DEVELOPING *ESCHERICHIA COLI* AS A PLATFORM TO STUDY PROTEIN GLYCOSYLATION AND SECRETION

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

Aravind Natarajan

August 2019

© 2019 Aravind Natarajan

ALL RIGHTS RESERVED

DEVELOPING ESCHERICHIA COLI AS A PLATFORM TO STUDY PROTEIN GLYCOSYLATION AND SECRETION

Aravind Natarajan, Ph. D.

Cornell University 2019

Escherichia coli is an exceptional model organism, tolerant to additions and overhauls of metabolic pathways, and recombinant expression of diverse proteins making it a tractable chassis for both the study of complex biological systems and production of therapeutics. However, lab strains of *E. coli* are unable to perform critical posttranslational modifications, like glycosylation, that are important for the study and production of protein therapeutics. In this work, we 1) endow *E. coli* with the catalytic potential to produce a diverse panel of human glyco-epitopes, including immunologically relevant MUC1 VNTR motifs, bearing mucin-type *O*-glycans through a novel recombinant pathway constituted by an *N*-acetyl galactosamine transferase, PglC from *A. baumannii*, a galactosyl transferase, WbwC from *E. coli* O104, and *O*-oligosaccharyl transferases from *N. gonorrhoeae* and *N. meningitidis.,* 2) developed a facile platform for the production of *N*-glycoproteins through the chromosomal engineering of *E. coli* at the Enterobacterial common antigen (ECA) and O-polysaccharide synthesis (O-PS) loci, creating a platform for the expression of three unique glycans – modified *C. jejuni* heptasaccharide, mammalian tri-mannosyl core, and *E. coli* O56 O-antigen, and four different glycoproteins, scFv13-R4, hGH, RNAse A and glucagon, bearing the modified *C. jejuni* glycan. Therefore, this work facilitates the study of glycosylation pathways and glycoconjugate products using *E. coli*.

Similarly, we created yet another platform for the high-throughput study of bacterial protein secretion, a system of relevance to pathogenesis and applications in bioprocessing. Through a novel survival-selection scheme, we screened 3,895 single

gene knockouts mutants of *E. coli* to identify deletions of *fliH, gfcC, yaiW and ydfI* as leading to hypersecretory phenotype of YebF and diverse fusion proteins.

While research to augment our understanding of biological processes towards improving quality of life is important, the inherent inequality of our society limits access to the fruits of such discoveries. We believe equity in higher education can empower communities through socio-economic mobility, and therefore explore the use of podcasts, an emerging medium, to amplify the voice of role models to support students from traditionally under-represented identities persist in STEM. This work lays the foundation for follow up studies, harnessing podcasts to foster inclusion and equity.

BIOGRAPHICAL SKETCH

Aravind Natarajan was born to Dr. Gurumurti Natarajan and Mrs. Anuradha Natarajan in the coastal city of Chennai (erstwhile Madras) in Tamil Nadu, India. His first foray into research was through a summer workshop on 'Bioresources and Biotechnology' at the M.S. Swaminathan Research Foundation in the summer of 2004.

He earned a Bachelor of Science in Biochemistry with minors in Chemistry and Zoology from the University of Madras in Tamil Nadu, India. Here he was recognized for ranking second in his program and for graduating First Class with Distinction (equivalent of a *summa cum laude*). During his undergraduate studies, Aravind was also a recipient of the prestigious Summer Research Fellowship, jointly awarded by the Indian Academy of Sciences, Indian National Science Academy and the National Academy of Sciences (akin to the National Research Council in the US). This enabled him to work with Professor L.S. Shashidhara at the Indian Institute for Science Education and Research in Pune, India on 'Expressing and purifying Noggin from Hydra in *Escherichia coli*'.

Aravind then went on to earn a Master of Science with a major in Genomics and minors in Cell Biology, Chemistry, Chronobiology, Neurogenetics, and Structural Biology from Center for Excellence in Genomic Sciences at School of Biological Sciences in the Madurai Kamaraj University. Here he was awarded a Gold medal for ranking first in the program, and graduated First Class with Distinction (equivalent of a *summa cum laude*). Over the summer of 2010, he worked with Professor Samit Adhya at the Indian Institute for Chemical Biology, on the 'Trimeric RNA import complex: Expression, purification, reconstitution and transfection studies on the Hep G2 cell line'. Aravind's Masters' thesis project titled 'A race between codon misreading and

information suppression: Elucidation of novel properties associated with *ppm*, an allele of *rspD*, and *supE44*, a classic Amber suppressor in *E. coli*" was guided by Professor Hussain Munavar, who was a great source of inspiration to him. While in graduate school, Aravind also led a team of peers to the second place in the Biotechnology Entrepreneurship Students Team competition (akin to the iGEM competition in the US), hosted by the Department of Biotechnology (Government of India) and the Association of Biotech Led Enterprises for proposing a 'Nanocomposite biokit for the effective detection of Tetanus Toxin'.

In 2012, Aravind embarked on his Doctoral studies in Microbiology with a minor in Molecular Biology and Biochemical Engineering at Cornell University. Here he worked with Professor Matthew DeLisa, studying protein secretion and glycosylation. He was awarded the Cornell Fellowship for meritorious graduate students in 2012, and the Procter & Gamble Award in 2018 for his academic accomplishments. He also received the Cornelia Ye Outstanding Teaching Assistant Award in 2018, awarded to only two teaching assistants across Cornell University every year. His proactive engagement and advocacy for equity in higher education is highlighted through being inducted into the prestigious Edward A. Bouchet Honor Society, and being awarded the Excellence in Leadership award by the Office of Inclusion and Student Engagement and the Outstanding Graduate or Professional Student award by the Cornell Asian Pacific Islander Student Union. Aravind is also the recipient of an Engaged Graduate Student grant to pursue community engaged research, and of a Robert Frederick Smith Initiative award to found a science podcast titled, Science Blender. Finally, Aravind was nominated as the Degree Marshal for the Graduate School during the Commencement event, and as a speaker in the International Student Graduation ceremony in 2019.

Dedicated to Indhira.

Thank you for always being there,

and filling our lives with so much love and joy.

ACKNOWLEDGMENTS

I am indebted to my parents, Mrs. Anuradha Natarajan and Dr. Gurumurti Natarajan, who have made many sacrifices to provide me the support and education that has made me who I am today. I am also grateful to my brother, Mr. Abenav Natarajan for being a pillar of support and a source of joy. Mrs. Divya Tantry, my sister-in-law, has been a blessing to the family, bringing joy abound. I am grateful to all of you for your unwavering belief in me.

Professor Matthew DeLisa has been a valuable mentor and role model, shaping me to be a better scientist in ever so many ways - from experimental design to the final presentation of my work. I sincerely appreciate your guidance. Professor Linda Nicholson has been an inspiration and a calming presence in the course of my doctoral work. Thank you for being thoughtful and guiding me through many key steps of my doctoral endeavor, particularly during my postdoctoral search. It has been very rewarding to have Professor Matthew Paszek on my committee. Your expertise in mammalian *O*-glycosylation has been a great asset, and our conversations have inspired my science. Further, your generosity in sharing experiences leading to a faculty position and thoughts on inclusion in higher education have been insightful.

I am fortunate to have worked alongside many phenomenal members of the DeLisa Research Group who have been pivotal to my success in grad school. I will cherish our many conversations, CTB outings, and discussions at group meetings. I am particularly grateful to Professor Jason Boock, who was a key mentor, supporting my transition to the group. He continues to be generous with his time, sharing wisdom from his experiences as guidance to my career. Thanks to the many graduate and

undergraduate students that have worked with me, providing the opportunity to grow with them, and supporting my research.

I have had the privilege of having found a home in two fields at Cornell – in the Department of Microbiology (Micro), and the Robert Frederick Smith School of Chemical and Biomolecular Engineering (CBE). In Micro, I owe thanks to many faculty members who have taken a keen interest in supporting my career, specifically Professor Anthony Hay and Professor Stephen Winans who both took time to sit down and discuss different aspects of my career, and provided me valuable guidance. I am indebted to Dr. Chantal Koechli, Dr. Olga Lastovetsky, and Dr. Erin Eggleston who were senior to me in the program and helped my transition to the American education system. I sincerely appreciate Dr. Koechli, Dr. Lastovesky and their partners Mr. Dave Flanelly and Mr. Marc Barker, for looking out for me and creating a welcoming space here in Ithaca. Ms. Lillian Henry, Ms. Patti Butler, Ms. Shirley Cramer, and Ms. Tina Daddona have been an invaluable support. At CBE, I am thankful to many faculty members for welcoming me to the department and for ever so many pleasant and insightful conversations. Specifically, I am grateful to Professor Susan Daniel who has been an ally, advocating for me and supporting my career aspirations in incredible ways. I am fortunate for the opportunity to work closely with her on the Diversity and Inclusion Program and the ARISE study. I am appreciative of ChEGSA and several batches of CBE graduate students who have supported my time at Cornell. I will cherish our joys playing intramural sports, our softball championships and the many memories in Olin Hall. I am also thankful to many staff at CBE who have supported my journey, including Ms. Johanna Tuttle, Ms. Karen Prosser, Ms. Lucinda Allen, Mr. Matthew Pitcher, and Mr. Paul Pelletier.

I am indebted to Director Anitra McCarthy, Dean Barbara Knuth, Dean Jan Allen, Dean Janna Lamay, Dean Jason Kahabka, Dean Sara Hernandez, and Dean Vijay Pendakur for the opportunity to work with them on so many initiatives and to learn from the process. I am also thankful to members of the GPSA for many memorable conversations and learning moments.

Dr. Anne Poduska, Ms. Linda Tompkins, Dr. Lisa Schneider-Bentley and peers from the Engineering Learning Initiatives office have been invaluable to my growth as an instructor and communicator. Thank you for being incredibly thoughtful.

I am thankful for the opportunity to work with Dr. Kristen Elmore who has been invaluable to the ARISE project and wonderfully kind and cheerful person to work with.

I am grateful to my friends from so many walks of life at Cornell and in Ithaca for having enriched my life and supported my journey. Specifically, Ms. Lauren Beckles has been a godsend, looking out for me, and nourishing me with care and good health. Thanks to Mr. Vinay Varghese and Mrs. Anindita Bagchi for always being there for me, bringing me happiness and cheering for me.

I am fortunate to have been funded through many engagements over the years, that have supported my academic endeavors and professional growth. Specifically, I am grateful for the Smith Initiative that enabled me to create a podcast, Science Blender, for the Engaged Graduate Student grant towards the ARISE project, to the Graduate School for the Cornell Fellowship of meritorious students and a travel grant that enabled my participation in a workshop at the Complex Carbohydrate Research Center, to the Engineering Learning Initiatives for the TA Fellow position, and finally to Micro and

CBE for TA appointments that have all supplemented the RAship provided by Professor DeLisa over the course of my doctoral program.

Ms. Janani Hariharan has been a blessing, a ray of sunshine and a force that has makes me a better person with every moment we spend together. Thank you for tolerating my crazy work-life balance, and for bringing me incredible joy.

I want to acknowledge that this body of work was carried out in lands stolen from Indigenous people, specifically from the Cayuga Nation. I pledge to use the privilege of my education to create welcoming spaces, to empower Indigenous people and more broadly support people from all identities.

TABLE OF CONTENTS

ABSTRACT i

BIOGRAPHICAL SKETCH iii

ACKNOWLEDGMENTS vi

TABLE OF CONTENTS x

LIST OF FIGURES xiv

LIST OF TABLES xvii

CHAPTER 1: Metabolic engineering of glycoprotein biosynthesis in bacteria 1

1.1 Abstract 1

1.2 Introduction 1

1.3 Design and assembly of orthogonal glycosylation pathways 5

1.4 Optimization of glycosylation pathway performance 11

1.5 Cell-free metabolic engineering of glycoprotein biosynthesis 17

1.6 Concluding remarks 21

1.7 Summary 22

1.8 Abbreviations 23

1.9 Funding 23

1.10 Competing interests 24

1.11 References 24

CHAPTER 2: Glyco-recoded *Escherichia coli*: Recombineering-based genome editing of native polysaccharide biosynthesis gene clusters 36

2.1 Abstract 36

2.2 Introduction 37

2.3 Results 40

2.3.1. Recombinase-mediated genomic replacement of native oligosaccharide gene clusters 40

2.3.2. Protein glycosylation from single-copy expression of PglB 44

2.3.3. Heterologous glycans expressed from the genome and transferred to proteins 47

2.3.4. Double integrant requires only acceptor protein plasmid for glycosylation 50

2.4 Discussion 56

2.5 Materials and methods 60

2.5.1. Strains and plasmids 60

2.5.2. Chromosomal integration of DNA 63

2.5.3. Cell growth 63
2.5.4. Protein expression, periplasmic extraction and purification 64
2.5.5. Immunoblotting and dot blotting analysis 64
2.5.6. Flow cytometric analysis 65
2.6 Author contributions 66
2.7 Conflicts of interest 66
2.8 Acknowledgements 66
2.9 References 66

CHAPTER 3: Engineering an orthogonal *O*-linked protein glycosylation pathway in bacteria for producing tumor-associated glycoconjugates 72
3.1 Abstract 72
3.2 Introduction 72
3.3 Results 74
3.3.1. Engineered assembly of T antigen on lipid carrier 74
3.3.2. Site-specific installation of T antigen on acceptor proteins 76
3.3.3. O-OSTs glycosylate an expanded set of acceptor proteins with T antigens 80
3.3.4. Expression of immunologically relevant MUC1 glycotopes 83
3.3.5. In vitro O-glycosylation with the T antigen 85
3.4 Discussion 87
3.5 Materials and Methods 89
3.5.1. Bacterial strains and growth conditions 89
3.5.2. Plasmid construction 89
3.5.3. Protein analysis 94
3.5.4. In vitro glycoprotein synthesis 95
3.5.5. Flow cytometric analysis 96
3.6 References 97

CHAPTER 4: An engineered survival-selection assay for extracellular protein expression uncovers hypersecretory phenotypes in *Escherichia coli* 104
4.1 Abstract 104
4.2 Introduction 105
4.3 Results 108
4.3.1. An Engineered Selection for Extracellular Protein Expression 108
4.3.2. Selection of Gain-of-Function Mutations That Increase Host Secretion Capacity. 112

4.3.3. Mutant Strains Promote Hypersecretion of YebF Passenger Protein 119
4.3.4. Mutant Strains Promote Hypersecretion of YebF Bearing N-Linked Glycans 119
4.4 Discussion 122
4.5 Materials and Methods 126
4.5.1. Bacterial Strains and Growth Conditions 126
4.5.2. Plasmid Construction 127
4.5.3. Selective Plating of Bacteria 127
4.5.4. Genetic Selection of Keio Knockout Library 128
4.5.5. Protein Analysis 129
4.5.6. Characterization of Outer Membrane Permeability 130
4.6 Author Contributions 130
4.7 Acknowledgments 130
4.8 References 130

CHAPTER 5: ARISE: Amplifying Role models as Inspiration for STEM Education 136
5.1 Abstract 136
5.2 Introduction 137
5.3 Methods 138
5.4 Measures 140
5.4.1. Anticipated belonging in science 140
5.4.2. Self-efficacy in science 140
5.4.3. Perceptions of one's effort expenditure relative to other students in science class 141
5.4.4. Intentions to pursue a career in science 141
5.4.5. Intentions to pursue additional higher education 141
5.4.6. Identification with the podcast scientist 142
5.4.7. Science learning 142
5.4.8. Stereotypes of scientists 142
5.4.9. Podcast feedback 143
5.5 Results 144
5.5.1. Listeners of shared demographic identity identify with the scientist 144
5.5.2. Sense of belonging and self-efficacy are unchanged in either case 146
5.5.3. SO podcast catalyzes increased scientific learning 148
5.5.4. Podcast feedback highlights appetite for science content 149
5.5.5. Correlation of posttest variables marks difference in effect of the SO vs PN case 153
5.6 Discussion 154

5.7 Acknowledgments 156
5.8 Funding 156
5.9 References 157

CHAPTER 6: Future perspectives 162
6.1 Chromosomal engineering of glycosylation pathways in *E. coli* 162
6.1.1. Synthesize therapeutically relevant, activated monosaccharides for glycosylation reactions 163
6.1.2. Take a systems-level approach to streamline the metabolic flux of host strains for glycoprotein synthesis 164
6.2 Expanding O-glycoprotein expression in *E. coli* 165
6.2.1. Comprehensively characterizing the acceptor sequence specificity of O-OSTs 166
6.2.2. Expressing sialylated variants of the T and Tn antigens 166
6.3 Understanding YebF 167
6.4 Using a data driven approach to create podcasts to foster equity in higher education 168
6.5 References 169

APPENDIX A: Complete set of peptide motifs and proteins tested for *O*-glycosylation 173
APPENDIX B: Characterizing soluble expression of glycosyl-transferases 175
APPENDIX C: Estimated marginal means of post-test measures based on gender and race 186

LIST OF FIGURES

Figure 1-1: Protein glycosylation pathways reconstituted in *E. coli*..... 3

Figure 1-2: Cell-free metabolic engineering (CFME) of glycoprotein biosynthesis. . 17

Figure 2-1: Glyco-recoding the genome of *E. coli* K-12..... 41

Figure 2-2: Characterization of gene expression from O-PS and EPA sites. 43

Figure 2-3: Characterization of genomically-integrated glycan biosynthesis pathways. 45

Figure 2-4: Site-specific genomic integration of functional PglB. 47

Figure 2-5: Site-specific genomic integration of orthogonal glycan biosynthesis pathway..... 49

Figure 2-6: Dual-locus integration of complete protein glycosylation machinery. 51

Figure 2-7: Expression of different glycoprotein targets requires a single plasmid. .. 53

Figure 2-8: Growth analysis of strains expressing MBP-GCGDQNAT acceptor protein. 54

Figure 3-1: Schematic illustration of proposed core 1 *O*-glycosylation pathway..... 75

Figure 3-2: FACS based assay identifying *Ec*WbwC as the most active GalT. 77

Figure 3-3: Expression of *O*-glycosylated MBP^{MOOR} protein bearing the T and Tn antigens. 79

Figure 3-4: Expanded panel of protein carriers bearing the T antigen..... 80

Figure 3-5: Glycosylation of motifs derived from native glycoproteins..... 82

Figure 3-6: Synthesis of authentic human MUC1 glycotopes. 84

Figure 3-7: *In vitro O*-glycosylation with T antigen. 86

Figure 4-1: Graphical abstract illustrating the survival-selection scheme. 104

Figure 4-2: An engineered survival-selection assay for extracellular expression of proteins. 109

Figure 4-3: Linking extracellular secretion of YebF with resistance to β- lactam antibiotics. 111

Figure 4-4: Mock library selection for secretion of YebF to the culture medium. ... 113

Figure 4-5: Analysis of previously isolated mutations that enhance extracellular secretion. 115

Figure 4-6: Isolation of mutations that enhance extracellular secretion of YebF. 117

Figure 4-7: Characterization of isolated mutant strains. 120

Figure 4-8: Hypersecretion of structurally diverse substrate proteins. 121

Figure B-1: Testing GALNTs and GalTs using protein extraction and western blot. 175

Figure B-2: Analyzing human and bacterial sialyl transferases across five different host strains using protein extraction and western blot. 176

Figure B-3: Studying a panel of multifunctional STs, and STs that catalyze the attachment of sialic acid through an α(2,8) linkage using protein extraction and western blot. 177

Figure B-4: Characterizing a panel of STs that catalyze the attachment of sialic acid through an α(2,3) linkage using protein extraction and western blot. 178

Figure B-5: Testing a panel of STs that catalyze the attachment of sialic acid through an α(2,6) linkage using protein extraction and western blot. 179

Figure B-6: Representative spot titer images that capture the degree of soluble expression of GALNTs and GalTs through resistance to Carbenicillin. 180

Figure B-7: Survival curves for serially diluted E. coli MC4100 cells expressing GalNacTs 180

Figure B-8: Survival curves for serially diluted *E. coli* MC4100 cells expressing GalTs 181

Figure B-9: Representative spot titer images that capture the degree of soluble expression of STs listed below (batch 1) through resistance to Carbenicillin. 181

Figure B-10: Representative spot titer images that capture the degree of soluble expression of STs listed below (batch 2) through resistance to Carbenicillin. 182

Figure B-11: Representative spot titer images that capture the degree of soluble expression of STs listed below (batch 2) through resistance to Carbenicillin. 183

Figure B-12: Survival curves for serially diluted *E. coli* MC4100 cells expressing STs (set 1) 183

Figure B-13: Survival curves for serially diluted *E. coli* MC4100 cells expressing STs (set 2) 184

Figure B-14: Survival curves for serially diluted *E. coli* MC4100 cells expressing STs (set 3) 184

Figure B-15: Survival curves for serially diluted *E. coli* MC4100 cells expressing STs (set 4) 185

LIST OF TABLES

Table 2-1: Comparative cell growth and protein yields for different bacterial strains. 55

Table 2-2: Bacterial strains and plasmids used in this study 61

Table 3-1: Plasmids used in the current study. 91

Table 5-1: Summary of participant distribution across community colleges partners and experimental case. 139

Table 5-2: Summary of measures evaluated through this study. 143

Table 5-3: Analysis of sense of identification with podcast by race/ ethnicity. 145

Table 5-4: Excerpts from feedback provided by participants towards the PN and SO episodes included in the study. 150

Table 5-5: Correlation matrix of posttest variables across the SO and PN cases. 153

Table 6-1: Estimation of CMP-NeuNAc concentration through LC-MS. 163

Table A-1: List of proteins presented to *Ng*PglO and *Nm*PglL that were not glycosylated by either *O*-OSTs. 173

Table C-1: Estimated Marginal Means (and standard errors of the means) of post-test measures based on group comparison ANCOVA analyses considering condition, gender, and race/ethnicity as predictors (*controlled for pre-test scores). 186

Table C-2: Estimated Marginal Means (and standard errors of the means) of post-test measures based on group comparison ANCOVA analyses considering condition, and intersection of gender with African American race/ethnicity as predictors (*controlled for pre-test scores). 187

CHAPTER 1

Metabolic engineering of glycoprotein biosynthesis in bacteria[1]

1.1 Abstract

The demonstration more than a decade ago that glycoproteins could be produced in *Escherichia coli* cells equipped with the *N*-linked protein glycosylation machinery from *Campylobacter jejuni* opened the door to using simple bacteria for the expression and engineering of complex glycoproteins. Since that time, metabolic engineering has played an increasingly important role in developing and optimizing microbial cell glyco-factories for the production of diverse glycoproteins and other glycoconjugates. It is becoming clear that future progress in creating efficient glycoprotein expression platforms in bacteria will depend on the adoption of advanced strain engineering strategies such as rational design and assembly of orthogonal glycosylation pathways, genome-wide identification of metabolic engineering targets, and evolutionary engineering of pathway performance. Here, we highlight recent advances in the deployment of metabolic engineering tools and strategies to develop microbial cell glyco-factories for the production of high-value glycoprotein targets with applications in research and medicine.

1.2 Introduction

Protein glycosylation is the process of covalently linking a complex carbohydrate known as a glycan to an amino acid side chain of an acceptor protein(*4*). It is one of the most common post-translational protein modifications and is present in all domains of life including bacteria(*5*) and archaea(*6*). In the case of bacteria, the existence of enzymatic pathways for attaching glycans to asparagine residues (*N*-linked

---

1This chapter has been previously published: Natarajan A., Jaroentomeechai T., Li M., Glasscock C.J., and DeLisa M.P. (2018). Metabolic engineering of glycoprotein biosynthesis in bacteria. Emerging Topics in Life Sciences, 2 (3) 419-432. DOI: 10.1042/ETLS20180004. URL: http://www.emergtoplifesci.org/content/2/3/419

glycosylation) or to serine or threonine residues (*O*-linked glycosylation) is firmly established and reviewed extensively(*7-10*). For both types, two distinct modes of glycan attachment have been observed in bacteria ***(Figure 1-1)***: (i) *en bloc* transfer where the glycan is first assembled on a lipid carrier molecule as a lipid-linked oligosaccharide (LLO) and then transferred from the lipid to an acceptor protein(*11*) and (ii) sequential glycosylation whereby the glycan is assembled directly on the protein by successive addition of monosaccharides by glycosyltransferase (GT) enzymes, typically occurring in the cytoplasm(*12-16*). For the former mode, glycan assembly is catalyzed by GTs on the cytoplasmic side of the inner membrane, while glycan transfer to protein is catalyzed by an integral membrane protein known as an oligosaccharyltransferase (OST) on the periplasmic side of the inner membrane. Importantly, both sequential and *en bloc* glycosylation mechanisms from diverse bacterial species have been at least partially if not completely reconstituted in laboratory strains of *Escherichia coli*(*17-20*), thereby providing important experimental proof-of-concept that this genetically tractable host organism can be leveraged for engineering novel glycoproteins and gaining access to unexplored regions of glycoprotein structural space as described in the sections that follow.

Before discussing specific examples, it is important to consider the distinct advantages that arise from producing glycoproteins in *E. coli* compared with eukaryotic expression hosts. For one, the viability of *E. coli* (or most bacteria for that matter), which do not natively possess glycosylation pathways, does not depend on protein glycosylation. This is not the case in eukaryotes where glycosylation is generally important for growth and development and, in the specific cases of *N*-glycosylation and *O*-mannosylation, essential for viability(*20*). Hence, manipulation of these native

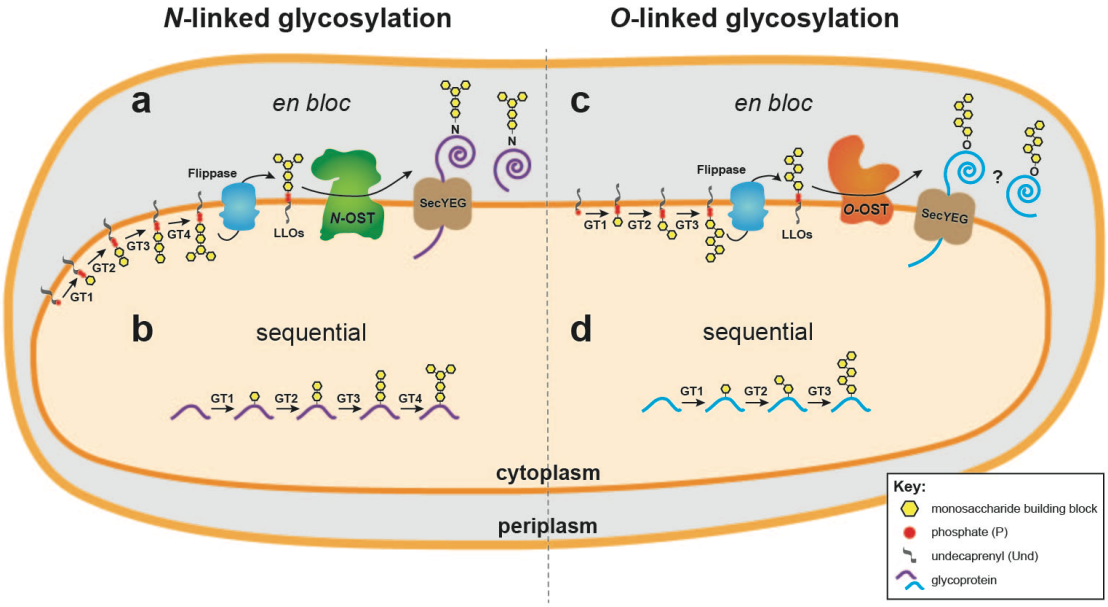


Figure 1-1: Protein glycosylation pathways reconstituted in E. coli.
Schematic illustration of the various different modes of protein glycosylation that have been functionally transferred to E. coli cells. (a) Periplasmic N-glycosylation in which a glycan is assembled on undecaprenyl pyrophosphate (Und-PP) on the cytoplasmic face of the inner membrane by a series of glycosyltransferases (GTs 1-4), flipped to the periplasmic face of the membrane by a native flippase enzyme, and then transferred en bloc to by an N-oligosaccharyltransferase (N-OST) to asparagine residues in acceptor proteins that are translocated into the periplasm. The N-OST is able to transfer glycans to proteins before folding is complete, co-translationally, or after folded is completed, post-translationally (1-3). (b) Periplasmic O-glycosylation is mechanistically similar to periplasmic N-glycosylation but an O-oligosaccharyltransferase (O-OST) is responsible for en bloc transfer of the glycan to serine or threonine residues in acceptor proteins. Little is known at present about the folding state of acceptor proteins recognized by the O-OST. (c) N-linked and (d) O-linked cytoplasmic glycosylation in which glycans are sequentially assembled on the acceptor protein directly, with initiation catalyzed by either an N-glycosyltransferase (NGT) or O-glyosyltransferase (OGT), respectively, followed by glycan elaboration by different glycosyltransferases.

glycosylation pathways (*i.e.* genetic deletion or overexpression of glycosyltransferase genes) in eukaryotic cells can severely compromise the overall fitness of the host(*21, 22*). Another major advantage that stems from the absence of native glycosylation systems is that *E. coli* provides a 'clean chassis' in which orthogonal protein

glycosylation pathways can be installed without interference from endogenous GTs, thereby leading to more uniformly glycosylated protein products(*23-25*).

It should also be pointed out, however, that the absence of natural glycosylation pathways also presents significant engineering challenges that relate to: (i) assembling functional pathways of GTs that catalyze the biosynthesis of a desired glycan either on a lipid carrier or on a protein target; (ii) accumulating sufficient levels of necessary building blocks (*i.e.* nucleotide-activated sugars, acceptor proteins); and (iii) conjugating glycan to protein with high transfer efficiency so as to maximize major performance indices, namely titer, yield, and productivity. In this review, we discuss how an emerging marriage between bacterial glycoprotein engineering and contemporary metabolic engineering is furnishing an armamentarium of new tools and strategies for overcoming the challenges associated with developing industrially competitive microbial cell glyco-factories. Of particular interest are recent advances and opportunities in the construction of artificial glycosylation pathways, genome-wide identification of metabolic engineering targets, creation of synthetic circuits for fine-tuning and controlling glyco-enzyme expression, and multiplex genome editing and evolutionary engineering for optimizing desirable glyco-phenotypes. Just as metabolic engineering has been instrumental in developing microorganisms for the production of various chemicals and materials to achieve a sustainable chemical industry, we anticipate that metabolic engineering of designer glycosylation pathways will similarly advance bacterial production platforms for rapidly and economically supplying structurally uniform glycoprotein therapeutics and vaccines for a wide range of human and animal diseases.

1.3 Design and assembly of orthogonal glycosylation pathways

The first bacterial glycoprotein engineering platform was described by Aebi and co-workers who functionally transferred the entire protein glycosylation (*pgl*) locus discovered in *Campylobacter jejuni*, which encodes a complete *N*-glycosylation pathway, into a laboratory strain of *E. coli*(*26*). The OST in this pathway, *C. jejuni* PglB (*Cj*PglB), catalyzed *en bloc* transfer of the preassembled *C. jejuni* glycan from the native lipid carrier in *E. coli*, undecaprenyl pyrophosphate (Und-PP), to an asparagine residue of an acceptor protein ***(Figure 1-1a)***. This pioneering system has since been leveraged as a model to mechanistically dissect the bacterial glycosylation mechanism, which shares important functional similarities with the eukaryotic process(*9*), and as a starting point for further glycoprotein engineering. For example, the compatibility of PglB with different glycan substrates has been explored by heterologous expression of diverse lipopolysaccharide biosynthesis pathway that also proceed via an Und-PP intermediate(*27*). These studies revealed that PglB has a relaxed specificity towards its glycan substrate(*28, 29*), minimally requiring an acetamido group at the C2 position of the reducing-end sugar(*29*). The substrate protein preferences for PglB have similarly been explored using *E. coli* cells engineered with the *pgl* locus. These studies identified a short consensus sequence (sequon) D/E-X1-N-X2-S/T (where X1, X2 ≠ Pro) that is strictly required for substrate recognition by *Cj*PglB(*1*) and that can be engineered into flexible secondary structures within a protein(*1, 30*) or at the N- or C-termini of the protein(*31*), enabling glycosylation of a wide range of heterologous proteins by simple introduction of a glycosylation tag. An important prerequisite for the glycosylation of these proteins is targeting to the periplasmic space, although substrate proteins that subsequently localize to the outer membrane or extracellular medium are also suitable candidates for PglB-mediated glycosylation(*31*).

Importantly, these mechanistic studies provided the requisite insight for making customized recombinant bacterial glycoproteins, where potentially any protein of interest could be modified with any glycan moiety at a desired position simply by co-expression of *Cj*PglB, a pathway for assembling sugars of interest on Und-PP, and the desired acceptor protein modified to contain one or more glycosylation sequons. Indeed, one of the first applications to leverage the unique ability of PglB to tolerate a broad range of glycan substrates and transfer these to a short, specific motif embedded in heterologous proteins was the production of glycoconjugate vaccines (for reviews, see refs(*32-34*)). Specifically, by transferring the locus encoding an *O*-polysaccharide antigen or a capsular polysaccharide into *E. coli* cells carrying a copy of *Cj*PglB and an appropriate antigenic carrier protein, glycoproteins can be produced that elicit an immune response to the corresponding polysaccharide epitope found on the surface of a pathogenic bacterium(*33, 35-38*) and some of these have even advanced into human clinical trials(*39-41*). It is also possible to transfer *Cj*PglB and a carrier protein into a pathogenic host strain, resulting in rerouting of the pathogen's naturally produced Und-PP-linked polysaccharide onto the carrier(*42-44*), which is a viable alternative in cases where the genetic locus encoding the polysaccharide pathway is undetermined or locus is recalcitrant to cloning.

The aforementioned examples take advantage of glycan structures that are built through the enzymatic potential of a single organism, either in the native host or through heterologous transfer of intact gene clusters to *E. coli*. In a notable departure from this approach, several groups have reported the rational design of artificial operons comprising GTs from different organisms that co-ordinate the assembly of desired glycan structures — including those produced by human cells — on Und-PP. For example, engineering of a heterologous pathway comprising GTs from four different

bacterial species enabled production of a glycoprotein bearing the human Lewis X (Lex) antigen(*45*). Along similar lines, our group has engineered a synthetic pathway comprising four GTs that enabled the production of the eukaryotic mannose$_3$-N-acetylglucosamine$_2$ ($Man_3GlcNAc_2$) glycan(*3*), which serves as the core structure common to all eukaryotic *N*-glycans including those produced in humans(*46*). In a notable departure from previous studies that relied on bacterial GTs for *N*-glycan biosynthesis, we expressed four GTs that were sourced from *Saccharomyces cerevisiae* and functionally reconstituted in the cytoplasm of *E. coli*. To increase the pool of GDP-Man required by the yeast mannosyltransferases, Alg1 and Alg2, we performed genome editing using the λ Red recombineering system(*47*) to delete the gene encoding GDP-mannose dehydratase (GMD), which converts GDP-Man to GDP-4-keto-6-deoxymannose in the first step of GDP-L-fucose (GDP-Fuc) synthesis. In cells producing Und-PP-linked $Man_3GlcNAc_2$, *Cj*PglB was capable of transferring the trimannosyl core glycan to a range of acceptor proteins including bovine RNase A and a human Fc domain when the native acceptor site in these glycoproteins was changed to a bacterial sequon (DQNAT). Importantly, this was the first report to describe the engineering of an orthogonal eukaryotic *N*-glycosylation pathway in *E. coli*. While the initial product yields were low (~50 μg/l), this has now been improved through metabolic engineering strategies discussed below.

To date, the majority of PglB-mediated glycoprotein engineering studies employ a two- or even three- plasmid system to transfer all of the requisite machinery along with the acceptor protein into *E. coli*. A consequence of this multi-plasmid approach is the metabolic burden placed on the host that can cause slow growth, plasmid loss, and product degradation by up-regulation of stress responses. To remedy this issue, efforts to integrate PglB into the genome have been described albeit with glycosylation

efficiency that is well below 50%(*48*). We recently performed a systematic glyco-recoding of the *E. coli* genome using a targeted genome editing strategy to site-specifically insert PglB and the glycosyation pathway enzymes in place of native oligosaccharide operons, namely the *O*-polysaccharide antigen and enterobacterial common antigen, that are unnecessary for viability or glycosylation(*49*). By recombineering the complete *N*-glycosylation machinery to these two genomic loci, our host strain only required transformation with a single plasmid encoding the acceptor protein. As a result, cells reached significantly higher final densities and volumetric productivity was greatly increased. Moreover, glycosylation efficiency in these strains was at or near 100%, rivaling or exceeding the best glycosylation that could be obtained from plasmid-based systems.

While PglB-mediated glycosylation remains the archetypical glycoprotein engineering platform, several noncononical modes of glycosylation have been functionally reconstituted in *E. coli*. Most extensively studied among these is periplasmic *O*-oligosaccharyltransferase (*O*-OST)-mediated glycosylation, which is mechanistically similar to bacterial *N*-glycosylation including glycan assembly on Und-PP and *en bloc* transfer by the *O*-OST to acceptor proteins ***(Figure 1-1b)***. The first periplasmic *O*-glycoprotein engineering platforms in *E. coli* involved the *O*-OSTs PilO from *Pseudomonas aeruginosa* and PglL from *Neisseria meningitides*(*17*). In contrast to *Cj*PglB, which prefers glycans having an acetamido group at the C2 position of the reducing-end sugar(*29*) and that lacks a β1–4 linkage between the first two sugars(*50*), PglL appears able to transfer a wider range of sugars(*51*), suggesting that it may be useful for a broader range of applications. For example, direct transfer of PglL along with a suitable carrier protein into attenuated pathogens resulted in *O*-linked glycoproteins bearing antigenic polysaccharide structures including some that were

known to be incompatible with the more restrictive PglB enzyme(*52, 53*). Importantly, several of the *O*-glycoprotein conjugates evoked specific and protective immune responses in laboratory animals. These same authors identified a minimal optimal *O*-linked recognition (MOOR) motif comprising eight residues that, when flanked by additional hydrophilic flanking sequences, was able to direct PglL-mediated glycosylation to heterologous proteins(*53*). It is worth mentioning that the *O*-OSTs from diverse *Bacteroidetes* appear to modify proteins at a minimal 3-residue motif, D-(S/T)-(A/L/V/I/M/T)(*54*), which is enticing from a glycoprotein engineering standpoint; however, the broad utility of these particular enzymes will depend on their glycan substrate specificity, which has yet to be determined.

Perhaps, the simplest platform for glycoprotein engineering is based on sequential *N*-glycosylation ***(Figure 1-1c)*** whereby *E. coli* cells expressing the *N*-glycosyltransferase from *Actinobacillus pleuropneumoniae* (*Ap*NGT) use nucleotide-activated glucose (Glc) to glycosylate asparagines of target proteins expressed in the cytoplasm(*18*). A useful feature of *Ap*NGT is its relaxed substrate specificity, requiring only a minimal N-X-S/T sequon that allows glycosylation of heterologous glycoproteins such as human erythropoietin at their native *N*-glycosylation sites. While a single *N*-linked Glc is of limited utility on its own, several groups have used this monosaccharide for priming the biosynthesis of glycan polymers including dextran(*55*) and polysialic acid(*56*) that might find use in vaccine applications or half-life extension of protein drugs.

Cytoplasmic pathways for sequential *O*-glycosylation in *E. coli* have also been described ***(Figure 1-1d)***. In one instance, the complete pathway for stepwise glycosylation of *Streptococcus parasanguinis* Fap1 was transferred to *E. coli*, including

an initiating *O*-glycosyltransferase (OGT) complex comprising GtfA and GtfB that generates *O*-linked GlcNAc and several additional GTs that elaborate the glycan(*19*). Efforts to re-engineer the glycan have yet to be described, but a 25-residue glycosylation sequence has been identified that might permit glycosylation of heterologous proteins bearing this motif(*57*). In a related example, human UDP-GalNAc/polypeptide UDP-GalNAc transferase (GalNAc-T), an OGT that transfers *N*-acetylgalactosamine (GalNAc) from UDP-GalNAc to certain serines or threonines was expressed in the cytoplasm of *E. coli* to execute the initial step of human, mucin-type *O*-glycosylation(*58*). Because nonpathogenic *E. coli* strains produce UDP-GlcNAc, but not UDP-GalNAc, co-expression of the UDP-GlcNAc C-4 epimerase WbpP from *P. aeruginosa* was required to convert UDP-GlcNAc to UDP-GalNAc. In cells producing UDP-GalNAc, GalNAc-T was able to glycosylate several fusion proteins containing GalNAc-T recognition sequences derived from eukaryotic *O*-glycoproteins including the human mucin-based peptide, CDNKPAPGSTAPPAA, and the *Sus scrofa* plasminogen-based peptide, PPPTSGPT. The authors subsequently used the *O*-linked GalNAc as a handle for *in vitro* chemical modification with poly-ethylene glycol (PEG), yielding a glyco-PEGylated protein product. While not the focus of the study, one can envision the installation of GalNAc as a starting point for the bacterial biosynthesis of more elaborate mucin-type glycans, including the tumor-associated T-antigen disaccharide (Galβ1,3GalNAcα) and sialic acid-containing versions thereof. A key feature of this approach is the ability to form fully human glycan linkages, which is not currently possible with the native bacterial NGT and OGT enzymes described above and thus limits their biotechnological utility.

1.4 Optimization of glycosylation pathway performance

In tandem with the creation of orthogonal glycosylation pathways in *E. coli*, significant efforts have been made towards optimizing the overall glycolipid and glycoprotein yield, glycosylation efficiency (or site occupancy), and product homogeneity of these platforms. These can be broadly categorized into efforts that (i) improve bioprocess conditions through manipulation of media compositions, temperature, and pathway induction; (ii) optimize host strains and pathway enzyme expression levels to increase flux towards protein, lipid, and glycan synthesis; and (iii) engineer enzymes that constitute the pathway. Each of these approaches has been shown to improve glycosylation pathway performance by relieving bottlenecks in specific steps, paving the way to a combined, holistic approach to make new bacterial glycosylation pathways commercially viable.

In the case of optimizing bioprocess conditions, many relatively straightforward approaches have been reported to improve glycosylation pathway performance. For example, during the production of a glycoconjugate vaccine, it was found that biomass production was particularly sensitive to induction of the *Cj*PglB OST, revealing a clear need for separation of biomass growth and glycosylation pathway expression(*59*). In addition, a 30–50-fold improvement in glycosylation of carrier proteins (up to 24 mg/l of glycosylated carrier protein) was achieved by optimizing a fed-batch reactor feeding strategy and medium composition. A subsequent study provided further evidence that induction timing of *Cj*PglB is critical for glycosylation efficiency, with a delay between PglB induction and carrier protein expression proving to be optimal(*60*). It was suggested that functional expression and proper insertion of *Cj*PglB at the time of carrier protein translocation may allow better glycan transfer, possibly before carrier folding. These authors also showed that optimizing Mg^{2+} in the media had a small effect on

glycosylation efficiency. Along similar lines, the use of autoinduction instead of IPTG-based induction modestly improved *Cj*PglB-mediated glycosylation of a different target protein with either *C. jejuni* or *Haemophilus influenzae* glycans(*61*). It is widely recognized that the use of autoinduction media helps overcome the inherent instability of inducible strains, thereby reducing metabolic burden(*62*). This burden can cause the levels of endogenously produced nucleotide-activated sugars to be limiting, in which case supplementation of the media with the corresponding sugar can help to alleviate the bottleneck(*60*). In certain instances, supplementation is required because the sugar is natively absent, such as with sialic acid in K-12 strains of *E. coli*(*23*). Supplementing bacterial growth medium with monosaccharides also presents opportunities for metabolic labeling with azido-sugars, resulting in the incorporation of azide functional groups into newly synthesized polysaccharides. For example, metabolic labeling of *Helicobacter pylori* with an acetylated azido GlcNAc analog was used to identify glycoproteins in this bacterium(*63*) while labeling of *Bacteroides fragilis* with an acetylated azido GalNAc analog enabled imaging of bacterial polysaccharides and tracking of host–microbiota interactions(*64*). Such labeling strategies could also be used in high-throughput screens to optimize protein glycosylation efficiency in *E. coli*.

In addition to bioprocess considerations, optimization of host strains and pathway enzyme expression levels offers a powerful route to improving glycosylation pathway performance. This can be accomplished by adding, deleting, up-regulating, or knocking down the expression of host genes. A simple yet effective strategy is to rationally identify host enzymes that are known to catabolize desired substrates and delete the genes encoding these enzymes, as illustrated by deletion of *gmd* to increase GDP-Man pools in *E. coli*(*3*), discussed above. Other examples include deletion of *waaL*, which encodes the *E. coli O*-antigen ligase, to eliminate unwanted transfer of

Und-PP-linked substrates to lipid A-core(*28*) and deletion of the *E. coli nanA* gene to remove sialic acid aldolase activity and ensure that sialic acid supplemented in the media is not catabolized and instead remains available for activation by the CMP-sialic acid synthetase and subsequent incorporation into polysialic acid(*23*). The same rational approach has been used to identify target genes whose overexpression improves glycosylation. For instance, overexpression of the genes encoding the initiating glycosyltransferase WecA has been reported to improve glycosylation efficiency with the *C. jejuni* glycan by PglB(*65*) while overexpression of the genes encoding phosphomannomutase (ManB) and mannose-1-phosphate guanylyl transferase (ManC) in the GDP-Man pathway helped to increase production of the $Man_3GlcNAc_2$ glycan by nearly 50-fold to 13.9 μg/l(*66*). As discussed above, placement of glycosylation enzymes including *Cj*PglB into the genome has been shown to improve glycosylation pathway performance by reducing burden on the host(*48, 49*). Codon optimization has also been used to improve expression of PglB, which positively affected glycosylation efficiency with the *C. jejuni* glycan in *E. coli*(*65*). In a more nuanced approach, a recent study by our group has shown that manipulating expression level by promoter engineering, in combination with overexpression of ManB and ManC, led to production of up to 10 times more glycosylated acceptor protein (to ∼14 mg/l) (*66*). This large improvement represented a two-orders-of-magnitude improvement compared with our earlier report(*3*) and rivaled the yield (25 mg/l) reported for *E. coli*-based production of carrier proteins glycosylated with bacterial polysaccharides(*59*).

To identify less obvious metabolic engineering targets that affect glycosylation, several groups have turned to computer-aided flux balance analysis and high-throughput experimental approaches. For example, we used a constraint-based model of *E. coli* metabolism in combination with heuristic optimization to design gene knockout strains

that overproduced the *C. jejuni* glycan precursor(*67*). Predicted knockouts were subsequently constructed in *E. coli* and experimentally verified to improve glycan production. In a related example, shotgun proteomics was combined with metabolic network analysis to forward engineer *E. coli* for improved glycosylation of the *C. jejuni* AcrA glycoprotein(*68*). This approach revealed that enhancing flux through the glyoxylate cycle, for example by up-regulating isocitrate lyase expression, could increase the amount of glycosylated protein by up to 3-fold. The same group described an inverse metabolic engineering strategy in which screening of genomic DNA libraries was used to identify genes (*malQ, rseP, ptsA, rffE,* and *dxs*) whose overexpression in *E. coli* improved production of the heterologously expressed *C. jejuni* glycan and overall glycoprotein yield, albeit without significantly enhancing glycosylation efficiency(*69*).

Efforts to improve glycosylation pathway performance have more recently expanded to include enzyme engineering strategies. Both rational design and random mutagenesis-based approaches have proved to be particularly useful for tailoring important properties of glycosylation pathway enzymes including catalytic activity, specificity, and selectivity, thereby relieving system constraints that limit bacterial glycoprotein engineering platforms. To address the restricted sequon specificity of *Cj*PglB, we employed a rational approach whereby structure-guided mutagenesis of *Cj*PglB active-site residues was used to generate a combinatorial library that was screened using a high-throughput replica-plating strategy termed glycoSNAP (glycosylation of secreted *N*-linked acceptor proteins)(*70*). Library screening yielded mutants of *Cj*PglB with broadly relaxed specificity, which probably arose from a more open peptide-binding pocket. Interestingly, when these mutations were mapped onto *Campylobacter lari* PglB, the relaxed acceptor sequon specificity was maintained, suggesting that these mutations are broadly applicable to many homologs. This

structure-guided PglB engineering approach was complemented with the metagenomic mining of *Cj*PglB homologs that also exhibited relaxed sequon specificity, in particular the PglB enzyme from *Desulfovibrio gigas*(*71*). Additional efforts for structure-guided engineering of *Cj*PglB uncovered mutants with 16-fold increased activity to a non-native polysaccharide substrate(*72*).

As a complement to structure-guided approaches, directed evolution has been used to engineer specific traits of glycosylation enzymes. The success of directed evolution for glyco-enzymes hinges on the non-trivial task of developing genetic screens or selections that enable isolation of rare mutants from large populations ($10^{10–12}$ clones) by creating a crucial genotype-to-glycophenotype linkage. In seminal work, Withers and co-workers engineered the sugar-transfer activity of the sialyltransferase CstII from *C. jejuni* using a high-throughput assay that relied on the fact that fluorescently labeled acceptor sugars can freely diffuse across the bacterial cell membrane, but the transfer products are selectively trapped in cells. Using this assay, a combinatorial library containing more than a million CstII variants was screened via fluorescence-activated cell sorting, and a mutant exhibiting 400-fold higher catalytic activity was discovered(*73*). Similar mutagenesis and screening approaches have since been developed to evolve other GTs such as the *Neisseria meningitidis* serogroup B polysialyltransferase(*74*) and the *C. jejuni* β-1,3-galactosyltransferase CgtB(*75*).

Directed evolution in the context of PglB-mediated glycosylation has been performed using enzyme-linked immunosorbent assay (ELISA)-based detection of periplasm-extracted glycoproteins in 96-well plates, uncovering inactivating mutations in *Cj*PglB and corroborating the necessity of a strictly conserved WWDYG motif(*76*). However, the ELISA-based screening approach, while effective, was limited to

relatively small libraries containing ~5000 or fewer clones and thus incapable of deeply penetrating protein sequence space as is desirable in directed evolution experiments. Another concern with extract- or cell lysis-based screens(*69, 76*) is their potential to be confounded by the prevalence of glycan intermediates that have not been transferred to proteins (for example, LLOs in bacterial cell membranes), increasing the likelihood for false-positive hits. These issues can be overcome using genetic assays where final glycosylated products are effectively segregated from their precursors such as in cell-surface display of glycoconjugates(*3, 31, 66*), glycophage display(*77, 78*), and glycoSNAP(*70*), with the latter two methods having already shown promise for screening rationally designed combinatorial libraries of acceptor sequons and glycosylation enzymes.

