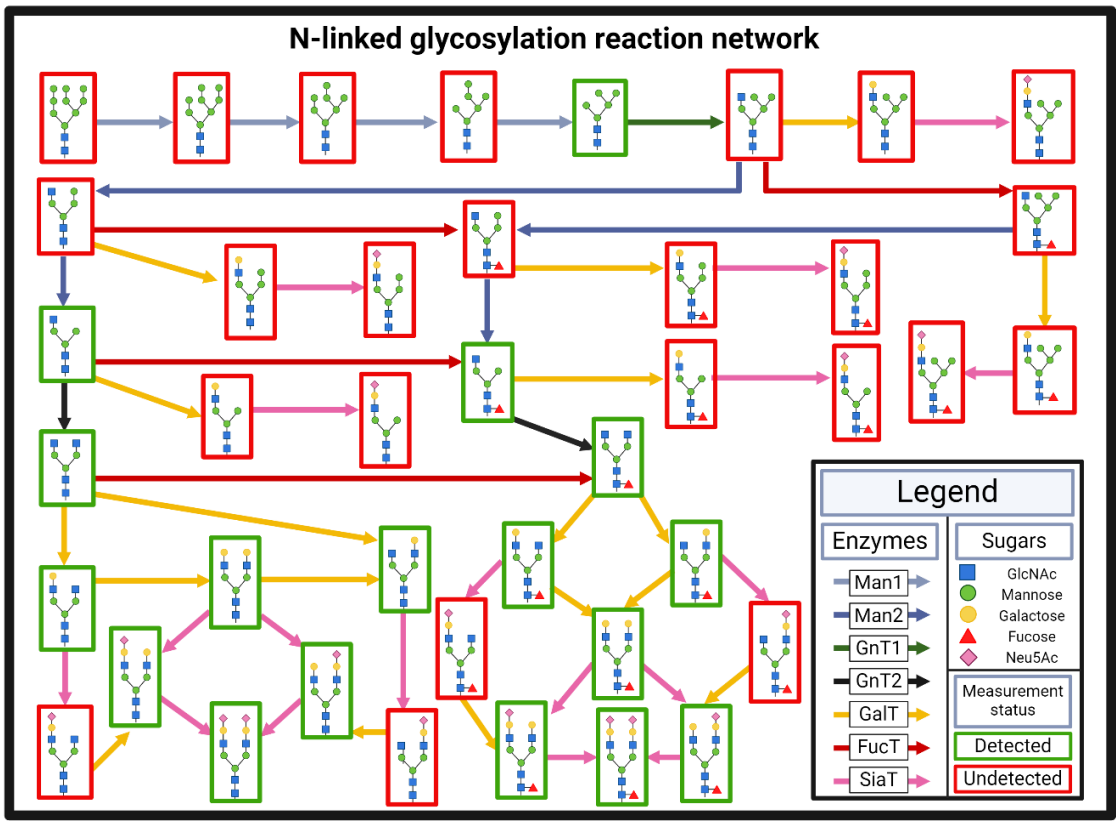

Figure 3.1: N-linked glycosylation reaction network to produce glycans observed in experimental data.

To develop a mechanistic model for site-specific N-linked glycosylation, we examined the mechanism of site-specific N-linked glycosylation (Figure 3.2A). The process of translation and assembly of the IgG from mRNA starts in the endoplasmic reticulum (ER). Addition of glycans to the protein occurs co-translationally. For IgG molecules with Fab and Fc glycosylation, this occurs to the heavy chain and light chain fragments independently as the proteins are translated. This is followed by trimming of the glucose residues by glucosidase in the ER and subsequent folding of the glycoprotein [89]. Heavy chain dimer assembly occurs first, followed by disulfide bond formation to the light chains [231]. Further trimming of the glycan on the IgG molecule

by mannosidase starts in the ER but most of the processing of the glycans occurs in the Golgi. Based on that mechanistic information, mathematical models for N-linked glycosylation assume that the glycans entering the Golgi are a mixture of the high mannose species M8 and M9 [95, 113]. The fully assembled IgG molecule enters the Golgi. Hence, both the Fab and Fc glycans have the same residence time in the Golgi and are exposed to the same enzymes and nucleotide sugar concentrations. The only difference between the two sites is their interaction with the N-linked glycosylation enzymes. Once the basal glycan has been added co-translationally in the ER, accessibility of a glycosyltransferase enzyme to the glycan at a particular site is dependent on the tertiary structure of the protein as well as local interactions with specific amino acids around the glycan [232]. To capture this in a mechanistic model, both sites have to be exposed to the same enzyme and nucleotide sugar concentrations, while differences in kinetic rate constant values can be used to mimic the different interactions between the glycosylation sites and the glycosylation enzymes.

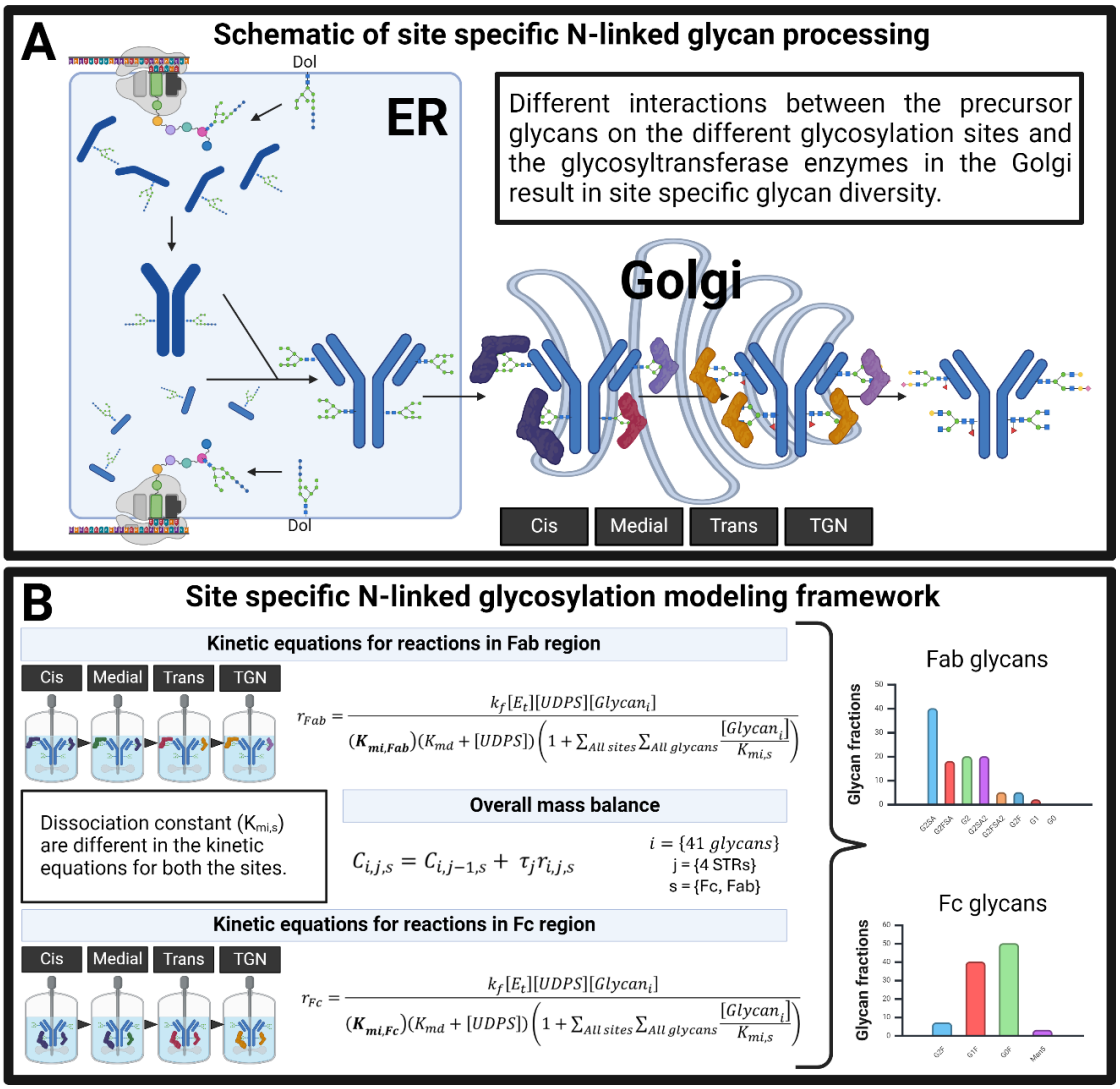

Figure 3.2: Summary of glycosylation process to build the mathematical model for the site-specific N-linked glycosylation processing in the Golgi. A) Site-specific N-linked glycosylation process. B) Site-specific N-linked glycosylation model.

In this study, we approximate the Golgi as a series of four STRs [94]. The specific mAb production rate (q_P) has been used to determine the total amount of N-glycans (C_{tot}) going into the Golgi by multiplying it with the number of glycosylation

sites per IgG (4 in this case), residence time (τ) and dividing by the volume of each compartment (v) in the Golgi according to the Equation (3.1).

$$C_{tot} = \frac{4 \times q_P \times \tau}{v} \tag{3.1}$$

The model requires solving the mass balance equations for each glycan at each site in the STRs. The concentration of a particular glycan indexed by *i* in tank *j* at the glycosylation site *s* is given by Equation (3.2). The net rate of production of glycan *i* in tank *j* at site *s* is represented by $r_{i,j,s}$.

$$C_{i,j,s} = C_{i,j-1,s} + \tau_j r_{i,j,s} \tag{3.2}$$

We have used Michaelis-Menten kinetics (Equation (3.3)), [94, 113] to model the enzymatic kinetics of the glycosyltransferases:

$$r = \frac{k_f[E_t][UDPS][Glycan_i]}{(K_{mi,s})(K_{md} + [UDPS])\left(1 + \sum_{All\ sites}\sum_{All\ glycans}\frac{[Glycan_i]}{K_{mi,s}}\right)} \tag{3.3}$$

k_f is the forward rate constant, $K_{mi,s}$ and K_{md} are the dissociation constants for the substrate glycan *i* at glycosylation site *s* and the donor co-substrate (nucleotide sugar). [UDPS] and $[E_t]$ represent the concentrations of nucleotide sugars and enzymes in the Golgi. The only site dependent parameter in Equation (3.3) is $K_{mi,s}$. Differences in interactions between the glycans in Fc and Fab glycosylation sites and the glycosyltransferase enzymes are captured by using two sets of $K_{mi,s}$ values (s = {Fc, Fab}). The net rate of production of a glycan is determined by adding all rates of reaction (Equation (3.3)) involving that glycan (Figure 3.1). The net rate of production of a glycan is included in the mass balance of Equation (3.2). This leads to a system of nonlinear equations containing mass balances for glycans in the Fc region with one set of $K_{mi,Fc}$ values resulting in 164 equations: 4 STRs with 41 equations per STR. Mass

balances for Fab glycans with a different set of $K_{mi,Fab}$ values add another 164 equations. The resulting system of 328 nonlinear equations was solved using the Levenberg-Marquardt algorithm [233] implemented in Python with the library SciPy [234]. A schematic of the site-specific N-linked glycosylation model has been shown in Figure 3.2B.

3.3.2 N-linked glycosylation parameters

Enzyme concentrations, forward rate constant values, nucleotide sugar dissociation constants, and default nucleotide sugar concentrations were taken from the literature [94, 113, 116] and are summarized in supplementary Tables B2 and B3. The UDP-Gal dissociation constant was determined from data on galactose feeding experiments in the literature [23]. The Golgi volume and residence time of proteins in the Golgi were also taken from the literature [94, 113]. The different cell culture conditions used in this study led to variations in cell specific antibody titer production rates, which are summarized in Table 3.2. The distribution of enzymes in each compartment of the Golgi was adapted from the literature calculations of enzyme activities in the various Golgi compartments [230, 235]. These values are shown in supplementary Table B4. An even distribution of M9 and M8 glycans was assumed to exit the ER [94]. Calculation of intracellular nucleotide sugar concentrations involved assuming the volume of a CHO-K1 cell to be 1600 μm^3 [43].

Table 3.2: Culture conditions impact cell specific antibody titer production rates for the VRC01 cell line.

Condition	Cell specific antibody titer production, q_P (pg/cell/day)	Reference
Bioreactor culture at pH 6.75	5.5	Chapter 2 [224]
Bioreactor culture at pH 7.00	8.9	
Bioreactor culture at pH 7.25	11.8	
Shake flask culture from Table 3.1	3.5	Measured in this study (Chapter 3)

3.3.3 Regression of the model parameters to experimental data

The dataset contains experimental measurements from multiple days of three bioreactor experiments at different pH conditions and several measurements from shake flask experiments at different galactose feeding concentrations. This results in a total of 32 measurements including biological duplicates from the 16 experimental observations. The diversity of experimental measurements has been highlighted in Figure 3.3. Calculations of glycosylation indices were performed according to Equations (3.4), (3.5), and (3.6). F1 and F0 represent glycans with and without fucose residue. G2, G1, and G0 represent glycans with two galactose residues, one galactose

residue, and no galactose. S2, S1, and S0 represent glycans with two sialic acid (5-N-acetylneuraminic acid) residues, one sialic acid residue, and no sialic acid.

$$Fucosylation\ index\ \% = \ 100 \times \frac{F1}{F0 + F1} \tag{3.4}$$

$$Galactosylation\ index\ \% = \ 100 \times \frac{2 \times G2 + G1}{2\ \times G2 + 2\ \times G1 + 2\ \times G0} \tag{3.5}$$

$$Sialylation\ Index\ \% = \ 100 \times \frac{2\ \times S2 + S1}{2\ \times S2 + 2\ \times S1 + 2\ \times S0} \tag{3.6}$$

N-linked glycosylation measurements at a specific culture time represent the cumulative glycan fractions produced from the inoculation of the bioreactor until the sampling time. To determine the differences in glycan fractions produced between two sampling events, the data were transformed according to Equation (3.7) to yield the adjusted glycan fractions. GF_{time1} represents the glycan fraction measured at the previous sample time and GF_{time2} represents the glycan fractions measured at the current time point. This transformation has been commonly employed in the literature for time series of glycosylation data [236, 237]. The original glycan fractions, titers, and adjusted glycan fractions are plotted in supplementary Figures B1 to B10.

$$Adjusted\ glycan\ fractions = \frac{(GF_{time2} \times titer_{time2}) - (GF_{time1} \times titer_{time1})}{titer_{time2} - titer_{time1}} \tag{3.7}$$

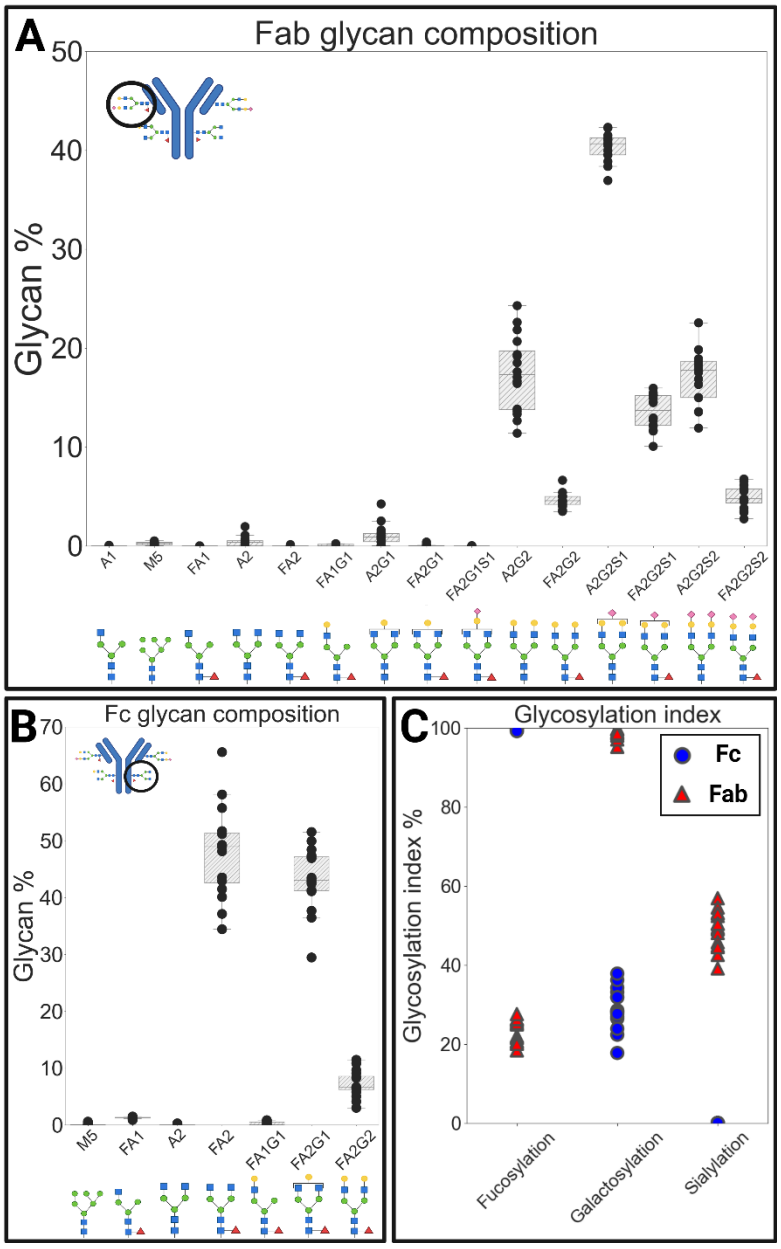

Figure 3.3: Cell culture process conditions impact N-linked glycosylation fractions and glycosylation indices. Data from 32 experimental measurements (including biological duplicates) resulted in 16 data points displayed in this boxplot. A) Fab glycan fractions. B) Fc glycan fractions. C) Fab and Fc glycosylation indices.

The parameters that need to be regressed have been divided into two groups. Group 1 consists of molecule dependent parameters ($K_{mi,s}$) that are not dependent on process or culture conditions. The $K_{mi,s}$ values for both sites for Man1, and Man2 could not be determined from the experimental data as the substrates for these enzymes were below the experimental detection limit. Hence, these reactions were assumed to be very fast and their $K_{mi,s}$ value was set to 5 μM for both the sites, indicating very high accessibility to the sites. It is well known that the galactosyltransferase enzyme activities are different for galactosylation in the α-1,3 arm of the glycan compared to the α-1,6 arm of the glycan as the two FA2G1 isomers have different abundance [238]. Hence, it is common to regress multiple dissociation constant values for the galactosyltransferase enzyme [94, 95]. The ratio of $K_{mi,s}$ in the α-1,3 arm to $K_{mi,s}$ in the α-1,6 arm were constrained to be the same for the Fab and Fc region. The results of regression of these molecule-dependent parameters are presented in Section 3.4.1.

The second group of parameters are related to the process or culture conditions. Process or culture conditions can impact the glycosyltransferase enzyme expression levels, nucleotide sugar transporter enzyme expression levels, and/or intracellular nucleotide sugar concentrations. Changing the enzyme concentration [E_t] in Equation (3.3) can be used as a proxy to mimic an increase or decrease in enzyme activities that could result from changing enzyme expression levels, nucleotide sugar transporter expression levels, nucleotide sugar concentrations, or other inhibitors of glycosylation such as ammonia. To model the effect of process conditions for each experimental measurement, the [E_t] was regressed for each process condition. The results of this regression are plotted in the supplementary Figure B11 illustrating an excellent agreement. This result demonstrates that once the $K_{mi,s}$ values are fixed, changing the

enzyme activity of GnT1, FucT, GalT, and SialT can explain the variations observed in the data. The regression was carried out by minimizing the sum of squares due to regression (SSR; Equation (3.8)) between the model predicted glycan fractions $f(K_{mi}, [E_t])$ and the experimental measurements.

$$SSR = \sum_{All\ experiments} (f(K_{mi}, [E_t]) - Experimental\ glycan\ fraction)^2 \quad (3.8)$$

The results from the regression led to determining parameter values that are discussed in Sections 3.4.1, 3.4.2, 3.4.3, and 3.4.4. Galactose addition shake flask experiments 3 and 4 from Table 3.1 were left out of the regression and were instead used to test the predictive capabilities of the model in (Section 3.4.4).

3.3.4 Sensitivity analysis

The complex interactions of intracellular nucleotide sugar concentrations, glycosyltransferase enzyme expression levels, and nucleotide sugar transporter enzyme expression levels and their individual contributions to glycosylation are difficult to model, and not entirely understood. Hence, an approximation has been made to utilize the total enzyme concentration variable as a proxy to mimic the overall contribution of intracellular nucleotide sugar concentrations, glycosyltransferase enzyme levels, and nucleotide sugar transporter enzyme levels. Data in Figure 3.3 show that fucosylation, galactosylation and sialylation were significantly affected by process conditions. Very low levels of high mannose content were observed across all the experimental measurements (Figure 3.3). Hence, we decided to study the impact of increase or decrease in fucosyltransferase (FucT), galactosyltransferase (GalT), and sialyltransferases (SiaT) activities on the different glycan fractions. There is also evidence that these enzymes could impact each other. For example, in the literature

galactose supplementation has resulted in an increase in sialylation due to the presence of more galactosylated glycan substrates for SiaT [239].

Another parameter that is commonly used as input in glycosylation models is cell specific mAb production rate (q_P). There is experimental and computational evidence that q_P can impact N-linked glycosylation [94, 237, 240]. In general, increased q_P results in less mature glycoforms. This is because increased q_P results in increased concentrations of glycans in the Golgi. If the concentration of glycan in the Golgi is much larger than the dissociation constant ($K_{mi,s}$), then according to Michaelis-Menten kinetics (Equation (3.3)) the rate of reaction does not depend on the concentration of the glycan and results in the maximum possible reaction rate, thus leading to saturation of reaction rates (depending on dissociation constant values). Further increase in glycan concentrations does not result in increased reaction rates and leads to lower conversion of glycans to more mature glycoforms [94].

Hence, to quantify the effect of q_P, FucT, GalT, and SialT on the fucosylation, galactosylation, and sialylation indices in both the Fc and Fab region, a sensitivity analysis was performed. This analysis also improves the interpretability of the model results by more clearly demonstrating the relationship between model independent variables (q_P, FucT, GalT, and SialT) and outputs. The minimum and maximum values of q_P, FucT activity, GalT activity, and SiaT activity (inputs) were determined from the regression results. Understanding the relationship between four input parameters on the Fc fucosylation, Fc-galactosylation, Fc-sialylation, Fab-fucosylation, Fab-galactosylation, and Fab-sialylation indices led to 6 different problems. For each of these problems, a uniform sample vector was generated using Saltelli sampling [241] over the range of all four inputs. The site-specific glycosylation model was solved 10240

times (4 parameters and 1024 initial samples lead to 10240 unique parameter combinations) per problem. The resulting Fab and Fc glycosylation indices for each of the unique 10240 simulations were analyzed using Sobol sensitivity analysis [242] to determine first order, second order and total order indices. The first order index provides information on the variance of the output that can be explained by a single input variable. The second order index provides information on the effect of interactions between two inputs on the variance of output. The total order index is the sum of the first order index and the second order index. The sampling and sensitivity analysis were implemented using the Sensitivity Analysis Library (SALib) in Python [243].

3.4 Results and discussion

3.4.1 Model parameters account for site-specific glycosylation heterogeneity

The dissociation constants ($K_{mi,s}$) values for the Fc and Fab region (Table 3.3) were used to explore process agnostic interactions between glycosyltransferase enzymes and N-linked glycosylation site that led to the differences in glycan fractions observed in the Fc and Fab region (Figure 3.3). Large dissociation constant ($K_{mi,s}$) values represent poor accessibility of glycosyltransferase enzyme to the glycan at a particular site (Equation (3.3)). Low values of dissociation constant ($K_{mi,s}$) represent good accessibility of glycosyltransferase enzyme to the glycan at a particular site (Equation (3.3)). Higher reaction rates (Equation (3.3)) are observed for glycosyltransferase enzymes that can easily access the glycans at a particular site.

Table 3.3: Values of dissociation constants from regressing the model to experimental data in the Fab and Fc region for various glycosyltransferase enzymes.

Enzyme	$K_{mi,Fc}$ (μM)	$K_{mi,Fab}$ (μM)
GnT1	68	136
GnT2	152	20
FucT	2.1	1543
GalT	22000 (α-1,3 arm) 3400 (α-1,6 arm)	168 (α-1,3 arm) 26 (α-1,6 arm)
SiaT	1.00×10^9 (α-1,3 arm) 1.37×10^8 (α-1,6 arm)	3490 (α-1,3 arm) 478 (α-1,6 arm)

As seen in Figure 3.3 and supplementary Figures B1 and B2, M5 fractions were not observed in the Fc region for the majority of the data points, but a small amount of M5 species was observed in the Fab region consistently. The regression results are consistent, since the dissociation constant ($K_{mGnT1,Fc}$) value (68 μM) in the Fc region indicates good accessibility and the dissociation constant ($K_{mGnT1,Fab}$) value in the Fab region was determined to be 136 μM, indicating reduced accessibility to the glycans in the Fab region. Figure 3.3 and the supplementary Figures B1 and B2 show that the Fc region and Fab region consisted of A1 and FA1 glycan species, which are the substrates for GnT2. The experimentally observed fractions of these glycans in the Fab region were significantly lower than those in the Fc region, indicating increased accessibility to glycans in the Fab region. As a result, parameter regression provided us with a value of 152 μM ($K_{mGnT2,Fc}$) in the Fc region and 20 μM ($K_{mGnT2,Fab}$) in the Fab region.

Figure 3.3C shows nearly 100% fucosylation in the Fc region but 15 to 30% fucosylation in the Fab region, indicating very high accessibility of fucosyltransferase to the glycans in the Fc region and poor access to the glycans in the Fab region. Hence, regression of the dissociation constants gave a value of 2.1 μM ($K_{mFucT,Fc}$) for the Fc region and 1543 μM ($K_{mFucT,Fab}$) for the Fab region. The opposite trend was noticed for

the galactosylation index (Figure 3.3C). The galactosylation index in the Fab region was very high (nearly 100%) compared to the galactosylation index in the Fc region (15 to 40 %). Only biantennary glycan structures were observed for this mAb. The regression results gave a value of 22000 μM ($K_{mGalT\ \alpha\text{-}1,3\ ,Fc}$) for the α-1,3 arm and 3400 μM ($K_{mGalT\ \alpha\text{-}1,6\ ,Fc}$) for the α-1,6 arm in the Fc region. The increased accessibility to the Fab region is reflected in the dissociation-constant values of 168 μM ($K_{mGalT\ \alpha\text{-}1,3\ ,Fab}$) for the α-1,3 arm and 26 μM ($K_{mGalT\ \alpha\text{-}1,6\ ,Fab}$) for the α-1,6 arm.

Similarly, for the sialic acid transferase, two dissociation values had to be regressed as sialylation can also occur on both the branches of the glycan. The dissociation constant of 10^9 μM ($K_{mSiaT\ \alpha\text{-}1,3\ ,Fc}$) was set for the α-1,3 arm to indicate no accessibility of the sialic acid transferase to the glycans in the Fc region as no sialylation was observed in the Fc region. Parameter regression resulted in a value of 1.37×10^8 μM ($K_{mSiaT\ \alpha\text{-}1,6\ ,Fc}$) for the α-1,6 arm. Dissociation constant values of 3490 μM ($K_{mSiaT\ \alpha\text{-}1,3\ ,Fab}$) and 478 μM ($K_{mSiaT\ \alpha\text{-}1,6\ ,Fab}$) in the Fab region indicate great accessibility of the sialyltransferases enzyme for the glycan structures. The differences in $K_{mi,s}$ values reflect the non-uniform accessibility of the different glycosyltransferases to the glycans on the two glycosylation sites. This is not surprising as the amino acid sequence around the glycosylation sites are very different (supplementary Table B5) and there are differences in tertiary structure of the protein backbone in the Fab and Fc regions.

3.4.2 Sensitivity analysis improves model interpretability

3.4.2.1 Sensitivity analysis suggests that Fab fucosylation index is affected by q_P and FucT activity

The first problem described in Section 3.3.4 involved studying the impact of q_P, FucT, GalT, and SiaT on the Fab fucosylation index. The sensitivity analysis resulted

in high first order and total order sensitivity indices for the model inputs q_P and FucT on the model output Fab fucosylation index (Figure 3.4A). This suggests that variations in the Fab fucosylation index are correlated to q_P and FucT activity. The total order index was equal to the first order index (Figure 3.4A). This suggests that there are no second order interactions between q_P and FucT on the Fab fucosylation index. While it is obvious that FucT activity is correlated to the Fab fucosylation index, it was not intuitive that q_P also might impact, based on the model, the Fab fucosylation index in the data set used here. After analyzing the results of the sensitivity analysis, we quantified that the linear correlation between experimentally measured q_P and the experimentally measured Fab fucosylation index has a Pearson correlation coefficient of -0.83. A Pearson correlation coefficient of -0.83 indicates that, according to the experimentally measured data, there is a high negative correlation between q_P and Fab fucosylation index. Increase in q_P results in reduction in Fab fucosylation index and increase in FucT activity results in increase in Fab fucosylation index (Figure 3.4B).

3.4.2.2 Sensitivity analysis suggests that Fab and Fc galactosylation are predominantly impacted by GalT activity and weakly affected by q_P

The relationship between the model inputs and the galactosylation index in both regions can be explained by the high first order index for GalT and a low first order index value for q_P (Figure 3.4C and 3.4E). This would mean that q_P has a small effect on Fab galactosylation and most of the variation in the data can be explained by changes to GalT activity. Interestingly, this was the only case with non-zero second order effects, suggesting that Fab galactosylation can be impacted by both q_P and GalT with a second order index of 0.044. A second order index value of 0.044 indicates that variations in Fab galactosylation are also weakly correlated to the interaction between q_P and GalT.

For Fc galactosylation, the first order index was equal to the total order index, thus suggesting that there are no second order interactions between q_P and GalT on the Fc galactosylation index. Surface plot visualization (Figure 3.4D) suggests that there is a small decrease in the Fc galactosylation index with increasing q_P and a drastic decrease in Fc galactosylation index with reduction in GalT activity. Figure 3.4F shows that the Fab galactosylation was nearly 100% for most of the input values. However, at high q_P values and low GalT activity, there was a drastic drop in Fab galactosylation index to as low as 85%. This rapid drop in Fab galactosylation (highlighted in Figure 3.4F) occurs when normalized GalT activity drops below 0.9 and q_P values are greater than 10 pg/cell/day. The interactive effect of q_P and GalT on Fab galactosylation was not surprising as the second order sensitivity analysis index (0.044) was not negligible.

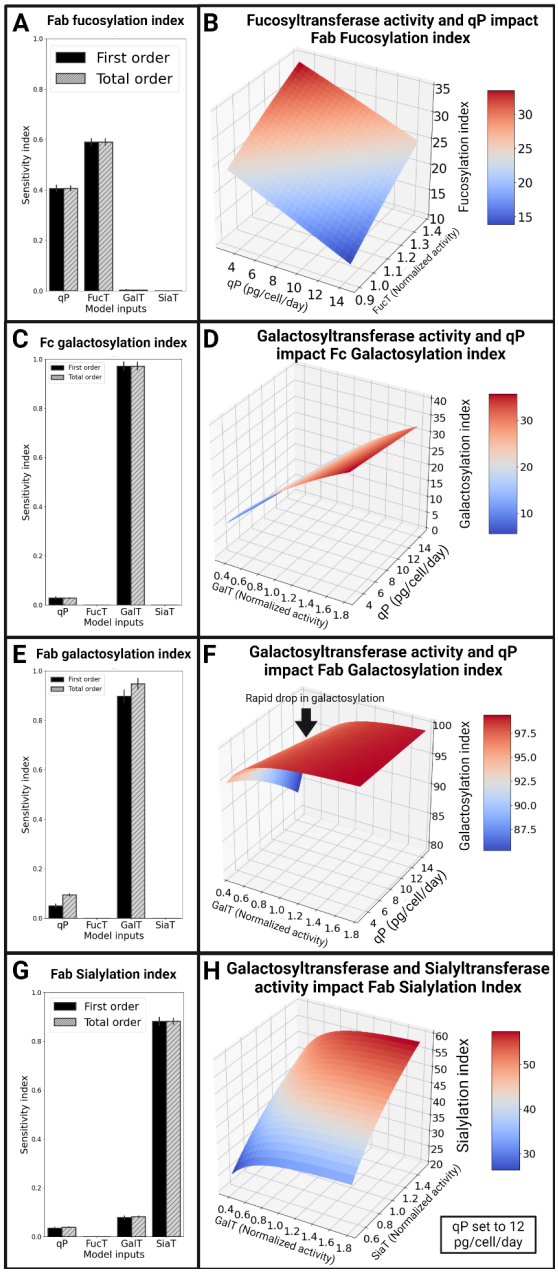

Figure 3.4: Sobol sensitivity analysis reveals the impact of model inputs (q_P, FucT, GalT, and SiaT activities) on the model outputs (glycosylation indices at both antibody sites). Figures 3.4A, 3.4C, 3.4E, and 3.4G depict the first order and total order indices of the sensitivity analysis. Figure 3.4B, 3.4D, 3.4F, and 3.4H depict 3D surface plots to visualize the impact of the most important inputs on the model outputs. Rapid drop in Fab galactosylation at high q_P and GalT activity has been highlighted in Figure 3.4F with an arrow.

3.4.2.3 The model suggests that Fab sialylation is predominantly based on the SiaT activity but can also be influenced by GalT activity and q_P

The high first order index for SiaT (Figure 3.4G) indicates that most of the variation in Fab sialylation index can be attributed to SiaT activity, but small values of first order index were also observed for q_P and GalT. This suggests that a small fraction of the variation in the Fab sialylation index can also be attributed to q_P and GalT activity. The first order indices were equal to the total order indices (Figure 3.4G), indicating no significant interaction effects between the model inputs on the Fab sialylation index. The surface plots shown in Figure 3.4H indicate that Fab sialylation index is impacted by GalT activity in the range (0.4 to 0.9) at a fixed q_P value (12 pg/cell/day). The surface plot on the impact of q_P and GalT on Fab galactosylation index (Figure 3.4F) shows that at high q_P (12 pg/cell/day) and low GalT activity (below 0.9), the Fab galactosylation starts dropping (highlighted in Figure 3.4F), thus resulting in a drop in galactosylated substrate concentrations for the SiaT enzyme. This suggests that, for a high q_P (values greater than 10 pg/cell/day), normalized GalT activity approximately below 0.9 can impact Fab sialylation, but for large GalT activity (greater than 0.9), Fab sialylation is only driven by the SiaT activity.

