SYSTEMATIC APPROACH TO CHARACTERIZE AND DEVELOP A CELL FREE GLYCOPROTEIN REMODELING PLATFORM

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Yi- Wen Liu

December 2018

ABSTRACT

Glycosylation has profound impacts in many biological processes, such as immune response, inflammation, cell- cell communication, and host- pathogen interaction. The importance of glycosylation is further accentuated as more than 70% of approved protein-based drugs are asparagine-linked (N-linked) glycoproteins. In many cases, the efficacy, safety, and stability of glycoprotein drugs are dictated by its glycan structure. Despite its importance, advancement in glycoscience and glycoengineering is hindered, largely due to insufficient understanding of glycan biosynthesis and bioprocess, as well as, a lack of platform for producing homogeneous designed glycan structures on proteins. Current glycoprotein expression platforms, such as using Chinese hamster ovary (CHO) cell line, cannot avoid heterogeneous glycoform with high batch to batch variability due to their endogenous glycosylation pathways that are sensitive to a change in culture environment. Moreover, the complexity of metabolic network inside cells has made it difficult to understand and engineer specific glycosylation pathways to produce designer glycoproteins. To address those challenges systematically, we proposed to integrate experimental and mathematical methods, such as *E. coli*- based cell free glycoprotein synthesis (CFGpS) and constrained based flux balance analysis (FBA) to characterize and optimize glycoprotein synthesis. This platform will in the future allow us to robustly and systematically engineer novel glycosylation pathways. To approach that, we first developed a platform for enzymatically remodeling glycan structures on therapeutic relevant N-glycoproteins. This technology allowed us to have direct control of the glycoform and to produce homogeneous glycoproteins for structure- function relationship study. As a whole, we anticipated our mathematical and experimental approach could facilitate the fundamental understanding in glycoscience and could provide a novel approach for producing glycoprotein therapeutics.

BIOGRAPHICAL SKETCH

Yi- Wen Liu has received her B.S degree in chemical engineering at the University of Rochester in 2016. Her work with Dr. Rudi Fasan on designing a macrocyclic peptide inhibitor of the Sonic Hedgehog/Patched interaction at University of Rochester was published in JACS in 2017. Following her B.S degree, she studied her M.S degree in chemical engineering under Dr. Matthew DeLisa and Dr. Jeffrey Varner at Cornell University in 2018. Her projects focused on developing an *in vitro* platform for producing design glycoproteins and studying glycoprotein synthesis in *E. coli* based cell free glycoprotein synthesis using sequence specific flux balance analysis.

ACKNOWLEDGMENTS

This work was dedicated to Dr. Matthew DeLisa and Dr. Jeffrey Varner's research groups, my friends and my parents. This work will not be accomplished without the supports from Dr. DeLisa and Dr. Varner. Specially, Thapakorn Jaroentomeechai and Michael Vilkhovoy had been very helpful mentors throughout my time at Cornell.

TABLE OF CONTENTS

Biography Sketch 3
Acknowledgements 4
Table of Contents 5-6
List of Figures 7
List of Abbreviation 8
Preface 9
Reference 53-56

Chapter 1. Introduction to glycosylation and flux balance analysis (FBA) 10-18
- 1.1 Introduction
- 1.2 Protein glycosylation
- 1.3 Bacterial glycoengineering
- 1.4 Glycan remodeling technology
- 1.5 Cell free protein synthesis (CFPS)
- 1.6 Constrained based flux balance analysis (FBA)

Chapter 2. The development of *in vitro* glycoprotein remodeling system 19-36
- 2.1 Introduction
  - *2.1.1 Production of linear polysialic acid- containing glycoprotein*
  - *2.1.2 Production of biantennary glycoprotein*
- 2.2 Results for bottom-up design of *in vitro* glycan remodeling platform
  - *2.2.1 Generation of N-linked glycoprotein precursor*
  - *2,2,2 Engineering and expressing glycosyltransferase- GntI, GntII, CstII, and PST^{NM}*
  - *2.2.3 Activity of N-acetylglucosaminyltransferases- GntI and GntII*
  - *2.2.4 Activity of α-2,3-sialyltransferase- CSTII*
- 2.3 Discussion and future direction
- 2.4 Methods
  - *2.4.1 Strains and plasmids*
  - *2.4.2 Expression and purification of model glycoprotein precursors*
  - *2.4.3 Expression and purification of GnTs, CSTII, and PST^{NM}*
  - *2.4.4 Synthesis of GlcNAc precursor*
  - *2.4.5 Synthesis of $GlcNac_2Man_3GlcNac_2$-MBP-Glucagon*
  - *2.4.6 Synthesis of $(Neu5Ac)_2$ $GlcNac_2Man_3GlcNac_2$-A1AT and $(Neu5Ac)_n(Neu5Ac)_2$ $GlcNac_2Man_3GlcNac_2$-A1AT*
  - *2.4.7 SDS-PAGE gel and Weston Blot*
  - *2.4.8 Click Reaction*

Chapter 3. Sequence Specific Flux Balance Analysis (ssFBA) on glycoprotein synthesis using cell free protein synthesis (CFPS) in combination with *in vitro* glycoprotein synthesis 37-52
- 3.1 Introduction
- 3.2 Results- ssFBA on the synthesis of $GalNAc_4GlcNAc$-sfGFP using CFPS following by *in vitro* glycoprotein synthesis.

3.2.1 Validation of model predictions
3.2.2 Analysis of the CFPS performance
3.3 Discussion and future direction
3.4 Methods
3.4.1 Cell free protein synthesis
3.4.2 In vitro glycosylation reactions
3.4.3 Constrained based sequence specific flux balance analysis
3.4.4 Calculation of energy efficiency
3.4.5 Calculation of carbon yield
3.4.6 Calculation of maximum and minimum bound for fluxes
3.4.6 Calculation of maximum and minimum bound for fluxes
3.4.5 Quantification of amino acids

LIST OF FIGURES

Figure 1.1 A schematic for constrained based flux balance analysis (FBA). 18

Figure 2.1 Proposed pathways for *in vitro* glycan remodeling platforms. 20

Figure 2.2 A schematic for the synthesis of glycoprotein bearing single (GlcNAc) structure. 22

Figure 2.3 Glycoproteins bearing ($GalNAc_4GlcNAc$) (a,b) and ($Man_3GlcNAc_2$) (c) expression detected by immunoblots. 25

Figure 2.4 Immunoblots for verifying the activity of EndoS and α-N-acetylgalactosaminidase. 26

Figure 2.5 Immunoblots for detecting **α**-2,3 sialyltransferase, CSTII and **α**-2,8-polysialyltransferase, PST^{NM} fused to MBP against histidine tag. 27

Figure 2.6 Immunoblot for verifying the activity of MBP-GntI and MBP-GntII against histidine tag. 28

Figure 2.7 In gel fluorescence detection for the activity of GntI using $Man_3GlcNAc_2$-MBP-Glucagon. 28

Figure 2.8 In gel fluorescence detection for the activity of **α**-2,3 sialyltransferase, CSTII using human alpha-1-antitrypsin,A1AT, which has biantenarry structure attached. 29

Figure 3.1 A panel of 11 metabolites concentration over 20 minutes CFPS reactions. 42

Figure 3.2 Pie charts for carbon yields. 43

Figure 3.3 Pie charts for demonstrating the energy efficiency spent on the following metabolic reactions: glycolysis (gly), amino acids synthesis (AA), anapleoretic (ana), chorismate (cho), energy degradation (en), and protein synthesis (prot). 44

CHAPTER 1

INTRODUCTION TO GLYCOSYLATION AND FLUX BALANCE ANALSYSIS (FBA)

1.1 Introduction:

Glycosylation is one of the most common post translational modification (PTMs) on polypeptide chain in nature. The three main types of glycosylation, in which carbohydrate moieties are covalently attached to polypeptide chains are: *N*-linked glycosylation, *O*-linked glycosylation, and *C*-mannosylation glycosylation. Each type of glycosylation has glycans attach to the amide nitrogen of asparagine residues (*N*-glycosylation), to the hydroxyl oxygen of serine or threonine residues (*O*- glycosylation), and rarely to the indole C2 carbon of Tryptophan through C-C linkage (*C*-mannosylation) [1]. Among these three types of glycosylation, Asparagine- linked (*N*-linked) glycosylation is the most diverse and complex in structure and is one of the most frequent PTMs in eukaryotic cells. *N*-linked glycosylation is greatly involved in protein function, stability, folding, host-immune responses, and cell- cell and cell- virus interactions [2-4]. The importance of *N*-glycosylation is further reinforced as that more than 50% of the current protein-based therapeutics are *N*-linked glycoproteins [5]. As demands for recombinant monoclonal antibodies (rmAbs) products have increased significantly since 1986, many studies have done on studying and engineering mAbs [6-8]. An important finding is that correct glycosylation pattern can increase antibody or other glycoprotein stability and serum half- life, decrease its immunogenicity, and increase its binding affinity to antigens [9, 10]. Although many efforts have been made in the past decades to understand how glycosylation affects protein structure- function, the inability to precisely control glycan structures and the heterogeneity production have made it difficult [5].

Currently, mammalian based expression systems are preferable for glycoprotein therapeutic production in biopharmaceutical industry. Chinese Hampster Ovary (CHO) cells are by far the

most commonly used cell line [5]. It is involved in more than 70% of recombinant glycoconjugate therapeutics production, especially recombinant monoclonal antibody. CHO cells have the ability to produce properly folded protein with human- like glycoform, which is more likely to be compatible and bioactive in humans. They also have robust growth in serum-free suspension culture with substantial yield. Despite many advancements have been made in the past decades in engineering mammalian cell lines as platforms for producing glycoproteins, their endogenous glycosylation machinery makes it impossible to avoid glycoform heterogeneity leading to unwanted immune responses. This also makes the process of deciphering the effects of glycoforms on protein structure and function difficult [7, 11].

The discovery of bacterial glycosylation pathways opens up potential avenue for scientists to engineer and produce homogeneous glycoproteins on demand. In early 2000s, bacterium *N*-linked glycosylation was found in *Campylobacter jejuni* and its glycosylation gene cluster, *pgl*, was incorporated into *Escherichia coli* [2]. The ability to produce glycoproteins in *E. coli* has many advantages. Natively, *E. coli* lacks glycosylation machinery. Thus, by using *E. coli* as a host, we can avoid the possibility of introducing immunogenic epitopes produced by endogenous glycosylation pathway that mammalian cells have. Moreover, *E. coli* has a faster cell growth rate than mammalian cell lines. Since the incorporation of *C. jejuni pgl* pathway into *E. coli*, scientists have been able to use this system to engineer and produce proteins with various glycan structures attached [12-14]. Despite various *in vivo* expression platforms for glycoprotein synthesis are available, the inability to precisely control glycan structure of the end products has made it difficult to analyze the structure- function relationship of glycoproteins. In addition, the complex metabolic networks of cells have made the optimization of pathway challenging.

System engineering approach has been utilized to understand and optimize metabolic pathways of cells toward maximizing recombinant product formation. Instead of manipulating

single element of the system, system biology studies the responses of microorganism to genetic changes using data and model-driven approaches. A wide range of simulation methods have been developed to analyze biological systems. These include kinetic approach that requires the knowledge of kinetic parameters of each enzyme, and stoichiometric approach that relies on stoichiometric matrix and constrains [15]. Shuler and coworkers pioneered in constructing large scale single cell models across multiple organisms that incorporated multiple catabolic and macromolecular synthesis pathways constrained by experimental kinetic parameters [16]. Unlike kinetic models, cybernetic model was first developed to predict microbial growth and behavior on multiple substrates and metabolic engineering strategies [17]. Although these two types of modeling strategies have been used widely and successfully in understanding and predicting metabolic changes, they are usually nonlinear systems that have complicated mathematics, and require a large set of experimental parameters that are expensive to implement.

Constrained based flux balance analysis (FBA) is a widely used stoichiometric based mathematical approach to study genome wide metabolic network toward an optimal objective. FBA simplifies kinetic systems to simple linear algebraic system of mass balances. This modeling strategy assumes the system is operating at pseudo- steady state [18, 19]. It is commonly implemented for *in vivo* system study and optimization, but as mentioned above, *in vivo* systems do not allow precise control of metabolite inputs; and, metabolites may be redirect to account for cell growth.

Cell free protein synthesis (CFPS) technology, however, allows us to bypass the complex metabolic networks and growth factors within cells. CFPS is an *in vitro* protein production platform using biological machinery in a non-living cell environment [20]. This technology has been receiving growing attention as a tool for synthetic biology research and also for therapeutic production [21-23]. Today, it is a widely used tool for studying protein structure-

function relationship [20]. Therefore, in this work, we proposed to combine CFPS technology and mathematical approach, FBA, to understand metabolic networks involved in glycoprotein synthesis and to produce structurally- defined glycoproteins. To achieve that goal, we first aimed to develop an *in vitro* enzymatic glycan remodeling platform where we could produce homogeneous glycoproteins for structure- function relationship study. Then, we aimed to characterize the glycoprotein synthesis performance in our CFPS lysate systematically using FBA approach.

1.2 Protein glycosylation

N-linked glycosylation is one of the most essential and common PTMs found in eukaryotic cells. This process involves the synthesis of preassembled oligosaccharide($Glc_3Man_9GlcNAc_2$) in the lumen of endoplasmic reticulum (ER) of eukaryotes. Multisubunit enzymes called oligosaccharyltransferase (OTase) then catalyzes the transfer of the preassembled $Glc_3Man_9GlcNAc_2$ moiety from the lipid carrier dolichyl pyrophosphate (Dol-PP) to an asparagine residue in the sequon Asn-X-Ser/Thr (N-X-S/T), of nascent polypeptide chains, where X is any amino acid residue except for proline [2, 3, 24].

