EXPLORATION OF MICROFLUIDIC TECHNOLOGIES AND EUKARYOTIC OLIGOSACCHARYLTRANSFERASES TO IMPROVE *ESCHERICHIA COLI*-BASED GLYCOPROTEIN PRODUCTION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Alicia Kimberly Aquino

August 2022

EXPLORATION OF MICROFLUIDIC TECHNOLOGIES AND EUKARYOTIC OLIGOSACCHARYLTRANSFERASES TO IMPROVE *ESCHERICHIA COLI*-BASED GLYCOPROTEIN PRODUCTION

Alicia Aquino, Ph. D.

Cornell University 2022

A key feature in manufacturing biologics is glycosylation, a posttranslational modification where a glycan is attached to amino acid side chains of proteins. Glycosylation gives rise to diverse chemical structures that affect drug properties such as stability, activity, and immunogenicity. Traditionally, glycoproteins are produced in eukaryotic cells where glycosylation is not template based, giving rise to heterogeneous mixtures of glycoforms that are difficult to parse out. In response to this challenge, cell-free technologies have emerged that enable production of glycoproteins outside of the cell. In my dissertation work, I address two bottlenecks of cell-free production of glycoproteins. First, current cell-free technologies use a "one-pot" system where integration of complex reaction pathways can lead to competing reactions and generation of side products. In my work, I have integrated cell-free technologies into a microfluidic system where individual bioprocesses are compartmentalized to enable greater control and optimization at each step. Second, to expand the range of human-like glycoproteins that can be produced, I examine single subunit OSTs from eukaryotic sources for use in *E. coli* that I believe will be amenable to cell-free systems. Together, these advancements bring us closer to the ability to perform bottom-up construction of designer, structurally uniform glycoproteins for study and use as therapeutics.

BIOGRAPHICAL SKETCH

Alicia Aquino was born in Massachusetts to Heesoon Kim and Francisco Aquino. She is one of two siblings and has an older brother, Alexander Aquino. She grew up in Marlborough, MA, graduating high school in 2013. She then attended Worcester Polytechnic Institute (WPI), majoring in Chemical Engineering with a minor in Nanoscience. During her freshman year at WPI, Alicia met her now fiancé, Norbert Hugger. Starting in 2014, she participated in undergraduate research in Terri Camesano's Lab, where she worked with Lindsay Lozeau on developing delivery methods for antimicrobial peptides. Through this experience, Alicia became interested in research and entrepreneurship and was inspired to pursue a PhD. With help from Lindsay and Professor Camesano, Alicia applied and was awarded an NSF Graduate Research Fellowship for 2017-2022. After graduating WPI with a B.S. in Chemical Engineering, she joined Cornell University as a PhD student in Chemical and Biomolecular Engineering in 2017. There, she joined the lab of Matthew DeLisa. During her time at Cornell, Alicia was awarded an NIH Chemistry-Biology Interface training grant and the Austin Hooey Graduate Research Excellence Award. She was also involved in outreach activities in the department as a producer in the Science Blender podcast and on the CBE Women board.

DEDICATION

This work is dedicated to

my mother, father, and fiancé for their continuous support

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Matt DeLisa, for his guidance and support. His mentorship has been instrumental in my growth and achievements as a scientist during my PhD. Additionally, I must thank everyone in the DeLisa research group who have been a source of both emotional and intellectual support. I thank Tommy Jaroentomeechai for getting me started in the lab and training me in all the basic molecular biology techniques that I have relied on for my dissertation work. I thank Alexandra Pang who has worked with me in the lab since she was in high school and now as an undergraduate student at Cornell. I also thank Sean Chung, Kevin Weyant, and Jinjoo Jung for always taking the time to chat with me about research or other topics. Additionally, I would like to thank current and former members of the glyco-subgroup May Taw, Xiaolu Zheng, Aravind Natarajan, Daniel Kwon, Belen Sotomayor, Sophia Hulbert, Erik Bidstrup, Jiazhen Chen, and Keehun Kim for their support and input. Outside of the DeLisa lab, I would also like to thank the old Golgi buddies Zach Manzer, Ferra Pinnock, Han-Yuan Liu, and Rohit Singh from the Daniel lab who played a large role in my transition into the PhD student life during my first year. Thank you as well to everyone that I have worked with in CBE Women and Science Blender over the years.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH iv
DEDICATION v
ACKNOWLEDGMENTS vi
1 INTRODUCTION 1
1.1 The basics of protein glycosylation 1
1.1.1 Cellular compartmentalization of glycosylation in eukaryotes 1
1.1.2 Protein glycosylation in bacteria 3
1.1.3 Glycoprotein therapeutics 6
1.2 Engineering *E. coli* to produce glycoproteins 7
1.2.1 Transfer of *C. jejuni* protein glycosylation system into *E. coli* 8
1.2.2 Glycosylation pathway remodeling in *E. coli* 9
1.3 Cell-free technologies for glycoproteins synthesis 10
1.4 Applications of microfluidics in glycan and protein synthesis 14
1.4.1 Synthesis of glycans in microfluidics 15
1.4.2 Protein production in microfluidics 17
1.5 Single subunit eukaryotic OSTs 18
1.5.1 Kinetoplastid STT3s do not integrate into an OST complex 19
1.5.2 Glycan and protein specificities of kinetoplastid STT3s 21
1.5.3 Kinetoplastid STT3s for cell-based and cell-free glycoprotein production 23
1.6 Summary 25
References 26
2 GLYCOSYLATION-ON-A-CHIP: A FLOW-BASEDMICROFLUIDIC SYSTEM FOR CELL-FREE GLYCOPROTEIN BIOSYNTHESIS 34
2.1 Introduction 34
2.2 Results 38
2.2.1 Design of a modular microfluidic platform for continuous glycoprotein production 38
2.2.2 Continuous-flow CFPS module improves protein production 40
2.2.3 Tethered OST enzyme enables a continuous-flow glycosylation module .. 45
2.2.4 IMAC module enables continuous enrichment of product proteins 53
2.3 Discussion 57
2.4 Materials and Methods 60

2.5 Acknowledgements 68
References 70
3 KINETOPLASTID SINGLE SUBUNIT OSTS ENABLE TRANSFER OF EUKARYOTIC *N*-GLYCANS IN *E. COLI* 74
3.1 Introduction 74
3.2 Results 79
3.2.1 Analysis of the eukaryotic single subunit STT3 panel 79
3.2.2 Identification of eukaryotic STT3s that transfer $Man_3GlcNAc_2$ from UndPP to a DQNAT consensus sequon 83
3.2.3 Top hits from the eukaryotic single subunit OST activity screen exhibit preference for negatively charged amino acids in the -2 position 85
3.3 Discussion 87
3.4 Materials and Methods 89
3.5 Acknowledgements 93
References 94
4 FUTURE DIRECTIONS 97
4.1 Introduction 97
4.2 Discussion 98
4.2.1 Further exploration of microfluidic devices for glycoprotein synthesis 98
4.2.2 Continued exploration of single subunit eukaryotic STT3s 99
4.2.3 Integration of eukaryotic STT3s into cell-free systems 100
References 101

1 INTRODUCTION

1.1 The basics of protein glycosylation

A key feature in many proteins produced by the cell is glycosylation, a post-translational modification where a complex sugar group, called a glycan, is attached to amino acid side chains of proteins[1]. The two most common sites for attachment are at the nitrogen atom of the asparagine residue in *N*-linked glycosylation and at the oxygen atom of a serine or threonine residue in *O*-linked glycosylation[2]. These glycans are composed of a variety of sugar monomers that are connected by glycosidic linkages that, when linked together, form diverse and complex branching structures. Because of the complexity of glycans, glycosylation confers unique properties to proteins that modify and expand their functionalities. These structural and functional changes have important implications in health and disease and by understanding the parts and pieces of how protein glycosylation works, we can better engineer glycoproteins with enhanced therapeutic properties. This dissertation work specifically focuses on investigating and engineering *N*-linked glycosylation in bacterial and cell-free systems

1.1.1 Cellular compartmentalization of glycosylation in eukaryotes

In eukaryotic cells, *N*-linked glycoprotein biosynthesis occurs through a series of enzymatic reactions that are spatially and temporally controlled in the endoplasmic reticulum (ER) and the Golgi apparatus[3]. In the ER, the glycan structure is built on a lipid carrier from the bottom up at the cytosolic side of the organelle by enzymes called glycosyltransferases (GTs). The glycan is then flipped into the lumen of the ER where it is further modified by GTs and added to a protein by the oligosaccharyltransferase

(OST). Specifically, the OST catalyzes the transfer of an oligosaccharide from a lipid carrier to the asparagine residue of a protein at a minimal amino acid consensus sequence defined as N-X-S/T, where X is any amino acid except for proline. In higher eukaryotes, the glycan structure that is attached to the protein with the highest specificity is glucose$_3$mannose$_9$*N*-acetylglucosamine$_2$ ($Glc_3Man_9GlcNAc_2$), as shown in **Figure 1.1a**. Once attached, the glycan structure guides quality control mechanisms that ensure proper folding of the protein within the ER. Specifically, trimming of two Glc residues at the non-reducing end of the glycan by glycosidases enables binding of the glycoprotein to lectins that serve as chaperones that aid in protein folding. Removal of the third Glc residue then releases the glycoprotein from the chaperones. If the glycoprotein is still not properly folded and remains in the ER for a long enough time, two Man residues are eventually cleaved from the glycan which acts as a signal to send the misfolded protein to the proteosome for degradation. Properly folded glycoproteins, usually with a $Man_9GlcNAc_2$ or $Man_8GlcNAc_2$ glycan attached, are trafficked to the Golgi apparatus.

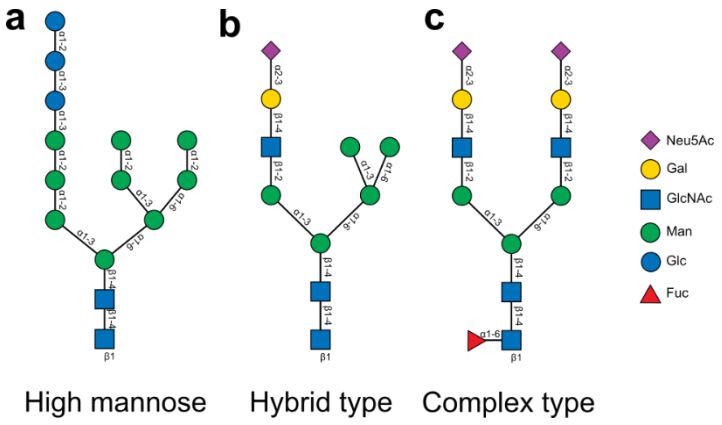


Figure 1.1. *N*-glycan types. a) High mannose, **b)** hybrid type, and c) complex type glycan structures, including linkage positions. Images were produced using GlycoGlyph.

The Golgi apparatus consists of three cisternae— cis, medial, and trans— that contain different glycosylation enzymes. The first stop for the glycoprotein after leaving the ER is the cis-Golgi where the high mannose structures are trimmed down to $Man_5GlcNAc_2$. The glycoprotein is then shuttled to the medial-Golgi where complex type glycans are synthesized. Notably, a GlcNAc residue must be added to the α1-3Man branch before the two Man residues attached to the α1-6Man branch can be trimmed and remodeled with GlcNAc. The resulting $GlcNAc_2Man_3GlcNAc_2$ structure is the simplest glycan defined as a complex type glycan. Failure to remove the two Man residues of the α1-6Man branch will result in hybrid type glycan structures such as the one shown in **Figure 1.1b**. Further complicating the glycan structure, additional GlcNAc residues can be installed on both α-Man branches as well as at the β-Man of the core in a position called the bisecting GlcNAc. These glycans will then be further processed by addition of a fucose (Fuc) residue to the core of the glycan structure in the medial-Golgi and elongation of the GlcNAc residues by attachment of galactose (Gal) and sialic acid residues— *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc)— in the trans-Golgi, to synthesize a mature complex type glycan like the one shown in **Figure 1.1c**. Depending on the organism, the types of glycan structures, linkages of the sugar monomers, and variety of sugars present will vary[4].

1.1.2 Protein glycosylation in bacteria

Although, historically, protein glycosylation was thought to only occur eukaryotes, it is now evident that protein glycosylation also occurs in both bacteria[5] and

archaea[6]. Prokaryotic cells do not have organelles, so instead, protein glycosylation is localized at the cytoplasmic membrane and occurs through a series of enzymatic reactions. Similar to what occurs in the ER of eukaryotic cells, in prokaryotes the glycan is built up on a lipid carrier at the cytoplasmic side of the membrane and then flipped to the other side of the membrane where it is accessible to the active site of the OST. In gram-negative bacteria, this side of the membrane is the periplasm and is located between the inner and outer membrane. Additional differences are the lipid carrier and glycan structures among the three domains of life. In eukaryotes and archaea, the lipid carrier is usually dolichol pyrophosphate (DolPP), but in bacteria, the lipid carrier is undecaprenol pyrophosphate (UndPP)[7]. In eukaryotes, *N*-glycans all contain a core $Man_3GlcNAc_2$, but this is not present in prokaryotes.

In 1999, Szymanski et al. were the first to report evidence for a general system for N-linked protein glycosylation in bacteria[8]. In this work, a genetic locus from *Campylobacter jejuni* was systematically mutated to show that deletion of a portion of these genes resulted in changes to glycoproteins (flagellin), but not to surface glycolipids (lipopolysaccharide). This indicated that this set of genes was involved in protein glycosylation. Since then, the *C. jejuni* system has become the most well-studied bacterial protein glycosylation pathway. In 2002, mass and NMR spectrometry were used to determine the structure of the *C. jejuni* glycan to be *N*-acetylgalactosamine$(GalNAc)_5$(Glc)bacillosamine(Bac)[9]. The identity and functions of the enzymes encoded in the protein glycosylation locus, *pgl*, that synthesize the $GalNAc_5$(Glc)Bac structure were then determined in 2005-2006[10–12]. These studies

showed that Bac is synthesized by the enzymes PglD, PglE, and PglF and attached to UndPP on the cytoplasmic side of the membrane by PglC. Five Gal residues are then added to Bac by PglA, PglJ, and PglH and Glc is added to the third Gal by PglI before the heptasaccharide is flipped to the periplasmic side of the membrane by PglK, as shown in **Figure 1.2**. In the periplasm, the glycan is then transferred to the asparagine residue of a protein by the OST, PglB. Significantly from an engineering standpoint, in 2002 the *C. jejuni N*-glycosylation system was reconstituted in *E. coli*[13], where it has been used to synthesize a variety of recombinant glycoprotein structures for study and biotherapeutic uses. This topic will be described in more detail in **Section 1.2**.

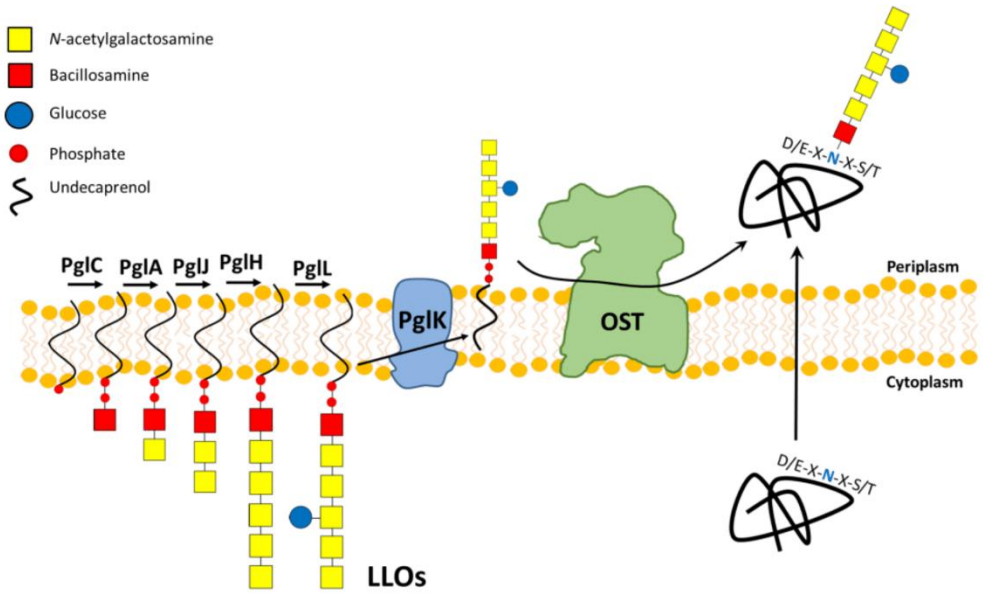


Figure 1.2. *N*-glycosylation system in *Campylobacter jejuni*. The *C. jejuni* heptasaccharide is built on undecaprenol pyrophosphate by enzymes in the cytoplasm, then flipped into the periplasm where the glycan is transferred to the protein by the oligosaccharyltransferase, PglB. Figure made by DeLisa lab.

1.1.3 Glycoprotein therapeutics

Aside from the role of glycans in biology, the glycosylation pattern is one the most important criteria for developing a safe and effective therapeutic glycoprotein[14]. Because of the diversity and complexity of the glycan structure, protein glycosylation gives rise to diverse chemical structures that affect glycoprotein drug properties such as stability, biological activity, and immunogenicity[15–18]. An early example of how glycosylation patterns affect drug properties comes from studies of erythropoietin (EPO), a glycoprotein therapeutic used for the treatment of anemia. In 1991, Wasley et al. demonstrated that EPO with incompletely processed *N*-linked glycans exhibited a 500-fold decrease in activity *in vivo* compared to EPO where no changes were made to the glycosylation patterns[19]. They postulated that this decrease in activity could be due to both the increased rate of clearance as well as a positive effect of EPO *N*-glycosylation on glycoprotein activity. A decade later, a glycoengineered hyperglycosylated recombinant human EPO, darbepoetin alfa, was developed and patented by Amgen Inc that increased the serum half-life and potency of EPO[19]. This improvement was made by engineering two additional glycosylation sites into EPO that increased both the size for the glycoprotein which reduced serum clearance by kidney filtration and the sialic acid content which reduced clearance of the glycoprotein by asialoglycoprotein receptors.

Another important and well-studied example of how glycans affect glycoprotein drug properties is in the immunoglobulin G (IgG) fragment crystallizable (Fc) region. Monoclonal antibodies (mAb) are one of the fastest growing biologics in past years for

the treatment of cancers and autoimmune diseases[20]. In the antibody, the *N*-glycan on the Fc region is an important modulator of antibody functionality. For example, sialic acid capping at the non-reducing end of the glycan improves circulation time of the drug, while removal of the core-fucose enhances antibody binding with the Fc receptor of immune cells which in turn will boost cell-killing through a process called antibody-dependent cell-mediated cytotoxicity (ADCC)[17]. Additionally, when using non-human mammalian hosts such as Chinese Hamster Ovary (CHO) cells to produce glycoprotein therapeutics, non-human glycan motifs such as galactose-alpha-1,3-galactose (alpha-Gal) and *N*-glycolylneuraminic acid (Neu5Gc) can be synthesized on the glycoprotein which produce immunogenetic effects[21]. These examples highlight the importance of glycans in biopharmaceutical industry and the ability to precisely control glycosylation on these protein drugs is one of major goals in glycoscience.

1.2 Engineering *E. coli* to produce glycoproteins

Traditionally, glycoprotein therapeutics are produced in eukaryotic cells because they have the native machinery required for glycosylation. In the biopharmaceutical industry, glycoproteins are commonly produced in CHO cells, a mammalian cell-line where glycosylation patterns closely resemble those of humans. However, a key production challenge is the variability of glycosylation distributions that arise from small changes to cell culture conditions. Furthermore, glycosylation is an important part of cellular function, so modifications to existing glycosylation machinery can have adverse effects on the cell.

To circumvent this issue, *E. coli*, which have no native glycosylation machinery,

have been proposed as an alternative host for producing recombinant glycoproteins because they provide a chassis for engineering in only the desired protein glycosylation pathways. The idea is that this will limit the range of glycoforms that can be produced and facilitate synthesis of more homogenous glycoproteins. Additionally, *E. coli* are a well-studied host with a long history of use for recombinant and biotherapeutic protein production because of their fast growth, relatively low costs, and range of simple genetic modification techniques[22]. In this section, I will highlight key milestones and challenges in engineering *E. coli* to produce a variety of glycoprotein structures.

1.2.1 Transfer of *C. jejuni* protein glycosylation system into *E. coli*

In 2002, Wacker, Linton, et al. demonstrated the functional transfer of *Campylobacter jejuni* protein *N*-glycosylation system into *E. coli*[23]. In this seminal work, the authors co-expressed the glycan biosynthesis gene cluster, *pgl*, and a *C. jejuni* periplasmic protein, AcrA, in *E. coli* and demonstrated that transfer of the glycan onto AcrA was dependent on the OST, PglB. This work set the foundation for the production of glycoproteins in *E. coli*. Since then, more has been learned about the specific mechanisms of the *C. jejuni* glycosylation system. In 2006, Kowarik et al. systematically mutated amino acids around the glycosylation sites in AcrA and found that there was an expanded *N*-glycosylation consensus sequence when using the bacterial OST[24]. In eukaryotes, the sequon is canonically N-X-S/T, but in the *C. jejuni* system, the authors found evidence for an expanded sequon, D/E-X-N-X-S/T. The expanded consensus sequence required in the *Campylobacter* glycosylation system has thus far limited the ability to glycosylate many biotherapeutic proteins at their natural

sites in *E. coli*. However, PglBs from *Desulfovibrio, Helicobacter,* and deep sea vent bacterial species have been identified that have more relaxed specificities[25–27].

1.2.2 Glycosylation pathway remodeling in *E. coli*

Although a more stringent amino acid sequon specifity is required, *Cj*PglB has proven to have relaxed glycan selectivity which has enabled the production of a variety of therapeutically relevant glycoproteins in *E coli*. In 2005, Feldman et al. demonstrated the ability of *Cj*PglB to transfer the O7 and O11 antigens from *E. coli* and *Pseudomonas aeruginosa*, respectively, onto a protein in *E. coli*[28]. O-antigens are polysaccharides found at the outer end of the lipopolysaccharide (LPS) of Gram-negative bacteria and contribute to antigenic variability of bacteria. Thus, they have been studied for their use as antigens in glycoconjugate vaccines. Normally, the O-antigen is transferred from UndPP to the lipid A core of the LPS by the enzyme, WaaL. The authors hypothesized that PglB could instead transfer the O-antigen structure from UndPP to a protein. Therefore, using a ΔwaaL mutant strain of *E. coli*, they co-expressed plasmids encoding for *Cj*PglB, the O-antigen biosynthesis pathway, and AcrA and were successfully able to transfer the O-antigen onto AcrA. This was the first evidence showing the relaxed glycan specificity of *Cj*PglB as well as the ability of *E. coli* to produce therapeutically relevant glycoproteins. Since then, a variety of other glycoconjugate vaccines to protect against bacteria have been produced in *E. coli*[29].

In addition to bacterial glycans, progress has been made towards the synthesis of more human-like glycan structures. In 2012, Valderrama-Rincon et al. reported

synthesis of glycoproteins bearing the $Man_3GlcNAc_2$ glycan structure[30]. In this work, a combination of native *E. coli* and recombinantly expressed *S. cerevisiae* glycan synthesis enzymes were used to build the $Man_3GlcNAc_2$ structure on UndPP on the cytoplasmic side of the inner membrane. The glycan was then flipped into the periplasm by a native *E. coli* flippase where it was accessible to *Cj*PglB for transfer onto the protein. This is significant because $Man_3GlcNAc_2$ is the human core *N*-glycan structure and could serve as the base for further remodeling into complex type *N*-glycans found on many biotherapeutic glycoproteins. Alternatively, Schwarz et al. proposed a method for producing eukaryotic-like structures using an engineered *C. jejuni* glycan synthesis pathway[31]. In this system, a bacterial glycan was installed on the protein that contained a GlcNAc-asparagine linkage. This glycan could then be trimmed down to a single GlcNAc which could be remodeled to produce eukaryotic N-glycans. Some other examples of human-like structures produced in *E. coli* include the Lewis x antigen[32], human blood group B[33] antigens, and terminally sialyated N-glycans[34]. The variety of glycans that can be installed onto proteins by *Cj*PglB highlights the relaxed glycan specificity, and thus versatility of this enzyme for production of therapeutic glycoproteins.

1.3 Cell-free technologies for glycoproteins synthesis

Cell-free glycosylation pathway assembly involves the use of glycosylation enzymes— purified from cells or produced from cell-free crude extracts— to sequentially build diverse glycan structures. These reactions are mediated by OSTs and GTs that use lipid-linked oligosaccharides (LLOs) or nucleotide activated sugars,

respectively, as substrates to glycosylate proteins. The advantages of recreating glycosylation pathways in cell-free systems are enhanced control over local reaction environment conditions, decoupling of glycosylation from cell viability, and the ability to use enzymes from any host source in the same pathway.

Early work in this space has leveraged glycoengineered strains of *E. coli* to generate glycosylation components for use *in vitro*. This is because bacterial *N*-glycosylation systems are relatively simple, and thus amendable to expression outside the cell. In 2005, Glover et al. demonstrated that purified *Cj*PglB could be used to transfer a synthetic donor, Und-PP-disaccharide, onto a synthetic peptide acceptor[35]. Additionally, they showed that the complete *C. jejuni* LLO could be synthesized *in vitro* using purified enzymes[36]. Following this work, Kowarik et al. used the *C. jejuni in vitro* glycosylation system to evaluate the ability of *Cj*PglB to glycosylate a model protein in either a folded or unfolded state[37]. From this work, they postulated that *N*-glycosylation in bacteria occurs in flexible parts of folded proteins in contrast to in higher eukaryotes where glycosylation is coupled to translational machinery and thus occurs in unfolded proteins. Further establishing *in vitro* glycosylation as a standard tool, Jaffee et al. developed a protocol that greatly increased yield and stability of purified *Cj*PglB, a large integral membrane protein, for its use in glycoengineering and glycobiology[38].

Since then, *in vitro* glycosylation technologies have been coupled with cell-free protein synthesis (CFPS) with the goal of enabling all reactions components to be generated cell-free. CFPS is an innovative manufacturing approach that involves mixing biological enzymes obtained from crude extracts and essential metabolites together in

'one-pot' to produce batches of high-value biologics[39,40]. CFPS has become a well-established method that has been shown to be promising for on-demand manufacturing of drugs, affinity conjugates, vaccines, and antimicrobial peptides[41]. While crude extracts have been generated from a variety of eukaryotic organisms, such as mammalian[42,43], yeast[44], plant[45,46], and insect[47] cells for protein synthesis, precise control over glycosylation reactions has yet to be demonstrated in these systems.

As a step towards that goal, in 2012, Guarino et al. explored coupling *in vitro* glycosylation with cell-free protein synthesis[48]. In this work, the authors supplemented cell-free extracts generated from *E. coli* with purified *Cj*PglB and extracted *Cj*LLOs. This enabled cell-free production of a protein substrate that could then be glycosylated by interaction with the *C. jejuni* machinery. Taking this idea a step further, in 2018, Jaroentomeechai, Stark, et al. developed single-pot platform for cell-free glycoprotein synthesis (CFGpS) where *Cj*PglB and *Cj*LLOs were enriched in *E. coli* cells prior to the preparation of cell-free extracts[49]. This simplified the production process by bypassing the need for purification and extraction of glycosylation components and allowed for cell-free production of glycosylated proteins simply through addition of a plasmid to the CFGpS lysate. Since then, this method has since been applied to produce glycoconjugate vaccines for use at point-of-care[50].

