CELL-FREE SYNTHETIC GLYCOBIOLOGY SYSTEM AS A NOVEL ROUTE TO CUSTOMIZED GLYCOTHERAPEUTICS AND VACCINES

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

Thapakorn Jaroentomeechai

August 2021

CELL-FREE SYNTHETIC GLYCOBIOLOGY SYSTEM AS A NOVEL ROUTE TO CUSTOMIZED GLYCOTHERAPEUTICS AND VACCINES

Thapakorn Jaroentomeechai, Ph. D.

Cornell University 2021

Glycans and glycosylated biomolecules are directly involved in almost every biological process as well as the etiology of most major diseases. Hence, glycoscience knowledge is essential to efforts aimed at addressing fundamental challenges in understanding and improving human health. While much progress has been made, there remains an urgent need for new tools that can overexpress structurally uniform glycans and glycoconjugates in the quantities needed for characterization and that can be used to mechanistically dissect the enzymatic reactions and multi-enzyme assembly lines that promote their construction. To address this technology gap, we develop the cell-free synthetic glycobiology system as a simplified and highly modular framework to investigate, prototype, and engineer pathways for glycan biosynthesis and protein glycosylation outside the confines of living cells. First, we engineered a novel cell-free glycoprotein synthesis (CFGpS) system that seamlessly integrates protein biosynthesis with asparagine-linked (*N*-linked) or serine/threonine-linked (*O*-linked) protein glycosylation. This technology leveraged a glyco-optimized *Escherichia coli* strain to source crude extracts that were selectively enriched with glycosylation components. The resulting extracts enabled a one-pot reaction scheme

for efficient and site-specific glycosylation of therapeutic proteins including biologically-active human erythropoietin. Next, we expanded the utility of the cell-free glycosylation system for an *in vitro* bioconjugate vaccine expression (iVAX) in lysates derived from detoxified, nonpathogenic *E. coli*. We demonstrated that iVAX synthesized vaccines against *Franciscella tularensis* subsp. tularensis (type A) strain Schu S4 conferred complete protection in an intranasal mouse model of *F. tularensis* infection. Finally, we developed a collection of the cell-free glycan remodeling modules, enabled by the creation of a glycosyltransferase high-level expression library. The glycan remodeling modules are facile and highly modular, providing an efficient platform for rapid biosynthesis of authentic, complex human *N*-glycans. Together, our cell-free synthetic glycosylation system effectively broadens the glycoengineering toolbox and is anticipated to facilitate fundamental understanding in glycoscience and make possible new applications in on-demand biomanufacturing of glycoprotein therapeutics and vaccines.

BIOGRAPHICAL SKETCH

Thapakorn Jaroentomeechai was raised by parents Ruangsak and Chaweewan Tomeechai in Bangkok and Phitsanulok, Thailand. He attended Anuban Phitsanulok Elementary School where he developed an interest in impromptu speech competition and dreamt of becoming a professional MC. Thapakorn then attended middle and high school at the Bodindecha (Sing Singhaseni) II School where he was encouraged to explore his passion and aptitude for science by Mrs. Oun-a-nan, Mrs. Wongdee, Mr. Vi-jitra, and Dr. Sombatsompop. In grade 9th, Thapakorn was selected as a finalist for the National Math and Science Olympiad program. He also attended International Asian-Pacific Science Conference as one of the Thai national representatives for junior scientist. From grade 10th-12th, Thapakorn was selected as a National Junior Science Talent Program scholar. During these times, Thapakorn was quite fond of chemistry and began to perform chemistry experiments in his bedroom. Following several experiment[1]-gone-wrong incidents which involved fire alarm, a destruction of his mom's favorite plants, and several complaints from neighbors, Thapakorn and his family agreed it would be better if he received a formal training in science anywhere but home. As a result, after graduating high school with *Summa Cum Laude*, Thapakorn joined the Department of Chemistry at Mahidol University as Sri-Trang-Thong scholar. There, he was fortunate to secure an undergraduate research position under supervision of Dr. Vuthichai Ervithayasuporn. In addition, he spent several months working in the Dr. Masafumi Unno research group at the Gunma University, Japan. Thapakorn also served as the class president and later as the student government president for the Faculty of Science at Mahidol University.

[1] One of which was the carbon snake experiment, which is the dehydration reaction of sugar by sulfuric acid. Perhaps he found an excitement in the carbohydrate science before he even knew it!

Upon graduation with *Cum Laude* from Mahidol University with B.S. in Chemistry in 2011, Thapakorn was awarded the Royal Thai Government Fellowship to pursue higher education in the area of his choice. Thapakorn began his graduate degree in the School of Chemical and Biomolecular Engineering at Cornell University in 2013 where he joined Dr. Matthew P. DeLisa research group. He also spent a stint as research assistant in the Antibody Discovery group at Biogen Inc., under the supervision of Dr. Hua Wang. During his time at Cornell, Thapakorn was awarded Samuel C. Fleming Family Graduate Fellowship and won the Austin Hooey Graduate Research Excellence Recognition Award. In 2021, Thapakorn was awarded a fellowship from the European Molecular Biology Organization to begin his postdoctoral research position in the Dr. Henrik Clausen's group at the Copenhagen Center for Glycomics, University of Copenhagen, where he will continue to pursue his passion in glycobiology research.

To my mom, whose words encourage.

To my dad, whose actions inspire.

And to my sister, whose love always stays with me.

ACKNOWLEDGMENTS

My intellectual journey to attain a research doctorate is long and arduous, yet not a solitary one. Throughout my time at, and prior, Cornell, I was deeply indebted to many, whose kindness, support, and friendship have made this journey a more tolerable and meaningful one. To mom and dad, thank you for your incredible support throughout my life and for ensuing I have had the chance to pursue my dreams growing up. To my sister, your courage and free spirit are always inspiring, and I look forward to welcoming your baby girl to our family soon. To my cousins, relatives, and family friends, thank you for your supports and for always being kind to our family throughout these years, especially while I was 8,544 miles away from home.

To my advisor Dr. Matthew DeLisa, it was my great pleasure and a privilege to learn under your tutelage. Matt, thank you for your excellent vision and continued mentorship. Transitioning to glycobiology was one of the best decisions in my life. However, without your guidance, I would not have gotten far in the complex world of glycobiology. I am also very grateful to Dr. Michael Jewett for his scientific advice, insightful discussion, and wonderful collaboration throughout my entire PhD program. I also would like to thank my past and current members of my thesis committee, Dr. Susan Daniel, Dr. Yimon Aye, and Dr. Julius Lucks, for their helpful career advice and scientific suggestions.

I am forever thankful to my undergraduate advisor, Dr. Vuthichai Ervithayasuporn. Aj. Golf has been one of my best role models for a scientist, a teacher, and a researcher. Aj. Golf, thank for always encouraging me to be a better version of myself and never give up. Also, your enthusiasm in research and teaching is contagious.

To past and present members of the DeLisa group, thank for making our laboratory an incredible working environment. To my DLRG mentors, Drs. Anne, Yi, Dario, Taylor, M-P, May, Bunyarit, and Aravind, I would not have made it through my early years without your technical training and suggestions, and I will forever be grateful for everything you have taught me. I would like to specifically thank Dr. Bunyarit for introducing me to the DLRG and being my first mentor in the lab. P' Pao, thanks for letting me use your bench in my first year and always being a great brother. I wish you every success in life and career. To my DLRG cohorts, Drs. Morgan, Emily, Cameron, Xiaolu, Tyler, and Kevin, we have made it! Thank for your friendship, advice, and solace during beer hours in these past years. I would like to extend my gratitude to my CBE cohort, Siyu, Zhu, and Taewon. You guys have always been a great friend who have provided support during tough time. I will fondly miss our time in the Olin hall.

To the special committee of Olin 303 –you have made my life in the laboratory never been boring. May, you are a calming force of our group. Our crossword puzzle game was always my favorite break while waiting for the experiment. Xiaolu, I always appreciate your straightforward and insightful comments for my projects. Thanks for brining so much joy to our group. Also, your memory is amazing. M-P is one of the most hard-working and meticulous scientists I ever know. Thanks for being a great role model for the group and for all the discussion in science and many other topics. Aravind, your enthusiasm never runs dry, and I value your friendship and support. Mingji, thank for always being kind and helpful. I will miss the time I played with your son, Eric. Sean, you have become great friend and colleague, and I really enjoy our conversation throughout the years.

To younger DLRG members, Alicia, Natalia, Tina, Weiyao, Erik, Primit, Sophia, Connor, Belen, Azmain, Suthara, and Abhishek, you are all personable and

great junior. I look forward to your success in graduate school and beyond. I was also fortunate to work with several bright, young scientists during my tenure in Olin hall. Amy, Olivia, Silvi (Adrian), Ben, Daniel K., and Ricardo, you all have made me proud in your unique way. I have learnt a great deal from mentoring each of you. Many of you have become my friends and I will always root for your success.

Numerous collaborators, many of whom have become my dear friends, have made the cell-free glycosylation project possible. Dr. Jessica, thanks for a wonderful collaboration and friendship since the beginning of my PhD. I hope we will continue to have a fruitful collaboration in the future. I also would like to thank many members of the Jewett group including Drs. James, Weston, Jasmine, and Katherina for their scientific inputs. Dr. Han-Yuan, I have learnt a great deal about membrane bilayer from you. We tackled some of the most challenging research projects together. Thank you for your endless enthusiasm and optimism. I wish you and your wife, Yuna, an incredible success in the future. To many member of the Daniel group, Dr. Rohit, Zachary, and Ferra, thanks for being both great collaborator and friends. You guys are also the awesome neighbor. Dr. Yung-Fu, I always appreciate your kindness and for allowing me to perform immunological experiments in your lab. Bryce, you have taught me so much about cryoEM. Thanks for always analyzing protein particles for us. I look forward to us solving the protein structure together! Dr. Sheng, Dr. Ruchika, Bob, and Beth at the Cornell Biotechnology Center, I can't thank each of you enough to share with me your knowledge and experience in mass spectrometry analysis, which is critical for making the body of this work possible. Dr. Joshua at Glycobia, your skills and knowledge in chemical glycobiology is truly inspiring and I will forever appreciate your kindness to me whenever I go to Glycobia. David Mace, thank you for introducing me to the world of biotechnology company startup. I wish you and Swiftscale team a continuing success in the future.

I am also fortunate to make friends with many outside the Olin hall. To my Thai student cohort at Cornell, Dr. Kitthikun (Obb), Dr. Chalermpat (Nong), Dr. Siriphat (Fay), Dr. Visarut (Earth), Bua, Ploy, and Billy, thank you for making me feel home whenever we see each other. I cherished our memories throughout the years at Cornell. Dr. Chinawat, Dr. Sunireee, Dr. Kullachate, P' Nual, P' Rawinthira, and P' Tom, thank you for taking care of me and always reaching out with kindness. To all Thai brothers and sisters I get to know at Cornell, you guys are amazing people and I will miss you dearly. Dr. Sang, you are one of my best friends from Cornell. From the time we introduced ourselves in the back of the room during Mass and Energy Balance class, you have been one of the most sincere and supportive friends I ever have, and I am looking forward for our life-long friendship. Daniel R., you have taught me how to live a fuller and richer life. Taking a trip with you was a highlight of my life and I look forward to travelling the world with you in the future. Alex, Irene, and Michael, who would have thought being in the same Chinese classes with you guys was just a beginning of many fun experiences. Thanks for being a great classmate and friends. Derek S., thanks for being an incredible best friend throughout these years. Your kind-heart, patience, and thoughtfulness have touched me in every which way I could imagine. You inspire me to be a better version of myself every day and I look forward to seeing where the future will take us.

This work was supported by numerous grants from the National Science Foundation, the National Institutes of Health, the Defense Threat Reduction Agency, and the Bill and Melinda Gates Foundation. Specific grant numbers are provided at the end of each chapter. Thapakorn was supported by a Royal Thai Government Fellowship and a Cornell Fleming Graduate Scholarship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ApoAI	apolipoprotein AI
CFGpS	cell-free glycoprotein synthesis
CFPS	cell-free protein synthesis
CFSG	cell-free synthetic glycobiology
CHO	Chinese hamster ovary cell line
C. jejuni	*Campylobacter jejuni*
CMP	cytidine monophosphate
CRM197	cross reacting materials 197
DNA	deoxyribonucleic acid
Dol	dolichol, dolichyl
E. coli	*Escherichia coli*
ELISA	enzyme-linked immunosorbent assay
EPO	erythropoietin
ER	endoplasmic reticulum
F. tularensis	*Francisella tularensis*
FDA	U.S. Food and Drug Administration
Fuc	fucose
FucT	fucosyltransferase
Gal	galactose
GalNAc	*N*-acetylgalactosamine
GalT	galactosyltransferase
GDP	guanosine diphosphate
GH	glycosyl hydrolasae
Glc	glucose
GlcA	glucoronic acid
GlcNAc	*N*-acetylglucosamine
GnT	*N*-Acetylglucosaminyltransferase
GT	glycosyltransferase
HEK293	human embryonic kidney 293 cell line
HPLC	high-performance liquid chromatography
iVAX	*in vitro* conjugate vaccine expression
LLOs	lipid-linked oligosaccharides
LPS	lipopolysaccharide
mAb	monoclonal antibody
MALDI-MS	matrix-assisted laser desorption/ ionization mass spectrometry
Man	mannose
ManT	mannosyltransferase
MBP	maltose binding protein
Neu5Ac	*N*-Acetylneuraminic acid
NGT	*N*-linked glycosyltransferase
OGT	*O*-linked glycosyltransferase
O-PS	*O*-polysaccharides
OST	oligosaccharyltransferase
P	phosphate
PCR	polymerase chain reaction
pgl	protein glycosylation locus
ppGalNAcT	polypeptide *N*-acetylgalacto-saminyltransferase
scFv	single chain variable fragment
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sfGFP	superfolder green fluorescent protein
SIMPLEx	solubilization of integral membrane protein with high level of expression
SRP	signal recognition particle
ST	sialyltransferase
UDP	uridine diphosphate
Und	undecaprenol, undecaprenyl
Xyl	xylose

CHAPTER 1

INTRODUCTION[1]

1.1 Glycoscience: it is time for a sweet transformation.

Humans love sugar. From the evolutionary perspective, carbohydrate-rich foods provide an efficient calories source, and contribute significantly to a survival of our ancestors, Hominidae or great apes, in nature. It is not surprising that we develop favor toward sugar as well as its natural and man-made products including fruit, syrup, honey, beer, wine, etc. A systematic study of sugar has dated back to the late 19 century where Emil Fischer attempted to synthesize glycan using esterification reaction and devised a stereoisomeric diagram for drawing monosaccharides (Fischer, 1890), research that won him a Nobel Prize in Chemistry in 1949. Subsequent breakthrough in the field was a study by Karl Landsteiner (Nobel Prize in Chemistry in 1930) who discovered ABO blood group system and its significance in blood transfusion. The following work by Morgan and Watkins in 1952 revealed carbohydrate as factor determinant for blood group in human, which in turn has sparked further interest in carbohydrate research. In 1949, L.F. Leloir (Nobel Prize in Chemistry in 1970) identified the role of nucleotide sugars in the biosynthesis of glycan and laid a foundational work in understanding glycan processing enzymes (Figueroa

1 Parts of this chapter have been published in the Journal of Frontiers in Chemistry:

Jaroentomeechai, T., Taw, M.N., Li, M., Aquino, A., Agashe, N., Chung, S., Jewett, M.C., and DeLisa, M.P. (2020) Cell-free synthetic glycobiology: designing and engineering glycomolecules outside the cell. *Front Chem.* 8: 645.

et al., 2021; Leloir LF, 1953). Following these seminal works, many other types of glycan and its associated biosynthesis pathway have been identified and characterized. The knowledge in structure and biosynthesis of natural glycan and glycosylated molecule formed the basis for the development of the glycoscience discipline.

By today's definition, glycoscience is a study of structure, synthesis, and function of carbohydrates. Carbohydrates or glycans are ubiquitous in nature and can be found across all domains of life – from microorganism including prion to animal and plant (Varki, 2017a). Glycans participate in a myriad of biological processes including development, immunity, homeostasis, and pathogenicity (Helenius and Aebi, 2001b; Ohtsubo and Marth, 2006; Peixoto et al., 2019; Shental-Bechor and Levy, 2009; Skropeta, 2009). These processes involve direct glycan recognition or indirect effect whereby glycans provide specific effect on the recipient biomolecules. Glycan also features prominently in diseases, for examples, tumor cells express atypical level of glycan on its surface (Peixoto et al., 2019). Carbohydrates are also the important part of materials such as cellulose and fiber. Therefore, our knowledge of glycoscience will expand our understanding in human health and disease, generating new materials, and discovering more efficient energy source (Walt, 2012).

Modern glycoscience primarily comprised of glycochemistry and glycobiology. A research in the 19th century laid foundation and established an important knowledge that: (i) glycans are present prevalently on cell surface, extracellular matrix, and inside the cell; (ii) the structure, in particular stereoisomer, of glycans are important for their

function and antigenicity. In addition, glycan structures are genetically determined; and (iii) a change in the glycan structures can result in human and animal diseases. These knowledges have spawned interests in glycan synthesis, glycoenzyme characterization, and glycan binding molecule discovery. Such research interest, in turn, has led to a discovery of new class of glycan structures, glycosylation pathway, and glycan-associated disease as evidenced by an exponentially increase in scientific literatures related to glycoscience (**Figure 1-1**).

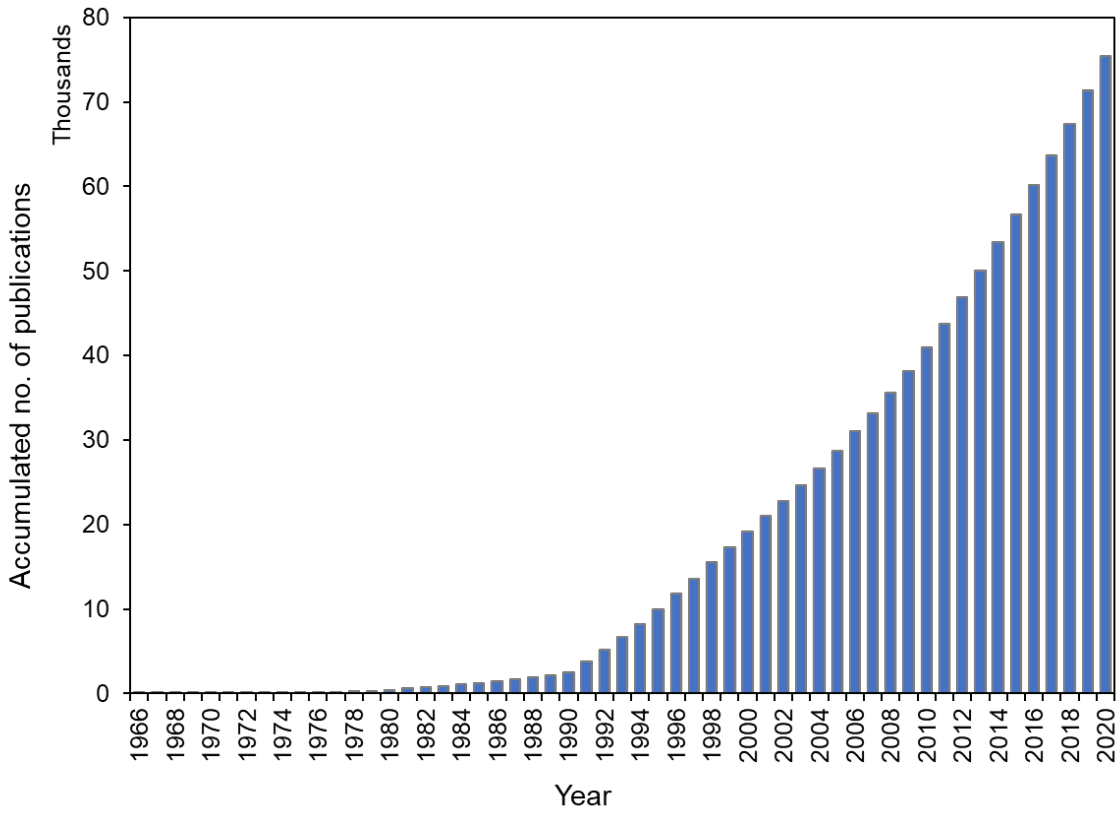


Figure 1-1. Number of publications including keywords 'Glycan' or 'Glycosylation' in the database. Both keywords were used to simultaneously search within the Thomson Reuters' Web of Science portal. Publication contains both keywords is counted only once. Note that many publications before 1945 are not listed in the database, hence, are not included in this dataset. Data was retrieved using built-in analytical research on February 23, 2021.

These developments, notwithstanding, the technology for study carbohydrates is significantly behind their nucleic acids and protein counterparts. In nature, molecular evolution has relatively conserved approach to the synthesis of polynucleotides and polypeptides, as evidenced by the fact that they are synthesized by 5 nucleotides and 20 amino acids[2], respectively. Moreover, these molecular building blocks are generally strung together in a linear fashion with similar covalent linkages. In stark contrast, glycans can be synthesized from over 100 different monosaccharide building blocks that can be connected in many ways. Both constitutional and stereo-isomers commonly exists in the carbohydrate chemistry. This offers both advantages and challenges in synthesis, characterization, and analysis of glycan. Moreover, glycans are seldom a sole player in biological processes as they are typically found linked with other biomolecule complements including small molecule, lipid, protein, and nucleic acid. Hence the study of carbohydrate does not encompass only the glycan itself but also the underlying biomolecule that it connects to. Glycan and its entire complements within an organism are referred to as glycome. A systematic study of glycome within a given organism is termed glycomics, which is a subject of intensive study in the past few decades.

The overarching goal in the current glycoscience community is to develop new tools and methodologies for characterize, synthesize, and engineer complex glycans

2 These numbers refer only to the canonical building blocks. Many organisms synthesize biopolymers using modified nucleotides and/or amino acids. Such topic is beyond the scope of this dissertation.

and glycomolecules. Akin to the history of DNA and protein studies, the ability to read, write, and edit glycans seem a daunting task but the potential reward from deciphering the glycocode in nature are astonishing. It is time for the scientific community to transform glycoscience into a major discipline in order to realize its fullest potential in basic and translational science.

1.2 Protein glycosylation in nature.

Glycoprotein is ubiquitous in nature. Indeed, protein modification by carbohydrate is one of the most common protein post-translational modifications across kingdom of life. More than 40 different types of carbohydrate-to-protein linkages have been identified to date (Varki, 2017b). Among these, glycan installation at the asparagine (*N*-linked) and serine/threonine (*O*-linked) residues constitutes the greatest proportion of glycoproteins (Spiro, 2002). Other notable types include C-mannosylation at the tryptophan amino acid, glycosaminoglycans, and glycosylphosphatidylinositol (GPI) anchors to peptide backbones (Gagneux and Varki, 1999; Spiro, 2002). Glycan associated proteins are found ubiquitously in both extra- and intracellular matrix as well as at the membrane interface. In mammalian, glycans and glycoconjugates are assembled from 14 basic monosaccharides (Herget et al., 2008) that give rise to a vast number of possible carbohydrate structures. In prokaryotes, the diversity of monosaccharide building blocks increases significantly to over 140 structures, hence their glycan structures are even more diverse. Glycan diversity in

both Eukarya and Prokarya are also differ in term of carbohydrate size, branching index, and glycan charge distributions (Egorova et al., 2015; Herget et al., 2008). While they share some similarities in the glycan biosynthesis and protein glycosylation processes, eukaryotic and prokaryotic cells display distinctions in terms of glycan biosynthesis localization, available modes of protein glycosylation, and amino acid sequences to which glycan can be attached.

1.2.1 Eukaryote protein glycosylation pathway.

Eukarya comprises of diverse organisms, ranging from single cell organism such as yeast and most protists to complex animals including human. This vast diversity also displays in the protein glycosylation within these organisms. For example, primitive eukaryotes (protozoa) utilize single subunit oligosaccharyltransferase enzyme, called Stt3, for *N*-glycosylating their proteins. In contrast, Stt3s in all higher eukaryotes (mammals) are multi-subunit protein complexes where Stt3 catalysts are accompanied by eight membrane proteins that are necessary to carry Stt3's glycosylation functions (Bai et al., 2018). In human, the established glycoproteome is generated through 16 distinct glycosylation pathways that have been characterized based on the type of protein-carbohydrate linkage, monosaccharide at the reducing end of the glycan, as well as glycosyltransferase enzyme that catalyzes protein-carbohydrate linkage formation (Schjoldager et al., 2020). Collectively, there are about 700 genes encoding for enzymes, transporters, and chaperones that work in

glycoprotein biosynthesis (Hansen et al., 2020; Moremen et al., 2012). This represents about 3-4% of the entire human genome. Among these, over 200 genes are encoded for glycosyltransferases (GTs) which are the key enzymes in complex glycan biosynthesis. GTs transfer monosaccharide unit from donor molecule to the glycone or aglycone acceptor. Most GTs exhibit sugar donor, acceptor, and linkage specificity, a unique feature that can be exploited in the glycoengineering application. Most GTs are type II transmembrane proteins that reside at the membrane interface within the lumen of the endoplasmic reticulum (ER) and the lumina of the cis-, medial-, or trans-cisternae of the Golgi apparatus (**Figure 1-2**). GTs' localization within ER/Golgi secretory pathway is quite specific, and the dislocation of some GTs can lead to human diseases. Many of these GTs are characterized as Leloir GTs for their ability to utilize nucleotide sugar as substrate. Some ER-resident GTs are multi pass membrane proteins and they typically use lipid-linked oligosaccharides (LLOs) substrates, specifically mannose or glucose-linked dolichol pyrophosphate. Notably, several evidences suggest glycosylation also play a role in regulating function and localization of GTs, its own synthesis enzyme (Mikolajczyk et al., 2020). In addition, some complex GTs are also discovered and being studied for its biological significance (Kellokumpu et al., 2016). The number of new glycosylation pathway and its associated GT genes are still being discovered (Hirata et al., 2018; Larsen et al., 2017; Yoshida-Moriguchi and Campbell, 2015).

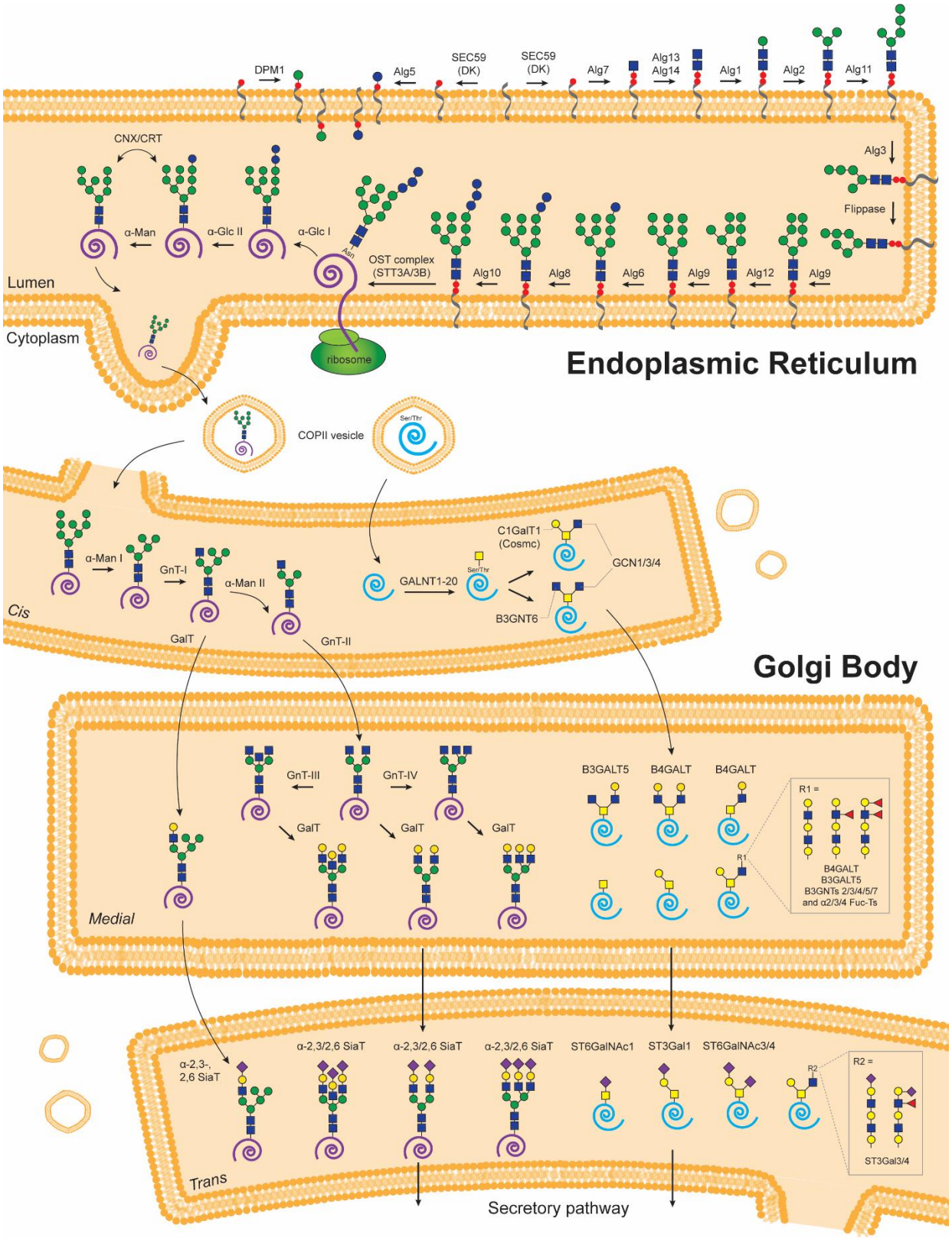

DPM1
Alg5
SEC59 (DK)
SEC59 (DK)
Alg7
Alg13 Alg14
Alg1
Alg2
Alg11
Alg3
Flippase
CNX/CRT
α-Man
α-Glc II
α-Glc I
Asn
OST complex (STT3A/3B)
Alg10
Alg8
Alg6
Alg9
Alg12
Alg9
Lumen
Cytoplasm
ribosome
Endoplasmic Reticulum
COPII vesicle
Ser/Thr
α-Man I
GnT-I
α-Man II
GALNT1-20
Ser/Thr
C1GalT1 (Cosmc)
GCN1/3/4
B3GNT6
Cis
GalT
GnT-II
Golgi Body
GnT-III
GnT-IV
GalT
GalT
GalT
B3GALT5
B4GALT
B4GALT
R1 =
R1
B4GALT
B3GALT5
B3GNTs 2/3/4/5/7
and α2/3/4 Fuc-Ts
Medial
α-2,3-, 2,6 SiaT
α-2,3/2,6 SiaT
α-2,3/2,6 SiaT
α-2,3/2,6 SiaT
ST6GalNAc1
ST3Gal1
ST6GalNAc3/4
R2
R2 =
ST3Gal3/4
Trans
Secretory pathway

Figure 1-2. Overview of major protein glycosylation processes in human. *N*-linked and *O*-linked protein glycosylation constitutes a major portion of glycoproteins in human. Both *N*- and *O*-glycoprotein biogenesis happen in the secretory pathway which includes endoplasmic reticulum (ER) and Golgi body. In the *N*-linked glycosylation pathway, several GTs contribute to the assembly of the high-mannose glycan on the dolichol pyrophosphate lipid carrier. In ER, some GTs are Leloir GT and utilize nucleotide active sugar as substrate while others require lipid-liked monosaccharide as sugar donor. The preassembled tetradecasaccharides is transferred onto nascent polypeptide arisen from ribosome complex associated with the ER. This glycan is then further processed to generate handle for the quality control of the protein folding through calnexin and calreticulin (CNX/CRT) cycle. *N*-glycoproteins that pass this quality control will then trafficked to the cis-Golgi via COPII transport vesicle. In the Golgi body, hundreds of GTs work in concert to process *N*-glycans. This process including glycan trimming and core extension that happen at both cis- and medial-Golgi. Finally, many *N*-glycans are modified at their non-reducing end with the sialic acid cap, a process that happens at the late or trans-Golgi. Together, these *N*-glycans editing steps generate hybrid and complex-type *N*-glycans with almost 100 possible structures. For mucin-type *O*-glycosylation pathway, the process begins in the early Golgi by the installation of single GalNAc residue on polypeptide. Mucin molecules carrying GalNAc are then elaborated, either with Gal or GlcNAc residue to generate common core *O*-glycans. These core glycans are then transverse through medial Golgi where their structures are further extended. Notably, some O-glycans are used as a primer to generate glycopolymer including poly-LacNAc, an important glycan structure for host-pathogen interaction. *O*-glycan on mucin can also be installed with sialic acid cap at the late Golgi. Fully folded protein carrying mature *N*- and/or *O*-linked glycans will exit Golgi body and enter the next stage of secretory pathway that will deliver glycoprotein to the cell surface. While not presence in this Figure, glycan on glycoprotein can be further modified with sulfate and phosphate moieties. Glycan structures appear in this schematic are the representative of some common structures. Text in section 1.2.1 provides further detail on these protein glycosylation steps. Symbol nomenclature of glycan can be found in the Figure 1-3.

Protein glycosylation happens within the secretory pathway, nucleus, cytoplasm, and mitochondria of all eukaryotic cells. Complex carbohydrates in eukaryotic cells typically are derived from 10 monosaccharide building blocks[3] including glucose (Glc), galactose (Galactose), *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), fucose (Fuc), mannose (Man), xylose (Xyl), glucuronic acid (GlcA), neuraminic acid (NeuAc), and ribose (Rib), whose structures, exception for ribose are shown in **Figure 1-3**. These monosaccharides are used from dolichol lipid carrier or nucleotide sugar donor. Glycan can be attached: (i) to the nitrogen amine of the asparagine amino acid to form *N*-linked protein glycosylation; (ii) to the oxygen atom of the hydroxy group within serine or threonine amino acids to form *O*-linked protein glycosylation; (iii) to the carbon atom within indole ring of the tryptophan amino acid to form C-linked protein glycosylation; and (iv) between phospholipid and protein to form glypidated biomolecules. Many protein glycosylation processes are initiated within the ER. The initiation of some *O*-linked glycosylation including mucin-type O-glycosylation and Xylose-O happens in early Golgi while GlcNAc-O is typically form in the nucleus and cytoplasm (Reily et al., 2019). After initiation step, many glycans undergo extensive glycan trimming, elaboration, capping, sulfation, and phosphorylation. These processes are orchestrated by several glycosyltransferases and

3 Section 1.2 stated that, in mammals, glycoproteins are assembled from 14 basic monosaccharides. These 14 monosaccharides arise from 10 basic units with the additional 4 represents basic units with different anomers and ring patterns.

glycan processing enzymes and happen in mid- and late Golgi to yield fully matured carbohydrate structures (**Figure 1-2**).

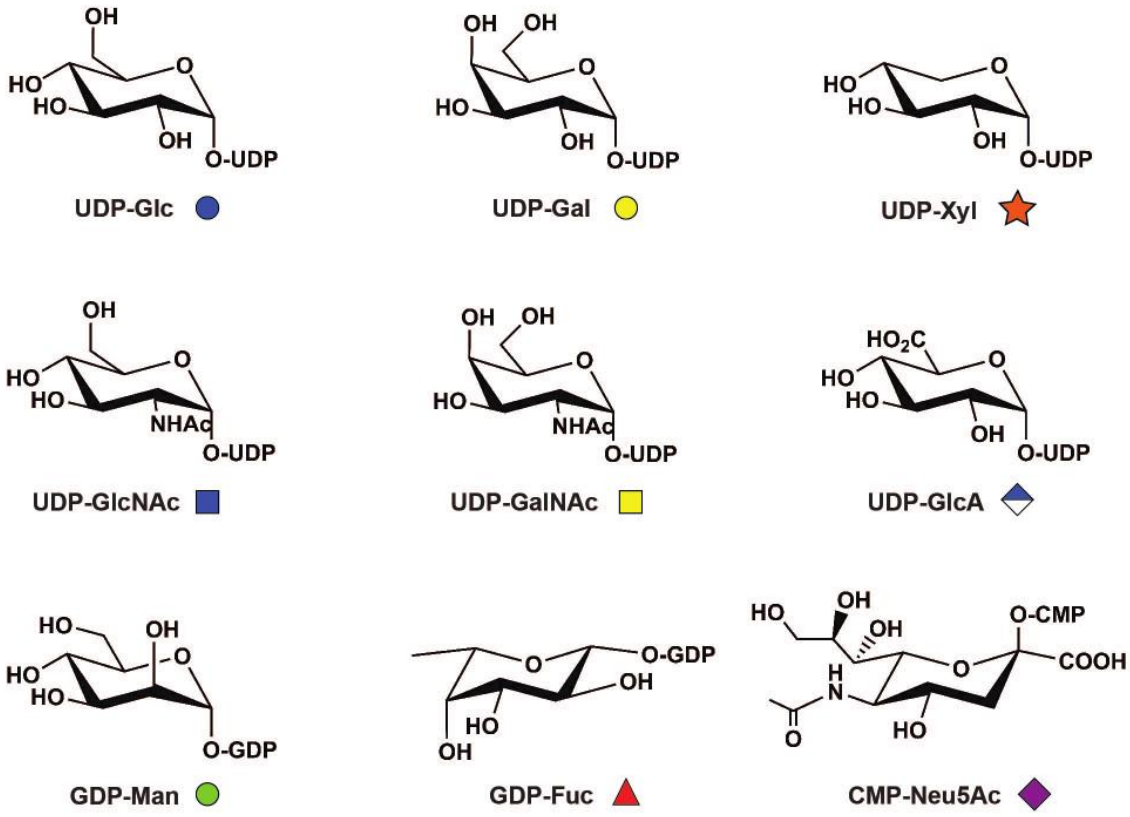


Figure 1-3. Common nucleotide sugar structures used in producing glycan and glycoconjugate in mammalian cells. Shorthand notation along with the symbol nomenclature are provided under the chemical structure. Structures of nucleotide bases and phosphate groups are omitted for clarity. Note that the structure of ribose which is found prevalently as part of the nucleotide is not shown as this sugar is typically not part of the glycoproteins and glycolipids in mammalian cells. Abbreviations: ATP, adenosine triphosphate; ADP, adenosine diphosphate; NTP, nucleoside triphosphate; PPi, inorganic pyrophosphate; and CTP, cytidine triphosphate.

Eukaryotic N-linked protein glycosylation

Asparagine-linked protein glycosylation in higher organism is extremely complicated. In human, this process begins by the assemble of tetradecasaccharides ($Glc_3Man_9GlcNAc_2$-) onto dolichyl-pyrophosphate carrier in the process called dolichol pathway (Burda and Aebi, 1999) (**Figure 1-2**). Dolichols are α-saturated lipid molecules consisting of 14-17 isoprene units, synthesized in the ER from farnesyl-pyrophosphate (Schenk et al., 2001). Dolichol phosphate (Dol-P) is a substrate of Alg7, an *N*-acetylglucosamine-phosphate transferase that modifies Dol-P to Dol-PP-GlcNAc (Kukuruzinska and Robbins, 1987). Hence the first step in the *N*-linked oligosaccharides biosynthesis. Subsequently, a series of GTs namely Alg13/Alg14, Alg1, Alg2, and Alg11 sequentially elaborates Dol-PP-GlcNAc to Dol-PP-$GlcNAc_2Man_5$ (Gao et al., 2004; Gao et al., 2005). Dolichyl pyrophosphate containing heptasaccharides is then flipped by Rft1p flippase to the lumen of the ER membrane (Helenius et al., 2002). In the ER lumen, mannosyltransferases (Alg3, Alg9, and Alg12) and glucosyltransferases (Alg6, Alg8, and Alg10) then complete Dol-PP-$Glc_3Man_9GlcNAc_2$ biosynthesis using both Dol-P-Man and Dol-P-Glc substrates (Burda and Aebi, 1999).

Pre-assembled Dol-PP-$Glc_3Man_9GlcNAc_2$ is utilized by the oligosaccharyltransferase (OST) complex containing STT3A and STT3B catalytic subunits along with several associated subunits and protein adaptor. Recent study demonstrates distinct role of STT3A and STT3B in catalyzing co-translational and post-

translational protein *N*-glycosylation, respectively (Ramirez et al., 2019). OST complex catalyzes a transfer of $Glc_3Man_9GlcNAc_2$ "core" glycan onto protein at the consensus sequon N-X-S/T, where X is any amino acid except proline. A subset of newly synthesize polypeptide chain contains signal peptide recognized by the signal recognition particle or SRP (Halic and Beckmann, 2005). The SRP will direct actively-translated polypeptide to the translocon machinery which will then transports polypeptide across ER membrane where *N*-linked glycosylation takes place (Chavan et al., 2005). It is worth noting that while the presence of glycosylation sequon or primary protein structure predominantly determines glycosylation site occupancy within a given protein, local secondary protein structure which dictate accessibility to the glycosylation sequon also plays an important role in both glycosylation site occupancy and downstream glycan processing (Chen et al., 2005; Thaysen-Andersen and Packer, 2012).

Following OST-catalyzed glycosylation, two terminal Glc residues are cleaved by glucosidase enzymes. This glucoses trimming is required for protein to enter quality control cycle facilitated by chaperone-like lectins calnexin and calreticulin. These lectins work in concert with Erp57 co-chaperone with thiol-oxidoreductase activity to ensure proper fold of the protein (Helenius and Aebi, 2001a). Eventually, folded protein that has passed this quality checkpoint will be translocated to the Golgi after its *N*-glycan is remodeled to $Man_8GlcNAc_2$ structure. In contrast, misfolded protein will be recognized by mannosidase-like lectin Mnl1p and subjected to the ER-

associated degradation (ERAD) pathway (Meusser et al., 2005). *N*-glycan processing within Golgi involves several glycosidases and glycosyltransferases that trim and elaborate the glycan. In human, *N*-glycan processing results in the addition of diverse monosaccharide units including GlcNAc, Gal, Neu5Ac, and GalNAc to the glycoproteins to produce branched and complex glycan structures (Schjoldager et al., 2020). The *N*-glycan process is highly dynamic and largely controlled by the spatiotemporal arrangement of the glycan processing enzyme at the Golgi membrane as well as the availability of sugar substrates (Reily et al., 2019).

Eukaryotic O-linked protein glycosylation

Eukaryotic *O*-linked protein glycosylation typically proceeds in a processive manner, which begins by the attachment of a single monosaccharide to the serine or threonine of the acceptor proteins. Other monosaccharides are then elaborated from the first sugar one at a time until the *O*-glycan reaches mature structure. The first monosaccharide in this process can be GalNAc, GlcNAc, Man, Glc, Fuc, Ara, Xyl, or Gal (Corfield, 2017). Several *O*-linked protein glycosylation occurs within the lumen of the Golgi with the exceptions including *O*-linked mannosylation, which initiates in the ER (Lommel and Strahl, 2009), and *O*-linked GlcNAcylation, which happens within cytosol and nucleus (Yang and Qian, 2017). Unlike *N*-glycosylation, there is no consensus glycosylation sequon for protein *O*-glycosylation. It has been hypothesized that *O*-glycosite occupancy depends on factors including local concentration of the

GTs, local protein secondary structures, and the availability of sugar substrates (Van den Steen et al., 1998). In recent years, computational biology and machine learning approach have been utilized to better predict *O*-glycosylation sites in the glycome of several organism (Cai et al., 2011; Chen et al., 2015; Hassan et al., 2015).

Mucin type *O*-glycosylation is one of the most common type of protein *O*-glycosylation in mammals. Mucins is a family of large glycoproteins found on the apical surfaces of many epithelia and some hematopoietic cells (Gendler, 2001). Mucins are heavily decorated with *O*-glycans attached on serine/threonine (tyrosine) amino acid residues, and dense patterns of *O*-glycans form a physical network resembling a polymeric meshwork on the cell surface. Traditionally, this mucinous layer has been considered to be a mere physical barrier that protects cells from mechanical disturbance or against pathogenic invasions (Gendler, 2001). Emerging evidence, however, demonstrates that mucin and mucin-like proteins through their *O*-glycans serve as molecular signals that inform critical cellular and multicellular behaviors, and dysregulation of glycan biosynthesis on mucins leads to a signal redistribution that is directly linked to disease promotion (Beatson et al., 2016; Bhatia et al., 2019). *O*-glycan on mucin is synthesized by polypeptide *N*-acetylgalactosaminyltransferase (ppGalNAcT), a family of 20 isoenzymes with distinct and partially overlapping mucin substrate specificity (Bennett et al., 2012; de las Rivas et al., 2019). ppGalNAcT modifies mucin with simple GalNAc residue to generate GalNAcα1–O-Ser/Thr, a structure known as the cancer-associated Tn antigen (Bhatia et al., 2019). This first GalNAc

residue is then elaborated through core extension, branching, and capping steps within *cis-*, *medial-* and *trans*-Golgi to generate diverse mucin *O*-glycan structures including glycopolymer such as polyLacNAc containing fucose and sialic acid caps (**Figure 1-2**). In the past decades, *O*-glycomucins have received considerable attention due to a discovery that many cancers such as pancreas, breast, lung, and colon carcinoma overexpress mucin with aberrant, truncated O-glycans including Tn (GalNAc-α-O-Ser/Thr), STn (Neu5Acα2-6-Tn), and T (Galβ1-3-Tn) antigens on the cell surface (Beatson et al., 2016; Bhatia et al., 2019; Kasprzak and Adamek, 2019; Venturi et al., 2019). In addition, it has been shown that overexpression of mucin with aberrant O-glycans synergistically confer oncogenic features to the malignant cells including proliferation, immunosuppression, and metastasis (Kufe, 2009). Importantly, in the extracellular matrix, Tn-, STn-, and T-mucin are found exclusively on cancerous tissues (Kufe, 2009; Taylor-Papadimitriou et al., 1999), which lead to the hypothesis that aberrant O-glycosylated mucin can serve as a safe targets for cancer immunotherapy (Steentoft et al., 2018). In fact, one particular monoclonal antibody (mAb) named 5E5, that binds specifically to Tn-MUC1 epitope with high affinity (Sorensen et al., 2006), has recently been reformatted into a chimeric antigen receptor T-cells (CAR-Ts), and the pre-clinal study has demonstrated the ability of the 5E5 CAR-T to potently eradicate tumor xenografts in mouse model (Posey et al., 2016).

1.2.2 Prokaryote protein glycosylation pathway.

While progress in study protein glycosylation in eukaryote has been made since the work of Neuberger in 1930s who studied glycopeptide from ovalbumin (Neuberger, 1938), it was long thought that protein glycosylation does not exist in prokaryotes (Abu-Qarn et al., 2008). In 1970s, S-layer surface glycoproteins were first described in the halophile *Halobacterium salinarum* (Mescher and Strominger, 1976) and in the gram-positive clostridia (Sleytr and Thorne, 1976). While these S-layer glycoproteins represent the first example of prokaryotic glycoproteins, their glycan structures and linkages are markedly distinct than the eukaryotic counterparts. Not until 1990s to 2000s when cell-surface *O*-glycoprotein appendages such as pili and flagella (Power et al., 2000) and general protein *N*-glycosylation pathway were discovered in bacterium (Szymanski and Wren, 2005) that the scientific community fully accepted the existence of prokaryotic protein glycosylation system. Interestingly, shortly after the first report of prokaryote protein *N*-glycosylation system, several other protein glycosylation types were described. To date, protein glycosylation in prokaryote covers both *N*- and *O*-linked type as well as both *en bloc* and possessive mechanisms (**Figure 1-4**).

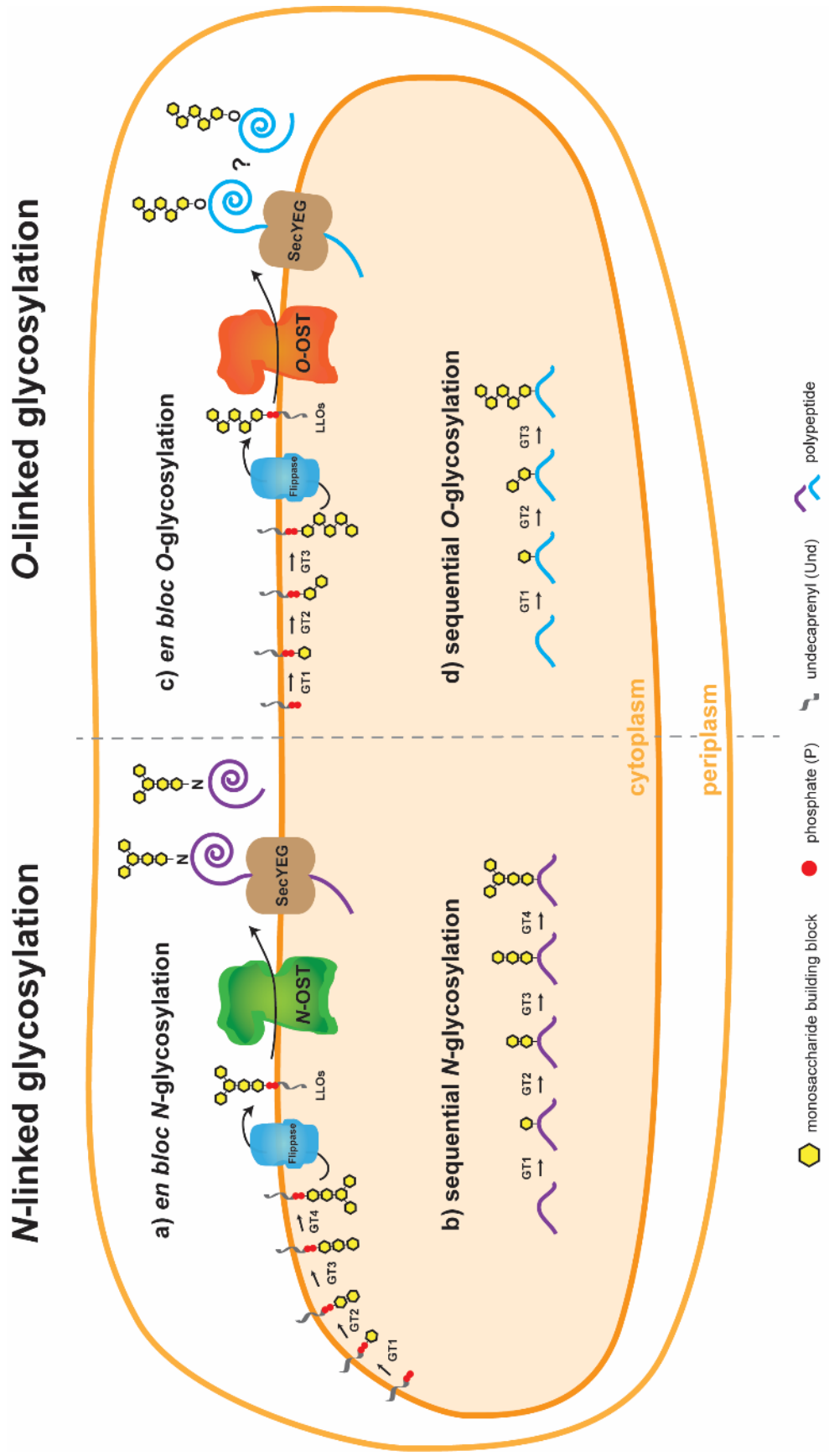

N-linked glycosylation
O-linked glycosylation
a) en bloc N-glycosylation
b) sequential N-glycosylation
c) en bloc O-glycosylation
d) sequential O-glycosylation
SecYEG
N-OST
O-OST
Flippase
LLOs
GT1
GT2
GT3
GT4
cytoplasm
periplasm
monosaccharide building block
phosphate (P)
undecaprenyl (Und)
polypeptide

Figure 1-4

Figure 1.4 Protein glycosylation processes in prokaryotes. For the biosynthesis of *N*-linked glycoproteins, **(a)** Bacteria including *Campylobacter, Helicobacter,* and *Desulfovibrio* species utilize *en bloc* transfer mechanism where *N*-glycan is preassembled on lipid-carrier – undecaprenyl pyrophosphate (Und-PP) in the cytoplasm of the bacteria. This lipid-linked glycan is then flipped into periplasm. Single subunit OST will then transfer glycan from Und-PP carrier onto the asparagine residue within the glycosylation sequon, which is generally more stringent than eukaryotic counterpart. Bacterial OST can perform both co- and post-translational modifications of target protein given that a glycosylation sequon is presence and accessible by the OST. **(b)** Bacteria including *H. influenzae* and *A. pleuropneumoniae* possess atypical[4] glycosylation mechanism where glycan is sequentially built directly on the acceptor protein within the cytoplasm. The glycosylation sequon for the cytoplasmic *N*-glycosylation is similar to the eukaryotic sequon - Asn-X-Ser/Thr where X is any amino acids but proline. For producing *O*-linked glycoproteins, **(c)** diverse gram-negative bacteria including *Neisseria, Burkholderia,* and *Acinetobacter* species contain one or more putative O-OSTs which follows *en bloc* glycan transfer mechanism. Their key enzymes, *O*-OST, transfer glycan from Und-PP to a variety of protein including a large family of pillin protein. Interestingly, many *O*-OSTs have extreme relaxed glycan donor specificity and can utilize a variety of glycan and lipid structures as substrate. **(d)** Sequential O-glycosylation in bacteria is diverse. A family of bacterial autotransporter heptosyltransferase (BAHT) is a known cytoplasmic O-glycosyltransferase enzyme that are found in several gram-negative bacteria including *E. coli*. These O-GTs modify several protein domains of the type V secretion pathway, and this modification is hypothesized to be important for the virulence of several pathogens including enterotoxigenic (ETEC) *E. coli*. Several gram-positive bacteria such as *Streptococci* and *Staphylococci* species express serine-rich repeat proteins (SRRP) that are heavily glycosylated with *O*-glycans. This modification contributes significantly to the adhesion efficiency to host cells. SRRP glycosylation begins with the formation of GtfA-GtfB complex that initiates a transfer of GlcNAc residue to the SRRP. This GlcNAc moiety provides a primer for further glycan elaboration by several other GTs. Analogously, the SRRP glycosylation is similar to the mucin *O*-glycosylation in mammals. [This figure appeared in Natarajan, A., Jaroentomeechai, T., Li, M., Glasscock, C.J. and DeLisa, M.P. (2018) Metabolic engineering of glycoprotein synthesis in bacteria. Emerg Top Life Sci 2: 419-432.]

[4] We called this 'atypical' since processive *N*-glycosylation mechanism is not found in higher organism.

Prokaryotic N-linked protein glycosylation

The first *bona fide N*-linked protein glycosylation in gram negative bacteria is discovered in *Campylobacter jejuni* in 1999 (Szymanski et al., 1999). This gastroenteritis pathogen is found to contain a genetic locus responsible for the biosynthesis of glycoproteins including PEB3 and CgpA. These proteins are highly immunogenic and bind strongly to soybean agglutinin lectin (Linton et al., 2002). Glycan on PEB3 was not removed by β-elimination reaction which suggested glycan was attached to PEB3 via *N*-glycosylation linkage (Linton et al., 2002). Subsequent studies using mass spectrometry (MS) and nuclear magnetic resonance (NMR) on glycopeptide elucidated glycan structures to be heptasaccharides containing GalNAc-α1,4-GalNAc-α1,4-[Glc-β-1,3]-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-diNAcBac where diNAcBac is 2,4-diacetamido-2,4,6-trideoxy-D-glucopyranose (Linton et al., 2002; Szymanski et al., 2003; Young et al., 2002) (See **Figure 1-5** for the structure of this glycan). Importantly, this linear poly-GalNAc glycan is conserved among several *Campylobacter* species (Nothaft and Szymanski, 2010) and is structurally distinct from eukaryotic *N*-glycans, which contain branching structures. Altogether, these breakthrough studies helped establish the existence of general but unique *N*-glycosylation pathway in bacteria.

The aforementioned genetic locus in *C. jejuni* was later named protein glycosylation locus (*pgl*). Using bioinformatic analysis, it was hypothesized that the *pgl* contains five putative glycosyltransferases which are named *pglA, pglC, pglH, pglI,* and *pglJ*. *pgl* also encodes *pglD, pglE,* and *pglF* genes that involve in monosaccharide

biosynthesis. The *gne* gene encodes a bifunctional UDP-Glc/GlcNAc 4 epimerase that can convert UDP-Glc/Gal to UDP-GlcNAc/GalNAc (Bernatchez et al., 2005). Another gene in the *pgl* named *pglK* contains sequence of the flippase enzyme, an ATP-binding cassette (ABC) transporter protein that translocates glycan from the cytoplasmic to periplasmic space. Finally, *pglB* gene encodes protein with high homology to the Stt3p subunit within the yeast OST complex, and hence PglB is hypothesized to be a single subunit oligosaccharyltransferase (Linton et al., 2002) (**Figure 1-4a**).

Shortly after its discovery, the *pgl* was functionally expressed in *E. coli* along with the *C. jejuni* periplasmic proteins AcrA and PEB3, which are natively *N*-glycosylated. The AcrA and PEB3 were successfully produced and glycosylated in the engineered *E. coli* system (Wacker et al., 2002). This provides an evidence that *pgl* itself is sufficient to carry protein *N*-glycosylation. More importantly, tandem mass spectrometry analysis of the glycopeptides revealed *C. jejuni* glycan was attached to the asparagine amino acid of the DFNVS and DFNR peptides in the PEB3 and AcrA, respectively. These peptide sequences are similar to the *N*-glycosylation consensus sequon in the eukaryote and lead to an important postulation that the mechanism of the prokaryotic *N*-glycosylation is similar to that of eukaryote. Specifically, the first step in this pathway involves glycan assembly on a lipid carrier by the activity of GTs (PglA, PglC, PglH, PglI) and monosaccharide synthases (PglD, PglE, PglF). This glycan assembly step is analogous to the Dol-PP-$Glc_3Man_9GlcNAc_2$ assembly within ER of the eukaryote. Full-length lipid-linked oligosaccharides (LLOs) is then flipped by PglK

into the periplasmic space, similar to how LLOs is flipped from cytoplasmic space to ER lumen. Finally, the central enzyme in this pathway, PglB, then transfers a preassembled glycan onto acceptor protein containing glycosylation sequon (Wacker et al., 2002). In retrospect, this is quite a remarkable assumption as this hypothesis has proven to be almost entirely correct by a series of study over the next decades.

The glycoengineered *E. coli* system described by Wacker *et al* has offered a facile framework for interrogating the *pgl* pathway in model organism. Due to the lack in native protein glycosylation pathway, the viability of *E. coli* is not affected by the alteration in the glycosylation system. In stark contrast, protein glycosylation dictates several biological processes including cell growth and development in the eukaryotes. Hence manipulation of the native glycosylation pathways can severely compromise cell viability (Barnard et al., 2004; Jiang et al., 2015). Moreover, the absence of native glycosylation system renders *E. coli* a "clean" chassis in which orthogonal glycosylation pathways can be introduced without interference from endogenous glycosylation components. These advantages make glycoengineered *E. coli* an attractive platform for interrogating glycosylation pathway of interest as well as for producing a more uniformly glycosylated protein products (Baker et al., 2013a; Keys and Aebi, 2017; Merritt et al., 2013).

Following seminal works by Szymanski and Aebi groups, detailed characterizations of various genes within the *pgl* pathway were reported. As mentioned earlier, PglD, PglE, and PglF are monosaccharide synthase and the *in vitro*

study had confirmed their coordinate role in the biosynthesis of diNAcBac sugar (Glover et al., 2005a; Rangarajan et al., 2008). Through mutation studies in glycoengineered *E. coli*, specific roles of the PglC, PglA, PglJ, and PglI GTs in the *C. jejuni* heptasaccharides biosynthesis were established (Linton et al., 2005) (**Figure 1-5**). Interestingly, the exact mechanism of how PglH adds three GalNAc monosaccharide to the growing glycan chain was elucidated using recently using biochemical and cryo-EM studies (Ramirez et al., 2018). Subsequent works using *in vitro* reconstitution of these GTs further provided insight how these enzymes work at the membrane interface to build oligosaccharide on the lipid carrier (Glover et al., 2005a; Weerapana et al., 2005).

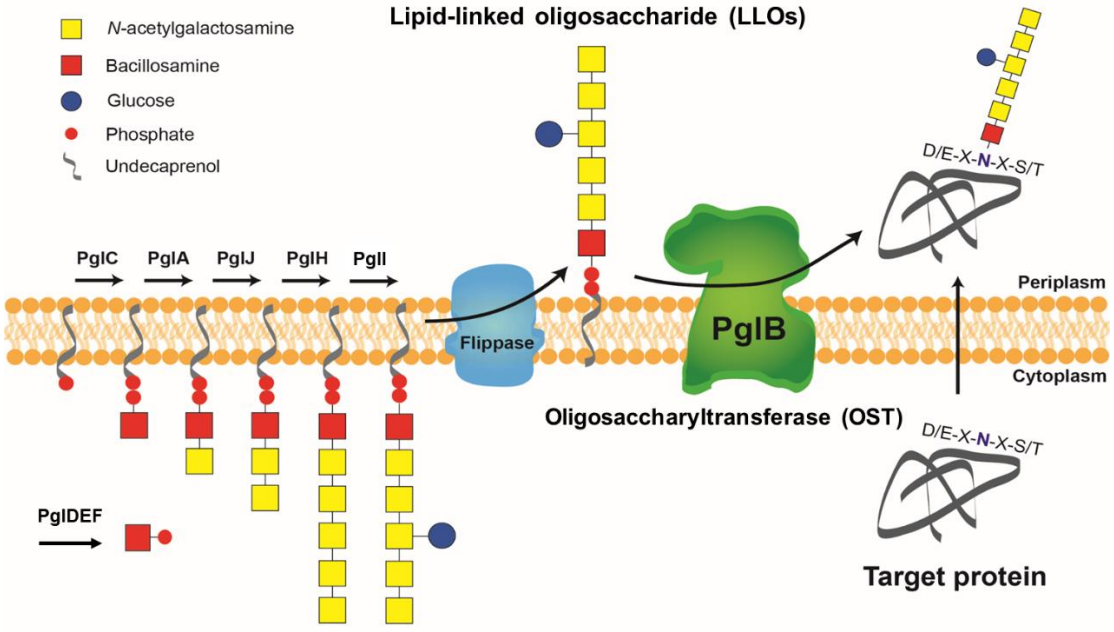


Figure 1-5. *N*-linked protein glycosylation pathway in *Campylobacter jejuni*. Within the cytoplasmic space, PglE and PglF modify UDP-GlcNAc nucleotide sugar to form UDP-2-acetamido-4-amino-2,4,6-trideoxy-α-D-glycopyranose (Schoenhofen et al., 2006). PglD which is identified as acetyltransferase then convert the UDP-4-amino-sugar to UDP-N,N′-

diacetylbacillosamine (bacillosamine). PglC transfers bacillosamine phosphate onto undecaprenol monophosphate to generate an initial lipid-linked oligosaccharide (LLOs) structure. Successive reactions catalyzed by PglA, PglJ, and PglH extend this initial structure with a linear chain of 5 GalNAc monosaccharides. Then PglI adds a branching glucose to complete the biosynthesis of the *C. jejuni* heptasaccharide glycan. LLOs is a substrate of flippase, an ABC transporter, that transport LLOs across inner membrane into the periplasm. PglB, a single-subunit oligosaccharyltransferase, will form a protein complex comprised of lipid-linked heptasaccharide and acceptor peptide with D/E-X_{-1}-N-X_{+1}-S/T (where X ≠ Pro) glycosylation sequon. The spatial arrangement within this complex facilitates a transfer of heptasaccharide glycan from LLOs to the asparagine of the acceptor peptide. Accumulated evidence has shown that PglB can modify acceptor proteins both co- and post-translationally, although the accessibility of the sequon post-protein folding is required. Finally, in the absence of acceptor peptide, PglB has been shown to spontaneously catalyze glycan release from LLOs to generate free oligosaccharides (FOS).

Within the *pgl* pathway, perhaps the existence of *pglB* gene and its corresponding PglB enzyme is the most important finding. PglB shares significant homology with the Stt3 enzyme within the OST complex of the yeast (Yan and Lennarz, 2002) and its ability to transfer glycan from polyprenol-phosphate lipid carrier has been demonstrated both *in vivo* and *in vitro* (Kowarik et al., 2006a; Wacker et al., 2002). PglB also contains WWDYG amino acids, a conserved motif present in all OST enzymes and is required for the glycan transfer activity (Wacker et al., 2002). Importantly, *C. jejuni* PglB requires a more stringent glycosylation sequon and the presence of the acidic amino acid such as aspartic or glutamic acid at the second position upstream from the

asparagine5 is necessary for its substrate recognition (Kowarik et al., 2006b; Wacker et al., 2006). This minus-two rule has been now well-understood, owning to the availability of several bacterial OST structures in complex with their peptide substrates (Lizak et al., 2011b; Lizak et al., 2014; Napiorkowska et al., 2017). However, a more relaxed substrate specificity can be observed in the OSTs from other *Campylobacter* and bacterial species (Ollis et al., 2015c). Prior to a discovery of its bacterial homologs, oligosaccharyltransferase enzyme was found only in eukaryote and archaea and there were several questions concerning the origin and evolution of the *N*-linked protein glycosylation pathway (Gagneux and Varki, 1999; Lombard, 2016). With the discovery of putative bacterial OST, it has now become clear that polyprenol-phosphate glycosylation pathway exists in all domains of life and the eukaryotic *N*-glycosylation pathway may contain enzymes from both bacterial and proteoarchaeal origins (Lombard, 2016).

The polyprenol-phosphate *N*-glycosylation pathway is generally called an *en bloc* glycosylation as it involves a transfer of fully assembled oligosaccharides onto target protein (**Figure 1-4a**). While the *en bloc* glycosylation is a main subject of study for this dissertation, it is important to mention that a few gram-negative bacteria including *Haemophilus influenzae* and *Actinobacillus pleuropneumoniae* contain another unique *N*-glycosylation pathway where monosaccharide (Glc or Gal) is added sequentially and directly onto protein (Choi et al., 2010; Gross et al., 2008). This

5 Hence minus-two rule.

sequential glycosylation system occurs in the cytoplasmic space, a distinction from the OST-mediated glycosylation which happens in the periplasm (**Figure 1-4b**). The key catalyst in this pathway, *N*-glycosyltransferase or NGT, has a similar substrate specificity as eukaryote N-glycosylation sequon (N-X-S/T), although a recent evidence suggests that the neighboring amino acid up to +/- 6 position from the asparagine might play a role in substrate recognition of the NGT (Kightlinger et al., 2018). In recent years, this sequential *N*-glycosylation pathway has been functionally characterized and transferred into *E. coli*, providing a unique glycoengineered system for producing novel glycans and glycoconjugates (Keys et al., 2017; Tytgat et al., 2019).

Prokaryotic O-linked protein glycosylation

O-linked glycosylation in bacteria is diverse and encompasses both the *en bloc* and processive mechanisms. Periplasmic *O*-oligosaccharyltransferase (OST)-mediated protein glycosylation pathway was first described in *Neisseria meningitidis* and *Pseudomonas aeruginosa* (Faridmoayer et al., 2007). The key enzyme *O*-OST in this pathway was found to operate in an *en bloc* transfer mechanism and could transfer preassembled glycan from the undecaprenyl-PP carrier onto an acceptor protein pillin (**Figure 1-4c**). Interestingly, *O*-OST from *N. meningitidis* displays an extreme glycan substrate promiscuity and can recognize a range of structurally-different oligosaccharides, including mono-, di-, and polysaccharide, as well as diverse lipid carrier (Faridmoayer et al., 2008), suggesting its potential use in glycoengineering

applications. Besides *N. meningitidis*, a similar *en bloc O*-glycosylation pathway has also been described in *Neisseria gonorrhoeae* (Hartley et al., 2011), *Burkholderia cepacia* K56-2 (Lithgow et al., 2014), *Vibrio cholerae*, and *Burkholderia thailandensis* (Gebhart et al., 2012).

Cytoplasmic *O*-glycosylation pathways are prevalent in both gram positive and gram negative bacteria (Schaffer and Messner, 2017). Common bacterial components including flagellins, pilins, fimbriae, and adhesins have been found to contain carbohydrate moieties. Many studies have elucidated that these glycans are installed on proteins in the sequential process within the cytoplasm of the bacteria (**Figure 1-4d**). For example, a complete biosynthesis pathway for a processive *O*-glycosylation in *Streptococcus parasanguinis* has been characterized and is comprised of GtfA and GtfB complex that initiates GlcNAc-serine linkage within Fap1, a serine-rich repeat glycoprotein (SRRPs) (Zhu et al., 2016). Owning to a longer history since its discovery, excellent reviews on the subject have been published (Schaffer and Messner, 2017; Tan et al., 2015).

1.3 Synthetic glycobiology.

The term 'synthetic glycobiology' was first used to describe the redesign of GT assembly lines for the production of specific glycan structures using protein engineering and chemical approaches (Czlapinski and Bertozzi, 2006). This initial definition referred narrowly to the exploitation of Golgi-resident GTs to engineer protein glycosylation inside and on the surface of eukaryotic cells, as exemplified by a

number of notable glycoengineering studies in yeast (Choi et al., 2003; Hamilton et al., 2003) and more recently in mammalian cells (Chang et al., 2019; Meuris et al., 2014). These successes notwithstanding, simpler, cell-viability independent systems that permit bottom-up assembly of prescribed glycosylation pathways and controllable biosynthesis of designer glycomolecules are of great scientific and technological interest, and have the potential to be transformative. In this vein, Aebi and coworkers pioneered the first bacterial glycoprotein expression platform by transferring the *N*-linked glycosylation machinery from *Campylobacter jejuni* into laboratory strains of *Escherichia coli*, giving the latter the ability to transfer glycans site-specifically onto acceptor proteins (Wacker et al., 2002). Following this seminal work, numerous additional heterologous glycosylation systems have been functionally reconstituted in *E. coli* (Feldman et al., 2005a; Hug et al., 2011; Ihssen et al., 2010; Keys et al., 2017; Schwarz et al., 2011a; Shang et al., 2016; Tytgat et al., 2019; Valderrama-Rincon et al., 2012), giving this simple organism the ability to produce a diverse array of complex glycomolecules. Hence, a more current definition of synthetic glycobiology is the purposeful alteration or rational construction of *any* glycosylation system using chemical and molecular biological approaches in conjunction with metabolic pathway engineering tools. Such synthetic systems have been instrumental in increasing our understanding of glycosylation networks and producing desired glycans and glycoconjugates.

1.3.1 Synthetic glycobiology goes cell-free.

While the majority of synthetic glycobiology efforts to date have involved living organisms, recent years have seen the emergence of cell-free systems as a new platform for synthetic glycobiologists to investigate and manipulate glycosylation outside of cells, leading to the birth of an entirely new field that we call cell-free synthetic glycobiology. Although still in its infancy, cell-free synthetic glycobiology has already helped to uncover the underlying mechanisms governing a variety of glycosylation reactions and enabled preparation of structurally-defined glycomolecules. The origins of this new field can be traced back almost 60 years ago when cell-free biology was used to decipher the genetic code (Matthaei et al., 1962; Nirenberg and Matthaei, 1961). Since that time, cell-free biology has matured into a well-established field in biological research (Carlson et al., 2011; Dudley et al., 2015; Hammerling et al., 2020; Martin et al., 2018; Silverman et al., 2020), undergoing a technological renaissance in the early 21st century (Jewett et al., 2008a; Jewett and Swartz, 2004c; Shimizu et al., 2001) that has catalyzed significant improvements in batch reaction yields (Caschera and Noireaux, 2014; Des Soye et al., 2019), operational volumes (Yin et al., 2012; Zawada et al., 2011), standardization of protocols (Kim et al., 2019; Kwon and Jewett, 2015a; Silverman et al., 2019), availability of various active lysate systems (Des Soye et al., 2018; Perez et al., 2016b), and the ability to incorporate non-standard amino acids (Goerke and Swartz, 2009; Lee et al., 2019; Shimizu et al., 2005), and post-translational modifications (Jaroentomeechai et al., 2018b; Oza et al.,

2015b) into biomolecule products. As a result, cell-free systems are now widely used, sometimes in tandem with cell-based systems, to produce complex biomolecules (Matthies et al., 2011), to prototype and optimize metabolic pathways (Casini et al., 2018; Dudley et al., 2019; Karim et al., 2020; Karim and Jewett, 2016; Lim and Kim, 2019; O'Kane et al., 2019), for molecular sensing (Meyer et al., 2019; Pardee et al., 2014; Takahashi et al., 2018; Thavarajah et al., 2020), and to build and implement genetic networks (Swank et al., 2019; Takahashi et al., 2015). Owing to its open nature, cell-free reactions provide unprecedented flexibility to directly and precisely control compositions and conditions of a given system. Eliminating the biological membrane boundary also facilitates the integration of cell-free reactions with high-throughput screening tools (Su et al., 2016; Zhang et al., 2019), real-time monitoring, and automation (Georgi et al., 2016), resulting in significant reductions in design-build-test (DBT) timelines. Furthermore, the ability to harness cellular machineries without any impediments due to cell viability provides an opportunity to synthesize products and engineer biochemical pathways that otherwise exceed cellular toxicity tolerance (Kai et al., 2015; Thoring et al., 2017). Collectively, these versatile features are precisely what make cell-free platforms especially attractive for both mechanistic discovery and technological applications in glycoscience.

Currently, the utility of cell-free synthetic biology in glycoscience spans from a production of nucleotide-activated monosaccharide building blocks to large glycoproteins carrying complex polysaccharides. Characteristic features of these

approaches are the synthesis of glycosylation components and the assembly of these components into functional glycosylation pathways that produce glycosylated molecules of interest in a well-controlled environment without the use of intact, living cells (**Figure 1-6**). In its simplest form, cell-free synthetic biology involves using purified enzymes to catalyze specific glycosylation reactions in vitro (Natarajan, 2018; Yu and Chen, 2016). Alternatively, to circumvent the time- and labor-intensive process of enzyme purification, in situ production of glycosylation enzymes in cell-free lysates has been used to assemble multi-step glycosylation reactions (Kightlinger et al., 2019; Yu and Chen, 2016). These biosynthesis-focused methods will be discussed in detail below, along with the use of cell-free synthetic biology as a tool to characterize and evolve glycosylation enzymes and pathways for designer functions. Collectively, these greatly simplified platforms offer exquisite control over reaction fluxes and compositions, which in turn become powerful tools to understand the biotransformation of glycomolecules.

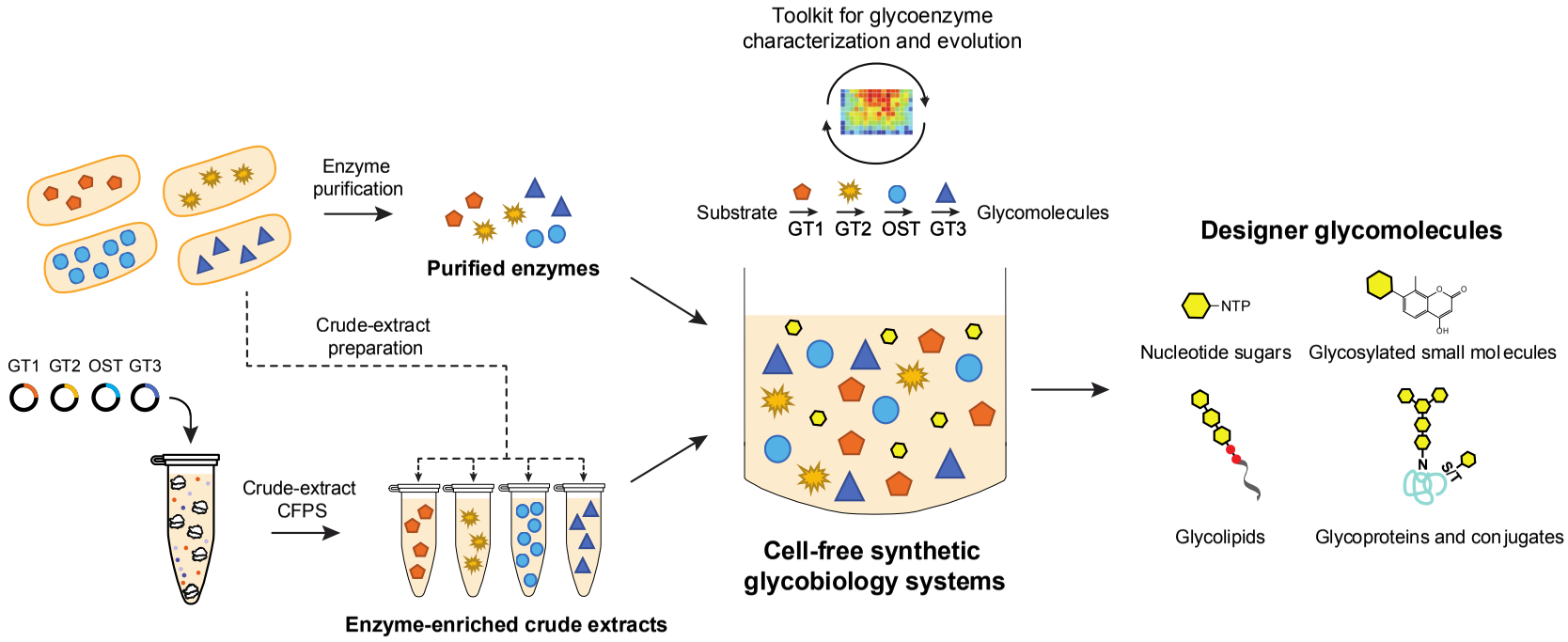

Toolkit for glycoenzyme characterization and evolution
Enzyme purification
Purified enzymes
Substrate
GT1 GT2 OST GT3
Glycomolecules
Designer glycomolecules
Crude-extract preparation
GT1 GT2 OST GT3
Crude-extract CFPS
Enzyme-enriched crude extracts
Cell-free synthetic glycobiology systems
NTP
Nucleotide sugars
Glycosylated small molecules
Glycolipids
Glycoproteins and conjugates

Figure 1-6

Figure 1-6: Cell-free synthetic glycobiology systems for on-demand biosynthesis of designer glycomolecules. Specific glycosylation reactions/pathways are assembled *in vitro* to synthesize designer glycomolecules including nucleotide sugar building blocks, small molecule glycosides, glycans, glycolipids, and glycoproteins. Glycoenzymes such as glycosyltransferases (GTs) and oligosaccharyltransferases (OSTs) can be prepared from living cells in a purified format or from cell-free protein synthesis (CFPS) expression platforms. Purified enzymes and/or enzyme-enriched crude lysates can then be combined in sequential reactions or in a single-pot to catalyze a prescribed glycosylation reaction. The open nature of cell-free systems allows each of the reaction components and their environment to be precisely controlled and monitored in real time, which in turn, facilitates the design-build-test cycle for an optimal output. Such open systems also provide great modularity as well as offer a unique possibility to integrate with high-throughput screening tools for constructing novel glycosylation pathways or evolving glycosylation enzymes.

1.3.2 Cell-free technologies for assembling glycans and glycolipids.

Complex carbohydrates or glycans in their unconjugated form are valuable reagents, finding use in both fundamental research and biomedical applications. An outstanding example is the use of structurally-defined free glycans to construct glycoarrays that enable high-throughput screening of molecular interactions between glycan epitopes and carbohydrate-binding entities including proteins and even whole organisms (Rillahan and Paulson, 2011). Since their first report (Blixt et al., 2004; Fukui et al., 2002), glycoarrays have proven to be tremendously useful for the discovery of antibodies, lectins, and immune receptors against carbohydrate antigens as well as for determining the substrate-specificity of various GTs (Blixt et al., 2008; Wen et al., 2018). To date, the Consortium for Functional Glycomics (CFG) and the Glycosciences

Laboratory at Imperial College have developed two of the largest glycoarray libraries, respectively consisting of approximately ~609 mammalian glycans (McQuillan et al., 2019) and ~796 neoglycolipid glycan structures (Li and Feizi, 2018; Palma et al., 2014). Another important application of free glycans is in the synthesis and development of conjugate vaccines, which are particularly effective against various bacterial pathogens (Moeller et al., 2018).

One of the major impediments to using free glycans as described above is accessibility of pure glycans at sufficient quantities. Initially, glycan libraries were obtained from natural sources such as microbes, plants, or animal products. In this process, glycans were separated from their bioconjugates through chemical or enzymatic hydrolysis followed by tedious, multistep purifications (Rillahan and Paulson, 2011). However, the high diversity of glycan structures present in natural samples makes it very difficult to acquire highly pure compounds using this approach. An alternative to harvesting glycans from natural sources is the use of chemical synthesis methods to generate free glycans from simple monosaccharide precursors. Chemical synthesis typically involves performing iterative rounds of glycosylation reactions utilizing a protecting group scheme that enables functionalization of a single hydroxyl group for sugar attachment. However, such *de novo* synthesis requires lengthy organic chemistry procedures, often necessitating highly specialized individuals and instrumentation. Some of these limitations have been alleviated by the introduction of automated solid-phase oligosaccharide synthesizers for the rapid

synthesis of glycans as described by Seeberger and coworkers (Plante et al., 2001). By adopting solid-phase synthesis, excess amounts of glycosyl donor can be used to drive reactions to completion and the removal of unwanted side products or reagents can be done in a single wash step. Since the time of its inception, the technology has now matured into a fully commercial system known as Glyconeer 2.1 (Hahm et al., 2017). These developments notwithstanding, the chemical synthesis of glycans remains a significant challenge due to the complexity in achieving stereo- and regio-selective synthesis. Selecting appropriate protective groups to achieve the desired glycosidic linkage remains one of the main hurdles and becomes more difficult as the complexity of the glycan architecture increases.

To circumvent the need for protecting group manipulation, the development of cell-free glycan synthesis systems that leverage enzymes such as GTs, GHs, and other glycan-processing enzymes is an attractive alternative. Enzymatic glycosylation permits precise stereo- and regio-controlled synthesis with high conversions using unprotected monosaccharides as substrates. Reactions generally proceed under mild, aqueous conditions without the need for toxic and harsh organic reagents. Using bio- and/or chemoenzymatic synthesis tools, several natural and engineered glycan libraries have recently been constructed including asymmetric multi-antennary *N*-glycans (Wang et al., 2013b), glycosphingolipid glycans (Yu et al., 2016), authentic human type *N*-glycans (Hamilton et al., 2017; Li et al., 2015), *O*-mannosyl glycans (Meng et al., 2018; Wang et al., 2018), human milk oligosaccharides (HMOs) (Prudden

et al., 2017; Xiao et al., 2016), and tumor-associated antigens ('t Hart et al., 2019; Li et al., 2019a). Similar strategies have been adopted for cell-free enzymatic synthesis of glycolipid libraries including those from bacterial (Glover et al., 2005a), animal (Stubs et al., 2010), and human origins (Li et al., 2019b). Many of these glycan and glycolipid libraries have been employed to construct glycan microarrays for profiling glycan-binding molecules such as lectins and antibodies as well as for gaining mechanistic insights into glycosylation reactions.

Cell-free enzymatic glycan synthesis can also be integrated with automated systems for more expeditious glycan assembly. To achieve this goal, several developments that simplify purification processes and increase conversion efficiencies have recently been reported. For example, the Linhardt group demonstrated the use of a fluorous tag to capture heparin sulfate products directly from solution (Cai et al., 2014). Additional advances include a photocleavable linker that enables chemoenzymatic synthesis of tumor-associated glycan epitopes (Bello et al., 2015) and an ion-exchange purification technique that aids in cell-free biosynthesis of HMOs (Zhu et al., 2017). These advances, along with many others, have been instrumental in realizing the goal of a fully automated enzymatic glycan synthesizer, several of which have now been reported or are in late stages of development. For example, Nishimura and coworkers developed an artificial "Golgi apparatus" to prepare sialyl Lewis X derivatives using a dendrimer-based solid support (Matsushita et al., 2010). Their process took four days and provided an overall yield of 16%. One of the main

challenges of the Golgi apparatus was that it required multiple filtration-purification steps that hindered its efficiency. To address these challenges, Wang and colleagues recently combined a thermosensitive polymer with a commercially available peptide synthesizer to mediate automated glycan assembly (Zhang et al., 2018). Their system was able to prepare several blood group antigens and ganglioside glycans with yields ranging from 27-38% within 1-2 days. Coincidentally, Boons and coworkers simultaneously developed a similar automated system using a set of water-soluble sulfate tags for a catch-and-release synthesis strategy (Li et al., 2019c). The sulfate tags were compatible with a range of glycosylation enzymes and, more importantly, were readily adapted to a custom-designed automated glycosynthesizer. Using this fully automated platform, quantitative amounts of complex glycans including gangliosides, HMOs, poly-*N*-acetyllactosamine (poly-LacNAc) derivatives, and *N*-glycans could be prepared in a less labor- and time-intensive process. Despite the small number of examples, the use of automated enzymatic glycan synthesis platforms shows significant promise, both as stand-alone systems and in combination with automated chemical synthesis (Fair et al., 2015). With rapid developments in automation, instrumentation, solid-support matrices, reliable tags and linkers, as well as a growing collection of accessible glycosylation enzymes, a fully mature and reliable enzymatic glycan synthesizer capable of synthesizing virtually any complex carbohydrate structure appears to be within reach.

1.3.3 Cellular glycoengineering to control glycoprotein biosynthesis.

More than 40 different types of carbohydrate-to-protein linkages have been identified to date. Among these, glycan installation at the asparagine (*N*-linked) and serine/threonine (*O*-linked) residues constitutes the greatest proportion of glycoproteins (Spiro, 2002). Protein glycosylation is highly dynamic and the glycan profile is controlled both spatially and temporally by the amino acid sequence, the local structural conformation of the glycosylation site, and the expression level of glycoenzymes at different stages of cellular development (Colley et al., 2015). Thus, glycoproteins are generally found in nature as a mixture of glycoforms sharing the same protein backbone but a variety of glycan structures. This intrinsic heterogeneity makes it challenging to decipher how specific glycoforms impact the structure and function of a modified protein. It has also been proven to be a major impediment for the development of glycoprotein-based therapeutics as the consistent ratio and identity of glycoforms are essential for reproducible clinical efficacy and safety of the biologic (Wang and Lomino, 2012). To address these challenges, a variety of glycoengineering approaches have been reported that involve the design and construction of molecular, cellular, and whole-organism systems with tunable glycosylation.

There is a long history of cellular glycoengineering in eukaryotes including in mammalian cells, plants, and yeasts (Bertozzi et al., 2009). Among these, the glycoengineering of Chinese hamster ovary (CHO) cultures has dominated the field as it is still the most commonly used host cell line in the biopharmaceutical industry

(Walsh, 2018). Many groups have explored glycosylation control using genetic manipulation to overexpress genes encoding glycoenzymes such as Golgi-resident GTs (Son et al., 2011; Weikert et al., 1999) (**Figure 1-7a**). Small molecule inhibitors targeting glycoenzymes such as kifunensine (Elbein et al., 1990) and swainsonine (Elbein et al., 1981) have also been successfully used to regulate a protein's glycoform in CHO culture (Ehret et al., 2019). More recently, systems biology and bioinformatics tools have been used to model glycosylation reaction networks in order to explore and quantify how perturbations to glycosylation parameters affect the cell (Neelamegham and Liu, 2011). Coupling this insight with precise genome editing tools will offer unprecedented freedom to glycoengineer organisms with greater control over glycoprotein products. In 2015, a landmark achievement in this regard was reported by the Clausen group whereby quantitative genomics data and precise genome editing was used to generate a panel of CHO cells with specific glycosyltransferase gene knock-outs (KO) (Yang et al., 2015). These glycoengineered CHO cells were used to screen and identify GT genes that play a major role in regulating protein *N*-glycosylation within the CHO cell glycome. Such knowledge, in turn, provided a blueprint for genetic reconstruction of CHO cells to generate variants with desirable homogeneous glycosylation capacities including those producing human-like α2,6-linked sialic acid capping glycoforms on therapeutic proteins such as human IgG and EPO (Caval et al., 2018; Schulz et al., 2018; Yang et al., 2015). Another notable example from the Weiss group explored the use of CRISPR/Cas9 to implement synthetic gene

circuits in CHO cells, allowing tunable *N*-glycan profiles of CHO culture-derived IgGs in a small molecule concentration-dependent manner (Chang et al., 2019). More recently, Narimatsu and coworkers at the Copenhagen Center for Glycomics further extended the utility of precise gene editing to create a library of validated CRISPR/Cas9 guide RNA targeting constructs for all human glycosyltransferase genes (Narimatsu et al., 2018). This gRNA library was subsequently applied to create a cell-based array displaying the human glycome in HEK293 cells (Narimatsu et al., 2019). A library of human cells displaying defined glycan structures will be a valuable resource for dissecting glycan biosynthesis and glycomolecule interaction within native physiological context (Narimatsu et al., 2019). It should be pointed out that advances in cell-based glycoengineering has extended beyond mammalian cells, with significant achievements toward homogeneous glycoforms production being reported in other eukaryotes including yeast (Hamilton et al., 2003; Wildt and Gerngross, 2005), microalgae (Barolo et al., 2020), insect cells (Toth et al., 2014), and plant cell cultures (Hurtado et al., 2020; Montero-Morales and Steinkellner, 2018). Comprehensive reviews of the glycoengineering approaches developed in these eukaryotic systems have been published elsewhere (Hamilton and Zha, 2015; Heffner et al., 2018).

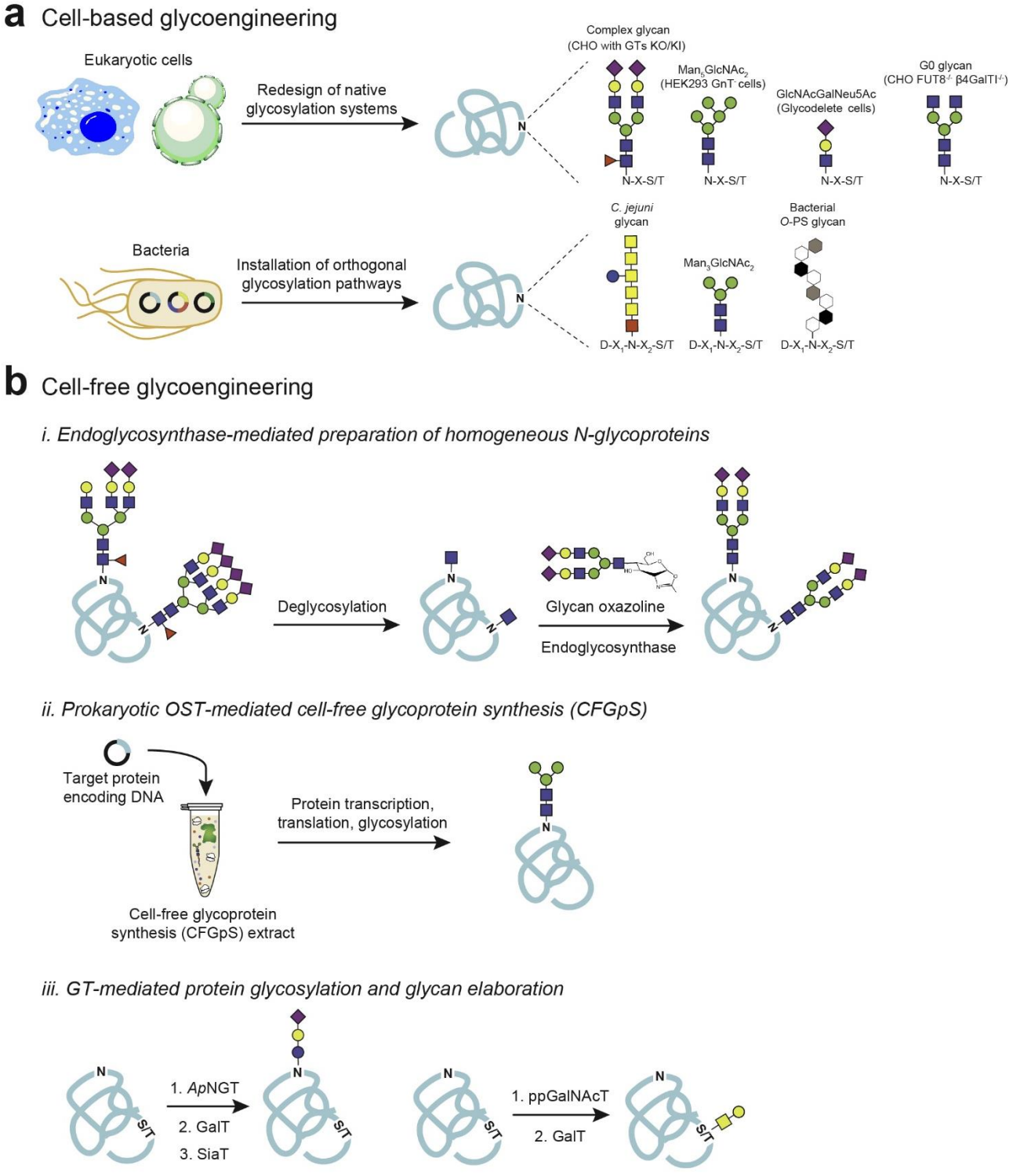

a Cell-based glycoengineering
Eukaryotic cells
Redesign of native glycosylation systems
Complex glycan (CHO with GTs KO/KI)
Man5GlcNAc2 (HEK293 GnT- cells)
GlcNAcGalNeu5Ac (Glycodelete cells)
G0 glycan (CHO FUT8-/- β4GalTI-/-)
N-X-S/T
Bacteria
Installation of orthogonal glycosylation pathways
C. jejuni glycan
Man3GlcNAc2
Bacterial O-PS glycan
D-X1-N-X2-S/T
b Cell-free glycoengineering
i. Endoglycosynthase-mediated preparation of homogeneous N-glycoproteins
Deglycosylation
Glycan oxazoline
Endoglycosynthase
ii. Prokaryotic OST-mediated cell-free glycoprotein synthesis (CFGpS)
Target protein encoding DNA
Protein transcription, translation, glycosylation
Cell-free glycoprotein synthesis (CFGpS) extract
iii. GT-mediated protein glycosylation and glycan elaboration
1. ApNGT
2. GalT
3. SiaT
1. ppGalNAcT
2. GalT

Figure 1-7: Glycoengineering strategies for producing homogeneous glycoproteins. (a) Cell-based glycoengineering involves redesign of native protein glycosylation pathways in the host organism to control glycoforms. Precise genome editing strategies such as ZFNs and CRISPR/Cas9 systems have been employed to knock-out (KO) and knock-in (KI) glycosyltransferase (GT) genes to alter endogenous glycosylation networks for producing proteins bearing desirable glycoforms such as $Man_5GlcNAc_2$, GlcNAcGalNeu_5Ac, G0, and complex-types with sialic acid capping glycoforms. Alternatively, orthogonal expression of protein glycosylation pathways in *E. coli* can be utilized to produce proteins that are site-specifically modified with bacterial *N*-glycans, eukaryotic glycans such as $Man_3GlcNAc_2$, and bacterial *O*-polysaccharide (*O*-PS) antigens. **(b)** Cell-free glycoengineering using glycoenzymes including **i.** endoglycosynthases (ENGases), **ii.** prokaryotic oligosaccharyltransferases (OSTs), and **iii.** glycosyltransferases (GTs). For ENGase-mediated glycosylation, glycoproteins bearing heterogenous *N*-glycoforms are deglycosylated using specific glycosyl hydrolases (GHs) to generate monosaccharide handle such as GlcNAc at the native glycosylation site. Pre-synthesized glycan structures containing an oxazoline functional group at the reducing end are then used as glycosyl donor in a reaction catalyzed by ENGase to remodel glycans to homogeneity. For prokaryotic OST-mediated glycosylation, cell-free extract is generated from glycoengineered *E. coli* such that the extract is enriched with all necessary gene transcription, protein translation, and protein glycosylation machineries. Supplementing extracts with DNA encoding target protein co-activates protein synthesis and site-specific protein glycosylation. For GT-mediated protein glycosylation, sequential glycosylation reactions are carried out, beginning with installation of an initial monosaccharide on the protein using a specific GT such as *Ap*NGT that installs Glc on asparagine residues and ppGalNAcT that modifies serines or threonines with GalNAc. The monosaccharide primer can then be extended, directly on glycoprotein, by a series of specific glycosyltransferase such as GalT and SiaT to generate a final glycoform.

Not to be outdone, glycoengineering in prokaryotes has emerged as an attractive strategy for cell-based production of homogenous glycoproteins (**Figure 1-7a**). The discovery of a *bona fide N*-linked protein glycosylation pathway in the mucosal

bacterium *C. jejuni* (Gross et al., 2008; Szymanski et al., 1999), and its functional reconstitution in *E. coli* (Wacker et al., 2002), laid the foundation for the development of a bacterial glycoengineering system. Owing to its lack of any native protein glycosylation systems, *E. coli* offers a blank canvas on which prescribed, orthogonal glycosylation pathways can be assembled without concern over interference from endogenous glycoenzymes. Combined with its fast growth, ease of genetic manipulation, and the ability to express a wide range of recombinant proteins, *E. coli* cells equipped with glycosylation machinery are capable of biosynthesizing designer glycoproteins bearing various therapeutically-important glycan epitopes such as the eukaryotic core *N*-glycan $Man_3GlcNAc_2$ (Glasscock et al., 2018; Valderrama-Rincon et al., 2012), bacterial *O*-polysaccharide (O-PS) antigen structures (Feldman et al., 2005a), human blood group antigens (Hug et al., 2011; Shang et al., 2016), authentic human *O*-glycans (Du et al., 2018; Natarajan et al., 2020b), and polysialic acid-containing glycans (Keys et al., 2017; Tytgat et al., 2019). Taken together, efforts in cellular glycoengineering have yielded a variety of expression platforms, both prokaryotic and eukaryotic, for producing glycoproteins with chemically-defined carbohydrate structures. Further improvement of the existing methods as well as the invention of entirely new technologies are anticipated to expand the glycoprotein expression toolkit available to scientists and engineers.

1.3.4 Cell-free glycoengineering approach for structurally-defined glycoprotein biosynthesis.

Recent advents in facile, precised gene editing tools have transformed the field of cellular glycoengineering and served as essential tools for creating a myriad of engineered organisms with desirable glycosylation capacities (Steentoft et al., 2014; Tejwani et al., 2018). Despite these advancements, carrying out cell-based glycoengineering is non-trivial as it is constrained by a multitude of factors including cell viability, cellular complexity, and difficulties in precisely tuning the expression of glycosylation components. On the other hand, cell-free approaches are not restricted by these cellular limitations and can provide more stringent control over glycan assembly and installation reactions to obtain highly pure, structurally-defined glycoproteins. Many of the early efforts in cell-free glycoengineering focused on total synthesis of glycoproteins using native chemical ligation and/or chemoselective ligation (Wang and Davis, 2013), and significant progress on this front has been made as documented in reports describing the assembly of large and complex glycoproteins including the α- and β-subunits of human hormone (Aussedat et al., 2012; Nagorny et al., 2012), interferon (Sakamoto et al., 2012), RNase C (Piontek et al., 2009a; Piontek et al., 2009b), and human erythropoietin (Wang et al., 2013a). In parallel, enzyme-mediated cell-free glycoprotein synthesis is emerging as a tool to complement chemical methods for synthesizing homogeneous glycoproteins. As mentioned earlier, enzymatic glycosylation offers precise control over stereo- and regio-chemistry

without the need for chemical protecting groups, making it especially attractive for preparative-scale biosynthesis of complex glycoproteins. In the text that follows, we present three major cell-free enzymatic approaches for preparing glycan-defined glycoproteins.

a) Endoglycosynthase-mediated preparation of homogeneous N-glycoproteins.

GHs are a class of glycoenzymes responsible for breaking specific glycosidic bonds in glycomolecules. They exhibit dual functionalities depending on whether a water molecule (hydrolysis reaction) or an activated –OH group of another carbohydrate acceptor (transglycosylation reaction) attacks an enzyme-substrate complex during catalysis (Li and Wang, 2018). The latter activity has pointed to the potential use of GHs in preparing glycoproteins by the *en bloc* transfer of pre-synthesized glycans from a glycosyl donor onto an acceptor protein (**Figure 1-7b, scheme i**). The most commonly used acceptor is a protein containing a single GlcNAc moiety, which can be generated *via* chemical synthesis or by enzymatic glycan trimming of glycoproteins derived from mammalian (Giddens et al., 2016; Goodfellow et al., 2012), yeast (Liu et al., 2018), or microbial cultures (Schwarz et al., 2010).

Following initial attempts to use transglycosylation to synthesize glycosides (Kobayashi et al., 1991), two seminal discoveries have significantly propelled progress in the field: (i) the generation of GH mutants called glycosynthases that favor transglycosylation over hydrolysis (Mackenzie et al., 1998); and (ii) the use of sugar

oxazolines as glycosyl donors for glycosynthases, which dramatically improves transglycosylation yields (Umekawa et al., 2008). To date, more than a dozen GHs including β-glycosidases (GH 1), α-galactosidases (GH 35, 36), α-fucosidases (GH 29), and endohexosaminidases (GH 18, 20, 25, 56, 84, and 85) have been cataloged and transformed into glycosynthases (Danby and Withers, 2016). Among these, endo-β-*N*-acetylglucosaminidases (ENGases) such as those isolated from *Arthrobacter protophorminae* (EndoA), *Mucor hiemalis* (EndoM), *Streptococcus species* (EndoD, S, and H), and *Elizabethkinga meningoseptica* (EndoF3) of the GH 18 and 85 families, have attracted the greatest attention due to their ability to cleave between the chitobiose core of *N*-glycans (Barreaud et al., 1995). A series of ENGase-mutants with improved transglycosylation activity has been isolated and successfully used for convergent synthesis and glycan remodeling of diverse *N*-glycoproteins including Saposin C glycopeptide with a complex-type *N*-nonasaccharide (Hojo et al., 2012), RNase B bearing a high mannose-type *N*-glycan (Amin et al., 2011; Takegawa et al., 1995), glycosylated insulin (Tomabechi et al., 2010), glycosylated HIV peptide antigen (Amin et al., 2013), fibrinogen (Giddens et al., 2016), and, most notably, IgG-Fc with a homogenous glycoform (Fan et al., 2012).

