ADVANCING CELL CULTURE ENGINEERING THROUGH MECHANISTIC MODEL OPTIMIZATION

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE
UNIVERSITY OF MINNESOTA
BY

Conor Michael O'Brien

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisers: Professor Wei-Shou Hu And Professor Prodromos Daoutidis

April 2020

Acknowledgements

First and foremost, I would like to thank my advisors Dr. Wei-Shou Hu and Dr. Prodromos Daoutidis for their constant support and mentorship during my time in their labs. They have helped me grow as a scientist but just as importantly as a person throughout my Ph.D. program. It has been and honor and a privilege to work with them over the past several years, and I will carry their lessons with me.

I would also like to thank my thesis committee members, Dr. Qi Zhang and Dr. Samira Azarin in the Department of Chemical Engineering and Materials Science, and Dr. Douglas Mashek in the Department of Biochemistry, Molecular Biology and Biophysics for their feedback and comments during my defense but also throughout my Ph.D.

Additionally, I would like to thank current and former members of the groups for our many great discussions and collaborations: Tung Le, Arpan Bandyopadhyay, Kyoungho Lee, Sofie O'Brien, Jennifer One, Andrew Allman, Udit Gupta, Christopher Stach, Meghan McCann, Kevin Ortiz-Rivera, Zion Lee, Thu Phan, David Chau, Michael Zachar, Dong Seong Cho, and Yen-An Lu. I would also like to thank Dr. Bhanu Chandra Mulukutla for our discussions throughout writing our review from which I learned much about metabolism and how to critically read papers. I would also like to thank Dr. Qi Zhang, Dr. Douglas Mashek, and Dr. Michael Smanski for our collaboration and your constructive feedback on my research.

I have been privileged to have made so many great friends who have supported me through graduate school: Jen, Matt, Jake, Tiff, Josh, Meghan, Andrew, Mike, Beth, Ryan, Erika, Jen, and George. You have made these past several years deeply enjoyable.

Finally, I would like to thank my family for their many years of support. To my wife Sofie, your support and love has made each day brighter. To my parents, who supported me for so many years until I could reach my potential. In no way could I have made it here without you.

Dedication

To my parents Jake and Susan and my wife Sofie.

Abstract

Over the past few decades, the emergence of new classes of treatments, including protein therapeutics, gene therapies, and cell therapies, has ushered in a new era of medicine. Unlike small molecule therapeutics, these treatments are produced in or consist of cells, typically mammalian in origin. Processes have been developed to produce many of these drugs at large scale, often in stirred tank bioreactors. Significant effort has driven staggering increases in the productivity of these processes, enabling economical manufacturing, and the potential to drive down costs and make drugs more widely available.

However, the bioreactor is not a natural environment for cells isolated from a multicellular mammalian organism. Many biological regulations are carried over from the cells' origin and can result in numerous undesirable behaviors manifesting in the dense, highly productive reactor environment. In certain culture stages, or in the case of excess nutrient supply, cells will secrete undesirable metabolites including lactate, ammonia, and many byproducts of amino acid metabolism. These compounds can retard cell growth, or otherwise alter the potency or productivity of the cultures.

Traditional biologics process development employs the use of statistical design of experiments, often encompassing several reactors run in parallel for multiple rounds of experiments over a few months. There is thus substantial room for improvement for both the outcome of the development process, such as an increase in titer, and the time it takes to complete the development stage. Given that cell culture processes share intrinsic similarities in their underlying mechanistic behavior, there exists significant opportunity to reduce the overall number of experiments needed for process development, scaling, and diagnostics using models rather than treating cell culture processes as a black box.

In this thesis, we present the case for the use of mathematical optimization of mechanistic models to accurately describe cell culture processes and augment their behavior. We first outline recent advances in understanding of metabolic regulation and homeostasis. Cell signaling and metabolic networks interact over multiple time-scales and through multiple means, resulting in cell metabolism with nonlinear behavior that is consequently context-dependent.

In the following sections of this work, we then develop an optimization framework which can efficiently be used for the design of experiments to rewire cellular metabolism through metabolic engineering, or to otherwise understand the biological requirements of different metabolic phenomena. This framework is first applied to the Warburg effect, a century-old unsolved problem of rapid lactate production in proliferating cells to identify which enzymes may be altered to mitigate the lactate production. This framework in then applied to the problem of hepatic gluconeogenesis to study metabolic disease. As the expression of the enzymes specific to gluconeogenesis is not sufficient for glucose production, we explore what other requirements exist for the synthesis of glucose from different substrates.

The next portion discusses the construction and optimization of a bioprocess model which includes metabolism, signaling, cell growth, and the reactor environment. This model is fit to a manufacturing-scale dataset to explore the origins of process variability and potential mitigation strategies.

In the final segment of this thesis, we explore another aspect of protein therapeutics: product quality. A model of N-glycosylation is optimized in conjunction with successive rounds of experimentation with the goal of improving the galactose content on an antibody. This work highlights the benefits of feeding back experimental data to refine model parameters for better design and prediction of subsequent experiments.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABBREVIATION	DEFINITION
13BPG	1,3-bisphosphoglycerate
2HG	2-hydroxyglutarate
2PG	2-phosphoglycerate
2-P-LACTATE	2-phospho-L-lactate
3PG	3-phosphoglycerate
6PG	6-phosphogluconate
6PGDH	6-phosphogluconate dehydrogenase
ACLY	ATP citrate lyase
ACO	aconitase
ADHFE	alcohol dehydrogenase iron containing
ALDO	aldolase
AMPK	AMP-activated protein kinase
ASNS	asparagine synthetase
ASPG	asparaginase
CARM	co-activator-associated arginine methyltransferase
CHO	Chinese hamster ovary
CS	citrate synthase
DHAP	dihydroxyacetone phosphate
E4P	erythrose 4-phosphate
ENO	enolase
F16BP	fructose 1,6-bisphosphate
F26BP	fructose 2,6-bisphosphate
F6P	fructose 6-phosphate
FH	fumarase
G3P	glyceraldehyde 3-phosphate
G3PS	glycerol-3-phosphate shuttle
G6P	glucose 6-phosphate
G6PD	glucose-6-phosphate dehydrogenase
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GLCNAC	N-acetylglucosamine
GLDH	glutamate dehydrogenase
GLS	glutaminase
GLUT	glucose transporter, also known as the SLC2A family
GLY3P	glycerol 3-phosphate
GOT	glutamate-oxaloacetate transaminase
GPD	glycerol-3-phosphate dehydrogenase
GPR	G protein-coupled receptor
GPT	glutamate-pyruvate transaminase

HIF	hypoxia-inducible factor
HK	hexokinase
IDH	isocitrate dehydrogenase
K/P	Kinase to phosphatase ratio of PFKFB
LDH	lactate dehydrogenase
MAS	malate-aspartate shuttle
MDH	malate dehydrogenase
ME	malic enzyme
MPC	mitochondrial pyruvate carrier
MTOR	mammalian target of rapamycin
NDRG3	N-myc downstream regulated gene 3
OAA	oxaloacetate
OGDC	oxoglutarate dehydrogenase complex
OXPHOS	oxidative phosphorylation
PC	pyruvate carboxylase
PDHC	pyruvate dehydrogenase complex
PDK	pyruvate dehydrogenase kinase
PEP	phosphoenolpyruvate
PEPCK	phosphoenolpyruvate carboxykinase (PCK)
PFK	phosphofructokinase
PFKFB	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase
PGI	phosphoglucoisomerase
PGK	phosphoglycerate kinase
PGM	phosphoglycerate mutase
PGP	phosphoglycolate phosphatase
PHD	prolyl hydroxylase domain-containing
PK	pyruvate kinase
PPP	pentose phosphate pathway
PRPP	phosphoribosyl pyrophosphate
PSAT	phosphoserine transaminase
R5P	ribose 5-phosphate
RPE	ribulose-phosphate 3-epimerase
RPI	ribose 5-phosphate isomerase
RU5P	ribulose 5-phosphate
S7P	seduheptulose 7-phosphate
SAICAR	phosphoribosylaminoimidazolesuccinocarboxamide
SCS	succinyl-CoA synthetase
SDH	succinate dehydrogenase
SLC	solute carrier
TA	transaldolase
TCA CYCLE	tricarboxylic acid cycle

TKT	transketolase
TPI	triose-phosphate isomerase
TRX	thioredoxin
TXNIP	thioredoxin interacting protein
UDP	Uridine diphosphate
VDUP	vitamin D3 upregulated protein
X5P	xylulose 5-phosphate
αKG	α-ketoglutarate

1. Introduction

1.1. Advancing cell culture engineering through optimization

Cell culture engineering, or the improvement of biotechnological processes, has transformed modern medicine by amplifying the production of complex biopharmaceuticals, creating many new avenues of treatment for a host of diseases. Currently, industrial cell culture is largely used to produce protein therapeutics (also known as biologics) and vaccines, but cell and gene therapies are rapidly evolving fields which show great potential to transform patients' lives.

Over the past few decades, significant strides have been made in the standardization of practices in cell culture from cell handing, to media, and even bioreactor design. Automation of bioreactor operation has also enabled the collection of massive datasets and large gains in culture productivity and reduced variability. Design of experiments based on statistical models is now pervasive through the drug development process, simplifying workflows and improving outcomes. However, the treatment of cells as a black box will ultimately limit advancement in cell culture development and understanding of the mechanisms governing cell behavior. The development of new drugs and the decrease in their costs should not be held back by such black box treatment, especially in an age where the use of sophisticated models is standard practice amongst many other industries.

Cell behavior in the bioreactor environment is dictated by many factors. One of the most critical among these factors is central metabolism. Central metabolism is highly regulated at the metabolite level through allosteric regulation, at the protein level through post-translational modifications by signaling proteins, and at the transcriptional level. This multilayered regulation allows cells to reach internal homeostasis in a wide variety of conditions. While many of these regulations are essential for cell proliferation, mammalian cells did not evolve to grow efficiently

in a cell-dense bioreactor. Either through careful control of operation conditions, or engineering of the cells themselves, there is significant headroom to increase culture productivity.

Most cells in bioprocessing are given large quantities of glucose, amino acids, and other compounds to meet cellular needs for growth and biosynthesis. As the culture progresses, cells grow, produce antibody, and secrete waste products. Notably, the production of large quantities of lactate is a nearly universal phenomena of fast-growing cells, known as the Warburg effect. Balancing cellular needs while not providing excess nutrients is one of many ways to alter cell culture processes to drastically improve culture performance.

Metabolism is further linked to product quality, notably *N*-glycosylation for many antibody therapeutics, as it provides numerous precursors for the generation of nucleotide sugars which serve as substrates for the process. Metabolism is also linked to the bioreactor state: as lactate and carbon dioxide are produced during cell culture, the pH drops, which must be countered by the addition of base, typically sodium hydroxide or sodium carbonate. Carbon dioxide produced by cells must also be stripped out of the medium by aeration and agitation of the bioreactor, although large reactors often suffer from limited transport capacity and carbon dioxide accumulation during the cell dense portion of the culture. In the late stage of culture, lactate can be consumed, increasing the pH and necessitating the use of acid to again control pH at the desired level. These control events maintain characteristics of the bioreactor to maintain an ideal environment for cell growth, except that they increase the osmolarity of the culture, with the potential to result in a poor extracellular environment due to osmotic pressure in the late stage of the culture. Osmolarity can further trigger cell stress and subsequently alter cell metabolism in undesirable ways. Thus, metabolism and reactor control can form a vicious cycle.

The interaction of cells with the reactor environment, coupled with highly nonlinear behavior originating from the complexity of metabolic enzymes, inevitably means that over-simplified

models will have limited predictive performance in guiding cell culture process development. As mechanistic, first-principles models contain the underlying biological and physical relationships key to observed biological behavior, they can have vastly improved predictive capacity. However, mechanistic relationships are not known for all processes relevant to cell culture, and even when models are available their nonlinearity and stiffness makes them difficult to work with.

Due to the nonlinearity of biological models, exploration by simulation can be very limited, even for systems where physical intuition may be applied. These limitations are overcome in this work using mathematical optimization, a natural tool for rewiring metabolic pathways and for parameter estimation to fit models to experimental data.

In this thesis, optimization methods are developed to make working with mechanistic biological models computationally tractable, and then applied to designing metabolic engineering experiments to rewire central metabolism, understanding the requirements of metabolic phenotypes, modeling manufacturing scale processes to understand the origins of process variability, and improving product quality.

1.2. Scope of thesis

The focus of this thesis is on the development and application of optimization methods on biological models of metabolism, glycosylation, and more broadly the reactor environment.

In the first portion of the thesis, we overview recent advances in the understanding of metabolism and cell signaling networks, and how their interactions drive metabolic homeostasis. Metabolism is regulated over different timescales: at the short timescale through interactions with metabolites, over longer timescales by post-translational modifications, and over long times by transcriptional regulation as downstream effects from cell signaling. We also highlight the bidirectional nature of these interactions; with metabolism both regulating and being regulated by

key cell signaling proteins. By understanding the key regulations in metabolism, we are better equipped to build and optimize bioprocess-relevant metabolic models.

In the following chapters, we will discuss the development of a framework for optimization of nonlinear, stiff biological models to understand or alter key metabolic phenomena. This algorithm is then applied to a mechanistic model of metabolism and used to design metabolic engineering experiments to rewire the Warburg effect. In this way, potential combinations of enzymes are identified which through alteration of their expression level can result in reduced or eliminated lactate production at the fast growth stage.

This metabolic model is then extended through the addition of several key enzymes found in hepatocyte liver cells which facilitate the process of gluconeogenesis. Gluconeogenesis, or the production of glucose from several different potential carbon substrates, is an essential part of glucose homeostasis in the body whose dysregulation is associated with diabetes and fatty liver disease. However, simply expressing the essential enzymes in gluconeogenesis does not result in an appreciable rate of glucose synthesis. We explore the differential metabolic requirements for the utilization of the primary gluconeogenic substrates, both in isolation and in tandem, to illuminate the mechanisms governing gluconeogenesis in the body.

We then combine the original metabolic model discussed in the first section with a mechanistic model of reactor behavior and empirical models of cell growth and signaling. The model is then fit to a set of manufacturing scale CHO cell culture data. This dataset has inherent variability, with high and low performance runs (as measured by titer) diverging in metabolic behavior in the late stage of culture despite initially having very similar behavior. The low titer runs begin to produce significant quantities of lactate, which requires input of additional base to counter the pH, forming unfavorable conditions for cell growth, decreasing both viable cell density, and final titer. The

combined hybrid mechanistic-empirical model is used to hypothesize the origins of this variability and to explore process level changes which could mitigate the undesirable behavior.

Finally, a model of *N*-glycosylation is fit to a set of experimental data for a CHO cell line producing an immunoglobulin G, a common class of protein therapeutics, with the aim of improving product quality though altering the glycan profile of the antibody. This model is used in conjunction with rounds of single and multi-gene glycoengineering experiments to predict potential gene changes which may have beneficial outcomes. The systems biology approach, combining modeling and rounds of experimental design was able to identify engineered cell lines with substantially increased galactosylation, a highly desired product quality attribute for many protein therapeutics.

1.3. Thesis organization

This thesis is divided into 6 chapters. Chapter 2 covers a broad overview of metabolic regulation, highlighting the two-way interactions between metabolic and cell signaling networks. Chapter 3 details a framework developed for the optimization of stiff biological models. This framework is applied to study and rewire the Warburg effect, with the aim to reduce lactate production during the rapid proliferation stage. Chapter 4 discusses application of this framework to study gluconeogenesis in an extended metabolic model. The liver has evolved to synthesize glucose from multiple carbon sources, depending on availability, but the requirements for each are not straightforward, but are unraveled through optimization. Chapter 5 presents an extension of the metabolic model to include the reactor environment and cell signaling. This model is then used to hypothesize about the origins of manufacturing process variability. By extending a mechanistic model with empirical components, its scope can be expanded while maintaining its biological origins. Detailed phenomena arising from cell signaling regulation and interaction of metabolism with the changing reactor environment are explored. Chapter 6 contains a discussion of

optimization to fit a model of *N*-glycosylation to experimental data to improve product quality through genetic engineering. Appendix A contains a discussion of the development of a visualization tool for *O*-glycosylation networks.

2. Regulation of metabolic homeostasis in cell culture bioprocesses

Reproduced with permission from O'Brien, C. M., Mulukutla, B.C., Mashek D. G., & Hu, W. S. (2020). Regulation of metabolic homeostasis in cell culture bioprocesses. *Trends in Biotechnology*, (In Press).

2.1. Introduction: Importance of homeostasis in central metabolism in biologics production

The pathways involved in energy metabolism, glycolysis, the tricarboxylic acid (TCA) cycle, and the pentose phosphate pathway (PPP) are ubiquitous in all cells. These pathways serve to meet the diverse metabolic needs of different tissues. This is possible because based on their needs tissues express distinct combinations of isozymes which are subject to different types of regulation that determine their kinetic behavior. Most cells in tissues are naturally in a quiescent state and as a result metabolize glucose chiefly through oxidation to carbon dioxide (CO_2). Intriguingly, upon becoming proliferative they also acquire a high glycolytic metabolic phenotype and increase lactic acid production, a phenomenon known as the Warburg effect. Such high glycolytic metabolism occurs despite aerobic conditions and is sometimes referred to as aerobic glycolysis.

Mammalian cells are the core of cell culture-based manufacturing for protein biologics, vaccines, and live-cell therapeutics. Biopharmaceuticals are valued at over US$188 billion per annum globally [1]. In cell culture processes, cell metabolism is the driver for changes in the chemical environment. Cells under different culture conditions may exhibit aerobic glycolysis or a more oxidative metabolism. A switch of glucose metabolism from a glycolytic to an oxidative state

is correlated with high productivity. Importantly, the productivity and glycosylation pattern of protein biologics is affected by the metabolic state of the production cell.

Central metabolism (Figure 2.1) is highly regulated and adjusts its fluxes dynamically to meet the diverse cellular needs of biosynthesis and energy generation under different conditions. Regulation of glucose metabolism takes place at the enzyme activity level via allosteric regulation and post-translational modifications, at the transcriptional level by signaling pathways and growth control, and at the spatial localization level. Together, these different levels of regulation achieve a homeostatic state that responds to environmental changes and growth needs. Understanding the regulation of glucose metabolism increases our ability to enhance process productivity and product quality. This article reviews recent findings on the regulation of glucose metabolism with the aim of better guiding the metabolism of cells in culture.

2.2. Regulation of enzyme activity over short timescales

The first layer of control of cellular energy metabolism comprises allosteric regulation exerted, over short timescales, on enzymes by metabolites upstream and downstream of the enzyme. These allosteric interactions modulate the activity of the target enzyme in accordance with the level of regulating metabolites. The universal pathways (Figure 2.1) gain their characteristics in different cells by expressing different combinations of isoforms. Different isoforms of an enzyme have distinct kinetic properties and respond to allosteric feedforward or feedback regulation in a tissue-specific manner. Classical metabolic nodes in glycolysis that are subject to allosteric regulations include HK, PFK, PFKFB, and PK (Figure 2.2**).** This regulation and its implications for metabolic flux have been discussed in greater detail in previous reviews [2, 3]. For cells in culture the dominant isoforms are HK1/2, PFKL, PFKFB3 and PKM2, but other isoforms are often coexpressed. Activation of PFK (L or M) by F26BP and activation of PKM2 by F16BP play key

roles in elevating glycolytic flux in aerobic glycolysis. The accumulation of PEP and G6P inhibits the activity of PFKFB and HK, respectively.

Allosteric regulation influences the outflow of intermediates of glycolysis to other biosynthetic pathways. The muscle (M) isoform of PK has two alternatively spliced gene products, PKM1 and PKM2. PKM1 exists as an active tetramer. PKM2 is activated by F16BP, serine, and phosphoribosylaminoimidazole-succinocarboxamide (SAICAR), among many other metabolites [4, 5]. F16BP and serine binding promote a switch from a PKM2 inactive dimer to an active tetrameric state, whereas SAICAR binding to PKM2 dimer was reported to convert the dimer into an active dimer form (Figure 2.2) [5]. F16BP signals glucose availability. Serine is synthesized from the glycolytic intermediate 3-phosphoglycerate (3PG) and is a precursor for further amino acid synthesis, and SAICAR is an intermediate of *de novo* purine nucleotide synthesis which accumulates under glucose starvation. SAICAR signals the need for increased glycolytic flux, whereas serine shortage decreases PKM2 activity to increase 3PG level for its diversion to serine biosynthesis [4, 6] (serine metabolism is reviewed in [7]).

In addition to classical allosteric regulation, the activity of enzymes and other proteins is also modulated by other chemical modifications. This section summarizes some of the recent findings on this regulation.

2.2.1. Regulation through post-translational modifications

Protein phosphorylation plays a key role in regulating glucose metabolism, such as controlling glycolytic and oxidative flux through pyruvate kinase (PK) M2 [8]. In addition to phosphorylation, glycolytic enzymes are subject to other post-translational modifications, such as methylation, addition of N-acetylglucosamine (*O*-GlcNAcylation), and acetylation that modulate their metabolic activities.

Methylation of PKM2 mediated by CARM1

Coactivator-associated arginine methyltransferase 1 (CARM1) transfers a methyl group to arginine residues in histones, transcription factors and co-regulators. The reversible methylation of PKM2 by CARM1 has been reported to heighten its activity and shift metabolism towards aerobic glycolysis, thus promoting tumorigenesis [9]. CARM1 also suppresses oxidative phosphorylation (OXPHOS) possibly by altering carbon fluxes into the mitochondria or by reducing mitochondrial influx of calcium from the endoplasmic reticulum, a key process for OXPHOS activity [10]. For cell culture bioprocessing, Chinese hamster ovary (CHO) cells are most commonly used to develop production cell lines. CARM1 is expressed at high levels in CHO cells [11].

O-GlcNAcylation of PKM2, PFK, and G6PD

O-GlcNAcylation, a reversible addition of N-acetylglucosamine (GlcNAc) to a serine/threonine residue, plays a key role in transcriptional, epigenetic, and post-translational regulation [12]. The level of *O*-GlcNAcylation is a balance between *O*-GlcNAc transferase and *O*-β-N-acetylglucosaminidase, and is affected by the availability of the substrate, UDP-GlcNAc. It has been proposed that *O*-GlcNAc serves as a nutrient sensor because the synthesis of UDP-GlcNAc requires ATP, uridine, glucose, glutamine, and acetyl-CoA [12]. *O*-GlcNAcylation regulates activities of PKM2 [13], phosphofructokinase (PFK) [14], and glucose-6-phosphate dehydrogenase (G6PD) [15] (Figure 2.2). *O*-GlcNAcylation of PKM2 destabilizes its active tetrameric form, suppresses its activity and promotes its nuclear translocation [13]. Suppression of PKM2 activity results in accumulation of upper glycolytic intermediates [3]. *O*-GlcNAcylation of PFK suppresses its activity by abolishing allosteric activation by fructose 2,6-bisphosphate (F26BP) and ATP, thus suppressing glycolysis and channeling additional flux towards the PPP [14]. *O*-GlcNAcylation of G6PD increases its activity, which is the rate-limiting enzyme leading to the PPP [15]. Simultaneous *O*-GlcNAcylation of PKM2, PFK, and G6PD results in increased PPP flux, thus enhancing nucleotide and NADPH synthesis.

2.2.2. Nonspecific metabolite modifications

Non-enzymatic chemical modifications by metabolites

In addition to enzyme-mediated methylation and *O*-GlcNAcylation, the activity of enzymes can also be modified by intermediates of energy metabolism. The glycolytic intermediate 1,3-bisphosphoglycerate (13BPG) was found to participate in a spontaneous reaction to form 3-phophoglyceryl-lysine residues on some proteins, including near to the active site of glycolytic enzymes such as aldolase (ALDO) A/B, triose-phosphate isomerase (TPI), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), enolase (ENO) 1, phosphoglycerate mutase (PGAM) 1, and PKM2 (Figure 2.2) [16]. This covalent modification reduces the activity of these enzymes, and therefore affects the glycolysis flux and diversion of its intermediates for anabolism. Interestingly, the level of this modification correlates with extracellular glucose levels (and intracellular 13BPG), which highlights the potential of 13BPG as a flux-modulating signal.

Non-specific reactions by metabolic enzymes modulate fluxes

Some enzymes in energy metabolism exhibit substrate non-specificity (‘sloppiness’) [17]. They can convert metabolites that are structurally similar to their specific substrates to byproducts. In some cases the byproduct regulates the activity of other enzymes. GAPDH can catalyze a side reaction that converts erythrose-4-phosphate, an intermediate in the PPP, to 1,4-bisphosphoerythronate, which is subsequently converted to 4-phosphoerythronate by phosphoglycerate kinase (PGK). 4-phosphoerythronate inhibits 6-phosphogluconate dehydrogenase (6PGDH) and thus reduces flux through the PPP (Figure 2.2). In similar fashion to GAPDH, PKM2, which converts phosphoenolpyruvate (PEP) to pyruvate, can also catalyze phosphorylation of lactate to generate 2-phospho-L-lactate (2-P-lac) [17]. 2-P-lac inhibits the kinase activity of the bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFKFB), leading to reduced levels of F26BP, the allosteric activator of PFK, resulting in reduced

flux through glycolysis (Figure 2.2). It was found that phosphoglycolate phosphatase (PGP) catalyzes the dephosphorylation of both 4-phosphoerythronate and 2-P-lac, and thus reduces their accumulation and negative impact on cellular metabolism [17]. PGP is expressed in CHO cells [11].

2-Hydroxyglutarate from multiple reactions regulates metabolism

Another metabolite, L-2-hydroxyglutarate (L-2HG) is generated from α-ketoglutarate (αKG) by a non-specific reaction catalyzed by lactate dehydrogenase (LDH) or malate dehydrogenase (MDH) 1/2, an activity that is increased under acidic conditions [18]. D-2HG, an enantiomer of L-2HG, is produced by alcohol dehydrogenase iron-containing 1 (ADHFE1) [19]. Overexpression of ADHFE1 leads to increased levels of metabolites including D-2HG and intermediates in the TCA cycle and glycolysis [19]. In addition, D-2HG can also be produced from αKG by mutated isocitrate dehydrogenase enzymes [cytosolic isocitrate dehydrogenase (IDH1) and mitochondrial IDH2], as seen in some cancers (reviewed in [20]).

In addition to affecting glucose metabolism, both L-2HG and D-2HG interact with and inhibit αKG-dependent dioxygenase enzymes, including DNA methyl cytosine dioxygenases, histone demethylase, and prolyl hydroxylase domain-containing (PHD) enzymes [20]. Under normoxic conditions, PHD hydroxylates hypoxia-inducible factor (HIF)-1α, leading to its ubiquitination and degradation. The suppression of PHD activity stabilizes HIF-1α and enhances glucose metabolism. All the above-mentioned enzymes are expressed in CHO cells and 2HG has been detected in CHO cell cultures, suggesting that glucose metabolism in CHO cells may be modulated by 2HG [11].

Unlike, 4-phosphoerythronate and 2-P-lac, which are produced in the cytosol and act on a cytosolic enzyme, 2HG is produced either in the cytosol by LDH, MDH1, IDH1, or ADHFE1, or in mitochondria by MDH2 or the mutant form of IDH2. HIF-1α exerts its regulatory effects in the nucleus. 2HG formed in mitochondria is likely exported by the mitochondrial citrate transporter

(SLC25A1). Importantly, metabolites produced in different cellular compartments rely on transporters that are necessary for communication and for exerting regulatory effects across compartments.

2.3. Communication between mitochondrial and cytosolic compartments

2.3.1. Material flow across the cytosol and mitochondria to maintain redox homeostasis

The cytosolic pyruvate formed from glucose through glycolysis is primarily divided into two fluxes that either enter the mitochondria through mitochondrial pyruvate carrier (MPC) or generate lactate through LDH (**Figure 2.3**). The oxidation of one pyruvate in the mitochondria generates three CO_2, six net reducing equivalents (four NADH and one $FADH_2$ in mitochondria, and one NADH from glycolysis in the cytosol), and consumes three O_2. The flux of pyruvate through MPC must be accompanied by an approximately equal molar flux of the reducing equivalent into mitochondria through the malate/aspartate shuttle (MAS) to maintain the NAD^+/NADH balance in glycolysis. The LDH reaction may occur in both directions, generating or consuming lactate and NAD^+. The MAS consists of a set of reactions that take up cytosolic NADH by regenerating it to NAD^+ and transferring the captured reducing equivalents to the mitochondrial NAD^+ pool ultimately producing NADH. In an idealized situation the MAS does not incur net gain or loss of each of its constituents [αKG, malate, oxaloacetate (OAA), aspartate, and glutamate]. In reality, these compounds are each part of a larger pool (Figure 2.1) and the concentration of each compound in each compartment is constrained by its generation and consumption by other biochemical reactions. Ultimately, the NAD^+/NADH ratio in the mitochondria and cytosol will affect the fluxes involved in redox homeostasis profoundly. The MAS is the dominant mechanism of NADH

shuttling [21], although an alternative reducing equivalent transfer pathway, the glycerol-3-phosphate shuttle, may also contribute.

Note that the glycolytically generated NADH eventually must contribute its reducing equivalents to an electron acceptor, either oxygen in OXPHOS or an organic compound. In the latter case, this takes place via reduction and excretion of the metabolite (e.g. lactate), out of the cell. It has been reported that the cytosolic conversion of NADH to NAD^+ can also be facilitated by reductive carboxylation. Under some conditions of impaired mitochondrial oxidation of NADH, glutamine is converted to αKG in the cytosol and then to citrate through IDH1 by reductive carboxylation, consuming a CO_2 molecule and NADPH. Citrate is then split to become OAA and acetyl-CoA. OAA is then converted to malate by oxidizing NADH through cytosolic MDH1 [22]. In this case malate or another compound can be expected to be produced in stoichiometric amounts to balance the reducing equivalents.

Cytosolic concentrations of NADH and NAD^+ are determined by the balance of the generation of NADH from glycolysis and, the regeneration of NAD^+ by LDH and MAS (Figure 2.3). The mitochondrial levels of NADH and NAD^+ are mainly governed by NADH generation by the TCA cycle, the action of MAS, and the oxidation of NADH to NAD^+ by OXPHOS. It is not well understood how changes in NADH and NAD^+ concentrations in one compartment impact on the redox state and key redox-dependent fluxes in other compartments. Some protein-protein interactions also lead to substrate channeling, altering the expected fluxes and potentially favoring some metabolic routes over others.

On perturbing the NADH levels separately in the cytosol and mitochondria, it was found that only the mitochondrial NADH perturbation impacted the total NAD^+/NADH ratio [23]. The authors further suggested that perturbing the NAD^+/NADH ratio in the mitochondria changes the cytosolic NAD^+/NADH ratio, but not the reverse. Interestingly, lowering cytosolic or mitochondrial NADH

levles did not affect the rates of glucose consumption or lactate production, although it resulted in increased secretion of pyruvate and aspartate.

2.3.2. Localization of enzymes of energy metabolism

Total protein concentration in the cytosol and mitochondria are very high. Many enzymes may not be homogeneously distributed. Some form complexes with one another through protein-protein interactions. These associations create channels to allow the product of one enzyme that is also the substrate of the associated enzyme to translocate preferentially between two consecutive reactions in the metabolic pathway. Notable examples are the MDH2/CS complex that transfers OAA to form citrate in the first reaction of the TCA cycle [24], and the channeling of NADH from GAPDH to LDH in the GAPDH/LDH complex [25].

In addition to protein-protein association, glycolytic enzymes may also be localized in cellular compartments. Almost all enzymes of glycolysis (at least some isoforms) and several TCA cycle enzymes have been reported to translocate to the nucleus, although the mechanism of nuclear translocation and the function of many such translocations are not known (reviewed in [26]). In some cases, nuclear localization serves a key biochemical role. The pyruvate dehydrogenase complex that converts pyruvate to acetyl-CoA in mitochondria translocates to the nucleus where acetyl-CoA serves as the substrate for histone acetylation. Aldolase A was shown to translocate to the nucleus where it was shown to play a noncatalytic role [27]. The possible nonenzymatic role was corroborated by the observation that yeast aldolase is associated with RNA polymerase III in the nucleus [28]. GAPDH is among the most highly expressed cellular proteins. It has been reported to localize to the mitochondria where it plays a role in apoptosis [29]. In addition, GAPDH translocates to nucleus under hyperglycemia and plays a role in cell cycle progression [30]. PKM2, the only PK reported to translocate to the nucleus, uses its substrate PEP to act as a protein kinase,

and has diverse substrates including AKTS1 whose phosphorylation subsequently activates mTORC1 [31].

2.4. Material flow across the cytosol and mitochondria to maintain carbon balance

During cell growth, cataplerotic reactions direct TCA cycle intermediates to other pathways for the biosynthesis of various cellular precursors. The carbon deficit caused by this diversion is replenished by anaplerotic reactions that infuse carbon from the cytosol to maintain the activity of the TCA cycle. Cataplerosis, anaplerosis, and the MAS share many common intermediates that are present in multiple compartments. These intermediates shuttle between the mitochondria and the cytosol through the mitochondrial membrane at high fluxes. Their intercompartmental transport is facilitated by mitochondrial solute carrier proteins (SLC25 family, http://slc.bioparadigms.org/). Homeostasis between cataplerosis and anaplerosis is thus crucial because an imbalance could either deplete or flood the TCA cycle of its intermediates, thereby perturbing OXPHOS.

Cataplerosis

A major cataplerotic flux is the export of citrate from the mitochondria to the cytosol for conversion to acetyl-CoA through the citrate-malate antiporter (SLC25A1). In the cytosol, citrate is converted by ATP citrate lyase (ACLY) to acetyl-CoA and OAA to supply acetyl-CoA for lipid synthesis and protein acetylation. Mitochondrial OAA, αKG, and pyruvate are diverted for biosynthesis of aspartate, glutamate and alanine, respectively (reviewed in [32]). Another efflux route from the TCA cycle is via conversion of OAA to PEP by the action of phosphoenolpyruvate carboxykinase (PCK) 2. PEP is transported to the cytosol by the citrate-malate transporter (SLC25A1) by exchange of malate [33]. In the cytosol, PEP supplements the intermediates of the lower half of glycolysis that provides carbon for the synthesis of biomass precursors including serine [34] and glycerol 3-phosphate.

Anaplerosis

For cultured cells, most of the anaplerotic carbon comes from glutamine and asparagine via αKG and OAA (Figure 2.3). Glutamine and asparagine are first converted to glutamate and aspartate by glutaminase and asparaginase, respectively. Glutamate then undergoes a transaminase or dehydrogenase reaction to generate αKG that enters the TCA cycle, whereas aspartate enters as OAA through a transaminase reaction.

Deamination of glutamine to glutamate can occur by the action of the mitochondrial/cytosolic glutaminase *gls*, or by the mitochondria-only isoform *gls2* [35]. Interestingly, although GLS is found in the mitochondria, the mitochondrial transporter for glutamine has not been identified. Nonetheless, the presence of GLS in the mitochondria suggests that glutamine passes the mitochondrial membrane. Myc, an oncoprotein that induces growth, has been shown to induce *gls* gene expression [36]. The *gls* encoded mitochondrial localized glutaminase C (GAC) isoform has been implicated in cancer [37], and the expression of *gls2* has been reported to be induced by tumor-suppressor protein, p53 [38, 39]. CHO cells predominantly express *gls*.

Glutamate is converted to αKG by glutamate dehydrogenase (GLDH) (using isoforms GLUD1/2, which also generate ammonium and NADH) or by one of many transaminases, including glutamate-oxaloacetate transaminase (GOT), glutamate-pyruvate transaminase (GPT), or phosphoserine transaminase (PSAT) that transfer the amino group of glutamate to OAA, pyruvate, or phosphopyruvate to form aspartate, alanine, or phosphoserine, respectively. The glutamine (or glutamate) anaplerotic flux is tightly regulated by TCA cycle activity [40]. It has been reported that rapidly proliferating cells preferentially use transaminases that couple glutamine anaplerosis to the synthesis of non-essential amino acids. By contrast, quiescent cells favor the use of the mitochondrial form of GLUD1 [41].

Because glutamate, aspartate, OAA, and αKG are involved in the MAS as well as in anaplerosis, the localization of deamination reactions can affect the concentration gradient across the mitochondrial membrane and MAS redox transfer activity (Figure 2.3). GLUD1 and GLUD2 are both localized to the mitochondria [42], whereas transaminases have both cytosolic (GOT1, GPT1 and PSAT1) and/or mitochondrial (GOT2 and GPT2) forms. CHO cells express GOT1, GOT2, GPT2, PSAT1 and GLUD1 [11].

Asparagine is converted to aspartate (and ammonium) by cytoplasmic asparaginase. Aspartate is transported into the mitochondria and is converted to OAA by the action of a transaminase. In mammalian systems, asparaginase is encoded by two genes, *aspg* and *asrgl1* [43]. Both genes are expressed at a moderate level in CHO cells [11]. When glutamine is depleted or when cells are cultured in glutamine free environments (as in the case of CHO cells using the glutamine synthetase expression system), asparagine becomes an essential amino acid [43].

Glutamate and aspartate cross the mitochondrial membrane through the antiporters SLC25A12 or SLC25A13. In addition, glutamate carriers GC1 (also known as SLC25A22) or GC2 (also known as SLC25A18) are symporters that co-transport glutamate and a proton into the mitochondria [44]. The pH gradient across the mitochondrial membrane favors influx of glutamate.

There are other fluxes that replenish carbon in the TCA cycle which do not involve amino acids. For example, when OAA availability in mitochondria is insufficient to react with pyruvate imported from the cytosol to form citrate, a portion of pyruvate is diverted to form OAA by the action of mitochondrial pyruvate carboxylase (PC). Alternatively, under conditions of glucose depletion and low cytosolic pyruvate supply, cells synthesize pyruvate from anaplerotically generated malate through the action of mitochondrial malic enzyme [40]. These fluxes have been reported to range from zero to a significant proportion of total carbon input into the TCA cycle [45].

2.5. Crosstalk between cell signaling and metabolism

Regulation via allosteric and chemical modifications, as well as through inter-compartmental flow of material, as described above, generally occurs over short timespans to maintain metabolic network homeostasis at the metabolite and flux level. By contrast, some metabolites (e.g., 2HG) exert a broader-range and longer timescale effects through modifications of non-metabolic proteins (e.g., PHD). Metabolic enzymes can also translocate in a short timescale to the nucleus and other intracellular locales where they play a metabolic role or participate in signaling or transcriptional regulations (**Box 2**). By comparison, cell signaling regulatory networks often act on longer timescale via transcription and often link glucose metabolism to global processes such as energy production, amino acid metabolism, and growth. However, cell signaling can also result in protein phosphorylation that elicits cellular responses on a short timescale.

The major signaling regulators that integrate metabolism with other physiological cues are shown in Figure 2.4. AKT and mammalian target of rapamycin (mTOR) regulate glucose metabolism in response to growth factor stimulation and nutrient availability [46, 47]. AMP-activated protein kinase (AMPK) responds to changes in the energy state of the cell by inhibiting anabolism and promoting catabolism. Thioredoxin interacting protein (TXNIP) relays environmental and other cellular signals to modulate glucose metabolism. c-Myc, a proto-oncogene which frequently escapes regulation in cancer, acts as a transcription factor to induce the expression of genes encoding enzymes across multiple pathways including glycolysis, the TCA cycle, and glutamine anaplerosis [48]. HIF-1α regulates cellular metabolism in response to hypoxic conditions and other physiological cues, and promotes glycolytic metabolism [49]. It has recently become clear that signaling regulation is bidirectional in that signaling pathways not only modulate metabolism, but also respond to the metabolic state of the cell. This crosstalk helps cells to establish homeostasis in response to changing cellular needs and environmental signals.

2.5.1. mTOR regulation of metabolic state and regulation by amino acids

mTOR is a serine/threonine protein kinase that forms the complexes mTORC1 and mTORC2 (reviewed in [50]). Broadly, mTORC1 is activated by a sufficient supply of nutrients, amino acids, and growth factors. Activated mTORC1 stimulates cell growth by inducing lipid, nucleotide, and protein synthesis. mTORC1 also enhances glucose metabolism and represses autophagy (Figure 2.4). mTORC1 stimulates purine synthesis through the tetrahydrofolate pathway, upregulates *de novo* pyrimidine synthesis (reviewed in [51]), and regulates rRNA synthesis, thus linking anabolism and cell growth [52]. Through insulin/insulin-like growth factor stimulation, mTORC2 regulates cell proliferation and survival by phosphorylation of AKT [53]. Activated AKT stimulates glycolysis by regulating multiple glycolytic enzymes, as well as lipid metabolism by activating ACLY [54], details of which have been reviewed previously [2].

mTORC1 activation takes place in response to environmental signals and nutrients. The availability of amino acids, including leucine, glutamine [55], arginine [56], and methionine, promotes translocation of mTORC1 from the cytoplasm to lysosomes where it is activated by insulin/growth factor stimulation [51].

2.5.2. AMPK regulation of energy metabolism

AMPK is a protein kinase that serves as a sensor of cellular energy status and regulates cellular metabolism – typically by inhibiting anabolism and stimulating catabolism under energy stress (reviewed in [57]). In addition to sensing energy status, AMPK also senses glucose depletion independent of energy status by monitoring the levels of the non-F16BP (fructose 1,6-bisphosphate)-bound form of the glycolytic enzyme, aldolase, that reflects glucose concentration and intracellular flux [58]. F16BP-unbound aldolase promotes the formation of a lysosomal complex leading to phosphorylation and activation of AMPK.

AMPK activation stimulates catabolic pathways such as the uptake and metabolism of glucose, fatty acid oxidation, and autophagy. Concurrently, AMPK activation suppresses anabolic pathways including protein, fatty acid, and rRNA synthesis [59]. AMPK activation increases glucose uptake by suppressing of TXNIP protein levels [60] as discussed below. AMPK was recently shown to also act more broadly on the epigenetic state of cells, affecting both histone acetylation and DNA methylation (reviewed in [61]).

From a metabolic regulation perspective, AMPK senses cellular energy status and glucose availability, whereas mTORC1 senses the availability of amino acids and glucose, and activation of both takes place on the lysosome membrane. Interestingly, AMPK can abolish mTORC1 activation through direct phosphorylation of a component of mTORC1 or by activating an upstream inhibitor of mTORC1 (AMPK is reviewed in [62]).

2.5.3. MondoA:Mlx-TXNIP-GLUT1 regulation of glucose uptake

TXNIP (also known as VDUP1), was first identified as a negative regulator of the function of thioredoxin (TRX) as a reducing agent [63]. TXNIP was later identified as a negative regulator of glucose uptake into cells, which is independent of its role in moderating TRX activity [64]. Oddly, however, TXNIP transcript expression hinges upon glucose availability. The presence of glucose increases cytosolic glucose 6-phosphate (G6P) levels, which promotes nuclear translocation of the MondoA:Mlx transcription factor complex, where it binds to the upstream promoter of TXNIP and induces its expression [65]. Induction of TXNIP suppresses glucose uptake by concurrently reducing the transcript level and promoting the endocytosis of GLUT1 (also known as SLC2A1) [66]. TXNIP and glucose transport thus form a negative feedback loop on cellular glucose uptake rate (Figure 2.4).

TXNIP integrates other environmental and cell signaling cues to regulate glucose metabolism (Figure 2.4). Lactic acidosis induces TXNIP expression [67]. An acidic environment in the cytosol

enhances mitochondrial ATP generation, which in turn intensifies G6P generation from mitochondria-bound hexokinase (HK) triggering TXNIP expression [68]. Both lactic acid accumulation and an acidic cytosol thus suppress glucose uptake. TXNIP is phosphorylated under low-energy conditions by AMPK [66], and in response to insulin stimulation by AKT [69], leading to its degradation and inhibition of endocytosis of the glucose transporters GLUT1 and GLUT4 (also known as SLC2A4), thus enhancing glucose uptake. Further, activation of mTOR has been shown to reduce TXNIP expression by preventing the dimerization of the MondoA:Mlx complex [70]. c-Myc has also been shown to repress TXNIP [71].

2.5.4. Metabolites as signaling molecules

The above discussed modulation of MondoA:Mlx-TXNIP-mediated metabolic regulation by G6P and lactate is an example of how metabolites can affect signaling and the reciprocal regulation of metabolism by signaling. Lactate also acts as a signaling molecule by binding to N-myc downstream regulated gene 3 (NDRG3) protein [72]. Under normal conditions, NDRG3 is degraded via PHD2 driven hydroxylation and subsequent ubiquitylation. Lactate, when present in sufficient quantities (5-50 mM), binds to and stabilizes NDRG3 and prevents its degradation, leading to activation of Raf-ERK signaling and stimulation of cell growth and angiogenesis. NDRG3 is expressed in CHO and HEK293 cell lines (transcriptome data for HEK293 are available at https://www.proteinatlas.org/) [11, 73]. Lactate routinely accumulates to high levels in cell culture bioprocesses that may also experience hypoxic environments to varying degrees in large-scale bioreactors, and NDGR3 signaling could come into play under such conditions.

2.5.5. Signaling through extracellular metabolites

The cell signaling regulators of metabolism discussed above translate intracellular into regulatory actions. It was recently shown that metabolites which accumulate extracellularly participate in autocrine/paracrine signaling [74]. Lactate may act as one such signal through its

binding to G protein-coupled receptor (GPR81) in adipocytes [75, 76]. Binding of lactate to GPR81 inhibits adenylyl cyclase activity in a G_i-dependent fashion, resulting in suppression of cAMP levels, which further suppresses lipolysis activities and increases glycolysis activity (reviewed in [74]). GPR81 is expressed in multiple cancer cell lines and, specifically, its expression has been shown to play a key role in the survival of pancreatic cancer cell lines under low-glucose and high-lactate conditions [77]. GPR81 displays little to no expression in CHO or HEK293 cells [11, 73]. There are other GPRs expressed in mammalian cells that respond to different types of extracellular metabolites including butyrate, succinate, free fatty acids, and amino acids [74]. It is not clear whether these GPRs play a role in cell culture processes.

2.6. Engineering energy metabolism

Cultured cells produce high levels of metabolites, especially lactate and ammonium. For cell culture processes in biomanufacturing, the high level of accumulated metabolites pose a challenge because of their negative effect on cell growth and productivity. The list of undesirable metabolites extends to intermediates of TCA cycle and the amino acid catabolic pathways [78]. The root cause of such excessive metabolite accumulation has been attributed to the "wasteful" aerobic glycolysis. Various approaches have been taken to reduce hexose consumption and lactate production to enhance growth and productivity.

Process level changes include supplying cells with a sugar (such as galactose) that is only taken up by cells slowly, thus reducing the glycolytic flux and lactate production [79]. Reduced glucose consumption and lactate production have also been achieved by controlled feeding of glucose to maintain it at low levels in the medium, as monitored by oxygen consumption, pH control, and Raman spectroscopy [80, 81].

Metabolic engineering has also been utilized. Overexpression of GLUT5 (also known as SLC2A5) enabled cells to slowly consume fructose while growing at a normal rate [82].

Knockdown of LDHA failed to generate cell lines that displayed no lactate production but grew at a normal rate [83, 84]. Simultaneous LDH and pyruvate dehydrogenase kinase (PDK) knockdown [85], and knockout of LDH in conjunction with PDK knockdown were found to retard growth or were lethal in CHO cells [86]. Different means of controlling lactate production have been reviewed recently [87].

Overexpression of the yeast form of pyruvate carboxylase (PYC2, mammals lack the cytosolic form of this enzyme) allowed simultaneous diversion of pyruvate and the reducing equivalents generated by glycolysis into the mitochondria. Pyruvate is first converted to OAA which is reduced to malate in the cytosol by MDH1 and then transported to the mitochondria for entry to the TCA cycle. Overexpression of PYC2 alone [88] or in conjunction with MDH2 [89] has been demonstrated to alter cell growth and metabolic characteristics, including some reduction of lactate production during batch culture.