While generally the first step of protein *N*-glycosylation involves the OST, another group of enzymes that can be used to attach carbohydrates to the asparagine residue of proteins are *N*-glycosyltransferases (NGTs). The most well-characterized NGT is derived from *Actinobacillus pleuropneumoniae* (*Ap*NGT)[51]. These are soluble,

single subunit enzymes that catalyze the addition of nucleotide activated Glc and Gal to the N-X-S/T consensus sequence on an acceptor protein[52]. Using *Ap*NGT, Kightlinger et al. demonstrated attachment of a Glc primer to a protein that could then be elaborated chemoenzymatically using enzymes generated cell-free[53]. This work provides an alternative strategy for building glycoproteins without the use of OSTs and LLOs.

After the initial transfer of a carbohydrate to the protein— either through use of an OST or NGT— the glycan can then be selectively modified by enzymes such as GTs that add nucleotide activated monosaccharides to the glycan and glycosidases that remove sugars from the glycan. These enzymes are amenable to *in vitro* systems and have been used to build glycosylation pathways in one-pot multienzyme (OPME) systems for chemoenzymatic addition of the nine common mammalian monosaccharides[54]. The combination of all these tools to build glycosylation pathways outside of the cell has been termed cell-free synthetic glycobiology (CFSG)[55], as illustrated in **Fig. 1.3**. In CFSG, glycosylation enzymes and substrates can be extracted from cells or expressed by CFPS and combined to yield a variety of designer glycomolecules. The advantages of this technology are enhanced control over local reaction environment conditions, decoupling of glycosylation from cell viability, and the ability to use enzymes from any host source in the same pathway.

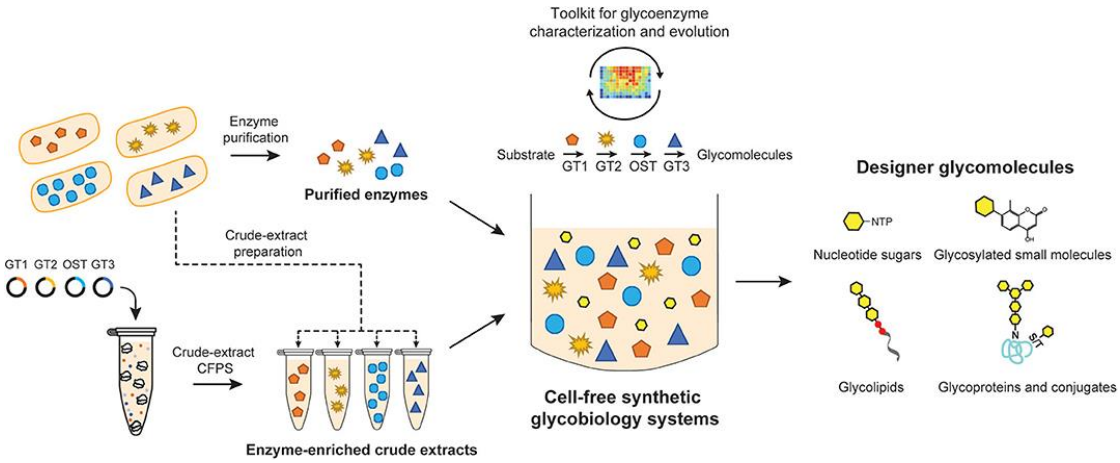


Figure 1.3. Overview of cell-free synthetic glycobiology strategy. Protein substrates and glycosylation enzymes are purified from cells or generated by cell-free protein synthesis (CFPS) and combined in sequential reactions or in a single-pot to catalyze a prescribed glycosylation reaction. This figure is from Jaroentomeechai, T. et al. *Front. Chem.* **8**, (2020)[55] and is under the Creative Commons Attribution License (CC BY).

1.4 Applications of microfluidics in glycan and protein synthesis

More than thirty years ago, microfluidics were developed alongside microelectronics through advancements in photolithography that enabled precise fabrication of micro and nano-scale designs into materials such as silicon, glass, and poly (methyl methacrylate)[56]. However, these devices, which require specialized equipment and clean room facilities to manufacture, were not accessible to many researchers until strategies were developed that make use of one silicon wafer to serve as a mold to produce many devices. This strategy, pioneered by George Whitesides and coworkers, uses poly(dimethylsiloxane) (PDMS) as a material for cheaper and faster device fabrication[57]. Since then, the use of microfluidic devices has expanded and now includes applications in areas of biology such as synthetic biology[58,59], biophysics[60,61] and cell biology[62]. The advantages of using microfluidics as tools to study and

recapitulate biological processes are the small volumes that minimize waste, length and time scales similar to those *in vivo*, spatiotemporal control, and ability to monitor processes through real-time high resolution imaging[63]. The applications of microfluidics are broad, but here I focus specifically on their use as microreactors for production of glycans and proteins.

1.4.1 Synthesis of glycans in microfluidics

Two approaches used for synthesis of *N*-glycans in microfluidic devices are chemical and chemoenzymatic synthesis[64]. The major challenge in glycan synthesis is linking together sugar monomers in a stereo- and regioselective manner. Commonly, monosaccharides are composed of five or more carbons that form a ring. Each carbon is linked to a hydroxyl group where a glycosidic linkage can form between sugars. Therefore, in chemical synthesis, the general scheme for this is a series of protection and deprotection reactions. All hydroxyl groups that are not participating in the glycosidic bond need to be shielded by a protecting group. When a free hydroxyl group is desired in the final structure, a permanent protecting group such as benzyl ethers is used whereas when a hydroxyl group needs to be exposed during synthesis, a temporary protecting group such as esters is used[65,66]. These protecting groups can then be removed to create a nucleophile that will participate in a glycosylation reaction or free hydroxyl groups in the final deprotected glycan structure. For this process, oftentimes the growing glycan chain is tethered to a solid support to simplify the isolation of the products from the reagent. Because of the open and cyclic nature of this process, chemical synthesis

of glycans has been amenable to automation[67] and computer-aided approaches[68]. Using these methods, a variety of mammalian, bacterial, fungal, plant, and unnatural glycan structures have been generated[69]. Based on this chemical synthesis approach, microfluidics has been used to enable efficient mixing and rapid heat transfer of the organic synthesis reactions and to serve as a small-scale model for reaction optimization. For example, in work by Koichi Fukase's lab, microfluidics has been used to improve production of asparagine linked glycans containing α-sialylation[70] and other synthetic glycoconjugates[71].

In chemoenzymatic synthesis of glycans, glycosylation enzymes are used to form the precise linkages between monosaccharides at physiological conditions without the need for hazardous or expensive chemicals. These enzymes have been identified from organisms across all domains of life and are catalogued in databases like the Carbohydrate-Active Enzyme (CAZy) database[72] that facilitate selection of enzymes to synthesize a particular product. For chemoenzymatic synthesis of glycans, microfluidics enables reuse of enzymes through tethering or recycling strategies and optimization of enzymatic reactions at small scales. A common format for these systems is flow-based transport of substrates through enzymes that are either immobilized to the tubular surface of the microfluidic device or to beads in a packed bed system. For example, in work by Ono et al., a three chamber microfluidic device was fabricated where each compartment contained glycoenzymes tethered to beads[64]. A syringe pump was then used to push the substrates through the device to enable production of a glycosaminoglycan linkage region tetrasaccharide. Another format is movement of

magnetic carriers attached to substrates[73] or enzymes[64] by electromagnetic fields. For example, in work by Martin et al., enzymes were compartmentalized in separate chambers, analogous to what occurs in the Golgi. Heparan sulfate was immobilized on magnetic nanoparticles that could then be moved between compartments to enable sequential glycosylation[73].

1.4.2 Protein production in microfluidics

There has been increasing interest in the biopharmaceutical industry to implement continuous manufacturing technologies[74] that grant greater control over reaction variables, are amenable to automation, and are more flexible to changes in market demand compared to batch reactors[75,76]. Therefore, as scaled-down models of flow systems, microfluidic devices are a potential tool for prototyping flow-based technologies. As mentioned previously, biologics are commonly produced in cells, so this transition requires a new toolset for continuous upstream and downstream bioprocessing[77]. To this end, microfluidics has been proposed to be used as bioprocess development tools for culture and monitoring of cells, product extraction and purification, and product analysis, but development of most of these technologies is still in early stages[78,79]. Circumventing the need for many of these new tools are cell-free systems that have enabled production of biologics outside of cells. Using microreactors, Scott Retterer's lab has demonstrated CFPS in continuous flow systems and have shown up to a six-fold increase in protein production[80,81]. As cell-free technologies advance to

include protein glycosylation, the range of biologics that can be synthesized in flow-based microreactors will expand.

1.4.3 Mimicking cellular compartmentalization using microfluidics

Although the use of microfluidics for recapitulating cellular processes is vast, in this dissertation work, I focus on the ability of microfluidics to offer compartmentalization for glycoprotein manufacturing. In cells, protein glycosylation is highly coordinated through compartmentalization in the ER and the Golgi. This enables spatiotemporal control where substrates and enzymes are co-localized in compartments at high concentrations that improve reaction efficiency and conversion and occur in a specific sequence of reaction steps when the right substrates are available. Thus, for glycoprotein synthesis, microfluidics offers a promising technology to mimic the spatiotemporal control naturally found in cells that cannot be achieved in single pot reaction mixtures. Indeed, integration of cell-free glycoprotein synthesis systems into a microfluidic device has the potential to enable compartmentalization of key bioprocesses into separate chips where they can be individually optimized, tunable reaction times that are controlled by the flowrate or length of the channels, reusability of the biocatalyst through tethering to the device, and potentially opening an avenue for continuous manufacturing of glycoproteins that are currently limited to production as batch reactions in cells.

1.5 Single subunit eukaryotic OSTs

Although glycoengineering strategies in *E. coli* have focused only on

incorporating bacterial PglBs, OSTs are present across all domains of life and could potentially be used in bacterial and cell-free systems. In higher eukaryotes, the OST is a multisubunit transmembrane protein composed of eight subunits[82,83]. Only five of the eight subunits are required for cell survival. One of the essential subunits is the catalytic domain called the STT3. The STT3 has thirteen transmembrane helices followed by a C-terminal soluble domain that contains the catalytic pocket. The other essential subunits are involved in substrate recognition, association with oxidoreductases or the translocon, and stabilizing the complex. Although these eukaryotic enzymes would potentially be better at transferring human-like glycan structures to eukaryotic proteins at their natural sites, the complexity of these multisubunit OST structures will make it challenging to express them recombinantly. In contrast, in some lower eukaryotes the OST is composed of just a single subunit that is homologous to the STT3 domain of higher eukaryotes. These single subunit OSTs are found in kinetoplastids which are unicellular, parasitic protists. Although no homologs to the other OST subunits are found in their genomes, there are multiple copies of the STT3 gene. Duplication of the STT3 enzymes is thought to expand glycosylation ability of the host where each STT3 has different selectivity for glycan and protein substrates. The current understanding and functions of protozoan single subunit STT3s, in particular from *Leishmania* and *Trypanosoma*, will be the focus of this section.

1.5.1 Kinetoplastid STT3s do not integrate into an OST complex

To study the functions and substrate specificities of these enzymes, many have

recombinantly expressed protozoan STT3s in *Saccharomyces cerevisiae*, a model organism where glycosylation pathways and the OST structure are better understood. The first such study was done in 2005 by Shams-Eldin et al. who showed that over-expressing *Toxoplasma gondii* STT3 in *S. cerevisiae* could restore viability of a strain of yeast where the native *Sc*STT3 gene was deleted[84]. In yeast, *N*-glycosylation is required for the growth and survival of the cell, so this meant that the single subunit *Tg*STT3 could effectively replace the function of native *Sc*STT3 which is the catalytic component of the eight subunit *Sc*OST. This raises the question of whether protozoan ssOSTs function independently of the yeast OST complex or integrate into the complex in place of *Sc*STT3.

In 2006, Castro et al. performed a similar experiment in which they replaced native *Sc*STT3 with *Trypanosoma cruzi* STT3 and showed that the yeast were still viable[85]. In this study, the authors found that *Tc*STT3 co-precipitated with another *Sc*OST subunit, Ost1p, and that the glycan substrate specificity of *Tc*STT3 changed depending on whether it was expressed in the context of yeast or *T. cruzi*. These results suggested that *Tc*STT3 was integrating into the *Sc*OST complex, and that the preferential transfer of the glycan is determined by the entire OST complex, not the catalytic subunit. However, these findings were not true of all protozoan STT3s. In 2008-2009, studies examining STT3s from *Leishmania major* in the context of yeast were published[86,87]. There are four paralog STT3 genes in *L. major* that will be labeled here as *Lm*STT3A, *Lm*STT3B, *Lm*STT3C, and *Lm*STT3D. These studies found that *Lm*STT3A, *Lm*STT3B, and *Lm*STT3D could replace the function of *Sc*STT3, but not

*Lm*STT3C. Notably, when the essential *Sc*OST subunits were individually knocked out along with *Sc*STT3, expression of *Lm*STT3A[87], *Lm*STT3B[87], and *Lm*STT3D[86,87] could restore viability of the yeast cells. Additionally, in blue native gel electrophoresis experiments, the *Lm*STT3s did not appear to migrate with *Sc*OST subunits[86] or only a small amount integrated into the OST complex[87]. Together, these results suggest that *L. major* STT3s do not integrate into the yeast OST, but instead, operate independently and replace the function of the entire *Sc*OST complex.

1.5.2 Glycan and protein specificities of kinetoplastid STT3s

Another question that arises about these protozoan STT3s is whether they have relaxed or narrow glycan substrate and protein sequon specificities. Although the study of *Trypanosoma cruzi* STT3 suggested that the glycan specificity is not dependent on the STT3[85], the studies of *Leishmania major* STT3s showed evidence that there is both glycan and peptide sequence specificities. When using *Lm*STT3A, *Lm*STT3B, or *Lm*STT3D to replace *Sc*STT3 for glycosylation of native yeast cell wall proteins, mass spectrometry showed that the relative glycosylation occupancy of the peptide glycosylation sites changed depending on which *Lm*STT3 paralog was expressed[86]. When the ability of the *Lm*STT3s to glycosylate carboxypeptidase Y (CPY), a model protein with four glycosylation sites, was interrogated, both studies found that *Lm*STT3D was able to fully glycosylate CPY while *Lm*STT3A and *Lm*STT3B could not[86,87]. Furthermore, after knockout of *alg* genes that truncate the native $Glc_3Man_9GlcNAc_2$ in yeast, expression of *Lm*STT3D was found to increase transfer of

$Man_{5-7}GlcNAc_2$[86] and expression of *Lm*STT3A was found to increase transfer of $Man_{5-6}GlcNAc_2$ and $Man_9GlcNAc_2$[87]. Together, these results suggest that each *Lm*STT3 has its own protein and glycan substrate specificity.

Expanding on this early work, studies of the three STT3 paralogs from *Trypanosoma brucei* have more thoroughly characterized substrate specificities. In initial studies by Izquierdo et al., the authors examined *Tb*STT3A and *Tb*STT3B in the context of its native host, *T. brucei*, and *Tb*STT3B and *Tb*STT3C in *S. cerevisiae*[88]. From these studies, they found that *Tb*STT3A selectively transferred $Man_5GlcNAc_2$ to acidic regions of the protein, *Tb*STT3B transferred $Man_9GlcNAc_2$ to neutral or basic regions of the protein (albeit less selectively) and *Tb*STT3C could use the yeast $Glc_3Man_9GlcNAc_2$ as a substrate (though yeast expressing *Tb*STT3C in place of *Sc*STT3 exhibited a severe growth defect) and had more selectivity for acidic peptide acceptors[88]. In a follow-up study, the authors further specified that *Tb*STT3B selectively transferred longer high mannose glycans ($Man_{8-9}GlcNAc_2$) over shorter high mannose glycans ($Man_{5-7}GlcNAc_2$) while *Tb*STT3A only transferred shorter high mannose glycans[89]. Later, the protein substrate specificities of *Tb*STT3A and *Tb*STT3B were more systematically examined using computational methods and experiments using a synthetic peptide where the charge of amino acids ten amino acids upstream and downstream of the glycosylation site were varied[90]. The results from this study corroborated previous results where *Tb*STT3A has preference for acidic residues around the glycosylation site and *Tb*STT3B preferred hydrophobic and positively charged residues. With greater understanding of the substrate specificities of the *Tb*STT3s, a

2017 study demonstrated that *Tb*STT3A could restore glycosylation in yeast cells with an *Sc*STT3 deletion if it was also paired with *alg* mutations that produce truncated glycans[91]. Using this strategy, the substrate specificities of all the *Tb*STT3s could be interrogated in the same host. In this study it was found that *Tb*STT3A and *Tb*STT3C prefer upstream acidic residues while *Tb*STT3B showed no particular preference. Additionally, *Tb*STT3C exhibited the most relaxed glycan specificity where it transferred all tested structures ($Man_{5-7}GlcNAc_2$, $Man_9GlcNAc_2$, and $Glc_3Man_9GlcNAc_2$) while *Tb*STT3A only transferred $Man_5GlcNAc_2$ and *Tb*STT3B transferred all except $Man_5GlcNAc_2$.

1.5.3 Kinetoplastid STT3s for cell-based and cell-free glycoprotein production

Kinetoplastid STT3s are promising OSTs for use in cell-based or cell-free glycoprotein production systems because they are single subunit enzymes that can be more readily expressed in model organisms or stabilized outside of the cell compared to their multisubunit eukaryotic OST counterparts. For cell-based platforms, *Lm*STT3D in particular has been explored to improve glycosylation site occupancy of monoclonal antibodies in various yeast[92,93] and plant[94] species because of its relatively broad glycan and protein substrate specificities[86,87]. Yeast natively do not produce complex glycan structures, so much work has focused on engineering yeast to produce more humanized glycoproteins[95]. Along this vein, in 2012, Choi et al. expressed *Lm*STT3A, *Lm*STT3B, *Lm*STT3C, and *Lm*STT3D in glycoengineered *Pichia pastoris* and tested their ability to improve glycosylation of anti-HER2 and anti-RSV antibodies[92]. In these experiments

they found that only expression of *Lm*STT3D significantly improved *N*-glycan site occupancy of the antibody to greater than 99% and did not appear to significantly affect the *N*-glycan composition. In 2013, Nasab et al. engineered *S. cerevisiae* to produce $Man_3GlcNAc_2$ that could be subsequently modified by recombinantly expressed GTs to produce complex-type *N*-glycans[93]. In this work, protozoan STT3s from *Leishmania major*, *Leishmania brasiliensis*, *Leishmania infantum*, and *Trypanosoma brucei* were investigated for their ability to transfer the truncated $Man_3GlcNAc_2$ structure. They found that *Lb*STT3_1, *Lb*STT3_3, and *Lm*STT3D improved glycosylation of the model glycoprotein, CPY with $Man_3GlcNAc_2$. Following this OST screen, *Lb*STT3_3 was chosen as the best candidate and enabled production of a monoclonal antibody, HyHEL-10, glycosylated with complex type glycans. Like in yeast, glycosylation systems in plants also need to be engineered to produce humanized glycan structures[96] and glycan site occupancy with these engineered structures could be improved with protozoan STT3s. In 2018, Castilho et al. expressed LmSTT3D in *Nicotiana benthamiana*, and showed that this improved glycosylation site occupancy on various therapeutic glycoproteins such as immunoglobulin (Ig) G, IgA, IgE, EPO-Fc, and IFN- γ[94].

For cell-free systems, *Tb*STT3A has mainly been explored. In 2017, Ramírez et al. expressed STT3s from *L. braziliensis*, *L. infantum*, *L. major*, and *T. brucei* in human embryonic kidney (HEK293) cells and found that *Tb*STT3A exhibited the highest expression levels[97]. Therefore, they chose this STT3 to study *in vitro*. In these studies, turnover rates were calculated for *Tb*STT3A when glycosylating synthetic peptides with varied amino acids in the glycosylation sequon and synthetic lipid carriers with varied

lengths (C10, C15, C20 and C25) and double bond stereochemistry coupled to a chitobiose moiety. The results showed that for *Tb*STT3A, turnover rates were highest for negatively charged amino acids at the -2 position and decreased with positively charged amino acids at this position[97], similar to what was observed *in vivo*[88–91]. When comparing the effect of changes to the synthetic lipids, it was found that the length of the polyprenyl tail affected their apparent affinity, with longer lengths increasing K_M (though the difference between the k_{cat} values was, at maximum, 2-fold). The stereochemistry of the double bonds did not appear to have a significant effect. In a follow-up, the authors also showed that *Tb*STT3A transferred $GlcNAc_2$, $Man_3GlcNAc_2$, and $Man_5GlcNAc_2$ onto a peptide at the same rate[98]. This was interesting because although *Tb*STT3A has been shown to prefer $Man_5GlcNAc_2$ over $Man_9GlcNAc_2$[88–91], it does not seem to have a preference among the shorter glycans. In addition to investigating the substrate specificities *in vitro*, cell-free systems have also incorporated *Tb*STT3A for chemoenzymatic synthesis of LLOs. In a 2020 study by Rexer et al., mannopentaose-di-(N-acetylglucosamine) was synthesized chemoenzymatically from low-cost substrates. Purified *Tb*STT3A could then be used to attach the glycan to a model peptide.

1.6 Summary

Protein glycosylation is a fundamental biological process that has important effects in drug design. Thus, the ability to produce glycoproteins with a desired, uniform glycan structure would facilitate their study and use as therapeutics. In eukaryotic cells that have been traditionally used for glycoprotein production, protein glycosylation is

highly coordinated through compartmentalization in the ER and the Golgi. However, in these systems, the ability to control glycosylation patterns is limited because glycosylation is tied to cell health, so any changes to the existing glycosylation machinery can negatively affect cell growth and viability. As an alternative manufacturing approach, studies have examined the ability of bacteria, which do not require glycosylation, to produce glycoproteins. In particular, the general protein glycosylation system from *C. jejuni* has been extensively studied and was successfully transferred into *E. coli* where it has enabled the production of a variety of therapeutic glycoproteins. Using these glycoengineered *E. coli*, cell-free technologies have been developed where protein synthesis and glycosylation components are extracted and combined into a single pot that enable more direct control over glycosylation reactions.

In Chapter 2 of this dissertation work, I explore the use of microfluidic technologies to improve the production of glycoproteins cell-free. The idea is that compartmentalization of each reaction step will offer greater control over the process conditions. In Chapter 3, to expand the range of glycoproteins that can be produced in these systems, I then explore the use of single subunit OSTs from eukaryotic sources that I believe will be amenable to *E. coli*-based cell-free systems. Together, these advancements bring us closer to the ability to perform bottom-up construction of designer, structurally uniform glycoproteins for study and use as therapeutics.

References

1. Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* (2001). doi:10.1126/science.291.5512.2364

2. Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **1**, 1–5 (2011).
3. Stanley P, Moremen KW, Lewis NE, et al. N-Glycans. In: Varki A, Cummings RD, Esko JD, et al., editors. Essentials of Glycobiology [Internet]. 4th edition. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022. Chapter 9. Available from: https://www.ncbi.nlm.nih.gov/books/NBK579964/
4. Clausen, H., Wandall, H. H., DeLisa, M. P., Stanley, P. & Schnaar, R. L. *Glycosylation Engineering*. *Essentials of Glycobiology* (2022).
5. Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: Sweeter than ever. *Nat. Rev. Microbiol.* **8**, 765–778 (2010).
6. Calo, D., Kaminski, L. & Eichler, J. Protein glycosylation in Archaea: Sweet and extreme. *Glycobiology* **20**, 1065–1076 (2010).
7. Jarrell, K. F. *et al.* N-Linked Glycosylation in Archaea: a Structural, Functional, and Genetic Analysis. *Microbiol. Mol. Biol. Rev.* **78**, 304–341 (2014).
8. Szymanski, C. M., Ruijin, Y., Ewing, C. P., Trust, T. J. & Guerry, P. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol. Microbiol.* **32**, 1022–1030 (1999).
9. Young, N. M. *et al.* Structure of the N-linked glycan present on multiple glycoproteins in the gram-negative bacterium, Campylobacter jejuni. *J. Biol. Chem.* **277**, 42530–42539 (2002).
10. Kelly, J. *et al.* Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer. *J. Bacteriol.* **188**, 2427–2434 (2006).
11. Linton, D. *et al.* Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. *Mol. Microbiol.* **55**, 1695–1703 (2005).
12. Alaimo, C. *et al.* Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25**, 967–976 (2006).
13. Wacker, M. *et al.* N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science (80-.).* **298**, 1790–1793 (2002).
14. Dammen-Brower, K. *et al.* Strategies for Glycoengineering Therapeutic Proteins. *Front. Chem.* **10**, 1–27 (2022).
15. Imperiali, B. & O'Connor, S. E. Effect of N-linked glycosylatian on glycopeptide and glycoprotein structure. *Current Opinion in Chemical Biology* (1999). doi:10.1016/S1367-5931(99)00021-6
16. Hebert, D. N., Lamriben, L., Powers, E. T. & Kelly, J. W. The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nature Chemical Biology* (2014). doi:10.1038/nchembio.1651
17. Lin, C. W. *et al.* A common glycan structure on immunoglobulin G for enhancement of effector functions. *Proc. Natl. Acad. Sci. U. S. A.* (2015). doi:10.1073/pnas.1513456112
18. Wolfert, M. A. & Boons, G. J. Adaptive immune activation: Glycosylation does matter. *Nat. Chem. Biol.* **9**, 776–784 (2013).

