

**MODELING, ESTIMATION, AND CONTROL OF GLYCOSYLATION IN
MONOCLONAL ANTIBODIES PRODUCED IN CHO CELLS**

by

Devesh Radhakrishnan

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 Engineering

Fall 2016

© 2016 Devesh Radhakrishnan
All Rights Reserved

ProQuest Number: 10243412

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10243412

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

**MODELING, ESTIMATION, AND CONTROL OF GLYCOSYLATION IN
MONOCLONAL ANTIBODIES PRODUCED IN CHO CELLS**

by

Devesh Radhakrishnan

Approved: _____
Abraham M. Lenhoff, Ph.D.
Chair of the Department of Chemical and Biomolecular Engineering

Approved: _____
Babatunde A. Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: _____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

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:

Babatunde A. Ogunnaike, 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:

Anne S. Robinson, 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:

Prasad Dhurjati, 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:

Maciek Antoniewicz, 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

First and foremost, I wish to thank my *gurus*, my advisers, Professors Babatunde Ogunnaike and Anne Robinson, whose untiring efforts have ensured that I grow into an independent researcher. To paraphrase the 15th century Indian mystic, Kabir, “The *guru* diligently cures the flaws of the student like a potter tending to an unbaked pot – with a loving, supporting hand on the inside, while beating it on the outside”. In the last six years, my advisers have molded my mind and character with great care and love, and I owe my personal gratitude to both of them. Professor Ogunnaike accepted me into his group and inspired me with his vision for this project. He granted me the freedom to explore various facets of the research problem and helped me every time I faltered. He provided me with every opportunity to present my research at national and international conferences and encouraged both academic and industrial collaborations that enabled me to gain valuable feedback from my peers. By setting a high bar on the quality of my written work, he taught me to communicate my ideas with greater clarity and I am very grateful to him for his patient and extensive feedback on my work. I consider myself fortunate to have received his guidance during my doctoral studies and I hope to emulate the high research and ethical standards that he has set as a teacher, mentor, and a researcher.

Professor Robinson has been a strong support system for me during my doctoral studies and has always motivated and encouraged me, even when things did not go as planned. Her guidance and advice have been invaluable in the completion of my research work. She has been a kind, considerate, and patient mentor and despite

the geographical challenges of being in different states and time zones, she has always been available to help me with my queries and to discuss research ideas. I am indebted to her for all the help and guidance that I have received from her over the years.

My committee members have been instrumental in enhancing the quality of this dissertation with their critical observations and valuable feedback. Professor Prasad Dhurjati encouraged me to examine my problem statement from a broad perspective, while Professor Maciek Antoniewicz ensured that I didn't lose sight of the nuances involved in this complex research problem. Their questions helped me examine the various facets of my work in great detail and I thank them for their support and guidance. I also wish to thank Professor Michael Betenbaugh for his invaluable inputs to my research work over the past several years.

I am deeply indebted to my colleague and former member of the Ogunnaike research group, Dr. Melissa St.Amand, who patiently trained me in performing cell culture experiments and glycan analysis. She helped ease my transition into the Ogunnaike group and inspired me with her perseverance, diligence, hard work, and dedication. I am thankful to her for all her encouragement and valuable feedback and I continue to be amazed by her creative energy. This project also gave me an opportunity to work with and mentor some of the brightest undergraduate students at UD. In particular, I would like to thank James DeChiara for his help in developing the glycan assay. He bears the mark of a great researcher in the making and I am sure that his enthusiasm for research and his persistence will hold him in good stead as he pursues his future academic endeavors. Kevin Tran, Andrew Bitner, Joanna Adadevoh and Meng Ren have worked on different computational aspects of this project and I thank them for the exciting research interactions that we have shared. I would also like

to thank the visiting scholars in the Ogunnaike group, Dr. Elisabeth Bludau and Dr. Jing Wang, with whom I have had several fruitful discussions on the glycosylation project.

I owe many thanks to past and present members of the Ogunnaike group - Dr. Jacob McGill, Dr. Qian Gou, Dr. Zachary Whiteman, Dr. Chia-Hung Tsai, Dr. James Park, Dr. Robert Lovelett, Dr. Daniel Cook, and Aalap Verma. I couldn't ask for a better cohort of researchers to work with and I am thankful to them for creating such a vibrant and stimulating research environment in the group. The visiting professors in the Ogunnaike group, Dr. Arun Tangirala, and Dr. Peter Verheijen, have been extremely generous in their advice, support, and mentorship and I thank them for their encouragement in my personal and professional activities. I would be remiss not to thank Crystal Maccari, for managing to squeeze me in for an appointment with Dr. Ogunnaike every time I requested, and Lisa Henriksen, for making sure that I never had issues with purchasing equipment and supplies for the lab.

I am also thankful to my colleagues and classmates in the Department of Chemical Engineering at UD. In particular, I would like to thank the members and PIs of the Sullivan group, the Roberts group, the Colby group, and the Kloxin group at Colburn and the Lee group at the Delaware Biotechnology Institute (DBI) for sharing their research facilities. I also wish to thank Leila Choe at DBI for analyzing the glycan samples on the MALDI and for being such a positive person to work with. The Department of Chemical Engineering also provided me with an opportunity to teach the Process Dynamics and Control course that helped me gain a renewed appreciation and understanding of the subject matter. I thank Professor Fraser Russell and Professor Abraham Lenhoff for this opportunity. The staff in Colburn, in particular, Kathie

Young, Chil Alba, and Mary Walsh have been extremely friendly and approachable and I am grateful to them for their numerous favors.

I would not have had the mental and emotional fortitude to face my biggest fears and make it through the last stages of my dissertation work, but for the sessions with my counselor, Dr. Troy Lea at the Center for Counseling and Student Development. I am indebted to him for helping me through some of my toughest moments in graduate school.

On a more personal note, I would like to thank my roommate of four years, Nikodimos Gebreselassie for the kindness, warmth, and cheer that he exuded, and for sharing the delicious Ethiopian food that his mom sent; Srimoyee Dasgupta, Aditi Swarup, and Vinit Choudhary for being so graciously welcoming when I first came to Delaware and extending that love and affection for all these years; Ben Kremkow and Jillian Emerson whose friendship I value and appreciate greatly. I would also like to thank Archana for being a constant friend through the highs and the lows; and Drs. Brian Moreno and Michail Stamatakis, bereft of whose company, I would not have had the opportunity to expand my horizons or discover myself. If these foreign shores felt like home, it was due to the companionship that I experienced with the Batata warriors, especially Dr. Aasma Khan, Dr. Sona Lakshme Balasubramaniam, Dr. Hemanth Akkiraju, and Soma Dash who helped me create some of my happiest memories here in Delaware. Abhinav Jain and Anagha Kulkarni are the family that I have chosen and they have given me incomparable love and affection. They know me better than I know myself and I am happy to have them in my life. My friends from IIT Bombay, Drs. Amit Patil, Omkar Deshmukh, Arijit Sarkar, Meenesh Singh, Harsh Katkar, and Dwaipayan Dasgupta continue to inspire me with their achievements and

successes. Harsh and DD toiled late into the night proofing and suggesting edits to the dissertation and I am thankful for their assistance. Their companionship has made me a better version of myself and I am forever indebted to them.

Finally, I would not have made it this far or achieved any of my dreams had it not been for the unconditional love, support, and prayers from my family. Mahesh, Tejal, and Hitansh fill each day of my life with joy and happiness. My parents, Dr. Naliyat Kavungal Radhakrishnan and Mrs. Chandrika Radhakrishnan, are my strength, my hope, my dreams, my joy, my love, and my life. My success stems from the boundless love they have given me, the limitless blessings they shower upon me, and the countless sacrifices that they have made for me. I am forever grateful to them for everything that they have done for me. Thank you.

TABLE OF CONTENTS

LIST OF TABLES	xiv
LIST OF FIGURES	xvi
ABSTRACT	xix

Chapter

1	INTRODUCTION	1
1.1	Background.....	1
1.2	Motivation – Quality Considerations in the Biopharmaceutical Industry.....	4
1.3	Proposed Framework for Glycosylation Control	6
1.4	Monoclonal Antibodies – An Overview.....	7
1.4.1	Antibody Structure and Function	7
1.4.2	Antibody Expression Systems	10
1.4.3	Quality Attributes of Monoclonal Antibodies	13
1.5	Glycosylation – An Overview	15
1.5.1	Glycosylation Pathway and Glycan Biosynthesis	15
1.5.2	Effects of Glycosylation on Protein Quality	18
1.5.3	Factors Affecting Glycosylation.....	20
1.6	Dissertation Overview	22
2	MULTI-SCALE MODEL OF ANTIBODY GLYCOSYLATION	24
2.1	Introduction	24
2.2	Mathematical Models for Biological Processes	26
2.2.1	Modeling Cell Growth in Mammalian Cell Culture.....	26
2.2.2	Modeling Glycosylation	28
2.2.3	Modeling Cell Growth and Antibody Glycosylation	34
2.3	Development of an In-house Multi-scale Model of Glycosylation	36

2.3.1	Batch Experiments – Materials and Methods.....	37
2.3.2	Macro-scale Model Development	38
2.3.3	Coupling the Macro-scale Model with the Micro-scale Model...	47
2.3.4	Micro-scale Model Development	48
2.4	Validation using Fed-batch Model with Pulse Feeding	54
2.4.1	Fed-batch Experiments – Materials and Methods	54
2.4.2	Modifications to Multi-scale Model	55
2.5	Summary and Conclusions	59
3	CONTROLLING THE GLYCOSYLATION PROFILE USING TIME-DEPENDENT MEDIA SUPPLEMENTATION	61
3.1	Introduction	61
3.2	Materials and Methods	64
3.2.1	Cell Culture	64
3.2.2	Experimental Design	65
3.2.3	Glycan Permethylated Assay	66
3.2.4	Glycosylation Index.....	68
3.2.5	Controllability Analysis.....	69
3.3	Results	70
3.3.1	Early Addition of EDTA is Detrimental to Cell Growth and Reduces Antibody Titer.....	70
3.3.2	Early Addition of MnCl ₂ Alters the Glycan Distribution Significantly	73
3.3.3	The Type of Media Supplement and the Time of Addition both showed Statistically Significant Effects on Glycan Distribution..	77
3.4	Discussion.....	79
3.5	Summary and Conclusions	98
4	A CONTROLLER DESIGN FRAMEWORK FOR CONTROLLING THE GLYCOSYLATION PROFILE	101
4.1	Introduction	101
4.2	Materials and Methods	103
4.2.1	Computational Model of Glycosylation	103
4.2.2	Cell Culture	104

4.2.3	Glycan Permethylation Assay	105
4.2.4	Statistical Design of Experiments	106
4.2.5	Controllability Analysis.....	109
4.2.6	Controller Design under Nominal Conditions.....	110
4.2.7	Controller Design under Model-Plant Mismatch Conditions....	112
4.3	Results and Discussions	114
4.3.1	Controllability Analysis.....	114
4.3.2	Set-point Tracking under Nominal Conditions	126
4.3.3	Set-point Tracking under Model-Plant Mismatch Conditions ..	129
4.4	Summary and Conclusions	133
5	GLYCOSYLATION ANALYSIS: OBSERVABILITY AND STATE ESTIMATION.....	136
5.1	Introduction	136
5.2	Glycan and Glycoproteomic Analysis – An Overview	138
5.2.1	In-house Glycan Assay Development	144
5.3	Designing Novel Glycan Assays using Observability Analysis.....	145
5.3.1	Selecting Measurement Variables for Analysis	150
5.3.2	Results from Observability Analysis.....	153
5.4	Estimating Glycan States.....	158
5.4.1	Overview of State Estimation.....	159
5.4.2	System and Model Description	162
5.4.3	Algorithm for Designing the State Estimator	165
5.4.4	Estimating Glycan States using a Constant, Time-invariant State Estimator.....	168
5.5	Summary and Conclusions	173
6	SUMMARY, CONCLUSIONS, AND FUTURE WORK	174
6.1	Summary and Conclusions	174
6.2	Future Work.....	178
6.2.1	Expanding the Scope of the Multi-scale Model	178
6.2.2	Multi-attribute Analytics and Control	181

6.2.3	Validating the State-estimator and Implementing On-line Glycosylation Control	183
REFERENCES		185
Appendix		
A	EXPERIMENTALLY OBSERVED GLYCAN SPECIES	203
B	MIXED LEVEL DESIGN OF EXPERIMENTS	205

LIST OF TABLES

Table 1.1:	List of approved monoclonal antibodies expressed in CHO cells.....	12
Table 1.2:	List of enzymes participating in glycosylation reactions	17
Table 2.1:	Parameter values used in the macro-scale model	44
Table 2.2:	Reaction rules for generating the glycosylation reaction network based on KB2005 model	49
Table 2.3:	Kinetic mechanisms for different glycosylation enzymes	50
Table 2.4:	Parameters to estimate the distribution of different glycosylation enzymes along the length of the Golgi apparatus.....	52
Table 2.5:	Kinetic constants used in the rate expressions for different glycosyltransferase enzymes arranged as per the reaction rule.....	52
Table 3.1:	Experimental conditions tested in mixed factorial design.....	68
Table 3.2:	Glycosylation index (GI) and fucosylation index (FI) for each experimental condition	90
Table 3.3:	Singular values obtained from the SVD analysis of the matrix of significant coefficients	92
Table 4.1:	Plackett-Burman experimental design with high and low concentrations of different glycosyltransferase enzymes in the dynamic mathematical model for glycosylation	107
Table 4.2:	Plackett-Burman experimental design with high and low concentrations of individual amino acids added to the cell culture media	108
Table 4.3:	Singular values obtained from singular value decomposition of the matrix of significant factor coefficients with (a) glycosyltransferase enzyme coefficients as inputs; and (b) amino acid supplements as inputs	119

Table 5.1:	List of new measurement variables and their definitions.....	151
Table 5.1:	Measurement delay and sampling frequencies for different measurements	162

LIST OF FIGURES

Figure 1.1:	A representative image of the antibody structure.....	8
Figure 1.2:	Distribution of approved mAbs by cell line used.....	11
Figure 1.3:	Representative examples of different types of glycan isoforms that contribute to antibody microheterogeneity.....	18
Figure 2.1:	Overview of the in-house multi-scale model.	36
Figure 2.2:	Nutrient and by-product concentration profiles for batch culture.	45
Figure 2.3:	Viable cell density and total cell density profiles for batch culture..	46
Figure 2.4:	Antibody concentration profile for batch culture.	46
Figure 2.5:	Glycan distribution profile obtained from the multi-scale model.	54
Figure 2.6:	Nutrient and by-product concentration profiles for fed-batch culture with daily 5 mL pulse feeding..	56
Figure 2.7:	Viable cell density and total cell density profiles for fed-batch culture with daily 5 mL pulse feeding.	57
Figure 2.8:	Antibody concentration profile for fed-batch culture with 5 mL daily pulse feeding.....	58
Figure 2.9:	Glycan distribution profile obtained from the multi-scale model for fed-batch case.	58
Figure 3.1:	Viable cell concentration data for CHO-K1 cells under different media supplementation conditions	72
Figure 3.2:	Average antibody titer for the 16 experimental conditions.....	73
Figure 3.3:	Average relative glycan percentage of IgG1 glycans produced in CHO-K1 cells under different experimental conditions.....	75

Figure 3.4:	Heat map of significant factor coefficients ($\alpha = 0.05$) obtained from ANOVA.	78
Figure 3.5:	Comparing fractional differences in the glycoforms for different experimental cases.....	86
Figure 3.6:	Comparison of experimental data and model fit for the glycan distribution profile obtained from (a) control flask; and (b) when MnCl_2 is added on D0.	88
Figure 3.7:	Graphical representation of the coefficients associated with the first five input and output modes (with $\sigma_i \geq \sigma^*=0.5$) obtained from controllability analysis.	96
Figure 4.1:	Control block diagram indicating different components of the glycosylation control scheme.	114
Figure 4.2:	Heat map of significant factor coefficients obtained by analysis of variance of the glycan distribution data.	116
Figure 4.3:	Graphical representation of the coefficients associated with the first four input and output modes with glycosyltransferase enzymes as the input factors..	122
Figure 4.4:	Graphical representation of the coefficients associated with the first four input and output modes with glycosyltransferase enzymes as the input factors.	125
Figure 4.5:	Controller performance under nominal conditions.	128
Figure 4.6:	P-type and PI-type controller performance under model-plant mismatch conditions.....	132
Figure 5.1:	Different analytical techniques for glycoprotein characterization..	139
Figure 5.2:	Inference maps based on SCC decomposition of the glycan reaction network obtained from DK2011 model.....	149
Figure 5.3:	Experimental and predicted value of the glycan distribution using observability analysis with four measurements.....	154
Figure 5.4:	Comparing experimental and predicted values obtained from glycosylation analysis with four, five, and six measurements.	155

Figure 5.5: Root mean squared error (RMSE) of the glycan prediction for different scenarios plotted against the corresponding condition number of the measurement matrix..	156
Figure 5.6: Root mean squared error (RMSE) for glycan predictions with each of the 39 cases tested plotted against the corresponding condition number.	157
Figure 5.7: General structure of a state estimator	160
Figure 5.8: Flowchart depicting state estimation algorithm	167
Figure 5.9: State estimator performance under noise-free conditions.	170
Figure 5.10: State estimator performance in the presence of model and measurement noise.	172
Figure 6.1: Schematic representation of systematic model development.	179
Figure 6.2: Thermal unfolding from DSC measurements for different glycan isoforms.	182
Figure 6.3: Strategy for on-line control of glycosylation.	184

ABSTRACT

Monoclonal antibodies (mAbs) are a class of commercially valuable biopharmaceuticals that are used for treating diseases such as psoriasis, rheumatoid arthritis, and multiple types of cancer. A vast majority of these biotherapeutics are expressed in mammalian cell lines such as Chinese Hamster Ovary (CHO) cells to enable post-translational modifications that generate human-like protein structures. One such post-translational modification that results in structural and pharmacological changes in the protein is N-linked glycosylation, involving the addition and subsequent modification of an oligosaccharide to the protein backbone. The non-template driven, enzymatic modification of the attached oligosaccharide yields a heterogeneous distribution of glycan isoforms, altering the immunogenicity, stability and half-life of the mAb, and hence the final drug product quality. Maintaining the desired product quality of mAbs in the presence of process variations during manufacturing has been difficult for a variety of reasons, including: (i) a lack of quantitative understanding of the effect of input factors on product quality attributes; (ii) the absence of on-line or real-time measurements of quality attributes as these are monitored infrequently or using time-consuming assays; (iii) the lack of effective control strategies that incorporate these infrequent measurements (as and when they become available) to regulate product quality. To ensure product safety and therapeutic efficacy, regulatory agencies are encouraging manufacturers to monitor and control the drug product quality, specifically maintaining the glycan distribution within an acceptable range. The overall goal of this dissertation, therefore, is to

develop a rational framework to *model* the effect of different input factors on the glycosylation profile, *estimate* the glycan distribution using a dynamic mathematical model supplemented with infrequent measurements, and *control* the final glycosylation profile in monoclonal antibodies produced in CHO cells.

As the glycosylation profile in mAbs is influenced by several process variables spanning multiple scales – from operating conditions at the bioreactor (macro) scale, to factors at cellular (meso) scale and organelle (micro) scale – we developed an integrated multi-scale model of glycosylation and validated the model predictions using experimental results obtained with an in-house cell line. The model serves as a useful link between nutrient concentrations and cell growth at the macro-scale and the glycosylation profile at the micro-scale.

In parallel, we used a holistic approach that combined factorial design of experiments and a novel computational technique to identify the various combinations of glycan species that are affected by dynamic media supplementation and to quantify mathematically how they are affected. Our experiments demonstrated the importance of taking into consideration the time of addition of trace media supplements, not just their concentrations, and the corresponding mathematical analysis provided insight into what supplements to add, when, and how much, in order to induce specific changes in the glycosylation profile.

We developed a two-step framework to control the glycosylation profile by first generating quantitative input-output relationships using the previously described holistic approach and then designing proportional (P) and proportional integral (PI) controllers based on this quantitative input-output relationship. The set-point tracking performance of these P and PI controllers was evaluated via simulations under

nominal conditions (i.e. when the model is assumed to be representative of the actual ‘plant’ or process) and model-plant mismatch conditions. Our results demonstrated that the developed framework can be implemented to design glycosylation controllers to achieve a desired target glycosylation profile under different conditions.

The P and PI controllers that we have developed are suited for batch-to-batch control as they depend on the final glycosylation profile. To achieve real-time control of glycosylation we require real-time information of the glycan distribution obtained from glycan assays; however, current glycan assays are infrequent and characterized by long analysis times. We address this limitation in glycosylation analysis using two approaches: (i) by formulating a rational framework based on observability analysis to guide the development of novel assays that can simplify glycan analysis or reduce analysis time; and (ii) by designing a state estimator to predict the glycan distribution profile in the absence of measurements using the previously developed multi-scale model and updating those predictions as and when measurements become available.

The framework developed in this dissertation will form the basis of an online control scheme to control the final glycosylation profile in the product, thereby achieving consistent product quality.

Chapter 1

INTRODUCTION

1.1 Background

The global market for pharmaceuticals is predicted to grow to \$1.2 trillion by 2017, with nearly 20% of the market share being dominated by biologics such as monoclonal antibodies (mAbs), hormones, and therapeutic enzymes. Of these, monoclonal antibodies have steadily increased their market dominance – with US sales rising from \$8.29 billion in 2005 to \$24.6 billion in 2012 – accounting for nearly 39% of total biologics sales in 2012 (Aggarwal 2007; Aggarwal 2014). In 2015, global sales of the top 5 best-selling mAbs (Humira®, Remicade®, Rituxan®, Avastin®, and Herceptin®), prescribed for treating breast cancer, colon cancer, Crohn’s disease to rheumatoid arthritis, exceeded US \$45 billion, indicating the high commercial value of these biotherapeutics. Following the commercialization of the first mAb product in 1986, there has been a spurt in the development of mAb products with over fifty monoclonal antibodies receiving approval by the FDA, four of them in the first half of 2016 alone, and nearly 300 molecules in active development (Ecker et al. 2015; Elvin et al. 2013; Reichert 2012).

However, despite the therapeutic benefits offered by these drugs, they present a financial burden on the healthcare system, with increased drug pricing becoming a matter of concern for both patients and policy makers (Araújo et al. 2016; Kesselheim et al. 2016). In an approach to curb these increasing prices, regulatory agencies are encouraging the development of biosimilars or biologics that are similar but not

identical to the innovator molecule. With the anticipated patent cliff expected to open up nearly half of the biologics market by 2022, manufacturers are exploring this growth opportunity in both developed and emerging markets (Deloitte Touche Ltd. 2015). The first biosimilar mAb the US market, Celltrion's Inflectra, which is a biosimilar to Janssen's Remicade, received FDA approval in early 2016 and is expected to alter the drug pricing landscape for mAb therapeutics. Similar disruptions in drug pricing have been observed in Norway where the launch of the biosimilar Remsima with discounts of up to 69% on the prices of the innovator, Remicade resulted in increased switching to the biosimilar (Stanton 2015).

As biosimilars share an identical amino acid sequence with the innovator molecule but differ only in their quality profiles, regulatory agencies are seeking extensive characterization of the complete quality attributes of the drug. The advent of biosimilars notwithstanding, there is a shift in the focus from improving productivity, which has dramatically increased with host cell optimization, to maintaining consistent product quality across batches, in order to improve overall profitability (Kelley 2009). Over the past decade, regulatory agencies have also been encouraging the biopharmaceutical industry to implement the 'Quality by Design' framework, wherein quality related to drug product safety and efficacy is in-built into every stage of the process, as per ICH Q8 guidelines (del Val et al. 2010; Rathore et al. 2010; Rathore and Winkle 2009). Maintaining drug product quality to ensure product safety and efficacy has therefore come to occupy the center stage of the biopharmaceutical industry.

One important determinant of mAb quality is N-glycosylation, a post-translational modification of the antibody in which an oligosaccharyltransferase

complex in the endoplasmic reticulum adds a sugar substrate (glycan) to the Asn-X-Ser/Thr motif in the heavy chain of the mAb (where X is any amino acid other than Pro). As the mAb traverses the Golgi complex, the attached oligosaccharide is subjected to a series of non-template driven enzymatic modifications mediated by the localized glycosyltransferase enzymes in the different Golgi compartments (Cumming 2003; Kornfeld and Kornfeld 1985). The intricate dynamics of multiple glycosyltransferase enzymes determine the eventual fate of the core glycan and result in the formation of a diverse array of glycan isoforms that affect the immunogenicity, effector functions, and the pharmacokinetic properties of the mAb, and consequently the final drug product quality (Berger et al. 2012; del Val et al. 2010; Liu 2015). Thus, there is considerable motivation for manufacturers to understand, characterize, and, if necessary, modulate the glycoform distribution in mAbs in order to maintain a consistent glycan profile and to meet the quality standards established by regulatory agencies worldwide (FDA 2006; Harmonised Tripartite Guideline 2009; Read et al. 2011; Wacker et al. 2011). However, manipulating the glycan distribution effectively requires (i) identifying the factors that can influence the glycan distribution and quantifying the degree to which these factors affect the concentration of the glycoform species; (ii) characterizing the glycan distribution profile online or in real-time; and (iii) controlling the glycosylation profile in the face of potential process variations during manufacturing. Therefore, the **overall objective of this dissertation is to provide a rational framework to quantitatively model, estimate, and control glycosylation in monoclonal antibodies produced in mammalian cells.**

The rest of the chapter is organized as follows: In §1.2, current quality considerations in the biopharmaceutical industry are discussed and the challenges

associated with controlling glycosylation during manufacturing are identified. In §1.3, we introduce our proposed framework for controlling glycosylation. Next, §1.4 and §1.5 provide a concise overview on monoclonal antibodies and glycosylation. Finally, in §1.6 we briefly introduce the contents of the remaining chapters of this dissertation.

1.2 Motivation – Quality Considerations in the Biopharmaceutical Industry

Deviations in drug product quality can result in compromised drug safety or efficacy, adversely affecting patient health. Hence, drug manufacturers are required to adhere to the quality standards and guidelines issued by drug regulatory agencies like the Food and Drug Administration (FDA). In conventional biopharmaceutical manufacturing, drug product quality testing is performed at the end of a series of production, purification, and formulation stages to evaluate if the end product meets the desired target quality profile. This ‘quality by testing’ (QbT) paradigm has ensured that marketed drugs meet specific quality standards, but it limits manufacturers to operating within certain ranges to ensure consistent product quality. Further, measuring product quality at the end of a batch generates, virtually no information about the effect of intermittent process conditions and operating parameters on the quality attributes of the drug (del Val et al. 2012). To encourage pharmaceutical innovation and improve the overall quality of the manufactured drug products, the FDA released a guidance document on a regulatory framework for Process Analytical Technology (PAT) in 2004, stating

[PAT] . . . is a system for designing, analyzing, and controlling manufacturing through timely measurements . . . of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality . . . [Quality] cannot be tested into products; it should be built-in or should **be** by design.

Thus, the goal of PAT is to ensure consistent product quality during manufacturing by designing well-understood and controlled processes (FDA 2004; Watts 2004). The concepts established in PAT lead to a marked shift in attitudes towards quality within the biopharmaceutical industry and resulted in a new “Quality by Design” (QbD) paradigm involving greater process characterization and risk assessment during manufacturing to ensure product quality.

Implementing PAT and QbD principles to design well-characterized processes requires three main activities during process development:

Design – The design phase starts early in process development where a critical quality attribute (CQA) is identified along with the critical process parameters (CPP) that affect it. In the context of monoclonal antibody development, glycosylation is a critical quality attribute as it influences a wide range of antibody properties including stability and effector functions. The final glycan distribution profile in an antibody is also affected a large number of input factors (see §1.5 for details) at different system scales – ranging from conditions at the bioreactor (macro) scale to conditions at the cellular (meso) scale and organelle (micro) scale. The current process understanding of how factors at each of these scales affect glycosylation has been generated from qualitative experiments and there is a need to generate quantitative input-output relationships in a systematic fashion.

Analyze – The analysis phase involves the design or selection of suitable analyzers that monitor CQAs and CPPs within a reasonable time-scale compared to the process time, so as to facilitate real-time decision making. Current glycosylation characterization assays however, are associated with long sample preparation times and/or infrequent measurements. In the absence of real-time measurements, there is a

need to develop techniques to infer or estimate the glycosylation profile to enable glycosylation control.

Control – In the control phase, the process understanding generated in the design phase, coupled with the CQA and CPP measurements available from the analysis phase are used to manipulate the CPPs in order to ensure consistent quality. Controlling glycosylation on-line remains a challenge due to the lack of process understanding and the absence of real-time measurements. Even when such measurements become available there currently exist no control schemes that can be implemented to control the glycosylation profile.

1.3 Proposed Framework for Glycosylation Control

Previously, we had outline a strategic vision for online quality control in the biopharmaceutical industry (St.Amand et al 2012). The work presented in this dissertation builds on this approach and addresses current unmet challenges associated with implementing PAT in upstream mAb development. Here, I propose to develop a rational framework for *modeling*, *estimating*, and *controlling* glycosylation in mAbs produced in CHO cells. The three main aspects of this work fall under the PAT framework as follows:

Design – To control glycosylation, it is necessary to generate quantitative understanding of the effect of different input factors on the glycosylation profile. In this work, we propose two approaches to generate such quantitative input-output relationships. First, a multi-scale model of glycosylation will be developed to evaluate the effect of bioreactor operating conditions on the glycosylation profile of an IgG1 producing CHO-K1 cell line. In parallel, we will identify different media supplements

that affect the glycosylation profile and quantify their effect on the glycan distribution using a combined experimental and computational approach.

Analysis – To overcome challenges associated with delayed and infrequent glycan measurements, we will develop novel techniques that can infer or estimate the glycan distribution profile based on the multi-scale model developed in the design phase. Additionally, we will evaluate rational approaches to design simpler or faster glycan assays.

Control – To control the final glycosylation profile observed at the end of the batch, we will design and implement controllers for set-point tracking using the quantitative input-output relationship established using controllability analysis.

The framework presented in this dissertation will contribute to the development of an online glycosylation control scheme that can be implemented to achieve real-time glycosylation control. In the following section, we briefly review monoclonal antibodies and glycosylation to appreciate their underlying complexity and understand why controlling glycosylation can be challenging.

1.4 Monoclonal Antibodies – An Overview

1.4.1 Antibody Structure and Function

Immunoglobulins (Igs) or antibodies are serum proteins whose main function is to bind and eliminate antigens in the body by inactivation or by triggering an inflammatory response (Wright and Morrison 1997). The core structure of these antibodies consists of two identical heavy chains and two identical light chains that together bind to form a Y-shaped structure as shown in Figure 1.1. The N terminals of the heavy and light chains consist of a variable region followed by the first constant

region of the heavy chain. Together, this variable region and constant region form the antigen binding Fab domain, while the remaining constant heavy chain regions form the Fc stem or the crystallizable stem. It is the Fab domain that recognizes and binds to the antigen, while the Fc domain is responsible for triggering the immune function resulting in the elimination of the antigen.

In humans, five major classes of immunoglobulins are known: IgG, IgM, IgA, IgE, and IgD. Of these IgG (or immunoglobulin *gamma*, named for the gamma heavy chain) is known to have four different isotypes – IgG1, IgG2, IgG3 and IgG4, whereas IgA (immunoglobulin *alpha*) has two known isotypes – IgA1 and IgA2, yielding a total of nine different immunoglobulin types with widely differing biological activities, structures and even relative abundances in human sera.

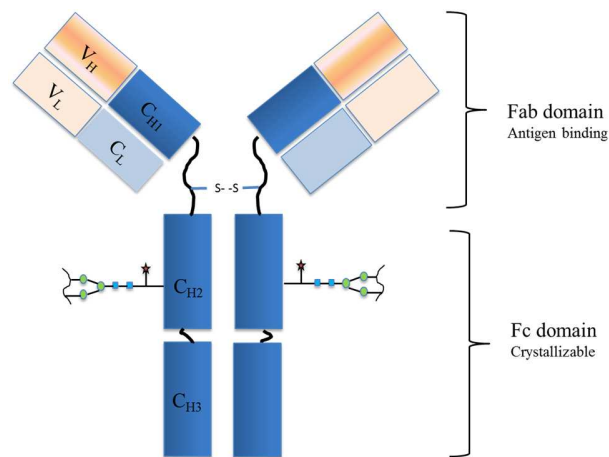


Figure 1.1: A representative image of the antibody structure. The image shows the two light chains (L) and the two heavy chains (H) that comprise the core structure of the antibody, with the variable (V) and constant (C) regions, and the antigen binding (Fab) and crystallizable (Fc) domains. Representative glycan species are also shown attached to two glycosylation sites on the CH2 domain.

Monoclonal antibodies (mAbs) are antibodies that are produced from cells originating from a single cell clone and have a very high specificity for a particular antigen (Nelson et al. 2000). Although they were long purported to be ‘magic bullets’ for selectively targeting diseases, it wasn’t until the development of hybridoma technology that use of mAb therapies became practical (Adams and Weiner 2005). Following decades of research, the first mAb product received regulatory approval in 1986 and today there are over 300 mAb products in various stages of development (Ecker et al. 2015). Therapeutic mAbs use a combination of mechanisms in triggering cytotoxic response at the cellular level such as

1. Antibody dependent cell cytotoxicity (ADCC) – where the Fab domain binds to the antigen on the diseased cell and the Fc domain binds to Fc receptor on the surface of natural killer cells; and
2. Complement dependent cytotoxicity (CDC) – where the binding of the mAbs exposes binding sites in the C1 domain, initiating the formation of a complement cascade that triggers the release of chemotactic factors.

Another mode of action involves targeting the signaling events leading to cellular proliferation by blocking the interaction of extracellular ligands with cell surface receptors. mAbs have also been used to deliver cytotoxic payloads as evidenced by the development of novel antibody drug conjugates (ADC) (Scott et al. 2012; Weiner 2007). This versatility in mAb function and structure has resulted in the approval of mAbs for a variety of indications ranging from colorectal cancer, breast cancer, non-Hodgkin’s lymphoma, and multiple sclerosis to rheumatoid arthritis, allergic asthma, and plaque psoriasis. Various strategies for developing the next generation of mAbs are being investigated at present to increase the applicability of

mAbs to more indications and to enhance their overall functionality (Beck et al. 2010; Nelson et al. 2010).

1.4.2 Antibody Expression Systems

Therapeutic mAbs can be synthesized in several different mammalian cell expression systems such as Chinese Hamster Ovary (CHO) cell lines, murine myeloma cell lines NS0, Sp2/0, with over half of all currently approved mAbs are produced in Chinese Hamster Ovary (CHO) cell lines (see Figure 1.2) (data accessed online from Drugs@FDA). Although *Pichia pastoris* and *E. coli* have been explored as potential expression systems, so far they have been used for generating antibody fragments. Two such Fab fragments that are expressed in *E.coli* and have received regulatory approval are Cimzia® and Lucentis® that are prescribed for rheumatoid arthritis and macular degeneration, respectively. Human cell lines such as human embryonic kidney (HEK) and a retinal cell line PER.C6 are also being investigated as potential cell lines to generate antibody with post-translational modifications that are identical to those found in humans. In the future, humanized mAb products from such cell lines might also gain regulatory approval.

The standard process for development of a stable mammalian cell line for expressing mAbs involves developing an expression vector containing the antibody heavy and light chain genes as well as selectable markers, followed by transfection by electroporation, lipofection transfection, and post-transfection cell line screening and selection for high productivity cells (Li et al. 2010). Commonly observed metabolic selectable markers for CHO cell lines are dihydrofolate reductase (DHFR) and glutamine synthase (GS) with methotrexate (MTX) or methionine sulfoximine (MSX) acting as the selective reagent.

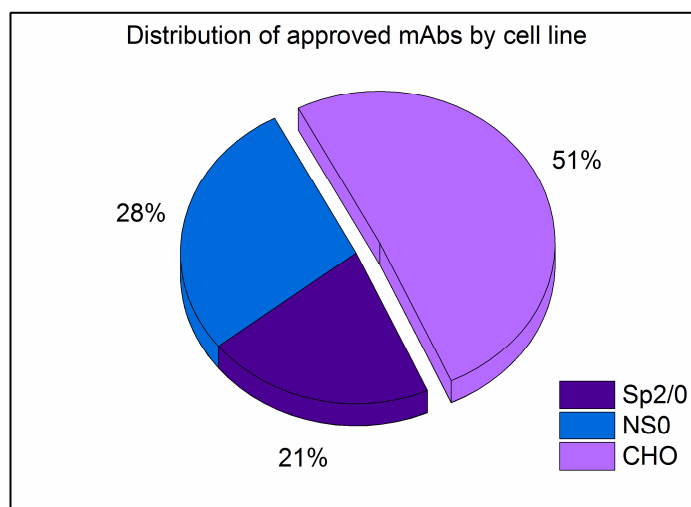


Figure 1.2: Distribution of approved mAbs by cell line used. Over half of currently approved mAbs are expressed in CHO cell lines, with the remainder being expressed in murine myeloma cell lines such as NS0 and Sp2/0

The preference for CHO as the host expression system for recombinant therapeutic proteins arises from a variety of reasons. Regulatory agencies have greater confidence in the safety of CHO-based therapeutic products due to the decades long research and safety testing that has been carried out on this commercial cell line. From a manufacturing perspective, the availability of powerful gene amplification systems, such as DHFR-mediated gene amplification, helps improve specific productivity in these cell lines thereby driving up overall profitability. In recent years, antibody titers of up to 1 g/L for batch cultures and from 1-10 g/L for fed-batch cultures have been reported for CHO based mAb production processes, indicating the extensive development that has taken place in this field. Additionally, the ease of adapting CHO cells to suspension cultures that are required for large scale glycan production makes them a preferred host for most therapeutics (Jayapal et al. 2007; Kim et al. 2012; Kunert and Reinhart 2016). Table 1.1 lists all currently approved mAbs that are

manufactured in CHO cell lines along with the corresponding indications and manufacturer.

Table 1.1: List of approved monoclonal antibodies expressed in CHO cells

mAb	Trade name	Select indications	Manufacturer
Adalimumab	Humira	Rheumatoid arthritis, Crohn's disease	AbbVie Inc.
Alemtuzumab	Campath	B-cell chronic lymphocytic leukemia	Genzyme
Alemtuzumab	Lemtrada	Multiple sclerosis	Sanofi Genzyme
Alirocumab	Praluent	Cholesterol lowering	Regeneron
Atezolizumab	Tecentriq	Urothelial carcinoma	Genentech
Bevacizumab	Avastin	Colon, lung, ovarian, kidney, brain cancer	Genentech
Daratumumab	Darzalex	Multiple myeloma	Janssen
Denosumab	Prolia	Post-menopausal osteoporosis	Amgen
Denosumab	Xgeva	Prevent fracture, spinal cord compression	Amgen
Evolocumab	Repatha	High Cholesterol treatment	Amgen
Ipilimumab	Yervoy	Stage III melanoma	BMS
Ixekizumab	Taltz	Moderate to severe plaque psoriasis	Eli Lilly
Mepolizumab	Nucala	Asthma	Glaxo-Smith Kline
Nivolumab	Opdivo	Non-small cell lung cancer	BMS
Obinutuzumab	Gazyva	Untreated chronic lymphocytic leukemia	Genentech
Omalizumab	Xolair	Allergic asthma,	Genentech
Panitumumab	Vectibix	Colorectal cancer	Amgen
Pembrolizumab	Keytruda	Melanoma, Non-small cell lung cancer	Merck
Pertuzumab	Perjeta	Metastatic breast cancer	Genentech
Rituximab	Rituxan	Non-Hodgkin's lymphoma	Genentech
Secukinumab	Cosentyx	Plaque psoriasis, psoriatic arthritis	Novartis
Siltuximab	Sylvant	Multicentric Castleman's Disease	Janssen
Tocilizumab	Actemra	Rheumatoid arthritis	Genentech
Trastuzumab	Herceptin	Her2+ breast cancer	Genentech
Vedolizumab	Entyvio	Crohn's disease	Takeda

In addition to the advantages listed above, CHO cells present the distinct advantage of having post-translational modification machinery that enables the formation of structures commonly observed in human cells, thereby ensuring biocompatibility between the products manufactured in CHO cells and human beings (Raju 2003). In the following section the different quality attributes of mAbs are discussed with a special focus on glycosylation.

1.4.3 Quality Attributes of Monoclonal Antibodies

The first generation of mAbs that were produced in murine cell lines presented immunogenic challenges in clinical evaluations, leading researchers to investigate the role of both the antibody sequence as well as post-translational modifications on antibody quality and effector functions. As mentioned above, the efficient post-translational machinery in CHO cells results in the formation of structural isoforms of the antibody that closely resemble human-like structures, thereby ensuring greater biocompatibility than mAbs generated in other cell lines. As some of these post-translational modifications are known to affect drug product quality, regulatory agencies require manufacturers to ensure they be consistent so that the drugs meet specified quality requirements. Developing the next generation of mAbs, including bispecific antibodies and antibody drug conjugates (ADCs), also requires a thorough understanding of the multiple quality attributes of monoclonal antibodies. Finally, the arrival of ‘biosimilars’ – biological drugs whose amino acid sequences are identical to that of the reference product, but the quality attributes are only similar (not identical) – along with the rapid advancement made in analytical characterization techniques, has increased the focus on the analyzing and identifying the factors that affect the quality attributes of antibodies. In the quality by design paradigm, only those physical and

chemical changes that affect drug product safety or efficacy are designated as critical quality attributes (CQAs) whose levels should be maintained within defined limits (Goetze et al. 2010; Rathore and Winkle 2009). Although quality attributes will vary between processes and products, a few quality attributes of monoclonal antibodies that are known to affect antibody activity in general are discussed here in brief (Gramer 2014; Liu et al. 2008; Zhong and Wright 2013):

Aggregation – The presence of aggregates or misfolded proteins is known to induce adverse immunological responses in patients and must hence such aggregated proteins must be cleared from the final drug product by implementing appropriate downstream purification strategies. Protein aggregation can occur due to exposed hydrophobic patches on the protein or due to changes in operating conditions. The formation of such aggregates lowers the efficiency of the process and reduces product yield and hence, appropriate strategies must be implemented to reduce protein aggregation.

Glycation – Glycation is the natural attachment of a reducing sugar to the amine group of lysine side chains via a nonenzymatic condensation reaction. Glycation of antibodies can take place during cell culture or *in vivo* upon storage with lactose. The resulting glycated antibody can exhibit lowered immunoreactivity and increase the drug product heterogeneity.

Cysteine variants – Monoclonal antibodies contain interchain and intrachain disulphide bonds, which, if disturbed, can cause heterogeneity and disulphide bond scrambling in mAbs. Reduced mAb potency has been observed due to the presence of incomplete disulphide pairing on the protein.

Apart from these chemical and physical changes to the protein, glycosylation, one of the most commonly observed post-translational modification, is considered to be a highly critical determinant of protein quality and is discussed in the following section

1.5 Glycosylation – An Overview

Glycosylation is one of the most commonly observed post-translational modifications in eukaryotic cells which results in the addition of an oligosaccharide to the protein backbone. The formation of the carbohydrate-peptide bond can be classified into five major groups: (i) N-glycosydic bonds; (ii) O-glycosydic bonds; (iii) C-glycosydic bonds; (iv) P-glycosydic bonds; and (v) Glypiation (Spiro 2002). While each of these carbohydrate-peptide linkages affect and alter protein functionality and vary by cell line and protein type, we limit our discussion to N-linked glycosylation, or the β -glycosylamine linkage of a GlcNAc (N-acetylglucosamine) to an asparagine (Asn) residue of a tripeptide Asn-X-Ser/Thr consensus sequence, where X is any amino acid except proline (Butler 2006; Cumming 2003; Gramer 2014; Kornfeld and Kornfeld 1985).

1.5.1 Glycosylation Pathway and Glycan Biosynthesis

N-linked glycosylation is initiated on the cytosolic face of the ER membrane when a nucleotide sugar donor, UDP-GlcNAc transfers a GlcNAc phosphate (GlcNAc-P) to a dolichol phosphate (dol-P) on the ER membrane resulting in the formation of a dolichol pyrophosphate N-acetylglucosamine (dol-P-P-GlcNAc). Subsequently, another GlcNAc group is added to the dolichol linked structure followed by the addition of five mannose groups by GDP-Mannose. The dolichol

linked structure then flips to the luminal face of the ER where four additional mannose residues are attached followed by three glucose residues. The fourteen sugar (Glc₃Man₉GlcNAc₂) oligosaccharide is transferred *en bloc* by an oligosaccharyl transferase (OST) enzyme to the asparagine residue of the Asn-X-Ser/Thr consensus sequon on a protein that is being translocated through the ER membrane. Following cleavage of the three glucose units on the oligosaccharide structure by the α -glucosidases in the ER, the newly synthesized protein is transferred to the Golgi for further processing (Stanley et al. 2009).

As the protein is transported through the Golgi compartment, the non-template driven enzymatic processing of the attached oligosaccharide by the different mannosidases and glycosyltransferase enzymes localized in the different regions of the Golgi apparatus. A partial list of different enzymes involved in the glycosylation pathway is listed in Table 2. The initial processing of the oligosaccharide involves the clipping of the mannose groups by mannosidase I (ManI) to form a five mannose oligosaccharide, at which point it a GlcNAc residue is added to the α -1,3-mannose by GnTI enzyme in the cis-Golgi region forming a hybrid glycan. The mannosidase in the cis-Golgi compartment cleaves majority of the glycans to produce a glycan with the core 3-mannose structure. Subsequent transfer of different glycans in the other compartments of the Golgi result in the formation of mature, complex glycans and hybrid glycans containing galactose, fucose, and sialyl residues. However, not all glycans are fully processed and the secreted glycoprotein consists of a heterogeneous distribution of different glycoforms or glycan isoforms (microheterogeneity) that impart different properties to the glycosylated protein. Here, it is important to distinguish between *microheterogeneity*, which arises due to the multiplicity of glycan

isoforms attached at a particular glycosylation site, and *macroheterogeneity*, which arises due to the variable occupation of a particular glycosylation site.

Macroheterogeneity is of particular concern in glycosylated proteins like tissue plasminogen activator (tPA) and erythropoietin (EPO) that have multiple glycosylation binding sites, and glycosylation site occupancy can compromise efficacy (Hossler et al. 2009). While both macroheterogeneity and microheterogeneity influence protein function and quality, the focus of our work will be on glycan microheterogeneity.

Table 1.2: List of enzymes participating in glycosylation reactions

Enzyme	Name
ManI (E.C. 3.2.1.113)	Mannosyl-oligosaccharide 1,2- α -mannosidase
ManII (E.C. 3.2.1.114)	Mannosyl-oligosaccharide 1,3-1,6- α -mannosidase
FucT (E.C. 2.4.1.68)	Glycoprotein- α -6-L-fucosyltransferase
GalT (E.C. 2.4.1.38)	B-N-Acetylglucosaminylglycopeptide β -1,4-galactosyltransferase
GnTI (E.C. 2.4.1.101)	α -1,3-mannosyl-glycoprotein 2- β -N-acetylglucosaminyltransferase
GnTII (2.4.1.143)	α -1,6-mannosyl-glycoprotein 2- β -N-acetylglucosaminyltransferase
GnTIII (2.4.1.144)	β -1,4-mannosyl-glycoprotein 4- β -N-acetylglucosaminyltransferase
GnTIV (2.4.1.145)	α -1,3-mannosyl-glycoprotein 4- β -N-acetylglucosaminyltransferase
GnTV (2.4.1.155)	α -1,6-mannosyl-glycoprotein 6- β -N-acetylglucosaminyltransferase
GnTE (2.4.1.149)	N-acetyllactosaminide β -1,3-N-acetylglucosaminyltransferase
SiaT (2.4.99.6)	N-acetyllactosaminide α -2,3-sialyltransferase

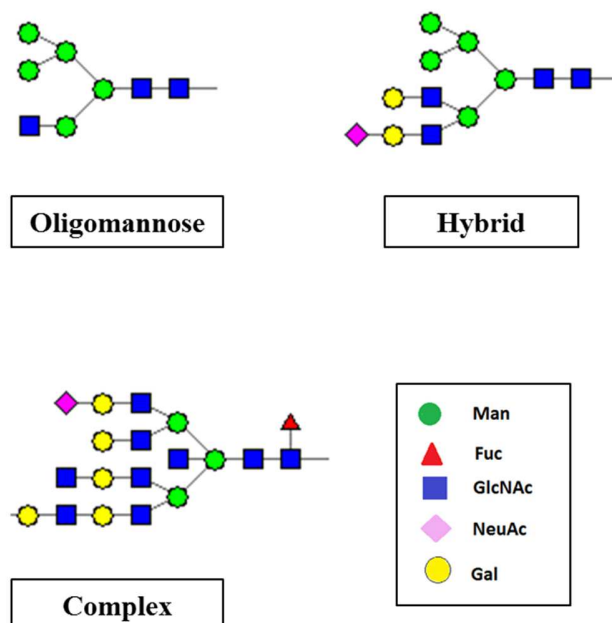


Figure 1.3: Representative examples of different types of glycan isoforms that contribute to antibody microheterogeneity. All structures have been drawn using the GlycoForm software (McDonald et al. 2010)

1.5.2 Effects of Glycosylation on Protein Quality

The heterogeneous distribution of different glycan species at a specific glycan binding site or glycan microheterogeneity influences several properties of the therapeutic protein. Changes in protein glycosylation can have significant impact on physiological processes and diseases such as hemostasis and thrombosis (Preston et al. 2013), allergies and autoimmunity (Karsten et al. 2012), tumor cell sensitivity (Mendelsohn et al. 2007), aging (Dall'Olio et al. 2013), in addition to other congenital disorders of glycosylation. At the protein level, glycosylation is known to improve protein stability and local structure (Imperiali and O'Connor 1999). The effect of glycosylation on protecting approved therapeutics, including IgG like antibodies, from protein instabilities such as proteolytic degradation, oxidation, precipitation,

aggregation, chemical denaturation, and thermal denaturation has been well documented and has led to the development of glycosylation as a strategy to develop better antibody therapies (Jefferis 2009; Onitsuka et al. 2014; Solá and Griebenow 2010; Solá and Griebenow 2009; Zheng et al. 2014). Glycans also play an intrinsic role in determining protein pharmacokinetics (PK) which in turn determines the efficacy of the protein (Sethuraman and Stadheim 2006). Glycans influence both the size and the net charge on the protein and can alter the clearance of the protein from serum. For instance, the terminal sialic group shields the protein from the asialoglycoprotein receptor on the surface of hepatocyte cells, thereby increasing serum half-life (Berger et al. 2012). In the case of monoclonal antibodies, a recent study suggested that the oligomannose species were cleared at a much faster rate than the fucosylated biantennary species thereby affecting the drug PK (Alessandri et al. 2012). Thus, manufacturers would have to ensure that the total oligomannose fractions in the antibody be reduced to enhance the serum half-life of the mAb.