Most genetic manipulation work to reduce glycolytic flux has employed constitutive promoters. Interventions that succeeded in changing cell metabolism typically also resulted in reduced growth. However, cells appear to be more amenable to a low-flux metabolic state when the growth rate is lower. Considering that the Warburg effect is an inherent property of virtually all rapidly proliferating cells, the use of dynamic or inducible gene expression to increase target gene transcription when the growth rate is decreased warrants exploration for engineering cell metabolism.

2.7. Concluding remarks

The metabolic state of cells in biomanufacturing has a profound effect on product quality and productivity. Better understanding of metabolic regulation will facilitate better control of the process. In this article we have highlighted recent findings on the regulation of glucose metabolism in cultured cells from a perspective of homeostasis among pathways and between compartments.

Cellular metabolism has evolved to respond to an immense number of environmental conditions through multidimensional regulation networks. The regulatory responses observed in experimental studies are inevitably context-dependent; nevertheless, they can provide important clues. A key question in such a review is always the relevance of the findings to cell culture bioprocesses (see Outstanding Questions).

Transcriptomic and genomic data for cell lines of interest are now readily attainable. Targeted proteomics has advanced to allow simultaneous quantitative assessment of variations in enzymes, isozymes, and transporters at the protein level to complement the RNA-seq data (targeted proteomics reviewed in [90]). Metabolites that were recently found to influence metabolism can be assessed by targeted metabolomics to verify their role in cell culture bioprocesses [91]. Isotopic carbon labeling of molecules such as glucose, glutamine, and asparagine will facilitate metabolic flux calculations (reviewed in [92-94]). Advances are also being made to examine metabolite levels in cytosolic and mitochondrial compartments [95, 96]. Deeper exploration using an integrated multi-omic approach can identify the regulatory mechanisms that are most relevant to industrial cell lines and processes. The findings of the past few years have also illustrated the complexity and high dimensionality of glucose metabolism in mammalian cells. To gain deeper physiological insights, multiomic information needs to be integrated into a mechanistic kinetic model. Such a systems approach will be necessary to better understand the regulation of metabolic homeostasis and to control cellular metabolism in bioprocesses.

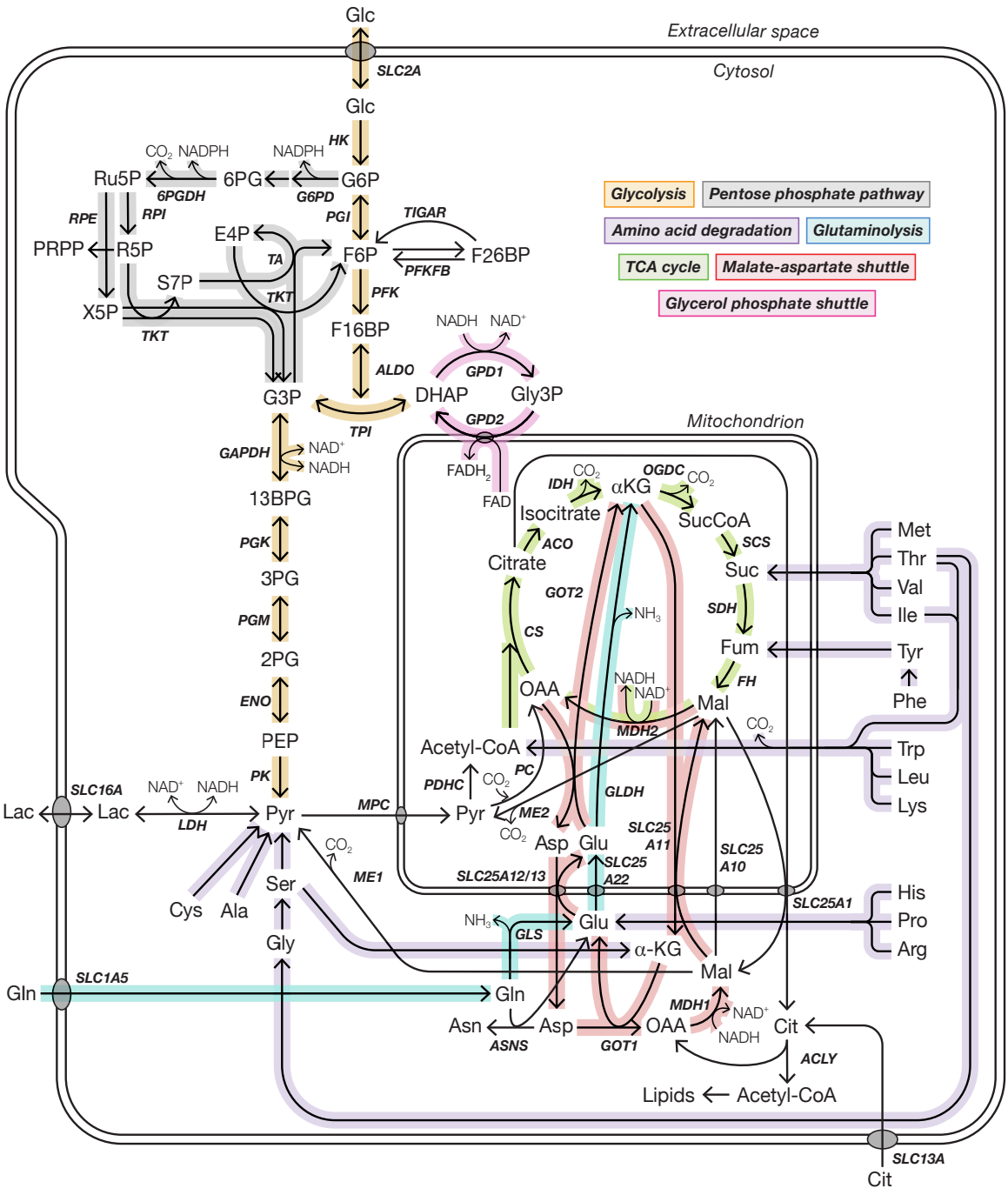


Figure 2.1: Pathways in central metabolism.

The key metabolic pathways discussed are highlighted in color. Imported glucose is converted through glycolysis to pyruvate and generates two NADH and two ATP. This NADH must be regen-

erated to NAD+ to drive glycolysis. Both pyruvate and NADH are split into two streams, one forms lactate and NAD+ at 1:1 molar ratio, the other pyruvate stream enters the mitochondria through the MPC transporter, accompanied by NADH regeneration to NAD+ via MAS and the G3PS. The LDH reaction, lactate excretion, and the flux through MAS/G3PS regenerates NAD+ to allow glycolysis to continue. The flux of pyruvate imported into the mitochondria and that of combined MAS/G3PS is also equal molar. The PPP branches from glycolysis and serves to produce NADPH and supply ribose for downstream biosynthetic pathways. TKT, TA, and isomerase reactions allow two F6P and one G3P to convert to three Ru5P to balance the material flow of PPP. Pyruvate imported into the mitochondria can be used to fuel the TCA cycle, driving the production of large quantities of ATP. Intermediates in the TCA cycle can be used to supply other pathways, referred to as cataplerotic reactions. To sustain the TCA cycle, anaplerotic flux replenishes TCA cycle intermediates chiefly from the breakdown of glutamine or other amino acids, β-oxidation of fatty acids, or alternative metabolic routes of pyruvate metabolism. Because metabolism is compartmentalized, a reaction may take place in the cytosol, mitochondria, or both, depending on the expression of localized enzymes and transporters.

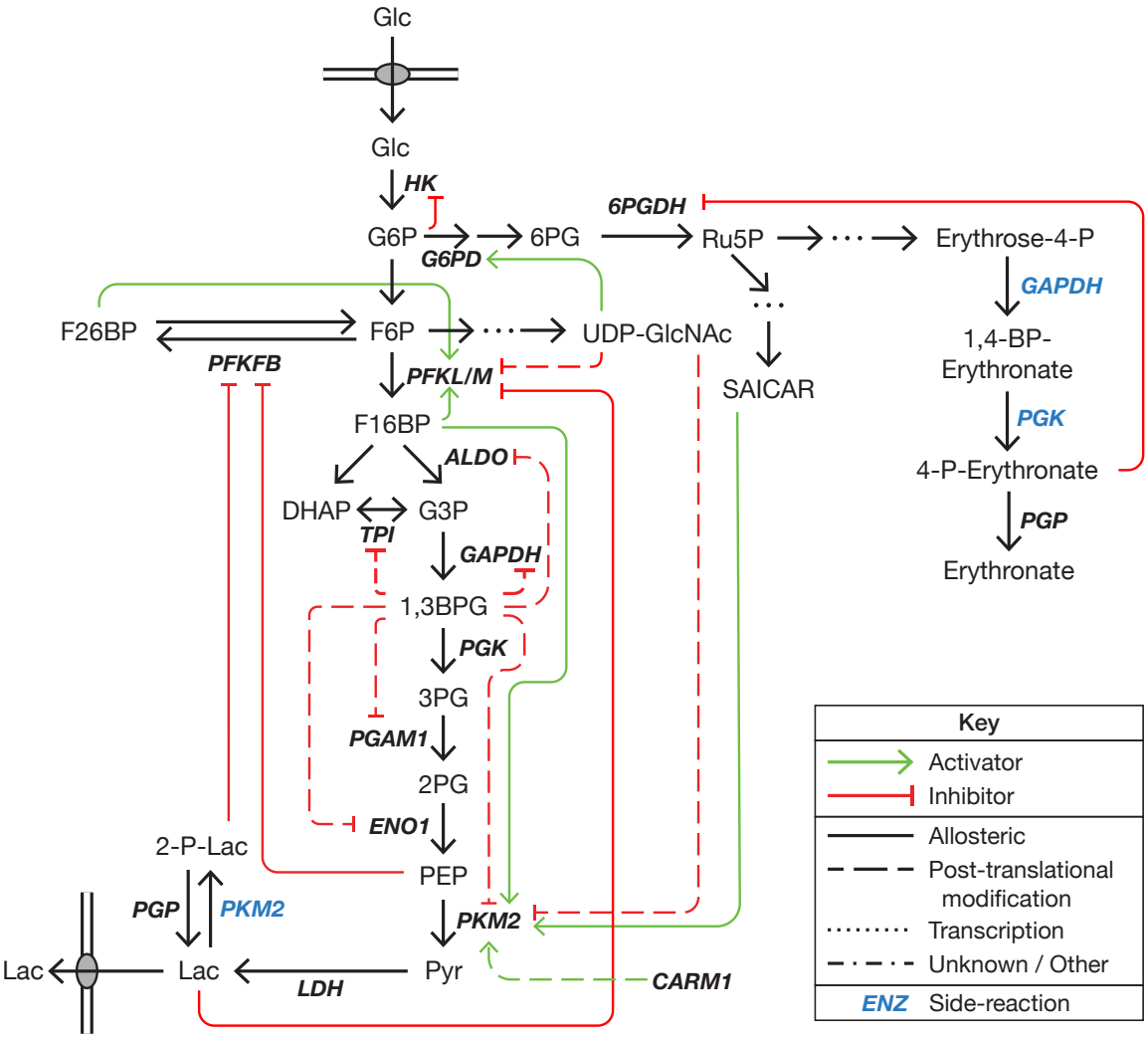


Figure 2.2: Regulation of metabolism through allostery and post-translational modifications.

The regulation of glycolysis flux by allosteric actions of metabolites centers on HK, PFK, PFKFB, and PK. Shown in the figure are the isozymes commonly expressed in fast growing cells. HK is inhibited by its product, G6P. PFK is activated by F26BP as well as by its product F16BP, and some isoforms are inhibited by lactate. The bifunctional enzyme PFKFB is notably inhibited by PEP. PKM2 is regulated by numerous products of glycolysis and branching pathways including F16BP, SAICAR, and serine (not shown). Post-translational modification of HK, PFK, and PKM2 through O-GlcNAcylation inhibits these enzymes, enabling UDP-GlcNAc, an indicator of nutrient availability to regulate glycolysis flux. 13BPG modifies several enzymes non-specifically and affects their

activity. Side reactions of metabolic enzymes can generate metabolites that modulate enzyme activity, including the inhibition of PFKFB by 2-P-lac produced by PKM2 and the inhibition of 6PGDH by 4-P-erythronate produced by PGK, which regulates the flux through the PPP. Enzymes catalyzing non-specific or side reactions other than their primary products are marked in blue.

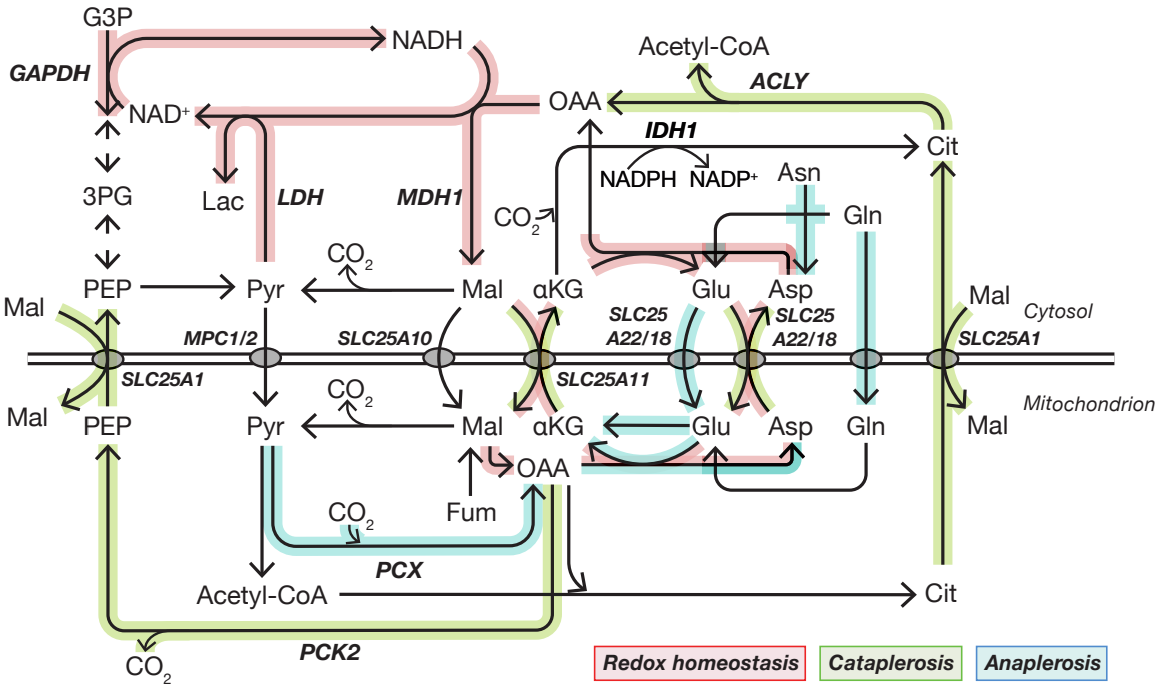


Figure 2.3: Material exchange and redox balance between the cytosol and mitochondria.

Metabolites are exchanged across the mitochondrial membrane for both energy supply and biosynthetic purposes. To maintain the redox balance, glycolytic flux must be balanced by an equimolar regeneration of NAD+ through a redox transfer shuttle, or by transferring the potential to another compound such as lactate or malate (through reductive carboxylation). The TCA cycle likewise must be sustained through anaplerotic reactions because metabolites are siphoned off towards other biosynthetic pathways through cataperotic reactions. Reactions important for maintaining redox balance in the cytosol, anaplerosis, and cataplerosis are highlighted.

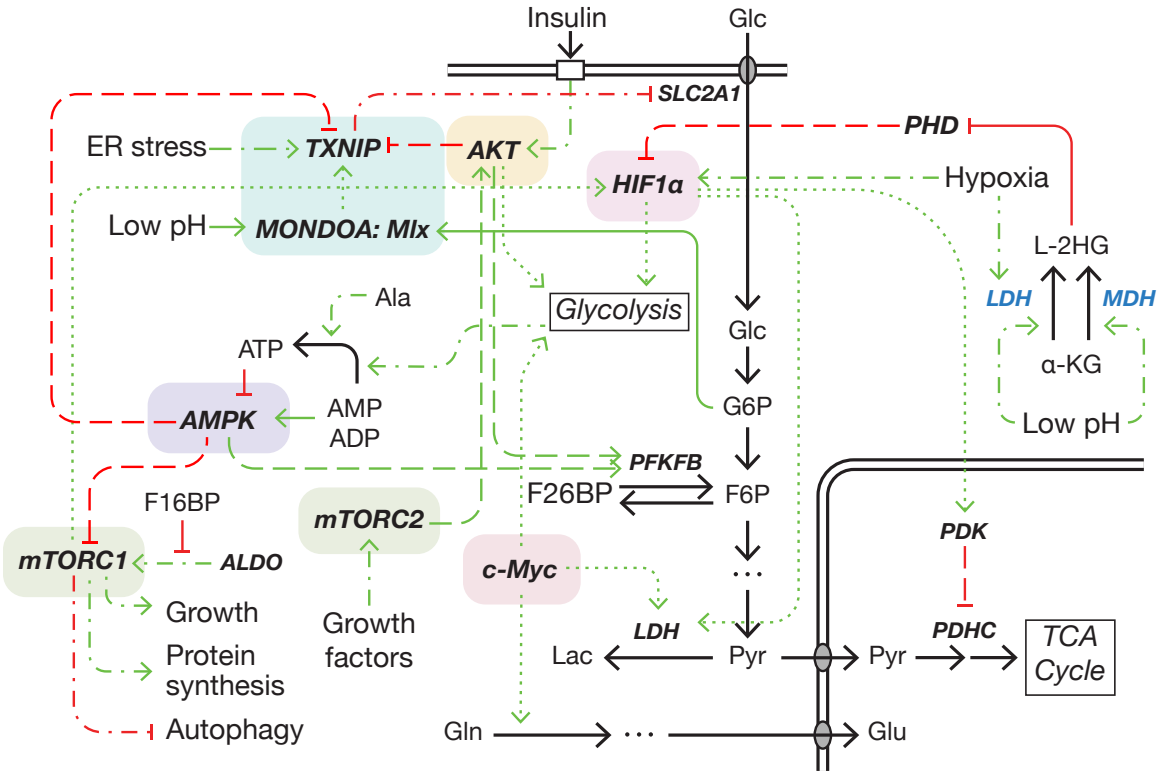


Figure 2.4: Interaction of cell signaling and metabolism.

Many cellular signaling networks both regulate, and are regulated by, central carbon metabolism. The figure shows some of the simplified interactions between AMPK, AKT, c-Myc, HIF1α, mTORC1/2, TXNIP/MondoA:M1x, and central metabolism. AMPK is regulated by the cellular energy state through the relative abundances of AMP, ADP, and ATP. It increases the kinase activity of PFKFB and inhibits mTORC1 under low-glucose conditions. mTORC1 is also regulated by non-F16BP-bound aldolase, activating downstream pathways related to growth, protein synthesis, and HIF-1α. HIF-1α also responds to environmental conditions such as stress and low pH through binding to 2HG, a side product of several metabolic enzymes under adverse conditions, and which subsequently alters the oxidative state of cells by transcriptionally promoting lactate production and inhibiting pyruvate metabolism in the mitochondria through PDK. MondoA:M1x responds to metabolic signals as well as to environmental conditions such as endoplasmic reticulum stress and low pH, subsequently activating TXNIP and inhibiting the translocation of SLC2A1 to the cell membrane.

3. Kinetic model optimization and its application to mitigating the Warburg effect through multiple enzyme alterations

Reproduced with permission from O'Brien, C., Allman, A., Daoutidis, P., & Hu, W. S. (2019). Kinetic model optimization and its application to mitigating the Warburg effect through multiple enzyme alterations. *Metabolic engineering*, *56*, 154-164.

3.1. Introduction

Engineering of metabolic pathways has been practiced for decades. Mutagenesis and selection practiced since the mid-twenty century has contributed greatly to the rise of the biochemical industry that produces various amino acids, organic acids, and natural products. More recently recombinant DNA coupled with knowledge-based pathway design and controlling-step relaxation has allowed humans to harness the tremendous metabolic capabilities in living systems. However, the exploitation is still largely limited to microbial organisms. Metabolic alterations in mammalian cells still face large hurdles because of the complexity of metabolic regulation.

Glucose is metabolized through the central metabolic pathways of glycolysis, the pentose phosphate pathway (PPP), and the tricarboxylic acid (TCA) cycle. In addition to supplying energy, these pathways also provide biosynthetic precursors for cell growth. These pathways are highly conserved in living systems and subject to extensive regulation. However, in mammals energy metabolism is even more complex than in microorganisms due to tissue specific enzyme isoform expression and allosteric regulation [97]. The dysfunction of glucose metabolism is implicated in various metabolic diseases. In cancer cells and in many fast-growing cells, glucose is rapidly converted to lactate concurrently with oxidative metabolism – a phenomenon often referred to as aerobic glycolysis and the Warburg effect [98-100]. In the manufacturing of biologics, excessive

lactate production in the late stage of culture can affect the productivity and glycosylation of proteins produced [101, 102]. An ideal metabolic state is one with low glycolytic flux, reduced lactate production, and a high proliferation rate.

Numerous attempts have been made to engineer glucose metabolism in cell lines to test hypotheses of the Warburg effect or to improve glucose metabolism for cell bioprocessing. Examples include the overexpression of a cytosolic form of pyruvate carboxylase (PYC2) [103], knockdown of lactate dehydrogenase (LDH) [104], knockdown of glucose transporters [105], overexpression of a mutant form of 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase (PBKFB) that has a low kinase activity [106], knockdown of TP53-inducible glycolysis and apoptosis regulator (TIGAR) [107, 108], isoform switch from pyruvate kinase (PK) M2 to M1 [109], and knockout of PKM2 [110]. For greater control, multiple enzyme expressions can be altered simultaneously, such as the knockdown of both LDH and pyruvate dehydrogenase kinase (PDK) [85]. These efforts have not be able to eliminate the Warburg effect to allow for fast cell proliferation without aerobic glycolysis. We hypothesize that the high energy provision and biosynthetic precursor supply needed for rapid cell growth requires achieving a new homeostatic state that can only be met by optimal alteration of multiple enzymes of energy metabolism pathways. As biological systems often do not rely on a single rate controlling enzyme [111], appropriation of metabolic intermediates to branching pathways likely requires more complex intervention than at a single step [112-114]. However, selecting combinations of enzyme changes that alter the glycolytic behavior while still meeting the constraints of growth requirements is a daunting task.

Central metabolism is biochemically well characterized and well suited for systematic exploration of its behavior using mathematical models [115-117]. The selection of appropriate combinations of enzyme alterations can be aided by mathematical optimization of the model to

achieve defined objectives [118, 119]. Optimization is a powerful mathematical tool which guarantees that decisions made are the best possible with respect to a given objective. Previous work has applied optimization for model structure determination and parameter estimation [120], as well as identification of the optimal and minimal number of genetic alterations for maximizing yield [112, 113].

Optimization problems are typically classified by the presence of continuous or integer variables and linear or nonlinear constraints. Linear optimization problems are typically easier to solve, and a linear formulation for identifying metabolic engineering targets can be achieved by formulating the biochemical reactions as a stoichiometric model [121-123]. Because integer variables are used to keep track of whether an enzyme changes expression or not, a formulation with stoichiometric models is a mixed integer linear program (MILP). CosMos, for example, uses a bilevel optimization approach to identify combinations of flux changes in a stoichiometric flux network, and uses a penalty to control the number of changes made to the network [121]. Stoichiometric model optimization has been successfully applied to force the co-utilization of glucose and xylose in *Escherichia coli* for biofuel applications [124]. Kinetic metabolic models can capture the many nonlinear regulations present in higher organisms, but are typically ill-conditioned due to biological time scales, rates, and concentrations varying over multiple orders of magnitude [125], adding considerable numerical difficulty to finding a solution [126]. The nonlinear nature of these models necessitates solving mixed-integer nonlinear programs (MINLP) [113]. These are challenging problems to solve, and different methods and solvers have been developed to reduce computational cost [127].

To avoid solving a challenging MINLP when considering a kinetic model, a variety of simplified model structures have been used to allow for MILP formulation for kinetic models [120, 128]. Under the condition that metabolic regulations keep the change in concentrations of many

metabolites small [114, 129], simplified model forms may be adopted in place of mechanistic models. However, such simplified model structures may not be predictive for larger metabolic changes [130]. Yet other studies have combined different model types to take advantage of the size of stoichiometric models and the detail of kinetic models. k-Optforce, for example, formulates a MINLP to solve a combined stoichiometric and kinetic model to identify the minimum interventions to maximize a production rate using either a single, or two-step optimization method depending on the complexity of the model [118].

In this work, we take a novel approach to select metabolic engineering targets without resorting to solving an MINLP; instead of reformulating nonlinear constraints, we eliminate the need for integer variables to identify which genes are altered through the use of convex penalty terms in the objective function. These terms are analogous in purpose to the penalty terms and constraints used in the above works by allowing the identification of a small number of genetic targets to ensure that the results obtained are experimentally feasible. By eliminating integer variables, we are able to optimize nonlinear kinetic models of large size while retaining the additional detail present in mechanistic models. The resulting nonlinear program (NLP) reduces the computational time required for solution and allows for a comprehensive evaluation of alternatives. We apply this novel optimization framework to an ill-conditioned, nonlinear kinetic model of glucose metabolism to identify minimal sets of genetic interventions that can redirect the distribution of metabolic flux. Here, we target the Warburg effect by minimizing lactate production while maintaining cellular requirements for proliferation and provide a detailed analysis of the resulting optimized metabolic states. This case study explores the potential link between the Warburg effect and cell growth through precursor supply, and demonstrates the insight gained in each step of the optimization framework. The optimization points to paths of mitigating the Warburg effect by simultaneous engineering of at least three enzymes in energy metabolism.

3.2. Methods

3.2.1. Glucose metabolism model

The kinetic model used for this work is taken from a previous model developed to describe glucose metabolism in cultured mammalian cells using the same parameters and rate equations as previously published [116]. In this model we consider multiple isoforms for three enzymes in glycolysis: hexokinase (HK), phosphofructokinase (PFK), and PK using information previously detailed [97]. These isoforms have different kinetic parameter values and are subject to different allosteric regulations. Changes in their relative expression level alter the cell's metabolic behavior. In the model, the dominant isoforms were chosen as HK1, PFKM, and PKM2. In the optimization other isoforms can be overexpressed. Under certain conditions this kinetic model exhibits bistable glycolysis flux [97]. Two stable steady states of glucose fluxes then coexist: a high flux state, characterized by a high rate of glucose consumption and correspondingly high lactate production, and a low flux state, characterized by low glucose uptake and low lactate production or even lactate consumption.

The model encompasses glycolysis, the (PPP) in the cytosol, the TCA cycle in the mitochondria, and molecular exchange reactions between the cytosol and mitochondria (Figure 3.1). The extracellular concentrations of glucose and lactate, and the intracellular concentration of glutamine are held constant. The dilution effect of cell growth is assumed to be negligible. The kinetic model is of the general form:

$$\frac{dC_i}{dt} = \sum_{j \in S_i} \nu_{ij} r_i \tag{1}$$

where i denotes metabolites, j denotes an enzyme and the reaction it catalyzes, ν_{ij} is the stoichiometric coefficient of species i in reaction j, r_j is the rate of reaction j, C_i the concentration of species i, and S_i the set of all reactions in which C_i participates. As this model has both cytosolic

and mitochondrial compartments, species which exist in both compartments are considered as separate species in the different compartments.

The reaction rates are of the following form:

$$r_j = E_j \cdot f\left(E_{0,j}, k_{cat,j}, \boldsymbol{K}_j, \boldsymbol{C}\right) \tag{2}$$

where $E_{0,j}$ is the original enzyme concentration in the model for enzyme j, $k_{cat,j}$ is the catalytic rate constant, $\boldsymbol{K}_j$ is a set of kinetic constants corresponding to r_j, $\boldsymbol{C}$ is the set of concentrations of species present in the rate expression, and E_j is the relative expression of enzyme j compared to the wild type model value. The different E_j comprise the decision variables in the optimization problem. Values of E_j of 1, > 1, and < 1 correspond to the original, increased, and decreased concentration compared to the wild type expression of enzyme j, respectively. In this study, extracellular concentrations of 5 mM and 1 mM for glucose and lactate respectively are used as physiologically relevant conditions for optimization [131]. Additionally, as this mechanistic model does not have an explicit formulation for growth, here we implicitly account for growth through constraints on the flux of ATP and the concentration of intermediates that supply precursors for cell growth, as will be further discussed in the following sections.

3.2.2. Problem formulation

An optimization strategy was formulated to identify a combination of changes to the relative enzyme expressions E_j that will achieve reduced lactate production at steady state while maintaining metabolic requirements for fast growth.

Objective function

Finding feasible steady state solutions for such an ill-conditioned model requires a good initial guess. Additionally, a full exploration of the enzyme expression space to identify candidate enzymes for modification would typically require non-convex constraints or formulation as an

integer optimization problem, which would add considerable computational complexity. To overcome these difficulties, the objective function is formulated as

$$\min\left[r_{LDH} + a \cdot \sum_i \left(\frac{dC_i}{dt}\right)^2 + b \cdot \sum_j \left(\log_{10} E_j\right)^2\right] \tag{3}$$

Here, r_{LDH} is the rate of lactate production, which is the primary metabolic objective. The second term is a penalty weighted by a to facilitate the finding of a feasible steady state, which removes the need for a good initial guess of the rates and concentrations. The last term is another penalty consisting of relative enzyme levels and a penalty weight b. The weighting factor b can be modulated to adjust the tradeoff between a reduction in the primary metabolic objective r_{LDH} and the overall magnitude of change in the enzyme expression variables E_j. Increasing the value of b will decrease the extent to which different enzyme expression values are changed in optimized states. A squared logarithm of these terms is chosen so that the penalty for each E_j is symmetric around a value of one, which represents their nominal values. The penalty of a reduction (knockdown) of an E_j then has the same value as that of an increase (exogenous expression).

Decision variables and metabolic constraints

The decision variables for the optimization are the relative enzyme expression levels, E_j. They are constrained in the range of

$$0.01 \leq E_j \leq 100 \tag{4}$$

except for reactions where large increases in maximal reaction velocity are infeasible. For example, in the case of pyruvate dehydrogenase complex (PDHC), a multi-unit protein complex, the upper bound is set to ten, as the activity is often increased indirectly by manipulating the expression of its inhibitor, PDK. We will denote the activity of PDHC instead of specifying the specific protein

to be engineered. Additionally, a number of metabolic constraints are imposed to represent the requirements of a growing cell in the context of the model:

(1) The production rate of ribose-5-phosphate and NADPH through the PPP is at least equal to the high flux state by setting the rate of the first step of the PPP to ensure an adequate supply of NADPH and precursors for important biosynthetic pathways:

$$r_{G6PD} \geq 1\ mM/hr \tag{5}$$

(2) The maximum rate of ribose-5-phosphate production is restricted to no more than approximately fivefold higher than the original high flux metabolism to prevent excessive carbon flux through the PPP:

$$r_{PRPP} \leq 5\ mM/hr \tag{6}$$

(3) The flux of ATP citrate lyase is set to be no higher than that observed in the high flux reference states to prevent an excess flow of carbon to the synthesis of fatty acids/lipids:

$$r_{CLY} \leq 15\ mM/hr \tag{7}$$

(4) The energy generation is no lower than that seen in a high flux state to support any additional energetic requirements in the growth state:

$$\begin{aligned} 2.5 * \left(NADH_{gen,m} - \left(NADH_{gen,m}\right)_{high\ flux}\right) + 1.5 * \left(FADH_{2,gen,m} - \left(FADH_{2,gen,m}\right)_{high\ flux}\right) \\ + \left(ATP_{gly} - \left(ATP_{gly}\right)_{high\ flux}\right) - \left(ATP_{cons} - (ATP_{cons})_{high\ flux}\right) \geq 0 \end{aligned} \tag{8}$$

These terms correspond to the ATP generation in the mitochondria through the flux of NADH and $FADH_2$, the production of ATP through glycolysis, and the loss of ATP through those reactions which consume ATP, respectively. The detailed terms are as follows:

$$NADH_{gen} = r_{pdhc} + r_{idh} + r_{gdh} + r_{akgd} + r_{mdh2} + r_{me2}$$

$$FADH_{2,gen} = r_{sdh}$$

$$ATP_{gly} = r_{pgk} + r_{pk} - r_{hk} - r_{pfk}$$

$$ATP_{cons} = r_{cly} + r_{pc} \quad (9)$$

(5) An upper bound for the pyruvate flux into the mitochondria is set at a level two times higher than that seen in a high flux state to prevent an excessive rate of mitochondrial activity:

$$r_{PYRH} \leq 90\ mM/hr \quad (10)$$

(6) The synthesis of serine, a key nutrient for rapidly proliferating cells, from 3-phosphoglycerate (3PG) [132-135], is kept above a basal level by setting 3PG concentration to be at least that of the high flux state:

$$C_{3PG,C} \geq 0.2\ mM \quad (11)$$

An implicit assumption is that the flux directed toward serine biosynthesis is small compared to the overall carbon flux through glycolysis.

(7) The concentration range of metabolites are constrained as:

$$0.1\ \mu M \leq C_i \leq 5\ mM \quad (12)$$

except for a few compounds that are known to already be present at a higher or lower level in our model, such as intracellular lactate. These constraints were chosen to maintain a physiologically relevant state and otherwise prevent changes to metabolism which could result in reduced cell growth or even cell death. In total, this optimization problem has 266 variables and 177 constraints. Simulations were performed on a combination of a desktop computer using an Intel i7-4790 processor and at the Minnesota Supercomputing Institute (MSI) using Intel Xeon Nehalem processors.

3.2.3. Optimization framework

The metabolic reaction network, together with the objective function and metabolic constraints, form a nonlinear program (NLP), for which open-source solvers are available. The size and non-convexity of this problem preclude finding the global optimum, so local solvers are employed. The problem was solved in the software General Algebraic Modeling System (GAMS), using the solver Sparse Nonlinear OPTimizer (SNOPT) [136]. The framework discussed below is schematically shown in Figure 3.2.

Overall optimization strategy

The locally optimal solutions obtained are strongly dependent on the initial guesses used for the decision variables. To adequately sample the solution space, multiple random initializations are performed on the relative enzyme expression variables E_j to find different local optima. As it is difficult to find feasible steady state solutions starting from a random initial guess, the second term is used to help guide the solution towards steady state before enforcing it as a constraint. This enables easier exploration of the complex nonlinear space, with a larger fraction of random initializations resulting in locally optimal solutions. After each initialization, optimization is performed with $a = 10^{-3}$, i.e. applying only a small penalty to deviations from steady state (Figure 3.2A and 3.3). After obtaining a locally optimal solution, a is increased iteratively by an order of magnitude until reaching $a = 10^{3}$. By using a to guide the model towards a steady state, the difficulty of finding feasible solutions at steady state is reduced. A final iteration then includes Eq. 13 as a constraint to explicitly require steady state:

$$\frac{dC_i}{dt} = 0 \; \forall \; i \tag{13}$$

In this study, the objective function is to reduce the rate of lactate production. A wide range of feasible glucose uptake rates satisfy the constraints. The optimization will favor solutions with

a low glucose flux, as they have less flux to direct away from lactate production. To identify glucose flux states, we formulated a multi-objective optimization problem using ϵ-constrained optimization [137]. The rate of glucose uptake, r_{glut1} was constrained to be above a value ϵ:

$$r_{glut1} \geq \epsilon \tag{14}$$

By increasing the value of ϵ, local optima with different glucose uptake rates and different lactate production rates are found. The initial value of ϵ is set to zero and is increased by $0.5\, mM/hr$ every 32 iterations. Random initialization is performed 3840 times for each optimization loop, to gain a thorough sampling of the different locally optimal solutions possible. This loop is described schematically in Figure 3.2A.

Identification of key enzyme expression changes

The optimization seeks combinations of relative enzyme levels that achieve the posed objective. The number of possible such combinations is inordinately high. We therefor devise a two-stage optimization framework to identify combinations of a small number of enzymes that achieve the objective.

First, optimization is performed using different penalty weights b spanning several orders of magnitude. For each value of b, E_j are randomly initialized a few thousand times and locally optimal E_j^* values are identified (Figures 3.2A and 3.2B). The enzyme penalty term of Eq. 3 consists of a penalty weight and the contribution of all enzymes. The average contribution of the relative expression level of enzyme j to the penalty term at a given b, $P_{E_j}(b)$, is expressed as:

$$P_{E_j}(b) = \frac{1}{n_b} \sum_{k=1}^{n_b} \left(\log_{10} E_{j,k}^*\right)^2 \tag{15}$$

where n_b is the number of local optima found and k denotes a locally optimal solution. How P_{E_j} changes as b increases is used as an indicator for the impact of the relative expression level of enzyme j to the reduction of lactate production as will be described in the Results section. For subsequent identification of key enzymes, a value of b is selected such that optimal reduction of lactate is attained with most P_{E_j} relatively small, indicating that only a subset of enzymes expressions are appreciably changing during optimization.

Next, for the chosen b, the enzymes with the highest P_{E_j} are considered for further analysis. To determine the order of importance, enzymes are ranked based on their calculated P_{E_j}. Enzymes with the largest P_{E_j} are those which change more consistently or by a larger magnitude. Those enzymes with large P_{E_j} as the penalty parameter b increases therefore must have an impact on the metabolic objective which outweighs the penalty levied by the enzyme change to the objective function. P_{E_j} can thus be used to identify enzymes whose expression level alteration has the more profound effect on the reduction of lactate production.

In the second stage the optimal modifications to enzymes sets are determined. Enzyme sets are formed by choosing enzymes in rank order to form sets of increasing numbers of enzymes. For each set, all E_j are fixed at 1 except for the selected enzymes and the penalty for enzyme change was removed by setting $b = 0$ before further optimization, as shown in Figure 3.2B. By repeating this procedure for different enzyme sets, the potential of different enzyme combinations to reduce lactate can be assessed. These enzyme sets are analyzed both by looking at the enzyme compositions of their Pareto optimal solutions, and through the use of t-distributed stochastic neighbor embedding (t-SNE) for dimensionality reduction to compare enzyme changes and the impact on reaction flux and metabolite concentrations [138].

3.3. Results

3.3.1. Identification of key enzymes

A total of 54 enzymes and transporters were allowed to increase or decrease in relative expression in the optimization. The penalty weight, b, was varied over four orders of magnitude (0.01 to 100). The frequency distribution of the relative levels for each enzyme at each b value was computed. Shown in Figure 3.4A are such distributions for E^*_{LDH} and E^*_{GLAST} (for enzymes LDH and the glutamate aspartate transporter, GLAST (SLC1A3)), at three b values. Among the local optima, E^*_{LDH} has a wide distribution at $b = 0.01$, with a large fraction of optima being highly amplified or suppressed. As b increases to 1, E^*_j begin to reduce their magnitude and cluster around their wild type values of 1. As b increases further to 100, the propensity of LDH enzyme level to change in the optimal solutions diminishes. In contrast, at a low value of $b = 0.01$, E^*_{GLAST} similarly takes on a wide range of values although only in the direction of amplification, but as b increases, the distribution E^*_{GLAST} quickly clusters around its starting value ($\log_{10} E^*_{GLAST} = 0$). The data thus suggest that LDH is less sensitive to the penalty weight than GLAST and likely to play a more significant role in the optimization.

The P_{E_j} of all the enzymes and transporters at different b values is shown in Figure 3.4B, with the curves for LDH and GLAST shown in red and yellow respectively. The behavior shown in Figure 3.4A for LDH and GLAST is reflected in the P_{E_j} plot. P_{E_j} is high in low b regions for both, decreases quickly to near zero for GLAST but remains at a relatively high value for LDH for moderate b. A large portion of enzymes behave similarly to GLAST, with large P_{E_j} at low penalty weight b that decreases precipitously as b increases. Importantly, a few enzymes can be seen to behave similarly to LDH, retaining a high P_{E_j} value at high b. These enzymes, like LDH, would have a high impact on reducing lactate production. At the largest b values even enzymes which have significant contribution cease to change.

The average lactate production of all optima obtained at different b values is shown in Figure 3.4C. When b is small, the average lactate production is low, but begins to sharply increase when $b > 10$. The contribution of the enzyme penalty and lactate production terms is shown in Figure 3.5 to visualize the impact of changing b. Choosing a sufficiently large value of b is important to eliminate those enzymes with little impact on lactate production while retaining key enzymes needed to achieve low lactate production. We therefore choose $b = 10$ to capture the most important enzymes, as it is the highest value of b which retains low LDH flux while having only a small amount of enzymatic change. To ensure that this choice of b is appropriate, alternative values of b have also been analyzed in the same method as below, and their results are shown in Figure 3.6.

To identify enzymes which play key roles in reducing lactate production, enzymes are ranked according to their P_{E_j} at $b = 10$ (Figure 3.7A). P_{E_j} for most enzymes is close to zero. Only a small number of enzymes, including LDH and PDHC, have a high P_{E_j}. The top ranked enzymes are then formed into sets for further optimization as will be described in the next section.

3.3.2. Optimization of enzyme combinations

The enzymes with the top P_{E_j} shown in Figure 3.7A are combined into five sets of the top one, two, three, five, and six enzymes in rank order and subjected to optimization of expression change. The penalty b is set to zero, while E_j values for all other enzymes are fixed at the baseline value of 1. The rate of lactate production of each optimum is plotted against the reciprocal glucose consumption rate (Figure 3.7B). Among all the optima obtained for each set of enzymes, only the Pareto efficient optima, or those with the lowest lactate production rate for a given glucose consumption are plotted. Thus, any optimum with a lower lactate production would have a lower glucose consumption. The dotted lines in Figure 3.7B indicate the molar ratio of lactate produced to glucose consumed ($\Delta L/\Delta G$). The complete conversion of glucose to lactate, or the maximum

$\Delta L/\Delta G$, is 2.0. At a typical high flux state the $\Delta L/\Delta G$ is higher than 1.4 [139]. With only the expression level of the top ranked LDH being optimized, or the top two of LDH and PDHC, no local optima achieve significant reduction of lactate production. With the three enzyme combination of LDH, PHDC, and SCS (succinyl-CoA synthetase), $\Delta L/\Delta G$ decreases below 0.7. The addition of ALD (aldolase) and GOT1 (aspartate aminotransferase) in the five-enzyme case sees lactate production nearly eliminated with concurrent reduction of glucose consumption. When ME1 (malic enzyme) is added to the optimization for the six-enzyme case, the optima reach into regions of lactate consumption.

3.3.3. Optimal enzyme expression

The Pareto fronts for the top-five and top-six enzyme sets from Figure 3.7B are replotted in Figures 3.7C and 3.7F respectively with each optimum colored by lactate production rate. The E_j^*s of a local optimum are depicted as one line (Figures 3.7D, 3.7E and Figures 3.7G, 3.7H for the top-five and top-six cases respectively) and its lactate production rate is represented by the same color scale as in Figure 3.7C. The expression changes for some enzymes, such as the decrease of ME1 (Figures 3.7G and 3.7H) as well as the increase of PDHC activity, fall into a relatively narrow range for all optima across the whole Pareto front, while changes of some other enzymes take on a range of values, such as ALD (Figure 3.7E). Some enzymes even change their expression in opposite directions in different optima, as can be seen for LDH in the top-six enzyme set, where decrease or increase in relative expression is seen (Figure 3.7G).

Some features of enzyme level changes can be seen. For the top-five enzyme case the different degrees of lactate production reduction (Figures 3.7D and 3.7E) show similar profiles except that the magnitude of changes is larger when lactate production is almost eliminated (Figure 3.7E). Common among many solutions in the top-six enzyme case is the near elimination (10-100 fold decrease) of ME1, increase of PDHC activity, and increase in SCS level (Figures 3.7G and 3.7H).

3.3.4. Homeostasis and metabolite concentrations

The local optima of the top-six enzyme set were further analyzed using t-SNE. t-SNE is a nonlinear dimensionality reduction technique which can condense high dimensional data down into two or three dimensions where similar points are found close together and disparate points found farther apart. The six-enzyme set was chosen because the resulting optima cover a wide range of lactate production and consumption. The fluxes of glucose uptake (r_{GLUT1}), lactate production (r_{LDH}), G6P entry into the PPP (r_{G6PD}), and malic enzyme flux (r_{ME1}) in the cytosol are shown in Figures 3.8A-D. Glucose and lactate flux are indicative of the metabolic states. The metabolic optimal states shown vary significantly in both the rates of glucose consumption and lactate production, yielding a large range of $\Delta L/\Delta G$, and distinct regions which include local optima achieved by different metabolic strategies. Region I, marked in Figure 3.8A, has many of the lowest glucose consumption, highest lactate consumption states. Region II corresponds to a mix of metabolic states reflecting more moderate glucose flux. Finally, region III reflects both high and low glucose consumption, but corresponds to distinct changes in enzyme expression level.

Shown in Figures 3.8E-H are the relative change of intracellular concentration of key metabolites from the baseline high flux metabolic state. These metabolites, glucose-6-phosphate (G6P), fructose-1,6-bisphosphate (F16P) and fructose-2,6-bisphospate (F26P), exert major allosteric regulation in glycolysis (Figure 3.1B). The concentration of pyruvate is also shown, as it lies at a key branching point, intersecting carbon flow between glycolysis, lactate production, and the TCA cycle. As G6P inhibits HK, its highest values (found primarily in region I) correspond to the lowest glycolytic flux. Region I, and to a lesser extent region II, also have the largest decreases in F16P and F26P concentrations (Figures 3.8F and 3.8G).

The enzyme concentration adjustments which achieve these flux and concentration values are shown in Figures 3.8I-N. Certain enzymes are relatively consistent in their expression change: E^*_{SCS}

is overexpressed over a wide range of local optima, with little variation in its behavior at different regions (Figure 3.8M). Other enzymes, however, exhibit a wide range of behaviors. For example, E^*_{ALD} (Figure 3.8I) is overexpressed primarily in region I and part of region III. Region I contains the most extreme changes to these six enzyme expressions, corresponding to large changes to glucose and lactate flux.

Region III has decreased E^*_{LDH}, whereas region I as well as most other states have increased E^*_{LDH}. The overexpression of LDH in these states may facilitate lactate consumption due to the low driving force from the small extracellular concentration of lactate present in the optimization. Concurrently, region III represents those metabolic states where ME1 is not decreased (Figure 3.8N). These states correspond to increased pyruvate concentration (Figure 3.8H), whereas most of the states with ME1 decrease have a substantial decrease in pyruvate concentration, reflected in region I and II. The adjustment of ME1 also corresponds to malic enzyme flux (Figure 3.8D) and is inversely related to PPP flux (Figure 3.8C). This distinction demonstrates the presence of multiple distinct strategies for reducing $\Delta L/\Delta G$ present in the optimization.

3.3.5. Similar flux states from different enzyme expression

The fluxes of the optimized metabolic states are then compared to the original enzyme expression configuration ($E_j = 1\ \forall j$). The flux map for the original high flux metabolic state shown in Figure 3.9A is compared against Pareto efficient, optimal solutions shown in Figures 3.9B-D. While different locally optimal solutions may differ in certain fluxes or distributions, many optimal solutions share a core set of characteristics. Primarily, a relatively higher quantity of glucose is diverted to other pathways compared to the original metabolism, resulting in decreased lactate production as well as glucose consumption. Other changes common to the shown solutions are pyruvate flux to the mitochondria at the upper bound, along with increased malate-aspartate shuttle activity, compensating for reduced cytosolic NAD^+ regeneration through lactate production.