In contrast with the growing suite of genetic screens that have been reported, genetic selections for interesting glyco-phenotypes have proved more challenging because these require associating the glycosylation state of a protein with survival of a host that does not natively produce such protein products. At present, only one such selection has been described for bacterial glycosylation and is based on the commonly observed phenomenon of glycan-induced alterations in protein folding(*79, 80*). Specifically, TEM-1 β-lactamase was fused to a glycoprotein substrate in cells carrying the *C. jejuni* glycosylation machinery such that changes in periplasmic folding and stability caused by *N*-glycosylation of the substrate resulted in measurable differences in bacterial resistance to β-lactam antibiotics(*81*). The selection was then used to explore the effects of *Cj*PglB-mediated glycosylation on *in vivo* protein folding and stability and could potentially enable the isolation of modified proteins with advantageous folding properties. However, there have been no reports of using the assay for high-throughput screening of combinatorial libraries. Finally, it is worth mentioning that the application

of screens and selections for general protein folding and stability(*82, 83*), independent from glycosylation, could be used to overcome the well-documented insolubility and aggregation issues associated with heterologous expression of glycosylation enzymes in *E. coli*. Such solubility-enhanced, aggregation-resistant enzymes are prized for glycoprotein engineering applications.

1.5 Cell-free metabolic engineering of glycoprotein biosynthesis

Cell-free metabolic engineering (CFME) is emerging as a complementary method to *in vivo* approaches for accelerating the design, build, and test (DBT) cycles of biochemical pathways. From a metabolic engineering perspective, cell-free systems

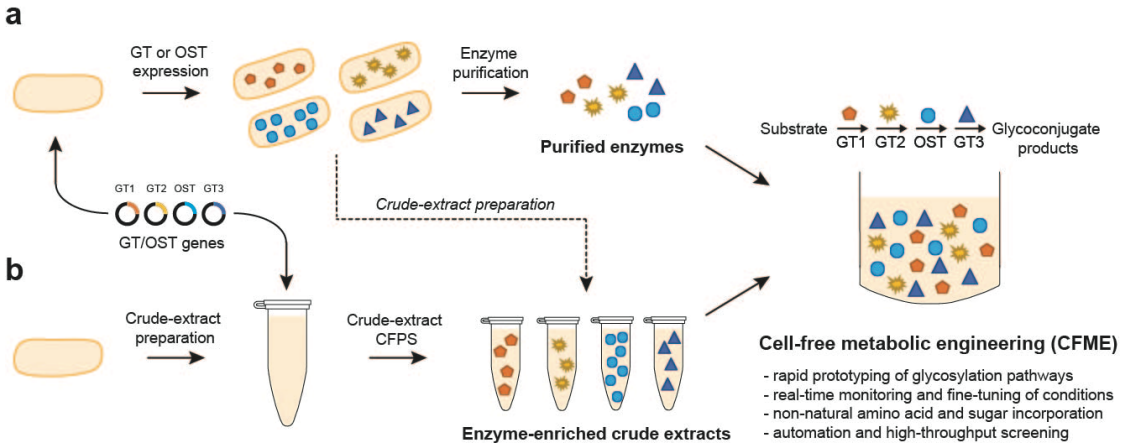


Figure 1-2: Cell-free metabolic engineering (CFME) of glycoprotein biosynthesis. *(a) Schematic of purified enzyme CFME in which glycosyltransferases (GTs) and/or oligosaccharyltransferases (OSTs) are overexpressed in living cells and then purified. The purified enzymes derived from these cells are then used to assemble specific glycosylation pathways in vitro. It should be noted that the process of molecular cloning, optimization of enzyme expression, and purification can be costly and time consuming. (b) Schematic of crude-extract CFME in which glycosylation enzymes are synthesized in crude extract, simply by augmenting cell lysate with DNA encoding the glycosylation enzyme of interest. Alternatively, cells transformed with DNA encoding the glycosylation enzyme of interest can be used to prepare enzyme-enriched crude extracts containing the GTs/OSTs (dashed arrow). Crude extracts enriched with a glycosylation enzyme are then combined, allowing rapid prototyping of novel metabolic pathways for glycan and glycoconjugate production.*

decouple cell survival and engineering goals, offering a unique opportunity to redirect nearly all carbon flux towards producing desired products. Prokaryote-based cell-free systems were introduced over a century ago, but have only recently been applied to glycan and glycoprotein biosynthesis with an eye towards maximizing product homogeneity and yield as well as characterizing the relevant enzymatic reactions. The two major types of CFME approaches are based on: (i) purified enzymes and (ii) crude extracts ***(Figure 1-2)***.

The principle of purified enzyme CFME is to express, purify[2], and combine individual enzymes in a manner that reconstitutes a biochemical pathway. This approach offers exquisite control over reaction flux, allowing the precise tuning of concentration, activity, and activation time of each component. Hence, this method has become a powerful tool to understand biotransformation, decipher the precise function and kinetics of GTs and OSTs, and identify key metabolites and feedback control of mechanisms that regulate glycan or glycoprotein biosynthesis. One of the first reports to achieve *in vitro* glycosylation pathway assembly involved combining *Cj*PglB, a synthetic glycolipid, and peptide substrates(*84*). Shortly thereafter, *Cj*PglB-mediated *N*-glycosylation of full-length prokaryotic proteins was successfully reconstituted *in vitro*(*1*). These seminal studies provided an *in vitro* framework for investigating fundamental questions about bacterial *N*-glycosylation. Combined with advances in mass spectrometry (MS)-based glycoprotein analysis and fluorescence-based *in vitro* activity assays, cell-free glycosylation using purified components has now become a standard tool for investigating glycan and polypeptide substrate specificity, kinetic profiles, and reaction mechanisms of bacterial GTs, OSTs, and glycan flippases(*71, 85*-

[2] Owing to a lack of robust membrane protein purification protocols, several early studies used crude-membrane extracted OSTs instead of purified enzymes in cell-free reaction. However, we consider these reactions as purified systems because the OST has been overexpressed and isolated from cells.

88). In addition, the open nature of cell-free systems permits the use of synthetic substrates to interrogate OST activity. For example, it was reported that *Cl*PglB can modify several synthetic peptide analogs(*89*), while *Cj*PglB can transfer monosaccharides attached to unnatural C20 isoprenoid lipid carriers onto protein substrates(*90*). These studies are not only valuable for deciphering bacterial protein *N*-glycosylation mechanisms but also broaden the repertoire of substrates that can be used with bacterial glycoprotein engineering systems and pave the way for a low-cost *in vitro* route to custom-made glycoproteins. While much of the work in purified enzyme CFME has centered on *Cj*PglB, other OSTs have proved amenable to cell-free applications including bacterial *O*-OSTs(*51, 91*), bacterial NGTs(*13, 92, 93*), single-subunit protozoan OSTs(*94*), archaeal OSTs(*95*), and even the multi-subunit eukaryote OST complex from *S. cerevisiae*(*96*).

Purified enzyme CFME is also a powerful method to construct pathways for glycan assembly on lipid carriers, which can later be used as a substrate by OSTs. Since glycan synthesis often depends on the concomitant expression of several GTs, sometimes derived from different organisms, the ability to monitor each of their activities both independently and in concert can speed up the DBT process towards pathway assembly. This approach has been exemplified in the construction of a multi-step enzymatic pathway for the biosynthesis of the *C. jejuni N*-linked glycan(*97-99*) and eukaryotic $GlcNAc_2Man_{1-5}$(*100, 101*) on lipid carriers, which are then transferred onto peptide or protein acceptors using their respective OSTs. Finally, two-stage glycosylation strategies that involve the *in vivo* production of an initial glycoprotein, whose glycans are then remodeled *in vitro* using purified GTs, have enabled the synthesis of glycoconjugates with complex glycan structures including Lex antigen(*45*) as well as $Man_9GlcNAc_2$ and biantennary complex-type *N*-glycans(*102*).

In contrast with purified protein CFME, crude extract systems bypass laborious protein production and purification by employing cell extracts containing relevant enzymes. The extract is made by simply collecting the supernatant from cell lysates enriched with protein synthesis machineries. Currently, the best-characterized and most widely adopted crude extract CFME is from *E. coli*, which can activate long enzymatic pathways of over eight enzymes with productivities exceeding 100 mg/ml-h and in scales of 100 l(*103-106*). The application of *E. coli* crude extract CFME in endeavors of glycoengineering is limited owing to the absence of endogenous glycosylation machinery. This was initially overcome through augmenting commercial *E. coli* S30 extracts with purified *C. jejuni N*-linked glycosylation components, yielding glycoproteins at ~1 mg/ml(*107*). Our more recent work reports the development of a fully integrated *E. coli* cell-free glycoprotein synthesis (CFGpS) technology(*108*). CFGpS enables a one-pot reaction for efficient and site-specific glycosylation of target proteins, requiring only the addition of DNA into the reaction. This platform is highly modular allowing the interchange of multiple distinct OSTs and structurally diverse LLOs, including the eukaryotic *N*-glycan $Man_3GlcNAc_2$.

The ability to rapidly synthesize and characterize GTs would greatly benefit the study and development of glycosylation systems. The bottleneck of expressing these enzymes, which are usually integral transmembrane proteins, in a cell-free system has recently been overcome(*109*). By supplying *E. coli* extracts with optimal concentrations of nanodisc lipid scaffolds, soluble expression of functional OSTs from the *Campylobacter species* has been achieved. The full potential of cell-free strategies towards characterizing GTs has recently been showcased through the glycosylation sequence characterization and optimization by rapid expression and screening (GlycoSCORES) technology, a high-throughput screening platform that couples crude

extract-based glycosylation with self-assembled monolayers for matrix-assisted desorption/ionization (SAMDI) MS(*93*). The automation of this platform should accelerate the pace of characterizing relevant enzymes. Data gathered through this strategy can be used to generate cell-free kinetic models(*67, 110, 111*) and incorporated into cell-free machine learning algorithms (*112, 113*) to guide the design of glycoprotein and glycan biosynthesis pathways.

1.6 Concluding remarks

The identification of several distinct *N*- and *O*-linked protein glycosylation systems in a variety of bacterial species and their functional transfer to genetically tractable *E. coli* cells has greatly enhanced the potential of bacterial systems for the generation of therapeutically relevant glycoconjugates. These bacterial glycoprotein engineering platforms have been employed to generate well-defined therapeutic compounds including the first conjugate vaccines produced entirely in bacteria, two of which have recently undergone successful phase I clinical trials(*39-41*).

Owing to their versatility and ease of manipulation, bacteria are an ideal host for the production of a diverse array of structurally defined polysaccharides and glycoconjugates that will be of interest as medical and industrial products. Furthermore, the low cost of goods associated with the culturing of bacterial strains, especially *E. coli*, opens up this technology to a far wider range of laboratories than existing chemical/chemoenzymatic synthesis methods or mammalian cell culture approaches. The findings from a recent report commissioned by the National Academy of Sciences state that ‘glycans play roles in almost every biological process and are involved in every major disease’ and further assert that ‘the development of transformative methods for the facile synthesis of carbohydrates and glycoconjugates should be a high priority’

(National Research Council 2012). Bacterial glycoengineering represents an emerging field with the potential to play a major role in meeting these goals. While recent work shows great promise for the design and construction of orthogonal glycosylation pathways in *E. coli*, there is still a critical need to bring major performance indices, namely titer, yield, and productivity to production scale by tackling bottlenecks in cell metabolism, pathway enzyme expression, and catalytic transformation of precursors to products. The marriage of metabolic engineering with bacterial cell-based and cell-free glycoprotein production systems is helping to overcome these bottlenecks by fine-tuning and optimizing the metabolic fluxes that lead to increased yield and concentration of desired glyco-products while limiting formation of unwanted byproducts, creating novel glycosylation enzymes by rational design and random mutagenesis-based strategies and integrating these engineered catalysts into synthetic gene regulatory circuits, increasing the tolerance of cells to the burden of expressing multi-enzyme glycosylation pathways, and even the development of cost-effective fermentation strategies and downstream processes. Complementing these experimental strategies are computational methods for characterizing glycosylation phenotypes from a holistic perspective and predicting gene targets throughout the whole cellular network that contribute to the enhanced production of glycoprotein targets. It is anticipated that the continued deployment of advanced metabolic engineering tools and strategies will increase the efficiency of strain development, leading to a veritable transformation of bacteria into living glyco-factories for scalable, bottom-up production of a wide array of complex glycoconjugates with applications in research and medicine.

1.7 Summary

- The identification of several distinct *N*- and *O*-linked protein glycosylation systems in a variety of bacterial species and their functional transfer to genetically tractable

E. coli cells has greatly enhanced the potential of bacterial systems for the generation of glycoconjugates.

- Owing to their versatility and ease of manipulation, bacteria are an ideal host for the production of a diverse array of structurally defined polysaccharides and glycoconjugates that will be of interest as medical and industrial products.
- The marriage of metabolic engineering with bacterial cell-based and cell-free glycoprotein production systems is increasing the efficiency of strain development, leading to a veritable transformation of bacteria into living glycol-factories for scalable, bottom-up production of a wide array of complex glycoconjugates.

1.8 Abbreviations

*Ap*NGT, *N*-glycosyltransferase from *Actinobacillus pleuropneumoniae*; CFGpS, cell-free glycoprotein synthesis; CFME, cell-free metabolic engineering; DBT, design, build, and test; ELISA, enzyme-linked immunosorbent assay; GalNAc, *N*-acetylgalactosamine; GalNAc-T, GalNAc transferase; Glc, glucose; glycoSNAP, glycosylation of secreted *N*-linked acceptor proteins; GMD, GDP-mannose dehydratase; GT, glycosyltransferase; Lex, Lewis X; LLO, lipid-linked oligosaccharide; $Man_3GlcNAc_2$, $mannose_3$-*N*-$acetylglucosamine_2$; ManC, mannose-1-phosphate guanylyl transferase; MS, mass spectrometry; OGT, *O*-glycosyltransferase; *O*-OST, *O*-oligosaccharyltransferase; OST, oligosaccharyltransferase; PEG, poly-ethylene glycol; Pgl, protein glycosylation; Und-PP, undecaprenol pyrophosphate.

1.9 Funding

This work was supported by the Defense Threat Reduction Agency (grant 11631647 to M.P.D.) and the National Science Foundation (grants # CBET-1159581, CBET-1264701, CBET-1402843, CBET-1605242, and MCB-1413563 to M.P.D.). T.J.

was supported by a Royal Thai Government Fellowship. C.J.G. was supported by a National Science Foundation Graduate Research Fellowship (grant # DGE-1144153).

1.10 Competing interests

M.P.D. has a financial interest in Glycobia, Inc. and Versatope, Inc. M.P.D.'s interests are reviewed and managed by Cornell University in accordance with their conflict of interest policies.

1.11 References

1. M. Kowarik *et al.*, **N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase.** *Science* 314, 1148-1150 (2006).
2. J. M. Silverman, B. Imperiali, **Bacterial N-Glycosylation Efficiency Is Dependent on the Structural Context of Target Sequons.** *J Biol Chem* 291, 22001-22010 (2016).
3. J. D. Valderrama-Rincon *et al.*, **An engineered eukaryotic protein glycosylation pathway in Escherichia coli.** *Nat Chem Biol* 8, 434 (2012).
4. F. Schwarz, M. Aebi, **Mechanisms and principles of N-linked protein glycosylation.** *Curr Opin Struct Biol* 21, 576-582 (2011).
5. C. M. Szymanski, B. W. Wren, **Protein glycosylation in bacterial mucosal pathogens.** *Nat Rev Microbiol* 3, 225-237 (2005).
6. J. Eichler, **Extreme sweetness: protein glycosylation in archaea.** *Nat Rev Microbiol* 11, 151-156 (2013).
7. H. Nothaft, C. M. Szymanski, **Protein glycosylation in bacteria: sweeter than ever.** *Nat Rev Microbiol* 8, 765-778 (2010).
8. H. Nothaft, C. M. Szymanski, **Bacterial protein N-glycosylation: new perspectives and applications.** *J Biol Chem* 288, 6912-6920 (2013).

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

10. J. A. Iwashkiw, N. F. Vozza, R. L. Kinsella, M. F. Feldman, **Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation.** *Mol Microbiol* 89, 14-28 (2013).

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

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

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

14. S. Grass *et al.*, **The Haemophilus influenzae HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis.** *Mol Microbiol* 48, 737-751 (2003).

15. Y. Chen, R. Seepersaud, B. A. Bensing, P. M. Sullam, T. A. Rapoport, **Mechanism of a cytosolic O-glycosyltransferase essential for the synthesis of a bacterial adhesion protein.** *Proc Natl Acad Sci U S A* 113, E1190-1199 (2016).

16. Q. Lu *et al.*, **An iron-containing dodecameric heptosyltransferase family modifies bacterial autotransporters in pathogenesis.** *Cell Host Microbe* 16, 351-363 (2014).

17. A. Faridmoayer, M. A. Fentabil, D. C. Mills, J. S. Klassen, M. F. Feldman, **Functional characterization of bacterial oligosaccharyltransferases**

involved in O-linked protein glycosylation. *J Bacteriol* 189, 8088-8098 (2007).

18. A. Naegeli *et al.*, **Molecular analysis of an alternative N-glycosylation machinery by functional transfer from Actinobacillus pleuropneumoniae to Escherichia coli.** *J Biol Chem* 289, 2170-2179 (2014).

19. F. Zhu *et al.*, **Engineering and dissecting the glycosylation pathway of a Streptococcal serine-rich repeat adhesin.** *J Biol Chem* 291, 27354-27363 (2016).

20. L. Lehle, S. Strahl, W. Tanner, **Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases.** *Angew Chem Int Ed Engl* 45, 6802-6818 (2006).

21. B. Jiang *et al.*, **Inactivation of a GAL4-like transcription factor improves cell fitness and product yield in glycoengineered Pichia pastoris strains.** *Appl Environ Microbiol* 81, 260-271 (2015).

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

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

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

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

26. M. Wacker *et al.*, **N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli.** *Science* 298, 1790-1793 (2002).

27. I. Hug, M. F. Feldman, **Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria.** *Glycobiology* 21, 138-151 (2011).

28. M. F. Feldman *et al.*, **Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli.** *Proc Natl Acad Sci U S A* 102, 3016-3021 (2005).

29. M. Wacker *et al.*, **Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems.** *Proc Natl Acad Sci U S A* 103, 7088-7093 (2006).

30. M. Kowarik *et al.*, **Definition of the bacterial N-glycosylation site consensus sequence.** *EMBO J* 25, 1957-1966 (2006).

31. A. C. Fisher *et al.*, **Production of secretory and extracellular N-linked glycoproteins in Escherichia coli.** *Appl Environ Microbiol* 77, 871-881 (2011).

32. V. S. Terra *et al.*, **Recent developments in bacterial protein glycan coupling technology and glycoconjugate vaccine design.** *J Med Microbiol* 61, 919-926 (2012).

33. J. Cuccui *et al.*, **Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis.** *Open Biol* 3, 130002 (2013).

34. J. Cuccui, B. Wren, **Hijacking bacterial glycosylation for the production of glycoconjugates, from vaccines to humanised glycoproteins.** *J Pharm Pharmacol*, (2014).

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

36. Z. Ma *et al.*, **Glycoconjugate vaccine containing Escherichia coli O157:H7 O-antigen linked with maltose-binding protein elicits humoral and cellular responses.** *PLoS One* 9, e105215 (2014).

37. M. Wacker *et al.*, **Prevention of Staphylococcus aureus infections by glycoprotein vaccines synthesized in Escherichia coli.** *J Infect Dis* 209, 1551-1561 (2014).

38. M. Wetter *et al.*, **Engineering, conjugation, and immunogenicity assessment of Escherichia coli O121 O antigen for its potential use as a typhoid vaccine component.** *Glycoconj J*, (2012).

39. A. Huttner *et al.*, **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* 17, 528-537 (2017).

40. M. S. Riddle *et al.*, **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* 23, 908-917 (2016).

41. C. F. Hatz *et al.*, **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 (2015).

42. J. A. Iwashkiw *et al.*, **Exploiting the Campylobacter jejuni protein glycosylation system for glycoengineering vaccines and diagnostic tools directed against brucellosis.** *Microb Cell Fact* 11, 13 (2012).

43. W. Shang *et al.*, **Production of human blood group B antigen epitope conjugated protein in Escherichia coli and utilization of the adsorption blood group B antibody.** *Microb Cell Fact* 15, 138 (2016).

44. A. E. Ciocchini *et al.*, **Development and validation of a novel diagnostic test for human Brucellosis using a glyco-engineered antigen coupled to magnetic beads.** *PLoS Negl Trop Dis* 7, e2048 (2013).

45. I. Hug *et al.*, **Exploiting bacterial glycosylation machineries for the synthesis of a Lewis antigen-containing glycoprotein.** *J Biol Chem* 286, 37887-37894 (2011).

46. T. Song, D. Aldredge, C. B. Lebrilla, **A method for in-depth structural annotation of human serum glycans that yields biological variations.** *Anal Chem* 87, 7754-7762 (2015).

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

48. B. Strutton, S. R. P. Jaffe, J. Pandhal, P. C. Wright, **Producing a glycosylating Escherichia coli cell factory: The placement of the bacterial oligosaccharyl transferase pglB onto the genome.** *Biochem Biophys Res Commun* 495, 686-692 (2018).

49. L. E. Yates, A. Natarajan, M. P. DeLisa, **Glyco-recoding Escherichia coli by recombineering-based genome editing of native polysaccharide biosynthesis gene clusters (submitted).** (2018).

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

51. A. Faridmoayer *et al.*, **Extreme substrate promiscuity of the Neisseria oligosaccharyl transferase involved in protein O-glycosylation.** *J Biol Chem* 283, 34596-34604 (2008).

52. P. Sun *et al.*, **Design and production of conjugate vaccines against S. Paratyphi A using an O-linked glycosylation system in vivo.** *NPJ Vaccines* 3, 4 (2018).

53. C. Pan *et al.*, **Biosynthesis of conjugate vaccines using an O-linked glycosylation system.** *MBio* 7, e00443-00416 (2016).

54. M. J. Coyne *et al.*, **Phylum-wide general protein O-glycosylation system of the Bacteroidetes.** *Mol Microbiol* 88, 772-783 (2013).

55. J. Cuccui *et al.*, **The N-linking glycosylation system from Actinobacillus pleuropneumoniae is required for adhesion and has potential use in glycoengineering.** *Open Biol* 7, (2017).

56. T. G. Keys *et al.*, **A biosynthetic route for polysialylating proteins in Escherichia coli.** *Metab Eng* 44, 293-301 (2017).

57. W. W. Shi *et al.*, **Structure of a novel O-linked N-acetyl-D-glucosamine (O-GlcNAc) transferase, GtfA, reveals insights into the glycosylation of pneumococcal serine-rich repeat adhesins.** *J Biol Chem* 289, 20898-20907 (2014).

58. G. E. Henderson, K. D. Isett, T. U. Gerngross, **Site-specific modification of recombinant proteins: a novel platform for modifying glycoproteins expressed in E. coli.** *Bioconjug Chem* 22, 903-912 (2011).

59. J. Ihssen *et al.*, **Production of glycoprotein vaccines in Escherichia coli.** *Microb Cell Fact* 9, 61 (2010).

60. M. M. Kampf *et al.*, **In vivo production of a novel glycoconjugate vaccine against Shigella flexneri 2a in recombinant Escherichia coli: identification of stimulating factors for in vivo glycosylation.** *Microb Cell Fact* 14, 12 (2015).

61. N. Ding *et al.*, **Increased glycosylation efficiency of recombinant proteins in Escherichia coli by auto-induction.** *Biochem Biophys Res Commun* 485, 138-143 (2017).

62. F. W. Studier, **Stable expression clones and auto-induction for protein production in E. coli.** *Methods Mol Biol* 1091, 17-32 (2014).

63. K. Champasa, S. A. Longwell, A. M. Eldridge, E. A. Stemmler, D. H. Dube, **Targeted identification of glycosylated proteins in the gastric pathogen Helicobacter pylori (Hp).** *Mol Cell Proteomics* 12, 2568-2586 (2013).

64. N. Geva-Zatorsky *et al.*, **In vivo imaging and tracking of host-microbiota interactions via metabolic labeling of gut anaerobic bacteria.** *Nat Med* 21, 1091-1100 (2015).

65. J. Pandhal *et al.*, **Systematic metabolic engineering for improvement of glycosylation efficiency in Escherichia coli.** *Biochem Biophys Res Commun* 419, 472-476 (2012).

66. C. J. Glasscock *et al.*, **A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation.** *Metab Eng* 47, 488-495 (2018).

67. J. A. Wayman, C. Glasscock, T. J. Mansell, M. P. DeLisa, J. D. Varner, **Improving designer glycan production in Escherichia coli through model-guided metabolic engineering.** *Metab Eng Commun*, (in revision) doi: https://doi.org/10.1101/160853 (2017).

68. J. Pandhal, S. Y. Ow, J. Noirel, P. C. Wright, **Improving N-glycosylation efficiency in Escherichia coli using shotgun proteomics, metabolic network analysis, and selective reaction monitoring.** *Biotechnol Bioeng* 108, 902-912 (2011).

69. J. Pandhal *et al.*, **Inverse metabolic engineering to improve Escherichia coli as an N-glycosylation host.** *Biotechnol Bioeng* 110, 2482-2493 (2013).

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

71. A. A. Ollis *et al.*, **Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences.** *Sci Rep* 5, 15237 (2015).

72. J. Ihssen *et al.*, **Increased efficiency of Campylobacter jejuni N-oligosaccharyltransferase PglB by structure-guided engineering.** *Open Biol* 5, 140227 (2015).

73. A. Aharoni *et al.*, **High-throughput screening methodology for the directed evolution of glycosyltransferases.** *Nat Methods* 3, 609-614 (2006).

74. T. G. Keys *et al.*, **Engineering the product profile of a polysialyltransferase.** *Nat Chem Biol* 10, 437-442 (2014).

75. G. Yang *et al.*, **Fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases.** *J Am Chem Soc* 132, 10570-10577 (2010).

76. J. Ihssen *et al.*, **Structural insights from random mutagenesis of Campylobacter jejuni oligosaccharyltransferase PglB.** *BMC Biotechnol* 12, 67 (2012).

77. C. Durr, H. Nothaft, C. Lizak, R. Glockshuber, M. Aebi, **The Escherichia coli glycophage display system.** *Glycobiology* 20, 1366-1372 (2010).

78. E. Celik, A. C. Fisher, C. Guarino, T. J. Mansell, M. P. DeLisa, **A filamentous phage display system for N-linked glycoproteins.** *Protein Sci* 19, 2006–2013 (2010).

79. D. Shental-Bechor, Y. Levy, **Folding of glycoproteins: toward understanding the biophysics of the glycosylation code.** *Curr Opin Struct Biol* 19, 524-533 (2009).

80. N. Mitra, S. Sinha, T. N. Ramya, A. Surolia, **N-linked oligosaccharides as outfitters for glycoprotein folding, form and function.** *Trends Biochem Sci* 31, 156-163 (2006).

81. T. J. Mansell, C. Guarino, M. P. Delisa, **Engineered genetic selection links in vivo protein folding and stability with asparagine-linked glycosylation.** *Biotechnol J* 8, 1445-1451 (2013).

82. A. C. Fisher, W. Kim, M. P. DeLisa, **Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway.** *Protein Sci* 15, 449-458 (2006).

83. J. T. Boock *et al.*, **Repurposing a bacterial quality control mechanism to enhance enzyme production in living cells.** *J Mol Biol* 427, 1451-1463 (2015).

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

85. T. Jaroentomeechai *et al.*, **A pipeline for studying and engineering single-subunit oligosaccharyltransferases.** *Methods Enzymol* 597, 55-81 (2017).

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

87. C. Perez *et al.*, **Structure and mechanism of an active lipid-linked oligosaccharide flippase.** *Nature* 524, 433-438 (2015).

88. S. Gerber *et al.*, **Mechanism of bacterial oligosaccharyltransferase: in vitro quantification of sequon binding and catalysis.** *J Biol Chem* 288, 8849-8861 (2013).

89. C. Lizak *et al.*, **Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation.** *Nat Commun* 4, 2627 (2013).

90. F. Liu *et al.*, **Rationally designed short polyisoprenol-linked PglB substrates for engineered polypeptide and protein N-glycosylation.** *J Am Chem Soc* 136, 566-569 (2014).

91. M. A. Musumeci, M. V. Ielmini, M. F. Feldman, **In vitro glycosylation assay for bacterial oligosaccharyltransferases.** *Methods Mol Biol* 1022, 161-171 (2013).

92. Q. Song *et al.*, **Production of homogeneous glycoprotein with multisite modifications by an engineered N-glycosyltransferase mutant.** *J Biol Chem* 292, 8856-8863 (2017).

93. W. Kightlinger *et al.*, **Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases.** *Nat Chem Biol*, (2018).

94. A. S. Ramirez *et al.*, **Characterization of the single-subunit oligosaccharyltransferase STT3A from Trypanosoma brucei using synthetic peptides and lipid-linked oligosaccharide analogs.** *Glycobiology* 27, 525-535 (2017).

95. S. Matsumoto *et al.*, **Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation.** *Proc Natl Acad Sci U S A* 110, 17868-17873 (2013).

96. R. Wild *et al.*, **Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation.** *Science* 359, 545-550 (2018).

97. M. D. Hartley, P. E. Schneggenburger, B. Imperiali, **Lipid bilayer nanodisc platform for investigating polyprenol-dependent enzyme interactions and activities.** *Proc Natl Acad Sci U S A* 110, 20863-20870 (2013).

98. A. S. Ramirez *et al.*, **Structural basis of the molecular ruler mechanism of a bacterial glycosyltransferase.** *Nat Commun* 9, 445 (2018).

99. K. J. Glover, E. Weerapana, B. Imperiali, **In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation.** *Proc Natl Acad Sci U S A* 102, 14255-14259 (2005).

100. A. S. Ramirez *et al.*, **Chemo-enzymatic synthesis of lipid-linked GlcNAc2Man5 oligosaccharides using recombinant Alg1, Alg2 and Alg11 proteins.** *Glycobiology*, 1-8 (2017).

101. T. F. T. Rexer *et al.*, **One pot synthesis of GDP-mannose by a multi-enzyme cascade for enzymatic assembly of lipid-linked oligosaccharides.** *Biotechnol Bioeng* 115, 192-205 (2018).

102. F. Schwarz *et al.*, **A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation.** *Nat Chem Biol* 6, 264-266 (2010).

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

104. J. E. Kay, M. C. Jewett, **Lysate of engineered Escherichia coli supports high-level conversion of glucose to 2,3-butanediol.** *Metab Eng* 32, 133-142 (2015).

105. J. F. Zawada *et al.*, **Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines.** *Biotechnol Bioeng* 108, 1570-1578 (2011).

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

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

108. T. Jaroentomeechai *et al.*, **Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery.** *Nat Commun* 9, 2686 (2018).

109. J. A. Schoborg *et al.*, **A cell-free platform for rapid synthesis and testing of active oligosaccharyltransferases.** *Biotechnol Bioeng* 115, 739-750 (2018).

110. M. Vilkhovoy *et al.*, **Sequence specific modeling of E. coli cell-free protein synthesis.** *ACS Synth Biol*, (in review) doi: https://doi.org/10.1101/139774 (2017).

111. N. Horvath *et al.*, **Toward a genome scale sequence specific dynamic model of cell-free protein synthesis in Escherichia coli.** *ACS Synth Biol*, (in review) doi: https://doi.org/10.1101/215012 (2017).

112. E. Karzbrun, J. Shin, R. H. Bar-Ziv, V. Noireaux, **Coarse-grained dynamics of protein synthesis in a cell-free system.** *Phys Rev Lett* 106, 048104 (2011).

113. F. Caschera *et al.*, **Coping with complexity: machine learning optimization of cell-free protein synthesis.** *Biotechnol Bioeng* 108, 2218-2228 (2011).

CHAPTER 2

Glyco-recoded Escherichia coli: Recombineering-based genome editing of native polysaccharide biosynthesis gene clusters[1]

2.1 Abstract

Recombineering-based redesign of bacterial genomes by adding, removing or editing large segments of genomic DNA is emerging as a powerful technique for expanding the range of functions that an organism can perform. Here, we describe a glyco-recoding strategy whereby major non-essential polysaccharide gene clusters in K-12 *Escherichia coli* are replaced with orthogonal glycosylation components for both biosynthesis of heterologous glycan structures and site-specific glycan conjugation to target proteins. Specifically, the native enterobacterial common antigen (ECA) and O-polysaccharide (O-PS) antigen loci were systematically replaced with ~9–10 kbp of synthetic DNA encoding *Campylobacter jejuni* enzymes required for asparagine-linked (*N*-linked) protein glycosylation. Compared to *E. coli* cells carrying the same glycosylation machinery on extrachromosomal plasmids, glyco-recoded strains attached glycans to acceptor protein targets with equal or greater efficiency while exhibiting markedly better growth phenotypes and higher glycoprotein titers. Overall, our results define a convenient and reliable framework for bacterial glycome editing that provides a more stable route for chemical diversification of proteins *in vivo* and effectively expands the bacterial glycoengineering toolkit.

---

[1] *This chapter has been previously published:* Yates L.E.*, Natarajan A.*, Li M.*, Hale M., Mills D.C., and DeLisa M.P. (2019). Glyco-recoded *Escherichia coli*: Recombineering-based genome editing of native polysaccharide biosynthesis gene clusters. Metabolic Engineering, 53, 59-68. DOI: 10.1016/j.ymben.2019.02.002.
URL: https://www.sciencedirect.com/science/article/pii/S1096717618304476?via%3Dihub
* These authors contributed equally to this work. Natarajan A. contributed to data in Figures 2-4, 2-8, 2-9, and Tables 2-1, 2-2.

2.2 Introduction

Protein glycosylation is one of the most common post-translational protein modifications and is present in all domains of life including bacteria(*2*) and archaea(*3*). In bacteria, processes for covalently attaching complex carbohydrates known as glycans to both asparagine residues via *N*-linked glycosylation and serine or threonine residues *via O*-linked glycosylation are now firmly established and reviewed extensively(*4-7*). Shortly after the discovery of *N*-glycosylation in the bacterial pathogen *Campylobacter jejuni*(*8*), Aebi and coworkers described the functional transfer of the entire *C. jejuni* protein glycosylation (*pgl*) locus into a laboratory strain of *Escherichia coli*(*9*). The resulting glycoprotein expression platform involved the assembly of the *C. jejuni* heptasaccharide glycan on the common bacterial lipid carrier, undecaprenyl pyrophosphate (Und-PP), followed by *en bloc* transfer of the preassembled glycan by the *C. jejuni* oligosaccharyltransferase (OST) PglB to an asparagine residue of a *C. jejuni* acceptor protein in the periplasm.

Since its inception, this reconstituted system has been used to understand the bacterial *N*-glycosylation mechanism and also its topologically and functionally similar counterpart in eukaryotes(*7*). For example, *E. coli* cells expressing the heterologous *C. jejuni* pathway have been used to demonstrate that *C. jejuni* PglB has a strict preference for the consensus sequence (sequon) D/E-X_1-N-X_2-S/T (where X_1, $X_2 \neq$ Pro), which is more specific than the N-X-S/T (where X ≠ Pro) sequon recognized by eukaryotic OSTs(*10*). By swapping *C. jejuni* PglB with orthologous PglB enzymes from different Gram-negative bacteria, it was determined that some bacterial OSTs exhibit more relaxed sequon specificity(*11*). Likewise, the ability to interchange different glycan biosynthesis pathways was used to demonstrate that *C. jejuni* PglB has relaxed specificity towards the glycan substrate(*12*).

Importantly, these and other mechanistic studies have paved the way for making customized glycoproteins in *E. coli*, whereby different proteins of interest can be site-specifically modified with desired glycan structures simply by co-expressing an OST, a pathway for assembling sugars of interest on Und-PP, and an acceptor protein target bearing one or more glycosylation sequons (reviewed in(*13, 14*)). For example, the production of bioconjugate vaccine candidates is made possible by leveraging PglB for the conjugation of O-polysaccharide (O-PS) or capsular polysaccharide (CPS) antigens to vaccine-relevant carrier proteins in *E. coli*(*15-19*). Along similar lines, we described the rational design and assembly of an orthogonal glycosylation pathway that enabled biosynthesis of the eukaryotic trimannosyl core glycan and its transfer to a panel of diverse acceptor proteins(*20*). For some of these proteins, glycan attachment was enabled by a DQNAT glycosylation tag (GlycTag), a preferred sequon for *C. jejuni* PglB(*21*) that can easily be grafted into regions of flexible secondary structure or appended at the N- or C-termini of a protein, leading to efficient glycosylation by PglB(*10, 22*).

In the majority of the aforementioned studies, glycosylation was achieved using a two- or even three-plasmid system to transfer all of the requisite machinery along with the acceptor protein into *E. coli*. This is in line with conventional metabolic engineering approaches where multi-gene biochemical pathways are often introduced into a host cell using multiple plasmids, which can autonomously replicate, attain high copy numbers, and allow strong gene expression. Unfortunately, such multi-plasmid approaches often impose a deleterious metabolic burden on the host that manifests as a decrease in fitness and a concomitant decline in productivity arising from factors such as plasmid loss and upregulation of proteolytic activity(*23-25*). Wright and coworkers recently described a simplified glycosylation platform where the gene encoding *C. jejuni* PglB was

integrated into the genome; however, the glycosylation efficiency achieved in this platform was only ~40% overall and nearly all of this corresponded to monoglycosylated product despite the presence of two glycosylation sites in the acceptor protein(*26*). Moreover, no apparent fitness advantage was gained because: (i) significant cellular biosynthetic capacity was diverted to expressing *pglB* from a strong constitutive promoter; and (ii) the *pglB* integrant strain still required the addition of two plasmids including one that encoded all of the glycan biosynthesis enzymes.

Here, we sought to engineer a robust glycoprotein engineering host strain that was capable of high-level, genetically stable expression of all heterologous glycosylation enzymes through de novo chromosomal engineering, thereby eliminating the need for extraneous plasmid-based expression of any pathway enzymes. To this end, we propose a glycomic recoding (glyco-recoding) strategy whereby native polysaccharide gene clusters in the *E. coli* genome are reassigned to create an alternative glycan code. Specifically, a targeted genome editing strategy was developed for site-specific insertion of PglB and the requisite glycosylation pathway enzymes in place of two native gene clusters encoding enzymes that produce the O-PS antigen and enterobacterial common antigen (ECA), both of which are unnecessary for cell viability or *N*-linked protein glycosylation. By recombineering the complete *N*-glycosylation machinery into these two genomic loci, our host strain only required transformation with a single plasmid encoding the acceptor protein. As a result, cells appeared to experience a significantly reduced metabolic burden, growing to significantly higher final densities and achieving much greater product titers. Moreover, glycosylation efficiency in these strains rivaled or exceeded the level of glycosylation that could be obtained from corresponding plasmid-based systems.

2.3 Results

2.3.1. Recombinase-mediated genomic replacement of native oligosaccharide gene clusters

To enable stable integration of glycosylation-relevant genes into targeted locations in the *E. coli* K-12 chromosome, we first developed two new shuttle plasmids based on pKD4(*27*). These plasmids were designed for homologous recombination-based insertion of genes-of-interest (GOIs) in place of the existing genomic loci encoding the ECA and O-PS biosynthetic pathways, respectively ***(Figure 2-1)***. These pathways were chosen for several reasons. First, neither pathway is essential for cell viability and both encode enzymes that are unnecessary for most glycoengineering efforts. In fact, in the case of the O-PS gene cluster, the *wbbL* gene is inactivated by an IS5 element that renders the locus functionally incomplete ***(Figure 2-1)***(*28*). Second, some of the pathway enzymes may deplete the pool of important substrates or otherwise divert intermediates to unwanted end products. For example, the ECA locus enzymes WecB and WecG directly compete for the UDP-N-acetylglucosamine (UDPGlcNAc) building block and GlcNAc-loaded undecaprenyl pyrophosphate (Und-PP) intermediate, respectively, which are key precursors in our engineered eukaryotic glycosylation pathway(*20*). Finally, both regions include an operon polarity suppressor, meaning they are well-suited to the expression of large clusters of genes.

To replace the O-PS gene cluster, plasmid pRecO-PS was created by cloning the flanking genes *glf* and *gnd* into pKD4 with an intervening multiple cloning site (MCS) for insertion of GOIs and a FLP recognition target (FRT)-flanked kanamycin resistance (Kan^R) marker for facile selection of integrants ***(Figure 2-1)***. The Kan^R cassette can then be removed using the FRT sites for marker-less strain construction. We targeted *glf* and *gnd* because all serotype-specific structural genes of the O-PS are located

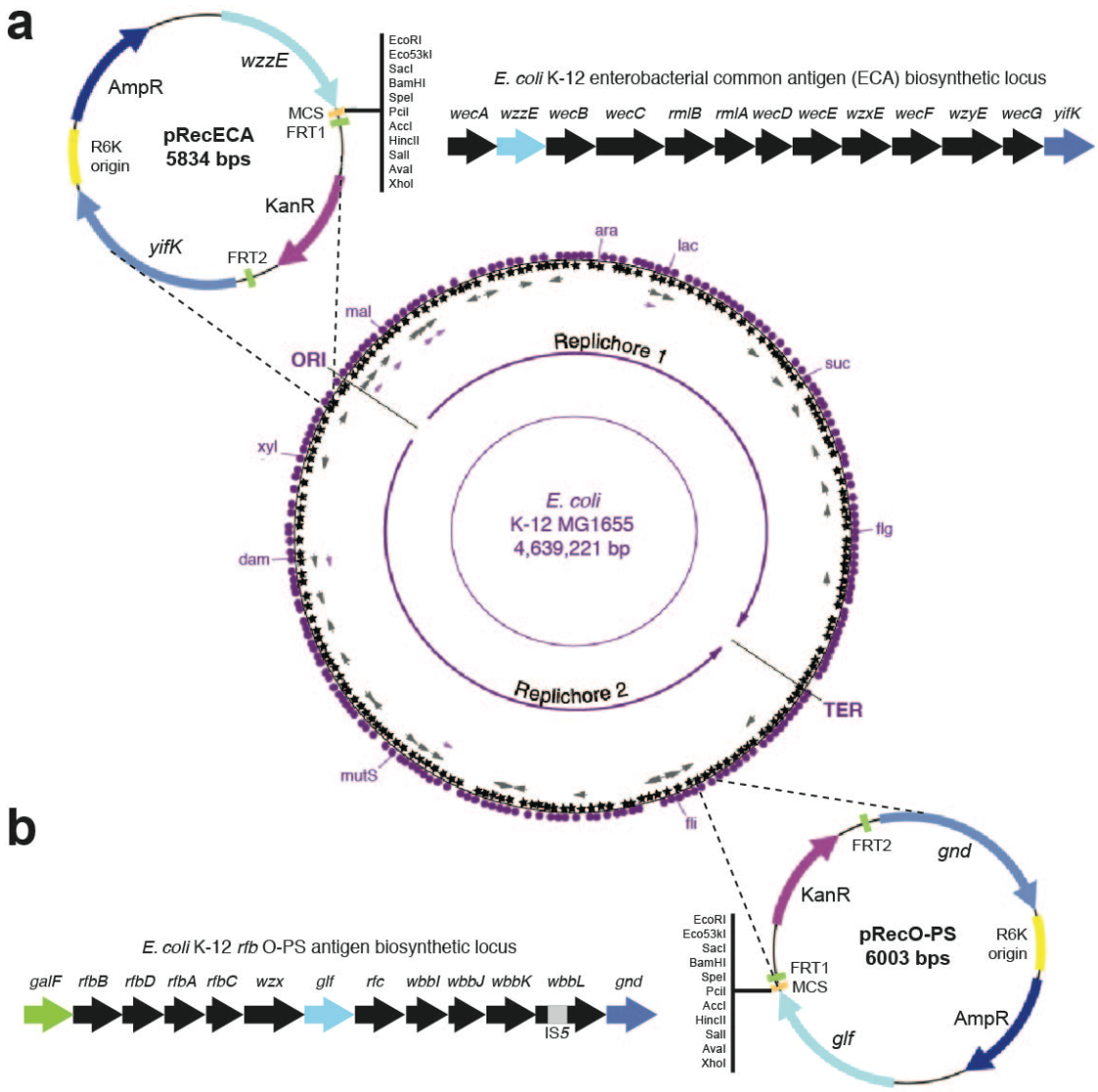


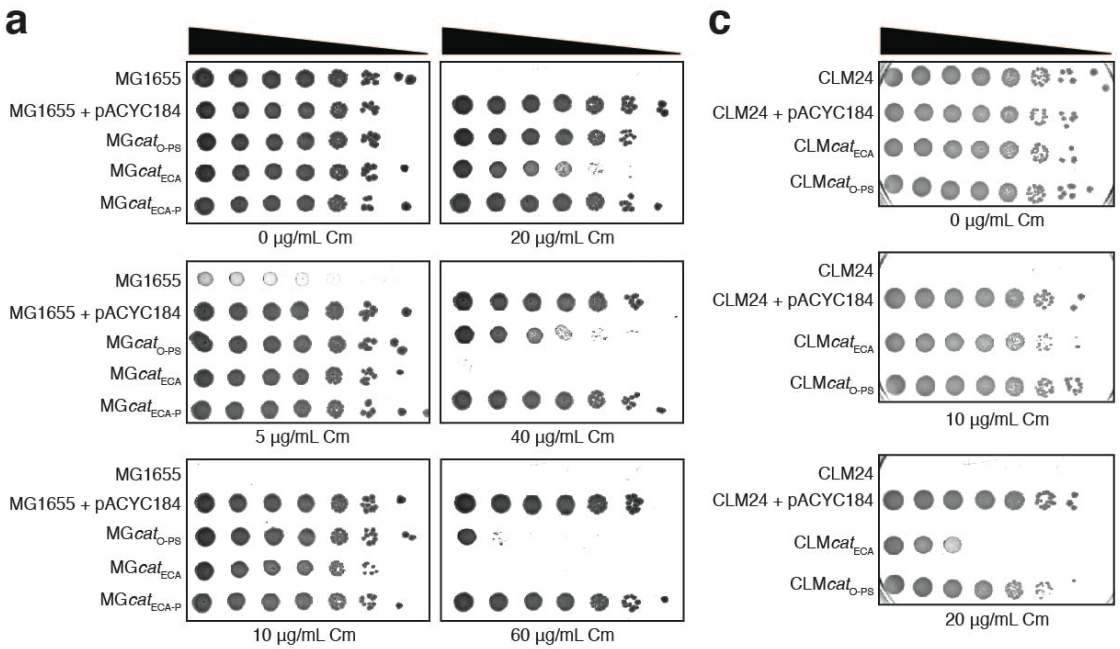
Figure 2-1: Glyco-recoding the genome of E. coli K-12.

(a) Organization of the enterobacterial common antigen (ECA) biosynthetic locus from E. coli K-12 strain MG1655 and the corresponding plasmid pRecECA for genomic replacement of the ECA cluster with glyco-genes/pathways of interest between the genes wzzE and yifK. (b) Same as (a) but instead showing organization of the rfb O-polysaccharide (O-PS) antigen biosynthetic locus from E. coli K-12 strain MG1655, including the IS5 inactivation of wbbL (i.e., the rfb-50 mutation), and the corresponding plasmid pRecO-PS for genomic replacement of the rfb cluster between glf and gnd. Circular map of the E. coli MG1655 chromosome showing positional context of the ECA and O-PS recombineering sites, which are located between ilv and cya near min 85 on the E. coli chromosome for ECA and between wca and his near min 45 on the E. coli chromosome for O-PS. ORI and TER, origin and terminus of replication, respectively.

between these genes(*29*). The intervening genes were judged as having the potential to

interfere with heterologously expressed pathways, whereas the O-PS flippase *wzx*, which is located immediately outside of this region, was deliberately retained to provide a membrane translocation function for pathways lacking a native flippase enzyme. To determine whether the O-PS locus was a viable integration site (i.e., whether the expression of integrated genes would be driven by the promoter in front of *galF*, the first gene of the operon), we cloned the gene encoding chloramphenicol acetyltransferase (CAT) into the MCS of pRecO-PS, yielding pRecO-PS_{cat}. Next, *E. coli* K-12 strain MG1655 that was initially transformed with a helper plasmid expressing λ-red recombinase was subjected to a second transformation to introduce a linear DNA fragment amplified from plasmid pRecO-PS_{cat} containing the cloned cat insert flanked by *glf-gnd*. Following selection on LB agar plates supplemented with Kan, a large number of colonies were recovered indicating a high integration efficiency. Randomly chosen Kan-resistant colonies were then counter selected on chloramphenicol (Cm)-containing plates and all were resistant to chloramphenicol. Because the pRecO-PS_{cat} plasmid itself does not confer Cm resistance, this is a good indication that (i) integration was successful and (ii) the chromosomal promoter immediately upstream of the insertion site is capable of driving transcription of inserted genes. To benchmark the level of CAT expression, we performed spot plating analysis where serial dilutions of a representative positive clone, strain MG$cat_{O\text{-}PS}$ for which the Kan^R cassette had been removed, was subjected to increasing concentrations of Cm. Whereas plasmid-free MG1655 cells were susceptible to all Cm concentrations tested, strain MG$cat_{O\text{-}PS}$ was resistant up to 60 μg/ml Cm, which compared favorably to the resistance exhibited by MG1655 carrying the chloramphenicol-resistant pACYC184 plasmid, which is maintained at ~15 copies per cell ***(Figure 2-2)***.