Changes in the glycan profile due to the attachment of non-human glycan species can trigger immunogenic reactions in patients, as has been observed due to the presence of the alpha-linked galactose on the variable region glycosylation of the antibody Erbitux (Gramer 2014). The terminal sugar in glycans can also influence a variety of mAb properties such as antibody resistance to papain degradation (Raju and Scallon 2007) and changes in the CDC activity (Raju 2008; Raju and Jordan 2012). Core fucosylation is another factor influencing antibody effector functions as the absence of core fucose has been shown to enhance antibody ADCC activity (Houde et al. 2010; Kanda et al. 2007). Some studies suggest that antibody hemi-glycosylation does not affect Fab mediated antigen binding or Fc receptor binding (Ha et al. 2011).

Therefore, there is considerable interest in identifying factors influencing glycosylation, designing assays to characterize the glycan distribution, and developing strategies to control the glycosylation profile. We discuss some of these aspects in the following section.

1.5.3 Factors Affecting Glycosylation

In the previous section we briefly discussed how variability in protein glycosylation can affect antibody stability and efficacy, and hence its final drug product quality. In the current section we look at factors that are known to affect the glycan distribution profile, while different glycan characterization techniques are discussed elsewhere in this dissertation.

The foremost cause of differences in protein glycosylation is the choice of expression system (Jenkins et al. 1996; Parekh 1991). Mammalian, murine, insect, and yeast cell line systems exhibit differences in their post-translational machinery and the right choice of cell line becomes an important factor in determining protein quality. Given the challenges associated with characterizing and controlling glycosylation, the focus has shifted to choosing cell lines or engineering cell lines to achieve desired glycan distribution profiles (Beck et al. 2008; Sethuraman and Stadheim 2006; van Berkel et al. 2009; Yoo et al. 2010). Having chosen a particular cell expression system, there can still be differences in the glycosylation profile during batch to batch operation due to differences in bioreactor conditions (Curling et al. 1990).

For instance, glycosylation is known to be affected by bioreactor operating conditions such as pH (Aghamohseni et al. 2014; Ivarsson et al. 2014; Muthing et al. 2003; Trummer et al. 2006; Yoon et al. 2005), temperature (Ahn et al. 2008; Clark et

al. 2004; Gawlitzek et al. 2009; Sou et al. 2015), hydrodynamic stress (Godoy-Silva et al. 2009), and dissolved oxygen (Kunkel et al. 1998; Restelli et al. 2006; Serrato et al. 2004). Nutrient conditions in the bioreactor, such as the concentration of ammonia (Borys et al. 1994; Chen and Harcum 2006; Gawlitzek et al. 2000; Grammatikos et al. 1998; Yang and Butler 2000), and glucose (Fan et al. 2015a; Fan et al. 2015b; Liu et al. 2014; Villacres et al. 2015) influence the final glycan distribution profile via distinct mechanisms. Similarly, glutamine substitution by glutamate or TCA cycle intermediates was shown to alter the galactosylation and sialylation profiles in different proteins (Ha and Lee 2014; Hong et al. 2010). The use of other hexoses, notably mannose and galactose, to alter the glycan distribution is also well studied (Huang et al. 2015; Liu et al. 2015; Slade et al. 2016).

Several modifications have been made to cell culture media to evaluate the effect of different components on the glycosylation profile. Apart from altering medium osmolality to alter fucose (Konno et al. 2012), different media supplements have been introduced such as nucleotide sugar precursors (Blondeel et al. 2015; Wong et al. 2010; Zhang et al. 2016a), surfactant (Clincke et al. 2011), sodium butyrate (Borys et al. 2010; Chen et al. 2011; Gawlitzek et al. 2009), and trace metals like manganese chloride ($MnCl_2$) (Grainger and James 2013; Gramer et al. 2011; Pacis et al. 2011; Surve and Gadgil 2015).

It is important to note that the list of factors affecting the glycosylation profile presented above is not exhaustive, but is representative of the different cell culture conditions that can alter the glycan profile. Further, the understanding generated from such empirical studies has been primarily qualitative, with no quantitative input-output relationship being developed. Thus, the modulation of the glycan distribution profile

due to of multiple factors at different scales via distinct mechanisms poses several challenges to controlling glycosylation in the face of process variations during manufacturing. To date, cell line engineering and media design remain the only options to control the glycosylation profile within acceptable limits. Newer therapeutics circumventing protein glycosylation are being considered, but such aglycosylated proteins require extensive cell engineering and considerable investment (Ju and Jung 2014; Jung et al. 2011).

Therefore, the work presented in this dissertation expands the available tools to control the glycosylation profile by generating fundamental quantitative relationships between inputs and outputs and then using the generated information in novel control schemes to ensure consistent glycan distribution profiles.

1.6 Dissertation Overview

The overall objective of this work is to develop an effective framework for modeling, estimating and controlling the glycosylation profile in monoclonal antibodies produced in CHO cells. In Chapter 2 an integrated multi-scale model of glycosylation is developed for an in-house CHO cell line grown under batch and fed-batch conditions. Chapter 3 takes a parallel approach to developing a quantitative understanding of the effect of media supplements on the glycan profile. While the role of media supplements on glycosylation is well-studied, the work presented here evaluates the effect of dynamic media supplementation using experimental and computational techniques using controllability analysis. We then design and implement glycosylation controllers using controllability analysis and evaluate the set-point tracking ability of glycosylation controllers in Chapter 4. In Chapter 5 the different challenges to developing an online glycosylation assay are discussed and two

different approaches based on observability analysis and state estimation are used to address the issue of delayed, infrequent measurements. Finally, the key findings are summarized in Chapter 6 along with a discussion of future directions.

Chapter 2

MULTI-SCALE MODELING OF ANTIBODY GLYCOSYLATION

2.1 Introduction

As seen in Chapter 1, the glycosylation profile in mAbs is a critical determinant of protein quality as it influences such quality attributes as antibody immunogenicity, stability, efficacy, and half-life. Therefore, to ensure drug product quality, it is important to ensure consistent glycosylation of mAbs during manufacturing. However, as we have noted previously, protein glycosylation is extremely complex, involving a series of non-template driven enzymatic reactions that can be influenced by a variety of factors spanning multiple ‘scales’ – from macro-scale properties such as bioreactor operating conditions and media composition, to meso-scale properties such as antibody productivity and nucleotide sugar donor concentrations at the cellular level, to micro-scale properties at the organelle level such as kinetic rates of glycosyltransferase and glycosidase enzymes in the Golgi apparatus of the cell. Thus, in order to develop any control strategy that will ensure consistent glycosylation, it is first necessary to identify the factors that influence glycosylation and develop a fundamental, quantitative input-output relationship between different input factors and the output glycan profile. This chapter and the following chapter demonstrate two parallel approaches taken to understand how the glycosylation profile is influenced by different input factors. In this chapter, we explore, by means of a fundamental mathematical model, the effect of bioreactor operating conditions on the glycosylation profile, while in the next, we use

controllability analysis to establish a quantitative relationship between different media supplements and the output glycosylation profile.

The motivation to develop a mathematical model of a complex biological system can stem from a need to organize the vast and disparate information available for a biological system into one coherent whole; or to understand the qualitative and phenomenological features of the system; or to develop predictive capabilities for a process of interest; or as means to synthesize existing empirical knowledge of a biological system and generate new insights about the underlying complex mechanisms (Bailey 1998). In the context of antibody glycosylation, generating mathematical models to describe the intricate multi-scalar interactions leading to changes in the glycan distribution will enable manufacturers to optimize process operating conditions based on a fundamental understanding of the effect that any process change would have on the final quality profile. Developing such predictive capabilities is vital to ensure the production of high value biotherapeutics with consistent quality. Further, such a mathematical model can also be useful when designing a control scheme to meet desired quality targets in the face of process variations during manufacturing. In this chapter, we develop a multi-scale model for glycosylation and test its performance for an in-house cell line under batch and fed-batch conditions.

The chapter begins with a brief review of various modeling approaches for biological processes, specifically focusing on models of cell growth in mammalian cell cultures and mathematical models for predicting the glycan distribution in glycoproteins, as described in §2.2. The use of these models in developing an in-house

multi-scale model for glycosylation is then presented in §2.3 and model results are compared to experimental data collected from batch and fed-batch experiments.

2.2 Mathematical Models for Biological Processes

2.2.1 Modeling Cell Growth in Mammalian Cell Culture

A variety of kinetic models have been used to simulate and understand cellular growth. Much like the models used to describe microbial cell growth, kinetic models for mammalian cell culture can be classified based on the level of detail incorporated in the model (Tziampazis and Sambanis 1994). Briefly, the various classes of models based on their structural classification are:

Unstructured, unsegregated models: A vast majority of the models used for describing cellular growth are unstructured and unsegregated. Based on an idealized representation of cell growth, these models assume that all cells in the culture behave in the same fashion (unsegregated) and treat the cell as a black box, with no accounting of cellular reactions within the different regions of the cell (unstructured). Commonly used empirical models that describe cell growth on the basis of logistic growth rate expressions (Goudar 2012) or Monod kinetics fall under the unstructured, unsegregated category of models (Shirsat et al. 2015).

Structured, unsegregated models: Single cell models (Sanderson et al. 1999; Sidoli et al. 2004; Wu et al. 1992) that take into account intercellular transport, cellular compartmentalization, and intracellular metabolic pathways, and macroscopic biological models (Baughman et al. 2010; Provost and Bastin 2004; Provost et al. 2006) that account for cellular metabolism are classified as structured, unsegregated kinetic models. Such models tend to be more complex, and hence, more

computationally demanding than unsegregated models. However, these single cell models do not account for any heterogeneity whatsoever in the cell population.

Unstructured, segregated models: Segregated models or population balance models describe cell growth while factoring in the heterogeneity in cellular populations. With the advent of experimental tools such as flow cytometry that can be used to sort cells based on cellular age, and with improved computational tools, the applicability of these models in characterizing cellular growth has increased (Jang and Barford 2000; Karra et al. 2010; Munzer et al. 2015a; Munzer et al. 2015b).

Segregated, structured models: The most realistic models of cell growth are those that take into account both the heterogeneity in cellular population as well as the internal metabolic structure (Sidoli et al. 2006). The high level of parameterization in these complex models necessitates the development of elaborate parameter reduction and estimation techniques, thereby increasing the associated computational challenges.

In addition to such a structural classification of models, it is possible to develop models that do not necessarily conform to one or the other type. For instance, Kontoravdi and coworkers (Kontoravdi et al. 2005) developed a hybrid structured/unstructured model for a mAb producing mammalian cell culture, where cell growth dynamics were modeled using unstructured Monod kinetics, but the antibody productivity was modeled by taking into account the formation of the light and heavy chains starting from an organelle level mass balance for mRNA (Bibila and Flickinger 1991). The model parameters were then identified using global sensitivity analysis and model predictions were compared to experimental data. The utility of such a combined approach was further demonstrated by Ho and coworkers who compared the model predictions from a hybrid structured/unstructured model and a

detailed single cell model to experimental data obtained from a mAb producing GS-NS0 cell line under hyperosmotic stress (Ho et al. 2006). They observed that predictions from the simpler hybrid model were comparable to those from the more computationally demanding single cell model and required simpler parameter estimation techniques. Similarly, data driven models using artificial neural networks (Marique et al. 2002) or based on Markov chain Monte Carlo methods (Xing et al. 2010) have been developed to describe cellular growth in mammalian cells.

While different forms of such structurally differentiated models have been used to model cell growth and antibody productivity in mammalian cell culture, the eventual choice of the model is dictated by the application for which the model is developed (Kontoravdi et al. 2010). A segregated, structured model can be used to understand cellular behavior, but the associated computational challenges render it impractical for process control and monitoring purposes, which are better served by simpler unstructured and unsegregated models. Thus, model complexity should be chosen appropriate to the task at hand.

Next we discuss the different approaches that have been used to develop qualitative and quantitative models of protein glycosylation.

2.2.2 Modeling Glycosylation

The inherent complexity in glycosylation reaction networks has provided significant challenges to modelers seeking to create appropriate mathematical representations of the underlying biological phenomenon. In overcoming these challenges, several models based on mechanistic and empirical approaches have been developed (Sha et al. 2016). Each of these glycan models has contributed significantly to our knowledge of the mechanisms associated with protein glycosylation and helped

quantify the changes in the glycosylation patterns in glycoproteins. Although models for both N-linked and O-linked glycosylation are widely available (Neelamegham et al. 2008; Puri and Neelamegham 2012), for the purposes of this study, we limit our discussion to models for N-linked glycosylation alone.

Glycosylation models have been developed to understand the diversity in glycan site occupancy in glycoproteins and the diversity in glycan structures and branching patterns at specific glycan sites on the protein backbone. Although models describing the branching patterns in observed in N-linked glycosylation reaction networks are more numerous, very few have examined the heterogeneity in glycan site occupancy (macroheterogeneity) that arises in glycoproteins with multiple glycan binding sites. For instance, the kinetic modeling framework developed by Shelikoff and coworkers (Shelikoff et al. 1996) modeled the co-translational transfer of the dolichol phosphate linked oligosaccharide to the nascent polypeptide. Although their model was largely qualitative in nature, it provided a framework to determine how glycosylation site occupancy depended on protein synthesis rates, and other parameters. In contrast to this structured kinetic model, Senger developed a predictive model using artificial neural networks (Senger and Karim 2005) to study site occupancy of recombinant tissue plasminogen activator (r-tPA) based on protein sequencing. Their goal was to understand how different protein residues influence the eventual glycan macroheterogeneity in glycoproteins with multiple binding sites. As with the structured kinetic model developed previously, the challenge of corroborating model predictions with experimental data has limited the applicability of models predicting glycan macroheterogeneity.

Mathematical models that predict the variability in glycan branching and the resulting microheterogeneity in glycoproteins have gained in popularity in recent years. These models quantify the non-template driven, enzymatic modification of the oligosaccharide attached to the protein backbone as the glycoprotein traverses the different regions of the Golgi apparatus before being secreted. One of the earliest kinetic models to capture the complexity of N-linked glycosylation networks was developed by Umaña and Bailey who wished to develop rational metabolic engineering strategies guided by effective mathematical models to achieve desired glycoform distributions (Umaña and Bailey 1997). In this seminal work, the mathematical model (referred to as UB1997) was developed considering mass balances for a simple glycosylation reaction network with 33 glycoforms participating in 33 reactions catalyzed by 7 glycosyltransferase enzymes. They factored the transport of proteins between different regions of the Golgi complex and modeled the individual enzymatic reactions using Michaelis-Menten rate expressions. The resulting oligosaccharide balances were generated using the glycan productivity rate which, in turn, was estimated from the glycoprotein productivity rate – a critical assumption that formed the basis of all subsequent kinetic models. Having estimated kinetic parameters from literature, the model was then analyzed at steady state to assess qualitative differences in glycan distribution as a function of antibody productivity as well as differences in the glycan distribution arising due to changes in glycan activity.

Krambeck and Betenbaugh expanded the scope and applicability of the UB1997 model by incorporating additional glycosyltransferase enzymes and glycan branching structures, thereby enhancing the predictive capability of the model to a very wide range of oligosaccharide structures (Krambeck and Betenbaugh 2005). The

resulting mathematical model (referred to as KB2005) predicted the formation of 7565 glycan structures participating in 22,871 reactions catalyzed by 11 glycosyltransferase enzymes. Like the UB1997 model before it, the KB2005 model too assumed that the Golgi apparatus could be modeled as a series of well-mixed reactors in sequence with bulk transport of cargo. The KB2005 model introduced a framework for automated generation of different glycosylation reaction networks (and consequently, various oligosaccharide sequences) by defining enzyme reaction rules. These rules vastly simplified the computational expense involved in generating large scale glycan reaction networks and enabled effective mathematical representation of different glycoforms using a nine-digit numbering system. The developed kinetic model was solved under steady state conditions using kinetic parameters estimated from literature and compared to experimental data. The model was also used to study the effect of increased glycan productivity on the glycosylation profile. In subsequent developments, the predictive capabilities of an expanded version of the KB2005 model were refined by adjusting model parameters based on mass spectrometric measurements (Krambeck et al. 2009) and transcriptomic data was integrated with the glycomic model to identify cell biomarkers (Bennun et al. 2013).

Contemporaneous to the development of the KB2005 model, which was based on the then prevailing vesicular transport model of the Golgi complex, studies of protein secretion in live cells seemed to suggest that the cisternal maturation model could account for secretory kinetics in the Golgi compartment better than the vesicular transport model (Losev et al. 2006; Matsuura-Tokita et al. 2006). To compare the differences in the glycan distribution arising due to the choice of vesicular transport model and the cisternal maturation model, Hossler and coworkers developed two

reaction schemes wherein they modeled the different compartments of the Golgi compartment as four continuous mixing tanks (CSTR) or four plug flow reactors (PFR) respectively (Hossler et al. 2007). The authors examined the effect of processing time, compartmentalization, and spatial localization of enzymes on the glycan distribution while studying the sensitivity of individual glycoforms to the concentrations of different enzymes. They noticed that the four PFR model provided a more realistic representation of the glycan distribution than the four CSTR model, indicating the effect of processing time on glycan microheterogeneity. The work done in this study is further enhanced when compared to past experimental data highlighting the effect of processing time on glycan microheterogeneity. For instance, Wang and coworkers (Wang et al. 1991) found that the amount of poly-acetyllactosamines in membrane glycoproteins increased with prolonged association with the Golgi at lower temperatures due to slower intracellular transport. This was similar to the work by Fuller and coworkers (Fuller et al. 1985) who had demonstrated a change in the galactosylation and sialylation of a G protein processed at lower temperatures, due to increased residence time in the Golgi.

In a further extension based on the cisternal maturation model, del Val and coworkers developed a model (referred to as DK2011) where the Golgi complex was viewed as a single PFR and obtained dynamic material balance for different glycoforms as well as nucleotide sugar donors (del Val et al. 2011). They also defined enzyme and transporter protein concentrations along the length of the Golgi apparatus to account for recycling of the different components along the Golgi compartment, while incorporating detailed kinetic mechanisms for each enzyme catalyzed reaction. Kinetic parameters were estimated from literature and unknown parameters were

estimated using optimization routines. The resulting model was matched with experimental data and compared to previous models. The DK2011 model provided a comprehensive dynamic glycosylation model that, when coupled to a model for nucleotide sugar donor (NSD) metabolism could link bioreactor and extracellular conditions to intracellular changes.

As most of these kinetic models require extensive parameter estimation, there have been attempts at reconciling experimental glycosylation distribution data to parameter free models. For instance, predictive glycan models have been developed using artificial neural networks (Senger and Karim 2008) that correlate the type of glycan attached to a particular site to protein structure and sequence. By examining protein databases and antibody structures, such a neural network model aims to reduce the need for using elaborate kinetic models. Recently, Spahn and coworkers developed a low parameter Markov chain model for predicting the glycan distribution in different glycoproteins and evaluated its efficacy to predict the effect of glycosyltransferase knockdowns on glycosylation (Spahn et al. 2016). By utilizing the reaction rules first elucidated in the KB2005 model, a probabilistic framework was developed for generating a glycosylation reaction network. The model does not account for nucleotide sugar precursors, but simply adjusts the transition properties in the Markov model to account for changes in the glycan distribution. Along similar lines, Kremkow and Lee have developed a parameter free glycosylation network model specifically for CHO model based on the CHO genome (Kremkow and Lee 2016).

While such predictive models have helped lay the framework for understanding the diversity in glycosylation reaction networks, there is a need to understand how bioreactor culture conditions influence the heterogeneity in

glycosylation distribution. The development of models linking extracellular conditions to changes in the intracellular conditions is reviewed in the next section.

2.2.3 Modeling Cell Growth and Antibody Glycosylation

An examination of the factors that influence glycosylation indicates that the eventual antibody glycan distribution profile is affected by both intracellular and extracellular factors as reviewed in §1.5.3. By understanding the mechanistic relationship between the input factors and the output glycan distribution, one can develop necessary schemes to manipulate the various input factors and alter the resulting glycosylation profile in the protein.

One such approach to understanding the effect of growth dynamics and antibody productivity on the final glycan distribution involves linking the ‘macro-scale’ models of cell growth with the ‘micro-scale’ models of glycan productivity. Kontoravdi and coworkers developed a dynamic model linking a hybrid structured/unstructured model for cell growth for a mAb producing cell culture with the UB1997 model for glycosylation (Kontoravdi et al. 2007). Although the model produced qualitative results, it was a useful predictive tool to study the effect of different feeding strategies on fed batch cell culture. With the development of detailed micro-scale models of glycosylation that could be linked to experimental data, it was now possible to use this framework to link extracellular conditions to the eventual glycosylation profile. Jedrzejewski and coworkers sought to develop one such modeling framework by focusing on an *in silico* model reconstruction of sugar nucleotide and nucleotide sugar donor (NSD) synthesis pathways to link the availability of NSDs to the final glycan distribution predicted by the DK2011 model (Jedrzejewski et al. 2014). They linked a dynamic cell growth model to a simplified

nucleotide model developed using Monod kinetics to a NSD synthesis model developed from the KEGG database and modeled using Michaelis-Menten kinetics, to the DK2011 glycosylation model. The resulting model prediction for the glycosylation profile was compared to experimental data from an antibody producing mammalian cell culture and it showed a fairly reasonable match. Recently, the influence of antibody productivity on the glycosylation profile was examined using this integrated modeling framework for fed-batch cultures (del Val et al. 2016).

In a different modeling approach, Kaveh and coworkers linked a dynamic model for cell growth to the glycosylation model proposed by Hossler et al (Ohadi et al. 2013). Recognizing that experimental measurements of glycan distribution in cell culture are representative of the cumulative amount of antibody generated, and that the kinetic models for glycosylation are instantaneous models, they converted the instantaneous glycan distribution to a cumulative measure by accounting for the cumulative antibody production at that time. Further by formulating a dynamic metabolic flux model for the extracellular metabolites, they attempted to link extracellular conditions to the glycosylation profile.

The models reviewed thus far have been developed using a wide range of assumptions and for different systems. However, different cell lines producing different glycoproteins (or even the same glycoprotein) exhibit differences in cellular behavior. Thus in the absence of a general model of cell growth and glycosylation it becomes necessary to develop a specific model with a particular set of experimental parameters. In the next section we present results from our multi-scale modeling efforts using an in-house cell line.

2.3 Development of an In-house Multi-scale Model of Glycosylation

To model the effect of cell growth conditions on the glycan distribution profile, we developed a multi-scale model to describe the cellular growth and glycosylation characteristics of an in-house cell line, wherein the macro-scale model was developed using an unstructured Monod model for cell growth and the micro-scale model was adapted from the DK2011 model. Model predictions were then compared to batch and fed-batch experimental data, as described in the following sub-sections.

Figure 2.1 provides a brief overview of the overall model structure which incorporates both the macro-scale and the micro-scale models.

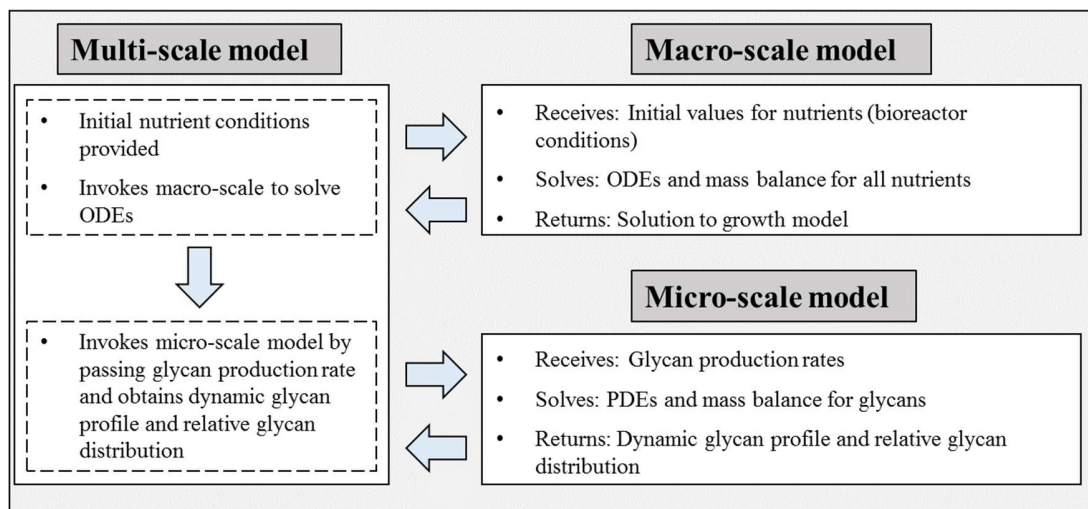


Figure 2.1: Overview of the in-house multi-scale model. The multi-scale model first invokes the macro-scale model to solve mass balances for all nutrients that returns the dynamic nutrient concentration profile and the antibody production rate. The glycan production rate is then calculated and the micro-scale model is invoked to obtain the dynamic glycan distribution profile and the relative distribution of individual glycoforms.

2.3.1 Batch Experiments – Materials and Methods

An IgG1 producing CHO-K1 cell line (gift of Genentech, CA) was used in our shake flask batch experiments. The cells were grown in suspension culture at 37° C with a 5% CO₂ overlay in vented cap Erlenmeyer shake flasks with 50 mL custom CD OPTICHO™ media (Thermo Fisher, MA) supplemented with 28 mM glucose, 4 mM glutamine, and an initial seeding density of 0.5 x 10⁶ cells/mL. Daily cell count measurements were taken using a hemocytometer, while metabolite (glutamine, glucose, glutamate and lactate) concentrations, media pH, and osmolality were measured using a Bioprofile 100+ analyzer (Nova Biomedical, MA). On day 8 after inoculation, the cells were centrifuged to harvest spent media from which the IgG1 antibody was then purified with a PhyNexus Benchtop MEA2 system using Protein A chromatography resin packed in a 2 mL PhyTip column (PhyNexus, CA). Antibody titer in the harvest and post-purification was quantified on a Thermo Scientific™ MAbPac Protein A chromatography column (12-micron particle size, 35x4.0 mm I.D., Thermo Fisher Scientific, MA) using an Agilent 1200 HPLC instrument. The purified antibody was then trypsinized at 37° C, followed by enzymatic deglycosylation using PNGase-F (ProZyme, CA) for a minimum of 16 hours at 37° C. The free separated glycans were captured on Hypersep Hyper Carb SPE cartridges and permethylated following the Ciucanu method using methyl iodide and NaOH in the presence of DMSO (Ciucanu and Costello 2003; Ciucanu and Kerek 1984). The labeled glycans were purified in a liquid-liquid extraction step with chloroform, dried and resuspended in 80% methanol, and spotted onto a MALDI/TOF plate with a DHB matrix. The labeled glycans were then analyzed using a 4800 MALDI TOF/TOF Analyzer (ABSciex) in positive ion, reflector mode. The data collected using the mass spectrometer was then exported to DataExplorer to obtain the peak heights for the

identified glycans (see Appendix A for the complete list of identified glycans). The relative glycan distribution was calculated from the peak heights of individual glycan species.

2.3.2 Macro-scale Model Development

The macro-scale model used in this work is an unstructured, unsegregated model based on Monod kinetics, with appropriate mass balance equations included to account for nutrient consumption.

In mammalian cell cultures, cell growth depends on the availability of two key nutrients, glucose and glutamine. Cells uptake glucose through the glycolytic pathway, resulting in the formation of lactate as a by-product, while the uptake of glutamine results in the formation of ammonia as a by-product. Both lactate and ammonia are known to inhibit cell growth in mammalian cells (Lao and Toth 1997). We take into account each of these factors in developing the expression for cell growth rate, μ , as

$$\mu = \mu_{\max,1} \frac{\text{Glc}}{(K_{\text{Glc}} + \text{Glc})} \frac{\text{Gln}}{(K_{\text{Gln}} + \text{Gln})} \frac{K_{i,\text{Lac}}}{(K_{i,\text{Lac}} + \text{Lac})} \frac{K_{i,\text{Amm}}}{(K_{i,\text{Amm}} + \text{Amm})} \quad 2.1$$

The cell death rate, which is inversely related to the growth rate, asymptotically reaches a maximum value as the growth rate goes to zero (Sanderson et al. 1999).

Thus

$$\mu_d = \mu_{d,\max} \frac{k_{d,0}}{(k_{d,1} + \mu)} \quad 2.2$$

The consumption rates for glucose and glutamine are obtained by considering a mass balance for the two nutrients. For glucose we obtain

$$\frac{dGlc}{dt} = -Glc \frac{1}{V} \frac{dV}{dt} - q_{Glc} X_v + \frac{F_{in} Glc_{in}}{V} - \frac{F_{out} Glc}{V} \quad 2.3$$

Here, V is the working volume in the bioreactor, F_{in} and F_{out} are the flowrates into and out of the bioreactor, with Glc_{in} denoting the concentration of glucose in the input feed. Under batch conditions, we set F_{in} and F_{out} as zero. Further, the specific consumption rate for glucose, q_{Glc} is given by

$$q_{Glc} = \frac{\mu}{Y_{X/Glc}} + m_{Glc} \quad 2.4$$

where, $Y_{X/Glc}$ is the yield coefficient and the maintenance coefficient for glucose, m_{glc} is given by

$$m_{Glc} = \frac{a_1 Glc}{a_2 + Glc} \quad 2.5$$

Similarly, glutamine consumption is given by

$$\frac{dGln}{dt} = -Gln \frac{1}{V} \frac{dV}{dt} - q_{Gln} X_v - K_{d,Gn} Gln + \frac{F_{in} Gln_{in}}{V} - \frac{F_{out} Gln}{V} \quad 2.6$$

with the specific glutamine consumption rate given by

$$q_{Gln} = \frac{\mu}{Y_{X/Gln}} + m_{Gln} \quad 2.7$$

where $Y_{X/Gln}$ is the yield coefficient of glutamine and m_{Gln} is the maintenance coefficient for glutamine.

Glucose uptake by the cells results in the formation of lactate as a by-product at a rate directly proportional to the glucose consumption rate, with the accumulation of lactate resulting in growth inhibition. However, some CHO cells are known to consume lactate under low glucose conditions under batch and fed-batch conditions (Ozturk et al. 1997) resulting in higher productivities. This has led investigators to study the metabolic changes that result in the onset of lactate consumption and its effect on total protein productivity (Le et al. 2012; Mulukutla et al. 2012; Mulukutla et al. 2015). Several strategies have also been devised to control the production of lactate by substituting galactose for glucose or by using pH controlled addition of glucose thereby seeking to enhance protein productivity (Altamirano et al. 2006; Gagnon et al. 2011). In our shake flask experiments too, we observe the consumption of lactate as the growth phase shifts under low glucose and glutamine conditions. Although both glucose and glutamine are vital for cell growth and maintenance in CHO cells, cells continue to grow even at limiting concentrations of glutamine (Sun and Zhang 2004). The rapid uptake and degradation of glutamine accompanied by the consumption of lactate results in a slowing down of cellular growth and the expression for cellular growths is now modeled as

$$\mu = \mu_{\max,2} \frac{\text{Glc}}{(K_{\text{Glc}} + \text{Glc})} \frac{\text{Lac}}{(K_{\text{Lac}} + \text{Lac})} \frac{K_{i,\text{Amm}}}{(K_{i,\text{Amm}} + \text{Amm})} \quad 2.8$$

Consequently, when we account for lactate mass balance, we incorporate terms for both lactate production rate as well as consumption. Thus:

$$\frac{d\text{Lac}}{dt} = -\text{Lac} \frac{1}{V} \frac{dV}{dt} + q_{\text{Lac}} X_v - q_{\text{cons}} X_v + \frac{F_{\text{in}} \text{Lac}_{\text{in}}}{V} - \frac{F_{\text{out}} \text{Lac}}{V} \quad 2.9$$

where the specific lactate production rate, q_{Lac} is given by:

$$q_{\text{Lac}} = Y_{\text{Lac}/\text{Glc}} q_{\text{Glc}} \quad 2.10$$

while the specific lactate consumption rate, q_{cons} is given by:

$$q_{\text{cons}} = \begin{cases} k, & \text{under exponential growth conditions} \\ \frac{\mu}{Y_{\text{X}/\text{Lac}}}, & \text{under low glucose conditions} \end{cases} \quad 2.11$$

The uptake of glutamine during cellular growth as well as the degradation of glutamine results in the formation of ammonia, which continues to accumulate even after glutamine depletion. This is likely due to the degradation of glutamate and the conversion of alanine to pyruvate as the cells consume lactate (Li et al. 2012). Thus, we write the mass balance for ammonia as

	$\frac{dA_{mm}}{dt} = -A_{mm} \frac{1}{V} \frac{dV}{dt} + q_{A_{mm}} X_v + K_{d,Gln} Gln - \frac{F_{out} A_{mm}}{V} + k_{amm}$	2.12
--	--	------

where the specific ammonia production rate is given by

$$q_{A_{mm}} = Y_{A_{mm}/Gln} q_{Gln} \quad 2.13$$

The viable cell density (X_v) depends on the growth rate as well as the death rate, while the total cell density (X_t) depends on the number of viable cells and the cell lysis rate K_{lysis} . Therefore, we get

$$\frac{dX_v}{dt} = \left(\mu - \mu_d - \frac{F_{out}}{V} - \frac{1}{V} \frac{dV}{dt} \right) X_v \quad 2.14$$

$$\frac{dX_t}{dt} = \mu X_v - \frac{F_{out}}{V} X_t - \frac{1}{V} \frac{dV}{dt} X_t - K_{lysis} (X_t - X_v) \quad 2.15$$

Several different approaches have been used to model protein productivity, using segregated or unsegregated, structured or unstructured models. For instance, an antibody synthesis model was developed for hybridoma cells that accounted for light and heavy chain protein and mRNA expression (Bibila and Flickinger 1991; Bibila and Flickinger 1993). Extensive factorial design experiments have been performed to identify appropriate cell culture media supplementation that can enhance protein productivity (Xu et al. 2014). Data mining approaches involving historic manufacturing data have also been used to identify factors that influence protein

productivity (Charaniya et al. 2010). Recent investigations of the metabolomic activity in different cell lines producing industrially relevant mAbs suggest a correlation between oxidative stress and the production of TCA cycle intermediates (Ishii et al. 2015; Templeton et al. 2013). However, for our modeling purposes, we have associated antibody titer with the glucose consumption rates. Thus, we get a mass balance for the antibody production rate as

$$\frac{d\text{MAb}}{dt} = -\text{MAb} \frac{1}{V} \frac{dV}{dt} + q_{\text{MAb}} X_v - \frac{F_{\text{out}} \text{MAb}}{V} \quad 2.16$$

where the specific antibody production rate, q_{MAb} is given by

$$q_{\text{MAb}} = Y_{\text{MAb}/\text{Glc}} q_{\text{Glc}} \quad 2.17$$

and $Y_{\text{MAb}/\text{Glc}}$ is the yield coefficient of the antibody on glucose.

The resulting set of ODEs was solved in MATLAB using an in-built ODE solver with a non-negativity constraint. Nutrient concentration from the batch experiments was measured on day 0 using the Bioprofile Analyzer and these values were passed to the ODE solver as initial conditions. The model parameters were obtained from literature and predictions for glucose, glutamine, and lactate were compared to the experimental values obtained from the batch run. Later these, parameters were optimized using the in-built optimizer in MATLAB. The parameters used in the macro-scale model have been listed in Table 2.1

Table 2.1: Parameter values used in the macro-scale model

Parameter	Value (in-house model)
Yield coefficient of biomass on glucose, $Y_{X/Glc}$, [cells/ mM]	1.40×10^9
Yield coefficient of biomass on glutamine, $Y_{X/Gln}$, [cells/ mM]	2.70×10^9
Yield coefficient of biomass on lactate, $Y_{X/Lac}$, [cells/ mM]	6.53×10^7
Yield coefficient of ammonia on glutamine, $Y_{Amm/Gln}$, [mM/ mM]	0.63
Yield coefficient of lactate on glucose, $Y_{Lac/Glc}$, [mM/ mM]	1.30
Yield coefficient of mAb on glucose, $Y_{MAb/Glc}$, [g/L/ mM]	5.55×10^{-3}
Constant for glutamine degradation, $K_{d,Gln}$, [hour ⁻¹]	9.60×10^{-3}
Monod constant for glucose, K_{Glc} , [mM]	0.14
Monod constant for lactate, K_{Lac} , [mM]	0.25
Monod constant for glutamine, K_{Gln} , [mM]	0.025
Constant for lactate inhibition, $K_{I,Lac}$, [mM]	171.76
Constant for ammonia inhibition, $K_{I,Amm}$, [mM]	28.48
Cell lysis rate, K_{lysis} , [hour ⁻¹]	0.02 – 0.06
Glutamine maintenance coefficient, $mgln$, [mM-hour ⁻¹ /cells]	4.25×10^{-15}
Constant for glucose maintenance coefficient, a_0 , [mM-hour ⁻¹ /cells]	2.25×10^{-10}
Constant for glucose maintenance coefficient, a_1 , [mM]	39.65
Maximum growth rate (exponential), μ_{max1} , [hour ⁻¹]	0.03
Maximum growth rate (stationary), μ_{max2} , [hour ⁻¹]	6.50×10^{-3}
Maximum death rate, $\mu_{d,max}$, [hour ⁻¹]	0.042
Death rate constant, k_{d0} , [hour ⁻¹]	4.54×10^{-4}
Death rate constant, k_{d1} , [hour ⁻¹]	5.00×10^{-3}

The resulting dynamic profiles for nutrient uptake and production are shown in Figure 2.2, while Figure 2.3 shows a comparison of the model predictions and the experimental measurement for the cellular viability. The model adequately captures

the trends in the nutrient uptake and metabolite production as well as the cell growth rates. Due to the low antibody concentrations seen in our experiments, mAb measurements are available for just two time points, on day 7 and day 8 at the end of the batch. The model predictions for antibody production compare well with the experimental data as shown in Figure 2.4.

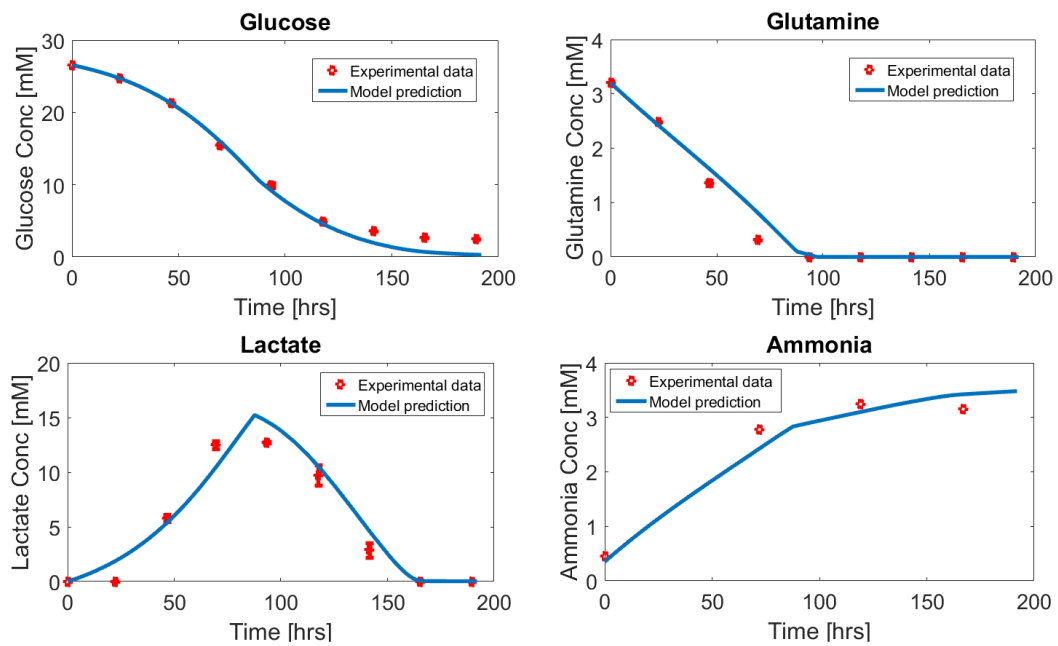


Figure 2.2: Nutrient and by-product concentration profiles for batch culture. The plot shows the model fit for glucose, glutamine, lactate, and ammonia. The blue solid line represents the model predictions, while the red circles are the experimental measurements.

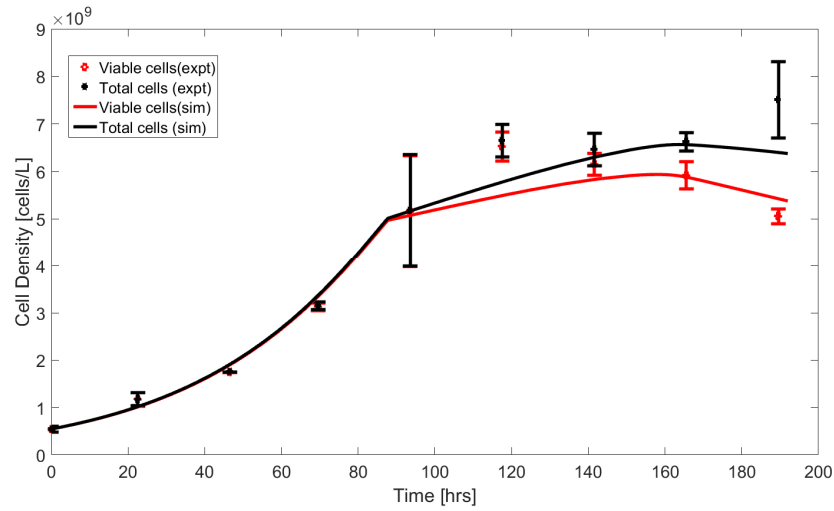


Figure 2.3: Viable cell density and total cell density profiles for batch culture. The solid black line represents the model predictions for the total cell density while the solid red line represents the model fit for viable cell density. The black and red circles represent the average of the experimentally observed total cell density and viable cell density respectively, while the error bars represent the range of observed values (n=2).

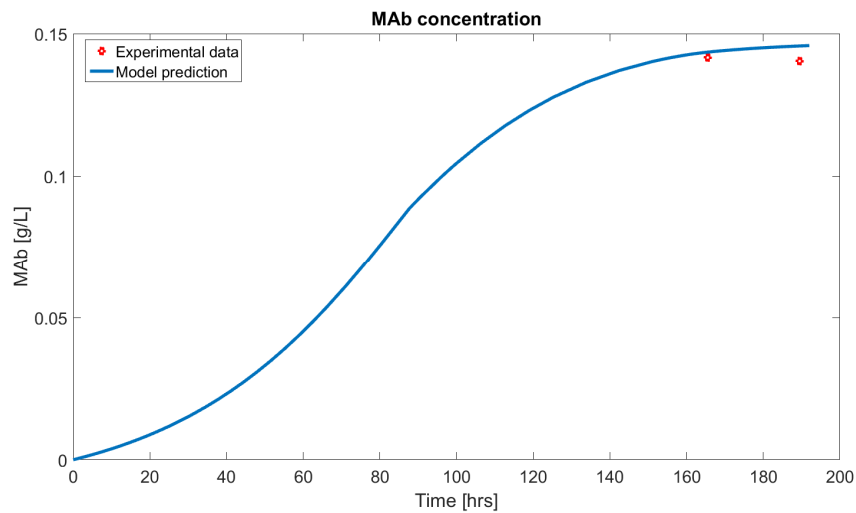


Figure 2.4: Antibody concentration profile for batch culture. The solid blue line represents the model simulations while the red circles represent the experimental data.

2.3.3 Coupling the Macro-scale Model with the Micro-scale Model

To couple the macro-scale model with the eventual glycan distribution in the output, we must estimate the glycan production rate from the antibody production rate. First we calculate the integral viable cell density (IVCD) which is calculated as

$$IVCD = \int_0^{t_n} X_v dt \quad 2.18$$

We numerically integrate the viable cell density in MATLAB using the cumulative trapezoidal function. Next, we evaluate the specific antibody productivity rate at different time intervals using

$$q_p|_{t=t_2} = \frac{MAb_{t_2} - MAb_{t_1}}{IVCD_{t_2} - IVCD_{t_1}} \quad 2.19$$

The low antibody titers seen in our cell line imply that the productivity that we observe can be as low as 0.2 pg/cell/day. Next by assuming that each antibody has two glycan binding sites, we calculate the glycan productivity rate as

$$q_{glyc} = q_p \frac{2}{MAb \text{ molecular weight}} \quad 2.20$$

The glycan productivity rate is then passed to the micro-scale model which is described in the following section.

2.3.4 Micro-scale Model Development

The micro-scale model used in the current work is adapted from the DK2011 model while the glycan reaction network is built using the reaction rules and the numbering system originally proposed in the KB2005 model. First, we generate the glycan reaction network using the reaction rules listed in Table 2.2. The columns list the enzyme (see Table 1.2 for enzyme names), the associated reaction rule, and the corresponding reaction leading to the formation of a particular glycan. The complete list of experimentally observed glycan isoforms and their structures are listed in Appendix A. The generation of different glycan isoforms per each reaction rule gets recorded in a nine-digit array, where every number is a unique representation of a glycan isoform. In this representation, *Man*, *Fuc*, and *Gal* refer to the digits corresponding to the number of mannose groups (which can range from 3 to 9), the fucosylation state (0 or 1), and the number of galactose residues (0 to 4) respectively. *Br2* and *Br4* correspond to the extension level in branch 2 (upper branch attached to the α -1,6-linked mannose of the tri-mannosyl core) and the extension level in branch 4 (lower branch attached to the α -1,3-linked mannose in the tri-mannosyl core), respectively. In generating the reaction network, we have not included the bisecting species (denoted by *Gnb*) which arise due to the action of GnTE enzyme, nor do we account for sialyl groups (*Sia*), which are added as extensions to the galactosylated branches by the action of SiaT.

Table 2.2: Reaction rules for generating the glycosylation reaction network based on KB2005 model

Enzyme	Rule	Reaction
ManI	Man>5	Man = Man-1;
ManII	Man>3 && Br4==1 && Man~4	Man = Man-1;
ManII	Man>3 && Br4==1 && Man==4	Man = Man-1;
FucT	Fuc==0 && Br4>0 && Gal==0 && Man==3	Fuc = Fuc+1;
GnTI	Br4==0 && Man==5	Br4 = Br4+1;
GnTII	Br2==0 && Man<4 && Br4==1 && Gnb==0	Br2 = Br2+1;
GalT	(Br2==1) && ~(Br2>0) && Man<4	Br2 = Br2+1 or Br4 = Br4+1; Gal = Gal+1;
GalT	(Br2==1 && Br4 ==1) && (Br2>0)	Br2 = Br2+1 or Br4 = Br4+1; Gal = Gal+1;

To generate the reaction network, we start with nine mannose oligosaccharide (labeled M9) which gets converted to a five mannose glycoform (M5) by the action of ManI enzyme, as per the first rule. To the M5 glycoform, the enzyme GnTI adds an N-acetylglucosamine (*GlcNAc*) at branch 4, i.e. at the α -1,3-mannose on the trimannosyl core. Accordingly, the nine-digit glycan modifier gets edited to account for the addition of the GlcNAc at Br4. Subsequently, other glycan isoforms are created based on the different reaction rules, leading to the formation of a glycan reaction network with 18 glycan isoforms participating in 20 reactions.

Having defined a complete glycan reaction network suitable for our purpose, we then calculate the complete glycosylation distribution profile using an adapted

version of the DK2011 model. Based on the cisternal maturation model, we approximate the Golgi apparatus as a single PFR system with no axial dispersion, constant linear velocity, and constant Golgi diameter to get the oligosaccharide mass balance as:

$$\frac{\partial[\text{Glyc}_j]}{\partial t} = -\frac{4q}{\pi D^2} \frac{\partial[\text{Glyc}_j]}{\partial z} - \sum v_{i,j} r_j \quad 2.21$$

where the kinetic rate expression r_j for each enzyme is as listed in Table 2.3.