Monoclonal antibodies (mAbs) continue to be one of the fastest growing classes of biotherapeutics (Walsh, 2018). All therapeutic mAbs contain an *N*-glycan at the conserved N297 residue within the Fc region. The impact of Fc glycan on the physicochemical properties and effector functions of mAbs has been well-documented

(Jefferis, 2009). Thus, the ability to obtain mAbs in a pure glycoform not only guarantees a reproducible route for producing safe biologics, but also opens the door for engineering more effective therapeutic mAbs. With this in mind, Lai-Xi Wang and coworkers isolated two EndoS mutants that could effectively remodel the glycan of an intact IgG (Huang et al., 2012). The utility of their approach was demonstrated through the remodeling of glycans on the therapeutic mAb, Rituximab, resulting in well-defined glycoforms including $Man_3GlcNAc_2$ (M3) azide-containing M3, $Gal_2GlcNAc_2Man_3GlcNAc_2$ (G2), and $NeuNAc_2Gal_2GlcNAc_2Man_3GlcNAc_2Fuc$ (G2FS2). The glycan remodeling reactions were efficient, yielding sufficient quantities of each glycoform to be examined for binding affinity with Fcγ receptors. Following this breakthrough, ENGase-mediated chemoenzymatic glycan remodeling has become widely adopted by many research groups to generate homogeneous glycoforms of therapeutic mAbs, including Rituximab (Lin et al., 2015) and Herceptin (Kurogochi et al., 2015; Liu et al., 2018), with a relatively large glycoform library. Recent work from the Davis group further improved the transglycosylation reaction conditions with reduced unfavorable side reactions such as chemical glycation. The most optimal reaction yielded the purest glycoform of Herceptin (~90%) to date (Parsons et al., 2016). Finally, the utility of chemoenzymatic transglycosylation was recently extended to install phosphorylated glycans (Priyanka et al., 2016) and to remodel *N*-linked glycans in the Fab region (Giddens et al., 2018). Importantly, the ability to generate relatively homogenous mAb glycoforms is providing insights into how specific carbohydrate

epitopes modulate conformational changes and effector functions of antibodies, including antibody-dependent cellular cytotoxicity (ADCC), complement-dependent cytotoxicity (CDC), and anti-inflammatory activities.

b) Prokaryotic OST-mediated cell-free glycoproteins biosynthesis.

The ability to reconstitute glycoprotein biosynthesis in a well-defined, cell-free environment has the potential to transform the study of glycoscience. In such a system, not only can a particular step in glycan assembly, glycan modification, and glycan installation on the protein be carefully interrogated, but it can also facilitate the construction of engineered glycosylation pathways for making specific glycoforms of a protein. Such systems are inspired by and borrow components from natural glycosylation mechanisms found in eukaryotes, and more recently in prokaryotes.

In eukaryotes, *N*-glycoprotein biosynthesis involves the transfer of a preassembled glycan ($Glc_3Man_9GlcNAc_2$) from a dolichyl-pyrophosphate carrier to an asparagine residue within the Asn-Xaa-Thr/Ser (where X ≠ Pro) consensus sequon of a nascent polypeptide chain by an oligosaccharyltransferase (OST) enzyme (Aebi, 2013). The precursor *N*-glycan on glycoproteins then undergoes a series of GH-mediated glycan trimming and GT-mediated glycan elaboration steps in the ER and Golgi to yield the final glycoform of the protein (Arigoni-Affolter et al., 2019; Berger, 1985). The OST is a key enzyme of this pathway and consists of a protein complex containing multiple transmembrane subunit proteins, including the catalytic subunit STT3

(Kelleher and Gilmore, 2006). Early work from the Coward and Imperiali groups devised *in vitro* glycosylation assays to gain mechanistic understanding of the substrate specificity and activity of the yeast OST (Tai and Imperiali, 2001; Xu and Coward, 1997). Many of these studies were done using crude extract-containing detergent-solubilized OSTs from yeast microsomes to catalyze glycan transfer from dolichyl lipid-linked oligosaccharides (LLOs) onto peptide acceptors containing a glycosylation motif (Sharma et al., 1981; Srinivasan and Coward, 2002). Due to the inherent structural complexity, the preparation of membrane-bound OST complexes has proven difficult, often leading to inactive or unstable enzymes. Recent advances in biochemical techniques, however, have now made it possible to obtain highly-pure and active OST complexes, including those from humans, for *in vitro* functional characterization and most importantly structural elucidation (Bai et al., 2018; Ramirez et al., 2019; Wild et al., 2018). Nevertheless, the preparation of glycoproteins by eukaryotic OST-mediated *in vitro* glycosylation has yet to be realized. Key impediments include the inaccessibility of dolichyl LLO libraries (Gibbs and Coward, 1999) and the uncertainty of whether eukaryotic OSTs, which operate co-translationally, can also post-translationally modify target proteins. To date, *in vitro* glycosylation reactions catalyzed by eukaryotic OSTs have only been performed with short synthetic peptide acceptors and it remains to be seen whether these enzymes can efficiently glycosylate fully folded-proteins *in vitro* (Bai et al., 2018; Ramirez et al., 2019; Wild et al., 2018).

Similar to eukaryotes, *N*-linked protein glycosylation in certain Proteobacteria such as *Campylobacter* and *Helicobacter* species involves *en bloc* transfer of glycans from undecaprenyl-pyrophosphate (Und-PP) glycolipids onto conserved glycosylation motifs within the protein chain (Nothaft and Szymanski, 2010; Szymanski and Wren, 2005). Bacterial OSTs share a conserved architecture with eukaryotic STT3s with the exception that bacterial OSTs are single-subunit enzymes (Dell et al., 2010; Szymanski et al., 1999). Shortly after the discovery of the first *bona fide N*-glycosylation system in *C. jejuni*, Aebi and colleagues demonstrated the functional transfer of the *C. jejuni* protein glycosylation locus (*pgl*) into *E. coli* (Wacker et al., 2002), which not only facilitated mechanistic studies of the pathway but opened the door to bacterial glycoengineering.

By leveraging glycoengineered strains of *E. coli*, early work demonstrated that the *C. jejuni* OST (hereafter *Cj*OST) has a more stringent substrate specificity than eukaryotic OSTs, requiring an extended glycosylation sequon, Asp/Glu-X_{-1}-Asn-X_{+1}-Ser/Thr (where X_{-1}, $X_{+1} \neq$ Pro) (Kowarik et al., 2006b). The so called 'minus-two rule' of the *Cj*OST, requiring an acidic amino acid residue at the -2 position of the glycosylation site, did not strictly apply to other bacteria, as several *Cj*OST homologs, such as those found in *Desulfovibrio, Helicobacter*, and deep sea vent bacterial species, were observed to have significantly relaxed substrate specificity (Mills et al., 2016; Ollis et al., 2015a). Regardless of their specific sequon preferences, these enzymes are capable of installing glycans onto sequons that have been engineered at the N- and C-termini and in flexible

regions of heterologous proteins (Fisher et al., 2011; Lizak et al., 2011a) and can glycosylate such heterologous proteins both in cell-based and cell-free systems (Kowarik et al., 2006a; Ollis et al., 2015a).

An attractive feature of bacterial *N*-glycosylation systems is their inherent simplicity, which makes them readily amenable to reconstitution outside the cell. Indeed, following the functional expression of the *C. jejuni pgl* locus in *E. coli* cells, the same glycosylation reaction was recapitulated *in vitro* by Imperiali and coworkers who showed that purified *Cj*OST was capable of transferring a glycan from a synthetic donor, Und-PP-disaccharide, onto a synthetic peptide acceptor (Glover et al., 2005c). Along similar lines, Aebi and coworkers described an *in vitro* glycosylation assay comprised of purified *Cj*OST, a purified acceptor protein, and LLOs bearing the *C. jejuni* heptasaccharide glycan (*Cj*LLOs) that were extracted from glycosylation-competent *E. coli* (Kowarik et al., 2006a). Using this greatly simplified and well-controlled *in vitro* system, they were able to evaluate the ability of *Cj*OST to glycosylate distinct folding states of a model acceptor protein, RNase A^{S32D}, leading to important insights about the preferred conformation (folded versus unfolded) of bacterial acceptor proteins and the timing (co- versus post-translational) of the bacterial glycosylation process (Kowarik et al., 2006a). Further, despite being large, integral membrane proteins with thirteen transmembrane segments (Lizak et al., 2011b), bacterial OSTs can be readily overexpressed and purified from a recombinant host like *E. coli,* and the robust protocols for large-scale purification of *Cj*OST and other bacterial

OST homologs are documented (Jaffee and Imperiali, 2013; Jaroentomeechai et al., 2017). Taken together, these developments have established *in vitro* glycosylation as one of the standard tools in bacterial glycobiology and glycoengineering.

Building on these advances, Guarino and DeLisa explored coupling bacterial-based *in vitro* glycosylation with *E. coli*-based cell-free protein synthesis (CFPS) technology (Guarino and DeLisa, 2012a). Specifically, they demonstrated that by supplementing either standard cell-free S30 extracts derived from *E. coli* or the PURE (protein synthesis using recombinant elements) system (Shimizu et al., 2001) with purified *Cj*OST and extracted *Cj*LLOs, it was possible to achieve efficient glycosylation of different model glycoprotein targets including the *C. jejuni* AcrA protein and a single chain fragment variable (scFv) antibody engineered with a C-terminal glycosylation sequon. **Chapter 2** in this dissertation will describe a development of a more integrated, single-pot platform for cell-free glycoprotein synthesis (CFGpS) (**Figure 1-7b, scheme ii**) in which S30 extracts were selectively enriched with both *Cj*OST and *Cj*LLOs, effectively bypassing the need for purification and extraction, respectively, of these essential glycosylation components (Jaroentomeechai et al., 2018b). When these glyco-enriched extracts were supplemented with plasmid DNA encoding different acceptor proteins including human erythropoietin, protein synthesis and *N*-glycosylation were co-activated in a manner that resulted in appreciable amounts of site-specifically modified target proteins that retained biological activity. Importantly, the system was demonstrated to be highly modular, allowing several different *Cj*OST homologs and

structurally-distinct glycans including the eukaryotic trimannosyl core glycan, $Man_3GlcNAc_2$, to be rapidly interchanged into the cell-free reaction. **Chapter 3** of this dissertation will outline utility of the CFGpS platform for cell-free conjugate vaccine synthesis (Stark et al., 2019), which takes advantage of the fact that *Cj*OST has a relaxed glycan substrate specificity and is capable of catalyzing transfer of O-PS antigens to yield conjugate vaccines (Feldman et al., 2005a; Terra et al., 2012). By developing S30 extracts from low-endotoxin *E. coli* cells expressing *Cj*OST and different O-PS structures, it was possible to decorate a panel of different FDA-approved protein carriers such as CRM197 and *Haemophilus influenza* protein D with pathogen-specific polysaccharides including the *O*-PS antigen from the highly virulent pathogen *Franciscella tularensis* subsp. *tularensis* (type A) strain Schu S4. Importantly, conjugates supplied by this cell-free technology were observed to elicit O-PS-specific antibodies and provided complete protection against pathogen challenge in immunized mice (Stark et al., 2019).

It should be pointed out that while the CFGpS systems described above rely on heterologous expression of OSTs and LLOs in cells *prior* to extract preparation, it should be possible to streamline these systems with cell-free biosynthesis of each glycosylation component. To this end, it has been demonstrated that full-length and active membrane-bound bacterial OSTs could be directly synthesized in cell-free lysate that was supplemented with nanodiscs lipid scaffold (Schoborg et al., 2018). It has also been shown that chemically-defined LLOs bearing the *C. jejuni* glycan can be generated

by *in vitro* assembly of a biosynthetic pathway comprised of purified GTs (Glover et al., 2005a). By integrating the biogenesis of OSTs and LLOs *in vitro* with cell-free glycoprotein synthesis platforms, we anticipate the creation of a simplified yet highly modular framework for furthering the study and exploitation of the bacterial glycosylation mechanism.

In addition to *N*-linked glycosylation, certain bacterial species including *Neisseria* and *Pseudomonas* possess *O*-linked protein glycosylation pathways that involve a similar *en bloc* glycan transfer mechanism (Faridmoayer et al., 2007). Utilizing cell-free reconstitution, a central enzyme in this pathway, *O*-OST, was found to have extremely broad substrate promiscuity, both in terms of the recognizable glycan structures and their lipid carriers (Faridmoayer et al., 2008; Musumeci et al., 2013b). These features make this class of OST enzymes especially attractive for biotechnological and biomedical applications. However, widespread use of *O*-OSTs for preparing useful glycoproteins has been hindered by the lack of a consensus glycosylation motif, which in turn limits our ability to perform *O*-glycosylation on heterologous targets. This hurdle was partially resolved recently with the rational design of a minimum optimal *O*-linked recognition (MOOR) motif that was recognized by the *O*-OST PglL from *Neisseria meningitidis* (Pan et al., 2016). The MOOR sequence, which is composed of 8 amino acids flanked by two hydrophilic motifs, was used to produce an *O*-linked conjugate vaccine against *Shigella flexneri*. Since *O*-OSTs can transfer a wide range of structurally-diverse *O*-polysaccharides (Faridmoayer et al.,

2008), the advent of the MOOR motif is expected to accelerate the use of *O*-OST-based glycosylation as a platform to produce and engineer conjugate vaccines against diverse bacterial pathogens. In an important first step towards cell-free *O*-glycoprotein biosynthesis, the DeLisa group has generated S30 extracts enriched with different *O*-OSTs and LLOs bearing short-chain human *O*-glycans (*e.g.*, Tn antigen, T antigen, and sialylated versions of both) (Natarajan et al., 2020b). The resulting glyco-enriched extracts were capable of synthesizing antigenically authentic glycoforms of human mucin 1 (MUC1), thereby providing a platform for construction of designer *O*-glycoproteins and further expanding the cell-free glycoprotein expression toolkit.

c) GT-mediated protein glycosylation and glycan elaboration.

Processive protein glycosylation is prevalent in nature with the archetype represented by vertebrate mucin-type *O*-glycosylation, a mechanism whereby the glycan is assembled directly on the protein by sequential addition of monosaccharides by GTs (Hang and Bertozzi, 2005). Mucin-type *O*-glycosylation is initiated by the formation of α-glycosidic bonds between GalNAc monosaccharides and Ser/Thr residues that are catalyzed by a specific enzyme in the polypeptide-*N*-acetylgalactosaminyl transferase (ppGalNAcT) family. This core structure, named Tn antigen, can then be extended *via* sequential addition of other monosaccharides including Gal, GlcNAc, and Neu5Ac by one or more of the ~30 different Golgi-resident GTs (Bennett et al., 2012). Since mucin-associated *O*-glycan structures are associated

with many types of cancer (Pinho and Reis, 2015), there is great interest in obtaining structurally-defined *O*-glycoproteins for the development of carbohydrate-based cancer vaccine candidates. One promising avenue has been chemoenzymatic synthesis for preparing large glycopeptides carrying cancer-related *O*-glycans including Tn and sialylated Tn antigens. For example, Clausen and coworkers pre-synthesized MUC1 peptides that were subsequently modified by a series of ppGalNAcT enzymes with differential glycosylation site preferences (Sorensen et al., 2006). The GalNAc moieties on the MUC1 glycopeptides were then elongated to T or sTn antigen using β3-Gal or ST6GalNAcI transferases, respectively (**Figure 1-7b, scheme iii**). The resulting *O*-glycosylated mucin peptides were subsequently used to immunize mice, leading to the elicitation of Tn/sTn antigen-specific antibodies that could recognize specific types of cancer cells. This study highlights the potential of cell-free glycoprotein synthesis approaches in the design and production of carbohydrate-based vaccine candidates. It is worth mentioning that while *O*-GalNAcylation of mucin has been the focus of intense research, many lesser studied types of *O*-linked glycosylation have been reported in recent years including the enzymatic transfer of GlcNAc, Man, Fuc, Glc, Gal, and Xyl sugars onto specific proteins such as Notch receptors and epidermal growth factor like (EGF) repeats (Bennett et al., 2012; Haltiwanger et al., 2015; Holdener and Haltiwanger, 2019; Varki, 2017a). Given our incomplete understanding of these relatively new types of *O*-glycosylation, it stands to reason that cell-free approaches will help to decipher their mechanisms and roles in biology.

Certain gram-negative γ-proteobacteria have been found to contain unique processive *N*- and *O*-linked protein glycosylation pathways (Ohuchi et al., 2000; Zhou and Wu, 2009). Among them, *N*-glycosyltransferase (NGT) from *Actinobacillus pleuropneumoniae* (*Ap*NGT) is the best characterized enzyme, which is capable of transferring Glc residues onto the same Asn-X-Ser/Thr motif used in canonical *N*-glycosylation that proceeds by the *en bloc* mechanism (Choi et al., 2010; Kawai et al., 2011; Naegeli et al., 2014a). The *Ap*NGT has been functionally transferred into *E. coli* (Naegeli et al., 2014b), providing a novel mode of bacterial glycoengineering (Keys et al., 2017). Indeed, several groups have leveraged *Ap*NGT to site-specifically *N*-glycosylate target proteins with a Glc moiety that serves as a "glycan primer" for further extension to defined glycoforms such to α-Gal, lactose, siallylactose, LacNAc, and Lewis-X structures by prescribed GTs (Kightlinger et al., 2019; Tytgat et al., 2019). In the work by Aebi, Keys and coworkers, multiple copies of these glycoepitopes could then be installed on the same target protein to create multivalent glycopolymers or equipped onto self-assembling polypeptides to produce megadalton glycoprotein assemblies (Tytgat et al., 2019). Such multivalent glycostructures could find applications in antibody discovery and the development of novel biomedical materials. In work done by the Jewett and DeLisa groups, a multi-pot reaction scheme was devised whereby each pot contained *E. coli* extract synthesizing a specific GT enzyme (Kightlinger et al., 2019). These reaction pots containing active GTs could then be combined in a sequence-specific manner to prototype designer glycosylation pathways

(**Figure 1-7b, scheme iii**). This modular technology, called glycosylation pathway assembly by rapid *in vitro* mixing and expression (GlycoPRIME), enabled the generation of 23 unique glycan epitopes whose pathways were successfully transferred into *E. coli* to bio-manufacture useful glycoproteins including the H1HA10 protein vaccine containing an α-Gal epitope. This study showcases the power of cell-free synthetic glycobiology as a versatile tool to design, build, test, and employ designer glycosylation pathways for the development and production of putative glycomedicines. It should be pointed out that while unique bacterial enzymes such as *Ap*NGT have been harnessed for processive glycan construction, similar strategies have been developed using *Cj*PglB. That is, even though *Cj*PglB is known for its ability to transfer preassembled glycan structures, it can also be used to install a single GlcNAc residue onto acceptor protein targets as was shown recently by Davis and colleagues (Liu et al., 2014). These authors designed a series of short polyisoprenol variants that were modified with a single GlcNAc monosaccharide and used these unnatural sugar-unnatural lipid conjugates to demonstrate that purified *Cj*OST could catalyze the formation of GlcNAc-ylated peptides and proteins. They further showed that these GlcNAc-ylated species could be extended with ENGases and GTs (*e.g.*, EndoA, β1,4-GalT) thereby demonstrating a novel *in vitro* route to tailor-made glycoproteins.

1.4 Dissertation overview.

Motivated by the continuing needs for new tools and methodologies for further improving our understanding of the glycosylation process in nature as well as for enabling novel applications in glycoscience and medicine, this dissertation sought to leverage recent advances in synthetic biology and glycoscience to create a suite of cell-free tools capable of design, build, and test natural and synthetic protein glycosylation pathway (**Figure 1-8a**).

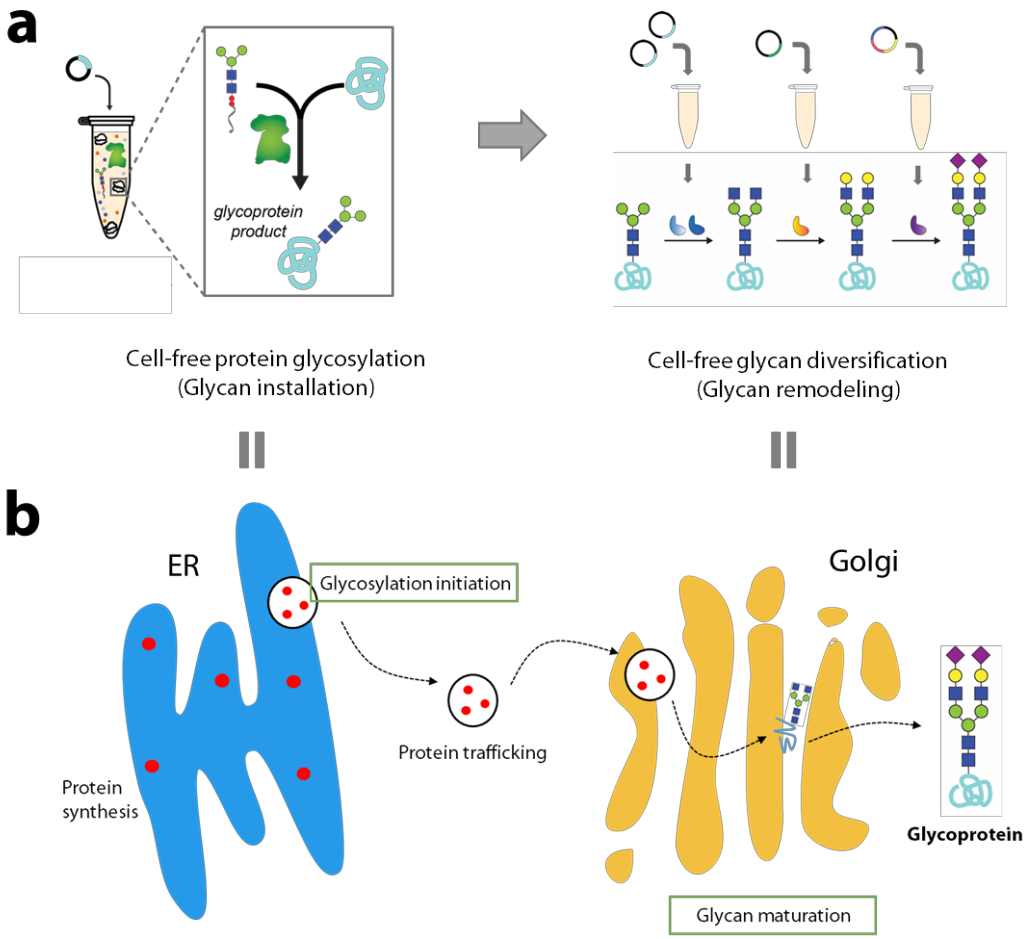


Figure 1-8. Nature-inspired cell-free synthetic glycobiology (CFSG) system. (**a**) To assemble structurally-defined glycans and glycomolecules, CFSG system is comprised of two stages. In

the first stage, cell-free protein glycosylation reaction leverages both protein synthesis and protein glycosylation machineries to biosynthesize target protein and simultaneously install prescribed glycan at the specific sequon within target protein. Glycoprotein products from the first stage is then transferred to glycan remodeling cascade where a series of glycosyltransferase is used to trim, edit, and elaborate glycan to the final structure of interest. This reaction cascade can be performed in a single or multi-pots reaction, depending on the compatibility of reaction components. (**b**) The first step in the CFSG is resembling natural processes of polypeptide biosynthesis and *N*-glycosylation initiation at the lumen of the endoplasmic reticulum (ER). The polypeptide containing *N*-glycan is then trafficked from ER to the Golgi body. Within the Golgi lumen, several reactions including *O*-glycan initiation, *N*- and *O*-glycan editing, and elaboration take place as part of the glycan maturation process. These reaction cascades in Golgi are similar to the glycan remodeling module of the CFSG.

Inspired by the elegant process of natural protein glycosylation, our Cell-Free Synthetic Glycobiology (CFSG) system is a two-stages process (**Figure 1-8a**). Akin to a natural process within the ER (**Figure 1-8b**), the first stage in CFSG involves cell-free biosynthesis and site-specific glycan installation of the target protein. Detailed works involving the creations and applications of the cell-free glycoprotein synthesis technology are described in **Chapter 2** and **Chapter 3**, respectively. The second stage in the CFSG would mimic natural glycan elaboration processes (**Figure 1-8a**-**b**). By leveraging protein engineering strategy and knowledge in system glycobiology, we generated glycan remodeling reaction cascades, akin to the glycan elaboration assembly line within the Golgi, that results in a homogenous production of several important glycans and glycoproteins. This work will be discussed in detail in **Chapter 4**. Finally, **Chapter 5** of this dissertation will summarize successful efforts in generating

CFSG tools, outline remaining challenges and knowledge gaps that need to be addressed, as well as provide a perspective on the use of the cell-free synthetic glycobiology technology to answer fundamental questions in glycoscience and to enable applications in on-demand biomanufacturing of glycoprotein therapeutics and vaccines.

1.5 Bibliographic note.

The introduction and body of this dissertation appear in several published works and presentations. The published venues will be indicated in the footnote at the beginning of each chapter. Texts and figures are subjected to change to comply with the dissertation format. All published works are co-authored with Dr. Matthew P. DeLisa and Dr. Michael C. Jewett.

This dissertation often uses pronoun 'we/our' to indicate collaborative work between the author, Dr. Matthew P. DeLisa, Dr. Michael C. Jewett, and several other collaborators.

1.6 References.

't Hart, I.M.E., Li, T., Wolfert, M.A., Wang, S., Moremen, K.W., and Boons, G.J. (2019). Chemoenzymatic synthesis of the oligosaccharide moiety of the tumor-associated antigen disialosyl globopentaosylceramide. Org Biomol Chem *17*, 7304-7308.

Abu-Qarn, M., Eichler, J., and Sharon, N. (2008). Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. Curr Opin Struct Biol *18*, 544-550.

Adiga, R., Al-adhami, M., Andar, A., Borhani, S., Brown, S., Burgenson, D., Cooper, M.A., Deldari, S., Frey, D.D., Ge, X., *et al.* (2018). Point-of-care production of therapeutic proteins of good-manufacturing-practice quality. Nat Biomed Eng.

Aebi, M. (2013). N-linked protein glycosylation in the ER. Biochim Biophys Acta *1833*, 2430-2437.

Ahn, J.H., Hwang, M.Y., Lee, K.H., Choi, C.Y., and Kim, D.M. (2007). Use of signal sequences as an in situ removable sequence element to stimulate protein synthesis in cell-free extracts. Nucleic Acids Res *35*, e21.

Albayrak, C., and Swartz, J.R. (2013). Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. Nucleic Acids Res *41*, 5949-5963.

Amin, M.N., Huang, W., Mizanur, R.M., and Wang, L.X. (2011). Convergent synthesis of homogeneous Glc1Man9GlcNAc2-protein and derivatives as ligands of molecular chaperones in protein quality control. J Am Chem Soc *133*, 14404-14417.

Amin, M.N., McLellan, J.S., Huang, W., Orwenyo, J., Burton, D.R., Koff, W.C., Kwong, P.D., and Wang, L.X. (2013). Synthetic glycopeptides reveal the glycan specificity of HIV-neutralizing antibodies. Nat Chem Biol *9*, 521-526.

Arigoni-Affolter, I., Scibona, E., Lin, C.W., Bruhlmann, D., Souquet, J., Broly, H., and Aebi, M. (2019). Mechanistic reconstruction of glycoprotein secretion through monitoring of intracellular N-glycan processing. Sci Adv *5*, eaax8930.

Ashok, A., Brison, M., and LeTallec, Y. (2017). Improving cold chain systems: Challenges and solutions. Vaccine *35*, 2217-2223.

Aussedat, B., Fasching, B., Johnston, E., Sane, N., Nagorny, P., and Danishefsky, S.J. (2012). Total synthesis of the alpha-subunit of human glycoprotein hormones: toward fully

synthetic homogeneous human follicle-stimulating hormone. J Am Chem Soc *134*, 3532-3541.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol *2*, 2006 0008.

Bai, L., Wang, T., Zhao, G., Kovach, A., and Li, H. (2018). The atomic structure of a eukaryotic oligosaccharyltransferase complex. Nature *555*, 328-333.

Baker, J.L., Celik, E., and DeLisa, M.P. (2013). Expanding the glycoengineering toolbox: the rise of bacterial N-linked protein glycosylation. Trends Biotechnol *31*, 313-323.

Barnard, G.C., Henderson, G.E., Srinivasan, S., and Gerngross, T.U. (2004). High level recombinant protein expression in Ralstonia eutropha using T7 RNA polymerase based amplification. Protein Expr Purif *38*, 264-271.

Barolo, L., Abbriano, R.M., Commault, A.S., George, J., Kahlke, T., Fabris, M., Padula, M.P., Lopez, A., Ralph, P.J., and Pernice, M. (2020). Perspectives for Glyco-Engineering of Recombinant Biopharmaceuticals from Microalgae. Cells *9*.

Barreaud, J.P., Bourgerie, S., Julien, R., Guespin-Michel, J.F., and Karamanos, Y. (1995). An endo-N-acetyl-beta-D-glucosaminidase, acting on the di-N-acetylchitobiosyl part of N-linked glycans, is secreted during sporulation of Myxococcus xanthus. J Bacteriol *177*, 916-920.

Bayburt, T.H., and Sligar, S.G. (2010). Membrane protein assembly into Nanodiscs. FEBS Lett *584*, 1721-1727.

Beatson, R., Tajadura-Ortega, V., Achkova, D., Picco, G., Tsourouktsoglou, T.D., Klausing, S., Hillier, M., Maher, J., Noll, T., Crocker, P.R., *et al.* (2016). The mucin MUC1 modulates the tumor immunological microenvironment through engagement of the lectin Siglec-9. Nat Immunol *17*, 1273-1281.

Bello, C., Wang, S., Meng, L., Moremen, K.W., and Becker, C.F. (2015). A PEGylated photocleavable auxiliary mediates the sequential enzymatic glycosylation and native chemical ligation of peptides. Angew Chem Int Ed Engl *54*, 7711-7715.

Bennett, E.P., Mandel, U., Clausen, H., Gerken, T.A., Fritz, T.A., and Tabak, L.A. (2012). Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology *22*, 736-756.

Berger, E.G. (1985). How Golgi-associated glycosylation works. Cell Biol Int Rep *9*, 407-417.

Bernatchez, S., Szymanski, C.M., Ishiyama, N., Li, J., Jarrell, H.C., Lau, P.C., Berghuis, A.M., Young, N.M., and Wakarchuk, W.W. (2005). A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in Campylobacter jejuni. J Biol Chem *280*, 4792-4802.

Bertozzi, C.R., Freeze, H.H., Varki, A., and Esko, J.D. (2009). Glycans in Biotechnology and the Pharmaceutical Industry. In Essentials of Glycobiology, nd, A. Varki, R.D. Cummings, J.D. Esko, H.H. Freeze, P. Stanley, C.R. Bertozzi, G.W. Hart, and M.E. Etzler, eds. (Cold Spring Harbor (NY)).

Bhatia, R., Gautam, S.K., Cannon, A., Thompson, C., Hall, B.R., Aithal, A., Banerjee, K., Jain, M., Solheim, J.C., Kumar, S., *et al.* (2019). Cancer-associated mucins: role in immune modulation and metastasis. Cancer Metastasis Rev *38*, 223-236.

Bhushan, R., Anthony, B.F., and Frasch, C.E. (1998). Estimation of group B streptococcus type III polysaccharide-specific antibody concentrations in human sera is antigen dependent. Infect Immun *66*, 5848-5853.

Blixt, O., Allin, K., Bohorov, O., Liu, X., Andersson-Sand, H., Hoffmann, J., and Razi, N. (2008). Glycan microarrays for screening sialyltransferase specificities. Glycoconj J *25*, 59-68.

Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M.E., Alvarez, R., Bryan, M.C., Fazio, F., Calarese, D., Stevens, J., *et al.* (2004). Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc Natl Acad Sci U S A *101*, 17033-17038.

Bogaert, D., Hermans, P.W., Adrian, P.V., Rumke, H.C., and de Groot, R. (2004). Pneumococcal vaccines: an update on current strategies. Vaccine *22*, 2209-2220.

Boles, K.S., Kannan, K., Gill, J., Felderman, M., Gouvis, H., Hubby, B., Kamrud, K.I., Venter, J.C., and Gibson, D.G. (2017). Digital-to-biological converter for on-demand production of biologics. Nat Biotechnol *35*, 672-675.

Brito, L.A., and Singh, M. (2011). Acceptable levels of endotoxin in vaccine formulations during preclinical research. J Pharm Sci *100*, 34-37.

Brodel, A.K., Sonnabend, A., Roberts, L.O., Stech, M., Wustenhagen, D.A., and Kubick, S. (2013). IRES-mediated translation of membrane proteins and glycoproteins in eukaryotic cell-free systems. PLoS One *8*, e82234.

Bundy, B.C., and Swartz, J.R. (2010). Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation. Bioconjug Chem *21*, 255-263.

Burda, P., and Aebi, M. (1999). The dolichol pathway of N-linked glycosylation. Biochim Biophys Acta *1426*, 239-257.

Cai, C., Dickinson, D.M., Li, L., Masuko, S., Suflita, M., Schultz, V., Nelson, S.D., Bhaskar, U., Liu, J., and Linhardt, R.J. (2014). Fluorous-assisted chemoenzymatic synthesis of heparan sulfate oligosaccharides. Org Lett *16*, 2240-2243.

Cai, Y., He, J., and Lu, L. (2011). Prediction of mucin-type O-glycosylation sites by a two-staged strategy. Mol Divers *15*, 427-433.

Carlson, E.D., Gan, R., Hodgman, C.E., and Jewett, M.C. (2012a). Cell-free protein synthesis: Applications come of age. Biotechnol Adv *30*, 1185-1194.

Carlson, E.D., Gan, R., Hodgman, C.E., and Jewett, M.C. (2012b). Cell-free protein synthesis: applications come of age. Biotechnol Adv *30*, 1185-1194.

Carlson, E.D., Gan, R., Hodgman, E.C., and Jewett, M.C. (2011). Cell-free protein synthesis: Applications come of age. Biotechnology Advances *30*.

Caschera, F., and Noireaux, V. (2014). Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. Biochimie *99*, 162-168.

Casella, C.R., and Mitchell, T.C. (2008). Putting endotoxin to work for us: monophosphoryl lipid A as a safe and effective vaccine adjuvant. Cell Mol Life Sci *65*, 3231-3240.

Casini, A., Chang, F.Y., Eluere, R., King, A.M., Young, E.M., Dudley, Q.M., Karim, A., Pratt, K., Bristol, C., Forget, A., *et al.* (2018). A Pressure Test to Make 10 Molecules in 90 Days: External Evaluation of Methods to Engineer Biology. J Am Chem Soc *140*, 4302-4316.

Caval, T., Tian, W., Yang, Z., Clausen, H., and Heck, A.J.R. (2018). Direct quality control of glycoengineered erythropoietin variants. Nat Commun *9*, 3342.

CDC (2019). CDC Vaccine Price List.

Celik, E., Ollis, A.A., Lasanajak, Y., Fisher, A.C., Gur, G., Smith, D.F., and DeLisa, M.P. (2015). Glycoarrays with engineered phages displaying structurally diverse oligosaccharides enable high-throughput detection of glycan-protein interactions. Biotechnol J *10*, 199-209.

Chang, M.M., Gaidukov, L., Jung, G., Tseng, W.A., Scarcelli, J.J., Cornell, R., Marshall, J.K., Lyles, J.L., Sakorafas, P., Chu, A.A., *et al.* (2019). Small-molecule control of antibody N-glycosylation in engineered mammalian cells. Nat Chem Biol *15*, 730-736.

Chavan, M., Yan, A., and Lennarz, W.J. (2005). Subunits of the translocon interact with components of the oligosaccharyl transferase complex. J Biol Chem *280*, 22917-22924.

Chen, B., Vogan, E.M., Gong, H., Skehel, J.J., Wiley, D.C., and Harrison, S.C. (2005). Determining the structure of an unliganded and fully glycosylated SIV gp120 envelope glycoprotein. Structure *13*, 197-211.

Chen, D.J., Osterrieder, N., Metzger, S.M., Buckles, E., Doody, A.M., DeLisa, M.P., and Putnam, D. (2010). Delivery of foreign antigens by engineered outer membrane vesicle vaccines. Proc Natl Acad Sci U S A *107*, 3099-3104.

Chen, L., Valentine, J.L., Huang, C.-J., Endicott, C.E., Moeller, T.D., Rasmussen, J.A., Fletcher, J.R., Boll, J.M., Rosenthal, J.A., Dobruchowska, J., *et al.* (2016a). Outer membrane vesicles displaying engineered glycotopes elicit protective antibodies. Proc Natl Acad Sci U S A.

Chen, L., Valentine, J.L., Huang, C.J., Endicott, C.E., Moeller, T.D., Rasmussen, J.A., Fletcher, J.R., Boll, J.M., Rosenthal, J.A., Dobruchowska, J., *et al.* (2016b). Outer membrane vesicles displaying engineered glycotopes elicit protective antibodies. Proc Natl Acad Sci U S A *113*, E3609-3618.

Chen, M.M., Glover, K.J., and Imperiali, B. (2007). From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. Biochemistry *46*, 5579-5585.

Chen, Y., Zhou, W., Wang, H., and Yuan, Z. (2015). Prediction of O-glycosylation sites based on multi-scale composition of amino acids and feature selection. Med Biol Eng Comput *53*, 535-544.

Choi, B.K., Bobrowicz, P., Davidson, R.C., Hamilton, S.R., Kung, D.H., Li, H., Miele, R.G., Nett, J.H., Wildt, S., and Gerngross, T.U. (2003). Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast Pichia pastoris. Proc Natl Acad Sci U S A *100*, 5022-5027.

Choi, K.J., Grass, S., Paek, S., St Geme, J.W., 3rd, and Yeo, H.J. (2010). The Actinobacillus pleuropneumoniae HMW1C-like glycosyltransferase mediates N-linked glycosylation of the Haemophilus influenzae HMW1 adhesin. PLoS One *5*, e15888.

Colley, K.J., Varki, A., and Kinoshita, T. (2015). Cellular Organization of Glycosylation. In Essentials of Glycobiology, rd, A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, *et al.*, eds. (Cold Spring Harbor (NY)), pp. 41-49.

Corfield, A. (2017). Eukaryotic protein glycosylation: a primer for histochemists and cell biologists. Histochem Cell Biol *147*, 119-147.

Crowell, L.E., Lu, A.E., Love, K.R., Stockdale, A., Timmick, S.M., Wu, D., Wang, Y.A., Doherty, W., Bonnyman, A., Vecchiarello, N., *et al.* (2018). On-demand manufacturing of clinical-quality biopharmaceuticals. Nat Biotechnol.

Cuccui, J., Thomas, R.M., Moule, M.G., D'Elia, R.V., Laws, T.R., Mills, D.C., Williamson, D., Atkins, T.P., Prior, J.L., and Wren, B.W. (2013). Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis. Open Biol *3*, 130002.

Czlapinski, J.L., and Bertozzi, C.R. (2006). Synthetic glycobiology: Exploits in the Golgi compartment. Curr Opin Chem Biol *10*, 645-651.

Danby, P.M., and Withers, S.G. (2016). Advances in Enzymatic Glycoside Synthesis. ACS Chem Biol *11*, 1784-1794.

Datsenko, K.A., and Wanner, B.L. (2000a). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640-6645.

Datsenko, K.A., and Wanner, B.L. (2000b). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640-6645.

de las Rivas, M., Lira-Navarrete, E., Gerken, T.A., and Hurtado-Guerrero, R. (2019). Polypeptide GalNAc-Ts: from redundancy to specificity. Current Opinion in Structural Biology *56*, 87-96.

Dell, A., Galadari, A., Sastre, F., and Hitchen, P. (2010). Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. Int J Microbiol *2010*, 148178.

Des Soye, B.J., Davidson, S.R., Weinstock, M.T., Gibson, D.G., and Jewett, M.C. (2018). Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens. ACS Synth Biol *7*, 2245-2255.

Des Soye, B.J., Gerbasi, V.R., Thomas, P.M., Kelleher, N.L., and Jewett, M.C. (2019). A Highly Productive, One-Pot Cell-Free Protein Synthesis Platform Based on Genomically Recoded Escherichia coli. Cell Chem Biol *26*, 1743-1754 e1749.

Du, T., Buenbrazo, N., Kell, L., Rahmani, S., Sim, L., Withers, S.G., DeFrees, S., and Wakarchuk, W. (2018). A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans. Cell Chem Biol.

Dudley, Q.M., Anderson, K.C., and Jewett, M.C. (2016). Cell-free mixing of Escherichia coli crude extracts to prototype and rationally engineer high-titer mevalonate synthesis. ACS synthetic biology *5*, 1578-1588.

Dudley, Q.M., Karim, A.S., and Jewett, M.C. (2015). Cell-free metabolic engineering: Biomanufacturing beyond the cell. Biotechnology Journal *10*, 69-82.

Dudley, Q.M., Nash, C.J., and Jewett, M.C. (2019). Cell-free biosynthesis of limonene using enzyme-enriched Escherichia coli lysates. Synth Biol (Oxf) *4*, ysz003.

Egorova, K.S., Kondakova, A.N., and Toukach, P.V. (2015). Carbohydrate Structure Database: tools for statistical analysis of bacterial, plant and fungal glycomes. Database (Oxford) *2015*.

Ehret, J., Zimmermann, M., Eichhorn, T., and Zimmer, A. (2019). Impact of cell culture media additives on IgG glycosylation produced in Chinese hamster ovary cells. Biotechnol Bioeng *116*, 816-830.

Elbein, A.D., Solf, R., Dorling, P.R., and Vosbeck, K. (1981). Swainsonine: an inhibitor of glycoprotein processing. Proc Natl Acad Sci U S A *78*, 7393-7397.

Elbein, A.D., Tropea, J.E., Mitchell, M., and Kaushal, G.P. (1990). Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. J Biol Chem *265*, 15599-15605.

Fair, R.J., Hahm, H.S., and Seeberger, P.H. (2015). Combination of automated solid-phase and enzymatic oligosaccharide synthesis provides access to alpha(2,3)-sialylated glycans. Chem Commun (Camb) *51*, 6183-6185.

Fan, S.Q., Huang, W., and Wang, L.X. (2012). Remarkable transglycosylation activity of glycosynthase mutants of endo-D, an endo-beta-N-acetylglucosaminidase from Streptococcus pneumoniae. J Biol Chem *287*, 11272-11281.

Faridmoayer, A., Fentabil, M.A., Haurat, F.M., Yi, W., Woodward, R., Wang, P., and Feldman, M.F. (2008). Extreme Substrate Promiscuity of the Neisseria Oligosaccharyl Transferase Involved in Protein O-Glycosylation. Journal of Biological Chemistry *283*, 34596-34604.

Faridmoayer, A., Fentabil, M.A., Mills, D.C., Klassen, J.S., and Feldman, M.F. (2007). Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. J Bacteriol *189*, 8088-8098.

Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., and Aebi, M. (2005a). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. Proc Natl Acad Sci U S A *102*, 3016-3021.

Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., and Aebi, M. (2005b). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. Proc Natl Acad Sci U S A *102*, 3016-3021.

Fernandez, S., Palmer, D.R., Simmons, M., Sun, P., Bisbing, J., McClain, S., Mani, S., Burgess, T., Gunther, V., and Sun, W. (2007). Potential role for Toll-like receptor 4 in mediating Escherichia coli maltose-binding protein activation of dendritic cells. Infect Immun *75*, 1359-1363.

Figueiredo, D., Turcotte, C., Frankel, G., Li, Y., Dolly, O., Wilkin, G., Marriott, D., Fairweather, N., and Dougan, G. (1995). Characterization of recombinant tetanus toxin derivatives suitable for vaccine development. Infect Immun *63*, 3218-3221.

Figueroa, C.M., Lunn, J.E., and Iglesias, A.A. (2021). Nucleotide-sugar metabolism in plants: the legacy of Luis F. Leloir. J Exp Bot *72*, 4053-4067.

Fischer, E. (1890). Synthese des Traubenzuckers. Mittheilungen *23*, 6.

Fisher, A.C., Haitjema, C.H., Guarino, C., Celik, E., Endicott, C.E., Reading, C.A., Merritt, J.H., Ptak, A.C., Zhang, S., and DeLisa, M.P. (2011). Production of secretory and

extracellular N-linked glycoproteins in Escherichia coli. Appl Environ Microbiol *77*, 871-881.

Frasch, C.E. (2009). Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. Vaccine *27*, 6468-6470.

Fukui, S., Feizi, T., Galustian, C., Lawson, A.M., and Chai, W. (2002). Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. Nat Biotechnol *20*, 1011-1017.

Fulop, M., Mastroeni, P., Green, M., and Titball, R.W. (2001). Role of antibody to lipopolysaccharide in protection against low- and high-virulence strains of Francisella tularensis. Vaccine *19*, 4465-4472.

Gagneux, P., and Varki, A. (1999). Evolutionary considerations in relating oligosaccharide diversity to biological function. Glycobiology *9*, 747-755.

Gao, X.D., Nishikawa, A., and Dean, N. (2004). Physical interactions between the Alg1, Alg2, and Alg11 mannosyltransferases of the endoplasmic reticulum. Glycobiology *14*, 559-570.

Gao, X.D., Tachikawa, H., Sato, T., Jigami, Y., and Dean, N. (2005). Alg14 recruits Alg13 to the cytoplasmic face of the endoplasmic reticulum to form a novel bipartite UDP-N-acetylglucosamine transferase required for the second step of N-linked glycosylation. J Biol Chem *280*, 36254-36262.

Garcia-Quintanilla, F., Iwashkiw, J.A., Price, N.L., Stratilo, C., and Feldman, M.F. (2014). Production of a recombinant vaccine candidate against Burkholderia pseudomallei exploiting the bacterial N-glycosylation machinery. Front Microbiol *5*, 381.

Gebhart, C., Ielmini, M.V., Reiz, B., Price, N.L., Aas, F.E., Koomey, M., and Feldman, M.F. (2012). Characterization of exogenous bacterial oligosaccharyltransferases in Escherichia coli reveals the potential for O-linked protein glycosylation in Vibrio cholerae and Burkholderia thailandensis. Glycobiology *22*, 962-974.

Gendler, S.J. (2001). MUC1, the renaissance molecule. J Mammary Gland Biol Neoplasia *6*, 339-353.

Georgi, V., Georgi, L., Blechert, M., Bergmeister, M., Zwanzig, M., Wustenhagen, D.A., Bier, F.F., Jung, E., and Kubick, S. (2016). On-chip automation of cell-free protein synthesis: new opportunities due to a novel reaction mode. Lab Chip *16*, 269-281.

Gibbs, B.S., and Coward, J.K. (1999). Dolichylpyrophosphate oligosaccharides: large-scale isolation and evaluation as oligosaccharyltransferase substrates. Bioorg Med Chem *7*, 441-447.

Giddens, J.P., Lomino, J.V., Amin, M.N., and Wang, L.X. (2016). Endo-F3 Glycosynthase Mutants Enable Chemoenzymatic Synthesis of Core-fucosylated Triantennary Complex Type Glycopeptides and Glycoproteins. J Biol Chem *291*, 9356-9370.

Giddens, J.P., Lomino, J.V., DiLillo, D.J., Ravetch, J.V., and Wang, L.X. (2018). Site-selective chemoenzymatic glycoengineering of Fab and Fc glycans of a therapeutic antibody. Proc Natl Acad Sci U S A *115*, 12023-12027.

Glasscock, C.J., Yates, L.E., Jaroentomeechai, T., Wilson, J.D., Merritt, J.H., Lucks, J.B., and DeLisa, M.P. (2018). A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation. Metab Eng *47*, 488-495.

Glover, K., Weerapana, E., and Imperiali, B. (2005a). In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. Proceedings of the National Academy of Sciences of the United States of America *102*, 14255-14259.

Glover, K.J., Weerapana, E., Numao, S., and Imperiali, B. (2005b). Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni.

Glover, K.J., Weerapana, E., Numao, S., and Imperiali, B. (2005c). Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. Chem Biol *12*, 1311-1315.

Goerke, A.R., and Swartz, J.R. (2009). High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. Biotechnol Bioeng *102*, 400-416.

Goodfellow, J.J., Baruah, K., Yamamoto, K., Bonomelli, C., Krishna, B., Harvey, D.J., Crispin, M., Scanlan, C.N., and Davis, B.G. (2012). An endoglycosidase with alternative glycan specificity allows broadened glycoprotein remodelling. J Am Chem Soc *134*, 8030-8033.

Goshima, N., Kawamura, Y., Fukumoto, A., Miura, A., Honma, R., Satoh, R., Wakamatsu, A., Yamamoto, J.-i., Kimura, K., Nishikawa, T., *et al.* (2008). Human protein factory for

converting the transcriptome into an in vitro–expressed proteome. Nature Methods *5*, 1011-1017.

Gross, J., Grass, S., Davis, A.E., Gilmore-Erdmann, P., Townsend, R.R., and St Geme, J.W., 3rd (2008). The Haemophilus influenzae HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. J Biol Chem *283*, 26010-26015.

Guarino, and DeLisa (2012a). A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. Glycobiology *22*.

Guarino, C., and DeLisa, M.P. (2012b). A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. Glycobiology *22*, 596-601.

Gurramkonda, C., Rao, A., Borhani, S., Pilli, M., Deldari, S., Ge, X., Pezeshk, N., Han, T.C., Tolosa, M., Kostov, Y., *et al.* (2018). Improving the recombinant human erythropoietin glycosylation using microsome supplementation in CHO cell-free system. Biotechnol Bioeng *115*, 1253-1264.

Hagelueken, G., Clarke, B.R., Huang, H., Tuukkanen, A., Danciu, I., Svergun, D.I., Hussain, R., Liu, H., Whitfield, C., and Naismith, J.H. (2015). A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. Nat Struct Mol Biol *22*, 50-56.

Haghi, F., Peerayeh, S.N., Siadat, S.D., and Montajabiniat, M. (2011). Cloning, expression and purification of outer membrane protein PorA of Neisseria meningitidis serogroup B. J Infect Dev Ctries *5*, 856-862.

Hahm, H.S., Schlegel, M.K., Hurevich, M., Eller, S., Schuhmacher, F., Hofmann, J., Pagel, K., and Seeberger, P.H. (2017). Automated glycan assembly using the Glyconeer 2.1 synthesizer. Proc Natl Acad Sci U S A *114*, E3385-E3389.

Halic, M., and Beckmann, R. (2005). The signal recognition particle and its interactions during protein targeting. Curr Opin Struct Biol *15*, 116-125.

Haltiwanger, R.S., Wells, L., Freeze, H.H., and Stanley, P. (2015). Other Classes of Eukaryotic Glycans. In Essentials of Glycobiology, rd, A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, A.G. Darvill, T. Kinoshita, N.H. Packer, *et al.*, eds. (Cold Spring Harbor (NY)), pp. 151-160.

Hamilton, B.S., Wilson, J.D., Shumakovich, M.A., Fisher, A.C., Brooks, J.C., Pontes, A., Naran, R., Heiss, C., Gao, C., Kardish, R., *et al.* (2017). A library of chemically defined human N-glycans synthesized from microbial oligosaccharide precursors. Sci Rep *7*, 15907.

Hamilton, S.R., Bobrowicz, P., Bobrowicz, B., Davidson, R.C., Li, H., Mitchell, T., Nett, J.H., Rausch, S., Stadheim, T.A., Wischnewski, H., *et al.* (2003). Production of complex human glycoproteins in yeast. Science *301*, 1244-1246.

Hamilton, S.R., and Zha, D. (2015). Progress in Yeast Glycosylation Engineering. Methods Mol Biol *1321*, 73-90.

Hammerling, M.J., Fritz, B.R., Yoesep, D.J., Kim, D.S., Carlson, E.D., and Jewett, M.C. (2020). In vitro ribosome synthesis and evolution through ribosome display. Nat Commun *11*, 1108.

Hang, H.C., and Bertozzi, C.R. (2005). The chemistry and biology of mucin-type O-linked glycosylation. Bioorg Med Chem *13*, 5021-5034.

Hansen, L., Husein, D.M., Gericke, B., Hansen, T., Pedersen, O., Tambe, M.A., Freeze, H.H., Naim, H.Y., Henrissat, B., Wandall, H.H., *et al.* (2020). A mutation map for human glycoside hydrolase genes. Glycobiology *30*, 500-515.

Hartley, M.D., Morrison, M.J., Aas, F.E., Borud, B., Koomey, M., and Imperiali, B. (2011). Biochemical characterization of the O-linked glycosylation pathway in Neisseria gonorrhoeae responsible for biosynthesis of protein glycans containing N,N'-diacetylbacillosamine. Biochemistry *50*, 4936-4948.

Hassan, H., Badr, A., and Abdelhalim, M.B. (2015). Prediction of O-glycosylation Sites Using Random Forest and GA-Tuned PSO Technique. Bioinform Biol Insights *9*, 103-109.

Hatz, C.F., Bally, B., Rohrer, S., Steffen, R., Kramme, S., Siegrist, C.A., Wacker, M., Alaimo, C., and Fonck, V.G. (2015). Safety and immunogenicity of a candidate bioconjugate vaccine against Shigella dysenteriae type 1 administered to healthy adults: A single blind, partially randomized Phase I study. Vaccine *33*, 4594-4601.

Hayes, C. (2012). Biomolecular Breadboards:Protocols:cost estimate. http://www.openwetware.org/wiki/Biomolecular_Breadboards:Protocols:cost_estimate.

Hebert, D.N., Lamriben, L., Powers, E.T., and Kelly, J.W. (2014). The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. Nat Chem Biol *10*, 902-910.

Heffner, K.M., Wang, Q., Hizal, D.B., Can, O., and Betenbaugh, M.J. (2018). Glycoengineering of Mammalian Expression Systems on a Cellular Level. Adv Biochem Eng Biotechnol.

Helenius, A., and Aebi, M. (2001a). Intracellular functions of N-linked glycans. Science (New York, NY) *291*, 2364-2369.

Helenius, A., and Aebi, M. (2001b). Intracellular functions of N-linked glycans. Science *291*, 2364-2369.

Helenius, J., Ng, D.T., Marolda, C.L., Walter, P., Valvano, M.A., and Aebi, M. (2002). Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. Nature *415*, 447-450.

Herget, S., Toukach, P.V., Ranzinger, R., Hull, W.E., Knirel, Y.A., and von der Lieth, C.W. (2008). Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans. BMC Struct Biol *8*, 35.

Hirata, T., Mishra, S.K., Nakamura, S., Saito, K., Motooka, D., Takada, Y., Kanzawa, N., Murakami, Y., Maeda, Y., Fujita, M., *et al.* (2018). Identification of a Golgi GPI-N-acetylgalactosamine transferase with tandem transmembrane regions in the catalytic domain. Nat Commun *9*, 405.

Hojo, H., Tanaka, H., Hagiwara, M., Asahina, Y., Ueki, A., Katayama, H., Nakahara, Y., Yoneshige, A., Matsuda, J., Ito, Y., *et al.* (2012). Chemoenzymatic synthesis of hydrophobic glycoprotein: synthesis of saposin C carrying complex-type carbohydrate. J Org Chem *77*, 9437-9446.

Holdener, B.C., and Haltiwanger, R.S. (2019). Protein O-fucosylation: structure and function. Curr Opin Struct Biol *56*, 78-86.

Hong, S.H., Ntai, I., Haimovich, A.D., Kelleher, N.L., Isaacs, F.J., and Jewett, M.C. (2014). Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation. ACS Synth Biol *3*, 398-409.

Huang, W., Giddens, J., Fan, S.Q., Toonstra, C., and Wang, L.X. (2012). Chemoenzymatic glycoengineering of intact IgG antibodies for gain of functions. J Am Chem Soc *134*, 12308-12318.

Hug, I., Zheng, B., Reiz, B., Whittal, R.M., Fentabil, M.A., Klassen, J.S., and Feldman, M.F. (2011). Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein. The Journal of biological chemistry *286*, 37887-37894.

Humphreys, G. (2011). Vaccination: rattling the supply chain (Bulletin of the World Health Organization: World Health Organization).

Hurtado, J., Acharya, D., Lai, H., Sun, H., Kallolimath, S., Steinkellner, H., Bai, F., and Chen, Q. (2020). In vitro and in vivo efficacy of anti-chikungunya virus monoclonal antibodies produced in wild-type and glycoengineered Nicotiana benthamiana plants. Plant Biotechnol J *18*, 266-273.

Huttner, A., Hatz, C., van den Dobbelsteen, G., Abbanat, D., Hornacek, A., Frolich, R., Dreyer, A.M., Martin, P., Davies, T., Fae, K., *et al.* (2017). Safety, immunogenicity, and preliminary clinical efficacy of a vaccine against extraintestinal pathogenic Escherichia coli in women with a history of recurrent urinary tract infection: a randomised, single-blind, placebo-controlled phase 1b trial. Lancet Infect Dis.

Ihssen, J., Kowarik, M., Dilettoso, S., Tanner, C., Wacker, M., and Thony-Meyer, L. (2010a). Production of glycoprotein vaccines in Escherichia coli. Microb Cell Fact *9*, 61.

Ihssen, J., Kowarik, M., Dilettoso, S., Tanner, C., Wacker, M., and Thony-Meyer, L. (2010b). Production of glycoprotein vaccines in Escherichia coli. Microb Cell Fact *9*, 61.

Imperiali, B., and O'Connor, S.E. (1999). Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. Curr Opin Chem Biol *3*, 643-649.

Jackson, K., Kanamori, T., Ueda, T., and Fan, Z.H. (2014). Protein synthesis yield increased 72 times in the cell-free PURE system. Integr Biol (Camb) *6*, 781-788.

Jaffee, M.B., and Imperiali, B. (2013). Optimized protocol for expression and purification of membrane-bound PglB, a bacterial oligosaccharyl transferase. Protein expression and purification *89*, 241250.

Jansson, P.E., Lindberg, B., Widmalm, G., and Leontein, K. (1987). Structural studies of the Escherichia coli O78 O-antigen polysaccharide. Carbohydr Res *165*, 87-92.

Jaroentomeechai, T., Stark, J.C., Natarajan, A., Glasscock, C.J., Yates, L.E., Hsu, K.J., Mrksich, M., Jewett, M.C., and DeLisa, M.P. (2018a). Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. Nat Commun *9*, 2686.

Jaroentomeechai, T., Stark, J.C., Natarajan, A., Glasscock, C.J., Yates, L.E., Hsu, K.J., Mrksich, M., Jewett, M.C., and DeLisa, M.P. (2018b). Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. Nat Commun *9*, 2686.

Jaroentomeechai, T., Zheng, X., Hershewe, J., Stark, J.C., Jewett, M.C., and DeLisa, M.P. (2017). A Pipeline for Studying and Engineering Single-Subunit Oligosaccharyltransferases. Methods Enzymol *597*, 55-81.

Jefferis, R. (2009). Recombinant antibody therapeutics: the impact of glycosylation on mechanisms of action. Trends Pharmacol Sci *30*, 356-362.

Jervis, A.J., Butler, J.A., Lawson, A.J., Langdon, R., Wren, B.W., and Linton, D. (2012). Characterization of the structurally diverse N-linked glycans of Campylobacter species. J Bacteriol *194*, 2355-2362.

Jewett, M.C., Calhoun, K.A., Voloshin, A., Wuu, J.J., and Swartz, J.R. (2008a). An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol *4*, 220.

Jewett, M.C., Calhoun, K.A., Voloshin, A., Wuu, J.J., and Swartz, J.R. (2008b). An integrated cell-free metabolic platform for protein production and synthetic biology. Molecular systems biology *4*, 220.

Jewett, M.C., and Swartz, J.R. (2004a). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnology and bioengineering *86*, 19-26.

Jewett, M.C., and Swartz, J.R. (2004b). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol Bioeng *86*, 19-26.

Jewett, M.C., and Swartz, J.R. (2004c). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol Bioeng *86*, 19-26.

Jiang, B., Argyros, R., Bukowski, J., Nelson, S., Sharkey, N., Kim, S., Copeland, V., Davidson, R.C., Chen, R., Zhuang, J., *et al.* (2015). Inactivation of a GAL4-like transcription factor improves cell fitness and product yield in glycoengineered Pichia pastoris strains. Appl Environ Microbiol *81*, 260-271.

Jin, C., Gibani, M.M., Moore, M., Juel, H.B., Jones, E., Meiring, J., Harris, V., Gardner, J., Nebykova, A., Kerridge, S.A., *et al.* (2017). Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of Salmonella Typhi: a randomised controlled, phase 2b trial. Lancet (London, England) *390*, 2472-2480.

Johnson, J.R. (1991). Virulence factors in Escherichia coli urinary tract infection. Clin Microbiol Rev *4*, 80-128.

Kai, L., Orban, E., Henrich, E., Proverbio, D., Dotsch, V., and Bernhard, F. (2015). Co-translational stabilization of insoluble proteins in cell-free expression systems. Methods Mol Biol *1258*, 125-143.

Kaiser, L., Graveland-Bikker, J., Steuerwald, D., Vanberghem, M., Herlihy, K., and Zhang, S. (2008). Efficient cell-free production of olfactory receptors: detergent optimization, structure, and ligand binding analyses. Proc Natl Acad Sci U S A *105*, 15726-15731.

Karim, A.S., Dudley, Q.M., Juminga, A., Yuan, Y., Crowe, S.A., Heggestad, J.T., Abdalla, T., Grubbe, W., Rasor, B., Coar, D., *et al.* (2020). In vitro prototyping and rapid optimization of biosynthetic enzymes for cellular design. bioRxiv.

Karim, A.S., and Jewett, M.C. (2016). A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. Metab Eng *36*, 116-126.

Kasprzak, A., and Adamek, A. (2019). Mucins: the Old, the New and the Promising Factors in Hepatobiliary Carcinogenesis. Int J Mol Sci *20*.

Kawai, F., Grass, S., Kim, Y., Choi, K.J., St Geme, J.W., 3rd, and Yeo, H.J. (2011). Structural insights into the glycosyltransferase activity of the Actinobacillus pleuropneumoniae HMW1C-like protein. J Biol Chem *286*, 38546-38557.

Kelleher, D.J., and Gilmore, R. (2006). An evolving view of the eukaryotic oligosaccharyltransferase. Glycobiology *16*, 47R-62R.

Kellokumpu, S., Hassinen, A., and Glumoff, T. (2016). Glycosyltransferase complexes in eukaryotes: long-known, prevalent but still unrecognized. Cell Mol Life Sci *73*, 305-325.

Keys, T.G., and Aebi, M. (2017). Engineering protein glycosylation in prokaryotes. Curr Opin Syst Biol *5*, 23-31.

Keys, T.G., Wetter, M., Hang, I., Rutschmann, C., Russo, S., Mally, M., Steffen, M., Zuppiger, M., Muller, F., Schneider, J., *et al.* (2017). A biosynthetic route for polysialylating proteins in Escherichia coli. Metab Eng *44*, 293-301.

Kiga, D., Sakamoto, K., Kodama, K., Kigawa, T., Matsuda, T., Yabuki, T., Shirouzu, M., Harada, Y., Nakayama, H., Takio, K., *et al.* (2002). An engineered Escherichia coli tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. Proc Natl Acad Sci U S A *99*, 9715-9720.

Kightlinger, W., Duncker, K.E., Ramesh, A., Thames, A.H., Natarajan, A., Stark, J.C., Yang, A., Lin, L., Mrksich, M., DeLisa, M.P., *et al.* (2019). A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. Nat Commun *10*, 5404.

Kightlinger, W., Lin, L., Rosztoczy, M., Li, W., DeLisa, M.P., Mrksich, M., and Jewett, M.C. (2018). Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. Nat Chem Biol.

Kim, D.M., and Swartz, J.R. (2001). Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. Biotechnol Bioeng *74*, 309-316.

Kim, J., Copeland, C.E., Padumane, S.R., and Kwon, Y.C. (2019). A Crude Extract Preparation and Optimization from a Genomically Engineered Escherichia coli for the Cell-Free Protein Synthesis System: Practical Laboratory Guideline. Methods Protoc *2*.

Knapp, K.G., Goerke, A.R., and Swartz, J.R. (2007). Cell-free synthesis of proteins that require disulfide bonds using glucose as an energy source. Biotechnol Bioeng *97*, 901-908.

Kobayashi, S., Kashiwa, K., Kawasaki, T., and Shoda, S. (1991). Novel method for polysaccharide synthesis using an enzyme: the first in vitro synthesis of cellulose via a nonbiosynthetic path utilizing cellulase as catalyst. J Am Chem Soc *113*, 3079-3084.

Kowarik, M., Numao, S., Feldman, M.F., Schulz, B.L., Callewaert, N., Kiermaier, E., Catrein, I., and Aebi, M. (2006a). N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. Science *314*, 1148-1150.

Kowarik, M., Young, M.N., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., *et al.* (2006b). Definition of the bacterial N-glycosylation site consensus sequence. The EMBO Journal *25*, 1957-1966.

Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., *et al.* (2006c). Definition of the bacterial N-glycosylation site consensus sequence. EMBO J *25*, 1957-1966.

Kufe, D.W. (2009). Mucins in cancer: function, prognosis and therapy. Nat Rev Cancer *9*, 874-885.

Kukuruzinska, M.A., and Robbins, P.W. (1987). Protein glycosylation in yeast: transcript heterogeneity of the ALG7 gene. Proc Natl Acad Sci U S A *84*, 2145-2149.

Kurogochi, M., Mori, M., Osumi, K., Tojino, M., Sugawara, S., Takashima, S., Hirose, Y., Tsukimura, W., Mizuno, M., Amano, J., *et al.* (2015). Glycoengineered Monoclonal Antibodies with Homogeneous Glycan (M3, G0, G2, and A2) Using a Chemoenzymatic Approach Have Different Affinities for FcgammaRIIIa and Variable Antibody-Dependent Cellular Cytotoxicity Activities. PLoS One *10*, e0132848.

Kwon, Y.-C., and Jewett, M.C. (2015a). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. Scientific Reports *5*, 8663.

Kwon, Y.C., and Jewett, M.C. (2015b). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. Sci Rep *5*, 8663.

L'vov, V.L., Shashkov, A.S., Dmitriev, B.A., Kochetkov, N.K., Jann, B., and Jann, K. (1984). Structural studies of the O-specific side chain of the lipopolysaccharide from Escherichia coli O:7. Carbohydrate research *126*, 249-259.

Lanctot, P.M., Gage, F.H., and Varki, A.P. (2007). The glycans of stem cells. Curr Opin Chem Biol *11*, 373-380.

Larsen, I.S.B., Narimatsu, Y., Joshi, H.J., Siukstaite, L., Harrison, O.J., Brasch, J., Goodman, K.M., Hansen, L., Shapiro, L., Honig, B., *et al.* (2017). Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. Proc Natl Acad Sci U S A *114*, 11163-11168.

Lee, J., Schwieter, K.E., Watkins, A.M., Kim, D.S., Yu, H., Schwarz, K.J., Lim, J., Coronado, J., Byrom, M., Anslyn, E.V., *et al.* (2019). Expanding the limits of the second genetic code with ribozymes. Nat Commun *10*, 5097.

Leloir LF, C.E. (1953). The enzymic synthesis of trehalose phosphate. Journal of the American Chemical Society *75*, 2.

Li, C., and Wang, L.X. (2018). Chemoenzymatic Methods for the Synthesis of Glycoproteins. Chem Rev *118*, 8359-8413.

Li, L., Liu, Y., Ma, C., Qu, J., Calderon, A.D., Wu, B., Wei, N., Wang, X., Guo, Y., Xiao, Z., *et al.* (2015). Efficient Chemoenzymatic Synthesis of an N-glycan Isomer Library. Chem Sci *6*, 5652-5661.

Li, P.J., Huang, S.Y., Chiang, P.Y., Fan, C.Y., Guo, L.J., Wu, D.Y., Angata, T., and Lin, C.C. (2019a). Chemoenzymatic Synthesis of DSGb5 and Sialylated Globo-series Glycans. Angew Chem Int Ed Engl *58*, 11273-11278.

Li, S.T., Lu, T.T., Xu, X.X., Ding, Y., Li, Z., Kitajima, T., Dean, N., Wang, N., and Gao, X.D. (2019b). Reconstitution of the lipid-linked oligosaccharide pathway for assembly of high-mannose N-glycans. Nat Commun *10*, 1813.

Li, T., Liu, L., Wei, N., Yang, J.Y., Chapla, D.G., Moremen, K.W., and Boons, G.J. (2019c). An automated platform for the enzyme-mediated assembly of complex oligosaccharides. Nat Chem *11*, 229-236.

Li, Z., and Feizi, T. (2018). The neoglycolipid (NGL) technology-based microarrays and future prospects. FEBS Lett *592*, 3976-3991.

Lim, H.J., and Kim, D.M. (2019). Cell-Free Metabolic Engineering: Recent Developments and Future Prospects. Methods Protoc *2*.

Lin, C.W., Tsai, M.H., Li, S.T., Tsai, T.I., Chu, K.C., Liu, Y.C., Lai, M.Y., Wu, C.Y., Tseng, Y.C., Shivatare, S.S., *et al.* (2015). A common glycan structure on immunoglobulin G for enhancement of effector functions. Proc Natl Acad Sci U S A *112*, 10611-10616.

Lingappa, V.R., Lingappa, J.R., Prasad, R., Ebner, K.E., and Blobel, G. (1978). Coupled cell-free synthesis, segregation, and core glycosylation of a secretory protein. Proc Natl Acad Sci U S A *75*, 2338-2342.

Linton, D., Allan, E., Karlyshev, A.V., Cronshaw, A.D., and Wren, B.W. (2002). Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in Campylobacter jejuni. Mol Microbiol *43*, 497-508.

Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M., and Wren, B.W. (2005). Functional analysis of the Campylobacter jejuni N-linked protein glycosylation pathway. Mol Microbiol *55*, 1695-1703.

Lithgow, K.V., Scott, N.E., Iwashkiw, J.A., Thomson, E.L., Foster, L.J., Feldman, M.F., and Dennis, J.J. (2014). A general protein O-glycosylation system within the Burkholderia cepacia complex is involved in motility and virulence. Mol Microbiol *92*, 116-137.

Liu, C.P., Tsai, T.I., Cheng, T., Shivatare, V.S., Wu, C.Y., Wu, C.Y., and Wong, C.H. (2018). Glycoengineering of antibody (Herceptin) through yeast expression and in vitro enzymatic glycosylation. Proc Natl Acad Sci U S A *115*, 720-725.

Liu, D., and Reeves, P.R. (1994). Escherichia coli K12 regains its O antigen. Microbiology *140 (Pt 1)*, 49-57.

Liu, F., Vijayakrishnan, B., Faridmoayer, A., Taylor, T.A., Parsons, T.B., Bernardes, G.J., Kowarik, M., and Davis, B.G. (2014). Rationally designed short polyisoprenol-linked PglB substrates for engineered polypeptide and protein N-glycosylation. J Am Chem Soc *136*, 566-569.

Lizak, C., Fan, Y.Y., Weber, T.C., and Aebi, M. (2011a). N-Linked glycosylation of antibody fragments in Escherichia coli. Bioconjug Chem *22*, 488-496.

Lizak, C., Gerber, S., Numao, S., Aebi, M., and Locher, K.P. (2011b). X-ray structure of a bacterial oligosaccharyltransferase. Nature *474*, 350-355.

Lizak, C., Gerber, S., Zinne, D., Michaud, G., Schubert, M., Chen, F., Bucher, M., Darbre, T., Zenobi, R., Reymond, J.L., *et al.* (2014). A catalytically essential motif in external loop 5 of the bacterial oligosaccharyltransferase PglB. J Biol Chem *289*, 735-746.

Lombard, J. (2016). The multiple evolutionary origins of the eukaryotic N-glycosylation pathway. Biol Direct *11*.

Lommel, M., and Strahl, S. (2009). Protein O-mannosylation: conserved from bacteria to humans. Glycobiology *19*, 816-828.

Lu, Z., Madico, G., Roche, M.I., Wang, Q., Hui, J.H., Perkins, H.M., Zaia, J., Costello, C.E., and Sharon, J. (2012). Protective B-cell epitopes of Francisella tularensis O-polysaccharide in a mouse model of respiratory tularaemia. Immunology *136*, 352-360.

Lydon, P., Zipursky, S., Tevi-Benissan, C., Djingarey, M.H., Gbedonou, P., Youssouf, B.O., and Zaffran, M. (2014). Economic benefits of keeping vaccines at ambient temperature during mass vaccination: the case of meningitis A vaccine in Chad. Bull World Health Organ *92*, 86-92.

Ma, Z., Zhang, H., Shang, W., Zhu, F., Han, W., Zhao, X., Han, D., Wang, P.G., and Chen, M. (2014). Glycoconjugate vaccine containing Escherichia coli O157:H7 O-antigen linked with maltose-binding protein elicits humoral and cellular responses. PloS one *9*, e105215.

Mackenzie, L.F., Wang, Q., Warren, R.A.J., and Withers, S.G. (1998). Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis. J Am Chem Soc *120*, 5583-5584.

Marshall, L.E., Nelson, M., Davies, C.H., Whelan, A.O., Jenner, D.C., Moule, M.G., Denman, C., Cuccui, J., Atkins, T.P., Wren, B.W., *et al.* (2018). An O-antigen glycoconjugate vaccine produced using protein glycan coupling technology is protective in an inhalational rat model of tularemia. J Immunol Res *2018*, 8087916.

Martin, R.W., Des Soye, B.J., Kwon, Y.C., Kay, J., Davis, R.G., Thomas, P.M., Majewska, N.I., Chen, C.X., Marcum, R.D., Weiss, M.G., *et al.* (2018). Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. Nat Commun *9*, 1203.

Matsuoka, K., Komori, H., Nose, M., Endo, Y., and Sawasaki, T. (2010). Simple screening method for autoantigen proteins using the N-terminal biotinylated protein library produced by wheat cell-free synthesis. J Proteome Res *9*, 4264-4273.

Matsushita, T., Nagashima, I., Fumoto, M., Ohta, T., Yamada, K., Shimizu, H., Hinou, H., Naruchi, K., Ito, T., Kondo, H., *et al.* (2010). Artificial Golgi apparatus: globular protein-like dendrimer facilitates fully automated enzymatic glycan synthesis. J Am Chem Soc *132*, 16651-16656.

Matthaei, J.H., Jones, O.W., Martin, R.G., and Nirenberg, M.W. (1962). Characteristics and composition of RNA coding units. Proc Natl Acad Sci U S A *48*, 666-677.

Matthies, D., Haberstock, S., Joos, F., Dotsch, V., Vonck, J., Bernhard, F., and Meier, T. (2011). Cell-free expression and assembly of ATP synthase. J Mol Biol *413*, 593-603.

McQuillan, A.M., Byrd-Leotis, L., Heimburg-Molinaro, J., and Cummings, R.D. (2019). Natural and synthetic sialylated glycan microarrays and their applications. Front Mol Biosci *6*, 88.

Meng, C., Sasmal, A., Zhang, Y., Gao, T., Liu, C.C., Khan, N., Varki, A., Wang, F., and Cao, H. (2018). Chemoenzymatic Assembly of Mammalian O-Mannose Glycans. Angew Chem Int Ed Engl *57*, 9003-9007.

Merritt, J.H., Ollis, A.A., Fisher, A.C., and DeLisa, M.P. (2013). Glycans-by-design: engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. Biotechnol Bioeng *110*, 1550-1564.

Mescher, M.F., and Strominger, J.L. (1976). Purification and characterization of a prokaryotic glycoprotein from the cell envelope of Halobacterium salinarium. J Biol Chem *251*, 2005-2014.

Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Dos Santos, A., Devos, S., Rousseau, F., Plets, E., Houthuys, E., *et al.* (2014). GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins. Nat Biotechnol *32*, 485-489.

Meusser, B., Hirsch, C., Jarosch, E., and Sommer, T. (2005). ERAD: the long road to destruction. Nat Cell Biol *7*, 766-772.

Meyer, A., Saaem, I., Silverman, A., Varaljay, V.A., Mickol, R., Blum, S., Tobias, A.V., Schwalm, N.D., 3rd, Mojadedi, W., Onderko, E., *et al.* (2019). Organism Engineering for the Bioproduction of the Triaminotrinitrobenzene (TATB) Precursor Phloroglucinol (PG). ACS Synth Biol *8*, 2746-2755.

Mikami, S., Kobayashi, T., Yokoyama, S., and Imataka, H. (2006). A hybridoma-based in vitro translation system that efficiently synthesizes glycoproteins. J Biotechnol *127*, 65-78.

Mikolajczyk, K., Kaczmarek, R., and Czerwinski, M. (2020). How glycosylation affects glycosylation: the role of N-glycans in glycosyltransferase activity. Glycobiology *30*, 941-969.

Mills, D.C., Jervis, A.J., Abouelhadid, S., Yates, L.E., Cuccui, J., Linton, D., and Wren, B.W. (2016). Functional analysis of N-linking oligosaccharyl transferase enzymes encoded by deep-sea vent proteobacteria. Glycobiology *26*, 398-409.

Moeller, T.D., Weyant, K.B., and DeLisa, M.P. (2018). Interplay of Carbohydrate and Carrier in Antibacterial Glycoconjugate Vaccines. Adv Biochem Eng Biotechnol.

Montero-Morales, L., and Steinkellner, H. (2018). Advanced Plant-Based Glycan Engineering. Front Bioeng Biotechnol *6*, 81.

Moore, S.J., MacDonald, J.T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., Kylilis, N., Bell, D.J., McClymont, D.W., Jensen, K., *et al.* (2018). Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. Proc Natl Acad Sci U S A *115*, E4340-E4349.

Moremen, K.W., Tiemeyer, M., and Nairn, A.V. (2012). Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol *13*, 448-462.

Moreno, S.N., Ip, H.S., and Cross, G.A. (1991). An mRNA-dependent in vitro translation system from Trypanosoma brucei. Mol Biochem Parasitol *46*, 265-274.

Murphy, T.W., Sheng, J., Naler, L.B., Feng, X., and Lu, C. (2019). On-chip manufacturing of synthetic proteins for point-of-care therapeutics. Microsyst Nanoeng *5*, 13.

Musumeci, M.A., Hug, I., Scott, N.E., Ielmini, M.V., Foster, L.J., Wang, P.G., and Feldman, M.F. (2013a). In vitro activity of Neisseria meningitidis PglL O-oligosaccharyltransferase with diverse synthetic lipid donors and a UDP-activated sugar. J Biol Chem *288*, 10578-10587.

Musumeci, M.A., Ielmini, M.V., and Feldman, M.F. (2013b). In Vitro Glycosylation Assay for Bacterial Oligosaccharyltransferases. Methods in molecular biology (Clifton, NJ) *1022*, 161-171.

Naegeli, A., Michaud, G., Schubert, M., Lin, C.W., Lizak, C., Darbre, T., Reymond, J.L., and Aebi, M. (2014a). Substrate specificity of cytoplasmic N-glycosyltransferase. J Biol Chem *289*, 24521-24532.

Naegeli, A., Neupert, C., Fan, Y.Y., Lin, C.W., Poljak, K., Papini, A.M., Schwarz, F., and Aebi, M. (2014b). Molecular analysis of an alternative N-glycosylation machinery by functional transfer from Actinobacillus pleuropneumoniae to Escherichia coli. J Biol Chem *289*, 2170-2179.

Nagorny, P., Sane, N., Fasching, B., Aussedat, B., and Danishefsky, S.J. (2012). Probing the frontiers of glycoprotein synthesis: the fully elaborated beta-subunit of the human follicle-stimulating hormone. Angew Chem Int Ed Engl *51*, 975-979.

Napiorkowska, M., Boilevin, J., Sovdat, T., Darbre, T., Reymond, J.L., Aebi, M., and Locher, K.P. (2017). Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. Nat Struct Mol Biol *24*, 1100-1106.

Narimatsu, Y., Joshi, H.J., Nason, R., Van Coillie, J., Karlsson, R., Sun, L., Ye, Z., Chen, Y.H., Schjoldager, K.T., Steentoft, C., *et al.* (2019). An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. Mol Cell *75*, 394-407 e395.

Narimatsu, Y., Joshi, H.J., Yang, Z., Gomes, C., Chen, Y.H., Lorenzetti, F.C., Furukawa, S., Schjoldager, K.T., Hansen, L., Clausen, H., *et al.* (2018). A validated gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferase genome. Glycobiology *28*, 295-305.

Natarajan, A., Jaroentomeechai, T., Cabrera, M., Mohammed, J., Cox, E.C., Young, O., Shajahan, A., Vilkhovoy, M., Vadhin, S., Varner, J.D., *et al.* (2020). Engineering orthogonal O-linked glycoprotein biosynthesis in bacteria yields human tumor-associated glycotopes. *(in review)*.

Natarajan, A., Jaroentomeechai, T., Li, M., Glasscock, Cameron J. , DeLisa, Matthew P. (2018). Metabolic engineering of glycoprotein biosynthesis in bacteria Emerging Topics in Life Sciences *2*, 419-432.

Needham, B.D., Carroll, S.M., Giles, D.K., Georgiou, G., Whiteley, M., and Trent, M.S. (2013). Modulating the innate immune response by combinatorial engineering of endotoxin. Proc Natl Acad Sci U S A *110*, 1464-1469.

Neelamegham, S., and Liu, G. (2011). Systems glycobiology: biochemical reaction networks regulating glycan structure and function. Glycobiology *21*, 1541-1553.

Neuberger, A. (1938). Carbohydrates in protein: The carbohydrate component of crystalline egg albumin. Biochem J *32*, 1435-1451.

Nirenberg, M.W., and Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A *47*, 1588-1602.

Nothaft, H., and Szymanski, C.M. (2010). Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol *8*, 765-778.

Novak, R.T., Kambou, J.L., Diomande, F.V., Tarbangdo, T.F., Ouedraogo-Traore, R., Sangare, L., Lingani, C., Martin, S.W., Hatcher, C., Mayer, L.W., *et al.* (2012). Serogroup A meningococcal conjugate vaccination in Burkina Faso: analysis of national surveillance data. Lancet Infect Dis *12*, 757-764.

O'Kane, P.T., Dudley, Q.M., McMillan, A.K., Jewett, M.C., and Mrksich, M. (2019). High-throughput mapping of CoA metabolites by SAMDI-MS to optimize the cell-free biosynthesis of HMG-CoA. Sci Adv *5*, eaaw9180.

Ohtsubo, K., and Marth, J.D. (2006). Glycosylation in cellular mechanisms of health and disease. Cell *126*, 855-867.

Ohuchi, T., Ikeda-Araki, A., Watanabe-Sakamoto, A., Kojiri, K., Nagashima, M., Okanishi, M., and Suda, H. (2000). Cloning and expression of a gene encoding N-glycosyltransferase (ngt) from Saccarothrix aerocolonigenes ATCC39243. J Antibiot (Tokyo) *53*, 393-403.

Ollis, A.A., Chai, Y., Natarajan, A., Perregaux, E., Jaroentomeechai, T., Guarino, C., Smith, J., Zhang, S., and DeLisa, M.P. (2015a). Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. Scientific Reports *5*, 15237.

Ollis, A.A., Chai, Y., Natarajan, A., Perregaux, E., Jaroentomeechai, T., Guarino, C., Smith, J., Zhang, S., and DeLisa, M.P. (2015b). Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. Sci Rep *5*, 15237.

Ollis, A.A., Chai, Y., Natarajan, A., Perregaux, E., Jaroentomeechai, T., Guarino, C., Smith, J., Zhang, S., and DeLisa, M.P. (2015c). Substitute sweeteners: Diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. Sci Rep *5*, 15237.

Ollis, A.A., Zhang, S., Fisher, A.C., and DeLisa, M.P. (2014a). Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. Nat Chem Biol *10*, 816-822.

Ollis, A.A., Zhang, S., Fisher, A.C., and DeLisa, M.P. (2014b). Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. Nat Chem Biol *10*, 816-822.

Oyston, P.C., Sjostedt, A., and Titball, R.W. (2004). Tularaemia: bioterrorism defence renews interest in Francisella tularensis. Nat Rev Microbiol *2*, 967-978.

Oza, J.P., Aerni, H.R., Pirman, N.L., Barber, K.W., ter Haar, C.M., Rogulina, S., Amrofell, M.B., Isaacs, F.J., Rinehart, J., and Jewett, M.C. (2015a). Robust production of recombinant phosphoproteins using cell-free protein synthesis. Nature Communications *6*, 8168.

Oza, J.P., Aerni, H.R., Pirman, N.L., Barber, K.W., Ter Haar, C.M., Rogulina, S., Amrofell, M.B., Isaacs, F.J., Rinehart, J., and Jewett, M.C. (2015b). Robust production of recombinant phosphoproteins using cell-free protein synthesis. Nat Commun *6*, 8168.

Palma, A.S., Feizi, T., Childs, R.A., Chai, W., and Liu, Y. (2014). The neoglycolipid (NGL)-based oligosaccharide microarray system poised to decipher the meta-glycome. Curr Opin Chem Biol *18*, 87-94.

Pan, C., Sun, P., Liu, B., Liang, H., Peng, Z., Dong, Y., Wang, D., Liu, X., Wang, B., Zeng, M., *et al.* (2016). Biosynthesis of Conjugate Vaccines Using an O-Linked Glycosylation System. mBio *7*, 16.

Pardee, K., Green, A.A., Ferrante, T., Cameron, D.E., DaleyKeyser, A., Yin, P., and Collins, J.J. (2014). Paper-based synthetic gene networks. Cell *159*, 940-954.

Pardee, K., Slomovic, S., Nguyen, P.Q., Lee, J.W., Donghia, N., Burrill, D., Ferrante, T., McSorley, F.R., Furuta, Y., Vernet, A., *et al.* (2016a). Portable, on-demand biomolecular manufacturing. Cell *167*, 248-259 e212.

Pardee, K., Slomovic, S., Nguyen, Peter Q., Lee, Jeong W., Donghia, N., Burrill, D., Ferrante, T., McSorley, Fern R., Furuta, Y., Vernet, A., *et al.* (2016b). Portable, on-demand biomolecular manufacturing. Cell *167*, 248-259.e212.

Parsons, T.B., Struwe, W.B., Gault, J., Yamamoto, K., Taylor, T.A., Raj, R., Wals, K., Mohammed, S., Robinson, C.V., Benesch, J.L., *et al.* (2016). Optimal Synthetic Glycosylation of a Therapeutic Antibody. Angew Chem Int Ed Engl *55*, 2361-2367.

Peixoto, A., Relvas-Santos, M., Azevedo, R., Santos, L.L., and Ferreira, J.A. (2019). Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks. Front Oncol *9*, 380.

Perez-Pinera, P., Han, N., Cleto, S., Cao, J., Purcell, O., Shah, K.A., Lee, K., Ram, R., and Lu, T.K. (2016). Synthetic biology and microbioreactor platforms for programmable production of biologics at the point-of-care. Nat Commun *7*, 12211.

Perez, J.G., Stark, J.C., and Jewett, M.C. (2016a). Cell-free synthetic biology: Engineering beyond the cell. Cold Spring Harb Perspect Biol.

Perez, J.G., Stark, J.C., and Jewett, M.C. (2016b). Cell-Free Synthetic Biology: Engineering Beyond the Cell. Cold Spring Harb Perspect Biol *8*.

Petsch, D., and Anspach, F.B. (2000). Endotoxin removal from protein solutions. J Biotechnol *76*, 97-119.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem *25*, 1605-1612.

Pinho, S.S., and Reis, C.A. (2015). Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer *15*, 540-555.

Piontek, C., Ring, P., Harjes, O., Heinlein, C., Mezzato, S., Lombana, N., Pohner, C., Puttner, M., Varon Silva, D., Martin, A., *et al.* (2009a). Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 1. Angew Chem Int Ed Engl *48*, 1936-1940.

Piontek, C., Varon Silva, D., Heinlein, C., Pohner, C., Mezzato, S., Ring, P., Martin, A., Schmid, F.X., and Unverzagt, C. (2009b). Semisynthesis of a homogeneous glycoprotein enzyme: ribonuclease C: part 2. Angew Chem Int Ed Engl *48*, 1941-1945.

Plante, O.J., Palmacci, E.R., and Seeberger, P.H. (2001). Automated solid-phase synthesis of oligosaccharides. Science *291*, 1523-1527.

Poehling, K.A., Talbot, T.R., Griffin, M.R., Craig, A.S., Whitney, C.G., Zell, E., Lexau, C.A., Thomas, A.R., Harrison, L.H., Reingold, A.L., *et al.* (2006). Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine. Jama *295*, 1668-1674.

Posey, A.D., Jr., Schwab, R.D., Boesteanu, A.C., Steentoft, C., Mandel, U., Engels, B., Stone, J.D., Madsen, T.D., Schreiber, K., Haines, K.M., *et al.* (2016). Engineered CAR T Cells Targeting the Cancer-Associated Tn-Glycoform of the Membrane Mucin MUC1 Control Adenocarcinoma. Immunity *44*, 1444-1454.

Power, P.M., Roddam, L.F., Dieckelmann, M., Srikhanta, Y.N., Cheng Tan, Y., Berrington, A.W., and Jennings, M.P. (2000). Genetic characterization of pilin glycosylation in Neisseria meningitidis. Microbiology (Reading) *146 (Pt 4)*, 967-979.

Prior, J.L., Prior, R.G., Hitchen, P.G., Diaper, H., Griffin, K.F., Morris, H.R., Dell, A., and Titball, R.W. (2003). Characterization of the O antigen gene cluster and structural analysis of the O antigen of Francisella tularensis subsp. tularensis. J Med Microbiol *52*, 845-851.

Priyanka, P., Parsons, T.B., Miller, A., Platt, F.M., and Fairbanks, A.J. (2016). Chemoenzymatic Synthesis of a Phosphorylated Glycoprotein. Angew Chem Int Ed Engl *55*, 5058-5061.

Prudden, A.R., Liu, L., Capicciotti, C.J., Wolfert, M.A., Wang, S., Gao, Z., Meng, L., Moremen, K.W., and Boons, G.J. (2017). Synthesis of asymmetrical multiantennary human milk oligosaccharides. Proc Natl Acad Sci U S A *114*, 6954-6959.

Qadri, F., Svennerholm, A.M., Faruque, A.S., and Sack, R.B. (2005). Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev *18*, 465-483.

Raetz, C.R., and Whitfield, C. (2002). Lipopolysaccharide endotoxins. Annu Rev Biochem *71*, 635-700.

Raman, R., Raguram, S., Venkataraman, G., Paulson, J.C., and Sasisekharan, R. (2005). Glycomics: an integrated systems approach to structure-function relationships of glycans. Nat Methods *2*, 817-824.

Ramirez, A.S., Boilevin, J., Biswas, R., Gan, B.H., Janser, D., Aebi, M., Darbre, T., Reymond, J.L., and Locher, K.P. (2017). Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. Glycobiology *27*, 525-535.

Ramirez, A.S., Boilevin, J., Mehdipour, A.R., Hummer, G., Darbre, T., Reymond, J.L., and Locher, K.P. (2018). Structural basis of the molecular ruler mechanism of a bacterial glycosyltransferase. Nat Commun *9*, 445.

Ramirez, A.S., Kowal, J., and Locher, K.P. (2019). Cryo-electron microscopy structures of human oligosaccharyltransferase complexes OST-A and OST-B. Science *366*, 1372-1375.

Reily, C., Stewart, T.J., Renfrow, M.B., and Novak, J. (2019). Glycosylation in health and disease. Nat Rev Nephrol *15*, 346-366.

Riddle, M.S., Kaminski, R.W., Di Paolo, C., Porter, C.K., Gutierrez, R.L., Clarkson, K.A., Weerts, H.E., Duplessis, C., Castellano, A., Alaimo, C., *et al.* (2016). Safety and immunogenicity of a candidate bioconjugate vaccine against Shigella flexneri 2a administered to healthy adults: a single blind, randomized phase I study. Clin Vaccine Immunol.

Rillahan, C.D., and Paulson, J.C. (2011). Glycan microarrays for decoding the glycome. Annu Rev Biochem *80*, 797-823.

Rothblatt, J.A., and Meyer, D.I. (1986). Secretion in yeast: reconstitution of the translocation and glycosylation of alpha-factor and invertase in a homologous cell-free system. Cell *44*, 619-628.

Roush, S.W., McIntyre, L., and Baldy, L.M. (2008). Manual for the surveillance of vaccine-preventable diseases. Atlanta: Centers for Disease Control and Prevention, 4.

Rudd, P.M., and Dwek, R.A. (1997). Glycosylation: heterogeneity and the 3D structure of proteins. Critical reviews in biochemistry and molecular biology *32*, 1-100.

Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A., and Dwek, R.A. (2001). Glycosylation and the immune system. Science *291*, 2370-2376.

Russell, J.A. (2006). Management of sepsis. N Engl J Med *355*, 1699-1713.

Sakamoto, I., Tezuka, K., Fukae, K., Ishii, K., Taduru, K., Maeda, M., Ouchi, M., Yoshida, K., Nambu, Y., Igarashi, J., *et al.* (2012). Chemical synthesis of homogeneous human glycosyl-interferon-beta that exhibits potent antitumor activity in vivo. J Am Chem Soc *134*, 5428-5431.

Salehi, A.S., Smith, M.T., Bennett, A.M., Williams, J.B., Pitt, W.G., and Bundy, B.C. (2016a). Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system. Biotechnol J *11*, 274-281.

Salehi, A.S., Smith, M.T., Bennett, A.M., Williams, J.B., Pitt, W.G., and Bundy, B.C. (2016b). Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system. Biotechnol J *11*, 274-281.

Schaffer, C., and Messner, P. (2017). Emerging facets of prokaryotic glycosylation. FEMS Microbiol Rev *41*, 49-91.

Schenk, B., Fernandez, F., and Waechter, C.J. (2001). The ins(ide) and out(side) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. Glycobiology *11*, 61R-70R.

Schjoldager, K.T., Narimatsu, Y., Joshi, H.J., and Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. Nat Rev Mol Cell Biol *21*, 729-749.

Schoborg, J.A., Hershewe, J.M., Stark, J.C., Kightlinger, W., Kath, J.E., Jaroentomeechai, T., Natarajan, A., DeLisa, M.P., and Jewett, M.C. (2017). A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. Biotechnol Bioeng.

Schoborg, J.A., Hershewe, J.M., Stark, J.C., Kightlinger, W., Kath, J.E., Jaroentomeechai, T., Natarajan, A., DeLisa, M.P., and Jewett, M.C. (2018). A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. Biotechnol Bioeng *115*, 739-750.

Schoenhofen, I.C., McNally, D.J., Vinogradov, E., Whitfield, D., Young, N.M., Dick, S., Wakarchuk, W.W., Brisson, J.R., and Logan, S.M. (2006). Functional characterization of dehydratase/aminotransferase pairs from Helicobacter and Campylobacter: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. J Biol Chem *281*, 723-732.

Schulz, M.A., Tian, W., Mao, Y., Van Coillie, J., Sun, L., Larsen, J.S., Chen, Y.H., Kristensen, C., Vakhrushev, S.Y., Clausen, H., *et al.* (2018). Glycoengineering design options for IgG1 in CHO cells using precise gene editing. Glycobiology *28*, 542-549.

Schwarz, F., Fan, Y.Y., Schubert, M., and Aebi, M. (2011a). Cytoplasmic N-glycosyltransferase of Actinobacillus pleuropneumoniae is an inverting enzyme and recognizes the NX(S/T) consensus sequence. The Journal of biological chemistry *286*, 35267-35274.

Schwarz, F., Huang, W., Li, C., Schulz, B.L., Lizak, C., Palumbo, A., Numao, S., Neri, D., Aebi, M., and Wang, L.X. (2010). A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat Chem Biol *6*, 264-266.

Schwarz, F., Lizak, C., Fan, Y.Y., Fleurkens, S., Kowarik, M., and Aebi, M. (2011b). Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. Glycobiology *21*, 45-54.

Shang, W., Zhai, Y., Ma, Z., Yang, G., Ding, Y., Han, D., Li, J., Zhang, H., Liu, J., Wang, P.G., *et al.* (2016). 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.

Sharma, C.B., Lehle, L., and Tanner, W. (1981). N-Glycosylation of yeast proteins. Characterization of the solubilized oligosaccharyl transferase. Eur J Biochem *116*, 101-108.

Shental-Bechor, D., and Levy, Y. (2009). Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. Current Opinion in Structural Biology *19*.

Shibutani, M., Kim, E., Lazarovici, P., Oshima, M., and Guroff, G. (1996). Preparation of a cell-free translation system from PC12 cell. Neurochem Res *21*, 801-807.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. Nature biotechnology *19*, 751-755.

Shimizu, Y., Kanamori, T., and Ueda, T. (2005). Protein synthesis by pure translation systems. Methods (San Diego, Calif) *36*, 299-304.

Silverman, A.D., Karim, A.S., and Jewett, M.C. (2020). Cell-free gene expression: an expanded repertoire of applications. Nat Rev Genet *21*, 151-170.

Silverman, A.D., Kelley-Loughnane, N., Lucks, J.B., and Jewett, M.C. (2019). Deconstructing Cell-Free Extract Preparation for in Vitro Activation of Transcriptional Genetic Circuitry. ACS Synth Biol *8*, 403-414.

Sinclair, A.M., and Elliott, S. (2005). Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci *94*, 1626-1635.

Skropeta, D. (2009). The effect of individual N-glycans on enzyme activity. Bioorg Med Chem *17*, 2645-2653.

Sleytr, U.B., and Thorne, K.J. (1976). Chemical characterization of the regularly arranged surface layers of Clostridium thermosaccharolyticum and Clostridium thermohydrosulfuricum. J Bacteriol *126*, 377-383.

Son, Y.D., Jeong, Y.T., Park, S.Y., and Kim, J.H. (2011). Enhanced sialylation of recombinant human erythropoietin in Chinese hamster ovary cells by combinatorial engineering of selected genes. Glycobiology *21*, 1019-1028.

Sorensen, A.L., Reis, C.A., Tarp, M.A., Mandel, U., Ramachandran, K., Sankaranarayanan, V., Schwientek, T., Graham, R., Taylor-Papadimitriou, J., Hollingsworth, M.A., *et al.* (2006). Chemoenzymatically synthesized multimeric Tn/STn MUC1 glycopeptides elicit cancer-specific anti-MUC1 antibody responses and override tolerance. Glycobiology *16*, 96-107.

Spiro, R.G. (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology *12*, 43R-56R.

Srichaisupakit, A., Ohashi, T., Misaki, R., and Fujiyama, K. (2015). Production of initial-stage eukaryotic N-glycan and its protein glycosylation in Escherichia coli. J Biosci Bioeng *119*, 399-405.

Srinivasan, A., and Coward, J.K. (2002). A biotin capture assay for oligosaccharyltransferase. Anal Biochem *306*, 328-335.

Stapleton, J.A., and Swartz, J.R. (2010). Development of an in vitro compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases. PLoS One *5*, e15275.

Stark, J.C., Jaroentomeechai, T., Moeller, T.D., Dubner, R.S., Hsu, K.J., Stevenson, T.C., DeLisa, M.P., and Jewett, M.C. (2019). On-demand, cell-free biomanufacturing of conjugate vaccines at the point-of-care. bioRxiv.

Stech, M., Nikolaeva, O., Thoring, L., Stocklein, W.F.M., Wustenhagen, D.A., Hust, M., Dubel, S., and Kubick, S. (2017). Cell-free synthesis of functional antibodies using a coupled in vitro transcription-translation system based on CHO cell lysates. Sci Rep *7*, 12030.

Steentoft, C., Bennett, E.P., Schjoldager, K.T., Vakhrushev, S.Y., Wandall, H.H., and Clausen, H. (2014). Precision genome editing: a small revolution for glycobiology. Glycobiology *24*, 663-680.

Steentoft, C., Migliorini, D., King, T.R., Mandel, U., June, C.H., and Posey, A.D., Jr. (2018). Glycan-directed CAR-T cells. Glycobiology *28*, 656-669.

Stefan, A., Conti, M., Rubboli, D., Ravagli, L., Presta, E., and Hochkoeppler, A. (2011). Overexpression and purification of the recombinant diphtheria toxin variant CRM197 in Escherichia coli. J Biotechnol *156*, 245-252.

Stefanetti, G., Okan, N., Fink, A., Gardner, E., and Kasper, D.L. (2019). Glycoconjugate vaccine using a genetically modified O antigen induces protective antibodies to Francisella tularensis. Proc Natl Acad Sci U S A *116*, 7062-7070.

Stubs, G., Rupp, B., Schumann, R.R., Schroder, N.W., and Rademann, J. (2010). Chemoenzymatic synthesis of a glycolipid library and elucidation of the antigenic epitope for construction of a vaccine against Lyme disease. Chemistry *16*, 3536-3544.

Su, Z., Brown, E.C., Wang, W., and MacKinnon, R. (2016). Novel cell-free high-throughput screening method for pharmacological tools targeting K+ channels. Proc Natl Acad Sci U S A *113*, 5748-5753.

Sullivan, C.J., Pendleton, E.D., Sasmor, H.H., Hicks, W.L., Farnum, J.B., Muto, M., Amendt, E.M., Schoborg, J.A., Martin, R.W., Clark, L.G., *et al.* (2016). A cell-free expression and purification process for rapid production of protein biologics. Biotechnol J *11*, 238-248.