Glycosylation is not only important in protein structure, function, stability, cell- cell and cell virus interaction, but its alternation can also lead to many diseases, such as cancer, cognitive disorder and mucosal diseases. This usually occurs when there is a change in glycosylation pattern or deficiency or overly producing of certain glycoform and enzymes. On many tumor cell surface, glycans are usually galactose terminated, which are suspected to be involved in the adhesion of metastatic cells [25]. The structure-function relationship present in diseases has potential for serving as therapeutic targets. But, despite many findings have identified glycan abnormal features of diseases, it is difficult to understand their mechanism and to find specific antibody [26].

1.3 Bacterial glycoengineering

Although *N*- linked glycosylation was once thought to present only in eukaryotes, the existence of glycosylation pathways has been discovered and established in bacteria and archaea [27]. Among them, *Campylobacter jejuni* glycosylation pathway is one of the most well studied in bacteria. The glycosylation machinery in *C. jejuni* is encoded in a single gene cluster called, *pgl*. This functional pathway was transferred into *E. coli* and served as microbial expression platform for glycoprotein synthesis. In this system, a heptasaccharide, GalNAc- α1,4-GalNAc- α1,4- (Glc-β 1,3)- GalNAc- α1,4- GalNAc- α1,4- GalNAc- α1,3-Bac (where Bac is bacillosamine) is synthesized on lipid carrier undecaprenylpyrophosphate (Und-PP) at the cytoplasm. The heptassacharide is later flipped to the periplasm by a flipase, PglK. Similar to eukaryotes, a *C. jejuni* OTase named, pglB, transfers the oligosaccharide onto an asparagine residue of a polypeptide that has a consensus sequence of D/E-X-N-X-S/T, where X is any amino acid except for proline. PglB has been shown to share similar structure as the eukaryotic OTase, Stt3. And, that both of these OTases are transmembrane proteins [1-3, 28]. The similarities between prokaryotic and eukaryotic glycosylation have motivated scientists to express proteins containing human-like or human- relevant glycoform in *E. coli*.

One engineering example using *E. coli* based expression platform is the synthesis of $Man_3GlcNAc_2$ glycoform. This glycofom is significant because it is a common consensus glycan core structure in human N- glycans. Dr. DeLisa's group has successfully produced and transferred $Man_3GlcNAc_2$ structure onto maltose binding protein (MBP) and single variant fragment antibody R4 (scFV13- R4) with a single C- terminal DQNAT glycosylation sequon via bacterial OTase pglB in *E. coli*. Although the efficiency was low, it has demonstrated possibility to engineer bacteria to express non-bacterial glycan structure [13].

1.4 Glycan remodeling technology

Chemoenzymatic and chemoselective remodeling platforms are alternative methods for producing structure- defined glycoproteins with high specificity. Chemoenzymatic platform is an *in vitro* based system that utilizes a class of endoglycosidases (EnGases) to ligate free glycan and GlcNAc- tag- protein to form homogeneous glycoprotein. This technology requires the synthesis of intact oligosaccharides with oxazolines group that mimics the transition state of donor substrate. Later, endo-β-N- acetylglucosaminidase (ENGase) catalyzes trans-glycosylation to transfer preassembled oligosaccharide oxazolines onto GlcNAc – containing peptide or protein in a single step chemical reaction [29, 30].

Chemoselective remodeling platform is another *in vitro* glycan remodeling platform. It utilizes site- directed mutagenesis in combination with chemical synthesis to introduce specific glycan structures onto protein. This strategy utilizes many different functional groups to facilitate formation of chemical bounds between protein and chemical synthesized glycan. One example is to introduce cysteine residue onto selected position of protein. The thiol group of cysteine residue is then reacted with glycol- methanethiosulfonate (glyco- MTS) reagents, and forms disulfide bond. Although both strategies are highly specific and can produce designed homogeneous glycoforms, they require synthetic chemical strategies, which can limit the complexity of glycan structure and can introduce chemical contamination [31].

1.5 Cell free protein synthesis (CFPS)

Cell free protein synthesis (CFPS) is a well-established and powerful technology for expressing proteins in cell lysate. CFPS has been used for decades as a research tools for fundamental and applied biology. Nirenberg and Mattaei used CFPS to decipher and elucidate the genetic code in 1961 [20, 32]. To date, CFPS is served as a tool for synthetic biology study and it is an emerging system for therapeutics production. Compared to *in vivo* protein

expression, CFPS has many advantages. CFPS produces proteins in cell lysate therefore decouples cell viability and production objective. Moreover, the absence of cell walls creates an open environment, which is convenient for direct monitoring, rapid sampling, and direct manipulation of protein production [33]. Another motivation for using CFPS is that it is portable and can produce desired therapeutics on demand. This system allows rapid production with a timeline of 1-2 days, whereas *in vivo* production usually takes about 1 week [21].

One important application of CFPS is that it has been successfully used for expressing membrane proteins in wheat germ, rabbit reticulocytes, and insect cells made lysate [34, 35]. However, in this study we aim to couple CFPS with *in vitro* glycosylation using *E. coli* based cell lysate to produce designer glycoproteins. *E. coli* lysate has many advantages such that it has low manufacturing cost, robust in high- pressure homogenizer, and rapid growth titer. Most importantly, *E. coli* lacks glycosylation pathway, which provides us a clean chassis for bottom-up glycoengineering as well as avoiding possible contamination from native glycosylation machinery.

Although CFPS has many benefits compared to *in vivo* expression systems, it is limited by its reaction size, product yield and efficiency. Because the production of CFPS depends largely on cell extracts, many studies have been made in optimizing cell extracts preparation and reaction conditions to overcome those obstacles. This opens up an opportunity to couple CFPS with mathematical models to study and optimize cell- free metabolism.

1.6 Constrained based flux balance analysis (FBA)

Constrained based flux balance analysis (FBA) is a widely used mathematical approach for studying biochemical networks, including genome scale networks [36]. Constrained based FBA uses linear programing method to predict productivity, yield, and mutant behavior for biochemical networks. This method is also a metabolic engineering method to identify the best

approach for producing desired products, such as chemicals and pharmaceutics. Constrained based FBA calculates metabolic fluxes across cells through deconstruction of metabolic networks into stoichiometric matrix under two assumptions: 1) The system must be at pseudo steady state. 2) the system is operating at its optimal to achieve a cellular objective. The problem is constrained by mass conservation equation and the upper and lower bounds of network metabolites. These constrains setup the allowable space for flux distribution (Figure 1.1). The next step of FBA is to setup a biological objective that the problem is interested in studying. Constraint based model is a powerful tool for estimating the performance of metabolic network using very few parameters [19, 37]. This method is a lot simpler than kinetic models that usually require kinetic parameters of enzymes and that the problem is usually non-linear.

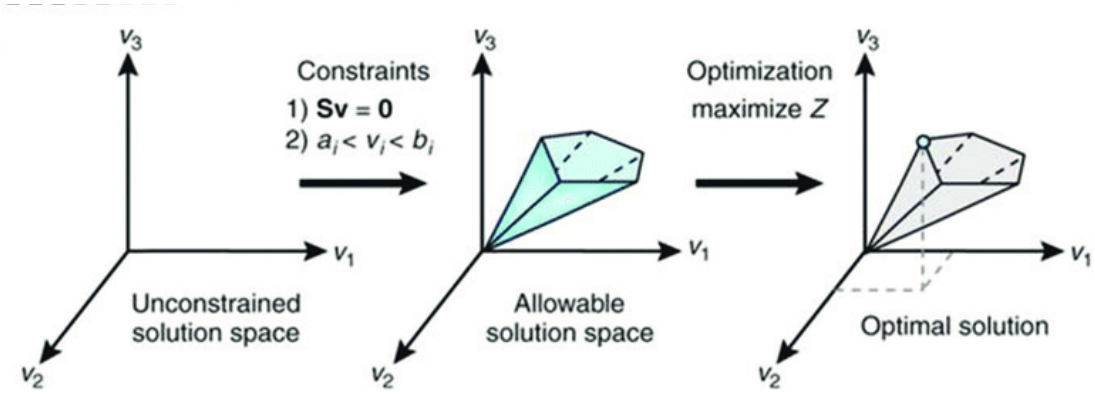


Figure 1.1 A schematic for constrained based flux balance analysis (FBA). S denotes for stoichiometric matrix and v denotes for metabolite fluxes. The unconstrained space is constrained by 1) the mass balance equation of metabolites and 2) the lower and upper bounds of metabolite fluxes. After an allowable space is defined, an objective needed to be defined in order to seek an optimal solution within the allowable space [19].

CHAPTER 2

THE DEVELOPMENT OF *IN VITRO* GLYCOPROTEIN REMODELING SYSTEM

2.1 Introduction

Glycosylation plays an essential role in eukaryotic cells. With incorrect glycoform, proteins tend to misfold, degrade, or can lead to diseases formation [38, 39]. Although many studies have done to engineer metabolic pathways in cells to produce designed glycoproteins [40, 41], heterogeneous glycofrom product cannot be avoided due to endogenous glycosylation pathways present in mammalian cells [42]. Some alternative methods are chemoenzymatic and chemoselective methods, which utilize chemical synthetic methods to produce preassembled glycan structures with functional groups and then attached to selected site through chemical reactions [29, 31, 43]. Despite these *in vitro* methods allow high specificity and homogeneous product formations, they are limited by the complexity of glycans that synthetic chemistry can produce. Therefore, herein we sought an alternative *in vitro* platform where we could bypass metabolic complexity of cells and also difficulties associated with chemical synthesis to produce homogeneous designed glycoproteins.

In this enzymatic glycan remodeling platform, we aimed to engineer and express functional enzymes to modify model glycoproteins. The advantage to combine *in vivo* expression methods with *in vitro* glycan remodeling technology is that we could produce designed glycoproteins homogeneously and in large quantity. The first step of this process was to synthesize model glycoprotein precursors. Later, sequential enzymatic reactions were setup using the glycoprotein precursors to form desired glycoforms. The two glycan structure precursors were 1) GlcNAc and 2) $Man_3GlcNAc_2$. These two glycan structures were chosen because majority of *N*- linked glycans begin with GlcNAc carbohydrate. And, most of human *N*-linked glycoproteins share this $Man_3GlcNAc_2$ structure as their consensus core structure. By

having these two glycan structures as our model glycoprotein precursors, we had the potential to enzymatically elaborate on those glycan structures to a wide range of other *N*- linked glycoproteins. In this study, two glycoforms were focused: 1) a liner polysialic acid structure and 2) a biantennary structure. In figure 2.1, two glycosylation pathways of interest were shown. The first pathway begins with protein with single GlcNAc precursor and ends with a single polysialic acidd structure. The second pathway begins with $Man_3GlcNAc_2$ glycan structure and ends with a biantennary structure.

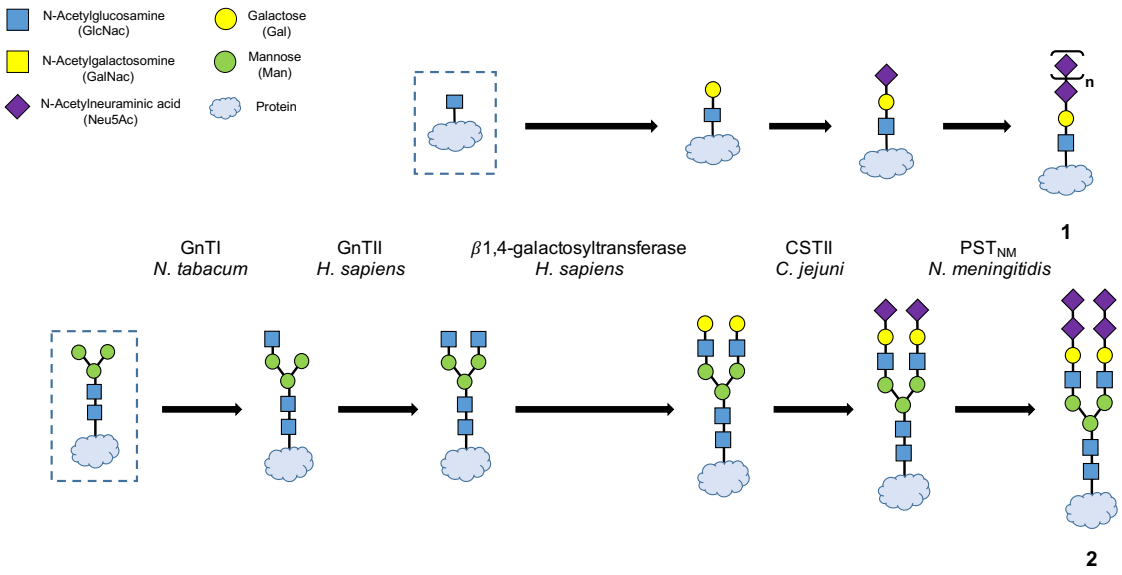


Figure 2.1 Proposed pathways for *in vitro* glycan remodeling platforms. Product 1 is polysialic acid structure and **Product 2** is biantennary structure. To synthesize **Product 1**, we utilized glycoprotein bearing single (GlcNAc) as our starting material, whereas for **Product 2**, we utilized glycoprotein bearing ($Man_3glcNAc_2$) glycan structure as out starting material. Enzymes involved in synthesizing these two pathways are N- acetylglucosaminyltransferases, 1) GntI and 2) GntII derived from *Nicotiana tabacum* and *Homo. Sapiens,* 3) β-1,4- galactosyltransferase from *Homo sapiens*, 4) **α**-2,3- sialyltransferase (CSTII) from *C. jejuni*, and 5) **α**-2,8- polysialyltransferase (PST^{NM}) from *Neisseria meningitides*.