3.4.3 Glycosylation model provides mechanistic insights into the effect of bioreactor pH on site specific N-linked glycosylation

3.4.3.1 Model predicted enzyme activities provide insights into how bioreactor pH affects Fab and Fc glycosylation

Regression of enzyme activity to experimental data from Chapter 2 [224] on the effect of bioreactor pH on Fab and Fc glycosylation (Figure 3.5A) resulted in computed enzyme activities displayed in Figure 3.5B. These computed enzyme activities for the bioreactor at pH 7.00, day 4 have been set to enzyme concentrations obtained from the

literature [94] and the relative enzyme activity values for all the data points have been normalized by these literature values.

The sensitivity analysis results discussed in Section 3.4.2 revealed that galactosylation is primarily driven by GalT activity. Thus, we can see large variations in the model-computed GalT activity (Figure 3.5B). Bioreactor pH 7.00 had the highest galactosylation throughout the culture duration, and as a result, it had the highest model computed GalT activity. Similarly, GalT activity increased with culture duration for bioreactor runs at pH 7.00 and 7.25 consistent with the increased galactosylation with culture duration. For bioreactor at pH 7.25 on day 4, the model computed activity for GalT was below 0.9. From the sensitivity analysis in Section 3.4.2 it was revealed that Fab galactosylation starts dropping in this range (0.4 to 0.9) for GalT at high q_P (Figure 3.4F). This can be confirmed from the experimental data. Bioreactor pH on day 4 exhibited the lowest Fab galactosylation (Figure 3.5A) while compared to all the experimental data points.

Surprisingly, there were very minimal differences in model computed FucT activity throughout the culture durations for all conditions even though large variations in experimental Fab fucosylation were observed. The model computed FucT activity was slightly higher for bioreactor pH 7.25 than pH 7.00, even though pH 7.25 resulted in the lowest fucosylation index. The sensitivity analysis performed in Section 3.4.2 revealed that variations in Fab fucosylation are negatively correlated to the q_P. Hence, higher q_P at pH 7.25 required higher FucT activity to predict fucosylation at the experimentally observed levels.

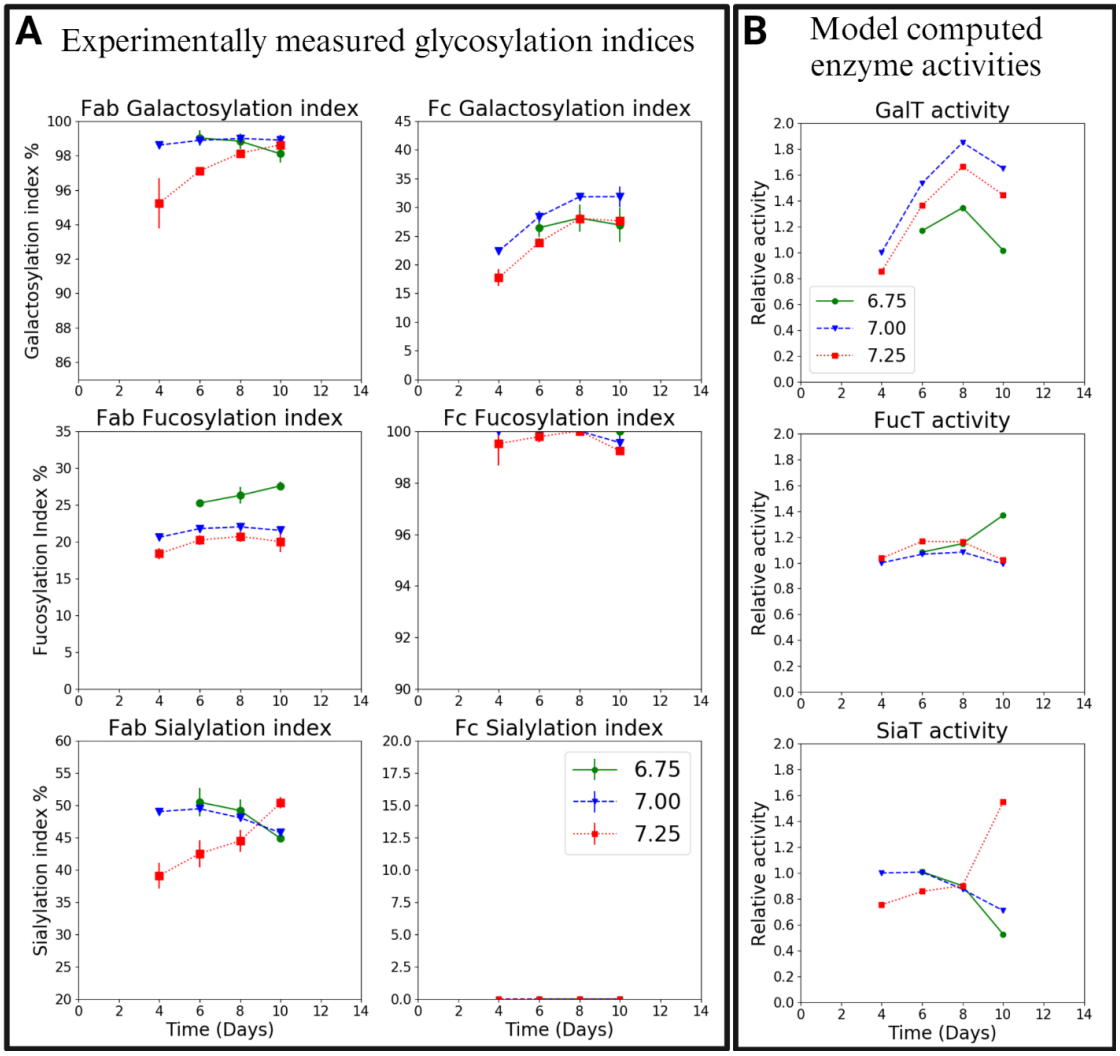

Figure 3.5: Model computed enzyme activities provide insights into the effect of bioreactor pH on site-specific N-linked glycosylation. Data shown for biological duplicates. A) Fab and Fc glycosylation indices. B) Model computed enzyme activities.

For pH 7.00 and 6.75 bioreactor experiments, the sialylation index and model computed SiaT activity are decreasing while the galactosylation index and model computed GalT activity are increasing. Sensitivity analysis revealed that after the galactosylation index reaches 100% in the Fab region, the galactosylation and

sialylation are not correlated. Changes in sialylation can be independent of galactosylation. The lowest sialylation index was observed for bioreactor operated at pH 7.25 on day 4. The model computed enzyme activities of SiaT and GalT at this data point indicate that there was a drop in Fab galactosylation that impacted Fab sialylation and the reduction in SiaT activity also impacted Fab sialylation.

3.4.3.2 Testing model capabilities using measurements of enzyme mRNA expression levels and intracellular nucleotide sugar concentrations

Experimental validation of model computed enzyme activities is complex as they are related to the glycosyltransferase enzyme expression level, nucleotide sugar transporter expression levels, intracellular nucleotide sugar concentration, and also inhibitors of glycosylation such as ammonia. In an effort to validate model computed activities, we measured the mRNA expression levels of seven N-linked glycosylation enzymes (GnT1, β4GalT1, β4GalT2, β4GalT3, ST3Gal3, ST3Gal4, and FUT8) and four nucleotide sugar transporters (SLC35A1, SLC35A2, SLC35C1, and SLC35D1). mRNA levels are not necessarily proportional to enzyme levels, but still, they provide a measure of likely enzyme protein levels and are widely used in this context in the literature. We also measured intracellular concentrations of UDP-Gal, GDP-Fuc, and CMP-Neu5Ac. The mRNA expression levels of β4GalT1, SLC35A2, FUT8, SLC35C1, ST3Gal3, and SLC35A1 are plotted in Figure 3.6A, while the mRNA expression levels of the rest of the enzymes are plotted in supplementary Figure B12.

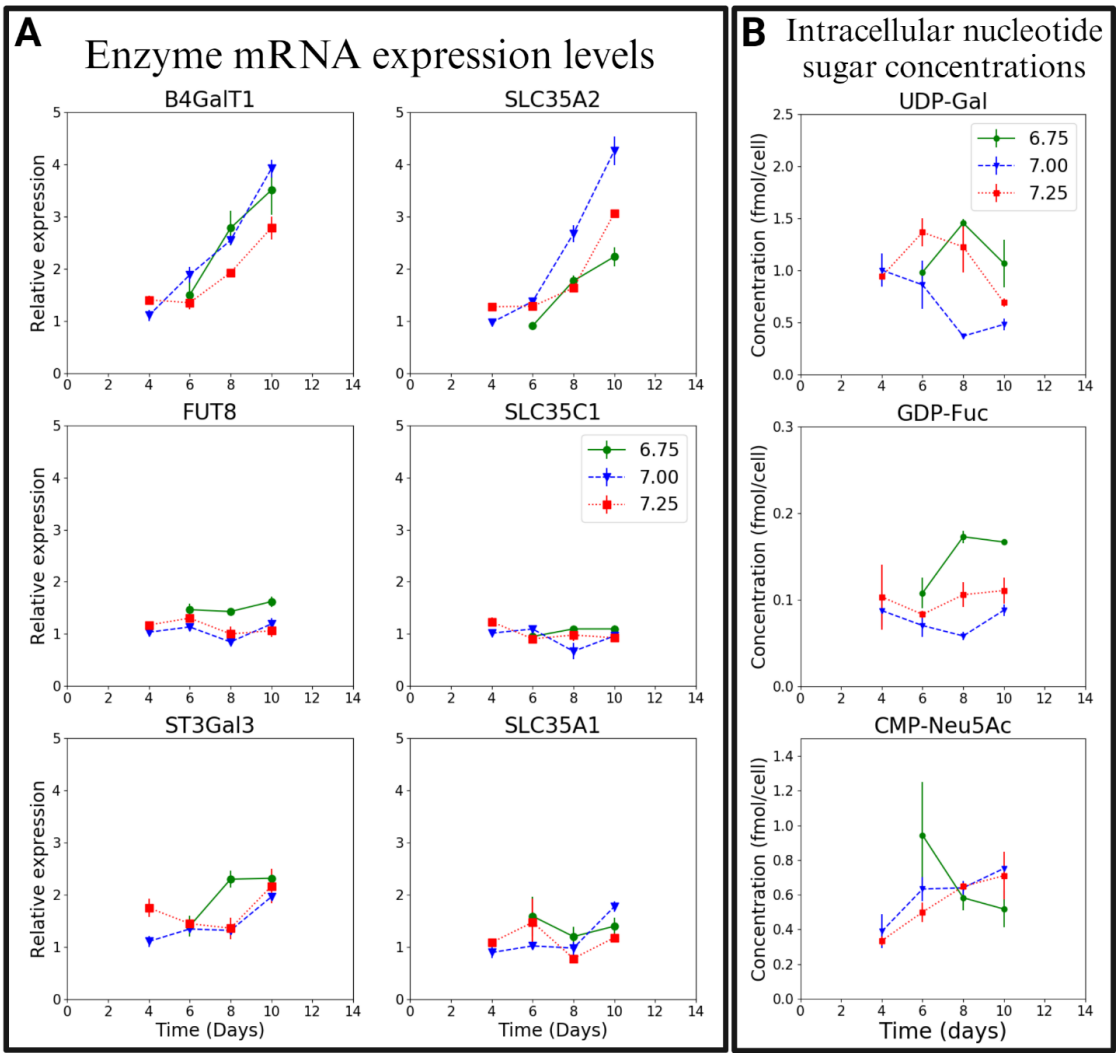

Figure 3.6: Bioreactor pH and culture duration impact enzyme mRNA expression levels and intracellular nucleotide sugar concentrations. Data shown for biological duplicates. A) Glycosylation enzyme mRNA expression levels. Data normalized to bioreactor pH 7.00 day 4. B) Measurements of intracellular nucleotide sugar concentrations.

According to Figure 3.5B, the model computed enzyme activity for GalT was increasing with culture duration and it was the highest at pH 7.00. Measurements of β4GalT1 and SLC35A2 (UDP-Gal transporter) mRNA levels agreed with these trends.

β4GalT1 has been shown to be the most dominant galactosyltransferase for N-linked glycosylation of recombinant glycoproteins in CHO cells [244]. Expression of β4GalT1 and SLC35A2 were highest at pH 7.00 and increased with increasing culture duration. Surprisingly, measurements of intracellular UDP-Gal concentrations did not agree with the glycosylation trends. This suggests that the interdependence between the cytosolic UDP-Gal concentrations and the UDP-Gal transporter can influence the availability of nucleotide sugar in the Golgi for N-linked glycosylation. In the context of the fact that ammonia is a known inhibitor of galactosylation [216], it should also be noted that the highest ammonia concentrations (supplementary Figure B13) were observed in bioreactors operated at pH 6.75.

The model computed enzyme activity for FucT showed little variation with culture duration (Figure 3.5B). This is consistent with experimental measurements of mRNA levels of FUT8 and SCL35C1 (GDP-Fuc transporter), and intracellular GDP-Fuc concentrations. Figure 3.6A shows that the mRNA levels of FUT8 and SCL35C1 did not vary with culture duration and showed minimal variations with bioreactor pH. At bioreactor pH 6.75 we measured slightly higher FUT8 mRNA expression levels and intracellular GDP-Fuc concentrations, which is consistent with model-predicted results.

The measurement of mRNA levels for ST3GalT3, ST3GalT4, SCL35A1 (CMP-Neu5Ac transporter), and intracellular CMP-Neu5Ac concentrations showed very limited agreement with the experimentally measured glycosylation data, as well as the model-computed enzyme activities. The experimentally measured data and the model computed enzyme activities (Figure 3.5) indicate that sialylation decreases with increasing culture duration for bioreactors at pH 6.75 and 7.00. However, the mRNA levels for ST3GalT3, ST3GalT4, SCL35A1 and the intracellular CMP-Neu5Ac

concentrations data (Figure 3.6) do not suggest a drop in sialylation. This is probably because sialylation is impacted by some other process parameter as well. For example, ammonia is known to significantly impact N-linked glycosylation [127]. More specifically, ammonia can negatively impact galactosylation and sialylation [73, 216]. In this context, we note that bioreactors operated at pH 6.75 and 7.00 had higher ammonia accumulation (supplementary Figure B13) than bioreactors operated at pH 7.25.

3.4.4 Testing the predictive capability of the model using galactose and $MnCl_2$ medium supplementation

The goal of this study is twofold. The first is to quantify the effect of galactose and $MnCl_2$ supplementation on site-specific N-linked glycosylation and the second is to demonstrate that the site-specific N-linked glycosylation model is capable of predicting the glycan fractions when these nutrients or cofactors are supplemented.

3.4.4.1 Galactose supplementation impacts intracellular nucleotide sugar synthesis

CHO cells were cultivated in biological duplicates for the five different conditions shown in Table 3.1. The first flask was extracted on day 4 of the culture to determine the glycosylation fractions prior to the start of galactose or $MnCl_2$ feeding. The cell culture data from the flasks are plotted in Figure 3.7 and supplementary Figures B14. Supplementation of galactose or $MnCl_2$ did not impact growth rates (Figure 3.7B), glucose uptake rates, lactate production rates or ammonia concentrations (supplementary Figure B14). The concentrations of galactose in the flask can be seen in Figure 3.7C. Increasing galactose addition by four-fold (flask 4 from Table 3.1) resulted in statistically significant (p value < 0.01) increases in galactose uptake rates (Figure

3.7C). The enzyme mRNA expression levels of all 11 targets were measured for all conditions (supplementary Figure B15). β4GalT1 and SLC35A2 expression increased from day 4 to day 9 but the expression levels were almost the same on day 9 for all different conditions (Figure 3.7D). Galactose or $MnCl_2$ supplementation did not result in changes to β4GalT1 and SLC35A2 mRNA expression levels. Galactose supplementation resulted in increase in intracellular UDP-Gal concentrations (Figure 3.7E) without impacting other nucleotide sugar concentrations (supplementary Figure B16). The q_P for all the different conditions were the same (Figure 3.7F).

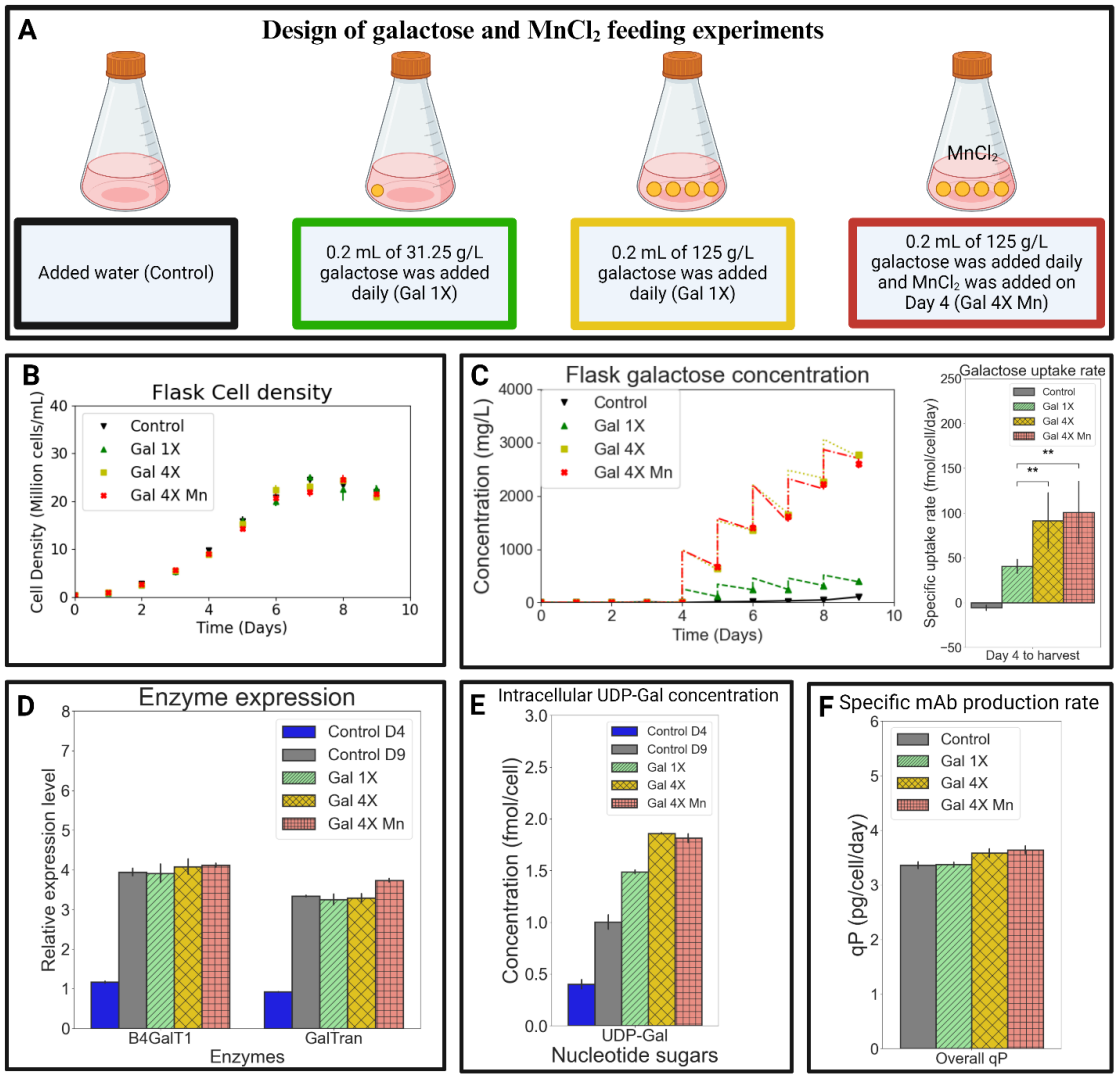

Figure 3.7: Galactose and $MnCl_2$ supplementation lead to increased galactose uptake rates and intracellular UDP-Gal concentrations without impacting β4GalT1 and SLC35A2 mRNA expression levels and the cell specific mAb production rate (q_P). Data shown for biological duplicates. ** represents p-value below 0.01 in a two-tailed T-test. A) Experimental design. B) Viable cell density. C) Galactose concentrations and galactose uptake rates. D) β4GalT1 and SLC35A2 mRNA expression levels. E) Intracellular UDP-Gal concentrations. F) Cell specific mAb production rates (q_P).

3.4.4.2 Galactose and $MnCl_2$ supplementation impacts Fc glycosylation without impacting Fab glycosylation

Site-specific glycan analysis showed that the 1X galactose supplementation condition resulted in a slight increase in Fc galactosylation as demonstrated by increases in FA2G2 and FA2G1 glycoforms and reduction in the FA2 form (Figure 3.8B). Adding 4X the amount of galactose led to more significant increases in FA2G2 and FA2G1 glycoforms and reduction in FA2 glycoforms. Addition of galactose and $MnCl_2$ led to the highest increase in Fc galactosylation by increasing FA2G2 and FA2G1 fractions and reducing FA2 fractions. We also observed increased galactosylation from day 4 to day 9 and increase in β4GalT1 and SLC35A2 mRNA expression levels from day 4 to day 9 (Figure 3.7D). This trend was also observed in bioreactor data shown earlier. Supplementation of galactose or $MnCl_2$ did not impact Fab glycosylation (Figure 3.8A).

3.4.4.3 Model predicts Fab and Fc glycan fractions after modulating GalT activity

The data from Figure 3.7 were used as input to the model to predict glycosylation in the Fab and Fc regions. The saturation constants were regressed earlier, the enzyme activities for FucT, SiaT, and GalT were adjusted based on the control experiment (flask 2 from Table 3.1). The results from this regression are shown in Figure 3.8 and labeled as 'Reg'. Figure 3.7D shows that there was no increase in β4GalT1 and SLC35A2 mRNA expression levels for all the different conditions. Thus, we assumed that the transport of intracellular nucleotide sugar to the Golgi can be scaled linearly with increasing intracellular UDP-Gal concentrations. A scaling factor of 40X has been assumed in the literature [94] and the same factor was used in our study. This means that the concentration of Golgi UDP-Gal is assumed to be 40X the intracellular UDP-Gal concentration. Manipulation of this one variable led to predictions of glycan

fractions in the Fab and Fc region for the 1X galactose addition and 4X galactose addition conditions as shown in Figure 3.8. The predictions (Figure 3.8; labeled as 'Pred') showed that FA2G2 and FA2G1 fractions in the Fc region increased while FA2 fractions decreased. The predictions also showed that there is no impact on Fab glycosylation as the galactosylation index is already 100%.

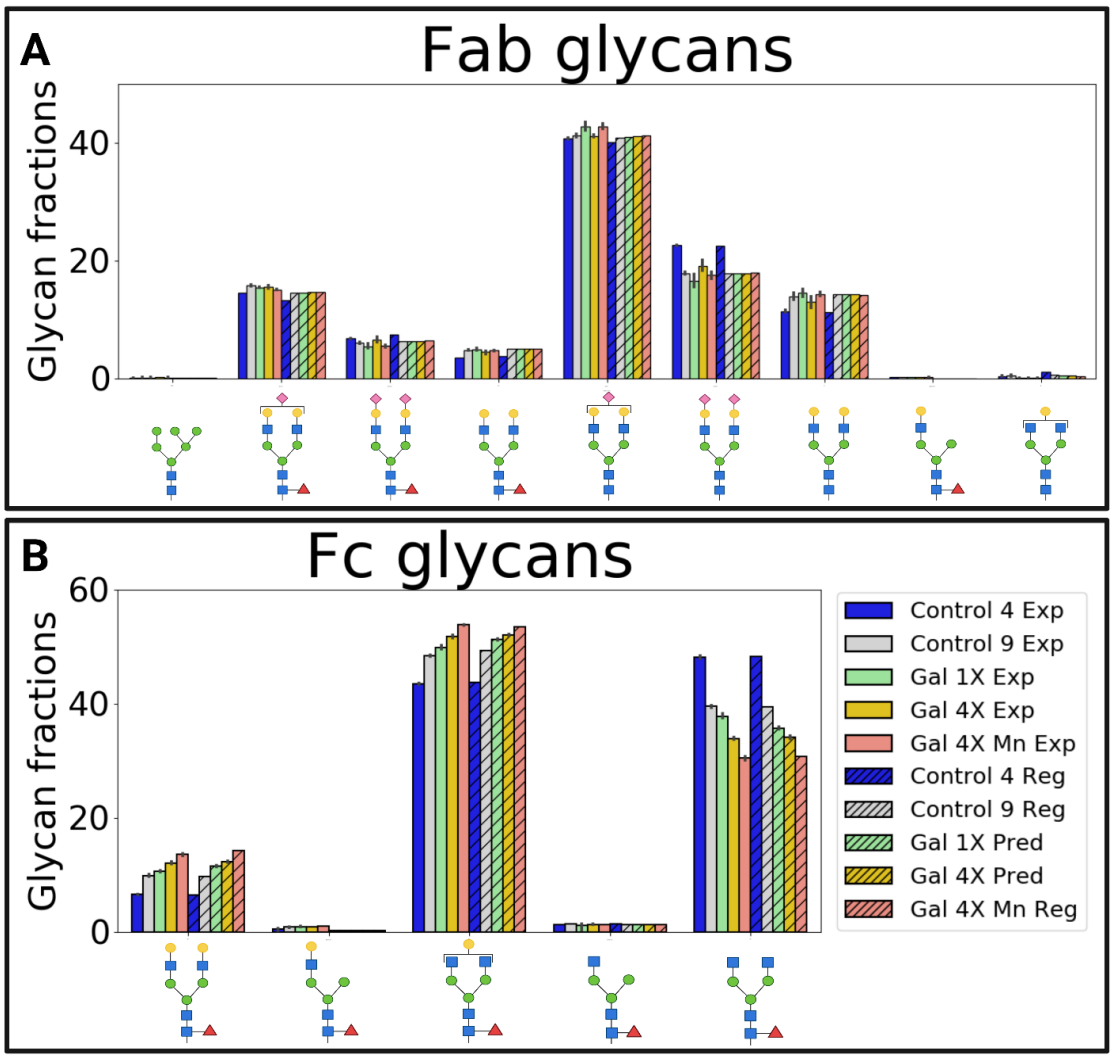

Figure 3.8: Site-specific N-linked glycosylation model predicts Fab and Fc glycan fractions in galactose and $MnCl_2$ supplemented cultures. Data shown for biological duplicates. Adjusted glycan fractions calculated from Equation (3.5). A) Fab glycan fractions. B) Fc glycan fractions.

The model was subsequently regressed to the $MnCl_2$ addition glycosylation data while using the same nucleotide sugar concentration measured for the Gal4X condition. In addition to the increased UDP-Gal concentration, model computed GalT activity had to be increased by 12% from the control condition to predict the increased

galactosylation observed when $MnCl_2$ was supplemented. Thus, $MnCl_2$ addition led to a 12% increase in GalT activity. $MnCl_2$ and galactose supplementation together led to the largest increase in galactosylation.

3.5 Conclusions

This chapter of the dissertation aimed to develop a kinetic modeling framework for site-specific N-linked glycosylation. The relationship between the model independent variables (q_P, GalT, FucT, and SiaT) and outputs (glycosylation indices at both the sites) was better understood by performing Sobol sensitivity analysis. This model was used to understand the difference in glycans observed in the two sites of the IgG molecule. The parameters of the mathematical model provided information on interactions between the glycosylation sites and the glycosyltransferase enzymes. These interactions largely influence the type of glycan observed at both the sites. The effect of bioreactor pH on site-specific N-linked glycosylation of the mAb produced from CHO VRC01 cells was mechanistically examined using the mathematical model and experimental measurements of enzyme expression and intracellular nucleotide sugar. The results showed us that the Fab fucosylation index is correlated to fucosyltransferase activity and q_P. The Fab and Fc galactosylation index is mostly influenced by galactosyltransferase activity and weakly correlated to q_P. The Fab sialylation index is largely influenced by sialyltransferases activity but if the galactosyltransferase activity falls below a certain threshold, then it can also impact sialylation. These findings improve our understanding of the effect of a critical process parameter on site-specific N-linked glycosylation in line with the quality by design framework.

The impact of increasing the activity of the galactosyltransferase enzyme on the site-specific N-linked glycans was experimentally studied by galactose and $MnCl_2$

supplementation. The mathematical model was utilized to predict the impacts of increasing intracellular UDP-Gal concentrations on site-specific N-linked glycosylation. Mathematical models in the literature can predict intracellular nucleotide sugar concentrations by modeling the kinetic rates of nucleotide sugar synthesis as a function of galactose uptake rates [73, 114]. The site-specific N-linked glycosylation model can be combined with these kinetic models of nucleotide sugar metabolism to predict dynamic trends of N-linked glycosylation at multiple sites. The model was also used to predict the effect of $MnCl_2$ on galactosyltransferase activity, showing the potential of this framework in guiding process development of novel therapeutics and biosimilars with multiple N-linked glycosylation sites.

Although this study was focused on an IgG with Fab and Fc glycans, this modeling framework can be applied to examine the site-specific heterogeneity of glycans of other protein therapeutics. If sufficient experimental data are available, more detailed kinetic equations accounting for various glycosylation modulators can also be easily incorporated into this site-specific glycosylation framework by modifying Equation (3.3).

Chapter 4

INCORPORATING PROCESS CONDITIONS INTO DYNAMIC METABOLIC FLUX ANALYSIS MODEL TRAINED ON FED-BATCH DATA TO PREDICT INTENSIFIED FED-BATCH AND PERFUSION BIOREACTOR PERFORMANCE

4.1 Introduction

Development of a process to produce mAbs in CHO cells involves cell line development, media optimization, feeding schedule optimization, bioreactor parameter optimization and scale up [7]. These processes are most often interlinked and require experimentation to identify the optimal operating conditions as a large number of parameters (pH, temperature, media composition, feeding schedules, and seeding cell density) can influence bioreactor performance. Changes in bioreactor temperature or pH results in changes in nutrient uptake rates or inhibitor production rates [118, 124]. Hence, expensive and laborious experiments are required to screen many different media formulations, feeding schedules, and bioreactor conditions to maximize mAb titers. Performing limited experiments in order to develop mathematical models and using simulations for process development can aid with reducing experiments and reduce time required for process development, hence reducing costs and improving the time to market. Mathematical models can help streamline and optimize the manufacturing and process development-based research and development tasks by directing desirable experiments to achieve optimal process conditions [9, 10].

Modeling of mammalian cell culture is complex as process performance can be impacted by concentrations of many metabolites (amino acids, sugars, vitamins, and

inhibitors) and process conditions (pH, temperature, dissolved oxygen). Further complexities arise as the concentrations of these metabolites can change with culture duration in fed-batch processes. Kinetic models provide a simple empirical framework that relates nutrient concentrations to their uptake rates and cell growth rates via semi-empirically constructed ordinary differential equations. Kinetic models can be solved to resolve time profiles of nutrients. Kinetic models have been applied to determine feeding schedules [20]. Stoichiometric models involve reaction networks that are used to understand the relationship between different metabolites. Stoichiometric models are non-parametric and rely on metabolic fluxes to make predictions. The non-parametric nature of stoichiometric models allows for incorporating many metabolites and many reactions. This has led to the development of genome scale models that have been used for media formulation [162, 168]. The combination of kinetic and stoichiometric models for mammalian cell metabolism can help project dynamic profiles while maintaining mechanistic fidelity [196]. Cell culture profiles also undergo switches in metabolism and these models have shown good applicability to model lactate and ammonia switch [62], hence providing a system that can be more dependable while making predictions outside the model training design space.

Experimental data (Chapter 2) has shown that bioreactor pH significantly affects CHO cell metabolism by impacting glucose uptake rates, lactate production rates, ammonia production rates, amino acid uptake rates, cell growth rates, and cell specific mAb production rates [224]. However, there are very few models in the literature that include the effect of bioreactor pH while modeling CHO cell metabolism, thus limiting their applicability.

In this work, we develop a mathematical model that can predict dynamic concentration profiles of glucose, lactate, ammonia, essential amino acids, nonessential amino acids, cell density, and mAb as a function of bioreactor pH, basal media composition, feed media composition, media supplementation strategy and seeding cell density. We have used a dynamic metabolic flux analysis (DMFA) model based on the basic principles published in the literature [62]. Literature DMFA models do not include the impacts of bioreactor pH and are capable of predicting dynamic profiles of concentrations of cell density, mAb, glucose, lactate, ammonia, and eight amino acids. The DMFA model constructed in our work is capable of predicting dynamic profiles of concentrations of cell density, mAb, glucose, lactate, ammonia and eighteen amino acids. To expand the DMFA model to include eighteen amino acids, we have curated a reaction network that contains reactions involved in amino acid metabolism. The model was trained on fed-batch data containing concentrations of cell densities, glucose, lactate, ammonia, amino acids, and antibody generated by culturing the CHO VRC01 cell line at pH 6.75, 7.00 and 7.25 [224], thus providing us with a model that is capable of predicting the cell culture performance as a function of bioreactor pH, media composition, feeding schedule, and cell density. Development of this model allowed us to assess various applications of mathematical modeling within the context of therapeutic glycoprotein process development.