Sun, Z.Z., Hayes, C.A., Shin, J., Caschera, F., Murray, R.M., and Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. J Vis Exp, e50762.

Swank, Z., Laohakunakorn, N., and Maerkl, S.J. (2019). Cell-free gene-regulatory network engineering with synthetic transcription factors. Proc Natl Acad Sci U S A *116*, 5892-5901.

Szymanski, C.M., Michael, F.S., Jarrell, H.C., Li, J., Gilbert, M., Larocque, S., Vinogradov, E., and Brisson, J.R. (2003). Detection of conserved N-linked glycans and phase-variable lipooligosaccharides and capsules from campylobacter cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. J Biol Chem *278*, 24509-24520.

Szymanski, C.M., and Wren, B.W. (2005). Protein glycosylation in bacterial mucosal pathogens. Nat Rev Microbiol *3*, 225-237.

Szymanski, C.M., Yao, R., Ewing, C.P., Trust, T.J., and Guerry, P. (1999). Evidence for a system of general protein glycosylation in Campylobacter jejuni. Mol Microbiol *32*, 1022-1030.

Tai, V.W., and Imperiali, B. (2001). Substrate specificity of the glycosyl donor for oligosaccharyl transferase. J Org Chem *66*, 6217-6228.

Takahashi, M.K., Hayes, C.A., Chappell, J., Sun, Z.Z., Murray, R.M., Noireaux, V., and Lucks, J.B. (2015). Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. Methods *86*, 60-72.

Takahashi, M.K., Tan, X., Dy, A.J., Braff, D., Akana, R.T., Furuta, Y., Donghia, N., Ananthakrishnan, A., and Collins, J.J. (2018). A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. Nat Commun *9*, 3347.

Takegawa, K., Tabuchi, M., Yamaguchi, S., Kondo, A., Kato, I., and Iwahara, S. (1995). Synthesis of neoglycoproteins using oligosaccharide-transfer activity with endo-beta-N-acetylglucosaminidase. J Biol Chem *270*, 3094-3099.

Tan, F.Y., Tang, C.M., and Exley, R.M. (2015). Sugar coating: bacterial protein glycosylation and host-microbe interactions. Trends Biochem Sci *40*, 342-350.

Tarui, H., Imanishi, S., and Hara, T. (2000). A novel cell-free translation/glycosylation system prepared from insect cells. J Biosci Bioeng *90*, 508-514.

Taylor-Papadimitriou, J., Burchell, J., Miles, D.W., and Dalziel, M. (1999). MUC1 and cancer. Biochim Biophys Acta *1455*, 301-313.

Tejwani, V., Andersen, M.R., Nam, J.H., and Sharfstein, S.T. (2018). Glycoengineering in CHO Cells: Advances in Systems Biology. Biotechnol J *13*, e1700234.

Terra, V.S., Mills, D.C., Yates, L.E., Abouelhadid, S., Cuccui, J., and Wren, B.W. (2012). Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design. J Med Microbiol *61*, 919-926.

Thavarajah, W., Silverman, A.D., Verosloff, M.S., Kelley-Loughnane, N., Jewett, M.C., and Lucks, J.B. (2020). Point-of-Use Detection of Environmental Fluoride via a Cell-Free Riboswitch-Based Biosensor. ACS Synth Biol *9*, 10-18.

Thaysen-Andersen, M., and Packer, N.H. (2012). Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. Glycobiology *22*, 1440-1452.

The Review on Antimicrobial Resistance, C.b.J.O.N. (2014). Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations.

Thoring, L., Dondapati, S.K., Stech, M., Wustenhagen, D.A., and Kubick, S. (2017). High-yield production of "difficult-to-express" proteins in a continuous exchange cell-free system based on CHO cell lysates. Sci Rep *7*, 11710.

Tomabechi, Y., Suzuki, R., Haneda, K., and Inazu, T. (2010). Chemo-enzymatic synthesis of glycosylated insulin using a GlcNAc tag. Bioorg Med Chem *18*, 1259-1264.

Toth, A.M., Kuo, C.W., Khoo, K.H., and Jarvis, D.L. (2014). A new insect cell glycoengineering approach provides baculovirus-inducible glycogene expression and increases human-type glycosylation efficiency. J Biotechnol *182-183*, 19-29.

Trotter, C.L., McVernon, J., Ramsay, M.E., Whitney, C.G., Mulholland, E.K., Goldblatt, D., Hombach, J., and Kieny, M.P. (2008). Optimising the use of conjugate vaccines to prevent disease caused by Haemophilus influenzae type b, Neisseria meningitidis and Streptococcus pneumoniae. Vaccine *26*, 4434-4445.

Tytgat, H.L.P., Lin, C.W., Levasseur, M.D., Tomek, M.B., Rutschmann, C., Mock, J., Liebscher, N., Terasaka, N., Azuma, Y., Wetter, M., *et al.* (2019). Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. Nat Commun *10*, 5403.

Umekawa, M., Huang, W., Li, B., Fujita, K., Ashida, H., Wang, L.X., and Yamamoto, K. (2008). Mutants of Mucor hiemalis endo-beta-N-acetylglucosaminidase show enhanced transglycosylation and glycosynthase-like activities. J Biol Chem *283*, 4469-4479.

Valderrama-Rincon, J.D., Fisher, A.C., Merritt, J.H., Fan, Y.Y., Reading, C.A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., and DeLisa, M.P. (2012). An engineered eukaryotic protein glycosylation pathway in Escherichia coli. Nat Chem Biol *8*, 434-436.

Valvano, M.A., and Crosa, J.H. (1989). Molecular cloning and expression in Escherichia coli K-12 of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of E. coli O7:K1. Infect Immun *57*, 937-943.

Van den Steen, P., Rudd, P.M., Dwek, R.A., and Opdenakker, G. (1998). Concepts and principles of O-linked glycosylation. Crit Rev Biochem Mol Biol *33*, 151-208.

Varki, A. (2017a). Biological roles of glycans. Glycobiology *27*, 3-49.

Varki, A. (2017b). Essentials of glycobiology, Third edition. edn (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Venturi, G., Ferreira, I.G., Pucci, M., Ferracin, M., Malagolini, N., Chiricolo, M., and Dall'Olio, F. (2019). Impact of sialyltransferase ST6GAL1 overexpression on different colon cancer cell types. Glycobiology.

Wacker, M., Feldman, M.F., Callewaert, N., Kowarik, M., Clarke, B.R., Pohl, N.L., Hernandez, M., Vines, E.D., Valvano, M.A., Whitfield, C., *et al.* (2006). Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. Proceedings of the National Academy of Sciences of the United States of America *103*, 7088-7093.

Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., *et al.* (2002). N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. Science *298*, 1790-1793.

Wacker, M., Wang, L., Kowarik, M., Dowd, M., Lipowsky, G., Faridmoayer, A., Shields, K., Park, S., Alaimo, C., Kelley, K.A., *et al.* (2014). Prevention of Staphylococcus aureus

infections by glycoprotein vaccines synthesized in Escherichia coli. J Infect Dis *209*, 1551-1561.

Wahl, B., O'Brien, K.L., Greenbaum, A., Majumder, A., Liu, L., Chu, Y., Luksic, I., Nair, H., McAllister, D.A., Campbell, H., *et al.* (2018). Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15. Lancet Glob Health *6*, e744-e757.

Walsh, G. (2018). Biopharmaceutical benchmarks 2018. Nat Biotechnol *36*, 1136-1145.

Walt, D. (2012). Transforming Glycoscience: A Roadmap for the Future. In Transforming Glycoscience: A Roadmap for the Future, D. Walt, ed. (Washington (DC): National Academies Press (US)).

Wang, L.X., and Davis, B.G. (2013). Realizing the Promise of Chemical Glycobiology. Chem Sci *4*, 3381-3394.

Wang, L.X., and Lomino, J.V. (2012). Emerging technologies for making glycan-defined glycoproteins. ACS Chem Biol *7*, 110-122.

Wang, P., Dong, S., Shieh, J.H., Peguero, E., Hendrickson, R., Moore, M.A.S., and Danishefsky, S.J. (2013a). Erythropoietin derived by chemical synthesis. Science *342*, 1357-1360.

Wang, S., Zhang, Q., Chen, C., Guo, Y., Gadi, M.R., Yu, J., Westerlind, U., Liu, Y., Cao, X., Wang, P.G., *et al.* (2018). Facile Chemoenzymatic Synthesis of O-Mannosyl Glycans. Angew Chem Int Ed Engl *57*, 9268-9273.

Wang, Z., Chinoy, Z.S., Ambre, S.G., Peng, W., McBride, R., de Vries, R.P., Glushka, J., Paulson, J.C., and Boons, G.J. (2013b). A general strategy for the chemoenzymatic synthesis of asymmetrically branched N-glycans. Science *341*, 379-383.

Weerapana, E., Glover, K.J., Chen, M.M., and Imperiali, B. (2005). Investigating bacterial N-linked glycosylation: Synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. Journal of the American Chemical Society *127*, 13766-13767.

Weerapana, E., and Imperiali, B. (2006). Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. Glycobiology *16*, 91R-101R.

Weikert, S., Papac, D., Briggs, J., Cowfer, D., Tom, S., Gawlitzek, M., Lofgren, J., Mehta, S., Chisholm, V., Modi, N., *et al.* (1999). Engineering Chinese hamster ovary cells to

maximize sialic acid content of recombinant glycoproteins. Nat Biotechnol *17*, 1116-1121.

Weintraub, A. (2003). Immunology of bacterial polysaccharide antigens. Carbohydr Res *338*, 2539-2547.

Wen, L., Edmunds, G., Gibbons, C., Zhang, J., Gadi, M.R., Zhu, H., Fang, J., Liu, X., Kong, Y., and Wang, P.G. (2018). Toward Automated Enzymatic Synthesis of Oligosaccharides. Chem Rev *118*, 8151-8187.

Wetter, M., Kowarik, M., Steffen, M., Carranza, P., Corradin, G., and Wacker, M. (2013). Engineering, conjugation, and immunogenicity assessment of Escherichia coli O121 O antigen for its potential use as a typhoid vaccine component. Glycoconj J *30*, 511-522.

WHO (2014). Temperature Sensitivity of Vaccines.

Wild, R., Kowal, J., Eyring, J., Ngwa, E.M., Aebi, M., and Locher, K.P. (2018). Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. Science.

Wildt, S., and Gerngross, T.U. (2005). The humanization of N-glycosylation pathways in yeast. Nat Rev Microbiol *3*, 119-128.

Wolfert, M.A., and Boons, G.J. (2013). Adaptive immune activation: glycosylation does matter. Nat Chem Biol *9*, 776-784.

Xiao, Z., Guo, Y., Liu, Y., Li, L., Zhang, Q., Wen, L., Wang, X., Kondengaden, S.M., Wu, Z., Zhou, J., *et al.* (2016). Chemoenzymatic Synthesis of a Library of Human Milk Oligosaccharides. J Org Chem *81*, 5851-5865.

Xu, T., and Coward, J.K. (1997). 13C- and 15N-labeled peptide substrates as mechanistic probes of oligosaccharyltransferase. Biochemistry *36*, 14683-14689.

Yan, Q., and Lennarz, W.J. (2002). Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. J Biol Chem *277*, 47692-47700.

Yang, X., and Qian, K. (2017). Protein O-GlcNAcylation: emerging mechanisms and functions. Nat Rev Mol Cell Biol *18*, 452-465.

Yang, Z., Wang, S., Halim, A., Schulz, M.A., Frodin, M., Rahman, S.H., Vester-Christensen, M.B., Behrens, C., Kristensen, C., Vakhrushev, S.Y., *et al.* (2015). Engineered CHO cells for production of diverse, homogeneous glycoproteins. Nat Biotechnol *33*, 842-844.

Yin, G., Garces, E.D., Yang, J., Zhang, J., Tran, C., Steiner, A.R., Roos, C., Bajad, S., Hudak, S., Penta, K., *et al.* (2012). Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. mAbs *4*, 217-225.

Yoshida-Moriguchi, T., and Campbell, K.P. (2015). Matriglycan: a novel polysaccharide that links dystroglycan to the basement membrane. Glycobiology *25*, 702-713.

Young, N.M., Brisson, J.R., Kelly, J., Watson, D.C., Tessier, L., Lanthier, P.H., Jarrell, H.C., Cadotte, N., St Michael, F., Aberg, E., *et al.* (2002). Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. J Biol Chem *277*, 42530-42539.

Yu, H., and Chen, X. (2016). One-pot multienzyme (OPME) systems for chemoenzymatic synthesis of carbohydrates. Org Biomol Chem *14*, 2809-2818.

Yu, H., Li, Y., Zeng, J., Thon, V., Nguyen, D.M., Ly, T., Kuang, H.Y., Ngo, A., and Chen, X. (2016). Sequential One-Pot Multienzyme Chemoenzymatic Synthesis of Glycosphingolipid Glycans. J Org Chem *81*, 10809-10824.

Zawada, J.F., Yin, G., Steiner, A.R., Yang, J., Naresh, A., Roy, S.M., Gold, D.S., Heinsohn, H.G., and Murray, C.J. (2011). Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines. Biotechnology and bioengineering *108*, 1570-1578.

Zhang, J., Chen, C., Gadi, M.R., Gibbons, C., Guo, Y., Cao, X., Edmunds, G., Wang, S., Liu, D., Yu, J., *et al.* (2018). Machine-Driven Enzymatic Oligosaccharide Synthesis by Using a Peptide Synthesizer. Angew Chem Int Ed Engl *57*, 16638-16642.

Zhang, Y., Minagawa, Y., Kizoe, H., Miyazaki, K., Iino, R., Ueno, H., Tabata, K.V., Shimane, Y., and Noji, H. (2019). Accurate high-throughput screening based on digital protein synthesis in a massively parallel femtoliter droplet array. Sci Adv *5*, eaav8185.

Zhou, M., and Wu, H. (2009). Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. Microbiology *155*, 317-327.

Zhu, F., Zhang, H., Yang, T., Haslam, S.M., Dell, A., and Wu, H. (2016). Engineering and Dissecting the Glycosylation Pathway of a Streptococcal Serine-rich Repeat Adhesin. J Biol Chem *291*, 27354-27363.

Zhu, H., Wu, Z., Gadi, M.R., Wang, S., Guo, Y., Edmunds, G., Guan, W., and Fang, J. (2017). Cation exchange assisted binding-elution strategy for enzymatic synthesis of human milk oligosaccharides (HMOs). Bioorg Med Chem Lett *27*, 4285-4287.

CHAPTER 2

SINGLE-POT GLYCOPROTEIN BIOSYNTHESIS USING A CELL-FREE TRANSCRIPTION-TRANSLATION SYSTEM ENRICHED WITH GLYCOSYLATION MACHINERY[6]

2.1 Abstract.

The emerging bacterial glycoengineering discipline has made it possible to produce designer glycans and glycoconjugates for use as vaccines and therapeutics. These developments notwithstanding, cell-based production of homogeneous glycoproteins remains a significant challenge due to the cell viability constraints and the inability to control glycosylation components at precise ratios in vivo. To address these challenges, we describe a novel cell-free glycoprotein synthesis (CFGpS) technology that seamlessly integrates protein biosynthesis with asparagine-linked protein glycosylation. This technology leverages a glyco-optimized Escherichia coli to source cell extracts that are selectively enriched with glycosylation components, including oligosaccharyltransferases (OSTs) and lipid-linked oligosaccharides (LLOs). The resulting extracts enable a one-pot reaction scheme for efficient and site-specific glycosylation of target proteins. CFGpS platform is highly modular, allowing the use of multiple distinct OSTs and structurally diverse LLOs. As such, we anticipate CFGpS

6 This chapter appears in the Nature Communication journal:

Jaroentomeechai, T.*, Stark, J.C.*, Natarajan, A., Glasscock, C.J., Yates, L.E., Hsu, K.J., Mrksich, M., Jewett, M.C. and DeLisa, M.P. (2018) Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat Commun* 9: 2686.

will facilitate fundamental understanding in glycoscience and make possible applications in on-demand biomanufacturing of glycoproteins.

2.2 Introduction.

Asparagine-linked (*N*-linked) protein glycosylation is one of the most common post-translational modifications in eukaryotes, and profoundly affects many important protein properties such as folding, *in vivo* stability, immunogenicity, and pharmacokinetics (Hebert et al., 2014; Helenius and Aebi, 2001a; Imperiali and O'Connor, 1999). The attached *N*-glycans can directly participate in a wide spectrum of biological processes such as immune recognition/response (Rudd et al., 2001; Wolfert and Boons, 2013) and stem cell fate (Lanctot et al., 2007). Moreover, the intentional engineering of protein-associated glycans can be used to manipulate the therapeutic properties of a protein such as enhancing its *in vivo* activity and half-life (Sinclair and Elliott, 2005).

At present, however, the inherent structural complexity of glycans and the corresponding difficulties producing homogeneously glycosylated proteins have slowed advances in our understanding of glycoprotein functions and limited opportunities for biotechnological applications. Indeed, glycan structural diversity and information density is immense (Werz et al., 2007), far exceeding that of DNA or proteins. Moreover, because glycan biosynthesis is neither template-driven nor genetically encoded, glycans cannot be produced from recombinant DNA technology.

Instead, *N*-glycans are naturally made by coordinated expression of multiple glycosyltransferases (GTs) across several subcellular compartments. This mode of biosynthesis combined with the lack of a strict proofreading system results in inherent glycan heterogeneity even at a single glycosylation site, and accounts for the large diversity of structures in the expressed glycan repertoire of a cell or organism (Raman et al., 2005; Rudd and Dwek, 1997). Further complicating matters is the paucity of structure-function relationships for GTs, which hinders the *a priori* prediction of glycan structure. Altogether, these factors have frustrated production of homogeneous glycans and glycoconjugates in biological systems and in turn restricted our capacity to elucidate the biochemical and biophysical effects of glycans on the target proteins to which they are attached. Thus, there is an unmet need for a simple technology capable of rapidly producing useful quantities of proteins featuring user-specified glycosylation for biochemical and structural biology studies.

Recent pioneering efforts in glycoengineering of cellular systems including mammalian (Meuris et al., 2014), yeast (Hamilton et al., 2003), and bacterial cells (Valderrama-Rincon et al., 2012) have expanded our ability to reliably synthesize chemically defined glycans and glycoproteins. Yet despite the promise of these systems, protein expression yields often remain low and design-build-test (DBT) cycles—iterations of re-engineering organisms to test new sets of enzymes—can be slow. One promising alternative to cell-based systems is cell-free protein synthesis (CFPS) in which the synthesis of proteins occurs *in vitro* without using intact, living

cells. Recently, a technical renaissance has revitalized CFPS systems to help meet increasing demands for simple and efficient protein synthesis, with *Escherichia coli*-based CFPS systems now exceeding grams of protein per liter reaction volume (Carlson et al., 2012), with the ability to support co- or post-translational modifications (*e.g.*, non-standard amino acid incorporation (Kiga et al., 2002; Martin et al., 2018; Oza et al., 2015a; Stapleton and Swartz, 2010); incorporation of metal centers (Stapleton and Swartz, 2010)). As a complement to *in vivo* expression systems, cell-free systems offer several potential advantages. First, the open nature of the reaction allows the user to directly influence the biochemical systems of interest. As a result, new components can be added or synthesized, and these can be maintained at precise concentrations (Albayrak and Swartz, 2013; Karim and Jewett, 2016). Second, cell-free systems bypass viability constraints making possible the production of complex proteins at titers that would otherwise be toxic in living cells (Kaiser et al., 2008). Third, processes that take days or weeks to design, prepare, and execute *in vivo* can be done more rapidly in a cell-free system(Dudley et al., 2016; Moore et al., 2018b), leading to high-throughput production campaigns on a whole-proteome scale (Goshima et al., 2008) with the ability to automate (Matsuoka et al., 2010). Indeed, recent efforts have shown the power of CFPS systems in accelerating DBT cycles (Dudley et al., 2016; Moore et al., 2018a).

Unfortunately, CFPS systems have been limited by their inability to co-activate efficient protein synthesis and glycosylation. The best characterized and most widely adopted CFPS systems use crude *E. coli* lysates to activate *in vitro* protein synthesis, but

these systems are incapable of making glycoproteins because *E. coli* lacks endogenous glycosylation machinery. Glycosylation is possible in some eukaryotic CFPS systems, including those prepared from insect cells (Tarui et al., 2000), trypanosomes (Moreno et al., 1991), hybridomas (Mikami et al., 2006), or mammalian cells (Brodel et al., 2013; Shibutani et al., 1996). However, these platforms are limited to endogenous machinery for performing glycosylation, meaning that (i) the possible glycan structures are restricted to those naturally synthesized by the host cells and (ii) the glycosylation process is carried out in a "black box" and thus difficult to engineer or control. Additionally, eukaryotic CFPS systems are technically difficult to prepare, often requiring supplementation with microsomes (Lingappa et al., 1978; Rothblatt and Meyer, 1986), and suffer from inefficient protein synthesis and glycosylation yields due to inefficient trafficking of nascent polypeptide chains to microsomes (Moreno et al., 1991; Rothblatt and Meyer, 1986).

Despite progress in eukaryotic cell-free systems, cell-free extracts from bacteria like *E. coli* offer a blank canvas for studying glycosylation pathways, provided they can be activated *in vitro*. A recent work from our group highlights the ability of CFPS to enable glycoprotein synthesis in bacterial cell-free systems by augmenting commercial *E. coli*-based cell-free translation systems (*e.g.*, New England Biolabs PURExpress, Promega *E. coli* S30 Extract System) with purified components from a bacterial *N*-linked glycosylation pathway (Guarino and DeLisa, 2012b). While these results established the possibility of *E. coli* lysate-based glycoprotein production, there are

several drawbacks of using purified glycosylation components that serve to limit system utility. First, preparation of the key glycosylation components required time-consuming and cost-prohibitive steps, namely purification of a topologically complex, multipass transmembrane oligosaccharyltransferase (OST) enzyme and organic solvent-based extraction of lipid-linked oligosaccharide (LLO) donors from bacterial membranes. These steps significantly lengthen the process development timeline, requiring 3-5 days each for isolation of the LLO and OST components, necessitate skilled operators and specialized equipment, and result in products that must be refrigerated and are stable for only a few months to a year. Second, glycoproteins were produced using a sequential translation/glycosylation strategy, which required 20 h for cell-free synthesis of the glycoprotein target and an additional 12 h for post-translational protein glycosylation.

Here, we addressed these drawbacks by developing an integrated cell-free glycoprotein synthesis (CFGpS) technology that bypasses the need for purification of OSTs and organic solvent-based extraction of LLOs. The creation of this streamlined CFGpS system was made possible by two important discoveries: (i) crude extract prepared from the glyco-optimized *E. coli* strain, CLM24, which lacks two native enzymes that potentially compete with engineered glycan biosynthesis reactions (Feldman et al., 2005a), is able to support cell-free protein expression and *N*-linked glycosylation; and (ii) OST- and LLO-enriched extracts derived from this CLM24 chassis strain are able to reproducibly co-activate protein synthesis and *N*-

glycosylation in a reaction mixture that minimally requires priming with DNA encoding the target glycoprotein of interest. Importantly, the CFGpS system decouples production of glycoprotein synthesis components (*i.e.*, OSTs, LLOs, translational machinery) and the glycoprotein target of interest, providing significantly reduced cell viability constraints compared to *in vivo* systems. The net result is a one-pot bacterial glycoprotein biosynthesis platform whereby different acceptor proteins, OSTs, and/or oligosaccharide structures can be functionally interchanged and prototyped for customizable glycosylation.

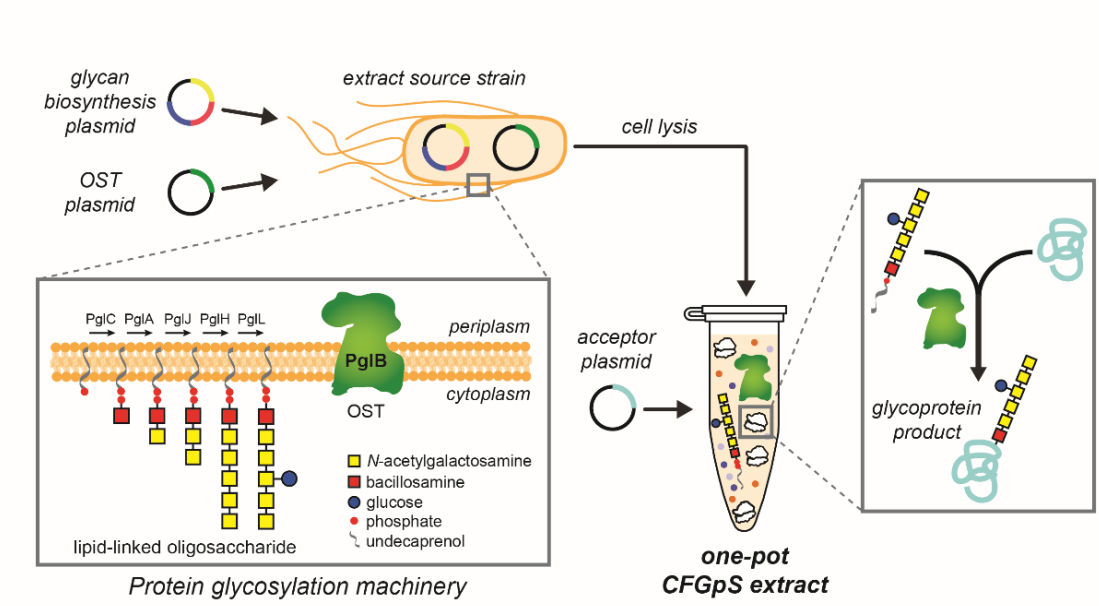


Figure 2-1. Schematic of single-pot CFGpS technology. Glyco-engineered *E. coli* that are modified with (i) genomic mutations that benefit glycosylation reactions and (ii) plasmid DNA for producing essential glycosylation components (*i.e.*, OSTs, LLOs) serve as the source strain for producing crude S30 extracts. Candidate glycosylation components can be derived from all kingdoms of life, including bacteria, and include single-subunit OSTs like *C. jejuni* PglB and LLOs bearing *N*-glycans from *C. jejuni* that are assembled on Und-PP by the Pgl pathway

enzymes. Following extract preparation by lysis of the source strain, one-pot biosynthesis of *N*-glycoproteins is initiated by priming the extract with DNA encoding the acceptor protein target of interest.

2.3 Results

2.3.1 Efficient CFGpS using extracts from glyco-optimized chassis strain.

To develop a one-pot glycoprotein synthesis system, the bacterial protein glycosylation locus (*pgl*) present in the genome of the Gram-negative bacterium *Campylobacter jejuni* was chosen as a model glycosylation system (**Figure 2-1**). This gene cluster encodes an asparagine-linked (*N*-linked) glycosylation pathway that is functionally similar to that of eukaryotes and archaea (Weerapana and Imperiali, 2006), involving a single-subunit OST, PglB, that catalyzes the *en bloc* transfer of a preassembled 1.4 kDa GlcGalNAc$_5$Bac heptasaccharide (where Bac is bacillosamine) from the lipid carrier undecaprenyl pyrophosphate (Und-PP) onto asparagine residues in a conserved motif (D/E-X$_{-1}$-N-X$_{+1}$-S/T, where X$_{-1}$ and X$_{+1}$ are any residues except proline) within acceptor proteins. PglB was selected because we previously showed that *N*-glycosylated acceptor proteins were reliably produced when cell-free translation kits were supplemented with (i) *C. jejuni* PglB (*Cj*PglB) purified from *E. coli* cells and (ii) LLOs extracted from glycoengineered *E. coli* cells expressing the enzymes for producing the *C. jejuni N*-glycan on Und-PP (*Cj*LLOs) (Guarino and DeLisa, 2012b). Additionally, PglB has been used in engineered *E. coli* for transferring eukaryotic

trimannosyl chitobiose glycans (mannose$_3$-*N*-acetylglucosamine$_2$, Man$_3$GlcNAc$_2$) to specific asparagine residues in target proteins (Valderrama-Rincon et al., 2012).

Establishing a CFGpS system first required crude cell extracts suitable for glycoprotein synthesis; hence, we selected *E. coli* strain CLM24 that was previously optimized for *in vivo* protein glycosylation (Feldman et al., 2005a). CLM24 has two attributes that we hypothesized would positively affect cell-free protein glycosylation. First, CLM24 does not synthesize O-polysaccharide antigen due to an inactivating insertion in w*bbL*, which encodes a rhamnosyl transferase that transfers the second sugar of the O16 subunit to UndPP (Liu and Reeves, 1994). Thus, absence of WbbL should allow uninterrupted assembly of engineered glycans, such as the *C. jejuni* heptasaccharide, on UndPP. Second, CLM24 cells lack the *waaL* gene, which encodes the ligase that transfers O-polysaccharide antigens from UndPP to lipid A-core. Because WaaL can also promiscuously transfer engineered glycans that are assembled on UndPP (Chen et al., 2016c; Valderrama-Rincon et al., 2012), the absence of this enzyme should favor accumulation of target glycans on UndPP.

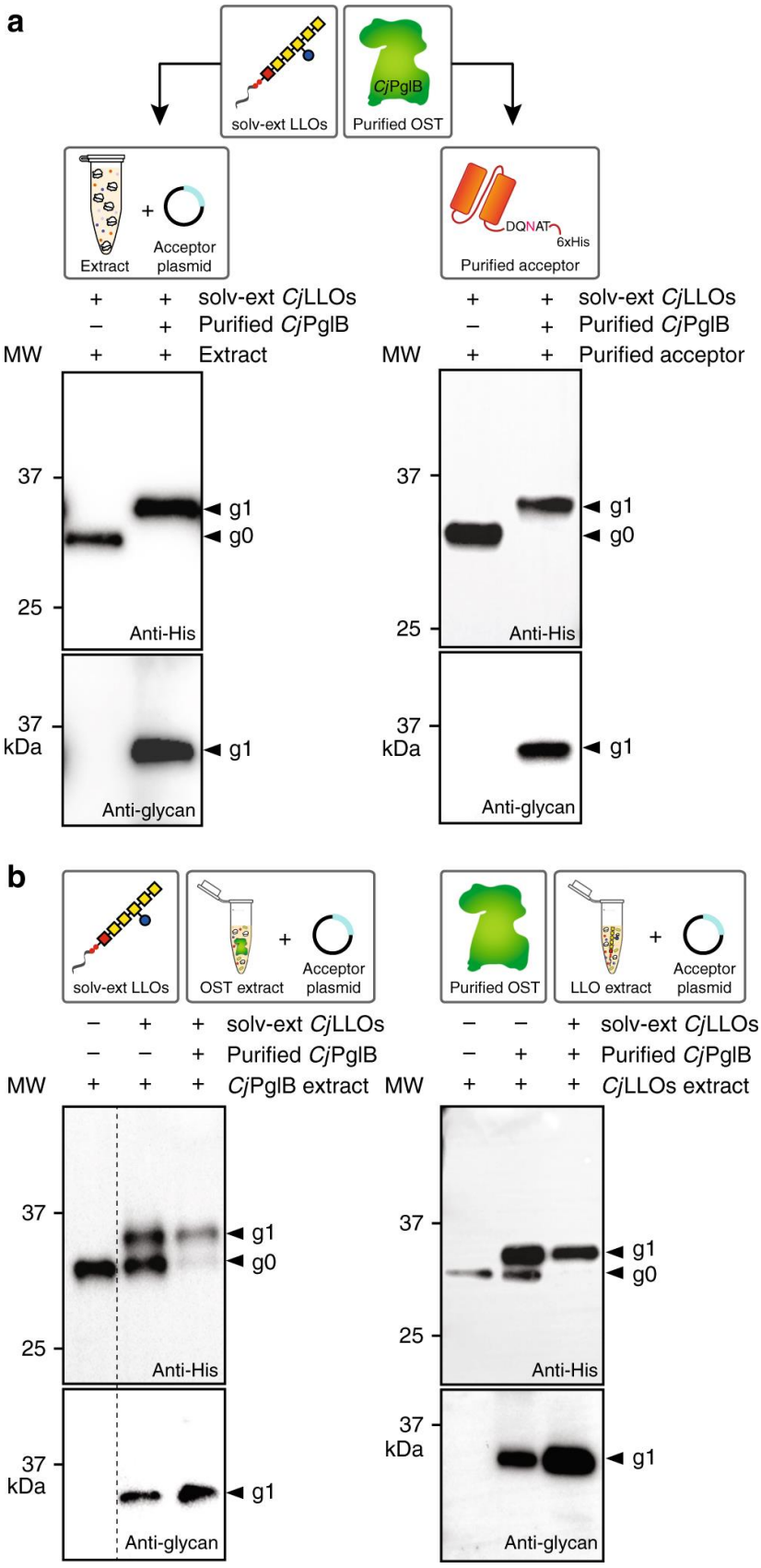

a
solv-ext LLOs
CjPglB
Purified OST
Extract
Acceptor plasmid
DQNAT
6xHis
Purified acceptor
+ + solv-ext CjLLOs
– + Purified CjPglB
MW + + Extract
MW + + Purified acceptor
37
25
g1
g0
Anti-His
37 kDa
Anti-glycan
b
solv-ext LLOs
OST extract
Acceptor plasmid
Purified OST
LLO extract
– + + solv-ext CjLLOs
– – + Purified CjPglB
MW + + + CjPglB extract
MW + + + CjLLOs extract

Figure 2-2. Extract from glyco-optimized chassis strain supports CFGpS. (a) (left) Western blot analysis of scFv13-R4^{DQNAT} produced by crude CLM24 extract supplemented with purified *Cj*PglB and organic solvent-extracted (solv-ext) *Cj*LLOs, and primed with plasmid pJL1-scFv13-R4^{DQNAT}. (right) Western blot analysis of *in vitro* glycosylation reaction using purified scFv13-R4^{DQNAT} acceptor protein that was incubated with purified *Cj*PglB and organic solvent-extracted (solv-ext) *Cj*LLOs. Control reactions (lane 1 in each panel) were performed by omitting purified *Cj*PglB. **(b)** (left) Western blot analysis of scFv13-R4^{DQNAT} produced by crude CLM24 extract selectively enriched with *Cj*PglB from heterologous overexpression from pSF-*Cj*PglB. (right) Western blot analysis of scFv13-R4^{DQNAT} produced by crude CLM24 extract selectively enriched with *Cj*LLOs from heterologous overexpression from pMW07-pglΔB. Reactions were primed with plasmid pJL1-scFv13-R4^{DQNAT} and supplemented with purified *Cj*PglB and organic solvent-extracted (solv-ext) *Cj*LLOs as indicated. Control reactions (lane 1 in each panel) were performed by omitting solv-ext *Cj*LLOs in (left) or purified *Cj*PglB (right) in (b). Blots were probed with anti-hexa-histidine antibody (anti-His) to detect the acceptor protein or hR6 serum (anti-glycan) to detect the *N*-glycan. Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of scFv13-R4^{DQNAT}. Molecular weight (MW) markers are indicated at left. Results are representative of at least three biological replicates.

To determine whether CLM24 could be used as a chassis strain to support integrated cell-free transcription, translation, and glycosylation, we first prepared crude S30 extract from these cells using a rapid and robust procedure for extract preparation based on sonication(Kwon and Jewett, 2015b). Then, 15-µL batch-mode, sequential CFGpS reactions were performed using CLM24 crude extract that was supplemented with the following: (i) an OST catalyst in the form of purified *Cj*PglB that was prepared as described previously (Guarino and DeLisa, 2012b); (ii) oligosaccharide donor in the form of *Cj*LLOs that were isolated by organic solvent extraction from the membrane fraction of glycoengineered *E. coli* cells as described

previously (Guarino and DeLisa, 2012b); and (iii) plasmid DNA encoding the model acceptor protein scFv13-R4^{DQNAT}, an anti-β-galactosidase (β-gal) single-chain variable fragment (scFv) antibody modified C-terminally with a single DQNAT motif (Valderrama-Rincon et al., 2012). The glycosylation status of scFv13-R4^{DQNAT} was analyzed by SDS-PAGE and immunoblotting with an anti-polyhistidine (anti-His) antibody or hR6 serum that is specific for the *C. jejuni* heptasaccharide glycan (Schwarz et al., 2011b). Following an overnight reaction at 30°C, highly efficient glycosylation was achieved as evidenced by the mobility shift of scFv13-R4^{DQNAT} entirely to the mono-glycosylated (g1) form in anti-His immunoblots and the detection of the *C. jejuni* glycan attached to scFv13-R4^{DQNAT} by hR6 serum (**Figure 2-2a**). For synthesis of scFv13-R4^{DQNAT}, the reaction mixture was modified to be oxidizing, through the addition of iodoacetamide and a 3:1 ratio of oxidized and reduced glutathione, demonstrating the flexibility of CFGpS reaction conditions for producing eukaryotic glycoprotein targets. The efficiency achieved in this CFGpS system rivaled that of an *in vitro* glycosylation reaction in which the scFv13-R4^{DQNAT} acceptor protein was expressed and purified from *E. coli*, and then incubated overnight with purified *Cj*PglB and extracted *Cj*LLOs (**Figure 2-2a**). As expected, when *Cj*PglB was omitted from the reaction, the scFv13-R4^{DQNAT} acceptor protein was produced only in the aglycosylated (g0) form. The results generated here with CLM24 extract are consistent with our earlier studies using an *E. coli* S30 extract-based CFPS system or purified translation machinery (Guarino and

DeLisa, 2012b), and establish that the *C. jejuni N*-linked protein glycosylation mechanism can be functionally reconstituted outside the cell.

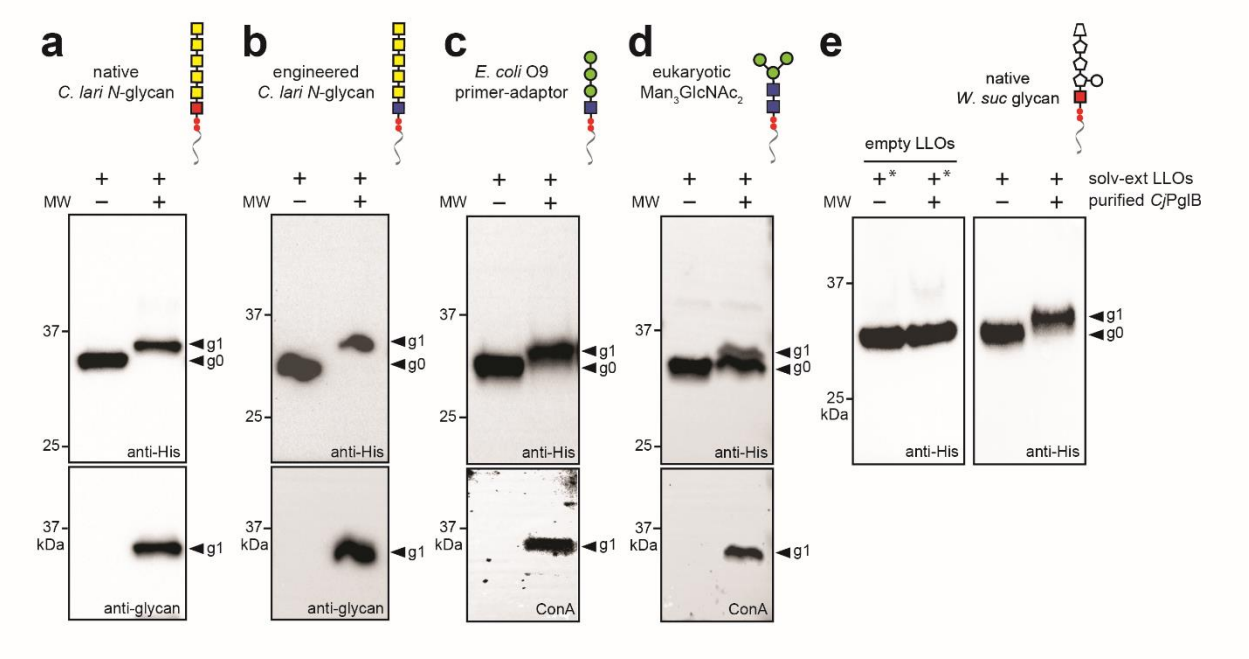


Figure 2-3. Expanding cell-free glycosylation with different oligosaccharide structures. Western blot analysis of *in vitro* glycosylation reaction products generated with purified scFv13-R4^{DQNAT} acceptor protein, purified *Cj*PglB, and organic solvent-extracted (solv-ext) LLOs from cells carrying: **(a)** plasmid pACYC*pgl4* for making the native *C. lari* hexasaccharide *N*-glycan; **(b)** plasmid pACYC*pgl2* for making the engineered *C. lari* hexasaccharide *N*-glycan; **(c)** plasmid pO9-PA for making the *E. coli* O9 'primer-adaptor' $Man_3GlcNAc$ structure; **(d)** plasmid pConYCGmCB for making the eukaryotic $Man_3GlcNAc_2$ *N*-glycan structure; and **(e)** fosmid pEpiFOS-5*pgl5* for making the native *W. succinogenes* hexasaccharide *N*-glycan. Reactions were run at 30°C for 16 h. Blots were probed with anti-His antibody to detect the acceptor protein and one of the following: hR6 serum that cross-reacts with the native and engineered *C. lari* glycans or ConA lectin that binds internal and non-reducing terminal α-mannosyl groups in the $Man_3GlcNAc$ and $Man_3GlcNAc_2$ glycans. Because structural determination of the *W. succinogenes N*-glycan is currently incomplete, and because there are no available antibodies,

the protein product bearing this *N*-glycan was only probed with the anti-His antibody. As an additional control for this glycan, we included empty LLOs prepared from the same host strain but lacking the pEpiFOS-5*pgl5* fosmid (left hand panel, "+" signs marked with an asterisk). Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of the scFv13-R4^{DQNAT} protein. Molecular weight (MW) markers are indicated at left. Results are representative of at least three biological replicates.

2.3.2 Expanding the glycan repertoire of cell-free glycosylation.

To date, only the *C. jejuni* glycosylation pathway has been reconstituted in vitro (Guarino and DeLisa, 2012), and it remains an open question whether our system can be reconfigured with different LLOs and OSTs. Therefore, to extend the range of glycan structures beyond the *C. jejuni* heptasaccharide, we performed glycosylation reactions in which the solvent-extracted *Cj*LLOs used above were replaced with oligosaccharide donors extracted from E. coli cells carrying alternative glycan biosynthesis pathways. These included LLOs bearing the following glycan structures: (i) native *C. lari* hexasaccharide N-glycan (Schwarz et al., 2011); (ii) engineered $GalNAc_5GlcNAc$ based on the *Campylobacter lari* hexasaccharide N-glycan (Schwarz et al., 2010); (iii) native *Wolinella succinogenes* hexasaccharide N-glycan containing three 216-Da monosaccharides and an unusual 232-Da residue at the nonreducing end (Jervis et al., 2012); (iv) engineered *E. coli* O9 primer-adaptor glycan, $Man_3GlcNAc$, that links the O-chain and core oligosaccharide in the lipopolysaccharide of several *E. coli* and *Klebsiella pneumoniae* serotypes (Hagelueken et al., 2015); and (v) eukaryotic trimannosyl core N-glycan, $Man_3GlcNAc_2$ (Valderrama-Rincon et al., 2012). Glycosylation of scFv13-

R4^{DQNAT} with each of these different glycans was observed to occur only in the presence of *Cj*PglB (**Figure 2-3**). It should be noted that 100% glycosylation conversion was observed for each of these glycans except for the $Man_3GlcNAc_2$ *N*-glycan, which had a conversion of ~40% as determined by densitometry analysis. While the reasons for this lower efficiency remain unclear, conjugation efficiency of the same $Man_3GlcNAc_2$ glycan to acceptor proteins in vivo was reported to be even lower (<5%) (Srichaisupakit et al., 2015; Valderrama-Rincon et al., 2012). Hence, transfer of $Man_3GlcNAc_2$ to acceptor proteins in vitro appears to overcome some of the yet-to-be-identified bottlenecks of in vivo glycosylation. This result is likely due to the opportunity with CFGpS to control the concentration of reaction components, for example, providing a higher local concentration of LLO donors. Importantly, scFv13-R4^{DQNAT} was uniformly decorated with a $Man_3GlcNAc_2$ glycan as evidenced by liquid chromatography-mass spectrometry (LC-MS). Specifically, the only major glycopeptide product to be detected was a triply-charged ion containing an N-linked pentasaccharide with m/z = 1032.4583, consistent with the Man3GlcNAc2 glycoform (**Supplementary Figure 6-1**). The tandem MS spectra for this triply-charged glycopeptide yielded an excellent y-ion series and a good b-ion series enabling conclusive determination of the tryptic glycopeptide sequence and attachment of the $Man_3GlcNAc_2$ glycoform at residue N273 of the scFv13-R4^{DQNAT} protein (**Supplementary Figure 6-2**). Taken together, these results demonstrate that structurally diverse glycans, including those that resemble

eukaryotic structures, can be modularly interchanged in cell-free glycosylation reactions.

2.3.3 Extracts enriched with OST enzymes or LLOs co-activate glycosylation.

To circumvent the need for exogenous addition of purified glycosylation components, we hypothesized that heterologous overexpression of OST or GT enzymes directly in the chassis strain would yield extracts that are selectively enriched with the requisite glycosylation components. This strategy was motivated by a recent metabolic engineering approach whereby multiple cell-free lysates were each selectively enriched with an overexpressed metabolic enzyme and then combinatorially mixed to construct an intact pathway (Dudley et al., 2016; Karim and Jewett, 2016). However, a fundamental difference in our system is the fact that the OST and LLOs are not soluble components but instead reside natively in the inner cytoplasmic membrane. This is potentially problematic because of the significant breakup of the cell membrane during S30 extract preparation. However, it has been established that fragments of the *E. coli* inner membrane reform into membrane vesicles, some of which are inverted but others that are orientated properly (Jewett et al., 2008b), and thus could supply the OST and LLOs in a functionally accessible conformation within the extract.

To test this hypothesis, we used a high-pressure homogenization method to prepare crude S30 extract from CLM24 cells carrying a plasmid-encoded copy of

*Cj*PglB such that the resulting cell-free lysates were selectively enriched with detectable quantities of full-length OST enzyme as confirmed by Western blot analysis (**Supplementary Figure 6-3a**). Similarly, crude S30 extract from CLM24 cells overexpressing the *C. jejuni* glycan biosynthesis enzymes produced lysate that was selectively enriched with *Cj*LLOs as confirmed by dot blot analysis with hR6 serum (**Supplementary Figure 6-3b**). It should be noted that the amount of *Cj*LLOs enriched in the crude extract rivaled that produced by the significantly more tedious organic solvent extraction method. Importantly, when 15-μL batch-mode sequential CFGpS reactions were performed using the OST-enriched crude extract that was supplemented with solvent extracted *Cj*LLOs and plasmid DNA encoding scFv13-R4^{DQNAT}, clearly detectable glycosylation of the acceptor protein was observed (**Figure 2-2b**). The conversion of acceptor protein to glycosylated product was ~50%; however, further supplementation with purified *Cj*PglB increased the conversion to nearly 100%, indicating that the amount of OST in the crude extract might have been limiting under the conditions tested. When similar CFGpS reactions were performed using the *Cj*LLOs-enriched crude extract supplemented with purified *Cj*PglB and plasmid DNA encoding scFv13-R4^{DQNAT}, >80% glycosylation of the acceptor protein was observed, which reached 100% when additional donor glycans were supplemented (**Figure 2-2b**).

2.3.4 CFGpS modularity enables glycosylation components to be rapidly interchanged.

Given the open nature of cell-free biosynthesis, we postulated that it should be possible to functionally interchange and prototype alternative biochemical reaction components. One straightforward way that this can be accomplished is by combining separately prepared extracts, each of which is selectively enriched with a given enzyme, such that the resulting reaction mixture comprises a functional biological pathway (Dudley et al., 2016; Karim and Jewett, 2016). As proof of this concept, separately prepared *Cj*LLO and *Cj*PglB extracts were mixed and subsequently primed with DNA encoding the scFv13-R4^{DQNAT} acceptor. The resulting mixture promoted efficient glycosylation of scFv13-R4^{DQNAT} as observed in Western blots probed with anti-His antibody and hR6 serum (**Figure 2-4a**). In addition to scFv13-R4^{DQNAT}, we also expressed a different model acceptor protein that was created by grafting a 21-amino acid sequence from the *C. jejuni* glycoprotein AcrA (Guarino and DeLisa, 2012b), which was further modified with an optimized DQNAT glycosylation site, into a flexible loop of superfolder GFP (sfGFP$^{217\text{-}DQNAT}$). The mixed lysate reaction scheme was able to glycosylate the sfGFP$^{217\text{-}DQNAT}$ acceptor protein with 100% conversion (**Figure 2-4a**). It is noteworthy that the high conversion observed for both acceptor proteins was achieved in mixed lysates without the need to supplement the reactions with purified OST or organic solvent-extracted *Cj*LLOs.

Next, we sought to demonstrate that the mixed lysate approach could be used to rapidly prototype the activity of four additional bacterial OSTs. Crude extracts were separately prepared from CLM24 source strains heterologously overexpressing one of the following bacterial OSTs: *Campylobacter coli* PglB (*Cc*PglB), *Desulfovibrio desulfuricans* PglB (*Dd*PglB), *Desulfovibrio gigas* PglB (*Dg*PglB), or *Desulfovibrio vulgaris* PglB (*Dv*PglB). The resulting extracts were selectively enriched with full-length OST proteins at levels that were comparable to *Cj*PglB (**Supplementary Figure 6-3a**). Each OST extract was mixed with the *Cj*LLO-enriched extract and then supplemented with plasmid DNA encoding sfGFP$^{217\text{-}DQNAT}$ or a modified version of this target protein where the residue in the -2 position of the acceptor sequon was mutated to alanine. Upon completion of CFGpS reactions, the expression and glycosylation status of sfGFP$^{217\text{-}DQNAT}$ and sfGFP$^{217\text{-}AQNAT}$ was followed by Western blot analysis, which revealed information about the sequon preferences for these homologous enzymes. For example, the mixed lysate containing *Cc*PglB was observed to efficiently glycosylate sfGFP$^{217\text{-}DQNAT}$ but not sfGFP$^{217\text{-}AQNAT}$ (**Figure 2-4b**). This activity profile for *Cc*PglB was identical to that observed for *Cj*PglB, which was not surprising based on its high sequence similarity (~81%) to *Cj*PglB. In contrast, lysate mixtures containing OSTs from *Desulfovibrio* sp., which have low sequence identity (~15-20%) to *Cj*PglB, showed more relaxed sequon preferences (**Figure 2-4b**). Specifically, *Dg*PglB-enriched extract mixtures modified both (D/A)QNAT motifs with nearly equal efficiency while mixed lysates containing *Dd*OST and *Dv*OST preferentially glycosylated the AQNAT sequon.

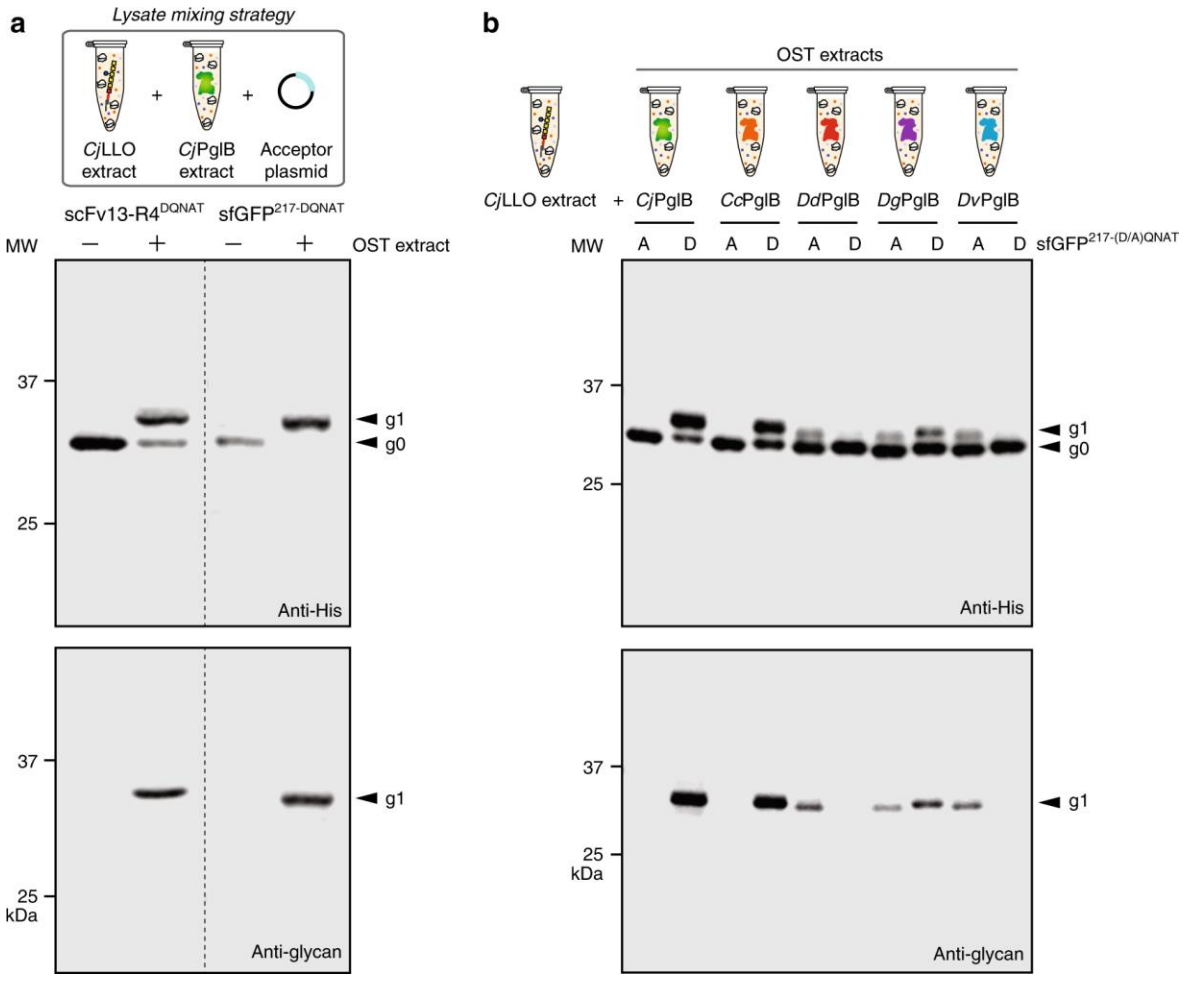


Figure 2-4. Mixing of CFGpS extracts enables rapid prototyping of different OST enzymes. **(a)** Western blot analysis of CFGpS reactions performed using lysate mixing strategy whereby *Cj*LLO lysate derived from CLM24 cells carrying pMW07-pglΔB was mixed with *Cj*PglB lysate derived from CLM24 cells carrying pSF-*Cj*PglB, and the resulting CFGpS mixture was primed with plasmid DNA encoding either scFv13-R4^{DQNAT} or sfGFP$^{217\text{-}DQNAT}$. **(b)** Western blot analysis of CFGpS reactions performed using *Cj*LLO lysate mixed with extract derived from CLM24 cells carrying a pSF plasmid encoding one of the following OSTs: *Cj*PglB, *Cc*PglB, *Dd*PglB, *Dg*PglB, or *Dv*PglB. Mixed lysates were primed with plasmid DNA encoding either sfGFP$^{217\text{-}DQNAT}$ (D) or sfGFP$^{217\text{-}AQNAT}$ (A). Blots were probed with anti-His antibody to detect the acceptor proteins (top panels) and hR6 serum against the *C. jejuni* glycan (bottom panels). Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of the acceptor proteins. Molecular weight (MW) markers are indicated at left. Results are representative of at least three biological replicates.

2.3.5 One-pot extract promotes efficient biosynthesis of diverse glycoprotein targets.

To create a fully integrated CFGpS platform that permits one-pot synthesis of *N*-glycoproteins without the need for supplementation of either purified OSTs or solvent-extracted LLOs (**Figure 1**), we produced crude S30 extract from CLM24 cells heterologously overexpressing *Cj*PglB and the *C. jejuni* glycan biosynthesis enzymes. The resulting extract was selectively enriched with both *Cj*PglB and *Cj*LLOs donor to an extent that was indistinguishable from the separately prepared extracts (**Supplementary Figure 6-3a and b**). Using this extract, CFGpS reactions were performed by addition of plasmid DNA encoding either scFv13-R4^{DQNAT} or sfGFP$^{217\text{-}DQNAT}$. In both cases, 100% protein glycosylation was achieved without the need for exogenous supplementation of separately prepared glycosylation components (**Figure 2-5a**). Independent extract preparations yielded identical results for both protein substrates, confirming the reproducibility of the CFGpS system (**Supplementary Figure 6-4a and b**). Importantly, the *in vitro* synthesized scFv13-R4^{DQNAT} and sfGFP$^{217\text{-}DQNAT}$ proteins retained biological activity that was unaffected by *N*-glycan addition (**Supplementary Figs. 6-5 and 6-6**). From the activity data, the yield of glycosylated scFv13-R4^{DQNAT} and sfGFP$^{217\text{-}DQNAT}$ proteins produced by the one-pot CFGpS system was determined to be ~20 mg L^{-1} and ~10 mg L^{-1}, respectively.

To determine whether human glycoproteins could be similarly produced in our one-pot system, we constructed plasmids for cell-free expression of human erythropoietin (hEPO) glycovariants in which the native sequons at residue N24 (22-

AENIT-26), N38 (36-NENIT-40) or N83 (81-LVNSS-85) were individually mutated to the optimal bacterial sequon, DQNAT (**Figure 2-5b**). CFGpS reactions were then initiated by priming the all-in-one extract with plasmid DNA encoding hEPO$^{22\text{-DQNAT-}26}$, hEPO$^{36\text{-DQNAT-}40}$, or hEPO$^{81\text{-DQNAT-}85}$. Western blot analysis revealed clearly detectable glycosylation of each hEPO glycovariant with 100% glycosylated product for the N24 and N38 sites and ~30-40% for the N83 site (**Figure 2-5b**). As with the model glycoproteins scFv13-R4^{DQNAT} and sfGFP$^{217\text{-DQNAT}}$ above, all three glycosylated hEPO variants retained biological activity that was indistinguishable from the activity measured for the corresponding aglycosylated counterparts, with yields in the ~10 mg L^{-1} range (**Supplementary Figure 6-7**). Collectively, these findings establish that one-pot CFGpS extracts are capable of co-activating protein synthesis and *N*-glycosylation in a manner that yields efficiently glycosylated proteins including those of human origin.

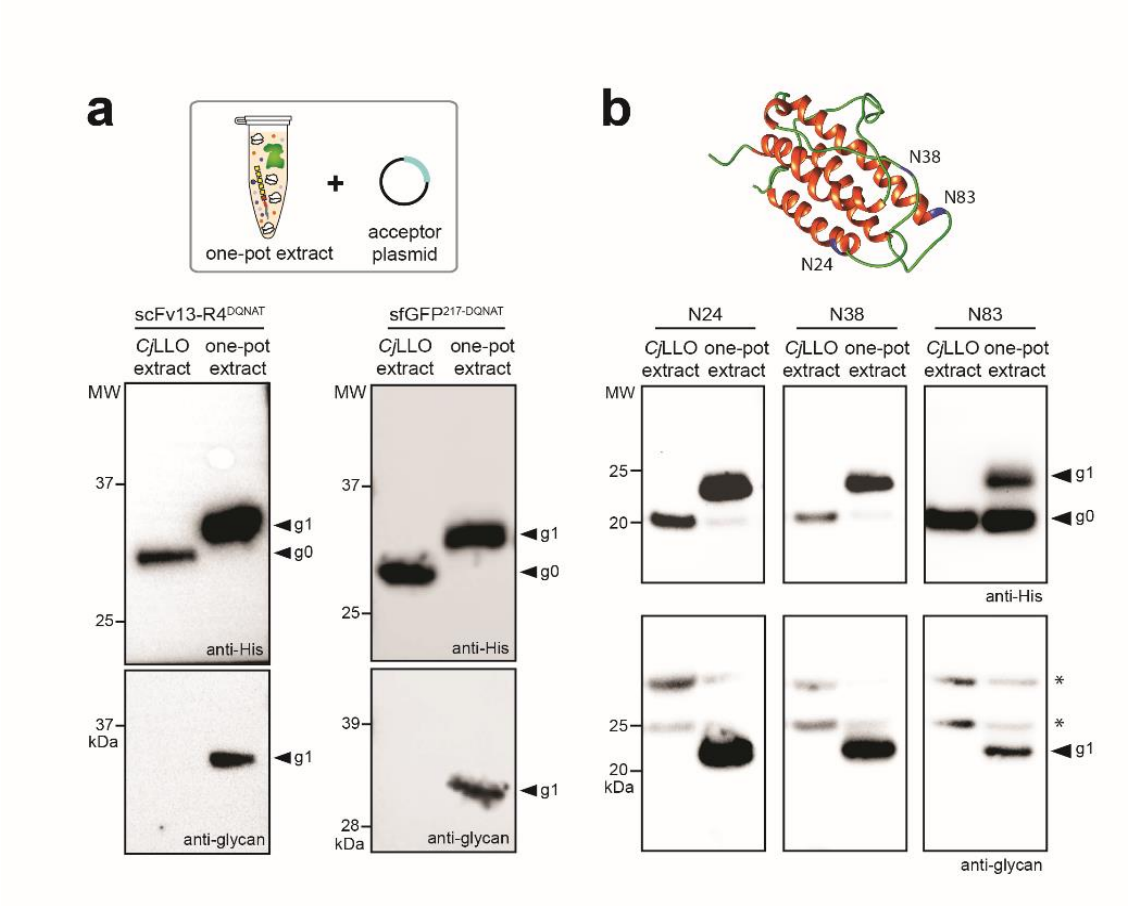


Figure 2-5. One-pot CFGpS using extracts selectively enriched with OSTs and LLOs. (a) Western blot analysis of scFv13-R4^{DQNAT} or sfGFP$^{217\text{-}DQNAT}$ produced by crude CLM24 extract selectively enriched with (i) *Cj*PglB from heterologous overexpression from pSF-*Cj*PglB and (ii) *Cj*LLOs from heterologous overexpression from pMW07-pglΔB. Reactions were primed with plasmid pJL1-scFv13-R4^{DQNAT} or pJL1-sfGFP$^{217\text{-}DQNAT}$. **(b)** Ribbon representation of human erythropoietin (PDB code 1BUY) with α-helixes and flexible loops colored in red and green, respectively. Glycosylation sites modeled by mutating the native sequons at N24 (22-AENIT-26), N38 (36-NENIT-40) or N83 (81-LVNSS-85) to DQNAT, with asparagine residues in each sequon colored blue. Image prepared using UCSF Chimera package.(Pettersen et al., 2004) Glycoengineered hEPO variants in which the native sequons at N24 (22-AENIT-26), N38 (36-NENIT-40) or N83 (81-LVNSS-85) were individually mutated to an optimal bacterial sequon, DQNAT (shown in blue). Western blot analysis of hEPO glycovariants produced by crude

CLM24 extract selectively enriched with (i) *Cj*PglB from heterologous overexpression from pSF-*Cj*PglB and (ii) *Cj*LLOs from heterologous overexpression from pMW07-pglΔB. Reactions were primed with plasmid pJL1-hEPO$^{22\text{-}DQNAT\text{-}26}$ (N24), pJL1-hEPO$^{36\text{-}DQNAT\text{-}40}$ (N38), or pJL1-hEPO$^{81\text{-}DQNAT\text{-}85}$ (N83) as indicated. All control reactions (lane 1 in each panel) were performed using *Cj*LLO-enriched extracts that lacked *Cj*PglB. Blots were probed with anti-hexa-histidine antibody (anti-His) to detect the acceptor proteins or hR6 serum (anti-glycan) to detect the *N*-glycan. Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of the protein targets. Asterisks denote bands corresponding to non-specific serum antibody binding. Molecular weight (MW) markers are indicated at left. Results are representative of at least three biological replicates (see **Supplementary Fig. 6-2** for replicate data).

2.4 Discussion.

In this work, we successfully created a technology for one-pot biosynthesis of *N*-linked glycoproteins in the absence of living cells. This was accomplished by uniting cell-free transcription and translation with the necessary reaction components for *N*-linked protein glycosylation through a process of crude extract enrichment. By preparing OST- and LLO-enriched crude S30 extracts from a glyco-optimized chassis strain, glycosylation-competent lysates were capable of supplying efficiently glycosylated target proteins, with conversion levels at or near 100% in most instances. The glycoprotein yields obtained for three structurally diverse proteins were in the 10-20 mg L^{-1} range, which compare favorably to some of the yields reported previously for these proteins in different CFPS kits or in-house generated extracts. For example, Jackson *et al*. produced 3.6 mg L^{-1} of GFP using the PURExpress system (Jackson et al., 2014), Stech *et al*. produced ~12 mg L^{-1} of an anti-SMAD2 scFv using a CHO cell-derived

lysate (Stech et al., 2017), Ahn *et al*. produced 55 mg L^{-1} of hEPO using an *E. coli*-derived S30 lysate (Ahn et al., 2007), and Gurramkonda, *et al.* produced ~120 mg L^{-1} of hEPO using a CHO cell-derived lysate supplemented with CHO microsomes (Gurramkonda et al., 2018).

Furthermore, this work represents the first demonstration of extract enrichment with catalytically active multipass transmembrane enzymes (and their corresponding lipid-linked substrates) without the need for domain truncation or supplementation of extra scaffold molecules (Schoborg et al., 2018), and provides a blueprint for other CFPS-based applications beyond glycosylation that involve this important class of proteins. Moreover, the ability of OST- or LLO-enriched crude extracts to co-activate glycosylation partially bypassed the need for costly, labor-intensive preparation of glycosylation components and paved the way for a modular single-pot CFGpS platform in which protein synthesis and *N*-linked glycosylation were integrated.

A major advantage of the CFGpS system developed here is the level of control it affords over each of the glycosylation components (*i.e.*, catalysts, substrates, and cofactors) in terms of important process variables such as relative concentration, timing of addition, overall reaction time, etc. Likewise, genome engineering of the chassis strain used to supply the extract, such as our recent report enhancing cell-free synthesis containing multiple, identical non canonical amino acids (Martin et al., 2018), makes it possible to eliminate inhibitory substances such as glycosidases that catalyze the undesired hydrolysis of glycosidic linkages. This user-level control provides an

opportunity to overcome system bottlenecks that effectively limit glycosylation efficiency as we showed with both the *C. jejuni* heptasaccharide and the eukaryotic $Man_3GlcNAc_2$ glycan. Moreover, the open nature of the CFGpS system could be further exploited in the future to introduce components that may otherwise be incompatible with chassis strain expression such as unusual and/or non-natural LLOs that cannot be assembled or flipped *in vivo*.

An additional advantage of the CFGpS system is that it does not rely on commercial cell-free kits to support protein synthesis. For comparison, the glycoproteins yields obtained here were ~10-20 mg L^{-1} in reactions costing ~$0.01-0.03 per μL (**Supplementary Table 6-1** and also (Sun et al., 2013)) versus previous kit-based (*e.g.*, Promega L110; NEB® E6800S) glycoprotein yields of ~100 mg L^{-1} (Guarino and DeLisa, 2012b) in reactions costing ~$1 per μL (Hayes, 2012). As a result, our system can synthesize ~1 μg glycoprotein/$ reagents compared to the previously published approach that can synthesize ~0.1 μg glycoprotein/$ reagents, representing an order of magnitude improvement in relative protein synthesis yields. It is also worth noting that this cost analysis does not take into account the cost of purifying OSTs or extracting LLOs that were used to supplement the commercial kits in our previous work (Guarino and DeLisa, 2012b). We anticipate this reduction in cost will encourage adoption of the CFGpS platform.

Perhaps the most important feature of the CFGpS platform is its modularity, which was evidenced by the interchangeability of: (i) OST enzymes from different

bacterial species; (ii) engineered LLOs with glycan moieties derived from bacteria and eukaryotes; and (iii) diverse acceptor protein targets including naturally occurring human *N*-glycoproteins with terminal or internal acceptor sequons. Importantly, enriched extracts could be readily mixed in a manner that enabled screening of an OST panel whose activities in CFGpS were in line with previously reported activities *in vivo* (Ollis et al., 2015c), thereby validating this lysate mixing strategy as a useful tool for rapid characterization of glycosylation enzyme function and for prototyping glycosylation reactions. In light of this modularity, we envision that lysate enrichment could be further expanded beyond the glycosylation components/substrates tested here. For example, extracts could be heterologously enriched with alternative membrane-bound or soluble OSTs that catalyze *N*-linked or *O*-linked glycosyl transfer reactions. Such biocatalyst swapping is expected to be relatively straightforward in light of the growing number of prokaryotic and eukaryotic OST enzymes that have been recombinantly expressed in functional conformations and used to promote *in vitro* glycosylation reactions (Glover et al., 2005b; Kightlinger et al., 2018; Lizak et al., 2011b; Musumeci et al., 2013a; Ollis et al., 2015c; Ramirez et al., 2017; Schoborg et al., 2018). Likewise, as newly engineered glycan biosynthesis pathways emerge (Merritt et al., 2013), these could be readily integrated into the CFGpS platform through heterologous expression of GTs in the chassis strain. The ability to modularly reconfigure and quickly interrogate glycosylation systems *in vitro* should make the CFGpS technology a useful new addition to the glycoengineering toolkit for increasing

our understanding of glycosylation and, in the future, advancing applications of on demand biomolecular manufacturing [Pardee et al., 2014; Salehi et al., 2016),(Pardee et al., 2016a; Sullivan et al., 2016)].

2.5 Materials and methods.

Bacterial strains and plasmids.

The following *E. coli* strains were used in this study: DH5α, BL21(DE3) (Novagen), CLM24, and Origami2(DE3) *gmd*::kan Δ*waaL*. DH5α was used for plasmid cloning and purification. BL21(DE3) was used for expression and purification of the scFv13-R4^{DQNAT} acceptor protein that was used in all *in vitro* glycosylation reactions. CLM24 is a glyco-optimized derivative of W3110 that carries a deletion in the gene encoding the WaaL ligase, thus facilitating the accumulation of preassembled glycans on Und-PP (Feldman et al., 2005a). CLM24 was used for purification of the *Cj*OST enzyme, organic solvent-based extraction of all LLOs bearing bacterial glycans, and the source strain for preparing extracts with and without selectively enriched glycosylation components. Origami2(DE3) *gmd*::kan Δ*waaL* was used for producing $Man_3GlcNAc_2$-bearing LLOs and was generated by sequential mutation with P1*vir* phage transduction using the respective strains from the Keio collection (Baba et al., 2006) as donors, which were obtained from the Coli Genetic Stock Center (CGSC). In brief, donor lysate was generated from strain JW3597-1 (Δ*rfaL734*::kan) and the resulting phage was used to infect Origami2(DE3) target cells. After plating transformants on LB plates containing kanamycin (Kan), successful transductants were

selected and their Kan resistance cassettes were removed by transforming with temperature-sensitive plasmid pCP20 (Datsenko and Wanner, 2000a). The resulting strain, Origami2(DE3) Δ*waaL*, was then used for subsequent deletion of the *gmd* gene according to an identical strategy but using donor strain JW2038-1 (Δ*gmd751*::kan).

All plasmids used in the study are listed in **Supplementary Table 6-2**. Plasmids constructed in this study were made using standard cloning protocols and confirmed by DNA sequencing. These included the following. Plasmid pJL1-scFv13-R4^{DQNAT} was generated by first PCR amplifying the gene encoding scFv13-R4^{DQNAT} from pET28a-scFv13-R4(N34L, N77L)^{DQNAT}, where the N34L and N77L mutations were introduced to eliminate putative internal glycosylation sites in scFv13-R4 (Ollis et al., 2015c). The resulting PCR product was then ligated between *Nco*I and *Sal*I restriction sites in plasmid pJL1, a pET-based vector used for CFPS (Schoborg et al., 2018). Plasmid pJL1-sfGFP$^{217\text{-}DQNAT}$ was generated by ligating a commercially-synthesized DNA fragment encoding sfGFP$^{217\text{-}DQNAT}$ (Integrated DNA Technologies) into pJL1. This version of sfGFP contains an additional GT insertion after K214, which extends this flexible loop before the final beta sheet (Bundy and Swartz, 2010). Into this flexible loop, immediately after T216, we grafted a 21 amino acid sequence containing the *C. jejuni* AcrA N123 glycosylation site (Guarino and DeLisa, 2012b), but with an optimal DQNAT sequon in place of the native AcrA sequon. Similar procedures were used to generate plasmids pJL1-sfGFP$^{217\text{-}AQNAT}$, pJL1-hEPO$^{22\text{-}DQNAT\text{-}26}$, pJL1-hEPO$^{36\text{-}DQNAT\text{-}40}$, and pJL1-hEPO$^{81\text{-}DQNAT\text{-}85}$. In the case of pJL1-hEPO$^{22\text{-}DQNAT\text{-}26}$, the gene for mature human

EPO was designed such that the native sequon at N24 was changed from 22-AENIT-26 to an optimal bacterial sequon, DQNAT. Identical cloning strategies were carried out to separately introduce optimal DQNAT motifs in place of the native hEPO sequons 36-NENIT-40 and 81-LVNSS-85. Recombinant expression of the *E. coli* O9 primer-adaptor glycan ($Man_3GlcNAc$) on Und-PP was achieved by cloning the genes encoding the WbdB and WbdC mannosyltransferase enzymes derived from *E. coli* ATCC31616 for assembling the glycan, and RfbK and RfbM, also derived from *E. coli* ATCC31616 for increasing the pool of available GDP-mannose, in *E. coli* MG1655. Plasmid pConYCGmCB was constructed by isothermal Gibson assembly and encodes an artificial operon comprised of: (i) the yeast glycosyltransferases Alg13, Alg14, Alg1, and Alg2 for $Man_3GlcNAc_2$ glycan biosynthesis (Valderrama-Rincon et al., 2012) and (ii) the *E. coli* enzymes phosphomannomutase (ManB) and mannose-1-phosphate guanylyltransferase (ManC), which together increase availability of GDP-mannose substrates for the Alg1 and Alg2 enzymes.

Protein expression and purification.

Purification of *Cj*PglB was performed according to a previously described protocol (Guarino and DeLisa, 2012b). Briefly, a single colony of *E. coli* CLM24 carrying plasmid pSN18 (Kowarik et al., 2006a) was grown overnight at 37°C in 50 mL of Luria-Bertani (LB; 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 5 g L^{-1} NaCl, pH 7.2) supplemented with ampicillin (Amp) and 0.2% (w/v %) D-glucose. Overnight cells were subcultured

into 1 L of fresh terrific broth (TB; 12 g L^{-1} tryptone, 24 g L^{-1} yeast extract, 0.4% (v/v %) glycerol, 10% (v/v %) 0.17 M KH_2PO_4/0.72 M K_2HPO_4 phosphate buffer), supplemented with Amp and grown until the absorbance at 600 nm (Abs_{600}) reached a value of ~0.7. The incubation temperature was adjusted to 16°C, after which protein expression was induced by the addition of L-arabinose to a final concentration of 0.02% (w/v %). Protein expression was allowed to proceed for 20 h at 16°C. Cells were harvested by centrifugation and then disrupted using a homogenizer (Avestin C5 EmulsiFlex). The lysate was centrifuged to remove cell debris and the supernatant was ultracentrifuged (100,000×g) for 2 h at 4°C. The resulting pellet containing the membrane fraction was fully resuspended with a Potter-Elvehjem tissue homogenizer in buffer containing 50 mM HEPES, 250 mM NaCl, 10% (v/v %) glycerol, and 1% (w/v %) n-dodecyl-β-D-maltoside (DDM) at pH 7.5. The suspension was incubated at room temperature for 1 h to facilitate detergent solubilization of *Cj*PglB from native *E. coli* lipids, which were removed by subsequent ultracentrifugation (100,000×g) for 1 h at 4°C. The supernatant containing DDM-solubilized *Cj*PglB was purified using Ni-NTA resin (Thermo) according to manufacturer's specification with the exception that all buffers were supplemented with 1% (w/v %) DDM. The elution fraction from Ni-NTA purification was then subjected to size exclusion chromatography (SEC) using an ÄKTA Explorer FPLC system (GE Healthcare) with Superdex 200 10/300 GL column. Purified protein was stored at a final concentration of 1-2 mg mL^{-1} in OST storage buffer (50 mM HEPES,

100 mM NaCl, 5% (v/v %) glycerol, 0.01% (w/v %) DDM, pH 7.5) at 4°C. Glycerol concentration in the sample was adjusted to 20% (v/v %) for long-term storage at -80°C.

Purification of acceptor protein scFv13-R4^{DQNAT} was carried out as described previously (Ollis et al., 2015c). Briefly, *E. coli* strain BL21(DE3) carrying plasmid pET28a-scFv13-R4(N34L, N77L)^{DQNAT} was grown in 1.0 L of TB supplied with kanamycin. The culture was incubated at 37°C until Abs_{600} reached ~0.7, at which point protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. Protein expression was allowed to proceed for 20 h at 25°C. Cells were harvested and disrupted identically as described above. The scFv13-R4^{DQNAT} protein was purified using Ni-NTA resin followed by SEC according to manufacturer's protocols. Protein was stored at a final concentration of 1-2 mg mL^{-1} in storage buffer (50 mM HEPES, 250 mM NaCl, 1 mM EDTA, pH 7.5) at 4°C.

Extraction of LLOs.

The protocol for organic solvent extraction of LLOs from *E. coli* membranes was adapted from a previously described protocol (Guarino and DeLisa, 2012b; Jaroentomeechai et al., 2017). In most cases, a single colony of strain CLM24 carrying a plasmid for target glycan biosynthesis (**Supplementary Table 6-2**) was grown overnight in LB media. The notable exceptions were LLOs bearing the *W. succinogenes* *N*-glycan (*Ws*LLOs), which were produced using DH5α cells carrying the pEpiFOS-

5*pgl5* fosmid (kindly provided by Dr. Markus Aebi), and LLOs bearing $Man_3GlcNAc_2$, which were produced using Origami2(DE3) *gmd*::kan Δ*waaL* cells carrying plasmid pConYCGmCB. Overnight cells were subcultured into 1.0 L of TB supplemented with an appropriate antibiotic and grown until the Abs_{600} reached ~0.7. The incubation temperature was adjusted to 30°C for biosynthesis of all glycans except for $Man_3GlcNAc_2$, which was adjusted to 16°C. For plasmid pMW07-pglΔB, protein expression was induced with L-arabinose at a final concentration of 0.2% (w/v %) while for fosmid pEpiFOS-5*pgl5* induction was with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1.0 mM. All other plasmids involved constitutive promoters and thus did not require chemical inducers. After 16 h, cells were harvested by centrifugation and cell pellets were lyophilized to complete dryness at -70°C. For extraction of *Cj*LLOs, native and engineered *Cl*LLOs, *E. coli* O9 primer-adaptor LLOs, and *Ws*LLOs, the lyophilisates were suspended in 10:20:3 volumetric ratio of $CHCl_3$:CH_3OH:H_2O solution and incubated at room temperature for 15 min to facilitate extraction of LLOs. For extraction of LLOs bearing $Man_3GlcNAc_2$ glycan, lyophilisate was successively suspended in 10:20 (v/v %) $CHCl_3$:CH_3OH solution, water, and 10:20:3 $CHCl_3$:CH_3OH:H_2O solution with 15 min of incubation at room temperature between each step. In each case, the final suspension was centrifuged (4000×g) for 15 min, after which the organic layer (bottom layer) was collected and dried with a vacuum concentrator followed by lyophilization. Lyophilisates containing active LLOs

were resuspended in cell-free glycosylation buffer (10 mM HEPES, pH 7.5, 10 mM $MnCl_2$, and 0.1% (w/v %) DDM) and stored at 4°C.

Preparation of crude S30 extracts.

CLM24 source strains were grown in 2xYTPG (10 g L^{-1} yeast extract, 16 g L^{-1} tryptone, 5 g L^{-1} NaCl, 7 g L^{-1} K_2HPO_4, 3 g L^{-1} KH_2PO_4, 18 g L^{-1} glucose, pH 7.2) until the Abs_{600} reached ~3. To generate OST-enriched extract, CLM24 carrying plasmid pSF-*Cj*PglB, pSF-*Cc*PglB, pSF-*Dd*PglB, pSF-*Dg*PglB, or pSF-*Dv*PglB (Ollis et al., 2015c) was used as the source strain. To generate LLO-enriched extract, CLM24 carrying plasmid pMW07-pglΔB was used as the source strain. To generate one-pot extract containing both OST and LLOs, CLM24 carrying pMW07-pglΔB and pSF-*Cj*OST was used as the source strain. As needed, the expression of glycosylation components was induced with L-arabinose at final concentration of 0.02% (w/v %). After induction, protein expression was allowed to proceed at 30°C to a density of OD_{600} ~3, at which point cells were harvested by centrifugation (5,000×g) at 4°C for 15 min. All subsequent steps were carried out at 4°C unless otherwise stated. Pelleted cells were washed three times in S30 buffer (10 mM tris acetate, 14 mM magnesium acetate, 60 mM potassium acetate, pH 8.2). After the last wash, cells were pelleted at 7000×g for 10 min and flash frozen on liquid nitrogen. To make lysate, cells were thawed and resuspended to homogeneity in 1 mL of S30 buffer per 1 g of wet cell mass. Cells were disrupted using an Avestin EmulsiFlex-B15 high-pressure homogenizer at 20,000-25,000 psi with a single passage.

The lysate was then centrifuged twice at 30,000×g for 30 min to remove cell debris. Supernatant was transferred to a new vessel and incubated with 250 rpm shaking at 37°C for 60 min to degrade endogenous mRNA transcripts and disrupt existing polysome complexes in the lysate. Following centrifugation (15,000×g) for 15 min at 4°C, supernatant was collected, aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C. S30 extract was active for about 3 freeze-thaw cycles and contained ~40 g L^{-1} total protein as measured by Bradford assay.

Cell-free glycoprotein synthesis.

For *in vitro* glycosylation of purified acceptor protein, reactions were carried out in a 50 μL volume containing 3 μg of scFv13-R4^{DQNAT}, 2 μg of purified *Cj*PglB, and 5 μg extracted LLOs (in the case of $Man_3GlcNAc_2$ LLOs, 20 μg was used) in *in vitro* glycosylation buffer (10 mM HEPES, pH 7.5, 10 mM $MnCl_2$, and 0.1% (w/v %) DDM). The reaction mixture was incubated at 30° C for 16 h. For crude extract-based expression of glycoproteins, a two-phase scheme was implemented. In the first phase, protein synthesis was carried out with a modified PANOx-SP system (Jewett and Swartz, 2004b). Specifically, 1.5 mL microcentrifuge tubes were charged with 15-μL reactions containing 200 ng plasmid DNA, 30% (v/v %) S30 extract and the following: 12 mM magnesium glutamate, 10 mM ammonium glutamate, 130 mM potassium glutamate, 1.2 mM adenosine triphosphate (ATP), 0.85 mM guanosine triphosphate (GTP), 0.85 mM uridine triphosphate (UTP), 0.85 mM cytidine triphosphate (CTP),

0.034 mg mL^{-1} folinic acid, 0.171 mg mL^{-1} *E. coli* tRNA (Roche), 2 mM each of 20 amino acids, 30 mM phosphoenolpyruvate (PEP, Roche), 0.33 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme-A (CoA), 4 mM oxalic acid, 1 mM putrescine, 1.5 mM spermidine, and 57 mM HEPES. For scFv13-$R4^{DQNAT}$ and $hEPO^{22\text{-}DQNAT\text{-}26}$, this phase was carried out at 30°C for 4 h under oxidizing conditions while for $sfGFP^{217\text{-}DQNAT}$ and $sfGFP^{217\text{-}AQNAT}$ this phase was carried out at 30°C for 5 min under reducing conditions. For oxidizing conditions, extract was pre-conditioned with 750 μM iodoacetamide in the dark at room temperature for 30 min and the reaction mix was supplied with 200 mM glutathione at a 3:1 ratio between oxidized and reduced forms. The active sfGFP yields from cell-free reactions were quantified by measuring fluorescence in-lysate and converting into concentration using a standard curve as previously described (Kwon and Jewett, 2015b). In the second phase, protein glycosylation was initiated by the addition of $MnCl_2$ and DDM at a final concentration of 10 mM and 0.1% (w/v %), respectively, and allowed to proceed at 30°C for 16 h. As needed, reactions were supplemented with 2 μg of purified *Cj*PglB (*i.e.*, for CFGpS with LLO-enriched extracts) or 5 μg solvent-extracted *Cj*LLOs (*i.e.*, for CFGpS with OST-enriched extracts). All reactions were stopped by adding Laemmli sample buffer containing 5% βME, after which samples were boiled at 100° C for 15 min and analyzed by SDS-PAGE and Western blotting.

Western blot analysis.

Samples containing 0.5 μg of acceptor protein were loaded into SDS-PAGE gels. Following electrophoretic separation, proteins were transferred from gels onto Immobilon-P polyvinylidene difluoride (PVDF) membranes (0.45 μm) according to manufacturer's protocol. Membranes were washed twice with TBS buffer (80 g L^{-1} NaCl, 20 g L^{-1} KCl, and 30 g L^{-1} Tris-base) followed by incubation for 1 h in blocking solution (50 g L^{-1} non-fat milk in TBST (TBS supplied with 0.05 % (v/v %) Tween-20)). After blocking, membranes were washed 4 times with TBST with 10 min incubation between each wash. A first membrane was probed with 6xHis-polyclonal antibody (Abcam, ab137839, 1:7500) that specifically recognizes hexahistidine epitope tags while a second replicate membrane was probed with one of the following: hR6 (1:10000) serum from rabbit that recognizes the native *C. jejuni* and *C. lari* glycan as well as engineered *C. lari* glycan or ConA-HRP (Sigma, L6397, 1:2500) that recognizes $Man_3GlcNac$ and $Man_3GlcNAc_2$. Probing of membranes was performed for at least 1 hour with shaking at room temperature, after which membranes were washed with TBST in the same manner as described above. For development, membranes were incubated briefly at room temperature with Western ECL substrate (BioRad) and imaged using a ChemiDocTM XRS+ System. OST enzymes enriched in extracts were detected by an identical SDS-PAGE procedure followed by Western blot analysis with a polyclonal antibody specific to the FLAG epitope tag (Abcam, ab49763, 1:7500). The

glycan component of LLOs enriched in extracts was detected by directly spotting 10 μL of extracts onto nitrocellulose membranes followed by detection with hR6 serum.

GFP fluorescence activity. The activity of cell-free-derived sfGFP was determined using an in-lysate fluorescence analysis as described previously (Kwon and Jewett, 2015b). Briefly, 2 μL of cell-free synthesized glycosylated sfGFP reaction was diluted into 48 μL of nanopure water. The solution was then placed in a Costar 96-well black assay plate (Corning). Excitation and emission wavelength for sfGFP fluorescence were at 485 and 528 nm, respectively.

Enzyme-linked immunosorbent analysis (ELISA).

Costar 96-well ELISA plates (Corning) were coated overnight at 4°C with 50 μl of 1 mg mL^{-1} *E. coli* β-gal (Sigma-Aldrich) in 0.05 M sodium carbonate buffer (pH 9.6). After blocking with 5% (w/v %) bovine serum albumin (BSA) in PBS for 3 h at room temperature, the plates were washed four times with PBST buffer (PBS, 0.05% (v/v %) Tween-20, 0.3% (w/v %) BSA) and incubated with serially diluted purified scFv13 R4 samples or soluble fractions of CFGpS lysates for 1 h at room temperature. Samples were quantified by the Bradford assay and an equivalent amount of total protein was applied to the plate. After washing four times with the same buffer, anti-6X-His-HRP conjugated rabbit polyclonal antibody (Abcam) in 3% PBST was added to each well for 1 h. Plates were washed and developed using standard protocols.

***In vitro* cell proliferation assay**.

Human erythroleukemia TF-1 cells (Sigma) that require granulocyte–macrophage colony-stimulating factor (GM-CSF), interleukin 3 (IL-3), or hEPO for growth and survival were used. Cells were maintained in RPMI-1640 media supplemented with 10% FBS, 50 U mL^{-1} penicillin, 50 mg mL^{-1} streptomycin, 2 mM glutamine, and 2 ng mL^{-1} GM-CSF at 37°C in a humidified atmosphere containing 5% CO_2. After 16 h incubation in RPMI-1640 media without GM-CSF, cells were counted, harvested, and resuspended in fresh media. $5x10^3$ TF-1 cells/well were seeded in a 96-well assay plate, and EPO standards or samples were added to final desired concentrations to each well. Cells were incubated with for 6 h in humid incubator before adding alamarBlue®. After 12 h, fluorescence signal was measured at 560 nm/590 nm excitation/emission wavelength.

2.6 Acknowledgements.

We thank Judith Merritt (Glycobia, Inc.) for providing plasmid pMW07-pglΔB, Bil Clemons (California Institute of Technology) for plasmids encoding various PglB homologs, and Markus Aebi for providing plasmid pACYCpgl2 and hR6 serum used in this work. We thank Mr. Robert Sherwood from the Cornell Proteomics and Mass Spectrometry Facility for his technical assistance acquiring the LC–MS/MS raw data files. We thank Weston Kightlinger and James Kath for providing plasmid pJL1-hEPO that was used to generate the EPO variants used in this study. Finally, we thank

Jasmine Hershewe for critical discussions and sharing of reagents and ideas. This work was supported by the Defense Threat Reduction Agency (GRANT11631647 to M.P.D., M.C.J., and M.M.), National Science Foundation (Grants # CBET 1159581 and CBET 1264701 both to M.P.D. and MCB 1413563 to M.P.D. and M.C.J.), the David and Lucile Packard Foundation, and the Dreyfus Teacher-Scholar program. T.J. was supported by a Royal Thai Government Fellowship. J.C.S. and C.J.G. were each supported by a National Science Foundation Graduate Research Fellowship.

2.7 References.

Ahn, J. H., Hwang, M. Y., Lee, K. H., Choi, C. Y., & Kim, D. M. (2007). Use of signal sequences as an in situ removable sequence element to stimulate protein synthesis in cell-free extracts. *Nucleic Acids Res, 35*(4), e21. doi:10.1093/nar/gkl917

Albayrak, C., & Swartz, J. R. (2013). Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res, 41*(11), 5949-5963. doi:10.1093/nar/gkt226

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., . . . Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol, 2*, 2006 0008. doi:10.1038/msb4100050

Brodel, A. K., Sonnabend, A., Roberts, L. O., Stech, M., Wustenhagen, D. A., & Kubick, S. (2013). IRES-mediated translation of membrane proteins and glycoproteins in eukaryotic cell-free systems. *PLoS ONE, 8*(12), e82234. doi:10.1371/journal.pone.0082234

Bundy, B. C., & Swartz, J. R. (2010). Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation. *Bioconjug Chem, 21*(2), 255-263. doi:10.1021/bc9002844

Carlson, E. D., Gan, R., Hodgman, C. E., & Jewett, M. C. (2012). Cell-free protein synthesis: applications come of age. *Biotechnol Adv, 30*(5), 1185-1194. doi:10.1016/j.biotechadv.2011.09.016

Chen, L., Valentine, J. L., Huang, C. J., Endicott, C. E., Moeller, T. D., Rasmussen, J. A., . . . DeLisa, M. P. (2016). Outer membrane vesicles displaying engineered glycotopes elicit protective antibodies. *Proc Natl Acad Sci U S A, 113*(26), E3609-3618. doi:10.1073/pnas.1518311113

Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A, 97*(12), 6640-6645. doi:10.1073/pnas.120163297

Dudley, Q. M., Anderson, K. C., & Jewett, M. C. (2016). Cell-free mixing of Escherichia coli crude extracts to prototype and rationally engineer high-titer mevalonate synthesis. *ACS Synth Biol, 5*(12), 1578-1588. doi:10.1021/acssynbio.6b00154

Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., . . . Aebi, M. (2005). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proc Natl Acad Sci U S A, 102*(8), 3016-3021. doi:10.1073/pnas.0500044102

Glover, K. J., Weerapana, E., Numao, S., & Imperiali, B. (2005). Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from Campylobacter jejuni. *Chem Biol, 12*(12), 1311-1315. doi:S1074-5521(05)00309-1

Goshima, N., Kawamura, Y., Fukumoto, A., Miura, A., Honma, R., Satoh, R., . . . Nomura, N. (2008). Human protein factory for converting the transcriptome into an in vitro–expressed proteome. *Nature Methods, 5*(12), 1011-1017. doi:10.1038/nmeth.1273

Guarino, C., & DeLisa, M. P. (2012). A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology, 22*(5), 596-601. doi:10.1093/glycob/cwr151

Gurramkonda, C., Rao, A., Borhani, S., Pilli, M., Deldari, S., Ge, X., . . . Rao, G. (2018). Improving the recombinant human erythropoietin glycosylation using microsome supplementation in CHO cell-free system. *Biotechnol Bioeng, 115*(5), 1253-1264. doi:10.1002/bit.26554

Hagelueken, G., Clarke, B. R., Huang, H., Tuukkanen, A., Danciu, I., Svergun, D. I., . . . Naismith, J. H. (2015). A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. *Nat Struct Mol Biol, 22*(1), 50-56. doi:10.1038/nsmb.2935

Hamilton, S. R., Bobrowicz, P., Bobrowicz, B., Davidson, R. C., Li, H., Mitchell, T., . . . Gerngross, T. U. (2003). Production of complex human glycoproteins in yeast. *Science, 301*(5637), 1244-1246. doi:10.1126/science.1088166

Hayes, C. (2012). Biomolecular Breadboards:Protocols:cost estimate. http://www.openwetware.org/wiki/Biomolecular_Breadboards:Protocols:cost_estimate.

Hebert, D. N., Lamriben, L., Powers, E. T., & Kelly, J. W. (2014). The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nat Chem Biol, 10*(11), 902-910. doi:10.1038/nchembio.1651

Helenius, A., & Aebi, M. (2001). Intracellular functions of N-linked glycans. *Science, 291*(5512), 2364-2369.

Imperiali, B., & O'Connor, S. E. (1999). Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol, 3*(6), 643-649. doi:S1367-5931(99)00021-6

Jackson, K., Kanamori, T., Ueda, T., & Fan, Z. H. (2014). Protein synthesis yield increased 72 times in the cell-free PURE system. *Integr Biol (Camb), 6*(8), 781-788. doi:10.1039/c4ib00088a

Jaroentomeechai, T., Zheng, X., Hershewe, J., Stark, J. C., Jewett, M. C., & DeLisa, M. P. (2017). A Pipeline for Studying and Engineering Single-Subunit Oligosaccharyltransferases. *Methods Enzymol, 597*, 55-81. doi:10.1016/bs.mie.2017.07.011

Jervis, A. J., Butler, J. A., Lawson, A. J., Langdon, R., Wren, B. W., & Linton, D. (2012). Characterization of the structurally diverse N-linked glycans of Campylobacter species. *J Bacteriol, 194*(9), 2355-2362. doi:10.1128/JB.00042-12

Jewett, M. C., Calhoun, K. A., Voloshin, A., Wuu, J. J., & Swartz, J. R. (2008). An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol, 4*, 220. doi:10.1038/msb.2008.57

Jewett, M. C., & Swartz, J. R. (2004). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol Bioeng, 86*(1), 19-26. doi:10.1002/bit.20026

Kaiser, L., Graveland-Bikker, J., Steuerwald, D., Vanberghem, M., Herlihy, K., & Zhang, S. (2008). Efficient cell-free production of olfactory receptors: detergent optimization, structure, and ligand binding analyses. *Proc Natl Acad Sci U S A, 105*(41), 15726-15731. doi:0804766105

Karim, A. S., & Jewett, M. C. (2016). A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metab Eng, 36*, 116-126. doi:10.1016/j.ymben.2016.03.002

Kiga, D., Sakamoto, K., Kodama, K., Kigawa, T., Matsuda, T., Yabuki, T., . . . Yokoyama, S. (2002). An engineered Escherichia coli tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proc Natl Acad Sci U S A, 99*(15), 9715-9720. doi:10.1073/pnas.142220099

Kightlinger, W., Lin, L., Rosztoczy, M., Li, W., DeLisa, M. P., Mrksich, M., & Jewett, M. C. (2018). Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. *Nat Chem Biol*. doi:10.1038/s41589-018-0051-2

Kowarik, M., Numao, S., Feldman, M. F., Schulz, B. L., Callewaert, N., Kiermaier, E., . . . Aebi, M. (2006). N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science, 314*(5802), 1148-1150. doi:10.1126/science.1134351

Kwon, Y. C., & Jewett, M. C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci Rep, 5*, 8663. doi:10.1038/srep08663

Lanctot, P. M., Gage, F. H., & Varki, A. P. (2007). The glycans of stem cells. *Curr Opin Chem Biol, 11*(4), 373-380. doi:10.1016/j.cbpa.2007.05.032

Lingappa, V. R., Lingappa, J. R., Prasad, R., Ebner, K. E., & Blobel, G. (1978). Coupled cell-free synthesis, segregation, and core glycosylation of a secretory protein. *Proc Natl Acad Sci U S A, 75*(5), 2338-2342.

Liu, D., & Reeves, P. R. (1994). Escherichia coli K12 regains its O antigen. *Microbiology, 140 (Pt 1)*, 49-57. doi:10.1099/13500872-140-1-49

Lizak, C., Gerber, S., Numao, S., Aebi, M., & Locher, K. P. (2011). X-ray structure of a bacterial oligosaccharyltransferase. *Nature, 474*(7351), 350-355. doi:10.1038/nature10151

Martin, R. W., Des Soye, B. J., Kwon, Y. C., Kay, J., Davis, R. G., Thomas, P. M., . . . Jewett, M. C. (2018). Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nat Commun, 9*(1), 1203. doi:10.1038/s41467-018-03469-5

Matsuoka, K., Komori, H., Nose, M., Endo, Y., & Sawasaki, T. (2010). Simple screening method for autoantigen proteins using the N-terminal biotinylated protein library produced by wheat cell-free synthesis. *J Proteome Res, 9*(8), 4264-4273. doi:10.1021/pr9010553

Merritt, J. H., Ollis, A. A., Fisher, A. C., & DeLisa, M. P. (2013). Glycans-by-design: engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. *Biotechnol Bioeng, 110*(6), 1550-1564. doi:10.1002/bit.24885

Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Dos Santos, A., . . . Callewaert, N. (2014). GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins. *Nat Biotechnol, 32*(5), 485-489. doi:10.1038/nbt.2885

Mikami, S., Kobayashi, T., Yokoyama, S., & Imataka, H. (2006). A hybridoma-based in vitro translation system that efficiently synthesizes glycoproteins. *J Biotechnol, 127*(1), 65-78. doi:S0168-1656(06)00535-9

Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., . . . Freemont, P. S. (2018). Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proc Natl Acad Sci U S A, 115*(19), E4340-E4349. doi:10.1073/pnas.1715806115

Moreno, S. N., Ip, H. S., & Cross, G. A. (1991). An mRNA-dependent in vitro translation system from Trypanosoma brucei. *Mol Biochem Parasitol, 46*(2), 265-274.

Musumeci, M. A., Hug, I., Scott, N. E., Ielmini, M. V., Foster, L. J., Wang, P. G., & Feldman, M. F. (2013). In vitro activity of Neisseria meningitidis PglL O-oligosaccharyltransferase with diverse synthetic lipid donors and a UDP-activated sugar. *J Biol Chem, 288*(15), 10578-10587. doi:10.1074/jbc.M112.432815

Ollis, A. A., Chai, Y., Natarajan, A., Perregaux, E., Jaroentomeechai, T., Guarino, C., . . . DeLisa, M. P. (2015). Substitute sweeteners: diverse bacterial

oligosaccharyltransferases with unique N-glycosylation site preferences. *Sci Rep, 5*, 15237. doi:10.1038/srep15237

Ollis, A. A., Zhang, S., Fisher, A. C., & DeLisa, M. P. (2014). Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nat Chem Biol, 10*(10), 816-822. doi:10.1038/nchembio.1609

Oza, J. P., Aerni, H. R., Pirman, N. L., Barber, K. W., Ter Haar, C. M., Rogulina, S., . . . Jewett, M. C. (2015). Robust production of recombinant phosphoproteins using cell-free protein synthesis. *Nat Commun, 6*, 8168. doi:10.1038/ncomms9168

Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., & Collins, J. J. (2014). Paper-based synthetic gene networks. *Cell, 159*(4), 940-954. doi:10.1016/j.cell.2014.10.004

Pardee, K., Slomovic, S., Nguyen, P. Q., Lee, J. W., Donghia, N., Burrill, D., . . . Collins, J. J. (2016). Portable, on-demand biomolecular manufacturing. *Cell, 167*(1), 248-259 e212. doi:10.1016/j.cell.2016.09.013

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem, 25*(13), 1605-1612. doi:10.1002/jcc.20084

Raman, R., Raguram, S., Venkataraman, G., Paulson, J. C., & Sasisekharan, R. (2005). Glycomics: an integrated systems approach to structure-function relationships of glycans. *Nat Methods, 2*(11), 817-824. doi:10.1038/nmeth807

Ramirez, A. S., Boilevin, J., Biswas, R., Gan, B. H., Janser, D., Aebi, M., . . . Locher, K. P. (2017). Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs. *Glycobiology, 27*(6), 525-535. doi:10.1093/glycob/cwx017

Rothblatt, J. A., & Meyer, D. I. (1986). Secretion in yeast: reconstitution of the translocation and glycosylation of alpha-factor and invertase in a homologous cell-free system. *Cell, 44*(4), 619-628. doi:0092-8674(86)90271-0

Rudd, P. M., & Dwek, R. A. (1997). Glycosylation: heterogeneity and the 3D structure of proteins. *Crit Rev Biochem Mol Biol, 32*(1), 1-100. doi:10.3109/10409239709085144

Rudd, P. M., Elliott, T., Cresswell, P., Wilson, I. A., & Dwek, R. A. (2001). Glycosylation and the immune system. *Science, 291*(5512), 2370-2376.