To replace the ECA locus, a nearly identical procedure was performed by first creating plasmid pRecECA, which utilized *wzzE* and *yifK* as the flanking genes ***(Figure 2-1).*** We chose *wzzE* as the upstream homology sequence because the first gene in the ECA pathway, *wecA*, encodes a UDP-GlcNAc:undecaprenyl-phosphate GlcNAc-1-



b

identifier	*sequence*	*measured strength*
BBa_J23100	ttgacggctagctcagtcctaggtacagtgctagc	1.00
BBa_J23101	tttacagctagctcagtcctaggtattatgctagc	0.70
BBa_J23110	tttacggctagctcagtcctaggtacaatgctagc	0.33

Figure 2-2: Characterization of gene expression from O-PS and EPA sites.
(a) Representative spot titer images for serially diluted E. coli MG1655 cells with: no plasmid, empty plasmid pACYC184, the cat gene integrated at O-PS site (MGcatO-PS), the cat gene integrated at ECA site (MGcatECA), and the cat gene integrated at ECA site with enhanced promoter (MGcatECA-P) as indicated. Resistance of cells was evaluated by serially diluting overnight cells in liquid LB and then plating 5 µl of diluted cells on LB agar in the presence (5-60 µg/ml) or absence (0 µg/ml) of Cm. Black triangle represents relative amount of cells plated. (b) Selected promoter sequences that were evaluated for their ability to enhance expression from the EPA site. Red letters indicate bases that were changed in the parental promoter, BBa_J23100, leading to decreased promoter strength as indicated at right. BBa_J23100 is a constitutive, intermediate-strength promoter from the Anderson promoter library (1). (c) Same as in (a) but with CLM24 as host strain.

phosphate transferase that catalyzes the formation of Und-PP-GlcNAc, which is a key intermediate in certain designer glycan structures(*20, 30*). While integration was similarly efficient based on Kan selection, the Cm resistance of a representative positive clone, strain MG*cat*$_{ECA}$, was significantly weaker than that observed for MG*cat*$_{O\text{-}PS}$ ***(Figure 2-2 a)***. The lower expression from the ECA site compared to the OP-S site was in line with previous reports that gene expression from the genome is dependent on the location of the genes(*31, 32*). We hypothesized that this weak resistance was due to poor expression of the cat gene from the native ECA promoter, so we modified pRecECA with a synthetic promoter to drive expression of cloned genes. Specifically, the constitutive, intermediate-strength BBa_J23110 promoter from the Anderson promoter library(*33*) ***(Figure 2-2 b)*** was introduced upstream of the MCS, yielding the new shuttle plasmid pRecECA-P. A derivative of this plasmid carrying the cloned cat insert was used to create strain MG*cat*$_{ECA\text{-}P}$, which exhibited resistance to Cm that resembled that of the control strains carrying the multi-copy pACYC184 plasmid ***(Figure 2-2 a)***. It is also worth mentioning that the shuttle plasmids were compatible with genomic integration of the cat gene to both loci in a different K-12 strain background, *E. coli* CLM24 ***(Figure 2-2 c)***, which is a commonly used host for bacterial glycoengineering studies(*12, 22, 26, 34*).

2.3.2. Protein glycosylation from single-copy expression of PglB

As a first step towards bacterial glycome editing, we focused on replacing either the O-PS or ECA gene cluster in MG1655 with the gene encoding *C. jejuni* PglB. This involved creating shuttle vectors pRecOPS*pglB*, pRecECA*pglB*, and pRecECA-P*pglB*, which were subsequently used to generate the strains MG*pglB*$_{O\text{-}PS}$, MG*pglB*$_{ECA}$, and MG*pglB*$_{ECA\text{-}P}$, respectively. For glycan expression, we generated plasmid pACYC*pgl9* that encodes a minimal 9-kb *pgl* pathway for making a modified *C. jejuni*

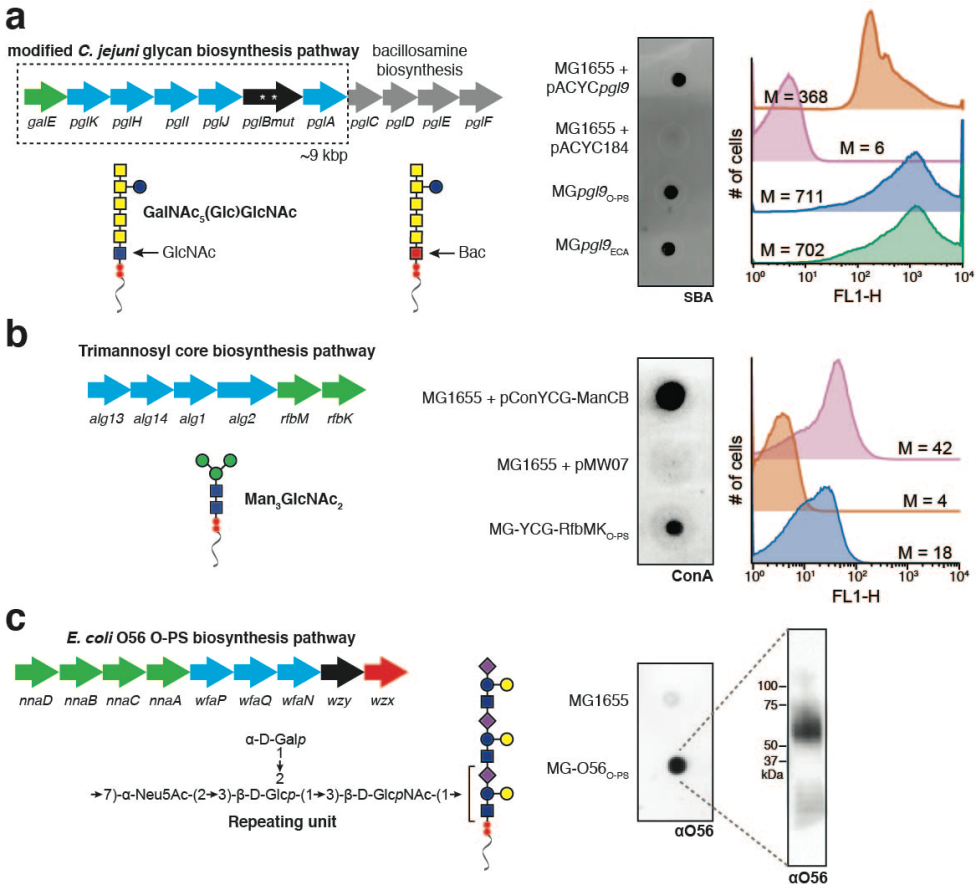


Figure 2-3: Characterization of genomically-integrated glycan biosynthesis pathways.

(a) Schematic of pgl9 pathway for making modified C. jejuni heptasaccharide glycan, GalNAc5(Glc)GlcNAc, on Und-PP. The pathway includes a mutant copy of pglB that is functionally inactive as indicated by white asterisks. Genes for bacillosamine biosynthesis (grey arrows) were excluded from the pathway, resulting in a glycan with GlcNAc as the reducing end sugar. Dot blot and flow cytometric analysis of intact cells using SBA lectin that recognizes terminal GalNAc residues. Stained cells included: MG1655 carrying pACYCpgl9; MG1655 carrying empty pACYC184; MGpgl9O-PS; and MGpgl9ECA as indicated. (b) Schematic of trimannosyl core ($Man_3GlcNAc_2$) biosynthesis pathway. Dot blot and flow cytometric analysis of intact cells probed with ConA lectin that recognizes sub-terminal and terminal Man residues. Stained cells included: MG1655 carrying pConYCG-ManCB; MG1655 carrying empty pMW07, and MG-YCG-RfbMKO-PS as indicated. Median fluorescence (M) values for histograms of SBA- and ConA-stained cells are provided for each. (c) Schematic of E. coli O56 O-PS biosynthesis pathway. Dot blot of intact cells and Western blot analysis of crude LPS preparation, both probed with serum against O56 OPS. Cells analyzed included: plasmid-free MG1655 and MG-O56O-PS as indicated.

heptasaccharide glycan, $GalNAc_5(Glc)GlcNAc$, on Und-PP(*30*) and includes a mutant

copy of *pglB* that is functionally inactive ***(Figure 2-3 a)***. When MG1655 cells were transformed with this plasmid, they were readily stained with the lectin soybean agglutinin (SBA) that recognizes terminal N-acetylgalactosamine (GalNAc) residues ***(Figure 2-3 a)***. The positive staining arises from the cell surface expression of the glycan, which is a consequence of the well-documented mechanism for transfer of Und-PP-linked glycans to lipid A core by the WaaL ligase and subsequent shuttling of the hybrid glycolipids to the outer membrane(*20, 35*). Next, the MG*pglB*$_{\text{O-PS}}$, MG*pglB*$_{\text{ECA}}$, and MG*pglB*$_{\text{ECA-P}}$ strains were co-transformed with this plasmid along with a plasmid for expressing the model acceptor protein spDsbA-scFv13-R4$^{\text{DQNAT}}$, which is comprised of an anti-β-galactosidase single-chain Fv antibody (scFv13-R4) modified N-terminally with an SRP-dependent signal peptide derived from *E. coli* DsbA (spDsbA) and C-terminally with an optimal DQNAT acceptor sequon(*20*). A C-terminal 6x-His tag was also introduced just after the DQNAT motif for immunodetection and purification. Following expression in these strains, the glycosylation status of scFv13-R4$^{\text{DQNAT}}$ was analyzed by immunoblotting with an anti-polyhistidine (anti-His) antibody or hR6 serum that is specific for the *C. jejuni* heptasaccharide glycan(*36*). Importantly, all three *pglB*-recoded strains were capable of glycosylating scFv13-R4$^{\text{DQNAT}}$ ***(Figure 2-4)***. The most efficient glycosylation (90%) occurred in MG*pglB*$_{\text{O-PS}}$ cells, which exceeded the glycosylation efficiency (61%) observed in non-recoded MG1655 that carried three plasmids, one for glycan biosynthesis, one for acceptor protein expression, and one for OST expression ***(Figure 2-4)***. Consistent with the cat integrants above, the least efficient glycosylation (60%) was observed for MG*pglB*$_{\text{ECA}}$ cells expressing PglB from the ECA locus; however, greatly increased glycosylation efficiency (82%) was achieved by introducing the BBa_J23110 promoter to drive *pglB* expression from the ECA region ***(Figure 2-4)***.

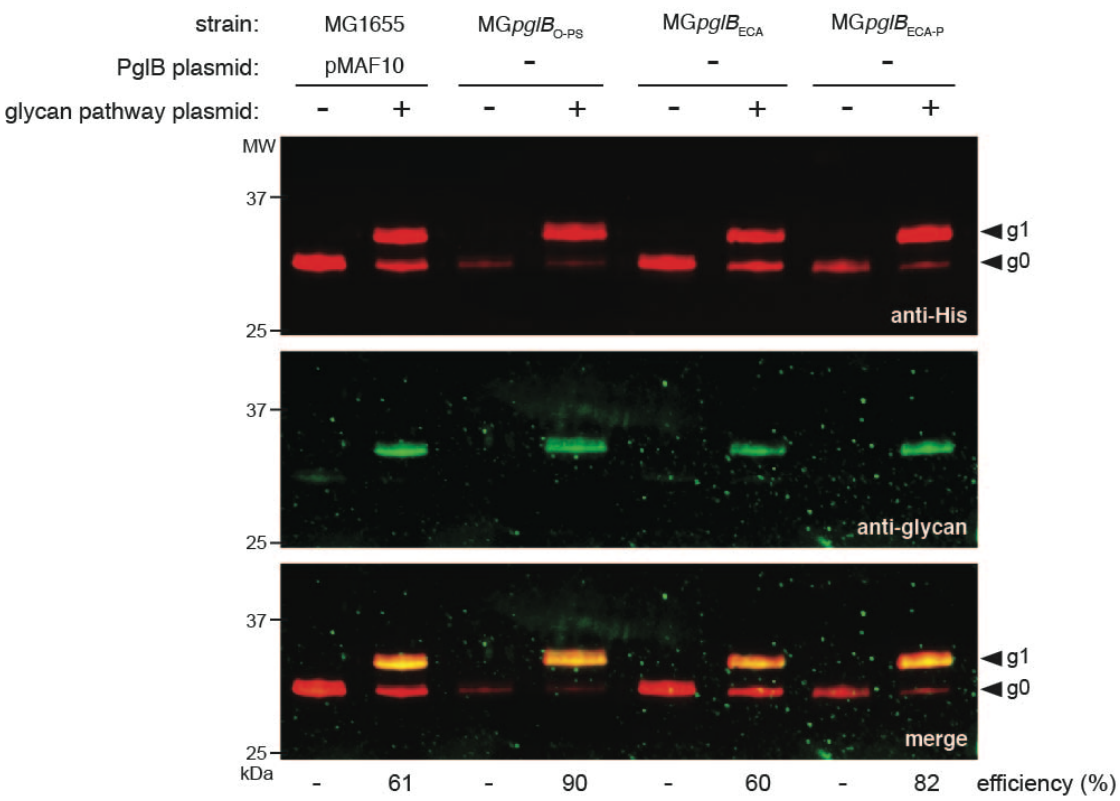


Figure 2-4: Site-specific genomic integration of functional PglB.
Western blot analysis of scFv13-R4^{DQNAT} acceptor protein in periplasmic fractions derived from: (lanes 1 and 2) E. coli MG1655 carrying pBS-spDsbA-scFv13-R4^{DQNAT}, pMAF10, and either pACYCpgl9 (+) or empty pACYC184 (−); and (lanes 3–8) E. coli MGpglB$_{O\text{-}PS}$, MGpglB$_{ECA}$, or MGpglB$_{ECA\text{-}P}$ cells carrying pBS-spDsbA-scFv13-R4^{DQNAT} along with either pACYCpgl9 (+) or empty pACYC184 (−). 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. Bottom panel depicts overlay of anti-His and hR6 blots (merge). Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of scFv13-R4^{DQNAT}. Molecular weight (MW) markers are indicated on the left. Results are representative of at least three biological replicates.

2.3.3. Heterologous glycans expressed from the genome and transferred to proteins

To glyco-recode the genome with orthogonal pathways for glycan biosynthesis, we next attempted to integrate the minimal *C. jejuni pgl* pathway into the chromosome of MG1655. To this end, the strains MG*pgl9*$_{O\text{-}PS}$ and MG*pgl9*$_{ECA}$ were generated in

which the O-PS and ECA loci, respectively, were each replaced with the 9-kb *pgl9* gene cluster for glycan biosynthesis. SBA lectin blotting and flow cytometric analysis of intact MG*pgl9*$_{\text{O-PS}}$ and MG*pgl9*$_{\text{ECA}}$ cells confirmed that each produced cell-surface GalNAc-terminal glycans at a level that was unexpectedly higher than that produced by MG1655 cells carrying pACYC*pgl9* as evidenced by dot blot analysis and flow cytometry ***(Figure 2-3 a)***. Next, each *pgl9* pathway-recoded strain was transformed with one plasmid encoding PglB and another encoding spDsbA-scFv13- R4$^{\text{DQNAT}}$, and glycosylation of the acceptor protein was analyzed by immunoblotting as described above. In both cases, the acceptor protein was glycosylated at a level (52–62%) that was nearly identical to glycosylation in the non-recoded MG1655 strain carrying three plasmids for glycan biosynthesis, OST, and acceptor protein ***(Figure 2-5)***.

To demonstrate the ease with which alternative glycan biosynthesis pathways can be inserted into the genome, we cloned a heterologous pathway for making eukaryotic $Man_3GlcNAc_2$ glycans(*20*) into pRecO-PS. The resulting pRecO-PS-YCGRfbMK plasmid encoded four yeast enzymes, Alg13, Alg14, Alg1, and Alg2, for assembling the $Man_3GlcNAc_2$ glycan on Und-PP and two enzymes, RfbM and RfbK, involved in the formation of GDP-mannose in *E. coli* O9:K30(*37*). These latter enzymes are isozymes of GDP-mannose pyrophosphorylase ManC and phosphomannomutase ManB, respectively, from *E. coli* K-12 that were used recently to enhance $Man_3GlcNAc_2$ levels(*35*). Importantly, the recoded strain MG-YCG-RfbMK$_{\text{O-PS}}$ was observed to produce $Man_3GlcNAc_2$ at comparable levels to MG1655 cells carrying plasmid pConYCG-ManCB ***(Figure 2-3 b)***.

Another important class of glycoproteins that have been produced in laboratory

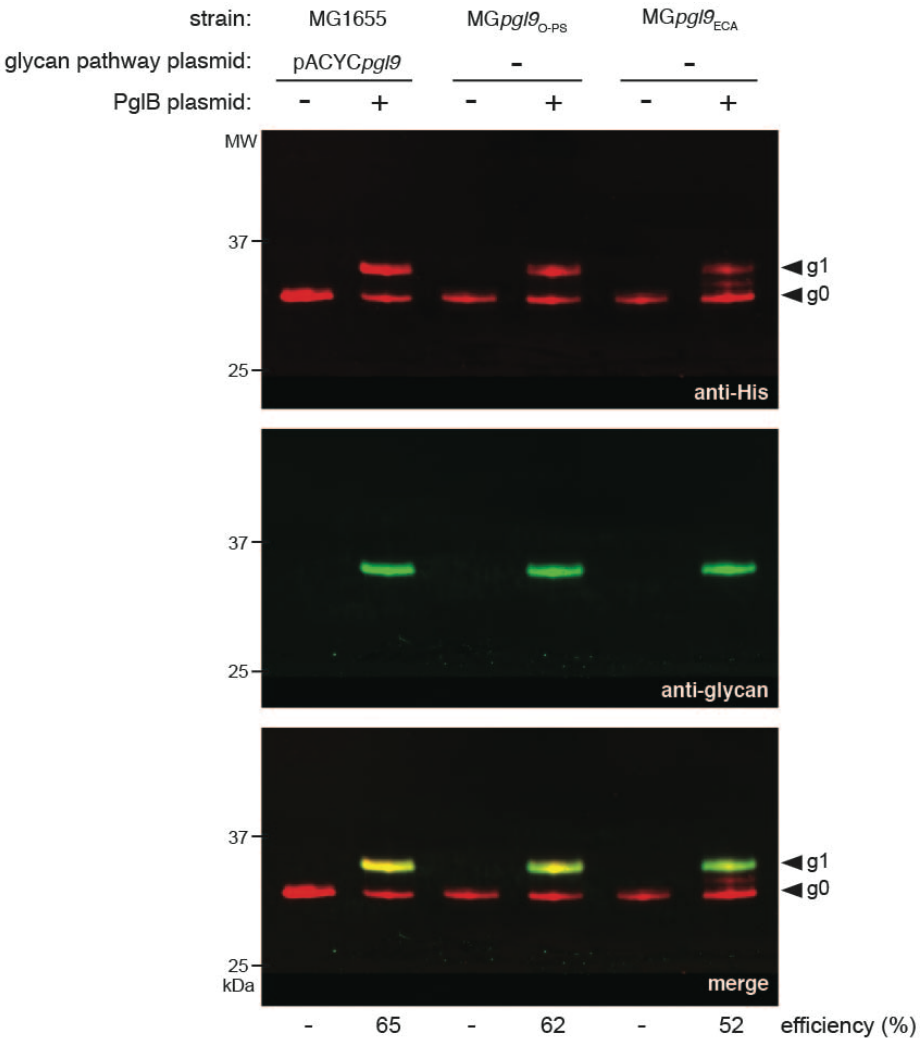


Figure 2-5: Site-specific genomic integration of orthogonal glycan biosynthesis pathway.

Western blot analysis of scFv13-R4^{DQNAT} acceptor protein in periplasmic fractions derived from: (lanes 1 and 2) E. coli MG1655 cells carrying pBS-spDsbA-scFv13-R4$^{DQNAT-GKG-His6}$, pACYCpgl9, and either plasmid pMAF10 (+) or empty pMLBAD (−); and (lanes 3–6) E. coli MGpgl9$_{O-PS}$ or MGpgl9$_{ECA}$ cells carrying pBS-spDsbA-scFv13-R4$^{DQNAT-GKG-His6}$ along with either pMAF10 (+) or empty pMLBAD (−). 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. Bottom panel depicts overlay of anti-His and hR6 blots (merge). Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of scFv13-R4^{DQNAT}. Molecular weight (MW) markers are indicated on the left. Results are representative of at least three biological replicates.

E. coli cells are bioconjugates, vaccine carrier proteins covalently modified with bacterial O-PS molecules or capsular polysaccharides(*38*). Thus, we next sought to demonstrate that the recoding strategy was compatible with the glycan biosynthesis pathways used to make bioconjugates. As proof-of-concept, we generated plasmid pRecO-PS-O56 by cloning a 9-gene locus from *E. coli* O56 that encodes enzymes for biosynthesis of an O-PS structure comprised of tetrasaccharide repeating units containing one residue of N-acetylneuraminic acid (Neu5Ac) and the common sugars GlcNAc, glucose and galactose ***(Figure 2-3 c)***. This plasmid was used to create recoded strain MG-O56$_{O\text{-}PS}$, which produced cell-surface polysaccharides that cross-reacted with O56-specific serum when analyzed by dot blot analysis and exhibited characteristic laddering when analyzed by Western blot analysis ***(Figure 2-3 c)***.

2.3.4. Double integrant requires only acceptor protein plasmid for glycosylation

Encouraged by our success generating single integrants with glycosylation-relevant GOIs integrated at two different genomic loci, we next attempted to engineer a double integrant in which the entire complement of protein glycosylation enzymes was integrated into the chromosome. For this purpose, we implemented a sequential recombineering strategy: first, λ-Red-mediated integration of the *pgl9* pathway into the O-PS locus of strain MG1655 using the pRecO-PS shuttle vector followed by removal of the Kan^R marker; second, λ-Red-mediated integration of *pglB* into the ECA locus of the markerless MG*pgl9*$_{O\text{-}PS}$ strain using the promoter-enhanced shuttle vector pRecECA-P. The result of this dual-targeting procedure was the creation of a marker-free double integrant strain, MG*pglB*$_{ECA\text{-}P}$*pgl9*$_{O\text{-}PS}$, that only required transformation with a single plasmid encoding the acceptor protein to achieve protein glycosylation

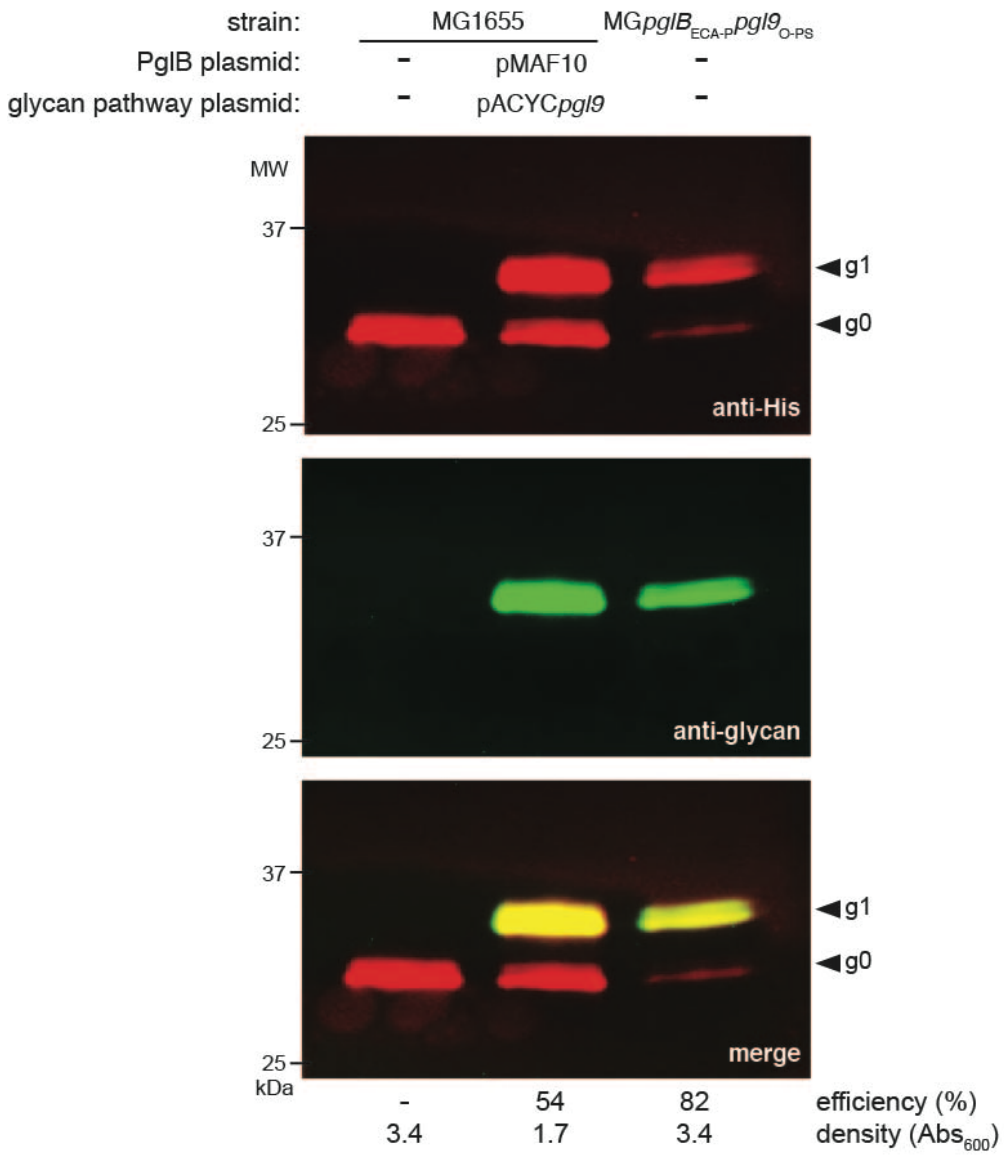


Figure 2-6: Dual-locus integration of complete protein glycosylation machinery. *Western blot analysis of scFv13-R4^{DQNAT} acceptor protein in periplasmic fractions derived from: (lane 1) E. coli MG1655 cells carrying pBS-spDsbA-scFv13-R4$^{DQNAT-GKG-His6}$ in the absence of glycosylation machinery; (lane 2) MG1655 carrying pBS-spDsbA-scFv13-R4$^{DQNAT-GKG-His6}$, pMAF10, and pACYCpgl9; and (lane 3) E. coliMGpglB$_{ECA-P}$pgl9$_{O-PS}$ carrying only pBS-spDsbA-scFv13-R4$^{DQNAT-GKG-His6}$. 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. Bottom panel depicts overlay of anti-His and hR6 blots (merge). Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of scFv13-R4^{DQNAT}. Molecular weight (MW) markers are indicated on the left. Results are representative of at least three biological replicates.*

(Figure 2-6). In fact, immunoblot analysis of scFv13-R4^{DQNAT} produced in this strain

confirmed efficient glycosylation (82%) that surpassed the level of glycosylation (54%) obtained with a prototypical three-plasmid glycoprotein system ***(Figure 2-6)***.

A major advantage of the double integrant is the ability to easily express different acceptor proteins simply by changing the target protein plasmid. To demonstrate this versatility here, we transformed MG*pglB*$_{\text{ECA-P}}$*pgl9*$_{\text{O-PS}}$ cells with a plasmid for expressing one of the following constructs: (i) spDsbA-MBP-GCG$^{\text{DQNAT}}$, a fusion between spDsbA, *E. coli* maltose-binding protein (MBP), and human glucagon (GCG; residues 1–29) followed by a C-terminal GlycTag(*35*); (ii) spDsbA-MBP-hGHv$^{138-\text{DQNAT}-142}$, a fusion between spDsbA-MBP and the placental variant of human growth hormone (hGHv) in which the naturally occurring glycosylation site at N140 was converted to an optimal bacterial sequon; and (iii) spDsbA-RNase A$^{32-\text{DQNAT}-36}$, a fusion between spDsbA and bovine pancreatic RNase A in which the naturally occurring glycosylation site at N34 was converted to an optimal bacterial sequon. All three of these target proteins were efficiently expressed in the double integrant strain, with glycosylation efficiencies matching those observed in the three-plasmid system ***(Figure 2-7)***. It is worth mentioning that the MBP-GCG$^{\text{DQNAT}}$ fusion protein was not detected in control MG1655 cells under the conditions tested, indicating poor intracellular stability in the absence of glycosylation. However, glycan addition in either the double integrant or the strain containing three plasmids greatly enhanced intracellular stability either by enhancing productive folding of the protein itself or rendering it less susceptible to proteolysis, or both. Collectively, our results indicate that diverse glycoprotein targets can be produced in the recoded strain without compromising the level of protein expression or glycosylation.

Another advantage of the glyco-recoded strain is the potential to reduce the plasmid-encoded metabolic burden that accompanies recombinant protein expression

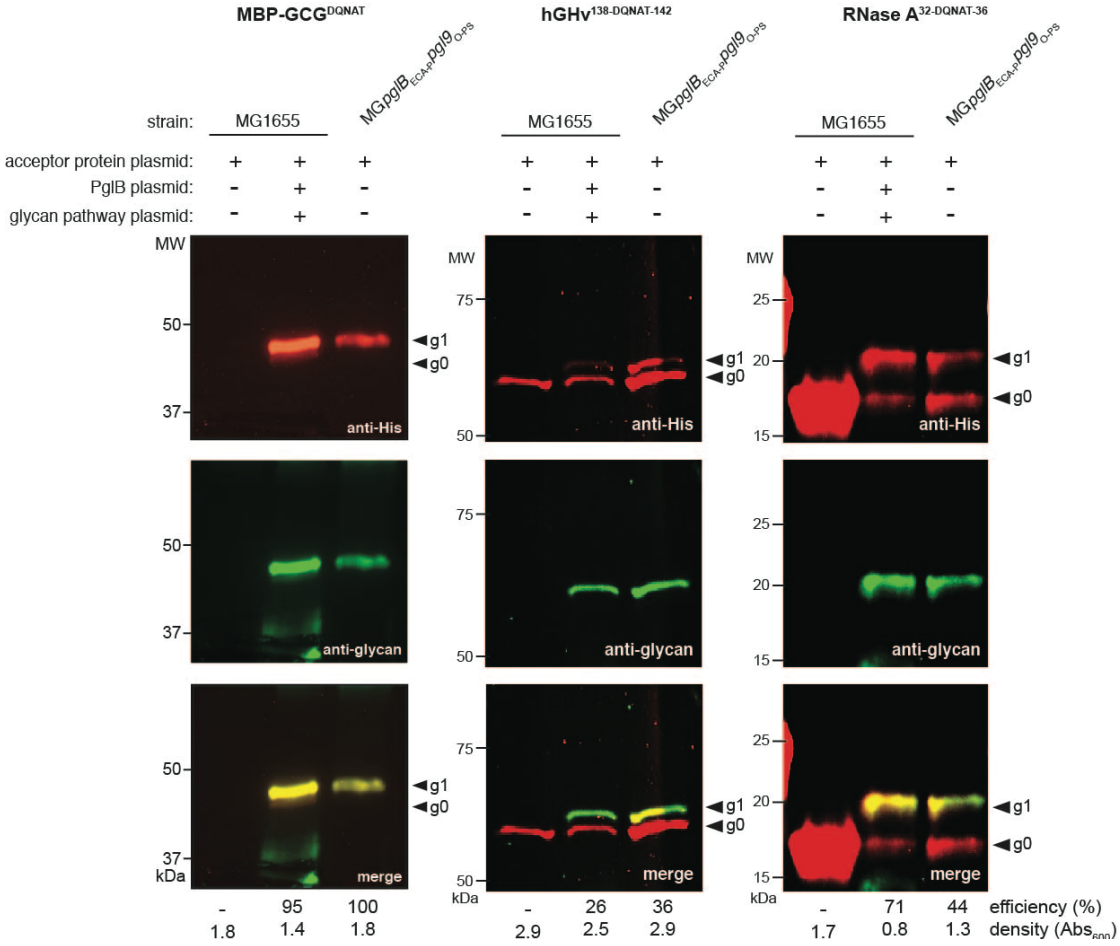


Figure 2-7: Expression of different glycoprotein targets requires a single plasmid. *Western blot analysis of periplasmic fractions derived from cells expressing MBP-GCG^{DQNAT} (left), MBP-hGHv$^{138-DQNAT-142}$ (center), and RNase A$^{32-DQNAT-36}$ (right) as indicated. Lane 1 in each blot corresponds to MG1655 cells carrying the acceptor protein plasmid (pTrc-spDsbA-MBP-GCG^{DQNAT}, pTrc-spDsbA-MBP-hGHv$^{138-DQNAT-142}$, or pTrc-spDsbA-RNase A$^{32-DQNAT-36}$) in the absence of glycosylation machinery; lane 2 in each blot corresponds to MG1655 cells carrying the acceptor protein plasmid along with pMAF10 and pACYCpgl9; and lane 3 in each blot corresponds to MGpglB$_{ECA-P}$pgl9$_{O-PS}$ carrying only the acceptor protein plasmid. 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. Bottom panel depicts overlay of anti-His and hR6 blots (merge). Arrows denote aglycosylated (g0) and singly glycosylated (g1) forms of the different acceptor proteins. Molecular weight (MW) markers are indicated on the left. Results are representative of at least three biological replicates.*

and is well-known to reduce both cell growth rate and productivity(*23*). Indeed, during expression of scFv13-R4^{DQNAT} above, the double integrant cells reached a density that was two times greater than the three plasmid-based system and comparable to MG1655 cells carrying just the acceptor protein plasmid ***(Table 2-1)***. Consistent with the difference in cell densities, the glycoprotein titers achieved with the double integrant were 2-fold higher than the titers reached with the three-plasmid strain ***(Table 2-1)***.

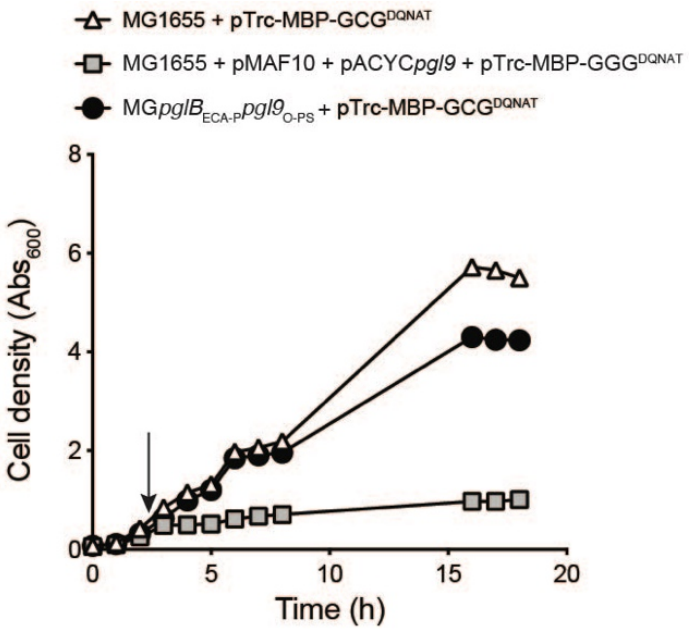


Figure 2-8: Growth analysis of strains expressing MBP-GCGDQNAT acceptor protein.
Cell density (Abs600) measured over 18 h for: MG1655 cells carrying pTrc-spDsbA-MBP-GCGDQNAT ; MG1655 carrying pTrc-spDsbA-MBPGCGDQNAT , pMAF10, and pACYCpgl9; and MGpglBECA-Ppgl9O-PS carrying only pTrc-spDsbA-MBP-GCGDQNAT as indicated. Arrow indicates induction of protein expression.

Likewise, for each of the three additional constructs, the double integrant strain reached densities that were always greater than the strain carrying three plasmids, with the most striking growth difference observed for MBP-GCG^{DQNAT} ***(Figure 2-7 and Table 2-1)***. A more thorough growth characterization revealed that the double integrant

grew similarly to the MG1655 cells expressing only MBP-GCG^{DQNAT} without any glycosylation enzymes, with both reaching much higher final densities (Abs_{600} > 4) than

MG1655 cells carrying three plasmids, which only managed to reach a final density of Abs_{600}≈1.0 over the 18-h experiment ***(Figure 2-8)***. The quantity of protein that could be purified from the double integrant was also comparable or better than the three-plasmid strain, with titers of MBP-GCG^{DQNAT} and MBP-hGHv$^{138-DQNAT-142}$ both ~2-fold greater than the amount recovered from the strain carrying three plasmids ***(Table 2-1)***.

Construct	Cell growth and protein yields[2]		
	MG1655	MG1655 + pACYC*pgl9* + pMAF10	MG*pglB*$_{ECA-P}$*pgl9*$_{O-PS}$
spDsbA-scFv13-R4$_{DQNAT}$	3.4	1.7	3.4
	0.6 ± 0.2 mg/L	0.3 ± 0.2 mg/L	0.6 ± 0.2 mg/L
spDsbA-MBP-GCG$_{DQNAT}$	1.8	1.4	1.8
	2.7 ± 0.2 mg/L	1.4 ± 0.3 mg/L	3.0 ± 0.6 mg/L
spDsbA-MBP-hGHv$_{138-DQNAT-142}$	2.9	2.5	2.9
	2.5 ± 0.2 mg/L	2.7 ± 0.1 mg/L	5.7 ± 0.2 mg/L
spDsbA-RNase A$_{32-DQNAT-36}$	1.7	0.8	1.3
	2.2 ± 0.1 mg/L	0.3 ± 0.1 mg/L	0.4 ± 0.1 mg/L

Table 2-1: Comparative cell growth and protein yields for different bacterial strains.

Interestingly, strains expressing the MBP-GCG^{DQNAT} construct grew similarly, indicating that poor growth was not the only factor responsible for the lower product titers in the three-plasmid strain background. It is also worth mentioning that the titers obtained using the double integrant were appreciably greater than those measured for MG1655 cells expressing only MBP-GCG^{DQNAT} or MBP-hGHv$^{138-DQNAT-142}$ without

[2] Final cell density of all cultures was recorded as Abs_{600} values taken after 16–18 h of overnight growth. Values are the average of three independent replicates and the standard error is less than 5%. Protein yields (mg/L) in italics were determined by purification of proteins using NiNTA chromatography and quantification using total protein assay. Values are the average of three independent replicates and the error is the standard deviation of the mean.

any glycosylation enzymes ***(Table 2-1)***, providing further support for enhanced intracellular stability afforded by glycan attachment.

2.4 Discussion

The discovery of *N*-linked protein glycosylation in bacteria(*8*) and its potential as a glycoprotein engineering platform has spawned new opportunities for metabolic engineering of glycoprotein synthesis(*5*). This is a particularly challenging endeavor because it requires seamless integration of two complex activities in living bacteria. The first is biosynthesis and conjugation of heterologous glycans, which involves metabolic pathway engineering to assemble and coordinately express a set of glycosyltransferases and an OST. The second is recombinant expression of a target acceptor protein in an appropriate subcellular compartment, which for *N*-linked protein glycosylation in Gram-negative bacteria is the periplasmic space. Typically, these activities are implemented through the use of plasmids that replicate autonomously and are maintained at a relatively high copy number when introduced into a host organism, most often a non-pathogenic bacterium such as *E. coli*. However, the orthogonal protein glycosylation system described here requires the expression of a large number of heterologous genes that necessitate the use of two or three plasmids. While plasmid-based systems offer some important advantages, most notably strong expression of the recombinant genes as a result of high copy number and powerful promoters, there are a number of potential drawbacks that may result from their use. For example, plasmids may be structurally or segregationally unstable and lost from the cells over several rounds of replication(*39, 40*). It is also well documented that multi-copy plasmids can significantly alter the normal metabolism of the host cell, further increasing the potential for plasmid instability. This metabolic burden has been linked to the expenditure of energy and resources related to plasmid replication(*24*) and translation of the plasmid-

encoded genes(*23*). This latter stress may be particularly significant when operons of heterologous genes are present in multiple copies, or when the resulting proteins have a substantial impact on the physiology of the cell, for example in the case of membrane associated or transmembrane proteins. Consequently, multi-copy plasmids can be problematic for certain metabolic engineering applications.

To circumvent these plasmid-related issues, a number of groups have reported the chromosomal integration of multi-gene pathways for producing heterologous metabolites(*31, 32, 41, 42*) and in some cases production titers were improved when the metabolic pathways were chromosomally expressed instead of from plasmids(*43, 44*). Inspired by these efforts, we developed a glyco-recoding strategy that involved replacing major non-essential polysaccharide gene clusters in the chromosome of *E. coli* K-12 strains with genes encoding carbohydrate biosynthesis and protein glycosylation machinery. Specifically, the genomic ECA and O-PS loci were systematically replaced with ~9–10 kbp of heterologous DNA encoding *C. jejuni* enzymes required for *N*-linked protein glycosylation. These native loci were targeted because, in addition to being non-essential, their gene products were judged as having the potential to interfere with the orthogonal glycosylation pathways. We found that the glyco-recoded double integrant strain, which carried the glycan biosynthesis locus at the O-PS site and the OST enzyme at the ECA site, attached glycans to acceptor protein targets with equal or greater efficiency than cells carrying the same glycosylation machinery on extrachromosomal plasmids. Importantly, the double integrant only required the maintenance of a single plasmid encoding the target protein, reducing the potential for loss of recombinant plasmids and thus improving the genetic stability of the system as well as simplifying the process by which the glycoprotein of interest is produced. While relatively efficient glycosylation was obtained in all cases that we tested, if required, further fine tuning

could be achieved by modifying or introducing additional promoter elements or by adjusting the gene copy numbers from single copy as described here to higher copy numbers as needed(*42*). Additionally, further genomic integration sites could be explored, allowing for an extension of our strategy.

In addition to achieving robust glycosylation, the double integrant strain exhibited a markedly better growth phenotype, which likely resulted from alleviating the metabolic burden that often accompanies plasmid-bearing cells(*25*). Because the poor growth of the plasmid-based glycoprotein expression host was observed to coincide with induction of PglB and the target acceptor protein, we suspect that the fitness penalty was likely caused by two energy intensive membrane processes: (i) the bilayer integration of PglB, a topologically complex integral membrane protein(*45*); and (ii) the SRP/Sec-mediated translocation of the acceptor protein. Regardless of the precise mechanism, these results clearly highlight the fitness advantage that is gained by integrating the glycosylation machinery into the genome, which eliminates issues associated with stability/ maintenance of multiple plasmids and effectively reduces expression levels of all pathway enzymes including those that may be deleterious to the host, in particular complex membrane proteins such as PglB.

It is worth mentioning that metabolic burden effects have been observed in other metabolic engineering applications. For example, the production of long-chain inorganic polyphosphate (polyP) and lycopene by recombinant expression of pathway enzymes from high-copy plasmids (> 100 copies per cell) in *E. coli* was met with significantly lower cell densities and lower overall product concentrations relative to low-copy plasmid-based systems (1–2 copies per cell) where genetic copy numbers of pathway enzymes were similar to that of the chromosome(*46*). Just as we observed here,

reducing the expression level of pathway enzymes resulted in high levels of product with little to no impact on cell growth. Collectively, these studies suggest that low gene copy number is a useful approach for metabolic engineering of diverse products, from small molecules to large glycoproteins, especially when one or more of the pathway substrates or intermediates is involved in the normal metabolism of the host organism.

Because our bacterial recombineering is based on homologous recombination, it allows insertion, deletion or alteration of any sequence in a precise manner and is not dependent on the presence of restriction sites in the target DNA. The only requirement for direct use of our constructs is the presence of sites on the genome with sufficient homology to the flanking genes on the recombineering constructs. Consequently, our pRecO-PS construct is compatible with any commonly used K-12-derived laboratory strains of *E. coli*, including MG1655 and W3110. Our pRecECA construct has a broader host range, and could potentially be used with other members of the *Enterobacteriaceae* family. If the recombination sites described in this study are not present or the homology is too low for efficient homologous recombination, our strategy can be readily adapted to other strains or species by replacing our recombination flanks with sequences found in similar genomic locations of the species of interest. While not evaluated here, there are at least four other non-essential polysaccharide biosynthesis gene clusters in *E. coli* K-12 strains(*47*) that could serve as glyco-recoding target sites for insertion of additional glyco-enzymes. These include: the *wca* operon for colanic acid polysaccharide biosynthesis(*48*); the *pgaABCD* operon encoding poly-β-1,6-N-acetylglucosamine (PNAG) biosynthesis(*49*); a type IV capsule biosynthesis pathway encoded by a pseudo-operon comprised of *gfcABCDE*, *etp*, and *etk* (this operon is nonfunctional in *E. coli* K-12 due to the presence of an IS1 element in its promoter region); and a pathway encoded by *yjbEFGH* that is predicted to be involved in

biosynthesis of an exopolysaccharide having unknown function. Furthermore, as the two loci targeted for recombineering here are generally suitable for the expression of large genetic loci, our approach is not limited to glycoengineering applications, but could easily be expanded to the expression of many other biosynthetic pathways.

Looking forward, we anticipate that glyco-recoded *E. coli* will be broadly useful for metabolic glycoprotein engineering, synthetic biology and industrial biotechnology given the genetic stability, high expression, and reduced need for plasmid maintenance. For therapeutic protein targets, this will require further glyco-engineering of the host to produce more therapeutically relevant human glycoforms that extend from the $Man_3GlcNAc_2$ glycan described here and previously(*20, 35*). For bioconjugate vaccine applications(*15-19*) pathogen-specific CPS or O-PS pathways could be integrated into the chromosome as we described here for the *E. coli* O56 O-PS structure. The prospect of long-term expression of orthogonal glycosylation pathways with minimal metabolic burden and reduced need for selectable markers provides a more stable route for chemical diversification of proteins *in vivo* that effectively expands the bacterial glycoengineering toolkit.

2.5 Materials and methods

2.5.1. Strains and plasmids

All *E. coli* strains and plasmids used in this study are listed in ***Table 2-2***. Strain MG1655 was used for all experiments unless noted otherwise. To make recombineering plasmids, flanking regions for recombination at either the O-PS or the ECA recombineering sites were amplified from *E. coli* genomic DNA and cloned on either side of the kanamycin resistance gene in pKD4 ***(Figure 2-1)***. Genes for integration were amplified and cloned into the engineered MCS on the recombineering plasmids. All

recombineering plasmids were confirmed by DNA sequencing and maintained in λpir + strains at all times. Plasmid pACYC*pgl9* was generated by PCR amplification of genes *gale* to *pglA* from pACYC(pgl_{mut}) followed by Gibson assembly into BamHI-digested pACYC184. The plasmid pTrc99A-spDsbA-MBPhGHv$^{138-DQNAT-142}$ was constructed using standard cloning procedures. Briefly, two gBlocks (Integrated DNA Technologies) were synthesized, one corresponding to MBP modified with an N-terminal DsbA signal peptide and the other corresponding to hGHv$^{138-DQNAT-142}$ with a C-terminal hexahistidine tag. The MBP gBlock was treated with SacI and BamHI and the hGHv gBlock with BamHI and HindIII, both of which were then inserted into plasmid pTrcS, a derivative of pTrc99A in which the *bla* gene was replaced with the *aadA* gene encoding spectinomycin resistance. A similar cloning strategy was adopted to construct plasmid pTrc-spDsbA-RNaseA$^{32-DQNAT-36}$.

Strain or plasmid	Relevant genotype	Reference
E. coli strains		
MG1655	F$^-$ λ$^-$ *ilvG$^-$ rfb-50 rph-1*	Laboratory stock
CLM24	F$^-$ λ$^-$ *rph-1* IN(*rrnD-rrnE*)1 Δ*waaL*	(*12*)
TransforMax™ EPI300™	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80Δ*lacZ*ΔM15 Δ*lacX74 recA1 endA1 araD139* Δ(*ara*, *leu*)7697 *galU galK* λ$^-$ *rpsL nupG trfA*	Epicentre Biotechnologies
MG*cat*$_{O\text{-}PS}$	MG1655 with *cat* gene at O-PS site	This study
MG*cat*$_{ECA}$	MG1655 with *cat* gene at ECA site	This study
MG*cat*$_{ECA\text{-}P}$	MG1655 with *cat* gene at ECA site with enhanced promoter sequence	This study
CLM*cat*$_{O\text{-}PS}$	CLM24 with *cat* gene at O-PS site	This study
CLM*cat*$_{ECA}$	CLM24 with *cat* gene at ECA site	This study
MG*pglB*$_{O\text{-}PS}$	MG1655 with *pglB* gene at O-PS site	This study
MG*pglB*$_{ECA}$	MG1655 with *pglB* gene at ECA site	This study
MG*pglB*$_{ECA\text{-}P}$	MG1655 with *pglB* gene at ECA site with enhanced promoter sequence	This study
MG*pgl9*$_{O\text{-}PS}$	MG1655 with truncated *C. jejuni pgl* locus integrated at O-PS site	This study
MG*pgl9*$_{ECA}$	MG1655 with truncated *C. jejuni pgl* locus integrated at ECA site	This study
MG-YCG-RfbMK$_{O\text{-}PS}$	MG1655 with *S. cerevisiae alg13*, *alg14*, *alg1*, *alg2* and *E. coli* O9 *rfbM-K* integrated at O-PS site	This study

MG-O56$_{O\text{-}PS}$	MG1655 with *E. coli* O56 O-PS locus integrated at O-PS site	This study
MG*pglB*$_{ECA\text{-}P}$*pgl9*$_{O\text{-}PS}$	MG1655 with *C. jejuni pglB* integrated at ECA site with enhanced promoter sequence and truncated *C. jejuni pgl* locus integrated at O-PS site	This study
Plasmids		
pKD4	Template plasmid for recombination constructs; R6K ori, Amp^R, Kan^R	(*27*)
pKD46	Encodes the λ-red recombinase; Amp^R	(*27*)
pE-FLP	Encodes the FLP recombinase; Amp^R	(*50*)
pACYC184	Cloning vector; Cm^R, Tet^R	Laboratory stock
pTrc99s	Derivative of pTrc99A, in which bla was replaced with aadA; SpR	Laboratory stock
pACYC(*pgl*$_{mut}$)	Encodes *C. jejuni pgl* locus containing a mutation in *pglB*; Cm^R	(*9*)
pACYC*pgl9*	Encodes truncated version of the *C. jejuni pgl* locus, containing a mutation in *pglB*; Cm^R	This study
pMLBAD	Expression vector; Tmp^R	(*51*)
pMAF10	pMLBAD expressing *C. jejuni* PglB; Tmp^R	(*12*)
pBS-spDsbA-scFv13-R4^{DQNAT}	pBS encoding scFv13-R4 with N-terminal signal peptide from *E. coli* DsbA and C-terminal GlycTag	(*11, 34*)
pTrc-spDsbA-MBP-GCG^{DQNAT}	pTrc99A expressing *E. coli* maltose-binding protein fused with human glucagon with N-terminal signal peptide from *E. coli* DsbA and C-terminal GlycTag; Amp^R	(*35*)
pTrc-spDsbA-MBP-hGHv$^{138\text{-}DQNAT\text{-}142}$	pTrc99S encoding E. coli MBP modi fused with placental variant of human growth hormone (hGHv) in which naturally occurring glycosylation site changed to DQNAT, with N-terminal DsbA signal peptide; Sp^R	This study
pTrc-spDsbA-RNase A$^{32\text{-}DQNAT\text{-}36}$	pTrc99S encoding bovine ribonuclease A (RNaseA) in which naturally occurring glycosylation site was changed to DQNAT, with N-terminal DsbA signal peptide; Sp^R	This study
pRecO-PS	Shuttle vector for cloning of genes for integration into the O-PS locus; Amp^R, Kan^R	This study
pRecECA	Shuttle vector for cloning of genes for integration into the ECA locus; Amp^R, Kan^R	This study

pRecECA-P	Shuttle vector for cloning of genes for integration into the ECA locus, with an additional promoter sequence for increased expression; Amp^R, Kan^R	This study
pRecO-PS*cat*	Chloramphenicol resistance gene cloned into pRecO-PS; Amp^R, Kan^R, Cm^R	This study
pRecECA*cat*	Chloramphenicol resistance gene cloned into pRecECA; Amp^R, Kan^R, Cm^R	This study
pRecECA-P*cat*	Chloramphenicol resistance gene cloned into pRecECA-P; Amp^R, Kan^R, Cm^R	This study
pRecO-PS*pglB*	*C. jejuni pglB* cloned into pRecO-PS; Amp^R, Kan^R	This study
pRecECA*pglB*	*C. jejuni pglB* cloned into pECA; Amp^R, Kan^R	This study
pRecECA-P*pglB*	*C. jejuni pglB* cloned into pECA-P; Amp^R, Kan^R	This study
pRecO-PS*pgl9*	Truncated version of *C. jejuni pgl* locus cloned into pRecO-PS; Amp^R, Kan^R	This study
pRecECA*pgl9*	Truncated version of *C. jejuni pgl* locus cloned into pRecECA; Amp^R, Kan^R	This study
pRecO-PS-YCG-RfbMK	*S. cerevisiae alg13*, *alg14*, *alg1*, *alg2* and *E. coli* O9 *rfbM-K* cloned into pRecO-PS; Amp^R, Kan^R	This study
pRecO-PS-O56	*E. coli* O56 O-PS locus cloned into pRecO-PS; Amp^R, Kan^R	This study

Table 2-2: Bacterial strains and plasmids used in this study

2.5.2. Chromosomal integration of DNA

Linear DNA required for integration was generated by long-range PCR using the recombineering plasmids as a template. The DNA was electroporated into the target strain expressing the λ-Red recombinase from pKD46 and putative successful integrants selected by overnight growth at 37 °C on LB agar plates containing kanamycin (Kan) at 50 μg/ml. Single colonies were inoculated into antibiotic-free media and grown for 6 h at 37 °C for removal of the pKD46 plasmid, and streaked onto separate LB agar plates containing ampicillin (Amp) at 100 μg/ml or Kan at 50 μg/ml. Amp^S/Kan^R colonies were picked, and analyzed for successful integration of the target genes.