Table 2.3: Kinetic mechanisms for different glycosylation enzymes

Enzyme	Kinetic mechanism	Rate expression
ManI, ManII	Michaelis-Menten with competitive and product inhibitions	$r_j = \frac{k_{f,j}[E_j][P_i]}{K_{m,i} \left(1 + \sum \frac{[P_k]}{K_{m,k}} \right)}$
GnTI, GnTII, GalT	Sequential order Bi-Bi kinetics with competitive and product inhibitions	$r_j = \frac{k_{f,j}[E_j][P_i][\text{UDP} - S_i]}{K_{m,i}K_{md,i} \left(1 + 2 \frac{[\text{UDP} - S_i]}{K_{md,i}} + \frac{[\text{UDP} - S_i]}{K_{md,i}} \sum \frac{[P_k]}{K_{m,k}} \right)}$
FucT	Random order Bi-Bi kinetics with competitive and product inhibitions	$r_j = \frac{k_{f,j}[E_j][P_i][\text{UDP} - S_i]}{K_{m,i}K_{md,i} \left(1 + 2 \frac{[\text{UDP} - S_i]}{K_{md,i}} + \left(1 + \frac{[\text{UDP} - S_i]}{K_{md,i}} \right) \sum \frac{[P_k]}{K_{m,k}} \right)}$

Next, as proposed in DK2011 model, we calculate the enzyme distribution along the length of the Golgi apparatus assuming it is normally distributed using

$$E_j(z) = E_{j,max} e^{-\frac{1}{2} \left(\frac{z - z_{j,max}}{w} \right)^2} \quad 2.22$$

The initial simulations are carried out using values provided in literature for peak enzyme concentrations (E_{max}), the mean value along the length of the Golgi (z_{max}), and the standard deviation (w) that are used in equation 2.22.

Next, we solve for the glycan distribution profile using the method of lines by dividing the length of the Golgi into different grid points. The glycan concentration at the entrance to the Golgi (or the first grid point) is the boundary condition, while we define initial concentrations for the glycan species in all the regions of the Golgi. Using the glycan productivity rate, we estimate the amount of glycans entering the Golgi and solve to obtain the dynamic glycan distribution profile the end of the batch culture. As the experimentally reported values are available to us as relative glycan percentages, we convert the glycosylation profile obtained from our simulations to a relative glycan distribution and compare the model simulations at the end of the batch with the experimentally observed glycan distribution. We then optimize the kinetic parameters in the model and the optimized kinetic parameters for the peak enzyme concentration and for the kinetic constants as listed in Tables 2.4 and 2.5, respectively. The model simulations and output glycan distribution is shown in Figure 2.5.

Table 2.4: Parameters to estimate the distribution of different glycosylation enzymes along the length of the Golgi apparatus

Enzyme	E _{max} (μM)	z _{max}	ω
FucT	0.16	11.60	1.51
GalT	1.00	17.15	0.90
GnTI	2.52	8.02	1.56
GnTII	1.07	10.94	1.56
ManI	2.45	5.63	1.57
ManII	1.00	8.58	1.15

Table 2.5: Kinetic constants used in the rate expressions for different glycosyltransferase enzymes arranged as per the reaction rule.

Enzyme	Rule	k_f , min ⁻¹	K_m , mM	K_{md} , mM
ManI	Man > 8	2734	60.5	0
ManI	Man > 7 && ~(Man>8)	2005	110	0
ManI	Man > 6 && ~(Man>7)	792	30.8	0
ManI	Man > 5 && ~(Man>6)	70	625	0
ManII	Man>3 && Br4==1 && Gnb==0 && Man~=4	1026	20	0
FucT	Fuc==0 && Br4>0 && Gnb==0 && Gal==0 && Man==3	225	250	46
GnTI	Br4==0 && Man==5	2768	260	170
GnTII	Br2==0 && Man<4 && Br4==1 && Gnb==0	56	190	960

Table 2.5: continued

Enzyme	Rule	k_f , min ⁻¹	K_m , mM	K_{md} , mM
GalT	(Br2==1) && ~(Gnb>0 && Br2>0) && Man<4	25	130	65
GalT	(Br4==1) && ~(Gnb>0 && Br2>0) && Man<4	25	130	65
ManII	Man>3 && Br4==1 && Gnb==0 && Man==4	43	20	0
GalT	(Br2==1 && Br4 ==1) && (Gnb>0 && Br2>0)	904	6280	65
GalT	(Br4==1 && Br4==1) && (Gnb>0 && Br2>0)	2444	6280	65

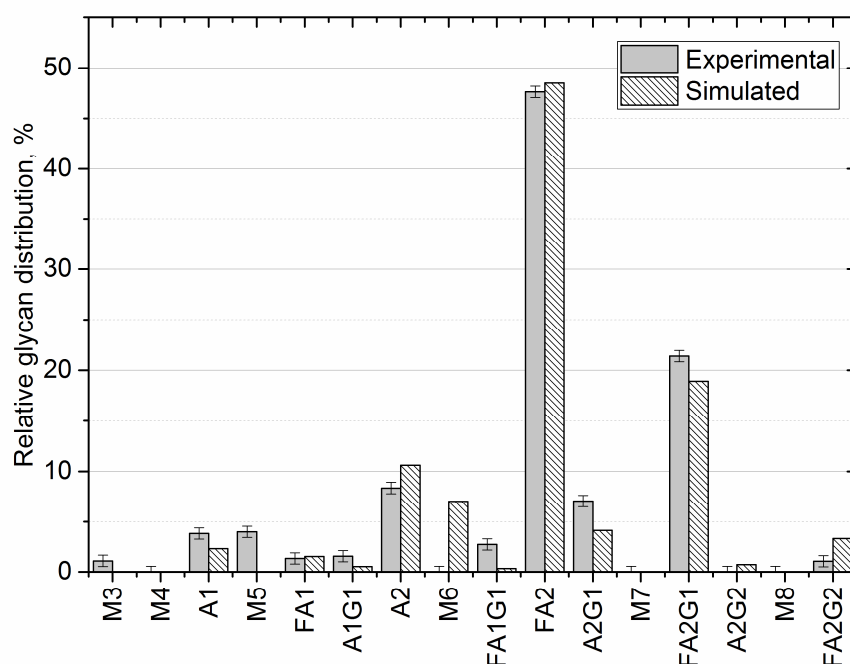


Figure 2.5: Glycan distribution profile obtained from the multi-scale model. Solid grey bars represent the experimentally observed glycan distribution while the cross-hatched bars represent model predictions.

2.4 Validation using Fed-batch Model with Pulse Feeding

2.4.1 Fed-batch Experiments – Materials and Methods

To assess the validity of the model structure and expand its overall applicability, we evaluated the model under fed batch conditions with pulse feeding. As described in § 2.3.1, we grew an IgG1 producing CHO-K1 cell line (gift of Genentech, CA) in suspension culture at 37° C with a 5% CO₂ overlay in vented cap Erlenmeyer shake flasks starting with 50 mL custom CD OPTICHO™ media (Thermo Fisher, MA) supplemented with 28 mM glucose, 4 mM glutamine, and an initial seeding density of 0.5 x 10⁶ cells/mL. Cell count measurements were taken using a

Nexcelom Cellometer Mini (Nexcelom Bioscience, MA), while metabolite (glutamine, glucose, glutamate and lactate) concentrations, media pH, and osmolality were measured using a Bioprofile 100+ analyzer (Nova Biomedical, MA). The cells were grown under batch conditions till day 4, after which 5 mL custom CD OPTICHO™ media supplemented with 28 mM glucose, 4 mM glutamine was added on a daily basis. Samples were withdrawn before and after every feeding and cell count and nutrient concentration measurements were noted. Antibody quantification, purification and subsequent glycan release, permethylation, and analysis was carried out as per the protocol listed earlier.

2.4.2 Modifications to Multi-scale Model

To account for the pulse addition at fixed time intervals, we model the time between the feeding as individual batch runs (Xing et al. 2010) and adjust the nutrient concentration at the start of each new batch by accounting for the volumetric change in the bioreactor due to the addition of fresh media daily. Thus the change in nutrient and metabolite concentrations after feeding is

$$[N_{\text{post}}] = ([N_{\text{pre}}] \times V_{\text{pre}} + 5 \times [N_{\text{feed}}])/V_{\text{post}} \quad 2.23$$

$$[M_{\text{post}}] = ([M_{\text{pre}}] \times V_{\text{pre}})/V_{\text{post}} \quad 2.24$$

where V_{pre} refers to the volume before adding feed and V_{post} is the volume after adding the feed. As the batch and fed-batch experiments were performed under slightly different conditions, we are able to match the new nutrient and cell viability profile as

well as the antibody concentration and the glycan distribution profile as shown in Figures 2.6 through 2.8 by making minor adjustments to the kinetic parameters.

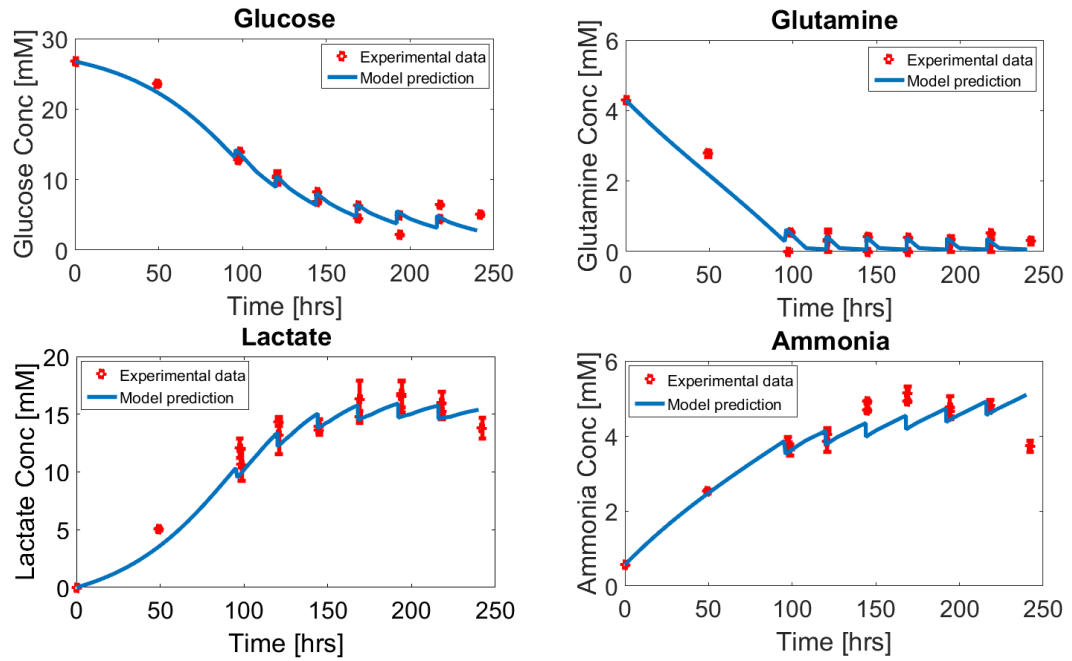


Figure 2.6: Nutrient and by-product concentration profiles for fed-batch culture with daily 5 mL pulse feeding. The plot shows the model fit for glucose, glutamine, lactate, and ammonia. The blue solid line represents the model predictions, while the red circles are the average experimental measurements taken from two trials. The error bars indicate the range of experimental values.

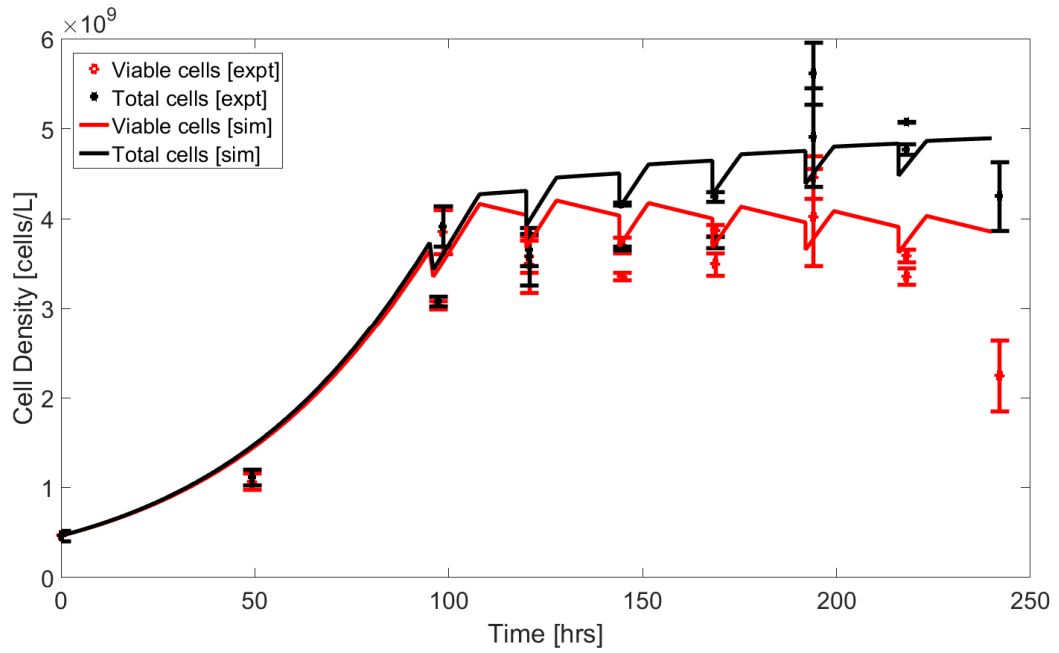


Figure 2.7: Viable cell density and total cell density profiles for fed-batch culture with daily 5 mL pulse feeding. The solid black line represents model predictions for total cell density while the solid red line represents model predictions for viable cell density. The black and red circles represent the average of the experimentally observed total cell density and viable cell density respectively, while the error bars represent the range of observed values (n=2)

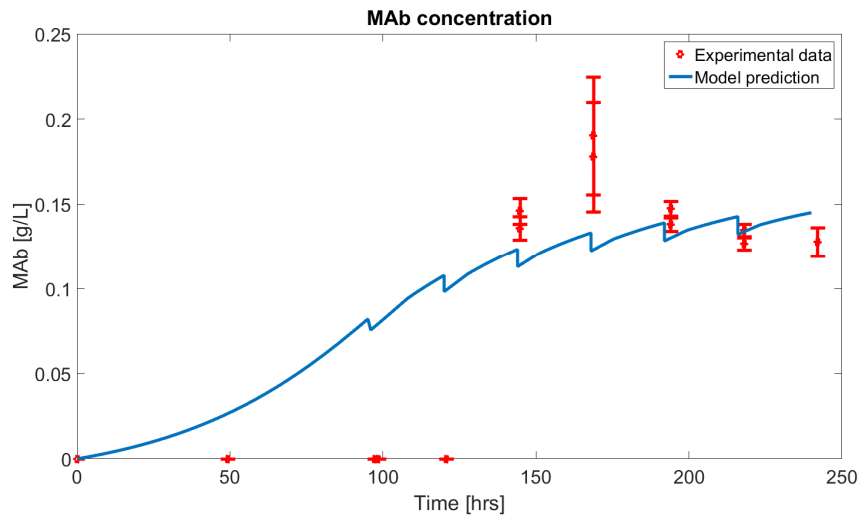


Figure 2.8: Antibody concentration profile for fed-batch culture with 5 mL daily pulse feeding. The solid blue line represents the model simulations while the red circles represent the average experimental data obtained from two trials. The error bars represent the range in the experimental values observed.

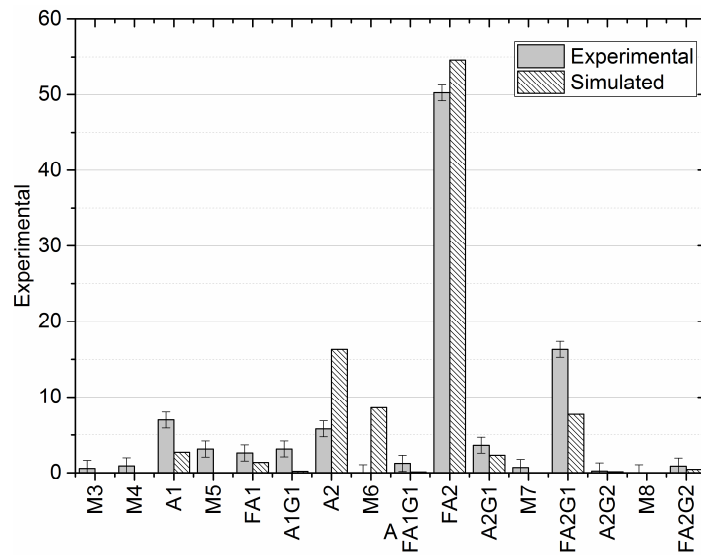


Figure 2.9: Glycan distribution profile obtained from the multi-scale model for fed-batch case. Solid grey bars represent the experimentally observed glycan distribution while the cross-hatched bars represent model predictions

Thus the developed mathematical model is useful in evaluating the glycan distribution under different conditions.

2.5 Summary and Conclusions

In this chapter, we developed a multi-scale model to describe cell growth kinetics and antibody glycosylation for an in-house cell line under batch and fed-batch conditions and compared the model predictions to experimental data. Our model captured the growth profiles at the macro-scale model and predicted the antibody productivity rate based on the total glucose consumption rate. The glycan productivity rate was then calculated from the antibody productivity rate and used in an adapted micro-scale model wherein the Golgi compartment was modeled as a plug flow reactor and the dynamic glycan distribution was obtained by solving the mass balance for each glycan species. The model parameters were optimized using the data from the batch bioreactor experiments and validated using fed-batch data. The models developed in this chapter will be used in Chapter 4 to test the design of batch to batch controllers for glycosylation control using the enzyme concentrations as varying inputs in the micro-scale model and in Chapter 5 to develop a state estimation scheme for predicting the glycan distribution profile in the absence of real time measurements. Chapter 6 will touch upon the different aspects of the model that can be improved in successive iterations.

While this model captures the effect of cell growth dynamics on the glycosylation profile, we note that the effects of cell culture media, another major macro-scale factor that affects glycosylation, have not been included in the predictive model. One of the prime reasons for not including cell culture media in our model is due to its highly complex composition which due to proprietary concerns, is often

unknown to the end user. In the following chapter, we use a parallel approach to understand how different media components affect the glycan distribution, we develop a quantitative relationship between time-dependent media supplementation and the glycosylation profile.

Chapter 3

CONTROLLING THE GLYCOSYLATION PROFILE USING TIME-DEPENDENT MEDIA SUPPLEMENTATION

3.1 Introduction

In Chapter 2, a multi-scale model of glycosylation was developed and tested using in-house batch and fed-batch data. The macro-scale model, based on Monod kinetics, accounted for the effect of cell culture conditions such as concentrations of glucose, glutamine, lactate, and ammonia on cell growth, viability, and the antibody productivity rate, which was subsequently used in the micro-scale model used to generate the glycosylation profile. Although the mathematical model gave a descriptive understanding of the effect of various macro-scale parameters on the cell culture performance, it did not consider a wide range of macro-scale factors that are known to affect glycosylation. For instance, previous studies have demonstrated that protein glycosylation can be influenced by various factors, such as pH (Ivarsson et al. 2014; Yoon et al. 2005), temperature (Ahn et al. 2008; Gawlitzek et al. 2009; Sou et al. 2015), dissolved oxygen (Kunkel et al. 1998; Serrato et al. 2004), ammonia (Borys et al. 1994), and media supplements such as nucleotide sugar precursors (Wong et al. 2010) and manganese chloride (MnCl_2) (Grainger and James 2013; Gramer et al. 2011; Pacis et al. 2011). Modulating the complete glycan distribution profile requires manipulating multiple input factors simultaneously, and to be effective, such action must be based on a thorough, holistic understanding of how these inputs individually and jointly affect various glycan species.

In the absence of detailed first-principles based, mathematical models that describe the effect of all cell culture process variables that can influence glycosylation, we rely upon statistical design of experiments to systematically generate empirical understanding, whereby input factors are judiciously varied simultaneously to generate data on the main and interaction effects they exert on all the output responses of interest. Such structural information indicates which inputs to manipulate, and by how much, in order to alter the relative concentrations of different glycan species appropriately. In most cases, however, the available inputs are fewer than the glycan species to be controlled, resulting in a system with insufficient degrees of freedom. Consequently, we must first answer a fundamental question: given a limited set of inputs, to what extent can we independently control the concentrations of *all* the desired glycan species? In other words, is the desired change in the glycan distribution achievable using the available inputs? We address this question using “controllability analysis”, by which we can determine quantitatively the extent to which the system is controllable. (Informally, a system is considered completely controllable if it is possible to drive the complete set of outputs from some initial value to any arbitrarily specified final, desired value by manipulating the available set of inputs.) Previously, the concept of output controllability was introduced and the controllability of the glycan reaction network was assessed using data generated from statistical design of experiments (St. Amand et al. 2014c). The applicability of controllability analysis was also illustrated by identifying glycan species whose concentrations can be controlled using such media supplements as MnCl_2 , galactose and NH_4Cl as manipulated variables (St. Amand et al. 2014b).

The role of different media supplements in modulating critical quality attributes of the mAb in general, and the glycan distribution profile in particular, has received considerable attention recently (Brühlmann et al. 2015). Typically, supplements such as MnCl_2 , that are known to affect the expression and activity of several glycosyltransferase enzymes, are added to the media at the start of the batch to alter the glycan distribution. However, over the course of the batch run, as the cells continue to grow and produce mAb molecules, changes in the cellular availability of supplements will influence not just the antibody productivity but also the activity of the glycosyltransferase enzymes, thus affecting the final glycan distribution. Hence, we postulate that it is possible to control the glycosylation profile in mAbs by introducing specific media supplements at different stages of cell growth. Specifically, we aim to identify the glycan species that can be controlled by adding MnCl_2 during lag, exponential, and stationary phases of cell growth, and quantify the effect of such time-dependent MnCl_2 additions on the glycan distribution. We postulate further that introducing a chelating agent to the media can alter the effect of MnCl_2 addition on the glycan distribution.

In this chapter we use a mixed factorial experimental design to add MnCl_2 and EDTA at various stages of cell growth and analyze the resulting data appropriately to quantify the effect of time-dependent media supplementation on the glycosylation profile in mAbs. Subsequently, we use controllability analysis to identify the glycan species whose relative percentages can be controlled effectively by introducing MnCl_2 and EDTA to the media at different time points, and quantify the effect of these time-dependent additions. Overall, our results highlight the importance of taking into account the dynamic nature of media supplementation, and presents concepts that can

be exploited to develop new strategies for controlling the glycosylation profile in mAbs.

In the following section, we discuss the experimental and modeling techniques used in this chapter. The results from our work are then presented in section §3.3 with a detailed discussion in section §3.4. The key findings are summarized in the last section of this chapter.

3.2 Materials and Methods

3.2.1 Cell Culture

All experiments were conducted using an IgG1 producing CHO-K1 cell line donated by Genentech, San Francisco, California. The cells were scaled up in a custom CD OptiCHO™ medium formulation (Thermo Fisher Scientific, Waltham, MA) that was supplemented with 4 mM glutamine, 5 g/L glucose and 25 nM MTX. The osmolality was adjusted to 300 mOsm by adding NaCl stock solution. The concentration of MnCl₂ in the media was adjusted using a 0.5 M stock solution (Sigma Aldrich). Similarly, a 0.5 M EDTA sterile stock solution was prepared and added to the media as required. The cells were inoculated with an initial seeding density of 0.5 x 10⁶ cells/mL in vented-cap Erlenmeyer shake flasks with a working volume of 50 mL and grown in batch in suspension in an incubator maintained at 37° C with a 5% CO₂ overlay, with supplements only as indicated by experimental design below. Cell count measurements were taken every two days using a hemocytometer. Metabolite (glutamine, glucose, glutamate and lactate) concentrations, media pH, and osmolality were measured using a Bioprofile 100+ analyzer (Nova Biomedical, Waltham, MA). Antibody titer was measured with an Agilent 1200 HPLC instrument using 1X PBS

buffer on a Thermo Scientific™ MAbPac Protein A chromatography column (12 micron particle size, 35x4.0 mm I.D., Thermo Fisher Scientific, Waltham, MA).

3.2.2 Experimental Design

The shake flask experiments were conducted according to a (2^2 , 3^2) mixed level experimental design for the following factors: (i) MnCl_2 concentration (high and low levels); (ii) EDTA concentration (high and low levels); (iii) time of addition of MnCl_2 (high, intermediate, and low levels); and (iv) time of addition of EDTA (high, intermediate and low levels). The concentration of MnCl_2 in the basal media corresponds to the low level condition (-1) for MnCl_2 , while the high level condition (+1) corresponds to the final concentration of MnCl_2 supplemented media (0.04 mM). Similarly, the low level condition (-1) for EDTA corresponds to “no EDTA” added to the media, while the high level (+1) corresponds to 0.08 mM EDTA added to the media. MnCl_2 and EDTA are added on day 0 (D0), day 3 (D3), or day 6 (D6) after inoculation, corresponding respectively to the low (-1), intermediate (0), and high (+1) levels. Thus, this full factorial mixed level (2^2 , 3^2) experimental design yields a total of 36 different possible shake flask conditions to be tested. However, the 36 conditions are not unique because some of the cases correspond to identical experimental conditions. For instance, 9 of the 36 conditions correspond to MnCl_2 and EDTA at low levels (-1), with the time of addition at low (-1), intermediate (0) and high levels (+1). The low level condition for MnCl_2 represents basal concentrations, while the low level for EDTA represents no EDTA supplementation. Thus, these 9 cases represent

identical conditions where the flask has basal levels of MnCl_2 with no EDTA supplementation either on D0, D3, or D6. A single flask (F1) was used for all nine cases and was treated as the control flask because of what the conditions represent — basal level of MnCl_2 and no EDTA supplementation on any day. It can be shown that there are in fact only 16 unique experimental cases/conditions, as listed in Table 3.1. Each condition was tested with two biological replicates. The glycan distribution profile was determined using the permethylation assay described below, and the resulting relative glycan percentages data obtained for each condition were analyzed in MINITAB using standard analysis of variance (ANOVA) to obtain the factor effects/coefficients and associated p -values.

3.2.3 Glycan Permethylation Assay

On day 8 after inoculation, the cells were centrifuged at 3000 rpm for 10 minutes and the spent media was harvested. The IgG1 antibody was then purified from the spent media using a PhyNexus Benchtop MEA2 system using Protein A chromatography resin packed in a 2 mL PhyTip column (PhyNexus, San Jose, CA). The glycan permethylation assay was then carried out with 100 microgram of the purified antibody using a previously described method (St. Amand et al. 2014b). Briefly, the antibody was first digested with trypsin (Promega, Madison, WI) for four hours in an incubator held at 37 °C, followed by enzymatic deglycosylation using PNGase-F (ProZyme, Hayward, CA) for a minimum of 16 hours at 37 °C. The free separated glycans were captured on Hypersep Hyper Carb SPE cartridges (Thermo

Fisher Scientific, Waltham, MA) and permethylated following the Ciucanu method using methyl iodide and NaOH in the presence of DMSO (Ciucanu and Costello 2003; Ciucanu and Kerek 1984). The permethylated glycans were purified in a liquid-liquid extraction step with chloroform (Sigma Aldrich, St. Lois, MO), dried and resuspended in 80% methanol (Sigma Aldrich, St. Lois, MO). The resuspended glycans were spotted onto a MALDI/TOF plate with a DHB matrix and analyzed using a 4800 MALDI TOF/TOF Analyzer (ABSciex) in positive ion, reflector mode. The data collected using the mass spectrometer was then exported to DataExplorer to obtain the peak heights for the identified glycans (see Appendix A). The relative glycan distribution in each sample was calculated from the sum of the peak heights for all the identified glycans in that sample.

Table 3.1: Experimental conditions tested in mixed factorial design

Experimental condition	MnCl ₂ conc. (mM)	EDTA conc. (mM)	Time of addition of MnCl ₂ †	Time of addition of EDTA †	Label
1	0.01	0	D0	D0	Control
2	0.01	0.08	D0	D0	ED D0
3	0.04	0.08	D0	D0	Mn D0/ED D0
4	0.04	0.08	D3	D0	Mn D3/ED D0
5	0.04	0.08	D6	D0	Mn D6/ED D0
6	0.01	0.08	D0	D3	ED D3
7	0.04	0.08	D0	D3	Mn D0/ED D3
8	0.04	0.08	D3	D3	Mn D3/ED D3
9	0.04	0.08	D6	D3	Mn D6/ED D3
10	0.01	0.08	D0	D6	ED D6
11	0.04	0.08	D0	D6	Mn D0/ED D6
12	0.04	0.08	D3	D6	Mn D3/ED D6
13	0.04	0.08	D6	D6	Mn D6/ED D6
14	0.04	0	D0	D0	Mn D0
15	0.04	0	D3	D0	Mn D3
16	0.04	0	D6	D0	Mn D6

† D0, D3 and D6 refer to Day 0, Day 3 and Day 6 after inoculation, respectively

3.2.4 Glycosylation Index

For each experimental condition, glycosylation indices were calculated from the relative percentages of individual glycan species (Ivarsson et al. 2014; Majid et al. 2007). For example, the galactosylation index (GI), defined as the percentage of mono- and di-galactosylated species in the total glycan distribution, was determined according to the following equation

$$GI = \frac{2 \times G_2 + G_1}{2 \times (G_0 + G_1 + G_2)} \% \quad 3.1$$

where G_0 is the sum of all agalactosylated species, G_1 is the sum of all monogalactosylated species, and G_2 is the sum of all digalactosylated species. Similarly, we calculated the fucosylation index (FI) for each distribution as

$$FI = \frac{F_1}{(F_0 + F_1)} \% \quad 3.2$$

where F_0 and F_1 are the sum of all afucosylated and fucosylated species, respectively.

3.2.5 Controllability Analysis

Using the technique presented in St.Amand et al (2014b), we perform controllability analysis to quantify the effect of time dependent media supplementation on the glycosylation profile in mAbs. Briefly, we note first that estimates of factor coefficients obtained after analyzing the mixed factorial design data correspond to the various “process gains”, defined as the change observed in the glycan distribution (output), $\Delta \mathbf{y}$, in response to a unit change in the input factor with which the coefficient in question is associated. By selecting statistically significant factor coefficients (at the significance level of $\alpha = 0.05$) and setting all non-significant coefficients to zero, we generate the process gain matrix \mathbf{K} so that

$$\Delta \mathbf{y} = \mathbf{K} \Delta \mathbf{u}$$

3.3

where $\Delta \mathbf{u}$ represents the change in the input factor. Singular value decomposition of the process gain matrix produces the diagonal singular value matrix, Σ , and the unitary matrices, \mathbf{W} and \mathbf{V}^T , that are subsequently used to obtain the orthogonal input (μ) and output (η) modes, which, along with the corresponding singular values are used to assess controllability.

3.3 Results

3.3.1 Early Addition of EDTA is Detrimental to Cell Growth and Reduces Antibody Titer

Figures 3.1 and 3.2 show the effect that introducing media supplements (MnCl₂ or EDTA) at different time points had on cell culture parameters and final antibody titer, compared to corresponding results obtained from a control flask (F1) which contained MnCl₂ at basal media concentrations and no EDTA.

Compared to the conditions in the control flask, early addition of EDTA on D0 reduced the cell density significantly, which is consistent with hampered cell growth. However, this effect was offset somewhat by introducing MnCl₂ in addition to EDTA on D0. The viable cell density (VCD) measured on day 4 for samples in which both MnCl₂ and EDTA were introduced on D0, was 1.31×10^9 cells/L, which was nearly three times as large as the value of the VCD in the flasks where only EDTA was added on D0 ($\sim 0.4 \times 10^9$ cells/L). Introducing MnCl₂ on D3 or D6 *after* the addition of EDTA on D0 did not improve the VCD. Similarly, when EDTA was introduced on D3, the VCD in samples with no additional MnCl₂ supplementation dropped sharply. By contrast, when the media was supplemented with MnCl₂ on D0 or D3, the VCD

was slightly higher than when the media was supplemented with just EDTA on D3. Supplementing the samples with MnCl_2 on D6 after the addition of EDTA did not improve the VCD significantly. In all cases, the observed VCD was generally higher than when EDTA was introduced on D0. Addition of EDTA on D6 had no impact on the VCD, regardless of the time of introduction of MnCl_2 . Similarly, as shown in Figure 3.1(d), early addition of MnCl_2 on D0 and D3 reduced the VCD, while addition of MnCl_2 on D6 did not alter the VCD. Thus, in summary, early addition of EDTA reduced cell viability in the absence of MnCl_2 supplementation, but the addition of MnCl_2 by itself did not alter cell viability significantly.

Figure 3.2 shows the effect of media supplementation on antibody titers. The average mAb titer in the control flask F1 (with no MnCl_2 or EDTA supplementation) was 0.13 g/L. The addition of EDTA to the media on D0 in the absence of MnCl_2 supplementation decreased the titer by about a fourth, to 0.03 g/L. This decrease in the titer was marginally offset when MnCl_2 was introduced on D0, D3, or D6, with earlier MnCl_2 supplementation resulting in higher titers than later supplementation. When EDTA was introduced on D3, the resulting titer was 0.10 g/L, three times higher than that observed with EDTA supplementation on D0. Further supplementing the media with MnCl_2 on D0, D3, or D6, increased the titer observed with EDTA supplementation on D3 to values comparable to that of the control case. Supplementing the flasks on D6 with EDTA increased titers even further to 0.15 g/L, beyond values obtained in the control case. The titer values were also higher when EDTA supplementation on D6 was combined with MnCl_2 supplementation on D0 (0.17 g/L), D3 (0.17 g/L) or D6 (0.16 g/L). Finally, MnCl_2 supplementation alone on D0, D3, or D6 increased the titer to an average of 0.14 g/L. Thus while early EDTA

supplementation has an adverse effect on the antibody titer, late EDTA addition improves the final titer. Consequently, we conclude that the addition of MnCl_2 to the cell culture media or to EDTA supplemented media enhances the titer.

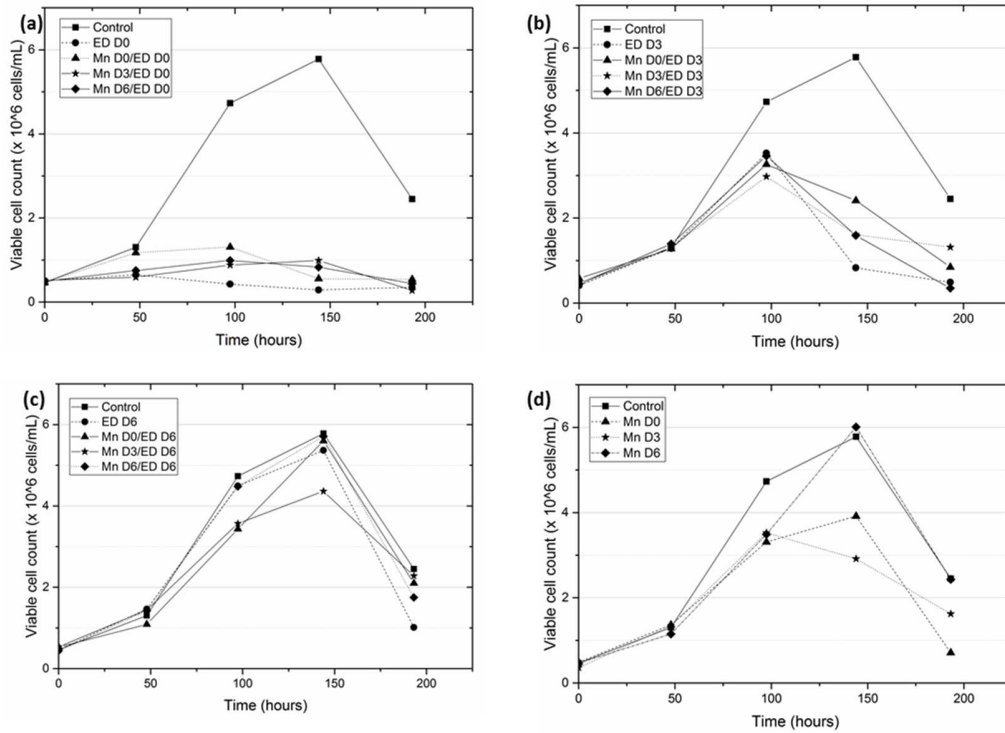


Figure 3.1: Viable cell concentration data for CHO-K1 cells under different media supplementation conditions. (a) When EDTA is added by itself on D0 (●), or in the presence of MnCl_2 on D0 (▲), D3 (*), and D6 (◆); (b) When EDTA is added by itself on D3 (●), or in the presence of MnCl_2 on D0 (▲), D3 (*), and D6 (◆); (c) When EDTA is added by itself on D6 (●), or in the presence of MnCl_2 on D0 (▲), D3 (*), and D6 (◆); and (d) When MnCl_2 is added on D0 (▲), D3 (*), and D6 (◆)

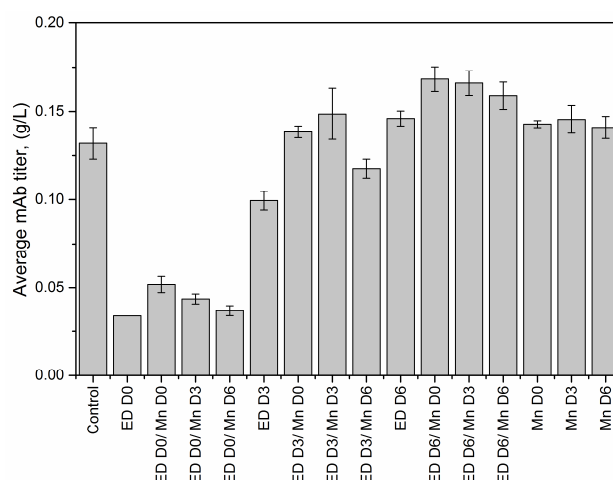


Figure 3.2: Average antibody titer for the 16 experimental conditions. Mn and ED refer to the media supplements MnCl_2 and EDTA, while D0, D3 and D6 refer to Day 0, Day 3 and Day 6 after inoculation, respectively. Error bars represent the range of biological replicates (n=2).

3.3.2 Early Addition of MnCl_2 Alters the Glycan Distribution Significantly

The addition of EDTA and MnCl_2 at different time points altered the glycan distribution, with earlier addition of MnCl_2 having the more significant effect. Figure 3.3 shows the effect of media supplementation on the glycan distribution, with panels 3.3(a), (b) and (c) showing the changes in the glycan profile as a result of media supplementation with EDTA on D0, D3, and D6, with or without additional MnCl_2 supplementation, and 3.3(d) showing the impact of MnCl_2 addition in the absence of EDTA supplementation.

Adding EDTA to the cell culture media on D0 (Figure 3.3(a)) decreased the amount of biantennary fucosylated species, FA2, by 5.58%, which was offset by a concomitant increase in its galactosylated isoforms, FA2G1 (4.52%) and FA2G2 (1.52%). The mannosylated species M5 also increased by 1.52% with a corresponding reduction of 1.79% in the concentration of the biantennary group A2 (which is

produced from M5 in the Golgi compartment). Adding MnCl_2 on D0, D3, or D6 to media in which EDTA was introduced on D0 resulted in a further increase in the relative concentrations of the galactosylated isoforms FA2G1 and FA2G2, in addition to a decrease in the relative percentage of FA2, A2, and A2G1 species, compared to the distribution in the control flask. However, the increase in FA2G1 and FA2G2 was more pronounced when MnCl_2 was introduced on D3 or D6 after adding EDTA. When EDTA was added on D3 (Figure 3.3(b)) the relative percentages of the biantennary species FA2, FA2G1, and A2G1 decreased, while the relative percentages of glycan species M5, A1, and FA1 increased. When MnCl_2 was added on D0 followed by EDTA supplementation on D3, the concentrations of M5 and FA1 increased, while the concentration of FA2 decreased. Additionally, there was an increase in the relative percentage of A2G1 that was offset by the decrease in the concentration of A1G1. When both EDTA and MnCl_2 were added to the media on D3, the change in the glycan distribution mirrors the trend observed when only EDTA was introduced on D3. The main difference occurs when MnCl_2 was added to the media *after* EDTA addition, i.e., on D6. Here, the relative percentages of M5, FA1, FA2G1, and FA2G2 increased while the relative percentages of A2, A2G1, and FA2 decreased, similar to the trends observed with the addition of MnCl_2 on D3 or D6 following EDTA supplementation on D0. Although the late addition of EDTA on D6 of the cell culture did not affect the glycan profile significantly (Figure 3.3(c)), for those glycan species whose relative percentages change, the trend was similar to that observed with EDTA supplementation on D0 and D3, i.e., an increase in the relative percentage of M5 and A1 accompanied by a decrease in that of FA2 and FA2G1. When the media was first supplemented with MnCl_2 on D0 or D3, followed by EDTA supplementation

on D6, the relative percentage of FA2 species decreased while the relative percentages of M5, A2 and FA2G1 species increased. A simultaneous addition of both EDTA and MnCl_2 toward the end of the batch increased the relative percentage of A2G1 while decreasing the relative amount of its fucosylated isoform FA2G1.

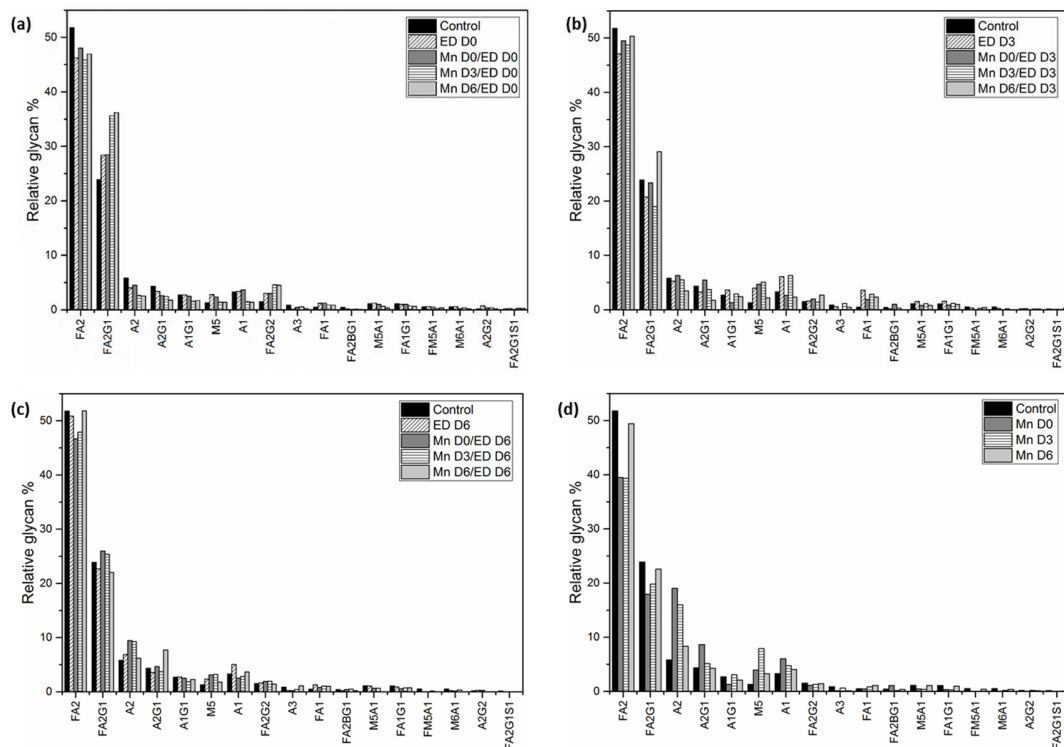


Figure 3.3: Average relative glycan percentage of IgG1 glycans produced in CHO-K1 cells under different experimental conditions. (a) When EDTA is added on D0 with no MnCl_2 supplementation or with MnCl_2 supplementation on D0, D3, and D6; (b) When EDTA added on D3 with no MnCl_2 supplementation or with MnCl_2 supplementation on D0, D3, and D6; (c) When EDTA added on D6 with no MnCl_2 supplementation or with MnCl_2 supplementation on D0, D3, and D6; and (d) When MnCl_2 is added on D0, D3, and D6

The solo effect of MnCl_2 on the glycan profile was observed by supplementing the media with MnCl_2 in the absence of EDTA on D0, D3 or D6 (Figure 3.3(d)). Adding MnCl_2 on D0 decreased the relative percentage of the FA2 species from the control value of 51.79% (with a 95% confidence interval range of $\pm 0.94\%$) to 39.50% ($\pm 1.55\%$). Similarly, the relative percentage of the monogalactosylated species FA2G1 drops from the control value of 23.87% ($\pm 0.94\%$) to 17.95% ($\pm 1.55\%$). This decrease in the fucosylated species was offset by an increase in the biantennary species A2 (which increased by 13.24%), the mannosylated species M5 (2.62%), and the galactosylated biantennary species A2G1 (4.29%). A similar trend was observed when MnCl_2 was introduced to the culture medium on D3, with the decrease in the fucosylated species being offset by a significant increase in the biantennary species A2 (which increased by 10.20%), the mannosylated species M5 (increasing by 6.59%) and only a marginal increase in the A2G1 species (by 0.76%). Adding MnCl_2 during the peak exponential phase (D6) resulted in similar changes to the glycan distribution, but these changes were smaller in comparison to the changes in the glycan distribution due to earlier addition of MnCl_2 . The concentrations of the fucosylated species FA2 and FA2G1 decreased by 2.36% and 1.28% respectively, with the corresponding increases in the concentrations of A2 and M5 species being 2.52% and 1.94%, respectively. While such qualitative discussions of the glycan distribution profiles may be instructive, for our purposes, a quantitative analysis relating the changes in the experimental conditions to the observed changes in the glycan profile is more informative.

3.3.3 The Type of Media Supplement and the Time of Addition both showed Statistically Significant Effects on Glycan Distribution

The concentrations of MnCl_2 and EDTA and the time of addition of each one constitute the four factors in the $(2^2, 3^2)$ mixed factorial experimental design. A full factorial experimental design gives rise to 36 experimental conditions and the resulting model in principle consists of 35 main and interaction effects. The statistically significant factor coefficients (at the significant level of $\alpha = 0.05$) estimated using ANOVA, were used to generate a “gain” matrix, whose elements represent “by how much” each output variable (relative glycan distribution) will change in response to a unit change in each input factor, including multi-factor interactions (since multiple combinations of single inputs are considered as valid inputs in this case). However, because the full factorial experiment consists of only 16 unique cases (see Materials and Methods), we must eliminate the redundant rows in our gain matrix resulting in a reduced gain matrix with 15 input factors, some of which are multi-factor combinations (see Appendix B for the complete gain matrix and the interactions) Figure 3.4 shows a heat map of the elements of the gain matrix, indicating which input factor affects which glycan, and the magnitude as well as the direction (increase or decrease) of each effect.

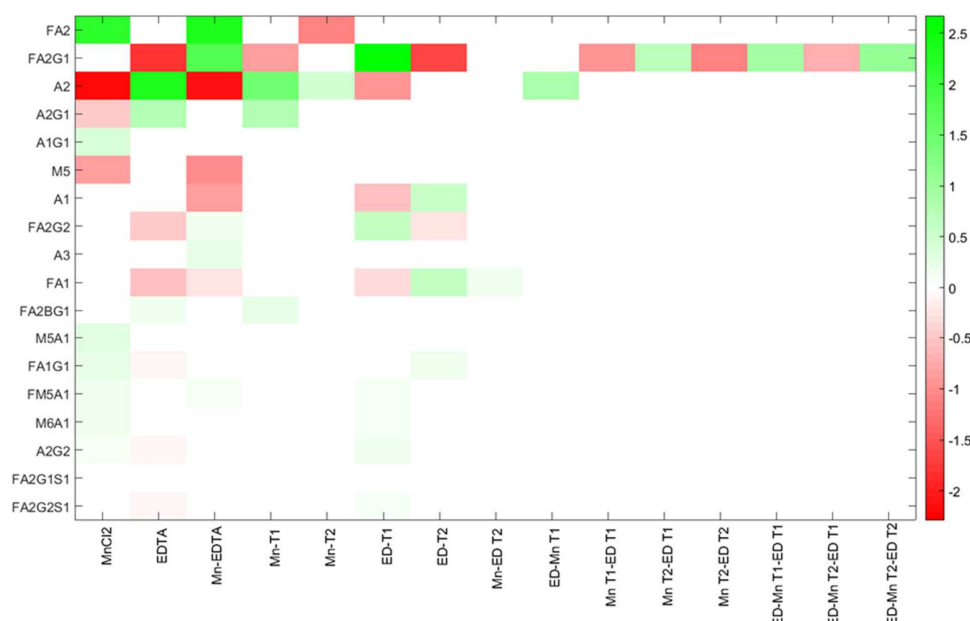


Figure 3.4: Heat map of significant factor coefficients ($\alpha = 0.05$) obtained from ANOVA. The input factors are listed along the horizontal axis, and individual glycan species along the vertical axis. The color red indicates a decrease in the concentration of a particular glycan while green indicates an increase. The color intensity represents the magnitude of the significant coefficient, with increasingly darker hues indicating increasingly larger magnitudes and progressively lighter hues indicating commensurately lower magnitudes.