Additionally, the flux from malate to pyruvate in the cytosol is somewhat decreased, decreasing flux of metabolites from the mitochondria back to pyruvate. The relative flux to the PPP has also increased in each of these solutions. The main differences for these solutions originate from the number of enzymes and extent of change of these enzymes in each state. Going from the three-enzyme set (Figure 3.9B) to the five-enzyme set (Figure 3.9C) and six-enzyme set (Figure 3.9D), many metabolite concentrations are more greatly altered. This suggests that more extreme departures from the high flux state (reduction of $\Delta L/\Delta G$) require more significant rewiring of metabolism. Interestingly, however, even with differences in both $\Delta L/\Delta G$ and metabolite concentrations, the internal flux distributions are comparable, indicating that multiple distinct strategies may yield similar metabolic phenotypes.

3.4. Discussion

In this work, we developed an optimization framework to identify sets of reactions in a metabolic network that can be modified to meet a metabolic objective function. A key feature of this approach is the use of a penalty term with an adjustable penalty weight b, to replace the use of integer variables and reduce the computational complexity. We took a two-step optimization approach. First, b was adjusted to identify top ranked enzymes that have the largest contribution to the reduction of lactate production. Optimization was then performed on sets of different combinations of the identified enzymes to find lowest lactate production attainable for each enzyme combination.

There have been several hypotheses on why cancer cells and rapidly proliferating cells have high glucose consumption and lactate production, including that the metabolic state provides a high level of ATP or a high flux of intermediates for biosynthesis [140]. These hypotheses are not mutually exclusive, and it is possible many are at least partially correct under certain physiological contexts. Here, we investigated the ability for a low lactate production state to

provide biosynthetic intermediates, and used multiple enzyme concentration adjustments to decouple those phenomena.

This approach identified multiple combinations of enzymes and expression profiles that meet the constraints imposed for growth while achieving low lactate production. These combinations of enzyme changes may be considered as targets for cell engineering to alter metabolism. This analysis indicated that multiple routes exist to rewire glucose metabolism in growing cells. It also showed that changes in single or small groups of enzymes could not substantially reduce lactate flux; however, by changing more enzymes simultaneously, low flux can be achieved.

The enzymes found to have a high impact on lactate production provide insight into metabolism in the context of the regulations present in the model. Three of the enzymes (LDH, PDHC, and ME1) with the highest P_{E_j} center around the pyruvate node in the metabolic pathway. Most of the optima identified have a decreased cytosolic pyruvate concentration compared to the original high flux state (Figure 3.8H), thereby also a decreased driving force for lactate production. The concurrent increase in PDHC helps drive pyruvate flux into mitochondria and into TCA cycle to sustain the ATP production rate above the required level. Lactate consumption is associated with over expression of LDH (Figure 3.8J) to compensate for the relatively low driving force presented by the low extracellular concentration used for this optimization. The cytosolic ME1 is also mostly reduced (Figure 3.8N). This leads to reduced flux of mitochondrial intermediates to cytosolic pyruvate (Figure 3.9). However, reduced flux through ME1 also reduces NADPH production, which can be compensated for by increased PPP flux (Figure 3.9), although this was not a specific objective of this optimization.

There have been a number of reports on the knockdown of LDH in hybridoma [141], CHO [85, 104], breast cancer [142, 143], and a panel of cancer and non-cancer cells [144]. While lactate production was reduced, the growth can also be significantly affected. The results in this study also

showed that LDH knockdown alone does not lead to a state that satisfies the constraints stipulated for cell proliferation with a low rate of lactate production. Concurrent with LDH knockdown and PDHC activity increase by knocking down of PDK has been reported to reduce lactate flux in late-stage culture when the growth rate was slow [85]. Knockout of LDH in conjunction with PDK knockdown was found to be lethal [86], suggesting that increasing PDHC activity alone is insufficient to support cell growth in the absence of lactate production. Our results suggest that additional changes are required for the rewiring of glucose metabolism while still supporting rapid growth.

This optimization identified three other enzymes as pivotal: SCS, GOT1, and ALD. SCS level increases in most optima (Figure 3.8M), possibly increasing TCA cycle flux. GOT1, increased or unchanged in all optima identified (Figure 3.8K), increases the capacity of the malate-aspartate shuttle that regenerates NAD^+ in the cytosol and alters the homeostatic concentration difference of malate-aspartate shuttle components. It likely enables a higher fraction of cytosolic NAD^+ to be regenerated through malate-aspartate shuttle and reduces that regenerated via lactate production (Figures 3.9A and 3.9B). That in turn increases the fraction of pyruvate channeled into mitochondria and reduces its flux to lactate. ALD, mostly overexpressed or unchanged (Figure 3.8I), increases the capacity of the conversion of F16P to the two trioses, thus decreasing F16P concentration (Figure 3.8F) and in turn reducing its allosteric activation of PFK and PKM2. Additionally, while bistability was only shown to exist over a subset of possible conditions [97], it is possible that some engineered cells still retain bistability. In this case, their metabolism can be guided to the desired state through control of culture conditions such as maintaining low glucose concentration [145].

The optimization framework described in this report successfully identified optima that minimize the lactate production of rapidly growing cells, thus perturbing the Warburg effect. The

framework can be easily adapted to different conditions or objectives to devise strategies for metabolic engineering or test metabolic hypotheses. With the molecular biology tools for combinatorial multiple gene amplification and knock down readily available, it is feasible to validate the results shown in this work experimentally. Precise control of gene expression level is not trivial; however, many identified enzymes have a wide range of gene expression level that can impart the desired lactate flux change. The set of gene expression changes identified is likely to vary depending both on the constraints imposed to the optimization as well as the starting point for the baseline enzyme expression. Thus, with an appropriately tuned model for a given cell line, this framework can be a valuable tool for cellular and pathway engineering.

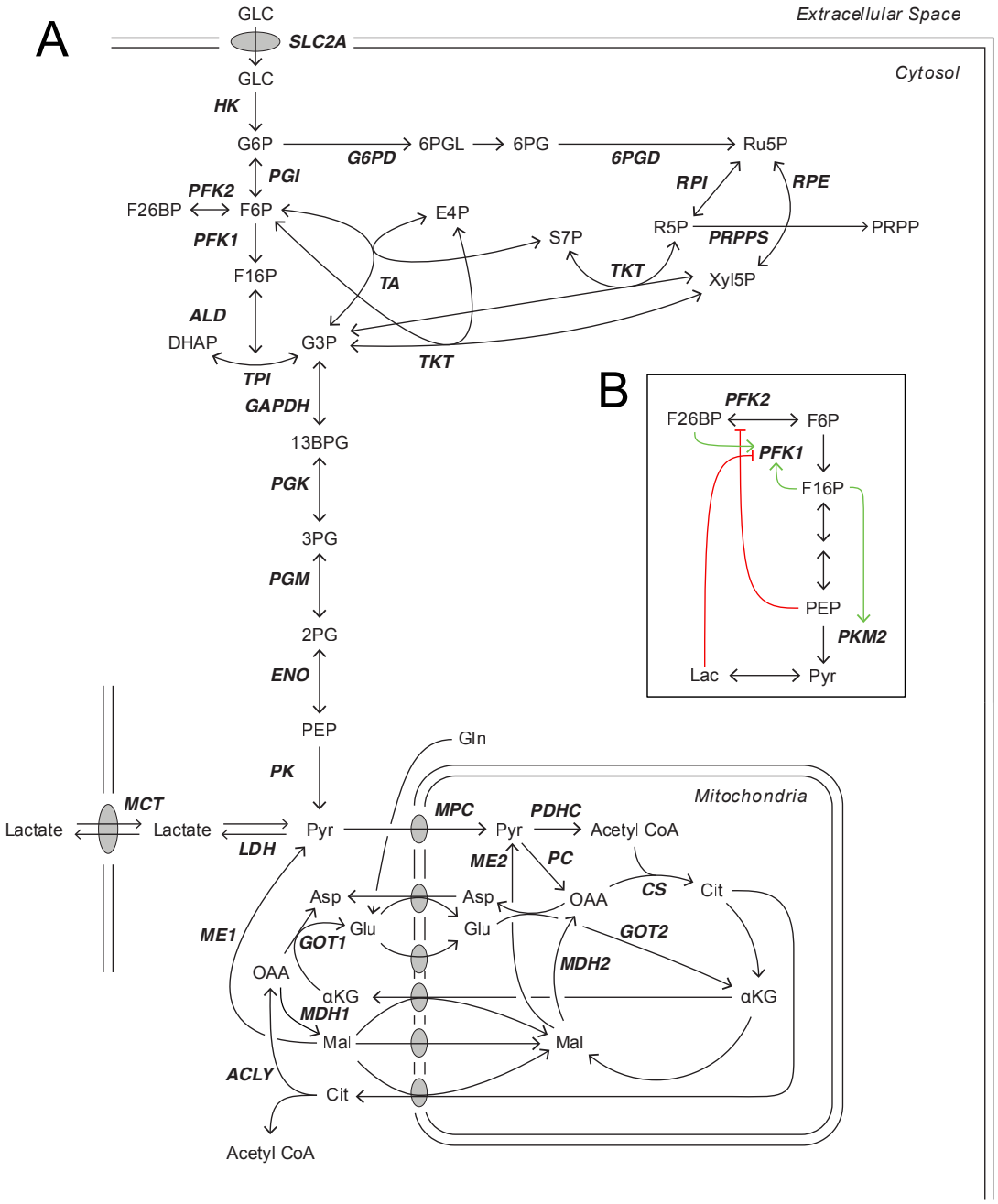


Figure 3.1: Glucose metabolism model.

(A) The kinetic model of central metabolism. Depicted are the pathways of glycolysis, the PPP, and the TCA cycle. The TCA cycle has been simplified for clarity. (B) Allosteric regulations for

key enzymes in glycolysis.

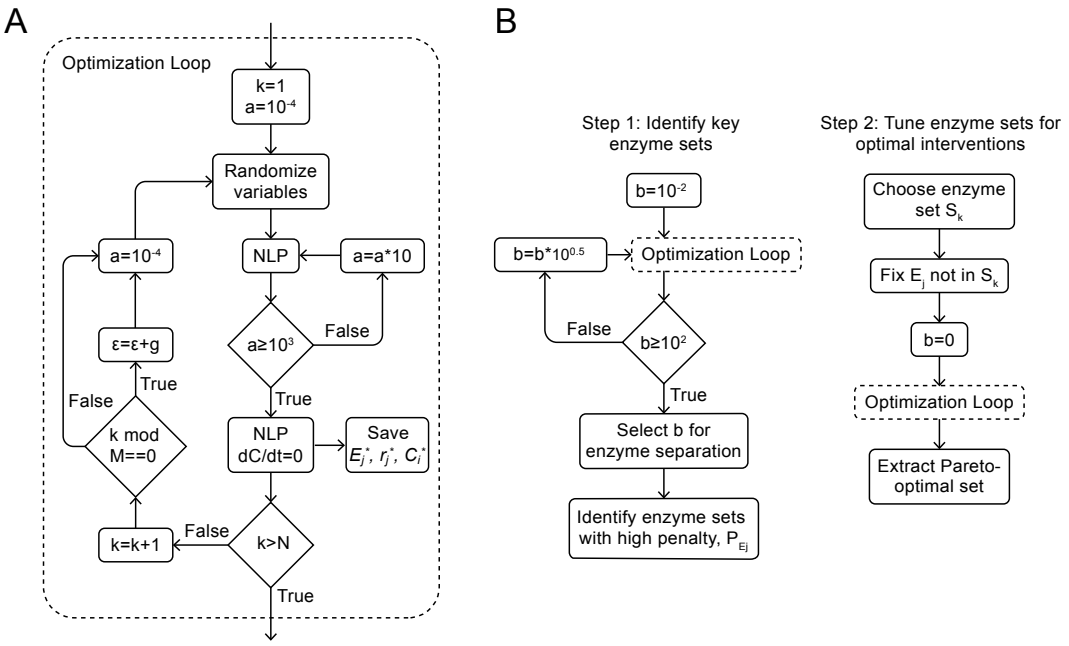


Figure 3.2: Optimization framework.

(A) Core optimization loop used in both steps of the optimization framework, involving random initialization followed by iterative solving to aid in finding the steady state value. Here N is the total number of iterations in the optimization loop, M is number of iterations per concentration of glucose, and g is the concentration step size of glucose for the ϵ constraint. (B) Optimization framework steps detailed. The first step aims to find the ideal value of penalty value b, and the second uses identified subsets of enzymes to find optimal metabolic interventions.

Algorithm 1 Optimization loop

$\epsilon \leftarrow 0$
for all $k \in 1 : N$ **do**
  Random initialization for all E_j in $[0.01, 100]$
  if k mod $M == 0$ **then**
    $\epsilon \leftarrow \epsilon + g$
  end if
  for all $a \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$ **do**
    Solve NLP to obtain E_j, r_j, C_i
    Use current E_j, r_j, C_i to initialize next solve
  end for
  Add constraint $\frac{dC_i}{dt} = 0$
  Solve NLP to obtain steady state E_j^*, r_j^*, C_i^*
  $soln \leftarrow E_j^*, r_j^*, C_i^*$
  Remove constraint $\frac{dC_i}{dt} = 0$
end for

Where N is the total number of iterations in the optimization loop, M is number of iterations per concentration of glucose, and g is the concentration step size of glucose for the ϵ constraint.

Algorithm 2 Step 1: Identify key enzymes

for all $b \in \{10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.5}, 10^{0}, 10^{0.5}, 10^{1}, 10^{1.5}, 10^{2}\}$ **do**
  Perform Algorithm 1 optimization loop using current b
end for
At a given value of b choose enzyme subsets

Algorithm 3 Step 2: Optimize enzyme subsets

for all Enzyme sets S_k **do**
  Perform Algorithm 1 optimization loop using $b = 0$ and all enzymes fixed except $E_j \in S_k$
end for
Extract Pareto optimal set of E_j^*, r_j^*, C_i^*

Figure 3.3: Mathematical Description of Algorithm.

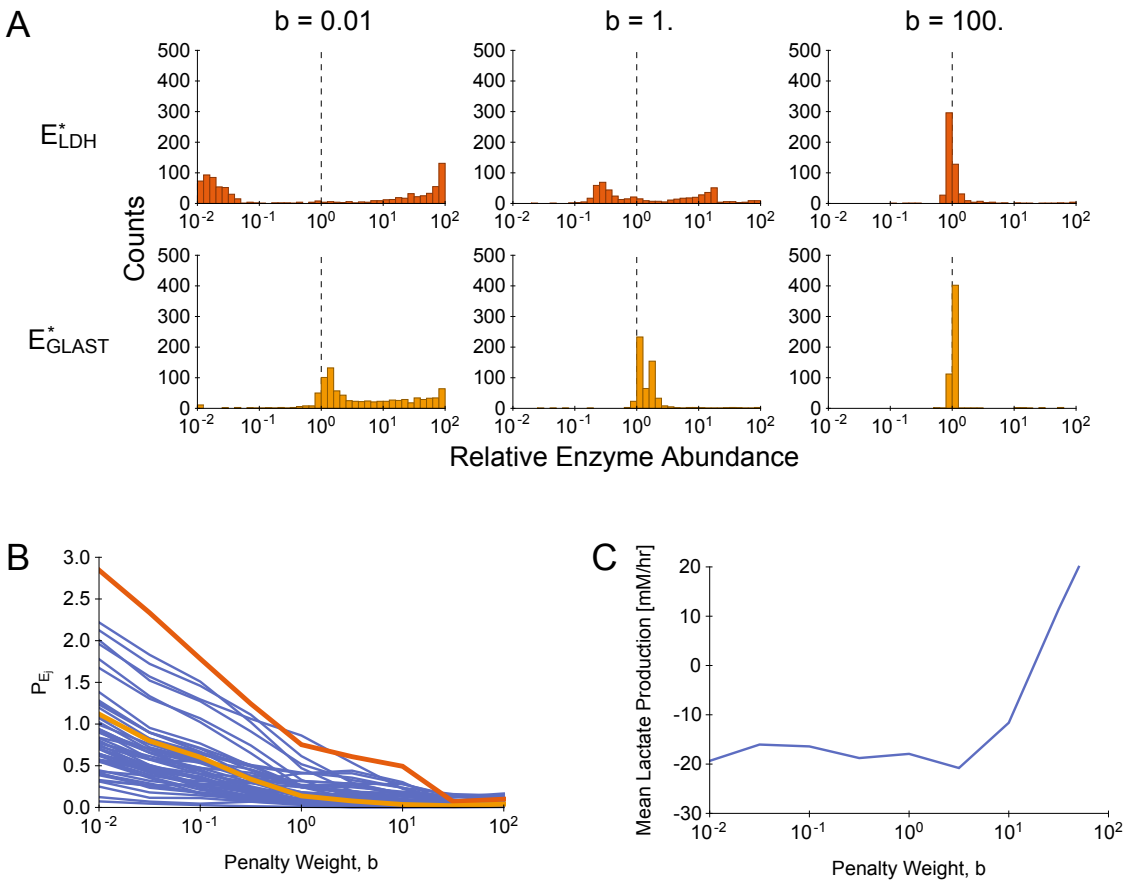


Figure 3.4: Identification of key enzymes.

Enzyme expression change and lactate response as a function of penalty weight. (A) Histograms of enzymes expression variable E_j values (organized by row) with different penalty weights (organized by column). E^*_{LDH} (top) and E^*_{GLAST} (bottom) are shown. Each histogram represents the relative enzyme level distribution for a particular value of b for all local optima identified. The dotted line at zero represents the level for the baseline metabolism. (B) P_{E_j} for all enzymes across different penalty weight values, where the red and yellow line correspond to E^*_{LDH} and E^*_{GLAST} as in (A), with the remaining enzymes shown in blue. (C) Mean lactate production at locally optimal solutions across different penalty weight values.

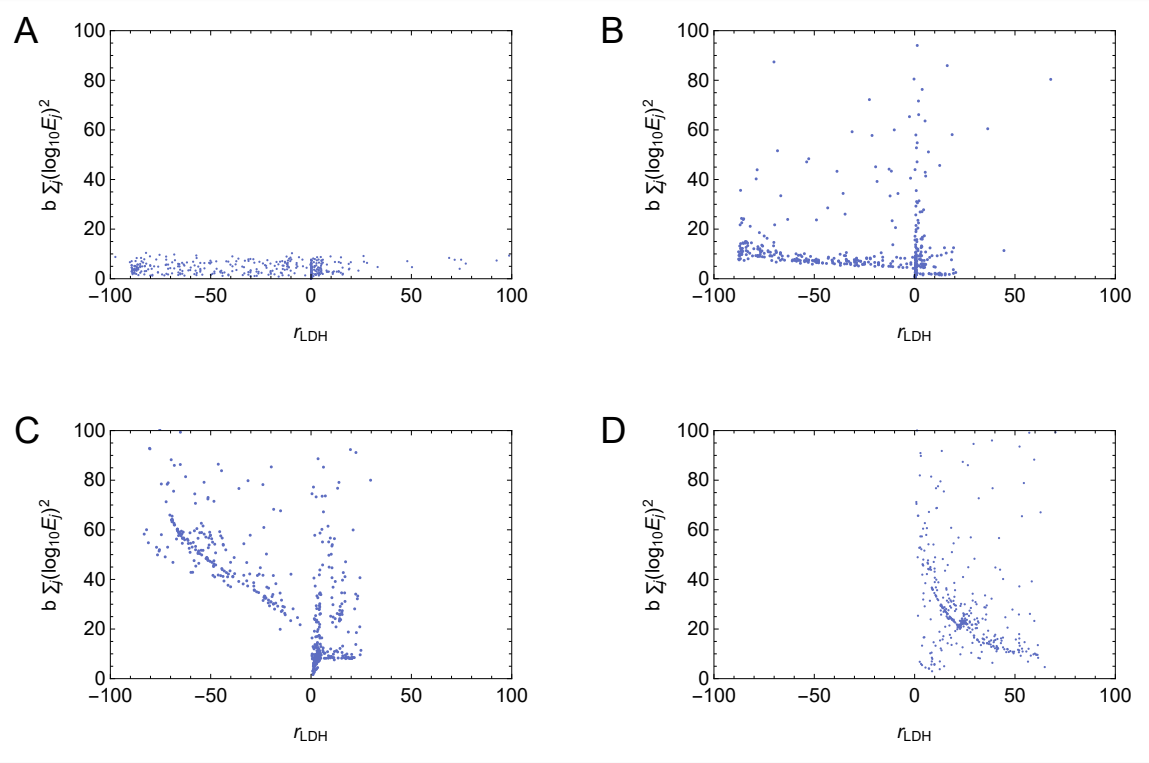


Figure 3.5: Objective function term values for different values of b.

Shown are $b = 10^{-1}$ (A), 10^{0} (B), 10^{1} (C), and 10^{2} (D).

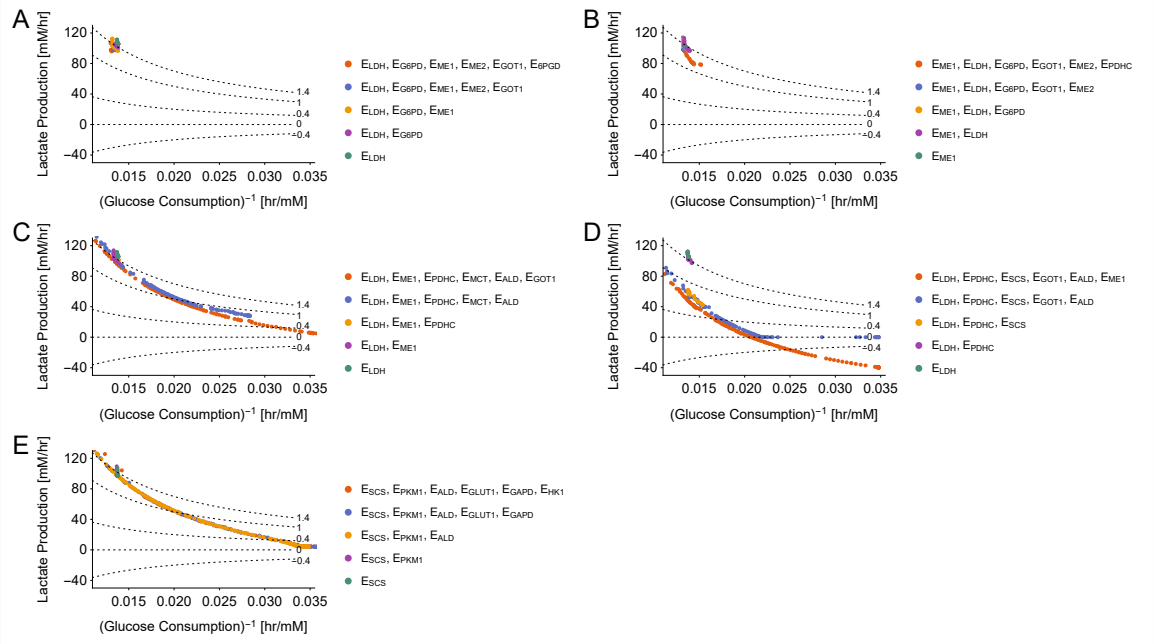


Figure 3.6: Optimization of enzymes sets chosen using different values of b.

Shown are the top ranked enzyme sets found by using $b = 10^{-0.5}$ (A), 10^{0} (B), $10^{0.5}$ (C), 10 (D), and $10^{1.5}$ (E). For those sets enzyme sets derived from $b < 10$, the performance in terms of lactate production at a given glucose consumption is worse. For $b = 10^{1.5}$, the three-enzyme case is able to achieve better performance than the set found from $b = 10$, but both the five- and six-enzyme sets have lower performance. Additionally, for $b = 10^{1.5}$, the five- and six-enzyme sets do not improve upon the performance from the three-enzyme set.

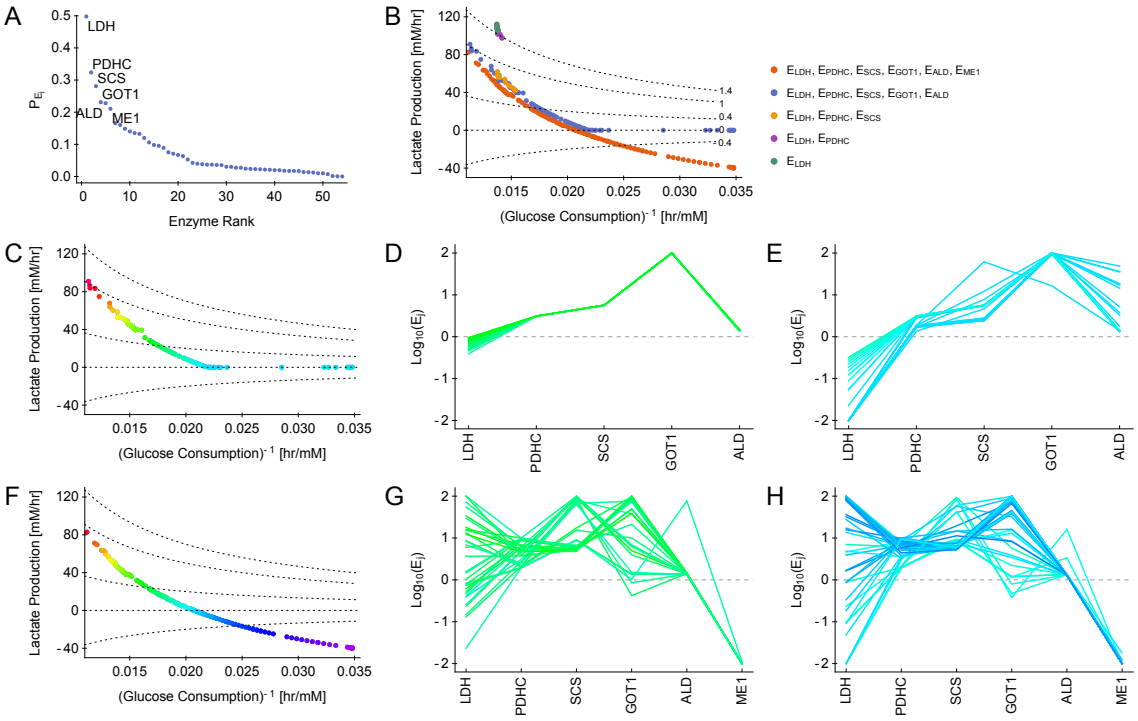


Figure 3.7: Optimization of enzyme combinations.

(A) Enzymes ordered by rank of P_{E_j}. (B) Pareto fronts for optimization of enzyme subsets chosen by P_{E_j} rank shown with contour lines representing ΔL/ΔG. The Pareto fronts for the five (C-E) and six (F-H) enzyme optimization cases from the subsets in (B), with data colored by lactate production rate, and dotted lines showing several $\Delta L/\Delta G$ values (left). Shown are the respective values of each optimized enzyme expression for low-lactate (D and G 10 mM/hr < r_{LDH} < 30 mH/hr) and no-lactate (E and H -10 mM/hr < r_{LDH} < 10 mH/hr) producing states, where each line represents a local optima whose color corresponds to the points along the Pareto front.

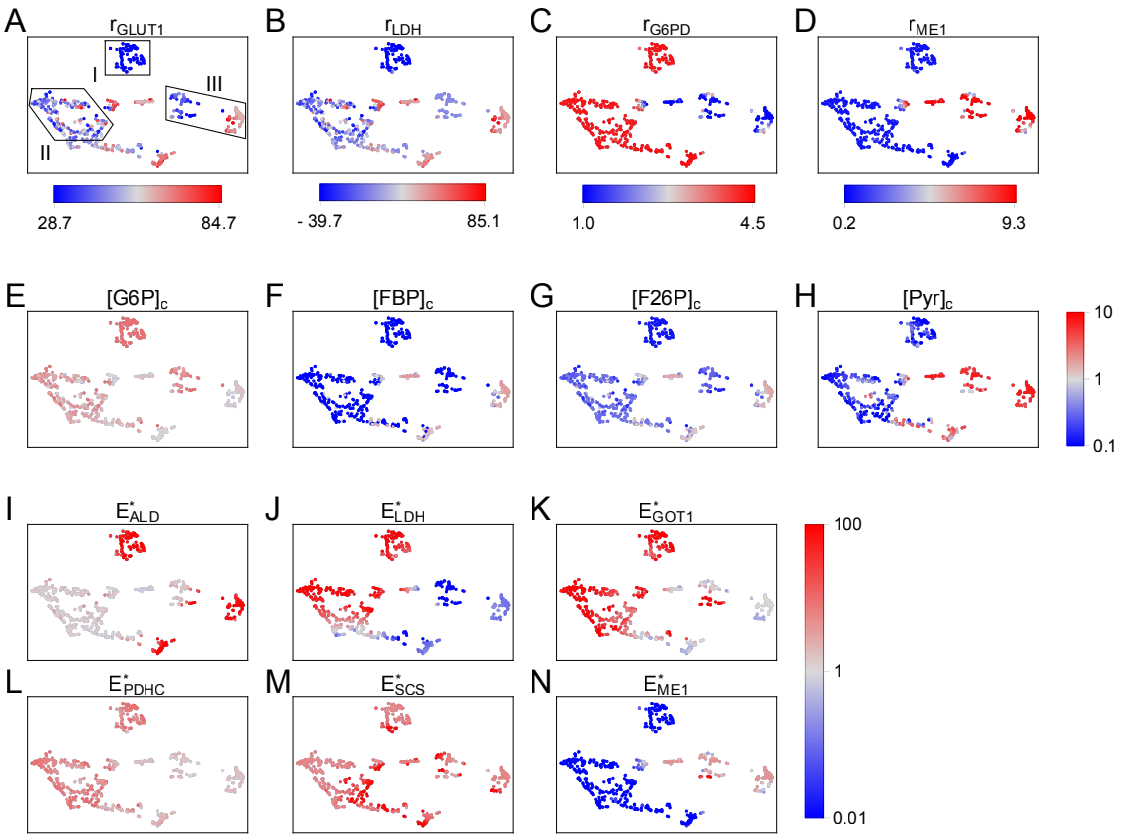


Figure 3.8: Homeostasis and metabolite concentrations.

t-SNE performed on the six-enzyme combination chosen by rank. (A-D) Absolute fluxes for selected reactions shown in absolute values in mM/hr. (E-H) Relative metabolite concentrations to the high flux state which have important allosteric interactions with metabolic enzymes (in the case of G6P, F16P, and F26P), and hold important positions in the reaction network for dictating the flow of metabolites (in the case of pyruvate (Pyr)). (I-N) Altered enzyme expressions of selected enzyme expression values. Increased rate, concentration, and enzyme expression is shown in red, while decreased is shown in blue. Regions I, II, and III are marked on (A) for discussion.

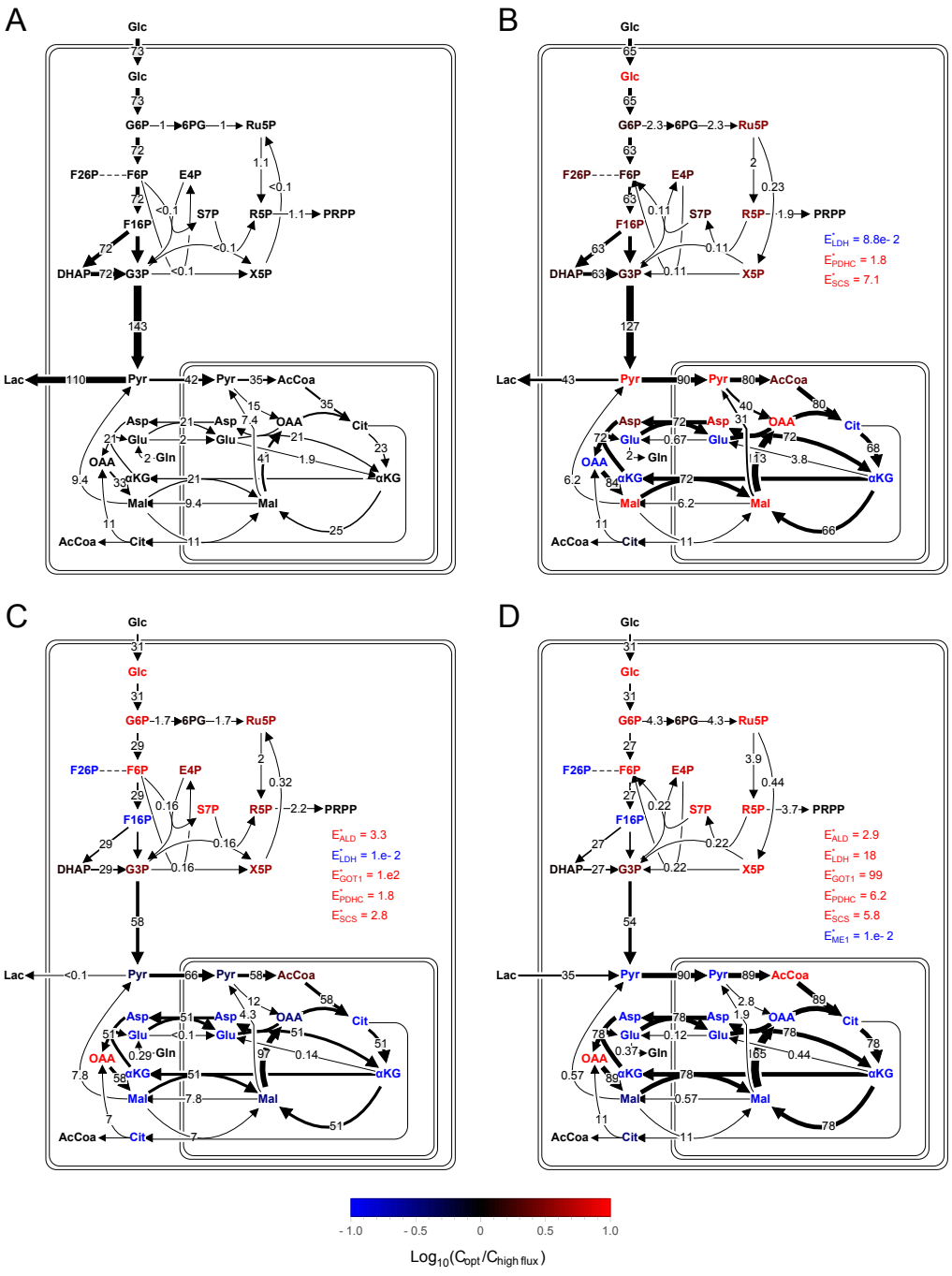


Figure 3.9: Similar flux states from different enzyme expression.

Comparison of optimized metabolism to the original metabolic state. Flux distributions for the

simulated extracellular glucose and lactate in a physiologically relevant range (near 5 mM glucose and 1 mM lactate). Shown are the flux distributions for the case of the original metabolism (A) and Pareto efficient optima for the lowest lactate production from the three-enzyme rank ordered set (B), a near zero lactate flux for the five- (C), as well as a lactate consumption state at similar glucose consumption with the six- (D) rank ordered enzyme sets. All reaction fluxes are given in mM/hr. The relative metabolite concentrations to the high flux state are depicted by color.

4. Understanding the metabolic requirements of gluconeogenesis through kinetic model optimization

4.1. Introduction

Gluconeogenesis, the synthesis of glucose from other carbon substrates, is one of several metabolic pathways key to controlling blood glucose levels. Thus, dysfunction or dysregulation of gluconeogenesis can result incorrect regulation of blood glucose, and is associated with a number of diseases, including non-alcoholic fatty liver disease [146] and diabetes [147].

While the carbon substrates for gluconeogenesis can come from many potential sources, a few substrates are primarily used to supply the production of glucose due to their participation in biological cycles with other tissues. One of these key substrates is lactate, which is released during exercise or other anaerobic conditions, largely from muscle, and is then turned back into glucose by the liver, a process called the Cori cycle [148]. Protein degradation also results in the release of amino acids, particularly alanine [149] and glutamine, which may be recycled through the liver to produce glucose. Glycerol, through the breakdown of fatty acids, also reenters central metabolism and serves as a gluconeogenic substrate [150].

The relative utilization of these different substrates depends upon the tissue (e.g. hepatic vs. kidney), the hormonal regulation in effect, and substrate availability. One estimate for hepatic tissue placed lactate as the primary source of glucose, followed by alanine, glycerol, and glutamine [151].

The expression of several enzymes is necessary to bypass irreversible reactions found in glycolysis and utilize these substrates for *de novo* glucose synthesis. Phosphoenolpyruvate carboxykinase (PEPCK) converts oxaloacetate (OAA) to phosphoenolpyruvate (PEP), bypassing pyruvate kinase. PEPCK has both mitochondrial and cytosolic isoforms (PCK1/2 respectively).

PEP generated in the mitochondrial by PCK2 requires transport into the cytosol, likely through the citrate-malate transporter.

Two other irreversible reactions, phosphofructokinase and hexokinase, are subsequently bypassed through the expression of FBPase and G6Pase, respectively. Concurrent expression of these sets of enzymes provides opportunity for futile cycling, where the forward and reverse reactions may simultaneously occur, greatly reducing the net flux emerging from that reaction step. This may in fact be beneficial for cells, as it allows glucose uptake and release to be fine-tuned through substrate concentration and regulation of the members of these futile cycles [152].

Pyruvate is generated and consumed by multiple competing reactions in both cytosol and mitochondria. One example of this competition is that between pyruvate carboxylase (PC), which converts mitochondrial pyruvate to OAA which can be subsequently converted to PEP and exported from the mitochondria, and pyruvate dehydrogenase complex (PDHC) which converts pyruvate to acetyl-CoA. Such examples of substrate competition may also play key roles in directing flux for gluconeogenesis.

Models of liver metabolism have been developed to study liver metabolic processes. For example, a kinetic model of liver metabolism was built to study the mechanisms of glucose regulation through gluconeogenesis and glycogen metabolism, using liver flux data to set the enzyme maximal rates [153]. This model focused on the hormonal phosphorylation regulation by insulin, glucagon, and epinephrine on key enzymes, as well as quantifying the relative contributions of gluconeogenesis and glycogenolysis towards glucose release from the liver. Another model of liver metabolism including more metabolic pathways likewise fit relative enzyme activities to experimental data in order to simulate the liver response to typical metabolic situations encountered by the liver, interactions with specific drugs, and the effects of metabolic disorders [154].

In this work, we adapt a model of central metabolism to study the process of gluconeogenesis. Gluconeogenesis, however, is not a straightforward pathway, and expression of the key enzymes alone is insufficient to synthesize glucose. In order to determine the metabolic requirements of gluconeogenesis, rather than fit a specific dataset of liver metabolism, we extend an optimization framework we previously developed in order identify the subset of enzymes required to change in activity or expression to synthesize glucose from the commonly utilized substrates: glycerol, lactate, alanine, and glutamine. Using this optimization framework, we are able to identify commonalities in enzyme activity required for the different carbon sources. The identification of these key nodes in gluconeogenesis aids to the understanding of the pathways, what can go wrong in disease states, and suggests potential targets for treatment of diseases related to dysregulation of blood glucose concentration.

4.2. Methods

In this study, we employ a modified version of a model of central glucose metabolism which was previous used to study the Warburg effect using an optimization framework we developed [155]. The metabolic model is expanded in this work to contain the essential enzymes for gluconeogenesis, and the computational framework is adapted for use in studying liver metabolism.

4.2.1. Adaptations to the kinetic model

In this work, we extend this metabolic model to include the key reactions for gluconeogenesis: PCK1, PCK2, a mitochondrial PEP transporter (PEPX), G6Pase, and FBPase, as depicted in Figure 4.1 with reaction equations and kinetic parameters adapted from previous models of liver metabolism [153, 154]. We also adjust key enzymes such as HK4/GCK regulation through GKRP to have appropriate regulation for liver metabolism [153]. In this model HK4, PFKL, and PKL are considered the dominant isoforms for the optimization, but other isoforms may be expressed simultaneously. Given that the ATP/ADP ratio is expected to be highly different between glycolytic

and gluconeogenic cell types, a dynamic balance of ATP production and consumption was also added to the model. In this way, the ratio of ATP/ADP is adjusted depending on the relative mitochondrial activity to gluconeogenic rate, and relative increases in the ATP/ADP ratio can serve as a driving force for gluconeogenesis.

4.2.2. Optimization framework

Here we adapt and extend the optimization framework we previously developed for identifying essential enzymes to rewire metabolic behaviors [155]. In this case, our objective function to determine the enzymes changes required for different gluconeogenic rates is:

$$\min \left[r_{GLUT} + b \cdot \sum_j \left(\log_{10} E_j \right)^2 \right] \tag{16}$$

To achieve gluconeogenesis, we define glucose production as a negative value through the glucose transporter r_{GLUT}. The final term penalizes adjustments to the relative enzyme abundances (or kinetic rates).

At appropriate values of b, this ensures that only the required adjustments are made to metabolic enzymes, while the remained are unchanged if they do not meaningfully contribute to the objective function. In our previous work, an additional term was used to help guide the optimization problem to a steady state, as most randomly chosen starting points for optimization are infeasible if the steady state constraint is immediately imposed. However, because we use the BARON solver for this work, a feasible initial guess is not required to identify an optimum, and instead the steady state constraint can be directly applied:

$$\frac{dC_i}{dt} = 0 \; \forall i \tag{17}$$

As BARON is used for solution in this work rather than a local optimizer, the optimization is run once for each penalty value to find an appropriate penalty value by which to select enzymes. After selecting a penalty value in which the majority of enzymes have ceased to change without significant degradation of the gluconeogenic rate, the penalty value of each enzyme can be written as:

$$P_{E_j}(b) = \left(\log_{10} E^*_{j,k}\right)^2 \tag{18}$$

The penalty value represents the change an individual enzyme can take on for a given b before further changes yield diminishing returns on the objective of producing glucose. A full explanation of the framework and its implementation can be found in our previous paper [155].

In the second stage of optimization, the penalty weight is set to zero, and only those enzymes which contribute substantially to the penalty are kept:

$$P_{E_j} \geq \delta \tag{19}$$

Then, all enzymes aside from those selected are fixed at their original values of 1. Then, in a series of optimizations, all possible combinations of enzymes are tested from this subset in increasing number: choose 1, choose 2, and so forth. In this way, we see what synergistic effects may exist to identify different potential routes of gluconeogenesis and also which combinations and expressions are most effective in altering the gluconeogenic rate.

4.2.3. Key modeling constraints

In this work, we assess the requirements for different gluconeogenic pathways by adjusting constraints for three total optimization problems:

1. All carbon sources: lactate, alanine, and glutamine are allowed to be used for *de novo* glucose synthesis.

2. Lactate: only lactate can be consumed, the rates of amino acid consumption are fixed at zero.
3. Amino acids: only alanine and glutamine can be consumed, the rate of the consumption of lactate is fixed at zero.

In all three of these problems, we consider that PDHC and PC undergo reciprocal regulation [156], and so the activity of PDHC is set to zero, and acetyl-CoA is supplied to the TCA cycle through the breakdown of fatty acids. Additionally, enzymes are generally allowed to vary 100-fold from their initial values:

$$0.01 \leq E_j \leq 100 \tag{20}$$

Except in the case of non-expressed isoforms, whose expression is normalized to the dominant isoforms expression, but subtracted by 1 so that an unchanged enzyme value of 1 corresponds to no expression, and the bounds are set as:

$$1 \leq E_j \leq 100 \tag{21}$$

Additionally, as NAD^+ and NADH are considered to interchange from a fixed total pool, the total concentration of these two key redox molecules is set as:

$$0.1\ mM \leq NAD^+ + NADH \leq 0.6\ mM \tag{22}$$

Additionally, the concentrations of metabolic species are given the following bounds:

$$10^{-6}\ mM \leq C_i \leq 50\ mM \tag{23}$$

We also consider the breakdown of 16-carbon chain fatty acids, providing a stoichiometric amount of glycerol to the rate of acetyl-CoA generation in the mitochondria from fatty acid breakdown:

$$24\, r_{glycerol} = r_{FAO} \tag{24}$$

4.2.4. Modeling and package details

This model was written in Python 3.7 using packages from the Anaconda distribution (https://www.anaconda.com/). In addition to these packages, we use Pyomo for optimization [157, 158] with the BARON solver [159], using IPOPT as the NLP subsolver [160]. This model has 175 constraints and 248 variables. This optimization was performed on a combination of a desktop using an AMD Ryzen 7 processor, and on AMD EPYC 7702 processors using resources from the Minnesota Supercomputing Institute.

4.3. Results

4.3.1. All enzyme optimization

In this optimization, 65 enzymes were allowed to fluctuate in their maximal reaction rate, through adjustments of the E_j parameters. By varying the penalty weight b over several orders of magnitude, different stages of the optimization emerged. As can be seen in Figure 4.2 shows an analysis of all three problems stated in the Methods, considering the potential glucose carbon substrates all together, and each separately. Distinct patterns emerged from the enzymes in response to adjustment of the penalty weight. At low penalty weights, many enzymes take on substantial penalty. As this penalty is increased, many enzymes return to their default expression $E_j \approx 1$, and few retain their altered expression. However, until moderate values of the penalty weight, the gluconeogenic rate is not degraded from the optimal values seen

Each of the individual carbon sources appears to be capable of producing approximately one-third of the total gluconeogenic flux found in the first problem. Figure 4.2A shows the penalty value, P_{E_j}, for each enzyme. The majority of enzymes have large changes (high P_{E_j}) at very low penalty weight value. Most of these changes diminish to near zero, or their original expression/activity value at no cost in terms of gluconeogenic rate as shown in Figure 4.2B, by around $b = 10^{-1}$. At high penalty weights, $b > 1$, the gluconeogenic rate falls of rapidly, as

enzyme changes essential towards achieving gluconeogenic flux are eliminated. Each of these problems is able to achieve a relatively consistent gluconeogenesis rate until penalty weights greater than 1 are applied, similar to the first problem, after which the rate falls off dramatically.

4.3.2. Identification of essential gluconeogenic enzymes

A penalty weight of $b = 1$ was then chosen for further analysis, as the highest penalty value without a substantial decrease in gluconeogenesis rate. The penalty value of each enzyme for all three problems is shown for $b = 1$ in Figure 4.3, ranked by penalty value in Problem 1. It can be seen by the tail-off in P_{E_j} at high b, that these problems share significant commonality in the required enzyme expression changes for gluconeogenesis, while most enzymes can be left unchanged. A filter was then applied to remove enzymes which have insignificant penalty value for all three problems, leaving only a subset of enzymes which have any appreciable change required for gluconeogenesis.

4.3.3. Testing combinations of gluconeogenic enzymes

Once the essential enzymes required for gluconeogenesis were identified, the second stage of optimization was performed with the penalty weight set to zero ($b = 0$), while only the enzymes identified above were allowed to vary in small groups. In order to comprehensively explore the potential synergistic effects of different optimization, all possible combinations of enzymes were tested from this initial list in sets of increasing size, using the no enzyme set as the base-case. These results are presented in Figure 4.4, which shows the results of these simulations, grouped by number of enzymes optimized, and ordered by ascending rate of gluconeogenesis.

Larger enzyme combinations were able to achieve higher gluconeogenic rates than smaller sets, however within each group relatively few enzymes stood out with the highest gluconeogenic rates, with the majority being significantly lower. This has two implications: (A) relatively few enzymes directly contribute towards the gluconeogenic rate, (B) enzymes have significant

interdependence in their contribution towards gluconeogenesis. In other words, some enzymes may require concurrent change in other enzymes rather than acting independently to achieve meaningful increase in the rate of gluconeogenesis.

4.4. Discussion

In this work, we have extended an optimization framework to identify key enzymes required for gluconeogenesis. Using optimization, we have reduced a list of 65 possible candidates to approximately 10, which can be further evaluated to identify only the essential changes required for utilization of the different common carbon substrates for *de novo* glucose synthesis.