2.5.3. Cell growth

MG1655 or genomically recoded integrant strains carrying one or three plasmids were grown in LB medium supplemented with the appropriate antibiotics. Briefly, 150

ml cultures were grown in 500 ml shake flasks at 37 °C. When cell density reached an absorbance at 600 nm (Abs_{600}) of ~0.5, 0.1mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 0.2% (w/v) arabinose were added to the cultures to induce expression of the target protein and plasmid-encoded PglB, respectively. Cell density (Abs_{600}) of each culture was periodically recorded over 18 h. For bacterial spot plating, overnight bacterial cultures were adjusted to a density of Abs_{600}=1.0, followed by 8 rounds of 1:10 serial dilution in LB broth. 5 µl of each dilution was spotted onto LB agar plates containing the desired antibiotic concentrations, and incubated for 16 h at 37 °C.

2.5.4. Protein expression, periplasmic extraction and purification

Overnight cultures were diluted 1:50 into fresh LB containing the required antibiotics and allowed to grow to an Abs_{600}=0.5, at which point the required inducers were added. Arabinose was added again after 5 h, and the cultures incubated for 16 h at 37 °C with shaking at 250 rpm. After incubation, the Abs_{600} of each culture was determined, and periplasmic extractions were performed as described(*12*) at a ratio of 100 µl extraction buffer per 1 ml of cell suspension at Abs_{600}=2. The proteins MBP-GCG^{DQNAT}, $MBPhGHv^{138-\ DQNAT-142}$, and $RNaseA^{32-DQNAT-36}$ were each purified from 100 ml of culture using Ni-NTA spin columns per manufacturer's instructions (Qiagen) and product titers were quantified using QuantiPro™ BCA Assay Kit (Sigma) also per the manufacturer's instructions.

2.5.5. Immunoblotting and dot blotting analysis

Periplasmic extractions were mixed with SDS loading dye and separated by SDS-PAGE using Mini-PROTEAN gels (Bio-Rad). After separation, proteins were transferred to nitrocellulose membranes and blocked with 2% milk in TBS. Primary and secondary antibodies were diluted in TBS-T and incubated with the membrane for 1 h

each. Where necessary, antibodies were multiplexed to enable two-color detection. After incubation, the membrane was washed three times with TBS-T, and imaged directly using a Chemidoc MP fluorescent imager (Bio-Rad). For dot blotting of intact cells, the density of the desired overnight culture was adjusted to Abs600=1, and 1 ml was harvested, washed, and resuspended in 1 ml PBS, followed by 6 1:10 serial dilutions in PBS. 2 µl of cell suspension was applied directly to a nitrocellulose membrane, and allowed to dry completely. Blocking and detection were then performed as above. HRP-conjugated lectins and O56 antisera derived from rabbits (Statens Serum Institut) were used for detection. Crude LPS samples were prepared as described(*19*). Electrophoretic separation and analysis were performed as described above for proteins.

2.5.6. Flow cytometric analysis

Flow cytometric analysis of glycoengineered cells was performed as described(*35*). Briefly, overnight cultures of each strain were grown in LB with relevant antibiotics. Cells were subcultured to an Abs_{600} of 0.1 in 10 ml LB and grown overnight at 30 °C. The next day, 1 ml of overnight culture was washed twice with 1 ml of PBS and resuspended in 500 µl PBS. All samples were diluted to an Abs_{600} of 0.2 in 250 µl PBS. Detection of the trimannosyl core glycan was performed with concanavalin A (ConA) lectin conjugated to AlexaFluor 488 (ThermoFisher), while detection of the *C. jejuni* heptasaccharide was performed with AlexaFluor 488-conjugated soybean agglutinin (SBA). Both lectins were diluted 1:500 in PBS, and 250 µl of diluted lectin was added to cells followed by incubation at 37 °C for 30 min. Cells were pelleted at 6000 X g for 4 mins, washed in 1 ml PBS, resuspended in 1 ml PBS and analyzed by flow cytometry using a FACS-Calibur flow cytometer (BD Biosciences). All experiments were performed in triplicate with the resulting data generated through CellQuest Pro 6.0 and analyzed on FlowJo 10.5 software.

2.6 Author contributions

L.E.Y. conceived and designed the strategy, designed and performed all research, analyzed all data, and wrote the paper. A.N. and M.L. designed and performed research, analyzed all data, and wrote the paper. M.H. and D.C.M. performed research. M.P.D. conceptualized and directed research, analyzed data, and wrote the paper.

2.7 Conflicts of interest

M.P.D. has a financial interest in Glycobia, Inc. and Versatope, Inc. M.P.D.'s interests are reviewed and managed by Cornell University in accordance with their conflict of interest policies. All other authors declare no other competing interests.

2.8 Acknowledgements

We thank Dr. Markus Aebi (ETH-Zürich) for providing strain CLM24, plasmids, and hR6 serum against the *C. jejuni* heptasaccharide glycan. This work was supported by the Defense Threat Reduction Agency (GRANT11631647 to M.P.D.) and the National Science Foundation (CBET-1605242, CBET-1159581 and CBET-1402843, all to M.P.D.)

2.9 References

1. F. Schwarz, M. Aebi, **Mechanisms and principles of N-linked protein glycosylation.** *Curr Opin Struct Biol* 21, 576-582 (2011).
2. C. M. Szymanski, B. W. Wren, **Protein glycosylation in bacterial mucosal pathogens.** *Nat Rev Microbiol* 3, 225-237 (2005).
3. J. Eichler, **Extreme sweetness: protein glycosylation in archaea.** *Nat Rev Microbiol* 11, 151-156 (2013).

4. J. A. Iwashkiw, N. F. Vozza, R. L. Kinsella, M. F. Feldman, **Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation.** *Mol Microbiol* 89, 14-28 (2013).

5. A. Natarajan, T. Jaroentomeechai, M. Li, C. J. Glasscock, M. P. DeLisa, **Metabolic engineering of glycoprotein synthesis in bacteria.** *Emerg Top Life Sci* 2, 419-432 (2018).

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

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

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

9. M. Wacker *et al.*, **N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli.** *Science* 298, 1790-1793 (2002).

10. M. Kowarik *et al.*, **Definition of the bacterial N-glycosylation site consensus sequence.** *EMBO J* 25, 1957-1966 (2006).

11. A. A. Ollis *et al.*, **Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences.** *Sci Rep* 5, 15237 (2015).

12. M. F. Feldman *et al.*, **Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli.** *Proc Natl Acad Sci U S A* 102, 3016-3021 (2005).

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

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

15. J. Cuccui *et al.*, **Exploitation of bacterial N-linked glycosylation to develop a novel recombinant glycoconjugate vaccine against Francisella tularensis.** *Open Biol* 3, 130002 (2013).

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

17. Z. Ma *et al.*, **Glycoconjugate vaccine containing Escherichia coli O157:H7 O-antigen linked with maltose-binding protein elicits humoral and cellular responses.** *PLoS One* 9, e105215 (2014).

18. M. Wacker *et al.*, **Prevention of Staphylococcus aureus infections by glycoprotein vaccines synthesized in Escherichia coli.** *J Infect Dis* 209, 1551-1561 (2014).

19. M. Wetter *et al.*, **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 (2013).

20. J. D. Valderrama-Rincon *et al.*, **An engineered eukaryotic protein glycosylation pathway in Escherichia coli.** *Nat Chem Biol* 8, 434 (2012).

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

22. A. C. Fisher *et al.*, **Production of secretory and extracellular N-linked glycoproteins in Escherichia coli.** *Appl Environ Microbiol* 77, 871-881 (2011).

23. W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis, D. S. Kompala, **Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria.** *Biotechnol Bioeng* 35, 668-681 (1990).

24. S. Birnbaum, J. E. Bailey, **Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant Escherichia coli.** *Biotechnol Bioeng* 37, 736-745 (1991).

25. A. H. Kachroo, M. Jayaram, P. A. Rowley, **Metabolic engineering without plasmids.** *Nat Biotechnol* 27, 729-731 (2009).

26. B. Strutton, S. R. P. Jaffe, J. Pandhal, P. C. Wright, **Producing a glycosylating Escherichia coli cell factory: The placement of the bacterial oligosaccharyl transferase pglB onto the genome.** *Biochem Biophys Res Commun* 495, 686-692 (2018).

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

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

29. G. Samuel, P. Reeves, **Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly.** *Carbohydrate Research* 338, 2503-2519 (2003).

30. F. Schwarz *et al.*, **A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation.** *Nat Chem Biol* 6, 264-266 (2010).

31. M. C. Bassalo *et al.*, **Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli.** *ACS Synth Biol* 5, 561-568 (2016).

32. J. A. Englaender *et al.*, **Effect of Genomic Integration Location on Heterologous Protein Expression and Metabolic Engineering in E. coli.** *ACS Synth Biol* 6, 710-720 (2017).

33. J. R. Kelly *et al.*, **Measuring the activity of BioBrick promoters using an in vivo reference standard.** *Journal of Biological Engineering* 3, 4 (2009).

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

35. C. J. Glasscock *et al.*, **A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation.** *Metab Eng* 47, 488-495 (2018).

36. F. Schwarz *et al.*, **Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo.** *Glycobiology* 21, 45-54 (2011).

37. P. Jayaratne *et al.*, **Cloning and analysis of duplicated rfbM and rfbK genes involved in the formation of GDP-mannose in Escherichia coli O9:K30 and participation of rfb genes in the synthesis of the group I K30 capsular polysaccharide.** *J Bacteriol* 176, 3126-3139 (1994).

38. J. Ihssen *et al.*, **Production of glycoprotein vaccines in Escherichia coli.** *Microb Cell Fact* 9, 61 (2010).

39. D. Noack *et al.*, **Maintenance and genetic stability of vector plasmids pBR322 and pBR325 in Escherichia coli K12 strains grown in a chemostat.** *Mol Gen Genet* 184, 121-124 (1981).

40. M. O'Connor, M. Peifer, W. Bender, **Construction of large DNA segments in Escherichia coli.** *Science* 244, 1307-1312 (1989).

41. C. Y. Ng, I. Farasat, C. D. Maranas, H. M. Salis, **Rational design of a synthetic Entner-Doudoroff pathway for improved and controllable NADPH regeneration.** *Metab Eng* 29, 86-96 (2015).

42. K. E. Tyo, P. K. Ajikumar, G. Stephanopoulos, **Stabilized gene duplication enables long-term selection-free heterologous pathway expression.** *Nat Biotechnol* 27, 760-765 (2009).

43. C. N. Santos, D. D. Regitsky, Y. Yoshikuni, **Implementation of stable and complex biological systems through recombinase-assisted genome engineering.** *Nat Commun* 4, 2503 (2013).

44. J. Wang *et al.*, **Engineering of a Highly Efficient Escherichia coli Strain for Mevalonate Fermentation through Chromosomal Integration.** *Appl Environ Microbiol* 82, 7176-7184 (2016).

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

46. K. L. Jones, S. W. Kim, J. D. Keasling, **Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria.** *Metab Eng* 2, 328-338 (2000).

47. M. Ionescu, S. Belkin, **Overproduction of exopolysaccharides by an Escherichia coli K-12 rpoS mutant in response to osmotic stress.** *Appl Environ Microbiol* 75, 483-492 (2009).

48. G. Stevenson, K. Andrianopoulos, M. Hobbs, P. R. Reeves, **Organization of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid.** *J Bacteriol* 178, 4885-4893 (1996).

49. Y. Itoh *et al.*, **Roles of pgaABCD genes in synthesis, modification, and export of the Escherichia coli biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine.** *J Bacteriol* 190, 3670-3680 (2008).

50. F. St-Pierre *et al.*, **One-step cloning and chromosomal integration of DNA.** *ACS Synth Biol* 2, 537-541 (2013).

51. M. D. Lefebre, M. A. Valvano, **Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in Burkholderia cepacia complex isolates.** *Appl Environ Microbiol* 68, 5956-5964 (2002).

CHAPTER 3

Engineering an orthogonal O-linked protein glycosylation pathway in bacteria for producing tumor-associated glycoconjugates [1]

3.1 Abstract

We have constructed a novel heterologous *O*-glycosylation pathway in *E. coli* that installs the monosaccharide Tn antigen and disaccharide T antigen onto serine residues in motifs from human glycoproteins that are natively found glycosylated. This platform employs Gne, a UDP-GlcNAc 4-epimerase from *Campylobacter jejuni*, PglC, a GalNAcT from *Acinetobacter baumannii*, and WbwC, a GalT from *Escherichia coli* O104 in concert with *O*-OSTs from *Neisseria gonorrhoeae* and *Neisseria meningitidis* to glycosylate 7 different acceptor proteins bearing an engineered glycosylation motif, and 8 authentic human glycotopes. We show that this system is functional both *in vivo* and *in vitro*, and on internal and terminal glycosites. Finally, we demonstrate the glycosylation of immunologically relevant epitopes derived from the VNTR sequence of MUC1 with tumor associated cancer antigens.

3.2 Introduction

Escherichia coli is burgeoning as host for the study of glycosylation mechanisms and production of glycoconjugates(*1, 2*). To this end, a lot of progress has been made by creating orthologous pathways that can generate *N*-linked glycoproteins bearing bacterial *O*-antigens, native bacterial glycans, and even mammalian core-glycans all linked to the asparagine residue on proteins, including at authentic human glycosylation sites. However, *O*-linked glycosylation, a parallel system of post-translational

[1] *This chapter is based on a manuscript in preparation:* Natarajan A., Cox E., Jaroentomeechai T., Cabrera M., and DeLisa M.P. Engineering an orthogonal *O*-linked protein glycosylation pathway in bacteria for producing tumor-associated glycoconjugates.

modification extensively found in eukaryotic hosts(*3, 4*) continues to remain vastly intractable.

O-linked glycosylation refers to the covalent linkage of sugars to predominantly serine or threonine amino acids *via* a hydroxyl bond. This represents a crucial modification that impacts the structure and function of many glycoproteins(*3, 4*). Unlike *N*-glycosylation, *O*-glycosylation has been challenging to recreate in non-native hosts since substrate specificities of the key native initiating enzymes from the ppGalNAcT family of glycosyltransferases remain poorly understood(*5, 6*), and until recently bacterial homologs were few and far between(*1*). However, recently, the human core-1 N-acetyl galactosamine transferase (ppGalNAcT1) has been successfully expressed in *E. coli* to produce recombinant *O*-glycoproteins(*7, 8*), with the key limitation that these orthologous pathways are only able to glycosylate proteins at the threonine amino acid.

Recent discovery of bacterial *O*-oligosaccharyltransferases (*O*-OST)(*9, 10*) that are able to transfer a diverse range of sugars(*9, 11*) onto serine residues on acceptor proteins has provided an opportunity to fully expand the glyco-potential of *E. coli* to produce human glycotopes bearing authentic human *O*-glycans. As a start, we aimed to transfer the core-1 mucin-type human glycan, defined by a GalNAc-Gal disaccharide (T antigen), since this core glycan can serve as a primer for subsequent elaboration. The T antigen and its derivatives are a commonly found human *O*-glycans(*3*), and bear great significance as a tumor associated cancer antigen(*12*). Therefore, we developed here a recombinant pathway for the transfer of the Tn and T antigens using the *O*-OSTs, PglO from *Neisseria gonorrhoeae*(*9*) and PglL *Neisseria meningitidis*(*10*). We test the scope of our platform to glycosylate a synthetic MOOR acceptor sequence(*13*) fused to 7 different acceptor proteins, each with a unique utility either as a reagent or of therapeutic

significance. Further, overcoming previous skepticism(*7*), we demonstrate the ability to glycosylate native human glycotopes that naturally bear the core-1 glycan, derived from erythropoietin, glycophorin C and the ectodomain of human MUC1. The ability to glycosylate an immunologically relevant 41-amino acid motif derived from the cancer-associated VNTR region of MUC1 with a tumor associated cancer antigen glycan is noteworthy.

3.3 Results

3.3.1. Engineered assembly of T antigen on lipid carrier

As a first step towards an orthogonal pathway for *O*-linked glycosylation of acceptor proteins in *E. coli*, we sought to assemble the most common *O*-glycan structure, galactosyl-β1-3-N-acetylgalactosamine (βGal-(1-3)-GalNAc; core 1) on the undecaprenyl-phosphate (Und-P) lipid carrier ***(Figure 3-1)***. This carrier has been exploited for assembling a range of heterologous glycan structures in *E. coli* including bacterial *N*- and *O*-glycans, bacterial *O*-antigens and eukaryotic N-glycans(*1, 2*). However, to our knowledge, assembly of a human-type *O*-glycan on Und-P has not been demonstrated. However, we identified three major impediments: (i) competition for the Und-P substrate from *E. coli*'s native phosphoglycosyltransferase WecA that utilizes UDP-N-acetylglucosamine (UDP-GlcNAc) to generate diphosphoundecaprenol-N-acetylglucosamine (Und-PP-GlcNAc)(*14*); (ii) the inability of any laboratory strain of *E. coli* to naturally produce UDP-N-acetylgalactosamine (UDP-GalNAc)(*15*); and (iii) the lack of an enzyme that has been proven to catalyze the synthesis of diphosphoundecaprenol-N-acetylgalactosamine (UndPP-GalNAc) in *E. coli*. To address these issues, we first generated a strain of *E. coli* lacking the *wecA* gene to eliminate formation of Und-PP-GlcNAc. Next, we generated a plasmid encoding the UDP-GlcNAc 4-epimerase from *Campylobacter jejuni* (*Cj*Gne), which is capable of

generating a robust pool of UDP-GalNAc in *E. coli*(*16*). To this plasmid, we added the gene encoding PglC from *Acinetobacter baumannii* ATCC 17978, which is known to specifically transfer UDP-GalNAc to Und-P in its native host(*17*).

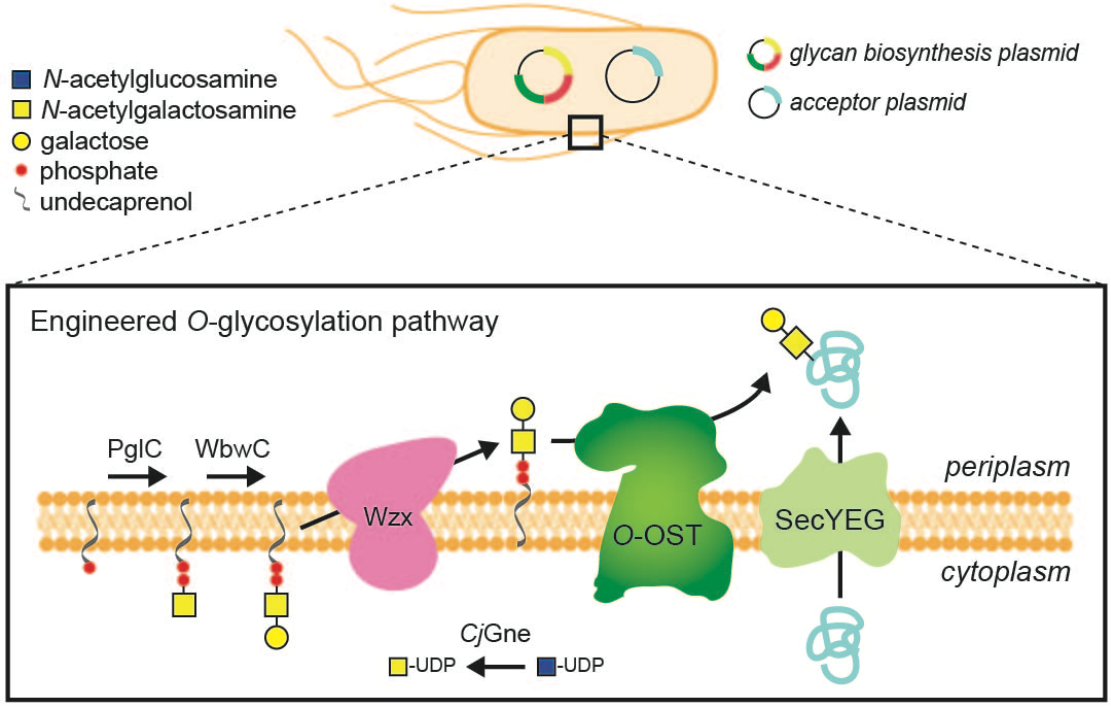


Figure 3-1: Schematic illustration of proposed core 1 O-glycosylation pathway. *CjGne creates a pool of UDP-GalNAc that serves as a substrate for AcPglC towards the synthesis of UndPP-GalNAc. This monosaccharide, the lipid-linked Tn antigen, is then elaborated by EcWbwC to generate UndPP-GalNAc-β1-3-Gal, the lipid-linked T antigen. This glycolipid is then flipped to the periplasmic face by Wzx and transferred en bloc to a serine amino acid on an acceptor protein by an O-OST.*

The addition of galactose requires a galactosyl transferase (GalT) capable of utilizing Und-PP-GalNAc as a substrate on the cytoplasmic phase of *E. coli*'s periplasmic membrane. To identify a candidate enzyme, we curated a diverse panel of GalTs from different organisms including: (i) the core 1 synthases, glycoprotein-N-acetylgalactosamine 3-β-galactosyltransferase 1 from *Homo sapiens*(*18*) (*Hs*C1GalT1) and *Drosophila melanogaster* (*Dm*C1GalT1)(*19*) each cloned without their respective 29- and 50-residue *N*-terminal transmembrane anchors; (ii) *Bifidobacterium infantis* D-

galactosyl-β1-3-N-acetyl-D-hexosamine phosphorylase (*Bi*GalHexNAcP),(*20*) (iii) the "S42" mutant of *C. jejuni* β1-3-galactosyltransferase (*Cj*CgtB) that has been engineered for improved catalytic activity,(*21*) (iv) and the β-1,3-galactosyltransferases derived from the *O*-antigen biosynthesis pathways of *E. coli* O86 (*Ec*WbnJ)(*22*) and *E. coli* O104 (*Ec*WbwC).(*23*) To rapidly screen the activity of these GalTs, we employed a flow cytometric assay developed previously by our group (*24, 25*). This assay takes advantage of the observation that lipid-linked glycans assembled on Und-P are flipped into the periplasm by the native *E. coli* flippase, Wzx, and transferred onto lipid A-core by the *O*-antigen ligase, WaaL ***(Figure 3-2a)***. Upon shuttling to the outer membrane, lipid A-core displays the attached glycan on the cell surface, where it can be readily detected by a fluorescently tagged antibody or lectin. Here, we assessed the catalytic activity of our GalT panel using *Arachis hypogaea* derived peanut agglutinin (PNA) lectin, which is well known to recognize the T antigen disaccharide (*26*), conjugated with fluorescein isothiocyanate (FITC). Flow cytometric analysis of PNA-FITC labeled cells expressing putative T antigen pathways identified *Ec*WbwC as the only enzyme in our panel with the ability to transfer galactose to Und-PP-linked GalNAc ***(Figure 3-2b, c)***; hence, co-expression of *Cj*Gne, *Ab*PglC, and *Ec*WbwC from plasmid pMW07-T served as the foundational *O*-glycosylation pathway for all further experiments.

3.3.2. Site-specific installation of T antigen on acceptor proteins

The next step in our orthogonal pathway requires an enzyme for transferring the T antigen from the lipid carrier to hydroxylated amino acids in a target protein ***(Figure 3-1)***. For this activity, we turned our attention to a family of bacterial *O*-oligosaccharyltransferases (*O*-OSTs) that catalyze *O*-linked glycosylation of their protein substrates *via* an *en bloc* mechanism in which preassembled glycans borne on Und-PP are flipped to the periplasm where they serve as substrates for the *O*-OST.(*27*)

The PglL enzyme from *Neisseria meningitidis* (*Nm*PglL) is one of the best characterized *O*-OSTs to date. Of particular relevance to our orthogonal pathway, *Nm*PglL is a promiscuous enzyme that exhibits remarkably relaxed glycan specificity, enabling it to transfer many glycans from Und-PP to acceptor proteins (*11, 28*). Moreover, *Nm*PglL can be functionally expressed in *E. coli* where it has been shown to

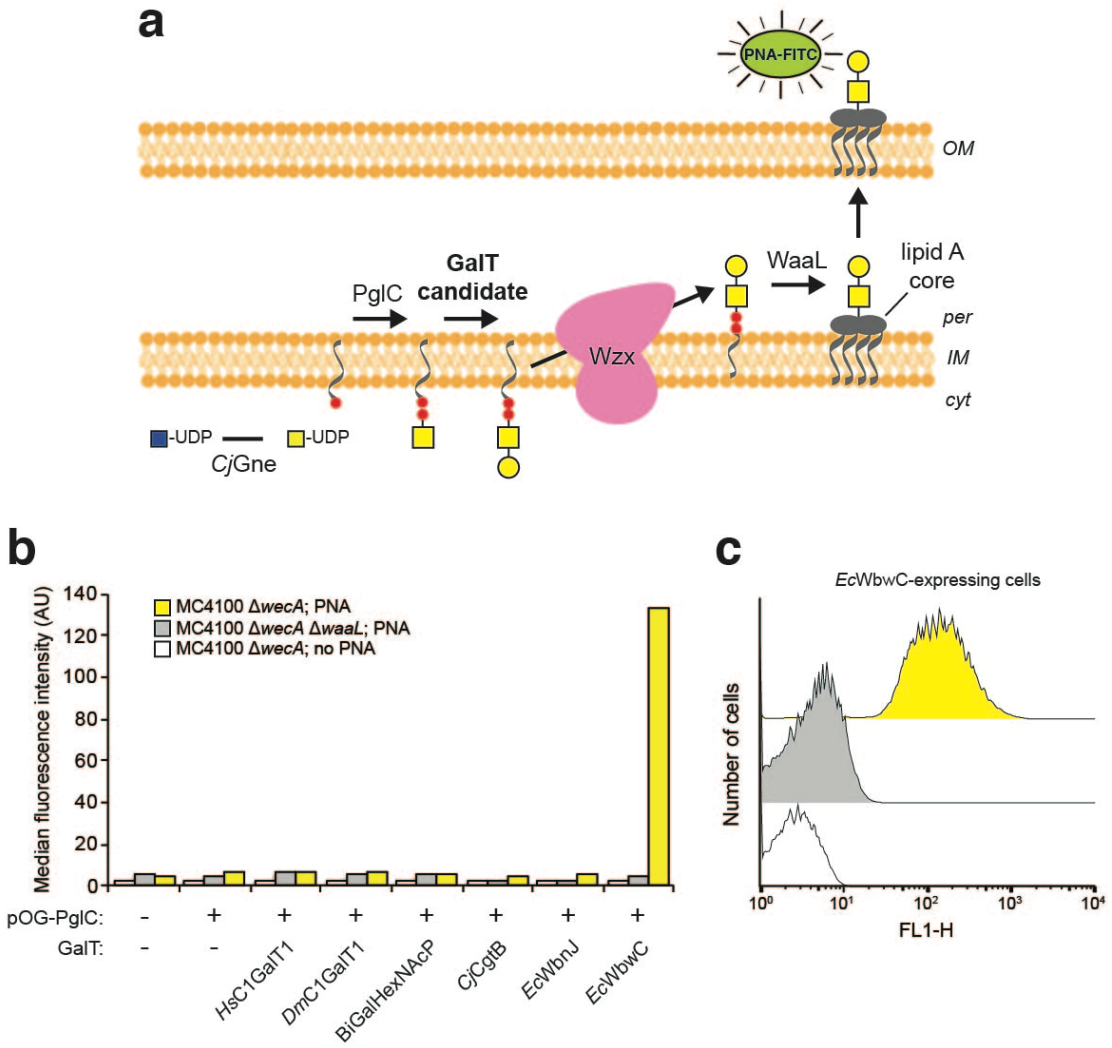


Figure 3-2: FACS based assay identifying EcWbwC as the most active GalT.
a. Strategy for the FACS assay, where glycan is transferred from UndPP to a lipid A core by WaaL in the periplasm and displayed on the surface of E. coli cells. The glycan was then probed using PNA::FITC. b. Histogram of the median fluorescence intensity measured by FACS. The complete experimental case is illustrated in yellow, with negative controls - deletion of waaL in grey and absence of PNA in white. c. Histogram of samples expressing EcWbwC marked clear shift in fluorescence.

transfer bacterial glycans to acceptor proteins (*11, 28*). Thus, we postulated that *Nm*PglL, or its homologs such as *N. gonorrhoeae* PglO (*Ng*PglO) that bears 95% sequence identity to *Nm*PglL, would be ideal candidates for outfitting proteins of interest with human *O*-glycans, starting with T antigen. To test this notion, we first added the gene encoding *Nm*PglL or *Ng*PglO to our *O*-glycosylation pathway yielding plasmids pMW07-T::Nm and pMW07-T::Ng. Next, for the acceptor protein, we created a pEXT20-based plasmid encoding an engineered version of *E. coli* maltose-binding protein (MBP) modified at its C-terminus with a MOOR (minimum optimal *O*-linked recognition) tag that was previously optimized for recognition by *Nm*PglL(*13*). The MOOR motif is comprised of an 8-residue core sequence (WPAAASAP) that mimics the Ser63 *O*-glycosylation site (*O*-glycosite) in the native substrate of *Nm*PglL, PilE, and also includes two non-native hydrophillic fragments of 10 and 7 amino acids, respectively, that flank the core ***(Figure 3-3a)***. The resulting *O*-glycosylation pathway and acceptor protein plasmids were used to co-transform *E. coli* strain W3110, which lacked *wecA* to eliminate competition with PglC as discussed above as well as *waaL* to ensure that Und-PP-linked glycans were not shunted to lipid A-core but instead remained available for the *O*-OST. These genome editing efforts ensured that metabolic flux was tilted in favor of our desired product. Following expression in these cells, MBP^{MOOR} was purified and subject to Western blot analysis, which revealed strong glycosylation of this engineered acceptor protein with the T antigen as evidenced by cross-reaction with PNA ***(Figure 3-3b)***. As expected, glycosylation was completely undetectable when either *Ng*PglO or *Nm*PglL was removed or the serine residue in the MOOR tag was substituted with glycine. Mass spectrometry confirmed the presence of a HexNAc-Hex disaccharide (data not shown) on the relevant Ser residue highlighted in red in ***Figure 3-3a***. Along the way, we also demonstrated the transfer of the monosaccharide N-acetyl-D-galactosamine, Tn antigen, onto the MBPMOOR acceptor

protein, in the absence of *Ec*WbwC through both *Ng*PglO and *Nm*PglL ***(Figure 3-3c)***. This was confirmed through cross reaction with lectin derived from *Vicia villosa* (VVA) that specifically recognizes the Tn antigen(*29*). Collectively, these results demonstrate the functionality of a novel engineered pathway for making and installing human *O*-glycans on proteins.

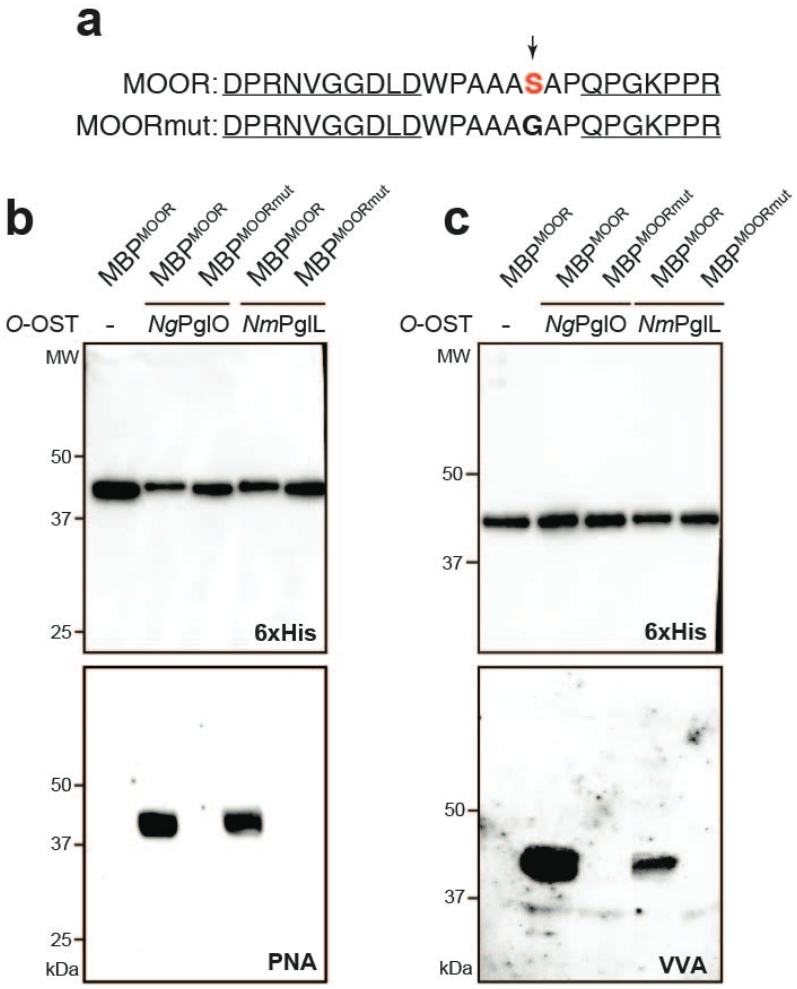


Figure 3-3: Expression of O-glycosylated MBP^{MOOR} protein bearing the T and Tn antigens.

a. MOOR and MOORmut sequences that were fused to MBP, highlighting the glycosylated serine residue confirmed by mass spectrometry in red and the hydrophilic flanking region underlined. b, c. Immunoblot analysis of Ni-NTA purified proteins, probed with anti-his antibody against the C-terminal polyhistidine tag on acceptor proteins, PNA lectin reactive to the T antigen disaccharide, and VVA lectin detecting the Tn antigen monosaccharide. O-OST status and identity of acceptor proteins are marked above the blots, and molecular weight markers (MW) to the left.

3.3.3. O-OSTs glycosylate an expanded set of acceptor proteins with T antigens

Next, we sought to determine the range of acceptor proteins that could be glycosylated by our orthogonal T antigen glycosylation pathway. This was first evaluated by grafting the MOOR tag onto the C-terminus of a range of different carrier proteins including: (i) YebF, a natively secreted *E. coli* protein that has previously been used for studying and engineering *N*-linked glycosylation in *E. coli* (*30-32*); (ii) *E. coli* glutathione-S-transferase (GST) that is widely used in pull-down assays to identify interacting biomolecules(*33*); (iii), single-chain antibody fragment scFv13-R4(*34*) and superfolder green fluorescent protein (sfGFP)(*35*), both of which have been *N*-

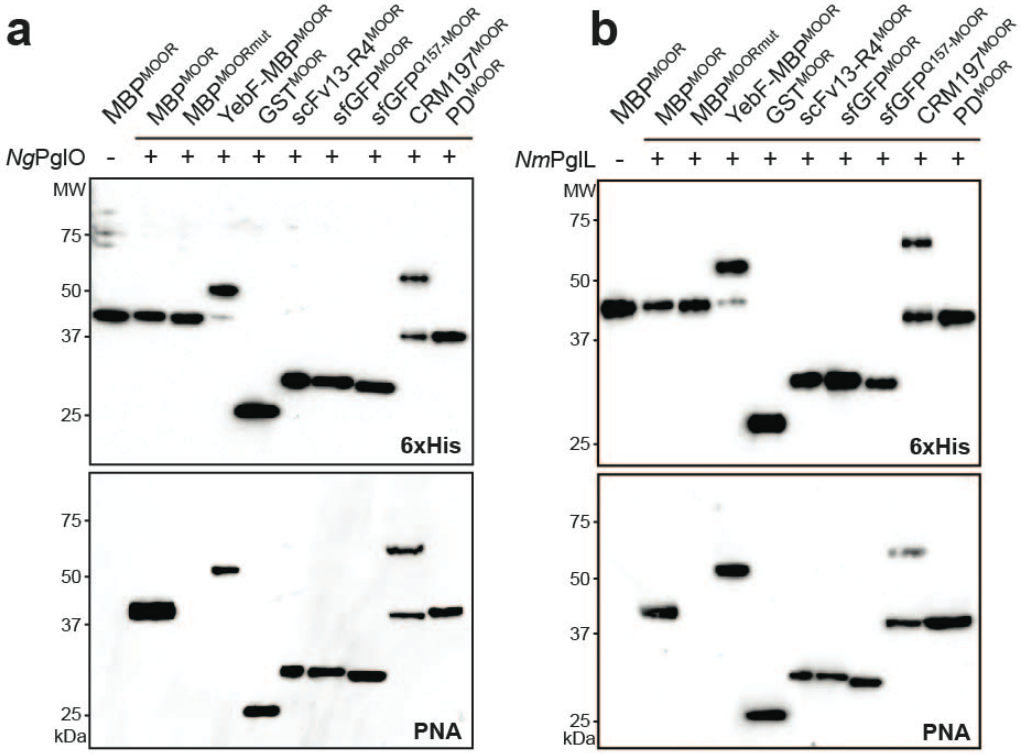


Figure 3-4: Expanded panel of protein carriers bearing the T antigen.
a, b. Immunoblot analysis of Ni-NTA purified proteins probed with anti-his antibody against the C-terminal polyhistidine tag on the acceptor proteins, and PNA lectin reactive to the T antigen disaccharide. Label above the blot indicates status of O-OST and identity of acceptor proteins that bears the C terminal MOOR tag. Molecular weight markers (MW) are indicated to the left.

glycosylated in glycoengineered *E. coli*(*25, 31*); and (iv) cross-reacting material 197 (CRM197), a genetically detoxified toxin from *Corynebacterium diphtheriae*, and protein D (PD), a surface lipoprotein from *Haemophilus influenzae*, both of which are used as carrier proteins in licensed glyco-conjugate vaccines(*36, 37*). Following expression and purification, all of these proteins cross-reacted with PNA in the presence of *Ng*PglO ***(Figure 3-4a)*** and *Nm*PglL ***(Figure 3-4b)***, confirming glycosylation with T antigen. We also engineered a variant of sfGFP in which the MOOR motif was grafted in an internal loop starting at Gln157, which was also observed to be glycosylated by both *O*-OSTs.

We further tested the generality of the pathway by swapping the 8-residue core sequence of the MOOR tag with a panel of different *O*-glycosylation motifs derived from human glycoproteins ***(Figure 3-5a)***, that natively bear the T antigen or its sialylated variants at these sites. These included 8 residues surrounding the: (i) Ser126 *O*-glycosite in human erythropoietin (EPO);(*38*) (ii) Ser24 *O*-glycosite in human glycophorin C (GPC), a surface glycoprotein found on red blood cells that marks the Gerbich antigen system(*39*). We also included 8 residues derived from the ectodomain of human mucin 1 (MUC1), which is expressed on the apical surface of glandular epithelial cells at low levels but following cancerous transformation is expressed at very high levels and with altered glycosylation patterns(*40*). Here, each of these human glycoprotein-derived sequences was glycosylated by *Ng*PglO ***(Figure 3-5b)*** whereas *Nm*PglL glycosylated the EPO and MUC1 sequences but only weakly glycosylated GPC ***(Figure 3-5c)***, revealing a subtle difference in substrate selectivity for these *O*-OSTs. Mass spectrometry confirmed the presence of a HexNAc-Hex disaccharide (data not shown) on the relevant Ser residue highlighted in red ***(Figure 3-6c)***. Taken together, these results highlight the ability of our glycoprotein engineering platform to post-

translationally modify motifs derived from natural human glycotopes, and beyond those derived from the native substrates of *Ng*PglO and *Nm*PglL.

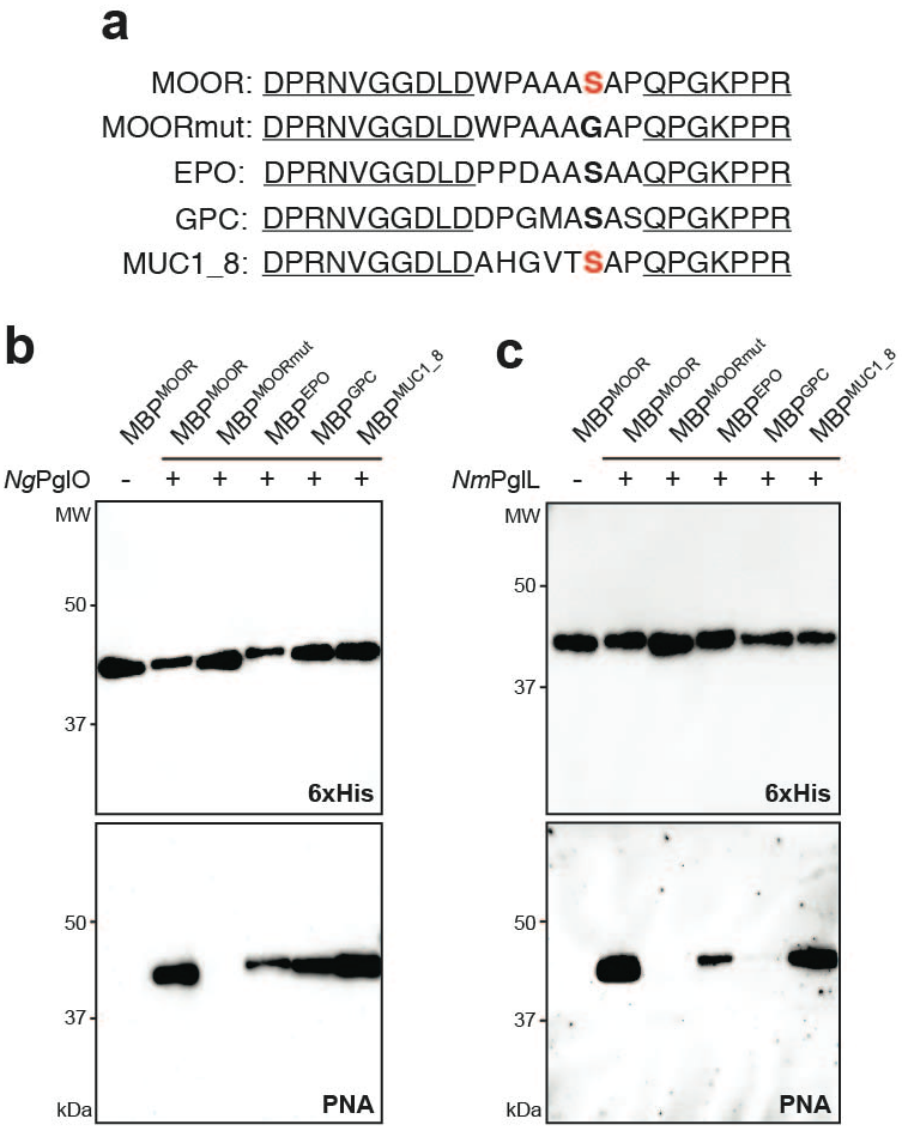


Figure 3-5: Glycosylation of motifs derived from native glycoproteins

a. Sequence of 8-mer motifs derived from authentic human O-glycosites and presented between the MOOR-derived flanking regions as C terminal fusions to MBP. Red highlights the glycosylated serine residue confirmed by mass spectrometry and the hydrophilic flanking region are underlined. b, c. Immunoblot analysis of Ni-NTA purified proteins, probed with anti-his antibody against the C-terminal polyhistidine tag on acceptor proteins, and PNA lectin reactive to the T antigen disaccharide. O-OST status and identity of acceptor proteins are marked above the blots, and molecular weight markers (MW) to the left.

3.3.4. Expression of immunologically relevant MUC1 glycotopes

Inspired by the ability of our glycoprotein engineering platform to glycosylate an 8-mer MUC1 peptide, we next attempted to generate more immunologically relevant MUC1 glycotopes. The extracellular domain of MUC1 contains a variable number of tandem repeats (VNTR) consisting of 20-120 repeats of a 20-amino acid sequence (PDTRPAPGSTAPPAHGVTSA) that contains five potential *O*-glycosylation sites (underlined)(*41*). In a number of different cancers, these sites are aberrantly glycosylated with truncated carbohydrate structures antigen including the T antigen as well as Tn (αGalNAc-Ser/Thr) and STn (αNeu5Ac-(2,6)-αGalNAc-Ser/Thr) antigens(*42*). Here, we created 4 additional sequences derived from the MUC1 VNTR region by incrementally extending the MUC1_8 motif by two amino acids at both ends and cloned each between the hydrophillic flanking regions at the C-terminus of our MBP carrier protein. The sequence designs ***(Figure 3-6a)*** included motifs based on the most frequent minimal epitopes of natural MUC1 recognized by IgG and IgM antibodies including the PPAHGVT, PDTRP, and RPAPGS(*43*) that feature in our MUC1_12, MUC1_16, and MUC1_24 designs, respectively, and on epitopes that bind to specific human MHC class I molecules including STAPPAHGV, SAPDTRPAP, and SAPDTRPAPG(*44*) that feature in MUC1_20 and MUC1_24. Western blot analysis revealed that each of these extended MUC1 motifs were glycosylated with T antigen by *Ng*PglO ***(Figure 3-6b)*** whereas *Nm*PglL did not detectably glycosylate the 12-residue MUC1 sequence and only weakly glycosylated the 24-residue MUC1 derivative ***(Figure 3-6c)***, consistent with the subtle differences in substrate selectivity for the *O*-OSTs observed above. Moreover, the MUC1_16, MUC1_20, and MUC1_24 constructs each cross-reacted with an anti-human MUC1 antibody that is widely used to label MUC1 in different cancer cell lines including several breast cancer lines(*45*) ***(Figure 3-***

6b), confirming the immunological relevance of these MUC1 peptides. The observation

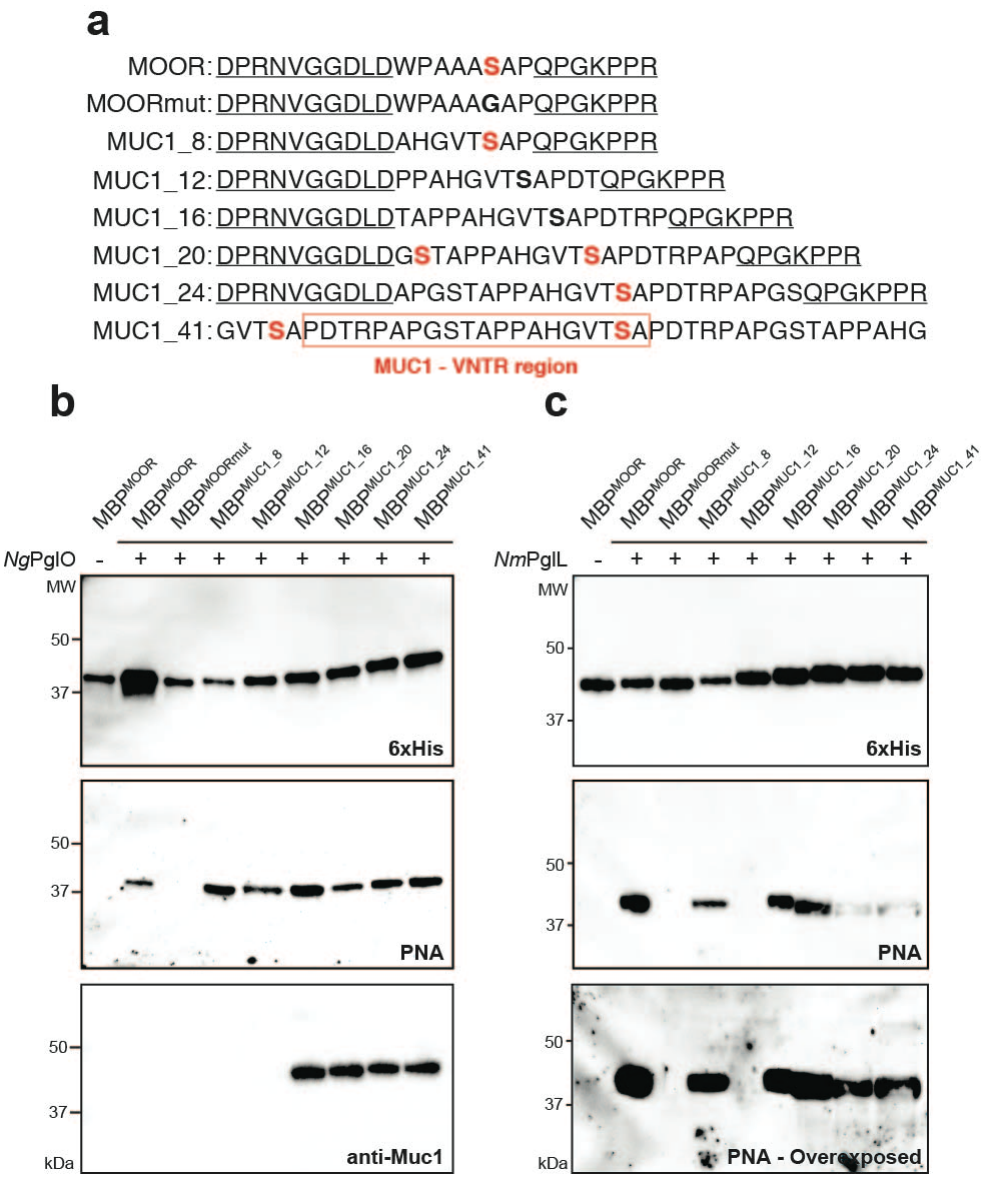


Figure 3-6: Synthesis of authentic human MUC1 glycotopes.
a. Sequences of MUC1 derived glycosylation motifs presented as fusions to MBP. Underline indicated the MOOR-based flanking, and red serines highlight the glycosylated residues confirmed by mass spectrometry. b, c. Immunoblot analysis of Ni-NTA purified proteins, probed with anti-his antibody against the C-terminal polyhistidine tag on acceptor proteins, PNA lectin reactive to the T antigen disaccharide, and anti-Muc1 antibody reactive to native human MUC1. O-OST status and identity of acceptor proteins are marked above the blots, and molecular weight markers (MW) to the left.

that *Nm*PglL and *Ng*PglO could recognize core motifs derived from human *O*-glycoproteins and that these motifs could be as a long as 24 residues suggested a greater flexibility than was first reported(*13*).