An examination of the heat map indicates that the glycan species FA2, FA2G1, and A2 are affected by most of the factors and their combinatorial interactions. For instance, the concentration of the most abundant glycoform, FA2, (which accounts for nearly 52% of the glycan concentration in the control sample) is affected by the concentration of MnCl_2 , the two-way interaction of MnCl_2 and EDTA, and the late stage addition of MnCl_2 . In particular, a unit change in the concentration of MnCl_2 causes an increase in the average concentration of FA2 (as indicated by the positive coefficient for the MnCl_2 effect), while introducing MnCl_2 on D3 causes a reduction in

the average concentration of FA2. By contrast, the monogalactosylated form FA2G1 is not affected by changes in the media concentration of MnCl_2 ; rather it is influenced by changes in the concentration of EDTA, the early addition of MnCl_2 and the two-way interaction of MnCl_2 and EDTA. Further, we observe that two of the interaction factors, Mn T1-ED T1 and Mn T2-ED T1, have statistically significant and opposing effects on the concentration of FA2G1. The factor Mn T1-ED T1 represents the two-way interaction of adding MnCl_2 on D0 and adding EDTA on D0, whereas Mn T2-ED T1 represents the two-way interaction of adding MnCl_2 on D3 when EDTA is introduced on D0. We observe therefore that the concentration of FA2G1 is affected by multiple input factors, including complex interactions between the amount of supplements added, and the times of addition of the supplements. One is thus able to assess the impact of each input factor on the response of all other individual glycan species in similar fashion. As indicated by the heat map, the effects of higher order interactions on most of the glycan species are negligible (if they exist at all) because estimates of the coefficients associated with most interaction effects are statistically insignificant.

3.4 Discussion

While media supplementation has been studied widely for its effect on the quality attributes of mAbs, such studies have been limited to the introduction of media supplements exclusively at the start of the culture, and the results, when quantitative, have yielded only isolated single factor relationships. The results from the current study show that introducing media supplements at different time points during cell culture does in fact have an effect on cell growth conditions and antibody

glycosylation distribution, and the effects can be quantified globally and potentially used to design effective control schemes.

Specifically, we have shown that earlier addition of EDTA is detrimental to cell growth and results in a decrease in antibody titer. When EDTA was added on D0 (at the inoculation stage) the peak viable cell density (VCD) remained close to the seeding density, indicating a hampering of cell growth. This result is consistent with the well-known fact that EDTA is toxic to cells (Kim et al. 2009b). The decline in the viable cell densities due to EDTA addition during the growth phase can also be attributed to the removal from the media (as a result of EDTA's chelating effect) such trace metals as Ca^{++} , Zn^{++} , etc. that are essential for cell survival. By contrast, when both EDTA and MnCl_2 were added on D0, the peak VCD increased two-fold. While this peak VCD is lower than the peak VCD of the control flask (5.78×10^6 cells/mL), the increased viability can be attributed to the presence of excess MnCl_2 titrating EDTA, resulting in reduced cytotoxicity. The increase in the cell viability due to the simultaneous addition of MnCl_2 and EDTA also resulted in higher mAb titers, compared to the titers observed when only EDTA was added.

When EDTA was added on D3, the cells were in the middle of the growth phase and the addition of the cytotoxic EDTA hampered further growth, leading to a steep decline in the cell viability beyond D3 (Figure 3.1(b)). By contrast, when MnCl_2 was added to the media (on D0 or D3) in the presence of EDTA, the cells did not experience a similar reduction in viability. As noted previously, simultaneous addition of EDTA and MnCl_2 also resulted in an improvement in the titer to a value comparable to that in the control flask. When EDTA was added on D6, the cells were already at the end of the growth phase and hence the introduction of EDTA did not

alter the cell viability. However, when EDTA was added on D6 along with MnCl_2 supplementation on D0, D3 or D6, the final titer values were higher than the titer value in the control flask. The increase in antibody titer in the presence of EDTA has been observed by others as well (Kao et al. 2010; McAtee et al. 2014) and is attributed to the inhibition of antibody reduction during cell lysis. Thus, later addition of EDTA combined with early addition of MnCl_2 could potentially improve titer without compromising cell viability. Further, analyzing the EOR titer data using ANOVA shows that the factor coefficients for the concentration of EDTA, the concentration of MnCl_2 , and time of addition of EDTA are statistically significant (at a significance level of $\alpha = 0.05$). The expected change in the EOR titer in response to a unit change in any of these factors is quantified by the magnitude of that factor coefficient, while the sign of the factor coefficient indicates the direction of change. Thus, for example, a unit positive step change in the concentration of EDTA (with a factor coefficient of -0.017) or time of addition of EDTA (-0.032) results in a decrease in EOR titer; a unit positive step change in the concentration of MnCl_2 (0.008) causes an increase in titer. Quantifying the effect of these input factors on the EOR titer provides a rational basis for selecting what particular supplement to add, and how much of it to add, at a given stage of cell culture in order to maximize product yield. However, any media supplementation strategy must meet not just the desired specifications for final titer but also for product quality, i.e., to be effective, the implemented media supplementation strategy must not alter the glycan distributions significantly.

The EOR titer represents the *total* amount of antibody accumulated at the end of the batch, while the measured glycan distribution represents the *relative* amount of each individual glycan isoform that has accumulated over the duration of the batch.

Now, the relative amount of individual glycan species is a function of antibody productivity and it changes over the course of the batch. In our case, the addition of different media supplements at different stages of batch cultures affected both viability and antibody titer; consequently, the observed change in the glycan distribution has been induced by dynamic media supplementation and changes in productivity. Therefore, to develop a mechanistic understanding of the effect of dynamic media supplementation on the glycosylation profile, we first decouple the effect of antibody productivity on the glycan distribution from the overall change observed at the end of the batch. One such decoupling approach involves estimating the mass fractions of specific glycoforms produced at different stages of cell culture using the expression (Fan et al. 2015b)

$$f_i = \frac{[mAb_i]_{t=t_1} - [mAb_i]_{t=t_2}}{[mAb_{Tot}]_{t=t_1} - [mAb_{Tot}]_{t=t_2}} \quad 3.4$$

for f_i , the fraction of mAb glycoform i produced in the time period $[t_1, t_2]$ relative to the total amount of antibody secreted in the same period. However, we cannot use this expression for our purpose because we only measured EOR titer and final glycan distribution, not intermittent antibody titer or glycan concentration. Consequently, we propose an alternative metric based solely on the final titer and glycan measurements.

To illustrate, consider the glycan distribution in the control flask and in the flask with $MnCl_2$ added on D6. In both flasks, the cell growth profile and antibody productivity will be the same until day 6, when $MnCl_2$ is introduced to the latter flask.

Thus the amount of i^{th} glycoform fractions accumulated between day 6 (D6) and the end of the run (EOR) for the two flasks can be written as

$$f_{Mn6/i} = \frac{[mAb_{Mn6/i}|_{t=EOR}] - [mAb_{Mn6/i}|_{t=D6}]}{[mAb_{Mn6/Tot}|_{t=EOR}] - [mAb_{Mn6/Tot}|_{t=D6}]} \quad 3.5$$

$$f_{control/i} = \frac{[mAb_{control/i}|_{t=EOR}] - [mAb_{control/i}|_{t=D6}]}{[mAb_{control/Tot}|_{t=EOR}] - [mAb_{control/Tot}|_{t=D6}]} \quad 3.6$$

Recognizing that the D6 values in equations (5) and (6) above are identical, we can eliminate the intermittent time point with simple arithmetic manipulations and obtain the *change* in the accumulation of the i^{th} glycoform based solely on EOR values, as

$$\Delta f_i = \frac{[mAb_{Mn6/i}|_{t=EOR}] - [mAb_{control/i}|_{t=EOR}]}{[mAb_{Mn6/Tot}|_{t=EOR}] - [mAb_{control/Tot}|_{t=EOR}]} \quad 3.6$$

Thus, this fractional difference allows us to group together different experimental conditions with similar antibody titers, making it possible to compare final glycan distributions and hence quantify the effect of individual media supplements on the glycan profile appropriately. Such analyses extended to other experimental conditions yield the comparative fractional difference in the glycoform distribution shown in Figure 3.5.

A comparison of the fractional difference in the glycan distribution in flasks where MnCl_2 is added to D0 EDTA supplemented flasks relative to the glycan distribution in D0 EDTA supplemented flasks (figure 5(a)), shows the following: an increase in the amount of FA2 (by nearly ~ 50% in all cases), FA2G1 (by 29% during D0 supplementation, 60% during D3 supplementation, and 122% during D6 supplementation), and FA2G2 (by 3%, 10%, and 20% respectively), with a relative decrease in A2 and M5 by 2% and 4% when MnCl_2 is added on D3, and nearly 14% when MnCl_2 is added on D6. A similar trend is observed in the fractional difference in the glycan distribution in flasks where MnCl_2 is added on D3 and D6 to D3 EDTA supplemented flasks (figure 5b). Here the fractional difference in the glycan distribution is calculated relative to the glycan distribution observed when EDTA is added on D3. Again, we notice an increase in FA2, FA2G1 and FA2G2, with a decrease in A2, M5, and A2G1 observed only when MnCl_2 is added on day 6 *after* the addition of EDTA on D3. Previous studies have shown that adding MnCl_2 produces an upregulation of galactosyltransferase enzymes (St. Amand et al. 2014b), and subsequently in increased galactosylation (Grainger and James 2013; Gramer et al. 2011). Hence, the increase in the amount of FA2G1 and FA2G2 species can be attributed to the effect of late stage manganese addition on the galactosyltransferase enzyme. Figure 3.5(c) shows fractional difference when EDTA is added on D6 after MnCl_2 has been added to the culture on D0 and D3. These differences are calculated relative to the glycan distribution observed due to the addition of MnCl_2 on D3 and D6, respectively. We note that the fractional difference in the fucosylated species FA2 and FA2G1 is positive when EDTA is added after MnCl_2 supplementation, indicating that the addition of EDTA increases the concentration of these species relative to their

respective concentrations in MnCl_2 supplemented cultures. Also, a comparison of the fractional difference in the glycan distribution when MnCl_2 was introduced on D0, D3, or D6 relative to the glycan distribution in the control flask (figure 3.5(d)), shows that the relative concentrations of FA2 and FA2G1 species decreased in flasks with D0, while late stage addition of MnCl_2 did not have a significant effect on the overall glycan distribution. Taken together, our findings indicate that the latter addition of EDTA reverses the changes in the glycan distribution induced by MnCl_2 .

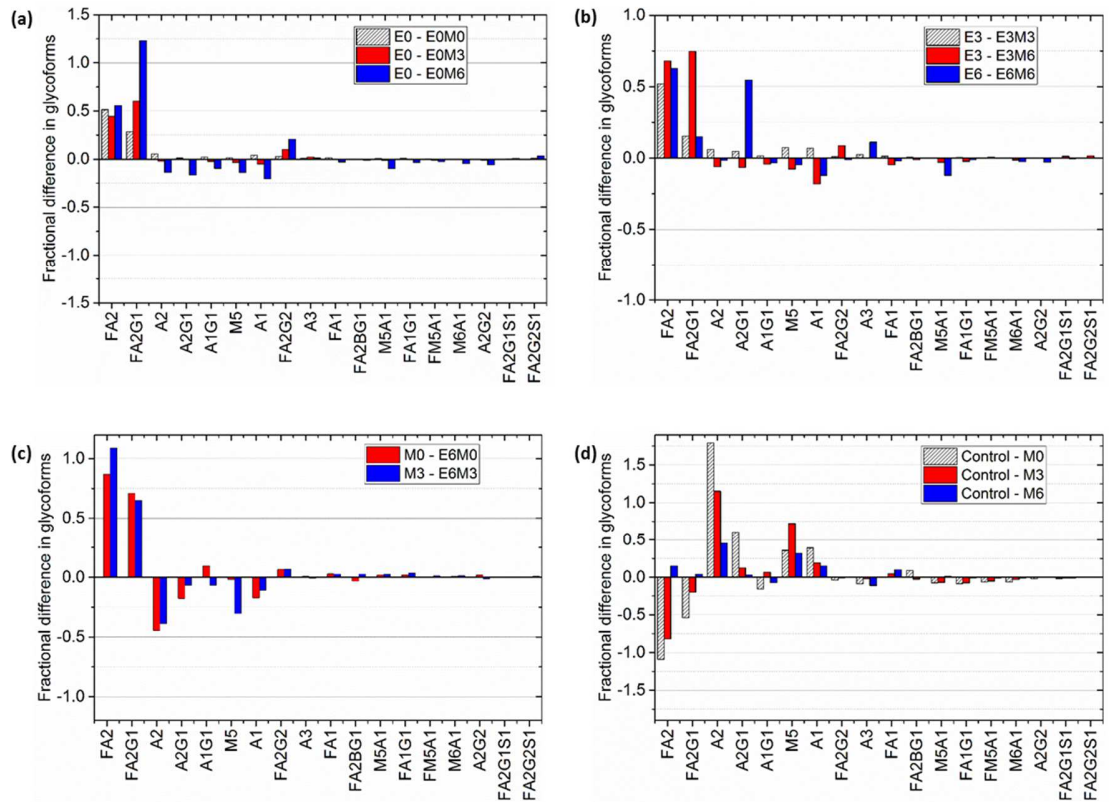


Figure 3.5: Comparing fractional differences in the glycoforms for different experimental cases. (a) When MnCl_2 is added on D0 (E0M0), D3 (E0M3), and D6 (E0M6) to EDTA D0 supplemented cultures. The fractional changes are calculated relative to the glycan distribution in EDTA D0 (labeled E0) cultures and no MnCl_2 supplementation. (b) When MnCl_2 is added on D3 (E3D3), and D6 (E3D6) to EDTA D3 supplemented cultures and on D6 (E6D6) to EDTA D6 supplemented cultures. The fractional changes are calculated relative to the glycan distribution in EDTA D3 (E3) and D6 (E6) cultures respectively. (c) When EDTA is added on D3 (E3M0) and D6 (E6M0) to MnCl_2 D0 cultures and on D6 (E6M3) to MnCl_2 D3 cultures. The fractional changes are calculated relative to the glycan distribution in MnCl_2 D0 and D3 (M3) cultures respectively. (d) when MnCl_2 is added on D0, D3, and D6 to control cultures. The fractional changes are calculated relative to the distribution in the control case.

Fractional difference analysis helps to identify which glycan species are altered as a result of the addition of specific media supplements, but not *why* those particular glycan species changed. To identify the kinetic mechanisms underlying these changes observed in the glycan distribution upon adding MnCl_2 to the media, we use an existing dynamic mathematical model of glycosylation previously described in §2.3.3 to test the proposition that the changes observed in the glycan distribution due to the addition of MnCl_2 are induced by changes in the concentrations of the glycosyltransferase enzymes. First we matched the simulated glycan profile to the experimentally determined glycan profile in the control flask using kinetic parameters obtained by the optimization subroutine. Next, we note that the addition of MnCl_2 on D0 of the cell culture results in a decrease in the relative abundance of fucosylated glycoforms, FA2 and FA2G1 with a corresponding increase in the concentration of their afucosylated isoforms, A2 and A2G1 species. The enzyme alpha-1,6-fucosyltransferase (FucT) is responsible for fucosylating glycan species in the N-glycan biosynthetic pathway. By reducing the total enzyme concentration for FucT in the kinetic model from 0.55 mM to 0.22 mM, we are able to simulate the change in the glycan distribution due to addition of MnCl_2 on D0, as shown in Figure 3.6. The simulations show that changes in the glycan distribution due to the addition of MnCl_2 can be attributed to the changes in the enzyme concentration of FucT.

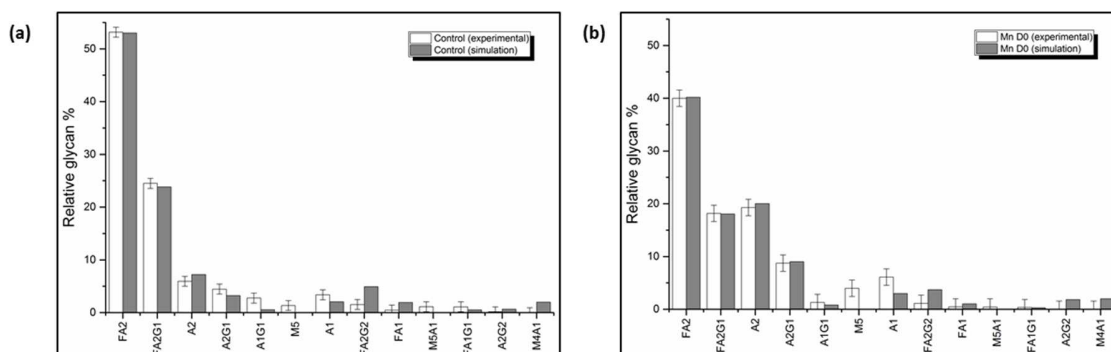


Figure 3.6: Comparison of experimental data and model fit for the glycan distribution profile obtained from (a) control flask; and (b) when MnCl_2 is added on D0.

In addition to fractional difference analysis, we use the glycan indices computed for each experimental condition, shown in Table 3.2, as a quantitative metric for quantifying and assessing the change in the final glycosylation profile caused by the addition of different media supplements. Specifically, a comparison of the individual glycan indices for each condition against the corresponding values under control conditions allows us to establish objectively that altering the availability of MnCl_2 in the media using a chelating agent reverses the changes in the glycan distribution. While MnCl_2 addition on D0 resulted in a decrease in the fucosylation index from 79.9% in the control flask to 60.5%, we see that subsequent introduction of EDTA on either D0, D3, or D6 reversed that trend. Early stage addition of EDTA on D0 increased the fucosylation index to 82.46%, while adding EDTA on D3 and D6 resulted in fucosylation indices of 78.8%, and 76.5% respectively. Similarly, the decrease in the galactosylation index, upon adding MnCl_2 on D0, from 17.9% in the control flask to 15.8%, could be reversed by subsequently adding EDTA on D3. Adding MnCl_2 to the media on D3 reduced the fucosylation index to 61.9%, but if we

then added EDTA on D6, the fucosylation index increased to 77.8%, which is comparable to the value of 79.9% in the control flask. It is important to note however, that the reversal in the glycan indices observed due to the addition of EDTA on D0 and D3 is achieved at the expense of reduced titer and reduced viability, as discussed above. The indication from our results is that changes in the glycan distribution due to MnCl_2 addition can be reversed only when EDTA is added to the media after MnCl_2 addition. Thus, the effect of MnCl_2 supplementation can be reversed, without decreasing productivity, by adding EDTA on D6 to MnCl_2 supplemented media, providing a means of ensuring higher productivity without altering glycan distribution

Table 3.2: Glycosylation index (GI) and fucosylation index (FI) for each experimental condition

Experimental condition [†]	Galactosylation Index (GI)	Fucosylation Index (FI)
Control	17.9	79.9
Mn D0	15.8	60.5
Mn D0/ ED D0	20.6	82.5
Mn D0/ ED D3	18.0	78.8
Mn D0/ ED D6	19.3	76.5
Mn D3	15.7	61.8
Mn D3/ ED D0	25.2	88.6
Mn D3/ ED D3	15.1	73.9
Mn D3/ ED D6	18.0	77.8
Mn D6	16.7	76.3
Mn D6/ ED D0	24.8	90.3
Mn D6/ ED D3	19.9	86.4
Mn D6/ ED D6	17.9	77.3
ED D0	21.6	81.0
ED D3	16.5	75.3
ED D6	16.9	77.7

Although such observations as these provide useful qualitative information about the system, they cannot be used to develop a control strategy; that requires a quantitative representation (and hence understanding) of the effects of media supplementation on glycan distribution. Such quantitative understanding can be

obtained via formal analysis of the experimental data using analysis of variance (ANOVA) to generate the process gain matrix, \mathbf{K} , as described in Materials and Methods. Singular value decomposition of the gain matrix \mathbf{K} produces a (diagonal) matrix of singular values, $\mathbf{\Sigma}$, and two unitary matrices, \mathbf{W} and \mathbf{V}^T . Together these three matrices provide a particularly insightful representation of the process information encapsulated in the gain matrix, \mathbf{K} : Equation 3.3 is transformed into a series of n individual and independent equations where, in each case, a linear combination of the original process input factors, with weighting coefficients from the matrix \mathbf{W} , now constituting an “input mode”, μ_i , is connected through the associated singular value σ_i , to the corresponding linear combination of the output glycans, (with weighting coefficients from the matrix \mathbf{V}), now constituting an output mode η_i (St. Amand et al. 2014c). Furthermore, as a result of this decoupling transformation, the magnitude of each singular value naturally quantifies the extent to which the output mode in question will change in response to a change in the corresponding input mode. Thus, the larger the value of σ_i , the greater will be the change in the corresponding output mode η_i as a result of changes in the input mode μ_i , so that output modes associated with larger values of σ_i will be more “controllable” than modes associated with smaller values of σ_i .

The first ten singular values ($\sigma_1 - \sigma_{10}$) for our experimental system are listed in Table 3.3 in decreasing order of magnitude. As modes associated with singular values of smaller magnitude are less controllable, we limit our analysis only to those modes that are practically controllable; we do this by using a threshold cutoff value, σ^* , arbitrarily selected to be 0.5 in this example, thereby limiting our analysis to the first five singular values. From a process control perspective, modes associated with

singular values below this threshold are considered to be of no practical importance since, for all intents and purposes, they are not controllable

Table 3.3: Singular values obtained from the SVD analysis of the matrix of significant coefficients

σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7	σ_8	σ_9	σ_{10}
6.22	3.68	2.21	0.80	0.61	0.45	0.41	0.31	0.10	0.05

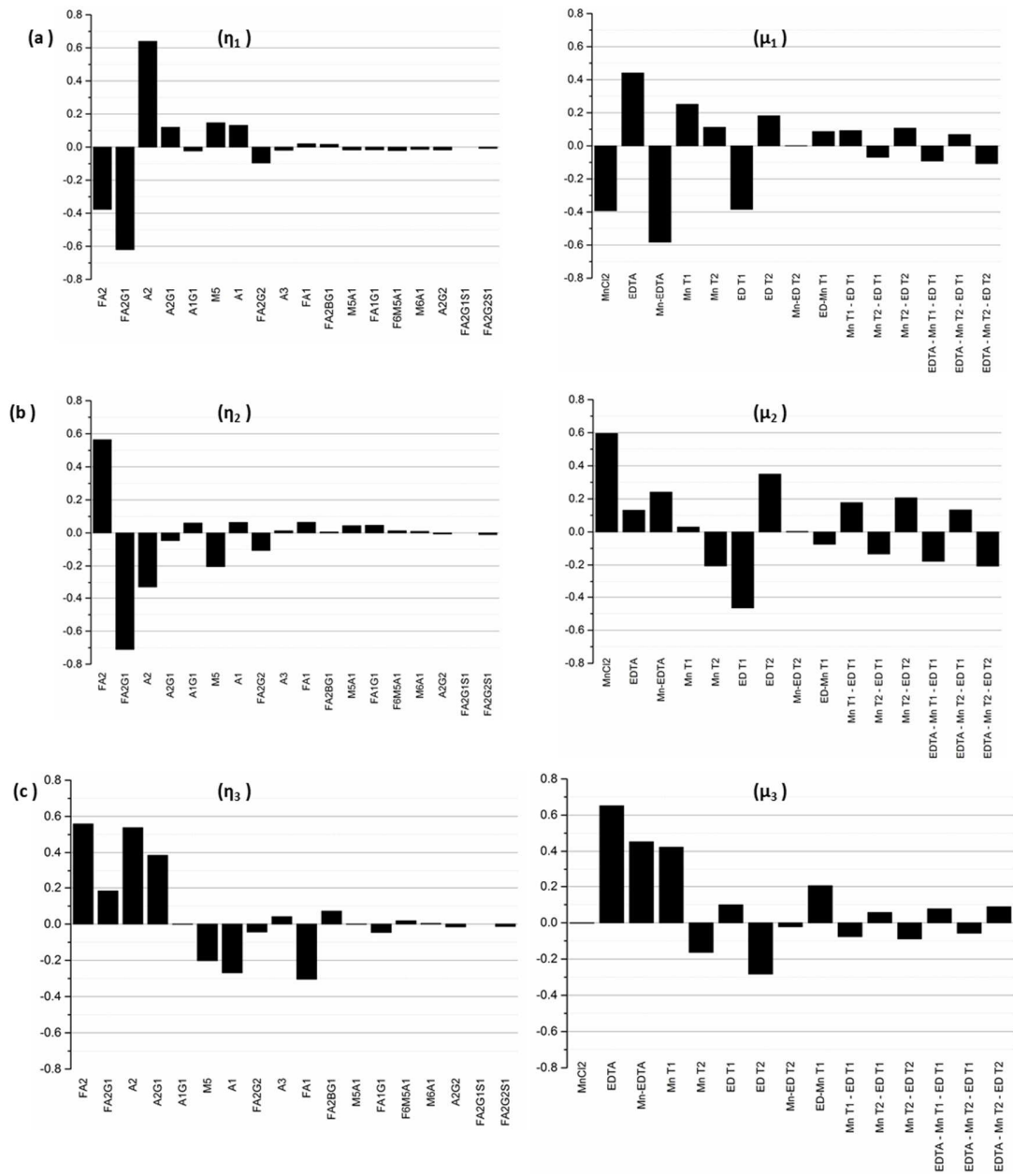
Next, since by definition, each input–output mode pair represents the linear combination of output glycan species that can be controlled by manipulating the specific input factors in the corresponding input mode, the coefficients of each output factor in the output modes and of each input factor in the input modes provide further information about the relative influences exerted by each original input factor on each output factor. Specifically, the coefficient of a particular output factor in a particular mode represents the magnitude by which the relative percentages of those particular glycans will change in response to a unit change in the input mode. On the other hand, the coefficient of a particular input factor in the associated input mode corresponds to the relative contribution of that input factor to the unit change the input mode in question. Thus, the inputs with the largest coefficients in an input mode represent the dominant factors and hence the largest contributors to the influence of that mode, while the output glycans with the largest coefficients in an output mode represent those species whose relative percentages will change the most under the influence of

the input modes. The input-output mode pairs and their associated coefficients are shown in Figure 3.7.

Because it is associated with the largest singular value of $\sigma_1 = 6.62$, the first output mode η_1 is the most controllable output mode. The value of σ_1 represents the change in the overall output mode η_1 resulting from a unit change in the input mode μ_1 . For this output mode, we note that the dominant glycan species are A2, with a coefficient of 0.64, followed by FA2G1 (with a coefficient of -0.62), FA2 (-0.38), M5 (0.14), and A1 (0.13), indicating that a unit positive change in the input mode μ_1 will result in an increase in A2, M5, and A1, accompanied by a decrease in FA2G1 and FA2, each in the amount indicated by the identified coefficients. The biantennary species A2, with the largest coefficient, is the most controllable glycan in the first mode, followed by FA2G1 and FA2. The indicated coupling between the glycans A2, FA2G1, and FA2 makes sense because an increase in the afucosylated glycoforms occurs at the expense of the fucosylated forms, as our experimental results show. The associated input mode μ_1 is a linear combination of different input factors representing the media supplements MnCl_2 and EDTA as well as the times of their addition. Mode μ_1 is primarily dominated by the interaction of MnCl_2 and EDTA, and the concentrations of EDTA and MnCl_2 , with associated coefficients -0.58, 0.44, and -0.39 respectively, indicating that the addition of these two media supplements has opposing individual effects on output mode 1. Early stage addition of EDTA and MnCl_2 , denoted by the factors EDT1 and MnT1, with associated coefficients -0.38 and 0.25 respectively, also exert important influences on the first output mode. Based on the different elements that comprise the first input mode, we note that one can control

the glycans in output mode η_1 by adjusting the concentrations of the two supplements at the early stages of the cell culture.

The next controllable output mode η_2 is associated with the singular value $\sigma_2 = 3.68$, and the linear combination of glycans represented by this mode is dominated by the glycan species FA2G1, FA2, A2, M5, and FA2G2. The coefficients associated with these glycans are -0.71, 0.56, -0.33, -0.20 and -0.11, respectively, indicating that a unit positive change in the input mode μ_2 will cause a relative increase in FA2 while causing the other glycan species to decrease. The increase in FA2 coupled with the decrease in FA2G1 and FA2G2 indicates that perturbations to the input mode μ_2 affect the galactosylated species particularly. Since the singular values are arranged in decreasing order, the influence of mode μ_2 on output mode η_2 is less than that of μ_1 on mode η_1 . The largest coefficients in mode μ_2 are associated with the input factors MnCl_2 (with a coefficient of 0.6), the early stage addition of EDTA denoted by factors EDT1 (with a coefficient of -0.47) and EDT2 (0.35), and the interaction effect of MnCl_2 and EDTA (0.24).



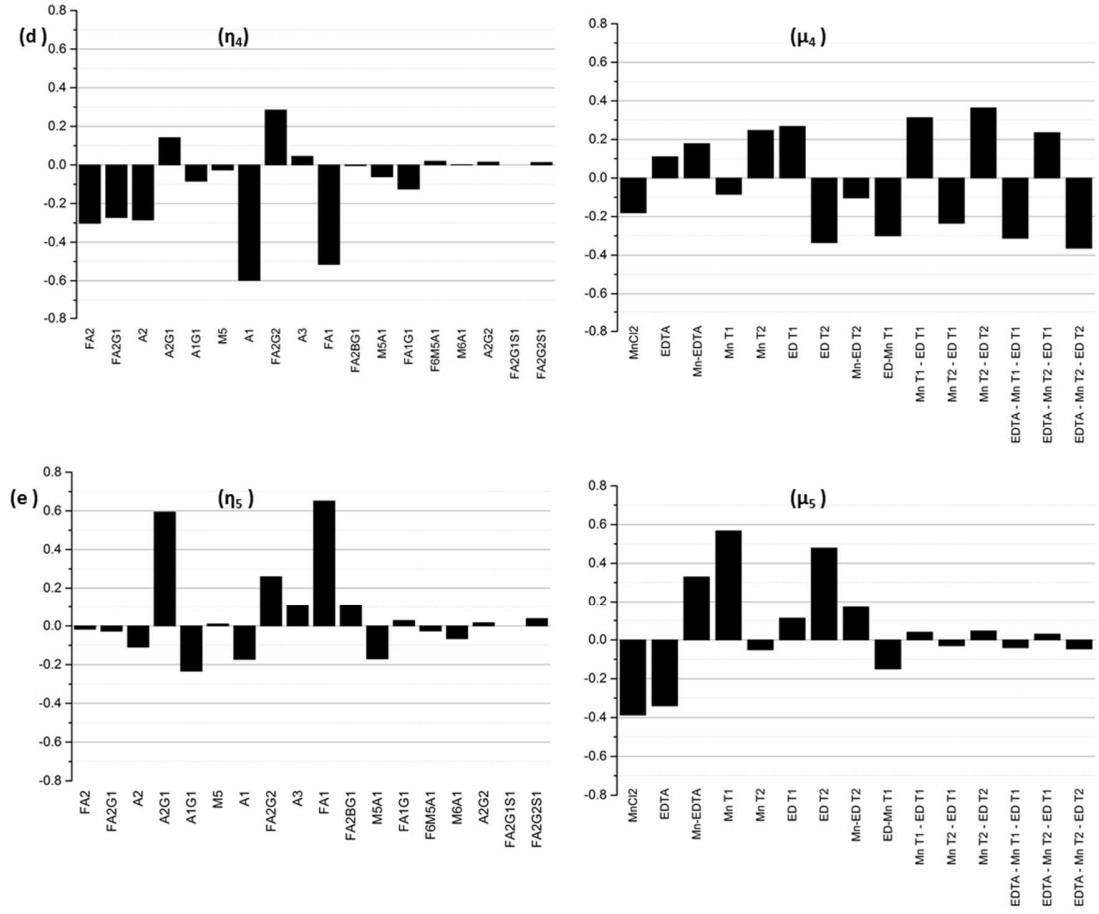


Figure 3.7: Graphical representation of the coefficients associated with the first five input and output modes (with $\sigma_i \geq \sigma^*=0.5$) obtained from controllability analysis. (a) Output mode η_1 and the corresponding input mode μ_1 ; (b) Output mode η_2 and the corresponding input mode μ_2 ; (c) Output mode η_3 and the corresponding input mode μ_3 ; (d) Output mode η_4 and the corresponding input mode μ_4 ; (e) Output mode η_5 and the corresponding input mode μ_5

A unit positive change to the input mode μ_3 causes the following changes in the relative concentrations of the glycan species that comprise output mode η_3 (with $\sigma_3 = 2.21$): A2, A2G1, FA2, and FA2G1 increase, and M5, A1, and FA1 decrease, simultaneously. This indicates that input mode μ_3 can be used to increase the

biantennary species, but at the expense of a (perfectly logical) concomitant decrease in the species that are upstream of these biantennary species. The input mode μ_3 is dominated by EDTA (with associated coefficient 0.65), the interaction effect of MnCl_2 and EDTA (with coefficient 0.45), and the early stage addition of MnCl_2 denoted by the factor MnT1). In each of the three input modes, the coefficients associated with the interaction effect of MnCl_2 and EDTA indicate the importance of this combination input factor in altering the concentrations of the glycan species associated with the respective modes.

The fourth and the fifth modes are less controllable, as they are associated with singular values of comparatively smaller magnitudes ($\sigma_4 = 0.80$ and $\sigma_5 = 0.61$). η_4 is dominated by glycans A1, A2, FA1, FA2, FA2G1, and FA2G2, while η_5 is dominated by FA1, A2G1, and FA2G2. The input mode μ_4 is dominated by the interaction effects Mn T2-ED T2 and $\text{EDTA}^*\text{Mn T2-ED T2}$ representing the interaction between the time of addition of the media supplements. By contrast, the predominant factors in mode μ_5 are MnT1 and EDT2 , which represent the addition of MnCl_2 on D0 and EDTA on D3 respectively.

It is worth mentioning that the controllability analysis presented here provides a quantification of the effect that the addition of specified amounts of particular media supplements and the respective times of addition jointly have on the output glycan distribution at the end of the batch. As discussed above, introducing these media supplements dynamically also results in quantifiable changes in the antibody titer. Thus, the dynamic supplementation strategies discussed here present a challenging problem involving a trade-off between product yield and product quality. To be effective, a control strategy based on these considerations, must therefore be carefully

designed to resolve these conflicts appropriately in order to optimize both the titer and the quality simultaneously.

In closing, we note that:

1. the relationships between time-dependent changes in media supplements and the corresponding changes in glycan distribution are (understandably) complex and not necessarily obvious or easily amenable to qualitative thinking; but
2. controllability analysis via singular value decomposition, and the resulting input-output mode pairs determined specifically for our experimental system, have enabled us to identify which input factors are best manipulated, in order to effect changes in the relative percentage of specific glycan species;
3. in addition, the coefficients in the equations representing the input and output modes allow us to quantify by how much we expect the glycan species to change in response to specific time-dependent media supplementation actions.

3.5 Summary and Conclusions

There is growing interest in evaluating the role of media supplements, especially MnCl_2 , in modulating glycoform distributions in mAbs. However, most media supplementation studies (where supplements are added to the media before starting the batch) do not take into consideration the impact of introducing media supplements at different stages of cell growth. In this chapter we have presented a systematic approach for evaluating the effects of time dependent media supplementation on the glycan profile, and provided a methodology for quantifying and analyzing the complex effects. Our results show that while it is important to consider which supplements are to be added to the media in order to alter specific glycan species, when they are to be added is just as important. In addition to this general observation about what to change and when, our results and analysis technique

also demonstrate how to quantify “by how much” to expect specific glycan species to change as a result of the changes in the media supplements.

For instance, we observe that early stage addition of MnCl_2 affects fucosylated species and alters the glycan distribution more significantly than a late stage addition. Similarly, early stage addition of EDTA affects not just the antibody titer, but also the relative percentages of biantennary and galactosylated species; late stage addition of EDTA does not alter the glycan distribution significantly. The glycan distribution profile is also affected by the addition of both EDTA and MnCl_2 to the media at different time points and a mechanistic understanding of the effect of individual media components on the glycan distribution can be developed by studying the fractional difference in the glycan distribution. In fact, our results demonstrate that changes in the glycan distribution profile due to the addition of MnCl_2 are not immutable; they can be reversed by adding EDTA after MnCl_2 has been added to the media. We then used controllability analysis to identify the specific combinations of input factors which, when manipulated, result in quantifiable changes in the relative percentage of specific glycan species. For the specific experimental system, our analyses show that A2, FA2G1 and FA2 are the most controllable glycan species whose concentrations can be changed by early stage supplementation of EDTA and late stage supplementation of MnCl_2 .

Traditionally, the composition of cell culture media is fixed prior to starting the batch. However, we have demonstrated that introducing supplements at different points in time can influence both the productivity and the quality attribute of the antibody. While we have examined only two specific media components in the current set of experiments, in principle, the systematic approach presented here can be

extended easily to other media components such as amino acids, or trace metals. In addition, the development of hydrogels that can induce a “slow release” or “timed release” of specific nutrients and media supplements should provide additional degrees of freedom that could be investigated in future experiments.

While this chapter has focused on establishing a rational framework for studying the influence of time-dependent media supplementation on the glycosylation profile, the techniques introduced here can be extended to tackle the complementary problem of designing and implementing appropriate control strategies to achieve desired glycan distribution profiles. In the following chapter, the design and implementation of controllers based on output controllability analysis is demonstrated for two different systems.

Chapter 4

A CONTROLLER DESIGN FRAMEWORK FOR CONTROLLING THE GLYCOSYLATION PROFILE

4.1 Introduction

In the preceding chapters, we developed a fundamental understanding of the effect of different input factors on the glycan distribution using two parallel approaches. First, in Chapter 2 we developed a multi-scale model of glycosylation to model the effect of macro-scale bioreactor conditions such as nutrient and metabolite concentrations on the glycan distribution profile. Next, we studied how media conditions affect the glycan distribution and identified, using controllability analysis, the specific combinations of different inputs that can be used to manipulate specific combinations of different glycan outputs. In this work, we use output controllability analysis to design proportional (P) and proportional-integral (PI) controllers to control the glycan distribution in mAbs.

We propose a two-step framework for controlling glycosylation based on output controllability analysis. First, we use statistical design of experiments to study the changes in the glycosylation pattern due to different input factors and then implement controllability analysis to identify specific inputs that can be used to control the glycosylation profile. As demonstrated in Chapter 3, output controllability analysis is used to identify the combination of input factors that alter the concentrations of specific combinations of glycan species, and to develop a quantitative input-output relationship. Using the structural information generated from

such an analysis, we design a control scheme for glycosylation control and illustrate its applicability using two specific case studies: (a) where the concentrations of specific glycosyltransferase enzymes are the inputs (or manipulated variables) used to control glycan distribution (output) and we represent the input-output relationship using the micro-scale model developed from a first principles based analysis of the glycan reaction network; and (b) where we add select amino acids to the cell culture media and use these cell culture media supplements as manipulated variables. The first case study is based on the micro-scale mathematical model originally presented in §2.3.3. We investigate the second case study as there is growing interest in examining the role of media additives and optimizing cell culture media to improve antibody titer and modulate product quality attributes (Blondeel et al. 2016; Jordan et al. 2013; Kildegaard et al. 2016; Kishishita et al. 2015; Kyriakopoulos and Kontoravdi 2014; Landauer 2014; Rouiller et al. 2014; Zhang et al. 2013). Cell culture media development is an area of active research, and specifically, the role of amino acids in altering cellular metabolism is being investigated to develop means to enhance cell culture productivity and titer (Gonzalez-Leal et al. 2011; Xing et al. 2011). However, the effect of adding different amino acids on the glycosylation profile and hence the overall quality is not well understood and there are often conflicting reports on the subject. For instance, the addition of asparagine to cell culture media has been reported to modulate the concentrations of galactosylated species of IgG1 (McCracken et al. 2014; Rives et al. 2015), while other studies have reported no effect of amino acids on the quality attributes of IgG3 (Read et al. 2013). In our work, we systematically investigate the effect of amino acid supplements using a factorial design approach and perform controllability analysis.

Based on output controllability analysis (that has been described in Chapter 3), we design P- and PI-type controllers, and then test via simulations the controller performance for set-point tracking. We then test, via simulations, the controller performance under nominal conditions by assuming that the process model accurately represents the process. However, due to the inherent complexity and non-linearity of the system, the linear process model generated by our analysis is bound to be different from the ‘true’ process and hence, we test the controller performance under model-plant mismatch conditions. Our results show that in each instance the designed controllers are able to track the set-point effectively, indicating that the two-step framework presented here is an effective method for controlling the glycosylation profile.

In §4.2, different experimental and computational approaches used to develop the glycosylation controllers are discussed, while in §4.3 our main findings are presented and the results are discussed. We summarize the work presented in this chapter in the final section of this chapter.

4.2 Materials and Methods

4.2.1 Computational Model of Glycosylation

The micro-scale glycosylation model introduced in Chapter 2 was suitably modified to study the changes in the glycan distribution induced by changes in the concentrations of the glycosyltransferase enzymes. Having defined the reaction network and obtained the mass balances for the 18 glycan species participating in 20 reactions based on the cisternal maturation model, we calculate the distribution of the different glycosylation enzymes along the length of the Golgi using the peak enzyme

concentrations (E_{\max}) listed here: $E_{\max, \text{FucT}} = 2.33 \mu\text{M}$; $E_{\max, \text{GalT}} = 1.00 \mu\text{M}$; $E_{\max, \text{GnTI}} = 2.52 \mu\text{M}$; $E_{\max, \text{GnTII}} = 1.07 \mu\text{M}$; $E_{\max, \text{ManI}} = 2.45 \mu\text{M}$; $E_{\max, \text{ManII}} = 1.00 \mu\text{M}$. The mass balance equations form a system of partial differential equations that was solved using the ode15s function in MATLAB to obtain the dynamic concentration profile for each glycan species. Next, using Plackett-Burman experimental design, the peak enzyme concentrations, E_{\max} , for all six glycosyltransferase enzymes were changed by $\pm 10\%$ and the new glycan distribution was obtained. The variance in the glycan distribution data was then analyzed in MINITAB using ANOVA.

4.2.2 Cell Culture

Frozen cells from a CHO-K1 cell line producing a model IgG1 (gift of Genentech, San Francisco, CA) were thawed and inoculated with a seeding density of 0.5×10^6 cells/mL in vented-cap Erlenmeyer shake flasks with a working volume of 50 mL, maintained in an incubator at 37°C with a 5% CO_2 overlay. Custom CD-OTPICHO™ media (Thermo Fisher Scientific, Waltham, MA) supplemented with 5 g/L glucose, 4 mM glutamine and 0.01 mM MnCl_2 was used as the cell culture media in this study.

Based on a survey of existing literature, eleven amino acids were chosen as additional cell culture supplements: valine, arginine, glutamic acid, proline, asparagine, serine, methionine, threonine, leucine, and lysine (Sigma Aldrich, St. Louis, MO). Stock solutions for each amino acid were prepared in deionized water and added to individual flasks as per the conditions specified in the experimental design section. In addition to these flasks, a control flask containing the basal media with no amino acid supplementation was also cultured. All 13 experimental conditions (12 corresponding to the Plackett Burman design + 1 control) were performed in

duplicate. pH, nutrient, and metabolite concentration in each flask was analyzed using a Bioprofile 100+ Analyzer (Nova Biomedical, Waltham, MA) and the viable cell count was measured manually using a hemocytometer. Antibody titer was estimated using a Thermo Scientific™ MAbPac Protein A chromatography column (12 micron particle size, 35x4.0 mm I.D., Thermo Fisher Scientific, Waltham, MA) on an Agilent 1200 HPLC instrument.

4.2.3 Glycan Permethylation Assay

Media containing the secreted antibody was harvested nine days after inoculation by spinning down the cells at 3000 rpm. The antibody was then purified using 2 mL PhyTip columns packed with Protein A chromatography resin and operated using PhyNexus Benchtop MEA2 system (PhyNexus, San Jose, CA). Next, after the tryptic digestion (Promega, Madison, WI) of the purified mAb at 37 °C for four hours, a 16 hour deglycosylation protocol was initiated using PNGase-F (ProZyme, Hayward, CA) to release the glycan groups attached to the mAb. The released glycans were separated from the residual enzyme and other peptides using Hypersep Hyper Carb SPE cartridges (Thermo Fisher Scientific, Waltham, MA) and permethylated following the Ciucanu method with methyl iodide and NaOH in the presence of DMSO (Ciucanu and Costello 2003; Ciucanu and Kerek 1984). The permethylated glycans were recovered by a liquid-liquid extraction step using chloroform (Sigma Aldrich, St. Lois, MO), dried, and resuspended in 80% methanol followed by purification on Sep Pak SPE C18 cartridges (3 cc Vac Cartridge, 200 mg Sorbent per Cartridge, 55-105 µm Particle Size, Waters Corporation, Milford, MA). Finally, the glycans were dried and resuspended in LC-MS grade methanol (Sigma Aldrich, St. Lois, MO) and spotted onto a MALDI-TOF plate with a DHB matrix to be

analyzed using a 4800 MALDI TOF/TOF Analyzer (ABSciex, Framingham, MA) in positive ion, reflector mode. The data collected from the mass spectrometer was then exported to Data Explorer to obtain the heights of the individual peaks in the sample. The relative glycan distribution in each sample was estimated from the percentage peak heights of each sample.

4.2.4 Statistical Design of Experiments

Plackett-Burman experimental design was used to assess the effect of changing the different inputs on the glycan distribution. Table 4.1 lists the experimental design used as well as the high and low concentration levels used for each of the glycosyltransferases to simulate the dynamic glycan profile. Table 4.2 lists all the experimental conditions tested with amino acid supplementation and specifies the high and low amino acid levels that were introduced to each flask. The variance in the data obtained from each experimental design was analyzed using ANOVA in MINITAB.