The production of therapeutic glycoproteins is moving away from the traditional fed-batch platforms to intensified fed-batch cultures of perfusion bioreactors. Intensified fed-batch cultures involve inoculating the bioreactors at high cell densities to achieve similar antibody titers to fed-batch processes with much shorter culture duration by skipping the initial growth phase of traditional fed-batch cultures. Shorter

run times are desirable as they can maximize the utilization of large scale production facilities. [245]. Perfusion bioreactors utilize cell retention devices and continuous media exchange to achieve large cell densities (greater than 100 million cells/mL) and can operate for much longer durations while continuously producing mAb. At high cell specific productivity, perfusion bioreactors can lead to better process economics but also lead to increased process complexity [246]. We utilized the DMFA model trained on fed-batch data to make predictions of concentrations of viable cells, glucose, lactate, ammonia, amino acids, and mAb titers in intensified fed-batch and perfusion cultures by changing the seeding cell density, bioreactor pH, feed media supplementation or process mode (perfusion). This was followed by performing intensified fed-batch and perfusion experiments to validate the model predictions and demonstrate the reliability of the DMFA model. The model was subsequently tested on predictions of completely different media by utilizing the availability of media composition for the AMBIC media [205].

4.2 Materials and methods

4.2.1 Cell culture

4.2.1.1 Cell line and media details

CHO-K1 VRC01 cell line (Clone A11 from the Vaccine Research Center at the National Institutes of Health) stably expressing a broadly neutralizing anti-HIV monoclonal antibody was used in this study. CHO cells were grown in 125 mL Erlenmeyer flasks at 30 mL culture volume for passage and inoculum preparation. Erlenmeyer flasks used for inoculation were seeded at 0.4 million cells/mL and incubated at 37 °C in an incubator configured to 85% relative humidity, 20% O_2 and

5% CO_2. HyClone Actipro media (Cytiva, catalog number: SH31039.02) supplemented with 6 mM L-glutamine was used as the basal media. HyClone Cell Boost 7a (Cytiva, catalog number: SH31119.02) supplement and HyClone Cell Boost 7b (Cytiva, catalog number: SH31120.01) supplement were used as feed media.

4.2.1.2 Protocol for fed-batch bioreactor operation

Fed-batch experimental data taken from Chapter 2 [224]. Fed-batch experiments were performed in 1 L Eppendorf BioFlo 120 bioreactors at 750 mL culture. The temperature of the bioreactor was controlled at 37 °C. The setpoint for dissolved oxygen was 40% and agitation was set to 90 rotations per minute using a pitched blade impeller. Three different bioreactor pH values (6.75, 7, and 7.25) were studied by running biological triplicates for each condition. The inoculation seeding cell density in the reactors was 0.4 million cells/mL. The bioreactor pH was controlled by sparging with CO_2 and pumping 6% sodium bicarbonate solution. A significant amount of sodium bicarbonate was added to the reactor for the pH 7.25 condition due to high lactate accumulation. The feeding protocol involved adding 22.5 mL of HyClone Cell Boost 7a and 2.25 mL of HyClone Cell Boost 7b daily from day 3 onwards. Starting on day 5, glucose concentration was increased to 9 g/L every day by spiking 45% (w/v) sterile glucose solution (Sigma Aldrich, catalog number: 50-99-7). Antifoam C was added when foaming was observed. Samples were collected daily for measurements of cell counts, glucose concentration, lactate concentration, ammonia concentration, amino acid concentration and titer. Cell counts and viability were measured by using the trypan blue assay on the DeNovix automatic cell counter.

4.2.1.3 Protocol for intensified fed-batch experiment

Intensified fed-batch (high seeding cell density) experiments were conducted by inoculating CHO cells in 1 L Eppendorf BioFlo 120 bioreactors (750 mL culture volume) at a seeding density of 5 million cells/mL. To demonstrate the predictive capability of the model, we decided to operate the intensified fed-batch bioreactor at a pH of 7.12. The pH was controlled by sparging with CO_2 and pumping 6% sodium bicarbonate solution when necessary. The dissolved oxygen was set to 40% and controlled by sparging with oxygen and air. The agitation was set to 90 RPM using a pitched blade impeller. The Actipro basal media along with the HyClone 7a and 7b feed media was used. Glutamine was only added initially to bring its concentration to 6 mM. On day 1 of the culture, 22.5 mL of HyClone Cell Boost 7a media and 2.25 mL of HyClone Cell Boost 7b media were added to the reactor. From day 2 onwards, 38 mL of HyClone Cell Boost 7b media and 3.8 mL of HyClone Cell Boost 7b media were added to the bioreactor. On day 1, glucose was supplemented to the reactor to bring the glucose concentration to 5 g/L. From day 2 onwards, glucose was supplemented to the reactor to maintain a concentration of 9 g/L. Culture was stopped when viability dropped below 80%. Since this was used for model validation, only a single experiment was performed. Daily measurements of glucose, lactate, ammonia, amino acids, antibody titer, cell density were collected.

4.2.1.4 Protocol for perfusion bioreactor operation

Perfusion experiments were performed in a 1 L Eppendorf BioFlo 320 bioreactor (Eppendorf) at a working volume of 864 mL. For cell retention, a 20 cm long modified polyether sulfone (mPES) hollow fiber filter (Repligen, USA) with 0.65 µm pore size

was operated in tangential flow filtration (TFF) mode at a target shear rate of 1000 s^{-1} with a centrifugal pump (PuraLev® i30MU, Levitronix, Switzerland).

The operational workflow followed a similar protocol to the fed-batch bioreactor experiments described earlier. The bioreactor was inoculated at a target seeding density of 0.4 million cells/mL in HyClone Actipro basal media with 0.6 mM supplemental L-glutamine. Batch mode operation ensued until day three of culture, after which perfusion was initiated at a rate of 1 vessel volume per day (vvd^{-1}) using a perfusion media formulation adapted from an existing fed-batch protocol (Actipro media supplemented with 3% Hyclone 7a and 0.3% Hyclone 7b media). The perfusion rate was set by the harvest pump and a scale interfaced with the media pump was used to maintain a constant working volume. Cell growth continued until a pre-determined viable cell density (VCD) setpoint, after which a daily manual cell bleed was initiated to maintain steady state at the desired setpoint. The VCD setpoint and required cell bleed were calculated from prior growth characterization studies in spin tube semi-perfusion cultures. The culture temperature was maintained at 37 °C. Bioreactor pH was controlled at 7.00 via CO_2 sparging. The perfusion media was prepared at a pH of 7.40, eliminating the need for external base addition to control the pH. The dissolved oxygen was maintained at 40%. The agitation rate was set to 110 rpm at inoculation and increased in increments of 20 – 40 rpm per day, reaching 270 rpm by the time the VCD setpoint was attained. Daily samples were collected for offline measurements of VCD, % viability, glucose, lactate, ammonia, amino acids, and mAb titer as described for the fed-batch experiments. During steady state operation, which continued until retention module fouling, sample collection took place before and after the daily manual cell bleed.

4.2.1.5 Measurement of substrate, metabolite and IgG concentrations

Measurements of glucose and lactate were performed using the YSI 2900 bioanalyzer. Ammonia concentrations and osmolality were measured using the BioProfile Flex bioanalyzer (Nova). Amino acid concentrations were measured using an Agilent HPLC 1260 instrument. Amino acid standards (Agilent, catalog number: 5061) and a HPLC column (AdvanceBio Amino Acid Analysis column) were purchased from Agilent (catalog number: 695975-322). A guard column was also purchased from Agilent (catalog number: 823750-946). The amino acids were derivatized with OPA for primary amino acids and FMOC for secondary amino acids as per the protocol provided by Agilent. Antibody titers were measured on the Agilent HPLC 1260 using protein A chromatography using a POROS A HPLC column (catalog number: 1502226; ThermoFisher). As per the protocol provided by ThermoFisher, mobile phase A consisted of 50 mM phosphate and 150 mM NaCl at pH 7.0. Mobile phase B consisted of 12 mM hydrochloric acid at pH 1.9. A UV detector (280 nm) was used to detect the mAbs. Additional method details are provided in the literature [224]. HPLC grade IgG standard (catalog number: MFCD00163923) was purchased from Sigma-Aldrich.

4.3 Model formulation

4.3.1 Reaction network

Performing metabolic flux analysis requires a reaction network that is representative of core metabolic reactions of CHO cells. In addition to a reaction network, metabolic flux analysis utilizes nonlinear objective functions such as the sum of squared differences between experimentally measured fluxes and the model computed flux along with the pseudo steady state assumption, resulting in constraints (mass balances) for the fluxes. Solving constrained nonlinear optimization problems is

challenging. Hence, it is required to design a reaction network and objective function that can be solved by using widely available nonlinear programming solvers that provide tractable solutions. To achieve this, we developed a reaction network consisting of reactions describing glycolysis, TCA cycle, amino acid metabolism, urea cycle, custom IgG formation, and biomass synthesis. We started with a reaction network from the literature consisting of 100 reactions and 72 metabolites [202]. Certain metabolites were not relevant for our applications and their reactions were removed. A few reactions consisted of very small or very large stoichiometric reaction coefficients; these reactions were scaled appropriately and lumped together whenever possible. This resulted in a reaction network consisting of 69 reactions and 42 metabolites shown in Figure 4.1 and supplementary Table C1.

Solving nonlinear optimization problems with linear constraints often involves finding the inverse of the stoichiometric matrix. These problems are challenging to solve if the stoichiometric matrix is ill-conditioned. Hence, the reaction network has been formulated to yield a low condition number (18). The IgG formation reaction was developed based on the amino acid sequence of the mAb used in this study. This reaction was scaled by dividing the stoichiometric coefficients of the amino acids with the total number of carbon atoms per molecule of mAb. Hence, the flux for the IgG reaction is also normalized by the same value while performing metabolic flux analysis.

Biomass (CHO cells) require lipids, proteins, RNA and DNA to replicate. Hence, the biomass reaction should reflect the composition of lipids, proteins, DNA and RNA. Literature reaction networks have utilized this information to calculate a biomass reaction for CHO cells that contains amino acids and glucose [72]. This reaction has been used for our study with slight modifications to reflect more recently published

CHO cell composition data [43]. CHO cell concentrations are measured in units of million cells per mL. Since the biomass reactions have been formulated per Carbon-mol (Cmol) of biomass, we would need to convert the viable cell density units to Cmols to generate the biomass flux in Cmols/cell/day for metabolic flux analysis. Hence, the biomass flux is calculated by utilizing 281 pg/cell and 24.12 g/Cmol of biomass [247].

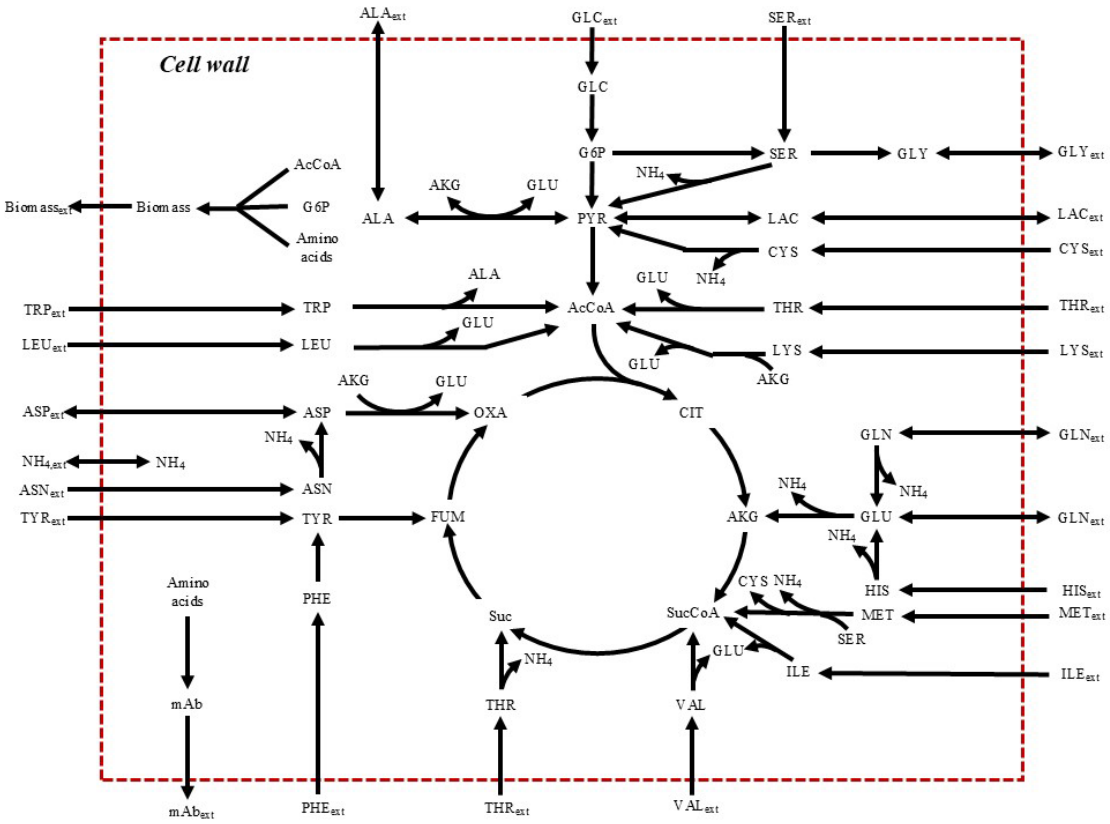

Figure 4.1: Schematic of reaction network used in the dynamic metabolic flux analysis model. Reactions with stoichiometric coefficients are listed in supplementary Table C1.

4.3.2 Kinetic equations

Dynamic metabolic flux analysis integrates stoichiometric modeling with kinetic models or a system of ordinary differential equations (ODEs). These kinetic equations are empirical in nature and various forms of kinetics have been used in the literature [24, 173]. We will be using kinetic equations of Monod type for the growth rates and yield coefficients for the metabolites. The kinetic equations used in this study are described in Table 4.1.

Table 4.1: List of kinetic equations.

Reaction	Equation	Equation number
60	$\mu_{GLC} = \frac{[GLC]}{K_{GLC} + [GLC]}$ $\mu_{GLU} = kdiauxic_{GLU} \frac{[GLU]}{K_{GLU} + [GLU]}$ $\mu_{GLN} = kdiauxic_{GLN} \frac{[GLN]}{K_{GLN} + [GLN]}$ $\mu_{ASP} = kdiauxic_{ASP} \frac{[ASP]}{K_{ASP} + [ASP]}$ $\mu_{ASN} = kdiauxic_{ASN} \frac{[ASN]}{K_{ASN} + [ASN]}$ $\mu_{max,OSMO} = \mu_{max}\left(1 - \frac{1}{1 + e^{(-(Osmolarity - OSMO_A) \times OSMO_B)}}\right)$ $\mu = \mu_{max,OSMO}\mu_{GLC}(\mu_{GLU} + \mu_{GLN})(\mu_{ASN} + \mu_{ASP})$	(4.1)
42	$v_{GLC} = \frac{\mu}{Y_{\frac{X}{GLC}}} + m_{GLC}$	(4.2)
43	$if\ [GLN] > 0: v_{LAC} = Y_{\frac{LAC}{GLC}} v_{GLC}$ $if\ [GLN] = 0: v_{LAC} = \frac{\mu}{Y_{\frac{X}{LAC}}}\left(\frac{[LAC]}{[LAC] + K_{S,LAC}}\right)$	(4.3)
44	$if\ [GLN] > 0: v_{ALA} = Y_{\frac{ALA}{GLC}} v_{ALA}$ $if\ [GLN] = 0: v_{ALA} = \frac{\mu}{Y_{\frac{X}{ALA}}}$	(4.4)

Table 4.1: Continued

45, 49, 50, 51, 53, 54, 55, 56, 59, 61, 62, 63	$v_S = \frac{\mu}{Y_{\frac{X}{S}} \times AAscaling(pH)}$ $S \in \{TRP, ARG, MET, SER, THR, TYR, HIS, ILE, LEU, LYS, PHE, VAL\}$	(4.5)
60	$v_{GLN} = \frac{\mu_{max,OSMO}\mu_{GLC}(\mu_{GLN})(\mu_{ASN} + \mu_{ASP})}{Y_{\frac{X}{GLN}}}$	(4.6)
47	$v_{ASN} = \frac{\mu_{max,OSMO}\mu_{GLC}(\mu_{GLN} + \mu_{GLU})(\mu_{ASN})}{Y_{\frac{X}{ASN}}} + m_{ASN}$	(4.7)
58	$if\ [GLN] > 0: v_{GLU} = Y_{\frac{GLU}{GLN}} v_{GLN}$ $if\ [GLN] = 0: v_{GLU} = m_{GLU}$	(4.8)
48	$if\ [GLN] > 0: v_{ASP} = Y_{\frac{ASP}{GLN}} v_{ASP}$ $if\ [GLN] = 0: v_{ASP} = m_{ASP}$	(4.9)
52	$if\ [GLN] > 0: v_{GLY} = Y_{\frac{GLY}{SER}} v_{SER}$ $if\ [GLN] = 0: v_{GLY} = \frac{\mu}{Y_{\frac{X}{GLY}}}$	(4.10)
46	$v_{AMM} = Y_{\frac{AMM}{GLN}} v_{GLN} - R_{AMM,1}\mu + R_{AMM,2}$	(4.11)
67	$v_{mAb} = q_{mAb}$	(4.12)
Cell death	$\mu_d = \mu_{death,1} + \mu_{death,2}\left(\frac{1}{1 + e^{(-(IVCD - P_{1,death,IVCD}) \times P_{2,death})}}\right)$	(4.13)

Equation (4.1) was used to calculate the cellular growth rates as a function of concentrations of key nutrients, bioreactor pH, and osmolality. CHO cell growth rates can be affected by depletion of nutrients. One of the conditions led to low glucose concentrations. Hence, a glucose term was included in the growth rate. The CHO VRC01 cell line grows rapidly when glutamine is supplemented but can also grow at reduced growth rates after glutamine depletion if sufficient glutamate is present in the culture. Similarly, asparagine depletion can also lead to reduced growth rates if aspartate is present in the cultures. Since the cells exhibit diauxic growth for cultures with glutamine-glutamate and asparagine-aspartate and these metabolites were depleted in the cultures, they have been included in the kinetic equations. Bioreactors operated at pH 6.75 led to very high osmolality. Hence, an osmolality term has been included in the maximum growth rate term to account for reduced growth rates.

Equation (4.2) was used to calculate the glucose uptake rates. The glucose uptake rate is split into its contribution towards growth as well as maintenance. CHO cells exhibited uptake of glucose even in the non-growth phases of the culture. Equation (4.3) was used to calculate the lactate uptake/secretion rates. Flux balance analysis revealed that the early phase of the culture leads to significant pyruvate accumulation due to high glycolysis flux and glutaminolysis [224]. Hence, the shift in metabolism of lactate has been attributed to glutamine depletion. Lactate is initially produced from glycolysis and consumed by the cells after depletion of glutamine. Equation (4.4) was used to calculate the alanine uptake/secretion rates. Alanine is produced from pyruvate and glutamate. Production of alanine was noticed in the early phase of the culture. Hence, the switch in metabolism of alanine from secretion to uptake has also been linked to glutamine depletion. Alanine is also produced from pyruvate that comes from

glycolysis. Hence, its secretion rate has been linked to glucose uptake rate. The essential amino acid uptake rates were calculated from Equation (4.5). The amino acids tryptophan, arginine, methionine, threonine, tyrosine, histidine, isoleucine, leucine, lysine, phenylalanine, and valine were modeled by using a simple yield coefficient ($Y_{\frac{X}{S}}$) as their uptake rates were proportional to the growth rates (supplementary Figure C1). Serine exhibited extremely high uptake in the early phase of the culture and lower uptake rates later in the culture (supplementary Figure C1). Hence, one yield coefficient ($Y_{\frac{X}{SER},A}$) was used for the early phase of the culture and another yield coefficient ($Y_{\frac{X}{SER},B}$) was used for the later phase of the culture. Higher bioreactor pH led to increases in uptake rates of all these amino acids. Hence all the yield coefficients were scaled by using an AAscaling(pH) term that is a quadratic function of bioreactor pH.

Equation (4.6) and (4.7) were used to calculate the glutamine and asparagine uptake rates. CHO cells exhibited diauxic glutamine-glutamate and asparagine-aspartate consumption. Hence, the glutamate contribution to growth rate was removed while calculating glutamine uptake rates. Similarly, the aspartate contribution to growth rate was removed while calculating asparagine uptake rates. Flux balance analysis has shown that glutamate and aspartate were produced from glutamine during the early phases of the culture [224]. Hence, the concentration of glutamine was used to link the switch in metabolism of glutamate and aspartate (Equation (4.8) and (4.9)). The production of glutamate and aspartate were linked to glutamine uptake rates. Glutamate and aspartate uptake rates did not change across the culture phases (supplementary Figure C1). Hence, constant terms ($m_{GLU}\ and\ m_{ASP}$) have been used to model their uptake.

Equation (4.10) has been used to model glycine metabolism. Since glycine was initially produced by the cells and subsequently consumed, its uptake and secretion rates were also linked to glutamine depletion. Flux balance analysis has shown high intracellular glycine production flux from serine [224]. Hence, the kinetic equation links glycine secretion flux to serine uptake rates. Complex ammonia metabolism was described using Equation (4.11). Ammonia was initially produced due to rapid glutamine consumption. Hence, the ammonia secretion rate has been linked to glutamine uptake rate. After glutamine depletion, the cell requires ammonia to synthesis glutamine for biomass synthesis (protein requirements). Hence, its uptake from the media was linked to the growth rates via the parameter $R_{AMM,1}$. The catabolism of all amino acids can also lead to ammonia secretion. This contribution was lumped into one term $R_{AMM,2}$.

The cell specific mAb production rate (Equation (4.12)) was assumed to be constant (q_{mAb}) throughout the culture duration. The cell death rates (Equation (4.13)) were split into two terms. The first term is a constant cell death rate that does not vary with culture duration. Cells in fed-batch cultures can exhibit reduced cell viability at the later stages of the culture even if all the nutrients are supplied, due to accumulation of many growth inhibitors [248]. Since these metabolites are difficult to measure, the cell death rates have been linked to the integral viable cell density (IVCD) in the literature [173] to indicate that, the larger the IVCD, the higher the concentration of all the inhibitors in the media. A sigmoid curve as a function of IVCD was fit in the second term of the cell death equation.

The equations shown above resulted in 52 model parameters. The experimental data (supplementary Figure C1) showed that the bioreactor pH significantly impacted

growth rates, glucose uptake rates, lactate production rates, titer production rates, amino acid uptake rates, and ammonia production rates. Hence out of these 52 model parameters, 6 parameters ($Y_{\frac{X}{GLC}}, \mu_{max}, Y_{\frac{LAC}{GLC}}, q_{mAb}, AAScaling, and\ R_{AMM,2}$) were modeled as a quadratic function of bioreactor pH according to Equation (4.14).

$$Parameter_i(pH) = A_i pH^2 + B_i pH + C_i \tag{4.14}$$

$$Parameter_i \in \left\{Y_{\frac{X}{GLC}}, \mu_{max}, Y_{\frac{LAC}{GLC}}, q_{mAb}, AAScaling, R_{AMM,2}\right\}$$

4.3.3 Dynamic metabolic flux analysis

The initial conditions to a dynamic metabolic flux analysis model are the cell seeding density, basal media concentration, feed media concentration, and feed media supplementation schedule. These initial conditions are fed to the kinetic equations described in Table 4.1 to calculate the uptake rates of nutrients, production rates of byproducts, mAb production rate, and growth rate. These reactions are listed from 42 to 67 in supplementary Table C1 and indexed by j. All the other reactions in the model are indexed by i. These rates ($v_{j,meas}$) computed from the kinetic equations are fed to the stoichiometric model described by the reaction network in Figure 4.1 and supplementary Table C1 to calculate all the rates via minimization of the following objective function.

$$Minimize \sum_i \left(\frac{v_i}{\sum_j v_{j,meas}}\right)^2 + \sum_j \left(\frac{v_j - v_{j,meas}}{v_{j,meas}}\right)^2 \tag{4.15}$$

$$Subject\ to\ the\ constraint:$$
$$S.v = 0$$
$$lb < v < ub$$

The nonlinear objective function described in Equation (4.15) has two components. The first is to minimize the sum of all the fluxes in the reaction network (except reactions 42 to 67). This term is normalized by the sum of all the kinetic expression predicted fluxes. Minimization of intracellular fluxes is based on the assumption that evolution with limited resources has placed pressure on living cells to operate with minimal effort [249], which has led to the development of parsimonious enzyme usage flux balance analysis (pFBA) that has widely been used to provide more realistic solutions to stoichiometric models [250]. This principle has been applied to model CHO cell metabolism in the literature. Minimizing the nutrient requirements after providing phenotypical constraints (constraining the biomass solution to the experimentally measured growth rate) has provided more accurate solutions to stoichiometric models [48]. Hence, we have decided to include this term to minimize the intracellular fluxes while providing phenotypical constraints via a second term that minimizes the difference between the kinetic equation predicted flux and the stoichiometric model predicted flux. Similar objective functions have been used in the literature to provide more realistic solutions to stoichiometric models [199].

The matrix of stoichiometric coefficients (S) was constructed from the reaction network. The lower bound (lb) and upper bound (ub) were constrained to be 0 to 100000 if the reaction is irreversible and -100000 to 100000 if the reaction is reversible. Reactions 42 to 67 represent the phenotypical uptake and secretion rates that are modeled via kinetic equations. Hence, these values should be constrained to the fluxes provided by the kinetic equations. Thus, the lower bound (lb) and upper bound (ub) were set to 50% to 150% of the kinetic equation predicted flux. If no feasible solution is found at these bounds, the bounds are increased by increments of 10% and the

stoichiometric model is solved till a feasible solution is found. The nonlinear optimization problem described in Equation (4.15) was solved by using IPOPT [251] implemented in Python by using pyomo [252] to determine all the intracellular fluxes (v).

The results from the stoichiometric model were assumed to be constant for 0.2 days and the new concentrations of metabolites in the media were calculated according to Equation (4.16) and (4.17).

$$[S]_{t+1} = [S]_t + v_{S,MFA} \times time\ step \times \left(\frac{[VCD]_t + [VCD]_{t+1}}{2}\right) \tag{4.16}$$

$$[VCD]_{t+1} = [VCD]_t e^{(\mu - \mu_d) \times time\ step} \tag{4.17}$$

The new concentration values were fed to the kinetic equations (Table 4.1) to determine the new fluxes at the next iteration point and the process was repeated. This process was repeated until harvest day or when viability dropped below 80%.

4.3.4 Model regression

The DMFA model was trained on experimental data from Chapter 2. The measurements included 864 data points and the difference in the model predicted value and the experimental value was used to determine the model parameters (θ) according to Equation 4.18.

$$SSR = \sum_{All\ expriments} \left(\frac{f(\theta, pH) - experimental\ value}{experimental\ value}\right)^2 \tag{4.18}$$

Equation 4.18 was minimized by using dual annealing (simulated annealing [253] with Nelder-Mead [254]) by using the Python library scipy [234]. The results of regression are plotted in Figure 4.2. The model parameters are tabulated in supplementary Table C2 and C3.

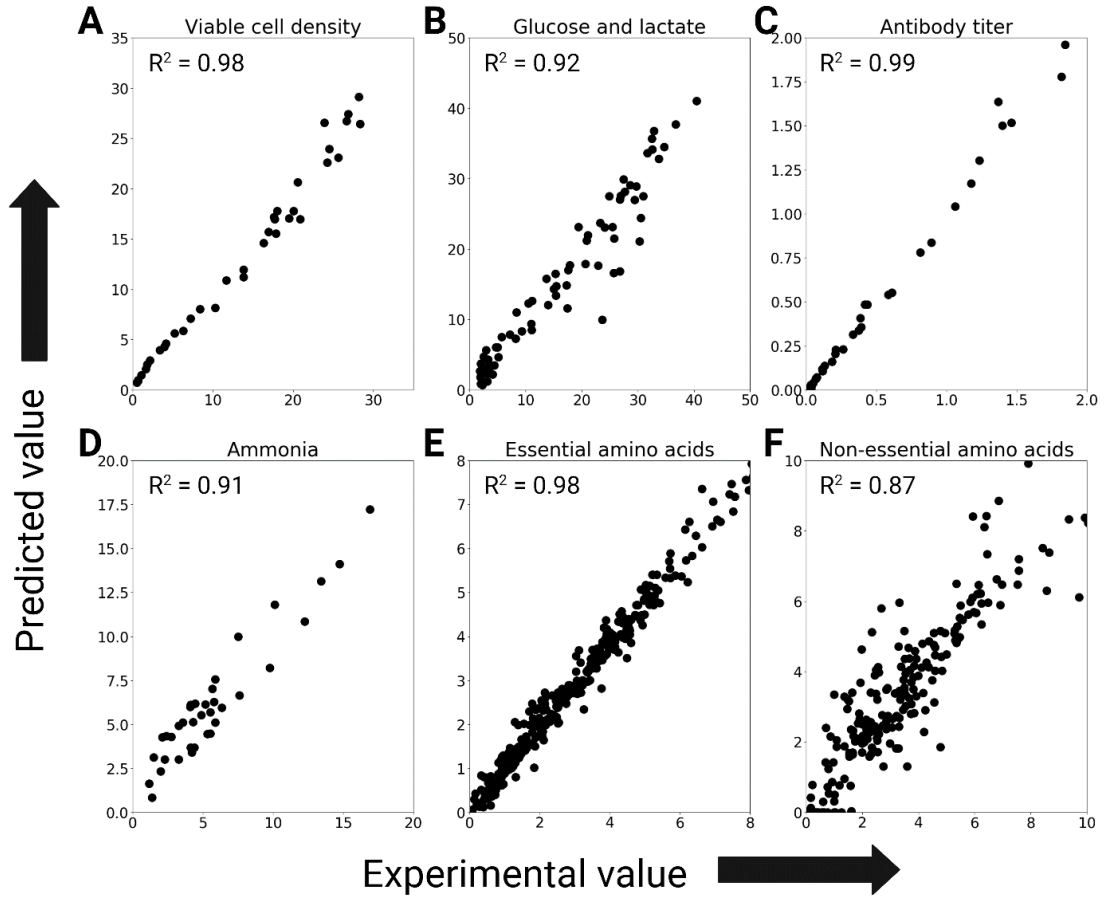

Figure 4.2: Scatter plot of experimental values vs the predicted values show that the model accurately captures variations in viable cell density, antibody titer, glucose concentrations, and essential amino acid concentrations. The model predictions can describe variations in concentrations of ammonia, lactate, and non-essential amino acids but with slightly reduced accuracy.

High R^2 values (greater than 0.95) were observed for regression related to viable cell density, mAb titers, and essential amino acid concentrations. Slightly lower values were observed for glucose, lactate, and ammonia concentrations due to modeling the complexities of lactate and ammonia metabolism. The lowest R^2 values were observed for non-essential metabolism, also due to a switch in metabolism of various non-

essential amino acids. Nevertheless, the R^2 values show that the model is capable of describing most of the variations in the experimental data. The spread of the experimental measurements also highlights a very diverse dataset.

4.4 Results and discussion

4.4.1 Bioreactor pH and nutrient depletion are used to predict growth rates

Equation (4.1) describes the growth rate (μ) as a function of concentrations of glucose, glutamine, glutamate, asparagine, and aspartate. The cells had the highest growth rate from day 0 to day 3 for all bioreactor conditions. This growth rate is fueled by rapid glutamine consumption. According to the protocol to grow the CHO VRC01 cells, it is recommended to add glutamine only at the start of the culture [205]. Depletion of glutamine (Figure 4.3D) is also successfully modeled; this depletion leads to reduction in growth rates of the cells. Equation (4.1) shows that the cell can still consume glutamate to produce glutamine for their protein requirements. Hence, growth rates do not drop to zero after glutamine deletion and grow at reduced rates until the depletion of asparagine (Figure 4.3E). The concentration of asparagine is also successfully modeled throughout the culture duration for all the pH values (Figure 4.3E). Bioreactors operated at pH 6.75 had reduced growth rates and nutrient uptake rates, which led to accumulation of nutrients and an increase in bioreactor osmolality. An increase in osmolality of CHO cell culture has shown to significantly reduce growth rates [206]. Hence, the osmolarity term in Equation (4.1) accounts for reduced growth rates. The final phase of the culture saw reduced growth rates (due to nutrient depletion or osmolality) and cell death. CHO cell fed-batch cultures eventually lead to cell death even if sufficient nutrients are supplied, because inefficiencies in cell metabolism can

lead to production of many inhibitory metabolites [248]. Hence, integral viable cell density was used in Equation (4.13) to model cell death (Figure 4.3B). The predictions of cell death seem overestimated but the experimental measurement of dead cell density using trypan blue does not account for lysed cells [255].

The mAb concentrations were successfully modeled (Figure 4.3C) using a single parameter (q_{mAb}) that is a quadratic function of bioreactor pH as shown in Equation (4.14).

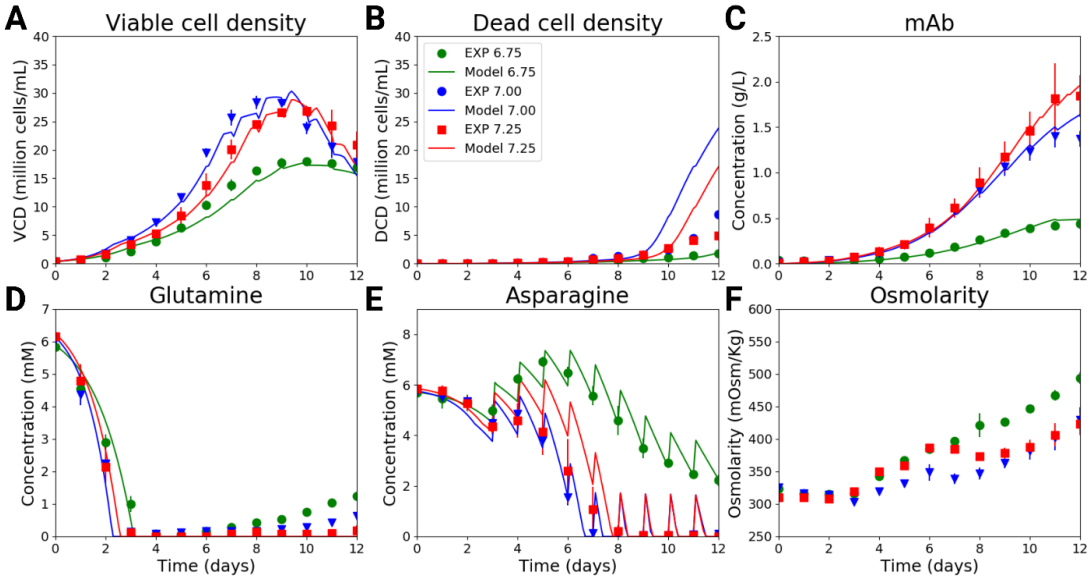

Figure 4.3: DMFA model can predict the effect of bioreactor pH, nutrient concentrations, and osmolality on growth rates and mAb concentrations. A) Viable cell density, B) Dead cell density, C) mAb concentrations, D) Glutamine concentrations, E) Asparagine concentrations, and F) Osmolality.