So far, glycosylation of the different *O*-glycosite motifs was achieved in the presence of hydrophilic flanking sequences (DPRNVGGDLD and QPGKPPR) that were previously shown to be required for recognition and efficient modification by *Nm*PglL(*13*). However, the requirement for these additional fragments effectively limits the authenticity of glycotopes that can be produced by this system. Therefore, we next determined whether the *O*-OSTs could recognize a 41-residue MUC1 sequence containing the previously glycosylated 24-residue MUC1 VNTR motif flanked with additional stretches of the MUC1 VNTR sequence rather than the original MOOR fragments ***(Figure 3-6a)***. Importantly, both *Ng*PglO and *Nm*PglL were able to transfer the T antigen to the MUC1_41 construct ***(Figure 3-6b, c)***. Mass spectrometry has confirmed the presence of a HexNAc-Hex disaccharide at all of the Ser residues highlighted in red in ***(Figure 3-6a)***. Overall, these results demonstrate that our orthogonal glycosylation pathway can generate authentic human *O*-glycoprotein epitopes without the need for hydrophilic flanking regions. Indeed, estimates of the grand average of hydropathy (GRAVY)(*46*) using the ProtParam server(*47*) on ExPASy revealed that both the N- and C- terminal flanking regions in the native MUC1_41 sequence are significantly more hydrophobic than those found in the MOOR sequence, extending our understanding of the acceptor sequence preferences for these *O*-OSTs.

3.3.5. In vitro O-glycosylation with the T antigen

Cell free modalities of glycosylation are gaining increasing attention as a tool to rapidly expand our understanding of biosciences, and for the on demand

biomanufacturing of glycoproteins products(*48*). Therefore, we next sought to test if our engineered pathway was functional *in vitro*. We prepared detergent extracts from cell lysates enriched with lipid-linked oligosaccharides (LLOs) bearing the T antigen and the *Ng*PglO enzyme. Upon addition of independently purified MBP^{MOOR} protein, we saw clear glycosylation through cross-reaction with PNA ***(Figure 3-7)***. Preparing reactions with either the $MBP^{MOORmut}$ or by skipping any of the required components

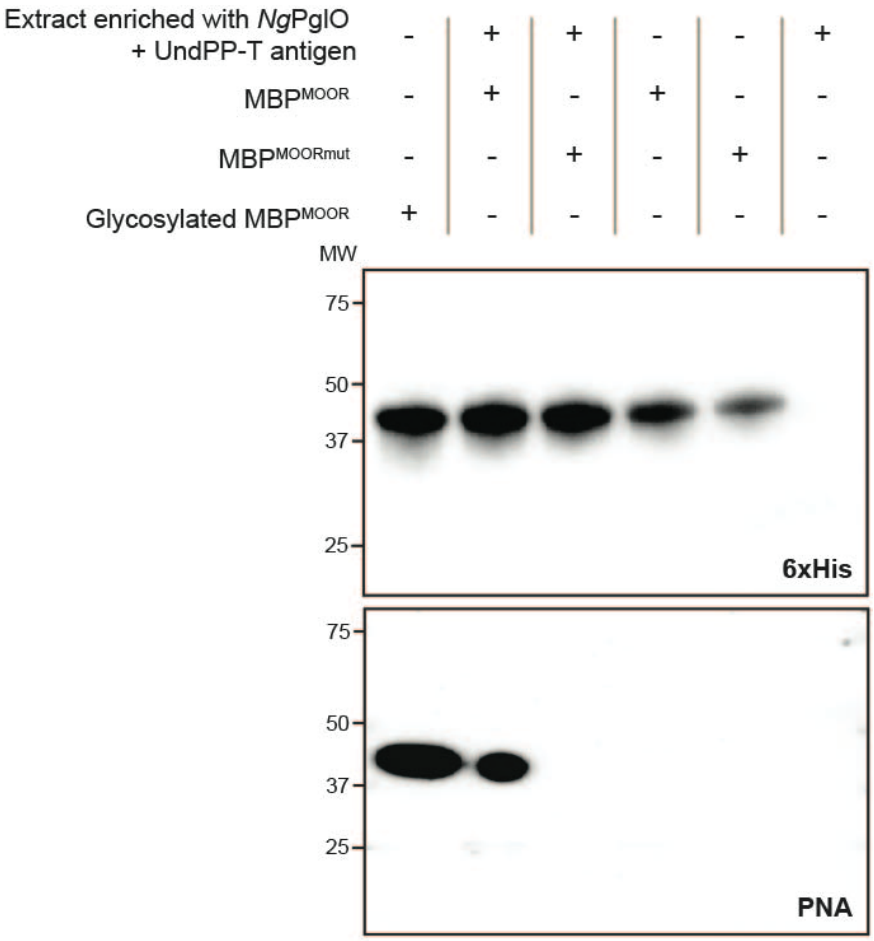


Figure 3-7: In vitro O-glycosylation with T antigen.
Immunoblot analysis of Ni-NTA purified proteins, from in vitro glycosylation reactions, probed with anti-his antibody against the C-terminal polyhistidine tag on acceptor proteins, and PNA lectin reactive to the T antigen disaccharide. Constituents of each reaction is listed above the blots, and molecular weight markers (MW) on the left. MBP^{MOOR}, $MBP^{MOORmut}$, and in vivo glycosylated MBP^{MOOR} are purified from cells. Extract enriched with NgPglO and UndPP-GalNAc-Gal are derived from cell lysate.

abrogated glycosylation. Therefore, this demonstrates that the engineered *O*-glycosylation pathway effectively produces glycoprotein products both in the context of *in vivo* and *in vitro* assays.

3.4 Discussion

We have here engineered a truly bacterial *O*-linked glycosylation pathway to express authentic human glycotopes bearing mucin type core 1 glycans. This platform proves to be versatile in glycosylating a diverse range of target proteins, bearing both bacterial and human derived *O*-glycosites, in terminal as well as internal locations on the target protein, and with both the Tn and T antigens. Specifically, the ability to glycosylate serine amino acids, truly expands the glycoengineering toolkit accessible through *E. coli.*

The orthologous pathway begins with the *in vivo* synthesis of UndPP-GalNAc from substrates UndP and UDP-GalNAc, through a unique enzymatic process unavailable in *E. coli*, this far. This in itself promises to spark the expression of a wide range of glycoconjugates that bear a GalNAc sugar in the reducing end. Next, we adapt flow cytometry to quantify the activity of a broad panel of GalTs from different organisms in high-throughput, assaying their ability to extend UndPP-GalNAc by a single galactose sugar. While flow cytometry has been used to engineer glycosyl-transferases(*49*) and whole glycosylation pathways(*24*) in *E. coli*, this extends the facile platform to focus on a specific class of GalTs that can catalyze the transfer of galactose onto a UndPP-GalNAc substrate, of relevance to efforts expanding *E. coli* for the synthesis of relevant bacterial and eukaryotic polysaccharides. Finally, once the UndPP-GalNAcβ(1-3) Gal lipid-linked oligosaccharide is expressed, we demonstrate its

transfer to protein targets using bacterial *O*-oligosaccharyl transferases, PglO from *Neisseria gonorrhoeae*(*9*) and PglL from *Neisseria meningitidis*(*10*).

The scope of any heterologous glycosylation system is limited by the range of acceptor proteins that it can target. Therefore, the ability to glycosylated 7 different carrier proteins, bearing the engineered MOOR glycosylation tag is valuable. Further, that *Ng*PglO and *Nm*PglL can glycosylate both the C-terminal and internally engineered tags is important. However, the true test of an assembled platform is the ability to glycosylate native glycotopes, authentic amino acid sequences derived from naturally glycosylated proteins of relevance. For instance, in the more well studied realm of bacterial *N*-glycoengineering, for many years until recently, expression of *N*-glycoproteins *via E. coli* was limited to bacterial and other engineered glycosylation motifs(*31, 50*). Here, we demonstrate that, despite previous skepticism(*7*), *Ng*PglO is able to glycosylate three human *O*-glycosites, that naturally bear the core 1 glycan, derived from EPO, GPC and the ectodomain of MUC1. Finally, both *Ng*PglO and *Nm*PglL are able to glycosylate authentic immunologically relevant epitopes derived from the tumor-associated VNTR sequence of human MUC1. Noteworthy, is that the epitopes glycosylated in this study are commonly found reactive towards IgG, IgM and human MHC class I molecules(*43, 44*).

Along the way, this work extends our understanding of the glycan and especially amino acid substrate specificities of these bacterial *O*-OSTs. Contrary to our previous understanding(*13*) we find that both enzymes are able to catalyze the transfer of glycans even in the absence of engineered hydrophilic flanking regions inspired by their native substrates. Further, the glycosylation of acceptor proteins fused to *E. coli*'s native secretory protein, YebF, promises to enable the extensive study of acceptor sequence

specificity of these promiscuous *O*-OSTs, and their evolution to target motifs hitherto inaccessible to bacterial *O*-glycoengineering.

3.5 Materials and Methods

3.5.1. Bacterial strains and growth conditions

E. coli strain DH5α was used for cloning and maintenance of plasmids. Strain W3110 *ΔwecA ΔwaaL* was constructed by deleting *wecA* from CLM24(*51*) through P1 phage transduction where strain JW3758-2(*Δrfe-735*::kan) from the Keio collection(*52*) was used as the donor. MC4100 *ΔwecA* and MC4100 *ΔwecA ΔwaaL* were used as the hosts towards flow cytometry. The latter strain was similarly generated from a lab stock of MC4100 *ΔwecA*, deleting *waaL* using P1 phage transduction where JW3597-1(*ΔrfaL734::kan*) from the Keio collection(*52*) was used as the donor. In both cases, the linked Kan resistance cassette was removed using the method prescribed by Datsenko and Wanner, by transforming with temperature-sensitive plasmid pCP20(*51*). Cultures were grown at 37 °C in LB containing 20 μg/ml chloramphenicol (Cm) and 100 μg/ml ampicillin (Amp) where required. 0.2% (w/v) D-glucose was added to cultures and plates unless when induced for protein production. Induction was always at mid-log phase with 0.1 mM isopropyl β-D-thiogalactoside (IPTG) and 0.2% (w/v) L-arabinose at 16 °C for 16 - 20 h.

3.5.2. Plasmid construction

All plasmids used in this study were made using standard cloning protocols, confirmed by DNA sequencing, and are listed in ***Table 3-1***. Genes coding for *Nm*PglL and *Ng*PglO were designed by reverse translating their amino acid sequence, codon optimized for expression in *E. coli* and synthesized through TWIST Biosciences with overlapping regions to facilitate recombination. The gene fragments were then

amplified by PCR to have overlapping ends and recombined with linearized vector pMW07-T using a modified lazy bones protocol(*54*). Briefly, 0.5 ml of an overnight yeast culture was pelleted and washed in sterile TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA). 0.4 mg of salmon sperm carrier DNA (Sigma), plasmid DNA and PCR products were added to the pellet along with 0.5 ml 'lazy bones' solution (40% polyethylene glycol MW 3350, 0.1 M lithium acetate, 10 mM Tris-HCl pH 7.5 and 1 mM EDTA). After vortexing for 1 min, this solution was incubated up to 4 d at room temperature. Cells were heat-shocked at 42 °C, pelleted and plated on selective medium. The pOG backbone is derived from pMW07(*25*) where the yeast genetic elements, namely origin and URA3 loci are deleted. Genes coding for the Gne epimerase and PglC glycosyltransferase are PCR amplified to have overlapping regions and cloned by Gibson Assembly using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB HiFi) to generate the pOG-PglC backbone. GalTs tested in this study were designed by reverse translating their respective amino acid sequences, codon optimized for expression in *E. coli* and synthesized through TWIST Biosciences with overlapping regions to facilitate recombination. These genes were then amplified by PCR and cloned into pOG-PglC by Gibson Assembly (NEB HiFi).

All acceptor proteins were cloned in a lab stock of pEXT20(*53*). Briefly, *E. coli*'s native maltose binding protein (MBP) was PCR amplified to contain a leader signal peptide, *sp*DsbA, for periplasmic secretion, and cloned into pEXT20 using restriction cloning between the *EcoRI* and *XbaI* sites. The MOOR sequence(*13*) was synthesized as a G block with a polyhistidine epitope tag at its C terminus and cloned between the *XbaI* and *HindIII* sites. All other acceptor proteins were synthesized by Integrate DNA Technologies as G blocks and cloned to replace MBP by Gibson Assembly (NEB HiFi) using the *EcoRI* and *XbaI* sites to linearize the backbone. All acceptor peptides were

synthesized by Integrate DNA Technologies as G blocks and cloned to replace the MOOR sequence by Gibson Assembly (NEB HiFi) using the *XbaI* and *HindIII* sites to linearize the backbone.

Plasmids	Description	Inducer	A/B^R	Reference
p-OG	Modified pMW07 plasmid without the yeast origin and URA3 selection marker	Ara	Cm	Lab stock
pOG-PglC	p-OG encoding Gne from *Campylobacter jejuni* and PglC from *Acinetobacter baumannii*	Ara	Cm	This study
pOG-*HsC1*GalT1	pOG-PglC encoding glycoprotein-N-acetylgalactosamine 3- β -galactosyltransferase 1 derived from *Homo sapiens,* cloned without the first 29 amino acids	Ara	Cm	This study
pOG-*DmC1*GalT2	pOG-PglC encoding glycoprotein-N-acetylgalactosamine 3- β -galactosyltransferase A, isoform B derived from *Drosophila melanogaster,* cloned without the first 50 amino acids	Ara	Cm	This study
pOG-BiGalHexNAcP	pOG-PglC encoding D-galactosyl-β1-3-N-acetyl-D-hexosamine phosphorylase derived from *Bifidobacterium longum subspecies Infantis* ATCC 15697	Ara	Cm	This study
pOG-*Cj*CgtB	pOG-PglC encoding mutant S42 of the β1-3-galactosyltransferase from *Campylobacter jejuni*	Ara	Cm	This study
pOG-*Ec*WbnJ	pOG-PglC encoding 1,3-α- N-acetylgalactosamine -diphosphoundecaprenol β-1,3-galactosyltransferase derived from *Escherichia coli* O86	Ara	Cm	This study
pOG-*Ec*WbwC	pOG-PglC encoding N-acetylgalactosamine -diphosphoundecaprenol β1,3-galactosyltransferases derived from *Escherichia coli* O104	Ara	Cm	This study
pMW07-T	pMW07 plasmid encoding Gne from *Campylobacter jejuni,* PglC from *Acinetobacter baumannii* and WbwC from Escherichia coli O104	Ara	Cm	Lab stock
pMW07-T::Ng	pMW07-T encoding PglO from *Neisseria gonorrhea*	Ara	Cm	This study
pMW07-T::Nm	pMW07-T encoding PglL from *Neisseria meningitidis*	Ara	Cm	This study

pMW07-Tn::Ng	pMW07 plasmid encoding Gne from *Campylobacter jejuni,* PglC from *Acinetobacter baumannii* and PglO from *Neisseria gonorrhea*	Ara	Cm	This study
pMW07-Tn::Nm	pMW07 plasmid encoding Gne from *Campylobacter jejuni,* PglC from *Acinetobacter baumannii* and PglL from *Neisseria meningitidis*	Ara	Cm	This study
pEXT20	Expression vector	IPTG	Amp	(*53*)
pEXT20-MBP^{MOOR}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing the 25 amino acid MOOR sequence	IPTG	Amp	This study
pEXT20-$MBP^{MOORmut}$	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an spDsbA leader peptide for periplasmic secretion and a C terminal fusion bearing the 25 amino acid MOOR sequence with the serine amino acid mutated to a glycine	IPTG	Amp	This study
pEXT20-MBP^{MOOR}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing the 25 amino acid MOOR sequence	IPTG	Amp	This study
pEXT20-YebF-MBP^{MOOR}	pEXT20 encoding the YebF from *Escherichia coli* with an *sp*DsbA leader peptide for extracellular secretion and a C terminal fusion with MBP and the 25 amino acid MOOR sequence	IPTG	Amp	This study
pEXT20-GST^{MOOR}	pEXT20 encoding the glutathione-S-transferase from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing the 25 amino acid MOOR sequence	IPTG	Amp	This study
pEXT20-scFv13-$R4^{MOOR}$	pEXT20 encoding a single-chain antibody fragment against β-galactosidase with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing the 25 amino acid MOOR sequence	IPTG	Amp	This study
pEXT20-$sfGFP^{MOOR}$	pEXT20 encoding the superfolder variant of the green fluorescent protein with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal	IPTG	Amp	This study

	fusion bearing the 25 amino acid MOOR sequence			
pEXT20-sfGFP$^{Q157\text{-}MOOR}$	pEXT20 encoding the superfolder variant of the green fluorescent protein with an *sp*DsbA leader peptide for periplasmic secretion and an internal 25 amino acid MOOR sequence at the amino acid Q157	IPTG	Amp	This study
pEXT20-CRM197^{MOOR}	pEXT20 encoding the cross-reacting material 197 from *Corynebacterium diphtheriae* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing the 25 amino acid MOOR sequence	IPTG	Amp	This study
pEXT20-PD^{MOOR}	pEXT20 encoding the Protein D from *Haemophilus influenzae* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing the 25 amino acid MOOR sequence	IPTG	Amp	This study
pEXT20-MBP^{EPO}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing an 8 amino acid motif derived from Erythropoietin flanked by the MOOR flanking sequence	IPTG	Amp	This study
pEXT20-MBP^{GPC}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing an 8 amino acid motif derived from Glycophorin C flanked by the MOOR flanking sequence	IPTG	Amp	This study
pEXT20-MBP^{MUC1_8}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing an 8 amino acid motif derived from the VNTR region of MUC1 flanked by the MOOR flanking sequence	IPTG	Amp	This study
pEXT20-MBP^{MUC1_12}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing a 12 amino acid motif derived	IPTG	Amp	This study

	from the VNTR region of MUC1 flanked by the MOOR flanking sequence			
pEXT20-MBP^{MUC1_16}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing a 16 amino acid motif derived from the VNTR region of MUC1 flanked by the MOOR flanking sequence	IPTG	Amp	This study
pEXT20-MBP^{MUC1_20}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing a 20 amino acid motif derived from the VNTR region of MUC1 flanked by the MOOR flanking sequence	IPTG	Amp	This study
pEXT20-MBP^{MUC1_24}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion bearing a 24 amino acid motif derived from the VNTR region of MUC1 flanked by the MOOR flanking sequence	IPTG	Amp	This study
pEXT20-MBP^{MUC1_41}	pEXT20 encoding the maltose binding protein from *Escherichia coli* with an *sp*DsbA leader peptide for periplasmic secretion and a C terminal fusion a 41 amino acid peptide derived from the VNTR region of the human MUC1 protein	IPTG	Amp	This study

Table 3-1: Plasmids used in the current study.
A/B^R refers to antibiotic resistance, Ara for L-Arabinose, IPTG for isopropyl β-D-thiogalactoside, Amp for Ampicillin and Cm for Chloramphenicol.

3.5.3. Protein analysis

Glycoprotein production was induced in 150 ml cultures in Lysogeny Broth medium (LB) for 16–20 h. Cells were pelleted at 8,000 rpm for 30 minutes at 4 °C, resuspended in 2 ml of lysis buffer containing 50 mM sodium phosphate, 300 mM

sodium chloride, and 10 mM imidazole. Samples were frozen at -80 °C overnight. Cells were then thawed, gently agitated at room temperature with 200 μg/ml of lysozyme (Sigma) for 15 minutes, and lysed by sonication. Lysed samples were then centrifuged at 8,000 rpm for 30 minutes at 4 °C and the supernatant was subject to Ni affinity purification using Qiagen's Ni-NTA spin columns in accordance with the manufacturer's protocol.

Purified protein samples were prepared in Invitrogen LDS Sample Buffer (Thermo Scientific) and resolved on Bolt SDS-polyacrylamide gels (Thermo Scientific). Following electrophoresis, proteins were transferred onto Immobilon-P polyvinylidene difluoride (PVDF) membranes (0.45 μm) according to manufacturer's protocol (Thermo Scientific). Western blotting used polyhistidine epitope tag–specific polyclonal antibodies (Abcam), the anti-Muc1 antibody (HMPV; BD Bioscience), or appropriate lectins (Vector labs) used as per suggested protocol. Lectins used were biotinylated and required a secondary Extravidin reagent conjugated to HRP (Sigma). Bio-Rad enhanced chemiluminescent (ECL) substrate was used for detection. All blots were visualized using a Chemidoc XRS+ system with Image Lab image capture software (Bio-Rad).

3.5.4. In vitro glycoprotein synthesis

Acceptor proteins, MBP^{MOOR} and $MBP^{MOORmut}$ were purified from 500 mL culture of BL21(DE3) cells using standard Ni affinity chromatography (Ni-NTA, ThermoFisher). Solubilized membrane extracts enriched with *Ng*PglO and LLOs containing UndPP-GalNAc-β(1-3)-Gal was performed as follows. CLM24 cells carrying pMW07-T, *Ng*PglO plasmid were grown overnight at 37 °C in LB media. The following day, cells were subcultured into 4.0 L LB media and allowed to grow at 37

°C until mid-log phase (Abs600 ~0.6). Culture was then induced for 20 h at 16 °C with 0.2% L-arabinose. Cells were harvested by centrifugation at 10,000 × g for 30 min at 4 °C, and then resuspended in buffer containing 50 mM Tris-HCl (pH 8.0) and 25 mM sodium chloride. Cells were lysed by passing cell suspension through a high-pressure homogenizer (Avestin) five times and the resulting lysate was centrifuged at 15,000 × g for 20 min at 4 °C. The supernatant was collected and subject to ultracentrifuge at 100,000 × g for 2 h at 4 °C. The resulting pellet corresponding to the membrane fraction was collected and resuspended in 3 mL of buffer containing 50 mM Tris-HCl (pH 7.0), 25 mM sodium chloride, and 0.1% (w/v) n-dodecyl-β-D-maltoside (DDM). Suspension and incubated with mild agitation at room temperature for 1 h, to enable the solubilization of *Ng*PglO and LLOs. Following incubation, the mixture was centrifuged at 16,000 × g for 1 h at 4 °C, and the supernatant was retained as a crude membrane extract. *In vitro* glycosylation of purified acceptors was carried out in 1.5 mL reactions containing 50 µg of purified acceptor protein and 1 mL of crude membrane extracted in reaction buffer containing 10 mM HEPES (pH 7.5), 10 mM manganese chloride, and 1% (w/v) DDM. The reaction was incubated at 30 °C for 16 h with mild tumbling. Upon reaction completion, acceptor proteins were purified from reaction mixture using Ni-NTA spin column (Qiagen).

3.5.5. Flow cytometric analysis

Flow cytometric analysis of GalT activity performed as previously described(*24*). Briefly, overnight cultures of each strain were grown in LB with relevant antibiotics. Cells were subcultured to an Abs600 of 0.1 in 10 ml LB and grown overnight at 30 °C. The next day, 1 ml of overnight culture was washed twice with 1 ml of PBS and resuspended in 500 µl PBS. All samples were diluted to an Abs600 of 0.2 in 250 µl PBS. Detection of the disaccharide T antigen was performed with peanut agglutinin

lectin (PNA) conjugated to fluorescein isothiocyanate (FITC) (Vector labs). PNA::FITC was diluted 1:500 in PBS, and 250 µl of diluted lectin was added to cells followed by incubation at 37 °C for 30 min. Cells were pelleted at 6000×g for 4 min, washed in 1 ml PBS, resuspended in 1 ml PBS and analyzed by flow cytometry using a FACSCalibur flow cytometer (BD Biosciences). All experiments were performed in triplicate with the resulting data generated through CellQuest Pro 6.0 and analyzed on FlowJo 10.5 software.

3.6 References

1. C. M. Harding, M. F. Feldman, **Glycoengineering bioconjugate vaccines, therapeutics, and diagnostics in E. coli.** *Glycobiology* 29, 519-529 (2019).
2. A. Natarajan, T. Jaroentomeechai, M. Li, C. J. Glasscock, M. P. DeLisa, **Metabolic engineering of glycoprotein synthesis in bacteria.** *Emerg Top Life Sci* 2, 419-432 (2018).
3. A. Varki, **Biological roles of glycans.** *Glycobiology* 27, 3-49 (2017).
4. N. Jentoft, **Why are proteins O-glycosylated?** *Trends in Biochemical Sciences* 15, 291-294 (1990).
5. K. G. Ten Hagen, T. A. Fritz, L. A. Tabak, **All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases.** *Glycobiology* 13, 1R-16R (2003).
6. T. A. Gerken, J. Raman, T. A. Fritz, O. Jamison, **Identification of Common and Unique Peptide Substrate Preferences for the UDP-GalNAc:Polypeptide α-N-acetylgalactosaminyltransferases T1 and T2 Derived from Oriented Random Peptide Substrates.** *Journal of Biological Chemistry* 281, 32403-32416 (2006).

7. T. Du *et al.*, **A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans.** *Cell Chemical Biology* 26, 203-212.e205 (2019).

8. G. E. Henderson, K. D. Isett, T. U. Gerngross, **Site-specific modification of recombinant proteins: a novel platform for modifying glycoproteins expressed in E. coli.** *Bioconjug Chem* 22, 903-912 (2011).

9. M. D. Hartley *et al.*, **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 (2011).

10. P. M. Power, K. L. Seib, M. P. Jennings, **Pilin glycosylation in Neisseria meningitidis occurs by a similar pathway to wzy-dependent O-antigen biosynthesis in Escherichia coli.** *Biochem Biophys Res Commun* 347, 904-908 (2006).

11. A. Faridmoayer *et al.*, **Extreme substrate promiscuity of the Neisseria oligosaccharyl transferase involved in protein O-glycosylation.** *J Biol Chem* 283, 34596-34604 (2008).

12. F.-G. Hanisch, S. Müller, **MUC1: the polymorphic appearance of a human mucin.** *Glycobiology* 10, 439-449 (2000).

13. C. Pan *et al.*, **Biosynthesis of conjugate vaccines using an O-linked glycosylation system.** *MBio* 7, e00443-00416 (2016).

14. J. Lehrer, K. A. Vigeant, L. D. Tatar, M. A. Valvano, **Functional characterization and membrane topology of Escherichia coli WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide.** *J Bacteriol* 189, 2618-2628 (2007).

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

16. S. Bernatchez *et al.*, **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 (2005).

17. C. M. Harding, M. F. Haurat, E. Vinogradov, M. F. Feldman, **Distinct amino acid residues confer one of three UDP-sugar substrate specificities in Acinetobacter baumannii PglC phosphoglycosyltransferases.** *Glycobiology* 28, 522-533 (2018).

18. T. Ju, K. Brewer, A. D'Souza, R. D. Cummings, W. M. Canfield, **Cloning and Expression of Human Core 1 β1,3-Galactosyltransferase.** *Journal of Biological Chemistry* 277, 178-186 (2002).

19. R. Muller *et al.*, **Characterization of mucin-type core-1 beta1-3 galactosyltransferase homologous enzymes in Drosophila melanogaster.** *Febs j* 272, 4295-4305 (2005).

20. H. Yu *et al.*, **Highly efficient chemoenzymatic synthesis of β1-3-linked galactosides.** *Chemical communications (Cambridge, England)* 46, 7507-7509 (2010).

21. G. Yang *et al.*, **Fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases.** *J Am Chem Soc* 132, 10570-10577 (2010).

22. W. Yi *et al.*, **Escherichia coli O86 O-Antigen Biosynthetic Gene Cluster and Stepwise Enzymatic Synthesis of Human Blood Group B Antigen Tetrasaccharide.** *Journal of the American Chemical Society* 127, 2040-2041 (2005).

23. S. Wang *et al.*, **Characterization of Two UDP-Gal:GalNAc-Diphosphate-Lipid β1,3-Galactosyltransferases WbwC from <span class="named-**

content genus-species" id="named-content-1">Escherichia coli Serotypes O104 and O5. *J Bacteriol* 196, 3122-3133 (2014).

24. C. J. Glasscock *et al.*, **A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation.** *Metab Eng* 47, 488-495 (2018).

25. J. D. Valderrama-Rincon *et al.*, **An engineered eukaryotic protein glycosylation pathway in Escherichia coli.** *Nat Chem Biol* 8, 434 (2012).

26. G. F. Springer, P. R. Desai, **Common precursors of human blood group MN specificities.** *Biochemical and Biophysical Research Communications* 61, 470-475 (1974).

27. J. A. Iwashkiw, N. F. Vozza, R. L. Kinsella, M. F. Feldman, **Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation.** *Mol Microbiol* 89, 14-28 (2013).

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

29. L. Grubhoffer, M. Ticha, J. Kocourek, **Isolation and properties of a lectin from the seeds of hairy vetch (Vicia villosa Roth).** *Biochem J* 195, 623-626 (1981).

30. A. A. Ollis, Y. Chai, M. P. DeLisa, **GlycoSNAP: A High-Throughput Screening Methodology for Engineering Designer Glycosylation Enzymes.** *Methods Mol Biol* 1321, 37-47 (2015).

31. A. C. Fisher *et al.*, **Production of secretory and extracellular N-linked glycoproteins in Escherichia coli.** *Appl Environ Microbiol* 77, 871-881 (2011).

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

33. **Detection of protein-protein interactions using the GST fusion protein pull-down technique.** *Nature Methods* 1, 275-276 (2004).

34. P. Martineau, P. Jones, G. Winter, **Expression of an antibody fragment at high levels in the bacterial cytoplasm11Edited by J. Karn.** *Journal of Molecular Biology* 280, 117-127 (1998).

35. J.-D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, **Engineering and characterization of a superfolder green fluorescent protein.** *Nat Biotechnol* 24, 79-88 (2006).

36. E. Kay, J. Cuccui, B. W. Wren, **Recent advances in the production of recombinant glycoconjugate vaccines.** *NPJ Vaccines* 4, 16 (2019).

37. F. Micoli, R. Adamo, P. Costantino, **Protein Carriers for Glycoconjugate Vaccines: History, Selection Criteria, Characterization and New Trends.** *Molecules (Basel, Switzerland)* 23, 1451 (2018).

38. P. H. Lai, R. Everett, F. F. Wang, T. Arakawa, E. Goldwasser, **Structural characterization of human erythropoietin.** *Journal of Biological Chemistry* 261, 3116-3121 (1986).

39. A. G. Maier *et al.*, **Plasmodium falciparum erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations.** *Nat Med* 9, 87-92 (2003).

40. M. A. Tarp, H. Clausen, **Mucin-type O-glycosylation and its potential use in drug and vaccine development.** *Biochim Biophys Acta* 1780, 546-563 (2008).

41. S. Gendler, J. Taylor-Papadimitriou, T. Duhig, J. Rothbard, J. Burchell, **A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats.** *Journal of Biological Chemistry* 263, 12820-12823 (1988).

42. C. Fu *et al.*, **Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen.** *Hla* 88, 275-286 (2016).

43. S. von Mensdorff-Pouilly *et al.*, **Reactivity of natural and induced human antibodies to MUC1 mucin with MUC1 peptides and n-acetylgalactosamine (GalNAc) peptides.** *International Journal of Cancer* 86, 702-712 (2000).

44. V. Apostolopoulos, V. Karanikas, J. S. Haurum, I. F. McKenzie, **Induction of HLA-A2-restricted CTLs to the mucin 1 human breast cancer antigen.** *The Journal of Immunology* 159, 5211-5218 (1997).

45. I. Keydar *et al.*, **Production and characterization of monoclonal antibodies identifying breast tumor-associated antigens.** *Proceedings of the National Academy of Sciences* 86, 1362-1366 (1989).

46. J. Kyte, R. F. Doolittle, **A simple method for displaying the hydropathic character of a protein.** *J Mol Biol* 157, 105-132 (1982).

47. J. M. Walker, SpringerLink (Online service). (Humana Press,, Totowa, N.J., 2005), pp. xviii, 988 p.

48. T. Jaroentomeechai *et al.*, **Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery.** *Nat Commun* 9, 2686 (2018).

49. A. Aharoni *et al.*, **High-throughput screening methodology for the directed evolution of glycosyltransferases.** *Nat Methods* 3, 609-614 (2006).

50. M. Kowarik *et al.*, **N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase.** *Science* 314, 1148-1150 (2006).

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

52. T. Baba *et al.*, **Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.** *Mol Syst Biol* 2, 2006.0008 (2006).

53. D. M. Dykxhoorn, R. St. Pierre, T. Linn, **A set of compatible tac promoter expression vectors.** *Gene* 177, 133-136 (1996).

54. R. M. Shanks, N. C. Caiazza, S. M. Hinsa, C. M. Toutain, G. A. O'Toole, **Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria.** *Appl Environ Microbiol* 72, 5027-5036 (2006).

CHAPTER 4

An engineered survival-selection assay for extracellular protein expression uncovers hypersecretory phenotypes in Escherichia coli[1]

4.1 Abstract

The extracellular expression of recombinant proteins using laboratory strains of *Escherichia coli* is now routinely achieved using naturally secreted substrates, such as YebF or the osmotically inducible protein Y (OsmY), as carrier molecules. However, secretion efficiency through these pathways needs to be improved for most synthetic biology and metabolic engineering applications. To address this challenge, we developed a generalizable survival-based selection strategy that effectively couples extracellular protein secretion to antibiotic resistance and enables facile isolation of rare

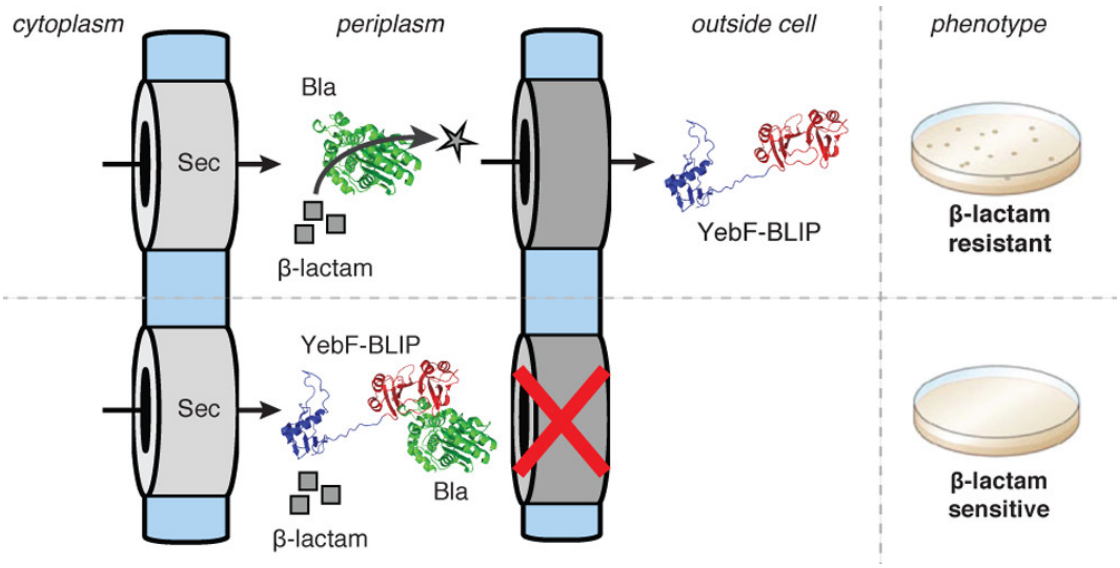


Figure 4-1: Graphical abstract illustrating the survival-selection scheme.

mutants from very large populations (i.e., 10^{10-12} clones) based simply on cell growth. Using this strategy in the context of the YebF pathway, a comprehensive library of *E. coli* single-gene knockout mutants was screened and several gain-of-function mutations

[1] *This chapter has been previously published:* Natarajan A., Jaroentomeechai T., Li M., Glasscock C.J., and DeLisa M.P. (2018). Metabolic engineering of glycoprotein biosynthesis in bacteria. Emerging Topics in Life Sciences, 2 (3) 419-432. DOI: 10.1042/ETLS20180004. URL: http://www.emergtoplifesci.org/content/2/3/419

were isolated that increased the efficiency of extracellular expression without compromising the integrity of the outer membrane. We anticipate that this user-friendly strategy could be leveraged to better understand the YebF pathway and other secretory mechanisms - enabling the exploration of protein secretion in pathogenesis as well as the creation of designer *E. coli* strains with greatly expanded secretomes - all without the need for expensive exogenous reagents, assay instruments, or robotic automation.

4.2 Introduction

Nonpathogenic strains of *Escherichia coli* are widely used for laboratory and industrial-scale biosynthesis of recombinant proteins, with expression usually taking place in either the cytoplasmic or periplasmic compartment(*1-4*). However, the inside of a bacterial cell is highly crowded with macromolecules, often reaching a concentration of 300−400 mg/mL(*5, 6*). As a result, many recombinant proteins misfold, aggregate, and form inclusion bodies, which require expensive, labor-intensive denaturation/refolding processes to obtain biologically active proteins(*1-4, 7*). Furthermore, intracellular product recovery is nontrivial given the large number of host proteins that accumulate in the cell alongside the protein of interest as well as host proteases that can degrade the protein product.

Expression of recombinant proteins in the extracellular medium, on the other hand, affords an opportunity to overcome these issues because the concentration of deleterious proteases and other contaminating cellular components, in particular lipopolysaccharide (endotoxin), is very low(*2*). Moreover, by sidestepping the need for cell membrane disruption, purification of extracellular proteins is greatly simplified and thus potentially more economical(*2*). It is also worth mentioning that extracellular protein expression enables certain applications that would otherwise not be possible

such as production of proteins that are cytotoxic to the host or otherwise recalcitrant to intracellular expression(*9*), enzymatic conversion of substrates that are inadequately taken up by the cell such as cellulose for consolidated bioprocessing(*10, 11*) or toxic compounds for bioremediation(*12*), and delivery of therapeutic proteins and peptides for treatment of disease(*13-16*).

However, whereas *E. coli* has been the "workhorse" for making recombinant proteins in the cytoplasmic and periplasmic compartments(*1-4, 7*), it has historically been overlooked for the directed secretion of proteins to the extracellular environment. A primary reason for this is that nonpathogenic laboratory strains of *E. coli* generally express only trace amounts of proteins in the culture medium under normal growth conditions(*17, 18*). Further, while specific transport systems (*e.g.*, type I, II and III secretion systems) from pathogenic bacteria can be functionally reconstituted in *E. coli*(*19, 20*), extracellular secretion through these pathways is often inadequate and efforts to engineer these sophisticated multicomponent systems for high-level protein export into the growth medium have so far met with little success(*3*). One promising alternative involves leveraging small, endogenous proteins, such as YebF and the osmotically inducible protein Y (OsmY), that are naturally secreted by laboratory *E. coli* strains(*21, 22*). Both of these proteins are first localized to the periplasmic compartment *via* the Sec pathway and, in the case of YebF, translocation across the outer membrane appears to involve OmpC and OmpF(*23*). Importantly, both OsmY and YebF have been used as carriers to deliver biotechnologically relevant "passenger" proteins linked to their C-termini into the culture medium without compromising integrity of the outer membrane(*8, 10, 11, 21, 24*). Unfortunately, the extracellular yields of many YebF or OsmY fusion proteins are often too low to be practically useful. For example, the use of engineered *E. coli* as a consolidated bioprocessing platform for

converting xylan to biodiesel was limited by the suboptimal extracellular activity of OsmY-xylanase fusions(*11*).

Recently, we developed a universal fluorescence-based genetic screen that was compatible with the YebF and OsmY mechanisms and allowed microtiter plate-based screening of desirable secretion phenotypes(*8*). In this study, the phenotype of each mutant was determined independently by growing each clone separately in microtiter (96 or 384 well) plates and measuring the fluorescence in each well by labeling of the secreted proteins with a biarsenical fluorescent compound (FlAsH). Unfortunately, the requirement to examine each member of the population severely limits the number of clones that can be assayed - even with the aid of robotic automation - and is a disadvantage of all genetic screens. In contrast, genetic selection bypasses the need to assay thousands of individual isolates, thereby enabling rapid screening of large libraries (10^{10-12} clones, limited mainly by the ability to make such libraries)(*25, 26*) without the need for expensive exogenous reagents, assay instruments, or robotic automation. However, while selection strategies allow the isolation of rare mutants from much larger cell populations, they can only be applied when a mutation confers a growth advantage. This has proven to be a major stumbling block for extracellular protein expression, which has so far eluded adaptation to a generalizable selection strategy owing to technical difficulties associated with linking this phenotype to cell growth.

Here, we engineered a generic survival-selection system that can be used to identify rare secretory phenotypes and to optimize those phenotypes by coupling extracellular secretion activity to the survival or death of the host bacteria. This was accomplished by genetically fusing YebF to the β-lactamase inhibitor protein (BLIP), a chimera that upon secretion out of cells offers a selective growth advantage in the

presence of β- lactam antibiotics and thus allows for the rapid isolation of mutations that enhance extracellular accumulation of YebF. Using this selection, we isolated several gain-of-function mutations from a comprehensive library of *E. coli* single-gene knockout mutants(*27*). These mutants were each confirmed to increase the efficiency of extracellular expression without compromising the integrity of the outer membrane. When secretion-optimized hosts were equipped with asparagine-linked (*N*-linked) glycosylation machinery, higher amounts of glycosylated YebF were detected in the culture medium, revealing a potentially useful compatibility of high-level YebF secretion with intracellular protein modification pathways and thus an opportunity to further expand the *E. coli* secretome.

4.3 Results

4.3.1. An Engineered Selection for Extracellular Protein Expression

We sought to create a genetic selection that can be interfaced with any extracellular secretion pathway and used to isolate desired secretory phenotypes from among a large number of nonsecreting or poorly secreting cells based solely on cell growth. Ideally, such a selection would couple the secretory phenotype with cell survival, while the non-secretory phenotype would be lethal. Our strategy exploited the β-lactamase inhibitor protein (BLIP) from *Streptomyces clavuligerus*, which is a potent inhibitor of a variety of β-lactamases (Bla) with Ki values in the sub-nanomolar to picomolar range(*28, 29*). Periplasmic expression of BLIP in *E. coli* has been shown to inhibit Bla activity in this compartment, thereby rendering cells susceptible to β-lactam antibiotics(*30, 31*). This β- lactam sensitive phenotype was recreated here through the pTrc-BLIP plasmid ***(Figure 4-2 a)***, a pTrc99A backbone encoding: (i) a re-engineered

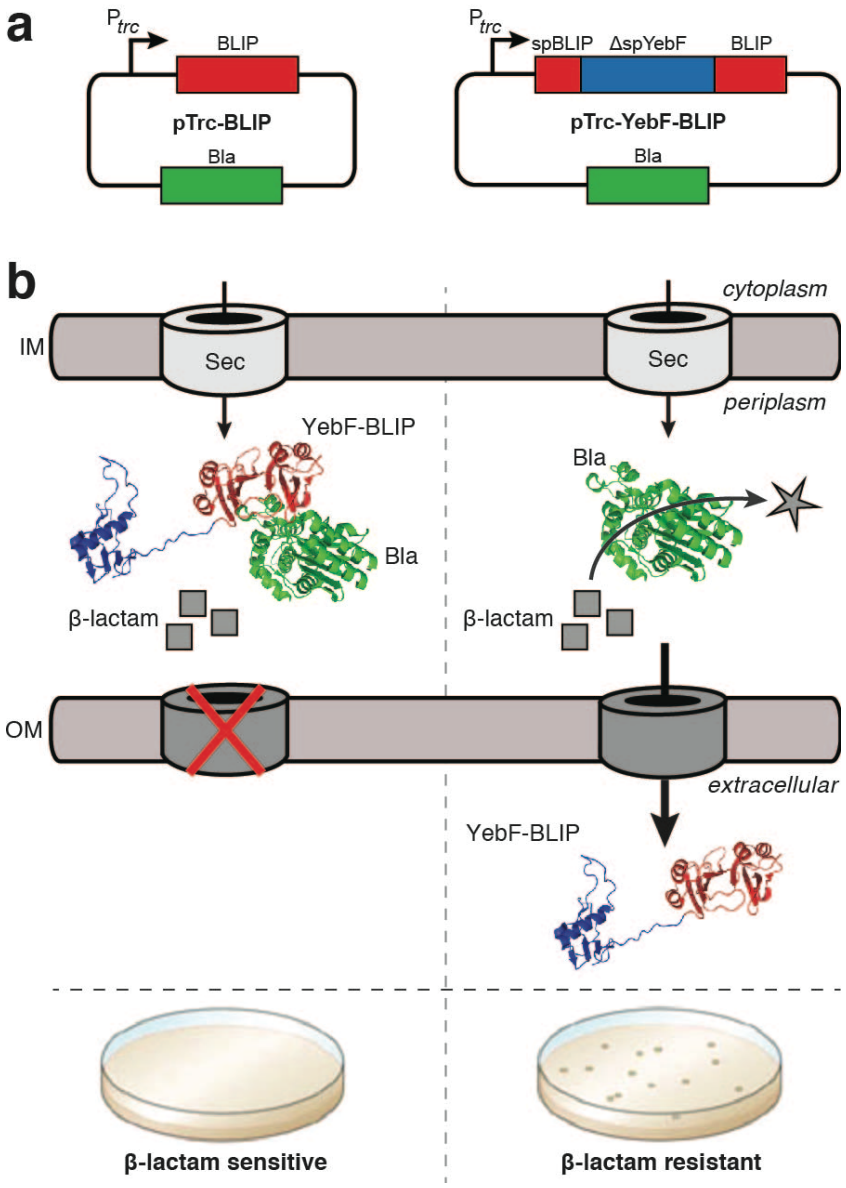


Figure 4-2: An engineered survival-selection assay for extracellular expression of proteins.

(a) Schematic of the pTrc-BLIP plasmid, encoding full-length BLIP, and pTrc-YebF-BLIP, encoding the YebF-BLIP chimera. Both are based on plasmid pTrc99A, which expresses TEM -1 Bla under the control of a constitutive promoter. (b) Schematic of the survival-selection strategy. Briefly, retention of YebF-BLIP in the periplasm due to inefficient or blocked extracellular secretion inhibits Bla activity and renders cells sensitive to β-lactam antibiotics (left panel). Conversely, YebF-mediated delivery of BLIP to the extracellular medium relieves the inhibition of Bla in the periplasm, thereby conferring increased antibiotic resistance to cells (right panel).

BLIP coding gene(*31*) under the control of the trc promoter; and (ii) TEM-1 Bla under

the control of a constitutive promoter. BW25113 *E. coli* cells co-expressing these two proteins were highly sensitive to carbenicillin (Carb) at a concentration of 25 μg/mL ***(Figure 4-3 a)***, demonstrating that periplasmically localized BLIP was a potent inhibitor of Bla activity. The observation that growth was uninhibited when the same cells were plated in the absence of Carb ***(Figure 4-3 a)*** indicated that antibiotic sensitivity was due to Bla inhibition and not an unrelated growth defect. Next, we linked this BLIP-mediated inhibition to extracellular protein expression. This involved construction of plasmid pTrc-YebFBLIP ***(Figure 4-2 a)***, a pTrc99A backbone encoding a chimera comprised of (i) the *N*-terminal Sec signal peptide from BLIP; (ii) a truncated version of *E. coli* YebF lacking its native N-terminal Sec signal peptide; and (iii) a similarly truncated version of BLIP lacking its native N-terminal Sec signal peptide. This YebF-BLIP chimera formed the basis for our survival selection ***(Figure 4-2 b)***, which is rationalized as follows. YebF-mediated delivery of BLIP to the extracellular medium should relieve the inhibition of Bla in the periplasm, thereby conferring cells with increased resistance to β-lactam antibiotics. Conversely, retention of YebF-BLIP in the periplasm due to inefficient or blocked extracellular secretion should inhibit Bla activity and render cells antibiotic sensitive. Indeed, when YebF-BLIP was co-expressed with Bla, BW25113 cells exhibited strong resistance to Carb that was on par with the resistance conferred by Bla expression from empty pTrc99A in the absence of BLIP ***(Figure 4-3 a)***. Control cells expressing ΔspYebF-BLIP, a chimera that was blocked for export out of the cytoplasm due to removal of the N-terminal Sec signal peptide, were similarly resistant to Carb ***(Figure 4-3 a)***.

To determine if these differences in antibiotic resistance were attributable to extracellular localization of the YebF-BLIP protein, cell-free supernatant fractions were harvested from cells and analyzed by immunoblotting. In agreement with the resistant

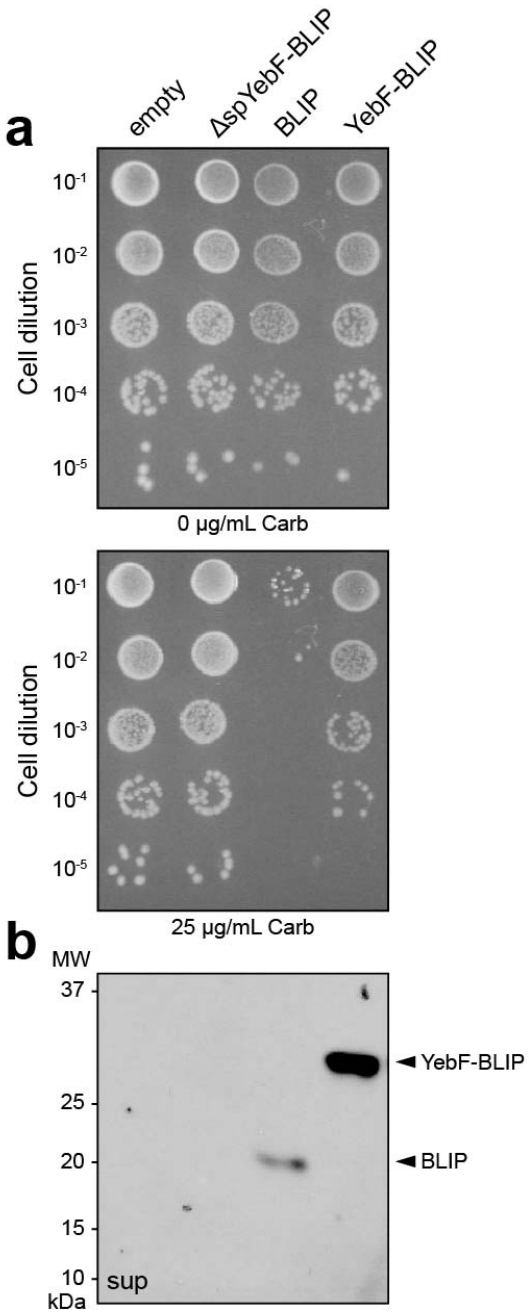


Figure 4-3: Linking extracellular secretion of YebF with resistance to β- lactam antibiotics.