Table 4.1: Plackett-Burman experimental design with high and low concentrations of different glycosyltransferase enzymes in the dynamic mathematical model for glycosylation

Run	FucT, (mM)	GalT, (mM)	GnTI, (mM)	GnTII, (mM)	ManI, (mM)	ManII, (mM)
1	2.56	0.90	2.78	0.96	2.21	0.90
2	2.56	1.10	2.27	1.18	2.21	0.90
3	2.09	1.10	2.78	0.96	2.70	0.90
4	2.56	0.90	2.78	1.18	2.21	1.10
5	2.56	1.10	2.27	1.18	2.70	0.90
6	2.56	1.10	2.78	0.96	2.70	1.10
7	2.09	1.10	2.78	1.18	2.21	1.10
8	2.09	0.90	2.78	1.18	2.70	0.90
9	2.09	0.90	2.27	1.18	2.70	1.10
10	2.56	0.90	2.27	0.96	2.70	1.10
11	2.09	1.10	2.27	0.96	2.21	1.10
12	2.09	0.90	2.27	0.96	2.21	0.90

Table 4.2: Plackett-Burman experimental design with high and low concentrations of individual amino acids added to the cell culture media

Run	Val, (mM)	Arg, (mM)	Glu, (mM)	Pro, (mM)	Gly, (mM)	Asn, (mM)	Ser, (mM)	Met, (mM)	Thr, (mM)	Leu, (mM)	Lys, (mM)
1	0.27	1.15	0.54	2.50	0.50	2.50	0.80	0.10	0.70	0.22	0.50
2	0.27	0.72	0.54	2.50	2.50	0.50	0.80	0.40	0.24	0.70	0.50
3	1.61	1.15	0.54	0.50	2.50	2.50	0.40	0.40	0.24	0.22	0.50
4	1.61	0.72	0.54	0.50	0.50	0.50	0.80	0.40	0.70	0.22	0.87
5	1.61	0.72	0.54	2.50	0.50	2.50	0.40	0.10	0.24	0.70	0.87
6	1.61	0.72	0.17	0.50	2.50	2.50	0.80	0.10	0.70	0.70	0.50
7	0.27	0.72	0.17	2.50	2.50	2.50	0.40	0.40	0.70	0.22	0.87
8	0.27	1.15	0.17	0.50	0.50	2.50	0.80	0.40	0.24	0.70	0.87
9	0.27	1.15	0.54	0.50	2.50	0.50	0.40	0.10	0.70	0.70	0.87
10	0.27	0.72	0.17	0.50	0.50	0.50	0.40	0.10	0.24	0.22	0.50
11	1.61	1.15	0.17	2.50	2.50	0.50	0.80	0.10	0.24	0.22	0.87
12	1.61	1.15	0.17	2.50	0.50	0.50	0.40	0.40	0.70	0.70	0.50

4.2.5 Controllability Analysis

As described in Chapter 3, we performed controllability analysis to quantify the effect of changing the input conditions on the glycosylation profile in mAbs. Briefly, we used Plackett-Burman factorial design to obtain different experimental conditions to be tested in the two cases. Accordingly, we performed multiple simulations of the dynamic model using the different glycosyltransferase enzyme concentrations as our inputs, and carried out shake flask experiments with varying concentrations of amino acids added to each flask. The variance in the final glycan distribution obtained from our simulations and from our experiments was analyzed using Minitab. The estimates of factor coefficients obtained from such an analysis represent the process gains or the change in the output glycan distribution ($\Delta \mathbf{y}$), when there is a unit change in the input factor ($\Delta \mathbf{u}$) associated with that coefficient. We selected statistically significant factor coefficients and set all other insignificant coefficients to zero to generate the process gain matrix \mathbf{G} , which gives us

$$\Delta \mathbf{y} = \mathbf{G} \Delta \mathbf{u} \quad 4.1$$

Next, by performing singular value decomposition of the process gain matrix, we obtained the diagonal singular value matrix, $\mathbf{\Sigma}$, and the unitary matrices, \mathbf{W} and \mathbf{V}^T , that were subsequently used to obtain the orthogonal input ($\boldsymbol{\mu}$) and output ($\boldsymbol{\eta}$) modes. Thus,

$$\Delta \mathbf{y} = \mathbf{W} \mathbf{\Sigma} \mathbf{V}^T \Delta \mathbf{u} \quad 4.2$$

$$\mathbf{W}^T \Delta \mathbf{y} = \mathbf{\Sigma} \mathbf{V}^T \Delta \mathbf{u} \quad 4.3$$

$$\Delta\boldsymbol{\eta} = \boldsymbol{\Sigma}\Delta\boldsymbol{\mu} \quad 4.4$$

As $\boldsymbol{\Sigma}$ is a diagonal matrix, we related the change in each output mode ($\Delta\eta_i$) to its corresponding input mode ($\Delta\mu_i$) using the associated singular value (σ_i) as:

$$\Delta\eta_i = \sigma_i\Delta\mu_i \quad 4.5$$

4.2.6 Controller Design under Nominal Conditions

Proportional controllers were designed based on the results obtained from controllability analysis. Starting with the steady state glycan distribution (\mathbf{y}_{ss}), controllers were designed to track the new set-point, or new glycan distribution, (\mathbf{y}_d). These variables were then transformed to output modes using the orthogonal matrix (\mathbf{W}) obtained by singular value decomposition of the gain matrix (\mathbf{G}). Thus,

$$\boldsymbol{\eta} = \mathbf{W}^T(\mathbf{y} - \mathbf{y}_{ss}) \quad 4.6$$

$$\boldsymbol{\eta}_d = \mathbf{W}^T(\mathbf{y}_d - \mathbf{y}_{ss}) \quad 4.7$$

The error ($\boldsymbol{\varepsilon}$) was calculated based on the difference between the output modes for the set point and the measurement as

$$\boldsymbol{\varepsilon} = \boldsymbol{\eta}_d - \boldsymbol{\eta} = \Delta\boldsymbol{\eta} \quad 4.8$$

The controller is then designed to obtain the particular set of inputs ‘ \mathbf{u} ’ which, when implemented on the process, will result in an output signal similar to the new set-point, thereby minimizing this error. To design the controller appropriately, we note that equation (5) can be rewritten as

$$\Delta\mu_i = \frac{\Delta\eta_i}{\sigma_i} \quad 4.9$$

Thus, we designed simple P-controllers using the individual singular values, using the control law given by

$$\Delta\boldsymbol{\mu} = \mathbf{K}_c \boldsymbol{\varepsilon} \quad 4.10$$

where the proportional gain of the controller, \mathbf{K}_c , is given by

$$\mathbf{K}_c = \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_n \end{pmatrix} \quad 4.11$$

Finally, we estimated the change in the input that, when implemented on the process, will track the set point effectively

$$\Delta\mathbf{u} = \mathbf{V}\Delta\boldsymbol{\mu} \quad 4.12$$

The final output glycan distribution, \mathbf{y}_{new} , corresponding to the new set of input conditions obtained from the controller was then calculated using the gain matrix from equation 4.1

$$\mathbf{y}_{\text{new}} = \mathbf{y}_{\text{ss}} + \mathbf{G}\Delta\mathbf{u} \quad 4.13$$

4.2.7 Controller Design under Model-Plant Mismatch Conditions

The glycosylation controllers were designed under nominal conditions, based on the assumption that the process model accurately defines the “true” physical process or the plant. However, in practical instances we do not have complete knowledge of the “true” model of the process resulting in model-plant mismatch. To simulate model-plant mismatch conditions normally distributed random error (Φ) was added to every element of the nominal gain matrix, (\mathbf{G}), to yield a modified gain matrix, (\mathbf{G}_{mod}). Thus,

$$\mathbf{G}_{\text{mod}} = \mathbf{G} + \Phi \quad 4.14$$

Singular value decomposition of the modified gain matrix yielded a new set of singular values, (Σ_{mod}) and new orthonormal matrices \mathbf{W}_{mod} and $\mathbf{V}_{\text{mod}}^T$. Proportional (P) and Proportional Integral (PI) glycosylation controllers were then designed using the new set of singular values; thus:

$$\mathbf{K}_{c,MPM} = \begin{pmatrix} 1/\sigma_{1,mod} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_{n,mod} \end{pmatrix} \quad 4.15$$

$$\mathbf{I}_{MPM} = \mathbf{K}_{c,MPM}/3 \quad 4.16$$

where $\mathbf{K}_{c,MPM}$ is the proportional gain of the P- and PI-controller and \mathbf{I}_{MPM} is the associated integral constant used in the PI-controller. The ability of the glycosylation controllers to track the output trajectory under model plant mismatch conditions was assessed iteratively, with every successive iteration representing a new batch. The control laws used for the P- and PI- controllers were

$$\boldsymbol{\mu}(k+1) = \boldsymbol{\mu}(k) + \mathbf{K}_{p,MPM} * \boldsymbol{\varepsilon}(i) \quad 4.17$$

$$\boldsymbol{\mu}(k+1) = \boldsymbol{\mu}(k) + \mathbf{K}_{p,MPM} * \boldsymbol{\varepsilon}(k) + \mathbf{I}_{MPM} * \sum_{m=1}^k \boldsymbol{\varepsilon}_m \quad 4.18$$

where $\boldsymbol{\mu}$ is the vector of input modes and ‘ k ’ denotes the current iteration. The corresponding change in the input, $\Delta \mathbf{u}$, and the final glycan distribution \mathbf{y}_{new} are determined after every iteration using equations 4.12 and 4.13 above. Figure 4.1 shows the control block diagram representing the glycosylation control scheme.

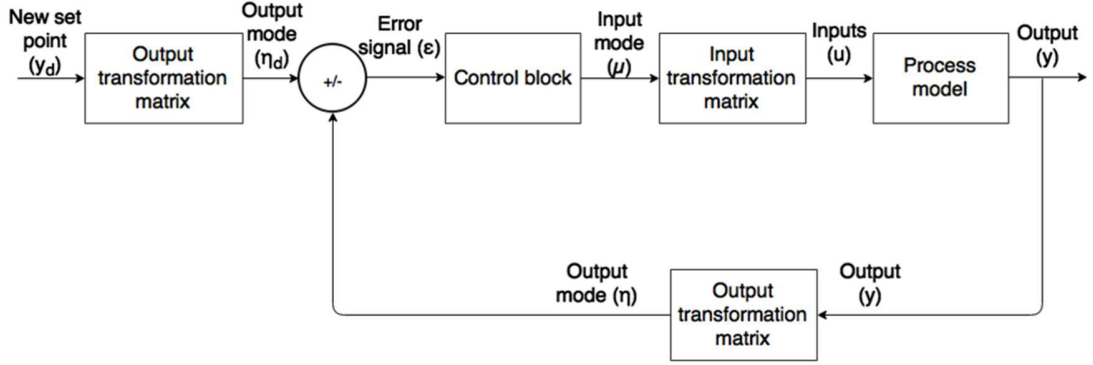


Figure 4.1: Control block diagram indicating different components of the glycosylation control scheme. The current glycan distribution measured at the output (\mathbf{y}) and the new glycan distribution set point (\mathbf{y}_d) are transformed to output modes ($\boldsymbol{\eta}$ and $\boldsymbol{\eta}_d$, respectively) using the output transformation matrix, (\mathbf{W}^T), obtained by the singular value decomposition of the process gain matrix, (\mathbf{G}). The error signal, ($\boldsymbol{\varepsilon}$), obtained from the difference of the two output modes is sent to the control block, where controllers designed using appropriate singular values ($\sigma_i \geq \sigma^*$) generate the input mode ($\boldsymbol{\mu}$) which is converted back to the set of inputs (\mathbf{u}) using the appropriate input transformation matrix, \mathbf{V}

4.3 Results and Discussions

4.3.1 Controllability Analysis

Using statistical design of experiments (DoE), we have systematically studied the effect of manipulated variables (i.e. enzyme concentrations in model simulations, and amino acid concentrations in cell culture media supplementation trials) on the glycan distribution profile in mAbs. However, the qualitative data obtained from our preliminary analysis is not sufficient to develop an effective control scheme to control the glycan distribution using the manipulated variables; instead, we need to establish a quantitative input output relationship between the amino acid or enzyme concentrations (manipulated variables) and the output glycan distribution. Such a quantitative understanding of the effect of adding a specific amino acid or altering

certain glycosyltransferase enzyme concentrations on the glycan distribution can be readily obtained by analyzing the variance in the glycan distribution data. In each case, the statistically significant input coefficients represent the magnitude (gain) associated with the output glycan species for unit changes in that specific input. Thus, using the coefficients obtained from ANOVA we construct a process gain matrix **G**, as previously described under materials and methods. The process gain matrices can be graphically represented using the heat maps shown in Figures 4.2(a-b), which indicate the magnitude as well as direction (increase or decrease) of change effected by a unit change in the input on the output glycan distribution profile. Figure 4.2(a) is a heat map representing the estimates of the statistically significant factor coefficients ($\alpha = 0.05$), with all other coefficients set to zero. We observe that the concentrations of the high mannose glycan species, M8, M7, M5, biantennary species A2 and its galactosylated isoforms A2G1, A2G2 are not affected by the change in the concentrations of any of the glycosyltransferase enzymes. Similarly, a change in the concentration of the input factor, α -1,6-fucosyltransferase (FucT) has no impact on any of the glycan species. The heat map also indicates that in the simulations tested, the biantennary fucosylated species FA2 and its galactosylated isoforms FA2G1 and FA2G2 are the most affected species.

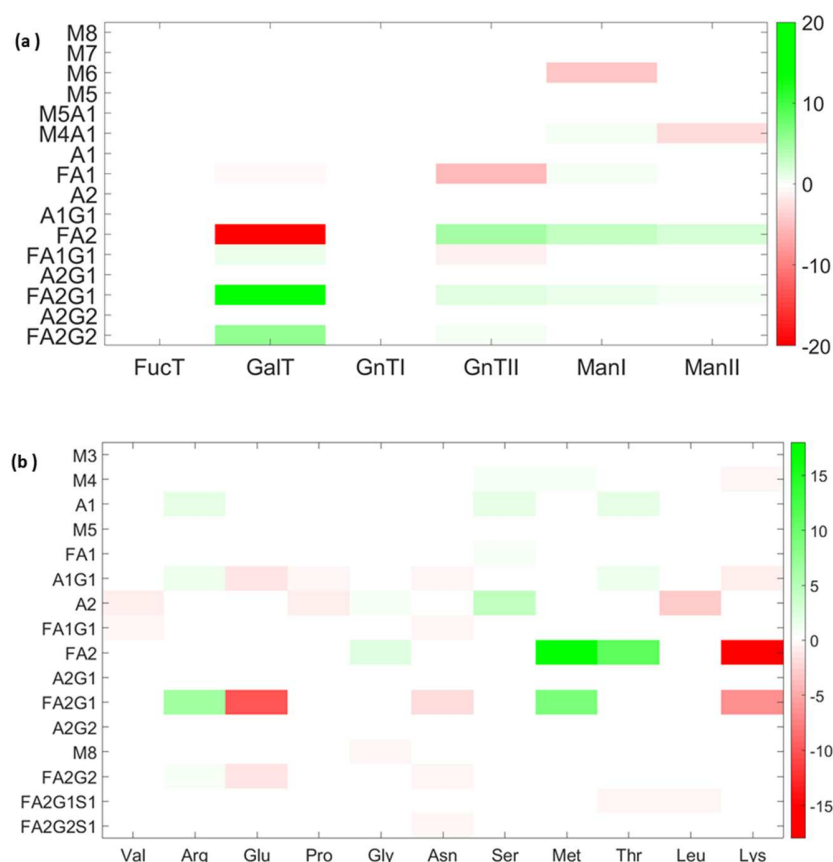


Figure 4.2: Heat map of significant factor coefficients obtained by analysis of variance of the glycan distribution data. The experimental data was obtained using (a) glycosyltransferase enzyme concentrations as inputs in the dynamic mathematical model of glycosylation (b) amino acid media supplements as inputs. The input factors are listed along the horizontal axis while the output glycan species obtained from the simulations are listed along the vertical axis. The color bar indicates the signage and magnitude of the factor coefficient, with red indicating a negative coefficient (i.e. a unit increase in an input decreases the relative concentration of that glycan) while green indicates a positive coefficient (i.e. a unit increase in an input increases the relative glycan concentration). The intensity of each color represents the magnitude with lighter hues indicating factor coefficients of lesser magnitude than those represented by darker hues

Figure 4.2(b) is a heat map representing the estimated gain coefficients obtained from analysis of variance of the glycan distribution data when different amino acids are added to the media. Here, we use a more “relaxed” criterion to estimate the statistically significant factor coefficients ($\alpha = 0.25$) as opposed to the conventionally used value of $\alpha = 0.05$. As α is the probability that we reject the null hypothesis when it is true, in this instance, we have chosen a higher risk of rejecting the null hypothesis (that the estimated factor coefficients are not significantly different from zero) in favor of the alternative. Our decision is informed by an understanding of the inherent variability in the experimental measurements involved in analyzing glycan distribution and the addition of 11 different amino acids to each individual flask. From Figure 4.2(b) we note that the concentrations of certain glycan species, such as M8, M5, A2G1 cannot be influenced by supplementing the cell culture media with any of the amino acids used in the study, whereas the concentrations of fucosylated glycoforms FA2 and FA2G1 can be altered to varying degrees using arginine, glutamic acid, methionine, threonine, and lysine. Each amino acid input in turn, influences the concentration of more than one glycan species, highlighting the close interdependence in the concentrations of different glycan species.

Further, by carrying out singular value decomposition of the process gain matrix \mathbf{G} , we obtain a diagonal matrix $\mathbf{\Sigma}$ containing the singular values, and the two unitary matrices \mathbf{W} and \mathbf{V}^T . As has been previously described under materials and methods, we can transform the original input output equation given in equation 4.1 to a series of n independent equations listed in equation 4.4, where $\mathbf{\eta}$, the output mode, represents a linear combination of the output glycan species weighted by the individual elements of \mathbf{W} ; the associated singular value σ (obtained from the diagonal

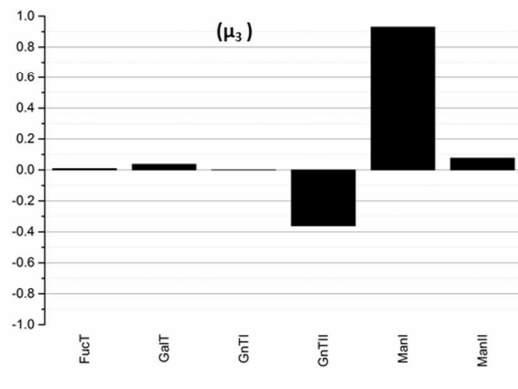
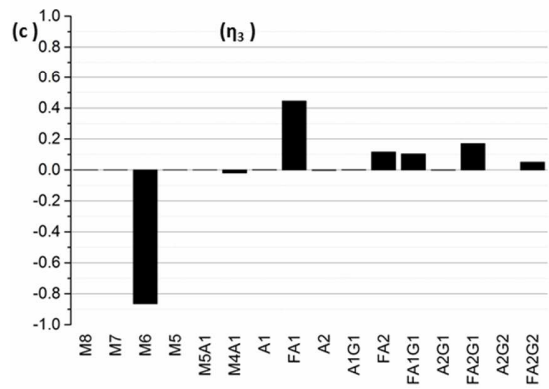
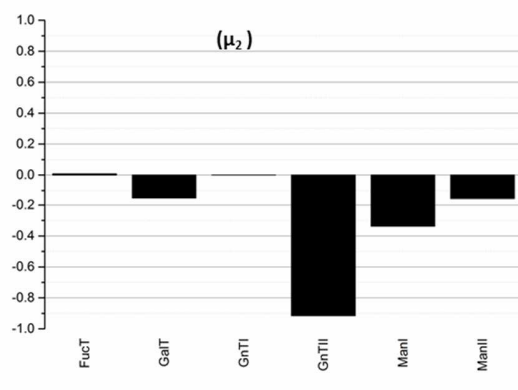
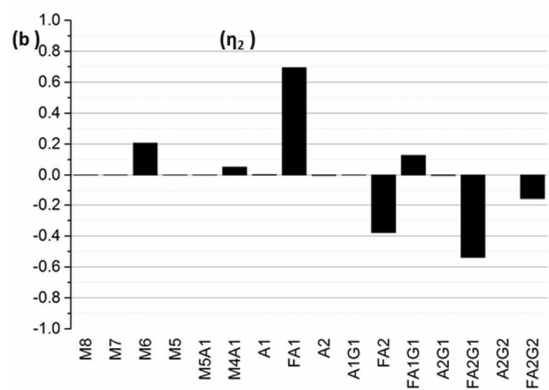
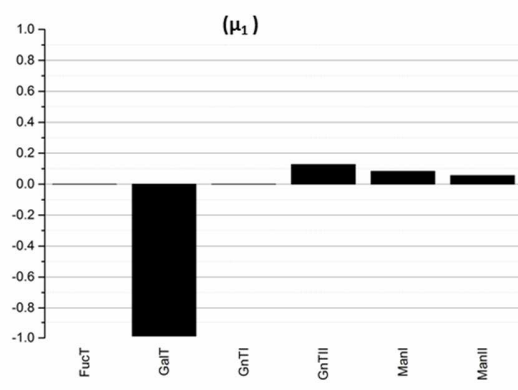
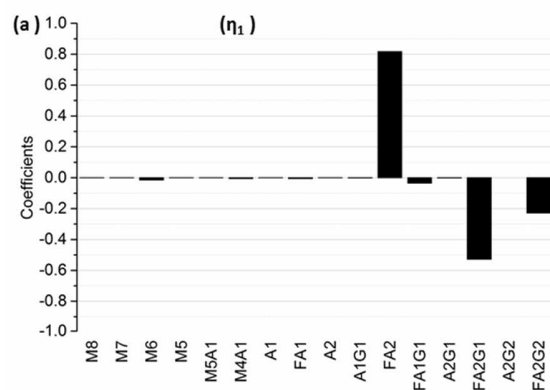
elements of the matrix Σ), quantifies the change in the output mode due to a unit change in the input mode μ , which is the linear combination of the inputs weighted by the unitary matrix V^T . As the magnitude of the singular value quantifies the extent of change in the particular output mode, we note that smaller values of σ are associated with output modes that are less controllable, and hence we choose to study only those modes that are above an arbitrarily chosen threshold cutoff value of σ^* and ignore all modes associated with lower singular values. After examining the singular values listed in Table 4.3 we choose a threshold cutoff value of $\sigma^* = 1$ and ignore all modes associated with lower singular values. Further, since each input-output mode pair represents the linear combination of glycans that can be controlled using different inputs, the associated coefficient of each input and output in any given mode is indicative of the relative influence of a particular input on the output. The coefficient of the output glycan is the magnitude by which its relative percentage changes in response to a unit change in the corresponding input mode, implying that the relative percentage of glycans with larger coefficients in a particular mode can be modulated more easily, making those glycans more controllable. Similarly, inputs with larger coefficients have a relatively larger contribution to a unit change in the associated mode and are hence considered as the dominant factors in that mode.

Table 4.3: Singular values obtained from singular value decomposition of the matrix of significant factor coefficients with (a) glycosyltransferase enzyme coefficients as inputs; and (b) amino acid supplements as inputs

Singular Value	(a)	(b)
σ_1	25.2	28.1
σ_2	7.1	11.8
σ_3	4.7	5.9
σ_4	2.7	2.9
σ_5	0.1	1.2
σ_6	0.0	0.6
σ_7	-	0.4
σ_8	-	0.3
σ_9	-	0.1
σ_{10}	-	0.1
σ_{11}	-	0.0

In Figure 4.3 we have plotted the input-output mode pairs relating the changes in the glycosyltransferase enzyme concentration to the changes in the output glycan distribution. Based on the selected threshold cutoff value of $\sigma^* = 1$, we select only the first four input-output mode pairs associated with the four largest singular values, $\sigma_1 - \sigma_4$. From the first output mode, η_1 (which is associated with the largest singular value, $\sigma_1 = 25.22$) we notice that the most controllable glycans are fucosylated biantennary glycan FA2 and its galactosylated glycoforms FA2G1 and FA2G2 as they have the largest coefficients (0.82, -0.53, and -0.23 respectively) indicating that a unit change in

the first input mode would result in an increase in the FA2 concentration and a decrease in the relative concentrations of its galactosylated isoforms. The corresponding input mode, μ_1 is dominated by GalT (-0.99), indicating that a decrease in the GalT enzyme levels would result in a decrease in the concentration of FA2G1 and FA2G2, as expected. The second output mode (with a singular value of $\sigma_2 = 7.12$) shows that a unit change in the input mode results in an increase in FA1 (with a coefficient of 0.69) and M6 (0.21) and a decrease in FA2G1 (-0.58) and FA2 (-0.38). The linear combination of enzymes associated with the second input mode are GnTII (with a coefficient of -0.92), ManI (-0.34), and GalT (-0.15). As expected, a decrease in GnTII would result in a decrease in the relative concentration of biantennary species FA2, while the reduction in GalT concentration causes the relative concentration of FA2G1 to decrease as well. Finally, we can study the third and fourth input-output mode pairs ($\sigma_3 = 4.69$ and $\sigma_4 = 2.74$, respectively) to assess which other glycan species can be modulated by modifying different combinations of glycosyltransferase enzymes.



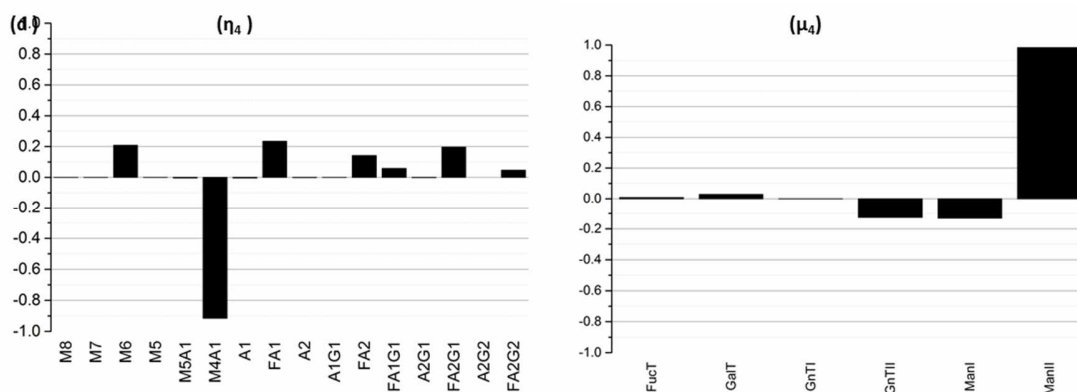
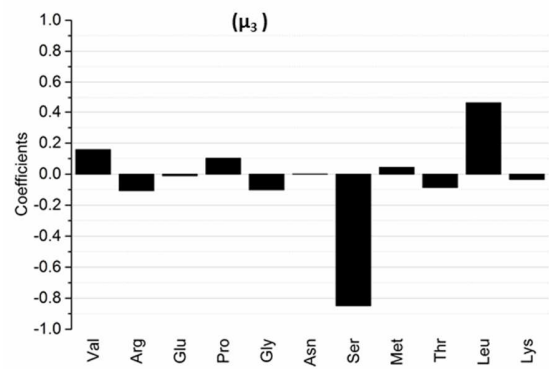
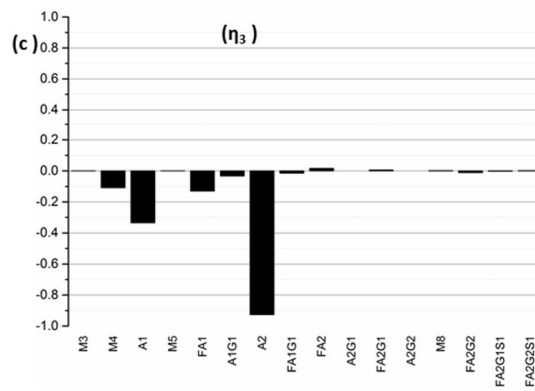
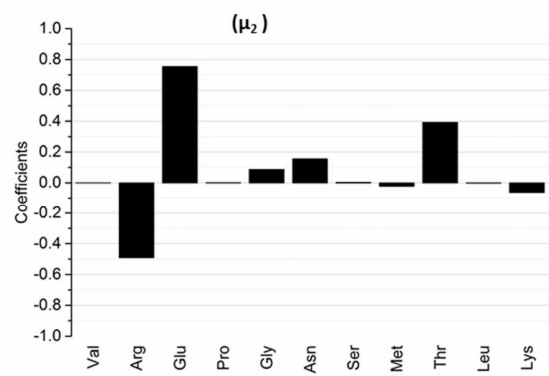
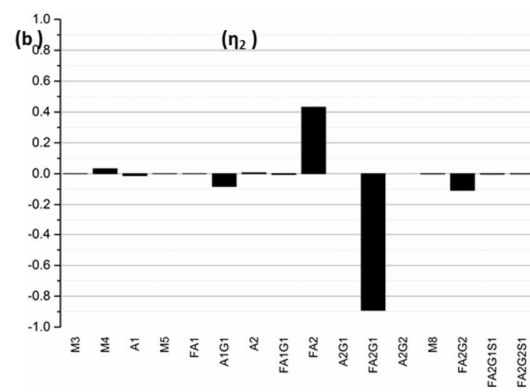
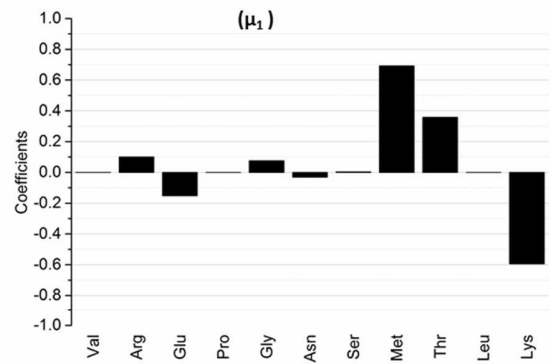
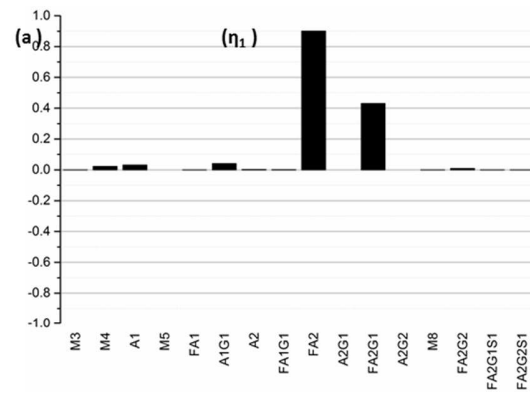


Figure 4.3: Graphical representation of the coefficients associated with the first four input and output modes with glycosyltransferase enzymes as the input factors. Shown here are the primary modes (with $\sigma_i \geq \sigma^*=0.5$) obtained from controllability analysis. (a) Output mode η_1 and the corresponding input mode μ_1 ; (b) Output mode η_2 and the corresponding input mode μ_2 ; (c) Output mode η_3 and the corresponding input mode μ_3 ; (d) Output mode η_4 and the corresponding input mode μ_4 .

A similar analysis to assess the effect of amino acid supplementation on the glycan distribution was carried out by studying the first five input and output modes (associated with singular values $\sigma_1 - \sigma_5 > \sigma^*$) plotted in Figure 4.4. The first output mode is associated with a singular value of $\sigma_1 = 28.1$ and is dominated by the coefficients of the glycan species FA2 (0.92) and FA2G1 (0.43), indicating that these two glycan species are the most controllable, with a unit change in the associated input mode resulting in an increase in their relative concentrations. The first input mode, μ_1 , representing the linear combination of amino acids that when added to the media will result in changes in the concentration of the glycans contained in the first output mode, is dominated by methionine (0.69), lysine (-0.59) and threonine (0.36). A unit change in the second input mode ($\sigma_2 = 11.8$) results in a decrease in FA2G1 (-0.89) and

FA2G2 (-0.11) and an increase in the agalactosylated isoform, FA2 (0.43). The linear combination of amino acids that make up the second input mode include glutamic acid (0.75), arginine (-0.49), and threonine (0.39). The third input-output mode pair with $\sigma_3 = 5.9$ indicates that the relative percentage of A2, A1, and FA1, with coefficients of -0.93, -0.33, and -0.13 respectively, are altered when a linear combination of all amino acids, dominated by serine (-0.85) and leucine (0.46) is changed. We can further analyze the fourth and fifth input-output mode pairs in a similar fashion. We note that no single amino acid dominates in these modes; instead, the change in the output glycan profile is due to the combined action of all the added amino acids.



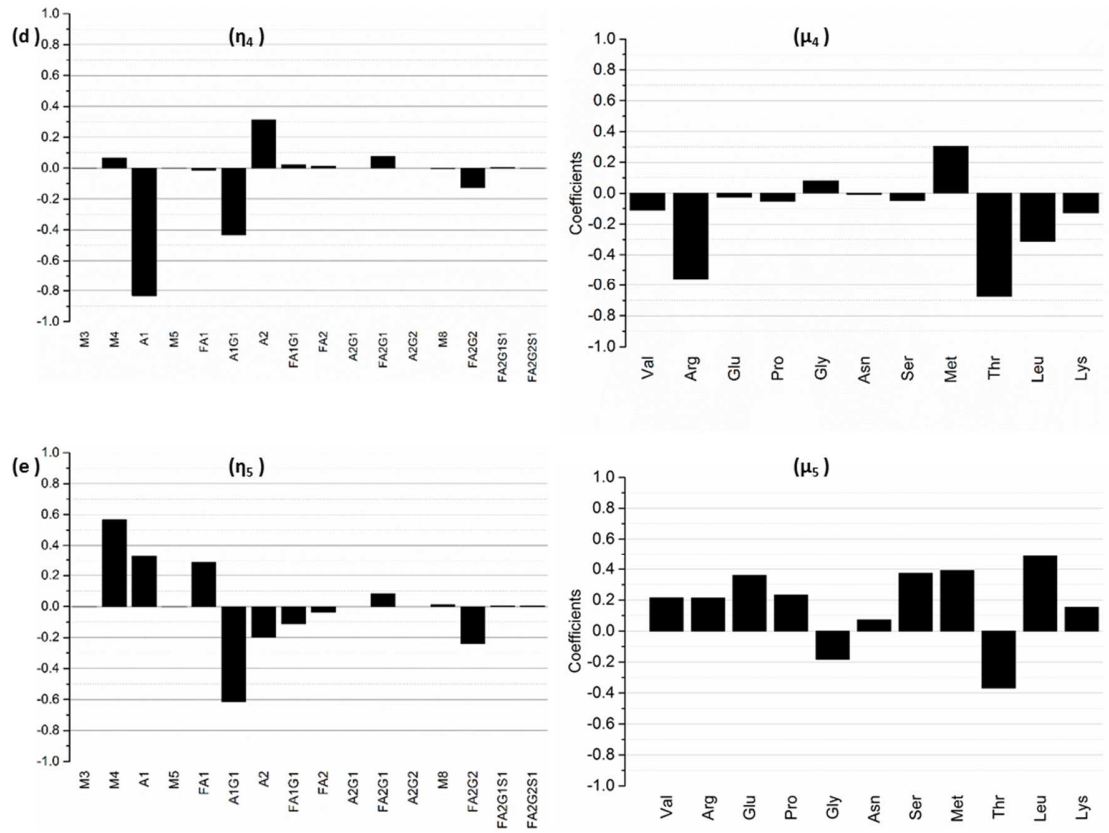


Figure 4.4: Graphical representation of the coefficients associated with the first four input and output modes with glycosyltransferase enzymes as the input factors. Shown here are the primary modes (with $\sigma_i \geq \sigma^*=0.5$) obtained from controllability analysis. (a) Output mode η_1 and the corresponding input mode μ_1 ; (b) Output mode η_2 and the corresponding input mode μ_2 ; (c) Output mode η_3 and the corresponding input mode μ_3 ; (d) Output mode η_4 and the corresponding input mode μ_4 ; (e) Output mode η_5 and the corresponding input mode μ_5

While the role of amino acids in directly regulating the glycan distribution is not well understood, some studies have established a correlation between glycosylation levels and the concentration of byproducts of amino acid biosynthesis. For instance, ornithine which is produced during the degradation of Arg, Pro, and Glu has been shown to modulate high mannose glycan levels (Kang et al. 2015).

Researchers have also noticed the effect of spermine and spermidine, by-products of Arg-Pro metabolism, on altering the activities of glycosyltransferase enzymes (Gréco et al. 2001). Therefore, the changes in the glycosylation profile seen in this experimental work may have accrued due to the by-products of the amino acid metabolism. Conversely, the change in the glycosylation levels seen in this work might have been induced solely due to the changes in the medium pH brought about by the addition of 11 different amino acids to each flask.

Despite not having a complete mechanistic understanding of the effects of amino acids on the glycosylation profile, the presented approach provides us with a quantitative input-output relationship and we can now assess how media supplementation affects the glycosylation profile.

4.3.2 Set-point Tracking under Nominal Conditions

Using the structural input-output relationship developed using controllability analysis, we design controllers to carry out set-point tracking of the glycan distribution as discussed under materials and methods. As each controller gain is determined by the corresponding singular value, and as we only consider singular values greater than a critical cutoff value, we note that the total number of controllers in each case will depend on the number of selected singular values. Thus, when we use glycosyltransferase enzymes as the manipulated variables, we obtain the controller gains from the first four singular values, whereas in the case with amino acid supplementation we use five different controllers based on the first five singular values. In each case, we generate the error signal (ϵ) based on the difference in the desired output mode, η_d , and the measured output mode, η , (obtained by transforming the corresponding desired glycan set-point (y_d) and measured glycan distribution (y),

respectively). Next using P-controllers, designed from select singular values, we obtain the new set of inputs that, when implemented on our process, will shift the process output to match the desired glycan set-point.

First, we assess the set point tracking capability of the controllers under nominal conditions with glycosyltransferase enzymes as the manipulated variables. The resulting output glycan distribution is plotted in Figure 4.5(a), showing that the controllers effectively track the new set-point and the output glycan distribution now matches the new set-point. The predicted set of input glycosyltransferase enzyme concentrations, plotted in Figure 4.5(b) are compared to the enzyme concentrations used in the original simulations to obtain the glycan distribution \mathbf{y}_d . Here we note that unlike the enzyme concentrations of GalT, GnTII, ManI and ManII, those of FucT and GnTI predicted by the controller do not match the enzyme concentrations originally used in the simulation. This discrepancy can be explained by assessing the input modes ($\mu_1 - \mu_4$), where we note that the coefficients of FucT and GnTI are very small in each mode, indicating that any change in their concentration will have negligible impact on the relative glycan distribution. Thus, even though the concentrations of FucT and GnTI predicted by the controller do not match those used in the simulation, we know that this will not influence the final glycan distribution

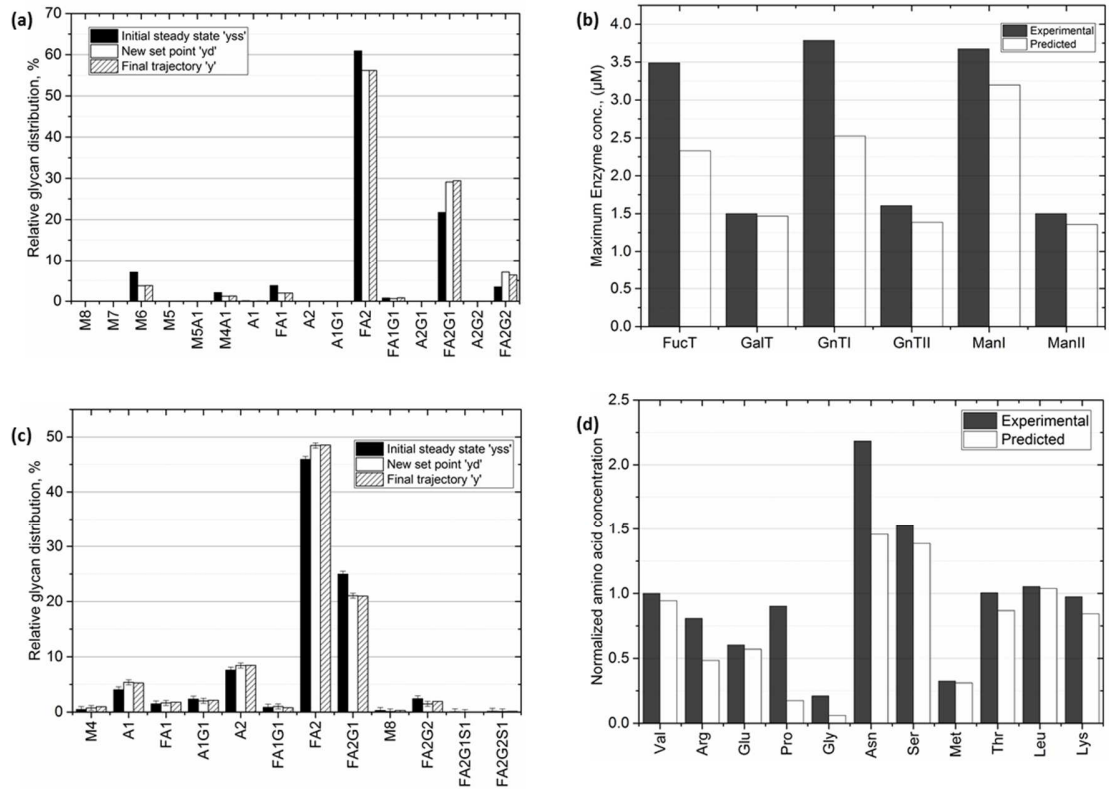


Figure 4.5: Controller performance under nominal conditions. (a) Relative glycan distribution when enzyme concentrations are used as inputs. The solid black bar represents the initial steady state, while the solid white bar represents the new set-point. The cross-hatched bar represents the final output achieved after taking control action. (b) Comparing experimental and predicted inputs. The black bar represents the enzyme concentration used to generate the new set-point and the white bar represents the inputs generated by the controller. (c) Relative glycan distribution when amino acid concentrations are used as inputs with the solid black bar representing the initial steady state, the solid white bar representing the new set-point, and the cross-hatched bar representing the final output achieved after taking control action. (b) Comparing experimental and predicted inputs. The black bar represents the amino acid concentrations (normalized to that of valine) used to generate the new set-point, while the white bar represents the inputs generated by the controller

Similarly, in Figure 4.5(c) we observe the set-point tracking capability of P-controllers where amino acid media supplements are used as manipulated variables. Here, the glycan distribution in the control flask with basal media and no amino acid supplementation is taken to be the measured glycan distribution, \mathbf{y} , while the glycan distribution observed in flask F1 from the experimental design is the desired glycan set-point, \mathbf{y}_d . Once again, using P-controllers designed using appropriate singular values, we are able to track the desired glycan set-point effectively. A comparison of the predicted and experimentally used amino acid concentrations, plotted in Figure 4.5(d), indicates that the concentrations of Asn and Pro do not match. An examination of the input modes indicates that the coefficients of Asn and Pro in the primary input modes is not large enough to alter the glycan distribution significantly.

4.3.3 Set-point Tracking under Model-Plant Mismatch Conditions

In the previous section, we assessed set-point tracking capabilities of the controller under nominal conditions, i.e. when the model is truly representative of the actual process. However, in reality, the true process model is not completely known and is estimated with some error. Hence, it is important to evaluate controller performance under such practical model plant mismatch conditions as well. As described under materials and methods section, we introduce random error to every element in the gain matrix obtained from ANOVA to generate a modified gain matrix. Using the singular values of this modified matrix we design P- and PI-controllers to track the new set-point by performing a series of iterative calculations (equation 4.14 to equation 4.18).

In Figure 4.6, we observe the set-point tracking ability of the P- and PI-controllers using glycosyltransferase enzyme concentrations in the dynamic glycan

model as the manipulated variables. We note that the output glycan distribution after 25 iterations using P- and PI-controllers matches the desired glycan distribution despite the controllers being designed under model plant mismatch conditions (Figure 4.6(a)), while the predicted inputs converge using either P- or PI- controllers (Figure 4.6(b)). A comparison of the sum of absolute errors estimated from the difference in the output glycan distribution and the desired glycan trajectory after each iteration using the P- and PI- controllers indicates that the sum of absolute errors associated with the P-controller decreases steadily and finally converges to fixed error, whereas the output from the PI controller oscillates before the sum of absolute errors eventually decreases. We observe similar trends when we use amino acid media supplements as the manipulated variables to control the glycan distribution under model plant mismatch conditions (Figure 4.7). Here, we note that both the P and PI controllers are able to effectively track the desired glycan trajectory and the predicted inputs under mismatch and nominal conditions are very similar. As seen in the case with enzyme concentrations as inputs under model-plant mismatch, a comparison of the controller performance shows that the sum of absolute error in the P-controller decreases steadily while the sum of absolute error in the PI-controller oscillates.

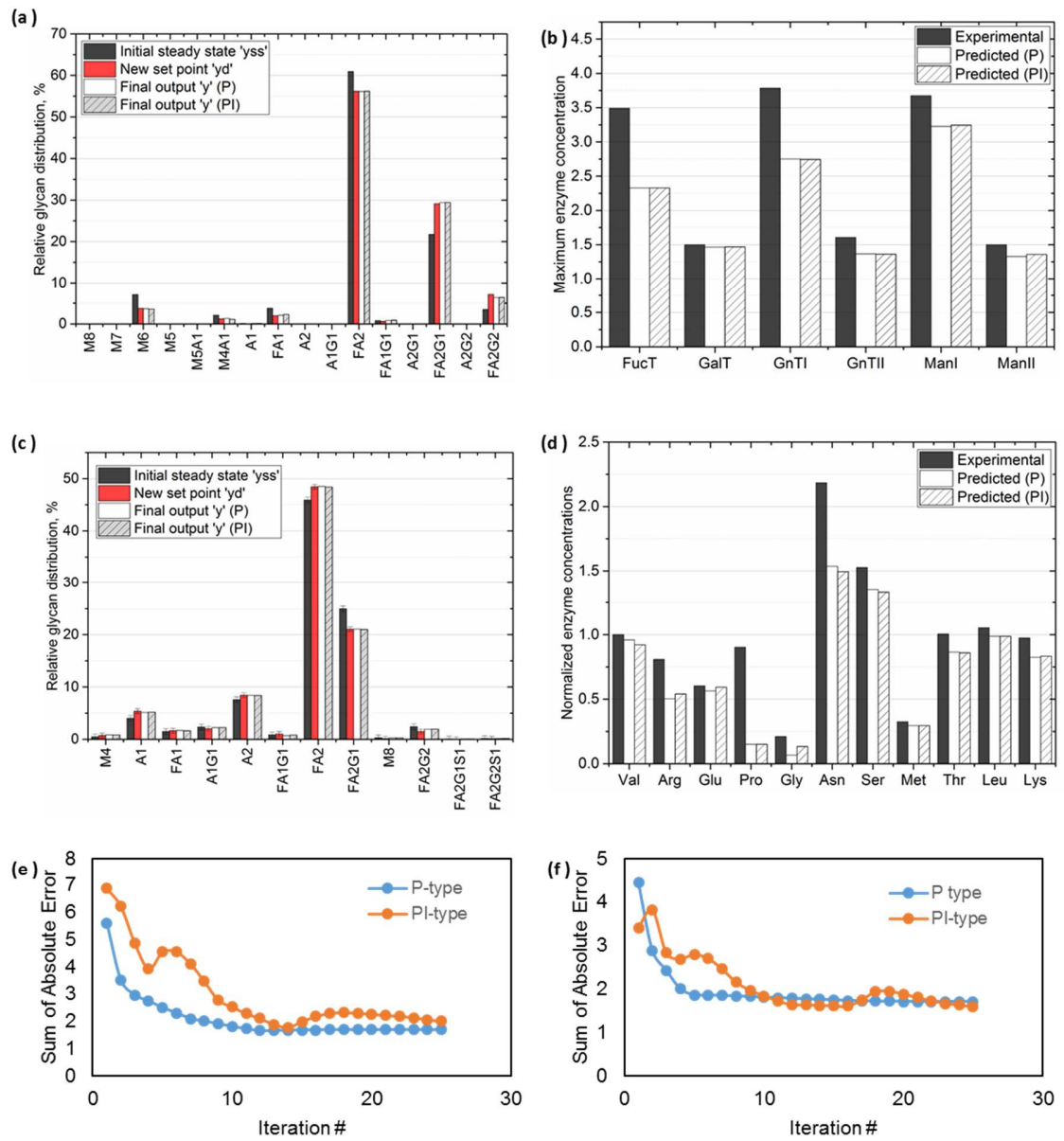


Figure 4.6: P-type and PI-type controller performance under model-plant mismatch conditions. (a) Relative glycan distribution when enzyme concentrations are used as inputs. The solid black bar represents the initial steady state, the solid red bar represents the new set-point, solid white bar is the final output using P-type and grey crosshatched bar is the final output using PI-type controller. (b) Comparing experimental and predicted inputs. The black bar represents the enzyme concentration used to generate the new set-point, solid white bar represents the inputs generated by the P-type, and crosshatched bar represents inputs from the PI-type controller. (c) Relative glycan distribution when amino acids are used as inputs. The solid black bar represents the initial steady state, the solid red bar represents the new set-point, solid white bar is the final output using P-type and grey crosshatched bar is the final output using PI-type controller. (d) Comparing experimental and predicted inputs. The black bar represents the amino acid concentrations (normalized to that of valine) used to generate the new set-point, solid white bar represents the inputs generated by the P-type, and crosshatched represents inputs by the PI-type controller. (e) Sum of absolute error for each successive iteration with P- and PI-type controllers with enzyme concentration as inputs. (f) Sum of absolute error for each successive iteration with P- and PI-type controllers with amino acid concentration as inputs.

In each case studied under model-plant mismatch conditions, we observe that the final glycan output converges to the desired trajectory regardless of the choice of controller used. This indicates robust controller behavior which stems from having used singular values of the gain matrix to determine the controller gains. The choice of select singular values greater than a cutoff value inherently eliminates the singularity in the system and results in a system with a low condition number (which is the ratio of the largest singular value to the smallest singular value), and hence, the designed controllers tend to be more robust, leading to effective convergence.

4.4 Summary and Conclusions

Changes in antibody glycosylation profile affect the final drug product quality significantly, leading researchers to investigate means to control, modulate and finely tune the glycan distribution profile. In this paper we have systematically developed and evaluated a two-step framework for controlling the glycosylation profile in mAbs. First, we identified a quantitative, structural input-output relationship by analyzing data obtained from statistically designed experiments. The structural relationship was then used to design controllers for two cases – (i) where the concentration of glycosyltransferase enzymes in the dynamic glycan model were manipulated to modulate the glycan distribution; and (ii) where different amino acid supplements added to the cell culture media were used as the manipulated variables. We then performed simulations to test the performance of our controller in trying to track a desired glycan set-point under two conditions: one, where we assumed complete knowledge of the process (i.e. nominal conditions) and another, under model plant mismatch conditions. In both these conditions, we observe that the controllers are able to track the set-point effectively and accurately. Even under model plant mismatch conditions, when we introduce random error in our gain matrix, we note that the glycan distribution generated by implementing the input signal from the designed proportional (P) and proportional integral (PI) controllers on the process closely matches the desired glycan output, as the sum of absolute error eventually converges. The robustness in the system is likely introduced by the elimination of process singularity, as a consequence of having chosen singular values above a specific cut-off value (σ^*).

Although it is easier to implement such glycosylation controllers in the second case (using media supplements) as opposed to the first (where glycosylation enzyme

concentrations were modified), it is instructive to demonstrate via simulations the applicability of the proposed framework to both these test cases. The development of glycosylation controllers using media supplements as input factors has a lot of significance to the field of media design. If at the end of a particular batch, the glycan distribution is not as desired and we wish to change the glycosylation profile, then, by implementing the glycosylation controller presented here, we can estimate how to modulate the concentrations of different media supplements in the successive batch so as to achieve the desired glycan distribution profile. Thus, we now have the ability to fine tune and modulate the final glycan profile by simply altering the media composition at the start of the batch. In the case of enzyme concentrations as inputs, the approach presented here provides a rational basis to carry out cell line engineering approaches to alter glycosyltransferase enzyme concentrations. The glycan enzyme concentrations can be targeted in a quantitative fashion to achieve the desired glycosylation profile at the end of the batch. Further, the structural information gained from this approach also indicates which enzyme concentrations have little to no impact on the glycosylation profile and hence, we can target select enzymes that can induce the desired change in the glycosylation profile. The glycosylation controllers designed using this study are thus robust and effective in controlling and modulating the glycan distribution under a variety of conditions.