2.1.1 Production of linear polysialic acid-containing glycoprotein

Linear polysialic acid, **α**-2,8 linked sialic acid (PolySia) is a modification on many *N*-glycans on specific cell surface proteins in central nervous (CNS) and immune systems. In CNS, PolySia is attached to neutral cell adhesion molecule (NCAM), where its hydrodynamic property inhibits cell- cell interactions [44]. Due to its large hydrodynamic volume, PolySia has been receiving as a potential "stealth polymer" for conjugation with protein drugs to improve therapeutics efficacy [45]. To synthesize proteins with linear polysialic acid, we utilized glycoprotein precursor with single GlcNAc. As proof of concept, we first employed already established *in vivo* expression method to express proteins, being glycosylated with 1) engineered *Campylobacter lari* glycan, ($GalNAc_4GlcNAc$) in *E. coli* or 2) complex biantennary *N*-glycan. Next, we treated proteins having $GalNAc_4GlcNAc$ structure with exoglycosydase, **α**-N-acetylgalactosomine and antibody having biantennary structure with endoglycosidase S (Endo S) to trim down to single GlcNAc moiety (Figure 2.2). Proteins that were used in this study were single chain fragment variable R4 (scFv13 R4), human growth hormone variable, (HGHv), and Herceptin. Both scFv13 R4 and HGHv were purified from *E. coli* with $GalNAc_4GlcNAc$ structure and Herceptin was purified from mammalian cell line 293F with biantennary structure before enzymatic treatment. The reason for choosing these three proteins was to show the possibility to modify therapeutically relevant proteins using our method. And, later of the process, we aimed to show PolySia could effectively improve the stability of those proteins through serum half-life study.

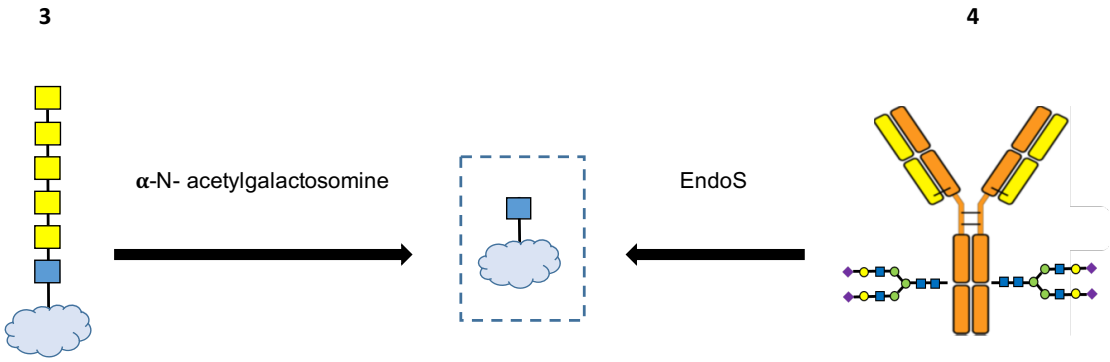


Figure 2.2 A schematic for the synthesis of glycoprotein bearing single (GlcNAc) structure. **Structure 3** is protein bearing $GalNAc_4GlcNAc$ purified from *E. coli in vivo* and can be trimmed down to single GlcNAc by using **α**-N- acetylgalactosomine. **Structure 4** is antibody bearing biantennary structure, which can be trimmed down using EndoS.

After obtaining protein precursors, enzymes needed for the pathways were identified from KEGG enzyme database. In figure 2.1, three enzymes were needed to synthesize **Product 1**: β-1,4- galactosyltransferase from *Homo sapiens*, **α**-2,3- sialyltransferase (CSTII) from *C. jejuni*, and **α**-2,8- polysialyltransferase (PST^{NM}) from *Neisseria meningitides*. In this study, we specifically focused on proving the activity of CSTII and PST^{NM}. Activity of CSTII was verified using Alpha- 1 antitrypsin (A1AT), which had been shown as a protein substrate for CSTII [12]. Therefore, we decided to repeat the same experiment to show our CSTII also retain its activity. Activity of PST^{NM} was tested using GlcNAc-scFv13 R4 with the treatment of β-1,4-galactosyltransferase and CSTII.

2.1.2 Production of biantennary glycoprotein

One of the common human-like glycans is N-linked biantennary glycan structure. It is usually present on the Fc domain of antibody. Most of time, a fucose moiety is attached to the first GlcNAc of the biantennary structure, which has been shown to lower the binding affinity of IgG1 to the FcγIII receptor. Many studies have been made to engineer cell lines or additional treatment to produce antibody with afucosylated biantennary structure [10, 46]. Therefore, it

was in our interest to also synthesize therapeutic relevant proteins with afucosylated biantennary structure. To achieve that goal, we first expressed our glycoprotein precursor, $Man_3GlcNAc_2$-MBP- Glucagon, in *E. coli* based on already established method [47]. Unlike protein precursor with single GlcNAc, $Man_3GlcNAc_2$- MBP- Glucagon did not require additional treatment. Enzymes needed to form the biantennary structure (**Product 2**) shown in figure 2.1 included N- acetylglucosaminyltransferases, GntI and GntII derived from *Nicotiana tabacum* and *Homo. sapiens*, β-1,4- galactosyltransferase, CSTII and PST^{NM}. To synthesize biantennary structure, we decided to utilize protein with $Man_3GlcNAc_2$ structure, a consensus structure for human-like N-glycosylation, as our starting material. Because $Man_3GlcNAc_2$ structure does not exist in nature, it requires a lot of engineering work to be synthesize in *E. coli*. Currently, with the knowledge we have in the lab, we glycosylated MBP-Glucagon, a peptide hormone, with $Man_3GlcNAc_2$ structure *in vivo* in *E. coli*. $Man_3GlcNAc_2$-MBP-glucagon has also been shown to be a good substrate for GntI and GntII [47].

2.2 Results for bottom-up design of in vitro glycan remodeling platform

2.2.1 Generation of N-linked glycoprotein precursor

In an effort to produce high yield homogeneous N- linked designer glycoprotein, we employed previously developed methods to first produce N-linked glycoproteins *in vivo* in *E. coli* [12], then followed by *in vitro* treatments to remodel the glycan structure into designer glycofrom (Figure 2.1). This method allowed us to produce glycoprotein precursor of interest with a high yield and with minimal genetic engineering work. Specifically, two glycoprotein precursors were focused in this project: 1) protein bearing a single GlcNAc and 2) protein bearing $Man_3GlcNAc_2$ structure. As described in the introduction, many human- liked N-linked glycoproteins usually begin with a single GlcNAc or share a consensus $Man_3GlcNAc_2$ structure. Therefore, it was in our interest to generate these two glycoforms as our starting glycoprotein precursors for our *in vitro* remodeling platform. And later we aimed to use glycoprotein

precursor bearing GlcNAc to produce linear polysialic acid structure, and glycoprotein precursor bearing $Man_3GlcNAc_2$ to produce biantennary structure.

Based on the *in vivo* glycoprotein synthesis methods developed previously, we expressed proteins bearing $(GalNAc)_4GlcNAc$, $Man_3GlcNAc_2$, or biantennary glycan structure in E. coli strains: CLM24 or Origami 2 ΔnanAΔwaaLΔgmd or in mammalian cell line: 293F (Figure 2.3). Protein bearing the later glycan structure, $Man_3GlcNAc$, was directly used in the *in vitro* assays to synthesize biantennary structure. However, for proteins bearing either $(GalNAc)_4GlcNAc$ or biantennary structure additional treatment with exoglycosidase, α-N-acetylgalactosaminidase, or endoglycosidase, EndoS, was required to produce glycoprotein precursors bearing GlcNAc (Figure 2.4). In this study, we were interested in investigating the structure and function relationship of therapeutically relevant proteins. Thus, we chose antibody fragment scFv13 R4 and human growth hormone variant (Hghv) as our model glycoprotein for synthesizing the linear polysialic acid structure, and MBP-Glucagon for synthesizing biantennary structure. We were also interested in preparing antibody bearing GlcNAc as our glycoprotein precursor. This would be beneficial in creating a repertoire of different designed therapeutics relevant glycoproteins for structure- function relationship study.

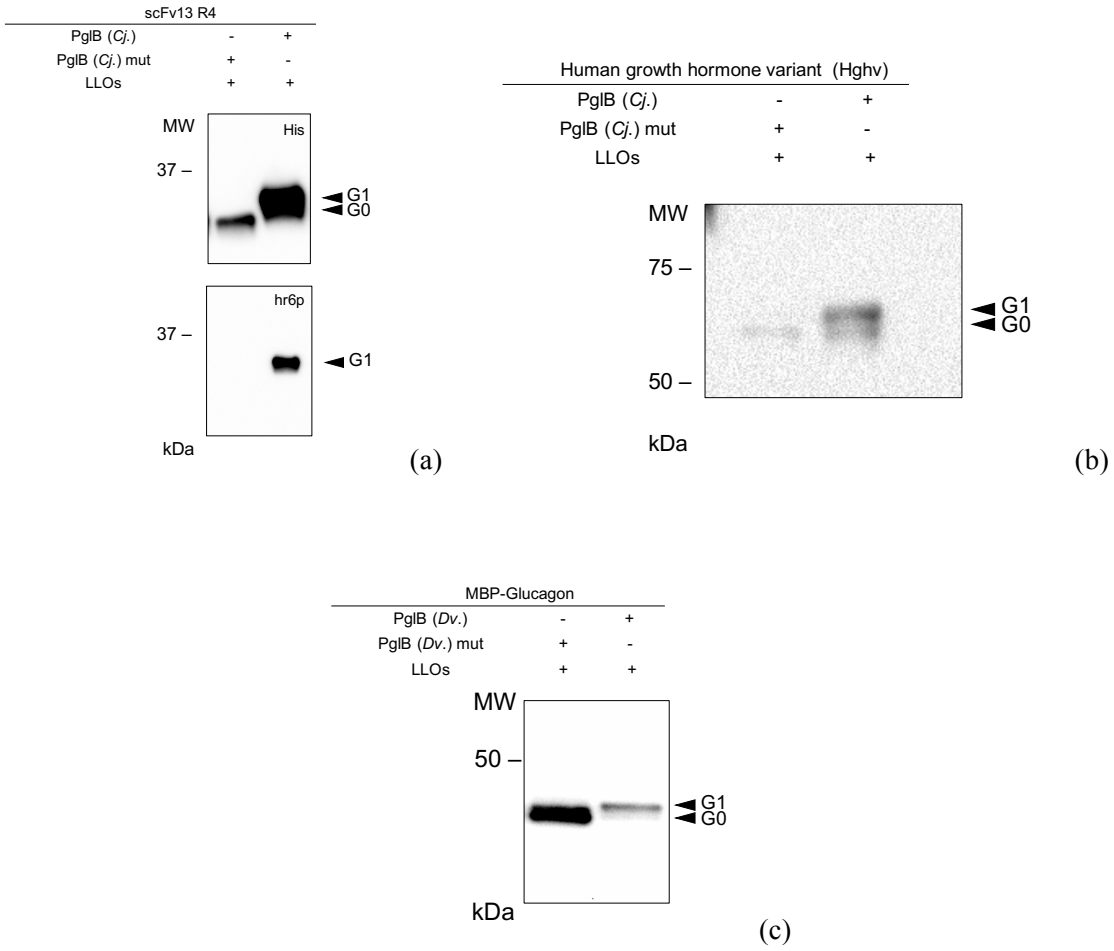


Figure 2.3 Glycoproteins bearing ($GalNAc_4GlcNAc$) (a,b) and ($Man_3GlcNAc_2$) (c) expression detected by immunoblots. a) Weston blot for detecting $GalNAc_4GlcNAc$-scFv13 R4 against histidine tag (up) and glycan serum, hr6p (down). G1 denotes for glycosylation band and G0 denotes for aglycosylated band. First lane was loaded with aglycosylated protein sample, and second lane was loaded sample with glycosylated sample. b) Weston blot for detecting $GalNAc_4GlcNAc$-MBP-Hghv against histidine tag. There is no result for detecting against hr6p serum due to shortage on serum material. (c) Expression result for $Man_3GlcNAc_2$-MBP-Glucagon against histidine tag. The first lane contained aglycosylated MBP-Glucagon, whereas the second lane contained $Man_3GlcNAc_2$-MBP-Glucagon. Note: Antibody used in this study were either purchase commercially or obtained from Emily Cox.