19. Amgen, O. & Oaks, T. erythropoiesis stimulating protein (NESP) y \ f y. **84**, 3–10 (2001).
20. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4305
21. Butler, M. & Spearman, M. The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr. Opin. Biotechnol.* **30**, 107–112 (2014).
22. Rosano, G. L., Morales, E. S. & Ceccarelli, E. A. New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Sci.* **28**, 1412–1422 (2019).
23. Wacker, M. *et al.* N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science (80-.).* (2002). doi:10.1126/science.298.5599.1790
24. Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966 (2006).
25. Ielmini, M. V. & Feldman, M. F. Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology* **21**, 734–742 (2011).
26. Ollis, A. A. *et al.* Substitute sweeteners: Diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. *Sci. Rep.* (2015). doi:10.1038/srep15237
27. Mills, D. C. *et al.* Functional analysis of N-linking oligosaccharyl transferase enzymes encoded by deep-sea vent proteobacteria. *Glycobiology* **26**, 398–409 (2016).
28. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* (2005). doi:10.1073/pnas.0500044102
29. Kay, E., Cuccui, J. & Wren, B. W. Recent advances in the production of recombinant glycoconjugate vaccines. *npj Vaccines* **4**, 1–8 (2019).
30. Valderrama-Rincon, J. D. *et al.* An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat. Chem. Biol.* **8**, 434–436 (2012).
31. Schwarz, F. *et al.* A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat. Chem. Biol.* **6**, 264–266 (2010).
32. Hug, I. *et al.* Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M111.287755
33. Shang, W. *et al.* Production of human blood group B antigen epitope conjugated protein in Escherichia coli and utilization of the adsorption blood group B antibody. *Microb. Cell Fact.* **15**, 138 (2016).
34. Zhu, J. *et al.* An Engineered Pathway for Production of Terminally Sialylated N-glycoproteins in the Periplasm of Escherichia coli. *Front. Bioeng. Biotechnol.* **8**, 1–9 (2020).
35. Glover, K. J., Weerapana, E., Numao, S. & Imperiali, B. Chemoenzymatic

synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. *Chem. Biol.* **12**, 1311–1316 (2005).
36. Glover, K. J., Weerapana, E. & Imperiali, B. In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14255–14259 (2005).
37. Kowarik, M. *et al.* N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science (80-.).* **314**, 1148–1150 (2006).
38. Jaffee, M. B. & Imperiali, B. Optimized protocol for expression and purification of membrane-bound PglB, a bacterial oligosaccharyl transferase. *Protein Expr. Purif.* **89**, 241–250 (2013).
39. Katzen, F., Chang, G. & Kudlicki, W. The past, present and future of cell-free protein synthesis. *Trends in Biotechnology* (2005). doi:10.1016/j.tibtech.2005.01.003
40. Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: Applications come of age. *Biotechnology Advances* (2012). doi:10.1016/j.biotechadv.2011.09.016
41. Pardee, K. *et al.* Portable, On-Demand Biomolecular Manufacturing. *Cell* (2016). doi:10.1016/j.cell.2016.09.013
42. Martin, R. W. *et al.* Development of a CHO-Based Cell-Free Platform for Synthesis of Active Monoclonal Antibodies. *ACS Synth. Biol.* (2017). doi:10.1021/acssynbio.7b00001
43. Ludgate, L. *et al.* Cell-Free Hepatitis B Virus Capsid Assembly Dependent on the Core Protein C-Terminal Domain and Regulated by Phosphorylation. *J. Virol.* (2016). doi:10.1128/jvi.00394-16
44. Schoborg, J. A., Hodgman, C. E., Anderson, M. J. & Jewett, M. C. Substrate replenishment and byproduct removal improve yeast cell-free protein synthesis. *Biotechnol. J.* (2014). doi:10.1002/biot.201300383
45. Tsuboi, T. *et al.* Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. *Infect. Immun.* (2008). doi:10.1128/IAI.01539-07
46. Buntru, M., Vogel, S., Spiegel, H. & Schillberg, S. Tobacco BY-2 cell-free lysate: An alternative and highly-productive plant-based in vitro translation system. *BMC Biotechnol.* (2014). doi:10.1186/1472-6750-14-37
47. Ezure, T., Suzuki, T. & Ando, E. A cell-free protein synthesis system from insect cells. *Methods Mol. Biol.* (2014). doi:10.1007/978-1-62703-782-2_20
48. Guarino, C. & Delisa, M. P. A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology* **22**, 596–601 (2012).
49. Jaroentomeechai, T. *et al.* Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat. Commun.* **9**, 2686 (2018).
50. Stark, J. C. *et al.* On-demand biomanufacturing of protective conjugate vaccines. *Sci. Adv.* **7**, 1–16 (2021).
51. Kawai, F. *et al.* Structural insights into the glycosyltransferase activity of the

Actinobacillus pleuropneumoniae HMW1C-like protein. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M111.237602
52. Schwarz, F., Fan, Y. Y., Schubert, M. & Aebi, M. Cytoplasmic N-glycosyltransferase of Actinobacillus pleuropneumoniae is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M111.277160
53. Kightlinger, W. *et al.* A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nat. Commun.* **10**, 5404 (2019).
54. Yu, H. & Chen, X. One-pot multienzyme (OPME) systems for chemoenzymatic synthesis of carbohydrates. *Organic and Biomolecular Chemistry* (2016). doi:10.1039/c6ob00058d
55. Jaroentomeechai, T. *et al.* Cell-Free Synthetic Glycobiology: Designing and Engineering Glycomolecules Outside of Living Cells. *Front. Chem.* **8**, (2020).
56. Convery, N. & Gadegaard, N. 30 years of microfluidics. *Micro and Nano Engineering* (2019). doi:10.1016/j.mne.2019.01.003
57. McDonald, J. C. *et al.* Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* (2000). doi:10.1002/(SICI)1522-2683(20000101)21:1<27::AID-ELPS27>3.0.CO;2-C
58. Szita, N., Polizzi, K., Jaccard, N. & Baganz, F. Microfluidic approaches for systems and synthetic biology. *Current Opinion in Biotechnology* (2010). doi:10.1016/j.copbio.2010.08.002
59. Damiati, S., Mhanna, R., Kodzius, R. & Ehmoser, E. K. Cell-free approaches in synthetic biology utilizing microfluidics. *Genes* (2018). doi:10.3390/genes9030144
60. Charmet, J., Arosio, P. & Knowles, T. P. J. Microfluidics for Protein Biophysics. *Journal of Molecular Biology* (2018). doi:10.1016/j.jmb.2017.12.015
61. Zheng, Y., Nguyen, J., Wei, Y. & Sun, Y. Recent advances in microfluidic techniques for single-cell biophysical characterization. *Lab on a Chip* (2013). doi:10.1039/c3lc50355k
62. Velve-Casquillas, G., Le Berre, M., Piel, M. & Tran, P. T. Microfluidic tools for cell biological research. *Nano Today* (2010). doi:10.1016/j.nantod.2009.12.001
63. Duncombe, T. A., Tentori, A. M. & Herr, A. E. Microfluidics: Reframing biological enquiry. *Nature Reviews Molecular Cell Biology* (2015). doi:10.1038/nrm4041
64. Heinzler, R., Hübner, J., Fischöder, T., Elling, L. & Franzreb, M. A compartmented flow microreactor system for automated optimization of bioprocesses applying immobilized enzymes. *Front. Bioeng. Biotechnol.* **6**, 1–10 (2018).
65. Palmacci, E. R., Plante, O. J., Hewitt, M. C. & Seeberger, P. H. Automated Synthesis of Oligosaccharides. *Helv. Chim. Acta* **86**, 3975–3990 (2003).
66. Pardo-Vargas, A., Delbianco, M. & Seeberger, P. H. Automated glycan assembly as an enabling technology. *Curr. Opin. Chem. Biol.* **46**, 48–55 (2018).
67. Hahm, H. S. *et al.* Automated glycan assembly using the Glyconeer 2.1

synthesizer. *Proc. Natl. Acad. Sci.* **114**, E3385–E3389 (2017).
68. Wu, C.-Y. & Wong, C.-H. Programmable One-Pot Glycosylation. in *Peptide-Based Materials* **310**, 223–252 (2010).
69. Guberman, M. & Seeberger, P. H. Automated Glycan Assembly: A Perspective. *J. Am. Chem. Soc.* **141**, 5581–5592 (2019).
70. Tanaka, K. *et al.* Synthesis of a Sialic Acid Containing Complex-Type N -Glycan on a Solid Support. *Chem. - An Asian J.* **4**, 574–580 (2009).
71. Tanaka, K., Miyagawa, T. & Fukase, K. Chemical N-Glycosylation by Asparagine under Integrated Microfluidic/Batch Conditions. *Synlett* **2009**, 1571–1574 (2009).
72. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495 (2014).
73. Martin, J. G. *et al.* Toward an artificial Golgi: Redesigning the biological activities of heparan sulfate on a digital microfluidic chip. *J. Am. Chem. Soc.* **131**, 11041–11048 (2009).
74. Fisher, A. C. *et al.* The Current Scientific and Regulatory Landscape in Advancing Integrated Continuous Biopharmaceutical Manufacturing. *Trends in Biotechnology* (2019). doi:10.1016/j.tibtech.2018.08.008
75. Malet-Sanz, L. & Susanne, F. Continuous flow synthesis. a pharma perspective. *Journal of Medicinal Chemistry* (2012). doi:10.1021/jm2006029
76. Poechlauer, P. *et al.* Pharmaceutical Roundtable Study Demonstrates the Value of Continuous Manufacturing in the Design of Greener Processes. *Org. Process Res. Dev.* (2013). doi:10.1021/op400245s
77. Hernandez, R. Continuous manufacturing: A changing processing paradigm. *BioPharm Int.* (2015).
78. Marques, M. P. & Szita, N. Bioprocess microfluidics: applying microfluidic devices for bioprocessing. *Current Opinion in Chemical Engineering* (2017). doi:10.1016/j.coche.2017.09.004
79. Titchener-Hooker, N. J., Dunnill, P. & Hoare, M. Micro biochemical engineering to accelerate the design of industrial-scale downstream processes for biopharmaceutical proteins. *Biotechnology and Bioengineering* (2008). doi:10.1002/bit.21788
80. Timm, A. C., Shankles, P. G., Foster, C. M., Doktycz, M. J. & Retterer, S. T. Characterization of extended channel bioreactors for continuous-flow protein production. *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* (2015). doi:10.1116/1.4932155
81. Timm, A. C., Shankles, P. G., Foster, C. M., Doktycz, M. J. & Retterer, S. T. Toward Microfluidic Reactors for Cell-Free Protein Synthesis at the Point-of-Care. *Small* (2016). doi:10.1002/smll.201502764
82. Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science (80-.).* **359**, 545–550 (2018).
83. Ramírez, A. S., Kowal, J. & Locher, K. P. Cryo-electron microscopy structures

of human oligosaccharyltransferase complexes OST-A and OST-B. *Science (80-.).* **366**, 1372–1375 (2019).
84. Shams-Eldin, H. *et al.* High-level expression of the Toxoplasma gondii STT3 gene is required for suppression of the yeast STT3 gene mutation. *Mol. Biochem. Parasitol.* **143**, 6–11 (2005).
85. Castro, O., Movsichoff, F. & Parodi, A. J. Preferential transfer of the complete glycan is determined by the oligosaccharyltransferase complex and not by the catalytic subunit. *Proc. Natl. Acad. Sci. U. S. A.* (2006). doi:10.1073/pnas.0607086103
86. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J. & Aebi, M. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. *Mol. Biol. Cell* (2008). doi:10.1091/mbc.E08-05-0467
87. Hese, K., Otto, C., Routier, F. H. & Lehle, L. The yeast oligosaccharyltransferase complex can be replaced by STT3 from Leishmania major. *Glycobiology* **19**, 160–171 (2009).
88. Izquierdo, L. *et al.* Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. *EMBO J.* **28**, 2650–2661 (2009).
89. Izquierdo, L., Mehlert, A. & Ferguson, M. A. The lipid-linked oligosaccharide donor specificities of Trypanosoma brucei oligosaccharyltransferases. *Glycobiology* **22**, 696–703 (2012).
90. Jinnelov, A., Ali, L., Tinti, M., Güther, M. L. S. & Ferguson, M. A. J. Single-subunit oligosaccharyltransferases of Trypanosoma brucei display different and predictable peptide acceptor specificities. *J. Biol. Chem.* **292**, 20328–20341 (2017).
91. Poljak, K. *et al.* Analysis of substrate specificity of Trypanosoma brucei oligosaccharyltransferases (OSTs) by functional expression of domain-swapped chimeras in yeast. *J. Biol. Chem.* (2017). doi:10.1074/jbc.M117.811133
92. Choi, B. K. *et al.* Improvement of N-glycan site occupancy of therapeutic glycoproteins produced in Pichia pastoris. *Appl. Microbiol. Biotechnol.* **95**, 671–682 (2012).
93. Nasab, F. P., Aebi, M., Bernhard, G. & Frey, A. D. A combined system for engineering glycosylation efficiency and glycan structure in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **79**, 997–1007 (2013).
94. Castilho, A. *et al.* An oligosaccharyltransferase from Leishmania major increases the N-glycan occupancy on recombinant glycoproteins produced in Nicotiana benthamiana. *Plant Biotechnol. J.* **16**, 1700–1709 (2018).
95. Wildt, S. & Gerngross, T. U. The humanization of N-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* **3**, 119–128 (2005).
96. Strasser, R., Altmann, F. & Steinkellner, H. Controlled glycosylation of plant-produced recombinant proteins. *Curr. Opin. Biotechnol.* **30**, 95–100 (2014).
97. Ramírez, A. S. *et al.* Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic

peptides and lipid-linked oligosaccharide analogs. *Glycobiology* **27**, 525–535 (2017).
98. Ramírez, A. S. *et al.* Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins. *Glycobiology* **27**, 726–733 (2017).

2 GLYCOSYLATION-ON-A-CHIP: A FLOW-BASEDMICROFLUIDIC SYSTEM FOR CELL-FREE GLYCOPROTEIN BIOSYNTHESIS

*This chapter appears in Aquino, A. K., Manzer, Z. A., Daniel, S., & DeLisa, M. P. "Glycosylation-on-a-chip: a flow-based microfluidic system for cell-free glycoprotein biosynthesis." *Frontiers in molecular biosciences* 8 (2021).

2.1 Introduction

Protein glycosylation is a major posttranslational modification where complex carbohydrates known as glycans are enzymatically added to amino acid sidechains of a protein at specific, regioselective positions. The potential information content encoded in these glycans greatly exceeds that of other biomacromolecules, with distinct glycan structures often playing critical roles in health and disease[3,4]. The attachment of glycans to asparagine residues, known as *N*-linked glycosylation, is the most abundant type of glycosylation and occurs in all domains of life[5]. This mode of glycosylation gives rise to diverse chemical structures that are well known to affect the biological and biophysical properties of a protein[6–9]. Because of these pronounced effects, there is a strong incentive to study glycosylation and leverage the resulting knowledge for the development of glycoengineered proteins with advantageous properties[10–13].

In eukaryotic *N*-glycosylation, glycans are first assembled by glycosyltransferases (GTs) in the cytosol and endoplasmic reticulum (ER), then transferred *en bloc* to the acceptor protein by an oligosaccharyltransferase (OST) in the endoplasmic reticulum, and finally elaborated to final structures as the protein is trafficked through the secretory pathway[14,15]. Thus, unlike the template-driven biosynthesis of DNA, RNA, and proteins, glycan biosynthesis is controlled by the

availability, abundance, and specificities of GTs and other enzymes involved in glycan synthesis and catabolism[16]. Because of the complexity of this multi-compartment, enzymatic process, products of natural protein glycosylation pathways are typically heterogeneous mixtures of glycoforms that can be difficult to isolate from the array of intermediate glycoforms and side products. As a step towards producing more homogeneous glycoprotein products, efforts have been made to better understand, control, and expand glycan synthesis in eukaryotic cell-based systems[17–20]. However, an inherent challenge of engineering existing glycosylation networks in eukaryotic cells is that *N*-linked glycosylation is an essential function, so modifications to these networks for the purpose of altering the target glycoprotein product can have adverse effects on the cell. Thus, even with the availability of powerful genome editing tools such as CRISPR-Cas for glycosylation engineering[21] there are strict limits on the extent of top-down engineering that one can achieve in eukaryotic host cells. As such, there remains a need for alternative methods to produce structurally uniform glycans in sufficient quantities for mechanistic studies and other downstream applications.

To this end, the emerging field of cell-free synthetic glycobiology has helped to expand the glycoengineering toolbox with new methods for synthesizing glycomolecules outside the confines of living cells[22,23]. In these approaches, glycosylation enzymes and substrates are synthesized and assembled *in vitro* to form multistep glycosylation pathways, with the simplest forms involving purified components such that the reaction composition is well-controlled[24]. Alternatively, glycosylation enzymes and substrates can be prepared by cell-free protein synthesis

(CFPS) individually[25] or in a single-pot reaction[26] to circumvent labor- and time-intensive protein purification steps. The advantages of these and other "open" formats for synthesis of glycans and glycoconjugates include enhanced control over reaction conditions, decoupling of glycosylation and protein synthesis from cell viability, and the ability to use enzymes from any and/or multiple host cells in the same system. Moreover, cell-free biomanufacturing is amenable to real time monitoring, automation, and continuous manufacturing systems.

In the context of CFPS, microfluidics offers a unique opportunity to build scaled-down models of integrated protein production systems in a format that enables precise and tunable spatiotemporal control, usage of small volumes that minimize waste, experimentation on length and time scales similar to those in cells, and in-line process monitoring through real-time, high resolution imaging[27,28]. Indeed, microfluidic systems have been shown to improve CFPS in many ways[29]. In particular, protein yields from microfluidic CFPS systems were measurably increased compared to those of traditional one-pot CFPS reactions as a result of greater heat and mass transfer[30] and the exchange of reactants and waste products through dialysis membranes[31] or engineered nanopores[32]. Furthermore, CFPS has been combined with affinity purification in integrated microfluidic systems, enabling efficient protein synthesis and capture[33,34]. With respect to cell-free synthetic glycobiology, there has only been one report describing the use of a microfluidic system in combination with a glycoenzyme[35]. In this seminal work, a digital microfluidics chip was used to merge a droplet containing the soluble GT enzyme D-glucosaminyl 3-*O*-sulfotransferase isoform-1 (3-OST-1) and

its adenosine 3′-phosphate 5′-phosphosulfate (PAPS) cofactor with a second droplet containing heparin sulfate (HS) glycans immobilized on magnetic nanoparticles. Following merging of the droplets on-chip, the HS-nanoparticles became enzymatically sulfated as determined by off-chip analysis of the immobilized HS glycans. To our knowledge, however, there have been no reports of microfluidics-based cell-free protein glycosylation.

Here, we developed a first-in-kind microfluidic device for achieving controllable biosynthesis of glycoproteins, which involved reconfiguring a one-pot method for cell-free glycoprotein synthesis (CFGpS)[26] into a microfluidic architecture. Our prototype involved spatiotemporally separating protein synthesis and protein glycosylation, akin to the subcellular compartmentalization that underlies the biosynthesis of glycoproteins in eukaryotic cells. Specifically, we modeled the cytosol and ER with a modular device that is capable of continuously synthesizing (module 1) and glycosylating (module 2) proteins, after which the post-translationally modified protein products were enriched from the reaction mixture by affinity capture (module 3) (**Fig. 1**). Our results demonstrate that the resulting device was capable of site-specifically glycosylating a model protein, namely superfolder green fluorescent protein (sfGFP), with a bacterial heptasaccharide glycan at a defined C-terminal acceptor site. Importantly, this work represents the first enzymatic glycosylation of a protein substrate in a microfluidic device and a critical first step on the path to building more complex reaction networks for *N*-linked protein glycosylation that more closely mimic the highly coordinated and compartmentalized process in eukaryotic cells.

2.2 Results

2.2.1 Design of a modular microfluidic platform for continuous glycoprotein production

The design of our microfluidic-based glycoprotein biosynthesis platform integrated three key processes: protein expression, protein glycosylation, and protein purification (**Fig. 1**). In the first

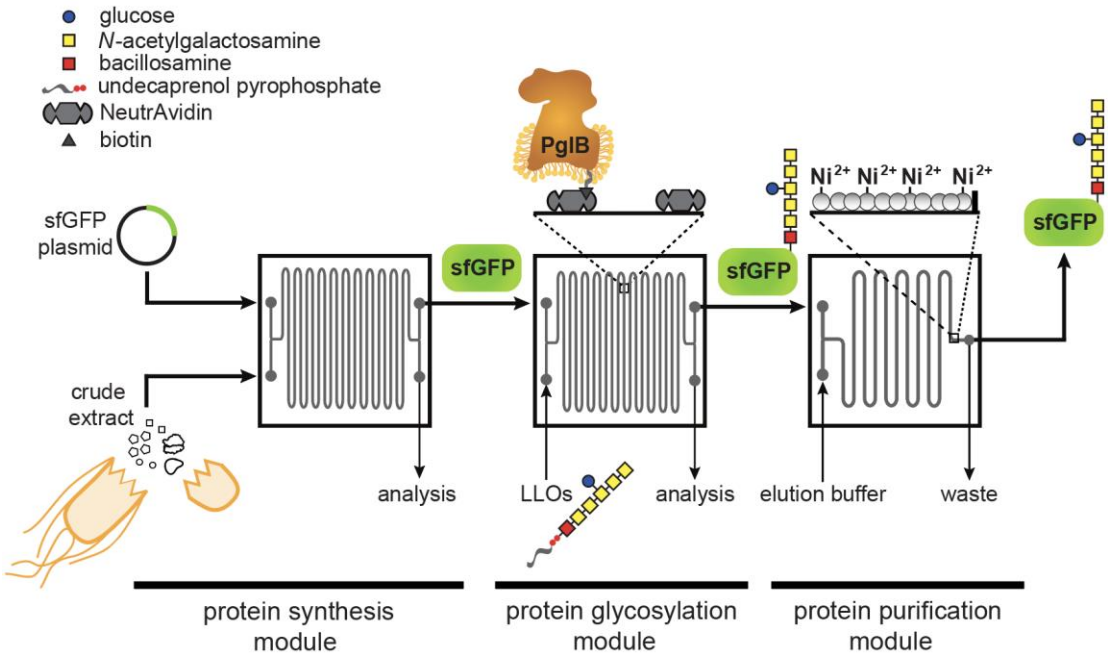


Figure 1. Schematic of glycosylation-on-a-chip system. The microfluidic platform integrates cell-free protein synthesis, glycosylation, and purification. In the first module of the device, one stream containing *E. coli* cell-free extract and a second stream containing plasmid DNA encoding the acceptor protein are combined at the inlet and mixed by diffusion as they travel through the channels. The product of the first chip is then delivered to a second module where it is subjected to an environment enriched with glycosylation machinery. In this case, glycosylation machinery is derived from *C. jejuni* and includes: (i) the OST enzyme, *Cj*PglB, that is tethered to the surface of the device and serves as the conjugating enzyme; and (ii) *Cj*LLOs comprised of undecaprenol-pyrophosphate-linked heptasaccharide from *C. jejuni* as the glycan donor. In the third module, protein product is isolated using immobilized metal affinity capture (IMAC). Depicted is the C. jejuni $GalNAc_5$(Glc)Bac heptasaccharide with reducing end bacillosamine (Bac; red square) followed by five *N*-acetylgalactosamine residues (GalNAc; yellow squares) and a branching glucose (Glc; blue circle). Structure drawn according to symbol nomenclature for glycans (SNFG; https://www.ncbi.nlm.nih.gov/glycans/snfg.html). Schematic created by AA.

module of the device, sfGFP bearing a C-terminal DQNAT glycosylation motif[36] that is optimally recognized by *Cj*PglB[37,38] and a hexahistidine tag was expressed using crude S30 extract derived from *E. coli*, which enabled transcription and translation of the target protein on chip. We chose sfGFP as the acceptor protein so that the protein production and purification processes could be visualized and easily quantified during optimization of the microfluidic system. Next, in the second module, site-specific glycosylation was achieved by subjecting the newly expressed sfGFP$^{DQNAT\text{-}6xHis}$ to components derived from a well-characterized bacterial *N*-linked glycosylation pathway, which occurs natively in the bacterium *C. jejuni* and has been functionally transferred to *E. coli*[2]. These components included *Cj*PglB as the glycan conjugating enzyme and *Cj*LLOs comprised of the *C. jejuni* GalNAc$_5$(Glc)Bac heptasaccharide linked to undecaprenol-pyrophosphate as the glycan donor. *Cj*PglB and its cognate *N*-glycan structure were chosen here for proof-of-concept experiments because of the high transfer efficiency that has been observed with these components both *in vivo*[39,40] and *in vitro*[20,26]. However, in a notable departure from previous works, we sought to site-specifically biotinylate *Cj*PglB and subsequently immobilize it in the device using biotin and streptavidin interactions, thereby enabling reuse of this important membrane protein biocatalyst[41]. Lastly, in the third module, the sfGFP$^{DQNAT\text{-}6xHis}$ product was selectively enriched using a microfluidic device loaded with affinity resin that facilitated reversible capture of the hexahistidine-tagged glycoprotein product. The modularity of the device was designed to enable optimization of each unit operation and to allow flexible biosynthesis of different glycoproteins by simply interchanging acceptor protein target

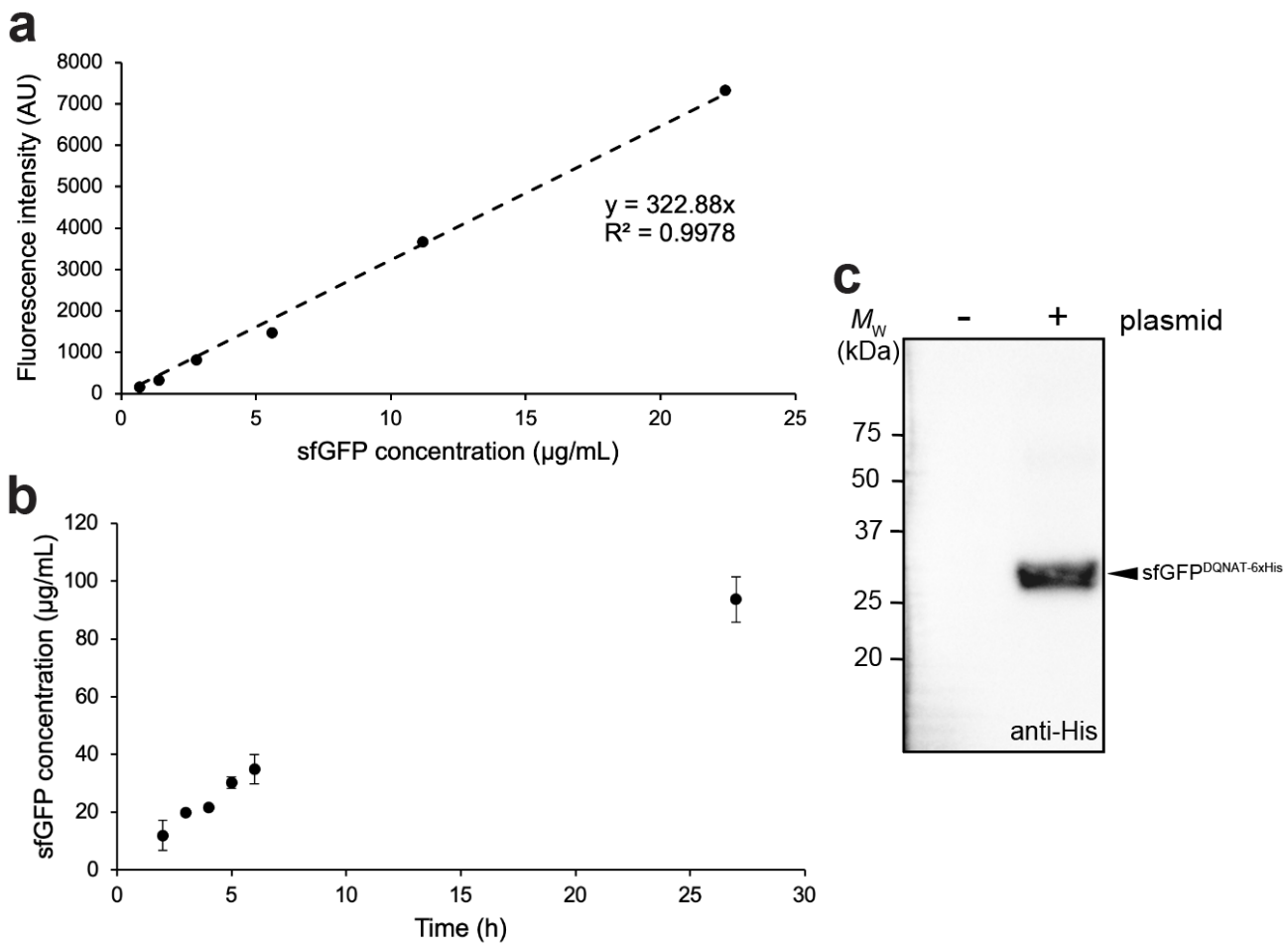
plasmids, glycosylation enzymes, LLO donors, affinity tags, and chromatography resins.