It is important to note, however, that the controller design implemented here based on output controllability analysis is dependent on the glycan distribution obtained at the end of the batch. Consequently, these controllers are designed as batch-to-batch controllers, where the control action is taken in the subsequent batch and a new glycan distribution profile is obtained. While this is a significant development

based on controllability analysis, in order to achieve our overall objective of online control of glycosylation, we need to design controllers that respond to the changes in the glycosylation profile in real-time. One of the biggest challenges in controlling glycosylation online is the lack of suitable, real-time glycan measurements as current glycan assays are typically performed offline and are associated with long time delays. To overcome these challenges, we explore two different techniques for estimating the glycan distribution in the following chapter. First, we consider how, using observability analysis and by exploiting the underlying network structure inherent in glycosylation, we can identify a subset of glycan species that can be measured to get an estimate of the complete glycan distribution. This approach can lead to the development of novel glycan assays. Next, we use the multi-scale model of glycosylation developed in Chapter 2 to estimate the glycan distribution profile when measurements are unavailable, with a periodic update of the glycan distribution as and when a measurement becomes available. Together, the approaches in the following chapter will lay the foundation for developing online glycosylation control schemes.

Chapter 5

GLYCOSYLATION ANALYSIS: OBSERVABILITY AND STATE ESTIMATION

5.1 Introduction

In the previous chapters we developed quantitative input-output structural relationships between manipulated variables and the glycosylation profile at the end of the batch run using controllability analysis. Further, we utilized these quantitative relationships to develop proportional (P) and proportional integral (PI) controllers to control the glycan distribution and demonstrated the ability of these controllers to track the new glycan set point under nominal and model plant mismatch conditions. However, as controllability analysis is based on changes in the glycan distribution profile in the antibody accumulated at the end of a batch run, the resulting P and PI type controllers designed from the controllability analysis are suited for batch-to-batch or iterative control rather than for real time control.

As stated in Chapter 1, glycosylation is a critical quality attribute and hence, it is necessary to monitor and control the glycosylation distribution during manufacturing to maintain it within acceptable levels (del Val et al. 2010). To achieve real time glycosylation control, it is necessary to have a detailed understanding of the influence of different input factors on the dynamic glycan distribution profile; and reliable on-line or at-line measurements of the glycan distribution profile, as what cannot be observed or measured cannot be controlled. Our efforts in Chapter 2 and 3 were directed towards generating a detailed understanding of how process variables

and media supplements at the bioreactor level influence the glycan distribution profile. In this chapter, we address the issues and challenges associated with analyzing the glycan distribution in mAbs.

The analysis of critical quality attributes and critical process parameters can be carried out *at-line*, which involves sample removal, isolation, and subsequent analysis close to the process stream; *on-line*, wherein the sample is removed from the process stream, analyzed, and returned to the process stream; *in-line*, which involves *in situ* analysis of the sample; or *off-line*, where the sample is removed and analyzed away from the process stream (Rathore et al. 2010). Although it would be ideal to measure all quality attributes such as glycosylation and aggregation on-line or at-line, under practical circumstances, most quality attributes are either measured infrequently, or, as is more commonly practiced, they are measured off-line at the end of the process. This is particularly true of glycan assays where the complexity and diversity of glycosylation profiles makes analysis a non-trivial process, often requiring multiple purification, labeling, and separation steps prior to being quantified (Marino et al. 2010). Consequently, detailed analysis of the glycan distribution profile using glycan assays becomes a laborious, time-consuming, and expensive endeavor requiring multiple instrumentation techniques. Presently, there exist no on-line assays for measuring the glycan distribution profile, while the available at-line glycan assays are used to obtain infrequent measurements that are associated with large time-delays. Thus, not only is there a requirement for developing faster, effective glycan assays, we also need techniques to infer the glycan distribution profile in the absence of frequent measurements. To accomplish both of these tasks, we rely on our understanding and knowledge of glycosylation presented in the previous chapters.

Here, we seek to address the following challenges associated with glycan analysis:

- How do we rationally develop novel glycosylation assays based on an understanding of the glycan reaction network?
- In the absence of real time measurements, how do we reliably estimate the glycan distribution profile using the available dynamic model of glycosylation?

First, we provide a concise review of different analytical methods for glycan characterization along with a brief description of in-house glycan assay development efforts. Next, in §5.3, a rational approach for developing novel glycan measurement assays based on observability analysis is presented. In §5.4, a novel state estimation technique is implemented to estimate the entire glycan distribution in the absence of frequent, real time measurements. Finally, the key findings are summarized at the end of the chapter.

5.2 Glycan and Glycoproteomic Analysis – An Overview

The diversity and heterogeneity in oligosaccharides attached to protein moieties has led to several challenging problems in the emerging field of glycobiology. However, aided by the development of newer and more robust analytical techniques, particularly in mass spectrometry, the scientific discipline of glycomics is witnessing a spurt in novel and high throughput analytical techniques for oligosaccharide characterization. The different analytical approaches for glycan characterization can be broadly classified as: (i) intact glycoprotein analysis; (ii) glycopeptide characterization; (iii) released glycan analysis; and (iv) monosaccharide analysis (Marino et al. 2010). The analytical technique chosen depends on the amount

of sample available, the level of characterization desired, and the nature of information required. Additionally, different applications might require orthogonal assays, or the use of multiple analytical techniques (Mittermayr et al. 2011; Roth et al. 2012). Figure 5.1 shows the different strategies that are available for glycan characterization along with the associated analytical technique.

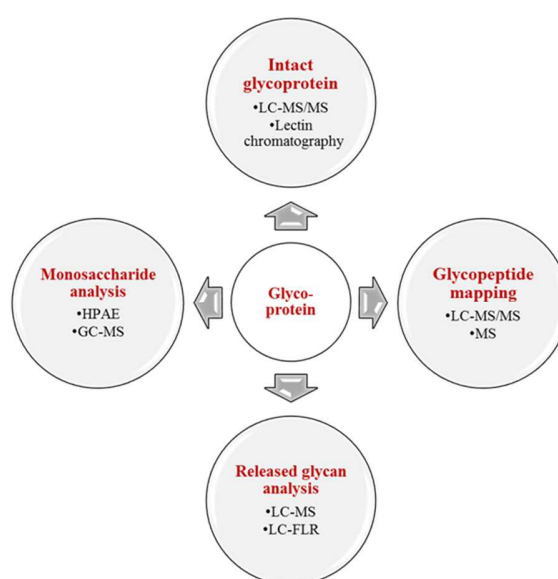


Figure 5.1: Different analytical techniques for glycoprotein characterization. These include characterizing the intact glycoprotein, trypsinizing the glycoprotein to produce glycopeptides that are analyzed separately, using glycosidases to release the glycans and then characterizing the individual glycan species, or breaking down the glycan fragment into individual monosaccharides to get a detailed understanding of the attached sugars.

Intact glycoprotein analysis can be considered a ‘top-level’ analysis wherein the glycoprotein is directly injected to the mass spectrometer (MS) with minimal sample preparation or separated using liquid chromatography before being injected to

the MS (Planinc et al. 2016; Zhang et al. 2016b). The advantage of such intact analysis is that it provides unique information on the attachment of specific glycan species to a particular glycosylation site on the protein backbone (Rosati et al. 2013). Lectin affinity chromatography is another powerful technique for isolating and analyzing intact glycoproteins, wherein lectins with high affinities towards particular oligosaccharides are immobilized to separation matrices such as agarose, resulting in techniques with high specificities (Mechref and Novotny 2002). The recent development of chemiluminescence-based lectin-binding assays have further expanded the applicability to high throughput glycan analysis (Onitsuka and Omasa 2015).

A ‘middle level’ approach involves proteolytic digestion of the glycoprotein that results in the formation of intact glycopeptides that are then separated, purified and analyzed using matrix assisted laser desorption/ionization time of flight (MALDI-TOF) MS (Bailey et al. 2005), or using electrospray ionization (ESI) MS following separation by liquid chromatography (LC) (Medzihradszky 2005). Such a mid-level analysis helps identify the glycosylation site and characterize the site-specific glycan heterogeneity in samples with multiple glycosylation sites. Note that, this approach at characterizing the glycopeptide differs from other techniques to identify the site *occupancy*. For instance, techniques have been developed using HPLC (St. Amand et al. 2014a) and Raman spectroscopy (Brewster et al. 2011) for characterizing the glycan macroheterogeneity, that are distinct from the techniques described here for glycan site heterogeneity. While the former techniques assess and quantify the level of deglycosylation in the sample, the latter characterize the glycan species at specific glycosylation sites.

The bottom level analysis involves proteolytic cleavage of the glycoprotein followed by a release of the N-glycans using enzymatic release (with glycosidase enzymes such as PNGase-F) or chemical release (using sodium borohydride). The released glycans are captured and derivatized with a chemical tag followed by spectroscopic analysis. While there are several different glycan labeling strategies, the commonly used techniques are based on reductive amination – wherein a label such as 2-amniobenzamide (2-AB), 2-aminobenzoic acid (2-AA), or 2-aminopyridine (2-AP) which contain primary amine groups react with the aldehyde group of the glycan forming a Schiff base; or are based on (per)-methylation – where hydrogens on the hydroxyl group, amine group, and carboxyl groups of the glycan are replaced by methyl groups from the added methyl iodide (Ruhaak et al. 2010). Labeled glycans are then analyzed using HPLC or MS.

Bottom-level analysis gives an in-depth characterization of the different glycan species present in the mAb and is required by all regulatory agencies. However, the detailed information received from a bottom level analysis comes at the expense of local information such as protein binding site, glycan pairing on the adjacent heavy chain glycosylation site, etc. Extensive research has been undertaken on the development of such assays that characterize the microheterogeneity in the antibody sample. As these assays require elaborate sample preparation, they are usually comprised of multiple process steps and efforts have been directed at optimizing each process step. For instance, following the purification of the glycoprotein, the enzymatic release of the glycan species requires overnight incubation. However, by using microwave based deglycosylation techniques, the overall deglycosylation time was reduced from 16 hours to a matter of minutes (Sandoval et al. 2007a; Sandoval et

al. 2007b). The subsequent choice of the labeling tag is dictated by the final analytical technique used for characterization. Different tags have their specific advantages and disadvantages and are optimized for each application. In our work, we have extensively used glycan permethylation as reported in the previous chapters to label the released glycans using the Ciucanu method (Ciucanu and Costello 2003; Ciucanu and Kerek 1984). Here the methyl iodide is added in the presence of a strong base such as sodium hydroxide followed by liquid-liquid extraction and solid phase extraction prior to analysis by MALDI-TOF (Kang et al. 2008; Lin and Lubman 2013). Alternatively, both 2-AB and 2-AA labels have been used as fluorescent tags followed by subsequent purification and glycan characterization using a variety of analytical techniques – such as capillary electrophoresis (Kamoda et al. 2006), normal phase HPLC (Royle et al. 2007), UPLC or ultrahigh performance liquid chromatography (Burnina et al. 2013), and HILIC or hydrophilic interaction liquid chromatography followed by MS (Shang et al. 2014). Label-free analytical methods using coupled orthogonal methods such as LC-MS and ¹H-NMR have also been implemented to characterize the glycan distribution in commercial mAbs (Wiegandt and Meyer 2014).

The relative comparison between different characterization methods such as Reverse Phase (RP)-UPLC, HILIC-UPLC and CE-LIF has also been studied, highlighting the disadvantages of using the RP-UPLC method (Adamczyk et al. 2014). Efforts have also been made to increase the throughput of these characterization methods by introducing automated sampling and analytical techniques thereby reducing sample time (Doherty et al. 2013; Tharmalingam et al. 2015). The methods listed so far are by no means exclusive, as a glycoprotein may be characterized by

more than one orthogonal technique (Klapoetke 2014; Stadlmann et al. 2008; Wagner-Rousset et al. 2008).

Monosaccharides are obtained by acid hydrolysis of the glycoprotein and can be analyzed using high performance anion exchange (HPAE) chromatography or by gas chromatography followed by mass spectrometry (GC-MS). They are often used to characterize the individual sugars that comprise any oligosaccharide.

With the advancement of analytical techniques and the increased understanding of the effect of glycosylation on other quality attributes such as aggregation and charge variants, there have been further advances in the development of multi-attribute methods wherein more than a single quality attribute is measured using orthogonal techniques (Rogers et al. 2015). Additionally, increasing regulatory requirements have also led to the development of several high-throughput techniques for glycan characterization at all levels (Aich et al. 2016). However, despite these advances, glycan analysis still remains an elaborate and time-consuming process. While a few assays have been developed to achieve near real-time analysis of glycan species (Tharmalingam et al. 2015), a majority of the techniques remain off-line or require elaborate sample preparation.

For instance, the glycan permethylation technique used in this thesis involves multiple sample preparation steps – starting from antibody purification, to glycan cleavage, capture, labeling, liquid-liquid extraction, and final clean-up – before they can be analyzed by MALDI-TOF. These sample preparation steps can take over 48-60 hours before the glycan distribution is fully characterized and hence, are not amenable for real time or near real-time glycan analysis. Some of the newer reagents being developed for faster glycan labeling remain prohibitively expensive and are more

suitable for final quality analysis before product release than for in-process use. There is thus a growing necessity to develop better glycan assays or, in the absence of glycan measurements, to have reliable inferential estimates of the glycan distribution. In the rest of this section, we present in brief, our efforts at developing an in-house HILIC based glycan assay followed by modeling approaches to address the challenges arising due to infrequent measurements.

5.2.1 In-house Glycan Assay Development

An in-house HILIC based glycan assay was developed for characterizing the glycan distribution profile in IgG1 samples obtained from our experiments. IgG1 was purified from spent media using a PhyNexus Benchtop MEA2 system using Protein A chromatography resin packed in a 2 mL PhyTip column. The antibody was then denatured using SDS followed by overnight enzymatic deglycosylation with PNGase-F at 37° C in a water batch. Microwave based deglycosylation techniques were also tested, but the technique was marred with issues of reproducibility. Following deglycosylation, the glycans were separated from the denatured protein by buffer exchanged followed by centrifugation, with the glycans collecting in the supernatant. The supernatant was then dried and labeled with 2-aminobenzamide (2-AB) at 65° C for 3 hours. Following glycan derivatization, excess reagent is washed off using a PhyNexus Benchtop MEA2 system with normal phase chromatography resin packed in a 2 mL PhyTip column. The labeled glycans are then analyzed on a Waters BEH XBridge Glycan column (250 mm, 4.6 mm o.d., 3.5 micron particles) following hydrophilic interaction chromatography (HILIC) on an Agilent 1100 system fitted with a fluorescence detector. The total glycan analysis time was further reduced by using non-linear gradient chromatography to enhance separation. The resulting glycan

peaks were characterized by comparing the retention times to the retention times for a 2-AB labeled dextran ladder that was analyzed using a similar elution gradient and the glycan peaks were then identified on the basis of their glucose unit (GU) values by searching the GlycoBase database (GlycoBase, NIBRT, Ireland).

The in-house assay reduces sample preparation and overall analysis time to a little over 48 hours in comparison to the longer permethylation assay. Thus, this assay can be potentially used for analyzing in-process samples. However, even with a reduced analysis time, we are faced with the issue of the long time delay associated with this assay. We can address this issue by either: (i) developing a different assay that would require fewer measurements to be made, thereby reducing sample time; or (ii) utilize a mathematical model to predict the glycan distribution taking into account the associated time delays. These challenges have been addressed in §5.3 and §5.4 respectively.

5.3 Designing Novel Glycan Assays using Observability Analysis

As our brief review shows, glycan analysis is an exciting and challenging frontier in glycobiology and has seen considerable development in the past couple of decades. The advancement of analytical tools along with the growing requirements from manufacturers and regulatory agencies alike to characterize the glycan distribution have led to the development of a variety of glycosylation assays. However, these assays require skilled personnel, expensive reagents, and/or detailed sample preparation prior to characterization. To characterize these diverse and complex glycosylation profiles, faster or simpler assays have to be developed that can be implemented on-line or at-line during manufacturing.

One characteristic feature of glycan species that has not been widely exploited in the development of current glycan assays is the underlying glycan reaction network. As noted in §1.5.1, although glycosylation is a non-template driven process, the localization of different glycosyltransferase enzymes causes specific glycan species to be produced in different regions of the Golgi apparatus. Further, the formation of late stage glycans is dependent on the glycan species that are produced prior to it. Thus, the sialic group will only be attached to a galactosylated glycan; the galactose group, in turn, only attaches to an N-acetylglucosamine precursor; and so on. Thus, the individual glycan species are interconnected forming an elaborate glycan reaction network whose topology can be studied to understand the impact of specific glycan species on other glycan intermediates. Such a topological study of the glycosylation reaction network has been studied previously to understand the modularity in these reaction networks and develop techniques for glycoengineering of therapeutic proteins (Kim et al. 2009a). Here, we use the information contained in such a glycan reaction network to guide glycan assay development. Essentially, we propose that, in lieu of measuring each individual glycan species, it might suffice to measure a judiciously selected set of glycans that contain information about the rest of the glycan reaction network. Assays developed for these “secondary” measurements would thus contain the relevant information needed to infer the glycan distribution profile.

The use of secondary or indirect measurements to estimate target variables is not new and the use of such soft sensor techniques for bioprocesses has gained a lot of traction in recent years (Mandenius and Gustavsson 2015). To be able to design and implement these soft sensors, we must first answer a critical question: for a given set of secondary (indirect) measurements, which combination would be suitable for

estimating the target variables (Golabgir et al. 2015)? In the context of our work, we frame this question as: which combination of glycan species, when measured together, would be suitable to estimate the complete glycan distribution profile? Having identified this combination of glycan species, we can then explore the feasibility of developing novel glycan assays targeted towards these specific information-rich glycan measurements.

To address this work, we rely upon observability analysis, which reflects the possibility of estimating the internal states of a system based on the input/data available over a finite time horizon (Ray 1984). In observability analysis we aim to infer the internal state or the glycan distribution for our system using the available output measurements, making it a dual to the controllability analysis that was previously described in §3.2.5. However, for a system as complex as glycan reaction network, where the enzyme kinetics for each glycan species are described using Michaelis-Menten reaction kinetics, we note that the use of standard observability analysis poses a mathematically intractable problem. Consider, for instance, a generic, time-invariant state space model describing the dynamics of our variable of interest, \mathbf{x} (the glycan distribution profile), as given by:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \mathbf{u}) \quad 5.1$$

where \mathbf{u} represents the associated inputs and the output measurements, \mathbf{y} (that we wish to monitor using the glycan assay) are given by:

$$\mathbf{y} = h(\mathbf{x}, \mathbf{u}) \quad 5.2$$

Applying standard observability analysis enables us to uniquely infer the initial state $\mathbf{x}(0)$, provided the following observability criteria is met – the observability matrix obtained as the Jacobian of the Lie derivative vector of the output must be full rank. Clearly, for a small glycan reaction network with just 20 glycan species, the associated Jacobian calculations become computationally challenging. Also, a complete observability analysis only serves to assess if our selected set of measurements renders the system observable, without providing any guidance as to the selection of specific sensors. As our objective is to identify the particular sensors whose measurement will aid in characterizing the complete glycan distribution, we seek alternative means of performing the observability analysis.

One such approach to identify the sensors that describe the internal states of a complex systems uses a graphical approach, wherein inference diagrams are drawn, connecting the different states that are interconnected and subsequently, strongly connected components (SCC) decomposition is performed on these inference diagrams to uniquely identify the nodes that can be selected as sensors (Liu et al. 2013). However, the challenge of applying such a graphical approach to our problem lies in the inherent linear nature of the glycan reaction network. Although the glycan species are interconnected, only *upstream* glycan species (i.e. glycans produced in the initial compartments of the Golgi apparatus such as high mannose glycans and biantennary glycans) have an effect on the *downstream* glycan species (i.e. glycans produced in the latter regions of the Golgi compartment such as complex glycans and hybrid glycans) and not vice-versa. Consequently, when we apply strongly connected components decomposition using the in-built *graphnoncomp* function in MATLAB for the 77 glycan reaction network presented in the DK2011 model (del Val et al.

2011), we note that every glycan in the resulting inference chart becomes a potential sensor, implying that we need to measure every single glycan under this scheme (see Figure 5.2). As the graphical approach does not assist us in sensor selection, we examine an alternate based on the stoichiometry of the system.

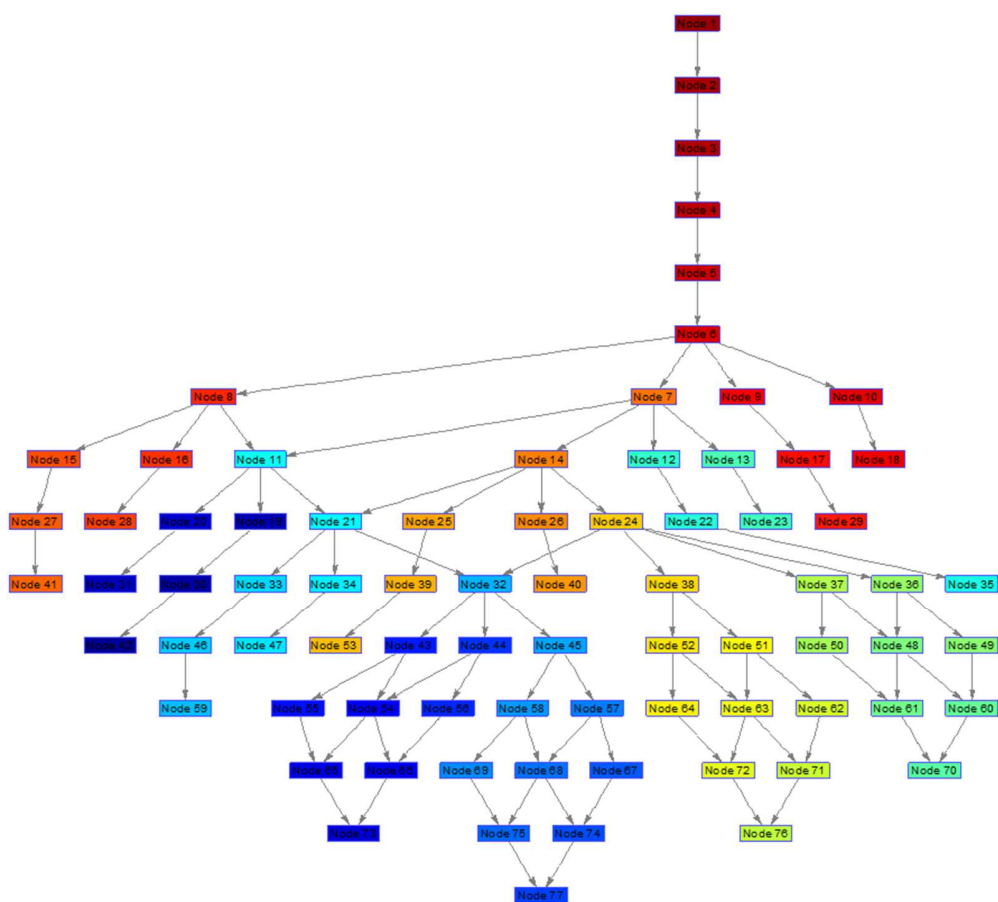


Figure 5.2: Inference maps based on SCC decomposition of the glycan reaction network obtained from DK2011 model. Every single glycan is identified as a potential sensor based on this map.

5.3.1 Selecting Measurement Variables for Analysis

To identify a subset of glycan species that can be measured to provide maximal information of the complete glycan distribution, we examined the stoichiometry of the glycan reaction network. The interdependencies of the glycan reaction network indicate that selecting any subset of glycans from the complete set of glycan species would pose subsequent computational challenges in defining the measurement matrix. Recall from equation 5.2, that the measurements are defined using a measurement model, \mathbf{h} , and are dependent on the inputs, \mathbf{u} , and the internal states, \mathbf{x} . Further, in Chapter 2, we noted that the mass balance on the glycan species involve coupled PDEs using detailed Michaelis-Menten rate equations. Thus selecting a particular subset of glycans as our output measurements, \mathbf{y} , and estimating the complete glycan distribution, \mathbf{x} , would result in a highly non-linear measurement model, \mathbf{h} . To circumvent this issue, we avoid selecting individual glycan species as our measurements and instead use ‘grouped variables’ (similar to the glycosylation indices presented in Chapter 3) as our measurement variables. For instance, our measurement variable ‘ F_1 ’ is defined as the sum of the relative distributions of all fucosylated glycans, while the sum of the relative distributions of all afucosylated glycans is defined as our measurement variable ‘ F_0 ’.

$$F_1 = \%FA2 + \%FA2G1 + \dots + \%FA2G2 + \dots \quad 5.3$$

$$F_1 = 0 \times \%M8 + 0 \times \%M7 + \dots + 1 \times \%FA2 + 1 \times \%FA2G1 + \dots \quad 5.4$$

The complete set of measurements used are listed in Table 5.1. Note that, in addition to the fucosylated and galactosylated measurement variables, we have introduced two additional variables – Man (M), the sum of relative distribution of all

glycans with greater than three mannose groups attached; and Ratio (R), the relative distribution of two most abundant glycans found in the system.

Table 5.1: List of new measurement variables and their definitions

Selected measurement	Definition
F ₀	Sum of relative distribution of afucosylated glycans
F ₁	Sum of relative distribution of fucosylated glycans
G ₀	Sum of relative distribution of agalactosylated glycans
G ₁	Sum of relative distribution of mono-galactosylated glycans
G ₂	Sum of relative distribution of bi-galactosylated glycans
Mannose, M	Sum of relative distribution of high mannose species (> 3)
Ratio (R)	Relative distribution of the two most abundant species

The biggest advantage of using ‘glycan groups’ as our measurement variables in lieu of individual glycan species is the resulting simplification of the measurement matrix (Equation 5.2). Instead of a highly non-linear measurement matrix, we now have a linear measurement matrix, where most elements are either 0 or 1. Thus, the linearized measurement model can be written as

$$\mathbf{y} = \mathbf{C}\mathbf{x} \quad 5.5$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdot & \cdot & c_{1n} \\ \cdot & \cdot & & & \cdot \\ c_{m1} & \cdot & \cdot & \cdot & c_{mn} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \quad 5.6$$

By this simplification, we reduce the issue of determining the complete glycan distribution profile from select measurements to a mathematically tractable matrix inversion problem. As the measurements included herein are sums of relative distributions, we are able to introduce additional constraints as given below

$$F_0 + F_1 = 100\% \quad 5.7$$

$$G_0 + G_1 + G_2 = 100\% \quad 5.8$$

$$x_i \geq 0 \quad 5.9$$

$$\sum x_i = 100\% \quad 5.10$$

These constraints include the non-negativity constraint on the glycan distribution and a summation constraint on the different measurement variables. The implication of the summation constraint is that we cannot choose to measure the fucosylation indices (F_0 , F_1) and the galactosylation indices (G_0 , G_1 , G_2) simultaneously, as the resulting measurement matrix would be rank deficient. Similarly, if we choose measurement variables such that the summation constraint in equation 5.10 is not satisfied, (for instance, by choosing F_0 , G_0 , G_1 , M , R), then we must modify the measurement matrix appropriately to satisfy that constraint. Regardless of how many measurement variables are chosen, the number of glycans to be estimated will be greater, i.e. in equation 5.6, the number of measurements, m , will be less than the number of glycans to be estimated, n . Thus the stoichiometric or measurement matrix, \mathbf{C} , is non-square and we solve for the glycan distribution using a non-negative least squares solver.

To assess the quality of glycan distribution predictions arising from our choice of measurement variables, we test the proposed calculation technique against experimental data. Briefly, an IgG1 producing CHO-K1 cell line was grown under standard growth conditions described in the previous chapters and the glycan distribution at the end of the batch was analyzed by permethylation followed by MALDI-TOF. A total of eighteen glycan species were experimentally observed and their relative glycan distribution was obtained. The experimental data was then used to calculate the values of the measurement variables F_0 , F_1 , G_0 , G_1 , G_2 , M , and R as defined in Table 5.1. Next, we estimate the glycan distribution using the non-negative least squares solver in MATLAB and compare the estimated glycan distribution to the experimentally observed glycosylation profile. The root mean square error (RMSE) is used to evaluate the quality of the predicted glycan distributions from our analysis, as demonstrated in the following section.

5.3.2 Results from Observability Analysis

To estimate the glycan distribution profile, we first choose a measurement vector with as few measurements as possible. Here, we choose to measure the fucosylation indices (F_0 , F_1) and the mannose value (M) in addition to the ratio (R) of the two most abundant glycan species. Figure 5.3 shows a comparison of the measured and estimated values.

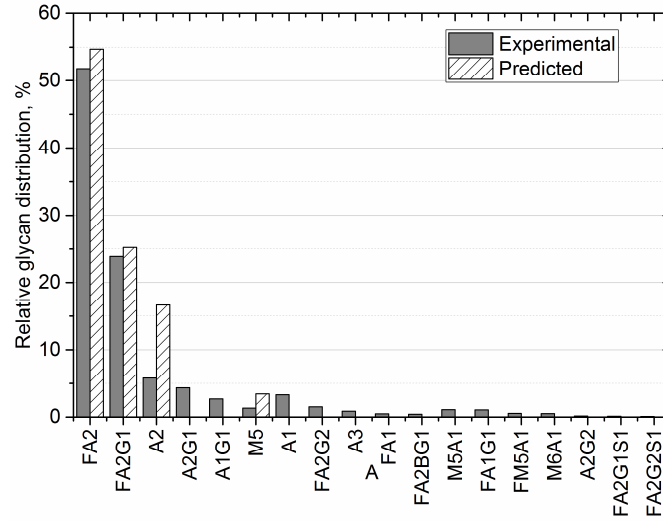


Figure 5.3: Experimental and predicted value of the glycan distribution using observability analysis with four measurements. The measurements included here are F_0 , F_1 , M , and R

We note that with just 4 measurement variables, the resulting estimate has an RMSE value of 13.27% as it closely accounts for most of the major glycan species. We expect that increasing the number of measurements will improve the quality of our estimate and accordingly, we choose to measure G_0 in addition to the measurement variables previously selected. Addition of a variable increases the rank of the measurement matrix and the RMSE decreases as expected to 9.27%. If we define an acceptable cutoff for the measurement assay at 10% RMSE, then an assay with just five measurements (F_0 , F_1 , G_0 , M , R) would be sufficient to get a reasonable estimate of the glycan distribution. If we now choose to include a sixth measurement, say G_1 , then the rank of the measurement matrix increases to 6 and the RMSE decreases to 9.09%. Thus the improvement in the quality of the estimate with the addition of a sixth measurement is only marginal in comparison to the improvement in the quality of the

estimate when a fifth measurement is added. Figure 5.4 plots the estimated glycan distribution for each scenario against the experimentally determined values.

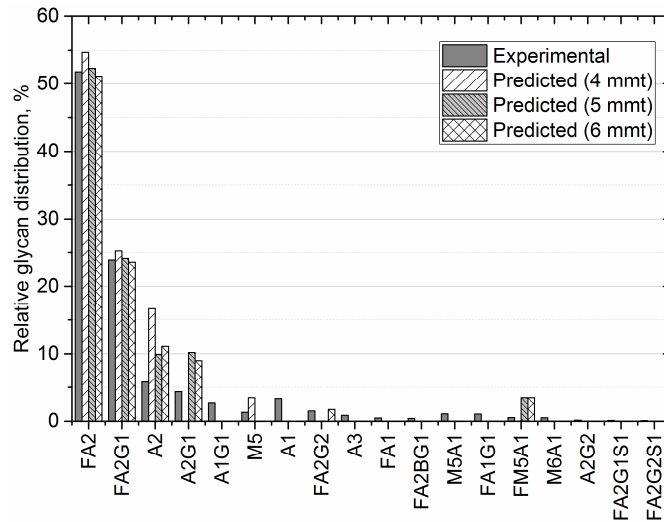


Figure 5.4: Comparing experimental and predicted values obtained from glycosylation analysis with four, five, and six measurements.

Additionally, we check to see if we can comment on the quality of the prediction *a priori* by checking specific metrics related to the measurement matrix. One such obvious metric that can be tested is the condition number of the measurement matrix. When we plot the RMSE against the condition number (Figure 5.5), we notice a general trend in that the increasing condition number tends to represent lower RMSE. This is expected as each additional measurement increases the rank of the matrix and the additional measurement improves the quality of the estimate. Thus, once we select a set of measurement variables and examine the

condition number, we might plausibly be able to judge if the quality of the estimate will be within the defined acceptable limit.

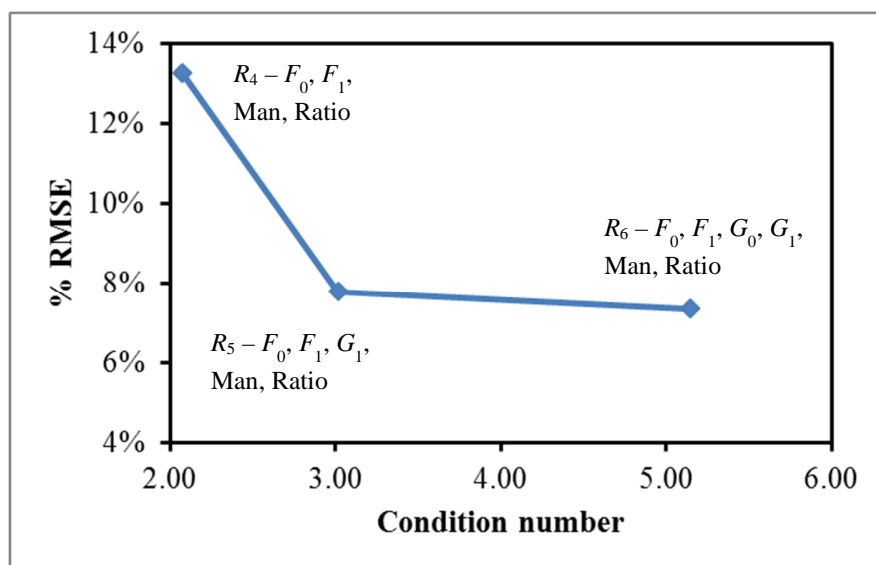


Figure 5.5: Root mean squared error (RMSE) of the glycan prediction for different scenarios plotted against the corresponding condition number of the measurement matrix. We see that with four measurements, the predicted measurements have a lower condition number but a higher RMSE, while with six measurements, the RMSE is lower.

Finally, we note that the choice of different measurements is a combinatorial problem. Assuming that the ratio (R) is measured each time, we can come up with forty different choices involving different combinations of variables that comprise the measurement vector. These include cases with four measurements, five measurements, and six measurements, as well as the rank deficient case where both fucosylation and galactosylation indices are included. In cases where the chosen measurements do not meet the summation constraint, the measurement matrix is

adjusted accordingly and the glycan distribution is then estimated. The resulting RMSE values for each of the 39 cases is plotted as a function of the matrix condition number.

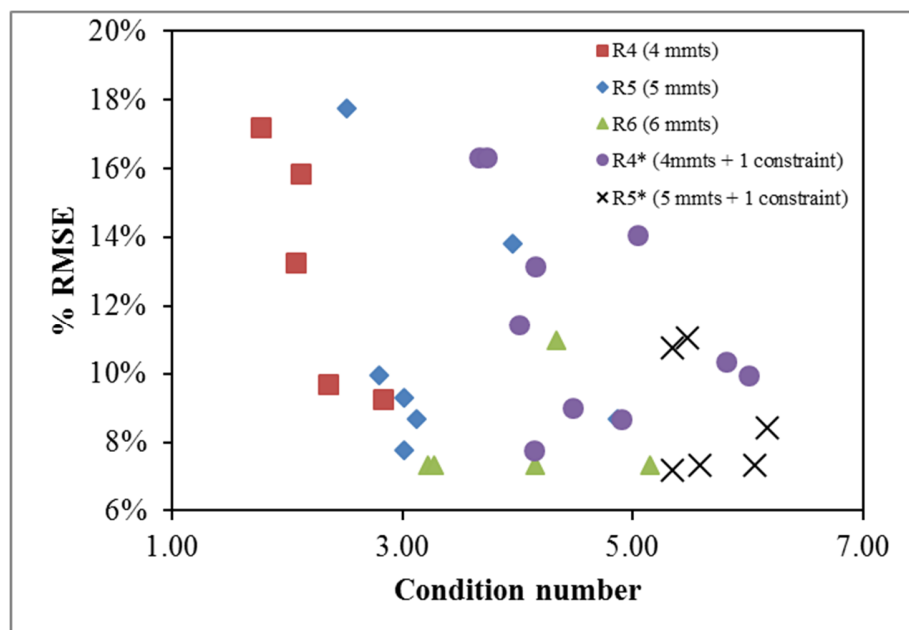


Figure 5.6: Root mean squared error (RMSE) for glycan predictions with each of the 39 cases tested plotted against the corresponding condition number. The red squares indicated cases with four measurements, the blue diamond represents cases with five measurements, the green triangle cases with six measurements. As some measurement choices do not meet the summation constraint, they are added to the measurement matrix. The purple circles consist of 4 measurements with the constraint added, while the black cross marks are for five measurements with a constraint added.

Our analysis shows that it is not just the number of measurements that matter, but also the choice of a measurement variable can have a significant impact on the overall estimate. For instance, for a given number of measurements, if we take an

average of the mean of squared errors, we note that the quality of the estimate improves when we exclude G₂ or Man as a measurement variable. This is likely because the information content in G₂ and Man is lower than that contained in the other measurement variables, as there are fewer high mannose and bi-galactosylated glycoforms.

Overall, the analysis presented here is a means to assess how measurement variables can be selected to obtain reasonably acceptable estimates of the glycan distribution. We must note that in each case we are estimating a large glycan distribution with a limited set of measurements, and hence there is some sacrifice in the quality of the estimate. However, the information content in all measurement variables is not the same and given a limited set of measurements, this method can be used to rationally design assays that measure the required variables.

While observability analysis helped us understand how novel assays can be developed, in the subsequent section, we address how the glycan distribution can be estimated using a dynamic model with intermittent measurement.

5.4 Estimating Glycan States

As discussed in §5.2, measuring the glycan distribution is laborious, expensive, and time-consuming; consequently, glycan measurements are made at infrequent intervals, with no reliable on-line monitoring of the glycosylation profile. Although progress is being made in monitoring real-time glycan distribution using some of the more advanced assays (Tharmalingam et al. 2015), the established methods for measuring the complete glycan distribution profile are still primarily offline. The absence of any commercially available assays for continuous, on-line monitoring of glycosylation, renders on-line control of glycosylation extremely difficult and

necessitates the development of alternate approaches to monitor the glycan distribution. In §5.3, we discussed how stoichiometric information of the glycan reaction network can be used to develop novel glycan assays based on output observability analysis. In this section, we address the issue of infrequent measurements directly by employing dynamic models described in detail in Chapter 2 to estimate the glycosylation profile.

5.4.1 Overview of State Estimation

State variables are those process variables that uniquely specify the internal condition (or state) of a process at any given time. In order to achieve effective control of a dynamic process, it is necessary to have reliable, real-time information on these state variables (Ogunnaike and Ray 1994b; Soroush 1998). However, in practice, not all state variables can be monitored on-line due to considerations of cost or analysis time, or even the availability of reliable sensors. The lack of frequent, on-line or at-line measurements of critical process variables is commonly encountered in the process industry, with routine analysis of samples being conducted offline and/or on an infrequent basis. Without on-line sensors for all state variables, regular feedback control of the process is rendered infeasible and hence, inferential techniques are used to reconstruct or estimate the state variables based on available measurements (Ogunnaike and Ray 1994a). The general structure of such a ‘state estimator’ is presented in Figure 5.7.

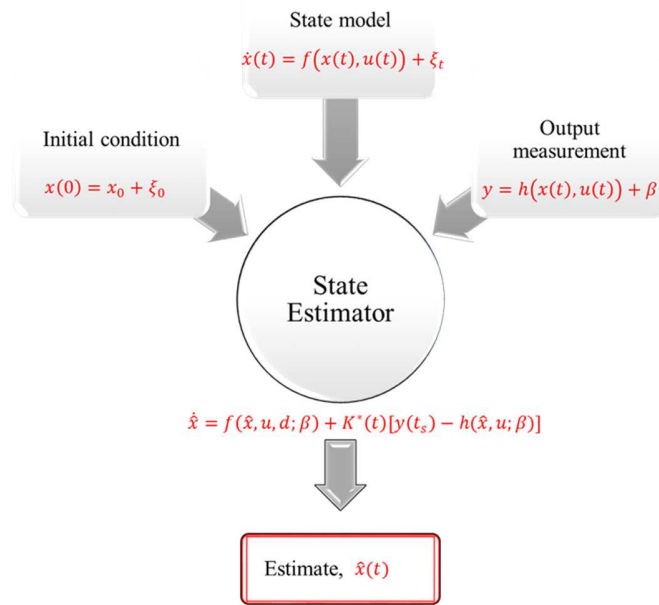


Figure 5.7: General structure of a state estimator. Adapted from Ogunnaike and Ray (1994b)

State estimators static or dynamic, deterministic (Dochain 2003) or Bayesian (stochastic) (Patwardhan et al. 2012) systems that are widely used to estimate the unmeasured states in chemical (Ogunnaike 1994; Zambare et al. 2003) and biochemical processes (Dewasme et al. 2015; Fernandes et al. 2015; Gudi et al. 1995; Tatiraju et al. 1999). The main components of a state estimator are as listed below.

A state estimator requires a dynamic model of the different states of the system \mathbf{x} , with process input \mathbf{u} , and model error ξ usually given by

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{u}) + \xi(t) \quad 5.11$$

The initial condition for the states, as represented by

$$\mathbf{x}(0) = \mathbf{x}_0 + \xi_0 \quad 5.12$$

The measurement device which produces the signal of measured values, \mathbf{y} ,

$$\mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, \beta) + \boldsymbol{\eta} \quad 5.13$$

where the measurement model depends on the parameters, β and the measurement signal contains noise, $\boldsymbol{\eta}$.

Based on these components, we develop a state estimator that takes the form

$$\frac{d\hat{\mathbf{x}}}{dt} = \mathbf{f}(\hat{\mathbf{x}}, \mathbf{u}) + \mathbf{K}^*(t)[\mathbf{y}(t) - \hat{\mathbf{y}}(t)] \quad 5.14$$

$$\hat{\mathbf{y}} = \mathbf{h}(\hat{\mathbf{x}}, \mathbf{u}, \beta) \quad 5.15$$

Here, the variables $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ refer to the on-line estimates of \mathbf{x} and \mathbf{y} respectively, and $\mathbf{K}(t)$ denotes the correction gain matrix. The first term in the state estimator is the prediction based on the process model, while the second term denotes the correction to the process model calculated as the difference between the actual measurement, \mathbf{y} , and the estimated value of the measurement signal, $\hat{\mathbf{y}}$ predicted by the model, $\mathbf{h}(\hat{\mathbf{x}}, \mathbf{u}, \beta)$. In the next section, we identify the state variables and measurements in our process.

5.4.2 System and Model Description

The multi-scale dynamic model for glycosylation has been described in detail in Chapter 2 of this thesis. Briefly, we use a macro-scale model developed using Monod kinetics to calculate the antibody productivity based on cellular growth. The glycan productivity, which is obtained from the antibody productivity in the macro-scale model, is used in the adapted DK2011 model to obtain the glycan distribution profile. The values of the process variables measured at the macro-scale (nutrient concentration, cell densities, antibody concentration) are available to us at the time scale of an hour, which when compared to the time required for analyzing the glycan distribution profile (> 48 hours), is practically insignificant. Thus, the macro-scale process variables are available at much faster rates and at higher frequencies than the micro-scale glycan distribution, as listed in Table 5.2

Table 5.2: Measurement delay and sampling frequencies for different measurements

Process Variable	Time delay (hours)	Sampling frequency (days)
Glucose concentration	0.0	1
Glutamine concentration	0.0	1
Lactate concentration	0.0	1
Ammonia concentration	0.0	1
Viable cell concentration, X_v	0.0	1
Total cell concentration, X_t	0.0	1
Antibody concentration	0.5	2
Glycan distribution	48.0	2

For the purposes of the present discussion, we evaluate the performance of the state estimator assuming that there is adequate knowledge of the macro-scale model.

Thus, we assume that there is *a priori* knowledge of cell growth profile, antibody productivity profile and hence, the glycan productivity profile. Such an assumption would be valid, considering that macro-scale concentration measurements are available at a significantly faster rate and a greater frequency than glycan measurements. Further, by assuming that the macro-scale model correctly predicts the antibody and glycan productivity rate (which form the inputs for the micro-scale model), we deduce that differences observed between the measured and the predicted glycan distribution profile are due to inadequacies in the micro-scale model alone. This assumption helps simplify our calculations and assess the performance of the state estimator. Before we get to designing the state estimator, however, we must first define the state and measurement variables and the attendant equations used in the dynamic state model.

The absolute concentrations of the different glycoform that accumulate in the flask over the period of the batch culture represent the internal state of the system that we wish to monitor, whereas typical glycan assays provide the relative distribution (or percentage distribution) of the individual glycan species in the flask at a particular sampling time. Thus the absolute accumulated concentrations of individual glycoforms are the state variables of interest (\mathbf{x}), whereas the relative glycan distribution (or percent glycan distribution) of each glycoform obtained from the glycan assay constitutes the measurement variable, \mathbf{y} . As the absolute concentrations of different glycan isoforms remain unmeasured and as glycan measurements are available only at infrequent intervals, we employ a dynamic model based state estimation technique to obtain real time estimates of the state variables.

Having defined our state and measurement variables, the resulting system equations that we use in our analysis are obtained as follows. We note that the accumulation of any specific glycoform (mAb_{*i*}) in the culture can be given by (del Val et al. 2016):

$$\frac{d[\text{mAb}_i]}{dt} = f_i \times q_{\text{mAb}} \times X_v \quad 5.16$$

where f_i is the fractional concentration of that mAb glycoform obtained from the micro-scale model, q_{mAb} and X_v are the antibody productivity and cell viability obtained from the macro-scale model, respectively. The glycan measurements are related to the glycan fraction state variables as:

$$\mathbf{y}(t) = \frac{100}{[\text{MAb conc}(t)]} \mathbf{x}(t), \% \quad 5.17$$

In the present analysis, we generate our model predictions using the aforementioned multi-scale model. However, lacking the experimental measurements to compare these model results to, we generate a set of simulated measurements using a modified version of the multi-scale model. This is accomplished by changing the concentrations of the glycosyltransferase enzymes in the micro-scale model and obtaining a new glycan distribution profile. The relative glycan distribution is calculated at specific time points and chosen as the simulated ‘measurement’ for our system. The algorithm for the state estimator and the associated calculations are discussed in the following section.

5.4.3 Algorithm for Designing the State Estimator

As described in the previous section, we implement the state estimator by assuming that the macro-scale model is known to us *a priori*, enabling us to predict the glycan distribution profile based on antibody productivity.

Starting with an initial concentration \mathbf{x}_0 , we solve equation 5.16, to obtain the estimate of the concentration of individual glycoforms accumulated from time $t = t_0$ to time $t = t_1$, and using the model listed in equation 5.17, we obtain the relative glycan distribution ($\hat{\mathbf{y}}$) at time $t = t_1$. At the first time instant, the solver checks to see if any measurements ($\mathbf{y}(t_1)$) are available, and if so, this measurement is compared to the model prediction at that time instant ($\hat{\mathbf{y}}(t_1)$), giving us the error, ϵ as

$$\epsilon|_{t=t_1} = \mathbf{y}(t_1) - \hat{\mathbf{y}}(t_1) \quad 5.18$$

We multiply this error by a correction gain matrix (\mathbf{K}) as given in equation 5.14 and add the innovation term to our prediction at that time instant. The correction gain matrix is generally specified according to the type of estimator used. For non-linear systems, a commonly employed state estimator is the extended Kalman filter (EKF), wherein the non-linear state equations are linearized around the current state estimate and an appropriate correction term (\mathbf{K}_{EKF}) is calculated using the process noise covariance matrix, the measurement noise covariance matrix, and the Jacobian of the process model (Fernandes et al. 2015). By contrast, the unscented Kalman filter (UKF) avoids linearization approach used in EKF and instead uses a Gaussian random variable to represent the state distribution, and calculates the Kalman correction gain using the covariance matrices. In our system, the high degree of non-linearity associated with the glycan reaction expressions makes it impractical to adopt

linearization techniques or evaluate the appropriate Jacobian of the process model. Instead, we use a constant, time-invariant correction gain matrix in all our simulations and evaluate the performance of the state estimator under these conditions.