By comparing results of the different problems, we were able to highlight enzymes which were required for each or any of the problems. Besides those enzymes which directly provide carbon substrate, highly altered enzymes include PCK1/2, PEP transporter (PEPX), ME1/2, ALD, GAPDH. The use of either PCK1 or PCK2 and the PEP transporter, appeared to be problem dependent, whether the carbon source originated from lactate in the cytosol, or from amino acids, which can be supplied in either the mitochondria or the cytosol. Gluconeogenic flux may also be routed through the mitochondria in some cases due to kinetic limitations and competing reactions for the cytosolic pyruvate pool. Additionally, both cytosolic and mitochondrial malic enzyme were found to have significant penalty weight. In this case, both enzymes were decreased from their nominal value, indicating that the additional competition they provide to the pyruvate pool against pyruvate carboxylase may be detrimental to gluconeogenesis.

In the second optimization stage, combinations of enzymes and their expression level changes are required for gluconeogenic rate were assessed. By significantly reducing the number of enzymes to test in the first stage, it become computationally tractable to test all possible combinations of enzymes. This examination revealed only specific enzyme combinations enabled

the greatest gluconeogenic rates, which further implies a strong interdependence between enzymes in their contribution towards gluconeogenesis.

4.5. Conclusion

In this work, we have applied and extended an optimization framework towards identifying the requirements for the important hepatic role of gluconeogenesis. As dysregulation of glucose control in the body is the cause of several diseases, it is of key interest to identify metabolic processes and targets for disease treatment. Here we identify enzymes in central metabolism that may play a key role in gluconeogenesis starting from the substrates of glycerol, lactate, and the amino acids alanine and glutamine.

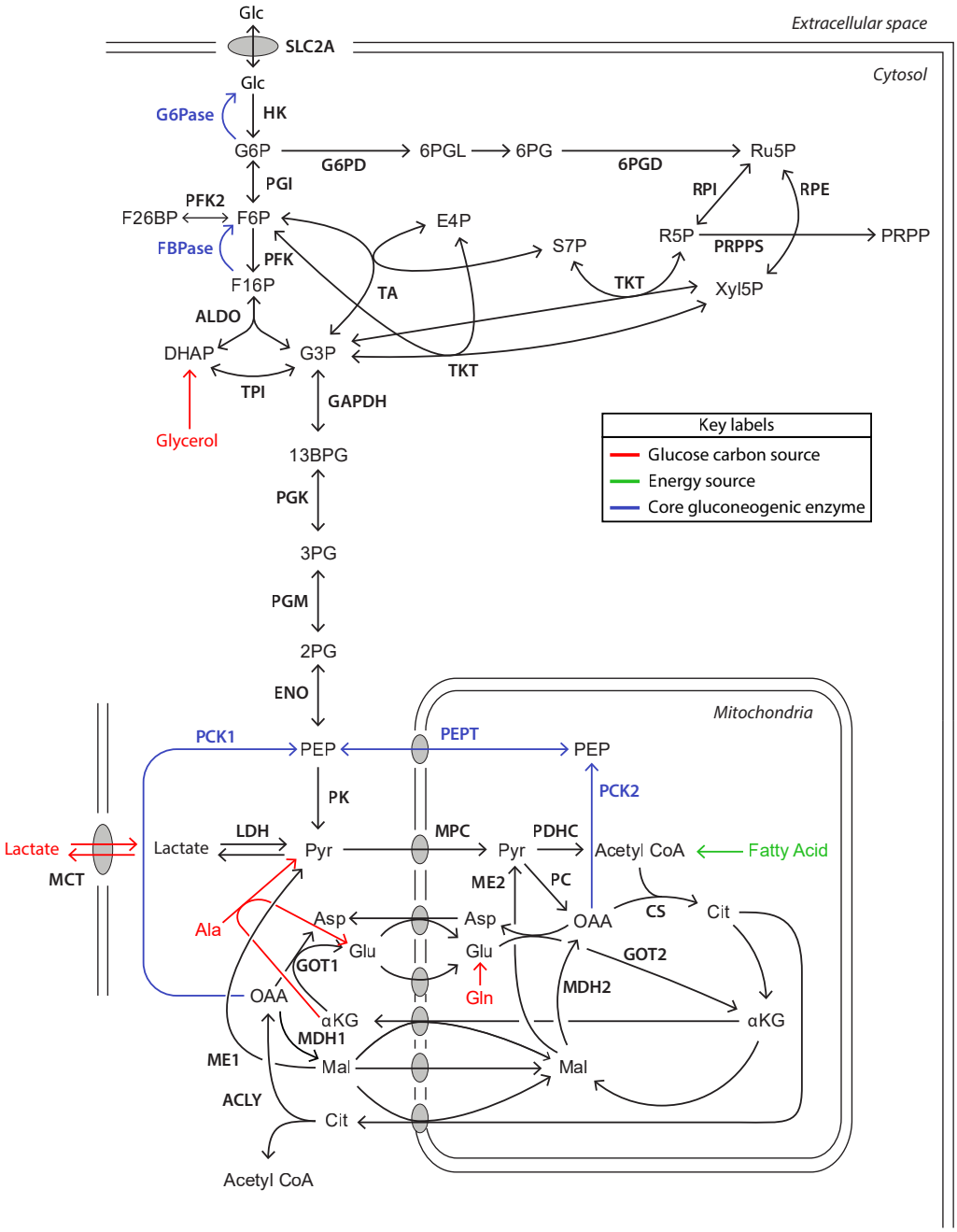


Figure 4.1: Kinetic model of gluconeogenesis.

Depicted are the central pathways of metabolism in the model, included glycolysis, the pentose phosphate pathway, and the TCA cycle. This model has been extended to include the key enzymes

in gluconeogenesis: PCK1/2, PEP transport, FBPase, and G6Pase. Additionally, glycerol, lactate, glutamine, and alanine are considered as carbon sources for gluconeogenesis. Fatty acid degradation is considered for supply of Acetyl-CoA in the mitochondria for energy generation.

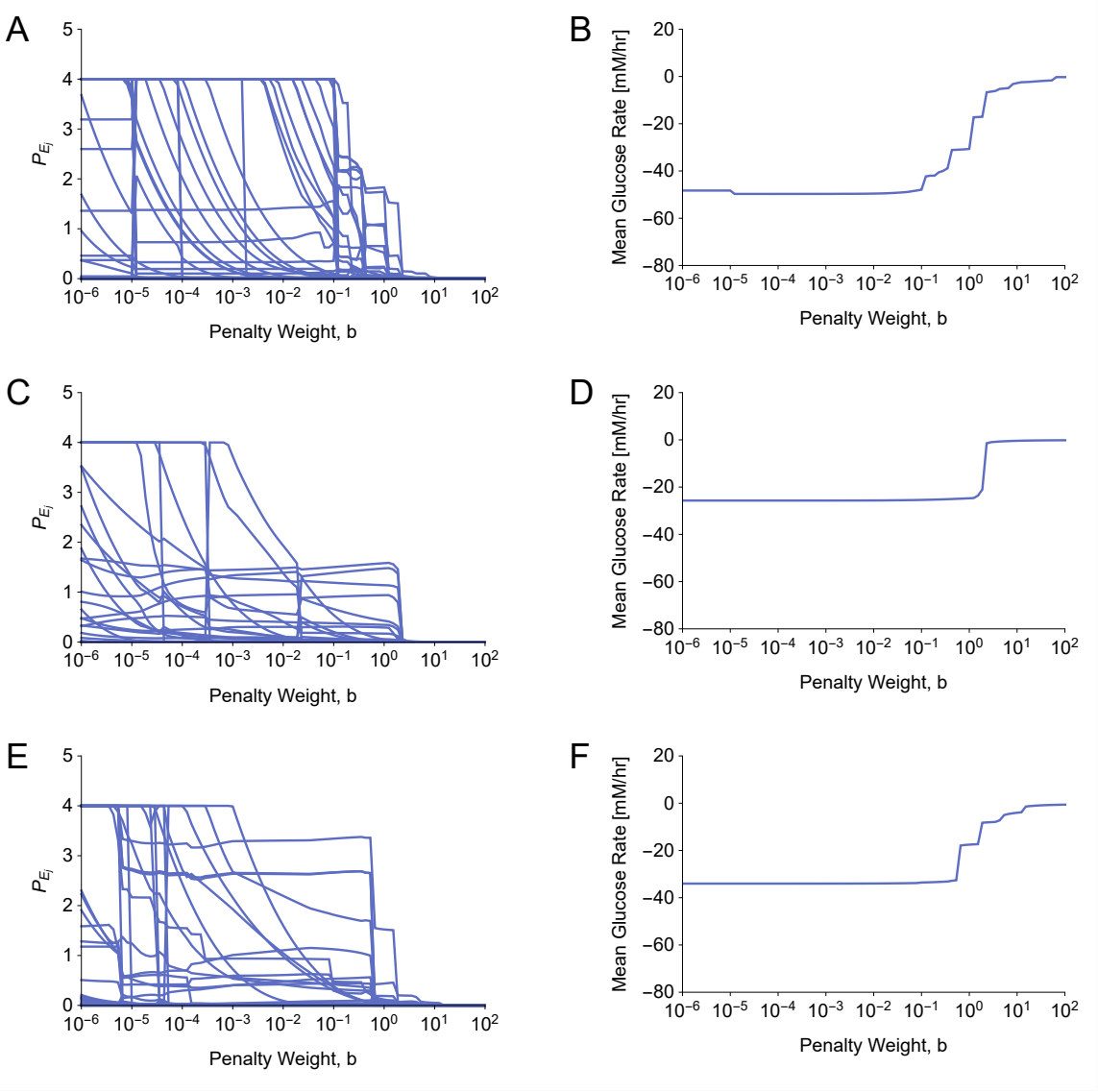


Figure 4.2: Rate of gluconeogenesis and enzyme penalty values over different values of b.

Shown are the enzyme penalty contributions P_{E_j} for all 65 enzyme expressions optimized, and their corresponding average rates of gluconeogenesis across the local optima found. Different carbon substrates are considered for four problems: (A,B) all substates, (C,D) lactate only, and (E,F) amino acids only.

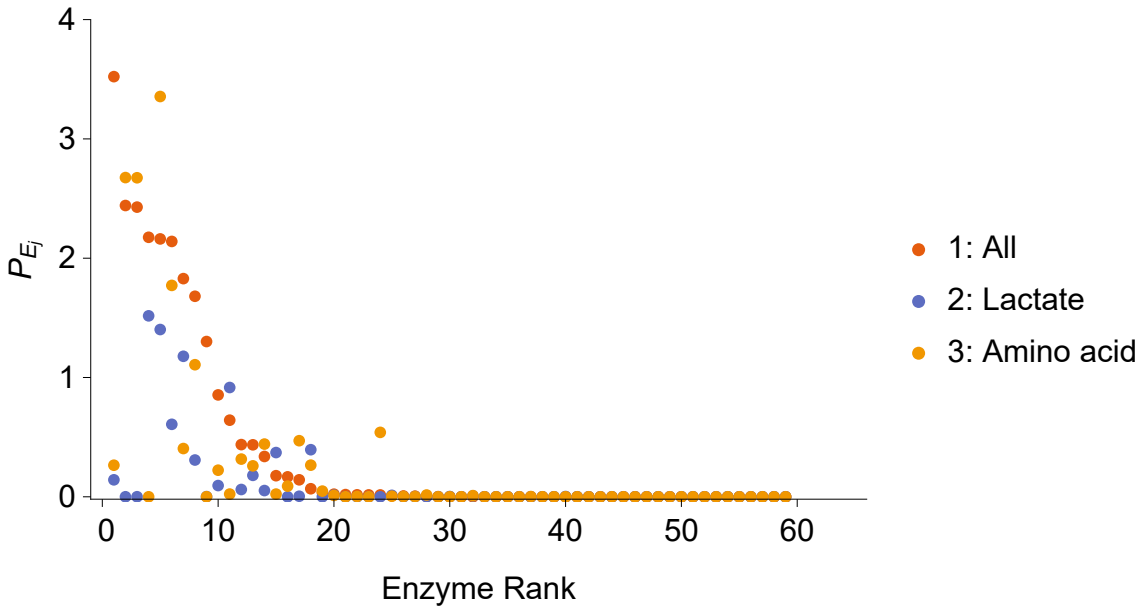


Figure 4.3: Enzyme penalty values at a fixed b.

The average penalty contribution for each enzyme at $b \approx 0.15$ is shown. Shown are the penalty values for each of the three problems. The enzymes are ranked by their penalty values for the all carbon source problem.

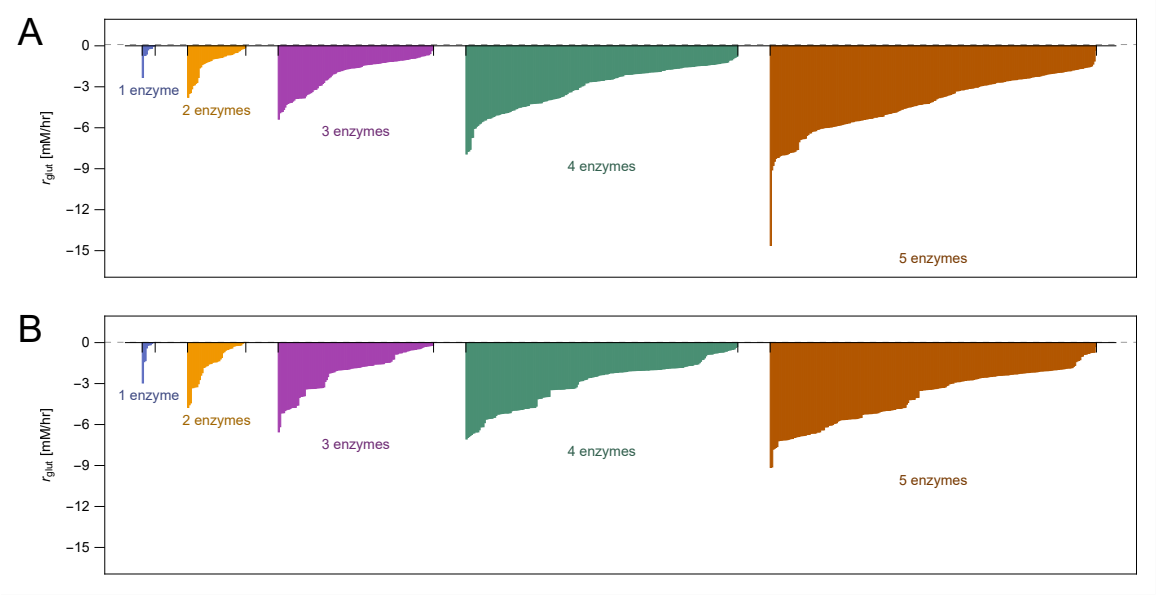


Figure 4.4: Testing enzyme combinations from the highly ranked enzymes.

The optimal gluconeogenic rate achievable from different numbers of enzymes allowed to change selected from the 10 enzyme subsets for the lactate only problem (A) and the amino acid only problem (B). The horizontal dashed line shows baseline case, taken to be the rate of gluconeogenesis achieved with no other enzymes allowed to vary.

5. A hybrid first principles-empirical bioprocess model for in silico process optimization

Reproduced from O'Brien, C. M., Zhang, Q., Daoutidis, P., & Hu, W. S. (2020). A hybrid first principles-empirical bioprocess model for in silico process optimization (In Preparation).

5.1. Introduction

Pharmaceuticals known as biologics, including protein therapeutics (e.g. blood clotting factors and antibodies), cells for cell therapy, and viruses for vaccination and gene therapy, are often produced by cell culture processes. The quality attributes for biologics are far more complex than for chemical drugs. The manufacturing process for clinical trials and commercial production must ensure consistent product quality and productivity in a product's life cycle. Product quality variability may arise in the production or cell culture process, in downstream recovery, or in the final product polishing. Minimizing the variability of the cell culture process is crucial to controlling the product quality. Unfortunately variability can occur when some processes are carried out in different sites, different reactors, or even over time in the same site and the same set of reactors [161]. Process variability may manifest in altered cell culture characteristics, including metabolism. Metabolic state is further linked with the time course profiles of cell growth, metabolites, and product [162], and affects product quality [11, 163]. There are many possible causes of process variability: fluctuations or change in sources of raw materials, operational variability, genomic or epigenomic alterations in the cell, and physical differences in reactors at different sites.

One of the key factors influencing process variability in cell culture is metabolism. The metabolic state of cells changes over time in a typical culture. In the early stage of rapid cell growth, cells also rapidly consume glucose which results in high lactate production. As the growth slows

down the metabolism may stay in a high glycolytic flux state or transit to a low glycolytic flux state with a reduced glucose consumption and little lactate production. The latter may be either producing or consuming lactate at a low specific rate, depending on the cell line and the process conditions. In a case study of manufacturing data obtained from the same process carried out at the same site, the time profiles of key process variables behaved rather similarly in the early stage, however, in some runs the mid-stage low flux metabolism returned to a high flux state, while in other cases it transitioned to lactate consumption [102].

In a non-steady state operation, metabolism drives growth, nutrient consumption, accumulation of waste metabolites, and pH change in the media. The reactor control algorithm responds to the environmental changes with actions that correct the deviation from the set point of the controlled variables. Meanwhile, other uncontrolled variables inevitably change. For example, addition of pH controlling agents increases the osmolarity in the media. This environmental change leads to alteration of metabolism and growth, which are linked in a vicious cycle with reactor control [164-166]. This cascade of the interactions of growth, metabolism, reactor control actions, and chemical environmental changes are likely to be the culprit of process variability observed even for manufacturing runs executed under the same operational conditions. Understanding the factors which lead to changes in process outcome and controlling them will enhance process robustness and product quality.

With the complexity of the interactions between these systems and the cascade of events occurring in a cell culture process, a systems approach of *in silico* modeling, simulation, and optimization framework will provide better process understanding. However, modeling a system of such complexity that incorporates the key processes of cell growth, metabolism, signaling, and the reactor environment presents significant challenges. With respect to central metabolism, a first-principles model of the allosteric regulation developed and applied to depict the shift of metabolic

states in cell culture [167]. The model was later extended to a bioreactor system model that considers metabolism, growth and concentration changes of glucose and lactate and used to illustrate metabolism driven occurrence of multiple steady states in continuous culture [145].

The nonlinearity and stiffness of mechanistic metabolic rate equations makes working with these types of models numerically challenging. Additionally, while a mechanistic model can be developed to describe cell metabolism, models describing how growth rate responds to the chemical environment and how metabolism responds to growth rate are unavoidably empirical. Consequently, previous bioprocess modeling approaches have used simplified, empirical descriptions of metabolism and growth to describe bioprocesses [168-171]. Some models have been further linked to other key pathways, such as glycosylation [172]. Different types of kinetic modeling for bioprocesses has been previously reviewed [173]. Empirical models require careful selection of rate laws to fit the observed data from a number of candidate functional forms [174]. Such approaches reduce computational difficulty but rely on functional forms to be fit to the experimental data in addition to identification of model parameters. Model reduction can be employed to further increase model tractability [175]. Even further abstracted from the biological mechanisms are data-driven, or black-box, models [176]. A hybrid of mechanistic and black-box models, referred to as “grey-box” models [177], can have a larger scope which has more capacity to extrapolate and test new process conditions. The different types of cell culture models and areas of bioprocessing where they may be very helpful has been recently reviewed [178].

Bioprocess modeling also faces the problem of limited availability of datasets. Before the product is launched commercially the number of runs is small. Once commercialized, the quantity of data at the manufacturing scale accumulates over time. However, the number of runs is still typically small relative to the number of parameters that are typically fit during bioprocess model construction. Workflows have been developed to improve parameter identifiability [179], but

significant challenges remain from the relative scarcity of data. Furthermore, such datasets are rarely available to the researchers.

In this paper, data taken from a previous work [102] was used to construct an *in silico* hybrid mechanistic-empirical bioprocess model. This model includes central metabolism, cell signaling, cell growth, and the reactor environment including carbon dioxide accumulation and pH control. The processes exhibit a degree of variability and a range of productivity. The inherent complexity of behavior creates difficulty in model fitting, but also increases parameter identifiability and reduces overfitting. By using first-principles mechanistic components where possible, the resulting model has a greater capacity to extrapolate to new conditions than a purely empirical model. This hybrid model is used to develop hypotheses regarding the origins of this process variability, explore operating conditions which may be altered to improve the process outcomes, and study process scaling through the effects of mass transfer limitations on the reactor environment.

5.2. Methods

Data used for this study includes 243 manufacturing runs of CHO cell fed-batch culture at the 12,000 liter scale published in a previous work [102, 180]. Runs were ranked by the product titer and the time profile of some process variables of the top 20 (blue) and bottom 20 runs (red) are shown in Figure 5.1. These runs exhibit metabolic dynamics and variability among different runs which emerges in the later stages of the runs with some runs shifting to lactate production, despite starting with relatively similar conditions. Note that maintaining the lactate consumption low flux state through the end of the culture is correlated to a higher product titer.

5.2.1. Treatment of experimental data

Data was processed using functions available in Mathematica to remove outliers and smooth the curves. Outliers were identified using the anomaly detection algorithm AnanomalyDetection, which fits a statistical distribution to the data at each timepoint and then removes the outlying

points. The missing points were then reconstructed using the SynthesizeMissingValues method, which similarly constructs a statistical distribution, but as a time-series rather than at each point separately, from the remaining data. Finally, a Gaussian blur was applied to provide a minimal amount of smoothing while maintaining real trends present in the runs. The data treatment process is illustrated in Figure 5.2. From treated data, specific growth rate was calculated to provide bounds for the cell growth portion of the model for both maximal growth rate and death rate.

5.2.2. Construction of the bioprocess model

The cell bioprocess system consists of a biotic phase of cells and an abiotic phase of culture fluid and its chemical contents and a gas phase. Cell metabolism occurs in the biotic phase whose volume fraction changes with the culture time due to growth and death. The biotic and abiotic phases interact with each other through nutrient supply to biotic phase and excretion of metabolites to the abiotic phase. The abiotic phase interacts with the environment outside the bioreactor system through the addition of feed nutrients and base for neutralizing pH, and addition or removal of CO_2 through aeration. The combined model thus consists of three component models: (1) a mechanistic metabolism model, (2) a cell growth model, (3) a reactor environment model, as shown in Figure 5.3 and detailed below.

5.2.3. Mechanistic Metabolism model

The metabolism model was based on a mechanistic kinetic model described previously [167]. The model consists of the pathways of glycolysis, the pentose phosphate pathway, the tricarboxylic acid cycle, and the malate-aspartate shuttle and includes the known allosteric regulations as depicted in simplified form in Figure 5.3B. The values of kinetic constants were derived from literature as described previously. The parameters corresponding to the expression level of enzymes are not known and vary by cell line. The values reported previously were used as the starting point for parameter estimation as discussed in the later sections. The metabolism model comprises 42

species and 50 reactions. As the experimental dataset used for this study is a fed-batch culture, glucose feeding was implemented in this model. To match the initial feed, 22 mM glucose was added at 90 h, and then in the late stage culture 10 mM glucose was added any time the concentration dropped below 15 mM, to approximately capture the observed feeding.

5.2.4. Cell signaling model

In addition to allosteric regulation the glycolytic flux is regulated by cell growth rate via the adjustment of the kinase and phosphatase activities (herein referred to as K/P ratio, where K is the maximal reaction rate of the forward reaction for forming F26BP, while P is that of the reverse reaction) of PFKFB by Akt and AMPK [181-183]. Akt is phosphorylated under fast growth and phosphorylated Akt activates K/P and increases glycolysis flux. In the model Akt activity is taken to vary proportionally to specific cell growth rate [184], using the form:

$$\alpha_{AKT} = \frac{\mu}{\mu + \mu_{max}} \tag{25}$$

AMPK responds to environmental stress conditions such as nutrient starvation, redox imbalance and osmolality, and activates glycolysis by adjusting the K/P ratio of PFKFB [185] which is given the form:

$$\alpha_{AMPK} = \left(1 + K_{AMPK}\left(\frac{Os + K_{AMPK,os} * (Os - Os_I)^2}{Os}(1 + [I]^n)\right)^{-1}\right)^{-1} \tag{26}$$

It has been reported that growth inhibitory metabolites accumulated at the high cell concentration practiced in contemporary CHO cell culture. Many of these metabolites originate from pathways involving aromatic and branched chain aliphatic amino acids [78]. We thus model the production of those not yet completely identified inhibitory metabolites as an inhibitor (I) whose concentration increases at a rate proportional to cell concentration:

$$\frac{dI}{dt} = C_{cell} r_I \tag{27}$$

The incorporation of unknown inhibitory as well as activating secreted products has been incorporated previously in the modeling of CHO cell growth [186]. Here, r_I is a constant rate of production of inhibitory compounds and is a fitting parameter of this optimization. AMPK and Akt thus provide positive and negative feedback control on PFKFB with a net multiplicative effect:

$$\frac{K}{P} \propto \left[\frac{K}{P}\right]_0 \alpha_{AKT} \alpha_{AMPK} \tag{28}$$

5.2.5. Growth model

The cell growth model considers both growth and death. The empirical growth model was based on Monod kinetics as was the model employed previously [187]. The growth rate is inhibited by lactate and decreases with excessive osmolality. But its dependence on glucose was neglected as the glucose concentration in culture stays high so that the growth rate is relatively insensitive to its fluctuations [188]. As described in the previous section the growth rate is assumed to be affected by the accumulation of the inhibitory metabolite I:

$$\mu = \mu_{max} \frac{K_{elac}}{K_{elac} + [lac]_e^2} \frac{1}{1 + [I]^{I_{exp}}} \frac{Os}{Os + K_{osmo}\,(Os - Os_I)^2} \tag{29}$$

Cell death is assumed to vary inversely with the specific cell growth rate, with a rate constant of zero at maximum growth, and $k_{d,max}$ when cell growth in fully inhibited:

$$k_d = k_{d,max} \frac{\mu_{max} - \mu}{\mu_{max}} \tag{30}$$

5.2.6. Reactor environment model

The reactor environment model describes the chemical dynamics of the abiotic phase of the culture fluid including the gas phase that is traveling through the bioreactor. The model consists of

the equations balancing the consumption and production of major nutrients (in this study glucose) and metabolites (lactate, CO_2 and its hydration products). The balance of CO_2 considers production from metabolism, exchange through interfacial transfer, hydration reactions and sodium carbonate addition. The reactor fluid, both liquid and gas phase, was assumed to be well mixed and an average overall mass transfer coefficient ($k_L a$) is used to represent the interfacial transfer characteristics of the entire reactor.

pH is maintained by base addition (sodium carbonate) that balances CO_2 hydration reactions, lactate production or consumption and base addition for pH control. Equilibrium among these species was assumed. The buffering effects of other medium components and cell mass was neglected. The model thus captures the major effector of pH change, but not other minor components. A proportional-integral (PI) controller is used to maintain pH at a constant level. The controller actions are to increase CO_2 level in the gas phase when the pH is above the set point and to add sodium carbonate when the pH drops below the set point.

Additionally, the dynamics of osmolarity resulting from changes in the total sum of dissociable species caused by metabolism, CO_2 exchange, and base addition for neutralizing pH was also considered. The osmolarity is estimated by adding to the initial experimentally measured value the contribution from metabolism (i.e. glucose, lactate), accumulation of CO_2 and related species, and a consumption term proportional to cell concentration and growth to account for consumption of metabolites (e.g. amino acids, phosphate) not considered in the model and incorporation into biomass. Since the medium is close to an ideal solution, the value obtained provides a good estimation.

5.2.7. Model integration

The component models (the mechanistic metabolism model, the empirical growth model and signaling model, and the mechanistic reactor environmental model) were integrated into a hybrid

mechanistic/empirical model by considering their interactions. Metabolism and growth are the major driver of environmental change in the bioreactor by consuming glucose and excreting lactate and CO_2. These chemical changes along with the reactor control actions of CO_2 sparging and base addition alter the osmolality. The fluctuations in the bioreactor environment trigger growth rate change, which in turn alter metabolism. Such reciprocating interactions among component systems give rise to the dynamic culture behavior that is captured by the model.

5.2.8. Parameter estimation and bioprocess simulation

A number of the model parameters related to the cell growth, the effect of signaling on K/P and several key enzyme concentrations, were estimated from the experimental data by minimizing the residual between the measured and simulated time profiles of glucose, lactate, osmolarity, and VCD. To reduce the size of the parameter estimation problem, the growth and metabolism portions of the model were fit separately before being combined into the process model. The parameter estimation was performed in MATLAB using methods from both the optimization and the global optimization toolboxes. The model fitting process is depicted in Figure 5.4.

As the cell growth model is small and non-stiff, parameter estimation was performed using the global optimization algorithm globalsearch to obtain initial values for 7 parameters by simulating for all 243 runs of data. The experimental time course data for glucose, lactate, and osmolarity served as model inputs, as well as parameters obtained from fitting the initial cell concentration model. The objective function was set to be the minimum sum of squared residuals comparing the model and experimental runs, with both datasets interpolated to time points at every hour.

To fit the metabolic model, given the complexity of the low titer runs which shift to lactate production, the initial goal was to recreate these two shifting behaviors. An initial screening was performed using a Latin-hypercube sampling (LHS) of 50,000 points for the 21 parameters to fit

for the metabolic model. The metabolic model was simulated with each of these 50,000 points, using time profiles for experimental VCD and osmolarity, as well as the starting concentrations of glucose and lactate as model input. For this initial case, only the average of the bottom 20 runs was used as input. To capture the dynamics of the three distinct phases of the experimental lactate profiles, the model slope of extracellular lactate at 10 h, 100 h, and 250 h was used to screen parameter sets from the LHS.

Initially, to reduce the size of the parameter space, parameters for which the lactate slope at these three timepoints was insensitive were eliminated from the design space. Further, the range over which the enzymes varied was significantly reduced to more densely sample the remaining parameter space. This eliminated six out of the 21 parameters from the parameter space.

A second LHS of 50,000 parameter sets was then sampled as above, but on the reduced parameter space. Then, the following criteria were applied to determine suitable starting points (cultures that start and end with lactate production, with a lower flux state in the middle of the culture) for local optimization:

$$r_{lac}|_{t=10} > r_{lac}|_{t=100} \wedge r_{lac}|_{t=100} < r_{lac}|_{t=250} \wedge r_{lac}|_{t=10} > 0 \wedge r_{lac}|_{t=250} > 0 \quad (7)$$

Finally, local optimization was performed using fmincon, regressing the extracellular glucose and lactate profiles to the experimental data for the average of the 10 lowest titer runs, as discussed above. This regression was performed over the first 160 h of culture to ensure that the three distinct metabolic phases of the culture were maintained.

Finally, the models of cell growth and metabolism were combined, and the simulation of osmolarity and pH was added. This combined model was fit first using local optimization and finally local sensitivity analysis was used to fine tune parameters from the input models against an average high and low run (calculated by averaging the top 20 and bottom 20 runs, respectively). The combined model has a total of 25 parameters which were estimated using the experimental

data, not including the six parameters left at their default values after elimination in the metabolic model fitting.

The simulations were performed on a mix of desktop computers using AMD Ryzen 7 and Intel Core i7 processors and using resources from the Minnesota Supercomputing Institute (MSI).

5.3. Results

5.3.1. Model parameter estimation using process data

The growth model is virtually segregated from other component models since the growth rate is only affected by but does not affect directly the environmental variables in the bioreactor. Hence, the VCD data from all 243 runs were used to fit the parameters related to the effect of lactate, inhibitor and osmolality in the growth model, without the need to simulate the metabolic model. A plot of the specific growth rate as a function of lactate concentration and osmolarity as predicted by the final parameters from the fitted model is shown in Figure 5.5.

A number of factors may affect the metabolic shift from a high glycolytic flux state to a low flux state, including allosteric regulation of enzymes, growth rate and the accumulation of lactate in culture [189]. The expression levels of glycolytic enzymes and other elements regulating glycolytic flux vary in different cell lines which may contribute to different dynamics of metabolic state among different cell lines [190, 191]. Hence the level of glycolytic enzymes, pyruvate dehydrogenase, and glucose-6-phosphate dehydrogenase in the metabolism model are included in the parameters to be optimized to capture the characteristics of the manufacturing runs. Glycolysis enzymes were considered because of their role in allosteric regulation of the metabolic state, while the pyruvate dehydrogenase complex was previously determined to have a significant impact on the metabolic state [155], and glucose-6-phosphate dehydrogenase is a key step controlling flux to the pentose phosphate pathway [192]. The concentrations of all other enzymes and transporters were kept constant. Additional fitting parameters include the pool size for the key redox molecules

NAD^+ and NADH and the empirical parameters relating to the action of Akt and AMPK on the K/P ratio of PFKFB as well as the initial K/P ratio for each culture.

The average of the bottom 20 runs were used to fit the parameters of the metabolism and signaling models after an initial LHS screening as discussed in the Methods section. During the LHS screening, the action of six parameters were determined to not have a significant impact on the metabolic state, and they were fixed at their original values, reducing the complexity of model fitting and increasing parameter identifiability.

As can be seen in Figure 5.1A the starting culture conditions exhibited some degree of variability. The variability reflects the operational difference in initial glucose and the amount of cells and lactate carried over from the preceding seed culture (the N-1 reactor of a seed train). Additionally the state of metabolism and growth of cells from the N-1 reactor may also vary. In a previous study on the same set of data, it was shown the culture characteristics of the seed train were correlated to the final product titer [102]. We hypothesize that the characteristics of the seed culture manifested in its metabolic state and growth rate at the time of transfer to the production culture. We further hypothesize that the metabolic state is reflected in the K/P value of PFKFB. The initial value of K/P of each run was thus chosen as a parameter to be obtained by fitting the manufacturing data. The K/P ratio affects the F26BP level which regulates the extent of activation of PFK and plays a pivotal role in controlling glycolysis flux. K/P for PFKFB is regulated by Akt reflecting the effect of growth rate on metabolic activity, and by AMPK as a downstream consequence of stress. Depending on the state of seed culture at the time of inoculation K/P may vary.

By setting K/P as a fitted parameter the model presents the effect of initial conditions on the metabolic outcome better. As shown in Figure 5.1C PCA on the initial conditions (concentrations

of cell, glucose, lactate and osmolality) did not segregate of top and bottom runs. In contrast by including the fitted K/P value, the top runs and bottom runs become well separated.

The optimized parameter values for the growth, metabolism and signaling models were then used as the initial parameter estimate in the subsequent parameter value optimization of the combined systems model. The final systems model parameters were obtained using local optimization followed by local sensitivity analysis using the sets of parameters obtained from the initial parameter optimization as input against the top 10 and bottom 10 runs, by titer. The values of these parameters at each stage are listed in table 5.1.

Table 5.1: Table of fitted parameters

	Stage 1	Stage 2	Stage 3
Parameter names	Growth only	Metabolism only	Systems model
mumax	0.027556275		0.031281425
kdmax	0.01478456		0.009817326
Kelac	401.7717894		6057.952037
n	1.104879891		1.422516545
qI	0.002800218		0.003235636
Kos	0.01662589		0.002296431
Os_I	275.001034		286.4204429
Kcellosmo			10.56
kla			132
qcell		0.001903858	0.001931022
Kampk		271.4802861	334.0053726
Kstress		1.10019837	2.158958976
E0perm		10.23919318	1
E0hk		1.295668885	0.306258642
E0pgi		16.47186531	3.032091283
E0pfkfb		1	1
E0pfk		1.061878445	26.02462719
E0ald		1.256256509	2.887318955
E0tpi		1.755544077	0.254882323
E0gapd		21.2414329	2.432657781
E0pgk		1	1
E0pgm		1	1
E0en		1	1

E0pk		0.167404949	0.19900826
E0ldh		1.928486827	1.962100387
E0mct		2.465842908	1
E0g6pd		1	1
E0pdhc		1	1
NAD^+/NADH pool		0.336992655	0.48
KBP (low)			98.88
KBP (high)		211.5260085	213.84

5.3.2. Simulation of process behavior and divergent metabolic behavior

The optimized parameters were applied to simulate the kinetics of growth and metabolism in the reactor using the initial concentration of cell, glucose and lactate and the optimized K/P for each of the top 20 and bottom 20 runs. The simulated time profiles are shown in Figure 5.6A. Importantly, the model is able to reproduce the key characteristics of the experimental metabolic data: (1) an initial period of high flux glucose metabolism accompanied by lactate production, (2) a shift down in the early phase of culture to a low flux state with reduced lactate production, and (3) divergence of metabolic behavior among runs: some maintain a low flux metabolism ranging from lactate consumption to very slow lactate production, while others shift back to a high flux state and produce large quantities of lactate. The time profiles of some process variables related to pH control and CO_2 are shown in Figure 5.7.

The parameter values estimated using only the top 20 and bottom 20 runs were used to simulate the behavior of the intermediate runs. For this test the initial values of cell, glucose, lactate concentration and osmolality of each run were used. For the initial K/P ratio a linear interpolated value of the top and the bottom value by rank of product titer was used as a first approximation. As shown in Figure 5.6B, the model simulation captured the intermediate behavior observed experimentally, with runs of low to moderate lactate production dominating the middle runs.

Among the parameters, the K/P ratio of PFKFB plays a particularly pivotal role in shaping the dynamics of metabolic behavior as described previously. It regulates F26BP level which activates the PFK activity. The product of PFK, F16BP, activates PKM2 downstream thus amplifies the

glycolytic flux. To further understand the mechanism leading to different metabolic behavior between the top and bottom runs, the contribution of Akt and AMPK, as well as K/P over the time course of the culture are shown in Figure 5.8. As the growth rate decreases over time K/P decreases with decreasing Akt activation. However, as osmolarity increases, especially upon feeding of nutrients, AMPK activation of PFKFB becomes prominent, resulting in an increase in K/P. The goodness of fit of the model is shown in Figure 5.9.

5.3.3. In silico experimentation – simulation of process alterations

The systems model was then used to investigate the effect of the alteration of some operating conditions on the process performance. In a previous study using the same set of data it was revealed that the process performance (product titer) is related to the initial conditions of the production reactor run, or the final conditions of the inoculation seed (N-1) reactor [102]. In our model the initial conditions of a culture also direct it to a different process outcome. Experimentally the initial osmolarity and lactate concentration of a culture can be modulated by partially removing or exchanging the culture fluid from the seed culture. We hence tested the effect of reducing the initial osmolality and lactate level on process performance.

Reducing the initial lactate level for the top 20 and bottom 20 runs to nearly zero did not affect the metabolic behavior. Both top and bottom runs maintained their overall time profile of major culture characteristics (Figure 5.10B). In contrast, a reduction in the initial osmolarity (Figure 5.10C) by 5% (top) and 10% (bottom) changed the behavior of some of the bottom 20 runs significantly, altering their glucose addition profile as the glucose consumption is reduced in a low flux state and enabled a metabolic shift to lactate consumption in some runs and reduced lactate production in others.

5.3.4. Effect of reactor scale

Cell culture process performance is potentially sensitive to changes in gas-liquid mass transfer characteristics of the reactor. The interfacial mass transfer capacity is affected by process parameters including aeration rate, agitation rate, equipment design and presence of surfactant, as well as by process scale. Changes in interfacial mass transfer characteristics influence CO_2 exchange rate, in turn changing pH and the addition of pH control agents (CO_2 and sodium carbonate), further affecting cell growth and metabolism. We evaluated the effect of changes in mass transfer capacity by increasing the mass transfer coefficient $k_L a$ by three-fold to simulate a scaling down reactor while maintaining the initial conditions the same (Figure 5.11A). With an increased $k_L a$ a reduced accumulation of CO_2 during the period of high cell concentration (data not shown). This consequently led to reduced base addition and a lower osmolarity, resulting in the shift of many runs from a high lactate accumulation to a low flux metabolism. It is also shown that a progressive increase of $k_L a$ also reduces final lactate accumulation (Figure 5.11B).

5.4. Discussion

5.4.1. A model that describes process dynamics

The cell bioprocess model described in this work connects reactor operation, including control actions, nutrients, and gas exchanges, with growth and metabolism to evaluate the cascading effects that stem from these interactions. The manufacturing data used for estimating model parameters exhibited significant variability in both final titer and several process variables. Due to the numerical complexity of the metabolic model employed for this work, the parameter fitting process was broken down into separate steps to reduce the number of parameters fitted simultaneously. These initial parameter sets were then used in the combined process model as a good initial guess, reducing the adjustments needed in the most complex stage of parameter fitting.

We note that the aim of the model was not to completely mimic each run as several factors hinder the reproduction of the manufacturing profile. In the simulation we imposed control actions for pH and glucose levels which are likely different from the actual ones as the controller settings in the manufacturing process were not available. That would likely render the fine details of chemical dynamics simulated different from the actual run.

The parameter-optimized model qualitatively reproduces the successive metabolic shifts and variability of metabolic behavior seen in the manufacturing data. Importantly, even though the parameter optimization was performed using only 20 sets of top and bottom runs, the model was able to reproduce the behavior in the intermediate runs.

Goodness of fit for the model was assessed using the lactate production during the final stage as a benchmark (Figure 5.9), given that all model simulations mirror the initial phase of lactate production followed by a low-flux state. As expected, the model fit to the highest and lowest titer runs was better, upon which the parameter estimation was initially performed. However, even the middle runs were predicted with reasonable accuracy, giving confidence in the extrapolation capacity of the model.

5.4.2. *In silico* evaluation of operating conditions

The behavior of the seed train was previously found to be correlated to the final performance of the manufacturing run. Consistent with that finding, the model identified that the initial cultural conditions had a major effect on the variability. Hence, a natural question to ask is whether intervention of initial cultural conditions can steer all cultures to switch to the desired metabolic behavior in the late stage. Two variables, lactate concentration and osmolality, were chosen to be varied at the start of the process as they were shown previously to be correlated to the productivity [102]. The two variables are also manipulatable by employing a perfusion culture in the N-1 seed reactor or, in the case of osmolality, by reducing the osmolality of the initial medium. Interestingly

reducing the initial lactate level has little effect on the performance of the culture (Figure 5.10B), serving mainly to drive additional lactate production for the low titer runs by relieving initial lactate inhibition. In contrast reducing the initial osmolality changed the lactate profile significantly and reduced lactate accumulation (Figure 5.10C).

Because of the high value of the reactor content and supply chain issues, experimentation on the manufacturing scale is rarely practiced. With a mechanistic hybrid model that is capable of simulating culture dynamics, this experimentation can be conducted *in silico*. Then, cell culture experiments can be executed in a scaled-down model bioreactor.

5.4.3. Model guided scale down experimentation

A rational scaling down must not merely reduce the volume of reactor and to multiplex experimental conditions. It must also consider physical and chemical variables affected by the scale, not all of which can be maintained constant during scale translation, and predict and mitigate their effects on the process dynamics. Using the parameter-optimized model we illustrated that enhancing CO_2 stripping by increasing interfacial mass transfer capacity can lead to better and more robust metabolic behavior and process performance. Since scaling down from a manufacturing reactor to a laboratory one is often accompanied by an increase in oxygen and carbon dioxide transfer capacity across the gas-liquid interface, the simulations also suggested that without adjusting process variables properly a small bioreactor may not reproduce the conditions leading to process variability. In other words, if one is to investigate the effect of reducing the osmolality in a manufacturing process using a laboratory bioreactor, the results will likely be confounded by simultaneous changes in process dynamics caused by different mass transfer. The model, by interconnecting metabolism, cell growth to reactor parameters that are sensitive to scale allows one to evaluate and mitigate the effect of scale translation on process performance.

In the model presented in this study, the most scale sensitive parameters are CO_2 and osmolarity. The model considers only the interplay between metabolism and cell growth, but not to product formation and product quality. With better understanding of cellular physiology, one may extend the model to consider the stress caused by hydrodynamics, or the effect of metabolic behavior on product glycosylation.

5.5. Concluding remarks

In this work we present an integrated cell culture process model that reproduces complex process dynamics and their variability. The model can be applied to explore the effect of altering process conditions and changes in scale. The incorporation of mechanistic metabolic model components into an integrated process model allows for reproduction of the dynamics of the process. However, the model is nonlinear and stiff, posing computational challenges for simulation and optimization. The stiffness hinders discretization of the differential equations in the model and prompts the use of black-box optimization methods for parameter estimation. Model reduction to remove fast times scales would reduce the computational complexity of evaluating the model [193-195] and enable the use of faster optimization solvers. The combined use of model reduction and advanced optimization techniques will advance the application of this approach in process research and enhance process robustness and shorten product development timelines. Nevertheless, here we were able to fix a complex set of manufacturing scale data to the model and use it to explore the origins of process variability as well as potential mitigation strategies, with great potential to accelerate process development, reactor scaling, and aid in manufacturing support activities.

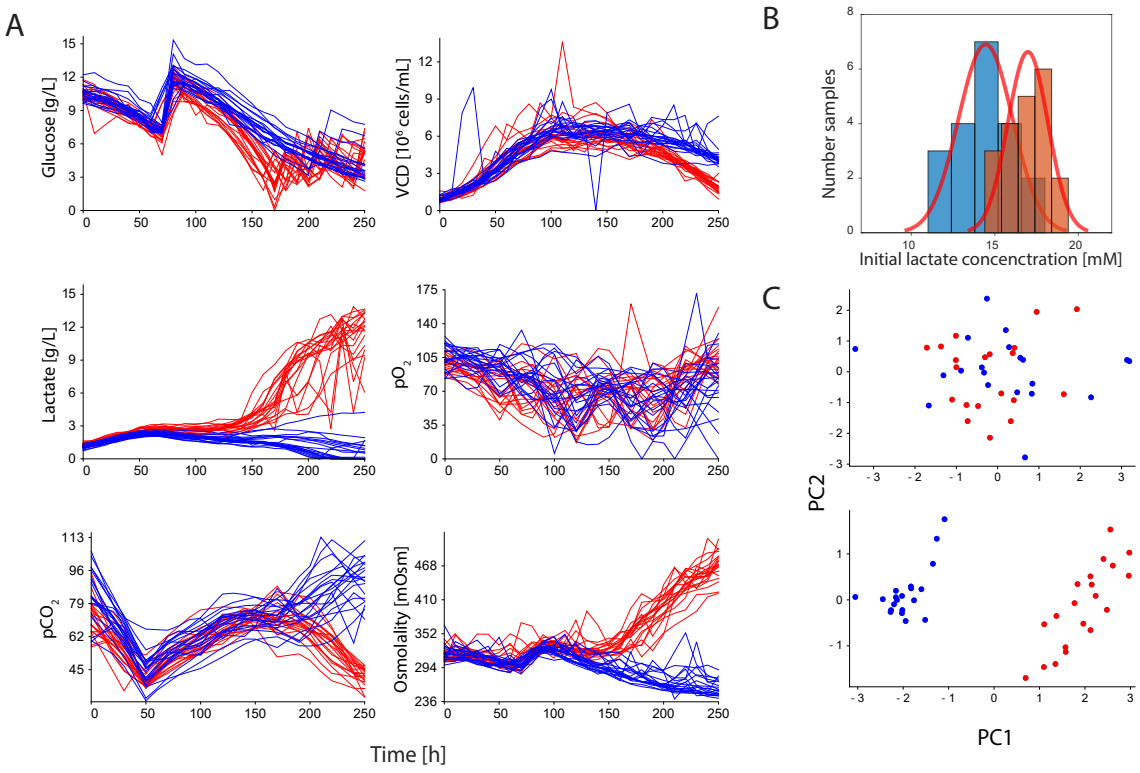


Figure 5.1: Overview of key process data characteristics.

Shown are the top 20 (blue) and bottom 20 (red) runs as ranked by final titer. (A) Key offline measurements at the 12,000L scale. (B) Initial lactate concentration for the top and bottom runs. (C) PCA applied to the initial (top panel) and final (bottom panel) conditions.