(a) Spot plate analysis of wt BW25113 cells harboring plasmid pTrc99A (empty) or a pTrc99A derivative encoding ΔspYebF-BLIP, full-length BLIP, or YebF-BLIP. Cells were serially diluted on LB agar plates with or without Carb as indicated. (b) Immunoblot analysis of supernatant (sup) fractions isolated from the same cells as in (a). Proteins were detected using an anti6x-His antibody. Molecular weight (MW) marker indicated on the left.

phenotype conferred to cells, the YebF-BLIP chimera accumulated at a high level in the

cell-free supernatant as revealed by a strong cross-reacting band in the immunoblot ***(Figure 4-3 b)***. In contrast, only a very low level of unfused BLIP was detected in the cell-free supernatant ***(Figure 4-3 b)***, consistent with the sensitive phenotype conferred by this construct. As expected, ΔspYebF-BLIP was undetectable in the supernatant fraction ***(Figure 4-3 b)***, confirming that secretion of BLIP to the culture medium relies on both the N-terminal signal peptide and the YebF domain. Collectively, these results demonstrate that the survival-based selection strategy effectively couples extracellular secretion to antibiotic resistance and enables facile discrimination of secretory phenotypes based solely on cell growth.

To determine whether cells expressing the YebF-BLIP chimera could be isolated from a large number of non-secreting cells, we performed mock selection experiments. Overnight cultures of BW25113 cells carrying either pTrc-BLIP or pTrc-YebF-BLIP were mixed in equal amounts and subsequently selected by plating the cell mixture on solid agar with or without antibiotic. A total of 10/10 colonies from the Carb containing plates carried the pTrc-YebF-BLIP plasmid insert, whereas only 3/10 colonies carried this plasmid when selected in the absence of Carb ***(Figure 4-4)***. These results provided proof-of-concept for the use of our selection tool in the routine isolation of rare secretory-competent cells from a large population of non-secretory cells.

4.3.2. Selection of Gain-of-Function Mutations That Increase Host Secretion Capacity.

Encouraged by the mock selection results, we next attempted to use the survival-based selection to isolate gain-of-function mutations that increased the efficiency of extracellular expression. It should be pointed out that isolating gain-of-function mutations involved the more challenging task of enriching cells exhibiting a hyper-

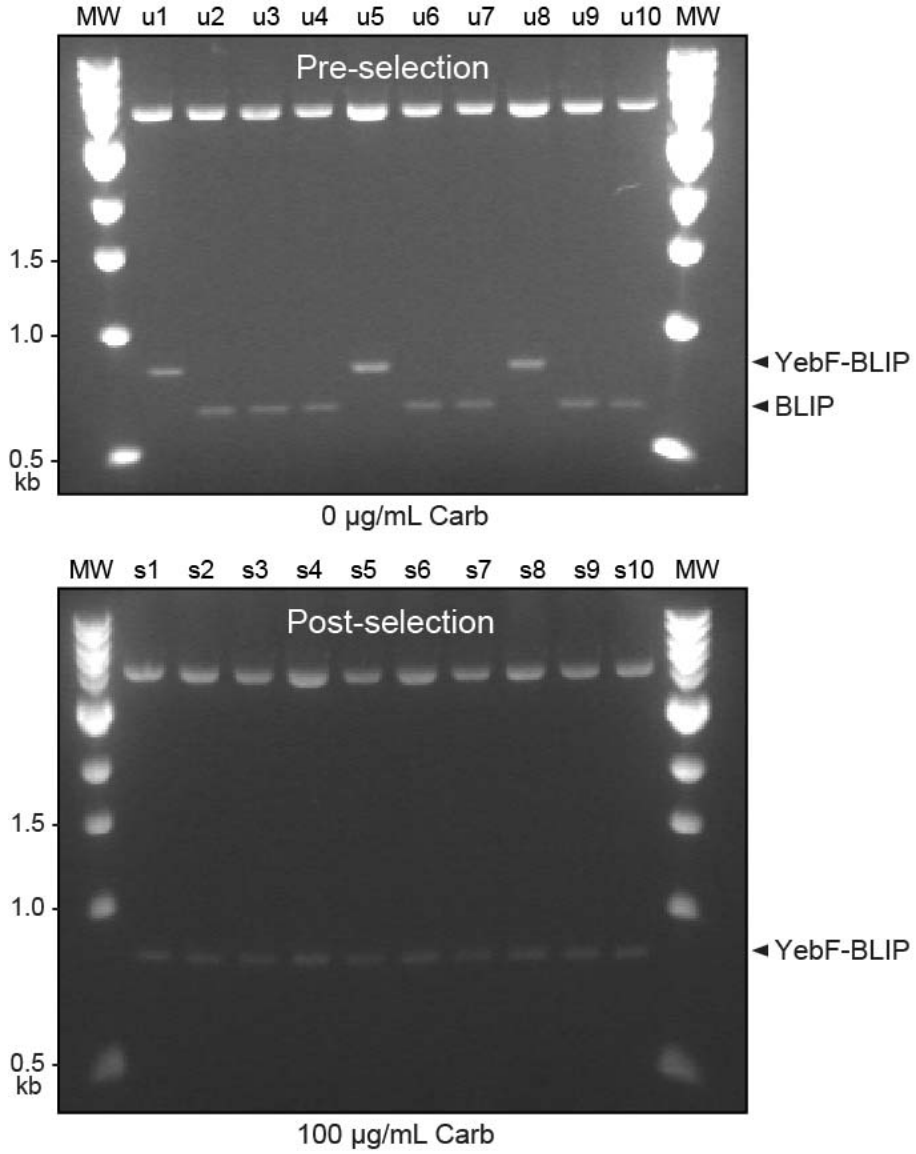


Figure 4-4: Mock library selection for secretion of YebF to the culture medium. *Overnight cultures of BW25113 cells carrying either pTrc-BLIP or pTrc-YebF-BLIP were mixed in equal amounts and subsequently selected by plating the cell mixture on solid LB agar with or without antibiotic as indicated. Plasmids were isolated from ten randomly chosen colonies plated under non-selective conditions (0 g/mL Carb; clones n1-n10) or selective conditions (100 g/mL Carb; clones s1-s10). Plasmids were analyzed by restriction enzyme digestion using EcoRI and HindIII to release either the BLIP or YebF-BLIP insert, with predicted sizes of 672 bp and 948 bp, respectively. Digestion products were then analyzed by agarose gel electrophoresis with molecular weight (MW) ladder run in the first and last lanes of each gel.*

functional phenotype from a large background of cells exhibiting a weaker but still

measurable secretion-positive phenotype. To this end, we hypothesized that selection pressure could be dialed to a level that would permit growth of only the cells exhibiting a hyper-functional phenotype. To test this hypothesis, BW25113 cells expressing BLIP, ΔspYebF-BLIP or YebF-BLIP were plated on a significantly higher concentration of Carb (100 μg/ mL). As was seen at the lower Carb concentration, BW25113 cells expressing BLIP remained antibiotic sensitive due to inhibition of Bla activity while BW25113 cells expressing ΔspYebF-BLIP, which was blocked for export out of the cytoplasm, remained strongly resistant ***(Figure 4-5 a)***. However, the resistance conferred to wild-type (wt) BW25113 cells by YebF-BLIP was markedly reduced at this higher level of Carb, with cell growth barely visible above that of the BLIP-expressing control cells ***(Figure 4-5 a)***. To determine whether increased YebF-BLIP secretion could render cells less susceptible to this level of Carb, we employed a panel of mutant *E. coli* strains that were previously isolated based on their ability to enhance YebF secretion using a plate-based fluorescence screen(*8*). In line with our hypothesis, expression of YebF-BLIP in BW25113 cells lacking *entC*, *entE*, *envZ*, *mzrA*, and *tnaA* resulted in antibiotic resistance that was significantly higher than that observed for wt BW25113 cells expressing YebF-BLIP or BLIP ***(Figure 4-5 a)***. It is also worth mentioning that the relative resistance observed for each mutant agreed qualitatively with the secretion levels measured previously by fluorescent labeling of YebF(*8*) ***(Figure 4-5 b)***. Importantly, these results establish the use of increased selection pressure as a plausible route to isolating gain-of-function mutations.

To isolate new mutants with enhanced secretion activity, we leveraged the Keio collection, a set of 3985 *E. coli* K-12 in-frame, single-gene knockout mutants each with

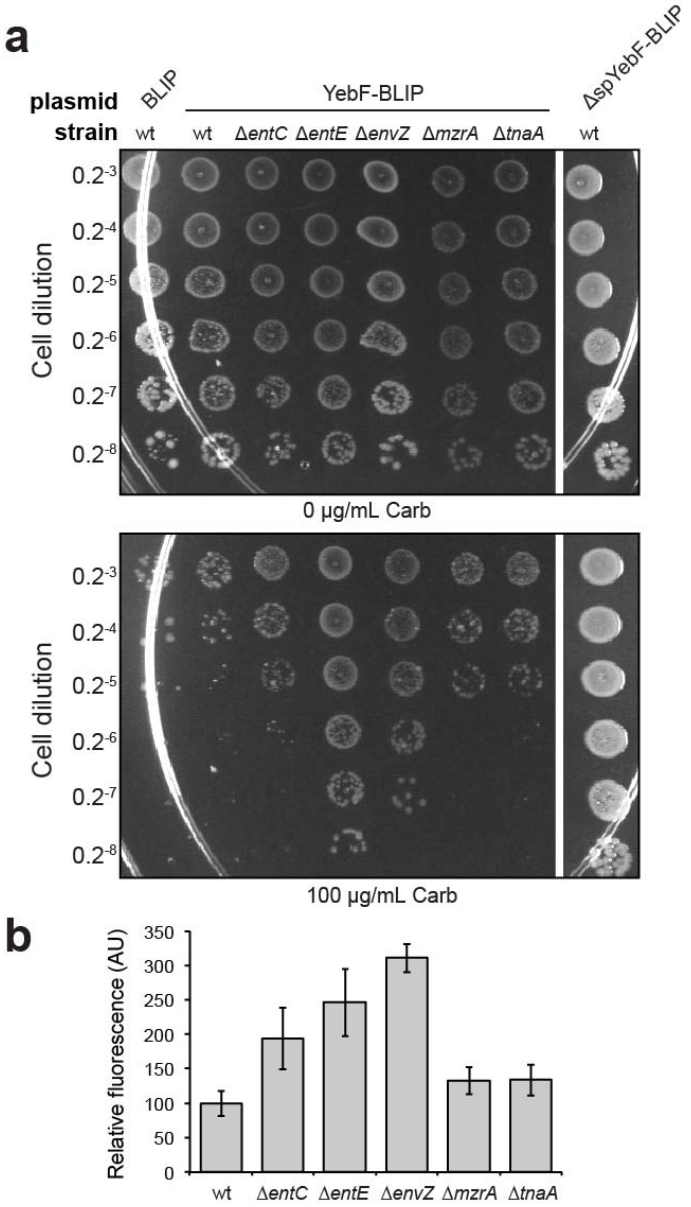


Figure 4-5: Analysis of previously isolated mutations that enhance extracellular secretion.

(a) Spot plate analysis of the following: wt BW25113 cells harboring plasmid pTrc-BLIP, pTrc-YebF-BLIP or spYebF-BLIP; and BW25113 mutant strains harboring plasmid pTrc-YebF-BLIP as indicated. Cells were serially diluted on LB agar plates with or without Carb as indicated. (b) FlAsH labeling of mutant strains in (a) performed using a 96-well plate-based fluorescence assay (8). All fluorescence data was normalized to the signal from wt BW25113 cells expressing YebF-FT (where FT = epitope tag that cross-reacts with FlAsH reagent) and is expressed as the mean ± standard error of the mean (SEM) of biological triplicates.

a precisely defined, single-gene deletion of a nonessential gene(*27*). The individual Keio

strains were pooled and subsequently transformed with plasmid pTrc-YebF-BLIP. The resulting cell library was cultured overnight, diluted 1/10 in fresh medium, and plated on solid Luria−Bertani (LB) agar supplemented with 100 μg/mL Carb. A total of 13 colonies were selected from the mutant library based on their growth in the presence of Carb. Of these, we successfully sequenced six isolates, which contained kanamycin marker insertions in the following nonessential genes: *fliH*, *gfcC*, *yaiW*, *ydfI* and *dsbA*, with *fliH* represented twice. To confirm the resistance phenotypes of these isolates, each was freshly cultured overnight and then grown on antibiotic-supplemented LB agar plates. In the presence of 100 μg/mL Carb, all of the mutant strains outgrew wt BW25113 cells carrying pTrc-YebF-BLIP, with the *ΔgfcC* and *ΔyaiW* mutants exhibiting the greatest resistance ***(Figure 4-6 a)***. In fact, the *ΔgfcC* and *ΔyaiW* cells were significantly more resistant than wt cells at Carb concentrations as high as 200 μg/ mL, with *ΔgfcC* cells growing as well as positive control cells expressing Bla from pTrc99A in the absence of BLIP ***(Figure 4-6 a)***. Importantly, all cells grew equally well in the absence of Carb ***(Figure 4-6 a)*** confirming that differences in resistance were not the result of general growth defects.

In light of the strong resistance observed in the *ΔgfcC* and *ΔyaiW* strain backgrounds, we focused our attention on these two mutants. To determine whether deletion of *gfcC* or *yaiW* led to increased accumulation of YebF in the culture medium, we performed immunoblot analysis on the supernatant fractions derived from cells expressing native YebF lacking the BLIP domain from pTrc-YebF(*8*). As early as 2 h after induction, the *ΔgfcC* and *ΔyaiW* mutants accumulated approximately 6- and 18-fold more YebF in the cell-free supernatant compared to wt BW25113 cells ***(Figure 4-6 b and c)***, which was entirely consistent with the relative antibiotic resistance observed for these strains. At this time point, the total intracellular YebF was significantly lower

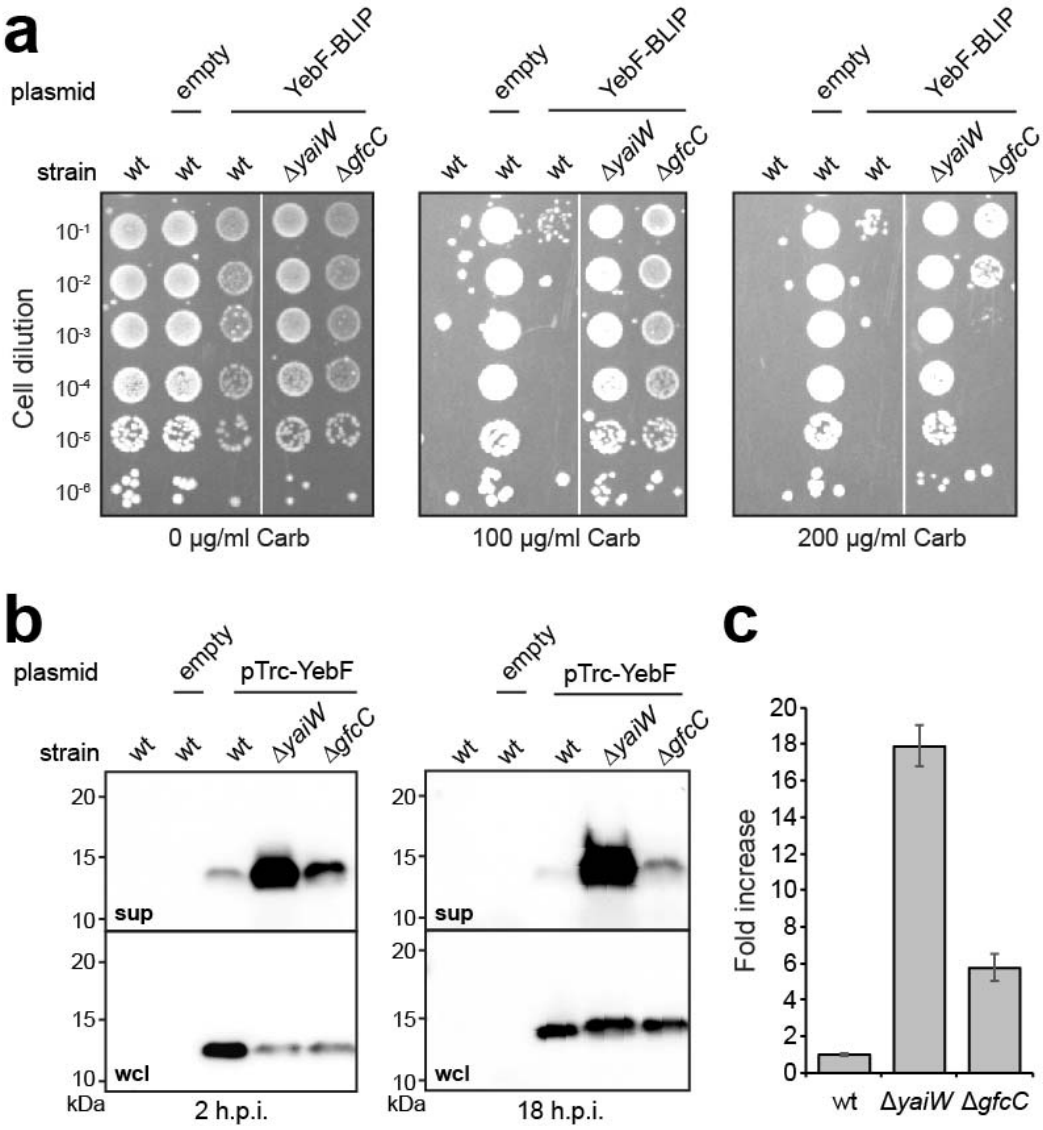


Figure 4-6: Isolation of mutations that enhance extracellular secretion of YebF. *(a) Spot plate analysis of: wt BW25113 cells without a plasmid (wt), wt BW25113 cells harboring pTrc99A (empty) or pTrc-YebF-BLIP (YebF-BLIP), and BW25113 ΔgfcC and ΔyaiW mutants harboring pTrc-YebF-BLIP. Cells were serially diluted on LB agar plates with or without Carb as indicated. (b) Immunoblot analysis of supernatant (sup) fractions and whole cell lysate (wcl) isolated from the same strains as in (a), either 2 or 18 hours post induction (h.p.i.) as indicated. Proteins were detected using an anti-6x-His antibody. Molecular weight (MW) marker indicated on the left. (c) Densitometry analysis corresponding to the 2 h.p.i. immunoblot from (b) was performed using ImageJ software. Fold increase was calculated by normalizing all densitometry values to the value for wt BW25113 cells harboring pTrc-YebF-BLIP.*

in the *ΔgfcC* and *ΔyaiW* mutants compared to wt BW25113 cells ***(Figure 4-6 b)***,

indicating that the higher levels of extracellular YebF produced by the mutants is the result of more efficient secretion out of cells. At later time points after induction, increased levels of extracellular YebF were again observed for the *ΔgfcC* and *ΔyaiW* mutants compared to wt BW25113 cells, while the intracellular YebF levels across all strains had equilibrated ***(Figure 4-6 b)***. The higher levels of YebF secretion in the mutant strains had little to no impact on cell viability, with growth rates and final cell concentrations for the *ΔgfcC* and *ΔyaiW* mutants comparable to those of wt BW25113 cells ***(Figure 4-7 a)***. Taken together, these results confirm that the increased resistance observed for the *ΔgfcC* and *ΔyaiW* mutants was the result of significantly greater expression of YebF in the culture medium. To gain some insight into the mechanism by which the selected mutations enhanced the extracellular accumulation of YebF, we characterized the envelope stability of each mutant strain. This involved assaying for detergent sensitivity and leakage of cytoplasmic and periplasmic components into the supernatant. BW25113 cells lacking *ompR* served as a positive control as these cells are known to be sensitive to detergents and to leak the cytoplasmic protein RNase I and periplasmic proteins maltose-binding protein (MBP) and DsbA into the culture medium(*8, 32*). Cells lacking *yaiW* were moderately defective, leaking MBP and other cellular proteins at a level that was similar to *ΔompR* cells in these analyses ***(Figure 4-7 b)***. Given that YaiW is a surface-exposed outer membrane lipoprotein(*33*), we speculate that the absence of this protein has a destabilizing effect on the outer membrane, resulting in passive diffusion of periplasmic proteins including YebF into the growth medium. In contrast, the *ΔgfcC* strain showed virtually no signs of membrane instability or permeability, except for a very low level of MBP leakage ***(Figure 4-7 b)***, suggesting that the increased resistance of this strain is the result of *bona fide* hypersecretion of YebF into the culture supernatant. GfcC is a periplasmic protein that is thought to interact with GfcD, a predicted β-barrel outer membrane lipoprotein, and together they

appear to affect capsule polysaccharide expression on the cell surface(*34*). Hence, the absence of this protein and a corresponding defect in capsule biosynthesis may somehow favor secretion across the outer membrane by the YebF pathway.

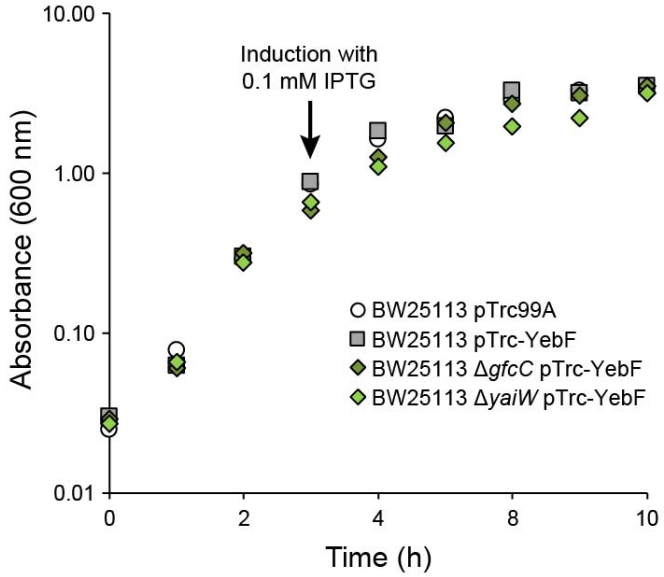
4.3.3. Mutant Strains Promote Hypersecretion of YebF Passenger Protein

An attractive feature of YebF and OsmY is their proven versatility as an extracellular protein expression tag that can deliver a wide range of structurally diverse proteins into the culture supernatant(*8, 10, 11, 21, 22, 24*). For example, it has been demonstrated that different cellulolytic enzymes from *Cellvibrio japonicus*, a Gram-negative bacterium known for efficient plant cell wall degradation(*35*), can be targeted to the culture medium via fusion to YebF or OsmY(*8, 10*). We also showed that mutant *E. coli* strains isolated by fluorescence-based plate screening were capable of more efficient extracellular secretion of these cellulolytic enzymes, including the β-glucosidase Cel3A(*8*). To determine whether the newly isolated *ΔgfcC* and *ΔyaiW* mutants could similarly improve the secretion of a cellulolytic passenger protein, we transformed these strains with a plasmid encoding *C. japonicus* Cel3A fused to YebF(*8*). Western blot analysis of the supernatant fractions revealed that both mutant strains were able to enhance extracellular secretion by more than 2.5-fold compared to wt cells ***(Figure 4-8 a)***.

4.3.4. Mutant Strains Promote Hypersecretion of YebF Bearing N-Linked Glycans

In addition to promoting secretion of structurally diverse passenger proteins, YebF can be efficiently modified in the cytoplasm or periplasm - prior to secretion out of the cell - with various post-translational modifications (PTMs) including biotinylation(*36*), disulfide bond formation(*8, 22, 37*), phosphopantetheinylation(*36*),

a



b

Characterization of outer membrane integrity and permeability

Host	BW25113	BW25113	BW25113	Δ*yaiW*	Δ*gfcC*	Δ*ompR**
Plasmid	-	pTrc99A	pTrc-YebF			
RNAseI leakage[a]	-	-	-	-	-	+
Detergent sensitivity[b]	-	-	-	++	-	-
MBP leakage[c]	-	-	-	++	+	++
Protein profile[d]	-	-	-	+	-	+

*positive control for leakage
[a]no clearing (-); clearing (+)
[b]resistant (-); mildly sensitive (+); sensitive (++)
[c]none detectable (-); low amount (+); high amount (++)
[d]relative protein profile similar to wt (-); similar to leaky Δ*ompR* (+)

Figure 4-7: Characterization of isolated mutant strains.
(a) Growth curve of wt BW25113 cells and mutants derived thereof carrying either plasmid pTrc99A or pTrc-YebF-BLIP as indicated. Protein expression was induced with 0.1 mM IPTG at the time indicated. Culture absorbance at 600 nm (Abs_{600}) was periodically recorded prior to and following induction over 10 h. (b) Outer membrane integrity and permeability for plasmid-free wt BW25113 cells compared to wt BW25113 cells and mutants derived thereof carrying plasmid pTrc99A or pTrc-YebF-BLIP as indicated. Assays for detergent sensitivity as well as leakage of RNase I, MBP, and intracellular proteins were performed as described previously(23).

and *N*-linked glycosylation(*38, 39*). In light of these findings, we wondered whether any of these PTMs, in particular *N*-linked glycosylation, could be coupled with hypersecretion to produce higher amounts of modified YebF in the culture medium.

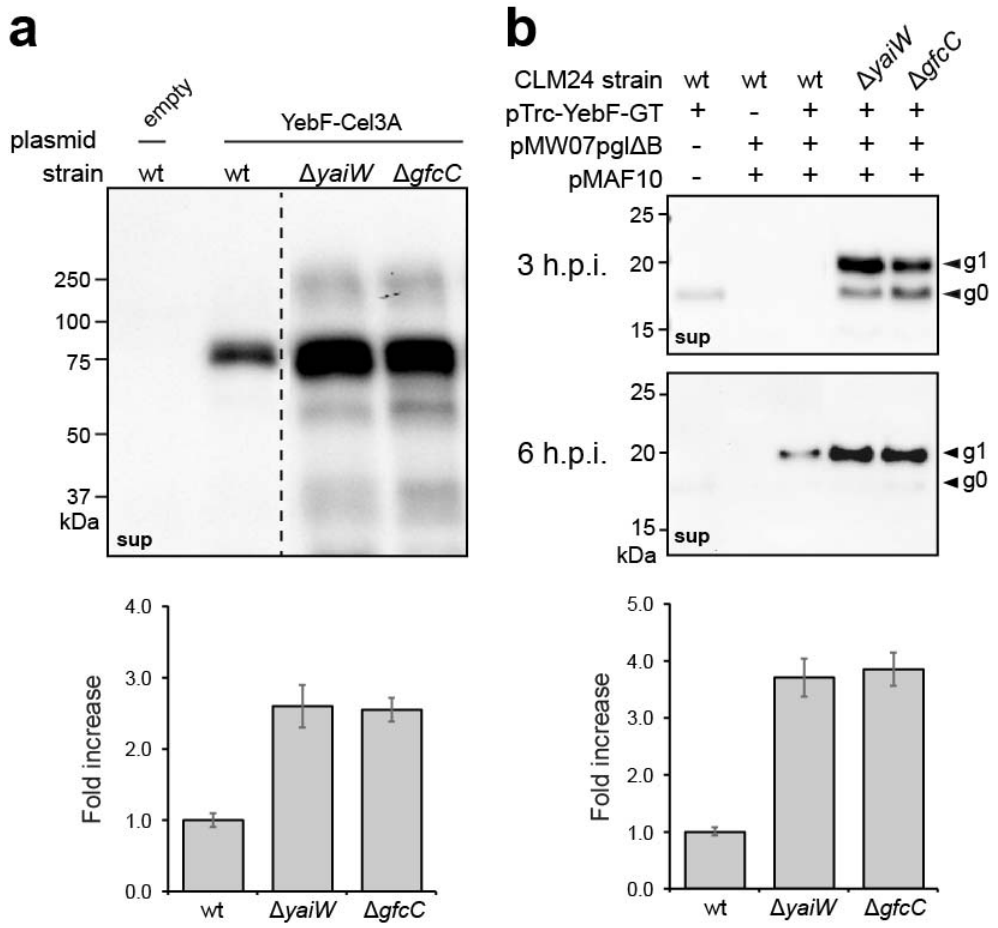


Figure 4-8: Hypersecretion of structurally diverse substrate proteins.
(a) Immunoblot analysis of supernatant (sup) fractions from the following: wt BW25113 cells harboring pTrc99A (empty) or pTrc-YebF-Cel3A (YebF-Cel3A); and BW25113 ΔgfcC and ΔyaiW mutants harboring pTrc-YebF-Cel3A. Proteins were detected using an anti-6x-His antibody. Molecular weight (MW) marker indicated on the left. Densitometry analysis corresponding to the immunoblot above was performed using ImageJ software. Fold increase was calculated by normalizing all densitometry values to the value for wt BW25113 cells harboring pTrc-YebF-Cel3A. (b) Immunoblot analysis of supernatant (sup) fractions isolated at 3 or 6 hours post induction (h.p.i.) from CLM24 cells or CLM24 ΔgfcC and ΔyaiW mutants harboring pTrc-YebF-GT, pMW07pglΔB, and pMAF10 as indicated. Proteins were detected using an anti-6x-His antibody. Singly glycosylated (g1) and aglycosylated (g0) forms of YebF indicated on the right; molecular weight (MW) marker indicated on the left. Densitometry analysis corresponding to the 3 h.p.i. immunoblot above was performed using ImageJ software. Fold increase was calculated by normalizing all densitometry values to the value for wt CLM24 cells harboring pTrc-YebF-GT, pMW07pglΔB, and pMAF10.

To generate *N*-glycosylated derivatives of YebF, we employed plasmid pTrc-YebF-GT that encodes a modified version of YebF carrying a bacterial glycosylation tag (GT) at its C-terminus(*40*). In addition, the *ΔgfcC* and *ΔyaiW* single-gene knockouts were genetically transferred into CLM24 cells, an *E. coli* K-12 strain that was previously optimized for *N*-linked glycosylation(*41*), by P1 transduction. The resulting mutant strains, CLM24 *ΔgfcC* and CLM24 *ΔyaiW*, were transformed with pTrc-YebF-GT along with plasmid pMW07*pglΔB*(*39*), which encodes enzymes from the *Campylobacter jejuni* protein *N*-glycosylation locus (*pgl*) that catalyze the biosynthesis of a heptasaccharide glycan, and plasmid pMAF10(*41*), which encodes the *C. jejuni* oligosaccharyltransferase PglB that catalyzes transfer of the heptasaccharide glycan to acceptor sites in target proteins. Following various durations of induction, the supernatant fractions derived from the glyco-competent *ΔgfcC* and *ΔyaiW* mutant strains contained significantly more glycosylated YebF than the supernatant derived from glycocompetent wt CLM24 ***(Figure 4-8 b)***. Importantly, in all cases where glycosylation machinery was present, the efficiency of glycosylation was relatively high as evidenced by the intensity of the band corresponding to the singly glycosylated protein (g1) compared to the aglycosylated protein (g0). Taken together, these results confirm that the hyper-secretory phenotypes associated *ΔgfcC* and *ΔyaiW* knockout are (i) maintained following genetic transfer of the alleles to different *E. coli* host strains; and (ii) entirely compatible with intracellular PTMs, namely *N*-linked glycosylation, that can effectively expand the structural diversity of protein substrates that are secreted to the extracellular environment.

4.4 Discussion

Here, we report a generalizable survival-selection assay that effectively links extracellular protein secretion with antibiotic resistance, thereby enabling isolation of

desired secretory phenotypes based solely on cell growth. When applied in the context of the YebF secretion pathway, this selection strategy permitted the isolation of several gain-of-function mutations that increased the efficiency of extracellular expression without compromising the integrity of the outer membrane. This is significant because, to date, there are only a few reported examples of using an evolutionary (i.e., screening- and selection-based) approach to isolate *E. coli* mutants with improved protein production characteristics(*42*). For extracellular protein expression, in particular, the lack of suitable screens and selections represents a major bottleneck to such evolutionary optimization, with most efforts limited to more tedious directed engineering (i.e., "one-gene-at-a-time") approaches such as deleting or co-expressing genes that often lead to little or no improvement in target protein secretion(*43*). We recently reported a generic fluorescence-based screen for extracellular protein expression(*8*); however, to our knowledge, this is the first report of a genetic selection for this challenging and important phenotype. Because selection is performed under conditions in which only the desired mutant can grow or its growth is at least strongly favored, much larger clonal populations can be examined than with a genetic screen(*25, 26*). Moreover, selection can be performed without the need for sophisticated instrumentation or costly reagents.

Another advantage of our selection is that, unlike several earlier screens for extracellular secretion that were based on the native activity (*e.g.*, lignocellulolysis) of the exoprotein substrate, our assay does not rely on the intrinsic activity of the secreted protein. As a result, enzymes with low specific activity or whose activity does not have a reliable assay can still be investigated and/or optimized using our selection platform provided that two important caveats are met. First, the chimera generated between the relatively small BLIP domain (167 amino acids) and the exoprotein of interest must

remain competent for extracellular secretion, as was observed here for YebF-BLIP. Second, the extracellular secretory pathway must be a two-step mechanism in which a secreted substrate is translocated first to the periplasm and then to the culture supernatant (*e.g.*, type II secretion). In support of the assay's adaptability, we have observed that the BLIP-based genetic selection is compatible with extracellular secretion of *Dickeya dadantii* PelB when co-expressed in an *E. coli* host strain equipped with the *D. dadantii* type II secretion machinery (unpublished observations)(*20*).

At present, the exact mechanism(s) by which the isolated mutations enhance extracellular YebF secretion remains unclear. Several of the knocked-out genes, namely *fliH*, *gfcC*, and *yaiW*, encode proteins that are either directly associated with the *E. coli* inner and outer membranes, or those that interact with membrane-associated biomolecules, suggesting that their deletion may serendipitously favor the secretion of YebF while still maintaining structural integrity. In the case of DsbA, a periplasmic oxidoreductase enzyme that associates with the transmembrane protein DsbB in the disulfide bond formation pathway, follow-on Western blot analysis of supernatant fractions isolated from the *ΔdsbA* strain resulted in no measurable extracellular YebF. While initially surprising, this result was entirely consistent with a previous report showing that deletion of *dsbA* caused a large decrease in YebF abundance(*37*). These authors further demonstrated that the two cysteine residues in mature YebF were oxidized in wt cells and reduced in a *dsbA* strain, leading to the conclusion that the significantly diminished YebF levels in *dsbA* cells is likely due to misfolding and clearance by proteases. In light of these observations, we speculate that isolation of the *ΔdsbA* mutant is best explained by mis-oxidation and proteolytic degradation of the periplasmic YebF-BLIP intermediate, thereby relieving inhibition of Bla and conferring a growth advantage to cells. Importantly, even though this mutant represents a false

positive in the context of engineering improved extracellular secretion of YebF, it provides direct evidence that our selection can be advantageously used to uncover mechanistic details about the secretion pathway being studied.

Previously, we reported that the yield of YebF in the supernatant of wt BW25113 cells is 0.4 mg/L(*8*) which is reasonable considering cultivation was performed in small shake flask cultures with no optimization of growth/induction conditions. Importantly, we predict yields in the 1−10 mg/L range for the *yaiW* and *gfcC* knockout strains isolated here based on densitometry analysis of secreted YebF ***(Figure 4-6 c)***. While these levels are not yet commercially viable, we note that transfer of such hypersecretion mutations to strains of *E. coli* that are more amenable to protein overexpression, such as BL21(DE3) carrying plasmids that employ strong T7 promoters, increased yield by an order of magnitude(*8*). In light of these observations, we anticipate that significantly higher yields could be achieved in the future by transferring the *yaiW* and *gfcC* mutations to more favorable expression host/ plasmid combinations and by optimizing the growth environment (i.e., fermentor vs shake flask) and cultivation conditions (i.e., fed-batch vs batch).

While efforts here focused on mutant strain discovery, further optimization efforts could focus on selecting mutagenic libraries of YebF or the outer membrane porins OmpC and OmpF through which YebF is thought to transit(*23*). These evolutionary strategies could be performed in the context of YebF alone, as we performed here, or could also be implemented in the context of a "passenger" protein of interest thereby enabling product-specific optimization. Such efforts are becoming increasingly important in light of the growing number of applications that involve the fusion of passenger proteins to YebF and its counterparts (e.g., OsmY, OmpA and

OmpF) including recombinant protein production(*8, 21, 22, 24, 44, 45*), consolidated bioprocessing(*8, 10, 11*), and drug delivery(*14*). We anticipate an even greater number of applications to emerge given that these proteins are compatible with important intracellular PTMs, such as biotinylation(*36*), disulfide bond formation(*8, 22, 37*), phosphopantetheinylation(*36*), and *N*-linked glycosylation(*38, 39*), that are installed prior to secretion out of the cell. Taken together, these opportunities underscore the importance of understanding and engineering the secretion capacity of *E. coli* and provide yet another example of "the unexhausted potential"(*46*) of this model organism.

4.5 Materials and Methods

4.5.1. Bacterial Strains and Growth Conditions

E. coli strain BW25113 (*lacI$_q$ rrnBT14 ΔlacZWJ16 hsdR514 ΔaraBAD$_{AH33}$ ΔrhaBAD$_{LD78}$*) and single-gene knockout mutants of BW25113 from the Keio collection(*27*) were used for genetic selection and related secretion experiments. *E. coli* strain CLM24(*41*) was used for *N*-linked glycosylation of YebF. Knockout mutants of CLM24 were generated by P1 transduction using the corresponding BW25113 single-gene knockout mutants as donors. All strains were routinely grown aerobically at 37 °C in LB medium, and antibiotics were supplemented at the following concentrations: carbenicillin (Carb, 100 μg/mL), ampicillin (Amp, 100 μg/mL), kanamycin (Kan, 50 μg/mL), chloramphenicol (Cm, 20 μg/mL), and trimethoprim (Tmp, 100 μg/mL). Heterologous protein expression was induced when the cells reached an absorbance at 600 nm (Abs_{600}) of ~0.5 by adding 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and/or 0.2% (w/v) arabinose to the media as noted. Glucose (0.2% w/v) was used to repress expression of BLIP fusion proteins from pTrc-YebF-BLIP during strain maintenance. To generate growth curves, overnight cultures were subcultured in 50 mL

fresh LB media and protein expression was induced with 0.1 mM IPTG when Abs_{600} reached ~0.5−1.0. Abs600 was recorded periodically as a function of time.

4.5.2. Plasmid Construction

All plasmids constructed in this study were derivatives of pTrc99A (Amp^R) (Amersham Pharmacia). To generate pTrc-YebF-BLIP, DNA encoding the signal peptide from *S. clavuligerus* BLIP was PCR-amplified from pluxBLIP and cloned between EcoRI and SacI of pTrc99A. It should be noted that the sequence encoding the native signal peptide was extensively modified to improve genetic stability *in vivo* without changing the amino acid sequence(*31*), including reduction of both GC content and homology of the periplasmic signal peptide region with that of TEM-1 β-lactamase. Next, DNA encoding mature YebF (lacking its native signal peptide) was PCR-amplified from *E. coli* genomic DNA and ligated between SacI and XbaI, which placed it just after the BLIP signal peptide. Finally, mature BLIP (lacking its native signal peptide) was PCR-amplified from pluxBLIP and ligated between XbaI and HindIII, which placed it just after the gene encoding mature YebF. Plasmids pTrc- ΔspYebF-BLIP and pTrc-BLIP were constructed by sequentially removing the DNA encoding the BLIP signal peptide and mature YebF, respectively, from pTrc-YebF-BLIP. A C-terminal 6x-His tag was included on all constructs. All primers were synthesized by Integrated DNA Technologies and all plasmids were verified by DNA sequencing. The following additional plasmids were used: pTrc-YebF(*8*), pTrc-YebF-Cel3A(*8*), pTrc-YebF-GT(*40*), pMW07*pglΔB*(*39*), and pMAF10(*41*).

4.5.3. Selective Plating of Bacteria

Selective plating of bacteria was performed as described(*47*). Briefly, BW25113 cells carrying pTrc99A, pTrc-BLIP, pTrc-YebF-BLIP, or pTrc-ΔspYebF-BLIP were

grown overnight at 37 °C in LB medium supplemented with Carb to maintain the plasmid. The next day, antibiotic resistance of bacteria was evaluated by spot plating 5 μL of serially diluted overnight cells that had been normalized in fresh LB to Abs_{600} of ~1.0 onto LB agar plates supplemented with Carb and IPTG. For mock selection experiments, overnight cultures of BW25113 cells carrying either pTrc-BLIP or pTrc-YebF-BLIP were mixed in equal amounts and subsequently selected by plating the cell mixture on LB agar plates supplemented with Carb and IPTG. Plasmids from ten randomly selected colonies were isolated and analyzed by restriction enzyme digestion using EcoRI and HindIII to release either the BLIP or YebF-BLIP insert, with predicted sizes of 672 bp and 948 bp, respectively. Digestion products were analyzed by agarose gel electrophoresis.

4.5.4. Genetic Selection of Keio Knockout Library

To generate a comprehensive gene knockout library, individual members of the Keio strain collection(*27*) were pooled into a single culture and transformed with pTrc-YebF-BLIP. Following transformation, the library was grown overnight and subsequently selected by plating the cell mixture on LB agar plates supplemented with Carb and IPTG. To identify mutations that conferred enhanced extracellular secretion, a range of Carb concentrations were tested to find conditions where individual library members were resistant but where separately plated control cells, namely wt BW25113 carrying pTrc-YebF-BLIP, were sensitive. Clones selected in this manner were chosen for further characterization. To determine the identity of the knocked-out gene in selected hits, arbitrary nested PCR was used to identify the genetic region flanking the insertion coding for the Kan resistance (Kan^R) cassette. A first round of PCR was performed using forward primer KRP1 (GAC CGC TAT CAG GAC ATA GCG TTG) and reverse degenerate primer ARB1A (GCC ACG CGT CGA CTA GTA CNN NNN

NNN NNA CGC C), ARB1B (GCC ACG CGT CGA CTA GTA CNN NNN NNN NNT GCG G), or ARB1C (GCC ACG CGT CGA CTA GTA CNN NNN NNN NNT CCG G). A first PCR reaction was performed by adding 1 μL of overnight culture of each selected hit directly to the reaction mix containing these primers. A second PCR reaction was performed using 2 μL from the first PCR, forward primer KRP2 (GTG CTT TAC GGT ATC GCC GCT C) specific to a region further downstream within the Kan^R cassette and arbitrary reverse primer ARB2 (GCC ACG CGT CGA CTA GTA C). The PCR product from the second reaction was gel purified and sequenced.

4.5.5. Protein Analysis

Cells expressing recombinant proteins were harvested after overnight induction as described above. Cell-free supernatant fractions were prepared by centrifugation at 5,000 g for 10 min, and the proteins were precipitated overnight with 1% TCA at 4°C. After centrifugation at 13,000 rpm for 30 min, the pellet was washed once with ice-cold 100% acetone and dissolved in 1 M Tris-HCl pH 8.0 and an equal volume of electrophoresis loading buffer. The periplasm was separated from the cytoplasm using the osmotic shock method(*48*). All proteins samples were loaded in equal amounts for separation on an SDS-PAGE gel and then transferred to a nitrocellulose membrane for Western blot analysis. The following primary antibodies were used: mouse anti6x-His (1:5000; Abcam) and mouse anti-MBP (1:2000, New England Biolabs). Pierce enhanced chemiluminescent (ECL) substrate (Thermo Scientific) was used for detection of bound antibodies. All blots were visualized using a Chemidoc XRS+ system with Image Lab image capture software (Bio-Rad). Densitometry analysis of blots was performed using ImageJ analysis software(*49*).

4.5.6. Characterization of Outer Membrane Permeability

Assays for detergent sensitivity and RNase I, MBP, and DsbA leakage were performed essentially as described before(*8, 32*).

4.6 Author Contributions

A.N. and C.H.H. designed research, performed research, analyzed data and wrote the paper. R.L. performed research and analyzed data. J.T.B. designed research. M.P.D. designed research, analyzed data and wrote the paper. The authors declare no competing financial interest.

4.7 Acknowledgments

We thank Dr. Anne Ollis for helpful discussions of the manuscript. Funding was provided by the DOE Great Lakes Bioenergy Research Center (GLBRC) Project 3.2.8, USDA NIFA Award # 2009-02202, and NSF CBET Award # 1159581 (all to M.P.D.), and NSF GK12 Award # DGE-1045513 "Grass Roots: Advancing education in renewable energy and cleaner fuels through collaborative graduate fellow/teacher/grade school student interactions" (to J.T.B.).

4.8 References

1. F. Baneyx, M. Mujacic, **Recombinant protein folding and misfolding in Escherichia coli.** *Nat Biotechnol* 22, 1399-1408 (2004).
2. J. H. Choi, S. Y. Lee, **Secretory and extracellular production of recombinant proteins using Escherichia coli.** *Appl Microbiol Biotechnol* 64, 625-635 (2004).
3. G. Georgiou, L. Segatori, **Preparative expression of secreted proteins in bacteria: status report and future prospects.** *Curr Opin Biotechnol* 16, 538-545 (2005).

4. G. Georgiou, P. Valax, **Expression of correctly folded proteins in Escherichia coli.** *Curr Opin Biotechnol* 7, 190-197 (1996).

5. R. J. Ellis, **Macromolecular crowding: an important but neglected aspect of the intracellular environment.** *Curr Opin Struct Biol* 11, 114-119 (2001).

6. S. B. Zimmerman, A. P. Minton, **Macromolecular crowding: biochemical, biophysical, and physiological consequences.** *Annu Rev Biophys Biomol Struct* 22, 27-65 (1993).

7. J. R. Swartz, **Advances in Escherichia coli production of therapeutic proteins.** *Curr Opin Biotechnol* 12, 195-201 (2001).

8. C. H. Haitjema *et al.*, **Universal genetic assay for engineering extracellular protein expression.** *ACS Synth Biol* 3, 74-82 (2014).

9. D. M. Widmaier *et al.*, **Engineering the Salmonella type III secretion system to export spider silk monomers.** *Mol Syst Biol* 5, 309 (2009).

10. G. Bokinsky *et al.*, **Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli.** *Proc Natl Acad Sci U S A* 108, 19949-19954 (2011).

11. E. J. Steen *et al.*, **Microbial production of fatty-acid-derived fuels and chemicals from plant biomass.** *Nature* 463, 559-562 (2010).

12. C. Yang, C. Song, R. Freudl, A. Mulchandani, C. Qiao, **Twin-arginine translocation of methyl parathion hydrolase in Bacillus subtilis.** *Environ Sci Technol* 44, 7607-7612 (2010).

13. F. Duan, K. L. Curtis, J. C. March, **Secretion of insulinotropic proteins by commensal bacteria: rewiring the gut to treat diabetes.** *Appl Environ Microbiol* 74, 7437-7438 (2008).

14. I. Y. Hwang *et al.*, **Reprogramming microbes to be pathogen-seeking killers.** *ACS Synth Biol* 3, 228-237 (2014).

15. A. Z. Reeves *et al.*, **Engineering Escherichia coli into a protein delivery system for mammalian cells.** *ACS Synth Biol* 4, 644-654 (2015).

16. H. Russmann *et al.*, **Delivery of epitopes by the Salmonella type III secretion system for vaccine development.** *Science* 281, 565-568 (1998).

17. E. Papanikou, S. Karamanou, A. Economou, **Bacterial protein secretion through the translocase nanomachine.** *Nat Rev Microbiol* 5, 839-851 (2007).

18. A. P. Pugsley, O. Francetic, **Protein secretion in Escherichia coli K-12: dead or alive?** *Cell Mol Life Sci* 54, 347-352 (1998).

19. J. H. Ham, D. W. Bauer, D. E. Fouts, A. Collmer, **A cloned Erwinia chrysanthemi Hrp (type III protein secretion) system functions in Escherichia coli to deliver Pseudomonas syringae Avr signals to plant cells and to secrete Avr proteins in culture.** *Proc Natl Acad Sci U S A* 95, 10206-10211 (1998).

20. S. Y. He, M. Lindeberg, A. K. Chatterjee, A. Collmer, **Cloned Erwinia chrysanthemi out genes enable Escherichia coli to selectively secrete a diverse family of heterologous proteins to its milieu.** *Proc Natl Acad Sci U S A* 88, 1079-1083 (1991).

21. Z. G. Qian, X. X. Xia, J. H. Choi, S. Y. Lee, **Proteome-based identification of fusion partner for high-level extracellular production of recombinant proteins in Escherichia coli.** *Biotechnol Bioeng* 101, 587-601 (2008).

22. G. Zhang, S. Brokx, J. H. Weiner, **Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in Escherichia coli.** *Nat Biotechnol* 24, 100-104 (2006).

23. G. Prehna *et al.*, **A protein export pathway involving Escherichia coli porins.** *Structure* 20, 1154-1166 (2012).

24. A. Kotzsch *et al.*, **A secretory system for bacterial production of high-profile protein targets.** *Protein Sci* 20, 597-609 (2011).

25. A. J. Link, K. J. Jeong, G. Georgiou, **Beyond toothpicks: new methods for isolating mutant bacteria.** *Nat Rev Microbiol* 5, 680-688 (2007).

26. H. A. Shuman, **Just toothpicks and logic: how some labs succeed at solving complex problems.** *J Bacteriol* 185, 387-390 (2003).

27. T. Baba *et al.*, **Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.** *Mol Syst Biol* 2, 2006.0008 (2006).

28. J. L. Doran, B. K. Leskiw, S. Aippersbach, S. E. Jensen, **Isolation and characterization of a beta-lactamase-inhibitory protein from Streptomyces clavuligerus and cloning and analysis of the corresponding gene.** *J Bacteriol* 172, 4909-4918 (1990).