The correction gain matrix is multiplied to the error determined from equation 18, to obtain the correction (or innovation) term which is then used to update the model as shown in equation 5.14. We then solve equation 5.16 for our state variables (glycoform concentration) from time $t = t_1$ to the next time instant, $t = t_2$ using the updated state estimate at time $t = t_1$. At the next time instant, $t = t_2$, we evaluate once again the availability of any measurements. If no measurements are available at time $t = t_2$, equation xx is solved till the next time instant $t = t_3$, without any correction to the model prediction at time $t = t_2$. This loop is iteratively performed till the end of the batch. Figure 5.8 shows a flowchart that briefly summarizes the algorithm implemented here for state estimation

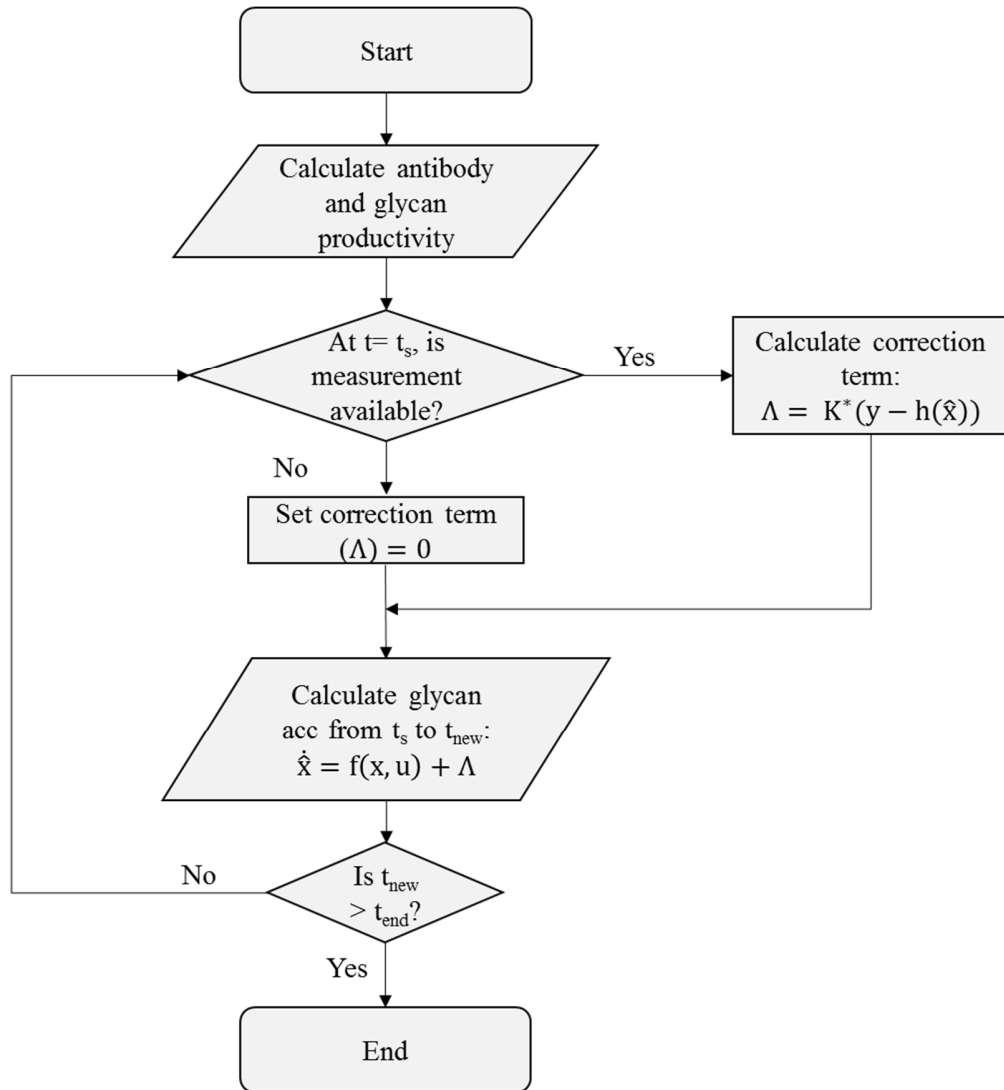


Figure 5.8: Flowchart depicting state estimation algorithm. At specific time instants, the availability of measurements is checked. If no new measurements are available, the state estimator uses the model to predict the glycan distribution. When a measurement becomes available, the model is updated and a new prediction is obtained. The algorithm is performed iteratively till the end of the batch.

5.4.4 Estimating Glycan States using a Constant, Time-invariant State Estimator

In this section, we will discuss the key findings obtained by implementing the designed state estimation technique. First, we evaluate the state estimator performance under nominal, noise-free conditions where, we assume that neither the model nor the measurement has any noise associated with it. To do this, we generate noise-free measurements using the multi-scale model at time points, $t = 58$ hours, 98 hours, 138 hours, and 178 hours. These time points are chosen assuming that the initial glycan measurement is made after about two and a half days after the cells were inoculated in the flask, followed by glycan measurements every 40 hours. Next, we follow the state estimation algorithm and generate model predictions from time $t = 0$ to time $t = t_f = 58$ hours. At 58 hours, the innovation (correction) term in the state estimator is activated and the model prediction gets updated, followed by subsequent integration of the state model till the time point at which the next measurement is available, i.e. 98 hours. Once again, at 98 hours, the measurement is compared to the estimate of the measured value as generated by the model predictions, and the correction term is activated accordingly. This process is repeated for time points 138 hours and 178 hours and the final glycan distribution profile predicted by the model is recorded. In these simulations, the correction matrix is defined as a diagonal matrix with a constant, time-invariant gain of 10^{-4} .

Additionally, to compare the efficacy of our state estimator, we obtain model predictions for the complete time period without including the state estimation technique, i.e. model predictions are not corrected at the specific time points at which measurements are available. Figure 5.9 plots the estimated values of the accumulated

glycan fractions for three glycan species A2, FA2, and M6 and compares them to the available measurements.

We note that in each case, the estimates obtained from the model prediction do not match the available measurements and hence, the model predictions must be corrected appropriately. When the first measurement becomes available at time $t = 58$ hours, we note that the activation of the correction matrix improves the subsequent estimate of the glycan distribution, and with each additional measurement, the final estimate is closer to the measured value of the glycan state. Figure 5.9(d) compares the sum of squared errors (SSE) between the estimated glycan distribution and the measured glycan distribution at each time point. Once again, we note that the sum of squared error between the estimated values and the measured values decreases when a state estimator is implemented.

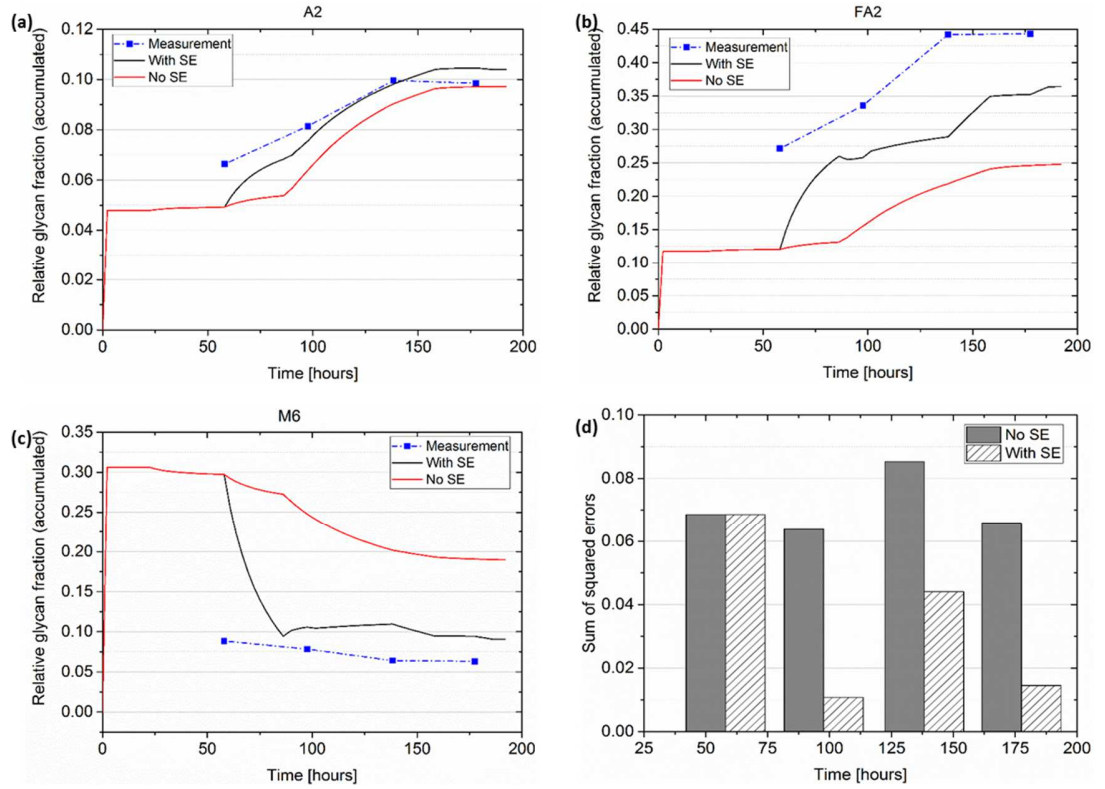


Figure 5.9: State estimator performance under noise-free conditions. Plots represent the accumulation of (a) A2; (b) FA2; and (c) M6. Solid red line represents the estimate without state estimation, the solid black line represents the estimate obtained by implementing the state estimation technique, while the blue dashed line with boxes represents the available measurements. (d) Sum of squared errors for estimates obtained with and without state estimation. The solid grey bars represent the SSE for estimates obtained without implementing the state estimation technique while the cross-hatched bars represent the SSE for estimates obtained after implementing state estimation

Next, we test the state estimator performance under conditions where we account for model and measurement noise. Here, we add a uniformly distributed random number, generated using the inbuilt '*rand*' function in MATLAB, to each measurement value and to each model prediction. However, the measurement

variables represent relative glycan fractions and are hence constrained to sum up to 1. Therefore, upon adding a random number to each measurement and model value, we must then renormalize the measurement and model values so that the inherent constraint (that the sum of glycan fractions must add up to 1) is satisfied.

Figure 5.10 compares the estimated value of the accumulated glycan fraction for glycoforms A2, FA2, and M6 to their respective measured values. Once again, we note that the estimates obtained by implementing the state estimator are closer to the actual measurement than the estimates obtained without implementing the state estimator. The sum of squared errors of the estimates obtained using the state estimator are also lesser than the sum of squared errors of estimates obtained without implementing the state estimator.

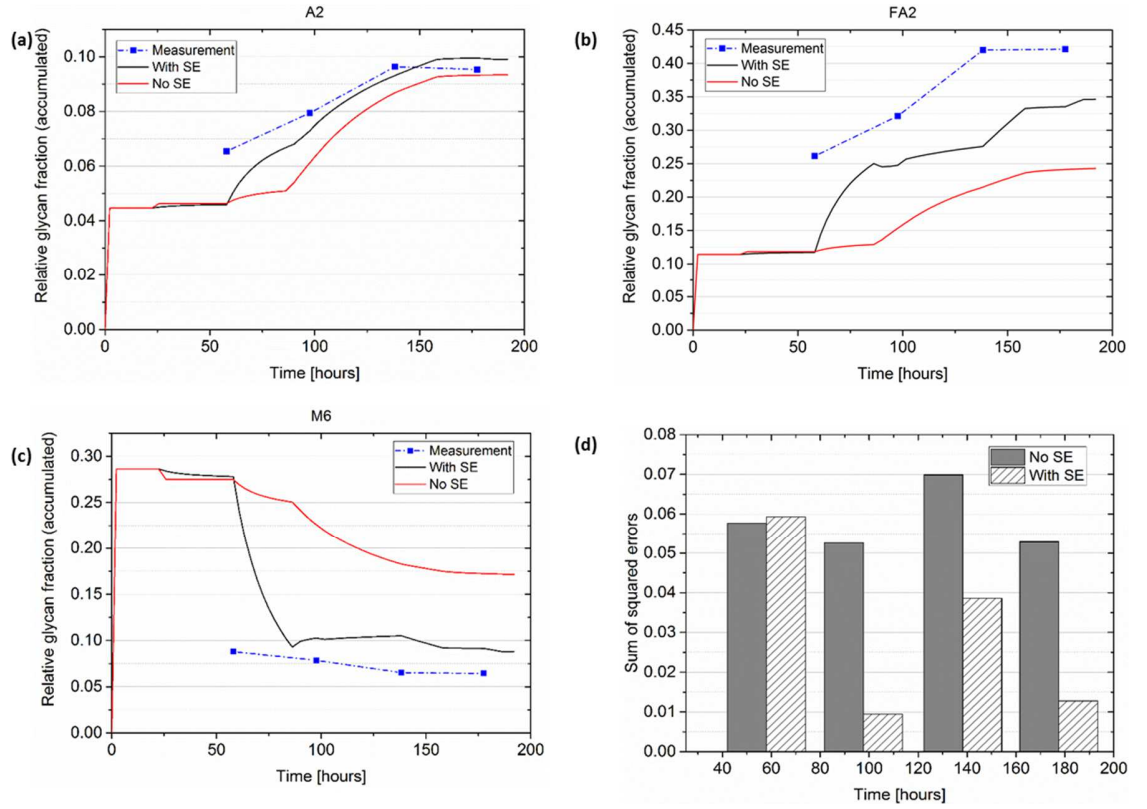


Figure 5.10: State estimator performance in the presence of model and measurement noise. As before, the plots represent the accumulation of (a) A2; (b) FA2; and (c) M6. Solid red line represents the estimate without state estimation, the solid black line represents the estimate obtained by implementing the state estimation technique, while the blue dashed line with boxes represents the available measurements. (d) Sum of squared errors for estimates obtained with and without state estimation. The solid grey bars represent the SSE for estimates obtained without implementing the state estimation technique while the cross-hatched bars represent the SSE for estimates obtained after implementing state estimation

In the analysis performed here, we have worked with a constant, time-invariant gain in the correction matrix. Further refinement of the state estimator can be carried out by using a varying correction gain matrix.

5.5 Summary and Conclusions

Glycan distribution is heterogeneous and diverse resulting in a wide array of oligosaccharide attachments that are naturally observed. The attachment of these diverse oligosaccharides to different glycoproteins alters their form, functionality, and structure and is hence critical to be monitored. While glycan analysis has been used primarily for off-line quality control to ensure batch-to-batch consistency and comparability, the adoption of quality by design (QbD) principles by manufacturers and regulatory agencies, and the emergence of biosimilars is increasing the need for at-line and on-line glycan characterization assays. However, current glycan characterization assays are laborious, expensive, and time-consuming and hence, most glycan assays are still performed off-line. The lack of on-line or real-time measurements further hinders the development of control schemes that can be developed to ensure on-line glycosylation control. To address the challenges associated with the lack of on-line measurements we have presented two techniques in this chapter. First, we developed a rational method based on observability analysis to identify glycan groups that can provide the maximum information about the glycan distribution profile, based solely on the stoichiometric information of the glycan reaction network. Next, we have demonstrated the efficacy of a state estimator scheme in estimating the glycan distribution profile in the absence of real time measurements. This work will form the foundation of future efforts to develop on-line control schemes for controlling glycosylation in mAbs.

Chapter 6

SUMMARY, CONCLUSIONS, AND FUTURE WORK

6.1 Summary and Conclusions

The work presented in this dissertation provides a robust framework: (a) for modeling the effect of different input factors on the glycosylation profile and obtaining quantitative input-output relationships; (b) for estimating the glycan distribution profile in the absence of real time glycan measurements; and (c) for controlling the final glycosylation profile based on an understanding of the underlying structural input-output relationship. The developed framework forms the basis of a rational approach to implement process analytical technology (PAT) in upstream cell culture operations to ensure glycosylation control.

As glycosylation is affected by a variety of factors at different system scales, we first developed a multi-scale model of glycosylation linking a macro-scale cell culture model to a micro-scale, kinetic model of glycosylation. The model predictions were compared to the experimental data obtained from in-house shake flask studies and the model parameters were optimized accordingly. The model was then validated under fed-batch conditions and found to be fairly representative of the growth dynamics under both batch and fed-batch conditions. The antibody productivity rate from the macro-scale model was used to obtain the glycan productivity rate which was used, subsequently, to predict the dynamic glycan distribution profile in the system. This multi-scale model serves as a quantitative link between cell growth conditions in

the bioreactor and the observed glycosylation profile and can predict the glycan distribution under different cell growth conditions.

However, one of the limitations of the developed multi-scale model is that the model accounts for the effects of cell culture nutrients on the glycan distribution but does not take into account the changes in the glycan distribution arising due to variations in cell culture media. Media composition is known to affect cell growth and product quality, but the composition of most commercially available cell culture media is proprietary information and is, therefore, unknown to the end user. Consequently, there exist few models that can capture the underlying mechanisms relating the vast majority of media components and the glycan distribution. In order to rationally quantify the effect of different media components on the glycosylation profile, we used a holistic approach combining factorial design of experiments and mathematical analysis as demonstrated in Chapter 3, wherein we evaluated the effect of dynamically introducing MnCl_2 and EDTA on the final glycan distribution using controllability analysis.

While the effect of MnCl_2 as a media supplement has been well studied, our work is a novel consideration into the effect of introducing multiple media supplements at different stages of cell culture. Specifically, we demonstrated through mixed-level factorial experiments that adding MnCl_2 early during cell culture has a greater impact on the glycan distribution profile than a late stage addition. Further, we showed that the changes in the glycosylation profile due to the addition of MnCl_2 to the media are not immutable and can be reversed by the judicious addition of EDTA. Specifically, the addition of EDTA after MnCl_2 has been added to the cell culture was shown to reverse the changes in the glycan distribution due to the addition of MnCl_2 .

However, the time of addition of EDTA also influenced cell growth and productivity, with early addition of EDTA having an adverse effect while late stage addition of EDTA enhanced the final antibody titer.

Next, by performing controllability analysis, we identified the combinations of inputs factors which, when manipulated, result in quantifiable changes in the relative percentage of specific glycan species. Our analyses showed that the most controllable glycan species in our experimental system were A2, FA2G1 and FA2 whose concentrations were affected by early stage supplementation of EDTA and late stage supplementation of MnCl_2 . Thus, we successfully demonstrated that the glycan distribution is affected not just by the concentration of the media supplement, but also by the time of introduction of the media supplement. While conventional strategies for media development include media preparation prior to the start of the batch, the experimental and computational approaches demonstrated in our work present a holistic approach to controlling the glycosylation profile using time-dependent media supplementation.

Controllability analysis allows us to identify *which* combination of input factors influence specific glycan species and by *how much*. Thus, by implementing controllability analysis, we generate a quantitative input-output relationship relating the multiple inputs that are available to the multiple outputs in the system. In Chapter 4, we used this structural relationship to design glycosylation controllers for two cases of practical significance: (i) where the input factors chosen were the enzyme concentrations in the micro-scale model developed in Chapter 2 and the changes in the glycosylation model were observed via simulations; and (ii) where the input factors were amino acid supplements added to cell culture media and the changes in the

glycosylation pattern were evaluated experimentally. In both cases we designed proportional and proportional integral controllers and evaluated the performance of the controllers under set-point tracking conditions. We observed that the controllers were able to effectively track the new set-point under both nominal conditions (i.e. when we assumed that the input-output model generated by our controllability analysis is a perfect representation of the ‘true’ process or the plant) and under the more realistic model-plant mismatch conditions. Significantly, we observed that by designing our controllers using select singular values obtained from controllability analysis, controller robustness is maintained in the face of significant model-plant mismatch. The glycosylation controllers developed in this work can form the basis of future cell engineering and media design efforts.

However, it must be noted that the glycosylation controllers designed on the basis of controllability analysis are batch-to-batch controllers, with the control action performed on every successive batch to ensure consistent glycan distribution. To achieve real time control of glycosylation, it is necessary to measure the glycosylation profile continuously or at reasonably frequent intervals compared to the time-scale of the process. As noted in our brief review in Chapter 5, current assays for glycosylation are performed offline, at irregular or less frequent intervals and are associated with long measurement delays. Consequently, in the absence of any real-time measurement of the glycan distribution there is a need to develop techniques to infer the glycan profile based on a process understanding of the system. We have addressed this challenge using two approaches. First, using observability analysis and by exploiting the underlying connectivity inherent in glycan reaction networks, we have identified a particular subset of glycan measurements that when measured together would give us

ample information about the complete glycoform distribution. Such an approach can lead to the rational development of shorter or simpler glycan assays that do not require measuring the entire glycosylation profile. Next, we have used the multi-scale model presented in Chapter 2 in a state estimation scheme to obtain estimates of the glycan distribution profile over the course of a batch run and compared these to a set of simulated ‘measurements’. As and when the measurements were available, the predicted estimates would be corrected and a new estimate would be obtained. In our simulations, we tested a constant, time-invariant correction gain matrix and the resulting estimates were closer to the measurement than the values obtained without state estimation. The state estimation performance remained robust even when evaluated against measurement noise. To our knowledge, this is the first such design of a state estimation scheme to achieve real-time estimates of the glycosylation profile. The state estimator designed here will be an integral component in the development of an online glycosylation controller.

In the following section, we briefly consider some avenues of research that can be considered to extend the work presented in this dissertation.

6.2 Future Work

6.2.1 Expanding the Scope of the Multi-scale Model

In this dissertation, we have established that the multi-scale modeling approach presented here is critical to our understanding of glycosylation and for obtaining reasonable estimates of the different glycan states. Therefore, enhancing the fidelity of the model will be vital to the development of robust controllers. Some of the approaches to extend the scope of the multi-scale model are listed below:

Mechanistic model development: In order to enhance the utility of a mathematical model, it is necessary to ensure that the model is valid and has acceptable predictive capabilities over a wide range of input conditions beyond those in the model identification set. To improve model reliability, it is necessary to perform informed experiments that can be used to refine the model systematically (Kontoravdi et al. 2010; Lencastre Fernandes et al. 2013). For biological systems, the resulting mathematical models typically contain several model parameters whose values are either based on historical manufacturing data, obtained from literature, or they are fit to suit the available experimental data. However, such an approach limits the utility of the model and hence, there is a need for systematic estimation of the model parameters, as shown in Figure 6.1

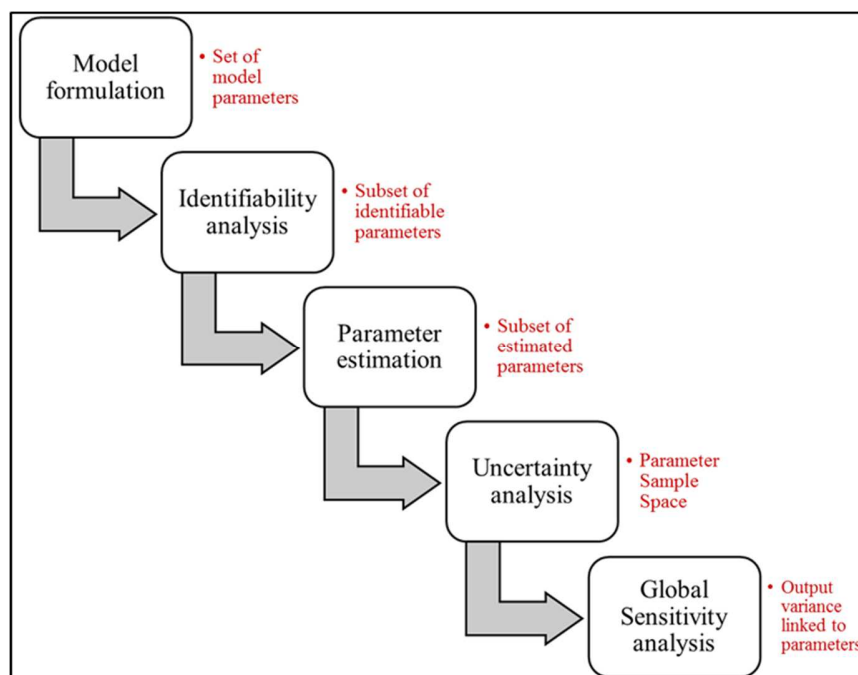


Figure 6.1: Schematic representation of systematic model development. Adapted from Lencastre Fernandes et al (2013).

Given that the number of model parameters in most mathematical models of biological systems far exceeds the available output or experimental measurements (for instance, the macro-scale model developed in this dissertation consists of 21 model parameters and 6 output measurements), it is important to understand which parameters are 'identifiable', i.e. the parameters whose values can be estimated with a high degree of certainty. Of the subset of identifiable parameters, only some are estimable using the available data. By performing global sensitivity analysis, the dependence of the output variance on the model parameters can be assessed. Such a systematic approach is necessary to increase model fidelity and robustness.

Incorporating the effect of multiple input factors: In this dissertation, cell growth and productivity were modeled considering substrate consumption and by-product inhibition rates. However, cell growth is also affected by the addition of different media supplements such as amino acids, temperature shifts, pH and shear rate. Each of these factors is also known to influence the glycosylation profile and hence an extended multi-scale model would incorporate the effect of each of these input factors at multiple scales.

Bioreactor pH is one such input factor that affects cell growth, productivity, as well as glycan distribution. During manufacturing, cellular metabolism and by-product formation results in a drift in bioreactor pH, which is controlled within a specific range to maintain cell growth at optimal levels by using complex buffer and base additions (Gramer and Ogorzalek 2007). By generating empirical or mechanistic models linking pH measurements at the cellular level to the intracellular pH (Wu et al. 1993) and then extending these model to assess organelle level pH changes, it would be possible to estimate how bioreactor pH can influence the pH in the Golgi, changes in which are

known to affect the glycosylation profiles and glycosyltransferase enzyme localization (Hassinen et al. 2011; Rivinoja et al. 2009; Rivinoja et al. 2012). Similar models can be built to relate other inputs such as temperature, dissolved oxygen content, shear rate, at the macro-scale to changes at the micro-scale, thereby providing us with manipulated variables at the macro-scale that can be used to fine-tune the glycosylation profile.

In the absence of extensive mechanistic models, controllability analysis can be used to generate quantitative input-output relationships between macro-scale factors and the glycan distribution profile. Ideally, the use of micro-bioreactors should be explored to perform the necessary factorial experiments under high throughput conditions. Further, if intermittent glycan measurements are available, the current steady state controllability analysis can be extended to quantify the effect of input parameters on the output glycan distribution as a function of time. By evaluating the total change in the glycosylation profile at a given point of time as well as the fractional difference in the glycan distribution profile between measurements, a dynamic gain matrix can be generated and mathematically analyzed to identify how different combinations of glycans can be controlled at different stages of cell culture.

6.2.2 Multi-attribute Analytics and Control

Protein therapeutics have multiple quality attributes that need to be monitored both in-process and offline. The increasing number of analytical tools employed for protein characterization have added to the cost and complexity of drug testing, necessitating the development of simpler, multi-attribute analytical methods (Rogers et al. 2015). However, existing analytical methods are laborious and expensive, with fewer options for in-process monitoring. To reduce the cost, complexity and analysis

time, a representative set of attributes that contain information about other quality attributes can be analyzed. The observability analysis technique developed in Chapter 5 can be used in determining if the choice of measurements would yield adequate information about the entire quality profile of the mAb.

For instance, thermal unfolding experiments – where spectroscopic techniques are used to determine changes in the conformation of a protein heated at a constant rate – are typically carried out to test the thermodynamic stability of proteins. Recent research has indicated (Zheng et al. 2014) that the thermal unfolding of antibodies differs based on the abundance of oligomannose groups attached to the constant heavy chain (CH2) domain (Fig. 6.2). Conversely, we can now examine changes in the thermal unfolding of the CH2 domain as a potential surrogate for mannose measurement. Such an analysis can lead to the design of rational multi-attribute assays.

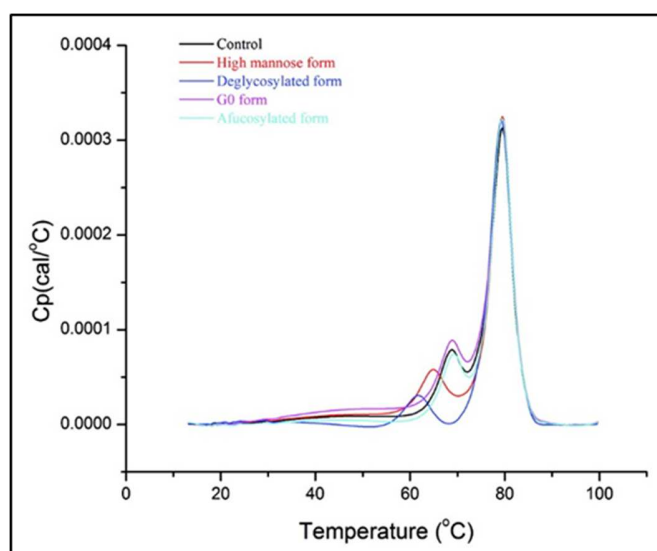


Figure 6.2: Thermal unfolding from DSC measurements for different glycan isoforms. Adapted from Zheng et al (2014)

Further, it will be interesting to consider how these multiple quality attributes can be controlled using common inputs. For instance, heterogeneity that arises from modifications to the antibody such as the loss of a lysine residue on the C-terminal of the antibody (C-terminal heterogeneity) and acidic and basic variants (charge variants), affects protein stability and binding. As these changes depend on the amino acid profile in the cell culture, we could evaluate if the addition of specific amino acid supplements to the media alter multiple quality attributes such as C-terminal heterogeneity and charge variants in the mAb. If so, how do we develop a framework for achieving *multiple objectives* of controlling charge and sequence variants using the *same set of inputs*? The answers to these questions will be critical in the manufacturing of generic biologics (biosimilars) that are expected to match every quality attribute of the innovator drug.

6.2.3 Validating the State-estimator and Implementing On-line Glycosylation Control

The state estimator designed in this dissertation is integral to the development of an on-line glycosylation control scheme. Presently, the state estimation has been designed using the multi-scale model. However, it will be essential to validate the state estimation with actual, real-time measurements taken in-process. The in-house glycan assay described in Chapter 5 will be useful in making intermittent measurements of the glycan distribution.

Having evaluated the validity of the state estimator, we can then implement the state-estimation scheme in an on-line glycosylation control scheme (St Amand et al. 2011). Figure 6.3 shows the proposed hierarchical multi-loop control scheme where the inner loop controllers maintain the bioreactor at defined in-process set-points and

the outer-loop controller takes control action to achieve the desired glycan set-point. The state estimator is used to predict the glycosylation profile on the basis of the multi-scale model and in-process measurements and is updated when the glycosylation measurement becomes available.

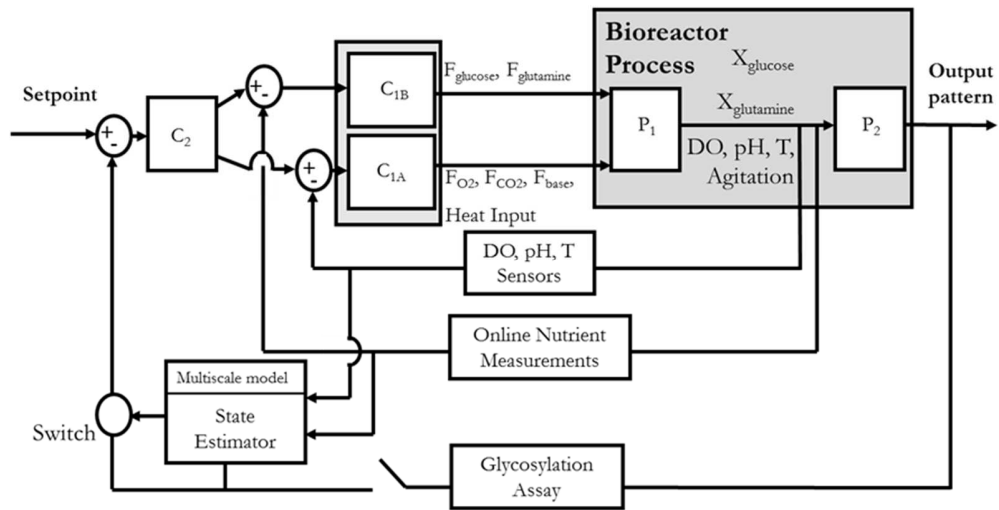


Figure 6.3: Strategy for on-line control of glycosylation. Adapted from St.Amand et al (2011)

The development and experimental validation of such an online glycosylation control scheme will establish a new basis for on-line quality control of biopharmaceuticals and will ensure consistent product quality during manufacturing.

REFERENCES

- Adamczyk B, Tharmalingam-Jaikaran T, Schomberg M, Szekrenyes A, Kelly RM, Karlsson NG, Guttman A, Rudd PM. 2014. Comparison of separation techniques for the elucidation of IgG N-glycans pooled from healthy mammalian species. *Carbohydrate Research* 389:174-185.
- Adams GP, Weiner LM. 2005. Monoclonal antibody therapy of cancer. *Nature Biotechnology* 23(9):1147-1157.
- Aggarwal S. 2007. What's fueling the biotech engine? *Nature Biotechnology* 25(10):1097-1104.
- Aggarwal S. 2014. What's fueling the biotech engine-2012 to 2013. *Nature Biotechnology* 32(1):32-39.
- Aghamohseni H, Ohadi K, Spearman M, Krahn N, Moo-Young M, Scharer JM, Butler M, Budman HM. 2014. Effects of nutrient levels and average culture pH on the glycosylation pattern of camelid-humanized monoclonal antibody. *Journal of Biotechnology* 186:98-109.
- Ahn WS, Jeon JJ, Jeong YR, Lee SJ, Yoon SK. 2008. Effect of Culture Temperature on Erythropoietin Production and Glycosylation in a Perfusion Culture of Recombinant CHO Cells. *Biotechnology and Bioengineering* 101(6):1234-1244.
- Aich U, Lakbub J, Liu A. 2016. State-of-the-art technologies for rapid and high-throughput sample preparation and analysis of N-glycans from antibodies. *Electrophoresis* 37(11):1468-1488.
- Alessandri L, Ouellette D, Acquah A, Rieser M, LeBlond D, Saltarelli M, Radziejewski C, Fujimori T, Correia I. 2012. Increased serum clearance of oligomannose species present on a human IgG1 molecule. *Mabs* 4(4):509-520.
- Altamirano C, Illanes A, Becerra S, Cairo JJ, Godia F. 2006. Considerations on the lactate consumption by CHO cells in the presence of galactose. *Journal of Biotechnology* 125(4):547-556.
- Araújo FC, Gonçalves J, Fonseca JE. 2016. Pharmacoeconomics of Biosimilars: What Is There to Gain from Them? *Current Rheumatology Reports* 18(8):1-9.
- Bailey JE. 1998. Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities. *Biotechnology Progress* 14(1):8-20.
- Bailey MJ, Hooker AD, Adams CS, Zhang SH, James DC. 2005. A platform for high-throughput molecular characterization of recombinant monoclonal antibodies. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 826(1-2):177-187.

- Baughman AC, Huang XQ, Sharfstein ST, Martin LL. 2010. On the dynamic modeling of mammalian cell metabolism and mAb production. *Computers & Chemical Engineering* 34(2):210-222.
- Beck A, Wagner-Rousset E, Bussat MC, Lokteff M, Klinguer-Hamour C, Haeuw JF, Goetsch L, Wurch T, Van Dorsselaer A, Corvaia N. 2008. Trends in Glycosylation, Glycoanalysis and Glycoengineering of Therapeutic Antibodies and Fc-Fusion Proteins. *Current Pharmaceutical Biotechnology* 9(6):482-501.
- Beck A, Wurch T, Bailly C, Corvaia N. 2010. Strategies and challenges for the next generation of therapeutic antibodies. *Nature Reviews Immunology* 10(5):345-352.
- Bennun SV, Yarema KJ, Betenbaugh MJ, Krambeck FJ. 2013. Integration of the Transcriptome and Glycome for Identification of Glycan Cell Signatures. *Plos Computational Biology* 9(1).
- Berger M, Kaup M, Blanchard V. 2012. Protein Glycosylation and Its Impact on Biotechnology. In: Hu WS, Zeng A-P, editors. *Genomics and Systems Biology of Mammalian Cell Culture*: Springer Berlin Heidelberg. p 165-185.
- Bibila T, Flickinger MC. 1991. A structured model for monoclonal antibody synthesis in exponentially growing and stationary phase hybridoma cells. *Biotechnology and Bioengineering* 37(3):210-226.
- Bibila TA, Flickinger MC. 1993. Use Of A Structured Kinetic-Model Of Antibody-Synthesis And Secretion For Optimization Of Antibody-Production Systems .1. Steady-State Analysis (*Biotechnology And Bioengineering*, Vol 39, PG 251, 1992). *Biotechnology and Bioengineering* 41(6):682-682.
- Blondeel EJM, Braasch K, McGill T, Chang D, Engel C, Spearman M, Butler M, Aucoin MG. 2015. Tuning a MAb glycan profile in cell culture: Supplementing N-acetylglucosamine to favour G0 glycans without compromising productivity and cell growth. *Journal of Biotechnology* 214:105-112.
- Blondeel EJM, Ho R, Schulze S, Sokolenko S, Guillemette SR, Slivac I, Durocher Y, Guillemette JG, McConkey BJ, Chang D and others. 2016. An omics approach to rational feed: Enhancing growth in CHO cultures with NMR metabolomics and 2D-DIGE proteomics. *Journal of Biotechnology* 234:127-138.
- Borys MC, Dalal NG, Abu-Absi NR, Khattak SF, Jing Y, Xing ZZ, Li ZJ. 2010. Effects of Culture Conditions on N-Glycolylneuraminic Acid (Neu5Gc) Content of a Recombinant Fusion Protein Produced in CHO Cells. *Biotechnology and Bioengineering* 105(6):1048-1057.
- Borys MC, Linzer DIH, Papoutsakis ET. 1994. Ammonia affects the glycosylation patterns of recombinant mouse placental lactogen-I by chinese hamster ovary cells in a pH-dependent manner. *Biotechnology and Bioengineering* 43(6):505-514.
- Brewster VL, Ashton L, Goodacre R. 2011. Monitoring the Glycosylation Status of Proteins Using Raman Spectroscopy. *Analytical Chemistry* 83(15):6074-6081.

- Brühlmann D, Jordan M, Hemberger J, Sauer M, Stettler M, Broly H. 2015. Tailoring recombinant protein quality by rational media design. *Biotechnology Progress* 31(3):615-629.
- Burnina I, Hoyt E, Lynaugh H, Li HJ, Gong B. 2013. A cost-effective plate-based sample preparation for antibody N-glycan analysis. *Journal of Chromatography A* 1307:201-206.
- Butler M. 2006. Optimisation of the cellular metabolism of glycosylation for recombinant proteins produced by mammalian cell systems. *Cytotechnology* 50(1-3):57-76.
- Charaniya S, Le H, Rangwala H, Mills K, Johnson K, Karypis G, Hu W-S. 2010. Mining manufacturing data for discovery of high productivity process characteristics. *Journal of Biotechnology* 147(3-4):186-197.
- Chen F, Kou TC, Fan L, Zhou Y, Ye ZY, Zhao L, Tan WS. 2011. The combined effect of sodium butyrate and low culture temperature on the production, sialylation, and biological activity of an antibody produced in CHO cells. *Biotechnology and Bioengineering* 16(6):1157-1165.
- Chen PF, Harcum SW. 2006. Effects of elevated ammonium on glycosylation gene expression in CHO cells. *Metabolic Engineering* 8(2):123-132.
- Ciucanu I, Costello CE. 2003. Elimination of Oxidative Degradation during the per-O-Methylation of Carbohydrates. *Journal of the American Chemical Society* 125(52):16213-16219.
- Ciucanu I, Kerek F. 1984. A simple and rapid method for the permethylation of carbohydrates. *Carbohydrate Research* 131(2):209-217.
- Clark KJR, Chaplin FWR, Harcum SW. 2004. Temperature effects on product-quality-related enzymes in batch CHO cell cultures producing recombinant tPA. *Biotechnology Progress* 20(6):1888-1892.
- Clincke MF, Guedon E, Yen FT, Ogier V, Roitel O, Goergen JL. 2011. Effect of Surfactant Pluronic F-68 on CHO Cell Growth, Metabolism, Production, and Glycosylation of Human Recombinant IFN-gamma in Mild Operating Conditions. *Biotechnology Progress* 27(1):181-190.
- Cumming DA. 2003. Pathways and functions of mammalian protein glycosylation. *Gene Transfer and Expression in Mammalian Cells* 38:433-455.
- Curling EMA, Hayter PM, Baines AJ, Bull AT, Gull K, Strange PG, Jenkins N. 1990. Recombinant Human Interferon-Gamma - Differences In Glycosylation And Proteolytic Processing Lead To Heterogeneity In Batch Culture. *Biochemical Journal* 272(2):333-337.
- Dall'Olio F, Vanhooren V, Chen CC, Slagboom PE, Wuhler M, Franceschi C. 2013. N-glycomic biomarkers of biological aging and longevity: A link with inflammaging. *Ageing Research Reviews* 12(2):685-698.
- del Val IJ, Fan Y, Weilguny D. 2016. Dynamics of immature mAb glycoform secretion during CHO cell culture: An integrated modelling framework. *Biotechnology Journal* 11(5):610-623.

- del Val IJ, Jedrzejewski PM, Exley K, Sou SN, Kyriakopoulos S, Polizzi KM, Kontoravdi C. 2012. Application of Quality by Design Paradigm to the Manufacture of Protein Therapeutics. *Glycosylation*:353-396.
- del Val IJ, Kontoravdi C, Nagy JM. 2010. Towards the Implementation of Quality by Design to the Production of Therapeutic Monoclonal Antibodies with Desired Glycosylation Patterns. *Biotechnology Progress* 26(6):1505-1527.
- del Val IJ, Nagy JM, Kontoravdi C. 2011. A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. *Biotechnology Progress* 27(6):1730-1743.
- Dewasme L, Fernandes S, Amribt Z, Santos LO, Bogaerts P, Vande Wouwer A. 2015. State estimation and predictive control of fed-batch cultures of hybridoma cells. *Journal of Process Control* 30:50-57.
- Dochain D. 2003. State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of Process Control* 13(8):801-818.
- Doherty M, Bones J, McLoughlin N, Telford JE, Harmon B, DeFelippis MR, Rudd PM. 2013. An automated robotic platform for rapid profiling oligosaccharide analysis of monoclonal antibodies directly from cell culture. *Analytical Biochemistry* 442(1):10-18.
- Ecker DM, Jones SD, Levine HL. 2015. The therapeutic monoclonal antibody market. *mAbs* 7(1):9-14.
- Elvin JG, Couston RG, van der Walle CF. 2013. Therapeutic antibodies: Market considerations, disease targets and bioprocessing. *International Journal of Pharmaceutics* 440(1):83-98.
- Fan Y, Del Val IJ, Muller C, Sen JW, Rasmussen SK, Kontoravdi C, Weilguny D, Andersen MR. 2015a. Amino Acid and Glucose Metabolism in Fed-Batch CHO Cell Culture Affects Antibody Production and Glycosylation. *Biotechnology and Bioengineering* 112(3):521-535.
- Fan YZ, Del Val IJ, Muller C, Lund AM, Sen JW, Rasmussen SK, Kontoravdi C, Baycin-Hizal D, Betenbaugh MJ, Weilguny D and others. 2015b. A multi-pronged investigation into the effect of glucose starvation and culture duration on fed-batch CHO cell culture. *Biotechnology and Bioengineering* 112(10):2172-2184.
- FDA US. 2004. PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. In: Research CfDEa, editor.
- FDA US. 2006. Guidance for industry; Q8 pharmaceutical development. US FDA, Rockville, MD, USA.
- Fernandes S, Richelle A, Amribt Z, Dewasme L, Bogaerts P, Vande Wouwer A. 2015. Extended and Unscented Kalman Filter design for hybridoma cell fed-batch and continuous cultures. *Ifac Papersonline* 48(8):1108-1113.
- Fuller SD, Bravo R, Simons K. 1985. An enzymatic assay reveals that proteins destined for the apical or basolateral domains of an epithelial cell line share the same late Golgi compartments. *The EMBO Journal* 4(2):297-307.

- Gagnon M, Hiller G, Luan YT, Kittredge A, DeFelice J, Drapeau D. 2011. High-End pH-Controlled Delivery of Glucose Effectively Suppresses Lactate Accumulation in CHO Fed-Batch Cultures. *Biotechnology and Bioengineering* 108(6):1328-1337.
- Gawlitsek M, Estacio M, Furch T, Kiss R. 2009. Identification of Cell Culture Conditions to Control N-Glycosylation Site-Occupancy of Recombinant Glycoproteins Expressed in CHO Cells. *Biotechnology and Bioengineering* 103(6):1164-1175.
- Gawlitsek M, Ryll T, Lofgren J, Sliwkowski MB. 2000. Ammonium alters N-glycan structures of recombinant TNFR-IgG: Degradative versus biosynthetic mechanisms. *Biotechnology and Bioengineering* 68(6):637-646.
- Godoy-Silva R, Chalmers JJ, Casnocha SA, Bass LA, Ma NN. 2009. Physiological Responses of CHO Cells to Repetitive Hydrodynamic Stress. *Biotechnology and Bioengineering* 103(6):1103-1117.
- Goetze AM, Schenauer MR, Flynn GC. 2010. Assessing monoclonal antibody product quality attribute criticality through clinical studies. *mAbs* 2(5):500-507.
- Golabgir A, Hoch T, Zhariy M, Herwig C. 2015. Observability Analysis of Biochemical Process Models as a Valuable Tool for the Development of Mechanistic Soft Sensors. *Biotechnology Progress* 31(6):1703-1715.
- Gonzalez-Leal IJ, Carrillo-Cocom LM, Ramirez-Medrano A, Lopez-Pacheco F, Bulnes-Abundis D, Webb-Vargas Y, Alvarez MM. 2011. Use of a plackett-burman statistical design to determine the effect of selected amino acids on monoclonal antibody production in CHO cells. *Biotechnology Progress* 27(6):1709-1717.
- Goudar CT. 2012. Analyzing the dynamics of cell growth and protein production in mammalian cell fed-batch systems using logistic equations. *Journal of Industrial Microbiology & Biotechnology* 39(7):1061-1071.
- Grainger RK, James DC. 2013. CHO Cell Line Specific Prediction and Control of Recombinant Monoclonal Antibody N-Glycosylation. *Biotechnology and Bioengineering* 110(11):2970-2983.
- Gramer MJ. 2014. Product Quality Considerations for Mammalian Cell Culture Process Development and Manufacturing. *Mammalian Cell Cultures for Biologics Manufacturing* 139:123-166.
- Gramer MJ, Eckblad JJ, Donahue R, Brown J, Shultz C, Vickerman K, Priem P, van den Bremer ETJ, Gerritsen J, van Berkel PHC. 2011. Modulation of antibody galactosylation through feeding of uridine, manganese chloride, and galactose. *Biotechnology and Bioengineering* 108(7):1591-1602.
- Gramer MJ, Ogorzalek T. 2007. A semi-empirical mathematical model useful for describing the relationship between carbon dioxide, pH, lactate and base in a bicarbonate-buffered cell-culture process. *Biotechnology and Applied Biochemistry* 47:197-204.
- Grammatikos SI, Valley U, Nimtz M, Conradt HS, Wagner R. 1998. Intracellular UDP-N-acetylhexosamine pool affects N-glycan complexity: A mechanism of

- ammonium action on protein glycosylation. *Biotechnology Progress* 14(3):410-419.
- Gréco S, Niepceon E, Hugueny I, George P, Louisot P, Biol M-C. 2001. Dietary Spermidine and Spermine Participate in the Maturation of Galactosyltransferase Activity and Glycoprotein Galactosylation in Rat Small Intestine. *The Journal of Nutrition* 131(7):1890-1897.
- Gudi RD, Shah SL, Gray MR. 1995. Adaptive multirate state and parameter estimation strategies with application to a bioreactor. *AIChE Journal* 41(11):2451-2464.
- Ha S, Ou YS, Vlasak J, Li Y, Wang SY, Vo K, Du Y, Mach A, Fang YL, Zhang NY. 2011. Isolation and characterization of IgG1 with asymmetrical Fc glycosylation. *Glycobiology* 21(8):1087-1096.
- Ha TK, Lee GM. 2014. Effect of glutamine substitution by TCA cycle intermediates on the production and sialylation of Fc-fusion protein in Chinese hamster ovary cell culture. *Journal of Biotechnology* 180:23-29.
- Harmonised Tripartite Guideline ICH. 2009. Pharmaceutical Development Q8 (R2). Current step 4.
- Hassinen A, Pujol FM, Kokkonen N, Pieters C, Kihlstrom M, Korhonen K, Kellokumpu S. 2011. Functional Organization of Golgi N- and O-Glycosylation Pathways Involves pH-dependent Complex Formation That Is Impaired in Cancer Cells. *Journal of Biological Chemistry* 286(44):38329-38340.
- Ho Y, Varley J, Mantalaris A. 2006. Development and analysis of a mathematical model for antibody-producing GS-NS0 cells under normal and hyperosmotic culture conditions. *Biotechnology Progress* 22(6):1560-1569.
- Hong JK, Cho SM, Yoon SK. 2010. Substitution of glutamine by glutamate enhances production and galactosylation of recombinant IgG in Chinese hamster ovary cells. *Applied Microbiology and Biotechnology* 88(4):869-876.
- Hossler P, Khattak SF, Li ZJ. 2009. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 19(9):936-949.
- Hossler P, Mulukutla BC, Hu WS. 2007. Systems Analysis of N-Glycan Processing in Mammalian Cells. *Plos One* 2(8).
- Houde D, Peng YC, Berkowitz SA, Engen JR. 2010. Post-translational Modifications Differentially Affect IgG1 Conformation and Receptor Binding. *Molecular & Cellular Proteomics* 9(8):1716-1728.
- Huang CJ, Lin H, Yang J. 2015. A Robust Method for Increasing Fc Glycan High Mannose Level of Recombinant Antibodies. *Biotechnology and Bioengineering* 112(6):1200-1209.
- Imperiali B, O'Connor SE. 1999. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Current Opinion in Chemical Biology* 3(6):643-649.
- Ishii Y, Imamoto Y, Yamamoto R, Tsukahara M, Wakamatsu K. 2015. Titer of trastuzumab produced by a Chinese hamster ovary cell line is associated with

- tricarboxylic acid cycle activity rather than lactate metabolism. *Journal of Bioscience and Bioengineering* 119(4):478-485.
- Ivarsson M, Villiger TK, Morbidelli M, Soos M. 2014. Evaluating the impact of cell culture process parameters on monoclonal antibody N-glycosylation. *Journal of Biotechnology* 188:88-96.
- Jang DJ, Barford JP. 2000. An unstructured kinetic model of macromolecular metabolism in batch and fed-batch cultures of hybridoma cells producing monoclonal antibody. *Biochemical Engineering Journal* 4(2):153-168.
- Jayapal KR, Wlaschin KF, Hu WS, Yap MGS. 2007. Recombinant protein therapeutics from CHO cells - 20 years and counting. *Chemical Engineering Progress* 103(10):40-47.
- Jedrzejewski PM, del Val IJ, Constantinou A, Dell A, Haslam SM, Polizzi KM, Kontoravdi C. 2014. Towards Controlling the Glycoform: A Model Framework Linking Extracellular Metabolites to Antibody Glycosylation. *International Journal of Molecular Sciences* 15(3):4492-4522.
- Jefferis R. 2009. Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery* 8(3):226-234.
- Jenkins N, Parekh RB, James DC. 1996. Getting the glycosylation right: Implications for the biotechnology industry. *Nat Biotech* 14(8):975-981.
- Jordan M, Voisard D, Berthoud A, Tercier L, Kleuser B, Baer G, Broly H. 2013. Cell culture medium improvement by rigorous shuffling of components using media blending. *Cytotechnology* 65(1):31-40.
- Ju MS, Jung ST. 2014. Aglycosylated full-length IgG antibodies: steps toward next-generation immunotherapeutics. *Current Opinion in Biotechnology* 30:128-139.
- Jung ST, Kang TH, Kelton W, Georgiou G. 2011. Bypassing glycosylation: engineering aglycosylated full-length IgG antibodies for human therapy. *Current Opinion in Biotechnology* 22(6):858-867.
- Kamoda S, Ishikawa R, Kakehi K. 2006. Capillary electrophoresis with laser-induced fluorescence detection for detailed studies on N-linked oligosaccharide profile of therapeutic recombinant monoclonal antibodies. *Journal of Chromatography A* 1133(1-2):332-339.
- Kanda Y, Yamada T, Mori K, Okazaki A, Inoue M, Kitajima-Miyama K, Kuni-Kamochi R, Nakano R, Yano K, Kakita S and others. 2007. Comparison of biological activity among nonfucosylated therapeutic IgG1 antibodies with three different N-linked Fc oligosaccharides: the high-mannose, hybrid, and complex types. *Glycobiology* 17(1):104-118.
- Kang P, Mechref Y, Novotny MV. 2008. High-throughput solid-phase permethylation of glycans prior to mass spectrometry. *Rapid Communications in Mass Spectrometry* 22(5):721-734.
- Kang S, Zhang Z, Richardson J, Shah B, Gupta S, Huang C, Jr., Qiu J, Le N, Lin H, Bondarenko PV. 2015. Metabolic markers associated with high mannose

- glycan levels of therapeutic recombinant monoclonal antibodies. *Journal of Biotechnology* 203:22-31.
- Kao Y-H, Hewitt DP, Trexler-Schmidt M, Laird MW. 2010. Mechanism of antibody reduction in cell culture production processes. *Biotechnology and Bioengineering* 107(4):622-632.
- Karra S, Sager B, Karim MN. 2010. Multi-Scale Modeling of Heterogeneities in Mammalian Cell Culture Processes. *Industrial & Engineering Chemistry Research* 49(17):7990-8006.
- Karsten CM, Pandey MK, Figge J, Taylor PR, Berger M, Strait R, Nimmerjahn F, Brown GD, Finkelman F, Koehl J. 2012. Galactosylated IgG1 ICs control C5aR-mediated inflammation via a novel interaction between FcγRIIB and Dectin-1. *Immunology* 137:96-96.
- Kelley B. 2009. Industrialization of mAb production technology The bioprocessing industry at a crossroads. *Mabs* 1(5):443-452.
- Kesselheim AS, Avorn J, Sarpatwari A. 2016. The high cost of prescription drugs in the united states: Origins and prospects for reform. *JAMA* 316(8):858-871.
- Kildegaard HF, Fan Y, Sen JW, Larsen B, Andersen MR. 2016. Glycoprofiling effects of media additives on IgG produced by CHO cells in fed-batch bioreactors. *Biotechnology and Bioengineering* 113(2):359-366.
- Kim J, Kim Y-G, Lee G. 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied Microbiology and Biotechnology* 93(3):917-930.
- Kim PJ, Lee DY, Jeong H. 2009a. Centralized Modularity of N-Linked Glycosylation Pathways in Mammalian Cells. *Plos One* 4(10).
- Kim WH, Kim J-S, Yoon Y, Lee GM. 2009b. Effect of Ca²⁺ and Mg²⁺ concentration in culture medium on the activation of recombinant factor IX produced in Chinese hamster ovary cells. *Journal of Biotechnology* 142(3-4):275-278.
- Kishishita S, Katayama S, Kodaira K, Takagi Y, Matsuda H, Okamoto H, Takuma S, Hirashima C, Aoyagi H. 2015. Optimization of chemically defined feed media for monoclonal antibody production in Chinese hamster ovary cells. *Journal of Bioscience and Bioengineering* 120(1):78-84.
- Klapoetke S. 2014. N-Glycosylation Characterization by Liquid Chromatography with Mass Spectrometry. *Monoclonal Antibodies: Methods and Protocols*, 2nd Edition 1131:513-524.
- Konno Y, Kobayashi Y, Takahashi K, Takahashi E, Sakae S, Wakitani M, Yamano K, Suzawa T, Yano K, Ohta T and others. 2012. Fucose content of monoclonal antibodies can be controlled by culture medium osmolality for high antibody-dependent cellular cytotoxicity. *Cytotechnology* 64(3):249-265.
- Kontoravdi C, Asprey SP, Pistikopoulos EN, Mantalaris A. 2005. Application of global sensitivity analysis to determine goals for design of experiments: An example study on antibody-producing cell cultures. *Biotechnology Progress* 21(4):1128-1135.