4.4.2 Bioreactor pH impacts glucose and lactate concentrations

Bioreactor pH significantly impacted glucose and lactate concentrations. The glucose uptake rate was modeled by using two parameters, representing the contribution of glucose to growth ($Y_{\frac{X}{GLC}}$) and the contribution of glucose to maintaining cellular metabolism (m_{GLC}). This formulation successfully predicts the increase in glucose uptake rate with increase in bioreactor pH (Figure 4.4A) throughout the culture duration.

Lactate production rates increased with increase in bioreactor pH, leading to accumulation of lactate in bioreactor pH 7.25 condition. Lactate also exhibited a shift in metabolism, as it was produced initially and consumed later on. The production of lactate was modeled as a function of glucose uptake rates (Equation (4.3)) with a yield coefficient $\left(Y_{\frac{LAC}{GLC}}\right)$ that is a function of bioreactor pH. The increase in glucose uptake rates and a higher value of the yield coefficient with increasing bioreactor pH led to successful prediction of lactate accumulation (Figure 4.4B). For this cell line, flux balance analysis has shown that glutaminolysis can lead to glycolysis flux being diverted to lactate [224]. Hence, the switch in metabolism of lactate was modeled as a function of glutamine concentration and according to Equation (4.3), depletion of glutamine leads to a switch in metabolism of lactate. This assumption led to successful predictions of lactate switch in pH 7.00 and pH 6.75 bioreactors but the prediction at pH 7.25 was not as accurate.

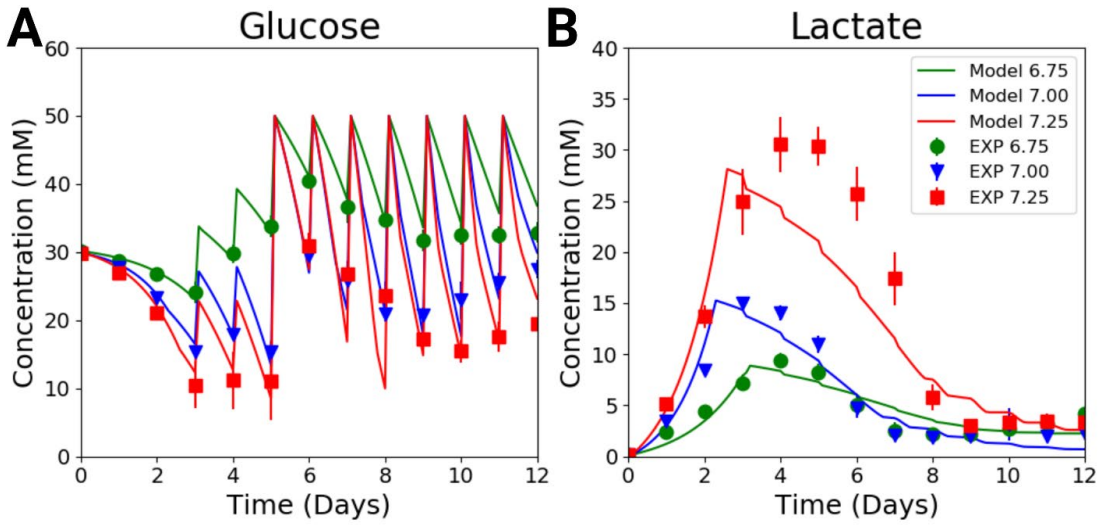

Figure 4.4: A) DMFA model can accurately predict the effect of bioreactor pH on glucose concentrations throughout the culture duration. B) The model can also predict the concentrations of lactate and shift in metabolism of lactate with slightly reduced accuracy.

4.4.3 Incorporation of bioreactor pH to model amino acid metabolism

Uptake rates of essential amino acids were modeled using yield coefficients ($Y_{\frac{X}{S}}$) described in Equation (4.5). Bioreactors operated at pH 7.25, however, exhibited slightly increased amino acid uptake rates. Hence, the yield coefficients ($Y_{\frac{X}{S}}$) were scaled accordingly as described in Equation (4.5). A total of eleven yield coefficients and three quadratic constants for the amino acid scaling factor ($AAScaling(pH)$) were used to model the concentrations of eleven amino acids (Figure 4.5). The increase in amino acid uptake rates at higher pH can lead to depletion of amino acids such as methionine (Figure 4.5A). This depletion has been successfully modeled without including a methionine term in the growth rate equation (Equation (4.1)), thus exhibiting the potential of dynamic metabolic flux analysis.

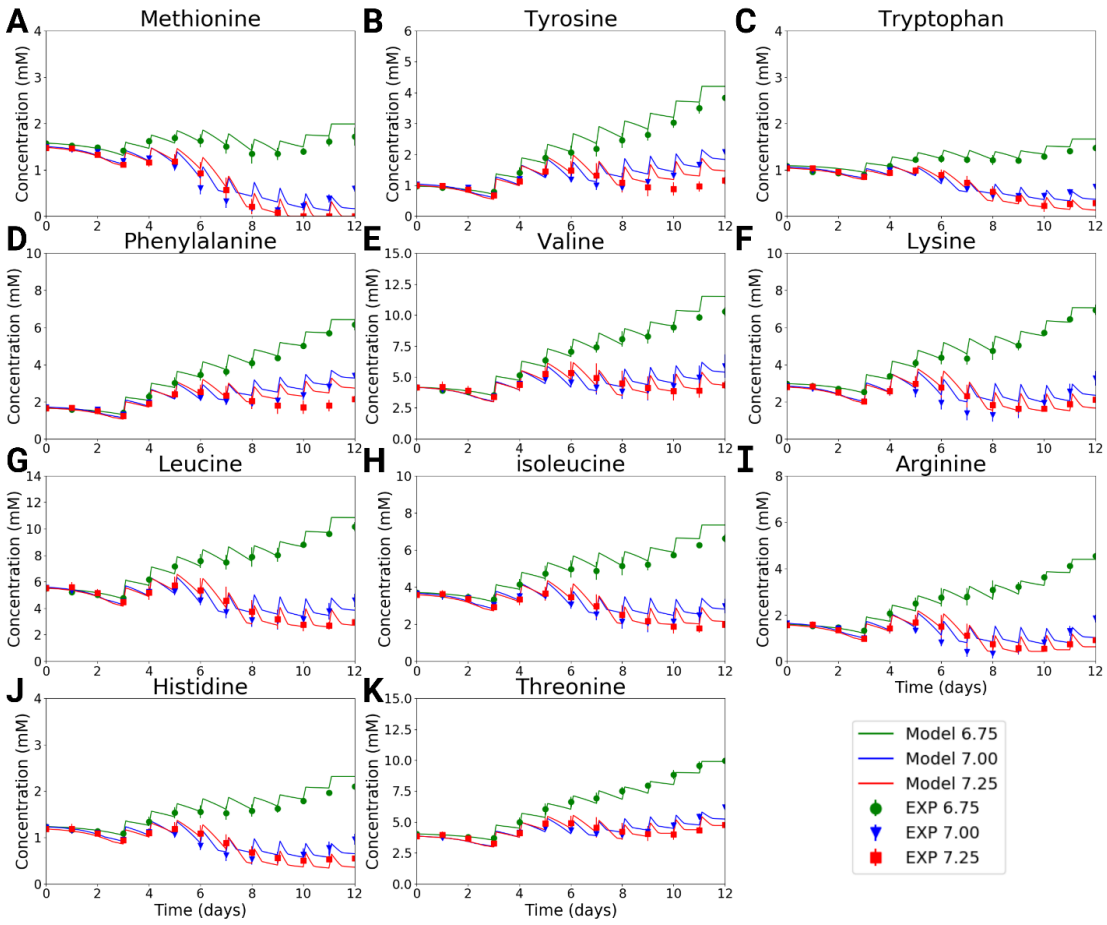

Figure 4.5: Essential amino acid concentrations can be successfully predicted for all pH conditions. Concentrations of A) Methionine, B) Tyrosine, C) Tryptophan, D) Phenylalanine, E) Valine, F) Lysine, G) Leucine, H) isoleucine, I) Arginine, J) Histidine, and K) Threonine.

Asparagine (Figure 4.3E) and glutamine (Figure 4.3D) exhibited only consumption and were not produced throughout the culture duration. Equation 4.6 and 4.7 successfully predict asparagine and glutamine concentrations and their effect on growth of the cells. Consumption of glutamine led to production of glutamate and aspartate in the early phase of the culture. Glutamate and aspartate were consumed after glutamine depletion, which has successfully been modeled (Figure 4.6A and 4.6B) by

Equation (4.8) and Equation (4.9). Alanine was also produced during the early phases of the culture. Since it is also a byproduct of glycolysis, excess pyruvate accumulation can also lead to alanine production during the early phases of the culture. Hence, its predictions are very similar to the lactate profiles and Equation (4.4) can successfully predict alanine (Figure 4.6C) concentrations. Prediction of glycine and serine were challenging for the model, and it appears that Equation (4.5) and (4.10) perform poorly at certain points of the culture, hence leading to reduced R^2 values (Figure 4.2F). Hence, caution should be exercised to utilize this model for prediction of serine and glycine concentrations. Glycine and serine metabolism seem to be regulated by something other than glutamine consumption and improved mechanistic understanding for the reason for the switch or the utilization of statistical modelling approaches to generate hybrid models might be useful.

Ammonia metabolism is also extremely complex as it can be produced from many sources. The effect of glutamine consumption on ammonia metabolism, the subsequent consumption of ammonia, and the switch again from consumption to production have been successfully modeled using Equation (4.11). However, the R^2 value of 0.91 (Figure 4.2D) tells us that ammonia predictions are not always accurate.

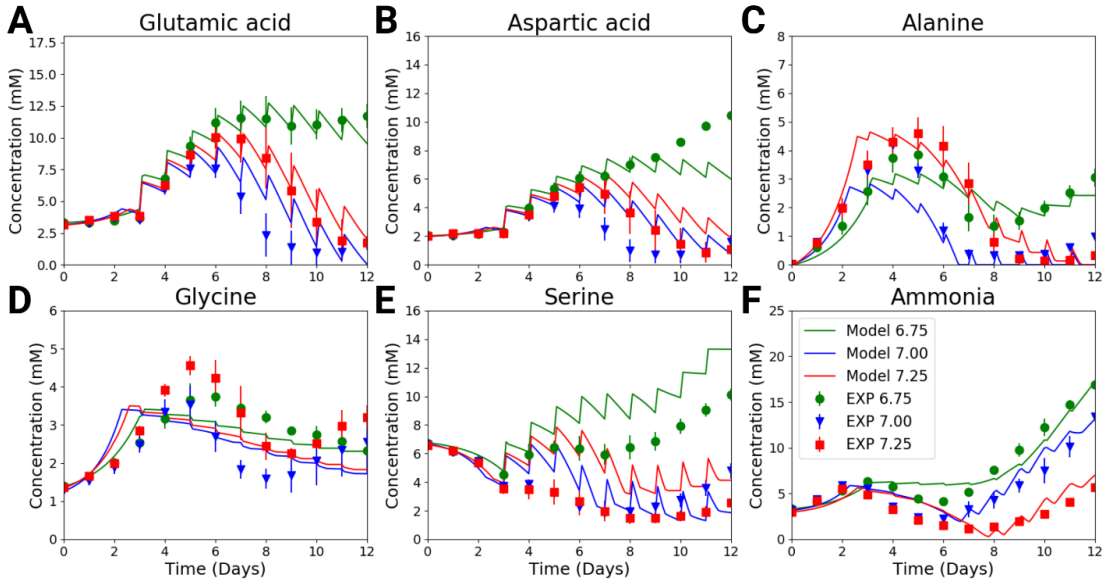

Figure 4.6: The model can successfully predict shift in metabolism in glutamic acid, aspartic acid, alanine, and ammonia. However, the predictions of concentrations of glycine and serine have low accuracy. Concentrations of A) Glutamic acid, B) Aspartic acid, C) Alanine, D) Glycine, E) Serine, and F) Ammonia.

4.4.4 Prediction of intensified fed-batch cultures

Section 4.4.1 to 4.4.3 highlighted the successful predictions of cell densities, concentrations of glucose, lactate, mAb, amino acids, and ammonia. However, the model predictions still need to be validated. Model validation requires experiments that were not used to train the model. The inputs to the model are bioreactor pH, seeding cell density, basal media concentrations, feed media concentrations, amount of feed media added, and feed media time. Hence, we decided to perform an experiment at a different bioreactor pH (7.12), seeding cell density (5 million cells/mL), and a new feed supplementation plan (discussed in Section 4.2.1.3) that led to increased nutrient feeding.

Utilizing N-1 perfusion to seed the bioreactor at higher cell density has led to the development of intensified fed-batch platforms of mAb production [245]. In addition to model validation, we wanted to demonstrate that our model is capable of predicting intensified fed-batch culture performance after training on fed-batch data and this can lead to demonstrating model applications such as derisking of intensified fed-batch process development.

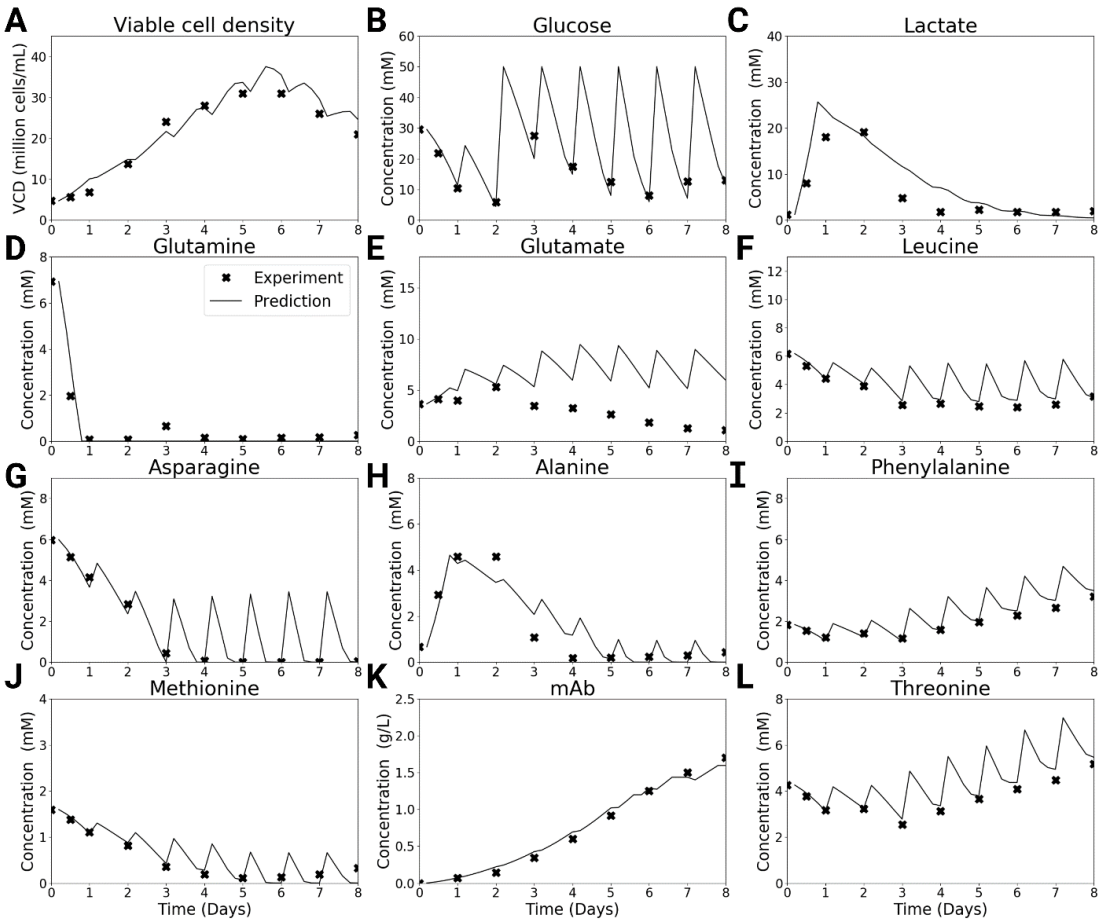

Figure 4.7: The model successfully predicts intensified fed-batch culture performance for concentrations of A) Viable cell density, B) Glucose, C) Lactate, D) Glutamine, E) Glutamate, F) Leucine, G) Asparagine, H) Alanine, I) Phenylalanine, J) Methionine, K) Monoclonal antibody, and L) Threonine.

Concentrations of metabolites that impact growth rates (glutamine and asparagine) were successfully modeled in intensified fed-batch cultures (Figure 4.7D and 7G). This led to successful prediction of cell densities at the early cell growth phase (driven by glutamine), late cell growth phase (driven by asparagine), and cell death phase in the intensified fed-batch experiment at pH 7.12 (Figure 4.7A). Bioreactor pH significantly impacted glucose uptake rates (Figure 4.4A). The predictive abilities of the model are demonstrated via successful glucose predictions (Figure 4.7B) throughout the culture duration in the intensified fed-batch process. Successful predictions of glutamine concentrations (Figure 4.7D) led to successful prediction of early shift in metabolism of lactate (Figure 4.7C), alanine (Figure 4.7H) and glutamate (Figure 4.6E). The increase in lactate accumulation at the pH and subsequent depletion was also successfully modeled. The concentration of alanine was also successfully modeled (Figure 4.7H) throughout the culture duration.

Glutamate uptake rates could not be successfully modeled after day 3 in this pH condition (Figure 4.7E). The R^2 square value for essential amino acid prediction is very high, hence it is not surprising that essential amino acid concentrations can be successfully modeled (Figure 4.7F, 4.7I, 7J and 7L). This highlights the capability of the model to derisk the process prior to experimentation. The experimentally validated model predictions show that phenylalanine (Figure 4.7I) and threonine (Figure 4.7L) are accumulating in the culture. The experimentally validated model predictions also show that methionine is depleted in the culture (Figure 4.7J), hence indicating the need to rebalance the media formulation to provide appropriate nutrient levels to prevent depletion or accumulation. For example, experimentally validated model predictions show that leucine is being supplied at the perfect level to maintain a constant

concentration in the culture (Figure 4.7F). The predictions of other amino acids and ammonia are shown in supplementary Figure C2. The model could successfully predict ammonia concentrations in the early phase of the culture but performed poorly in the later stages (supplementary Figure C2).

The most important experimentally validated model prediction is the mAb concentrations (Figure 4.7K). This shows that the intensified fed-batch cultures can achieve similar mAb titers (on day 8) while compared to the fed-batch culture (on day 12), hence reducing the time to operate large scale bioreactors and showing that the model can successfully be used to predict intensified fed-batch culture performance and can help derisk intensified fed-batch process development.

4.4.5 Prediction of perfusion bioreactor cultures

Perfusion bioreactor process development is also challenging and requires expensive experiments. Hence, we want to utilize the mathematical model (trained only on fed-batch data) to show that successful predictions of perfusion cultures can demonstrate the DMFA model's utility in perfusion bioreactor process development.

Figure 4.8D shows that the model accurately predicts glutamine depletion in the cultures. Subsequent growth is driven by glutamate consumption that is also accurately predicted (Figure 4.8E). Discrete cell bleed was used in the perfusion experiment described in Section 4.2.1.4. Hence, the predictions of viable cell density drop whenever the cell bleed occurs. The viable cell density in perfusion cultures with a target viable cell density of 20 million cells/mL has successfully been predicted (Figure 4.8A). The model also successfully predicts switch in metabolism of lactate (Figure 4.8C) and alanine (Figure 4.8H). The steady state concentrations of glucose and all the other amino acids have also been successfully predicted by the DMFA model. The mAb

concentrations (Figure 4.8K) have also been successfully predicted, hence demonstrating the ability of the model to aid perfusion process development by using fed-batch data.

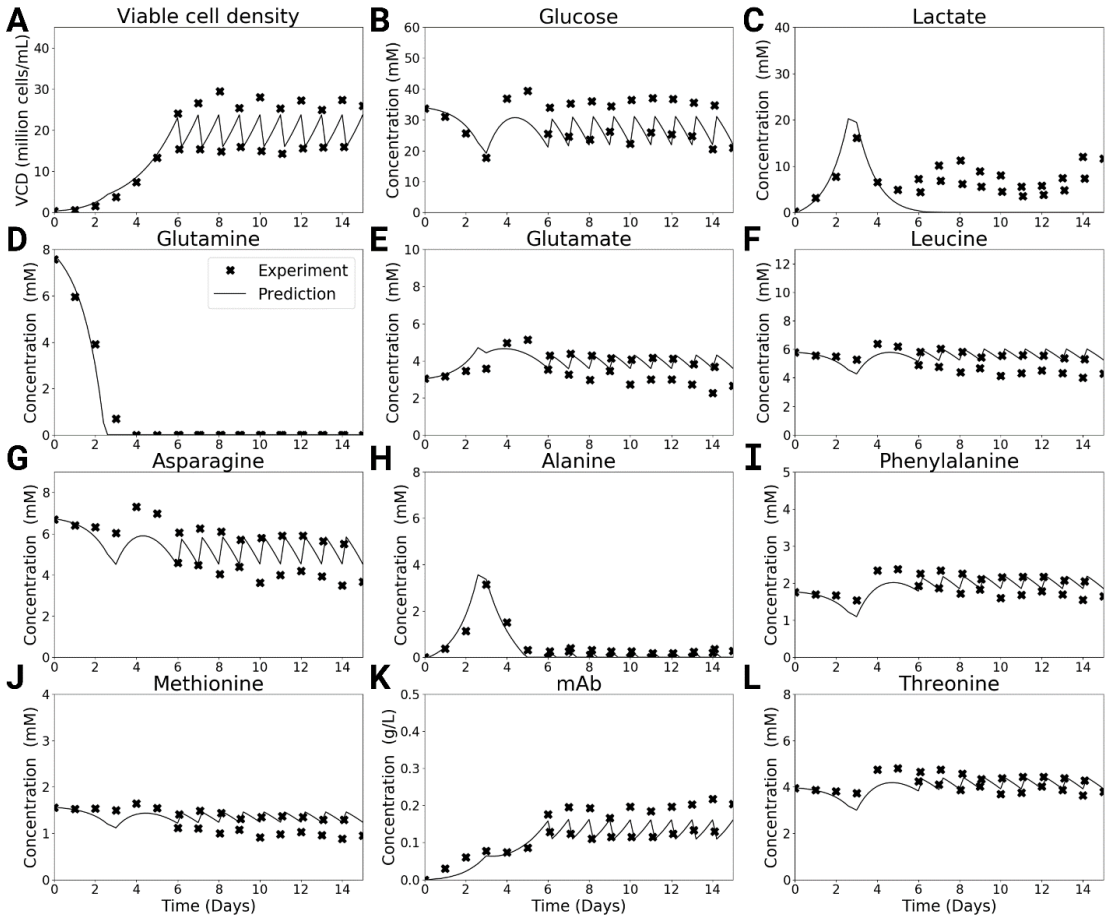

Figure 4.8: The model successfully predicts perfusion culture performance for concentrations of A) Viable cell density, B) Glucose, C) Lactate, D) Glutamine, E) Glutamate, F) Leucine, G) Asparagine, H) Alanine, I) Phenylalanine, J) Methionine, K) Monoclonal antibody, and L) Threonine.

4.4.6 Derisking cell culture media development

The previous sections have validated the model predictions under different bioreactor pH conditions and nutrient compositions. The model predictions have demonstrated that the model can successfully predict bioreactor performance in different bioreactor operating modes.

Another application of the mathematical model is to aid with development of cell culture media. Commercial media formulations are seldom reported in the literature. However, the publication of the composition of the chemically defined AMBIC media has led to a more open source approach to compare bioprocesses. The CHO VRC01 has been cultivated in the AMBIC media in the literature in a fed-batch bioreactor. This experiment involved seeding CHO VRC01 cells at 0.5 million cells/mL in the AMBIC basal media along with daily supplementation of AMBIC feed media. The measurements of cell density, glucose concentration, lactate concentration, glutamine concentration, glutamate concentration and ammonia concentration have also been taken from the literature [205]. We utilized this information along with the DMFA model to predict and validate the AMBIC media performance.

The model successfully predicts the peak viable cell density that can be achieved by this media (Figure 4.9A). The reason for reduction of cell growth can be attributed to depletion of asparagine (Figure 4.9G), valine (Figure 4.9K), and phenylalanine (Figure 4.9J). The model predictions of concentrations of glutamine, glutamate, ammonia, and mAb were also successful, highlighting that the model is capable of predicting new media formulation performances, prior to running experiments. Due to ease of measurement, cell density, glutamate, ammonia, mAb, glucose and lactate concentrations are commonly measured in cell culture experiments. Measurement of other amino acids adds to costs and they are not often measured. According to the

literature data, for the CHO VRC01 cell line, a peak viable cell density of 30 million cells was achieved using the ActiPro media and Hyclone feed media, but the prek viable cell density was less than 10 million cells/mL with the AMBIC basal and feed media [205]. The successful predictions of viable cell density (Figure 4.9A) according to the model are correlated with amino acids that are not typically measured, hence indicating that if a robust mathematical model has been developed for a cell line, it can be used to provide potential explanations for poor bioprocess performance.

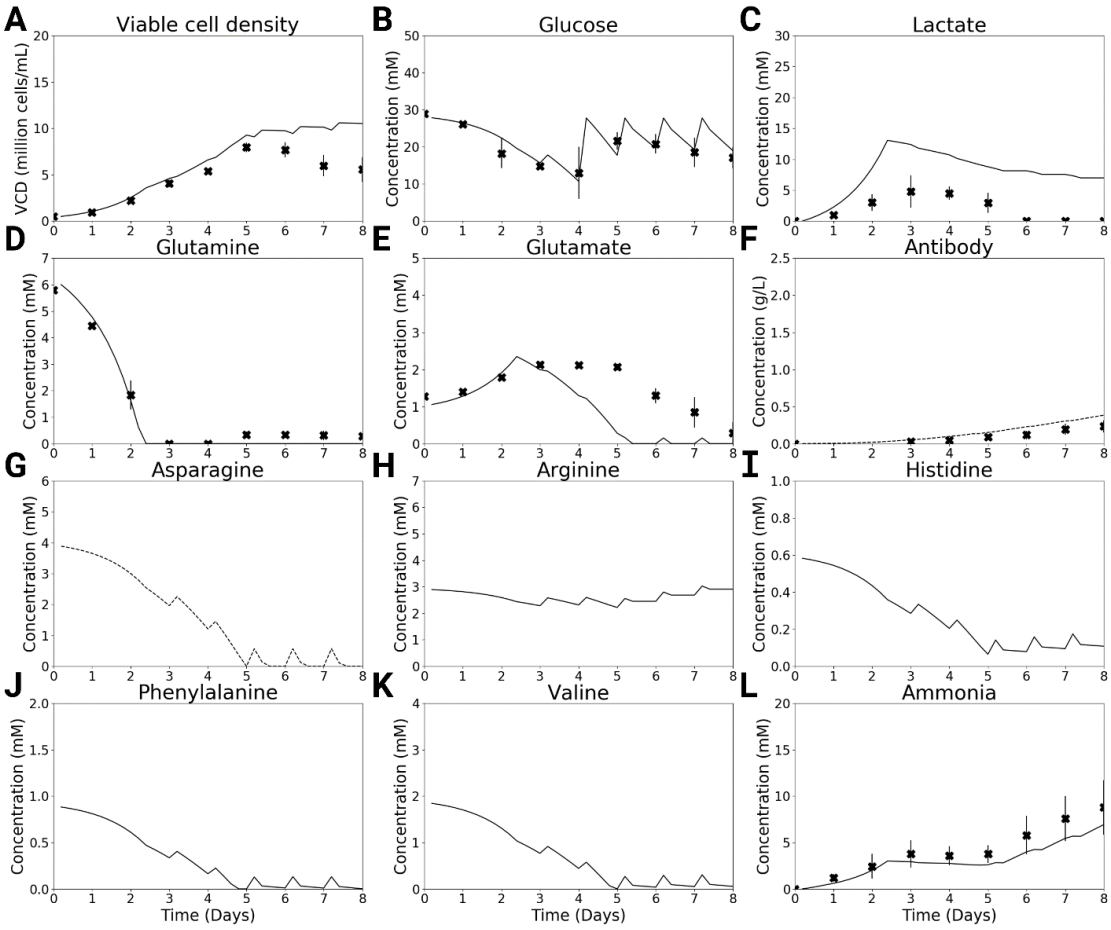

Figure 4.9: Predictions of CHO VRC01 fed-batch cultures while using the AMBIC basal and feed media along with experimental validation with literature experimental data. A) Viable cell density, B) Glucose, C) Lactate, D) Glutamine, E) Glutamate, F) Monoclonal antibody, G) Asparagine, H) Arginine, I) Histidine, J) Phenylalanine, K) Valine, and L) Ammonia.

4.5 Conclusions

The lack of detailed models of CHO cell metabolism that include the effect of bioreactor pH has been addressed in this work. Although dynamic metabolic flux analysis models exist in the literature [62], they do not include the effect of bioreactor pH and also only model a subset of amino acids. Hence, we developed a dynamic

metabolic flux analysis model to predict concentrations of cell density, glucose, lactate, amino acids, ammonia, and mAb as a function of bioreactor pH, seeding cell density, and nutrient composition. This model was trained on experimental data from fed-batch experiments run at three different bioreactor pH conditions. The model was subsequently used to predict intensified fed-batch culture and perfusion culture performance. This was followed by successful experimental validation, thus demonstrating the model's predictive capabilities. This also shows the potential of utilizing this model for process development of intensified fed-batch and perfusion experiments after training the model on fed-batch data.

Flowsheet models in the literature utilize mass balances to study the economic feasibility of fed-batch, intensified fed-batch, and perfusion bioreactors. The application of one model to predict different bioreactor operator modes can help provide more realistic constraints to flow sheet models [222]. An application of mathematical models is to use the model as a digital twin by predicting unmeasured metabolites after providing concentrations of a subset of measured metabolites. Additional work has to be done to demonstrate this application.

Chapter 5

CONCLUSIONS AND FUTURE WORK

5.1 Summary and conclusions

In this dissertation, various mathematical modeling approaches were developed, experimentally validated, and utilized for upstream process development. Majority of mathematical models in the literature utilize data generated from cell cultures in Erlenmeyer flasks. Thus, resulting in a data set from experiments where pH was not controlled. This poses a challenge to predict bioreactor culture performance in pH-controlled environments that are more representative of large-scale bioreactors.

To address this issue and generate a more realistic dataset, Chapter 2 involves cultivating CHO cells in pH-controlled bioreactors. Thus, the effect of bioreactor pH on CHO cell metabolism was quantified. Measurements of concentrations of glucose, lactate, ammonia, amino acids, mAb, and cell density showed that bioreactor pH significantly affects cell growth rates, glucose-lactate metabolism, ammonia accumulation, cell specific mAb production rates, and amino acid uptake rates. Various metabolites (glutamate, aspartate, lactate, alanine, glycine, and ammonia) exhibited shift in metabolism from secretion to uptake. The measured uptake and secretion rates of all the measured metabolites were used to constrain a stoichiometric model and flux balance analysis was utilized to understand the effect of bioreactor pH on various metabolites and explain the reasons for shift in metabolism of various metabolites.

The mAb used in this study exhibits Fab and Fc glycosylation. The effect of process parameters (pH, temperature, or dissolved oxygen) on site-specific N-linked

glycosylation of any glycoprotein has not been reported in the literature. However, the literature has shown that Fab and Fc glycosylation can impact product quality [108]. To improve our understanding of the impact of process parameter on site-specific N-linked glycosylation, we utilized glycopeptide mapping to determine the effect of bioreactor pH on site-specific N-linked glycosylation. Peptide mapping revealed that bioreactor pH affects fucosylation and sialylation in the Fab region and galactosylation in the Fc region. This data demonstrates that bioreactor pH can impact a glycosylation index in one site without impacting the same glycosylation index the other site.

Multiple mathematical models for glycosylation have been published in the literature [196]. However, none of them have modeled the site-specific glycosylation heterogeneities that can fundamentally arise from site-specific N-linked glycosylation processing in the Golgi. To understand the complexities of site-specific N-linked glycosylation in the Fab and Fc region that were measured Chapter 2, we developed a mathematical model for site-specific N-linked glycosylation in Chapter 3. A sensitivity analysis was performed to determine the relationship between the model independent variables (cell specific mAb production rate, galactosyltransferase enzyme concentration, sialyltransferases enzyme concentration, and fucosyltransferase enzyme concentration) and the model outputs (glycosylation indices in the Fab and Fc regions), thus enhancing model interpretability. This model was used to predict glycosylation enzyme activities at various experimental pH conditions. The model predicted enzyme activities were used to understand the effect of bioreactor pH on Fab and Fc glycan fractions. The model predicted enzyme activities were experimentally validated by measurements of glycosyltransferase RNA levels and intracellular nucleotide sugar concentrations. The site-specific glycosylation model was further tested by predicting

the effect of increasing galactosyltransferase enzyme activity on Fab and Fc glycan fractions. These predictions were successfully validated by performing galactose and $MnCl_2$ supplementation experiments, thus successfully developing and testing a site-specific N-linked glycosylation model.