29. N. C. Strynadka, S. E. Jensen, P. M. Alzari, M. N. James, **A potent new mode of beta-lactamase inhibition revealed by the 1.7 A X-ray crystallographic structure of the TEM-1-BLIP complex.** *Nat Struct Biol* 3, 290-297 (1996).

30. C. H. Collins, J. R. Leadbetter, F. H. Arnold, **Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR.** *Nat Biotechnol* 24, 708-712 (2006).

31. Y. Yokobayashi, F. H. Arnold, **A dual selection module for directed evolution of genetic circuits.** *Nat Comput* 4, 245-254 (2005).

32. A. J. McBroom, A. P. Johnson, S. Vemulapalli, M. J. Kuehn, **Outer membrane vesicle production by Escherichia coli is independent of membrane instability.** *J Bacteriol* 188, 5385-5392 (2006).

33. M. F. Arnold *et al.*, **Enteric YaiW is a surface-exposed outer membrane lipoprotein that affects sensitivity to an antimicrobial peptide.** *J Bacteriol* 196, 436-444 (2014).

34. K. Sathiyamoorthy, E. Mills, T. M. Franzmann, I. Rosenshine, M. A. Saper, **The crystal structure of Escherichia coli group 4 capsule protein GfcC reveals a domain organization resembling that of Wza.** *Biochemistry* 50, 5465-5476 (2011).

35. R. T. DeBoy *et al.*, **Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus.** *J Bacteriol* 190, 5455-5463 (2008).

36. N. Chen *et al.*, **Modified recombinant proteins can be exported via the Sec pathway in Escherichia coli.** *PLoS One* 7, e42519 (2012).

37. D. Vertommen *et al.*, **The disulphide isomerase DsbC cooperates with the oxidase DsbA in a DsbD-independent manner.** *Mol Microbiol* 67, 336-349 (2008).

38. A. C. Fisher *et al.*, **Production of secretory and extracellular N-linked glycoproteins in Escherichia coli.** *Appl Environ Microbiol* 77, 871-881 (2011).

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

40. A. A. Ollis *et al.*, **Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences.** *Sci Rep* 5, 15237 (2015).

41. M. F. Feldman *et al.*, **Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli.** *Proc Natl Acad Sci U S A* 102, 3016-3021 (2005).

42. S. Schlegel, P. Genevaux, J. W. de Gier, **Isolating Escherichia coli strains for recombinant protein production.** *Cell Mol Life Sci* 74, 891-908 (2017).

43. Z. Zheng, T. Chen, M. Zhao, Z. Wang, X. Zhao, **Engineering Escherichia coli for succinate production from hemicellulose via consolidated bioprocessing.** *Microb Cell Fact* 11, 37 (2012).

44. E. J. Seo, S. Weibel, J. Wehkamp, T. A. Oelschlaeger, **Construction of recombinant E. coli Nissle 1917 (EcN) strains for the expression and secretion of defensins.** *Int J Med Microbiol* 302, 276-287 (2012).

45. T. N. Wang, M. Zhao, **A simple strategy for extracellular production of CotA laccase in Escherichia coli and decolorization of simulated textile effluent by recombinant laccase.** *Appl Microbiol Biotechnol* 101, 685-696 (2017).

46. Z. D. Blount, **The unexhausted potential of E. coli.** *Elife* 4, (2015).

47. A. C. Fisher, W. Kim, M. P. DeLisa, **Genetic selection for protein solubility enabled by the folding quality control feature of the twin-arginine translocation pathway.** *Protein Sci* 15, 449-458 (2006).

48. M. P. DeLisa, D. Tullman, G. Georgiou, **Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway.** *Proc Natl Acad Sci U S A* 100, 6115-6120 (2003).

49. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, **NIH Image to ImageJ: 25 years of image analysis.** *Nat Methods* 9, 671-675 (2012).

CHAPTER 5

ARISE: Amplifying Role models as Inspiration for STEM Education[1]

5.1 Abstract

Podcasts are a burgeoning medium of science communication that have the potential to be highly impactful given the ease of accessibility and exponentially growing body of content. However, little is known about how to leverage this contemporary modality as an intervention to improve the sense of belonging and self-efficacy among students from traditionally under-represented identities in STEM (URMs). We hypothesize that weaving in a personal narrative encompassing the identity and experiences of a URM scientist along with their scientific accomplishments might inspire students from similar backgrounds and experiences to feel an increased sense of belonging in STEM, in addition to being educated by the scientific content of the podcast episode. For this study we chose to work with students from four community colleges, since, particularly for URMs, these institutions serve as the gateway to higher education. While there was no marked increase in the participants' sense of belonging, this study provides the first evidence that listeners of a podcast episode are more likely to identify with a speaker with shared demographic identities, even when their identity is not explicitly revealed. Further, listening to a shorter episode that focuses on just the science leads to greater retention of information. Overall, this study paves the way to follow up investigations to explore how to effectively leverage this new-age media to be an impactful intervention promoting access to science and the diversification of higher education especially in STEM. By showing that listeners tend to associate with speakers of shared identity, it also emphasizes the importance of diversity in the voices of scientists captured through science podcasts.

[1] *Pre-published data from a collaborative study:* Natarajan A., Elmore K.C., Weyant K., and Daniel S.

5.2 Introduction

Community colleges play a pivotal role in the education of more than 50% of American students, including a disproportionately high number of students from traditionally under-represented identities in STEM (URMs)(*1-3*) - women (58%) and ethnic minorities (38.5%) (*4, 5*). However, in science and engineering, alarmingly few women(*4, 6, 7*) and ethnic minorities complete their degrees(*4, 8*). URMs face the daunting challenge of isolation, being among few of their identity group pursuing a STEM education, and having even fewer ingroup members to look up to as role models(*9-11*). Yet, community colleges continue to be a crucial pipeline for higher education, especially among URMs(*1-5*). Therefore, developing and testing new approaches to encourage vulnerable students at community colleges that STEM is a place they can belong and acquire critical career skills is an essential step towards ushering in equity in higher education and spurring social mobility. In a bid to find an impactful strategy that is not hemmed by the traditional structures of academic programs, we explore podcasts.

Podcasts present a uniquely impactful opportunity to support the success of students by complementing the more traditional efforts of teaching and mentoring on community college campuses. They are presented in a primarily audio-based format, offering listeners the freedom to engage with a varying range of attention, often for no cost. This flexibility and unprecedented accessibility have fueled an exponential rise in podcast listeners, presenting a new opportunity for academic engagement with a reach beyond any other existing forums. In 2018, Edison Research found that, in the U.S., 124 million people have listened to a podcast (compared to 16.8 million students enrolled in post-secondary institutions in fall 2017(*12*)), roughly equally among men (52%) and women (48%), and 30% of people age 12-24 had listened to a podcast in the past

month(*13*). Further, given studies that clearly document the increased uptake of information provided through this medium(*14, 15*), it is no surprise that podcasts have made their mark in science communication(*16, 17*) and as a teaching tool in higher education(*18-22*). However, podcasts have predominantly been limited to three broad academic uses – a) disseminating course content such as recorded seminars or instructions, b) making available information through audiobooks, c) enhancing studying through active learning engagements(*23-25*). Further, there are no studies that employ podcasts as an intervention to increase retention of URMs in STEM education. While there is ample evidence that role models are important to inspire the success of URMs(*26-29*), nobody has studied how to present role models *via* podcasts.

Our hypotheses are that blending the personal narrative of a scientist – including information about their minority identity and experiences thereof leading to a Ph.D. in a STEM field – with their scientific accomplishments may 1) present a readily accessible role model for other URMs, 2) increase understanding of the scientific content through finding the scientist relatable. The current study tests these hypotheses in collaboration with four community colleges in the state of New York, across 160 student participants.

5.3 Methods

Participants were recruited from community college campuses in the State of New York ***(Table 5-1)***; only students who were enrolled in a STEM-related course were eligible to participate. Participants answered a pre-survey and, one week later, were asked to listen to a podcast and complete a post-survey. In total, 228 participants enrolled in the study, and 160 completed both the pre and post-surveys. The final analytic sample was 42% men (57% women, 2% other gender identity) and an average

age of 23.4 years old. Most participants were earlier in their college careers, 28% were first years, 47% were second years, 23% were in their third year or higher. When asked their racial group membership, 68% were white, 14% were black, 7% were Asian or Asian American, and 9% of participants reported their ethnicity as Hispanic/Latino.

Community College	**Number of participants**		
	SO	**PN**	**Total**
Brookdale Community College	2	2	4
CUNY Borough of Manhattan Community College	4	2	5
Onondaga Community College	27	31	57
SUNY Broome Community College	44	47	89
Total	77	82	159

Table 5-1: Summary of participant distribution across community colleges partners and experimental case.
SO refers to a science only narrative; PN is an enhanced version with personal narrative integrated with the science.

Participants were randomly assigned to one of two podcast treatment conditions and completed two assessments (pre- and post-treatment). The treatment of interest is in the content of the podcast; one group listened to a roughly 15-minute narrative focused only on the scientific content of an Engineering student's dissertation work (Science Only = SO podcast, n=78). The other group listened to a roughly 30-minute enhanced podcast in which the narrative includes not only scientific content but also a personal narrative about the student's identity and experiences thereof leading to higher education (Personal Narrative = PN podcast, n=82).

Both pre- & post-surveys measure anticipated belonging in science, self-efficacy in science, perceptions of one's effort expenditure relative to other students in science

class, intentions to pursue a career in science, anticipated belonging at an elite college, and intentions to pursue additional higher education. The post-test also includes a measure of identification with the podcast scientist and a judgment of the amount of personal information heard about the scientist. Participants who completed both the pre and post-test were rewarded with a $25 gift card.

Science Blender, a podcast produced by the Robert Frederick Smith School of Chemical and Biomolecular Engineering, Cornell University (CBE) provided us relevant audio content. Through feedback from experienced mentors and STEM faculty in the collaborating community colleges, we chose to use the story of Dr. Michael Paul Robinson, an African American male and recent graduate of CBE as the study stimulus, since we thought his story of not having a clear path to higher education might be most compelling and broadly relatable to our participants.

5.4 Measures

5.4.1. Anticipated belonging in science

Measured by the average of participants' agreement with three items: "I would get along well with people in a science department," "I feel I would belong in a science department," and "I would feel comfortable around people in a science department." Responses were recorded from 1 (Strongly disagree) to 7 (Strongly agree), and items scaled together at the pre-survey (α=.90; M=5.55, SD=1.51) and the post-survey (α=.83; M=5.60, SD=1.19).

5.4.2. Self-efficacy in science

Measured by a single item within a set of questions asking: "In general, how confident do you feel about your abilities in the following educational domains," of

which “science” was one. Responses were recorded from 1 (Not at all confident) to 5(very confident), Pre-survey *M*=3.87, *SD*=1.0; Post-survey *M*=3.84, *SD*=1.04.

5.4.3. Perceptions of one’s effort expenditure relative to other students in science class

Measured using four items: "Compared with other students, how much effort do you expend in your science classes?", "Compared with other students, to what extent to do you find the material and work in science classes challenging?", "Compared with other students, to what extent does science come easily and naturally to you?" (Reverse coded), and "Compared with other students, how much energy does it take you to succeed in science?" Responses were recorded from 1 (A lot less) to 5 (A lot more), and items scaled together at the pre-survey (α=.77; *M*=2.97, *SD*=.83) and the post-survey (α=.81; *M*=2.91, *SD*=.83).

5.4.4. Intentions to pursue a career in science

Measured with three items, which asked: “To what extent do you intend to: Pursue a science-related major; Pursue a science-related career; Pursue graduate study in science” on a scale that ranged from 1 (definitely will not) to 7 (definitely will). Items scaled together at the pre-survey (α=.92; *M*=5.42, *SD*=1.74) and the post-survey (α=.90; *M*=5.31, *SD*=1.73).

5.4.5. Intentions to pursue additional higher education

Measured with three items, which asked: “To what extent do you intend to: Apply to a 4-year college; Apply to an elite 4-year college like Cornell; Pursue graduate study” on a scale that ranged from 1 (definitely will not) to 7 (definitely will). Items did

not scale together in the pre-survey (α=.18), so analysis only focuses on post-survey measures (α=.73; *M*=5.12, *SD*=1.43).

5.4.6. Identification with the podcast scientist

To determine whether participants were sensitive to the greater amount of personal narrative information in the PN podcast, we included a three item measure of identification with the podcast scientist in the post-survey: "How much do you identify with the scientist you heard about in the podcast," "How much do you relate to the scientist you heard about in the podcast," "How much are you similar to the scientist in the podcast?" (α=.85; *M*=2.63, *SD*=.92). In addition, participants judged "how much scientific information" (*M*=3.64, *SD*=.90). and "how much personal information about the scientist" (*M*=2.90, *SD*=1.01) was "shared in the podcast episode that you listened to?" on a scale from 1 (none at all) to 5 (a great deal).

5.4.7. Science learning

We included a knowledge assessment based on the scientific content of the podcast which included three multiple choice items. We created a ***science learning measure***, which was a sum of each correct response and ranged from 0 to 3 (*M*=1.96, *SD*=.96; across participants, 29% scored 0-1, 36% scored 2, and 34% scored 3).

5.4.8. Stereotypes of scientists

Measured along two dimensions: warmth and competence. In the post-test survey only, participants were asked, "When you think about scientists in general, how would you rate them on the following traits?" Judgments were reported in response to the prompt, "Compared to the Average Person, Scientists are," and reported on a scale from 1 (Much less) to 7 (Much more). Traits included "warm" and "friendly", which

were averaged together to form a warmth stereotype (α=.83; M=4.41, SD=1.16). Traits "competent" and "capable" were also averaged together to form a competence stereotype of scientists (α=.75; M=5.94, SD=1.02).

Parameter	Features in		Adapted from
	Pre-test	Post-test	
Anticipated belonging in science	Yes	Yes	***(30, 31)***
Self-efficacy in science	Yes	Yes	***(32)***
Perceptions of one's effort expenditure relative to other students in science class	Yes	Yes	***(30)***
Intentions to pursue a career in science	Yes	Yes	***(32, 33)***
Intentions to pursue additional higher education	Yes	Yes	***(32, 33)***
Identification with the podcast scientist	No	Yes	***(32)***
Science learning measure	No	Yes	***(This study)***
Stereotypes of scientists	No	Yes	***(34)***
Podcast feedback	No	Yes	***(This study)***
Demographics	No	Yes	***(This study)***

Table 5-2: Summary of measures evaluated through this study.

5.4.9. Podcast feedback

Assessed across three items about interest in the podcast: "How interesting was the podcast episode that you listened to", "How likely are you to seek out similar episodes from this podcast series", and "How likely are you to recommend the podcast to people you know?". Reponses were reported on a scale from 1 to 5 with higher numbers reflecting more positive impressions of the podcast episode (α=.87; *M*=2.98, *SD*=1.03).

These measures are summarized in ***Table 5-2***. The Qualtrics platform was used to develop and distribute the survey instrument. The survey logic of the post-test ensured that participants could not progress in the survey until they had spent time equal to the length of their assigned podcast episode.

5.5 Results

5.5.1. Listeners of shared demographic identity identify with the scientist

Using two-group comparisons in ANCOVA, controlling for pre-test scores, we assessed whether listening to the PN podcast (compared to the SO podcast) increased identification with the scientist in the podcast. Greater identification with the speaker was our hypothesized mediator, yet there was no difference in identification with the podcast scientist between the PN podcast M=2.70, SD=.91, and the SO podcast, M=2.55, SD=.93; $F(1, 158)$=.98, p=.324.

However, people did detect the difference in personal information between the two conditions, serving as a manipulation check to ensure the PN condition was in fact perceived as markedly different from the SO condition. Participants in the PN condition (M=3.38, SD=.94) reported having heard more personal information about the scientist than those in the SO condition (M=2.40, SD=.81); $F(1, 157)$=49.72, p<.001.

We also explored whether participant race or gender was a factor in their identification with the scientist in the podcast, who was an African American man. Although these effects did not interact with PN/SO conditions, there was a main effect of race, such that non-white participants (M=2.99, SD=1.09) reported greater identification with the scientist in the podcast than white participants (M=2.46, SD=.78);

$F(1, 156)=11.46$, $p=.001$. Further, when analyzed more closely, participants who identified as Black or African American specifically related more with the protagonist in both cases. Additionally, there was a main effect of gender in which men ($M=2.82$, $SD=1.00$) identified more with the scientist than women ($M=2.49$, $SD=.84$); $F(1, 146)=6.79$, $p=.010$.

Case	Participants	Mean	SD
Science Only	White	2.388	0.81
	Black/ African American	3.179	1.25
	Other minority groups	2.633	0.81
Personal Narrative	White	2.554	0.77
	Black/ African American	3.533	1.03
	Other minority groups	2.667	1.04

Table 5-3: Analysis of sense of identification with podcast by race/ ethnicity.

Notably, identification based on shared identity, namely race and gender, between listeners and the protagonist was more influential than a narrative recollection of the individual's personal journey in education – one that may have resonated with the experiences of students at community colleges by virtue of shared humble beginnings. The grammatical and phonological patterns of speech that lead a listener to identify the ethnicity of a speaker is well documented in the study of sociolinguistics(*35*). Therefore, even in the SO case, where the identity of the scientist was not explicitly mentioned, the participants likely would have identified the speaker as an African American male. This explains why in a condition independent manner, non-white and male participants more clearly identified with the scientist.

5.5.2. Sense of belonging and self-efficacy are unchanged in either case

We next assessed whether listening to the PN podcast (compared to the SO podcast) improved anticipated belonging in science, increased self-efficacy in science, decreased the perceptions of one's effort expenditure relative to other students, and increased intentions to pursue a career in science or further education.

Sense of belonging is an important factor is the recruitment, persistence and success of students, particularly from under-served communities(*36-39*). Hence, we first assessed whether the podcast had any effect on anticipated belonging in science. Controlling for pre-survey scores, we found that there was no difference between the two podcast treatments; $F(1, 157)=.64$, $p=.424$. Further, no change occurred in anticipated belonging in science before ($M=5.46$, $SD=1.52$) and after listening the PN podcast ($M=5.51$, $SD=1.10$).

Self-efficacy is a sense of confidence about one's own skills in an area of study or work that is central to their ability to successfully pursue higher education in that realm(*40-43*). The podcast condition did not alter the participants' self-efficacy in science either; $F(1, 157)=0.00$, $p=.983$. Participants reported similar levels of self-efficacy in science both before ($M=3.79$, $SD=1.15$) and after listening the PN podcast ($M=3.80$, $SD=1.05$).

A complementary assessment of efficacy is to measure how much effort they perceive they need to expend relative to their peers specifically in science education(*30*). This also did not change as a result of which podcast treatment participants were randomly assigned to hear; $F(1, 157)=1.71$, $p=.193$. For those who heard the personal

narrative, their judgements of effort were similar both before (M=2.94, SD=.86) and after (M=2.95, SD=.80) listening to the PN podcast.

Interest in pursuing a career in science also did not change as a result of which podcast treatment participants were randomly assigned to hear; $F(1, 157)=2.57$, p=.111. For those who heard the personal narrative, their judgements of effort were similar both before (M=5.33, SD=1.86) and after (M=5.12, SD=1.78) listening to the PN podcast. Interest in pursuing additional higher education was also the same across the two conditions, $F(1, 158)=0.10$, p=.750.

Finally, we did not observe any differences in students' view of scientists across the two podcast treatments. Stereotypes of scientists as warm were similar in both the SO and the PN conditions, SO M=4.28, SD=1.12; PN M=4.54, SD=1.19; F(1, 158)=1.94, p=.165. Stereotypes of scientists as competent were not significantly different across the two podcast versions: SO M=5.84, SD=1.00; PN M=6.04, SD=1.03; F(1, 158)=1.50, p=.223. Overall, participants perceived scientists as more competent than warm, t(159)=15.05, p<.001.

Our hypothesis was that a sense of identification with a successful scientist from an under-represented identity may lead to an increased feeling of belonging and self-efficacy in STEM among students with shared experiences in community colleges. However, we found that beyond shared experiences such as those associated with being a first-generation graduate student, common identity in ethnicity and gender may be over-riding the ability to identify with and relate to the African American male protagonist. Therefore, in a participant population self-identified as 57% women and

68% white, it is somewhat unsurprising that these facets of belonging and self-efficacy are unmoved.

5.5.3. SO podcast catalyzes increased scientific learning

Despite the identical scientific content in both podcasts, participants in the PN condition perceived their podcast as sharing less scientific information (M=3.38, SD=.87) than those in the SO condition (M=3.91, SD=.86); $F(1, 156)=14.52$, $p<.001$. Therefore, we notice that the addition of a personal narrative seems to lead to the perspective that the scientific content may be more dilute. We next explored whether there was an effect of condition on performance on the science learning assessment, which included three questions based on the podcast's science content. Despite the identical scientific content, participants who listened to the shorter SO podcast (M=2.14, SD=.82) performed better than those in the PN podcast condition (M=1.78, SD=1.05); $F(1, 158)=5.20$, $p=.017$.

Narrative storytelling has been known to improve student engagement, cognition, retention of knowledge, and organizational structure across a wide range of disciplines, including physical sciences(*44-46*). Hence, we hypothesized that having the personal narrative of the scientist woven into their scientific accomplishments might assist in higher level of science learning and retention. However, we found that on the contrary, participants performed better at recalling facts from the shorter SO podcast. One significant difference between the SO and PN test cases is the length of the audio clips, where the former is roughly 15 minutes long and the latter is 30 minutes. Studies that analyze the attention span of students or listeners of a lecture(*47*) or a podcast(*48, 49*) are at best divided about the ideal length of an oratory engagement. Therefore, in the design of this study, we may have underestimated the influence of the difference in

length of the two audio clips. Perhaps this difference in length may have influenced science retention.

5.5.4. Podcast feedback highlights appetite for science content

Perhaps reflective of the difference in the content and length of the two podcasts, participants provided more positive feedback in response to the science only podcast compared to the personal narrative podcast, including metrics evaluating if they might seek out similar episodes and recommend these to people they know; PN condition (M=2.85, SD=1.00) and in SO (M=3.15, SD=1.04), $F(1,157)=4.40$, $p<.05$.

Perhaps reflective of the difference in the content and length of the two podcasts, participants provided more positive feedback in response to the science only podcast compared to the personal narrative podcast, including metrics evaluating if they might seek out similar episodes and recommend these to people they know; PN condition (M=2.85, SD=1.00) and in SO (M=3.15, SD=1.04), $F(1,157)=4.40$, $p<.05$.

We also included an open-ended question soliciting feedback on the podcast. This was meant to be used as a proxy to identify what aspects of the narratives stood out in the minds of the listeners enough to comment on. In the PN case, 14 participants commented on how Dr. Robinson's personal narrative was moving to them, while 3 participants commented on the scientific content of the podcast, 2 participants explicitly referred to both the science and the personal story, while 13 others provided general comments. Not surprisingly, the SO case did not receive any feedback that was personal in nature. Instead, the comments were general (n = 32) or focused on the science (n = 4) (randomly picked excerpts summarized in ***Table 5-4***).

PN; Science related excerpts	ID
"I liked it overall, it was easy enough to understand if you don't have a prominent science background, but complicated enough for people with the science background."	F, N/P/A/H/W
"It was a great podcast!!! I was also introduced into a very complex form of research. The research I do (sic) are with bNAbs to prevent HIV infection, so this was very relatable."	M, H
PN; Personal Story related excerpts	**ID**
"….interesting to hear how just one person changed someones (sic) life. It was a nice thing to listen to and gave me hope after I struggled my first year at college."	M, W
"I like that the scientist got to show his story in how it's like about looking forward in graduate school from his professors looking into him"	F, W
"I liked that the podcast made personal references to his life and wasn't just 100% science talk"	F, A
'…….the focus did seem to be more on humanizing scientists then it was on sharing interesting scientific ideas'	M, W
'I really enjoyed listening to a bit of MP's story and related to it'	F, A/H/W
'….It was interesting to hear how his background has affected his motivation and path…'	F, W
'…..Sharing a story like M.P was really inspiring because he is someone that I can relate to. It made me feel like I can go out there and do great things…..'	F, A/H

"…..Great to hear about a minority's inspirational rise through education. I can't relate to this at a great extent but I certainly understand where he's coming from….."	M, H
"Very inspiring!! I believe stories such as these motivative (sic) students especially students of color trying to make a positive impact in known to be white dominant fields. As a person of color….."	M, A
PN; Science and Personal story related excerpts	**ID**
"….I loved hearing MP's story as well as learning about his research…."	F, H
"It was interesting learning about new ways to more "ethically" produce antibodies which have great use in science. It was also great to hear the perspective and personal life of a young graduate scientist who struggles like everyone else."	F, H
PN; General excerpts	**ID**
"I enjoyed it overall"	M, W
"I think at points the podcast was a little difficult to pay attention to, it was easy to get distracted and stop listening, I think the subject matter was a bit dull for the common person."	F, W
"I thought it was ok ~ it was not made too difficult to follow and understand."	F, W
"It was an interesting episode but most of the information I was already aware of so it was also slightly boring."	M, W
SO; Science related excerpts	**ID**
"having visuals of the workings of antibodies"	F, W
"i really like the podcast it was really interesting i would like to learn more about antibodies"	F, A

SO; Personal Story related excerpts	ID
(None)	
SO; Science and Personal story related excerpts	**ID**
(None)	
SO ; General excerpts	**ID**
"I believe it was all right. I would recommend it more as a use of studying."	M, H
"I enjoyed the podcast a whole lot, I appreciate the endeavors of the scientific team……."	M, A
"I enjoyed the podcast, the information was laid out very clearly. The scientist clearly knew his subject matter very well and communicated it excellently."	M, W
"I like podcasts, it's a great way to learn on the go."	F, W
"I think the podcast is a great way to make science accessible."	F, W
"Is a bit dry, could use more common examples to keep audiences listening."	F, W

Table 5-4: Excerpts from feedback provided by participants towards the PN and SO episodes included in the study.

The gender and racial identity (ID) of the participant quoted is noted with gender (M for Male, F for Female, and O for Other) comma race (A for Black/ African American, H for Hispanic/ Latin American, N for American Indian/ Alaskan Native, P for Asian/ Pacific American, W for White American). Multi-racial individuals are denoted by all self-identified races separated by a back-slash. Quotes were randomly curated and then ascribed demographic markers.

The feedback from participants points to a keen interest among community college students in listening to science podcasts. Further, comments on the PN condition indicate that the personal narrative inspired at least some of the listeners. Multiple participants also referred to other podcasts that they enjoy listening to in their comments.

Especially given the uniquely vulnerable nature of students in community colleges, these results inspire a focused effort in introducing podcasts as an additional modality of instruction, extending opportunities to learn and be inspired beyond their limited time on campus.

5.5.5. Correlation of posttest variables marks difference in effect of the SO vs PN case

As summarized in ***Table 5-5***, the degree of listeners' identification with the scientist in the SO or PN case, correlated with a sense of belonging and a feeling of self-

	Variable	1	2	3	4	5	6	7	8
1	Identification with podcast scientist	--	.27*	.37**	**-.32****	.34**	.40***	**.27***	.46***
2	Belonging in science	.52***	--	.52***	-.31**	.56***	.23*	.26*	.41***
3	Self-efficacy in science	.43***	.58***	--	-.60***	.60***	.27*	.33**	.26*
4	Perceptions of effort expenditure relative to other students	**-.07**	-.04	-.35**	--	-.37**	-.14	-.33**	-.18
5	Intentions to pursue a career in science	.32**	.37**	.42***	.09	--	.36**	.28*	.36**
6	Intentions to pursue additional education	.26*	.18	.20	.04	.40***	--	.09	.29**
7	Podcast science knowledge	**-.00**	.10	.16	-.16	.10	-.00	--	.21
8	Podcast feedback	.47***	.43***	.44***	-.18	.34**	.21	.13	--

Table 5-5: Correlation matrix of posttest variables across the SO and PN cases. *Personal Narrative condition is above the diagonal (n=82), Science Only condition is below the diagonal (n=78). Key differences between these cases are highlighted in red. *p<.05, **p<.01, *** p<.001*

efficacy in science, intentions to pursue a career in science, and positive feedback towards the podcast. While the extent of this correlation with podcast feedback was comparable, among participants who listened to the SO podcast, identification with the protagonist was more closely associated with the increase in sense of belonging and self-efficacy in science. Uniquely, the extent of students' identification with the protagonist's portrayal in the PN podcast inversely correlated with the perceptions of effort expenditure, suggesting that the focus of the podcast story on the journey elicits feelings more in line with the effort and journey of education. Identification also correlated more closely with retention of science knowledge among those in the PN condition compared to the SO condition. Given the greater focus on narrative in the PN condition, feeling connected to the speaker may have been more impactful for attention and knowledge retention compared to the SO condition.

5.6 Discussion

Our hypothesis was that creating a podcast that clearly highlights the identity of a URM scientist, blended with their scientific achievements might inspire an increased sense of belonging and self-efficacy, thereby promoting persistence in STEM education. We found that even in the case of the SO condition that did not explicitly identify the demographics of the protagonist, non-white and male participants indicated an increased sense of identification with the scientist. This shows that sociolinguistic cues that come across in the course of a conversation through characteristic phonology, grammar among other patterns of speech are sufficient cues to promote a sense of similarity with the speaker. We also found that even though community college students may share key aspects of Dr. Robinson's lived experience, including being a first-generation graduate student with little expertise from his personal relationships to support his pursuit of higher education, the participants appear to have more strongly affiliated with race and

gender-based identities rather. Further, exploratory analysis of the bivariate correlations of posttest variables suggested that focusing on the journey and experiences of a scientist may draw attention to the effort associated with and journey in STEM education that can prove to be more difficult to capture comprehensively, given the greater diversity in personal experiences compared to gender or racial diversity.

Given this understanding, it is not surprising that our sample of predominantly female, and white students, reported only a limited sense of identification with the scientist in the podcast. Given our hypothesis that identification would function as a mediator, it logically follows that neither conditions had any impact on their sense of belonging or self-efficacy overall. Future studies should examine the effect of narratives among a population of students that predominantly match the portrayed scientist's gender and race.

However, it is reassuring to note that perhaps the format of traditional science podcasts need not change dramatically to include a personal narrative that overtly talks about the effects of an individual's identity and experiences thereof. Perhaps it is sufficient to host a diversity of scientists on brief episodes that talk about their scientific accomplishments, since sociolinguistic cues alone may paint a sufficient picture about the identity of the speaker. Thus, the inclusion of diverse speakers can communicate the message that anybody, irrespective of their identity can be successful and can belong in STEM.

Further, a single audio intervention may be insufficient to move the needle given the pervasive forces of socialization that undermine the self-worth and confidence of URMs among many other social challenges, impeding their mobility and success in

education. Therefore, there is a need for more comprehensive intervention approaches that are dedicated to capturing the successes of diverse scientists as role models.

Finally, in keeping with broader trends in the society, community college students clearly have an appetite for podcasts that complement their learning on campus. Given that URMs heavily rely of community colleges as a means to higher education, we are presented with a golden opportunity to support the persistence and educational ambitions of these vulnerable students, therefore ushering in equity and diversity in higher education.

5.7 Acknowledgments

The authors are grateful for the time and effort spent by collaborators from community colleges in New York - Mr. Joe Spence (Transfer Counselor and Chair, Counseling Services, SUNY Broome Community College), Prof. Lalitha Jayant (Professor, Science Department, CUNY Borough of Manhattan Community College), Prof. Justin Fiene (Biology Department, Onondaga Community College), Prof. Buffy Quinn (Chemistry & Physical Sciences Department, Onondaga Community College), and Prof. Daniel Lopez (Math department, Brookdale Community College) - towards guiding this study, recruiting student participants and being insightful resources. We thank Dr. Joseph Brown for his efforts in creating the original podcast episode that was adopted for this study, and Dr. Amanda Barrett Wittman for her inputs.

5.8 Funding

This Engaged Cornell work was supported by an Engaged Graduate Student Grant from the Office of Engagement Initiatives (Cornell University) awarded to A.N. and a Robert Frederick Smith Initiative grant awarded to S.D. The podcast episode used

in this study is derived from the Science Blender series, funded by a Robert Frederick Smith Initiative grant awarded to A.N. by the Robert Frederick Smith School of Chemical and Biomolecular Engineering (Cornell University).

5.9 References

1. P. R. Bahr, G. Jackson, J. McNaughtan, M. Oster, J. Gross, **Unrealized Potential: Community College Pathways to STEM Baccalaureate Degrees.** *The Journal of Higher Education* 88, 430-478 (2017).
2. J. N. Schinske *et al.*, Broadening Participation in Biology Education Research: Engaging Community College Students and Faculty. *Cbe-Life Sciences Education* 16, (2017).
3. X. Wang, K. Wickersham, **Women, Community Colleges, and STEM Careers.** *New Directions for Institutional Research* 2018, 73-90 (2018).
4. E. Hoffman, S. Starobin, F. S. Laanan, M. Rivera, Role of community colleges in STEM education: Thoughts on implications for policy, practice, and future research. 16, 85-96 (2010).
5. S. Starobin, F. S. Laanan, From community college to Ph.D.: Educational pathways in science, technology, engineering, and mathematics. 16, 67-84 (2010).
6. J. Lester, Women in male-dominated career and technical education programs at community colleges: Barriers to participation and success. 16, 51-66 (2010).
7. B. W.-L. Packard, J. L. Gagnon, O. LaBelle, K. Jeffers, E. Lynn, **Women's experiences in the stem community college transfer pathway.** 17, 129-147 (2011).
8. L. S. Hagedorn, W. Maxwell, P. Hampton, **Correlates of Retention for African-American Males in Community Colleges.** *Journal of College Student Retention: Research, Theory & Practice* 3, 243-263 (2001).

9. C. M. Loo, G. Rolison, **Alienation of Ethnic Minority Students at a Predominantly White University.** *The Journal of Higher Education* 57, 58-77 (1986).

10. K. L. Odom, L. M. Roberts, R. L. Johnson, L. A. Cooper, Exploring obstacles to and opportunities for professional success among ethnic minority medical students. *Acad Med* 82, 146-153 (2007).

11. M. F. Summers, F. A. Hrabowski, **Preparing Minority Scientists and Engineers.** *Science* 311, 1870-1871 (2006).

12. J. McFarland *et al.*, "The Condition of Education 2019 (NCES 2019-144)," (2019).

13. . (Edison Research), vol. 2018.

14. J. Johnson, L. Ross, W. Iwanenko, J. Schiffert, A. Sen, **Are podcasts effective at educating African American men about diabetes?** *Am J Mens Health* 6, 365-367 (2012).

15. S. S. Prakash, N. Muthuraman, R. Anand, Short-duration podcasts as a supplementary learning tool: perceptions of medical students and impact on assessment performance. *BMC Med Educ* 17, 167 (2017).

16. M. V. Dantas-Queiroz, L. C. P. Wentzel, L. L. Queiroz, **Science communication podcasting in Brazil: the potential and challenges depicted by two podcasts.** *An Acad Bras Cienc* 90, 1891-1901 (2018).

17. I. Picardi, S. Regina, **Science via podcast.** *Journal of Science Communication* 7, (2008).

18. C. Evans, The effectiveness of m-learning in the form of podcast revision lectures in higher education. *Computers & Education* 50, 491-498 (2008).

19. K. F. Hew, Use of audio podcast in K-12 and higher education: a review of research topics and methodologies. *Educational Technology Research and Development* 57, 333-357 (2009).

20. B. C. Jham, G. V. Duraes, H. E. Strassler, L. G. Sensi, **Joining the Podcast Revolution.** *Journal of Dental Education* 72, 278 (2008).

21. M. J. W. Lee, C. McLoughlin, A. Chan, **Talk the talk: Learner-generated podcasts as catalysts for knowledge creation.** *British Journal of Educational Technology* 39, 501-521 (2008).

22. E. Berger, Podcasting in Engineering Education: A Preliminary Study of Content, Student Attitudes, and Impact. *Innovate: Journal of Online Education* 4, (2007).

23. A. A. Carvalho *et al.* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009), pp. 417-426.

24. V. Fernandez, J. M. Sallan, P. Simo, in *Exploring Learning & Teaching in Higher Education,* M. Li, Y. Zhao, Eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2015), pp. 305-330.

25. V. Fernandez, P. Simo, J. M. Sallan, Podcasting: A new technological tool to facilitate good practice in higher education. *Computers & Education* 53, 385-392 (2009).

26. J. E. L. Shin, S. R. Levy, B. London, **Effects of role model exposure on STEM and non-STEM student engagement.** *Journal of Applied Social Psychology* 46, 410-427 (2016).

27. A. N. Herrera, Ed.D., University of the Pacific, Ann Arbor (2012).

28. I. R. Johnson, E. S. Pietri, F. Fullilove, S. Mowrer, **Exploring Identity-Safety Cues and Allyship Among Black Women Students in STEM Environments.** *Psychology of Women Quarterly* 43, 131-150 (2019).

29. S. D. Herrmann *et al.*, The Effects of a Female Role Model on Academic Performance and Persistence of Women in STEM Courses. *Basic and Applied Social Psychology* 38, 258-268 (2016).

30. J. L Smith, K. Lewis, L. Hawthorne, S. Hodges, When Trying Hard Isn't Natural: Women's Belonging With and Motivation for Male-Dominated STEM Fields As a Function of Effort Expenditure Concerns. (2012), vol. 39.

31. G. M. Walton, G. L. Cohen, **A question of belonging: Race, social fit, and achievement.** *Journal of Personality and Social Psychology* 92, 82-96 (2007).

32. J. G. Stout, N. Dasgupta, M. Hunsinger, M. A. McManus, STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *J Pers Soc Psychol* 100, 255-270 (2011).

33. M. Estrada, A. Woodcock, P. R. Hernandez, P. W. Schultz, Toward a Model of Social Influence that Explains Minority Student Integration into the Scientific Community. *J Educ Psychol* 103, 206-222 (2011).

34. K. Rios, Z. H. Cheng, R. R. Totton, A. F. Shariff, **Negative Stereotypes Cause Christians to Underperform in and Disidentify With Science.** *Social Psychological and Personality Science* 6, 959-967 (2015).

35. in *Language and Ethnicity,* C. Fought, Ed. (Cambridge University Press, Cambridge, 2006), pp. 19-42.

36. L. R. M. Hausmann, J. W. Schofield, R. L. Woods, Sense of Belonging as a Predictor of Intentions to Persist Among African American and White First-Year College Students. *Research in Higher Education* 48, 803-839 (2007).

37. M. Hoffman, J. Richmond, J. Morrow, K. Salomone, **Investigating "Sense of Belonging" in First-Year College Students.** *Journal of College Student Retention: Research, Theory & Practice* 4, 227-256 (2002).

38. D. R. Johnson *et al.*, Examining Sense of Belonging Among First-Year Undergraduates From Different Racial/Ethnic Groups. *Journal of College Student Development* 48, 525-542 (2007).

39. J. M. Ostrove, A. J. Stewart, N. L. Curtin, **Social Class and Belonging: Implications for Graduate Students' Career Aspirations.** *The Journal of Higher Education* 82, 748-774 (2011).

40. A. Carpi, D. M. Ronan, H. M. Falconer, N. H. Lents, Cultivating minority scientists: Undergraduate research increases self-efficacy and career ambitions for underrepresented students in STEM. *Journal of Research in Science Teaching* 54, 169-194 (2017).

41. M. M. Chemers, E. L. Zurbriggen, M. Syed, B. K. Goza, S. Bearman, **The Role of Efficacy and Identity in Science Career Commitment Among Underrepresented Minority Students.** *Journal of Social Issues* 67, 469-491 (2011).

42. G. Garibay, University of Southern California, (2016).

43. D. MacPhee, S. Farro, S. S. Canetto, Academic Self-Efficacy and Performance of Underrepresented STEM Majors: Gender, Ethnic, and Social Class Patterns. *Analyses of Social Issues and Public Policy* 13, 347-369 (2013).

44. S. Nathanson, Harnessing the Power of Story: Using Narrative Reading and Writing across Content Areas. (2006).

45. C. Hill, L. Baumgartner, Stories in Science: The Backbone of Science Learning. (2009).

46. D. M. Smith, Wayne State University, (2013).

47. N. A. Bradbury, Attention span during lectures: 8 seconds, 10 minutes, or more? *Adv Physiol Educ* 40, 509-513 (2016).

48. K. Lien, A. Chin, A. Helman, T. M. Chan, A Randomized Comparative Trial of the Knowledge Retention and Usage Conditions in Undergraduate Medical Students Using Podcasts and Blog Posts. *Cureus* 10, e2065-e2065 (2018).

49. T. Charney, in *Current; News for People in Public Media,* D. Sefton, Ed. (Current; News for People in Public Media, 2016).

CHAPTER 6

Future perspectives

6.1 Chromosomal engineering of glycosylation pathways in E. coli

Since seminal work by Aebi and co-workers(*1*) of transferring a protein glycosylation pathway from *Campylobacter jejuni* to *Escherichia coli* around two decades ago, the field of bacterial glyco-engineering has generated a powerful set of tools for the study of complex glycosylation mechanisms, and the expression and characterization of diverse glycoproteins (reviewed in (*2-4*)). Briefly, our group and others facilitated this rapid growth by i) Curating a broad panel of natural and engineered glycosyl transferases that can specifically transfer a glycan onto a diverse set of relevant glycosites, including on authentic human glycosylation motifs. ii) Generating a repertoire of polysaccharide synthesis pathways to construct different glycans either on the lipid carrier undecaprenol phosphate (UndP) as substrate for *en bloc* transfer onto proteins, or for synthesis directly on target proteins. This includes bacterial glycans, bacterial O-antigens, and authentic human core glycans. iii) More recently, as the diversity of polysaccharides of interest has extended to include monosaccharides traditionally unavailable for transfer in *E. coli* such as UDP-N-acetyl galactosamine (UDP-GalNAc), GDP-fucose (GDP-Fuc), CMP-N-acetyl neuraminic acid (CMP-NeuNAc), or GDP-mannose (GDP-Man), we have worked towards assembling homologous pathways that maintain an active pool of these sugars as required.

Through these combined efforts, we have now arrived at a powerful glycoengineering platform in *E. coli* capable of producing therapeutically relevant glycoprotein products(*2, 3*), albeit one that is complex and not easily accessible for wider use. Therefore, the glyco-recoding strategy described here is not only important

for the clear fitness benefits and improved yields it affords, but also to reduce the number of moving parts, approaching the state of a glyco-competant host that can be purchased off the rack and employed to yield specific defined glycoconjugate products. In keeping with this idea of developing *E. coli* as a universal glycoprotein synthesis platform that is also readily accessible for use on demand, I believe our chromosomal engineering strategy should be extended to:

6.1.1. Synthesize therapeutically relevant, activated monosaccharides for glycosylation reactions

Host	CMP-NeuNAc pathway source	Conc. (μM)
MC4100 *ΔwecA ΔwaaL*	*(Negative control)*	0.302
MC4100 *ΔwecA ΔnanA* NeuDBAC	NeuDBAC from *E. coli* K1 @ O-PS	163.796
MC4100 *ΔwecA Δwaal ΔnanA* NeuDBAC	NeuDBAC from *E. coli* K1 @ O-PS	172.846
MG1655	*(Negative control)*	0.37
MG1655 O56@O-PS	NeuDBAC from *E. coli* O56 @ O-PS	17.682
Shuffle T7 *ΔwecA ΔwaaL*	*(Negative control)*	0.588
Shuffle T7 *ΔwecA ΔwaaL ΔnanA* NeuDBAC	NeuDBAC from *E. coli* K1 @ O-PS	134.8
Shuffle T7 *ΔwecA ΔnanA* NeuDBAC	NeuDBAC from *E. coli* K1 @ O-PS	165.772
Shuffle T7 *ΔwecA ΔnanA ΔwaaL::Kan* NeuDBAC	NeuDBAC from *E. coli* K1 @ O-PS	161.482
W3110	*(Negative control)*	0.594
W3110 *Δwaal* NeuDBAC	NeuDBAC from *E. coli* K1 @ O-PS	53.19
W3110 *ΔwecA ΔwaaL* NeuDBAC	NeuDBAC from *E. coli* K1 @ O-PS	83.522
ZLKA	*(Negative control)*	0.25
ZLKA	pMLByCon	683.956

Table 6-1: Estimation of CMP-NeuNAc concentration through LC-MS.
ZLKA was gift from Dr. Eric Samain, and is defined in previous work(5).

The focus of our current work was to highlight the ability to build diverse glycans on UndP, and to produce glycoproteins through chromosomally encoded pathways. However, perhaps a more basic requirement for a universal glycosylation system is to generate a strain that always maintains a steady pool of activated monosaccharides, particularly those that feature in glycoproteins sought after by the industry or research community. Identifying a growing interest in sialic acid containing glycoconjugates, I have generated strains that maintain a pool of UDP-NeuNAc, quantified through LC-MS in collaboration with Dr. Michael Vilkhovoy ***(Table 6-1)***. This is a starting place to now extend to include pathways that specifically maintain UDP-GalNAc, GDP-Man and GDP-Fuc, three monosaccharides that feature heavily in mammalian glycoproteins, but are unavailable for transfer in *E. coli*(*3*). This effort will provide a more universal strain for future endeavors in glyco-engineering.

6.1.2. Take a systems-level approach to streamline the metabolic flux of host strains for glycoprotein synthesis

Historically, most efforts in bacterial glycoengineering have assembled recombinant pathways for heterologous expression of glycoproteins in *E. coli*, paying little attention to its context in the scheme of the host's metabolic flux that has evolved with the objectives of survival and propagation. Certainly, the consideration of the effect of expressing non-native biomolecules in bacterial hosts is an old trade in the field of metabolic engineering, and some early efforts have been made to either knockout host genes that hinder the production of a target glycoconjugate (reviewed in (*3*)), and to optimize the expression levels of glycosyl-transferases that constitute the recombinant pathway(*6*). However, in an effort to truly streamline the inherent biochemical fluxes towards creating an optimized glyco-competent *E. coli* strain, broader systems-level approaches that evolve the host in a concerted fashion through strategies such as

Multiplex Automated Genomic Engineering (MAGE)(*7*) informed by existing bioinformatic models(*8*) is much needed. Here too, a host with glycosylation pathways stably integrated in the chromosome will be a unique asset, and serve as a worthy starting place.

6.2 Expanding O-glycoprotein expression in E. coli

Eukaryotic *O*-linked protein glycosylation takes place in the golgi apparatus and is elaborately controlled through tight compartmentalization of the requisite enzymes and relevant activated monosaccharides, and through controlled transport of proteins across the various stacks of the golgi or to the proteasome for degradation depending on the state of glycosylation(*9*). Further, often, proteins decorated with *O*-glycans bear multiple polysaccharides at proximal residues, where the addition of the first glycan can lead to a cascading effect catalyzing further glycosylation(*10*). Finally, unlike in the case of *N*-linked glycosylation, the identification of a clear glycosylation motif for key eukaryotic glycosyltransferases has been elusive. Taken together, glyco-engineering of proteins at *O*-glycosites has proven to be intractable until recently(*11, 12*).

The apparent challenges to replicating the complex compartmentalized eukaryotic *O*-glycosylation pathway in bacteria has further shifted our reliance on the use of bacterial enzymes that are natively functional in simpler organizational schemes. While the first bacterial oligosaccharyl transferase discovered was a catalyst of *O*-linked glycosylation(*13*), this area of study has remained rather subdued until recently(*14*). The last decade has seen the identification of multiple *O*-glycosylation pathways in bacteria, presenting the opportunity to grow glyco-engineering efforts in *E. coli* to include *O*-glycosylation of proteins. Our demonstration here of the transfer of the Core 1 glycan

onto human *O*-glycosites by *Ng*PglO and *Nm*PglL is a significant progress in the area of bacterial glycoengineering. This paves the way to two key follow up studies:

6.2.1. Comprehensively characterizing the acceptor sequence specificity of O-OSTs

Our current understanding of the acceptor sequence specificity of *O*-OSTs is limited to a single study conducted on *Nm*PglL to identify the MOOR glycosylation sequence(*15*). However, our work has found that both *Ng*PglO and *Nm*PglL are able to effectively glycosylate acceptor proteins that deviate significantly from the MOOR sequence, including those that do not contain hydrophilic flanking regions that had been prescribed to be essential. Therefore, evidently, our understanding of the catalytic potential of these *O*-OSTs, and therefore, by extension, that of the *O*-glycosylation pathway prescribed here is limited. Extending our ability to glycosylate proteins *in vitro* and combining this with the high-throughput nature of the SAMDI-MS technology (self-assembled monolayers with matrix-assisted laser desorption-ionization mass spectrometry) will help us truly define the acceptor sequence guidelines for these *O*-OSTs, similar to previous work on *N*-glycosyl transferases(*16*).

6.2.2. Expressing sialylated variants of the T and Tn antigens

Sialylated Tn antigen, sialylated T antigen and disialylated T antigen are recognized as tumor associated cancer antigens owing to their over-expression on MUC1 through aberrations in the glycosylation machinery as a result of cancer transformation(*17, 18*). Further, these glycans are known to be physiologically relevant to cancer progression. Finally, these sialylated glycans are also widely found in *O*-glycoproteins in healthy individuals, influencing the stability and folding of relevant proteins(*19*). Therefore, there is great interest in the ability to produce glycoproteins

bearing these glycoforms. Combining the CMP-NeuNAc synthesis pathways described above with bacterial derived sialyl-transferases provide the most promising route to expressing this glycoprotein. Putative bacterial sialyltransferases include the α(2,6) sialyltransferase from *Vibrionaceae photobacterium* sp. JT-ISH-224(*20*), α(2,3)/α(2,6) lipooligosaccharide sialyltransferase from *Neisseria meningitidis* serotype L1(*21*), and α(2,3) sialyltransferase from *Escherichia coli* O104(*22*).

6.3 Understanding YebF

Effective secretion of proteins through YebF bears significant value to applications in bioprocessing(*23, 24*), and has recently gained interest in therapeutic interventions(*25, 26*). Briefly, the opportunity to secrete functional enzymes allows the catalysis of chemistries that are i) impermeable to the cell (*23, 24*) and ii) incompatible with the survival of the host. Further, recent explorations attempt to use *E. coli* as a reactive probiotic intervention, engineering the secretion of bioactive molecules as fusions to YebF, where expression and secretion can be linked to specific environmental cues(*25*). While there is a growing interest in *E. coli*'s native ability to secrete YebF, the exact mechanism of the transfer of this protein and its fusion partners from the periplasm to the extracellular milieu remains unknown. It has been suggested that the outer membrane porins, Omp F, C and X may be involved in the secretion(*27*). However, since the Omp transport systems are commonly found affiliated with the movement of small molecules that are less than 600 da(*28*), and YebF is able to effectively transfer fusions of well over 70 kDa, I remain skeptical of the role of porins. In an effort to identify the secretion pathway associated with YebF, we can create a fusion with β-lactamase and isolate mutants that exhibit increased resistance to β-lactam antibiotics. Briefly, β-lactam antibiotics inhibit cell growth by affecting cell wall synthesis. However, β-lactamase expressed in the periplasm breaks down β-lactams and

promotes antibiotic resistance. Therefore, this scheme will enable the selection of mutants that retain β-lactamase in the periplasm and abrogates/minimizes extracellular secretion, thereby helping identify candidate secretion mechanisms.