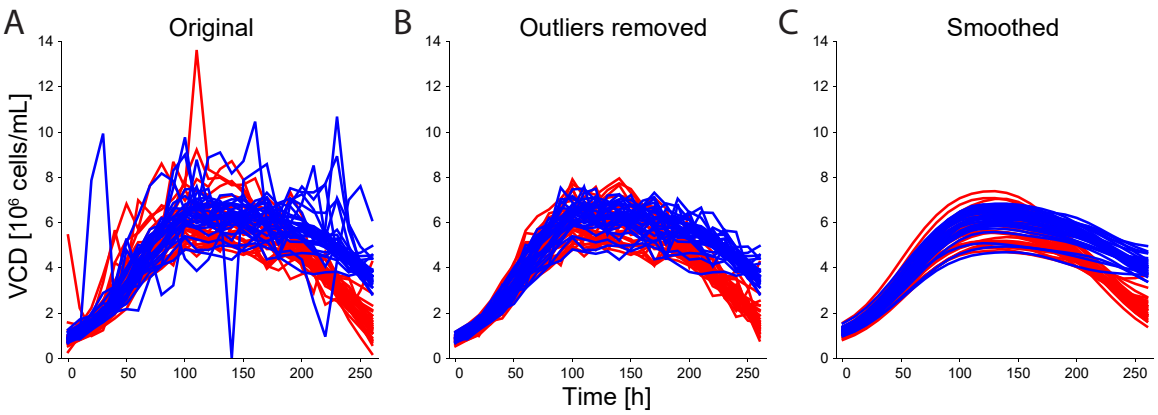


Figure 5.2: Schematic for hybrid mechanistic-empirical model used in this work.

(A) Overview of all model components: reactor liquid phase, reactor gas phase, biotic phase. (B) Cellular metabolic model. (C) Cell signaling regulations. (D) Depiction of the CO2 interactions in the cell, liquid, and has phases. (E) Overview of model size and number of model parameters.

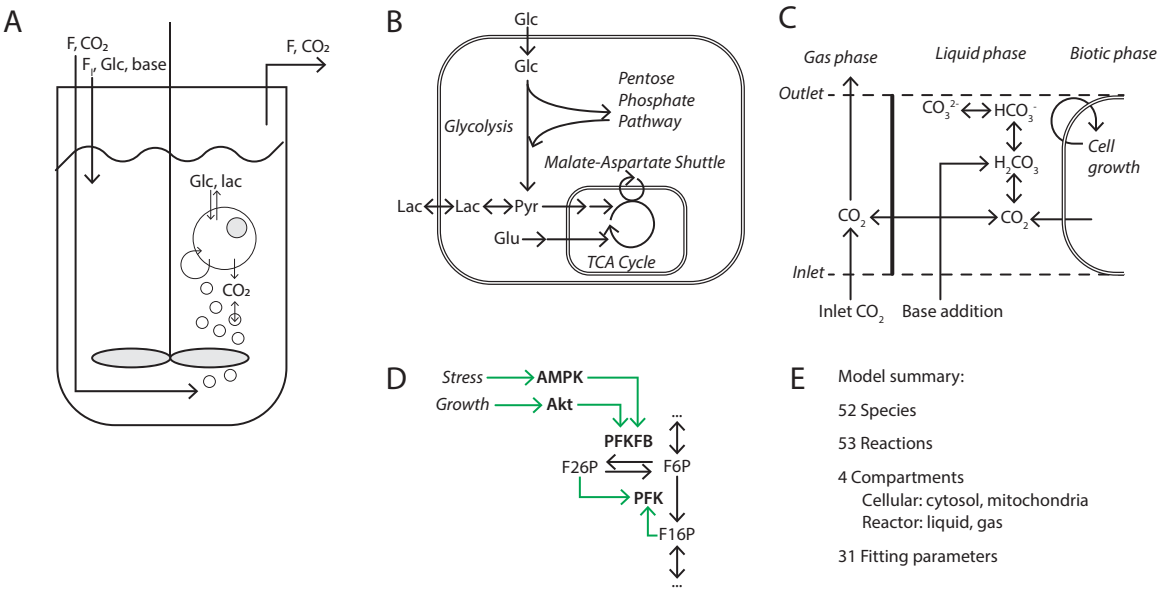


Figure 5.3: Data smoothing for model optimization.

(A) Original curves. (B) Anomaly detection to remove statistical anomalies from curves and data reconstruction to add missing data points. (D) Gaussian blur to smooth data for fitting and calculation of derived values such as specific growth rate.

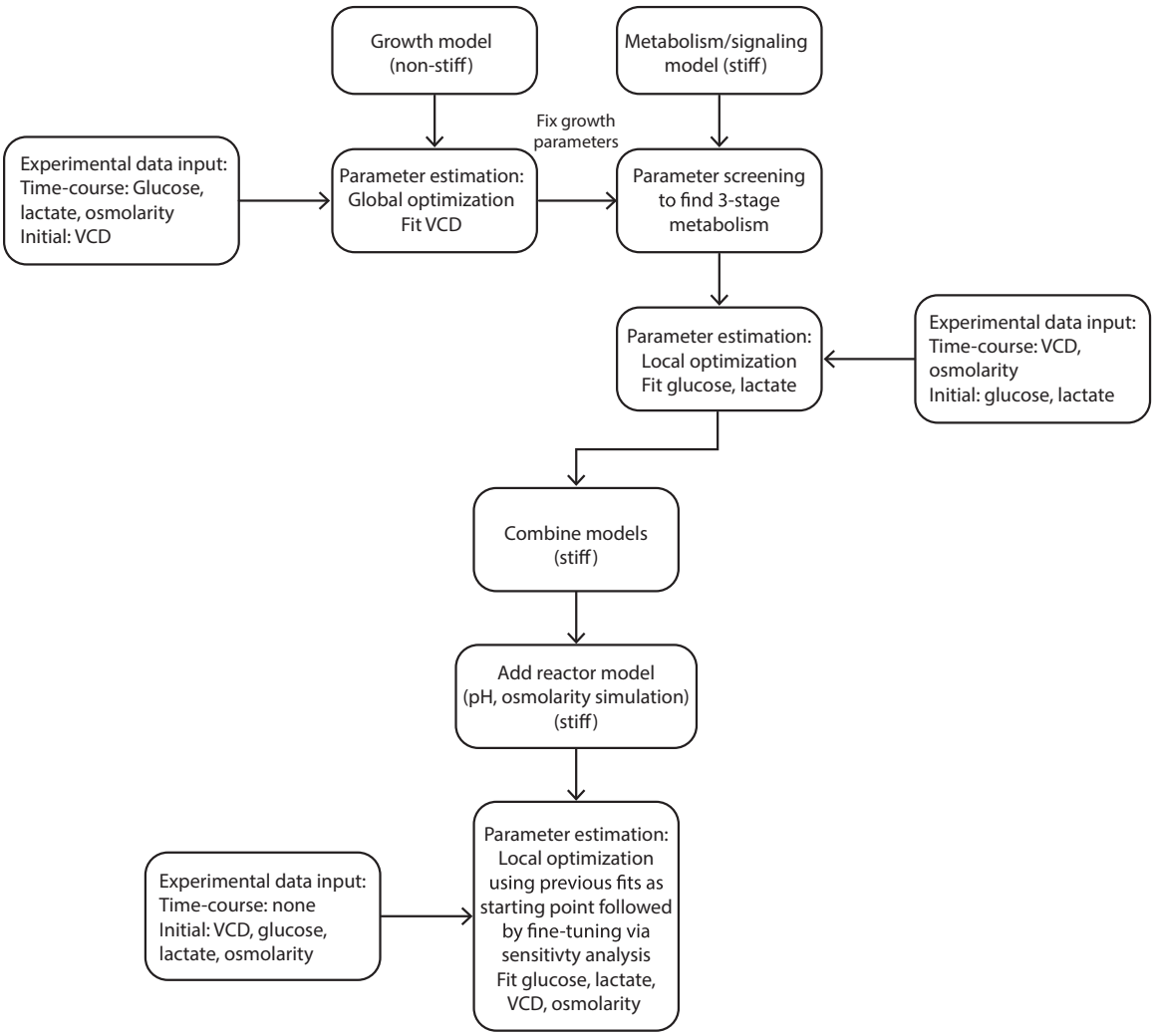


Figure 5.4: Flowchart for model fitting.

Complexity of parameter estimation was reduced by fitting parts of the model separately and using those fits as an initial guess for the combined model.

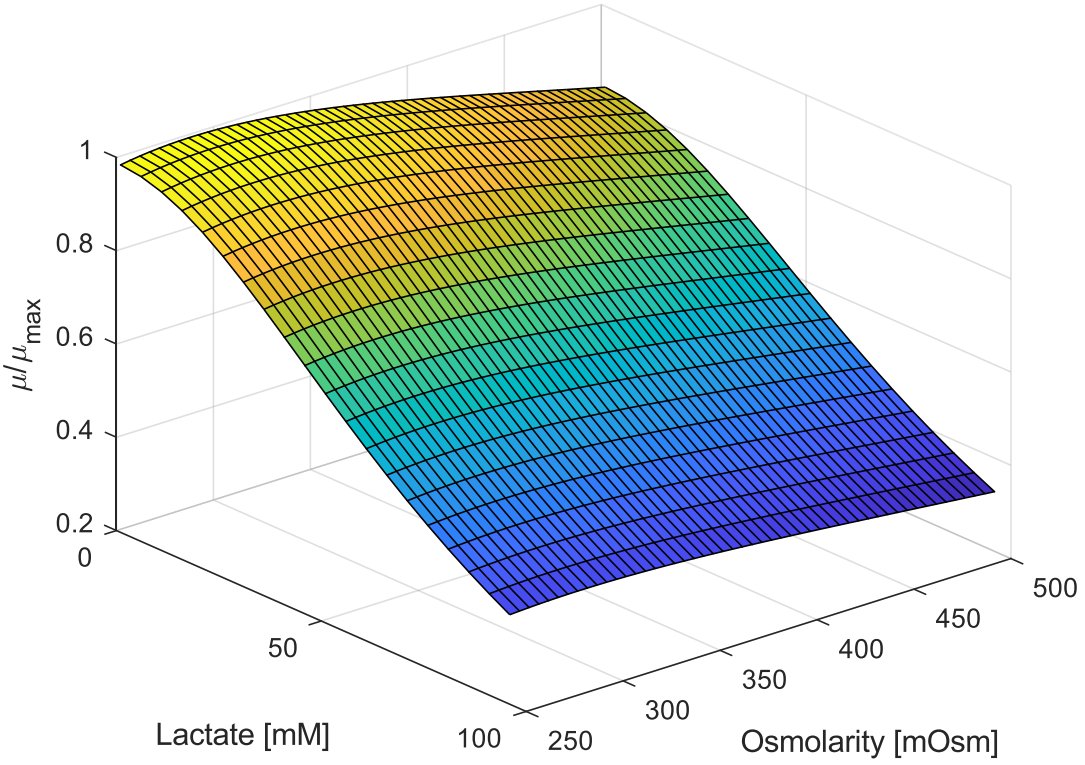


Figure 5.5: Relative growth rate as a function of lactate and osmolarity.

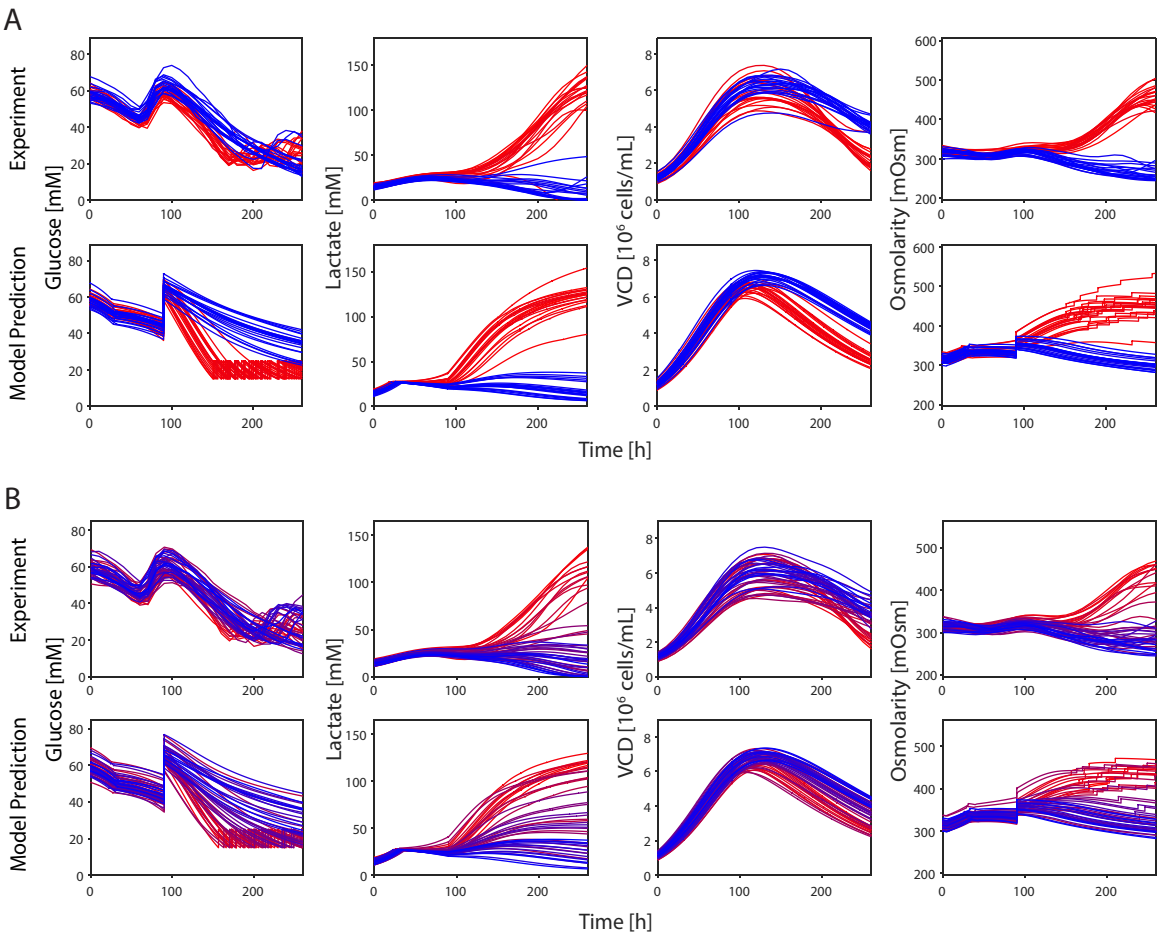


Figure 5.6: Combined metabolism, signaling, cell growth, and reactor model parameter fit.

(A) Comparison of the experimental data of the top 20 (blue) and bottom 20 (red) runs (top panels) to the model fit (bottom panels). The model exhibits the 3-stage behavior as seen in the experimental data: an initial period of high flux metabolism and lactate production, followed by an intermediate period of flow flux metabolism, and finally with some runs shifting to high flux in the late stage of the culture. (B) Comparison of a selection of the 243 runs (1 out of every 5 runs as ranked by titer, top panel) to model simulation (bottom panel). Line color is shown representing final titer, with high titer runs as blue and low titer runs as red.

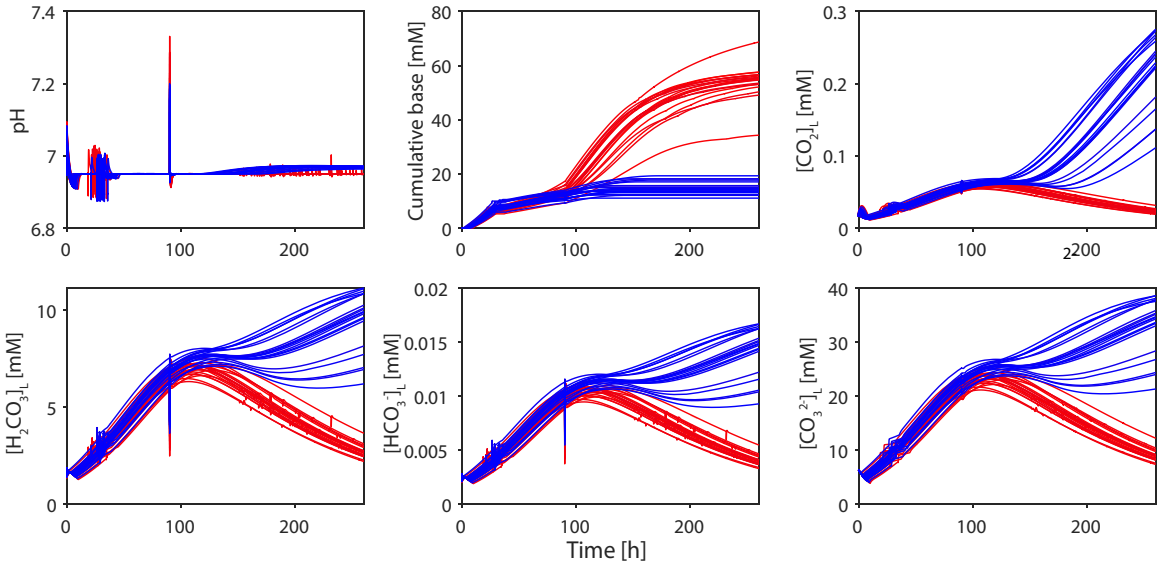


Figure 5.7: pH and carbon dioxide related parameters from the combined model fit corresponding to the experimental data shown in Figure 3.

Shown are the simulations corresponding to the top 20 (blue) and bottom 20 (red) runs.

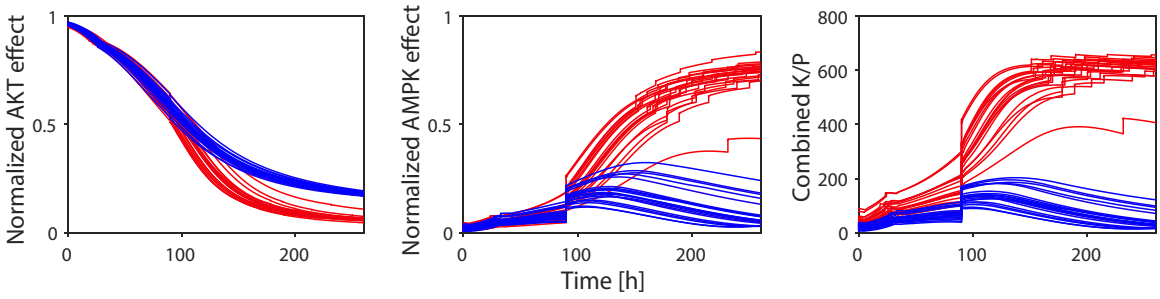


Figure 5.8: Cell signaling effect guides metabolic fates.

Akt, AMPK, and the combined effect on K/P (including the initial K/P which was set according to titer), as an approximation of cell stress in the seed train for the top 20 and bottom 20 runs.

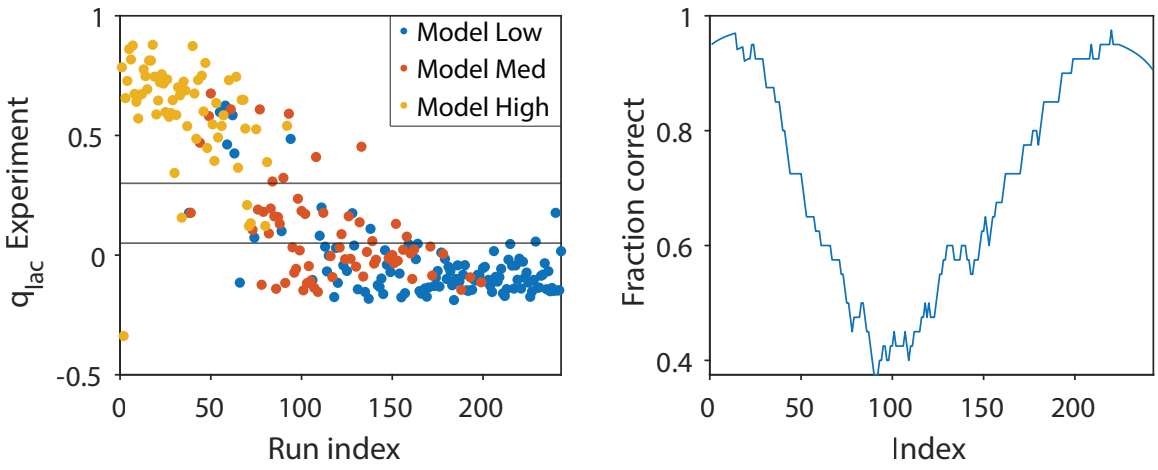


Figure 5.9: Model goodness of fit.

Average lactate flux for the final 120 hours of each culture is shown as a function of run index, as ranked by final titer, with horizontal lines denoting a high, medium, and low flux from top to bottom, and color representing the class assigned to the model points (left). Also shown is the moving window average of fraction of runs predicted in the correct class as a function of final titer (right).

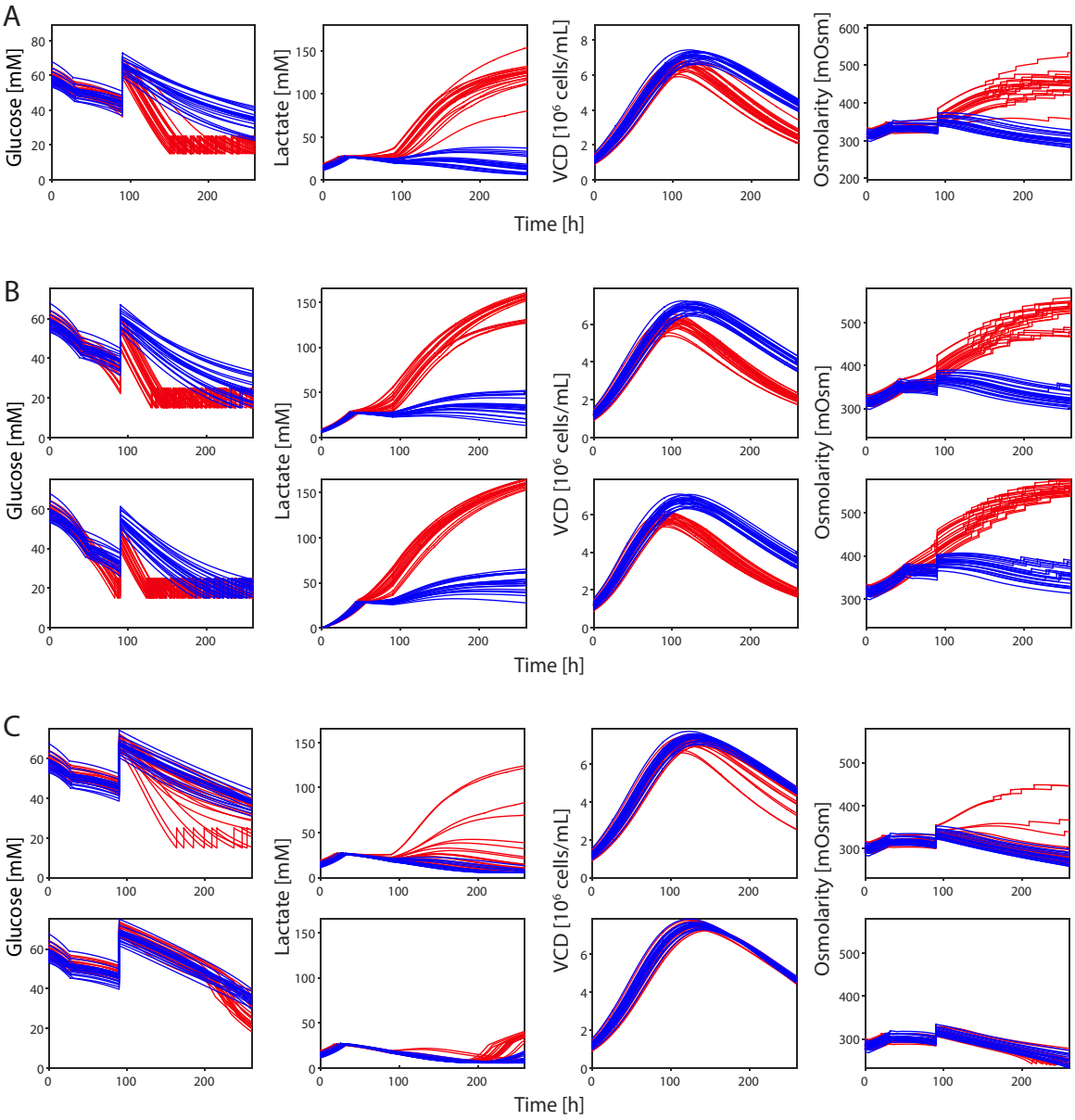


Figure 5.10: Simulation of altering reactor operation conditions.

(A) Baseline model parameter for reference. (B) Reduction of input lactate concentration. Lactate concentration reduced to 50% (top) and 0% (bottom) of original values. (C) Reduction of initial reactor osmolarity to 95% (top) and 90% (bottom) of original model values.

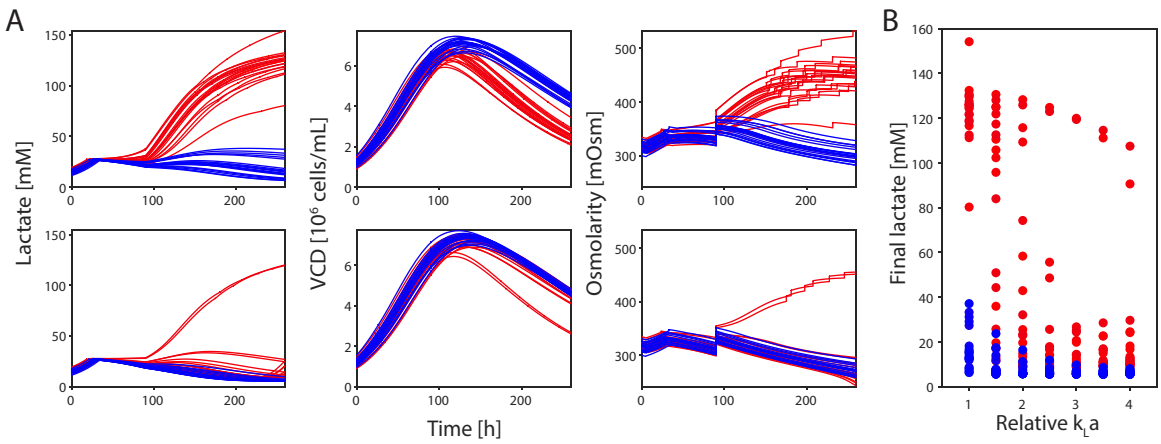


Figure 5.11: Simulation of the effect of altered reactor k_La on process performance.

(A) Original large-scale k_La (top panels) and a three-fold increased k_La (bottom panels). (B) Final lactate concentration as a function of k_La relative to large scale.

6. Model-driven engineering of N-linked glycosylation in Chinese Hamster Ovary cells

Reprinted with permission from Stach, C. S.*, McCann, M. G.*, O'Brien, C. M.*, Le, T. S., Somia, N., Chen, X., ... & Hu, W. S. (2019). Model-driven engineering of N-linked glycosylation in Chinese Hamster Ovary cells. *ACS synthetic biology*, *8*(11), 2524-2535.

* These authors contributed equally to this work.

6.1. Introduction

Mammalian cell lines have applications as biological factories for protein-based therapeutics, gene therapies, and vaccines. Towards this end, cell lines have been developed that are capable of industrial-scale production of human therapeutics and other valuable biomolecules [196]. Cell lines used for industrial production today are the result of iterative, empirically-guided improvement efforts that include random transgene integration, amplification, and screening [197]. Several aspects of cell physiology, including growth, metabolism, and protein-processing work together to contribute to product quality and overall production titers [198]. However, many naturally-evolved traits are not desirable for applications in industrial biomanufacturing. Developing next-generation mammalian cell factories will require reshaping cell physiology to improve product quality, reproducibility, and economy of production. We focus here on controlling protein glycosylation pathways.

Asparagine-(N-) linked glycosylation is an ubiquitous post-translational modification on therapeutic proteins. During protein synthesis, a 14-sugar oligosaccharide containing two N-acetylglucosamines (GlcNAc), nine mannoses, and three glucoses is transferred *en-bloc* to an asparagine residue [199]. This oligosaccharide is further modified by the addition and removal of

diverse sugar monomers in the endoplasmic reticulum and Golgi apparatus to produce the final glycan structure [200]. The final population of secreted proteins contains a mixture of N-linked glycans, and the structural composition of these glycans varies from one cell line to another. Importantly, the chemical structure of N-linked glycans affects physical and chemical stability, pharmacokinetic/pharmacodynamic properties, and ultimately biological activity of a therapeutic protein [201]. For example, galactose content influences complement-dependent cytotoxicity and affinity for C1q in the complement cascade [202-205]. In addition, high galactose content is required for sialic acid incorporation, which mediates the pro- or anti-inflammatory responses of therapeutic proteins containing the IgG Fc region [206]. Improved ability to control N-linked glycosylation will provide the opportunity to tune clinically-relevant properties of therapeutic proteins.

Systems-level approaches to understand protein processing have been used to demonstrate the dynamic nature of the glycosylation network via simulation models [207] and parameter fitting to experimental data [208], and to predict the outcomes of engineering efforts [209]. A challenge with such models is a relative limitation in data compared to the complexity of the systems, leading to potential overfitting [208]. Perturbation analyses, particularly those that vary multiple parameters at once, are rarely employed to validate systems-level predictions.

Previous efforts towards glycoengineering of therapeutic proteins have spanned from rational genetic manipulations to indirect methods such as medium supplementation and random mutagenesis [210]. Fucose incorporation was the target of several early glycoengineering efforts. Blocking fucose incorporation to N-glycans on human IgGs enhanced their binding to the FcγRIIIa receptor of natural killer cells and increased their antibody-dependent cell-mediated cytotoxicity (ADCC) activities. Cell lines were identified that naturally produced afucosylated product, but these were not suitable for industrial production [211]. Chemical mutagenesis and selection yielded

CHO cell lines lacking the ability to produce and incorporate GDP-fucose [212, 213]. Later, genes involved in GDP-fucose synthesis were rationally disrupted using sequential homologous recombination [214] and sequence-programmable nucleases [215]. Transcriptional regulation for glycoengineering was demonstrated using siRNA to knock down the levels of fucosyltransferase, Fut8 [216]. Lastly, heterologous expression of single transgenes has been shown to change N-linked glycan profiles. There are few examples wherein multiple glycosylation genes are overexpressed in concert which have been shown to improve engineering efforts [217]. The above examples are not comprehensive, but illustrate the diversity of approaches that have been applied to control glycan structure in CHO cells.

Here, we combine approaches from systems and synthetic biology to guide N-linked protein glycosylation in CHO cells. First, we characterized the glycan profile of IgG secreted by CHO-2C10, a CHO K1 derived strain that produces 18 pg/cell/day of the recombinant IgG [218]. We then used an iterative approach of systems-level modeling followed by genetic construct design, fabrication, and testing (Figure 6.1A) to increase the galactose content on the secreted IgG. This approach led to improvements in the systems-level model and allowed us to rationally increase galactose content by five-fold. Our pipeline for engineering multi-gene systems in mammalian cells is applicable to systems beyond N-linked glycosylation. For example, this iterative multi-gene engineering approach could be used to modify other metabolic or physiological properties of CHO cells to develop next generation industrial cell lines.

6.2. Materials and Methods.

6.2.1. Media and reagents

Escherichia coli NEB Stable cells (New England Biolabs #C3040H, Ipswitch, MA) were used for routine cloning and plasmid propagation. Difco LB-Miller (BD #244620) media was used for *E. coli* strain growth and maintenance. Antibiotic selection was done using Kanamycin (50mg/L,

Sigma-Aldrich, #K1377), Carbenicillin (100mg/L, IBIScientific, #IB02025), and Blasticidin (12mg/L, InvivoGen, # ant-bl-05). Primers and gene fragments (G-blocks) were ordered from Integrated DNA Technologies (Coralville, IA), while sequence verified genes were ordered from Twist Bioscience (San Francisco, CA). CHO-K1 cells were obtained from ATCC (CHO-K1, ATCC CCL-61) and cultured in F12K Medium (Kaighn's Modification, Gibco, Waltham, MA) supplemented with 10% Fetal Bovine Serum (Gibco, Waltham, MA) and incubated at 37 °C in 5% CO2.

6.2.2. Construction of expression vectors

Here we briefly describe construction of key plasmids. Each plasmid was transformed into chemically competent *E. coli* NEB Stable cells, selected on solid medium with appropriate antibiotics, isolated by plasmid mini-prep (Qiagen #27106, Valencia, CA), and confirmed by Sanger sequencing.

pCDS vector. pCDS was constructed via with a one-pot Golden Gate assembly as previously described [219] using 5 U *Bbs*I (New England Biolabs, R0539S) and 5 U T4 Ligase (Promega, # M1804) from two PCR fragments. The origin of replication and selectable Kanamycin marker was amplified using oligonucleotide primers (pCDS_vecF, pCDS_vecR) from plasmid pMJS1AE and lacZα gene product was amplified from plasmid pMJS1AE using oligonucleotide primers (pCDS_lacZF, pCDS_lacZR) that contain lacZ-specific sequences, a *Sap*I recognition site, 4-bp assembly scar sequence, and an *Aar*I recognition site.

pSG-DV vector. pSG-DV was constructed via an isothermal assembly reaction [220] from multiple PCR fragments (Integrated DNA Technologies, Coralville, IA). The lacZα gene product flanked on each side with *Bbs*I followed by *Aar*I restriction sites with 4- base pair "AB" assembly scars was PCR amplified from pMG-DV (CSS_ScarAF_LacZ, CSS_LacZ_ScarBR). The promoter driving Blasticidin resistance was amplified from pMG-DV as 2 fragments to remove an *Aar*I site

((CSS_ScarB_UBCpromF, CSS_UBCR_RemAarIR; fragment 1), (CSS_UBCF_RemAarIF, CSS_UBCR_Blast; fragment 2)). A portion of the vector backbone was amplified from pMG-DV as 2 fragments to remove a SapI site ((CSS_SynPolyAF, CSS_SpacerR_RemSap; fragment 1), (CSS_RemSapIF_Spacer, CSS_SpacerR_Beta-lac; fragment 2). The origin of replication was amplified from pMG-DV as 2 fragments to remove a *Sap*I site ((SpacerF, CSS_ColE1_RemSapIR; fragment 1), (CSS_Rem_SapIF_ColE1, CSS_ColE1R_ISce; fragment 2)). The remaining fragment of the vector backbone was amplified from pMG-DV using primers CSS_Vector_BB_ScarAR and CSS_ISceIF. Carbenicillin resistance marker was amplified from pMJS_GFP (CSS_Beta-LacF, CSS_Beta_LacR-Spacer). Blasticidin resistance marker was amplified from a G-block (CSS_G-Block Blasticidin) (CSS_BlastF, CSS_Blast_R_SynPolyA). Plasmids were assembled using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, #E2621S) following manufacturer's protocols. The *Aar*I and *Bbs*I restriction sites flanking the lacZα gene contained 4-base pair assembly scars; A-GGAG; B-TACT; C-TTGG; and D-AGGT. A lacZα gene containing BC (CSS_Vec_ScarBR, CSS_ScarB_LacZF, CSS_LacZ_ScarCR, CSS_ScarC_UBCF), or CD (CSS_Vec_ScarCR, CSS_ScarC_LacZF, CSS_LacZ_ScarDR, CSS_ScarD_UBCF) scars was also amplified and used in place of the lacZα "AB" fragment to generate variants. Three pSG-DV vectors designated as pSG-DVAB, pSG-DVBC, or pSG-DVCD were generated.

pMG-DV vector. pMG-DV (CHO-DV) was assembled using an isothermal assembly reaction from multiple PCR and G-block fragments (Integrated DNA Technologies, Coralville, IA). A fragment containing the origin of replication and selectable Kanamycin marker was amplified from plasmid pMC.CMV-MCS-EF1a-RFP-SV40pA (#1pMC_fwd, #1 pMC_rev). The lacZα gene product flanked on each side with *Bbs*I followed by *Aar*I restriction sites was synthesized (LacZ Fragment G_block) the with "AD" 4-bp scar combination. The Blasticidin selectable marker fragment was synthesized (Blast(R)_G-Block), while the promoter driving Blasticidin expression was amplified from pSF-CMV-Ub-Blast via PCR (#3 UbcP(a)_fwd, #4 Ubcp(b)-gcg_rev). The

fragments were assembled using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, #E2621S) following manufacturer's protocols. A BFP expressing variant of pMG-DV was also generated. The BFP was amplified from pCDSBFP (CSS_BFPswtchBFPF, CSS_BFPswtchBFPR). The vector backbone which includes lacZα flanked on each side with *Bbs*I followed by *Aar*I restriction the with "AD" 4-bp scar combination, Carbenicillin selectable marker, Blasticidin selectable marker, origin of replication and recombination sites was amplified from CHODVRFPSwitch (CSS_BFPswtchCHDVF, CSS_BFPswtchCHDVR).The fragments were assembled using one-pot SapI restriction digestion-ligation reaction from multiple PCR fragments (using 10 U *Sap*I (New England Biolabs, #R0569) and 5 U T4 ligase (Promega, # M1804).

CDS-part plasmids. Human CDSs were redesigned with silent mutations that eliminated recognition sites for *Aar*I, *Bbs*I, and *Sap*I restriction enzymes. CDSs were synthesized by Twist Bioscience (San Francisco, CA) with (5'- atgcaCACCTGCTACTA-) and (-TATGGGCAGGTGatgca-3') appended to the 5' and 3' ends, respectively, to enable modular cloning. CDSs were cloned into the pCDS vector via a one-pot *Aar*I restriction digestion-ligation reaction (GeneArt™ Type IIs Assembly Kit, Aar I ThermoFisher Scientific, # A15916) [219].

pFACS plasmids. pFACS plasmids were assembled using a one-pot *Aar*I restriction digestion-ligation reaction using a promoter, a GFP gene, and a polyadenylation sequence. The GFP gene was amplified from pMJS2-GFP with flanking *Aar*I and *Sap*I restriction sites with AATG and TGAT overhangs. The promoters and polyadenylation sequences were PCR amplified with oligonucleotides appending 4-base pair scar sites corresponding to A, B, or C and B, C, or D scars respectively for assembly into pSG-DVAB, pSG-DVBC, or pSG-DVCD. pCMV promoter was amplified from pMC.CMV-MCS-EF1a-RFP-SV40p (CSS_CMVF_ScarA, CSS_CMVR_AATG). pSV40 promoter was amplified from pCDNA3.1_Hygro (+) (CSS_SV40promF_ScarB, CSS_SV40promR_AATG). pEF1a promoter was PCR amplified from pMC.CMV-MCS-EF1a-

RFP-SV40p (CSS_EF1F_ScarC, CSS_EF1R_AATG). SV40pA polyadenylation signal was PCR amplified from pMC.CMV-MCS-EF1a-RFP-SV40p (CSS_SV40TermAF_tgat, CSS_SV40TermAR_ScarC). bGpA polyadenylation signal was PCR amplified from pCAG-Cre:GFP (CSS_rbglobTermAF_tgat and CSS_rbglobTermAR_ScarB or CSS_rbglobTermAR_ScarD).

Single-gene expression plasmids. Single-gene expression vectors were built by replacing the fluorophore gene in pFACS plasmids with a CDS from a CDS-part plasmid. For each reaction, a one-pot Golden Gate assembly was performed as previously described [219] using 10 U *Sap*I (New England Biolabs, #R0569) and 5 U T4 ligase (Promega, # M1804).

Multi-gene expression plasmids. Multi-gene plasmids were built by combining single-gene plasmids with flanking 'AB', 'BC', and 'CD' four-base overhangs, respectively, into the pMG-DV vector with 'AD' four-base overhangs. Each multi-gene plasmid assembly was performed with a one-pot Golden Gate assembly as previously described[24] using 5 U *Bbs*I (New England Biolabs, R0539S) and 5 U T4 Ligase (Promega, # M1804).

6.2.3. Quantification of cis-regulatory elements

The relative strength of promoter elements used to drive transgene expression was determined using fluorescent reporters in CHO-K1 cells. A promoter characterization vector (pCSSRFP, Figure 6.1B) with an invariant RFP expression cassette was used for copy-number normalization. CHO-K1 cells were transfected with 500ng of each construct using DNA-In CHO transfection reagent from MTI-GlobalStem (MTI-GlobalStem, 73781). Cells were seeded at 2 x 10^5 cells/well in a 24 well plate in 500uL of F12K medium supplemented with 10% FBS. At 24 hours post-seeding, 500 ng of DNA complexed with 7.5 µL DNA-In CHO in 250 µL Opti-MEM I medium (Thermo Fisher Scientific 31985062) was added dropwise to each well. Twenty-four hours post transfection cells were trypsinized and analyzed using flow cytometry. The GFP/RFP signal was

determined by plotting a best fit line through all cells gated as GFP+ and RFP+, and the slope of the line was used as the measure of mean promoter strength.

6.2.4. Generation of stable pools

Plasmids were linearized via *MauB*I (multi-gene plasmids), *Sph*I, or *Nde*I (single-gene plasmids) digestion and desalted with the Zymogen DNA Clean and Concentrator 5 kit (multi-gene plasmds) or the Qiaex II gel extraction kit (single-gene plasmids) prior to transfection into CHO-2C10. Linearized plasmids were introduced to the cells using DNA-In® CHO transfection reagent. One day prior to transfection, cells were seeded in 6-well plates at 1.6 x 10^5 cells/well in 3 mL of F12K medium, supplemented with 10% FBS. At 24h post-seeding, 2.5 µg of DNA complexed with 7.5 µL DNA-In® CHO in 250 µL Opti-MEM I medium was added dropwise to each well following a 15-minute incubation at room temperature. Medium was removed from the cells 24 hours post-transfection and replaced with 3 mL of fresh F12K supplemented with 10% FBS.

At 72 hours post-transfection, cells were transferred to 75 cm^2 tissue culture flasks and plated in F12K medium supplemented with 10% FBS and 12 µg/mL blasticidin. Blasticidin-containing medium was replaced every 3 days until cells became confluent. Next, cells were expanded in F12K medium with 5% FBS and 12 µg/mL blasticidin, at a 1:6 ratio every 3 days until five 175 cm^2 tissue culture flasks per condition were obtained.

6.2.5. Targeted integration of multi-gene cassettes

Site-specific integration of a multi-gene cassette into the genome of CHO-2C10 was achieved via dual recombinase mediated cassette exchange (RMCE) [218]. The genomic landing pad has the organization: pCMV-Lox511-IgG-miniFRT-IRES-dGFP-LoxP. Multi-gene plasmids containing minimal FRT and LoxP sites were co-transfected at a 1:4 ratio with recombinase-encoding plasmid,

pCRE2aFLP. pCRE2aFLP codes for flippase (Flpe) linked to CRE recombinase via a *Theoseaasigna* virus (T2A) linker [221] and was provided courtesy of Dr. Yuri Voziyanov.

Following transfection BFP-positive cells were sorted using fluorescence activated cell sorting (FACS) and isolated as clonal populations via limiting dilution. Clones were screened for correct integration of the glycosylation cassette in the landing pad via PCR.

6.2.6. Cell growth and IgG production

Cells in T-175 flasks reached confluency at 3 days. At this point, they were washed twice with PBS and cultured overnight in 25 mL serum-free medium (Ex-Cell CHO DHFR-; Sigma-Aldrich, C8862) supplemented with 2.5 mM L-Glutamine and 3.5 mL/L phenol red. The following day, media was replaced with 40 mL fresh serum-free medium. After 3 days, the cell supernatant was collected and 40 mL of fresh media was added to cells. This was then repeated with supernatant harvested at 3, 6, and 9 days post wash. Collected supernatant was filtered through a 0.22 μm filter and stored at -20°C.

6.2.7. Reverse-transcriptase – polymerase chain reaction

When seeding T-175 flasks, additional cells were collected for RNA analysis. RNA was extracted using the Qiagen RNeasy kit with DNAse treatment performed in-solution (10X Turbo™ DNase buffer; Invitrogen, AM2238). Complimentary DNA (cDNA) was then synthesized using the Invitrogen SuperScript™ III First-Strand Synthesis Supermix with random hexamer priming (Invitrogen, 11752050). With cDNA, 35 cycles of PCR were performed, and the resulting product visualized with a 1% agarose DNA gel containing SYBR Safe DNA Gel Stain (Invitrogen S33102) and UV light.

6.2.8. Determination of IgG concentration using Enzyme-linked Immunosorbant Assay

IgG concentration was determined using an ELISA assay in 96-well plate format as described previously [218].

6.2.9. Purification of IgG and cleavage of N-Glycans

Filtered supernatants collected from days 3 and 6 were immobilized Protein A column (GE, rProtein A Sepharose Fast Flow antibody purification resin 17127901) and eluted with 100mM Glycine, pH 3.0. Eluate was then pH adjusted to 8.0 with 1 M Tris-HCl (pH 9.0) before buffer exchange into 20 mM sodium phosphate buffer, pH 7.4. Next, 200 μg IgG was digested with PNGase-F (NEB P0705). Cleaved N-Glycans were purified by an ethanol precipitation of protein and subsequently lyophilized prior to analysis.

6.2.10. N-Glycan analysis via liquid chromatography (LC)

Lyophilized N-glycans were reconstituted in 2-amino benzamide labeling solution (Prozyme) and incubated at 65 ºC for 3 hours. After labeling, samples were analyzed on a Waters UPLC (H-class Bio System) equipped with a BEH-amide UPLC column (Waters, Milford, USA) and a fluorescence detector (Ex: 330 nm, Em: 420 nm). Mobile phase A was 100 mM ammonium formate (pH 4.5) and mobile phase B was acetonitrile. N-glycan separation was performed using a linear gradient from 75% B to 54% B in 40 minutes at 0.4 mL/min.

6.2.11. Statistical analysis

Comparison of multi-class glycan profiles was performed using a modified χ^2-test. Specifically, a custom conversion factor for comparing χ^2-values to p-values was needed to account for the fact that glycan profile data was recorded as a percentage and not a count value. Briefly, nine independent replicates of the starting cell line, CHO-2C10 were compared in all pairwise combinations using the same test and χ^2-values were plotted to fit a standard Probability

Distribution Function for a 2-degrees of freedom analysis. This fit produced a scaling factor that was used to determine p-values for all χ^2–tests.

6.2.12. Model refinement

The kinetic model is based on an assembly of four stirred tank reactors connected in series, with seven types of enzymes and four Golgi nucleotide sugar concentrations. The enzyme kinetics and nucleotide sugar concentrations starting parameter values for this model have been described previously[12]. Each compartment of the Golgi was considered to have a different enzyme composition, but for simplicity this distribution was held fixed and only the total enzyme amount was adjusted during parameter estimation. The transport of the nucleotide sugars was neglected, and were instead represented by a static pool in the Golgi compartment. Transport or other precursor supply limitation is thus modeled through a decreased Golgi concentration of nucleotide sugars. Additionally, retrograde transport was not considered as a part of this model, proteins flow only downstream. For parameter estimation, the relative concentration of these 11 components were used as the unknown parameters. The starting values for these parameters were allowed to vary between 0.01 and 100 times their original values.

Nonlinear regression was performed to fit the model to the baseline glycan profile of CHO-2C10 using the output of the final reactor in the model as the point of comparison. The Nelder-Mead simplex method was used to obtain the initial fit [222]. Due to the non-convexity of the model and the local nature of the optimization method, the optimization was performed multiple times from different starting points to obtain different fit sets of parameters. To reduce computation time of the optimization, initial points were selected as the lowest sum of squared residuals (SSR) starting points from a Latin hypercube sampling as implemented in MATLAB of 20,000 parameters sets where each of the 11 parameters varied between 0.01 and 100. Single gene results were used to further refine the model. This refinement was achieved by simulating single gene overexpression

for each identified parameter set for CHO-2C10. The SSR for each parameter set as a function of single gene overexpression was computed against each experiment. All modeling and parameter estimation were performed in MATLAB.