For the microfluidic device design, we aimed to create a system where protein synthesis, glycosylation, and protein purification could happen continuously in series at a fixed flow rate. Therefore, we fabricated individual chips to serve as building blocks that could be serially connected to increase the residence time of a particular process as needed. To test this design, we used an etched silicon wafer as a mold to fabricate channels in polydimethylsiloxane (PDMS) that was subsequently attached to glass slides. We chose PDMS because it enabled low-cost microfluidic fabrication that was sufficiently robust for device prototyping. Each microfluidic chip involved a serpentine channel design (width = 200 μm, depth = 120 μm, volume = 11 μL) that was inspired by previous work in which a similarly designed microfluidic bioreactor resulted in enhanced CFPS productivity[30]. Additionally, we hypothesized that long serpentine channels with a high surface area-to-volume ratio would promote efficient glycosylation by allowing sufficiently high levels of *Cj*PglB enzyme to be tethered to the device, thereby increasing the probability of contact with substrates. For the purification module, an immobilized metal affinity chromatography (IMAC) strategy was implemented whereby 25-μm posts were spaced apart from one another at the outlet of the device and the resulting channels were filled with Ni^{2+}-charged beads for efficient hexahistidine-tagged protein capture.

2.2.2 Continuous-flow CFPS module improves protein production

As a first test of our design, we measured the on-chip protein titers obtained from the

protein synthesis module following two modes of operation—batch and continuous flow—and compared these to the titers produced from one-pot reactions performed in standard microcentrifuge tubes. For these experiments, we generated crude S30 extract from *E. coli* strain BL21 Star^{TM} (DE3) using a low-cost, sonication-based method [42] and the resulting extract was primed with plasmid pJL1-sfGFP$^{DQNAT\text{-}6xHis}$ to drive the expression of sfGFP$^{DQNAT\text{-}6xHis}$. In a standard 15-µL, one-pot CFPS reaction using a microcentrifuge tube, we produced 11.9 µg/mL of sfGFP$^{DQNAT\text{-}6xHis}$ in two hours (**Fig. 2a** and **Supplementary Fig. 1**). To determine how the microfluidic environment affected sfGFP expression, we next performed batch-mode CFPS reactions in the first module of the microfluidic device. Specifically, the device was quickly filled with the same CFPS reaction mixture and fluorescence evolution was monitored in 30-min increments. When all CFPS components were present, fluorescence emission in the microfluidic channels gradually increased over time (**Fig. 2b**), corresponding to production of 9.4 µg/mL of sfGFP$^{DQNAT\text{-}6xHis}$ in two hours (**Fig. 2a**). This result confirmed that the microfluidic environment itself had little-to-no effect on batch-mode CFPS productivity. It is also worth noting that surface blocking within the device was sufficient to allow sfGFP$^{DQNAT\text{-}6xHis}$ clearance from the channels by simple rinsing.



Supplementary Figure 1. Characterization of sfGFP produced by CFPS. (a) Calibration curve correlating sfGFP$^{DQNAT-6xHis}$ concentration to fluorescence intensity. Protein concentration was determined by Bradford assay while fluorescence intensity was measured on a plate reader at 485 nm excitation and 510 nm emission. **(b)** Effect of reaction time on batch-mode CFPS titers of sfGFP$^{DQNAT-6xHis}$. Cell-free extract was used to express sfGFP$^{DQNAT-6xHis}$ from plasmid pJL1-sfGFP$^{DQNAT-6xHis}$ and fluorescence intensity was monitored over time as in (a). The concentration of sfGFP$^{DQNAT-6xHis}$ produced in each reaction was determined by quantifying fluorescence intensity above background at different dilutions of the product to ensure that measurements were in the linear range of the calibration curve. **(c)** Representative immunoblot of CFPS product using cell-free extract with (+) or without (-) addition of plasmid pJL1-sfGFP$^{DQNAT-6xHis}$ to the CFPS reaction mixture. Reactions were incubated overnight at 30°C and products were analyzed by immunoblot analysis in which membrane was probed with anti-polyhistidine (anti-His) antibody. Molecular weight (M_W) ladder is indicated at left of blot. Data were generated and analyzed by AA.

We next investigated the effect of continuous flow on CFPS-based sfGFP expression. To generate a device that could accommodate a two-hour residence time (and thus be directly comparable to the batch-mode experiments above), we created a multi-chip system by linking individual devices with short pieces of tubing. Two input streams, one containing plasmid, energy, salts, and metabolites and the other containing S30 extract and T7 polymerase, met at the inlet and were mixed via diffusion between the two parallel streams as they moved through the channels (**Fig. 2c**). In a four-chip system, corresponding to a two-hour residence time, we observed increasing fluorescence along the length of the channels from the inlet to the outlet corresponding to production of 38.3 μg/mL of $sfGFP^{DQNAT\text{-}6xHis}$ (**Fig. 2a** and **d**). Fluorescence across the width of the channels was uniform, indicating that the solution was well-mixed. Additionally, when comparing the fluorescence generation in two-, three-, and four-chip systems, corresponding to one, one and a half-, and two-hour residence times, respectively, we observed non-linear protein production with the maximum production rate occurring between one and one and a half hours (**Supplementary Fig. 2a** and **b**).

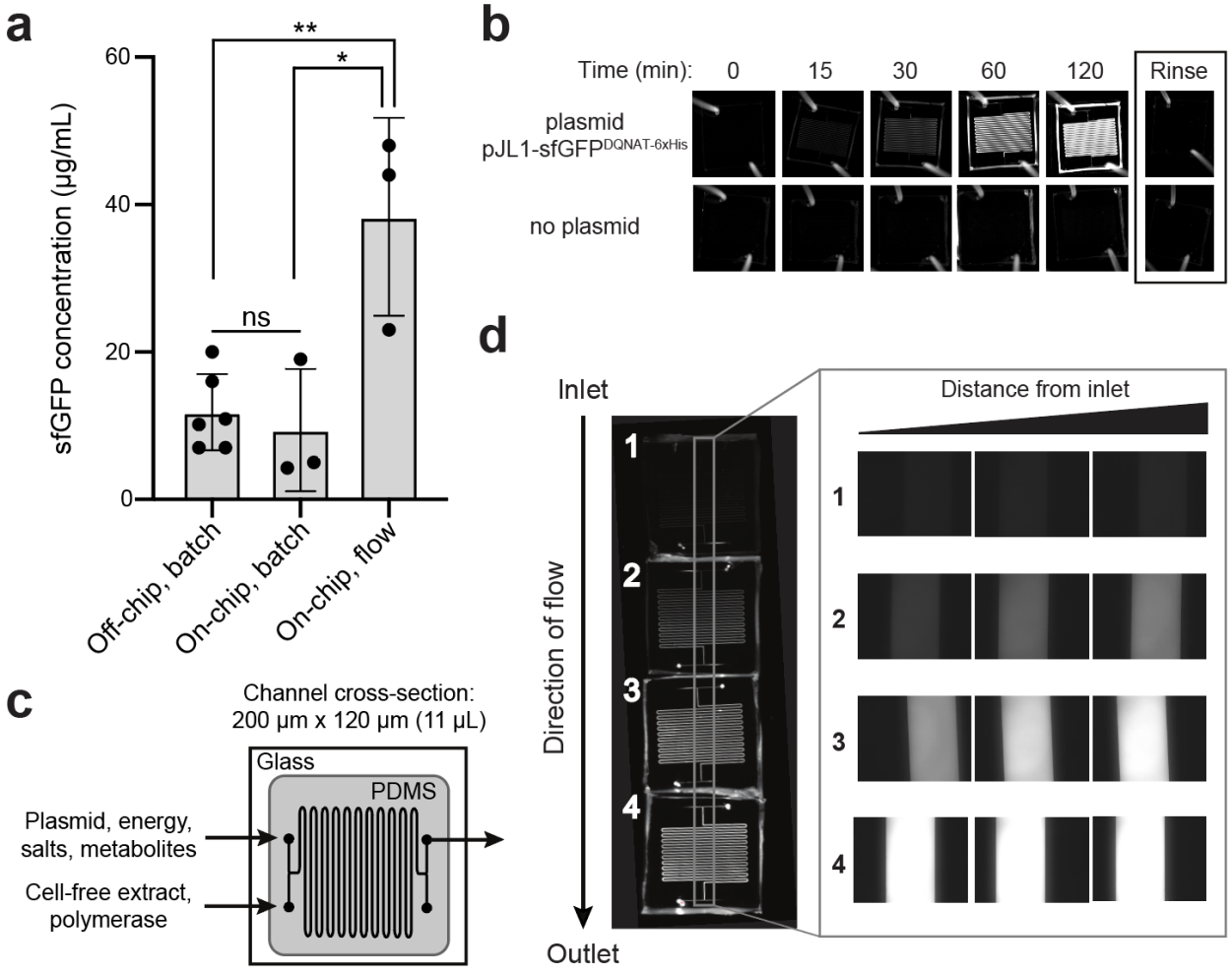


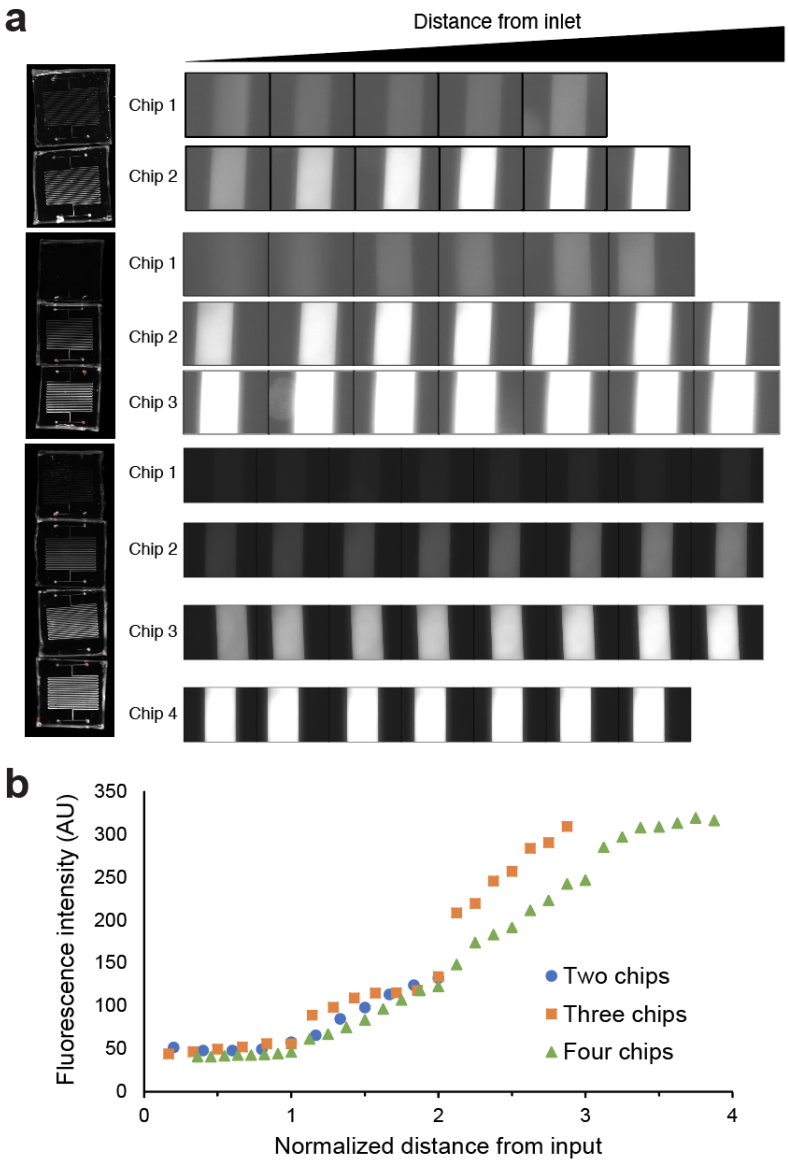
Figure 2. On-chip cell-free protein synthesis. (a) Mean sfGFP$^{DQNAT\text{-}6xHis}$ concentration produced from the following reactions: off-chip, batch mode in a microcentrifuge tube; on-chip, batch mode in the microfluidic device; and on-chip, continuous-flow mode in the microfluidic device. For the on-chip systems, measurements were made on samples collected at the outlet of the chips. Data are the average of biological replicates ($n \geq 3$), error bars represent standard deviation, and p values were determined by paired sample t-test (*, $p < 0.1$; **, $p < 0.01$; and ns, not significant). **(b)** Fluorescence imaging of batch-mode operation in which all CFPS components were mixed, flown into the microfluidic chip, and allowed to react over a two-hour period. Representative images showing sfGFP$^{DQNAT\text{-}6xHis}$ synthesis within the chip (top row) and a control experiment where plasmid was omitted from the CFPS reaction mixture (bottom row). **(c)** Serpentine channel microfluidic design for flow-based CFPS. The flow rate was set so that the reaction residence within each chip was 30 min. Cell-free extract and plasmid DNA were added at separate inlets so that protein synthesis was initiated inside the device. **(d)** Representative fluorescence images of continuous-flow mode in which four chips were linked together for a two-hour reaction residence time and reactants were flown into the two inlets. Inset shows expanded view of the regions within the gray box in the image at left. Cell-free protein synthesis components were generated and verified by AA. Microfluidic devices were fabricated and operated by ZM. Fluorescent images were generated by ZM. CFPS products were analyzed by AA

Importantly, the production rates in equivalent chips were similar, indicating that linking chips in series is a viable method for varying the residence time. Lastly, when comparing the titers of the two modes of on-chip operation relative to the microcentrifuge tube reaction for a two-hour residence time, we observed that sfGFP$^{DQNAT\text{-}6xHis}$ produced on-chip in batch mode was statistically similar to off-chip production in a microcentrifuge tube, whereas introducing flow to the system significantly improved production by several-fold compared to both batch operations (**Fig. 2a**). This increase in production has also been observed by others[30] and can be attributed to shorter diffusion lengths in the microfluidic channels.

2.2.3 Tethered OST enzyme enables a continuous-flow glycosylation module

In the protein glycosylation module, we sought to develop an OST tethering strategy that would allow for efficient protein glycosylation as the reaction substrates—the acceptor protein and LLOs— were continuously flown over the immobilized OST enzymes within the device. The advantage of OST tethering is that it enables creation of a local environment with a high concentration of OST enzyme that is reused in continuous operation. Such a reusable configuration is significant because OSTs are integral membrane proteins that are laborious and time consuming to produce in purified form[43]. For surface immobilization of *Cj*PglB, we leveraged avidin-biotin technology because it afforded the opportunity to site-specifically modify the OST with biotin such that enzymatic activity was minimally affected. To this end, an AviTag was genetically fused to the C-terminus of *Cj*PglB, providing a unique site for covalent biotin conjugation by separately prepared BirA enzyme. Biotinylation of *Cj*PglB was

confirmed by immunoblot analysis using commercial ExtrAvidin-Peroxidase that



Supplementary Figure 2. On-chip CFPS with continuous flow. (a) In continuous-flow mode, two, three, and four chips were linked together for one-, one and a half-, and two-hour reaction residence times, respectively. Fluorescence evolution within the channels corresponding to sfGFP$^{DQNAT\text{-}6xHis}$ production by CFPS was monitored from the inlet to the outlet. Representative fluorescence imaging of the series of microfluidic devices were taken using a ChemiDoc imaging system with blue epi illumination. Images of the channels were taken with a microscope using a blue light source. **(b)** On-chip sfGFP$^{DQNAT\text{-}6xHis}$ production as determined by quantifying mean fluorescence intensity measurements from each

microscope image in (a) plotted as a function of distance from the inlet of chip 1. Cell-free protein synthesis components were generated and products analyzed by AA. Microfluidic devices were fabricated and operated by ZM. Fluorescent images were generated by ZM. Fluorescence data was analyzed by AA.

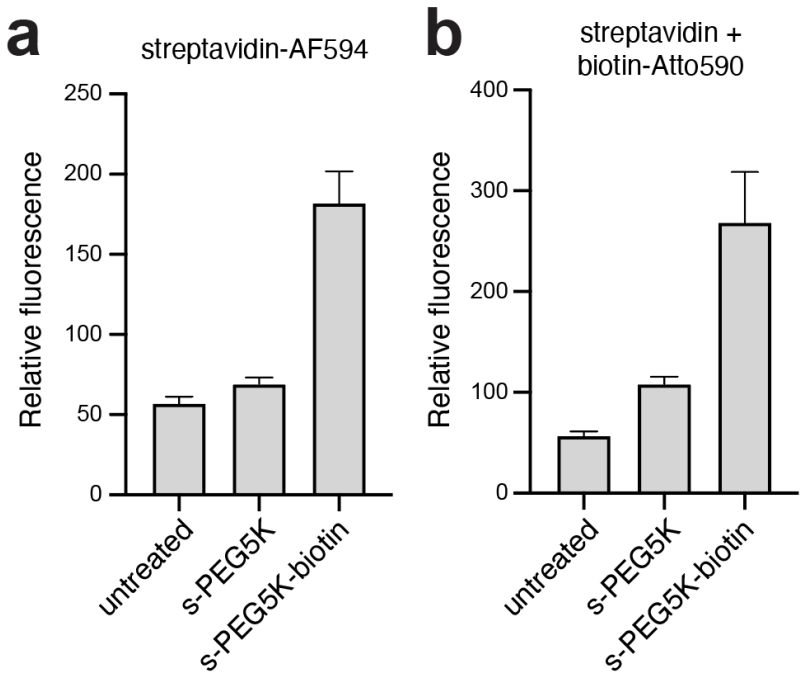
specifically detects biotin (**Fig. 3a**). To verify that enzymatic activity of *Cj*PglB-biotin was not diminished by this modification or subsequent tethering onto a solid support, we performed off-chip *in vitro* glycosylation (IVG) reactions in a microcentrifuge tube using purified sfGFP$^{DQNAT\text{-}6xHis}$ as acceptor protein, *Cj*LLOs as glycan donor, and either untethered *Cj*PglB-biotin or *Cj*PglB-biotin that was tethered to commercial streptavidin beads. Immunoblot analysis of the sfGFP$^{DQNAT\text{-}6xHis}$ produced in these reactions was performed using an anti-His antibody to detect the protein/glycoprotein and hR6 serum that specifically recognizes the *C. jejuni* heptasaccharide. These blots revealed 100% conversion of sfGFP^{DQNAT} to the glycosylated form (g1) in reactions with both untethered and tethered *Cj*PglB-biotin, but only when the microcentrifuge tube for the latter reactions was shaken to keep the beads well suspended in solution (**Fig. 3b**). In the absence of shaking, the beads were observed to sink to the bottom of the microcentrifuge tube so that *Cj*PglB-biotin was not well dispersed within the reaction mixture, thereby reducing glycosylation efficiency as evidenced by the detection of sfGFP^{DQNAT} in a predominantly aglycosylated form (g0). Importantly, these results confirmed that *Cj*PglB tolerated both site-specific biotinylation and tethering to a solid surface without any measurable loss in enzyme activity.

Encouraged by these results, we went on to investigate a strategy for surface tethering of *Cj*PglB-biotin within the channels of our microfluidic device. To provide an evenly distributed, functionalized surface having low non-specific adsorption of other biomolecules, we modified the surface of our device with a silane-PEG5000-

biotin moiety. This molecular weight of PEG has been shown to effectively reduce non-specific binding[44] and to improve surface coverage compared to traditional coupling methods[45]. Here, silane-PEG5000-biotin provided a highly selective binding surface that was observed to promote higher loading capacity compared to non-specific adsorption to non-biotinylated silane-PEG5000 when visualized with fluorescently labeled streptavidin **(Supplementary Fig. 3a**). Comparing the surface coverage of the functionalized PEG brush to that of the non-covalent random adsorption also showed that we had a 30% increase in streptavidin coverage, allowing us to load more enzyme onto the surface of the device. Next, unlabeled NeutrAvidin was immobilized on the silane-PEG5000-biotin surface and was observed to bind fluorescently labeled, free biotin (**Supplementary Fig. 3b**), indicating that unliganded binding sites in surface-bound NeutrAvidin, which has four putative biotin-binding pockets, were available to capture additional biotin groups. Collectively, these experiments confirmed that silane-PEG5000-biotin provided a highly selective, passivating surface that increased binding capacity.

To evaluate this tethering strategy in the context of *Cj*PglB, we coated the channels of our microfluidic device with silane-PEG5000-biotin, followed by the addition of streptavidin and then *Cj*PglB-biotin (**Fig. 3a**). To determine whether immobilization of *Cj*PglB in this manner resulted in a glycosylation-competent device, we first performed on-chip IVG reactions in batch mode without flow. This involved manually pushing IVG reaction components—sfGFP$^{DQNAT\text{-}6xHis}$ and *Cj*LLOs—over *Cj*PglB that was tethered in the microfluidic device. The sfGFP$^{DQNAT\text{-}6xHis}$ product was

collected from the chip and analyzed by immunoblotting, which revealed barely



Supplementary Figure 3. Protein immobilization on functionalized device surfaces. (a) Fluorescence emission from microfluidic devices prepared with either silane-PEG5000 (s-PEG5K) or silane-PEG5000-biotin (s-PEG5K-biotin) following incubation with streptavidin conjugated with Alexa-Fluor594. An untreated device surface (untreated) subjected to the same amount of streptavidin conjugated with Alexa-Fluor594 served as a negative control. **(b)** Same treated and untreated microfluidic devices as in (a) but following a two-step treatment with non-fluorescently labeled streptavidin followed by biotin conjugated with Atto590. Fluorescence measurements in (a) and (b) represent the average and standard deviation of six biological replicates taken as a line scan across the width of the microfluidic channel. All data were normalized to fluorescence emission measured for untreated device. Data were generated and analyzed by ZM.

detectable glycosylation that was significantly less efficient than the glycosylation observed for an on-chip, batch-mode control reaction performed concurrently in a microcentrifuge tube (**Supplementary Fig. 4a**). To determine if continuous flow would remedy this issue, we next flowed the IVG reaction components over the device-tethered *Cj*PglB across a series of chips, each with a reaction residence time of 30 min. In parallel, batch reactions in microcentrifuge tubes were conducted at the same time

for comparison. For these off-chip reactions, we calculated the maximum amount of enzyme that could theoretically be bound to the microfluidic surface and used that amount in the microcentrifuge-based reactions. It should be noted that this amount is likely higher than what is tethered within the device. The sfGFP$^{DQNAT\text{-}6xHis}$ products from these reactions were analyzed by immunoblotting as above, with readily detectable glycosylation occurring in the on-chip, continuous-flow system that was on par in terms of efficiency with the off-chip microcentrifuge reactions (**Fig. 3c** and **Supplementary Fig. 4b**). Interestingly, the addition of flow even appeared to enhance the reaction kinetics, akin to what was observed in the CFPS module.