The experiments performed in Chapter 2 show that bioreactor pH can significantly affect CHO cell metabolism, thus impacting glycoprotein titers. However, there is very limited literature on incorporating pH in mathematical models for CHO cell metabolism. In Chapter 4, we have developed a dynamic metabolic flux analysis model that is capable of predicting bioreactor performance (cell growth rates and antibody titers), if provided with the bioreactor pH, nutrient composition of the basal media, composition of feed media, feed media supplementation time, and initial cell density. The DMFA model utilizes kinetic equations consisting of experimentally regressed parameters, a subset of which are dependent on bioreactor pH via quadratic equations to provide fluxes of uptake/secretion of metabolites as a function of extracellular concentrations of metabolites. A reaction network that describes core reactions of CHO cell metabolism is constrained by the fluxes from the kinetic expression and the fluxes of all the reactions are analyzed by metabolic flux analysis. The solution of metabolic flux analysis is assumed to be constant over a short period of time and the new extracellular metabolite concentrations are calculated. These new concentrations are fed to the kinetic equations to iterate the modeling process until the harvest day of the culture, thus providing dynamic predictions of viable cell density, metabolite concentrations, and antibody titers. The capabilities of the developed model in biopharmaceutical process development have been demonstrated by making predictions of intensified fed-batch cultures, perfusion cultures, and cultures with

different media. Each of these predictions have been experimentally validated. Thus, demonstrating that we have developed a robust model.

5.2 Perspectives and recommendations

Ideally, we would like to utilize mathematical models to optimize cell culture processes to maximize the production of therapeutics with good quality attributes. The DMFA model developed in Chapter 4 can be used to determine the optimal operating conditions to maximize the titers. Including nucleotide sugar metabolism into the DMFA model and coupling the DMFA model with the site-specific glycosylation model developed in Chapter 3 can be used to maximize the titer while constraining the product quality (glycosylation fractions). Combining DMFA models with site-specific glycosylation models is not trivial, mostly due to poor understanding of the relationship between the intracellular nucleotide sugar concentration and the nucleotide sugar transporter (cytosol to Golgi) enzymes. Chapter 3 has shown that the intracellular nucleotide sugar concentrations and transporter enzyme concentrations can vary drastically in different cultures. Hence, experiments involving upregulating and downregulating the transporter enzyme expression are critical to help understand the relationship between the nucleotide sugar concentrations and enzyme concentrations. This functionality can be included to bridge the DMFA model to the N-linked glycosylation model.

Chapter 4 involved developing a complex DMFA model with 52 model parameters. Although the model predictions were successfully validated, increasing the size and quality of training data set to accurately estimate the model parameters is crucial to utilize the model for process optimization. Thus, model-based design of experiments approaches must be applied to accurately determine crucial model

parameters. Only a subset of DMFA model capabilities have been explored in Chapter 4. The model should subsequently be utilized as a digital twin [256] to predict concentrations of unmeasured metabolites while utilizing online or at-line measurements of a limited set of metabolites. Certain amino acids are known to have poor solubility in cell culture media [257], hence leading to changes in critical material attributes (media amino acid concentrations) that can impact the bioreactor performance. The model can be used to predict and derisk the impact of critical material attributes on the bioreactor performance.

BIBLIOGRAPHY

1. Shepard, H.M., et al., Developments in therapy with monoclonal antibodies and related proteins. Clinical Medicine, 2017. **17**(3): p. 220-232.

2. Mullard, A., *FDA approves 100th monoclonal antibody product.* Nature Reviews Drug Discovery, 2021. **20**(7): p. 491-495.

3. InfogenceGlobalResearch, *Global Biopharmaceutical Market (2021–2026) by Product Type, Therapeutic Application, Geography, Competitive Analysis and the Impact of Covid-19 with Ansoff Analysis, ID: 5397295*. 2021.

4. Howard, D.H., et al., *Pricing in the Market for Anticancer Drugs.* Journal of Economic Perspectives, 2015. **29**(1): p. 139-162.

5. Verma, V., et al., *A systematic review of the cost and cost-effectiveness studies of immune checkpoint inhibitors.* J Immunother Cancer, 2018. **6**(1): p. 128.

6. Pierpont, T.M., C.B. Limper, and K.L. Richards, *Past, Present, and Future of Rituximab-The World's First Oncology Monoclonal Antibody Therapy.* Front Oncol, 2018. **8**: p. 163.

7. Bielser, J.-M., et al., *Perfusion mammalian cell culture for recombinant protein manufacturing – A critical review.* Biotechnology Advances, 2018. **36**(4): p. 1328-1340.

8. Rathore, A.S. and H. Winkle, *Quality by design for biopharmaceuticals.* Nature Biotechnology, 2009. **27**(1): p. 26-34.

9. Chakrabarty, A., G.T. Buzzard, and A.E. Rundell, *Model-based design of experiments for cellular processes.* Wiley Interdiscip Rev Syst Biol Med, 2013. **5**(2): p. 181-203.

10. Sommeregger, W., et al., *Quality by control: Towards model predictive control of mammalian cell culture bioprocesses.* Biotechnol J, 2017. **12**(7).

11. Dhara, V.G., et al., *Recombinant Antibody Production in CHO and NS0 Cells: Differences and Similarities.* BioDrugs, 2018. **32**(6): p. 571-584.

12. Trummer, E., et al., *Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors.* Biotechnol Bioeng, 2006. **94**(6): p. 1033-44.

13. Tziampazis, E. and A. Sambanis, *Modeling of cell culture processes.* Cytotechnology, 1994. **14**(3): p. 191-204.

14. Ben Yahia, B., L. Malphettes, and E. Heinzle, *Macroscopic modeling of mammalian cell growth and metabolism.* Applied Microbiology and Biotechnology, 2015. **99**(17): p. 7009-7024.

15. Kyriakopoulos, S., et al., *Kinetic Modeling of Mammalian Cell Culture Bioprocessing: The Quest to Advance Biomanufacturing.* Biotechnology Journal, 2018. **13**(3): p. 1700229.

16. Shuler, M.L. and F. Kargi, *Bioprocess engineering : basic concepts*. 2nd ed. Prentice-Hall international series in the physical and chemical engineering sciences. 2002, Upper Saddle River, NJ: Prentice Hall.

17. Liebermeister, W. and E. Klipp, *Bringing metabolic networks to life: convenience rate law and thermodynamic constraints.* Theoretical Biology and Medical Modelling, 2006. **3**(1).

18. Jedrzejewski, P., et al., *Towards Controlling the Glycoform: A Model Framework Linking Extracellular Metabolites to Antibody Glycosylation.* International Journal of Molecular Sciences, 2014. **15**(3): p. 4492-4522.

19. López-Meza, J., et al., *Using simple models to describe the kinetics of growth, glucose consumption, and monoclonal antibody formation in naive and infliximab producer CHO cells.* Cytotechnology, 2016. **68**(4): p. 1287-1300.

20. Kiparissides, A., E.N. Pistikopoulos, and A. Mantalaris, *On the model-based optimization of secreting mammalian cell (GS-NS0) cultures.* Biotechnology and Bioengineering, 2015. **112**(3): p. 536-548.

21. Robitaille, J., J. Chen, and M. Jolicoeur, A Single Dynamic Metabolic Model Can Describe mAb Producing CHO Cell Batch and Fed-Batch Cultures on Different Culture Media. PLOS ONE, 2015. **10**(9): p. e0136815.

22. Naderi, S., et al., *Development of a mathematical model for evaluating the dynamics of normal and apoptotic Chinese hamster ovary cells.* Biotechnology Progress, 2011. **27**(5): p. 1197-1205.

23. Sha, S. and S. Yoon, *An investigation of nucleotide sugar dynamics under the galactose supplementation in CHO cell culture.* Process Biochemistry, 2019. **81**: p. 165-174.

24. Saa, P.A. and L.K. Nielsen, Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. Biotechnology Advances, 2017. **35**(8): p. 981-1003.

25. Kanehisa, M., *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.

26. Karp, P.D., et al., *The BioCyc collection of microbial genomes and metabolic pathways.* Briefings in Bioinformatics, 2019. **20**(4): p. 1085-1093.

27. Zupke, C. and G. Stephanopoulos, *Intracellular flux analysis in hybridomas using mass balances and in vitro13C nmr.* Biotechnology and Bioengineering, 1995. **45**(4): p. 292-303.

28. Stephanopoulos, G., A.A. Aristidou, and J.H. Nielsen, *Metabolic engineering : principles and methodologies*. 1998, San Diego: Academic Press.

29. García Sánchez, C.E. and R.G. Torres Sáez, *Comparison and analysis of objective functions in flux balance analysis.* Biotechnology Progress, 2014. **30**(5): p. 985-991.

30. Sha, S., et al., Mechanistic modeling and applications for CHO cell culture development and production. Current Opinion in Chemical Engineering, 2018. **22**: p. 54-61.

31. Ahn, W.S. and M.R. Antoniewicz, *Towards dynamic metabolic flux analysis in CHO cell cultures.* Biotechnol J, 2012. **7**(1): p. 61-74.

32. Deshpande, R., T.H. Yang, and E. Heinzle, *Towards a metabolic and isotopic steady state in CHO batch cultures for reliable isotope-based metabolic profiling.* Biotechnol J, 2009. **4**(2): p. 247-63.

33. Goudar, C., et al., Metabolic flux analysis of CHO cells in perfusion culture by metabolite balancing and 2D [13C, 1H] COSY NMR spectroscopy. Metabolic Engineering, 2010. **12**(2): p. 138-149.

34. Antoniewicz, M.R., *A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications.* Metabolic Engineering, 2021. **63**: p. 2-12.

35. Ahn, W.S. and M.R. Antoniewicz, *Parallel labeling experiments with [1,2-13C]glucose and [U-13C]glutamine provide new insights into CHO cell metabolism.* Metabolic Engineering, 2013. **15**: p. 34-47.

36. Kirsch, B.J., et al., *Metabolic analysis of the asparagine and glutamine dynamics in an industrial Chinese hamster ovary fed-batch process.* Biotechnology and Bioengineering, 2022. **119**(3): p. 807-819.

37. Nicolae, A., et al., *Non-stationary 13C metabolic flux analysis of Chinese hamster ovary cells in batch culture using extracellular labeling highlights metabolic reversibility and compartmentation.* BMC Systems Biology, 2014. **8**(1): p. 50.

38. Sacco, S.A. and J.D. Young, *13C metabolic flux analysis in cell line and bioprocess development.* Current Opinion in Chemical Engineering, 2021. **34**: p. 100718.

39. Shupletsov, M.S., et al., *OpenFLUX2: 13C-MFA modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments.* Microbial Cell Factories, 2014. **13**(1).

40. Yoo, H., et al., *Quantifying Reductive Carboxylation Flux of Glutamine to Lipid in a Brown Adipocyte Cell Line.* Journal of Biological Chemistry, 2008. **283**(30): p. 20621-20627.

41. Young, J.D., *INCA: a computational platform for isotopically non-stationary metabolic flux analysis.* Bioinformatics, 2014. **30**(9): p. 1333-1335.

42. Hefzi, H., et al., *A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism.* Cell Systems, 2016. **3**(5): p. 434-443.e8.

43. Széliová, D., et al., *What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines.* Metabolic Engineering, 2020. **61**: p. 288-300.

44. Széliová, D., et al., *Inclusion of maintenance energy improves the intracellular flux predictions of CHO.* PLoS Comput Biol, 2021. **17**(6): p. e1009022.

45. Széliová, D., et al., *Error propagation in constraint-based modeling of Chinese hamster ovary cells.* Biotechnology Journal, 2021. **16**(4): p. 2000320.

46. Schinn, S.M., et al., *A genome-scale metabolic network model and machine learning predict amino acid concentrations in Chinese Hamster Ovary cell cultures.* Biotechnol Bioeng, 2021. **118**(5): p. 2118-2123.

47. Schinn, S.M., et al., Systematic evaluation of parameters for genome-scale metabolic models of cultured mammalian cells. Metab Eng, 2021. **66**: p. 21-30.

48. Chen, Y., et al., An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells. npj Systems Biology and Applications, 2019. **5**(1).

49. Chen, Y., et al., A genome-scale nutrient minimization forecast algorithm for controlling essential amino acid levels in CHO cell cultures. Biotechnology and Bioengineering, 2022. **119**(2): p. 435-451.

50. Yeo, H.C., et al., *Enzyme capacity-based genome scale modelling of CHO cells.* Metabolic Engineering, 2020. **60**: p. 138-147.

51. Sou, S.N., et al., *How does mild hypothermia affect monoclonal antibody glycosylation?* Biotechnol Bioeng, 2015. **112**(6): p. 1165-76.

52. Saha, R., A. Chowdhury, and C.D. Maranas, *Recent advances in the reconstruction of metabolic models and integration of omics data.* Current Opinion in Biotechnology, 2014. **29**: p. 39-45.

53. Satish Kumar, V., M.S. Dasika, and C.D. Maranas, *Optimization based automated curation of metabolic reconstructions.* BMC Bioinformatics, 2007. **8**(1): p. 212.

54. Hosseini, Z. and S.-A. Marashi, *Discovering missing reactions of metabolic networks by using gene co-expression data.* Scientific Reports, 2017. **7**(1): p. 41774.

55. King, Z.A., et al., *BiGG Models: A platform for integrating, standardizing and sharing genome-scale models.* Nucleic Acids Research, 2016. **44**(D1): p. D515-D522.

56. Wishart, D.S., et al., *HMDB 4.0: the human metabolome database for 2018.* Nucleic Acids Research, 2018. **46**(D1): p. D608-D617.

57. Fouladiha, H., et al., Systematically gap-filling the genome-scale metabolic model of CHO cells. Biotechnol Lett, 2021. **43**(1): p. 73-87.

58. Ataman, M. and V. Hatzimanikatis, lumpGEM: Systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites. PLOS Computational Biology, 2017. **13**(7): p. e1005513.

59. Ataman, M., et al., redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. PLOS Computational Biology, 2017. **13**(7): p. e1005444.

60. Erdrich, P., R. Steuer, and S. Klamt, An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. BMC Systems Biology, 2015. **9**(1).

61. Provost, A., G. Bastin, and Y.-J. Schneider, *FROM METABOLIC NETWORKS TO MINIMAL DYNAMIC BIOREACTION MODELS.* IFAC Proceedings Volumes, 2007. **40**(4): p. 1-6.

62. Nolan, R.P. and K. Lee, *Dynamic model of CHO cell metabolism.* Metabolic Engineering, 2011. **13**(1): p. 108-124.

63. Quek, L.E., et al., *Metabolic flux analysis in mammalian cell culture.* Metab Eng, 2010. **12**(2): p. 161-71.

64. Sheikholeslami, Z., M. Jolicoeur, and O. Henry, Probing the metabolism of an inducible mammalian expression system using extracellular isotopomer analysis. J Biotechnol, 2013. **164**(4): p. 469-78.

65. Zamorano, F., et al., Dynamic metabolic models of CHO cell cultures through minimal sets of elementary flux modes. J Biotechnol, 2013. **164**(3): p. 409-22.

66. Kastelic, M., et al., Dynamic metabolic network modeling of mammalian Chinese hamster ovary (CHO) cell cultures with continuous phase kinetics transitions. Biochemical Engineering Journal, 2019. **142**: p. 124-134.

67. Antonakoudis, A., et al., Synergising stoichiometric modelling with artificial neural networks to predict antibody glycosylation patterns in Chinese hamster ovary cells. Computers & Chemical Engineering, 2021. **154**: p. 107471.

68. Carinhas, N., et al., (13) C-metabolic flux analysis of human adenovirus infection: Implications for viral vector production. Biotechnol Bioeng, 2017. **114**(1): p. 195-207.

69. Martínez-Monge, I., et al., *Metabolic flux balance analysis during lactate and glucose concomitant consumption in HEK293 cell cultures.* Biotechnol Bioeng, 2019. **116**(2): p. 388-404.

70. Niklas, J., et al., *Metabolism and metabolic burden by α1-antitrypsin production in human AGE1.HN cells.* Metab Eng, 2013. **16**: p. 103-14.

71. Hagrot, E., et al., *Novel column generation-based optimization approach for poly-pathway kinetic model applied to CHO cell culture.* Metabolic Engineering Communications, 2019. **8**: p. e00083.

72. Hagrot, E., et al., *Poly-pathway model, a novel approach to simulate multiple metabolic states by reaction network-based model – Application to amino acid depletion in CHO cell culture.* Journal of Biotechnology, 2017. **259**: p. 235-247.

73. Karst, D.J., et al., *Modulation and modeling of monoclonal antibody N-linked glycosylation in mammalian cell perfusion reactors.* Biotechnology and Bioengineering, 2017. **114**(9): p. 1978-1990.

74. Selvarasu, S., et al., Combined data preprocessing and multivariate statistical analysis characterizes fed-batch culture of mouse hybridoma cells for rational medium design. J Biotechnol, 2010. **150**(1): p. 94-100.

75. Vodopivec, M., et al., Metabolomic profiling of CHO fed-batch growth phases at 10, 100, and 1,000 L. Biotechnol Bioeng, 2019. **116**(10): p. 2720-2729.

76. Ben Yahia, B., et al., Segmented linear modeling of CHO fed-batch culture and its application to large scale production. Biotechnol Bioeng, 2017. **114**(4): p. 785-797.

77. Salim, T., et al., Using MVDA with stoichiometric balances to optimize amino acid concentrations in chemically defined CHO cell culture medium for improved culture performance. Biotechnol Bioeng, 2022. **119**(2): p. 452-469.

78. Banton, D., et al., Reduced scale model qualification of 5-L and 250-ml bioreactors using multivariant visualization and Bayesian inferential methods. Biotechnol Bioeng, 2020. **117**(5): p. 1337-1347.

79. Goldrick, S., et al., *Multivariate Data Analysis Methodology to Solve Data Challenges Related to Scale-Up Model Validation and Missing Data on a Micro-Bioreactor System.* Biotechnol J, 2020. **15**(3): p. e1800684.

80. Brinc, M. and A. Belič, Optimization of process conditions for mammalian fed-batch cell culture in automated micro-bioreactor system using genetic algorithm. J Biotechnol, 2019. **300**: p. 40-47.

81. Yeo, H.C., et al., Combined multivariate statistical and flux balance analyses uncover media bottlenecks to the growth and productivity of Chinese hamster ovary cell cultures. Biotechnol Bioeng, 2022. **119**(7): p. 1740-1754.

82. Hong, J.K., et al., Data-driven and model-guided systematic framework for media development in CHO cell culture. Metabolic Engineering, 2022. **73**: p. 114-123.

83. Park, S.-Y., et al., Characterizing Basal and Feed Media Effects on Mammalian Cell Cultures by Systems Engineering Approaches. IFAC-PapersOnLine, 2022. **55**(7): p. 31-36.

84. Schinn, S.M., et al., A genome-scale metabolic network model and machine learning predict amino acid concentrations in Chinese Hamster Ovary cell cultures. Biotechnology and Bioengineering, 2021. **118**(5): p. 2118-2123.

85. Ramos, J.R.C., et al., Genome-scale modeling of Chinese hamster ovary cells by hybrid semi-parametric flux balance analysis. Bioprocess Biosyst Eng, 2022. **45**(11): p. 1889-1904.

86. Khaleghi, M.K., et al., Synergisms of machine learning and constraint-based modeling of metabolism for analysis and optimization of fermentation parameters. Biotechnology Journal, 2021. **16**(11): p. 2100212.

87. Bayer, B., et al., Comparison of mechanistic and hybrid modeling approaches for characterization of a CHO cultivation process: Requirements, pitfalls and solution paths. Biotechnology Journal, 2023. **18**(1): p. 2200381.

88. Tsopanoglou, A. and I. Jiménez del Val, Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses. Current Opinion in Chemical Engineering, 2021. **32**: p. 100691.

89. Wang, Q., et al., *Glycoengineering of CHO Cells to Improve Product Quality*. 2017, Springer New York. p. 25-44.

90. Pothukuchi, P., et al., *Translation of genome to glycome: role of the Golgi apparatus.* FEBS Letters, 2019. **593**(17): p. 2390-2411.

91. Yu, L.X., et al., *Understanding Pharmaceutical Quality by Design.* The AAPS Journal, 2014. **16**(4): p. 771-783.

92. Hajba, L., et al., *On the glycosylation aspects of biosimilarity.* Drug Discovery Today, 2018. **23**(3): p. 616-625.

93. Umaña, P. and J.E. Bailey, *A mathematical model of N-linked glycoform biosynthesis.* Biotechnology and Bioengineering, 1997. **55**(6): p. 890-908.

94. Krambeck, F.J. and M.J. Betenbaugh, *A mathematical model of N-linked glycosylation.* Biotechnology and Bioengineering, 2005. **92**(6): p. 711-728.

95. Jimenez Del Val, I., J.M. Nagy, and C. Kontoravdi, A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. Biotechnology Progress, 2011. **27**(6): p. 1730-1743.

96. Zhang, L., et al., Glycan Residues Balance Analysis - GReBA: A novel model for the N-linked glycosylation of IgG produced by CHO cells. Metab Eng, 2020. **57**: p. 118-128.

97. Spahn, P.N., et al., *A Markov chain model for N-linked protein glycosylation – towards a low-parameter tool for model-driven glycoengineering.* Metabolic Engineering, 2016. **33**: p. 52-66.

98. Kotidis, P. and C. Kontoravdi, Harnessing the potential of artificial neural networks for predicting protein glycosylation. Metabolic Engineering Communications, 2020. **10**: p. e00131.

99. Yang, O. and M. Ierapetritou, mAb Production Modeling and Design Space Evaluation Including Glycosylation Process. Processes, 2021. **9**(2): p. 324.

100. Kremkow, B.G. and K.H. Lee, Glyco-Mapper: A Chinese hamster ovary (CHO) genome-specific glycosylation prediction tool. Metabolic Engineering, 2018. **47**: p. 134-142.

101. Hutter, S., et al., Glycosylation flux analysis reveals dynamic changes of intracellular glycosylation flux distribution in Chinese hamster ovary fed-batch cultures. Metabolic Engineering, 2017. **43**: p. 9-20.

102. Hutter, S., et al., Glycosylation Flux Analysis of Immunoglobulin G in Chinese Hamster Ovary Perfusion Cell Culture. Processes, 2018. **6**(10): p. 176.

103. Fung Shek, C., P. Kotidis, and M. Betenbaugh, *Mechanistic and data-driven modeling of protein glycosylation.* Current Opinion in Chemical Engineering, 2021. **32**: p. 100690.

104. Kotidis, P., et al., *DigiGlyc: A hybrid tool for reactive scheduling in cell culture systems.* Computers & Chemical Engineering, 2021. **154**: p. 107460.

105. Štor, J., et al., Towards rational glyco-engineering in CHO: from data to predictive models. Current Opinion in Biotechnology, 2021. **71**: p. 9-17.

106. Jefferis, R., *Glycosylation of recombinant antibody therapeutics.* Biotechnol Prog, 2005. **21**(1): p. 11-6.

107. van de Bovenkamp, F.S., et al., Variable Domain N-Linked Glycans Acquired During Antigen-Specific Immune Responses Can Contribute to Immunoglobulin G Antibody Stability. Front Immunol, 2018. **9**: p. 740.

108. Van De Bovenkamp, F.S., et al., *The Emerging Importance of IgG Fab Glycosylation in Immunity.* The Journal of Immunology, 2016. **196**(4): p. 1435-1441.

109. Xu, Q., et al., *A study of the possible role of Fab-glycosylated IgG in tumor immunity.* Cancer Immunol Immunother, 2021. **70**(7): p. 1841-1851.

110. Zhang, P., et al., *Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs.* Drug Discov Today, 2016. **21**(5): p. 740-65.

111. Batra, J. and A.S. Rathore, *Glycosylation of monoclonal antibody products: Current status and future prospects.* Biotechnology Progress, 2016. **32**(5): p. 1091-1102.

112. Reinhart, D., et al., *Bioprocessing of Recombinant CHO-K1, CHO-DG44, and CHO-S: CHO Expression Hosts Favor Either mAb Production or Biomass Synthesis.* Biotechnol J, 2019. **14**(3): p. e1700686.

113. Krambeck, F.J., et al., *Model-based analysis of N-glycosylation in Chinese hamster ovary cells.* PLOS ONE, 2017. **12**(5): p. e0175376.

114. Kotidis, P., et al., *Model-based optimization of antibody galactosylation in CHO cell culture.* Biotechnology and Bioengineering, 2019. **116**(7): p. 1612-1626.

115. Sou, S.N., et al., *Model-based investigation of intracellular processes determining antibody Fc-glycosylation under mild hypothermia.* Biotechnol Bioeng, 2017. **114**(7): p. 1570-1582.

116. Stach, C.S., et al., *Model-Driven Engineering of N-Linked Glycosylation in Chinese Hamster Ovary Cells.* ACS Synthetic Biology, 2019. **8**(11): p. 2524-2535.

117. Naik, H.M., N.I. Majewska, and M.J. Betenbaugh, *Impact of nucleotide sugar metabolism on protein N-glycosylation in Chinese Hamster Ovary (CHO) cell culture.* Current Opinion in Chemical Engineering, 2018. **22**: p. 167-176.

118. Yoon, S.K., et al., *Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37.0 C.* Biotechnology and bioengineering, 2005. **89**(3): p. 345-356.

119. McHugh, K.P., et al., *Effective temperature shift strategy development and scale confirmation for simultaneous optimization of protein productivity and quality in Chinese hamster ovary cells.* Biotechnology Progress, 2020. **36**(3).

120. Aghamohseni, H., et al., *A semi-empirical glycosylation model of a camelid monoclonal antibody under hypothermia cell culture conditions.* J Ind Microbiol Biotechnol, 2017. **44**(7): p. 1005-1020.

121. Bollati-Fogolín, M., et al., *Temperature reduction in cultures of hGM-CSF-expressing CHO cells: effect on productivity and product quality.* Biotechnol Prog, 2005. **21**(1): p. 17-21.

122. Fox, S.R., et al., *Maximizing interferon-gamma production by Chinese hamster ovary cells through temperature shift optimization: experimental and modeling.* Biotechnol Bioeng, 2004. **85**(2): p. 177-84.

123. Martínez, V.S., et al., Dynamic metabolic flux analysis using B-splines to study the effects of temperature shift on CHO cell metabolism. Metabolic Engineering Communications, 2015. **2**: p. 46-57.

124. Ahn, W.S., et al., Effect of culture temperature on erythropoietin production and glycosylation in a perfusion culture of recombinant CHO cells. Biotechnology and Bioengineering, 2008. **101**(6): p. 1234-1244.

125. Andersen, D.C. and C.F. Goochee, *THE EFFECT OF AMMONIA ON THE O-LINKED GLYCOSYLATION OF GRANULOCYTE-COLONY-STIMULATING FACTOR PRODUCED BY CHINESE-HAMSTER OVARY CELLS.* Biotechnology and Bioengineering, 1995. **47**(1): p. 96-105.

126. Borys, M.C., D.I.H. Linzer, and E.T. Papoutsakis, *CULTURE PH AFFECTS EXPRESSION RATES AND GLYCOSYLATION OF RECOMBINANT MOUSE PLACENTAL-LACTOGEN PROTEINS BY CHINESE-HAMSTER OVARY (CHO) CELLS.* Bio-Technology, 1993. **11**(6): p. 720-724.

127. Borys, M.C., D.I.H. Linzer, and E.T. Papoutsakis, *AMMONIA AFFECTS THE GLYCOSYLATION PATTERNS OF RECOMBINANT MOUSE PLACENTAL LACTOGEN-1 BY CHINESE-HAMSTER OVARY CELLS IN A PH-DEPENDENT MANNER.* Biotechnology and Bioengineering, 1994. **43**(6): p. 505-514.

128. Goochee, C.F. and T. Monica, *ENVIRONMENTAL-EFFECTS ON PROTEIN GLYCOSYLATION.* Bio-Technology, 1990. **8**(5): p. 421-427.

129. Gramer, M.J. and C.F. Goochee, *GLYCOSIDASE ACTIVITIES IN CHINESE-HAMSTER OVARY CELL LYSATE AND CELL-CULTURE SUPERNATANT.* Biotechnology Progress, 1993. **9**(4): p. 366-373.

130. Jiang, R., H. Chen, and S. Xu, *pH excursions impact CHO cell culture performance and antibody N-linked glycosylation.* Bioprocess and Biosystems Engineering, 2018. **41**(12): p. 1731-1741.

131. Hogiri, T., et al., *Optimization of a pH-shift control strategy for producing monoclonal antibodies in Chinese hamster ovary cell cultures using a pH-dependent dynamic model.* J Biosci Bioeng, 2018. **125**(2): p. 245-250.

132. Lee, A.P., et al., *Multi-omics profiling of a CHO cell culture system unravels the effect of culture pH on cell growth, antibody titer, and product quality.* Biotechnol Bioeng, 2021. **118**(11): p. 4305-4316.

133. Seidel, S., et al., *Oxygen Mass Transfer in Biopharmaceutical Processes: Numerical and Experimental Approaches.* Chemie Ingenieur Technik, 2021. **93**(1-2): p. 42-61.

134. Li, F., et al., *Cell culture processes for monoclonal antibody production.* MAbs, 2010. **2**(5): p. 466-79.

135. Anane, E., I.M. Knudsen, and G.C. Wilson, *Scale-down cultivation in mammalian cell bioreactors—The effect of bioreactor mixing time on the response of CHO cells to dissolved oxygen gradients.* Biochemical Engineering Journal, 2021. **166**: p. 107870.

136. Link, T., et al., *Bioprocess development for the production of a recombinant MUC1 fusion protein expressed by CHO-K1 cells in protein-free medium.* J Biotechnol, 2004. **110**(1): p. 51-62.

137. Brunner, M., et al., *The impact of pH inhomogeneities on CHO cell physiology and fed-batch process performance - two-compartment scale-down modelling and intracellular pH excursion.* Biotechnol J, 2017. **12**(7).

138. Brunner, M., et al., Investigation of the interactions of critical scale-up parameters (pH, pO(2) and pCO(2)) on CHO batch performance and critical quality attributes. Bioprocess Biosyst Eng, 2017. **40**(2): p. 251-263.

139. Restelli, V., et al., The effect of dissolved oxygen on the production and the glycosylation profile of recombinant human erythropoietin produced from CHO cells. Biotechnol Bioeng, 2006. **94**(3): p. 481-94.

140. Hippach, M.B., et al., *Fluctuations in dissolved oxygen concentration during a CHO cell culture process affects monoclonal antibody productivity and the sulfhydryl-drug conjugation process.* Biotechnology Progress, 2018. **34**(6): p. 1427-1437.

141. Hossler, P., S.F. Khattak, and Z.J. Li, *Optimal and consistent protein glycosylation in mammalian cell culture.* Glycobiology, 2009. **19**(9): p. 936-949.

142. Kunkel, J.P., et al., *Comparisons of the Glycosylation of a Monoclonal Antibody Produced under Nominally Identical Cell Culture Conditions in Two Different Bioreactors.* Biotechnology Progress, 2000. **16**(3): p. 462-470.

143. Kunkel, J.P., et al., *Dissolved oxygen concentration in serum-free continuous culture affects N-linked glycosylation of a monoclonal antibody.* J Biotechnol, 1998. **62**(1): p. 55-71.

144. Serrato, J.A., et al., *Heterogeneous conditions in dissolved oxygen affect N-glycosylation but not productivity of a monoclonal antibody in hybridoma cultures.* Biotechnology and Bioengineering, 2004. **88**(2): p. 176-188.

145. Sengupta, N., S.T. Rose, and J.A. Morgan, *Metabolic flux analysis of CHO cell metabolism in the late non-growth phase.* Biotechnology and Bioengineering, 2011. **108**(1): p. 82-92.

146. Farzan, P. and M.G. Ierapetritou, Integrated modeling to capture the interaction of physiology and fluid dynamics in biopharmaceutical bioreactors. Computers & Chemical Engineering, 2017. **97**: p. 271-282.

147. Ghorbaniaghdam, A., O. Henry, and M. Jolicoeur, *An in-silico study of the regulation of CHO cells glycolysis.* Journal of Theoretical Biology, 2014. **357**: p. 112-122.

148. Gao, Y., et al., *Combined metabolomics and proteomics reveals hypoxia as a cause of lower productivity on scale-up to a 5000-liter CHO bioprocess.* Biotechnology Journal, 2016. **11**(9): p. 1190-1200.

149. Clanton, T.L., *Hypoxia-induced reactive oxygen species formation in skeletal muscle.* Journal of Applied Physiology, 2007. **102**(6): p. 2379-2388.

150. Majmundar, A.J., W.J. Wong, and M.C. Simon, *Hypoxia-Inducible Factors and the Response to Hypoxic Stress.* Molecular Cell, 2010. **40**(2): p. 294-309.

151. Yao, T. and Y. Asayama, *Animal-cell culture media: History, characteristics, and current issues.* Reproductive Medicine and Biology, 2017. **16**(2): p. 99-117.

152. deZengotita, V.M., et al., *Selected amino acids protect hybridoma and CHO cells from elevated carbon dioxide and osmolality.* Biotechnology and Bioengineering, 2002. **78**(7): p. 741-752.

153. Kimura, R. and W.M. Miller, *Effects of elevated pCO(2) and/or osmolality on the growth and recombinant tPA production of CHO cells.* Biotechnology and Bioengineering, 1996. **52**(1): p. 152-160.

154. Kimura, R. and W.M. Miller, *Glycosylation of CHO-derived recombinant tPA produced under elevated pCO(2).* Biotechnology Progress, 1997. **13**(3): p. 311-317.

155. Schmelzer, A.E. and W.M. Miller, *Effects of osmoprotectant compounds on NCAM polysialylation under hyperosmotic stress and elevated pCO(2).* Biotechnology and Bioengineering, 2002. **77**(4): p. 359-368.

156. Zanghi, J.A., et al., *Bicarbonate concentration and osmolality are key determinants in the inhibition of CHO cell polysialylation under elevated pCO(2) or pH.* Biotechnology and Bioengineering, 1999. **65**(2): p. 182-191.

157. Brunner, M., et al., *Elevated pCO2 affects the lactate metabolic shift in CHO cell culture processes.* Engineering in Life Sciences, 2018. **18**(3): p. 204-214.

158. Aghamohseni, H., et al., *Effects of nutrient levels and average culture pH on the glycosylation pattern of camelid-humanized monoclonal antibody.* J Biotechnol, 2014. **186**: p. 98-109.