6.3. Results

6.3.1. DNA assembly pipeline and genetic parts for glycoengineering

We began by adapting a previously described pipeline for building multi-gene expression constructs to make it compatible with mammalian host cells [223]. A hierarchical DNA assembly strategy was used to assemble genetic elements (promoter/5'-UTRs, CDSs, 3'-UTRs) into monocistronic expression constructs and later into multi-gene expression constructs (Figure 6.1A). Cis-regulatory elements used in glycoengineering cassettes were quantitatively compared using a transient fluorescent reporter gene assay. Two biological replicates were analyzed and error bars represent the standard error of the mean (Figure 6.1B). Variant GFP expression was normalized by RFP expression to control for DNA copy number variation. Promoter hEF1a provided the highest expression strength with CMV and SV40 demonstrating the next highest and lowest expression, respectively. Expression was not impacted by changing 3'-UTR.

Coding DNA Sequences (CDSs) for 21 proteins involved in N-linked glycosylation (Figure 6.1C) were refactored to remove unwanted restriction recognition sites and synthesized. In some cases, multiple isoforms of a single gene were included. CDSs in our library include those involved with precursor oligosaccharide production, nucleotide-sugar biosynthesis, nucleotide-sugar transport into the Golgi, and extension of N-linked glycans in the Golgi (Figure 6.1D).

All single-gene and multi-gene plasmids contain vector backbones that allow for site-specific integration via dual recombinase mediated cassette exchange (RMCE) at FRT and LoxP flanked

sites in the genome. Alternatively, constructs could be randomly integrated by lipofection and selection. This pipeline allows for rapid production of expression constructs up to 12 kilobases.

6.3.2. Structural analysis of N-linked glycans on recombinant IgG

CHO-2C10 is a CHO-K1 derived cell line previously engineered to express human IgG from a construct integrated into the host cell genome [218]. The IgG heavy and light chains are expressed from a single CMV promoter with an intervening P2A (porcine teschovirus-1 2A) linker for translational coupling [224].

A baseline glycan profile (Figure 6.1C) was obtained for the CHO-2C10 cell line. IgG produced by the cells was purified by affinity chromatography, digested with PNGase F, and oligosaccharides were labeled with 2-aminobenzamide fluorophores as previously described [225]. Chemical structure of labeled glycans was determined by comparison to reference standards via ultra-high performance liquid chromatography (UPLC). The majority of IgG glycans were lacking galactose ('G0', 57.1% ± 4.6%), followed by monogalactosylated ('G1', 30.1%± 3.1%) and bigalactosylated ('G2', 11.9% ± 2.8%) compounds. Analysis of biological replicates revealed an average between-replicate standard deviation of less than 1% for the majority of the nineteen characterized glycan structures.

6.3.3. Model-driven glycoengineering

We mapped the IgG glycan profile to the glycosylation network using GlycoVis (Figure 6.2A) [226]. GlycoVis maps the network of possible glycan structures as nodes connected by the enzymes that catalyze their interconversion. Many of the glycans that are abundant on IgG from CHO-2C10 cells are immediately upstream of β-1,4-galactosyltransferase (β4GalT) mediated reactions, an enzyme encoded by seven distinct genes in the human genome (Figure 6.2B) [227].

A kinetic model for N-linked glycosylation [207] was used to predict how the glycan profile would be effected by altering glycosyltransferase and mannosidase concentrations as well as

nucleotide sugar concentrations in the Golgi. As a baseline, the model parameters were first fit to match the baseline glycan profile of the CHO-2C10 cell line via the following approach. Local parameter optimization was performed using a Nelder-Mead hill-climbing algorithm [228]. This algorithm identifies combinations of parameters that minimize the sum of squared residuals (SSR) between the experimental glycan profile and the composition of glycans in the final Golgi compartment in the kinetic model. The parameter estimation consisted of eleven variables: two mannosidases (ManI, ManII), five transferases (FucT, GnTI, GnTII, GalT, and SiaT) and four nucleotide sugars (GDP-fucose, UDP-GlcNAc, UDP-galactose, and CMP-sialic acid). As Nelder-Mead is a local optimization method, the optimization was repeated from 50 different starting points to produce 50 locally optimal parameter sets. Next, a local sensitivity analysis was performed on each parameter set. Specifically, we investigated the impact of a ± 10% perturbation of the intra-Golgi concentrations for each glycosyltransferase and nucleotide-sugar on the amount of galactose incorporation (Figure 6.2C). Increasing the concentration of galactosyltransferases (GalTs) and UDP-galactose were expected to have the largest impact on galactose incorporation (Figure 6.2D). The qualitative similarity in behavior and sensitivity of many of these parameter sets with regards to glycosylation (Figure 6.2C, D) implies that the major phenomena and limitations of the starting CHO-2C10 cells have been identified as far as the experimental data available can support. Impact of the number of initial starting points prior to Nelder-Mead optimization was determined by plotting how the mean Galactosylation sensitivity for each parameter would have changed if the number of starting points varied from 1-50. The results indicate that sampling 50 points is likely sufficient to capture the diversity of model behavior. Thus we believe the parameter sets identified are sufficient to perturb and further refine the model in the subsequent sections.

β4GalT1 and β4GalT2, two human galactosyltransferases, were selected as candidates for overexpression because of their previously reported importance to β-1,4 galactosylation [229-231]. In parallel, we selected genes encoding enzymes predicted to increase the concentration of UDP-

galactose in the Golgi, including biosynthetic enzymes (GALE, UGP2a/b, GALK1, GALT) and nucleotide sugar transporters (SLC35A2/D1). Together these represent the top targets based on kinetic modeling (Figure 6.2D), and their overproduction was predicted to increase the galactose content on IgG. Lastly, we selected two genes (MGAT2, DOLK) predicted to increase the concentration of the agalactosylated 'G0' glycans that are substrates for the galactosyltransferase.

6.3.4. Single-gene glycoengineering

We built eleven single-gene glycoengineering constructs by assembling candidate CDSs with characterized cis-regulatory elements. For rapid screening of overexpression constructs, we transfected CHO-2C10 cells and selected for random genomic integration events using the blasticidin resistance marker. Recombinant cells were isolated as non-clonal pools, and their glycan profile analyzed as described above (Figure 6.3A). Expression of glycosylation genes were confirmed with RT-PCR. Cell growth and IgG productivity of cell pools were not affected by the single-gene overexpression cassettes (data not shown).

The only single-gene perturbation that significantly changed galactose incorporation (based on a modified χ^2-test) was the overexpression of β4GalT1. That β4GalT1 would impact galactose incorporation is expected; every parameter combination tested in the kinetic model predicted that changing galactosyltransferase concentration would increase the amount of 'G1' and 'G2' glycans. β4GalT2 overexpression did not significantly change glycan composition, but this enzyme has been reported have a higher K_m than the β4GalT1 isoform [232], and thus it may be less efficient under some circumstances.

None of the cell pools engineered to overexpress genes involved in UDP-galactose synthesis or transport yielded a change in glycan galactose content. There are three explanations for this result: (i) overexpressing the biosynthetic or transport genes did not change either the protein expression level, or (ii) the UDP-galactose concentration in the Golgi, or (ii) that UDP-galactose

concentration is not limiting in CHO-2C10 cells. Regardless of the underlying reason, this suggests that the kinetic model can be improved using the new experimental data and used to better guide glycoengineering efforts. We pursued this with the following experiments.

6.3.5. Improving the kinetic glycosylation model

We sought to improve the kinetic model based on results from the single-gene overexpression experiment. The large number of model parameters relative to the amount of input data from the CHO-2C10 glycan profile likely resulted in overfitting of locally-optimal parameters, which is observed during the sensitivity analysis (**Figure 6.2C**). Notably, certain parameter sets incorrectly identified UDP-galactose concentrations as limiting galactose incorporation into the IgG glycans. The model performance with respect to UDP-galactose concentration was multi-modal (**Figure 6.2D**), suggesting that some parameter combinations are more accurate than others.

We then compared model predictions and experimental results at each of our 50 parameter combinations (Figure 6.3B). For different targets in the single-gene experiment, we calculated the sum of squared residuals between the single gene experimental results and the model predictions at different relative enzyme or nucleotide sugar concentrations. We note that although we confirmed overexpression of the introduced genes, the actual change in enzyme activity in our overexpression experiments is difficult to measure, as it is affected by protein concentrations and kinetics in each Golgi cisternae which gene expression cannot fully predict. By performing this analysis across a range of enzyme levels, we can observe differences in the robustness of parameter sets.

The glycosyltransferases and UDP-galactose biosynthesis/transport genes show stark differences regarding the model's robustness towards variations in input parameters. The model's ability to predict the single-gene overexpression results was virtually identical across all 50 parameter sets for β4GalT1/2 and MGAT2. For the seven UDP-galactose biosynthesis/transport

genes, only a small subset of parameter combinations (represented as blue traces in Figure 6.3B) generated predictions that agreed with overexpression data. We performed a principle components analysis on log-transformed values of the 50 parameter sets, and the retained parameter sets (blue traces and points in Figure 6.3B,C) separated from the removed sets (red traces and points) along a principle component that is dominated by CMP-sialic acid and UDP-galactose levels. In the retained sets, UDP-galactose levels are on average 225-fold greater than in the removed sets, and CMP-sialic acid levels are approximately 2-fold. This stark difference in UDP-galactose concentration between the two classes of parameter sets stems from the difference in K_m for galactosyltransferase and the Golgi concentration of UDP-galactose. For the model to be sensitive to UDP-galactose supply, the concentration must first be substantially decreased, so that the reaction kinetics are no longer in the saturation region. However, the single gene perturbation experiments imply that in the starting CHO-2C10 cells no such limitation exists.

The results in Figure 6.3B, C were used to down-select a total of thirty parameter sets for future modeling. The retained sets still include diverse values (Figure 6.3C), but these sets better predicted the experimental results, both for the starting CHO-2C10 cells and the single gene overexpression mutants.

These parameter subsets were used to simulate the effects of gene overexpression on galactose incorporation. This time, we allowed for multiple genes to be perturbed at the same time (Figure 6.4A). The narrower set of parameter combinations greatly reduces the variability among the predictions. The model now predicts that UDP-galactose levels are not limiting in CHO-2C10, but become limiting only upon GalT overexpression exerts some influence on galactose incorporation. This is apparent by the slight curve in the response surface (Figure 6.4A) along the UDP-galactose axis only at higher levels of GalT expression. This non-independence of gene expression on glycan structure is common in complex multi-gene systems and has been observed in other multi-gene

systems [223, 233, 234]. Importantly, the model suggests that multi-gene overexpression is important to produce greater changes in the galactose content of IgG glycans.

6.3.6. Glycoengineering through expression of multi-gene cassettes

Adjusting expression levels of individual metabolic enzymes can shift the pathway bottleneck (*i.e.* rate-limiting component) between intermediate steps without altering final glycosylation profiles. While the single-gene overexpression experiment suggested that low β4GalT1 levels in the original cell line limited the degree of galactose incorporation to IgG, our refined model suggested that UDP-galactose concentration may have increased effect after overexpression of galactosyltransferase (Figure 6.4A).

To address this, we designed, built, and tested multi-gene constructs that simultaneously overexpress β4GalT with genes that target UDP-galactose biosynthesis or transport (Figure 6.4B). In addition, several other multi-gene constructs were designed that overexpress genes involved in sialic acid precursor synthesis, sialic acid transfer, and increased glycan branching. The latter group was designed based on previous literature [217, 235, 236] and GlycoVis simulations, but not the kinetic model. Each multi-gene construct is comprised of three separate mono-cistronic expression cassettes. Each monocistron has a unique promoter (Figure. 6.1B) and the SV40 or rbGlob polyA tail. As with the single-gene overexpression experiment, these were originally integrated randomly to the CHO-2C10 genome and pools of transformed cells were isolated by selecting with blasticidin.

Our first design included β4GalT1, UDP-galactose transporter SLC35A2, and galactose kinase GALK1. This design did not lead to more galactose incorporation than β4GalT1 alone. Our next set of designs included the UDP-sugar transporter SLC35D1. SLC35D1 is reported as a UDP-N-acetylgalactosamine (UDP-GalNAc) transporter, although it also transports other UDP-sugars, including UDP-galactose, with lower efficiency [235]. Multi-gene expression constructs that

included SLC35D1 produced the highest rates of galactose incorporation, regardless of which UDP-galactose biosynthetic enzyme was included in the three-gene cassette.

The success rate of constructs engineered to increase galactose incorporation was higher than in the single-gene experiment. Each of the seven cell pools generated showed a significantly increased level of galactose incorporation (Figure. 6.4B) compared to the CHO-2C10 control. The best cells, overexpressing β4GalT1, SLC35D1, and GALK1, have 61.4 ± 6.7 % bi-galactosylated ('G2') glycans, compared to only 11.8 ± 2.7 % 'G2' in CHO-2C10 and 47.3 ± 5.5 % in the β4GalT1-overexpressing cell pools. Importantly, the improvement in galactose incorporation in our best three-gene cell line highlights the shifting bottlenecks that arise when engineering complex systems. Neither SLC35D1 nor GALK1 overexpression impacted galactose incorporation individually, but these were able to synergize the impact of β4GalT1 overexpression.

6.3.7. Reproducible glycoengineering using site-specific integration to a genomic landing pad

The decision to randomly integrate overexpression constructs to the genome was driven by low rates of dual Recombinase-Mediated Cassette Exchange (RMCE) into the target locus within CHO-2C10 [218]. However, we tested whether similar levels of glycan perturbation would be seen following integration of our best three-gene construct site-specifically to the genomic landing pad. The β4GalT1-SLC35D1-GALK1 construct was introduced to CHO-2C10 integration at FRT and loxP sites (Figure 6.4C). Several clonal cell lines were isolated by single-cell cloning and assayed as described above. Each clonal cell line showed high levels of 'G1'/'G2' glycans, ranging from 88% - 90% (Figure 6.4D). IgG titers and cell growth were not affected by the three-gene overexpression cassettes. This experiment highlights the reproducibility of this system.

6.4. Discussion

Glycosylation is a critical post-translation modification that affects the therapeutic properties of antibodies. Steering glycan structure towards a specific target is a long-standing challenge in biopharmaceutical production. In this study we employed a two-pronged approach using synthetic and systems biology to engineer increased galactosylation of IgG. We increased total galactose glycan species ('G1'+'G2') from 42.8% in the IgG-producing CHO-2C10 cell line to 90.4% in a targeted insertion engineered cell line. The fraction of bi-galactosylated glycan increased five-fold. Galactosylatyion is important in modulating the inflammatory potential of IgG and binding of C1q to initiate complement dependent cellular cytotoxicity [202-205]. Galactose incorporation is also a prerequisite for sialylation, which is a primary determining factor in serum retention [201].

In this study we sought to use systems-level models to directly guide construct design and optimization. We started by using empirical glycan composition data from an engineered IgG-producing CHO cell line to fit parameters from a kinetic model. The kinetic model treats the four Golgi compartments (cis, medial, trans, and TGN) as four continuous well-mixed reactors in series [207]. The model considers not only enzyme kinetics and concentrations of substrates in the Golgi, but also the distribution of enzymes between compartments. To simplify the initial parameter optimization, we fixed variables defining flow of substrates between compartments as well as the relative distribution of each enzyme among the different compartments. The only parameters we varied were the V_{max} of seven key glycosyltransferases and the Golgi concentration of four NDP-sugars. Even within these eleven variables, wildly different parameter combinations equally fit the experimental data. This speaks to the complexity of N-linked glycosylation but also to the paucity of fitting data present in the single endpoint analysis of IgG glycan profile.

Our strategy to discriminate between these differing parameter combinations was to perform a perturbation analysis of the model. By overexpressing several enzymes involved in NDP-sugar

biosynthesis, transport, or incorporation into N-linked glycans, we identified a narrower subset of parameter combinations that best predicted the experimental outcomes. This single-gene overexpression experiment (Figure 6.3) showed that increased transcription of none of the enzymes tested required for UDP-galactose biosynthesis or transport into the Golgi were the sole limiting factor in CHO-2C10 cells. From data generated here, we could not distinguish between several possible explanations of these results. However, in light of recent multi-omics systems analysis of IgG glycosylation in CHO cells, it is plausible that metabolites and/or cofactors that increase overall UDP-galactose flux into Golgi apparatus or the transferase activity itself were limiting [237]. It is also possible that expression levels of UDP-galactose biosynthesis enzymes, including GALE, GALK1, and GALT, do not correlate with UDP-galactose concentrations [237].

It was observed that pairing overexpression of β4GalT1 with genes aimed at increasing UDP-Gal supply increased galactosylation. The apparent increased sensitivity of galactosylation to UDP-Gal supply only upon β4GalT1 overexpression, implies that the increased utilization of UDP-Gal may have decreased its concentration in the Golgi below the region of saturation, thus allowing synergistic effects from enzymes which increase its supply. While two genes were paired with β4GalT1 overexpression in this work, it is plausible that overexpression of one of gene with β4GalT1 may have also changed glycosylation performed by the cell.

The high level of galactose incorporation afforded by three-gene constructs that include SLC35D1 may seem counterintuitive due to this transporter's annotation as a UDP-GalNAc transporter involved in O-glycosylation. Several explanations could explain this result. During CHO cell culture, UDP-GalNAc levels are known to rise considerably over the several day cultivation period [237, 238]. Like SLC35D1, SLC35A2 is promiscuous in the UDP-sugars it transports, and has activity with UDP-GalNAc as well as UDP-GlcNAc [239, 240]. These alternative substrates can act as competitive inhibitors of UDP-galactose transport through

SLC35A2. The impact of overexpressing SLC35D1 on galactose incorporation could come from relieving inhibition of the SLC35A2 channel through lowering the cytoplasmic concentration of UDP-GalNAc. Alternatively, the ability for SLC35D1 to function as a low efficiency UDP-galactose transporter could directly increase levels of UDP-galactose in the Golgi apparatus. Regardless of the explanation, SLC35D1 was able to reproducibly increase galactose incorporation to the IgG glycans in several different genetic designs. Compared to β4GalT1 alone from the single-gene experiment, in which 68% of the glycans were 'G1' or 'G2' (Figure 6.3A), adding SLC35D1 to β4GalT1 constructs increased 'G1'/'G2' levels to 78%-90% (Figure 6.4B,D).

While our kinetic model performed reasonably well at predicting the results from three-gene overexpression experiments (Figure 6.4), it is noteworthy that it's predictions are limited by availability of experimental data. Transcription of introduced genes were confirmed with RT-PCR, but level of overexpression and the kinetic effects of it were not. Additionally, the final optimal model parameters (*e.g.* apparent V_{max} values for glycosyltranferases) are likely influenced by the specific biochemistry of IgG glycosylation. One example of this, is that certain enzymes appear to have a kinetic preference for one branch over the other *in vitro*, whereas the model has equal preference, resulting in concentration differences [207]. Nevertheless, the model was able to recapitulate the approximate results of the experiments. Additionally, each optimized parameter set contains a low V_{max} value for the sialyltransferase, SiaT. The model produced this result because of the low frequency of sialic acid incorporation to IgG, which previous studies have suggested is due in part to steric blocking of the site of sialylation by bulky amino acids in the Fc region [241]. As such, the predictive ability of parameter combinations identified in this study is limited to IgG glycosylation, and not N-linked glycosylation in general. However, our approach of coupling kinetic modeling with single- and multi-gene over-expression should work for a variety of cellular processes.

The largest construct that we integrated to the CHO genome comprised 3 monocistronic expression cassettes in 9.1 kb of total sequence. When inserted into the genomic landing pad that included an IgG expression construct, the total amount of rDNA integrated site-specifically to the CHO genome was 13.7 kb. Modern tools for DNA synthesis and assembly allow for the construction of substantially larger constructs [220, 242]. Currently, transfection efficiency of large constructs limits the size/complexity of rDNA that can be routinely integrated to CHO genomes. While plasmid assembly has improved dramatically in the past decade, methods available for delivering large genetic constructs have not seen similar improvements. Effective delivery of 48 kb cosmid DNA to primary cells was described in the early 1990's. Recent state-the-art efforts, including delivery of rDNA constructs comprising ~40 kb to HEK293FT cells [242], are of the same order of magnitude. Delivery methods that allow for routine delivery of 100 kb or more in a single event, which are available for bacterial engineering [243], would dramatically accelerate multi-gene engineering workflows.

6.5. Conclusion

In conclusion, we present a workflow for iterative kinetic modeling and strain design/construction to engineer N-linked glycan structure in CHO cells. We demonstrated a proof of concept by rationally increasing galactose incorporation to human-IgG. We saw that coupling the systems- and synthetic biology workflows led to improvements in each. Our cell lines that incorporate high levels of galactose retain their good per cell productivity levels and represent platform cells for future glycoengineering efforts, for example to increase sialylation.

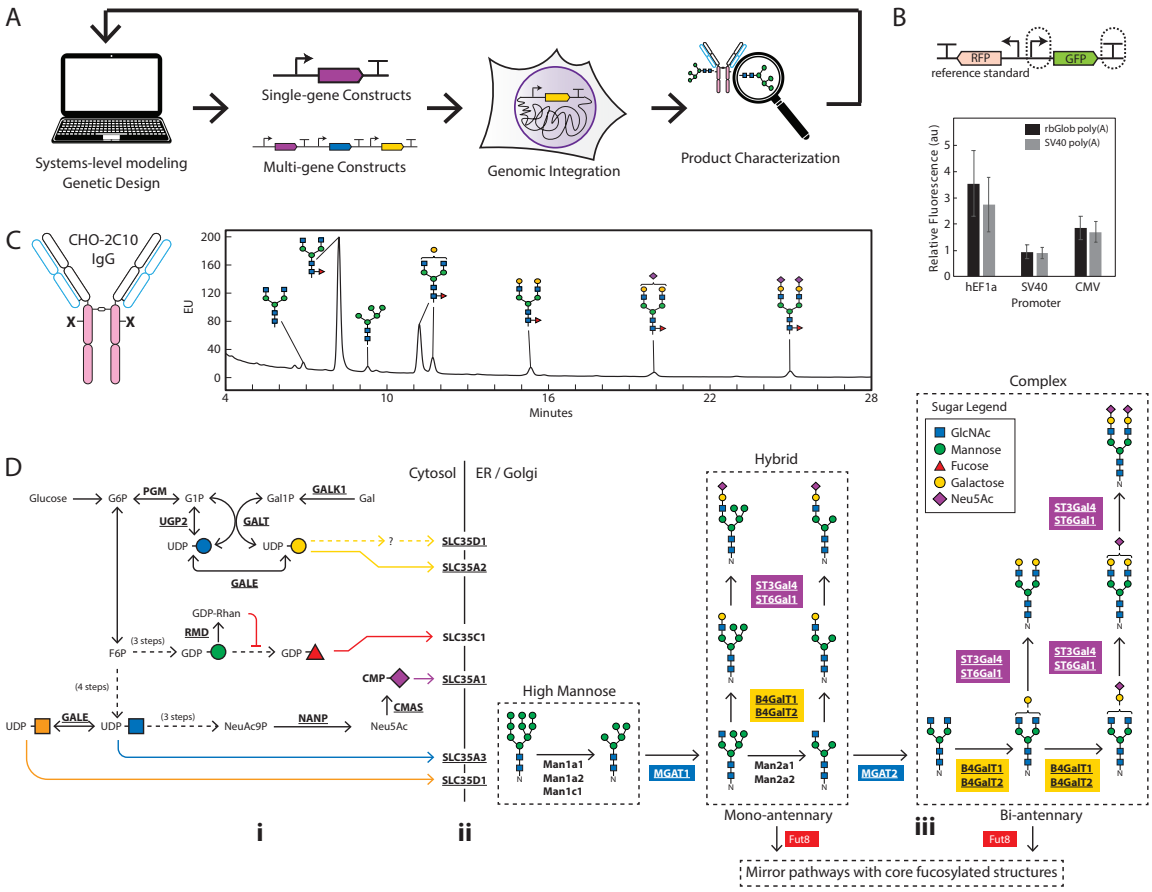


Figure 6.1: Design and assembly of glycoengineering cassettes.

(A) Schematic illustration of model-based iterative glycoengineering. Synthetic Biology Open Language (SBOL) iconography is used to represent genetic constructs here and throughout this manuscript. (B) Genetic design of reporter construct with red fluorescent protein (rfp) internal reference standard (top) and relative expression data (bottom) for mammalian promoters and 3'-UTR elements used in glycoengineering constructs. The variant promoter and terminator positions are marked with dashed boxes. (C) Representative UPLC trace for glycan structures cleaved from IgG isolated from CHO-2C10 cells (sugar legend in D). (D) Systems-level schematic of N-linked protein glycosylation, including (i) NDP-sugar biosynthesis in the cytoplasm, (ii) transport into the ER/Golgi lumen, and (iii) oligosaccharide extension and remodeling. Coding sequences for genes underlined are included in the CDS library.

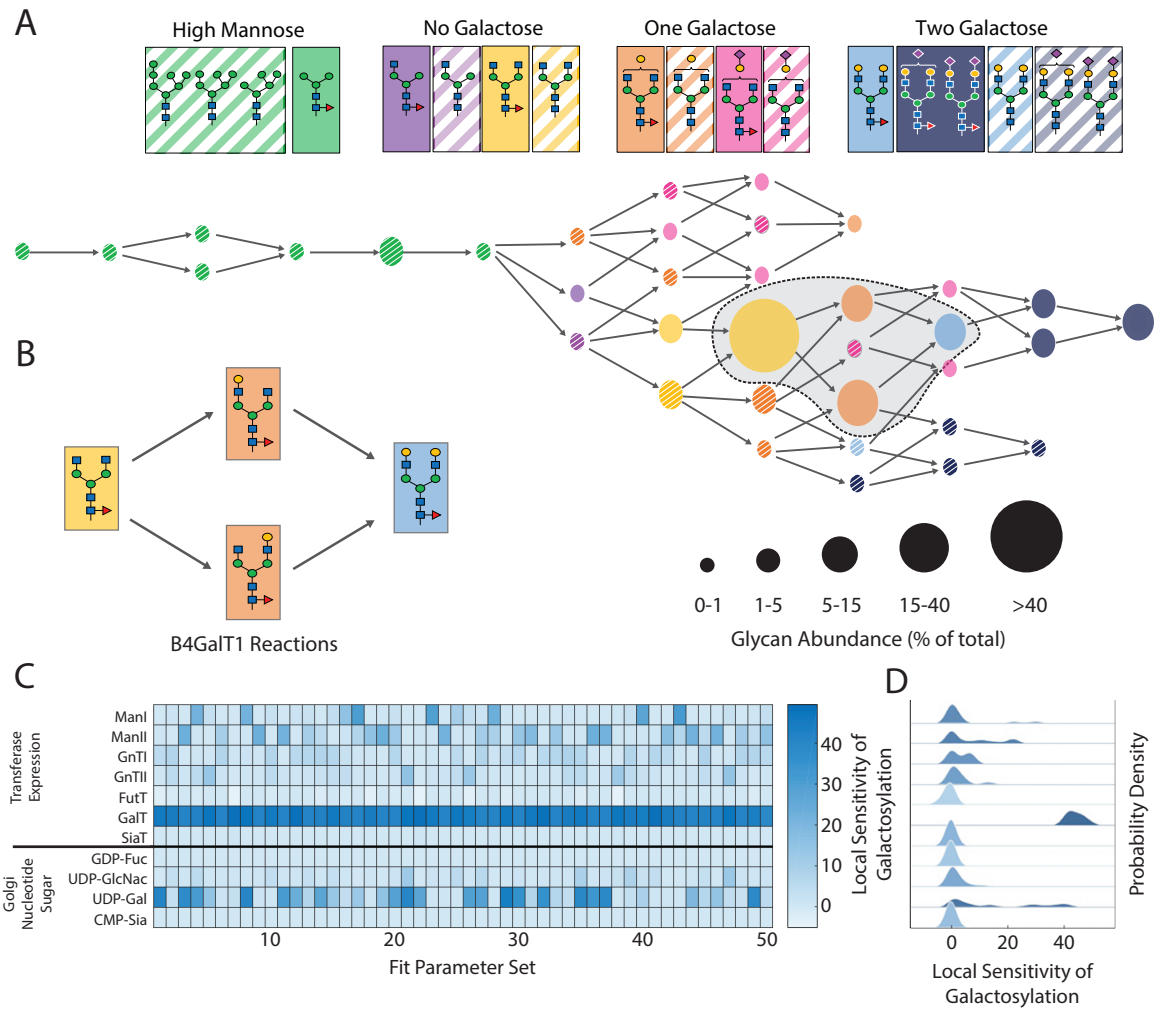


Figure 6.2: Model-driven design of single-gene overexpression constructs.

(A) Baseline glycan profile in CHO-2C10 visualized using GlycoVis network. Nodes in the network diagram are colored according to the legend above, with node size denoting glycan abundance from the baseline UPLC analysis. Arrow (edges) represent known chemical transformations catalyzed by glycotransferases or mannosidases. (B) Structures of predominant glycans, corresponding to the shaded sub-network in (A). Arrows in (B) correspond to B4GalT1-catalyzed reactions. (C) Simulated sensitivity analysis of kinetic N-linked glycosylation model to perturbations in glycotransferase, mannosidase, or nucleotide-sugar concentration. Heatmap shows predicted sensitivity of galactose content in total glycan profile. Galactosylation is defined as percentage of 'G1' gly-

cans plus 2x percentage of ‘G2’ glycans. (D) Probability density plot of local sensitivity analysis from (C), with plot order corresponding to rows in heatmap. Color intensity corresponds to mean sensitivity.

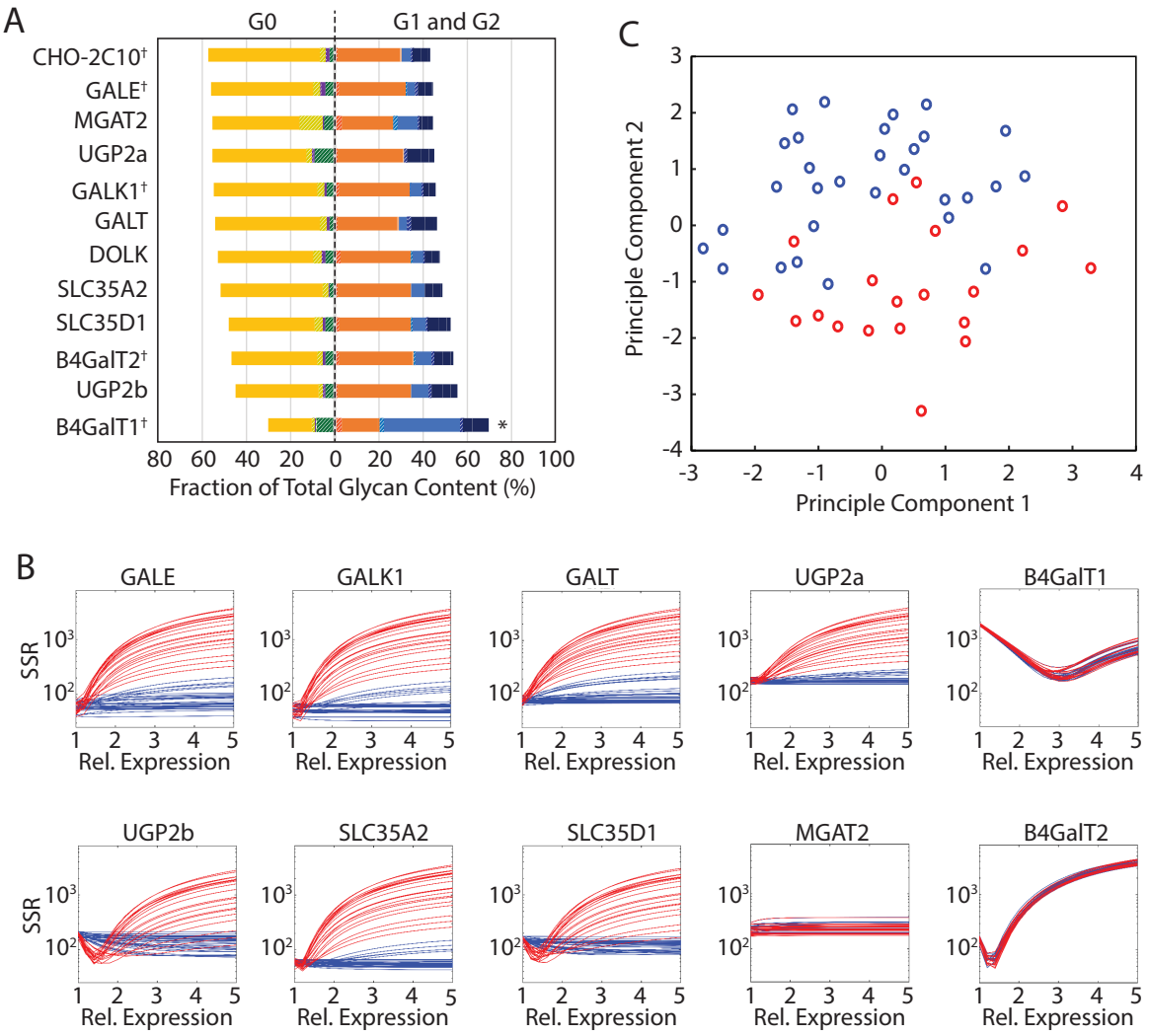


Figure 6.3: Single-gene overexpression and model refinement.

(A) Glycan profile for CHO-2C10 and single-gene overexpression cells. Bars are colored according to glycan structure using the legend in Figure 6.2 and are arranged to highlight the different total fractions of agalactosylated ('G0', left of dashed line) and mono- or bi-galactosylated glycans ('G1' and 'G2', respectively; right of dashed line). † denotes glycan analysis from triplicate experiments, * denotes significantly different galactose incorporation based on $\chi 2$-analysis, with p-value < 0.05 after Bonferonni correction for multiple comparisons (raw p-value < 0.0045). (B) Kinetic model robustness analysis for overexpression of UDP-galactose biosynthesis/transport genes and glycosyltransferases from single-gene overexpression experiment. Red and blue traces represent

different parameter combinations, with blue indicated the down-selected subset carried forward to future modeling. (C) Principle components analysis of 50 parameter sets with points colored as in (B).

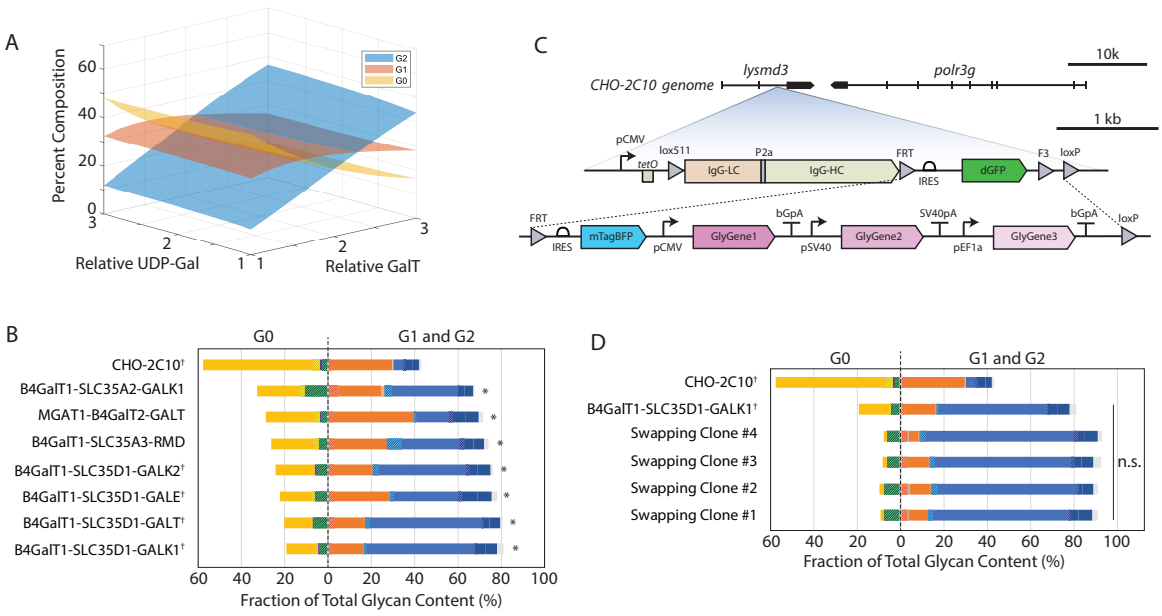


Figure 6.4: Model-driven multi-gene glycoengineering.

(A) Kinetic modeling of the interdependence of UDP-galactose levels and galactosyltransferase activity on galactose incorporation levels. (B) Glycan profile of original CHO-2C10 and engineered cells with randomly integrated three-gene glycosylation constructs. Bars are colored according to glycan structure using the legend in Figure 6.2 and are arranged to highlight the different total fractions of agalactosylated ('G0', left of dashed line) and mono- or bi-galactosylated glycans ('G1' and 'G2', respectively; right of dashed line). † denotes glycans measured from triplicate experiments, * denotes significantly different galactose incorporation based on ^{2}-analysis, with p-value < 0.05 after Bonferonni correction for multiple comparisons (raw p-value < 0.007). (C) Genomic location and organization of dual RMCE landing pad, including integration design for three-gene glycosylation constructs. (D) Glycan profile of original CHO-2C10 and engineered cell lines highlighting reproducibility of glycan perturbation following site-specific integration. n.s. denotes non-significant difference of swapping clones from randomly integrated B4GalT1-SLC35D1-GALK1 construct, based on ^{2}-analysis. All are significantly different from CHO-2C10.

7. Conclusions and future directions

In this thesis, we have utilized optimization on numerically challenging biological models, both steady state and dynamic, with the aim to provide insight into underlying metabolic mechanisms and as a targeted method to identify mitigations to undesirable behavior. The methods developed and explored in this document may be used to streamline biotechnological development by utilizing models capable of extrapolation, and with the capacity to provide mechanistic insight into complex biological problems in an automated fashion. We have demonstrated numerous uses of optimization, with applications ranging from metabolic engineering for cell line development to manufacturing support and diagnostics of process variability.

We first detailed an optimization framework, applicable to numerically challenging models which can be used to understand metabolic phenomena and rewire them. This framework poses an objective function which removes the need for integer variables when identifying specific enzymes to target within a metabolic pathway, changing the MINLP to an NLP and greatly decreasing the computational complexity of the resulting problem. We demonstrate, using a model of central metabolism tuned for a fast-growth, lactate-producing cell, that the Warburg phenomenon may be mitigated through the use of metabolic engineering to change the expression of three or more enzymes simultaneously. This reinforces the experimental observation that the one-at-a-time approaches which have primarily used so far are unlikely to enact significant changes to lactate production without retarding growth, thus suggesting that a more profound perturbation of metabolism is necessary. We then use this same framework to explore a disease relevant system: gluconeogenesis. As gluconeogenesis is a traditionally difficult metabolic process to model, optimization is well suited to identify the different enzymatic conditions required for the utilization of the substrates most used as carbon sources for synthesizing glucose. Understanding the

enzymatic requirements for this critical pathway may help to elucidate key potential targets when glucose homeostasis is dysregulated in the body.

The metabolic model was combined with a reactor level model, a cell growth model, and a cell signaling model to form a systems *in silico* bioprocess model. This model, after being appropriately fit to manufacturing-scale experimental data using local optimization combined with other numerical techniques, was then used to identify a potential origin of the process variably and suggest mitigation strategies to reduce the number of low performing runs. It was found that changes which to reduce the vicious cycle of metabolism, reactor control action, and osmolarity, such as lowering input osmolarity or likewise increasing mass transfer between the liquid and the gas phases such as would come from scaling down the culture, were most effective in reducing the frequency and extent of lactate production in the late stage of culture.

In the final portion of the thesis, a model of *N*-glycosylation was used in conjunction with multiple rounds of genetic engineering experiments to 'close the loop,' allowing both the model and the experimental design to inform each other. This work highlights the benefits of such a design, but also some of the limitations present in some biological models. The nature of glycosylation, including the specific configuration of the golgi compartments in the cells for this work as well as the relatively limited feedback and regulation in the model, resulted in model uncertainty and overfitting. While the overfitting was partially mitigated through the experimental design through perturbing single parameters in the first round, it remained a challenge which limited the overall extrapolation capacity of the model.

The work in this thesis can serve as a basis both to streamline biological simulation to enable more routine use of optimization, but also to extend and improve the models. In particular, the dynamic model discussed in this thesis required black-box optimization algorithms for parameter estimation, as the metabolic shifts and model stiffness prevented a straightforward discretization.

The use of model reduction techniques would enable a more rigorous exploration of the parameter space with reduced computational cost. Additionally, models contained herein may be extended to include additional reactions and regulations, particularly key interactions between metabolic enzymes and cell signaling as well as many amino acid metabolic pathways are neglected. Expansion of the model will enable additional optimization applications and study of the complex interplay between the highly dimensional regulatory networks that guide metabolic homeostasis.

8. Bibliography

1. Walsh, G. (2018) Biopharmaceutical benchmarks 2018. Nat Biotechnol 36 (12), 1136-1145.
2. Mulukutla, B.C. et al. (2010) Glucose metabolism in mammalian cell culture: new insights for tweaking vintage pathways. Trends in Biotechnology 28 (9), 476-484.
3. Mulukutla, B.C. et al. (2016) Regulation of Glucose Metabolism - A Perspective From Cell Bioprocessing. Trends in Biotechnology 34 (8), 638-651.
4. Chaneton, B. et al. (2012) Serine is a natural ligand and allosteric activator of pyruvate kinase M2. Nature 491 (7424), 458-462.
5. Yan, M. et al. (2016) Succinyl-5-aminoimidazole-4-carboxamide-1-ribose 5 '-Phosphate (SAICAR) Activates Pyruvate Kinase Isoform M2 (PKM2) in Its Dimeric Form. Biochemistry 55 (33), 4731-4736.
6. Keller, K.E. et al. (2012) SAICAR stimulates pyruvate kinase isoform M2 and promotes cancer cell survival in glucose-limited conditions. Science 338 (6110), 1069-72.
7. Amelio, I. et al. (2014) Serine and glycine metabolism in cancer. Trends Biochem Sci 39 (4), 191-8.
8. Hitosugi, T. et al. (2009) Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth. Sci Signal 2 (97), ra73.
9. Abeywardana, T. et al. (2018) CARM1 suppresses de novo serine synthesis by promoting PKM2 activity. Journal of Biological Chemistry 293 (39), 15290-15303.
10. Liu, F. et al. (2017) PKM2 methylation by CARM1 activates aerobic glycolysis to promote tumorigenesis. Nat Cell Biol 19 (11), 1358-1370.
11. Sumit, M. et al. (2019) Dissecting N-Glycosylation Dynamics in Chinese Hamster Ovary Cells Fed-batch Cultures using Time Course Omics Analyses. iScience 12, 102-120.
12. Hanover, J.A. et al. (2018) O-GlcNAc in cancer: An Oncometabolism-fueled vicious cycle. J Bioenerg Biomembr 50 (3), 155-173.
13. Wang, Y. et al. (2017) O-GlcNAcylation destabilizes the active tetrameric PKM2 to promote the Warburg effect. Proc Natl Acad Sci U S A 114 (52), 13732-13737.
14. Yi, W. et al. (2012) Phosphofructokinase glycosylation regulates cell growth and metabolism. Science 337, 975-980.
15. Rao, X. et al. (2015) O-GlcNAcylation of G6PD promotes the pentose phosphate pathway and tumor growth. Nat Commun 6, 8468.
16. Moellering, R.E. and Cravatt, B.F. (2013) Functional Lysine Modification by an Intrinsically Reactive Primary Glycolytic Metabolite. Science 341 (6145), 549-553.
17. Collard, F. et al. (2016) A conserved phosphatase destroys toxic glycolytic side products in mammals and yeast. Nat Chem Biol 12 (8), 601-7.
18. Nadtochiy, S.M. et al. (2016) Acidic pH Is a Metabolic Switch for 2-Hydroxyglutarate Generation and Signaling. Journal of Biological Chemistry 291 (38), 20188-20197.
19. Mishra, P. et al. (2018) ADHFE1 is a breast cancer oncogene and induces metabolic reprogramming. J Clin Invest 128 (1), 323-340.
20. Ye, D. et al. (2018) Metabolism, Activity, and Targeting of D- and L-2-Hydroxyglutarates. Trends Cancer 4 (2), 151-165.
21. Locasale, J.W. and Cantley, L.C. (2011) Metabolic Flux and the Regulation of Mammalian Cell Growth. Cell Metabolism 14 (4), 443-451.
22. Gaude, E. et al. (2018) NADH Shuttling Couples Cytosolic Reductive Carboxylation of Glutamine with Glycolysis in Cells with Mitochondrial Dysfunction. Molecular Cell 69 (4), 581-+.
23. Titov, D.V. et al. (2016) Complementation of mitochondrial electron transport chain by manipulation of the NAD(+)/NADH ratio. Science 352 (6282), 231-235.

24. Bulutoglu, B. et al. (2016) Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes. ACS Chem Biol 11 (10), 2847-2853.
25. Svedruzic, Z.M. and Spivey, H.O. (2006) Interaction between mammalian glyceraldehyde-3-phosphate dehydrogenase and L-lactate dehydrogenase from heart and muscle. Proteins 63 (3), 501-11.
26. Boukouris, A.E. et al. (2016) Metabolic Enzymes Moonlighting in the Nucleus: Metabolic Regulation of Gene Transcription. Trends in Biochemical Sciences 41 (8), 712-730.
27. Lew, C.R. and Tolan, D.R. (2012) Targeting of Several Glycolytic Enzymes Using RNA Interference Reveals Aldolase Affects Cancer Cell Proliferation through a Non-glycolytic Mechanism. Journal of Biological Chemistry 287 (51), 42554-42563.
28. Ciesla, M. et al. (2014) Fructose bisphosphate aldolase is involved in the control of RNA polymerase III-directed transcription. Biochimica Et Biophysica Acta 1843, 1103-1110.
29. Tarze, A. et al. (2007) GAPDH, a novel regulator of the pro-apoptotic mitochondrial membrane permeabilization. Oncogene 26 (18), 2606-20.
30. Jayaguru, P. and Mohr, S. (2011) Nuclear GAPDH: changing the fate of Muller cells in diabetes. J Ocul Biol Dis Infor 4 (1-2), 34-41.
31. He, C.-L. et al. (2016) Pyruvate Kinase M2 Activates mTORC1 by Phosphorylating AKT1S1. Scientific Reports 6, 21524.
32. Spinelli, J.B. and Haigis, M.C. (2018) The multifaceted contributions of mitochondria to cellular metabolism. Nat Cell Biol 20 (7), 745-754.
33. Palmieri, F. (2013) The mitochondrial transporter family SLC25: identification, properties and physiopathology. Mol Aspects Med 34 (2-3), 465-84.
34. Vincent, E.E. et al. (2015) Mitochondrial Phosphoenolpyruvate Carboxykinase Regulates Metabolic Adaptation and Enables Glucose-Independent Tumor Growth. Molecular Cell 60 (2), 195-207.
35. Katt, W.P. et al. (2017) A tale of two glutaminases: homologous enzymes with distinct roles in tumorigenesis. Future Medicinal Chemistry 9, 223-243.
36. Wise, D.R. et al. (2008) Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. Proc Natl Acad Sci U S A 105 (48), 18782-7.
37. Cassago, A. et al. (2012) Mitochondrial localization and structure-based phosphate activation mechanism of Glutaminase C with implications for cancer metabolism. Proc Natl Acad Sci U S A 109 (4), 1092-7.
38. Hu, W. et al. (2010) Glutaminase 2, a novel p53 target gene regulating energy metabolism and antioxidant function. Proc Natl Acad Sci U S A 107 (16), 7455-60.
39. Suzuki, S. et al. (2010) Phosphate-activated glutaminase (GLS2), a p53-inducible regulator of glutamine metabolism and reactive oxygen species. Proc Natl Acad Sci U S A 107 (16), 7461-6.
40. Yang, C.D. et al. (2014) Glutamine Oxidation Maintains the TCA Cycle and Cell Survival during Impaired Mitochondrial Pyruvate Transport. Molecular Cell 56 (3), 414-424.
41. Coloff, J.L. et al. (2016) Differential Glutamate Metabolism in Proliferating and Quiescent Mammary Epithelial Cells. Cell Metabolism 23 (5), 867-880.
42. Mastorodemos, V. et al. (2009) Human GLUD1 and GLUD2 glutamate dehydrogenase localize to mitochondria and endoplasmic reticulum. Biochem Cell Biol 87 (3), 505-16.
43. Pavlova, N.N. et al. (2018) As Extracellular Glutamine Levels Decline, Asparagine Becomes an Essential Amino Acid. Cell Metab 27 (2), 428-438 e5.
44. Palmieri, F. and Monne, M. (2016) Discoveries, metabolic roles and diseases of mitochondrial carriers: A review. Biochimica Et Biophysica Acta-Molecular Cell Research 1863 (10), 2362-2378.
45. Ahn, W.S. and Antoniewicz, M.R. (2013) Parallel labeling experiments with [1,2-(13)C]glucose and [U-(13)C]glutamine provide new insights into CHO cell metabolism. Metab Eng 15, 34-47.