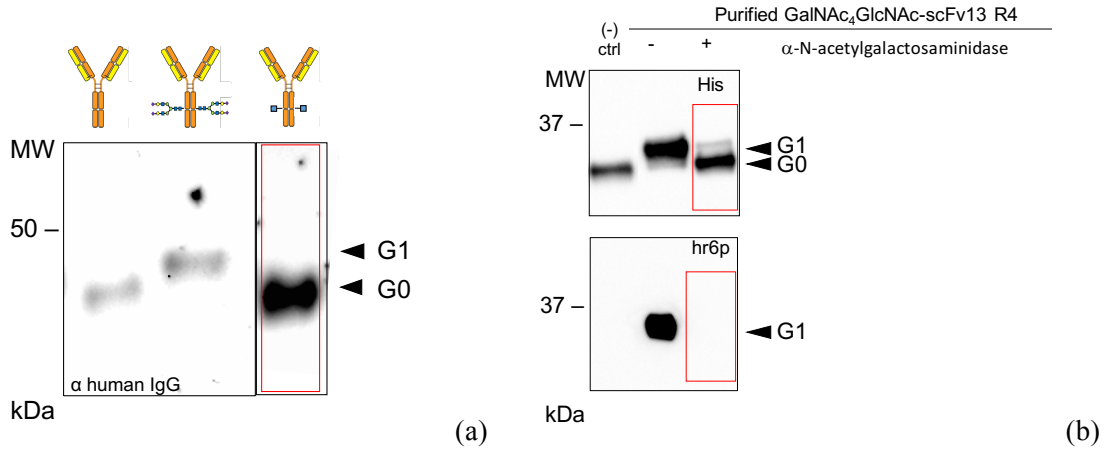


Figure 2.4 Immunoblots for verifying the activity of EndoS and α-N-acetylgalactosaminidase. (a) Detection of aglycosylated, glycosylated and trimmed Herceptin against human IgG antibody. From left to right were *in vivo* purified Herceptin treated with PNGase F, *in vivo* purified glycosylated Herceptin bearing complex biantennary glycan structure, and trimmed glycosylated Herceptin by EndoS separately. (b) Detection of a glycosylated, glycosylated, and trimmed scFv13 R4 against histidine tag (up) and glycan serum hr6p (down). From left to right were in vitro purified aglycosylated scFv13 R4, in vitro purified $GalNAc_4GlcNAc$-scFv13 R4, and GlcNAc-scFv13 R4 after trimming. Note: there is no result here showing the trimming activity for $GalNAc_4GlcNAc$-MBP-Hghv. It requires further experiments to consolidate the trimming data for $GalNAc_4GlcNAc$-MBP-Hghv.

2.2.2 Engineering and expressing glycosyltransferase- GntI, GntII, CSTII, and PST^{NM}

To produce glycoform **1** and **2**, several enzymes were selected to facilitate the *in vitro* remodeling process. N- acetylglucosaminyltransferase, GntI and GntII, transfer two GlcNAc onto $Man_3GlcNAc_2$; **α**-2,3 sialyltransferase, CSTII, transfers sialic acid onto glycoform bearing terminal galactose; **α**-2,8- polysialyltransferase, PST^{NM} polymerizes sialic acids onto glycoform bearing two terminal sialic acids. Most of the glycosyltransferases are known to be membrane anchor or membrane bound proteins. Thus, in this study, GntI and GntII were engineered to be soluble by fusing each of them to mannose binding protein (MBP) and remove their membrane anchor region [47]. CSTII was optimized based on [48]. Although CSTII is also a membrane anchor protein, it has been shown to be partially soluble when expressed. PST^{NM} is also a membrane anchor protein. It was engineered by removing its 19 amino acids at the *N*-terminal and fused to MBP [49]. MBP-GntI and MBP- GntII were separately expressed in MC4100 and Origami2 ΔnanAΔwaaLΔgmd::kan strains (Results not shown). CSTII and PST^{NM} were expressed in BL21(DE3) (Figure 2.5).

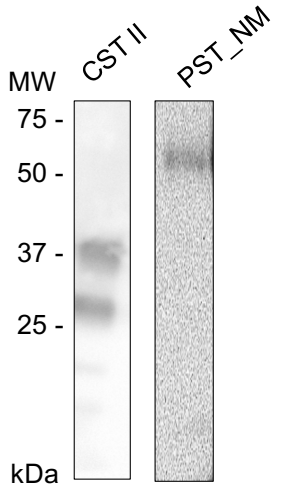


Figure 2.5 Immunoblots for detecting α-2,3 sialyltransferase, CSTII and α-2,8- polysialyltransferase, PST^{NM} fused to MBP against histidine tag. CSTII is about 34 kDa and MBP-PST^{NM} is about 61 kDa.

2.2.3 Activity of N-acetylglucosaminyltransferases- GntI and GntII

One focus of this work was to bioenzymatically synthesize complex biantennary glycan structure by using a common human N-linked intermediate glycan structure, $Man_3GlcNAc_2$, as our starting glycan precursor. To first test the activity of MBP-GntI and MBP-GntII, we supplied both of the enzymes into a reaction mixture where $Man_3GlcNac_2$MBP-glucagon UDP-GlcNAc were present. We observed a band shift detected by antibody against histidine tag via immunoblot for reaction containing MBP-GntI, but not MBP-GntII (Figure 2.6). To further confirmed the glycosylatransferases activity, we measured fluorescence generated by click chemistry method. In this experiment, we supplied MBP-GntI and MBP-GntII individually into different reaction mixture. We were only able to observe a band for reaction containing MBP-GntI under fluorescence (Figure 2.7), but not for the reaction containing MBP-GntII (result not shown). These results indicated GlcNAc residues were transferred onto the terminal α-1,3 mannoses successfully but not α-1,6.

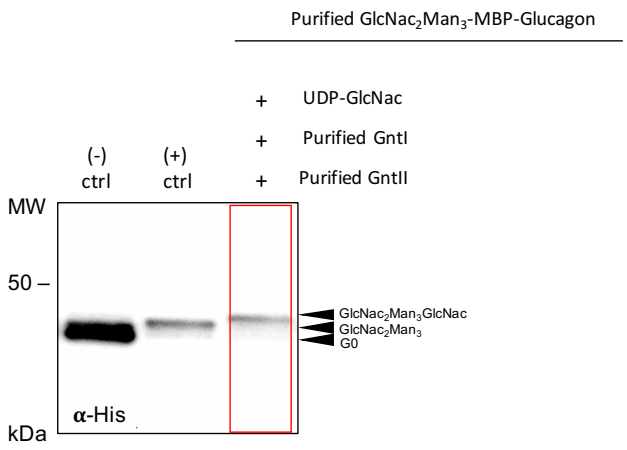


Figure 2.6 Immunoblot for verifying the activity of MBP-GntI and MBP-GntII against histidine tag. Lane one contained in vivo purified MBP-Glucagon; lane 2 contained in vivo purified $Man_3GlcNAc_2$-MBP-Glucagon; and lane 3 contained $GlcNAcMan_3GlcNAc_2$-MBP-Glucagon.

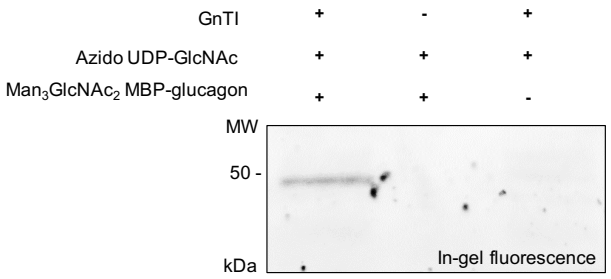


Figure 2.7 In gel fluorescence detection for the activity of GntI using $Man_3GlcNAc_2$-MBP-Glucagon. In lane one, reaction contained all the materials needed to glycosylate $Man_3GlcNAc_2$-MBP-Glucagon into $GlcNAcMan_3GlcNAc_2$-MBP-Glucagon. The second lane did not contain enzyme, MBP-GntI, and lane three did not contain any protein.

2.2.4 Activity of α-2,3-sialyltransferase- CSTII

To synthesize either complex biantennary or linear polysialic aicd structure, we needed to install sialic acid onto galactose residue first as polymerase enzyme, PST^{NM} requiring minimally two sialic acid residues as its substrate. Hence, we utilized the α-2,3-sialyltransferase from *C. jejuni* called CSTII to remodel our glycoprotein to install sialic acids. CSTII has the activity to transfer CMP-Neu5AC donor sugar onto a β-galactose acceptor sugar with an α-2,3 linkage and also onto an α- linked sialic aicd acceptor sugar with an α-2,8 linkage. Before testing whether CSTII had the activity to transfer CMP-Neu5Ac onto a β-galactose acceptor sugar using either GlcNAc-scFv13 R4 or GlcNAc- MBP-Hgh as our model protein, we assessed

its activity by transferring CMP-Neu5Ac onto an α-linked sialic acid acceptor sugar using natural human alpha-1-antitrypsin, A1AT (abcam). The activity of CSTII was verified through click reaction and visualized by fluorescence under Alexa 488 (Figure 2.8).

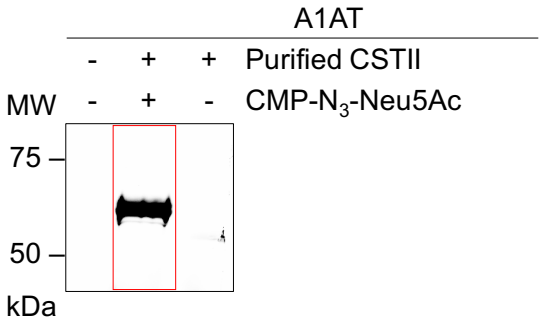


Figure 2.8 In gel fluorescence detection for the activity of α-2,3 sialyltransferase, CSTII using human alpha-1-antitrypsin,A1AT, which has biantenarry structure attached. In lane one where there was an absence of both CSTII and CMP-N_3-Neu5Ac, fluorescence was not detected. In lane two when A1At, CSTII, and CMP-N_3-Neu5Ac were present in the reaction, fluorescence was detected. In lane three, the reaction lacked VMP-N_3-Neu5Ac, fluorescence was again not detected.

2.3 Discussion and future direction

In this chapter, we have developed a platform to enzymatically remodel glycoproteins. The development of this platform included 1) generation of glycoprotein precursors 2) engineering and expressing glycosyltransferases in *E. coli*. 3) verifying the activity of glycosyltransferases through *in vitro* glycosylation activity assays. The enzymatic glycan remolding platform offers an opportunity to synthesize any desired *N*-linked glycoform homogeneously *in vitro*. This method avoids the presence of other glycoforms that usually occur in mammalian expression platform. Moreover, it bypasses the need to chemically synthesize glycan substrates, which was needed in chemoenzymatic and chemoselective remodeling platforms.

Herein, we have demonstrated methods for producing glycoprotein precursors with single GlcNAc moiety and also with $Man_3GlcNAC_2$ by combining *in vivo* expression methods and *in vitro* trimming methods. This combine approach for producing glycoprotein with single

GlcNAc allowed us to generate glycoprotein precursors with large quantity that is usually not able to be achieved using in vitro platform. More importantly, this allowed us to synthesize any *N*-linked glycoprotein of interest. For $Man_3GlcNAC_2$ structure, although we could express $Man_3GlcNAC_2$-MBP-Glucagon with significant yield and modify it to form $GlcNAcMan_3GlcNAC_2$-MBP-Glucagon using our enzymatic remodeling platform, there was a limitation toward this process. Proteins are generally difficult to be glycosylated with $Man_3GlcNAC_2$ structure because it does not exist in nature. Until now, MBP- Glucagon is by far the only protein that can be glycosylated in high efficiency with $Man_3GlcNAC_2$ structure using the current *E. coli* based *in vivo* glycoprotein expression system. This limits our *in vitro* glycoprotein remodeling platform to have a variety of glycoprotein substrates bearing $Man_3GlcNAC_2$ structre. However, recently efforts have been made to improve the in vivo glycosylation efficiency for $Man_3GlcNAC_2$ structure using *E. coli* [50]. This indicates that there are potentials to improve expression level for glycoproteins bearing $Man_3GlcNAC_2$ structure and open up other potential glycoprotein candidates for our *in vitro* glycoprotein remolding platform.

After the generation of glycoprotein precursors, we identified enzymes needed to produce polysialic acid structure and complex biantennary glycan structure using glycoprotein precursors either with single GlcNAc or with $Man_3GlcNAC_2$. We were able to engineer and purify MBP-GntI, MBP-GntII, CSTII, and PST^{NM} in soluble form in *E. coli*. Later, by performing *in vitro* glycosylation reactions, we obtained positive signals for the activity assays of MBP- GntI, MBP-GnTII, and CSTII through either western blotting or click chemistry. However, the activity assay for CSTII appeared to have weak signal compared to the background signal. Further improvements on the design of CSTII could potentially enhance CSTII enzymatic activity. To further confirm their activity and the purity of products, more detailed study, such as mass spectroscopy or NMR will need to be carried out. Moreover,

enzyme kinetic analysis can assist our understanding of enzyme activity to optimize our *in vitro* glycosylation reaction conditions and to be incorporated into model designs. The final goal of this project was to produce homogeneous designed glycoprotesin that can be used for studying glycoprotein structure-function relationship. Thus, after all the analysis, we aim to utilize the glycoproteins synthesized using this in vitro glycan remodeling platform to study the effects of different glycoform have on protein folding, function, serum half- life, and cytotoxicity to the cells. Experimental studies such as X- ray crystallography, NMR can help elucidating those objectives.