159. Ivarsson, M., et al., *Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation.* J Biotechnol, 2014. **188**: p. 88-96.

160. Nolan, R.P. and K. Lee, *Dynamic model for CHO cell engineering.* Journal of Biotechnology, 2012. **158**(1): p. 24-33.

161. Diab, S., et al., *Dynamic Simulation and Visualisation of pH-Modulated Fed-batch Fermentation for mAb Production from CHO Cell Cultures*. 2020, Elsevier. p. 1657-1662.

162. Huang, Z., et al., *CHO cell productivity improvement by genome-scale modeling and pathway analysis: Application to feed supplements.* Biochemical Engineering Journal, 2020. **160**: p. 107638.

163. Shirahata, H., et al., *Dynamic modelling, simulation and economic evaluation of two CHO cell-based production modes towards developing biopharmaceutical manufacturing processes.* Chemical Engineering Research and Design, 2019. **150**: p. 218-233.

164. Gutierrez, J.M., et al., *Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion.* Nature Communications, 2020. **11**(1): p. 68.

165. Ha, T.K., et al., *Factors affecting the quality of therapeutic proteins in recombinant Chinese hamster ovary cell culture.* Biotechnology Advances, 2022. **54**: p. 107831.

166. Kol, S., et al., *Multiplex secretome engineering enhances recombinant protein production and purity.* Nat Commun, 2020. **11**(1): p. 1908.

167. Ritacco, F.V., Y. Wu, and A. Khetan, *Cell culture media for recombinant protein expression in Chinese hamster ovary (CHO) cells: History, key components, and optimization strategies.* Biotechnology Progress, 2018. **34**(6): p. 1407-1426.

168. Fouladiha, H., et al., A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. Bioprocess and Biosystems Engineering, 2020. **43**(8): p. 1381-1389.

169. Nargund, S., J. Qiu, and C.T. Goudar, *Elucidating the role of copper in CHO cell energy metabolism using¹³C metabolic flux analysis.* Biotechnology Progress, 2015. **31**(5): p. 1179-1186.

170. Stone, A.T., et al., *Chemical speciation of trace metals in mammalian cell culture media: looking under the hood to boost cellular performance and product quality.* Curr Opin Biotechnol, 2021. **71**: p. 216-224.

171. Xing, Z., et al., Optimizing amino acid composition of CHO cell culture media for a fusion protein production. Process Biochemistry, 2011. **46**(7): p. 1423-1429.

172. Park, J.M., et al., *Flux variability scanning based on enforced objective flux for identifying gene amplification targets.* BMC Systems Biology, 2012. **6**(1): p. 106.

173. Ben Yahia, B., L. Malphettes, and E. Heinzle, *Predictive macroscopic modeling of cell growth, metabolism and monoclonal antibody production: Case study of a CHO fed-batch production.* Metabolic Engineering, 2021. **66**: p. 204-216.

174. Graham, R.J., H. Bhatia, and S. Yoon, *Consequences of trace metal variability and supplementation on Chinese hamster ovary (CHO) cell culture performance: A review of key mechanisms and considerations.* Biotechnol Bioeng, 2019. **116**(12): p. 3446-3456.

175. Alhuthali, S. and C. Kontoravdi, *Population balance modelling captures host cell protein dynamics in CHO cell cultures.* PLOS ONE, 2022. **17**(3): p. e0265886.

176. Savizi, I.S.P., et al., *Valine feeding reduces ammonia production through rearrangement of metabolic fluxes in central carbon metabolism of CHO cells.* Appl Microbiol Biotechnol, 2022. **106**(3): p. 1113-1126.

177. Ley, D., et al., Reprogramming AA catabolism in CHO cells with CRISPR/Cas9 genome editing improves cell growth and reduces byproduct secretion. Metabolic Engineering, 2019. **56**: p. 120-129.

178. Torkashvand, F. and B. Vaziri, *Main Quality Attributes of Monoclonal Antibodies and Effect of Cell Culture Components*, in *Iran Biomed J*. 2017. p. 131-41.

179. Leweke, S. and E. von Lieres, *Chromatography Analysis and Design Toolkit (CADET).* Computers & Chemical Engineering, 2018. **113**: p. 274-294.

180. Ding, C. and M. Ierapetritou, *A novel framework of surrogate-based feasibility analysis for establishing design space of twin-column continuous chromatography.* Int J Pharm, 2021. **609**: p. 121161.

181. Yang, O. and M. Ierapetritou, Application of PSE Methods on Monoclonal Antibody Productivity Improvement and Quality Control, in Computer Aided Chemical Engineering, Y. Yamashita and M. Kano, Editors. 2022, Elsevier. p. 2215-2220.

182. Yang, O., et al., Process Design and Comparison for Batch and Continuous Manufacturing of Recombinant Adeno-Associated Virus. Journal of Pharmaceutical Innovation, 2022.

183. Ding, C., et al., *Process design of a fully integrated continuous biopharmaceutical process using economic and ecological impact assessment.* Biotechnology and Bioengineering, 2022.

184. Chen, N., M.H. Bennett, and C. Kontoravdi, *Analysis of Chinese hamster ovary cell metabolism through a combined computational and experimental approach.* Cytotechnology, 2014. **66**(6): p. 945-966.

185. Erklavec Zajec, V., et al., Dynamic multiscale metabolic network modeling of Chinese hamster ovary cell metabolism integrating N-linked glycosylation in industrial biopharmaceutical manufacturing. Biotechnology and Bioengineering, 2021. **118**(1): p. 397-411.

186. Aehle, M., et al., *Increasing batch-to-batch reproducibility of CHO-cell cultures using a model predictive control approach.* Cytotechnology, 2012. **64**(6): p. 623-634.

187. Craven, S., J. Whelan, and B. Glennon, *Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller.* Journal of Process Control, 2014. **24**(4): p. 344-357.

188. Carswell, K.S. and E.T. Papoutsakis, *Culture of human T cells in stirred bioreactors for cellular immunotherapy applications: Shear, proliferation, and the IL-2 receptor.* Biotechnology and Bioengineering, 2000. **68**(3): p. 328-338.

189. Carswell, K.S. and E.T. Papoutsakis, *Extracellular pH affects the proliferation of cultured human T cells and their expression of the interleukin-2 receptor.* Journal of Immunotherapy, 2000. **23**(6): p. 669-674.

190. Carswell, K.S., J.W. Weiss, and E.T. Papoutsakis, *Low oxygen tension enhances the stimulation and proliferation of human T lymphocytes in the presence of IL-2.* Cytotherapy, 2000. **2**(1): p. 25-37.

191. Haddad, H. and E.T. Papoutsakis, Low oxygen tension and autologous plasma enhance T-cell proliferation and CD49d expression density in serum-free media. Cytotherapy, 2001. **3**(6): p. 435-447.

192. Haddad, H., et al., Molecular understanding of oxygen tension and patient-variability effects on ex vivo expanded T cells. Biotechnology and Bioengineering, 2004. **87**(4): p. 437-450.

193. Calmels, C., et al., *Application of a curated genome-scale metabolic model of CHO DG44 to an industrial fed-batch process.* Metabolic Engineering, 2019. **51**: p. 9-19.

194. Liu, L., Antibody glycosylation and its impact on the pharmacokinetics and pharmacodynamics of monoclonal antibodies and Fc-fusion proteins. J Pharm Sci, 2015. **104**(6): p. 1866-1884.

195. Ludwig, S.D., et al. A versatile design platform for glycoengineering therapeutic antibodies. Taylor & Francis.

196. Reddy, J.V., et al., Cell-culture process optimization via model-based predictions of metabolism and protein glycosylation. Biotechnology Advances, 2023. **67**: p. 108179.

197. Pan, X., et al., Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. Appl Microbiol Biotechnol, 2017. **101**(22): p. 8101-8113.

198. Ghaffari, N., et al., *Effects of cysteine, asparagine, or glutamine limitations in Chinese hamster ovary cell batch and fed-batch cultures.* Biotechnology Progress, 2020. **36**(2): p. e2946.

199. Jiménez Del Val, I., et al., *CHOmpact: A reduced metabolic model of Chinese hamster ovary cells with enhanced interpretability.* Biotechnol Bioeng, 2023. **120**(9): p. 2479-2493.

200. Ozturk, S.S. and B.O. Palsson, Chemical decomposition of glutamine in cell culture media: effect of media type, pH, and serum concentration. Biotechnol Prog, 1990. **6**(2): p. 121-8.

201. Sou, S.N., et al., *How does mild hypothermia affect monoclonal antibody glycosylation?* Biotechnology and Bioengineering, 2015. **112**(6): p. 1165-1176.

202. Zamorano, F., A.V. Wouwer, and G. Bastin, *A detailed metabolic flux analysis of an underdetermined network of CHO cells.* J Biotechnol, 2010. **150**(4): p. 497-508.

203. Carinhas, N., et al., *Metabolic signatures of GS-CHO cell clones associated with butyrate treatment and culture phase transition.* Biotechnol Bioeng, 2013. **110**(12): p. 3244-57.

204. Ebrahim, A., et al., *COBRApy: COnstraints-Based Reconstruction and Analysis for Python.* BMC Systems Biology, 2013. **7**(1): p. 74.

205. Cordova, L.T., et al., Generation of reference cell lines, media, and a process platform for CHO cell biomanufacturing. Biotechnology and Bioengineering, 2023. **120**(3): p. 715-725.

206. Alhuthali, S., P. Kotidis, and C. Kontoravdi, *Osmolality Effects on CHO Cell Growth, Cell Volume, Antibody Productivity and Glycosylation.* Int J Mol Sci, 2021. **22**(7).

207. Yoon, S.K., et al., *Effect of culture pH on erythropoietin production by Chinese hamster ovary cells grown in suspension at 32.5 and 37.0°C.* Biotechnology and Bioengineering, 2005. **89**(3): p. 345-356.

208. Alfarouk, K.O., et al., Glycolysis, tumor metabolism, cancer growth and dissemination. A new pH-based etiopathogenic perspective and therapeutic approach to an old cancer question. Oncoscience, 2014. **1**(12): p. 777-802.

209. Zagari, F., et al., *Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity.* N Biotechnol, 2013. **30**(2): p. 238-45.

210. Reitzer, L.J., B.M. Wice, and D. Kennell, *Evidence that glutamine, not sugar, is the major energy source for cultured HeLa cells.* J Biol Chem, 1979. **254**(8): p. 2669-76.

211. Naik, H.M., et al., *Chemical inhibitors of hexokinase-2 enzyme reduce lactate accumulation, alter glycosylation processing, and produce altered glycoforms in CHO cell cultures.* Biotechnology and Bioengineering, 2023. **120**(9): p. 2559-2577.

212. Yang, M. and M. Butler, *Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation.* Biotechnology and Bioengineering, 2000. **68**(4): p. 370-380.

213. Genzel, Y., et al., *Substitution of glutamine by pyruvate to reduce ammonia formation and growth inhibition of mammalian cells.* Biotechnol Prog, 2005. **21**(1): p. 58-69.

214. Lee, J.H., et al., Analysis of Golgi pH in Chinese hamster ovary cells using ratiometric pH-sensitive fluorescent proteins. Biotechnology and Bioengineering, 2019. **116**(5): p. 1006-1016.

215. Xu, P., et al., Effects of glutamine and asparagine on recombinant antibody production using CHO-GS cell lines. Biotechnology Progress, 2014. **30**(6): p. 1457-1468.

216. Synoground, B.F., et al., *Transient ammonia stress on Chinese hamster ovary (CHO) cells yield alterations to alanine metabolism and IgG glycosylation profiles.* Biotechnology Journal, 2021. **16**(7): p. 2100098.

217. Hong, J.K., et al., *In silico model-based characterization of metabolic response to harsh sparging stress in fed-batch CHO cell cultures.* J Biotechnol, 2020. **308**: p. 10-20.

218. Young, J.D., *Metabolic flux rewiring in mammalian cell cultures.* Curr Opin Biotechnol, 2013. **24**(6): p. 1108-15.

219. Luo, S. and B. Zhang, *Benchmark Glycan Profile of Therapeutic Monoclonal Antibodies Produced by Mammalian Cell Expression Systems.* Pharmaceutical Research, 2024. **41**(1): p. 29-37.

220. Anumula, K.R., *Quantitative glycan profiling of normal human plasma derived immunoglobulin and its fragments Fab and Fc.* J Immunol Methods, 2012. **382**(1-2): p. 167-76.

221. Flevaris, K. and C. Kontoravdi *Immunoglobulin G N-glycan Biomarkers for Autoimmune Diseases: Current State and a Glycoinformatics Perspective*. International Journal of Molecular Sciences, 2022. **23**, DOI: 10.3390/ijms23095180.

222. Malinov, N.G., et al., *End-to-end process flowsheet modeling for biopharmaceutical production: current state and future potential.* Current Opinion in Chemical Engineering, 2024. **45**: p. 101044.

223. Hutter, S., et al., *Glycosylation flux analysis reveals dynamic changes of intracellular glycosylation flux distribution in Chinese hamster ovary fed-batch cultures.* Metab Eng, 2017. **43**(Pt A): p. 9-20.

224. Reddy, J.V., et al., *Flux balance analysis and peptide mapping elucidate the impact of bioreactor pH on Chinese Hamster Ovary (CHO) cell metabolism and N-linked glycosylation in the Fab and Fc regions of the produced IgG.* bioRxiv, 2024: p. 2024.08.01.606220.

225. Wahrheit, J. and E. Heinzle, Quenching Methods for the Analysis of Intracellular Metabolites, in Animal Cell Biotechnology: Methods and Protocols, R. Pörtner, Editor. 2014, Humana Press: Totowa, NJ. p. 211-221.

226. del Val, I.J., et al., *An optimized method for extraction and quantification of nucleotides and nucleotide sugars from mammalian cells.* Analytical biochemistry, 2013. **443**(2): p. 172-180.

227. Dietmair, S., et al., Towards quantitative metabolomics of mammalian cells: development of a metabolite extraction protocol. Analytical biochemistry, 2010. **404**(2): p. 155-164.

228. Tomiya, N., et al., Determination of nucleotides and sugar nucleotides involved in protein glycosylation by high-performance anion-exchange chromatography: sugar nucleotide contents in cultured insect cells and mammalian cells. Anal Biochem, 2001. **293**(1): p. 129-37.

229. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method.* methods, 2001. **25**(4): p. 402-408.

230. Kotidis, P., et al., *CHOGlycoNET: Comprehensive glycosylation reaction network for CHO cells.* Metab Eng, 2023. **76**: p. 87-96.

231. Feige, M.J., L.M. Hendershot, and J. Buchner, *How antibodies fold.* Trends Biochem Sci, 2010. **35**(4): p. 189-98.

232. Wang, Q., et al., *The interplay of protein engineering and glycoengineering to fine-tune antibody glycosylation and its impact on effector functions.* Biotechnology and Bioengineering, 2022. **119**(1): p. 102-117.

233. Marquardt, D.W., *An algorithm for least-squares estimation of nonlinear parameters.* Journal of the society for Industrial and Applied Mathematics, 1963. **11**(2): p. 431-441.

234. Virtanen, P., et al., *SciPy 1.0: fundamental algorithms for scientific computing in Python.* Nature Methods, 2020. **17**(3): p. 261-272.

235. Arigoni-Affolter, I., et al., *Mechanistic reconstruction of glycoprotein secretion through monitoring of intracellular N-glycan processing.* Sci Adv, 2019. **5**(11): p. eaax8930.

236. Fan, Y., et al., *A multi-pronged investigation into the effect of glucose starvation and culture duration on fed-batch CHO cell culture.* Biotechnology and Bioengineering, 2015. **112**(10): p. 2172-2184.

237. Zalai, D., et al., *A control strategy to investigate the relationship between specific productivity and high-mannose glycoforms in CHO cells.* Appl Microbiol Biotechnol, 2016. **100**(16): p. 7011-24.

238. Cambay, F., et al., *Impact of IgG1 N-glycosylation on their interaction with Fc gamma receptors.* Current Research in Immunology, 2020. **1**: p. 23-37.

239. Liu, J., et al., Galactose supplementation enhance sialylation of recombinant Fc-fusion protein in CHO cell: an insight into the role of galactosylation in sialylation. World J Microbiol Biotechnol, 2015. **31**(7): p. 1147-56.

240. Hossler, P., B.C. Mulukutla, and W.-S. Hu, *Systems Analysis of N-Glycan Processing in Mammalian Cells.* PLOS ONE, 2007. **2**(8): p. e713.

241. Saltelli, A., *Making best use of model evaluations to compute sensitivity indices.* Computer Physics Communications, 2002. **145**(2): p. 280-297.

242. Sobol′, I.M., *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates.* Mathematics and Computers in Simulation, 2001. **55**(1): p. 271-280.

243. Herman, J. and W. Usher, *SALib: An open-source Python library for Sensitivity Analysis.* The Journal of Open Source Software, 2017. **2**.

244. Bydlinski, N., et al., *The contributions of individual galactosyltransferases to protein specific N-glycan processing in Chinese Hamster Ovary cells.* Journal of Biotechnology, 2018. **282**: p. 101-110.

245. Chen, C., H.E. Wong, and C.T. Goudar, *Upstream process intensification and continuous manufacturing.* Current Opinion in Chemical Engineering, 2018. **22**: p. 191-198.

246. Yang, O., S. Prabhu, and M. Ierapetritou, *Comparison between Batch and Continuous Monoclonal Antibody Production and Economic Analysis.* Industrial & Engineering Chemistry Research, 2019. **58**(15): p. 5851-5863.

247. Berrios, J., et al., *Continuous CHO cell cultures with improved recombinant protein productivity by using mannose as carbon source: Metabolic analysis and scale-up simulation.* Chemical Engineering Science, 2011. **66**(11): p. 2431-2439.

248. Kuang, B., et al., Identification of novel inhibitory metabolites and impact verification on growth and protein synthesis in mammalian cells. Metab Eng Commun, 2021. **13**: p. e00182.

249. Holzhütter, H.-G., *The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks.* European Journal of Biochemistry, 2004. **271**(14): p. 2905-2922.

250. Lewis, N.E., et al., *Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models.* Mol Syst Biol, 2010. **6**: p. 390.

251. Wächter, A. and L.T. Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming, 2006. **106**(1): p. 25-57.

252. Hart, W., et al., *Pyomo: Modeling and solving mathematical programs in Python.* Mathematical Programming Computation, 2011. **3**: p. 219-260.

253. Tsallis, C., *Possible generalization of Boltzmann-Gibbs statistics.* Journal of Statistical Physics, 1988. **52**(1): p. 479-487.

254. Nelder, J.A. and R. Mead, *A Simplex Method for Function Minimization.* The Computer Journal, 1965. **7**(4): p. 308-313.

255. Klein, T., et al., *Quantification of cell lysis during CHO bioprocesses: Impact on cell count, growth kinetics and productivity.* Journal of Biotechnology, 2015. **207**: p. 67-76.

256. Chen, Y., et al. *Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review*. Processes, 2020. **8**, DOI: 10.3390/pr8091088.

257. Hoang, D., et al., *Characterization of Chinese hamster ovary cell culture feed media precipitate.* Biotechnol Prog, 2021. **37**(5): p. e3188.

Appendix A

SUPPLEMENTARY MATERIAL TO CHAPTER 2

Table A1: This table contains data on uptake and secretion rates calculated from the transient metabolic profile data shown in Figure 2.1. This data has been used to perform flux balance analysis. The units of the uptake and secretion rates are in fmol/cell/day. The units for the growth rate or biomass production rate is in 1/day.

Metabolite	pH 6.75			pH 7			pH 7.25	
	Day 0 to 3	Day 4 to 7	Day 8 to 11	Day 0 to 3	Day 4 to 7	Day 8 to 11	Day 0 to 3	Day 4 to 7
Biomass	0.58 ± 0.045	0.41 ± 0.03	0.0 ± 0.0	0.79 ± 0.019	0.41 ± 0.024	0.0 ± 0.0	0.71 ± 0.033	0.44 ± 0.022
GLC	-2327 ± 586	-1301 ± 47	-1037 ± 82	-3420 ± 150	-1364 ± 72	-1105 ± 99	-4941 ± 300	-1669 ± 88
LAC	3096 ± 601	-229 ± 69	24.11 ± 9	4167 ± 126	-254 ± 21	1.21 ± 7	7255 ± 982	-199 ± 70
ALA	1002 ± 111	-112 ± 22	-1.7 ± 4.2	955 ± 85	-117 ± 12	-13 ± 0.7	1051 ± 194	-56.9 ± 12
GLY	456 ± 27	44 ± 8	-7 ± 8	318 ± 44	-15 ± 2	13 ± 6	404 ± 59	22 ± 21
SER	-769 ± 23	-270 ± 29	-62 ± 15	-709 ± 99	-207 ± 10	-58 ± 3	-740 ± 56	-266 ± 31
GLN	-1776 ± 244	5.68 ± 6	13 ± 1	-1773 ± 145	2 ± 1	4 ± 1	-1837 ± 179	2 ± 2
GLU	162 ± 16	-56 ± 39	-122 ± 19	167 ± 18	-143 ± 18	-109 ± 12	344 ± 77	-68 ± 24
ASN	-202 ± 36	-215 ± 9	-133 ± 15	-279 ± 54	-227 ± 10	-65 ± 2	-315 ± 20	-268 ± 28
ASP	129 ± 11	-63 ± 6	-15 ± 17	68 ± 10	-98 ± 11	-55 ± 2	129 ± 13	-58 ± 20
MET	-45 ± 24	-41 ± 2	-11 ± 4	-73 ± 14	-43 ± 3	-9 ±1	-80 ± 36	-45 ± 1
TYR	-51 ± 3	-44 ± 8	-13 ± 5	-79 ± 17	-48 ± 5	-13 ± 1	-65 ± 12	-50 ± 4
PHE	-74 ± 19	-52 ± 5	-15 ± 9	-88 ± 16	-59 ± 3	-20 ± 1	-69 ± 23	-65 ± 2
TRP	-86 ± 33	-18 ± 1	-7 ± 2	-63 ± 24	-23 ± 3	-7 ± 1	-85 ± 34	-25 ± 1
VAL	-145 ± 12	-86 ± 10	-44 ± 10	-178 ± 17	-110 ± 11	-40 ± 3	-151 ± 27	-115 ± 8
LEU	-237 ± 42	-106 ± 12	-38 ± 12	-274 ± 90	-129 ± 11	-43 ± 1	-243 ± 36	-140 ± 4
ILE	-155 ± 17	-76 ± 6	-26 ± 10	-144 ± 32	-65 ± 23	-27 ± 4	-150 ± 16	-93 ± 3
ARG	-118 ± 20	-54 ± 5	-13 ± 7	-136 ± 6	-71 ± 5	-13 ± 1	-132 ± 3	-71 ± 2
HIS	-52 ± 18	-29 ± 3	-9 ± 2	-81 ± 29	-33 ± 3	-9 ± 1	-56 ± 12	-34 ± 2
THR	-106 ± 49	-57 ± 6	-19 ± 11	-150 ± 52	-81 ± 7	-22 ± 1	-112 ± 25	-87 ± 2
LYS	-139 ± 9	-73 ± 7	-17 ± 10	-201 ± 47	-94 ± 7	-20 ± 1	-176 ± 10	-95 ± 3
PRO	-254 ± 88	-145 ± 21	-47 ± 9	-288 ± 98	-142 ± 17	-39 ± 4	-201 ± 96	-144 ± 15
AMM	1265 ± 159	-55 ± 21	158 ± 2	1013 ± 42	-22 ± 15	91 ± 1	997 ± 110	-71 ± 15
IgG	0.04 ± 0.005			0.06 ± 0.006			0.08 ± 0.008	

Table A2: Quantification of mAb titers using protein A chromatography. Flow rate 3 mL/min. Injection volume 20 µL.

Time (min)	Gradient (%B)
0	0
0.5	0
0.51	100
1.5	100
1.51	0
3.0	0

Table A3: Peptide mapping amino acid backbone

Chain	N-linked glycan site	Peptide sequence
Light chain	N70	WGPDYNLTISNLE (65 – 77)
Heavy chain	N301	EEQYNSTYR (297 – 305)

Table A4: Glycan modification library used in UNIFI for site-specific glycan analysis. Delta mass displayed to the nearest 10^{-4} Da .

Order	Modified name	Delta mass
1	Glycosylation G0F N	1444.5339
2	Glycosylation G1F N	1605.5867
3	Glycosylation G0F-GlcNAc N	1241.4545
4	Glycosylation G1F+SA N	1897.6281
5	Glycosylation G2F+2SA N	2350.8304
6	Glycosylation Man5 N	1216.4229
7	Glycosylation G2F+SA N	2059.7349
8	Glycosylation G2F N	1768.6395
9	Glycosylation G2+2SA	2204.7724
10	Glycosylation G2+SA	1913.6770
11	Glycosylation G2 N	1622.5816
12	Glycosylation G1 N	1460.5288
13	Glycosylation G0 N	1298.4760
14	Glycosylation G0-GlcNAc N	1095.3966
15	Glycosylation G1F-GlcNAc N	1403.5073
16	Glycosylation Man6 N	1378.4757
17	Glycosylation Man7 N	1540.5285
18	Glycosylation Man 8 N	1702.5813
19	Glycosylation Man 9 N	1864.6342
20	Glycosylation Man3-GlcNAc1	1095.3966
21	Glycosylation Man3-GlcNAc1-Fuc	1241.4545
22	Glycosylation Man3-GlcNAc1-Gal1	1257.4494
23	Glycosylation Man3-GlcNAc-1-Gal1-Fuc	1403.5073
24	Glycosylation Man3-GlcNAc1-Gal-SA1	1548.5448
25	Glycosylation Man3-GlcNAc1-Gal-SA1-Fuc	1694.6027
26	Glycosylation Man4-GlcNAc1	1257.4494
27	Glycosylation Man4-GlcNAc1-Fuc	1403.5073
28	Glycosylation Man4-GlcNAc1-Gal1	1419.5022
29	Glycosylation Man4-GlcNAc1-Gal1-Fuc	1565.5601
30	Glycosylation Man4-GlcNAc1-Gal1-SA1	1710.5977
31	Glycosylation Man4-GlcNAc1-Gal1-SA1-Fuc	1856.6556
32	Glycosylation Man5-GlcNAc1	1419.5022
33	Glycosylation Man5-GlcNAc1-Fuc	1565.5601
34	Glycosylation Man5-GlcNAc1-Gal1	1581.5551
35	Glycosylation Man5-GlcNAc1-Gal1-Fuc	1727.6130
36	Glycosylation Man5-GlcNAc1-Gal1-SA1	1872.6505

Table A5: Waters BioAccord LC-MS method parameters used for site-specific glycan analysis.

Mobile phase A	Water + 0.1% formic acid
Mobile phase B	Acetonitrile + 0.1% formic acid
Gradient	**0 to 2 min**, 99.0% A **2 to 65 min**, 99 to 60% A **65 to 68 min**, 60 to 30% A **68 to 70 min**, 30% A **70 to 71 min**, 30 – 99% A **71 to 72 min,** 99% A **72 to 80 min**, 99 to 60% A **80 to 82 min**, 60 to 20% A **82 to 86 min**, 20% A **86 to 87 min**, 20 to 99% A
Flow rate (mL/min)	0.250
TUV wavelength (nm)	214
RDa mode	Full scan with fragmentation
Mass range	Low (50 – 2000 m/z)
Scan rate (Hz)	1
Capillary voltage (kV)	1.20
Desolvation temperature (°C)	350
Lockmass	ACQUITY RDa waters connect Lockmass Kit (Waters part number 186009298)
Lockmass correction mode	Dynamic

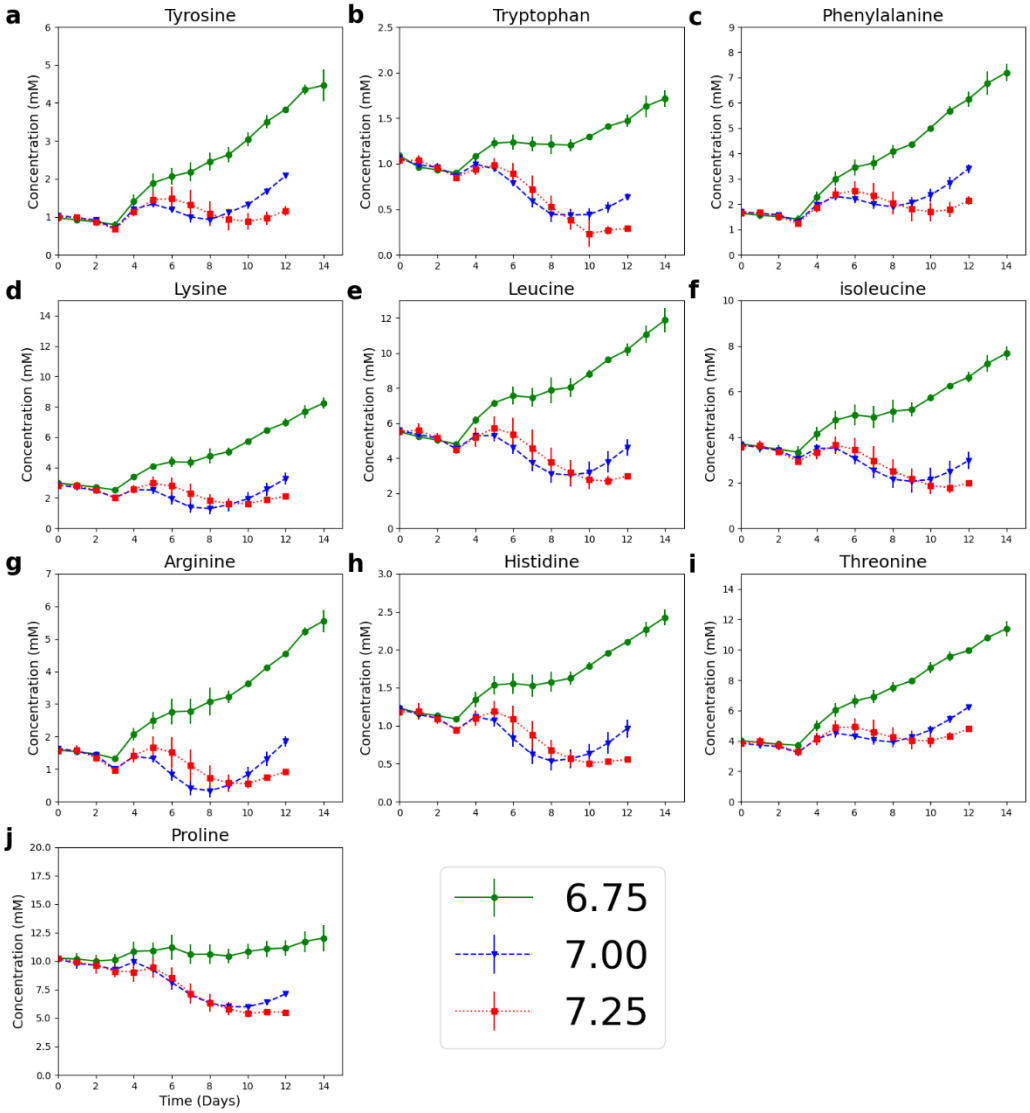

Figure A1: Effect of bioreactor pH and culture duration on amino acid concentrations.

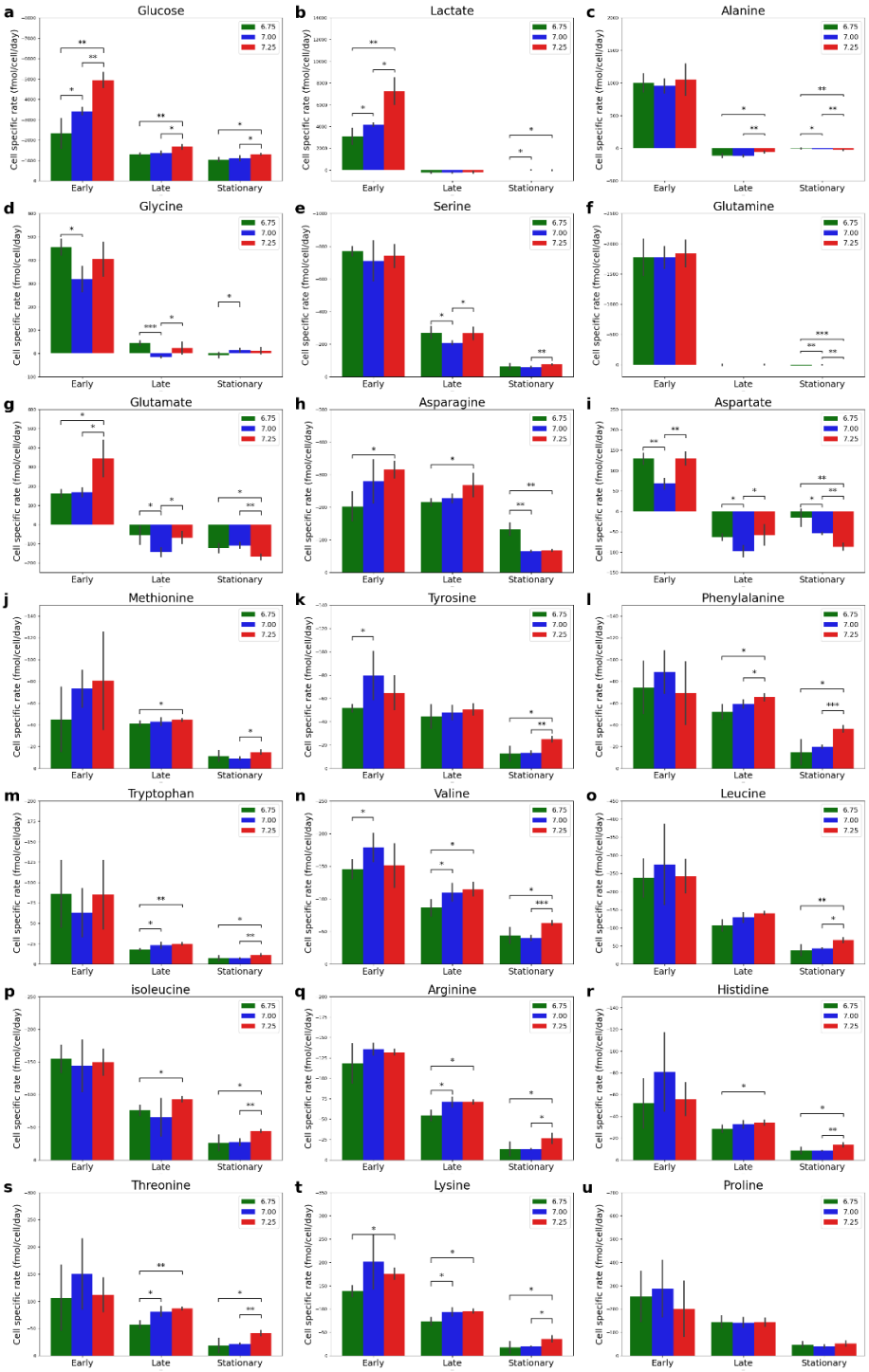

Figure A2: Effect of bioreactor pH and culture duration on metabolic uptake and production rates.