Salehi, A. S., Smith, M. T., Bennett, A. M., Williams, J. B., Pitt, W. G., & Bundy, B. C. (2016). Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system. *Biotechnol J, 11*(2), 274-281. doi:10.1002/biot.201500237

Schoborg, J. A., Hershewe, J. M., Stark, J. C., Kightlinger, W., Kath, J. E., Jaroentomeechai, T., . . . Jewett, M. C. (2018). A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. *Biotechnol Bioeng, 115*(3), 739-750. doi:10.1002/bit.26502

Schwarz, F., Huang, W., Li, C., Schulz, B. L., Lizak, C., Palumbo, A., . . . Wang, L. X. (2010). A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat Chem Biol, 6*(4), 264-266. doi:10.1038/nchembio.314

Schwarz, F., Lizak, C., Fan, Y. Y., Fleurkens, S., Kowarik, M., & Aebi, M. (2011). Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology, 21*(1), 45-54. doi:10.1093/glycob/cwq130

Shibutani, M., Kim, E., Lazarovici, P., Oshima, M., & Guroff, G. (1996). Preparation of a cell-free translation system from PC12 cell. *Neurochem Res, 21*(7), 801-807.

Sinclair, A. M., & Elliott, S. (2005). Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. *J Pharm Sci, 94*(8), 1626-1635.

Srichaisupakit, A., Ohashi, T., Misaki, R., & Fujiyama, K. (2015). Production of initial-stage eukaryotic N-glycan and its protein glycosylation in Escherichia coli. *J Biosci Bioeng, 119*(4), 399-405. doi:10.1016/j.jbiosc.2014.09.016

Stapleton, J. A., & Swartz, J. R. (2010). Development of an in vitro compartmentalization screen for high-throughput directed evolution of [FeFe] hydrogenases. *PloS one, 5*(12), e15275. doi:10.1371/journal.pone.0015275

Stech, M., Nikolaeva, O., Thoring, L., Stocklein, W. F. M., Wustenhagen, D. A., Hust, M., . . . Kubick, S. (2017). Cell-free synthesis of functional antibodies using a coupled in vitro transcription-translation system based on CHO cell lysates. *Sci Rep, 7*(1), 12030. doi:10.1038/s41598-017-12364-w

Sullivan, C. J., Pendleton, E. D., Sasmor, H. H., Hicks, W. L., Farnum, J. B., Muto, M., . . . Dresios, J. (2016). A cell-free expression and purification process for rapid production of protein biologics. *Biotechnol J, 11*(2), 238-248. doi:10.1002/biot.201500214

Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., & Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *J Vis Exp*(79), e50762. doi:10.3791/50762

Tarui, H., Imanishi, S., & Hara, T. (2000). A novel cell-free translation/glycosylation system prepared from insect cells. *J Biosci Bioeng, 90*(5), 508-514. doi:S1389-1723(01)80031-1 [pii]

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y. Y., Reading, C. A., Chhiba, K., . . . DeLisa, M. P. (2012). An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat Chem Biol, 8*(5), 434-436. doi:10.1038/nchembio.921

Weerapana, E., & Imperiali, B. (2006). Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology, 16*(6), 91R-101R.

Wolfert, M. A., & Boons, G. J. (2013). Adaptive immune activation: glycosylation does matter. *Nat Chem Biol, 9*(12), 776-784. doi:10.1038/nchembio.1403

2.8 Retrospective.

At the similar time that we published this work, Kubick group also reported cell-free glycosylation system based on *Spodoptera frugiperda* (Sf21)[7] that yielded *N*-glycosylated human EPO. As several cell-free protein synthesis platforms have become available, we anticipate these cell-free platforms will soon be integrated with various glycosylation system. Indeed, following the publication of this Chapter, we have further expanded the capability of the CFGpS system to biosynthesis *O*-linked glycoprotein with authentic human *O*-glycans[8]. Our collaborator has also recently reported the improvement of the quality of CFGpS lysate by optimizing lysate generation protocol which leads to an increasing accumulation of membrane vesicle enriched with OST enzyme[9]. In parallel to the OST-based cell-free glycosylation, sequential *N*-glycosylation system has also been reconstituted shortly after this work[10]. Together, these works in developing cell-free glycosylation systems represent an important stepping stone for glycosylation pathway reconstitution and are anticipated to increase an interest in using cell-free system for a study and application in glycobiology.

---

[7] Zemella, A., Thoring, L., Hoffmeister, C. *et al.* Cell-free protein synthesis as a novel tool for directed glycoengineering of active erythropoietin. *Sci Rep* **8,** 8514 (2018).

[8] Natarajan, A., Jaroentomeechai, T., Cabrera-Sánchez, M. *et al.* Engineering orthogonal human *O*-linked glycoprotein biosynthesis in bacteria. *Nat Chem Biol* **16,** 1062–1070 (2020).

[9] Hershewe, J.M., Warfel, K.F., Iyer, S.M. *et al.* Improving cell-free glycoprotein synthesis by characterizing and enriching native membrane vesicles. *Nat Commun* **12,** 2363 (2021).

[10] Kightlinger, W., Duncker, K.E., Ramesh, A. *et al.* A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nat Commun* **10,** 5404 (2019).

CHAPTER 3

ON-DEMAND, CELL-FREE BIOMANUFACTURING OF PROTECTIVE CONJUGATE VACCINES AT THE POINT-OF-CARE[11]

3.1 Abstract.

Conjugate vaccines are among the most effective methods for preventing bacterial infections. However, existing manufacturing approaches limit access to conjugate vaccines due to centralized production and cold chain distribution requirements. To address these limitations, we developed a modular technology for *in vitro* conjugate vaccine expression (iVAX) in portable, freeze-dried lysates from detoxified, nonpathogenic *Escherichia coli.* Upon rehydration, iVAX reactions synthesize clinically relevant doses of conjugate vaccines against diverse bacterial pathogens in one hour. We show that iVAX-synthesized vaccines against *Franciscella tularensis* subsp. *tularensis* (type A) strain Schu S4 protected mice from lethal intranasal *F. tularensis* challenge. The iVAX platform promises to accelerate development of new conjugate vaccines with increased access through refrigeration-independent distribution and portable production.

11 This chapter appears in the Science Advances journal:

Stark, J.C.*, Jaroentomeechai, T.*, Moeller, T.D., Dubner, R.S., Hsu, K.J., Stevenson, T.C., DeLisa, M.P., and Jewett, M.C. (2021) On-demand, cell-free biomanufacturing of conjugate vaccines at the point-of-care. *Sci Adv*. 7: 1-15.

3.2 Introduction.

Drug-resistant bacteria are predicted to threaten up to 10 million lives per year by 2050 (The Review on Antimicrobial Resistance, 2014), necessitating new strategies to develop and distribute antibiotics and vaccines. Conjugate vaccines, typically composed of a pathogen-specific capsular (CPS) or O-antigen polysaccharide (O-PS) linked to an immunostimulatory protein carrier, are among the safest and most effective methods for preventing life-threatening bacterial infections (Jin et al., 2017; Trotter et al., 2008; Weintraub, 2003). In particular, implementation of meningococcal and pneumococcal conjugate vaccines have significantly reduced the occurrence of bacterial meningitis and pneumonia worldwide (Novak et al., 2012; Poehling et al., 2006), in addition to reducing the frequency of antibiotic resistance in targeted strains (Roush et al., 2008). However, despite their proven safety and efficacy, global childhood vaccination rates for conjugate vaccines remain as low as ~30%, with lack of access or low immunization coverage accounting for the vast majority of remaining disease burden (Wahl et al., 2018). In addition, the 2018 WHO prequalification of Typhbar-TCV® to prevent typhoid fever represents the first conjugate vaccine approval in nearly a decade. In order to address emerging drug-resistant pathogens, new platform technologies to accelerate the development and global distribution of conjugate vaccines are urgently needed.

Contributing to the slow pace of conjugate vaccine development and distribution is the fact that these molecules are particularly challenging and costly to

manufacture. The conventional process to produce conjugate vaccines involves chemical conjugation of carrier proteins with polysaccharide antigens purified from large-scale cultures of pathogenic bacteria. Large-scale fermentation of pathogens results in high manufacturing costs due to associated biosafety hazards and process development challenges. In addition, chemical conjugation can alter the structure of the polysaccharide, resulting in loss of the protective epitope (Bhushan et al., 1998). To address these challenges, it was recently demonstrated that polysaccharide-protein conjugates can be made in *Escherichia coli* using protein-glycan coupling technology (PGCT) (Feldman et al., 2005a). In this approach, engineered *E. coli* cells covalently attach heterologously expressed CPS or O-PS antigens to carrier proteins via an asparagine-linked glycosylation reaction catalyzed by the *Campylobacter jejuni* oligosaccharyltransferase enzyme PglB (*Cj*PglB) (Cuccui et al., 2013a; Garcia-Quintanilla et al., 2014; Ihssen et al., 2010; Ma et al., 2014; Marshall et al., 2018b; Wacker et al., 2014; Wetter et al., 2013). Despite this advance, both chemical conjugation and PGCT approaches rely on living bacterial cells, requiring centralized production facilities from which vaccines are distributed via a refrigerated supply chain.

Refrigeration of conjugate vaccines is critical to avoid spoilage due to aggregate formation and significant loss of the pathogen-specific polysaccharide upon heating and freezing (Beresford et al., 2017; Frasch, 2009; Gao et al., 2014; Ho et al., 2000; Ho et al., 2002; WHO, 2014). Due to complexities and costs associated with cold chain refrigeration, vaccines with even short-term thermostability offer significant

advantages. The availability of effective and thermostable freeze-dried vaccines is cited as a key technological innovation that enabled the global eradication of smallpox, the only infectious disease to be eradicated to date (Hopkins, 1988). Development of MenAfriVac™, a meningococcal conjugate vaccine shown to remain active outside of the cold chain for up to 4 days, enabled increased vaccine coverage and an estimated 50% reduction in costs during vaccination in the meningitis belt of sub-Saharan Africa (Lydon et al., 2014). However, this required significant ($70M) investment in the development and validation of a thermostable vaccine. Further, conjugate vaccine thermostability varies for different O-PS antigens both across pathogens and between serotypes of the same pathogen, even in lyophilized formulations (Beresford et al., 2017; Gao et al., 2014; Ho et al., 2000; Ho et al., 2002). Thus, generalizable strategies to achieve thermostability in the context of current manufacturing and distribution strategies may well prove elusive. Broadly, the need for cold chain refrigeration creates economic and logistical challenges that limit the reach of vaccination campaigns and present barriers to the eradication of disease, especially in low and middle income countries (Ashok et al., 2017; Wahl et al., 2018).

Cell-free protein synthesis (CFPS) offers opportunities to both accelerate vaccine development and enable decentralized, cold chain-independent biomanufacturing by using cell lysates, rather than living cells, to synthesize proteins *in vitro* (Silverman et al., 2020). Importantly, CFPS platforms (i) enable point-of-care protein production, as relevant amounts of protein can be synthesized *in vitro* in just a

few hours, (ii) can be freeze-dried for distribution at ambient temperature and reconstituted by just adding water (Pardee et al., 2016b), and (iii) circumvent biosafety concerns associated with the use of living cells outside of a controlled laboratory setting. CFPS has recently been used to enable on-demand and portable production of aglycosylated protein subunit vaccines (Adiga et al., 2018; Pardee et al., 2016b). However, the production of efficacious glycoprotein products, which represent 70% of approved therapeutics (Sethuraman and Stadheim, 2006), from decentralized biomanufacturing platforms has not yet been demonstrated. As a result, there remains a need for additional technologies that enable decentralized production of glycosylated protein products, including conjugate vaccines.

To address this technological gap, here we describe the iVAX (*in vitro* conjugate vaccine expression) platform that enables rapid development and cold chain-independent biosynthesis of conjugate vaccines in cell-free reactions (**Figure 3-1**). iVAX was designed to have the following features. First, iVAX is fast, with the ability to produce multiple individual doses of conjugates in one hour. Second, iVAX is robust, yielding equivalent amounts of conjugate over a range of operating temperatures. Third, iVAX is modular, offering the ability to rapidly interchange carrier proteins, including those used in licensed conjugate vaccines, as well as conjugated polysaccharide antigens. We leverage this modularity to create an array of vaccine candidates targeted against diverse bacterial pathogens, including the highly virulent *Franciscella tularensis* subsp. *tularensis* (type A) strain Schu S4, enterotoxigenic (ETEC)

E. coli O78, and uropathogenic (UPEC) *E. coli* O7. Fourth, iVAX is shelf-stable, derived from freeze-dried cell-free reactions that operate in a just-add-water strategy. Fifth, iVAX is safe, leveraging lipid A engineering that effectively avoids the high levels of endotoxin present in wild-type *E. coli*. Our results demonstrate that a *F. tularensis* O-PS conjugate derived from freeze-dried, low-endotoxin iVAX reactions outperformed a conjugate produced using the established cell-based PGCT approach in its ability to elicit pathogen-specific antibodies. Moreover, the iVAX-derived conjugate afforded complete protection in a mouse model of intranasal *F. tularensis* infection. Overall, the iVAX platform offers a new way to deliver the protective benefits of an important class of antibacterial vaccines to both the developed and developing world.

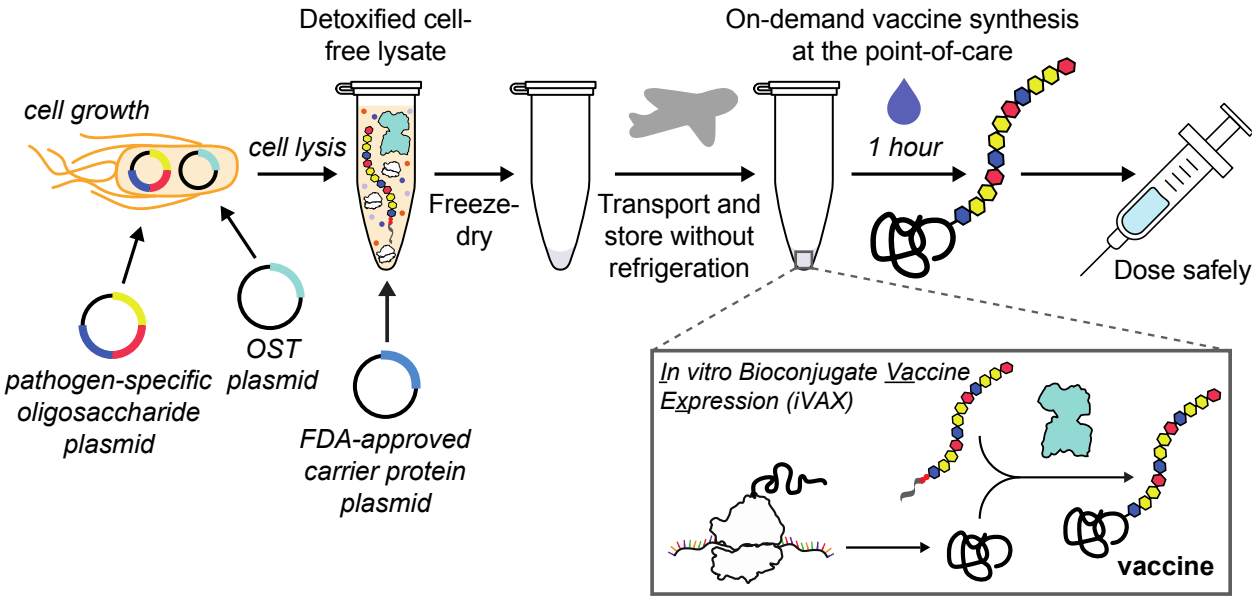


Figure 3-1. The iVAX platform enables on-demand and portable production of antibacterial vaccines. The *in vitro* vaccine expression (iVAX) platform provides a rapid means to develop and distribute conjugate vaccines against bacterial pathogens. Expression of pathogen-specific polysaccharides (e.g., CPS, O-PS) and a bacterial oligosaccharyltransferase enzyme in

engineered nonpathogenic *E. coli* with detoxified lipid A yields low-endotoxin lysates containing all of the machinery required for synthesis of bioconjugate vaccines. Reactions catalyzed by iVAX lysates can be used to produce bioconjugates containing licensed carrier proteins and can be freeze-dried without loss of activity for refrigeration-free transportation and storage. Freeze-dried reactions can be activated at the point-of-care via simple rehydration and used to reproducibly synthesize immunologically active bioconjugates in ~1 h.

3.3 Results.

3.3.1 *In vitro* synthesis of licensed vaccine carrier proteins.

To demonstrate proof-of-principle for cell-free bioconjugate vaccine production, we first set out to express a set of carrier proteins that are currently used in approved conjugate vaccines. Producing these carrier proteins in soluble conformations *in vitro* represented an important benchmark because their expression in living *E. coli* has proven challenging, often requiring multi-step purification and refolding of insoluble product from inclusion bodies (Haghi et al., 2011b; Stefan et al., 2011b), fusion of expression partners such as maltose-binding protein (MBP) to increase soluble expression (Figueiredo et al., 1995a; Stefan et al., 2011b), or expression of truncated protein variants in favor of the full-length proteins (Figueiredo et al., 1995a). In contrast, cell-free protein synthesis approaches have recently shown promise for difficult-to-express proteins (Perez et al., 2016a). The carrier proteins that we focused on here included nonacylated *H. influenzae* protein D (PD), the *N. meningitidis* porin protein (PorA), and genetically detoxified variants of the *Corynebacterium diphtheriae* toxin (CRM197) and the *Clostridium tetani* toxin (TT). We also tested

expression of the fragment C (TTc) and light chain (TTlight) domains of TT as well as *E. coli* MBP. While MBP is not a licensed carrier, it has demonstrated immunostimulatory properties (Fernandez et al., 2007) and when linked to O-PS was found to elicit polysaccharide-specific humoral and cellular immune responses in mice (Ma et al., 2014). Similarly, the TT domains, TTlight and TTc, have not been used in licensed vaccines, but are immunostimulatory and individually sufficient for protection against *C. tetani* challenge in mice (Figueiredo et al., 1995a). To enable glycosylation, all carriers were modified at their C-termini with 4 tandem repeats of an optimal bacterial glycosylation motif, DQNAT (Chen et al., 2007). A C-terminal 6xHis tag was also included to enable purification and detection via Western blot analysis. A variant of superfolder green fluorescent protein that contained an internal DQNAT glycosylation site (sfGFP$^{217\text{-}DQNAT}$) (Jaroentomeechai et al., 2018a) was used as a model protein to facilitate system development.

All eight carriers were synthesized *in vitro* with soluble yields of ~50.0-650 μg mL^{-1} as determined by ^{14}C-leucine incorporation (**Figure 3-2a**). In particular, the MBP^{4xDQNAT} and PD^{4xDQNAT} variants were nearly 100% soluble, with yields of ~500 μg mL^{-1} and 200 μg mL^{-1}, respectively, and expressed as exclusively full-length products according to Western blot and autoradiogram analysis (**Figure 3-2b, Supplementary Figure 7-1a**). Notably, similar soluble yields were observed for all carriers at 25°C, 30°C, and 37°C, with the exception of CRM197^{4xDQNAT} (**Supplementary Figure 7-1b**), which is known to be heat sensitive (WHO, 2014). These results suggest that our

method of cell-free carrier biosynthesis is robust over a 13°C range in temperature and could be used in settings where precise temperature control is not feasible.

The open reaction environment of our cell-free reactions enabled facile manipulation of the chemical and reaction environment to improve production of more complex carriers. For example, in the case of the membrane protein PorA^{4xDQNAT}, lipid nanodiscs were added to increase soluble expression (**Supplementary Figure 7-1c**). Nanodiscs provide a cellular membrane mimic to co-translationally stabilize hydrophobic regions of membrane proteins (Bayburt and Sligar, 2010). For expression of TT, which contains an intermolecular disulfide bond, expression was carried out for 2 hours in oxidizing conditions (Knapp et al., 2007), which improved assembly of the heavy and light chains into full-length product and minimized protease degradation of full-length TT (**Supplementary Figure 7-1d**). *In vitro* synthesized CRM197^{4xDQNAT} and TT^{4xDQNAT} were comparable in size to commercially available purified diphtheria toxin (DT) and TT protein standards and were reactive with α-DT and α-TT antibodies, respectively (**Supplementary Figure 7-1e, f**), indicating that both were produced in immunologically relevant conformations. This is notable as CRM197 and TT are FDA-approved vaccine antigens for diphtheria and tetanus, respectively, when they are administered without conjugated polysaccharides. Together, our results highlight the ability of CFPS to express licensed conjugate vaccine carrier proteins in soluble conformations over a range of temperatures.

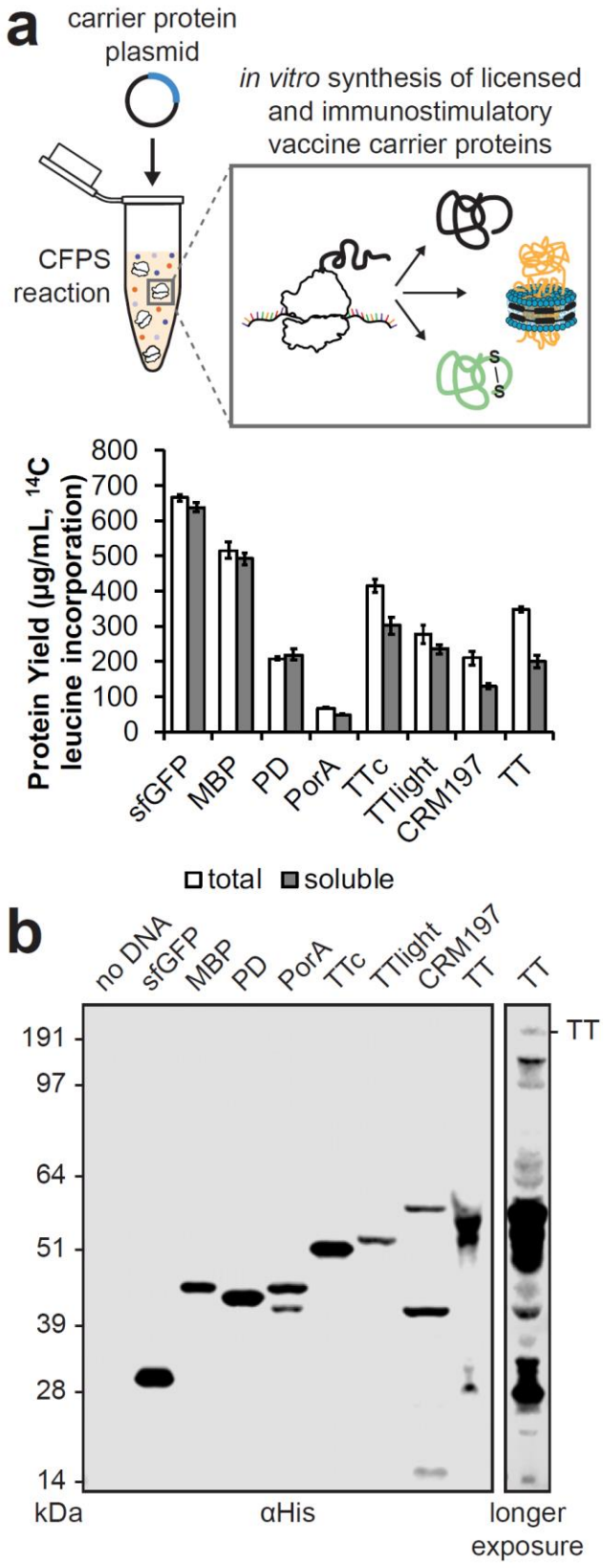

a
carrier protein plasmid
in vitro synthesis of licensed and immunostimulatory vaccine carrier proteins
CFPS reaction
S
S
Protein Yield (µg/mL, 14C leucine incorporation)
800
700
600
500
400
300
200
100
0
sfGFP
MBP
PD
PorA
TTc
TTlight
CRM197
TT
total
soluble
b
no DNA
sfGFP
MBP
PD
PorA
TTc
TTlight
CRM197
TT
TT
191
97
64
51
39
28
14
kDa
αHis
longer exposure

Figure 3-2. *In vitro* synthesis of licensed conjugate vaccine carrier proteins. (**a**) All four carrier proteins used in FDA-approved conjugate vaccines were synthesized solubly *in vitro*, as measured via ^{14}C-leucine incorporation. These include *H. influenzae* protein D (PD), the *N. meningitidis* porin protein (PorA), and genetically detoxified variants of the *C. diphtheriae* toxin (CRM197) and the *C. tetani* toxin (TT). Additional immunostimulatory carriers were also synthesized solubly, including *E. coli* maltose binding protein (MBP) and the fragment C (TTc) and light chain (TTlight) domains of TT. Values represent means and error bars represent standard deviations of biological replicates (n = 3). (**b**) Full length product was observed for all proteins tested via Western blot. Different exposures are indicated with solid lines. Molecular weight ladder is shown at left.

3.3.2 On-demand synthesis of bioconjugate vaccines.

We next sought to synthesize polysaccharide-conjugated versions of these carrier proteins by merging their *in vitro* expression with single-pot, cell-free glycosylation. As a model vaccine target, we focused on the highly virulent *Francisella tularensis* subsp. *tularensis* (type A) strain Schu S4, a gram-negative, facultative coccobacillus and the causative agent of tularemia. This bacterium is categorized as a class A bioterrorism agent due to its high fatality rate, low dose of infection, and ability to be aerosolized (Oyston et al., 2004). Vaccines targeting *F. tularensis* will likely need to be deployed rapidly using ring vaccination strategies in response to an outbreak, similar to those planned in the event of a smallpox attack (CDC, 2019b) and used to eradicate smallpox in the 1960s and 70s (Hopkins, 1988). We thus sought to develop rapidly deployable, thermostable conjugate vaccines against *F. tularensis* as an initial demonstration of the iVAX technology.

Although there are currently no licensed vaccines against *F. tularensis*, several studies have independently confirmed the important role of antibodies directed against *F. tularensis* LPS, specifically the O-PS repeat unit, in providing protection against the Schu S4 strain (Fulop et al., 2001; Lu et al., 2012). More recently, a bioconjugate vaccine comprising the *F. tularensis* Schu S4 O-PS (*Ft*O-PS) conjugated to the *Pseudomonas aeruginosa* exotoxin A ($EPA^{DNNNS\text{-}DQNRT}$) carrier protein produced using PGCT (Cuccui et al., 2013b; Marshall et al., 2018a) was shown to be protective against challenge with the Schu S4 strain in a rat inhalation model of tularemia (Marshall et al., 2018a). In light of these earlier findings, we investigated the ability of the iVAX platform to produce anti-*F. tularensis* bioconjugate vaccine candidates on-demand by conjugating the *Ft*O-PS structure to diverse carrier proteins *in vitro*.

The *Ft*O-PS is composed of the 826-Da repeating tetrasaccharide unit Qui4NFm-$(GalNAcAN)_2$-QuiNAc (Qui4NFm: 4,6-dideoxy-4-formamido-D-glucose; GalNAcAN: 2-acetamido-2-deoxy-D-galacturonamide; QuiNAc: 2-acetamido-2,6-dideoxy-D-glucose) (Prior et al., 2003b). We and other groups have previously shown that the authentic *Ft*O-PS structure can be synthesized in K-12 strains of *E. coli* via recombinant expression of the *Ft*O-PS biosynthetic pathway (Chen et al., 2016b; Cuccui et al., 2013a; Prior et al., 2003a). To glycosylate proteins with *Ft*O-PS, we produced an iVAX lysate from glycoengineered *E. coli* cells expressing the *Ft*O-PS biosynthetic pathway and the oligosaccharyltransferase enzyme *Cj*PglB (**Figure 3-3a**). This lysate, which contained lipid-linked *Ft*O-PS and active *Cj*PglB, was used to catalyze iVAX

reactions primed with plasmid DNA encoding sfGFP$^{217\text{-}DQNAT}$. Control reactions in which attachment of the *Ft*O-PS was not expected were performed with lysates from cells that lacked either the *Ft*O-PS pathway or the *Cj*PglB enzyme. We also tested reactions that lacked plasmid encoding the target protein sfGFP$^{217\text{-}DQNAT}$ or were primed with plasmid encoding sfGFP$^{217\text{-}AQNAT}$, which contained a mutated glycosylation site (AQNAT) that is not modified by *Cj*PglB (Kowarik et al., 2006c). In reactions containing the iVAX lysate and primed with plasmid encoding sfGFP$^{217\text{-}DQNAT}$, immunoblotting with anti-His antibody or a commercial monoclonal antibody specific to *Ft*O-PS revealed a ladder-like banding pattern (**Figure 3-3b**). This ladder is characteristic of *Ft*O-PS attachment, resulting from O-PS chain length variability through the action of the Wzy polymerase (Cuccui et al., 2013b; Feldman et al., 2005b; Prior et al., 2003b). Glycosylation of sfGFP$^{217\text{-}DQNAT}$ was observed only in reactions containing a complete glycosylation pathway and the preferred DQNAT glycosylation sequence (**Figure 3-3b**). This glycosylation profile was further reproducible across biological replicates from the same lot of lysate (**Figure 3-3c, left**) and using different lots of lysate (**Figure 3-3c, right**), with an average efficiency of conjugation with *Ft*O-PS of 69 ± 5% by densitometry analysis. *In vitro* protein synthesis and glycosylation was observed after 1 hour, with the amount of conjugated polysaccharide reaching a maximum between 0.75 and 1.25 hours (**Supplementary Figure 7-2**). Similar glycosylation reaction kinetics were observed at 37°C, 30°C, 25°C, and room temperature (~21°C), indicating that iVAX reactions are robust over a range of temperatures (**Supplementary Figure 7-2**).

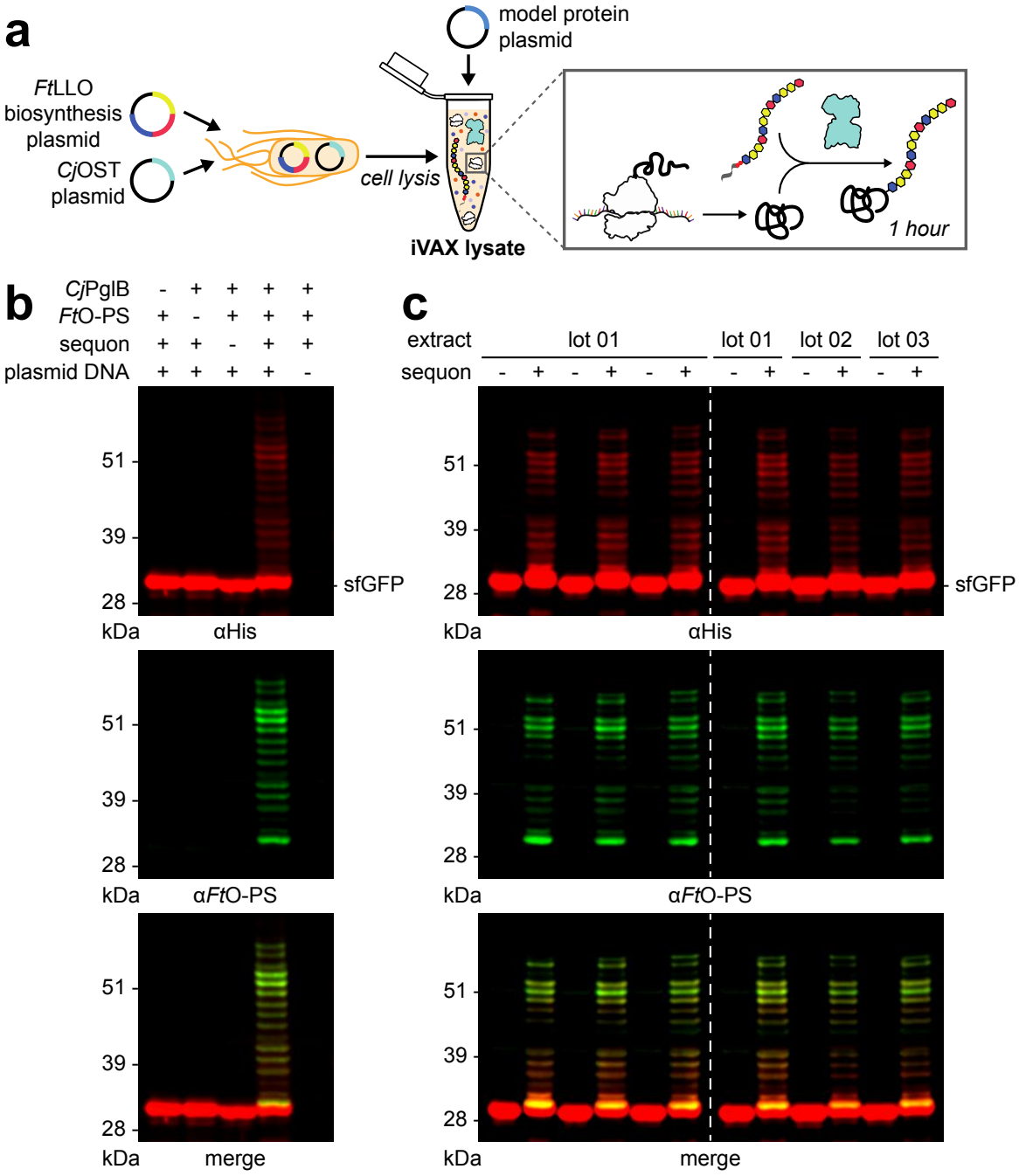


Figure 3-3. Reproducible glycosylation of proteins with *Ft*O-PS in iVAX lysates. (**a**) iVAX lysates were prepared from cells expressing *Cj*PglB and a biosynthetic pathway encoding *Ft*O-PS. (**b**) Glycosylation of sfGFP$^{217\text{-DQNAT}}$ with *Ft*O-PS was only observed when *Cj*PglB, *Ft*O-PS, and the preferred glycosylation sequence (sequon) were present in the reaction (lane 3). When plasmid DNA was omitted, sfGFP$^{217\text{-DQNAT}}$ synthesis was not observed. (**c**) Biological replicates of iVAX reactions producing sfGFP$^{217\text{-DQNAT}}$ using the same lot (**left**) or different lots (**right**) of

iVAX lysates demonstrated reproducibility of reactions and lysate preparation. Top panels show signal from probing with anti-hexa-histidine antibody (αHis) to detect the carrier protein, middle panels show signal from probing with commercial anti-*Ft*O-PS antibody (α*Ft*O-PS), and bottom panels show αHis and α*Ft*O-PS signals merged. Unless replicates are explicitly shown, images are representative of at least three biological replicates. Dashed lines indicate samples are from the same blot with the same exposure. Molecular weight ladders are shown at the left of each image.

Next, we investigated whether immunologically relevant carriers could be similarly conjugated with *Ft*O-PS in iVAX reactions. Following addition of plasmid DNA encoding MBP^{4xDQNAT}, PD^{4xDQNAT}, PorA^{4xDQNAT}, TTc^{4xDQNAT}, TTlight^{4xDQNAT}, CRM197^{4xDQNAT}, or the most common PGCT carrier protein, EPA$^{DNNNS-DQNRT}$ (Cuccui et al., 2013a; Ihssen et al., 2010; Marshall et al., 2018b; Wacker et al., 2014; Wetter et al., 2013), glycosylation of each with *Ft*O-PS was observed for iVAX reactions enriched with lipid-linked *Ft*O-PS and *Cj*PglB but not control reactions lacking *Cj*PglB (**Figure 3-4**). Notably, our attempts to synthesize the same panel of conjugates using the established PGCT approach in living *E. coli* yielded less promising results. Specifically, only limited expression of conjugates composed of PorA and CRM197, two of the carriers used in licensed conjugate vaccines, were achieved *in vivo* (**Supplementary Figure 7-3**). Collectively, these data indicate that iVAX may provide advantages over the established PGCT approach for production of conjugate vaccine candidates composed of diverse and potentially membrane-bound carrier proteins.

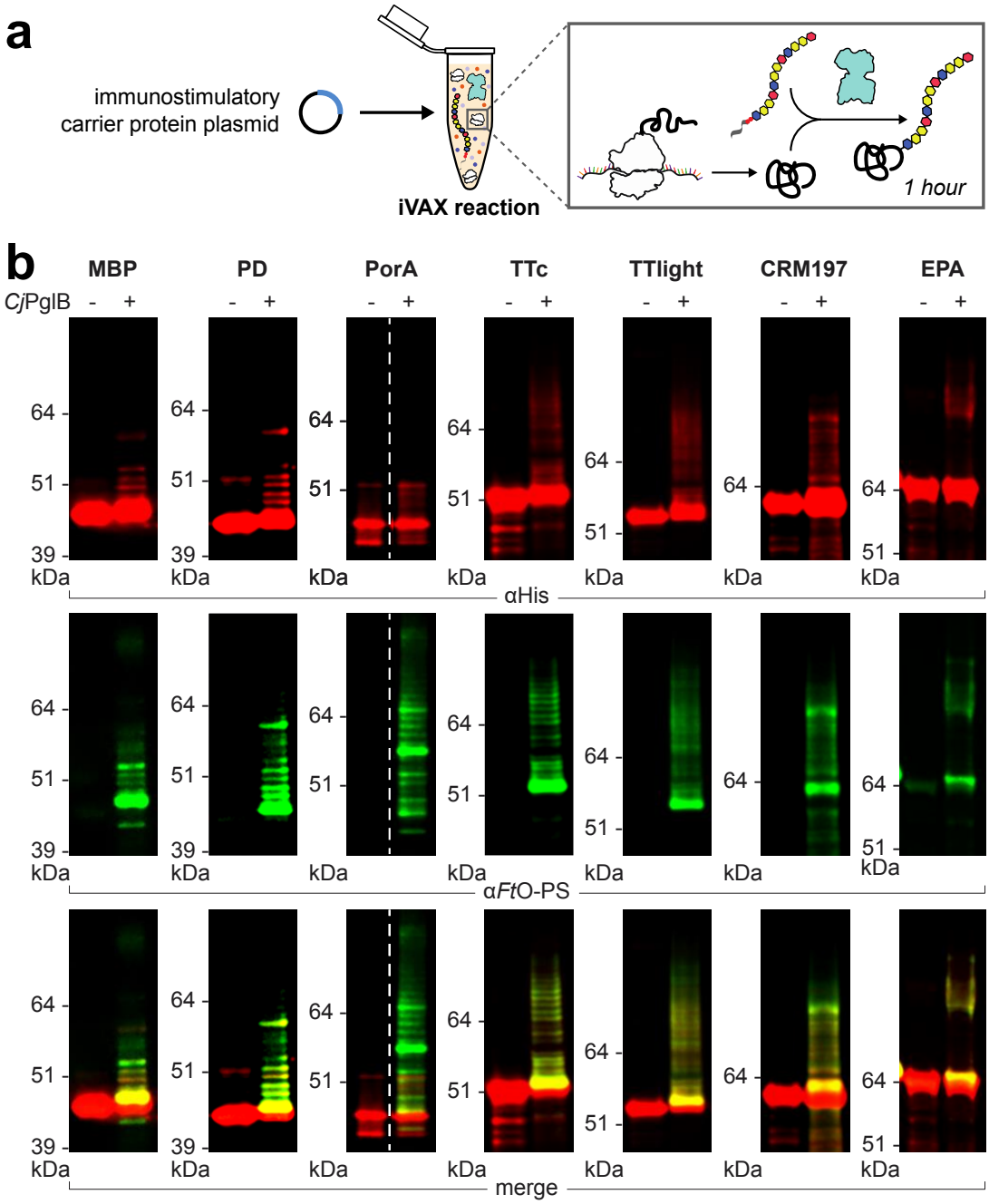


Figure 3-4. On-demand production of bioconjugates against *F. tularensis* using iVAX. (**a**) iVAX reactions were prepared from lysates containing *Cj*PglB and *Ft*O-PS and primed with plasmid encoding immunostimulatory carriers, including those used in licensed vaccines. (**b**) We observed on-demand synthesis of anti-*F. tularensis* bioconjugate vaccines for all carrier proteins tested. Bioconjugates were purified using Ni-NTA agarose from 1 mL iVAX reactions lasting ~1 h. Top panels show signal from probing with anti-hexa-histidine antibody (*α*His)

to detect the carrier protein, middle panels show signal from probing with commercial anti-*Ft*O-PS antibody (*αFt*O-PS), and bottom panels show αHis and *αFt*O-PS signals merged. Images are representative of at least three biological replicates. Dashed lines indicate samples are from the same blot with the same exposure. Molecular weight ladders are shown at the left of each image.

We next asked whether the yields of conjugates produced using iVAX were sufficient to enable production of relevant vaccine doses. To assess expression titers, we focused on $MBP^{4xDQNAT}$ and $PD^{4xDQNAT}$ because these carriers expressed *in vitro* with high soluble titers and as exclusively full-length protein (**Figure 3-2, Supplementary Figure 7-1a**). In addition, PD has been shown to be a safe and effective conjugate vaccine carrier protein (Palmu et al., 2013; Silfverdal et al., 2009) and may have advantages over DT and TT in generating robust immune responses to polysaccharide antigens (Burrage et al., 2002; Dagan et al., 1998; Dagan et al., 2004; Knuf et al., 2009). Recent clinical data show 1-10 μg doses of conjugate vaccine candidates are well-tolerated and effective in stimulating the production of antibacterial IgGs (Hatz et al., 2015; Huttner et al., 2017; Riddle et al., 2016). We found that reactions lasting ~1 hour produced ~20 μg mL^{-1}, or two 10-μg doses mL^{-1}, of *Ft*O-PS-conjugated $MBP^{4xDQNAT}$ and $PD^{4xDQNAT}$ as determined by ^{14}C-leucine incorporation and densitometry analysis (**Supplementary Figure 7-4a**). It should be noted that vaccines are currently distributed in vials containing 1-20 doses of vaccine to minimize wastage (Humphreys, 2011). Our yields indicate that multiple doses per mL can be synthesized in 1 hour using the iVAX platform.

To demonstrate the modularity of the iVAX approach for bioconjugate production, we sought to produce bioconjugates bearing O-PS antigens from additional pathogens including ETEC *E. coli* strain O78 and UPEC *E. coli* strain O7. *E. coli* O78 is a major cause of diarrheal disease in developing countries, especially among children, and a leading cause of traveler's diarrhea (Qadri et al., 2005), while the O7 strain is a common cause of urinary tract infections (Johnson, 1991). Like the *Ft*O-PS, the biosynthetic pathways for *Ec*O78-PS and *Ec*O7-PS have been described previously and confirmed to produce O-PS antigens with the repeating units $GlcNAc_2Man_2$ (Jansson et al., 1987) and Qui4NAcMan(Rha)GalGlcNAc (L'vov et al., 1984) (GlcNAc: *N*-acetylglucosamine; Man: mannose; Qui4NAc: 4-acetamido-4,6-dideoxy-D-glucopyranose; Rha: rhamnose; Gal: galactose), respectively. Using iVAX lysates from cells expressing *Cj*PglB and either the *Ec*O78-PS and *Ec*O7-PS pathways in reactions that were primed with $PD^{4xDQNAT}$ or $sfGFP^{217\text{-}DQNAT}$ plasmids, we observed carrier glycosylation only when both lipid-linked O-PS and *Cj*PglB were present in the reactions (**Supplementary Figure 7-4b, c**). These results demonstrate modular production of bioconjugates against multiple bacterial pathogens in our iVAX platform, enabled by compatibility of multiple heterologous O-PS pathways with *in vitro* carrier protein synthesis and glycosylation.

3.3.3 Endotoxin editing and freeze-drying yield iVAX reactions that are safe and portable.

A key challenge inherent in using any *E. coli*-based system for biopharmaceutical production is the presence of lipid A, or endotoxin, which is known to contaminate protein products and can cause lethal septic shock at high levels (Russell, 2006). As a result, the amount of endotoxin in formulated biopharmaceuticals is regulated by the United States Pharmacopeia (USP), US Food and Drug Administration (FDA), and the European Medicines Agency (EMEA) (Brito and Singh, 2011b). Because our iVAX reactions rely on lipid-associated components, such as *Cj*PglB and *Ft*O-PS, standard detoxification approaches involving the removal of lipid A (Petsch and Anspach, 2000) could compromise the activity or concentration of our glycosylation components in addition to increasing cost and processing complexities.

To address this issue, we adapted a previously reported strategy to detoxify the lipid A molecule through strain engineering (Chen et al., 2016a; Needham et al., 2013). In particular, the deletion of the acyltransferase gene *lpxM* and the overexpression of the *F. tularensis* phosphatase LpxE in *E. coli* has been shown to result in the production of nearly homogenous pentaacylated, monophosphorylated lipid A with significantly reduced toxicity but retained activity as an adjuvant (Chen et al., 2016a). This pentaacylated, monophosphorylated lipid A was structurally identical to the primary component of monophosphoryl lipid A (MPL) from *Salmonella minnesota* R595, an approved adjuvant composed of a mixture of monophosphorylated lipids (Casella and

Mitchell, 2008). To generate detoxified lipid A structures in the context of iVAX, we produced lysates from a Δ*lpxM* derivative of CLM24 that co-expressed *Ft*LpxE and the *Ft*O-PS glycosylation pathway (**Figure 3-5a**). Lysates derived from this chassis synthesize a detoxified pentaacylated, monophosphorylated lipid A molecule and exhibited significantly decreased levels of toxicity compared to strains expressing both unmodified (WT) or diphosphorylated pentaacylated lipid A (Δ*lpxM*) structures (**Figure 3-5b**) as measured by human TLR4 activation in HEK-Blue hTLR4 reporter cells (Needham et al., 2013). Importantly, the structural remodeling of lipid A did not affect the activity of the membrane-bound *Cj*PglB and *Ft*O-PS components in iVAX reactions (**Supplementary Figure 7-5a**). By engineering the chassis strain for lysate production, we produced iVAX lysates with endotoxin levels <1,000 EU mL^{-1}, within the range of reported values for commercial protein-based vaccine products (0.288-180,000 EU mL^{-1}) (Brito and Singh, 2011b). Further, *Ft*O-PS conjugates synthesized and affinity purified from detoxified iVAX lysates contained 0.21 ± 0.3 EU per 10-μg dose, which is well below endotoxin levels reported in commercial conjugate vaccines (<12 EU/dose) (Bolgiano et al., 2007; Brito and Singh, 2011a) (**Figure 3-5c**).

A major limitation of traditional conjugate vaccines is that they must be refrigerated (WHO, 2014), making it difficult to distribute these vaccines to remote or resource-limited settings. The ability to freeze-dry iVAX reactions for ambient temperature storage and distribution could alleviate the logistical challenges associated with refrigerated supply chains that are required for existing vaccines. To

investigate this possibility, detoxified iVAX lysates were used to produce *Ft*O-PS bioconjugates in two different ways: either by running the reaction immediately after priming with plasmid encoding the sfGFP$^{217\text{-}DQNAT}$ target protein or by running after the same reaction mixture was lyophilized and rehydrated (**Figure 3-5d**). In both cases, conjugation of *Ft*O-PS to sfGFP$^{217\text{-}DQNAT}$ was observed when *Cj*PglB was present, with modification levels that were nearly identical (average glycosylation efficiency of sfGFP$^{217\text{-}DQNAT}$ in reactions with and without lyophilization were 66 ± 7% and 69 ± 5% by densitometry, respectively) (**Figure 3-3c, 3-5e,** and **Supplementary Figure 7-6**). We also showed that detoxified, freeze-dried iVAX reactions can be scaled to 5 mL for production of *Ft*O-PS-conjugated MBP^{4xDQNAT} and PD^{4xDQNAT} in a manner that was reproducible from lot to lot and indistinguishable from production without freeze-drying (average glycosylation efficiencies of MBP^{4xDQNAT} and PD^{4xDQNAT} were 66 ± 10% and 70 ± 1% by densitometry, respectively) (**Supplementary Figure 7-5b, c**). In addition, freeze-dried reactions are stable under ambient temperature storage for at least 3 months, with no observable differences in protein synthesis or glycosylation activity (**Supplementary Figure 7-6**). The ability to lyophilize iVAX reactions, store reactions at ambient temperature, and manufacture bioconjugates without specialized equipment highlights the potential for portable, on-demand vaccine production.

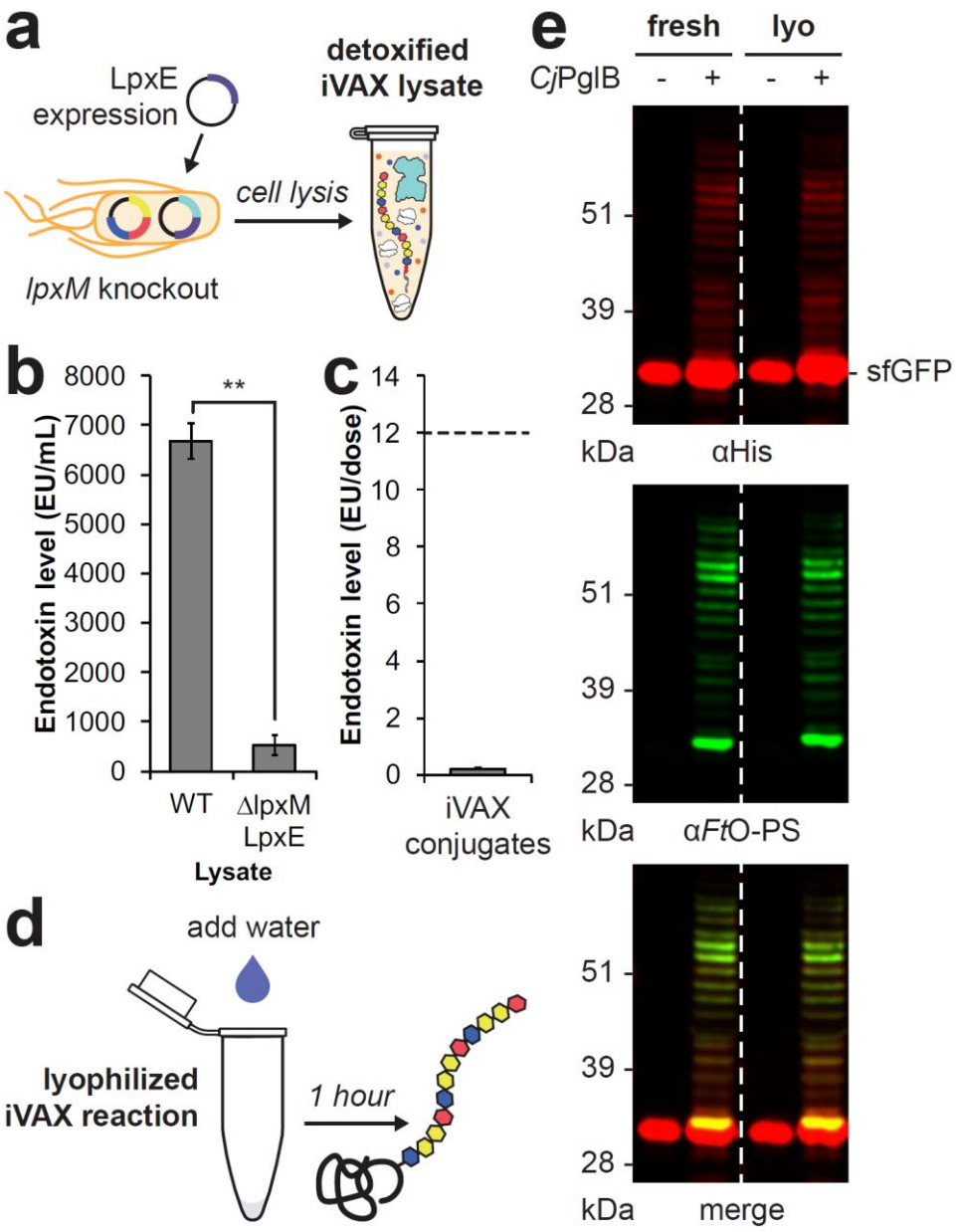


Figure 3-5. Detoxified, lyophilized iVAX reactions produce conjugate vaccines. (**a**) iVAX lysates were detoxified via deletion of *lpxM* and expression of *F. tularensis* LpxE in the source strain for lysate production. (**b**) The resulting lysates exhibited significantly reduced endotoxin activity, as measured by activation of human TLR4 in HEK-Blue hTLR4 reporter cells. ***p* = 0.003, as determined by two-tailed *t*-test. (**c**) *Ft*O-PS conjugate vaccines produced and purified from detoxified iVAX reactions contained 0.21 ± 0.3 EU/10 μg dose, as measured by human

TLR4 activation. Dashed line represents endotoxin levels reported in commercial conjugate vaccines (<12 EU/dose). (**d**) iVAX reactions producing sfGFP$^{217\text{-}DQNAT}$ were run immediately or following lyophilization and rehydration. (**e**) Glycosylation activity was preserved following lyophilization, demonstrating the potential of iVAX reactions for portable biosynthesis of conjugate vaccines. Top panel shows signal from probing with anti-hexa-histidine antibody (αHis) to detect the carrier protein, middle panel shows signal from probing with commercial anti-*Ft*O-PS antibody (α*Ft*O-PS), and bottom panel shows αHis and α*Ft*O-PS signals merged. Images are representative of at least three biological replicates. Molecular weight ladder is shown at the left of each image.

3.3.4 *In vitro* synthesized bioconjugates elicit pathogen-specific antibodies in mice.

We next evaluated the ability of iVAX-derived conjugates to elicit anti-*Ft*LPS antibodies in mice (**Figure 3-6a**). Importantly, we found that BALB/c mice receiving iVAX-derived *Ft*O-PS-conjugated MBP^{4xDQNAT} or PD^{4xDQNAT} produced high titers of *Ft*LPS-specific IgG antibodies, which were significantly elevated compared to the titers measured in the sera of control mice receiving PBS or unmodified versions of each carrier protein (**Figure 3-6b, Supplementary Figure 7-7**). Interestingly, the IgG titers measured in sera from mice receiving PGCT-derived MBP^{4xDQNAT} conjugates were similar to the titers observed in the control groups (**Figure 3-6b, Supplementary Figure 7-7**). Notably, both MBP^{4xDQNAT} and PD^{4xDQNAT} conjugates produced using iVAX elicited similar levels of IgG production and neither resulted in any observable adverse events in mice, confirming the modularity and safety of the technology for production of conjugate vaccine candidates.

We further characterized IgG titers by analysis of IgG1 and IgG2a subtypes and found that both iVAX-derived *Ft*O-PS-conjugated MBP^{4xDQNAT} and PD^{4xDQNAT} boosted production of IgG1 antibodies by >2 orders of magnitude relative to all control groups as well as to PGCT-derived MBP^{4xDQNAT} conjugates (**Figure 3-6c**). Observed IgG subclass titers elicited by iVAX-derived conjugates (IgG1 >> IgG2a) are further consistent with a Th2-biased response, which is characteristic of most conjugate vaccines (Bogaert et al., 2004), though additional studies are needed to confirm this immunological phenotype. Taken together, these results provide evidence that the iVAX platform supplies vaccine candidates that are capable of eliciting strong, pathogen- and polysaccharide-specific humoral immune responses.

3.3.5 iVAX-derived vaccines protect mice from intranasal *F. tularensis* challenge.

Finally, we tested the ability of iVAX-derived vaccines to protect mice in an intranasal model of *F. tularensis* infection (**Figure 3-6d**). We used an intranasal infection model because it represents the most challenging and relevant route of infection in potential bioterrorism attacks (Conlan et al., 2002; Sebastian et al., 2009). Mice were immunized with either iVAX- or PGCT-derived MBP^{4xDQNAT}, PD^{4xDQNAT}, or EPA$^{DNNNS-DQNRT}$ conjugates, as well as unmodified controls. Notably, across all three carrier proteins, only iVAX-derived vaccines elicited *Ft*O-PS-specific antibody titers that were significantly higher than those measured in the PBS immunized control group (**Figure 3-6e**). All immunized mice were challenged intranasally with 6000 cfu (60 times the

intranasal LD_{50}) of the virulent *F. tularensis* subsp. *holarctica* live vaccine strain (LVS) Rocky Mountain Laboratories. Importantly, all iVAX-derived vaccine candidates provided complete protection against intranasal challenge, which was indistinguishable from the protection conferred by PGCT-derived versions of the same vaccines (**Figure 3-6f-h**). These results demonstrate that the iVAX platform produces protective vaccine candidates that are at least as effective as those produced using a state-of-the-art biomanufacturing approach.

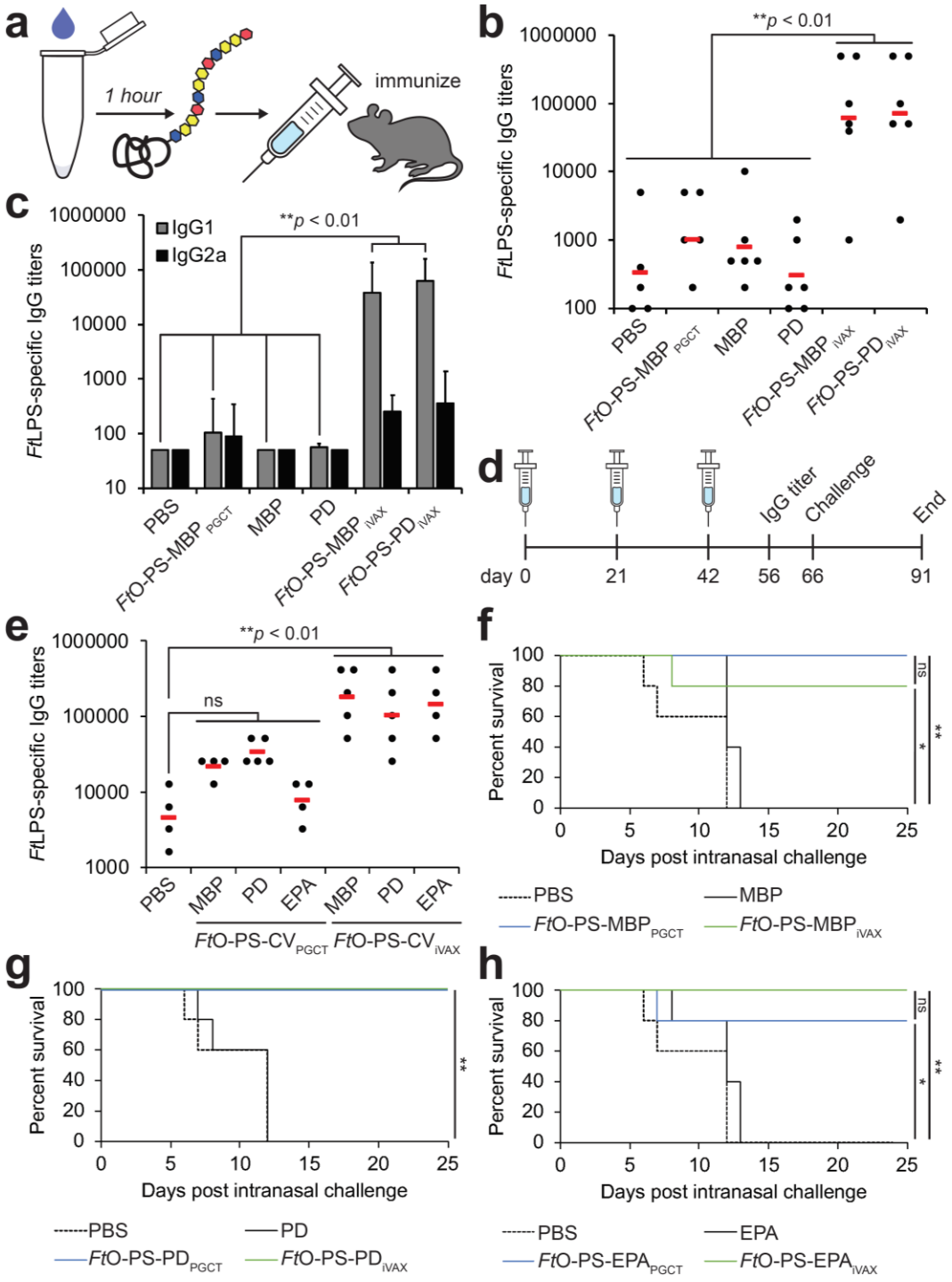


Figure 3-6. iVAX-derived conjugates elicit *Ft*LPS-specific antibodies and protect mice from lethal pathogen challenge. (**a**) Freeze-dried iVAX reactions assembled using detoxified lysates were used to synthesize anti-*F. tularensis* conjugate vaccines for immunization studies. (**b**)

Groups of BALB/c mice were immunized subcutaneously with PBS or 7.5 μg of purified, cell-free synthesized unmodified or *Ft*O-PS-conjugated carrier proteins. *Ft*O-PS-conjugated $MBP^{4xDQNAT}$ prepared in living *E. coli* cells using PCGT was used as a positive control. Each group was composed of six mice except for the PBS control group, which was composed of five mice. Mice were boosted on days 21 and 42 with identical doses of antigen. *Ft*LPS-specific IgG titers were measured by ELISA in endpoint (day 70) serum of individual mice (black dots) with *F. tularensis* LPS immobilized as antigen. Mean titers of each group are also shown (red lines). iVAX-derived conjugates elicited significantly higher levels of *Ft*LPS-specific IgG compared to all other groups (**$p < 0.01$, Tukey-Kramer HSD). (**c**) IgG1 and IgG2a subtype titers measured by ELISA from endpoint serum revealed that iVAX-derived conjugates boosted production of *Ft*O-PS-specific IgG1 compared to all other groups tested (**$p < 0.01$, Tukey-Kramer HSD). These results indicate that iVAX conjugates elicited a Th2-biased immune response typical of most conjugate vaccines. Values represent means and error bars represent standard errors of *Ft*LPS-specific IgGs detected by ELISA. (**d**) Groups of five BALB/c mice were immunized intraperitoneally with PBS or 10 μg of purified, cell-free synthesized unmodified or *Ft*O-PS-conjugated carrier proteins. *Ft*O-PS-conjugated carriers prepared in living *E. coli* cells using PCGT were used as positive controls. Mice were boosted on days 21 and 42 with identical doses of antigen. (**e**) On day 56, *Ft*LPS-specific IgG titers were measured by ELISA in serum of individual mice immunized with PBS or anti-*F. tularensis* conjugate vaccines (*Ft*O-PS-CV) (black dots) with *F. tularensis* LPS immobilized as antigen. Mean titers of each group are also shown (red lines). Only iVAX-derived conjugates elicited significantly higher levels of *Ft*LPS-specific IgG compared to PBS immunized controls across all carrier proteins tested (**$p < 0.01$, Tukey-Kramer HSD; CV: conjugate vaccine; ns: not significant). On day 66 mice were challenged intranasally with 6000 cfu (60 times the intranasal LD_{50}) *F. tularensis* subsp. *holarctica* LVS Rocky Mountain Laboratories and monitored for survival for an additional 25 days. Kaplan-Meier curves for immunizations with (**f**) $MBP^{4xDQNAT}$ (**g**) $PD^{4xDQNAT}$ and (**h**) $EPA^{DNNNS-DQNRT}$ as the carrier protein are shown. iVAX-derived vaccines protected mice from lethal pathogen challenge as effectively as vaccines synthesized using the state-of-the-art PGCT approach. (*$p < 0.05$; **$p < 0.01$, Fisher's exact test; ns: not significant).

3.4 Discussion.

In this work we have established iVAX, a cell-free platform for portable, on-demand, and scalable production of protective conjugate vaccines. We show that iVAX reactions can be detoxified to ensure the safety of conjugate vaccine products, freeze-dried for cold chain-independent distribution, and re-activated for high-yielding conjugate production by simply adding water. As a model vaccine candidate, we show that anti-*F. tularensis* conjugates produced in iVAX elicited pathogen-specific IgG antibodies and protected mice from lethal intranasal challenge with *F. tularensis*. Given the proven impact of thermostable meningococcal and smallpox vaccines in reducing or eradicating disease (Hopkins, 1988; Lydon et al., 2014), iVAX has the potential to significantly enhance vaccination efforts by reducing reliance on refrigerated supply chains.

The iVAX platform has several important features. First, iVAX is modular, which we have demonstrated through the interchangeability of (i) carrier proteins, including those used in licensed conjugate vaccines, and (ii) bacterial O-PS antigens from *F. tularensis* subsp. *tularensis* (type A) Schu S4, ETEC E. coli O78, and UPEC E. coli O7. The modularity of iVAX has the potential to enable rapid development of conjugate vaccines against diverse bacteria as well as multiple serotypes of a single bacterial pathogen that can be co-formulated to yield multivalent conjugate vaccines. Importantly, iVAX is the first example of enrichment of large and polymeric *O*-antigen carbohydrates as substrates for cell-free protein glycosylation. Moreover, to our

knowledge, this work represents the first demonstration of oligosaccharyltransferase-mediated O-PS conjugation to the carrier proteins used in licensed conjugate vaccine formulations, likely due to historical challenges associated with the expression of approved carriers in living *E. coli* (Figueiredo et al., 1995b; Haghi et al., 2011a; Stefan et al., 2011a).

Second, iVAX reactions are inexpensive, costing ~$12 mL^{-1} (**Supplementary Table 7-1**) with the ability to synthesize ~20 μg conjugate mL^{-1} in one hour (**Supplementary Figure 7-4a**). Assuming a dose size of 10 μg, our iVAX reactions can produce a vaccine dose for ~$6. For comparison, the CDC cost per dose for conjugate vaccines ranges from ~$9.50 for the *H. influenzae* vaccine ActHIB® to ~$75 and ~$118 for the meningococcal vaccine Menactra® and pneumococcal vaccine Prevnar 13®, respectively (CDC, 2019a).

Third, while conjugates derived from both living *E. coli* cells and iVAX PGCT protected mice from lethal challenge with *F. tularensis* LVS, iVAX-derived conjugates were significantly more effective at eliciting *Ft*LPS-specific IgGs than those derived from living *E. coli* cells using PGCT in the context of multiple carrier proteins (**Figure 3-6b, e**). Achieving high titers of polysaccharide-specific antibodies is broadly recognized as a correlate of conjugate vaccine efficacy (Beresford et al., 2017; Burrage et al., 2002; Dagan et al., 1998; Dagan et al., 2004; Gao et al., 2014; Hatz et al., 2015; Ho et al., 2000; Ho et al., 2002; Huttner et al., 2017; Knuf et al., 2009; Riddle et al., 2016). Given the importance of antibodies in protection against infection with the highly

virulent Schu S4 strain (Fulop et al., 2001; Lu et al., 2012), and the fact that antibody titers wane over time, achieving higher initial titers of *Ft*LPS-specific IgGs could provide protection against higher doses of pathogen and/or extend the duration of protection afforded by vaccination. Future comparative studies of iVAX- and PGCT-derived vaccines could provide a deeper understanding of immunogen features responsible for the enhanced *Ft*LPS-specific IgGs elicited by iVAX-derived vaccines, revealing design rules for the production of more effective conjugate vaccines.

Fourth, iVAX addresses a key gap in both cell-free and decentralized biomanufacturing technologies. Production of glycosylated products has not yet been demonstrated in cell-based decentralized biomanufacturing platforms (Crowell et al., 2018; Perez-Pinera et al., 2016) and existing cell-free platforms using *E. coli* lysates lack the ability to synthesize glycoproteins (Boles et al., 2017; Murphy et al., 2019; Pardee et al., 2016b; Salehi et al., 2016). While glycosylated human erythropoietin has been produced in a cell-free biomanufacturing platform based on freeze-dried Chinese hamster ovary cell lysates, its *in vivo* efficacy was not evaluated and challenges achieving efficient glycosylation were noted (Adiga et al., 2018). Further, this and the vast majority of other eukaryotic cell-free and cell-based systems rely on endogenous protein glycosylation machinery, and so are not compatible with conjugation of bacterial O-PS antigens. In contrast, the iVAX platform is enabled by our previous work to activate glycosylation in *E. coli* lysates that lack endogenous protein glycosylation pathways, which allows for bottom-up engineering of desired

glycosylation activities (Jaroentomeechai et al., 2018b). Here, we show that further development of this approach allows for rapid and portable production of protective conjugate vaccines. This required: (i) demonstration of modular cell-free synthesis and glycosylation of approved carrier proteins with O-PS antigens over a range of temperatures; (ii) rigorous assessment of reproducibility, scalability, and thermostability of reactions; (iii) detoxification of cell-free lysates; and (iv) evaluation of *in vivo* conjugate vaccine efficacy.

In summary, iVAX represents a platform technology for rapid development and on-demand, cold chain-independent biomanufacturing of conjugate vaccines. Conjugate vaccines are one of the safest and most effective methods for preventing bacterial infections (Jin et al., 2017; Trotter et al., 2008; Weintraub, 2003), but their development and distribution is limited by current manufacturing approaches. The iVAX platform addresses these limitations and further provides the first example, to our knowledge, of efficacious glycoprotein product synthesis in a decentralized manufacturing platform. iVAX joins an emerging set of technologies (Adiga et al., 2018; Boles et al., 2017; Crowell et al., 2018; Murphy et al., 2019; Pardee et al., 2016b; Perez-Pinera et al., 2016; Salehi et al., 2016) that have the potential to promote increased access to complex, life-saving drugs through decentralized production.

3.5 Materials and Methods.

Bacterial strains and plasmids.

E. coli NEB 5-alpha (NEB) was used for plasmid cloning and purification. *E. coli* CLM24 or CLM24 Δ*lpxM* strains were used for preparing cell-free lysates. *E. coli* CLM24 was used as the chassis for expressing bioconjugates *in vivo* using PGCT. CLM24 is a glyco-optimized derivative of W3110 that carries a deletion in the gene encoding the WaaL ligase, facilitating the accumulation of preassembled glycans on undecaprenyl diphosphate (Feldman et al., 2005b). CLM24 Δ*lpxM* has an endogenous acyltransferase deletion and serves as the chassis strain for production of detoxified cell-free lysates.

All plasmids used in the study are listed in **Supplementary Table 7-2**. Plasmids pJL1-MBP^{4xDQNAT}, pJL1-PD^{4xDQNAT}, pJL1-PorA^{4xDQNAT}, pJL1-TTc^{4xDQNAT}, pJL1-TTlight^{4xDQNAT}, pJL1-CRM197^{4xDQNAT}, and pJL1-TT^{4xDQNAT} were generated via PCR amplification and subsequent Gibson Assembly of a codon optimized gene construct purchased from IDT with a C-terminal 4xDQNAT-6xHis tag (Fisher et al., 2011) between the *NdeI* and *SalI* restriction sites in the pJL1 vector. Plasmid pJL1-EPA$^{DNNNS-DQNRT}$ was constructed using the same approach, but without the addition of a C-terminal 4xDQNAT-6xHis tag. Plasmids pTrc99s-ssDsbA-MBP^{4xDQNAT}, pTrc99s-ssDsbA-PD^{4xDQNAT}, pTrc99s-ssDsbA-PorA^{4xDQNAT}, pTrc99s-ssDsbA-TTc^{4xDQNAT}, pTrc99s-ssDsbA-TTlight^{4xDQNAT}, and pTrc99s-ssDsbA-EPA$^{DNNNS-DQNRT}$ were created via PCR amplification of each carrier protein gene and insertion into the pTrc99s vector between

the *NcoI* and *HindIII* restriction sites via Gibson Assembly. Plasmid pSF-*Cj*PglB-LpxE was constructed using a similar approach, but via insertion of the *lpxE* gene from pE (Needham et al., 2013) between the *NdeI* and *NsiI* restriction sites in the pSF vector. Inserts were amplified via PCR using Phusion® High-Fidelity DNA polymerase (NEB) with forward and reverse primers designed using the NEBuilder® Assembly Tool (nebuilder.neb.com) and purchased from IDT. The pJL1 vector (Addgene 69496) was digested using restriction enzymes NdeI and SalI-HF® (NEB). The pSF vector was digested using restriction enzymes NdeI and NotI (NEB). PCR products were gel extracted using an EZNA Gel Extraction Kit (Omega Bio-Tek), mixed with Gibson assembly reagents and incubated at 50°C for 1 hour. Plasmid DNA from the Gibson assembly reactions were transformed into E. coli NEB 5-alpha cells and circularized constructs were selected using kanamycin at 50 μg ml^{-1} (Sigma). Sequence-verified clones were purified using an EZNA Plasmid Midi Kit (Omega Bio-Tek) for use in CFPS and iVAX reactions.

Construction of CLM24 ΔlpxM strain.

E. coli CLM24 Δ*lpxM* was generated using the Datsenko-Wanner gene knockout method (Datsenko and Wanner, 2000b). Briefly, CLM24 cells were transformed with the pKD46 plasmid encoding the λ red system. Transformants were grown to an OD_{600} of 0.5-0.7 in 25 mL LB-Lennox media (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract and 5 g L^{-1} NaCl) with 50 μg mL^{-1} carbenicillin at 30°C, harvested and washed three times with

25 mL ice-cold 10% glycerol to make them electrocompetent, and resuspended in a final volume of 100 μL 10% glycerol. In parallel, a *lpxM* knockout cassette was generated by PCR amplifying the kanamycin resistance cassette from pKD4 with forward and reverse primers with homology to *lpxM*. Electrocompetent cells were transformed with 400 ng of the *lpxM* knockout cassette and plated on LB agar with 30 μg mL^{-1} kanamycin for selection of resistant colonies. Plates were grown at 37°C to cure cells of the pKD46 plasmid. Colonies that grew on kanamycin were confirmed to have acquired the knockout cassette via colony PCR and DNA sequencing. These confirmed colonies were then transformed with pCP20 to remove the kanamycin resistance gene via Flp-FRT recombination. Transformants were plated on LB agar with 50 μg mL^{-1} carbenicillin. Following selection, colonies were grown in liquid culture at 42°C to cure cells of the pCP20 plasmid. Colonies were confirmed to have lost both *lpxM* and the knockout cassette via colony PCR and DNA sequencing and confirmed to have lost both kanamycin and carbenicillin resistance via replica plating on LB agar plates with 50 μg mL^{-1} carbenicillin or kanamycin. All primers used for construction and validation of this strain are listed in **Supplementary Table 7-3**.

Cell-free lysate preparation.

E. coli CLM24 source strains were grown in 2xYTP media (10 g/L yeast extract, 16 g/L tryptone, 5 g/L NaCl, 7 g/L K_2HPO_4, 3 g/L KH_2PO_4, pH 7.2) in shake flasks (1 L scale) or a Sartorius Stedim BIOSTAT Cplus bioreactor (10 L scale) at 37°C. Protein

synthesis yields and glycosylation activity were reproducible across different batches of lysate at both small and large scale. To generate *Cj*PglB-enriched lysate, CLM24 cells carrying plasmid pSF-*Cj*PglB (Ollis et al., 2015b) was used as the source strain. To generate *Ft*O-PS-enriched lysates, CLM24 carrying plasmid pGAB2 (Cuccui et al., 2013b) was used as the source strain. To generate one-pot lysates containing both *Cj*PglB and *Ft*O-PS, *Ec*O78-PS, or *Ec*O7-PS, CLM24 carrying pSF-*Cj*PglB and one of the following bacterial O-PS biosynthetic pathway plasmids was used as the source strain: pGAB2 (*Ft*O-PS), pMW07-O78 (*Ec*O78-PS), and pJHCV32 (*Ec*O7-PS). *Cj*PglB expression was induced at an OD_{600} of 0.8-1.0 with 0.02% (w/v) L-arabinose and cultures were moved to 30°C. Cells were grown to a final OD_{600} of ~3.0, at which point cells were pelleted by centrifugation at 5,000xg for 15 min at 4°C. Cell pellets were then washed three times with cold S30 buffer (10 mM Tris-acetate pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate) and pelleted at 5000xg for 10 min at 4°C. After the final wash, cells were pelleted at 7000xg for 10 min at 4°C, weighed, flash frozen in liquid nitrogen, and stored at -80°C. To make cell lysate, cell pellets were resuspended to homogeneity in 1 mL of S30 buffer per 1 g of wet cell mass. Cells were disrupted via a single passage through an Avestin EmulsiFlex-B15 (1 L scale) or EmulsiFlex-C3 (10 L scale) high-pressure homogenizer at 20,000-25,000 psi. The lysate was then centrifuged twice at 30,000×g for 30 min to remove cell debris. Supernatant was transferred to clean microcentrifuge tubes and incubated at 37°C with shaking at 250 rpm for 60 min. Following centrifugation (15,000xg) for 15 min at 4°C, supernatant

was collected, aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C. S30 lysate was active for about 3 freeze-thaw cycles and contained ~40 g/L total protein as measured by Bradford assay.

Cell-free protein synthesis.

CFPS reactions were carried out in 1.5 mL microcentrifuge tubes (15 μL scale), 15 mL conical tubes (1 mL scale), or 50 mL conical tubes (5 mL scale) with a modified PANOx-SP system (Jewett and Swartz, 2004a). The CFPS reaction mixture consists of the following components: 1.2 mM ATP; 0.85 mM each of GTP, UTP, and CTP; 34.0 μg mL^{-1} L-5-formyl-5, 6, 7, 8-tetrahydrofolic acid (folinic acid); 170.0 μg mL^{-1} of *E. coli* tRNA mixture; 130 mM potassium glutamate; 10 mM ammonium glutamate; 12 mM magnesium glutamate; 2 mM each of 20 amino acids; 0.4 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); 57 mM HEPES; 13.3 μg mL^{-1} plasmid; and 27% v/v of cell lysate. For reaction volumes ≥1 mL, plasmid was added at 6.67 μg mL^{-1}, as this lower plasmid concentration conserved reagents with no effect on protein synthesis yields or kinetics. For expression of PorA, reactions were supplemented with nanodiscs at 1 μg mL^{-1}, which were prepared as previously described (Schoborg et al., 2017) or purchased (Cube Biotech). For expression of $CRM197^{4xDQNAT}$, CFPS was carried out

at 25°C for 20 hours, unless otherwise noted. For all other carrier proteins, CFPS was run at 30°C for 20 hours, unless otherwise noted.

For expression of $TT^{4xDQNAT}$, which contains intermolecular disulfide bonds, CFPS was carried out under oxidizing conditions. For oxidizing conditions, lysate was pre-conditioned with 750 μM iodoacetamide at room temperature for 30 min to covalently bind free sulfhydryls (-SH) and the reaction mix was supplemented with 200 mM glutathione at a 4:1 ratio of oxidized and reduced forms and 10 μM recombinant *E. coli* DsbC (Knapp et al., 2007).

***In vitro* bioconjugate vaccine expression (iVAX).**

For *in vitro* expression and glycosylation of carrier proteins in crude lysates, a two-phase scheme was implemented. In the first phase, CFPS was carried out for 15 min at 25-30°C as described above. In the second phase, protein glycosylation was initiated by the addition of $MnCl_2$ and DDM at a final concentration of 25 mM and 0.1% (w/v), respectively, and allowed to proceed at 30°C for a total reaction time of 1 hour. Protein synthesis yields and glycosylation activity were reproducible across biological replicates of iVAX reactions at both small and large scale. Reactions were then centrifuged at 20,000xg for 10 min to remove insoluble or aggregated protein products and the supernatant was analyzed by SDS-PAGE and Western blotting.

Purification of aglycosylated and glycosylated carriers from iVAX reactions was carried out using Ni-NTA agarose (Qiagen) according to manufacturer's

protocols. Briefly, 0.5 mL Ni-NTA agarose per 1 mL cell-free reaction mixture was equilibrated in Buffer 1 (300 mM NaCl 50 mM NaH_2PO_4) with 10 mM imidazole. Soluble fractions from iVAX reactions were loaded on Ni-NTA agarose and incubated at 4°C for 2-4 hours to bind 6xHis-tagged protein. Following incubation, the cell-free reaction/agarose mixture was loaded onto a polypropylene column (BioRad) and washed twice with 6 column volumes of Buffer 1 with 20 mM imidazole. Protein was eluted in 4 fractions, each with 0.3 mL Buffer 1 with 300 mM imidazole per mL of cell-free reaction mixture. All buffers were used and stored at 4°C. Protein was stored at a final concentration of 1-2 mg mL^{-1} in sterile 1xPBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) at 4°C.

Expression of bioconjugates in vivo using protein glycan coupling technology (PGCT).

Plasmids encoding bioconjugate carrier protein genes preceded by the DsbA leader sequence for translocation to the periplasm were transformed into CLM24 cells carrying pGAB2 and pSF-*Cj*PglB. CLM24 carrying only pGAB2 was used as a negative control. Transformed cells were grown in 5 mL LB media (10 g L^{-1} yeast extract, 5 g L^{-1} tryptone, 5 g L^{-1} NaCl) overnight at 37°C. The next day, cells were subcultured into 100 mL LB and allowed to grow at 37°C for 6 hours after which the culture was supplemented with 0.2% arabinose and 0.5 mM IPTG to induce expression of *Cj*PglB and the bioconjugate carrier protein, respectively. Protein expression was then carried

out for 16 hours at 30°C, at which point cells were harvested. Cell pellets were resuspended in 1 mL sterile PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) and lysed using a Q125 Sonicator (Qsonica, Newtown, CT) at 40% amplitude in cycles of 10 sec on/10 sec off for a total of 5 min. Soluble fractions were isolated following centrifugation at 15,000 rpm for 30 min at 4°C. Protein was purified from soluble fractions using Ni-NTA spin columns (Qiagen), following the manufacturer's protocol.

Western blot analysis.

Samples were run on 4-12% Bis-Tris SDS-PAGE gels (Invitrogen). Following electrophoretic separation, proteins were transferred from gels onto Immobilon-P polyvinylidene difluoride (PVDF) membranes (0.45 μm) according to the manufacturer's protocol. Membranes were washed with PBS (80 g L^{-1} NaCl, 0.2 g L^{-1} KCl, 1.44 g L^{-1} Na_2HPO_4, 0.24 g L^{-1} KH_2PO_4, pH 7.4) followed by incubation for 1 hour in Odyssey® Blocking Buffer (LI-COR). After blocking, membranes were washed 6 times with PBST (80 g L^{-1} NaCl, 0.2 g L^{-1} KCl, 1.44 g L^{-1} Na_2HPO_4, 0.24 g L^{-1} KH_2PO_4, 1 mL L^{-1} Tween-20, pH 7.4) with a 5 min incubation between each wash. For iVAX samples, membranes were probed with both an anti-6xHis tag antibody and an anti-O-PS antibody or antisera specific to the O antigen of interest, if commercially available. Probing of membranes was performed for at least 1 hour with shaking at room temperature, after which membranes were washed with PBST in the same

manner as described above and probed with fluorescently labeled secondary antibodies. Membranes were imaged using an Odyssey® Fc imaging system (LI-COR). CRM197 and TT production were compared to commercial DT and TT standards (Sigma) and orthogonally detected by an identical SDS-PAGE procedure followed by Western blot analysis with a polyclonal antibody that recognizes diphtheria or tetanus toxin, respectively. All antibodies and dilutions used are listed in **Supplementary Table 7-4**.

TLR4 activation assay.

HEK-Blue hTLR4 cells (Invivogen) were cultured in DMEM media, high glucose/L-glutamine supplement with 10% fetal bovine serum, 50 U mL^{-1} penicillin, 50 mg mL^{-1} streptomycin, and 100 μg mL^{-1} Normacin™ at 37°C in a humidified incubator containing 5% CO_2. After reaching ~50-80% confluency, cells were plated into 96-well plates at a density of 1.4×10^5 cells per mL in HEK-Blue detection media (Invivogen). Antigens were added at the following concentrations: 100 ng μL^{-1} purified protein; and 100 ng μL^{-1} total protein in lysate. Purified *E. coli* O55:B5 LPS (Sigma-Aldrich) and detoxified *E. coli* O55:B5 (Sigma-Aldrich) were added at 1.0 ng mL^{-1} and served as positive and negative controls, respectively. Plates were incubated at 37°C, 5% CO_2 for 10–16 h before measuring absorbance at 620 nm. Statistical significance was determined using paired *t*-tests.

Mouse immunization and F. tularensis challenge.

Groups of 5-6 six-week old female BALB/c mice (Harlan Sprague Dawley) were immunized with 100 μL PBS (pH 7.4) alone or containing purified aglycosylated MBP, *Ft*O-PS-conjugated MBP, aglycosylated PD, *Ft*O-PS-conjugated PD, aglycosylated EPA, or *Ft*O-PS-conjugated EPA, as previously described (Chen et al., 2010). The amount of antigen in each preparation was normalized to 7.5 μg to ensure that an equivalent amount of aglycosylated protein or bioconjugate was administered in each case. Purified protein groups formulated in PBS were mixed with an equal volume of incomplete Freund's Adjuvant (Sigma-Aldrich) before injection. Prior to immunization, material for each group (5 μL) was streaked on LB agar plates and grown overnight at 37°C to confirm sterility and endotoxin activity was measured by TLR4 activation assay. Each group of mice was boosted with an identical dosage of antigen 21 days and 42 days after the initial immunization. Mice were observed 24 and 48 hours after each injection for changes in behavior and physical health and no abnormal responses were reported.

For antibody titering studies, mice were immunized subcutaneously with vaccine or controls according to the protocol described above. Blood was obtained on day -1, 21, 35, 49, and 63 via submandibular collection and at study termination on day 70 via cardiac puncture.

For pathogen challenge studies, mice were immunized intraperitoneally with vaccine or controls according to the protocol described above. Mice were challenged

intranasally on day 66 with 6000 cfu (60 times the intranasal LD_{50}) *F. tularensis* subsp. *holarctica* LVS Rocky Mountain Laboratories. Survival was monitored for an additional 25 days following pathogen challenge, during which mice were examined daily for signs of disease and sacrificed according to the approved protocol when moribund. Statistical significance was determined via endpoint comparison using Fisher's exact test.

All procedures were carried out in accordance with Protocol 2012-0132 approved by the Cornell University Institutional Animal Care and Use Committee and/or Protocol 1305086 approved by the University of Iowa Animal Care and Use Committee.

Enzyme-linked immunosorbent assay.

F. tularensis LPS-specific antibodies elicited by immunized mice were measured via indirect ELISA using a modification of a previously described protocol (Chen et al., 2010). Briefly, sera were isolated from the collected blood draws after centrifugation at 5000xg for 10 min and stored at −20 °C; 96-well plates (Maxisorp; Nunc Nalgene) were coated with *F. tularensis* LPS (BEI resources) at a concentration of 5 μg mL^{-1} in PBS and incubated overnight at 4°C. The next day, plates were washed three times with PBST (PBS, 0.05% Tween-20, 0.3% BSA) and blocked overnight at 4°C with 5% nonfat dry milk (Carnation) in PBS. Samples were serially diluted by a factor of two in triplicate between 1:100 and 1:12,800,000 in blocking buffer and added to the plate for 2 hours at

37°C. Plates were washed three times with PBST and incubated for 1 hour at 37°C in the presence of one of the following HRP-conjugated antibodies (all from Abcam and used at 1:25,000 dilution): goat anti-mouse IgG, anti-mouse IgG1, and anti-mouse IgG2a. After three additional washes with PBST, 3,3′-5,5′-tetramethylbenzidine substrate (1-Step Ultra TMB-ELISA; Thermo-Fisher) was added, and the plate was incubated at room temperature in the dark for 30 min. The reaction was halted with 2 M H_2SO_4, and absorbance was quantified via microplate spectrophotometer (Tecan) at a wavelength of 450 nm. Serum antibody titers were determined by measuring the lowest dilution that resulted in signal 3 SDs above no serum background controls. Statistical significance was determined in RStudio 1.1.463 using one-way ANOVA and the Tukey–Kramer *post hoc* honest significant difference test.

Quantification and Statistical Analysis.

Quantification of cell-free protein synthesis yields

To quantify the amount of protein synthesized in iVAX reactions, two approaches were used. Fluorescence units of sfGFP were converted to concentrations using a previously reported standard curve (Hong et al., 2014). Yields of all other proteins were assessed via the addition of 10 μM L-^{14}C-leucine (11.1 GBq mmol^{-1}, PerkinElmer) to the CFPS mixture to yield trichloroacetic acid-precipitable radioactivity that was measured using a liquid scintillation counter as described previously (Kim and Swartz, 2001).

Statistical analysis

Statistical parameters including the definitions and values of *n*, *p*-values, standard deviations, and standard errors are reported in the figures and corresponding figure legends. Analytical techniques are described in the corresponding Materials and Methods section.

Data and Software Availability

All plasmid constructs used in this study including complete DNA sequences are deposited on Addgene (constructs 128389-128404).

3.6 Acknowledgements.

This work was supported by Defense Threat Reduction Agency (HDTRA1-15-10052/P00001 to M.P.D. and M.C.J.), National Science Foundation (Grants # CBET 1159581 and CBET 1264701 both to M.P.D. and MCB 1413563 to M.P.D. and M.C.J.), the David and Lucile Packard Foundation, and the Dreyfus Teacher-Scholar program. J.C.S. and T.C.S. were supported by National Science Foundation Graduate Research Fellowships. T.J. was supported by a Royal Thai Government Fellowship. R.S.D. was funded, in part, by the Northwestern University Chemistry of Life Processes Summer Scholars program.

3.7 References.

Adiga, R., Al-adhami, M., Andar, A., Borhani, S., Brown, S., Burgenson, D., Cooper, M.A., Deldari, S., Frey, D.D., Ge, X., *et al.* (2018). Point-of-care production of therapeutic proteins of good-manufacturing-practice quality. Nat Biomed Eng.

Ashok, A., Brison, M., and LeTallec, Y. (2017). Improving cold chain systems: Challenges and solutions. Vaccine *35*, 2217-2223.

Bayburt, T.H., and Sligar, S.G. (2010). Membrane protein assembly into Nanodiscs. FEBS letters *584*, 1721-1727.

Beresford, N.J., Martino, A., Feavers, I.M., Corbel, M.J., Bai, X., Borrow, R., and Bolgiano, B. (2017). Quality, immunogenicity and stability of meningococcal serogroup ACWY-CRM(197), DT and TT glycoconjugate vaccines. Vaccine *35*, 3598-3606.

Bhushan, R., Anthony, B.F., and Frasch, C.E. (1998). Estimation of group B streptococcus type III polysaccharide-specific antibody concentrations in human sera is antigen dependent. Infection and immunity *66*, 5848-5853.

Bogaert, D., Hermans, P.W., Adrian, P.V., Rumke, H.C., and de Groot, R. (2004). Pneumococcal vaccines: an update on current strategies. Vaccine *22*, 2209-2220.

Boles, K.S., Kannan, K., Gill, J., Felderman, M., Gouvis, H., Hubby, B., Kamrud, K.I., Venter, J.C., and Gibson, D.G. (2017). Digital-to-biological converter for on-demand production of biologics. Nature biotechnology *35*, 672-675.

Bolgiano, B., Mawas, F., Burkin, K., Crane, D.T., Saydam, M., Rigsby, P., and Corbel, M.J. (2007). A retrospective study on the quality of Haemophilus influenzae type b vaccines used in the UK between 1996 and 2004. Human vaccines *3*, 176-182.

Brito, L.A., and Singh, M. (2011). Acceptable levels of endotoxin in vaccine formulations during preclinical research. J Pharm Sci *100*, 34-37.

Burrage, M., Robinson, A., Borrow, R., Andrews, N., Southern, J., Findlow, J., Martin, S., Thornton, C., Goldblatt, D., Corbel, M., *et al.* (2002). Effect of vaccination with carrier protein on response to meningococcal C conjugate vaccines and value of different immunoassays as predictors of protection. Infection and immunity *70*, 4946-4954.

Casella, C.R., and Mitchell, T.C. (2008). Putting endotoxin to work for us: monophosphoryl lipid A as a safe and effective vaccine adjuvant. Cellular and molecular life sciences : CMLS *65*, 3231-3240.

CDC (2019a). CDC Vaccine Price List.

CDC (2019b). Ring Vaccination.

Celik, E., Ollis, A.A., Lasanajak, Y., Fisher, A.C., Gur, G., Smith, D.F., and DeLisa, M.P. (2015). Glycoarrays with engineered phages displaying structurally diverse oligosaccharides enable high-throughput detection of glycan-protein interactions. Biotechnology journal *10*, 199-209.

Chen, D.J., Osterrieder, N., Metzger, S.M., Buckles, E., Doody, A.M., DeLisa, M.P., and Putnam, D. (2010). Delivery of foreign antigens by engineered outer membrane vesicle vaccines. Proceedings of the National Academy of Sciences of the United States of America *107*, 3099-3104.

Chen, L., Valentine, J.L., Huang, C.-J., Endicott, C.E., Moeller, T.D., Rasmussen, J.A., Fletcher, J.R., Boll, J.M., Rosenthal, J.A., Dobruchowska, J., *et al.* (2016). Outer membrane vesicles displaying engineered glycotopes elicit protective antibodies. Proceedings of the National Academy of Sciences of the United States of America.

Chen, M.M., Glover, K.J., and Imperiali, B. (2007). From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. Biochemistry *46*, 5579-5585.

Conlan, J.W., Shen, H., Webb, A., and Perry, M.B. (2002). Mice vaccinated with the O-antigen of Francisella tularensis LVS lipopolysaccharide conjugated to bovine serum albumin develop varying degrees of protective immunity against systemic or aerosol challenge with virulent type A and type B strains of the pathogen. Vaccine *20*, 3465-3471.

Crowell, L.E., Lu, A.E., Love, K.R., Stockdale, A., Timmick, S.M., Wu, D., Wang, Y.A., Doherty, W., Bonnyman, A., Vecchiarello, N., *et al.* (2018). On-demand manufacturing of clinical-quality biopharmaceuticals. Nature biotechnology.

Cuccui, J., Thomas, R.M., Moule, M.G., D'Elia, R.V., Laws, T.R., Mills, D.C., Williamson, D., Atkins, T.P., Prior, J.L., and Wren, B.W. (2013). Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis. Open Biol *3*, 130002.

Dagan, R., Eskola, J., Leclerc, C., and Leroy, O. (1998). Reduced response to multiple vaccines sharing common protein epitopes that are administered simultaneously to infants. Infection and immunity *66*, 2093-2098.

Dagan, R., Goldblatt, D., Maleckar, J.R., Yaïch, M., and Eskola, J. (2004). Reduction of antibody response to an 11-valent pneumococcal vaccine coadministered with a vaccine containing acellular pertussis components. Infection and immunity *72*, 5383-5391.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America *97*, 6640-6645.

Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., and Aebi, M. (2005). Engineering N-linked

protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. Proc Natl Acad Sci U S A *102*, 3016-3021.

Fernandez, S., Palmer, D.R., Simmons, M., Sun, P., Bisbing, J., McClain, S., Mani, S., Burgess, T., Gunther, V., and Sun, W. (2007). Potential role for Toll-like receptor 4 in mediating Escherichia coli maltose-binding protein activation of dendritic cells. Infection and immunity *75*, 1359-1363.

Figueiredo, D., Turcotte, C., Frankel, G., Li, Y., Dolly, O., Wilkin, G., Marriott, D., Fairweather, N., and Dougan, G. (1995). Characterization of recombinant tetanus toxin derivatives suitable for vaccine development. Infection and immunity *63*, 3218-3221.

Fisher, A.C., Haitjema, C.H., Guarino, C., Celik, E., Endicott, C.E., Reading, C.A., Merritt, J.H., Ptak, A.C., Zhang, S., and DeLisa, M.P. (2011). Production of secretory and extracellular N-linked glycoproteins in Escherichia coli. Appl Environ Microbiol *77*, 871-881.

Frasch, C.E. (2009). Preparation of bacterial polysaccharide-protein conjugates: analytical and manufacturing challenges. Vaccine *27*, 6468-6470.

Fulop, M., Mastroeni, P., Green, M., and Titball, R.W. (2001). Role of antibody to lipopolysaccharide in protection against low- and high-virulence strains of Francisella tularensis. Vaccine *19*, 4465-4472.

Gao, F., Lockyer, K., Burkin, K., Crane, D.T., and Bolgiano, B. (2014). A physico-chemical assessment of the thermal stability of pneumococcal conjugate vaccine components. Human vaccines & immunotherapeutics *10*, 2744-2753.

Garcia-Quintanilla, F., Iwashkiw, J.A., Price, N.L., Stratilo, C., and Feldman, M.F. (2014). Production of a recombinant vaccine candidate against Burkholderia pseudomallei exploiting the bacterial N-glycosylation machinery. Front Microbiol *5*, 381.

Haghi, F., Peerayeh, S.N., Siadat, S.D., and Montajabiniat, M. (2011). Cloning, expression and purification of outer membrane protein PorA of Neisseria meningitidis serogroup B. Journal of infection in developing countries *5*, 856-862.

Hatz, C.F., Bally, B., Rohrer, S., Steffen, R., Kramme, S., Siegrist, C.A., Wacker, M., Alaimo, C., and Fonck, V.G. (2015). Safety and immunogenicity of a candidate bioconjugate vaccine against Shigella dysenteriae type 1 administered to healthy adults: A single blind, partially randomized Phase I study. Vaccine *33*, 4594-4601.

Ho, M.M., Bolgiano, B., and Corbel, M.J. (2000). Assessment of the stability and immunogenicity of meningococcal oligosaccharide C-CRM197 conjugate vaccines. Vaccine *19*, 716-725.

Ho, M.M., Mawas, F., Bolgiano, B., Lemercinier, X., Crane, D.T., Huskisson, R., and Corbel, M.J. (2002). Physico-chemical and immunological examination of the thermal stability of tetanus toxoid conjugate vaccines. Vaccine *20*, 3509-3522.

Hong, S.H., Ntai, I., Haimovich, A.D., Kelleher, N.L., Isaacs, F.J., and Jewett, M.C. (2014). Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation. ACS synthetic biology *3*, 398-409.

Hopkins, J.W. (1988). The eradication of smallpox: organizational learning and innovation in international health administration. J Dev Areas *22*, 321-332.

Humphreys, G. (2011). Vaccination: rattling the supply chain (Bulletin of the World Health Organization: World Health Organization).

Huttner, A., Hatz, C., van den Dobbelsteen, G., Abbanat, D., Hornacek, A., Frolich, R., Dreyer, A.M., Martin, P., Davies, T., Fae, K., *et al.* (2017). Safety, immunogenicity, and preliminary clinical efficacy of a vaccine against extraintestinal pathogenic Escherichia coli in women with a history of recurrent urinary tract infection: a randomised, single-blind, placebo-controlled phase 1b trial. The Lancet Infectious diseases.

Ihssen, J., Kowarik, M., Dilettoso, S., Tanner, C., Wacker, M., and Thony-Meyer, L. (2010). Production of glycoprotein vaccines in Escherichia coli. Microb Cell Fact *9*, 61.

Jansson, P.E., Lindberg, B., Widmalm, G., and Leontein, K. (1987). Structural studies of the Escherichia coli O78 O-antigen polysaccharide. Carbohydrate research *165*, 87-92.

Jaroentomeechai, T., Stark, J.C., Natarajan, A., Glasscock, C.J., Yates, L.E., Hsu, K.J., Mrksich, M., Jewett, M.C., and DeLisa, M.P. (2018). Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. Nat Commun *9*, 2686.

Jewett, M.C., and Swartz, J.R. (2004). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnology and bioengineering *86*, 19-26.

Jin, C., Gibani, M.M., Moore, M., Juel, H.B., Jones, E., Meiring, J., Harris, V., Gardner, J., Nebykova, A., Kerridge, S.A., *et al.* (2017). Efficacy and immunogenicity of a Vi-tetanus toxoid conjugate vaccine in the prevention of typhoid fever using a controlled human infection model of Salmonella Typhi: a randomised controlled, phase 2b trial. Lancet (London, England) *390*, 2472-2480.

Johnson, J.R. (1991). Virulence factors in Escherichia coli urinary tract infection. Clin Microbiol Rev *4*, 80-128.

Kim, D.M., and Swartz, J.R. (2001). Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. Biotechnology and bioengineering *74*, 309-316.

Kim, D.M., and Swartz, J.R. (2004). Efficient production of a bioactive, multiple disulfide-bonded protein using modified extracts of Escherichia coli. Biotechnology and bioengineering *85*, 122-129.

Knuf, M., Szenborn, L., Moro, M., Petit, C., Bermal, N., Bernard, L., Dieussaert, I., and Schuerman, L. (2009). Immunogenicity of routinely used childhood vaccines when coadministered with the 10-valent pneumococcal non-typeable Haemophilus influenzae protein D conjugate vaccine (PHiD-CV). Pediatr Infect Dis J *28*, S97-s108.

Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., *et al.* (2006). Definition of the bacterial N-glycosylation site consensus sequence. The EMBO journal *25*, 1957-1966.

L'vov, V.L., Shashkov, A.S., Dmitriev, B.A., Kochetkov, N.K., Jann, B., and Jann, K. (1984). Structural studies of the O-specific side chain of the lipopolysaccharide from Escherichia coli O:7. Carbohydrate research *126*, 249-259.

Lu, Z., Madico, G., Roche, M.I., Wang, Q., Hui, J.H., Perkins, H.M., Zaia, J., Costello, C.E., and Sharon, J. (2012). Protective B-cell epitopes of Francisella tularensis O-polysaccharide in a mouse model of respiratory tularaemia. Immunology *136*, 352-360.

Lydon, P., Zipursky, S., Tevi-Benissan, C., Djingarey, M.H., Gbedonou, P., Youssouf, B.O., and Zaffran, M. (2014). Economic benefits of keeping vaccines at ambient temperature during mass vaccination: the case of meningitis A vaccine in Chad. Bulletin of the World Health Organization *92*, 86-92.

Ma, Z., Zhang, H., Shang, W., Zhu, F., Han, W., Zhao, X., Han, D., Wang, P.G., and Chen, M. (2014). Glycoconjugate vaccine containing Escherichia coli O157:H7 O-antigen linked with maltose-binding protein elicits humoral and cellular responses. PloS one *9*, e105215.