2.4 Methods

2.4.1 Strains and plasmids

E. coli strain DH5α was used for cloning and site- directed mutagenesis while strains CLM24, Origami2 p, MC4100 (araD139 D(argF-lac)205 flb-5301 pstF25 rpsL150 deoC1 relA1), and BL21(DE3) (F^- ompT hsdSB(rB^- mB^-) gal dcm(DE3)) were used for expressing different proteins used in this project. Glycoproteins, such as $GalNAc_4GlcNAc$-scFv13 R4 and $GalNAc_4GlcNAc$- MBP-Hgh were expressed in CLM24; $Man_3GlacNAc_2$-MBP-glucagon was expressed in Origami2 Δ*nanA* Δ*waaL* Δ*gmd*::kan. MBP- GnTI and MBP-GnTII were separately transformed and expressed in MC4100 and Origami2 (Novagen). CSTII and PST^{NM} were expressed in BL21(DE3). Plasmid pMAF10 encoding for *C. jejuni* pglB derived from pMLBAD vector and plasmid pACYC-pgl2 comprising lipid-linked engineered *C. lari* glycan-$GalNac_4GlcNac$ were co- transformed into CLM24. Later, they were transformed again with plasmid encoding for the protein substrate, either pBS $R4^{DQNAT}$-6x His or pTrc-ssDsbA-MBP-TEV-hgh-6xHis. As a negative control, plasmid pMAF10 was replaced with pMAF10 mutant, which lacks the oligosaccharyl transferase activity. Plasmid pYCG-pglB (from *C. vulgaris*) was used to produce lipid-linked $Man_3GlcNac_2$. This plasmid was co- transformed with pManCB-

glucagon, which comprised of the linearized pMQ70 backbone and *E. coli manB* and *manC* genes. Genes encoded for GntI, GntII, PST_{NM} and CSTII were amplified from chromosomal DNA of *Nicotiana abacum*, *Homo sapiens*, *Neiseria maningitidis*, and *Campylobactor jejuni* separately. MBP-GntI and MBP-GntII constructs were kindly given by Glycobia [47]. Optimized CSTII gene [48] was cloned into PET28a with the digestion sties at NcoI and SalI. The construction of PST^{NM} gene was cloned from PET28a vector with a deletion of 19 amino residues at the N terminus, resulting an increase in solubility [49]. Gene $PST^{NM}\Delta19$ was inserted into XbaI and HindII sites of pMAL-c4x vector with N- terminal 6x His tag.

2.4.2 Expression and purification of model glycoprotein precursors

E. coli strain CLM24 was chemically transformed with pMAF10/pMAF10 mutant, pACYC-pgl2 and pBS- $R4^{DQNAT}$-6x His or pTRC99y MBP-hghv- 6xHis- NeuDBAC. Transformed cells were inoculate in a 10 mL LB. Cultures were grown at 37^{o}C overnight containing 100 μg/mL trimethoprim (Tmp), 80 μg/mL spectrinomycin (Spec), and 20 μg/mL chloramphenicol (Cm) or 100 μg/mL ampicillin (Amp). The 10 mL overnight culture was then transferred into 1L LB and shacked at 37^{o}C. At OD_{600} of 1.0, cells were induced with 0.2% (w/v) L-arabinose and 0.1 mM IPTG at 16-25^{o}C for 16-20 hours. *E. coli* strain Origami2 Δ*nanA* Δ*waaL* Δ*gmd*::kan carrying pYCG-pglB (from *C. vulgaris*) and pManCB-glucagon was inoculated in 10 mL LB containing 2% (v/v) glucose along with 20 μg/mL chloramphenicol (Cm) and 100 μg/mL ampicillin (Amp) at 37^{o}C. Overnight culture was then added to 1L TB supplemented with Cm and Amp and shacked at 37^{o}C until OD_{600} reached 1.5. Culture was then induced with 0.2% (w/v) L-arabinose at 30^{o}C for 12-16 hours followed by 4 hours induction by 0.1mM IPTG the next day. Cells were harvested by centrifugation at 8,000 x g for 20 min at 4^{o}C and resuspended in 10 mL of Buffer A (50 mM HEPES, 250 mM NaCl pH 7.5) per 1 g of wet cell mass. Cells were then lysed by Avestin C5 EmulsiFlix homogenizer at 17,000 psi for

three passes. Lysed cells were centrifuged at 30,000 x g for 30 min at 4°C to remove cell debris. Soluble fraction was collected and rotated in 10 mM imidazole and 0.25 mL Ni- NTA resin at 4°C for at least 1 hour. Sample was loaded into gravity column, washed with washing buffer (50 mM HEPES, 250 mM NaCl pH 7.5, 25 mM imidazole), and collected with elution buffer (50 mM HEPES, 250 mM NaCl pH 7.5, 300 mM imidazole). Protein was desalted using Zeba™ spinning desalting columns (ThermoFisher). Protein concentration was measured by Bradford assay (Bio-Rad).

2.4.3 Expression and purification of GnTs, CSTII, and PST^{NM}

E. coli strains MC4100, Origami were chemically transformed with either MBP-GnTI or MBP-GnTII. Both PET28a-CSTII^{opt} and MBP-PST^{NM}Δ19 were chemically transformed into *E. coli* strain BL21. Transformed cells containing GnTI, GntII, and PST^{NM}Δ19 were grown in 10 mL LB containing 100 μg/mL ampicillin (Amp) at 37°C overnight separately. Transformed cells containing CSTII was grown in kanamycin (Km) at 37°C overnight. Cultures were transferred into 1L LB media the next day and shacked at 37°C until OD_{600} reached 1.0. Culture was then induced by 0.2% (w/v) L-arabinose for GnTs and 0.1mM IPTG for CSTII and PST^{NM} at 16°C for 16-20 hours. CSTII and PST^{NM} were purified following the same purification method described in section 2.3.2. Cells containing MBP-GnTI and MBP-GnTII, however, were resuspended in 10 mL MBP binding buffer (20 mM Tris-HCl, 1mM EDTA, 200 mM NaCl, pH 7.4) per 1 g of wet cell mass. Resuspended cells were lysed by Avestin C5 EmulsiFlix homogenizer at 17,000 psi for three passes. Lysed cells were centrifuged at 30,000 x g for 30 min at 4°C to remove cell debris. Soluable fraction was collected and incubated by rotation in 4°C with 0.2 mL amylose resin for at least 1 hour. Later, sample was purified using gravity column and washed with MBP-binding buffer. Finally, protein was eluted with MBP elution

buffer (20 mM Tris-HCl, 1mM EDTA, 200 mM NaCl, pH 7.4, 10 mM maltose), and concentrated using 3K concentrator (ThermoFisher).

2.4.4 Synthesis of GlcNAc precursor

Synthesis of GlcNAc precursor, 1 μg of purified $GalNAc_4GlcNAc$- $R4^{DQNAT}$ or $GalNAc_4GlcNAc$-MBP- Hgh was treated with α-N- acetylgalactosominidase (New England BioLabs) following NEB's protocol. The reaction was carried out at 37°C for 1 hour followed by 10 min heat deactivation at 65°C. Synthesis of GlcNac precursor was preliminarily determined by Weston blot against histidine tag and against hr6p, an anti- engineered *C. lari* glycan serum (from Aebi lab).

2.4.5 Synthesis of $GlcNac_2Man_3GlcNac_2$-MBP-Glucagon

To produce $GlcNac_2Man_3GlcNac_2$-MBP-Glucagon, 1-5 μg of purified $Man_3GlcNac_2$-MBP-Glucagon was treated with 4 μg of MBP-GnT1 and MBP- GnT2 each and 2mM UDP-GlcNAc or 0.02μg/mL of UDP-N_3-GlcNAc in 20 mM HEPEs, 50 mM NaCl, 10 mM $MsnSO_4$ (pH 7.2) for 12 hours at 30°C. Product was detected by Weston blot against His tag. In addition, product produced by reacting with UDP- N_3- GlcNAc was later used to perform click reaction and detected through fluorescence under Alexa 488 mode.

2.4.6 Synthesis of $(Neu5Ac)_2$ $GlcNac_2Man_3GlcNac_2$-A1AT and $(Neu5Ac)_n(Neu5Ac)_2$ $GlcNac_2Man_3GlcNac_2$-A1AT

$(Neu5Ac)_2GlcNac_2Man_3GlcNac_2$-A1AT was synthesized by reacting 1-5 μg of Alpha-1- antitrypsin (A1AT) with 0.012 μg/mL CMP- N_3- Neu5Ac and 5 μg of purified CSTII in 50 mM sodium phosphate (pH 8) and 10 mM $MgCl_2$ at 37°C for 12 hours. Later, click reaction was performed to detect $(Neu5Ac)_2$ $GlcNac_2Man_3GlcNac_2$-A1AT under fluorescence. The product

was carried out to synthesize $(Neu5Ac)_n(Neu5Ac)_2GlcNac_2Man_3GlcNac_2$-A1AT by adding the previous reaction into 2 mM CMP- Neu5Ac (or 0.012 mM CMP- N_3- Neu5Ac) and 5 μg of purified $PST^{NM}\Delta19$, 50 mM sodium phosphate (pH 8) and 10 mM $MgCl_2$. The reaction was carried out at 30°C for 12 hours. The final product was detected through SDS-PAGE or fluorescence.

2.4.7 SDS-PAGE gel and Weston Blot

SDS-PAGE analysis were performed with Bolt 8% and 10% Bis-Tris gel (Thermo Fisher Sceientific) and Mini-PROTEAN® TGX™ 7.5%, 10%, and 12% Precast Gels in MOPS and SDS buffer accordingly. Most of the protein samples were boiled at 100°C for 10 minutes in 4x LDS protein dye with 10% 2-Mercaptoethanol, unless specified. For coomassie analysis, gels were stained with either R250 or G250 staining solution purchased from BioRad and destain with de-staining solution (water, methanol, and acetic acid in a ratio of 50/40/10 (v/v/v)) or deionize water. For Weston blot analysis, after gels finished running, they were transferred on to semi-transfer block or wet transfer block using membrane. Transferred image was block with 5% milk in TBST for minimally 30 minutes followed by 10 minutes washing step. Membranes were then incubated with antibody for one hour. Later, membranes were washed for 1 hour by changing TBST every other 10 minutes. If the samples did not require secondary antibody, they were imaged under Bio-Rad Molecular Imager® ChemiDoc™ XRS+ imaging system; otherwise, samples would require to repeat the incubation and wash steps with the secondary antibody.

2.4.8 Click Reaction

Click reaction was performed on samples that used azido sugar as their donor sugar substrate for the *in vitro* glycosylation assay. After *in vitro* glycosylation reaction was

completed, samples were treated with 100 mM iodoaceamide (IAM) and shaked at low speed at room temperature in the dark. Then, 100 μM carboxyrhodamine 110 DBCO dye was added to the reaction and incubated at room temperature in the dark for one hour. During this step, carboxyrhodamine 110 DBCO dye would interact with the azidol group. The samples were then boiled at 65^{o}C for 5 min in 4x LDS protein dye with 10% 2-Mercaptoethanol before loading on to SDS-PAGE gel (BioRad).

CHAPTER 3

SEQUENCE SPECIFIC FLUX BALANCE ANALYSIS (SSFBA) ON GLYCOPROTEIN SYNTHESIS USING CELL FREE PROTEIN SYSTHESIS (CFPS) IN COMBINATION WITH *IN VITRO* GLYCOPROTEIN SYSNTHESIS

3.1 Introduction

Metabolic networks are described by intricate enzymatic and chemical reactions and are usually not optimized for producing desired products in practical applications [51]. System biology aims to understand and manipulate biological processes by studying the responses of microorganism to genetic changes through data and model- driven approaches [15]. The first mathematical model introduced by Jacques Monod successfully describe cell growth [52]. However, Monod's model cannot handle situations when there is a change in environment [51]. Later, metabolic control analysis (MCA) was proposed by Kascer and Burns to analyze the effects of small change in enzyme concentration on flux distribution and intermediate concentrations of a metabolic pathway. MCA models require *a priori* knowledge of enzyme concentrations, which are difficult to acquire. Biochemical system theory (BST) was another mathematical theory attempted to describe metabolite changes. BST utilized rate constants as its sensitivity parameters, which involve power low expansions [53]. The experimental and mathematical difficulties involved in BST and MCA make them very expansive to compute. Cybernetic model describes microbial responses under multiple substrates environment in synthesizing multiple proteins under optimal condition [54]. The advantage of cybernetic network over MCA and BST is that it is able to adjust to environmental perturbation; however, this approach utilized only abstract models of the network.

As understanding of biology systems grew, metabolic networks also become more sophisticated. One of the pioneering whole cell models constructed by Shuler and coworkers was to understand cell growth in glucose minimal medium environment [16]. Since then, models have been developed to study various cells under various culture conditions. Although many of those models have shown to be successful, the difficulties involved in obtaining experimental data and the complexity involved in non-linear computation make them computational expansive.

Constraint based flux balance analysis (FBA) unlike kinetic models is a linear programing method based on stoichiometric matrix, flux constrains, and an objective function [19]. FBA models have been applied to design engineered strains to overexpress metabolites through alternating metabolic flux distribution. It is also utilized to rationally develop cell culture medium [18, 55]. All the FBA applications mentioned above share a common goal- to seek an optimal condition where cells can grow under minimal stress and can enhance product formation. FBA has been used in developing bacterial strain for biofuel and pharmaceutics production [56, 57]. An expansion on FBA has been developed to incorporate transcription and translation reactions for specific DNA and protein sequences. This modeling method is called sequence specific flux balance analysis (ssFBA). Both FBA, ssFBA, and other variants of FBA have been expanded into genome-scale and can capture the regulatory effects on change in metabolism [58]. FBA has been utilized mainly to study regulatory performance in *in vivo* system. But, as mentioned in chapter one and two, the "black- box" nature of *in vivo* systems has made it challenging to understand and rewire the metabolic networks of cells. Therefore, we sought an *in vitro* platform where we could study and understand the metabolic changes under different growing conditions. In particular, we utilized ssFBA to analyze metabolite fluxes of cells.