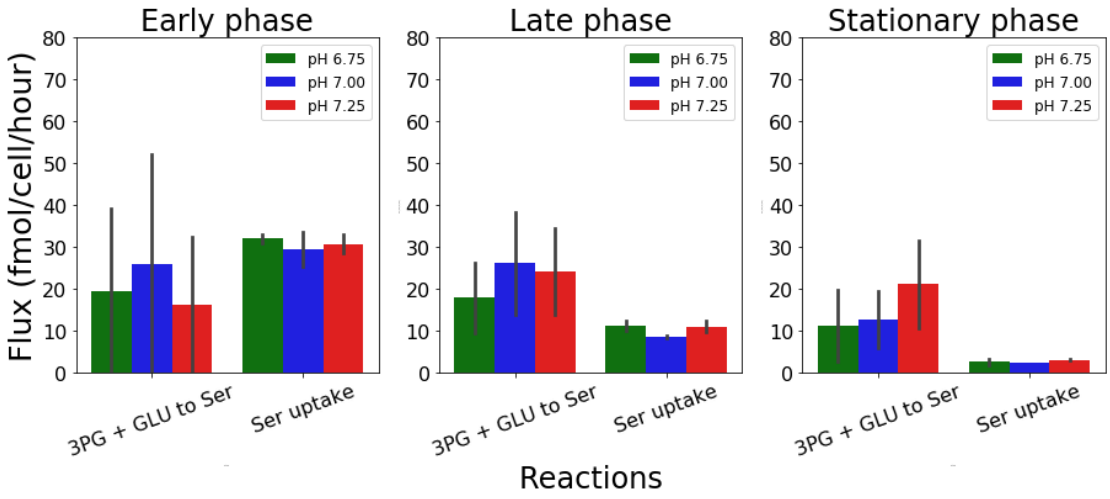

Figure A3: FVA identifies the sources of intracellular serine production during different phases of the culture at the three different pH conditions.

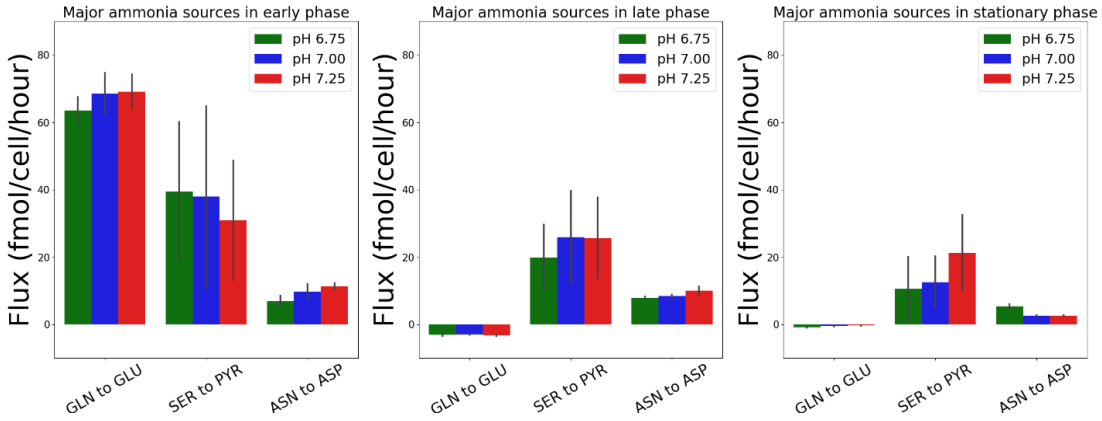

Figure A4: Major sources of intracellular ammonia production during different culture phases at the three different pH conditions.

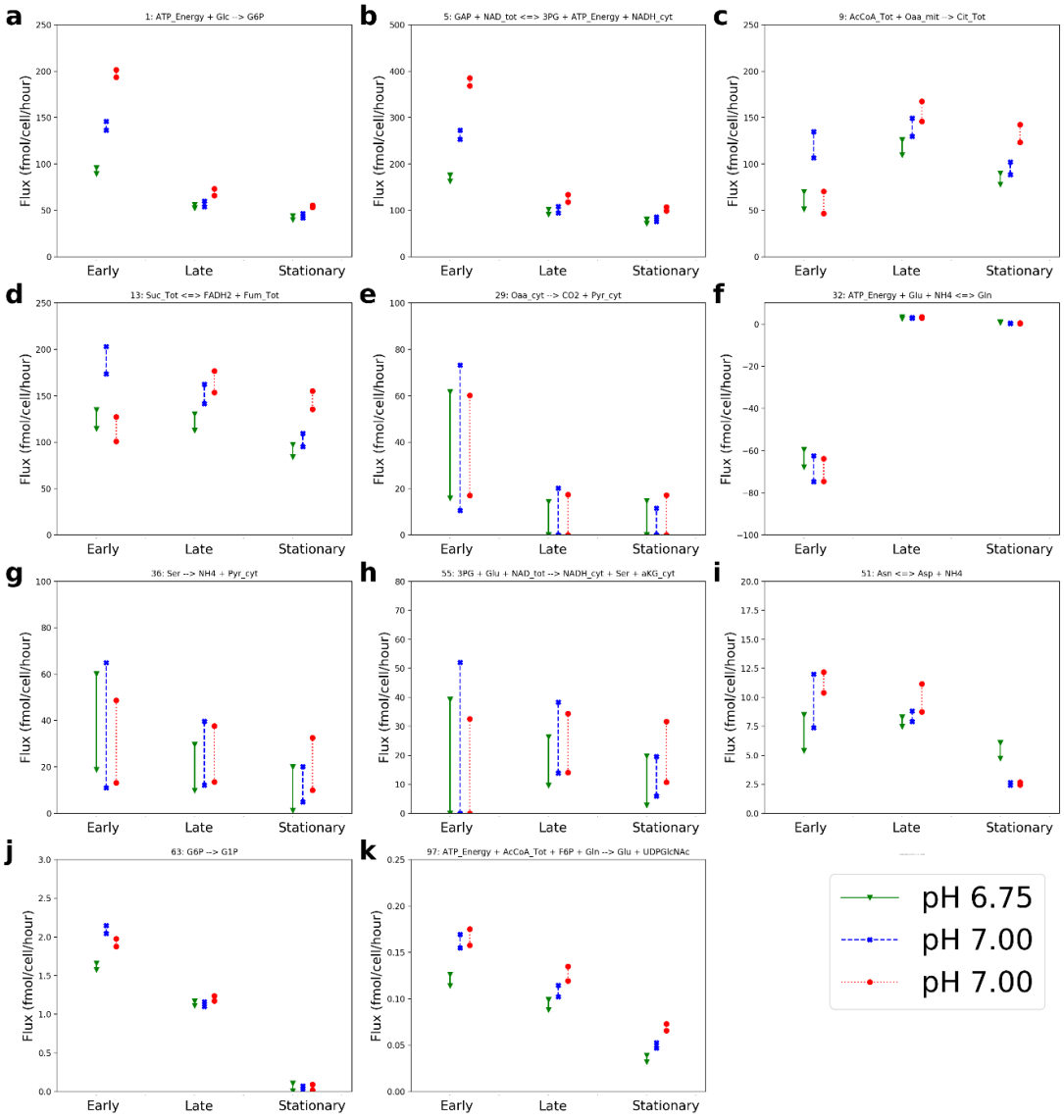

Figure A5: Flux variability analysis (FVA) shows that bioreactor pH and culture duration can impact glycolysis, oxaloacetate to pyruvate fluxes, conversion of serine to pyruvate, biomass precursors, nucleotide sugar synthesis rates and amino acid metabolism (asparagine, aspartate, glutamate, and glutamine).

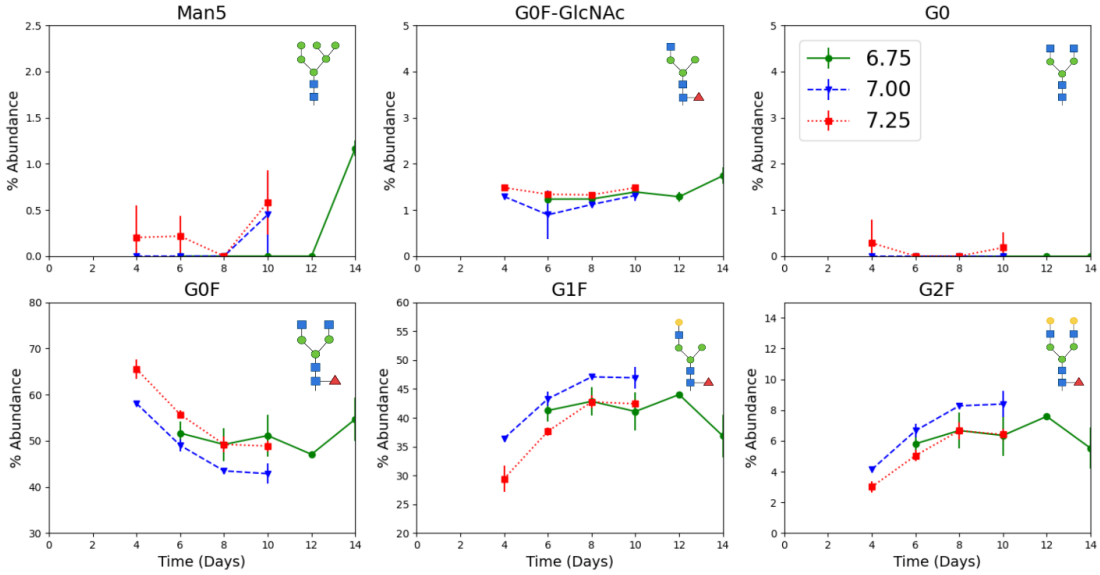

Figure A6: Dynamic N-linked glycosylation data for Fc region for all the detected glycans.

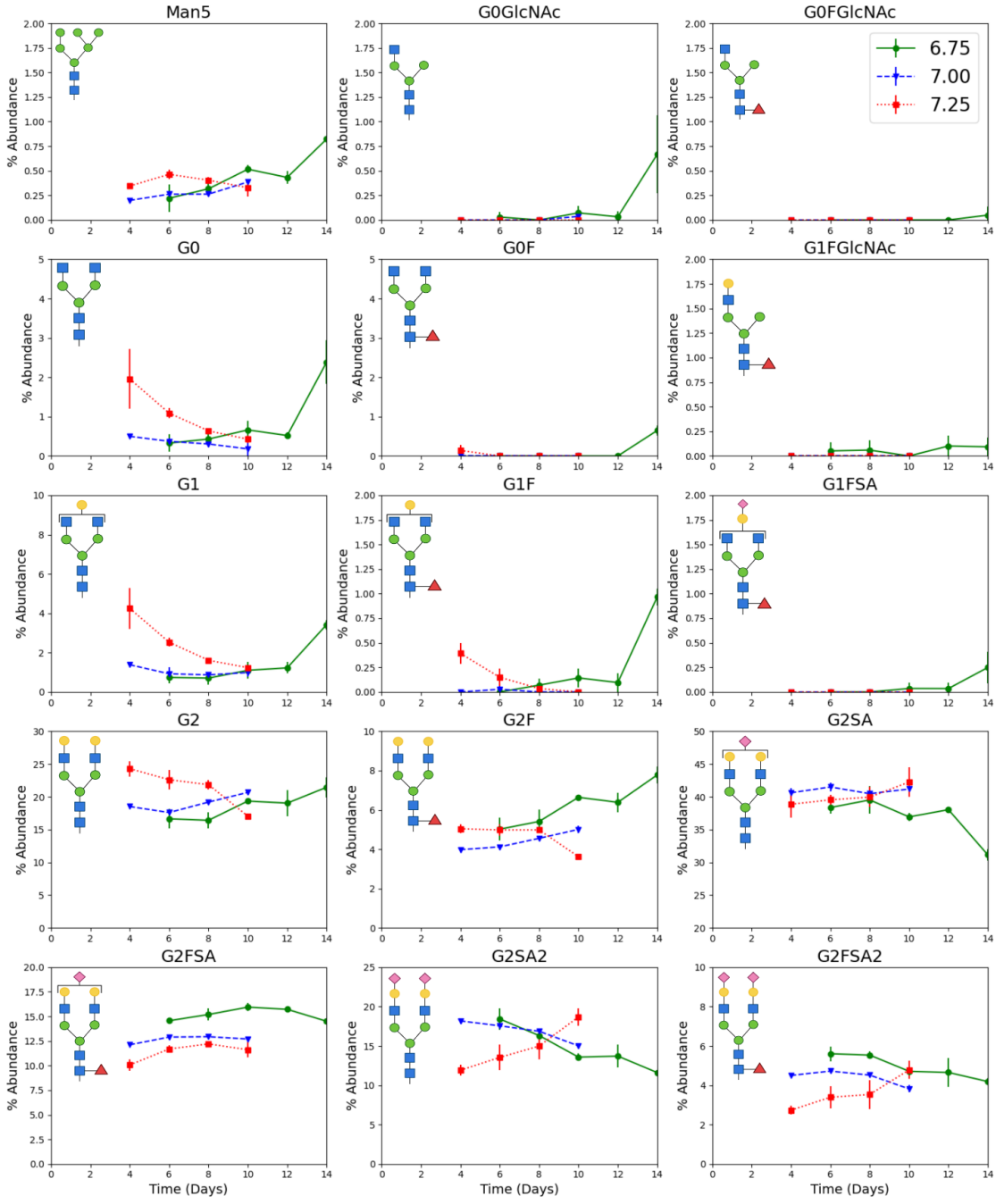

Figure A7: Dynamic N-linked glycosylation profiles for all the glycans detected in the Fab region.

Appendix B

SUPPLEMENTARY MATERIAL TO CHAPTER 3

Table B1: List of primers for qRT-PCR

Target	GeneID	Primer
N-acetylglucosaminyltransferase I (GnT1)	100682529	Forward – GCT TTG GGG TGC TAT CCT CT Reverse – AGG CTG GCA GGG TCA TCA TC
β-1,4-galactosyltransferase 1 (β4GalT1)	100689430	Forward – CTG GAC TTC GGT CCT AGT GC Reverse – GGA ATA ACG ACC GCC CAT CT
β-1,4-galactosyltransferase 2 (B4GalT2)	100689434	Forward – CTG GTC TTG TGG GCC GAG Reverse – TCC CGG TGT CTA AAG GGG AT
β-1,4-galactosyltransferase 3 (B4GalT3)	100689346	Forward – AAC TGC CAT AAT TGT GCC CC Reverse – TGC CAT ATG CAA GCT GCT G
UDP-Gal transporter (SLC35A2)	100689031	Forward – ACA CAC TCA AGC TCG CGG T Reverse – TGT CAC CTG GAA AGT GGC AG
ST3 β-galactoside α-2,3-sialyltransferase 3 (ST3Gal3)	100689187	Forward – CTT GGA AGT TGC ACT TGC TCC Reverse – GCC CAG CCG ATC ATA CTC TG
ST3 β-galactoside α-2,3-sialyltransferase 4 (ST3Gal4)	100689440	Forward – CTT GGC CCT ACA CCT CTG TG Reverse – CTT GGG AGA CAT TGT GCC CT
Fucosyltransferase 8 (FUT8)	100751648	Forward – GCC TGG GGG ACC TTA TTG TT Reverse – AGG GCC TTC TGG TAT TCG GA
GDP-Fuc transporter (SLC35C1)	100689382	Forward – CAT AGG CAC CAT CTT CGG GG Reverse – CAT CAG GGG CAA GAA GAG CA
UDP-GalNAc transporter (SLC35D1)	100751278	Forward – GAT TTT GTG CCA ACC AGC GT Reverse – CCT AGG GCA CCA TCG TGA AA
CMP-Neu5Ac transporter (SLC35A1)	100689322	Forward – ACCCAGGCAAGACACTACCT Reverse – GCTACCATCTGGGGTTGACT
β-actin	100689477	Forward – CCC CAT TGA ACA CGG CAT TG Reverse – TCT TTT CAC GGT TGG CCT TG

Table B2: Enzyme concentrations, forward rate constants and nucleotide sugar dissociation constants.

Enzyme	$[E_t]$ (μM)	k_f (min^{-1})	K_{md} (μM)
Man1	1.04	1924	0
Man2	1.04	1924	0
GnT1	3.03	990	170
GnT2	0.71	1320	960
FucT	1.00	253	46
GalT	0.44	8712	14500
SiaT	1.00	484	57

Table B3: Default nucleotide sugar concentrations.

Nucleotide sugar	Golgi concentration (μM)
UDP-GlcNAc	9143
UDP-Gal	3810
GDP-Fuc	5000
CMP-Neu5Ac	2286

Table B4: Golgi enzyme distributions in fractions

Enzyme	Tank 1	Tank 2	Tank 3	Tank 4
Man1	0.89	0.11	0	0
Man2	0.1	0.87	0.03	0
GnT1	0.49	0.51	0	0
GnT2	0	0.61	0.39	0
FucT	0.04	0.87	0.09	0
GalT	0	0	0.37	0.63
SiaT	0	0	0.23	0.77

Table B5: Glycosylation site and amino acid backbone

Chain	N-linked glycan site	Peptide sequence
Light chain	N70	WPDYNLTISNLE (66 – 77)
Heavy chain	N301	EEQYNSTYRVVS (297 – 308)

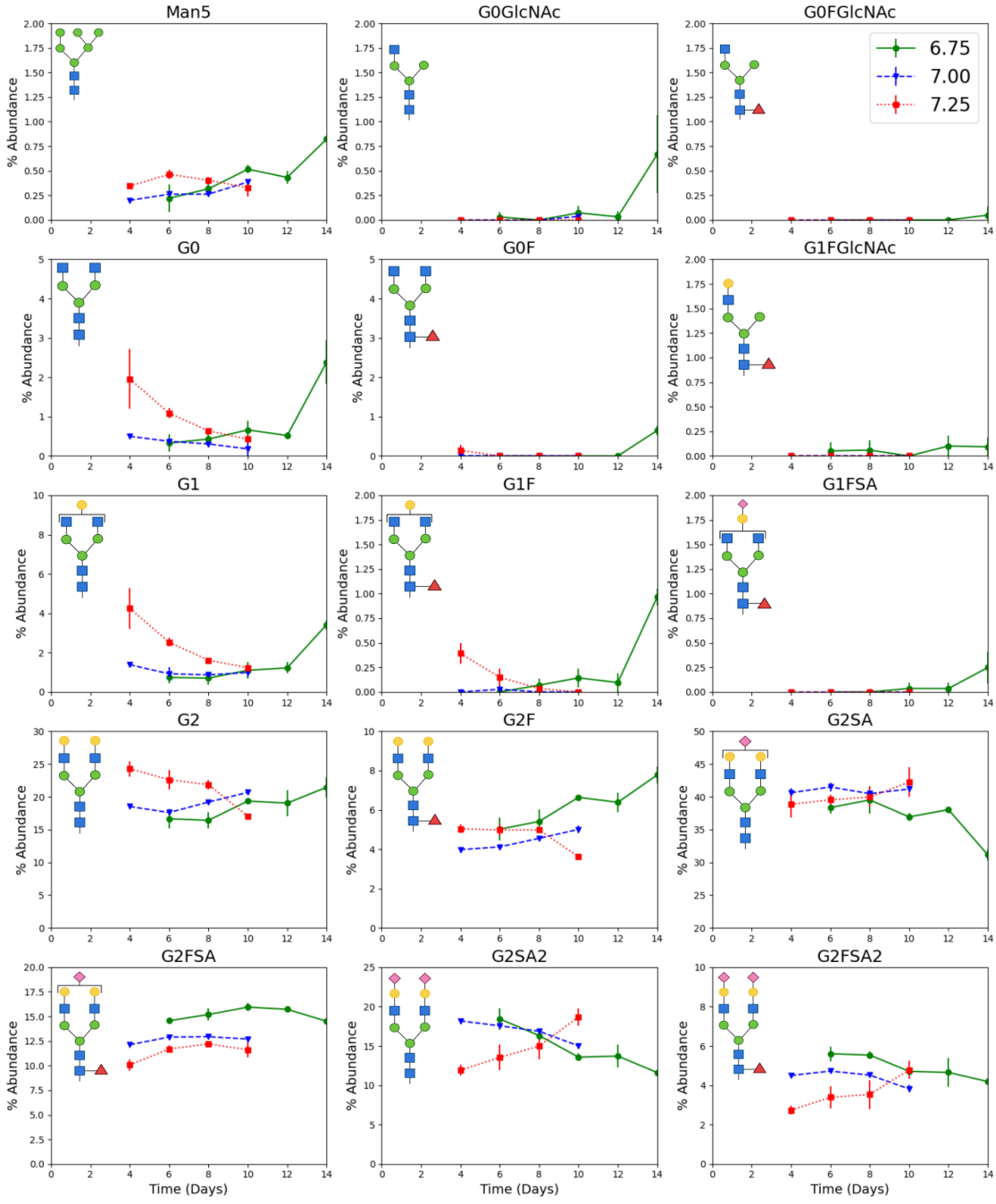

Figure B1: Impact of bioreactor pH on Fab glycan fractions

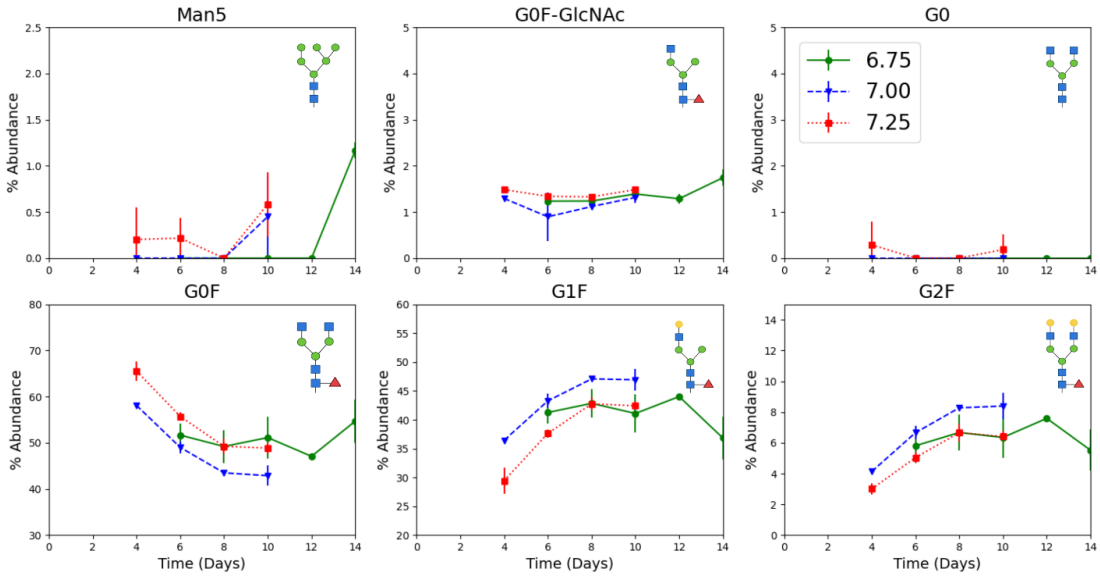

Figure B2: Impact of bioreactor pH on Fc glycan fractions.

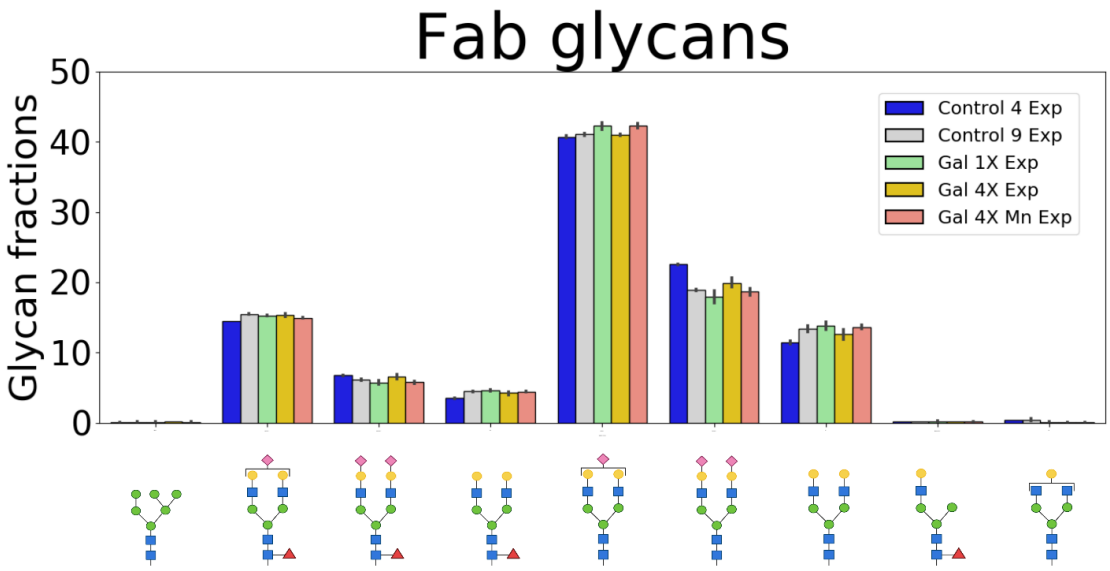

Figure B3: Original Fab glycosylation data from galactose and $MnCl_2$ supplemented experiments.

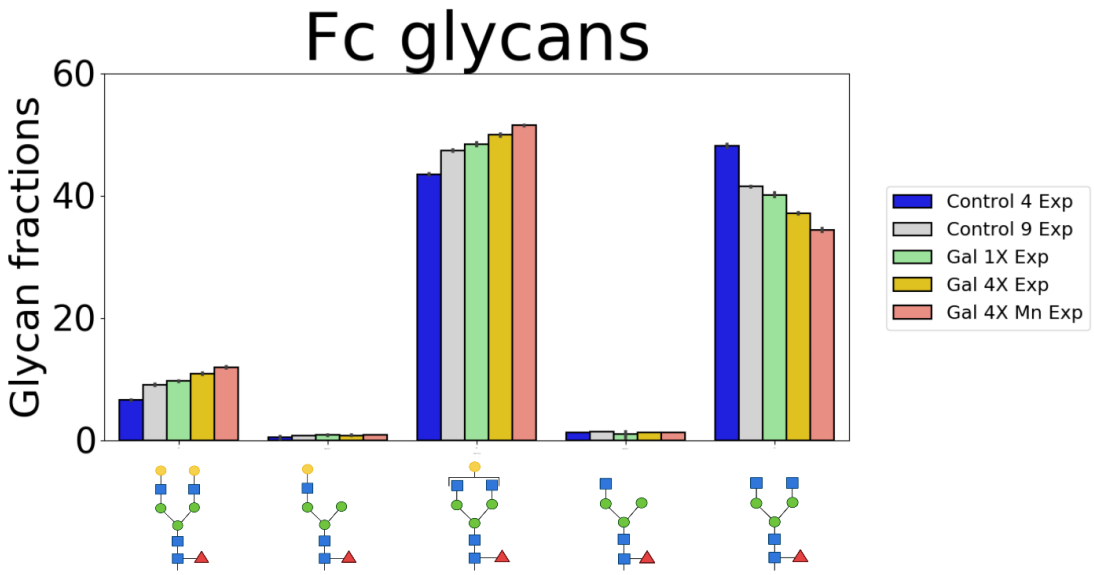

Figure B4: Original Fc glycosylation data from galactose and $MnCl_2$ supplemented experiments.

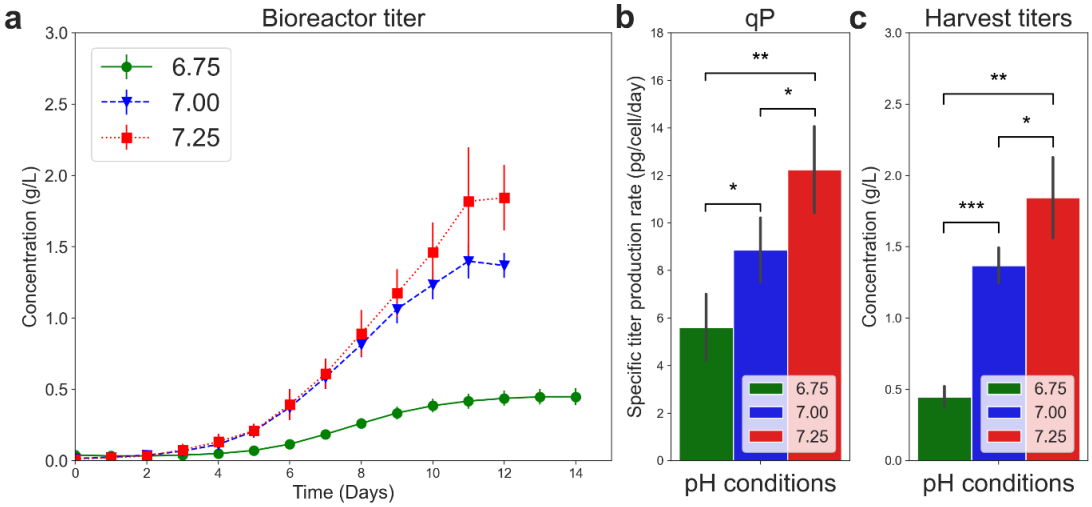

Figure B5: Bioreactor antibody titer data

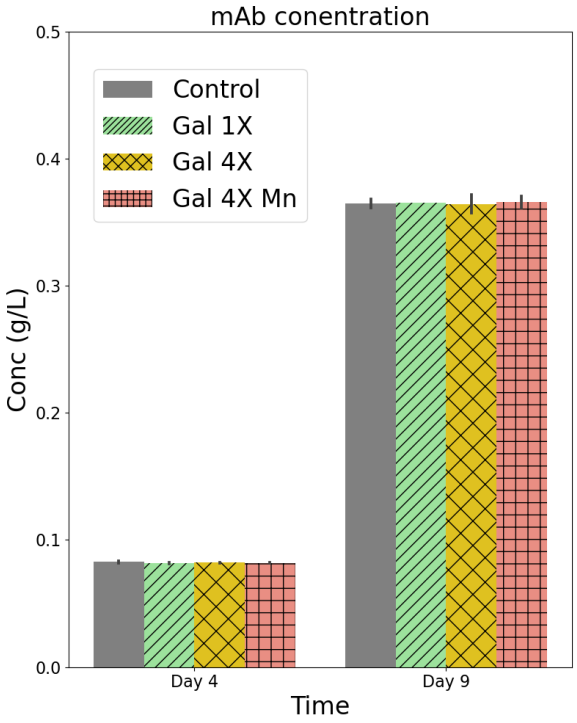

Figure B6: Galactose and $MnCl_2$ addition titer data

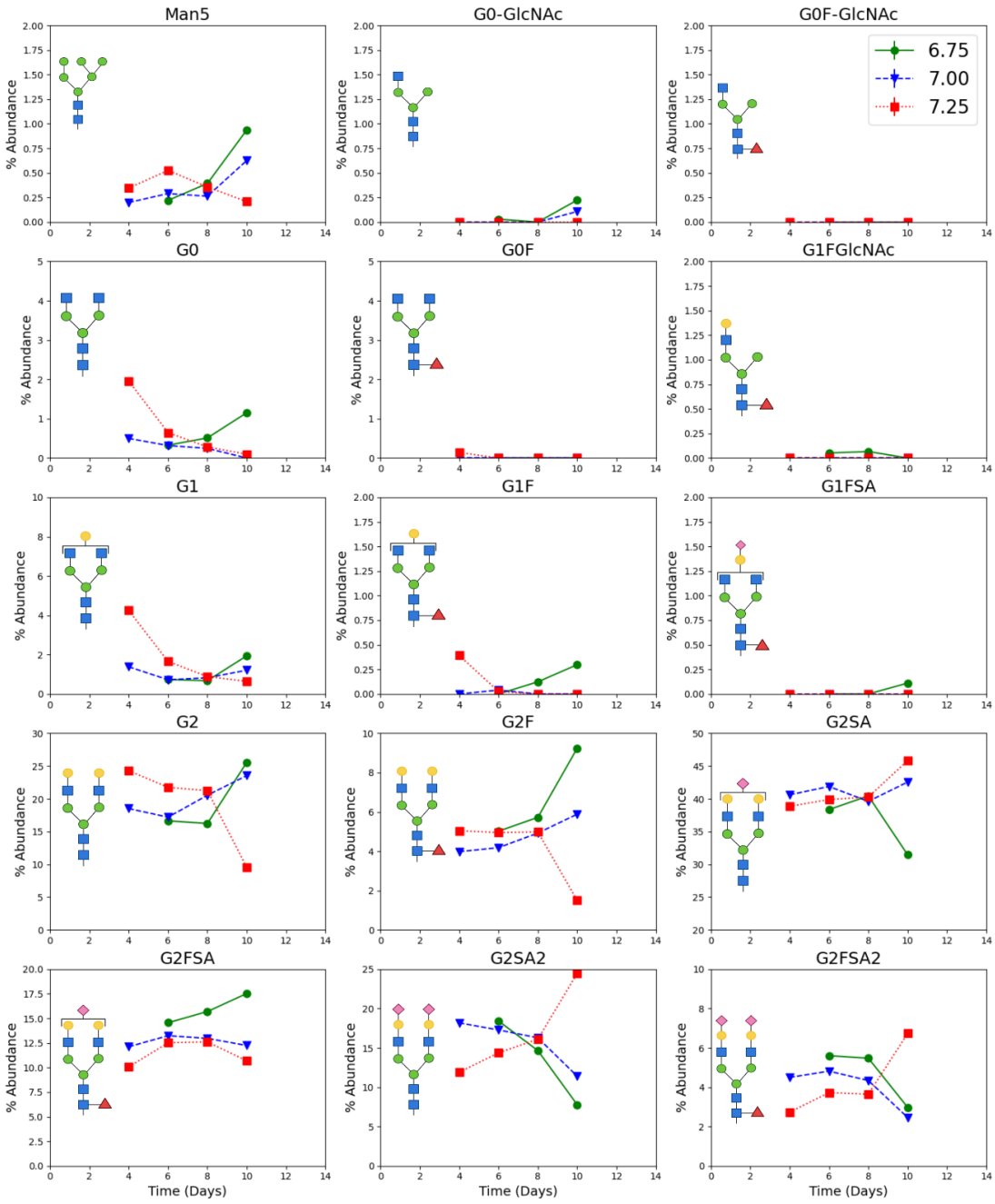

Figure B7: Adjusted glycan fraction bioreactor data for Fab region.