Marshall, L.E., Nelson, M., Davies, C.H., Whelan, A.O., Jenner, D.C., Moule, M.G., Denman, C., Cuccui, J., Atkins, T.P., Wren, B.W., *et al.* (2018). An O-antigen glycoconjugate vaccine produced using protein glycan coupling technology is protective in an inhalational rat model of tularemia. J Immunol Res *2018*, 8087916.

Maurin, M. (2015). Francisella tularensis as a potential agent of bioterrorism? Expert Rev Anti Infect Ther *13*, 141-144.

Murphy, T.W., Sheng, J., Naler, L.B., Feng, X., and Lu, C. (2019). On-chip manufacturing of synthetic proteins for point-of-care therapeutics. Microsyst Nanoeng *5*, 13.

Needham, B.D., Carroll, S.M., Giles, D.K., Georgiou, G., Whiteley, M., and Trent, M.S. (2013). Modulating the innate immune response by combinatorial engineering of endotoxin. Proceedings of the National Academy of Sciences of the United States of America *110*, 1464-1469.

Novak, R.T., Kambou, J.L., Diomande, F.V., Tarbangdo, T.F., Ouedraogo-Traore, R., Sangare, L., Lingani, C., Martin, S.W., Hatcher, C., Mayer, L.W., *et al.* (2012). Serogroup A meningococcal conjugate vaccination in Burkina Faso: analysis of national surveillance data. The Lancet Infectious diseases *12*, 757-764.

Ollis, A.A., Chai, Y., Natarajan, A., Perregaux, E., Jaroentomeechai, T., Guarino, C., Smith, J., Zhang, S., and DeLisa, M.P. (2015). Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. Sci Rep *5*, 15237.

Ollis, A.A., Zhang, S., Fisher, A.C., and DeLisa, M.P. (2014). Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. Nat Chem Biol *10*, 816-822.

Oyston, P.C., Sjostedt, A., and Titball, R.W. (2004). Tularaemia: bioterrorism defence renews interest in Francisella tularensis. Nat Rev Microbiol *2*, 967-978.

Palmu, A.A., Jokinen, J., Borys, D., Nieminen, H., Ruokokoski, E., Siira, L., Puumalainen, T., Lommel, P., Hezareh, M., Moreira, M., *et al.* (2013). Effectiveness of the ten-valent pneumococcal Haemophilus influenzae protein D conjugate vaccine (PHiD-CV10) against invasive pneumococcal disease: a cluster randomised trial. Lancet (London, England) *381*, 214-222.

Pardee, K., Slomovic, S., Nguyen, Peter Q., Lee, Jeong W., Donghia, N., Burrill, D., Ferrante, T., McSorley, Fern R., Furuta, Y., Vernet, A., *et al.* (2016). Portable, on-demand biomolecular manufacturing. Cell *167*, 248-259.e212.

Perez-Pinera, P., Han, N., Cleto, S., Cao, J., Purcell, O., Shah, K.A., Lee, K., Ram, R., and Lu, T.K. (2016). Synthetic biology and microbioreactor platforms for programmable production of biologics at the point-of-care. Nat Commun *7*, 12211.

Perez, J.G., Stark, J.C., and Jewett, M.C. (2016). Cell-free synthetic biology: Engineering beyond the cell. Cold Spring Harb Perspect Biol.

Petsch, D., and Anspach, F.B. (2000). Endotoxin removal from protein solutions. Journal of biotechnology *76*, 97-119.

Poehling, K.A., Talbot, T.R., Griffin, M.R., Craig, A.S., Whitney, C.G., Zell, E., Lexau, C.A., Thomas, A.R., Harrison, L.H., Reingold, A.L., *et al.* (2006). Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine. Jama *295*, 1668-1674.

Prior, J.L., Prior, R.G., Hitchen, P.G., Diaper, H., Griffin, K.F., Morris, H.R., Dell, A., and Titball, R.W. (2003). Characterization of the O antigen gene cluster and structural analysis of the O antigen of Francisella tularensis subsp. tularensis. J Med Microbiol *52*, 845-851.

Qadri, F., Svennerholm, A.M., Faruque, A.S., and Sack, R.B. (2005). Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev *18*, 465-483.

Riddle, M.S., Kaminski, R.W., Di Paolo, C., Porter, C.K., Gutierrez, R.L., Clarkson, K.A., Weerts, H.E., Duplessis, C., Castellano, A., Alaimo, C., *et al.* (2016). Safety and

immunogenicity of a candidate bioconjugate vaccine against Shigella flexneri 2a administered to healthy adults: a single blind, randomized phase I study. Clin Vaccine Immunol.

Roush, S.W., McIntyre, L., and Baldy, L.M. (2008). Manual for the surveillance of vaccine-preventable diseases. Atlanta: Centers for Disease Control and Prevention, 4.

Russell, J.A. (2006). Management of sepsis. N Engl J Med *355*, 1699-1713.

Salehi, A.S., Smith, M.T., Bennett, A.M., Williams, J.B., Pitt, W.G., and Bundy, B.C. (2016). Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system. Biotechnology journal *11*, 274-281.

Schoborg, J.A., Hershewe, J.M., Stark, J.C., Kightlinger, W., Kath, J.E., Jaroentomeechai, T., Natarajan, A., DeLisa, M.P., and Jewett, M.C. (2017). A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases. Biotechnology and bioengineering.

Sebastian, S., Pinkham, J.T., Lynch, J.G., Ross, R.A., Reinap, B., Blalock, L.T., Conlan, J.W., and Kasper, D.L. (2009). Cellular and humoral immunity are synergistic in protection against types A and B Francisella tularensis. Vaccine *27*, 597-605.

Sethuraman, N., and Stadheim, T.A. (2006). Challenges in therapeutic glycoprotein production. Curr Opin Biotechnol *17*, 341-346.

Silfverdal, S.A., Hogh, B., Bergsaker, M.R., Skerlikova, H., Lommel, P., Borys, D., and Schuerman, L. (2009). Immunogenicity of a 2-dose priming and booster vaccination with the 10-valent pneumococcal nontypeable Haemophilus influenzae protein D conjugate vaccine. Pediatr Infect Dis J *28*, e276-282.

Silverman, A.D., Karim, A.S., and Jewett, M.C. (2020). Cell-free gene expression: an expanded repertoire of applications. Nat Rev Genet *21*, 151-170.

Stefan, A., Conti, M., Rubboli, D., Ravagli, L., Presta, E., and Hochkoeppler, A. (2011). Overexpression and purification of the recombinant diphtheria toxin variant CRM197 in Escherichia coli. Journal of biotechnology *156*, 245-252.

The Review on Antimicrobial Resistance, C.b.J.O.N. (2014). Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations.

Trotter, C.L., McVernon, J., Ramsay, M.E., Whitney, C.G., Mulholland, E.K., Goldblatt, D., Hombach, J., and Kieny, M.P. (2008). Optimising the use of conjugate vaccines to prevent disease caused by Haemophilus influenzae type b, Neisseria meningitidis and Streptococcus pneumoniae. Vaccine *26*, 4434-4445.

Valvano, M.A., and Crosa, J.H. (1989). Molecular cloning and expression in Escherichia coli K-12 of chromosomal genes determining the O7 lipopolysaccharide antigen of a human invasive strain of E. coli O7:K1. Infection and immunity *57*, 937-943.

Wacker, M., Wang, L., Kowarik, M., Dowd, M., Lipowsky, G., Faridmoayer, A., Shields, K., Park, S., Alaimo, C., Kelley, K.A., *et al.* (2014). Prevention of Staphylococcus aureus infections by glycoprotein vaccines synthesized in Escherichia coli. J Infect Dis *209*, 1551-1561.

Wahl, B., O'Brien, K.L., Greenbaum, A., Majumder, A., Liu, L., Chu, Y., Luksic, I., Nair, H., McAllister, D.A., Campbell, H., *et al.* (2018). Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000-15. The Lancet Global health *6*, e744-e757.

Weintraub, A. (2003). Immunology of bacterial polysaccharide antigens. Carbohydrate research *338*, 2539-2547.

Wetter, M., Kowarik, M., Steffen, M., Carranza, P., Corradin, G., and Wacker, M. (2013). Engineering, conjugation, and immunogenicity assessment of Escherichia coli O121 O antigen for its potential use as a typhoid vaccine component. Glycoconj J *30*, 511-522.

WHO (2014). Temperature Sensitivity of Vaccines.

CHAPTER 4

A UNIVERSAL GLYCOENZYME BIOSYNTHESIS PIPELINE THAT ENABLES CELL-FREE CONSTRUCTION AND REMODELING OF GLYCANS[12]

4.1 Abstract.

Cell-free synthetic glycobiology has emerged as a simplified and highly modular framework to investigate, prototype, and engineer pathways for glycan biosynthesis and biomolecule glycosylation outside the confines of living cells. These developments notwithstanding, constructing *de novo* glycan biosynthesis pathways remains a significant challenge due to limited accessibility to functionally characterized glycoenzymes. To address these challenges, we describe a novel glycoenzyme biosynthesis pipeline that facilitates high-level, soluble production of functional glycosyltransferases (GTs) across multiple expression platforms. This technology leverages a protein engineering strategy to remodel GTs with an N-terminal decoy protein that prevents membrane insertion and a C-terminal amphipathic protein that effectively shields hydrophobic surfaces from the aqueous environment. We demonstrate the modularity and universality of the approach through the facile production of more than 100 difficult-to-express GTs across several diverse expression platforms including mammalian, yeast, and bacterial cells, as well

12 This chapter is in preparation for journal submission.

as cell-free protein synthesis. The utility of this pipeline was further revealed by using engineered GTs to construct cell-free biosynthesis pathways that furnished several human *N*-glycans including sialylated complex-type biantennary structures on the trastuzumab biologics. Overall, the platform described here provides a simplified yet highly effective strategy for producing large quantities of diverse, enzymatically active GTs that are expected to find use in structure-function studies as well as applications in biomanufacturing of complex glycans and glycomolecules.

4.2 Introduction.

An increasing appreciation of the biological roles and the therapeutic potentials of glycans has led to an interest in the synthesis of structurally-defined carbohydrate-containing molecules. Glycosyltransferases (GTs) are key enzymes in nature for producing diverse glycans and glycoconjugates. GTs represent a subclass of glycoenzyme that catalyzes a formation of glycosidic linkages by transfer sugar molecules from a donor substrate to an acceptor.(Lairson et al., 2008) These acceptors include simple and complex carbohydrates, lipid, protein, nucleic acid, and a large group of small molecules.(Lairson et al., 2008) A majority of GTs utilize nucleotide-activated sugar as donor substrate (Leloir GTs) but there exists several GTs that recognize phosphate- and lipid-phosphate sugar donor (non-Leloir GTs).(Ardevol and Rovira, 2015) GTs constitute a large number of enzyme, with each GT displaying distinct sugar donor, acceptor, and linkage specificity; hence GTs are a powerful

catalyst for accessing complex carbohydrates including oligosaccharides, polysaccharides, glycoproteins, and glycoconjugates.

An emerging discipline of bacterial glycoengineering has made it possible to characterize natural glycosylation pathway, and to engineer novel enzymes and pathways that catalyze prescribed glycosylation reactions. Simple bacteria including the laboratory strain of *Escherichia coli* do not possess natural protein glycosylation pathway. As a result, they offer an opportunity to investigate *de novo* protein glycosylation pathway without interference from the endogenous glycosylation components.(Baker et al., 2013b; Naegeli and Aebi, 2015) Owning to this unique feature, diverse natural and synthetic system including *en bloc N*-linked(Wacker et al., 2002), successive *N*-linked(Naegeli et al., 2014b), *en bloc O*-linked(Faridmoayer et al., 2007), and successive *O*-linked(Du et al., 2018) protein glycosylation pathways have been successfully constructed in *E. coli*. These glycoengineered *E. coli* have been instrumentally useful to elucidate how the glycan is assembled, and how glycan-to-protein transfer reaction occurs within each glycosylation pathway. Glycoengineered *E. coli* cells also offer a cost-effective, renewable sources for producing biotechnologically useful glycoproteins including: i) human-derived therapeutic glycoproteins carrying authentic eukaryotic *N*- and *O*-glycan structures(Glasscock et al., 2018; Natarajan et al., 2020a; Valderrama-Rincon et al., 2012); ii) glycoconjugate vaccines against various pathogenic bacteria(Faridmoayer et al., 2007; Lithgow et al., 2014; Vozza and Feldman, 2015); and iii) recombinant glycoprotein particles(Tytgat et

al., 2019). Furthermore, these same *E. coli* cells have been used as a source strain to generate crude cell extracts capable of catalyzing protein synthesis and protein glycosylation in a single-vessel cell-free reaction(Jaroentomeechai et al., 2018b). Notably, a well-controllable environment and high flexibility of the cell-free reaction renders such system a suitable platform for rapid-prototyping novel glycosylation pathways(Kightlinger et al., 2019) as well as for an on-demand biomanufacturing of useful glycoproteins and vaccines (Stark et al., 2021).

These developments notwithstanding, construction of protein glycosylation pathway in *E. coli* cells and *E. coli*-derived cell-free platform relies heavily on the successful expression of every glycosyltransferase gene that constitutes such pathway. Many GTs are difficult to recombinantly express. Natural protein glycosylation often takes place at the membrane interface, either between cytoplasm and periplasm in the gram-negative bacteria or between cytosol and ER/Golgi organelles within higher organisms. Therefore, many GTs are either peripheral or integral membrane proteins. They also exhibit a marked diversity in term of the membrane associated topologies including single transmembrane domains (TMDs) at the N- or C-terminal or within the internal domain, multipass TMDs, enzymes with the N-terminal signal sequences, and enzymes with the C-terminal ER retention sequences. These membrane elements render GTs water-insoluble, resulting in a poor recombinant expression yield, and necessitate membrane protein purification technique. The challenge in preparing GTs at yield is pronounced when analyzing the statistics from the Carbohydrate-Active

enZYmes Database (CAZy). This database currently classified GT genes into 114 families according to their amino acid sequence similarities. As of November 2018, CAZy comprises of ~459,666 entries, about 10-fold increase from the previous decade (**Supplementary Figure 8-1a**). This rapid increase has highlighted the advance in sequence identification of new GT genes. In stark contrast, less than 1% of the total GT proteins have been characterized, and only ~232 structures are available (**Supplementary Figure 8-1b**). A vast number of GTs with useful functionality has yet to be explored but new tools and methodologies for a facile preparation of these GT biocatalysts are needed to realize this goal.

To date, only a few universal strategies for recombinant GTs expression have been described. In one instance, ~50 human GTs were fused with proteins with internal repeats (PIR) to express as immobilized enzyme on the cell surface of *Saccharomyces cerevisiae.*(Shimma et al., 2006) More than 75% of surface-displayed GTs were active and were used to construct several defined glycans including Lewis x/y, H/A/B antigens, GM1b/GD1α, and core-fucosylated biantennary *N*-glycans.(Shimma et al., 2006) More recently, an impressive effort by Moremen *et al.* combined bioinformatics, glycoenzyme design, and protein engineering strategy to generate mammalian expression library of over 300 genes comprising of all known human glycosyl hydrolases (GHs), glycosyltransferases (GTs), and glycan modifying enzymes.(Moremen et al., 2018) More than 65% of the library members were successfully expressed and secreted from mammalian cells and the baculovirus-insect

cell system. The secreted yield was sufficient for functional characterization and resulted in the structure elucidation of the ST6GALNAC2.(Moremen et al., 2018) While the aforementioned platforms provide an access to a number of GTs, a universal platform for a high yield expression of GTs in a simple organism such as *E. coli* will greatly enhance our ability to characterize and utilize natural GTs as well as to develop engineered GTs for the glycan-related therapeutic applications.

Here we report an expression library for a production of diverse GTs in standard *E. coli* expression strains. We leveraged protein engineering strategy called Solubilization of the Integral Membrane Protein with high Level of Expression or SIMPLEx to render GTs water-soluble and suitable for an expression within *E. coli* cytoplasm. The SIMPLEx strategy is universal, resulting in a successful soluble expression of more than 95% of GT candidates, many of which are of human origin. This same strategy is also compatible with several expression formats including mammalian, yeast, and *E. coli*-based cell-free protein synthesis platforms. We also demonstrate the utility of our expression library by purifying a subset of GTs for functional characterization, for remodeling glycan on glycoproteins, and for constructing cell-free biosynthesis cascades that results in the homogenous production of authentic human biantennary *N*-glycans on therapeutic antibody. Together, we anticipate our expression library will be an important resource to accelerate our fundamental understanding of the glycoenzymes and to enable the development of novel glycosylation pathways and glycoconjugated molecules.

4.3 Results.

4.3.1 SIMPLEx promotes soluble overexpression of human ST6Gal1.

To generate a versatile and universal expression platform that provides access to unexplored regions of glycoenzyme space, we hypothesized that our SIMPLEx technology could be leveraged as a protein engineering tool for relieving bottlenecks associated with GT expression. As proof-of-concept, we chose human β-galactoside-α2,6-sialyltransferase 1 (*Hs*ST6Gal1), a sialytransferase belonging to the GT29 family. *Hs*ST6Gal1 consists of a short N-terminal cytoplasmic tail (CT), a transmembrane domain (TMD), a stem region that serves as a linker, and a large C-terminal catalytic domain that adopts a GT-A fold containing a seven-stranded central β-sheet flanked by α-helices (Kuhn et al., 2013). Overexpression of *Hs*ST6Gal1 has been documented in several cancer cell types (Garnham et al., 2019); hence, the ability to produce preparative amounts of *Hs*ST6Gal1 could help to understand its role in cancer biology and therapy.

To express this enzyme in the SIMPLEx architecture, we designed a tripartite *Hs*ST6Gal1 chimera in which its N-terminus was genetically fused to a water-soluble "decoy" protein, namely the *E. coli* maltose-binding protein lacking its N-terminal signal peptide (ΔspMBP), while its C-terminus was fused to an amphipathic "shield" protein, namely truncated human apolipoprotein A1 lacking its 43-residue globular N-terminal domain (ApoAI*), yielding ΔspMBP-*Hs*ST6Gal1-ApoAI* (hereafter SIMPLEx-*Hs*ST6Gal1) (**Figure 4-1a**). As removal of the transmembrane anchor segment is a

common practice to improve expression and solubility of mammalian GTs, we also generated the SIMPLEx-Δ26*Hs*ST6Gal1 variant in which 26 amino acids from the N-terminus of *Hs*ST6Gal1, comprising its CT and TMD, were genetically removed (**Figure 4-1a)**. The *Hs*ST6Gal1 enzyme contains 3 disulfide bonds in its native structure (Kuhn et al., 2013). Therefore, an engineered *E. coli* strain named SHuffle T7 Express (Lobstein et al., 2012), which possesses a more oxidizing cytoplasmic environment and expresses a cytoplasmic version of the disulfide bond isomerase DsbC, was selected as an expression host to facilitate cytoplasmic disulfide bond formation. Following expression of the two engineered chimeras in SHuffle T7 Express cells, we observed stable products corresponding to SIMPLEx-*Hs*ST6Gal1 and SIMPLEx-Δ26*Hs*ST6Gal1 that accumulated almost exclusively in the soluble cytoplasmic fraction (**Figure 4-1b**).

In stark contrast, no detectable expression of unfused *Hs*ST6Gal1 or Δ26*Hs*ST6Gal1 was seen in the soluble fraction and only minimal amounts were observed in the insoluble and detergent-solubilized fractions (**Figure 4-1b**), in agreement with previous findings that human sialyltransferases are poorly expressed in bacteria (Ortiz-Soto and Seibel, 2016a; Skretas et al., 2009). To demonstrate the importance of the decoy and shield domains, we also expressed chimeras lacking each of these elements. When the decoy protein was omitted, Δ26*Hs*ST6Gal1-ApoAI* partitioned almost entirely in the insoluble fraction, whereas omission of the shield protein resulted in accumulation ΔspMBP-Δ26*Hs*ST6Gal1 primarily in the soluble

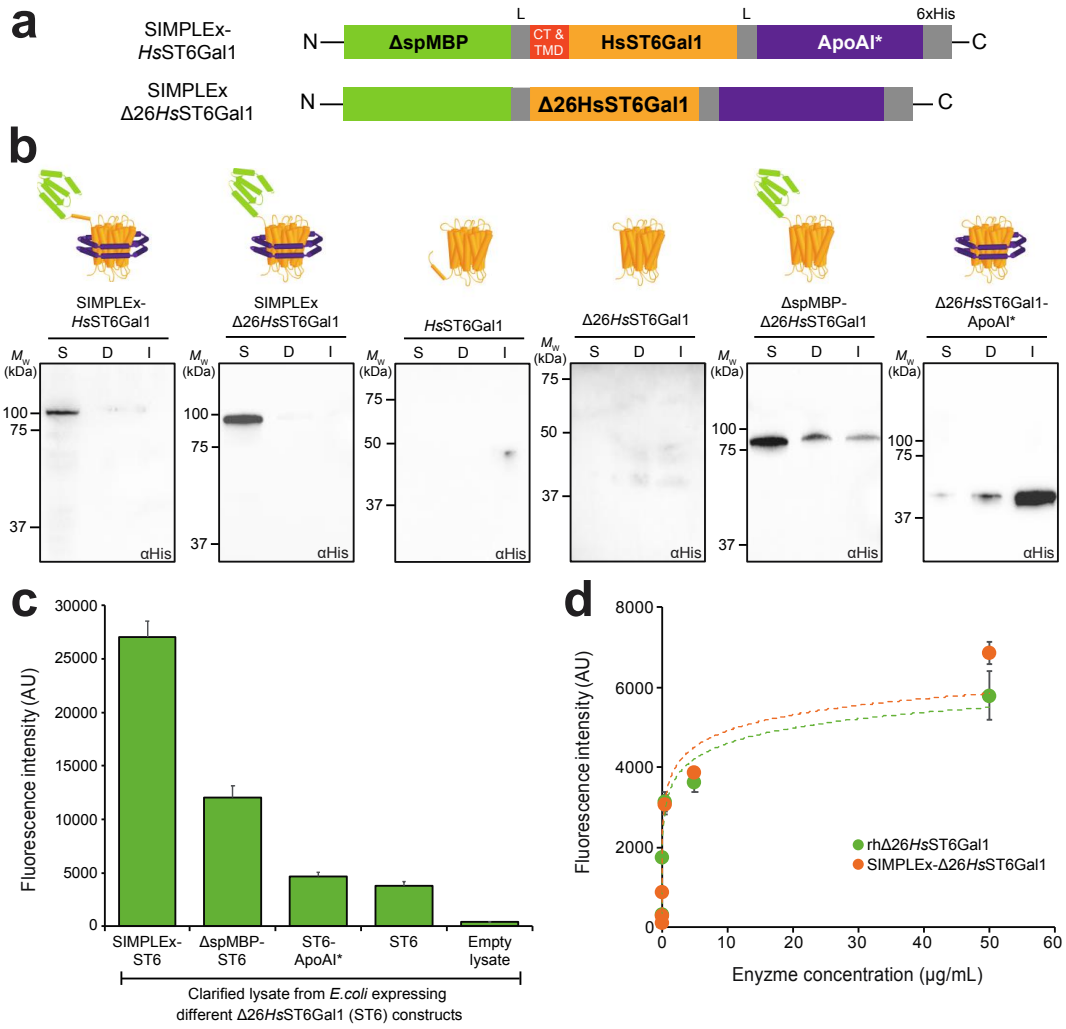


Figure 4-1. SIMPLEx promotes soluble expression of biologically-active *Hs*ST6Gal1. (a) Molecular design of the SIMPLEx constructs used in this study. Each construct consisted of an N-terminal ΔspMBP (green) and C-terminal ApoAI* (purple) that flanked *Hs*ST6Gal1 (yellow). The intervening flexible linkers (L, gray) were placed to connect ΔspMBP and ApoAI* to the *Hs*ST6Gal1 domain, and the hexahistidine tag (6xHis, gray) was placed at the C-terminus to facilitate detection and purification. The *Hs*ST6Gal1 variants studied here were wild-type (wt) *Hs*ST6Gal1 (top) and Δ26*Hs*ST6Gal1 (bottom), in which the cytoplasmic tail (CT, orange) and transmembrane domain (TMD, orange) were removed. (b) Western blot analysis of the soluble (S), detergent-solubilized (D), and insoluble (I) fractions prepared from *E. coli* SHuffle T7 Express *lysY* cells carrying plasmid pET28a(+) encoding SIMPLEx-*Hs*ST6gal1, *Hs*ST6Gal1, SIMPLEx-Δ26*Hs*ST6Gal1, ΔspMBP-Δ26*Hs*ST6Gal1, Δ26*Hs*ST6Gal1-ApoAI*, or Δ26*Hs*ST6Gal1,

as indicated. Blots were probed with anti-6xHis antibody. Results are representative of at least three biological replicates. Molecular weight (*M*w) markers are shown on the left. (c) Fluorescence (501/523 nm Ex/Em) corresponding to *in vitro* sialytransferase activity of the clarified lysates prepared from *E. coli* cells expressing one of the following constructs: SIMPLEx-Δ26*Hs*ST6Gal1, ΔspMBP-Δ26*Hs*ST6Gal1, Δ26*Hs*ST6Gal1-ApoAI, or Δ26*Hs*ST6Gal1, as indicated. Clarified lysate generated from *E. coli* cells carrying empty pET28a(+) plasmid was used as a negative control. (d) Fluorescence corresponding to *in vitro* sialytransferase activity of the purified SIMPLEx-Δ26*Hs*ST6Gal1 and commercial Δ26*Hs*ST6Gal1 standard as a function of enzyme concentration. Activity data are the mean of three biological replicates (starting from freshly transformed cells for each condition) and error bars represent standard error of the mean.

fraction, but with significant amounts also detected in the detergent-solubilized and insoluble fractions (**Figure 4-1b**). Our results confirm the importance of the decoy and shield in directing folding away from the membrane and promoting water solubility. Moreover, SIMPLEx-based expression in a redox-engineered bacterial host sidestepped the need for chaperones that occur uniquely in the mammalian secretory pathway and for *N*-glycosylation of the GT that is not required for activity but required for folding, stability and solubility of the enzyme (Chen and Colley, 2000; Meng et al., 2013).

4.3.2 Human ST6Gal1 in the SIMPLEx framework retains biological activity.

We next sought to determine whether the *Hs*ST6Gal1 enzyme in the SIMPLEx format was biologically active. Native ST6Gal1 catalyzes transfer of *N*-acetylneuramic acid (Neu5Ac) from CMP-Neu5Ac to the C6 hydroxyl group of terminal galactose-

containing acceptors. To evaluate this activity, we developed a bioorthogonal click chemistry-based assay for quantifying sialytransferase activity *in vitro*. Specifically, purified or clarified *E. coli* lysate containing SIMPLEx-Δ26*Hs*ST6Gal1 was used to transfer azido-Neu5Ac from CMP-activated glycosyl donor onto terminal galactoses within *N*-linked glycans of the alpha-1 antitrypsin (A1AT) serpin protein, which was first treated with neuraminidase to remove native sialic acids (**Supplementary Figure 8-3a**). The modified A1AT was then fluorescently labeled through a strain-promoted azide-alkyne cycloaddition (SPAAC) reaction using carboxyrhodamine 110 DBCO (**Supplementary Fig. 8-3a**). Following separation from the mixture using standard protein electrophoresis, fluorescent intensity of the labeled A1AT proteins, which corresponded to the degree of sialytransferase activity of the enzyme, could then be directly visualized and quantified by SDS-PAGE gel analysis (**Supplementary Figure 8-3b**) with commercial Δ26*Hs*ST6Gal1 serving as the standard for assay calibration (**Supplementary Figure 8-3c**).

Using clarified lysate generated from *E. coli* cells expressing SIMPLEx-Δ26*Hs*ST6Gal1 as a catalyst source, we detected strong fluorescence from the labeled A1AT, indicating high sialyltransferase activity in this lysate (**Figure 1c**; estimated to be 0.66 units per liter shake flask culture). As anticipated, clarified lysates containing either Δ26*Hs*ST6Gal1 or Δ26*Hs*ST6Gal1-ApoAI* yielded only a weak fluorescent signal (**Figure 1c**), which was consistent with the extremely poor soluble expression observed for these constructs that both lacked the ΔspMBP decoy. Interestingly, while the

presence of ΔspMBP led to soluble expression of ΔspMBP-Δ26*Hs*ST6Gal1, the clarified lysate activity corresponding to this construct was more than 50% lower than that measured for the SIMPLEx-Δ26*Hs*ST6Gal1 enzyme (**Figure 1b** and **1c**). These results suggest that a significant portion of the soluble ΔspMBP-Δ26*Hs*ST6Gal1 protein lacking ApoAI* is aggregates of misfolded enzyme, consistent with previous findings (Mizrachi et al., 2015) and indicative of the essential nature of both ΔspMBP and ApoAI* domains for producing soluble, active GT in the *E. coli* cytoplasm. More importantly, the *in vitro* activity of the purified SIMPLEx-Δ26*Hs*ST6Gal1 enzyme was indistinguishable from that of commercial Δ26*Hs*ST6Gal1 that was prepared using mouse myeloma cells (**Figure 1d**). The fact that ApoAI* does not measurably interfere with important C-terminal catalytic regions in Δ26*Hs*ST6Gal1 including sialyl motifs III (Tyr354–Gln357), S (Pro321–Phe343), and VS (His370–Glu375) suggests that its helical bundle structure provides sufficient flexibility to promote solubility while still allowing proper protein-glycan interactions that are necessary for native-like function of the enzyme.

4.3.3 SIMPLEx is a universal strategy for high-yield expression of diverse GTs.

Encouraged by the ability of SIMPLEx to promote human ST6Gal1 solubility in *E. coli* while preserving its biological activity, we next investigated whether the strategy could be extended to a larger collection of structurally diverse GTs. To this end, we compiled a set of 100 GT genes from diverse prokaryotic and eukaryotic organisms,

with an emphasis placed on those of human origin (**Supplementary Dataset 8-1**). These were organized according to their species of origin and activity, and included human sialytransferases (*Hs*SiaTs), human fucosyltransferases (*Hs*FUTs), human galactosyltransferases (*Hs*Gals), human glucosyltransferases (*Hs*GlcTs), human mannosyltransferases (*Hs*ManTs), human *N*-acetylgalactosyltransferases (*Hs*GalNAcTs), human *N*-acetylglucosaminyltransferases (*Hs*GlcNAcTs), other human and eukaryotic GTs, and prokaryotic GTs. Using the Uniprot database, we annotated these GTs based on the following characteristics: (i) N-terminal type II TMD; (ii) C-terminal type II TMD; (iii) multipass TMD; (iv) N-terminal signal sequence with C-terminal ER retention sequence; (v) multipass and internal TMDs; and (vi) cytosolic. It is known that N-/C-terminal TMDs as well as C-terminal ER retention domains in the mammalian GTs are used as membrane anchors, and are dispensable for catalytic activity, as was seen above for *Hs*ST6Gal1. We also observed that SIMPLEx-mediated solubility enhancement was independent of whether the TMD was present or absent. Therefore, we generally removed these TMD anchors from our design constructs. N-terminal signal sequences that natively route GTs to the proper secretory pathway, which was not necessary in the context of our *E. coli* cytoplasmic expression system, were also removed. GTs containing multipass and internal TMDs as well as predicted cytosolic GTs were designed as full-length genes. All designed constructs (see **Supplementary Dataset 8-1** for amino acid sequences) were cloned in T7 promoter-based expression vectors either as an unfused GT (full-length or truncated) with C-

terminal 6xHis tag (hereafter, GT^{6xHis}) or its tripartite SIMPLEx fusion counterpart (hereafter, SIMPLEx-GT).

The expression of all SIMPLEx-GTs was tested in small-scale, batch-mode microbial cultures. SHuffle T7 Express cells were used to produce enzymes containing previously observed or predicted disulfide bonds while BL21(DE3) cells were used to express enzymes without such bonds (**Supplementary Dataset 8-1**). Soluble, cytoplasmic expression of the SIMPLEx-GTs was profiled by immunoblot analysis of clarified lysate derived from *E. coli* cells expressing the respective constructs. Importantly, we detected strong expression for the vast majority of SIMPLEx-GT constructs with ~95% showing appreciable levels of soluble expression in the *E. coli* cytoplasm (**Figure 4-2**). It should be noted that this success rate was achieved under standard bacterial expression conditions (induction with 0.1 mM β-D-1-thiogalactopyranoside (IPTG) at 16 °C for 16-20 h in LB medium) that were identical for each construct and did not require any of the lengthy optimization trials that are commonly associated with expression campaigns employing bacteria. In stark contrast, only ~45% of the unfused GT^{6xHis} constructs could be detected in the soluble fraction under these same conditions, and in every case the level of soluble expression was lower compared to the SIMPLEx-GT counterpart (**Figure 4-2**). Subcellular fractionation analysis of select GT candidates revealed all SIMPLEx constructs accumulated predominantly in the soluble fraction whereas unfused versions of the enzymes

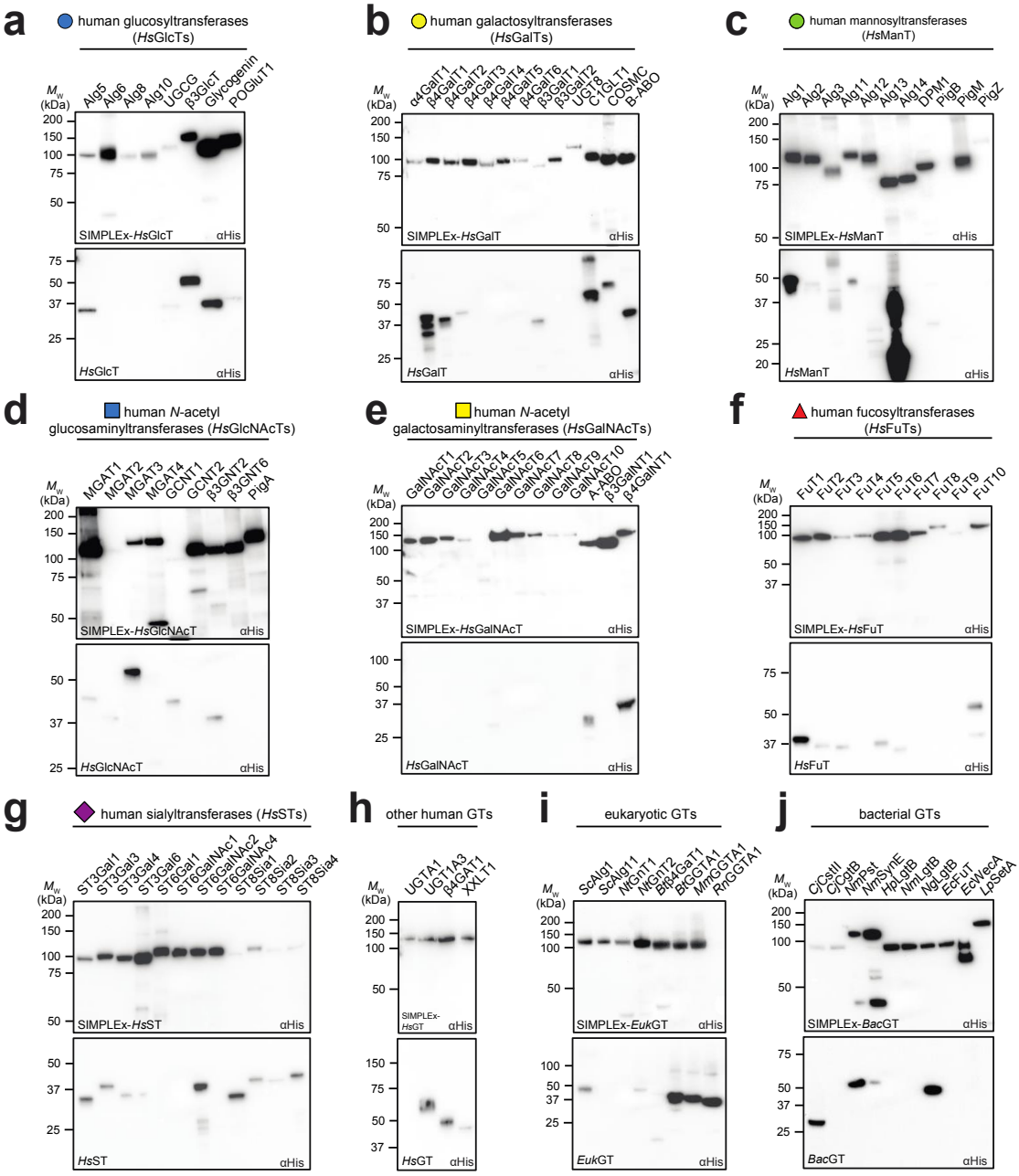


Figure 4-2. **Soluble expression of SIMPLEx-GT constructs in the *E. coli* cytoplasm.** A panel of ~100 glycoenzymes evaluated for soluble, cytoplasmic expression in the SIMPLEx framework. Immunoblot analysis of soluble fractions derived from either Shuffle T7 Express or BL21(DE3) cells carrying plasmids for either SIMPLEx-GT (top blot in each panel) or GT^{6xHis} (bottom blot in each panel) constructs. GT enzymes were clustered according to origin and activity as follows: (a) human glucosyltransferases (*Hs*GlcTs); (b) human galactosyltransferases (*Hs*GalTs);

(c) human mannosyltransferases (*Hs*ManTs); (d) human *N*-acetylglucosaminyltransferases (*Hs*GlcNAcTs); I human *N*-acetylgalactosaminyltransferases (*Hs*GalNAcTs); (f) human fucosyltransferases (*Hs*FuTs); (g) human sialyltransferases (*Hs*SiaTs), (h) other human glycosyltransferases (*Hs*GTs); (i) other eukaryotic glycosyltransferases (*Euk*GTs); and (j) bacterial glycosyltransferases (*Bac*GTs). Graphical representation of monosaccharides according to the SNFG system shown at the top of each group when possible. All blots were probed with anti-6xHis antibody to detect the protein. Results are representative of at least three biological replicates. Molecular weight (M_W) markers are indicated at left.

partitioned mostly in the insoluble or detergent-solubilized fractions (**Supplementary Figure 8-4**), consistent with the solubility profiles of the *Hs*ST6Gal1 enzyme constructs. Notably, whereas aberrant expression products such as aggregates and proteolytic degradants were prevalently detected among the GT^{6xHis} samples, the SIMPLEx-GT samples were largely devoid of such undesired products (**Figure 4-2**), highlighting added benefits of the SIMPLEx strategy in preventing aggregation and enhancing intracellular stability of target enzymes.

Another advantage of expressing GTs in the SIMPLEx framework is the potential to relieve cellular stress response due to accumulation of inclusion bodies or destabilization of the cytoplasmic membrane caused by high level expression of membrane proteins, both of which are well-known to negatively impact cell growth and productivity. Indeed, we consistently observed that cultures expressing SIMPLEx-GTs reached higher final cell densities than those expressing unfused GTs (**Supplementary Figure 8-5**). Likewise, titers of selected SIMPLEx-GT candidates produced in 1-L cultures were ~2–10 fold higher compared to their non-SIMPLEx

counterparts (**Supplementary Figure 8-6**). Taken together, these results demonstrate SIMPLEx as a universal strategy for high-yield, soluble expression of GTs having diverse origins, structures, and activities.

4.3.4 Correlates of successful GT expression in *E. coli*.

We next sought to identify the protein features that correlated with soluble protein expression by comparing physicochemical properties of the proteins such as molecular weight (M_W), isoelectric point (pI), and the amino acid content. We first assigned an expression score to each of the SIMPLEx-GT and GT^{6xHis} constructs based on their soluble expression profiles (**Supplementary Figure 8-7a**). We then used the expression score to bin these proteins into four groups: non-expressor (score 0); weak expressor (score 1); medium expressor (score 2); and strong expressor (score 3). According to this classification, ~95% of the SIMPLEx-GTs were identified as expressible (score ≥ 1), with more than 50% falling into the medium-to-high expressor groups. On the contrary, more than 50% of the GT^{6xHis} constructs were identified as non expressors, with the majority of the others classifying as weak expressors (**Supplementary Figure 8-7a**). A scatter plot relating protein M_w, excluding added mass from ΔspMBP and ApoAI* domains, with solubility score calculated by Protein-Sol, a web tool for predicting protein solubility from sequence (Hebditch et al., 2017), revealed that expressible SIMPLEx-GT constructs clustered within a 25–60 kDa range whereas expressible GTs^{6xHis} constructs were clustered in a narrower 25–40 kDa range

(**Supplementary Figure 8-7b**). This observation prompted us to further investigate the relationship between soluble expression of the protein and its M_w. To this end, we categorized all GTs into one of three size groups: small (M_w < 40 kDa), medium (M_w = 40–60 kDa), and large (M_w > 60 kDa). We then calculated weighted average expression score ($\overline{Ex}_w$) for each size group within the SIMPLEx-GT or GT^{6xHis} datasets. For GT^{6xHis} constructs, a significant decrease in $\overline{Ex}_w$ was observed as protein M_w increased, with no soluble expression for large proteins (**Supplementary Figure. 8-8a**), consistent with the observation that the bacterial translation machinery has evolved to express shorter polypeptides (Netzer and Hartl, 1997) and that expression of larger eukaryotic proteins in bacteria frequently leads to misfolding and aggregation (Dyson et al., 2004; Marston, 1986). On the contrary, $\overline{Ex}_w$ was high for all SIMPLEx-GT constructs, with no significant difference between small- and medium-sized proteins and only a small decrease in $\overline{Ex}_w$ for large-sized proteins (**Supplementary Figure 8-8a)**. These results suggest that the SIMPLEx framework helps to overcome the protein size barrier that typically restricts successful expression in *E. coli*. Unfortunately, attempts to identify additional parameters such as protein pI and amino acid components that correlated with expressibility did not yield conclusive results (**Supplementary Figure 8-7b** and **8-8b**). Nonetheless, the data presented here reveal important design parameters that could guide efforts to express novel GT enzymes in the future.

4.3.5 Efficient production of SIMPLEx-GTs across diverse expression platforms.

To further expand the utility of the platform and demonstrate its flexibility, we attempted to produce SIMPLEx fusions in a range of popular expression platforms including: (i) *E. coli*-based cell-free protein synthesis (CFPS); (ii) *Saccharomyces cerevisiae* strain SBY49; and (iii) human embryonic kidney (HEK) 293T cells. Using appropriate expression vectors for each system, the SIMPLEx-*Hs*Δ26ST6Gal1 construct was produced at significant levels in the soluble fraction across all three systems, whereas little to no expression was detected in the soluble fraction for the unfused *Hs*Δ26ST6Gal1 construct lacking the ΔspMBP and ApoAI* domains (**Figure 4-3**). While SIMPLEx-*Hs*Δ26ST6Gal1 expressed in *E. coli* CFPS was also detected in the insoluble fraction, the amount of product that partitioned in the soluble fraction was significantly higher (**Figure 4-3a**). For cell-based expression of SIMPLEx-*Hs*Δ26ST6Gal1, both yeast and human cells yielded products that accumulated almost exclusively in the soluble fractions (**Figure 4-3b** and **4-3c**), in line with the cell-based *E. coli* expression results observed above. Importantly, these results clearly demonstrate the compatibility of the SIMPLEx strategy with microbial, mammalian, and cell-free expression, and the ease with which it was adapted to these systems.

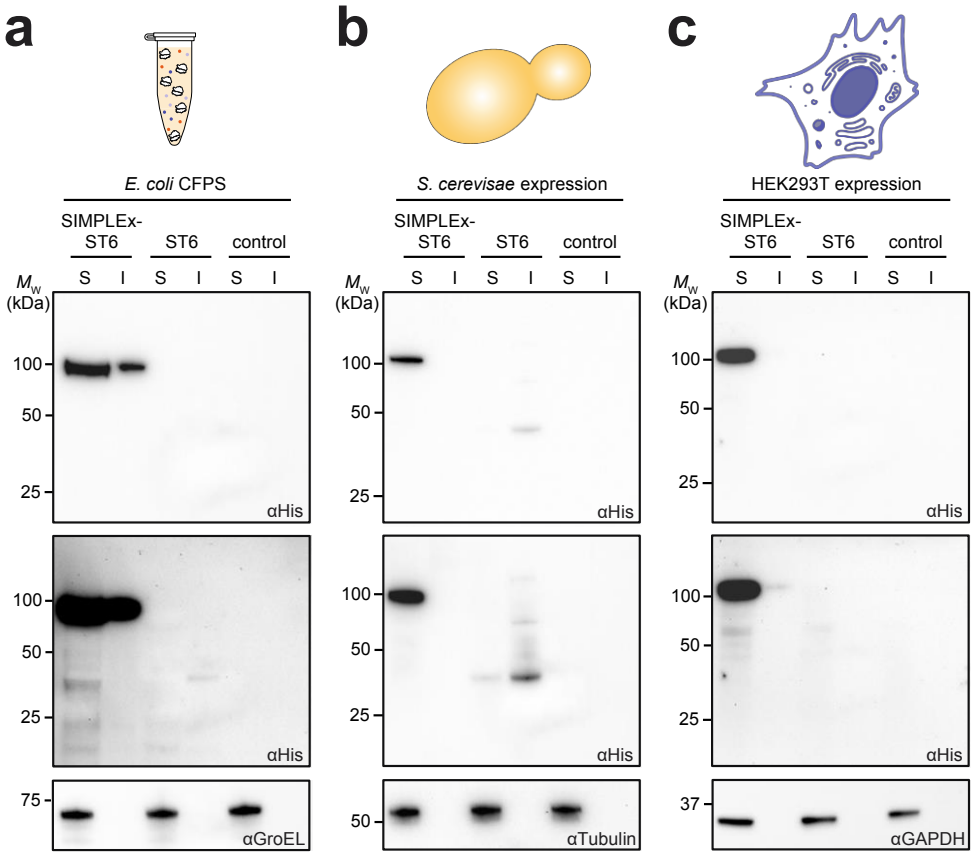


Figure 4-3. Compatibility of SIMPLEx-GTs with diverse expression platforms. Immunoblot analysis of the soluble (S) and insoluble (I) fractions derived from: (a) cell-free protein synthesis (CFPS) using crude S30 extract prepared from *E. coli* BL21(DE3); (b) cell-based expression using *S. cerevisiae* strain SBY49; and (c) cell-based expression using HEK 293T cells. All three systems involved plasmids for expressing either SIMPLEx-*Hs*Δ26ST6Gal1 (SIMPLEx-ST6) or unfused Δ26ST6Gal1 (ST6). Empty plasmid was used as a negative control (control) in each case. Blots were probed with anti-hexahistidine (*α*His) antibody with longer exposures (*α*His - long) provided to better identify protein products with low expression. An equivalent amount of total protein was loaded in each lane. To confirm, GroEL, tubulin, and GAPDH, which are commonly used housekeeping proteins in *E. coli*, yeast, and mammalian cells, respectively, were used as immunoblot loading controls. Loading control blots were probed with antibody specific to each housekeeping protein, as indicated. Results are representative of at least three biological replicates. Molecular weight (*M*w) markers are shown on the left.

4.3.6 Cell-free glycan construction using SIMPLEx-GTs yields human *N*-glycans.

Pure and chemically-defined glycans are valuable reagents for fundamental studies and biomedical applications in glycoscience. Cell-free bio/chemoenzymatic synthesis has recently emerged as one of the most effective platforms to access large repertoires of glycans (Hamilton et al., 2017; Li and Wang, 2018; Li et al., 2019d), especially for complex structures that are otherwise difficult to obtain by conventional chemical synthesis. These approaches, however, are usually restricted by the availability of the synthesis glycoenzymes, many of which cannot be recombinantly expressed or purified at scale. To address this gap, we sought to demonstrate the utility of SIMPLEx-GTs as catalysts for the construction of customized glycan structures via a previously described bioenzymatic synthesis approach (Hamilton et al., 2017). As proof-of-concept, we devised two glycoenzyme pathways for *de novo* biosynthesis of a library of human hybrid- and complex-type *N*-glycans starting from a mannose$_3$-*N*-acetylglucosamine$_2$ (Man$_3$GlcNAc$_2$) primer (**Figure 4-4a**). To generate this primer, we leveraged a glycoengineered *E. coli* strain carrying a heterologous biosynthesis pathway for the production of undecaprenyl-linked Man$_3$GlcNAc$_2$ glycan (Valderrama-Rincon et al., 2012). Following glycolipid extraction from these cells, Man$_3$GlcNAc$_2$ (M3; glycan **1**) was removed from undecaprenol by mild acid hydrolysis and purified to homogeneity as confirmed by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) analysis (**Figure 4-4b**). Using **1** as a primer, sequential glycan elaboration was carried out using

purified SIMPLEx-*Hs*Δ29GnTI and SIMPLEx-*Hs*Δ29GnTII, yielding hybrid-type glycan **2** and complex-type glycan **3** (also known as G0), respectively (**Figure 4-4a**). Further elaboration of glycan **3** with galactose residues was achieved using SIMPLEx-*Hs*Δ44β4GalT1 to generate glycan **4** (G2), which was subsequently elaborated using SIMPLEx-*Hs*Δ26ST6Gal1 to produce glycan **5** (G2S1) and glycan **6** (G2S2), the mono- and di-sialyl complex-type *N*-glycans, respectively (**Figure 4-4a** and **Supplementary Figure 8-9a**). Alternatively, glycan **3** was first fucosylated using SIMPLEx-*Hs*Δ30FUT8 to generate glycan **7** (G0F), which was then further elaborated to yield glycan **8** (G2F), glycan **9** (G2S1F), and glycan **10** (G2S2F) using a similar bioenzymatic strategy (**Figure 4-4a** and **Supplementary Figure 8-10a**). Importantly, we observed a near 100% conversion for most enzymatic steps as evidenced by MALDI-TOF MS spectra corresponding to each reaction (**Figure 4-4b**). Due to the unstable nature of sialic acid containing glycans in MALDI-TOF MS analysis, nano-scale reverse phase chromatography and tandem MS (nano LC-MS/MS) was used to further confirm the identity of sialylated glycans **5**, **6**, **9**, and **10** (**Supplementary Figure 8-9b**,**c** and **8-10b**,**c**).

It is also worth noting that while conversions from mono- to di-sialyated products were less than ideal (**Figure 4-4b**, and **Supplementary Figure 8-9a** and **8-10a**), this phenomenon has been well documented (Barb et al., 2009; Tayi and Butler, 2018) and arises from the fact that human ST6Gal1 exhibits a preference for α1–3Man-β1,2-GlcNAc-β1,4-Gal (hereafter α1–3Man branch) and readily installs sialic acid on this branch first (Barb et al., 2009).

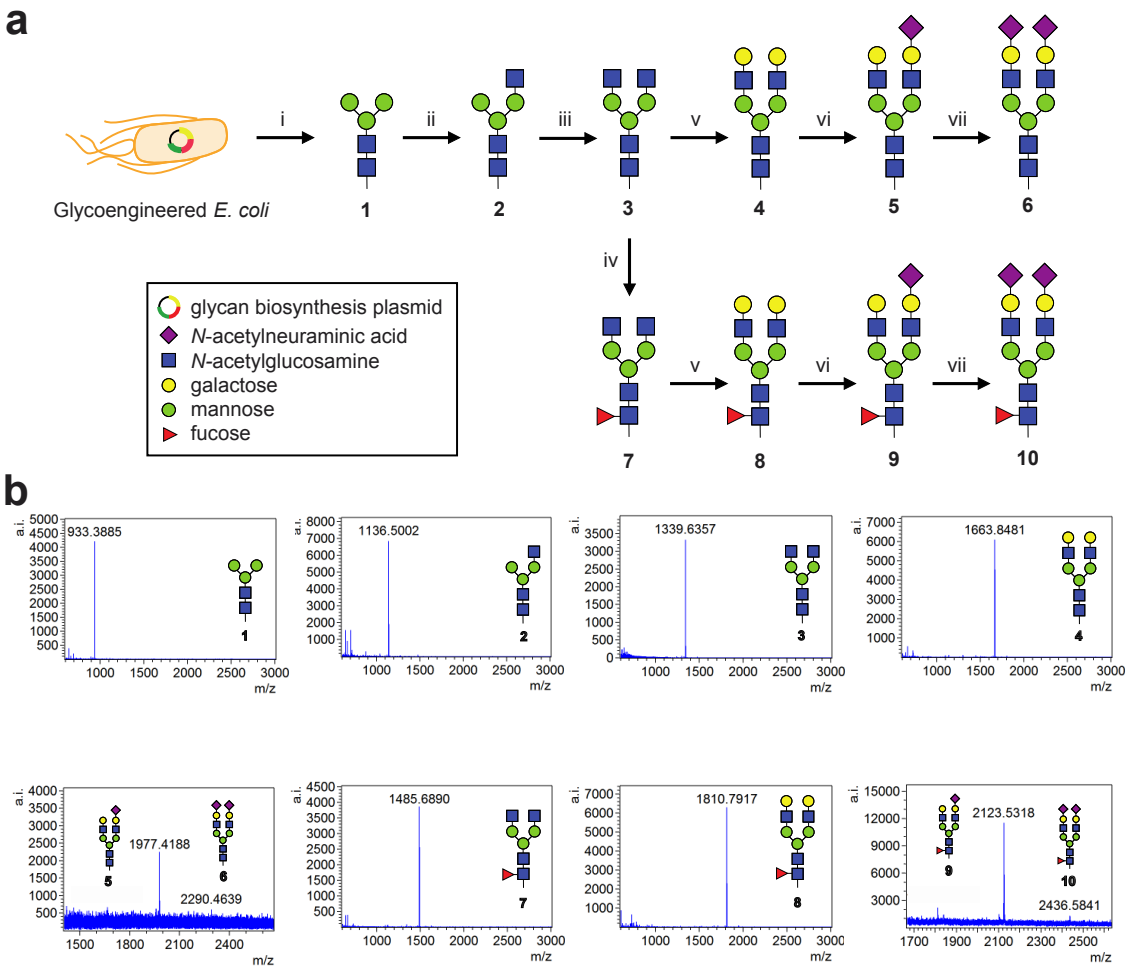


Figure 4-4. Cell-free construction of hybrid- and complex-type *N*-glycans using SIMPLEx-GTs. (a) Schematic of bioenzymatic routes to hybrid- and complex-type *N*-glycan structures. $Man_3GlcNAc_2$ glycan (M3; glycan **1**) derived from glycoengineered *E. coli* carrying eukaryotic trimannosyl core *N*-glycan biosynthesis pathway was used as primer for glycan construction. Subsequent cell-free glycan elaboration reactions yielded the following *N*-glycan structures: **2**; **3** (G0); **4** (G2); **5** (G2S); **6** (G2S2); **7** (G0F); **8** (G2F); **9** (G2S1F); and **10** (G2S2F). Glycan notation follows IgG's glycans short name system. See **Supplementary Table 2** for complete glycan list with chemical structures. Synthesis steps: (i) non-enzymatic acid hydrolysis; (ii) SIMPLEx-*Hs*Δ29GnT1; (iii) SIMPLEx-*Hs*Δ29GnT2; (iv) SIMPLEx-*Hs*Δ30FUT8; (v) SIMPLEx-*Hs*Δ44β4GalT1; (vi, vii) SIMPLEx-*Hs*Δ26ST6Gal1. (b) MALDI-TOF MS spectra of glycan **1** starting material and enzymatically-derived product glycans **2-10**. Mass (*m*/*z*) peaks for glycans **6** and **10** marked with red circle.

Subsequent sialyation of α1–6Man-β1,2-GlcNAc-β1,4-Gal (hereafter α1–6Man branch) is found to be very slow (Barb et al., 2009). Indeed, nano LC-MS/MS revealed our sialyation reactions yielded an approximate 5:1 ratio between mono- and di-sialyated products (**Supplementary Figure 8-9a** and 8-**10a**), which agreed well with previous studies (Barb et al., 2009; Tayi and Butler, 2018).

4.3.7 Cell-free glycan remodeling of therapeutic glycoproteins using SIMPLEx-GTs.

Glycoform manipulation is an emerging strategy for improving pharmacokinetics and pharmacodynamics of therapeutic glycoproteins such as monoclonal antibodies (Wang and Lomino, 2012; Wang et al., 2019). The remodeling of protein-linked glycans can be readily achieved using one or more GTs; however, the limited availability of requisite enzymes for customizing glycan structures represents a barrier to widespread adoption. To demonstrate this concept, we employed members from our library of SIMPLEx-GTs to alter the glycan profiles on several biomedically-relevant glycoproteins. Remodeling reactions included: (i) SIMPLEx-*Hs*Δ26ST6Gal1-mediated sialylation of the *N*-glycoforms on α_1-antitrypsin (A1AT), a serpin used in prophylactic treatment of the genetic disorder α_1-antitrypsin deficiency ; (ii) SIMPLEx-*Hs*Δ36FUT7-mediated fucosylation of the *N*-glycoforms on A1AT; (iii) SIMPLEx-*Hs*Δ35ST6GalNAc1-mediated sialylation of the *O*-glycoforms on bovine submaxillary mucin (BSM), a glycoprotein with potential uses as a biocompatible material and drug delivery vehicle; (iv) SIMPLEx-*Hs*Δ29GnT1-catalyzed GlcNAc transfer onto

$Man_3GlcNAc_2$ glycans present on MBP-GCG^{DQNAT}, a fusion between *E. coli* MBP and human glucagon (GCG, residue 1-29) followed by a C-terminal DQNAT glycosylation tag (Glasscock et al., 2018); and (v) SIMPLEx-*Hs*Δ24ppGalNAcT2-catalyzed *O*-GalNAcylation on MBP-GCSF, a fusion between *E. coli* MBP and human granulocyte colony-stimulating factor. In all cases, SIMPLEx-GTs readily remodeled their glycoprotein substrates, installing their respective monosaccharides in 1-h reactions that were monitored using a bioorthogonal click chemistry-based assay, as described above for sialyltransferase activity, with either a fluorophore or biotin reporter for glycan labeling (**Supplementary Figure 8-11**). It should be noted that significantly increased activity was observed for SIMPLEx-*Hs*Δ36FUT7 when the *N*-glycans on A1AT were pre-treated with neuraminidase to remove native sialic acid. This observation was in line with earlier reports and reveals that subtle differences in substrate specificity can be directly investigated using GTs within the SIMPLEx framework.

4.3.8 SIMPLEx glycoenzymes yield therapeutic IgG bearing homogenous N-glycoforms.

Fc N-glycan is a critical modulator for the effector function of the therapeutic antibody. Hence efforts aim at generating a structurally-defined N-glycan on IgG-Fc are expected to improve our understanding of the glycan role in human immunity as well as to broaden the applications of IgG-based biologics. To this end, we leveraged a facile access to the human glycoenzymes to generate therapeutic IgG Trastuzumab

bearing homogeneous N-glycan structures (**Figure 4-5a**). First, Trastuzumab was prepared from the Expi293F™ GnTI-, a glycoengineered cell line that homogeneously produces N-glycoprotein bearing $Man_5GlcNAc_2$ glycan (glycan **11**). Glycosidase sensitivity assay coupled with LC-MS analysis of the intact antibody was used to confirm the identity of glycan on the Trastuzumab (**Supplementary Figure 8-12**). Next, SIMPLEx-*Hs*Δ29GnTI was used to install N-acetylglucosamine on the ⌾1,3-man branch of **11** to generate $GlcNAcMan_5GlcNAc_2$ glycan (glycan **12**) directly on Trastuzumab (**Figure 4-5b**). Human Golgi Man2A1 was then used to remove two terminal mannose residues on the α1,6-man branch of **12** and yielded Trastuzumab bearing glycan **2**. Subsequent cell-free glycan remodeling reactions using SIMPLEx-*Hs*Δ29GnTII and SIMPLEx-*Hs*Δ44β4GalT1 furnished Trastuzumab with glycan **3** and **4**, respectively. Finally, SIMPLEx-*Hs*Δ26ST6Gal1 was used to cap glycan **4** with sialic acid moiety to generate glycan **5** and **6** (**Figure 4-5b**).

In addition to a construction of the authentic human N-glycans, we have successfully generated unnatural glycan structures on the IgG-Fc. SIMPLEx-*Hs*Δ29GnTI was used to install *N*-azidoacetylglucosamine, a synthetic monosaccharide containing azide moiety, and yielded glycan 13 on Trastuzumab. This azide group provides a versatile chemical handle for regiospecific bioconjugation via bioorthogonal click chemistry. We leveraged this concept to site-specifically install fluorescent reporter or biotin group on Trastuzumab's N-glycan. Importantly, SIMPLEx glycoenzymes were able to afford homogeneous N-glycan structures on IgG-Fc as

evidenced by a complete conversion from the mass of substrate to product in an MS analysis (**Figure 4-5b** and **Supplementary Figure 8-13**). Taken together, our IgG-Fc glycan construction and remodeling results showcase the propensity of SIMPLEx to furnish soluble GTs with full biological activity and thus should prove useful to increase our understanding of these enzymes as well as to enable an on-demand biomanufacturing of customized glycomolecules.

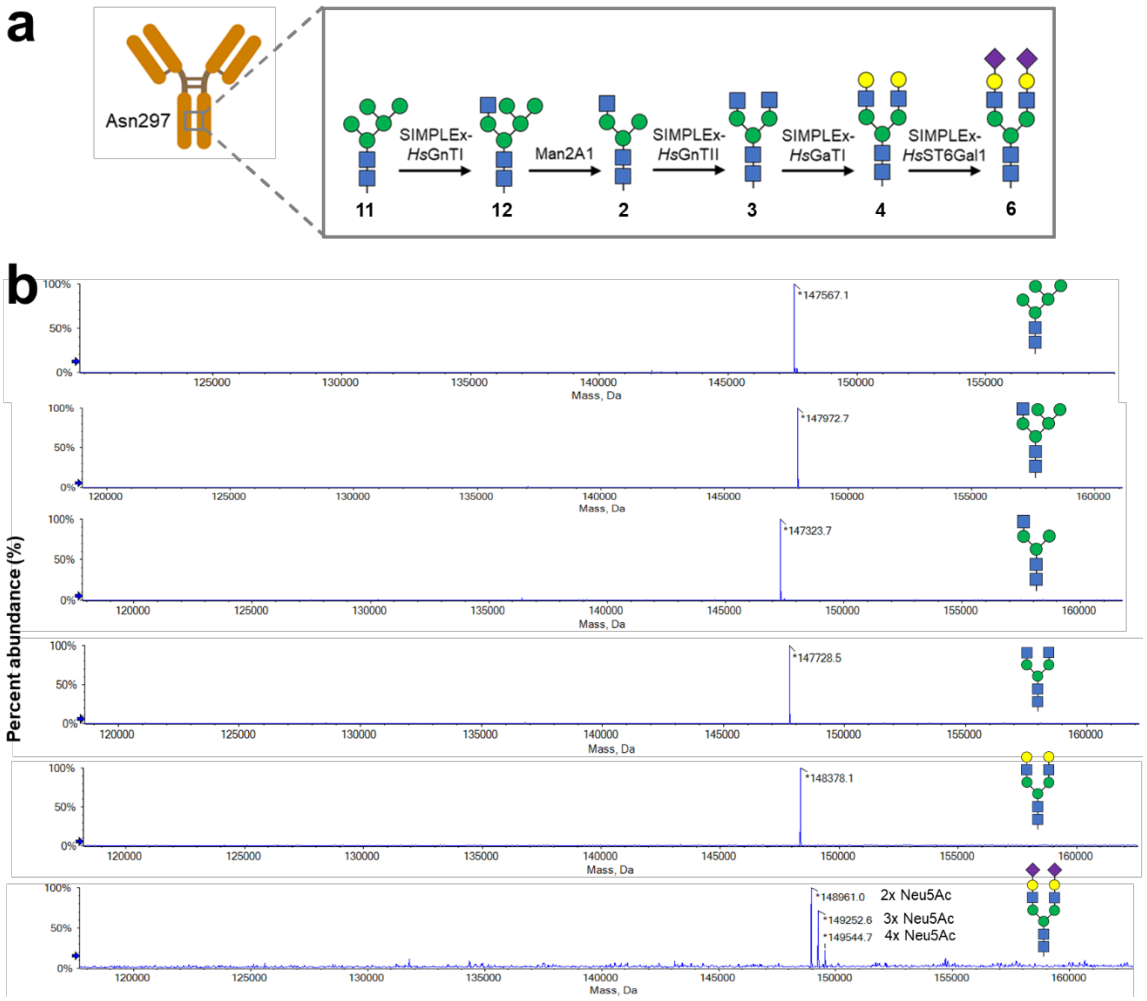


Figure 4-5 SIMPLEx glycoenzymes remodel IgG-Fc N-glycan on Trastuzumab. (a) Schematic of bioenzymatic routes to hybrid- and complex-type N-glycan structures linked to the asparagine 297 of the antibody. Trastuzumab bearing Man5GlcNAc2 glycan (M5; glycan 11) derived from glycoengineered HEK293F lacking GnTI activity was used as a glycan primer. Subsequent cell-free glycan remodeling reactions yielded the following N-glycan structures: **12** (M5+GlcNAc); **2** (G0-GlcNAc); **3** (G0); **4** (G2); and **6** (G2S2). Glycan notation follows IgG's glycans short name system. See **Supplementary Table 11-2** for complete glycan list with chemical structures. SIMPLEx glycoenzyme for each synthesis step is provided above reaction arrow. (b) LC-MS spectra using intact antibody analysis of Trastuzumab bearing glycan 11 as a starting material and enzymatically-derived product glycans **2**-**4**, **6**, and **12**. Structure of the anticipated N-glycan products is provided at the end of each spectrum.

4.4 Discussion.

In this work, we created a universal expression library for producing large quantities of GTs from a standard bacterial culture. Using our understanding of the membrane protein biogenesis, we use multipronged approach to engineered GT targets. First, we design GT construct where its membrane associated elements that are not necessary for their catalytic activity are truncated. The removal of highly hydrophobic element at the N-terminal is expected to prevent ribonucleoprotein particle from co-translationally inserts polypeptide into inner membrane through signal recognition particle (SRP) pathway.(Luirink and Sinning, 2004) It is important to note that a removal of the N-terminal highly hydrophobic peptide alone might not be sufficient. As many GTs are type II membrane protein, they often contain several mildly or medium hydrophobic regions within their polypeptide including around the

stem region at the N-terminus. Even a mildly hydrophobic peptide at the N-terminus can be recognized by FtsY receptor as signal sequence, and will be subjected to a membrane insertion *via* the SecB pathway.(Randall and Hardy, 2002) Indeed, this hypothesis was partially confirmed as we exclusively detected several GTs with truncated N-terminal TMD in the insoluble or detergent-solubilized fractions of the *E. coli* cells. Therefore, our second prong is to fuse a highly water-soluble protein at the N-terminus of the TMD-truncated GT. We dub this N-terminal fused protein a decoy protein from its role in rerouting the biogenesis of GT from inner membrane to soluble protein pathways. We rationale the N-terminal hydrophilic peptide of the decoy protein will effectively prevent ribosome-SRP complex formation, the first key step in the inner membrane protein biogenesis. In this work, we select truncated *E. coli* MBP variant as the decoy protein due to its relatively high solubility (protein solubility score: 0.586 vs 0.45, which is the population average of the soluble *E. coli* protein).(Hebditch et al., 2017; Niwa et al., 2009) In addition, MBP is one of the most popular protein fusions as it has proven to improve solubility, stability, and expression yield of difficult-to-express proteins including several GTs.(Lauber et al., 2015; Ortiz-Soto and Seibel, 2016b; Skretas et al., 2009) While the benefits of MBP fusion in this work is in line with several other reports(Lauber et al., 2015; Ortiz-Soto and Seibel, 2016b; Skretas et al., 2009), we observe an appreciable amount of MBP-fused GT expressed as an insoluble aggregate in the *E. coli* cytoplasm. This observation has indeed prompted us to further design GT variants by fusing its C-terminus with the

amphipathic protein, in this case truncated human ApoAI. As mentioned earlier, native GTs typically contain several hydrophobic regions to stabilize membrane localization, but these hydrophobic patches can also initiate the aggregate formation in aqueous solution. We reason that amphipathic protein, when placed in proximity with GTs, will help stabilize GTs in water *via* hydrophobic interaction, a similar mechanism with the use of nanodisc to solubilize membrane protein.(Grinkova et al., 2010) It is worth noting that while there are several methods to prevent protein aggregation including amino acid mutation or supplementing solution with detergent, these methods often require extensive optimization to identify optimal condition for individual protein.(Wagner et al., 2006) In contrast, our method is more universal and can be readily apply to new GT targets, simply by cloning a new gene into an expression template. The universality of the SIMPLEx strategy was exemplified by a creation of 100 GTs expression library with >95% successful expression using only one expression condition in *E. coli* culture. More importantly, GTs expressed within the SIMPLEx framework were found to be functionally active and, in the case of ST6Gal1, had similar catalytic profile with native enzyme.

It is intriguing to establish the main determinants for a successful expression of protein from its amino acid sequence. By analyzing expression data at hand, we concluded that SIMPLEx fusion relaxed M_w limit for the cytoplasmic expression in *E. coli*. Specifically, it improves expression success rate of the small (M_w < 40 kDa) and medium (M_w = 40–60 kDa) size GTs by more than 70%. Most notably, SIMPLEx design

allows the expression of several large (M_w > 60 kDa) size GTs, whose expression is completely absent without SIMPLEx fusion partners. The median protein size within the *E. coli* proteome is at ~30-40 kDa, (Brocchieri and Karlin, 2005) and recombinant expression of large protein (M_w > 60-70 kDa) in *E. coli* is known to be challenging. While our further analysis using solubility score and isoelectric point of the GTs do not yield any definite trends, the mechanism behinds expression enhancement through SIMPLEx fusion is an interesting question that warrants further investigation.

One advantage of the SIMPLEx strategy lies in its expression host flexibility. While we initially developed SIMPLEx-GT design with *E. coli* as a recombinant host in mind, to our surprise, we found this same design improved recombinant expression across different platforms including yeasts, HEK293 cells, and *E. coli*-based cell-free protein synthesis (CFPS). Protein expression in yeasts and HEK293 cells are well-established and both platforms are amenable to a scale-up in large culture. As eukaryotes, these cells are able to perform complex protein folding and protein post-translational modifications including *N*- and *O*-linked glycosylation, some of which are important for the biological function for a subset of GTs.(Mikolajczyk et al., 2020) As an emerging platform, *E. coli*-based CFPS offers an unprecedented speed and simplicity for biosynthesizing protein targets. These features make CFPS a suitable platform for a high-throughput screening of GT's function.(Kightlinger et al., 2018) The compatibility of SIMPLEx-GTs in diverse expression platforms is expected to accelerate characterization of novel GTs as well as broaden their applications in glycoscience.

In this work, we demonstrated a utility of our SIMPLEx glycoenzymes for a cell-free construction of chemically-defined N-glycans. We used glycoengineered E. coli to generate a large quantity of Man3GlcNAc2 N-glycan as a starting material.10 We then used E. coli-derived SIMPLEx-GTs to complete cell-free biosynthesis of the complex biantennary N-glycans including those containing core-fucose and sialic acid caps. This workflow, starting from SIMPLEx-GTs expression, purification, and utilization in the cell-free reactions to efficiently generate the entire library of complex N-glycans, is achieved within 5-7 days. Most importantly, this same strategy was applied to afford a production of homogenous N- and O-glycan structures on several therapeutically-relevant glycoproteins including full-length Trastuzumab. Looking forward, we anticipate the platform described here could find use in the scalable biosynthesis of diverse glycoenzymes, that will help expanding our knowledge and application in glycobiology and glycomedicine.

4.5 Materials and Methods.

Strains and cell lines.

All bacterial, yeast and mammalian cells used in this study are listed in **Supplementary Table 8-1**. *E. coli* strain DH5α was used for all molecular biology and plasmid storage. *E. coli* strain BL21(DE3) and its derivative SHuffle T7 Express *lysY* (New England Biolabs) were used for all protein expression and purification. Luria-Bertani medium (LB) was used to culture *E. coli* in all experiments and was

supplemented with appropriate antibiotics for plasmid maintenance. The final concentration for each antibiotic used was: 50 μg/mL kanamycin, 20 μg/mL chloramphenicol, and 100 μg/mL ampicillin. Yeast strain SBY49 was kindly provided by Dr. Scott Emr (Cornell University). Yeast cells were grown in complex yeast extract peptone dextrose (YPD) medium or yeast nitrogen base (YNB) medium without amino acids supplemented with uracil dropout amino acids (-URA media) for plasmid maintenance. HEK293T cells were obtained from ATCC and cultured in DMEM supplemented with 10% FetalClone (VWR), 4.5 g/L glucose and L-glutamine, and 1% (w/v) penicillin-streptomycin-amphotericin B (Thermo Fisher Scientific). Expi293F™ GnTI- cells were obtained from GibcoTM. Cells were cultured in Expi293™ Expression Medium supplemented with 1% (w/v) penicillin-streptomycin-amphotericin B (ThermoFisher Scientific). 293 cells were maintained in 37 °C incubator with 5% CO2 and 90% relative humidity.

Plasmid construction.

All plasmids used in this study are listed in **Supplementary Table 8-1**. The collection of prokaryotic and eukaryotic glycoenzymes was selected from the CAZy database (Lombard et al., 2014). Amino acid sequences were all extracted from Uniprot database (UniProt, 2019). Each GT coding region was examined for membrane domains from UniProt database to determine the TMD topology. For type II membrane proteins, cytoplasmic and helical TMD segments were truncated. Stem regions were

generally retained, unless stated otherwise. For other classes, N-terminal signal sequences and C-terminal transmembrane segments were generally removed. GTs with internal or multipass TMD were expressed as full-length proteins. Amino acid sequences of full-length and truncated variants of all GTs in this study are provided in **Supplementary Dataset 8-1**. All GT genes were codon-optimized for expression in *E. coli* using GeneArt software (Thermo Fisher Scientific). These genes were then synthesized and ligated into the previously described SIMPLEx plasmid (Mizrachi et al., 2015) to generate SIMPLEx-GT plasmids having the form pET28a(+)-MBP-(*Nde*I)-GT-(*Eco*RI)-ApoAI*-6xHis. PCR was used to amplify each GT gene with flanking *Nco*I and *Not*I restriction sites, and then ligated into pET28a(+) vector to create plasmids for expression of unfused GT^{6xHis} constructs having the form pET28a(+)-(*Nco*I)-GT-(*Not*I)-6xHis. All PCR reactions were performed using 0.1 μM gene-specific primers, 50 ng DNA template, and Phusion® High-Fidelity DNA Polymerase (New England Biolabs). Ligation products were used to chemically transform *E. coli* DH5α, and the transformation cultures were plated on LB-agar plates containing kanamycin. Clones were selected and screened by colony PCR using 2x-OneTaq quickload master mix (New England Biolabs). Successful clones were confirmed by Sanger sequencing at the Cornell Biotechnology Resource Center. Due to incompatibility of DNA restriction sites, plasmids used for expression in yeast and mammalian cells were constructed using Gibson assembly. Briefly, standard PCR was used to amplify target genes containing 20-25 bp homologous regions with vector at both ends. 50 ng of linearized

vector and 150 ng of amplified insert were then combined in a Gibson Assembly Master Mix (New England Biolabs) and incubated for 1 h. Assembly reactions were then used to transform *E. coli* DH5α, after which clones were screened and confirmed according to a similar procedure as described above.

Small-scale expression and subcellular fractionation.

Plasmids encoding SIMPLEx-GT and unfused GT^{6xHis} constructs were used to transform one of these *E. coli* strains: BL21(DE3) for GTs containing no disulfide bonds and SHuffle T7 Express *lysY* for GTs contain predicted or confirmed disulfide bonds. Small 5.0-mL LB cultures of *E. coli* harboring a SIMPLEx-GT or GT^{6xHis} plasmid were grown to an optical density at 600 nm (OD_{600}) of ~0.6-0.8 and then induced with IPTG to a final concentration of 0.1 mM. Protein expression proceeded for 18 h at 16 °C, after which culture volumes equivalent to OD_{600} of 2.0 were harvested. Media was removed by centrifugation and the resulting cell pellet was resuspended in 1 mL phosphate buffer saline (PBS). Cells were lysed using a Q125 Sonicator (Qsonica) with a 3.175-mm diameter probe at a frequency of 20 kHz and 40% amplitude. Lysate was then centrifuged at 15,000xg for 30 min at 4 °C. Supernatant was collected as the soluble fraction. Pellet was then resuspended in 1 mL PBS containing 1% (v/v) Triton X-100. The suspension was incubated for 1 h at 4 °C to allow partitioning of membrane proteins into Triton X-containing buffer. Following ultracentrifugation at 100,000xg for 1 h at 4 °C, supernatant was collected as the detergent-solubilized fraction, while the

pellet was taken as the insoluble fraction. For yeast expression, SBY49 cells were grown in -URA media at 30 °C until OD_{600} reached ~0.6-0.8, after which protein expression was induced with galactose to a final concentration of 2% (w/v). Protein expression was performed for 22 h at 30 °C. Yeast cells were lysed by vortexing the cell suspension with glass beads in PBS containing zymolyase enzyme. For mammalian cell expression, 2.0 mL of HEK293T cells at ~80% confluency in a 6-well plate were transfected with 2 μg plasmid DNA using jetPRIME® transfection reagent (Polyplus Transfection). After transfection, cells were maintained in an incubator at 37 °C with 5% CO_2 and 90% relative humidity for 36 h, after which they were harvested. HEK293T cells were lysed by tip sonication. Subcellular fractionation analysis for yeast and HEK293T cells was performed similarly as described above. All samples were stored at -20 °C until further analysis.

Cell-free protein synthesis.

E. coli lysate was prepared according to an established protocol (Kwon and Jewett, 2015a). Briefly, *E. coli* strain BL21(DE3) was cultured in 2xYTPG media (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, 7 g/L potassium phosphate monobasic, 3 g/L potassium phosphate dibasic and 18 g/L glucose) at 37 °C with 0.5 mM IPTG until OD_{600} reached ~1.0. Cells were then harvested and washed twice with cold S30 buffer (10 mM tris-acetate pH 8.2, 14 mM magnesium acetate and 60 mM potassium acetate). The resulting pellet was stored at -80 °C until used. To prepare crude extract, pellets were

thawed on ice and resuspended with S30 buffer (1 mL per gram cell pellet). Cells were lysed using a Q125 Sonicator with a 3.175-mm diameter probe at a frequency of 20 kHz and 40% amplitude until the total energy input reached 1500 J. Lysate was then centrifuged twice at 30,000xg at 4 °C for 30 min. Supernatant was then collected, aliquoted, and stored at -80 °C until used. Cell-free synthesis of SIMPLEx-GT and unfused GT^{6xHis} constructs was performed using the modified PANOx-SP system (Jewett and Swartz, 2004c). Specifically, S30 lysate was pre-conditioned with 750 μM iodoacetamide in the dark at room temperature for 30 min and then lysate was supplemented with 200 mM glutathione at a 3:1 ratio between oxidized and reduced forms. Then, 200 ng plasmid DNA was introduced into cell-free protein synthesis reaction containing 30% (v/v) S30 lysate and the following: 12 mM magnesium glutamate, 10 mM ammonium glutamate, 130 mM potassium glutamate, 1.2 mM adenosine triphosphate (ATP), 0.85 mM guanosine triphosphate (GTP), 0.85 mM uridine triphosphate (UTP), 0.85 mM cytidine triphosphate (CTP), 0.034 mg/mL folinic acid, 0.171 mg/mL *E. coli* tRNA (Roche), 2 mM each of 20 amino acids, 30 mM phosphoenolpyruvate (PEP, Roche), 0.33 mM nicotinamide adenine dinucleotide (NAD), 0.27 mM coenzyme-A (CoA), 4 mM oxalic acid, 1 mM putrescine, 1.5 mM spermidine, and 57 mM HEPES. The synthesis reaction was carried out at 30 °C for 6 h, after which the sample was centrifuged at 15,000xg for 30 min at 4 °C. Supernatant was collected and stored at -20 °C until further analysis.

Human MAN2A1 expression and purification.

An expression construct encoding the truncated catalytic domain of human MAN2A1 was generated as an NH2-terminal fusion protein in the pGEn2 expression vector. The NH2-terminal fusion sequences were composed of a 25-amino acid signal sequence, an His8 tag, AviTag, the "superfolder" GFP coding region, the 7-amino acid recognition sequence of the tobacco etch virus (TEV) protease followed by the human MAN2A1 catalytic domain sequence (UniProt Q16706, residues 27-1144). Recombinant human MAN2A1 was expressed as a soluble secreted protein by transient transfection of suspension culture HEK293-F cells (FreeStyleTM 293-Fcells, Thermo Fisher Scientific) maintained at 0.5–3.0x106 cells/ml in a humidified CO2 platform shaker incubator at 37°C with 50% humidity. Transient transfection was performed using HEK293F cells at a density of 2.5-3.0x106 cells/ml in expression medium comprised of a 9:1 ratio of FreestyleTM293 expression medium (Thermo Fisher Scientific) and EX-Cell expression medium including Glutamax (Sigma-Aldrich). Transfection was initiated by the addition of plasmid DNA and polyethyleneimine as transfection reagent (linear 25-kDa polyethyleneimine, Polysciences). Twenty-four hours' post-transfection the cell cultures were diluted with an equal volume of fresh media supplemented with valproic acid (2.2 mM final concentration) and protein production was continued for an additional 5 days at 37°C. The cell cultures were harvested, clarified by sequential centrifugation at 1200 rpm for 10 min and 3500 rpm for 15 min at 4°C, and passed through a 0.8 μM filter (Millipore). The enzyme preparation was

adjusted to contain 25 mM HEPES, 20 mM imidazole, 300 mM NaCl, pH 7.5, and subjected to Ni-NTA Superflow (Qiagen) chromatography using a column pre-equilibrated with 25 mM HEPES, 300 mM NaCl, 20 mM imidazole, pH 7.5 (Buffer I). Following loading of the sample, the column was washed with 3 column volumes of Buffer I followed by 3 column volumes of Buffer I containing 50 mM imidazole, and eluted with Buffer I containing 300 mM imidazole at pH 7.0. The protein buffer was exchanged using an ultrafiltration pressure cell (Millipore) with a 10-kDa molecular mass cutoff membrane and adjusted to 1 mg/ml with buffer containing 20 mM HEPES, 100 mM NaCl, pH 7.0, 0.05% sodium azide, and 10% glycerol. The final protein preparation was aliquoted in 100 μl volume (1 mg/ml) and stored at -80°C until use.

Antibody expression and purification from the glycoengineered mammalian culture.

HEK293F GnTI- (ThermoFisher Scientific) was maintained according to the vendor's protocol. Following at least three passages, cells were washed and resuspended at three million cells per mL concentration. pVITRO1-Trastuzumab-IgG1/κ was prepared from E. coli culture and the purified plasmid was flowed through endotoxin removal column to remove endotoxin contaminant. Plasmid DNA-cationic lipid complex was then generated using LipofectamineTM Transfection Reagent (ThermoFisher Scientific) and was slowly added into culture media with gentle mixing. The DNA/cationic-lipid/number of cells were scaled linearly according to vendor's

protocol. Cells were maintained in an 37oC incubator shaker for 24 hours prior being supplemented with Expression Enhancer Reagents (ThermoFisher Scientific). Cell cultures were maintained at the same condition for another 5 days to allow an accumulation of antibody in the culture supernatant. Cells were then removed by centrifugation at 1000 xg for 5 mins and supernatant was filtered through 0.2-micron bottle-top filter. Supernatant was then mixed with 1xPBS at 1:1 (v/v) ratio. This solution was flowed through Mabselect SuRe resin (Sigma Aldrich) twice to allow antibody capturing via an interaction with protein A/G. Following extensive washing with 1xPBS, captured antibody was eluded from the resin using glycine solution (pH 2.0) and deposited directly into a neutralizing buffer (Tris-HCl pH 8.5). The antibody product was then buffer exchanged into 1xPBS supplemented with 0.01% sodium azide. Antibody was stored at 4oC and was stable at the described condition for at least a month.

Immunoblot analysis.

Samples were prepared by combining with NuPAGE™ 4xLDS Sample Buffer (Invitrogen) supplemented with 2.5% β-mercaptoethanol and then boiled at 100 °C for 10 min. Samples equivalent to OD_{600} of 0.375 for small-scale expression or 15 μL of CFPS reaction were loaded into each well of Bolt™ 8% Bis-Tris Plus Gels (Thermo Fisher Scientific). Following electrophoretic separation and transfer to Immobilon-P polyvinylidene difluoride (PVDF) membranes (0.45 μm), blots were washed with TBS

buffer (80 g/L NaCl, 20 g/L KCl, and 30 g/L Tris-base) followed by a 1-h incubation in blocking solution (50 g/L non-fat milk in TBS supplemented with 0.05% (v/v%) Tween-20; TBST). Blots were then washed 4 times with TBST in 10-min intervals and probed with primary antibodies. Secondary antibodies were used as needed. Blots were then washed as above. Imaging of blots was performed using a ChemiDoc™ XRS+System following a brief incubation with Western ECL substrate (Bio-Rad). Vendor, catalog number, and specific dilution of all antibodies used in this study are provided in **Supplementary Table 8-3**.

Bioorthogonal click chemistry-based GT assay.

Strain-promoted alkyne-azide cycloaddition was used to assess the activity of SIMPLEx-GTs towards glycoprotein substrates. In a typical reaction, a 1.5-mL microcentrifuge tube was charged with 20 μL of reaction mixture consisting of 1 μg purified SIMPLEx-GT or 50 μg cell lysate, 3 μg purified acceptor glycoprotein substrate, and 10 mM nucleotide-activated monosaccharide donor modified with an azide functional group. Depending on the GT reactions, these nucleotide-activated monosaccharide donors included UDP-GlcNAz, UDP-GalNAz, GDP-AzFuc, and CMP-AzNec5Ac (all from R&D Systems). Following an incubation in a 37 °C water bath for 1 h, reaction mixtures were supplemented with 2-iodoacetamide (Sigma-Aldrich) at 100 mM final concentration and incubated in the dark at room temperature for 1 h. Then, 100 mM final concentration of carboxyrhodamine 110 or biotin(PEG)$_4$

conjugated dibenzocyclooctyne-amines (Click Chemistry Tools) in *N*,*N*-dimethylformamide (DMF) was supplemented into the reaction mixture. Strain-promoted click reactions were carried out at 37 °C for 2 h. Samples were then combined with 4xLDS Sample Buffer (Invitrogen) supplemented with 2.5% β-mercaptoethanol and heated at 65 °C for 5 min. Following SDS-PAGE analysis, in-gel fluorescence from carboxyrhodamine110-linked glycans on glycoproteins was measured using a ChemiDoc™ MP Imaging System (Bio-Rad) with 501/523 nm $\lambda_{ex}/\lambda_{em}$. Biotin-linked glycans on glycoproteins were analyzed following immunoblot analysis using horseradish peroxidase conjugated streptavidin (Sigma-Aldrich) in a similar manner as described above for immunoblot analysis.

Protein purification and yield quantitation.

A single colony was selected from a transformation plate and grown overnight in LB media at 37 °C. The next day, cells were subcultured 5% into 1 L of fresh LB media. Cells were grown at 37 °C until OD_{600} reached ~0.6-0.8, after which IPTG was supplemented into culture at 0.1 mM final concentration. Protein expression proceeded at 16 °C for 18 h. Unless otherwise noted, all purification procedures were performed at 4 °C. Cells were harvested, resuspended in PBS supplemented with 10% (v/v) glycerol, and lysed by passing the cell suspension through an Emulsiflex C5 homogenizer (Avestin) twice at 15,000 psi maximum pressure. Supernatant was collected following centrifugation at 15,000xg for 30 mins and then incubated with 300

μL pre-washed HisPur™ Ni-NTA resin (Thermo Fisher Scientific) at 4 °C for 1 h. The suspension was loaded onto an Econo-Pac® gravity flow chromatography column (Bio-Rad) and resin was washed with 6 column volumes HisPur wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0). The target protein was eluted with HisPur elusion buffer (50 mM NaH_2PO_4, 300 mM NaCl, 300 mM imidazole, pH 8.0). Sample was then buffer exchanged into PBS using Zeba spin desalting columns, 7K MWCO (Thermo Fisher Scientific). Protein concentration was determined using Bradford assay (Bio-Rad). Expression yield was a representative of three biological replicates.

All other purification was performed as described above but with amylose resin (NEB) instead of Ni-NTA resin. Clarified lysate was incubated with 300 μL pre-washed amylose resin with rotation for 2 h at 4 °C. The suspension was loaded onto an Econo-Pac® gravity column (Bio-Rad) and resin was washed with 6 column volumes of amylose column buffer (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4). The target protein was eluted with amylose elusion buffer (10 mM maltose in column buffer). Protein purity and concentration were determined by Coomassie staining and Bradford assay (both from Bio-Rad), respectively. Proteins were kept at 4 °C for 2 weeks. For longer term storage at -80 °C, protein solution was supplemented with 10% (v/v) glycerol and 0.02% (w/v) sodium azide as a cryogenic agent and bacteriostat, respectively.

Cell growth analysis.

To facilitate high-throughput cell growth measurements, three individual colonies corresponding to each construct were seeded into 96-deep well plates (Eppendorf) where each well contained 100 μL LB media. Culture plates were then sealed using plate sealer and placed in an incubator shaker at 37 °C for 16 h. Then, 5 μL of the overnight culture was subcultured into fresh 100 μL LB media and incubated for 8 h, after which IPTG was supplemented to a final concentration of 0.1 mM. Protein expression proceeded at 16 °C for 18 h. To measure OD_{600}, 10 μL of each sample was mixed with 90 μL DI water in a Costar 96-well assay plate (Corning) and OD_{600} of all samples was measured in an Infinite M1000Pro spectrophotometer (Tecan).

Bioenzymatic glycan synthesis.

All glycans and nucleotide-activated monosaccharide substrates were prepared in DI water and stored at -20 °C. Glycan **1** was prepared essentially as described (Hamilton et al., 2017). To synthesize glycan **2**, 5 μg of glycan **1** was incubated with 20 μg/mL SIMPLEx-*Hs*Δ29GnTI and 10 mM UDP-GlcNAc (Sigma-Aldrich) in GnT buffer (20 mM HEPES, 50 mM NaCl, 10 mM $MnCl_2$, pH 7.2) at 37 °C for 16 h. To synthesize glycan **3**, glycan **2** was incubated with 80 μg/mL SIMPLEx-*Hs*Δ29GnTII and 20 mM UDP-GlcNAc in GnT buffer at 37 °C for 36 h. Glycan **3** was then incubated with 20 μg/mL SIMPLEx-*Hs*Δ44β4GalT1 and 10 mM UDP-Gal (Sigma-Aldrich) in GalT buffer (20 mM HEPES, 150 mM NaCl, 10 mM $MnCl_2$, pH 7.5) at 37 °C for 16 h to produce

glycan **4**. Sialic acid terminal glycans **5** and **6** were synthesized by incubating glycan **4** with 20 μg/mL SIMPLEx-*Hs*Δ26ST6Gal1 and 20 mM CMP-Neu5Ac (Sigma-Aldrich) in SiaT buffer (50 mM sodium phosphate, 150 mM NaCl, 10 mM $MgCl_2$, pH 8.0) at 37 °C for 16 h. Glycan **7** was synthesized by incubating glycan **4** with 20 μg/mL SIMPLEx-*Hs*Δ30FuT8 and 10 mM GDP-fucose (Sigma-Aldrich) in FUT buffer (100 mM MES, 10 mM $MgCl_2$, pH 7.0) at 37 °C for 16 h. Glycans **8**, **9**, and **10** were synthesized sequentially from glycan **7** using SIMPLEx-*Hs*Δ44β4GalT1 and SIMPLEx-*Hs*Δ26ST6Gal1 as described above for glycans **4**, **5**, and **6**. Reaction progress was monitored by MALDI-TOF MS. Reaction clean-up and glycan purification were performed as described previously (Hamilton et al., 2017). Because sialic acid is subject to MS induced in-source and metastable decay, successful biosynthesis of glycans **5**, **6**, **9**, and **10** was verified by nano LC-MS/MS analysis.

Bioenzymatic glycan remodeling on glycoproteins.

Unless noted otherwise, all glycoprotein remodeling reactions were performed at 37 °C for 1 h prior to bioorthogonal labeling reaction as described above. The sialytransferase activity of SIMPLEx-*Hs*ST6Gal1 and SIMPLEx-*Cj*CstII was assessed using human A1AT as glycoprotein acceptor substrate. A total of 3 μg of recombinant A1AT (R&D Systems) was treated with 20 U/μL α2-3,6,8,9 neuraminidase A (NEB) in a 10-μL reaction at 37 °C for 2 h to remove terminal sialic acid residues on A1AT glycans. Reaction mixtures were then heated at 85 °C for 15 min to inactivate

neuramidase A. Neuramidase A-treated A1AT was then incubated with SIMPLEx-*Hs*ST6Gal1 or SIMPLEx-*Cj*CstII and CMP-AzNec5Ac in SiaT buffer. Sialyltransferase activity of SIMPLEx-*Hs*ST6GalNAc1 was evaluated in a similar manner but neuramidase-treated bovine submaxillary glands mucin (Sigma-Aldrich) was used as the glycoprotein substrate. *N*-acetylglucosaminyltransferase activity of SIMPLEx-*Hs*GnTI was assessed using MBP-GCG^{DQNAT}, a fusion between *E. coli* MBP and human glucagon (residues 1-29) followed by a C-terminal DQNAT glycosylation tag (Glasscock et al., 2018). The MBP-GCG^{DQNAT} was glycosylated with $Man_3GlcNAc_2$ using glycoengineered *E. coli* as described previously (Glasscock et al., 2018). Purified MBP-GCG^{DQNAT} was incubated with SIMPLEx-*Hs*GnTI and UDP-GlcNAz in GnT buffer. Fucosyltransferase activity was evaluated by incubating A1AT or neuramidase A-treated A1AT with SIMPLEx-*Hs*FUT7 and GDP-AzFuc in FUT buffer.

Chromatography and mass spectrometry.

Hydrophilic interaction liquid chromatography (HILIC) was carried out using an Exion HPLC system with built-in autosampler (SCIEX). The free glycan samples were reconstituted in buffer A (80%: 20% acetonitrile: water), filtered with 0.22 μm spin filter (Corning) and loaded onto a Kinetex HILIC column (2.6 μm, 2.6x150 mm; Phenomenex) with 80% ACN/20% water as buffer A and 50 mM NH_4FA with pH 4.4 as buffer B. The LC was performed using a gradient from 80-0% of buffer B in 7 min at a flow rate 400 μL/min.

All LC-MS/MS analysis was carried out using an SCIEX X500B QTOF (SCIEX) mass spectrometer equipped with an electrospray ion source and coupled with an Exion HPLC system. Each reconstituted sample (2 μL) was injected onto a Kinetex HILIC column (2.6 μm, 2.6x 150 mm; Phenomenex). The free glycans were eluted in a 9-min gradient of 80% to 0% (80% ACN/20% water) at 400 nL/min., followed by a 3-min hold at 80% (80% ACN/20% water) for re-equilibration. The SCIEX X500 QTOF was operated in positive ion mode with ESI voltage set at 5.0 kV, ion source gas 1, gas 2 = 50 psi, curtain gas = 35 and CAD gas = 7 and source temperature of 350 °C. Calibration was done using positive calibrant with CDS system. The instrument was operated in MS full scan mode from *m*/*z* range from 200-2000 followed by multiple reaction monitoring high-resolution (MRM-HR) scan from 0-12 min at two different collision energies of 20 and 35 V with DP = 20 V and accumulation time of 0.25 s. MS survey scans for the mass range of *m*/*z* 200-2000 with DP = 20 V, CE = 7 V and accumulation time of 0.25 s and MS/MS MRM-HR scans at the same DP voltage and CE = 20 V and with Q1 unit resolution. All MS and MS/MS raw spectra from each sample obtained by MRM-HR scan were analyzed by SCIEX OS 1.4 data analysis system. XIC spectra were extracted from MS full scan with each MRM transition. The glycan structure was annotated manually using GlycanMass-ExPAsy tool.

Physicochemical data collection and analysis.

The name, amino acid sequence, structure availability (full-length or partial), and predicted post-translational modifications (*i.e.*, disulfide bonds and protein glycosylation) for each GT enzyme were retrieved from the Uniprot database (UniProt, 2019). GT family members were annotated from the CAZy database (Lombard et al., 2014). Amino acid sequences of full length, truncated, and SIMPLEx-fused GTs were compiled in FASTA format. The M_w and pI were calculated in average resolution setting using the ExPASy Bioinformatics resource portal (Wilkins et al., 1999). Solubility prediction score was calculated using CamSol Intrinsic version 2.1 (Sormanni et al., 2015). The expression scores for all constructs were annotated based on immunoblots in **Supplementary Figure 8-7a**. Correlation between protein properties (M_w, pI, solubility prediction score, and expression score) were analyzed using R software version 3.4.2. Specifically, scatter plots between protein properties, generated with data points colored according to expression score, were used to examine any possible correlations. For the correlation between expression score and pI, datasets were analyzed as a function of expression score movement. Scatter plots comparing the pI of SIMLPEx-GT versus GT^{6xHis} constructs were created and the data were colored according to the change in expression score. A similar approach was used to analyze the correlation between expression score and solubility prediction score. Because no statistical significance for the correlation between expression score and both pI and solubility prediction score was observed, general observations from the

plots were described instead. For the correlation between expression score and M_w, data were categorized into 3 groups: M_w < 40 kDa, M_w = 40-60 kDa, and M_w > 60 kDa, before weighted average of expression score for each group was calculated. Both SIMPLEx-GT and GT^{6xHis} constructs were binned using the same criteria since the added mass from SIMPLEx fusion was constant for all constructs. Welch's *t*-test was used to analyze statistical significance for categorical datasets.

Statistical analysis.

Unless stated otherwise, experiments were performed with at least three biological replicates and at least three technical measurements. All data were reported as average values with error bars representing standard error of the mean (s.e.m.). Statistical significance was determined by Welch's *t*-test and *p*-values of <0.05 were considered significant. All graphs were generated in Microsoft Excel or R software version 3.4.2.

Data availability.

All data generated in the study are included in this article and its supplementary information files. Additional information and materials are available from the corresponding author upon reasonable request.

4.6 Acknowledgements.

We would like to thank Dr. Bernard Henrissat for providing statistical data of the GT genes from the CAZy database. We thank Dr. Scott Emr for providing the yeast strain and corresponding expression vector. We thank Dr. Sudeep Banjade, Dr. May Taw, and Dr. Morgan Ludwicki for assistance and critical discussions regarding mammalian cell expression. This work was supported by the Bill and Melinda Gates Foundation (OPP1217652 to M.C.J. and M.P.D.), Defense Threat Reduction Agency (HDTRA1-15-10052 and HDTRA1-20-10004 to M.C.J. and M.P.D.), National Science Foundation (CBET-1159581, CBET-1264701, CBET-1936823 to M.P.D. and MCB 1413563 to M.C.J. and M.P.D.), and National Institutes of Health (1R01GM137314 and 1R01GM127578 to M.P.D.). The work was also supported by seed project funding (to M.P.D.) through the National Institutes of Health-funded Cornell Center on the Physics of Cancer Metabolism (supporting grant 1U54CA210184). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. T.J. was supported by a Royal Thai Government Fellowship and a Cornell Fleming Graduate Scholarship.

4.7 References.

Ardevol, A., and Rovira, C. (2015). Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. J Am Chem Soc *137*, 7528-7547.

Baker, J.L., Çelik, E., and DeLisa, M.P. (2013). Expanding the glycoengineering toolbox: the rise of bacterial N-linked protein glycosylation. Trends in Biotechnology *31*, 313323.

Barb, A.W., Brady, E.K., and Prestegard, J.H. (2009). Branch-specific sialylation of IgG-Fc glycans by ST6Gal-I. Biochemistry *48*, 9705-9707.

Brocchieri, L., and Karlin, S. (2005). Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Res *33*, 3390-3400.

Chen, C., and Colley, K.J. (2000). Minimal structural and glycosylation requirements for ST6Gal I activity and trafficking. Glycobiology *10*, 531-583.

Du, T., Buenbrazo, N., Kell, L., Rahmani, S., Sim, L., Withers, S.G., DeFrees, S., and Wakarchuk, W. (2018). A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans. Cell Chem Biol.

Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L., and McCafferty, J. (2004). Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. BMC Biotechnol *4*, 32.

Faridmoayer, A., Fentabil, M.A., Mills, D.C., Klassen, J.S., and Feldman, M.F. (2007). Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. J Bacteriol *189*, 8088-8098.

Garnham, R., Scott, E., Livermore, K.E., and Munkley, J. (2019). ST6GAL1: A key player in cancer. Oncol Lett *18*, 983-989.

Glasscock, C.J., Yates, L.E., Jaroentomeechai, T., Wilson, J.D., Merritt, J.H., Lucks, J.B., and DeLisa, M.P. (2018). A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation. Metab Eng *47*, 488-495.

Grinkova, Y.V., Denisov, I.G., and Sligar, S.G. (2010). Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. Protein Eng Des Sel *23*, 843-848.

Hamilton, B.S., Wilson, J.D., Shumakovich, M.A., Fisher, A.C., Brooks, J.C., Pontes, A., Naran, R., Heiss, C., Gao, C., Kardish, R., *et al.* (2017). A library of chemically defined human N-glycans synthesized from microbial oligosaccharide precursors. Sci Rep *7*, 15907.

Hebditch, M., Carballo-Amador, M.A., Charonis, S., Curtis, R., and Warwicker, J. (2017). Protein-Sol: a web tool for predicting protein solubility from sequence. Bioinformatics *33*, 3098-3100.