- Kontoravdi C, Asprey SP, Pistikopoulos EN, Mantalaris A. 2007. Development of a dynamic model of monoclonal antibody production and glycosylation for product quality monitoring. *Computers & Chemical Engineering* 31(5-6):392-400.
- Kontoravdi C, Pistikopoulos EN, Mantalaris A. 2010. Systematic development of predictive mathematical models for animal cell cultures. *Computers & Chemical Engineering* 34(8):1192-1198.
- Kornfeld R, Kornfeld S. 1985. Assembly of Asparagine-Linked Oligosaccharides. *Annual Review of Biochemistry* 54(1):631-664.
- Krambeck FJ, Bennun SV, Narang S, Choi S, Yarema KJ, Betenbaugh MJ. 2009. A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. *Glycobiology* 19(11):1163-1175.
- Krambeck FJ, Betenbaugh MJ. 2005. A mathematical model of N-linked glycosylation. *Biotechnology and Bioengineering* 92(6):711-728.
- Kremkow B, Lee K. CHO-specific recombinant protein glycosylation reaction network. In: Robert Kiss G, Sarah Harcum CU, Jeff Chalmers OSU, editors. ECI Symposium Series; 2016; Palm Springs, CA.
- Kunert R, Reinhart D. 2016. Advances in recombinant antibody manufacturing. *Applied Microbiology and Biotechnology* 100(8):3451-3461.
- Kunkel JP, Jan DCH, Jamieson JC, Butler M. 1998. Dissolved oxygen concentration in serum-free continuous culture affects N-linked glycosylation of a monoclonal antibody. *Journal of Biotechnology* 62(1):55-71.
- Kyriakopoulos S, Kontoravdi C. 2014. A framework for the systematic design of fed-batch strategies in mammalian cell culture. *Biotechnology and Bioengineering* 111(12):2466-2476.
- Landauer K. 2014. Designing Media for Animal Cell Culture: CHO Cells, the Industrial Standard. *Animal Cell Biotechnology: Methods and Protocols*, 3rd Edition 1104:89-103.
- Lao MS, Toth D. 1997. Effects of ammonium and lactate on growth and metabolism of a recombinant Chinese hamster ovary cell culture. *Biotechnology Progress* 13(5):688-691.
- Le H, Kabbur S, Pollastrini L, Sun ZR, Mills K, Johnson K, Karypis G, Hu WS. 2012. Multivariate analysis of cell culture bioprocess data-Lactate consumption as process indicator. *Journal of Biotechnology* 162(2-3):210-223.
- Lencastre Fernandes R, Bodla VK, Carlquist M, Heins A-L, Eliasson Lantz A, Sin G, Gernaey KV. 2013. Applying Mechanistic Models in Bioprocess Development. In: Mandenius C-F, Titchener-Hooker JN, editors. *Measurement, Monitoring, Modelling and Control of Bioprocesses*. Berlin, Heidelberg: Springer Berlin Heidelberg. p 137-166.
- Li F, Vijayasankaran N, Shen A, Kiss R, Amanullah A. 2010. Cell culture processes for monoclonal antibody production. *Mabs* 2(5):466-479.

- Li JC, Wong CL, Vijayasankaran N, Hudson T, Amanullah A. 2012. Feeding lactate for CHO cell culture processes: Impact on culture metabolism and performance. *Biotechnology and Bioengineering* 109(5):1173-1186.
- Lin Z, Lubman DM. 2013. Permethylated N-Glycan Analysis with Mass Spectrometry. In: Matthiesen R, editor. *Mass Spectrometry Data Analysis in Proteomics*. Totowa, NJ: Humana Press. p 289-300.
- Liu B, Spearman M, Doering J, Lattova E, Perreault H, Butler M. 2014. The availability of glucose to CHO cells affects the intracellular lipid-linked oligosaccharide distribution, site occupancy and the N-glycosylation profile of a monoclonal antibody. *Journal of Biotechnology* 170:17-27.
- Liu HC, Gaza-Bulseco G, Faldu D, Chumsae C, Sun J. 2008. Heterogeneity of monoclonal antibodies. *Journal of Pharmaceutical Sciences* 97(7):2426-2447.
- Liu JT, Wang J, Fan L, Chen XN, Hu DD, Deng XC, Poon HF, Wang HB, Liu XP, Tan WS. 2015. Galactose supplementation enhance sialylation of recombinant Fc-fusion protein in CHO cell: an insight into the role of galactosylation in sialylation. *World Journal of Microbiology & Biotechnology* 31(7):1147-1156.
- Liu L. 2015. Antibody Glycosylation and Its Impact on the Pharmacokinetics and Pharmacodynamics of Monoclonal Antibodies and Fc-Fusion Proteins. *Journal of Pharmaceutical Sciences*:n/a-n/a.
- Liu YY, Slotine JJ, Barabasi AL. 2013. Observability of complex systems. *Proceedings of the National Academy of Sciences of the United States of America* 110(7):2460-2465.
- Losev E, Reinke CA, Jellen J, Strongin DE, Bevis BJ, Glick BS. 2006. Golgi maturation visualized in living yeast. *Nature* 441(7096):1002-1006.
- Ltd. DTT. 2015. 2016 Global Life Sciences Outlook: Moving forward with cautious optimism. Deloitte Touche Tohmatsu Ltd. p A1-A28.
- Majid FAA, Butler M, Al-Rubeai M. 2007. Glycosylation of an immunoglobulin produced from a murine hybridoma cell line: The effect of culture mode and the anti-apoptotic gene, bcl-2. *Biotechnology and Bioengineering* 97(1):156-169.
- Mandenius CF, Gustavsson R. 2015. Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics. *Journal of Chemical Technology and Biotechnology* 90(2):215-227.
- Marino K, Bones J, Kattla JJ, Rudd PM. 2010. A systematic approach to protein glycosylation analysis: a path through the maze. *Nature Chemical Biology* 6(10):713-723.
- Marique T, Cherlet M, Hendrick V, Godia F, Kretzmer G, Werenne J. 2002. A simplified general artificial neural network for the modelization of culture kinetics of different CHO strains. *Animal Cell Technology : Basic & Applied Aspects*, Vol 12:73-76.
- Matsuura-Tokita K, Takeuchi M, Ichihara A, Mikuriya K, Nakano A. 2006. Live imaging of yeast Golgi cisternal maturation. *Nature* 441(7096):1007-1010.

- McAtee AG, Templeton N, Young JD. 2014. Role of Chinese hamster ovary central carbon metabolism in controlling the quality of secreted biotherapeutic proteins. *Pharmaceutical Bioprocessing* 2(1):63-74.
- McCracken NA, Kowle R, Ouyang A. 2014. Control of galactosylated glycoforms distribution in cell culture system. *Biotechnology Progress* 30(3):547-553.
- McDonald A, Tipton K, Stroop C, Davey G. 2010. GlycoForm and Glycologue: two software applications for the rapid construction and display of N-glycans from mammalian sources. *BMC Research Notes* 3(1):173.
- Mechref Y, Novotny MV. 2002. Structural investigations of glycoconjugates at high sensitivity. *Chemical Reviews* 102(2):321-369.
- Medzihradszky KF. 2005. Characterization of protein N-glycosylation. *Mass Spectrometry: Modified Proteins and Glycoconjugates* 405:116-138.
- Mendelsohn R, Cheung P, Berger L, Partridge E, Lau K, Datti A, Pawling J, Dennis JW. 2007. Complex N-glycan and metabolic control in tumor cells. *Cancer Research* 67(20):9771-9780.
- Mittermayr S, Bones J, Doherty M, Guttman A, Rudd PM. 2011. Multiplexed Analytical Glycomics: Rapid and Confident IgG N-Glycan Structural Elucidation. *Journal of Proteome Research* 10(8):3820-3829.
- Mulukutla BC, Gramer M, Hu WS. 2012. On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. *Metabolic Engineering* 14(2):138-149.
- Mulukutla BC, Yongky A, Grimm S, Daoutidis P, Hu WS. 2015. Multiplicity of Steady States in Glycolysis and Shift of Metabolic State in Cultured Mammalian Cells. *Plos One* 10(3).
- Munzer DGG, Ivarsson M, Usaku C, Habicher T, Soos M, Morbidelli M, Pistikopoulos EN, Mantalaris A. 2015a. An unstructured model of metabolic and temperature dependent cell cycle arrest in hybridoma batch and fed-batch cultures. *Biochemical Engineering Journal* 93:260-273.
- Munzer DGG, Kostoglou M, Georgiadis MC, Pistikopoulos EN, Mantalaris A. 2015b. Cyclin and DNA Distributed Cell Cycle Model for GS-NS0 Cells. *Plos Computational Biology* 11(2).
- Muthing J, Kemminer SE, Conradt HS, Sagi D, Nimtz M, Karst U, Peter-Katalinic J. 2003. Effects of buffering conditions and culture pH on production rates and glycosylation of clinical phase I anti-melanoma mouse IgG3 monoclonal antibody R24. *Biotechnology and Bioengineering* 83(3):321-334.
- Neelamegham S, Marathe D, Liu G, Matta K. 2008. Systems Level Modeling of Cellular Glycosylation Reaction Networks: O-Linked Glycan Formation on Natural Selectin Ligands. *Glycobiology* 18(11):974-974.
- Nelson AL, Dhimolea E, Reichert JM. 2010. Development trends for human monoclonal antibody therapeutics. *Nature Reviews Drug Discovery* 9(10):767-774.
- Nelson PN, Reynolds GM, Waldron EE, Ward E, Giannopoulos K, Murray PG. 2000. Demystified ...: Monoclonal antibodies. *Molecular Pathology* 53(3):111-117.

- Ogunnaike BA. 1994. Online modeling and predictive control of an industrial terpolymerization reactor. *International Journal of Control* 59(3):711-729.
- Ogunnaike BA, Ray WH. 1994a. *Process Dynamics, Modeling, and Control*. New York: Oxford University Press.
- Ogunnaike BA, Ray WH. 1994b. *Process Dynamics, Modeling, and Control*: Oxford University Press.
- Ohadi SK, Aghamohseni H, Gädke J, Moo-Young M, Legge RL, Scharer J, Budman HM. Novel dynamic model to predict the glycosylation pattern of monoclonal antibodies from extracellular cell culture conditions; 2013 2013. p 30-35.
- Onitsuka M, Kawaguchi A, Asano R, Kumagai I, Honda K, Ohtake H, Omasa T. 2014. Glycosylation analysis of an aggregated antibody produced by Chinese hamster ovary cells in bioreactor culture. *Journal of Bioscience and Bioengineering* 117(5):639-644.
- Onitsuka M, Omasa T. 2015. Rapid evaluation of N-glycosylation status of antibodies with chemiluminescent lectin-binding assay. *Journal of Bioscience and Bioengineering* 120(1):107-110.
- Ozturk SS, Jorjani P, Taticek R, Lowe B, Shackleford S, LadehoffGuiles D, Thrift J, Blackie J, Naveh D. 1997. Kinetics of glucose metabolism and utilization of lactate in mammalian cell cultures. *Animal Cell Technology: from Vaccines to Genetic Medicine*:355-360.
- Pacis E, Yu M, Autsen J, Bayer R, Li F. 2011. Effects of cell culture conditions on antibody N-linked glycosylation—what affects high mannose 5 glycoform. *Biotechnology and Bioengineering* 108(10):2348-2358.
- Parekh RB. 1991. Mammalian cell gene expression: protein glycosylation. *Current opinion in biotechnology* 2(5):730-734.
- Patwardhan SC, Narasimhan S, Jagadeesan P, Gopaluni B, Shah SL. 2012. Nonlinear Bayesian state estimation: A review of recent developments. *Control Engineering Practice* 20(10):933-953.
- Planinc A, Bones J, Dejaegher B, Van Antwerpen P, Delporte C. 2016. Glycan characterization of biopharmaceuticals: Updates and perspectives. *Analytica Chimica Acta* 921:13-27.
- Preston RJS, Rawley O, Gleeson EM, O'Donnell JS. 2013. Elucidating the role of carbohydrate determinants in regulating hemostasis: insights and opportunities. *Blood* 121(19):3801-3810.
- Provost A, Bastin G. 2004. Dynamic metabolic modelling under the balanced growth condition. *Journal of Process Control* 14(7):717-728.
- Provost A, Bastin G, Agathos SN, Schneider YJ. 2006. Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. *Bioprocess and Biosystems Engineering* 29(5-6):349-366.
- Puri A, Neelamegham S. 2012. Understanding Glycomechanics Using Mathematical Modeling: A Review of Current Approaches to Simulate Cellular Glycosylation Reaction Networks. *Annals of Biomedical Engineering* 40(4):816-827.

- Raju S. 2003. Glycosylation variations with expression systems. *BioProcess International* 1:44-53.
- Raju TS. 2008. Terminal sugars of Fc glycans influence antibody effector functions of IgGs. *Current Opinion in Immunology* 20(4):471-478.
- Raju TS, Jordan RE. 2012. Galactosylation variations in marketed therapeutic antibodies. *Mabs* 4(3):385-391.
- Raju TS, Scallon B. 2007. Fc glycans terminated with n-acetylglucosamine residues increase antibody resistance to papain. *Biotechnology Progress* 23(4):964-971.
- Rathore AS, Bhambure R, Ghare V. 2010. Process analytical technology (PAT) for biopharmaceutical products. *Analytical and Bioanalytical Chemistry* 398(1):137-154.
- Rathore AS, Winkle H. 2009. Quality by design for biopharmaceuticals. *Nat Biotech* 27(1):26-34.
- Read EK, Bradley SA, Smitka TA, Agarabi CD, Lute SC, Brorson KA. 2013. Fermentanomics informed amino acid supplementation of an antibody producing mammalian cell culture. *Biotechnology Progress* 29(3):745-753.
- Read EK, Park JT, Brorson KA. 2011. Industry and regulatory experience of the glycosylation of monoclonal antibodies. *Biotechnology and Applied Biochemistry* 58(4):213-219.
- Reichert JM. 2012. Marketed therapeutic antibodies compendium. *mAbs* 4(3):413-415.
- Restelli V, Wang MD, Huzel N, Ethier M, Perreault H, Butler M. 2006. The effect of dissolved oxygen on the production and the glycosylation profile of recombinant human erythropoietin produced from CHO cells. *Biotechnology and Bioengineering* 94(3):481-494.
- Rives LM, Benga C, Zeng X; AbbVie, Inc.(North Chicago, IL), assignee. 2015. Methods to control protein heterogeneity. USA.
- Rivinoja A, Hassinen A, Kokkonen N, Kauppila A, Kellokumpu S. 2009. Elevated Golgi pH Impairs Terminal N-Glycosylation by Inducing Mislocalization of Golgi Glycosyltransferases. *Journal of Cellular Physiology* 220(1):144-154.
- Rivinoja A, Pujol FM, Hassinen A, Kellokumpu S. 2012. Golgi pH, its regulation and roles in human disease. *Annals of Medicine* 44(6):542-554.
- Rogers RS, Nightlinger NS, Livingston B, Campbell P, Bailey R, Balland A. 2015. Development of a quantitative mass spectrometry multi-attribute method for characterization, quality control testing and disposition of biologics. *mAbs* 7(5):881-890.
- Rosati S, van den Bremer ETJ, Schuurman J, Parren P, Kamerling JP, Heck AJR. 2013. In-depth qualitative and quantitative analysis of composite glycosylation profiles and other micro-heterogeneity on intact monoclonal antibodies by high-resolution native mass spectrometry using a modified Orbitrap. *Mabs* 5(6):917-924.
- Roth Z, Yehezkel G, Khalaila I. 2012. Identification and Quantification of Protein Glycosylation. *International Journal of Carbohydrate Chemistry* 2012:10.

- Rouiller Y, Perilleux A, Vesin MN, Stettler M, Jordan M, Broly H. 2014. Modulation of mAb quality attributes using microliter scale fed-batch cultures. *Biotechnology Progress* 30(3):571-583.
- Royle L, Radcliffe CM, Dwek RA, Rudd PM. 2007. Detailed Structural Analysis of N-Glycans Released From Glycoproteins in SDS-PAGE Gel Bands Using HPLC Combined With Exoglycosidase Array Digestions. In: Brockhausen I, editor. *Glycobiology Protocols*. Totowa, NJ: Humana Press. p 125-143.
- Ruhaak LR, Zauner G, Huhn C, Bruggink C, Deelder AM, Wuhrer M. 2010. Glycan labeling strategies and their use in identification and quantification. *Analytical and Bioanalytical Chemistry* 397(8):3457-3481.
- Sanderson CS, Barford JP, Barton GW. 1999. A structured, dynamic model for animal cell culture systems. *Biochemical Engineering Journal* 3(3):203-211.
- Sandoval WN, Arellano F, Arnott D, Raab H, Vandlen R, Lill JR. 2007a. Rapid removal of N-linked oligosaccharides using microwave assisted enzyme catalyzed deglycosylation. *International Journal of Mass Spectrometry* 259(1-3):117-123.
- Sandoval WN, Pham V, Ingle ES, Liu PS, Lill JR. 2007b. Applications of microwave-assisted proteomics in biotechnology. *Combinatorial Chemistry & High Throughput Screening* 10(9):751-765.
- Scott AM, Wolchok JD, Old LJ. 2012. Antibody therapy of cancer. *Nature Reviews Cancer* 12(4):278-287.
- Senger RS, Karim MN. 2005. Variable site-occupancy classification of N-linked glycosylation using artificial neural networks. *Biotechnology Progress* 21(6):1653-1662.
- Senger RS, Karim MN. 2008. Prediction of N-linked glycan branching patterns using artificial neural networks. *Mathematical Biosciences* 211(1):89-104.
- Serrato JA, Palomares LA, Meneses-Acosta A, Ramirez OT. 2004. Heterogeneous conditions in dissolved oxygen affect N-glycosylation but not productivity of a monoclonal antibody in hybridoma cultures. *Biotechnology and Bioengineering* 88(2):176-188.
- Sethuraman N, Stadheim TA. 2006. Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology* 17(4):341-346.
- Sha S, Agarabi C, Brorson K, Lee D-Y, Yoon S. 2016. N-Glycosylation Design and Control of Therapeutic Monoclonal Antibodies. *Trends in Biotechnology*.
- Shang TQ, Saati A, Toler KN, Mo JM, Li HY, Matlosz T, Lin X, Schenk J, Ng CK, Duffy T and others. 2014. Development and Application of a Robust N-Glycan Profiling Method for Heightened Characterization of Monoclonal Antibodies and Related Glycoproteins. *Journal of Pharmaceutical Sciences* 103(7):1967-1978.
- Shelikoff M, Sinskey AJ, Stephanopoulos G. 1996. A modeling framework for the study of protein glycosylation. *Biotechnology and Bioengineering* 50(1):73-90.

- Shirsat N, Mohd A, Whelan J, English NJ, Glennon B, Al-Rubeai M. 2015. Revisiting Verhulst and Monod models: analysis of batch and fed-batch cultures. *Cytotechnology* 67(3):515-530.
- Sidoli FR, Asprey SP, Mantalaris A. 2006. A Coupled Single Cell-Population-Balance Model for Mammalian Cell Cultures. *Industrial & Engineering Chemistry Research* 45(16):5801-5811.
- Sidoli FR, Mantalaris A, Asprey SP. 2004. Modelling of mammalian cells and cell culture processes. *Cytotechnology* 44(1-2):27-46.
- Slade PG, Caspary RG, Nargund S, Huang CJ. 2016. Mannose metabolism in recombinant CHO cells and its effect on IgG glycosylation. *Biotechnology and Bioengineering* 113(7):1468-1480.
- Solá RJ, Griebenow K. 2010. Glycosylation of Therapeutic Proteins: An Effective Strategy to Optimize Efficacy. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* 24(1):9-21.
- Solá RJ, Griebenow KAI. 2009. Effects of Glycosylation on the Stability of Protein Pharmaceuticals. *Journal of pharmaceutical sciences* 98(4):1223-1245.
- Soroush M. 1998. State and parameter estimations and their applications in process control. *Computers & Chemical Engineering* 23(2):229-245.
- Sou SN, Sellick C, Lee K, Mason A, Kyriakopoulos S, Polizzi KM, Kontoravdi C. 2015. How Does Mild Hypothermia Affect Monoclonal Antibody Glycosylation? *Biotechnology and Bioengineering* 112(6):1165-1176.
- Spahn PN, Hansen AH, Hansen HG, Arnsdorf J, Kildegaard HF, Lewis NE. 2016. A Markov chain model for N-linked protein glycosylation - towards a low-parameter tool for model-driven glycoengineering. *Metabolic Engineering* 33:52-66.
- Spiro RG. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12(4):43R-56R.
- St Amand M, M., Millili PG, McCabe MM, Ogunnaike BA. 2011. Strategic Vision for Integrated Process Analytical Technology and Advanced Control in Biologics Manufacturing. *PAT Applied in Biopharmaceutical Process Development And Manufacturing*: CRC Press. p 9-28.
- St. Amand MM, Ogunnaike BA, Robinson AS. 2014a. Development of at-line assay to monitor charge variants of MAbs during production. *Biotechnology Progress* 30(1):249-255.
- St. Amand MM, Radhakrishnan D, Robinson AS, Ogunnaike BA. 2014b. Identification of manipulated variables for a glycosylation control strategy. *Biotechnology and Bioengineering* 111(10):1957-1970.
- St. Amand MM, Tran K, Radhakrishnan D, Robinson AS, Ogunnaike BA. 2014c. Controllability Analysis of Protein Glycosylation in Cho Cells. *PLoS ONE* 9(2):e87973.
- Stadlmann J, Pabst M, Kolarich D, Kunert R, Altmann F. 2008. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics* 8(14):2858-2871.

- Stanley P, Schachter H, Taniguchi N. 2009. N-glycans. In: Varki A, Cummings R, Esko J, al. e, editors. *Essentials of Glycobiology*. 2nd Edition ed: Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Stanton D. 2015. Biosimilar discounts and switching will wipe-out J&J's Remicade in Norway, says regulator.
- Sun XM, Zhang YX. 2004. Glutamine cannot support recombinant CHO cell growth and maintenance in the absence of glucose. *Process Biochemistry* 39(6):717-720.
- Surve T, Gadgil M. 2015. Manganese Increases High Mannose Glycoform on Monoclonal Antibody Expressed in CHO When Glucose is Absent or Limiting: Implications for Use of Alternate Sugars. *Biotechnology Progress* 31(2):460-467.
- Tatiraju S, Soroush M, Mutharasan R. 1999. Multi-rate nonlinear state and parameter estimation in a bioreactor. *Biotechnology and Bioengineering* 63(1):22-32.
- Templeton N, Dean J, Reddy P, Young JD. 2013. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. *Biotechnology and Bioengineering* 110(7):2013-+.
- Tharmalingam T, Wu CH, Callahan S, Goudar CT. 2015. A Framework for Real-Time Glycosylation Monitoring (RT-GM) in Mammalian Cell Culture. *Biotechnology and Bioengineering* 112(6):1146-1154.
- Trummer E, Fauland K, Seidinger S, Schriebl K, Lattenmayer C, Kunert R, Vorauer-Uhl K, Weik R, Borth N, Katinger H and others. 2006. Process parameter shifting: Part I. Effect of DOT, pH, and temperature on the performance of Epo-Fc expressing CHO cells cultivated in controlled batch bioreactors. *Biotechnology and Bioengineering* 94(6):1033-1044.
- Tziampazis E, Sambanis A. 1994. Modeling of cell culture processes. *Cytotechnology* 14(3):191-204.
- Umaña P, Bailey JE. 1997. A mathematical model of N-linked glycoform biosynthesis. *Biotechnology and Bioengineering* 55(6):890-908.
- van Berkel PHC, Gerritsen J, Perdok G, Valbjorn J, Vink T, van de Winkel JGJ, Parren P. 2009. N-Linked Glycosylation is an Important Parameter for Optimal Selection of Cell Lines Producing Biopharmaceutical Human IgG. *Biotechnology Progress* 25(1):244-251.
- Villacres C, Tayi VS, Lattova E, Perreault H, Butler M. 2015. Low glucose depletes glycan precursors, reduces site occupancy and galactosylation of a monoclonal antibody in CHO cell culture. *Biotechnology Journal* 10(7):1051-1066.
- Wacker C, Berger CN, Girard P, Meier R. 2011. Glycosylation profiles of therapeutic antibody pharmaceuticals. *European Journal of Pharmaceutics and Biopharmaceutics* 79(3):503-507.
- Wagner-Rousset E, Bednarczyk A, Bussat MC, Colas O, Corvaia N, Schaeffer C, Van Dorselaer A, Beck A. 2008. The way forward, enhanced characterization of therapeutic antibody glycosylation: Comparison of three level mass

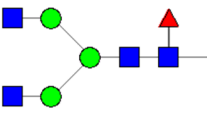
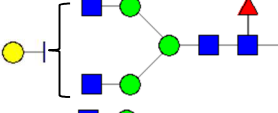
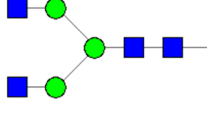
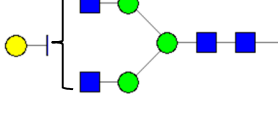
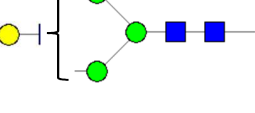
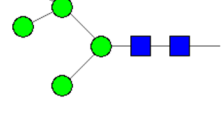
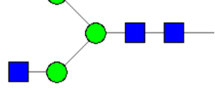
- spectrometry-based strategies. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 872(1-2):23-37.
- Wang WC, Lee N, Aoki D, Fukuda MN, Fukuda M. 1991. The poly-N-acetyllactosamines attached to lysosomal membrane glycoproteins are increased by the prolonged association with the Golgi complex. *Journal of Biological Chemistry* 266(34):23185-23190.
- Watts C. PAT–A framework for Innovative Pharmaceutical Development Manufacturing and Quality Assurance; 2004 2004.
- Weiner LM. 2007. Opinion - Building better magic bullets - improving unconjugated monoclonal antibody therapy for cancer. *Nature Reviews Cancer* 7(9):701-706.
- Wiegandt A, Meyer B. 2014. Unambiguous Characterization of N-Glycans of Monoclonal Antibody Cetuximab by Integration of LC-MS/MS and ¹H NMR Spectroscopy. *Analytical Chemistry* 86(10):4807-4814.
- Wong NSC, Wati L, Nissom PM, Feng HT, Lee MM, Yap MGS. 2010. An investigation of intracellular glycosylation activities in CHO cells: Effects of nucleotide sugar precursor feeding. *Biotechnology and Bioengineering* 107(2):321-336.
- Wright A, Morrison SL. 1997. Effect of glycosylation on antibody function: implications for genetic engineering. *Trends in biotechnology* 15(1):26-32.
- Wu P, Ray NG, Shuler ML. 1992. A Single-Cell Model for CHO Cells. *Annals of the New York Academy of Sciences* 665(1):152-187.
- Wu P, Ray NG, Shuler ML. 1993. A computer model for intracellular pH regulation in Chinese hamster ovary cells. *Biotechnology Progress* 9(4):374-384.
- Xing Z, Bishop N, Leister K, Li ZJ. 2010. Modeling kinetics of a large-scale fed-batch CHO cell culture by Markov chain Monte Carlo method. *Biotechnology Progress* 26(1):208-219.
- Xing ZZ, Kenty B, Koyrakh I, Borys M, Pan SH, Li ZJ. 2011. Optimizing amino acid composition of CHO cell culture media for a fusion protein production. *Process Biochemistry* 46(7):1423-1429.
- Xu P, Dai XP, Graf E, Martel R, Russell R. 2014. Effects of Glutamine and Asparagine on Recombinant Antibody Production Using CHO-GS Cell Lines. *Biotechnology Progress* 30(6):1457-1468.
- Yang M, Butler M. 2000. Effects of ammonia on CHO cell growth, erythropoietin production, and glycosylation. *Biotechnology and Bioengineering* 68(4):370-380.
- Yoo EM, Yu LJ, Wims LA, Goldberg D, Morrison SL. 2010. Differences in N-glycan structures found on recombinant IgA1 and IgA2 produced in murine myeloma and CHO cell lines. *Mabs* 2(3):320-334.
- Yoon SK, Choi SL, Song JY, Lee GM. 2005. Effect of culture pH on erythropoietin production by chinese hamster ovary cells grown in suspension at 32.5 and 37.0 degrees C. *Biotechnology and Bioengineering* 89(3):345-356.
- Zambare N, Soroush M, Ogunnaike BA. 2003. A method of robust multi-rate state estimation. *Journal of Process Control* 13(4):337-355.

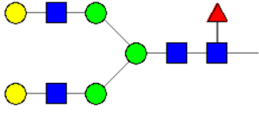
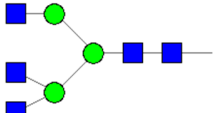
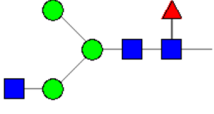
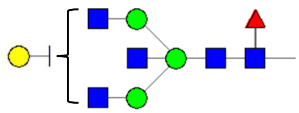
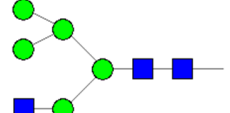
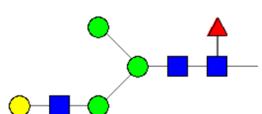
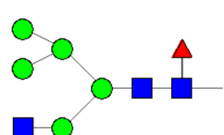
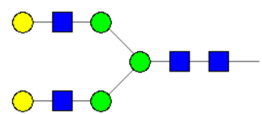
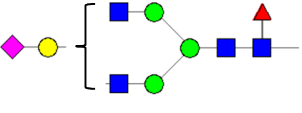
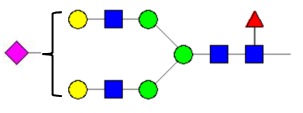
- Zhang A, Tsang VL, Markely LR, Kurt L, Huang Y-M, Prajapati S, Kshirsagar R. 2016a. Identifying the differences in mechanisms of mycophenolic acid controlling fucose content of glycoproteins expressed in different CHO cell lines. *Biotechnology and Bioengineering*:n/a-n/a.
- Zhang HF, Wang HB, Liu M, Zhang T, Zhang J, Wang XJ, Xiang WS. 2013. Rational development of a serum-free medium and fed-batch process for a GS-CHO cell line expressing recombinant antibody. *Cytotechnology* 65(3):363-378.
- Zhang PQ, Woen S, Wang TH, Liao B, Zhao S, Chen C, Yang YS, Song ZW, Wormald MR, Yu CF and others. 2016b. Challenges of glycosylation analysis and control: an integrated approach to producing optimal and consistent therapeutic drugs. *Drug Discovery Today* 21(5):740-765.
- Zheng K, Yarmarkovich M, Bantog C, Bayer R, Patapoff TW. 2014. Influence of glycosylation pattern on the molecular properties of monoclonal antibodies. *mAbs* 6(3):649-658.
- Zhong X, Wright JF. 2013. Biological Insights into Therapeutic Protein Modifications throughout Trafficking and Their Biopharmaceutical Applications. *International Journal of Cell Biology* 2013:19.

Appendix A

EXPERIMENTALLY OBSERVED GLYCAN SPECIES

Glycan structures Glycan structures drawn using GlycoForm software with sugar symbol set specified as per Consortium of Functional Glycomics. The masses correspond to the permethylated masses for the respective glycan peaks observed following mass spectrometry.

Glycan species	Structure	Mass (m/z)
FA2		1835.92
FA2G1		2040.02
A2		1661.83
A2G1		1865.93
A1G1		1620.8
M5		1579.78
A1		1416.7

FA2G2		2244.12
A3		1906.9
FA1		1590.79
FA2BG1		2285.2
M5A1		1824.9
FA1G1		1794.89
FM5A1		1998.9
A2G2		2070.036
FA2G1S1		2401.19
FA2G2S1		2605.29

Appendix B

MIXED LEVEL DESIGN OF EXPERIMENTS

The accompanying table lists the full factorial ($2^2, 3^2$) experimental design as well as the labels associated with each experimental condition.

Experiment	MnCl ₂ level	EDTA level	MnCl ₂ Addition	EDTA Addition	Label
1	-1	-1	-1	-1	Control
2	-1	-1	-1	0	
3	-1	-1	-1	+1	
4	-1	-1	0	-1	
5	-1	-1	0	0	
6	-1	-1	0	+1	
7	-1	-1	+1	-1	
8	-1	-1	+1	0	
9	-1	-1	+1	+1	
10	-1	+1	-1	-1	ED D0
11	-1	+1	0	-1	
12	-1	+1	+1	-1	

Table continued

13	-1	+1	-1	0	ED D3
14	-1	+1	0	0	
15	-1	+1	+1	0	
16	-1	+1	-1	+1	ED D6
17	-1	+1	0	+1	
18	-1	+1	+1	+1	
19	+1	-1	-1	-1	Mn D0
20	+1	-1	-1	0	
21	+1	-1	-1	+1	
22	+1	-1	0	-1	Mn D3
23	+1	-1	0	0	
24	+1	-1	0	+1	
25	+1	-1	+1	-1	Mn D6
26	+1	-1	+1	0	
27	+1	-1	+1	+1	
28	+1	+1	-1	-1	ED D0/ Mn D0
29	+1	+1	-1	0	ED D3/ Mn D0
30	+1	+1	-1	+1	ED D6/ Mn D0
31	+1	+1	0	-1	ED D0/Mn D3

Table continued

32	+1	+1	0	0	ED D3/ Mn D3
33	+1	+1	0	+1	ED D6/ Mn D3
34	+1	+1	+1	-1	ED D0/Mn D6
35	+1	+1	+1	0	ED D3/Mn D6
36	+1	+1	+1	+1	ED D6/Mn D6

†D0, D3, D6 refer to the time of addition of the supplement MnCl₂ (Mn) or EDTA (ED) on day 0, day 3, or day 6 respectively.

Gain matrix generated from statistically significant ($p \leq 0.05$) coefficients obtained from ANOVA of the full factorial design experimental data

	FA2	FA2G1	A2	A2G1	A1G1	M5	A1	FA2G2	A3	FA1	FA2BG1	M5A1	FA1G1	FM5A1	M6A1	A2G2	FA2G1S1	FA2G2S1
MnCl2	2.17	0.00	-2.21	-0.50	0.37	-0.86	0.00	0.00	0.00	0.00	0.00	0.27	0.22	0.13	0.13	0.07	0.00	0.00
EDTA	0.00	-1.79	2.34	0.80	0.00	0.00	0.00	-0.48	0.00	-0.55	0.12	0.00	-0.09	0.00	0.00	-0.07	0.00	-0.05
Mn Time 1	0.00	-0.87	1.45	0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mn Time 2	-1.08	0.00	0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ED Time 1	0.00	2.67	-0.92	0.00	0.00	0.00	-0.57	0.59	0.00	-0.32	0.00	0.00	0.00	0.10	0.09	0.12	0.00	0.06
ED Time 2	0.00	-1.66	0.00	0.00	0.00	0.00	0.51	-0.27	0.00	0.62	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00
MnCl2 – EDTA	2.35	1.76	-2.12	0.00	0.00	-0.99	-0.87	0.18	0.20	-0.23	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
MnCl2 - Mn Time1	0.00	0.87	-1.45	-0.78	0.00	0.00	0.00	0.00	0.00	0.00	-0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MnCl2 - Mn Time2	1.08	0.00	-0.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MnCl2 - ED Time1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MnCl2 - ED Time2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EDTA - Mn Time 1	0.00	0.00	0.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EDTA - Mn Time 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
EDTA - ED Time 1	0.00	-2.67	0.92	0.00	0.00	0.00	0.57	-0.59	0.00	0.32	0.00	0.00	0.00	-0.10	-0.09	-0.12	0.00	-0.06
EDTA - ED Time 2	0.00	1.66	0.00	0.00	0.00	0.00	-0.51	0.27	0.00	-0.62	0.00	0.00	-0.15	0.00	0.00	0.00	0.00	0.00

Table continued

[illegible]

Main and interaction effects and corresponding confounding factors

EDTA		
MnCl ₂		
MnCl ₂ – EDTA		
ED T1	=	-1*EDTA - ED T1
ED T2	=	-1*EDTA - ED T2
Mn T1	=	-1*MnCl ₂ - Mn T1
Mn T2	=	-1*MnCl ₂ - Mn T2
Mn T1 - ED T1	=	-1*MnCl ₂ - Mn T1 - ED T1
Mn T2 - ED T1	=	-1*MnCl ₂ - Mn T2 - ED T1
Mn T2 - ED T2	=	-1*MnCl ₂ - Mn T2 - ED T2
Mn T1 - ED T2	=	-1*MnCl ₂ - Mn T1 - ED T2
EDTA - Mn T1	=	-1*MnCl ₂ -EDTA-Mn T1
MnCl ₂ - ED T2	=	-1*MnCl ₂ -EDTA-ED T2
EDTA - Mn T2	=	-1*MnCl ₂ -EDTA-Mn T2
EDTA - Mn T1 - ED T1	=	-1*MnCl ₂ - EDTA - Mn T1 - ED T1
EDTA - Mn T2 - ED T1	=	-1*MnCl ₂ - EDTA - Mn T2 - ED T1
EDTA - Mn T2 - ED T2	=	-1*MnCl ₂ - EDTA - Mn T2 - ED T2
EDTA - Mn T1 - ED T2	=	-1*MnCl ₂ - EDTA - Mn T1 - ED T2
MnCl ₂ - ED T1	=	-1*MnCl ₂ -EDTA-ED T1

“Reduced” gain matrix (**K**) obtained by eliminating the redundant rows from the gain matrix listed in Table A2

[illegible]

Unitary matrix **W** obtained from SVD of gain matrix **K**

	η_1	η_2	η_3	η_4	η_5	η_6	η_7	η_8	η_9	η_{10}	η_{11}	η_{12}	η_{13}	η_{14}	η_{15}	η_{16}	η_{17}	η_{18}
FA2	-0.38	0.56	0.56	-0.30	-0.02	0.04	0.35	-0.10	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FA2G1	-0.62	-0.71	0.18	-0.27	-0.03	-0.03	-0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A2	0.64	-0.33	0.54	-0.29	-0.11	0.00	0.01	-0.32	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A2G1	0.12	-0.05	0.39	0.14	0.59	0.19	-0.11	0.60	-0.02	-0.01	0.02	-0.06	0.02	-0.19	-0.06	-0.07	0.00	-0.03
A1G1	-0.02	0.06	0.00	-0.08	-0.23	0.40	-0.23	0.01	-0.19	-0.44	0.39	0.44	0.25	-0.25	-0.05	0.01	0.00	-0.10
M5	0.15	-0.20	-0.20	-0.03	0.01	0.14	0.82	0.22	0.13	-0.25	-0.06	0.12	0.21	0.05	0.11	0.01	0.00	-0.01
A1	0.13	0.06	-0.27	-0.60	-0.17	0.25	0.01	0.34	-0.20	0.46	-0.06	0.03	-0.20	-0.08	-0.17	-0.09	0.00	-0.03
FA2G2	-0.10	-0.11	-0.04	0.28	0.26	0.57	0.20	-0.50	-0.24	0.31	-0.01	-0.04	-0.06	-0.09	-0.11	-0.16	0.00	-0.10
A3	-0.02	0.01	0.04	0.04	0.11	-0.22	-0.06	-0.05	0.06	0.26	-0.48	0.65	0.38	-0.06	-0.24	-0.03	0.00	-0.09
FA1	0.02	0.06	-0.30	-0.51	0.65	-0.10	-0.08	-0.31	0.04	-0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FA2BG1	0.02	0.01	0.07	0.00	0.11	0.19	-0.09	0.08	-0.34	0.00	-0.06	0.26	-0.07	0.77	0.24	0.29	0.00	0.11
M5A1	-0.02	0.04	0.00	-0.06	-0.17	0.29	-0.17	0.01	-0.14	-0.32	-0.72	-0.31	0.17	-0.11	0.16	-0.10	0.00	0.18
FA1G1	-0.02	0.05	-0.05	-0.13	0.03	0.24	-0.18	-0.01	0.45	0.35	0.17	-0.18	0.58	0.12	0.39	0.01	0.00	-0.07
FM5A1	-0.02	0.01	0.02	0.02	-0.03	0.18	-0.10	0.00	0.47	-0.09	-0.10	0.34	-0.50	0.09	0.31	-0.50	0.00	-0.06
M6A1	-0.01	0.01	0.00	0.00	-0.07	0.25	-0.08	0.02	0.40	-0.18	-0.03	-0.16	0.03	0.41	-0.73	0.01	0.00	-0.07
A2G2	-0.02	-0.01	-0.02	0.01	0.02	0.23	-0.02	-0.04	0.33	0.03	-0.17	0.10	-0.30	-0.29	0.06	0.79	0.00	-0.05
FA2G1S1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
FA2G2S1	-0.01	-0.01	-0.01	0.01	0.04	0.08	0.01	-0.04	0.13	0.09	0.13	0.14	0.03	-0.09	-0.10	-0.03	0.00	0.95

Diagonal matrix of singular values (Σ) obtained from SVD of reduced gain matrix \mathbf{K}

[illegible]

Unitary matrix \mathbf{V}^T obtained from SVD of reduced gain matrix \mathbf{K}

	MnCl ₂	EDTA	Mn- EDTA	Mn T1	Mn T2	ED T1	ED T2	Mn- ED T2	ED- Mn T1	Mn T1 - ED T1	Mn T2 - ED T1	Mn T2 - ED T2	EDTA - Mn T1 - ED T1	EDTA - Mn T2 - ED T1	EDTA - Mn T2 - ED T2
μ_1	-0.39	0.44	-0.58	0.25	0.11	-0.38	0.18	0.00	0.09	0.09	-0.07	0.11	-0.09	0.07	-0.11
μ_2	0.60	0.13	0.24	0.03	-0.21	-0.47	0.35	0.00	-0.07	0.18	-0.13	0.21	-0.18	0.13	-0.21
μ_3	0.00	0.65	0.45	0.42	-0.16	0.10	-0.28	-0.02	0.20	-0.08	0.06	-0.09	0.08	-0.06	0.09
μ_4	-0.18	0.11	0.18	-0.08	0.25	0.27	-0.34	-0.10	-0.30	0.31	-0.23	0.36	-0.31	0.23	-0.36
μ_5	-0.39	-0.34	0.33	0.57	-0.05	0.11	0.48	0.17	-0.15	0.04	-0.03	0.05	-0.04	0.03	-0.05
μ_6	0.49	-0.06	-0.48	0.48	-0.09	0.52	-0.03	-0.04	0.01	0.05	-0.04	0.06	-0.05	0.04	-0.06
μ_7	0.26	0.09	0.12	0.13	0.91	-0.02	0.16	0.03	-0.03	-0.08	0.06	-0.09	0.08	-0.06	0.09
μ_8	-0.01	-0.28	0.08	0.04	0.12	-0.05	-0.13	0.16	0.87	0.13	-0.10	0.15	-0.13	0.10	-0.15
μ_9	-0.05	0.37	0.03	-0.41	-0.04	0.52	0.60	0.06	0.23	0.03	-0.02	0.03	-0.03	0.02	-0.03
μ_{10}	-0.06	-0.11	0.07	0.07	0.02	-0.01	0.15	-0.96	0.16	-0.01	0.00	-0.01	0.01	0.00	0.01
μ_{11}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.16	0.46	0.49	0.27	-0.46	-0.49
μ_{12}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.52	-0.43	0.18	0.54	0.43	-0.18
μ_{13}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.73	-0.03	-0.04	0.68	0.03	0.04
μ_{14}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	-0.45	0.00	0.54	-0.45
μ_{15}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.45	-0.54	0.00	-0.45	-0.54