46. Manning, B.D. and Toker, A. (2017) AKT/PKB Signaling: Navigating the Network. Cell 169 (3), 381-405.
47. Weichhart, T. (2018) mTOR as Regulator of Lifespan, Aging, and Cellular Senescence: A Mini-Review. Gerontology 64 (2), 127-134.
48. Hsieh, A.L. et al. (2015) MYC and metabolism on the path to cancer. Semin Cell Dev Biol 43, 11-21.
49. Eales, K.L. et al. (2016) Hypoxia and metabolic adaptation of cancer cells. Oncogenesis 5 (1), e190-e190.
50. Mossmann, D. et al. (2018) mTOR signalling and cellular metabolism are mutual determinants in cancer. Nature Reviews Cancer 18, 744-757.
51. Saxton, R.A. and Sabatini, D.M. (2017) mTOR signaling in growth, metabolism, and disease. Cell 168 (960-976.).
52. Valvezan, A.J. et al. (2017) mTORC1 Couples Nucleotide Synthesis to Nucleotide Demand Resulting in a Targetable Metabolic Vulnerability. Cancer Cell 32 (5), 624-+.
53. Hagiwara, A. et al. (2012) Hepatic mTORC2 activates glycolysis and lipogenesis through Akt, glucokinase, and SREBP1c. Cell Metab 15 (5), 725-38.
54. Lee, J.V. et al. (2014) Akt-Dependent Metabolic Reprogramming Regulates Tumor Cell Histone Acetylation. Cell Metabolism 20 (2), 306-319.
55. Jewell, J.L. et al. (2015) Differential regulation of mTORC1 by leucine and glutamine. Science 347 (6218), 194-198.
56. Chantranupong, L. et al. (2016) The CASTOR Proteins Are Arginine Sensors for the mTORC1 Pathway. Cell 165 (1), 153-164.
57. Hardie, D.G. et al. (2012) AMPK: a nutrient and energy sensor that maintains energy homeostasis. Nat Rev Mol Cell Biol 13 (4), 251-62.
58. Zhang, C.S. et al. (2017) Fructose-1,6-bisphosphate and aldolase mediate glucose sensing by AMPK. Nature 548 (7665), 112-116.
59. Herzig, S. and Shaw, R.J. (2018) AMPK: guardian of metabolism and mitochondrial homeostasis. Nat Rev Mol Cell Biol 19 (2), 121-135.
60. Garcia, D. and Shaw, R.J. (2017) AMPK: Mechanisms of Cellular Energy Sensing and Restoration of Metabolic Balance. Molecular Cell 66 (6), 789-800.
61. Gongol, B. et al. (2018) AMPK: An Epigenetic Landscape Modulator. International Journal of Molecular Sciences 19 (10), 19.
62. Lin, S.C. and Hardie, D.G. (2018) AMPK: Sensing Glucose as well as Cellular Energy Status. Cell Metabolism 27 (2), 299-313.
63. Patwari, P. et al. (2006) The interaction of thioredoxin with Txnip. Evidence for formation of a mixed disulfide by disulfide exchange. J Biol Chem 281 (31), 21884-91.
64. Patwari, P. et al. (2009) Thioredoxin-independent regulation of metabolism by the alpha-arrestin proteins. J Biol Chem 284 (37), 24996-5003.
65. Stoltzman, C.A. et al. (2008) Glucose sensing by MondoA : Mlx complexes: A role for hexokinases and direct regulation of thioredoxin-interacting protein expression. Proc Natl Acad Sci U S A 105 (19), 6912-6917.
66. Wu, N. et al. (2013) AMPK-Dependent Degradation of TXNIP upon Energy Stress Leads to Enhanced Glucose Uptake via GLUT1. Molecular Cell 49 (6), 1167-1175.
67. Chen, J.L.Y. et al. (2010) Lactic Acidosis Triggers Starvation Response with Paradoxical Induction of TXNIP through MondoA. Plos Genetics 6 (9).
68. Wilde, B.R. et al. (2019) Cellular acidosis triggers human MondoA transcriptional activity by driving mitochondrial ATP production. Elife 8.
69. Waldhart, A.N. et al. (2017) Phosphorylation of TXNIP by AKT Mediates Acute Influx of Glucose in Response to Insulin. Cell Reports 19 (10), 2005-2013.

70. Kaadige, M.R. et al. (2015) MondoA-Mlx Transcriptional Activity Is Limited by mTOR-MondoA Interaction. Molecular and Cellular Biology 35 (1), 101-110.
71. Shen, L.L. et al. (2015) Metabolic reprogramming in triple-negative breast cancer through Myc suppression of TXNIP. Proc Natl Acad Sci U S A 112 (17), 5425-5430.
72. Lee, D.C. et al. (2015) A Lactate-Induced Response to Hypoxia. Cell 161 (3), 595-609.
73. Uhlen, M. et al. (2015) Proteomics. Tissue-based map of the human proteome. Science 347 (6220), 1260419.
74. Husted, A.S. et al. (2017) GPCR-Mediated Signaling of Metabolites. Cell Metab 25 (4), 777-796.
75. Ahmed, K. et al. (2010) An autocrine lactate loop mediates insulin-dependent inhibition of lipolysis through GPR81. Cell Metab 11 (4), 311-9.
76. Liu, C. et al. (2009) Lactate inhibits lipolysis in fat cells through activation of an orphan G-protein-coupled receptor, GPR81. J Biol Chem 284 (5), 2811-22.
77. Roland, C.L. et al. (2014) Cell Surface Lactate Receptor GPR81 Is Crucial for Cancer Cell Survival. Cancer Res 74 (18), 5301-5310.
78. Mulukutla, B.C. et al. (2019) Metabolic engineering of Chinese hamster ovary cells towards reduced biosynthesis and accumulation of novel growth inhibitors in fed-batch cultures. Metab Eng 54, 54-68.
79. Altamirano, C. et al. (2006) Considerations on the lactate consumption by CHO cells in the presence of galactose. J Biotechnol 125 (4), 547-56.
80. Gagnon, M. et al. (2011) High-end pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures. Biotechnol Bioeng 108 (6), 1328-37.
81. Matthews, T.E. et al. (2016) Closed Loop Control of Lactate Concentration in Mammalian Cell Culture by Raman Spectroscopy Leads to Improved Cell Density, Viability, and Biopharmaceutical Protein Production. Biotechnol Bioeng 113 (11), 2416-2424.
82. Wlaschin, K.F. and Hu, W.S. (2007) Engineering cell metabolism for high-density cell culture via manipulation of sugar transport. J Biotechnol 131 (2), 168-176.
83. Jeong, D.W. et al. (2006) Effects of lactate dehydrogenase suppression and glycerol-3-phosphate dehydrogenase overexpression on cellular metabolism. Mol Cell Biochem 284 (1-2), 1-8.
84. Noh, S.M. et al. (2016) Reduction of ammonia and lactate through the coupling of glutamine synthetase selection and downregulation of lactate dehydrogenase-A in CHO cells. Applied Microbiology and Biotechnology, 1-11.
85. Zhou, M.X. et al. (2011) Decreasing lactate level and increasing antibody production in Chinese Hamster Ovary cells (CHO) by reducing the expression of lactate dehydrogenase and pyruvate dehydrogenase kinases. J Biotechnol 153 (1-2), 27-34.
86. Yip, S.S.M. et al. (2014) Complete Knockout of the Lactate Dehydrogenase A Gene is Lethal in Pyruvate Dehydrogenase Kinase 1, 2, 3 Down-Regulated CHO Cells. Molecular Biotechnology 56 (9), 833-838.
87. Torres, M. et al. (2018) Process and metabolic engineerign perspectives of lactate production in mammalian cell cultures. Curr Opinon in Chemical Engineering 22, 184-190.
88. Fogolin, M.B. et al. (2004) Impact of temperature reduction and expression of yeast pyruvate carboxylase on hGM-CSF-producing CHO cells. J Biotechnol 109 (1-2), 179-191.
89. Wilkens, C.A. and Gerdtzen, Z.P. (2015) Comparative Metabolic Analysis of CHO Cell Clones Obtained through Cell Engineering, for IgG Productivity, Growth and Cell Longevity. Plos One 10 (3).
90. Vidova, V. and Spacil, Z. (2017) A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. Anal Chim Acta 964, 7-23.
91. Mulukutla, B.C. et al. (2017) Identification and control of novel growth inhibitors in fed-batch cultures of Chinese hamster ovary cells. Biotechnol Bioeng 114 (8), 1779-1790.

92. Antoniewicz, M.R. (2018) A guide to (13)C metabolic flux analysis for the cancer biologist. Exp Mol Med 50 (4), 19.
93. Jang, C. et al. (2018) Metabolomics and Isotope Tracing. Cell 173 (4), 822-837.
94. Lu, W. et al. (2017) Metabolite Measurement: Pitfalls to Avoid and Practices to Follow. Annu Rev Biochem 86, 277-304.
95. Chen, W.W. et al. (2016) Absolute Quantification of Matrix Metabolites Reveals the Dynamics of Mitochondrial Metabolism. Cell 166 (5), 1324-1337 e11.
96. Nonnenmacher, Y. et al. (2017) Analysis of mitochondrial metabolism in situ: Combining stable isotope labeling with selective permeabilization. Metab Eng 43 (Pt B), 147-155.
97. Mulukutla, B.C. et al. (2014) Bistability in Glycolysis Pathway as a Physiological Switch in Energy Metabolism. Plos One 9 (6).
98. Gatenby, R.A. and Gillies, R.J. (2004) Why do cancers have high aerobic glycolysis? Nature Reviews Cancer 4 (11), 891-899.
99. Lunt, S.Y. and Vander Heiden, M.G. (2011) Aerobic Glycolysis: Meeting the Metabolic Requirements of Cell Proliferation. Annual Review of Cell and Developmental Biology, Vol 27 27, 441-464.
100. Chen, Z. et al. (2007) The Warburg effect and its cancer therapeutic implications. Journal of Bioenergetics and Biomembranes 39 (3), 267-274.
101. Lao, M.S. and 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.
102. Le, H. et al. (2012) Multivariate analysis of cell culture bioprocess data--lactate consumption as process indicator. J Biotechnol 162 (2-3), 210-23.
103. Toussaint, C. et al. (2016) Metabolic engineering of CHO cells to alter lactate metabolism during fed-batch cultures. Journal of Biotechnology 217, 122-131.
104. Kim, S.H. and Lee, G.M. (2007) Down-regulation of lactate dehydrogenase-A by siRNAs for reduced lactic acid formation of Chinese hamster ovary cells producing thrombopoietin. Applied Microbiology and Biotechnology 74 (1), 152-159.
105. Paredes, C. et al. (1999) Modification of glucose and glutamine metabolism in hybridoma cells through metabolic engineering. Cytotechnology 30 (1-3), 85-93.
106. Donthi, R.V. et al. (2004) Cardiac expression of kinase-deficient 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase inhibits glycolysis, promotes hypertrophy, impairs myocyte function, and reduces insulin sensitivity. Journal of Biological Chemistry 279 (46), 48085-48090.
107. Bensaad, K. et al. (2006) TIGAR, a p53-inducible regulator of glycolysis and apoptosis. Cell 126 (1), 107-120.
108. Cheung, E.C. et al. (2013) TIGAR Is Required for Efficient Intestinal Regeneration and Tumorigenesis. Developmental Cell 25 (5), 463-477.
109. Christofk, H.R. et al. (2008) The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature 452 (7184), 230-U74.
110. Israelsen, W.J. et al. (2013) PKM2 Isoform-Specific Deletion Reveals a Differential Requirement for Pyruvate Kinase in Tumor Cells. Cell 155 (2), 397-409.
111. Kacser, H. and Burns, J.A. (1981) The Molecular-Basis of Dominance. Genetics 97 (3-4), 639-666.
112. Hatzimanikatis, V. et al. (1996) Analysis and design of metabolic reaction networks via mixed-integer linear optimization. Aiche Journal 42 (5), 1277-1292.
113. Polisetty, P.K. et al. (2008) Yield optimization of regulated metabolic systems using deterministic branch-and-reduce methods. Biotechnol Bioeng 99 (5), 1154-1169.
114. Millard, P. et al. (2017) Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in Escherichia coli. Plos Computational Biology 13 (2).

115. Konig, M. et al. (2012) Quantifying the Contribution of the Liver to Glucose Homeostasis: A Detailed Kinetic Model of Human Hepatic Glucose Metabolism. Plos Computational Biology 8 (6).
116. Mulukutla, B.C. et al. (2015) Multiplicity of Steady States in Glycolysis and Shift of Metabolic State in Cultured Mammalian Cells. Plos One 10 (3).
117. Rizzi, M. et al. (1997) In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae .2. Mathematical model. Biotechnology and Bioengineering 55 (4), 592-608.
118. Chowdhury, A. et al. (2014) k-OptForce: Integrating Kinetics with Flux Balance Analysis for Strain Design. Plos Computational Biology 10 (2).
119. Banga, J.R. (2008) Optimization in computational systems biology. Bmc Systems Biology 2.
120. Liu, P.K. and Wang, F.S. (2008) Inference of biochemical network models in S-system using multiobjective optimization approach. Bioinformatics 24 (8), 1085-1092.
121. Cotten, C. and Reed, J.L. (2013) Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. Biotechnology Journal 8 (5), 595-604.
122. Yang, L. et al. (2011) EMILiO: A fast algorithm for genome-scale strain design. Metabolic Engineering 13 (3), 272-281.
123. Zomorrodi, A.R. et al. (2012) Mathematical optimization applications in metabolic networks. Metabolic Engineering 14 (6), 672-686.
124. Gawand, P. et al. (2013) Novel approach to engineer strains for simultaneous sugar utilization. Metabolic Engineering 20, 63-72.
125. Gerdtzen, Z.P. et al. (2004) Non-linear reduction for kinetic models of metabolic reaction networks. Metabolic Engineering 6 (2), 140-154.
126. Costa, R.S. et al. (2016) Kinetic modeling of cell metabolism for microbial production. Journal of Biotechnology 219, 126-141.
127. Sendin, J.O.H. et al. (2010) Multi-objective mixed integer strategy for the optimisation of biological networks. Iet Systems Biology 4 (3), 236-248.
128. Vera, J. et al. (2010) Optimization of biochemical systems through mathematical programming: Methods and applications. Computers & Operations Research 37 (8), 1427-1438.
129. Visser, D. et al. (2004) Optimal re-design of primary metabolism in Escherichia coli using linlog kinetics. Metabolic Engineering 6 (4), 378-390.
130. Vital-Lopez, F.G. et al. (2006) A computational procedure for optimal engineering interventions using kinetic models of metabolism. Biotechnology Progress 22 (6), 1507-1517.
131. Boron, W.F., ; Boulpaep, E. L. (2008) Medical Physiology, 2nd edn., Elsevier Health Sciences.
132. DeBerardinis, R.J. (2011) Serine Metabolism: Some Tumors Take the Road Less Traveled. Cell Metabolism 14 (3), 285-286.
133. Locasale, J.W. et al. (2011) Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis. Nature Genetics 43 (9), 869-U79.
134. Maddocks, O.D.K. et al. (2016) Serine Metabolism Supports the Methionine Cycle and DNA/RNA Methylation through De Novo ATP Synthesis in Cancer Cells. Molecular Cell 61 (2), 210-221.
135. Snell, K. et al. (1987) The Modulation of Serine Metabolism in Hepatoma 3924a during Different Phases of Cellular Proliferation in Culture. Biochemical Journal 245 (2), 609-612.
136. Gill, P.E. et al. (2005) SNOPT: An SQP algorithm for large-scale constrained optimization. Siam Review 47 (1), 99-131.
137. Chinchuluun, A. and Pardalos, P.M. (2007) A survey of recent developments in multiobjective optimization. Annals of Operations Research 154 (1), 29-50.
138. van der Maaten, L. and Hinton, G. (2008) Visualizing Data using t-SNE. Journal of Machine Learning Research 9, 2579-2605.

139. Ahn, W.S. and Antoniewicz, M.R. (2011) Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. Metab Eng 13 (5), 598-609.
140. Liberti, M.V. and Locasale, J.W. (2016) The Warburg Effect: How Does it Benefit Cancer Cells? Trends in Biochemical Sciences 41 (3), 211-218.
141. Chen, K.Q. et al. (2001) Engineering of a mammalian cell line for reduction of lactate formation and high monoclonal antibody production. Biotechnology and Bioengineering 72 (1), 55-61.
142. Fantin, V.R. et al. (2006) Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance (vol 9, pg 425, 2006). Cancer Cell 10 (2), 172-172.
143. Wang, Z.Y. et al. (2012) LDH-A silencing suppresses breast cancer tumorigenicity through induction of oxidative stress mediated mitochondrial pathway apoptosis. Breast Cancer Research and Treatment 131 (3), 791-800.
144. Allison, S.J. et al. (2014) Identification of LDH-A as a therapeutic target for cancer cell killing via (i) p53/NAD(H)-dependent and (ii) p53-independent pathways. Oncogenesis 3.
145. Yongky, A. et al. (2015) Mechanism for multiplicity of steady states with distinct cell concentration in continuous culture of mammalian cells. Biotechnol Bioeng 112 (7), 1437-45.
146. Sunny, N.E. et al. (2011) Excessive hepatic mitochondrial TCA cycle and gluconeogenesis in humans with nonalcoholic fatty liver disease. Cell metabolism 14 (6), 804-810.
147. Basu, R. et al. (2005) Obesity and type 2 diabetes impair insulin-induced suppression of glycogenolysis as well as gluconeogenesis. Diabetes 54 (7), 1942-1948.
148. Garcia, C.K. et al. (1994) Molecular characterization of a membrane transporter for lactate, pyruvate, and other monocarboxylates: implications for the Cori cycle. Cell 76 (5), 865-873.
149. Felig, P. (1973) The glucose-alanine cycle. Metabolism 22 (2), 179-207.
150. Neeland, I.J. et al. (2017) Effects of visceral adiposity on glycerol pathways in gluconeogenesis. Metabolism 67, 80-89.
151. Meyer, C. et al. (2003) Relative importance of liver, kidney, and substrates in epinephrine-induced increased gluconeogenesis in humans. Am J Physiol Endocrinol Metab 285 (4), E819-26.
152. Hue, L. and Hers, H.-G. (1974) Utile and futile cycles in the liver. Biochemical and biophysical research communications 58 (3), 540-548.
153. Konig, M. et al. (2012) Quantifying the contribution of the liver to glucose homeostasis: a detailed kinetic model of human hepatic glucose metabolism. PLoS Comput Biol 8 (6), e1002577.
154. Berndt, N. et al. (2018) HEPATOKIN1 is a biochemistry-based model of liver metabolism for applications in medicine and pharmacology. Nat Commun 9 (1), 2386.
155. O'Brien, C. et al. (2019) Kinetic model optimization and its application to mitigating the Warburg effect through multiple enzyme alterations. Metab Eng 56, 154-164.
156. Weinberg, M.B. and Utter, M.F. (1979) Effect of thyroid hormone on the turnover of rat liver pyruvate carboxylase and pyruvate dehydrogenase. Journal of Biological Chemistry 254 (19), 9492-9499.
157. Hart, W.E. et al. (2017) Pyomo-optimization modeling in python, Second Edition edn., Springer.
158. Hart, W.E. et al. (2011) Pyomo: modeling and solving mathematical programs in Python. Mathematical Programming Computation 3 (3), 219.
159. Kılınç, M.R. and Sahinidis, N.V. (2018) Exploiting integrality in the global optimization of mixed-integer nonlinear programming problems with BARON. Optimization Methods and Software 33 (3), 540-562.
160. Wächter, A. and Biegler, L.T. (2006) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical programming 106 (1), 25-57.

161. Ramanan, S. and Grampp, G. (2014) Drift, evolution, and divergence in biologics and biosimilars manufacturing. BioDrugs 28 (4), 363-72.
162. Templeton, N. et al. (2013) Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. Biotechnol Bioeng 110 (7), 2013-24.
163. Fan, Y. et al. (2015) Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. Biotechnol Bioeng 112 (3), 521-35.
164. Xu, S. et al. (2018) Probing lactate metabolism variations in large-scale bioreactors. Biotechnol Prog 34 (3), 756-766.
165. Abu-Absi, S. et al. (2014) Cell culture process operations for recombinant protein production. Adv Biochem Eng Biotechnol 139, 35-68.
166. Xing, Z. et al. (2017) A carbon dioxide stripping model for mammalian cell culture in manufacturing scale bioreactors. Biotechnol Bioeng 114 (6), 1184-1194.
167. Mulukutla, B.C. et al. (2015) Multiplicity of steady states in glycolysis and shift of metabolic state in cultured mammalian cells. PLoS One 10 (3), e0121561.
168. Amribt, Z. et al. (2013) Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures. Biochemical Engineering Journal 70, 196-209.
169. Grilo, A.L. and Mantalaris, A. (2019) A Predictive Mathematical Model of Cell Cycle, Metabolism, and Apoptosis of Monoclonal Antibody-Producing GS-NS0 Cells. Biotechnol J 14 (11), e1800573.
170. Hernández Rodríguez, T. et al. (2019) Predicting industrial-scale cell culture seed trains–A Bayesian framework for model fitting and parameter estimation, dealing with uncertainty in measurements and model parameters, applied to a nonlinear kinetic cell culture model, using an MCMC method. Biotechnology and Bioengineering 116 (11), 2944-2959.
171. Alhuthali, S. et al. (2017) Multi-stage population balance model to understand the dynamics of fed-batch CHO cell culture. In Computer Aided Chemical Engineering (Espuña, A. et al. eds), pp. 2821-2826, Elsevier.
172. Kotidis, P. et al. (2019) Constrained global sensitivity analysis for bioprocess design space identification. Computers & Chemical Engineering 125, 558-568.
173. Kyriakopoulos, S. et al. (2018) Kinetic Modeling of Mammalian Cell Culture Bioprocessing: The Quest to Advance Biomanufacturing. Biotechnology Journal 13 (3), 1700229.
174. Mašić, A. et al. (2017) Shape constrained splines as transparent black-box models for bioprocess modeling. Computers & Chemical Engineering 99, 96-105.
175. Kappatou, C.D. et al. (2018) Model-Based Dynamic Optimization of Monoclonal Antibodies Production in Semibatch Operation—Use of Reformulation Techniques. Industrial & Engineering Chemistry Research 57 (30), 9915-9924.
176. Solle, D. et al. (2017) Between the poles of data-driven and mechanistic modeling for process operation. Chemie Ingenieur Technik 89 (5), 542-561.
177. Hong, M.S. et al. (2018) Challenges and opportunities in biopharmaceutical manufacturing control. Computers & Chemical Engineering 110, 106-114.
178. Narayanan, H. et al. (2020) Bioprocessing in the Digital Age: The Role of Process Models. Biotechnology Journal 15 (1), 1900172.
179. Ulonska, S. et al. (2018) Workflow for Target-Oriented Parametrization of an Enhanced Mechanistic Cell Culture Model. Biotechnol J 13 (4), e1700395.
180. Charaniya, S. et al. (2010) Mining manufacturing data for discovery of high productivity process characteristics. J Biotechnol 147 (3-4), 186-97.
181. Duran, J. et al. (2009) Pfkfb3 is transcriptionally upregulated in diabetic mouse liver through proliferative signals. FEBS J 276 (16), 4555-68.
182. Marsin, A.S. et al. (2000) Phosphorylation and activation of heart PFK-2 by AMPK has a role in the stimulation of glycolysis during ischaemia. Curr Biol 10 (20), 1247-55.

183. Novellasdemunt, L. et al. (2013) Akt-dependent activation of the heart 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (PFKFB2) isoenzyme by amino acids. J Biol Chem 288 (15), 10640-51.
184. Martini, M. et al. (2014) PI3K/AKT signaling pathway and cancer: an updated review. Ann Med 46 (6), 372-83.
185. Barnes, K. et al. (2002) Activation of GLUT1 by metabolic and osmotic stress: potential involvement of AMP-activated protein kinase (AMPK). J Cell Sci 115 (Pt 11), 2433-42.
186. Möller, J. et al. (2018) Model-based identification of cell-cycle-dependent metabolism and putative autocrine effects in antibody producing CHO cell culture. Biotechnology and Bioengineering 115 (12), 2996-3008.
187. Gambhir, A. et al. (1999) Analysis of the use of fortified medium in continuous culture of mammalian cells. Cytotechnology 31 (3), 243-54.
188. Mulukutla, B.C. et al. (2016) Regulation of Glucose Metabolism - A Perspective From Cell Bioprocessing. Trends Biotechnol 34 (8), 638-651.
189. Mulukutla, B.C. et al. (2012) On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. Metab Eng 14 (2), 138-49.
190. Luo, J. et al. (2012) Comparative metabolite analysis to understand lactate metabolism shift in Chinese hamster ovary cell culture process. Biotechnol Bioeng 109 (1), 146-56.
191. Zagari, F. et al. (2013) Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity. N Biotechnol 30 (2), 238-45.
192. Yang, H.-C. et al. (2019) The redox role of G6PD in cell growth, cell death, and cancer. Cells 8 (9), 1055.
193. Gerdtzen, Z.P. et al. (2004) Non-linear reduction for kinetic models of metabolic reaction networks. Metab Eng 6 (2), 140-54.
194. Rao, S. et al. (2014) A model reduction method for biochemical reaction networks. BMC systems biology 8 (1), 52.
195. Snowden, T.J. et al. (2017) A combined model reduction algorithm for controlled biochemical systems. BMC systems biology 11 (1), 17.
196. Walsh, G., Biopharmaceutical benchmarks 2014, Nature Biotechnology, Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 2014, p. 992.
197. Browne, S.M. and Al-Rubeai, M., Selection methods for high-producing mammalian cell lines, Trends in Biotechnology, 2007, pp. 425-432.
198. Wells, E. and Robinson, A.S., Cellular engineering for therapeutic protein production: product quality, host modification, and process improvement, Biotechnology Journal, 2017.
199. Kornfeld, R., Assembly of Asparagine-Linked Oligosaccharides, Annual Review of Biochemistry, 2002, pp. 631-664.
200. Helenius, A. and Aebi, M., Roles of N-Linked Glycans in the Endoplasmic Reticulum, Annual Review of Biochemistry, 2004, pp. 1019-1049.
201. Dwek, R.A., Biological importance of glycosylation., Developments in biological standardization, 1998, pp. 43-47.
202. Boyd, P.N. et al., The effect of the removal of sialic acid, galactose and total carbohydrate on the functional activity of Campath-1H, Molecular Immunology, 1995, pp. 1311-1318.
203. Hodoniczky, J. et al., Control of recombinant monoclonal antibody effector functions by Fc N-glycan remodeling in vitro, Biotechnology Progress, 2005, pp. 1644-1652.
204. Raju, T.S., Terminal sugars of Fc glycans influence antibody effector functions of IgGs, Current Opinion in Immunology, 2008, pp. 471-478.
205. Wright, A. and Morrison, S.L., Effect of C2-associated carbohydrate structure on Ig effector function: studies with chimeric mouse-human IgG1 antibodies in glycosylation mutants of Chinese hamster ovary cells, J Immunol, 1998, pp. 3393-3402.

206. Kaneko, Y. et al., Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation, Science, 2006, pp. 670-673.
207. Hossler, P. et al., Systems analysis of N-glycan processing in mammalian cells, PLoS ONE, 2007.
208. Jimenez del Val, I. et al., A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus, Biotechnology Progress, 2011, pp. 1730-1743.
209. Krambeck, F.J. and Betenbaugh, M.J., A mathematical model of N-linked glycosylation, Biotechnology and Bioengineering, 2005, pp. 711-728.
210. Clausen, H. et al., Glycosylation Engineering, Essentials of Glycobology, 2017.
211. Shinkawa, T. et al., The Absence of Fucose but Not the Presence of Galactose or Bisecting N -Acetylglucosamine of Human IgG1 Complex-type Oligosaccharides Shows the Critical Role of Enhancing Antibody-dependent Cellular Cytotoxicity * those produced by Chinese hamster ovary (CHO) cell, 2003, pp. 3466-3473.
212. Phenotypes, L.-r. et al., Lectin-Resistant CHO Cells : Selection of Four New Pea, 1986, pp. 51-62.
213. Shields, R.L. et al., Lack of Fucose on Human IgG1 N -Linked Oligosaccharide Improves Binding to Human Fc RIII and Antibody-dependent Cellular Toxicity *, 2002, pp. 26733-26740.
214. Yamane-ohnuki, N. et al., Establishment of FUT8 Knockout Chinese Hamster Ovary Cells : An Ideal Host Cell Line for Producing Completely Defucosylated Antibodies With Enhanced Antibody-Dependent Cellular Cytotoxicity, 2004.
215. Malphettes, L. et al., Highly Efficient Deletion of FUT8 in CHO Cell Lines Using Zinc-Finger Nucleases Yields Cells That Produce Completely Nonfucosylated Antibodies, 2010, pp. 774-783.
216. Imai-nishiya, H. et al., a new strategy for generating fully non-fucosylated therapeutic antibodies with enhanced ADCC, 2007, pp. 1-13.
217. Jeong, Y.T. et al., Enhanced sialylation of recombinant erythropoietin in CHO cells by human glycosyltransferase expression, Journal of Microbiology and Biotechnology, 2008.
218. O'Brien, S.A. et al., Single Copy Transgene Integration in a Transcriptionally Active Site for Recombinant Protein Synthesis, Biotechnology Journal, 2018.
219. Hsu, S.Y. and Smanski, M.J., Designing and implementing algorithmic DNA assembly pipelines for multi-gene systems, Methods in Molecular Biology, 2018, pp. 131-147.
220. Gibson, D.G. et al., Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome., Science (New York, N.Y.), 2008, pp. 1215-1220.
221. Kim, J.H. et al., High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice, PLoS ONE, 2011.
222. Lagarias, J.C. et al., Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions, SIAM J. on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, 1998, pp. 112-147.
223. Smanski, M.J. et al., Functional optimization of gene clusters by combinatorial design and assembly, Nature Biotechnology, Nature Publishing Group, 2014, pp. 1241-1249.
224. Szymczak, A.L. et al., Correction of multi-gene deficiency in vivo using a single'self-cleaving'2A peptide–based retroviral vector, Nature biotechnology, 2004, pp. 589-594.
225. Bigge, J.C. et al., Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid, Analytical Biochemistry, 1995, pp. 229-238.
226. Hossler, P. et al. (2006) GlycoVis: Visualizing glycan distribution in the protein N-glycosylation pathway in mammalian cells. Biotechnology and bioengineering 95 (5), 946-960.
227. Qasba, P.K. et al., Structure and Function of β-1,4-Galactosyltransferase, Current drug targets, 2008, pp. 292-309.
228. Nelder, J.A. and Mead, R., A Simplex Method for Function Minimization, The Computer Journal, 1965, pp. 308-313.

229. Dekkers, G. et al., Multi-level glyco-engineering techniques to generate IgG with defined Fc-glycans, Scientific Reports, 2016.
230. Keusch, J. et al., The effect on IgG glycosylation of altering $β$1,4-galactosyltransferase-1 activity in B cells, Glycobiology, 1998, pp. 1215-1220.
231. Yang, Z. et al., Engineered CHO cells for production of diverse, homogeneous glycoproteins, Nature Biotechnology, 2015, pp. 842-844.
232. Bydlinski, N. et al., The contributions of individual galactosyltransferases to protein speci fi c N- glycan processing in Chinese Hamster Ovary cells, Journal of Biotechnology, Elsevier, 2018, pp. 101-110.
233. Heinsch, S.C. et al., Simulation Modeling to Compare Optimization Strategies for Metabolic Engineering, 2018, pp. 1-8.
234. Kauffman, S., The Origins of Order, Oxford University Press, Oxford, 1993.
235. Muraoka, M. et al., Molecular characterization of human UDP-glucuronic acid / UDP- N - acetylgalactosamine transporter , a novel nucleotide sugar transporter with dual substrate speci ¢ city, 2001, pp. 87-93.
236. Wong, N.S.C. et al., Enhancing recombinant glycoprotein sialylation through CMP-sialic acid transporter over expression in Chinese hamster ovary cells, Biotechnology and Bioengineering, 2006.
237. Sumit, M. et al., Dissecting N-Glycosylation Dynamics in Chinese Hamster Ovary Cells Fed-batch Cultures using Time Course Omics Analyses Dissecting N-Glycosylation Dynamics in Chinese Hamster Ovary Cells Fed-batch Cultures using Time Course Omics Analyses, ISCIENCE, Elsevier Inc., 2019, pp. 102-120.
238. Kochanowski, N. et al., Intracellular nucleotide and nucleotide sugar contents of cultured CHO cells determined by a fast , sensitive , and high-resolution, 2006, pp. 243-251.
239. Hadley, B. et al., Structure and function of nucleotide sugar transporters : Current progress, CSBJ, Elsevier B.V., 2014, pp. 23-32.
240. Song, Z., Roles of the nucleotide sugar transporters (SLC35 family) in health and disease, Molecular Aspects of Medicine, 2013, pp. 590-600.
241. Yu, X. et al., Engineering Hydrophobic Protein − Carbohydrate Interactions to Fine-Tune Monoclonal Antibodies, 2013.
242. Guye, P. et al., Rapid , modular and reliable construction of complex mammalian gene circuits, 2013, pp. 3-8.
243. Jones, A.C. et al., Phage P1-Derived Artificial Chromosomes Facilitate Heterologous Expression of the FK506 Gene Cluster, 2013.
244. Ellies, L.G. et al. (1998) Core 2 oligosaccharide biosynthesis distinguishes between selectin ligands essential for leukocyte homing and inflammation. Immunity 9 (6), 881-890.
245. Mitoma, J. et al. (2007) Critical functions of N-glycans in L-selectin-mediated lymphocyte homing and recruitment. Nature Immunology 8 (4), 409-418.
246. Julien, S. et al. (2005) Stable expression of sialyl-Tn antigen in T47-D cells induces a decrease of cell adhesion and an increase of cell migration. Breast Cancer Research and Treatment 90 (1), 77-84.
247. Pinho, S. et al. (2007) Biological significance of cancer-associated sialyl-Tn antigen: Modulation of malignant phenotype in gastric carcinoma cells. Cancer Letters 249 (2), 157-170.
248. Johansson, M.E.V. et al. (2011) Composition and functional role of the mucus layers in the intestine. Cellular and Molecular Life Sciences 68 (22), 3635-3641.
249. Dube, S. et al. (1988) Glycosylation at Specific Sites of Erythropoietin Is Essential for Biosynthesis, Secretion, and Biological Function. Journal of Biological Chemistry 263 (33), 17516-17521.
250. Thim, L. et al. (2010) Purification and characterization of a new recombinant factor VIII (N8). Haemophilia 16 (2), 349-359.

251. van Schooten, C.J.M. et al. (2007) Variations in glycosylation of von Willebrand factor with O-linked sialylated T antigen are associated with its plasma levels. Blood 109 (6), 2430-2437.
252. Houel, S. et al. (2014) N- and O-Glycosylation Analysis of Etanercept Using Liquid Chromatography and Quadrupole Time-of-Flight Mass Spectrometry Equipped with Electron-Transfer Dissociation Functionality. Analytical Chemistry 86 (1), 576-584.
253. Van den Steen, P. et al. (1998) Concepts and principles of O-linked glycosylation. Critical Reviews in Biochemistry and Molecular Biology 33 (3), 151-208.
254. Gill, D.J. et al. (2013) Initiation of GalNAc-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. Proc Natl Acad Sci U S A 110 (34), E3152-E3161.
255. Liu, G. et al. (2013) Glycosylation Network Analysis Toolbox: a MATLAB-based environment for systems glycobiology. Bioinformatics 29 (3), 404-406.
256. Liu, G. and Neelamegham, S. (2014) A Computational Framework for the Automated Construction of Glycosylation Reaction Networks. Plos One 9 (6).
257. Kawano, S. et al. (2005) Prediction of glycan structures from gene expression data based on glycosyltransferase reactions. Bioinformatics 21 (21), 3976-3982.
258. McDonald, A.G. et al. (2016) A Knowledge-Based System for Display and Prediction of O-Glycosylation Network Behaviour in Response to Enzyme Knockouts. Plos Computational Biology 12 (4).
259. Rangarajan, S. et al. (2012) Language-oriented rule-based reaction network generation and analysis: Description of RING. Computers & Chemical Engineering 45, 114-123.
260. Rangarajan, S. et al. (2012) Language-oriented rule-based reaction network generation and analysis: Applications of RING. Computers & Chemical Engineering 46, 141-152.
261. Gupta, U. et al., Automated Network Generation and Analysis of Biochemical Reaction Pathways Using RING, In Press.
262. Zhang, L. et al. (2014) UDP-N-Acetyl-Alpha-D-Galactosamine: Polypeptide N-Acetylgalactosaminyltransferases (ppGalNAc-Ts). In Handbook of Glycosyltransferases and Related Genes (Taniguchi, N. et al. eds), pp. 495-511, Springer Japan.
263. Weininger, D. (1988) Smiles, a Chemical Language and Information-System .1. Introduction to Methodology and Encoding Rules. Journal of Chemical Information and Computer Sciences 28 (1), 31-36.
264. Cheng, K. et al. (2017) DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information. Glycobiology 27 (3), 200-205.
265. Uhlen, M. et al. (2017) A pathology atlas of the human cancer transcriptome. Science 357 (6352), 660-+.
266. Muller, S. and Hanisch, F.G. (2002) Recombinant MUC1 probe authentically reflects cell-specific O-glycosylation profiles of endogenous breast cancer mucin - High density and prevalent core 2-based glycosylation. Journal of Biological Chemistry 277 (29), 26103-26112.
267. Niemela, R. et al. (1998) Complementary acceptor and site specificities of Fuc-TIV and Fuc-TVII allow effective biosynthesis of sialyl-TriLex and related polylactosamines present on glycoprotein counterreceptors of selectins. J Biol Chem 273 (7), 4021-6.
268. Nishihara, S. et al. (1999) Alpha1,3-fucosyltransferase 9 (FUT9; Fuc-TIX) preferentially fucosylates the distal GlcNAc residue of polylactosamine chain while the other four alpha1,3FUT members preferentially fucosylate the inner GlcNAc residue. Febs Letters 462 (3), 289-94.
269. Turunen, J.P. et al. (1995) De novo expression of endothelial sialyl Lewis(a) and sialyl Lewis(x) during cardiac transplant rejection: superior capacity of a tetravalent sialyl Lewis(x) oligosaccharide in inhibiting L-selectin-dependent lymphocyte adhesion. J Exp Med 182 (4), 1133-41.
270. Kudelka, M.R. et al. (2016) Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells. Nature Methods 13 (1), 81-+.

271. Ikehara, Y. et al. (1999) Cloning and expression of a human gene encoding an N-acetylgalactosamine-alpha2,6-sialyltransferase (ST6GalNAc I): a candidate for synthesis of cancer-associated sialyl-Tn antigens. Glycobiology 9 (11), 1213-24.
272. Niemela, R. et al. (1998) Complementary acceptor and site specificities of Fuc-TIV and Fuc-TVII allow effective biosynthesis of sialyl-TriLex and related polylactosamines present on glycoprotein counterreceptors of selectins. Journal of Biological Chemistry 273 (7), 4021-4026.
273. Olson, F.J. et al. (2005) A MUC1 tandem repeat reporter protein produced in CHO-K1 cells has sialylated core 1 O-glycans and becomes more densely glycosylated if coexpressed with polypeptide-GalNAc-T4 transferase. Glycobiology 15 (2), 177-91.
274. Holgersson, J. and Lofling, J. (2006) Glycosyltransferases involved in type 1 chain and Lewis antigen biosynthesis exhibit glycan and core chain specificity. Glycobiology 16 (7), 584-593.
275. Hossler, P. et al. (2006) GlycoVis: visualizing glycan distribution in the protein N-glycosylation pathway in mammalian cells. Biotechnology and Bioengineering 95, 946-960.
276. Krambeck, F.J. et al. (2009) A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. Glycobiology 19 (11), 1163-75.

9. Appendix A: An integrated platform for mucin-type *O*-glycosylation network generation and visualization

Reproduced with permission from: Le, T.*, O'Brien, C.*, Gupta, U.*, Sousa, G., Daoutidis, P., Hu, W.S.. An integrated platform for mucin-type O-glycosylation network generation and visualization. Biotechnology and Bioengineering. 2019; 116: 1341– 1354. https://doi.org/10.1002/bit.26952

* These authors contributed equally to this work.

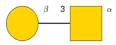
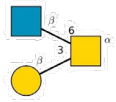
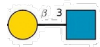
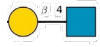
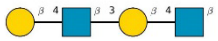
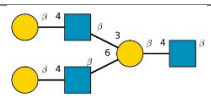
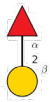
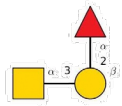
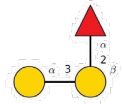
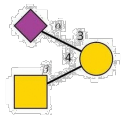
9.1. Introduction

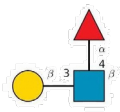
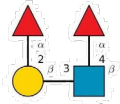
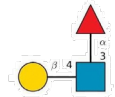
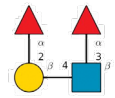
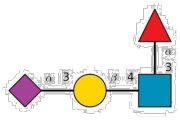
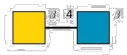
Mucin-type O-glycans (hereinafter referred to as O-glycans) are oligosaccharide moieties that are covalently attached to the Ser or Thr residues of the protein. O-glycans are abundantly present on cell surface proteins. They modulate the cell's interactions with the surrounding environments and other cells. O-glycans play an important role in cell-cell recognition in the immune system [244, 245]. Aberrant O-glycans on surface proteins of cancer cell lines change their adhesion characteristics and promote migration, invasion, and metastasis [246, 247]. O-glycans on secreted mucins contribute to the physical and chemical properties of mucins, enabling them to act as a general shield, protecting cells or tissues from direct contact with microorganisms [248].

O-glycans are also present in a number of glycoprotein therapeutics including recombinant erythropoietin [249], factor VIII [250], von Willebrand factor [251], and tumor necrosis factor receptor-Fc fusion protein [252]. They affect the stability, propensity to aggregation and pharmacokinetic properties of protein therapeutics (reviewed in [253]). Structural analysis of O-glycans is more challenging than N-glycans. Unlike N-glycosylation there is no consensus amino acid sequence for O-glycosylation. In general, O-glycosylation of proteins is less characterized than N-glycosylation.

O-glycosylation occurs mainly in the Golgi apparatus, albeit in tumor cells the initiation step might occur in the endoplasmic reticulum (ER) [254]. O-glycosylation is initiated by the transfer of an *N*-acetylgalactosamine (GalNAc) residue to the hydroxyl group of Ser or Thr in a peptide sequence, forming the Tn antigen structure (GalNAcαSer/Thr). The extension of the Tn antigen structure forms four major core structures (core 1-4). Core 1 is formed by the addition of a galactose to Tn antigen. The addition of an *N*-acetylglucosamine to the GalNAc of core 1 forms core 2. Core 3 results from the addition of a GlcNAc to Tn antigen. Core 4 is formed by branching of core 3 from its GalNAc with a β1-6GlcNAc. Core structures can be further extended to form complex O-glycans with linear or branched chains. A typical O-glycan structure may contain ABO and Lewis blood group determinants, polysialic, i and I antigens. Repetition of LacNAc (type 1, type 2 and type 3) chains provides frameworks for further elaboration of O-glycan with additional monosaccharides or the described functional groups (see Table A.1 for a list of common antigens).

Table 9.1. Classification of glycans in the O-glycosylation network of CHO cells based on the epitopes borne.

Epitope	SNFG representation	Percentage of glycan species bearing the epitope (%)			
		CHO	HUVEC	T47D	MCF7
Core 1 (and extended core 1) (a.k.a. type 3 LacNAc chain)		5.1	1.2	96.6	2.8
Core 2 (and extended core 2)		94.1	98.3	0	96
Type 1 LacNAc chain		7.8	0	0	0
Type 2 LacNAc chain		47	92.6	0	76.7
i-antigen		68.6	90	48.3	82.2
I antigen		0	70.7	0	0
Blood group O/H		0	84.9	55.2	85.4
Blood group A		0	0	0	0
Blood group B		0	0	0	0
Sd^a/Cad		49	0	13.8	0

Le^A		**0**	**0**	**0**	**0**
Le^B		0	0	0	0
Le^X		19.6	53.2	6.9	35.1
Le^Y		0	15	6.9	18.9
sLe^X		38.4	0	0	0
LacdiNAc		20.4	0	6.9	20

Most O-glycosylation enzymes belong to one of the five groups: *N*-acetylgalactosaminyltransferases (GalNAcTs), *N*-acetylglucosaminlyltransferases (GlcNAcTs), galactosyltransferases (GalTs), sialyltransferases (SiaTs), and fucosyltransferases (FucTs). Each enzyme group can be further divided into several subgroups with distinct substrate specificities. For example, the polypeptide α-GalNAcT enzyme family catalyzes the addition of GalNAc to a Ser/Thr residue to initiate O-glycan synthesis on the protein backbone, while other GalNAcTs transfer GalNAc to GlcNAc or Gal to extend core structures. Similarly, some GlcNAcTs add GlcNAc to the GalNAc residue of Tn antigen to form core structures while others add GlcNAc to Gal, forming I or i antigen. GalTs add Gal to GalNAc or GlcNAc to form core 1 structure or type 1 and type 2 chains. SiaTs add *N*-Acetylneuraminic acid (Neu5Ac) to GalNAc and Gal to form α6- and α3-sialyl antigens. FucTs add fucose (Fuc) to Gal or GlcNAc to form various blood group

antigens. The wide array of O-glycosylation enzymes and their manifold substrate specificities make O-glycosylation a highly diversified and complex biochemical network.

The expression pattern of O-glycosylation enzymes is cell and tissue-specific, making the O-glycosylation network also cell and tissue-specific. Computational tools prove useful to generate tissue- or cell-specific networks. A MATLAB-based framework was developed to construct the O-glycosylation network from five different glycosyltransferase activities using a machine-readable definition for enzyme class with functional groups, linkages and substrate specificity [255, 256]. The resulting network was used to predict possible reaction paths leading to O-glycans on P-selectin glycoprotein ligand-1 (PSGL-1) of a human promyeloid cell line. A glycosyltransferase reaction library was constructed based on the substrate specificity reported previously in the literature [257]. Using this reaction library, a repertoire of possible glycan structures was predicted from the set of glycosyltransferases expressed in a human carcinoma cell line. More recently, [258] introduced a pattern-matching algorithm to generate the O-glycosylation networks based on 25 common glycosyl- and sulfotransferase activities. Through *in-silico* enzyme knockouts, the authors demonstrated the roles of each enzyme in the O-glycoform micro-heterogeneity.