Jaroentomeechai, T., Stark, J.C., Natarajan, A., Glasscock, C.J., Yates, L.E., Hsu, K.J., Mrksich, M., Jewett, M.C., and DeLisa, M.P. (2018). Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. Nat Commun *9*, 2686.

Jewett, M.C., and Swartz, J.R. (2004). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnology and bioengineering *86*, 19-26.

Kightlinger, W., Duncker, K.E., Ramesh, A., Thames, A.H., Natarajan, A., Stark, J.C., Yang, A., Lin, L., Mrksich, M., DeLisa, M.P., *et al.* (2019). A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. Nat Commun *10*, 5404.

Kightlinger, W., Lin, L., Rosztoczy, M., Li, W., DeLisa, M.P., Mrksich, M., and Jewett, M.C. (2018). Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases. Nat Chem Biol.

Kuhn, B., Benz, J., Greif, M., Engel, A.M., Sobek, H., and Rudolph, M.G. (2013). The structure of human alpha-2,6-sialyltransferase reveals the binding mode of complex glycans. Acta Crystallogr D Biol Crystallogr *69*, 1826-1838.

Kwon, Y.-C., and Jewett, M.C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. Scientific Reports *5*, 8663.

Lairson, L.L., Henrissat, B., Davies, G.J., and Withers, S.G. (2008). Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem *77*, 521-555.

Lauber, J., Handrick, R., Leptihn, S., Durre, P., and Gaisser, S. (2015). Expression of the functional recombinant human glycosyltransferase GalNAcT2 in Escherichia coli. Microb Cell Fact *14*, 3.

Li, C., and Wang, L.X. (2018). Chemoenzymatic Methods for the Synthesis of Glycoproteins. Chem Rev *118*, 8359-8413.

Li, W., McArthur, J.B., and Chen, X. (2019). Strategies for chemoenzymatic synthesis of carbohydrates. Carbohydr Res *472*, 86-97.

Lithgow, K.V., Scott, N.E., Iwashkiw, J.A., Thomson, E.L., Foster, L.J., Feldman, M.F., and Dennis, J.J. (2014). A general protein O-glycosylation system within the Burkholderia cepacia complex is involved in motility and virulence. Mol Microbiol *92*, 116-137.

Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P., and Berkmen, M. (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. Microb Cell Fact *11*, 56.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res *42*, D490-495.

Luirink, J., and Sinning, I. (2004). SRP-mediated protein targeting: structure and function revisited. Biochim Biophys Acta *1694*, 17-35.

Marston, F.A. (1986). The purification of eukaryotic polypeptides synthesized in Escherichia coli. Biochem J *240*, 1-12.

Meng, L., Forouhar, F., Thieker, D., Gao, Z., Ramiah, A., Moniz, H., Xiang, Y., Seetharaman, J., Milaninia, S., Su, M., *et al.* (2013). Enzymatic basis for N-glycan sialylation: structure of rat alpha2,6-sialyltransferase (ST6GAL1) reveals conserved and unique features for glycan sialylation. J Biol Chem *288*, 34680-34698.

Mikolajczyk, K., Kaczmarek, R., and Czerwinski, M. (2020). How glycosylation affects glycosylation: the role of N-glycans in glycosyltransferase activity. Glycobiology *30*, 941-969.

Mizrachi, D., Chen, Y., Liu, J., Peng, H.M., Ke, A., Pollack, L., Turner, R.J., Auchus, R.J., and DeLisa, M.P. (2015). Making water-soluble integral membrane proteins in vivo using an amphipathic protein fusion strategy. Nat Commun *6*, 6826.

Mizrachi, D., Robinson, M.P., Ren, G., Ke, N., Berkmen, M., and DeLisa, M.P. (2017). A water-soluble DsbB variant that catalyzes disulfide-bond formation in vivo. Nat Chem Biol *13*, 1022-1028.

Moremen, K.W., Ramiah, A., Stuart, M., Steel, J., Meng, L., Forouhar, F., Moniz, H.A., Gahlay, G., Gao, Z., Chapla, D., *et al.* (2018). Expression system for structural and functional studies of human glycosylation enzymes. Nat Chem Biol *14*, 156-162.

Naegeli, A., and Aebi, M. (2015). Current Approaches to Engineering N-Linked Protein Glycosylation in Bacteria. Glyco-Engineering: Methods and Protocols *1321*, 3-16.

Naegeli, A., Neupert, C., Fan, Y.Y., Lin, C.W., Poljak, K., Papini, A.M., Schwarz, F., and Aebi, M. (2014). Molecular analysis of an alternative N-glycosylation machinery by

functional transfer from Actinobacillus pleuropneumoniae to Escherichia coli. J Biol Chem *289*, 2170-2179.

Natarajan, A., Jaroentomeechai, T., Cabrera-Sanchez, M., Mohammed, J.C., Cox, E.C., Young, O., Shajahan, A., Vilkhovoy, M., Vadhin, S., Varner, J.D., *et al.* (2020). Engineering orthogonal human O-linked glycoprotein biosynthesis in bacteria. Nat Chem Biol *16*, 1062-1070.

Netzer, W.J., and Hartl, F.U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature *388*, 343-349.

Niwa, T., Ying, B.W., Saito, K., Jin, W., Takada, S., Ueda, T., and Taguchi, H. (2009). Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. Proc Natl Acad Sci U S A *106*, 4201-4206.

Ortiz-Soto, M.E., and Seibel, J. (2016a). Expression of functional human sialyltransferases ST3Gal1 and ST6Gal1 in Escherichia coli. PLoS One *11*, e0155410.

Ortiz-Soto, M.E., and Seibel, J. (2016b). Expression of Functional Human Sialyltransferases ST3Gal1 and ST6Gal1 in Escherichia coli. Plos One *11*.

Randall, L.L., and Hardy, S.J. (2002). SecB, one small chaperone in the complex milieu of the cell. Cell Mol Life Sci *59*, 1617-1623.

Shimma, Y., Saito, F., Oosawa, F., and Jigami, Y. (2006). Construction of a library of human glycosyltransferases immobilized in the cell wall of Saccharomyces cerevisiae. Appl Environ Microbiol *72*, 7003-7012.

Skretas, G., Carroll, S., DeFrees, S., Schwartz, M.F., Johnson, K.F., and Georgiou, G. (2009). Expression of active human sialyltransferase ST6GalNAcI in Escherichia coli. Microb Cell Fact *8*, 50.

Sormanni, P., Aprile, F.A., and Vendruscolo, M. (2015). The CamSol method of rational design of protein mutants with enhanced solubility. J Mol Biol *427*, 478-490.

Stark, J.C., Jaroentomeechai, T., Moeller, T.D., Hershewe, J.M., Warfel, K.F., Moricz, B.S., Martini, A.M., Dubner, R.S., Hsu, K.J., Stevenson, T.C., *et al.* (2021). On-demand biomanufacturing of protective conjugate vaccines. Sci Adv *7*.

Tayi, V.S., and Butler, M. (2018). Solid-Phase Enzymatic Remodeling Produces High Yields of Single Glycoform Antibodies. Biotechnol J *13*, e1700381.

Tytgat, H.L.P., Lin, C.W., Levasseur, M.D., Tomek, M.B., Rutschmann, C., Mock, J., Liebscher, N., Terasaka, N., Azuma, Y., Wetter, M., *et al.* (2019). Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. Nat Commun *10*, 5403.

UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res *47*, D506-D515.

Valderrama-Rincon, J.D., Fisher, A.C., Merritt, J.H., Fan, Y.Y., Reading, C.A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., and DeLisa, M.P. (2012). An engineered eukaryotic protein glycosylation pathway in Escherichia coli. Nat Chem Biol *8*, 434-436.

Vozza, N.F., and Feldman, M.F. (2015). Glyco-engineering O-Antigen-Based Vaccines and Diagnostics in E. coli. springer *1321*, 57-70.

Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., *et al.* (2002). N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. Science *298*, 1790-1793.

Wagner, S., Bader, M.L., Drew, D., and de Gier, J.W. (2006). Rationalizing membrane protein overexpression. Trends Biotechnol *24*, 364-371.

Wang, L.X., and Lomino, J.V. (2012). Emerging technologies for making glycan-defined glycoproteins. ACS Chem Biol *7*, 110-122.

Wang, L.X., Tong, X., Li, C., Giddens, J.P., and Li, T. (2019). Glycoengineering of antibodies for modulating functions. Annu Rev Biochem *88*, 433-459.

Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. Methods Mol Biol *112*, 531-552.

CHAPTER 5

CONCLUSIONS

In this dissertation, we developed cell-free tools and platforms for expression and characterization of both bacterial and mammalian glycoenzymes, for rapid prototyping of novel protein glycosylation pathways, and for on-demand biomanufacturing of useful glycoproteins and glycoconjugate vaccines. This work resulted in the significant progress in the field of synthetic glycobiology and led to a creation of the whole new subfield concerning cell-free glycoprotein synthesis.

In Chapter 2, by combining our knowledge in bacterial glycoengineering and cell-free biology, we made the first important progress toward fully-integrated cell-free synthetic glycobiology (CFSG) platform by creating *E. coli* lysates capable of catalyzing *N*-glycoproteins biosynthesis in a single pot cell-free reaction. Subsequent effort has equipped our *E. coli* lysates with the ability to prepare *O*-glycoproteins carrying authentic human *O*-glycans, although this is beyond the scope of the current dissertation. In Chapter 3, we further engineered these glycocompetent *E. coli* lysates to furnish glycoconjugate vaccines against various gram-negative bacterial pathogens. We employed genome engineering tool and lysate lyophilization method to broaden the applications of our glycosylation lysates toward on-demand biomanufacturing of vaccines at the point-of-care. Together, the works described in Chapter 2-3 represent successful creation and application of the first module in our cell-free synthetic glycobiology system, as outline in the section 1.4.

In Chapter 4, we described the first step in engineering cell-free glycan elaboration, the second module in the CFSG pipeline. Glycosyltransferases (GTs) are an important catalyst for accessing diverse glycan and glycosylated molecules, yet this class of enzymes is known to be challenging to recombinantly express. Realizing this bottleneck, we adopt a novel approach for improving expression and solubility of GT enzymes. By leveraging protein engineering strategy called solubilization of the integral membrane protein with high level of expression (SIMPLEx), we successfully expressed over 100 GTs from diverse organisms and with distinct functions. A collection of GTs generated here is an important resource to improve our understanding of these GTs as well as to access customized complex glycan structures.

Glycoscience has an immense impact in basic biology and applied biotechnology. As such, new and expanded toolkits are required to help transform the field of glycoscience and realize its full potential across biology, chemistry, and material science. This dissertation represents a small, yet important, effort to push the boundary of glycoscience with a power of cell-free synthetic biology. By bridging these two fields, we begin to systematically study natural glycosylation pathways in a well-defined environment within cell-free reactions. In addition, we are able to use our cell-free platforms to bottom-up engineer synthetic glycosylation system. Beyond the works in this dissertation, we look forwards to further engineering our CFSG system by: i) integrating tools in cell-free biology including new energy recharging system and membrane vesicle retention techniques to improve protein synthesis yield and

glycosylation efficiency; ii) adopting computational, statistical analysis, and machine-learning approach to design-build-test new CFSG platforms with improved functionalities; iii) developing microfluidic- and robotic-based cell-free glycosylation system which would allow spatial and temporal arrangements of the glycosylation reactions, akin to the natural processes within the ER and Golgi organelles; and iv) coupling the existing glycoprotein synthesis reactions with other metabolic cascades including nucleotide-activated sugar and lipid-linked oligosaccharides biosynthesis. These improvements are anticipated to unlock a full potential of cell-free synthetic glycobiology system and further our advancement in glycoscience and its myriad applications.

CHAPTER 6

APPENDIX A[13]

Supplementary information: *Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery*

Supplemental Table 6-1. Cost analysis of CFGpS reactions.

Component	Cost ($/μL rxn)	Supplier	Product No
$Mg(Glu)_2$	negligible	Sigma	49605
NH_4Glu	negligible	MP	02180595
KGlu	negligible	Sigma	G1501
ATP	negligible	Sigma	A2383
GTP	0.000265	Sigma	G8877
UTP	0.000230	Sigma	U6625
CTP	0.000200	Sigma	C1506
Folinic acid	0.0000206	Sigma	47612
tRNA	0.000215	Roche	10109541001
Amino acids	negligible	homemade	
PEP	0.00179	Roche	10108294001
NAD	negligible	Sigma	N8535-15VL
CoA	0.000336	Sigma	C3144
Oxalic acid	negligible	Sigma	P0963
Putrescine	negligible	Sigma	P5780

13 This chapter appears in the Nature Communication journal:

Jaroentomeechai, T.*, Stark, J.C.*, Natarajan, A., Glasscock, C.J., Yates, L.E., Hsu, K.J., Mrksich, M., Jewett, M.C. and DeLisa, M.P. (2018) Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat Commun* 9: 2686.

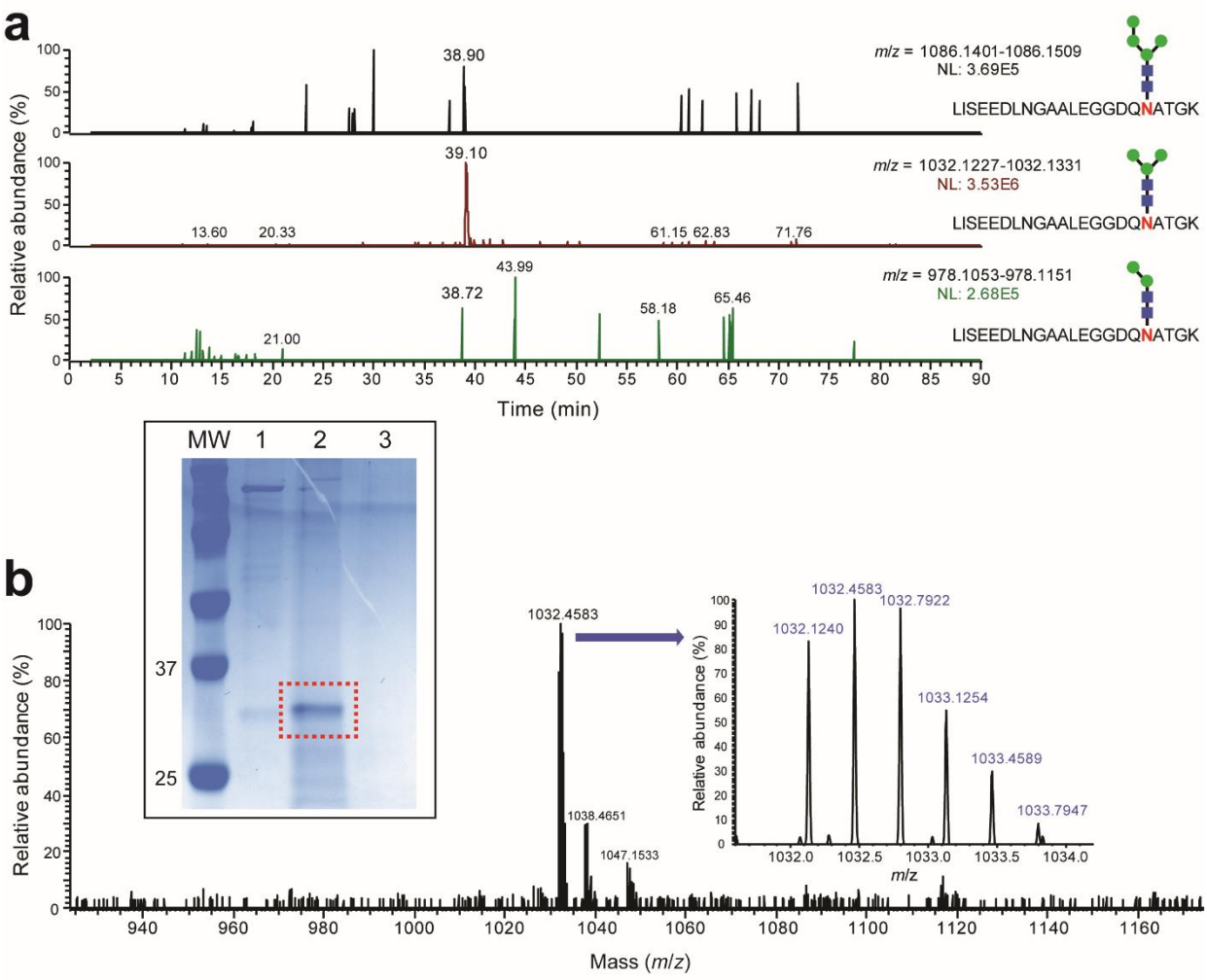
Spermidine	negligible	Sigma	S2626
HEPES	negligible	Sigma	H3375
$MnCl_2$	negligible	Sigma	63535
DDM	0.000358	Anatrace	D310S
Plasmid	negligible	user-supplied	
Extract	0.00737	homemade	
Total	**0.0108**	**$/μL rxn**	

The total cost to assemble CFGpS reactions is ~$0.01 per μL. In the table, amino acid cost accounts for 2 mM each of the 20 canonical amino acids purchased individually from Sigma. Extract cost is based on a single employee making 50 mL lysate from a 10 L fermentation, assuming 30 extract batches per year and a 5-year equipment lifetime. Component source is included in the table if it is available to purchase directly from a supplier. Homemade or user-supplied components cannot be purchased directly and must be prepared by the end user according to procedures described in the Methods section.

Supplementary Table 6-2. Plasmids used in the CFGpS study.

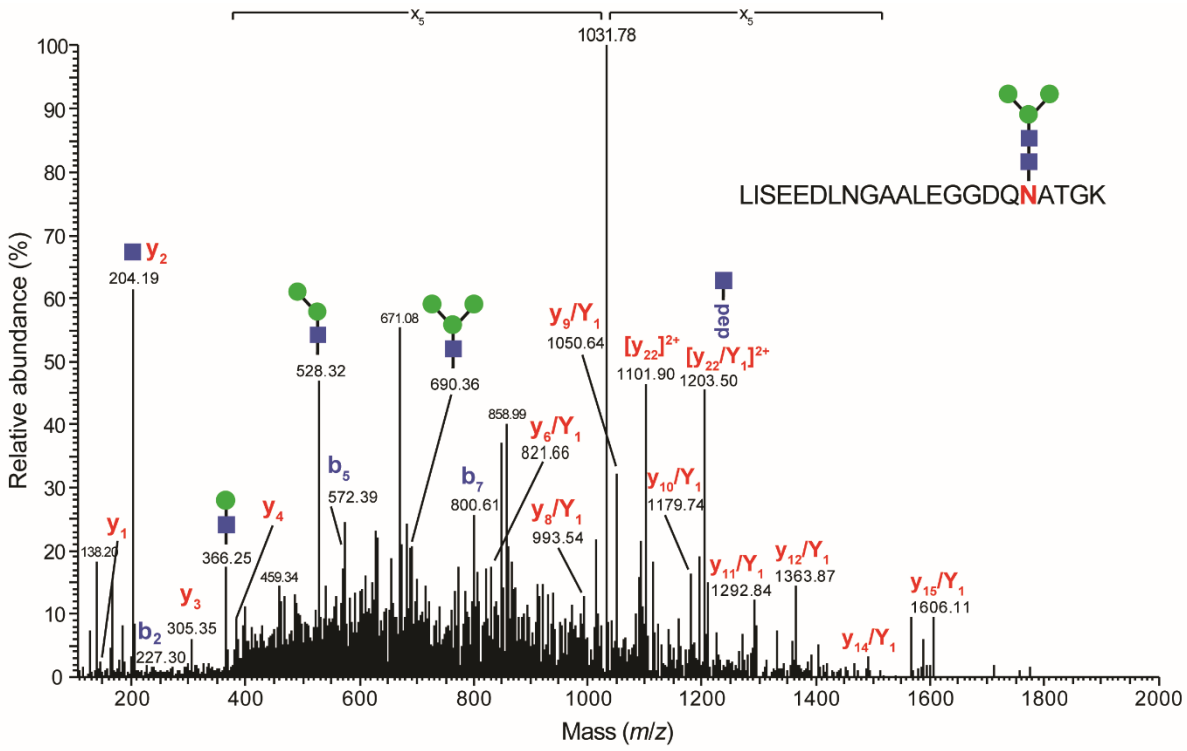
Plasmids	Description	Reference
pSN18	modified pBAD expression plasmid encoding *C. jejuni pglB* with a C-terminal decahistidine affinity tag.	[(Kowarik et al., 2006a)]
pET28a-scFv13-R4 (N34L, N77L)^{DQNAT}	pET28a(+) plasmid encoding scFv13-R4 modified with a C-terminal DQNAT glycosylation tag and two mutations (N34, N77L) to eliminate putative internal glycosylation sites.	[(Ollis et al., 2015c)]
pMW07-pglΔB	pMW07 plasmid encoding *C. jejuni* protein glycosylation locus (*pgl*) with complete in-frame deletion of *Cj*PglB	[(Ollis et al., 2014a)]
pACYC*pgl2*	pACYC plasmid encoding modified *C.lari* hexasaccharide glycan biosynthesis gene cluster lacking bacillosamine biosynthesis genes	[(Schwarz et al., 2010)]
pACYC*pgl4*	pACYC plasmid encoding native *C.lari* hexasaccharide hexasaccharide glycan biosynthesis gene cluster	[(Schwarz et al., 2011b)]
pEpiFOS-5*pgl5*	pEpiFOS-5 encoding the *Wolinella succinogenes* hexasaccharide glycan biosynthesis gene cluster cloned in the Eco72 site	Lab stock
pConYCG-mCB	pMW07 plasmid encoding $Man_3GlcNAc_2$ glycan biosynthesis genes and *manCB* genes for GDP-mannose biosynthesis	Lab stock
pJL1-scFv13-R4^{DQNAT}	pJL1 plasmid encoding scFv13-R4 (N34L, N77L)^{DQNAT}	This study
pJL1-sfGFP$^{217\text{-}DQNAT}$	pJL1 plasmid encoding superfolder GFP modified after residue T216 with 21 amino acid insertion containing	This study

	the *C. jejuni* AcrA N123 glycosylation site but with an optimal DQNAT sequon	
pJL1-sfGFP$^{217\text{-AQNAT}}$	same as pJL1-sfGFP^{DQNAT} but with AQNAT sequon	This study
pJL1-hEPO$^{22\text{-DQNAT-26}}$	pJL1 plasmid encoding human erythropoietin with native glycosylation motif surrounding N22 mutated to DQNAT	This study
pJL1-hEPO$^{36\text{-DQNAT-40}}$	pJL1 plasmid encoding human erythropoietin with native glycosylation motif surrounding N38 mutated to DQNAT	This study
pJL1-hEPO$^{81\text{-DQNAT-85}}$	pJL1 plasmid encoding human erythropoietin with native glycosylation motif surrounding N83 mutated to DQNAT	This study
pSF-*Cj*PglB	pSN18 derivative encoding *C. jejuni* PglB with C-terminal FLAG epitope tag	[(Ollis et al., 2014a)]
pSF-*Cc*PglB	pSN18-derivative encoding *C. coli* PglB with C-terminal FLAG epitope tag	[(Ollis et al., 2014a)]
pSF-*Dd*PglB	pSN18-derivative encoding *D. desulfuricans* PglB with C-terminal FLAG epitope tag	[(Ollis et al., 2014a)]
pSF-*Dg*PglB	pSN18-derivative encoding *D. gigas* PglB with C-terminal FLAG epitope tag	[(Ollis et al., 2014a)]
pSF-*Dv*PglB	pSN18-derivative encoding *D. vulgaris* PglB with C-terminal FLAG epitope tag	[(Ollis et al., 2014a)]

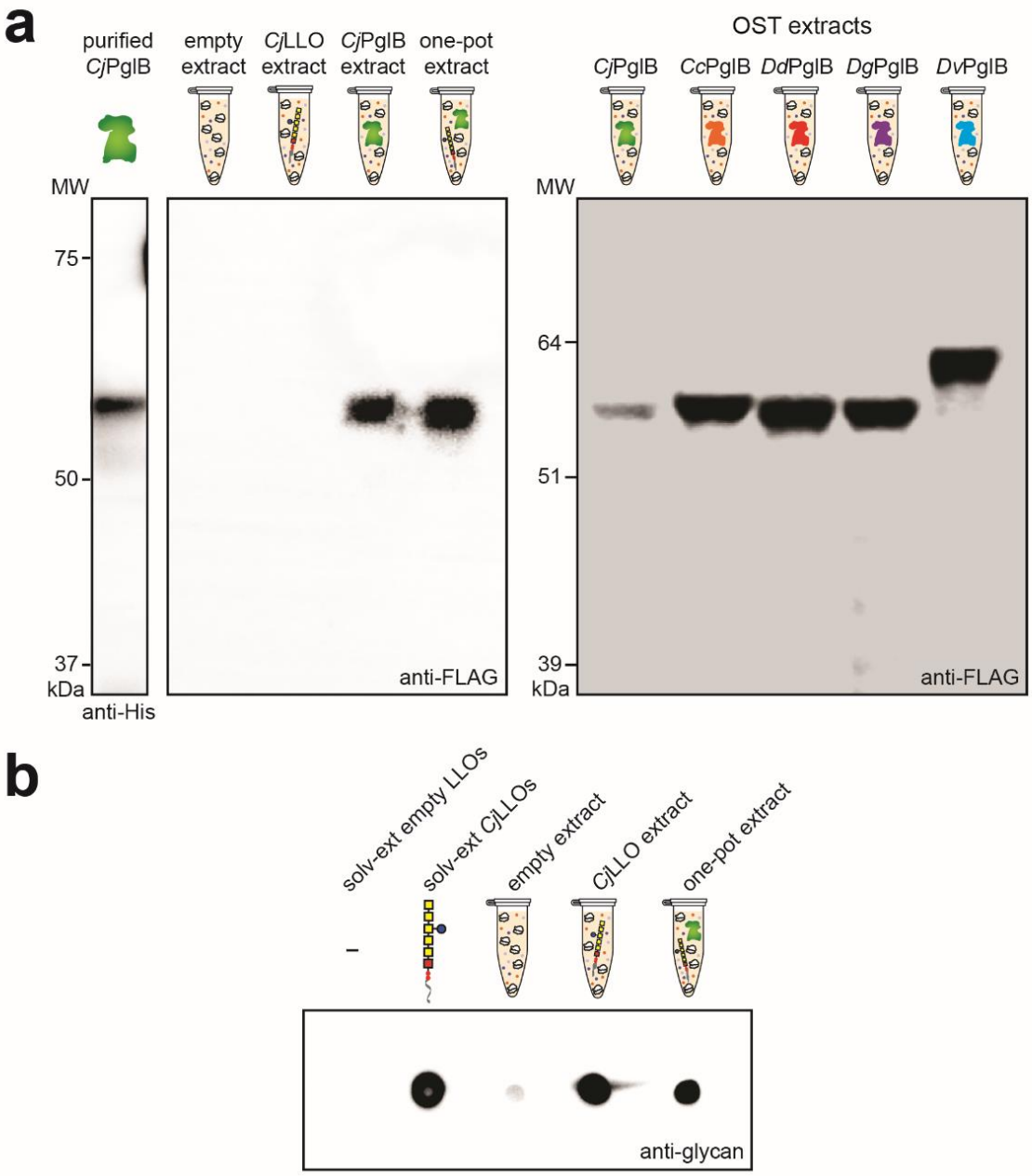


Supplementary Figure 6-1. MS analysis of scFv13-R4^{DQNAT} glycosylated with Man$_3$GlcNAc$_2$. Ni-NTA-purified scFv13-R4^{DQNAT} was subjected to *in vitro* glycosylation in the presence of purified *Cj*PglB and organic solvent-extracted Man$_3$GlcNAc$_2$ LLOs, and then directly loaded into an SDS-PAGE gel. Following staining of gel with Coomassie Brilliant Blue G-250 (inset), the glycosylated band (lane 2, indicated by red box) was excised and submitted for MS analysis. Controls included in vitro glycosylation reaction performed with solvent-extracted empty LLOs (lane 1) and complete *in vitro* glycosylation reaction mixture lacking purified scFv13-R4^{DQNAT} acceptor protein (lane 3). Molecular weight (MW) ladder loaded on the left. **(a)** Three

extracted ion chromatograms (XIC) corresponding to mass ranges for three possible glycopeptide products having masses consistent with the expected $Man_3GlcNAc_2$ (middle), as well as $Man_4GlcNAc_2$ (top) and $Man_2GlcNAc_2$ (bottom) attached to N273 site of scFv13- $R4^{DQNAT}$ (mass tolerance at 5 ppm). The individually normalized level (NL) for each glycoform indicates that only a $Hex_3HexNAc_2$ glycoform, which eluted at 39.10 min with NL of 3.53E6, was decently detected in the sample (middle). A trace amount of a $Hex_4HexNAc_2$ glycoform form eluted at 38.9 min with NL of 2.96E5 (top), but no $Hex_2HexNAc_2$ glycoform was detected. **(b)** MS spectrum of the detected glycopeptide containing an N-linked pentasaccharide consistent with $Man_3GlcNAc_2$ at m/z = 1032.4583. The MS inset shows an expanded view of the glycopeptide ion with triple charge.

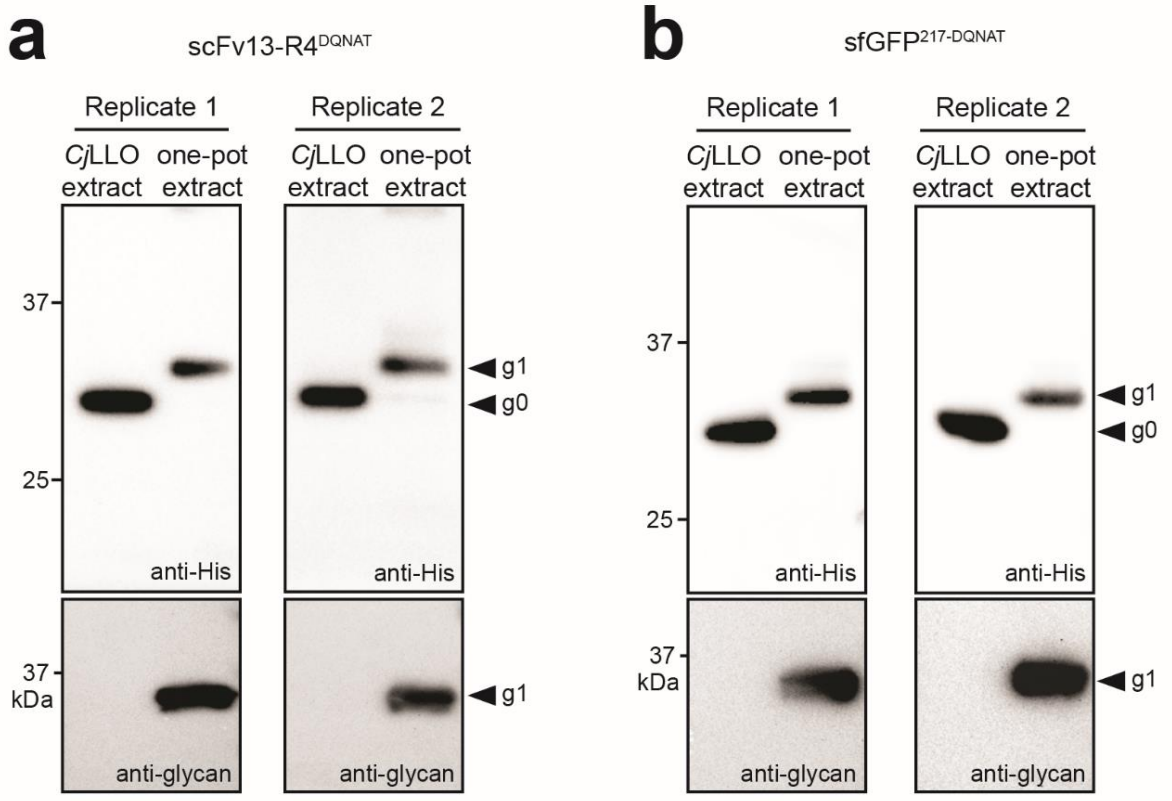


Supplementary Figure 6-2. Tandem mass spectrometry of scFv13-R4^{DQNAT} glycosylated with Man$_3$GlcNAc$_2$. MS/MS spectrum of the triply-charged precursor (m/z 1032.12), identifying the glycopeptide with core pentasaccharide (Hex$_3$HexNAc$_2$) attached to residue N273 (shown in red) in scFv13-R4^{DQNAT}. A series of y-ions covering from y1 to y4 and a second series of yions with the added mass of 203.08 Da at N273 site were found covering from y6/Y1 to y15/Y1, leading to the confident identification of tryptic peptide 256-LISEEDLNGAALEGGDQNATGK-277 and providing direct evidence for HexNAc as the innermost monosaccharide (Y1) attached to the N273 site. This result is also consistent with the previous observation that a relatively tight bond exists for the Y1-peptide compared to the fragile internal glycan bonds.

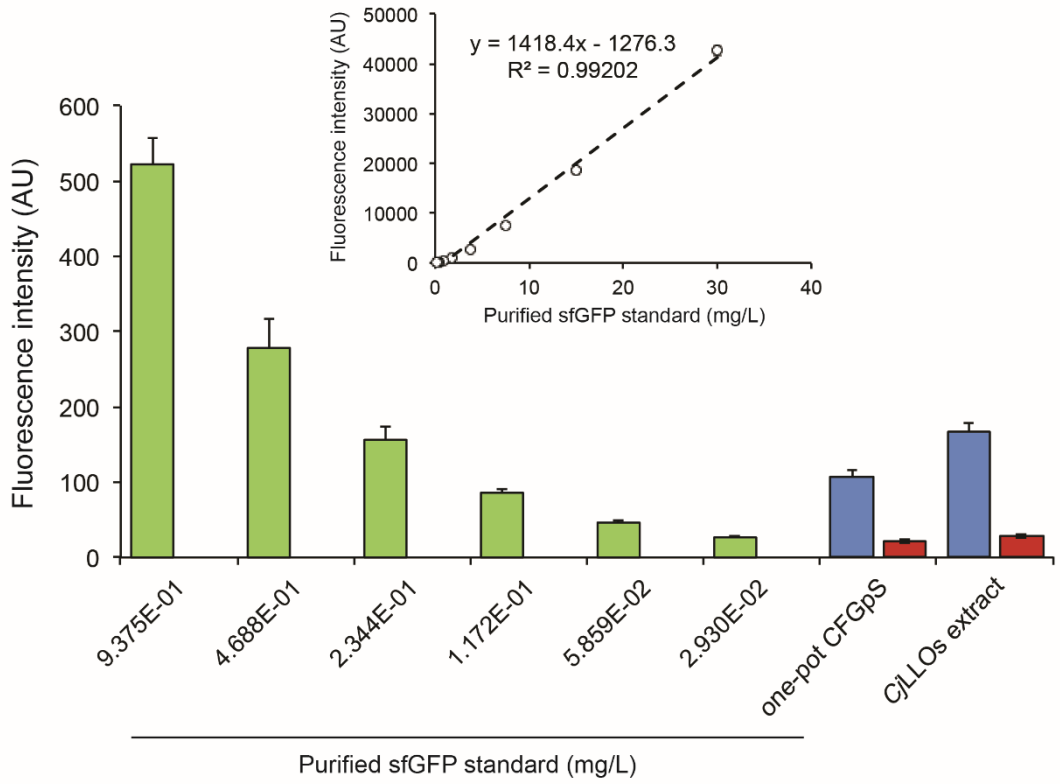


Supplementary Figure 6-3. Crude cell extracts are enriched with glycosylation machinery. (a) Western blot analysis of *Cj*PglB in the following samples: (left-hand panel) 1 μg of purified *Cj*PglB; (center panel) crude cell extracts derived from CLM24 cells with no plasmid (empty extract), CLM24 cells carrying pMW07-pglΔB (*Cj*LLO

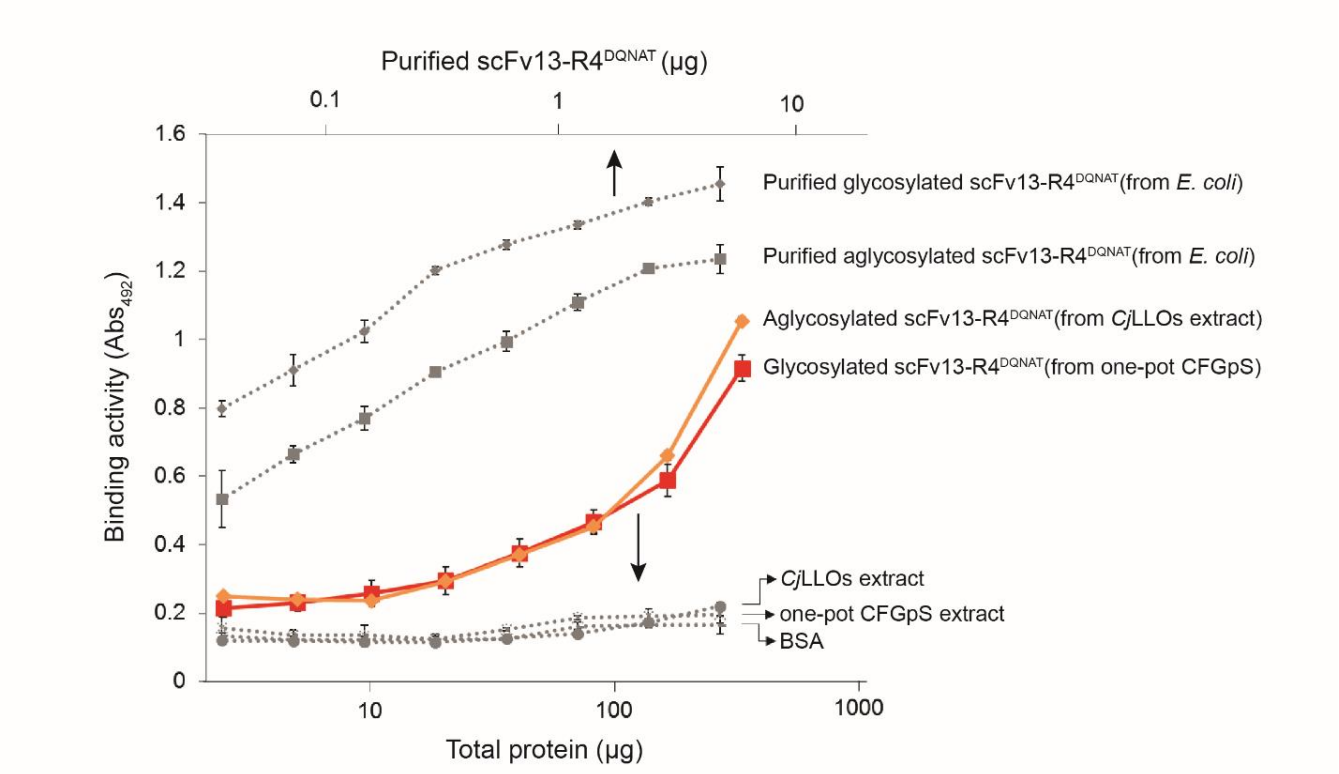
extract), CLM24 cells carrying pSF-*Cj*PglB (*Cj*PglB extract) or CLM24 cells carrying pMW07-pglΔB and pSF-*Cj*PglB (one-pot extract); and (right-hand panel) crude cell extracts derived from CLM24 cells carrying pSF-based plasmids encoding different PglB homologs as indicated. Blots were probed with anti-His antibody and anti-FLAG antibody as indicated. Molecular weight (MW) markers are indicated at left. Results are representative of at least three biological replicates. **(b)** Dot blot analysis of LLOs in the following samples: organic solvent extract from membrane fractions of CLM24 cells with no plasmid (solv-ext empty LLOs) or from CLM24 cells carrying plasmid pMW07-pglΔB (solv-ext *Cj*LLOs); crude cell extracts derived from CLM24 cells with no plasmid (empty extract), CLM24 cells carrying pMW07-pglΔB (*Cj*LLO extract) or CLM24 cells carrying pMW07-pglΔB and pSF-*Cj*PglB (one-pot extract). 10 µl of extracted LLOs or crude cell extract was spotted onto nitrocellulose membrane and probed with hR6 serum (anti-glycan).



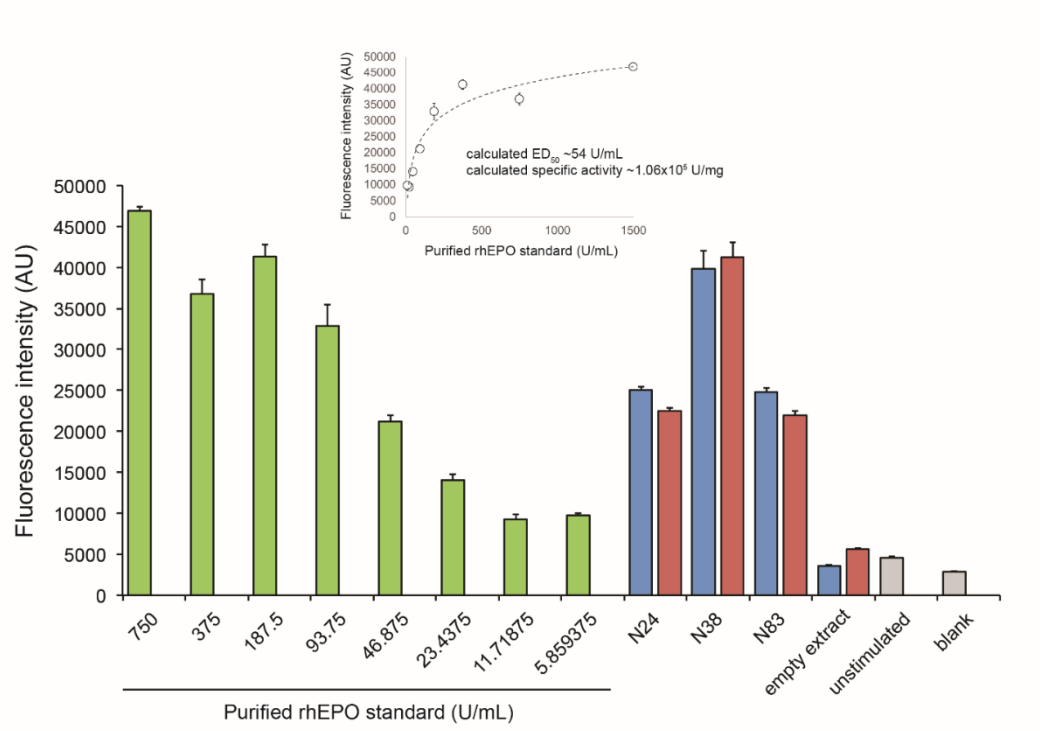
Supplementary Figure 6-4. **Independent biological replicates for one-pot CFGpS reactions.** Western blot analysis replicated twice for both the **(a)** scFv13-R4^{DQNAT} and (b) sfGFP$^{217\text{-}DQNAT}$ acceptor proteins produced using crude CLM24 extract selectively enriched with (i) *Cj*PglB from heterologous overexpression from pSF-*Cj*PglB and (ii) *Cj*LLOs from heterologous overexpression from pMW07-pgl B. Each replicate experiment involved charging freshly prepared cell-free extracts with freshly purified pJL1- scFv13-R4^{DQNAT} or pJL1-sfGFP$^{217\text{-}DQNAT}$ plasmid DNA. Control reactions (lane 1 in each panel) were performed using *Cj*LLO enriched extracts that lacked *Cj*PglB. Blots were probed with anti-hexa-histidine antibody (anti-His) to detect acceptor proteins or hR6 serum (anti-glycan) to detect the N-glycan. Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of the protein targets. Molecular weight (MW) markers are indicated at left



Supplementary Figure 6-5. CFGpS expression of active sfGFP. In-lysate fluorescence activity for glycosylated (one-pot CFGpS) and aglycosylated (*Cj*LLOs extract) sfGFP$^{217\text{-}DQNAT}$ produced in cell-free reactions charged with plasmid pJL1- sfGFP$^{217\text{-}DQNAT}$ (blue) or with no plasmid DNA (red). Following 2-h reactions, cell-free reactions containing glycosylated and aglycosylated sfGFP$^{217\text{-}DQNAT}$ were diluted 10 times with water and then subjected to fluorescence measurement. Excitation and emission wavelengths for sfGFP were 485 and 528 nm, respectively. Calibration curve was prepared by measuring fluorescence intensity of aglycosylated sfGFP$^{217\text{-}DQNAT}$ expressed and purified from *E. coli* cells and mixed with empty extract. Linear regression analysis (inset) was used to calculate the concentration of glycosylated sfGFP$^{217\text{-}DQNAT}$ in the samples, which was determined to be ~10 mg L^{-1} . Data are the average of three biological replicates and error bars represent the standard deviation of the mean.



Supplementary Figure 6-6. CFGpS expression of active scFv antibody fragment. Antigen-binding activity for -gal-specific scFv13-R4^{DQNAT} measured by ELISA with *E. coli*-gal as immobilized antigen. The scFv13-R4^{DQNAT} acceptor was produced as a glycosylated protein in one-pot CFGpS (red) or an aglycosylated protein in control extracts containing *Cj*LLOs but not *Cj*PglB (orange). Extracts were primed with plasmid pJL1- scFv13-R4^{DQNAT}. Positive controls included the same scFv13-R4^{DQNAT} protein produced in vivo by recombinant expression in *E. coli* in the presence (glycosylated) or absence (aglycosylated) of glycosylation machinery. Negative controls included extracts without plasmid and BSA. Comparing to signals from purified protein, the concentration of glycosylated scFv13-R4^{DQNAT} was determined to be ~20 mg L^{-1}. Data are the average of three biological replicates and error bars represent the standard deviation of the mean.



Supplementary Figure 6-7. CFGpS-derived hEPO glycovariants stimulate cell proliferation. Stimulation of human erythroleukemia TF-1 cell proliferation following incubation with purified rhEPO standard or hEPO variants produced in cell-free reactions. For CFGpS-derived hEPO glycovariants, TF-1 cells were treated with either glycosylated hEPO variants produced in one-pot CFGpS (blue) or aglycosylated hEPO variants produced in control extracts containing *Cj*LLOs but not *Cj*PglB (red). To produce the hEPO variants, extracts were primed with plasmid pJL1-hEPO$^{22\text{-DQNAT-}26}$ (N24), pJL1-hEPO$^{36\text{-DQNAT-}40}$ (N38), or pJL1- hEPO$^{81\text{-DQNAT-}85}$ (N83). For positive control rhEPO samples, cells were treated with serial dilutions of commercial rhEPO that was purified from CHO cells and thus glycosylated (green). TF-1 cells incubated with empty extracts or PBS (unstimulated) served as negative controls while RPMI media without cells was used as the blank. Regression analysis (inset) was performed to determine the concentration of hEPO variants in the samples, which was found to be at ~10 mg L^{-1}. Data are the average of three biological replicates and error bars represent the standard deviation of the mean.

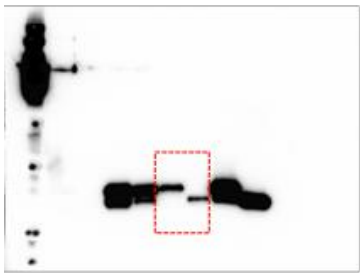


Figure 2a (left panel, anti-His blot, flipped)

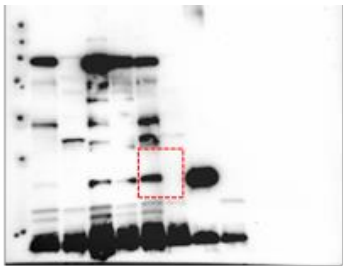


Figure 2a (left panel, anti-Glycan blot, flipped)

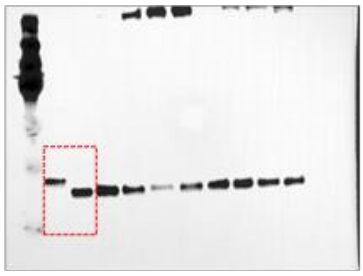


Figure 2a (right panel, anti-His blot, flipped)

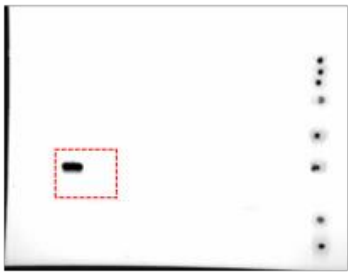


Figure 2a (right panel, anti-Glycan blot, flipped)

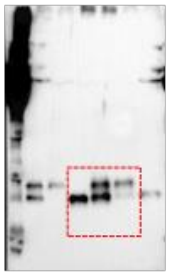


Figure 2b (left panel, anti-His blot)

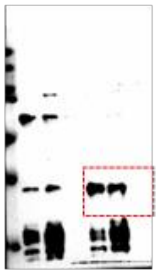


Figure 2b (left panel, anti-Glycan blot)

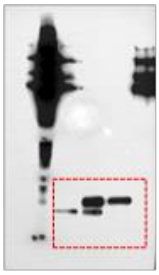


Figure 2b
(right panel, anti-His blot)

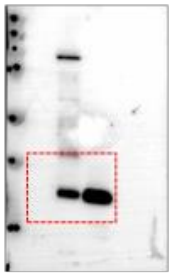


Figure 2b
(right panel, anti-Glycan blot)

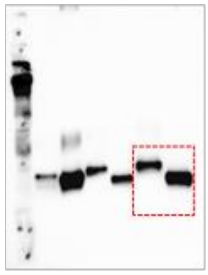


Figure 3a
(anti-His blot, flipped)

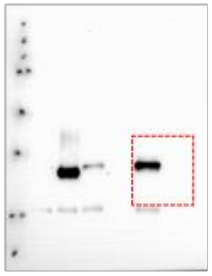


Figure 3a
(anti-Glycan blot, flipped)

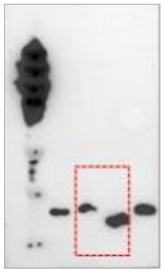


Figure 3 b
(anti-His blot, flipped)

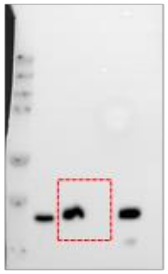


Figure 3b
(anti-Glycan blot, flipped)

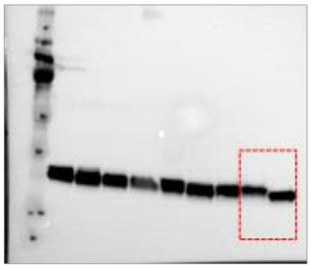


Figure 3c
(anti-His blot, flipped)

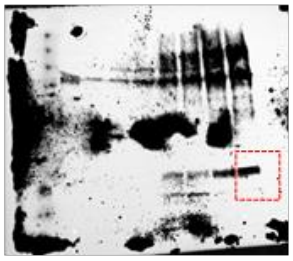


Figure 3c
(ConA blot, flipped)

Supplementary Figure 6-8. Full images of immunoblot presented in the main text. Cropped area is indicated by red box. Some blots are cropped and then flipped to make data presentation consistent throughout the manuscript. When this happened, blot legend is marked as "flipped".

Figure 3d
(anti-His blot, flipped)

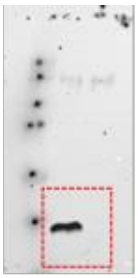


Figure 3d
(ConA blot, flipped)

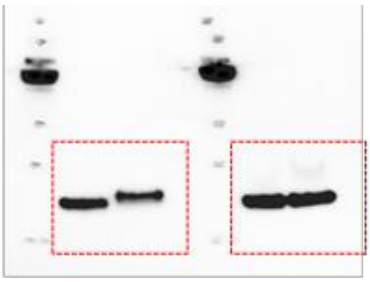


Figure 3e
(anti-His blot)

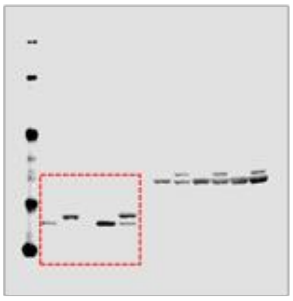


Figure 4a
(anti-His blot)

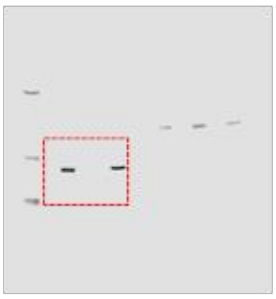


Figure 4a
(anti-Glycan blot)

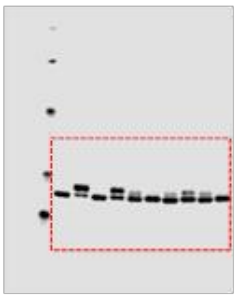


Figure 4b
(anti-His blot)

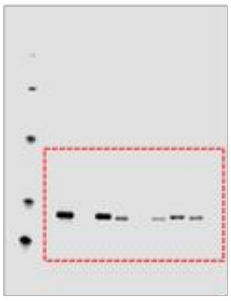


Figure 4b
(anti-Glycan blot)

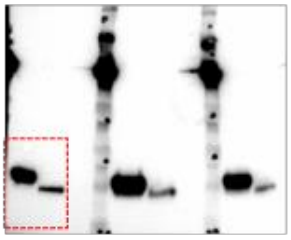


Figure 5a
(anti-His blot for scFv13 R4, flipped)

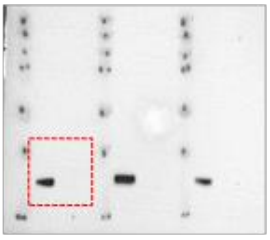


Figure 5a
(anti-Glycan blot for scFv13 R4, flipped)

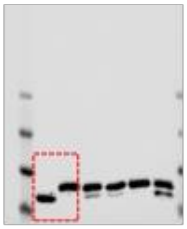


Figure 5a
(anti-His blot for sfGFP)

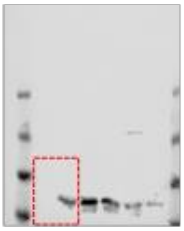


Figure 5a
(anti-Glycan blot for sfGFP)

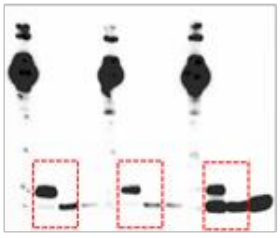


Figure 5b
(anti-His blot, flipped)

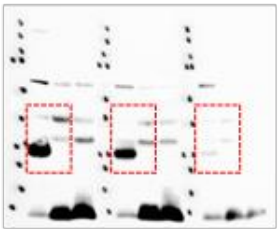


Figure 5b
(anti-Glycan blot, flipped)

Supplementary Figure 8 (cont'd). Full images of immunoblot presented in the main text. Cropped area is indicated by red box. Some blots are cropped and then flipped to make data presentation consistent throughout the manuscript. When this happened, blot legend is marked as "flipped".

CHAPTER 7

APPENDIX B[14]

Supplementary information: *On-demand, cell-free biomanufacturing of protective conjugate vaccines at the point-of-care*

Supplementary Table 7-1. Cost analysis for iVAX reactions. The total cost to assemble iVAX reactions is calculated below. A 1 mL iVAX reaction produces two 10 μg vaccine doses and can be assembled for $11.75. In the table, amino acid cost accounts for 2 mM each of the 20 canonical amino acids purchased individually from Sigma. Lysate cost is based on a single employee making 50 mL lysate from a 10 L fermentation, assuming 30 lysate batches per year and a 5-year equipment lifetime. Component source is also included in the table if it is available to purchase directly from a supplier. Homemade components cannot be purchased directly and must be prepared according to procedures described in the Materials and Methods section.

Component	Cost ($/mL rxn)	Supplier	Product No
$Mg(Glu)_2$	<0.00	Sigma	49605
NH_4Glu	<0.00	MP	02180595
KGlu	<0.00	Sigma	G1501
ATP	0.01	Sigma	A2383
GTP	0.27	Sigma	G8877
UTP	0.23	Sigma	U6625
CTP	0.20	Sigma	C1506
Folinic acid	0.02	Sigma	47612

14 14 This chapter appears in the Science Advances journal:

Stark, J.C.*, Jaroentomeechai, T.*, Moeller, T.D., Dubner, R.S., Hsu, K.J., Stevenson, T.C., DeLisa, M.P., and Jewett, M.C. (2021) On-demand, cell-free biomanufacturing of conjugate vaccines at the point-of-care. *Sci Adv*. 7: 1-15.

tRNA	0.21	Roche	10109541001
Amino acids	<0.00	homemade	
PEP	1.79	Roche	10108294001
NAD	0.07	Sigma	N8535-15VL
CoA	0.34	Sigma	C3144
Oxalic acid	<0.00	Sigma	P0963
Putrescine	<0.00	Sigma	P5780
Spermidine	<0.00	Sigma	S2626
HEPES	<0.00	Sigma	H3375
$MnCl_2$	<0.00	Sigma	63535
DDM	0.36	Anatrace	D310S
Plasmid	0.88	homemade	
Lysate	7.37	homemade	
Total	**11.75**	**$/mL rxn**	
	5.88	**$/dose**	

Supplementary Table 7-2. Plasmids used in the iVAX study.

Plasmid	Description	Source
pSF-*Cj*PglB	*C. jejuni* PglB with a C-terminal 1xFLAG epitope tag in pSF, a modified pBAD expression vector	(Ollis et al., 2014b)
pGAB2	*F. tularensis* O-PS antigen gene cluster in pLAFR1	(Cuccui et al., 2013b)
pMW07-O78	*E. coli* O78 antigen gene cluster in pMW07	(Celik et al., 2015)
pJHCV32	*E. coli* O7 antigen gene cluster in pVK102	(Valvano and Crosa, 1989)
pKD46	Encodes λ red system for recombineering	(Datsenko and Wanner, 2000b)
pKD4	Encodes kanamycin resistance cassette with upstream and downstream FRT sites	(Datsenko and Wanner, 2000b)
pCP20	Encodes *flp* for Flp-FRT recombination	(Datsenko and Wanner, 2000b)
pSF-*Cj*PglB-LpxE	*C. jejuni* PglB with a C-terminal 1xFLAG epitope tag and *F. tularensis* LpxE in pSF	This work; Addgene 128389
pJL1-sfGFP$^{217\text{-}DQNAT}$	Superfolder green fluorescent protein variant modified after residue T216 with 21 amino acid insertion containing the *C. jejuni* AcrA N123 glycosylation site but with an optimal DQNAT glycosylation sequence and a C-terminal 6xHis tag	(Jaroentomeechai et al., 2018a)
pJL1-sfGFP$^{217\text{-}AQNAT}$	Same as pJL1 sfGFP$^{217\text{-}DQNAT}$, but with an AQNAT glycosylation sequence that is not modified by *Cj*PglB	(Jaroentomeechai et al., 2018a)
pJL1-MBP^{4xDQNAT}	*E. coli* maltose binding protein with a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pJL1, a T7-driven *in vitro* expression vector	This work; Addgene 128390
pJL1-PD^{4xDQNAT}	*H. influenzae* protein D with a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pJL1	This work; Addgene 128391

pJL1-PorA^{4xDQNAT}	*N. meningitidis* PorA porin protein with a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pJL1	This work; Addgene 128392
pJL1-TTc^{4xDQNAT}	Fragment C domain of *C. tetani* toxin with a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pJL1	This work; Addgene 128393
pJL1-TTlight^{4xDQNAT}	Light chain variant of *C. tetani* toxin containing an inactivating E234A mutation in the enzyme active site with a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pJL1	This work; Addgene 128394
pJL1-CRM197^{4xDQNAT}	*C. diphtheriae* toxin variant with an inactivating G52E mutation in the enzyme active site with a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pJL1	This work; Addgene 128395
pJL1-TT^{4xDQNAT}	*C. tetani* toxin variant containing an inactivating E234A mutation in the enzyme active site with a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pJL1	This work; Addgene 128396
pJL1-EPA$^{DNNNS-DQNRT}$	*P. aeruginosa* exotoxin A containing a DNNNS glycosylation site at residue 242 and a DQNRT glycosylation site at residue 384 and a C-terminal 6xHis tag in pJL1	This work; Addgene 128397
pTrc99s-ssDsbA-MBP^{4xDQNAT}	*E. coli* maltose binding protein with an N-terminal DsbA signal sequence for periplasmic translocation and a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pTrc99s	This work; Addgene 128398
pTrc99s-ssDsbA-PD^{4xDQNAT}	*H. influenzae* protein D with an N-terminal DsbA signal sequence for periplasmic translocation and a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in Trc99s	This work; Addgene 128399
pTrc99s-ssDsbA-PorA^{4xDQNAT}	*N. meningitidis* PorA porin protein with an N-terminal DsbA signal sequence for periplasmic translocation and a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pTrc99s	This work; Addgene 128400
pTrc99s-ssDsbA-TTc^{4xDQNAT}	Fragment C domain of *C. tetani* toxin with an N-terminal DsbA signal sequence for periplasmic	This work; Addgene 128401

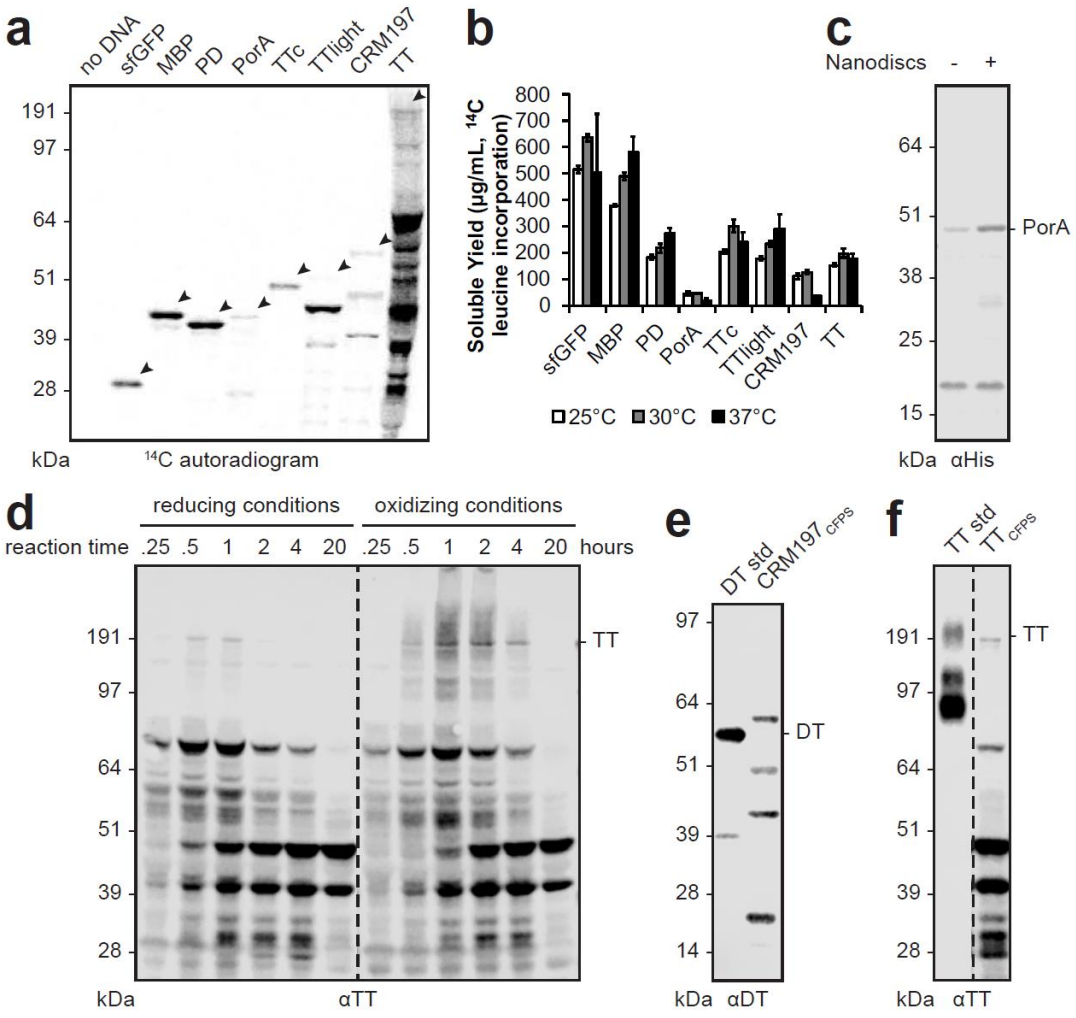
	translocation and a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pTrc99s	
pTrc99s-ssDsbA-TTlight^{4xDQNAT}	Light chain variant of *C. tetani* toxin containing an inactivating E234A mutation in the enzyme active site with an N-terminal DsbA signal sequence for periplasmic translocation and a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pTrc99s	This work; Addgene 128402
pTrc99s-ssDsbA-CRM1974xDQNAT	*C. diphtheriae toxin variant with an inactivating G52E mutation in the enzyme active site with an N-terminal DsbA signal sequence for periplasmic translocation and a C-terminal 4xDQNAT glycosylation tag and a 6xHis tag in pTrc99s*	This work; Addgene 128403
pTrc99s-ssDsbA-EPA$^{DNNNS-DQNRT}$	*P. aeruginosa* exotoxin A containing a DNNNS glycosylation site at residue 242 and a DQNRT glycosylation site at residue 384 with an N-terminal DsbA signal sequence for periplasmic translocation and a C-terminal 6xHis tag in pTrc99s	This work; Addgene 128404

Supplementary Table 7-3. Primers used to generate CLM24 Δ*lpxM*. Primers used to construct and verify the CLM24 Δ*lpxM* strain are listed below. KO primers were used for amplification of the kanamycin resistance cassette from pKD4 with homology to *lpxM*. Seq primers were used for colony PCRs and sequencing confirmation of knockout strains.

Primer Name	DNA Sequence (5′ to 3′)
lpxM KO for	TACACTATCACCAGATTGATTTTTGCCTTATCCGAAACTGGAAAAGCATGGTGTAGGCTGGAGCTGCTTC
lpxM KO rev	GCGAAGGCCTCTCCTCGCGAGAGGCTTTTTTATTTGATGGGATAAAGATCCATATGAATATCCTCCTTAGTTCCTATTC
lpxM seq for	AGTACCGGCTTTTTTTATTTGG
lpxM seq rev	CTAATACCACGCGTATTTTAACG

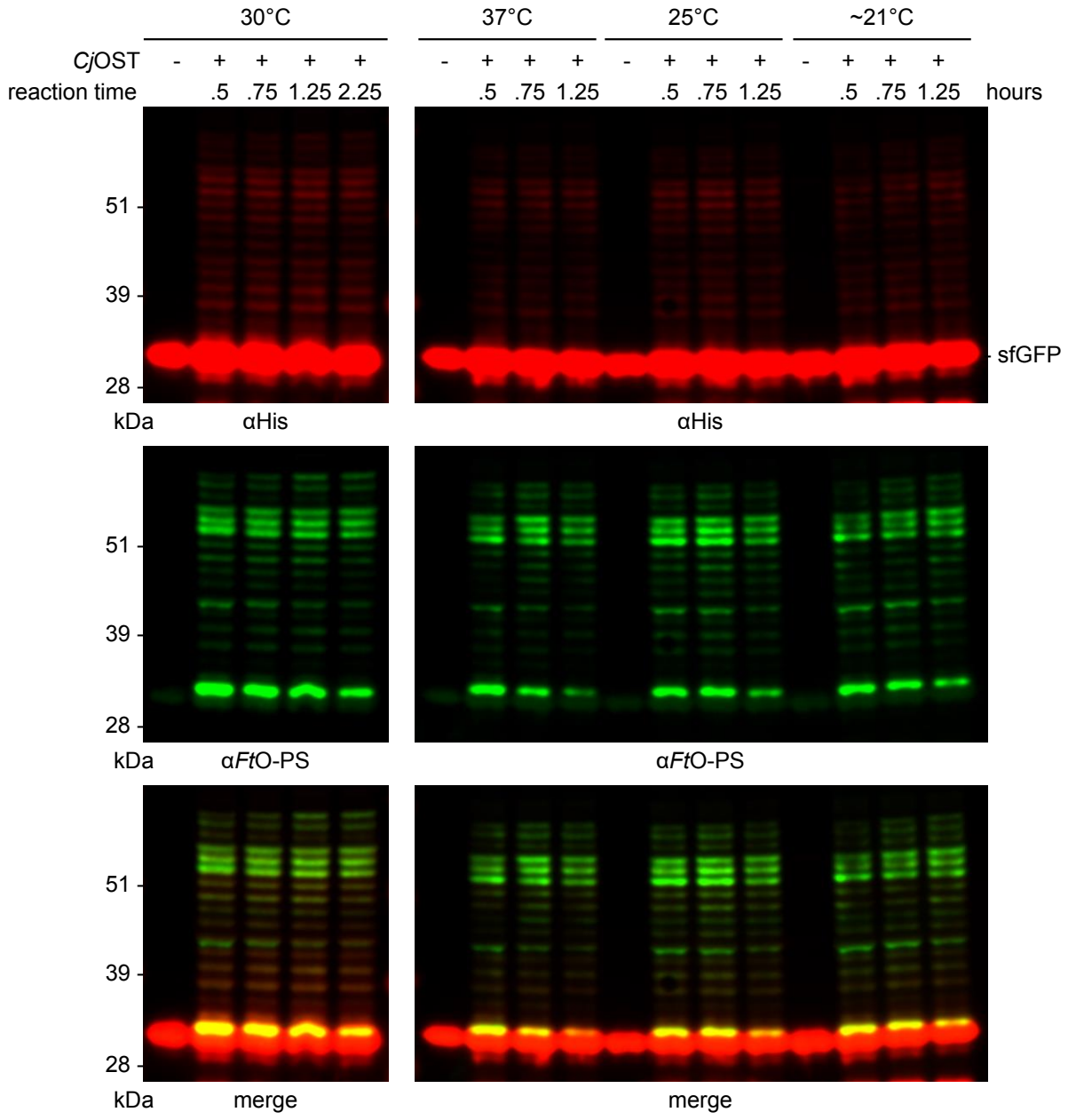
Supplementary Table 7-4. Antibodies and antisera used in the iVAX study.

Target	Source	Dilution
Rabbit pAb to 6xHis epitope tag	Abcam	1:7500
Mouse mAb FB11 to *F. tularensis* LPS	Abcam	1:5000
Rabbit pAb to *E. coli* O78 antigen	Abcam	1:2500
Rabbit pAb to *C. diphtheriae* toxin	Abcam	1:2000
Rabbit pAb to *C. tetani* toxin	Abcam	1:2000
Goat anti-rabbit IgG IR dye 680	LI-COR	1:15000-1:10000
Goat anti-rabbit IgG IR dye 800	LI-COR	1:15000-1:10000
Goat anti-mouse IgG IR dye 800	LI-COR	1:15000-1:10000
Goat anti-mouse IgG HRP	Abcam	1:25,000
Goat anti-mouse IgG1 HRP	Abcam	1:25,000
Goat anti-mouse IgG2a HRP	Abcam	1:25,000



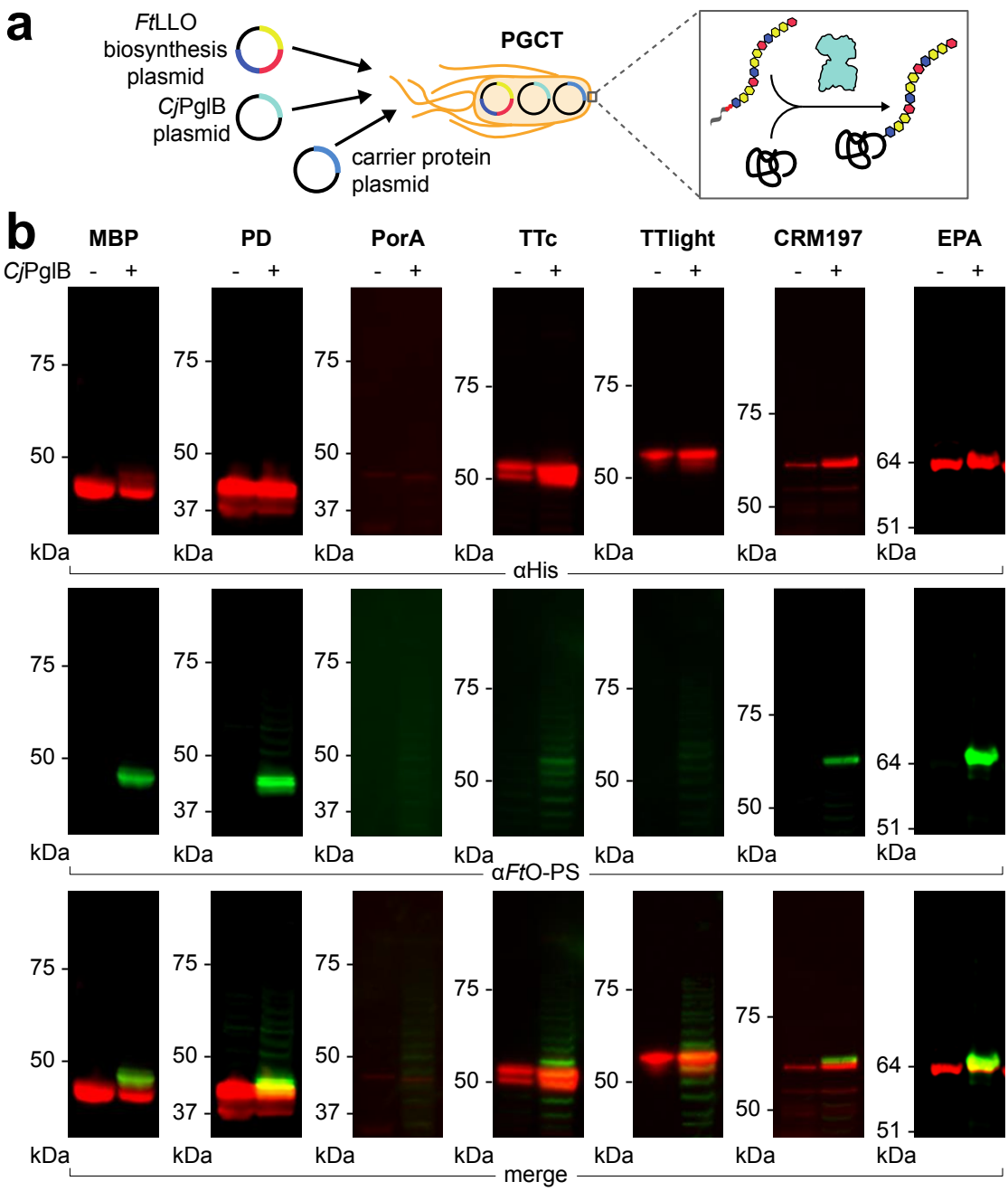
Supplementary Figure 7-1. *In vitro* synthesis of licensed conjugate vaccine carrier proteins is possible over a range of temperatures and can be readily optimized. (**a**) ^{14}C autoradiogram analysis of soluble fractions from CFPS reactions shows no evidence of endogenous protein translation (no DNA control) and some fraction of full-length protein synthesis for all carriers tested. (**b**) With the exception of CRM197, all carriers expressed with similar soluble yields at 25°C, 30°C, and 37°C, as measured by ^{14}C-leucine incorporation. Values represent means and error bars represent standard deviations of biological replicates (n = 3). (**c**) Soluble expression of PorA was

improved through the addition of lipid nanodiscs to the reaction. (**d**) Expression of full-length TT was enhanced by (i) performing *in vitro* protein synthesis in oxidizing conditions to improve assembly of the disulfide-bonded heavy and light chains into full-length TT and (ii) allowing reactions to run for only 2 h to minimize protease degradation. (**e**) CRM197 and (**f**) TT produced in CFPS reactions are detected with α-DT and α-TT antibodies, respectively, and are comparable in size to commercially available purified DT and TT protein standards (50 ng standard loaded). Images are representative of at least three biological replicates. Dashed line indicates samples are from the same blot with the same exposure. Molecular weight ladders are shown at the left of each image.



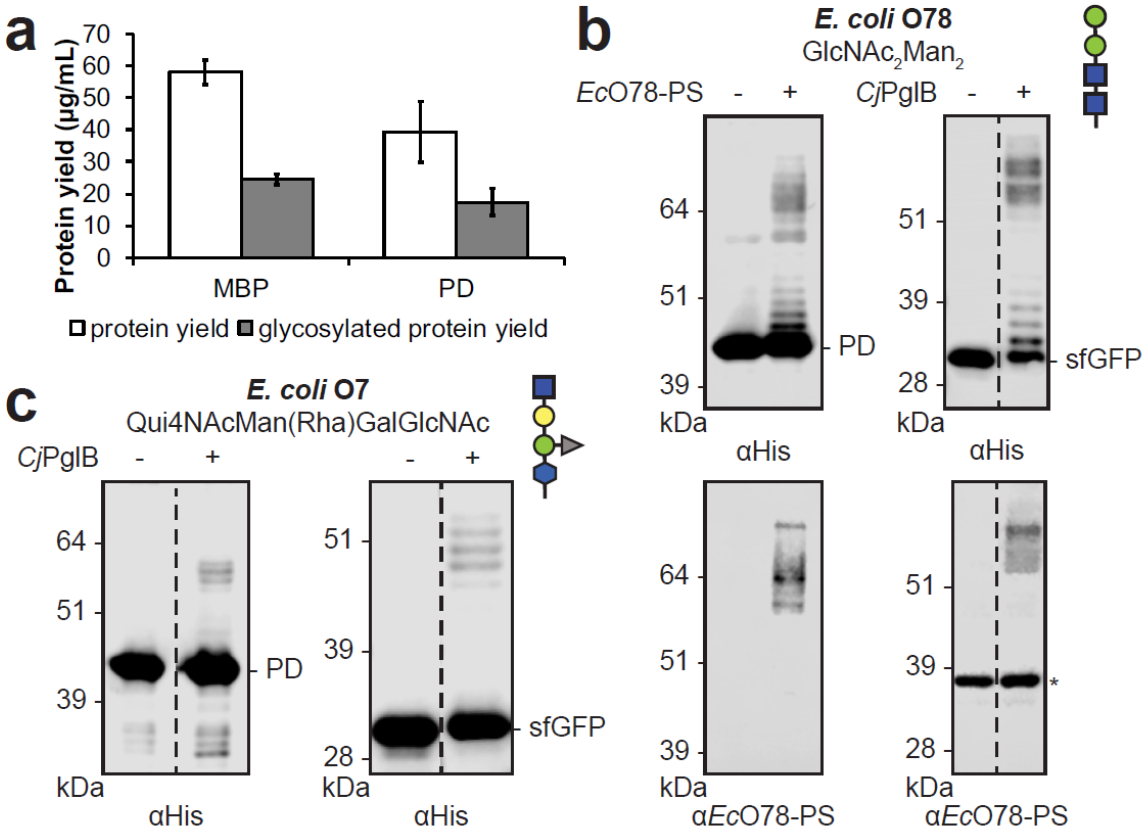
Supplementary Figure 7-2. Glycosylation in iVAX reactions occurs in 1 h over a range of temperatures. Kinetics of *Ft*O-PS glycosylation at 30°C (**left**), 37°C, 25°C, and room temperature (~21°C) (**right**) are comparable and show that protein synthesis and glycosylation occur in the first hour of the iVAX reaction. These results demonstrate

that the iVAX platform can synthesize conjugates over a range of permissible temperatures. Top panels show signal from probing with anti-hexa-histidine antibody (*α*His) to detect the carrier protein, middle panels show signal from probing with commercial anti-*Ft*O-PS antibody (*αFt*O-PS), and bottom panels show *α*His and *αFt*O-PS signals merged. Images are representative of at least three biological replicates. Molecular weight ladders are shown at the left of each image.



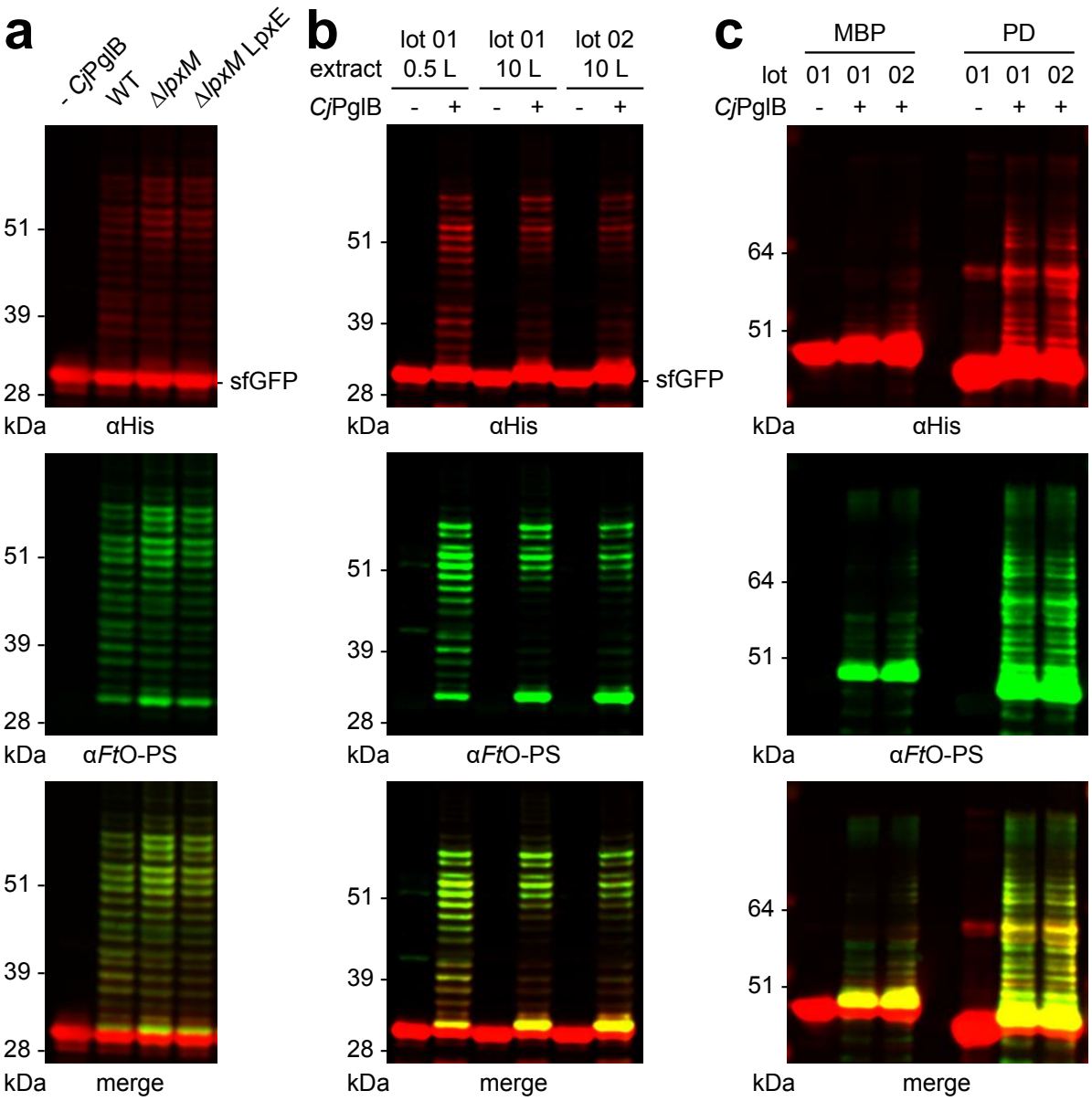
Supplementary Figure 7-3. Production of conjugate vaccines against *F. tularensis* using PGCT in living *E. coli*. (**a**) Conjugates were produced via PGCT in CLM24 cells expressing *Cj*PglB, the biosynthetic pathway for *Ft*O-PS, and a panel of immunostimulatory carriers including those used in licensed vaccines. (**b**) We observed low expression of the licensed conjugate vaccine carrier proteins PorA and CRM197, compared to iVAX-derived samples. Top panels show signal from probing

with anti-hexa-histidine antibody (αHis) to detect the carrier protein, middle panels show signal from probing with commercial anti-*Ft*O-PS antibody (α*Ft*O-PS), and bottom panels show αHis and α*Ft*O-PS signals merged. Images are representative of at least three biological replicates. Molecular weight ladders are shown at the left of each image.



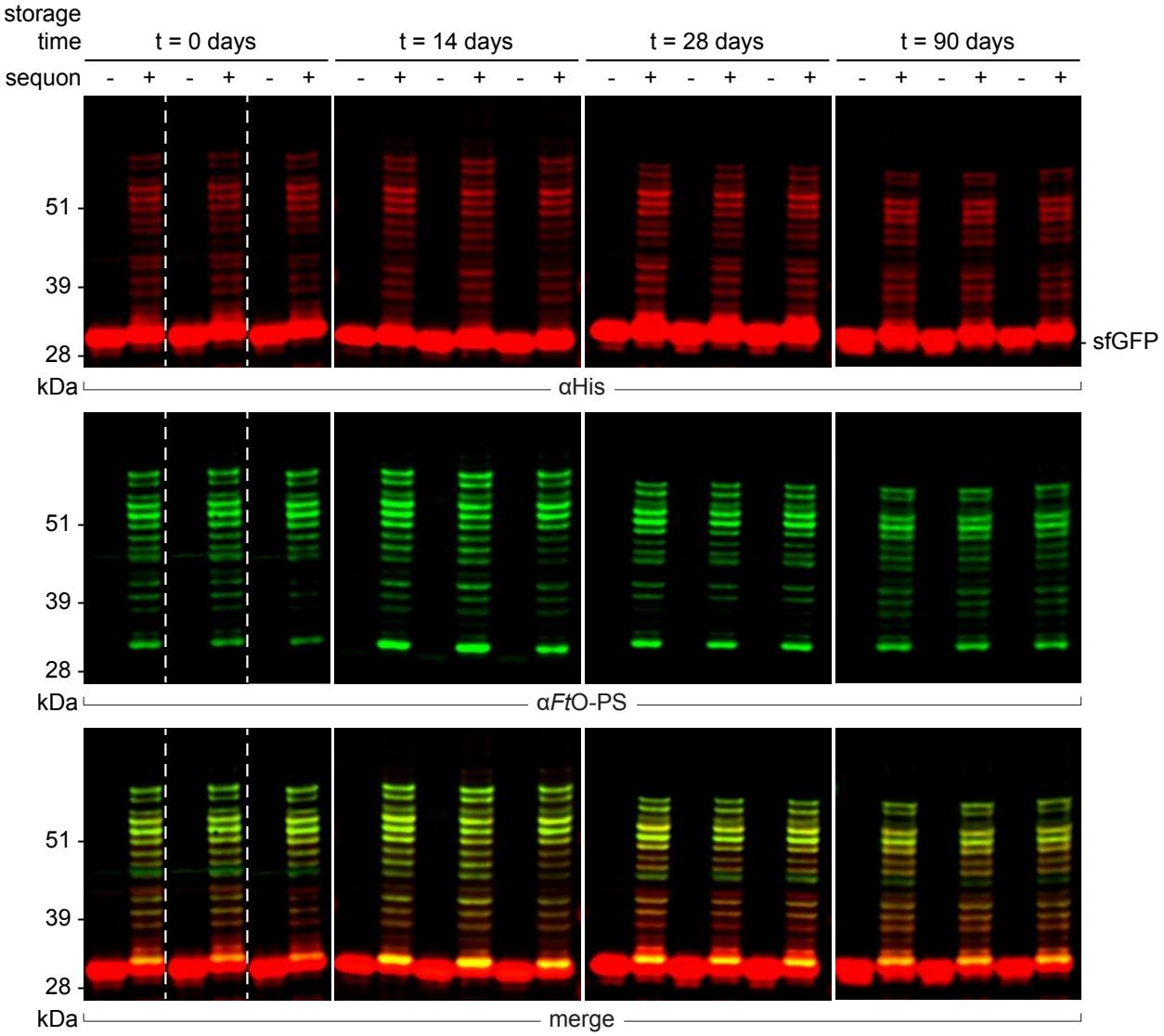
Supplementary Figure 7-4. The iVAX platform is modular and can be used to synthesize clinically relevant yields of diverse conjugate vaccines. (**a**) Protein synthesis and glycosylation with *Ft*O-PS were measured in iVAX reactions producing MBP^{4xDQNAT} and PD^{4xDQNAT}. After ~1 h, reactions produced ~40 μg mL^{-1} protein, as measured via ^{14}C-leucine incorporation, of which ~20 μg mL^{-1} was glycosylated with *Ft*O-PS, as determined by densitometry. Values represent means and error bars represent standard errors of biological replicates (n = 2). To demonstrate modularity, iVAX lysates were prepared from cells expressing *Cj*PglB and biosynthetic pathways for either (**b**) the *E. coli* O78 antigen or (**c**) the *E. coli* O7 antigen and used to synthesize PD^{4xDQNAT} (**left**) or sfGFP$^{217\text{-}DQNAT}$ (**right**) conjugates. The structure and composition of the repeating monomer unit for each antigen is shown. Both polysaccharide antigens are compositionally and, in the case of the O7 antigen, structurally distinct compared

to the *F. tularensis* O antigen. Blots show signal from probing with anti-hexa-histidine antibody (αHis) to detect the carrier protein. If a commercial anti-O-PS serum or antibody was available, it was used to confirm the identity of the conjugated O antigen (α-*Ec*O78 blots, panel **b**). Asterisk denotes bands resulting from non-specific serum antibody binding. Images are representative of at least three biological replicates. Dashed lines indicate samples are from nonadjacent lanes of the same blot with the same exposure. Molecular weight ladders are shown at the left of each image.



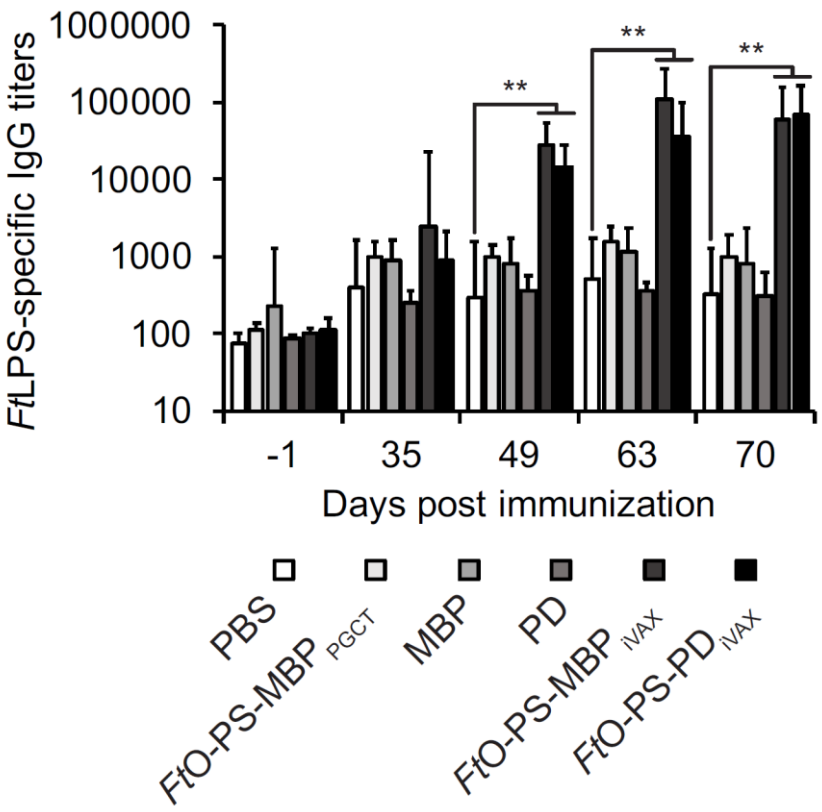
Supplementary Figure 7-5. Detoxified lysate production and freeze-dried reactions scale reproducibly. (**a**) iVAX lysates containing *Cj*PglB and *Ft*O-PS were prepared from wild-type CLM24, CLM24 Δ*lpxM*, or CLM24 Δ*lpxM* cells expressing *Ft*LpxE. Lysates from engineered strains retained the ability to glycosylate sfGFP$^{217\text{-DQNAT}}$ with *Ft*O-PS. (**b**) Fermentations to produce endotoxin-edited iVAX lysates were scaled from 0.5 L to 10 L. We observed similar levels glycosylation at large and small scale and across different batches of lysate from 10 L fermentations (average glycosylation

efficiencies of sfGFP$^{217\text{-}DQNAT}$ using lysates from 0.5 L and 10 L fermentations were 69 ± 5% and 69 ± 1% by densitometry, respectively). (**c**) For immunizations, we prepared two lots of *Ft*O-PS-conjugated MBP^{4xDQNAT} and PD^{4xDQNAT} from 5 mL freeze-dried iVAX reactions. We observed similar levels of purified protein (~200 μg) and *Ft*O-PS modification (66 ± 11% for MBP^{4xDQNAT} and 70 ± 1% for PD^{4xDQNAT} by densitometry) across both carriers and lots of material. Top panels show signal from probing with anti-hexa-histidine antibody (*α*His) to detect the carrier protein, middle panels show signal from probing with commercial anti-*Ft*O-PS antibody (*αFt*O-PS), and bottom panels show *α*His and *αFt*O-PS signals merged. Images are representative of at least three biological replicates. Molecular weight ladders are shown at left.



Supplementary Figure 7-6. iVAX reactions are stable under ambient temperature storage conditions for at least 3 months. Stability of iVAX reactions was tested over a period of 3 months using economic packaging materials. Specifically, we vacuum sealed reactions using a commercial FoodSaver® appliance with Dri-Card™ desiccant cards enclosed to prevent rehydration of the iVAX pellets. At each time point, n = 3 reactions were rehydrated with plasmid encoding sfGFP$^{217\text{-}AQNAT}$ (- sequon) or sfGFP$^{217\text{-}DQNAT}$ (+ sequon). The average amount of protein synthesized across all replicates and time points was 19.78 ± 3.12 μg mL^{-1} (n = 12). These results show that iVAX reactions are stable at room temperature for at least 3 months, with no observable differences in

protein synthesis or glycosylation activity, highlighting their potential for portable, on-demand conjugate vaccine production. Top panels show signal from probing with anti-hexa-histidine antibody (αHis) to detect the carrier protein, middle panels show signal from probing with commercial anti-*Ft*O-PS antibody (α*Ft*O-PS), and bottom panels show αHis and α*Ft*O-PS signals merged. Dashed lines indicate samples are from nonadjacent lanes of the same blot with the same exposure. Molecular weight ladders are shown at the left of each image.



Supplementary Figure 7-7. *Ft*LPS-specific antibody titers in vaccinated mice over time. Six groups of BALB/c mice were immunized subcutaneously with PBS or 7.5 μg of purified, cell-free synthesized aglycosylated MBP^{4xDQNAT}, *Ft*O-PS-conjugated MBP^{4xDQNAT}, aglycosylated PD^{4xDQNAT}, or *Ft*O-PS-conjugated PD^{4xDQNAT}. *Ft*O-PS-conjugated MBP^{4xDQNAT} prepared in living *E. coli* cells using PCGT was used as a positive control. Each group was composed of six mice except for the PBS control group, which was composed of five mice. Mice were boosted on days 21 and 42 with identical doses of antigen. *Ft*LPS-specific IgG titers were measured by ELISA in serum collected on day -1, 35, 49, 63, and 70 following initial immunization. iVAX-derived bioconjugates elicited significantly higher levels of *Ft*LPS-specific IgG compared to compared to the PBS control group in serum collected on day 35, 49, and 70 of the study (**$p < 0.01$, Tukey-Kramer HSD). Values represent means and error bars represent standard errors of *Ft*LPS-specific IgGs detected by ELISA.

CHAPTER 8

APPENDIX C[15]

Supplementary information: *A universal glycoenzyme biosynthesis pipeline that enables cell-free construction and remodeling of glycans*

Supplementary Table 8-1. Strains, cell lines, and plasmids used in the SIMPLEx study.