Cell free protein synthesis (CFPS) is an *in vitro* system for expressing proteins in non-living cell environment, which allows us to bypass metabolic burden on cell growth and allows direct control of metabolites [59]. CFPS has been receiving growing attention as a tool for synthetic biological research due to its open environment without interference of cell wall and cell growth and its ability for direct metabolite control. It has been used as a tool for deciphering genetic code, a method for incorporating unnatural amino acids into proteins, and a high-throughput method for screening labeled- protein functions [60-62]. Despite CFPS has many disadvantages in terms of its yield, expenses, and reaction size, many advancements have been made to improve CFPS system. It has been considered as a novel platform for producing therapeutic proteins in industrial size and quality [63, 64]. More recently, researchers have started to couple CFPS with *in vitro* glycosylation reaction to produce glycoproteins [33]. The open environment cell free protein glycoprotein synthesis offered allows us to interrogate the chemical environment and identify the genetic manipulation needed for producing desired glycoprotein.

In Chapter two, we demonstrated an *in vitro* glycoprotein remolding platform. To further understand and analyze protein synthesis to create a platform for homogeneous designed glycoprotein production, we aimed to produce proteins using CFPS and implemented a mathematical model to analyze change in metabolites for protein expression as a proof of concept in this chapter. We chose super folded green fluorescence protein (sfGFP) as our model protein and expressed it in CFPS. We implemented our mathematical model, ssFBA, to analyze change in metabolites at different time points. A core *E. coli* metabolic model, describing glycolysis, pentose phosphate pathway, energy metabolism, amino acid biosynthesis, and degradation was implemented into the model. The incorporation of amino acid synthesis and degradation augmented FBA to describe specific transcription and translation processes with specific promoter. In particular, we studied two cases, 1) a constrained case where the bounds

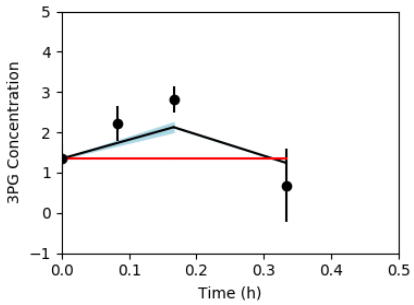
were limited by experimental data, and 2) an unconstrained case where there was no restriction set on the bounds.

3.2 Results- ssFBA on the synthesis of $GalNac_4GlcNAc$-sfGFP using CFPS following by in vitro glycoprotein synthesis

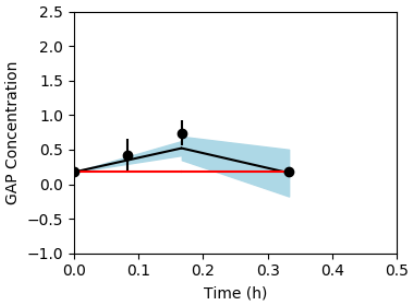
3.2.1 Validation of model predictions

Super folded green fluorescence protein (sfGFP) was expressed in CFPS under T7 promoter using PEP as the carbon and energy source. After the protein was expressed in CFPS, *in vitro* glycosylation reaction was directly performed. sfGFP was then glycosylated with ($GalNAc_4GlcNAc$) structure. Metabolites were quantified by LC-MS methods. Due to inconsistency between each sample preparation, some experimental data was optimized by removing some of those outliers. Mean and standard deviations were recalculated after optimization and those were the data used in our ssFBA model. In this study, proteins were only translated for 20 minutes in order to reach the maximum glycosylation efficiency. Samples were quenched at four different time points and at each time point, there were three duplicates.

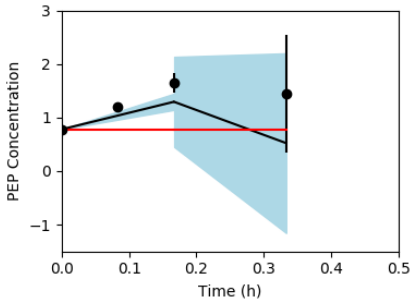
We validated our ssFBA model with the experimental measurements obtained from LC-MS. We specifically measured 20 organic acids, but only 11 of those were implemented in our ssFBA model to constrain to bounds. For the constrained case, the problem was set up to evaluate the data from 0 to 10 minutes followed by another calculation from 10 to 20 minutes based on the trends observed from the experimental data. Figure 3.1 shows a panel of 11 metabolites including experimental data, simulation for constrained and non-constrained cases. Most of the simulations for the constrained case behaved in similar trends as the experimental data. Especially, the experimental data for GAP, PEP, pyruvate, and malate fell in the 95% interval region of the simulation. However, the mathematical simulation did not correlate well with some of the metabolite, such as fumarate, glutamate, lactate and acetate.


3PG Concentration
Time (h)

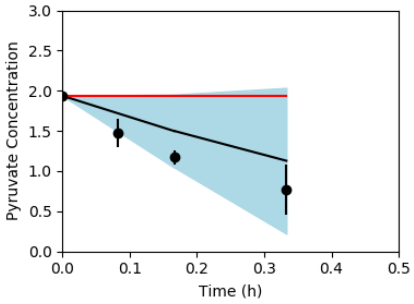
(a)


GAP Concentration
Time (h)

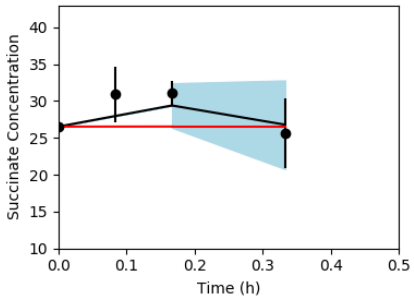
(b)


PEP Concentration
Time (h)

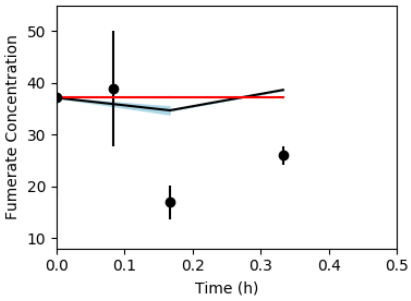
(c)


Pyruvate Concentration
Time (h)

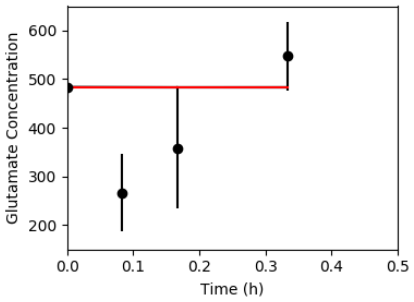
(d)


Succinate Concentration
Time (h)

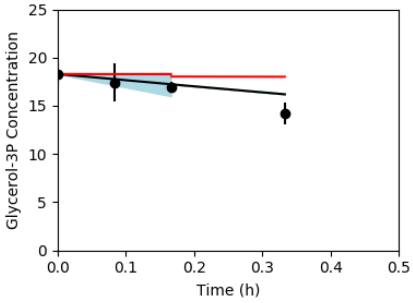
(e)


Fumerate Concentration
Time (h)

(f)


Glutamate Concentration
Time (h)

(g)


Glycerol-3P Concentration
Time (h)

(h)

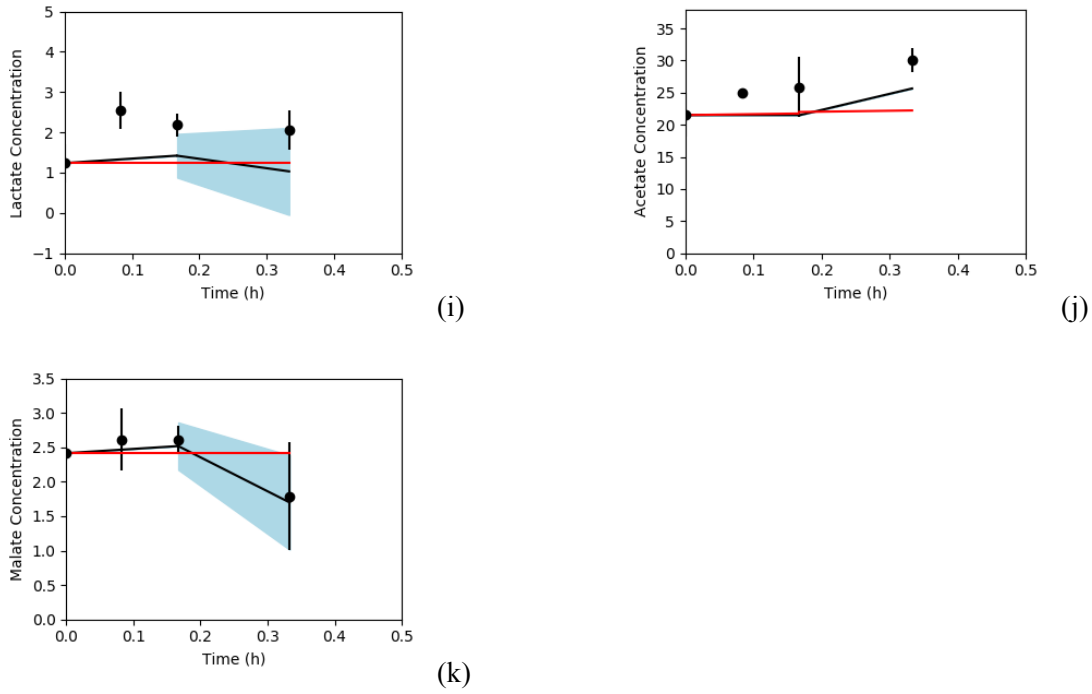


Figure 3.1 A panel of 11 metabolites concentration over 20 minutes CFPS reactions. Experimental data is denoted with black dots with error black bars. Red line denotes for the unconstrained case. Black line denotes for the constrained case. The blue shaded area represents the calculated 95% confidence interval for the constrained case. Each of the graph represents the concentration change over 20 minutes interval. From (a)-(k) is data 3GP, GAP, PEP, pyruvate, succinate, fumerate, glutamate, glycerol-3-phosphate, lactate, acetate, and malate.

3.2.2 Analysis of the CFPS performance

In order to understand the performance of our CFPS system, we analyzed the carbon yield and energy efficiency for constrained and unconstrained cases. The carbon yield for the constrained case from 0-10 minutes was 7.3% and 9.6% from 10-20 minutes, whereas the carbon yield for the unconstrained case was 96.9%. In figure 3.2, the carbon yield distribution was illustrated in pie charts format. As seen in Figure 3.2, in the unconstrained case (Figure 3.2 (a)), there was a 54.3% accumulation of sfGFP, whereas in the unconstrained case, from 0-10 minutes, there was 81% succinate accumulation; and from 10-20 minutes, there was about 52% of fumarate accumulation. This indicates that for the first 20 minutes of cell free protein synthesis, most of the carbons were spent on making organic acid byproducts.

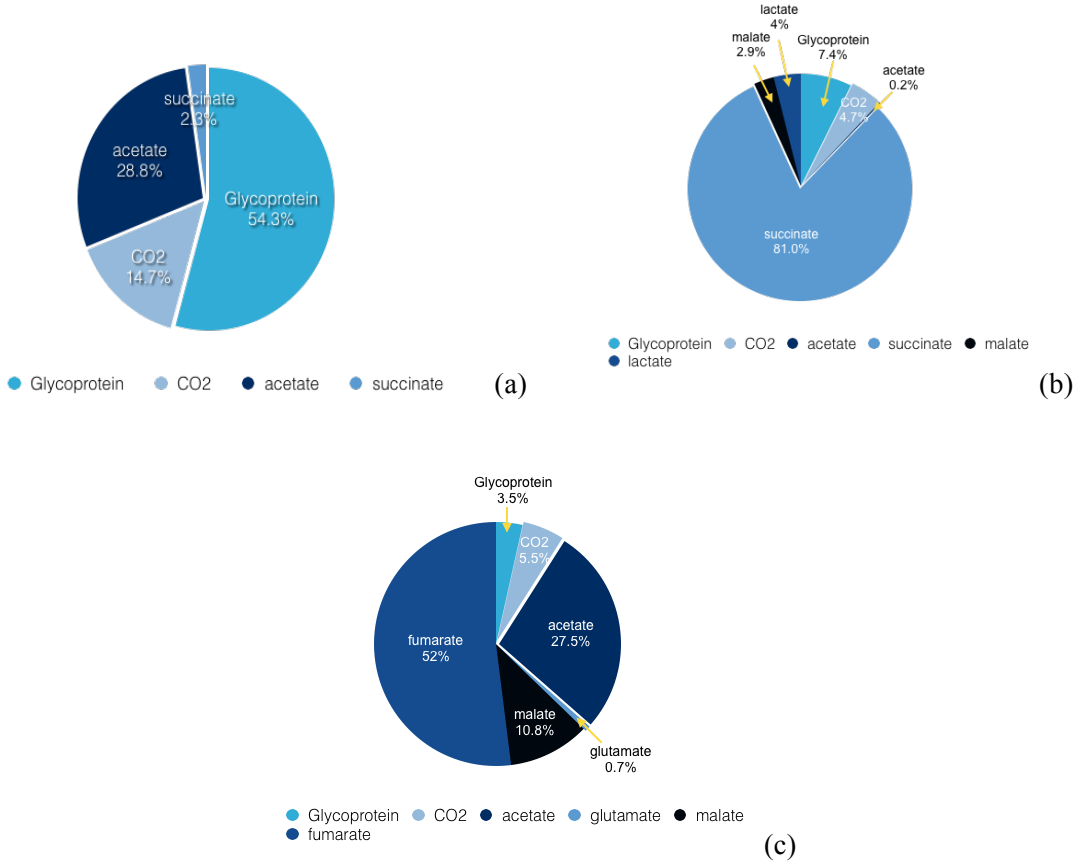


Figure 3.2 Pie charts for carbon yields. (a) is for the unconstrained case. (b) is for the experimental constrained case from 0-10 minutes. (c) is for the experimental constrained case from 10- 20 minutes.