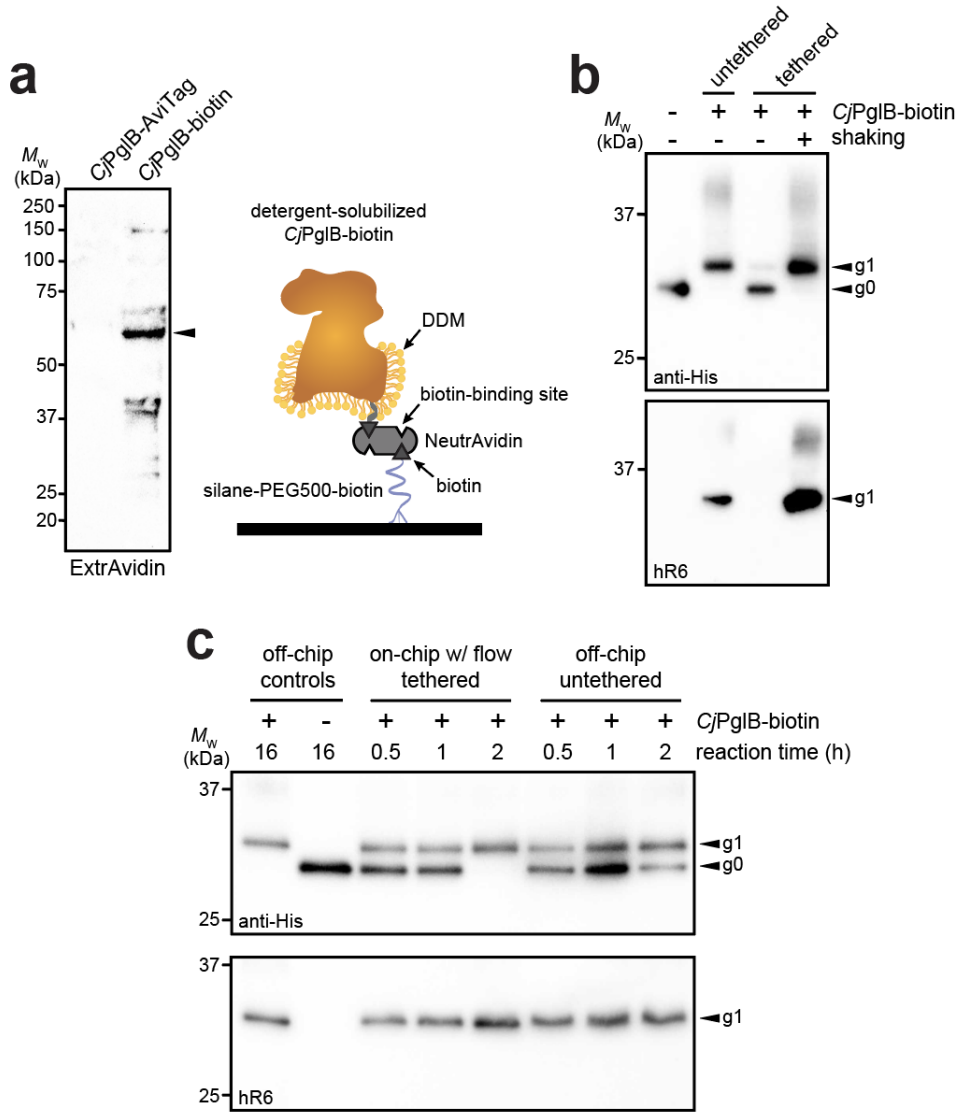
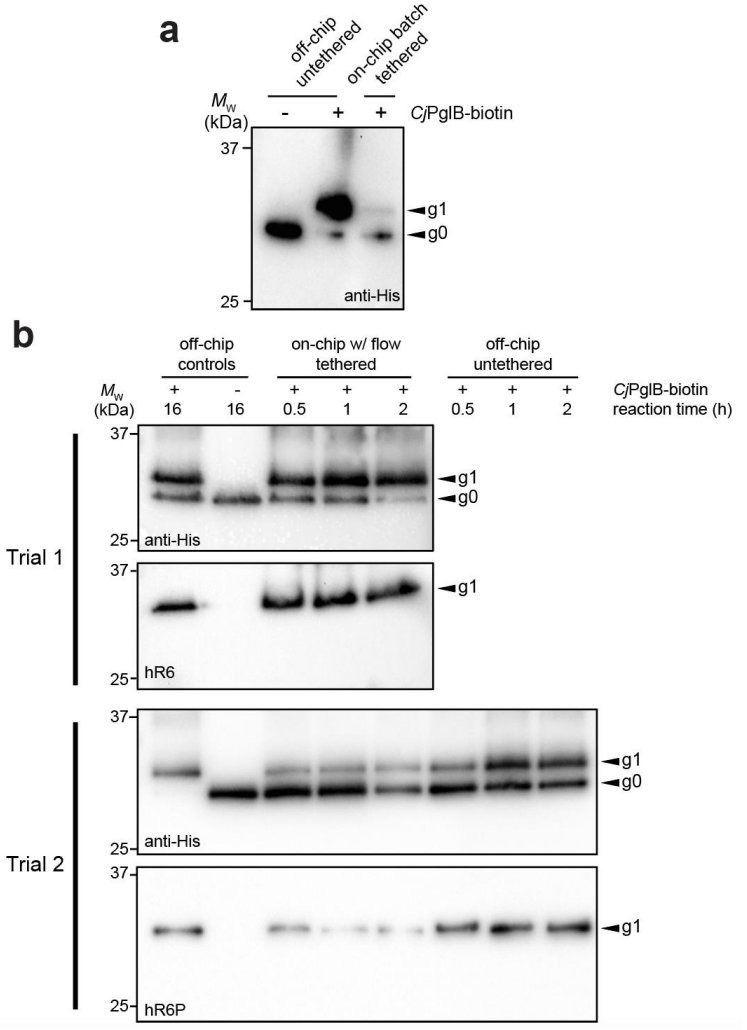


Figure 3. On-chip protein glycosylation. (a) Immunoblot analysis of *Cj*PglB bearing C-terminal AviTag that was subjected to biotinylation by treatment with BirA-containing lysate. Blot was probed with ExtrAvidin-Peroxidase that specifically detects biotin. Arrow denotes the expected molecular weight of *Cj*PglB-biotin. Schematic at right illustrates the tethering strategy used to immobilize *Cj*PglB-biotin generated in (a) within the channels of the microfluidic device. Schematic of *Cj*PglB-biotin tethering system. Silane-PEG5000-biotin was used to modify the surface of hydroxylated glass. Neutravidin, which has four binding sites with high affinity for biotin, was used to link *Cj*PglB-biotin to the surface of the device. **(b)** Immunoblot analysis of IVG reaction products generated in microcentrifuge tubes containing detergent-solubilized *Cj*PglB-biotin (untethered) or detergent-solubilized *Cj*PglB-biotin immobilized on streptavidin-coated beads (tethered). In the case of the latter, batch-mode reactions were performed with (+) or without (-) shaking as indicated. Blots were probed with an anti-polyhistidine (anti-His) antibody that recognized the C-terminal 6xHis tag on sfGFP$^{DQNAT\text{-}6xHis}$ and hR6 serum that specifically recognizes the *C. jejuni* heptasaccharide glycan. **(c)** Immunoblot analysis of IVG reaction products generated using the on-chip, continuous-flow system with detergent-solubilized *Cj*PglB-biotin immobilized in the device

channels (on-chip tethered) or the off-chip microcentrifuge system with detergent-solubilized *Cj*PglB-biotin free in solution (off-chip untethered). For the on-chip system, IVG components were flown through the channels, and the product was collected from the device outlet. Products from overnight microcentrifuge reactions in the presence (+) or absence (-) of *Cj*PglB-biotin were included as controls for glycosylation efficiency. Blots were probed identically as in (b). Arrows in (b) and (c) denote the monoglycosylated (g1) or aglycosylated (g0) sfGFP$^{DQNAT\text{-}6xHis}$ products in each blot. Molecular weight (M_W) markers are indicated at left of all blots. See Supplementary Information for uncropped versions of the images. *In vitro* glycosylation components were generated and products verified by AA. Microfluidic devices were fabricated and operated by ZM. Samples were analyzed by AA and ZM.



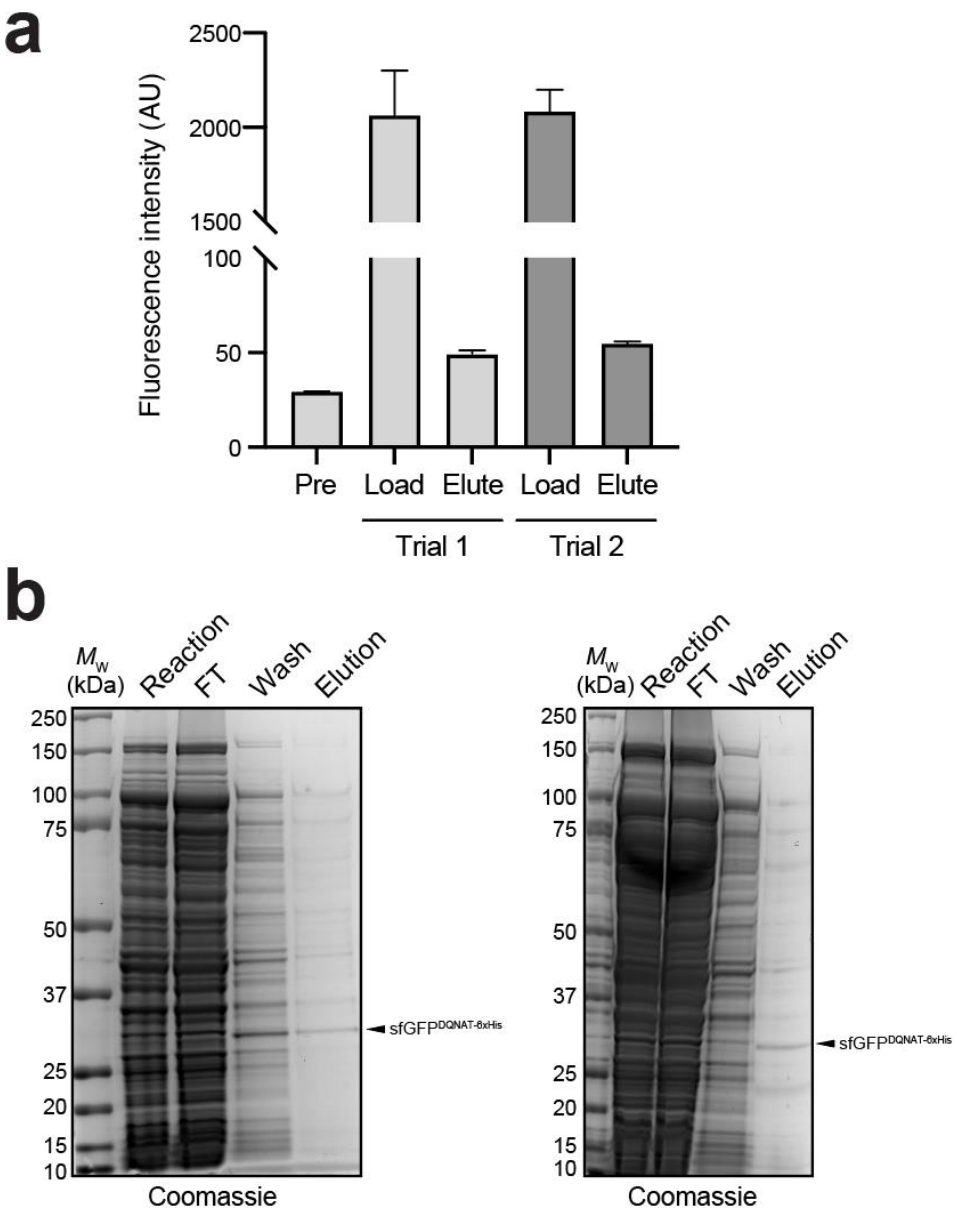
Supplementary Figure 4. On- and off-chip glycosylation in batch and continuous-flow modes of operation. (a) Immunoblot analysis of sfGFP$^{DQNAT\text{-}6xHis}$ glycosylation following either batch mode-mode IVG performed in microfluidic device with surface-tethered *Cj*PglB-biotin (on-chip batch, tethered) or in microcentrifuge tube with *Cj*PglB-biotin in solution (off-chip, untethered). For the latter mode of operation, a control reaction was performed without *Cj*PglB-biotin (-) as indicated. Reactions in all cases were carried out for 16 hours. **(b)** Biological replicates for on-chip continuous-flow mode experiments

performed identically to those described in Figure 3c of the main manuscript. *In vitro* glycosylation components were generated and products verified by AA. Microfluidic devices were fabricated and operated by ZM. Samples were analyzed by AA and ZM.

2.2.4 IMAC module enables continuous enrichment of product proteins

In the third module of our device, we sought to capture polyhistidine-tagged $sfGFP^{DQNAT\text{-}6xHis}$ using an affinity capture strategy. By selectively binding our target protein, unwanted cellular debris, cofactors, and other waste products generated from the upstream reactions can be easily removed by flow-based rinsing. The glycoprotein product can then be recovered by elution with buffer containing a high concentration of imidazole. Using a design based on earlier works[33,46], we prepared a PDMS microfluidic device with posts at the outlet that could be packed with commercial Ni^{2+}-charged beads, thereby enabling on-chip IMAC (**Fig. 4a**). To test this strategy, we attempted to purify $sfGFP^{DQNAT\text{-}6xHis}$ in CFPS reaction mixtures that were flowed through the device with the initial exit stream collected as the flowthrough. Next, we switched the inlet stream to buffer for washing the IMAC resin and removing any non-specifically bound proteins. Finally, we eluted the hexahistidine-tagged protein product using imidazole. The loading and elution steps were monitored by fluorescence imaging of the device (**Supplementary Fig. 5a**) while the composition of each purification fraction was analyzed by SDS-PAGE analysis (**Fig. 4b** and **Supplementary Fig. 5b**). Based on multiple trials, we achieved 78 ± 10% purity in the final product (**Fig. 4b**). This range of purities is to be expected because metal-binding proteins and histidine-rich regions in proteins are naturally present in cells which bind to the nickel resin and elute along with the his-tagged protein of interest. While this purity may not be acceptable for

human therapeutic purposes, a product produced in our system could potentially be used in animal studies if using detoxified cell-free lysate[47]. It should be noted that more complicated device configurations may improve the overall capture efficiency; nonetheless, our results are comparable to other microfluidic capture strategies[34]. To determine the efficiency of product capture, we measured the fluorescence of each fraction and calculated the percent $sfGFP^{DQNAT\text{-}6xHis}$ that was present. While there was some variation in the capture efficiency, we reproducibly captured 45 ± 14% of total produced $sfGFP^{DQNAT\text{-}6xHis}$ (**Fig. 4c**). This simple strategy for protein purification provides a convenient way to obtain a purified final protein product using inexpensive reagents and gentle elution conditions. Because of the modularity of our design, other types of resin (e.g., glycan-binding affinity reagents) could be used in place of, or in addition to the set-up shown here depending on the desired separation. Additionally, multiple devices could be connected for larger scale purifications.



Supplementary Figure 5. On-chip purification of glycosylated sfGFP$^{DQNAT-6xHis}$. (a) Fluorescence imaging of purification module in which Ni^{2+}-charged beads were packed into the device channels and used to capture glycosylated sfGFP$^{DQNAT-6xHis}$. Background fluorescence within the channels was measured before sample loading (Pre) and compared to fluorescence after sfGFP$^{DQNAT-6xHis}$ was flown in the device and beads were thoroughly washed (Load) as well as fluorescence within the channels after 300 mM imidazole buffer was flown through the device (Elute). Trials 1 and 2 were performed subsequently using the same device, indicating that the purification module is reusable. **(b)** Biological replicates for on-chip purification experiments performed identically to those described in Figure 4b of the main manuscript. Arrow indicates the expected molecular weight of sfGFP$^{DQNAT-6xHis}$. Supplementary Figure 5 data were generated by ZM. Cell-free protein synthesis components were generated by AA.

Microfluidic devices were fabricated and operated by ZM.

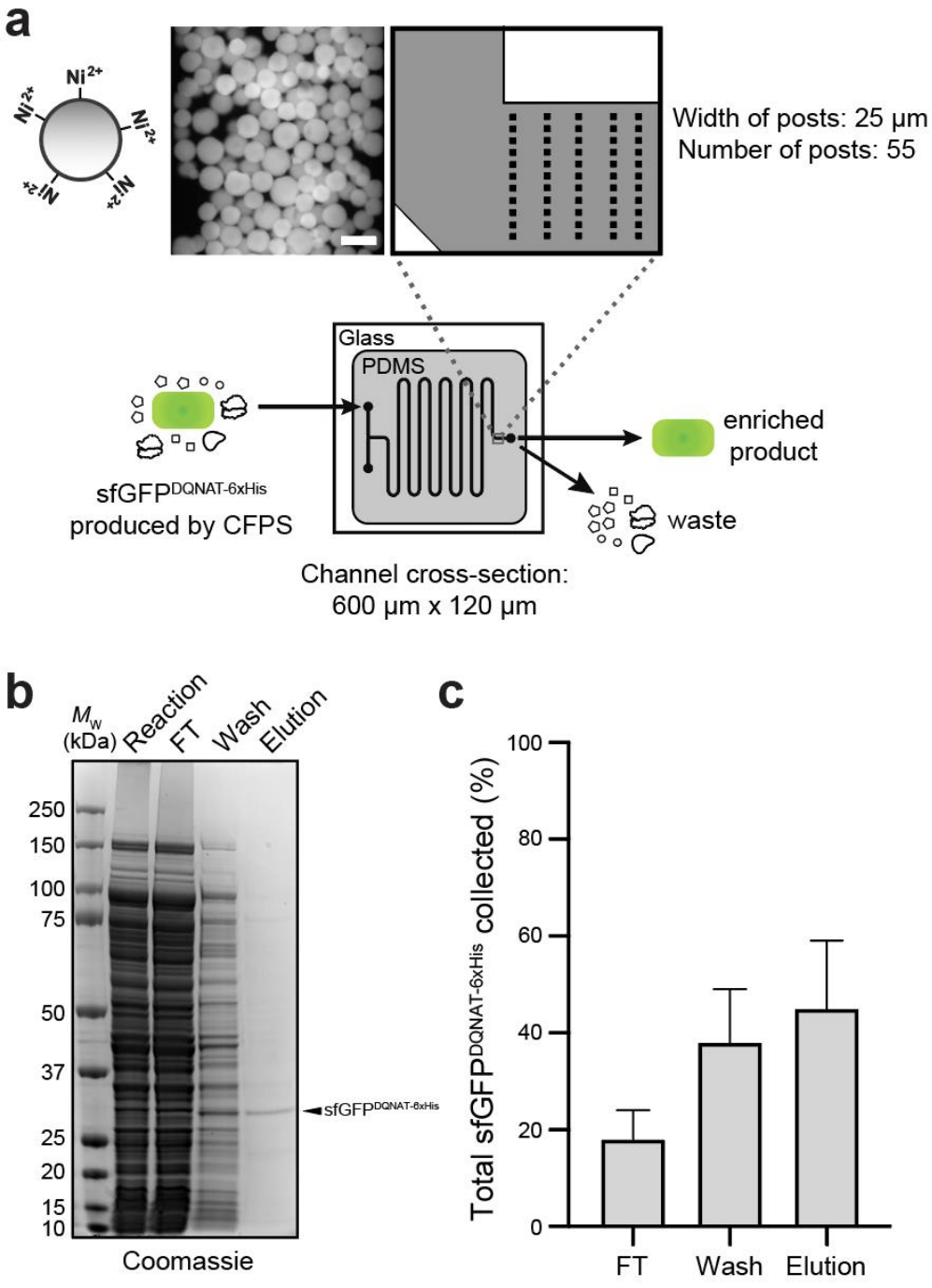


Figure 4. On-chip enrichment of CFPS product. (a) Schematic of purification module. The channels

were designed to be 600-μm wide with fifty-five 25-μm posts at the outlet to accommodate Ni^{2+}-functionalized beads. Shown at left is a representative fluorescence microscopy image of Ni^{2+}-charged beads bound to $sfGFP^{DQNAT\text{-}6xHis}$ within the device. After completion of CFPS reaction, product is pushed through the beads to allow for hexahistidine-tagged protein to bind to Ni^{2+} and flowthrough (FT) fraction is collected. Beads are then washed to remove any non-specifically bound proteins and wash fraction is collected. Finally, protein product is recovered through addition of buffer containing high concentration of imidazole and collected as elution fraction. **(b)** Representative Coomassie-stained SDS-PAGE gel comparing the protein composition of purification fractions as indicated. Arrow denotes the expected molecular weight of $sfGFP^{DQNAT\text{-}6xHis}$. Molecular weight ($M_W$) ladder is indicated at left. **(c)** Comparison of the amount of $sfGFP^{DQNAT\text{-}6xHis}$ in each purification fraction represented as percentage of the total amount of $sfGFP^{DQNAT\text{-}6xHis}$ collected. Data are the mean of biological replicates (n = 4) and error bars represent standard error of the mean. See Supplementary Information for uncropped versions of the images. Figure 4 data were generated by ZM. Cell-free protein synthesis components were generated by AA. Microfluidic devices were fabricated and operated by ZM. Purification fractions were analyzed by ZM.

2.3 Discussion

In this work, we designed and fabricated a microfluidic platform for flow-based, cell-free production of a model *N*-linked glycoprotein. This was accomplished in a modular system where protein synthesis, glycosylation, and purification were compartmentalized and individually optimized. In this approach, production rates were increased for continuous-flow processes compared to batch processes and protein production occurred at a faster rate than glycosylation. Importantly, the device was capable of glycosylating 100% of the added acceptor protein within two hours. For the device design, we used PDMS as an inexpensive material for prototyping the system and commercially available reagents in our enzyme tethering strategy. We believe this will simplify adoption of this strategy in other laboratories for testing flow-based glycoprotein manufacturing systems. One of the most significant developments in this work was the demonstration that the pivotal glycosylation catalyst, *Cj*PglB, could be successfully immobilized within the device while maintaining high glycosylation efficiency. As a multi-pass transmembrane protein with regions in the membrane

portion that are required for activity[48], *Cj*PglB is challenging to express and purify; hence, the opportunity to reuse this enzyme in a continuous fashion should help to relieve a major bottleneck related to mechanistic studies of this enzyme and its biotechnological exploitation. Furthermore, the ability to achieve 100% glycosylation efficiency within the device allowed the glycoprotein product to be purified in a single step using IMAC. We anticipate that for less efficiently glycosylated proteins, an additional purification step using immobilized lectins or antibodies that specifically bind to the glycan could be implemented for glycoprotein enrichment.

For the proof-of-concept studies performed herein, we selected the *C. jejuni N*-linked glycosylation system as a model because of the flexibility of *Cj*PglB as a stand-alone, single-subunit OST[41] that has proven to be compatible with a diverse array of glycan donors and acceptor protein substrates including some with therapeutic potential. To date, *Cj*PglB has been used to generate glycoproteins bearing bacterial[26,49,50] and smaller human-type glycans[26,51–54], and has enabled cell-free, one-pot systems for making *N*- and *O*-linked glycoproteins[26,55] as well as antibacterial conjugate vaccines[47]. In many of these cases, glycosylation was achieved through modification of C-terminal or surface-accessible internal sites that could be post-translationally modified in a cell-free glycosylation system. When transitioning to the production of human therapeutic glycoproteins, an important consideration is that in some instances, a native internal glycosylation site will be buried when the protein folds and thus it will become necessary to more seamlessly integrate protein synthesis and glycosylation within the same compartment, or else slow protein folding, such that glycans can be installed on

proteins prior to the completion of the folding process.

While not directly demonstrated in this work, cell-free strategies such as the glycosylation-on-a-chip platform described here could eventually provide access to glycoproteins that are modified with larger, complex-type *N*-glycans that mimic the structures commonly found on many human glycoprotein drugs such as monoclonal antibodies. This could be achieved by one-step *en bloc* transfer of fully assembled complex-type *N*-glycans or could instead be subdivided into discrete, compartmentalized modules. For example, we previously developed methods for *Cj*PglB-mediated transfer of the eukaryotic trimannosyl core *N*-glycan, mannose$_3$-*N*-acetylglucosamine$_2$ ($Man_3GlcNAc_2$), onto acceptor proteins both *in vivo* and *in vitro*[26,53]. The on-chip transfer of preassembled $Man_3GlcNAc_2$ glycans onto acceptor protein targets could serve as a first modular step that could be followed in subsequent modules by a series of immobilized GTs for elaborating the protein-linked $Man_3GlcNAc_2$ to discrete human-like *N*-glycan structures[56]. Alternatively, the ability of *Cj*PglB to transfer a single *N*-acetylglucosamine (GlcNAc) or diGlcNAc structure onto a target peptide[57] provides a minimal glycan primer that could serve as an earlier starting point for single-enzyme transglycosylation using synthetic oligosaccharide oxazolines as donor substrates[58] or multi-enzyme, cell-free glycan construction[25]. Importantly, our demonstration that *Cj*PglB can be immobilized in a microfluidic architecture without loss of catalytic activity is a critical first step to enabling any of these advanced strategies and paves the way for continuous production of a variety of therapeutically relevant glycoprotein products.

There has been increasing interest in the pharmaceutical industry to implement continuous manufacturing technologies that afford greater control over reaction variables, are amenable to automation, and are more flexible to changes in market demand compared to batch reactors[59–61]. Therefore, as a scaled-down model of flow-based systems, many researchers have investigated the use of microfluidic devices as microreactors for organic synthesis of pharmaceuticals[62,63]. Although biopharmaceuticals represent almost half of newly FDA approved therapeutics[64], production of these more complex molecules by chemical means for incorporation into flow systems has been limited. Hence, our work expands the capability of microfluidic systems to now include production of *N*-linked glycoproteins by leveraging the open-box format of cell-free systems in a manner that provides spatiotemporal control over reactions, residence times, and concentrations. Looking forward, we anticipate that the flow-based glycoprotein production platform established here will inspire deeper exploration of cell-free technologies for continuous biomanufacturing of biologics.

2.4 Materials and Methods

Bacterial strains and plasmids

E. coli strain DH5α (lab stock) was used for all molecular biology. *E. coli* strain BL21 Star™ (DE3) (Novagen) was used for expression and purification of sfGFP containing a C-terminal glycosylation tag[36] and polyhistidine tag (sfGFP$^{DQNAT\text{-}6xHis}$), which was used for *in vitro* glycosylation reactions. *E. coli* strain BL21 Star™ (DE3) was also used for expression of the enzyme BirA, which was used for biotinylation of the *Campylobacter jejuni* OST enzyme PglB (*Cj*PglB), and for preparing crude S30

extract. *E. coli* strain CLM24[49] was used for expression and purification of *Cj*PglB while *E. coli* strain SCM6[2] was used for preparation of lipid-linked oligosaccharides bearing *C. jejuni* heptasaccharide glycans (*Cj*LLOs) .

For both cell-free and cell-based expression of sfGFP$^{DQNAT\text{-}6xHis}$, the pJL1-sfGFP$^{DQNAT\text{-}6xHis}$ plasmid[26] was used. Plasmid pTrc99a-BirA (lab stock) was used for expression of the BirA enzyme. Plasmid pSPI01A-*Cj*PglB encoding *Cj*PglB with a C-terminal AviTag was constructed as follows. First, the *Cj*PglB^{10xHis} gene was PCR amplified from plasmid pSN18[38] and the resulting PCR product was then ligated between the NdeI and EcoRI restriction sites in plasmid pSPI01A[65], a vector containing the AviTag after the EcoRI cut site. All plasmids were confirmed by DNA sequencing at the Biotechnology Resource Center of the Cornell Institute of Biotechnology.

Protein expression, biotinylation, and purification

Preparation of lysates containing *Cj*PglB with a C-terminal AviTag was performed according to previously published methods[20,26]. Briefly, a colony of *E. coli* CLM24 carrying plasmid pSPI01A-*Cj*PglB was grown overnight in 5 mL of Luria-Bertani (LB) media supplemented with chloramphenicol. Overnight cultures were then subcultured into 1 L of terrific broth (TB; 24 g/L yeast extract, 12 g/L tryptone, 8 mL glycerol, 10% (v/v) 0.72 M K_2HPO_4/0.17 M KH_2PO_4 buffer) supplemented with chloramphenicol. Cells were grown at 37°C until an optical density at 600 nm (OD_{600}) of ~0.6 and then induced with 100 μM isopropyl β-D-1-thiogalactopyranoisde (IPTG) for 18 h at 16°C. Cells were harvested by centrifugation, after which the pellet was resuspended in Buffer 1 (25 mM TrisHCl, 250mM NaCl, pH 8.5) and lysed using a C5

Emulsiflex homogenizer (Avestin). The lysate was centrifuged to remove cellular debris and the supernatant was ultracentrifuged at 120,000 × g for 1 h at 4°C. The resulting pellet was manually resuspended using a Potter-Elvehjem tissue homogenizer into Buffer 2 (25 mM TrisHCl, pH 8.5, 250 mM NaCl, 1% (w/v) n-dodecyl-β-D-maltoside (DDM), and 10% (v/v) glycerol). Once fully resuspended, the solution was rotated at room temperature to facilitate solubilization of the protein and then ultracentrifuged again at 120,000 × g for 1 h at 4°C.

To prepare BirA-containing lysate, BL21(DE3) cells carrying pTrc99a-BirA were grown overnight and then subcultured into 250 mL of LB media supplemented with kanamycin. Upon reaching an OD_{600} of ~0.6, cells were induced with 100 µM IPTG for 18 h at 30°C. Cells were harvested, resuspended in Buffer 1, lysed by homogenization, and centrifuged to remove cellular debris. To prepare biotinylated *Cj*PglB (*Cj*PglB-biotin), *Cj*PglB-containing lysate was mixed with BirA-containing lysate and 5 mM biotin, 10 mM $MgCl_2$, 10 mM ATP, and 1 EDTA-free protease inhibitor cocktail tablet (Thermo Scientific). The mixture was rotated overnight at 4°C to allow time for biotinylation. *Cj*PglB-biotin was then enriched using HisPur Ni-NTA resin (Thermo Scientific) according to manufacturer's recommendations and the elution fraction was desalted with buffer containing 50 mM HEPES, pH 7.5, 100 mM NaCl, 5% (v/v) glycerol, and 0.05% (w/v) DDM.

To prepare $sfGFP^{DQNAT\text{-}6xHis}$, BL21(DE3) cells carrying plasmid pJL1-$sfGFP^{DQNAT\text{-}6xHis}$ were grown overnight and subcultured in LB media supplemented with kanamycin. Upon reaching an OD_{600} of ~0.6, cells were induced with 100 µM

IPTG for 18 h at 30°C. Cells were collected, resuspended in buffer containing 50 mM NaH_2PO_4, pH 8, and 300 mM NaCl and lysed as above. The $sfGFP^{DQNAT\text{-}6xHis}$ was purified using HisPur Ni-NTA resin as above. The final protein was desalted using buffer containing 20 mM HEPES, pH 7.5, 500 mM NaCl, and 1 mM EDTA.