Strain	Genotype	Source
Bacterial strain		
DH5α	F- (Φ80ΔlacZDM15,) Δ(lacIZYA-argF)U169 recA1 endA1 hsdR17(rk-,mk+) phoA supE44 λ- thi-1 gyrA96 relA1	Laboratory stock
BL21DE3	F- ompT gal dcm lon hsdSB(rB- mB-) λ(DE3)	Laboratory stock
SHuffle® T7 Express lysY	MiniF lysY (CamR) / fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB sulA11 R(mcr-73::miniTn10–TetS)2 [dcm] R(zgb-210::Tn10 –TetS) endA1 Δgor Δ(mcrC-mrr)114::IS10	NEB
Yeast strain		
SBY49	MATa ; pep4Δ ::LEU2 ; prb1Δ ::LEU2 ; ura3-52 his3-200 leu2-3,112 lys2-801 suc2-9	Scott D. Emr's lab
Cell line		
HEK293T		Laboratory stock

15 This chapter is in preparation for submission.

Plasmid	Description	Source
pET28a(+)	T7 lac promoter; Kan^R	Novagen
pET28a-SIMPLEx-GFP	ΔspMBP-GFP, and ApoAI*-6xHis, cloned in pET28a; Kan^R	1
pcDNA3	CMV promoter; Amp^R	Laboratory stock
pJL1-sfGFP	T7 lac promoter, cell-free expression vector; Kan^R	2
pYS338	CPS promoter, Galactose inducible vector; Amp^R	Scott D. Emr's lab
pET28a-SIMPLEx-Δ25 *Hs*FUT1	Chimeric gene comprised of *E. coli* MBP lacking its signal peptide (K27-T395); ΔspMBP or cMBP), truncated human FUT1 (Δ1-25), and truncated human ApoAI (Δ1-43; ApoAI*) followed by 6x-His tags, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*FUT2	ΔspMBP-truncated human FUT2 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ34 *Hs*FUT3	ΔspMBP-truncated human FUT3 (Δ1-34), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ172 *Hs*FUT4	ΔspMBP-truncated human FUT4 (Δ1-172), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ34 *Hs*FUT5	ΔspMBP-truncated human FUT5 (Δ1-34), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ34 *Hs*FUT6	ΔspMBP-truncated human FUT6 (Δ1-34), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ36 *Hs*FUT7	ΔspMBP-truncated human FUT7 (Δ1-36), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ30 *Hs*FUT8	ΔspMBP-truncated human FUT8 (Δ1-30), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ32 *Hs*FUT9	ΔspMBP-truncated human FUT9 (Δ1-32), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study

pET28a-SIMPLEx-Δ31 *Hs*FUT10	ΔspMBP-truncated human FUT10 (Δ1-31), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ24 *Hs*FUT11	ΔspMBP-truncated human FUT11 (Δ1-24), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ26 *Hs*POFUT1	ΔspMBP-truncated human POFUT1 (Δ1-26), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ34 *Hs*ST3Gal1	ΔspMBP-truncated human ST3Gal1 (Δ1-34), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*ST3Gal3	ΔspMBP-truncated human ST3Gal3 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ26 *Hs*ST3Gal4	ΔspMBP-truncated human ST3Gal4 (Δ1-26), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ25 *Hs*ST3Gal6	ΔspMBP-truncated human ST3Gal6 (Δ1-25), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ26 *Hs*ST6Gal1	ΔspMBP-truncated human ST6Gal1 (Δ1-26), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*ST6Gal1	ΔspMBP-human ST6Gal1, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ35 *Hs*ST6GalNAc1	ΔspMBP-truncated human ST3GalNAc1 (Δ1-35), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*ST6GalNAc2	ΔspMBP-truncated human ST3GalNAc2 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ27 *Hs*ST6GalNAc4	ΔspMBP-truncated human ST3GalNAc4 (Δ1-27), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ48 *Hs*ST8Sia1	ΔspMBP-truncated human ST8Sia1 (Δ1-48), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ23 *Hs*ST8Sia2	ΔspMBP-truncated human ST8Sia2 (Δ1-23), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study

pET28a-SIMPLEx-Δ33 *Hs*ST8Sia3	ΔspMBP-truncated human ST8Sia3 (Δ1-33), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ20 *Hs*ST8Sia4	ΔspMBP-truncated human ST8Sia4 (Δ1-20), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*ppGalNAcT1	ΔspMBP-truncated human ppGalNAcT1 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ24 *Hs*ppGalNAcT2	ΔspMBP-truncated human ppGalNAcT2 (Δ1-24), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ37 *Hs*ppGalNAcT3	ΔspMBP-truncated human ppGalNAcT3 (Δ1-37), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ35 *Hs*ppGalNAcT4	ΔspMBP-truncated human ppGalNAcT4 (Δ1-35), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ35 *Hs*ppGalNAcT5	ΔspMBP-truncated human ppGalNAcT5 (Δ1-35), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*ppGalNAcT6	ΔspMBP-truncated human ppGalNAcT6 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ29 *Hs*ppGalNAcT7	ΔspMBP-truncated human ppGalNAcT7 (Δ1-29), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ29 *Hs*ppGalNAcT8	ΔspMBP-truncated human ppGalNAcT8 (Δ1-29), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*ppGalNAcT9	ΔspMBP-truncated human ppGalNAcT9 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ31 *Hs*ppGalNAcT10	ΔspMBP-truncated human ppGalNAcT10 (Δ1-31), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ43 *Hs*B3GALNT1	ΔspMBP-truncated human B3GALNT1 (Δ1-43), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study

pET28a-SIMPLEx-Δ25 *Hs*B4GALNT1	ΔspMBP-truncated human B4GALNT1 (Δ1-25), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ53 *Hs*-A-group	ΔspMBP-truncated human A-ABO (Δ1-53), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ43 *Hs*A4GalT	ΔspMBP-truncated human A4GalT (Δ1-43), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ26 *Hs*B3GalT1	ΔspMBP-truncated human B3GalT1 (Δ1-26), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ45 *Hs*B3GalT2	ΔspMBP-truncated human B3GalT2 (Δ1-45), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ44 *Hs*B4GalT1	ΔspMBP-truncated human B4GalT1 (Δ1-44), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ36 *Hs*B4GalT2	ΔspMBP-truncated human B4GalT2 (Δ1-36), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ31 *Hs*B4GalT3	ΔspMBP-truncated human B4GalT3 (Δ1-31), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ38 *Hs*B4GalT4	ΔspMBP-truncated human B4GalT4 (Δ1-38), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ35 *Hs*B4GalT5	ΔspMBP-truncated human B4GalT5 (Δ1-35), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ35 *Hs*B4GalT6	ΔspMBP-truncated human B4GalT6 (Δ1-35), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ53 *Hs*-B-group	ΔspMBP-truncated human B-ABO (Δ1-53), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ20 *Hs*UGT8	ΔspMBP-human UGT8, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ29 *Hs*C1GLT	ΔspMBP-human C1GLT, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study

pET28a-SIMPLEx-*Hs*COSMC	ΔspMBP-human COSMC, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ23 *Hs*Alg1	ΔspMBP-truncated human Alg1 (Δ1-23), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*Alg2	ΔspMBP-human Alg2, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*Alg3	ΔspMBP-human Alg3, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ40 *Hs*Alg11	ΔspMBP-truncated human Alg11 (Δ1-40), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*Alg12	ΔspMBP-human Alg12, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*Alg13	ΔspMBP-human Alg13, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ24 *Hs*Alg14	ΔspMBP-truncated human Alg14 (Δ1-24), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*DPM1	ΔspMBP-human DPM1, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*PIGM	ΔspMBP-human PIGM, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*PIGB	ΔspMBP-human PIGB, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*PIGZ	ΔspMBP-human PIGZ, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*Alg5	ΔspMBP-truncated human Alg5 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*Alg6	ΔspMBP-human Alg6, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study

pET28a-SIMPLEx-*Hs*Alg8	ΔspMBP-human Alg8, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*Alg10	ΔspMBP-human Alg10, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*UGCG	ΔspMBP-human UGCG, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ27 *Hs*B3GLCT	ΔspMBP-truncated human B3GLCT (Δ1-27), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*GLYG	ΔspMBP-human Glycogenin, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*POGLUT1	ΔspMBP-human POGLUT1, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ29 *Hs*GnT1	ΔspMBP-truncated human GnT1 (Δ1-29), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ29 *Hs*GnT2	ΔspMBP-truncated human GnT2 (Δ1-29), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ23 *Hs*GnT3	ΔspMBP-truncated human GnT3 (Δ1-23), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ27 *Hs*GnT4a	ΔspMBP-truncated human GnT4a (Δ1-27), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ32 *Hs*GCNT1	ΔspMBP-truncated human GCNT1 (Δ1-32), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ23 *Hs*GCNT2	ΔspMBP-truncated human GCNT2 (Δ1-23), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*B3GNT2	ΔspMBP-truncated human B3GNT2 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ31 *Hs*B3GNT6	ΔspMBP-truncated human B3GNT6 (Δ1-31), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study

pET28a-SIMPLEx-*Hs*PIGA	ΔspMBP-human PIGA, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ25 *Hs*UGT1A1	ΔspMBP-truncated human UGT1A1 (Δ1-25), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ36 *Hs*UGT1A3	ΔspMBP-truncated human UGT1A3 (Δ1-36), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ36 *Hs*B4GAT1	ΔspMBP-truncated human B4GAT1 (Δ1-36), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ28 *Hs*XXLT1	ΔspMBP-truncated human XXLT1 (Δ1-28), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Cj*CstII	ΔspMBP-*C. jejuni* CstII, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Nm*Pst	ΔspMBP-*N. meningitidis* PolysiaT, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Cj*CgtB	ΔspMBP-*C. jejuni* CgtB, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hp*LgtB	ΔspMBP-*H. pylori* GalT, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Nm*LgtB	ΔspMBP-*N. meningitidis* GalT, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Ng*LgtB	ΔspMBP-*N. gonorrhea* GalT, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Ec*WbgL	ΔspMBP-*E. coli* WbgL, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Ec*WecA	ΔspMBP-*E. coli* WecA, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Lp*SetA	ΔspMBP-*L. pneumophila* SetA, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study

pET28a-SIMPLEx-*Nm*SynE	ΔspMBP-*N. meningitidis* SynE, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ34 *Sc*Alg1	ΔspMBP-truncated *S. cerevisae* Alg1 (Δ1-34), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Sc*Alg2	ΔspMBP-*S. cerevisae* Alg2, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ45 *Sc*Alg11	ΔspMBP-truncated *S. cerevisae* Alg11 (Δ1-45), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ30 *Nt*GnTI	ΔspMBP-truncated *N. tabacum* GnTI (Δ1-30), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ35 *Nt*GnTII	ΔspMBP-truncated *N. tabacum* GnTII (Δ1-35), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ22 *Bt*GGTA1	ΔspMBP-truncated *B. 302aurus* GGTA1 (Δ1-35), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ60 *Mm*GGTA1	ΔspMBP-truncated *M. musculus* GGTA1 (Δ1-60), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ58 *Rn*GGTA1	ΔspMBP-truncated *R. norvergicus* GGTA1 (Δ1-58), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-Δ44 *Bt*B4GalT1	ΔspMBP-truncated *B. 302aurus* B4GalT1 (Δ1-35), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*CDK4	ΔspMBP-human cyclin-dependent kinase 4, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*CDKN2A	ΔspMBP-human cyclin-dependent kinase inhibitor 2A, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*EGFR_TK	ΔspMBP-truncated human EGFR (P694-G1022), and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*FOS	ΔspMBP-human proto-oncogene c-Fos, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study

pET28a-SIMPLEx-*Hs*GATA2	ΔspMBP-human endothelial transcription factor GATA, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*JUN	ΔspMBP-human transcription factor JUN, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*MMP1	ΔspMBP-human Interstitial collagenase, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pET28a-SIMPLEx-*Hs*Insulin	ΔspMBP-human insulin, and ApoAI*-6xHis, cloned in pET28a; Kan^R	This study
pcDNA3-SIMPLEx-Δ26 *Hs*ST6Gal1	ΔspMBP-truncated human ST6Gal1 (Δ1-26), and ApoAI*-6xHis, cloned in pcDNA3; Amp^R	This study
pYS338-SIMPLEx-Δ26 *Hs*ST6Gal1	ΔspMBP-truncated human ST6Gal1 (Δ1-26), and ApoAI*-6xHis, cloned in pYS338; Amp^R	This study
pJL1-SIMPLEx-Δ26 *Hs*ST6Gal1	ΔspMBP-truncated human ST6Gal1 (Δ1-26), and ApoAI*-6xHis, cloned in pJL1; Kan^R	This study
pET28a-Δ25 *Hs*FUT1	Truncated human FUT1 (Δ1-25) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*FUT2	Truncated human FUT2 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ34 *Hs*FUT3	Truncated human FUT3 (Δ1-34) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ172 *Hs*FUT4	Truncated human FUT4 (Δ1-172) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ34 *Hs*FUT5	Truncated human FUT5 (Δ1-34) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ34 *Hs*FUT6	Truncated human FUT6 (Δ1-34) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ36 *Hs*FUT7	Truncated human FUT7 (Δ1-36) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ30 *Hs*FUT8	Truncated human FUT8 (Δ1-30) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ32 *Hs*FUT9	Truncated human FUT9 (Δ1-32) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ31 *Hs*FUT10	Truncated human FUT10 (Δ1-31) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ24 *Hs*FUT11	Truncated human FUT11 (Δ1-24) with C-terminal 6x-His tags in pET28a; Kan^R	This study

pET28a-Δ26 *Hs*POFUT1	Truncated human POFUT1 (Δ1-26) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ34 *Hs*ST3Gal1	Truncated human ST3Gal1 (Δ1-34) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*ST3Gal3	Truncated human ST3Gal3 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ26 *Hs*ST3Gal4	Truncated human ST3Gal4 (Δ1-26) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ25 *Hs*ST3Gal6	Truncated human ST3Gal6 (Δ1-25) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ26 *Hs*ST6Gal1	Truncated human ST6Gal1 (Δ1-26) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*ST6Gal1	Human ST6Gal1 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-MBP-Δ26 *Hs*ST6Gal1	ΔspMBP-truncated human ST6Gal1 (Δ1-26) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ26 *Hs*ST6Gal1-ApoAI	Truncated human ST6Gal1 (Δ1-26) with ApoAI*-6xHis in pET28a; Kan^R	This study
pET28a-Δ35 *Hs*ST6GalNAc1	Truncated human ST6GalNAc1 (Δ1-35) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*ST6GalNAc2	Truncated human ST6GalNAc2 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ27 *Hs*ST6GalNAc4	Truncated human ST6GalNAc4 (Δ1-27) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ48 *Hs*ST8Sia1	Truncated human ST8Sia1 (Δ1-48) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ23 *Hs*ST8Sia2	Truncated human ST8Sia2 (Δ1-23) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ33 *Hs*ST8Sia3	Truncated human ST8Sia3 (Δ1-33) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ20 *Hs*ST8Sia4	Truncated human ST8Sia4 (Δ1-20) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*ppGalNAcT1	Truncated human ppGalNAcT1 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ24 *Hs*ppGalNAcT2	Truncated human ppGalNAcT2 (Δ1-24) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ37 *Hs*ppGalNAcT3	Truncated human ppGalNAcT3 (Δ1-37) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ35 *Hs*ppGalNAcT4	Truncated human ppGalNAcT4 (Δ1-35) with C-terminal 6x-His tags in pET28a; Kan^R	This study

pET28a-Δ35 *Hs*ppGalNAcT5	Truncated human ppGalNAcT5 (Δ1-35) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*ppGalNAcT6	Truncated human ppGalNAcT6 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ29 *Hs*ppGalNAcT7	Truncated human ppGalNAcT7 (Δ1-29) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ29 *Hs*ppGalNAcT8	Truncated human ppGalNAcT8 (Δ1-29) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*ppGalNAcT9	Truncated human ppGalNAcT9 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ31 *Hs*ppGalNAcT10	Truncated human ppGalNAcT10 (Δ1-31) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ43 *Hs*B3GALNT1	Truncated human B3GALNT1 (Δ1-43) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ25 *Hs*B4GALNT1	Truncated human B4GALNT1 (Δ1-25) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*-A-group	Human A-ABO with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ43 *Hs*A4GALT	Truncated human A4GALT (Δ1-43) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ26 *Hs*B3GalT1	Truncated human B3GalT1 (Δ1-26) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ45 *Hs*B3GalT2	Truncated human B3GalT2 (Δ1-45) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ44 *Hs*B4GalT1	Truncated human B4GalT1 (Δ1-44) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ36 *Hs*B4GalT2	Truncated human B4GalT2 (Δ1-36) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ31 *Hs*B4GalT3	Truncated human B4GalT3 (Δ1-31) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ38 *Hs*B4GalT4	Truncated human B4GalT4 (Δ1-38) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ35 *Hs*B4GalT5	Truncated human B4GalT5 (Δ1-35) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ35 *Hs*B4GalT6	Truncated human B4GalT6 (Δ1-35) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*-B-group	Human B-ABO with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ20 *Hs*UGT8	Truncated human UGT8 (Δ1-20) with C-terminal 6x-His tags in pET28a; Kan^R	This study

pET28a-Δ29 *Hs*C1GLT	Truncated human C1GLT (Δ1-29) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*COSMC	Human COSMC with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ23 *Hs*Alg1	Truncated human Alg1 (Δ1-23) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*Alg2	Human Alg2 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*Alg3	Human Alg3 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ40 *Hs*Alg11	Truncated human Alg11 (Δ1-40) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*Alg12	Human Alg12 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*Alg13	Human Alg13 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ24 *Hs*Alg14	Truncated human Alg14 (Δ1-24) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*DPM1	Human DPM1 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*PIGM	Human PIGM with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*PIGB	Human PIGB with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*PIGZ	Human PIGZ with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*Alg5	Truncated human Alg5 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*Alg6	Human Alg6 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*Alg8	Human Alg8 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*Alg10	Human Alg10 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*UGCG	Human UGCG with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ27 *Hs*B3GLCT	Truncated human B3GLCT (Δ1-27) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*GLYG	Human glycogenin with C-terminal 6x-His tags in pET28a; Kan^R	This study

pET28a-*Hs*POGLUT1	Human POGLUT1 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ29 *Hs*GnT1	Truncated human GnT1 (Δ1-29) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ29 *Hs*GnT2	Truncated human GnT2 (Δ1-29) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ23 *Hs*GnT3	Truncated human GnT3 (Δ1-23) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ27 *Hs*GnT4a	Truncated human GnT4a (Δ1-27) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ32 *Hs*GCNT1	Truncated human GCNT1 (Δ1-32) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ23 *Hs*GCNT2	Truncated human GCNT2 (Δ1-23) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*B3GNT2	Truncated human B3GnT2 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ31 *Hs*B3GNT6	Truncated human B3GnT6 (Δ1-31) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*PIGA	Human PIGA with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ25 *Hs*UGT1A1	Truncated human UGT1A1 (Δ1-25) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ36 *Hs*UGT1A3	Truncated human UGT1A3 (Δ1-36) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ36 *Hs*B4GAT1	Truncated human B4GAT1 (Δ1-36) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ28 *Hs*XXLT1	Truncated human XXLT1 (Δ1-28) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Cj*CstII	*C. jejuni* CstII with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Nm*Pst	*N. meningitidis* PolysiaT with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Cj*CgtB	*C. jejuni* CgtB with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hp*LgtB	*H. pylori* GalT with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Nm*LgtB	*N. meningitidis* GalT with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Ng*LgtB	*N. gonorrhea* GalT with C-terminal 6x-His tags in pET28a; Kan^R	This study

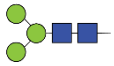
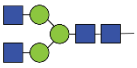
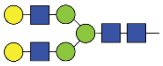
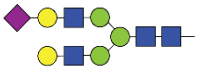
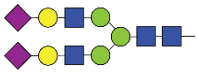
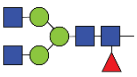
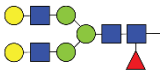
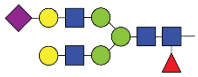
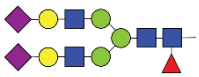
pET28a-*Ec*WbgL	*E. coli* WbgL with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Ec*WecA	*E. coli* WecA with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Lp*SetA	*L. pneumophila* SetA with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Nm*SynE	*N. meningitidis* SynE with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ34 *Sc*Alg1	Truncated *S. cerevisae* Alg1 (Δ1-34) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Sc*Alg2	*S. cerevisae* Alg2 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ45 *Sc*Alg11	Truncated *S. cerevisae* Alg11 (Δ1-45) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ30 *Nt*GnTI	Truncated *N. tabacum* GnTI (Δ1-30) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ35 *Nt*GnTII	Truncated *N. tabacum* GnTII (Δ1-35) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ22 *Bt*GGTA1	Truncated *B. 308aurus* GGTA1 (Δ1-35) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ60 *Mm*GGTA1	Truncated *M. musculus* GGTA1 (Δ1-60) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ58 *Rn*GGTA1	Truncated *R. norvergicus* GGTA1 (Δ1-58) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-Δ44 *Bt*B4GalT1	Truncated *B. 308aurus* B4GalT1 (Δ1-35) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*CDK4	Human cyclin-dependent kinase 4 with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*CDKN2A	Human cyclin-dependent kinase inhibitor 2A with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*EGFR_TK	Truncated human EGFR (P694-G1022) with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*FOS	Human proto-oncogene c-Fos with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*GATA2	Human endothelial transcription factor GATA with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*JUN	Human transcription factor JUN, and ApoAI*-6xHis with C-terminal 6x-His tags in pET28a; Kan^R	This study

pET28a-*Hs*MMP1	Human Interstitial collagenase, and ApoAI*-6xHis with C-terminal 6x-His tags in pET28a; Kan^R	This study
pET28a-*Hs*Insulin	Human insulin with C-terminal 6x-His tags in pET28a; Kan^R	This study
pcDNA3 Δ26 *Hs*ST6Gal1	Truncated human ST6Gal1 (Δ1-26) with C-terminal 6x-His tags in pcDNA3; Amp^R	This study
pYS338-Δ26 *Hs*ST6Gal1	Truncated human ST6Gal1 (Δ1-26) with C-terminal 6x-His tags in pYS338; Amp^R	This study
pJL1- Δ26 *Hs*ST6Gal1	Truncated human ST6Gal1 (Δ1-26) with C-terminal 6x-His tags in pJL1; Kan^R	This study

References

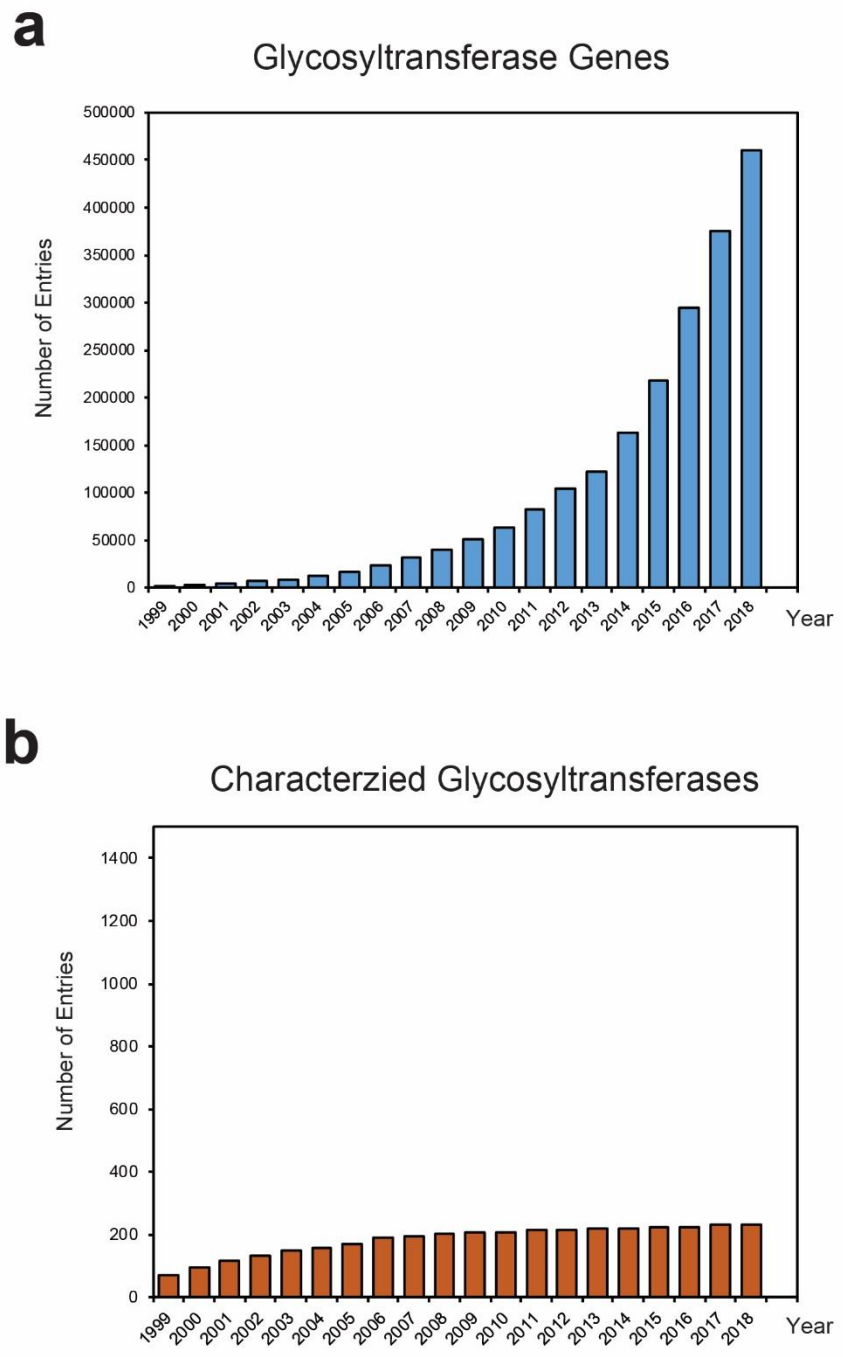
1. Mizrachi, D. *et al*. Making water-soluble integral membrane proteins in vivo using an amphipathic protein fusion strategy. *Nat Commun* 6, 6826 (2015).
2. Stark, J.C. *et al*. BioBits Bright: A fluorescent synthetic biology education kit. *Sci Adv 4*, eaat5107 (2018).

Supplementary Table 8-2: *N*-glycan structures produced in this study

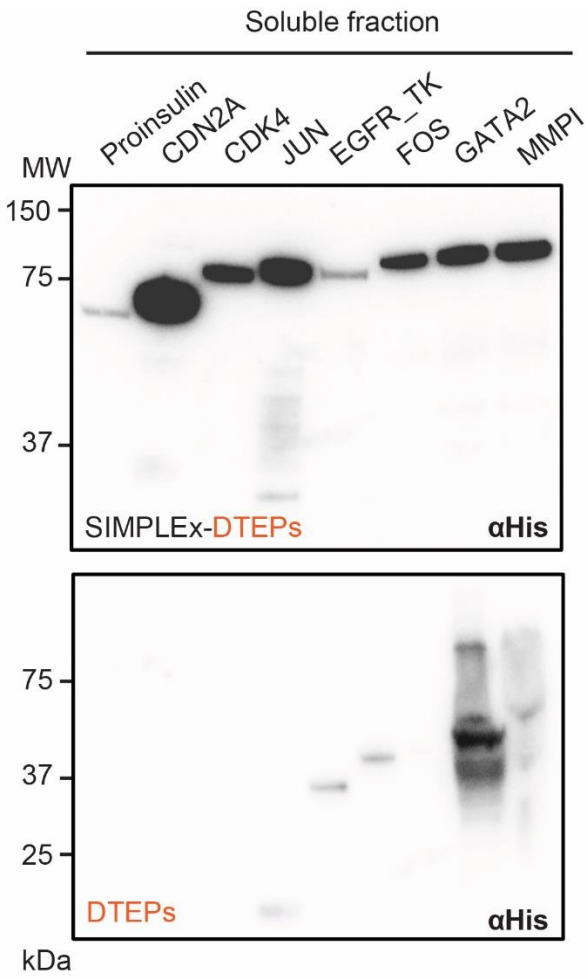
Glycan Number	Symbolic Representation	Oxford Notation	Structure
1		M3	Manα1-6(Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
2		A1	Manα1-6(GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAc β1-4GlcNAc
3		A2	GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Man β1-4GlcNAcβ1-4GlcNAc
4		A2G2	Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-2 Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAc
5		A2G2S1	Galβ1-4GlcNAcβ1-2Manα1-6(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4 GlcNAc
6		A2G2S2	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6 (Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3) Manβ1-4GlcNAcβ1-4GlcNAc
7		FA2	GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3) Manβ1-4GlcNAcβ1-4(Fucα1-6)GlcNAc
8		FA2G2	Galβ1-4GlcNAcβ1-2Manα1-6(Galβ1-4GlcNAcβ1-2 Manα1-3)Manβ1-4GlcNAcβ1-4(Fucα1-6)GlcNAc
9		FA2G2S1	Galβ1-4GlcNAcβ1-2Manα1-6(Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4 (Fucα1-6)GlcNAc
10		FA2G2S2	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-6 (Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-3) Manβ1-4(Fucα1-6)GlcNAcβ1-4GlcNAc

Supplementary Table 8-3: Antibodies and binding molecule used in the SIMPLEx study

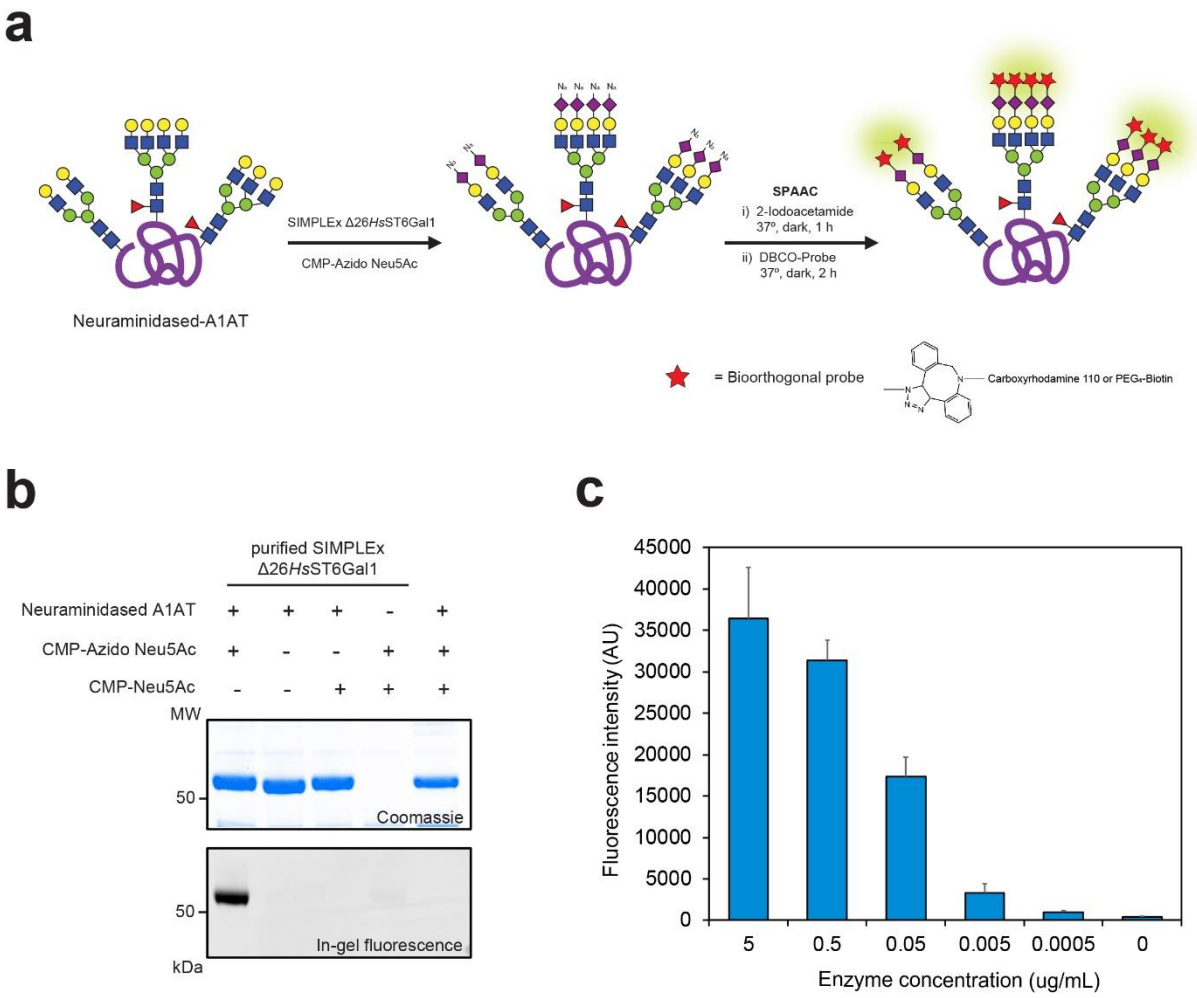
Target	Source	Cat. No.	Dilution
Rabbit pAb to 6x-His epitopeTag	ThermoFisher	PA1-983B	1:1000
Goat anti-Rabbit IgG H&L (HRP)	abcam	ab6721	1:5000
Mouse anti-GAPDH	Calbiochem	CB1001	1:5000
Rabbit anti-GroEL	Sigma	G6532	1:30000
Rabbit anti-alpha tubulin	abcam	ab184970	1:30000
ExtrAvidin®–Peroxidase	Sigma	E2886	1:5000



Supplementary Figure 8-1. Glycosyltransferase database as retrieved from the Carbohydrate active enzyme (CAZy) database. Accumulative numbers of (**a**) glycosyltransferase gene and (**b**) characterized glycosyltransferases in the CAZy database as a function of time from 1999 to 2018. Data was provided by CAZy database on November 2018.

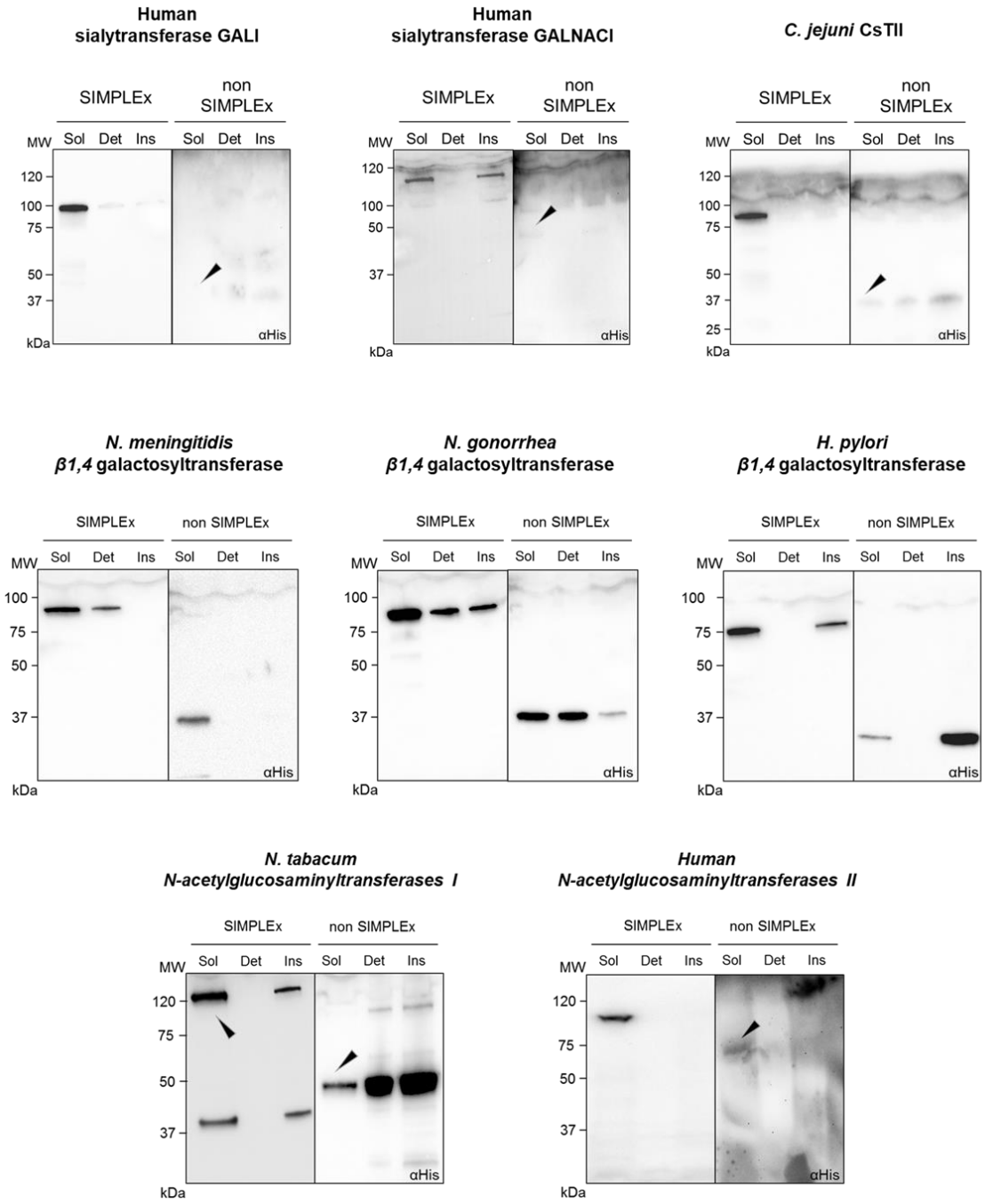


Supplementary Figure 8-2. SIMPLEx fusion rescues soluble expression of diverse difficult-to-express proteins (DTEPs). Immunoblot analysis of the soluble fraction prepared from *E. coli* cells expressing human-derived DTEPs as a SIMPLEx fusion (Top panel; SIMPLEx-DTEPs) or wild-type protein (Bottom panel; DTEPs). *E. coli* BL21(DE3) strain was used to express transcription factors (GATA2, JUN, and FOS), cyclin-dependent kinase and inhibitor (CDK4, CDN2A), while SHuffle T7 Express *lysY* strain was used to express collagenase (MMP1), epidermal growth factor receptor tyrosine kinase (EGFR-TK), and cytokine (proinsulin, INS). Blots were probed with anti-hexa-histidine (6xHis) antibody. Results are representative of at least three biological replicates. Molecular weight (MW) markers are shown on the left.

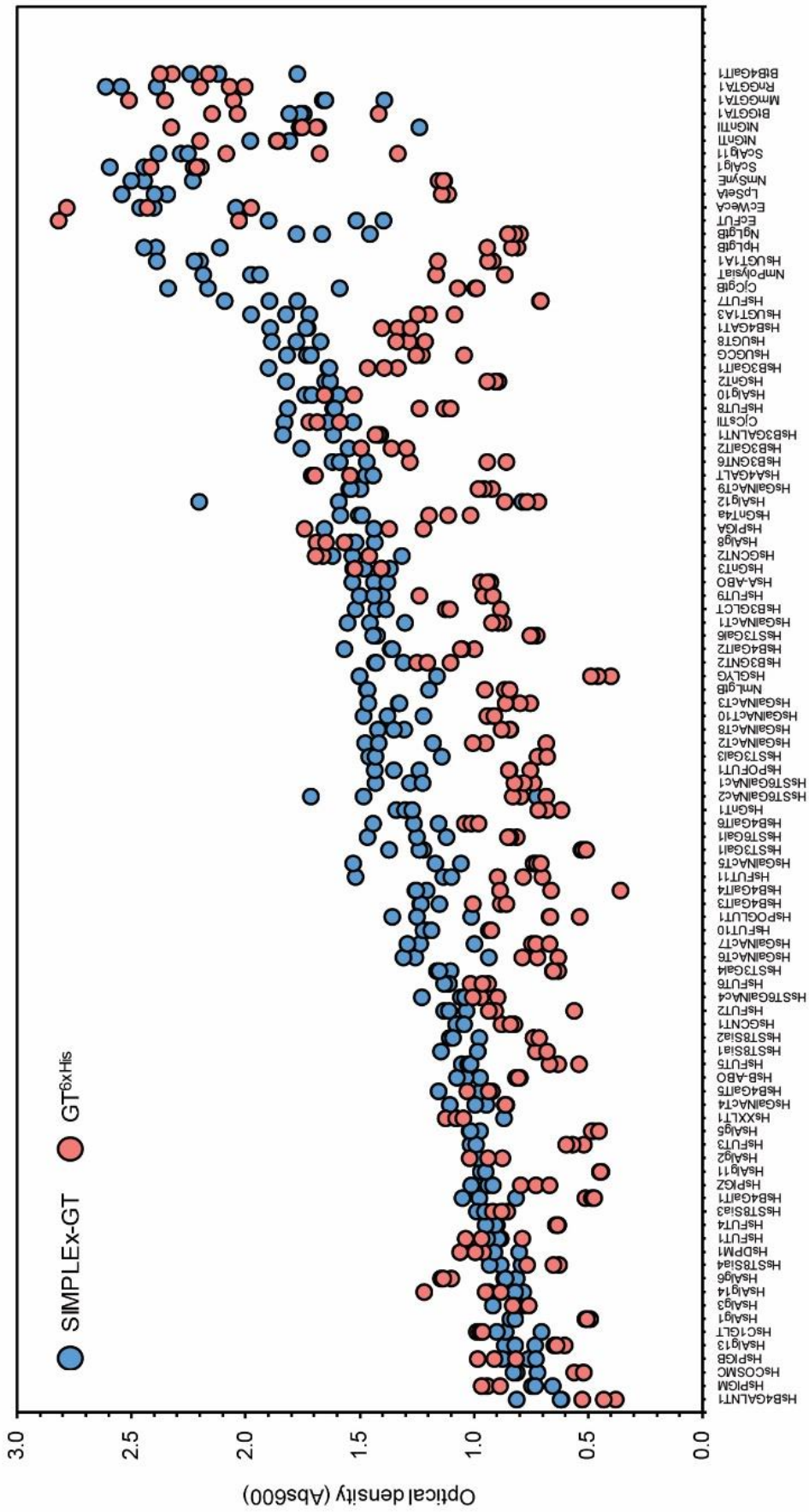


Supplementary Figure 8-3. Bioorthogonal chemistry-based *in vitro* sialytransferase assay characterizes SIMPLEx-*Hs*Δ26ST6Gal1. (**a**) Schematic of *in vitro* sialyltransferase assay. Purified human alpha-1 antitrypsin (A1AT), which was first treated with α2-3,6,8,9 Neuraminidase to remove native sialic acid, was used as glycoprotein precursor. SIMPLEx-*Hs*Δ26ST6Gal1 was then used to transfer azido-Neu5Ac from CMP-azido-Neu5Ac onto terminal galactose of the *N*-glycans on A1AT. Depicted glycans are the representative glycoforms of the native human A1AT. The azido (N_3-) functional group on Neu5Ac provided a chemical handle on the glycoprotein that could be readily conjugated with carboxyrhodamine 110 fluorophore or PEG_4-biotin reporters *via* strain-promoted azide-alkyne cycloaddition (SPAAC)

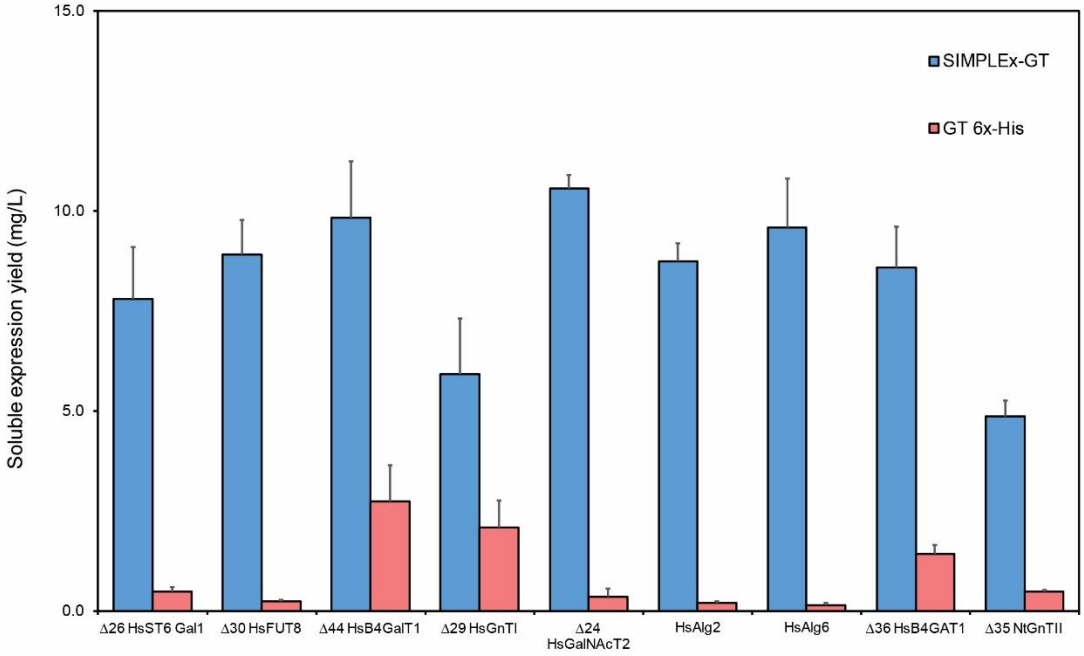
using dibenzocyclooctyne group (DBCO) as reactive alkyne. (**b**) Representative result of the *in vitro* sialyltransferase assay. Following a labeling of edited-glycans on A1AT glycoprotein with carboxyrhodamine 110, reaction mixture was separated on SDS-PAGE gel. Fluorescence signal of the labeled glycoprotein was measured at 501/523 nm $\lambda_{ex}/\lambda_{em}$ using fluorescent imager. Gel stained with Coomassie dyes was used as a loading control. Results are representative of at least three biological replicates. Molecular weight (MW) markers are shown on the left. (**c**) Fluorescent intensity measured *in vitro* sialytransferase activity of the commercial, purified Δ26*Hs*ST6Gal1 derived from NS0 Mouse myeloma cell line as a function of enzyme concentration. Reaction without enzyme was used as negative control. Activity data are the mean of three replicates and error bars represent standard error of the mean.



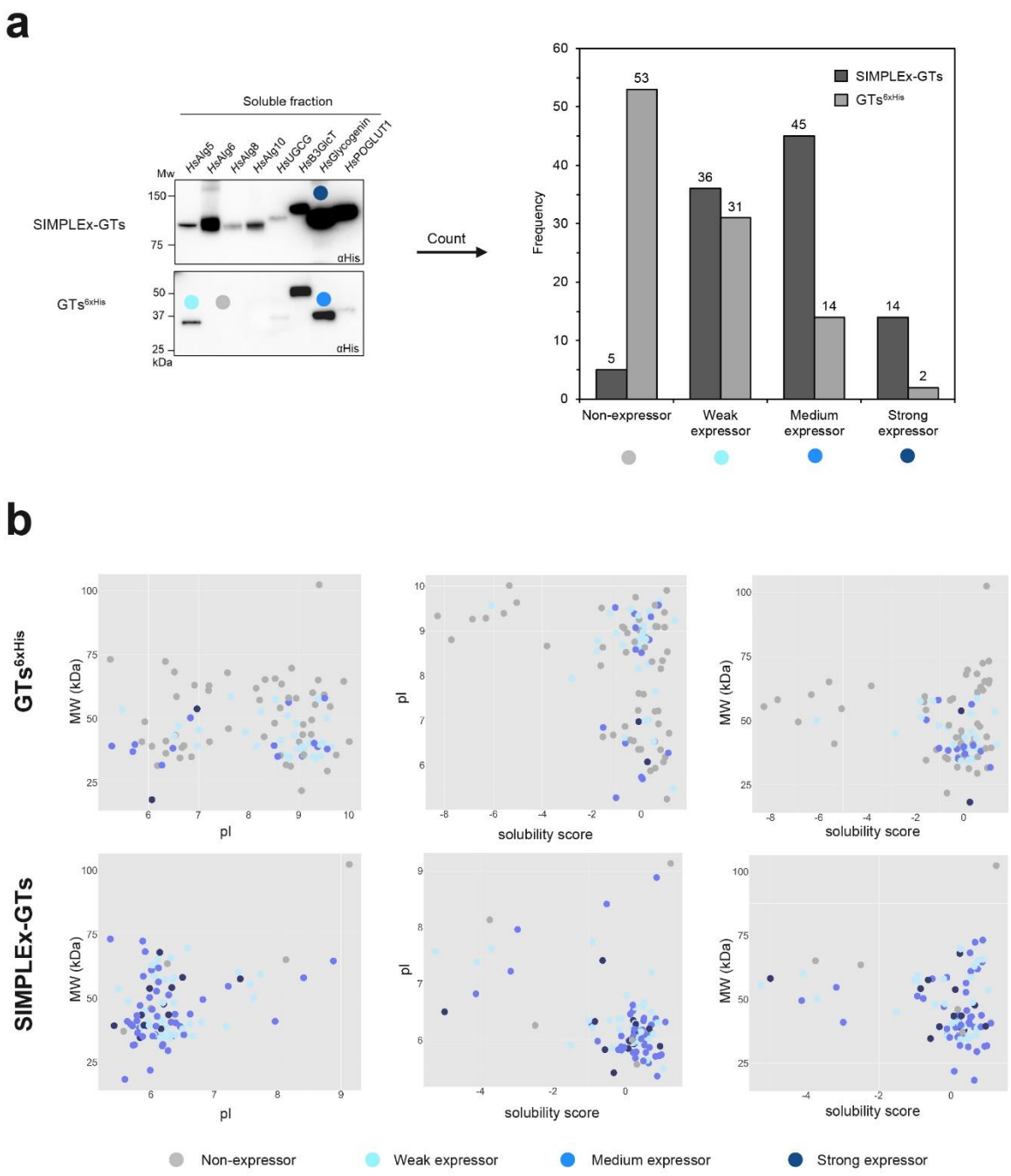
Supplementary Figure 8-4. SIMPLEx fusion promotes solubility and expression level of diverse GTs. Western blot analysis of the soluble (Sol), detergent-solubilized (Det), and insoluble (Ins) fractions prepared from *E. coli* cells carrying plasmid pET28a(+) encoding either SIMPLEx or WT versions of the GTs. The name of each GT is provided on top of each immunoblot.


SIMPLEx-GT
GT6xHis
Optical density (Abs600)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
HsB4GALNT1
HsPIGM
HsCOSMC
HsPIGB
HsA1g13
HsC1GLT
HsA1g1
HsA1g3
HsA1g14
HsA1g6
HsST8Sia4
HsDPM1
HsFUT1
HsFUT4
HsST8Sia3
HsB4GalT1
HsPIGZ
HsA1g11
HsA1g2
HsFUT3
HsA1g5
HsXXLT1
HsGalNAcT4
HsB4GalT5
HsB-ABO
HsFUT5
HsST8Sia1
HsST8Sia2
HsGCNT1
HsFUT2
HsST6GalNAc4
HsFUT6
HsST3Gal4
HsGalNAcT6
HsGalNAcT7
HsFUT10
HsPOGLUT1
HsB4GalT3
HsB4GalT4
HsFUT11
HsGalNAcT5
HsST3Gal1
HsST6Gal1
HsB4GalT6
HsGnT1
HsST6GalNAc2
HsST6GalNAc1
HsPOFUT1
HsST3Gal3
HsGalNAcT2
HsGalNAcT8
HsGalNAcT10
HsGalNAcT3
NmLgtB
HsGLYG
HsB3GNT2
HsB4GalT2
HsST3Gal6
HsGalNAcT1
HsB3GLCT
HsFUT9
HsA-ABO
HsGnT3
HsGCNT2
HsA1g8
HsPIGA
HsGnT4a
HsA1g12
HsGalNAcT9
HsA4GALT
HsB3GNT6
HsB3GalT2
HsB3GALNT1
CjCst1I
HsFUT8
HsA1g10
HsGnT2
HsB3GalT1
HsGCG
HsUGT8
HsB4GAT1
HsUGT1A3
HsFUT7
CjCgtB
NmPolysiaT
HsUGT1A1
HpLgtB
NgLgtB
EcFUT
EcWecA
LpSelA
NmSynE
ScAlg1
ScAlg11
NtGnTI
NtGnTII
BtGGTA1
MmGGTA1
RnGGTA1
BtB4GalT1

Supplementary Figure 8-5. Cell density of *E. coli* cultures expressing glycosyltransferase enzymes. Three representative data points for cultures expressing SIMPLEx-GTs were labeled in blue, while those of cultures expressing GTs^{6xHis} were labeled in red. All data points for each GT enzymes were plotted on the same axis. Final cell density was recorded as Abs_{600} values taken after 16-18 h of overnight growth. Results are representative of at least three biological replicates.

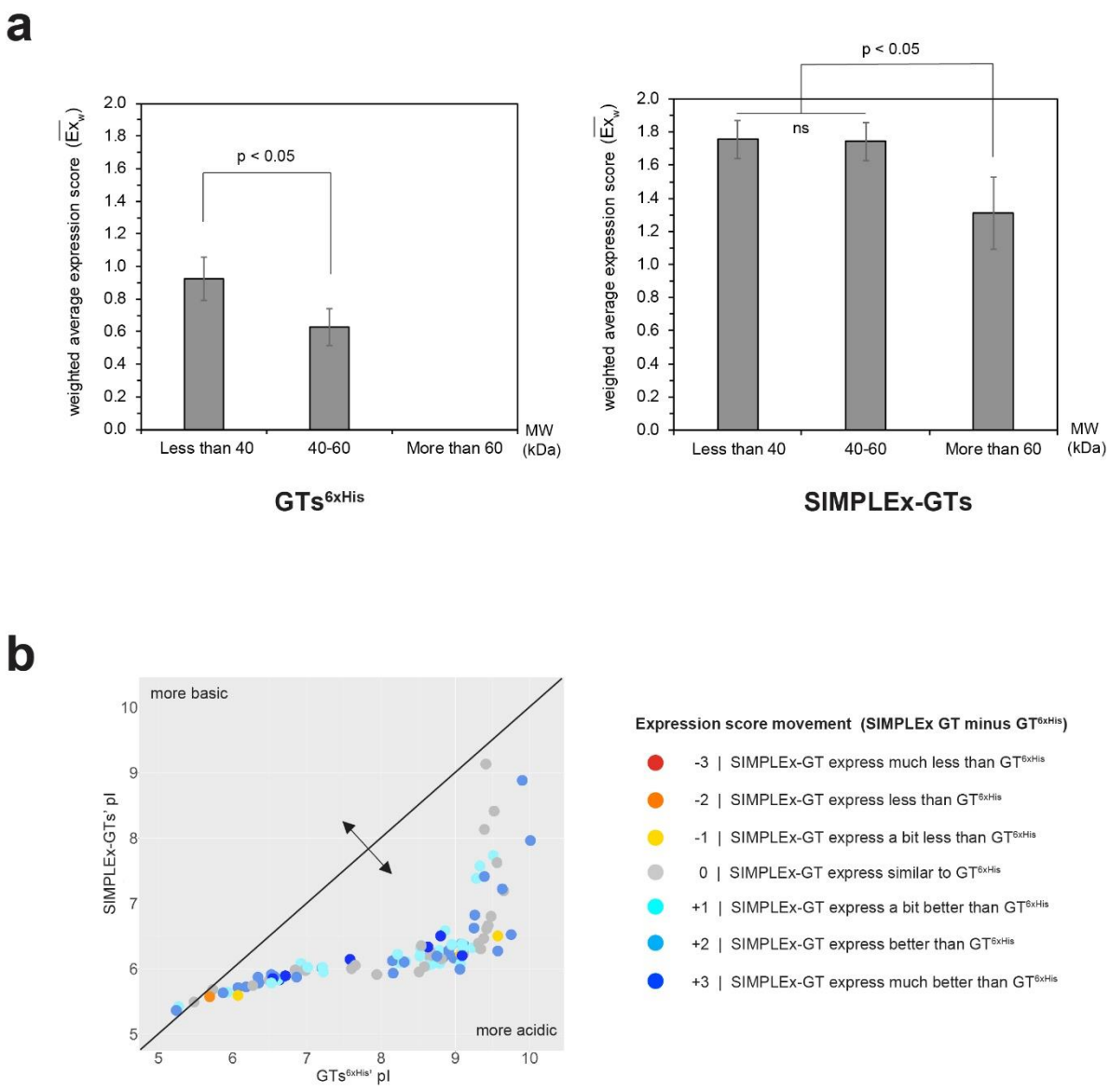


Supplementary Figure 8-6. Yield quantification for different SIMPLEx-GTs. Protein soluble yields (mg/L) were determined by purification of proteins using Ni-NTA chromatography from 1-L cultures of *E. coli* carrying each SIMPLEx-GT construct as indicated. Cultures of *E. coli* expressing GT^{6xHis} construct served as control. Yield values are representative of at least three biological replicates and error bars represent standard error of the mean.



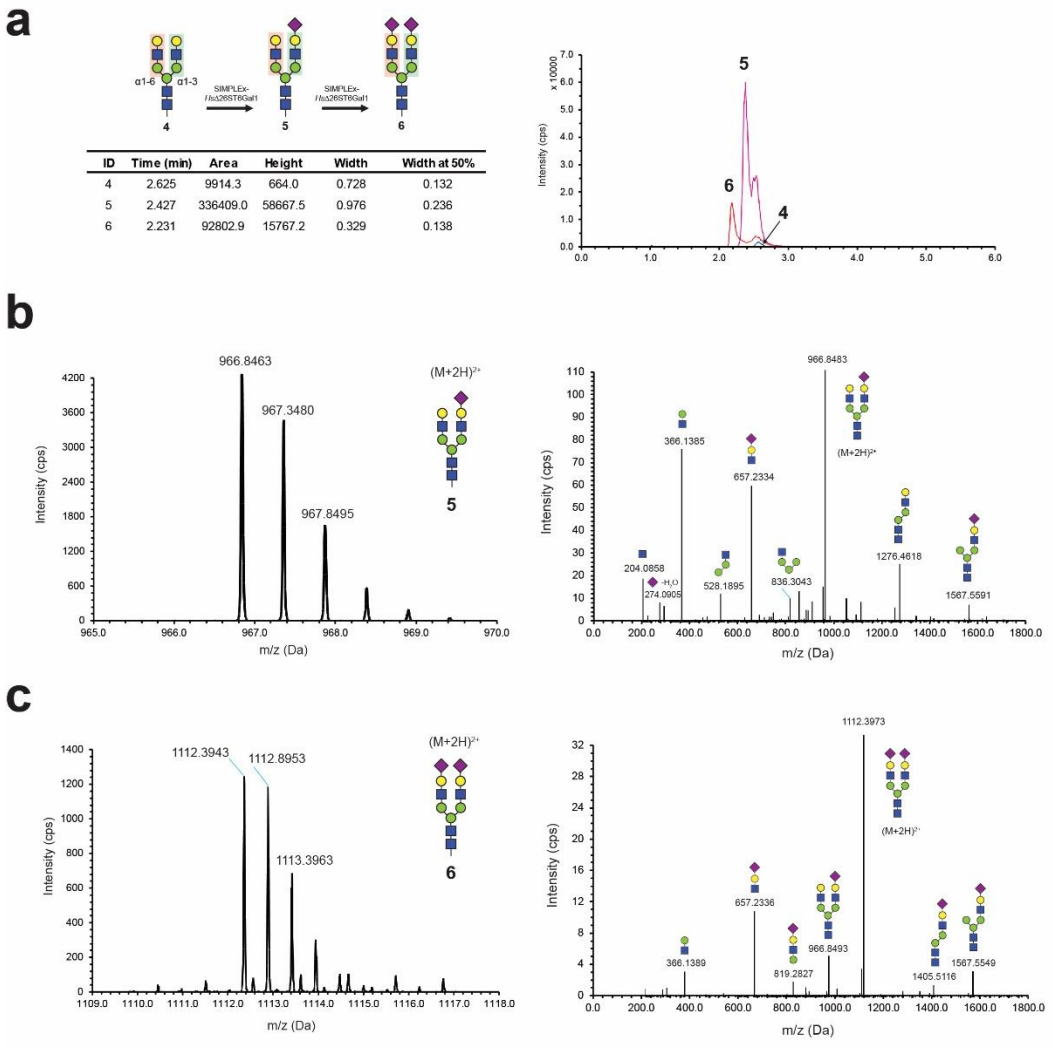
Supplementary Figure 8-7. Expression profile of SIMPLEx-GTs identifies physicochemical properties that correlates with successful expression. (**a**) Representative blots demonstrate the expression score assignment. Upon inspecting band intensity of its immunoblot, each GT, as native enzyme or within SIMPLEx

format, was categorized as non-expressor (score 0; grey circle), weak expressor (score 1; cyan circle), medium expressor (score 2; light blue circle), and strong expressor (score 3; dark blue circle). Bar graph summarizes a total number of members for each category, comparing between SIMPLEx-GTs and GTs^{6xHis}. (**b**) Scatter plots between permutations of the following physicochemical properties: (i) protein's molecular weight excluding added mass from ΔspMBP and ApoAI* domains; (ii) protein's isoelectric point; and (iii) protein's solubility score as calculated by Protein-Sol server. Data point was labeled according to its expressor category. Plots were generated using R version 3.4.2 software.

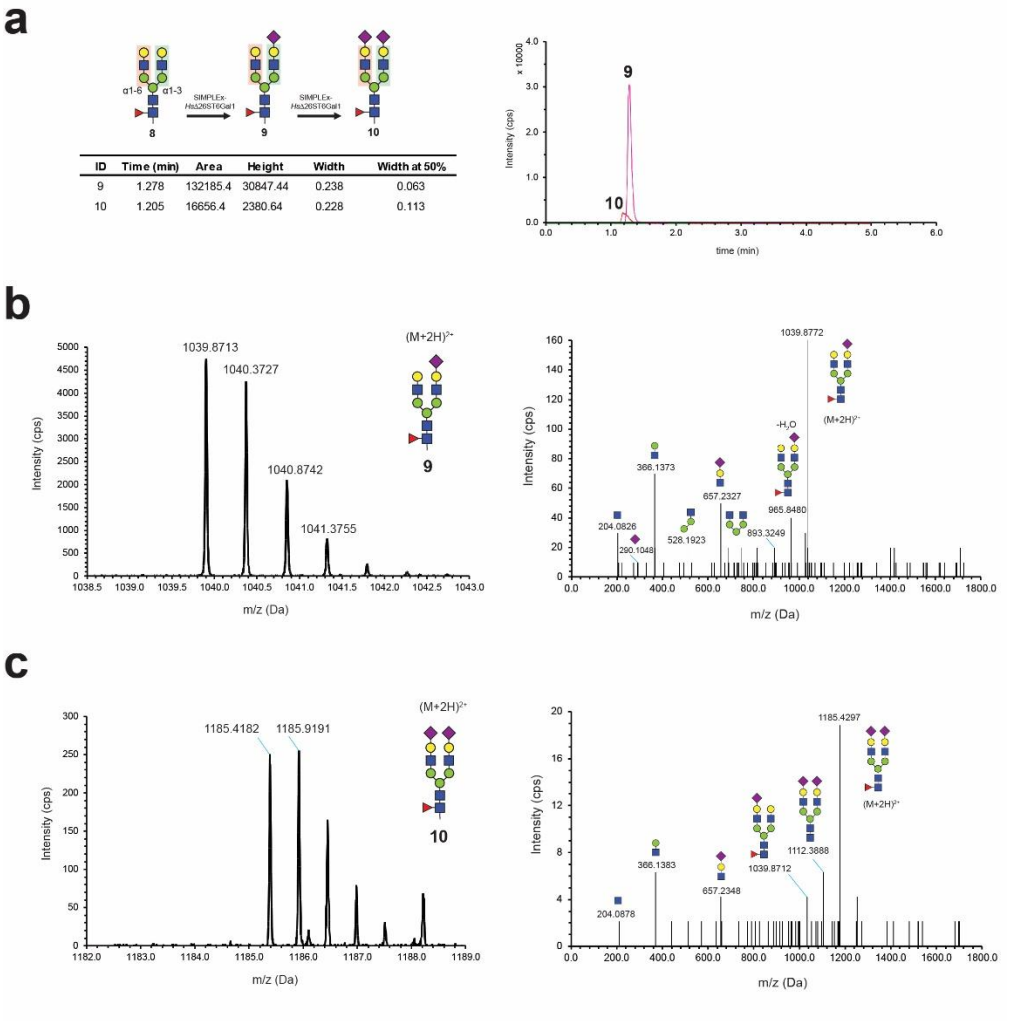


Supplementary Figure 8-8. Relationship between protein physicochemical properties and soluble expression. (**a**) Bar graphs comparing weighted average expression score ($\overline{Ex}_w$) between the small, medium, and large-size GTs expressed as GT^{6xHis} or SIMPLEx-GT. Criteria for protein size classification was indicated on the x-axis. Noted that added molecular weight from ΔspMBP and ApoAI* domains in the SIMPLEx construct were excluded during size classification. Error bar represented standard error of mean. Statistical significance was determined by Welch's t-test and

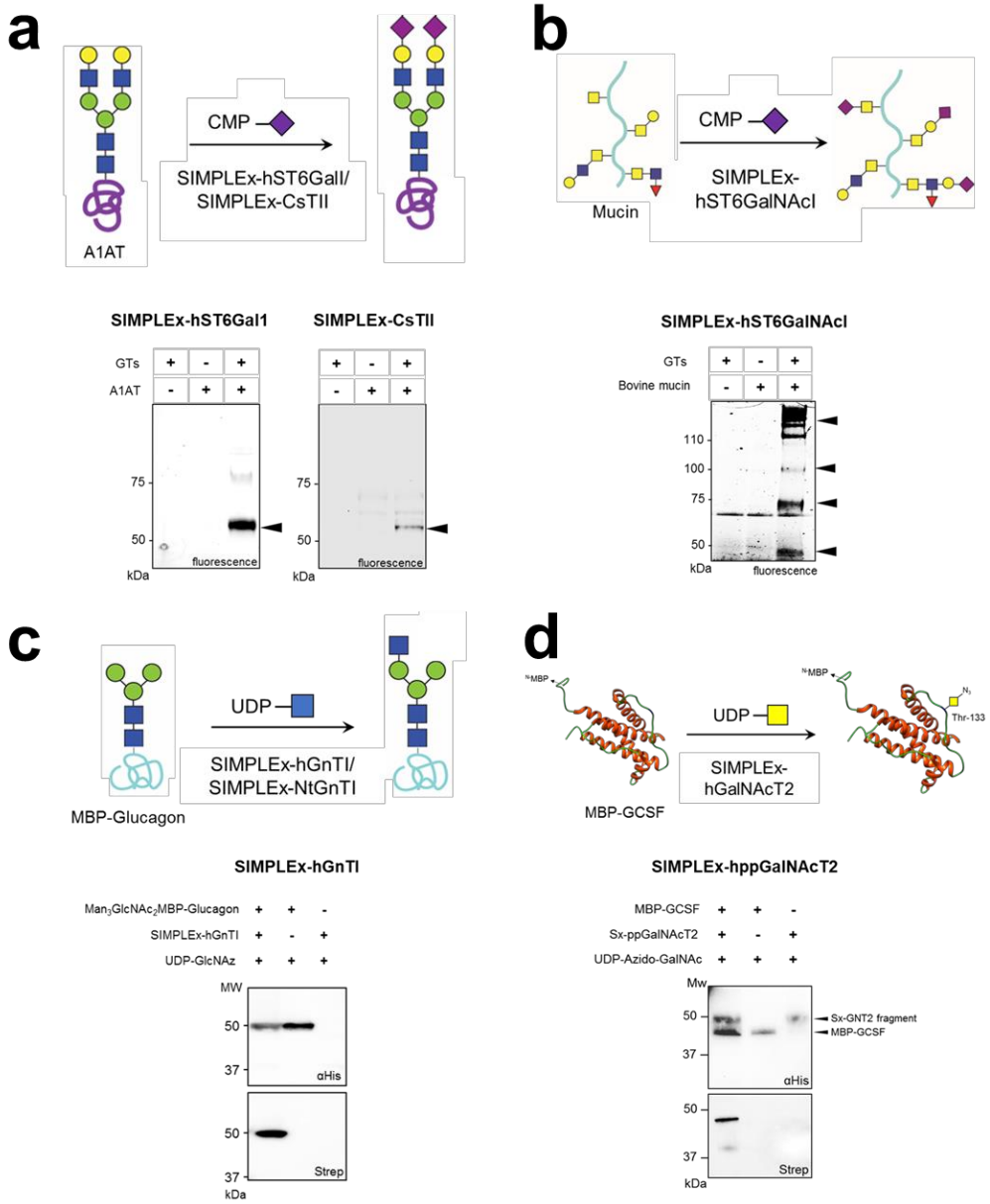
the p-values of less than 0.05 were considered significant. (**b**) Scatter plot between protein's isoelectric point (pI) of the wild-type and SIMPLEx-fused GT. Data point was labeled according to its expression score movement as indicated in the legend. Diagonal line (black color) is a theoretical line where there is no change in protein's pI before and after SIMPLEx fusion. Arrows represent the change toward more basic or more acidic proteins, as indicated, following SIMPLEx fusion. Plots were generated using R version 3.4.2 software.



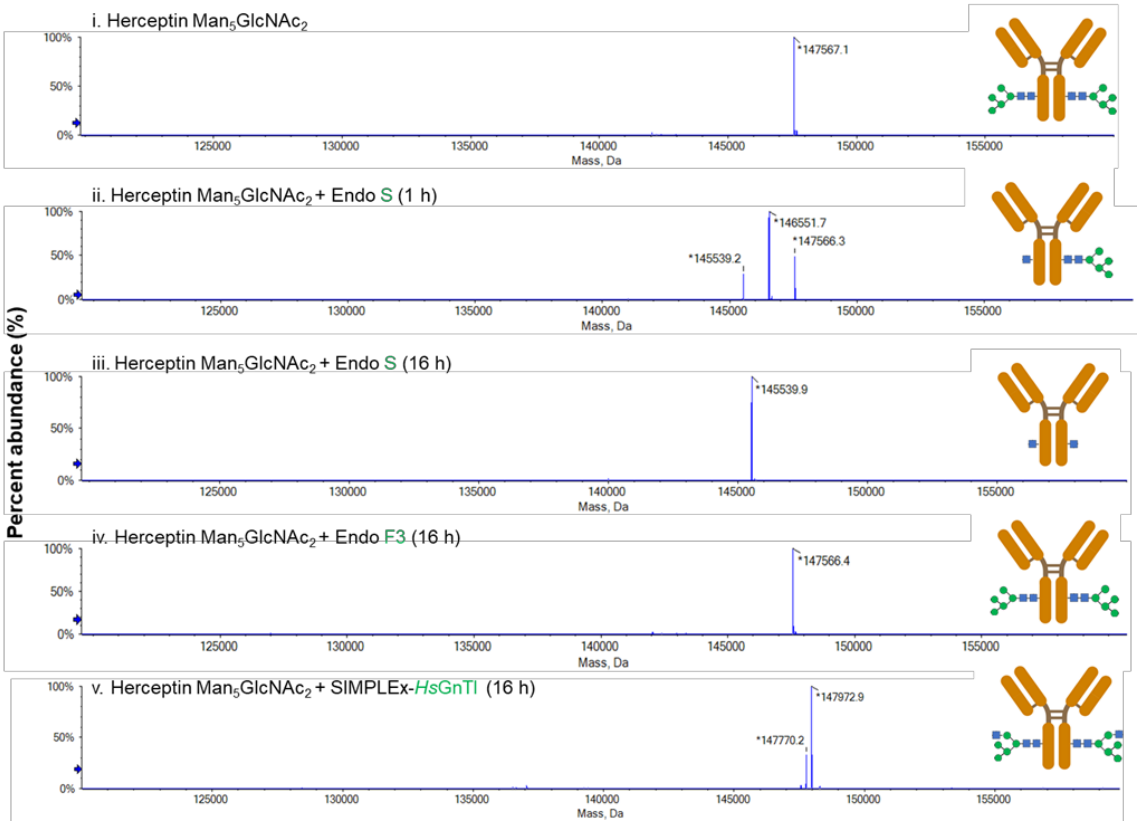
Supplementary Figure 8-9. LC-MS/MS analysis of sialylated *N*-glycans from the cell-free glycan remodeling reaction. (**a**) HILIC-LC-MS chromatogram of the cell-free reaction installing sialic acids on glycan **4** using SIMPLEx-*Hs*Δ26ST6Gal1. Schematic reaction for the stepwise conversion of glycan **4** to **5** and **6** was provided. A1–3Man branch, which is a preferred substrate for *Hs*ST6Gal1, was highlighted in green and the less-preferred *α*1–6Man branch was highlighted in red. (**b**) MS and MS/MS spectrum of the doubly charged glycan at *m*/*z* 966.8463. Positive ion MS/MS fragmentation pattern confirmed the identity of A2G2S1 product. (**c**) MS and MS/MS spectrum of the doubly charged glycan at *m*/*z* 1112.3943. Positive ion MS/MS fragmentation pattern confirmed the identity of A2G2S2 product.



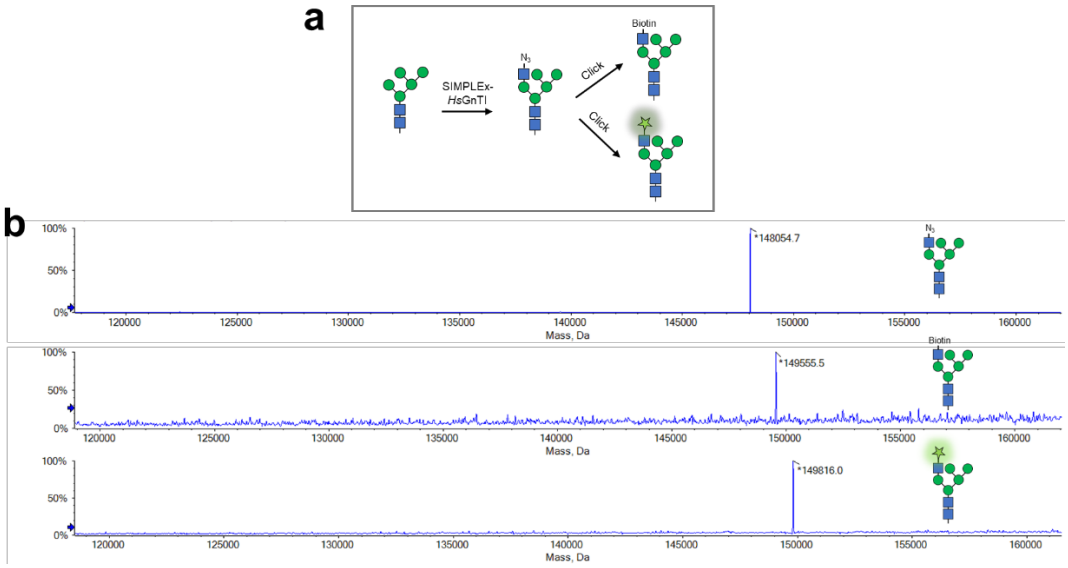
Supplementary Figure 8-10. LC-MS/MS analysis of sialylated, core-fucosylated *N*-glycans from the cell-free glycan remodeling reaction. (**a**) HILIC-LC-MS chromatogram of the cell-free reaction installing sialic acids on glycan **8** using SIMPLEx-*Hs*Δ26ST6Gal1. Schematic reaction for the stepwise conversion of glycan **8** to **9** and **10** was provided. A1–3Man branch, which is a preferred substrate for *Hs*ST6Gal1, was highlighted in green and the less-preferred *α*1–6Man branch was highlighted in red. (**b**) MS and MS/MS spectrum of the doubly charged glycan at *m*/*z* 1039.8713. Positive ion MS/MS fragmentation pattern confirmed the identity of A2G2S1 product. (**c**) MS and MS/MS spectrum of the doubly charged glycan at *m*/*z* 1185.4182. Positive ion MS/MS fragmentation pattern confirmed the identity of A2G2S2 product.



Supplementary Figure 8-11. SIMPLEx-GTs remodel glycan on therapeutic glycoproteins. Cell-free reaction modifies: **(a)** *N*-glycan on A1A1 using SIMPLEx-ST6Gal1; **(b)** *O*-glycan on mucin using SIMPLEx-ST6GalNAc1; **(c)** *N*-glycan on glucagon peptide using SIMPLEx-GnTI; and (d) granulocyte colony-stimulating factor (GCSF) using SIMPLEx-GalNAcT2. All reactions were followed using biorthogonal click reaction.



Supplementary Figure 8-12. Glycosidase sensitivity assay coupled with LC-MS analysis confirms *N*-glycan identity on Trastuzumab. Intact protein analysis mode in the LC-MS was used to detected (i) full-length Trastuzumab bearing a single glycoform. (ii) Upon treating this IgG with Endoglycosidase S, new mass peaks were detected, and these peaks were corresponded to the loss of one or two Man5GlcNAc molecules, respectively. (iii) A longer incubation yielded single mass peak corresponding to the IgG losing two Man5GlcNAc molecules. (iv) Trastuzumab bearing Man5 glycan was insensitive to the Endo F3, an endoglycosidase that cleaves complex *N*-glycan. (v) Trastuzumab bearing Man5 glycan can be modified by GnTI enzyme which recognizes Man5 glycan as its natural substrate.



Supplementary Figure 8-13. SIMPLEx glycoenzyme-mediated bioorthogonal conjugation of useful chemical moieties on Trastuzumab. (a) schematic representation of cell-free reactions modify full-length antibody with unnatural molecules. (b) LC-MS analysis of Trastuzumab bearing either azido-N-acetylglucosamine, biotin, or fluorescent reporter.

Supplementary Dataset 8-1. Amino acid sequences of glycosyltransferase in the SIMPLEx-GTs expression library

Name	full length sequence (FASTA)
HsFUT1	>FUT1 MWLRSHRQLCLAFLLVCVLSVIFFLHIHQDSFPHGLGLSILCPDRRLVTPPVAIFCLPGT AMGPNASSSCPQHPASLSGTWTVYPNGRFGNQMGQYATLLALAQLNGRRAFILPAMHAAL APVFRITLPVLAPEVDSRTPWRELQLHDWMSEEYADLRDPFLKLSGFPCSWTFFHHLREQ IRREFTLHDHLREEAQSVLGQLRLGRTGDRPRTFVGVHVRRGDYLQVMPQRWKGVVGDSA YLRQAMDWFRARHEAPVFVVTSNGMEWCKENIDTSQGDVTFAGDGQEATPWKDFALLTQC NHTIMTIGTFGFWAAYLAGGDTVYLANFTLPDSEFLKIFKPEAAFLPEWVGINADLSPLW TLAKP
HsFUT2	>FUT2 MLVVQMPFSFPMAHFILFVFTVSTIFHVQQRLAKIQAMWELPVQIPVLASTSKALGPSQL RGMWTINAIGRLGNQMGEYATLYALAKMNGRPAFIPAQMHSTLAPIFRITLPVLHSATAS RIPWQNYHLNDWMEEEYRHIPGEYVRFTGYPCSWTFYHHLRQEILQEFTLHDHVREEAQK FLRGLQVNGSRPGTFVGVHVRRGDYVHVMPKVWKGVVADRRYLQQALDWFRARYSSLIFV VTSNGMAWCRENIDTSHGDVVFAGDGIEGSPAKDFALLTQCNHTIMTIGTFGIWAAYLTG GDTIYLANYTLPDSPFLKIFKPEAAFLPEWTGIAADLSPLLKH
HsFUT3	>FUT3 MDPLGAAKPQWPWRRCLAALLFQLLVAVCFFSYLRVSRDDATGSPRAPSGSSRQDTTPTR PTLLILLWTWPFHIPVALSRCSEMVPGTADCHITADRKVYPQADTVIVHHWDIMSNPKSR LPPSPRPQGQRWIWFNLEPPPNCQHLEALDRYFNLTMSYRSDSDIFTPYGWLEPWSGQPA HPPLNLSAKTELVAWAVSNWKPDSARVRYYQSLQAHLKVDVYGRSHKPLPKGTMMETLSR YKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRSNYERFLPPDAFIHVDDFQSPK DLARYLQELDKDHARYLSYFRWRETLRPRSFSWALDFCKACWKLQQESRYQTVRSIAAWF T
HsFUT4	>FUT4 MRRLWGAARKPSGAGWEKEWAEAPQEAPGAWSGRLGPGRSGRKGRAVPGWASWPAHLAL AARPARHLGGAGQGPRPLHSGTAPFHSRASGERQRRLEPQLQHESRCRSSTPADAWRAEAA LPVRAMGAPWGSPTAAAGGRRGWRRGRGLPWTVCVLAAAGLTCTALITYACWGQLPPLPW ASPTPSRPVGVLLWWEPFGGRDSAPRPPPDCRLRFNISGCRLLTDRASYGEAQAVLFHHR DLVKGPPDWPPPWGIQAHTAEEVDLRVLDYEEAAAAAAEALATSSPRPPGQRWVWMNFESP SHSPGLRSLASNLFNWTLSYRADSDVFVPYGYLYPRSHPGDPPSGLAPPLSRKQGLVAWV VSHWDERQARVRYYHQLSQHVTVDVFGRGGPGQPVPEIGLLHTVARYKFYLAFENSQHLD YITEKLWRNALLAGAVPVVLGPDRANYERFVPRGAFIHVDDFPSASSLASYLLFLDRNPA VYRRYFHWRRSYAVHITSFWDEPWCRVCQAVQRAGDRPKSIRNLASWFER
HsFUT5	>FUT5 MDPLGPAKPQWLWRRCLAGLLFQLLVAVCFFSYLRVSRDDATGSPRPGLMAVEPVTGAPN GSRCQDSMATPAHPTLLILLWTWPFNTPVALPRCSEMVPGAADCNITADSSVYPQADAVI VHHWDIMYNPSANLPPPTRPQGQRWIWFSMESPSNCRHLEALDGYFNLTMSYRSDSDIFT PYGWLEPWSGQPAHPPLNLSAKTELVAWAVSNWKPDSARVRYYQSLQAHLKVDVYGRSHK PLPKGTMMETLSRYKFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRSNYERFLPP DAFIHVDDFQSPKDLARYLQELDKDHARYLSYFRWRETLRPRSFSWALAFCKACWKLQQE SRYQTVRSIAAWFT
HsFUT6	>FUT6 MDPLGPAKPQWSWRCCLTTLLFQLLMAVCFFSYLRVSQDDPTVYPNGSRFPDSTGTPAHS IPLILLWTWPFNKPIALPRCSEMVPGTADCNITADRKVYPQADAVIVHHREVMYNPSAQL PRSPRRQGQRWIWFSMESPSHCWQLKAMDGYFNLTMSYRSDSDIFTPYGWLEPWSGQPAH PPLNLSAKTELVAWAVSNWGPNSARVRYYQSLQAHLKVDVYGRSHKPLPQGTMMETLSRY KFYLAFENSLHPDYITEKLWRNALEAWAVPVVLGPSRSNYERFLPPDAFIHVDDFQSPKD LARYLQELDKDHARYLSYFRWRETLRPRSFSWALAFCKACWKLQEESRYQTRGIAAWFT
HsFUT7	>FUT7 MNNAGHGPTRRLRGLGVLAGVALLAALWLLWLLGSAPRGTPAPQPTITILVWHWPFTDQP PELPSDTCTRYGIARCHLSANRSLLASADAVVFHHRELQTRRSHLPLAQRPRGQPWVWAS

	MESPSHTHGLSHLRGIFNWVLSYRRDSDIFVPYGRLEPHWGPSPPLPAKSRVAAWVVSNF QERQLRARLYRQLAPHLRVDVFGRANGRPLCASCLVPTVAQYRFYLSFENSQHRDYITEK FWRNALVAGTVPVVLGPPRATYEAFVPADAFVHVDDFGSARELAAFLTGMNESRYQRFFA WRDRLRVRLFTDWRERFCAICDRYPHLPRSQVYEDLEGWFQA
HsFUT8	>FUT8 MRPWTGSWRWIMLILFAWGTLLFYIGGHLVRDNDHPDHSSRELSKILAKLERLKQQNEDL RRMAESLRIPEGPIDQGPAIGRVRVLEEQLVKAKEQIENYKKQTRNGLGKDHEILRRRIE NGAKELWFFLQSELKKLKNLEGNELQRHADEFLLDLGHHERSIMTDLYYLSQTDGAGDWR EKEAKDLTELVQRRITYLQNPKDCSKAKKLVCNINKGCGYGCQLHHVVYCFMIAYGTQRT LILESQNWRYATGGWETVFRPVSETCTDRSGISTGHWSGEVKDKNVQVVELPIVDSLHPR PPYLPLAVPEDLADRLVRVHGDPAVWWVSQFVKYLIRPQPWLEKEIEEATKKLGFKHPVI GVHVRRTDKVGTEAAFHPIEEYMVHVEEHFQLLARRMQVDKKRVYLATDDPSLLKEAKTK YPNYEFISDNSISWSAGLHNRYTENSLRGVILDIHFLSQADFLVCTFSSQVCRVAYEIMQ TLHPDASANFHSLDDIYYFGGQNAHNQIAIYAHQPRTADEIPMEPGDIIGVAGNHWDGYS KGVNRKLGRTGLYPSYKVREKIETVKYPTYPEAEK
HsFUT9	>FUT9 MTSTSKGILRPFLIVCIILGCFMACLLIYIKPTNSWIFSPMESASSVLKMKNFFSTKTDY FNETTILVWVWPFGQTFDLTSCQAMFNIQGCHLTTDRSLYNKSHAVLIHHRDISWDLTNL PQQARPPFQKWIWMNLESPTHTPQKSGIEHLFNLTLTYRRDSDIQVPYGFLTVSTNPFVF EVPSKEKLVCWVVSNWNPEHARVKYYNELSKSIEIHTYGQAFGEYVNDKNLIPTISTCKF YLSFENSIHKDYITEKLYNAFLAGSVPVVLGPSRENYENYIPADSFIHVEDYNSPSELAK YLKEVDKNNKLYLSYFNWRKDFTVNLPRFWESHACLACDHVKRHQEYKSVGNLEKWFWN
HsFUT10	>FUT10 MVRIQRRKLLASCLCVTATVFLLVTLQVMVELGKFERKEFKSSSLQDGHTKMEEAPTHLN SFLKKEGLTFNRKRKWELDSYPIMLWWSPLTGETGRLGQCGADACFFTINRTYLHHHMTK AFLFYGTDFNIDSLPLPRKAHHDWAVFHEESPKNNYKLFHKPVITLFNYTATFSRHSHLP LTTQYLESIEVLKSLRYLVPLQSKNKLRKRLAPLVYVQSDCDPPSDRDSYVRELMTYIEV DSYGECLRNKDLPQQLKNPASMDADGFYRIIAQYKFILAFENAVCDDYITEKFWRPLKLG VVPVYYGSPSITDWLPSNKSAILVSEFSHPRELASYIRRLDSDDRLYEAYVEWKLKGEIS NQRLLTALRERKWGVQDVNQDNYIDAFECMVCTKVWANIRLQEKGLPPKRWEAEDTHLSC PEPTVFAFSPLRTPPLSSLREMWISSFEQSKKEAQALRWLVDRNQNFSSQEFWGLVFKD
HsFUT11	>FUT11 MAAGPIRVVLVLLGVLSVCAASGHGSVAEREAGGEAEWAEPWDGAVFRPPSALGAVGVTR SSGTPRPGREEAGDLPVLLWWSPGLFPHFPGDSERIECARGACVASRNRRALRDSRTRAL LFYGTDFRASAAPLPRLAHQSWALLHEESPLNNFLLSHGPGIRLFNLTSTFSRHSDYPLS LQWLPGTAYLRRPVPPPMERAEWRRRGYAPLLYLQSHCDVPADRDRYVRELMRHIPVDSY GKCLQNRELPTARLQDTATATTEDPELLAFLSRYKFHLALENAICNDYMTEKLWRPMHLG AVPVYRGSPSVRDWMPNNHSVILIDDFESPQKLAEFIDFLDKNDEEYMKYLAYKQPGGIT NQFLLDSLKHREWGVNDPLLPNYLNGFECFVCDYELARLDAEKAHAASPGDSPVFEPHIA QPSHMDCPVPTPGFGNVEEIPENDSWKEMWLQDYWQGLDQGEALTAMIHNNETEQTKFWD YLHEIFMKRQHL
HsPOFUT1	>POFUT1 MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKL LNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPE KRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQ RFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLR IGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRS LDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTA FVKRERDLQGRPSSFFGMDRPPKLRDEF
HsST3Gal1	>ST3GAL1 MVTLRKRTLKVLTFLVLFIFLTSFFLNYSHTMVATTWFPKQMVLELSENLKRLIKHRPCT CTHCIGQRKLSAWFDERFNQTMQPLLTAQNALLEDDTYRWWLRLQREKKPNNLNDTIKEL FRVVPGNVDPMLEKRSVGCRRCAVVGNSGNLRESSYGPEIDSHDFVLRMNKAPTAGFEAD VGTKTTHHLVYPESFRELGDNVSMILVPFKTIDLEWVVSAITTGTISHTYIPVPAKIRVK QDKILIYHPAFIKYVFDNWLQGHGRYPSTGILSVIFSMHVCDEVDLYGFGADSKGNWHHY WENNPSAGAFRKTGVHDADFESNVTATLASINKIRIFKGR
HsST3Gal3	>ST3GAL3 MGLLVFVRNLLLALCLFLVLGFLYYSAWKLHLLQWEEDSNSVVLSFDSAGQTLGSEYDRL

	GFLLNLDSKLPAELATKYANFSEGACKPGYASALMTAIFPRFSKPAPMFLDDSFRKWARI REFVPPFGIKGQDNLIKAILSVTKEYRLTPALDSLRCRRCIIVGNGGVLANKSLGSRIDD YDIVVRLNSAPVKGFEKDVGSKTTLRITYPEGAMQRPEQYERDSLFVLAGFKWQDFKWLK YIVYKERVSASDGFWKSVATRVPKEPPEIRILNPYFIQEAAFTLIGLPFNNGLMGRGNIP TLGSVAVTMALHGCDEVAVAGFGYDMSTPNAPLHYYETVRMAAIKESWTHNIQREKEFLR KLVKARVITDLSSGI
HsST3Gal4	>ST3GAL4 MVSKSRWKLLAMLALVLVVMVWYSISREDRYIELFYFPIPEKKEPCLQGEAESKASKLFG NYSRDQPIFLRLEDYFWVKTPSAYELPYGTKGSEDLLLRVLAITSSSIPKNIQSLRCRRC VVVGNGHRLRNSSLGDAINKYDVVIRLNNAPVAGYEGDVGSKTTMRLFYPESAHFDPKVE NNPDTLLVLVAFKAMDFHWIETILSDKKRVRKGFWKQPPLIWDVNPKQIRILNPFFMEIA ADKLLSLPMQQPRKIKQKPTTGLLAITLALHLCDLVHIAGFGYPDAYNKKQTIHYYEQIT LKSMAGSGHNVSQEALAIKRMLEMGAIKNLTSF
HsST3Gal6	>ST3GAL6 MRGYLVAIFLSAVFLYYVLHCILWGTNVYWVAPVEMKRRNKIQPCLSKPAFASLLRFHQF HPFLCAADFRKIASLYGSDKFDLPYGMRTSAEYFRLALSKLQSCDLFDEFDNIPCKKCVV VGNGGVLKNKTLGEKIDSYDVIIRMNNGPVLGHEEEVGRRTTFRLFYPESVFSDPIHNDP NTTVILTAFKPHDLRWLLELLMGDKINTNGFWKKPALNLIYKPYQIRILDPFIIRTAAYE LLHFPKVFPKNQKPKHPTTGIIAITLAFYICHEVHLAGFKYNFSDLKSPLHYYGNATMSL MNKNAYHNVTAEQLFLKDIIEKNLVINLTQD
HsST6Gal1	>ST6Gal1 MIHTNLKKKFSCCVLVFLLFAVICVWKEKKKGSYYDSFKLQTKEFQVLKSLGKLAMGSDS QSVSSSSTQDPHRGRQTLGSLRGLAKAKPEASFQVWNKDSSSKNLIPRLQKIWKNYLSMN KYKVSYKGPGPGIKFSAEALRCHLRDHVNVSMVEVTDFPFNTSEWEGYLPKESIRTKAGP WGRCAVVSSAGSLKSSQLGREIDDHDAVLRFNGAPTANFQQDVGTKTTIRLMNSQLVTTE KRFLKDSLYNEGILIVWDPSVYHSDIPKWYQNPDYNFFNNYKTYRKLHPNQPFYILKPQM PWELWDILQEISPEEIQPNPPSSGMLGIIIMMTLCDQVDIYEFLPSKRKTDVCYYYQKFF DSACTMGAYHPLLYEKNLVKHLNQGTDEDIYLLGKATLPGFRTIHC
HsST6GalNAc1	>ST6GalNAc1 MRSCLWRCRHLSQGVQWSLLLAVLVFFLFALPSFIKEPQTKPSRHQRTENIKERSLQSLA KPKSQAPTRARRTTIYAEPVPENNALNTQTQPKAHTTGDRGKEANQAPPEEQDKVPHTAQ RAAWKSPEKEKTMVNTLSPRGQDAGMASGRTEAQSWKSQDTKTTQGNGGQTRKLTASRTV SEKHQGKAATTAKTLIPKSQHRMLAPTGAVSTRTRQKGVTTAVIPPKEKKPQATPPPAPF QSPTTQRNQRLKAANFKSEPRWDFEEKYSFEIGGLQTTCPDSVKIKASKSLWLQKLFLPN LTLFLDSRHFNQSEWDRLEHFAPPFGFMELNYSLVQKVVTRFPPVPQQQLLLASLPAGSL RCITCAVVGNGGILNNSHMGQEIDSHDYVFRLSGALIKGYEQDVGTRTSFYGFTAFSLTQ SLLILGNRGFKNVPLGKDVRYLHFLEGTRDYEWLEALLMNQTVMSKNLFWFRHRPQEAFR EALHMDRYLLLHPDFLRYMKNRFLRSKTLDGAHWRIYRPTTGALLLLTALQLCDQVSAYG FITEGHERFSDHYYDTSWKRLIFYINHDFKLEREVWKRLHDEGIIRLYQRPGPGTAKAKN
HsST6GalNAc2	>ST6GalNAc2 MGLPRGSFFWLLLLLTAACSGLLFALYFSAVQRYPGPAAGARDTTSFEAFFQSKASNSWT GKGQACRHLLHLAIQRHPHFRGLFNLSIPVLLWGDLFTPALWDRLSQHKAPYGWRGLSHQ VIASTLSLLNGSESAKLFAPPRDTPPKCIRCAVVGNGGILNGSRQGPNIDAHDYVFRLNG AVIKGFERDVGTKTSFYGFTVNTMKNSLVSYWNLGFTSVPQGQDLQYIFIPSDIRDYVML RSAILGVPVPEGLDKGDRPHAYFGPEASASKFKLLHPDFISYLTERFLKSKLINTHFGDL YMPSTGALMLLTALHTCDQVSAYGFITSNYWKFSDHYFERKMKPLIFYANHDLSLEAALW RDLHKAGILQLYQR
HsST6GalNAc4	>ST6GalNAc4 MKAPGRLVLIILCSVVFSAVYILLCCWAGLPLCLATCLDHHFPTGSRPTVPGPLHFSGYS SVPDGKPLVREPCRSCAVVSSSGQMLGSGLGAEIDSAECVFRMNQAPTVGFEADVGQRST LRVVSHTSVPLLLRNYSHYFQKARDTLYMVWGQGRHMDRVLGGRTYRTLLQLTRMYPGLQ VYTFTERMMAYCDQIFQDETGKNRRQSGSFLSTGWFTMILALELCEEIVVYGMVSDSYCR EKSHPSVPYHYFEKGRLDECQMYLAHEQAPRSAHRFITEKAVFSRWAKKRPIVFAHPSWRTE
HsST8Sia1	>ST8Sia1 MSPCGRARRQTSRGAMAVLAWKFPRTRLPMGASALCVVVLCWLYIFPVYRLPNEKEIVQG VLQQGTAWRRNQTAARAFRKQMEDCCDPAHLFAMTKMNSPMGKSMWYDGEFLYSFTIDNS TYSLFPQATPFQLPLKKCAVVGNGGILKKSGCGRQIDEANFVMRCNLPPLSSEYTKDVGS KSQLVTANPSIIRQRFQNLLWSRKTFVDNMKIYNHSYIYMPAFSMKTGTEPSLRVYYTLS

	DVGANQTVLFANPNFLRSIGKFWKSRGIHAKRLSTGLFLVSAALGLCEEVAIYGFWPFSV NMHEQPISHHYYDNVLPFSGFHAMPEEFLQLWYLHKIGALRMQLDPCEDTSLQPTS
HsST8Sia2	>ST8Sia2 MQLQFRSWMLAALTLLVVFLIFADISEIEEEIGNSGGRGTIRSAVNSLHSKSNRAEVVIN GSSSPAVVDRSNESIKHNIQPASSKWRHNQTLSLRIRKQILKFLDAEKDISVLKGTLKPG DIIHYIFDRDSTMNVSQNLYELLPRTSPLKNKHFGTCAIVGNSGVLLNSGCGQEIDAHSF VIRCNLAPVQEYARDVGLKTDLVTMNPSVIQRAFEDLVNATWREKLLQRLHSLNGSILWI PAFMARGGKERVEWVNELILKHHVNVRTAYPSLRLLHAVRGYWLTNKVHIKRPTTGLLMY TLATRFCKQIYLYGFWPFPLDQNQNPVKYHYYDSLKYGYTSQASPHTMPLEFKALKSLHE QGALKLTVGQCDGAT
HsST8Sia3	>ST8Sia3 MRNCKMARVASVLGLVMLSVALLILSLISYVSLKKENIFTTPKYASPGAPRMYMFHAGFR SQFALKFLDPSFVPITNSLTQELQEKPSKWKFNRTAFLHQRQEILQHVDVIKNFSLTKNS VRIGQLMHYDYSSHKYVFSISNNFRSLLPDVSPIMNKHYNICAVVGNSGILTGSQCGQEI DKSDFVFRCNFAPTEAFQRDVGRKTNLTTFNPSILEKYYNNLLTIQDRNNFFLSLKKLDG AILWIPAFFFHTSATVTRTLVDFFVEHRGQLKVQLAWPGNIMQHVNRYWKNKHLSPKRLS TGILMYTLASAICEEIHLYGFWPFGFDPNTREDLPYHYYDKKGTKFTTKWQESHQLPAEF QLLYRMHGEGLTKLTLSHCA
HsST8Sia4	>ST8Sia4 MRSIRKRWTICTISLLLIFYKTKEIARTEEHQETQLIGDGELSLSRSLVNSSDKIIRKAG SSIFQHNVEGWKINSSLVLEIRKNILRFLDAERDVSVVKSSFKPGDVIHYVLDRRRTLNI SHDLHSLLPEVSPMKNRRFKTCAVVGNSGILLDSECGKEIDSHNFVIRCNLAPVVEFAAD VGTKSDFITMNPSVVQRAFGGFRNESDREKFVHRLSMLNDSVLWIPAFMVKGGEKHVEWV NALILKNKLKVRTAYPSLRLIHAVRGYWLTNKVPIKRPSTGLLMYTLATRFCDEIHLYGF WPFPKDLNGKAVKYHYYDDLKYRYFSNASPHRMPLEFKTLNVLHNRGALKLTTGKCVKQ
HsGalNAcT1	>GALNACT1 MRKFAYCKVVLATSLIWVLLDMFLLLYFSECNKCDEKKERGLPAGDVLEPVQKPHEGPGE MGKPVVIPKEDQEKMKEMFKINQFNLMASEMIALNRSLPDVRLEGCKTKVYPDNLPTTSV VIVFHNEAWSTLLRTVHSVINRSPRHMIEEIVLVDDASERDFLKRPLESYVKKLKVPVHV IRMEQRSGLIRARLKGAAVSKGQVITFLDAHCECTVGWLEPLLARIKHDRRTVVCPIIDV ISDDTFEYMAGSDMTYGGFNWKLNFRWYPVPQREMDRRKGDRTLPVRTPTMAGGLFSIDR DYFQEIGTYDAGMDIWGGENLEISFRIWQCGGTLEIVTCSHVGHVFRKATPYTFPGGTGQ IINKNNRRLAEVWMDEFKNFFYIISPGVTKVDYGDISSRVGLRHKLQCKPFSWYLENIYP DSQIPRHYFSLGEIRNVETNQCLDNMARKENEKVGIFNCHGMGGNQVFSYTANKEIRTDD LCLDVSKLNGPVTMLKCHHLKGNQLWEYDPVKLTLQHVNSNQCLDKATEEDSQVPSIRDC NGSRSQQWLLRNVTLPEIF
HsGalNAcT2	>GALNACT2 MRRRSRMLLCFAFLWVLGIAYYMYSGGGSALAGGAGGGAGRKEDWNEIDPIKKKDLHHSN GEEKAQSMETLPPGKVRWPDFNQEAYVGGTMVRSGQDPYARNKFNQVESDKLRMDRAIPD TRHDQCQRKQWRVDLPATSVVITFHNEARSALLRTVVSVLKKSPPHLIKEIILVDDYSND PEDGALLGKIEKVRVLRNDRREGLMRSRVRGADAAQAKVLTFLDSHCECNEHWLEPLLER VAEDRTRVVSPIIDVINMDNFQYVGASADLKGGFDWNLVFKWDYMTPEQRRSRQGNPVAP IKTPMIAGGLFVMDKFYFEELGKYDMMMDVWGGENLEISFRVWQCGGSLEIIPCSRVGHV FRKQHPYTFPGGSGTVFARNTRRAAEVWMDEYKNFYYAAVPSARNVPYGNIQSRLELRKK LSCKPFKWYLENVYPELRVPDHQDIAFGALQQGTNCLDTLGHFADGVVGVYECHNAGGNQ EWALTKEKSVKHMDLCLTVVDRAPGSLIKLQGCRENDSRQKWEQIEGNSKLRHVGSNLCL DSRTAKSGGLSVEVCGPALSQQWKFTLNLQQ
HsGalNAcT3	>GALNACT3 MAHLKRLVKLHIKRHYHKKFWKLGAVIFFFIIVLVLMQREVSVQYSKEESRMERNMKNKN KMLDLMLEAVNNIKDAMPKMQIGAPVRQNIDAGERPCLQGYYTAAELKPVLDRPPQDSNA PGASGKAFKTTNLSVEEQKEKERGEAKHCFNAFASDRISLHRDLGPDTRPPECIEQKFKR CPPLPTTSVIIVFHNEAWSTLLRTVHSVLYSSPAILLKEIILVDDASVDEYLHDKLDEYV KQFSIVKIVRQRERKGLITARLLGATVATAETLTFLDAHCECFYGWLEPLLARIAENYTA VVSPDIASIDLNTFEFNKPSPYGSNHNRGNFDWSLSFGWESLPDHEKQRRKDETYPIKTP TFAGGLFSISKEYFEYIGSYDEEMEIWGGENIEMSFRVWQCGGQLEIMPCSVVGHVFRSK SPHSFPKGTQVIARNQVRLAEVWMDEYKEIFYRRNTDAAKIVKQKAFGDLSKRFEIKHRL QCKNFTWYLNNIYPEVYVPDLNPVISGYIKSVGQPLCLDVGENNQGGKPLIMYTCHGLGG

	NQYFEYSAQHEIRHNIQKELCLHAAQGLVQLKACTYKGHKTVVTGEQIWEIQKDQLLYNP FLKMCLSANGEHPSLVSCNPSDPLQKWILSQND
HsGalNAcT4	>GALNACT4 MAVRWTWAGKSCLLLAFLTVAYIFVELLVSTFHASAGAGRARELGSRRLSDLQKNTEDLS RPLYKKPPADSRALGEWGKASKLQLNEDELKQQEELIERYAINIYLSDRISLHRHIEDKR MYECKSQKFNYRTLPTTSVIIAFYNEAWSTLLRTIHSVLETSPAVLLKEIILVDDLSDRV YLKTQLETYISNLDRVRLIRTNKREGLVRARLIGATFATGDVLTFLDCHCECNSGWLEPL LERIGRDETAVVCPVIDTIDWNTFEFYMQIGEPMIGGFDWRLTFQWHSVPKQERDRRISR IDPIRSPTMAGGLFAVSKKYFQYLGTYDTGMEVWGGENLELSFRVWQCGGKLEIHPCSHV GHVFPKRAPYARPNFLQNTARAAEVWMDEYKEHFYNRNPPARKEAYGDISERKLLRERLR CKSFDWYLKNVFPNLHVPEDRPGWHGAIRSRGISSECLDYNSPDNNPTGANLSLFGCHGQ GGNQFFEYTSNKEIRFNSVTELCAEVPEQKNYVGMQNCPKDGFPVPANIIWHFKEDGTIF HPHSGLCLSAYRTPEGRPDVQMRTCDALDKNQIWSFEK
HsGalNAcT5	>GALNACT5 MNRIRKFFRGSGRVLAFIFVASVIWLLFDMAALRLSFSEINTRVIKEDIVRRERIGFRVQ PDQGKIFYSSIKEMKPPLRGHGKGAWGKENVRKTEESVLKVEVDLDQTQRERKMQNALGR GKVVPLWHPAHLQTLPVTPNKQKTDGRGTKPEASSHQGTPKQTTAQGAPKTSFIAAKGTQ VVKISVHMGRVSLKQEPRKSHSPSSDTSKLAAERDLNVTISLSTDRPKQRSQAVANERAH PASTAVPKSGEAMALNKTKTQSKEVNANKHKANTSLPFPKFTVNSNRLRKQSINETPLGS LSKDDGARGAHGKKLNFSESHLVIITKEEEQKADPKEVSNSKTKTIFPKVLGKSQSKHIS RNRSEMSSSSLAPHRVPLSQTNHALTGGLEPAKINITAKAPSTEYNQSHIKALLPEDSGT HQVLRIDVTLSPRDPKAPGQFGRPVVVPHGKEKEAERRWKEGNFNVYLSDLIPVDRAIED TRPAGCAEQLVHNNLPTTSVIMCFVDEVWSTLLRSVHSVINRSPPHLIKEILLVDDFSTK DYLKDNLDKYMSQFPKVRILRLKERHGLIRARLAGAQNATGDVLTFLDSHVECNVGWLEP LLERVYLSRKKVACPVIEVINDKDMSYMTVDNFQRGIFVWPMNFGWRTIPPDVIAKNRIK ETDTIRCPVMAGGLFSIDKSYFFELGTYDPGLDVWGGENMELSFKVWMCGGEIEIIPCSR VGHIFRNDNPYSFPKDRMKTVERNLVRVAEVWLDEYKELFYGHGDHLIDQGLDVGNLTQQ RELRKKLKCKSFKWYLENVFPDLRAPIVRASGVLINVALGKCISIENTTVILEDCDGSKE LQQFNYTWLRLIKCGEWCIAPIPDKGAVRLHPCDNRNKGLKWLHKSTSVFHPELVNHIVF ENNQQLLCLEGNFSQKILKVAACDPVKPYQKWKFEKYYEA
HsGalNAcT6	>GALNT6 MRLLRRRHMPLRLAMVGCAFVLFLFLLHRDVSSREEATEKPWLKSLVSRKDHVLDLMLEA MNNLRDSMPKLQIRAPEAQQTLFSINQSCLPGFYTPAELKPFWERPPQDPNAPGADGKAF QKSKWTPLETQEKEEGYKKHCFNAFASDRISLQRSLGPDTRPPECVDQKFRRCPPLATTS VIIVFHNEAWSTLLRTVYSVLHTTPAILLKEIILVDDASTEEHLKEKLEQYVKQLQVVRV VRQEERKGLITARLLGASVAQAEVLTFLDAHCECFHGWLEPLLARIAEDKTVVVSPDIVT IDLNTFEFAKPVQRGRVHSRGNFDWSLTFGWETLPPHEKQRRKDETYPIKSPTFAGGLFS ISKSYFEHIGTYDNQMEIWGGENVEMSFRVWQCGGQLEIIPCSVVGHVFRTKSPHTFPKG TSVIARNQVRLAEVWMDSYKKIFYRRNLQAAKMAQEKSFGDISERLQLREQLHCHNFSWY LHNVYPEMFVPDLTPTFYGAIKNLGTNQCLDVGENNRGGKPLIMYSCHGLGGNQYFEYTT QRDLRHNIAKQLCLHVSKGALGLGSCHFTGKNSQVPKDEEWELAQDQLIRNSGSGTCLTS QDKKPAMAPCNPSDPHQLWLFV
HsGalNAcT7	>GALNT7 MRLKIGFILRSLLVVGSFLGLVVLWSSLTPRPDDPSPLSRMREDRDVNDPMPNRGGNGLA PGEDRFKPVVPWPHVEGVEVDLESIRRINKAKNEQEHHAGGDSQKDIMQRQYLTFKPQTF TYHDPVLRPGILGNFEPKEPEPPGVVGGPGEKAKPLVLGPEFKQAIQASIKEFGFNMVAS DMISLDRSVNDLRQEECKYWHYDENLLTSSVVIVFHNEGWSTLMRTVHSVIKRTPRKYLA EIVLIDDFSNKEHLKEKLDEYIKLWNGLVKVFRNERREGLIQARSIGAQKAKLGQVLIYL DAHCEVAVNWYAPLVAPISKDRTICTVPLIDVINGNTYEIIPQGGGDEDGYARGAWDWSM LWKRVPLTPQEKRLRKTKTEPYRSPAMAGGLFAIEREFFFELGLYDPGLQIWGGENFEIS YKIWQCGGKLLFVPCSRVGHIYRLEGWQGNPPPIYVGSSPTLKNYVRVVEVWWDEYKDYF YASRPESQALPYGDISELKKFREDHNCKSFKWFMEEIAYDITSHYPLPPKNVDWGEIRGF ETAYCIDSMGKTNGGFVELGPCHRMGGNQLFRINEANQLMQYDQCLTKGADGSKVMITHC NLNEFKEWQYFKNLHRFTHIPSGKCLDRSEVLHQVFISNCDSSKTTQKWEMNNIHSV
HsGalNAcT8	>GALNT8 MMFWRKLPKALFIGLTLAIAVNLLLVFSSKGTLQNLFTGGLHRELPLHLNKRYGAVIKRL SHLEVELQDLKESMKLALRQQENVNSTLKRAKDEVRPLLKAMETKVNETKKHKTQMKLFP HSQLFRQWGEDLSEAQQKAAQDLFRKFGYNAYLSNQLPLNRTIPDTRDYRCLRKTYPSQL