6.4 Using a data driven approach to create podcasts to foster equity in higher education

Access to higher education can be a valuable catalyst for bringing about socio-economic mobility(*29, 30*), therefore fostering equity in the larger community. However, the path to higher education is lined with challenges thwarting the persistence of students from traditionally under-represented identities in STEM (URMs)(*31-33*). One key factor is the burden of isolation - by definition URMs are often one among few others that share a common identity and experiences thereof, thereby constantly made to like an outsider in academic spaces heavily influenced by majority populations(*33-35*). Often, institutions try to combat this by providing the few URM faculty in their roster a disproportionate responsibility of being visible, encouraging service related activities. This hidden burden borne by URM faculty further undermines their ability to succeed in their present roles(*36*), leading to a vicious cycle that exacerbates the original concern of underrepresentation of certain identities in academic institutions. Therefore, it is important to think outside the box to break this cycle. One strategy is to highlight the personal journey and successes of URMs through podcasts. Briefly, podcasts are a growing medium of information with the unique advantage of being widely accessible, often at no additional costs beyond reliable access to the internet or radio, and allowing engagement with a wide range of attention - both passively and actively. By capturing the diverse experiences and successes of URMs in a form that can exist in perpetuity, I propose we can create role models for aspiring students to look up to, while also minimizing the hidden burden on URM faculty. Further, individuals that are captured

in the podcast also expend a minimal effort of 2 to 4 hours interviewing and leave with a permanent record of their accomplishments. While, this idea, in keeping with current knowledge of the importance of role models to the success of URMs(*37-39*) and the salient features of a podcast clearly holds immense potential, there exists no formula for how to most effectively present role models through this modality. Therefore, in the presently growing phase of the era of podcasts, there is an urgent need to carefully explore how to best leverage this medium to support equity in higher education.

6.5 References

1. M. Wacker *et al.*, **N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli.** *Science* 298, 1790-1793 (2002).
2. C. M. Harding, M. F. Feldman, **Glycoengineering bioconjugate vaccines, therapeutics, and diagnostics in E. coli.** *Glycobiology* 29, 519-529 (2019).
3. J. H. Merritt, A. A. Ollis, A. C. Fisher, M. P. DeLisa, **Glycans-by-design: Engineering bacteria for the biosynthesis of complex glycans and glycoconjugates.** *Biotechnol Bioeng* 110, 1550-1564 (2013).
4. A. Natarajan, T. Jaroentomeechai, M. Li, C. J. Glasscock, M. P. DeLisa, **Metabolic engineering of glycoprotein synthesis in bacteria.** *Emerg Top Life Sci* 2, 419-432 (2018).
5. N. Fierfort, E. Samain, **Genetic engineering of Escherichia coli for the economical production of sialylated oligosaccharides.** *J Biotechnol* 134, 261-265 (2008).
6. C. J. Glasscock *et al.*, **A flow cytometric approach to engineering Escherichia coli for improved eukaryotic protein glycosylation.** *Metab Eng* 47, 488-495 (2018).
7. H. H. Wang *et al.*, **Programming cells by multiplex genome engineering and accelerated evolution.** *Nature* 460, 894 (2009).

8. J. A. Wayman, C. Glasscock, T. J. Mansell, M. P. DeLisa, J. D. Varner, **Improving designer glycan production in Escherichia coli through model-guided metabolic engineering.** *Metab Eng Commun* 9, e00088 (2019).

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

10. K. G. Ten Hagen, T. A. Fritz, L. A. Tabak, **All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases.** *Glycobiology* 13, 1R-16R (2003).

11. T. Du *et al.*, **A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans.** *Cell Chemical Biology* 26, 203-212.e205 (2019).

12. G. E. Henderson, K. D. Isett, T. U. Gerngross, **Site-specific modification of recombinant proteins: a novel platform for modifying glycoproteins expressed in E. coli.** *Bioconjug Chem* 22, 903-912 (2011).

13. P. Castric, **pilO, a gene required for glycosylation of Pseudomonas aeruginosa 1244 pilin.** *Microbiology* 141 (Pt 5), 1247-1254 (1995).

14. J. A. Iwashkiw, N. F. Vozza, R. L. Kinsella, M. F. Feldman, **Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation.** *Mol Microbiol* 89, 14-28 (2013).

15. C. Pan *et al.*, **Biosynthesis of conjugate vaccines using an O-linked glycosylation system.** *MBio* 7, e00443-00416 (2016).

16. W. Kightlinger *et al.*, **Design of glycosylation sites by rapid synthesis and analysis of glycosyltransferases.** *Nat Chem Biol*, (2018).

17. C. Fu *et al.*, **Tumor-associated antigens: Tn antigen, sTn antigen, and T antigen.** *Hla* 88, 275-286 (2016).

18. F.-G. Hanisch, S. Müller, **MUC1: the polymorphic appearance of a human mucin.** *Glycobiology* 10, 439-449 (2000).

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

20. Y. Kakuta *et al.*, **Crystal structure of Vibrionaceae Photobacterium sp. JT-ISH-224 alpha2,6-sialyltransferase in a ternary complex with donor product CMP and acceptor substrate lactose: catalytic mechanism and substrate recognition.** *Glycobiology* 18, 66-73 (2008).

21. L. Y.-C. Lin *et al.*, **Structure and Mechanism of the Lipooligosaccharide Sialyltransferase from Neisseria meningitidis.** *Journal of Biological Chemistry* 286, 37237-37248 (2011).

22. D. Czuchry *et al.*, **Identification and Biochemical Characterization of the Novel alpha2,3-Sialyltransferase WbwA from Pathogenic Escherichia coli Serotype O104.** *J Bacteriol* 197, 3760-3768 (2015).

23. T. N. Wang, M. Zhao, **A simple strategy for extracellular production of CotA laccase in Escherichia coli and decolorization of simulated textile effluent by recombinant laccase.** *Appl Microbiol Biotechnol* 101, 685-696 (2017).

24. Y. Xiao *et al.*, **Neurospora crassa tox-1 Gene Encodes a pH- and Temperature-Tolerant Mini-Cellulase.** *J Agric Food Chem* 64, 4751-4757 (2016).

25. E. J. Seo, S. Weibel, J. Wehkamp, T. A. Oelschlaeger, **Construction of recombinant E. coli Nissle 1917 (EcN) strains for the expression and secretion of defensins.** *Int J Med Microbiol* 302, 276-287 (2012).

26. D. Quintero *et al.*, **Co-Expression of a Chimeric Protease Inhibitor Secreted by a Tumor-Targeted Salmonella Protects Therapeutic Proteins from Proteolytic Degradation.** *J Microbiol Biotechnol* 28, 2079-2094 (2018).

27. G. Prehna *et al.*, **A protein export pathway involving Escherichia coli porins.** *Structure* 20, 1154-1166 (2012).

28. H. Nikaido, **Porins and specific channels of bacterial outer membranes.** *Molecular Microbiology* 6, 435-442 (1992).

29. N. Grawe, **Education and Economic Mobility.** (2008).

30. R. Haskins, H. Holzer, R. I. Lerman, **Promoting Economic Mobility by Increasing Postsecondary Education.** (2009).

31. D. MacPhee, S. Farro, S. S. Canetto, **Academic Self-Efficacy and Performance of Underrepresented STEM Majors: Gender, Ethnic, and Social Class Patterns.** *Analyses of Social Issues and Public Policy* 13, 347-369 (2013).

32. J. McFarland *et al.*, **"The Condition of Education 2019 (NCES 2019-144),"** (2019).

33. K. L. Odom, L. M. Roberts, R. L. Johnson, L. A. Cooper, **Exploring obstacles to and opportunities for professional success among ethnic minority medical students.** *Acad Med* 82, 146-153 (2007).

34. C. M. Loo, G. Rolison, **Alienation of Ethnic Minority Students at a Predominantly White University.** *The Journal of Higher Education* 57, 58-77 (1986).

35. M. F. Summers, F. A. Hrabowski, **Preparing Minority Scientists and Engineers.** *Science* 311, 1870-1871 (2006).

36. J. E. Rodríguez, K. M. Campbell, L. H. Pololi, **Addressing disparities in academic medicine: what of the minority tax?** *BMC medical education* 15, 6-6 (2015).

37. S. D. Herrmann *et al.*, **The Effects of a Female Role Model on Academic Performance and Persistence of Women in STEM Courses.** *Basic and Applied Social Psychology* 38, 258-268 (2016).

38. I. R. Johnson, E. S. Pietri, F. Fullilove, S. Mowrer, **Exploring Identity-Safety Cues and Allyship Among Black Women Students in STEM Environments.** *Psychology of Women Quarterly* 43, 131-150 (2019).

39. J. E. L. Shin, S. R. Levy, B. London, **Effects of role model exposure on STEM and non-STEM student engagement.** *Journal of Applied Social Psychology* 46, 410-427 (2016).

APPENDIX A

Complete set of peptide motifs and proteins tested for O-glycosylation

Full length proteins	
flDsbA-flEPO	
spMBP-flEPO	
spDsbA-flEPO	
flMBP-flEPO	
Native human glycoproteins	***Peptide motifs***
Choriogonadotropin beta chain-1	LPGPSDTP
Choriogonadotropin beta chain-2	LPSPSRLP
Choriogonadotropin beta chain-3	KAPPPSLP
Choriogonadotropin beta chain-4	DSSSSKAP
Complement C1 inhibitor	ILEVSSLP
Glycophorin A	EIVSISAS
Granulocyte-macrophage colony-stimulating factor-1	APARSPSPSTQP
Granulocyte-macrophage colony-stimulating factor-2	APARSPSP
Ig alpha-1 chain C region - 1	PCPVPSTP
Ig alpha-1 chain C region - 2	PTPSPSTP
Ig delta chain C region	PKAQASSV
Immunoglobulin heavy constant alpha 1-1	PCPVPSTP
Immunoglobulin heavy constant alpha 1-2	PTPSPSTP

Insulin-like growth factor binding protein 6 precursor	NPGTSTTP
Insulin-like growth factor II	SERDVSTP
Interferron gamma	QGVGVTET
Interferron gamma (Modified)	QGVGVSET
Kininogen	MPPISPAP
Lubricin	TTKSAPTT
Lysosome-associated membrane glycoprotein 1	SPSPSPVP
Phosphatidylcholine-sterol Acyltransferase	SPTASPEP
Plasminogen	TTPPPSSG
Plasminogen - 2	TPPPSSGP
RANTES	YSSDTTPCCCSNPVCREY INSLEMS
Thrombopoeitin	LTLSSPAP
Tissue factor plasminogen inhibitor	SLTPQSTK
Tissue kallikrein	VRSSSRTP
TNF-alpha	VRSSSRTP

Table A-1: List of proteins presented to NgPglO and NmPglL that were not glycosylated by either O-OSTs.

a. 4 full length (fl) proteins of EPO were tested as fusions to either the general secretor pathway (sec) signal peptides (sp) derived from DsbA or MBP, of full length DsbA or MBP. b. List of peptide motifs that capture native human O-glycosites that bear the core-1 glycan or its derivatives. These were all tested as fusions to spDsbA-YebF and presented in the context of the hydrophilic flanking regions recommended as part of the MOOR sequence.

APPENDIX B

Characterizing soluble expression of glycosyl-transferases

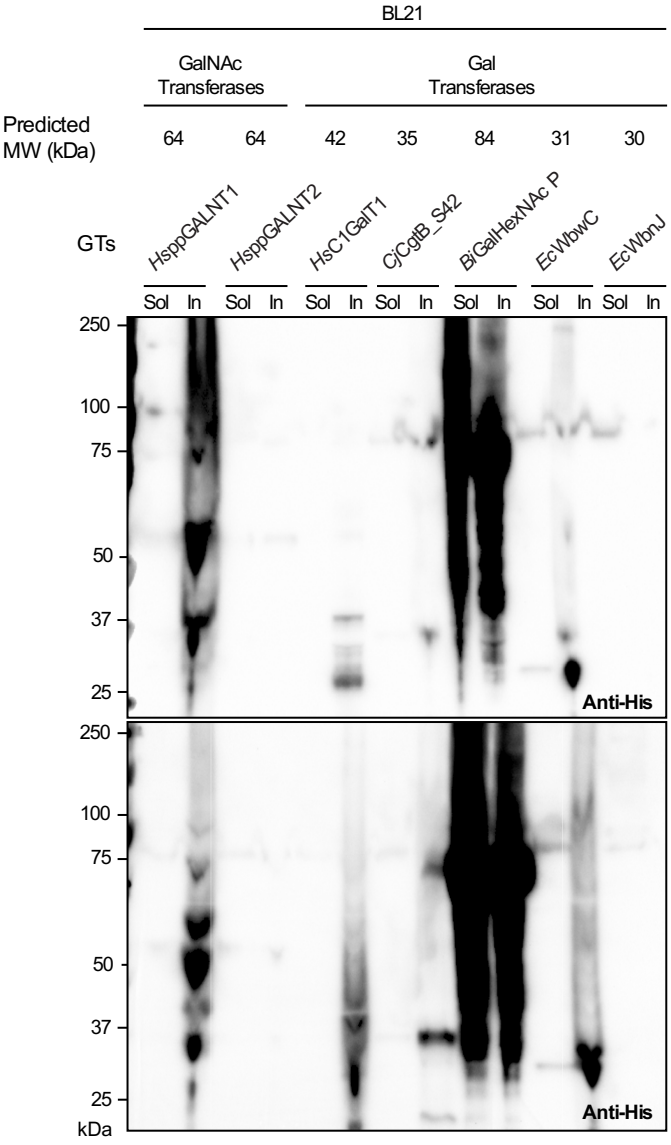


Figure B-1: Testing GALNTs and GalTs using protein extraction and western blot. *HsppGALNT1 - Homo sapiens ppGalNAcT1; HsppGALNT2 - Homo sapiens ppGalNAcT2; HsC1GalT1 - Homo sapiens core-1 GalT1; CjCgtB_S42 - evolved mutant of CgtB from Campylobacter jejuni; Bifidobacterium longum subsp. infantis ATCC 15697 GalHexNAcP; Escherichia coli O104 WbwC; Escherichia coli O86 WbnJ. Lower blot is overexposed. Sol refers to protein from the soluble fraction; In refers to protein extracted from the insoluble fraction.*

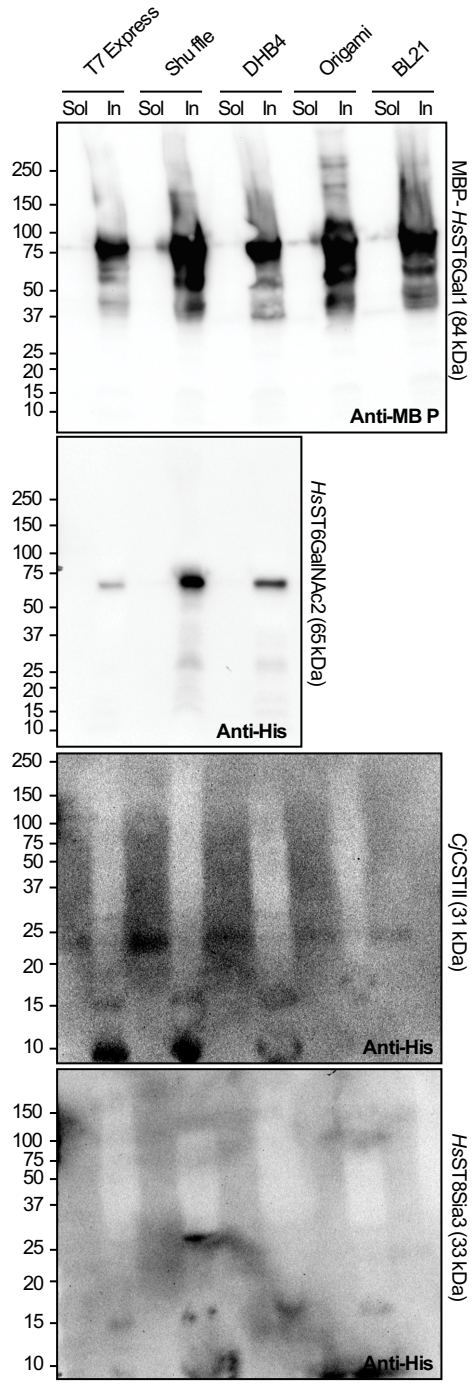


Figure B-2: Analyzing human and bacterial sialyl transferases across five different host strains using protein extraction and western blot.

HsST6GalI, HsST6GalNAc2 and HsST8Sia3 from Homo sapiens, and CjCSTII from Campylobacter jejuni. Their expected molecular weights are indicated in brackets. Sol refers to protein from the soluble fraction; In refers to protein extracted from the insoluble fraction.

Note: Panels of enzymes were curated by Aravind Natarajan and cloned into pET28a for protein expression towards extraction and western blots. Genes were later sub-cloned into pSALect through assistance from Ms. Marielisa Cabrera and Ms. Shannon Murphy. Spot plates presented here were generated by Mr. Robert Lee.

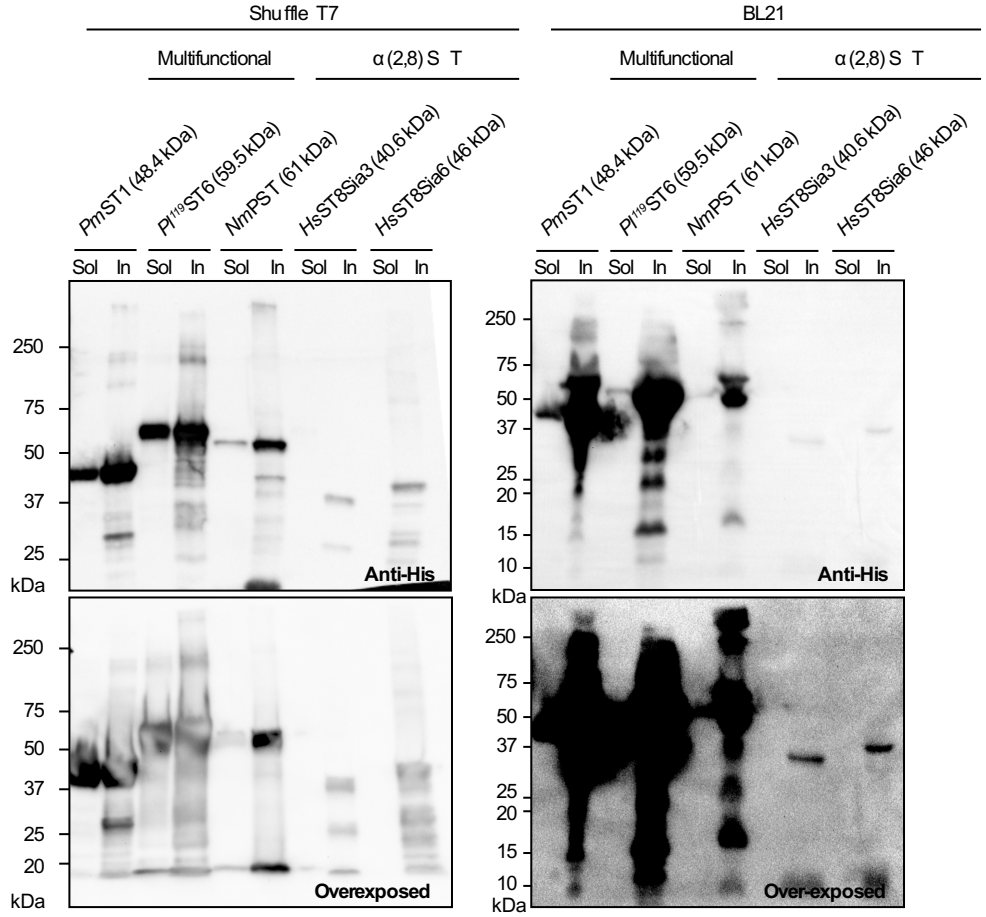


Figure B-3: Studying a panel of multifunctional STs, and STs that catalyze the attachment of sialic acid through an α(2,8) linkage using protein extraction and western blot.

PmST1 from Pasturella multicoda; Pl^{119}ST6 from Photobacterium leiognathi strain JT-SHIZ-119; NmPST from Neisseria meningitidis; HsST8Sia3 and HsST8Sia6 from Homo sapiens. Sol refers to protein from the soluble fraction; In refers to protein extracted from the insoluble fraction.

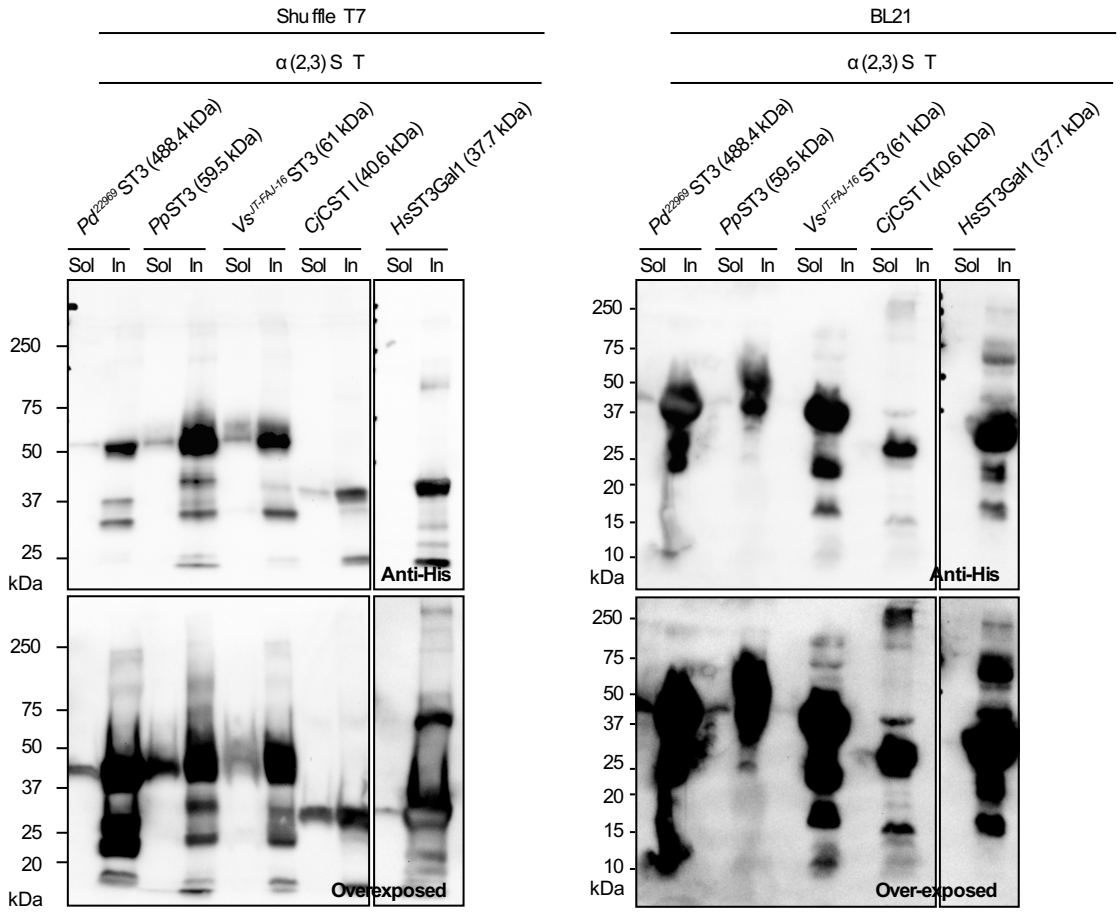


Figure B-4: Characterizing a panel of STs that catalyze the attachment of sialic acid through an α(2,3) linkage using protein extraction and western blot.

Pd^{22969}ST3 from Pasteurella dagmatis strain DSM 22969; PpST3 from Photobacterium phosphoreum; $Vs^{JT\text{-}FAJ\text{-}16}$ST3 from Vibrio sp. JT-FAJ-16; CjCSTI from Campylobacter jejuni; HsST3Gal1 from Homo sapiens. Sol refers to protein from the soluble fraction; In refers to protein extracted from the insoluble fraction.

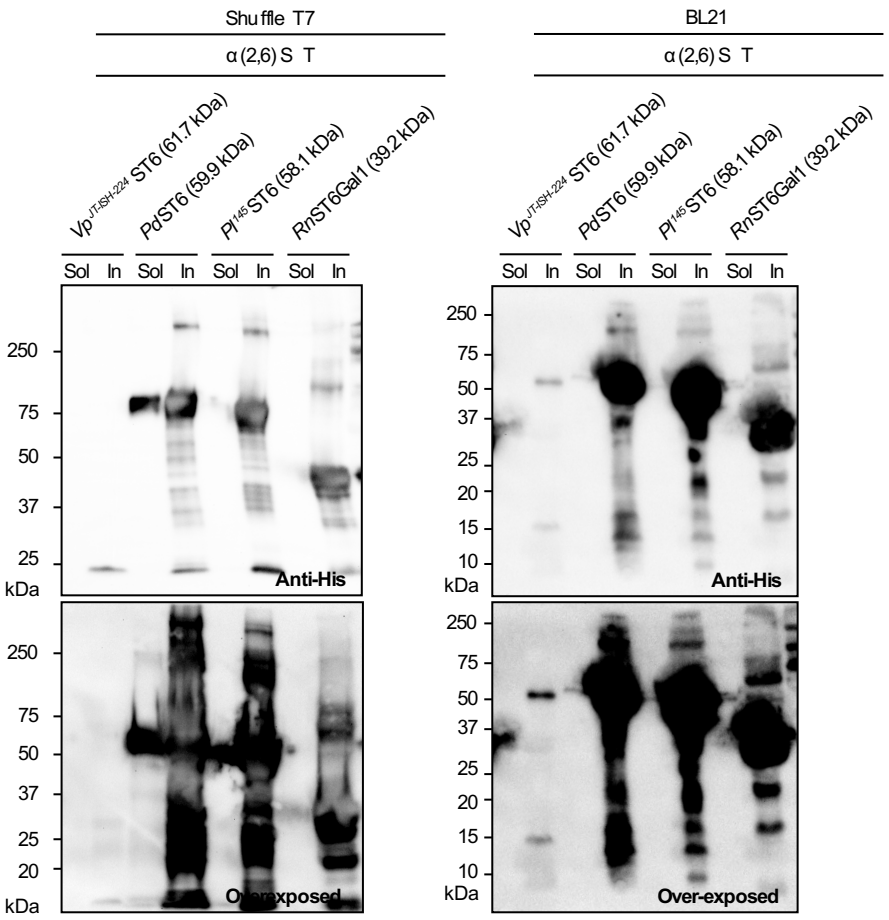


Figure B-5: Testing a panel of STs that catalyze the attachment of sialic acid through an α(2,6) linkage using protein extraction and western blot.

$Vp^{JT\text{-}ISH\text{-}224}$ST6 from Vibrionaceae Photobacterium sp. JT-ISH-224; PdST6 from Photobacterium damselae; Pl^{145}ST6 from Photobacterium leiognathi strain JT-ISH-224; RnST6Gal1 from Rattus norvegicus. Sol refers to protein from the soluble fraction; In refers to protein extracted from the insoluble fraction.

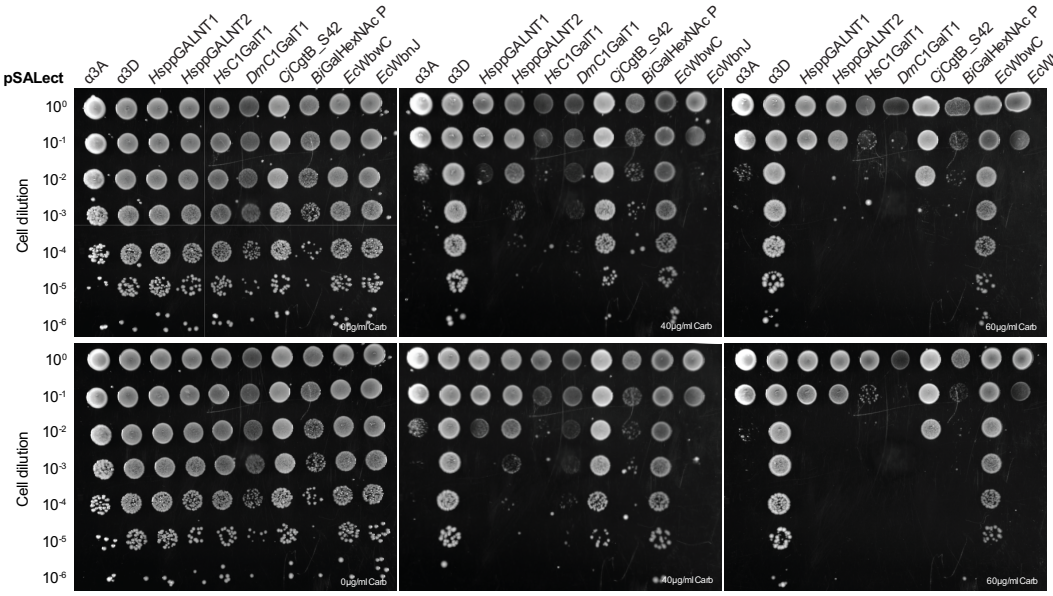


Figure B-6: Representative spot titer images that capture the degree of soluble expression of GALNTs and GalTs through resistance to Carbenicillin.

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3AD are synthetic proteins designed to be aggregation-prone and highly soluble respectively; HsppGALNT1 - Homo sapiens ppGalNAcT1; HsppGALNT2 - Homo sapiens ppGalNAcT2; HsC1GalT1 - Homo sapiens core-1 GalT1; DmC1GalT1 - Drosophila melanogaster core-1 GalT1; CjCgtB_S42 - evolved mutant of CgtB from Campylobacter jejuni; Bifidobacterium longum subsp. infantis ATCC 15697 GalHexNAcP; Escherichia coli O104 WbwC; Escherichia coli O86 WbnJ.

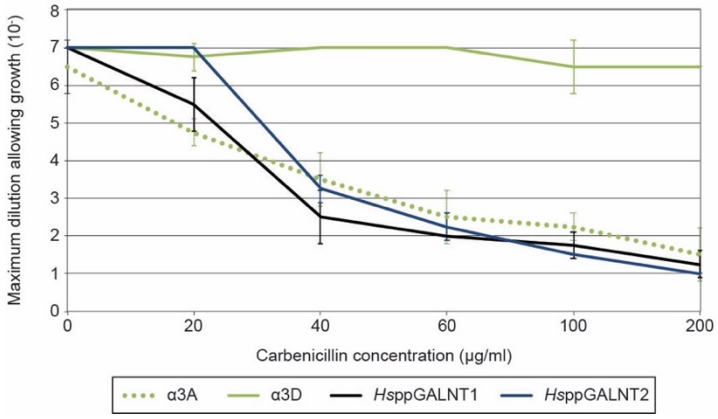


Figure B-7: Survival curves for serially diluted E. coli MC4100 cells expressing GalNacTs

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3D are synthetic proteins designed to be aggregation-prone and highly soluble respectively; HsppGALNT1 - Homo sapiens ppGalNAcT1; HsppGALNT2 - Homo sapiens ppGalNAcT2.

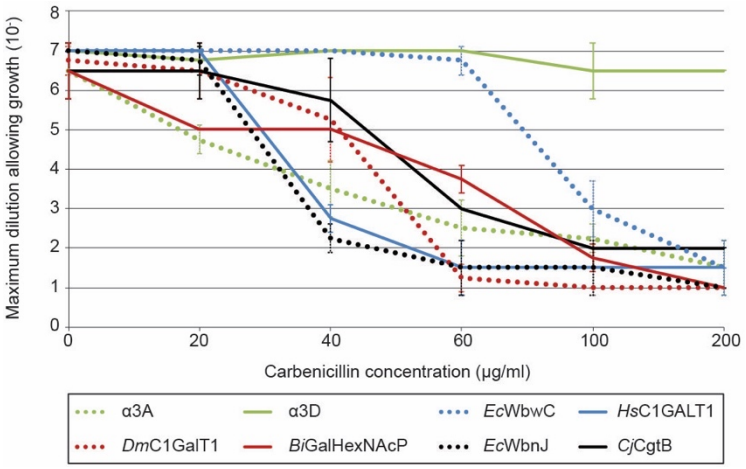


Figure B-8: Survival curves for serially diluted E. coli MC4100 cells expressing GalTs

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3D are synthetic proteins designed to be aggregation-prone and highly soluble respectively; CjCgtB_S42 - evolved mutant of CgtB from Campylobacter jejuni; Bifidobacterium longum subsp. infantis ATCC 15697 GalHexNAcP; Escherichia coli O104 WbwC; Escherichia coli O86 WbnJ.

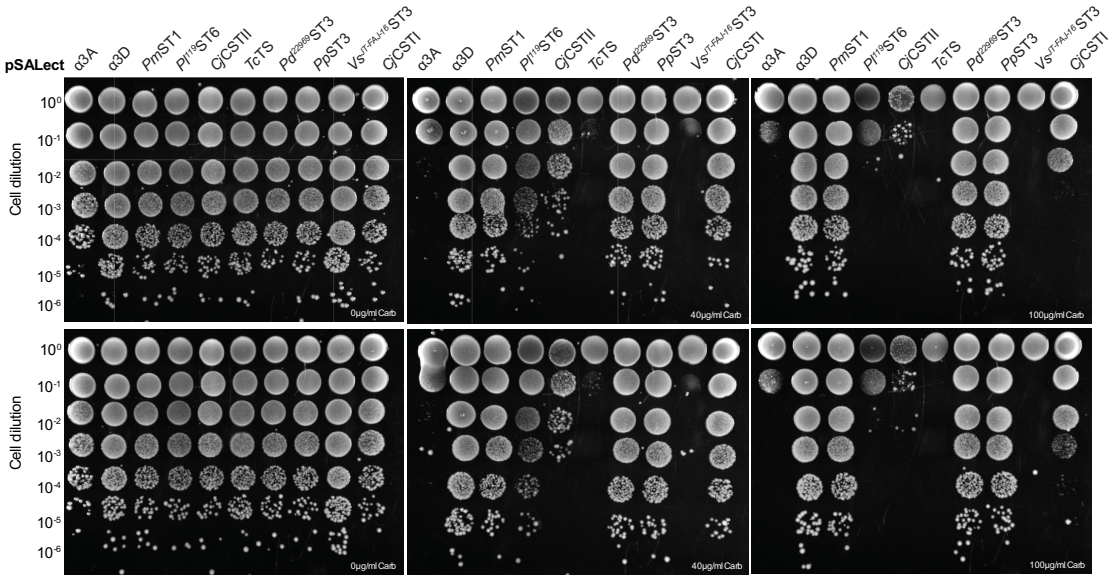


Figure B-9: Representative spot titer images that capture the degree of soluble expression of STs listed below (batch 1) through resistance to Carbenicillin.

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3AD are synthetic proteins designed to be aggregation-prone and highly soluble respectively; PmST1 from Pasturella multicoda; Pl^{119}ST6 from Photobacterium leiognathi strain JT-SHIZ-119; CjCSTII from Campylobacter jejuni; TcTs from Trypanosoma cruzi; Pd^{22969}ST3 from Pasteurella dagmatis strain DSM 22969; PpST3 from Photobacterium phosphoreum; $Vs^{JT\text{-}FAJ\text{-}16}$ST3 from Vibrio sp. JT-FAJ-16; CjCSTI from Campylobacter jejuni.

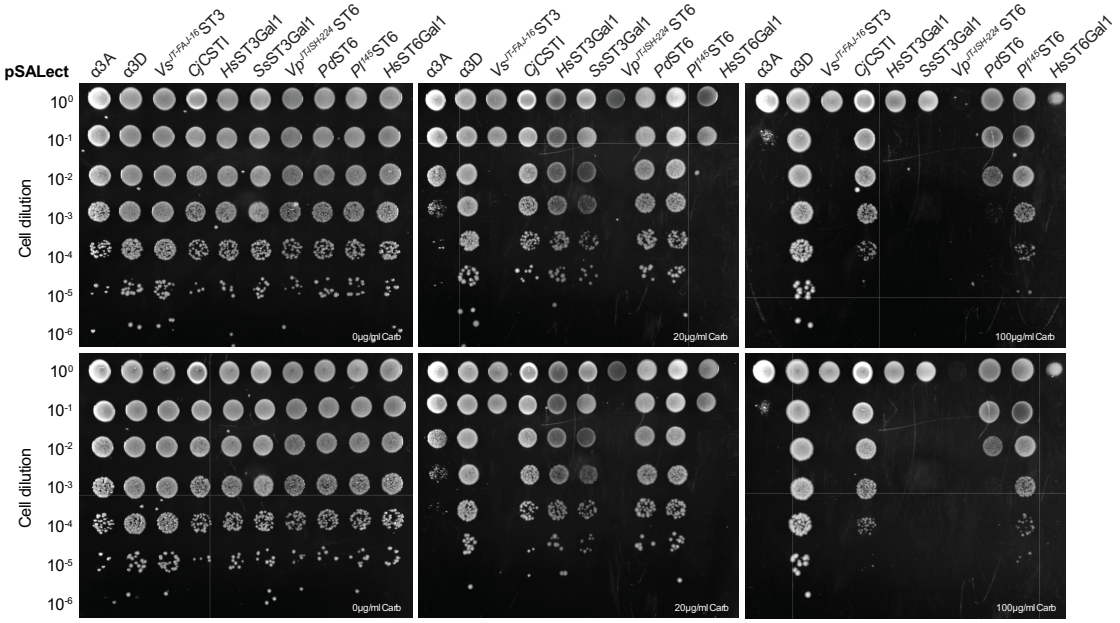


Figure B-10: Representative spot titer images that capture the degree of soluble expression of STs listed below (batch 2) through resistance to Carbenicillin.

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3AD are synthetic proteins designed to be poorly and highly soluble respectively; $Vs^{JT\text{-}FAJ\text{-}16}ST3$ from Vibrio sp. $JT^{\text{-}FAJ\text{-}16}$; CjCSTI from Campylobacter jejuni; HsST3Gal1 from Homo sapiens; SsST3Gal1 from Sus scrofa; $Vp^{JT\text{-}ISH\text{-}224}ST6$ from Vibrionaceae Photobacterium sp. JT-ISH-224; PdST6 from Photobacterium damselae; $Pl^{145}ST6$ from Photobacterium leiognathi strain JT-ISH-224; HsST6Gal1 from Homo sapiens.

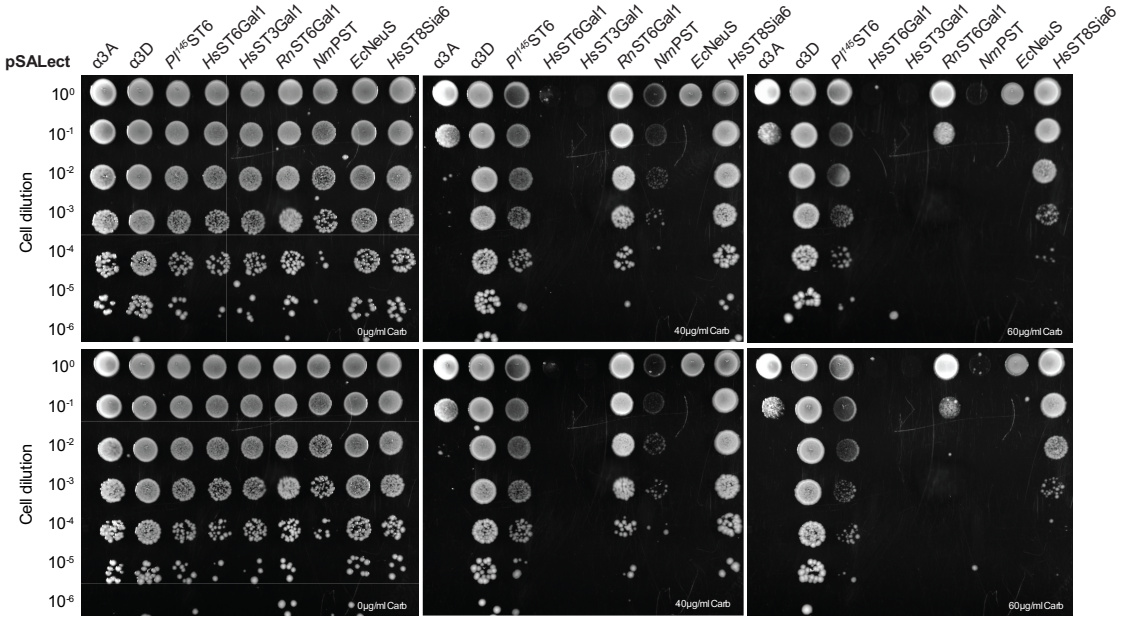


Figure B-11: Representative spot titer images that capture the degree of soluble expression of STs listed below (batch 2) through resistance to Carbenicillin.

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3AD are synthetic proteins designed to be aggregation-prone and highly soluble respectively; Pl^{145}*ST6 from Photobacterium leiognathi strain JT-ISH-224; HsST6Gal1 from Homo sapiens; HsST3Gal1 from Homo sapiens;* RnST6Gal1 from Rattus norvegicus; *NmPST from Neisseria meningitidis; EcNeuS from Escherichia coli; HsST8Sia6 from Homo sapiens.*

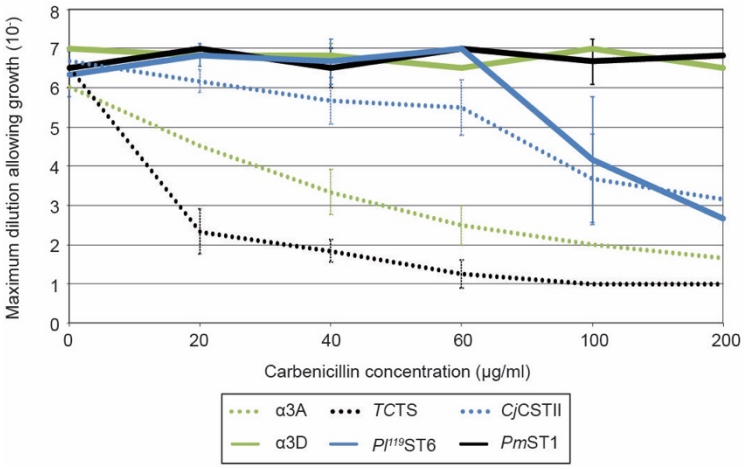


Figure B-12: Survival curves for serially diluted E. coli MC4100 cells expressing STs (set 1)

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3D are synthetic proteins designed to be aggregation-prone and highly soluble respectively; TcTs from Trypanosoma cruzi; PmST1 from Pasturella multicoda; CjCSTII from Campylobacter jejuni; Pl^{119}*ST6 from Photobacterium leiognathi strain JT-SHIZ-119.*

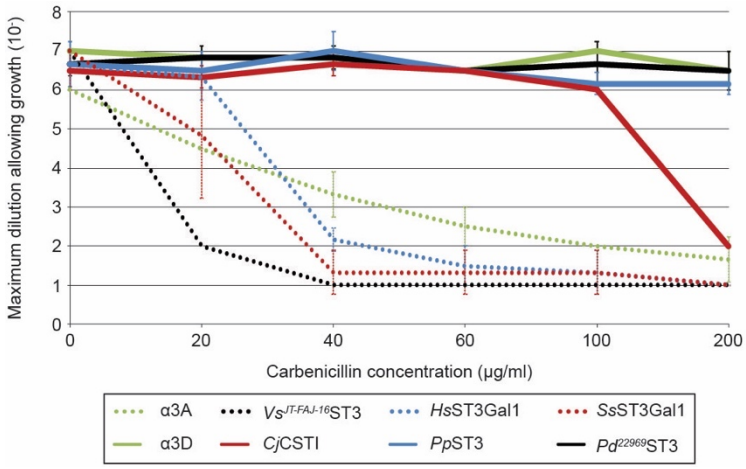


Figure B-13: Survival curves for serially diluted E. coli MC4100 cells expressing STs (set 2)

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3D are synthetic proteins designed to be aggregation-prone and highly soluble respectively; $Vs^{JT\text{-}FAJ\text{-}16}$ST3 from Vibrio sp. JT-FAJ-16; CjCSTI from Campylobacter jejuni; HsST3Gal1 from Homo sapiens; PpST3 from Photobacterium phosphoreum; SsST3Gal1 from Sus scrofa; Pd^{22969}ST3 from Pasteurella dagmatis strain DSM 22969.

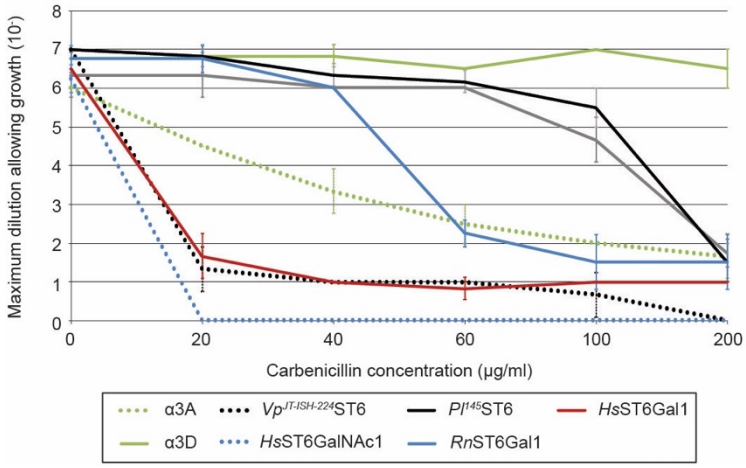


Figure B-14: Survival curves for serially diluted E. coli MC4100 cells expressing STs (set 3)

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3D are synthetic proteins designed to be aggregation-prone and highly soluble respectively; VpJT-ISH-224ST6 from Vibrionaceae Photobacterium sp. JT-ISH-224; HsST6GalNAc1 from Homo sapiens; Pl145ST6 from Photobacterium leiognathi strain JT-ISH-224; RnST6Gal1 from Rattus norvegicus; HsST6Gal1 from Homo sapiens.

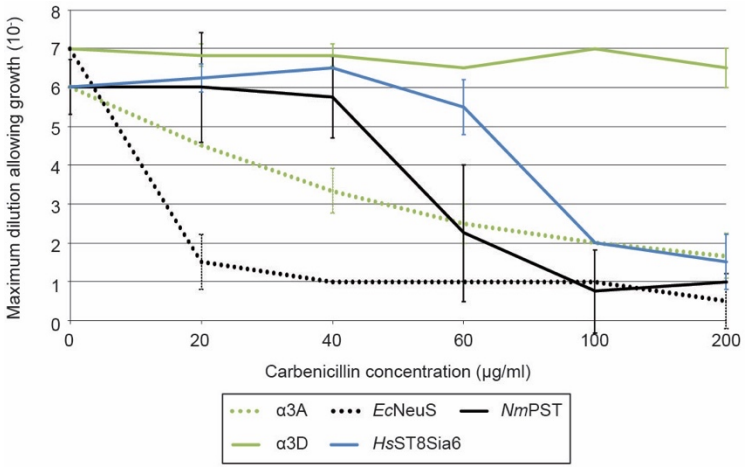


Figure B-15: Survival curves for serially diluted E. coli MC4100 cells expressing STs (set 4)

All proteins are expressed from pSALect as fusions to β-lactamase. α3A and α3D are synthetic proteins designed to be aggregation-prone and highly soluble respectively; EcNeuS from Escherichia coli; HsST8Sia6 from Homo sapiens, NmPST from Neisseria meningitidis.

APPENDIX C

Estimated marginal means of post-test measures based on gender and race

	Male	Female		White (racial majority)	Black/African American	Non-white racial minority (other than Black/African American)
Mean(*SE*)	(n=66)	(n=90)		(n=108)	(n=23)	(n=25)
Belonging in science*	5.65(.18)	5.54(.15)		5.59(.106)	5.72(.24)	5.47(.24)
Self-efficacy in science*	3.76(.12)	3.86(.10)		3.81(.07)	3.85(.16)	3.77(.15)
Perceptions of effort expenditure relative to other students*	2.85(.09)	2.97(.07)		2.91(.05)	3.14(.12)	2.69(.12)
Intentions to pursue a career in science*	5.23(.14)	5.39(.11)		5.21(.08)	5.59(.18)	5.14(.18)
Intentions to pursue additional education	5.25(.24)	5.43(.20)		4.94(.14)	5.38(.32)	5.70(.31)
Stereotypes of Scientists as Competent	5.75(.17)	6.07(.14)		5.91(.10)	6.15(.23)	5.67(.22)
Stereotypes of Scientists as Warm	4.60(.19)	4.49(.16)		4.33(.11)	4.97(.26)	4.34(.25)
Identification with podcast scientist	3.03(.14)	2.66(.12)		2.48(.08)	3.27(.19)	2.78(.18)
Podcast science knowledge	1.91(.16)	1.89(.13)		1.95(.09)	1.85(.21)	1.89(.21)
Podcast feedback	3.23(.17)	2.97(.13)		2.88(.10)	3.55(.22)	2.87(.21)

Table C-1: Estimated Marginal Means (and standard errors of the means) of post-test measures based on group comparison ANCOVA analyses considering condition, gender, and race/ethnicity as predictors (*controlled for pre-test scores).

M(*SE*)	African American Male (n=10)	African American Female (n=13)
Belonging in science	5.93(.38)	5.50(.30)
Self-efficacy in science	3.72(.24)	3.97(.20)
Perceptions of effort expenditure relative to other students	3.08(.19)	3.21(.15)
Intentions to pursue a career in science	5.68(.28)	5.50(.23)
Intentions to pursue additional education	4.94(.50)	5.82(.40)
Stereotypes of Scientists as Competent	5.77(.35)	6.53(.28)
Stereotypes of Scientists as Warm	5.02(.40)	4.91(.32)
Identification with podcast scientist	3.44(.29)	3.11(.24)
Podcast science knowledge	1.98(.33)	1.71(.27)
Podcast feedback	3.54(.34)	3.56(.28)

Table C-2: Estimated Marginal Means (and standard errors of the means) of post-test measures based on group comparison ANCOVA analyses considering condition, and intersection of gender with African American race/ethnicity as predictors (*controlled for pre-test scores).