Here, we report a framework for the generation and analysis of O-glycosylation network using RING (Rule Input Network Generator) and O-GlycoVis, a newly developed visualization program specifically for O-glycosylation. RING was previously used for studies of chemical [259, 260] and biochemical [261] systems. A unique feature of RING is its English-like reaction language that describes substrate specificity and additional constraints on the formation of glycan products. Based on this user-input information, RING can predict all possible O-glycosylation reactions and glycan structures generated thereof. O-GlycoVis then uses these inputs to display the network topology and overlaid it with glycan structure information. O-GlycoVis can also predict possible glycan structures in the network from an input monosaccharide composition, identify the reaction paths

leading to those glycans, and then map the relative abundance of glycans in an input profile to the reaction paths. O-GlycoVis is a Windows-based program, written in MATLAB® and freely available upon request.

9.2. Materials and methods

The program consists of three main modules: reaction network generator (RING), glycan structure builder, and network visualizer as described below (Figure 9.1).

9.2.1. Rule input network generator (RING)

For the generation of the O-glycosylation network, the inputs to the network generator contain information on (1) initial reactants, (2) substrate specificity of all enzymes (reaction rules), (3) global constraints on glycans being formed, and (4) other post-processing instructions. These inputs are written in an English-like reaction language. Based on these inputs, the network generator then iteratively applies all the rules to the initial reactants and the products generated thereof. The output from the network generator is a list of all possible glycans and reactions consistent with the rules supplied by the user.

The initial reactants in the O-glycosylation model include a Ser/Thr residue on a polypeptide backbone and five nucleotide sugars (UDP-GalNAc, UDP-GlcNAc, UDP-Gal, CMP-Neu5Ac, and GDP-Fuc). The five nucleotide sugars provide the monosaccharides to be added to the extending O-glycans. An extending glycan on the protein may have more than one terminal monosaccharide, each of which may have more than one carbon available, that can receive a nucleotide sugar and form a glycosidic bond. Furthermore, multiple nucleotide sugars may be capable of forming a glycosidic bond with the same hydroxyl carbon of the recipient monosaccharide. For example, UDP-Gal can be transferred to c3 or c4 of the GlcNAc residue on extending glycans to form β1-3 or β1-4 bond, respectively. The carbon c3 of GlcNAc can receive GDP-Fuc, UDP-GalNAc, UDP-Gal to form α1-3 or β1-3 bond. To deal with this complexity, each nucleotide sugar was represented

as the sugar designator, plus designators for the type(s) of glycosidic bond that its carbonyl carbon can form as well as the carbons that are available for further glycan extension (Figure 9.2A).

A reaction rule defines the structure requirements for two substrates, the receiving glycan to be extended and the incoming nucleotide sugar. The rule also specifies the carbon position of the glycosidic bond formed. The model consists of 32 reaction rules that describe the substrate specificity of common mammalian O-glycosyltransferases. Sulfotransferases were not considered in this work. The rules were derived from previous studies reported in the literature. Isozymes with similar substrate specificities were considered as one enzyme. For example, ST3GAL1, 2 were lumped because they both prefer the Gal residue of Galβ1-3GalNAc. Similarly, ST3GAL4, 6 preferentially act on the Gal residue of type 2 chains, were also combined. The effects of peptide sequence and charge on enzyme activity were neglected. Therefore, all the ppGalNAcTs were considered as having the same enzyme activity [262]. Figure 9.2B-D illustrates the implementation of one rule in the O-glycosylation model, whereby the structural requirements for two substrates are defined. One substrate is a glycan to be extended and the other is a nucleotide sugar.

Global constraints were implemented to restrict the formation of glycans to be within certain size or pattern. Most of these constraints prevent glycans with repetitive units to populate the network since they do not provide additional structural information. Five global constraints were imposed in the model: (1) the maximal length of LacNAc chains on any glycan is two repeats; (2) a hybrid structure between type 1 and type 2 chains cannot be formed; (3) two consecutive I-antigens are not allowed in the system; (4) each monosaccharide can be linked to at most three other monosaccharides; (5) any carbon of a monosaccharide is connected with only one other monosaccharide. In this work the number of repeat of LacNAc chains on any glycan is restricted to two. This significantly reduces the computational time required to generate networks. The

constraint can be relaxed if so desired, however the computational need for network generation will increase and limited additional information on the network may be gained from the relaxation.

Other post-processing instructions: Using the input, the program generates the reaction network and a list of all possible glycans. Each glycan is represented in a way analogous to SMILES [263], assigned a unique identification number (ID) and classified based on the borne epitopes (or structural patterns) defined in the post-processing instructions. All the glycan strings and their IDs are used as inputs to the glycan structure builder module. The relationship between glycans is stored in a matrix. Each row of the matrix corresponds to each reaction that occurred in the O-glycosylation network. Four columns of the matrix provide the IDs of the substrate and product glycans, the names and indices of rules involved in that reactions. The relationship matrix is an input of the network visualizer, aiding synthesis pathway identification.

9.2.2. Glycan structure builder

The format of output glycan strings from RING is not commonly used for glycan representation. To facilitate glycan visualization the program converts the output strings from RING to the standard IUPAC-condensed nomenclature (Figure 9.3). The DrawGlycan-SNFG software [264] then converts all the output strings to graphical representations of glycans and stores them as image files. Each image file is automatically named according to the ID of the glycan. ImageMagick then post-processes these image files for better visualization (e.g. crop and convert to background-transparent images).

9.2.3. Network visualizer (O-GlycoVis)

O-GlycoVis is a MATLAB based program developed for visualization of the O-glycosylation network. Using the glycan classification results as the input, O-GlycoVis interfaces with GraphViz to display the distribution of glycan epitopes in the O-glycosylation network. In the network display, each node represents a glycan and each edge depicts a reaction. Node colors (except for

white) represent the epitopes that a glycan bears. If a glycan carries multiple epitopes, the node will be segmented into parts, each of which will be colored accordingly. Edges are colored by enzyme activities unless otherwise specified.

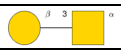
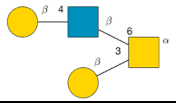
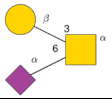
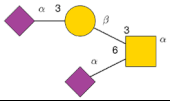
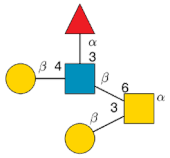
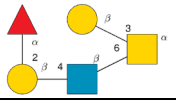
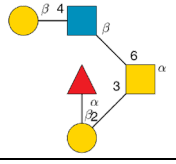
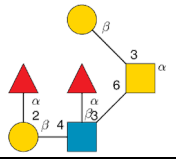
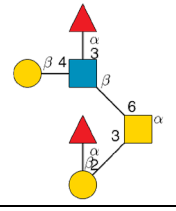
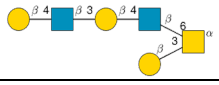
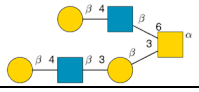
O-GlycoVis also predicts the synthesis pathway of input glycans. In this application, the module first retrieves the user-input information on the glycans and their abundances. Based on the monosaccharide composition of input glycans, the program identifies their IDs. An algorithm is then used to identify all the immediate reactants leading to the input glycans (Figure 9.4). This process is repeated until the algorithm reaches the initial reactant glycan of the network. All the input glycans and their reactants are stored in a matrix whose two columns are the IDs of the substrate and product glycans. Each row of the matrix is a reaction step leading to the input glycan profiles. The matrix is used to create a pathway map, in which edges are labeled by the responsible reaction rules and nodes are colored by glycan abundances (percentages of the total glycan input).

9.3. Results

9.3.1. O-Glycan Distribution in Breast Cancer Cells (T47D and MCF7)

Using the program, the O-glycosylation networks of two breast cancer cell lines T47D and MCF7 were predicted based on the glycosylation enzymes expression profile as revealed by the RNA-Seq data available from the Human Protein Atlas version 18 [265]. The predicted network for T47D only consists of 29 glycans and 30 reactions. On the other hand, that for MCF7 has 858 glycans and 1906 reactions. The structure of the epitopes of the O-glycans predicted in the two networks and the percent of glycans bearing each epitope are shown in Table A.1. Note that many glycans bear more than one epitope. The vast difference in the number of glycans generated between the two cell lines is also reflected in their epitope distribution.

Table 9.2. Glycan profiles of secretory MUC1 reported for T47D and MCF7 cell lines

Glycan ID		Structure	Relative amount (%)	
[266]	Current work		T47D	MCF 7
1	6		6.9	12.9
2	340		0	54.3
4	790		45.9	2.3
5	88		24.6	2.9
8	807		22.7	0
9	528		0	7.9
10	101		0	4.6
10’	342		0	4.6
11	34		0	1.1
11’	534		0	1.1
12	177		0	2.5
12’	339		0	2.5

The glycan profiles predicted for these two cell lines were compared to the literature reported glycans (Table A.2). The program predicted all the reported structures (4 and 11 for T47D and MCF7, respectively) and identified all pathways leading to such glycans. Figure 9.5A shows the enormous network predicted for MCF7, highlighting only the route leading to the reported glycans in T47D and MCF7 cells. Figure 9.5B and C show a simplified version of the output from O-Glycovis for T47D and MCF7 cells, respectively. The abundance level of the glycans is indicated by colors. The difference between the two networks is mainly due to the presence of core 2 enzymes, GCNT1 and GCNT3, in MCF7 cells line. Without those two enzymes the size of the MCF7 network is significantly reduced to somewhat similar to that of the T47D (i.e. 25 glycans and 26 reactions).

9.3.2. The representation of other epitopes in O-glycosylation network

The epitope(s) borne in each glycan was identified using the program and the glycans predicted by the program was classified by epitopes (Table A.1). The glycans generated for MCF7. encompasses eight common epitopes: (1) (extended) core 1, (2) (extended) core 2, (3) type 2 LacNAc, (4) LacdiNAc structure or one of the following antigens: (5) blood group O/H, (66) Lewis X (Le^X), (7) 7) Lewis Y (Le^Y) and (88) i antigen. Core 3 and core 4 structures are not present in the network because MCF7 cells have only very low expression of the β3-GnT6 enzyme. Similarly, I antigen was not formed because IGnT enzymes (GCNT2 and GCNT3) are not expressed. Lewis Y (Le^Y), blood groups AB, Le^A, Le^B, and SLe^A were also absent because α2-FucTs (FUT1 and FUT2) and α4-FucT (FUT3) are not expressed.

The number of glycans with i antigen far exceeds that with LacdiNAc structure (Table A.1). The i antigen is formed by the addition of a β1,4Gal to GlcNAc residues of glycans through B4GALT enzymes catalyzed reactions. LacdiNAc structure, however, is formed by the addition of a β1,4GalNAc to GlcNAc residues, catalyzed by B4GALNACT3,4. Interestingly, although

B4GALNACT3,4 has a broader substrate specificity than B4GALT, the number of glycans bearing LacdiNAc is far less than those having i antigen. After GalNAc is added to the GlcNAc residue to form LacdiNAc, the chain growth is terminated. However, once the i antigen is formed, its nascent Gal can be further extended with GlcNAc, Neu5Ac or GalNAc by B3GNT, ST3GAL, and B4GALNACT2, respectively. This increases the number of glycans having i antigen.

The program encapsulates substrate specificity as illustrated by the generation of glycans bearing Le^X antigen but not SLe^X because of the substrate specificity of FUT4 and FUT7 (Table A.1). FUT4 preferentially fucosylates the inner GlcNAc of type 2 chains to form Le^X [267, 268]. The enzyme prefers glycans having at least two consecutive LacNAc motifs (Galβ1-4GlcNAcβ1-). Because the imposed global constraints limits the length of LacNAc chains to two repeats (see Materials and Methods), less than 20% of glycans in the network are the preferred glycan substrate of FUT4. Because of its low transcriptional abundance in MCF7, FUT7 was excluded from the network. FUT7 accepts substrates with at least one LacNAc to fucosylate the distal GlcNAc of sialylated glycans form SLe^X [267, 269]. If FUT7 is present the generated glycans will have a much wider diversity.

9.3.3. O-glycan distribution in human umbilical vein endothelial cells (HUVEC)

The O-glycosylation network for HUVEC was constructed based on the gene expression (RNAseq) data of glycosylation enzymes in the HUVEC TERT2 cells available from the Human Protein Atlas version 18 [265]. The program generated a network of 515 glycans in 1210 reactions. Glycans in the network cover eight epitopes: (1) core 1 and (2) core 2 structures, (3) type 2 chain and other antigens including (4) Le^X, (5) Le^Y, (6) i, (7) I, and (8) blood group O/H. Similar to the O-glycosylation network of MCF7 cells, this network did not contain core 3 or 4 structure, blood group A or B, Le^A, and Le^B antigens because the responsible enzymes were not expressed. In addition, type 1 chain, blood group Sd^a/Cad, sLe^X, extended core 1 and LacdiNAc structures were

not present because B3GALT5, B4GALNT2, FUT7, B3GNT3 and B4GALNT3,4 enzymes were not expressed. The abundance levels of predicted glycans harboring various epitopes are compared to networks predicted for other cells (Table A.1). Similar to CHO cells, core 2 glycans are much more plentiful than core 1.

We compared the generated network with the reported O-glycan profile of HUVEC [270]. Again, the program predicted far more glycans than detected. The program predicted 38 out of the 50 reported glycan structures (including isomers). O-Glycovis is programmed to identify the possible reaction paths from the starting glycan to marked target glycans. Using O-Glycovis, the possible reaction paths leading to the literature detected glycans along the generated a network are shown in Figure 9.6A. Nodes along the reaction paths are colored yellow, and the edges are colored by the enzyme catalyzing the reaction step.

The program also generates an interactive graph that allows zooming in on a particular region of the network and visualizing the reactions that traverse a particular node with the structures of substrate and product glycans being displayed (Figure 9.6B,C).

Among the twelve glycans that were not predicted by O-Glycovis, nine carries poly-LacNAc chains with three repeats of Galβ1-4GlcNAc, that exceeded the maximal length of two repeats imposed in the global constraints. These constraints were not relaxed for this network, due to the considerable increase in network size resulting from longer chains and corresponding computational expense. Glycan #3 and #5 are formed by the transfer of Neu5Ac to GalNAc of asialylated core 1 structures, catalyzed by ST6GALNAC1,2 [271]. These two enzymes were not expressed in HUVEC TERT2 cells whose transcript profile was used to construct the network. All the reactions leading to glycan #30 were represented as the enzyme input into RING, it is formed by sequential sialylation and fucosylation of the β1-6 arm of an extended core 2 structure. Glycan #30 is absent in the predicted network because of the specification of reaction rules. Two synthesis

routes lead to its formation with the reversed order of the two enzymes: (1) FUT4 → -ST3GAL4, and (2) ST3GAL4 → -FUT4. In the input rules, the route 1, sialylation of the terminal Gal by ST3GAL4, was blocked by prior fucosylation of the sub-terminal GlcNAc by FUT4 [265]. Another rule blocks route 2 because FUT4 activity toward sialyl-LacNAc chain was reported to be very weak [272]. By relaxing those rules, glycan #30 can be generated.

9.3.4. O-glycan distribution in Chinese Hamster Ovary (CHO) Cells

The O-glycosyltransferase enzymes expressed in CHO cells were compiled using the RNA-Seq data of CHO-DG44 and CHO-K1 lines. The enzymes that are expressed at the transcript level in both cell lines were used to construct the O-glycosylation network. The network generated consists of 5 glycan species and 4 reactions, and is shown in Figure 9.7A. Only glycans of the (extended) core 1 epitope and the Tn- antigen are present in the network. The limited diversity in the CHO O-glycan network reflects the lack of expression of key O-glycosylation enzymes, especially GCNT1, which prevents the formation of core 2 structures. The expression of GCNT1 alone could more than triple the number of possible O-glycan species, extending the network to include core 2 and type 2 structures, as well as i antigen, in addition to the original core 1 species.

We compared the O-glycosylation network generated for CHO cells to the reported O-glycans on the recombinant MUC1(1.7TR)-IgG2a fusion protein [273] produced in CHO cells. Only core 1, mono- and di-sialyl core 1 glycans were seen in the reported profile, which is consistent with the glycan profile generated for CHO and its relatively limited expression of O-glycosylation enzymes. All reported detected glycans were predicted in the network generated by the program.

9.3.5. Visualization of O-glycan epitopes with O-glycovis

The O-glycosylation enzymes gives a relatively simple glycan network confining to core 1 structure. By expressing even a few more additional enzymes, the network is dramatically expanded. We used such an expanded network to illustrate another feature of the program that

highlights the glycans with common epitopes (Figure 9.7B-G). A very large fraction of glycans have more than one epitope. These glycans are represented as nodes with multiple colors, as seen in the terminal glycan g_T (the circled nodes in Figure 9.7B-G). For reference Core 1 glycans constitute about 5% of the total number of species generated. Vastly wider diversity is seen in core 2 glycans (Figure 9.7B). Far more glycans have type 2 chain than type 1 chain (Figure 9.7B). Fewer glycans have type 1 chains reflecting the different substrate specificity of β3-GalT5 and β4-GalT. β3-GalT5 has a narrow substrate specificity acting only on the GlcNAc residues of core 2 (GlcNAcβ1-6GalNAc) and core 3 (GlcNAcβ1-3GalNAc) structures to form type 1 chains [274]. In comparison, β4-GalT, catalyzes the formation of Galβ1-4GlcNAc of type 2 chains, has a broader substrate specificity. It accepts the GlcNAc residues of core 2/3 structures and β1,3GlcNAc residues of type 2 chains. Similar to the MCF7 network, the number of glycans with i antigen exceeds that with LacdiNAc structure (Figure 9.7D). The substrate specificity of FUT4 and FUT7 lead to a fewer number glycans with the Le^X antigen than SLe^X (Figure 9.7E).

The number of glycans bearing Sd^a/Cad antigen is equivalent to that having type 2 chain and far exceeds that with type 1 chain (Figure 9.7F-G). Interestingly, about half of the Sd^a/Cad carrying glycans also have type 2 chain and vice versa. On the other hand, less than 5% of glycans having Sd^a/Cad carry type 1 chain, indicated by dual-color nodes in Figure 9.5F. It should be noted that all the glycans having both Sd^a/Cad and type 1 or type 2 chain are core 2 glycans with two arms (β1-6 and β1-3) on its structure. Sd^a/Cad is usually formed on one arm while type 1 or type 2 chain is on the other arm.

9.4. Discussion

In this study, we used RING to generate O-glycosylation reaction networks and integrate the output into O-GlycoVis for visualization and analysis. The user inputs into RING are rules on the substrate specificity and the reaction of the glycosylation enzymes involved. We used transcriptome

data to decide on the enzymes to be included for each cell type evaluated. This is easily modified by users for any new cell line. RING predicts possible glycan structures generated in the O-glycosylation networks. Using the output from RING, O-GlycoVis generates network graphs that allow zooming into particular regions or nodes and displaying reactions that traverse one node with the structures of substrate and product glycans being shown. From an input monosaccharide composition, O-GlycoVis predicts the associated glycan structures and identifies the synthetic routes leading to each glycan. The abundance level of each glycan from an input profile can be overlaid onto the pathway map, assisting quantitative analysis of the glycan profile.

The program bears some degree of similarity to GlycoVis, the visualization tool for N-glycosylation network [275]. However, there is a major difference between the two programs. N-glycosylation of recombinant therapeutic proteins follows a relatively fixed pathway of fixed number of enzymes and reactions. Although some enzymes are represented by many isozymes with different substrate specificity, the basic network structure is not affected by those isoforms. In O-glycosylation, the number of enzymes is very large and the network generated in different cells or tissues can be vastly different as seen in the examples presented above. Hence, O-GlycoVis is integrated with RING to facilitate O-glycosylation network generation and visualization in a tissue or cell-specific fashion.

A number of computational tools have been developed for the generation of glycosylation networks. A reaction pattern library was developed, which specifies the acceptor and donor monosaccharides and the linkage formed between them for a number of glycosyltransferases [257]. GNAT uses class-based inheritance to describe the enzyme substrate specificity [255]. Specifically, structural requirements on the substrate and product glycans of each glycosylation reaction are defined in different fields of an enzyme class or sub-class. O-Glycologue utilizes a set of pattern-matching rules to describe enzyme substrate specificity [258]. The program generates O-

glycosylation networks by iterative application of such rules upon glycan strings. Glycan strings are represented using a one-letter code for monosaccharide and linkage. O-Glycologue is usable via a web tool where users may knock out reactions and gain information about structures. A similar approach was used to generate N-glycosylation networks, in which glycan structures are represented using modified LinearCode nomenclature [276]. RING can perform network generation with a similar capability to the aforementioned tools. A key difference in RING is the use of the English-like reaction language combined with rule based sugar addition. This language has simple descriptions of glycan structures and substrate specificity of glycosylation enzymes, and was written so that rules can be easily added or modified to generate new networks, and does not rely on a fixed set of enzymes or constraints. Like other network generating tools, the network generation step of RING may require a long computational time for a very large network that has a large number of global constraints imposed.

Instead of generating a comprehensive network for a wide range of tissues and species, this work illustrate the use of the program for generating O-glycosylation network generation for four cell types. A number of reactions, including terminal addition of sulfate and polysialic acid formation, were not included in this demonstration. Furthermore, the glycan network generation is influenced by the nucleotide sugar substrates considered. In this work, we did not consider uncommon substrates such as Neu5Gc. Neu5Gc can be take the place of Neu5Ac in glycans synthesized in non-human mammalian species. Importantly, we have provided a framework with O-Glycovis which can be built upon to tailor the network generation for new applications or more complete networks.

It is worth noting that the program and other network generation programs predict the intracellular biosynthesis network. Whereas the experimental glycan measurement only detects those on the secreted proteins or cell surface proteins. The generated networks certainly display

more glycan species than the experimental measurement. This program allows one to visualize the observed glycan species in the generated network and trace the possible reaction paths leading to the observed glycans. It can help to identify possible bottlenecks or constriction points caused by protein structural in biosynthetic pathways.

O-glycans are implicated in a variety of biological processes and human diseases. The structure of O-glycans affects the quality of therapeutic proteins. Better O-glycan characterization and a better control of O-glycosylation in protein and cell production will help generate high-quality products. The platform reported here can help predict glycans generated in any particular cell line or tissue to facilitate the identification of glycans. Its versatility in projecting the network and tracing the reaction path will facilitate the development of glycoengineering strategies.

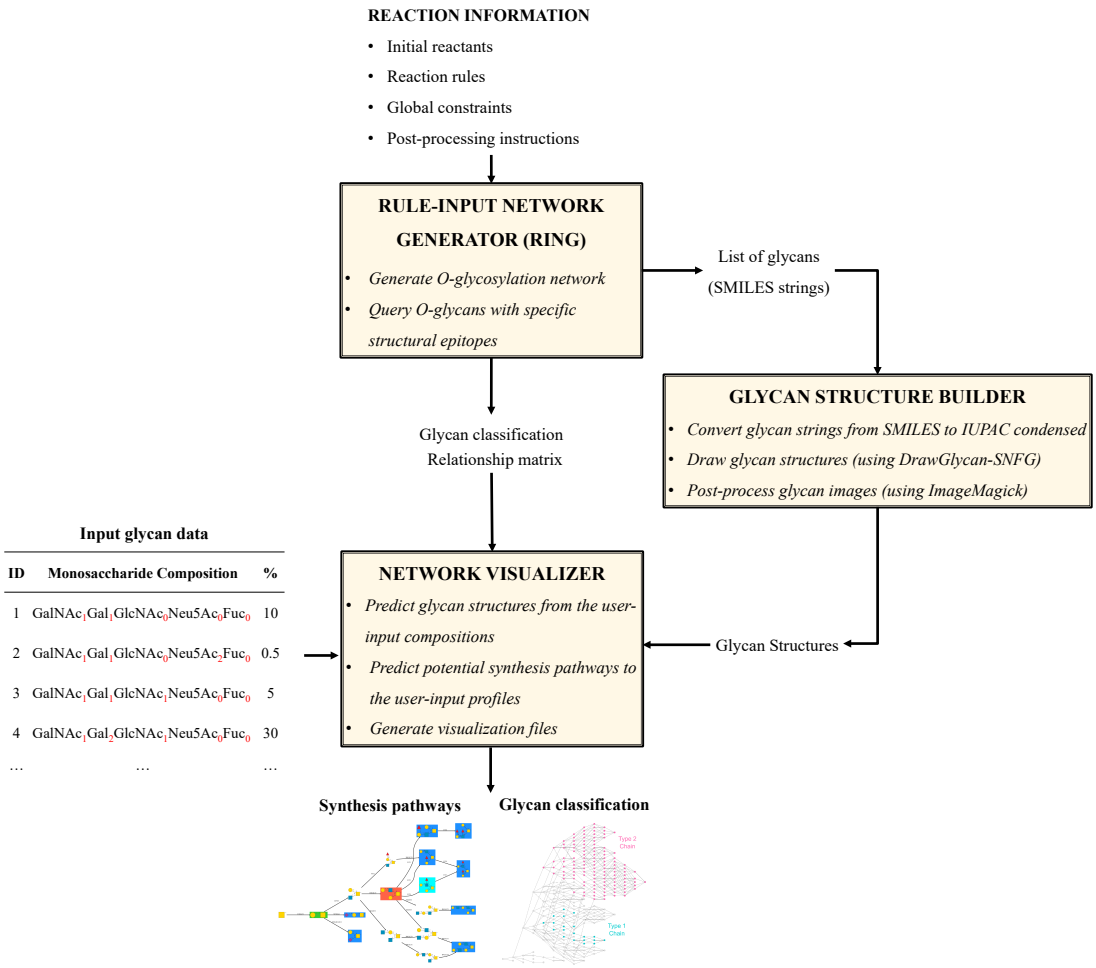


Figure 9.1: Input and output schematic for the platform

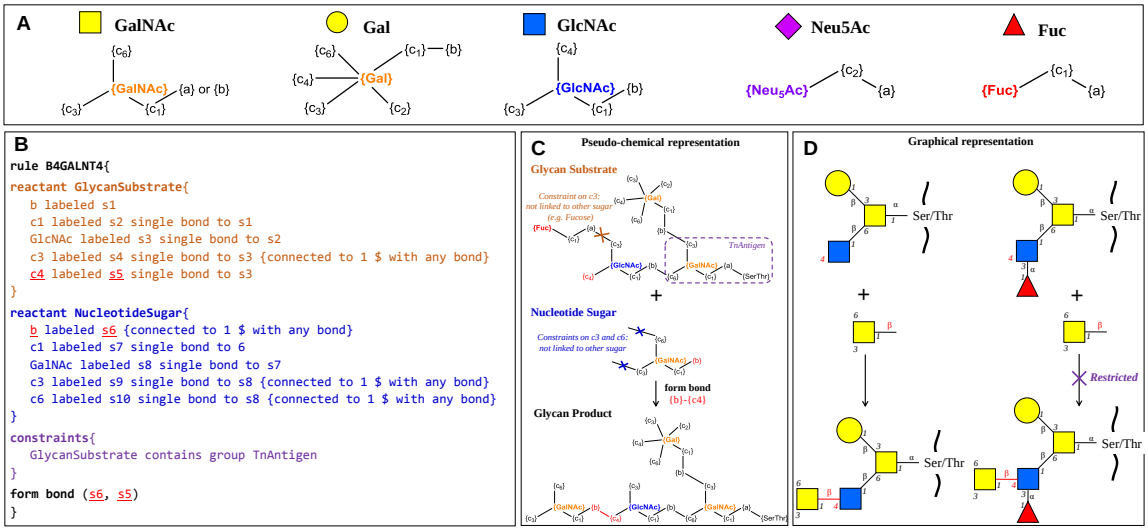


Figure 9.2: Overview of visualization

(A) Graphical and pseudo-chemical representations of nucleotide sugars. (B) An example of rule implementation in the O-glycosylation model. The rule describes structural requirements for the glycan to be extended (GlycanSubstrate) and the incoming nucleotide sugar (NucleotideSugar) in a reaction catalyzed by β1,4 N-acetylgalactosaminyltransferase 4 (B4GALNT4) enzyme. The Glycan Substrate must contain a terminal GlcNAc and the Tn antigen (GalNAcα-Ser/Thr) substructure. The Nucleotide Sugar must be UDP-GalNAc, shortened as GalNAc with an overhanging β-glycosidic "bond" connected to the carbonyl carbon c1 of GalNAc. The hydroxyl carbons c3 and c6 of GalNAc must not be linked with any other sugar. If all the requirements are satisfied, a β-glycosidic bond will be formed between s5 (hydroxyl carbon c4) of the Glycan Substrate and s6 (the overhanging β-glycosidic "bond") of the Nucleotide Sugar as stated in the "form bond" module. The product glycan will contain the GalNAcβ1-4GlcNAc substructure. (C) and (D) Pseudo-chemical and graphical representations of a reaction following the enzymatic rule defined in 2A. The pseudo-chemical representation was generated using output from RING.

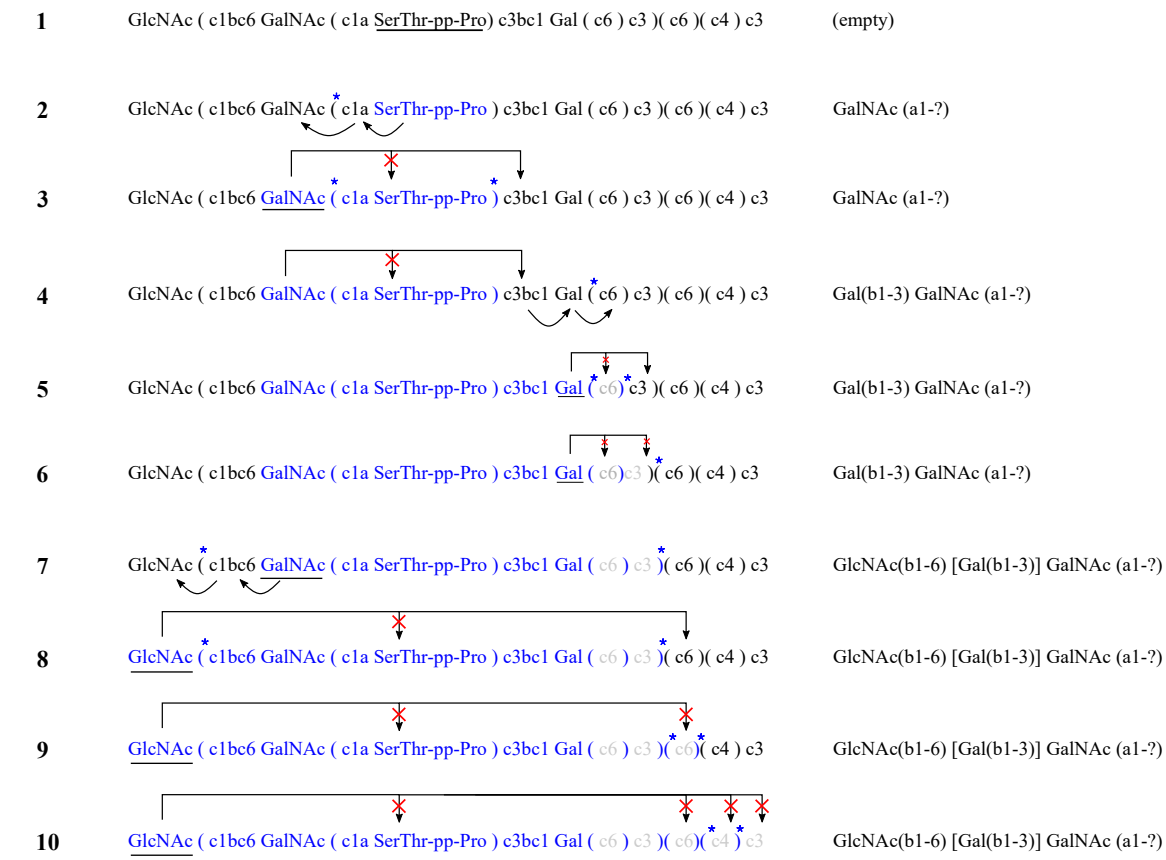


Figure 9.3: Overview of renaming algorithm for glycan string processing

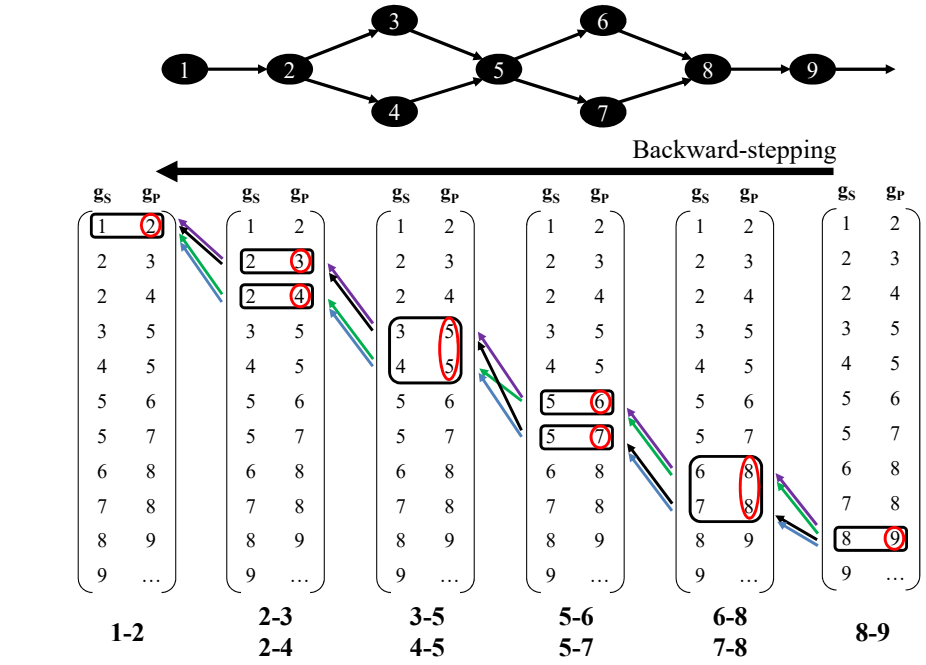


Figure 9.4: Reaction pathway tracing algorithm.

Backward-stepping algorithm used in the program to identify all potential reaction pathways to synthesize an O-glycan (i.e. glycan #9). gs and gp refer to substrate and product glycans of each glycosylation reaction. Each color represents one possible reaction path.

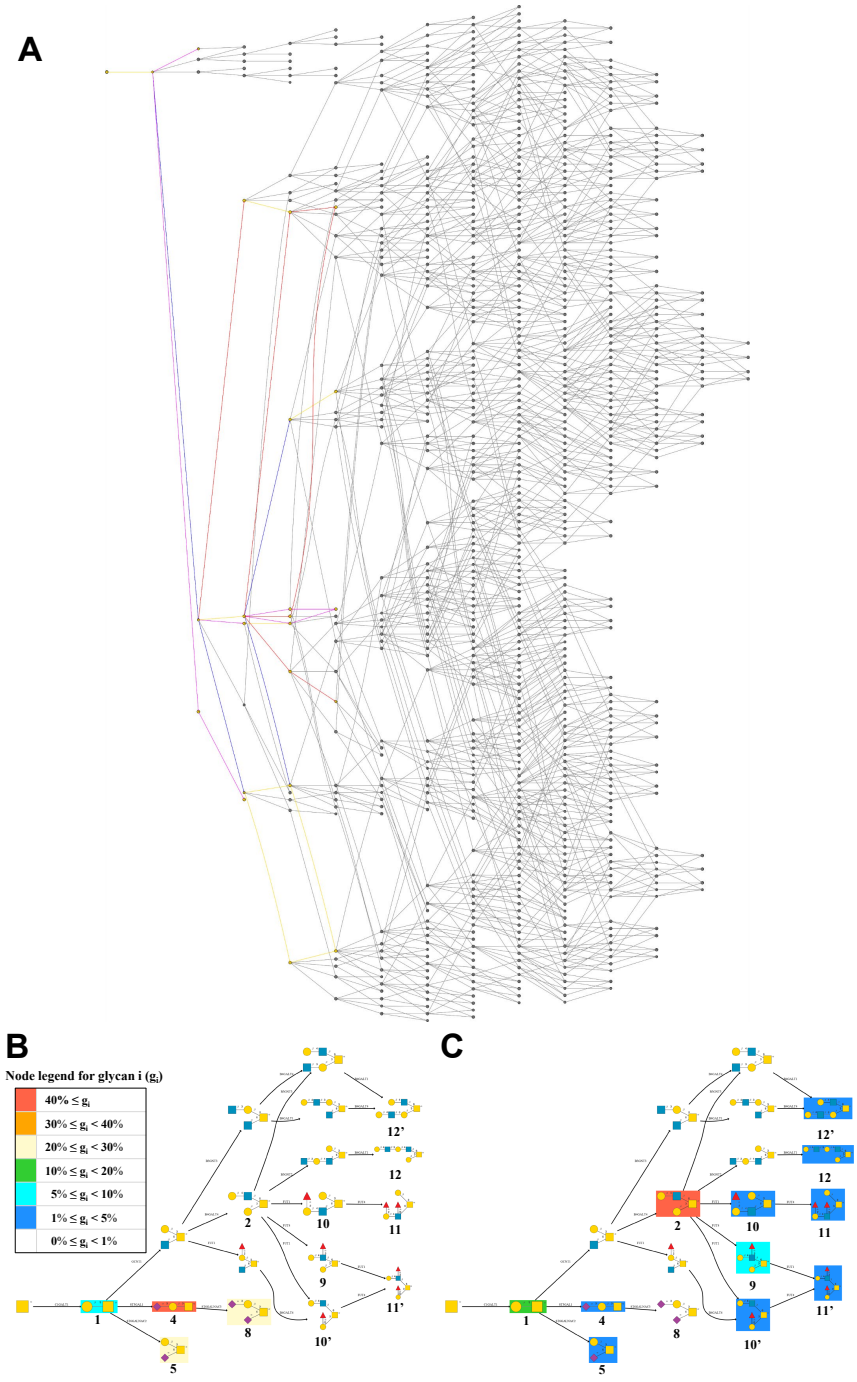


Figure 9.5: O-glycosylation in T47D and MCF7.

The program output using O-glycan profiling data from two breast cancer cell lines T47D and MCF7. (A) The O-glycosylation network predicted for MCF7 cells. Pathways proceed from left to

right. The reaction paths leading to the reported glycans in T47D and MCF7 cells are highlighted. Nodes along the reaction paths are colored yellow, and the edges are colored by the enzyme catalyzing the reaction step. (B) & (C) Pathway maps of cellular O-glycan biosynthesis in T47D (B) and MCF7 cells (C), respectively. Glycan structures are highlighted according to their abundance levels in the glycan profile. The numbers below glycan structures are their corresponding IDs listed in the first column of Table 2.

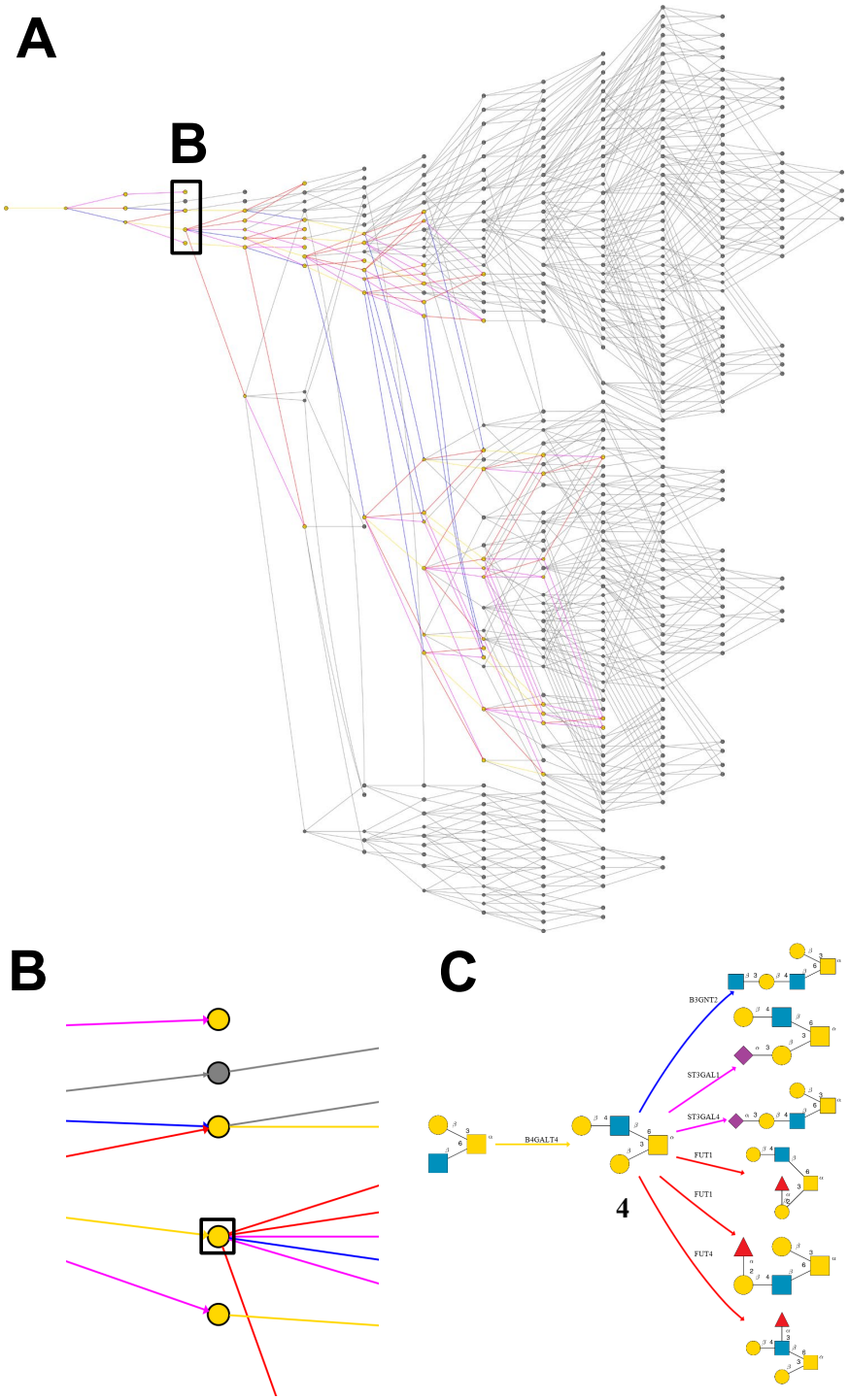


Figure 9.6: HUVEC O-glycan profile

The program output using the HUVEC O-glycome profiling data. (A) The O-glycosylation network predicted for HUVEC cells. Pathways proceed from left to right. The reaction paths leading to the reported glycans are highlighted. Nodes along the reaction paths are colored yellow, and the edges are colored by the enzyme catalyzing the reaction step. (i.e. Gal-transferase: yellow; GlcNAc-transferase: blue; Fuc-transferase: red; Neu5Ac-transferase: purple). (B) Magnification view of the rectangular region B in A. (C) Magnification view of the rectangular region in B.

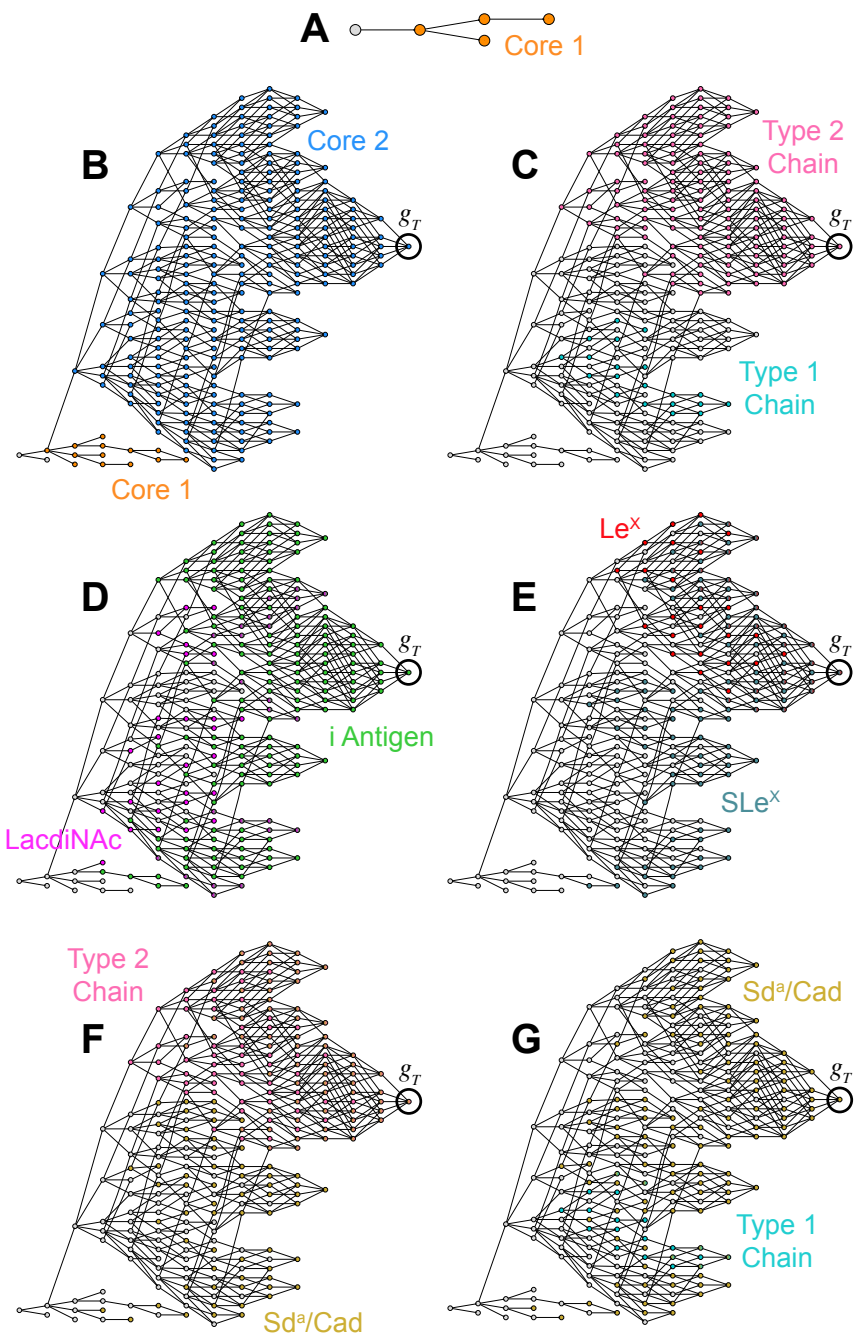


Figure 9.7: Epitope labeling on O-glycan networks.

(A) O-glycosylation network for CHO cells. (B)-(G) O-glycosylation network for an engineered cell line. Highlighted nodes represent glycans bearing (B) core 1 (orange) and core 2 (blue) struc-

tures, (C) type 1 (teal) and type 2 (pink) chains, (D) LacdiNAc (purple) and i antigen (green), (E) Lewis X (red) and sialyl-Lewis X (aqua), (F) type 2 chain and Sda/Cad antigen (yellow), (G) type 1 chain and Sda/Cad antigen. The circled node represents a terminal glycan gT in the network. This node has multiple colors, indicating that the corresponding glycan carries more than one epitope. The interactive network graph allows zooming in on the circled node to display its glycan structure (as shown in Figure S2).