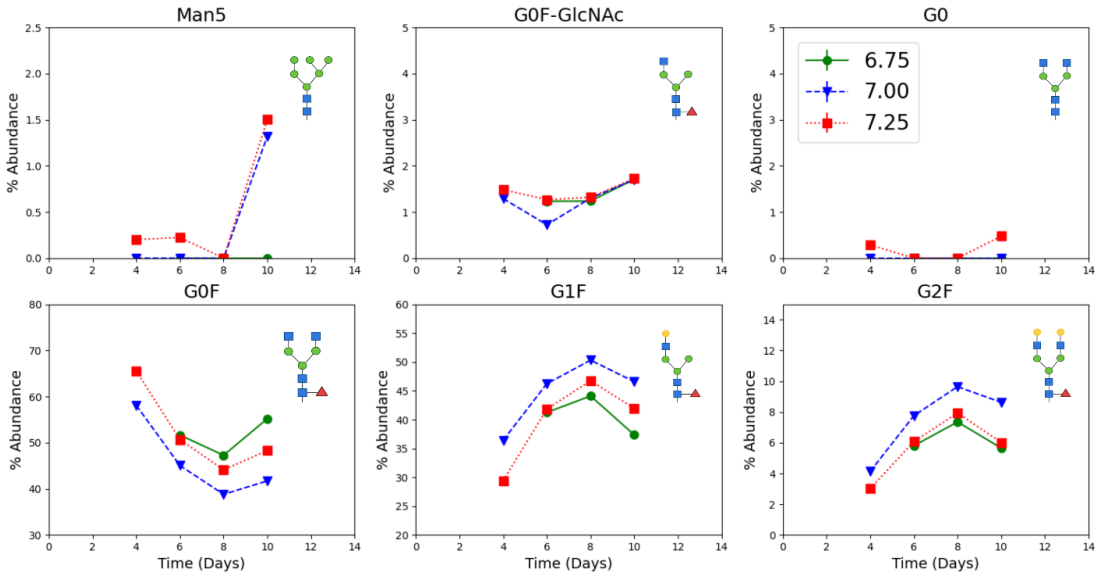

Figure B8: Adjusted glycan fraction bioreactor data for Fc region.

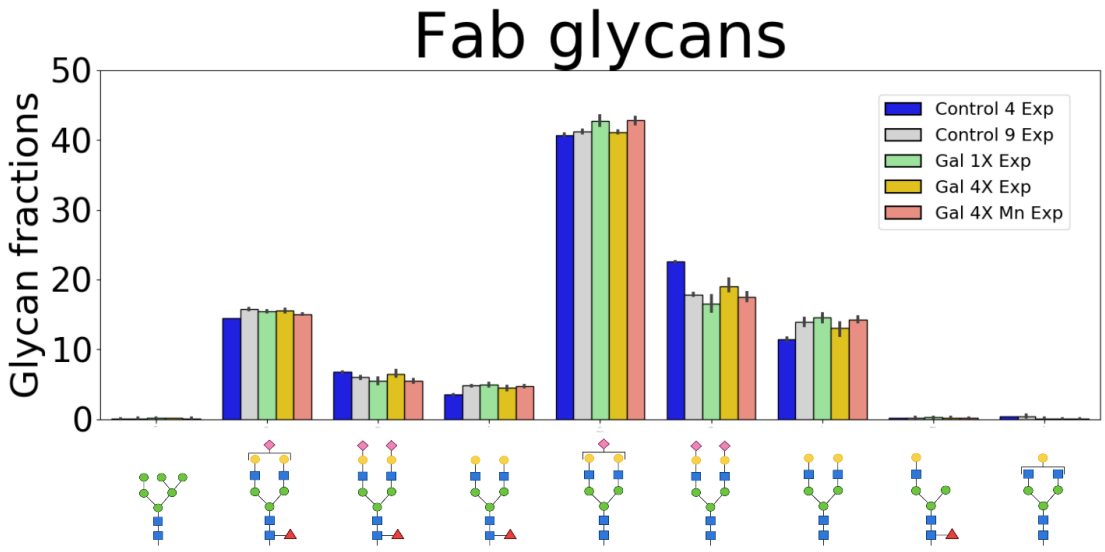

Figure B9: Adjusted glycan fraction galactose and $MnCl_2$ supplementation data for Fab region.

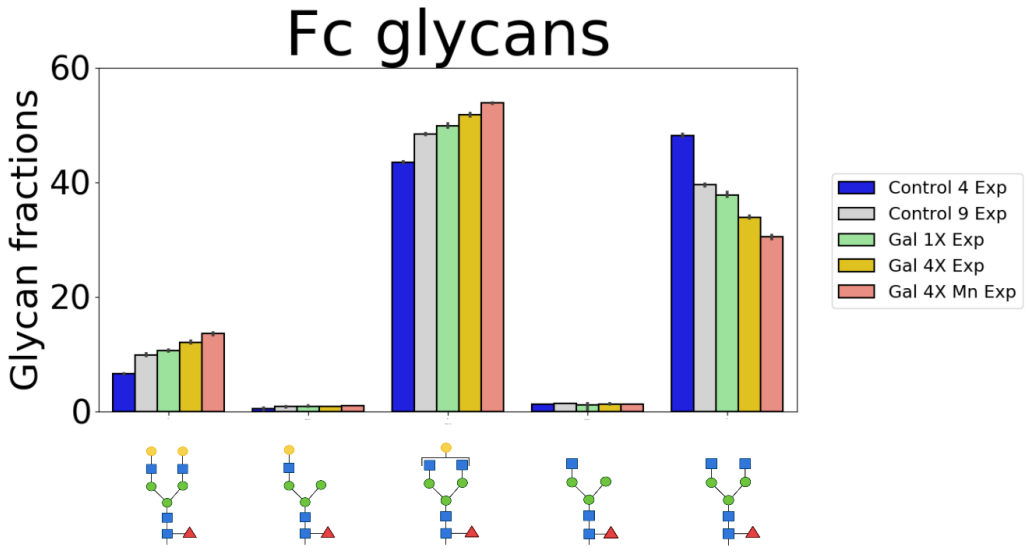

Figure B10: Adjusted glycan fraction galactose and $MnCl_2$ supplementation data for Fc region.

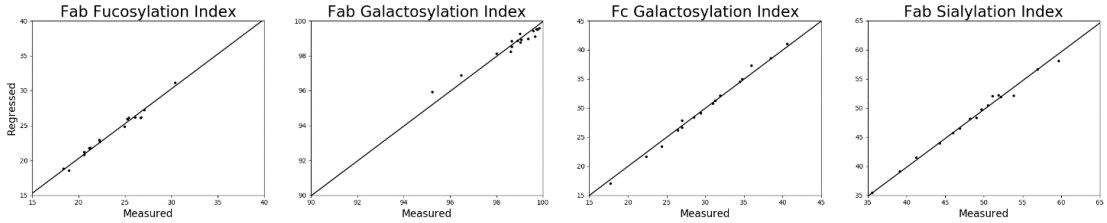

Figure B11: The regressed and experimentally measured data have been plotted to show the results of regression.

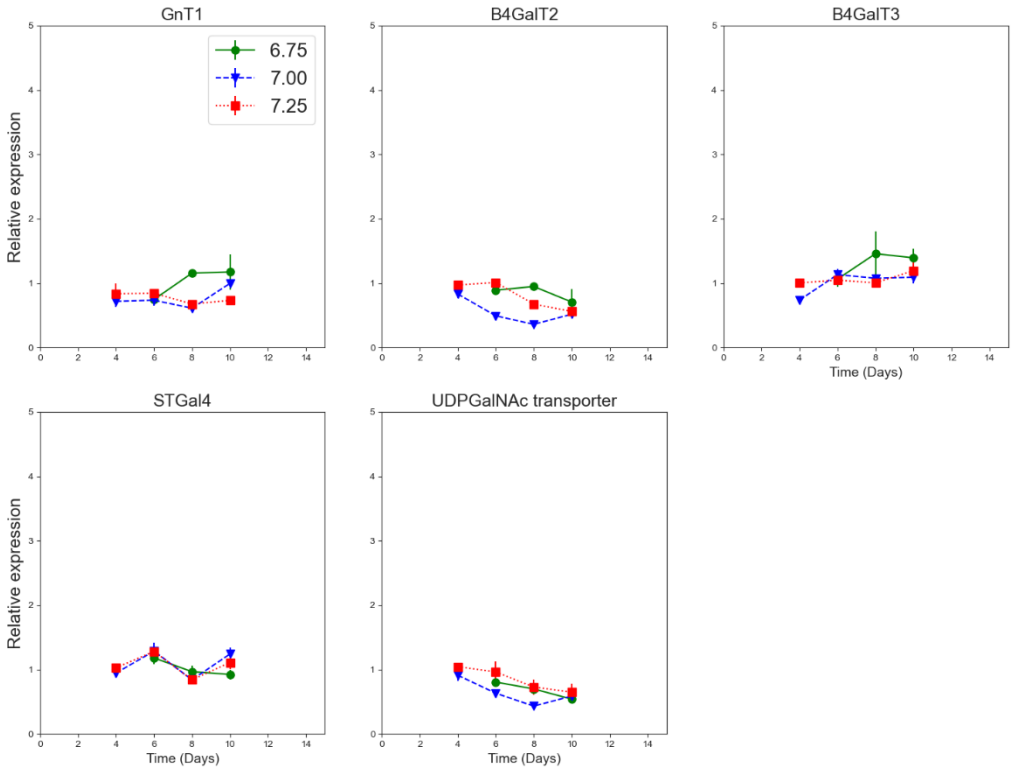

Figure B12: Enzyme expression levels for bioreactor samples.

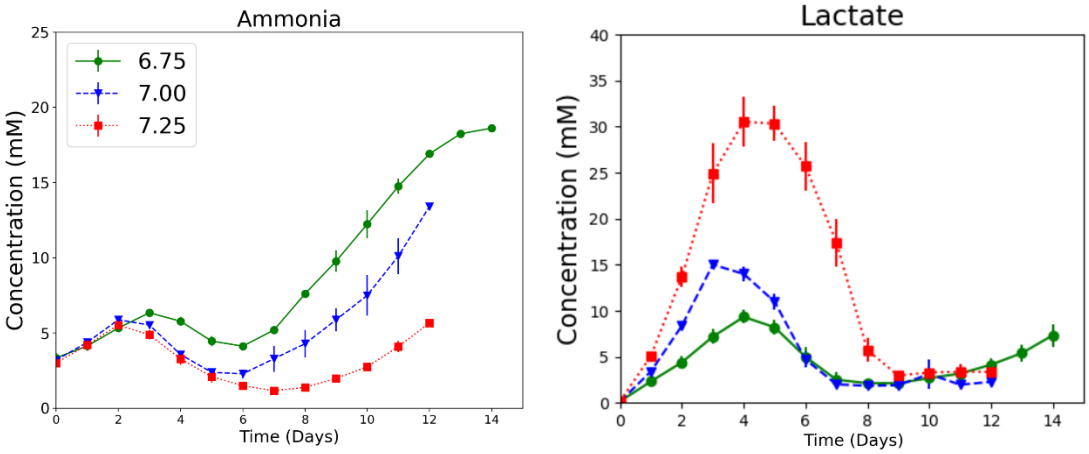

Figure B13: Bioreactor ammonia and lactate concentrations.

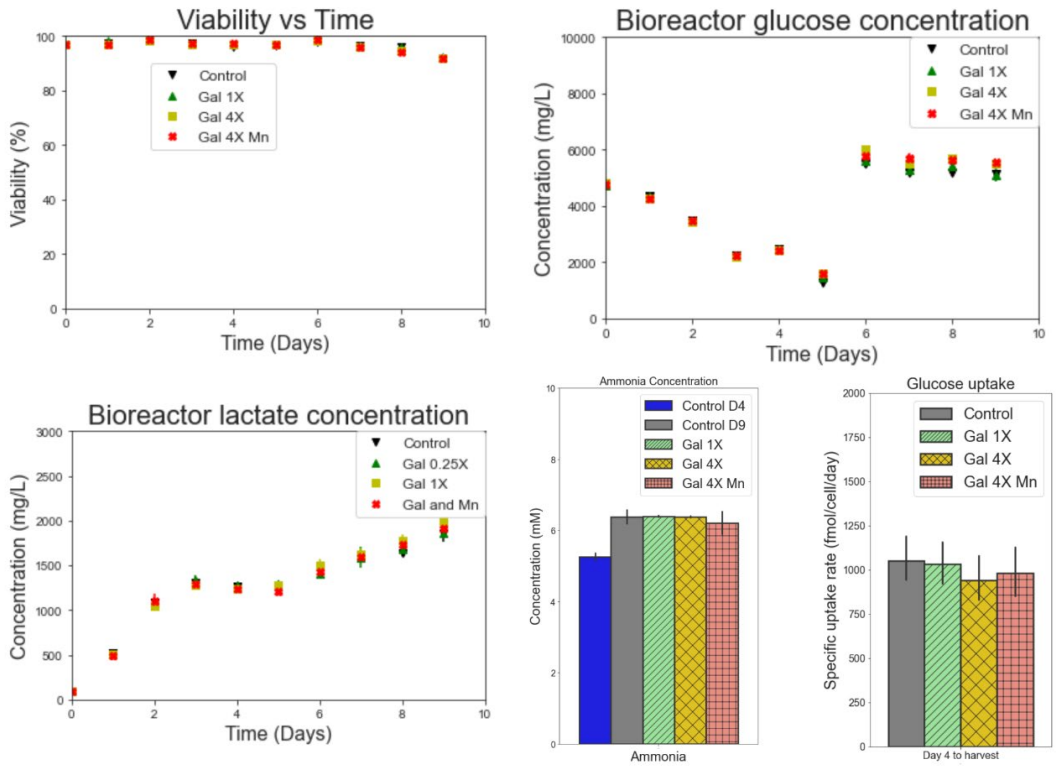

Figure B14: Additional cell culture data from galactose and $MnCl_2$ feeding experiments.

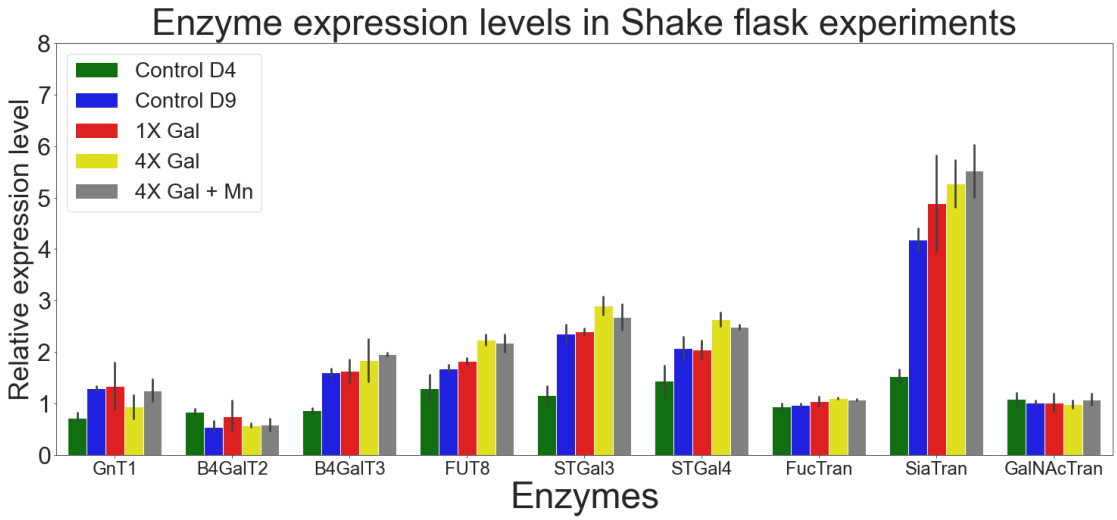

Figure B15: Glycosylation enzyme expression levels from galactose and $MnCl_2$ supplemented cultures.

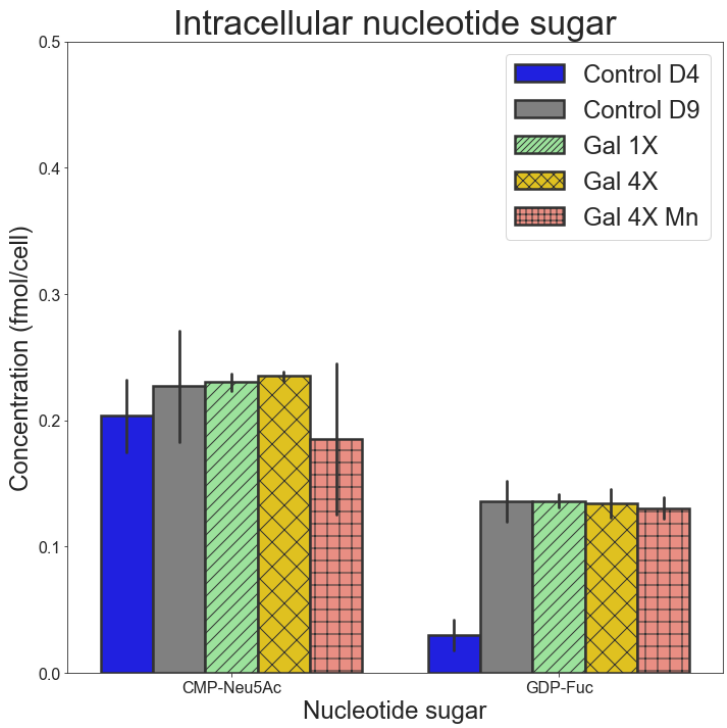

Figure B16: Measurements of nucleotide sugars in galactose and $MnCl_2$ supplementation experiments.

Appendix C

SUPPLEMENTARY MATERIAL TO CHAPTER 4

Table C1: Reaction network

Reaction number	Reaction
1	GLC + ATP -> G6P
2	G6P -> 2 PYR + 2 NADH + 3 ATP
3	PYR + NADH -> LAC
4	PYR + GLU -> ALA + AKG
5	PYR -> AcCoA + CO2 + NADH
6	AcCoA + OXA -> CIT
7	CIT -> CO2 + NADH + AKG
8	AKG -> NADH + CO2 + SucCoA
9	SucCoA -> ATP + Suc
10	Suc -> Fum + FADH2
11	Fum -> Pyr + CO2
12	Fum -> OXA + NADH
13	TRP -> 2 AcCoA + 4 CO2 + 2 NADH + 1 ALA
14	ASN -> NH3 + ASP
15	ASP + AKG -> OXA + GLU
16	PHE + NADH -> TYR
17	TYR + AKG -> GLU + CO2 + Acetoac + FUM
18	SER + NH3 + NADH + CO2 -> 2 GLY
19	LEU + AKG + ATP -> GLU + Acetoac + NADH + FADH2 + AcCoA
20	THR -> GLY + NADH + AcCoA
21	HIS -> GLLU + NH3
22	GLN -> GLU + NH3
23	ARG + AKG -> GLU + NADH

Table C1: Continued

24	GLU -> AKG + NH3 + NADH
25	CYS -> PYR + NH3
26	VAL + ATP + AKG -> GLU + CO2 + NADH + FADH2 + SucCoA
27	ILE + ATP + AKG -> GLU + AcCoA + 2 NADH + FADH2 + SucCoA
28	THR -> NADH + NH3 + Suc
29	MET + SER + ATP -> CYS + NADH + SucCoA + NH3
30	Acetoac + SucCoA -> 2 AcCoA + Suc
31	2 AKG + LYS -> 2 GLU + 2 AcCoA + 3 NADH + 1 FADH2
32	1 NADH2 + 0.5 O2 -> 2.5 ATP
33	1 FADH2 + 0.5 O2 -> 1.5 ATP
34	ARG -> ORN + Urea
35	ORN + NH3 + CO2 + ATP -> CLN
36	ASP + CLN + ATP -> FUM + ARG
37	ORN + AKG -> 2 GLU + NADH
38	G6P + 2 GLU -> 2 SER + 2 AKG + NADH
39	Ser -> PYR + NH3
40	0.01259 ALA + 0.01025 ARG + 0.00705 ASN + 0.01509 ASP + 0.00312 CYS + 0.01505 GLN + 0.00457 GLU + 0.01645 GLY + 0.00308 HIS + 0.00603 ILE + 0.01368 LEU + 0.01154 LYS + 0.00308 MET + 0.00851 PHE + 0.01123 SER + 0.00855 THR + 0.00223 TRP + 0.00467 TYR + 0.00834 VAL + 0.6138 ATP + 0.07733 NADH + 0.05304 AcCoA + 0.01716 G6P -> 0.00812 CO2 + 0.0039 FUM + Biomass
41	0.0096 ALA + 0.0084 ARG + 0.008 ASN + 0.0086 ASP + 0.0056 CYS + 0.0105 GLU + 0.0095 GLN + 0.0142 GLY + 0.0037 HIS + 0.0053 ILE + 0.0148 LEU + 0.0128 LYS + 0.0022 MET + 0.0068 PHE + 0.0232 SER + 0.0152 THR + 0.004 TRP + 0.0087 TYR + 0.0183 VAL + 0.816 ATP -> IgG
42	GLC_{ext} -> GLC
43	LAC_{ext} -> LAC
44	ALA_{ext} -> ALA

Table C1: Continued

45	TRP_{ext} -> TRP
46	$NH3_{ext}$ -> NH3
47	ASN_{ext} -> ASN
48	ASP_{ext} -> ASP
49	PHE_{ext} -> PHE
50	TYR_{ext} -> TYR
51	SER_{ext} -> SER
52	GLY_{ext} -> GLY
53	MET_{ext} -> MET
54	THR_{ext} -> THR
55	ILE_{ext} -> ILE
56	VAL_{ext} -> VAL
57	CYS_{ext} -> CYS
58	GLU_{ext} -> GLU
59	ARG_{ext} -> ARG
60	GLN_{ext} -> GLN
61	HIS_{ext} -> HIS
62	LEU_{ext} -> LEU
63	LYS_{ext} -> LYS
64	O_{2ext} -> O_2
65	CO_{2ext} -> CO_2
66	$Biomass_{ext}$ -> Biomass
67	IgG_{ext} -> IgG
68	Urea -> $Urea_{ext}$
69	ATP -> ATP_{sink}

Table C2: Parameter values

Parameter	Value
m_{GLC}	903
$K_{S,GLC}$	4.15
$K_{S,GLN}$	1e-6
$K_{S,GLU}$	1e-6
$K_{S,ASN}$	0.14
$K_{S,ASP}$	0.55
$kdiauxic_{GLN}$	1
$kdiauxic_{GLU}$	1
$kdiauxic_{ASN}$	1
$kdiauxic_{ASP}$	0.153
$OSMO\_A$	431.5
$OSMO\_B$	0.1
$Y_{\frac{X}{LAC}}$	5
$K_{S,LAC}$	30
$Y_{\frac{ALA}{GLC}}$	0.32
$Y_{X/ALA}$	50
$Y_{\frac{X}{SER},A}$	14
$Y_{\frac{X}{SER},B}$	21.5
$Y_{\frac{GLY}{SER}}$	0.92
$Y_{\frac{X}{GLY}}$	560
$Y_{\frac{X}{GLN}}$	2.6
m_{GLU}	145
$Y_{\frac{GLU}{GLN}}$	0.2
$Y_{\frac{X}{ASN}}$	30
m_{ASN}	105
m_{ASP}	80

Table C2: Continued

$Y_{\frac{X}{MET}}$	121
$Y_{\frac{X}{TYR}}$	105
$Y_{\frac{X}{TRP}}$	195
$Y_{\frac{X}{PHE}}$	80
$Y_{\frac{X}{VAL}}$	40.7
$Y_{\frac{X}{LYS}}$	55
$Y_{\frac{X}{LEU}}$	35
$Y_{\frac{X}{ILE}}$	55
$Y_{\frac{X}{ARG}}$	72
$Y_{\frac{X}{HIS}}$	148
$Y_{\frac{X}{THR}}$	55
$Y_{\frac{AMM}{GLN}}$	0.6
$R_{AMM,1}$	0.05
$P_{1,OSMO}$	431.5
$P_{2,OSMO}$	0.1
$\mu_{death,1}$	0.01
$\mu_{death,2}$	0.4
$P_{1,death,IVCD}$	120
$P_{2,death}$	0.1

Table C3: Parameter values (quadratic function of pH)

Parameter	A	B	C
$Y_{\frac{X}{GLC}}$	-7.05	89.55	275.86
μ_{max}	-3.717	52.478	-184.03
$Y_{\frac{LAC}{GLC}}$	-4	57.8	-206.8
q_{mAb}	-168	3114	-13096
$AA\ scaling\ ratio$	-1.36	18.7	-63.26
$R_{AMM,2}$	-56	634	-1508

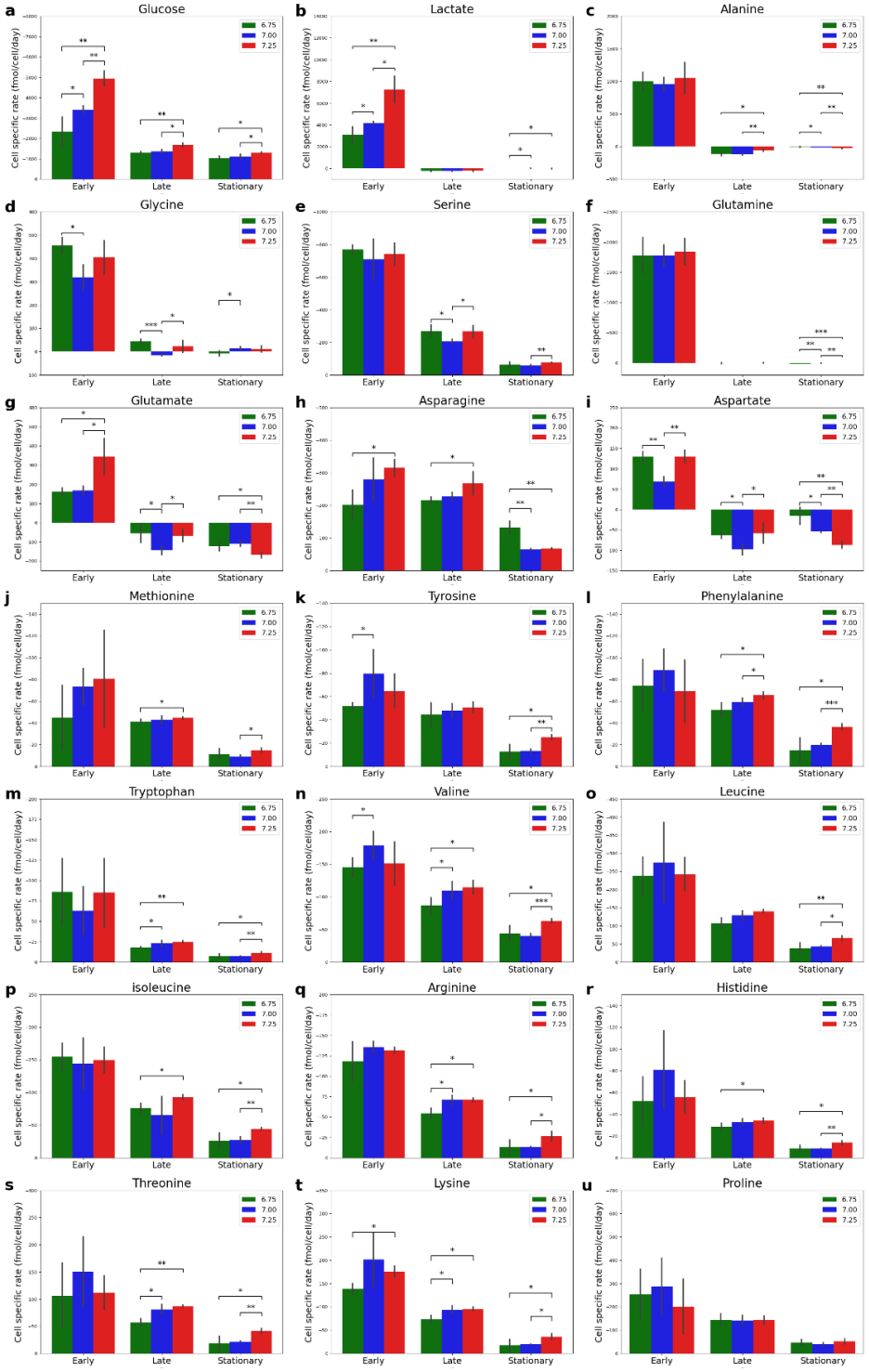

Figure C1: Uptake and secretion rates of measured metabolites

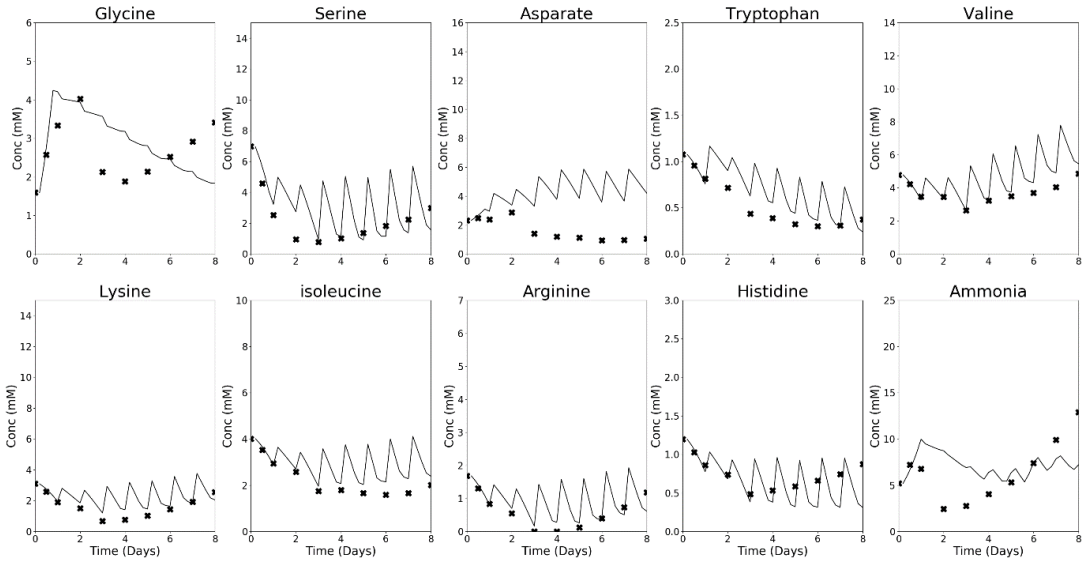

Figure C2: Other amino acid predictions in intensified fed-batch cultures.

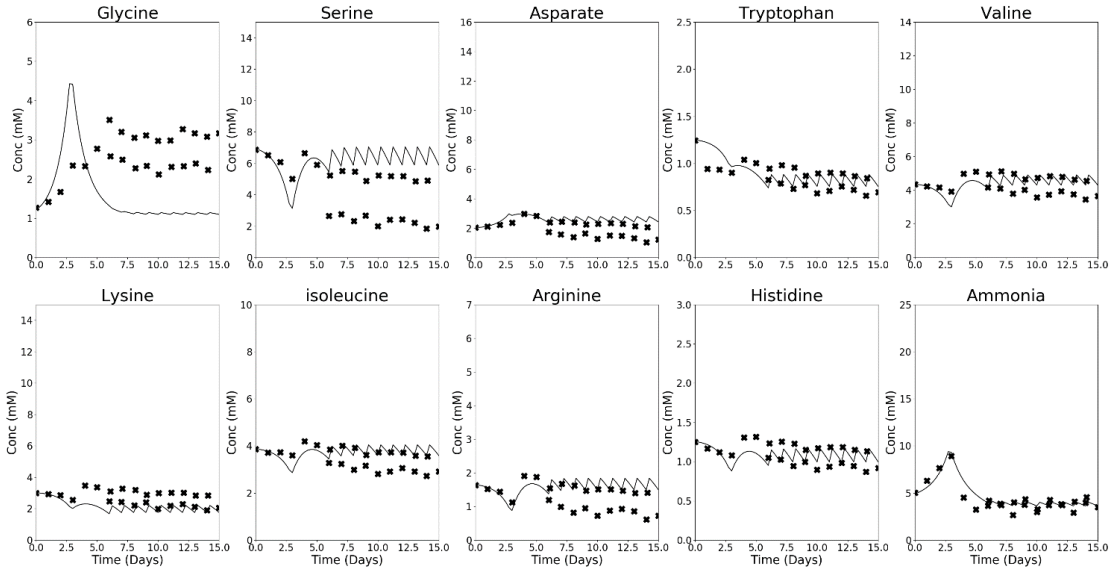

Figure C3: Other amino acid predictions in perfusion cultures.