Solvent extraction of CjLLOs

*Cj*LLOs were prepared by organic solvent-based extraction according to a protocol that was adapted from previous methods[26,66]. Briefly, SCM6 cells carrying plasmid pMW07-pglΔB[39] were grown overnight in LB media supplemented with chloramphenicol. Cells were then subcultured into 1 L of TB media, grown at 37°C until reaching an OD_{600} of ~0.7, then induced with a final concentration of 0.2% (w/v) *L*-arabinose for 16 h at 30°C. After induction, cells were harvested by centrifugation, the pellet re-suspended in methanol, and the cells dried for two days at room temperature. After drying, the cells were collected and subsequently suspended in 12 mL 3:2 (v/v) chloroform:methanol, 20 mL water, and 18 mL 10:10:3 (v/v/v) chloroform:methanol:water. After each step, sonication was used to facilitate extraction of LLOs. After the first two sonication steps, centrifugation was used to separate shorter sugars and water-soluble compounds in the supernatant from the pellet. After the final step, centrifugation was used to pellet the cellular debris and the supernatant was collected and dried at room temperature. After drying, the LLOs were resuspended in buffer containing 10 mM HEPES, pH 7.5, and 0.01% (w/v) DDM and stored at -20°C.

Fabrication of microfluidic devices

Microfluidic masters were made on silicon wafers according to standard photolithography protocols at the Cornell NanoScale Science and Technology Facility. Briefly, SPR220-7.0 photoresist was spun onto silicon wafers and exposed using an ABM Contact Aligner. Wafers were developed using Microposit MIF 300. Coated wafers were etched to the desired depth using a Unaxis 770 Deep Silicon Etcher, which was confirmed by using a Tencor P10 profilometer. Remaining photoresist was removed via plasma cleaning, and a coating of (1H,1H,2H,2H-perfluorooctyl) trichlorosilane (FOTS) was applied using a MVD-100 to allow for easy removal of polydimethylsiloxane (PDMS). Microfluidic devices were made by pouring degassed PDMS (mixed 1:10 with crosslinker) and curing for 5 h at 60°C. PDMS molds were cleaned with ethanol and MilliQ water, before being dried with nitrogen gas. Final devices were assembled after oxygen plasma cleaning at 700 μm for 25 sec and sealed with a Piranha washed (70/30 (v/v) H_2SO_4/ H_2O_2 for 10 min) glass coverslip. Devices were placed in a 70°C oven for 10 min to promote bonding of the PDMS to the glass.

Cell-free protein synthesis

S30 crude extracts for CFPS reactions were prepared using a simple sonication-based method[42]. Briefly, BL21(DE3) cells were grown in 1 L of 2xYTPG media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 7 g/L K_2HPO_4, 3 g/L KH_2PO_4, 20 g/L glucose) and harvested upon reaching an OD_{600} of ~3.0. Cell mass was washed three times in Buffer A (10 mM Tris-acetate, pH 8.2, 14 mM magnesium acetate, 60 mM potassium glutamate and 2 mM dithiothreitol) then resuspended in a ratio of 1 mL of

Buffer A to 1 g wet cell mass. The resuspended cells were sonicated with an optimal energy input (reported based on the volume obtained after resuspending cells) and centrifuged at 30,000 × g to obtain S30 extract, and the supernatant stored at -80°C. No run-off reaction was needed for the BL21(DE3) extract.

CFPS reactions consisted of a mixture of components at a final concentration of 13 ng/μL plasmid DNA, 40% (v/v) S30 crude extract, 1.2 mM adenosine triphosphate (ATP), 0.85 mM guanosine triphosphate (GTP), 0.85 mM uridine triphosphate (UTP), 0.85 mM cytidine triphosphate (CTP), 34 μg/mL L-5-formyl-5, 6, 7, 8-tetrahydrofolic acid (folinic acid); 170 μg/mL of *E. coli* tRNA mixture, 130 mM potassium glutamate, 10 mM ammonium glutamate, 12 mM magnesium glutamate, 2 mM each of 20 amino acids, 0.33 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme-A (CoA), 1.5 mM spermidine, 1 mM putrescine, 4 mM sodium oxalate, 33 mM phosphoenolpyruvate (PEP), 100 μg/mL T7 RNA polymerase.

For CFPS in a microcentrifuge tube, 15-μL reactions were conducted in 1.5-mL microtubes in a 30°C incubator. For CFPS on-chip batch reactions, CFPS reaction mixtures were manually inserted into the microfluidic device using a syringe and incubated at 30°C in a moist environment to prevent evaporation. For CFPS on-chip reactions with continuous flow, two mixtures were prepared— one containing S30 crude extract and T7 RNA polymerase and the other containing the rest of the CFPS components— that when combined contained all components diluted to the final concentrations of a standard CFPS reaction. These mixtures were flown into the microfluidic device using a syringe pump where the reactants had a total residence time

in each chip of 30 min.

Preparation of functionalized surfaces

Silane-PEG5000-biotin (Nanocs, Inc.) was dissolved in 95% (w/w) ethanol in water according to the manufacturer's recommendations. The solution was manually pushed into the microfluidic devices and left to react for 2 h at room temperature. Devices were flushed with 100 μL of MilliQ water and then PBS at a flowrate of 10 μL/min. A solution of 100 μg/mL NeutrAvidin (Thermo Scientific) was then introduced and left to bind to the surface for 1 h. For loading of purified *Cj*PglB-biotin, the devices were rinsed with PBS and then buffer containing 50 mM HEPES, 100 mM NaCl, 5% (v/v) glycerol, and 0.01% (w/v) DDM at pH 7.5. The purified *Cj*PglB-biotin was then introduced into the device and allowed to bind overnight at 4°C; unbound enzyme was rinsed away before use.

On- and off-chip glycosylation

For off-chip *in vitro* glycosylation (IVG) reactions, mixtures consisted of components at a final concentration of 17 μg/mL of purified $sfGFP^{DQNAT\text{-}6xHis}$, 170 μg/mL solvent-extracted *Cj*LLOs, 10 mM $MnCl_2$, and 0.1% (w/v) DDM. For microcentrifuge tube reactions, IVG reaction mixtures were supplemented with 170 μg/mL purified *Cj*PglB-biotin to a final volume of 30 μL and reactions were conducted in 1.5 mL microtubes in a 30°C incubator.

For on-chip glycosylation experiments, purified *Cj*PglB-biotin was immobilized on the functionalized surface of the device and the IVG reaction mixture was

continuously pushed through the channels using a syringe pump with a total residence time of 30 min per chip. The reaction was heated by placing the microfluidic chip on a hot plate to maintain the internal temperature of the device at 30°C and confirmed by using a thermocouple in a similar arrangement. The product was then collected at the outlet of the device and either saved for analysis or recirculated through the device again to measure the effect of increasing residence times.

On-chip purification

The microfluidic device used for protein purification was designed with a series of posts at the end of the channel to entrap chromatography resin in the channel. For each device, we manually introduced Ni-charged profinity resin (Bio-Rad) into the channels before use. CFPS reactions expressing sfGFP$^{DQNAT\text{-}6xHis}$ were then introduced into the inlet of the device with a total residence time of 30 min per chip and the outlet was collected and analyzed as the flowthrough fraction. The device was rinsed with PBS containing 10 mM imidazole at a flowrate of 2 µL/min and any unbound protein was collected and analyzed as the wash fraction. Finally, the target protein was eluted from the resin with PBS containing 300 mM imidazole at a flow rate of 2 µL/min. The fluorescence of each fraction was analyzed using a microplate reader to determine the sfGFP$^{DQNAT\text{-}6xHis}$ concentration and assayed for purity using standard SDS-PAGE with Coomassie blue staining.

Immunoblot analysis

For immunoblot analysis of IVG products and *Cj*PglB-biotin, samples were

diluted 3:1 in 4× NuPAGE LDS sample buffer (Invitrogen) supplemented with 10% beta-mercaptoethanol (v/v). IVG products were boiled at 100°C for 10 min while *Cj*PglB-biotin samples were held at 65°C for 5 min. The treated samples were subjected to SDS-polyacrylamide gel electrophoresis on Bolt™ 12% and 4-12% Bis-Tris Plus Protein Gels (Invitrogen). The separated protein samples were then transferred to polyvinylidene difluoride (PVDF) membranes. Following transfer, the membranes containing IVG samples were blocked with 5% (w/v) milk in TBST (TBS, 0.1% (v/v) Tween 20) and then probed with horseradish peroxidase (HRP) conjugated anti-His antibody (1: 5,000) (Abcam, catalog # ab1187) or rabbit polyclonal serum, hR6, that is specific for the *C. jejuni* heptasaccharide glycan (1:10,000) (kindly provided by Markus Aebi) for 1 h. To detect hR6 serum antibodies, goat anti-rabbit IgG conjugated to HRP (1:5,000) (Abcam, catalog # ab205718) was used as the secondary antibody. The membranes containing *Cj*PglB-biotin samples were blocked with 5% (w/v) bovine serum albumin (BSA) in TBST and then probed with ExtrAvidin-Peroxidase (1:10,000) (MilliporeSigma, catalog # E2886) for 1 h. After washing five times with TBST for 5 min, the membranes were visualized with Clarity ECL substrate (Bio-Rad) using a ChemiDoc™ MP Imaging System (Bio-Rad).

2.5 Acknowledgements

AA and ZM are co-first authors of the manuscript and contributed equally to the experimental design, generation of data, and data analysis. All authors contributed to project conceptualization, writing, and editing.

We thank Markus Aebi for providing strains CLM24 and SCM6 as well as hR6

serum used in this work. The authors also thank the Genomics Facility of the Biotechnology Resource Center at the Cornell Institute of Biotechnology for help with sequencing experiments and Thapakorn Jaroentomeechai, Han-Yuan Liu, Weston Kightlinger, and Mike Jewett for helpful discussions related to the manuscript.

This work was supported by the Defense Threat Reduction Agency (HDTRA1-15-10052 and HDTRA1-20-10004 to MD), National Science Foundation (CBET-1159581, CBET-1264701, CBET-1936823, and MCB-1413563 to MD; CMMI-1728049 to SD and MD), and National Institutes of Health (1R01GM127578 to MD). AA was supported by the National Science Foundation Graduate Research Fellowship (Grant No. DGE-1650441). AA and ZM were supported by a Chemical-Biology Interface (CBI) training grant from the National Institute of General Medical Sciences of the National Institutes of Health (T32GM138826). ZM was supported by a School of Chemical and Biomolecular Engineering Fleming Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. This work was performed in part at the Cornell NanoScale Facility, a member of the National Nanotechnology Coordinated Infrastructure (NNCI), which is supported by the National Science Foundation (Grant No. NNCI-2025233).

References

1. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).
2. Dube, D. H. & Bertozzi, C. R. Glycans in cancer and inflammation - Potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **4**, 477–488 (2005).
3. Abu-Qarn, M., Eichler, J. & Sharon, N. Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* **18**, 544–550 (2008).
4. Lin, C. W. *et al.* A common glycan structure on immunoglobulin G for enhancement of effector functions. *Proc. Natl. Acad. Sci. U. S. A.* (2015) doi:10.1073/pnas.1513456112.
5. Imperiali, B. & O'Connor, S. E. Effect of N-linked glycosylatian on glycopeptide and glycoprotein structure. *Current Opinion in Chemical Biology* (1999) doi:10.1016/S1367-5931(99)00021-6.
6. Wolfert, M. A. & Boons, G. J. Adaptive immune activation: Glycosylation does matter. *Nature Chemical Biology* (2013) doi:10.1038/nchembio.1403.
7. Hebert, D. N., Lamriben, L., Powers, E. T. & Kelly, J. W. The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nature Chemical Biology* (2014) doi:10.1038/nchembio.1651.
8. Beckham, G. T. *et al.* Harnessing glycosylation to improve cellulase activity. *Curr. Opin. Biotechnol.* **23**, 338–345 (2012).
9. Wang, L. X., Tong, X., Li, C., Giddens, J. P. & Li, T. Glycoengineering of antibodies for modulating functions. *Annu. Rev. Biochem.* **88**, 433–459 (2019).
10. Berti, F. & Adamo, R. Antimicrobial glycoconjugate vaccines: An overview of classic and modern approaches for protein modification. *Chem. Soc. Rev.* **47**, 9015–9025 (2018).
11. Van Landuyt, L., Lonigro, C., Meuris, L. & Callewaert, N. Customized protein glycosylation to improve biopharmaceutical function and targeting. *Curr. Opin. Biotechnol.* **60**, 17–28 (2019).
12. Helenius, A. & Aebi, M. Intracellular functions of N-linked glycans. *Science* (2001) doi:10.1126/science.291.5512.2364.
13. Nilsson, T., Au, C. E. & Bergeron, J. J. M. Sorting out glycosylation enzymes in the Golgi apparatus. *FEBS Lett.* **583**, 3764–3769 (2009).
14. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyl transferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
15. Hossler, P., Khattak, S. F. & Li, Z. J. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* **19**, 936–949 (2009).
16. Butler, M. & Spearman, M. The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr. Opin. Biotechnol.* **30**, 107–112 (2014).
17. Bosch, D. & Schots, A. Plant glycans: Friend or foe in vaccine development? *Expert Rev. Vaccines* **9**, 835–842 (2010).
18. Guarino, C. & Delisa, M. P. A prokaryote-based cell-free translation system that

efficiently synthesizes glycoproteins. *Glycobiology* **22**, 596–601 (2012).
19. Chang, M. M. *et al.* Small-molecule control of antibody N-glycosylation in engineered mammalian cells. *Nat. Chem. Biol.* **15**, 730–736 (2019).
20. Jaroentomeechai, T. *et al.* Cell-free synthetic glycobiology: designing and engineering glycomolecules outside of living cells. *Front. Chem.*
21. Kightlinger, W., Warfel, K. F., Delisa, M. P. & Jewett, M. C. Synthetic Glycobiology: Parts, Systems, and Applications. *ACS Synth. Biol.* **9**, 1534–1562 (2020).
22. Yu, H. & Chen, X. One-pot multienzyme (OPME) systems for chemoenzymatic synthesis of carbohydrates. *Org. Biomol. Chem.* **14**, 2809–2818 (2016).
23. Kightlinger, W. *et al.* A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nat. Commun.* (2019) doi:10.1038/s41467-019-12024-9.
24. Jaroentomeechai, T. *et al.* Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat. Commun.* (2018) doi:10.1038/s41467-018-05110-x.
25. Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442**, 368–373 (2006).
26. Duncombe, T. A., Tentori, A. M. & Herr, A. E. Microfluidics: Reframing biological enquiry. *Nature Reviews Molecular Cell Biology* (2015) doi:10.1038/nrm4041.
27. Georgi, V. *et al.* On-chip automation of cell-free protein synthesis: new opportunities due to a novel reaction mode. *Lab Chip* **16**, 269–281 (2016).
28. Timm, A. C., Shankles, P. G., Foster, C. M., Doktycz, M. J. & Retterer, S. T. Characterization of extended channel bioreactors for continuous-flow protein production. *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* (2015) doi:10.1116/1.4932155.
29. Jackson, K., Kanamori, T., Ueda, T. & Hugh Fan, Z. Protein synthesis yield increased 72 times in the cell-free PURE system. *Integr. Biol.* **6**, 781 (2014).
30. Timm, A. C., Shankles, P. G., Foster, C. M., Doktycz, M. J. & Retterer, S. T. Toward Microfluidic Reactors for Cell-Free Protein Synthesis at the Point-of-Care. *Small* **12**, 810–817 (2016).
31. Xiao, X. *et al.* Integration of cell-free protein synthesis and purification in one microfluidic chip for on-demand production of recombinant protein. *Biomicrofluidics* **12**, (2018).
32. Murphy, T. W., Sheng, J., Naler, L. B., Feng, X. & Lu, C. On-chip manufacturing of synthetic proteins for point-of-care therapeutics. *Microsystems Nanoeng.* **5**, 13 (2019).
33. Martin, J. G. *et al.* Toward an artificial Golgi: Redesigning the biological activities of heparan sulfate on a digital microfluidic chip. *J. Am. Chem. Soc.* **131**, 11041–11048 (2009).
34. Fisher, A. C. *et al.* Production of secretory and extracellular N-linked glycoproteins in Escherichia coli. *Appl. Environ. Microbiol.* (2011)

doi:10.1128/AEM.01901-10.
35. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* (2005) doi:10.1073/pnas.0500044102.
36. Wacker, M. *et al.* N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science (80-.).* (2002) doi:10.1126/science.298.5599.1790.
37. Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* (2006) doi:10.1038/sj.emboj.7601087.
38. Ikonomova, S. P., He, Z. & Karlsson, A. J. A simple and robust approach to immobilization of antibody fragments. *J. Immunol. Methods* (2016) doi:10.1016/j.jim.2016.04.012.
39. Kowarik, M. *et al.* N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science (80-.).* **314**, 1148–1150 (2006).
40. Ollis, A. A., Zhang, S., Fisher, A. C. & Delisa, M. P. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat. Chem. Biol.* (2014) doi:10.1038/nchembio.1609.
41. Kwon, Y. C. & Jewett, M. C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci. Rep.* (2015) doi:10.1038/srep08663.
42. Gerber, S. *et al.* Mechanism of bacterial oligosaccharyltransferase: In vitro quantification of sequon binding and catalysis. *J. Biol. Chem.* **288**, 8849–8861 (2013).
43. Perregaux, E. *et al.* Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. *Sci. Rep.* **5**, 1–13 (2015).
44. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* (2011) doi:10.1038/nature10151.
45. Schoborg, J. A. *et al.* A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnol. Bioeng.* (2018) doi:10.1002/bit.26502.
46. Teramura, Y., Kuroyama, K. & Takai, M. Influence of molecular weight of PEG chain on interaction between streptavidin and biotin-PEG-conjugated phospholipids studied with QCM-D. *Acta Biomater.* **30**, 135–143 (2016).
47. Baranowska, M. *et al.* Protein attachment to silane-functionalized porous silicon: A comparison of electrostatic and covalent attachment. *J. Colloid Interface Sci.* **452**, 180–189 (2015).
48. Zilberzwige-Tal, S. *et al.* Programmable On-Chip Artificial Cell Producing Post-Translationally Modified Ubiquitinated Protein. *Small* **1901780**, 1901780 (2019).
49. Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat. Struct. Mol. Biol.* (2017) doi:10.1038/nsmb.3491.
50. Ihssen, J. *et al.* Production of glycoprotein vaccines in Escherichia coli. *Microb.*

Cell Fact. (2010) doi:10.1186/1475-2859-9-61.
51. Hug, I. *et al.* Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein. *J. Biol. Chem.* (2011) doi:10.1074/jbc.M111.287755.
52. Shang, W. *et al.* Production of human blood group B antigen epitope conjugated protein in Escherichia coli and utilization of the adsorption blood group B antibody. *Microb. Cell Fact.* (2016) doi:10.1186/s12934-016-0538-z.
53. Valderrama-Rincon, J. D. *et al.* An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat. Chem. Biol.* **8**, 434–436 (2012).
54. Zhu, J. *et al.* An Engineered Pathway for Production of Terminally Sialylated N-glycoproteins in the Periplasm of Escherichia coli. *Front. Bioeng. Biotechnol.* **8**, 1–9 (2020).
55. Natarajan, A. *et al.* Engineering orthogonal human O-linked glycoprotein biosynthesis in bacteria. *Nat. Chem. Biol.* **16**, 1062–1070 (2020).
56. Stark, J. C. *et al.* On-demand biomanufacturing of protective conjugate vaccines. *Sci. Adv.* **7**, (2021).
57. Hamilton, B. S. *et al.* A library of chemically defined human N-glycans synthesized from microbial oligosaccharide precursors. *Sci. Rep.* (2017) doi:10.1038/s41598-017-15891-8.
58. Liu, F. *et al.* Rationally designed short polyisoprenol-linked PglB substrates for engineered polypeptide and protein N-glycosylation. *J. Am. Chem. Soc.* (2014) doi:10.1021/ja409409h.
59. Schwarz, F. *et al.* A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat. Chem. Biol.* **6**, 264–266 (2010).
60. Malet-Sanz, L. & Susanne, F. Continuous flow synthesis. a pharma perspective. *Journal of Medicinal Chemistry* (2012) doi:10.1021/jm2006029.
61. Poechlauer, P. *et al.* Pharmaceutical Roundtable Study Demonstrates the Value of Continuous Manufacturing in the Design of Greener Processes. *Org. Process Res. Dev.* (2013) doi:10.1021/op400245s.
62. Fisher, A. C. *et al.* The Current Scientific and Regulatory Landscape in Advancing Integrated Continuous Biopharmaceutical Manufacturing. *Trends in Biotechnology* (2019) doi:10.1016/j.tibtech.2018.08.008.
63. Porta, R., Benaglia, M. & Puglisi, A. Flow Chemistry: Recent Developments in the Synthesis of Pharmaceutical Products. *Organic Process Research and Development* (2016) doi:10.1021/acs.oprd.5b00325.
64. Wohlgemuth, R., Plazl, I., Žnidaršič-Plazl, P., Gernaey, K. V. & Woodley, J. M. Microscale technology and biocatalytic processes: Opportunities and challenges for synthesis. *Trends in Biotechnology* (2015) doi:10.1016/j.tibtech.2015.02.010.
65. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4305.

3 KINETOPLASTID SINGLE SUBUNIT OSTS ENABLE TRANSFER OF EUKARYOTIC *N*-GLYCANS IN *E. COLI*

Alicia K. Aquino[1], Jiazhen Chen[1], Alexandra Pang[1], and Matthew P. DeLisa[1,2]

[1]Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, United States
[2]Cornell Institute of Biotechnology, Cornell University, 130 Biotechnology Building, Ithaca, NY 14853 USA

3.1 Introduction

Glycoproteins are one of the fastest growing classes of biotherapeutics with monoclonal antibodies dominating the market[1]. Protein glycosylation is a post-translational modification where a complex sugar, called a glycan, is attached to amino acid side chains of a protein. The most common type of glycosylation is *N*-linked where the site for glycan attachment is the asparagine residue. In eukaryotic cells, these glycans arise from a series of enzymatic reactions localized in the endoplasmic reticulum and the Golgi. This process gives rise to a heterogeneous mixture of glycans with complex structures that modify and expand the functionalities of glycoproteins. It is now established that glycosylation is one of the most important criteria for developing safe and effective therapeutics[2]. For example, incomplete capping with sialic acid at the non-reducing end of the glycan can lead to faster clearance of the drug while non-human glycan motifs can produce immunogenetic effects[3]. For these reasons, glycoproteins therapeutics are manufactured in mammalian cells where glycosylation patterns closely resemble those in humans. However, a major production challenge in these systems is the variability of glycosylation distributions that arise from small changes to cell culture conditions. Furthermore, glycosylation is tied to cell health, so efforts to modify native

glycosylation machinery can have adverse effects on the cell. This makes it difficult to obtain glycoproteins with the desired therapeutic properties.

As an alternative to eukaryotic cells, *Escherichia coli* have been proposed as a host for glycoprotein production. *E. coli* have a long history of use for recombinant and biotherapeutic protein production because of their low costs, fast growth, and availability of simple genetic modification techniques[4]. Although *E. coli* lack native protein glycosylation machinery, this clean slate provides an opportunity to engineer in only the desired glycosylation pathways in a system where glycosylation is decoupled from cell viability. To this end, a major advancement in the field has been the functional transfer of the *N*-glycosylation system from *Campylobacter jejuni* into *E. coli*[5]. In this system, the *C. jejuni* bacterial heptasaccharide is built stepwise on an undecaprenol pyrophosphate (UndPP) lipid carrier at the cytosolic side of the inner membrane in *E. coli*. The glycan is then flipped into the periplasm where protein glycosylation occurs. The periplasm is an oxidative environment that facilitates formation of disulfide bonds and can be reached by fusion of an acceptor protein to signal peptides that direct the protein through the Sec-translocon. Together, these features make *E. coli* a promising host for the production of recombinant glycoproteins.

The key enzyme in protein *N*-glycosylation is the oligosaccharyltransferase (OST) which transfers a preassembled glycan *en bloc* from a lipid carrier at the minimal consensus sequon, N-X-S/T, where X is any amino acid except proline. In *C. jejuni*, the OST, *Cj*PglB, has been shown to have a relaxed glycan substrate specificity where the only known requirement is a HexNAc monosaccharide at the reducing end[6]. This has

enabled the production of glycoproteins bearing a variety of therapeutically relevant glycan structures, including bacterial *O*-polysaccharide antigens[7,8], human blood group B antigens[9], Lewis x antigens[10], terminally sialylated glycans[11,12], and the eukaryotic core *N*-glycan, mannotriose-di-(*N*-acetyl-glucosamine) ($Man_3GlcNAc_2$)[13,14]. Although the glycan specificity is relaxed, *Cj*PglB has a more stringent protein specificity compared to eukaryotic OSTs. For efficient transfer, *Cj*PglB has an additional amino acid requirement at the -2 position upstream of the asparagine residue for an expanded sequon, D/E-X-N-X-S/T[15]. This has limited the ability of *Cj*PglB to glycosylate therapeutic proteins at their natural sites. To address this bottleneck, PglBs from alternative sources have been explored. Notably, *Desulfovibrio, Helicobacter*, and deep sea vent bacterial species have been identified that have more relaxed specificities[16–18], however, the ability of these enzymes to transfer therapeutically relevant glycan structures has not been reported.

Although efforts to find alternative OSTs for glycoprotein production in *E. coli* have exclusively focused on OSTs from prokaryotic sources, eukaryotic OSTs, which generally have more relaxed sequon specificities, could potentially be used to expand the glycosylation repertoire in *E. coli*. In higher eukaryotes, the OST is a transmembrane protein composed of eight subunits[19]. The catalytic domain, called the STT3, is homologous to PglB and is composed of thirteen transmembrane helices followed by a C-terminal soluble domain that contains the catalytic pocket. The other subunits of the OST are involved in substrate recognition, association with oxidoreductases and the translocon, and stabilizing the complex. The complexity of these multisubunit OST

structures would potentially make it challenging to express them recombinantly. However, in some lower eukaryotes the OST is composed of just a single subunit. In kinetoplastids, which are unicellular, parasitic protists, the OST exists as a single subunit enzyme that is homologous to the STT3 domain of higher eukaryotes and to PglB in bacteria. While OSTs from higher eukaryotes diversify their glycosylation ability through incorporation of multiple subunits, protozoan single subunit OSTs do this through duplication of the STT3 gene where each STT3 paralog has different glycan and protein substrate specificities. These substrate specificities are well-studied in *Trypanosoma brucei*, where there are three STT3 paralogs, *Tb*STT3A, *Tb*STT3B, and *Tb*STT3C. *Tb*STT3A has been shown to have a preference for acidic residues around the glycosylation site and selectively transfers high mannose type glycans lacking the c-branch (seven or fewer mannoses)[20–24]. On the other hand, *Tb*STT3B has been shown to have no specific preference[23] or to prefer sites with hydrophobic and positively charged residues in the local environment[22] and preferentially transfers high mannose glycans with the c-branch (above seven mannoses)[20,21,23]. *Tb*STT3C has been found to also prefer glycosylation sites at acidic regions[20,23], but exhibited the broadest glycan substrate specificity with the ability to transfer all high mannose type glycan structures[23]. Substrate specificities have also been explored for the four paralog STT3s from *Leishmania major*, though they are not as well characterized. Of these, *Lm*STT3D was found to have the broadest glycan and protein substrates specificities and could complement the deletions in all essential OST subunits in yeast[25,26]. Because of these properties, *Lm*STT3D has been used to improve glycosylation site occupancy of

monoclonal antibodies in various yeast[27,28] and plant[29] recombinant glycoprotein production systems.