Another parameter used to analyze the CFPS performance was energy efficiency. Energy efficiency was analyzed by different metabolite reaction groups, including glycolysis (gly), amino acids synthesis (AA), anapleoretic (ana), chorismate (cho), energy degradation (en), and protein synthesis (prot). In the unconstrained case, the maximum energy efficiency was 100%. For constrained case, the maximum energy efficiency spent on protein synthesis was about 74% ± 15% from 0 to 10 minutes and 12% ± 3.5%. In order to investigate where the energy was spent on, we then looked into the energy efficiency for each of the categories listed above. For the first 10 minutes of the reaction, 74% of energy was spent on protein synthesis and 22.8% was spent on glycolysis. But after the first 10 minutes, 68% of the energy was spent

on energy degradation, 18.7% was spent on glycolysis, and only about 12.6% was spent on protein synthesis (Figure 3.3).

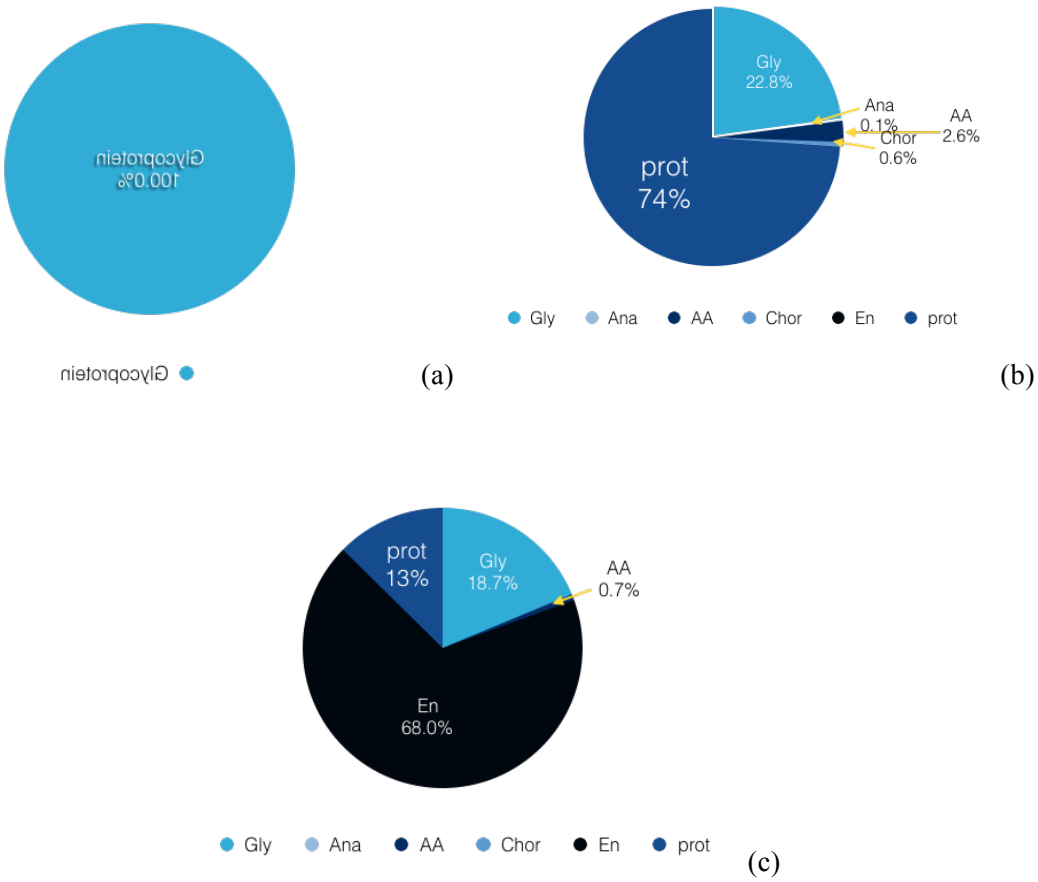


Figure 3.3 Pie charts for demonstrating the energy efficiency spent on the following metabolic reactions: glycolysis (gly), amino acids synthesis (AA), anapleoretic (ana), chorismate (cho), energy degradation (en), and protein synthesis (prot). (a) is for the unconstrained case. (b) is for experimental constrained case from 0-10 minutes. (c) is for the experimental constrained case from 10-20 minutes.

3.3 Discussion and future direction

In this chapter, we were able to synthesized sfGFP using CFPS and utilized the metabolites data to implement ssFBA model specifically describing sfGFP synthesis. For most of the metabolites, we were able to show close correlation between the experimental constrained simulation and the experimental data. However, there were some cases where the experimental data fell out of the calculated 95% interval. This can be due to experimental error, but further analysis will require to better train the model to predict sfGFP production.

Based on the analysis of the CFPS production, we noticed that most of the carbon was spent on succinate production in the first 10 minutes. And from 10 to 20 minutes, most of the carbons was spent on fumarate production. From the energy efficiency data, most of the energy was spent on protein synthesis in the first 10 minutes with a portion of energy spent on glycolysis. However, from 10 to 20 minutes, most of the energy was spent on energy degradation reactions. Usually transcription and translation will require about 10 to 12 hours. The reason we only setup our CFPS reaction for 20 minutes was based on previously observed trend by the DeLisa group that sfGFP could achieve highest glycosylation efficiency when CFPS reaction only occurred for 20 minutes. And, based on the result seen in chapter 3, further analysis can be done to further deciphering where the energy spent on and where the carbons are directed to form during different time points. This study can in the future help engineering E. coli cell lines to improve glycoprotein production.

3.4 Method

3.4.1 Cell free protein synthesis

CFPS reaction was performed using the S30 cell extract in a total volume of 15 μL. The S30 cell extract was prepared using BL21(DE3) cell line. The reaction mixture contained 1.2mM ATP, 0.85 mMof GTP, UTP, and CTP each, 34.0 μL/mL L-t-formyl-5,6,7,8 tetrahydrofolic aicd (folinic acid), 179.9 μL/mL of E. coli tRNA mixture, 130 mM potassium glutamate; 10 mM ammonium glutamate, 12 mM magnesium glutamate, 2mM of each 20 amino acids, 10 μM of L-[^{14}C(U)]-leucine, 0.33 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme-A (CoA), 1.5mM spermidine, 1mM putreschine, 4mM sodium oxalate, 33mM

phosphoenolpyruvate (PEP), 250 ng of sfGFP^{DQNAT} plasmid, 100 μg/mL T7 RNA polymerase and 27% v/v of cell extract. Cell free protein synthesis reactions were carried out at 30^{o}C for 0, 5, 10, and 20 minutes in triplicates. Translation only happened for 20 minutes was to be able to reach maximum glycosylation efficiency.

3.4.2 In vitro glycosylation reactions

After CFPS were completed, they were carried out to perform *in vitro* glycosylation at 30^{o}C for two hours. The final volume of the in vitro glycosylation reaction was 20 μL with 15 μL of CFPS, 2 μL of reaction buffer (10 mM HEPES, pH 7.5, 10 mM $MnCl_2$, and 0.1% (w/v%) DDM), and …. PglB.

3.4.3 Constrained based sequence specific flux balance analysis

Constrained based sequence specific flux balance analysis (ss- FBA) was used here to understand the performance of *E. coli* based cell free glycoprotein synthesis. This method coupes transcription and translation reactions into metabolic network. A core E. coli cell free stoichiometric network describing glycolysis, pentose phosphate pathway, energy metabolism, amino acid biosynthesis and degradation was developed from the *i*AF1260 reconstruction of K-12 MG1655 *E. coli*. This network contains 264 reactions and 146 species. In this problem, our model glycoprotein was $GalNAc_4GlcNAc$-sfGFP^{DQNAT}. FBA method requires two assumptions. First, cells must operate at pseudo steady state. Second, the cells must operate optimally to satisfy an objective. The problem then can be setup by decomposing the metabolic network into stoichiometric matrix following by linear programming problem:

$$\max_{\upsilon_1 \ldots \upsilon_R} \sum_{i=1}^{R} c_i \nu_i$$

$$S\upsilon = 0$$

$$L_i \le \sum_{j=1}^{R} S_{ij}\upsilon_j \le U_i$$

$$\alpha_j \le \upsilon_j \le \beta_j$$

where c is the objective vector, S represents the stoichiometric matrix of metabolites with each row describing one metabolite and each column describing a metabolic reaction, and ν denotes metabolic flux vector where α and β are the lower and upper bounds for the individual flux values. The problem was setup to maximize the rate of protein translation. In stoichiometric matrix, each element is denoted by σ_{ij}. When $\sigma_{ij}>0$, species i is produced by reaction j, whereas when $\sigma_{ij}<0$, species i is consumed by reaction j. In the case when $\sigma_{ij} = 0$, species i does not participate in reaction j. To account for transcription in the network, additional constraints on transcription and translation are required. The bounds on the transcription rate was modeled as:

$$\omega_x = V_x^{\max}\left(\frac{G_p}{K_x + G_p}\right)$$

where ω_x is the transcription rate; G_p denotes the concentration of the gene encoding the protein of interest; and, K_x denotes the transcription saturation coefficient. The maximum transcription rate V_x^{max} was calculated by:

$$V_x^{\max} = [R_T(\frac{V_x}{I_G})u(k))]$$

where R_T denotes the RNA polymerase concentration (nM), V_x denotes the RNA polymerase elongation rate (nt/hour), and I_G is the gene length (nt). The effective function of promoter activity, u(k) (dimensionless, $0<u(k)<1$) was taken from Moon *et al*, where k denotes promoter specific parameters. In this study we utilized T7 promoter. The promoter function for the T7 promoter, u_{T7}, was given by:

$$u_{T7} = \frac{K_{T7}}{1 + K_{T7}}$$

where K_{T7} denotes T7 RNA polymerase binding constant.

The translation rate (ω_T) was bounded by:

$$0 \le \omega_T \le V_T^{\max}\left(\frac{mRNA^*}{K_T + mRNA^*}\right)$$

where $mRNA^*$ represents the mRNA aboundance, and K_T represents a translationa saturation constant. The maximum translation rate V_T^{max} was calculated by:

$$V_T^{\max} = K_P R_T\left(\frac{\upsilon_T}{l_p}\right)$$

where K_P denotes the polysome amplification constant, ν_T denotes the ribosome elongation rate (amino acid per hour), and l_P denotes the number of amino acids in the protein of interest.

3.4.4 Calculation of energy efficiency

Energy efficiency (E) is calculated as the ratio of protein production to glucose consumption. They were formulated in terms of equivalent ATP molecules:

$$E = \frac{\omega_X(2 \times ATP_X + GTP_X) + \omega_T(ATP_T + GTP_T + CTP_T + UTP_T)}{q_{GLC} \times ATP_{GLC}}$$

where ATP_T, GTP_T, CTP_T, UTP_T represent the stoichiometric coefficients of energy species for the transcription of the protein of interest; ATP_X, GTP_X represent the stoichiometric coefficients of ATP and GTP for the translation of the protein of interest, q_{GLC} denotes for glucose uptake rate, and ATP_{GLC} denotes the equivalent ATP number produced per glucose.

3.4.5 Calculation of carbon yield

Carbon yield (Y_C^{POI}) was calculated as the ratio of carbon produced as the protein of interest divided by the carbon consumed as reactants (glucose and amino acids):

$$Y_C^{POI} = \frac{q_{POI} * C_{POI}}{\sum_{i=1}^{R} q_{mi} \times C_{mi}}$$

where q_{POI} denotes the flux of the protein of interest produced, C_{POI} denotes carbon number of the protein of interest, R denotes the number of reactants, q_{mi} denotes the uptake flux of the i^{th} reactant, and C_{mi} denotes the carbon number of the i^{th} reactant.

3.4.6 Calculation of maximum and minimum bound for fluxes

At each time point, a minimum and a maximum were obtained by subtracting or adding the standard deviation to the actual data point:

$$\max = [metabolite] + stdv.$$
$$\min = [metabolite] - stdv.$$

The upper bound and lower bound of each metabolite was calculated based on the following equations:

$$lower\ bound = \frac{y_{2,min} - y_{1,max}}{x_{2,min} - x_{1,max}}$$

$$upper\ bound = lower\ bound + (\frac{y_{2,max} - y_{1,min}}{x_{2,max} - x_{1,min}} - lower\ bound) \times random$$

where x represents time and y represents metabolite concentration.

3.4.4 Quantification of organic acids

The protocol for quatifying organic acid was taken from ….(paper) with some modification on cone voltages.

LC-MS system:

The LC-MS system consisted of a UPLC system (Acquity H-Class, Waters) and a source mass spectrometer (QDA detector, Waters). This system was controlled by Empower 3 software (Waters). Seperation was performed on a Acquity CSH C18 Column (2.1mm x 150 mm, Waters).

Solvent A: 5mM tert- butylamine (TBA) aqueous solution adjust pH to 5.0 with acetic acid.

Solvent B: 5 mM TBA in 100% acetonitrile.

1. Light labeling reagent: 6M $^{12}C_6$-aniline solution (pH 4.5) by mixing 5.500 mM of aniline with 2.250 mL of water and 2.250 mL ofconcentrated hydrochloric acid (HCl). Vortex well and measure pH. If ecessary adjust pH with HCl. The solution can be stored at 4^{o}C for up to 2 months.