	PSLSVILIFVNEALSIIQRAITSIINRTPSRLLKEIILVDDFSSNGELKVHLDEKIKLYN QKYPGLLKIIRHPERKGLAQARNTGWEAATADVVAILDAHIEVNVGWAEPILARIQEDRT VIVSPVFDNIRFDTFKLDKYELAVDGFNWELWCRYDALPQAWIDLHDVTAPVKSPSIMGI LAANRHFLGEIGSLDGGMLIYGGENVELSLRVWQCGGKVEILPCSRIAHLERHHKPYALD LTAALKRNALRVAEIWMDEHKHMVYLAWNIPLQNSGIDFGDVSSRMALREKLKCKTFDWY LKNVYPLLKPLHTIVGYGRMKNLLDENVCLDQGPVPGNTPIMYYCHEFSSQNVYYHLTGE LYVGQLIAEASASDRCLTDPGKAEKPTLEPCSKAAKNRLHIYWDFKPGGAVINRDTKRCL EMKKDLLGSHVLVLQTCSTQVWEIQHTVRDWGQTNSQ
HsGalNAcT9	>GALNT9 MAVARKIRTLLTVNILVFVGIVLFSVYCRLQGRSQELVRIVSGDRRVRSRHAKVGTLGDR EAILQRLDHLEEVVYNQLNGLAKPIGLVEGPGGLGQGGLAATLRDDGQEAEGKYEEYGYN AQLSDRISLDRSIPDYRPRKCRQMSYAQDLPQVSVVFIFVNEALSVILRSVHSVVNHTPS QLLKEVILVDDNSDNVELKFNLDQYVNKRYPGLVKIVRNSRREGLIRARLQGWKAATAPV VGFFDAHVEFNTGWAEPALSRIREDRRRIVLPAIDNIKYSTFEVQQYANAAHGYNWGLRC MYIIPPQDWLDRGDESAPIRTPAMIGCSFVVDREYFGDIGLLDPGMEVYGGENVELGMRV WQCGGSMEVLPCSRVAHIERTRKPYNNDIDYYAKRNALRAAEVWMDDFKSHVYMAWNIPM SNPGVDFGDVSERLALRQRLKCRSFKWYLENVYPEMRVYNNTLTYGEVRNSKASAYCLDQ GAEDGDRAILYPCHGMSSQLVRYSADGLLQLGPLGSTAFLPDSKCLVDDGTGRMPTLKKC EDVARPTQRLWDFTQSGPIVSRATGRCLEVEMSKDANFGLRLVVQRCSGQKWMIRNWIKH ARH
HsGalNAcT10	>GALNT10 MRRKEKRLLQAVALVLAALVLLPNVGLWALYRERQPDGTPGGSGAAVAPAAGQGSHSRQK KTFFLGDGQKLKDWHDKEAIRRDAQRVGNGEQGRPYPMTDAERVDQAYRENGFNIYVSDK ISLNRSLPDIRHPNCNSKRYLETLPNTSIIIPFHNEGWSSLLRTVHSVLNRSPPELVAEI VLVDDFSDREHLKKPLEDYMALFPSVRILRTKKREGLIRTRMLGASVATGDVITFLDSHC EANVNWLPPLLDRIARNRKTIVCPMIDVIDHDDFRYETQAGDAMRGAFDWEMYYKRIPIP PELQKADPSDPFESPVMAGGLFAVDRKWFWELGGYDPGLEIWGGEQYEISFKVWMCGGRM EDIPCSRVGHIYRKYVPYKVPAGVSLARNLKRVAEVWMDEYAEYIYQRRPEYRHLSAGDV AVQKKLRSSLNCKSFKWFMTKIAWDLPKFYPPVEPPAAAWGEIRNVGTGLCADTKHGALG SPLRLEGCVRGRGEAAWNNMQVFTFTWREDIRPGDPQHTKKFCFDAISHTSPVTLYDCHS MKGNQLWKYRKDKTLYHPVSGSCMDCSESDHRIFMNTCNPSSLTQQWLFEHTNSTVLEKF NRN
HsB3GALNT1	>B3GALNT1 MASALWTVLPSRMSLRSLKWSLLLLSLLSFFVMWYLSLPHYNVIERVNWMYFYEYEPIYR QDFHFTLREHSNCSHQNPFLVILVTSHPSDVKARQAIRVTWGEKKSWWGYEVLTFFLLGQ EAEKEDKMLALSLEDEHLLYGDIIRQDFLDTYNNLTLKTIMAFRWVTEFCPNAKYVMKTD TDVFINTGNLVKYLLNLNHSEKFFTGYPLIDNYSYRGFYQKTHISYQEYPFKVFPPYCSG LGYIMSRDLVPRIYEMMGHVKPIKFEDVYVGICLNLLKVNIHIPEDTNLFFLYRIHLDVC QLRRVIAAHGFSSKEIITFWQVMLRNTTCHY
HsB4GALNT1	>B4GALNT1 MWLGRRALCALVLLLACASLGLLYASTRDAPGLRLPLAPWAPPQSPRRPELPDLAPEPRY AHIPVRIKEQVVGLLAWNNCSCESSGGGLPLPFQKQVRAIDLTKAFDPAELRAASATREQ EFQAFLSRSQSPADQLLIAPANSPLQYPLQGVEVQPLRSILVPGLSLQAASGQEVYQVNL TASLGTWDVAGEVTGVTLTGEGQADLTLVSPGLDQLNRQLQLVTYSSRSYQTNTADTVRF STEGHEAAFTIRIRHPPNPRLYPPGSLPQGAQYNISALVTIATKTFLRYDRLRALITSIR RFYPTVTVVIADDSDKPERVSGPYVEHYLMPFGKGWFAGRNLAVSQVTTKYVLWVDDDFV FTARTRLERLVDVLERTPLDLVGGAVREISGFATTYRQLLSVEPGAPGLGNCLRQRRGFH HELVGFPGCVVTDGVVNFFLARTDKVREVGFDPRLSRVAHLEFFLDGLGSLRVGSCSDVV VDHASKLKLPWTSRDAGAETYARYRYPGSLDESQMAKHRLLFFKHRLQCMTSQ
HsA-ABO	>A-ABO MAEVLRTLAGKPKCHALRPMILFLIMLVLVLFGYGVLSPRSLMPGSLERGFCMAVREPDH LQRVSLPRMVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTV FAIKKYVAFLKLFLETAEKHFMVGHRVHYYVFTDQPAAVPRVTLGTGRQLSVLEVRAYKR WQDVSMRRMEMISDFCERRFLSEVDYLVCVDVDMEFRDHVGVEILTPLFGTLHPGFYGSS REAFTYERRPQSQAYIPKDEGDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVW HDESHLNKYLLRHKPTKVLSPEYLWDQQLLGWPAVLRKLRFTAVPKNHQAVRNP
HsA4GALT	>A4GALT MSKPPDLLLRLLRGAPRQRVCTLFIIGFKFTFFVSIMIYWHVVGEPKEKGQLYNLPAEIP

	CPTLTPPTPPSHGPTPGNIFFLETSDRTNPNFLFMCSVESAARTHPESHVLVLMKGLPGG NASLPRHLGISLLSCFPNVQMLPLDLRELFRDTPLADWYAAVQGRWEPYLLPVLSDASRI ALMWKFGGIYLDTDFIVLKNLRNLTNVLGTQSRYVLNGAFLAFERRHEFMALCMRDFVDH YNGWIWGHQGPQLLTRVFKKWCSIRSLAESRACRGVTTLPPEAFYPIPWQDWKKYFEDIN PEELPRLLSATYAVHVWNKKSQGTRFEATSRALLAQLHARYCPTTHEAMKMYL
HsB3GalT1	>B3GalT1 MASKVSCLYVLTVVCWASALWYLSITRPTSSYTGSKPFSHLTVARKNFTFGNIRTRPINP HSFEFLINEPNKCEKNIPFLVILISTTHKEFDARQAIRETWGDENNFKGIKIATLFLLGK NADPVLNQMVEQESQIFHDIIVEDFIDSYHNLTLKTLMGMRWVATFCSKAKYVMKTDSDI FVNMDNLIYKLLKPSTKPRRRYFTGYVINGGPIRDVRSKWYMPRDLYPDSNYPPFCSGTG YIFSADVAELIYKTSLHTRLLHLEDVYVGLCLRKLGIHPFQNSGFNHWKMAYSLCRYRRV ITVHQISPEEMHRIWNDMSSKKHLRC
HsB3GalT2	>B3GalT2 MLQWRRRHCCFAKMTWNAKRSLFRTHLIGVLSLVFLFAMFLFFNHHDWLPGRAGFKENPV TYTFRGFRSTKSETNHSSLRNIWKETVPQTLRPQTATNSNNTDLSPQGVTGLENTLSANG SIYNEKGTGHPNSYHFKYIINEPEKCQEKSPFLILLIAAEPGQIEARRAIRQTWGNESLA PGIQITRIFLLGLSIKLNGYLQRAILEESRQYHDIIQQEYLDTYYNLTIKTLMGMNWVAT YCPHIPYVMKTDSDMFVNTEYLINKLLKPDLPPRHNYFTGYLMRGYAPNRNKDSKWYMPP DLYPSERYPVFCSGTGYVFSGDLAEKIFKVSLGIRRLHLEDVYVGICLAKLRIDPVPPPN EFVFNHWRVSYSSCKYSHLITSHQFQPSELIKYWNHLQQNKHNACANAAKEKAGRYRHRK LH
HsB4GalT1	>B4GalT1 MRLREPLLSGSAAMPGASLQRACRLLVAVCALHLGVTLVYYLAGRDLSRLPQLVGVSTPL QGGSNSAAAIGQSSGELRTGGARPPPPLGASSQPRPGGDSSPVVDSGPGPASNLTSVPVP HTTALSLPACPEESPLLVGPMLIEFNMPVDLELVAKQNPNVKMGGRYAPRDCVSPHKVAI IIPFRNRQEHLKYWLYYLHPVLQRQQLDYGIYVINQAGDTIFNRAKLLNVGFQEALKDYD YTCFVFSDVDLIPMNDHNAYRCFSQPRHISVAMDKFGFSLPYVQYFGGVSALSKQQFLTI NGFPNNYWGWGGEDDDIFNRLVFRGMSISRPNAVVGRCRMIRHSRDKKNEPNPQRFDRIA HTKETMLSDGLNSLTYQVLDVQRYPLYTQITVDIGTPS
HsB4GalT2	>B4GalT2 MSRLLGGTLERVCKAVLLLCLLHFLVAVILYFDVYAQHLAFFSRFSARGPAHALHPAASS SSSSSNCSRPNATASSSGLPEVPSALPGPTAPTLPPCPDSPPGLVGRLLIEFTSPMPLER VQRENPGVLMGGRYTPPDCTPAQTVAVIIPFRHREHHLRYWLHYLHPILRRQRLRYGVYV INQHGEDTFNRAKLLNVGFLEALKEDAAYDCFIFSDVDLVPMDDRNLYRCGDQPRHFAIA MDKFGFRLPYAGYFGGVSGLSKAQFLRINGFPNEYWGWGGEDDDIFNRISLTGMKISRPD IRIGRYRMIKHDRDKHNEPNPQRFTKIQNTKLTMKRDGIGSVRYQVLEVSRQPLFTNITV DIGRPPSWPPRG
HsB4GalT3	>B4GalT3 MLRRLLERPCTLALLVGSQLAVMMYLSLGGFRSLSALFGRDQGPTFDYSHPRDVYSNLSH LPGAPGGPPAPQGLPYCPERSPLLVGPVSVSFSPVPSLAEIVERNPRVEPGGRYRPAGCE PRSRTAIIVPHRAREHHLRLLLYHLHPFLQRQQLAYGIYVIHQAGNGTFNRAKLLNVGVR EALRDEEWDCLFLHDVDLLPENDHNLYVCDPRGPRHVAVAMNKFGYSLPYPQYFGGVSAL TPDQYLKMNGFPNEYWGWGGEDDDIATRVRLAGMKISRPPTSVGHYKMVKHRGDKGNEEN PHRFDLLVRTQNSWTQDGMNSLTYQLLARELGPLYTNITADIGTDPRGPRAPSGPRYPPG SSQAFRQEMLQRRPPARPGPLSTANHTALRGSH
HsB4GalT4	>B4GalT4 MGFNLTFHLSYKFRLLLLLTLCLTVVGWATSNYFVGAIQEIPKAKEFMANFHKTLILGKG KTLTNEASTKKVELDNCPSVSPYLRGQSKLIFKPDLTLEEVQAENPKVSRGRYRPQECKA LQRVAILVPHRNREKHLMYLLEHLHPFLQRQQLDYGIYVIHQAEGKKFNRAKLLNVGYLE ALKEENWDCFIFHDVDLVPENDFNLYKCEEHPKHLVVGRNSTGYRLRYSGYFGGVTALSR EQFFKVNGFSNNYWGWGGEDDDLRLRVELQRMKISRPLPEVGKYTMVFHTRDKGNEVNAE RMKLLHQVSRVWRTDGLSSCSYKLVSVEHNPLYINITVDFWFGA
HsB4GalT5	>B4GalT5 MRARRGLLRLPRRSLLAALFFFSLSSSLLYFVYVAPGIVNTYLFMMQAQGILIRDNVRTI GAQVYEQVLRSAYAKRNSSVNDSDYPLDLNHSETFLQTTTFLPEDFTYFANHTCPERLPS MKGPIDINMSEIGMDYIHELFSKDPTIKLGGHWKPSDCMPRWKVAILIPFRNRHEHLPVL FRHLLPMLQRQRLQFAFYVVEQVGTQPFNRAMLFNVGFQEAMKDLDWDCLIFHDVDHIPE SDRNYYGCGQMPRHFATKLDKYMYLLPYTEFFGGVSGLTVEQFRKINGFPNAFWGWGGED

	DDLWNRVQNAGYSVSRPEGDTGKYKSIPHHHRGEVQFLGRYALLRKSKERQGLDGLNNLN YFANITYDALYKNITVNLTPELAQVNEY
HsB4GalT6	>B4GalT6 MSVLRRMMRVSNRSLLAFIFFFSLSSSCLYFIYVAPGIANTYLFMVQARGIMLRENVKTI GHMIRLYTNKNSTLNGTDYPEGNNSSDYLVQTTTYLPENFTYSPYLPCPEKLPYMRGFLN VNVSEVSFDEIHQLFSKDLDIEPGGHWRPKDCKPRWKVAVLIPFRNRHEHLPIFFLHLIP MLQKQRLEFAFYVIEQTGTQPFNRAMLFNVGFKEAMKDSVWDCVIFHDVDHLPENDRNYY GCGEMPRHFAAKLDKYMYILPYKEFFGGVSGLTVEQFRKINGFPNAFWGWGGEDDDLWNR VHYAGYNVTRPEGDLGKYKSIPHHHRGEVQFLGRYKLLRYSKERQYIDGLNNLIYRPKIL VDRLYTNISVNLMPELAPIEDY
HsB-ABO	>B-ABO MAEVLRTLAGKPKCHALRPMILFLIMLVLVLFGYGVLSPRSLMPGSLERGFCMAVREPDH LQRVSLPRMVYPQPKVLTPCRKDVLVVTPWLAPIVWEGTFNIDILNEQFRLQNTTIGLTV FAIKKYVAFLKLFLETAEKHFMVGHRVHYYVFTDQPAAVPRVTLGTGRQLSVLEVRAYKR WQDVSMRRMEMISDFCERRFLSEVDYLVCVDVDMEFRDHVGVEILTPLFGTLHPGFYGSS REAFTYERRPQSQAYIPKDEGDFYYLGGFFGGSVQEVQRLTRACHQAMMVDQANGIEAVW HDESHLNKYLLRHKPTKVLSPEYLWDQQLLGWPAVLRKLRFTAVPKNHQAVRNP
HsUGT8	>UGT8 MKSYTPYFILLWSAVGIAKAAKIIIVPPIMFESHMYIFKTLASALHERGHHTVFLLSEGR DIAPSNHYSLQRYPGIFNSTTSDAFLQSKMRNIFSGRLTAIELFDILDHYTKNCDLMVGN HALIQGLKKEKFDLLLVDPNDMCGFVIAHLLGVKYAVFSTGLWYPAEVGAPAPLAYVPEF NSLLTDRMNLLQRMKNTGVYLISRLGVSFLVLPKYERIMQKYNLLPEKSMYDLVHGSSLW MLCTDVALEFPRPTLPNVVYVGGILTKPASPLPEDLQRWVNGANEHGFVLVSFGAGVKYL SEDIANKLAGALGRLPQKVIWRFSGPKPKNLGNNTKLIEWLPQNDLLGHSKIKAFLSHGG LNSIFETIYHGVPVVGIPLFGDHYDTMTRVQAKGMGILLEWKTVTEKELYEALVKVINNP SYRQRAQKLSEIHKDQPGHPVNRTIYWIDYIIRHNGAHHLRAAVHQISFCQYFLLDIAFV LLLGAALLYFLLSWVTKFIYRKIKSLWSRNKHSTVNGHYHNGILNGKYKRNGHIKHEKKVK
HsC1GLT	>C1GLT MASKSWLNFLTFLCGSAIGFLLCSQLFSILLGEKVDTQPNVLHNDPHARHSDDNGQNHLE GQMNFNADSSQHKDENTDIAENLYQKVRILCWVMTGPQNLEKKAKHVKATWAQRCNKVLF MSSEENKDFPAVGLKTKEGRDQLYWKTIKAFQYVHEHYLEDADWFLKADDDTYVILDNLR WLLSKYDPEEPIYFGRRFKPYVKQGYMSGGAGYVLSKEALKRFVDAFKTDKCTHSSSIED LALGRCMEIMNVEAGDSRDTIGKETFHPFVPEHHLIKGYLPRTFWYWNYNYYPPVEGPGC CSDLAVSFHYVDSTTMYELEYLVYHLRPYGYLYRYQPTLPERILKEISQANKNEDTKVKLGNP
HsCOSMC	>COSMC MLSESSSFLKGVMLGSIFCALITMLGHIRIGHGNRMHHHEHHHLQAPNKEDILKISEDER MELSKSFRVYCIILVKPKDVSLWAAVKETWTKHCDKAEFFSSENVKVFESINMDTNDMWL MMRKAYKYAFDKYRDQYNWFFLARPTTFAIIENLKYFLLKKDPSQPFYLGHTIKSGDLEY VGMEGGIVLSVESMKRLNSLLNIPEKCPEQGGMIWKISEDKQLAVCLKYAGVFAENAEDA DGKDVFNTKSVGLSIKEAMTYHPNQVVEGCCSDMAVTFNGLTPNQMHVMMYGVYRLRAFG HIFNDALVFLPPNGSDND
HsAlg1	>ALG1 MAASCLVLLALCLLLPLLLLGGWKRWRRGRAARHVVAVVLGDVGRSPRMQYHALSLAMHG FSVTLLGFCNSKPHDELLQNNRIQIVGLTELQSLAVGPRVFQYGVKVVLQAMYLLWKLMW REPGAYIFLQNPPGLPSIAVCWFVGCLCGSKLVIDWHNYGYSIMGLVHGPNHPLVLLAKW YEKFFGRLSHLNLCVTNAMREDLADNWHIRAVTVYDKPASFFKETPLDLQHRLFMKLGSM HSPFRARSEPEDPVTERSAFTERDAGSGLVTRLRERPALLVSSTSWTEDEDFSILLAALE KFEQLTLDGHNLPSLVCVITGKGPLREYYSRLIHQKHFQHIQVCTPWLEAEDYPLLLGSA DLGVCLHTSSSGLDLPMKVVDMFGCCLPVCAVNFKCLHELVKHEENGLVFEDSEELAAQL QMLFSNFPDPAGKLNQFRKNLRESQQLRWDESWVQTVLPLVMDT
HsAlg2	>ALG2 MAEEQGRERDSVPKPSVLFLHPDLGVGGAERLVLDAALALQARGCSVKIWTAHYDPGHCF AESRELPVRCAGDWLPRGLGWGGRGAAVCAYVRMVFLALYVLFLADEEFDVVVCDQVSAC IPVFRLARRRKKILFYCHFPDLLLTKRDSFLKRLYRAPIDWIEEYTTGMADCILVNSQFT AAVFKETFKSLSHIDPDVLYPSLNVTSFDSVVPEKLDDLVPKGKKFLLLSINRYERKKNL TLALEALVQLRGRLTSQDWERVHLIVAGGYDERVLENVEHYQELKKMVQQSDLGQYVTFL RSFSDKQKISLLHSCTCVLYTPSNEHFGIVPLEAMYMQCPVIAVNSGGPLESIDHSVTGF LCEPDPVHFSEAIEKFIREPSLKATMGLAGRARVKEKFSPEAFTEQLYRYVTKLLV

HsAlg3	>ALG3 MAAGLRKRGRSGSAAQAEGLCKQWLQRAWQERRLLLREPRYTLLVAACLCLAEVGITFWV IHRVAYTEIDWKAYMAEVEGVINGTYDYTQLQGDTGPLVYPAGFVYIFMGLYYATSRGTD IRMAQNIFAVLYLATLLLVFLIYHQTCKVPPFVFFFMCCASYRVHSIFVLRLFNDPVAMV LLFLSINLLLAQRWGWGCCFFSLAVSVKMNVLLFAPGLLFLLLTQFGFRGALPKLGICAG LQVVLGLPFLLENPSGYLSRSFDLGRQFLFHWTVNWRFLPEALFLHRAFHLALLTAHLTL LLLFALCRWHRTGESILSLLRDPSKRKVPPQPLTPNQIVSTLFTSNFIGICFSRSLHYQF YVWYFHTLPYLLWAMPARWLTHLLRLLVLGLIELSWNTYPSTSCSSAALHICHAVILLQL WLGPQPFPKSTQHSKKAH
HsAlg11	>ALG11 MAAGERSWCLCKLLRFFYSLFFPGLIVCGTLCVCLVIVLWGIRLLLQRKKKLVSTSKNGK NQMVIAFFHPYCNAGGGGERVLWCALRALQKKYPEAVYVVYTGDVNVNGQQILEGAFRRF NIRLIHPVQFVFLRKRYLVEDSLYPHFTLLGQSLGSIFLGWEALMQCVPDVYIDSMGYAF TLPLFKYIGGCQVGSYVHYPTISTDMLSVVKNQNIGFNNAAFITRNPFLSKVKLIYYYLF AFIYGLVGSCSDVVMVNSSWTLNHILSLWKVGNCTNIVYPPCDVQTFLDIPLHEKKMTPG HLLVSVGQFRPEKNHPLQIRAFAKLLNKKMVESPPSLKLVLIGGCRNKDDELRVNQLRRL SEDLGVQEYVEFKINIPFDELKNYLSEATIGLHTMWNEHFGIGVVECMAAGTIILAHNSG GPKLDIVVPHEGDITGFLAESEEDYAETIAHILSMSAEKRLQIRKSARASVSRFSDQEFE VTFLSSVEKLFK
HsAlg12	>ALG12 MAGKGSSGRRPLLLGLLVAVATVHLVICPYTKVEESFNLQATHDLLYHWQDLEQYDHLEF PGVVPRTFLGPVVIAVFSSPAVYVLSLLEMSKFYSQLIVRGVLGLGVIFGLWTLQKEVRR HFGAMVATMFCWVTAMQFHLMFYCTRTLPNVLALPVVLLALAAWLRHEWARFIWLSAFAI IVFRVELCLFLGLLLLLALGNRKVSVVRALRHAVPAGILCLGLTVAVDSYFWRQLTWPEG KVLWYNTVLNKSSNWGTSPLLWYFYSALPRGLGCSLLFIPLGLVDRRTHAPTVLALGFMA LYSLLPHKELRFIIYAFPMLNITAARGCSYLLNNYKKSWLYKAGSLLVIGHLVVNAAYSA TALYVSHFNYPGGVAMQRLHQLVPPQTDVLLHIDVAAAQTGVSRFLQVNSAWRYDKREDV QPGTGMLAYTHILMEAAPGLLALYRDTHRVLASVVGTTGVSLNLTQLPPFNVHLQTKLVL LERLPRPS
HsAlg13	>ALG13MKCVFVTVGTTSFDDLIACVSAPDSLQKIESLGYNRLILQIGRGTVVPEPFSTESFTLDV YRYKDSLKEDIQKADLVISHAGAGSCLETLEKGKPLVVVINEKLMNNHQLELAKQLHKEG HLFYCTCSTLPGLLQSMDLSTLKCYPPGQPEKFSAFLDKVVGLQK
HsAlg14	>ALG14 MVCVLVLAAAAGAVAVFLILRIWVVLRSMDVTPRESLSILVVAGSGGHTTEILRLLGSLS NAYSPRHYVIADTDEMSANKINSFELDRADRDPSNMYTKYYIHRIPRSREVQQSWPSTVF TTLHSMWLSFPLIHRVKPDLVLCNGPGTCVPICVSALLLGILGIKKVIIVYVESICRVET LSMSGKILFHLSDYFIVQWPALKEKYPKSVYLGRIV
HsDPM1	>DPM1 MASLEVSRSPRRSRRELEVRSPRQNKYSVLLPTYNERENLPLIVWLLVKSFSESGINYEI IIIDDGSPDGTRDVAEQLEKIYGSDRILLRPREKKLGLGTAYIHGMKHATGNYIIIMDAD LSHHPKFIPEFIRKQKEGNFDIVSGTRYKGNGGVYGWDLKRKIISRGANFLTQILLRPGA SDLTGSFRLYRKEVLEKLIEKCVSKGYVFQMEMIVRARQLNYTIGEVPISFVDRVYGESK LGGNEIVSFLKGLLTLFATT
HsPIGM	>PIGM MGSTKHWGEWLLNLKVAPAGVFGVAFLARVALVFYGVFQDRTLHVRYTDIDYQVFTDAAR FVTEGRSPYLRATYRYTPLLGWLLTPNIYLSELFGKFLFISCDLLTAFLLYRLLLLKGLG RRQACGYCVFWLLNPLPMAVSSRGNADSIVASLVLMVLYLIKKRLVACAAVFYGFAVHMK IYPVTYILPITLHLLPDRDNDKSLRQFRYTFQACLYELLKRLCNRAVLLFVAVAGLTFFA LSFGFYYEYGWEFLEHTYFYHLTRRDIRHNFSPYFYMLYLTAESKWSFSLGIAAFLPQLI LLSAVSFAYYRDLVFCCFLHTSIFVTFNKVCTSQYFLWYLCLLPLVMPLVRMPWKRAVVL LMLWFIGQAMWLAPAYVLEFQGKNTFLFIWLAGLFFLLINCSILIQIISHYKEEPLTERIKYD
HsPIGB	>PIGB MRRPLSKCGMEPGGGDASLTLHGLQNRSHGKIKLRKRKSTLYFNTQEKSARRRGDLLGEN IYLLLFTIALRILNCFLVQTSFVPDEYWQSLEVSHHMVFNYGYLTWEWTERLRSYTYPLI FASIYKILHLLGKDSVQLLIWIPRLAQALLSAVADVRLYSLMKQLENQEVARWVFFCQLC SWFTWYCCTRTLTNTMETVLTIIALFYYPLEGSKSMNSVKYSSLVALAFIIRPTAVILWT PLLFRHFCQEPRKLDLILHHFLPVGFVTLSLSLMIDRIFFGQWTLVQFNFLKFNVLQNWG TFYGSHPWHWYFSQGFPVILGTHLPFFIHGCYLAPKRYRILLVTVLWTLLVYSMLSHKEF

	RFIYPVLPFCMVFCGYSLTHLKTWKKPALSFLFLSNLFLALYTGLVHQRGTLDVMSHIQK VCYNNPNKSSASIFIMMPCHSTPYYSHVHCPLPMRFLQCPPDLTGKSHYLDEADVFYLNP LNWLHREFHDDASLPTHLITFSILEEEISAFLISSNYKRTAVFFHTHLPEGRIGSHIYVY ERKLKGKFNMKMKF
HsPIGZ	>PIGZ MQICGSSVASVAAGTSFQVLGPVCWQQLDLKMAVRVLWGGLSLLRVLWCLLPQTGYVHPD EFFQSPEVMAEDILGVQAARPWEFYPSSSCRSVLFPLLISGSTFWLLRLWEELGPWPGLV SGYALLVGPRLLLTALSFALDGAVYHLAPPMGADRWNALALLSGSYVTLVFYTRTFSNTI EGLLFTWLLVLVSSHVTWGPTRKEPAPGPRWRSWLLGGIVAAGFFNRPTFLAFAVVPLYL WGTRGATNPGLKSLTREALVLLPGAALTAAVFVATDSWYFSSPATSRNLVLTPVNFLHYN LNPQNLARHGTHARLTHLAVNGFLLFGVLHAQALQAAWQRLQVGLQASAQMGLLRALGAR SLLSSPRSYLLLLYFMPLALLSAFSHQEARFLIPLLVPLVLLCSPQTQPVPWKGTVVLFN ALGALLFGCLHQGGLVPGLEYLEQVVHAPVLPSTPTHYTLLFTHTYMPPRHLLHLPGLGA PVEVVDMGGTEDWALCQTLKSFTRQPACQVAGGPWLCRLFVVTPGTTRRAVEKCSFPFKN ETLLFPHLTLEDPPALSSLLSGAWRDHLSLHIVELGEET
HsAlg5	>ALG5 MAPLLLQLAVLGAALAAAALVLISIVAFTTATKMPALHRHEEEKFFLNAKGQKETLPSIW DSPTKQLSVVVPSYNEEKRLPVMMDEALSYLEKRQKRDPAFTYEVIVVDDGSKDQTSKVA FKYCQKYGSDKVRVITLVKNRGKGGAIRMGIFSSRGEKILMADADGATKFPDVEKLEKGL NDLQPWPNQMAIACGSRAHLEKESIAQRSYFRTLLMYGFHFLVWFLCVKGIRDTQCGFKL FTREAASRTFSSLHVERWAFDVELLYIAQFFKIPIAEIAVNWTEIEGSKLVPFWSWLQMG KDLLFIRLRYLTGAWRLEQTRKMN
HsAlg6	>ALG6 MEKWYLMTVVVLIGLTVRWTVSLNSYSGAGKPPMFGDYEAQRHWQEITFNLPVKQWYFNS SDNNLQYWGLDYPPLTAYHSLLCAYVAKFINPDWIALHTSRGYESQAHKLFMRTTVLIAD LLIYIPAVVLYCCCLKEISTKKKIANALCILLYPGLILIDYGHFQYNSVSLGFALWGVLG ISCDCDLLGSLAFCLAINYKQMELYHALPFFCFLLGKCFKKGLKGKGFVLLVKLACIVVA SFVLCWLPFFTEREQTLQVLRRLFPVDRGLFEDKVANIWCSFNVFLKIKDILPRHIQLIM SFCSTFLSLLPACIKLILQPSSKGFKFTLVSCALSFFLFSFQVHEKSILLVSLPVCLVLS EIPFMSTWFLLVSTFSMLPLLLKDELLMPSVVTTMAFFIACVTSFSIFEKTSEEELQLKS FSISVRKYLPCFTFLSRIIQYLFLISVITMVLLTLMTVTLDPPQKLPDLFSVLVCFVSCL NFLFFLVYFNIIIMWDSKSGRNQKKIS
HsAlg8	>ALG8 MAALTIATGTGNWFSALALGVTLLKCLLIPTYHSTDFEVHRNWLAITHSLPISQWYYEAT SEWTLDYPPFFAWFEYILSHVAKYFDQEMLNVHNLNYSSSRTLLFQRFSVIFMDVLFVYA VRECCKCIDGKKVGKELTEKPKFILSVLLLWNFGLLIVDHIHFQYNGFLFGLMLLSIARL FQKRHMEGAFLFAVLLHFKHIYLYVAPAYGVYLLRSYCFTANKPDGSIRWKSFSFVRVIS LGLVVFLVSALSLGPFLALNQLPQVFSRLFPFKRGLCHAYWAPNFWALYNALDKVLSVIG LKLKFLDPNNIPKASMTSGLVQQFQHTVLPSVTPLATLICTLIAILPSIFCLWFKPQGPR GFLRCLTLCALSSFMFGWHVHEKAILLAILPMSLLSVGKAGDASIFLILTTTGHYSLFPL LFTAPELPIKILLMLLFTIYSISSLKTLFRKEKPLFNWMETFYLLGLGPLEVCCEFVFPF TSWKVKYPFIPLLLTSVYCAVGITYAWFKLYVSVLIDSAIGKTKKQ
HsAlg10	>AlG10 MAQLEGYCFSAALSCTFLVSCLLFSAFSRALREPYMDEIFHLPQAQRYCEGHFSLSQWDP MITTLPGLYLVSVGVVKPAIWIFAWSEHVVCSIGMLRFVNLLFSVGNFYLLYLLFHKVQP RNKAASSIQRVLSTLTLAVFPTLYFFNFLYYTEAGSMFFTLFAYLMCLYGNHKTSAFLGF CGFMFRQTNIIWAVFCAGNVIAQKLTEAWKTELQKKEDRLPPIKGPFAEFRKILQFLLAY SMSFKNLSMLFCLTWPYILLGFLFCAFVVVNGGIVIGDRSSHEACLHFPQLFYFFSFTLF FSFPHLLSPSKIKTFLSLVWKHGILFLVVTLVSVFLVWKFTYAHKYLLADNRHYTFYVWK RVFQRYAILKYLLVPAYIFAGWSIADSLKSKPIFWNLMFFICLFIVIVPQKLLEFRYFIL PYVIYRLNITLPPTSRLVCELSCYAIVNFITFYIFLNKTFQWPNSQDIQRFMW
HsUGCG	>UGCG MALLDLALEGMAVFGFVLFLVLWLMHFMAIIYTRLHLNKKATDKQPYSKLPGVSLLKPLK GVDPNLINNLETFFELDYPKYEVLLCVQDHDDPAIDVCKKLLGKYPNVDARLFIGGKKVG INPKINNLMPGYEVAKYDLIWICDSGIRVIPDTLTDMVNQMTEKVGLVHGLPYVADRQGF AATLEQVYFGTSHPRYYISANVTGFKCVTGMSCLMRKDVLDQAGGLIAFAQYIAEDYFMA KAIADRGWRFAMSTQVAMQNSGSYSISQFQSRMIRWTKLRINMLPATIICEPISECFVAS

	LIIGWAAHHVFRWDIMVFFMCHCLAWFIFDYIQLRGVQGGTLCFSKLDYAVAWFIRESMT IYIFLSALWDPTISWRTGRYRLRCGGTAEEILDV
HsB3GLCT	>B3GLCT MRPPACWWLLAPPALLALLTCSLAFGLASEDTKKEVKQSQDLEKSGISRKNDIDLKGIVF VIQSQSNSFHAKRAEQLKKSILKQAADLTQELPSVLLLHQLAKQEGAWTILPLLPHFSVT YSRNSSWIFFCEEETRIQIPKLLETLRRYDPSKEWFLGKALHDEEATIIHHYAFSENPTV FKYPDFAAGWALSIPLVNKLTKRLKSESLKSDFTIDLKHEIALYIWDKGGGPPLTPVPEF CTNDVDFYCATTFHSFLPLCRKPVKKKDIFVAVKTCKKFHGDRIPIVKQTWESQASLIEY YSDYTENSIPTVDLGIPNTDRGHCGKTFAILERFLNRSQDKTAWLVIVDDDTLISISRLQ HLLSCYDSGEPVFLGERYGYGLGTGGYSYITGGGGMVFSREAVRRLLASKCRCYSNDAPD DMVLGMCFSGLGIPVTHSPLFHQARPVDYPKDYLSHQVPISFHKHWNIDPVKVYFTWLAP SDEDKARQETQKGFREEL
HsGLYG	>GLYG MTDQAFVTLTTNDAYAKGALVLGSSLKQHRTTRRLVVLATPQVSDSMRKVLETVFDEVIM VDVLDSGDSAHLTLMKRPELGVTLTKLHCWSLTQYSKCVFMDADTLVLANIDDLFDREEL SAAPDPGWPDCFNSGVFVYQPSVETYNQLLHLASEQGSFDGGDQGILNTFFSSWATTDIR KHLPFIYNLSSISIYSYLPAFKVFGASAKVVHFLGRVKPWNYTYDPKTKSVKSEAHDPNM THPEFLILWWNIFTTNVLPLLQQFGLVKDTCSYVNVLSDLVYTLAFSCGFCRKEDVSGAI SHLSLGEIPAMAQPFVSSEERKERWEQGQADYMGADSFDNIKRKLDTYLQ
HsPOGLUT1	>POGLUT1 MEWWASSPLRLWLLLFLLPSAQGRQKESGSKWKVFIDQINRSLENYEPCSSQNCSCYHGV IEEDLTPFRGGISRKMMAEVVRRKLGTHYQITKNRLYRENDCMFPSRCSGVEHFILEVIG RLPDMEMVINVRDYPQVPKWMEPAIPVFSFSKTSEYHDIMYPAWTFWEGGPAVWPIYPTG LGRWDLFREDLVRSAAQWPWKKKNSTAYFRGSRTSPERDPLILLSRKNPKLVDAEYTKNQ AWKSMKDTLGKPAAKDVHLVDHCKYKYLFNFRGVAASFRFKHLFLCGSLVFHVGDEWLEF FYPQLKPWVHYIPVKTDLSNVQELLQFVKANDDVAQEIAERGSQFIRNHLQMDDITCYWE NLLSEYSKFLSYNVTRRKGYDQIIPKMLKTEL
HsGnT1	>MGAT1 MLKKQSAGLVLWGAILFVAWNALLLLFFWTRPAPGRPPSVSALDGDPASLTREVIRLAQD AEVELERQRGLLQQIGDALSSQRGRVPTAAPPAQPRVPVTPAPAVIPILVIACDRSTVRR CLDKLLHYRPSAELFPIIVSQDCGHEETAQAIASYGSAVTHIRQPDLSSIAVPPDHRKFQ GYYKIARHYRWALGQVFRQFRFPAAVVVEDDLEVAPDFFEYFRATYPLLKADPSLWCVSA WNDNGKEQMVDASRPELLYRTDFFPGLGWLLLAELWAELEPKWPKAFWDDWMRRPEQRQG RACIRPEISRTMTFGRKGVSHGQFFDQHLKFIKLNQQFVHFTQLDLSYLQREAYDRDFLA RVYGAPQLQVEKVRTNDRKELGEVRVQYTGRDSFKAFAKALGVMDDLKSGVPRAGYRGIV TFQFRGRRVHLAPPLTWEGYDPSWN
HsGnT2	>MGAT2 MRFRIYKRKVLILTLVVAACGFVLWSSNGRQRKNEALAPPLLDAEPARGAGGRGGDHPSV AVGIRRVSNVSAASLVPAVPQPEADNLTLRYRSLVYQLNFDQTLRNVDKAGTWAPRELVL VVQVHNRPEYLRLLLDSLRKAQGIDNVLVIFSHDFWSTEINQLIAGVNFCPVLQVFFPFS IQLYPNEFPGSDPRDCPRDLPKNAALKLGCINAEYPDSFGHYREAKFSQTKHHWWWKLHF VWERVKILRDYAGLILFLEEDHYLAPDFYHVFKKMWKLKQQECPECDVLSLGTYSASRSF YGMADKVDVKTWKSTEHNMGLALTRNAYQKLIECTDTFCTYDDYNWDWTLQYLTVSCLPK FWKVLVPQIPRIFHAGDCGMHHKKTCRPSTQSAQIESLLNNNKQYMFPETLTISEKFTVV AISPPRKNGGWGDIRDHELCKSYRRLQ
HsGnT3	>MGAT3 MKMRRYKLFLMFCMAGLCLISFLHFFKTLSYVTFPRELASLSPNLVSSFFWNNAPVTPQA SPEPGGPDLLRTPLYSHSPLLQPLPPSKAAEELHRVDLVLPEDTTEYFVRTKAGGVCFKP GTKMLERPPPGRPEEKPEGANGSSARRPPRYLLSARERTGGRGARRKWVECVCLPGWHGP SCGVPTVVQYSNLPTKERLVPREVPRRVINAINVNHEFDLLDVRFHELGDVVDAFVVCES NFTAYGEPRPLKFREMLTNGTFEYIRHKVLYVFLDHFPPGGRQDGWIADDYLRTFLTQDG VSRLRNLRPDDVFIIDDADEIPARDGVLFLKLYDGWTEPFAFHMRKSLYGFFWKQPGTLE VVSGCTVDMLQAVYGLDGIRLRRRQYYTMPNFRQYENRTGHILVQWSLGSPLHFAGWHCS WCFTPEGIYFKLVSAQNGDFPRWGDYEDKRDLNYIRGLIRTGGWFDGTQQEYPPADPSEH MYAPKYLLKNYDRFHYLLDNPYQEPRSTAAGGWRHRGPEGRPPARGKLDEAEV
HsGnT4a	>MGAT4A MRLRNGTVATALAFITSFLTLSWYTTWQNGKEKLIAYQREFLALKERLRIAEHRISQRSS ELNTIVQQFKRVGAETNGSKDALNKFSDNTLKLLKELTSKKSLQVPSIYYHLPHLLKNEG

	SLQPAVQIGNGRTGVSIVMGIPTVKREVKSYLIETLHSLIDNLYPEEKLDCVIVVFIGET DIDYVHGVVANLEKEFSKEISSGLVEVISPPESYYPDLTNLKETFGDSKERVRWRTKQNL DYCFLMMYAQEKGIYYIQLEDDIIVKQNYFNTIKNFALQLSSEEWMILEFSQLGFIGKMF QAPDLTLIVEFIFMFYKEKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRFRPSLFQ HVGLHSSLSGKIQKLTDKDYMKPLLLKIHVNPPAEVSTSLKVYQGHTLEKTYMGEDFFWA ITPIAGDYILFKFDKPVNVESYLFHSGNQEHPGDILLNTTVEVLPFKSEGLEISKETKDK RLEDGYFRIGKFENGVAEGMVDPSLNPISAFRLSVIQNSAVWAILNEIHIKKATN
HsGCNT1	>GCNT1 MLRTLLRRRLFSYPTKYYFMVLVLSLITFSVLRIHQKPEFVSVRHLELAGENPSSDINCT KVLQGDVNEIQKVKLEILTVKFKKRPRWTPDDYINMTSDCSSFIKRRKYIVEPLSKEEAE FPIAYSIVVHHKIEMLDRLLRAIYMPQNFYCIHVDTKSEDSYLAAVMGIASCFSNVFVAS RLESVVYASWSRVQADLNCMKDLYAMSANWKYLINLCGMDFPIKTNLEIVRKLKLLMGEN NLETERMPSHKEERWKKRYEVVNGKLTNTGTVKMLPPLETPLFSGSAYFVVSREYVGYVL QNEKIQKLMEWAQDTYSPDEYLWATIQRIPEVPGSLPASHKYDLSDMQAVARFVKWQYFE GDVSKGAPYPPCDGVHVRSVCIFGAGDLNWMLRKHHLFANKFDVDVDLFAIQCLDEHLRH KALETLKH
HsGCNT2	>GCNT2 MMGSWKHCLFSASLISALIFVFVYNTELWENKRFLRAALSNASLLAEACHQIFEGKVFYP TENALKTTLDEATCYEYMVRSHYVTETLSEEEAGFPLAYTVTIHKDFGTFERLFRAIYMP QNVYCVHLDQKATDAFKGAVKQLLSCFPNAFLASKKESVVYGGISRLQADLNCLEDLVAS EVPWKYVINTCGQDFPLKTNREIVQYLKGFKGKNITPGVLPPDHAVGRTKYVHQELLNHK NSYVIKTTKLKTPPPHDMVIYFGTAYVALTRDFANFVLQDQLALDLLSWSKDTYSPDEHF WVTLNRIPGVPGSMPNASWTGNLRAIKWSDMEDRHGGCHGHYVHGICIYGNGDLKWLVNS PSLFANKFELNTYPLTVECLELRHRERTLNQSETAIQPSWYF
HsB3GNT2	>B3GNT2 MSVGRRRIKLLGILMMANVFIYFIMEVSKSSSQEKNGKGEVIIPKEKFWKISTPPEAYWN REQEKLNRQYNPILSMLTNQTGEAGRLSNISHLNYCEPDLRVTSVVTGFNNLPDRFKDFL LYLRCRNYSLLIDQPDKCAKKPFLLLAIKSLTPHFARRQAIRESWGQESNAGNQTVVRVF LLGQTPPEDNHPDLSDMLKFESEKHQDILMWNYRDTFFNLSLKEVLFLRWVSTSCPDTEF VFKGDDDVFVNTHHILNYLNSLSKTKAKDLFIGDVIHNAGPHRDKKLKYYIPEVVYSGLY PPYAGGGGFLYSGHLALRLYHITDQVHLYPIDDVYTGMCLQKLGLVPEKHKGFRTFDIEE KNKNNICSYVDLMLVHSRKPQEMIDIWSQLQSAHLKC
HsB3GNT6	>B3GNT6 MAFPCRRSLTAKTLACLLVGVSFLALQQWFLQAPRSPREERSPQEETPEGPTDAPAADEP PSELVPGPPCVANASANATADFEQLPARIQDFLRYRHCRHFPLLWDAPAKCAGGRGVFLL LAVKSAPEHYERRELIRRTWGQERSYGGRPVRRLFLLGTPGPEDEARAERLAELVALEAR EHGDVLQWAFADTFLNLTLKHLHLLDWLAARCPHARFLLSGDDDVFVHTANVVRFLQAQP PGRHLFSGQLMEGSVPIRDSWSKYFVPPQLFPGSAYPVYCSGGGFLLSGPTARALRAAAR HTPLFPIDDAYMGMCLERAGLAPSGHEGIRPFGVQLPGAQQSSFDPCMYRELLLVHRFAP YEMLLMWKALHSPALSCDRGHRVS
HsPIGA	>PIGA MACRGGAGNGHRASATLSRVSPGSLYTCRTRTHNICMVSDFFYPNMGGVESHIYQLSQCL IERGHKVIIVTHAYGNRKGIRYLTSGLKVYYLPLKVMYNQSTATTLFHSLPLLRYIFVRE RVTIIHSHSSFSAMAHDALFHAKTMGLQTVFTDHSLFGFADVSSVLTNKLLTVSLCDTNH IICVSYTSKENTVLRAALNPEIVSVIPNAVDPTDFTPDPFRRHDSITIVVVSRLVYRKGI DLLSGIIPELCQKYPDLNFIIGGEGPKRIILEEVRERYQLHDRVRLLGALEHKDVRNVLV QGHIFLNTSLTEAFCMAIVEAASCGLQVVSTRVGGIPEVLPENLIILCEPSVKSLCEGLE KAIFQLKSGTLPAPENIHNIVKTFYTWRNVAERTEKVYDRVSVEAVLPMDKRLDRLISHC GPVTGYIFALLAVFNFLFLIFLRWMTPDSIIDVAIDATGPRGAWTNNYSHSKRGGENNEI SETR
HsXXLT1	>XXLT1 MGLLRGGLPCARAMARLGAVRSHYCALLLAAALAVCAFYYLGSGRETFSSATKRLKEARA GAPAAPSPPALELARGSVAPAPGAKAKSLEGGGAGPVDYHLLMMFTKAEHNAALQAKARV ALRSLLRLAKFEAHEVLNLHFVSEEASREVAKGLLRELLPPAAGFKCKVIFHDVAVLTDK LFPIVEAMQKHFSAGLGTYYSDSIFFLSVAMHQIMPKEILQIIQLDLDLKFKTNIRELFE EFDSFLPGAIIGIAREMQPVYRHTFWQFRHENPQTRVGGPPPEGLPGFNSGVMLLNLEAM RQSPLYSRLLEPAQVQQLADKYHFRGHLGDQDFFTMIGMEHPKLFHVLDCTWNRQLCTWW RDHGYSDVFEAYFRCEGHVKIYHGNCNTPIPED

HsUGT1A1	>UGT1A1 MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSMLGAIQQLQQRGHEIVVLAPD ASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFENDSFLQRVIKTYKKIKKDS AMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIVAQYLSLPTVFFLHALPCSLE FEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFSQNFLCDVVYSPYATLASEFLQR EVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVGGINCLHQNPLSQEFEAYINASGEH GIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQNDL LGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMTS EDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAHD LTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
HsB4GAT1	>B4GAT1 MQMSYAIRCAFYQLLLAALMLVAMLQLLYLSLLSGLHGQEEQDQYFEFFPPSPRSVDQVKAQL RTALASGGVLDASGDYRVYRGLLKTTMDPNDVILATHASVDNLLHLSGLLERWEGPLSVSVFA ATKEEAQLATVLAYALSSHCPDMRARVAMHLVCPSRYEAAVPDPREPGEFALLR SCQEVFDKLARVAQPGINYALGTNVSYPNNLLRNLAREGANYALVIDVDMVPSEGLWRGL REMLDQSNQWGGTALVVPAFEIRRARRMPMNKNELVQLYQVGEVRPFYYGLCTPCQAPTN YSRWVNLPEESLLRPAYVVPWQDPWEPFYVAGGKVPTFDERFRQYGFNRISQACELHVAG FDFEVLNEGFLVHKGFKEALKFHPQKEAENQHNKILYRQFKQELKAKYPNSPRRC
HsUGT1A3	>UGT1A3 MATGLQVPLPWLATGLLLLLSVQPWAESGKVLVVPIDGSHWLSMREVLRELHARGHQAVVLT PEVNMHIKEENFFTLTTYAISWTQDEFDRHVLGHTQLYFETEHFLKKFFRSMAMLNNMSLVYH RSCVELLHNEALIRHLNATSFDVVLTDPVNLCAAVLAKYLSIPTVFFLRNIPCDL DFKGTQCPNPSSYIPRLLTTNSDHMTFMQRVKNMLYPLALSYICHAFSAPYASLASELFQ REVSVVDILSHASVWLFRGDFVMDYPRPIMPNMVFIGGINCANRKPLSQEFEAYINASGE HGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVKWLPQND LLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTLNVLEMT SEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPHLRPAAH DLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
CjCsTII	>CjCSTII MKNNIIIAGNGSSLKEIDYSRLPNDFDVFRCNQFYFEDKYYLGKKCKAVFYNSTLFFEQYYTLKH LIQNQEYETELIMCSNFNQAHLENENFLKNFYDYFPDAHLGYDFFKQLKDFNAYFKFHEIYFNQ RITSGIYMCAVAIALGYKEIYLSGIDFYQNGSSYAFDTKQKNLLKLAPNFKNDNSHYIGHSKNTD IKALEFLEKTYKIKLYCLCPNSLLANFIELAPNLNSNFTIQEKNNYTKDILIPSKEAYKKFTKMTKT QEILKNNIYYKLTKDLLRLPSDIKHYFKGK
NmPolysiaT	>NmPST MGLKKIKKALFQPKKFFQDSMWLTTSPFYLTPPRNNLFVISNLGQLNQVQSLIKIQKLTNNLLVI LYTSKNLKMPKLVHQSANKNLFESIYLFELPRSPNNITPKKLLYIYRSYKKILNIIQPAHLYMLSFT GHYSYLISIAKKKNITTHLIDEGTGTYAPLLESFSYHPTKLERYLIGNNLNIKGYIDHFDILHVPFPE YAKKIFNAKKYNRFFAHAGGISINNNIANLQKKYQISKNDYIFVSQRYPISDDLYYKSIVEILNSIS LQIKGKIFIKLHPKEMGNNYVMSLFLNMVEINPRLVVINEPPFLIEPLIYLTNPKGIIGLASSSLIYTP LLSPSTQCLSIGELIINLIQKYSMVENTEMIQEHLEIIKKFNFINILNDLNGVISNPLFKTEETFETLL KSAEFAYKSKNYFQAIFYWQLASKNNITLLGHKALWYYNALYNVKQIYKMEYSDIFYIDNISVD FHSKDKLTWEKIKHYYYSADNRIGRDR
CjCgtB	>CjCgtB MGFKISIILPTYNVEQYIARAIESCIKQTFKDIEIIVVDDCGNDNSINIAKEYSKKDKRIKIIHNEKNI GLLRARYEGVKVANSPYIMFLDPDDYLELNACEECIKILDEQDEVDLVFFNAIVESNVISYKKFDF NSGFYSKKEFVKKIIAEKNLYWTMWGKLIRKKLYLEAFASLRLEKDVKINMAEDVLLYYPMLSQ AQKIAYMNCNLYHYVPNNNSICNTKNEVGVKNNIQDLQLVLNYLRQNYILNKYCSVLYVLIKY LLYIQIYKIKRTKLMVTLLAKINILTLKILFKYKKFLKQC
HpLgtB	>HpLgtB MRVFIISLNQKVCDQFGLVFRDTTTLLHNINATHHKAQIFDAIYSKTFESELHPLVKKHLHPYFIT QNIKDMGITTNLISRVSKFYYALKYHAKFMSFGELGCYASHYSLWEKCIELNEP ICILEDDITLKEDFKEGLDFLEKHIQELGYARLMYLLYDANVKSEPLSHKNHEIQERVGI IKAYSHGVGTQGYVITPKIAKVFKKCSRKWVVPVDTIMDATFIHGVKNLVLQPFVIADDE QISTIARKEEPYSSKIALMRKLHFKYLKYWQFV
NmLgtB	>NmLgtB MQNHVISLASAAERRAHIADTFGRHGIPFQFFDALMPSERLEQAMAELVPGLSAHPYLSGVEKA CFMSHAVLWKQALDEGLPYITVFEDDVLLGEGAEKFLAEDAWLQERFDPDTAFIVR

	LETMFMHVLTSPSGVADYCGRAFPLLESEHWGTAGYIISRKAMRFFLDRFAALPPEGLHP VDLMMFSDFFDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRLLNRKQQRRDS PANTFKHRLIRALTKISREREKRRQRREQFIVPFQ
NgLgtB	>NgLgtB MQNHVISLASAAERRAHIAATFGSRGIPFQFFDALMPSERLERAMAELVPGLSAHPYLSGVEKA CFMSHAVLWEQALDEGVPYIAVFEDDVLLGEGAEQFLAEDTWLQERFDPDSAFVVR LETMFMHVLTSPSGVADYGGRAFPLLESEHCGTAGYIISRKAMRFFLDRFAVLPPERLHP VDLMMFGNPDDREGMPVCQLNPALCAQELHYAKFHDQNSALGSLIEHDRRLNRKQQWRDS PANTFKHRLIRALTKIGRREREKRRQRREQLIGKIIVPFQ
EcFUT	>EcFUT MSIIRLQGGLGNQLFQFSFGYALSKINGTPLYFDISHYAENDDHGGYRLNNLQIPEEYLQ YYTPKINNIYKFLVRGSRLYPEIFLFLGFCNEFHAYGYDFEYIAQKWKSKKYIGYWQSEH FFHKHILDLKEFFIPKNVSEQANLLAAKILESQSSLSIHIRRGDYIKNKTATLTHGVCSL EYYKKALNKIRDLAMIRDVFIFSDDIFWCKENIETLLSKKYNIYYSEDLSQEEDLWLMSL ANHHIIANSSFSWWGAYLGTSASQIVIYPTPWYDITPKNTYIPIVNHWINVDKHSSC
EcWecA	>EcWecA MNLLTVSTDLISIFLFTTLFLFFARKVAKKVGLVDKPNFRKRHQGLIPLVGGISVYAGIC FTFGIVDYYIPHASLYLACAGVLVFIGALDDRFDISVKIRATIQAAVGIVMMVFGKLYLS SLGYIFGSWEMVLGPFGYFLTLFAVWAAINAFNMVDGIDGLLGGLSCVSFAAIGMILWFD GQTSLAIWCFAMIAAILPYIMLNLGILGRRYKVFMGDAGSTLIGFTVIWILLETTQGKTH PISPVTALWIIAIPLMDMVAIMYRRLRKGMSPFSPDRQHIHHLIMRAGFTSRQAFVLITL AAALLASIGVLAEYSHFVPEWVMLVLFLLAFFLYGYCIKRAWKVARFIKRVKRRLRRNRG GSPNLTK
LpSetA	>LpSetA MYKIYSYLGWRIDMKTENLPQAGQEAQIDKKIHFIWVGHIMPQKNIQVVSEWAEKNPGYETII WVDKKIAPAKELDLFILDMKSKGITVKDINEEGVCRDSIRHELDQESPNYGMVSDML RLNILAAEGGIYLDSDILCSAPFPDEIYAPFGFLLSPWSQGANNTLCNDIILCSKGNQII QQLADAIEQSYIARDSFEFTHEYASMKETKGERIAKTLGVTGPGFLFHQLKKMGILNDKS EMEAIHWELQDQRYLIDGSVKEPDYFYVPQNNTNDASWVPSIKRPGIENMSFQERLENAV QLIAFDIQKTGLFNLDHYANELKVKQNSWCIAAETSPELKPDSYLLIRPRDKTGEWTLYY VDEDKKLNPVTLPVIKGAIKLSEVSDPLRKFHTLLSQVSDPVNPTAHELKQIGRALIELK PRQDEWHCKNKWSGAEEIAQELWQRITSNETLRAQIKQCFTQFESLKPRVAELGLTRASG AGTEVEAHESTVKEQEIISQNTVGEEGTKEKNSVQLASENSSDEKIKTAHDLIDEIIQDV IQLDGKLGLLGGNTRQLEDGRVINIPNGAAMIFDDYKKYKQGELTAESALESMIKIAKLS NQLNRHTFFNQRQPETGQFYKKVAAIDLQTTIAAEYDNNHGLRI
NmSynE	>NmSynE MLQKIRKALFHPKKFFQDSQWFATPLFSSFAPKSNLFIISTFAQLNQAHSLTKMQKLKNNLLVIL YTTQNMKMPKLIQKSVDKELFSVTYMFELPRKPGIVSPKKFLYIQRGYKKLLKTI QPAHLYVMSFAGHYSSLLSLAKKMNITTHLVEEGTATYAPLLESFTYKPTKFEQRFVGNN LHQKGYFDKFDILHVAFPEYAKKIFNANEYHRFFAHSGGISTSQSIAKIQDKYRISQNDY IFVSQRYPVSDEVYYKTIVETLNQMSLRIEGKIFIKLHPKEMENKNIMSLFLNMVTINPR LVVINEPPFLIDPLIYLTTPKGIIGLTSTSIVYTPLLSPTTQCLSIGQIVIDSIHHTAQQ ENTALIEEHLEIVKQFDFIKILSSIEDGIDTNSFKTEETLEMLLKSAEYAYKNKNFYQAI FYWQLASNNDLSVLGYKSLWYYNALNKVKQNYKMKYLEINYIERISLYFNDKDKMIWQNI KNDFFKYSLCNQ
ScAlg1	>ScAlg1 MFLEIPRWLLALIILYLSIPLVVYYVIPYLFYGNKSTKKRIIIFVLGDVGHSPRICYHAI SFSKLGWQVELCGYVEDTLPKIISSDPNITVHHMSNLKRKGGGTSVIFMVKKVLFQVLSI FKLLWELRGSDYILVQNPPSIPILPIAVLYKLTGCKLIIDWHNLAYSILQLKFKGNFYHP LVLISYMVEMIFSKFADYNLTVTEAMRKYLIQSFHLNPKRCAVLYDRPASQFQPLAGDIS RQKALTTKAFIKDYIRDDFDTEKGDKIIVTSTSFTPDEDIGILLGALKIYENSYVKFDSS LPKILCFITGKGPLKEKYMKQVEEYDWKRCQIEFVWLSAEDYPKLLQLCDYGVSLHTSSS GLDLPMKILDMFGSGLPVIAMNYPVLDELVQHNVNGLKFVDRRELHESLIFAMKDADLYQ KLKKNVTQEAENRWQSNWERTMRGLKLIH
ScAlg11	>ScAlg11 MGSAWTNYNFEEVKSHFGFKKYVVSSLVLVYGLIKVLTWIFRQWVYSSLNPFSKKSSLLNRAVA SCGEKNVKVFGFFHPYCNAGGGGEKVLWKAVDITLRKDAKNVIVIYSGDFVNGENV TPENILNNVKAKFDYDLDSDRIFFISLKLRYLVDSSTWKHFTLIGQAIGSMILAFESIIQ

	CPPDIWIDTMGYPFSYPIIARFLRRIPIVTYTHYPIMSKDMLNKLFKMPKKGIKVYGKIL YWKVFMLIYQSIGSKIDIVITNSTWTNNHIKQIWQSNTCKIIYPPCSTEKLVDWKQKFGT AKGERLNQAIVLAQFRPEKRHKLIIESFATFLKNLPDSVSPIKLIMAGSTRSKQDENYVK SLQDWSENVLKIPKHLISFEKNLPFDKIEILLNKSTFGVNAMWNEHFGIAVVEYMASGLI PIVHASAGPLLDIVTPWDANGNIGKAPPQWELQKKYFAKLEDDGETTGFFFKEPSDPDYN TTKDPLRYPNLSDLFLQITKLDYDCLRVMGARNQQYSLYKFSDLKFDKDWENFVLNPICK LLEEEERG
NtGnTI	>NtGnT1 MRGNKFCCDFRYLLILAAVAFIYTQMRLFATQSEYADRLAAAIEAENHCTSQTRLLIDQISLQQG RIVALEEQMKRQDQECRQLRALVQDLESKGIKKLIGNVQMPVAAVVVMACNRADY LEKTIKSILKYQISVASKYPLFISQDGSHPDVRKLALSYDQLTYMQHLDFEPVHTERPGE LIAYYKIARHYKWALDQLFYKHNFSRVIILEDDMEIAPDFFDFFEAGATLLDRDKSIMAI SSWNDNGQMQFVQDPYALYRSDFFPGLGWMLSKSTWDELSPKWPKAYWDDWLRLKENHRG RQFIRPEVCRTYNFGEHGSSLGQFFKQYLEPIKLNDVQVDWKSMDLSYLLEDNYVKHFGD LVKKAKPIHGADAVLKAFNIDGDVRIQYRDQLDFENIARQFGIFEEWKDGVPRAAYKGIV VFRYQTSRRVFLVGHDSLQQLGIEDT
NtGnTII	>NtGnT2 MASCKKARIKDAAFRRLVSLALITVSGVVLLLILLRTNTVSNLINSYTNESSESFESIEE AKVSNTHLTRIPKLPLQNELSILLSKRNQMPPRNLNLYSKLAKDHVVIVLYVHNRPQYLQ IVIDSLSRVEGISETLLIVSHDGYFEEMNKIVESIKFCQVKQIFAPYSPHIFSDSFPGVS PKDCKEKDDPVEKHCEGMPDQYGNHRSPKIVSLKHHWWWMMNTVWDGLEVTRHHSGHILF IEEDHFIYPNAYRNMQLLAELKSKKCPDCYAANLAPSEVKLRGEGWECLVAERMGNVGYA FNRTVWRKIHKKAAEFCSFDDYNWDITMWSTVYPSFGSPVYTLRGSRTSAVHFGKCGLHQ GHNRNVACMDNGGVNIVVDEIDKVANIKPEWEVRKYEHQAGYKAGFKGWGGWGDVRDREL CMEFAKMYI
BtGGTA1	>BtGGTA1 MNVKGKVILSMLVVSTVIVVFWEYIHSPEGSLFWINPSRNPEVGGSSIQKGWWLPRWFNNGYHE EDGDINEEKEQRNEDESKLKLSDWFNPFKRPEVVTMTKWKAPVVWEGTYNRAVLDNYYAKQ KITVGLTVFAVGRYIEHYLEEFLTSANKHFMVGHPVIFYIMVDDVSRMPLIELGP LRSFKVFKIKPEKRWQDISMMRMKTIGEHIVAHIQHEVDFLFCMDVDQVFQDKFGVETLG ESVAQLQAWWYKADPNDFTYERRKESAAYIPFGEGDFYYHAAIFGGTPTQVLNITQECFK GILKDKKNDIEAQWHDESHLNKYFLLNKPTKILSPEYCWDYHIGLPADIKLVKMSWQTKE YNVVRNNV
MmGGTA1	>MmGGTA1 MITMLQDLHVNKISMSRSKSETSLPSSRSGSQEKIMNVKGKVILLMLIVSTVVVVFWEYVNRIPEV GENRWQKDWWFPSWFKNGTHSYQEDNVEGRREKGRNGDRIEEPQLWDWFNPKNRPDVLTV TPWKAPIVWEGTYDTALLEKYYATQKLTVGLTVFAVGKYIEHYLEDFLESADMY FMVGHRVIFYVMIDDTSRMPVVHLNPLHSLQVFEIRSEKRWQDISMMRMKTIGEHILAHI QHEVDFLFCMDVDQVFQDNFGVETLGQLVAQLQAWWYKASPEKFTYERRELSAAYIPFGE GDFYYHAAIFGGTPTHILNLTRECFKGILQDKKHDIEAQWHDESHLNKYFLFNKPTKILS PEYCWDYQIGLPSDIKSVKVAWQTKEYNLVRNNV
RnGGTA1	>RnGGTA1 MVTMPQDLHVKVSMSRSKSETSLLSSRSGSQEKIMNVKGKIILSVLMVSTVLVVFWEYVNRTHSY QEEDIERAREKGRNGDSIVEPQLWDWFNPKNRPEVLTVTPWKAPIVWEGTYDTAL LEKYYARQKITVGLTVFAVGKYIEHYLEDFLESANKYFMVGHRVIFYVMMDDTSRMPAVH LSPLHSLQVFEIRSEKRWQDISMMRMKTIGEHILDHIQHEVDFLFCMDVDQVFQDNFGVE TLGQLVAQLQAWWYKASPDEFTYERRELSAAYIPFGEGDFYYHAAVFGGTPVHILNLTRE CFKGILQDKKHDIEAQWHDESHLNKYFLFNKPTKILSPEYCWDYHIGLPSDIKNVKIAWQ TKEYNLVRSNV
BtB4GalT1	>B4GT1 MKFREPLLGGSAAMPGASLQRACRLLVAVCALHLGVTLVYYLAGRDLRRLPQLVGVHPPLQGS SHGAAAIGQPSGELRLRGVAPPPPLQNSSKPRSRAPSNLDAYSHPGPGPGPGSNLTS APVPSTTTRSLTACPEESPLLVGPMLIEFNIPVDLKLVEQQNPKVKLGGRYTPMDCISPH KVAIIIPFRNRQEHLKYWLYYLHPILQRQQLDYGIYVINQAGESMFNRAKLLNVGFKEAL KDYDYNCFVFSDVDLIPMNDHNTYRCFSQPRHISVAMDKFGFSLPYVQYFGGVSALSKQQ FLSINGFPNNYWGWGGEDDDIYNRLAFRGMSVSRPNAVIGKCRMIRHSRDKKNEPNPQRF DRIAHTKETMLSDGLNSLTYMVLEVQRYPLYTKITVDIGTPS

CHAPTER 9

APPENDIX D[16]

Technical notes and the detailed protocols for the development of cell-free synthetic glycobiology system.

9.1 Materials

Media

1. Luria-Bertani (LB) broth: 1.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, autoclave sterile.
2. LB agar: add 1.5% (w/v) agar to LB prepared as above. Aliquot 30 mL of LB agar per 150 mm petri dish into a plastic conical tube. Add appropriate antibiotics and sterile 0.2% (w/v) D-glucose. Mix and pour into petri dishes. For induction plates, omit glucose and add sterile 0.2% (w/v) L-arabinose and 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG).
3. Terrific broth (TB): 1.2% (w/v) tryptone, 2.4% (w/v) yeast extract, 0.4% (v/v) glycerol, 10% (v/v) phosphate buffer (0.17 M KH_2PO_4, 0.72 M K_2HPO_4), autoclave

16 Parts of this chapter appeared in the Methods in Enzymology journal:

Jaroentomeechai, T., Stark, J.C., Jewett, M.C. and DeLisa, M.P. (2017) A pipeline for studying and engineering single-subunit oligosaccharyltransferases. *Methods Enzymol* 597: 55-81.

Several other parts in this chapter are under preparation for submission to Natuer Protocol journal.

sterile. Autoclave phosphate buffer separately from other components and add to the broth prior use.

4. 2xYTPG broth: 1.6% (w/v) tryptone, 1.0% (w/v) yeast extract, 0.5% (w/v) NaCl, 1.8% (w/v) glucose, 0.7% (w/v) K_2HPO_4, 0.3% (w/v) KH_2PO_4, adjust pH to 7.3 with 5 N KOH and autoclave sterile. Autoclave 40% (w/v) glucose stock separately and add to the broth prior use.

Media supplements

1. Transformation and storage solution (1×TSS): supplement LB broth with 10% (w/v) polyethylene glycol (PEG)-8000, 5% (v/v) dimethylsulfoxide (DMSO), and 20 mM $MgSO_4$; adjust pH to 6.5 with HCl, and autoclave sterile.
2. Antibiotics: Ampicillin (Amp) is used at 100 μg/mL. To make a 1,000× stock, mix 1 g in 10 mL nanopure water. Chloramphenicol (Cam) is used at 20 μg/ mL. To make a 1,000× stock, dissolve 0.2 g in 10 mL ethanol. Trimethoprim (Tp) is used at 100 μg/mL. To make a 500× stock, dissolve 0.5 g in 10 mL DMSO. Kanamycin (Km) is used at 50 mg/mL. To make a 1,000× stock, dissolve 0.5 g in 10 mL nanopure water. All antibiotic stock is filter sterile.
3. Inducers/repressors: 20% (w/v) L-arabinose stock, 20% (w/v) D-glucose, and 0.1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG) stock. All inducer/repressor stocks are made in nanopure water and filter sterilize.

S30 extract preparation

1. Avestin EmulsiFlex B15 (volumes < 15 mL) or C3 (>15 mL) high-pressure homogenizer.
2. High-speed centrifuge and rotor capable of spinning at 30,000xg.
3. 1× S30 extract buffer: 10 mM TrisOAc, 14 mM $Mg(OAc)_2$, 60 mM KOAc, pH 8.2, filter sterile.
4. Sterile 50 mL falcon tube, 15 mL disposable conical tubes, and 1.5 mL microcentrifuge tube.
5. Aluminum foil.
6. Liquid nitrogen and dewar.

Producing ssOST in CFPS supplemented with POPC-nanodiscs

1. S30 extract
2. Stock solutions: All stock solution is dissolved in nanopure water and filter sterile. The aliquots of the stocks are flash-frozen and stored at -80°C. Keep working stocks at -20°C for several months.
    a. 15× salt solution (SS): 180 mM $Mg(Glu)_2$, 150 mM $NH_4(Glu)$, and 1.950 M $K(Glu)_2$.
    b. 15× master mix (MM) stock: 18 mM adenosine triphosphate (ATP), 12.75 mM guanosine triphosphate (GTP), 12.75 mM uridine triphosphate (UTP), 12.75

mM cytidine triphosphate (CTP), 0.51 mg/mL folinic acid, and 2.559 mg/mL *E. coli* tRNA (Roche).

c. 15× reagent mix (RM) stock: 50 mM amino acids mix, 1 M phosphoenolpyruvate (PEP, Roche), 100 mM nicotinamide adenine dinucleotide (NAD), 50 mM coenzyme-A (CoA), 1 M oxalic acid, 250 mM putrescine, 250 mM spermidine, and 1 M HEPES.

3. Purified 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) nanodiscs at 15 mg/mL stock concentration, prepared according to the standard protocol (Bayburt et al., 2002).
4. Nuclease-free water.
5. Sterile 1.5 mL microcentrifuge tube.

Protein purification and crude membrane extracts containing ssOST or LLOs

1. Buffer A Resuspend buffer: 50 mM HEPES, 250 mM NaCl, pH 7.5, filter sterile with 0.2 μm bottle-top filter.
2. Pierce™ Protease Inhibitor Tablets, EDTA-free.
3. RNase-Free DNase I from EpiCentre.
4. Avestin EmulsiFlex C5 homogenizer.
5. Buffer B: buffer A supplied with 10% (v/v) Glycerol and 1% (w/v) n-Dodecyl β-D-maltoside (DDM), at pH 7.5 and filter sterile.
6. Centrifuge and ultracentrifuge with rotor capable of spinning at 100,000xg.

7. Potter-Elvehjem tissue homogenizer.
8. Ni-NTA affinity resin.
9. Gravity column for affinity purification.
10. Nickel affinity purification buffers:

For preparing ssOST:	*For preparing scFv13-R4*:
Buffer C: 1 M Imidazole in buffer B.	Buffer P: 1 M Imidazole in buffer A.
Buffer D: 20 mM Imidazole in buffer B.	Buffer Q: 10 mM Imidazole in buffer A.
Buffer E: 60 mM Imidazole in buffer B.	Buffer R: 60 mM Imidazole in buffer A.
Buffer F: 250 mM Imidazole in buffer B.	Buffer S: 250 mM Imidazole in buffer A.

11. ÄKTA fast protein purification (FPLC) system with SuperDex-200 size exclusion chromatography column.
12. Size exclusion buffers using with ÄKTA system:

For preparing ssOST:	*For preparing scFv13-R4*:
Buffer G: 50 mM HEPES, 100 mM NaCl, 5% (v/v) glycerol, 0.01% (w/v) DDM, pH 7.5, filter sterile and degas for 5 min.	Buffer T: 50 mM HEPES, 100 mM NaCl, 1 mM EDTA, pH 7.5, filter sterile and degas for 5 min.

13. 3K MWCO protein concentrator column.
14. BioRad Bradford protein concentration assay.
15. BioRad RC DC™ (reducing agent and detergent compatible) protein concentration assay.

Extraction of LLOs

1. Lyophilizer.
2. 30 mL PTFE-conical tube.
3. Extracting solution: 10:20:3 (v/v/v) chloroform:methanol:water. Measure and mix all solvent in glassware since chloroform will dissolve plastic. Store the solution in a capped bottle at all time to prevent concentration change due to evaporation of chloroform and methanol.

Caution! Always wear gloves and use fume hood when handling the extracting solution since chloroform is toxic.

4. Clean metal spatula.
5. 15 mL clean glass vial.
6. Vacuum concentrator machine that withstands organic solvent.

IVG reaction

1. Sterile 1.5 mL microcentrifuge tube.
2. 10× IVG buffer stock: 100 mM HEPES, 100 mM $MnCl_2$, 1% (w/v) DDM, pH 7.5 in nanopure water and filter sterile.
3. Sterile nanopure water.
4. 30°C stationary water bath.

CFGpS reaction

1. Sterile 1.5 mL microcentrifuge tube.
2. Sterile nanopure water.
3. Similar buffer as CFPS set up in previous section
4. *N*-glycosylation activation solution: To prepare 50 mL of the solution, dissolve 0.247 g of manganese chloride tetrahydrate and 0.05 g n-Dodecyl β-D-maltoside in 40 mL MilliQ water. Mix well and adjust the final volume to 50 mL with MilliQ water. This buffer is stable at 4 °C for 2 months. For longer term storage, store at -20 ºC.
5. *O*-glycosylation activation solution: To prepare 50 mL of the solution, dissolve 1.71 g sucrose and 11.25 mg tetracycline in 40 mL MilliQ water. Mix well and adjust the final volume to 50 mL with MilliQ water. Store this buffer at -20 ºC.

Lectin blot and Western blot analysis of glycosylation products

1. Standard apparatus for SDS-PAGE and immunoblotting analysis.
2. Immobilon-P PVDF 0.45μm membrane.
3. Tris-buffered saline (TBS): dissolve 80.0 g of NaCl, 20.0 g of KCl, and 30.0 g of Tris base in 800 mL of nanopure water, bring volume to 1 L and autoclave.
4. Tris-buffered saline, 0.05 % Tween-20 (TBST): Add 100 mL of 10× TBS to 900 mL of nanopure water. Add 500 μL Tween-20.
5. Albumin from bovine serum (BSA): 5% (w/v) in TBST for blocking solution, 3% (w/v) in TBST for lectin blotting. When using lectin, BSA/TBST is a preferred blocking solution than milk/TBST to prevent interaction between lectin and milk oligosaccharides.
6. Non-fat dry milk Nestle®: 5% (w/v) in TBST for blocking solution for all other antibodies.
7. Anti-6xHis Tag® antibody peroxidase conjugate (His-HRP, from Abcam): 0.5 μg/mL in 5% milk/TBST.
8. Soybean agglutinin peroxidase conjugate (SBA-HRP): 0.5 μg/mL in 1% BSA/TBST (or other lectin or antibody specific for the glycan of choice).
9. Rabbit serum containing anti-*C.jejuni* heptasaccharide glycan antibody (HR6P-Rabbit) [Note 1]: 0.5 μg/mL in 5% milk/TBST.
10. Goat Anti-Rabbit IgG H&L peroxidase conjugate (Rabbit-HRP, from Abcam): 0.05 μg/mL in 5% milk/TBST.

11. BioRad Clarity™ Western ECL Substrate.
12. % Coomassie Brilliant Blue membrane stain solution: dissolve 0.1 g of Coomassie Blue R250 in 50 mL of methanol (MeOH), 7 mL of acetic acid, and 43 mL of nanopure water. Stain can be saved and reused multiple times.
13. Destain solution: 50 % MeOH in nanopure water. Discard destain following hazardous waste protocols.
14. BioRad ChemiDoc™ XRS+ System.

9.2 Methods

9.2.1 Preparation of ssOST by CFPS

S30 extract preparation

Days 0-2

1. Grow *E. coli* strain BL21 Star (DE3) in a shake flask or fermenter to OD_{600} ~3 in 2×YTPG media. Add 1 mM IPTG at OD_{600} 0.6-0.8 to induce expression of T7 RNA polymerase.
2. Harvest cells by centrifugation at 5000xg for 15 min at 4°C. Wash cells 3x in 25 mL of S30 buffer (vortex to resuspend) and pellet by centrifugation at 5000xg for 10 min at 4°C. After the last resuspension, pellet cells at 8000×g for 10 min at 4°C and flash-freeze on liquid nitrogen. Pellet can be stored at -80°C or used directly.

Day 3 preparing extract

Keep the extract on ice at all times unless noted otherwise. Work seamlessly. All equipment in contact with lysate should be pre-equilibrated to 4°C.

3. Remove cell pellet from -80°C and add 1 mL of S30 buffer per 1 g of wet cell mass. Dislodge pellet from the wall of the bottle. Vortex to resuspend to homogeneity.
4. Disrupt cells using Avestin EmulsiFlex-B15 (lysis volumes <15 mL) or C3 (>15 mL) high-pressure homogenizer at 20,000-25,000 psi. Pass the cells only once. Cell lysis is the key step in extract preparation and could be alternatively performed using sonication (Kwon and Jewett, 2015b).
5. Centrifuge lysate at 30,000×g for 30 min at 4°C to remove cell debris.
6. Immediately pipette the supernatant into new centrifuge tubes and centrifuge again at the same setting.
7. Immediately pipette supernatant into 1.5-mL microcentrifuge tubes.
8. Pre-incubation: wrap the microcentrifuge tubes in aluminum foil and incubate at 37°C in a shaker (~120-250 rpm) for 60 min.
9. Clarification: centrifuge at 15,000×g for 15 min at 4°C. Immediately pipette the supernatant into 15 mL disposable conical tubes and place on ice.
10. Immediately make 50 μL aliquot and 1-2 mL volume stocks of the cell extract. Flash freeze in liquid nitrogen and store at -80°C.
11. Perform a Bradford assay to measure total protein concentration (usually ~40 g/L). S30 extract performance is maintained for approximately 3 freeze-thaw cycles.

9.2.2 Producing ssOST in CFPS supplied with POPC-nanodiscs

1. Calculate the appropriate volumes of each reagent according to total number of reactions. Ensure that all components are at the concentrations listed.
2. Thaw all the reagents on ice. Set up microcentrifuge tubes on ice, one for each cell-free reaction and one for reaction premix.
3. CFPS reaction is performed with a modified, reducing PANOx-SP system (Jewett and Swartz, 2004d):

CFPS components	Volume per 1 reaction (μL)
15x SS	1
15x MM	1
15x RM	1
pJL1 plasmid encoding ssOST	200 ng
S30 extract	4
15 mg/mL POPC nanodiscs	1
Nuclease-freee water	Bring final volume to 15 μL

   a. To make a premix, combine the appropriate amounts of 15×SS, 15×MM, 15×RM, plasmid and nuclease-free water in reaction premix tube. Vortex and quickly spin down the tube [Note 4].

b. Then add the appropriate amounts of POPC nanodiscs and S30 extract.

c. Gently pipette the mixture up and down to thoroughly mix all components, but make sure to minimize bubble formation.

4. Aliquot 15 μL premix into individual microcentrifuge tube.
5. Briefly centrifuge CFPS-reaction tubes to ensure all liquid is held at the bottom of the tube.
6. Incubate reaction at 30°C for 6 h in a stationary water bath. The CFPS reaction containing ssOST in nanodisc can be loaded directly into IVG reaction for rapid glycosylation screening.

9.2.3 Preparation of purified and crude membrane extract glycosylation components

Preparation of crude membrane extract or purified ssOST enzyme

Days 0-1

1. Grow *E. coli* strain CLM24 carrying pSN18 plasmid in 50 mL LB supplied with ampicillin and 0.2% D-glucose overnight at 37°C.
2. Subculture 1:20 from the overnight culture into a fresh 1.0 L TB supplied with ampicillin. Grow with shaking at 220 rpm until OD_{600} reaches 0.4–0.5.
3. Adjust incubation temperature to 16°C and leave the culture with shaking for an hour.

4. Induce protein expression with 0.02% L-arabinose. Incubate at 16°C overnight.

Day 2 prepare membrane extract containing active *Cj*OST

5. The next day, harvest cell by centrifugation at 8,000×g for 10 min at 4°C. Wash cell pellet by resuspending with 200 mL buffer A and centrifuge again at the same setting. Discard supernatant. Collect pellet and determine wet cell mass. Pellet can be saved in -80°C fridge for a month or used directly.
6. Resuspend cell pellet using 10 mL buffer A per 1 g wet cell mass. Add EDTA-free protease inhibitor to prevent protein degradation. Add DNAse to reduce sample viscosity. Use standard manufacturer's protocol.
7. Pre-equilibrate Avestin homogenizer with ice-cold buffer A. Disrupt cells using Avestin C5 EmulsiFlex homogenizer at 17,000 psi for 3 passes.
8. Centrifuge lysate at 30,000×g for 30 min at 4°C to remove cell debris.
9. Collect and ultracentrifuge supernatant at 100,000×g for 2 h at 4°C to isolate membrane fraction.
10. Collect pellet containing membrane fraction and *Cj*OST. Resuspend pellet in 20 mL buffer B using Potter-Elvehjem tissue homogenizer. Make sure to fully resuspend the pellet [Note 5]. Transfer homogenized sample into sterile 50 mL conical tube. Add protease inhibitor cocktail into sample and incubate with shaking (120 rpm) at room temperature for an hour. The DDM detergent in buffer B will extract and solubilize *Cj*OST from bacterial membrane.

11. Ultracentrifuge sample at 100,000×g for an hour at 4°C. The supernatant now contains detergent-solubilized *Cj*OST enzyme.

Alternatively: To prepare crude membrane extract containing active *Cj*OST, centrifuge sample at 20,000×g for an hour at 4°C. Collect and immediately add protease inhibitor into supernatant after centrifugation. The crude membrane extract is active at 4°C for one week. We have demonstrated the use of this method to prepare several active ssOSTs that can modifty targeted protein acceptor. In addition, crude membrane extract containing active LLOs can be prepared in a similar method [Note 6].

12. Add 0.4 mL buffer C into supernatant to adjust imidazole concentration to 20 mM.
13. Equilibrate 0.5 mL Ni-NTA resin by washing with ice-cold buffer D at 5 times bed volume. Add pre-equilibrated Ni-NTA resin into supernatant, incubate with rolling overnight at 4°C.

Day 3 purification by affinity and size exclusion chromatography (SEC)

Keep sample on ice at all times unless noted otherwise. All equipment in contact with sample should be pre-equilibrated to 4°C.

14. Load sample into clean gravity column at the flowrate of 0.5 mL/min.
15. Wash resin with 5 bed volumes of buffer D, followed by 5 bed volumes of buffer E. Then elude protein with 7 bed volumes of buffer F. Keep all the fractions for analysis by Coomasie blue.

16. Pre-equilibrate SuperDex-200 SEC column connecting ÄKTA-FPLC system with ice-cold buffer G. Load eluent fraction into sample loop. Inject sample through SEC column. Collect and combine fractions with size corresponding to *Cj*OST (84 kDa) together.
17. Concentrate protein to 1-2 mg/mL final concentration using 3K MWCO protein concentrator column. Add glycerol to the sample at 20% (v/v) concentration. Aliquot and store *Cj*OST at -80°C for 4-5 months.
18. Determine protein concentration and sample purity with RC/DC assay and Coomassie blue protein stain, respectively.

Purification of acceptor protein scFv13-R4(N34L, N77L)$^{DQNAT\text{-}6xHis}$

1-8 These steps are essentially the same as protocol described in section 3.2.1 above, with a few exceptions. *E. coli* strain BL21 Star (DE3) carrying pET28a(+)-scFv13-R4(N34L, N77L)$^{DQNAT\text{-}6xHis}$ plasmid is used. The inducer for pET-based vector is IPTG at 0.1 mM final concentration. Kanamycin antibiotic is used at 100 ug/mL.

9. Adjust the imidazole concentration in the supernatant to 10 mM Imidazole with buffer P.
10. Equilibrate 0.25 mL Ni-NTA resin by washing with ice-cold buffer Q at 5 times bed volume. Add pre-equilibrated Ni-NTA resin into supernatant, incubate with rolling at room temperature for an hour.

Keep sample on ice at all times unless noted otherwise. All equipment in contact with sample should be pre-equilibrated to 4°C.

11. Load sample into clean gravity column at the flowrate of 0.5 mL/min.
12. Wash resin with 5 bed volumes of buffer Q, followed by 5 bed volumes of buffer R. Then elude protein with 7 bed volumes of buffer S. Keep all the fractions for analysis by Coomasie blue.
13. Pre-equilibrate SuperDex-200-SEC column connecting ÄKTA-FPLC system with ice-cold buffer T. Load eluent fraction into sample loop. Inject sample through SEC column. Collect and combine fractions with size corresponding to scFv13-R4 (29 kDa) together.
14. Concentrate protein to 1-2 mg/mL final concentration using 3K MWCO protein concentrator column. Add glycerol to the sample at 10% (v/v) concentration. Aliquot and store scFv13-R4 at -80°C for 6 months.
15. Determine protein concentration and sample purity with Bradford assay and Coomassie blue protein stain, respectively.

Extraction of LLOs bearing C. jejuni glycan (adapted from (Guarino and DeLisa, 2012b))

Days 0-2

1-5 These steps are essentially the same as protocol described in section 3.2.1, with a few exceptions. *E. coli* strain CLM24 carrying pMW07-pglΔB plasmid is used. The inducer for this plasmid is L-arabinose at 0.2% final concentration. Chloramphenicol antibiotic is used at 20 ug/mL.

6. Use clean spatula to scrap cell pellet and transfer to clean 50 mL conical tubes. Freeze-dry cell pellets to complete dryness at -70°C with lyophilizer (usually takes ~2 days).

Day 4

7. Weigh and combine lyophilisate into a sterile 30 mL PTFE-conical tube. Use clean spatula to break dried pellet into small fractures.
8. Add 20 mL 10:20:3 (v/v/v) chloroform:methanol:water extracting solution into the tube and incubate with shaking for 30 min at room temperature.
9. Centrifuge the mixture at 4000×g for 15 min at 4°C.
10. Transfer organic fraction (bottom layer) to a clean 15 mL glass vial. Remove chloroform and methanol with vacuum concentrator at room temperature (usually take ~4-5 h).
11. Place the vial into freeze-dry unit to remove residue water at -70°C overnight.

Day 5

12. Lyophilisate now contains active lipid-linked oligosaccharide (*Cj*LLOs). Weigh lyophilisate mass. Dried LLOs can be stored at -80°C for 6 months.
13. Resuspend lyophilisate at 1.0 mL 1× IVG buffer per 1.0 mg lyophilisate dried weight. The resuspension should look yellowish. Transfer the mixture to a sterile microcentrifuge tube, spin down briefly, aliquot and store soluble fraction containing active *Cj*LLOs in -20°C for up to 2 months.

9.2.4 IVG setup

Day 1

1. In a sterile 1.5-mL microcentrifuge tube, add following reagents:
   a. 3 μg purified antibody fragment scFv13-R4. Alternatively, N-terminal TAMRA-labelled peptide at 8.5 μM can be used as an acceptor substrate [Note 7].
   b. 2 μg purified *Cj*OST or 25 μL crude membrane extract containing active *Cj*OST or 25 μL CFPS-nanodisc reaction containing active *Cj*OST.
   c. 5 μg extracted *Cj*LLOs.
   d. 5 μL 10× IVG buffer.
   e. Bring final volume to 50 μL with sterile nanopure water.
2. In addition, it is necessary to set up control reactions to prevent fault-positive result. A typical reaction set is as follow:

IVG components	Sample	Control		
Protein/peptide acceptor	+	+	+	-
*Cj*OST	+	+	-	+
*Cj*LLOs	+	-	+	+

3. Incubate the reaction tube in stationary water bath at 30°C for 16 h.

Day 2

4. Centrifuge reaction tube at 10,000×g for 15 min at 4°C.

5. Collect soluble fraction. Reaction is stopped by adding Laemmli sample buffer. Keep sample at -20°C for analysis by SDS-PAGE followed by immunoblotting.

9.2.5 Cell-free reaction assembly and lyophilization

1. Assemble reactions for *in vitro* glycoprotein expression. For expression of aglycosylated proteins (e.g., conjugate vaccine carrier proteins), lysate from CLM24 Δ*lpxM* lacking either the *Cj*PglB oligosaccharyltransferase and/or a glycan biosynthetic pathway should be used [cell-free protein synthesis (CFPS) reactions]. Conversely, for production of glycoproteins, lysate from CLM24 Δ*lpxM* cells expressing both oligosaccharyltransferase and/or an O antigen biosynthetic pathway should be used. Reactions can be run at 15.0 μL scale in 1.5 mL microcentrifuge tubes, 1.0 mL in 15.0 mL conical tubes, or 5.0 mL scale in 50 mL conical tubes. (Note 9)
2. Reactions are run using a modified PANOx-SP system. All reagents and assembled reactions should be kept on ice at all times. Recipe for assembly of a 15 μL reaction is given below:

    a. Optional: Reagent mix, master mix, and salt solution can be pre-assembled in large quantities to speed reaction assembly and reduce error from pipetting small volumes.

Reagents	[Stock]		Volume (uL)- 1X	[Reagents]$_f$	
15X SS	15	*x*	1.00	1.00	*x*
$Mg(Glu)_2$	180	*mM*		*12*	*mM*
$NH_4(Glu)$	150	*mM*		*10*	*mM*
K(Glu)	1950	*mM*		*130*	*mM*
Master mix	15	*x*	1.00	1.00	*x*
ATP	18	*mM*		*1.2*	*mM*
GTP	12.75	*mM*		*0.850*	*mM*
UTP	12.75	*mM*		*0.850*	*mM*
CTP	12.75	*mM*		*0.850*	*mM*
Folinic acid	0.51	*mg/mL*		*0.034*	*mg/mL*
tRNA	2.559	*mg/mL*		*0.171*	*mg/mL*
Reagent mix			2.30		
20 amino acids	50	*mM*	0.60	2.00	*mM*
PEP	1000	*mM*	0.50	33.33	*mM*

NAD	100	*mM*	0.06	0.33	*mM*
CoA	50	*mM*	0.08	0.27	*mM*
Oxalic acid	1000	*mM*	0.06	4.00	*mM*
Putrescine	250	*mM*	0.06	1.00	*mM*
Spermidine	250	*mM*	0.09	1.50	*mM*
HEPES	1000	*mM*	0.86	57.00	*mM*
T7 RNA polymerase	5	*mg/mL*	0.30	0.10	*mg/mL*
Lysate			4.00		
Nuclease-free water			4.40		
		V_T=	**15.00**		

Note on optimizing cell-free reaction conditions for specific proteins.

a) For expression of proteins containing disulfide bonds[17], reactions can be carried out under oxidizing conditions, as previously reported[28]. To achieve oxidizing

[17] Protein databases such as UniProt and PDB can be used to determine whether a target protein contains disulfide bonds in its native state. If this information is not available a priori, prediction tools such as Disulfind (https://bio.tools/disulfind) and DiANNA (http://clavius.bc.edu/~clotelab/DiANNA/) can be used.

conditions, pre-condition lysate with 750 μM iodoacetamide at room temperature for 30 min to covalently bind free sulfhydryls (-SH), including the active site cysteines of the thioredoxin reductase (trxB) and glutathione reductase (gor) enzymes that represent the primary disulfide bond reducing enzymes in the *E. coli* cytoplasm. The CFPS reaction mix was then supplemented with 200 mM glutathione at a 4:1 ratio of oxidized and reduced forms and 10 μM recombinant *E. coli* DsbC.

b) For expression of membrane proteins, hydrophobic transmembrane domains can be stabilized by supplementing the cell-free reaction with nanodiscs or other membrane mimics. For example, we typically express PorA protein using cell-free lysate that is supplemented with nanodiscs at 1 μg mL^{-1} (Cube Biotech). Importantly, an appropriate size of the nanodisc is critical for a successful cell-free expression of the membrane proteins and, therefore, should be predetermined.

9.2.6 Lyophilization of cell-free reactions

1. Once assembled, flash-freeze reactions on liquid nitrogen. Poke holes in the lid of the tube or cover the tube with foil with holes in it to allow for proper lyophilization. Lyophilize reactions at 100 mTorr and −80 °C overnight or until fully freeze-dried.

9.2.7 Cell-free protein synthesis

1. Rehydrate lyophilized reactions with plasmid encoding the protein of interest at 6.67 μg mL^{-1} in the original reaction volume of nuclease-free water. (Note 10)
2. Incubate at 25-37 ºC for 1-20 hours. For best results, incubate reactions in a pre-warmed heat block with water in the wells to ensure consistent incubation temperature and optimal heat transfer.
3. Centrifuge reactions at 20,000 xg for 10 min and transfer the supernatant to a clean microcentrifuge tube.

9.2.8 Glycoprotein and conjugate vaccine synthesis

1. Rehydrate lyophilized iVAX reactions with plasmid encoding the protein of interest at 6.67 μg mL^{-1} in the original reaction volume of nuclease-free water.
2. Incubate at 25-37 ºC for 10-15 min. (Note 11)
3. For *N*-glycoprotein and conjugate vaccine biosynthesis, add 1 uL *N*-glycosylation activation solution per 15 uL of cell-free reaction. Mix well and return samples to an incubator for a total reaction time of 1 hour. Similarly, add 1 uL of O-glycosylation activation solution per 15 uL of cell-free reaction to initiate O-glycoprotein biosynthesis. Mix well and return samples to an incubator for a total reaction time of 16 hour. (Note 12)
4. Centrifuge reactions at 20,000 xg for 10 min and transfer the supernatant to a clean microcentrifuge tube.

9.2.9 Western blot analysis

1. Run reaction soluble fraction on 4-12% Bis-Tris SDS-PAGE gels (Invitrogen).
2. Transfer onto Immobilon-P polyvinylidene difluoride (PVDF) membranes (0.45 μm) using a semi-dry transfer cell (Bio-Rad) or similar.
3. Wash membranes with PBS and incubate in Odyssey® Blocking Buffer (LI-COR) at room temperature for 1 hour or overnight at 4 ºC.
4. Wash membranes 6 times with PBST for 5 min.
5. Probe membranes with both an anti-6xHis tag antibody and an antibody specific to the glycan/O antigen of interest, if available for at least 1 hour at room temperature or overnight at 4 ºC.
6. Wash membranes 6 times with PBST for 5 min.
7. Probe with appropriate fluorescently labeled secondary antibodies for at least 1 hour at room temperature.
8. Wash membranes 6 times with PBST for 5 min.
9. Image using an Odyssey® Fc imaging system (LI-COR), or similar. CRM197 and TT can also be detected antibodies recognizing diphtheria or tetanus toxin, respectively.

9.3 Notes.

1. Rabbit serum containing *C.jejuni* heptasaccharide glycan-specific antibody (hR6P) is made in-house and generously provided from Prof. Markus Aebi at ETH-Zürich, Switzerland.
2. Usually at least two transformations were done for each set to allow for plating of different cell densities or to ensure at least one plate with even spreading (the best one or both can be chosen to proceed with the assay).
3. This rinsing step was found to be important for cleaner blots so do not skip. For all blotting steps, it is important that the shaking evenly covers the membrane with the buffers. Insufficient shaking will result in uneven signal that will make it difficult to pick positive hits.
4. If different targeted ssOSTs will be produced in CFPS-nanodisc reaction, omit plasmid in premix and add each plasmid into individual reaction later.
5. Solubilization is a critical step in extracting active ssOST from the *E. coli* membrane. It is important to completely homogenize the sample and allow sufficient incubation time with DDM detergent to maximize extracting efficiency.
6. Similarly, crude membrane extract LLOs can be prepared the same way. Prepare 1 L TB culture with *E. coli* strain CLM24 carrying pMW07-pglΔB plasmid. After protein expression and cell harvesting, disrupt cell with Emulsiflex C5 homogenizer. Ultracentrifuge supernatant to isolate membrane fraction. Following solubilization membrane fraction with DDM detergent, centrifuge resuspend at

20,000×g for an hour at 4°C. Collect supernatant containing active LLOs. The crude membrane extract is active at 4°C for one week.

7. We use commercial N-terminal-TAMRA-GDQNATAF peptide substrate in our assay. In-house synthesized peptide with similar sequence can also be used as a glycosylation acceptor molecule.
8. The glycosylated protein will migrate slower in the SDS-PAGE gel due to the additional mass of the attached glycan, and on the anti-His immunoblot it will appear as a band slightly higher than the unmodified protein (if glycosylation efficiency is less than 100%, two bands will be apparent). The glycosylated form can be confirmed by appearance of a corresponding band on the glycan blot.
9. Reactions can also be run at the 15.0 μL scale in PCR tubes to increase throughput. However, as this changes the surface area to volume ratio of the reaction, initial rates and total yields can be lower. Yields should be assessed for each protein of interest.
10. Plasmid DNA purity can significantly affect *in vitro* protein synthesis yields. Midi- or maxi-prepped DNA is preferred. Elute DNA following purification in nuclease-free water. We recommend measuring and documenting 260 nm/280 nm and 260 nm/230 nm absorbance ratios to assess DNA purity.
11. Optimization: The length of this initial incubation period can impact glycosylation efficiencies by determining how much carrier protein is expressed before glycosylation. The glycosylation efficiency differs for S12 and S30 lysates due to the

concentration of membrane vesicles containing glycosylation machinery present in the lysate. Initial incubation times for S12 extracts can be extended out to 60 minutes.

12. Optimization: Incubation time can be extended up to 16 hours with a minimum of 45 minutes following addition of N-glycosylation activation solution.