While kinetoplastid single subunit STT3s have been used to substitute or augment OST activity in higher eukaryotes, they have yet to be explored in bacterial systems. A major difference between eukaryotic and prokaryotic protein glycosylation environments is the composition of the lipid bilayer and thus, the lipid carrier for the oligosaccharide. In eukaryotic cells, the lipid carrier is dolichol pyrophosphate (DolPP) while in prokaryotes the lipid carrier is UndPP. One of the differences between these lipid structures is the length of the chain. In eukaryotes, mammalian DolPP is composed of 18-21 isoprene units, plant and fungi have 15-16 isoprene units, and trypanosomatids have 10-12 isoprene units[30]. In prokaryotes, UndPP is composed of 11 isoprene units. Additionally, DolPP has one saturated unit at the terminus containing the hydroxyl group while UndPP has a double bond. Whether these differences influence the activity of STT3s was explored *in vitro* in a study that measured the turnover rate of *Tb*STT3A when transferring a disaccharide from a variety of synthetic lipid carrier. This study found that while a decrease in the number of isoprene units of the lipid tail decreased the affinity of *Tb*STT3A for the substrate, neither the double bond chemistry nor length of the lipid had a significant impact on the turnover rate[31]. This suggests that the difference in lipid structure should not inhibit the activity of the protozoan STT3s in *E. coli*.

In this study, we tested the activity of a panel of kinetoplastid single subunit STT3s from *Leishmania* and *Trypanosoma* in the context of *E. coli*. As the eukaryotic

glycan substrate, we used a previously developed glycosylation pathway where $Man_3GlcNAc_2$ is synthesized at the cytosolic side of the membrane by a combination of endogenous *E. coli* and recombinant *Saccharomyces cerevisiae* glycosylation enzymes[13,14]. This glycan is then flipped into the periplasm where it can interact with the active site of the STT3s. The advantage of transferring $Man_3GlcNAc_2$ is that this core structure serves as the base of all eukaryotic *N*-linked glycan structures. After the initial transfer onto a protein, this structure can then be elaborated to synthesize complex glycans found on glycoprotein therapeutics. Using this system, we demonstrate efficient glycosylation of a model protein by the kinetoplastid STT3s. We then investigate the sequon specificities of the most efficient OST candidates. This work represents the first time that eukaryotic OSTs have been functionally incorporated into *E. coli* and provides new OST candidates for production of glycoprotein therapeutics.

3.2 Results

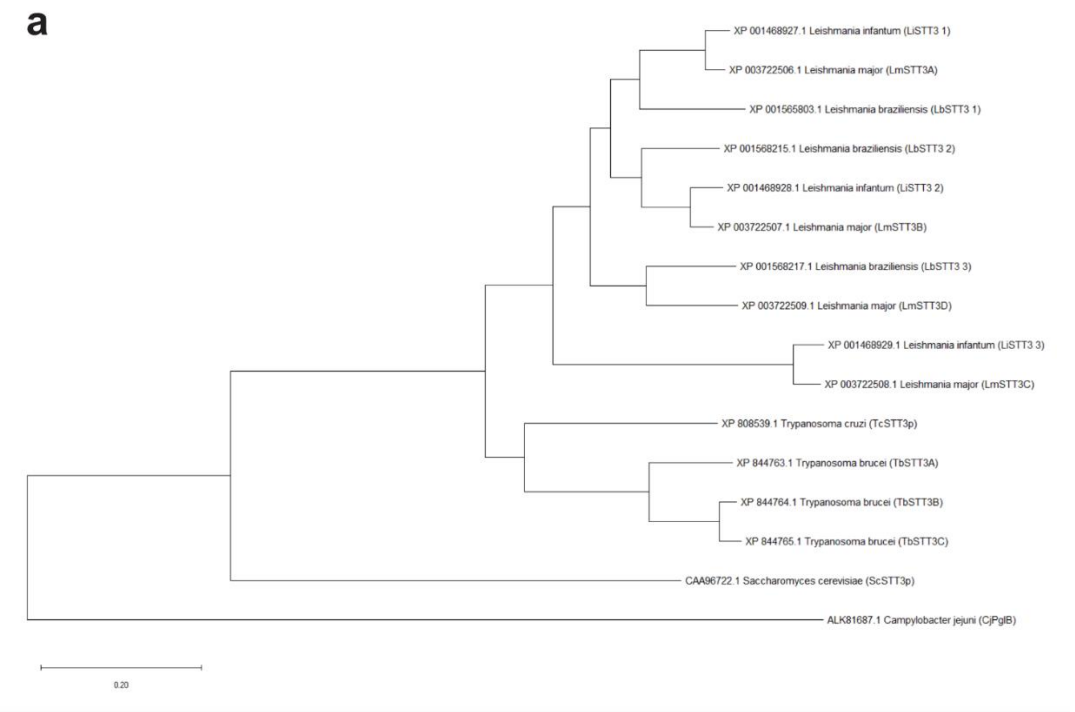
3.2.1 Analysis of the eukaryotic single subunit STT3 panel

In this study, fourteen single subunit kinetoplastid OSTs were examined from the species *Leishmania braziliensis*, *Leishmania infantum*, *Leishmania major*, *Trypanosoma brucei*, and *Trypanosoma cruzi*. These STT3s have previously been studied in the context of yeast for their ability to replace the function of the *Saccharomyces cerevisiae* STT3 subunit[20–23,25,26,28,32], and were investigated here for their ability to function in *E. coli*. In *L. braziliensis*, *L. infantum*, and *T. brucei* there are three STT3 genes, in *L. major* there are four, and in *T. cruzi* there is one. To gain a better view of how these STT3 paralogs compare across species, we generated a phylogenetic

tree, shown in **Fig.1a**. Although the kinetoplastid STT3 paralogs are more closely related to *S. cerevisiae* STT3, they only share 28-33% sequence homology. This range is similar to the sequence identities between the kinetoplastid STT3s and *C. jejuni* PglB which share 21-36% homology. Among the *Leishmania* STT3 paralogs, the STT3_1s clusters with STT3A, the STT3_2s cluster with STT3B, and STT3_3s, STT3C and STT3D cluster together in the phylogenetic tree and have the highest sequence homology with each other. This suggests that these paralogs may share common functions, and thus similar substrate specificities, across species. The pairs that exhibited the highest sequence identities were *Li*STT3_1 and *Lm*STT3A (95%), *Li*STT3_2 and *Lm*STT3B (93%), *Li*STT3_3 and *Lm*STT3C (93%), and *Tb*STT3B and *Tb*STT3C (95%). *L. braziliensis* STT3s shared the closest identities with *L. infantum* STT3s with highest homology between *Lb*STT3_1 and *Li*STT3_1 (79%), *Lb*STT3_2 and *Li*STT3_2 (83%), and *Lb*STT3_3 and *Lm*STT3D (80%). *Tc*STT3p shared the lowest sequence identity with all kinetoplastid STT3s with the highest range of 61-62% homology with the *Tb*STT3s.

Another feature of the STT3s that we compared were glycosylation sites on the OST. In the *S. cerevisiae* STT3 subunit there is a *N*-glycosylation motif, DNNTWNNT, downstream of the conserved OST active site, WWD. Within this motif, the N539 residue is glycosylated with $Man_8GlcNAc_2$ and is essential for the activity of the OST[33]. Similarly, *C. jejuni* PglB has also been found to be a glycoprotein and is glycosylated at the N534 site at the DYNQS motif[34]. However, the effect of this glycan structure on *Cj*PglB activity is unclear since *Cj*PglB is active in *E. coli* that lack the biosynthesis

pathway for the *C. jejuni* glycan and can transfer a variety of non-native glycan structures[7–14]. In *Trypanosoma cruzi*, while the eukaryotic DNNTWNNT motif where glycosylation is essential is absent, the N590 residue, located downstream of the WWD motif, has been shown to be glycosylated[35]. In a sequence alignment of the kinetoplastid OSTs with *Sc*STT3 and *Cj*PglB, this same NRT motif is found in all kinetoplastid STT3 paralogs (**Fig. 1b**). This suggests that they are also glycosylated, though the function and essentiality of the glycan structure is unknown.



b

```
ALK81687.1(CjPglB)        WWDYGYPVRYYSDVKTLVDGGKHLGKDNFFPSFALSKD-EQAAANMARLSVEYT----EKSFYAPQNDILKTDILQA--------MMKD-------YNQSNVDLFLA---------SLSKP-- 548
CAA96722.1(ScSTT3p)       WWDYGYQIGGMADRTTLVDNNTWNNTHIAIVGKAMASPEEKSYEILKEHDVDYVLVIFGGLIGFGGDDINKFLWMIRISEGIWPEEIKERDF-YTAEGEYRV---------------------- 616
XP_808539.1(TcSTT3p)      WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKLLTSPVAKAHL-LIRHLADYVLIWTGSR----AEDLMKSPHMARIGNSVYRDICPEDDPLCSNFGFEDYD---------------------- 675
XP_844763.1(TbSTT3A)      WWDYGYQITGIGNRTTLADGNTWSHKHIATIGKMLTSPVKESHA-LIRHLADYVLIWAGED----RGDLLKSPHMARIGNSVYRDMCSEDDPRCRQFGFEGGD---------------------- 704
XP_844764.1(TbSTT3B)      WWDYGYQITGIGNRTTLADGNTWNHEHIATIGKMLTSPVKESHA-LIRHLADYVLIWAGYD----GSDLLKSPHMARIGNSVYRDICSEDDPLCTQFGFYSGD---------------------- 701
XP_844765.1(TbSTT3C)      WWDYGYQITGIGNRTTLADGNTWNHEHIATIGKMLTSPVKESHA-LIRHLADYVLIWAGED----RGDLRKSRHMARIGNSVYRDMCSEDDPLCTQFGFYSGD---------------------- 701
XP_001468929.1(LiSTT3_3)  WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLIFAGDTY---FSDLNRSPHMARIGNSVYRDICPH-DPLCSRFVLQKRPKAAAAKRSRHVSVDELEEEDN 675
XP_003722508.1(LmSTT3C)   WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLISAGDTY---FSDLNRSPMMARIGNSVYHDICPD-DPLCSQFVLQKRPKAAAAKRSRHVSVDALEEDDT 675
XP_001568217.1(LbSTT3_3)  WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLIWAGQG-----GDLMKSPHMARIGNSVYHDICPN-DPLCQHFGFYKND---------------------- 731
XP_003722509.1(LmSTT3D)   WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVVEAHS-LVRHMADYVLIWAGQS-----GDLMKSPHMARIGNSVYHDICPD-DPLCQQFGFHRND---------------------- 732
XP_001568215.1(LbSTT3_2)  WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLIWAGQG-----GDLMKSPHMARIGNSVYHDICPN-DPLCQHFGFYE-D---------------------- 653
XP_001468928.1(LiSTT3_2)  WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLIWAGQS-----GDLMKSPHMARIGNSVYHDICPH-DPLCQQFGFYRND---------------------- 662
XP_003722507.1(LmSTT3B)   WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLIWAGQS-----GDLMKSPHMARIGNSVYHDICPD-DPLCQQFGFHRND---------------------- 662
XP_001565803.1(LbSTT3_1)  WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLIWAGQG-----GDLMKSPHMARIGNSVYHDICPN-DPLCQHFGFYE-D---------------------- 740
XP_001468927.1(LiSTT3_1)  WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLIWAGQS-----GDLMKSPHMARIGNSVYHDICPH-DPLCQQFGFYRND---------------------- 741
XP_003722506.1(LmSTT3A)   WWDYGYQITGIGNRTSLADGNTWNHEHIATIGKMLTSPVAEAHS-LVRHMADYVLIWAGQS-----GDLMKSPHMARIGNSVYHDICPD-DPLCQQFGFHRND---------------------- 741
```

Figure 1. Sequence analysis of single subunit oligosaccharyltransferases (OSTs). a) Phylogenetic tree of the kinetoplastid STT3s explored in this study compared with the well-studied OSTs, *Campylobacter jejuni* PglB and putative *Saccharomyces cerevisiae* STT3p. The tree was constructed using MEGA11 software and the evolutionary history was inferred using the Neighbor-Joining method. **b)** Sequence alignment by Clustal Omega of the OSTs analyzed in **a)** showing potential glycosylation sites on the enzyme. Known glycosylation sites are highlighted in gray. Possible glycosylation sites are bolded. The WWD motif common to all OST catalytic domains is underlined.

3.2.2 Identification of eukaryotic STT3s that transfer $Man_3GlcNAc_2$ from UndPP to a DQNAT consensus sequon

To screen the activity of the eukaryotic STT3s, we tested glycosylation of a model protein, MBP-glucagon (MBP-GCG) bearing the optimal *C. jejuni* glycosylation sequon at the C-terminal with the core eukaryotic glycan structure, $Man_3GlcNAc_2$. The method used here has previously been optimized for complete glycosylation of the acceptor protein by *Cj*PglB[14]. After expression of the OST candidates with the glycosylation machinery, we observed in the ConA blot glycosylation of the acceptor protein by the positive control, *Cj*PglB, as well as by *Lb*STT3_2, *Lb*STT3_3, and *Li*STT3_1 (**Fig. 2a**). Additionally, weak glycosylation bands could be observed in samples glycosylated by *Lm*STT3A, *Lm*STT3D, *Tb*STT3A, and *Tb*STT3C while no glycosylation is observed in samples glycosylated by *Tb*STT3p and the negative control where no OST was expressed. In the anti-His blot, we observed that productivity of the acceptor protein was reduced and could not be detected at the correct molecular weight when co-expressed with *Lb*STT3_1, *Lm*STT3B, *Lm*STT3C, *Li*STT3_2, *Li*STT3_3, and *Tb*STT3B, so no conclusions could be drawn about their activity under these conditions. The results from the PNGase F treatment in **Fig. 2b** further confirms that $Man_3GlcNAc_2$ is indeed attached to the MBP-GCG^{DQNAT} in samples glycosylated by *Cj*PglB, *Lb*STT3_2, *Lb*STT3_3, and *Li*STT3_1 and the observed band shift downwards after cleavage of the glycan suggests that there is high site occupancy of the glycan on the acceptor protein.

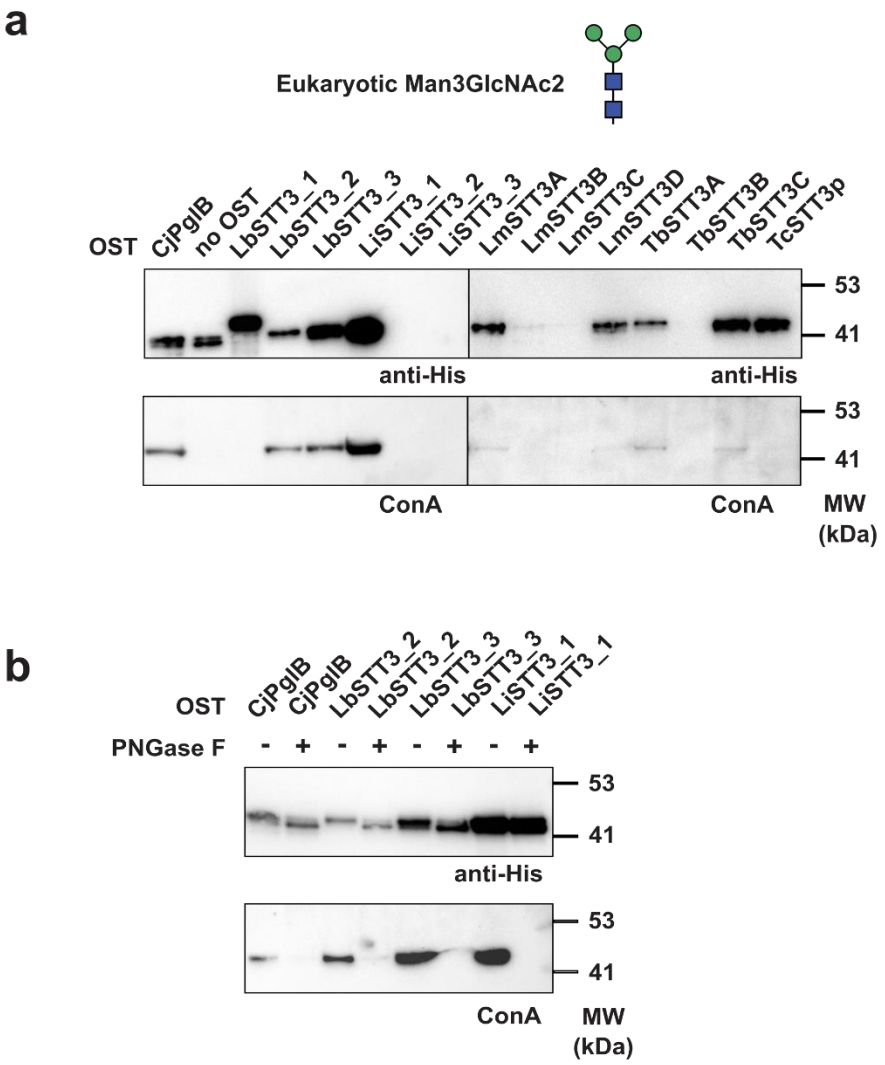


Figure 2. Screen of STT3 activity. a) Western blot analysis MBP-glucagon^{DQNAT} samples purified from *E. coli* that were glycosylated with $Man_3GlcNAc_2$ by the panel of single subunit eukaryotic oligosaccharyltransferases (OSTs). Expression of the glycosylation machinery with *Campylobacter jejuni* PglB serves as the positive control and no OST expression serves as the negative control. The anti-His blot probes for the acceptor protein, MBP-glucagon^{DQNAT}, that contains a 6x histidine tag at the C-terminal. The ConA blot probes for mannose in the mannotriose-di-(*N*-acetyl-glucosamine) ($Man_3GlcNAc_2$) glycan structure. **b)** Western blot analysis of $Man_3GlcNAc_2$ glycosylated MBP-glucagon^{DQNAT} samples from **a)** that were glycosylated by *Cj*PglB, *Leishmania braziliensis* STT3_2 (*Lb*STT3_2), *Leishmania braziliensis* STT3_3 (*Lb*STT3_3), and *Leishmania infantum* STT3_1 (*Li*STT3_1) were treated with PNGase F. PNGase F cleaves *N*-linked glycans from the protein at the linkage between the asparagine residue and the reducing end *N*-acetylglucosamine.

3.2.3 Top hits from the eukaryotic single subunit OST activity screen exhibit preference for negatively charged amino acids in the -2 position

To examine the substrate specificities of the single subunit OSTs, we generated a panel of MBP-GCG^{XQNAT} variants where the -2 sequon position was varied to all twenty amino acids. Each library member was then individually screened with the top single subunit OSTs from **Fig. 2a** for their ability to transfer the $Man_3GlcNAc_2$ glycan structure. The results show that each of the top STT3s had selectively towards an acidic residue in the -2 position (**Fig. 3a-c**) and did not appear to glycosylate the acceptor protein when the -2 position amino acid was anything other than D or E. This preference is identical to what is observed for PglBs from *C. jejuni* and *C. lari*[17]. For *Cl*PglB, it has been found that high specificity for a negatively charged amino acid in the -2 position results from residue R331 which forms a salt bridge with the D or E, tightening binding of PglB to the peptide[36]. Therefore, to gain a better understanding of why *Lb*STT3_2, *Lb*STT3_3, and *Li*STT3_1 exhibited a strong preference for negatively charged amino acids at the -2 position, we compared the structure of *Cl*PglB[37] with predicted 3D structures of the STT3s. The results show that while the interaction between R331 of *Cl*PglB and the -2 position of the peptide sequon is clear, there is not a similar region of positive charge in the active site of the STT3s (**Fig. 4**). On the contrary, this location contains mostly negative and neutral charge. Another difference is that the active site of the STT3s appears to be more open than that of *Cl*PglB. This may suggest that there is less interaction between the sequon and OST charged residues and the apparent selectivity of the STT3s for negatively charged amino acids at the -2 sequon position may not be explained only by this region.

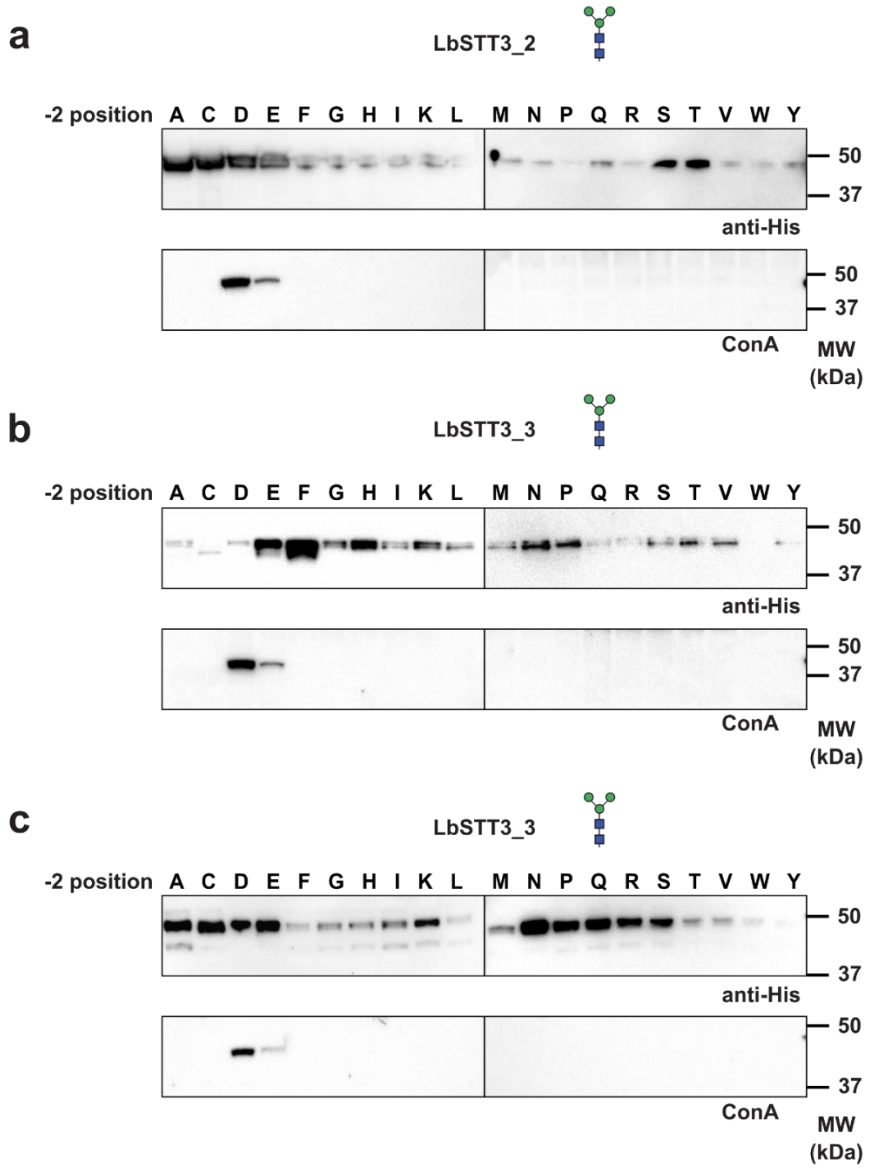


Figure 3. Sequon specificities of top STT3 candidates. A 20 member library of MBP-glucagon^{XQNAT} variants where the -2 position of the glycosylation sequon was changed to the every amino acid was used to screen the acceptor protein substrate specificity of **a)** *Leishmania braziliensis* STT3_2 (*Lb*STT3_2), **b)** *Leishmania braziliensis* STT3_3 (*Lb*STT3_3), and **c)** *Leishmania infantum* STT3_1 (*Li*STT3_1) using the mannotriose-di-(*N*-acetyl-glucosamine) ($Man_3GlcNAc_2$) glycan structure. The anti-His blot probes for MBP-glucagon^{XQNAT} which contains a 6x histidine tag at the C-terminal. The ConA blot probes for mannose in $Man_3GlcNAc_2$. Figure 3c data were generated by JC.

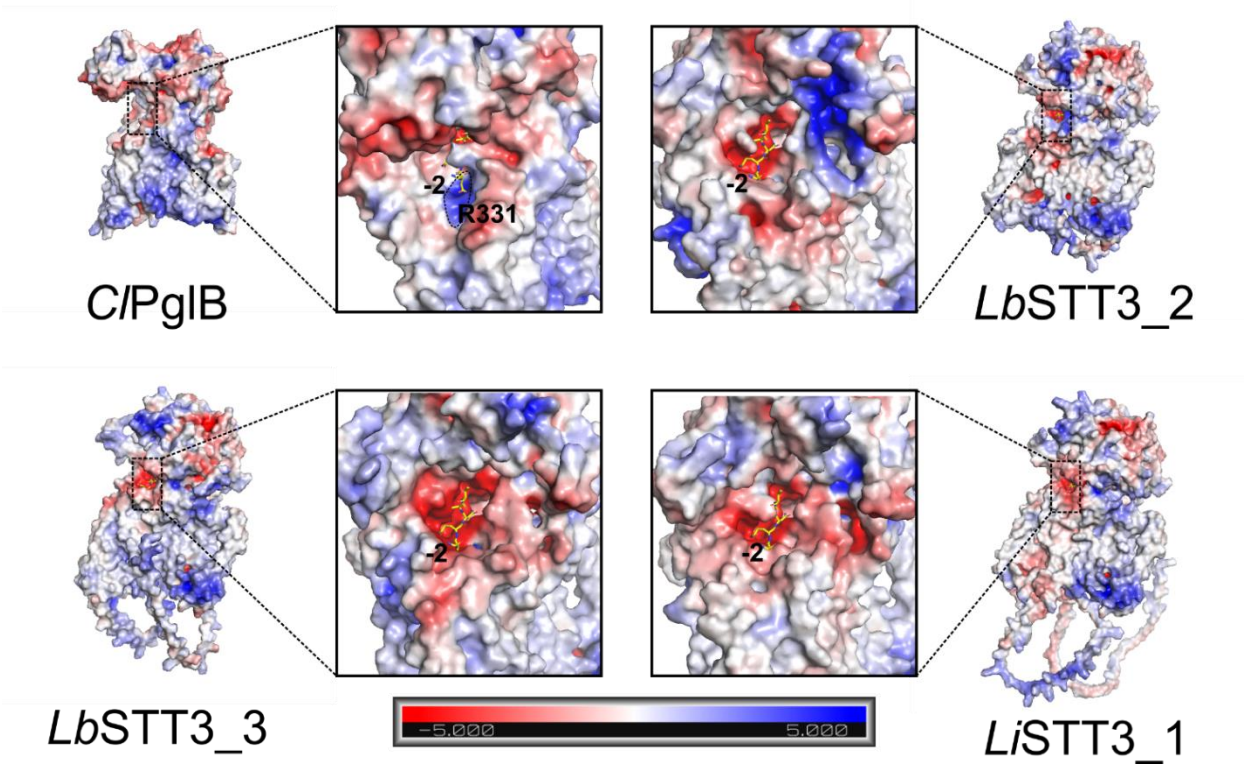


Figure 4. Comparison of OST active sites. The 3D structures of *Leishmania braziliensis* STT3_2 (*Lb*STT3_2), *Leishmania braziliensis* STT3_3 (*Lb*STT3_3), and *Leishmania infantum* STT3_1 (*Li*STT3_1) were modeled using ColabFold and aligned to the molecular structure of *Campylobacter lari* PglB from PDB 5ogl. Based on these structures, the electrostatics surface visualizations were created using the APBS Electrostatics Plugin in PyMOL The acceptor peptide, GDQNAT, from PDB 5ogl was modeled in the STT3 structures based on the alignment of the STT3s with *Cl*PglB. The R331 residue of *Cl*PglB and the location of the -2 position residue, D, of the sequon are labeled. The legend box was generated by the PyMOL APBS Electrostatics Plugin indicates the electrostatic charge from -5 to +5 where red regions of the protein indicate negatively charged residues and blue regions indicated positively charged residues. Predicted 3D structures of STT3s were generated by Dr. Sai Pooja.