2. EDC solution: N-(3-dimethylaminopropyl)-N'- ethylcarbodiimide hydrochloric (200mg/mL) dissolved in Milli-Q water.

3. Triethylamine (TEA)

Sample solution consisted 5 μL of quenched CFgPS reaction and 95 μL of autoclaved H_2O in 1.5 mL Eppendorf tube. 10 μL of EDC solution was added into the sample solution. Vortexed aniline solution until it was well mix before adding 10 μL into the sample preparation. Reaction mixture was incubated on rocker at low speed at room temperature for 2 hours. Then, 2-5 μL of TEA was added to the sample solution and centrifuged at 13,500xg for 2 min. Transferred the aqueous portion of the sample into an auto- sampler vial.

LC- MS setup:

Starting up by setting an elution gradient from 5- 70% Solvent B over 4 min, further increase to 100% Solvent B in 0.5 min, and held at 100% Solvent B for 2.5 min. Flowrate was set at 0.3 mL/ min with injection volume of 1 μL. Column was preconditioned by pumping the starting mobile phase mixture for 10 min. LC-MS chromatograms were acquired in negative ion mode under the following conditions: capillary voltage of 15 V, dry temperature at 520 ^{o}C, and an acquisition range of m/z 100-800.

3.4.5 Quantification of amino acids

A solution containing a mixture of 19 amino acids is drivatized with Waters AccQ-Tawq Ulta amino acid analysis kit (Waters). Sample was prepared by adding 80 μL amino acids (Amino Acid H- Standard, Waters) with 10 μL of buffer solution (Waters) followed by 10 μL of reagent (Waters). The reaction mixture was incubated at 55^{o}C fo 10 min. Later the sample was separated by UPLC with an Acquity Amide C18 Column (2.1 mm x 150 mm, Waters) and analyzed with a TUV detector at 260 mm. Amino acid s were detected and quantified based on known retention times.

REFERENCE

1. Nothaft, H. and C.M. Szymanski, *Protein glycosylation in bacteria: sweeter than ever.* Nature Reviews. Microbiology, 2010. **8**(11): p. 765-78.
2. Kowarik, M., et al., *N-Linked Glycosylation of Folded Proteins by the Bacterial Oligosaccharyltransferase.* Science, 2006. **314**(5802): p. 1148.
3. Wacker, M., et al., *N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into <em>E. coli</em&gt.* Science, 2002. **298**(5599): p. 1790.
4. Croset, A., et al., *Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells.* Journal of Biotechnology, 2012. **161**(3): p. 336-348.
5. Sethuraman, N. and T.A. Stadheim, *Challenges in therapeutic glycoprotein production.* Current Opinion in Biotechnology, 2006. **17**(4): p. 341-346.
6. Brorson, K. and A.Y. Jia, *Therapeutic monoclonal antibodies and consistent ends: terminal heterogeneity, detection, and impact on quality.* Current Opinion in Biotechnology, 2014. **30**: p. 140-146.
7. Ecker, D.M., S.D. Jones, and H.L. Levine, *The therapeutic monoclonal antibody market.* mAbs, 2015. **7**(1): p. 9-14.
8. Solá, R.J. and K. Griebenow, *Glycosylation of Therapeutic Proteins: An Effective Strategy to Optimize Efficacy.* BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy, 2010. **24**(1): p. 9-21.
9. Liu, L., *Antibody Glycosylation and Its Impact on the Pharmacokinetics and Pharmacodynamics of Monoclonal Antibodies and Fc-Fusion Proteins.* Journal of Pharmaceutical Sciences, 2015. **104**(6): p. 1866-1884.
10. Reusch, D. and M.L. Tejada, *Fc glycans of therapeutic antibodies as critical quality attributes.* Glycobiology, 2015. **25**(12): p. 1325-1334.
11. Zhu, J., *Mammalian cell protein expression for biopharmaceutical production.* Biotechnology Advances, 2012. **30**(5): p. 1158-1170.
12. Ollis, A.A., et al., *Engineered oligosaccharyltransferases with greatly relaxed acceptor site specificity.* Nature chemical biology, 2014. **10**(10): p. 816-822.
13. Valderrama-Rincon, J.D., et al., *An engineered eukaryotic protein glycosylation pathway in Escherichia coli.* Nature chemical biology, 2012. **8**(5): p. 434-436.
14. Ihssen, J., et al., *Production of glycoprotein vaccines in Escherichia coli.* Microbial Cell Factories, 2010. **9**: p. 61-61.
15. Hansen, A.S.L., et al., *Systems biology solutions for biochemical production challenges.* Current Opinion in Biotechnology, 2017. **45**: p. 85-91.
16. M., D.M., et al., *Computer model for glucose-limited growth of a single cell of Escherichia coli B/r-A.* Biotechnology and Bioengineering, 1984. **26**(3): p. 203-216.
17. J., V. and R. D., *Metabolic Engineering from a Cybernetic Perspective. 1. Theoretical Preliminaries.* Biotechnology Progress, 1999. **15**(3): p. 407-425.
18. Lee, J.M., E.P. Gianchandani, and J.A. Papin, *Flux balance analysis in the era of metabolomics.* Briefings in Bioinformatics, 2006. **7**(2): p. 140-150.

19. Orth, J.D., I. Thiele, and B.Ø. Palsson, *What is flux balance analysis?* Nature Biotechnology, 2010. **28**: p. 245.
20. Carlson, E.D., et al., *Cell-free protein synthesis: Applications come of age.* Biotechnology Advances, 2012. **30**(5): p. 1185-1194.
21. Pardee, K., et al., *Portable, On-Demand Biomolecular Manufacturing.* Cell, 2016. **167**(1): p. 248-259.e12.
22. Zawada, J.F., et al., *Microscale to Manufacturing Scale-up of Cell-Free Cytokine Production—A New Approach for Shortening Protein Production Development Timelines.* Biotechnology and Bioengineering, 2011. **108**(7): p. 1570-1578.
23. Hong, S.H., Y.-C. Kwon, and M.C. Jewett, *Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis.* Frontiers in Chemistry, 2014. **2**: p. 34.
24. Burda, P. and M. Aebi, *The dolichol pathway of N-linked glycosylation.* Biochimica et Biophysica Acta (BBA) - General Subjects, 1999. **1426**(2): p. 239-257.
25. Oliveira-Ferrer, L., K. Legler, and K. Milde-Langosch, *Role of protein glycosylation in cancer metastasis.* Seminars in Cancer Biology, 2017. **44**: p. 141-152.
26. *Glycosylation and Disease*, in *eLS*.
27. Igura, M., et al., *Structure-guided identification of a new catalytic motif of oligosaccharyltransferase.* The EMBO Journal, 2008. **27**(1): p. 234-243.
28. Szymanski, C.M. and B.W. Wren, *Protein glycosylation in bacterial mucosal pathogens.* Nature Reviews Microbiology, 2005. **3**: p. 225.
29. Li, B., et al., *A Highly Efficient Chemoenzymatic Approach toward Glycoprotein Synthesis.* Organic Letters, 2006. **8**(14): p. 3081-3084.
30. Li, B., et al., *Highly Efficient Endoglycosidase-Catalyzed Synthesis of Glycopeptides Using Oligosaccharide Oxazolines as Donor Substrates.* Journal of the American Chemical Society, 2005. **127**(27): p. 9692-9693.
31. Enrica, C., et al., *Chemoselective Glycosylation of Peptides through S-Alkylation Reaction.* Chemistry – A European Journal, 2018. **24**(23): p. 6231-6238.
32. Chong, S., *Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications.* Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.], 2014. **108**: p. 16.30.1-16.30.11.
33. A., S.J., et al., *A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases.* Biotechnology and Bioengineering, 2018. **115**(3): p. 739-750.
34. Isaksson, L., et al., *Expression screening of membrane proteins with cell-free protein synthesis.* Protein Expression and Purification, 2012. **82**(1): p. 218-225.
35. Rita, S., et al., *Synthesis of membrane proteins in eukaryotic cell-free systems.* Engineering in Life Sciences, 2013. **13**(1): p. 39-48.
36. Lewis, N.E., H. Nagarajan, and B.O. Palsson, *Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods.* Nature Reviews Microbiology, 2012. **10**: p. 291.

37. Schilling, C.H., D. Letscher, and B.Ø. Palsson, *Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective.* Journal of Theoretical Biology, 2000. **203**(3): p. 229-248.
38. Dube, D.H. and C.R. Bertozzi, *Glycans in cancer and inflammation — potential for therapeutics and diagnostics.* Nature Reviews Drug Discovery, 2005. **4**: p. 477.
39. Imperiali, B. and S.E. O'Connor, *Effect of N-linked glycosylation on glycopeptide and glycoprotein structure.* Current Opinion in Chemical Biology, 1999. **3**(6): p. 643-649.
40. von Horsten, H.H., et al., *Production of non-fucosylated antibodies by co-expression of heterologous GDP-6-deoxy-d-lyxo-4-hexulose reductase.* Glycobiology, 2010. **20**(12): p. 1607-1618.
41. Naoko, Y.O., 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.* Biotechnology and Bioengineering, 2004. **87**(5): p. 614-622.
42. Lalonde, M.-E. and Y. Durocher, *Therapeutic glycoprotein production in mammalian cells.* Journal of Biotechnology, 2017. **251**: p. 128-140.
43. Yang, Q. and L.-X. Wang, *Chapter Nine - Chemoenzymatic Glycan Remodeling of Natural and Recombinant Glycoproteins*, in *Methods in Enzymology*, B. Imperiali, Editor. 2017, Academic Press. p. 265-281.
44. Johnson, C.P., et al., *Direct Evidence That Neural Cell Adhesion Molecule (NCAM) Polysialylation Increases Intermembrane Repulsion and Abrogates Adhesion.* Journal of Biological Chemistry, 2005. **280**(1): p. 137-145.
45. Keys, T.G., et al., *A biosynthetic route for polysialylating proteins in Escherichia coli.* Metabolic Engineering, 2017. **44**: p. 293-301.
46. Royston, J., *Glycosylation of Recombinant Antibody Therapeutics.* Biotechnology Progress, 2005. **21**(1): p. 11-16.
47. Hamilton, B.S., et al., *A library of structurally homogeneous human N-glycans synthesized from microbial oligosaccharide precursors.* bioRxiv, 2017.
48. Chiu, C.P.C., et al., *Structural analysis of the sialyltransferase CstII from Campylobacter jejuni in complex with a substrate analog.* Nature Structural &Amp; Molecular Biology, 2004. **11**: p. 163.
49. El Maarouf, A., et al., *Enzymatic Engineering of Polysialic Acid on Cells in Vitro and in Vivo Using a Purified Bacterial Polysialyltransferase.* Journal of Biological Chemistry, 2012. **287**(39): p. 32770-32779.
50. Glasscock, C.J., et al., *A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation.* Metabolic Engineering, 2018. **47**: p. 488-495.
51. Bailey, J., *Toward a science of metabolic engineering.* Science, 1991. **252**(5013): p. 1668-1675.
52. Monod, J., *The Growth of Bacterial Cultures.* Annual Review of Microbiology, 1949. **3**(1): p. 371-394.

53. Fell, D.A., *Metabolic control analysis: a survey of its theoretical and experimental development.* Biochemical Journal, 1992. **286**(Pt 2): p. 313-330.
54. Doshi, P., R. Rengaswamy, and K.V. Venkatesh, *Modelling of microbial growth for sequential utilization in a multisubstrate environment.* Process Biochemistry, 1997. **32**(8): p. 643-650.
55. Edwards, J.S., R.U. Ibarra, and B.O. Palsson, *In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data.* Nature Biotechnology, 2001. **19**: p. 125.
56. Atsumi, S., T. Hanai, and J.C. Liao, *Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels.* Nature, 2008. **451**: p. 86.
57. Ro, D.-K., et al., *Production of the antimalarial drug precursor artemisinic acid in engineered yeast.* Nature, 2006. **440**: p. 940.
58. Allen, T.E. and B.Ø. Palsson, *Sequence-Based Analysis of Metabolic Demands for Protein Synthesis in Prokaryotes.* Journal of Theoretical Biology, 2003. **220**(1): p. 1-18.
59. Zemella, A., et al., *Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems.* Chembiochem, 2015. **16**(17): p. 2420-2431.
60. Kiyoshi, O., D.N. E., and O. Gottfried, *Cell-free synthesis of 15N-labeled proteins for NMR studies.* IUBMB Life, 2005. **57**(9): p. 615-622.
61. Patel, K.G. and J.R. Swartz, *Surface Functionalization of Virus-Like Particles by Direct Conjugation Using Azide−Alkyne Click Chemistry.* Bioconjugate Chemistry, 2011. **22**(3): p. 376-387.
62. Nirenberg, M.W. and J.H. Matthaei, *THE DEPENDENCE OF CELL- FREE PROTEIN SYNTHESIS IN E. COLI UPON NATURALLY OCCURRING OR SYNTHETIC POLYRIBONUCLEOTIDES.* Proceedings of the National Academy of Sciences of the United States of America, 1961. **47**(10): p. 1588-1602.
63. Zichel, R., et al., *Efficacy of a Potential Trivalent Vaccine Based on Hc Fragments of Botulinum Toxins A, B, and E Produced in a Cell-Free Expression System.* Clinical and Vaccine Immunology, 2010. **17**(5): p. 784-792.
64. Swartz, J., *Developing cell-free biology for industrial applications.* Journal of Industrial Microbiology and Biotechnology, 2006. **33**(7): p. 476-485.