3.3 Discussion

The OST is the key enzyme protein *N*-glycosylation and is a bottleneck in the development of *E. coli* as a platform for therapeutic, eukaryotic-like glycoprotein production. To date, only bacterial OSTs have been explored in *E. coli* for use in building glycosylation pathways and only *Cj*PglB has been shown to have the ability to transfer eukaryotic-like glycan structures. Here, we identified single subunit eukaryotic

OSTs from kinetoplastids that could be successfully expressed in *E. coli* and showed activity in transferring the core eukaryotic $Man_3GlcNAc_2$ glycan structure to a model acceptor protein. The most efficient OSTs identified in this study were *Lb*STT3_2, *Lb*STT3_3, and *Li*STT3_1. Interestingly, although they are eukaryotic enzymes, these STT3s showed high selectivity for glycosylation at sites where D or E was present in the -2 sequon position. This selectivity could not be explained by 3D structural models of the electrostatic properties at the active site of the STT3s. At the locations where the -2 position of the acceptor protein would be, the surrounding charges in the STT3s were negative or neutral. However, the exact positioning of the peptide in the STT3 active site is not a given and other factors may be involved in the acceptor peptide affinity and binding.

While the STT3s identified here that showed the highest glycosylation efficiency also had an expanded glycosylation sequon, this is likely a result of the method we used to screen the STT3s. Had we used a different glycosylation sequon or glycan in the initial screen, we may have identified other STT3s with higher glycosylation efficiencies. Additionally, while we chose to investigate only the STT3s that produced the highest ConA signal, testing the STT3s that produced a weak signal could have potentially yielded STT3s with more relaxed sequon specificities overall. Although the exact sequon specificities of *Lb*STT3_2, *Lb*STT3_3, and *Li*STT3_1 that we examined here are not well defined in the literature, we did observe that there were some similarities between our results and what has been observed by others previously. For example, *Tb*STT3A and *Tb*STT3C have been shown to preferentially glycosylate

regions with acidic amino acids in the local environment and are capable of transferring truncated high mannose structures[38], including $Man_3GlcNAc_2$ for *Tb*STT3A[24]. In our results using $Man_3GlcNAc_2$ and a DQNAT acceptor sequon, we see that samples glycosylated by *Tb*STT3A and *Tb*STT3C have some activity, albeit low levels of glycosylation were observed.

Overall, the results presented here represent the first time that eukaryotic OSTs have been shown to be active in *E coli*. Although the STT3s that we examined here do not offer any additional benefits over *Cj*PglB, kinetoplastid STT3 paralogs each have different glycan and protein substrate specificities. Therefore, depending on the target glycosylation site, different STT3 candidates may be better suited for each situation. Next steps should focus on screening the OST panel against a specific therapeutic glycoprotein target, such as the QYNST site of the fragment crystallizable (Fc) region in monoclonal antibodies. Efficient glycosylation of a therapeutic protein with $Man_3GlcNAc_2$ at the natural site would be the first step in biosynthesis of eukaryotic glycoproteins in *E. coli*.

3.4 Materials and Methods

Bacterial strains and plasmids

E. coli strain DH5α (lab stock) was used for plasmid cloning and purification. *E. coli* strain MC4100 gmd::kan ΔwaaL, which contains knockouts that improve enrichment of the $Man_3GlcNAc_2$ glycan, was developed previously[13] and was used here for all glycoprotein production experiments. Plasmids pConYCGmCB containing the $Man_3GlcNAc_2$ and GDP-mannose biosynthesis enzymes and pTrc-spDsbA-MBP-

GCG^{DQNAT} containing maltose binding protein (MBP) fused to human glucagon (GCG) with a C-terminal optimized bacterial glycosylation tag and 6xHis tag were cloned previously[14].

Plasmid construction

All protozoan single subunit STT3 genes were PCR amplified from plasmids or gblocks and cloned into pMLBAD[39] between EcoRI and NcoI restriction sites with GS-FLAG-GS-10xHis at the C-terminal using Gibson assembly. Template plasmids containing the genes for LbSTT3_1, LbSTT3_2, LiSTT3_1, LiSTT3_2, and LiSTT3_3 were amplified from pAX327, pAX316, pAX318, pAX319, pAX320, respectively, kindly provided by the Frey lab[28]. Template plasmids containing codon optimized genes for LbSTT3_3, LmSTT3B, LmSTT3C, LmSTT3D, TbSTT3A, TbSTT3B, TbSTT3C in the pJL1 vector were kindly provided by the Jewett lab (work not published). The template plasmid for TcSTT3p in the pMLBAD vector was kindly provided by the Parodi lab[40]. The template sequence for LmSTT3A was codon optimized for *E. coli* strain K12 using the IDT Codon Optimization Tool and purchased as a gBlock® gene fragment (Integrated DNA Technologies). To generate pMLBAD-CjPglB, the gene for CjPglB was amplified from pMAF10[7] and cloned back into pMLBAD between EcoRI and NcoI with GS-FLAG-GS-10xHis at the C-terminal using Gibson assembly. To generate the library of pTrc-spDsbA-MBP-GCG^{XQNAT} glycosylation tag variants, a combination of site-saturation mutagenesis[41] and site-directed mutagenesis was used.

Protein expression, glycosylation, and enrichment

MC4100 gmd::kan ΔwaaL cells carrying pConYCG_ManCB, pTrc-spDsbA-MBP-GCG^{XQNAT}, and one of the pMLBAD-OST plasmids were transformed, one plasmid at a time, into chemically competent cells and plated on Luria-Bertani (LB) agar supplemented with 20 µg/mL chloramphenicol, 100 µg/mL trimethoprim, 100 µg/mL ampicillin, 50 µg/mL kanamycin, and 0.2% D-glucose. Three colonies were then inoculated into 2-5 mL of LB supplemented with 0.2% D-glucose and the same antibiotics and grown overnight at 37°C. The next day, 2 mL of the overnight culture was subcultured into 100 mL of LB supplemented with antibiotics in a 500 mL Erlenmeyer flask and grown at 37°C and 250 rpm to an OD_{600} of 1.5-2. The cells were induced with 0.2% L-arabinose to initiate expression of the OST and grown overnight (16-20 hours) at 30°C. The next day, the cultures were induced with 100 µM IPTG to initiate expression of the acceptor protein and grown for another 8 hours at 30°C. The cells were then harvested by centrifugation and frozen at -20°C until further processing.

To extract the acceptor protein, cells were resuspended in 10 mL buffer containing 50 mM NaH2PO4, pH 8, and 300 mM NaCl and lysed by homogenization using a C5 Emulsiflex homogenizer (Avestin). The lysate was centrifuged to remove cellular debris, the lysate buffer was adjusted to contain 10 mM imidazole, and the solution was rotated with HisPur Ni-NTA resin (Thermo Scientific) for one hour at 4°C. Following incubation of the histidine-tagged protein with Ni-NTA beads, the lysate was poured into a gravity flow column (Bio-Rad), washed with buffer containing 50 mM NaH2PO4, pH 8, 300 mM NaCl, and 10 mM imidazole, and eluted with the same buffer

containing 300 mM imidazole. The protein elution fractions were stored at 4°C for analysis by Western blotting.

Western blot analysis

For Western blot analysis of MBP-GCG, the protein elution fraction was diluted 3:1 in 4x NuPAGE LDS sample buffer (Invitrogen) supplemented with 10% beta-mercaptoethanol (v/v). Sample was boiled at 100°C for 10 min and subjected to SDS-polyacrylamide gel electrophoresis on 8% Bolt™ Bis-Tris Plus Protein Gels (Invitrogen). The proteins were then transferred to polyvinylidene difluoride (PVDF) membranes. For analysis of membranes by anti-His antibodies, the membranes were blocked overnight with 5% (w/v) milk in TBST (TBS, 0.1% (v/v) Tween 20) and then probed with horseradish peroxidase (HRP) conjugated anti-His antibody (1: 5,000) (Abcam, catalog # ab1187) for 1 hour. For analysis of membranes by ConA, the membranes were blocked overnight with 5% (w/v) bovine serum albumin (BSA) in TBST. The membranes were then incubated with 2 µg/mL ConA-biotin (C2272) resuspended in PBS pH 6.8 containing 10µM $CaCl_2$ and 10µM $MnCl_2$, followed by incubation with ExtrAvidin- Peroxidase (1:10,000) (MilliporeSigma, catalog #E2886) for 1 h. Between blocking and incubation steps, the membranes were washed three times with TBST for 5 min. Prior to imaging, the membranes were washed five times with TBST for 5 min and then visualized with Clarity ECL substrate (Bio-Rad) using a ChemiDocTM MP Imaging System (Bio-Rad). For analysis by anti-His, roughly 1.5-2 µg of purified protein was loaded into each lane. For analysis by ConA, roughly 15-20 µg was loaded.

Generation of molecular models

The structures of *Lb*STT3_2, *Lb*STT3_3, and *Li*STT3_1 were modeled using ColabFold[42] through homology search of MMseq2 with AlphaFold2. The 3D structures were aligned to the molecular structure of *Cl*PglB from PDB 5ogl. Based on these structures, the electrostatics surface visualizations were created using the APBS Electrostatics Plugin in PyMOL The acceptor peptide, GDQNAT, from PDB 5ogl was modeled in the STT3 structures based on the alignment of the STT3s with *Cl*PglB.

3.5 Acknowledgements

This chapter represents unpublished work by AA, JC, AP, and MD. AA is the first author and led the experimental design, generation of data, and data analysis. JC contributed to generation of data shown in Figure 3c. AP contributed to cell preparation and culturing that led to data generated in this work. AA and MD contributed to project conceptualization.

We thank Maithili Deshpande for helping to clone members of the SST3 panel library during her rotation in the lab. We thank Dr. Sai Pooja from the Gray lab at Johns Hopkins University for providing PDB files of the predicted 3D structures of LbSTT3_2, LbSTT3_3, and LiSTT3_1 aligned to the crystalized structure of ClPglB using ColabFold. The authors thank the Genomics Facility of the Biotechnology Resource Center at the Cornell Institute of Biotechnology for help with sequencing experiments

References

1. Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4305
2. Van Landuyt, L., Lonigro, C., Meuris, L. & Callewaert, N. Customized protein glycosylation to improve biopharmaceutical function and targeting. *Curr. Opin. Biotechnol.* **60**, 17–28 (2019).
3. Dammen-Brower, K. *et al.* Strategies for Glycoengineering Therapeutic Proteins. *Front. Chem.* **10**, 1–27 (2022).
4. Rosano, G. L., Morales, E. S. & Ceccarelli, E. A. New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Sci.* **28**, 1412–1422 (2019).
5. Wacker, M. *et al.* N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science (80-.).* **298**, 1790–1793 (2002).
6. Wacker, M. *et al.* Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7088–7093 (2006).
7. Feldman, M. F. *et al.* Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* (2005). doi:10.1073/pnas.0500044102
8. Harding, C. M. & Feldman, M. F. Glycoengineering bioconjugate vaccines, therapeutics, and diagnostics in E. coli. *Glycobiology* (2019). doi:10.1093/glycob/cwz031
9. Shang, W. *et al.* Production of human blood group B antigen epitope conjugated protein in Escherichia coli and utilization of the adsorption blood group B antibody. *Microb. Cell Fact.* **15**, 138 (2016).
10. Hug, I. *et al.* Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M111.287755
11. Keys, T. G. *et al.* A biosynthetic route for polysialylating proteins in Escherichia coli. *Metab. Eng.* **44**, 293–301 (2017).
12. Zhu, J. *et al.* An Engineered Pathway for Production of Terminally Sialylated N-glycoproteins in the Periplasm of Escherichia coli. *Front. Bioeng. Biotechnol.* **8**, 1–9 (2020).
13. Valderrama-Rincon, J. D. *et al.* An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat. Chem. Biol.* (2012). doi:10.1038/nchembio.921
14. Glasscock, C. J. *et al.* A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation. *Metab. Eng.* **47**, 488–495 (2018).
15. Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966 (2006).
16. Ielmini, M. V. & Feldman, M. F. Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology* **21**, 734–742

(2011).
17. Ollis, A. A. *et al.* Substitute sweeteners: Diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. *Sci. Rep.* (2015). doi:10.1038/srep15237
18. Mills, D. C. *et al.* Functional analysis of N-linking oligosaccharyl transferase enzymes encoded by deep-sea vent proteobacteria. *Glycobiology* **26**, 398–409 (2016).
19. Wild, R. *et al.* Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science (80-.).* **359**, 545–550 (2018).
20. Izquierdo, L. *et al.* Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. *EMBO J.* **28**, 2650–2661 (2009).
21. Izquierdo, L., Mehlert, A. & Ferguson, M. A. The lipid-linked oligosaccharide donor specificities of Trypanosoma brucei oligosaccharyltransferases. *Glycobiology* **22**, 696–703 (2012).
22. Jinnelov, A., Ali, L., Tinti, M., Güther, M. L. S. & Ferguson, M. A. J. Single-subunit oligosaccharyltransferases of Trypanosoma brucei display different and predictable peptide acceptor specificities. *J. Biol. Chem.* **292**, 20328–20341 (2017).
23. Poljak, K. *et al.* Analysis of substrate specificity of Trypanosoma brucei oligosaccharyltransferases (OSTs) by functional expression of domain-swapped chimeras in yeast. *J. Biol. Chem.* **292**, 20342–20352 (2017).
24. Ramírez, A. S. *et al.* Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins. *Glycobiology* **27**, 726–733 (2017).
25. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J. & Aebi, M. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. *Mol. Biol. Cell* (2008). doi:10.1091/mbc.E08-05-0467
26. Hese, K., Otto, C., Routier, F. H. & Lehle, L. The yeast oligosaccharyltransferase complex can be replaced by STT3 from Leishmania major. *Glycobiology* **19**, 160–171 (2009).
27. Choi, B. K. *et al.* Improvement of N-glycan site occupancy of therapeutic glycoproteins produced in Pichia pastoris. *Appl. Microbiol. Biotechnol.* **95**, 671–682 (2012).
28. Nasab, F. P., Aebi, M., Bernhard, G. & Frey, A. D. A combined system for engineering glycosylation efficiency and glycan structure in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **79**, 997–1007 (2013).
29. Castilho, A. *et al.* An oligosaccharyltransferase from Leishmania major increases the N-glycan occupancy on recombinant glycoproteins produced in Nicotiana benthamiana. *Plant Biotechnol. J.* **16**, 1700–1709 (2018).
30. Mule, S. N. *et al.* Protein glycosylation in: Leishmania spp. *Mol. Omi.* **16**, 407–424 (2020).
31. Ramírez, A. S. *et al.* Characterization of the single-subunit

oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology* (2017). doi:10.1093/glycob/cwx017
32. Castro, O., Movsichoff, F. & Parodi, A. J. Preferential transfer of the complete glycan is determined by the oligosaccharyltransferase complex and not by the catalytic subunit. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 14756–14760 (2006).
33. Li, G., Yan, Q., Nita-Lazar, A., Haltiwanger, R. S. & Lennarz, W. J. Studies on the N-glycosylation of the subunits of oligosaccharyl transferase in Saccharomyces cerevisiae. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M410969200
34. Bokhari, H., Maryam, A., Shahid, R. & Siddiqi, A. R. Oligosaccharyltransferase PglB of Campylobacter jejuni is a glycoprotein. *World J. Microbiol. Biotechnol.* **36**, 1–12 (2020).
35. Atwood, J. A. *et al.* Glycoproteomics of Trypanosoma cruzi trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *J. Proteome Res.* **5**, 3376–3384 (2006).
36. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* (2011). doi:10.1038/nature10151
37. Napiórkowska, M. *et al.* Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat. Struct. Mol. Biol.* (2017). doi:10.1038/nsmb.3491
38. Poljak, K. *et al.* Analysis of substrate specificity of Trypanosoma brucei oligosaccharyltransferases (OSTs) by functional expression of domain-swapped chimeras in yeast. *J. Biol. Chem.* (2017). doi:10.1074/jbc.M117.811133
39. Lefebre, M. D. & Valvano, M. A. Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in Burkholderia cepacia complex isolates. *Appl. Environ. Microbiol.* **68**, 5956–5964 (2002).
40. Castro, O., Movsichoff, F. & Parodi, A. J. Preferential transfer of the complete glycan is determined by the oligosaccharyltransferase complex and not by the catalytic subunit. *Proc. Natl. Acad. Sci. U. S. A.* (2006). doi:10.1073/pnas.0607086103
41. Li, A., Acevedo-Rocha, C. G. & Reetz, M. T. Boosting the efficiency of site-saturation mutagenesis for a difficult-to-randomize gene by a two-step PCR strategy. *Appl. Microbiol. Biotechnol.* **102**, 6095–6103 (2018).
42. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, (2022).

4 FUTURE DIRECTIONS

4.1 Introduction

The costs of glycoprotein therapeutics, and in particular monoclonal antibody treatments, are exceptionally high, in part due to the high manufacturing costs. These costs arise from the high costs of mammalian cell culture as well as the difficulty in controlling and isolating desired glycoforms. This places a large financial burden on patients who need these life-saving treatments and the healthcare system and makes it challenging to produce glycoproteins outside of centralized manufacturing systems at large biopharmaceutical companies. To develop cheaper, alternative manufacturing strategies, our lab focuses on engineering cell-free and *E. coli*-based glycoprotein production platforms.

In this dissertation work, *E. coli*, which have no native glycosylation systems, were engineered to produce glycoproteins. Glycoengineered *E. coli* were then used to generate machinery for cell-free systems or were used as a production platform in cell-based methods. In cell-free manufacturing, glycosylation machinery was engineered into *E. coli* and then extracted to directly manufacture glycoproteins outside of the cell. For this, we used microfluidic chips that separated key processes- synthesis of the protein, addition of a sugar group to the protein, and purification of the resulting glycoprotein- into separate compartments that were individually optimized. Such a system provides a platform for small-batch, decentralized production of glycoprotein therapeutics and can serve as a laboratory tool for studying how uniform glycan structures affect glycoprotein function. In cell-based manufacturing, glycosylation

machinery was engineered into the periplasmic space, a compartment between the inner and outer membranes of the *E. coli*, where the target protein was efficiently glycosylated. For this project, we identified single subunit oligosaccharyltransferases from protozoa to manufacture more human-like glycoproteins in *E. coli*. However, there are still many areas of this research that need further investigation.

4.2 Discussion

4.2.1 Further exploration of microfluidic device for glycoprotein synthesis

In Chapter 2, a microfluidic device was developed that compartmentalized three bioprocesses— protein synthesis, protein glycosylation, and protein enrichment. The results showed that introduction of flow into the system improved protein synthesis and that tethering of the biocatalyst, PglB, enabled efficient glycosylation within the device. To expand upon this work, different microfluidic designs could be explored to improve each of the process steps. For example, to improve mixing, designs that introduce chaotic flows such as the staggered herringbone mixer[1] could be tested. To increase the reaction rate, a packed bed design could be used to increase the amount of enzyme present in the system. Additionally, although we focused here mainly on the benefits of flow and spatiotemporal control, microfluidic technologies can also be used to extract kinetic parameters and for real-time monitoring of reaction components and conditions. These data would improve the ability to model and optimize the reaction conditions at each step. Lastly, a microfluidic system for glycoprotein production could potentially be used for bedside or pharmacy side production of small quantities of therapeutics. This would be beneficial for personalized medicine that needs to be tailored for each

patient. For example, the glycan structure could be changed to try to reduce immunogenicity of the drug for each patient or to produce a drug for rare diseases where it is not profitable to manufacture at a large scale.

4.2.2 Continued exploration of single subunit eukaryotic STT3s

In Chapter 3, we demonstrated, for the first time, that eukaryotic OSTs can function in the context of *E. coli*. This has opened avenues for both further study of STT3s and their use in development of protein glycosylation pathways in *E. coli*. In particular, we showed that three of the STT3 candidates preferentially transferred the $Man_3GlcNAc_2$ glycan to sequons with D or E in the -2 position. This begs the question of whether the other STT3 candidates also exhibit this same protein substrate specificity or whether they are more relaxed. One way to test this would be to rescreen the STT3 panel using a different glycosylation sequence. For example, to glycosylate monoclonal antibodies at their native sites, the STT3s would need to be able to glycosylate a QYNST site. Additionally, looking at the original screen, there were some hits that showed lower glycosylation efficiency of the DQNAT site than the top three STT3 hits. It is possible that although these OSTs have lower efficiency of transferring to the DQNAT site, they may have more relaxed sequon specificities overall.

Another question that was not explored is what the glycan substrate specificities of the STT3s are. Kinetoplastid STT3s natively transfer high mannose structures[2,3], but many STT3 paralogs are able to transfer truncated glycans[4,5]. Notably, *Tb*STT3A is able to transfer a disaccharide, $GlcNAc_2$[5]. This suggests that some STT3s may have relaxed glycan substrate specificities and could potentially be used to transfer complex *N*-

glycans *en bloc*, reducing the number of glycan remodeling steps. To synthesize a minimal complex *N*-glycan, two additional enzymes, *N*-acetylglucosaminyltransferase I (GnTI) and II (GnTII), would need to be added to the $Man_3GlcNAc_2$ biosynthesis pathway. To generate this pathway, it may be beneficial to use the protozoan glycosyltransferases from *T. brucei*, *Tb*GnTI and *Tb*GnTII. This is because, unlike what occurs in higher eukaryotes, in *T. brucei Tb*GnTII is not dependent on prior action by *Tb*GnTI to function and $Man_3GlcNAc_2$ is a better substrate for *Tb*GnTI than $Man_5GlcNAc_2$ which is the preferred substrate for GnTI in mammalian cells[6].

4.2.3 Integration of eukaryotic STT3s into cell-free systems

One of the main motivations for the exploration of OSTs that are composed of just a single subunit is that these enzymes are likely to be translatable to cell-free systems. To tie together ideas from Chapters 2 and 3, STT3s that efficiently transfer eukaryotic glycans could be integrated into a microfluidic device, like how PglB was tethered in Chapter 2. Using the $Man_3GlcNAc_2$ structure as a base, the glycan can then be further elaborated by the glycosyltransferases, GnTI, GnTII, galactosyltransferase (GalT), and sialyltransferase (SiaT), to build a complex *N*-glycans commonly found on therapeutic proteins (**Fig. 4.1**). The spatiotemporal control provided by the microfluidic system could potentially improve conversion efficiency at each step to enable production of structurally uniform eukaryotic glycoprotein therapeutics.

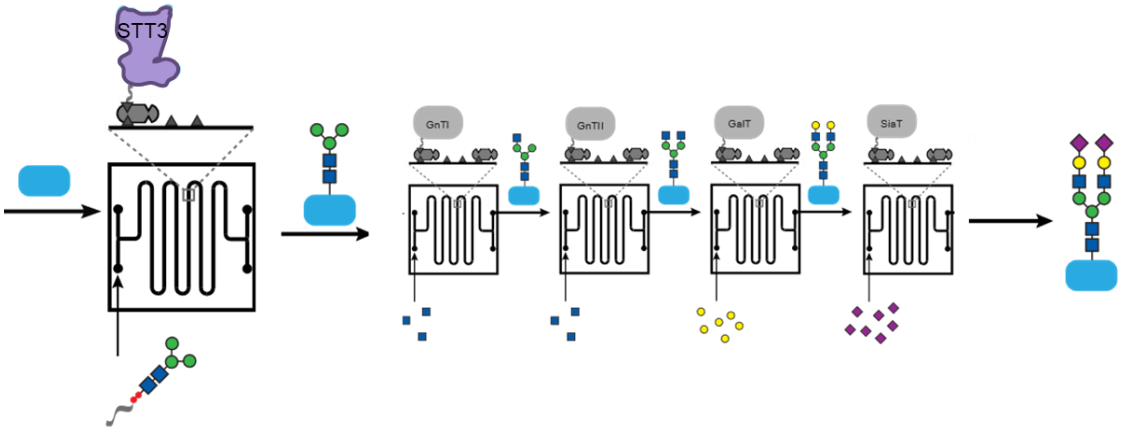


Figure 4.1 Schematic of microfluidic system for sequential glycosylation reactions. Schematic of modular system of sequential, compartmentalized glycosylation reactions facilitated by immobilized glycosylation enzymes. In the first step, $Man_3GlcNAc_2$ is attached to the acceptor protein by the STT3. It is then sequentially modified into a sialylated complex *N*-glycan by the action of glycosyltransferases.

References

1. Stroock, A. D., Dertinger, S. K., Ajdari, A., Mezic, I., Stone, H. A., & Whitesides, G. M. Chaotic mixer for microchannels. *Science*, *295*(5555), 647-651 (2002).
2. Izquierdo, L., Mehlert, A. & Ferguson, M. A. The lipid-linked oligosaccharide donor specificities of Trypanosoma brucei oligosaccharyltransferases. *Glycobiology* **22**, 696–703 (2012).
3. Mule, S. N. *et al.* Protein glycosylation in: Leishmania spp. *Mol. Omi.* **16**, 407–424 (2020).
4. Nasab, F. P., Aebi, M., Bernhard, G. & Frey, A. D. A combined system for engineering glycosylation efficiency and glycan structure in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* (2013). doi:10.1128/AEM.02817-12
5. Ramírez, A. S. *et al.* Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins. *Glycobiology* **27**, 726–733 (2017).
6. Damerow, M., Rodrigues, J. A., Wu, D., Güther, M. L. S., Mehlert, A., & Ferguson, M. A. Identification and functional characterization of a highly divergent N-acetylglucosaminyltransferase I (TbGnTI) in Trypanosoma brucei. *Journal of Biological Chemistry*, *289*(13), 9328-9339 (2014).